Speech Perception

Perceiving social variation:
why do people believe non-native speakers are difficult to understand?

College of Arts and Sciences
Department of Linguistics

End of our course!

- Please upload your writing assignment to the dropbox by 11:59pm tonight
  - https://www.dropbox.com/request/TY7MnSRIs0VWKylwhasG

Is the first C Same or Different?

Is the first C Same or Different?

Where is this little guy going?

Cross language (and cross dialect?) perception

- So far we have assumed native-like competence in the language being perceived

- But what happens when the listener and the talker have different language backgrounds?

- What do our theories predict?

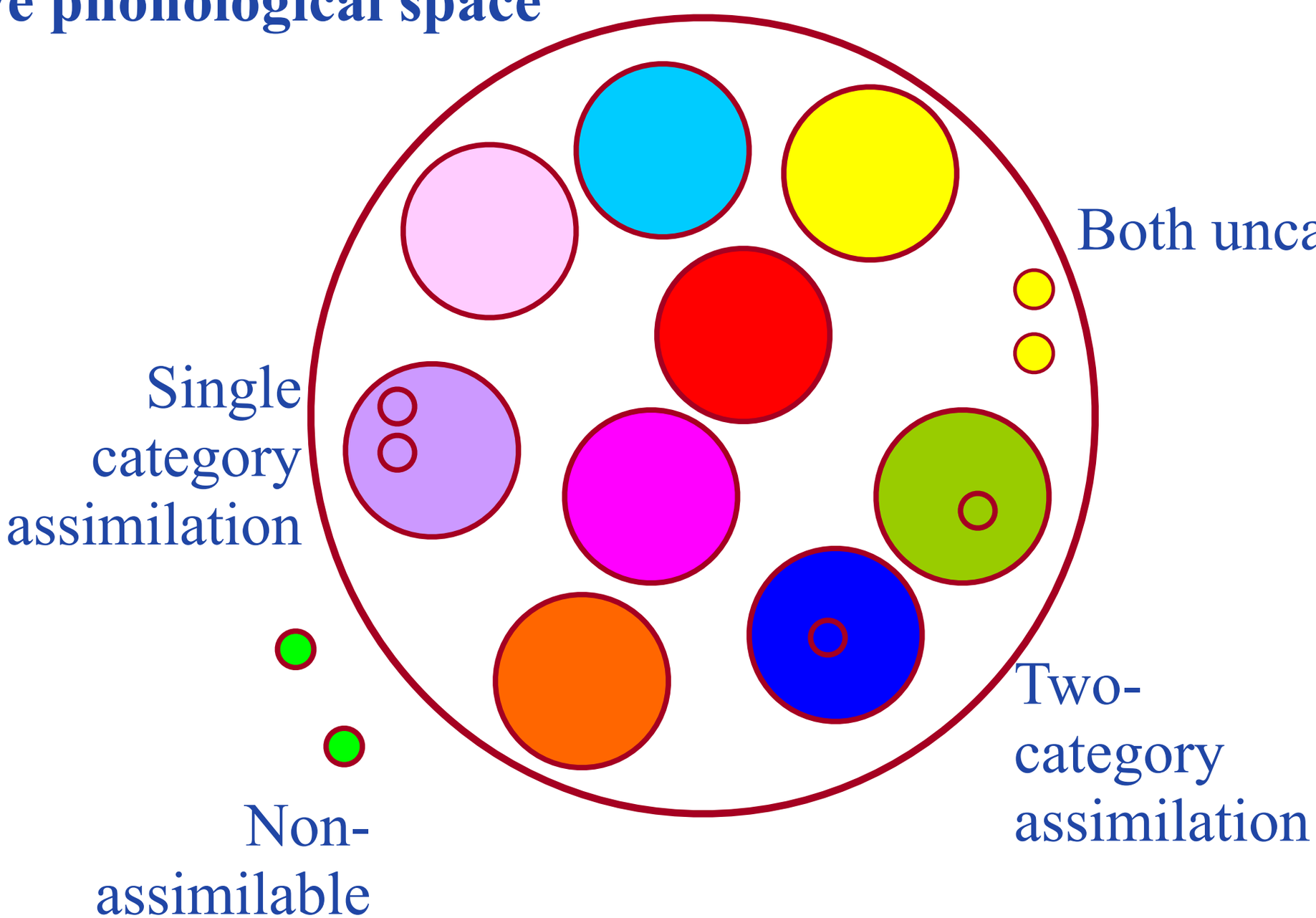
Perception across languages

- Based on listeners' native language(s)/dialect(s), can we arrive at a predictive theory for how easily a given phonetic distinction will be discriminated?
    - Perceptual Assimilation Model (Best 1995, 2009)
    - Flege's Speech Learning Model (Flege 1995, 2007)
- Fundamental assumptions: L2 discrimination depends in part on phonetic similarity (articulatory, acoustic) of invariants in the non-native sound(s) to invariants in native speech sounds.

The Perceptual Assimilation Model

- Patterns of perceptual assimilation of non-native segments by *naïve* listeners:
    - Assimilated to a native category (e.g. Norwegian /y/ to English /u/)
    - Assimilated as an uncategorizable speech sound (e.g. German /x/ for English listeners)
    - Not assimilated to speech (e.g. Zulu clicks for listeners from non-click languages)

The Perceptual Assimilation Model



Empirical support from Zulu contrast discrimination by American English listeners:

- **Non-assimilated (NA)**: Excellent discrimination of click contrasts (heard as non-speech)

- **Two category (TC)**: Excellent discrimination of voiced vs. voiceless lateral fricatives (assimilated to “shla” vs. “zhla”)

- **Category goodness difference (CG)**: Good discrimination of velar voiceless aspirated vs. ejective stops (variable assimilation to /k/)

- **Single category (SC)**: Poor discrimination of voiced bilabial plosive versus implosive (both assimilated to /b/)

The Speech Learning Model

- Like the Perceptual Assimilation Model, Flege's Speech Learning Model also posits that L1 will influence L2 perception

- Not all sounds are equally easy to understand

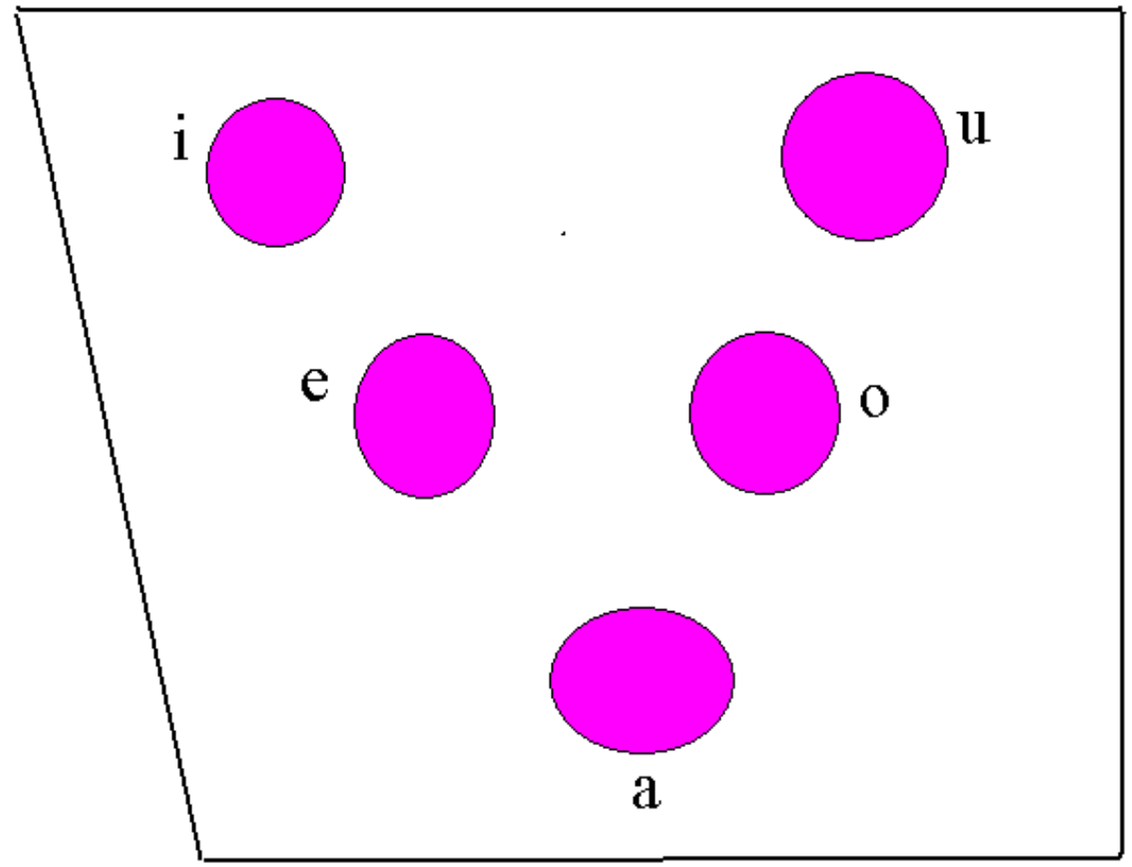
- No Critical Period: the resources an infant used to acquire their L1 are available throughout the lifespan

