Holistic perception of voice quality matters more than L1 when judging speaker similarity in short stimuli

Eugenia San Segundo, Paul Foulkes & Vincent Hughes

University of York

voice and identity

0101101ID1

SST 2016
Parramatta, Australia
7-9 December

Arts & Humanities Research Council

1. Introduction

- Voice quality (VQ)

  Quasi-permanent quality resulting from a combination of long-term laryngeal and supralaryngeal features (Laver, 1980) – **broad definition**

IDIOSYNCRATIC

FORENSIC PHONETICS!

- forensic speaker comparison
- earwitness evidence

APPROACH

Articulatory/**Perceptual/**Acoustic

naïve listeners experts

- holistic -featural

→ Differences in speaker similarity ratings by **native vs non-native** listeners?

2. Hypothesis

→ naïve listeners will rely on *holistic VQ perception* in order to judge similarity between speakers...

…. regardless of their L1
i.e. no native language advantage (cf. Perrachione et al. 2009)

- **when?** under controlled conditions of speaker similarity
- **what?** short speech samples
- **why?** VQ = only resource available for listeners to judge speaker similarity

3. Materials and method

3.1. Subjects

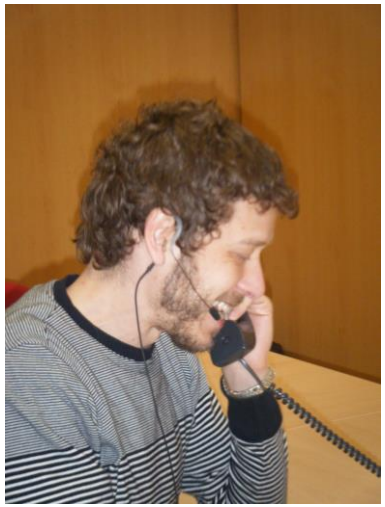
5 pairs male MZ twins:

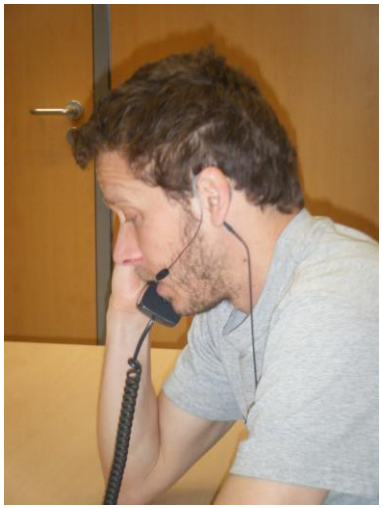
- native Spanish (Madrid)
- no voice pathologies
- similar sounding:
  1. similar age

     mean 21, sd 3.7
  2. similar mean F0

     mean 113 Hz, sd 13 Hz
  3. similar VQ

     expert (featural) assessment





3. Materials and method

3.1. Subjects

	VPA settings										
Speaker	Labial	Mandibular	Lingual Tip	Lingual body	Velopharyngeal	Pharyngeal	Larynx Height	VT tension	Larynx tension	Phonation type	SMC
AGF	0	0	0	0	0	0	2	1	1	1	
SGF	0	1	0	0	1	0	2	1	1	1	
Match	1	0	1	1	0	1	1	1	1	1	*0.8*

– using a simplified version of the VPA scheme:

e.g. mandibular setting

(close – neutral – open)

– Similarity Matching Coefficients

$$SMC = \frac{\text{number of setting matches}}{\text{number of settings}}$$

3. Materials and method

3.2. Stimuli and listeners

Stimuli

- approximately 3 secs
- from spontan. conversations
  - interlocutor = controlled
  - same speaking style
- declarative sentences
  - different ling. content
  - diverse neutral topics

Listeners

- 20 native Spanish speakers
  - age range 22-51; mean 33

- 20 native English speakers
  - age range 19-35; mean 25
  - **no knowledge of Spanish!**

