Homophone discrimination based on speaker-specific learning

Chelsea Sanker

Linguistic Society of America

8 January 2021

Phonetic details in the phonological representation

- Phonetic details can be part of the phonological representation or implementation rules
- Reflected in language-specific realizations of the same sound (e.g. Lieberman 1970; Keating 1985)
- And shifts in those details in convergence (e.g. Nielsen 2011) and perceptual learning (e.g. Kraljic & Samuel 2006)

Phonetic details in the lexical representation?

- Can phonetic details also be part of the lexical representation?
- Homophones provide a test
  - Homophone mates can exhibit significant acoustic differences (Gahl 2008; Guion 1995)
  - But listeners generally cannot discriminate between them (Bond 1973; Sanker 2019)

Speaker-specific learning

- Discrimination might require speaker-specific learning, and thus require
  - Stimuli from a single speaker
  - Items that are unambiguous from the context
- Exposure to a speaker could improve familiarity with that speaker's phonological system and other systematic patterns like frequency-based reduction (Gahl 2008)
- However, learning production patterns of less systematic factors like emotional valence (Nygaard et al. 2009) might require exposure to the particular words

This study

I present results from a word identification task:

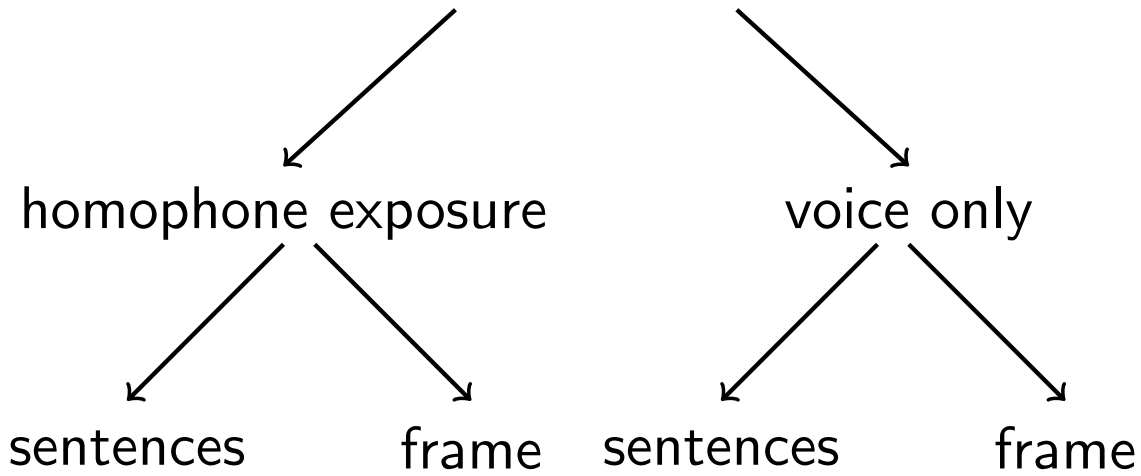
- How does exposure to a speaker influence listeners' accuracy in identifying homophones?
- Can listeners distinguish between homophone mates when they have heard the same voice producing different tokens of the same words?
- Can listeners distinguish between homophone mates when they have heard the voice, but not saying these particular words?

Listeners

- 128 native speakers of American English (mean age 39.0; 68 male, 59 female, 1 nonbinary)
- No reported speech or hearing disorders
- Run online though Qualtrics, with participants paid through Amazon Mechanical Turk

Task design

- Words presented auditorily in isolation for identifications
- All stimuli produced by one individual
- Four conditions, based on (1) type of exposure in training and (2) production context that stimuli were extracted from:



Training Phase

In the training phase:

- Participants heard a word and identified it as one of two **phonologically distinct** orthographic response options
- Two training conditions:
  1. Training pairs included words that would appear in the homophone task (e.g. *night*, *neat*)
  2. Training pairs only included words that would not appear in the homophone task (e.g. *pipe*, *peep*)

Test Phase

In the test phase:

