Conducting speech perception experiments online: Some tools, success, and challenges

Rachel M. Theodore, Ph.D.
Department of Speech, Language, and Hearing Sciences
Connecticut Institute for the Brain and Cognitive Sciences
University of Connecticut

NIDCD NSF UCONN UNIVERSITY OF CONNECTICUT slaplab.uconn.edu

@UConnNBL

osf.io/8krg3/

For a more in-depth tutorial, visit:

https://tinyurl.com/Online-Perception-Tutorial

Tools

Prolific

- https://www.prolific.co
- Online participant pool
- Large, diverse sample
- Researchers can apply filters to determine who has access to a study
- Prolific is built to provide high quality data and promote ethical treatment of participants
- Researchers are charged a fee based on payment to participant

Gorilla

- https://gorilla.sc
- Software to build experiments
- Server to host online studies
- If you can dream it, Gorilla can build it
- Supports collaboration, open materials, version control, data management
- Free to build; charged a "token" to download data for each subject

Headphone screen

- Woods et al. (2017), *Attention, Perception, & Psychophysics*
- Six-trial screen; 5 correct responses == "Pass"
- Task is choosing which of three tones is quietest
- Tone sequences manipulate phase across stereo channels
- Vetting data show reasonable sensitivity for detecting headphone use, in my opinion

/mcdermottLab

General procedures to facilitate success

- Design the experiment to be only as long as needed, we aim for ≤ 20 minutes
  - Data quality is better for shorter tasks
  - Subjects make their own break(s) in longer tasks
- Convert sound files to MP3 and image files to JPEG; *provide clear instructions regarding auto-play and headphone requirements*
- Give people at least two chances to pass the headphone screen, with a friendly reminder of headphone requirement between screens
- Pay well; we compensate at $10/hour

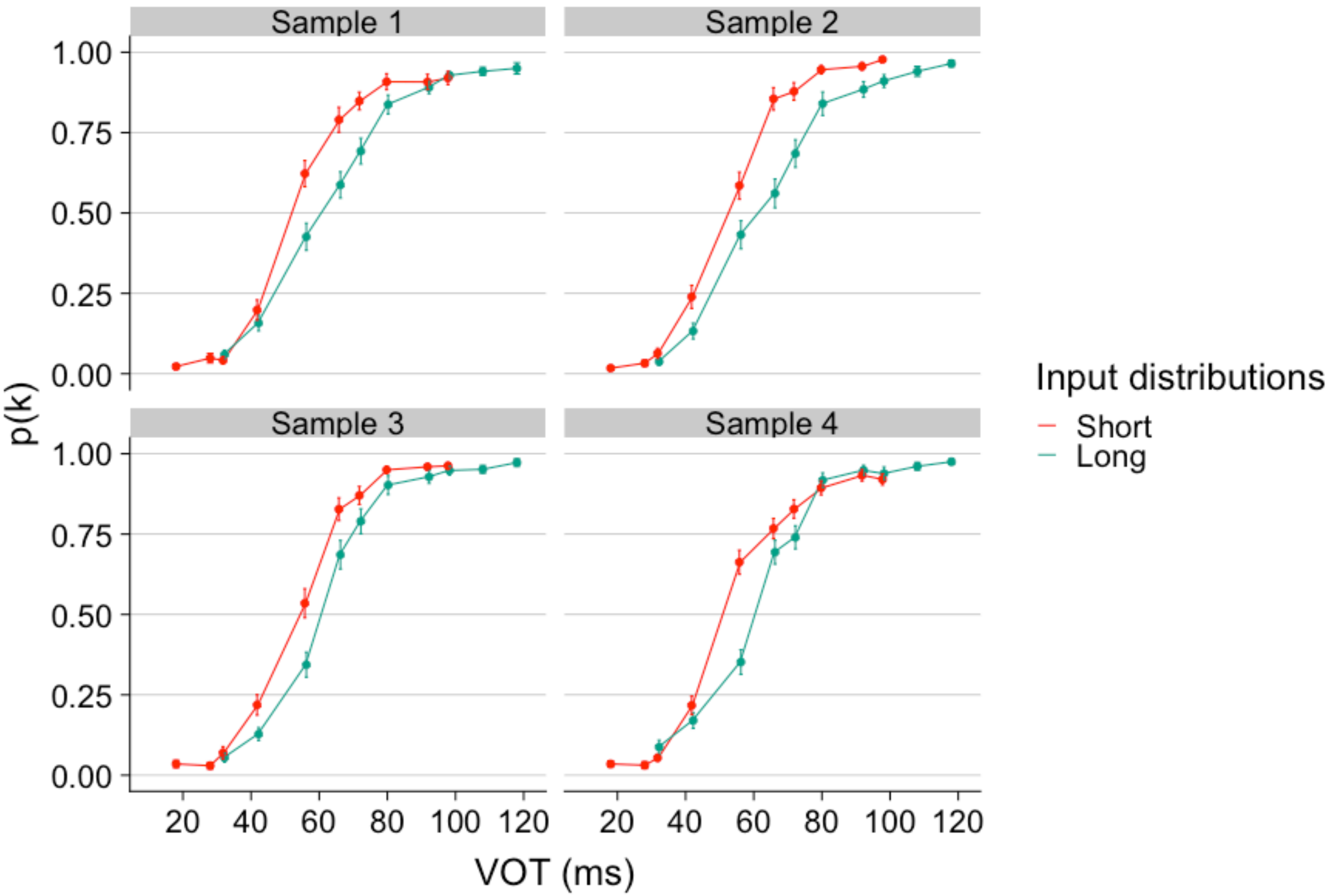
Success 1: Categorical perception/distributional learning

Block 1

- 152 trials of phonetic ID for tokens drawn from a VOT continuum to form either short or long VOT input distributions

Block 2

- 152 trials of phonetic ID for tokens drawn from a VOT continuum to form either short or long VOT input distributions



To achieve sample (n = 320), we excluded n = 52 due to failure to perform the task and n = 27 due to failure to pass headphone screen; attrition = 20%.

