

Style in Time

Online perceptions of sociolinguistic cues

Daisy Leigh

Dissertation Oral, Stanford Linguistics

July 21, 2021



We use *styles* — combinations of socially meaningful speech sounds — in constructing *personas*

Few local, Beijing-specific features (e.g. rhotacization)



Full tone in unstressed syllables

Long, intense bursts

Falsetto



Peripheral vowels

Raised, rounded LOT



Velarized //

Extreme F0 values

Social meanings are mutable

Social meanings are mutable



I'm watchin'
TV

(ING)
-in' vs. -ing

friendly,
chill

Social meanings are mutable

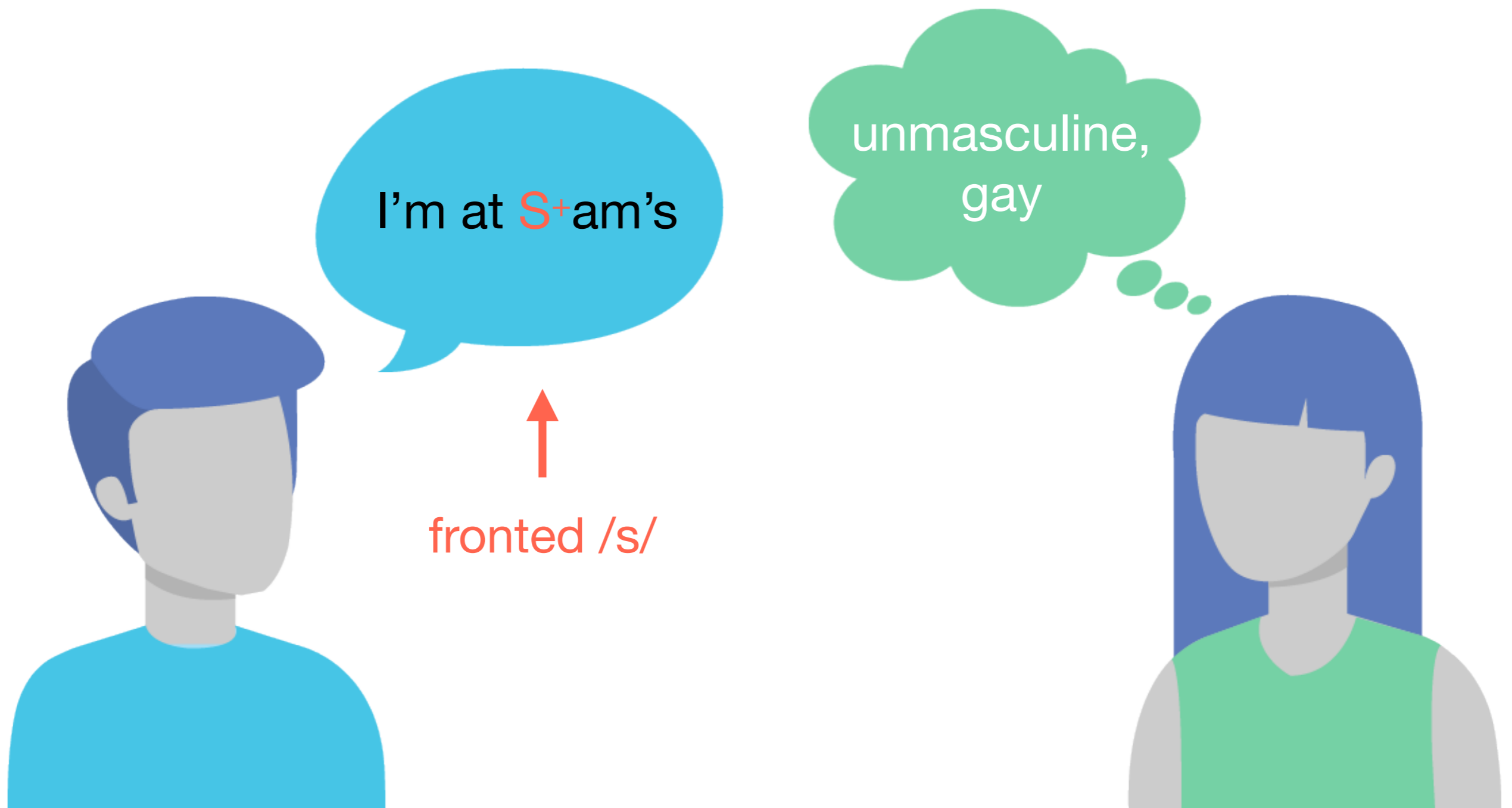


I'm watchin'
TV

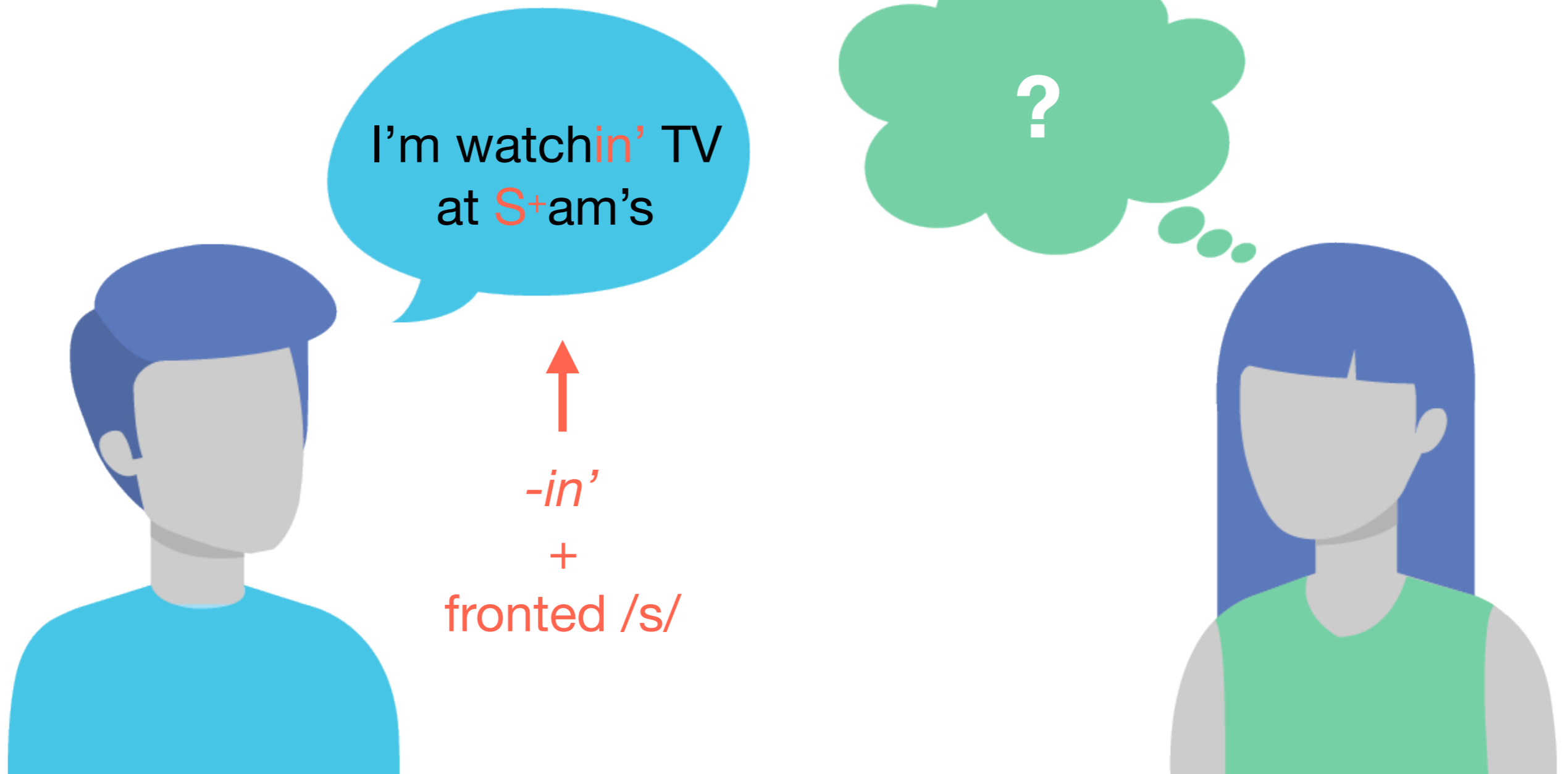


unintelligent,
slacker

Social meanings are mutable



Social meanings are mutable



As listeners, we integrate the meaning contributions of sociolinguistic cues with all the other social impressions that arise when listening to someone talk

... how and when do we do this?

How do we recognize sociolinguistic cues as comprising a particular style, produced by a particular kind of person?

Three Questions

- 1 Do listeners reconcile socioindexical cues when making online inferences about speaker persona?
- 2 How do existing expectations about a speaker modulate listeners' inferences?
- 3 How do listeners reconcile conflicting socioindexical cues?

1 Do listeners reconcile socioindexical cues when making online inferences about speaker persona?

(ING)

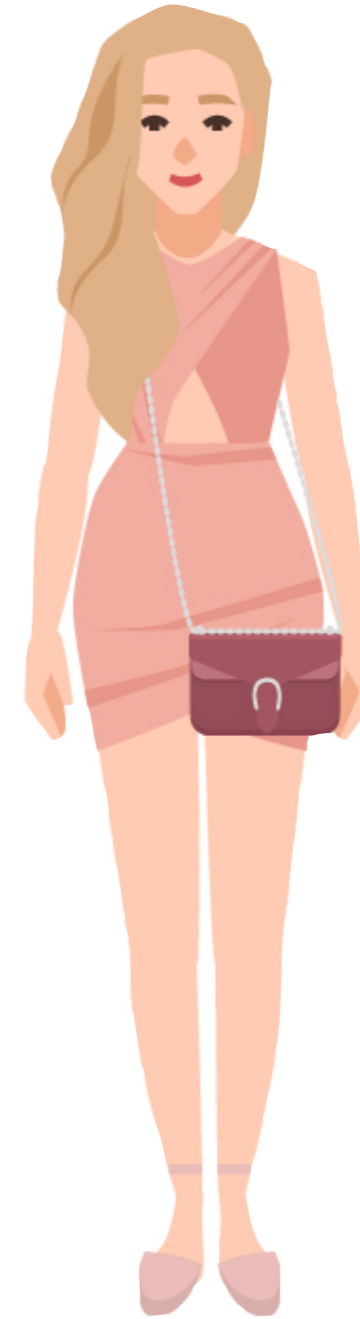
-in' vs. -ing



'Tough'

(HRT)

High Rising Terminals
vs. Declaratives

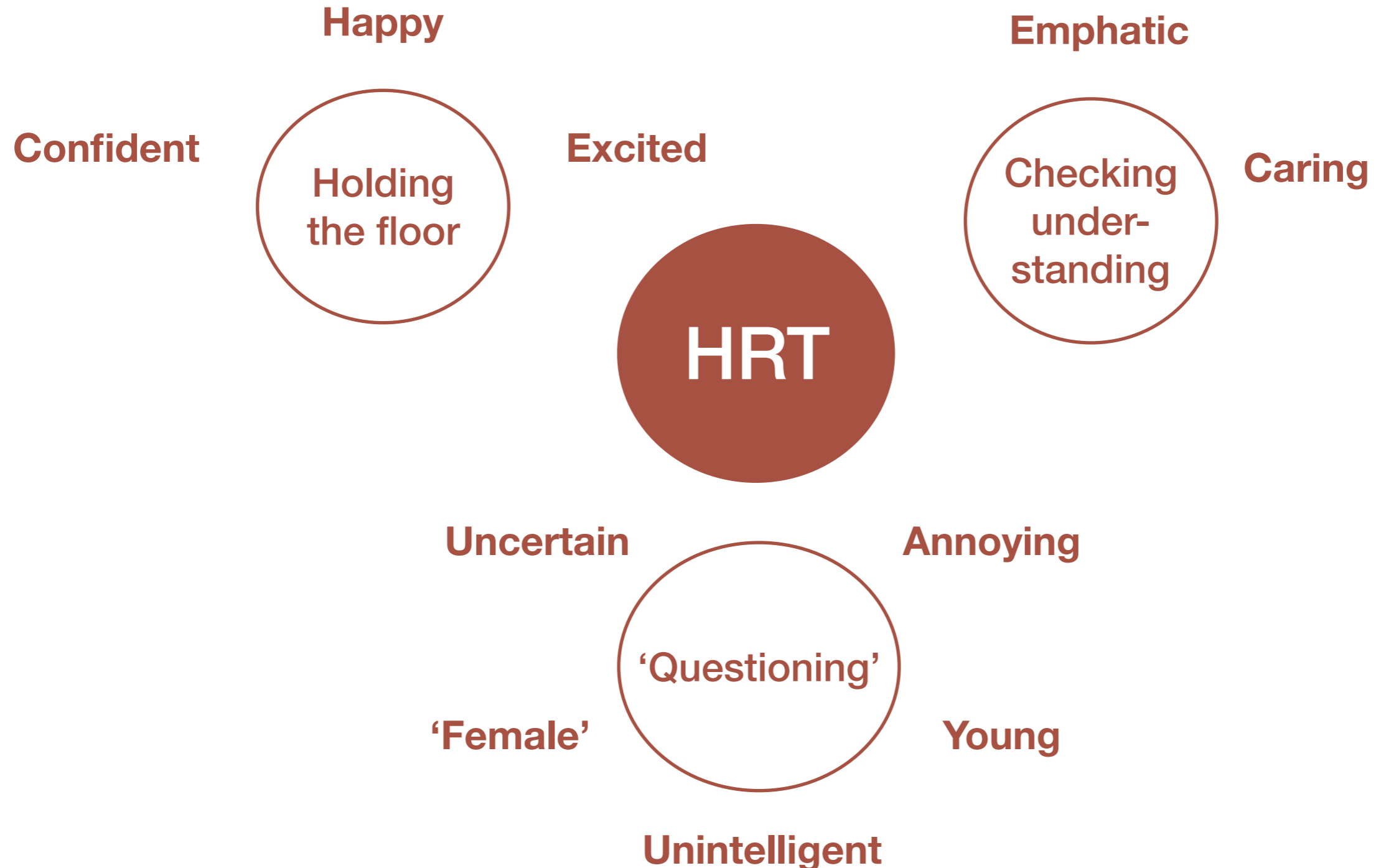


'Valley Girl'

(HRT)

High Rising Terminals vs. Declaratives

e.g. I'm talking about the dog ↗



(HRT)

High Rising Terminals vs. Declaratives

e.g. I'm talking about the dog ↗



Frank & Moon Zappa
Valley Girl



(ING)
-in' vs. -ing

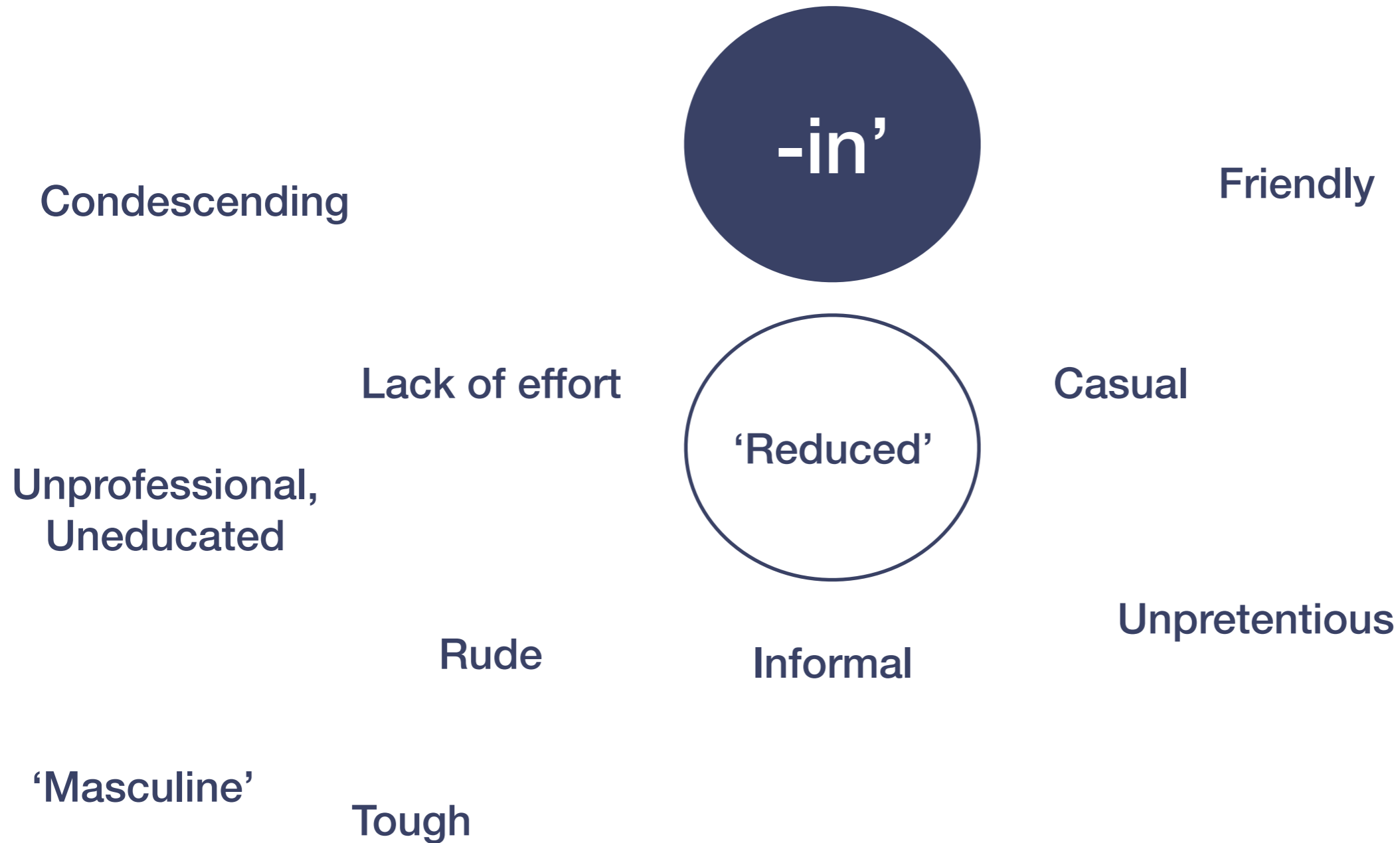
(HRT)
High Rising Terminals
vs. Declaratives



(ING)

-in' vs. -ing

*e.g. I'm talk**in'** about the dog*

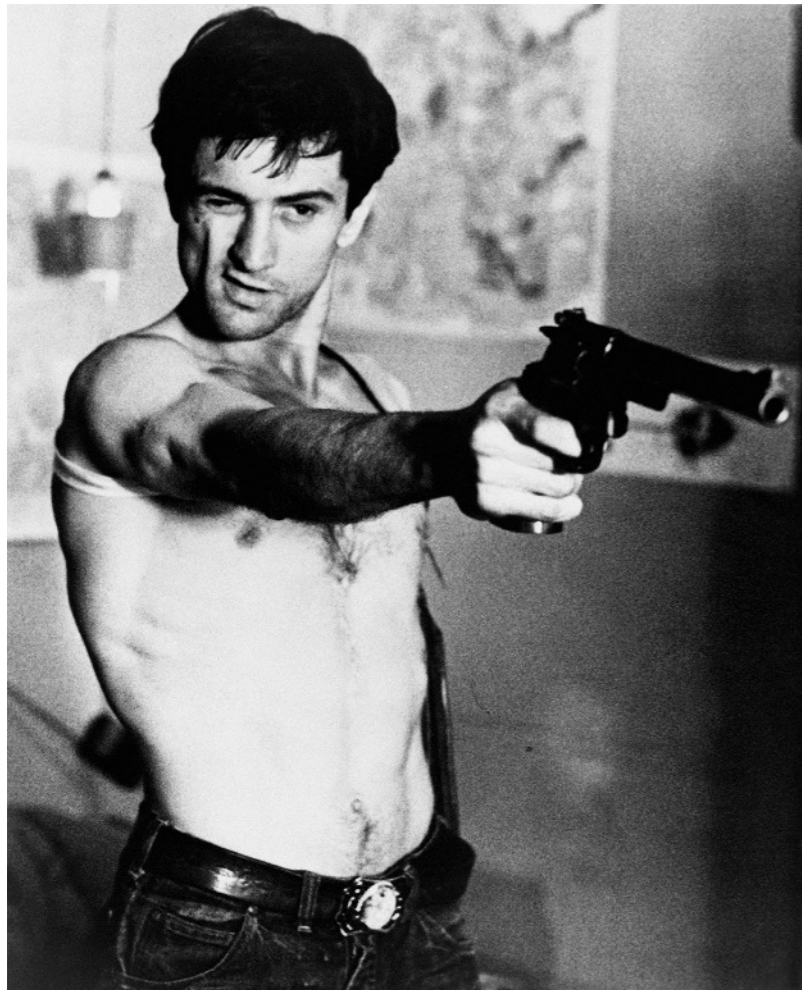


(ING)

-in' vs. -ing

*e.g. I'm talk**in'** about the dog*

Condescending



Effort



Casual



Unpretentious

Informal

(ING)

-in' vs. -ing



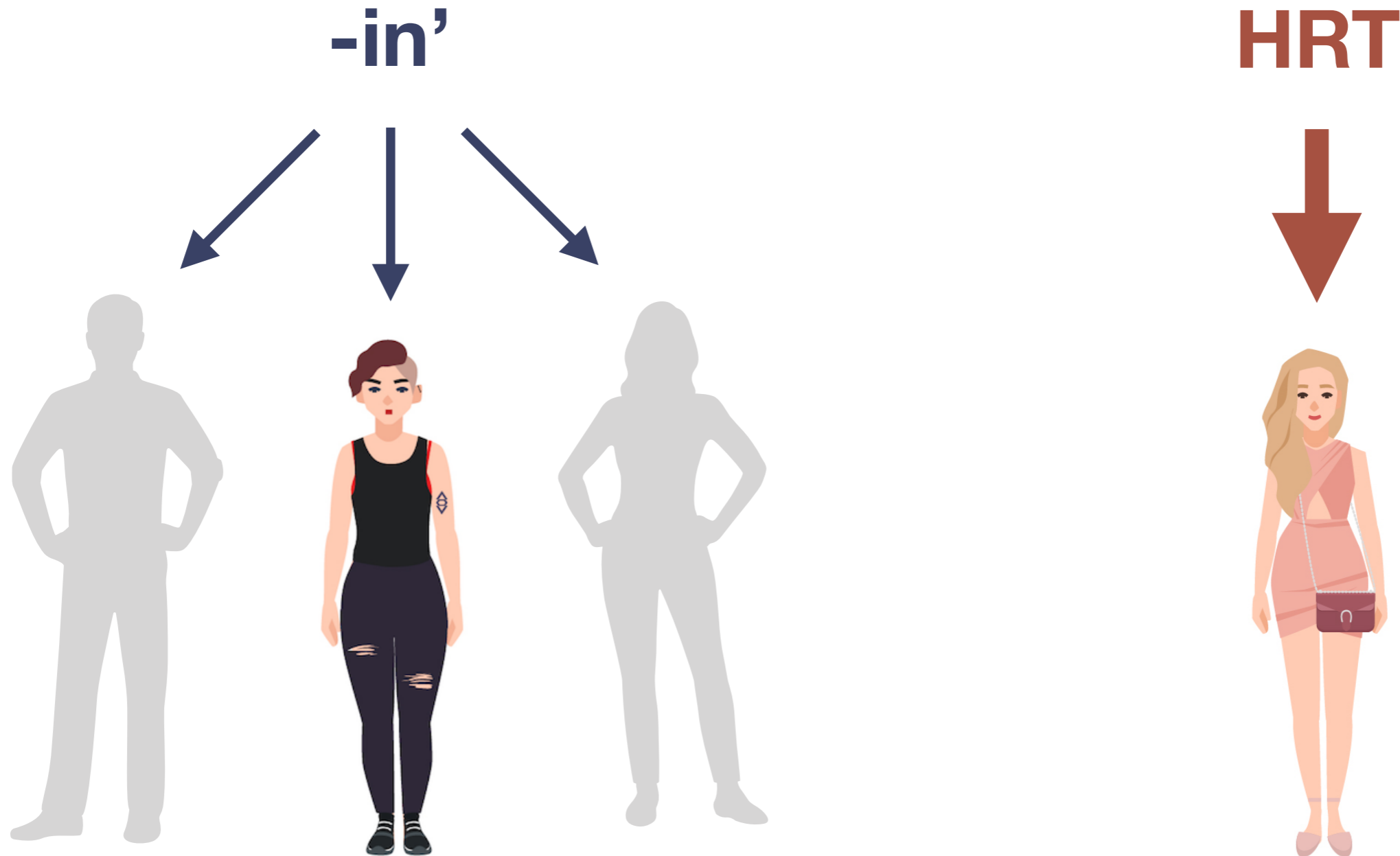
(HRT)

High Rising Terminals vs. Declaratives



1

Do listeners reconcile socioindexical cues when making online inferences about speaker persona?

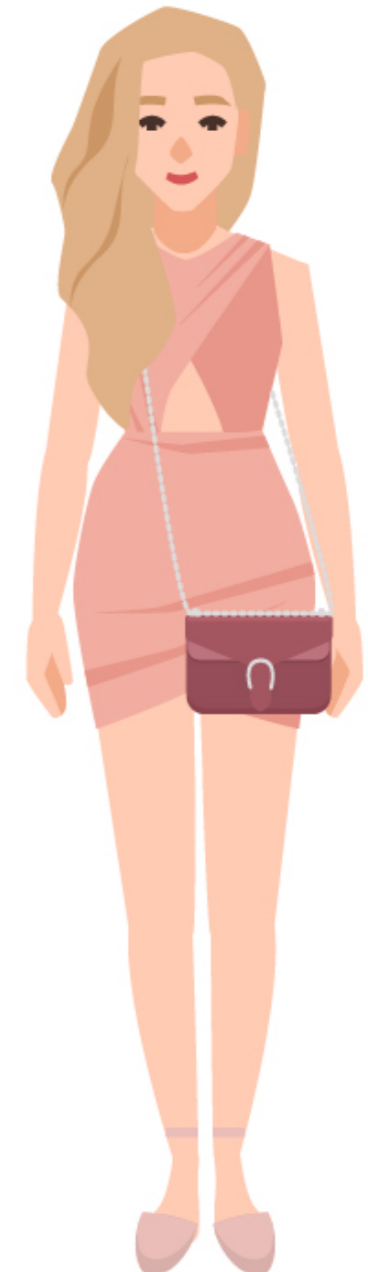


Select the speaker you think was **more likely** to have produced the utterance



 categorical speaker selections

 eye movements



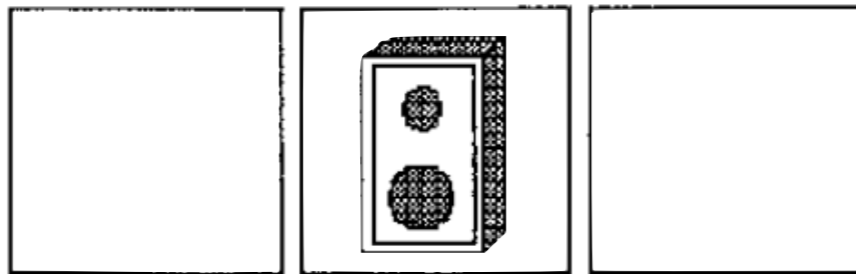
Participants:



The Visual World Paradigm

*Pick up the **beaker***

Speaker



Beaker

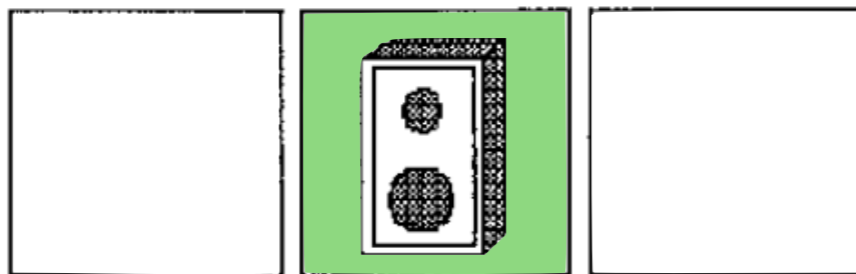
Beetle

Carriage

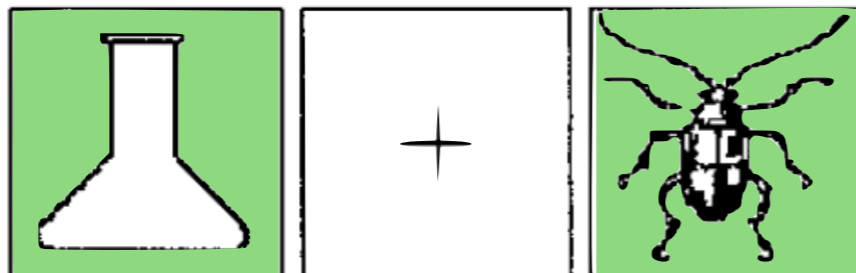
The Visual World Paradigm

Pick up the ...