- One critical difference from the Perceptual Assimilation Model is that the SLM presumes (increasingly) *experienced* L2 listeners

Illustration: perceived dissimilarity

Let's imagine that there are 5 vowels in the L1, depicted here by ellipses in a 2 dimensional high-low vs. front-back vowel space

Our imaginary language is similar to real languages such as Spanish.

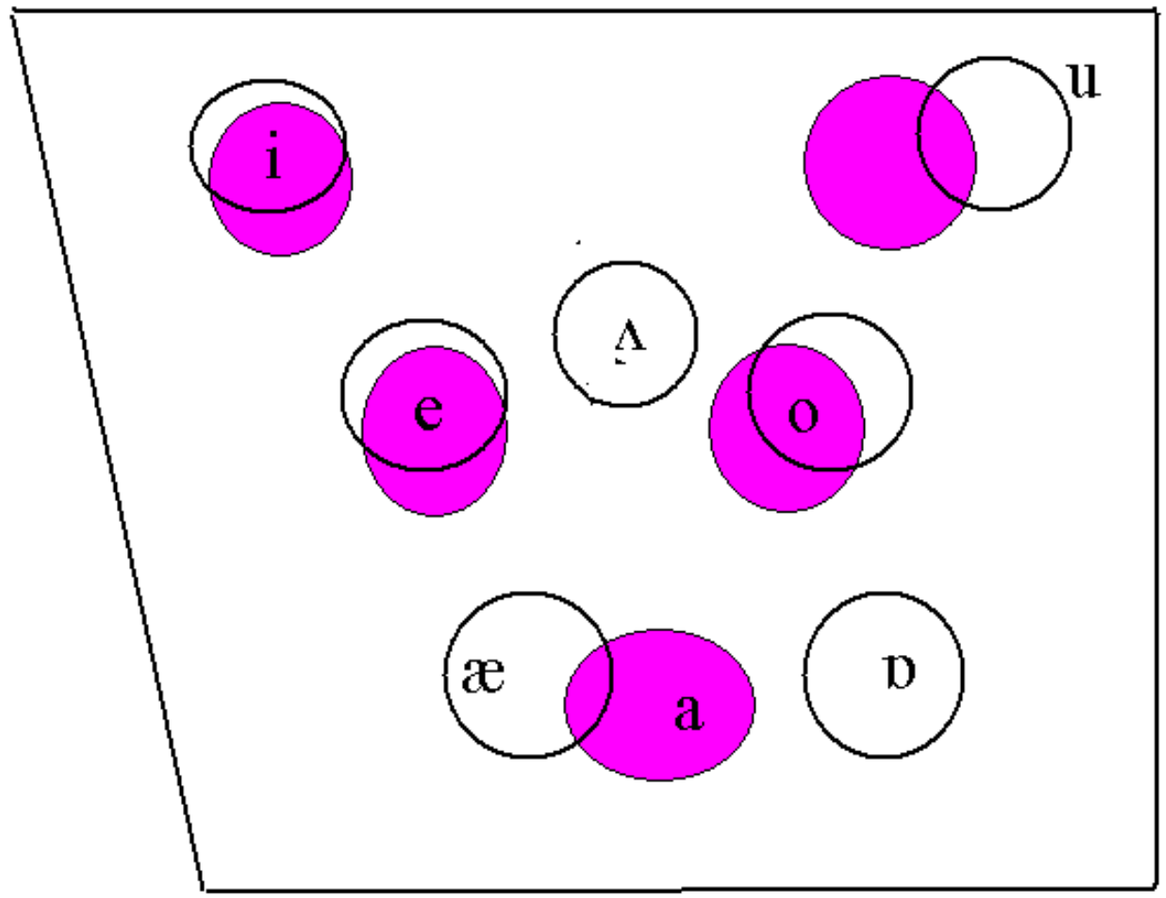


From Flege, ASA Keynote 2005

Illustration: perceived dissimilarity

Let's suppose that the L2 has 7 vowels and that perception of vowels of the L2, like those of the L1, are based entirely on center formant frequency values (no role of either duration or formant movement patterns)

Whether L2 learners will treat a vowel in the L2 as "new" will emerge over time. This determination can not be made by looking at plots of acoustic data.



but see Levy & Strange (2008)!

From Flege, ASA Keynote 2005

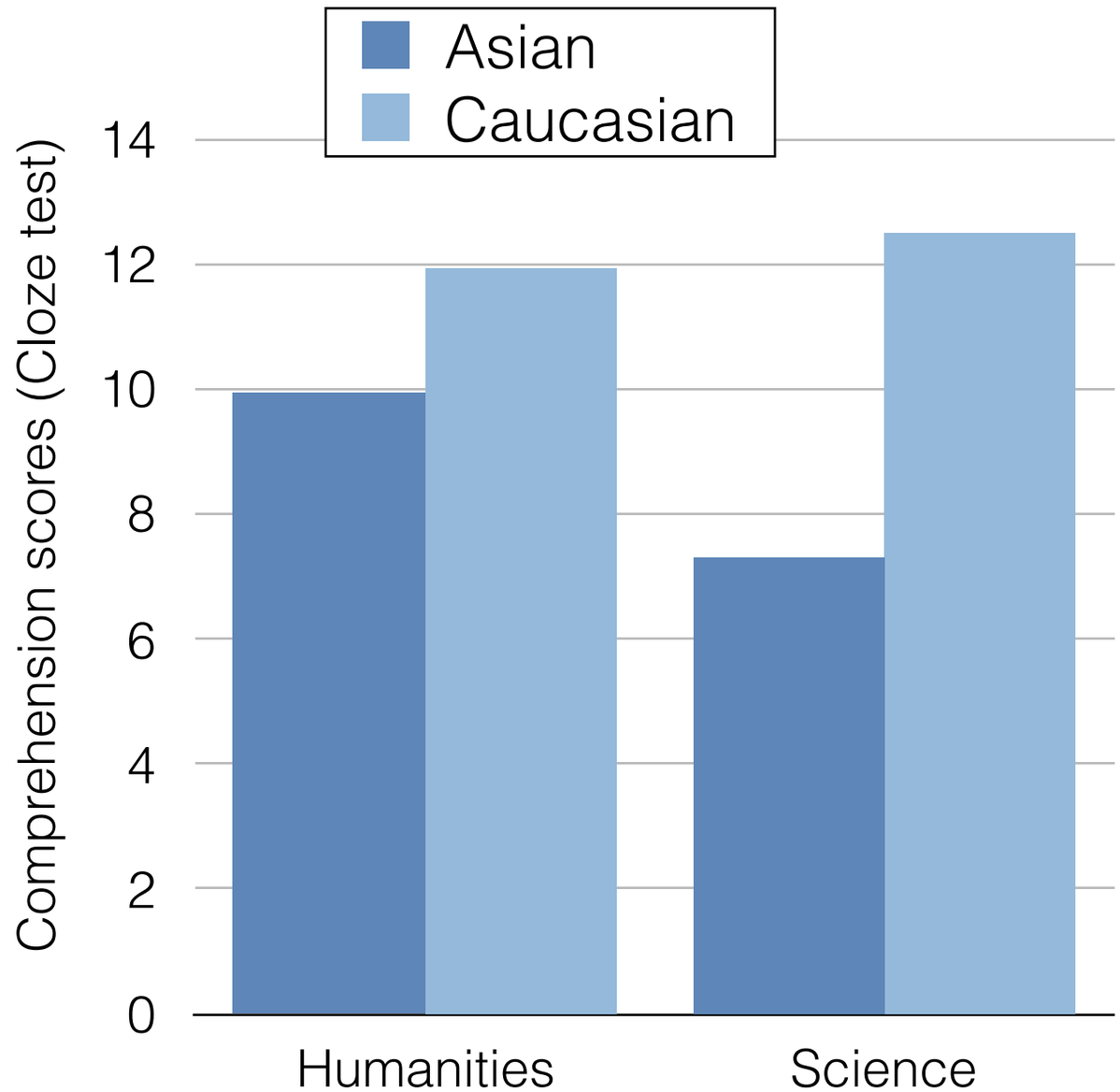
So what do exemplar models predict?

