Perceptual Learning in Spanish: Implications for vowel-specific learning trends

By Nikolai Andrés Schwarz-Acosta
April 26, 2025

The 43rd West Coast Conference on Formal Linguistics (WCCFL 43)
University of Washington, Seattle, WA

Introduction

Overview of Perceptual Learning

What is perceptual learning/adaptation?

People hear new speech patterns

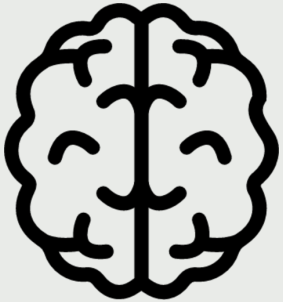
What is perceptual learning/adaptation?

People hear new speech patterns

People learn these patterns

What is perceptual learning/adaptation?

People hear new speech patterns



People learn these patterns

/a/=[e]

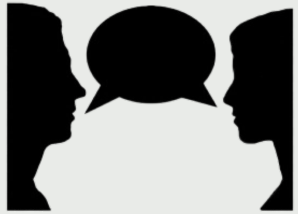
Remapping of phonemic boundaries

What is perceptual learning/adaptation?

People hear new speech patterns

People learn these patterns

/a/=[e]



Remapping of phonemic boundaries

Listeners understand speech with these patterns

What is perceptual learning/adaptation?

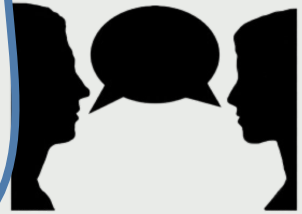
Exposure

People hear new speech patterns

/a/=[e]

Remapping of phonemic boundaries

People learn these patterns



Listeners understand speech with these patterns

What is perceptual learning/adaptation?

Exposure

People hear new speech patterns

/a/=[e]

Remapping of phonemic boundaries

People learn these patterns

Testing

Listeners understand speech with these patterns

Different Levels of Perceptual Learning

Different Levels of Perceptual Learning

Lexical Level:

Different Levels of Perceptual Learning

Lexical Level: listeners learn word-level representations separate from phonemic patterns

Different Levels of Perceptual Learning

Lexical Level: listeners learn word-level representations separate from phonemic patterns

Participants are more likely to recognize accented words if they have heard the word before with the accent (Weatherholtz, 2015; Maye et al., 2008).

Different Levels of Perceptual Learning

Lexical Level: listeners learn word-level representations separate from phonemic patterns

Participants are more likely to recognize accented words if they have heard the word before with the accent (Weatherholtz, 2015; Maye et al., 2008).

e.g. Participants hear the word ‘dog’ /pero/ in a story as [peru] → they are more likely to recognize this as a word than ‘point’ /punto/ as [puntu]

Different Levels of Perceptual Learning

Phoneme Level:

Different Levels of Perceptual Learning

Phoneme Level: listeners learn phonemic-specific changes that can be applied novel words and speakers.

Different Levels of Perceptual Learning

Phoneme Level: listeners learn phonemic-specific changes that can be applied novel words and speakers.

Participants hear words with a phonemic shift, they are able to recognize other words with this shift (Weatherholtz, 2015; Maye et al., 2008).

Different Levels of Perceptual Learning

Phoneme Level: listeners learn phonemic-specific changes that can be applied novel words and speakers.

Participants hear words with a phonemic shift, they are able to recognize other words with this shift (Weatherholtz, 2015; Maye et al., 2008).

e.g. Participants hear a story that contains words with a /o/ → [u] shift and are able recognize other words and speakers with this shift, regardless of lexical item exposure.

Different Levels of Perceptual Learning

Lexical Level: listeners learn word-level representations separate from phonemic patterns.

Phoneme Level: listeners learn phonemic-specific changes that can be applied novel words and speakers.

Factors affecting perceptual learning

There are multiple potential factors that impact the speed:

- **Accentedness** (e.g. Derwing & Munro, 1997; Witteman et al., 2013)
- **Multiple speakers** (e.g. Luthra et al., 2021; Weatherholtz, 2015; Xie et al., 2018)
- **Social factors** (e.g. Babel & Russell, 2015; McLaughlin & Van Engen, 2023; Staggs et al., 2022; Vaughn, 2019)

What can perceptual learning tell us about phonology?

Phonemic Representations

There are many theories that posit what constitutes and establishes a phoneme.

Phonemes should be distinctive in the phonology, but should be adaptable to various phonetic inputs.

“Categorical representations should enable flexible utilization of multiple levels of the speech network to improve speech perception in noise.” (Baese-Berk et al., 2022, pg. 3032)

What does perceptual learning tell us?

Research into perceptual learning can corroborate the claim that these representations are flexible to phonetic input.

Novel phonetic realizations can be applied to already existing phonological realizations, providing further grounding that these are phonological categories within the mind.

What’s missing?

Previous Gaps

Research mainly focuses adaptation to different accents in English.

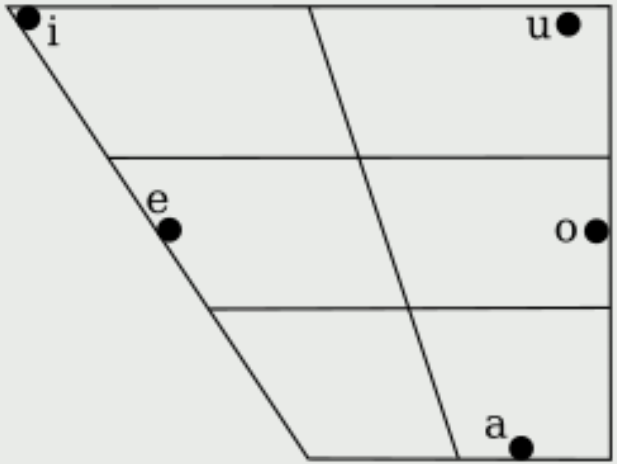
There seems to be an assumption that perceptual learning is inherent and consistent across languages.

The current study researches the perceptual learning of two vowel shifts in Spanish for two reasons:

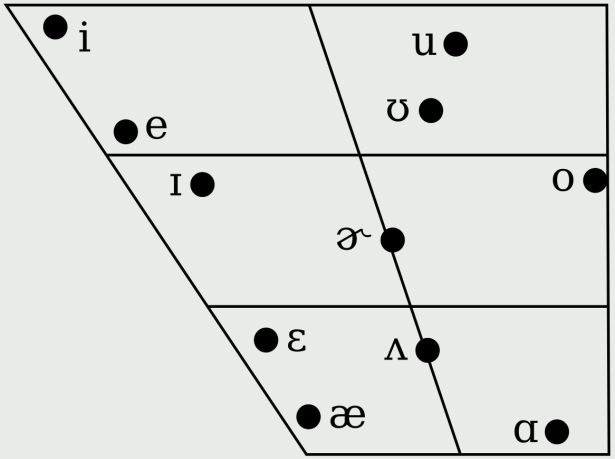
- Spanish has a very different vowel space
- Spanish listeners are rarely exposed to cross-phonemic variation

Spanish vs English vowel space

Spanish has a 5-vowel system



English has a ~11-vowel system



Vowel Variation

English (Labov, 2010):

- Substantial dialectal vowel variation
- Has variation across vowel phonemes
- Vowel variation is something people are regularly exposed to

Vowel Variation

English (Labov, 2010):

- Substantial dialectal vowel variation
- Has variation across vowel phonemes
- Vowel variation is something people are regularly exposed to

Spanish (Hualde, 2014):

- Almost no vocalic variation
- Listeners mainly listen to consonant or prosodic variation

The Current Study

What am I asking?

Q: Do Spanish listeners behave similarly to English listeners when adapting to a vowel chain shift?

H: Spanish listeners will adapt at different rates than English speakers due to the lack of vowel variation and bigger vowel space.

Previous studies on vowel chain shift adaptation

Maye et al., (2008):

Weatherholtz (2015):

Previous studies on vowel chain shift adaptation

Maye et al., (2008):

- Looked at two vowel shifts:
  - Front vowel lowering
  - Front vowel raising

- Listeners heard a story with one of these shifts and then completed a lexical decision task.

- Learning is found to have an effect in both conditions, but more so in the lowering condition.

Weatherholtz (2015):

Previous studies on vowel chain shift adaptation

Maye et al., (2008):

- Looked at two vowel shifts:
  - Front vowel lowering
  - Front vowel raising
- Listeners heard a story with one of these shifts and then completed a lexical decision task.
- Learning is found to have an effect in both conditions, but more so in the lowering condition.

Weatherholtz (2015):

- Looked at various vowel shifts
- Similar methods to Maye et al., (2008)
- Found learning in all conditions and generalization to some degree in all conditions.
- Exposed words (“trained words”) were recognized at higher rates.

Experiment 1

Taken from Weatherholtz (2015) and Maye et al., (2008)

Methods: Exposure

Exposure Phase: Participants listened to a story in the exposure phase in one of six conditions:

Time ↓ Shift →	Shifted	Unshifted
2 minutes	*2 minutes shifted*	*2 minutes unshifted*
5 minutes	*5 minutes shifted*	*5 minutes unshifted*
10 minutes	*10 minutes shifted*	*10 minutes unshifted*

When talking about exposure, I will refer to shifted as people who listened to the story with the shift.

Taken from Weatherholtz (2015) and Maye et al., (2008)

Methods: Testing Learning

Lexical Decision Task: Listeners were asked to judge audio clips on their lexicality,

i.e. Participants hear an audio clip → then, were asked to press 1 on their keyboard if it was a word and 0 if not.

Taken from Weatherholtz (2015) and Maye et al., (2008)

Methods: Testing Learning

Three types of audio clips were played during this task:

- Critical Words: words with the vowel shift (n = 60)
  - Trained vs not (n = 20 vs 40)
  - e.g. ‘dog’ /pero/ → [paru]

- Control Words: words without the vowel shift (n = 100)
  - e.g. ‘dog’ /pero/ → [pero]

- Control Nonwords: phonologically-licit maximally nonwords (n = 60)
  - e.g. /plima/ → [plima]

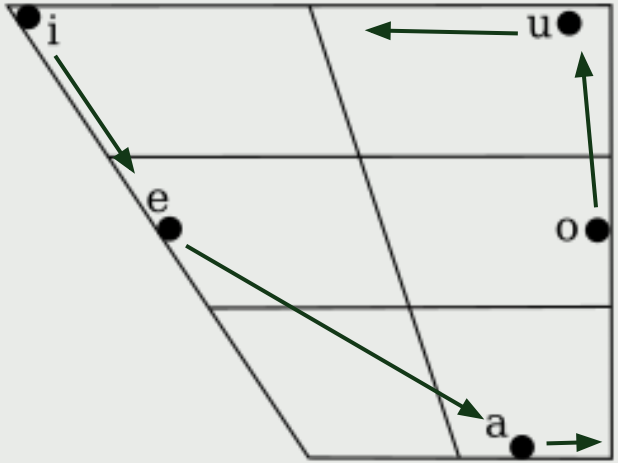
Stimuli

A counter-clockwise shift of the vowel space was implemented. For example,

- ‘pine’ /pino/ → [penu]

Vowels were shifted using praat-parselmouth. (Jadoul et al., 2018; Boersma & Weenink, 2025)

The speaker was an L1 Mexican Spanish speaker.



Results

How I will be talking about the data

Exposure condition:

- Shifted
- Unshifted

Dependent Variable:

Endorsement Rate (i.e. rate of ‘word’ response)

Statistical Model:

Bayesian Logistic Regression

Lexical items:

- Control words
- Control nonwords (maximal nonwords)
- Critical words
  - Trained (were present in the story)
  - New (were not present in the story)

Vowels:

Referring to the phonemic vowel
(i.e. the /e/→[a] shift will be referred to as ‘e’)

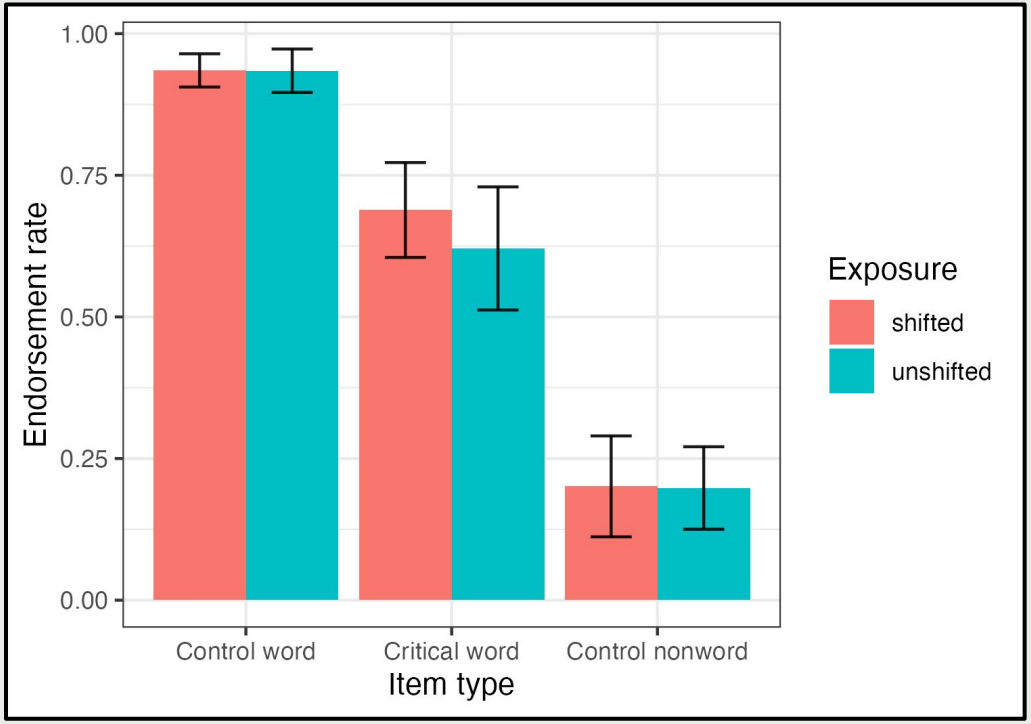
Endorsement Rates

109 L1 Mexican Spanish speakers were recruited.

Across conditions and word type.

