Italian monozygotic twins' speech: a preliminary forensic investigation

Alice Albanesi*, Sonia Cenceschi**, Chiara Meluzzi***, Alessandro Trivillini**

*Institute of Forensic Sciences, Milano, Italy

** Digital Forensic Service, DTI, University of Applied Sciences and Arts of Southern Switzerland

*** University of Pavia, Pavia, Italy

The main purpose of this preliminary study is to investigate whether it is possible to distinguish a speaker from his co-twin in compressed audio recordings. The twins speech similarity degree depends on the changing sum of an anatomical inheritance with environmental and social factors which contribute to sculpt their personality [1]. However, despite the debate on these factors, the monozygotic twins' voices represent a crucial point in forensics: beyond the practical cases involving siblings, they represent the most extreme physical similarity between two different speakers, and consequently the very lowest limits of between-speaker variation [2-3]. As a consequence, they are an excellent starting point to study a number of key topics in forensics such as the acoustic features influencing speaker's recognition accuracy, or the auditory discrimination of voices.

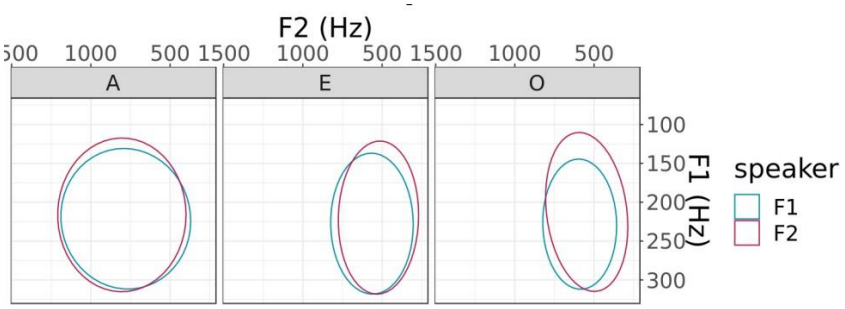
A rich international scientific literature exists [e.g. 4—8] but it tends to drastically decrease for Romance languages, where *more forensic corpora should be created, for example considering low quality recordings* [3]. Despite the presence of several Italian speech corpora [e.g. 9—11], nothing available seems to exist concerning twins' speech. A few investigations focus on the Italian twins perceptual acoustic discrimination [e.g. 12-13] but their results are still limited due to the low number of samples and because the results do not deal with the forensic perspective.

This first explorative work aims at investigating the speech of a small selected corpus of twins' speech. The first question to be addressed is whether and according to which variables Italian twins speech could be distinguished. Secondly, whether and to what extent these (dis)similarities are reduced or enhanced in controlled vs. spontaneous speech.

This preliminary study is conducted on a small corpus consisting in 6 pairs of monozygotic twins (3M and 3F), with the same sociolinguistic characteristics (high level of education, born and living in the north-west of Italy, L1 Italian) aged between 20 and 25 y.o. Each speaker was asked to record a list of 31 elicited sentences [14] and undergo a short interview. The similarity of prosodic profile was aimed at reducing the involved variables as much as possible to ensure repeatability, and to lay the ground for future studies.

In forensic contexts, audio files are often collected with different technologies, and are typically characterized by very low quality because of difficult and noisy recording conditions (e.g. recording with micro recorder in a restaurant), narrow frequency bands, and compressed formats (e.g. telephone interception). In order to mimic a forensic setting, we recorded with four different smartphone, creating m4a files (48-44 kHz, 16 bit), later converted to *.wav 44.1 kHz 16 bit for Praat processing (spectral band cut at 15kHz). Recordings have been realized in a silent room, but without specific details regarding environmental soundproofing. Vowels /a/, /e/ and /o/ were manually annotated on a PRAAT tier (both in sentences and interviews). In order to emulate as closely as possible forensic conditions (short or partially unusable audio), all possible vowels were taken [15] in order to start observing their overall variability on each task. For setting the vowel's left and right boundaries we based on the beginning and end of the second formant (not selecting if it wasn't clear to listening) obtaining about 440 vowels for typology on each task. After the annotation, we automatically extracted the following acoustic parameters at the midpoint using a Praat script: F0, F1, F2 and F3 of the target vowels. Then, formants values have been visually inspected through the web application Visible Vowels [16] as suggested in [17], with the statistical analysis carried out with IBM SPSS 20.