Speaker



Beaker



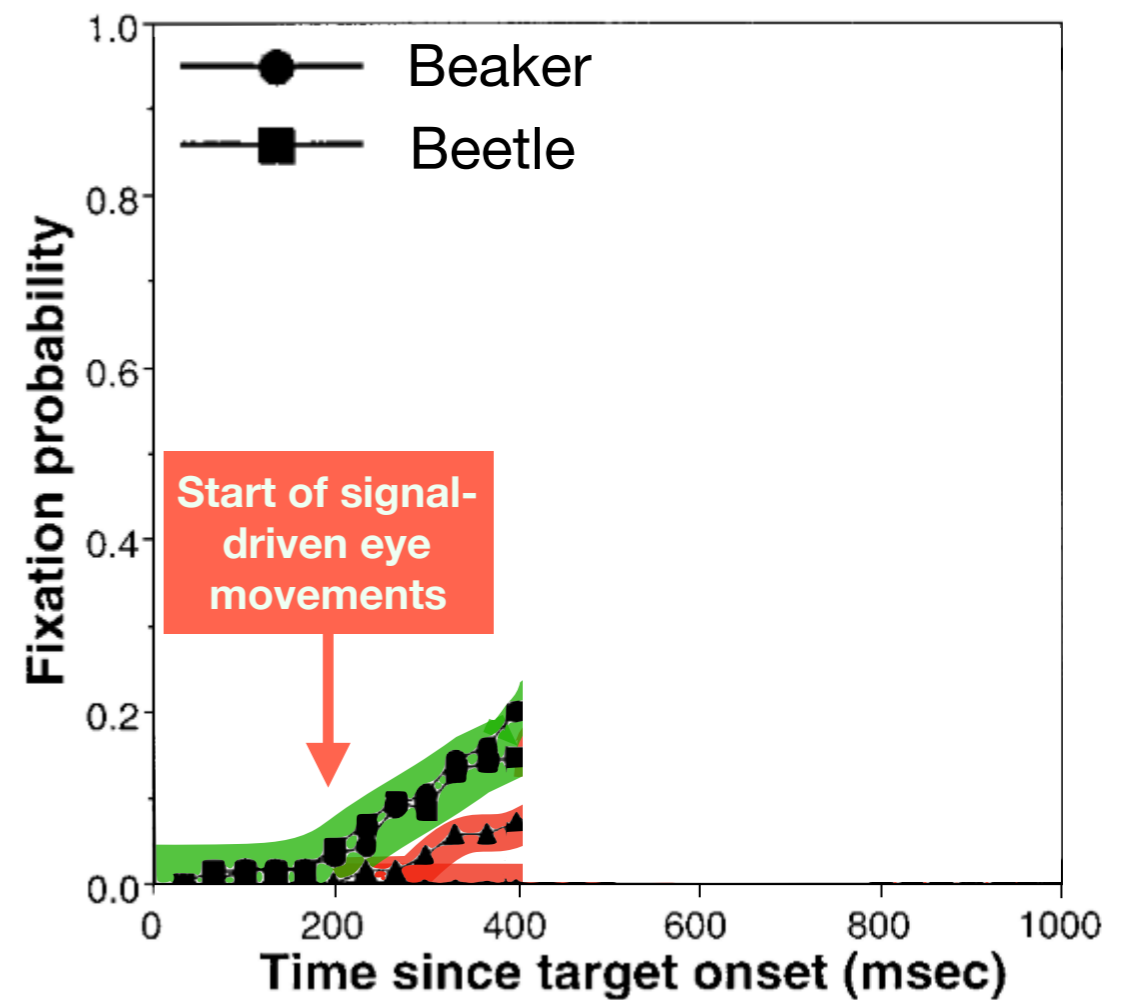
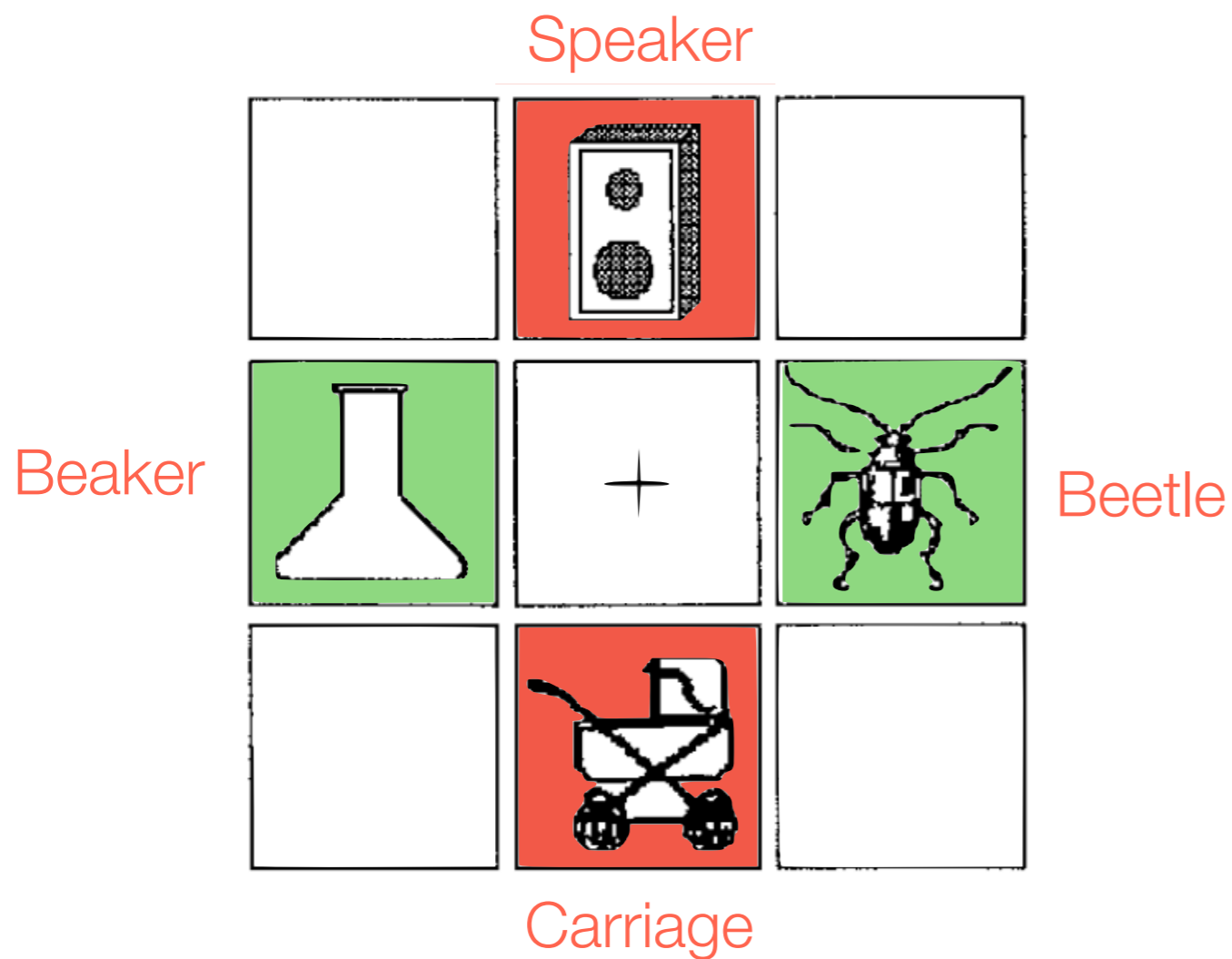
Beetle

Carriage



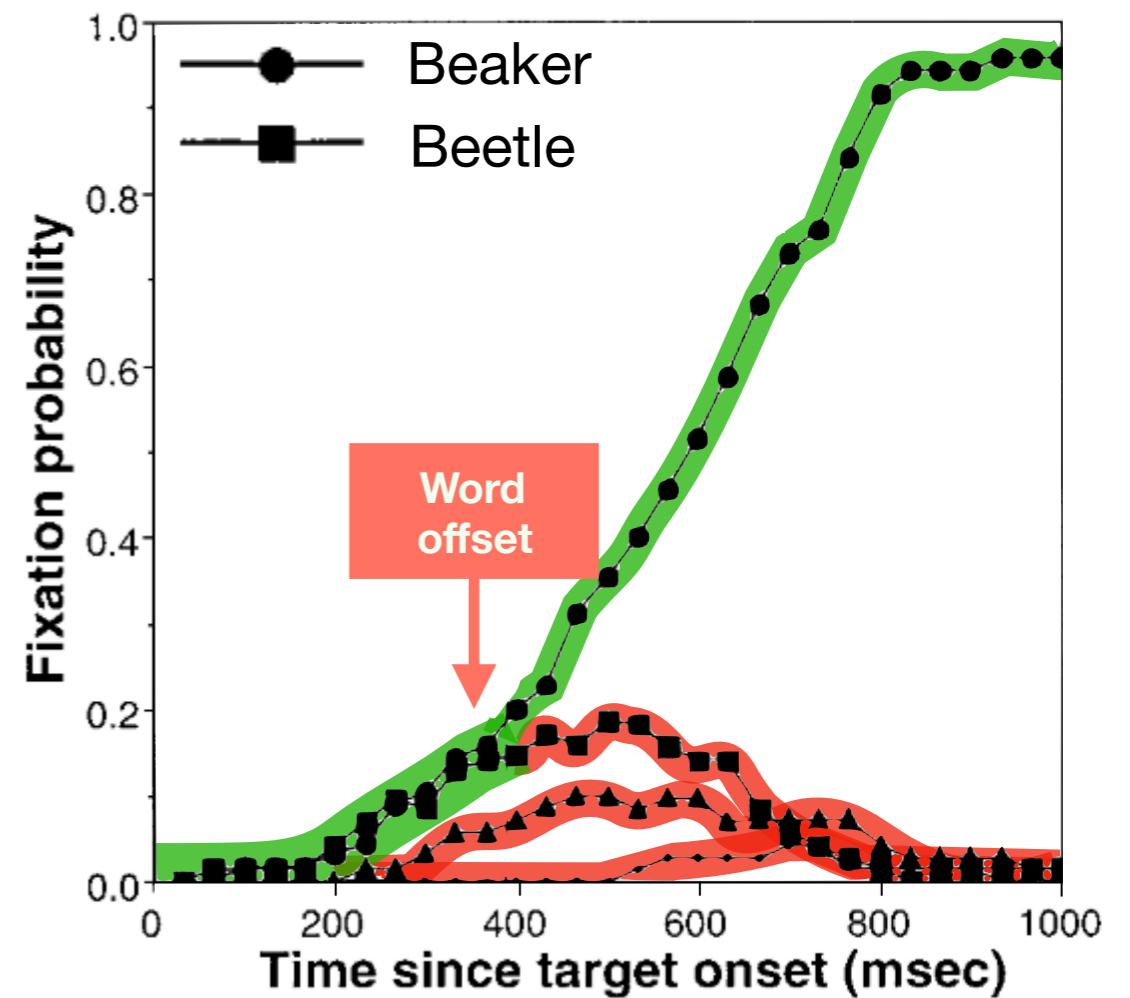
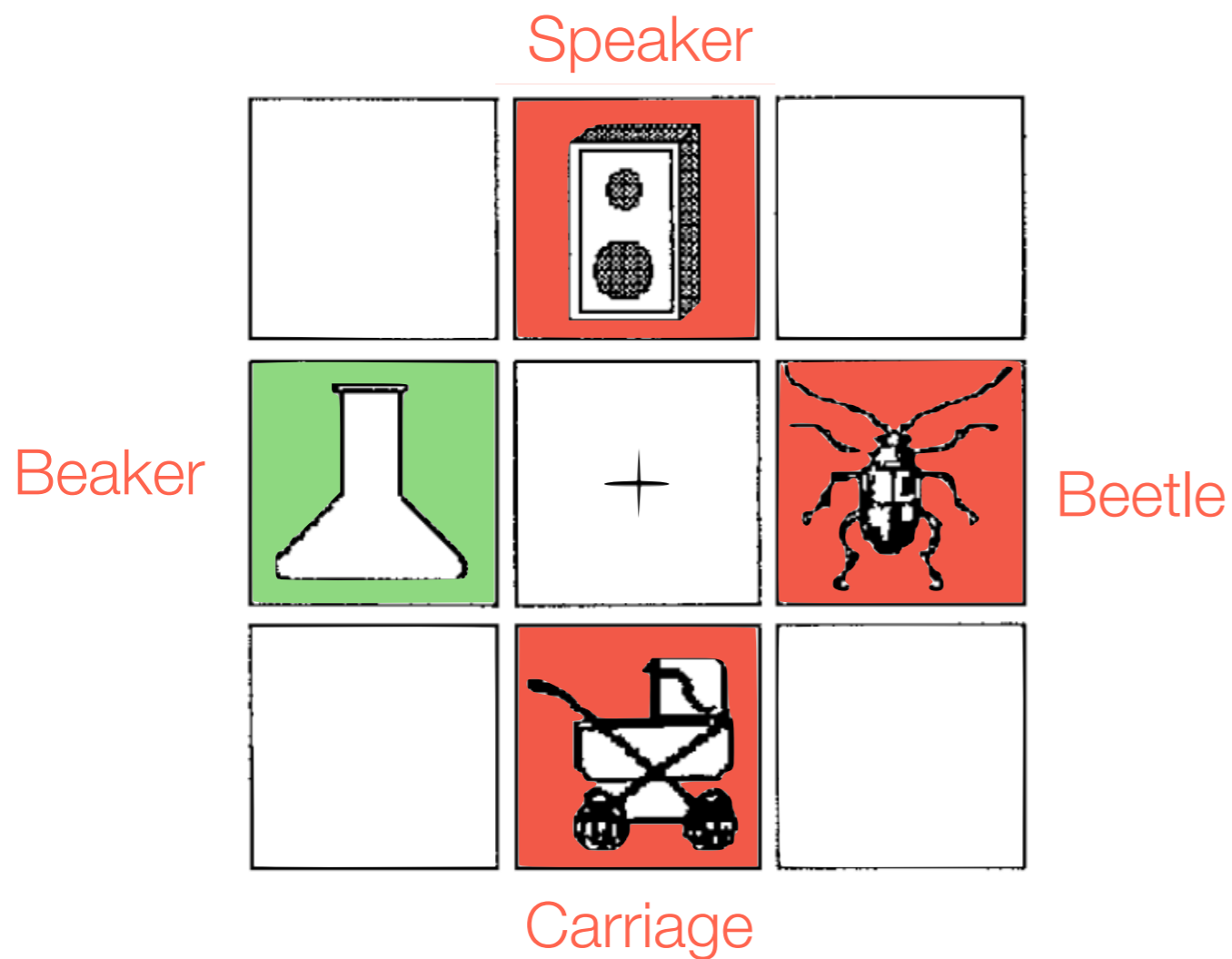
The Visual World Paradigm

*Pick up the **b...***



The Visual World Paradigm

Pick up the **beaker**



Why eye movements?

Closely time-locked to the linguistic signal

Fine-grained, automatic, and implicit

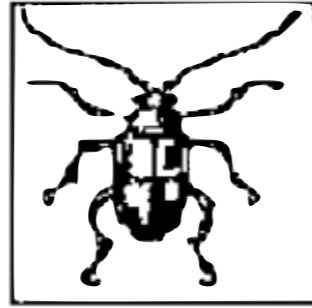
A proxy for ‘belief’

(more looks to a referent \approx greater probability it's the intended one)

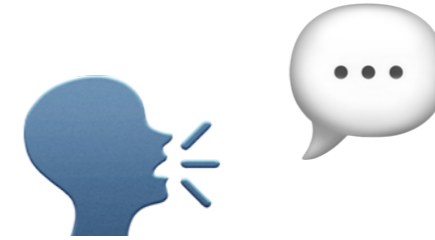
A sensitive measure of listeners'
online, unfolding linguistic inferences

Implicit, online processing of linguistic material

Pick up the ...



Explicit, offline evaluations of speakers



How _____ do you think the speaker is?

Implicit, online evaluations of speakers



1

Do listeners reconcile socioindexical cues when making online inferences about speaker persona?

2 x eye-tracking experiments

1 x critical voice

Critical stimuli

"People are ..."

Exp. 1a (ING)

Listeners hear (ING) cues



Exp. 1b (HRT)

Listeners hear (HRT) cues



2 x filler voices

Male voice

People are eating.



Female voice

People are eating.



Conducted remotely using WebGazer.js

1

Do listeners reconcile socioindexical cues when making online inferences about speaker persona?

Predictions

Exp 1a

-in'

More Tough selections
More Tough looks
after cue onset
(compared to **-ing**)

Exp 1b

HRT

Fewer Tough selections
Fewer Tough looks
after cue onset
(compared to **Decl.**)

Magnitude of
-in' (vs. *-ing*) effect



Magnitude of
HRT (vs. Decl.) effect

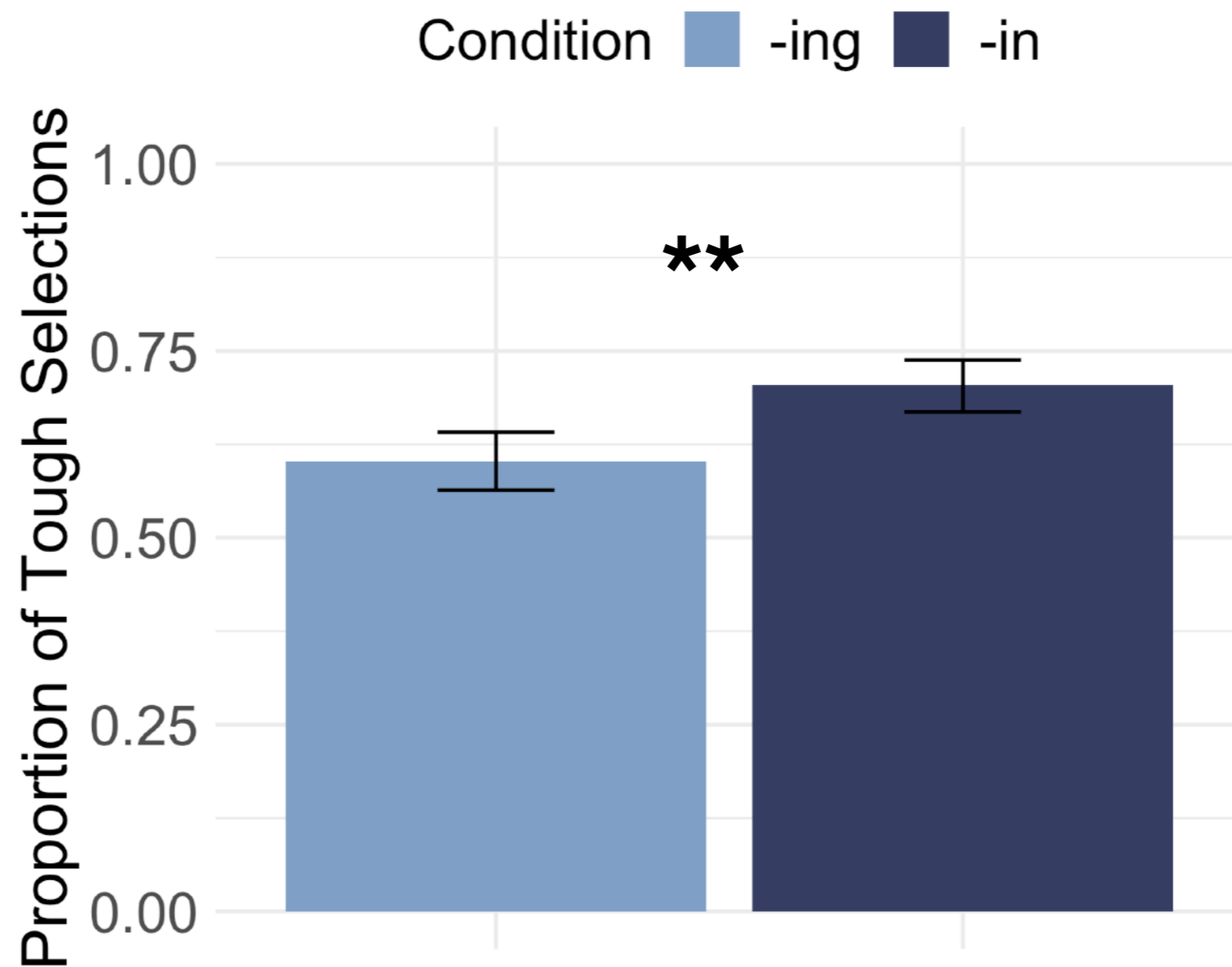


1 Do listeners reconcile socioindexical cues when making online inferences about speaker persona?

Critical trials: categorical selections

Exp. 1a: (ING)

N = 121



1

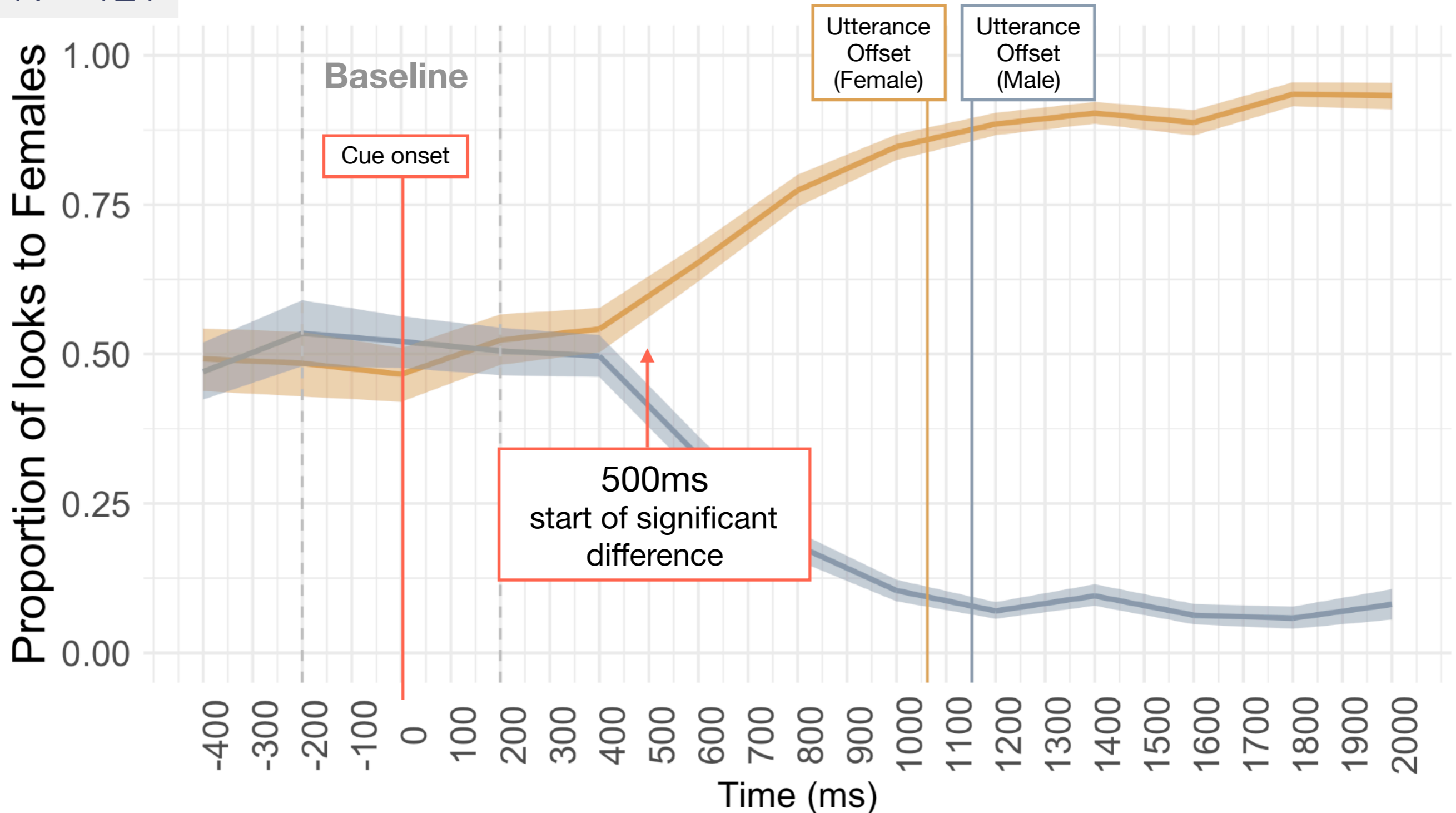
Do listeners reconcile socioindexical cues online when making inferences about speaker persona?

Sanity check! Results from filler trials

Exp. 1a: (ING)

N = 121

Condition + Female voice + Male voice



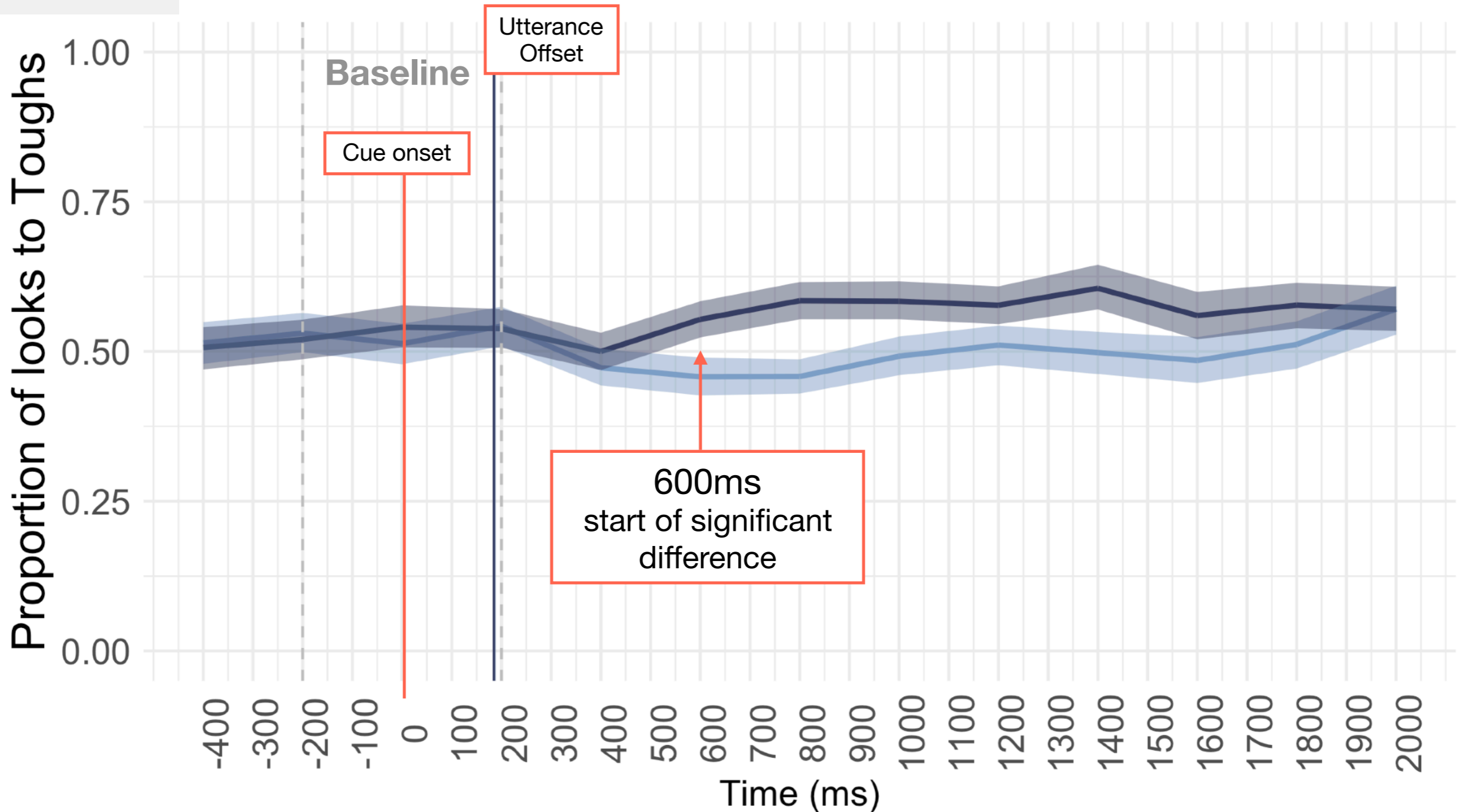
1 Do listeners reconcile socioindexical cues online when making inferences about speaker persona?

