Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 00 (2013) 000–000

Procedia
Social and Behavioral Sciences

www.elsevier.com/locate/procedia

V International Conference on Corpus Linguistics (CILC2013)

A Phonetic Corpus of Spanish Male Twins and Siblings: Corpus Design and Forensic Application

Eugenia San Segundo Fernández*

Phonetics Laboratory, Spanish National Research Council, Calle Albasanz 26-28, Madrid 28037 Spain

Abstract

Having noted a lack of oral corpora of Spanish twins' voices, we have collected a database which consists of the following male speakers: 24 monozygotic twins, 10 dizygotic twins, 8 non-twin brothers, and 12 non-twin reference-population speakers. Twenty-two voice-related studies about twins have been reviewed in order to account for the average sample size of previous twin corpora, as well as their data collection method. Within a forensic-phonetic approach, the study of twins' voices is particularly relevant, as these (especially monozygotic twins) represent the most extreme physical similarity in human beings. Distinguishing them is therefore a challenge in speaker identification.

Selection and peer-review under responsibility of CILC2013.

Keywords: phonetics; corpus design; twins, biometrics; forensic; Spanish

1. Introduction

This brief introduction has a two-folded aim: first describing the two types of twins that, together with the non-twin siblings and reference population, make up this corpus, and secondly, introducing the discipline of forensic phonetics, in which the study of twins' voices is especially relevant.

* Corresponding author. Tel.: +34-628-03-04-21
E-mail address: eugeniasansegundo@gmail.com

Selection and peer-review under responsibility of CILC2013.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 00 (2013) 000–000

Procedia
Social and Behavioral Sciences

www.elsevier.com/locate/procedia

1.1. Types of twins and the importance of the twin method

It seems useful to begin this section by noting that there are two basic types of twins:

- Monozygotic (MZ) or identical twins occur when a single egg is fertilized to form one zygote, which then divides into two separate embryos.
- Dizygotic (DZ) or non-identical twins usually occur when two eggs are independently fertilized by two different sperm cells.

A more accurate but complex classification of twins would take into account that there are a few subgroups within MZ and DZ twins; we can distinguish twins who develop within the same or separate amniotic cavities, twins who have distinct or the same chorions, and a couple of further distinctions. However, for the sake of simplification, and for our purposes in this paper, it seems enough to consider a general classification between MZ and DZ twins. While the former share the 100% of their genes, fraternal twins (dizygotic, DZ) share half their genetic information.

These are only two of the four types of speakers who have taken part as subjects in our corpus. Non-twin siblings are the third type of speakers. It is important to take into account that full siblings (i.e. of the same father and the same mother) are genetically the same as DZ twins: they share 50% of their genes. The fourth group of speakers who make up our corpus has been called the reference population[1].

The classical twin method compares the resemblance of MZ and DZ twins and is widely used in fields such as medicine and psychology to study certain diseases and to estimate the environmental and heritable factors that contribute to certain behavioral traits or "complex system of features" (Decoster, Van Gysel, Vercammen, & Debruyne, 2001), like voice. As Segal (1990) states, there are several designs that have been used in twin research: the classic twin method and variations of this. The traditional one, which compares resemblance within MZ twin pairs to resemblance within DZ twin pairs, assumes equal environmental influences for both types of twins: "Greater resemblance within MZ twin pairs, relative to DZ twin pairs, is consistent with (although not proof of) a genetic explanation for the trait under investigation" (Segal, 1990: 613).

[1] The relevance of the reference population (friends or work colleagues in this case) for Likelihood-Ratio-based forensic studies has been explained elsewhere (Morrison, 2010). Besides fulfilling the age and dialect criteria, the only requisite for their participation in this study was that they have to come to the recordings in pairs either with a friend or a work colleague. The importance of this requisite lies in the search for a speaking style similar, and thus comparable, to that found in the conversations between twins, usually characterized by their spontaneity due to a close long-term relationship.

Selection and peer-review under responsibility of CILC2013.

