

## Introduction

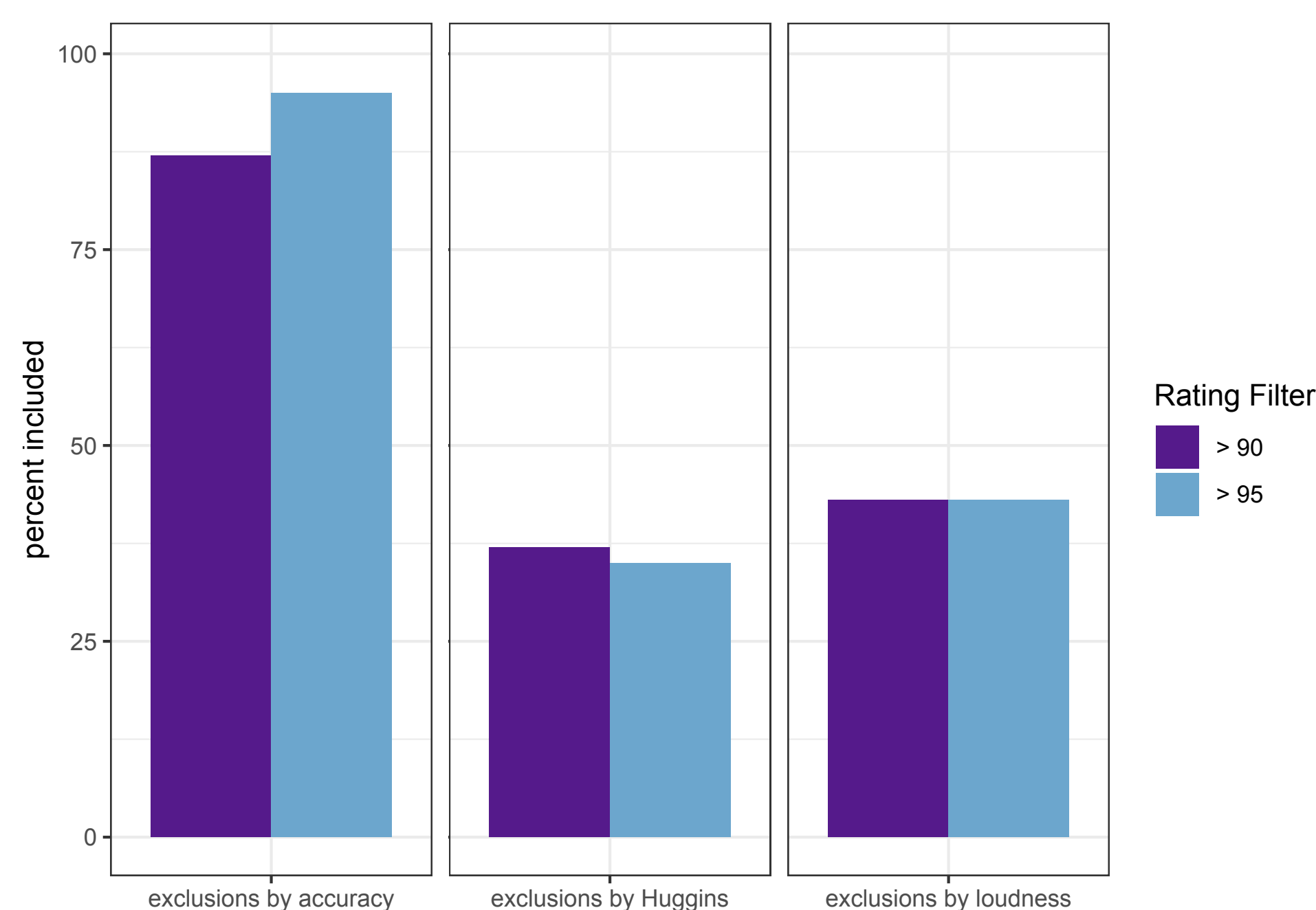
- Headphone checks have become an assumed part of best practices in online studies.
- Are they actually useful?
- *Are headphones better than other speakers?* Headphones can reduce environmental noise, but are not always better (Olive, Khonsaripour, & Welti, 2018).
- *Do headphones increase consistency across listeners?* There is substantial variation across headphones (Cooke & García Lecumberri 2021).

