# Narratives

## fMRI data for evaluating models of naturalistic language comprehension

Samuel A. Nastase
Princeton Neuroscience Institute
Princeton University
111.04—SfN 2019

@samnastase

# The *Narratives* data collection

**Sharing the "long tail" of neuroscience data**

Public, well-curated datasets have been tremendously beneficial in fields like machine learning; e.g., ImageNet, MNIST, COCO, CIFAR.

The re-use of publicly shared fMRI datasets from the INDI consortia has saved an estimated $1.7 billion in data generation costs—and this doesn't even include HCP, OpenNeuro, ABCD, or UK Biobank.

Naturalistic paradigms—e.g., movie-watching, story-listening—yield rich data with high re-use potential.

Ferguson et al., *Nat Neurosci*, 2014
Milham et al., *Nat Commun*, 2018
DuPre et al., *PsyArXiv*, 2019

# The *Narratives* data collection

"Pie Man" by Jim O'Grady
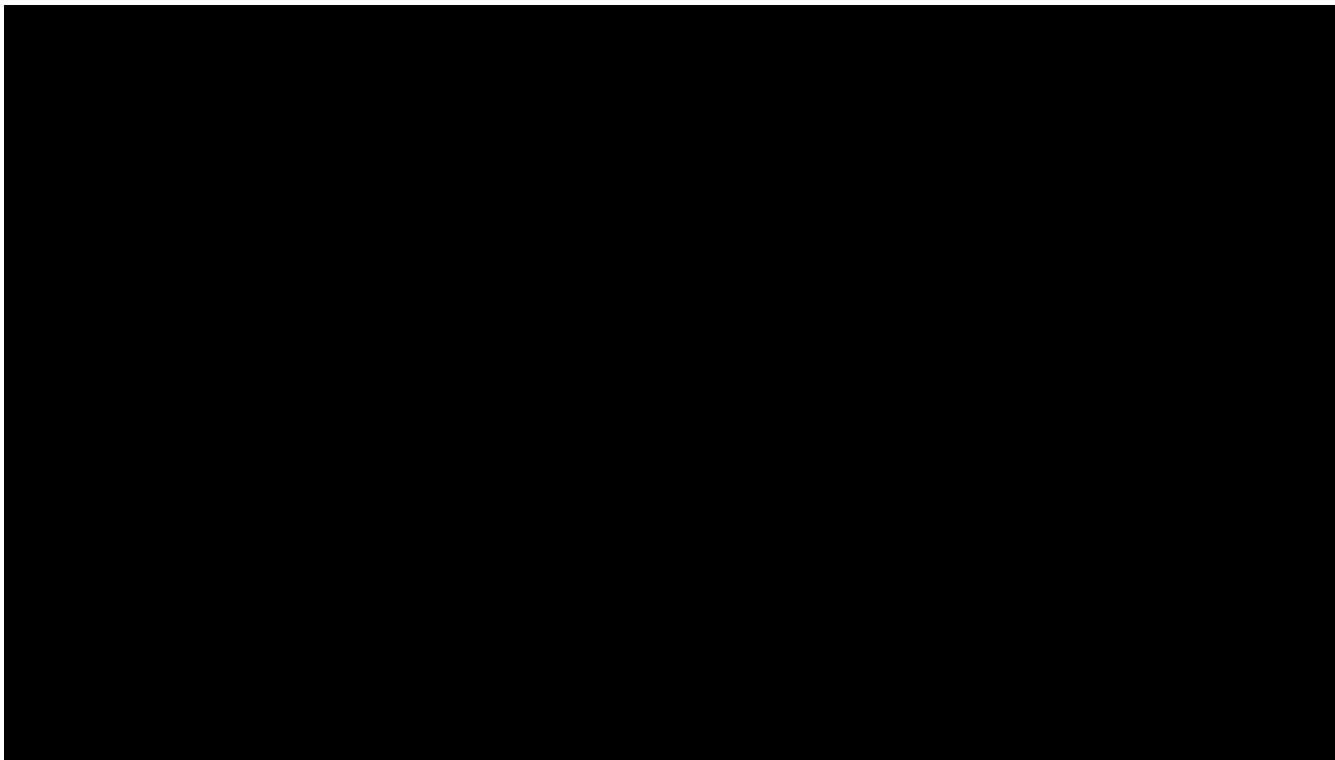
# The *Narratives* data collection

"Pie Man" by Jim O'Grady

I began my illustrious career in journalism in the Bronx where I toiled as a hard-boiled reporter for the Ram, the student newspaper at Fordham University. And one day I'm walking toward the campus center and out comes the elusive Dean McGowan, architect of a policy to replace Fordham's traditionally working- to middle-class students with wealthier, more prestigious ones. So I whip out my notebook and I go up to him and I say, "Dean McGowan, is it true that Fordham University plans to raise inflation rate, and if so, wouldn't that be a betrayal of its mi me and he says, "Listen up, punk." And right then, there's a blur in the corner of my eye which becomes this figure holding a cream pie which becom mashing a cream pie into Dean McGowan's face. And then covered with cream. So I give him a moment, and then I say, "Dean McGowan, would you care to comment on this latest attack?" And the Dean says, Fuck you." So I race back to the newsroom with my scoop

28 diverse spoken story stimuli ranging from ~3 minutes to ~56 minutes (mean ≈ 10 minutes) for a total of ~5 hours of unique stimuli.

315 unique subjects (mean age = 22.4 years, range = 18–53 years; 183 reported female) participating in 788 functional scans with accompanying anatomical data.

In total, over 350,000 TRs of story-listening fMRI data with accompanying stimuli—that is, ~150 hours or 6.2 days.

# The *Narratives* data collection

Data are organized into a machine-readable format according to the BIDS standard with exhaustive metadata derived from the original DICOMs.

Anonymized subject labels are linked across sessions and include demographic and behavioral variables including age, gender, condition, and comprehension score.