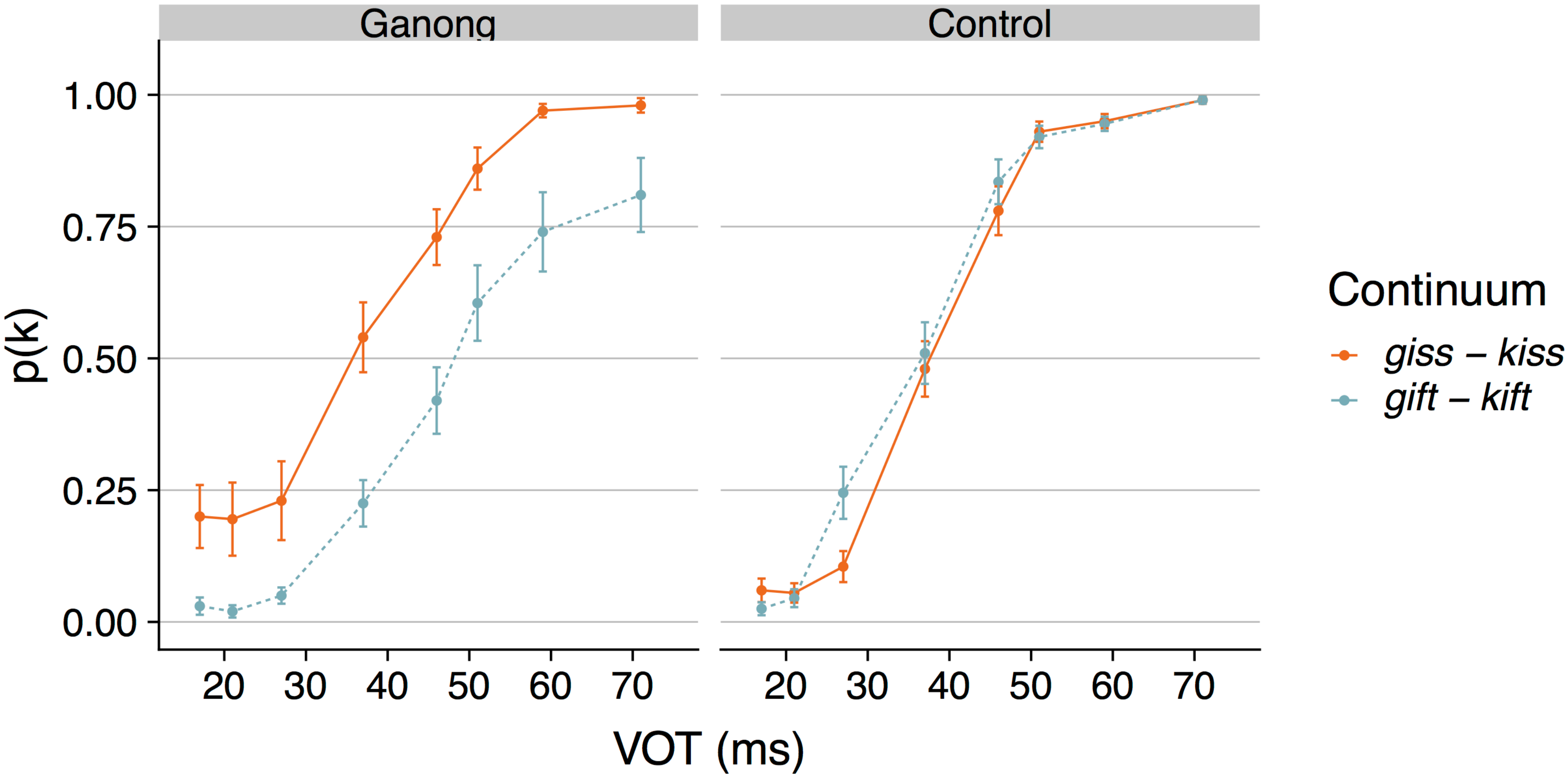
Success 2: Ganong effect

Block: Ganong

- 160 trials of phonetic ID for *gift-kift* and *giss-kiss* VOT continua

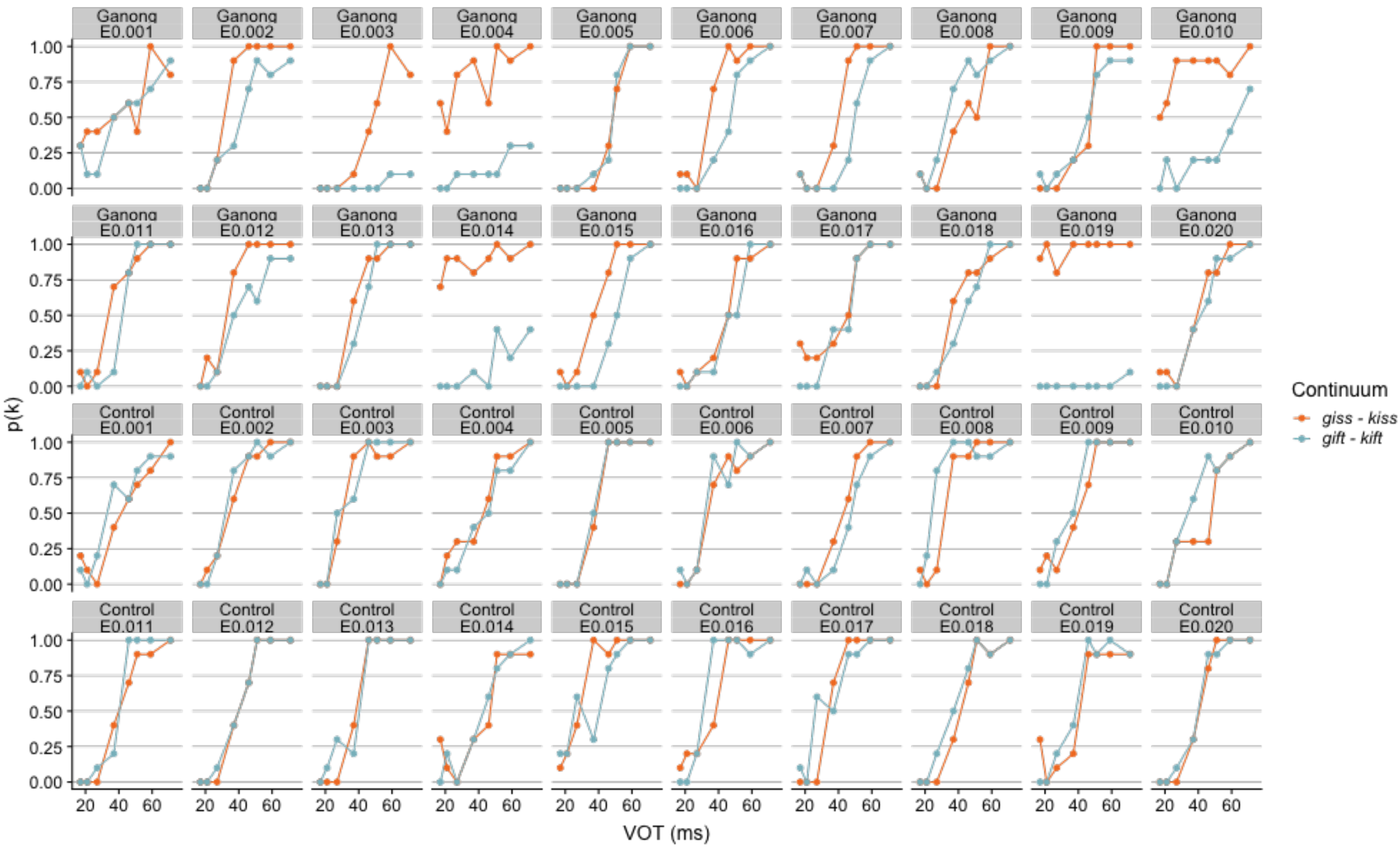
Block: Control

- 160 trials of phonetic ID for the same VOT continua, excising disambiguating lexical information



To achieve sample (n = 20), we excluded n = 0 due to failure to perform the task and n = 3 due to failure to pass headphone screen; attrition = 13%.

Success 2: Ganong effect