**This Study:** An online replication of three phonological patterns, testing whether excluding participants with two headphone check methods makes results clearer:

- The Huggins check improves results for how spectral tilt influences perception.
- No other results were improved by either headphone check.

## Initial Filtering

**Approval Rating:** Participants' Prolific approval rating predicts their accuracy identifying naturally produced consonants, but not their headphone check results:



**Instructions about device:** When task description excluded phones from the list of suitable devices, no participants report using phones for the task (vs 30% without this restriction)

## Methodology

**Participants:** 120 native speakers of American English, online through Prolific

**Filters:** Approval rating on Prolific (>90%, >95%);

Device suggestions in task description (phone included, phone excluded)

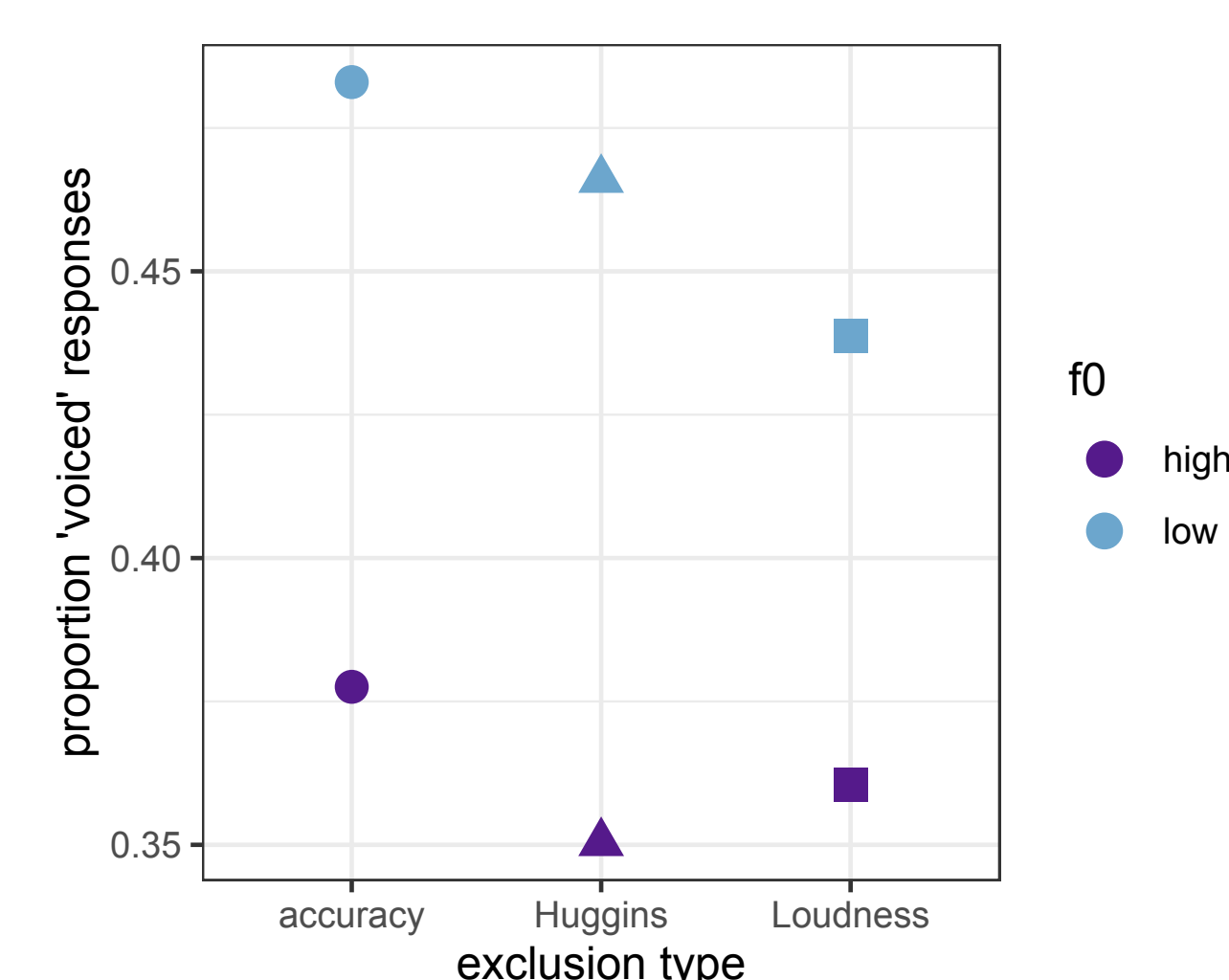
- Phonological perception tasks:
  1. Choosing the written form matching items with ambiguous VOT (e.g. *best, pest*) – higher following F0 should increase voiceless onset identifications (cf. Haggard et al 1970)
  2. Choosing the written form matching items with ambiguous F1 (e.g. *pit, pet*) after exposure to items with shifted F1 – exposure to raised/lowered F1 should produce a corresponding shift in subsequent perception (cf. Ladefoged & Broadbent 1957)
  3. Categorizing vowels as long or short in duration – lower spectral tilt should increase perceived vowel duration (cf. Sanker 2020)
- Audio checks:
  1. Huggins Pitch headphone check (Milne et al 2021) – threshold 5/6 correct
  2. Dichotic Loudness headphone check (Woods et al 2017) – threshold 5/6 correct
  3. Accuracy identifying consonants (e.g. *bud, bug; theft, heft*) – threshold 87% (21/24)