Critical trials: time course data

Exp. 1a: (ING)

N = 121

Condition + -ing + -in

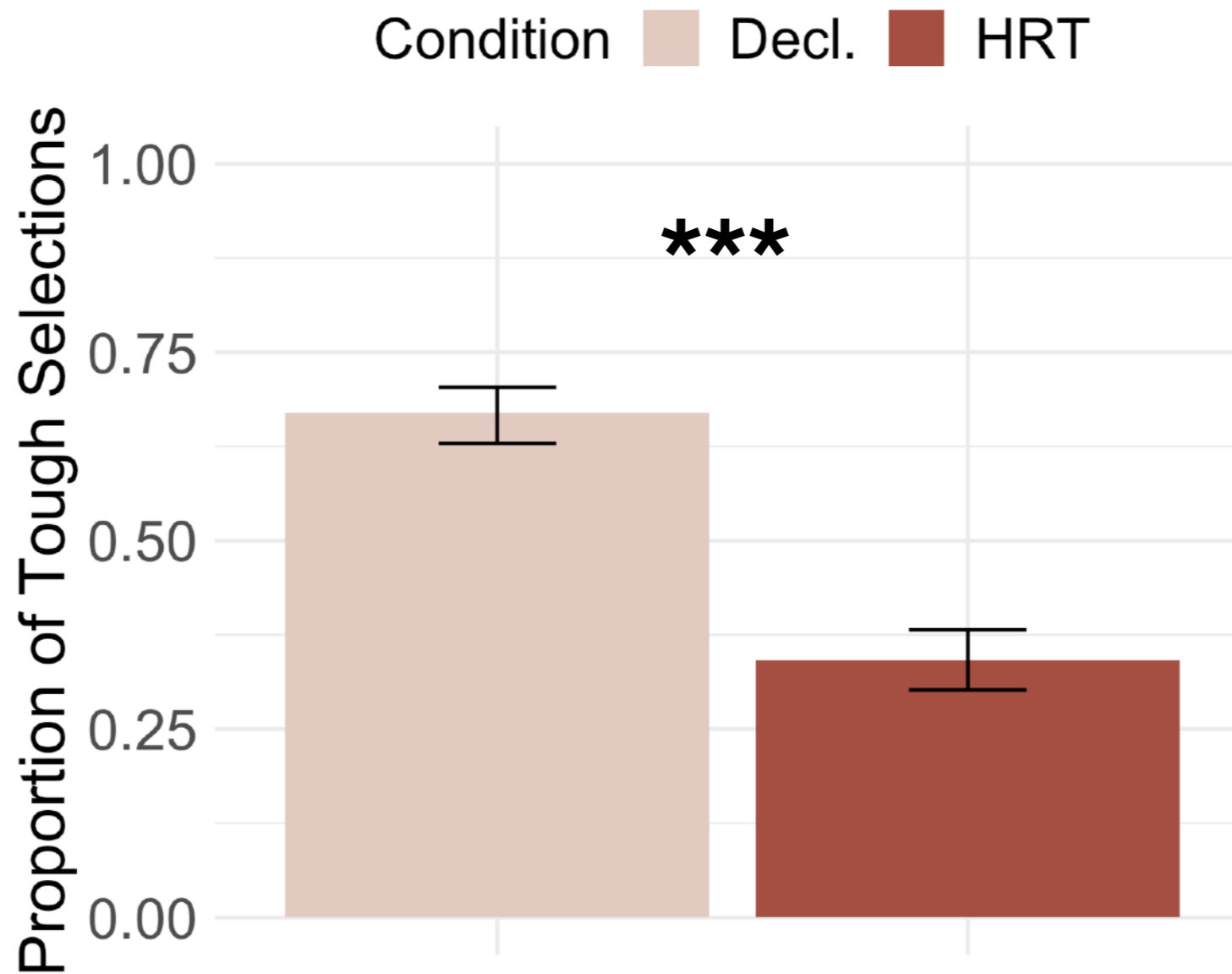


1 Do listeners reconcile socioindexical cues when making online inferences about speaker persona?

Critical trials: categorical selections

Exp. 1b: (HRT)

N = 119



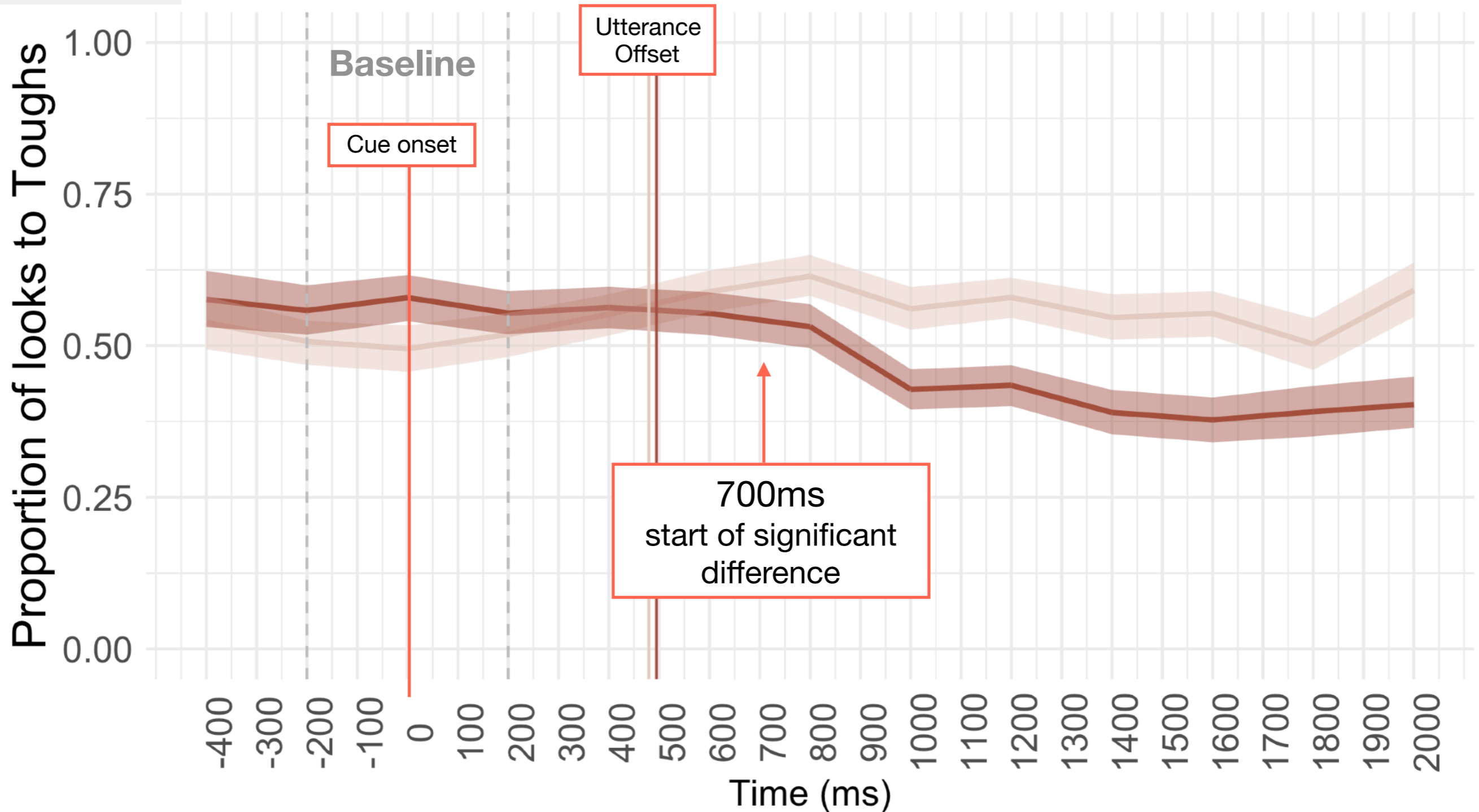
1 Do listeners reconcile socioindexical cues when making online inferences about speaker persona?

Critical trials: time course data

Exp. 1b: (HRT)

N = 119

Condition + Decl. + HRT



Summary

- ✓ Both cues give rise to effects in the expected directions
- ✓ Qualitatively, magnitude of (HRT) effect greater than (ING) effect
- ✓ The paradigm is capable of capturing listeners' online inferences about speakers (caveat: timings)

1

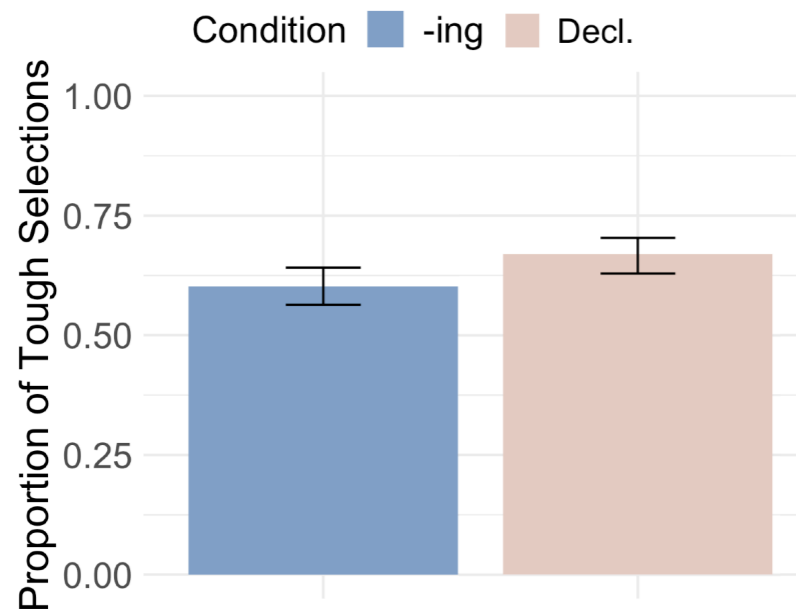
Do listeners reconcile socioindexical cues when making online inferences about speaker persona?

Conclusions and Questions

Listeners do use socioindexical cues to draw inferences about a speaker's social persona online

Tentatively: the more *informative* the cue, the greater it will shift listeners' beliefs — both offline and online

... Would these results generalize to *other* voices?



... or is it just because the voice sounds a little Tougher than Valley Girl?

2

How do existing expectations about a speaker modulate cue uptake?

4 x critical voices

Stimuli created from existing utterances in the NSP Corpus
All were white, college-aged students, from a variety of US dialect regions

Critical stimuli

“I’m talking about the ...”

example stimuli

-in’	I’m talk in ’ about the beam ↘
HRT	I’m talk ing about the beam ↗
-ing/Decl.	I’m talk ing about the beam ↘

2 x eye-tracking experiments

Exp. 2a (ING)

Listeners hear only (ING) cues

-in' I'm talk**in**' about the beam ↘

-ing/Decl. I'm talk**ing** about the beam ↘

Per voice, listeners heard:

2 x -in'

2 x -ing/Decl

Exp. 2b (HRT)

Listeners hear only (HRT) cues

HRT I'm talk**ing** about the beam ↗

-ing/Decl. I'm talk**ing** about the beam ↘

Per voice, listeners heard:

2 x HRT

2 x -ing/Decl

Predictions

Exp 2a

-in'

More Tough selections
 More Tough looks
 after cue onset
 (compared to **-ing**)

Exp 2b

HRT

Fewer Tough selections
 Fewer Tough looks
 after cue onset
 (compared to **Decl.**)

Magnitude of
-in' (vs. *-ing*) effect



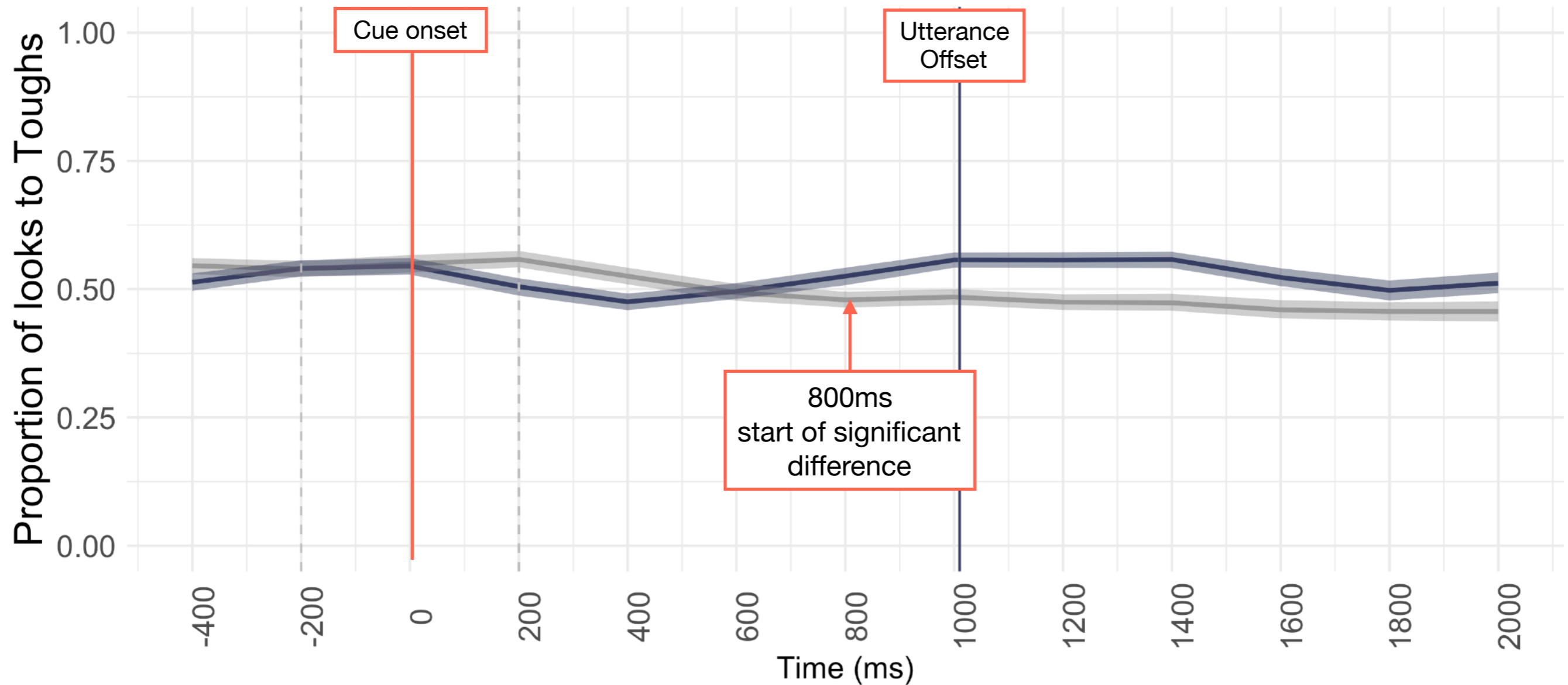


Magnitude of
 HRT (vs. Decl.) effect

Aggregate effects

Exp. 2a: (ING)

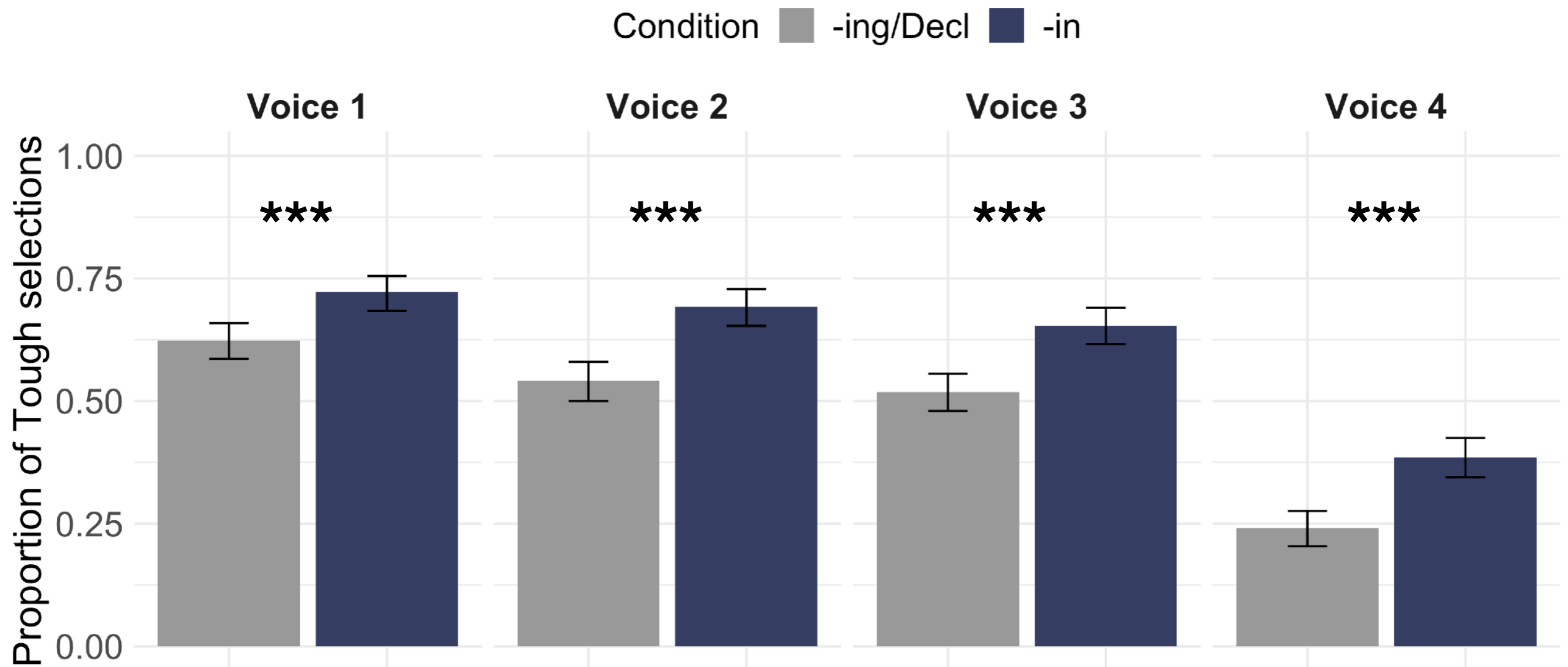
N = 322

Condition  -ing/Decl  -in

Voice-specific effects

Exp. 2a: (ING)

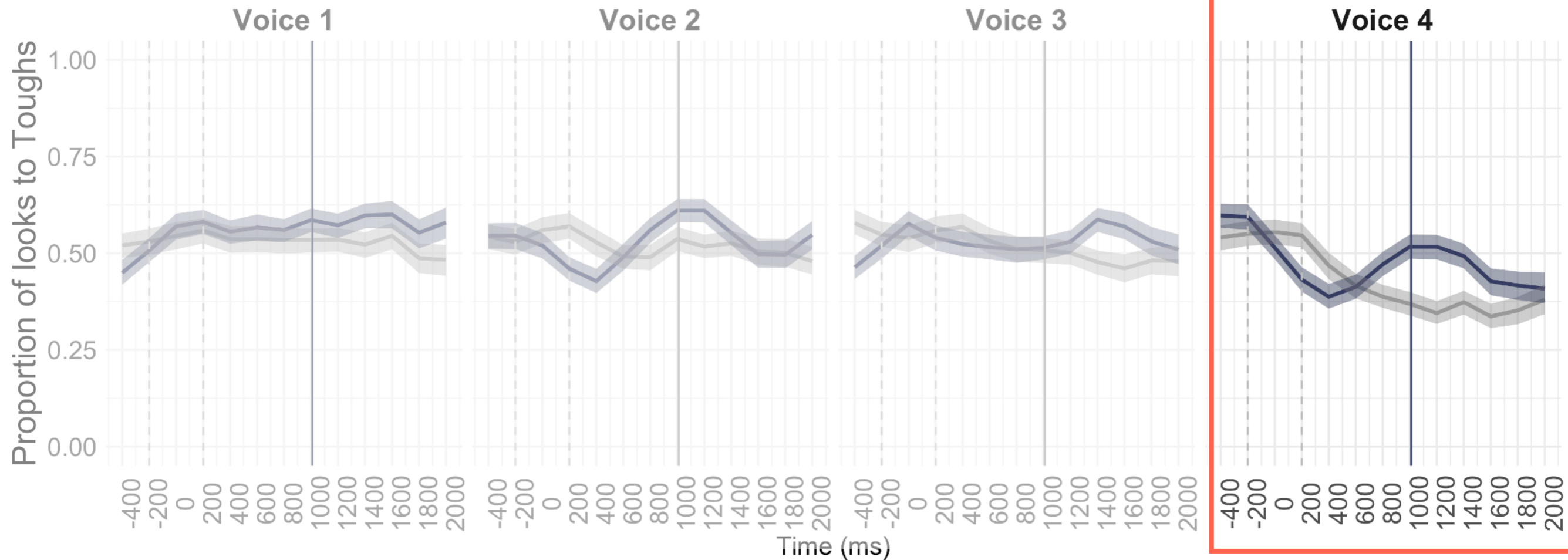
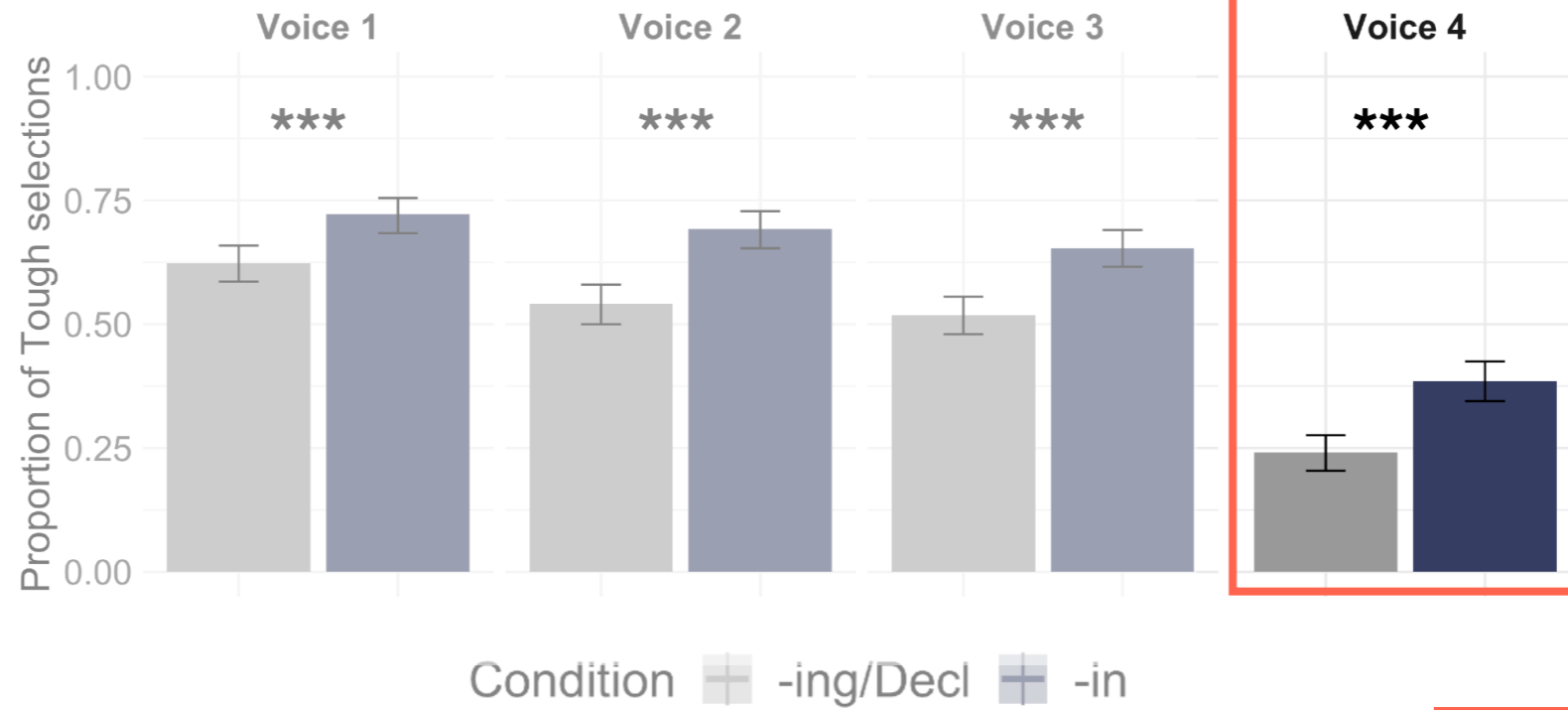
N = 322



Voice-specific effects

Exp. 2a:
(ING)

