



# Measures of acoustic and perceptual contrastiveness and nativelikeness in assessing segmental pronunciation development

Joan C. Mora  
Universitat de Barcelona



# Our L2 speech team

C. Aliaga-Garcia



M. Adrian



N. Fullana



J. Frank



K. Fraser



N. Gavalda



J.C. Mora



I. Mora-Plaza



M. Ortega



G. Sosa



# Aim

Discuss methodological issues in assessing pronunciation accuracy (focusing mainly on L2 vowel contrasts):

- between testing times after short pronunciation training interventions (HVPT, explicit/implicit phonetic instruction)
- between task performance conditions (task complexity)
- speech elicitation conditions (controlled vs spontaneous)

- contrastiveness vs. nativelikeness
- acoustic vs. perceptual measures
- controlled vs. spontaneous speech
- over-time vs. across conditions

# Assessing L2 pronunciation accuracy

## Pronunciation improvement over time

### Learning contexts & populations

- Immigrant populations in L2 environment
- Instructed SLA in L2 environment
- Short-term immersion (study abroad / intensive courses)
- Classroom language instruction in L1 environment

### L2 pronunciation instruction / training interventions

- Explicit pronunciation instruction, communicative form-focused pronunciation instruction, task-based pronunciation interventions
- Lab-based phonetic training (HVPT).



pronunciation accuracy  
Improvement  
(how much?)

## Pronunciation accuracy differences across conditions

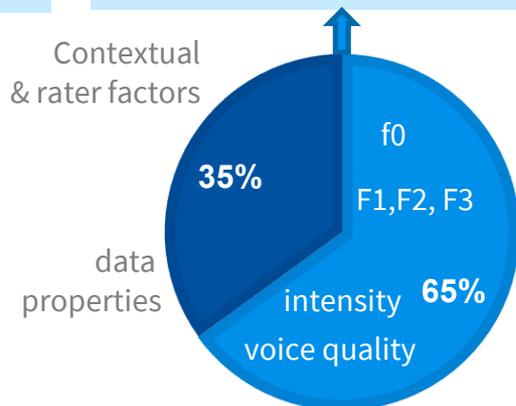
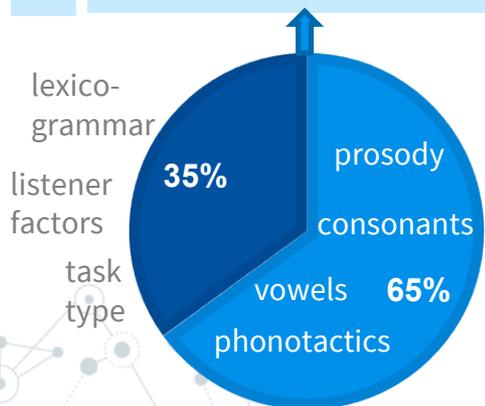
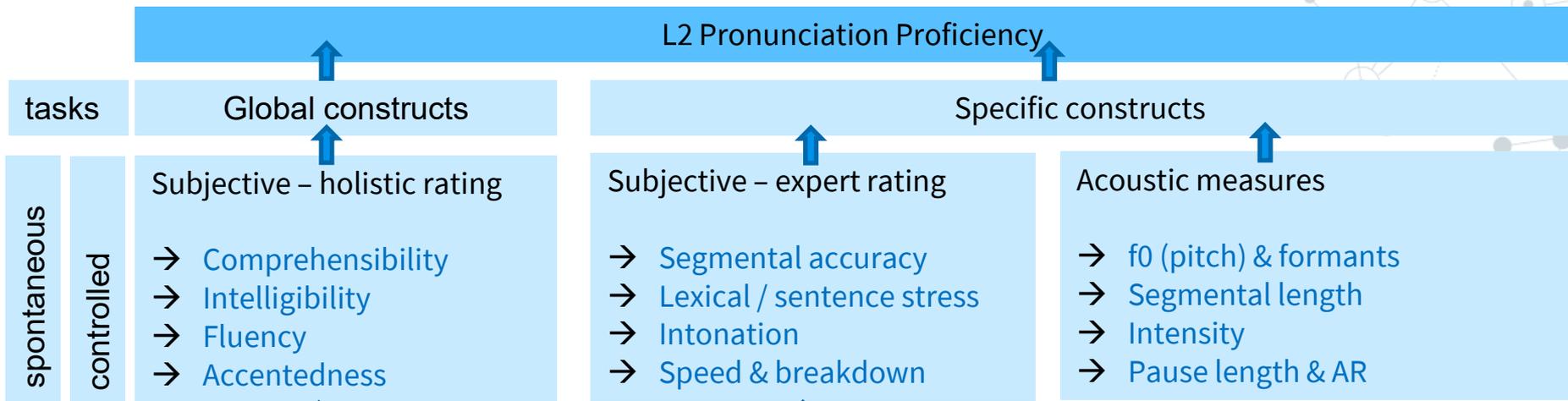
### Speaking task features

- Task features (simple vs. complex / easy vs. difficult / ...)
- Speech features: controlled (read) vs. spontaneous



pronunciation accuracy  
differences  
(how large?)

# Assessing L2 pronunciation



**Factors explaining effectiveness of L2 pronunciation instruction:**

- Scoring method (perceptual judgements vs. acoustic measures)
- global / specific focus
- test type: spontaneous / controlled

(77 studies 1982-2017)

## Assessing L2 pronunciation accuracy

- L2 pronunciation instruction / training interventions
- Task features / speech elicitation

Instruction / tasks focus on one / few targets (e.g. L2 vowel contrast such as French /u/-/y/, Spanish /d/-/r/, English /i/-/ɪ/, /æ/-/ʌ/, /æ/-/ɛ/, /r/-/l/)

- Pronunciation accuracy improvement : **small**
- Pronunciation accuracy differences : **small**

**Capturing small changes in pronunciation accuracy is informative:**

- Are pronunciation instruction / training interventions effective?
- Do pronunciation accuracy measures vary as a function of speaking modes and task types?

→ What kind of pronunciation accuracy measures may be more sensitive in capturing changes in L2 pronunciation accuracy?

# Pronunciation accuracy measures for L2 vowels

## Judgments by native speakers

- Accentedness ratings
- Comprehensibility ratings
- Identification + goodness ratings
- Discrimination + rated dissimilarity

- What do judges pay attention to?
- Individual sounds / contrasts too short?
- Overall accent vs. accuracy of specific sounds?
- Lexical bias in words?
- Little sensitivity to changes over time?

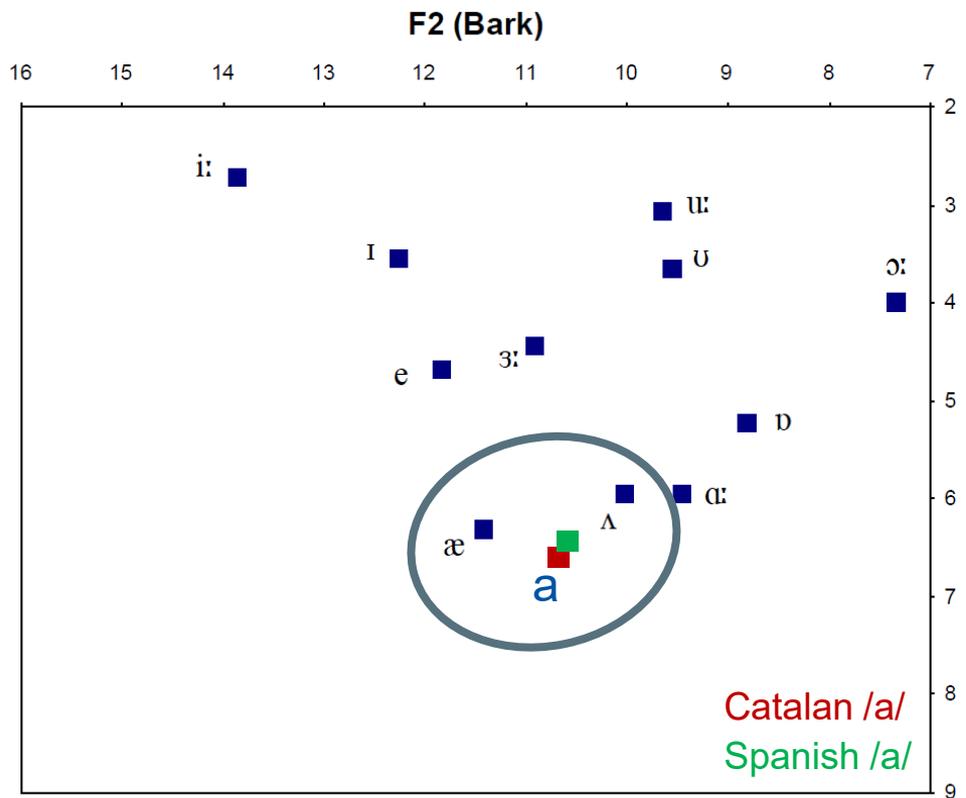
## Acoustic measurements

- Frequency & duration
- Formant transitions
- Distances btw L and NS Vs
- Distances btw Vs within Ls

- Ecological validity?
- Slow, time-consuming
- Which measure/s (formants, duration)?
- Normalization procedures
- Baseline? Native speakers (NS)?

