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.

Eugenia San Segundo Fernández

University of York

The similarity of twin pairs, whether dizygotic (DZ) or monozygotic (MZ), has attracted the attention of many scientists over the years and across sundry disciplines; Speech Sciences are not an exception. Due to the multidisciplinary nature of the study of voice itself, state-of-the-art research on twins' voices has benefitted from the enriching approaches of –among others– speech therapists and otoralyngologists (e.g. Decoster et al., 2000; Cielo, Agustini and Finger, 2012), acousticians, computer scientists and engineers (e.g. Homayounpour and Chollet, 1995; Scheffer et al. 2004; Ariyaeenia et al. 2008), as well as linguists and phoneticians (e.g. Nolan and Oh, 1996; Loakes, 2006; Künzel, 2010, Weirich, 2011). The literature review carried out for this thesis includes around 40 studies dealing with voice (dis)similarities in twins and non-twin siblings, where a special focus has been placed on their forensic application. As a result of this review, some conclusions can be drawn; for instance: (1) the main objectives of the investigations are either to find a genetic component in the variation of certain voice parameters by searching differences between MZ and DZ twin pairs, or testing whether it is possible to identify twin speakers by distinguishing them from their respective cotwins; (2) the average number of speakers recruited for this type of studies is quite low; large numbers of participants (above 60) are found only exceptionally and mainly thanks to the prior existence of a twin register or corpus; (3) only a minority of studies takes into account forensically realistic conditions such as channel mismatch, non-contemporaneous speech samples or different speaking styles, including spontaneous speech; and (4) few studies focus on non-twin siblings. To our knowledge, this thesis represents the first study considering both types of twins (MZ and DZ) together with non-twin siblings, as well as the first investigation into the phonetic and acoustic characteristics of this type of speakers in Standard Peninsular Spanish or North-Central peninsular variety.

On the other hand, an overview of the main methodologies in Forensic Speaker Comparison (FSC) nowadays (Cambier-Langeveld, 2007; Gold and French, 2011; French and Stevens, 2013) reveals the lack of consensus over the comparison methods and most frequently analyzed parameters in this discipline. This lack of methodological agreement may have positively triggered the constant search for new voice parameters –including the fusion and combination of them– which could allow for robust speaker comparison, provided that they fulfill certain criteria (e.g. Wolf, 1972; Nolan, 1983). As a result of this review of the state of the art in FSC, it was considered that this thesis should adopt a hybrid perspective which combined traditional phonetic-acoustic parameters with features which are characteristic of automatic methods (e.g. cepstral coefficients); this

comprehensive and combined approach being recommended for instance in Künzel (2011) and already suggested in Rose (2002) or Künzel and González-Rodríguez (2003), among other authors. In an attempt to place this investigation within current likelihood-ratio approaches to FSC, the results of the three different analyses that have been carried out are expressed in Likelihood Ratios (LRs).

For the 54 male speakers recorded *ad hoc* for this study (24 MZ twins, 10 DZ twins, 8 non-twin siblings and 12 unrelated speakers) on two different occasions (2-4 week time lapse), a three-folded methodological approach has been taken. On the one hand, I labeled and analyzed the F1-F3 formant trajectories of 19 Spanish vocalic sequences (VS), comprising both diphthongs and hiatuses, i.e, all the possible combinations of two different vowels in this language, except /ou̯/ for being considered a rare diphthong in Spanish (Aguilar, 2010). Secondly, several naturally sustained [eː] tokens were extracted from the speakers' spontaneous vowel fillers –or hesitation markers– and their glottal source characteristics were analyzed. More accurately, a vector of 68 different glottal parameters was created using *BioMet®Soft* (e.g. Gómez et al., 2014) with features ranging from classical perturbation parameters such as jitter and shimmer to cepstral coefficients estimated from the glottal source power spectral density (PSD). The other parameters could be classified in either singularities of the glottal source PSD, biomechanical estimates of the vocal fold mass, tension and losses, time-based glottal source coefficients, glottal gap coefficients or tremor coefficients. The formant-dynamic approach and the glottal-source perspective were complemented with an automatic speaker recognition analysis carried out with the software *Batvox*. For this third approach, speech fragments of around 120 seconds of net speech were extracted (per speaker and recording session) from an informal interview with the researcher.

The objective of this thesis has been to investigate the phonetic-acoustic similarities and differences in three main speaker groups: monozygotic (MZ) twins, dizygotic (DZ) twins and non-twin siblings. From a forensic-phonetic perspective, the study of this type of speakers is very relevant, as they represent extreme examples of physical similarity. Distinguishing their voices poses a well-recognized challenge in the forensic realm. Yet, there is an interest in this investigation *per se*, as the study of genetically identical speakers (MZ twins) and their comparison with non-genetically-identical siblings (DZ twins and non-twin siblings), on the one hand, and with a reference population of unrelated speakers, on the other hand, has been proposed as a useful method for gaining insight into the contribution of nature (genes) and nurture (education, environmental influences) in a speaker's voice characteristics.