Mean endorsement rates.

Error bars are 1 standard deviation away from the mean.



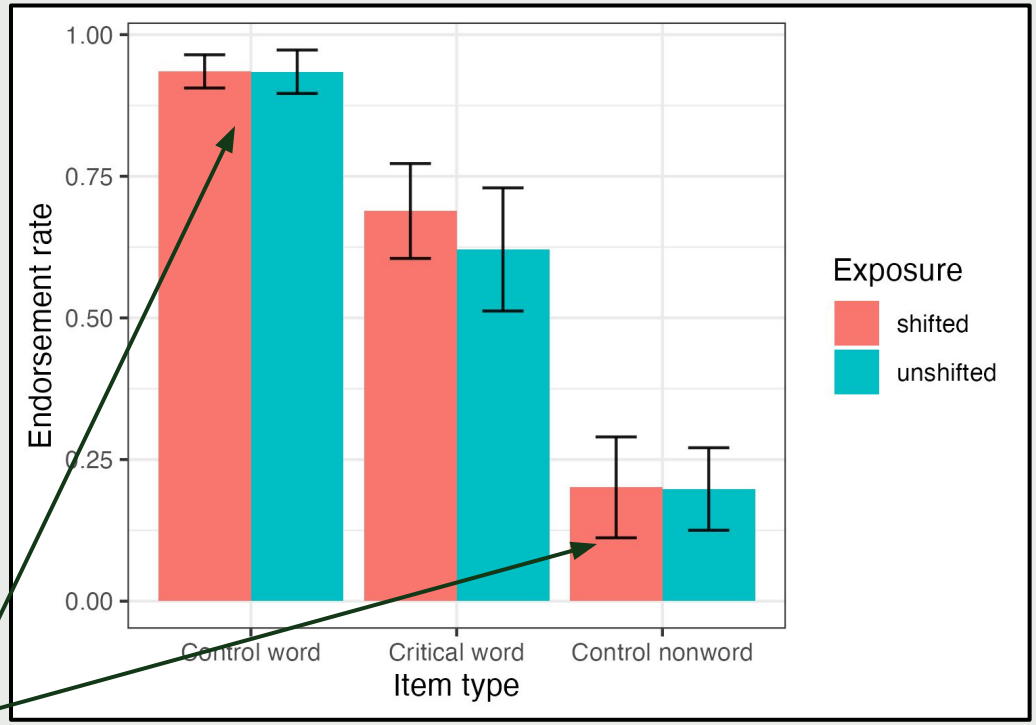
Endorsement Rates

120 L1 Mexican Spanish speakers were recruited.

Across conditions and word type.

Mean endorsement rates.

Error bars are 1 standard deviation away from the mean.



Expected outcome

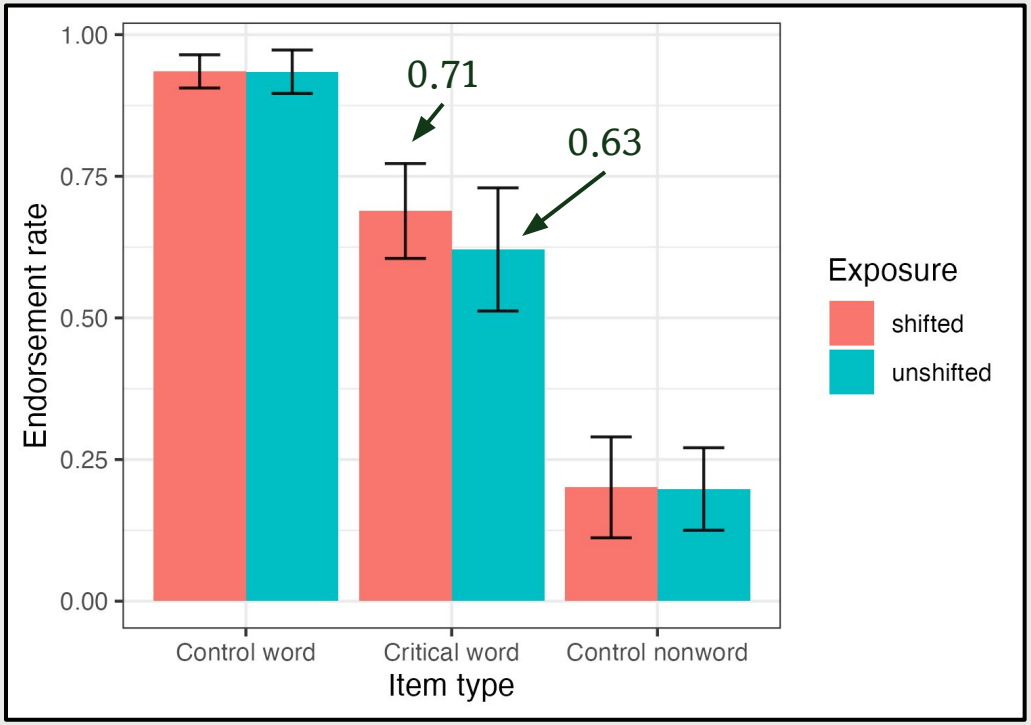
Endorsement Rates

120 L1 Mexican Spanish speakers were recruited.

Across conditions and word type.

Mean endorsement rates.

Error bars are 1 standard deviation away from the mean.



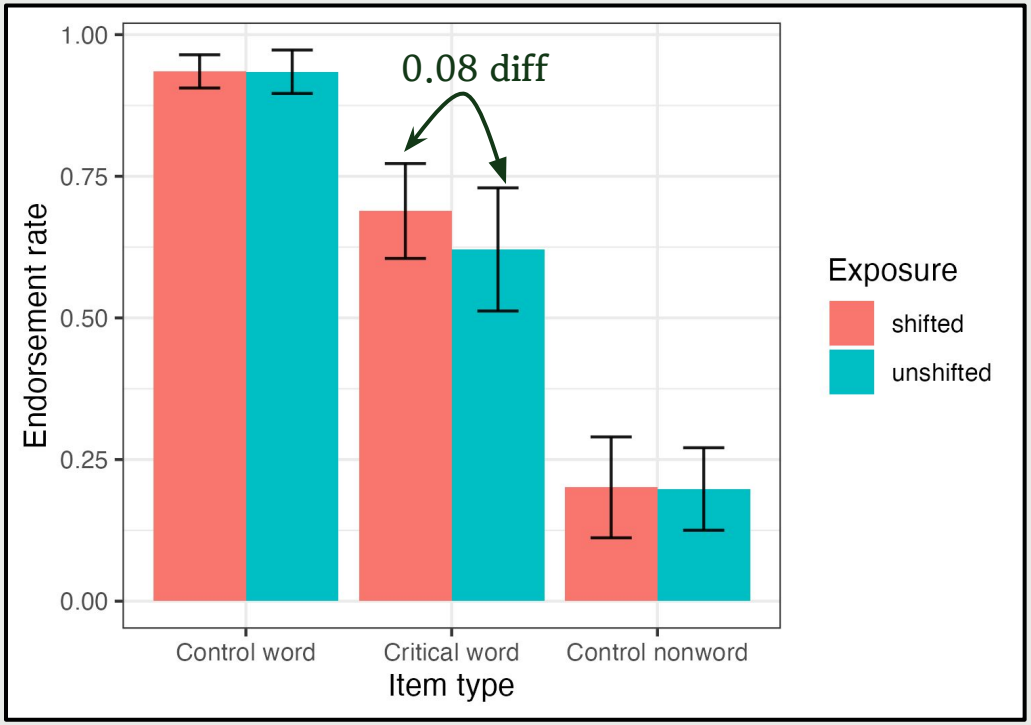
Endorsement Rates

120 L1 Mexican Spanish speakers were recruited.

Across conditions and word type.

Mean endorsement rates.

Error bars are 1 standard deviation away from the mean.



Endorsement Rates

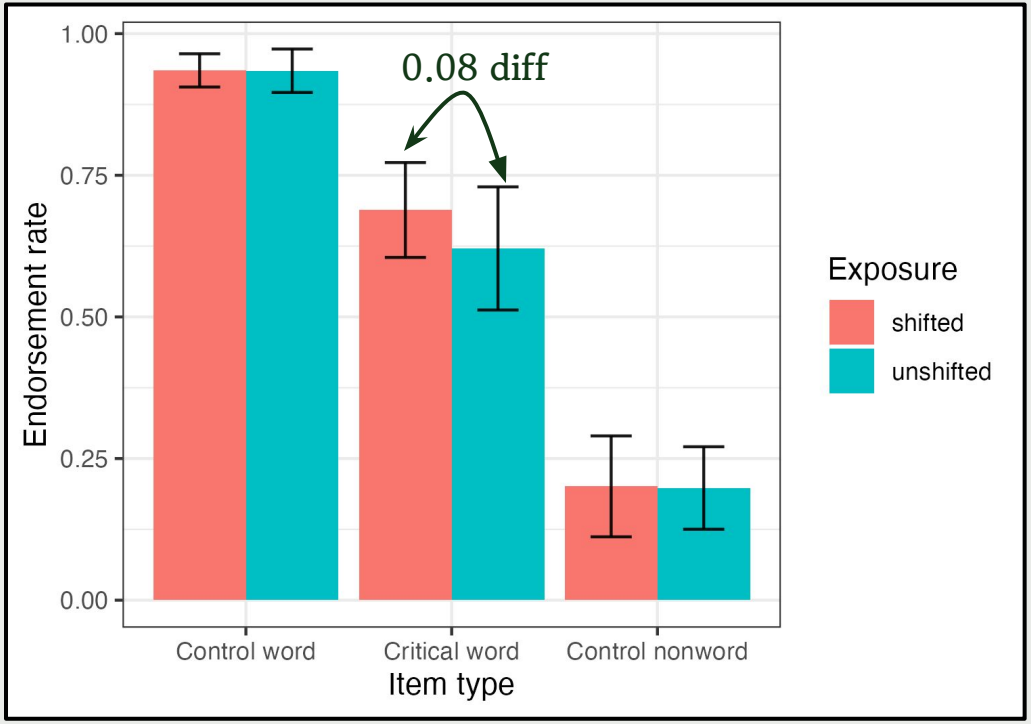
120 L1 Mexican Spanish speakers were recruited.

Across conditions and word type.

Mean endorsement rates.

Error bars are 1 standard deviation away from the mean.

<u>This is not a big change</u> compared to previous studies



Endorsement Rates

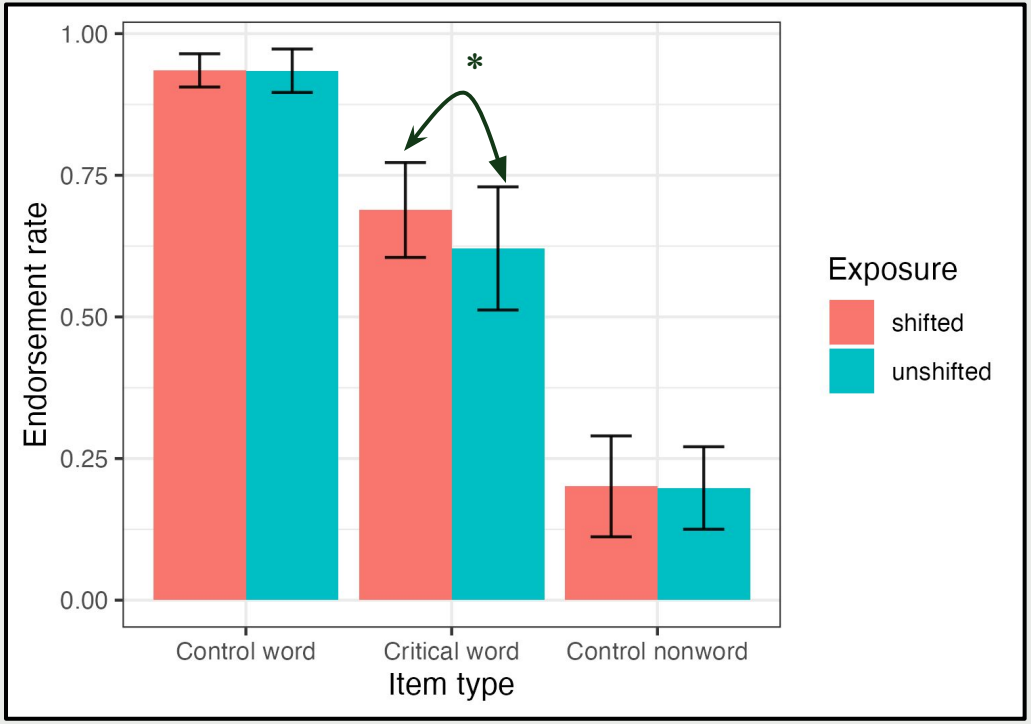
120 L1 Mexican Spanish speakers were recruited.