Rubin (1992)

Anti-Foreign Bias?

	***SAE Accent**
Asian Face	Incongruous
Caucasian	Congruous

Chinese Accent?



Task

- **Saw** the face of the purported speaker
- **Heard** a series of 60 Chinese-accented sentences presented in -4 dB SNR multitalker babble
- **Transcribed** as accurately as possible what they heard

A response was coded as correct if the listener typed the final ta
word:

e.g. Elephants are big animals

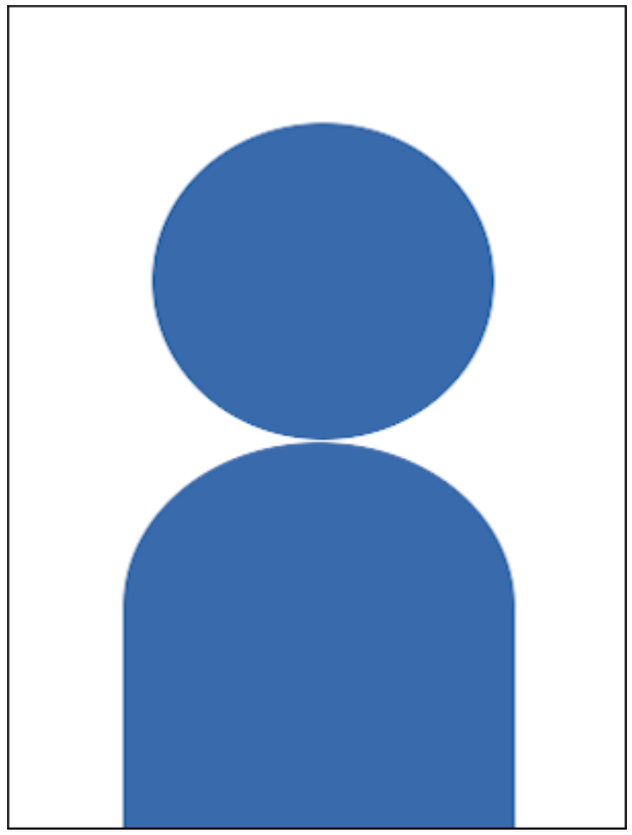
We pointed at the bird

The war plane dropped a bomb

Visual Stimuli: Face

Asian Face

congruous condition



Silhouette

control

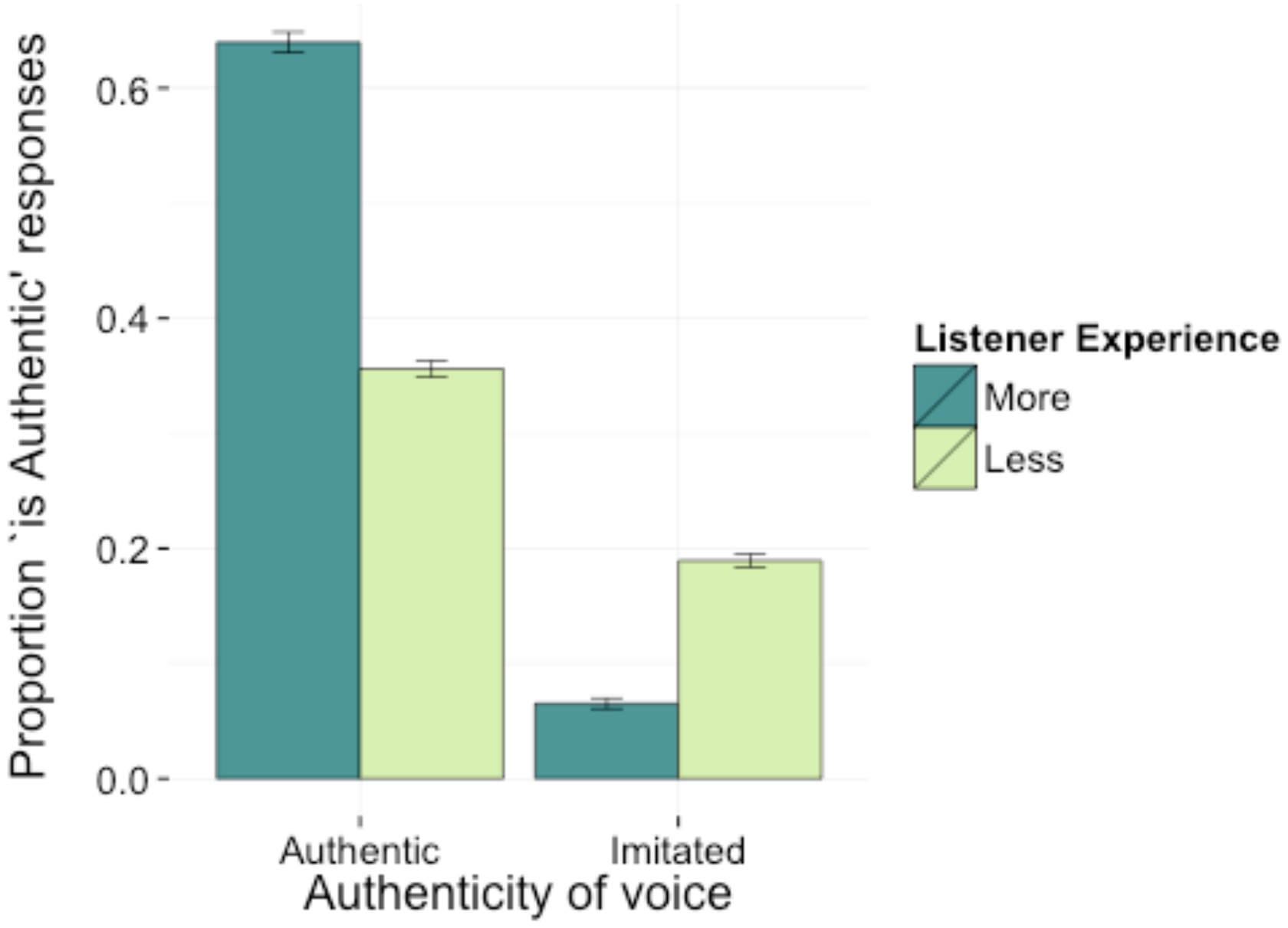
Caucasian Face

incongruous condition

Social categories depend on the listener

- 57 **Less Experienced** listeners from the University of Michigan
    - Less Experienced listeners self-reported* little to no known interaction with L1 Mandarin speakers speaking English
- 30 **More Experienced** listeners from UC Berkeley
    - More Experienced listeners self-reported* as Heritage Mandarin speakers who reported little to no fluency in Mandarin but extensive experience with L1 Mandarin speakers speaking English

Assessing Experience



Competing hypotheses

Exemplar

- **Variation is stored in memory and linked to social category.**
  (e.g. Johnson, 2006; Munson, 2010; Docherty & Foulkes, 2013).