Original stimuli will be supplemented with phoneme-, word- and sentence-level time-stamped transcripts.

| Story | Duration | Subjects |
|---|---|---|
| "Pie Man" | 07:03 | 82 |
| "Tunnel Under the World" | 25:34 | 23 |
| "Lucy" | 09:02 | 16 |
| "Pretty Mouth and Green My Eyes" | 11:17 | 40 |
| "Milky Way" | 06:50 | 53 |
| "Slumlord" | 15:03 | 18 |
| "Reach for the Stars One Small Step at a Time" | 13:45 | 18 |
| "It's Not the Fall That Gets You" | 09:45 | 18 |
| "Merlin" | 15:15 | 36 |
| "Sherlock" | 18:01 | 36 |
| The "Schema" suite | 23:17 | 31 |
| "Shapes" | 07:38 | 59 |
| "The 21st Year" | 55:38 | 25 |
| "Pie Man (PNI)" | 06:57 | 46 |
| "Running from the Bronx (PNI)" | 09:21 | 47 |
| "I Knew You Were Black" | 13:21 | 46 |
| "The Man Who Forgot Ray Bradbury" | 13:57 | 46 |

# The *Narratives* data collection

```
CHANGES                        sub-023   sub-053   sub-083   sub-113   sub-143   sub-173   sub-203   sub-233   sub-263   sub-293
code                           sub-024   sub-054   sub-084   sub-114   sub-144   sub-174   sub-204   sub-234   sub-264   sub-294
dataset_description.json       sub-025   sub-055   sub-085   sub-115   sub-145   sub-175   sub-205   sub-235   sub-265   sub-295
derivatives                    sub-026   sub-056   sub-086   sub-116   sub-146   sub-176   sub-206   sub-236   sub-266   sub-296
participants.json              sub-027   sub-057   sub-087   sub-117   sub-147   sub-177   sub-207   sub-237   sub-267   sub-297
participants.tsv               sub-028   sub-058   sub-088   sub-118   sub-148   sub-178   sub-208   sub-238   sub-268   sub-298
README                         sub-029   sub-059   sub-089   sub-119   sub-149   sub-179   sub-209   sub-239   sub-269   sub-299
stimuli                        sub-030   sub-060   sub-090   sub-120   sub-150   sub-180   sub-210   sub-240   sub-270   sub-300
sub-001                        sub-031   sub-061   sub-091   sub-121   sub-151   sub-181   sub-211   sub-241   sub-271   sub-301
sub-002                        sub-032   sub-062   sub-092   sub-122   sub-152   sub-182   sub-212   sub-242   sub-272   sub-302
sub-003                        sub-033   sub-063   sub-093   sub-123   sub-153   sub-183   sub-213   sub-243   sub-273   sub-303
sub-004                        sub-034   sub-064   sub-094   sub-124   sub-154   sub-184   sub-214   sub-244   sub-274   sub-304
sub-005                        sub-035   sub-065   sub-095   sub-125   sub-155   sub-185   sub-215   sub-245   sub-275   sub-305
sub-006                        sub-036   sub-066   sub-096   sub-126   sub-156   sub-186   sub-216   sub-246   sub-276   sub-306
sub-007                        sub-037   sub-067   sub-097   sub-127   sub-157   sub-187   sub-217   sub-247   sub-277   sub-307
sub-008                        sub-038   sub-068   sub-098   sub-128   sub-158   sub-188   sub-218   sub-248   sub-278   sub-308
sub-009                        sub-039   sub-069   sub-099   sub-129   sub-159   sub-189   sub-219   sub-249   sub-279   sub-309
sub-010                        sub-040   sub-070   sub-100   sub-130   sub-160   sub-190   sub-220   sub-250   sub-280   sub-310
sub-011                        sub-041   sub-071   sub-101   sub-131   sub-161   sub-191   sub-221   sub-251   sub-281   sub-311
sub-012                        sub-042   sub-072   sub-102   sub-132   sub-162   sub-192   sub-222   sub-252   sub-282   sub-312
sub-013                        sub-043   sub-073   sub-103   sub-133   sub-163   sub-193   sub-223   sub-253   sub-283   sub-313
sub-014                        sub-044   sub-074   sub-104   sub-134   sub-164   sub-194   sub-224   sub-254   sub-284   sub-314
sub-015                        sub-045   sub-075   sub-105   sub-135   sub-165   sub-195   sub-225   sub-255   sub-285   sub-315
sub-016                        sub-046   sub-076   sub-106   sub-136   sub-166   sub-196   sub-226   sub-256   sub-286
sub-017                        sub-047   sub-077   sub-107   sub-137   sub-167   sub-197   sub-227   sub-257   sub-287
sub-018                        sub-048   sub-078   sub-108   sub-138   sub-168   sub-198   sub-228   sub-258   sub-288
sub-019                        sub-049   sub-079   sub-109   sub-139   sub-169   sub-199   sub-229   sub-259   sub-289
sub-020                        sub-050   sub-080   sub-110   sub-140   sub-170   sub-200   sub-230   sub-260   sub-290
sub-021                        sub-051   sub-081   sub-111   sub-141   sub-171   sub-201   sub-231   sub-261   sub-291
sub-022                        sub-052   sub-082   sub-112   sub-142   sub-172   sub-202   sub-232   sub-262   sub-292
```