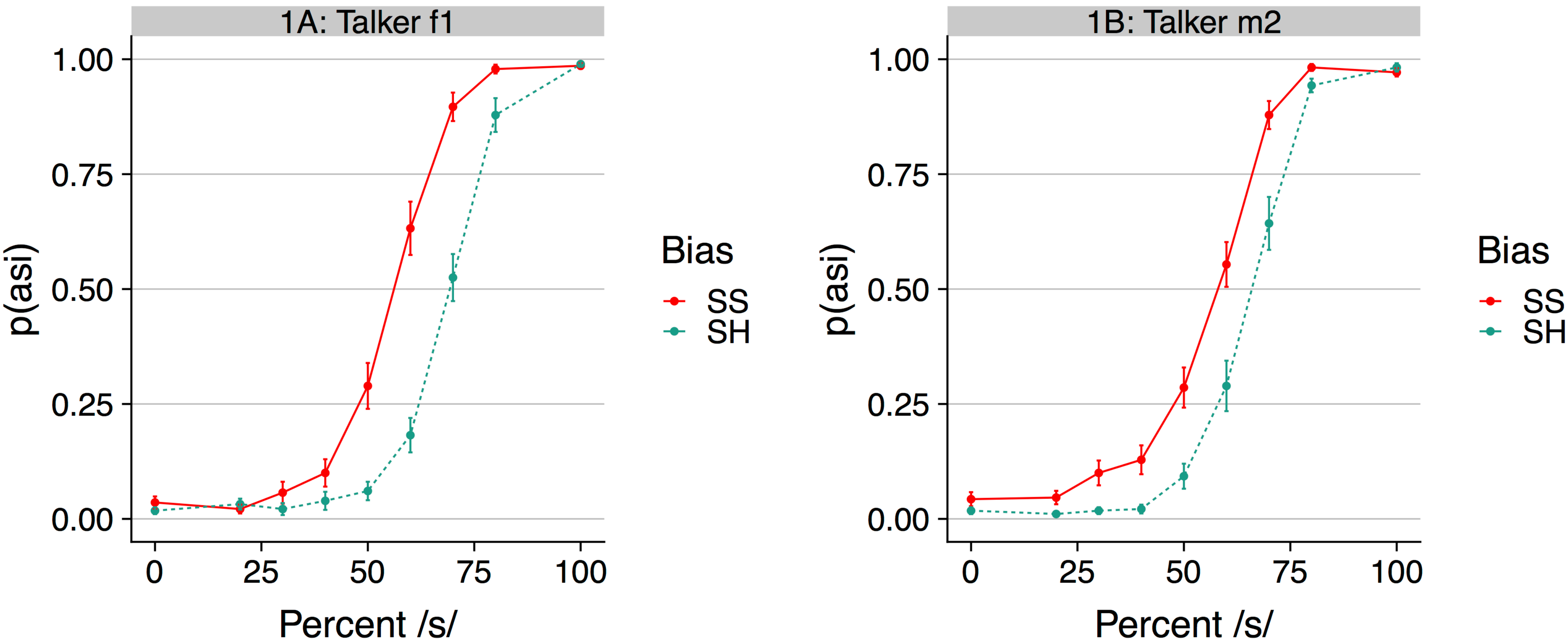
Success 3: Lexically guided perceptual learning

Block: Exposure

- 200 trials of a lexical decision task for word and nonword stimuli; critical ambiguous productions embedded in either /s/ or /ʃ/ biasing contexts

Block: Test

- 72 trials of phonetic ID for tokens drawn from an /asi/-/aʃi/ continuum



To achieve sample (n = 560), we excluded n = 32 due to failure to perform the task and n = 112 due to failure to pass headphone screen; attrition = 20%.