Bias

- **Seeing a Chinese face causes Caucasian Americans to pay less attention.**
  (e.g. Rubin, 1992; Kang & Rubin, 2009; Lippi-Green, 2011)

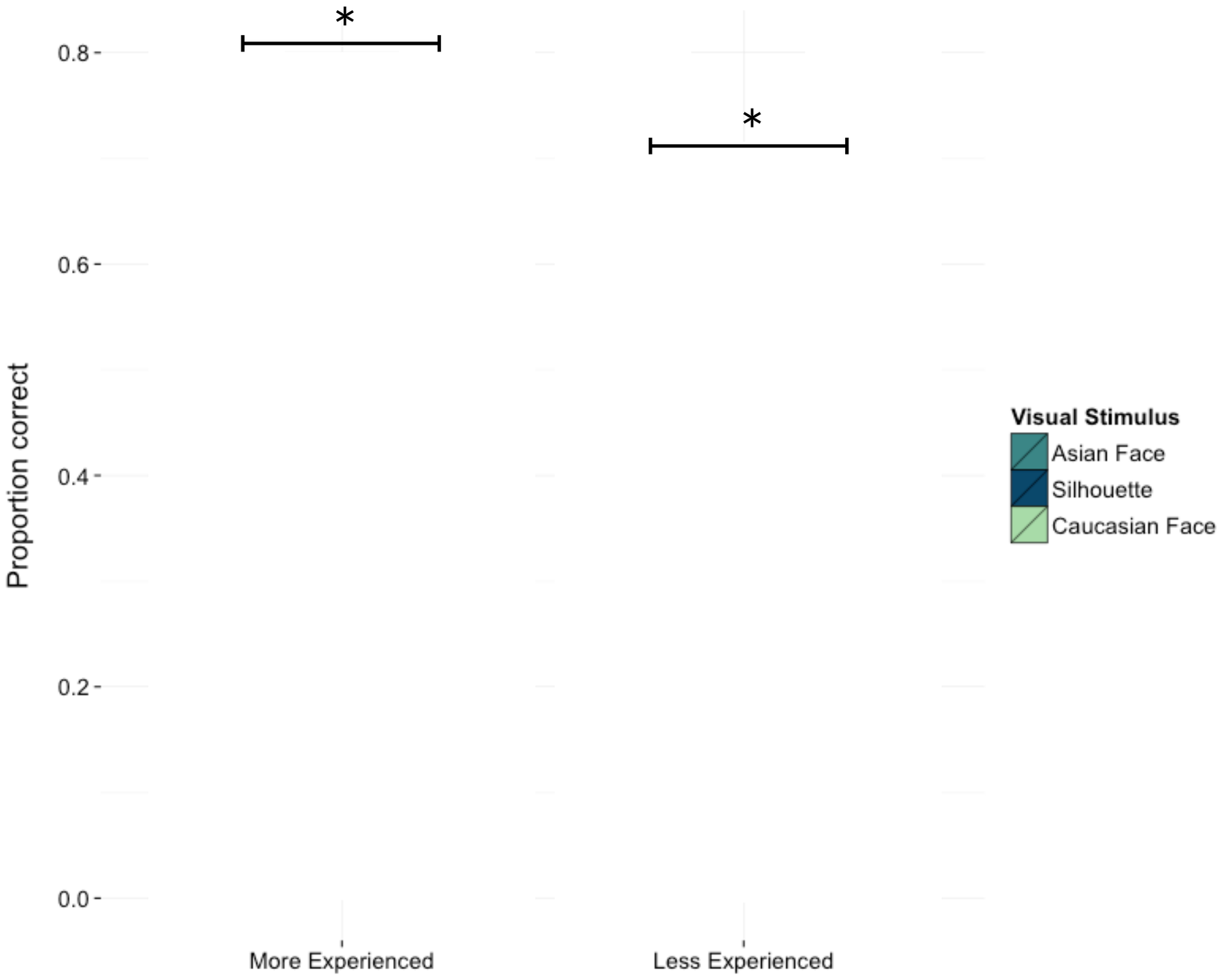
Competing predictions

Exemplar

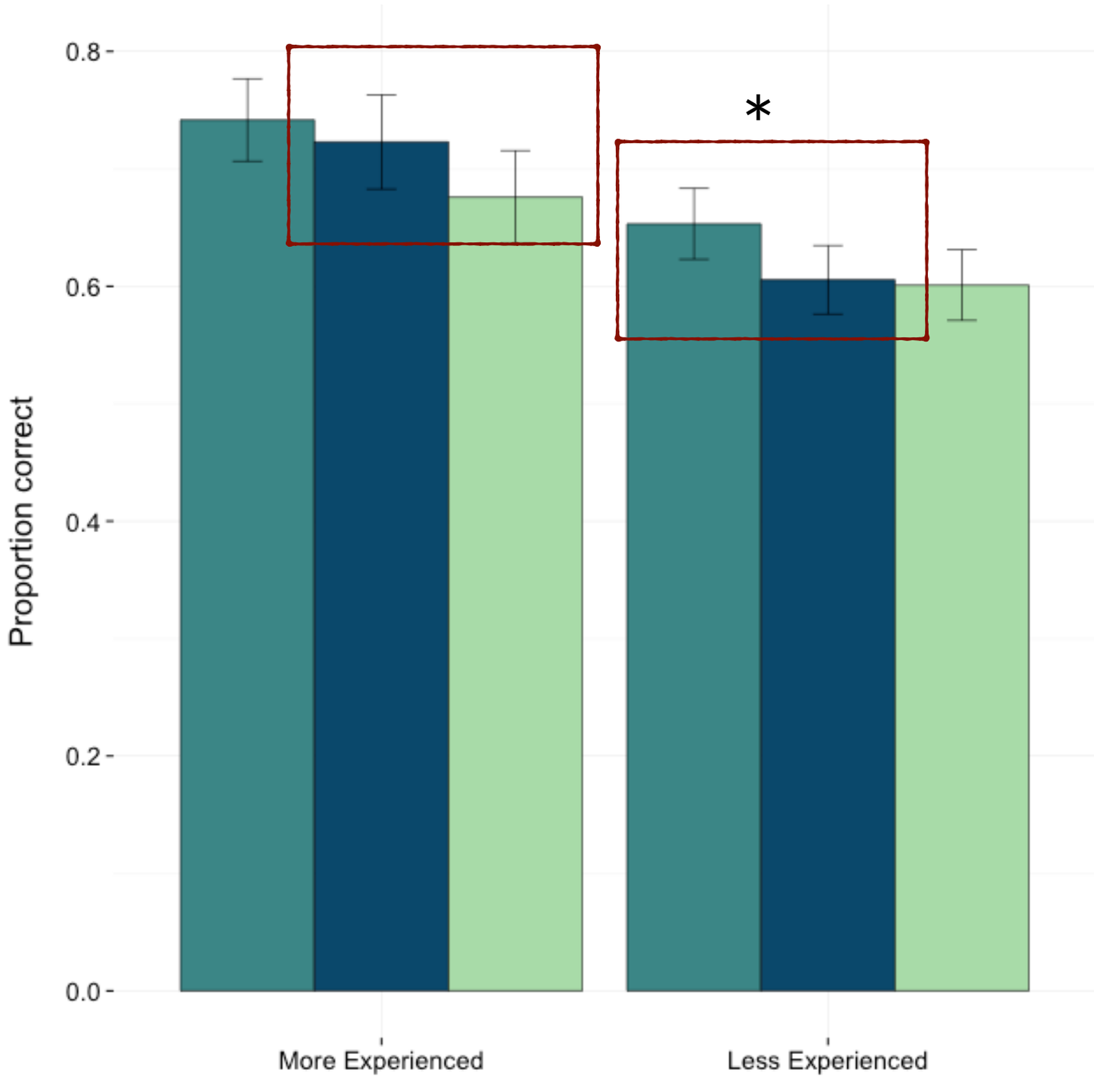
- **Variation is stored in memory and linked to social category.**
  (e.g. Johnson, 2006; Munson, 2010; Docherty & Foulkes, 2013).
  - Higher transcription accuracy when shown an Asian face
  - Lower transcription accuracy when shown a Caucasian face
  - Greater overall benefit for More Experienced listeners

Bias

- **Seeing a Chinese face causes Caucasian students to pay less attention.**
  (e.g. Rubin, 1992; Kang & Rubin, 2009; Lippi-Green, 2011)
  - Higher transcription accuracy when shown a Caucasian face
  - Lower transcription accuracy when shown an Asian face


*
*
0.8
0.6
0.4
0.2
0.0
Proportion correct
More Experienced
Less Experienced
Visual Stimulus
Asian Face
Silhouette
Caucasian Face

- The inclusion of a control condition allows us to compare baseline performance for the two groups
- **Inhibition**: More experience, Caucasian face impairs transcription
- **Facilitation**: Less experience, Asian face improves transcription



Discussion: Bias?