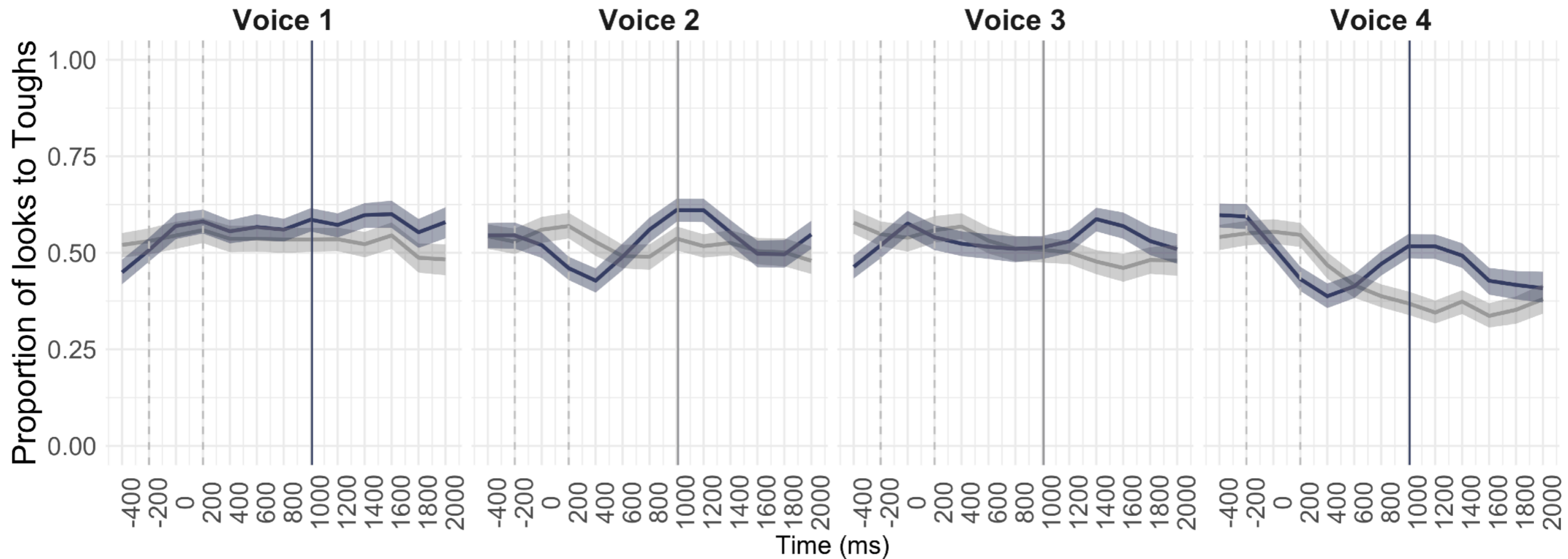
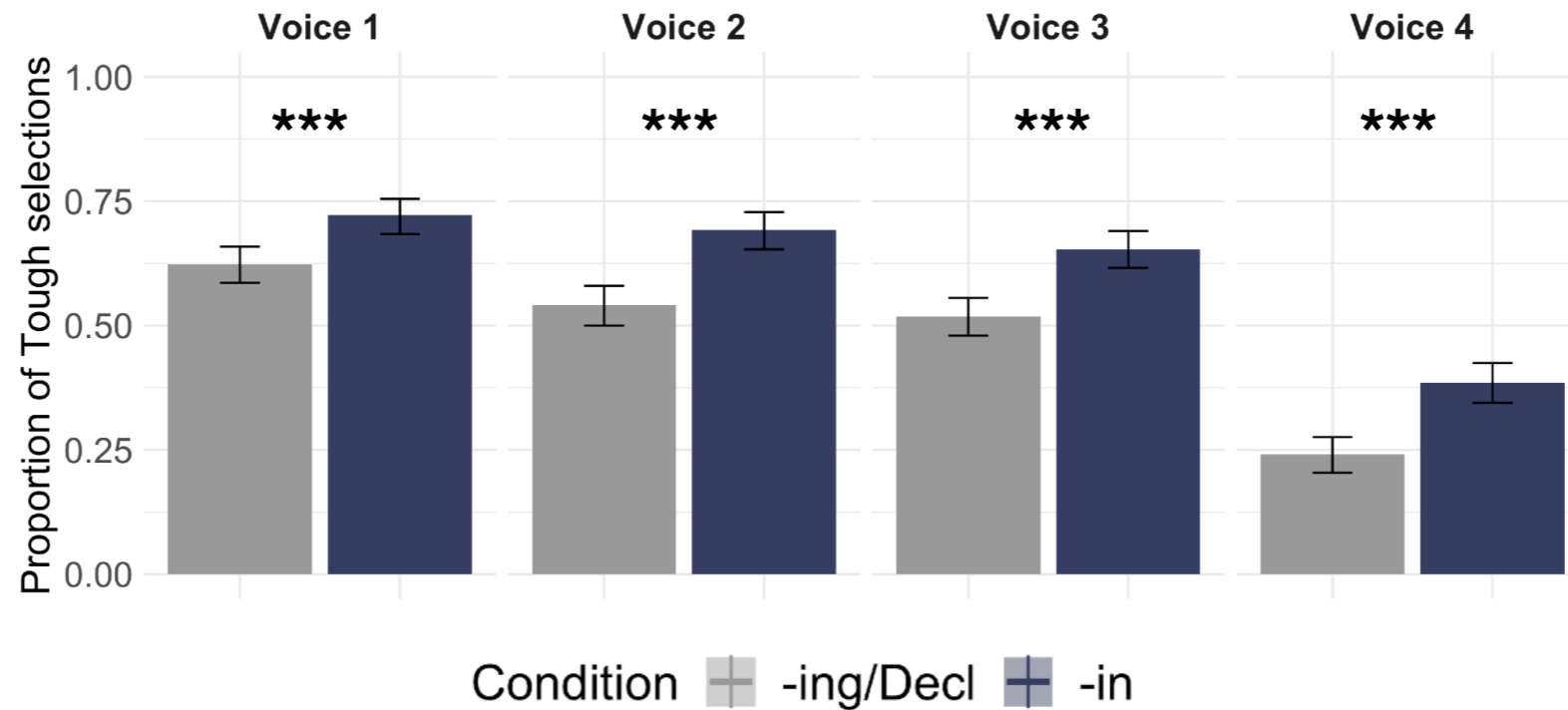
N = 322



Voice-specific effects

Exp. 2a:
(ING)

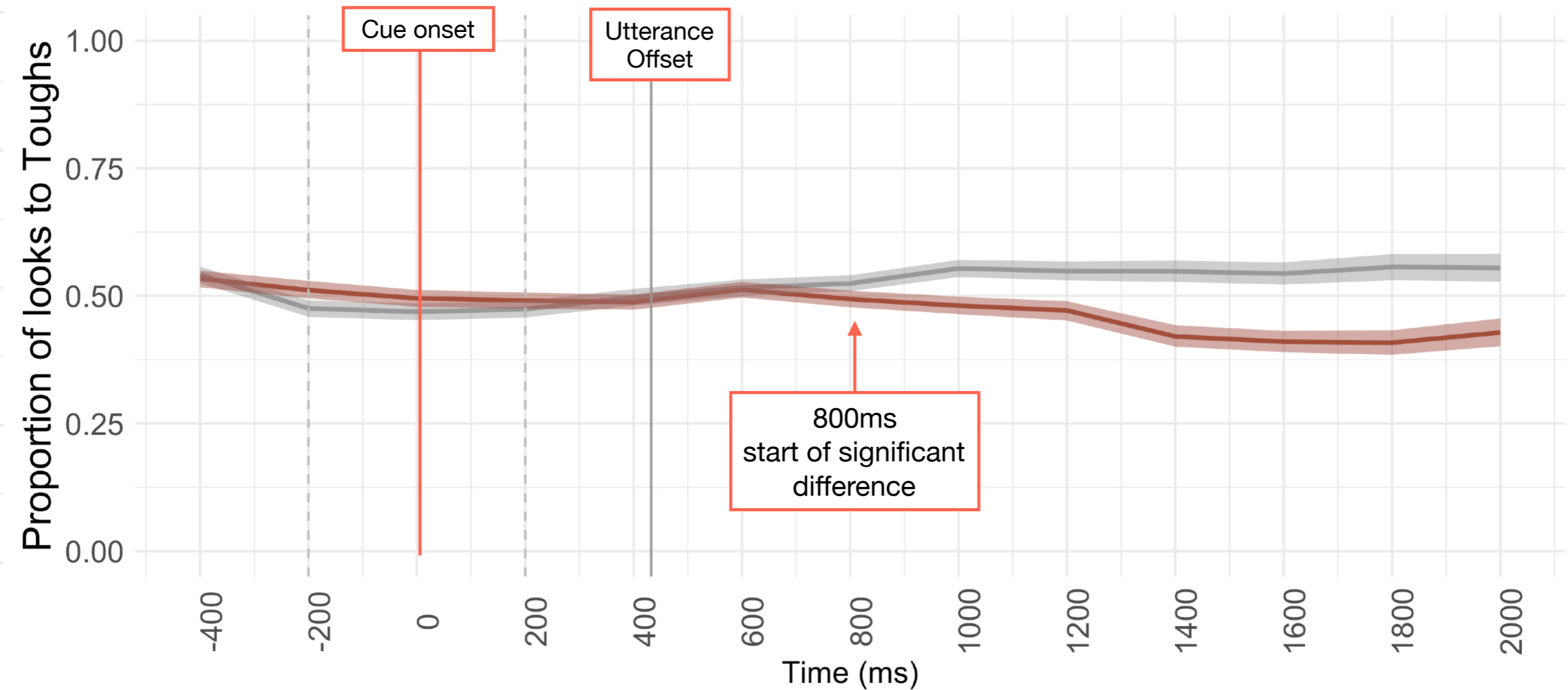
N = 322



Aggregate effects

Exp. 2b: (HRT)

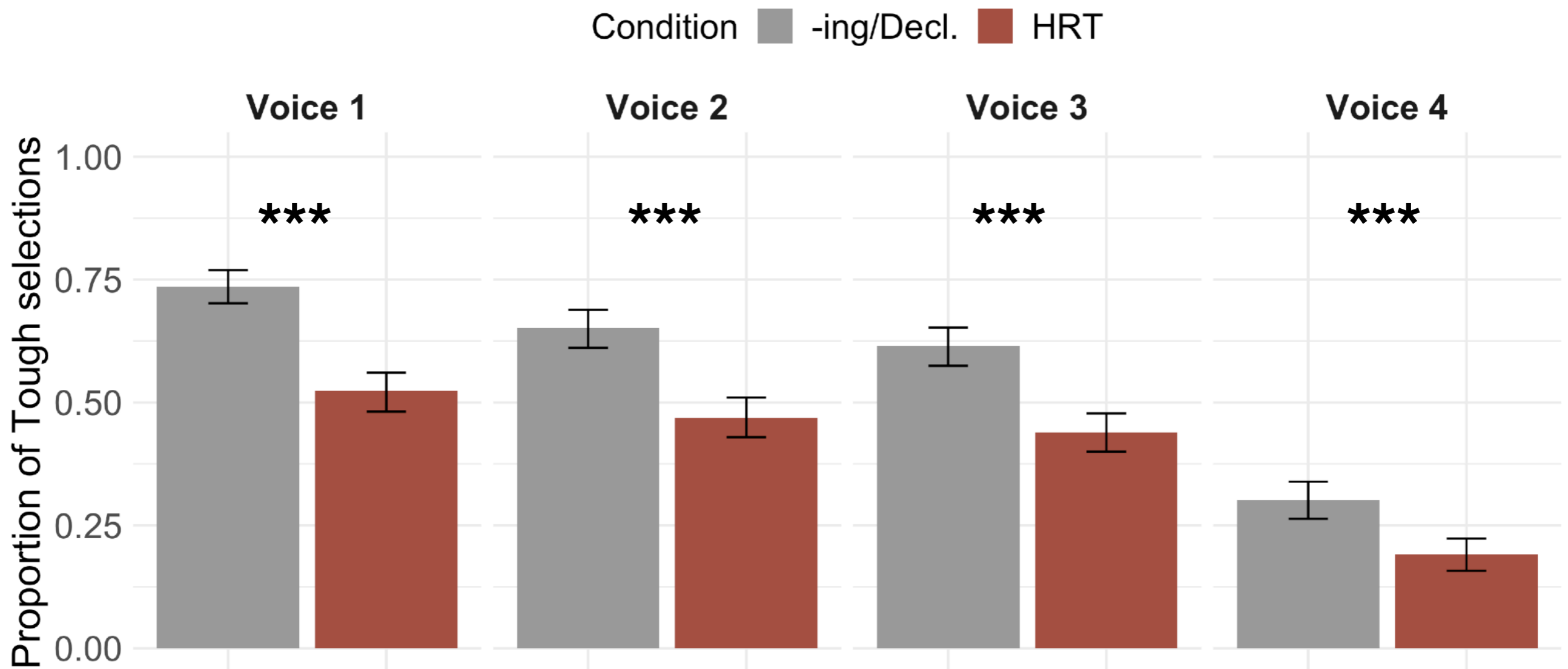
N = 321

Condition  -ing/Decl.  HRT

Voice-specific effects

Exp. 2b: (HRT)

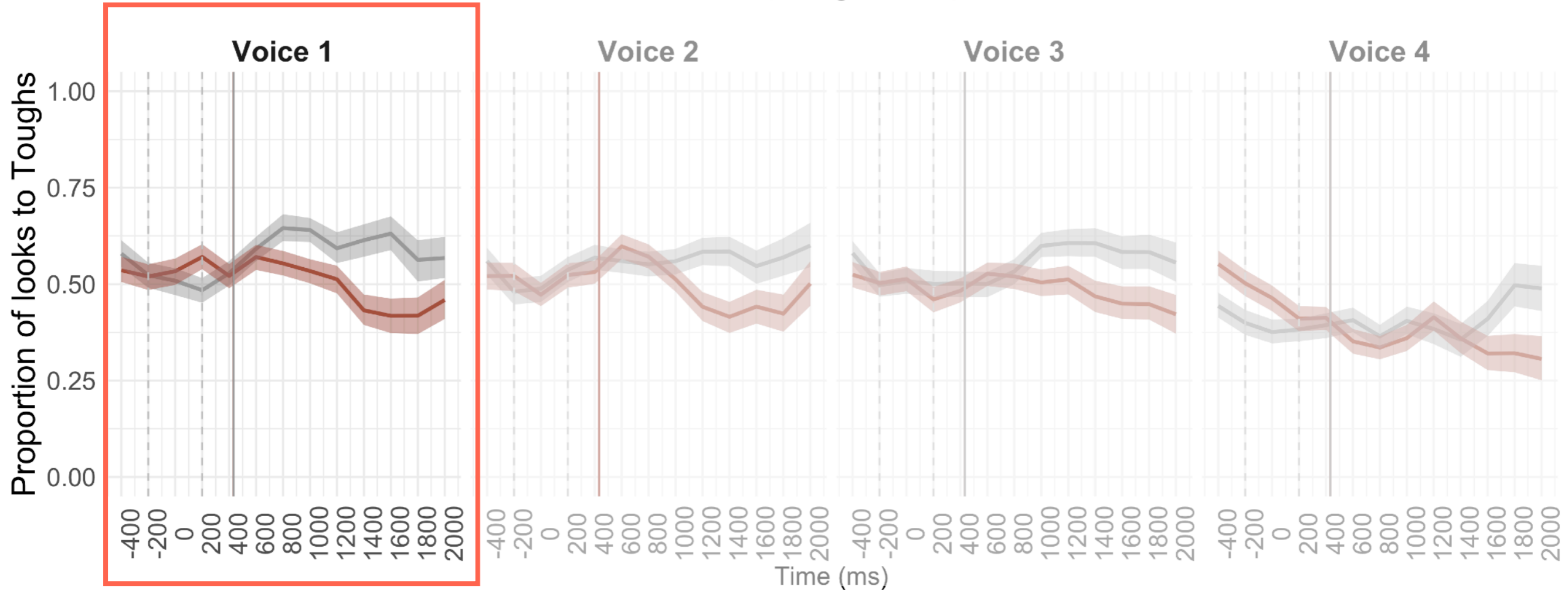
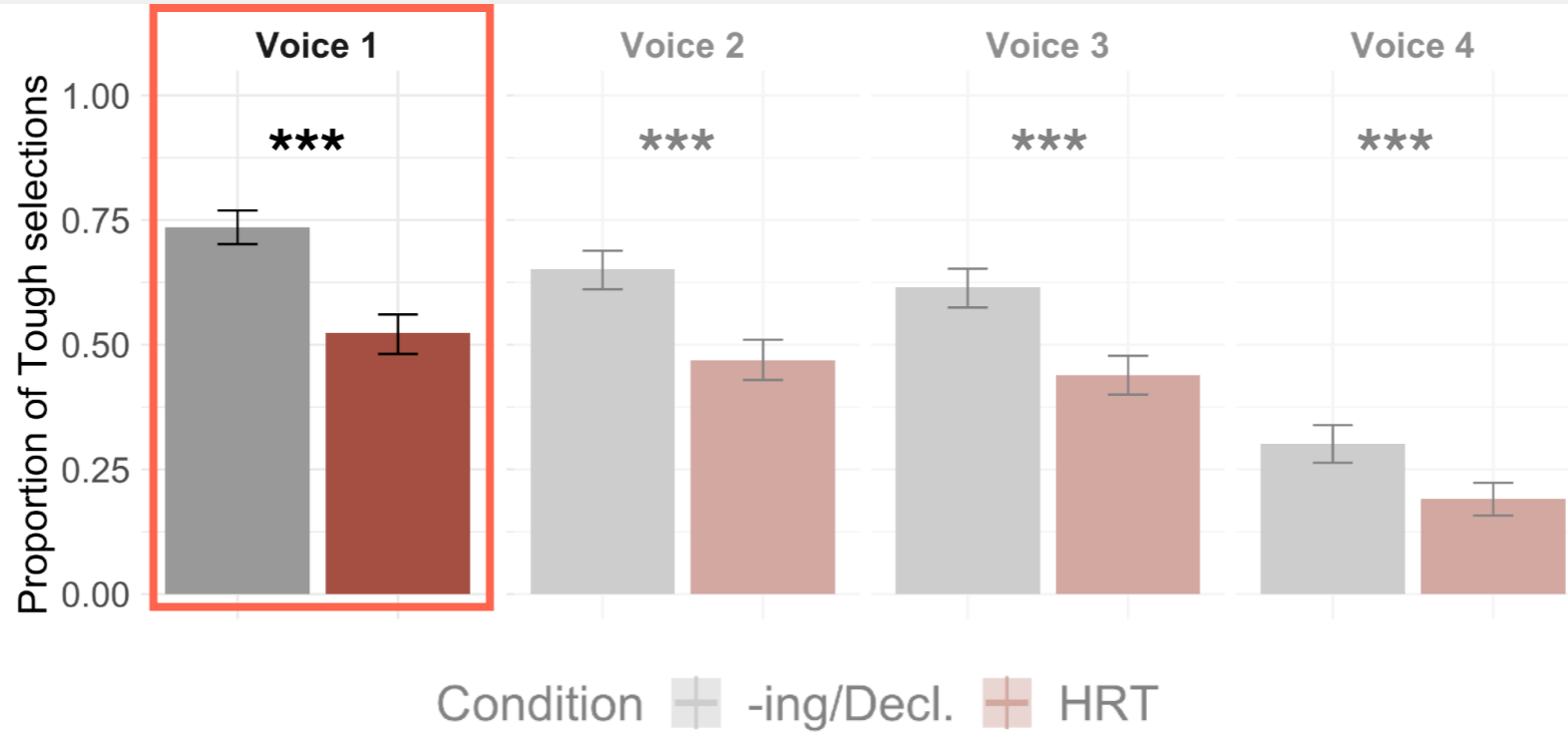
N = 321



Voice-specific effects

Exp. 2b:
(HRT)

N = 321



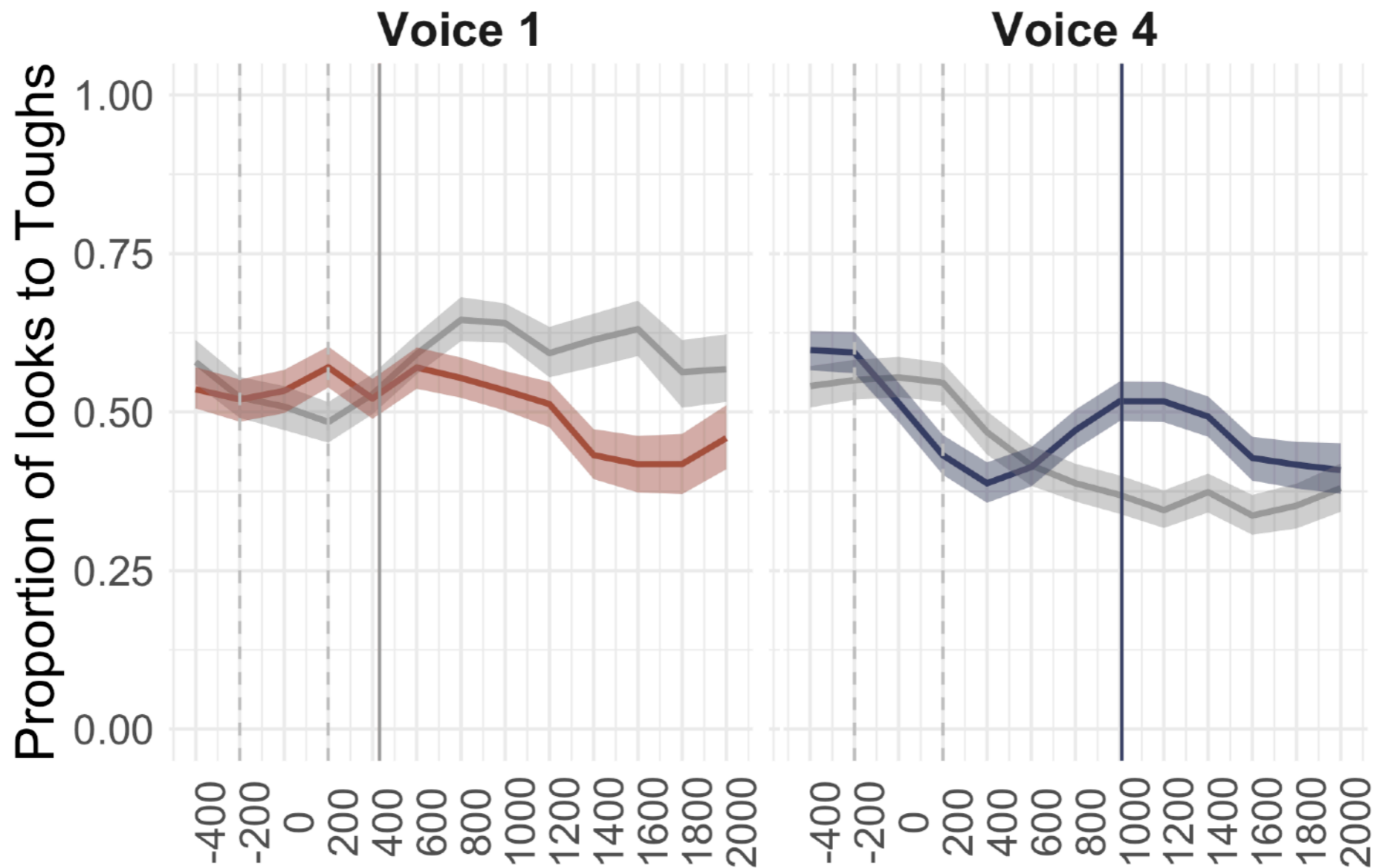
Summary

Aggregate effects of cue replicated in online and offline measures

Listeners weighed the meaning contributions of the cues against their existing expectations about speaker persona

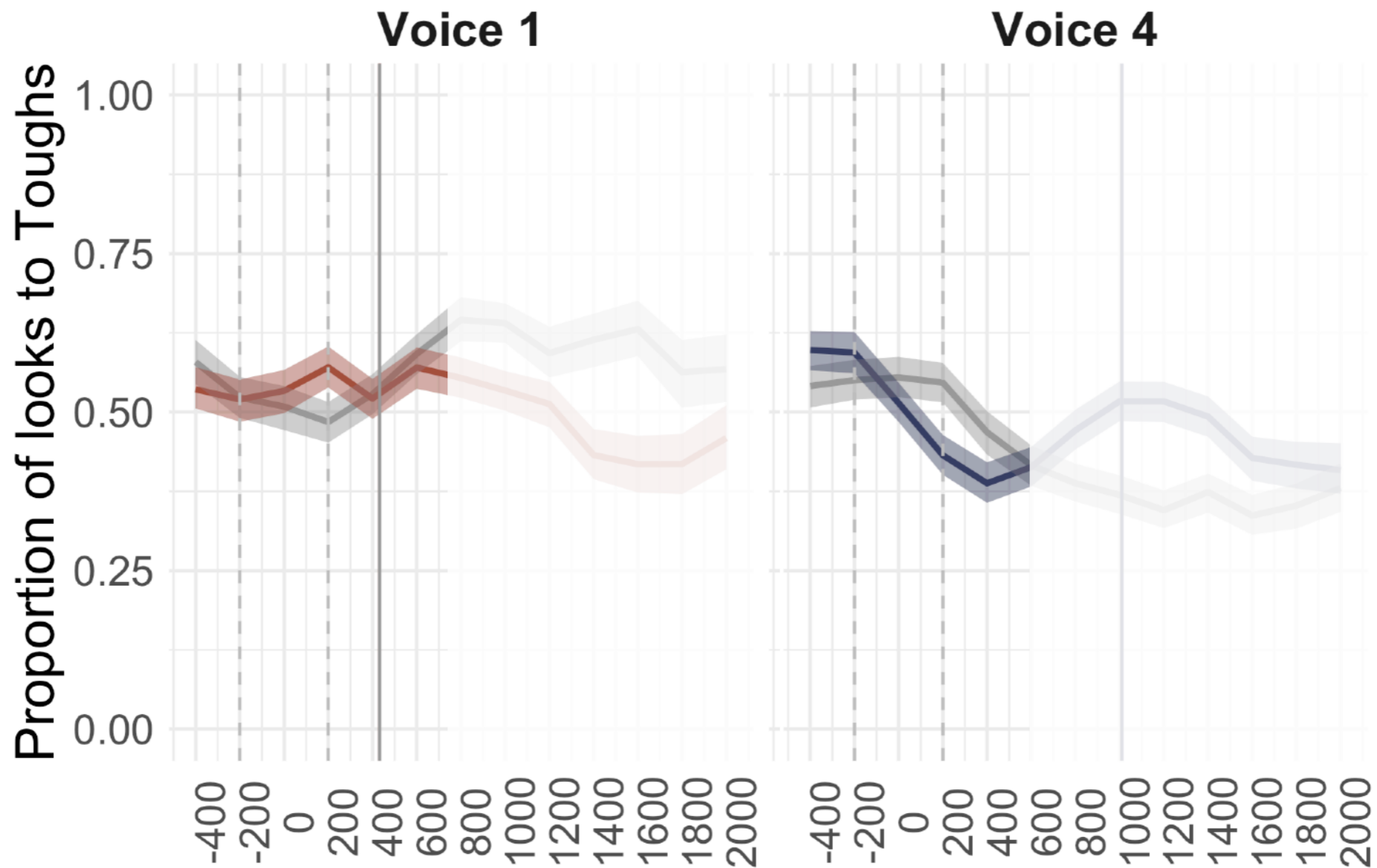
Discussion

No significant differences across voices in categorical effects, but timecourse effects suggest differences