Success 4: Perceptual learning for vocoded speech

Block: Pre-test

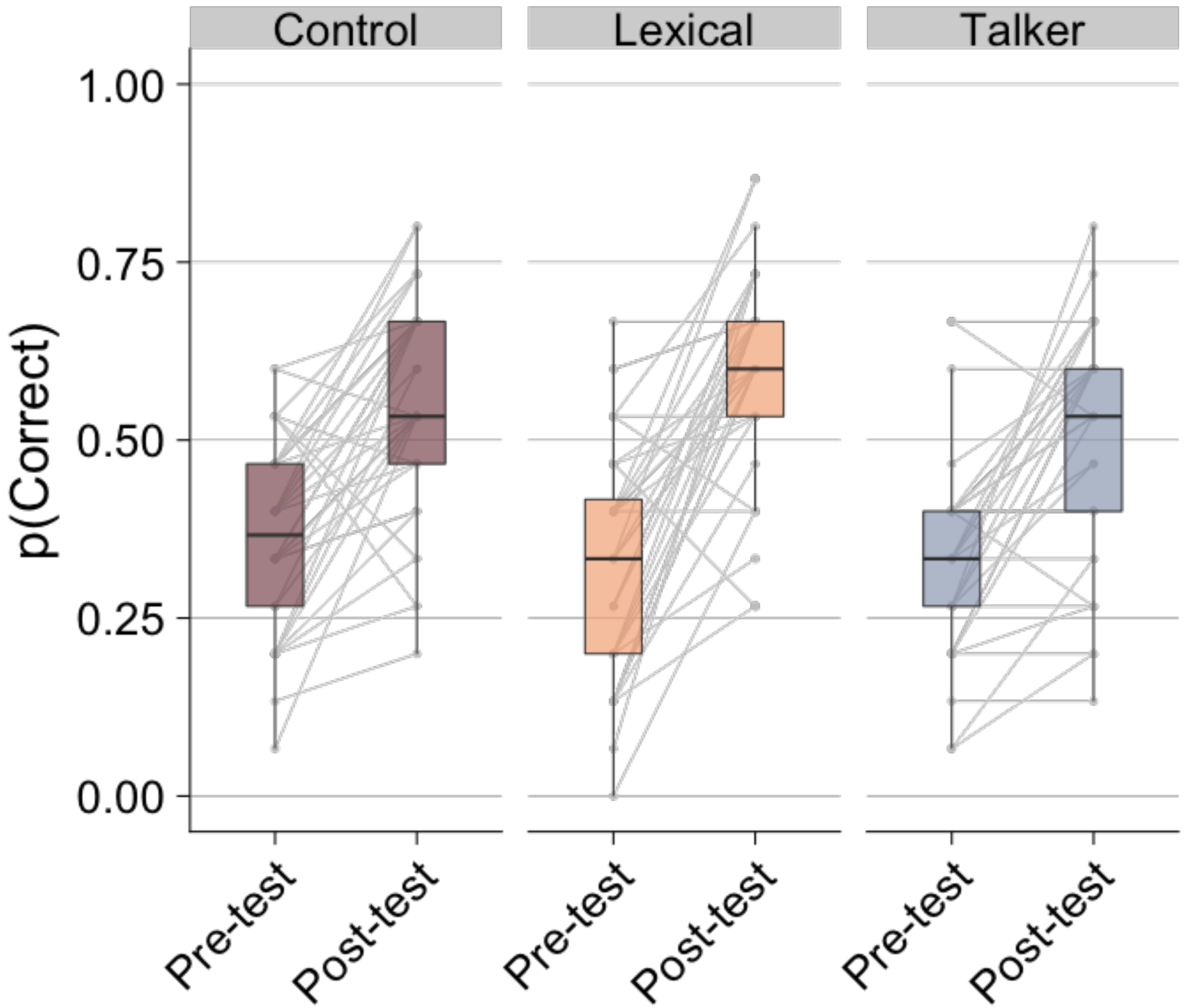
- 30 trials of a transcription task for vocoded sentences w/o feedback

Block: Training

- 150 trials with vocoded sentences
  - *Control*: Sentence transcription w/o feedback
  - *Lexical:* Sentence transcription w/ feedback
  - *Talker*: Talker ID w/ feedback