## Main Results

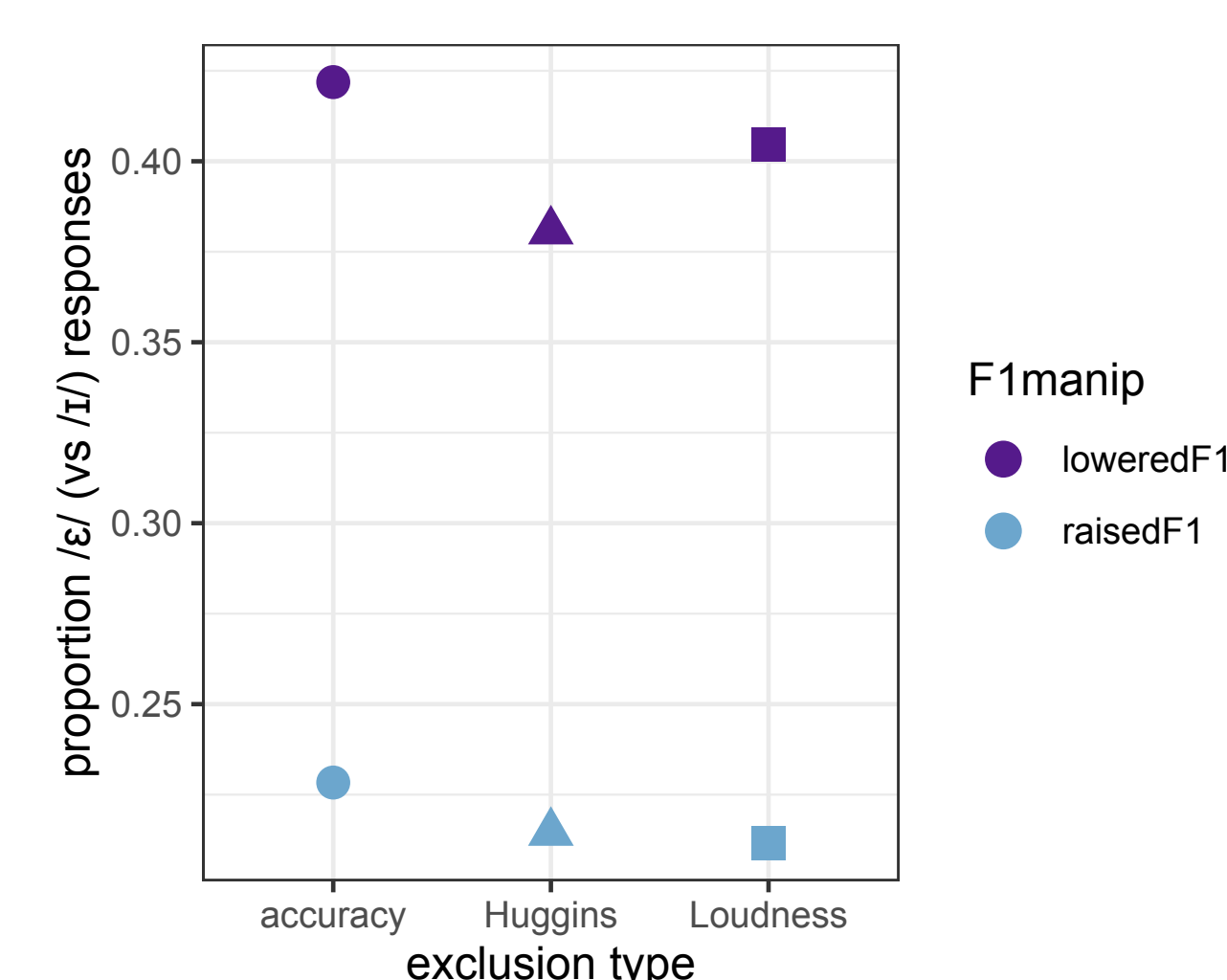
All tasks replicated the predicted results using an accuracy-based audio/attention check. Most results were very similar across the different exclusion methods – both headphone checks and exclusions based on accuracy in clear items

