Voice Biometrical Match of Twin and non-Twin Siblings

Eugenia SanSegundo[1], Pedro Gómez-Vilda[2]

[1]Phonetics Lab., Inst. of Lang. Lit. and Anthrop., Spanish National Research Council (CSIC)
C/ Albasanz 26-28, 28037 Madrid, Spain
[1]NeuVox Laboratory, Center for Biomedical Technology, Universidad Politécnica de Madrid,
Campus de Montegancedo, s/n, 28223 Pozuelo de Alarcón, Madrid
[2]Escuela Superior de Canto, C/ San Bernardo 44, 28015, Madrid
e-mails: eugeniasansegundo@gmail.com,pedro@pino.datsi.fi.upm.es

Abstract. The similarity in twins' voices has been always an intriguing issue in forensic speech matching, but has become a serious matter of research only recently. The present work is a preliminary study of exploratory character describing the similarities of monozygotic and dizygotic phonation under the point of view of vocal fold biomechanics, in contrast to other siblings' speech and unrelated speakers. Estimates of biomechanical parameters obtained from vowel fillers are used to produce bilateral matches between MZ and DZ twins and siblings, and unrelated speakers. These results show interesting relationships regarding genetic load and ambient factors in the adoption of phonation styles.

Keywords: voice production, forensic pattern match, phonation styles, glottal source features, voice quality.

1 Introduction

Recent studies in voice quality are conducted towards the evaluation of phonation performance in relation to either professional voice care, or in meta-acoustic knowledge (neurological deterioration, emotion detection, etc.) These fields of study are becoming more and more demanded nowadays. The aim of the present work is to study the similarities and differences of phonation characteristics in twins' voices, including monozygotic (MZ) as well as dizygotic (DZ) twins. A reference to previous work on twin voice quality analysis and vocal performance of interest for this research is that of Van Lierde et al. 1. The quality measurements used were perceptual GRBAS, breathing performance, fundamental frequency, jitter and shimmer, and the Dysphonia Severity Index (linear combination of highest pitch, lowest loudness, max. phonation time and rel. jitter). However, the study focused only on monozygotic siblings (MZ). Another relevant reference is that of Cielo et al. 2. Although the twin sample used is quite small (2 MZ pairs, one per gender) their analysis is interesting as far as they tackle some features not been considered in twins' voice studies before, namely vocal onset and harmonic characterization. While the results for maximum phonation time showed significant differences between twins, no such differentiation was found regarding vocal onset, fundamental frequency or intensity. The work of

Fuchs et al. 3 found that the voices of MZ twins showed more similarity among themselves than those of non-similar speakers regarding vocal range, highest and lowest fundamental frequency, prosodic pitch line, maximum intensity, number of overtones and intensity vibrato. The present work is intended to include biomechanical marks of relevance in the biometrical description of phonation 4. The working hypothesis is that phonation cycle quotients and biomechanics may offer differentiation capabilities among MZ, DZ and control speakers not explored already. The paper is organized as follows: A brief description of the materials and methods used in the study is given in section **Error! Reference source not found.**. In section **Error! Reference source not found.** results obtained from the bilateral tests and matches of 16 male speakers are given discussed. Conclusions are presented in section 4.

2 Materials and Methods

Recordings from 16 male subjects of spontaneous free discourse in Spanish were taken at a sampling rate of 44,100 Hz and 16 bits using HQ microphones in an isolated room. The distribution of speakers was the following: 2 pairs of MZ, 2 pairs of DZ, 2 pairs of non-twin siblings and 2 pairs of controls (non-relatives). Spontaneous fillers (long vowels maintained for more than 200 ms around vowel [ɛ] produced inadvertently by speakers of Spanish in words like “que”, “de”, or in hesitation marks like “eeh...” etc.) were used in the study. Each speaker was recorded twice (2 sessions) separated by a 3-week interval. Speech recordings were around 10 min long. An average of 8-10 fillers was extracted from each recording. A set of biomechanical parameters as body and cover dynamic mass and stiffness was estimated from the spectral description of the glottal source reconstructed by inverse filtering. The inter-cycle unbalances of these parameters were also estimated. Open, Close and Return Quotients were added to the parameter set as well as Contact, Adduction and Permanent Gap Defects. The parameter set was completed with jitter, shimmer, NHR and Mucosal Wave ratio to produce a feature vector of 65 parameters. A set of pair-wise parameter matching experiments was carried out by likelihood ratio contrasts used in forensic voice matching 6. The test is based on two-hypothesis contrasts: that the conditional probability between voice samples Z_a and Z_b (from to the subjects under test) is larger than the conditional probability of each subject to a Universal Speaker’s Model Z_U in terms of logarithmic likelihood ratios

