AUTOMATIC SPEAKER RECOGNITION OF SPANISH SIBLINGS: (MONOZYGOTIC AND DIZYGOTIC) TWINS AND NON-TWIN BROTHERS

Eugenia San Segundo[1]*; Hermann Künzel*[2]

[1]Department of Linguistics and Language Science, University of York
[2]Department of Phonetics, University of Marburg
eugenia.sansegundo@york.ac.uk; kuenzelh@uni-marburg.de

ABSTRACT

The performance of the automatic speaker recognition (ASR) system Batvox™ (Version 4.1) has been tested with a male population of 24 monozygotic (MZ) twins, 10 dizygotic (DZ) twins, eight non-twin siblings and 12 unrelated speakers (aged 18–52 with Standard Peninsular Spanish as their mother tongue). Since the cepstral features in which this ASR system is based depend largely on anatomical–physiological foundations, we hypothesized that such features ought to be gene-dependent. Therefore, higher similarity values should be found in MZ twins (100% shared genes) than in DZ twins, in brothers (B) or in a reference population of unrelated speakers (US).

Results corroborated the expected decreasing scale MZ > DZ > B > US since the similarity coefficients yielded by the automatic system for these speakers decreased exactly in the same direction as the kinship degree of the four speaker groups diminishes. This suggests that the system features are to a great extent genetically conditioned and that they are hence useful and robust for comparing speech samples of known and unknown origin, as found in legal cases. Furthermore, the 9.9% EER (Equal Error Rate) obtained when testing MZ pairs lies around the same value (11% EER) found in Künzel (2010) with German twins.

Keywords: forensic phonetics; twins; siblings; automatic speaker recognition; Spanish

RESUMEN

Hemos utilizado el sistema de reconocimiento automático Batvox™ (versión 4.1) con una población de hablantes masculinos compuesta de 24 gemelos monocigóticos, 10 gemelos dicigóticos, ocho hermanos no gemelares y 12 hablantes no emparentados (edades comprendidas entre 18 y 52 años, con español centropeninsular como lengua materna). Puesto que los parámetros cepstrales en los que se basa Batvox™ dependen en gran medida de las bases anatómicas y fisiológicas del tracto vocal del hablante, se propuso que estos debían estar influenciados genéticamente.

Esta hipótesis se pudo corroborar, puesto que los coeficientes de similitud arrojados por el sistema automático decrecen exactamente en la misma dirección en la que disminuye el grado de parentesco de las parejas de hablantes, es decir: gemelos monocigóticos, dicigóticos, hermanos no gemelares y hablantes no emparentados. Esto es, los gemelos monocigóticos obtuvieron valores más altos que los dicigóticos; estos, a su vez, mayores que los hermanos no gemelares, y, finalmente, estos últimos mayores que los hablantes no emparentados.

Estos resultados sugieren que los parámetros en los que está basado este sistema de reconocimiento están condicionados en gran medida por aspectos genéticos, y, por tanto, resultan útiles y robustos para la comparación de muestras de voz dubitadas e indubitadas que encontramos en un caso típicamente forense. Por otro lado, el EER (*Equal Error Rate*) del 9 % que se obtuvo en las comparaciones exclusivamente de gemelos monocigóticos supone un valor muy similar al hallado en estudios anteriores con gemelos monocigóticos alemanes, como Künzel (2010): EER del 11 %.

Palabras clave: fonética judicial; gemelos; reconocimiento automático; español

1. INTRODUCTION

1.1. The forensic relevance of twins and non-twin siblings

It is widely acknowledged that distinguishing twins poses a major challenge in the field of forensics because these individuals are physically very similar. For instance, biometrics such as fingerprints (Jain, Prabhakar & Pankanti, 2002) or palmprints (Kong, Zhang & Lu, 2006) have often been investigated in twins to study the subtle differences frequently observed between them. Similarly, researchers have investigated behavioral characteristics of twins such as handwriting (Srihari, Huang, & Srinivasan, 2008). In the same way that handwriting depends on physiology as much as on behavioral factors like training and habits, the foundations of speaker recognition are largely grounded on the idea that a voice is determined not only by anatomical structure but also by nonbiological or behavioral factors. These factors include mainly social or dialectal aspects but other environmental influences are possible. Nolan and Oh (1996, p. 39) highlighted that aspects of personal voice quality are determined by anatomical inheritance, mimicking traits from other people, or else, they are arbitrarily chosen in order to mark someone's personality. This *organic-learned dichotomy* (Nolan, 1997; Nolan & Oh, 1996) may be a good translation in phonetic terms of the well-known *nature–nurture dichotomy*, first outlined by Sir Francis Galton in 1875 (Galton 1875, in Segal 1993, p. 45).

This distinction, nature vs. nurture, has resulted in fruitful twin research in many disciplines, where heritability or concordance rates are calculated for certain traits in order to determine whether these could be genetically influenced. This happens when there is greater similarity on that trait between monozygotic (MZ) twin pairs than between dizygotic (DZ) twins. MZ twin pairs share 100% of their alleles and DZ twins, on average, share only half their genetic information, whereas both types of twin pairs share essentially the same prenatal and postnatal environments (Stromswold 2006, p. 334). This is the essence of the classical twin design, which requires that an important assumption be made: the *equal environment assumption* (EEA), i.e., it is assumed that the two twin types have similar environmental experience. [1] A number of studies have investigated the differences in MZ and DZ twins to assess the effect of genetic factors in voice (see Section 1.2), but—to the best of our knowledge—the joint consideration of MZ, DZ and non-twin siblings[2] has not been approached in phonetic studies before.

Acknowledging the existence of these two "forces", i.e., nature and nurture (alternatively also referred to as *organic* and *learned factors*, respectively) to explain the (dis)similarities between twins does not mean that their relative influence or importance can be clearly separated. Moreover, there is a third element, *epigenetics*, which is often neglected in twins' studies even though it usually comes into play to explain how changes in gene expression caused by mechanisms other than changes in the underlying DNA sequence can cause divergence in twins, which may account for strikingly dissimilarities between MZ twins. See, for instance, how a particular epigenetic process called DNA methylation (Martino et al. 2013; Philips, 2008) is reported to make the expression of genes weaker or stronger.

[1] From the EEA we can draw that the excess of similarity (for an investigated parameter) exhibited by MZ twins that is not present in DZ pairs must be due to genetic causes. Although we have taken advantage of this principle for our study, a strict application of the twin methodology would require the use of *heritability estimates* or *concordance rates*, in which the expected elevated similarity in MZs over DZs is often reported, depending on whether it is a continuous or a dichotomous trait (see Tomblin & Buckwalter, 1998).

[2] Monozygotic twins (also called identical) develop from one zygote that splits and forms two embryos, while dizygotic (also called fraternal) develop from two separate eggs that are fertilized by two separate sperm cells (Del Abril Alonso et al., 2009, p. 90). Full brothers are male siblings with the same father and the same mother.

A recent study (Felson, 2014) has aimed at undertaking a comprehensive evaluation of the EEA, which has often caused some skepticism amongst researchers. Felson presents evidence that suggests that neither extreme of the opposing views is correct, and that the truth lies somewhere in the middle. In other words, it seems that although environmental similarity may not have been adequately measured in some sociology-related twin studies, "the resulting bias is likely modest" (Felson, 2014, p. 184). Therefore it could be argued that despite its limitations twin research is still greatly encouraged nowadays to shed light on the interplay of genetic and environmental factors. Particularly referring to the difficult task of searching for genetic influences of the voice, Sataloff pointed out that "the complexities of genetic research in humans have left most of the relevant questions unanswered" (1995, p. 17).