Across conditions and word type.

Mean endorsement rates.

Error bars are 1 standard deviation away from the mean.

<u>This is not a big change</u> compared to previous studies



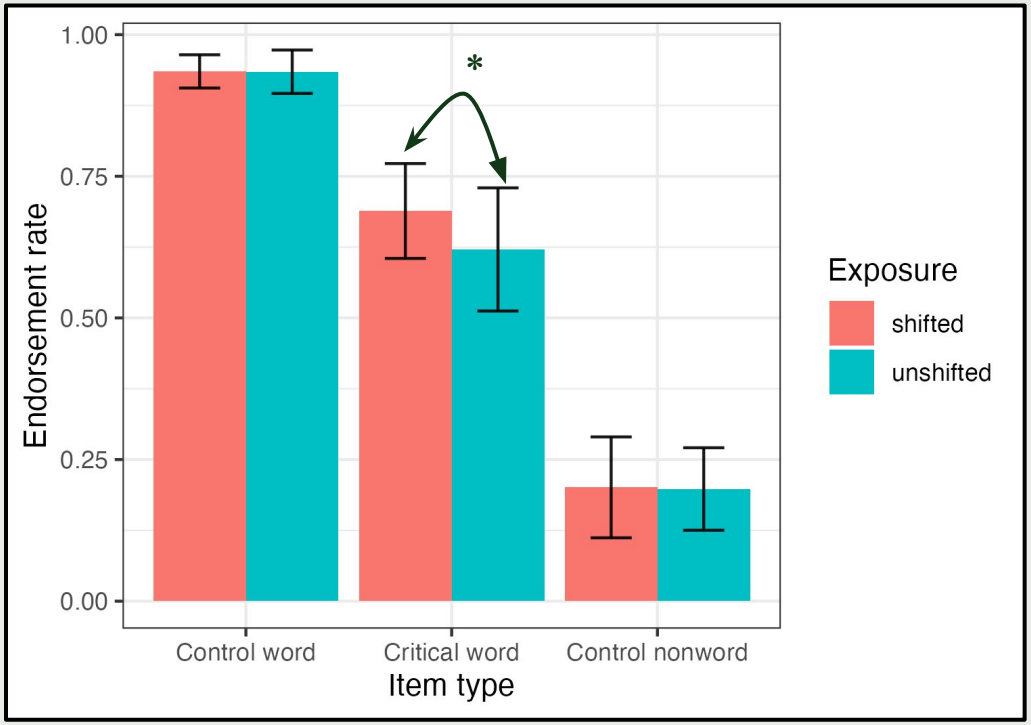
Endorsement Rates

120 L1 Mexican Spanish speakers were recruited.

Across conditions and word type.

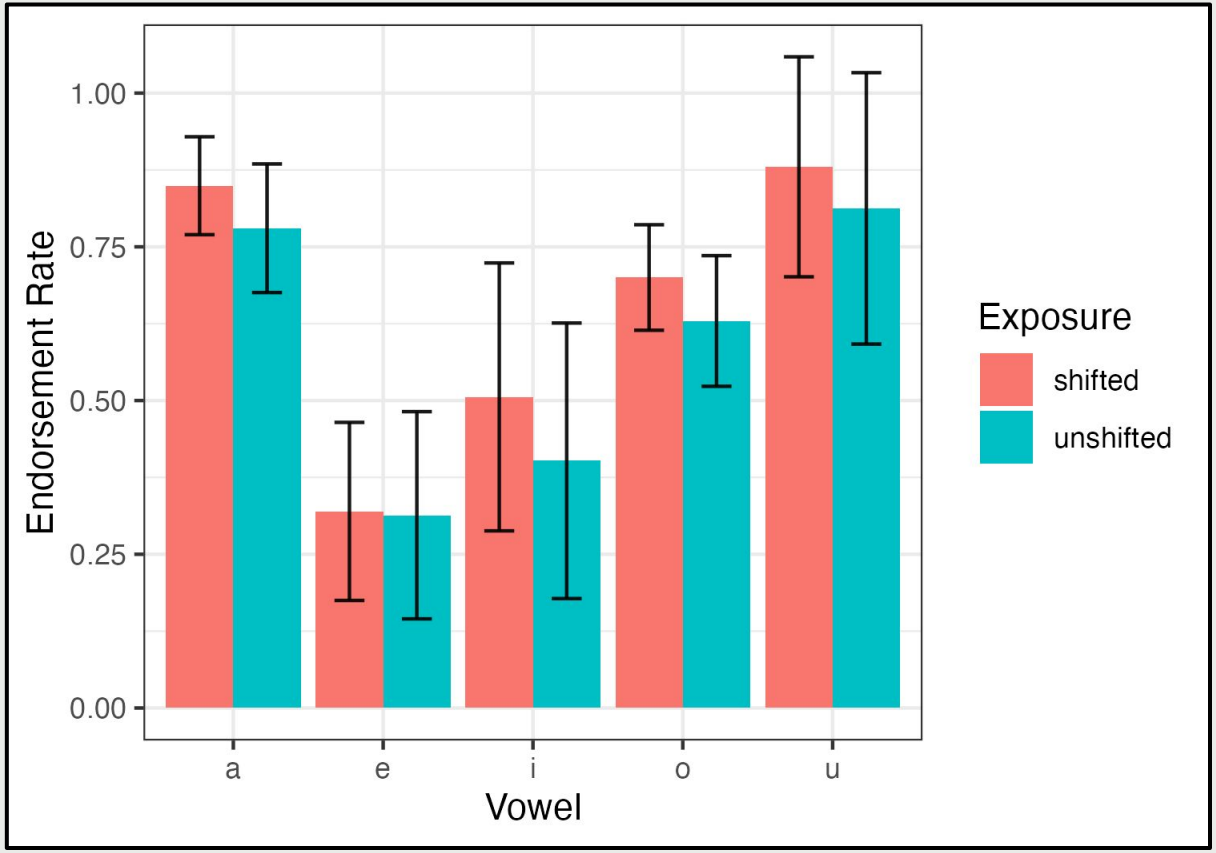
This is not a big change compared to previous studies

Time was not a reliable predictor of endorsement rates.



