Prosody can help distinguish identical twins: implications for forensic speaker comparison

Eugenia San Segundo[1], *Lei He*[2], *and Volker Dellwo*[2]
[1]*Department of Language and Linguistic Science, University of York, UK.*
eugenia.sansegundo@york.ac.uk
[2]*Institute of Computational Linguistics, University of Zurich, Switzerland.*
{lei.he|volker.dellwo}@uzh.ch

Poster presentation

Various researchers have shown an interest in the voice similarity of identical twins. However, results across studies are hardly comparable since the number of speakers, gender, speaking style and, most importantly, forensic comparison methods tend to differ. Therefore, it is difficult to assess the relative importance of different systems or the value of a set of acoustic features over others for identification purposes. Exceptionally we find studies, such as Künzel (2010) and San Segundo and Künzel (2015), which use the same ASR system, speaking task, sample duration and forensic output (EER). This facilitates the comparison of results: in this case system performance with German and Spanish twins. However, most studies are characterised by its heterogeneity in research design. With non-twin investigations, a similar situation occurs. For instance, the DyViS corpus has been extensively used in forensic research but the specific corpus task may differ across studies, different number of speakers may be selected for analysis, or the samples may present different quality (high quality vs. telephone filtered). This makes cross-study comparisons difficult.

The Twin Corpus (San Segundo, 2014) presents the advantage that most investigations have focused on Task 5: semi-directed spontaneous conversation between researcher and twins. Although the methodological approaches differ (Table 1), the same twin pairs were always considered, in high quality recordings. This makes new approaches to twins more easily comparable with results from previous methodologies, even though the expression of conclusions is different. This remains the main comparison challenge.

The new approach to twin research that we propose here is based on temporal parameters. We will analyze a range of rhythmic metrics related to the variability and proportion of duration between consonant and vocalic segments (Dellwo et al., 2015), as well as several syllabic measures such as intensity differences between consecutive syllables that have previously shown to play an important role in between-speaker rhythmic differences (He and Dellwo, 2016). The potential of these prosodic measures is promising, as they cover suprasegmental aspects of speaker idiosyncrasy that are more or less independent of traditional acoustic features such as formant frequencies. Anatomically, identical twins are so similar (e.g. their resonance cavities and vocal folds physiognomy) that previously tested systems based on spectral and glottal characteristics sometimes failed to distinguish them. If prosodic features could prove useful to tell twins apart when other systems fail, their inclusion in hybrid forensic comparison systems would be highly recommended. Further investigations into how to combine the output of these systems with LRs or scores yielded by ASR systems should then be necessary too.

Table 1. Investigations using Task 5 of the Twin Corpus. Results per twin pair: [1]San Segundo and Künzel (2015); [2]San Segundo and Gómez-Vilda (2014), [3]San Segundo and Mompeán (2017). Gray-shaded cells for misidentified twins by each system.

Methodological approach (output)			
Twin Pair	MFCCs[1] (scores)	Glottal features[2] (LLRs)	Vocal Profile Analysis[3] (Euclidean distances)
01	2.59	-0.1	0.8
02	2.65	-1.0	0.7
03	3.45	5.8	0.8
04	3.79	0	0.4
05	3.53	0.2	0.5
06	3.20	0.6	0.7
07	2.31	12.1	0.6
08	3.54	9.9	0.8
09	2.66	12.6	0.5
10	0.64	2.9	0.6
11	4.93	-1.5	1
12	1.34	-14.6	0.3

References

Dellwo, V.; Leemann, A.; Kolly, M-J. (2015). Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *J. Acoust Soc Am.*, **137**(3):1513-1528.

He, L.; Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech, Language and the Law,* **23**: 245-275.

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

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* [Doctoral dissertation]. Menéndez Pelayo International University & Spanish National Research Council.

San Segundo, E.; Gómez-Vilda, P. (2014). Evaluating the forensic importance of glottal source features through the voice analysis of twins and non-twin siblings. *Language and Law/Linguagem e Direito,* **1**:22-41.

San Segundo, E.; Künzel, H. J. (2015). Automatic speaker recognition of Spanish siblings: (monozygotic and dizygotic) twins and non-twin brothers. *Loquens*, **2**(2):e021.

San Segundo, E.; Mompeán, J.A. (2017). A Simplified Vocal Profile Analysis Protocol for the Assessment of Voice Quality and Speaker Similarity. *Journal of Voice*. DOI: http://dx.doi.org/10.1016/j.jvoice.2017.01.005