Glottal source parameters for forensic voice comparison: An approach to voice quality in twins' voices

Eugenia San Segundo Fernández
Phonetics Laboratory, Consejo Superior de Investigaciones Científicas, Madrid, Spain
eugenia.sansegundo@cchs.csic.es

Background

Research on twins' voices has been often acknowledged in forensic phonetic studies (Nolan and Oh 1996, Loakes 2006, Künzel 2010) as being useful to test to what extent a forensic parameter is robust to distinguish between very similar speakers. Identical twins (monozygotic, MZ) share the 100% of their genes while fraternal twins (dizygotic, DZ) share half their genetic information. The nature-nurture dichotomy –that is to say, how genes and environment influence in all areas of people living– has usually been referred to in Phonetics as the organic vs. learned dichotomy (Wolf 1972) and the validity of it has been called into question. Thus, Nolan (1997) states that the organic-learned dichotomy fails to show the true complexity of speaker individuality because in phonetic features, like formant frequencies, both the vocal tract and dialectal vowel differences manifest and overlap. Besides, there is no agreement as to in which kind of twin pairs (MZ or DZ twins) there is less variation. As regards physiological aspects (jaw and teeth size, position, etc.), variations are smaller in MZ twins than in DZ twins (Lundström 1948). However, there are studies (Stromswold 2006) claiming that due to perinatal hardships, more often experienced by MZ twins than by DZ, the former will tend to be more linguistically different from one another than the latter.

Objectives

The study reported here is part of a larger project which will use a combined method to analyse twins' voices, comprising: analysis of glottal source parameters, analysis of formant-frequency trajectories in vocalic sequences and automatic analysis using BATVOX system. In this communication we will present results from the first perspective, analysing glottal source features like jitter, shimmer and biomechanical estimates of vocal fold mass, stiffness, losses and unbalance.

Materials and method

20 Spanish male twins (6 MZ pairs and 4 MZ pairs, aged 17-36) were recorded on two occasions, separated by 2-4 weeks. Each recording session comprised 4 speaking tasks (semi-structured spontaneous conversation between twin pairs, information exchange task to elicit certain phonetic units, reading of two passages, and semi-structure interview of each twin with the researcher). The data for the voice quality analysis in this study were extracted from the last speaking task. It was conducted over the telephone and it consisted on each twin holding a conversation with the researcher about what they recalled from their conversation with his brother in the first task.

Figure 1. Identical twin pair in separate quite rooms holding a telephone conversation (task one) and recorded with flat-frequency response omnidirectional ear-mounted microphones.

In order to extract voice source properties we have used the software package GLOTTEX® (Gómez-Vilda *et al*. 2008). The main advantages of GLOTTEX® are (1) that it is capable of splitting vocal from glottal information, which enables richer glottal source analysis, and (2) that, although originally developed for voice pathology detection, this tool works as a non-invasive method based on inverse filtering of the acoustic signal generated by the speaker's vocal tract.

Cross-validated likelihood ratios were calculated with a reference population of 32 Spanish male speakers using the multivariate kernel density formula described in Aitken & Lucy (2004) and implemented in Morrison (2007). The variables entered into the formula were the glottal features extracted from naturally sustained [e] in hesitation speech (speaking task 4: interview).

Results and discussion

Preliminary results show that the biomechanical estimates of both the vocal fold body and the vocal fold cover are forensically useful. In view of the results obtained, the validity of eliciting hesitation speech (e.g. pause fillers) for forensic research can be discussed as regards the necessity of having recordings with relatively long vowels in order to obtain enough glottal cycles for the voice source features extraction.

Elicitation of hesitation speech is possible in the interview task since speakers have to do an effort to remember what they had been talking to his twin at the beginning of the recording (more than an hour of time span between tasks). Since the total duration of each recording session (comprising several short tasks and instructions) was longer than 90 minutes, and due to the fact that the researcher was never present at the recording place, we consider that an effort has been made to avoid the "observer's paradox" and to ensure that by the end of the recording session the speakers had settled down to a pattern approximating their everyday interactional style.

High-quality recordings have been used in this study. However, experimental work is ongoing to investigate voice source features in very similar speakers with recordings having been channel-degraded (both GSM and landline telephone). Furthermore, we contemplate the recording of a total of 10 MZ twin pairs and 10 DZ twin pairs besides the collection of a larger reference population.

Acknowledgement

This project is supported by a grant awarded by the *IAFPA.*

References

Aitken, C.G.G. and D. Lucy. (2004). Evaluation of trace evidence in the form of multivariate data. *Applied Statistics.,* **54**, 109–122.

Gómez-Vilda, P., A. Álvarez, L.M. Mazaira, R. Fernández-Baillo, V. Nieto, R. Martínez, C. Muñoz and V. Rodellar. (2008). Decoupling vocal tract from glottal source estimates in speaker's identification. *Language Design (Special Issue)*, 111–118.

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

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

Lundström, A. (1948). *Tooth size and occlusion in twins*. Basel: Karger.

Morrison, G.S. (2007). Matlab implementation of Aitken & Lucy's (2004) forensic likelihood-ratio software using multivariate-kernel-density estimation. [Software]

Nolan, F. (1997). Speaker recognition and forensic phonetics. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences,* 744–767. Oxford: Blackwell.

Nolan, F. and T. Oh (1996). Identical twins, different voices. *Forensic Linguistics,* **3**, 39–49.

Stromswold, K. (2006). Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition,* **101**, 333–384.

Wolf, J. (1972). Efficient acoustic parameters for speaker recognition. *J. Acoust. Soc. Am., **51**, 2044–2056*.