Vowel-Specific Rates

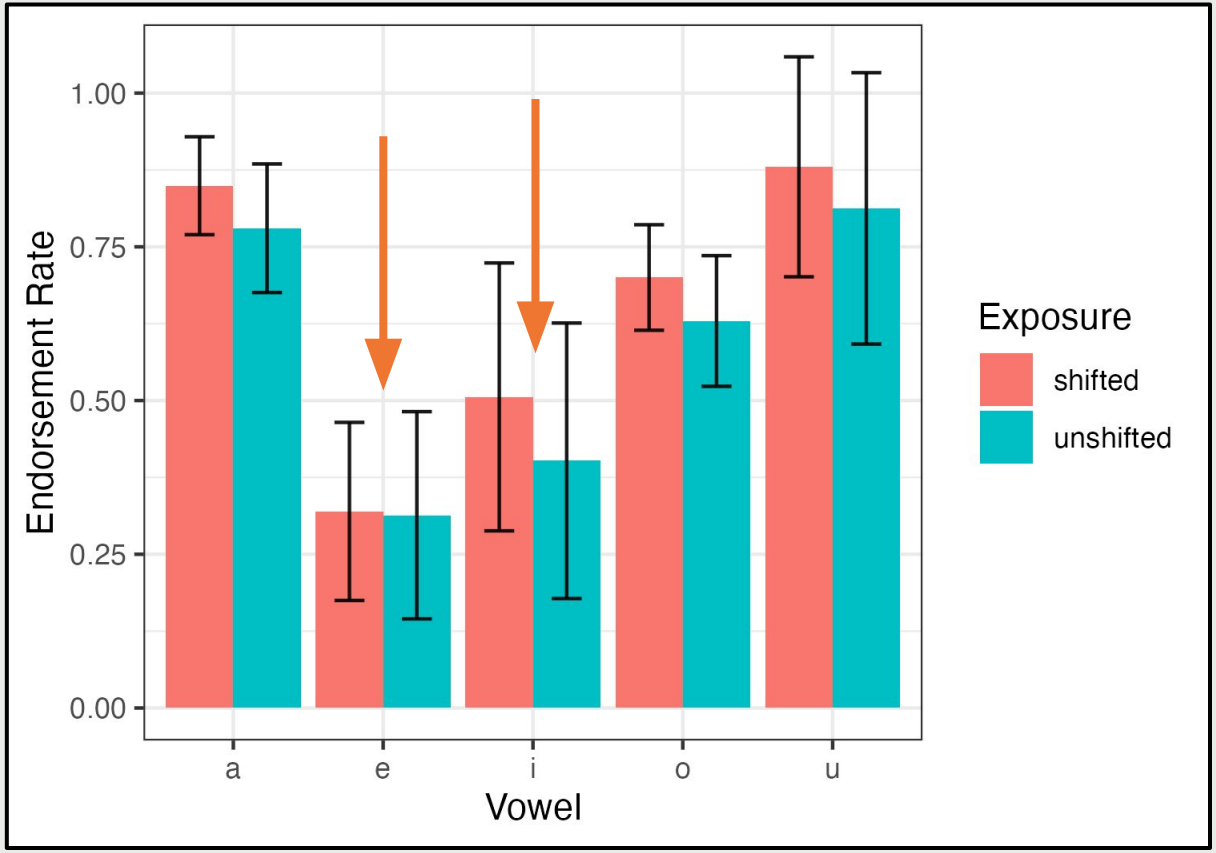
Vowels by condition



/e/ and /i/ have the lowest endorsement rates

Vowel-Specific Rates

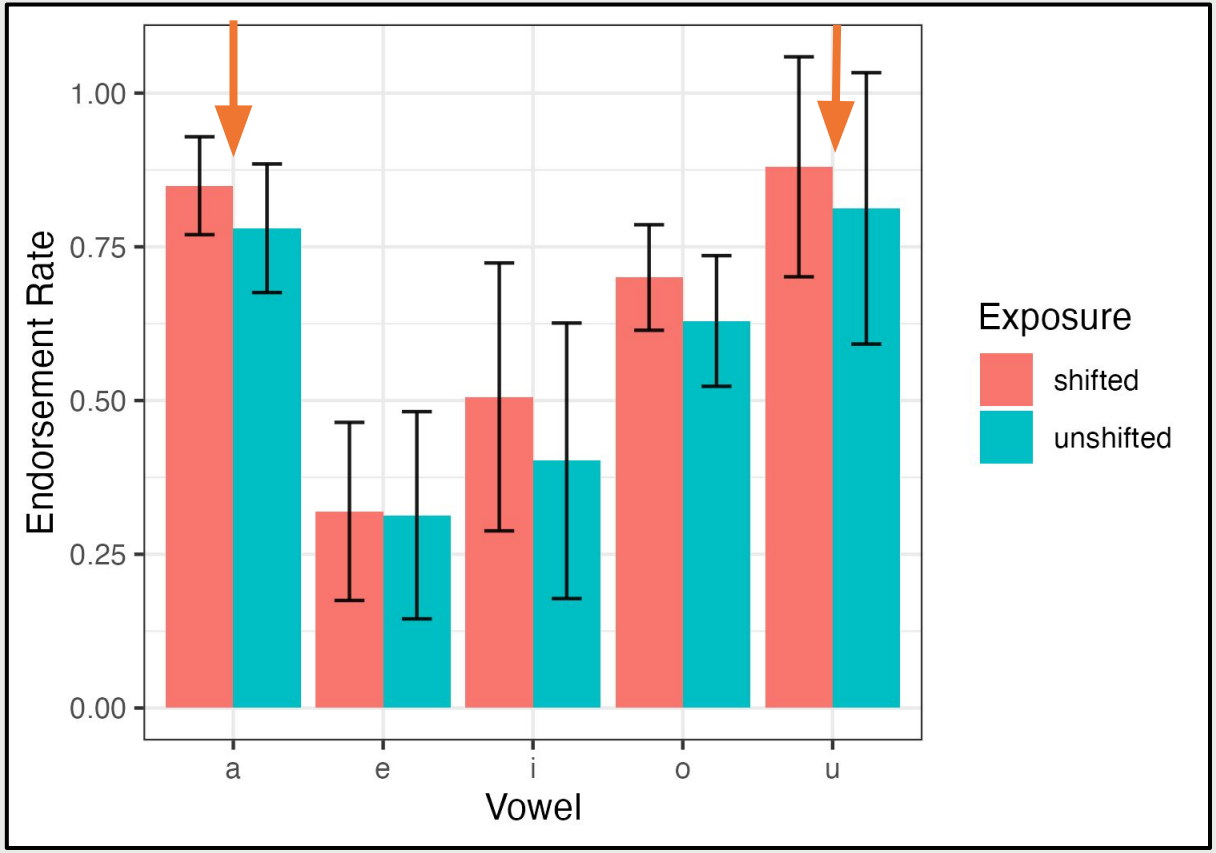
Vowels by condition



/a/ and /u/ have the highest endorsement rates

Vowel-Specific Rates

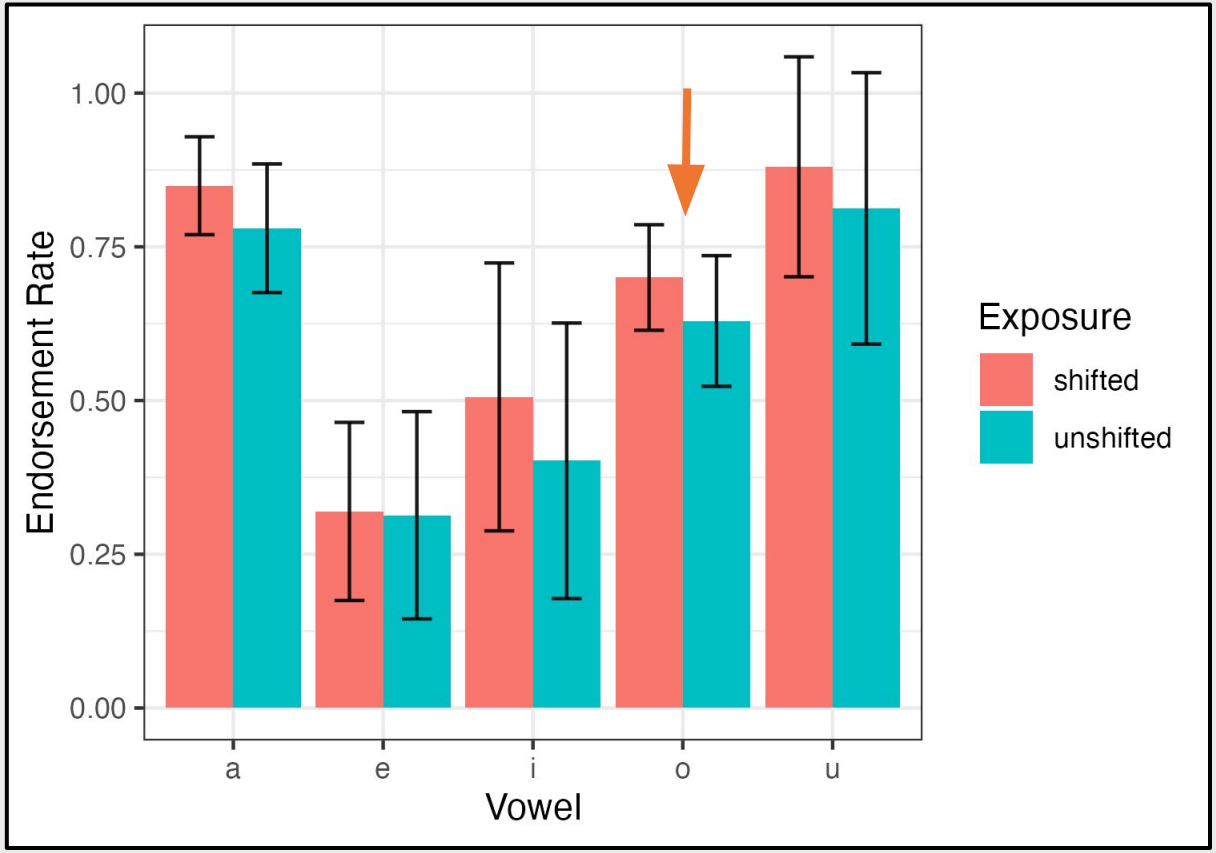
Vowels by condition



/o/ is somewhere in the middle

Vowel-Specific Rates

Vowels by condition



Bayesian Model for vowels in critical words

response ~ exposure * vowel * trained + (1 | Trial) + (1 | Participant)

Model Results:

- All vowels were significantly different from each other, except for /a/ and /u/.
- Exposure does not change the effect of vowel in model except for /a/:
  - However, pairwise comparisons show that /i/ and /o/ may have higher endorsement rates in the shifted condition
- Trained words were not found to be significant

Interim discussion

Vowels have different endorsement rates and potentially different learning effects.

There seems to be a learning effect overall, but not as big of an effect of suggested in previous research, also only present in pairwise comparison.

A more targeted approach needs to be taken to understand the vowel trends.

Experiment 2

Changes to the methodology

Two separate vowel shifts were analyzed separately:

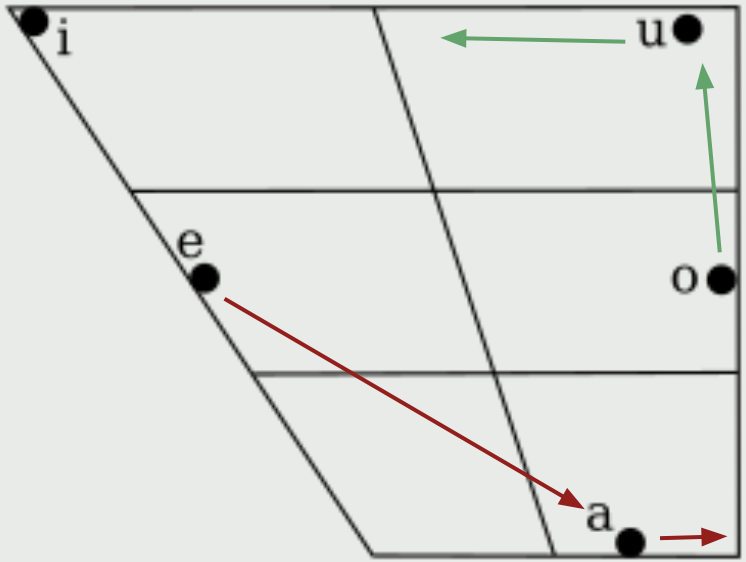
- Front Vowel Shift (FVS)
- Back Vowel Shift (BVS)

Exposure phase was only 5 minutes

Critical items were divided into two groups:

- Cross-phonemic (i.e. /e/ or /o/)
- Not cross-phonemic (i.e. /a/ or /u/)

80 participants of the same demographic as Exp 1 were recruited.

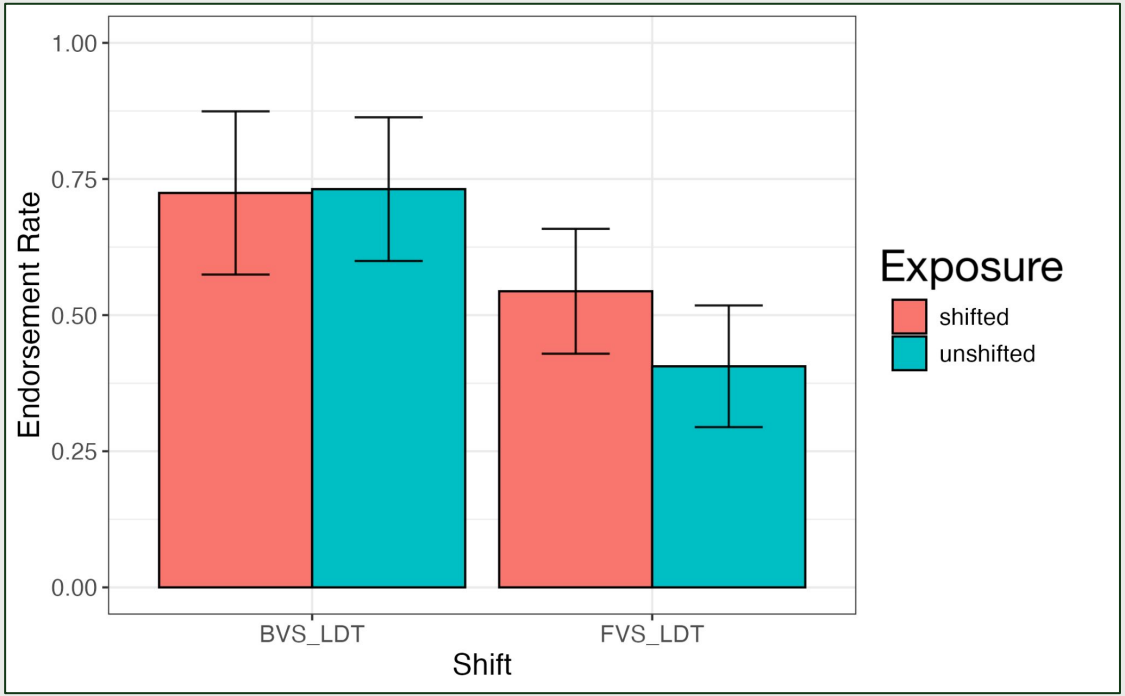


Results

Endorsement Rates across the two shifts

The BVS group shows no effect of exposure

The FVS group does show an effect of exposure

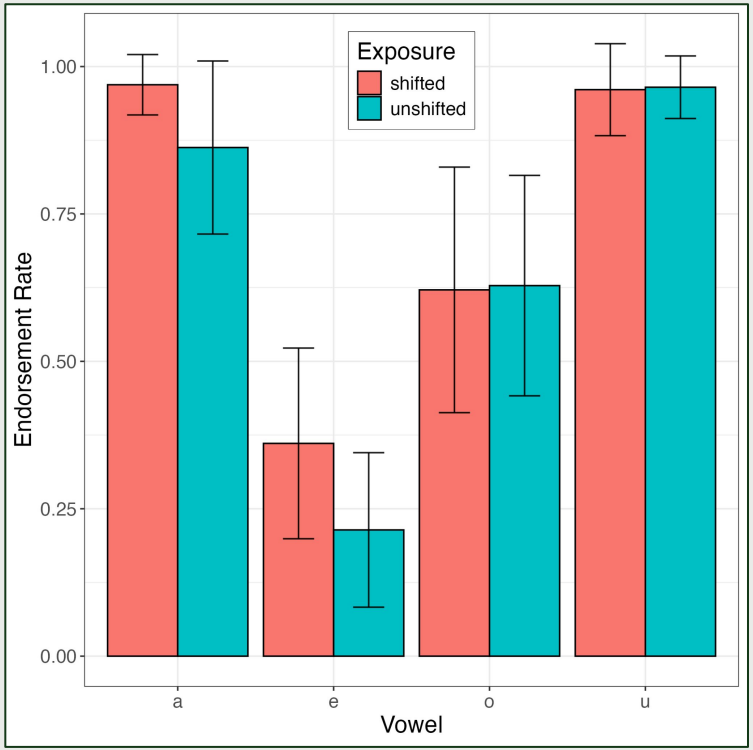


Vowel-Specific Endorsement Rates

Similar to last experiment in average endorsement rates.

Only /a/ show significant effects of exposure while /e/, /o/ and /u/ do not.

Additionally, /e/ and /o/ have lower endorsement rates overall compared to the two other vowels

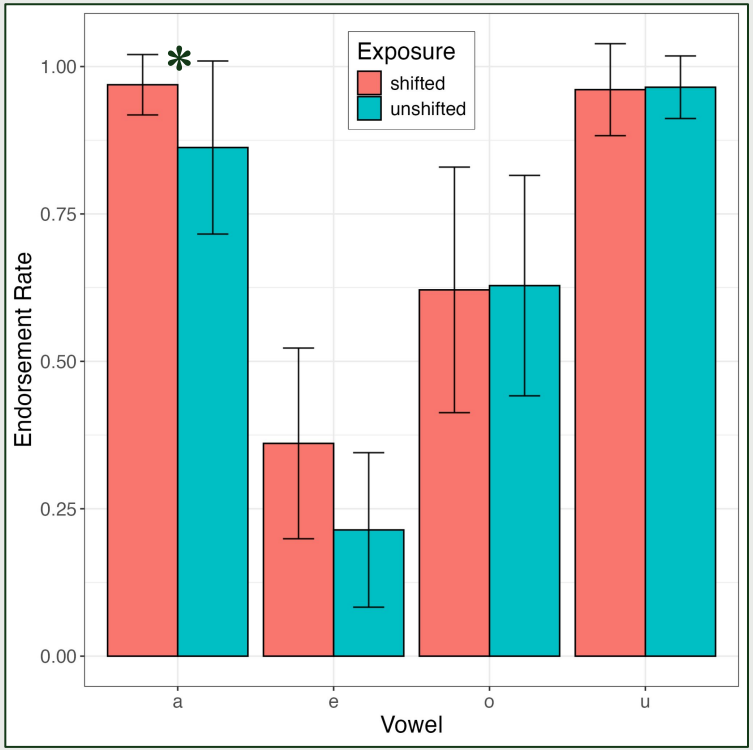


Vowel-Specific Endorsement Rates

Similar to last experiment in average endorsement rates.

Only /a/ show significant effects of exposure while /e/, /o/ and /u/ do not.