$$LLR = \log\left[\frac{p(Z_b \mid Z_a)}{\sqrt{p(Z_a \mid Z_U)p(Z_b \mid Z_U)}}\right]; \qquad (1)$$
$$p(Z_b \mid Z_a) = \Gamma_a(Z_b); \quad p(Z_a \mid Z_U) = \Gamma_U(Z_a); \quad p(Z_b \mid Z_U) = \Gamma_U(Z_b)$$

where conditional probabilities have been evaluated using Gaussian Mixture Models (Γ_a, Γ_b, Γ_U) from each vector subset. Intra-speaker tests used recordings from different sessions. A priori expectations assume that MZ will show the largest LLR's, followed by DZ, then by non-twin siblings; non-related speakers expected to show the lowest LLR's.

3 Results and Discussion

Three main types of results are expected: a) those LLR (log-likelihood ratios) consistent with a priori expectations; b) a group of results which are not in agreement with a priori expectations; c) and probably the most important group being those allowing insightful discussion concerning the influence of genetic endowment and environmental factors in the type of speakers analyzed. Regarding the first group, all intra-speaker comparisons yielded positive and relatively large LLRs (from 5.2 to 50.1), except one (-14.5). As far as the inter-speaker comparisons are concerned, all LLRs (discarding the twin- and non-twin sibling comparisons, which are referred in a section apart) yielded negative values, except in one speaker (LLR= 3.5). Consequently the unexpected case of intra-speaker comparison and the unexpected case of inter-speaker comparison already mentioned above are found in the second group of results. Looking at the third group of results, for which no a priori assumptions were formulated since it is the first time that these biometrical parameters are used to test this kind of speakers, the following values have been obtained: For the two MZ pairs, as well as for the two DZ pairs, the LLRs are positive (57.3, 7.1, 34.9, 41.1, respectively), while the comparisons for non-twin siblings yielded different results depending on the pair under consideration: one pair yields an LLR of 19.1 while the other gives -32.4. Something similar happens in the comparison of unrelated speakers. In one case a strong mismatch is produced (LLR of -45.9) while in the other a weak match is obtained (LLR=3.5). These results are depicted in Fig. 1 and summarized in Table 1.

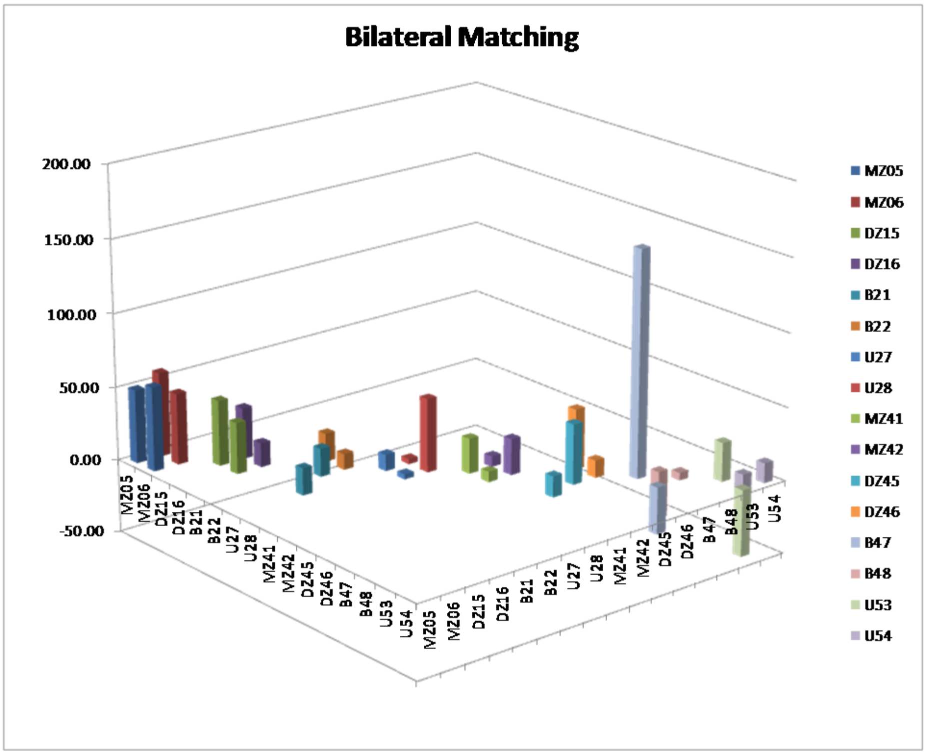

Fig. 1 Summary of the results for the different tests. MZ: Monozygotics; DZ: Dizygotics; B: Related Siblings (Brothers); U: Unrelated Speakers;

Table 1. Summary of the results for the different tests. MZ: Monozygotics; DZ: Dizygotics; RS: Related Siblings; US: Unrelated Speakers; (I): intra-speaker tests; (O): inter-speaker tests. Divided columns are used for each pair member.