# The *Narratives* data collection

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CHANGES | sub-023 | sub-053 | sub-083 | sub-113 | sub-143 | sub-173 | sub-203 | sub-233 | sub-263 | sub-293 |
| code | sub-024 | sub-054 | sub-084 | sub-114 | sub-144 | sub-174 | sub-204 | sub-234 | sub-264 | sub-294 |
| dataset_description.json | sub-025 | sub-055 | sub-085 | sub-115 | sub-145 | sub-175 | sub-205 | sub-235 | sub-265 | sub-295 |
| derivatives | sub-026 | sub-056 | sub-086 | sub-116 | sub-146 | sub-176 | sub-206 | sub-236 | sub-266 | sub-296 |
| participants.json | sub-027 | sub-057 | sub-087 | sub-117 | sub-147 | sub-177 | sub-207 | sub-237 | sub-267 | sub-297 |
| participants.tsv | sub-028 | sub-058 | sub-088 | sub-118 | sub-148 | sub-178 | sub-208 | sub-238 | sub-268 | sub-298 |
| README | sub-029 | sub-059 | sub-089 | sub-119 | sub-149 | sub-179 | sub-209 | sub-239 | sub-269 | sub-299 |
| stimuli | sub-030 | sub-060 | sub-090 | sub-120 | sub-150 | sub-180 | sub-210 | sub-240 | sub-270 | sub-300 |
| sub-001 | sub-031 | sub-061 | sub-091 | sub-121 | sub-151 | sub-181 | sub-211 | sub-241 | sub-271 | sub-301 |
| sub-002 | sub-032 | sub-062 | sub-092 | sub-122 | sub-152 | sub-182 | sub-212 | sub-242 | sub-272 | sub-302 |
| sub-003 | sub-033 | sub-063 | sub-093 | sub-123 | sub-153 | sub-183 | sub-213 | sub-243 | sub-273 | sub-303 |
| sub-004 | sub-034 | sub-064 | sub-094 | sub-124 | sub-154 | sub-184 | sub-214 | sub-244 | sub-274 | sub-304 |
| sub-005 | sub-035 | sub-065 | sub-095 | sub-125 | sub-155 | sub-185 | sub-215 | sub-245 | sub-275 | sub-305 |
| sub-006 | sub-036 | sub-066 | sub-096 | sub-126 | sub-156 | sub-186 | sub-216 | sub-246 | sub-276 | sub-306 |
| sub-007 | sub-037 | sub-067 | sub-097 | sub-127 | sub-157 | sub-187 | sub-217 | sub-247 | sub-277 | sub-307 |
| sub-008 | sub-038 | sub-068 | sub-098 | sub-128 | sub-158 | sub-188 | sub-218 | sub-248 | sub-278 | sub-308 |
| sub-009 | sub-039 | sub-069 | sub-099 | sub-129 | sub-159 | sub-189 | sub-219 | sub-249 | sub-279 | sub-309 |
| sub-010 | sub-040 | sub-070 | sub-100 | sub-130 | sub-160 | sub-190 | sub-220 | sub-250 | sub-280 | sub-310 |
| sub-011 | sub-041 | sub-071 | sub-101 | sub-131 | sub-161 | sub-191 | sub-221 | sub-251 | sub-281 | sub-311 |
| sub-012 | sub-042 | sub-072 | sub-102 | sub-132 | sub-162 | sub-192 | sub-222 | sub-252 | sub-282 | sub-312 |
| sub-013 | sub-043 | sub-073 | sub-103 | sub-133 | sub-163 | sub-193 | sub-223 | sub-253 | sub-283 | sub-313 |
| sub-014 | sub-044 | sub-074 | sub-104 | sub-134 | sub-164 | sub-194 | sub-224 | sub-254 | sub-284 | sub-314 |
| sub-015 | sub-045 | sub-075 | sub-105 | sub-135 | sub-165 | sub-195 | sub-225 | sub-255 | sub-285 | sub-315 |
| sub-016 | sub-046 | sub-076 | sub-106 | sub-136 | sub-166 | sub-196 | sub-226 | | | |
| sub-017 | sub-047 | sub-077 | sub-107 | sub-137 | sub-167 | sub-197 | sub-227 | | | sub-287 |
| sub-018 | sub-048 | sub-078 | sub-108 | sub-138 | sub-168 | sub-198 | sub-228 | | | |
| sub-019 | sub-049 | sub-079 | sub-109 | sub-139 | sub-169 | sub-199 | sub-229 | | | |
| sub-020 | sub-050 | sub-080 | sub-110 | sub-140 | sub-170 | sub-200 | sub-230 | | | |
| sub-021 | sub-051 | sub-081 | sub-111 | sub-141 | sub-171 | sub-201 | sub-231 | | | |
| sub-022 | sub-052 | sub-082 | sub-112 | sub-142 | sub-172 | sub-202 | sub-232 | sub-262 | | sub-292 |

Gorgolewski et al., *Sci Data*, 2016

BIDS
BRAIN IMAGING DATA STRUCTURE

# The *Narratives* data collection

`participants.tsv`

| participant_id | age | sex | task | condition | comprehension |
|---|---|---|---|---|---|
| sub-001 | 22,23 | F,F | pieman,tunnel | n/a | n/a |
| sub-002 | 20 | M | pieman | n/a | n/a |
| sub-003 | n/a | F | pieman | n/a | n/a |
| sub-004 | 24,24 | M,M | pieman,tunnel | n/a | n/a |
| ⋮ | | | | | |
| sub-101 | 22,22 | M,M | prettymouth,milkyway | paranoia,original | 0.926,0.821 |
| sub-102 | 27,27 | M,M | prettymouth,milkyway | paranoia,vodka | 0.963,0.857 |
| sub-103 | 19,19 | F,F | prettymouth,milkyway | paranoia,vodka | 0.889,0.821 |
| sub-104 | 18,18 | F,F | prettymouth,milkyway | paranoia,original | 0.963,1.0 |
| ⋮ | | | | | |

**BIDS**
BRAIN IMAGING DATA STRUCTURE

# The *Narratives* data collection

**participants.tsv**

| participant_id | age | sex | task | condition | comprehension |
|---|---|---|---|---|---|
| sub-001 | 22,23 | F,F | pieman,tunnel | n/a | n/a |
| sub-002 | 20 | M | pieman | n/a | n/a |
| sub-003 | n/a | F | pieman | n/a | n/a |
| sub-004 | 24,24 | M,M | pieman,tunnel | n/a | n/a |
| ⋮ | | | | | |
| sub-101 | 22,22 | M,M | prettymouth,milkyway | paranoia,original | 0.926,0.821 |
| sub-102 | 27,27 | M,M | prettymouth,milkyway | paranoia,vodka | 0.963,0.857 |
| sub-103 | 19,19 | F,F | prettymouth,milkyway | paranoia,vodka | 0.889,0.821 |
| sub-104 | 18,18 | F,F | prettymouth,milkyway | paranoia,original | 0.963,1.0 |
| ⋮ | | | | | |

**$ ls sub-001/func/**

```
sub-001_task-pieman_bold.json        sub-001_task-tunnel_bold.json
sub-001_task-pieman_bold.nii.gz      sub-001_task-tunnel_bold.nii.gz
sub-001_task-pieman_events.tsv       sub-001_task-tunnel_events.tsv
```