Additionally, /e/ and /o/ have lower endorsement rates overall compared to the two other vowels

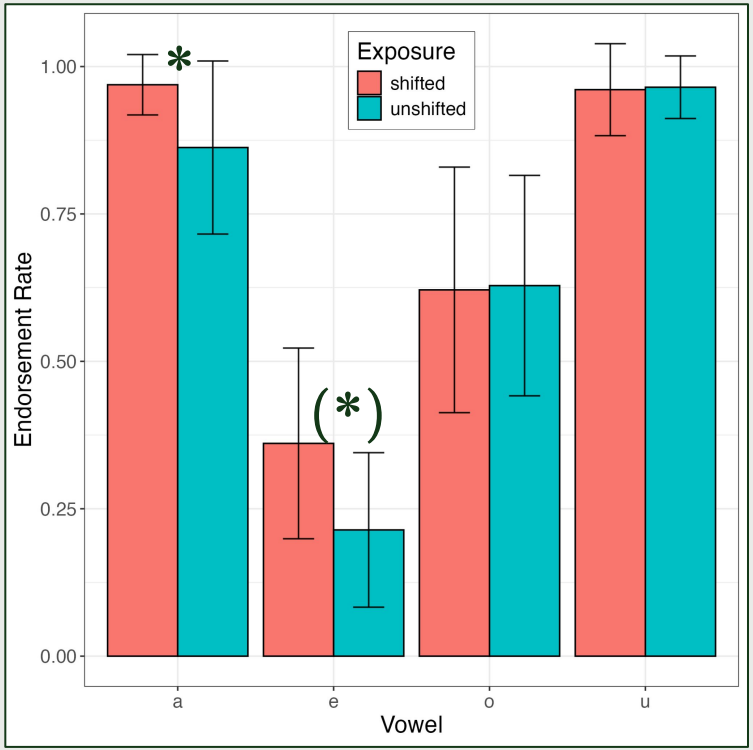


Vowel-Specific Endorsement Rates

Similar to last experiment in average endorsement rates.

Only /a/ show significant effects of exposure while /e/, /o/ and /u/ do not.

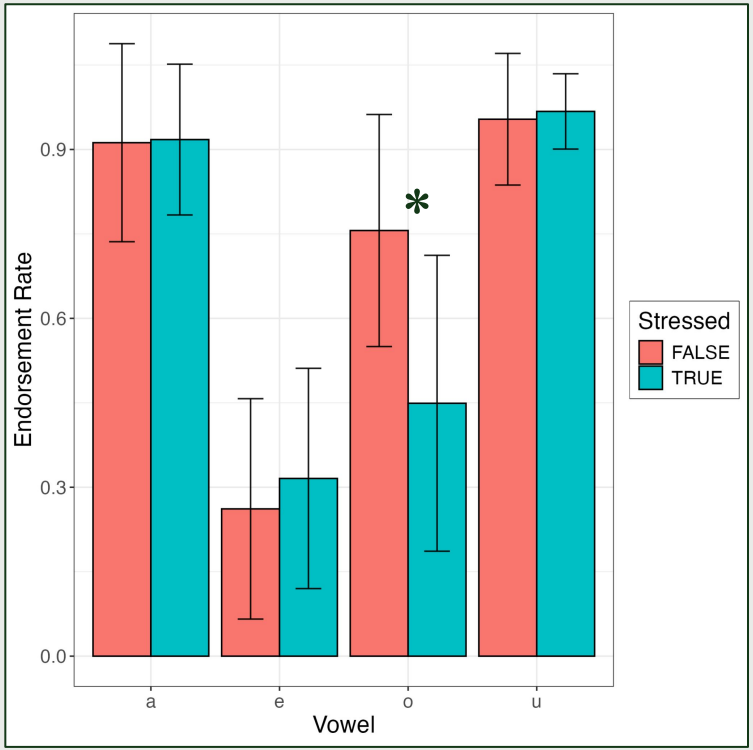
(*) /e/ exposure is found to be significant only in pairwise



Vowel-Specific Endorsement Rates: stress

Stress was a significant predictor of endorsement rates only for /o/:

- Stressed /o/ is endorsed at lower rates.



Discussion & Conclusion

Discussion

Findings:

- Exposure has less of an effect on Spanish listeners than previously reported for English listeners.
- Different vowels may have different breadths of acceptable variation.
- Stress may be important in some vowel representations.
- It is still unclear what conditions learning for Spanish listeners.
- No effect of presence in the story for critical words.

Difference from English

There are two main reasons why Spanish listeners may be behaving differently:

Difference from English

There are two main reasons why Spanish listeners may be behaving differently:

1. The Spanish vowels space is more dispersed, meaning more movement to create this vowel shift.

Difference from English

There are two main reasons why Spanish listeners may be behaving differently:

1. The Spanish vowels space is more dispersed, meaning more movement to create this vowel shift.

2. Less exposure to variation. Spanish listeners adapt at lower rates because they do not need to. Vowel variation is not expected, so there is no reason to retain these variants.

Difference from English

There are two main reasons why Spanish listeners may be behaving differently:

1. The Spanish vowels space is more dispersed, meaning more movement to create this vowel shift.

2. Less exposure to variation. Spanish listeners adapt at lower rates because they do not need to. Vowel variation is not expected, so there is no reason to retain these variants.

These two reasons are not divorced from each other. One can influence the other.

Vowel Differences

Vowel Differences

Functional load: /o/ and /u/ have much less minimal pairs than /e/ and /a/ causing more lexical competitors. This could make listeners less willing to accept [a] as a pronunciation of /e/.

Vowel Differences

Functional load: /o/ and /u/ have much less minimal pairs than /e/ and /a/ causing more lexical competitors. This could make listeners less willing to accept [a] as a pronunciation of /e/.

Individual phoneme plasticity: Some representations may just have stricter boundaries. This could be due to exposure to variation generally, or vowel space characteristics.

Vowel Differences

Functional load: /o/ and /u/ have much less minimal pairs than /e/ and /a/ causing more lexical competitors. This could make listeners less willing to accept [a] as a pronunciation of /e/.

Individual phoneme plasticity: Some representations may just have stricter boundaries. This could be due to exposure to variation generally, or vowel space characteristics.

(Perceptual) distance: /e/ and /a/ may be perceptually further away from each other than /o/ and /u/ causing less willingness to accept them as variants of each other.

No effect of Trained Items

No effect of Trained Items

It may be really hard for listeners to understand the speech given in the exposure phase.

→ It may be hard for the listeners to parse out every word given.

No effect of Trained Items

It may be really hard for listeners to understand the speech given in the exposure phase.

→ It may be hard for the listeners to parse out every word given.

Listeners may focus on learning the shift itself (i.e. phonemic level learning), rather than the individual lexical items (i.e. lexical level learning) to make it more generalizable.

Conclusion

Whether this is due to differences in vowel space, or experience with vowel variation remains unclear.

We can see that Spanish listeners behave differently than English listeners in previous studies.

The way categories exist may differ based on the listener and/or language.

Overall, a more fine-grained assessment about the listener-dependent effects on perceptual learning needs to be undertaken.

Acknowledgements

Molly Babel for guidance on conceptualisation and design.

Justin Davidson and Keith Johnson for being my readers.

Various members of the Linguistics, and Spanish and Portuguese departments for feedback on the presentation.

Santiago Barreda and Molly Babel for helping with the modelling.

Email: niko.andres@berkeley.edu

Gracias

Appendix

Bayesian Model: Experiment 1 conditions

Model information: response ~ item * time * exposure + (1 | participant) + (item | trial)

Family: Bernoulli

Priors:
intercept - normal(0,4)

itemControlWords - normal(1,2)

exposureunshifted - normal(-1,2)

Term	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-1.97	0.28	-2.5	-1.43	1.01	721	1593
itemControlword	5.76	0.32	5.13	6.39	1.01	872	1651
itemCriticalword	3.64	0.35	2.95	4.33	1	590	1407
time2min	0.27	0.2	-0.13	0.66	1	2247	3926
time5min	0.26	0.2	-0.13	0.64	1	2123	4468
exposureunshifted	0.16	0.21	-0.25	0.57	1	2090	3770
itemControlword:time2min	-0.12	0.16	-0.42	0.19	1	5832	8973
itemCriticalword:time2min	-0.54	0.14	-0.8	-0.27	1	6165	8271
itemControlword:time5min	-0.15	0.15	-0.46	0.14	1	6320	8580
itemCriticalword:time5min	-0.38	0.13	-0.63	-0.12	1	5861	8065
itemControlword:exposureunshifted	-0.18	0.16	-0.5	0.13	1	5218	7786
itemCriticalword:exposureunshifted	-0.89	0.14	-1.17	-0.61	1	4868	7639
time2min:exposureunshifted	-0.35	0.29	-0.93	0.23	1	2080	3941
time5min:exposureunshifted	-0.21	0.29	-0.77	0.38	1	1831	3469
itemControlword:time2min:exposureunshifted	0.38	0.22	-0.06	0.81	1	5316	8028
itemCriticalword:time2min:exposureunshifted	0.74	0.19	0.35	1.11	1	5179	8244
itemControlword:time5min:exposureunshifted	0.24	0.23	-0.2	0.68	1	5665	8405
itemCriticalword:time5min:exposureunshifted	0.37	0.19	-0.01	0.76	1	4965	7441

Bayesian Model: Experiment 1 vowels

Model information:

Syntax: exposure * vowel + exposure * trained + (1 | trial) + (1 | Participant)
Priors:

- intercept - normal(0,4)
- exposureshifted - normal(0.1603214, 0.2103827)

Family: Bernoulli

Term	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	1.79	0.09	1.61	1.97	1	4819	7619
exposureunshifted	-0.28	0.11	-0.5	-0.05	1	5029	6968
Vowele	-2.63	0.12	-2.87	-2.39	1	7677	8945
Voweli	-1.67	0.1	-1.87	-1.47	1	8155	8845
Vowelo	-0.78	0.08	-0.94	-0.62	1	7293	8495
Vowelu	0.27	0.22	-0.15	0.71	1	10286	8077
TrainedTRUE	-0.05	0.07	-0.19	0.08	1	10232	9469
exposureunshifted:Vowele	0.2	0.17	-0.14	0.54	1	7134	9016
exposureunshifted:Voweli	-0.12	0.14	-0.39	0.16	1	8421	9122
exposureunshifted:Vowelo	-0.02	0.11	-0.23	0.2	1	7065	7938
exposureunshifted:Vowelu	-0.14	0.29	-0.71	0.42	1	9702	8145
exposureunshifted:TrainedTRUE	0.06	0.1	-0.13	0.26	1	10206	8883

Bayesian Model: Experiment 2 Critical Items

Model information: **not** sum-coded

Syntax: response ~ vowel * stressed * exposure + (1 | Participant) + (vowel|word) + (exposure | trial)

Priors:

- intercept - normal(0,4)
- exposureunshifted - normal(-1,1)

Family: Bernoulli

Intercept:

- /a/
- Unstressed
- Shifted

Parameter	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	3.4	0.56	2.31	4.51	1	5665	7020
vowele	-4.15	0.62	-5.39	-2.94	1	5895	7779
vowelo	-1.53	0.7	-2.91	-0.18	1	5563	7125
vowelu	0.47	0.91	-1.26	2.3	1	8075	7892
stressedyes	0.07	0.64	-1.17	1.34	1	4854	6934
conditionunshifted	-1.24	0.53	-2.28	-0.19	1	5108	7035
vowele:stressedyes	-0.02	0.77	-1.55	1.49	1	5565	6657
vowelo:stressedyes	-2.21	0.84	-3.87	-0.55	1	5779	7728
vowelu:stressedyes	0.31	1.05	-1.77	2.36	1	9027	8998
vowele:exposureunshifted	0.14	0.56	-0.97	1.27	1	5699	7715
vowelo:exposureunshifted	0.86	0.64	-0.4	2.1	1	5133	7241
vowelu:exposureunshifted	1.31	0.94	-0.5	3.19	1	8740	8967
stressedyes:exposureunshifted	0.08	0.63	-1.16	1.28	1	4578	7071
vowele:stressedyes:exposureunshifted	0.36	0.7	-0.99	1.74	1	5005	7013
vowelo:stressedyes:exposureunshifted	0.62	0.71	-0.76	2.01	1	5406	7674
vowelu:stressedyes:exposureunshifted	-0.38	1.09	-2.53	1.73	1	10307	9843

Inclusion/Exclusion Criteria

Participants were recruited via Prolific (prolific.com) and completed the experiment on Gorilla (Anwyl-Irvine et al., 2020)

Participants were all L1 speakers of Mexican Spanish living in Mexico. They were Spanish-dominant and had not lived out of Mexico for more than 5 years.