exclusion method	f0 on voicing	F1 manip exposure	tilt on duration
consonant ident accuracy > 87%	0.468 **	-1.69 ***	0.49 *
Huggins check pass	0.554 **	-1.53 ***	0.98 ***
dichotic loudness pass	0.41 *	-1.84 ***	0.48

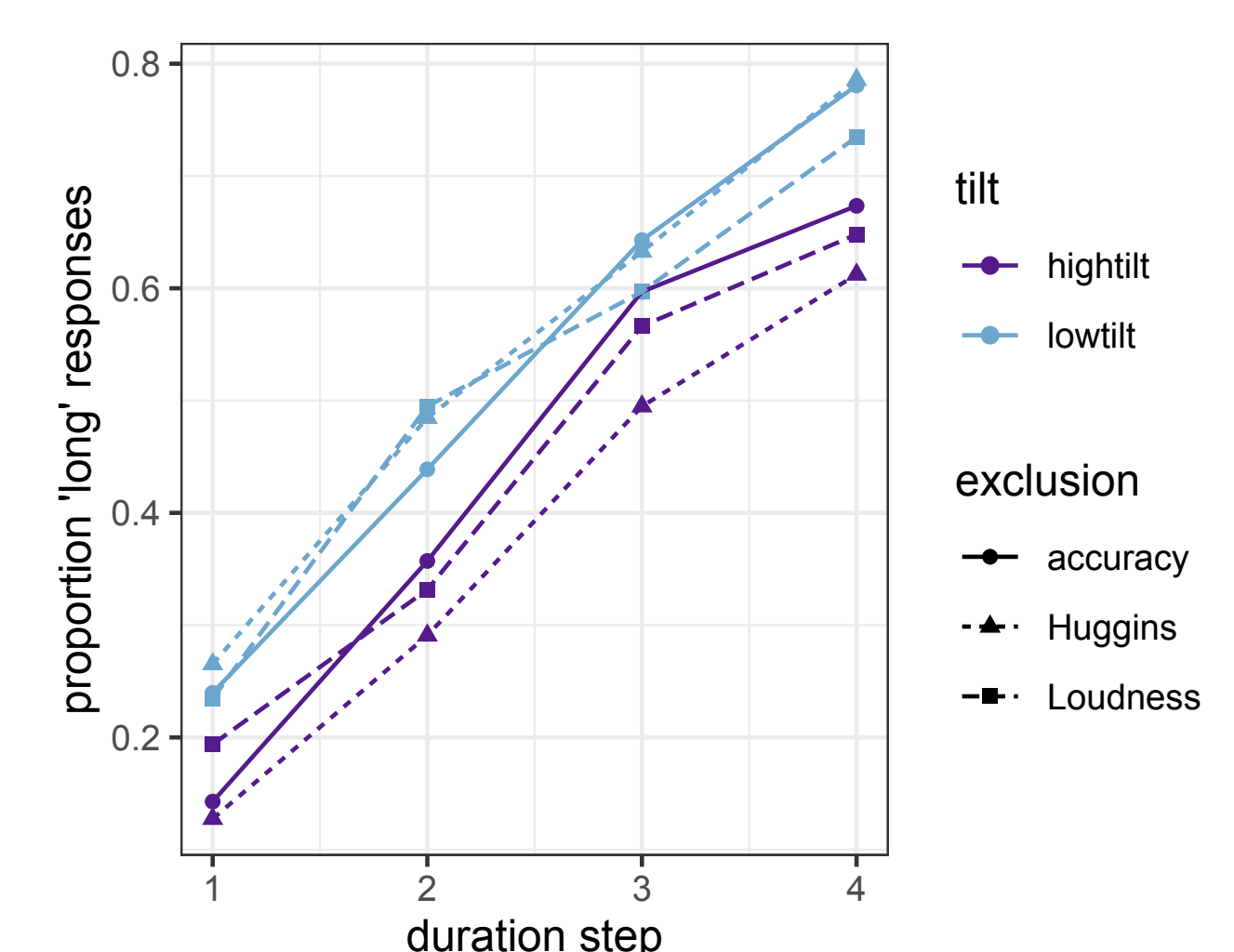
**Table 1:** Results of logistic mixed effects models for the factor of interest in each task (estimate and significance). The same model was run on the data with three different participant exclusion methods; all models used 49 participants (the number who passed the Huggins check).



