“Who’s calling, please?” Is there speaker-specific information in twins’ vowels?

Pieternel van Braak and Willemijn Heeren
Department of Languages, Literature and Communication
Utrecht University, Utrecht, The Netherlands
p.c.vanbraak@students.uu.nl; w.f.l.heeren@uu.nl

An anecdote: the father of a monozygotic twin pair at times seems unable to identify his adult twin daughters over the phone, even though identification in direct communication yields no problems. Since this is an anecdote, there is no evidence that acoustic differences indeed crucially contribute to the perceptual difference between transmission channels. However, anatomical and environmental similarities between twins, in combination with the filtering of telephone transmission, may influence the speech signal in such a way that the father cannot distinguish between his twin daughters. We explored if transmission channel characteristics of direct versus telephone recording influence how well a monozygotic twin pair is identified by listeners who know them well. We addressed this question by means of a small-scale perception test in which a third, non-related speaker was also included, to compare speaker confusion between twin and non-twin speakers.

Assuming that auditory speaker identification is possible over both transmission channels, it is expected that speaker-specific information is present in the speech. However, since monozygotic twins are characterized by maximum anatomical similarity and maximally equal environmental experience (cf. San Segundo Fernández, 2014), differences in acoustic information may be small. For this reason, twin speech was deemed suited for testing the distinctive power of speaker comparison parameters. We focused on vocalic content by including both long-term formants (LTFs) across vowels and (long-term) formants by vowel. LTFs are proposed as a useful measure in speaker comparison studies (cf. Nolan & Grigoras, 2005; Moos, 2010; Gold, French & Harrison, 2013, a.o.). Due to the assumed minimum inter-speaker variability between monozygotic twins and the generalizing nature of LTFs, the latter may not distinguish between such speakers. Given that several studies suggest that certain categories of vowels might contain more speaker-specific information than the vowel inventory as a whole (e.g. Loakes, 2004; Dukiewicz, 1970; Pickett, 2003), we also examined the (relative) positions of speakers’ individual vowels in the vowel space.

As materials for analysis and testing, speech from four demographically comparable, female native speakers of Dutch was recorded (one monozygotic twin pair, two unrelated speakers), aged 20 to 25. Each speaker twice read a standard text in Dutch containing all vowels of the language, and speech was recorded directly as well as over the telephone. The two samples of read speech also allowed for within-speaker comparisons. By speaker, all vocalic intervals were identified and labeled, and formant measurements were taken to compute LTFs and LTFDs by-vowel and across vowels.

Perception results showed that confusion was higher in telephone than in studio speech, but contrary to expectations, confusion was not higher between twin than non-twin speakers. Moreover, confusion differed between the two non-twin speaker pairs. Acoustic results suggest that LTFDs between twins were not more comparable than between non-twin pairs of speakers. In contrast with earlier research, front vowels did not more clearly differentiate between speakers, but particular other vowels in the current data set did: /e:, i:, u/. The twins’ vowel spaces were especially comparable for half-open front to central vowels /e:, ɛ, ɪ, ɵ, ə/.

References

Dukiewicz, L. (1970). Frequency band dependence of speaker. In Jassem, W. (ed.), Speech Analysis and Synthesis, Vol.II (pp. 41-50). Warsaw: Polish Academy of Sciences.

Gold, E., French, P. and Harrison, P. (2013). Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework. Proceedings of Meetings on Acoustics, 19, pp. 1-7.

Loakes, D. (2004). Front vowels as speaker-specific: Some evidence from Australian English. In Proceedings of the Australian International Conference on Speech Science, pp. 289-294.

Moos, A. (2010). Long-term formant distribution as a measure of speaker characteristics in read and spontaneous speech. The Phonetician, 101, pp. 7-24.

Nolan, F. and Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. Journal of Speech, Language and the Law, 12, pp. 143-173.

Pickett, C. (2003). Forensic speaker identification: An acoustic phonetic study. Unpublished first class honours thesis: Linguistics program, Monash University.

San Segundo Fernández, 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 Thesis, UIMP.