Studies on twins' voices are undertaken for at least two main reasons. On the one hand, this type of studies can reveal—for the investigated voice characteristics—how the results of pairwise comparisons vary depending of the type of speaker considered. The comparison is usually between MZ twins and DZ twins; San Segundo (2014) also proposed drawing comparisons against non-twin siblings and unrelated speakers. The genetic influence of the analyzed voice characteristics is apparent when higher similarity is observed in MZ twins than in DZ twins, non-twin siblings or unrelated speakers. On the other hand, the relevance of twin studies to Forensic Phonetics[3] in particular lies in the search for robust[4] voice characteristics that could facilitate the discrimination of very similar speakers.[5] Hence, these four speaker groups (MZ twins, DZ twins, non-twin siblings and unrelated speakers) are proposed for testing the performance of a speaker-comparison system. As can be observed, the two highlighted aspects are strongly linked, since a set of characteristics may be robust for speaker comparison as far as they are maximally influenced by the speaker's genetic endowment and minimally due to learned factors, the latter favoring voice disguise or imitation. The predominance of genes over environment is clearly related to the two most repeated (and probably important) criteria in the identification of characteristics for Forensic Speaker Comparison (FSC), namely that these characteristics should be as consistent as possible for each speaker (low within-speaker variability) and that they should exhibit large variation amongst speakers (high between-speaker variability). Among others, these criteria were already outlined by Wolf (1972) and Nolan (1983) in the phonetic realm, but they also appear in the literature specifically related to *automatic speaker recognition* (ASR). For example, Kinnunen and Li (2010) refer to the same characteristics for an ideal ASR system.

[3]A definition of Forensic Phonetics has been provided by different authors (e.g., Jessen, 2008; Künzel, 1994; Nolan, 1997; Rose, 2002). What all these definitions have in common is that they specify for the discipline of Phonetics the general definition of Forensics as the application of scientific knowledge to legal problems. *Forensic Phonetics* would then be the application of Phonetics aimed at solving any type of legal issue (see San Segundo, 2014). One of the most typical forensic cases where a phonetic expert is involved is one in which has to compare the voice of an offender (i.e., speech samples of an unknown speaker) with the voice of a suspect or several suspects (i.e., speech samples of known origin). It is widely accepted nowadays to refer to this kind of task as *Forensic Speaker Comparison* (FSC). Other possible tasks which a phonetician may be requested to perform for forensic purposes are described, for instance, in Foulkes & French (2012).

[4] Robustness is usually associated to a degradation factor, and could be defined as the reluctance of a system to lose performance when certain degradation factor is present. For our study, genetic similarity is seen as the degradation factor.

[5]While the typical question that a forensic phonetician has to answer in a FSC case is: "How much more likely the magnitude of the difference between samples is if they came from the same speaker than from different speakers?" (Rose, 2002, p. 89), in the case of siblings' voices the question would have to be formulated in a slightly different way. For example, as pointed out by Feiser (2009), "not uncommonly the question posed in court is whether a given unknown recording could have been spoken by the subject's brother(s) instead of the subject himself. Other than being a possible legal strategy, this question suggests itself because siblings often have similar sounding voices" (2009, p. 1).

1.2. Literature review: twins and ASR

From a literature review of around 30 voice-related twin studies (San Segundo, 2014), we can draw some interesting conclusions. For instance, it seems that previous phonetic studies focusing on twins have aimed at basically one of the following objectives (see San Segundo, 2015): (a) trying to find a genetic component in the variation of certain voice characteristics by searching differences between MZ and DZ twin pairs (e.g., Debruyne, Decoster, Van Gijsel, & Vercammen, 2002; Przybyla, Horii, & Crawford, 1992) or else, in a forensic scenario, (b) creating a system capable of discriminating between MZ and DZ twins (e.g., Forrai & Gordos, 1983) or, more frequently, testing whether it is possible to distinguish a speaker from his/her co-twin (e.g., Ariyaeeinia, Morrison, Malegaonkar, & Black, 2008; Homayounpour & Chollet, 1995; Künzel, 2010; Loakes, 2006; Nolan & Oh, 1996; Scheffer, Bonastre, Ghio, & Teston 2004). For a thorough discussion of the results derived from previous twin studies, see San Segundo (2014), where previous works have been classified in four groups depending on whether they represent perceptual, acoustic, articulatory or automatic (ASR) approaches.

While most of the studies undertaken from an acoustic perspective focus on *traditional* phonetic characteristics, as described in Künzel (2011) and Rose (2006)—for example, fundamental frequency (f_0), formant patterns or temporal characteristics such as word duration, vowel duration or *Voice Onset Time* (VOT)—, research into laryngeal features and phonation characteristics derived from the glottal waveform has been very limited. Classical distortion characteristics such as jitter and shimmer have only occasionally been explored in twins (van Lierde, Vinck, De Ley, Clement, & Van Cauwenberge 2005; Weirich & Lancia, 2011). More recently, some investigations on twins' voices (San Segundo, 2012; San Segundo & Gómez-Vilda, 2013; San Segundo 2014; San Segundo & Gómez-Vilda, 2015) have analyzed a considerably larger number of glottal features, on the basis of the voice analysis methodology described in Gómez-Vilda et al. (2007), which relies on the decoupling of the vocal tract from the glottal source estimates.

If we focus on ASR studies in particular, this approach to twins' voices has not been extensively developed, in comparison with other acoustic studies investigating specific segmental features. The main objectives of the ASR studies reviewed in San Segundo (2014) are one of the following: a) comparing the performance of ASR systems with the ability of familiar and non-familiar listeners to discriminate twins (Homayounpour & Chollet, 1995); (b) testing if an ASR system is able to detect correctly the twin pair of a speaker (Scheffer et al. 2004), or (c) in general, testing the intra-speaker, inter-speaker and intra-pair similarity of twins, for example in terms of Likelihood Ratios (LRs) or similarity coefficients. In this last research line we find two recent studies, namely Kim (2009) and Künzel (2010). Since both use the same ASR system that we are using in our study, we will devote an important part of this section to the description of their objectives and main findings.

Kim (2009) studied 22 Korean female twin pairs (17 MZ, including one triplet and five DZ) using Agnitio Voice Biometrics' Batvox™ (Version 3.0). Two different speaking styles—text reading and spontaneous interview—were used. The results of this investigation showed that every twin speaker was correctly identified in the same speaking style condition (when models and test files were *read* speech). According to the author, this would suggest that, at least in ASR, the same speaking style setting should be provided in order to get more confident results. Noteworthy of this study is also that in nine out of 22 pairs, intra-twin LRs in the same speaking style condition were higher than intra-speaker LRs in different speaking style condition. This situation is highly undesirable in a forensic context, where inter-speaker variation should be larger than intra-speaker variation (Wolf, 1972).

Künzel (2010) is the most recent study on automatic speaker recognition in which a Bayes-based system (Batvox™, Version 3.1)

was used to calculate LR distributions for inter-speaker, intra-pair and intra-speaker comparisons. A total of 35 German MZ pairs (26 female and nine male) participated in this study and two different tests were designed. In the first one, both target voices consisted of the same read text, while in the second one the speaker models were built from spontaneous speech samples but read speech samples were used as targets. The results showed that in the first experiment the automatic system allowed a perfect distinction of each member of a male twin pair (i.e., 0% of *Equal Error Rate*; EER) and 0.5% EER for female twin pairs. In the second experiment, the EER rose to 11% for male twin pairs and 4.4% for female twin pairs. These values represent the crossover point in the *Tippett plot* for the inter-speaker/intra-speaker LR distributions. However, the results for female twins are worse when considering intra-pair/intra-speaker distributions (19% EER in the first experiment and 48% EER for the second experiment). Therefore, the performance of the system was clearly superior for male than for female voices. The author's explanation for this phenomenon is that "as a consequence of the higher fundamental frequency of female voices the spacing of the harmonics is less dense than for male voices, which in turn yields less speech sound- and speaker information in the spectrum" (Künzel, 2010, p. 270). This becomes clearer if we bear in mind that the spectrum is used for the extraction of the *mel-frequency cepstrum coefficients* (MFCCs), which are the features in which this automatic system used is based upon.

Finally, references to siblings' voices within an automatic approach are almost inexistent except for the study of Charlet and Lecha (2007), which tested a text-dependent speaker recognition system with 33 families, finding that the son was highly confused with his brother. The implication is that someone could be a good impostor of his brother, making this type of speakers especially relevant in forensic studies and thus justifying not only the study of twins but also of non-twin siblings.

2. DATA AND METHOD

This section provides some details about the subjects recruited for this investigation, the data collection method and the characteristics of the speech samples analyzed. The methodology for carrying out the speaker comparison is also described, including the different stages of the automatic system Batvox™ as well as a description of the method for the measurement of system performance.

2.1. Data collection

This investigation is part of a larger research project (San Segundo, 2014); more details about the corpus of twin and non-twin subjects can be found in San Segundo (2013b, 2014). The automatic analysis that we present here is based on speech samples extracted from the fifth corpus task: informal interview with the researcher.