**BIDS**
BRAIN IMAGING DATA STRUCTURE

**sub-001_task-pieman_bold.json**

```
{
  "CogAtlasID":
"https://www.cognitiveatlas.org/task/id/trm_4c8991fadfe01",
  "DeviceSerialNumber": "45031",
  "DwellTime": 5.4e-06,
  "EchoTime": 0.028,
  "EffectiveEchoSpacing": 0.000390001,
  "FlipAngle": 64,
  "InstitutionAddress": "Washington Rd, Building 25,
Princeton, NJ 08540, USA",
  "InstitutionName": "Princeton University",
  "InstitutionalDepartmentName": "Princeton Neuroscience
Institute",
  "MagneticFieldStrength": 3,
  "Manufacturer": "Siemens",
  "ManufacturersModelName": "Skyra",
  "NumberOfVolumesDiscardedByScanner": 3,
  "ParallelReductionFactorInPlane": 2,
  "ParallelReductionType": "GRAPPA",
  "PartialFourier": 1,
  "PhaseEncodingDirection": "j-",
  "PixelBandwidth": 1445,
  "PulseSequenceDetails": "%SiemensSeq%_ep2d_bold",
  "PulseSequenceType": "Gradient Echo EPI",
  "ReceiveCoilActiveElements": "HE1-4",
  "ReceiveCoilName": "HeadNeck_20",
  "RepetitionTime": 1.5,
  "ScanOptions": "FS",
  "ScanningSequence": "EP",
  "SequenceName": "_epfid2d1_64",
  "SequenceVariant": "SK",
  "SliceTiming": [
    0,
    0.7775,
    0.055,
    0.8325,
    0.11,
    0.89,
    0.1675,
    0.945,
    0.2225,
    1,
    0.2775,
    1.055,
    0.3325,
    1.11,
    0.39,
    1.1675,
    0.445,
    1.2225,
    0.5,
    1.2775,
    0.555,
    1.3325,
    0.61,
    1.3875,
    0.6675,
    1.445,
    0.7225
  ],
  "SoftwareVersions": "syngo_MR_D11",
  "StationName": "AWP45031",
  "TaskDescription": "Passively listened to audio story 'pieman'",
  "TaskName": "pieman",
  "TotalReadoutTime": 0.0245701
}
```

COBIDAS Initiative
Nichols et al., *Nat Neurosci*, 2017

# The *Narratives* data collection

**participants.tsv**

| participant_id | age | sex | task | condition | comprehension |
|---|---|---|---|---|---|
| sub-001 | 22,23 | F,F | pieman,tunnel | n/a | n/a |
| sub-002 | 20 | M | pieman | n/a | n/a |
| sub-003 | n/a | F | pieman | n/a | n/a |
| sub-004 | 24,24 | M,M | pieman,tunnel | n/a | n/a |
| ⋮ | | | | | |
| sub-101 | 22,22 | M,M | prettymouth,milkyway | paranoia,original | 0.926,0.821 |
| sub-102 | 27,27 | M,M | prettymouth,milkyway | paranoia,vodka | 0.963,0.857 |
| sub-103 | 19,19 | F,F | prettymouth,milkyway | paranoia,vodka | 0.889,0.821 |
| sub-104 | 18,18 | F,F | prettymouth,milkyway | paranoia,original | 0.963,1.0 |
| ⋮ | | | | | |

**sub-001_task-pieman_events.tsv**

| onset | duration | trial_type | stim_file |
|---|---|---|---|
| 0.0 | 13.0 | music | pieman_audio.wav |
| 15.0 | 422.0 | story | pieman_audio.wav |

**BIDS**
BRAIN IMAGING DATA STRUCTURE

# The *Narratives* data collection

**participants.tsv**

| participant_id | age | sex | task | condition | comprehension |
|---|---|---|---|---|---|
| sub-001 | 22,23 | F,F | pieman,tunnel | n/a | n/a |
| sub-002 | 20 | M | pieman | n/a | n/a |
| sub-003 | n/a | F | pieman | n/a | n/a |
| sub-004 | 24,24 | M,M | pieman,tunnel | n/a | n/a |
| ⋮ | | | | | |
| sub-101 | 22,22 | M,M | prettymouth,milkyway | paranoia,original | 0.926,0.821 |
| sub-102 | 27,27 | M,M | prettymouth,milkyway | paranoia,vodka | 0.963,0.857 |
| sub-103 | 19,19 | F,F | prettymouth,milkyway | paranoia,vodka | 0.889,0.821 |
| sub-104 | 18,18 | F,F | prettymouth,milkyway | paranoia,original | 0.963,1.0 |
| ⋮ | | | | | |

**sub-201_task-schema_run-1_events.tsv**

| onset | duration | trial_type | stim_file |
|---|---|---|---|
| 5.0 | 185.0 | movie | n/a |
| 196.57 | 179.0 | story | vinny_audio.wav |
| 380.86 | 190.0 | movie | n/a |
| 576.13 | 187.0 | story | upintheair_audio.wav |

**BIDS**
BRAIN IMAGING DATA STRUCTURE

# Data provenance

All stages of data processing version-controlled using DataLad, allowing researchers to analyze raw or processed data.

Full history of data collection is visible via `git log`



Code: https://github.com/snastase/narratives
http://datalad.org — discover
http://github.com/datalad/datalad — contribute
http://handbook.datalad.org — learn
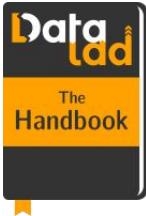
Halchenko et al., 2019

```
commit ea46e14fd0dc949acd654947e136fe2c354ef780 (HEAD -> master)
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Thu Oct 17 00:40:28 2019 -0400

    Updated some missing condition labels (prettymouth)

commit a46de0940b96057d7b0297a48ab2f7a8d08a1c1e
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Tue Oct 15 18:06:08 2019 -0400

    Added audio files to stimuli subdataset

commit a1019e628f4e6a40730941d3ec467e15fbdc8560
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Tue Oct 15 15:38:40 2019 -0400

    Added subdataset for stimuli

commit e93be6b469e7e664abc2dd6c0790002a9d0171c0
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Mon Oct 14 18:45:51 2019 -0400

    Added subdatasets for code and derivatives

commit bee66631b402f4a048244c8c3cb1ebda2ad33c3a
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Mon Oct 14 18:35:07 2019 -0400

    Populated raw NIfTIs and metadata (BIDS valid)

commit daf3b89009b9bde93fabffdb2f11d2788ea44688
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Mon Oct 14 17:39:19 2019 -0400

    Instruct annex to add text files to Git

commit bc03620e50865195228e46c1ab3da3a2fa87e6cf
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Mon Oct 14 17:39:17 2019 -0400

    [DATALAD] new dataset
```