A first analysis shows that all pairs present almost identical distributions for all vowels with slight differences in spontaneous speech. F0 and F3 are really similar, with minimal oscillation that can be explained by expressive variations. The figure shows an example of strong similarities in the F1-F2 unnormalized distribution for the first female pair in controlled speech. In spontaneous speech the distributions remain displaced in the same way, with a very slight shift of the lower F2 boundary towards the low frequencies for /e/ and /o/.



In the following months we aim at expanding the corpus and integrating our analysis with other acoustic parameters such as jitter and shimmer, but also dynamic variation of F2, since the second formant has proved to be a robust indicator of between-speaker variation in sociophonetics as well as in clinical phonetic research [18]. Moreover, vowels observation will be refined in order to observe in detail the proximity to certain consonants and their dynamic behavior (over 7 time steps). A parallel investigation concerns the perceptual aspect: once the analysis of the corpus has been completed, a perceptual test will be prepared to understand whether the hearing system validates (or not) the results.

References

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

[2] Loakes, D. (2008). A forensic phonetic investigation into the speech patterns of identical and non-identical twins. International Journal of Speech, Language and the Law, 15(1), 97-100.

[3] Fernández, E. S. S. (2013). A phonetic corpus of Spanish male twins and siblings: Corpus design and forensic application. Procedia-Social and Behavioral Sciences, 95, 59-67.

[4] van Braak, P., & Heeren, W. F. L. (2015). "Who's calling, please?" Is there speaker-specific information in twins' vowels?, Bachelor's thesis.

[5] Johnson, K., & Azara, M. (2000). The perception of personal identity in speech: Evidence from the perception of twins' speech. Unpublished manuscript.

[6] Sebastian, S. (2013). An investigation into the voice of identical twins. Otolaryngology online journal, 3(2), 9-15.

[7] Zuo, D., & Mok, P. P. K. (2015). Formant dynamics of bilingual identical twins. Journal of Phonetics, 52, 1-12.

[8] Van, W. G., Vercammen, J., & Debruyne, F. (2001). Voice similarity in identical twins. Acta oto-rhino-laryngologica Belgica, 55(1), 49-55. PERCEZIONE

[9] Falcone, M., & Barone, A. (2000). FOCUS: un corpus vocale di voci simili per lo studio della identificazione del parlatore in ambito forense. 28 Convegno AIA.

[10] Falcone, M., & Gallo, A. (1996, October). The" siva" speech database for speaker verification: Description and evaluation. In Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96 (Vol. 3, pp. 1902-1905). IEEE.

[11] Cresti, E., & Moneglia, M. (Eds.). (2005). C-ORAL-ROM: integrated reference corpora for spoken romance languages (Vol. 15). John Benjamins Publishing.

[12] Giannini, A. (1989). Test acustico-percettivo su voci di gemelli. In atti del xvii convegno nazionale aia (pp. 427-432).

[13] Gedda, L., Bianchi, A., & Bianchi-Neroni, L. (1958). La Voce dei Gemelli—I. Prova di identificazione intrageminale della voce in 104 coppie (58 MZ e 46 DZ). Acta geneticae medicae et gemellologiae: twin research, 4(2), 121-130.

[14] Prieto, P., Borràs-Comes, J., & Roseano, P. (Coords.) (2010-2014). Interactive Atlas of Romance Intonation. Web page: <http://prosodia.upf.edu/iari/>.

[15] Rhodes, R. W. (2012). Assessing the strength of non-contemporaneous forensic speech evidence (Doctoral dissertation, University of York) p.121.

[16] Heeringa, W., van de Velde, H. (2018). Visible Vowels: a Tool for the Visualization of Vowel Variation. In Proceedings CLARIN Annual Conference 2018, 8 - 10 October, Pisa, Italy.

[17] Cenceschi, S., Meluzzi, C. (2020). The variability of vowels' formants in forensic speech. In IEEE Instrumentation & Measurement Magazine, Vol.24 (in press).

[18] Meluzzi, C. (2021, to appear). Sound Spectrography, in M.J. Ball (ed.) Manual of Clinical Phonetics, London: Routledge, pp. 418-443.