- What kind of acoustic distances?
- What kind of perceptual distances?
- contrastiveness vs. nativelikeness: how do we relate them?
- acoustic distances vs. perceptual distances: how do we relate them?

# Contrastiveness vs. nativelikeness in L2 vowel contrasts



*/æ/ vs. /ʌ/ (cap vs. cup)*  
 Spanish/Catalan learners of English  
 → Map it onto Sp/Cat /a/  
 → L2 confusable vowels

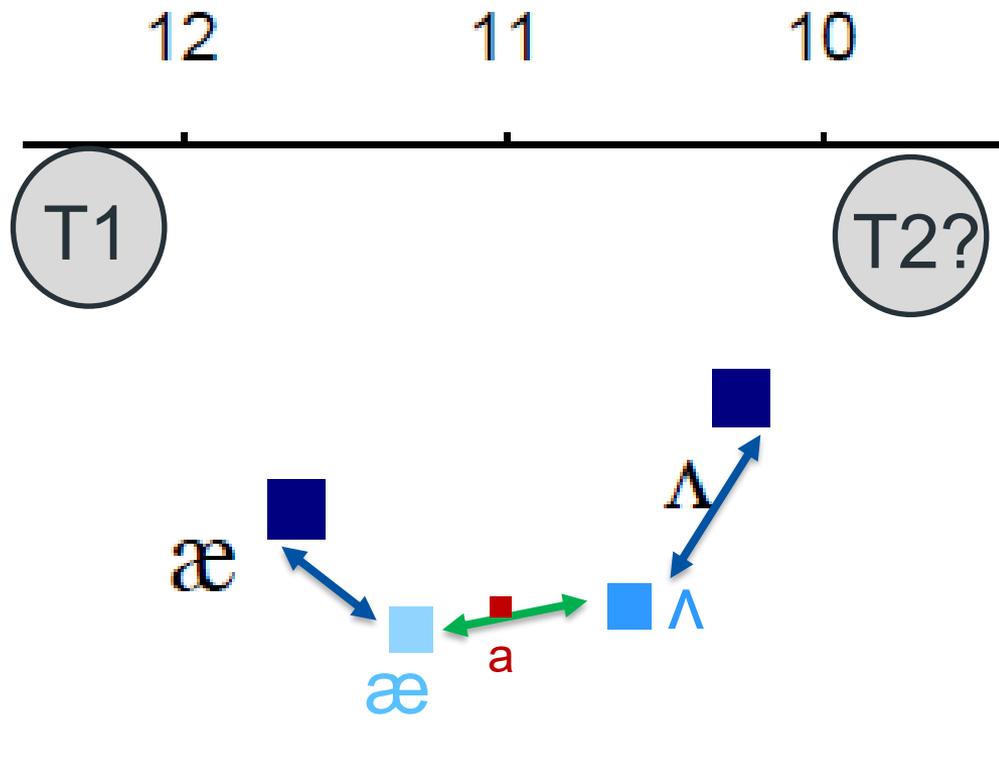
% identification (goodness rating)

		English stimuli	
		æ	ʌ
Listeners' response			
L1-Spanish	a	99 (5.5)	97 (5.4)
L1-Catalan	a	99 (5.6)	98 (4.7)

(Martínez-Celdrán, E., Fernández-Planas, A. M., & Carrera-Sabaté, J.. 2003; Recasens & Espinosa, 2006)

Cebrian (2019, 2021)

# Phonetic training intervention on English /æ/-/ʌ/ (HVPT)



→ measure possible changes in acoustic distances between testing times

→ How would acoustic distances change over time?

→ 3 examples in our data

- ↔ Distance æ - ʌ
- ↔ Distance æ - æ & ʌ - ʌ learner (L) - native (NS)

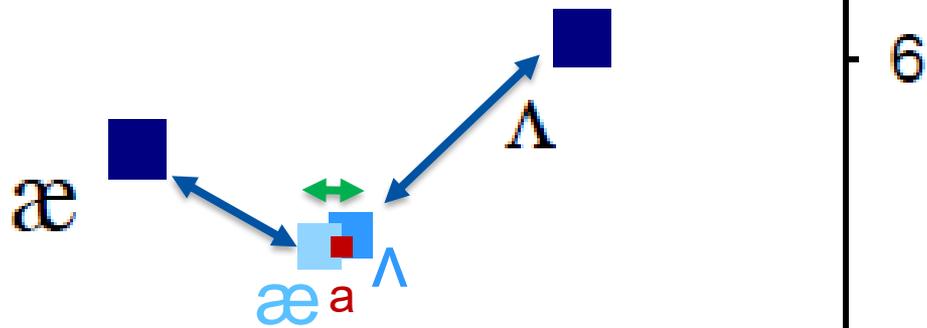
# Example 1

12

11

10

T1



Distinctiveness vs. Natielkeness:  
/æ/ - /ʌ/ vs. L-NS  
→ different levels of distinctiveness and potential training effects.

1. /æ/ & /ʌ/ as Cat a at T1

↔ Distance æ - ʌ  
↔ Distance æ - æ & ʌ - ʌ  
learner (L) - native (NS)

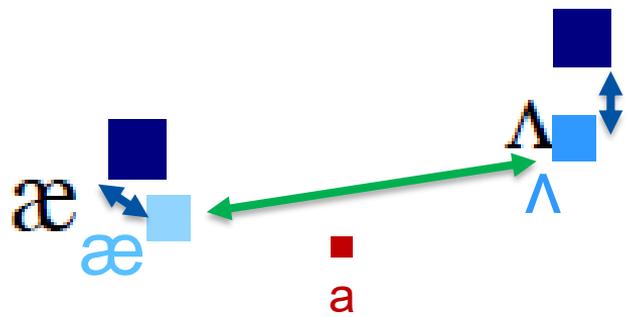
# Example 1

12

11

10

T2



Compared to T1 both vowels have become more accurate!

Distinctiveness vs. Nativelikeness:  
1. æ & ʌ target-like at T2

6 large distance æ - ʌ  
 small L-NS distance

7 Distance æ - ʌ  
 Distance L-NS  
æ - æ & ʌ - ʌ

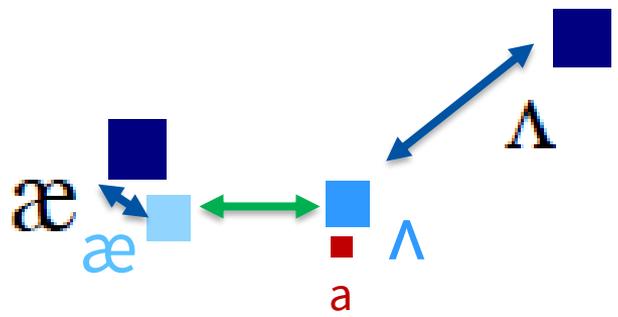
## Example 2

12

11

10

T1



Distinctiveness vs. Nativelikeness:  
2. /ʌ/ but not /æ/ as Cat a at T1

6

medium /æ/-/ʌ/ distance  
large L-NS distance for /ʌ/  
small L-NS distance for /æ/

7

Distance æ - ʌ  
Distance L-NS  
æ - æ & ʌ - ʌ

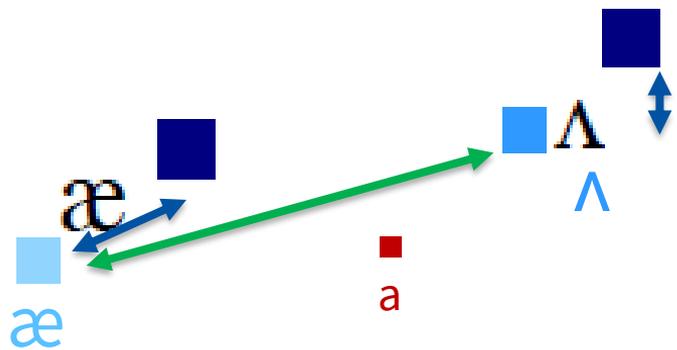
# Example 2

12

11

10

T2



Compared to T1 /æ/ has become less accurate!

Distinctiveness vs. Nativelikeness:  
2. /ʌ/ more target-like at T2  
/æ/ **less!** target-like at T2

6  $\longleftrightarrow$  larger distance /æ - ʌ/  
 $\longleftrightarrow$  smaller L-NS distance for /ʌ/  
larger L-NS distance for /æ/

7  $\longleftrightarrow$  Distance æ - ʌ  
 $\longleftrightarrow$  Distance L-NS  
æ - æ & ʌ - ʌ

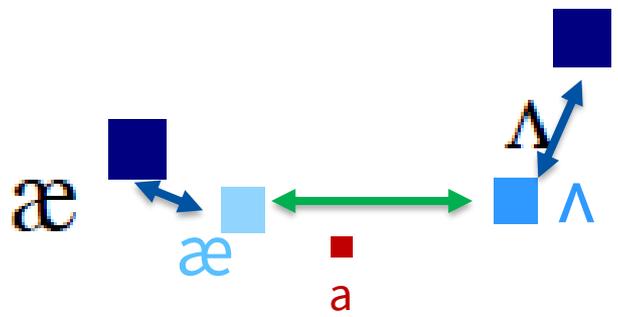
# Example 3

12

11

10

T1



Distinctiveness vs. Nativelikeness:  
3. /æ/ and /ʌ/ quite target-like at T1

6 quite large distance æ - ʌ  
 rather small L-NS distance

7 Distance æ - ʌ  
 Distance L-NS  
æ - æ & ʌ - ʌ

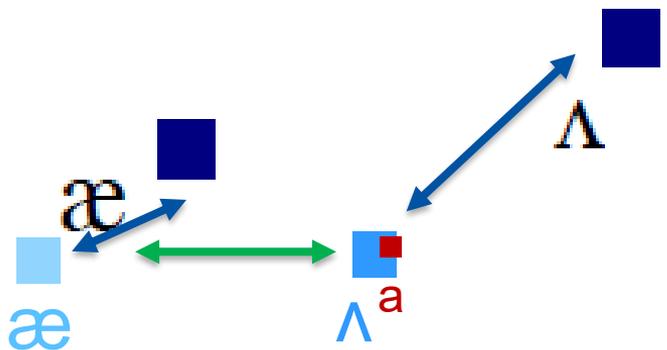
### Example 3

12

11

10

T2



Compared to T1 both vowels have become **less** accurate!