# Derivatives

Machine-readable file organization and rich metadata allow for automated processing using containerized "BIDS Apps", e.g.:

**MRIQC**: advancing the automatic prediction of image quality in MRI from unseen sites

**fMRIPrep**: a robust preprocessing pipeline for functional MRI

Gorgolewski et al., *PLOS Comp Biol*, 2017
Esteban et al., *PLOS One*, 2017
Esteban et al., *Sci Data*, 2019

# Derivatives—MRIQC
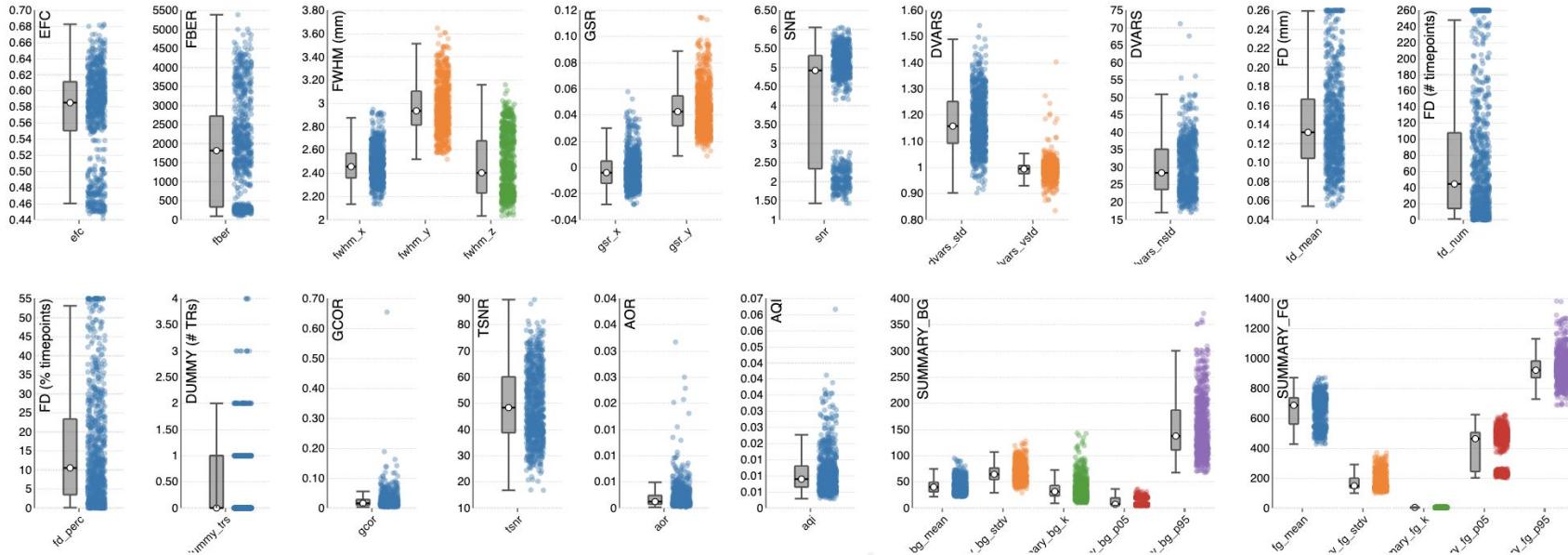
# Derivatives—MRIQC

Image quality metrics (IQMs) include:

**Temporal signal-to-noise ratio (TSNR)**: median = 48.5

**Head motion (framewise displacement; FD)**: median = 0.13 mm

**Spatial smoothness (FWHM)**: median ≈ 2.4 mm
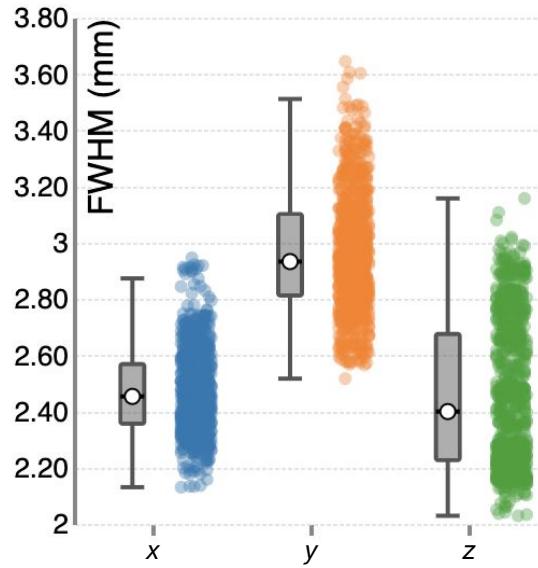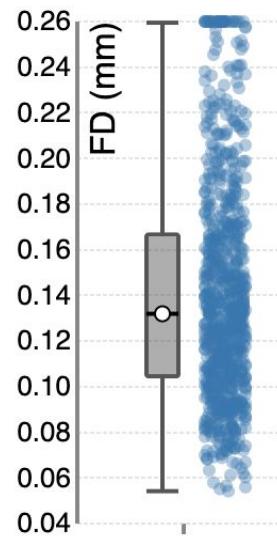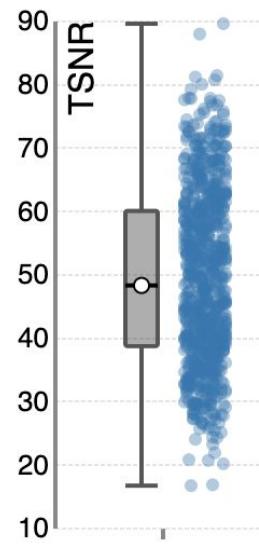
# Derivatives—fMRIPrep

All MRI data were preprocessed using fMRIPrep, a state-of-the-art pipeline for BIDS-formatted data.