Match/Case	MZ (I)		MZ(O)	DZ(I)		DZ(O)	RS(I)		RS(O)	US(I)		US(O)
LLR	50.1	48.4	57.3	44.5	16.3	34.9	**-18.8**	10.7	19.1	11.1	50.2	3.5
Match/Mismatch	High	High	High	High	High	High	**High**	High	High	High	High	Low
LLR	24	24.6	7.1	**-14.5**	11.8	41.1	154.6	5.2	**-32.4**	26.5	13.5	**-45.9**
Match/Mismatch	High	High	Low	**High**	High	High	High	Low	**High**	High	High	**High**

4 Conclusions

The most interesting finding is that there is a consistency in the results obtained within pairs in the case of both MZ twins and DZ twins: the results for pairs 1 (MZ) and 2 (MZ) show large positive matches, and the same occurs in the case of DZ twins: pairs 3 (DZ) and 4 (DZ) produce also large positive matches. Although at first sight, these results it might contrary-to-the-fact (i.e. the system produces large matches from two different speakers), they must be interpreted in a different way. We suggest that the parameters that have been used in such comparisons show a great influence of both genetic and environmental factors. If only the comparisons of MZ twin pairs had yielded large matches, the only explanation possible would be genetic influence. However, the fact that similar values are obtained for MZ and DZ twins cannot lead to that conclusion. The impact of external factors (like a similar living and educational environment, same age, etc.) must be more relevant than it may be thought a priori in this kind of voice studies. This reflection may be reinforced by the fact that opposing trends are observed for the non-twin sibling pairs. The comparison results from one of them looks more similar to the values obtained from non-related pairs, while the results from the other sibling pair is closer to the trend followed systematically by MZ and DZ twins. Further research would be necessary especially in order to study the role of the specific parameters (out of the 65 possible features) intervening in the results from each comparison. Likewise, it seems vital to consider a reanalysis with more speakers.

Acknowledgments. This work is being supported by an FPU grant from the Ministry of Education, a grant from the International Association for Forensic Phonetics and Acoustics, and by research grants TEC2009-14123-C04-03 and TEC2012-38630-C04-04 from Plan Nacional de I+D+i, Ministry of Economy and Competitiveness of Spain.

References

1. Van Lierde, K. M., Vinck, B., De Ley, S., Clement, G., and Van Cauwenberge, P. "Genetics of vocal quality characteristics in monozygotic twins: a multiparameter approach", Journal of Voice, 19 (4), 2005, pp. 511-518.
2. Cielo, C. A., Agustini, R. and Finger, L. S., "Características vocais de gêmeos monozigóticos", Revista CEFAC, 14 (6), 2012, pp. 1234-1241 (in Portuguese, summary in English).
3. Fuchs, M., Oeken, J., Hotopp, T., Täschner, R., Hentschel, B. and Behrendt, W., "Die Ähnlichkeit monozygoter Zwillinge hinsichtlich Stimmleistungen und akustischer Merkmale und ihre mögliche klinische Bedeutung", HNO, 48 (6), 2000, pp. 462-469.
4. Gómez, P., Fernández, R., Rodellar, V., Nieto, V., Álvarez, A., Mazaira, L. M., Martínez, R, and Godino, J. I.: Glottal Source Biometrical Signature for Voice Pathology Detection. *Speech Comm.*, (51) 2009, pp. 759-781.
5. Gómez, P., Rodellar, V., Nieto, V., Martínez, R., Álvarez, A., Scola, B., Ramírez, C., Poletti, D., and Fernández, M.: BioMet®Phon: A System to Monitor Phonation Quality in the Clinics. Proc. eTELEMED 2013: The Fifth Int. Conf. on e-Health, Telemedicine and Social Medicine, Nice, France, 2013, 253-258.
6. González, J., Rose, P., Ramos, D., Toledano, D. T. and Ortega, J., "Emulating DNA: Rigurous Quantification of Evidential Weight in Transparent and Testable Forensic Speaker Recognition", IEEE Trans. On Audio, Speech and Lang. Proc., 15 (7), 2007, pp. 2104-2115.