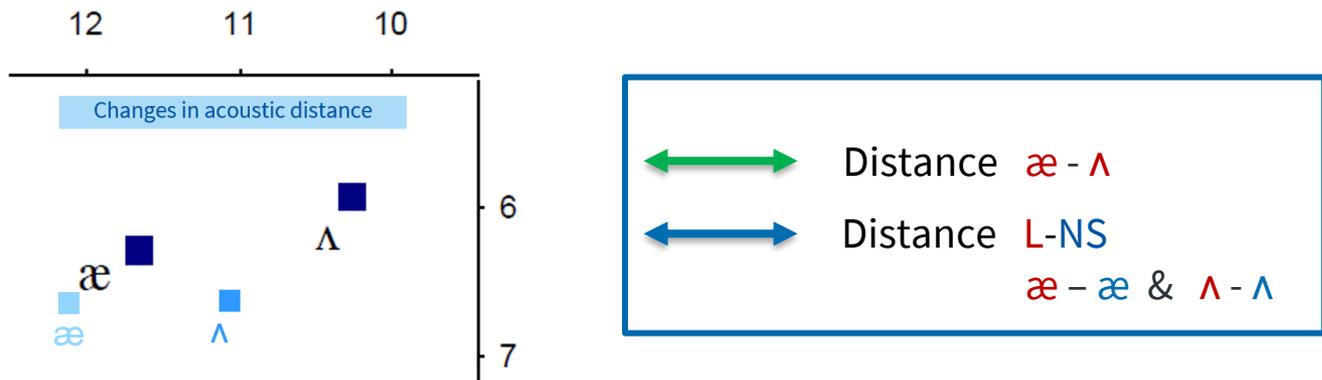
Distinctiveness vs. Nativelikeness:

2. /ʌ/ **less!** target-like at T2  
 /æ/ **less!** target-like at T2

6 same distance /æ/ - /ʌ/  
 larger L-NS distance for /æ/ and /ʌ/.

7 Distance æ - ʌ  
 Distance L-NS  
 æ - æ & ʌ - ʌ

## Examples 1, 2, 3: after a phonetic training intervention

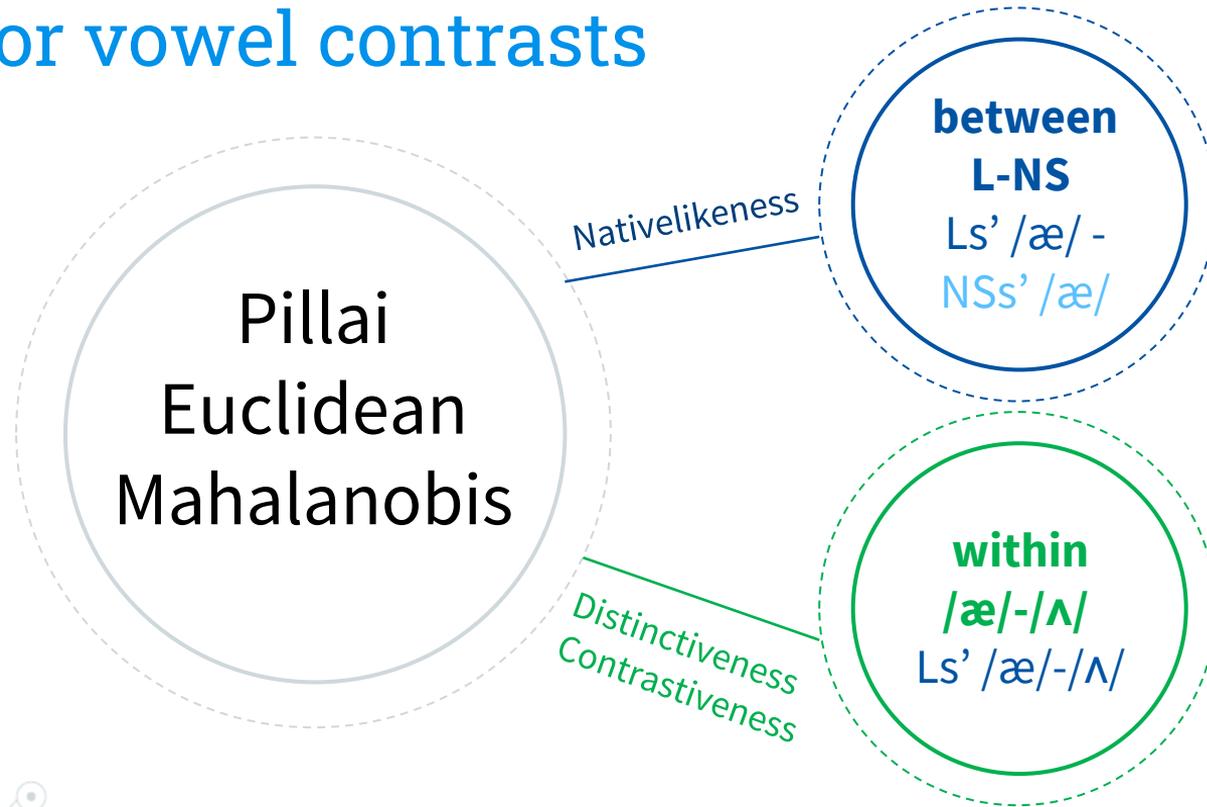


Changes in acoustic distance between contrasting vowels (distinctiveness æ - ʌ) do NOT necessarily correspond to changes in how close learners get to native speakers (nativeness L-NS)

Acoustic

Perceptual

# Acoustic distance measures for vowel contrasts

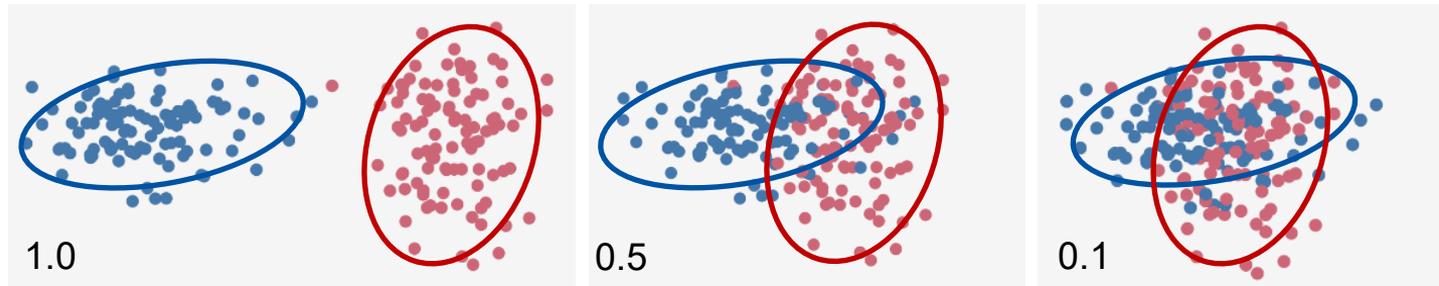


See also **LDA (linear discriminant analysis) classification accuracy scores**, which measure amount of overlap between realisations of different phonological categories

## Pillai scores

### Degree of overlap between two distributions (0-1)

Outcome of multivariate analysis of variance (MANOVA)

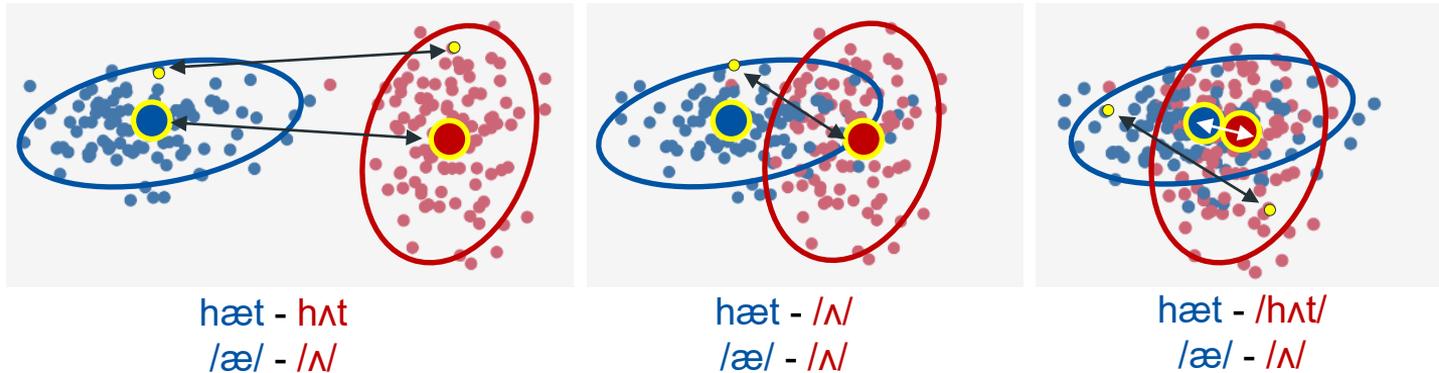


adapted from Stanley (2021)

- does not compute distances between data points
- 1 overlap measure per participant and vowel contrast
- may include other dimensions as co-variates (e.g. duration)

## Euclidean distances

Distance between two points from two distributions (e.g. in Bark), or between one point and a Mean, or between two Means.