- Participants heard a word and identified it as one of two orthographically distinct **homophone mates** (e.g. *night*, *knight*)
- Each participant heard both homophone mates of each pair, presented in randomized order
- The training stimuli and test stimuli were always distinct tokens, produced by the same speaker
- Two conditions of stimuli:
  1. Test stimuli extracted from a frame sentence, e.g. *The word is sun.*
  2. Test stimuli extracted from naturalistic sentences, e.g. *We took a picture of the sun.*

Hypothesis: Acclimation to speakers

Hypothesis A: Listeners become familiar with a speaker and can subsequently make predictions about systematic patterns like frequency-based reduction, which could produce above-chance discrimination of homophone mates

Hypothesis: Learning particular words by speaker

Hypothesis B: Listeners become familiar with how a speaker says particular words, including differences between homophone mates based on factors like emotional valence, which could produce above-chance discrimination of homophone mates only if listeners have previously heard the speaker saying those words

Hypothesis: No learning

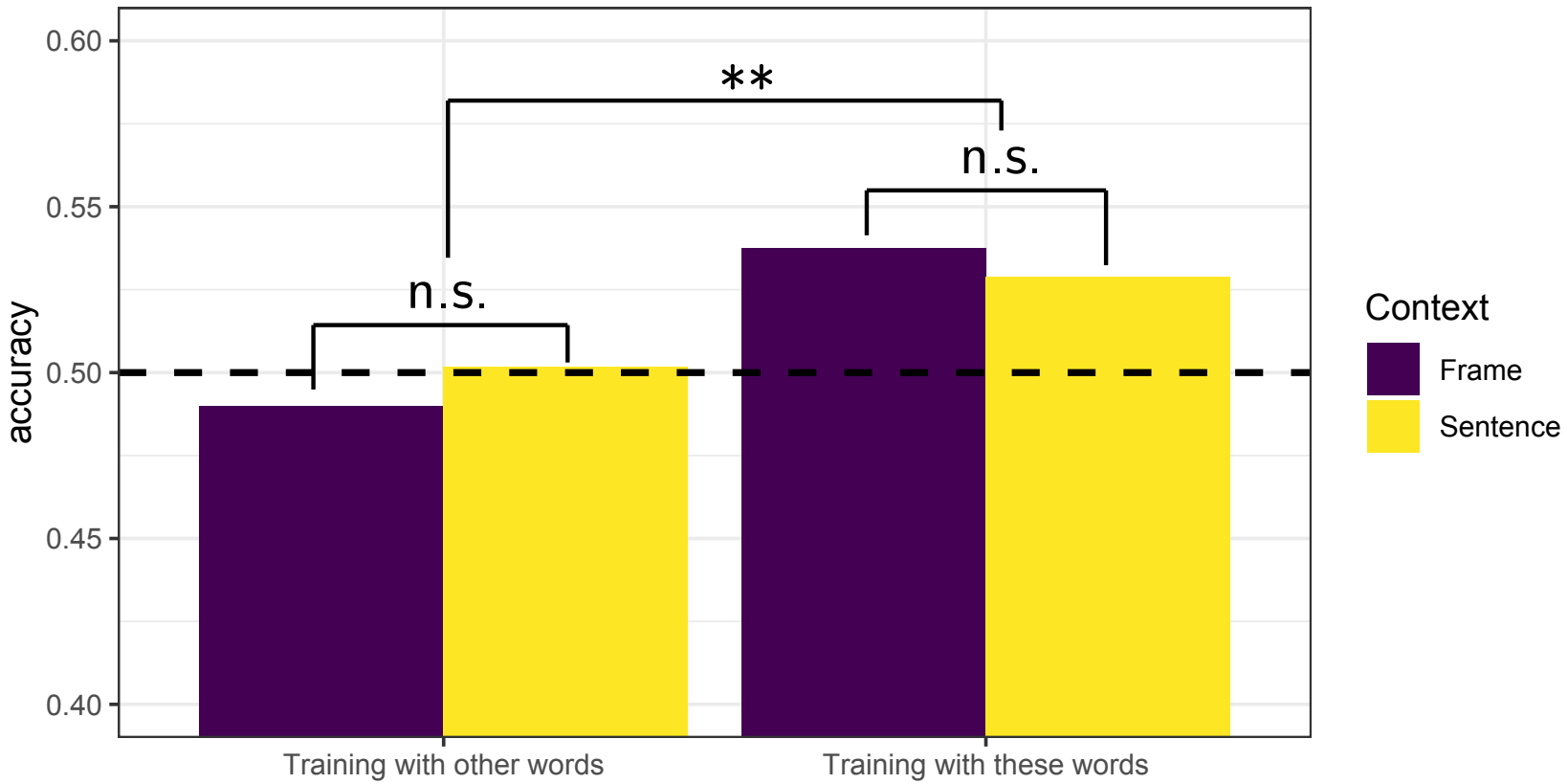
Hypothesis C: No amount of exposure to a speaker will improve discrimination between homophone mates, because there are no systematic phonological differences

