Matching Twin and non-Twin Siblings from Phonation Characteristics

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

[1]Phonetics Laboratory, Institute of Language, Literature and Anthropology
Spanish National Research Council (CSIC)
C/ Albasanz 26-28, 28037 Madrid, Spain
[1]NeuVox Laboratory, Center for Biomedical Technology, Universidad Politécnica de Madrid,
Campus de Montegancedo, s/n, 28223 Pozuelo de Alarcón, Madrid
e-mails: eugenia.sansegundo@cchs.csic.es,pedro@pino.datsi.fi.upm.es

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

Keywords: voice production, forensic pattern matching, phonation styles, glottal source features, twins.

1 Introduction

Recent studies in voice quality are conducted towards the evaluation of phonation performance in relation to either professional voice care, or in meta-acoustic knowledge (neurological deterioration, emotion detection, etc.) These fields of study are becoming more and more demanded nowadays. The aim of the present work is to study the similarities and differences of phonation characteristics in twins' voices, including monozygotic (MZ) as well as dizygotic (DZ) twins. A reference to previous work on twin voice quality analysis and vocal performance of interest for this research is that of Van Lierde et al. [1]. The quality measurements used were perceptual GRBAS, breathing performance, fundamental frequency, jitter and shimmer, and the Dysphonia Severity Index (linear combination of highest pitch, lowest loudness, max. phonation time and rel. jitter). However, the study focused only on monozygotic siblings (MZ). Another relevant reference is that of Cielo et al. [2]. Although the twin sample used is quite small (2 MZ pairs, one per gender) their analysis is interesting as far as they tackle some features not been considered in twins' voice studies before, namely vocal onset and harmonic characterization. While the

results for maximum phonation time showed significant differences between twins, no such differentiation was found regarding vocal onset, fundamental frequency or intensity. The work of Fuchs et al. [3] found that the voices of MZ twins showed more similarity among themselves than those of non-similar speakers regarding vocal range, highest and lowest fundamental frequency, prosodic pitch line, maximum intensity, number of overtones and intensity vibrato. The study of twins' voices can be approached from many perspectives. Stemming from a typical phonetic division, existing studies may be classified into one of these three fields: perception, acoustics or articulation. Some of the acoustic-related studies which specifically deal with voice-quality or glottal parameters have been reviewed [4]. Since perceptual or articulatory approaches are less relevant for the purpose of this presentation, we will consider a fourth group of studies: those which have investigated twins' voices from an automatic perspective. The automatic recognition system designed by Scheffer et al. [5] was able to identify twins with a good performance (85% of correct identifications). For the parameterization of the acoustic signal the method used was MFCC (Mel Frequency Cepstrum Coefficients). The 15% error of this automatic system (speakers who were not correctly detected as twins of their actual twins) would suggest, according to the authors, that "the twin of a speaker is not necessarily the most difficult impostor for an automatic speaker recognition system" ([5]: 2). The automatic system by Ariyaeeinia et al. [6] used LPCC (Linear Predictive Coding-Derived Cepstral) parameters, and the speaker representation was based on adapted Gaussian Mixture Models (GMMs). The results showed that the use of long test utterances led to smaller error rates than the use of short test utterances. Both KyungWha [7] and Künzel [8] used *Batvox*; although the former studied Korean females twin pairs (17 MZ, including 1 triplet and 5 DZ) and the latter studied German male and female twin pairs. The results in [7] showed that every twin speaker was correctly identified in the same speaking style condition (when models and test files were "read" speech). The performance of the system in [8] 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" ([8], p. 270). The present work is intended to concentrate in studying phonation marks (including biomechanical parameters) of relevance in the biometrical description of phonation [10, 11]. The working hypothesis is that phonation cycle quotients and biomechanics may offer differentiation capabilities among MZ, DZ and control speakers not explored already. The paper is organized as follows: A brief description of the materials and methods used in the study is given in section 2. In section 3 results obtained from the bilateral tests and matches of 16 male speakers are given discussed. Conclusions are presented in section 4.