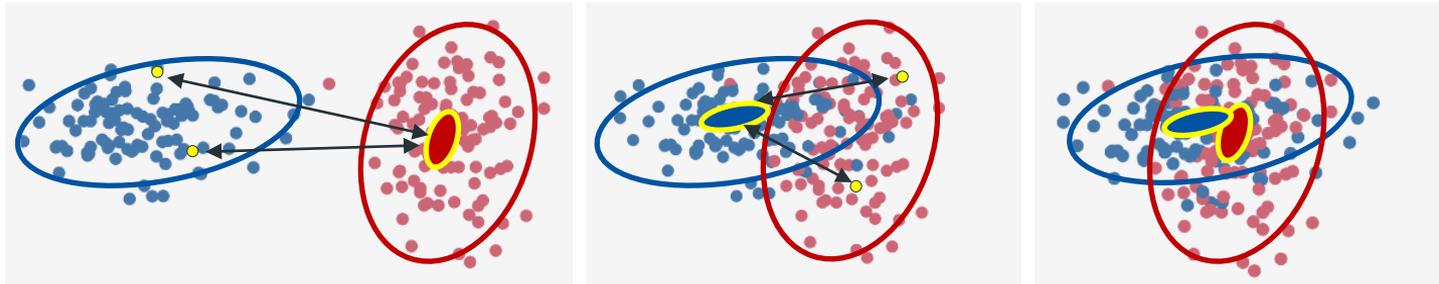


- distances between data points (MPs) or Means
- distances may be computed by item
- Means don't take into account the shape of the distribution

e.g. Flege, Bohn & Jang (1997), Kewley-Port & Atal (1989), Iverson & Evans (2009), Iverson & Kuhl (1996)

# Mahalanobis distances

## Distance (in SDs) between a point and the centroid of a distribution



*hæt* - /ʌ/

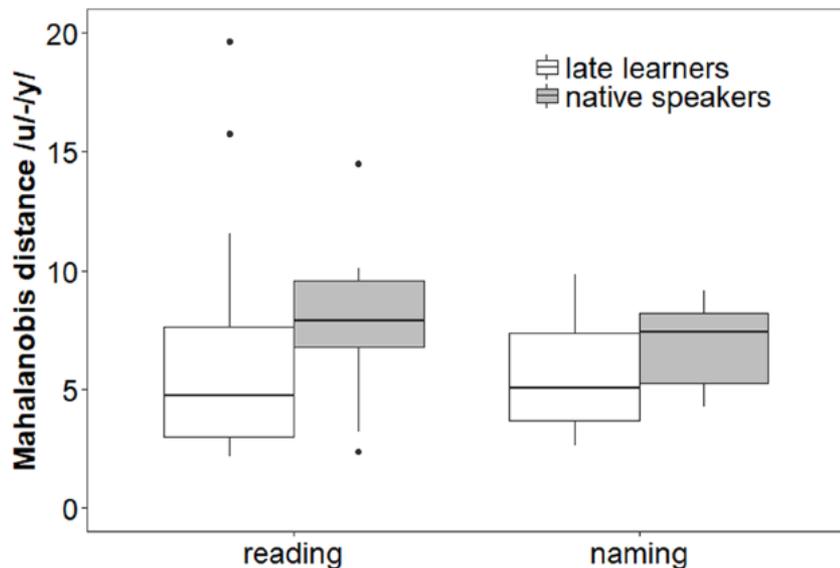
*/hæt/* - /æ/

- it is a unitless, scale-invariant measure
- the centroid takes into account the shape of the distribution (i.e. it reflects token variability)
- distances may be computed by item

# Contrastiveness vs. Nativelikeness (not both)

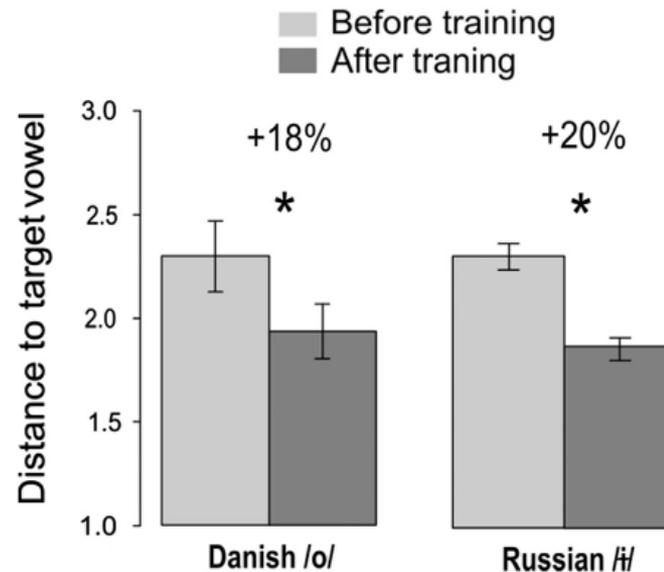
## Gains in pronunciation accuracy after HVPT interventions

Distances between contrasting vowels (Mahalanobis)



Melnik-Leroy, Turnbull, & Peperkamp (2021)

Distances (Mahalanobis) to target vowels  
Learner (L) – Native Speaker (NS)



Kartushina et al. (2014, 2015, 2016)

## Contrastiveness vs. Nativelikeness: how do we relate them?

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/–/y/ contrast. *Second Language Research*, 38(3), 581-605.

In order to measure the acoustic distance between the /u/ and /y/ categories we used the Mahalanobis distance metric, which measures the number of standard deviations from a point to the mean of a distribution. For each vowel contrast for each participant, we computed the mean Mahalanobis distance between each token and the distribution of the other category. Thus, for every participant we obtained the mean Mahalanobis distance from each /u/ token to the entire /y/ category and from each /y/ token to the entire /u/ category. Finally, we summed these two distances to obtain an individual measure of the distance between the two categories. A larger distance is indicative of a better separation between the two vowels, and, by hypothesis, of a higher production accuracy. The mean individual distance scores are shown in Figure 2.