Block: Post-test

- 30 trials of a transcription task for vocoded sentences w/o feedback

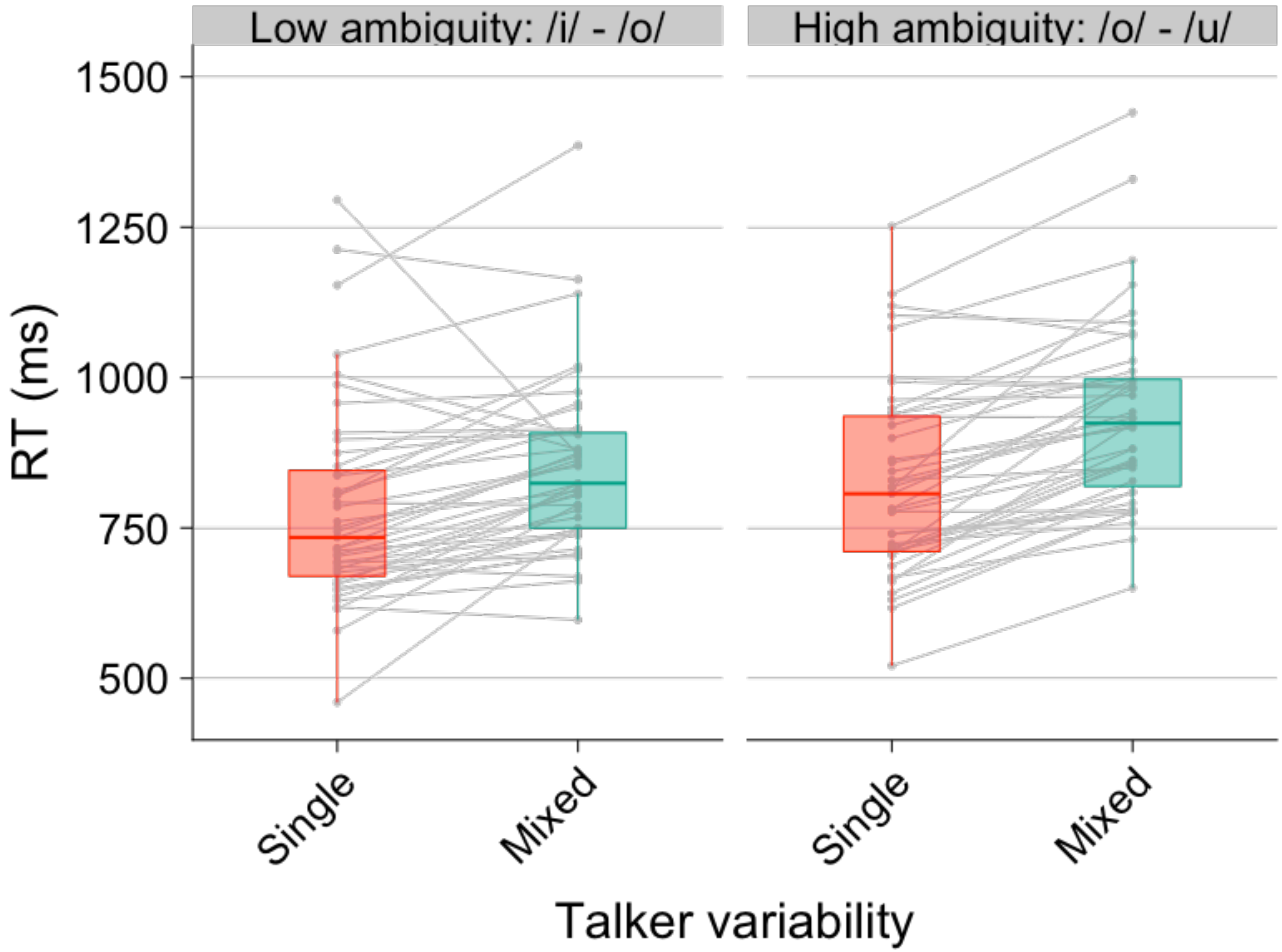


To achieve sample (n = 108), we excluded n = 2 due to failure to perform the task and n = 12 due to failure to pass headphone screen; attrition = 11%.

Success 5: Talker normalization/phonemic ambiguity

Blocks: Word ID

- 160 trials of speeded word ID distributed across four separate blocks, formed by crossing phonemic ambiguity and talker variability



To achieve current sample (n = 44), we excluded n = 4 due to failure to perform the task and n = 10 due to failure to pass headphone screen; attrition = 24%.

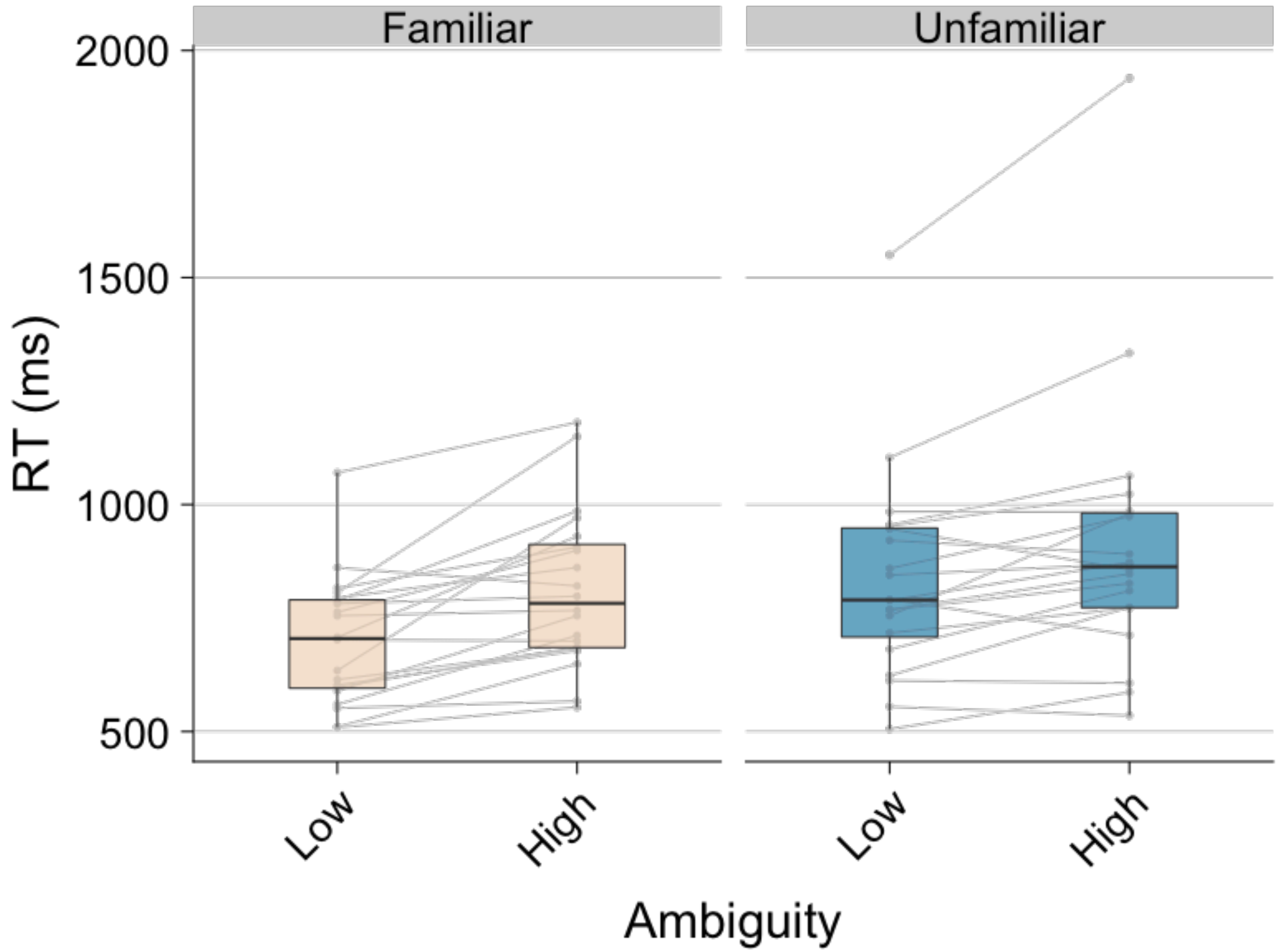
Success 6: Talker familiarity

Block: Familiarization

- 40 trials of talker ID task w/ feedback; Familiar vs. Unfamiliar listener groups