2 Materials and Methods

Recordings from 40 male native speakers of Spanish (holding a spontaneous conversation) were taken at a sampling rate of 44,100 Hz and 16 bits using HQ

microphones in an isolated room. The distribution of speakers was as follows: 7 MZ pairs, 5 DZ pairs, 4 pairs of non-twin siblings and 4 pairs of controls (non-relatives). Spontaneous fillers (long vowels maintained for more than 200 ms around vowel [ε] produced inadvertently by speakers of Spanish in words like "que", "de", or in hesitation marks like "eeh..." etc.) were used in the study. Each speaker was recorded twice (2 sessions) separated by a 3-week interval. Speech recordings were around 10 min long. An average of 8-10 fillers was extracted from each recording.

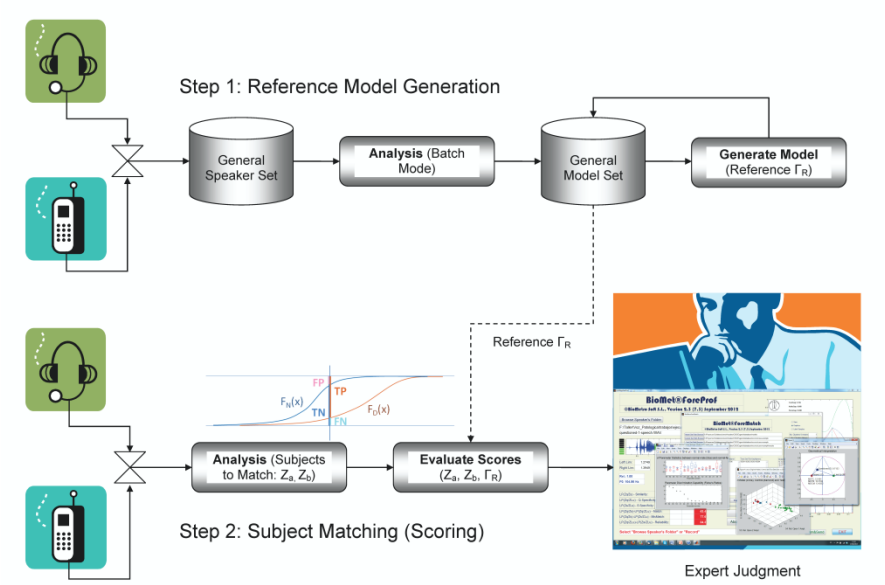

Fig. 1 Twins' Voice Matching Experimental Framework.

A set of biomechanical parameters as body and cover dynamic mass and stiffness was estimated from the spectral description of the glottal source reconstructed by inverse filtering. The inter-cycle unbalances of these parameters were also estimated. Open, Close and Return Quotients were added to the parameter set as well as Contact, Adduction and Permanent Gap Defects. The parameter set was completed with jitter, shimmer, NHR and Mucosal Wave ratio to produce a feature vector of 65 parameters referred to as $\mathbf{x}_{sij}$, where s refers to the subject, i is for the session, and j determines the filler. A set of pair-wise parameter matching experiments was carried out by likelihood ratio contrasts used in forensic voice matching [6]. The test is based on two-hypotheses contrasts: that the conditional probability between voice samples $\mathbf{Z}_a=\{\mathbf{x}_{aij}\}$ and $\mathbf{Z}_b=\{\mathbf{x}_{bij}\}$ (from the two subjects under test, a and b) is larger than the conditional probability of each subject to a Reference Speaker's Model Γ_R in terms of logarithmic likelihood ratios

$$\lambda_{ab} = \log\left[\frac{p(\mathbf{Z}_b \mid \Gamma_a)}{\sqrt{p(\mathbf{Z}_a \mid \Gamma_R)p(\mathbf{Z}_b \mid \Gamma_R)}}\right] \tag{1}$$

where conditional probabilities have been evaluated using Gaussian Mixture Models (Γ_a, Γ_b, Γ_R) constructed using available material from each speaker's vector subset and the reference set as:

$$p(\mathbf{Z}_b \mid \Gamma_a) = \Gamma_a(\mathbf{Z}_b); \quad p(\mathbf{Z}_a \mid \Gamma_R) = \Gamma_R(\mathbf{Z}_a); \quad p(\mathbf{Z}_b \mid \Gamma_R) = \Gamma_R(\mathbf{Z}_b) \tag{2}$$

Having these backgrounds stated, the Forensic Voice Evidence Evaluation Framework will then be a two-step process (refer to Fig. 1):