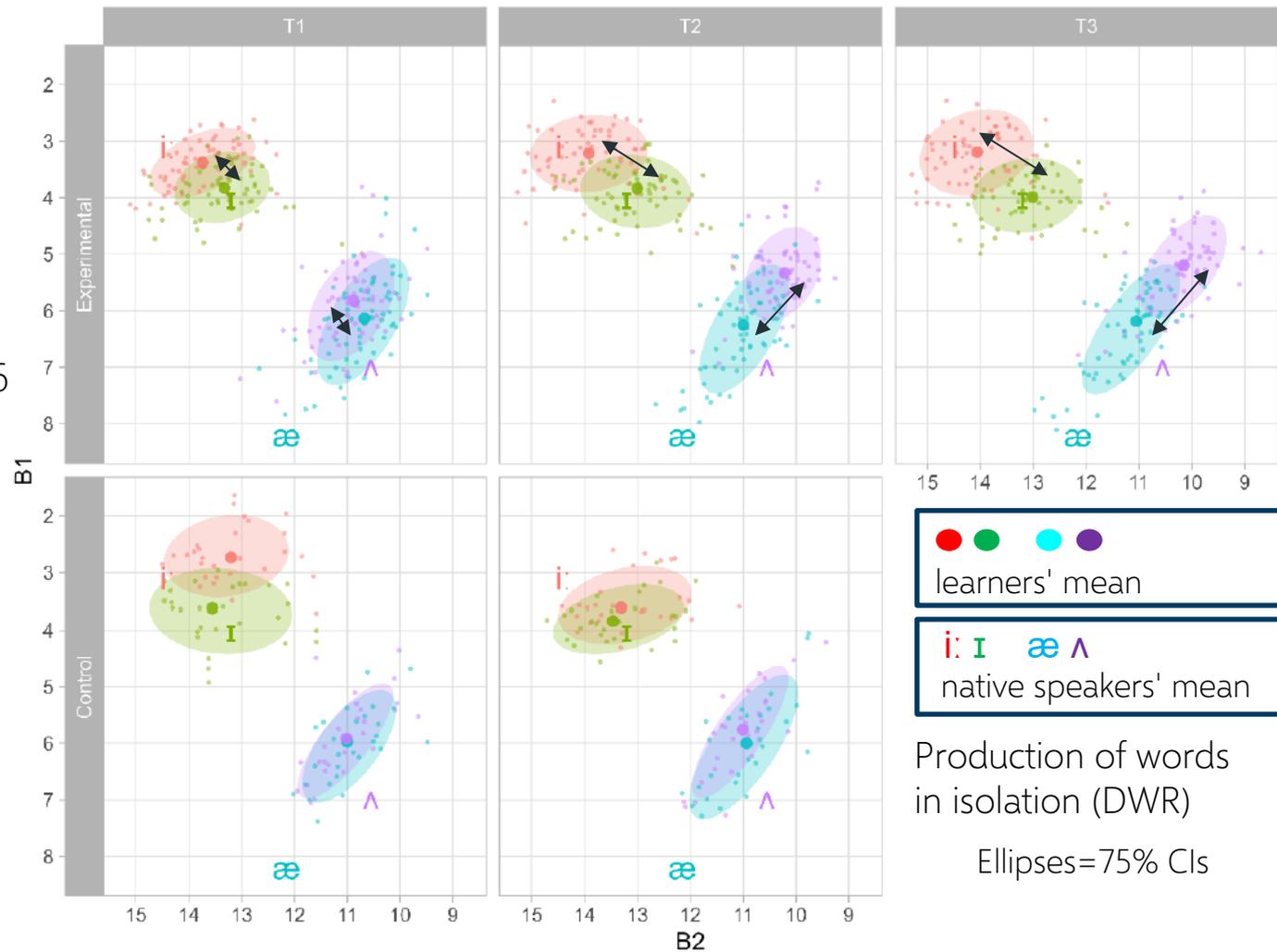


Mora-Plaza (2023)

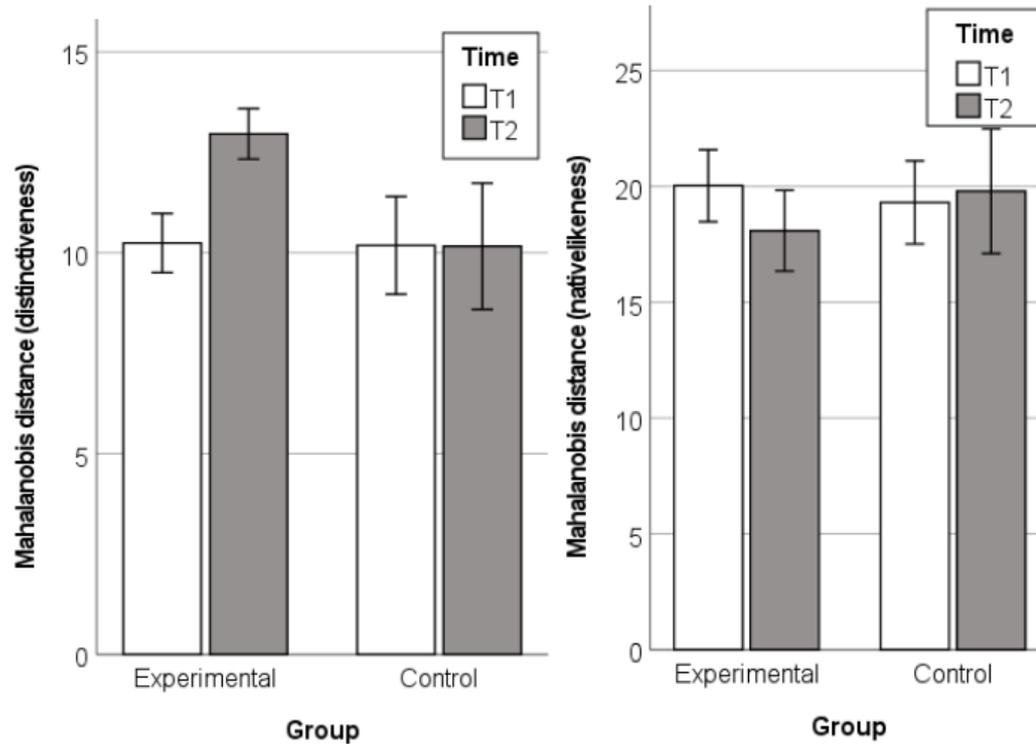
## TBPT intervention using pronunciation-focused communicative tasks.

→ 92 L1-Sp/Cat instructed EFL school learners aged 16  
Experimental  $N=63$   
Control  $N=29$

→ Exp. improved sig. in  
- contrastiveness  
- nativelikeness

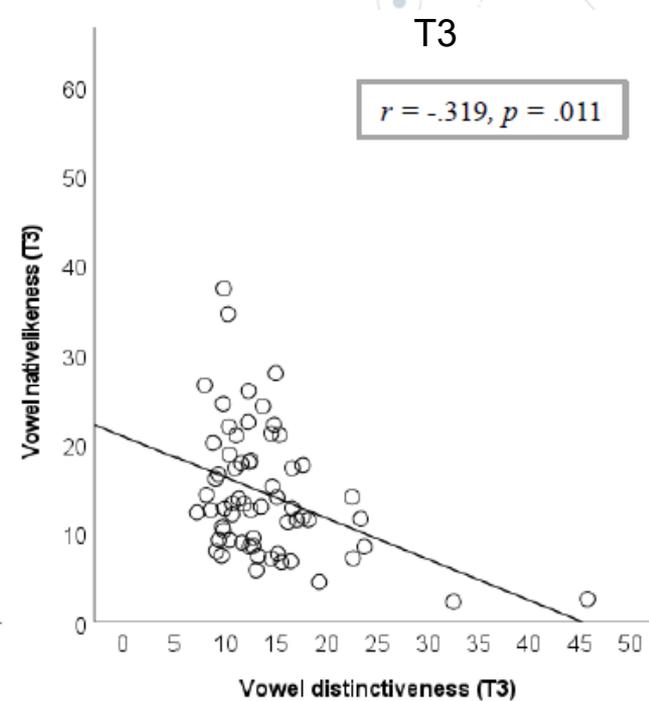
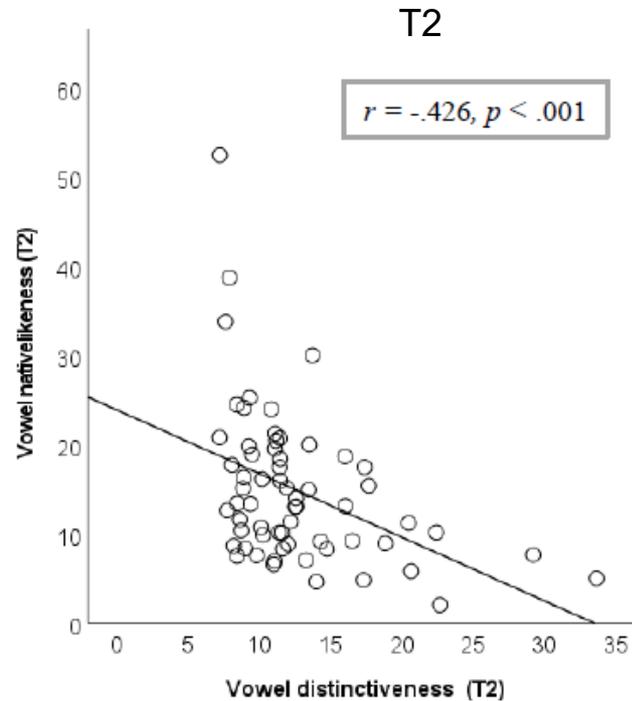
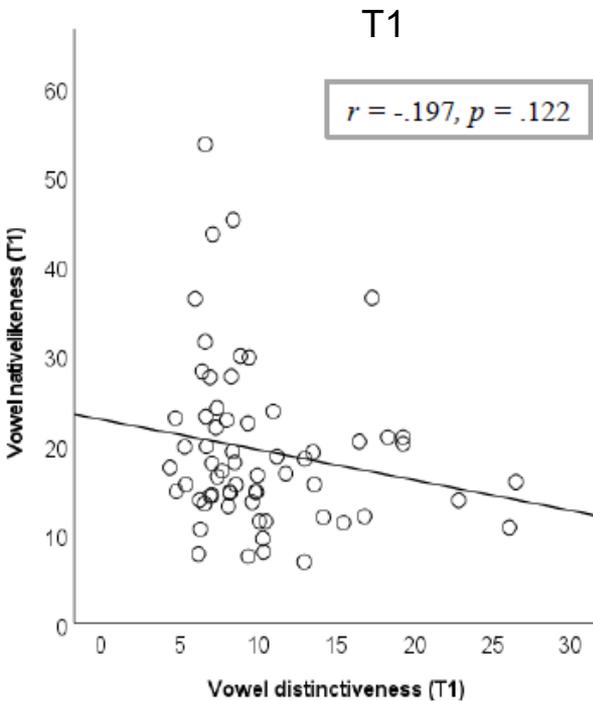