- Listeners were *most* accurate when shown an Asian Face
- The strong prediction of the Bias Hypothesis of Rubin is not supported by these results
- The problem with the bias hypothesis is that studies like Rubin (1992) forget that *SAE voices and Caucasian faces carry social information too!!

	*SAE Accent	**Chinese**
Asian Face	Incongruous	Congruous
Caucasian Face	Congruous	Incongruous

Discussion: Exemplars?

- Listeners were most accurate when shown an Asian Face
- More Experienced listeners *were* more accurate, overall, than Less Experienced listeners
- However, there was no interaction of Face and Experience, listeners saw the same degree of apparent improvement even if they basically knew nothing about what a Chinese accent sounds like!
- This suggests the need for a more nuanced, complicated view of what counts as experience…

Social categories depend on the listener

Preston (1996) offers a model of linguistic awareness that is useful when thinking about the role of the listener in social category awareness and linkages to phonetic detail

- **Detail** 'Asian' versus 'Chinese' versus 'Mandarin' vs 'Beijing'
- **Accuracy** Ability to distinguish Chinese from Japanese, Authentic from imitated, etc.
- **Availability** Ability to discuss particular features of a social category or specific linguistic features associated with that category
- **Control** Ability to produce (either in imitation or identity performance) cues to social category

“Perception” seems to happen at multiple levels

- In this class we have been concerned with an approach to “speech perception” that inherits from a phonetic and, ultimately, psychoacoustic tradition

- You will also see literature from sociolinguistic and anthropological traditions that mean something quite different by the words “speech perception” and it can be difficult to reconcile findings across these traditions.

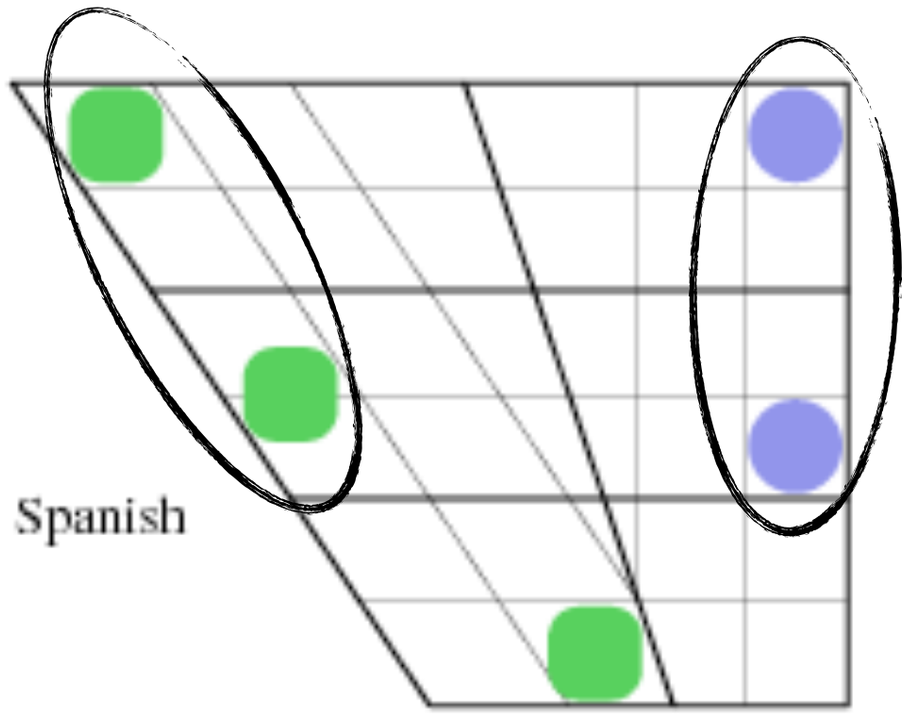
McGowan and Babel, in press

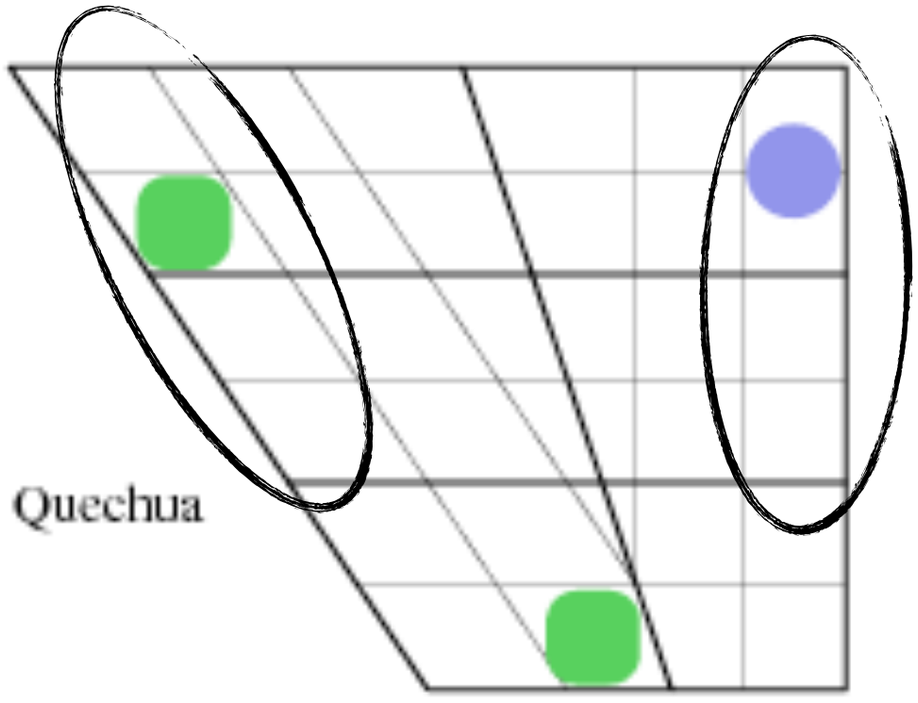
0 200 400 600 800 1000 Kilometers
0 200 400 600 Miles
60°
10°
20°
Rio Branco
Pucallpa
Tingo María
Huallaga Valley
Húanuco
Cerro de Pasco
Montaña (Eastern) Slopes
Lima
Callao
Huancayo
Cord. Oriental
Ayacucho
Cuzco
Pisco
Ica
Cordillera Occidental (W
Yungas
Highlands
Puno
Lake Titicaca
El Alto
La Paz
Arequipa
Ilo
Tacna
Arica
Oruro
Altiplano
Lake Poopó
Sucre
Potosí
Tarija
Cochabamba
Chaparé
Beni R.
BOLIVIA
The Oriente
BRAZIL
Cuiabá
Pantanal Wetland
Santa Cruz
Carreria
Campo Grande
PARAGUAY
Paraguay R.
Ciudad del Este
Asunción
ARGENTINA
Salta
CHILE
Antofagasta
Tropic of Capricorn
Current
CIFIC
CEAN

Social Context

- Quechua, the non-prestige code, remains closely associated with rurality, poverty, indigeneity, etc., but also with intimacy, locality, an idealized past, and traditional activities such as weaving and cooking. This can be seen in the higher incidence of Quechua contact features in some types of contexts (informal, conversational, joking) than in others (formal, meeting-style, serious).
- Speakers who identify themselves as Spanish-speakers may actually have a wide range of Quechua abilities, from true native fluency to knowledge of a few words, a bit of Mock Quechua, or punchlines of jokes.