1.2. Forensic phonetics

Stemming from the definition of forensics (in general) as the use of scientific knowledge to legal problems, such as the analysis of physical evidence from a crime scene, forensic phonetics is then the application of phonetics aimed at solving legal problems, being the most typical example one of voice comparison. In such a case, the forensic phonetician compares the voice of the offender with the voice of a suspect or several suspects. Although the goal is eventually identifying the offender, nowadays the term "speaker comparison" is preferred to "speaker identification" since the specific task of the forensic phonetician is offering an answer to the question "How much more likely are the observed differences between the known and questioned samples to occur under the hypothesis that the questioned sample has the same origin as the known sample than under the hypothesis that it has a different origin? The answer to this question is quantitatively expressed as a likelihood ratio." (Morrison, 2010: 16-17).

Imagine a burglar telling a colleague, on the phone, the details of a robbery. This conversation is intercepted by the police and, although the speaker may deny being one of the interlocutors of the conversation, he becomes a suspect of an actual theft which has been carried out some days after that conversation. The police would then obtain a voice sample of this speaker, together with the voice samples of more suspects who may be involved in the case. What the phonetician would be requested to do in this situation is analyzing and comparing the different (offender and suspect/s) voices with the goal of identifying the offender. There are some other tasks which forensic phonetics deals with (e.g. voice line-ups, disputed utterances, recordings' authentication, or speaker profiling). However, it is the task of speaker comparison in which the study of twins is most useful.

Twins (especially MZ pairs) represent the most extreme physical similarity in human beings. Since this resemblance also applies to their voice, distinguishing them is a challenge in speaker identification. In real casework, it is not extremely often to find twins as criminals but it is also not so unlikely as one may *a priori* think. The importance of studying twins' voices in forensic phonetics goes beyond this application, which is not a trivial question in any case[2]. The key purpose of carrying out research on twins' voices is that it could shed light on the limits of between-speaker and within-speaker variation.

As early as 1972 Wolf established some criteria for selecting a forensic-phonetic parameter: among others, it should yield the highest possible between-speaker variability and the minimum within-speaker variability. In other words, this means that certain parameters should show more differences when comparing two voice samples of the same woman recorded in two different moments than when comparing one of her voice samples with the voice of another woman. Around 40 years after Wolf (1972), there are many different phonetic and acoustic parameters that are not only studied by researchers all around the world but also considered in forensic casework (Gold & French, 2011). However, much research is still needed to find out the most useful and robust ones for speaker identification. There is yet some agreement that a single parameter is not enough to identify a speaker and probably it will never be. Indeed, the best methodology seems to be a combined approach. Our own corpus, as we will see in next pages, is actually designed to embrace different acoustic parameters from mixed or combined perspectives, and again, which is more important, to test these parameters in very challenging conditions of speaker similarity.

[2] In a recent piece of news (dated from 13th February 2013 in the Spanish newspaper *El País*) we could read about the arrest of two MZ twins in France charged of six rapes and sexual assaults. Even though the victims claimed that the aggression took place by only one person, the police could not determined, on the basis of the DNA found in the sperm cell, which one of the two twins committed the crime. The reason set out by the police was that the DNA is the same for identical twins. Yet the newspaper article specified that the French police would need to pay around one million Euros for a very complex and specific DNA test which would reveal the identity of the offender. However, according to other studies, no DNA test seems to allow the distinction of identical twins, at least in cases such as the one reported in Künzel (2011: 274). Hence the importance of carrying out research in disciplines like forensic phonetics, which relies on pieces of evidence other than DNA, like voice samples, which have not been proved to be totally identical in MZ twins.

2. Literature review

Explaining the impact of twin research in forensic disciplines different from phonetics lies beyond our scope. The list of recent papers in diverse forensic fields which have shown an interest in twins is quite large: we find research on the discriminability of twins on the basis of their handwriting (Srihari, Huang, & Srinivasan, 2008), their DNA (Rubocki, McCue, Duffy, Shepard, Shepherd, & Wisecarver, 2001) and their fingerprints (Jain, Prabhakar, & Pankanti, 2002).