Participants were excluded from the analysis if they scored less than 80% accurate on control trials (i.e. control words and maximal nonwords)

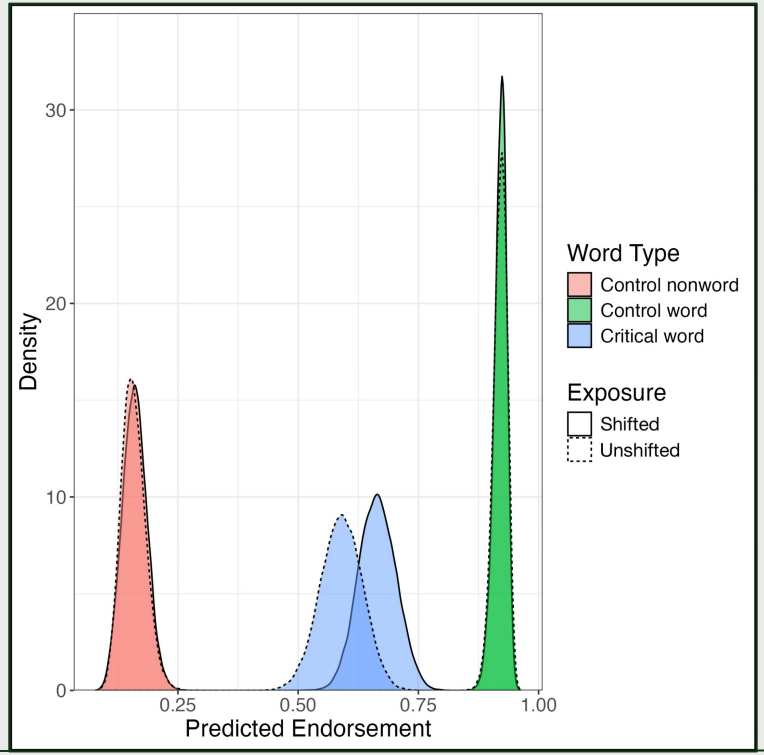
4 critical items were excluded from the analysis as the synthesis warped other phones in the word (e.g. ‘north’ /norte/ mean to be [nurta], but was perceived as [murta]). This was deduced in a separate transcription task.

Bayesian Model

response ~ response ~ item * time * exposure + (1 | participant) + (item | trial)

Model Results:

- Time was not found to be significant

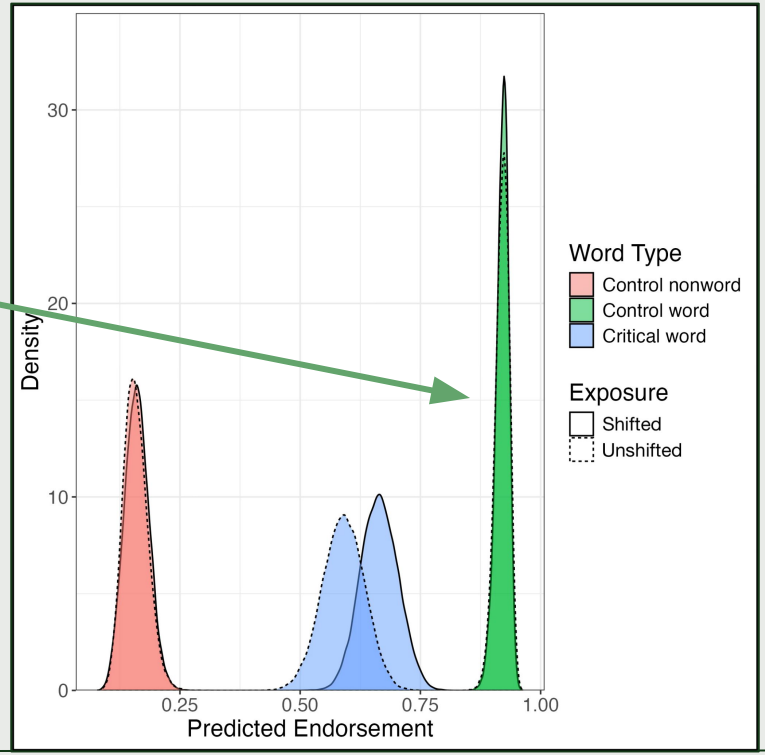


Bayesian Model

response ~ response ~ item * time * exposure + (1 | participant) + (item | trial)

Model Results:

- Time was not found to be significant
- Control items are predicted to have higher endorsement rate

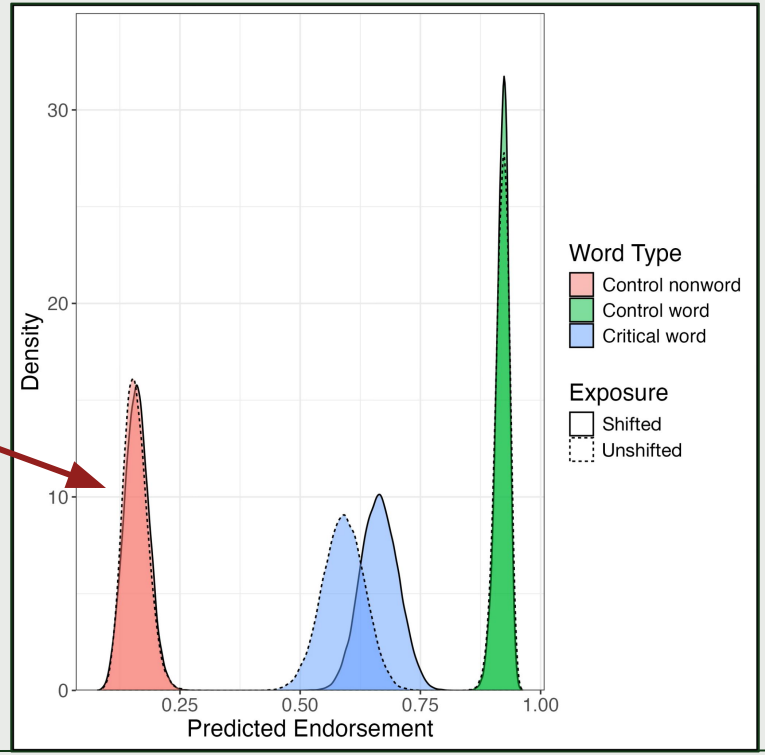


Bayesian Model

response ~ item * time * exposure + (1 | participant) + (item | trial)

Model Results:

- Time was not found to be significant
- Control words are predicted to have higher endorsement rate
- Control nonwords are predicted to have lower endorsement rates

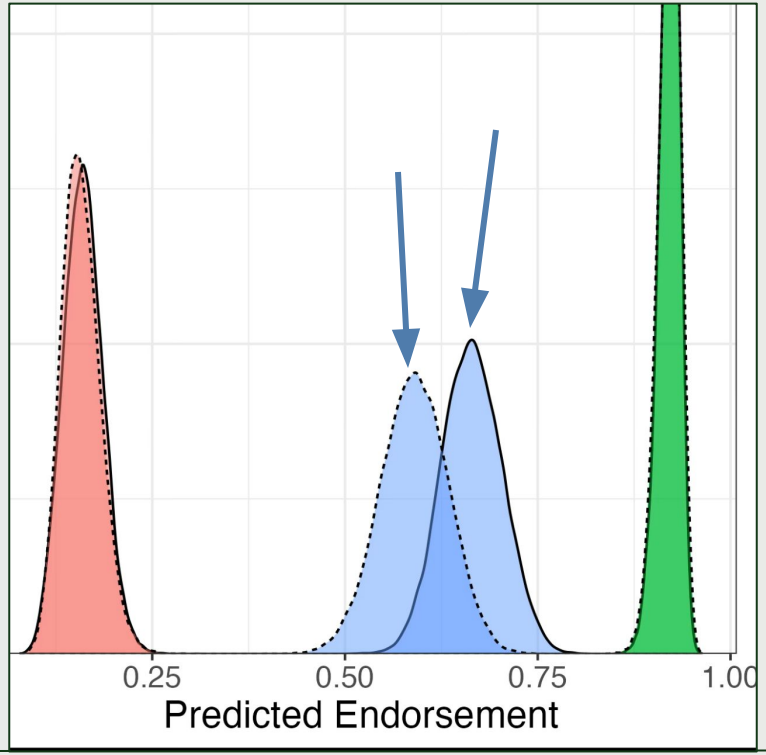


Bayesian Model

response ~ item * time * exposure + (1 | participant) + (item | trial)

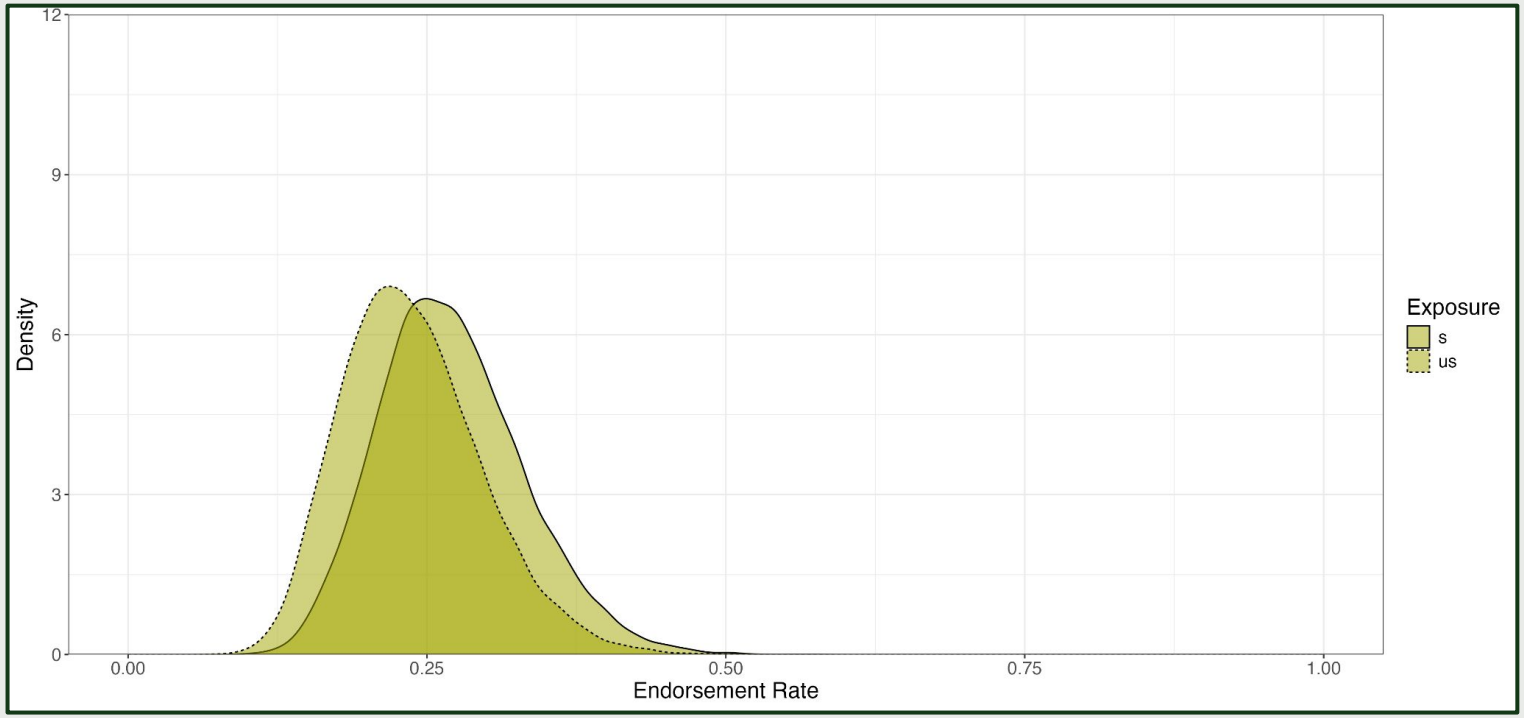
Model Results:

- Time was not found to be significant
- Control words are predicted to have higher endorsement rate
- Control nonwords are predicted to have lower endorsement rates
- Zooming into critical words: they are significantly affected by exposure condition *though not to the same degree as the lit.