- Step 1. Model Generation. A model Γ_R representative of the general segment of population to be considered (male subjects between 18-52 years-old) was created for reference. For such a set of speakers $\mathbf{Z}_R=\{\mathbf{x}_{Rjk}\}$ was collected. This set is used to create a Gaussian Mixture Model defined in general as $\Gamma_R=\{\mathbf{w}_R, \boldsymbol{\mu}_R, \mathbf{C}_R\}$, $\mathbf{w}_R$, $\boldsymbol{\mu}_R$ and $\mathbf{C}_R$ being the set of weights, averages and covariance matrices associated to each Gaussian Probability Distribution in the set. In what follows a single Gaussian Reference Model will be assumed.
- Step 2. Score Evaluation. It is assumed that the set of the material under evaluation will be composed of different samples of parameterized voice in matrix form $\mathbf{Z_a}=\{\mathbf{x}_{aj}\}$, where $1\leq j\leq J_a$ is the sample index, each sample being on its turn a vector of M parameters (features, or observations) from vowel-like segments (fillers) conveniently parameterized $\mathbf{x}_{aj}=\{x_{aj1}\ldots x_{ajM}\}$. Similarly, the set of the correspondent speaker to be matched will be composed of different samples of parameterized voice $\mathbf{Z_b}=\{\mathbf{x}_{bj}\}$, where $1\leq j\leq J_b$ on its turn will be the sample index, each sample being a vector of M parameters also from vowel-like segments $\mathbf{x}_{bj}=\{x_{bj1}\ldots x_{bjM}\}$.

The estimation of the conditioned probability of a sample from the material under evaluation from speaker a $\mathbf{x_{aj}}$ to be associated to speaker *b*'s model will be evaluated as:

$$\Pr(\mathbf{x_{bj}} \mid \Gamma_a) = \frac{1}{(2\pi)^{M/2}\left|\mathbf{C_a}\right|^{Q}} e^{-1/2\left(\mathbf{x_{bj}}-\boldsymbol{\mu_a}\right)^T \mathbf{C_s}^{-1}\left(\mathbf{x_{bj}}-\boldsymbol{\mu_a}\right)} \tag{3}$$

Similarly the conditioned probability of a sample from the material from speaker *a* to be associated to the Reference Model will be given as:

$$\Pr(\mathbf{x_{aj}} \mid \Gamma_R) = \frac{1}{(2\pi)^{M/2}\left|\mathbf{C_R}\right|^{Q}} e^{-1/2\left(\mathbf{x_{aj}}-\boldsymbol{\mu_R}\right)^T \mathbf{C_s}^{-1}\left(\mathbf{x_{aj}}-\boldsymbol{\mu_R}\right)} \tag{4}$$

Finally the conditioned probability of a sample from material from speaker *b* to be associated to the Reference Model will be given as:

$$\Pr(\mathbf{x_{bj}} \mid \Gamma_R) = \frac{1}{(2\pi)^{M/2}\left|\mathbf{C_R}\right|^{Q}} e^{-1/2\left(\mathbf{x_{bj}}-\boldsymbol{\mu_R}\right)^T \mathbf{C_s}^{-1}\left(\mathbf{x_{bj}}-\boldsymbol{\mu_R}\right)} \tag{5}$$

For a full description of this methodology the interested reader may check [12].

Intra-speaker tests used recordings from different sessions. A priori expectations assume that MZ will show the largest LLRs, followed by DZ, then by non-twin siblings; non-related speakers expected to show the lowest LLRs.

3 Results and Discussion

The study covered results from 40 subjects, two sessions per subject taken separated by an interval of 3 weeks in between. The composition of the sample was the following: 14 subjects are MZ siblings in 7 pairs (numbered as 01-02, 03-04, 05-06, 07-08, 09-10, 11-12 and 33-34), 10 subjects are DZ siblings in 5 pairs (corresponding to speakers numbered as 13-14, 15-16, 17-18, 19-29 and 45-46), 8 subjects are non-twin brothers (BS) in 4 pairs (numbered as 21-22, 23-24, 47-48 and 49-50) and 8 subjects are not known to have any familiar relationship (US), grouped also as 4 pairs (25-26, 27-28, 29-30 and 31-32). Speakers were matched by pairs in: a) intra-speaker tests comparing parameters from different sessions (I: intra-speakers), b) inter-speaker tests by pairs (O: inter-speakers). The results of the matching tests are summarized in Table 1.