Our interest in this section centers in reviewing the main voice-related studies about twins. This will be especially useful to account for the average sample size of previous twin corpora and their data collection method. As concerns the methodological perspectives followed, a first classification should be made distinguishing between articulatory, acoustic and perceptual approaches to the voices of twins. It goes without saying that some research papers combine several perspectives. Besides, the topic of twins' voices has also been considered under an automatic approach. We will limit our literature review to the acoustic approach, as it is the most extensive and (together with the automatic perspective) the most closely related to the forensic task of speaker comparison. For a more detailed account of studies focusing on twins' voices, see San Segundo (2013b). These are the main aspects that we would like to highlight after having carried out a literature review of around 22 acoustic studies:

- The degree of similarities and dissimilarities found in twins' voice and speech is not uniform across twin pairs (Loakes, 2006). Furthermore, results depend not only on the specific twin pair under consideration but also on the acoustic parameter selected for discrimination. Therefore, many different parameters have been studied to assess twins' (dis)similarities. In our literature review we have found the following ones: *fundamental frequency (f0)* and related parameters, such as Vocal Fundamental Frequency and long-term f0 (Przybyla, Horii, & Crawford, 1992; Decoster, Van Gysel, Vercammen, & Debruyne, 2001; Debruyne, Decoster, Van Gijsel, & Vercammen, 2003); *spectral parameters* (Alpert, Kurtzberg, Pilot, & Friedhoff, 1963; Flach, Schwickardi, & Steinert, 1968); *Voice Onset Time* (Forrai & Gordos, 1983; Ryalls, Shaw, & Simon, 2004); *coarticulation patterns* (Nolan & Oh, 1998; Whiteside & Rixon, 2003); *temporal parameters* (Whiteside & Rixon, 2001); *formant patterns* (Loakes, 2006) and *formant transitions* (Weirich, 2011); *plosives frication* (Loakes, 2006); voice quality parameters like *jitter*, *shimmer* and *HNR* (Fuchs, Oeken, Hotopp, Täschner, Hentschel, & Behrendt, 2000; Van Lierde, Vinck, De Ley, S., Clement, G., & Van Cauwenberge 2005; Cielo, Agustini, & Finger, 2010).
- As regards the average sample size of the corpora used for the above mentioned studies, only a few of them consider a relatively large amount of speakers. While the average number of subjects recruited is 26 twin pairs[3], it is worth mentioning that twin participation ranges from 1 pair to 202 pairs, being the standard deviation 41 and the mode 1. Since we find only exceptionally a high number of speakers, like 202, 117, 62 or 60, we think that the median (12 twins) is a more truthful indicator of the number of twin participants that we usually find in voice-related studies.
- Most of the studies focus on both MZ and DZ pairs, still many of them recruited only MZ twins as participants, while only a minority of the studies reviewed have considered the analysis of non-twin siblings. Among those related to siblings from an acoustic perspective, we find a common interest in comparing MZ twin pairs and sex-matched siblings in Whiteside & Rixon (2003) and Papay (2007) while Feiser (2009) focused only on siblings. As far as we know, no study has considered so far the inclusion of both types of twins together with siblings as a relevant population to study (dis)similarities of voice features.
- Concerning the forensic application of twins' studies in the phonetic realm, only Nolan & Oh (1998), Loakes (2006) and Feiser (2009) stem clearly from a forensic point of view, as they themselves state it. Yet, for the corpus design of their studies, only Loakes (2006) takes into account certain forensic-realistic criteria, like presence of channel-mismatch in the recordings or existence of non-contemporaneous speech samples.

[3] For this count, we have considered all the studies about twins reviewed in San Segundo (2013b), not distinguishing between those related to perception, acoustics or articulation.

