

## Perceptual similarity of identical twins across different L1 listeners: the importance of voice quality in Forensic Phonetics

Eugenia San Segundo<sup>1</sup>, Paul Foulkes<sup>1,2</sup>, Peter French<sup>1,2</sup>, Vincent Hughes<sup>1</sup> and Olaf Köster<sup>3</sup>

<sup>1</sup>Department of Language and Linguistic Science, University of York, UK.

<sup>2</sup>J P French Associates, York, UK.

<sup>3</sup>BKA (Federal Criminal Policy Office, Germany).

A round robin test was recently carried out by the *Bundeskriminalamt* (BKA) to evaluate the proficiency of experts performing speaker identification tasks. The test was approached mostly as an auditory evaluation since the technical characteristics of the recordings prevented most experts to carry out specific acoustic analyses. The speakers under comparison were a pair of female German twins. It is well known that these speakers challenge voice recognition because of their similarity (San Segundo 2014). This is true for automatic speaker recognition but also, from a broader phonetic perspective, in identification/recognition tasks performed by both human experts and naïve listeners. Interestingly, the BKA test suggests that the lack of native knowledge of the language spoken by the twins was not a disadvantage for telling these apart. We cannot draw strong conclusions from this test given the idiosyncrasies of the single twin pair used: advanced age and having lived in different dialectal regions. However, informal feedback gathered from the participants suggests that voice quality (VQ) – approached holistically rather than analytically– was the main cue used by non-native listeners (mainly Dutch) to distinguish the twins. This could be due to the fact that the non-native experts lacked formal training in the dialectal differences across Germany. The BKA test calls for the design of a perceptual experiment of a different nature which could shed light on how listeners of different L1s perform when assessing the voice of very similar-sounding speakers. In this study we have tested, with a larger twin sample and under controlled conditions of age and dialect, whether the different L1 of listeners affect the perceptual distances between speakers. Following McDougall (2013), the degree of perceived similarity is measured using Multidimensional Scaling (MDS).

We selected 10 speakers from the corpus collected by San Segundo (2014): 5 pairs of male monozygotic (MZ) twins, all native speakers of Standard Peninsular Spanish (SPS) with no voice or speech pathologies. The original corpus contains 54 speakers (aged 18-54), so for the selection of the 10 speakers of this experiment, some criteria were established. We controlled for similar age (mean: 21, sd: 3.7) and similar mean f0 (mean: 113 Hz; sd: 13 Hz). We also considered the Euclidean distance (ED) between each speaker and his twin in the study by San Segundo & Mompéan (2016). These EDs were based on the perceptual assessment of their VQ using a simplified version of the Vocal Profile Analysis scheme (Laver 1980). The mean EDs between twins in each pair was 0.6 (in Similarity Matching Coefficients these values could range between 0 and 1), which indicates that their VQ was quite similar: on average, at least 6 VQ settings were shared. We set up a Multiple Forced Choice perceptual experiment with 90 different-speaker pairings, i.e. each speaker is compared with everyone else except himself. The stimuli were speech samples of around three seconds, extracted from semi-directed spontaneous conversations. The 90 pairings were presented in random order to listeners, who were asked to indicate the degree of similarity of each stimuli pair on a scale 1 to 5. Two different groups of listeners were recruited to take part in this experiment: 20 native Spanish speakers and 20 native English speakers. The results of this test will potentially reveal the perceptual importance of VQ to distinguish very similar speakers. The implications of these findings will be evaluated in light of the results of the BKA test.

### References

San Segundo (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*, Doctoral dissertation, Menéndez Pelayo International University & Spanish National Research Council.

- San Segundo & 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.
- Laver, J. (1980) *The Phonetic Description of Voice Quality*. Cambridge: CUP
- McDougall, K. (2013). Assessing perceived voice similarity using Multidimensional Scaling for the construction of voice parades, *IJSL*, Vol. 20 (2): 163-172.