2.1.1. Subjects and recording characteristics

Our corpus of speakers is made up of 24 MZ twins, 10 DZ twins, eight brothers and 12 unrelated speakers with no kinship relationship (friends or work colleagues). The importance of the first three speaker types has been explained in the introduction. The fourth group of speakers was recruited with the aim of creating a reference population, whose relevance for Likelihood-Ratio-based forensic studies has been acknowledged on numerous occasions in the literature (Morrison, 2010).[6] Friends or work colleagues were preferred instead of complete strangers in order to match as closely as possible the speaking style found in the conversations between brothers, characterized by their spontaneity due to a long-term relationship. The age of the speakers ranged between 18–52 years (mean: 28.96).

[6]Eventually, a cohort of 31 Spanish male speakers was used as background population (spontaneous conversation and high-quality recordings), coming from Batvox™ databases (see Section 2.2.1) because a minimum of 25 speakers is required using this system. However, the group of twelve unrelated speakers served to compare the matching scores of MZ, DZ and non-twin brothers with speakers without any type of genetic relationship.

The age difference between the brothers in each pair varied between four and 11 years. They were all male speakers of North-Central Peninsular Spanish with no speech pathologies or hearing difficulties. All speakers were recorded on two different occasions in order to account for intra-speaker variability. These two recording sessions were separated by 2–3 weeks, which served to obtain non-contemporaneous speech samples.

Participants came in pairs (either with their twin or friend) to the recording sessions, which took place in the Phonetics Laboratory of the Spanish National Research Council. They were recorded with omnidirectional condenser microphones (head-mounted device) with flat frequency response. Recording specifications were: 44,1 kHz sample rate, 16-bit resolution and mono channel. Speakers were recorded in two different (acoustically isolated) rooms where they could communicate via landline telephone for certain cooperative tasks. Even though the recordings are high quality (telephone-degraded at a later stage), this set-up replicated forensic realistic conditions at the same time that it minimized the "observer's paradox" (Labov, 1972) by avoiding the presence of the researcher at the place of the recording.

2.1.2. Speech samples

Speech samples were extracted from the fifth task of the corpus fully described in San Segundo (2013b, 2014). In this speaking task (informal interview with the researcher), the researcher is at one end of the telephone and one member of each speaker pair at a time is at the other end of the telephone. In this interview, lasting around 10 minutes, the researcher asks each of the interviewees about any of the topics that they have been discussing with their twin/friend in the first task. Originally intended to elicit hesitation markers (i.e., vowel fillers) from the speakers, which could then facilitate glottal analyses (e.g., San Segundo & Gómez-Vilda, 2013, 2015), this corpus task was also considered the most appropriate for the ASR analysis. On the one hand, conversations here are long enough to allow the extraction of at least 120 seconds of net speech per speaker. According to Künzel (2010), this is the recommended duration of a voice sample to be analyzed using the ASR system Batvox^{TM}. On the other hand, this corpus task presents the advantage of having the same interlocutor in all conversations, i.e., the researcher. This levelled the speaking style of all speakers to the same degree of spontaneity/formality.[7]

The speech fragments (120 s of duration on average) were extracted from the audio files belonging to the first and the second recording session of each speaker (average duration of five min). The speech material chosen for further analyses was selected from approximately the middle of the audio file, in order to avoid the beginning of the conversation, where the speaker has not already settled to his ordinary speaking style. Prior to the labeling and extraction using Praat (Version 5.3.79), the audio files were first aurally examined in order to remove extraneous noise, laughter, clicks, cough, etc., following the recommendations in Künzel (2010, p. 256).

2.2. Analysis tools and method

2.2.1. ASR analysis

For the ASR analysis, we have used the software Batvox^{TM} (Version 4.1), which is based on parameters related to the resonances of the vocal tract, basically cepstral coefficients. One of the main assets of automatic systems is that between-sample differences in the speech content are not relevant because ASR systems exploit the voice itself and disregard the linguistic content of the utterances to a great extent. While this does not mean that Batvox^{TM} is independent of the language mismatch between utterances to

[7] The importance of the same interlocutor is strongly linked to the theory of accommodation (Giles, Coupland & Coupland, 1991). More recently, a fast-growing research line investigating convergence and imitation patterns in speech occurring between speakers in the course of conversational interactions (see e.g., Pardo, 2006; Pickering & Garrod, 2004; Trouvain & Truong, 2012), provides further evidence that speaker interlocutors actually converge in a number of phonetic features.

compare—which is not our case—, it still holds true that the relatively small influence of the linguistic content makes the extraction of speaker samples relatively easy, as there is no need for comparable phonetic units between speakers (in contrast with most traditional phonetic features). An overview of the first stages of a typical ASR system follows (see Kinnunen & Li, 2010, pp. 2–3[8]):

- *Parameter extraction*: transformation of the raw signal into feature vectors in which speaker-specific properties are emphasized and statistical redundancies suppressed.
- *Speaker modeling*: the feature vectors extracted from the training utterance of a speaker are used to train a speaker model, which is then stored in the system database. The *Gaussian mixture model* (GMM; Reynolds, Quatieri & Dunn, 2000; Reynolds & Rose, 1995) would be the most popular model for text-independent recognition, according to Kinnunen and Li (2010, p. 4).

Focusing on Batvox™ in particular, its main characteristics are, as explained in Künzel and Alexander (2014, p. 247): a 38-dimensional feature vector consisting of 19 MFCCs plus their deltas, GMM-Channel-Factor analysis for the compensation of speaker models (Kenny, Boulianne, Ouellet, & Dumouchel, 2005) and nuisance attribute projection (Campbell, Campbell, Reynolds, Singer, & Torres-Carrasquillo, 2006) for the test files.

A comparison between the statistical model for the reference speaker and the results for the target speaker's model is carried out. The similarity score obtained after this procedure is then weighed using a reference population. For this study, the system was set to *identification mode*,[9] where results are indicated as normalized scores that can be used to calculate False Alarms (FA) and False Rejections (FR) rates, and eventually, EERs. This identification mode of operation was deemed the most appropriate for the purpose of this investigation (see *Batvox 4.1 Basic User Manual*, 2013). As reference population, a cohort of 31 Spanish male speakers was used (from Batvox™ databases), with characteristics matching those of the recordings in the twin corpus: male speakers, spontaneous conversations and high-quality recordings.

The following tests were carried out:

- *Intra-speaker comparisons* (*matches* or *target trials*): each speaker's session one was compared with the same speaker's session two.
- *Inter-speaker comparisons* (*non-matches* or *impostor trials*): each speaker's session one was compared with all other speakers' session two.
- *Intra-pair comparisons:* each speaker's first session was compared with the first session of his sibling or conversation partner in the case of unrelated speakers.

The first two types of comparisons served to test the general performance of the comparison system without taking into account the fact that some speakers are MZ, DZ or non-twin siblings. Yet, in order to investigate the magnitude of the *sibling effect*, the third type of test is also necessary.

2.2.2. Performance measures

Assessing the output accuracy of a forensic-comparison system is a very relevant aspect in forensic sciences. Several measures and graphical ways have therefore been developed to evaluate such accuracy: for instance, the *log-likelihood-ratio cost* (C_{llr}), originally envisaged for its use in automatic speaker recognition (Brümmer & du Preez, 2006; van Leeuwen & Brümmer, 2007) but also applied in forensic-comparison studies based in traditional acoustic parameters (e.g.,

[8] More detailed information can be found in Künzel (2010: 253–4) where he cites relevant bibliographic references in this field (Drygajlo, 2007; Gonzalez-Rodriguez, Fierrez-Aguilar, & Ortega-Garcia, 2003; Przybocki, Martin, & Le, 2007; Ramos, 2007).

[9] Note that, depending on the author followed, this type of recognition task could be named differently (e.g., *verification task*). Cf. Bimbot et al. (2004).

Gonzalez-Rodriguez, Rose, Ramos, Toledano, & Ortega-Garcia, 2007; Morrison & Kinoshita, 2008). Besides, *Tippett plots* (Meuwly, 2001) have also been used as a graphical method to present the output of forensic systems and to assess its accuracy. In our study we have used EER, an accepted measure of the performance of an identification (also used in Künzel, 2010, or Künzel & Alexander, 2014, for the performance testing of Batvox™). The EER represents the point of intersection of matches and non-matches. Consequently, an EER of 0% indicates that there is no overlap of matches and non-matches, so neither FA nor FR occur. EERs were calculated using the Biometrics 1.2 software (Biometrics 1.2, 2012).