Logistic mixed effects model for accuracy of homophone identifications

	Estimate	Std. Error	t value	p value
(Intercept)	0.15	0.058	2.6	**0.0097**
Cond NoHomExposure	-0.19	0.08	-2.4	**0.016**
Context Sentence	-0.035	0.08	-0.43	0.66
Cond NoHomExposure * Cont Sentence	0.082	0.11	0.73	0.47

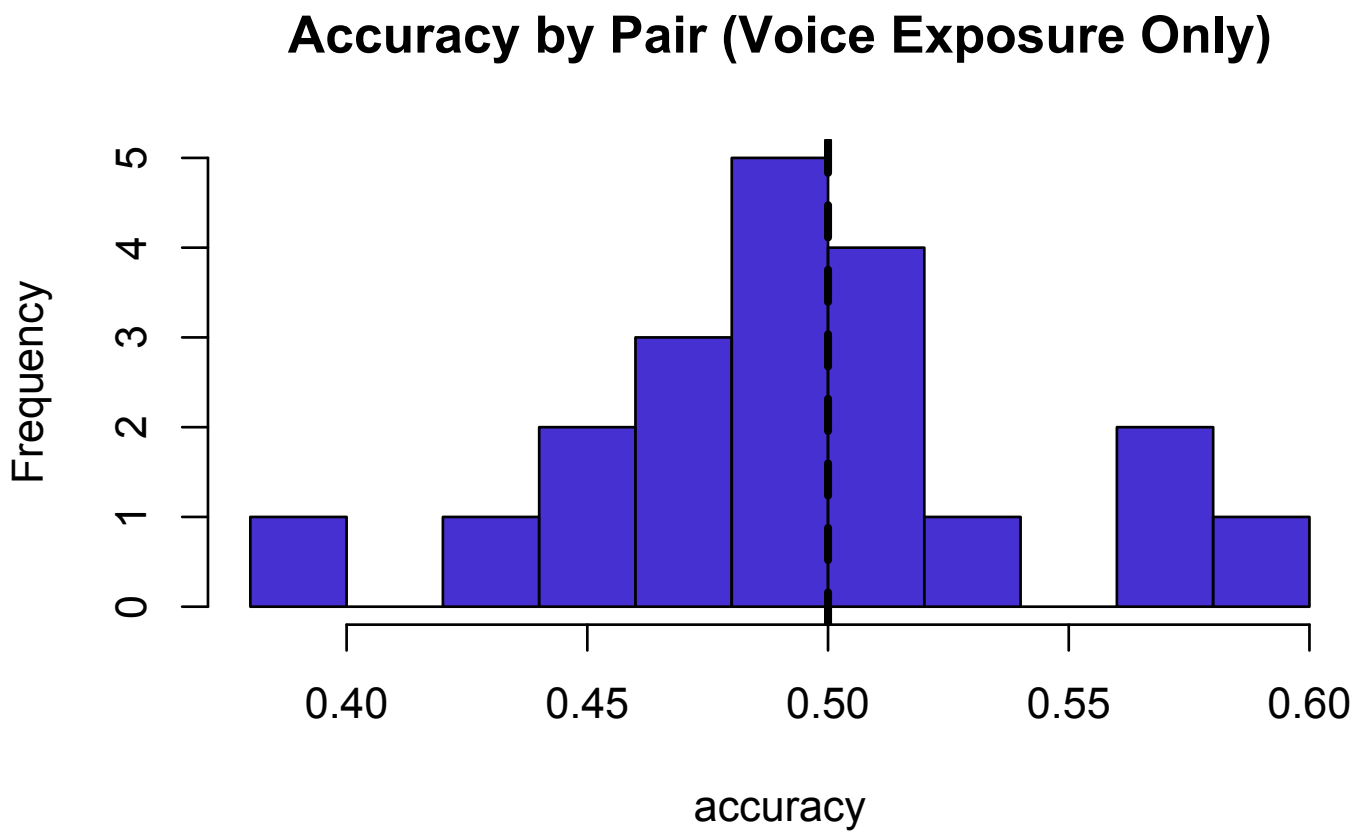
Intercept: Condition = HomophoneExposure, Context = FrameSentence