Motion correction, slice-timing correction

Nonlinear spatial normalization to MNI space using ANTs

Surface reconstruction and spatial normalization to FreeSurfer's fsaverage

"Fieldmap-less" susceptibility distortion correction
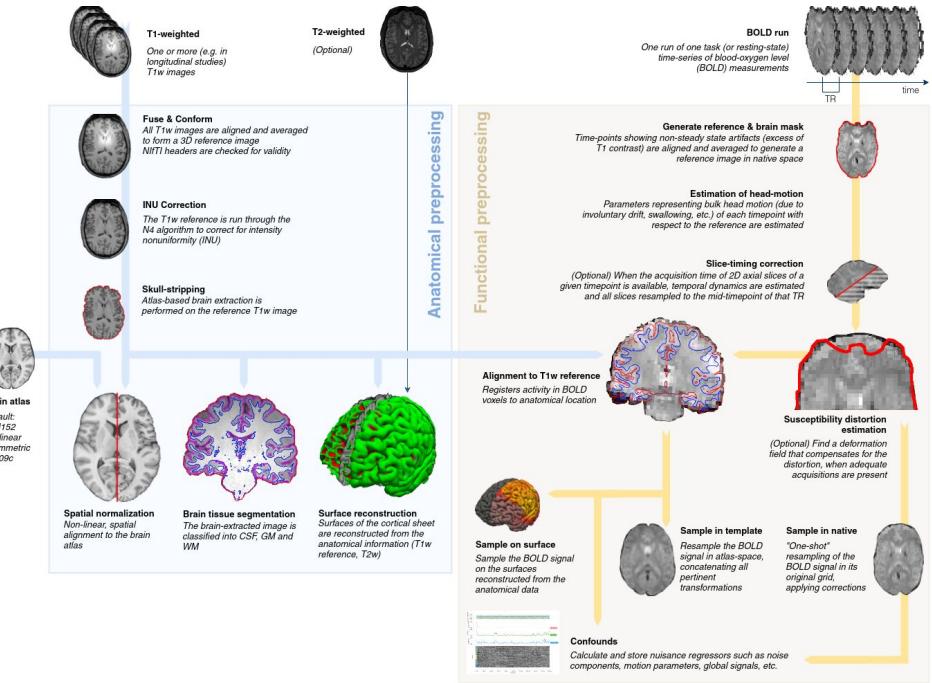
Regression with confound regressors including using AFNI's 3dTproject: polynomial detrending, high-pass filtering, head motion parameters, framewise displacement, and aCompCor principal component time series from CSF



Esteban et al., *Sci Data*, 2019
Cox, *NeuroImage*, 2012

# Stimuli

All stimuli and transcripts are freely available online and provided for non-commercial scholarly research under *fair use* and *fair dealing* provisions.

We use the **Gentle** forced-alignment software to extract phoneme-, word-, and sentence-level time-stamped transcripts.

I began my illustrious career in journalism in the Bronx where I toiled as a hard-boiled reporter for the Ram, the student newspaper at Fordham University. And one day I'm walking toward the campus center and out comes the elusive Dean McGowan, architect of a policy to replace Fordham's traditionally working- to middle-class students with wealthier, more prestigious ones. So I whip out my notebook and I go up to him and I say, "Dean McGowan, is it true that Fordham University plans to raise tuition substantially above the inflation rate, and if so, wouldn't that be a betrayal of its mission?" And he stops and looks at me and he says, "Listen up, punk." And right then, there's a blur in the corner of my eye

DuPre et al., *PsyArXiv*, 2019
Ochshorn & Hawkins, 2019

# Intersubject correlation

**Measuring shared responses**

Whole-brain leave-one-subject-out intersubject correlation analysis.



pieman

Intersubject correlation

0.4

0.1

Nastase et al., *Soc Cog Affect Neurosci*, 2019
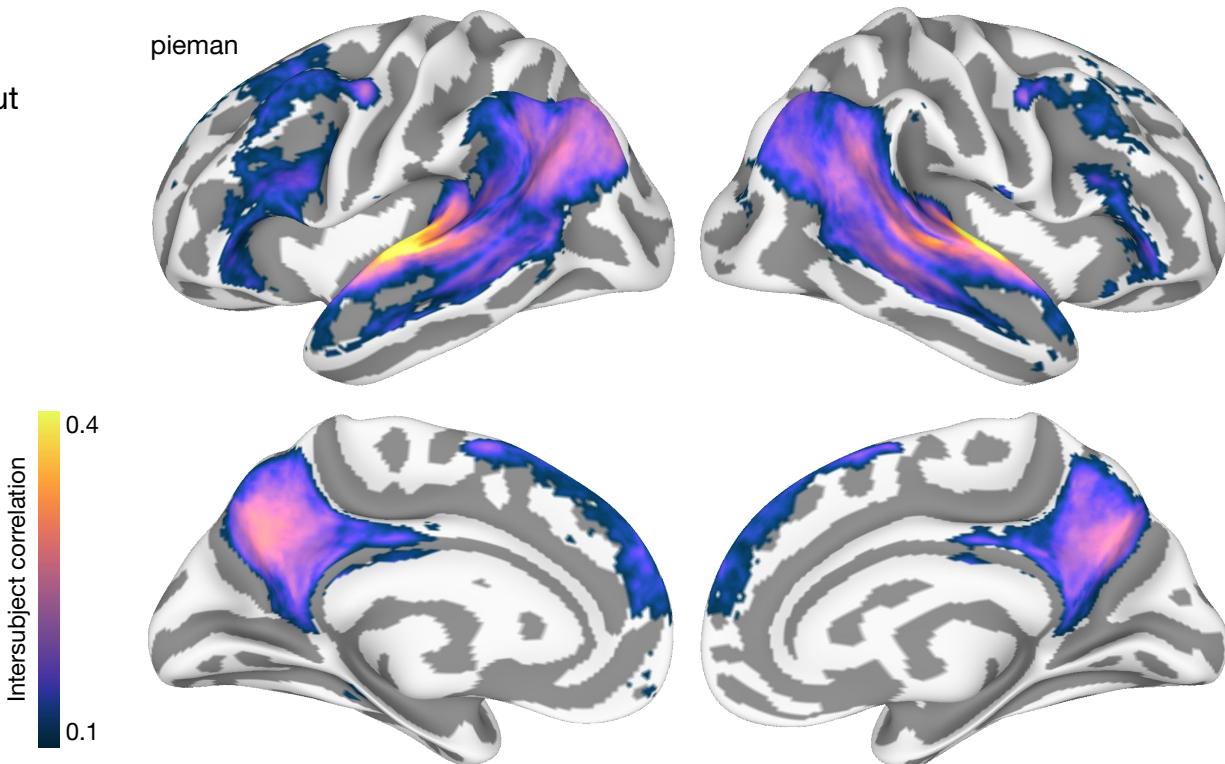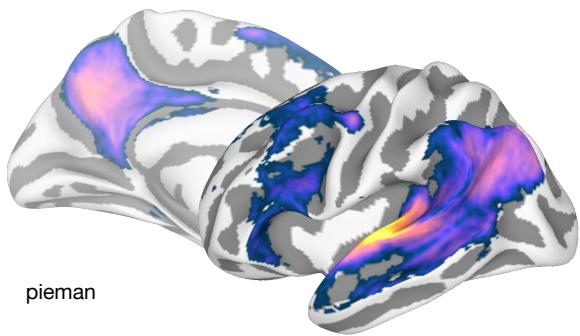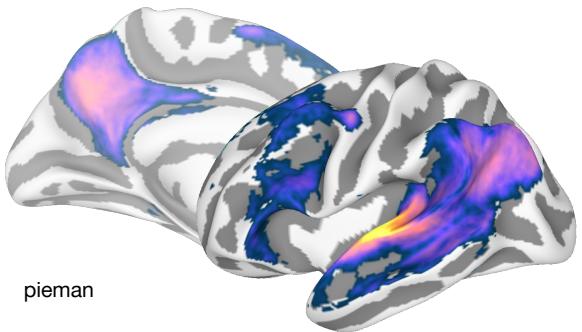
pieman