As far as the results are concerned, all the parameters tested for this investigation proved to be genetically conditioned since the hypothesized decreasing scale *MZ > DZ > non-twin siblings > unrelated speakers* was observed in the comparison values resulting from the three types of analyses. The rare discordant results –for specific intra-pair comparisons– were thoroughly discussed and could be explained in a great extent by physical causes such as the presence of certain pathologies or by the existence of divergences created by the twins in an attempt at differentiation from their cotwins (Segal, 1990). It seems then that the intratwin mimetism and accommodation found for some twins coexist with the voluntary tendency towards diverging in other twins. While these sociolinguistic aspects deserve further investigation, the fact that the matching scores in

MZ comparisons were in average higher than in DZ comparisons, and that these in turn yielded higher matching scores than non-twin siblings, whose matching results were above those obtained by unrelated speakers suggests that the analyzed parameters are genetically influenced. A voice characteristic which depends largely on the genetic endowment of the individual is expected to be robust for its use in a typical FSC scenario, because it will exhibit large between-speaker variation while remaining as consistent as possible for each speaker. Other findings of this investigation are specific for each of the analysis approaches. For instance, out of the two parametric procedures used for the curve fitting of the VS formant trajectories (polynomial functions and DCT functions), I found no outperformance of one method over the other when comparing their accuracy. Yet, cubic polynomials and third-degree DCT functions were found to better correlate with the original formant trajectories than their second-degree counterparts. In relation to the glottal-source analysis, its main potential lies in the fact that certain features of the 68 feature-vector seem very speaker-specific even if difficult to relate to physical characteristics while others would be more explanatory or illustrative before a court although exhibiting less discriminatory potential. Yet, this seems in agreement with the classical trade-off between automatic and traditional features already described by Rose (2006).

References

Aguilar, L. (2010). *Vocales en grupo*. Madrid: Arco/Libros.

Ariyaeeinia, A., Morrison, C., Malegaonkar, A., & Black, S. (2008). A test of the effectiveness of speaker verification for differentiating between identical twins. *Science & Justice*, 48(4), 182-186.

Cambier-Langeveld, T. (2007). Current methods in forensic speaker identification: Results of a collaborative exercise. *International Journal of Speech, Language and the Law, 14*(2), 223-243.

Cielo, C., Agustini, R., & Finger, L. (2012). Vocal features of monozygotic twins. *Revista CEFAC, 14*(6), 1234-1241.

Decoster, W., Van Gysel, A., Vercammen, J., & Debruyne, F. (2000). Voice similarity in identical twins. *Acta Oto-Rhino-Laryngologica Belgica, 55*(1), 49-55.

French, P., & Stevens, L. (2013). Forensic speech science. In M. Jones & R. Knight, *The Bloomsbury companion to phonetics* (pp. 183-197). London: Bloomsbury.

Gold, E., & French, P. (2011). An international investigation of forensic speaker comparison practices. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, China (pp. 1254-1257).

Gómez, P., San Segundo, E. Mazaira, L.M., Álvarez, A., & Rodellar, V. (2014). Using dysphonic voice to characterize speaker's biometry. *Language and Law / Linguagem e Direito*, Vol. 1(2), 42-66.

Homayounpour, M., & Chollet, G. (1995). Discrimination of voices of twins and siblings for speaker verification. In *Proceedings of Eurospeech* (pp. 345-348).

Künzel, H. J. (2010). Automatic speaker recognition of identical twins. *International Journal of Speech, Language and the Law, 17*(2), 251-277.

Künzel, H. J. (2011). La prueba de voz en la investigación criminalística. *Ciencia Forense, INACIPE-Academia Iberoamericana de Criminalística y Estudios Forenses*, *1*(1), 37-50.

Künzel, H.J., & González-Rodríguez, J. (2003). Combining automatic and phonetic-acoustic speaker recognition techniques for forensic applications. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1619-1622).

Loakes, D. (2006). *A forensic phonetic investigation into the speech patterns of identical and non-identical twins* (Doctoral dissertation). University of Melbourne.

Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Nolan, F., & Oh, T. (1996). Identical twins, different voices. *International Journal of Speech Language and the Law, 3*(1), 39-49.

Rose, P. (2002). *Forensic speaker identification*. London: Taylor & Francis.

Rose, P. (2006a). Technical forensic speaker recognition: Evaluation, types and testing of evidence. *Computer Speech & Language*, *20*(2), 159-191.

Scheffer, N., Bonastre, J., Ghio, A., & Teston, B. (2004). Gémellité et reconnaissance automatique du locuteur. *Actes des Journées d'Étude sur la Parole (JEP)*, 445-448.

Segal, N. (1990). The importance of twin studies for individual differences research. *Journal of Counseling & Development*, 68(6), 612-622.

Weirich, M. (2011). *The influence of NATURE and NURTURE on speaker-specific parameters in twins' speech: Articulation, acoustics and perception.* (Doctoral dissertation). Humbold-Universität zu Berlin.

Wolf, J. (1972). Efficient acoustic parameters for speaker recognition. *The Journal of the Acoustical Society of America*, *51*(6B), 2044-2056.