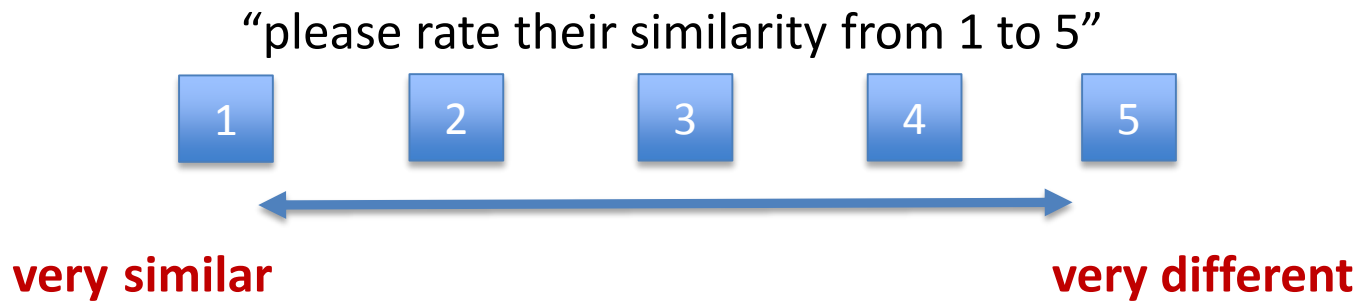
3. Materials and method

3.3. Design of perceptual test

- *MFC* Praat experiment

  90 different-speaker pairings – random order

- Instructions for listeners:



- Test duration = 15 min (break every 30 stimuli)
- **Listeners were not told that the test included twin pairs!**

3. Materials and method

3.4. Analysis methods

- **Multidimensional Scaling (MDS)**

  → to visualize degree of perceived similarity

  → to detect meaningful dimensions that explain observed (dis)similarities

- **Mixed-effects modelling**

  → to fit models to the similarity ratings

  — **Fixed effects** (predictors):

  - Listener language
  - SMC between speakers in the target trial
  - Reaction time
  - Twins – whether speakers were twins or not

  — **Random effects**:

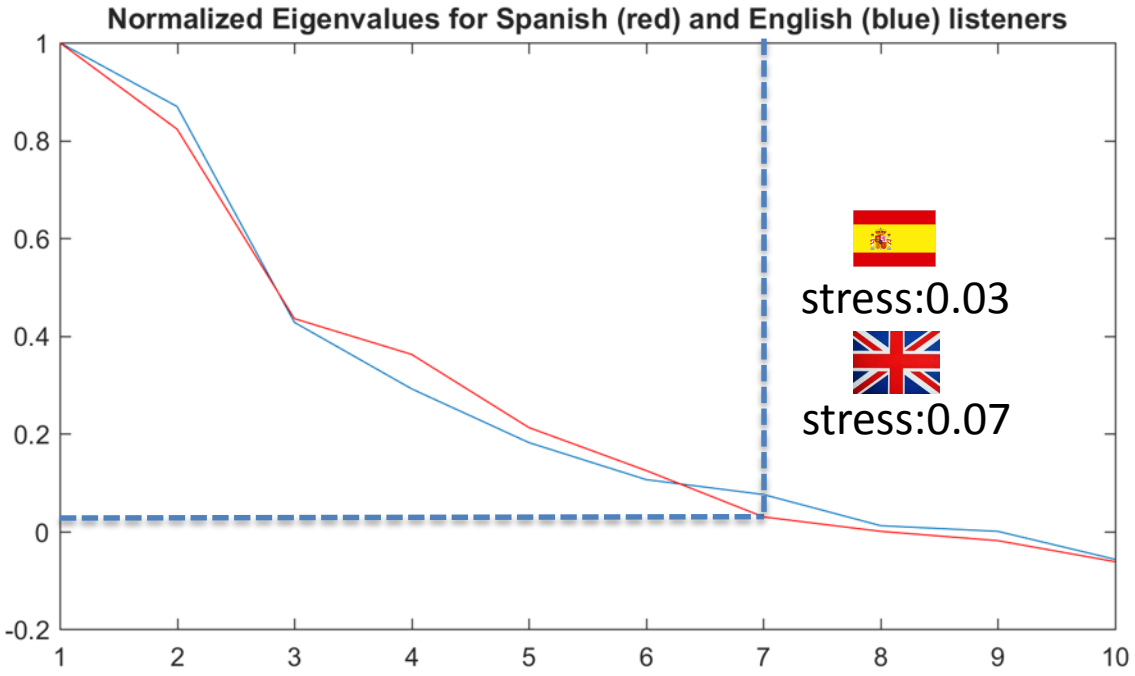
  - Listeners
  - Trial

  (target sp. comparison)