- The findings of the different studies carried out so far are diverse and depend on their main research objectives, which could be summarized as: (1) trying to find a genetic component in the variation of certain voice parameters *(e.g*. Alpert, Kurtzberg, Pilot, & Friedhoff, 1963; Przybyla, Horii, & Crawford, 1992; Debruyne, Decoster, Van Gijsel, & Vercammen, 2003); or else, in a forensic context, (2) trying to test whether it is possible to identify a speaker, distinguishing him/her from his/her co-twin (Loakes, 2006; Nolan & Oh, 1998). If we look at the results, for the first research objective, there is some agreement in the difficulty to discern the influence of genetic factors from the influence of shared environment (at least for certain parameters, like *f0*: Debruyne, Decoster, Van Gijsel, & Vercammen, 2003). Regarding the forensic application, although some findings might be language-dependant, for example the good discriminatory results of formant transitions in certain German sequences (Weirich, 2011), other results might be more universal, like Loakes' (2006) finding that F3 is the most speaker-specific formant frequency.
- A final word should be dedicated to the methods used for the recruitment of twins. In most cases, no information is available about this specific aspect in the publications reviewed. This is not a trivial issue since for the collection of any twin corpus, like the one presented here, researchers usually face many difficulties to find a population sample large enough to provide reliable statistical results. This is due to the scarceness of twins in the general population. In most articles where the number of twins recruited were 40 or above, a previously-collected twin register or database was available for research, like in Gedda, Bianchi, Bianchi-Neroni (1960), Przybyla, Horii, & Crawford (1992), or Van Lierde, Vinck, De Ley, S., Clement, G., & Van Cauwenberge (2005). In two cases (Scheffer, Bonastre, Ghio, & Teston, 2004; and Künzel, 2011), the authors used twins' voice samples previously gathered by TV programs which had showed an interest in twins' similarities. In all other cases, either the data collection method is unspecified or the corpus is supposed to have been collected "ad hoc" for the publication.

3. Corpus design

Having noted a total lack of oral corpora of Spanish twins' voices, we have collected our own corpus which consists of the following male speakers: 24 MZ twins, 10 DZ twins[4], 8 non-twin brothers and 12 speakers with no kinship relationship, which will serve as a reference population. The ages of the speakers recruited for this corpus ranged between 18 and 52 years old (mean age: 28.96). They had no voice pathologies or hearing difficulties. Among all the possible speaker recruitment methods, we used these: mailing lists in several universities and research centers, notice boards in places like public libraries, *Facebook* events, *Twitter* messages, and the so-called snowball method, mentioned in Debruyne, Decoster, Van Gijsel, & Vercammen (2003) and consisting in obtaining new twins' contact addresses from participating twins, already recruited. The participants in our corpus could be compensated for their participation thanks to a grant awarded to this research project by the IAFPA (International Association for Forensic Phonetics and Acoustics).

Concerning the dialectal aspects, the language variety spoken by all the subjects was Standard Peninsular Spanish. They were all native speakers of this variety, coming from different regions in Spain. However, the majority of them were born and lived either in Madrid or in nearby cities, like Ávila, Salamanca, etc. The dialectal variable was therefore tightly controlled. Besides, one of the requirements for recruiting the subjects of our study was that they had to be male speakers. The reasons for establishing this criterion were:

- In real forensic cases, there is a higher incidence of crimes committed by men.
- For the study of one sex group, it is necessary (according to the method we have adopted) to record not only twins and brothers but also a reference population of normophonic speaker of the same sex. Therefore, the

[4] The twins were asked to check (via official documents) whether they were MZ or DZ twins. For cases where zygosity was unclear, DNA testing was conducted (by the *Instituto Nacional de Toxicología y Ciencias Forenses* -Spanish National Toxicology and Forensic Sciences Institute).

inclusion of female speakers in this study would have implied the consideration of a further variable which would have doubled the number of speakers necessary: MZ and DZ twins, sisters and a group of normophonic female speakers. Consequently, this study is limited to one sex.