pieman
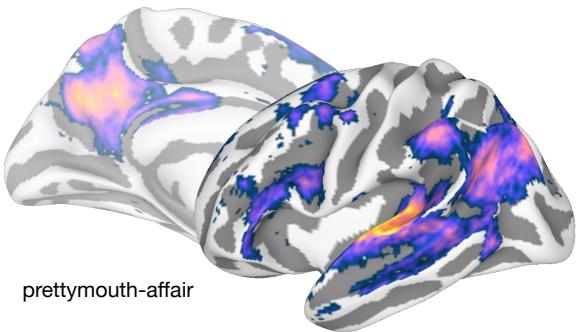
prettymouth-affair
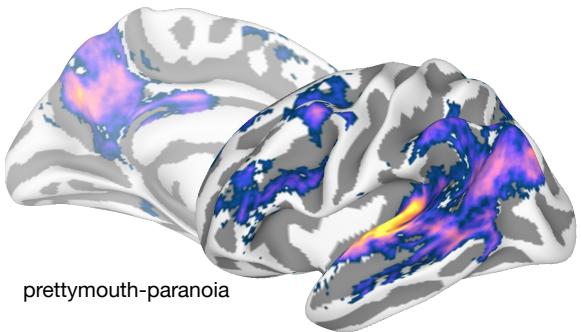
prettymouth-paranoia
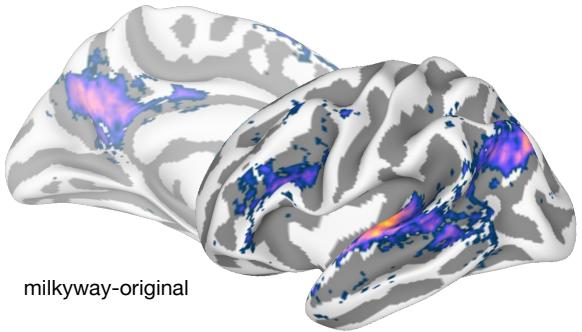
milkyway-original
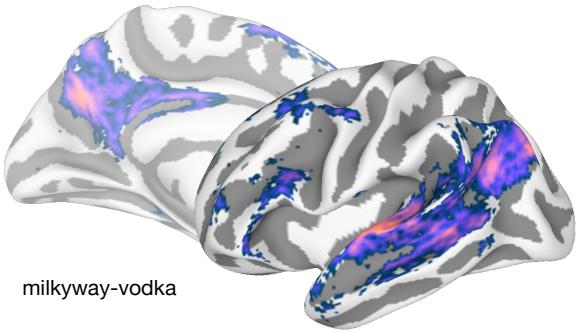
milkyway-vodka
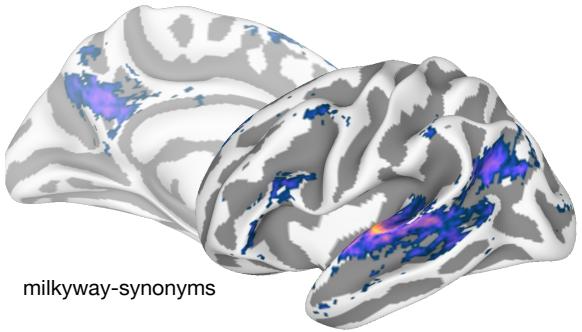
milkyway-synonyms
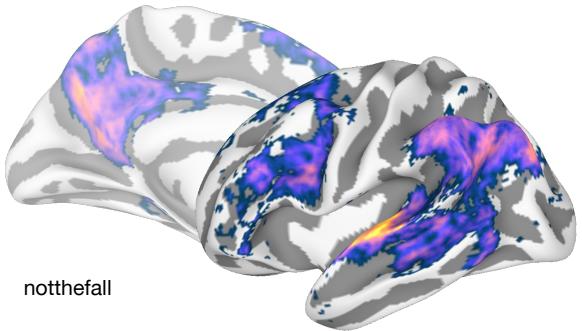
notthefall
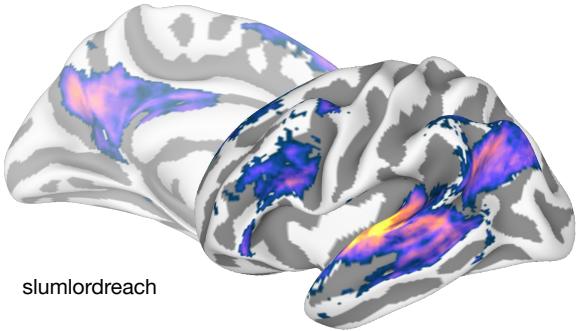
slumlordreach
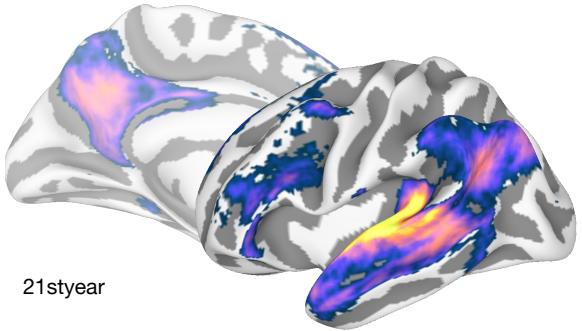
21styear

# Going forward

**The "Narratives" collection will continue to evolve…**

Time-stamped phoneme-, word-, and sentence-level transcripts coming soon!

