Speaker-similarity perception of Spanish twins and non-twins by native speakers of Spanish, German and English

Eugenia San Segundo, Almut Braun, Vincent Hughes and Paul Foulkes
Dept. Language & Linguistic Science, University of York, UK

BACKGROUND & OBJECTIVE

Previous investigations

- Native listeners have an advantage over non-native listeners in speaker identification tasks [1,2]

- **But**…

✓ Identification of voices is possible when stimuli are random phonemes with no semantic meaning [3]

✓ No native language effect when f0 controlled (British English, German, Chinese, Japanese) [4]

→ So listeners seem to pay attention to cues in a voice which do not require knowledge of the speaker's language

➢ Prosodic characteristics such as **voice quality** (VQ) remain largely unexplored in this area

→ little difference between Spanish and English listeners when rating similarity of twin and non-twin pairs [5]

Interpretation

✓ Naïve listeners relied on holistic VQ perception regardless of their L1 → no native language advantage.

✓ At least with…

- … short speech samples
- … controlled conditions of speaker similarity

➢ VQ seemed to be the only resource available for judging speaker similarity.

WHAT'S NEW?

As a follow-up to [5], we have:

- widened the scope of the study to include 20 German listeners (=20 English and 20 Spanish listeners)
- added the variable *linguistic-expertise*

	SPANISH	GERMAN	ENGLISH
Linguists	8	10	15
Total	20	20	20

Table 1: Participants of the perceptual experiment

Hypothesis: differences in voice similarity ratings for twin and non-twin pairs will depend on:

✓ listeners' L1
✓ reaction time
✓ linguistic expertise

MATERIALS & METHOD

Speakers

- 5 pairs of male monozygotic twins
- native speakers of Standard Peninsular Spanish
- selected from a larger twin corpus [6]

Selection criteria:

➢ similar age (mean: 21 y.o., sd: 3.7)
➢ similar mean f0 (mean: 113 Hz, sd: 13)
➢ similar Euclidean distance (ED) between each speaker and his twin [7] to select only the most similar-sounding pairs

✓ ED based on VQ perceptual assessment using a simplified version of the Vocal Profile Analysis (VPA) scheme [5,8]

	LABIAL	MANDIBULAR	LINGUAL TIP	LINGUAL BODY	VELOPHARYNGEAL	PHARYNGEAL	LARYNX HEIGHT	VT TENSION	LARYNX TENSION	PHONATION TYPE
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

Table 2: Example of calculation of ED / SMC
e.g. mandibular setting: close (1) – neutral (0) – open (2)

Stimuli

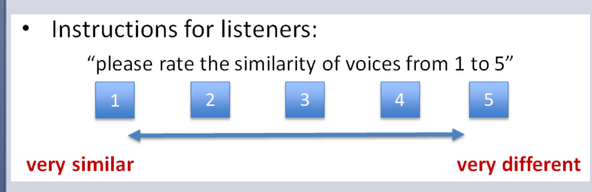
- speech samples (~ 3 seconds)
- semi-directed spontaneous conversations
- declarative sentences of different linguistic content (neutral topics)

Listeners

- native speakers of Spanish (age range 22-51, mean 33)
- native speakers of German **with no** knowledge of Spanish (age range 22-49, mean 29.3)
- native speakers of English with **no knowledge of Spanish** (age range 19-35, mean 25)

Design of perceptual test

- Praat *Multiple Forced Choice* experiment
- 90 different-speaker pairings, i.e. each speaker compared with everyone else.
- stimuli presented in random order



- listeners were not told that the stimuli included twin pairs!
- test on a PC with HQ headphones.
- short pre-test for familiarisation.

RESULTS

Ordinal mixed effects modelling

→to fit models to the similarity ratings

- predictors:

✓ *listeners' L1*
✓ *reaction time*
✓ *linguistic expertise*
+ [*whether speakers were twins or not*]

1

Regardless of **listeners' L1**, twins were judged as more similar than non-twins.

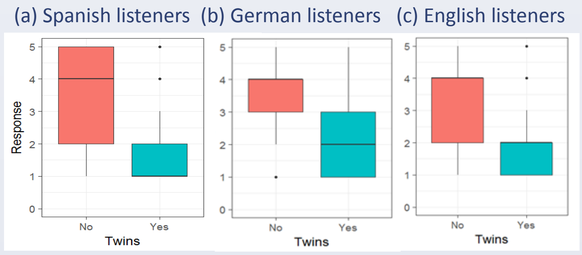


Fig. 1: Perceptual ratings as a function of speaker type.

2

Reaction time had no effect on the ratings given by English listeners.

- Conversely, both Spanish and German listeners were more likely to respond with 5 (very different) if their reaction time was short and more likely to respond with 1 (very similar) when the reaction time was long.