Accuracy by condition



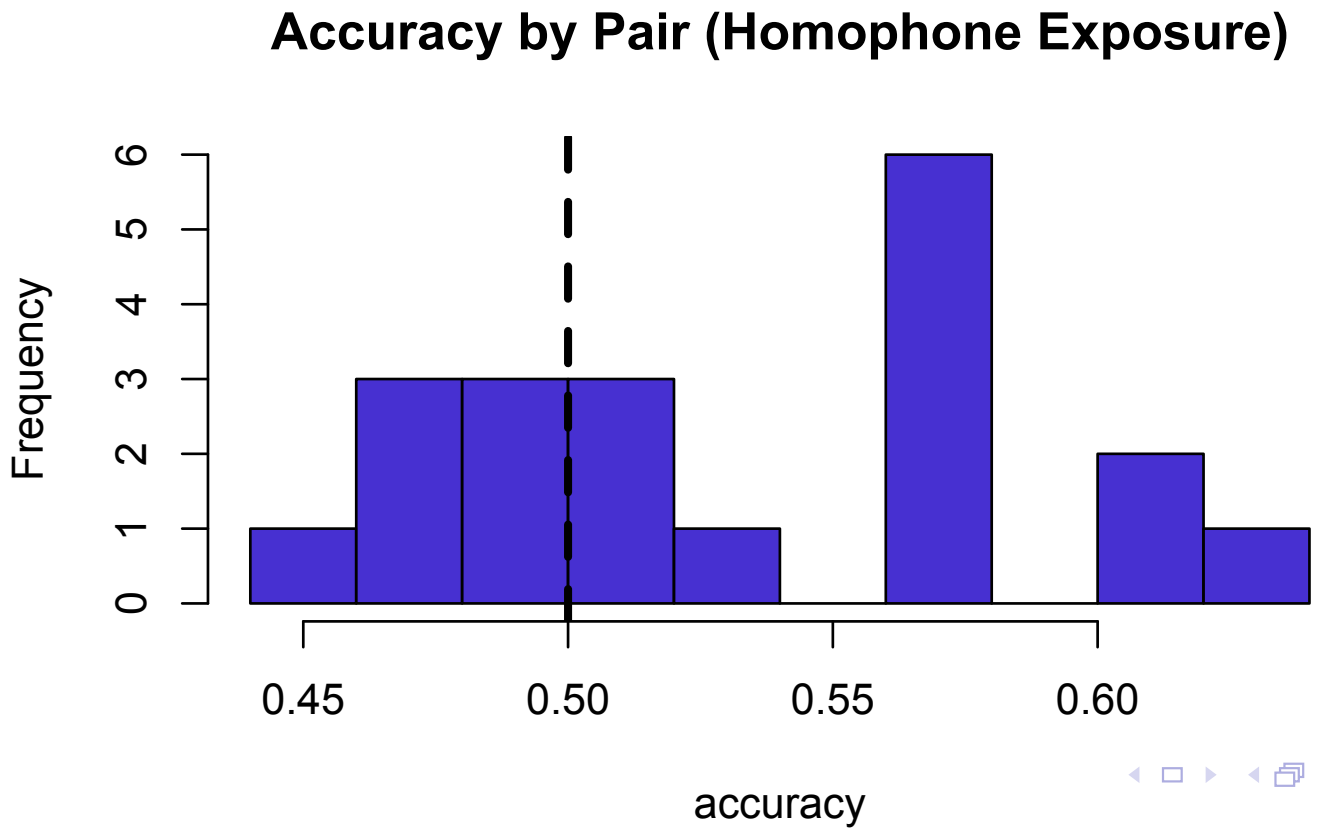
Accuracy by pair (Voice Exposure Only)