Table 1. Summary of the results for the different tests. MZ: Monozygotics; DZ: Dizygotics; RS: Related Siblings; US: Unrelated Speakers; (I): intra-speaker tests; (O): inter-speaker tests. Divided columns are used for each pair member. Cases: xxvyy means speaker xx versus speaker yy. Matches: Strong Likeness (SL): $\lambda \geq 1$; Weak Likeness (WL): $-1 \leq \lambda < 1$; Unlikeness (UL): $\lambda < -1$. In bold: results contrary to hypotheses H1 and H2 (MZ should be SL or WL, Intraspeaker's should be SL or WL). Hypothesis color code: H1; H2; H3; H4; H5; ~H1-5.

	MZ (I)		MZ(O)	DZ(I)		DZ(O)	RS(I)		RS(O)	US(I)		US(O)
Cases	01v01/02v02		01v02	13v13/14v14		13v14	21v21/22v22		21v22	25v25/26v26		25v26
LLR	2.4	-0.5	-0.0	6.4	-0.7	1.7	0.3	5.9	-3.5	**-42.2**	-0.7	-11.2
Match	SL	WL	WL	SL	WL	SL	WL	SL	UL	**UL**	WL	UL
Cases	03v03/04v04		03v04	15v15/16v16		15v16	23v23/24v24		23v24	27v27/28v28		27v28
LLR	**-1.1**	**-8.3**	-1.0	**-8.7**	5.2	-3.2	6.4	-0.3	0.7	10.2	11.9	-9.7
Match	**UL**	**UL**	WL	**UL**	SL	UL	SL	WL	WL	SL	SL	UL
Cases	05v05/06v06		05v06	17v17/18v18		17v18	47v47/48v48		47v48	29v29/30v30		29v30
LLR	12.5	6.1	5.8	1.6	4.3	-10.1	2.9	**-1.2**	-5.5	-0.2	7.5	-13.2
Match	SL	SL	SL	SL	SL	UL	SL	**UL**	UL	WL	SL	UL
Cases	07v07/08v08		07v08	19v19/20v20		19v20	49v49/50v50		49v50	31v31/32v32		31v32
LLR	12.0	6.6	12.1	0.6	**-7.7**	-0.4	**-1.3**	**-2.5**	1.6	6.1	5.2	-12.7
Match	SL	SL	SL	WL	**UL**	WL	**UL**	**UL**	SL	SL	SL	UL
Cases	09v09/10v10		09v10	45v45/46v46		45v46						
LLR	**-7.0**	23.0	12.6	-1.0	0.0	3.4						
Match	**UL**	SL	SL	WL	WL	SL						
Cases	11v11/12v12		11v12									
LLR	4.3	14.1	**-14.6**									
Match	SL	SL	**UL**									
Cases	33v33/34v34		33v34									
LLR	**-5.0**	0.2	0.6									
Match	**UL**	WL	WL									

The hypotheses tested were the following:

H1. It is expected that each speaker will produce large matching scores in intra-speaker tests.

H2. It is expected that MZ twins will show large matching scores also in inter-speaker tests.

H3. It is expected that DZ twins will show also large matching scores in inter-speaker tests although not that large as in H1 or H2.
H4. It is expected that BS will show matching scores over the background baseline (fixed empirically at $\lambda = -10$).
H5. It is expected that US will show matching scores aligned with the background baseline.

The baseline is defined by a reference background set composed of 20 speakers (set B). Scores are qualified as Strong Likeness if above 1, Weak Likeness if between 1 and -1 and Unlikeness if below -1.

Those results contradicting the strongest hypotheses (H1 and H2) are marked in bold. Four speakers out of the total of 40 appear to be in the limit of H1 (03, 48, 49 and 50), five others show strong intra-speaker dissimilarity (04, 09, 15, 20 and 33), and one shows very strong self-dissimilarity (25), therefore 10 out of 40 do not fulfil H1. The rest of the speakers show weak or strong self-similarity in inter-session tests, fulfilling H1. Regarding H2 we find only one case in which the hypothesis is not fulfilled (11 vs 12), out of 7 pairs. Hypothesis 3 is not fulfilled in one case (17 vs 18) out of 5 cases. Regarding H4 all four cases of non-twin brothers fulfil the hypothesis. In the case of unrelated subjects only one pair is slightly over the baseline threshold (27 vs 28) out of 4 cases, the rest of the pairs fulfil hypothesis H5. From the results summarized in Table 1 the ones affecting only to MZ siblings deserve special attention. To illustrate their results more conveniently they have been depicted in Fig. 2.

