Oral corpus of Spanish male twins and siblings:

Forensic-phonetic application and further uses

Eugenia San Segundo Fernández

Laboratorio de Fonética, Consejo Superior de Investigaciones Científicas

The importance of studying twins has been acknowledged in very diverse scientific disciplines; not surprisingly, twin similarities have intrigued human beings across cultures and centuries. The classical twin method compares the resemblance of monozygotic (MZ) and dizygotic (DZ) twins and is widely used not only to study diseases but also to estimate the environmental and heritable factors that contribute to certain behavioral traits or "complex system of features" (Decoster et al. 2001), such as voice. While MZ twins share all their genes, DZ twins share, on average, half their genes.

Most authors (e.g. Künzel, 2011) interested in this kind of studies point out the difficulty in recruiting twins who volunteer to participate in experiments. This is the reason why, having noted a lack of oral corpora of Spanish twins' voices, we have collected a database which (so far) consists of the following male speakers: 24 MZ twins, 10 DZ twins, 8 non-twin brothers, and 12 non-twin reference-population speakers.

Thirty-three voice-related studies about twins have been reviewed in order to account for the average sample size of previous twin corpora. Their data collection method as well as the main results obtained in such studies will be presented.

Within a forensic-phonetic approach, the study of twins' voices is particularly relevant, as these (especially MZ twins) represent the most extreme physical similarity in human beings. Distinguishing them is therefore a challenge in speaker identification. The inclusion of DZ twins and siblings in forensic corpora should not be neglected, as its usefulness has been previously acknowledged, among others, by Weirich (2012) and Feiser (2009), respectively.

Previous biometric studies support the great interest aroused by twins: we find research on the discriminability of the handwriting (Srihari, Huang and Srinivasan, 2008), DNA (Rubocki et al. 2001) and fingerprints (Jain, Prabhakar and Pankanti, 2002) of twins.

As regards the corpus presented in this paper, four different speaking tasks were designed: (1) semi-structured spontaneous conversation between twin pairs, (2) information exchange task to elicit certain phonetic units, (3) reading of two text passages –with and without voice disguise-, and (4) semi-structure interview between each twin and the researcher. Furthermore, the speakers were instructed to carry out a voice control technique in order to measure their MPT (Maximum Phonation Time) and their respiratory and phonatory efficiency. All the speakers were recorded on two occasions, separated by 2-4 weeks. Their recordings are available both in high quality and degraded via telephone filtering.

We would like to highlight that the corpus that we are presenting is, to our knowledge, the first oral corpus of Spanish male twins and siblings. The main forensic-phonetic application of

these recordings will be detailed in this paper but other linguistic uses will be suggested too. Using the speech samples recorded in this corpus, some publications have already appeared (San Segundo, 2012; Gil y San Segundo, 2013) and we hope to encourage new research on twins in the field of Linguistics by means of collaborations when this corpus becomes –upon request– an open resource available for scientists.

References

Decoster, W., Van Gysel, A., Vercammen, J. and Debruyne, F. (2001) Voice similarity in identical twins, *Acta oto-rhino-laryngologica belg*., 55: 49-55.

Feiser, H.S. (2009) Acoustic similarities and differences in the voices of same-sex siblings. In *Proc. 18th Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA)* 2009, Cambridge, UK.

Gil, J. and San Segundo, E. (2013) El disimulo de la cualidad de voz en fonética judicial: un estudio perceptivo para un caso de hiponasalidad, en A. Palacios (Ed.) Panorama de la Fonética Española Actual (In Press).

Jain, A., Prabhakar, S., Pankanti, S. (2002) On the similarity of identical twin fingerprints. Pattern Recognition, 35(1): 2653-63.

Künzel, H. (2010). Automatic Speaker Recognition of Identical Twins, *International Journal of Speech Language and the Law*, 17, 2, 251-277.

Rubocki, R., McCue, B., Duffy, K., Shepard, K., Shepherd, S., and Wisecarver, J. Natural DNA mixtures generated in fraternal twins in utero, Journal of Forensic Science, 46(1): 120-5.

San Segundo, E. (2012) Glottal source parameters for forensic voice comparison: An approach to voice quality in twins voices, *Proc. 21th Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA)* 2012, Santander, Spain.

Srihari, S., Huang, C. and Srinivasan, H. (2008) On the Discriminability of the Handwriting of Twins, Journal of Forensic Science, 53, 2: 430-446.

Weirich, M. (2012) The influence of NATURE and NURTURE on speaker-specific parameters in twins' speech: Articulation, acoustics and perception. PhD thesis. Humboldt-Universität zu Berlin.