When the training didn't include the test words, by-pair accuracy was close to normally distributed, centered near 0.5



Accuracy by pair (Homophone Exposure)

When training included the test words, half the distribution of by-pair accuracy is mostly symmetrical around 0.5, but the other pairs have substantially higher accuracy



Accuracy by pair

- The distribution suggests that the accuracy might be largely driven by certain pairs
  - sail - sale 56%
  - brake - break 57%
  - flea - flee 57%
  - whale - wail 57%
  - week - weak 57%
  - night - knight 58%
  - mail - male 60%
  - write - right 61%
  - steel - steal 63%

Word specific detail

- When trained on tokens of different words produced by the speaker, accuracy was at chance
- Listeners don't have pre-existing expectations about how homophone mates should differ
- Learning is not based on having a reference point to evaluate phonological contrasts or other systematic effects like frequency-conditioned reduction

Word specific AND speaker specific detail

- Accuracy was above chance after exposure to tokens of the same words produced by the speaker
- This must reflect non-systematic word-specific characteristics as produced by that speaker, which cannot be reliably predicted from other words
- Recall that accuracy was only slightly higher than chance (53%); this is not a phonological contrast

So what is the status of such details?

- The results suggest word-specific memories
  - Listeners recognize particular tokens that have been presented previously (e.g. Hintzman et al. 1972)
  - Token learning can also impact similar but non-identical tokens (e.g. Church & Schacter 1994)
- Listeners also can learn patterns particular to certain speakers (e.g. Krajlic & Samuel 2007)

Exemplar memories

- Exemplar memories of word specific details as produced by a particular speaker (cf. Pierrehumbert 2002, Goldinger 1998)
- Also phonetic details associated with phonological categories, connected across words and across speakers
- Word-specific patterns can be apparent with extensive exposure (e.g. Rochet-Capellan & Ostry 2011), or because they draw on pre-existing associations

References

Bond, Z. 1973. The perception of sub-phonemic phonetic differences. *Language and Speech*, 16, 351–355.

Church, B. & Schacter, D. 1994. Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 521–533.

Gahl, S. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3), 474–498.

Guion, S. 1995. Word frequency effects among homonyms. *Texas Linguistic Forum*, 35, 103–116.

Hintzman, D., Block, R., & Inskeep, N. 1972. Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, 11, 741–749.

Keating, P. 1985. Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115–131). Academic Press.

Kraljic, T., & Samuel, A. 2006. Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–178.

Kraljic, T., & Samuel, A. 2007. Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15.

Lieberman, P. 1970. Towards a unified phonetic theory. *Linguistic Inquiry*, 1(3), 307–322.

Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142.

Nygaard, L., Herold, D., & Namy, L. 2009. The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, 33, 127–146.