# Contrastiveness and Nativelikeness

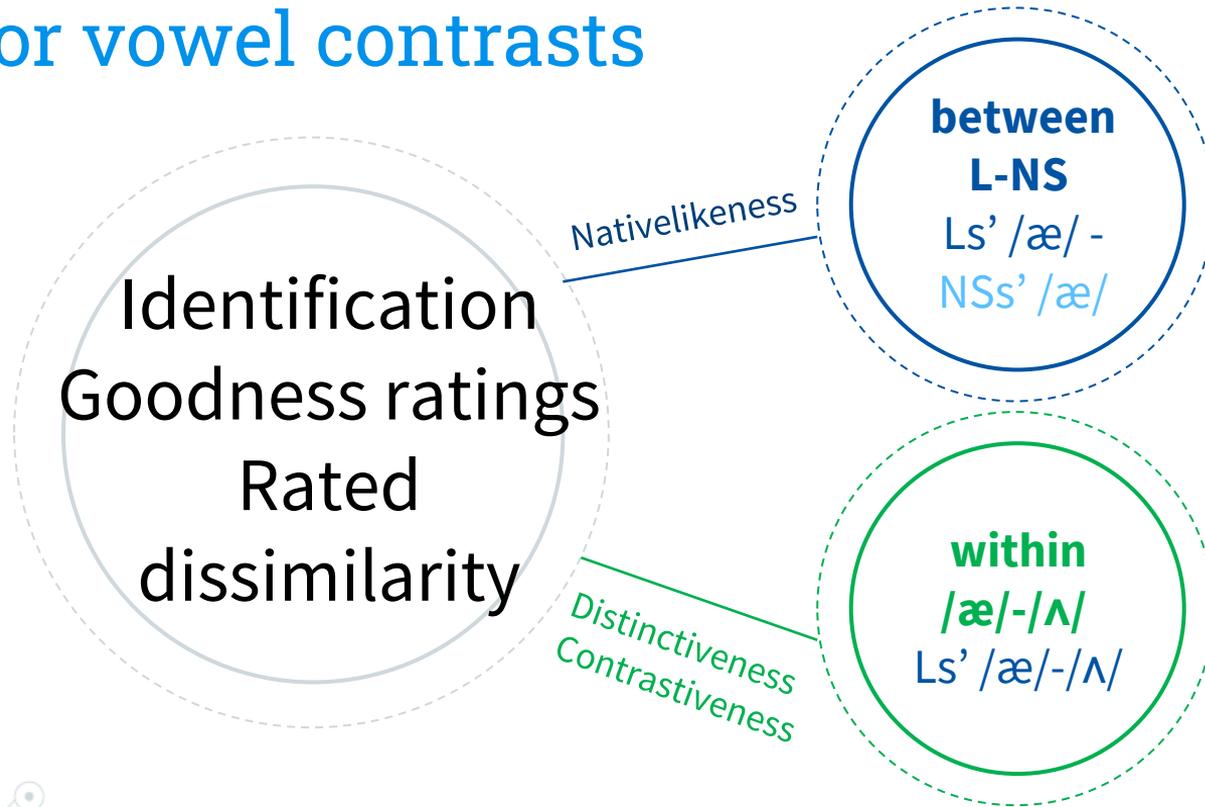


# Contrastiveness and Nativelikeness

words in isolation (DWR)



# Perceptual distance measures for vowel contrasts



See also *LDA (linear discriminant analysis) classification accuracy scores*, which measure amount of overlap between realisations of different phonological categories

# Contrastiveness and Nativelikeness

## Assessing L2 pronunciation gains after HVPT intervention

Comparing Ls' productions (auditory stimuli) to native listeners' representations through goodness ratings: → **nativelikeness distance**

Goodness ratings:

hard to identify = 1 ----- 9 = easy to identify

SOUND	CONTROL		DIS		ID	
	PRE	POST	PRE	POST	PRE	POST
/æ/	4.7 (2.7)	4.1 (1.8)	4.6 (1.8)	5.8 (2.8)	4.0 (2.5)	4.5 (2.1)
/ʌ/	2.0 (2.2)	1.8 (2.8)	3.3 (2.6)	3.3 (2.2)	2.3 (1.7)	3.4 (3.2)
/i:/	5.1 (1.7)	4.9 (3.2)	5.8 (2.1)	6.1 (2.6)	6.3 (1.7)	6.6 (3.2)
/ɪ/	4.8 (2.5)	4.4 (2.9)	3.9 (2.5)	3.9 (2.9)	4.7 (2.7)	5.1 (2.2)
/ɜ:/	6.1 (1.8)	5.9 (2.8)	6.3 (2.5)	6.4 (2.1)	5.9 (2.8)	6.5 (1.6)

Table 5. Median ratings obtained for each vowel per group (standard deviations are given in parentheses).

# Contrastiveness vs. Nativelikeness

## Identification + goodness ratings (nativelikeness)

æ	ʌ
<i>back</i>	<i>buck</i>
<i>bad</i>	<i>bud</i>
<i>bag</i>	<i>bug</i>
<i>cap</i>	<i>cup</i>
<i>cat</i>	<i>cut</i>
<i>fan</i>	<i>fun</i>
<i>hat</i>	<i>hut</i>
<i>lack</i>	<i>luck</i>
<i>mad</i>	<i>mud</i>
<i>match</i>	<i>much</i>
<i>pan</i>	<i>pun</i>
<i>sack</i>	<i>suck</i>

		Perceptual Distance Score	
		<i>r</i>	<i>p</i>
Pillai	between	.054	.612
	within	-.344	.001
Euclidean	between	.119	.260
	within	-.301	.004
Mahalanobis	between	.073	.491
	within	-.225	.031

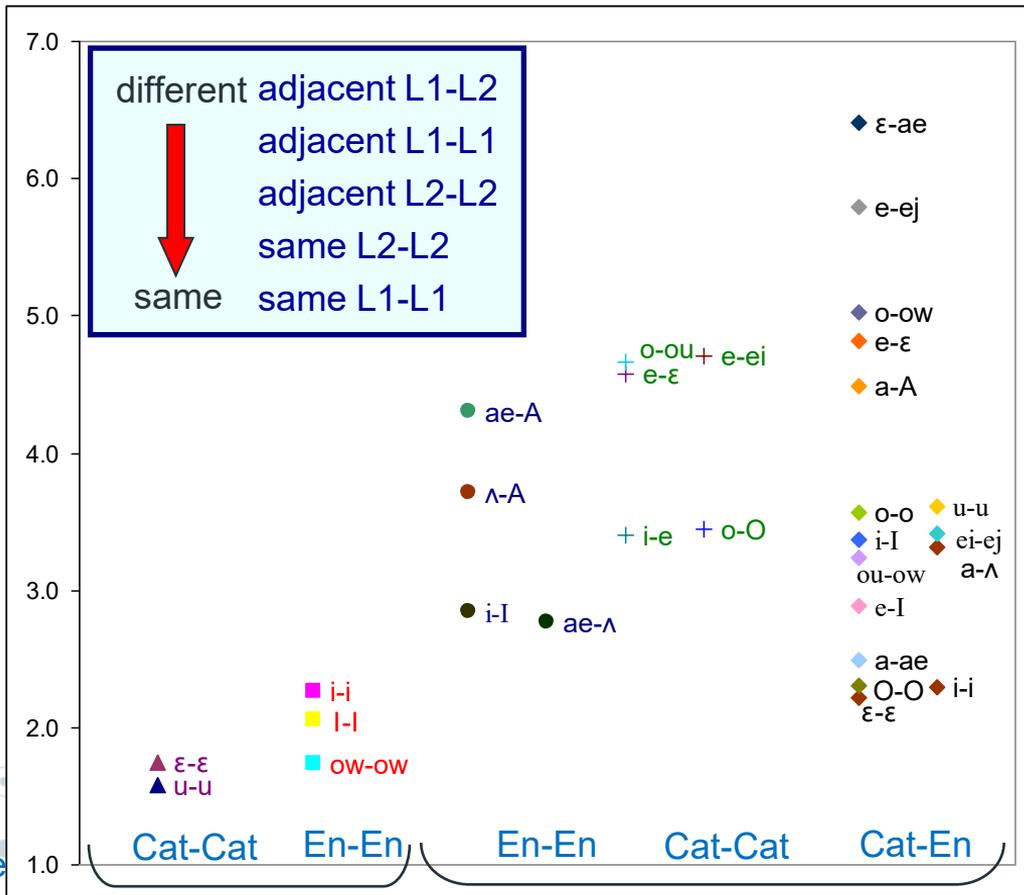
The more distinctly learners produce /æ/ from /ʌ/, the more accurate these productions are judged to be. BUT relationships are weak.

Word repeated	Listener's Identification	Goodness	Perceptual distance Score
<i>cap</i>	<i>cap</i>	9	1
	<i>cap</i>	8	2
	<i>cap</i>	7	3
	<i>cap</i>	6	4
	<i>cap</i>	5	5
	<i>cap</i>	4	6
	<i>cap</i>	3	7
	<i>cap</i>	2	8
	<i>cap</i>	1	9
	<i>cup</i>	1	10
	<i>cup</i>	2	11
	<i>cup</i>	3	12
	<i>cup</i>	4	13
	<i>cup</i>	5	14
	<i>cup</i>	6	15
	<i>cup</i>	7	16
	<i>cup</i>	8	17
	<i>cup</i>	9	18

→ rated dissimilarity?

# Dissimilarity ratings

different



same

Same-vowel pairs

Adjacent-vowel pairs

Dissimilarity ratings by NSs assessing perceived dissimilarity btw L1 realizations and L2 realizations of /æ/ and /Λ/?

## Features of speaking tasks (controlled vs. spontaneous)

### Controlled speaking tasks

- Read aloud (words, sentences, passages)
- Delayed word repetition
- Delayed sentence repetition

Spontaneous speaking tasks: how "controlled" do they need to be for pronunciation assessment of vowel contrasts?