Quechua vs. Spanish vowel systems

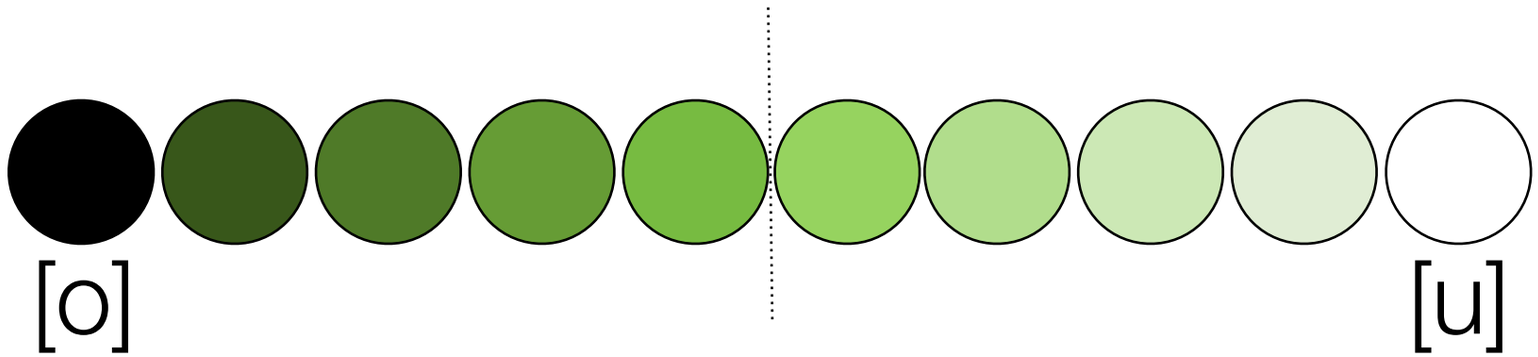




The Stimuli

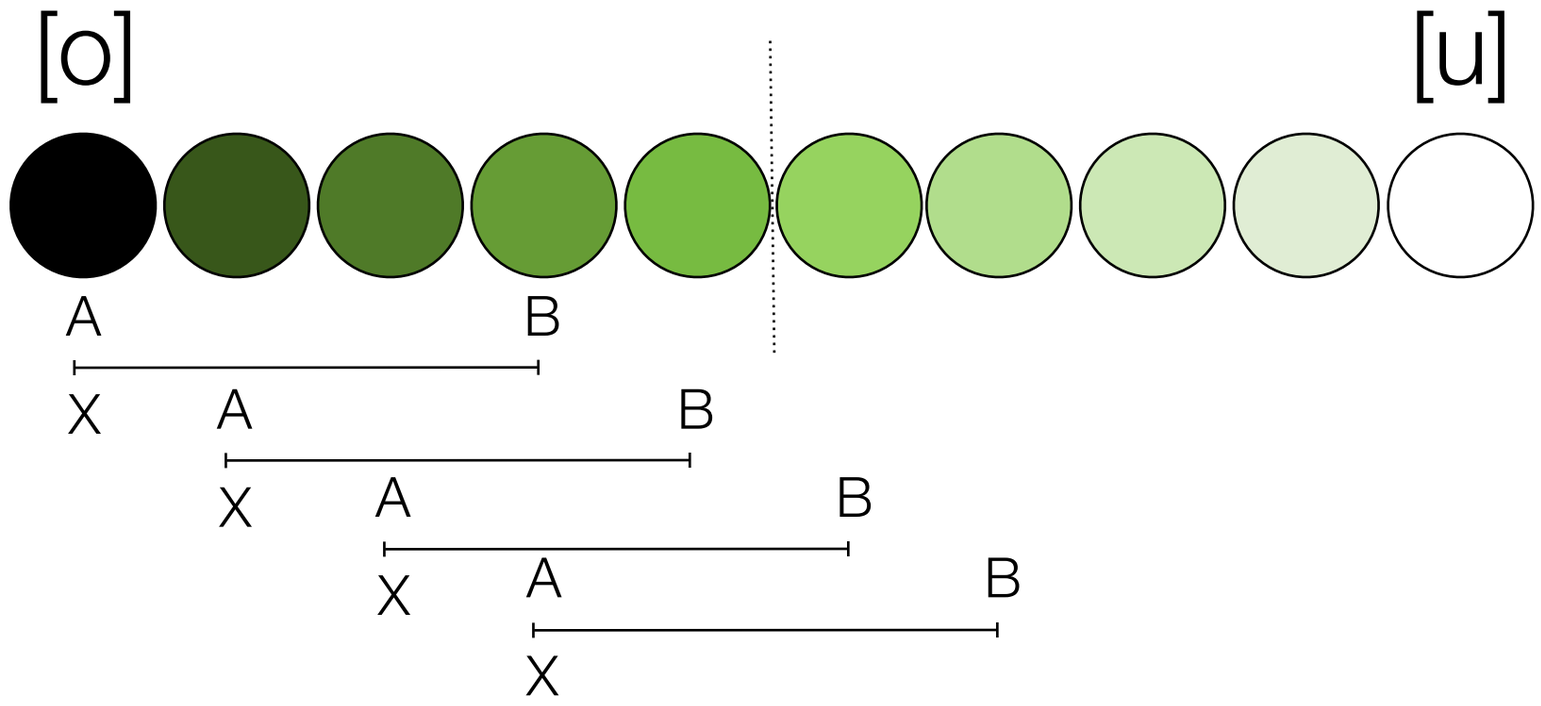
- Synthesized /e/-/i/ and /o/-/u/ continua using Praat
    - peca ~ pica
    - pesa ~ pisa
    - soda ~ suda
    - moda ~ muda

Synthesis Visualized

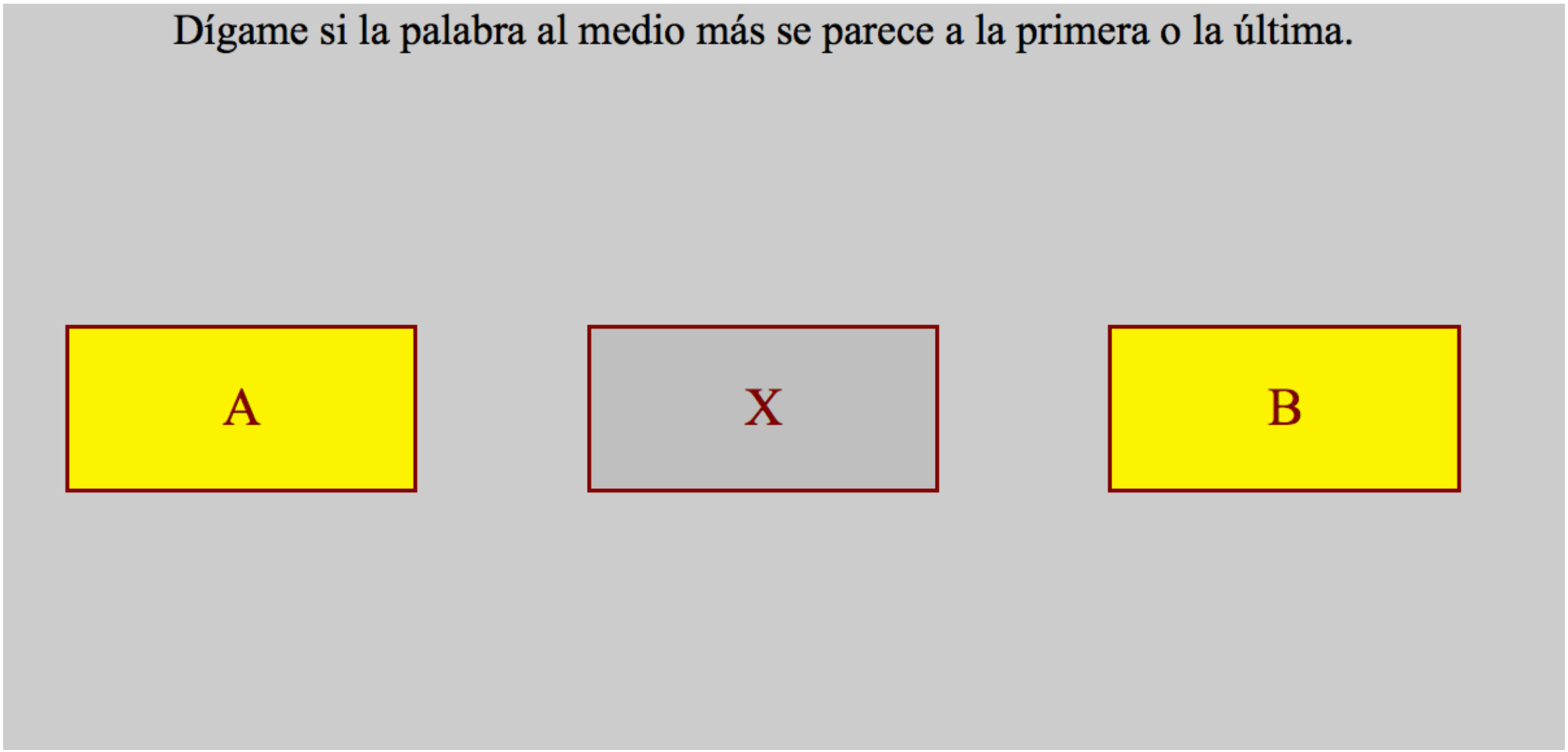


- Two separate continua were synthesized for each word pair to ensure that 'same' comparisons in the AXB task were not between two identical audio files but between two synthesized