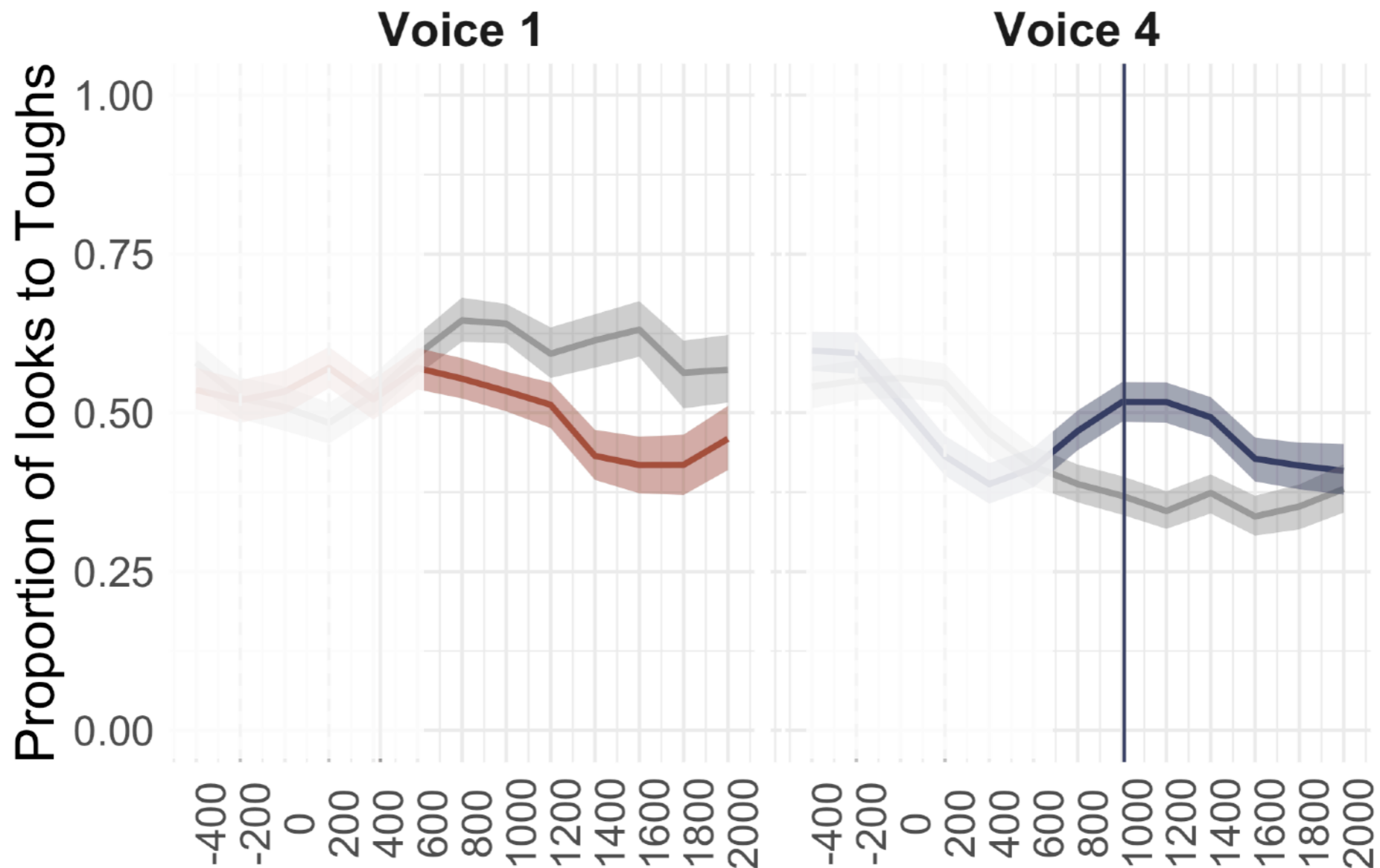
Discussion

No significant differences across voices in categorical effects, but timecourse effects suggest differences



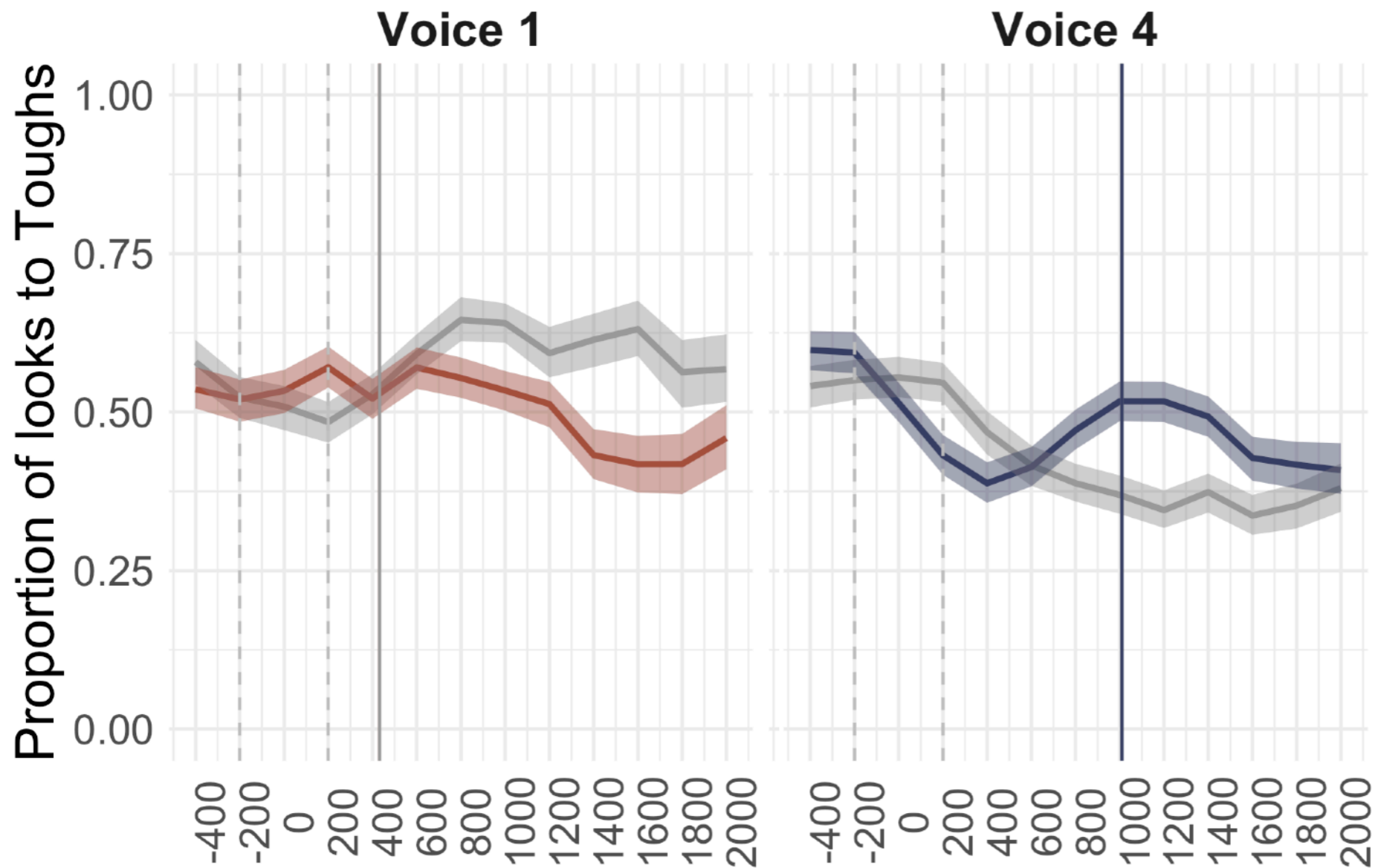
Discussion

No significant differences across voices in categorical effects, but timecourse effects suggest differences



Discussion

No significant differences across voices in categorical effects, but timecourse effects suggest differences



Categorical selections not fine-grained enough?

Surprisal effects were short lived?

3

How do listeners reconcile conflicting socioindexical cues?

Do listeners *revise* their impressions?

In the face of conflicting information, do listeners place more *weight on* some cues more than others?

1 x eye-tracking experiment

-in'	I'm talk in ' about the beam ↘	Tough-congruent
HRT	I'm talk ing about the beam ↗	Valley-congruent
-ing/Decl.	I'm talk ing about the beam ↘	Cue-clash
in'/HRT	I'm talk in ' about the beam ↗	Cue-clash

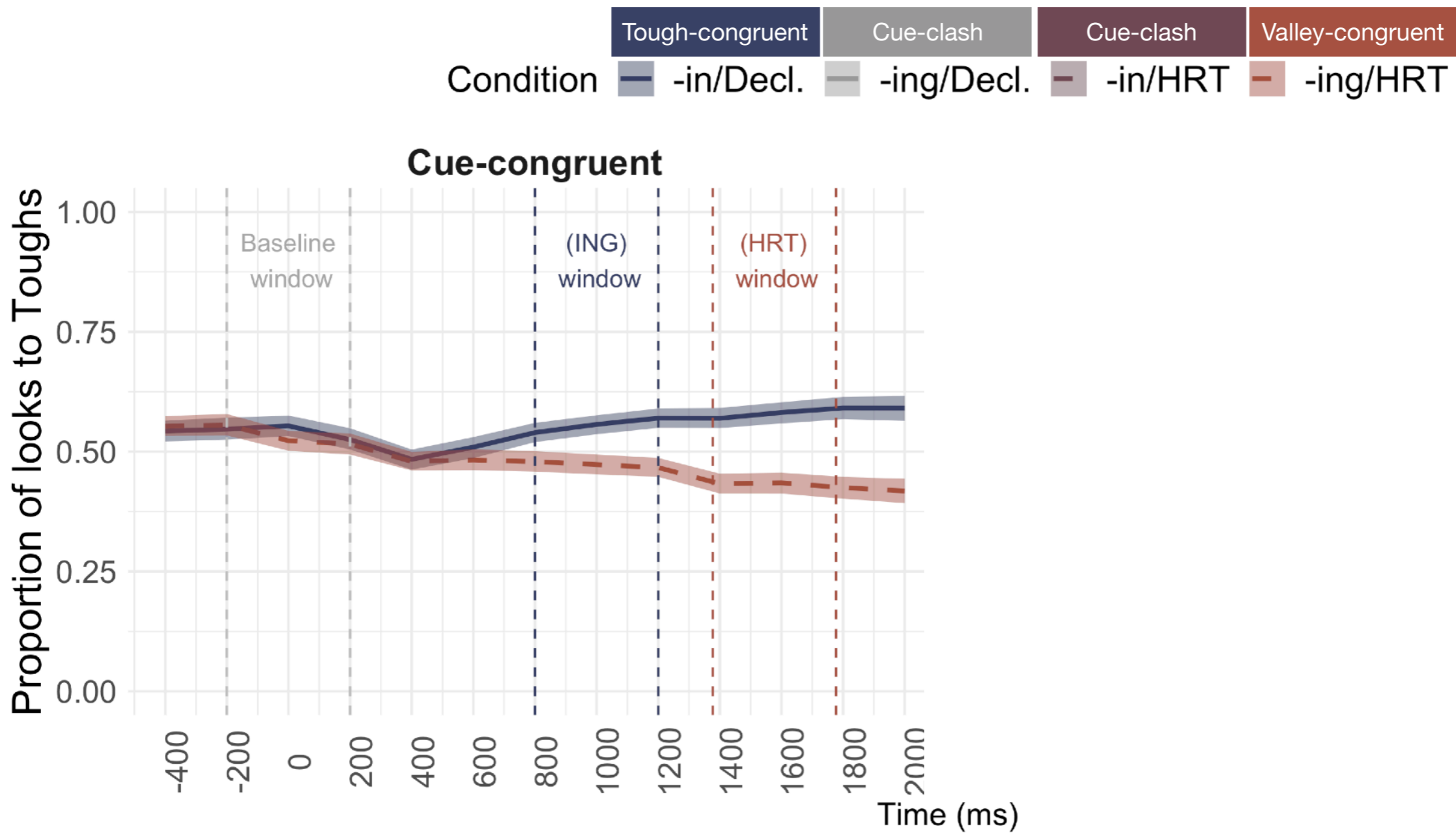
Per voice, listeners heard one of each:

-in'

HRT

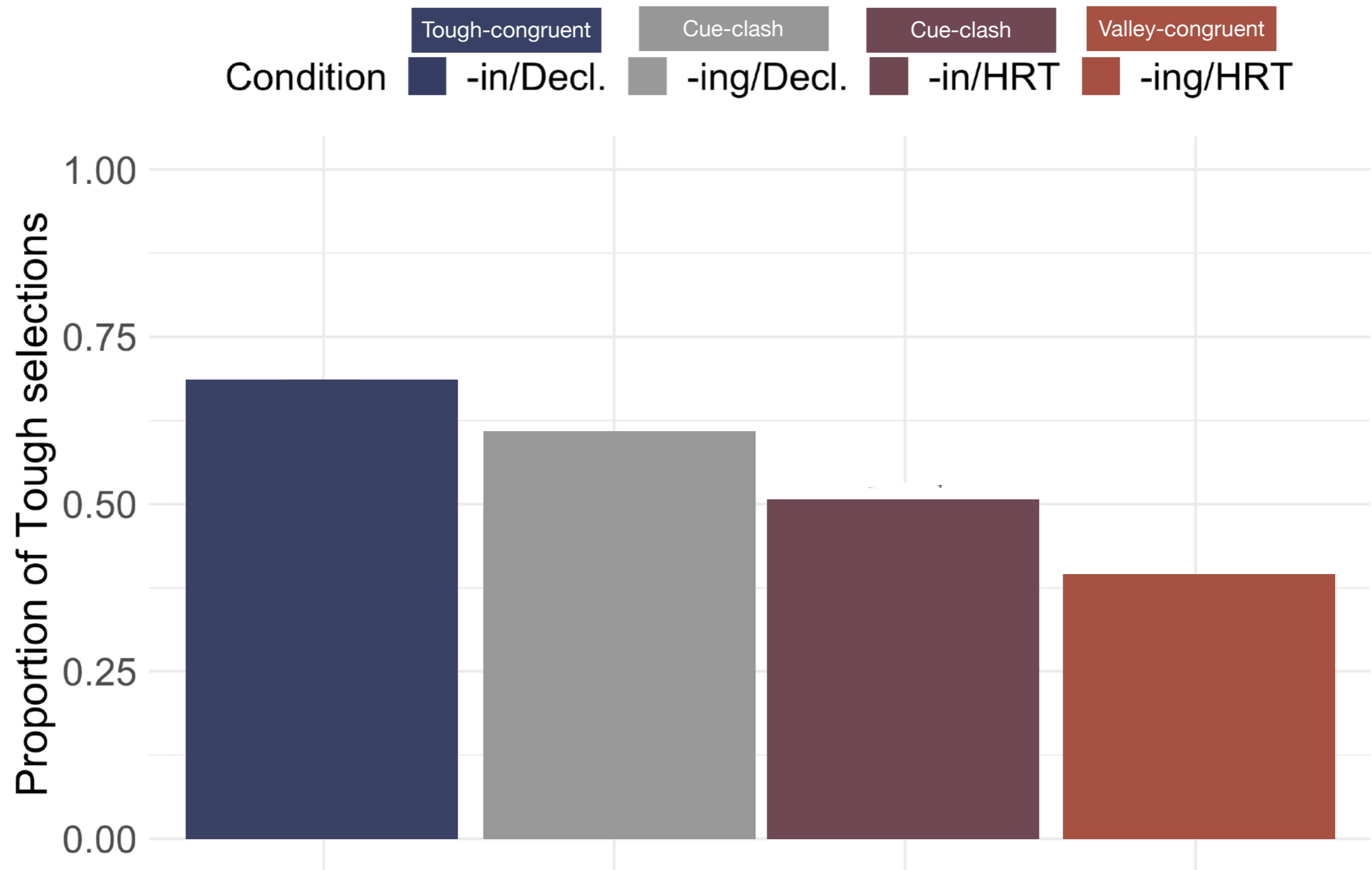
-ing/Decl.

in'/HRT

Do listeners *revise* their impressions?

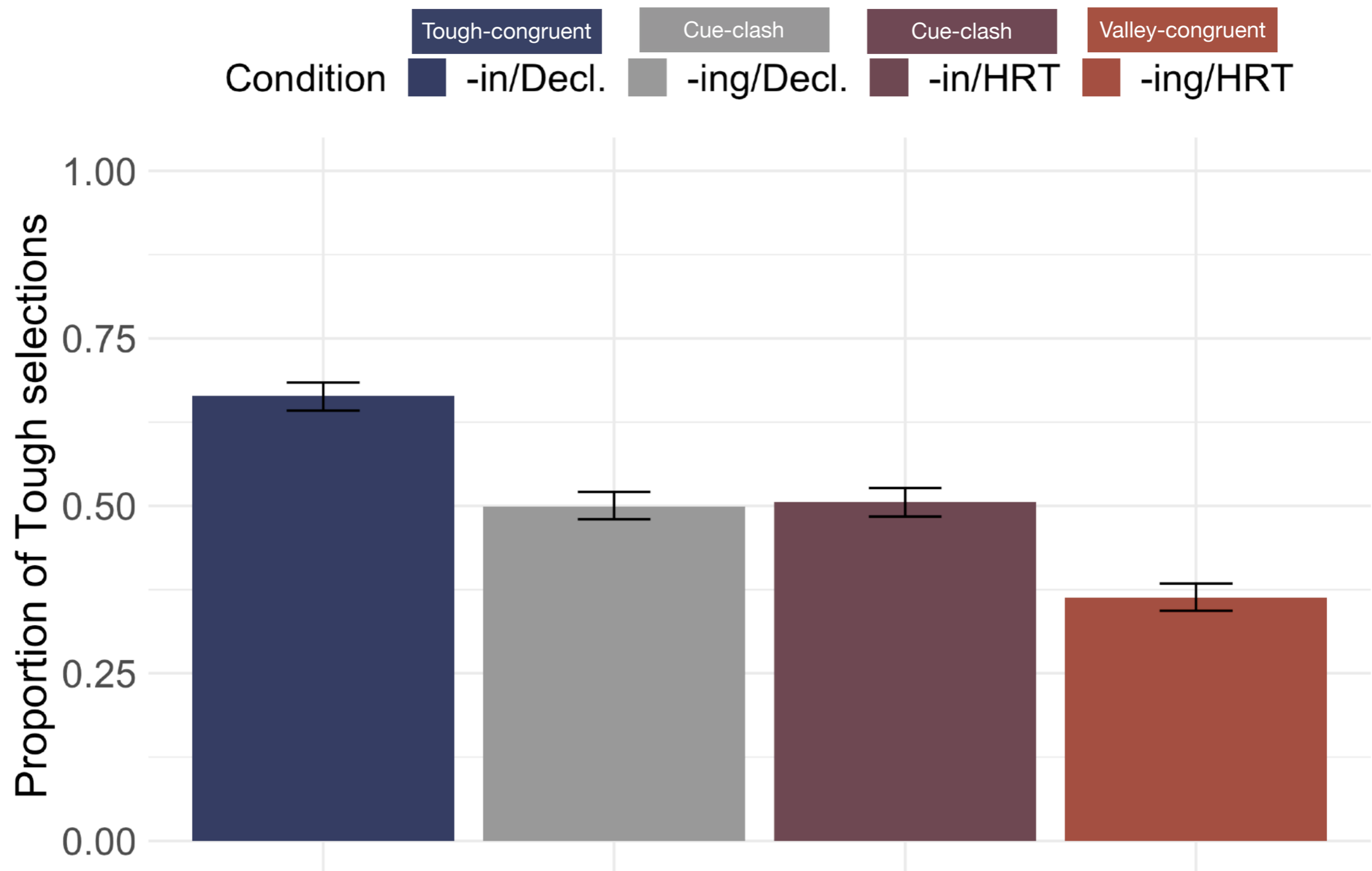
Do listeners *weight* some cues more than others?

Aggregate categorical *predictions*



Do listeners *weight* some cues more than others?

Aggregate categorical *results*



Do listeners *weight* some cues more than others?

Voice-specific categorical results

Condition

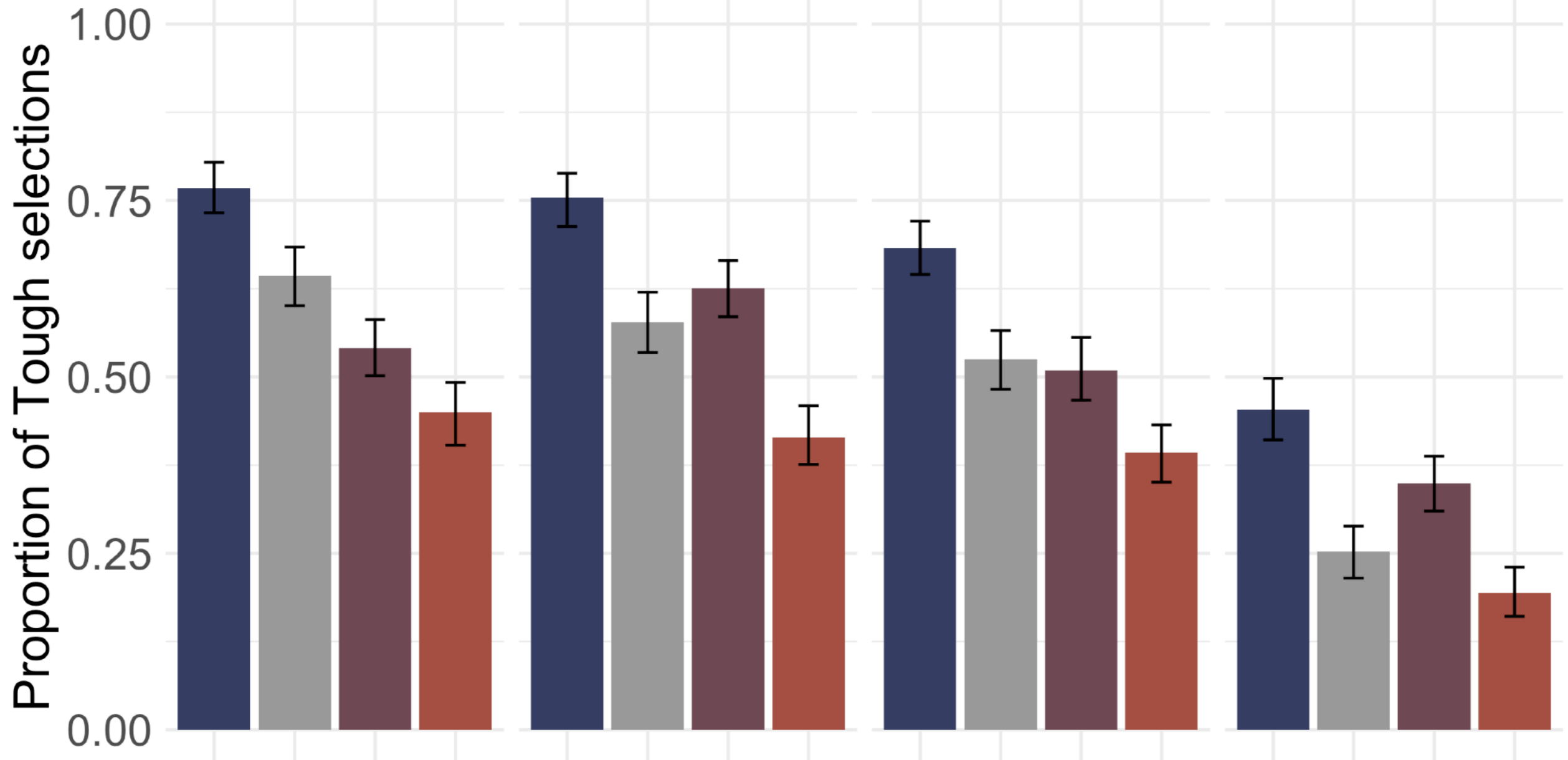
Condition	Tough-congruent	Cue-clash	Cue-clash	Valley-congruent
-in/Decl.	Dark Blue	Grey	Dark Purple	Dark Red
-ing/Decl.	Dark Blue	Grey	Dark Purple	Dark Red
-in/HRT	Dark Blue	Grey	Dark Purple	Dark Red
-ing/HRT	Dark Blue	Grey	Dark Purple	Dark Red

Voice 1

Voice 2

Voice 3

Voice 4



Do listeners *weight* some cues more than others?

Voice-specific categorical results

Condition

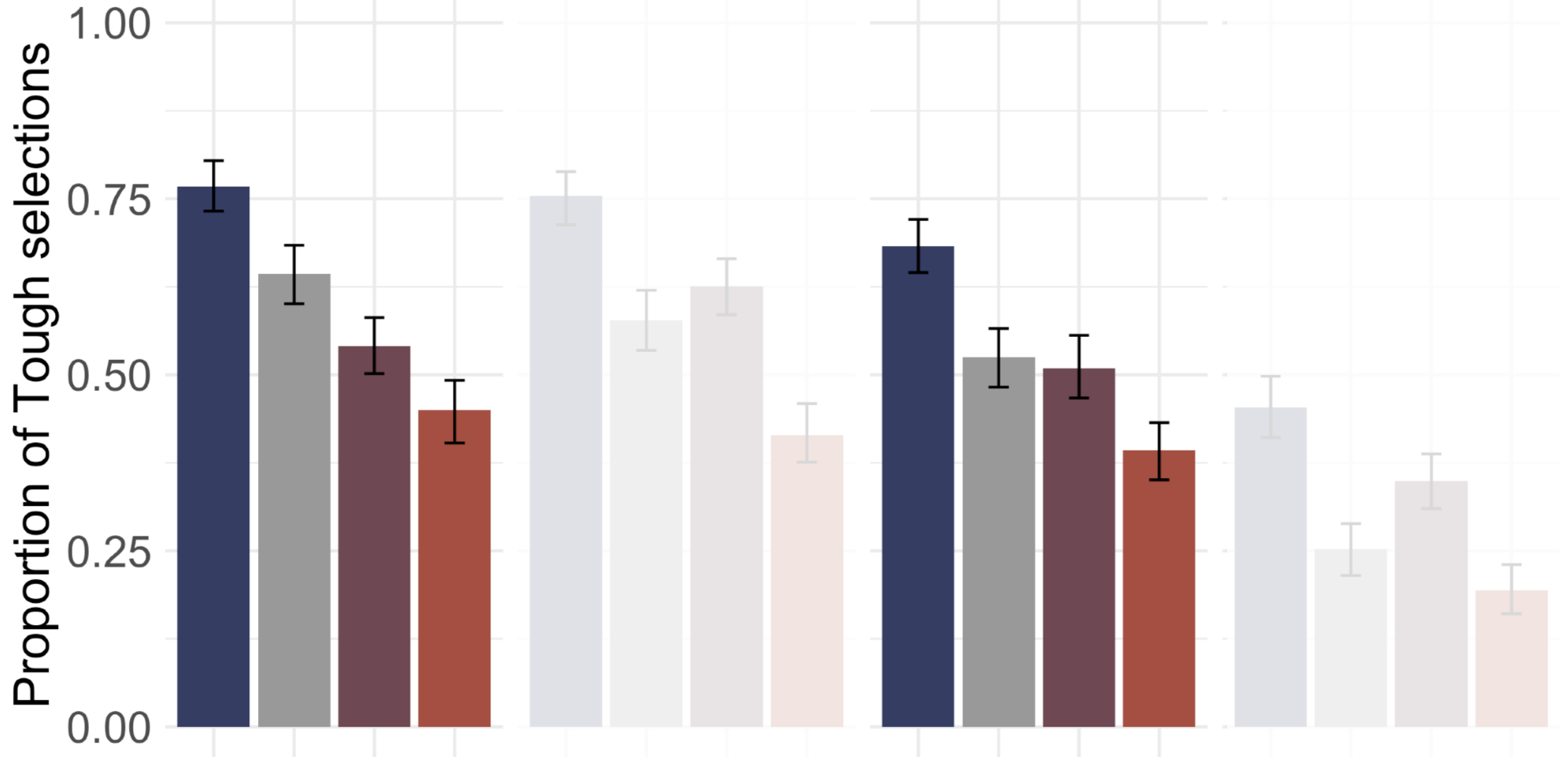
Tough-congruent	Cue-clash	Cue-clash	Valley-congruent
-in/Decl.	-ing/Decl.	-in/HRT	-ing/HRT

Voice 1

Voice 2

Voice 3

Voice 4



Do listeners *weight* some cues more than others?