- Picture-based oral narratives
- Sentence elicitation tasks
- Problem solving tasks

## Alignment of training tasks and testing tasks

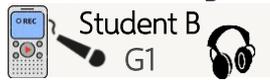
# Interactive Map Task (pair work) (pronunciation-focused)

Student A gives directions using the street names  
Student B follows directions clicking on the correct parcel.  
Street names were **clickable**.

14 parcels to pick (change colours when picked)  
12 target contrasting street names



Monitor A    Monitor B



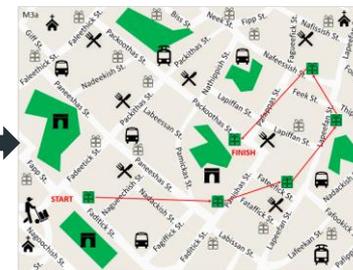
Monitor A    Monitor B



Simple



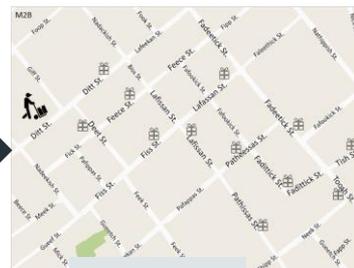
+ Complex



++ Complex



Simple



+ Complex



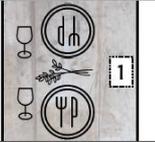
++ Complex

Testing: Delayed nonword repetition → no changes in /i:/ and /I/

# Problem-solving task (the dinner-table task)

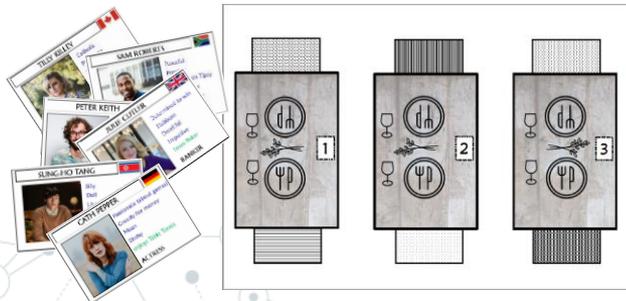
## Pronunciation un-focused

Simple

<b>TILLY KILLEY</b>  Catholic Pragmatic Talented artist Charismatic <i>into Wine Tasting</i> <b>POLITICIAN</b>	<b>PETER KEITH</b>  Gentle & open-minded Humble Simple & Kind Keen on martial arts <i>Does Kung fu</i> <b>TEACHER</b>	<b>SAM ROBERTS</b>  Peaceful Popular Funny when Tippy Trustworthy <i>likes Painting</i> <b>POLITICIAN</b>
 1	 2	 3
<b>SUNG-HO TANG</b>  Silly Dull Liberal Dreamy <i>practices Kickboxing</i> <b>JUDGE</b>	<b>CATH PEPPER</b>  Passionate (about games) Greedy for money Mean Bitchy <i>enjoys Table Tennis</i> <b>ACTRESS</b>	<b>JULIE CUTLER</b>  Determined to win Stubborn Disdainful Impulsive <i>loves Poker</i> <b>BANKER</b>

Consider this sitting arrangement carefully.  
Justify why it would not work.

- Provide as many REASONS as you can think of in terms of their PERSONALITY traits, PROFESSION and HOBBIES.



Your task is to come up with a NEW sitting arrangement that generates PLEASANT conversations and makes the dinner party successful.

- Provide as many REASONS as you can think of in terms of their PERSONALITY traits, PROFESSION and HOBBIES.

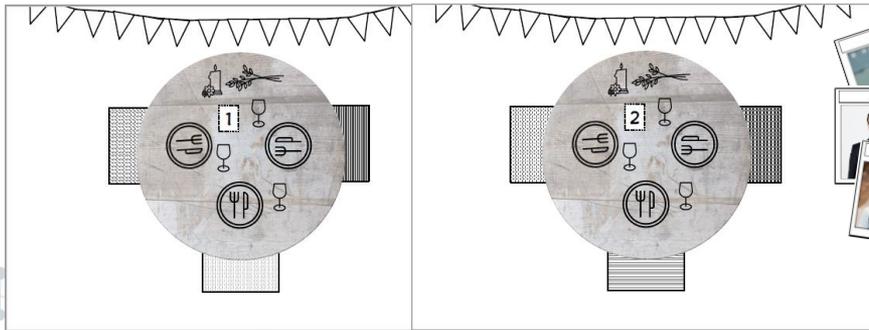
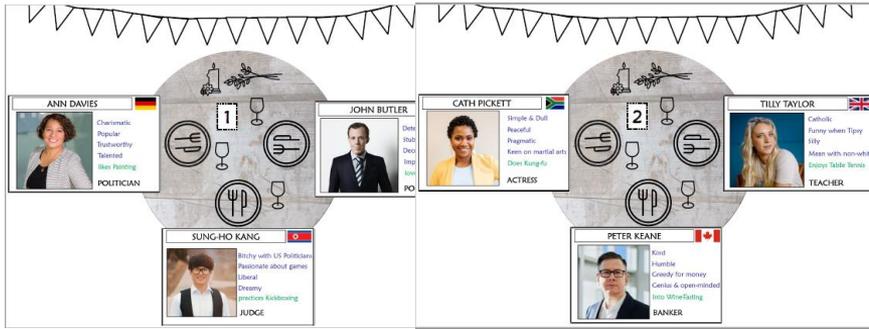
# Problem-solving task (the dinner-table task)

## Pronunciation un-focused

Complex

Cognitive complexity manipulations:  
→ 3 vs. 2 characters  
→ simple vs. complex personality

No focus on pronunciation  
→ 4-10 min of speech.



# Problem-solving task (the dinner-table task)

## Pronunciation un-focused

- Vowel contrasts (/i:/-/ɪ/, /æ/-/ʌ/)
- VOT in /p/, /t/, /k/

### Simple

Name	SAM ROBERTS	JULIE CUTLER	CATH PEPPER	TILLY KILLEY	SUNG-HO TANG	PETER KEITH		N	N
Personality/ideology	peaceful	determined to win	passionate (esp. about games)	catholic	silly	genius (& open-minded)	æ	9	p 10
	popular	stubborn	greedy for money	pragmatic	dull	humble	ʌ	9	t 9
	funny when tipsy	deceitful	mean	talented artist	liberal	simple & kind	i:	9	k 8
	trustworthy	impulsive	bitchy	charismatic	dreamy	keen on martial arts	ɪ	9	
Occupation	POLITICIAN	BANKER	ACTRESS	POLITICIAN	JUDGE	TEACHER			
Interests	Painting	Poker	Table Tennis	Wine-tasting	kickboxing	kung fu			
FLAGS	South Africa	UK	Germany	Canada	Korea	United States			

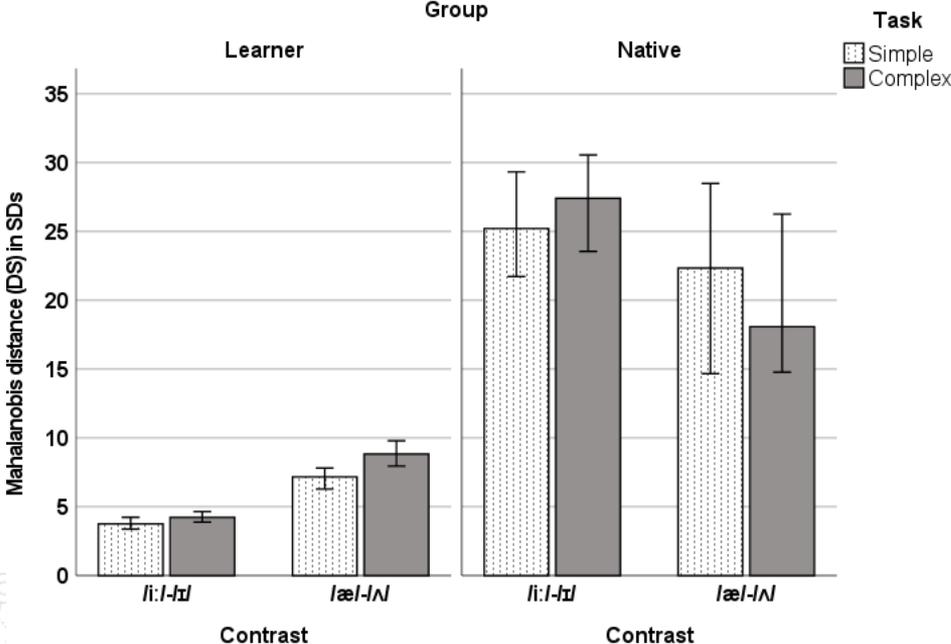
### Complex

Name	ANN DAVIES	JOHN BUTLER	CATH PICKETT	TILLY TAYLOR	SUNG-HO KANG	PETER KEANE		N	N
Personality/ideology	popular	determined to win	simple & dull	catholic	bitchy with US politicians	kind	æ	9	p 10
	charismatic	stubborn	keen on martial arts	funny when tipsy	passionate about games	humble	ʌ	9	t 9
	talented	deceitful	peaceful	silly	liberal	greedy for money	i:	9	k 8
	trustworthy	impulsive	pragmatic	mean with non-white	dreamy	genius & open-minded	ɪ	9	
Occupation	POLITICIAN	POLITICIAN	ACTRESS	TEACHER	JUDGE	BANKER			
Hobbies/Preferences	Painting	Poker	kung fu	Table Tennis	kickboxing	Wine-tasting			
FLAGS	Germany	United States	South Africa	UK	Korea	Canada			