Standardized, authoritative preprocessed fMRI derivatives (e.g., via fMRIPrep)

Growing collection of low- and high-level stimulus features (e.g., via NeuroScout)

de la Vega et al., 2019

# Going forward

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS  Check for updates

## The revolution will not be controlled: natural stimuli in speech neuroscience

Liberty S. Hamilton[a,b] and Alexander G. Huth[c,d]

[a]Communication Sciences & Disorders, Moody College of Communication, The University of Texas at Austin, Austin, USA; [b]Department of Neurology, Dell Medical School, The University of Texas at Austin, Austin, USA; [c]Department of Neuroscience, The University of Texas at Austin, Austin, USA; [d]Department of Computer Science, The University of Texas at Austin, Austin, USA

Hamilton & Huth, *Lang Cogn Neurosci*, 2018
Nastase et al., *bioRxiv*, 2019

# Going forward

**Research questions for naturalistic speech… what can I do with these data?**

Intersubject correlation (ISC) analysis and intersubject network estimation (ISFC)

Hasson et al., *Science*, 2004
Simony et al., *Nat Commun*, 2016

Hyperalignment and shared response models (SRM)

Haxby et al., *Neuron*, 2011
Chen et al., *NeurIPS*, 2015

Forwarding encoding models and inverted coding models

Santoro et al., *PLOS Comp Biol*, 2014
Huth et al., *Nature*, 2016

# Going forward



Nastase et al., *bioRxiv*, 2019

# Going forward



Nastase et al., *bioRxiv*, 2019
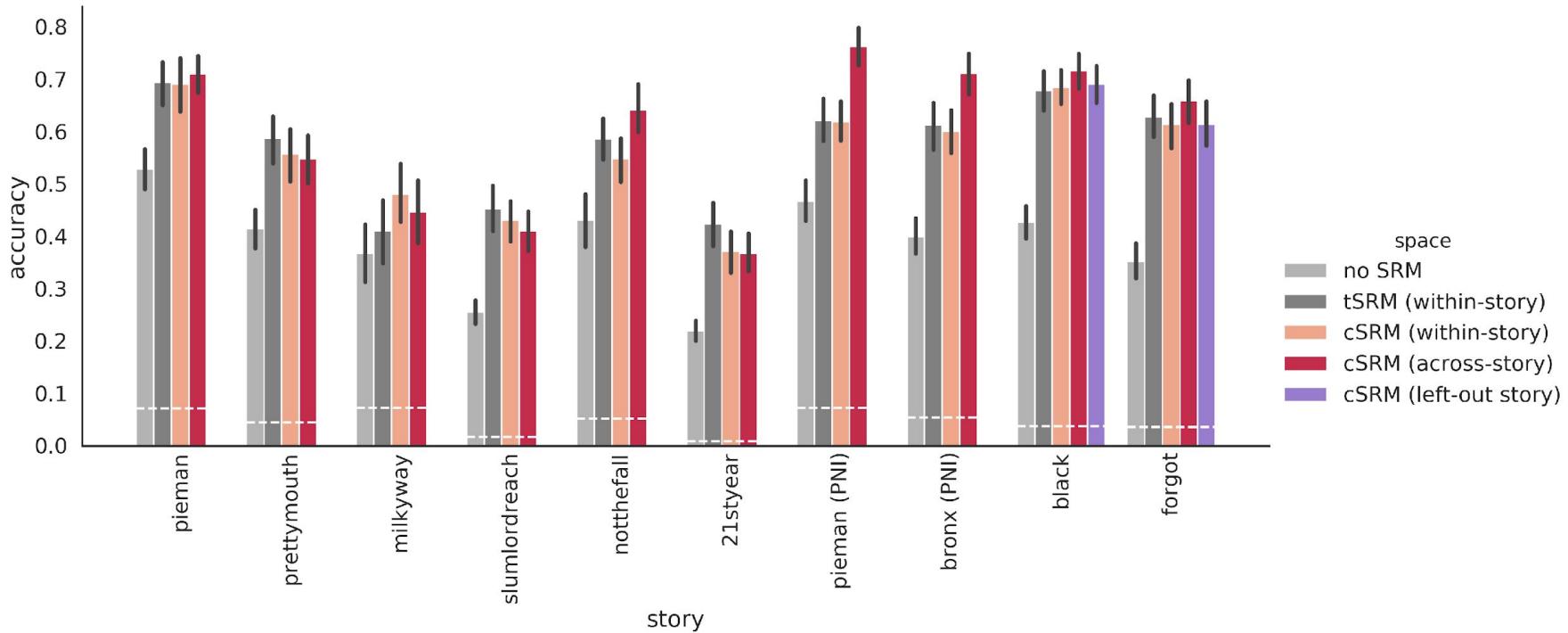
# The *Narratives* data collection

**An invitation to all "research parasites"**

28 diverse spoken story stimuli ranging from ~3 minutes to ~56 minutes (mean ≈ 10 minutes) for a total of ~5 hours of unique stimuli.

315 unique subjects (mean age = 22.4 years, range = 18–53 years; 183 reported female) participating in 788 functional scans with accompanying anatomical data.

In total, over 350,000 TRs of story-listening fMRI data with accompanying stimuli—that is, ~150 hours or 6.2 days.

**NIH** National Institute of Mental Health

ReproNim

Samuel A. Nastase, Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen, Christopher J. Honey, Yaara Yeshurun, Mor Regev, Mai Nguyen, Claire H. C. Chang, Christopher Baldassano, Olga Lositsky, Erez Simony, Michael A. Chow, Yuan Chang Leong, Paula P. Brooks, Emily Micciche, Gina Choe, Ariel Goldstein, Yaroslav O. Halchenko, Kenneth A. Norman, Uri Hasson