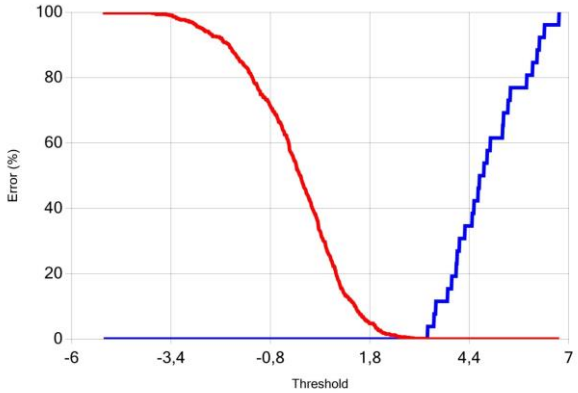
3. RESULTS

3.1. Overall system performance

As explained above, we carried out three types of tests, which yielded results for intra-speaker, inter-speaker and intra-pair comparisons. If we first look at the results for intra-speaker and inter-speaker comparisons alone, we see that similarly high coefficients of recognition are obtained for all the pooled four speaker types (MZ, DZ, B and US). This can be observed in Figure 1, which shows a 0% EER. The input values for the creation of this figure were of two types:

- *Matches* (blue line): the values were obtained from the comparison of each speaker's session one with his own session two.
- *Non-matches* (red line): the values were obtained from the comparison of each speaker's session one and all other speakers' session two.[10]

Figure 1: Cumulative distribution of scores for same-speaker comparisons or matches (blue) and different-speaker comparisons or non-matches (red).



The 0% EER indicates that there is no overlap of matches and non-matches, so neither FA nor FR occur. This shows that the overall system performance with high-quality recordings and without taking into account the sibling effect (intra-pair comparisons) is perfect.

3.2. Sibling effect

When taking into account also intra-pair comparisons, in addition to matches and non-matches, the recognition coefficients are expected to be much lower, as the comparison is not between the same individuals. However, different patterns were observed depending on the type of speaker (MZ, DZ, B or US). This can be seen in Table 1, where the values obtained are classified per speaker (i.e., his intra-speaker coefficients) and per speaker pair (i.e., their intra-pair coefficients), depending on whether they are MZ, DZ, B or US. As it can be observed in this table, all intra-speaker comparisons yield similarly high coefficients of recognition. In relation to the intra-pair comparisons, Table 1 is useful to observe the different values obtained by different speaker pairs, i.e., the performance of the system can be analyzed per speaker or per speaker pair. The fact that the speakers in this investigation are not very numerous is an advantage in order to carry out this kind of detailed examination. For instance, if we look at within-group

[10] To avoid comparing a speaker with his sibling or conversational partner, at least in this first analysis which does not take into account the sibling effect, only the even members of each speaker pair were selected, both for the matches and non matches. That is, only speakers 02, 04, 06, etc., were used in the analysis. Following the methodology described in Künzel (2010), in order to facilitate this task, one member of the twin pairs was labeled *red* (the odd numbers) and the other member was labeled *blue* (the even numbers). Figure 1 shows the EER (0%) using the blue speakers. The same test was repeated using only the red speakers and a very similar EER was obtained (0.07%).

differences, the value of MZ pair 39–40 (0.64) is very different from the other pairs' coefficients (much higher in average).

If we are interested in the behavior of the groups in general, and not specifically in each pair, Table 2 and its corresponding figure (Figure 2) are more insightful and probably more appropriate to assess the system performance depending on the speaker type. According to the information in Table 2, MZ intra-pair comparisons yield the highest values (i.e., the dissimilarity is the lowest, so they are the most similar speakers). From the average values obtained by the MZ pairs to the coefficient values yielded for US, we observe a gradation from largest to lowest, all through the average values of the DZ intra-pair comparisons and the B intra-pair comparisons. This trend is thus in agreement with our hypothesis, where we predicted the following scale (from more to less similar): MZ > DZ > B > US. In other words, the coefficient grading goes in the same direction as the "magnitude" of kinship relationship.

We have added to Table 2 the average coefficients obtained in (MZ) intra-speaker comparisons. As expected, these same-speaker comparisons yield the highest coefficients. The inclusion of these matches in the table is intended to serve as a baseline to which the rest of (intra-pair) coefficients can be compared, under the assumption that nobody could be more similar to anyone than to himself, although some exceptions may occur in the case of MZ twins, as we describe in Section 3.3.

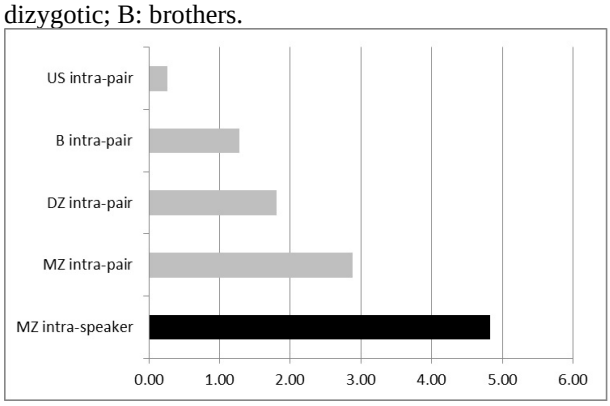
Table 1: Summary of the results for the different comparison tests. MZ: Monozygotic twins; DZ: Dizygotic twins; B: Brothers; US: Unrelated Speakers. Divided columns are used in the intra-speaker scores for each pair member. Cases: xxvyy means speaker xx versus speaker yy.

	MZ			DZ			B			US		
	Intra-speaker		Intra-pair	Intra-speaker		Intra-pair	Intra-speaker		Intra-pair	Intra-speaker		Intra-pair
Cases	01v01/02v02		01v02	13v13/14v14		13v14	21v21/22v22		21v22	25v25/26v26		25v26
Score	4.22	3.48	3.79	5.25	6.17	3.77	4.51	6.24	0.64	4.93	4.47	0.39
Cases	03v03/04v04		03v04	15v15/16v16		15v16	23v23/24v24		23v24	27v27/28v28		27v28
Score	4.82	4.79	2.65	4.27	4.87	2.53	7.76	5.27	3.31	3.99	4.29	0.64
Cases	05v05/06v06		05v06	17v17/18v18		17v18	47v47/48v48		47v48	29v29/30v30		29v30
Score	4.29	4.95	3.45	5.13	6.35	0.18	5.53	4.63	0.79	5.29	5.42	-0.66
Cases	07v07/08v08		07v08	19v19/20v20		19v20	49v49/50v50		49v50	31v31/32v32		31v32
Score	4.23	4.14	2.31	3.51	5.46	2.17	2.78	3.31	0.36	2.92	4.67	0.25
Cases	09v09/10v10		09v10	45v45/46v46		45v46				51v51/52v52		51v52
Score	3.64	4.06	2.66	3.44	3.83	0.40				3.80	3.52	0.71
Cases	11v11/12v12		11v12							53v53/54v54		53v54
Score	3.24	5.29	1.34							4.03	5.22	0.22
Cases	33v33/34v34		33v34									
Score	4.55	6.06	3.20									
Cases	35v35/36v36		35v36									
Score	6.44	3.94	4.93									
Cases	37v37/38v38		37v38									
Score	5.41	4.52	3.54									
Cases	39v39/40v40		39v40									
Score	6.05	6.74	0.64									
Cases	41v41/42v42		41v42									
Score	4.68	5.9	3.53									
Cases	43v43/44v44		43v44									
Score	4.43	4.08	2.59									

Table 2: Average coefficients per speaker type and test type. All the intra-pair values per speaker type but also the intra-speaker values for MZ twins (last row) are shown, in order to highlight the grading in values (from lowest to largest), where the lowest means more dissimilar and the largest, more similar.