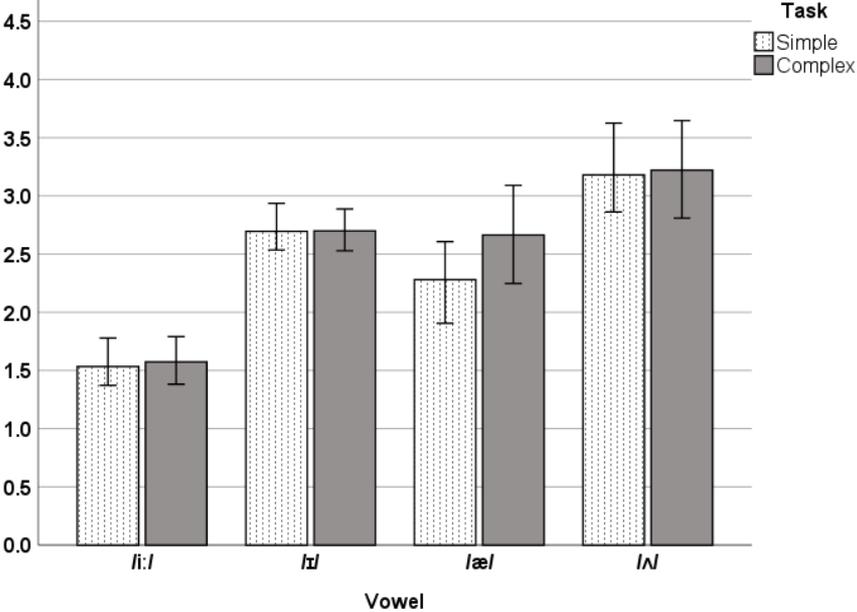
# Pronunciation: vowel quality (/i:/-/ɪ/, /æ/-/ʌ/)

## Mahalanobis distances

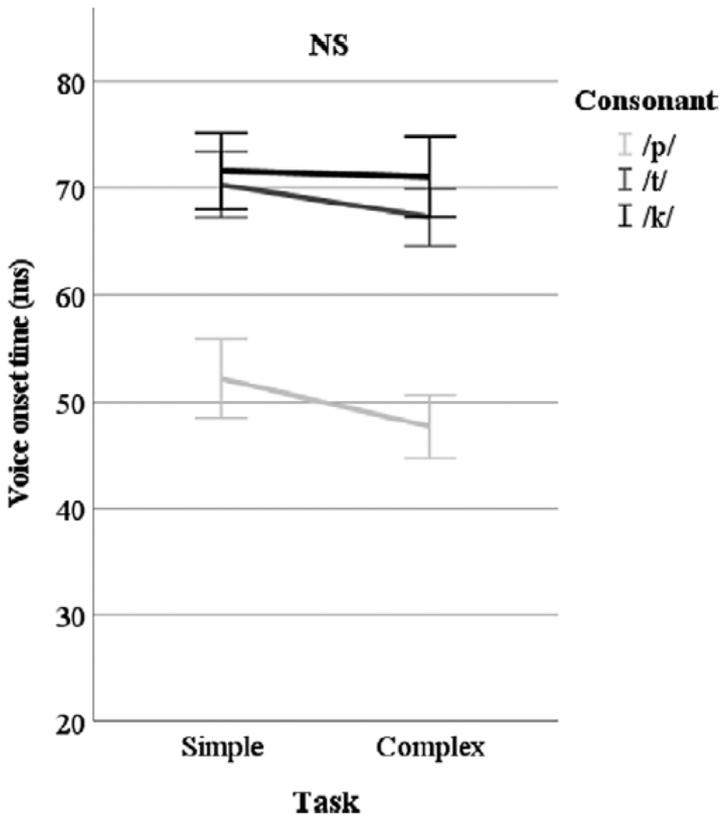
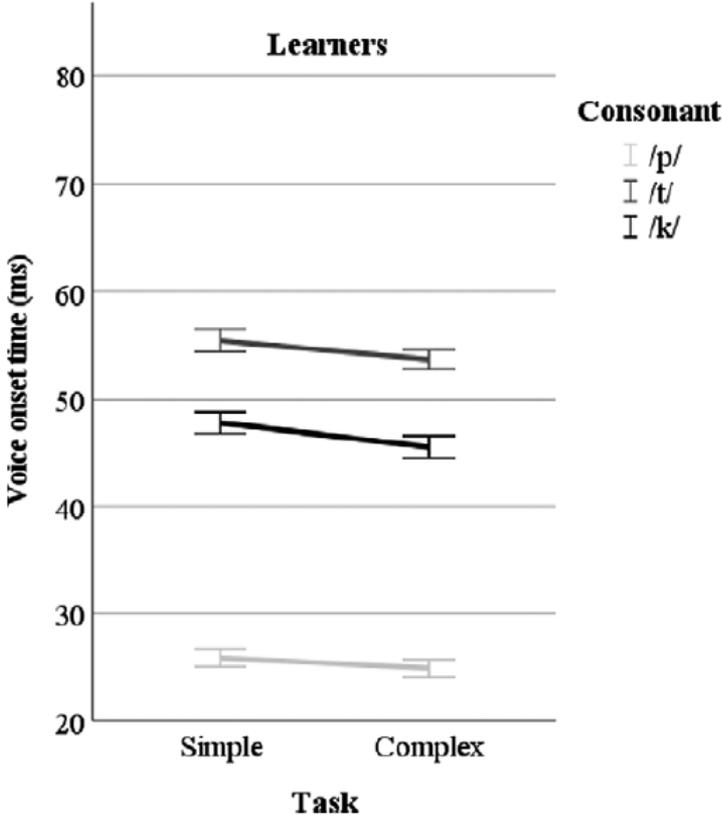
### Contrastiveness



### Nativeness



# Pronunciation: VOT in /p/, /t/, /k/



# Training to imitate an English accent on your L1 (L2AIL1)

→ Communicative role-play tasks focusing on 4 target pronunciation features of English Spanish learners have difficulty with:  
/p/-/t/-/k/ (VOT); /v/; /əʊ/; vowel reduction in unstressed syllables



VOT

L2AIL1

VS.  
L2I



# Training to imitate an English accent on your L1 (L2AIL1)

→ Communicative role-play tasks focusing on 4 target pronunciation features of English Spanish learners have difficulty with:  
/p/-/t/-/k/ (VOT); /v/; /əʊ/; vowel reduction in unstressed syllables (VR)

## L2AIL1

- Botella de aceite de **oliva**
- Bebidas típicas (cerveza, vermut, vino, cava)
  - **Cerveza** local
  - **Vermut**
  - **Vino**
  - **Cava**
- **Vela** decorativa
- Turrones típicos de **Navidad**
- **Vasos** de vidrio con grabados de los monumentos representativos de la ciudad
- **Llavero** de la bandera de Barcelona
- **Vestido** flamenco



## The perfect gift from Barcelona

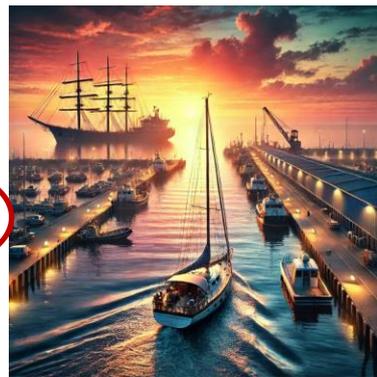
## L2I



- **Vintage** postcard
- Bottle of **cava**
- Music **Festival** badge from Primavera Sound
- Journal with **velvet** cover featuring iconic Barcelona
- Barcelona-themed **vanilla**-scented candle
- Decorative **vase** with hand-painted designs
- **Vermouth**
- Sagrada familia pendant in **silver**
- Specialty wine **vinegar** from local vineyards
- Gourmet chocolate bars featuring Mediterranean ingredients like **lavender**

# Testing: make a sentence describing the picture using the words given

## L2AIL1 Read aloud



boat  
port



van  
furniture



bingo  
prize



bar  
salsa dancing

"Llegué a Barcelona una tarde cálida y, tras dejar mi maleta, pedí una cerveza en un café mientras observaba la vida en la calle. Luego, pasé por un parque y vi a un hombre maduro sentado en un banco con su perro blanco, que movía la cola al verme pasar.

Por la noche, compré un tiquet en la taquilla para un espectáculo de flamenco. Una mujer con un vestido rojo me entregó la entrada con una sonrisa. La música y el baile me dejaron sin aliento.

Después, tomé un taxi hasta un club de salsa. Aunque no soy un gran bailarín, una mujer me animó a intentarlo. Entre risas y pasos torpes, me dejé llevar por el ritmo, disfrutando cada instante de mi aventura en Barcelona."

## Discussion and conclusion

Assessing segmental accuracy changes in the production of L2 vowel contrasts (/æ/-/ʌ/) between testing times or task conditions:

- distinctiveness / contrastiveness + nativelikeness
- both acoustic and perceptual measures
- both specific (rated dissimilarity) and global (accentedness, comprehensibility) perceptual measures
- spontaneous rather than read speech

## How can automated assessment help us?



# Thank you!

## Any questions?

mora@ub.edu