The Task: AXB



‘Tell me if the middle word sounds

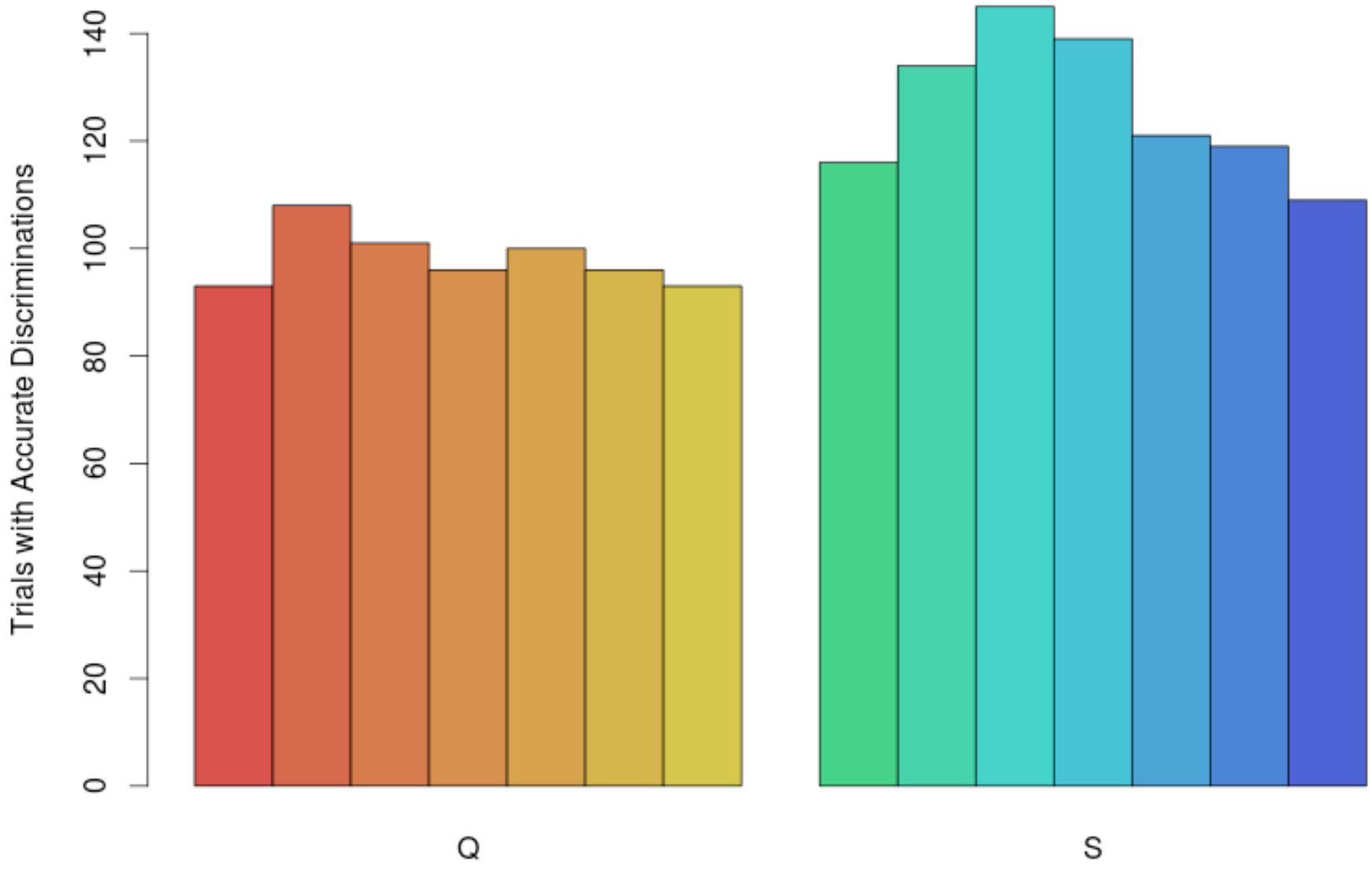


Guise Presentations

Initial
QUECHUA soda/suda pisa/pesa
SPANISH soda/suda pisa/pesa

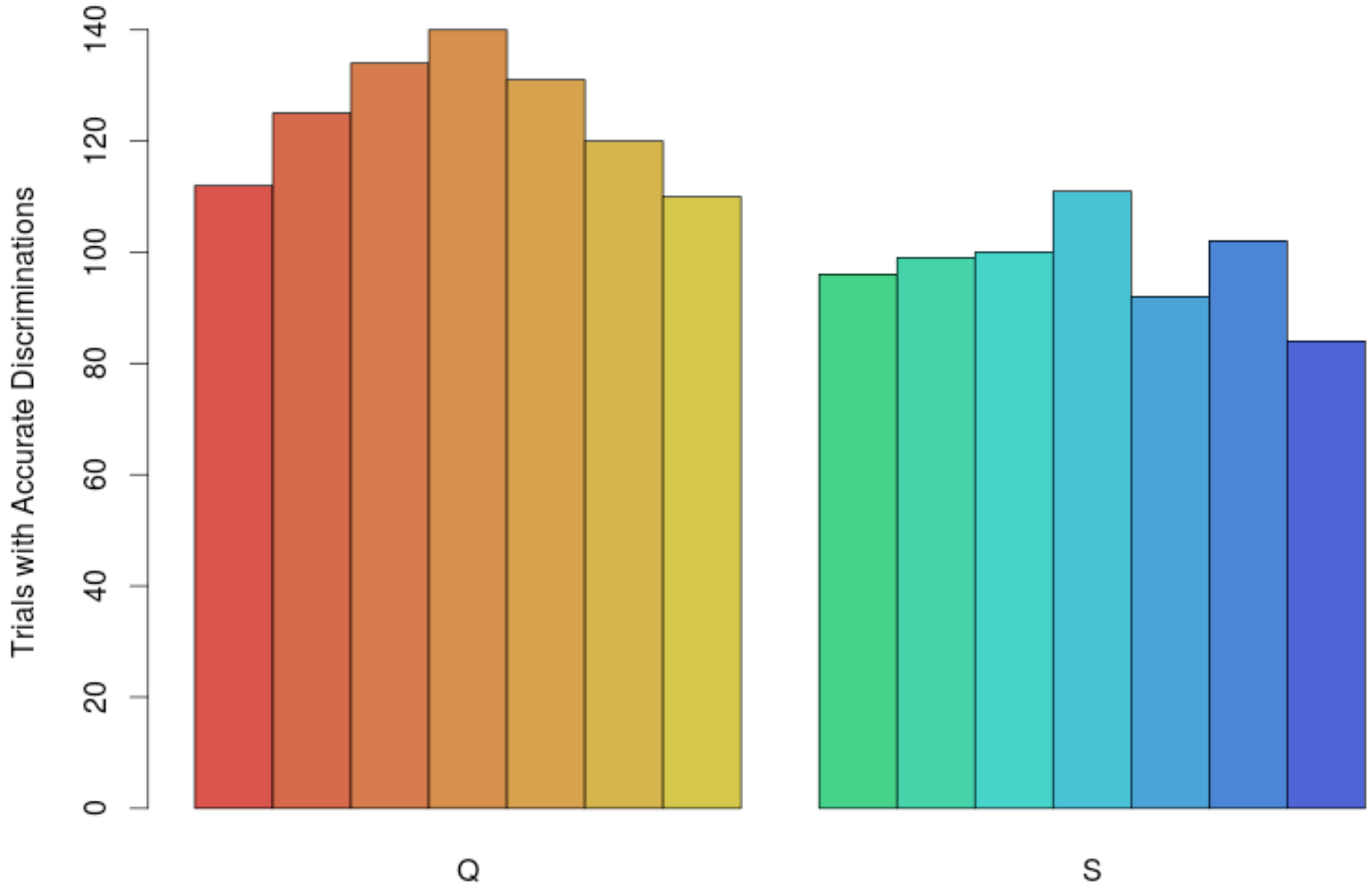
Predictions

- Manipulating expectation of speaker identity will alter listener category boundaries for /e/-/i/ and /o/-/u/.
    - Category boundaries between the vowel pairs should be relatively clear when participants believe the speaker is Spanish-dominant.
- Creating an expectation of Quechua-dominant speech will broaden listener vowel categories resulting in more gradient/less categorical discrimination
    - More gradient and more variable boundaries should emerge when participants believe the speaker is Quechua-dominant due to perceived variation in vowel production among Quechua-dominant speakers
- High Degree of awareness. Listeners should really know about this vowel difference.