- Female voices are more difficult to study in a forensic context where the voice samples for comparison usually come from telephone-intercepted calls. Since the pitch of female speakers is higher than that of male speakers, the telephone filter might affect worse female speech, due to the fact that the telephone channel is a band-pass filter which cuts off frequencies below 300 Hz and above 3,400 kHz. Yet, in general some formant frequencies are more affected than others, especially F3 and F4.

For the design of our corpus, we have distinguished four types of tasks to be carried out by the speakers participating in this study:

- Semi-structured spontaneous conversation between twins: They hold a telephone conversation (see figure 1) of about 10 minutes talking about some topics suggested by the researcher.
- Fax exchange to elicit specific vocalic sequences: The purpose of this task is that, throughout an exchange of fax samples, the speakers use certain words of interest mainly because they contain phonetic units that we want to study, namely diphthongs and hiatuses, which have already been considered forensically relevant in Spanish (San Segundo, 2010).
- Reading of two phonetically-balanced texts, which are read in two different ways: one without any kind of disguise (normal reading) while in the other type of reading the speakers have to pinch their nose.
- Informal interview with the researcher, intended to elicit a more formal speaking style, among other purposes.

Finally, the corpus is made up of a "vocal control technique" consisting in asking the speakers to sustain both the vowel [a] and the consonant [s] sound as long as possible. With this vocal technique, which was repeated three times, two measures could be calculated: *Maximum Phonation Time (MPT)* and *s/a Ratio*, "two traditionally popular indirect clinical measures of respiratory integrity and laryngeal valving efficiency" (Aronson & Bless, 2009: 148), which will be especially interesting for the analysis of voice quality parameters.

Following the recommendations in Morrison, Rose, & Zhang (2012), intended to suggest the collection of a corpus as forensically realistic as possible, three key aspects of our corpus are:

- Each speaker is recorded on two occasions, separated by 2-4 weeks. This is important in order to account for within-speaker variation.
- The recordings were made with high-quality microphones but, on a later step, they were also telephone-filtered – thus simulating real forensic situations, where recordings are intercepted by the police and they are telephone-degraded.
- An effort is made to elicit different speaking styles: spontaneous style between twins and friends, formal interview-style between the speaker and the researcher, reading style, and so on. Besides, this also allows for eliciting different acoustic parameters, such as naturally sustained vowels in hesitation speech or specific vocalic sequences in the fax-exchange task.

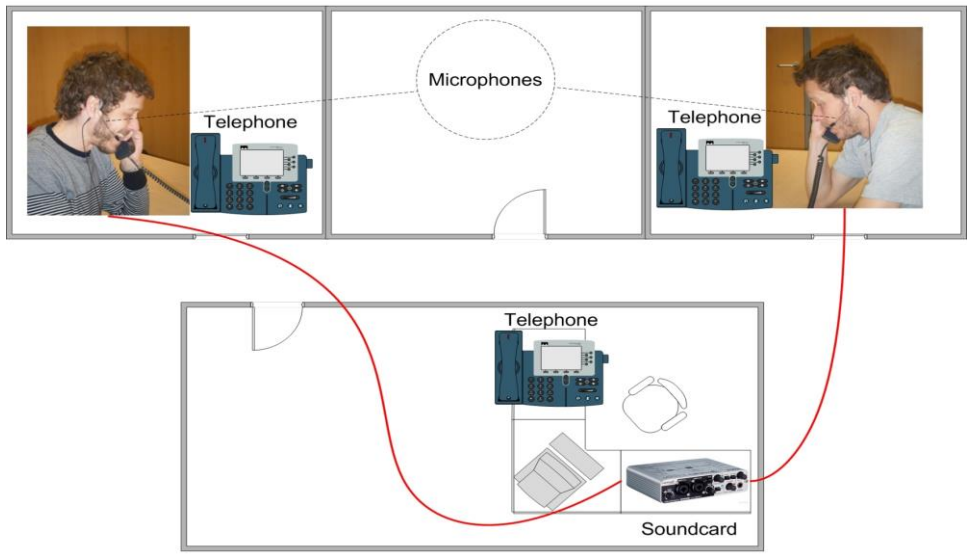

Fig. 1. Data collection set-up

In figure 1 we can see the set-up for the recordings: each speaker is placed in a different room, where they are acoustically isolated from each other but can communicate via telephone. There is another room from where the researcher monitors the recordings.