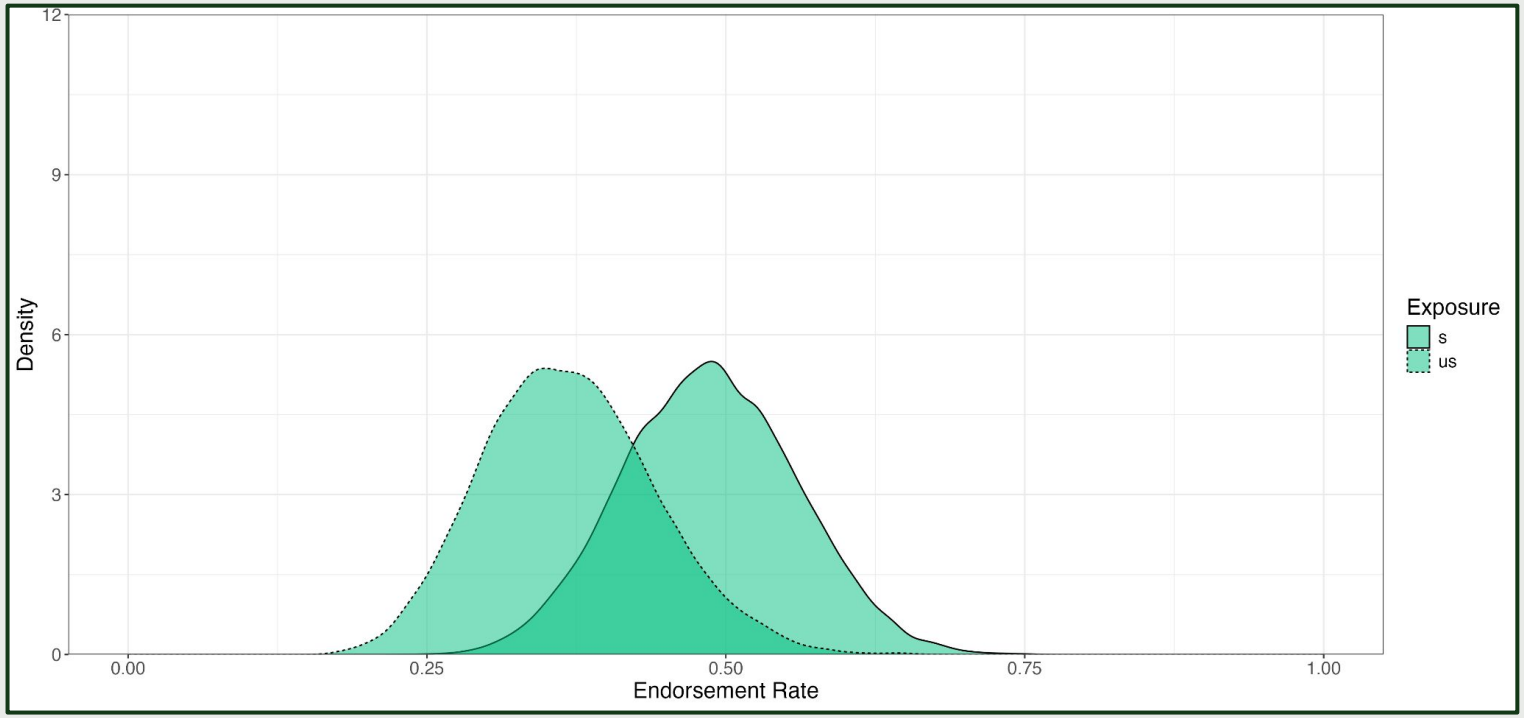
Bayesian Model for vowels in critical words

Model Results: /e/



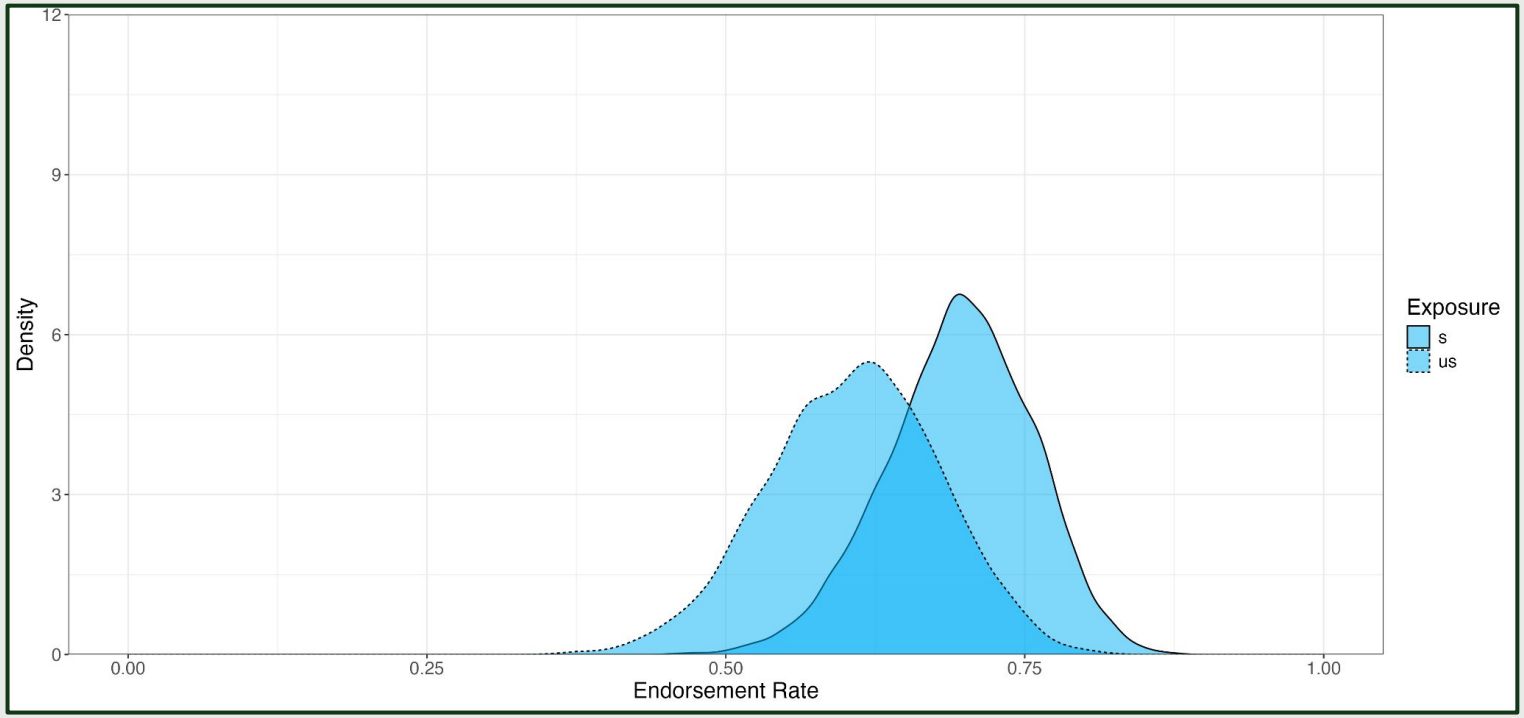
Bayesian Model for vowels in critical words

Model Results: /i/



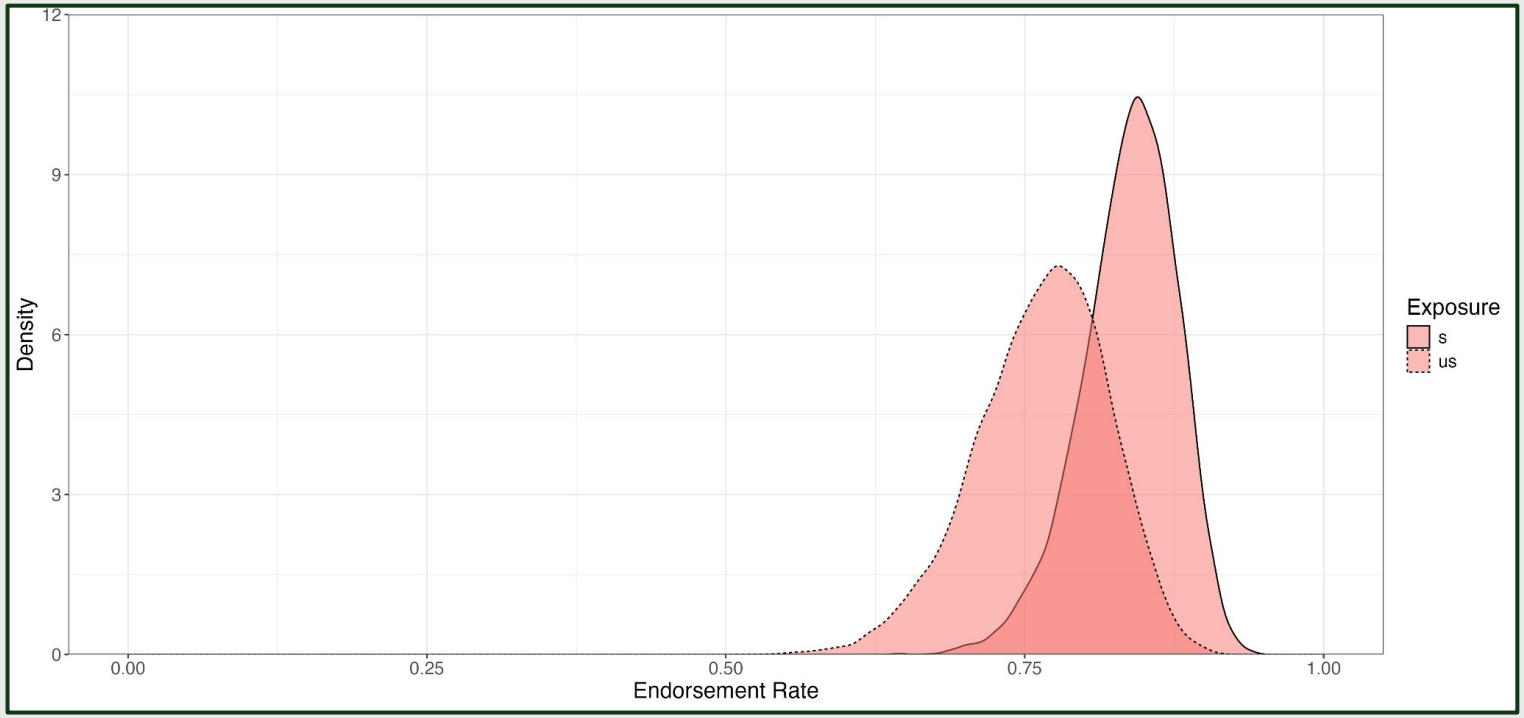
Bayesian Model for vowels in critical words

Model Results: /o/



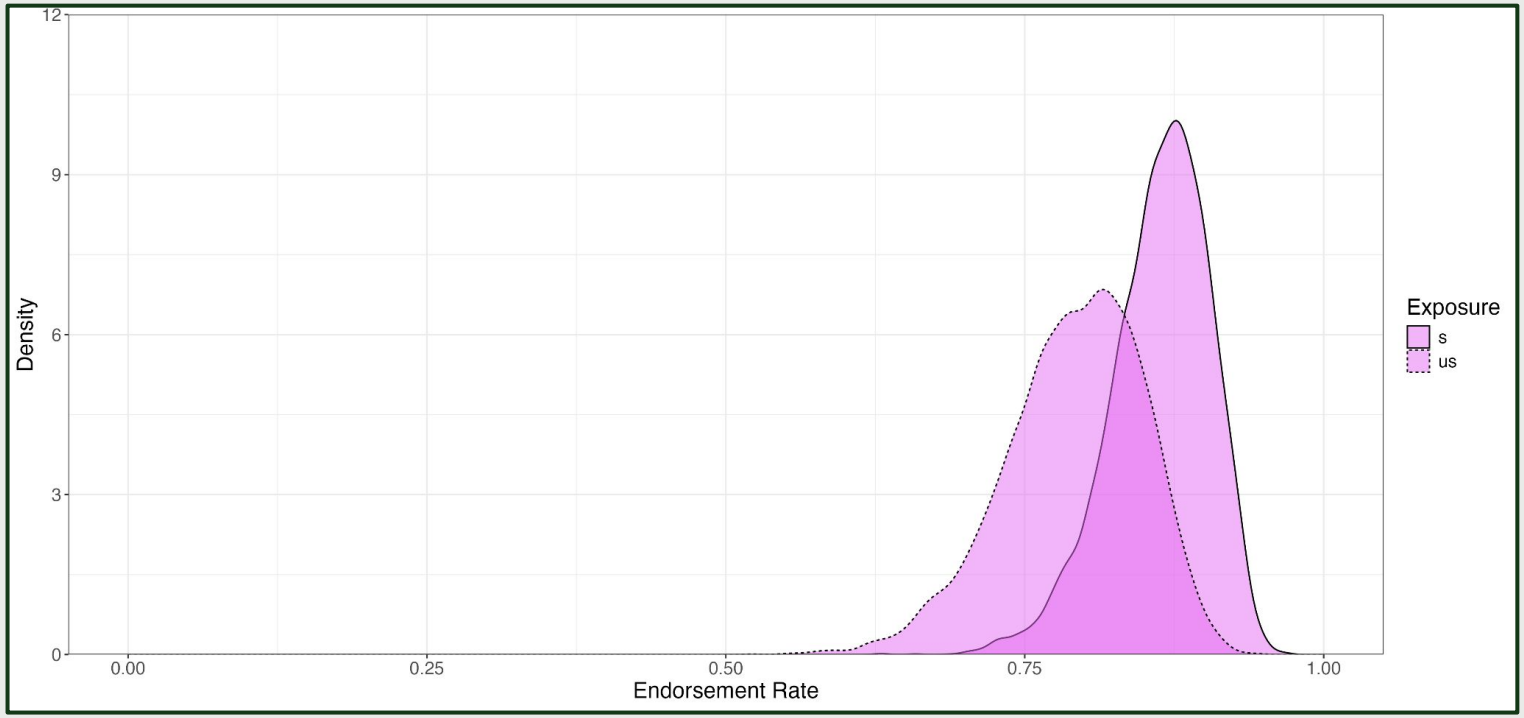
Bayesian Model for vowels in critical words