Block: Test

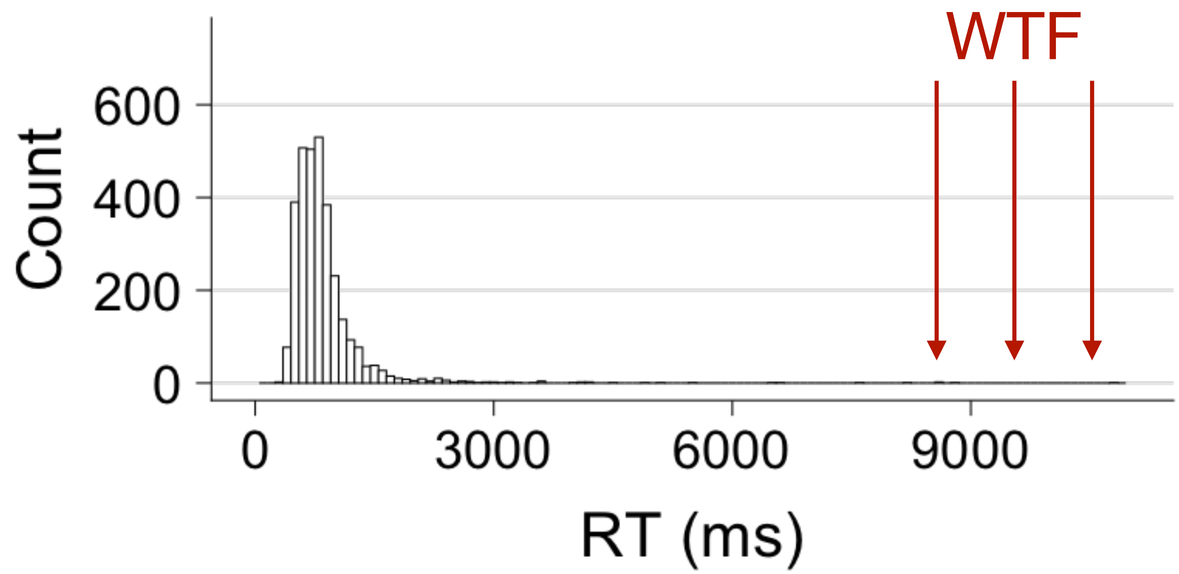
- 80 trials of speeded word ID distributed across two separate blocks; one for a low ambiguity contrast (*heed* vs. *hoed*) and one for a high ambiguity contrast (*hoed* vs. *who'd*)



To achieve current sample (n = 40), we excluded n = 3 due to failure to perform the task and n = 12 due to failure to pass headphone screen; attrition = 27%.

Success 6: Talker familiarity

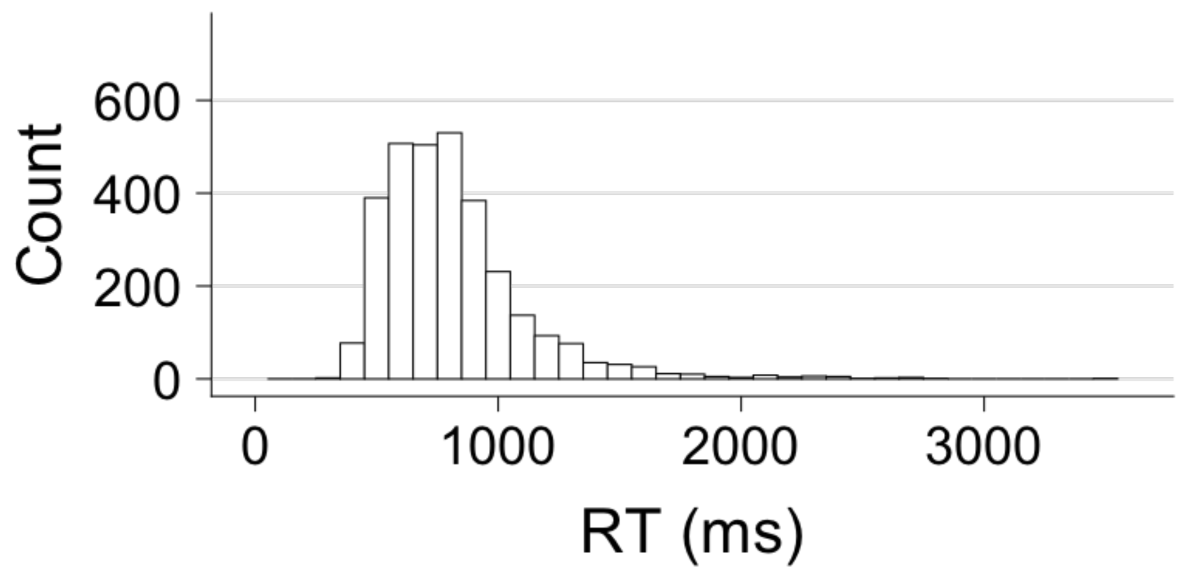
Distribution of RTs for 3,140 correct responses

