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

Department of Language and Linguistic Science, University of York, UK
{eugenia.sansegundo|almut.braun|vincent.hughes|paul.foulkes}@york.ac.uk

Poster

Most previous studies on speaker identification suggest that native listeners have an advantage over non-natives (Köster and Schiller 1997; Perrachione et al. 2009). Other investigations, however, suggest that it is possible to identify voices successfully when stimuli are random phonemes with no semantic meaning and not belonging to any language (Bricker and Pruzansky 1966). Therefore, listeners seem to pay attention to cues in a voice which do not require knowledge of the speaker's language, for instance suprasegmental aspects. Ho (2007) found no native language effect when comparing British English and Chinese listeners in a speaker identification task where F0 was modified. This pointed to F0 as a language-independent factor for voice identification. Other suprasegmental features, such as voice quality (VQ), remain largely unexplored in this area. San Segundo, Foulkes and Hughes (2016) found little difference between Spanish and English listeners when rating speaker similarity of twin and non-twin pairs. The results suggested that similar listening strategies operate in naïve listeners (i.e. holistic approach to voice quality) in order to judge speaker similarity, regardless of the listener's L1.

As a follow-up to San Segundo, Foulkes and Hughes (2016), in this investigation we have widened the scope of the original study to also include 20 German listeners (in addition to 20 English and 20 Spanish listeners), and we have also added the variables linguistic-phonetic expertise and musical training (Table 1).

Table 1. Participants of the perceptual experiment.

	Spanish	German	English
Degree in Linguistics	8/20	10/20	15/20
Musical training	6/20	18/20	18/20

The stimuli and the experimental design were the same as in San Segundo, Foulkes and Hughes (2016). Five pairs of male MZ twins were selected from the Twin Corpus (San Segundo 2014); all of them native speakers of Standard Peninsular Spanish. Three criteria were established in order to select only the most similar-sounding twin pairs from the corpus: (i) similar age (mean: 21, sd: 3.7); (ii) similar mean F0 (mean: 113 Hz, sd: 13 Hz); and (iii) similar Euclidean distance (EDs) between each speaker and his twin. EDs took the form of Similarity Matching Coefficients (SMCs) and were based on the perceptual assessment of their VQ using the Vocal Profile Analysis (VPA) scheme.

A Multiple Forced Choice experiment was set up in Praat with 90 different-speaker pairings, i.e. each speaker compared with everyone else. Stimuli were presented in random order and listeners had to indicate the degree of similarity of each stimuli pair on a scale 1 (very similar) to 5 (very different). Listeners were not told that the stimuli included twin pairs. Ordinal mixed effects modelling was used to fit models to the similarity ratings with the following

predictors: listener language, SMCs between the speakers, reaction time, whether speakers were twins or not, and whether the listener had a degree in Linguistics or not. The variable music was removed from the predictors because of its strong association with the variable language (6 Spanish listeners with musical training vs. 18 for German and 18 for English).

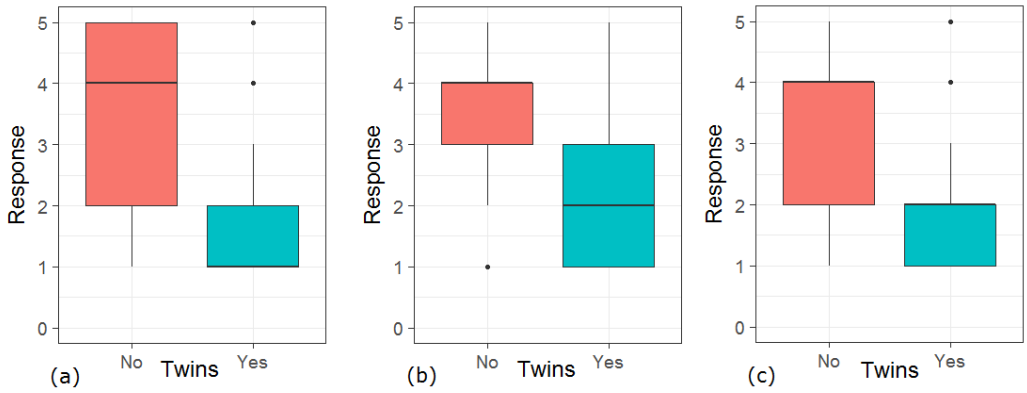

Figure 1 Perceptual ratings (from 1 to 5) as a function of speaker type (twin or non-twin pairs). Results for (a) Spanish, (b) German and (c) English listeners.

Both language-independent and language-dependent results were found. On the one hand, regardless of the listener language, twins were judged as being much more similar than non-twin pairs (Figure 1). On the other hand, the interaction between language and reaction time shows different effects depending on the language. Namely, 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 while the longer both listener groups took to respond, the more likely they were to respond with 1 (very similar).

Overall, listeners with a linguistics degree made a bigger distinction between twin and non-twin pairs. For linguists, there was no overlap of their rating distributions for twin and non-twin pairs (median rating for twins was '1' in the case of linguists vs. '2' in non-linguists). Linguists might have used a more analytical strategy to make their similarity judgments while the listeners without a degree in linguistics might have used a more holistic strategy.

References

Bricker, P. D., and S. Pruzansky. (1966). Effects of stimulus content and duration on talker identification, *Journal of the Acoustical Society of America*, **40**:1441-1449.

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

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

Perrachione, T.K., Pierrehumbert, J.B. and 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.

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. http://www.cervantesvirtual.com/nd/ark:/59851/bmcm9293

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