4. Results

- **MDS analysis**

  scree plot: relative magnitude of the sorted Eigenvalues

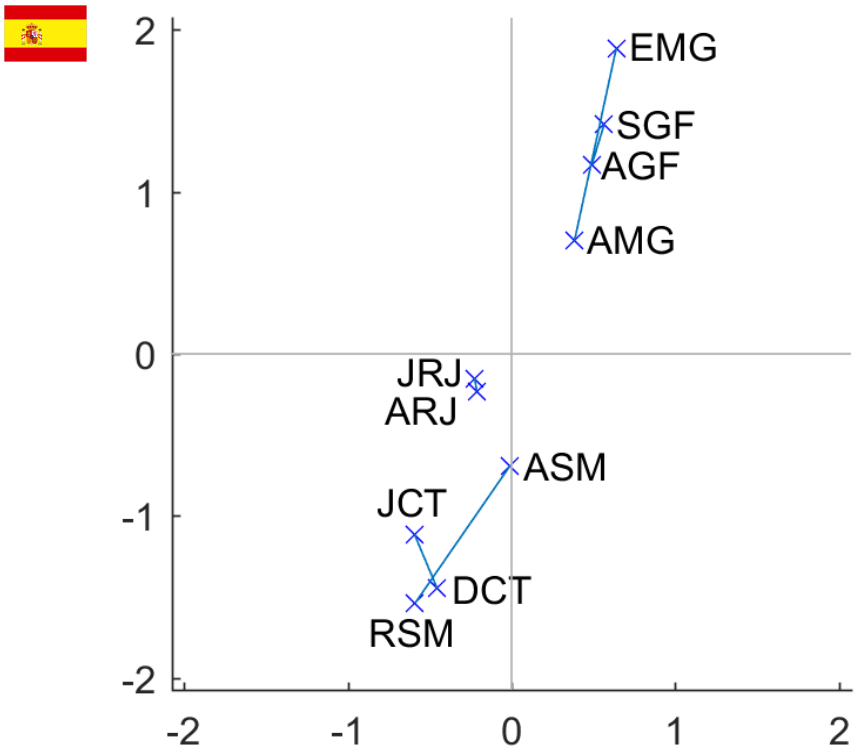


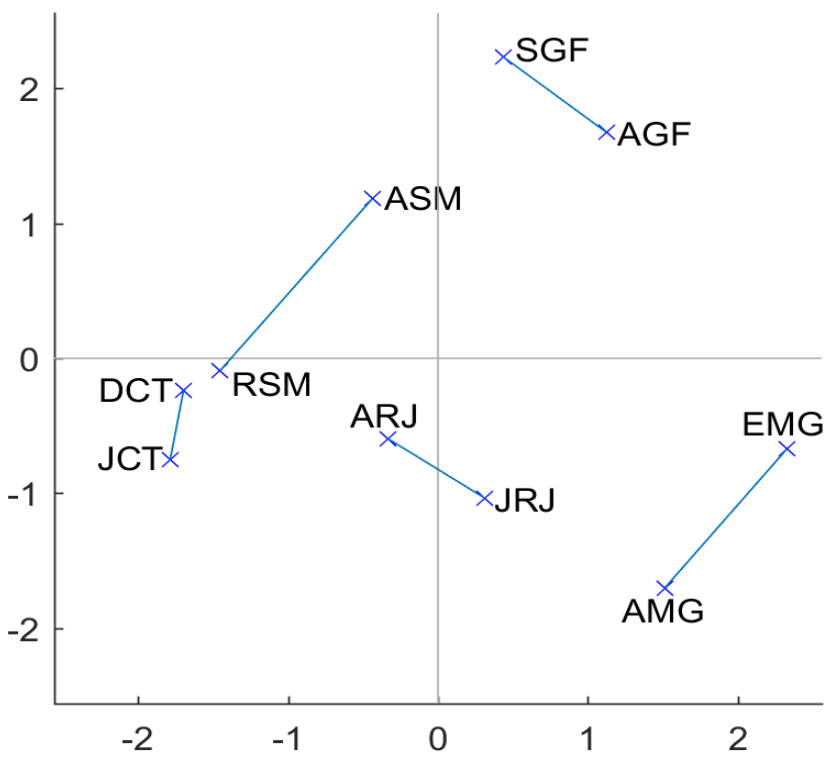
*stress = goodness-of-fit criterion to minimize.
Rule of thumb: <0.1 is excellent; >0.20 is poor

4. Results

- **MDS plots** (2D)