4. Examples of forensic usages of this corpus

Using the speech samples recorded in this corpus, some investigations have already been carried out (San Segundo, 2012; San Segundo, 2013a; Gil & San Segundo, 2013). For instance, based on the fourth task of the corpus, it was possible to obtain several pause fillers, from which biomechanical estimates of both the vocal fold body and the vocal fold cover could be extracted (San Segundo, 2012). In order to extract voice source properties (jitter, shimmer, estimates of the dynamic mass and stiffness of the vocal folds, etc.) we used the software package GLOTTEX® (Gómez Vilda et al., 2008). 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. Elicitation of hesitation speech was possible in the interview task since speakers had to do an effort to remember what they had been talking to his twin in the first corpus task, which took place 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, the "observer's paradox" (Labov, 1972) could be avoided. Therefore we ensured that by the end of the recording session the speakers had settled down to a pattern approximating their everyday interactional style. Since the results showed that at least the biomechanical estimates of both the vocal fold body and the vocal fold cover are forensically useful to distinguish twins and non-twin siblings, we propose the use of elicited hesitation speech (e.g. pause fillers) for forensic research. Its usefulness lies in the necessity of having recordings with relatively long vowels in which enough glottal cycles can be obtained for the posterior extraction of voice source features.

In San Segundo (2013a), our research interest lied in a type of non-verbal vocalization which has not received special attention in forensic phonetics yet: laughter. We tested listeners' performance in a perceptual task consisting in the identification of their own laughter and that of their siblings/twins and found that MZ twins correctly identified laughter stimuli in 67% of the cases (slightly above chance level), while DZ twins and brothers

obtained better results: 92% and 93% of identification rates, respectively. This different performance would be significant at $p<0.05$. However, the size of the sample in this study was not high enough to provide reliable significance results. We had hypothesized that MZ pairs would perform worse at this identification task than DZ pairs since MZ co-twins share 100% of their genetic information while DZ co-twins, as well as full brothers, share 50% of their genes. Yet further research is needed to investigate, on a larger sample of speakers, whether these results are due to genetic similarity or to shared social environment. As concerns the methodological aspects of this experiment, these are the details of the test design: Several months after the male twins (2 MZ twin pairs and 2 DZ twin pairs) and non-twin siblings (2 brother pairs) were recorded[5], they carried out a forced-choice identification task in which they listened to 24 stimuli consisting of laughter instances (6 x pair x 2 repetitions) extracted from their own recordings. They were asked to identify whose was the laugh at each time: his own or his sibling's. The main forensic relevance of this study lies in the fact that laughter (together with other non-linguistic features like tongue-clicking, audible breathing and throat clearing) is actually considered in forensic casework by the majority of international experts, as stated in Gold & French (2011). However, little research has been carried out to test the true idiosyncratic nature of these verbal vocalizations.

Finally, in Gil & San Segundo (2013), we made use of the third corpus task: reading of two phonetically-balanced texts, which were read in two different ways: one without any kind of disguise (normal reading) while in the other type of reading the speakers have to pinch their nose. We wanted to test how harmful this type of voice disguise is for speaker identification. For that purpose, we carried out a perceptual experiment with 28 listeners having to identify in a forced-choice test which speaker (among four possible ones) they were listening at each stimulus. The listeners had previously familiarized with the voices of these four speakers, extracted from the non-twin speakers of our corpus. The results of our study showed that there are no significant differences between the results obtained by the listeners in the perceptual test where the stimulus were extracted from pinched-nose speech samples and the results obtained in the perceptual test with normal-condition speech samples. Another purpose of this study was testing the forensic relevance of other variables, namely two psychological effects known as 'primacy' and 'recency'. We found that these did not affect the identification results of the listeners, at least in our experiment design.