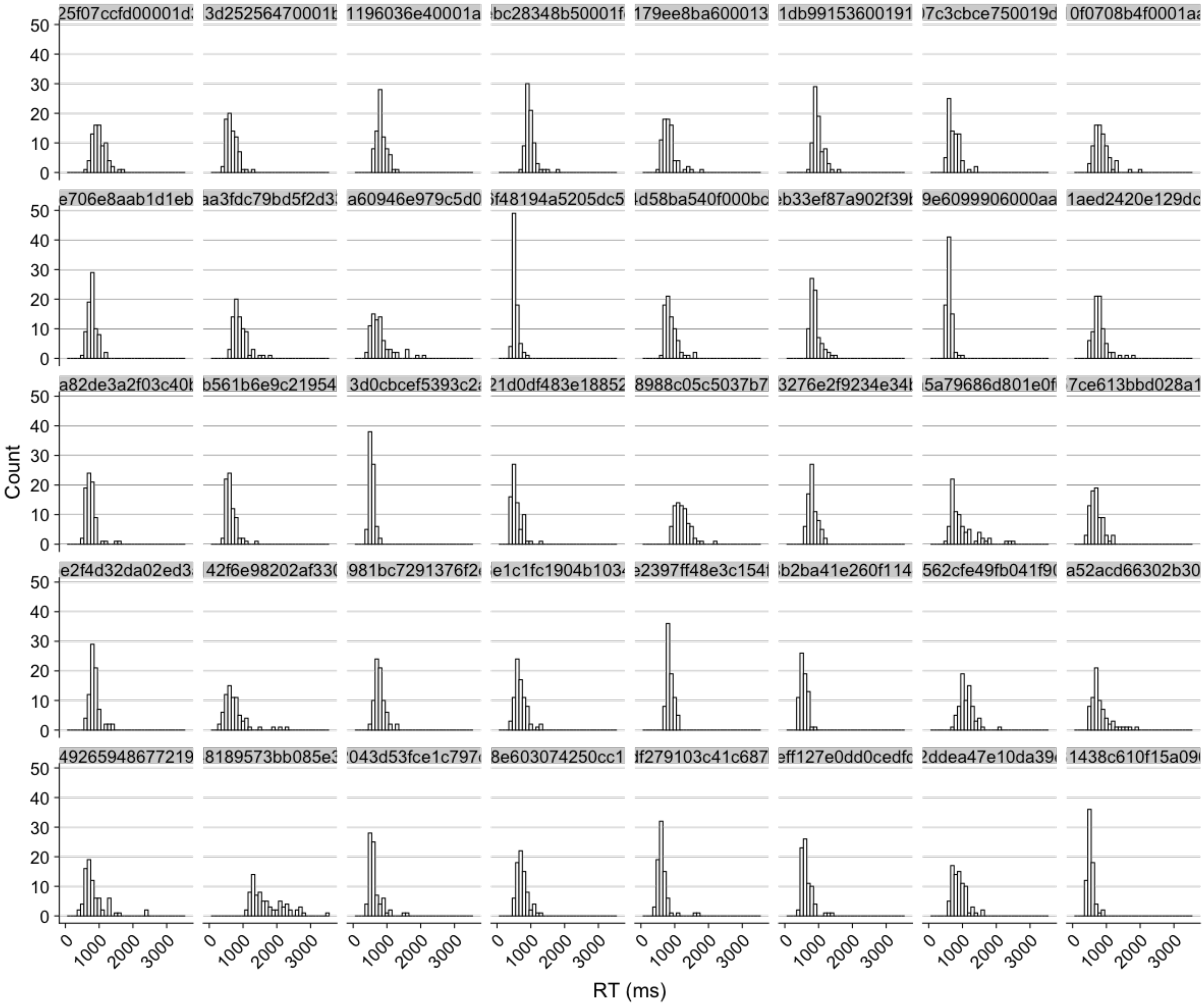


Distribution of RTs excluding:

- RTs > 5000 ms
- RTs exceeding 3 SDs of each subject's mean RT

Removes 1.8% of the data (56 of 3140 trials)

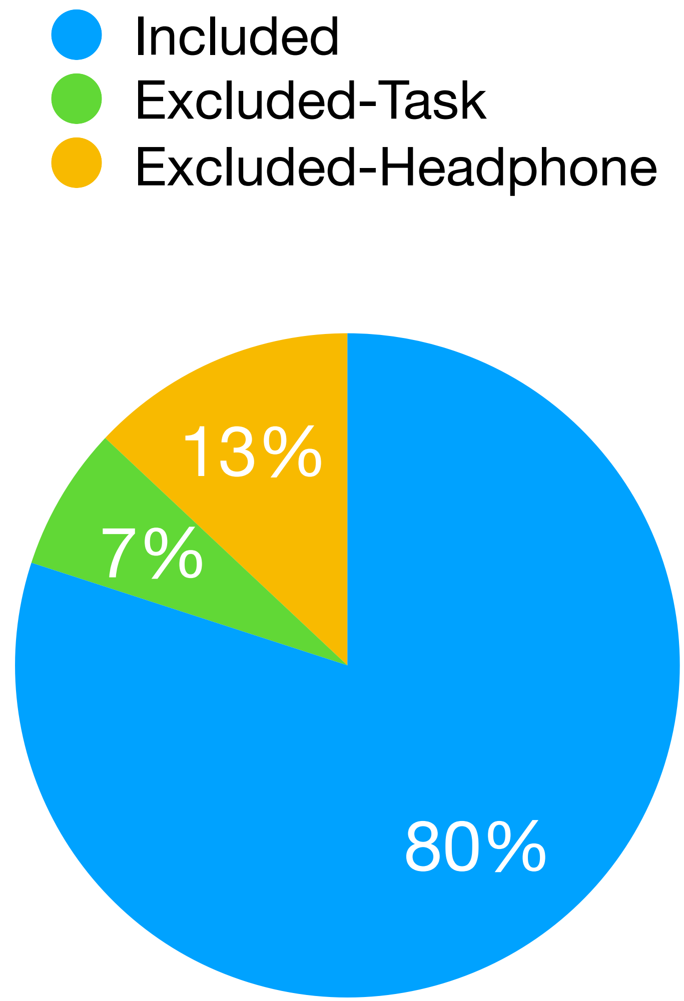




25f07ccfd00001d
3d25256470001b
1196036e40001a
bc28348b50001f
179ee8ba600013
1db99153600191
7c3cbce750019d
0f0708b4f0001a
e706e8aab1d1eb
aa3fdc79bd5f2d3
a60946e979c5d0
f48194a5205dc5
d58ba540f000bc
eb33ef87a902f39
9e6099906000aa
1aed2420e129dc
a82de3a2f03c40
b561b6e9c21954
3d0cbcef5393c2
21d0df483e18852
8988c05c5037b7
3276e2f9234e34
5a79686d801e0f
7ce613bbd028a1
e2f4d32da02ed3
42f6e98202af33
981bc7291376f2
e1c1fc1904b103
e2397ff48e3c154
b2ba41e260f114
562cfe49fb041f9
a52acd66302b30
49265948677219
8189573bb085e3
043d53fce1c797
8e603074250cc1
df279103c41c687
eff127e0dd0cedf
2ddea47e10da39
1438c610f15a09
Count
0
10
20
30
40
50
0
1000
2000
3000
RT (ms)