Voice-specific categorical results

Condition

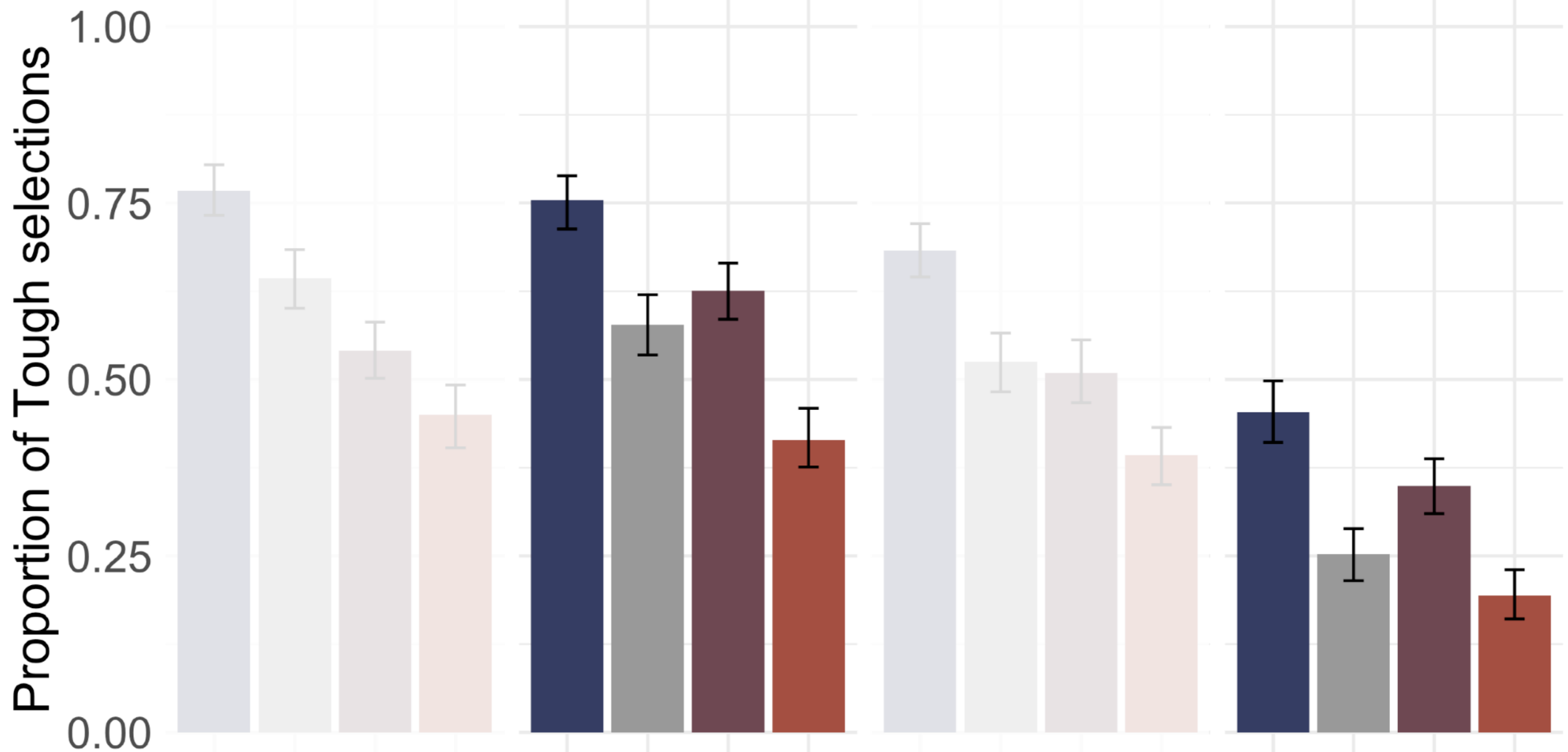
Tough-congruent	Cue-clash	Cue-clash	Valley-congruent
-in/Decl.	-ing/Decl.	-in/HRT	-ing/HRT

Voice 1

Voice 2

Voice 3

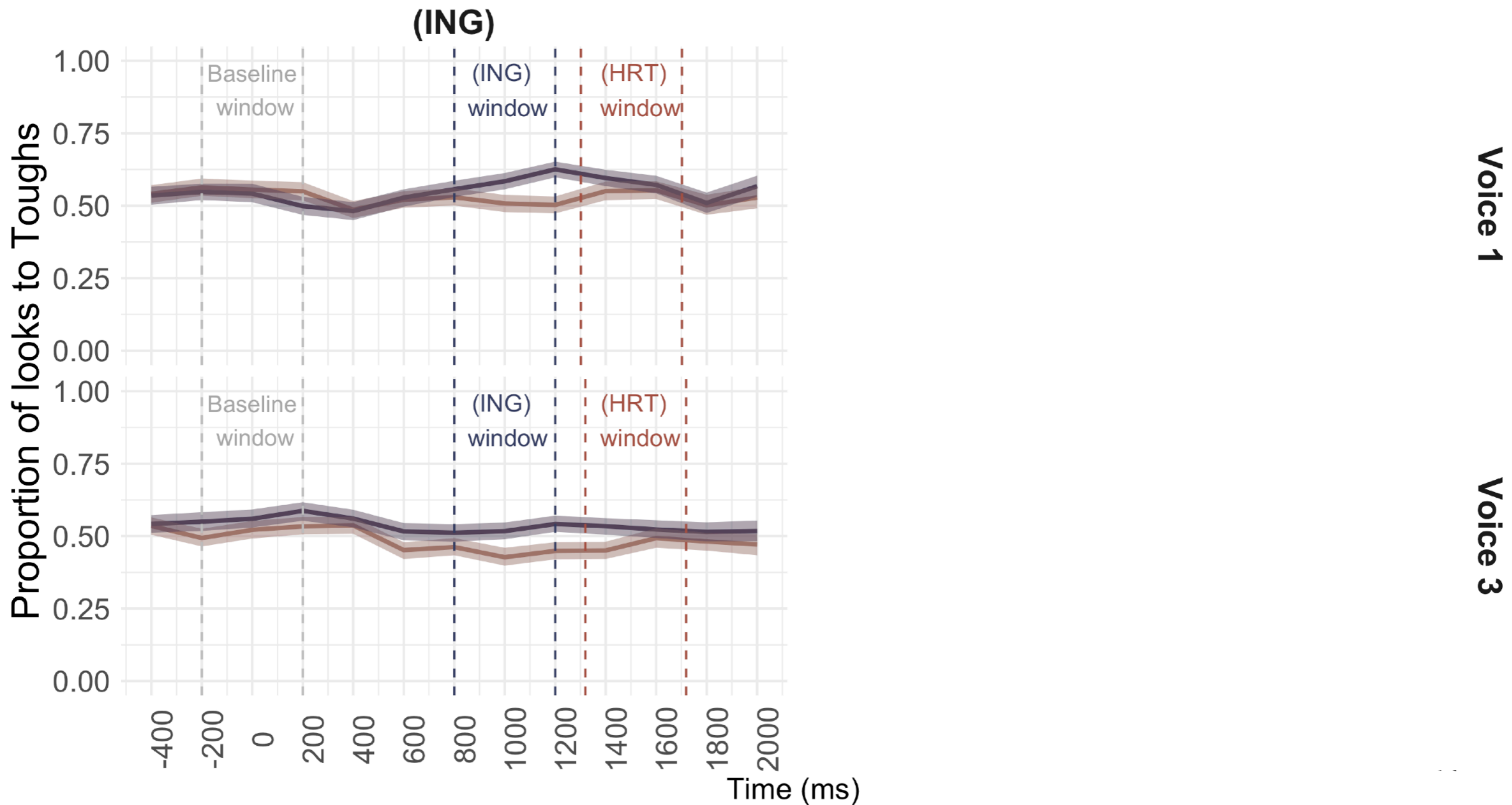
Voice 4



Do listeners *weight* some cues more than others?

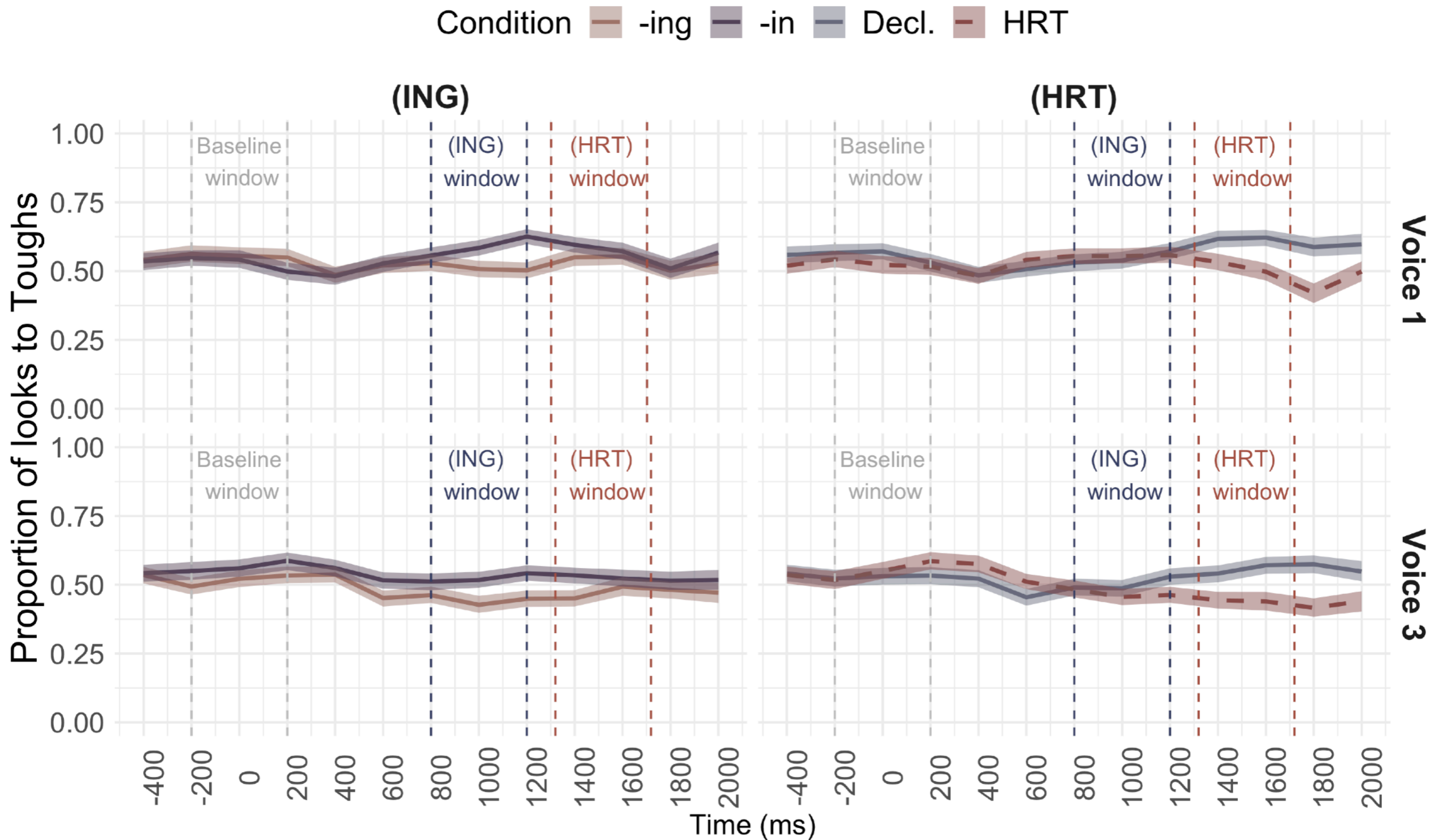
Voices 1 and 3: time course results

Condition ■ -ing ■ -in ■ Decl. ■ HRT



Do listeners *weight* some cues more than others?

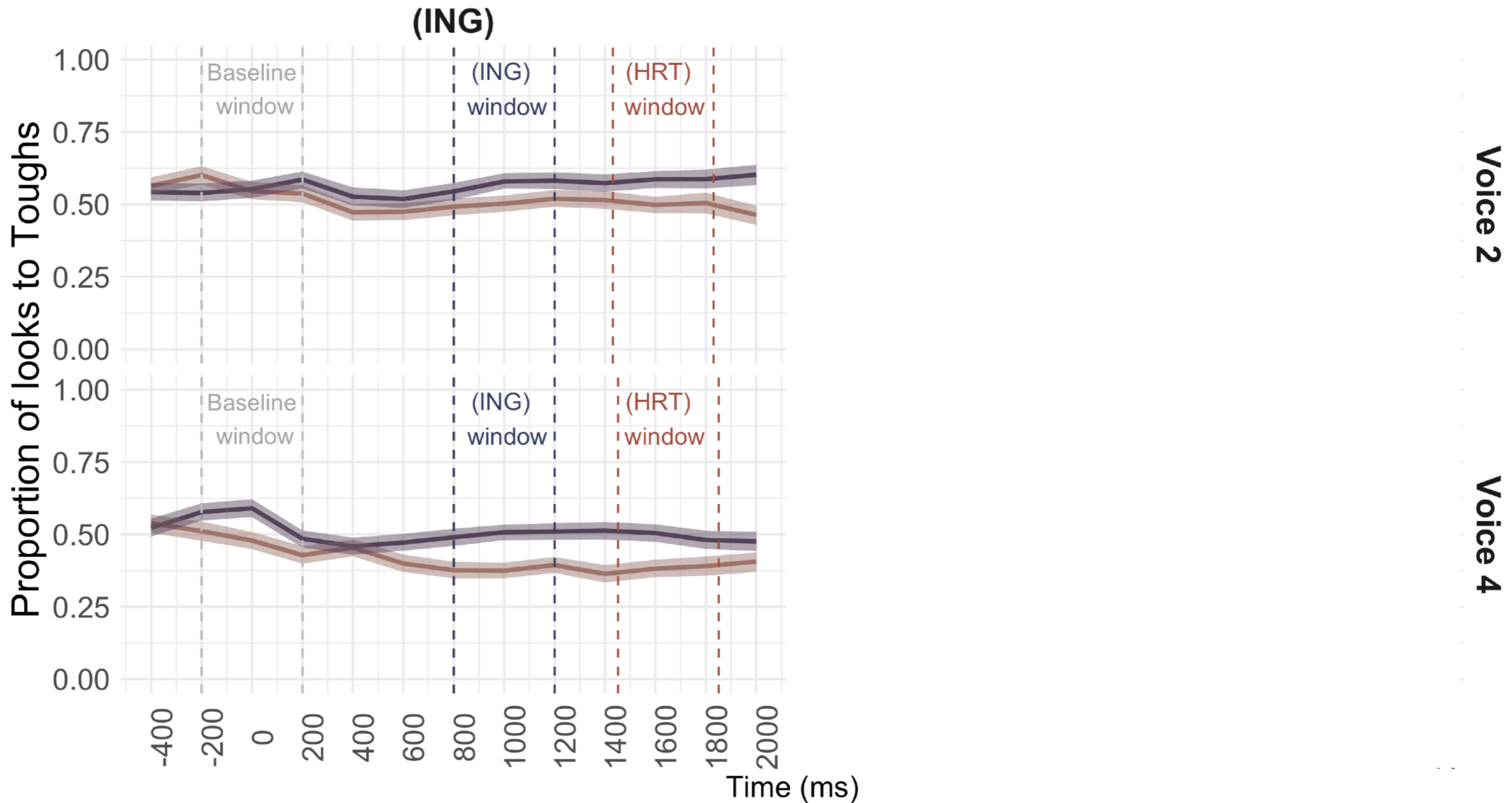
Voices 1 and 3: time course results



Do listeners *weight* some cues more than others?

Voices 2 and 4: time course results

Condition ■ -ing ■ -in ■ Decl. ■ HRT



Why these voice-specific cue weights?

An 'easy' answer:

'acoustic salience'

A less 'easy', less quantifiable, and more interesting answer:

For Voices 1 and 3, the specific *productions* of HRT indexed this *specific* Valley Girl to a greater extent than for Voices 2 and 4

... and vice versa for Voices 2 and 4

Three Questions

- 1 Do listeners reconcile socioindexical cues when making online inferences about speaker persona?
- 2 How do existing expectations about a speaker modulate listeners' inferences?
- 3 How do listeners reconcile conflicting socioindexical cues?

Three Answers

- 1 Listeners reconcile the contributions of socioindexical cues to draw inferences about a speaker's persona, as and when they talk
... but they can always *revise* their impressions
- 2 They do this probabilistically, by weighting the contribution of a cue against their existing expectations
- 3 *How* these weights are derived may be dependent on the broader, cultural socioindexical informativity of a cue — but also, its contributions within a specific voice

Some Implications

What listeners take notice of, and the *degree* to which particular cues shift perceptions and beliefs can inform sociolinguistic theories of how meanings circulate and are reproduced

Sociolinguistic reasoning may well behave like *other* kinds of reasoning

...these methods provide a workable template for further investigation 😊

My Brilliant Committee



Meghan Sumner



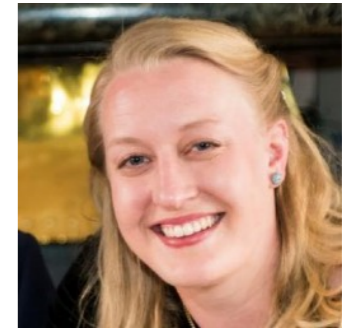
Judith Degen



Rob Podesva



Penny Eckert



Kathryn
Meyer Olivarius



Stanford
Sociolinguists
(especially
Diss Group!)

Thanks for listening!

Bonus slides

Image norming Exp. 1

Free choice associations

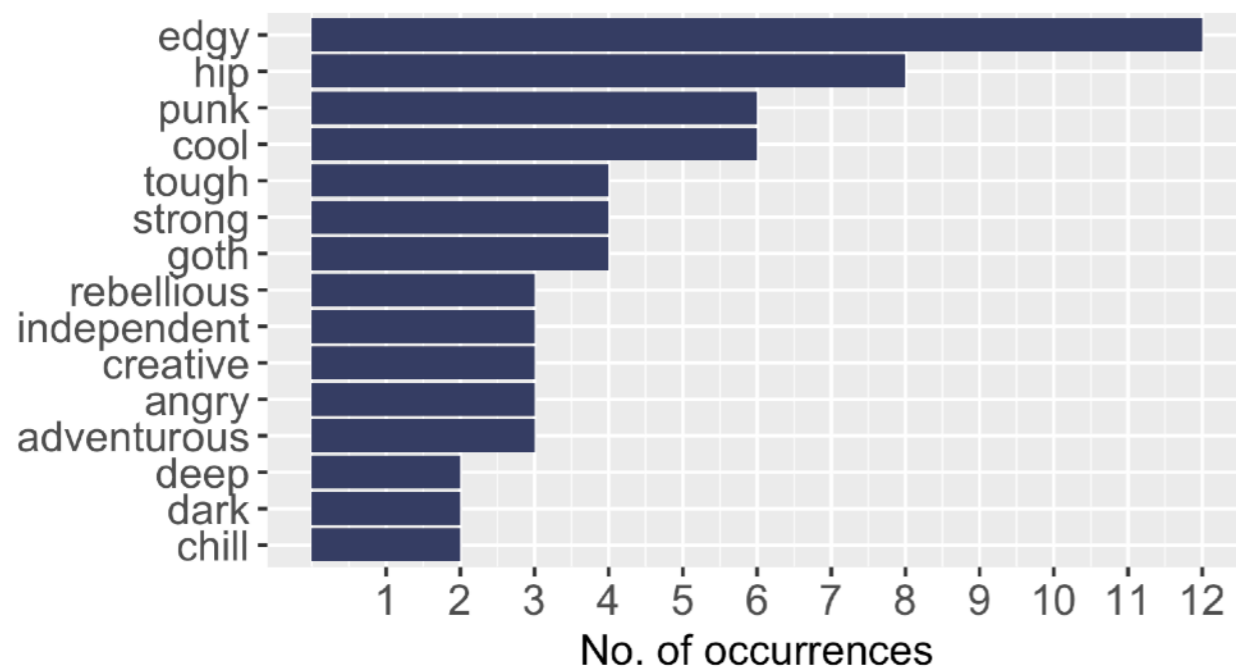


Adjective 1:

Adjective 2:

Adjective 3:

Tough adjectives



Valley Girl adjectives

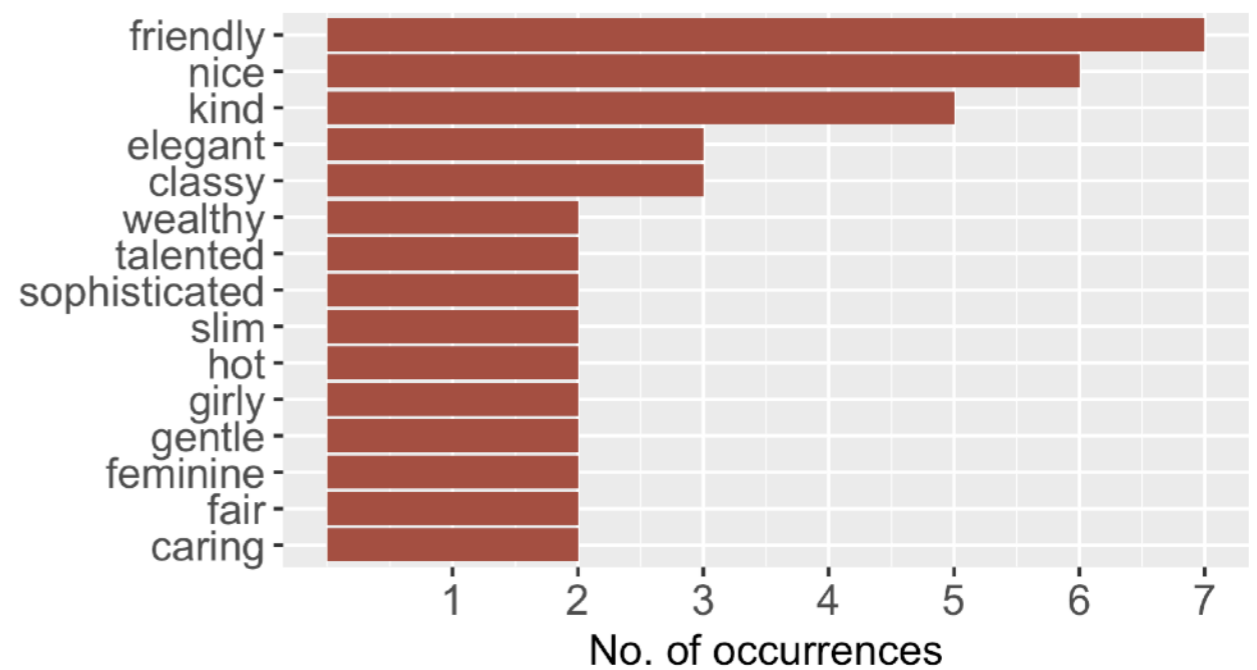


Image norming Exp. 2

Click on all the people that look like they are:

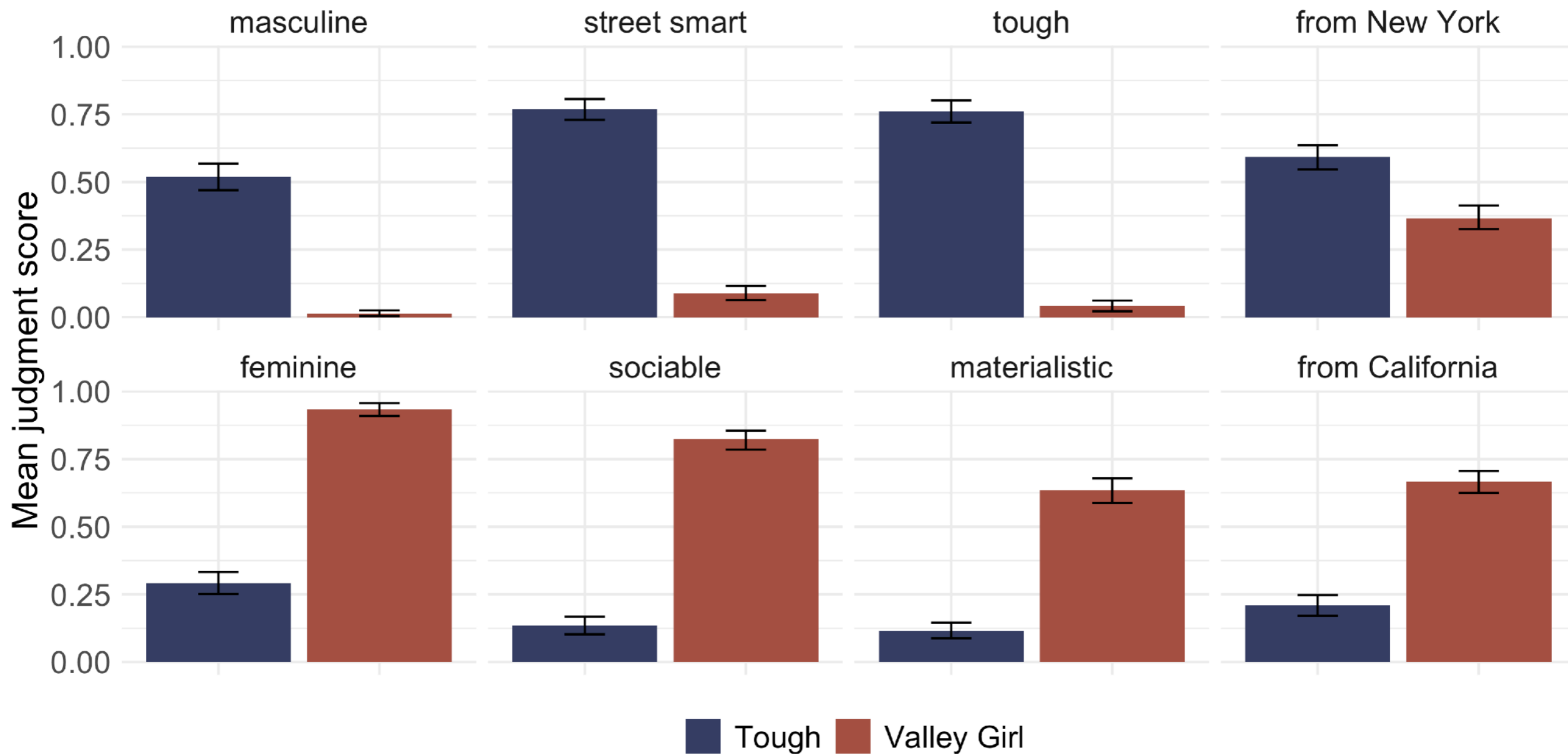
sociable



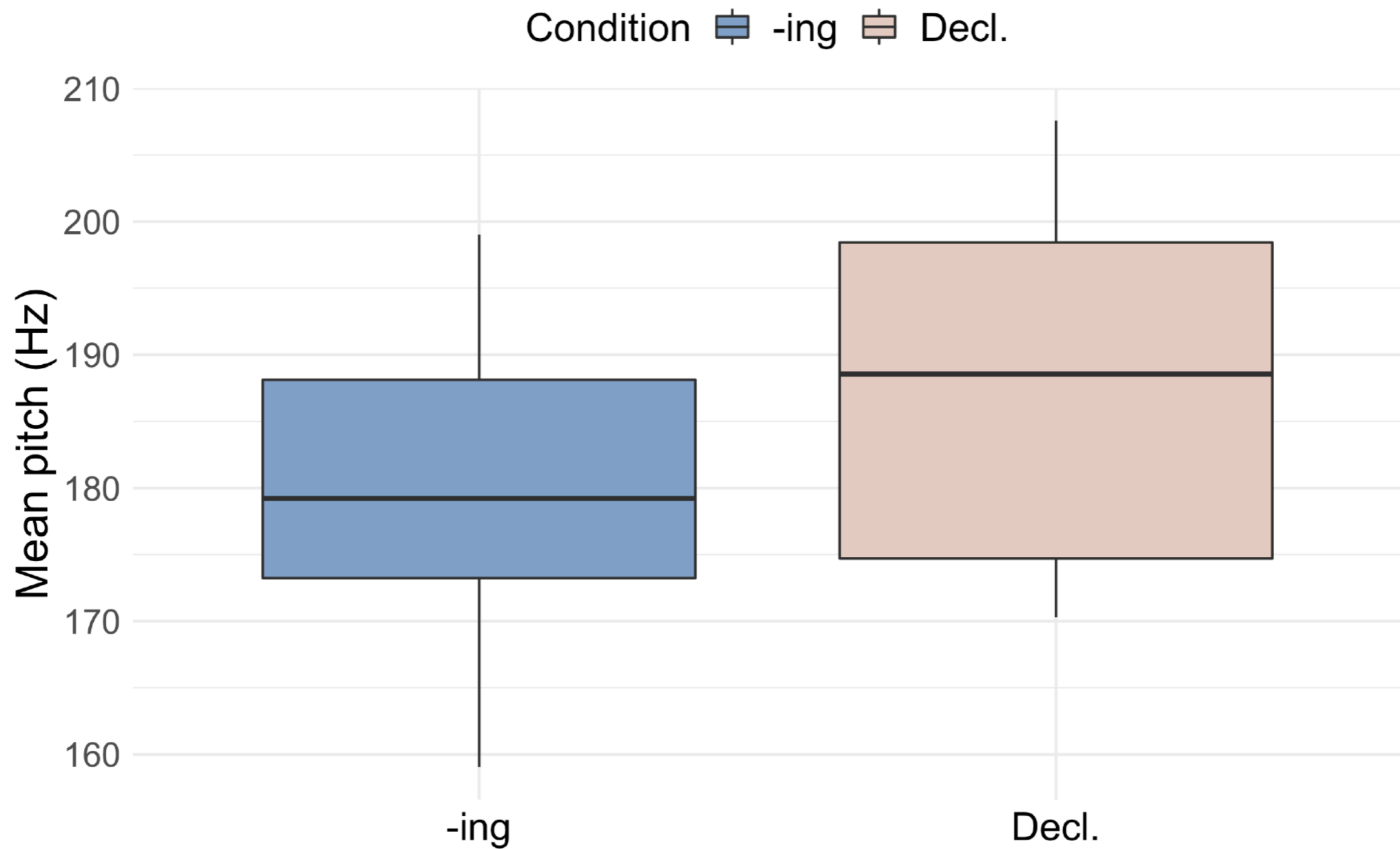
If you didn't select any of the people, explain why here:

Continue

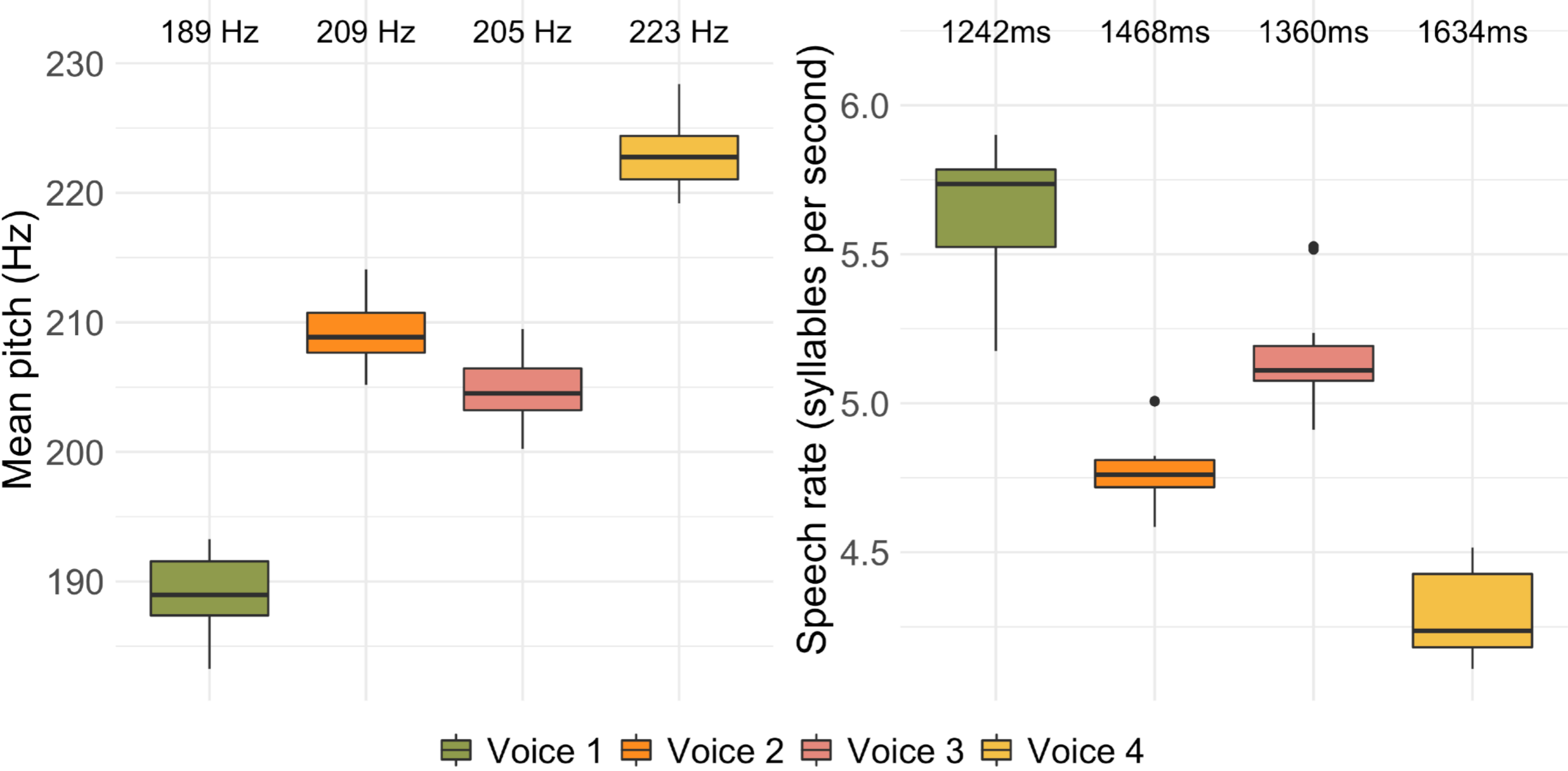
Image norming Exp. 2



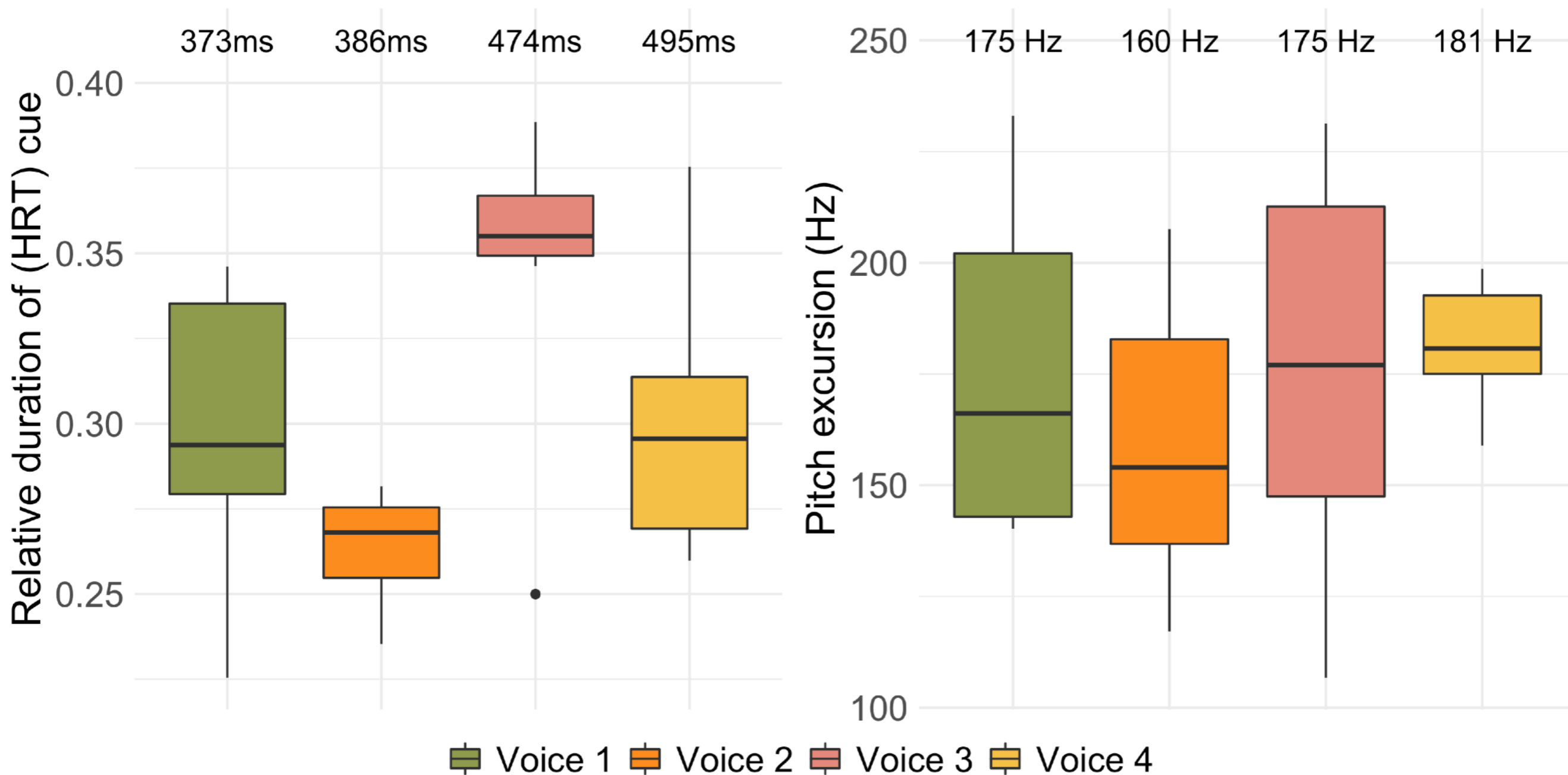
Single critical voice stimuli: Difference in pitch between *-ing* and Declarative tokens



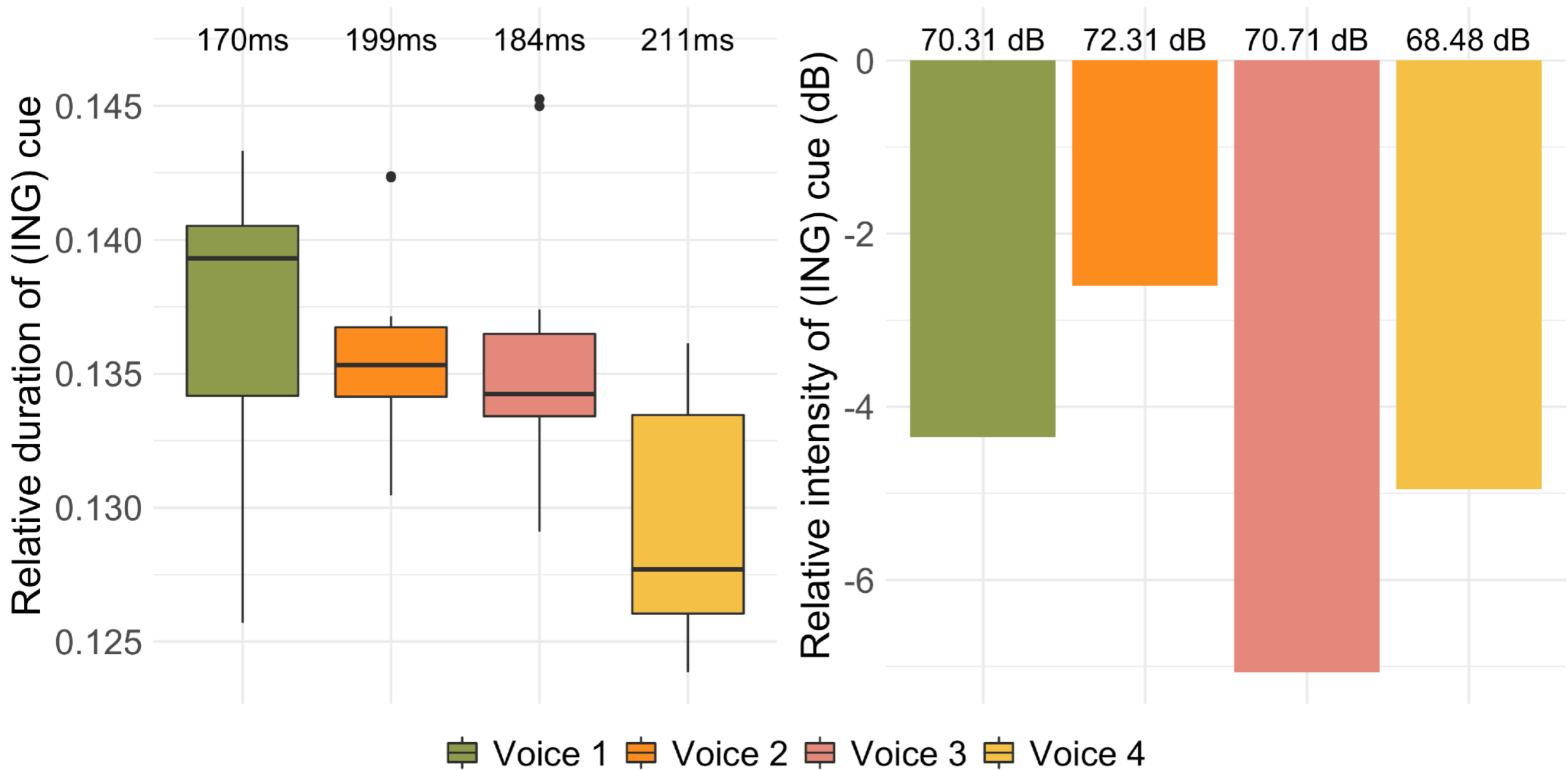
Four critical voices stimuli: Speech rate and pitch



Four critical voices stimuli: HRT cue duration and pitch excursion

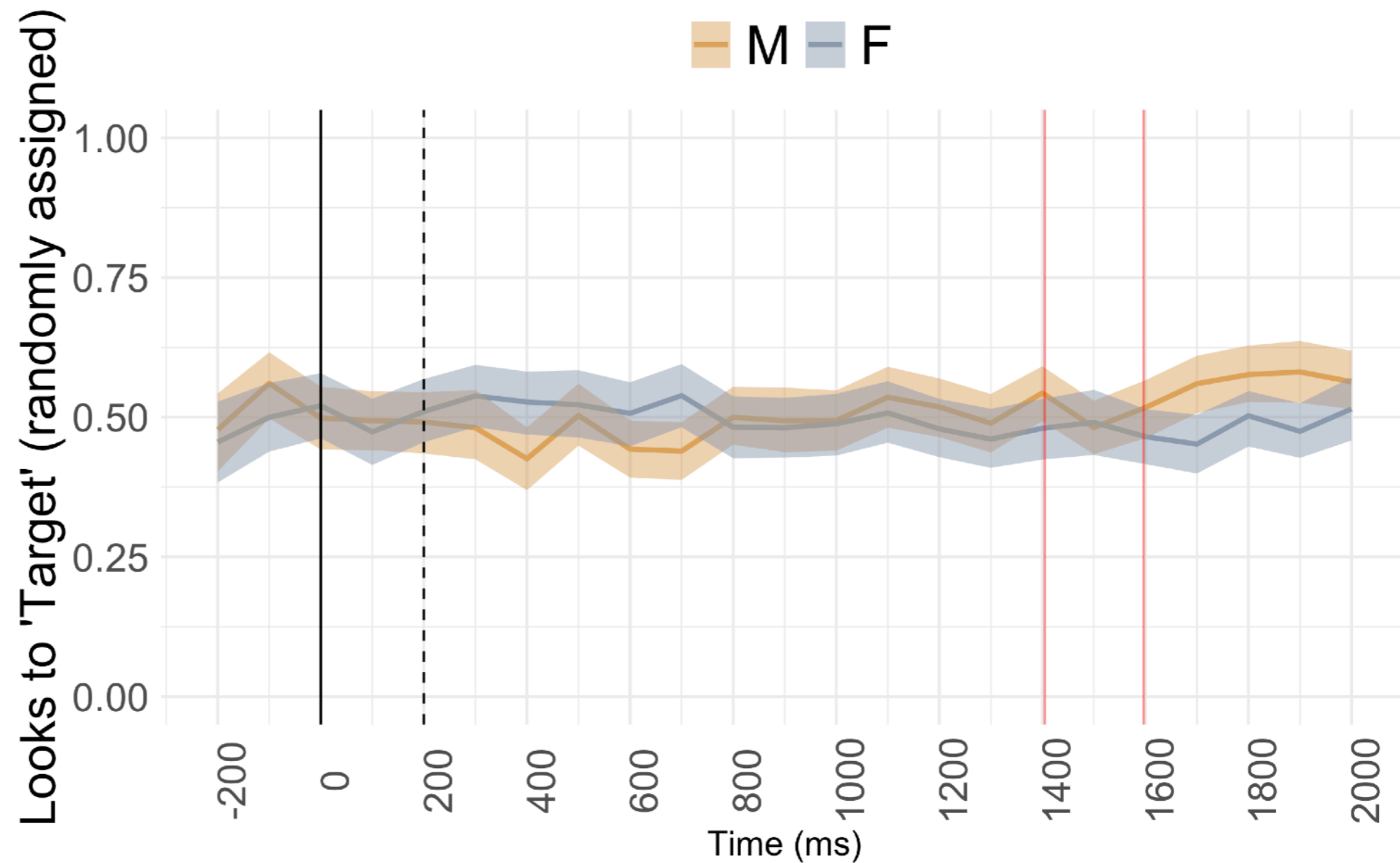


Four critical voices stimuli: *-in'* cue duration and pitch excursion



'Ambiguous' filler trials

Participants either heard a male voice and saw two males, or a female voice and two females



Statistical details

Exps 1 - 2

Categorical selection data

Mixed effects logistic regression models predicting log-odds of selecting the Tough speaker, given presence of -in' (vs. ing) or HRT (vs. Declarative)

Timecourse data

- Bayesian mixed effects logistic regression models predicting log-odds of looking at the Tough speaker (vs. Valley Girl) in each 100ms interval after cue onset.
- 200ms either side of cue onset was taken as a 'Baseline' with which to compare looks in subsequent windows.
- Condition (-in' vs. ing, HRT vs. Declarative), Interval (Baseline vs. 100ms interval), and their interaction were included as predictors.

Random effects structure

Maximal random effects structure justified by the design. Minimally, this included random intercepts for participant, item and speaker, and random by-speaker and by-participant slopes for condition.

1

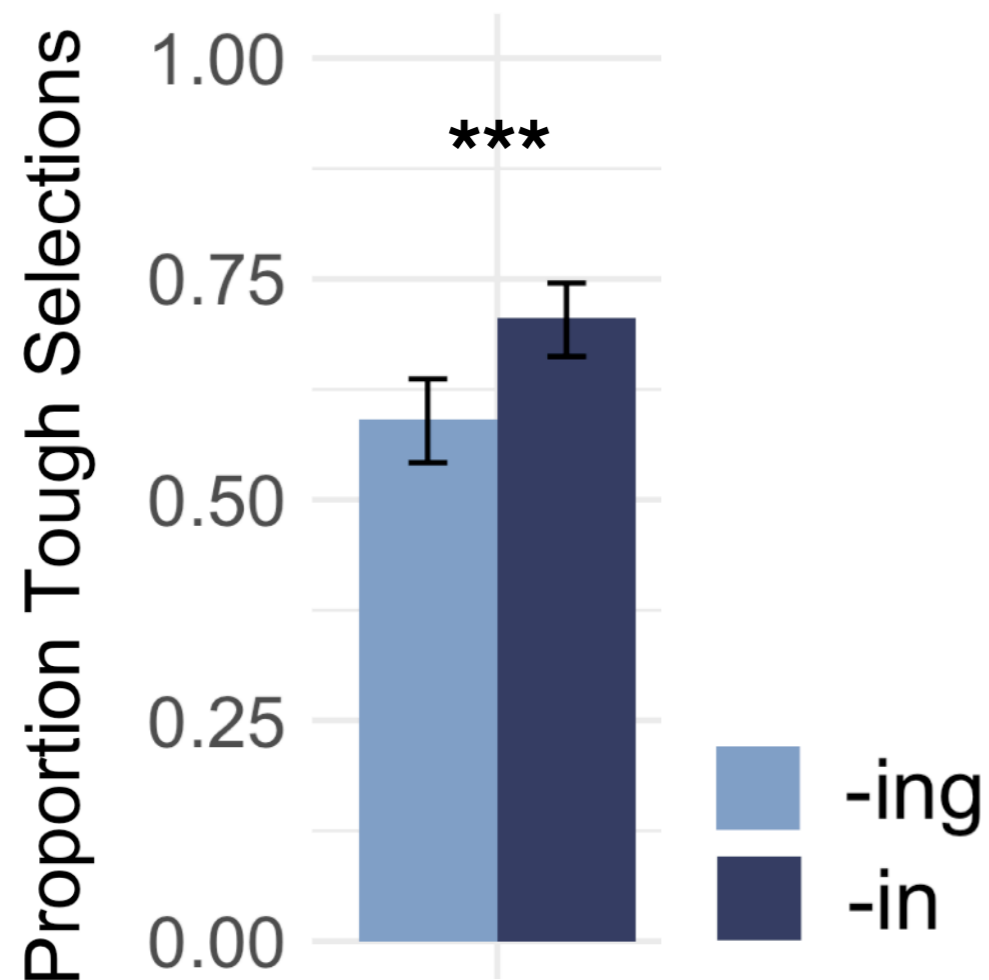
How does the strength of the indexical link between cue and persona modulate cue uptake?

3 x norming experiments (no eye tracking)

Exp. 1

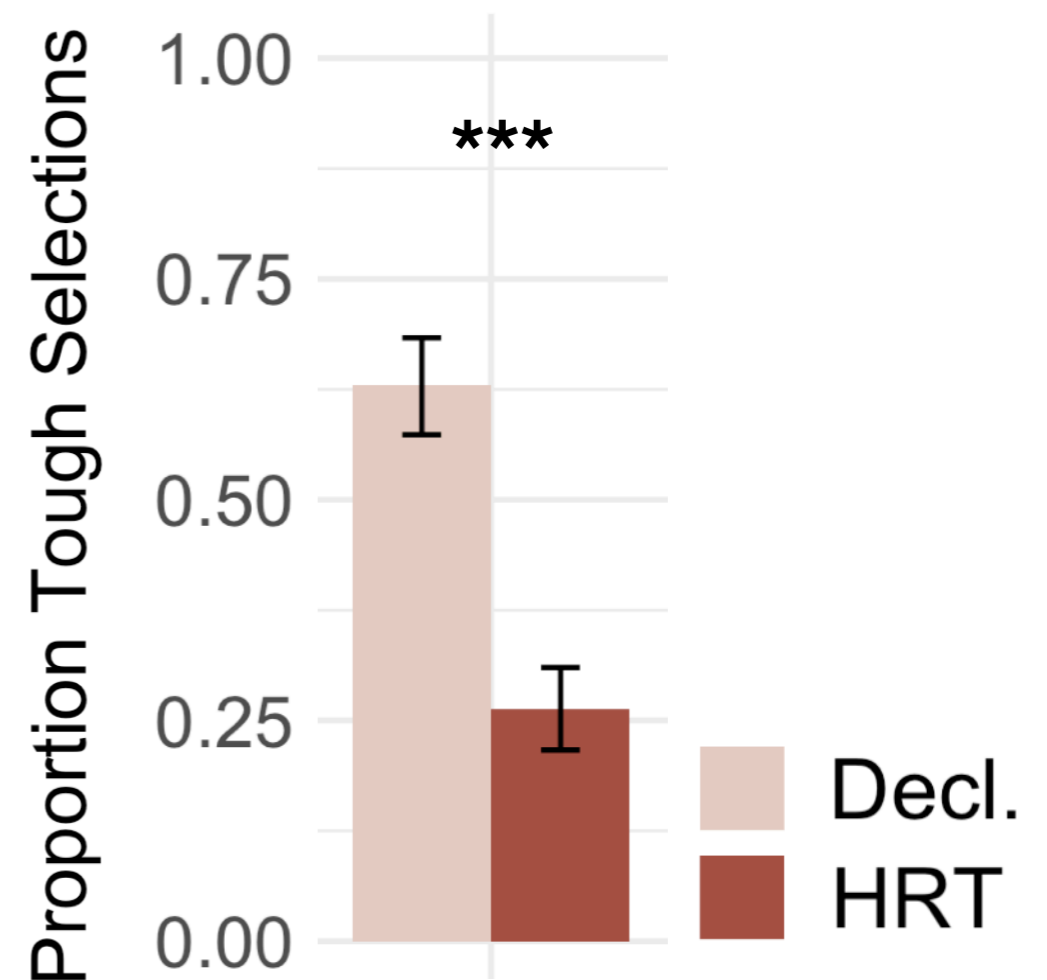
Exp. 1a

Listeners hear only (ING) cues



Exp. 1b

Listeners hear only (HRT) cues



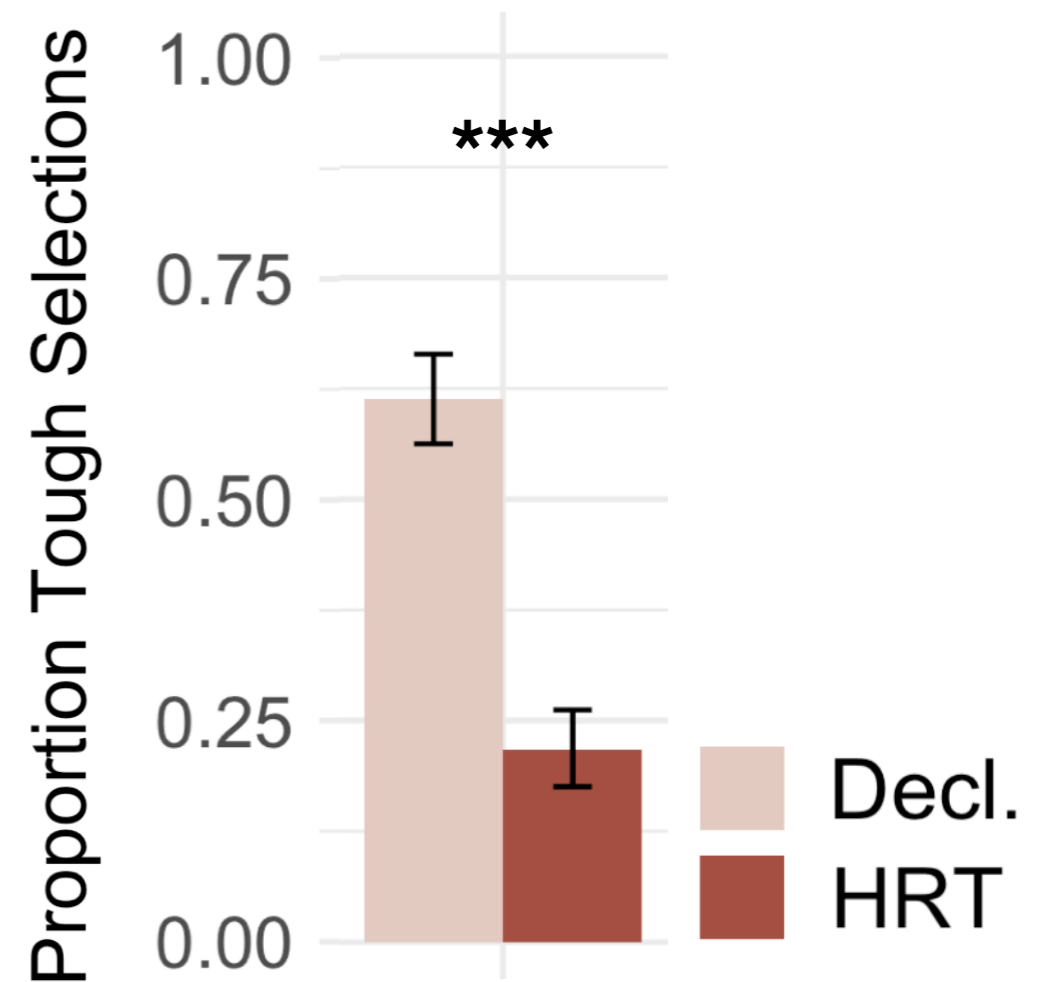
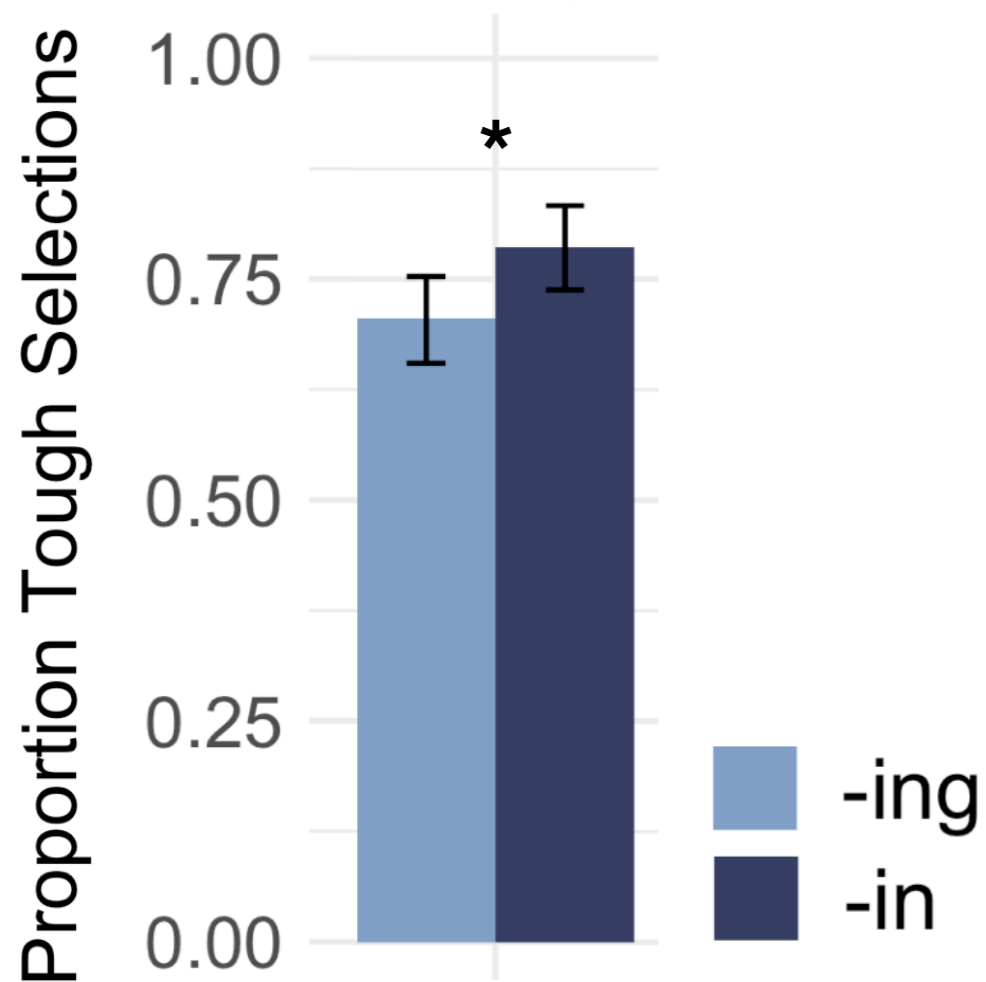
1

How does the strength of the indexical link between cue and persona modulate cue uptake?