5. Conclusions and further research

According to the literature review that we have carried out, it is clear that twins' studies are relevant in phonetics but much research is still needed in a particular field: forensic phonetics. That means that an effort should be made to fulfill forensically-realistic characteristics in the design and collection of future corpora, like considering telephone-degraded voice samples, different kinds of voice disguise or multiple-session recordings, which will allow intra-speaker comparisons. It seems especially necessary to investigate the different phonetic and acoustic parameters that we have reviewed but on a Spanish twin population, which has not been recruited with any phonetic purpose so far. From the fourth section of this article, some further conclusions can be drawn. With the oral corpus which is being presented, several presentations and publications have already appeared with the aim of characterizing the voices of MZ twins, DZ twins and non-twin siblings. The results obtained would hopefully help us understand better the influence of genetic endowment and environmental factors on certain voice features and to investigate the within-speaker and between-speaker variation, which is so important in forensic phonetics. Besides, the inclusion of a reference population in this corpus has allowed us to analyze listeners' performance in a perceptual test which does not take into account kinship relationships but which does consider other key aspects in the forensic application of phonetics, namely the influence of voice disguise in speaker identification (Gil & San Segundo, 2013).

In the future we consider a three-folded analysis of the twin and non-twin voices recorded in this corpus, comprising (1) a glottal source analysis from vowels naturally sustained in hesitation speech; (2) an analysis of formant transitions in 118 vocalic sequences, and (3) an automatic voice analysis using the automatic speaker

[5] The corpus task used here is the first one (semi-structured spontaneous telephone conversation with their respective siblings).

recognition system *Batvox*[6], which makes uses of around 120 seconds of spontaneous conversation per speaker. This multiple perspective will allow us to shed light on the limits of between-speaker and within-speaker variation, by testing different phonetic parameters under challenging conditions of speaker similarity, with the purpose of finding whether they may be forensically useful. For this aim, we will take into account both recording sessions of a speaker and not only their high quality recordings but also the telephone-filtered voice samples.

Furthermore, our intention is to encourage new research on twins in the field of linguistics, and more specifically in forensic phonetics. Therefore, we hope to achieve this purpose by means of collaborations with other researchers when this corpus becomes –upon request- an open resource available for scientists.

Acknowledgements

This research has been possible thanks to a grant awarded 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). I would also like to extend special thanks to Dr. Antonio Alonso from the *Instituto Nacional de Toxicología y Ciencias Forenses (Madrid).*

References

Aitken, C., & Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data. *Applied Statistics, 53 (4),* 109-122.

Alpert, M., Kurtzberg, R., Pilot, M., & Friedhoff, A. (1963). Spectral characteristics of the voices of twins. *Acta Genet. Med. Gemellol, 12 (4),* 335-342.

Aronson, A., & Bless, D. (2009). *Clinical voice disorders* (4th ed.). New York: Thieme Medical Publishers, Inc.

Cielo, C., Agustini, R., & Finger, L. (2010). Características vocais de gêmeos monozigóticos (Vocal features of monozygotic twins), *Revista CEFAC, 14 (6),* 1234-1241.

Debruyne, F., Decoster, W., Van Gijsel, A., & Vercammen, J. (2003). Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice, 16 (4),* 466-471.

Decoster, W., Van Gysel, A., Vercammen, J., & Debruyne, F. (2001). Voice similarity in identical twins. *Acta oto-rhino-laryngologica bel*., *55,* 49-55.

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

Flach, M., Schwickardi, H., & Steinert, R. (1968). Zur Frage des Einflusses erblicher Faktoren auf den Stimmklang (Zwillingsuntersuchungen). *Folia phoniat*., *20,* 369-278.

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

Fuchs, M., Oeken, J., Hotopp, T., Täschner, R., Hentschel, B., & Behrendt, W. (2000). Die Ähnlichkeit monozygoter Zwillinge hinsichtlich Stimmleistungen und akustischer Merkmale und ihre mögliche klinische Bedeutung. *HNO, 48,* 462-469.