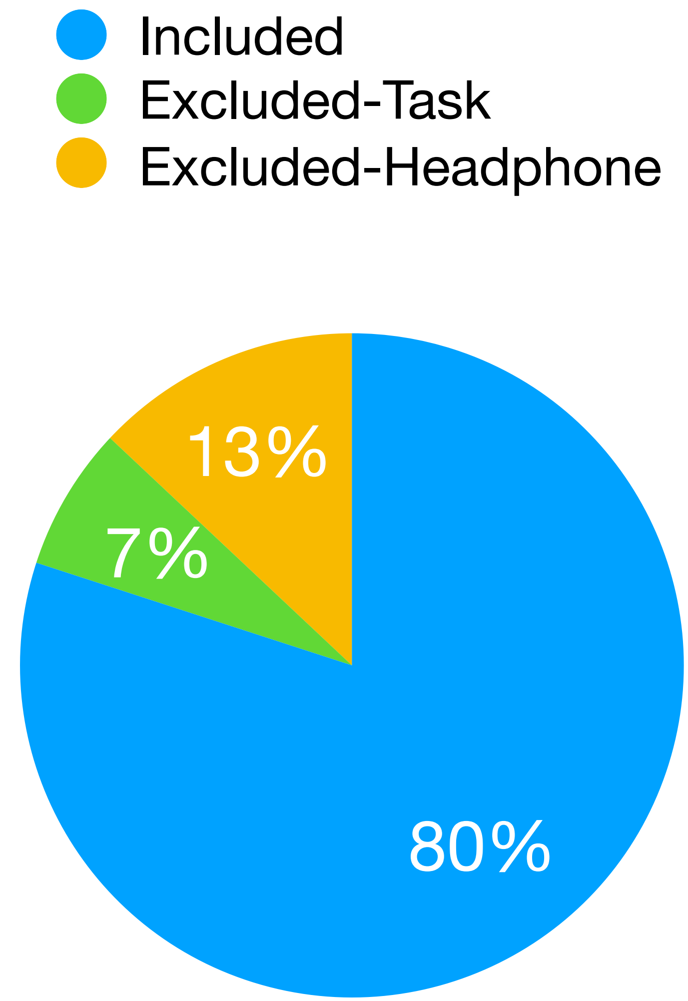
Challenges: Headphone compliance

- Headphone compliance is the greatest source of attrition
- Loss of 176 participants for the studies presented here
- *Specificity* of the Woods et al. (2017) screen?
- Dear Prolific, please let us compensate people for the headphone screen, and then route them out of the study



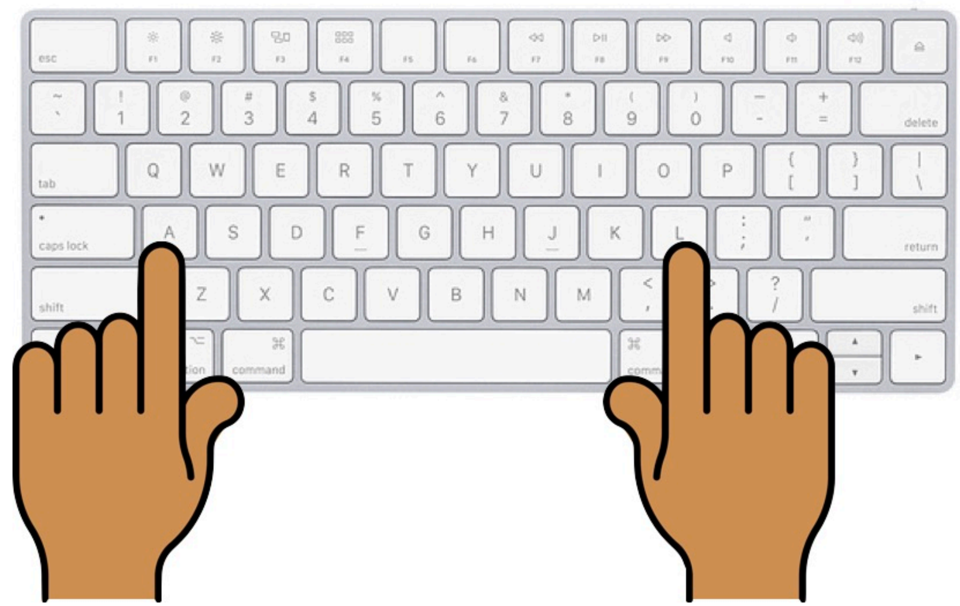
Challenges: Bots/low-effort participants

- Bots and low-effort participants are rare in Prolific, and fairly easy to detect
  - Are many RTs < 5 ms?
  - Are RTs *too* consistent?
  - Is accuracy at chance?
  - Do you see logistic response functions where expected?
- When possible, design studies that support bot detection



Challenges: RT as a dependent measure

- RT experiments pose unique challenges
  - Timing accuracy/ variability for sound presentation; how can we constrain it?
  - Need to develop clear, *a priori* inclusion and outlier criteria
  - Use within-subjects manipulations when you can



Behavioural
Science Online
#BeOnline2020

slaplab.uconn.edu

@UConnNBL

osf.io/8krg3/

For a more in-depth tutorial, visit:
https://tinyurl.com/Online-Perception-Tutorial