stress: 0.8

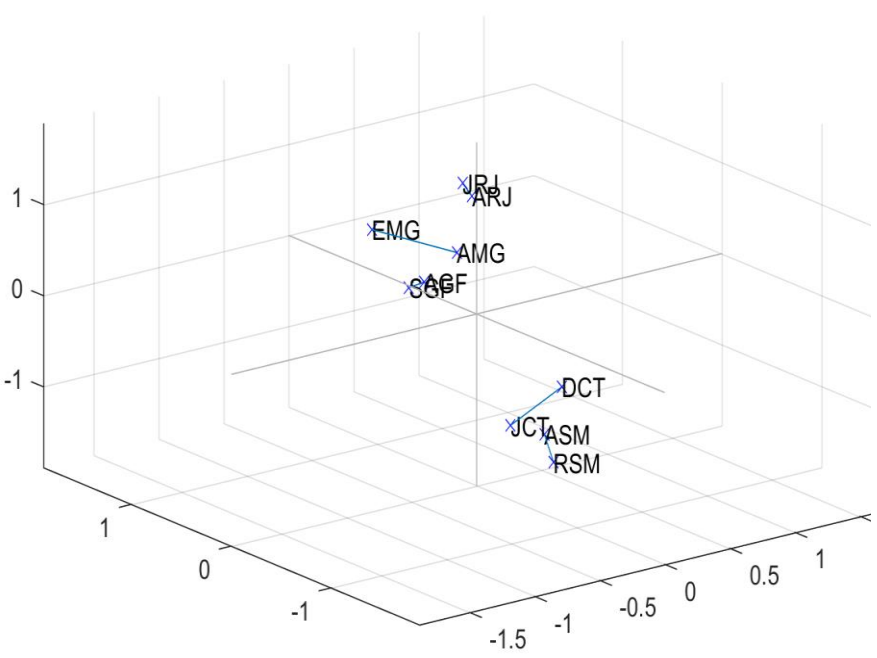


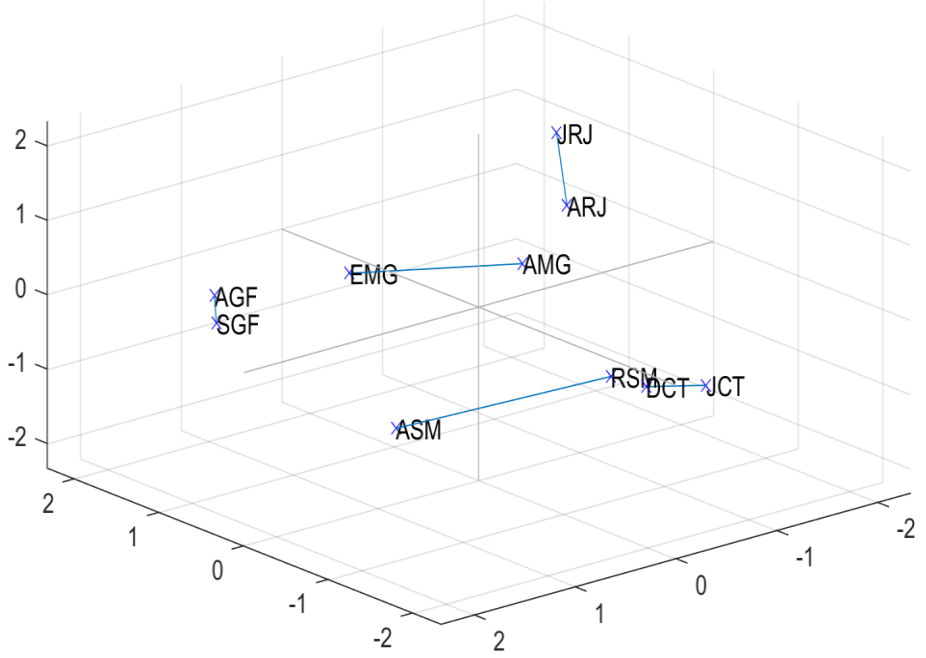


4. Results

- **MDS plots** (3D)

stress: 0.4





4. Results

- Intra-pair EDs based on 7D

speakers → / listeners ↓	AGF SGF	DCT JCT	ARJ JRJ	ASM RSM	AMG EMG
Spanish	0.341	0.343	0.345	0.369	0.607
English	0.264	0.219	0.349	0.435	0.445