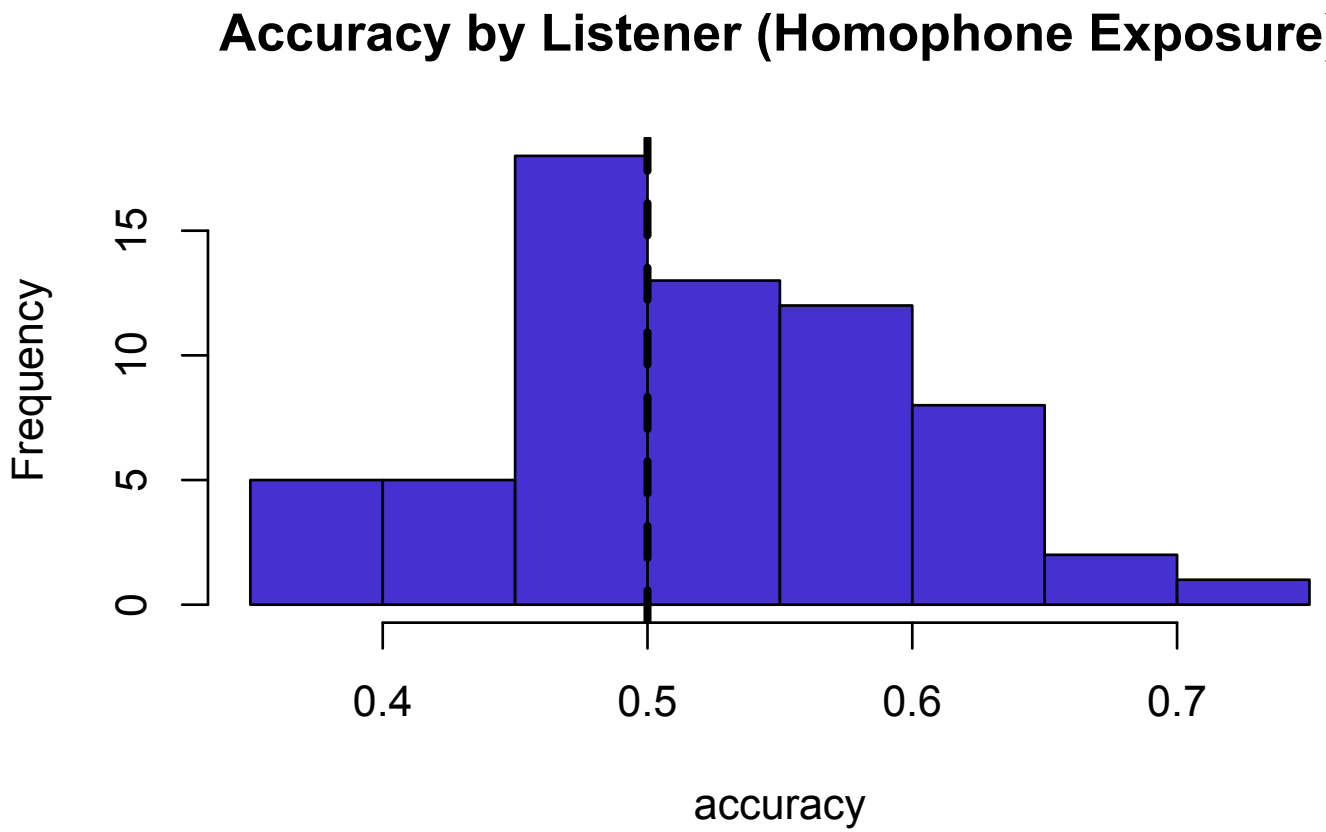
Pierrehumbert, J. 2002. Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 101–139). Berlin: Mouton de Gruyter.

Rochet-Capellan, A. & Ostry, D. 2011. Simultaneous acquisition of multiple auditory-motor transformations in speech. *Journal of Neuroscience*, 31(7), 2657–2662.

Sanker, C. 2019. Effects of lexical ambiguity, frequency, and acoustic details in auditory perception. *Attention, Perception, & Psychophysics*, 81, 323–343.