Speaker type	Test type	Average coefficient
Unrelated speakers (US)	Intra-pair	0.26
Non-twin brothers (B)		1.28
Dyzigotic twins (DZ)		1.81
Monozygotic twins (MZ)		2.89
Monozygotic twins (MZ)	Intra-speaker	4.83

Figure 2: Grading of average coefficients from US (Unrelated Speakers) to MZ (Monozygotic) intra-speaker comparisons: the larger the value, the more similarity. Grey is used for intra-pair comparisons while black is used for intra-speaker comparisons; DZ: dizygotic; B: brothers.



3.3. Special case study: MZ twins

The MZ intra-pair comparisons deserve special consideration. As they represent the cases of highest similarity in human beings, they have been more often studied than the other types of kinship relationships considered in this investigation. In the case of FSC carried out using automatic recognition methods, the existence of previous studies that have also used Batvox™ for the voice comparison of MZ twins gives us the opportunity to compare our results with previous findings.

For the MZ twins participating in our study, we have considered useful to compare the coefficients obtained by each speaker in the intra-speaker (IS) comparisons with the coefficients obtained by these same speakers in the intra-pair (IP) comparisons. Table 3 contains this information, extracted from the general results shown in Table 1.

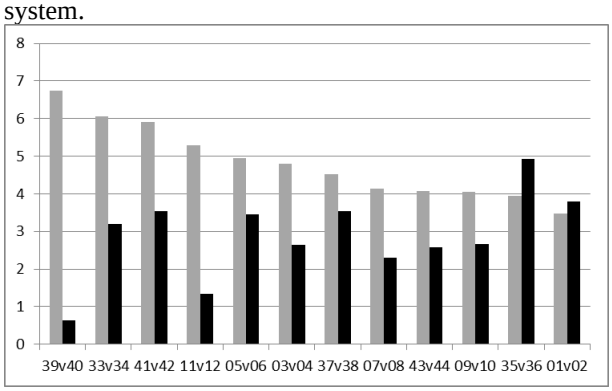
Table 3: For each of the MZ twin pairs, we show the *IS–IP value*, calculated as the difference between the intra-speaker (IS) comparison coefficient and the intra-pair (IP) comparison coefficient. Cases: xxvyy means speaker xx versus speaker yy. Only two out of 12 cases (greyshaded) show negative values.

MZ pair	IS comparison coefficient	IP comparison coefficient	IS–IP difference
01v02	3.48	3.79	-0.31
03v04	4.79	2.65	2.14
05v06	4.95	3.45	1.50
07v08	4.14	2.31	1.83
09v10	4.06	2.66	1.40
11v12	5.29	1.34	3.95
33v34	6.06	3.20	2.86
35v36	3.94	4.93	-0.99
37v38	4.52	3.54	0.98
39v40	6.74	0.64	6.10
41v42	5.9	3.53	2.37
43v44	4.08	2.59	1.49

We have calculated an IS–IP value to measure the difference between the IS comparison coefficient and the IP comparison coefficient. This has been done per speaker and speaker pair. Note however that for the IS coefficients, we have only taken into account the values obtained by one member of the pair: the twin member with the even number in his pair (i.e., 02, 04, 06, 08, etc.). The selection of the IS coefficients of the odd pairs did not yield any negative value. That is the reason why we show the results of the even numbers; as explained above, the interest of this calculation lies in finding any possible speaker pair subject to discrimination errors by the system under test.

As shown in Table 3, only two cases out of twelve MZ pairs show a negative value in their IS–IP value, meaning that the IP coefficient is larger than the IS coefficient. This implies that in these two cases the automatic system Batvox™ would not be able to discriminate between one twin and the other. In positive values, we can say that in 83.3% of the total MZ cases, the system identifies an identical twin without falsely accepting his co-twin. In Figure 3 we draw the IP and IS coefficient values per MZ twin pair, in IS-decreasing order to show how the trend "large IS–small IP" is followed in all cases except in the last two, corresponding to the MZ pairs 01v02 and 35v36, as we could also observe in Table 3. These two pairs account for the 16.7% not confirming the hypothesis that IS comparisons are always larger than MZ IP comparisons. However, as we will discuss below, this small percentage is in agreement with previous studies.

Figure 3: IS–IP difference per speaker pair. We show in the x-axis the 12 MZ (monozygotic) pairs and in the y-axis the coefficient values for IS (intra-speaker) comparisons (grey) and IP (intra-pair) comparisons (black). Only the two last twin pairs would not be discriminated by the system.



The two specific cases of MZ twins that were not be recognized by the system explain the 9.9% EER obtained in Figure 4, where the line for matches (right line) is used in this case for intra-pair comparisons (only MZ) and the line for non-matches (left curve) represents the inter-speaker comparisons.

In Figure 5, we have added the curves in Figure 1, which showed the overall system performance. The line further to the right (black) is for IS comparisons of all the speakers in the corpus, and the other right line (blue) represents the IP comparisons, only for MZ. In this new figure, one can distinguish a left-shift from the general IS-curve to the MZ IP-curve, which indicates the performance deterioration from a situation where the system has to recognize same speakers to a situation where identical-twin recognition takes place. The lines for the non-matches in both cases (compare the two curves rising to the left) are practically identical. In both cases, they represent different-speaker comparisons, while in one case (yellow curve, i.e., non-matches in Figure 1) these tests compared the first session of each speaker with the first session of all the other speakers in our corpus; and in the other case (red line, i.e, non-matches in Figure 4), the different-speaker tests were obtained from comparing each speaker's first session with all the other speakers' second session.

Figure 4: Cumulative distribution of scores for intra-pair (IP) comparisons or matches (blue) and inter-speaker (IS) comparisons or non-matches (red). The EER obtained is 9.9%, indicating that some overlap between matches and non-matches exist.

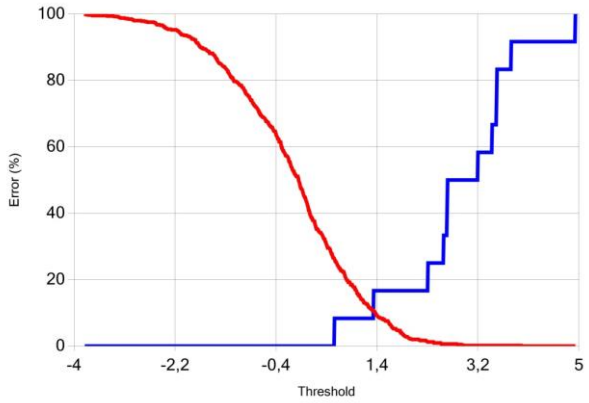
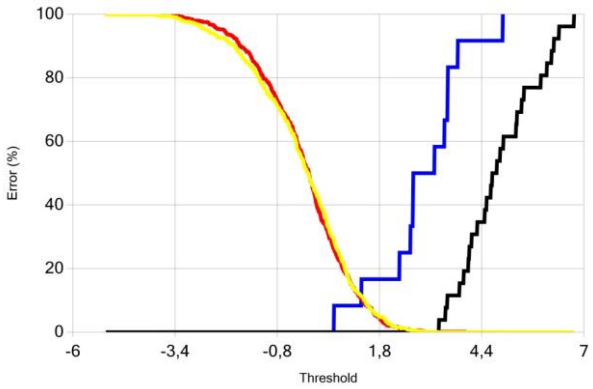


Figure 5: *Lines rising to the right*: cumulative distributions of scores for all-speakers intra-speaker (IS) comparisons (black line) and MZ intra-pair (IP) comparisons (blue line, crossing at the EER 9.9%). *Curves rising to the left* (yellow and red): both represent the cumulative distribution of IP comparisons or non-matches.[11]



4. DISCUSSION

Several aspects can be discussed in relation to the results obtained with the automatic system Batvox™. On the one hand, we have tested the overall system performance with our speakers as *tests* and *models*, i.e., without taking into account the fact that part of these speakers are twins or siblings. This test has yielded intra- and inter-speaker comparisons. In other words: matches (for same-speaker comparisons) and non-matches (for different speaker comparisons). The 0% EER obtained for this

[11]The only difference between both curves rising to the left in Figure 5 is that one (yellow) compared first session of every speaker with first session of all other speakers, while the other (red) compared first session of every speaker with second session of all other speakers.

first test shows that there were no FA or FR, which indicates a perfect performance of the system.