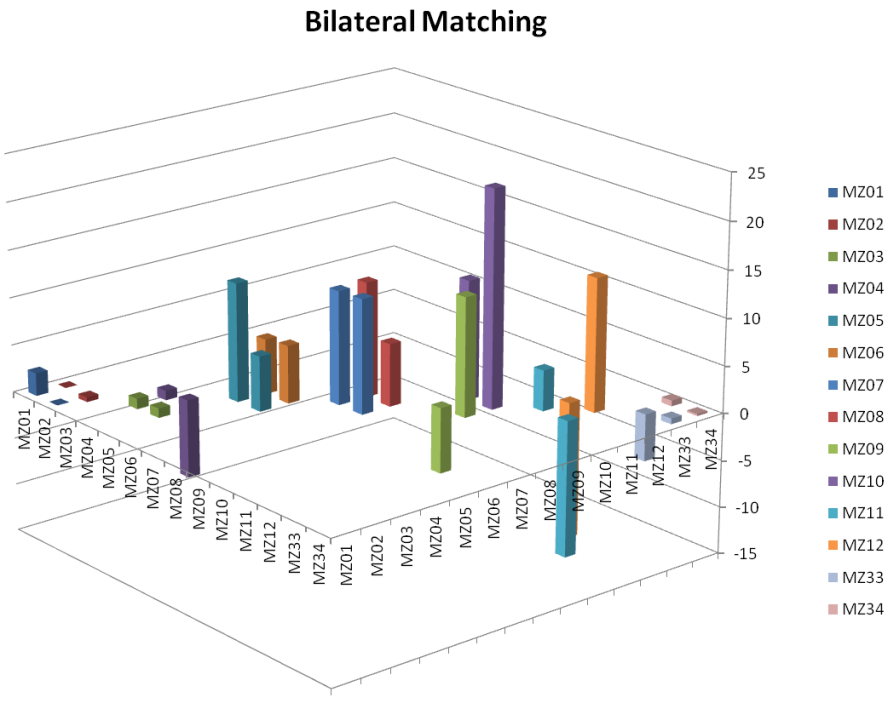

Fig. 2 Summary of the results for the MZ tests.

The 3 intra-speaker tests out of 14 which do not fulfil H1 may be clearly seen as relatively large negative columns (04, 09 and 33) as well as one inter-speaker test not fulfilling H2 (11 vs 12). Two twin pairs show good fulfilment of H1 and H2 (05, 06, 07 and 08), another twin pair do show a weak fulfilment of H1 and H2 (01 and 02), two twin pairs show weak fulfilment of H2, and irregular fulfilment of H1 (03, 04, 33 and 34) another twin pair shows strong fulfilment of H2 and irregular fulfilment of H1 (09 and 10) and another pair shows good fulfilment of H1 but irregular fulfilment of H2 (11 and 12). Some words have to be said about intra-speaker fulfilment of H1: it is unclear why 10 out of 40 speakers do show self-unlikeness in a larger or smaller extent when one session phonation is tested against another. Several reasons have been considered, as changes in phonation due to emotional stress or even temporary pathological conditions. Excluding weak self-unlikeness the number of cases would be 6 out of 40, which is still a large figure. Possibly some normalization on the selection of the speaker's most characteristic phonation patterns could help in reducing this apparently large number. Regarding H2 the number of unfulfillments seems smaller (1 out of 7 pairs). Reasons for dissimilarities in MZ within-pair comparisons seem somehow different. The most plausible reason that we can pinpoint is the nature-nurture dichotomy: in other words, the behavioural component of phonation as opposed to genetic reasons (essentially, phonation characteristics may be due to learned styles as much as to biological imprinted patterns).

4 Conclusions

The results of the study show some interesting considerations. Regarding H1 it seems that there are certain speakers who do not show strong intra-speaker similarity (6 out of 40 are in this situation). The immediate reflection is if these could be labelled as "goats" in Doddington's Zoo [13]. As far as H2 is concerned it seems that most MZ twins show reasonable inter-speaker (within-pair) similarity except in one pair out of 7. Whether this could be due to behavioural rather than to genetic factors is an open question. In DZ twins (H3) the situation is similar (only 1 out of 5 pairs show low inter-speaker scores). Non-twin brothers fulfil H4 relatively well, since all 4 pairs considered showed scores over the background baseline. Finally non-relative subjects showed scores well around the background baseline giving a good description of what would be considered the normal situation in unrelated speakers.