**Fig 1:** Onset voicing categorization by following f0 and exclusion type



**Fig 2:** Vowel identifications by F1 exposure manipulation and exclusion type



**Fig 3:** Duration categorization by spectral tilt and exclusion type

## Conclusions

### Results Summary

- Headphone checks can be useful for certain acoustic characteristics; the Huggins check results in a clearer effect of spectral tilt on perceived duration
- However, headphone checks are not useful for all tasks, and not all headphone checks have the same effect

### Takeaways

- Headphone checks exclude many potential participants, and for many studies they do not improve the data
- There are many sources of variation across participants which will not be reduced by headphone checks, such as:
  - Low effort or limited attention, e.g. listening to music or watching movies (Chandler, Mueller, & Paolacci, 2014)
  - Misrepresentation of native language in order to access studies (Peer et al 2021)
- A headphone check will not decrease the number of participants necessary for a reasonably powered study

## Selected References

Chandler, J., Mueller, P., & Paolacci, G. 2014. Nonnaivete among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behav Res Methods*, 46(1), 112–130.

Cooke, M., & García Lecumberri, M. L. 2021. How reliable are online speech intelligibility studies with known listener cohorts? *Journal of the Acoustical Society of America*, 150(2), 1390–1401.

Haggard, M., Ambler, S., & Callow, M. 1970. Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 47(2B), 613–617.

Ladefoged, P., & Broadbent, D. E. 1957. Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29(1), 98–104.

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. *Behav Res Methods*, 53(4), 1551–1562.

Olive, S., Khonsaripour, O., & Welti, T. 2018. A survey and analysis of consumer and professional headphones based on their objective and subjective performances. In *Proceedings of the 145th Audio Engineering Society Convention*.

Peer, E., David, R., Andrew, G., Zak, E., & Ekaterina, D. 2021. Data quality of platforms and panels for online behavioral research. *Behav Res Methods*.

Sanker, C. 2020. A perceptual pathway for voicing-conditioned vowel duration. *Laboratory Phonology*, 11(1), Art. 18.

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. 2017. Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79, 2064–2072.