On a second test, we introduced the concept of intra-pair (IP) comparison while taking into account the fact that out of the 54 speakers considered, 24 were MZ twins, 10 were DZ twins, eight were non-twin siblings and 12 were unrelated speakers. The results of comparing each speaker with his pair corroborated the hypothesis that higher similarity values would be found in MZ twins than in DZ twins, in siblings or in unrelated speakers. On average, higher coefficients were obtained by MZ IP-comparisons, followed by DZ twins, brothers and unrelated speakers, in that order. This is the scale that we expected taking into account the degree of shared genes and shared environmental factors by pairs in these four speaker types (see Section 1).

Finally, when the IP comparison values only for the MZ twins were compared with the non-matches, we obtained a 9.9% EER, so a left-shift was observed in Figure 5 from the general IS-curve to the MZ IP-curve. This represents the deterioration in the system performance from a situation where the recognition is between same speakers to a situation where identical-twin recognition takes place. These results could be compared with the 11% EER obtained by Künzel (2010), who also studied MZ twins. Although he studied both male and female twins, and two speaking styles (read speech and spontaneous speech) we have considered here only the results for male twins and spontaneous speech. The male participants in Künzel's study were nine MZ pairs while in our investigation there are 12 pairs. Yet the EER percentages are very similar, indicating that the rate of false acceptance of other twin by this system is around 10%. Having a closer look at the data for the individual twin pairs (i.e., comparing the IP and the IS values), Künzel found that some speakers were more easily identified than others. Our study also points in this direction, as the coefficients in the IS and IP comparisons differ between pairs, sometimes considerably (see Table 3 and Figure 3). In fact, as it follows from the literature review carried out in San Segundo (2014)—and summarized in the introduction to this article—this heterogeneity appears as a common factor in most studies on twins' voices. Previous analyses derived from the same corpus of Spanish twins showed the same phenomenon, namely that different twin pairs exhibit different results when an IP comparison is carried out, regardless of the type of phonetic-acoustic examination, be it formant trajectories (San Segundo, 2014) or glottal characteristics (San Segundo & Gómez Vilda, 2013). Indeed, this need not be a characteristic exclusively linked to twins but common in speaker recognition. As Doddington, Liggett, Martin, Przybocki, and Reynolds (1998) explain, different speaker typologies could be established on the basis on how easily recognized/imitated speakers are. This implies that, in terms of FA and FR, "a considerable amount of the errors in an experiment, may be linked to only a few speakers" (Künzel, 2010, p. 264).

Apart from Künzel (2010), the other study that has analyzed twins' voices using Batvox™ (Version 3.0) focused only on female voices (Kim, 2009), so the results in that study are not comparable with ours. From the investigation of Künzel (2010) we know that there is an important sex-related difference in the performance of the automatic system, this being superior for male as compared to female voices (see Section 1.2). Yet, it is worth-mentioning that Kim (2009) also found that in nine out of 22 cases, twins could be misidentified. She specifically refers to a situation where intra-twin LRs in the same speaking style condition were higher than intra-speaker LRs in different speaking style condition.

5. CONCLUSIONS

It is well known that the vocal tract is made up of different cavities (oral, nasal and pharyngeal). Each of these cavities has a resonance profile, which is supposed to be somehow typical and idiosyncratic for each speaker, at least similarly to what happens with other parts of the human anatomy, which are more or less individual (Künzel, 2010, p.

40). Automatic methods in general (as explained above), and Batvox^{TM} specifically, extract a set of features representing the resonance profile of the vocal cavities of a speaker (MFCCs) and creates a multidimensional vector. These are the kind of parameters (*low-level features*) used in this type of analysis, in contrast with *high-level features*, which would refer to other linguistic aspects that also serve to characterize a speaker, such as intonation patterns, pausing behavior, jargon, sociolect, regional coloring, etc. (see Kinnunen & Li, 2010; Künzel & Alexander, 2014). No separation of linguistic or phonetic units is made, therefore, under the automatic approach. This is why Jessen (2008, p. 699) classifies this type of automatic methods as *holistic*: "The distribution of the MFCCs over the entire course of the recording of a speaker is determined. (…) no segmentation of the speech stream into different linguistic categories, such as consonants, vowels or syllables is performed" (2008, p. 699).[12]

According to what has just been explained, we hypothesized that the cepstral features in which this ASR system is based would be strongly gene-dependent, as they depend largely on anatomical–physiological foundations. Therefore, higher similarity values should be found in MZ twins (100% shared genes) than in DZ twins, in brothers (B) or in a reference population of unrelated speakers (US). To the best of our knowledge, this represents the first investigation into the voice characteristic of Spanish twins and non-twin siblings from an ASR perspective. Previous studies (San Segundo, 2010a; San Segundo, 2010b; San Segundo, 2012; San Segundo, 2013a; and San Segundo & Gómez-Vilda, 2013, 2015) have tackled FSC of this set of twins and non-twin speakers from different points of view (mainly glottal analyses and formant trajectories).

The most important conclusion that can be drawn from this analysis is that—as we have hypothesized—the similarity coefficients yielded by the automatic system Batvox^{TM} decrease exactly as the kinship relationship of the speaker pairs decreases. In other words, the score sorting from largest to smallest resulted in the following scale of values: MZ > DZ > B > US.

In the introduction to this investigation we explained our reasons for sustaining the hypothesis that higher similarity values (hence worse recognition) would be found in MZ IP-comparisons than in DZ IP-comparisons. In turn, these speakers would be more similar than non-twin brothers (B) and the latter more similar than unrelated speakers (US). The justification for this lies in the fact that MZ twins share 100% of their genetic information and in general they also share educational and environmental backgrounds, while DZ twins share 50% of their genes but usually the same external influences as MZ twins. Sharing the same genetic information as DZ twins, brothers are supposed to share less environmental characteristics due to the age gap; and finally unrelated speakers share neither their genes nor their environmental background. This reasoning gives rise to the scale: MZ > DZ > B > US, where ">" means "more similar than"; for the aim of our investigation, at least in voice terms. To our knowledge, this is the first time that this hypothesis has been tested for an automatic system using the four types of speakers mentioned (MZ, DZ, B and US). The underlying idea behind this hypothesis is not foreign to phonetic studies, however. For instance, Künzel (2010, p. 251) sustains that "the more similar the geometry of two vocal tracts is, the more similar will be the respective similarity coefficients, or LRs" and that "this problem is particularly relevant to related speakers, most extremely for identical (MZ) twins" (Künzel, 2010, p. 251). As a matter of fact, the issue of how the comparison of very similar speakers can affect the recognition performance of an automatic

[12]As explained in Jessen (2008, p. 699), "as a means of smoothing the spectral shape and of making the outcome more realistic psycho-acoustically, the spectrum is then passed through a filterbank based on the non-linear Mel scale. The logarithms of the filter coefficients are transferred to the cepstrum by application of the Discrete Cosine Transform. The resulting vectors are now called cepstral coefficients."

system has been investigated before, albeit almost exclusively using MZ twins as participants.

When comparing our results with previous findings by other authors who have tested the same automatic system with twins, we have been able to corroborate the widely reported finding in the ASR literature that some speakers are simply more easily identified than others. The 9.9% EER in our study corresponding to two out of 12 MZ twins who would be misidentified, is comparable to the 11% EER in Künzel (2010), indicating that confusion or non-distinction between twins occurred. The issue of the "striking performance inhomogeneities among speakers within a population" was already raised by Doddington et al. (1998) and we already referred to it in the glottal analysis described in San Segundo (2014), where some cases (16.6%) were found of speakers exhibiting large self-unlikeness (i.e., they were very dissimilar when comparing their first and second recording session).

To sum up, testing the performance of an ASR system using identical twins implies a strong reduction of inter-speaker variation and, as explained by Künzel (2010), this is a most challenging task since "the *a priori* chances for a target voice to be very similar to the reference voice is much larger than within a set of unrelated speakers" (2010, p. 269). We agree with him in considering that "a system that identifies an identical twin without falsely accepting the other twin is probably fit for use in the forensic environment" (Künzel, 2010, p. 274). The explanation for this seems logical: the system works even when it is being tested in a disadvantageous situation, which could be compared with a situation where there is channel distortion or cross-language samples to compare. All these are challenging situations. However, a real case where twins' voices ought to be compared is not the most frequent situation in a forensic setting, basically because of the low incidence of twin births (rate of identical twins is four per thousand; fraternal rate is 22.8 per thousand). Yet, the importance of investigating twins' voices goes beyond this pragmatic view, i.e., it is relevant per se, regardless of how many real cases involve the comparison of twins. First, the comparative study of MZ and DZ twins can reveal the genetic influence of the parameters under study (see EEA, Section 1.1). Hence the importance of carrying out studies with both types of twins, not only MZ twins. The finding that certain voice parameters are genetically marked entails a good performance of any system that would be based on such parameters because the typical speakers for comparison would be usually genetically unrelated, which means that the system would be good at separating them. Second, the consideration of further types of kinship relationships, apart from MZ and DZ twins, such as non-twin siblings can help clarify certain under-researched issues, such as the interplay between genetic and environmental influences in voice.