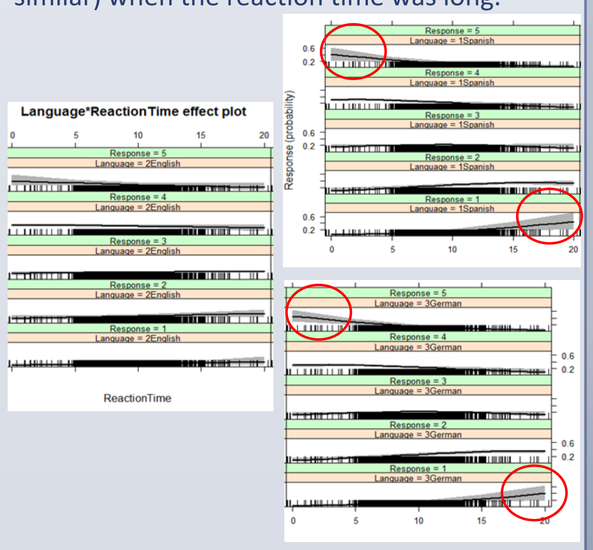


Fig. 2: Interaction of language and reaction time

3

Listeners with a degree in Linguistics made a clearer distinction between twin and non-twin pairs than listeners without a degree in Linguistics.

(See boxplots: no overlap between rating distributions in the case of linguists).

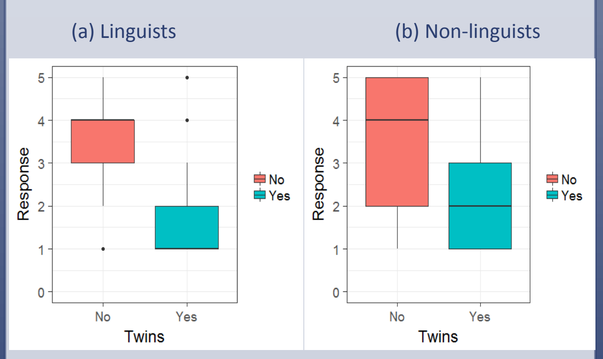


Fig. 3: Perceptual ratings as a function of speaker

DISCUSSION

Language-independent effects

1. native ≈ non-native

✓ without knowledge of Spanish, both English and German listeners rated paired voice samples of twins as sounding more similar to each other than paired voice samples of unrelated speakers

→ same as Spanish listeners
→ no native language advantage

2. linguists ≠non-linguists

✓ linguists kept twins and non-twins further apart in their similarity ratings than listeners without a degree in Linguistics

✓ spread of similarity ratings = larger within the group of non-linguists than within the group of linguists (regardless of L1)

Language-dependent effects

➢ reaction time – effect on Spanish and German listeners only!

CONCLUSIONS

1. No native language advantage

→no clear explanation for the lack of effect of reaction time in English listeners

2. Differences between linguists (experts) and non-linguists (naïve)

→ linguists might have used a more analytical strategy to make their similarity judgments while listeners without a degree in linguistics might have used a more holistic strategy

→ expert vs. naïve ratings: needs more tests regarding analytic vs. holistic approach to VQ

REFERENCES

[1] Köster, O. & N.O. Schiller (1997). Different influences of the native language of a listener on speaker recognition, *Forensic Linguistics*, 4:8-28.

[2] Perrachione, T.K., Pierrehumbert, J.B. & P.C.M. Wong (2009). Differential neural contributions to native- and foreign-language talker identification, *Journal of Experimental Psychology: Human Perception and Performance*, 35:1950-1960.

[3] Bricker, P. D., & S. Pruzansky. (1966). Effects of stimulus content and duration on talker identification, *JASA*, 40:1441-1449.

[4] Ho, C.-T. (2007). *Is pitch a language-independent factor in forensic speaker identification?*, MA diss., University of York.

[5] San Segundo, E.; Foulkes, P. & V. Hughes. (2016). Holistic perception of voice quality matters more than L1 when judging speaker similarity in short stimuli, *Proc. 16th SST Australasian Conference*. University of Western Sydney, Australia (pp. 309-312).

[6] San Segundo, E. (2014). *Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics*, PhD diss., Alicante: Biblioteca Virtual Miguel de Cervantes, 2017.

[7] San Segundo, E. & J.A. Mompeán. (2016). Voice quality similarity based on a simplified version of the Vocal Profile Analysis: A preliminary approach with Spanish speakers including identical twin pairs, *Sociolinguistics Symposium* 21, University of Murcia, Spain, 15-18 June.

[8] Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press

ACKNOWLEDGEMENTS

Thanks to Olaf Köster for initial discussions about this topic and for providing useful bibliographic references.