Gedda, L., Bianchi, A., & Bianchi-Neroni, L. (1960). La voce dei gemelli. Prova di identificazione intrageminale della voce in 104 coppie (58 MZ e 46 DZ). *Acta Geneticae Medicae et Gemellologiae, 4 (2),* 121-130.

Gil, J., & San Segundo, E. (2013). El disimulo de la cualidad de voz en fonética judicial: un estudio perceptivo para un caso de hiponasalidad. In A. Palacios (Ed.), *Panorama de la Fonética Española Actual* (In Press).

Gold, E., & French, P. (2011). An International Investigation of Forensic Speaker Comparison Practices, 17th International Congress of Phonetic Sciences, Hong Kong, 17-21 August 2011.

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

Jain, A., Prabhakar, S., & Pankanti, S. (2002). On the similarity of identical twin fingerprints. *Pattern Recognition, 35(1),* 2653-63.

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

Labov, W. (1972). *Sociolinguistic Patterns*. Oxford: Blackwell.

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

[6] Manufactured by Agnitio S.A., Madrid, Spain. For detailed information, readers are referred to the following document: www.agnitio.es/ingles/files/BATVOXpresentation.PDF.

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

Morrison, G. (2010). Forensic voice comparison. In I. Freckelton & H. Selby (Eds.) *Expert Evidence* (Ch. 99). Sydney, Australia: Thomson Reuters.

Morrison, G., Rose, P., & Zhang, C. (2012). Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice. *Australian Journal of Forensic Sciences*, 44, 155-167.

Nolan, F., & Oh, T. (1998). Identical twins, different voices. *Forensic Linguistics, 3*, 39-49.

Papay, K. (2007). Hereditary phonetic parameters of the human voice. *Magyar Nyelvor (Hungarian Language Guardian), 131 (3)*, 306-315.

Przybyla, B., Horii, Y., & Crawford, M. (1992). Vocal fundamental frequency in a twin sample: looking for a genetic effect. *Journal of Voice*, 6(3), 261-266.

Rubocki, R. ,McCue, B., Duffy, K., Shepard, K., Shepherd, S., & Wisecarver, J. (2001). Natural DNA mixtures generated in fraternal twins in utero. *Journal of Forensic Science, 46 (1)*, 120-125.

Ryalls, J., Shaw, H., & Simon, M. (2004). Voice Onset Time Production in Older and Younger Female Monozygotic Twins. *Folia Phoniatr. Logop., 56*, 165-169.

San Segundo, E. (2010). Parametric representations of the formant trajectories of Spanish vocalic sequences for likelihood-ratio-based forensic voice comparison. *Journal of the Acoustical Society of America, 128*, 2394.

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

San Segundo, E. (2013a). Guess who is laughing: A perceptual experiment on twin and non-twin siblings' identification, 31st International Conference AESLA, 2013, La Laguna, Spain.

San Segundo, E. (2013b). PhD Dissertation (Unpublished).

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

Segal, N. (1990). The importance of twin studies for individual differences research. *Journal of Counseling and Development, 68*, 612-622.

Srihari, S., Huang, C. & Srinivasan, H. (2008) On the Discriminability of the Handwriting of Twins. Journal of Forensic Science, 53 (2): 430-446.

Van Lierde, K., 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.

Weirich, M. (2012). The influence of NATURE and NURTURE on speaker-specific parameters in twins' speech: Articulation, acoustics and perception. PhD thesis. Humboldt-Universität zu Berlin.

Whiteside, S., & Rixon, E. (2001). Speech Patterns of Monozygotic Twins: An Acoustic Case Study of Monosyllabic Words. *The Phonetician, 84*, 9-22.

Whiteside, S., & Rixon, E. (2003). Speech characteristics of monozygotic twins and a same-sex sibling: an acoustic case study of coarticulation patterns in read speech. *Phonetica, 60 (4)*, 273-297.

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