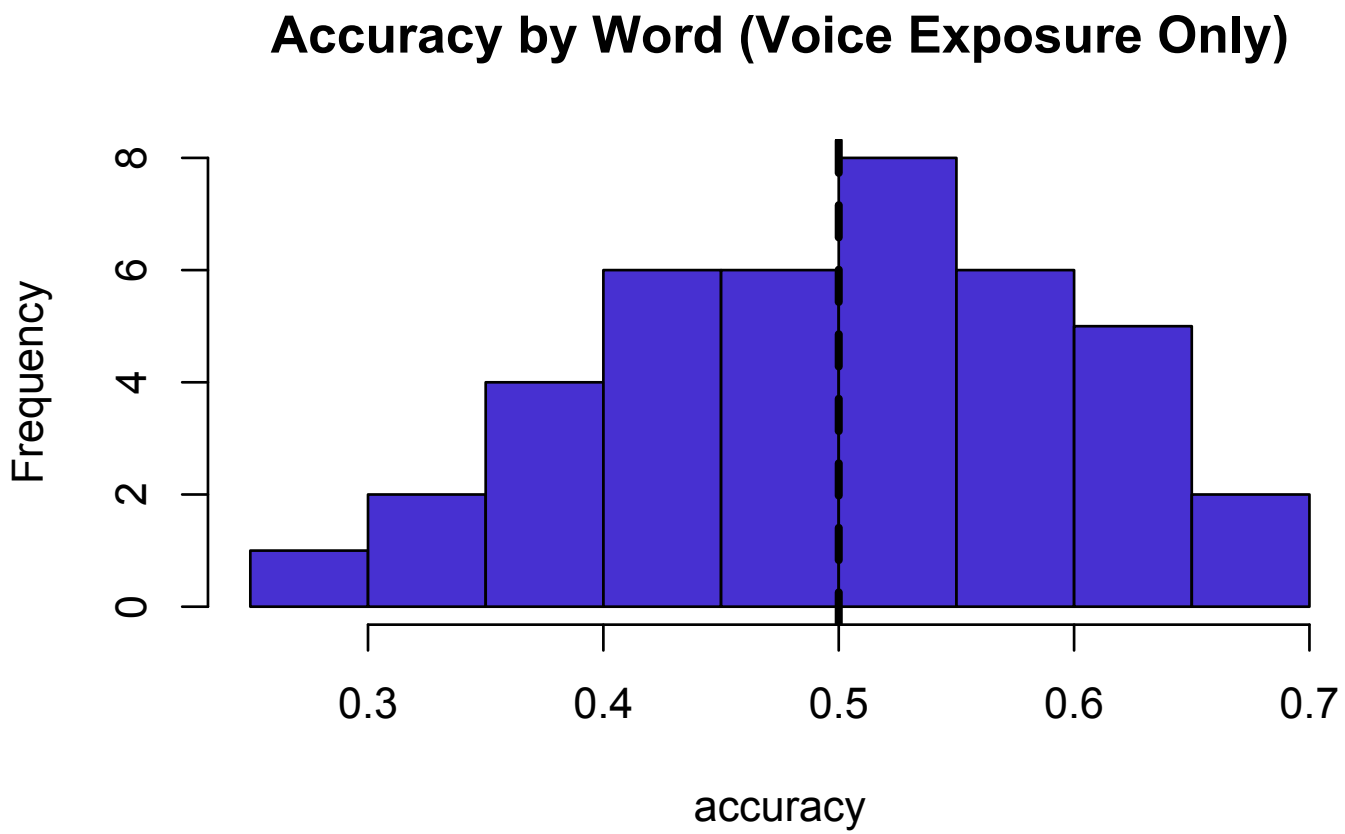
Accuracy by listener (Homophone Exposure)

When training included the test words, the peak is around 0.5, but there are many participants in the 0.5-0.6 range



Accuracy by word (Voice Exposure Only)

When the training didn't include the test words, by-word accuracy was close to normally distributed, centered near 0.5



Accuracy by word (Homophone Exposure)

When the training included the test words, by-word accuracy had a more distinct peak, substantially above 0.5