An eye-tracking replication without an eye tracker: Capturing predictive sentence processing of accented speech via the internet

Adam A. Bramlett, Chun-Ying Tu, Xiaohan Liu, Seth Wiener

Correspondence email: abramlet@andrew.cmu.edu

Carnegie Mellon University

Motivation

Reproducibility crisis suggestions[2]

- **More** replications need to be done
- Replicate **recent** studies to increase reach
- Replicate with **single change**
- **Openly available** experiment items and data
- Less replication is done with expensive methods (e.g. eye-tracking, neuro-imaging)

Why did we choose Porretta et al.(2020)[1]?

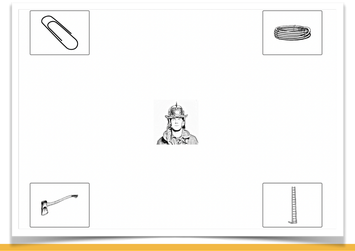
- Eye-tracking is expensive $$$
  E.g., Hardware, software and labs
- Expense limits access to highly funded researchers
- COVID-19 -> web based eye-tracking tools
- But, are the tools sufficiently developed for the language science world?
- Perfect opportunity for replication

Figure 1: VWP paradigm example

Unrestrictive: The fireman will **need** the ladder.

Restrictive: The fireman will **climb** the ladder.

Spoken in American and Chinese accented English



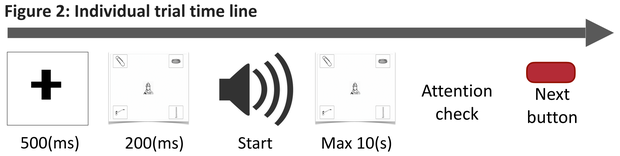
Porretta et al. (2020)

- Visual world paradigm
- Replicates [5], **with two accents (American and Chinese)**
- 60 items, two speakers, two sentences types: (un)restricted
- Figure 1 shows the typical trial with a talker in the center, a target, two competitors, and distractor in the visual world paradigm (VWP).

RQ 1: To what extent does accented speech affect predictive processing?

RQ 2: To what extent does experience affect processing of accented speech?

Figure 2: Individual trial time line



Data Processing

Cleaning and data manipulation

Data Preparation

- Aggregate data frame created by joining across participant eye fixation data (3,000 data sheets)
- Audio files, sentences, object sets, and image ID were then added as columns from Porretta et al.'s data via the OSF
- A dichotic pitch screen test [6] confirmed participants wore headphones
- n=49 (82% ≈ participants kept)

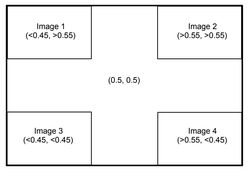
Determining Image Viewed

- Image viewed during each eye tracking recording determined by eye fixation coordinates (x, y) defined by quadrants, see figure 3.

Creating Time Bins

- Auxiliary timestamps were added to the aggregate data to facilitate binning of time intervals due to variable frame rate across trial and across participant

Fig 3: Definition of quadrant



Choice of time interval bin size

- Bin size of 50 used in analysis (average of 92 data points eye fixation measurements per bin)
- Larger bin size -> more data points within bin, less data points across time
- Smaller bin size -> less data points within bin, more data points across time

Figure 4: Frame rate by participant and item over time 50 ms bin size

frame rate by item (blue) and participant (yellow)

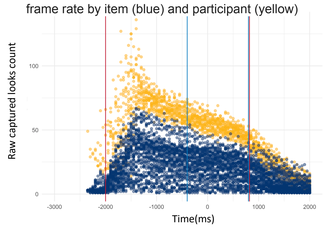


Table 1: Data collection and quality comparison

Our Replication	Porretta el al. (2020)
Web cam using Gorilla Web gazer[3]	Eye-link 1000 plus
Variable frame rate	Fixed consistent frame rate
$\bar{x}$ ≈ 20 Hz measurement	$\bar{x}$ ≈920Hz measurement
Looks to specific location converted to looks	Proportion of looks in 50ms window (e.g., 23 of 50)

Data Loss

Gorilla rejection

Quality removal

IRB consent: -1 participant

Headphone check [6]: -8 participants

Eye calibration: -23 participants

Timed out -5 participants

Dishonest: -2 participants

Low frame rate: -2 participants

Low accuracy: -7 participants

Bad item: -5 items

Completion ≈ 16 min

Time limit = 90 min

Statistics and Modeling

Looks to target object image

- Dependent variable of interest was binary *looks to target object image* (1 if looking, 0 otherwise)
- Logistic generalized additive mixed model was used to analyze the data with logistic link function, equivalent to modeling logit-transformed response probability with identity link function.