From the results of our investigation, we suggest that the cepstral parameters on which the automatic system Batvox^{TM} is based are genetically influenced. It is well known that these features relate to the geometry of the vocal tract, so some physical similarity between twins is expected to be encoded in DNA. Yet, the different use and configuration of the vocal apparatus could be exploited by twins in different ways, which could leave a generous margin for IP variation (Loakes, 2006; Nolan & Oh, 1996). These different usage preferences—more related to learned aspects than to inborn characteristics—might be the key to explain the two out of 12 twin cases that were misidentified by the system, accounting for the 9.9% EER.

All in all, as a direction for future work, it has not been mentioned so far that neither the group of MZ twins nor the DZ twin group are homogenous as far as their genes are concerned. MZ twins can be monochorionic or dichorionic, depending on whether they share the same placenta or have two different placentas instead; they can also be monoamniotic or diamnotic, depending on whether they share the same amniotic sac or not. How this can affect the differences found between one twin pair and another, as well as the influence of epigenetics in twin

differences, has not been fully addressed in twin voice literature yet. For instance, the fact that spontaneous mutations tend to occur more often in dichorionic MZ twins makes them more likely to differ genetically than monochorionic MZ twins (see Stromswold, 2006). Whether the existence of different types of MZ twins affects their voice similarity or not is an open research question, which, in any case, would require specific DNA testing to obtain detailed information about the zygosity of the twin pairs.

As regards epigenetics, future research focusing on twins' voices should pay more attention to this concept, which we briefly introduced in Section 1.1. Although only two "forces" are typically mentioned in the twin literature to explain the (dis)similarities in twins voices, namely, genetic and environmental factors, the often-neglected third factor, i.e, epigenetics (which explains the alteration in the expression of specific genes caused by mechanisms other than changes in the underlying DNA sequence) may play an important role in our understanding of the striking dissimilarities found for some twin pairs.

6. ACKNOWLEDGEMENTS

This research has been possible thanks to a doctoral grant awarded to the first author by the Spanish Ministry of Education (*Beca FPU-Programa Nacional de Formación de Profesorado Universitario, BOE 11-07-2009*) and also thanks to a grant awarded by the IAFPA (International Association for Forensic Phonetics and Acoustics) to the project "Forensic comparison of Spanish twins and non-twin brothers".

7. REFERENCES

Agnitio Voice Biometrics (2013). Batvox 4.1 Basic User Manual [Computer software].

Ariyaeeinia, A., Morrison, C., Malegaonkar, A., & Black, S. (2008). A test of the effectiveness of speaker verification for differentiating between identical twins. *Science & Justice*, *48*(4), 182–186. http://dx.doi.org/10.1016/j.scijus.2008.02.002

Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., … & Reynolds, D. A. (2004). A tutorial on text-independent speaker verification. *EURASIP Journal on Advances in Signal Processing*, *2004* (4), 1–22. http://dx.doi.org/10.1155/S1110865704310024

Brümmer, N., & du Preez, J. (2006). Application-independent evaluation of speaker detection. *Computer Speech & Language*, *20*(2–3), 230–275. http://dx.doi.org/10.1016/j.csl.2005.08.001

Campbell, W. M., Campbell, J. P., Reynolds, D. A, Singer, E., & Torres-Carrasquillo, P. A. (2006). Support vector machines for speaker and language recognition. *Computer Speech & Language*, *20*(2–3), 210–229. http://dx.doi.org/doi:10.1016/j.csl.2005.06.003

Charlet, D., & Lecha, V. P. (2007). Voice biometrics within the family: Trust, privacy and personalisation. In J. Filipe, H. Coelhas, M., & Saramago, (Eds.): *E-business and telecommunication networks: Second International Conference, ICETE 2005, Vol. 3* (pp. 93–100). Berlin: Springer. http://dx.doi.org/10.1007/978-3-540-75993-5_8

Debruyne, F., Decoster, W., Van Gijsel, A., & Vercammen, J. (2002). Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice*, *16*(4), 466–471. http://dx.doi.org/10.1016/S0892-1997(02)00121-2

Del Abril Alonso, Á., Ambrosio Flores, E., de Blas Calleja, M. d. R., Caminero Gómez, Á., García Lecumberri, C., & de Pablo González, J. M. (2009). *Fundamentos de psicobiología*. Madrid: Sanz y Torres.

Doddington, G., Liggett, W., Martin, A., Przybocki, M., & Reynolds, D. (1998). SHEEP, GOATS, LAMBS and WOLVES: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. *Proceedings of the International Conference on Spoken Language* (*ICSLP '98*), paper 0608.

Drygajlo, A. (2007). Forensic automatic speaker recognition [Exploratory DSP]. *IEEE Signal Processing Magazine*, *24*(2), 132–135. http://dx.doi.org/10.1109/MSP.2007.323278

Feiser, H. S. (2009). *Acoustic similarities and differences in the voices of same-sex siblings*. Paper presented at the 18th Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA), Cambridge, UK.

Felson, J. (2014). What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Social Science Research*, *43*, 184–199. http://dx.doi.org/doi:10.1016/j.ssresearch.2013.10.004

Forrai, G., & Gordos, G. (1983). A new acoustic method for the discrimination of monozygotic and dizygotic twins. *Acta paediatrica Academiae Scientiarum Hungarica*, *24*(4), 315–322.

Foulkes, P., & French, J. P. (2012). Forensic speaker comparison: A linguistic–acoustic perspective. In P. Tiersma & L. M. Solan (Eds.), *Oxford handbook of language and law*, 557–572. Oxford: Oxford

University Press. http://dx.doi.org/10.1093/oxfordhb/9780199572120.013.0041

Galton, F. (1875). The history of twins, as a criterion of the relative powers of nature and nurture (rev. ed.). *Journal of the Anthropological Institute of Great Britain and Ireland*, *5*, 391–406.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge: Cambridge University Press.

Gómez-Vilda, P., Fernández-Baillo, R., Nieto, A., Díaz, F., Fernández-Camacho, F. J., Rodellar, V., … & Martínez, R. (2007). Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. *Journal of Voice*, *21*(4), 450–476. http://dx.doi.org/10.1016/j.jvoice.2006.01.008

Gonzalez-Rodriguez, J., Fierrez-Aguilar, J., Ortega-Garcia, J. (2003). Forensic identification reporting using automatic speaker recognition systems. *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP '03*), *2*, 93–96.

Gonzalez-Rodriguez, J., Rose, P., Ramos, D., Toledano, D. T., & Ortega-Garcia, J. (2007). Emulating DNA: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(7), 2104–2115. http://dx.doi.org/10.1109/TASL.2007.902747.

Homayounpour, M. M., & Chollet, G. (1995). Discrimination of voices of twins and siblings for speaker verification. In *Proceedings of the 4th European Conference on Speech Communication and Technology* (*EUROSPEECH 1995*), 345–348.

Jain, A. K., Prabhakar, S., & Pankanti, S. (2002). On the similarity of identical twin fingerprints. *Pattern Recognition*, 35(11), 2653–2663. http://dx.doi.org/10.1016/S0031-3203(01)00218-7

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, *2*(4), 671–711. http://dx.doi.org/10.1111/j.1749-818X.2008.00066.x

Kenny, P., Boulianne, G., Ouellet, P., & Dumouchel, P. (2005). Factor analysis simplified. *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05*), *1*, 637–640.

Kim, K. (2009). *Automatic Speaker Identification of Korean Male Twins*. Paper presented at the 18th Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA), Cambridge, UK.