3 x norming experiments (no eye tracking)

Exp. 2

Listeners hear both (ING) and (HRT) cues.
The order they hear them is random.



1

How does the strength of the indexical link between cue and persona modulate cue uptake?

3 x norming experiments (no eye tracking)

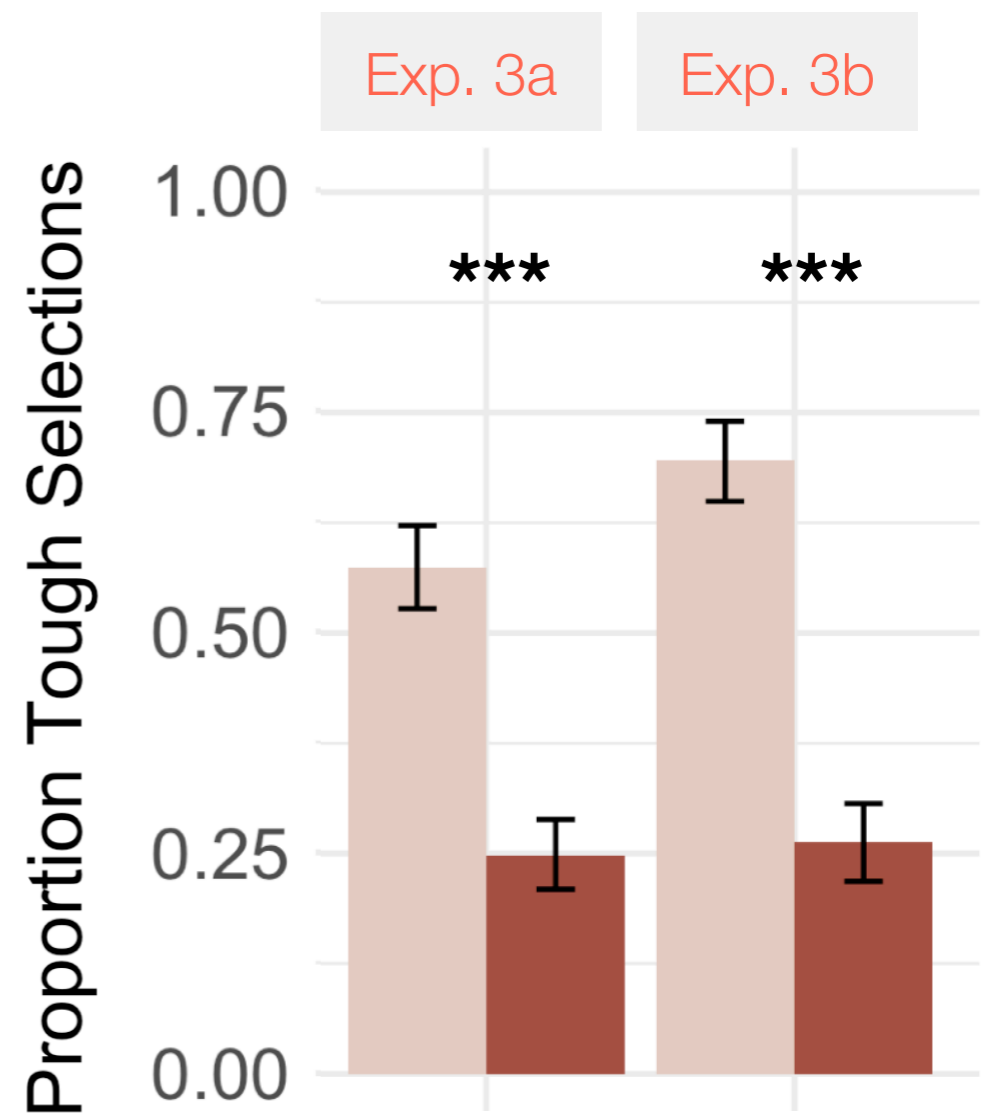
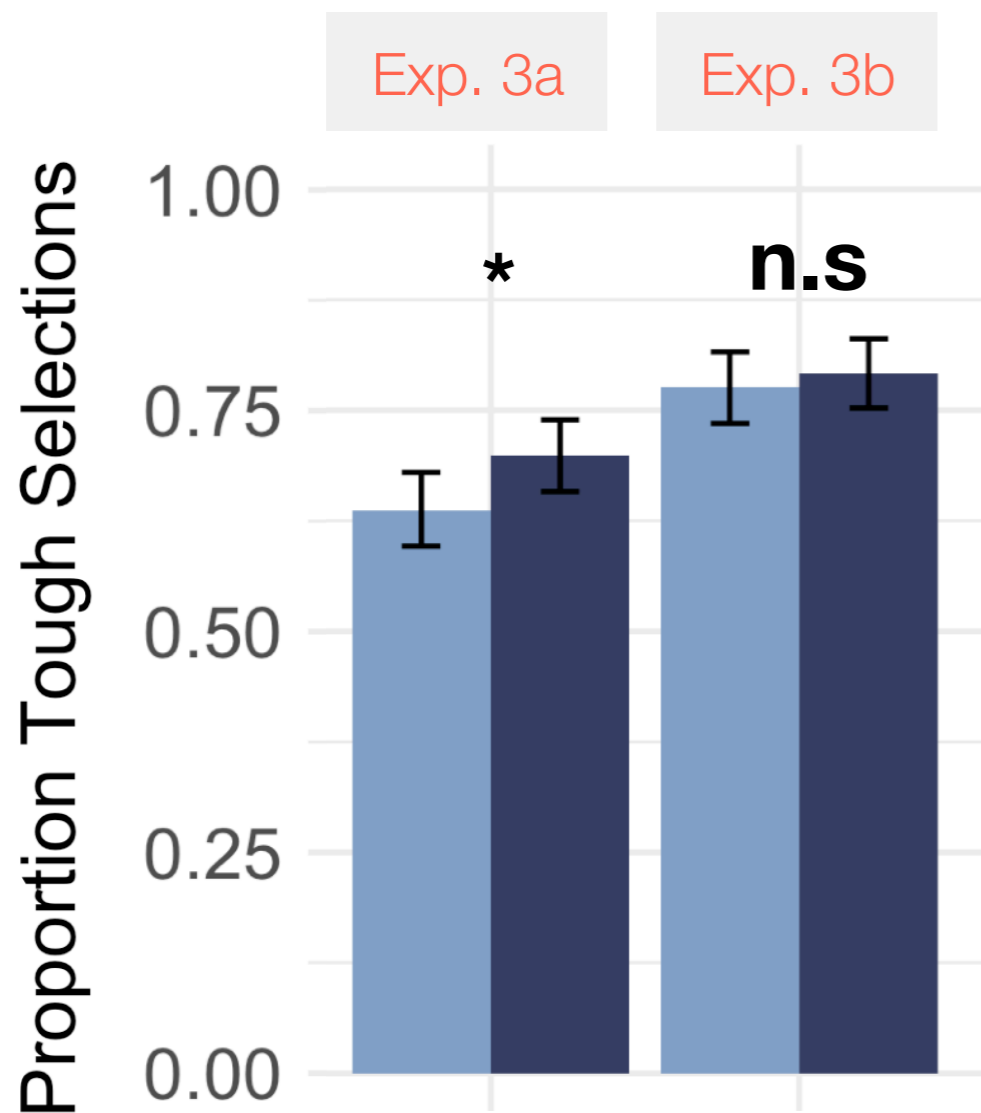
Exp. 3

Exp. 3a: Listeners hear only (ING) cues, then only (HRT) cues

Exp. 3b: Listeners hear only (HRT) cues, then only (ING) cues

(ING)

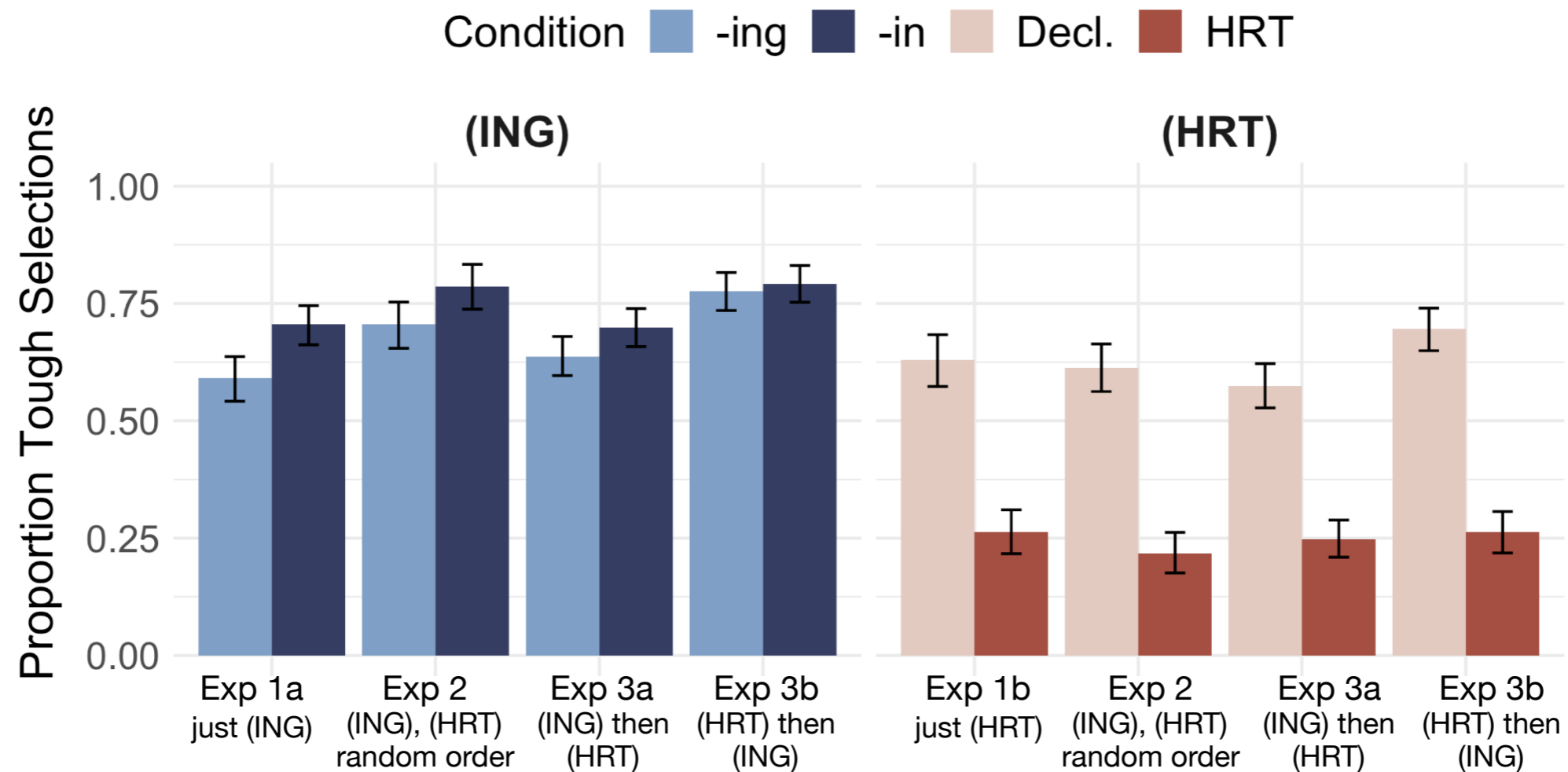
(HRT)



1

How does the strength of the indexical link between cue and persona modulate cue uptake?

3 x norming experiments (no eye tracking)



✓ Both cues give rise to effects in the expected directions

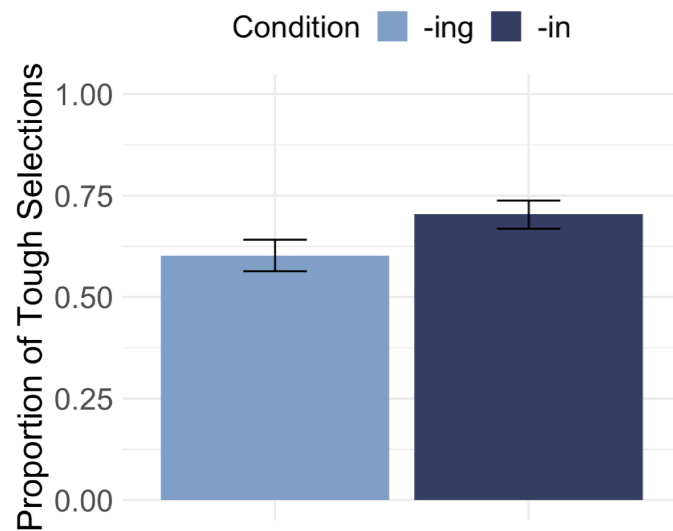
✓ Magnitude of effect greater for HRT than *-in*'

Listeners may have been using the informativity of the HRT cue to reason probabilistically about cue alternatives

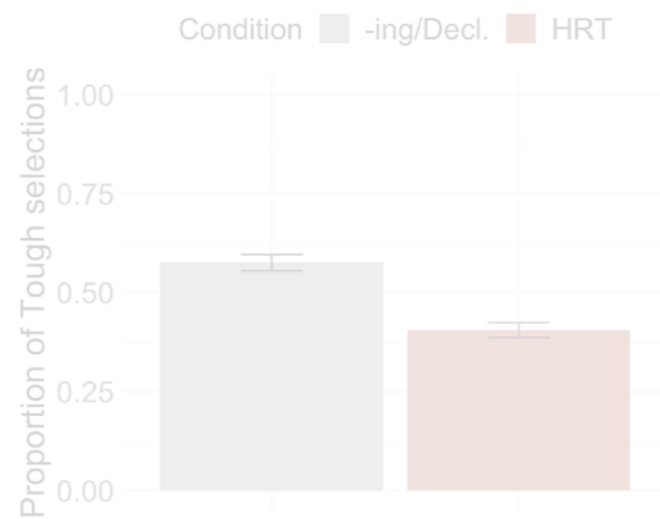
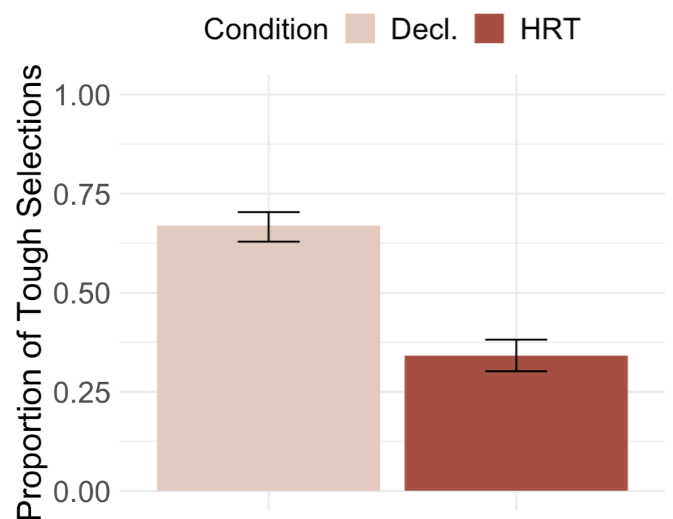
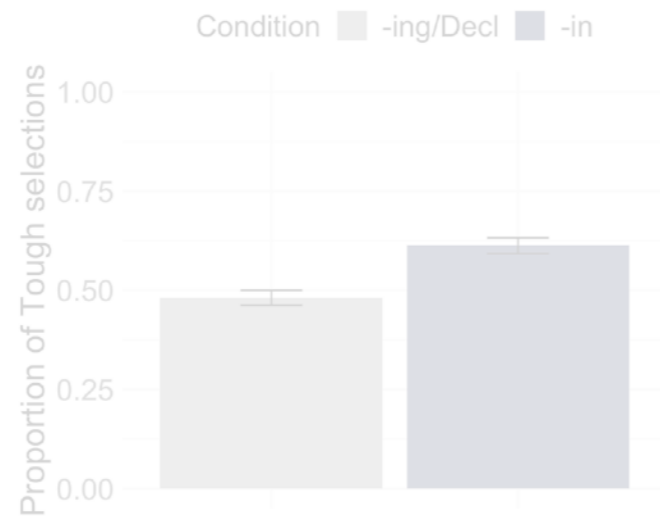
Discussion

Why was the difference in effect size between (ING) and (HRT) more muted compared to Exp. 1?

Exp 1



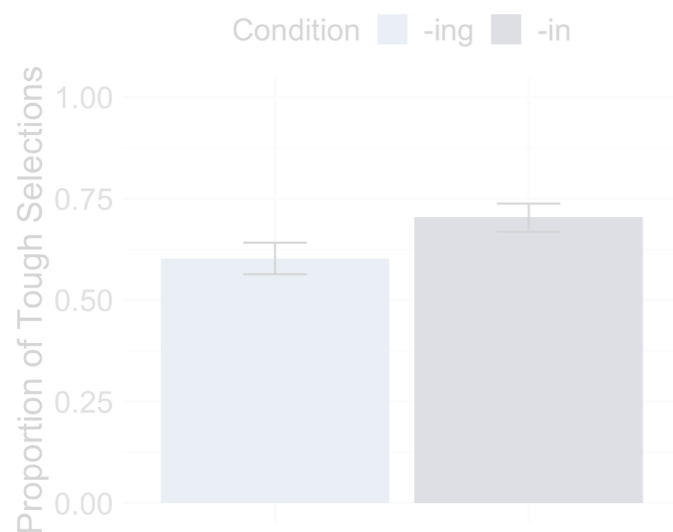
Exp 2



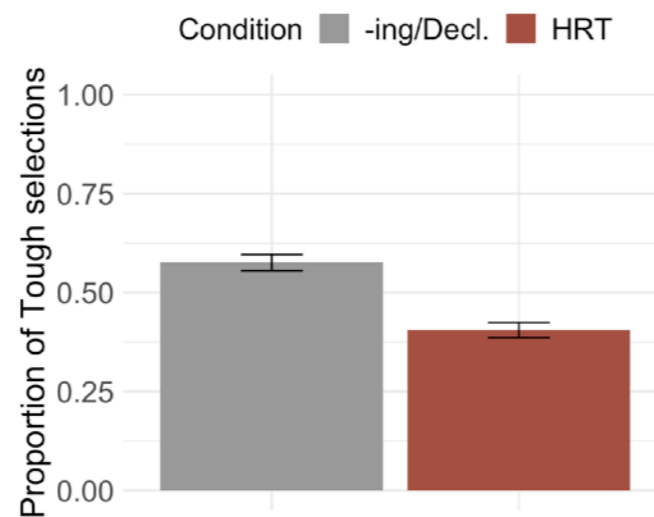
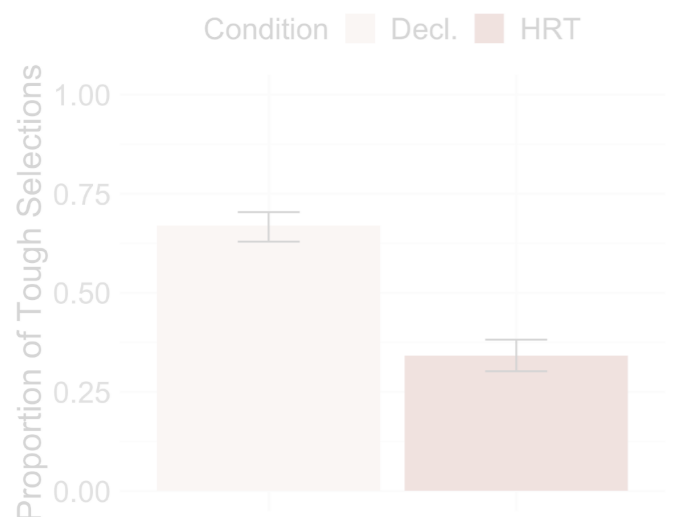
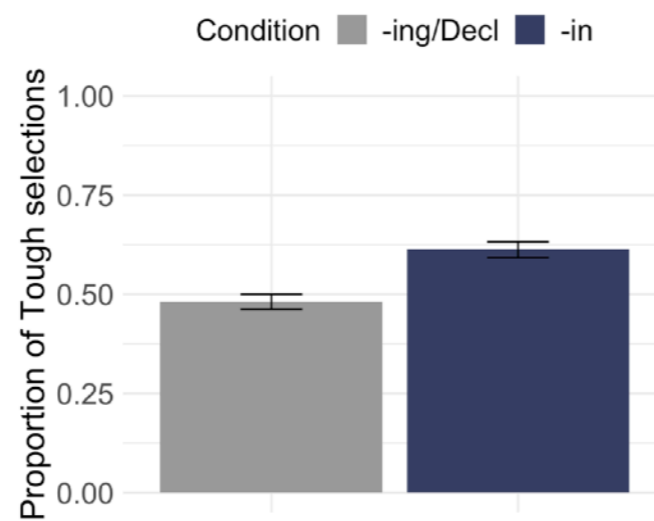
Discussion

Why was the difference in effect size between (ING) and (HRT) more muted compared to Exp. 1?

Exp 1

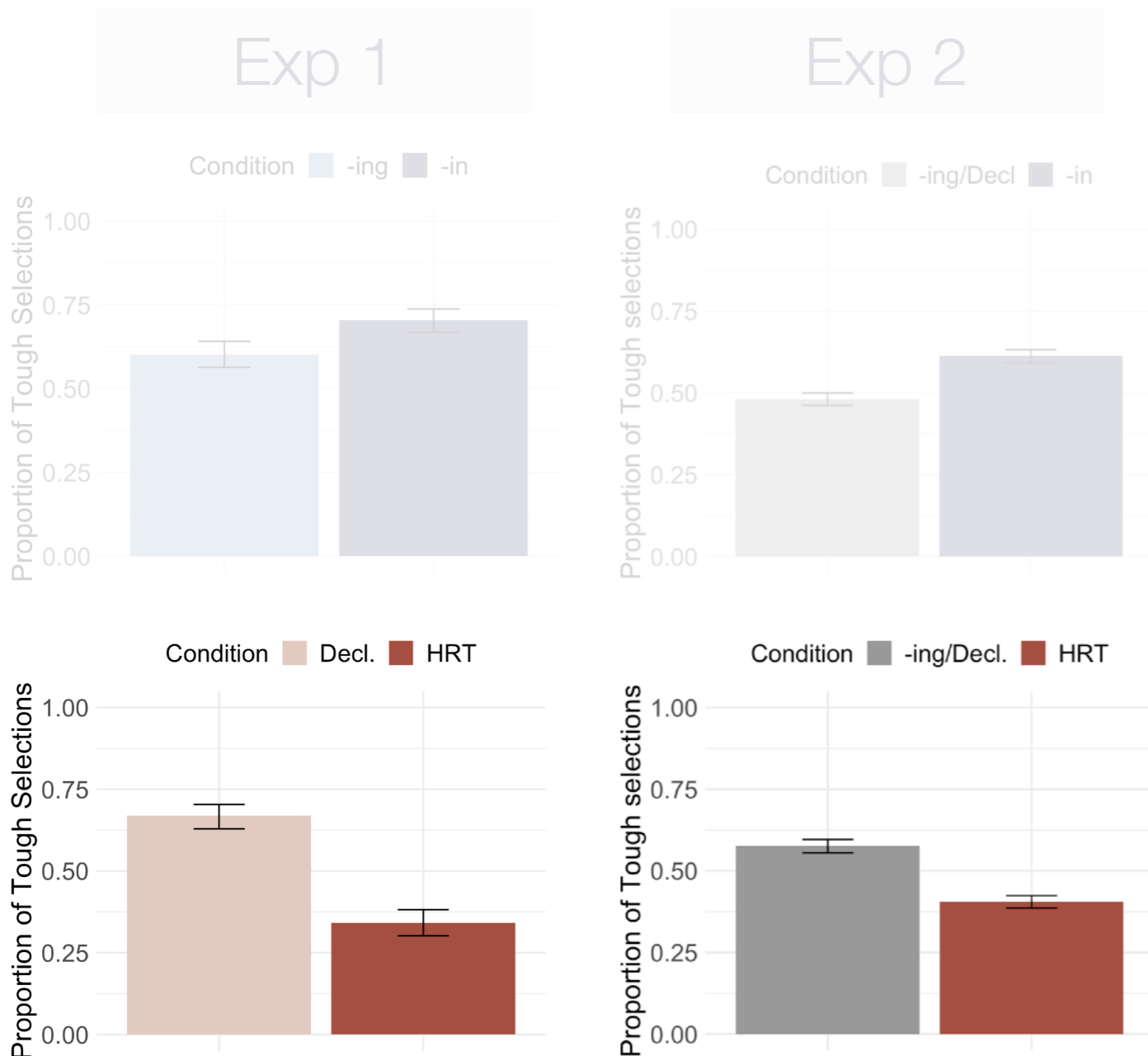


Exp 2



Discussion

Why was the difference in effect size between (ING) and (HRT) more muted compared to Exp. 1?



Less salient / less convincingly Valley Girl?

Here, *voice* could be used as a cue to speaker identity.

In Exp 1, listeners *capitalized* on the HRT cue given relative sparsity of information.