most similar

most different

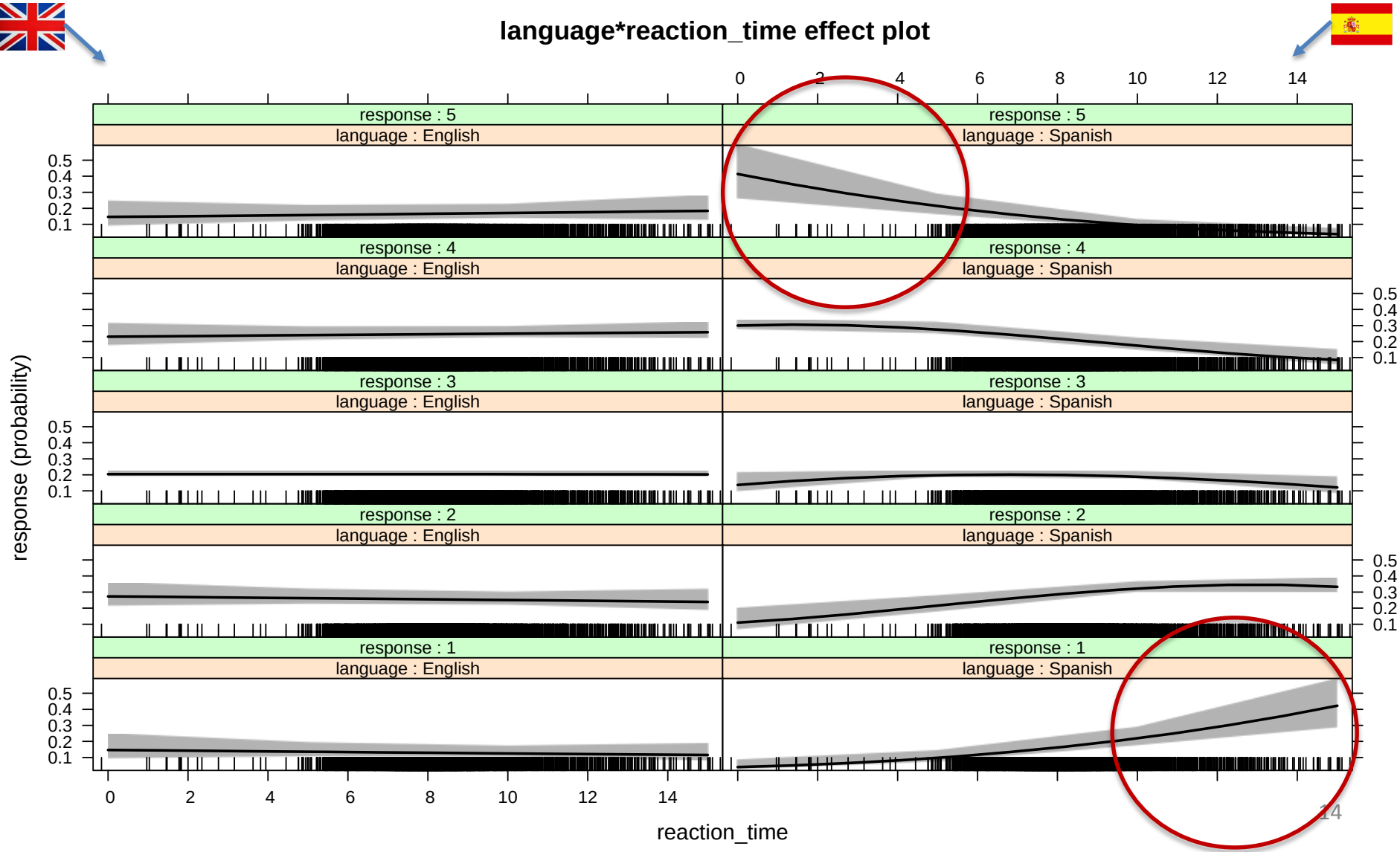
4. Results

- **Mixed-effects modelling**
  - Best model → all fixed effects + interactions
  - Significant interactions:

    ✓Language * Reaction time
    ✓Reaction time * Twins
    ✓SMC * Twins

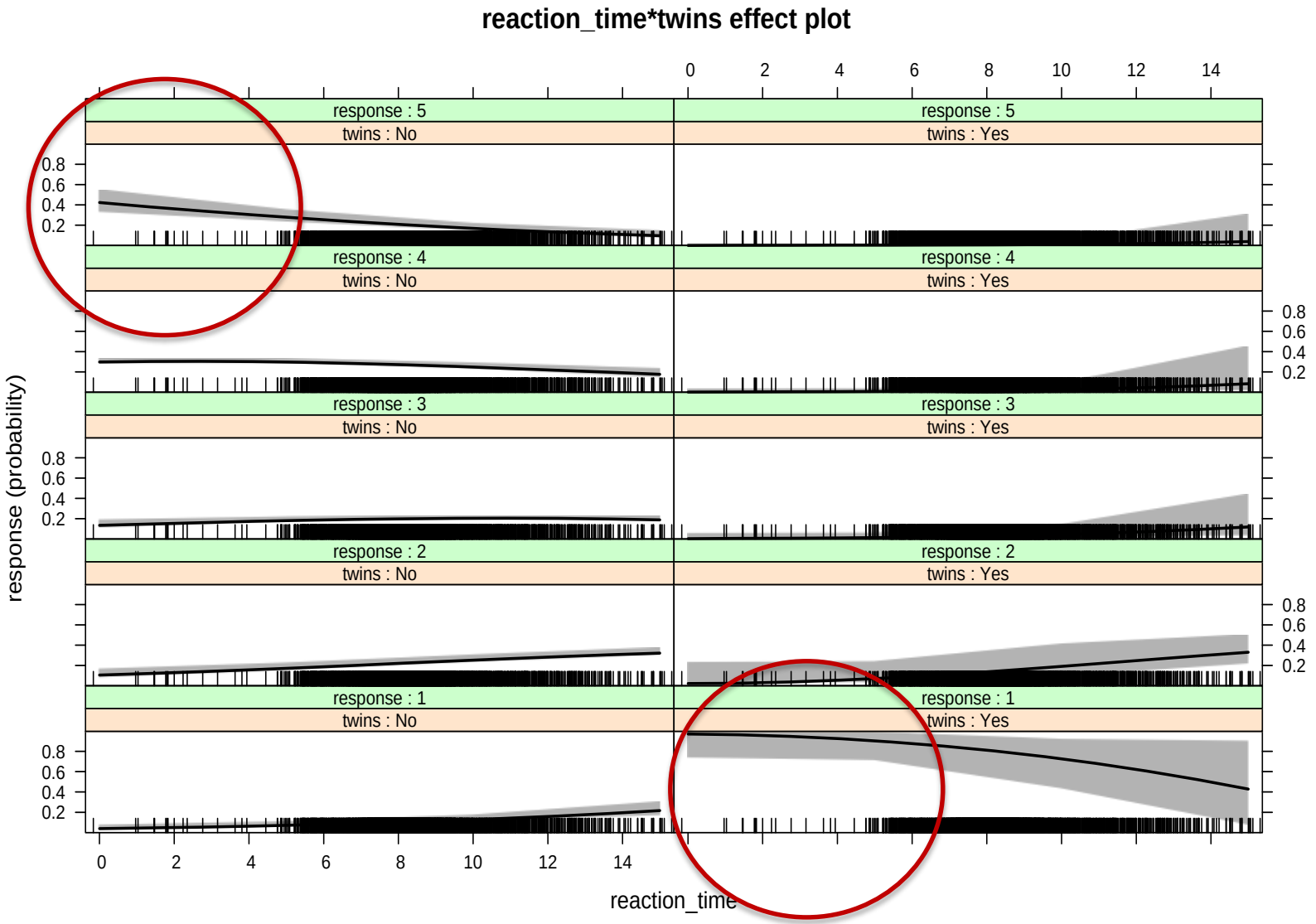
4. Results

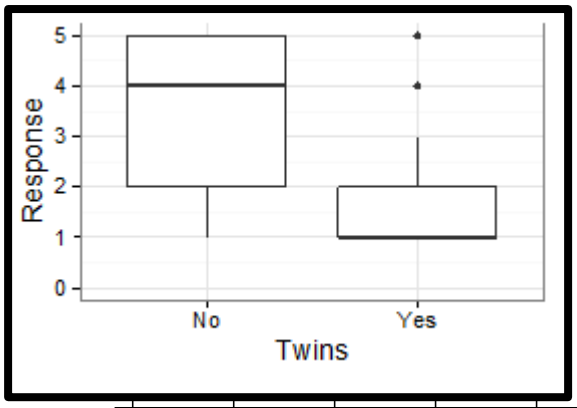
✓ Language * Reaction time



4. Results

✓ Reaction time * Twins
(language independent effects!)





✓ SMC * Twins

(language independent effects!)

smc*twins effect plot

response (probability)

response : 5 | twins : No
response : 5 | twins : Yes
response : 4 | twins : No
response : 4 | twins : Yes
response : 3 | twins : No
response : 3 | twins : Yes
response : 2 | twins : No
response : 2 | twins : Yes
response : 1 | twins : No
response : 1 | twins : Yes

0.2 0.4 0.6 0.8

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8

smc

5. Discussion

- MDS

- optimal configuration = 7D space
  - lowest possible stress value
  - confirms VQ multidimensionality (Kreiman & Sidtis 2011)
- from most similar….

	VPA settings										
Speaker	Labial	Mandibular	Lingual Tip	Lingual body	Velopharyngeal	Pharyngeal	Larynx Height	VT tension	Larynx tension	Phonation type	SMC
AGF	0	0	0	0	0	0	2	1	1	1	
SGF	0	1	0	0	1	0	2	1	1	1	
Match	1	0	1	1	0	1	1	1	1	1	*0.8*
DCT	0	2	0	2	0	2	0	2	1	1	
JCT	0	0	0	0	2	2	2	0	1	1	
Match	1	0	1	0	0	1	0	0	1	1	*0.5*

different weight?