Initial Presentation

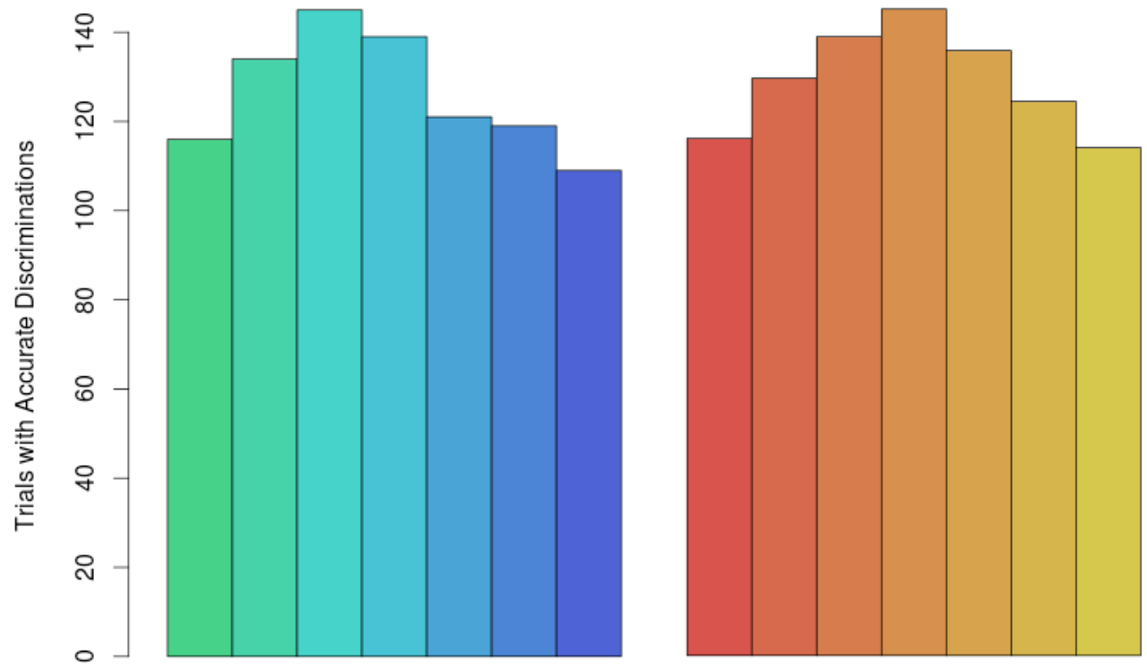
Discrimination by Speaker Guise



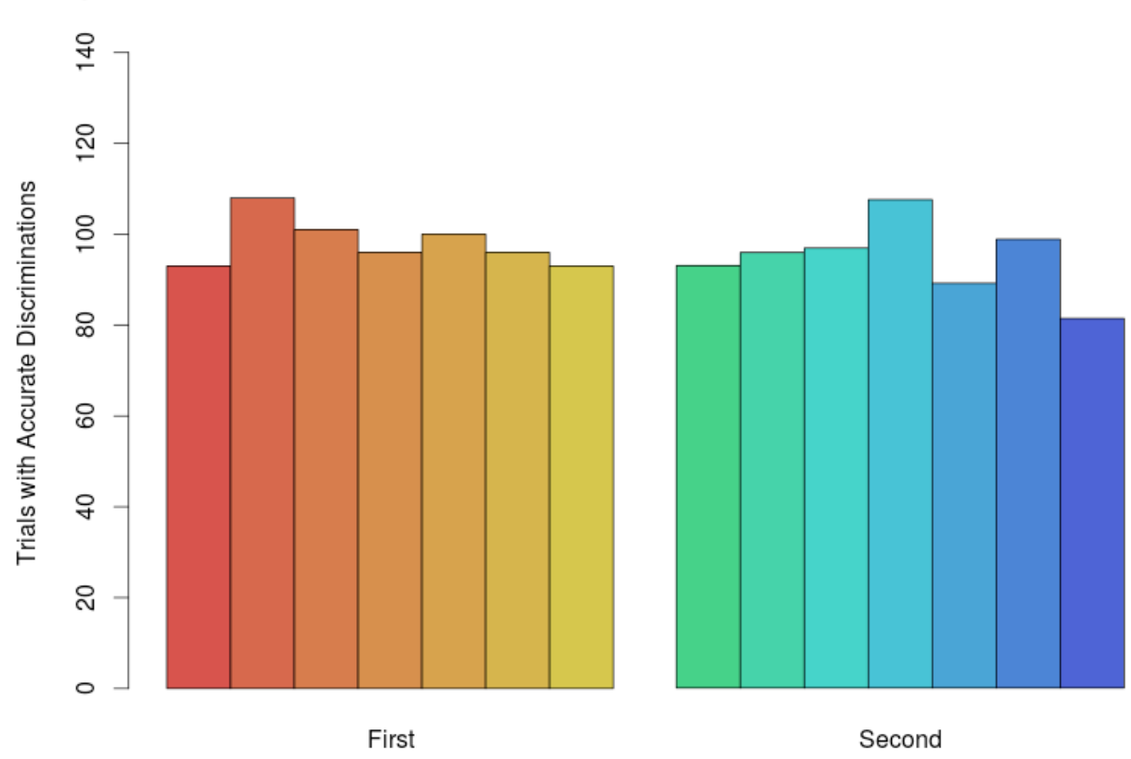
Second Presentation

Discrimination by Speaker Guise

Spanish Guise First:



Quechua Guise First:



Interviews: Consultant 1

- Quechua speaker: Has a “primary or secondary” education
- Spanish speaker: Has a “mid-level” education; has probably finished high school.

Interviews: Consultant 2

- Quechua guise: “I don’t know…I see her as a person who has studied through fourth grade or so.”
- Spanish guise: “I think [she has studied] because she speaks clearly.”

Interviews: Consultant 3

- “[The Spanish guise] let you understand what she said better. And [the Quechua guise] very little.”

Interviews: Consultant 4

- “There was a little [difference]…the [Quechua guise] I had a hard time understanding, but maybe it’s because I was getting used to it…the [Spanish guise] was clearer.”

Interviews: Consultant 5

- “The [Quechua guise] was more accentuated…As if she were trying to pronounce the /u/ and the /o/ at the same time.”

Interviews: Consultant 6

- “The tone was clearer [in the Spanish guise]. Different. She made more of a difference [between the words]. [The Quechua guise] was more muffled. [The Spanish guise] was clearer. Her voice carried better.”

Interviews: Consultant 7

- Spanish guise: “The words she says are very clear.”
- Quechua guise: “Seems like she’s from the city.”

Interviews: Consultant 8

- Quechua guise: “I don’t think [she has studied]. She’s probably studied some but only the a certain point. She hasn’t finished everything.” … “I think she has studied, to be able to determine all these things.”
- Spanish guise: “I think [she has studied]. She makes the difference [between the word pairs].”

Interviews: Consultant 9

- Quechua guise: “[I think she has studied] because she has a difference in her speech. Someone who hasn’t studied talks another way.”
- Spanish guise: “She must be from Santa Cruz. Someone from Cochabamba speaks differently. Someone from Santa Cruz speaks a little clearer…in Santa Cruz they speak clearer, because it’s legitimate Spanish.”
- About the task: “Peca…is altered vocabulary. Someone from Potosí or Cochabamba, sometimes they mess up and say “peca” instead of “pica.”

Discussion

- Qualitative measures of sociolinguistic awareness obtained through interviews and in conversation are different from experimental measures of sociolinguistic awareness obtained through a perception task.
- Claims about processing of fine phonetic detail based on high level or post hoc tasks are unlikely to reflect the processing level accurately.

- What people are able to *perceive* (in our sense in this class) does not necessarily match what they believe or report they hear.

The main question of speech perception:

How do listeners interpret the acoustic –or, rather, auditory– signal as linguistic forms?