Kinnunen, T., & Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication*, *52*(1), 12–40. http://dx.doi.org/10.1016/j.specom.2009.08.009

Kong, A. W.-K., Zhang, D., & Lu, G. (2006). A study of identical twins' palmprints for personal verification. *Pattern Recognition*, *39*(11), 2149–2156. http://dx.doi.org/doi:10.1016/j.patcog.2006.04.035

Künzel, H. J. (1994). Current approaches to forensic speaker recognition. In *Proceedings of the ESCA Workshop on Automatic Speaker Recognition, Identification, and Verification*, 135–141.

Künzel, H. J. (2010). Automatic speaker recognition of identical twins. *International Journal of Speech, Language and the Law*, *17*(2), 251–277. http://dx.doi.org/10.1558/ijsll.v17i2.251

Künzel, H. J., & Alexander, P. (2014). Forensic automatic speaker recognition with degraded and enhanced speech. *Journal of the Audio Engineering Society*, *62*(4), 244–253. http://dx.doi.org/10.17743/jaes.2014.0014

Labov, W. (1972). The transformation of experience in the narrative syntax. In W. Labov, *Language in the inner city: Studies in the Black English Vernacular* (pp. 354–396). Philadelphia, PA: University of Philadelphia Press.

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

Martino, D., Loke, Y. J., Gordon, L., Ollikainen, M., Cruickshank, M. N., Saffery, R., Craig, J. M. (2013). Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biology*, *14*(5): R42. http://dx.doi.org/10.1186/gb-2013-14-5-r42

Meuwly, D. (2001). *Reconnaissance de locuteurs en sciences forensiques: l'apport d'une approche automatique*. PhD dissertation, University of Laussane.

Morrison, G. S. (2010). Forensic voice comparison. In I. Freckelton & H. Selby (eds.), *Expert evidence* (Chapter 99). Sydney: Thomson Reuters.

Morrison, G. S., & Kinoshita, Y. (2008). Automatic-type calibration of traditionally derived likelihood ratios: Forensic analysis of Australian English /o/ formant trajectories. In *Proceedings of the 9th INTERSPEECH Conference 2008*, 1501–1504.

Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Nolan, F. (1997). Speaker recognition and forensic phonetics. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 744–767). Oxford: Blackwell. http://dx.doi.org/10.1111/b.9780631214786.1999.00025.x

Nolan, F., & Oh, T. (1996). Identical twins, different voices. *International Journal of Speech Language and the Law*, *3*(1), 39–49. http://dx.doi.org/10.1558/ijsll.v3i1.39

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the*

Acoustical Society of America, *119*(4), 2382–2393. http://dx.doi.org/10.1121/1.2178720

Philips, T. (2008). The role of methylation in gene expression, *Nature Education*, *1*(1), 116.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. http://dx.doi.org/10.1017/S0140525X04000056

Przybocki, M. A., Martin, A. F., & Le, A. N. (2007). NIST speaker recognition evaluations utilizing the Mixer corpora—2004, 2005, 2006. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(7), 1951–1959. http://dx.doi.org/10.1109/TASL.2007.902489

Przybyla, B. D., Horii, Y., & Crawford, M. H. (1992). Vocal fundamental frequency in a twin sample: Looking for a genetic effect. *Journal of Voice*, *6*(3), 261–266. http://dx.doi.org/doi:10.1016/S0892-1997(05)80151-1

Ramos, D. (2007). *Forensic evaluation of the evidence using automatic speaker recognition systems* (Doctoral dissertation). Universidad Autónoma de Madrid. Retrieved from http://hdl.handle.net/10486/1774

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, *10*(1–3), 19–41. http://dx.doi.org/10.1006/dspr.1999.0361

Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, *3*(1), 72–83. http://dx.doi.org/10.1109/89.365379

Rose, P. (2002). *Forensic speaker identification*. London: Taylor & Francis.

Rose, P. (2006). Technical forensic speaker recognition: Evaluation, types and testing of evidence. *Computer Speech & Language*, *20*(2–3), 159–191. http://dx.doi.org/doi:10.1016/j.csl.2005.07.003

San Segundo, E. (2010a). Parametric representations of the formant trajectories of Spanish vocalic sequences for likelihood-ratio-based forensic voice comparison. *The Journal of the Acoustical Society of America*, *128*(4), 2394. http://dx.doi.org/10.1121/1.3508586

San Segundo, E. (2010b). Variación inter e intralocutor: Parámetros acústicos segmentales que caracterizan fonéticamente a tres hermanos. *Interlingüística*, *21*, 352–363.

San Segundo, E. (2012). *Glottal source parameters for forensic voice comparison: An approach to voice quality in twins' voices*. Paper presented at the 21st Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA), Santander, Spain.

San Segundo, E. (2013a). *Guess who is laughing: A perceptual experiment on twin and non-twin siblings' identification*. Paper presented at the 31st International Conference AESLA (Asociación Española de Lingüística Aplicada). San Cristóbal de La Laguna: Universidad de La Laguna.

San Segundo, E. (2013b). A phonetic corpus of Spanish male twins and siblings: Corpus design and forensic application. *Procedia–Social and Behavioral Sciences*, *95*, 59–67. http://dx.doi.org/doi:10.1016/j.sbspro.2013.10.622

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*, Ph.D. thesis, Consejo Superior de Investigaciones Científicas-Universidad Internacional Menéndez Pelayo, Spain.

San Segundo, E. (2015). 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 [Thesis Abstract], *International Journal of Speech Language and the Law*, *22*(2), 249–253. http://dx.doi.org/10.1558/ijsll.v22i2.28821

San Segundo, E., & Gómez-Vilda, P. (2013). Voice biometrical match of twin and non-twin siblings. In C. Manfredi (Ed.), *Models and analysis of vocal emissions for biomedical applications: 8th International Workshop, Firenze, Italy, 2013*, (pp. 253–256). Retrieved from http://digital.casalini.it/9788866554707

San Segundo, E., Gómez-Vilda, P. (2015). 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*(2), 22–41.

Sataloff, R. T. (1995). Genetics of the voice. *Journal of Voice*, 9(1), 16–19. http://dx.doi.org/doi:10.1016/S0892-1997(05)80218-8

Scheffer, N., Bonastre, J.-F., Ghio, A., & Teston, B. (2004). Gémellité et reconnaissance automatique du locuteur. *Actes des XXV Journées d'Étude sur la Parole* (*JEP*), 445–448.

Segal, N. L. (1993). Implications of twin research for legal issues involving young twins. *Law and Human Behavior*, *17*(1), 43–58.

Srihari, S., Huang, C., & Srinivasan, H. (2008). On the discriminability of the handwriting of twins. *Journal of Forensic Sciences*, *53*(2), 430–446. http://dx.doi.org/10.1111/j.1556-4029.2008.00682.x

Stromswold, K. (2006). Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition*, *101*(2), 333–384. http://dx.doi.org/doi:10.1016/j.cognition.2006.04.007

Tomblin, J. B., & Buckwalter, P. P. (1998). Heritability of poor language achievement among twins. *Journal of Speech, Language, and Hearing Research*, *41*, 188–189. http://dx.doi.org/doi:10.1044/jslhr.4101.188

Trouvain, J., & Truong, K. P. (2012). *Convergence of laughter in conversational speech: Effects of*

quantity, temporal alignment and imitation. Paper presented at the International Symposium on Imitation and Convergence in Speech, Aix-en-Provence, France.

van Leeuwen, D. A., & Brümmer, N. (2007). An introduction to application-independent evaluation of speaker recognition systems. In C. Müller (Ed.), *Speaker classification I: Fundamentals, features, and methods* (pp. 330–353). Heidelberg: Springer–Verlag. http://dx.doi.org/10.1007/978-3-540-74200-5_19

Van Lierde, K. M., Vinck, B., De Ley, S., Clement, G., & Van Cauwenberge, P. (2005). Genetics of vocal quality characteristics in monozygotic twins: A multiparameter approach. *Journal of Voice*, *19*(4), 511–518. http://dx.doi.org/10.1016/j.jvoice.2004.10.005

Weirich, M., & Lancia, L. (2011). Perceived auditory similarity and its acoustic correlates in twins and unrelated speakers. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 17-Hong Kong*), 2118–2121.

Wolf, J. J. (1972). Efficient acoustic parameters for speaker recognition. *The Journal of the Acoustical Society of America*, *51*(6B), 2044–2056. http://dx.doi.org/10.1121/1.1913065