…to least similar twins

	VPA settings										
Speaker	Labial	Mandibular	Lingual Tip	Lingual body	Velopharyngeal	Pharyngeal	Larynx Height	VT tension	Larynx tension	Phonation type	SMC
AMG	0	1	0	2	2	2	1	1	1	0	
EMG	0	1	0	2	0	2	1	1	0	0	
Match	1	1	1	1	0	1	1	1	0	1	*0.8*
ASM	1	0	0	0	1	0	2	0	1	1	
RSM	1	0	0	0	1	0	2	2	2	0	
Match	1	1	1	1	1	1	1	0	0	0	*0.7*

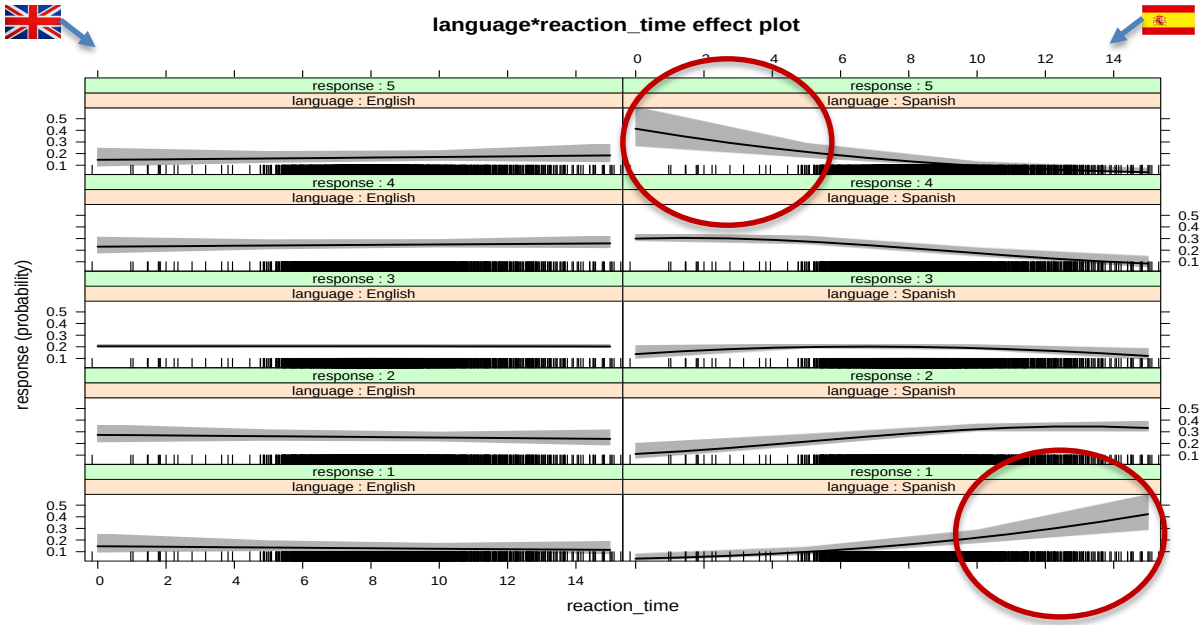
same cue prominence?

5. Discussion

- Mixed Effects Modelling

- mostly language-independent effects
  - notably: twins rated as more similar than non-twins
- …but one language-dependent effect:

mean. 0.84 s
sd: 0.18



mean. 0.82 s
sd: 0.14

6. Conclusions

- **aim** → explore the role of holistic VQ perception in speaker similarity ratings
- **results** → native ≈ non-native ratings of similarity
  - no native advantage - short stimuli + homogeneous population (same accent, similar age, etc.)
  - VQ = available resource
- **possible implications** in earwitness testimony
- **future studies**:
  - interrelationships between
    - (naïve) holistic VQ perception
    - (expert) featural VQ perception
      - → different salience
      - → weigthing methods

Thanks! Questions?

THE UNIVERSITY of York

J P French λssociates

Forensic speech and acoustics laboratory