Accent and prediction

- Primary independent variable of interest: factors *talker, verb type* (restricting vs. non-restricting) and the interaction between the two variables
- Significant intercept (reference level: native non-restricting) and restricting verb type indicate that prediction in restricting verbs only occurs with the native accent

Table 2: Model 1 output

	Estimate	Std.Error	p.value
(Intercept)	-1.102	0.106	<2e-16***
talkerNonNativeMale	-0.056	0.136	0.684
verb_typeRestricting	0.355	0.135	0.009 **
talkerNonNativeMale:verb_typeRestricting	0.159	0.190	0.404

Prediction and experience (non-native)

- Same as model above but with the fixed effect of experience

Table 3: Model 2 output

	Estimate	Std.Error	p.value
(Intercept)	-0.922	0.073	<2e-16***
experience	0.015	0.002	7.73e-16***

Figure 5: Online replication (left) vs. in person (right)

Right visual is from Porretta et al.'s (2020)

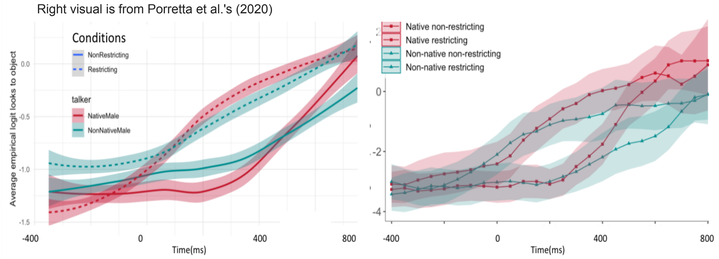


Table 4: Findings comparison

Our Replication	Porretta el al. (2020)
Prediction effect for native *and* non-native accented speech in restrictive condition	Prediction effect for native-accented speech in restrictive conditions [1]
Online experiment reaches broader population of participants [4]	All participants recruited from the University of Windsor

Results

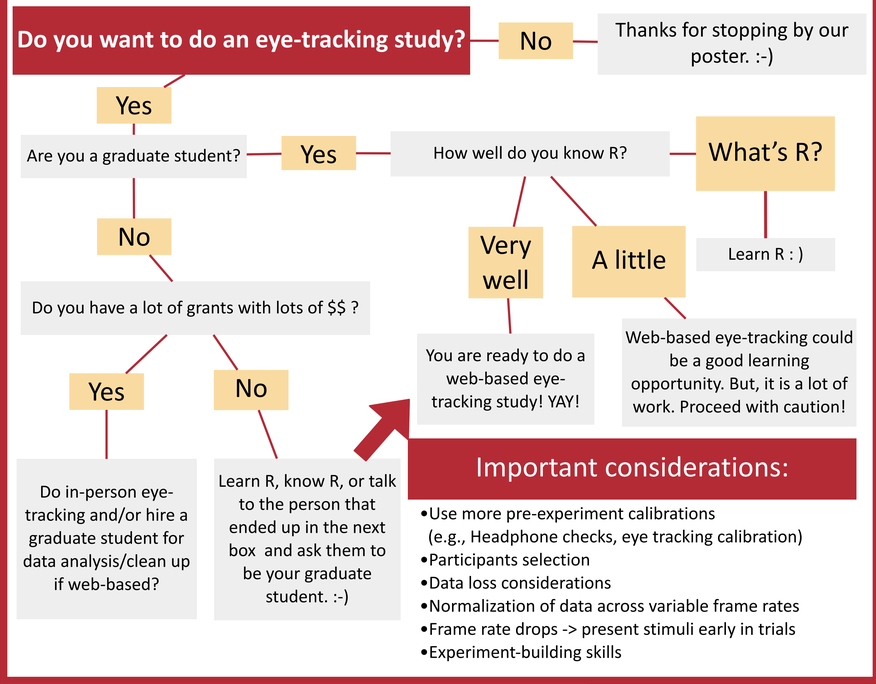
Table 5: Results comparison

Our Replication	Porretta el al. (2020)
(Un)restrictive verb-type modulates prediction overall (β=0.36, SE=0.14, p=.009**)	(Un)restrictive verb -type modulates prediction for native speech
Experience modulates prediction in non-native accented speech. (β=0.02, SE=0.002, p<.001***)	Experience modulates prediction in non-native accented speech

- The difference may be due to a more diverse sample/more experienced sample (Chinese accent familiarity)
- Web-based eye tracking and R skills are a good alternative for: expensive gear/labs/software.

Takeaways

Figure 6: A flowchart for planning an eye-tracking study



Important considerations:

- Use more pre-experiment calibrations (e.g., Headphone checks, eye tracking calibration)
- Participants selection
- Data loss considerations
- Normalization of data across variable frame rates
- Frame rate drops -> present stimuli early in trials
- Experiment-building skills

References

[1] Porretta, V., Buchanan, L., & Järvikivi, J. (2020). When processing costs impact predictive processing: The case of foreign-accented speech and accent experience. *Attention, Perception, & Psychophysics, 82*(4), 1558-1565.

[2] Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning, 68*(2), 321-391.

[3] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods, 52*, 388-407.

[4] Prolific (www.prolific.co), Accessed November 2022.

[5] Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

[6] Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, 53, 1551-1562.

Play with our data here

Learn more about what we do here