Model Results: /a/



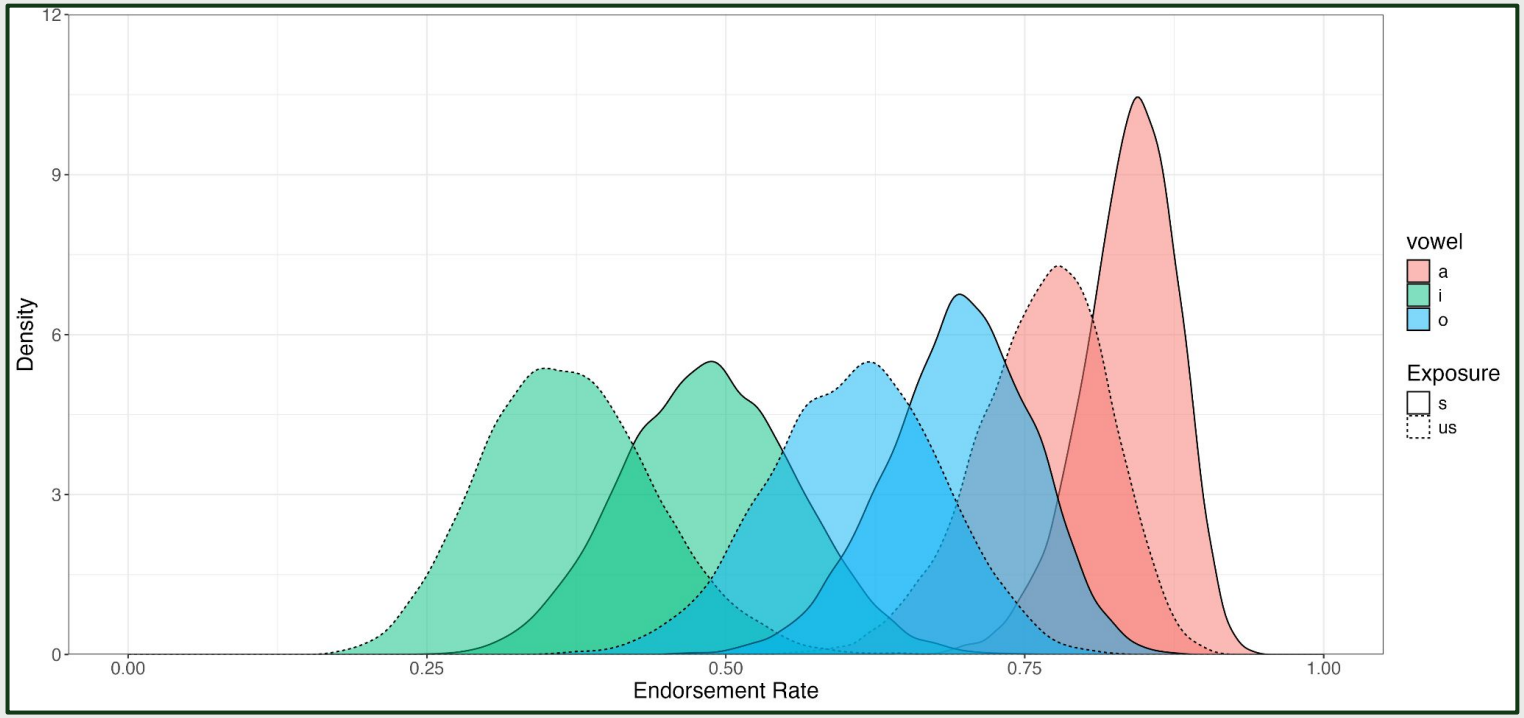
Bayesian Model for vowels in critical words

Model Results: /u/



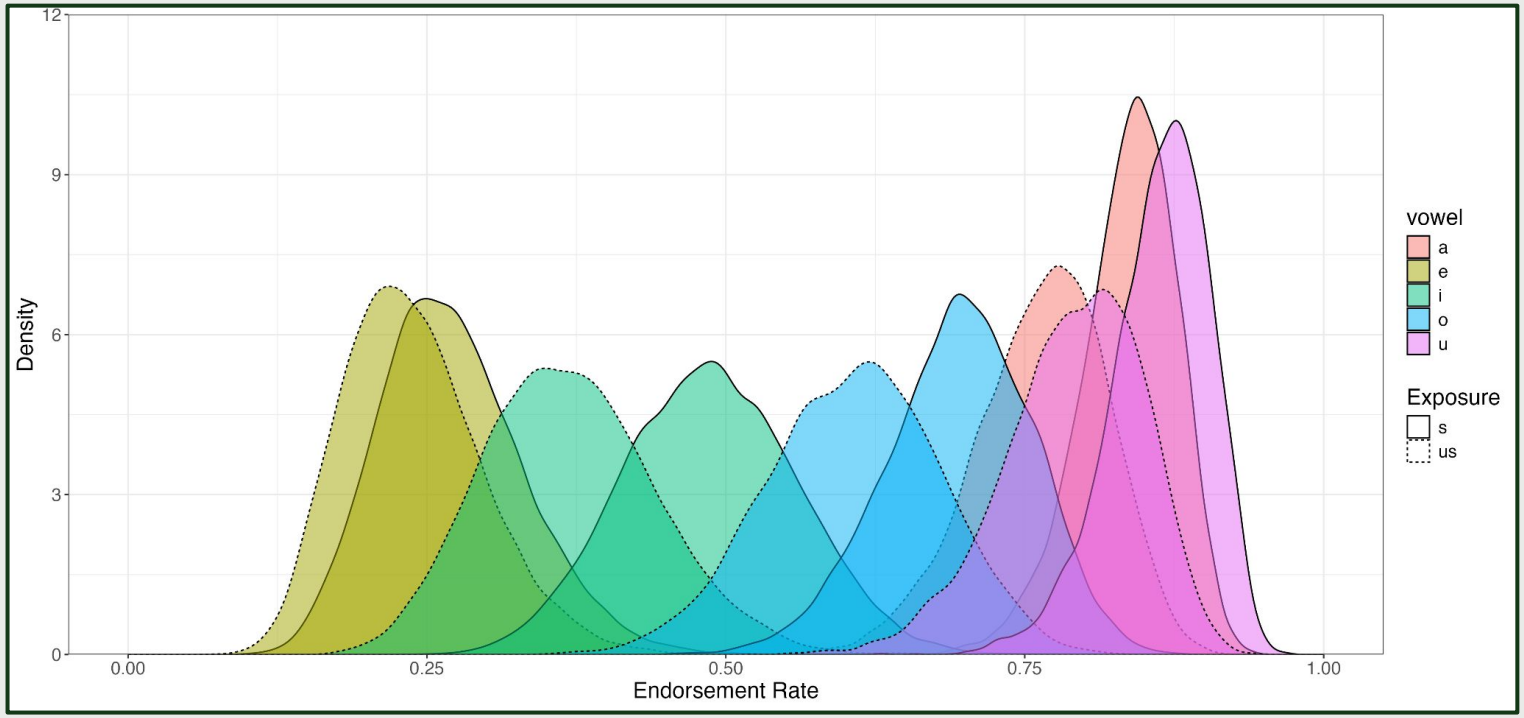
Bayesian Model for vowels in critical words

Model Results: significant exposure interaction



Bayesian Model for vowels in critical words

Model Results: all

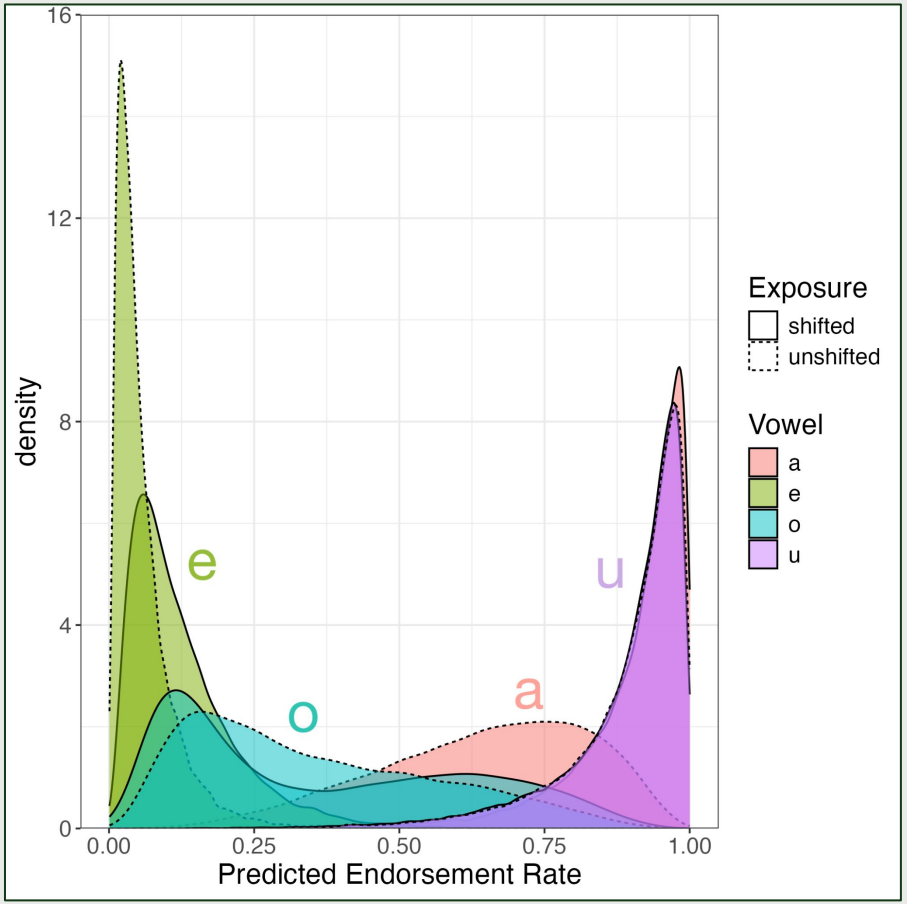


Bayesian Model: Critical words

vowel * stress * exposure + (1 | Participant) + (condition | word) + (1 | Trial)

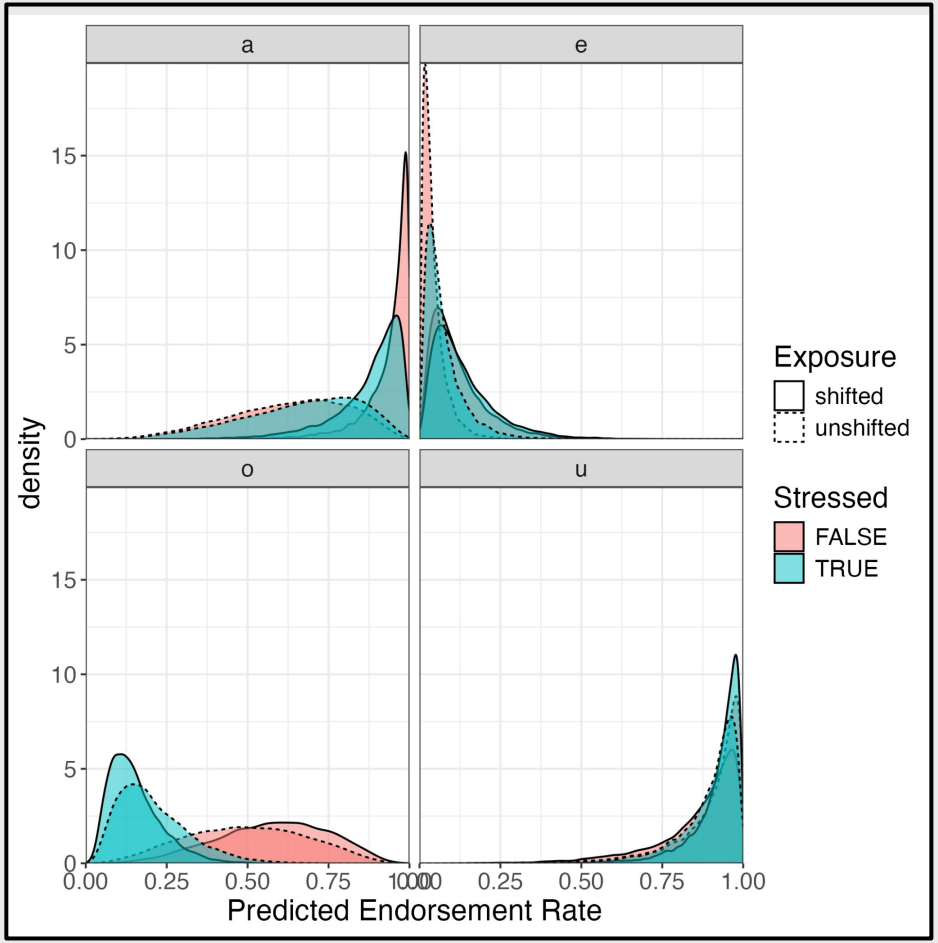
Model Results:

- Exposure had an overall positive effect
- All vowels significantly differed from each other
- Exposure seemed to not have an additional effect depending on the vowel in the base model.
  - Pairwise comparisons revealed positive effects of exposure for /e/ and /a/



Stress

Stress was not found to be a significant predictor for most vowels, except for /o/.



Stress

Stress was not found to be a significant predictor for most vowels, except for /o/.

Stressed /o/ was found to be less likely to be endorsed than unstressed /o/ regardless of exposure

Just /o/ →