A possible complementary explanation involves the parameters sensitive to the study out of the 65 set considered. It seems that the parameters that have been used in such comparisons show a great influence of both genetic and environmental factors. If only the comparisons of MZ twin pairs had yielded large matches, the only explanation possible would be genetic influence. However, the fact that similar values are obtained for MZ and DZ twins cannot lead to that conclusion. The impact of external factors (like a similar living and educational environment, same age, etc.) must be more relevant than it may be thought a priori in this kind of voice studies.

Further research would be necessary especially in order to study the role of the specific parameters (out of the 65 possible features) intervening in the results from each comparison. Likewise, it seems vital to consider a reanalysis with more speakers.

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

References

1. Van Lierde, K. M., Vinck, B., De Ley, S., Clement, G., and Van Cauwenberge, P. "Genetics of vocal quality characteristics in monozygotic twins: a multiparameter approach", Journal of Voice, 19 (4), 2005, pp. 511-518.
2. Cielo, C. A., Agustini, R. and Finger, L. S., "Características vocais de gêmeos monozigóticos", Revista CEFAC, 14 (6), 2012, pp. 1234-1241 (in Portuguese, summary in English).
3. Fuchs, M., Oeken, J., Hotopp, T., Täschner, R., Hentschel, B. and Behrendt, W., "Die Ähnlichkeit monozygoter Zwillinge hinsichtlich Stimmleistungen und akustischer Merkmale und ihre mögliche klinische Bedeutung", HNO, 48 (6), 2000, pp. 462-469.
4. San Segundo, E. and Gómez-Vilda, P. Voice Biometrical Match of Twin and non-Twin Siblings. 1st Multidisciplinary Conference of Users of Voice, Speech and Singing, Las Palmas de Gran Canaria, 27-28 June 2013 (to appear).
5. Scheffer, N., Bonastre, J-F., Ghio, A. and Teston, B. (2004) *Gémellité et reconnaissance automatique du locuteur*, Actes, Journées d'Etude sur la Parole (JEP), 445-448.
6. Ariyaeeinia, A., Morrison, C., Malegaonkar, A. and Black, S. (2008) A test of the effectiveness of speaker verification for differentiating between identical twins, Science and Justice, 48, 182-186.
7. KyungWha, K. (2010) Automatic speaker identification of Korean female twins, Proc. 19th Annual Conference of the International Association for Forensic Phonetics and Acoustics, 18-21 July 2010, Trier, Germany.
8. Künzel, H. (2010) Automatic speaker recognition of identical twins, International Journal of Speech Language and the Law, 17 (2): 251-277.
9. Gómez, P., Fernández, R., Rodellar, V., Nieto, V., Álvarez, A., Mazaira, L. M., Martínez, R, and Godino, J. I.: Glottal Source Biometrical Signature for Voice Pathology Detection. *Speech Comm.*, (51) 2009, pp. 759-781.
10. Gómez, P., Rodellar, V., Nieto, V., Martínez, R., Álvarez, A., Scola, B., Ramírez, C., Poletti, D., and Fernández, M.: BioMet®Phon: A System to Monitor Phonation Quality in the Clinics. Proc. eTELEMED 2013: The Fifth Int. Conf. on e-Health, Telemedicine and Social Medicine, Nice, France, 2013, 253-258.
11. González, J., Rose, P., Ramos, D., Toledano, D. T. and Ortega, J., "Emulating DNA: Rigurous Quantification of Evidential Weight in Transparent and Testable Forensic Speaker Recognition", IEEE Trans. On Audio, Speech and Lang. Proc., 15 (7), 2007, pp. 2104-2115.
12. Gómez, P., Mazaira, L. M., Hierro, J. A. and Nieto, R.: Distance Metrics in Voice Forensic Evidence Evaluation using Dysphonia-relevant Parameters. VI Jornadas de Reconocimiento Biométrico de Personas, Las Palmas de Gran Canaria, January 26-27, 2012.
13. Doddington, G., Liggett, W., Martin, A., Przybocki, M., & Reynolds, D.: Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. NIST, Gaithersburg, MD, 199.