Analysis of emotional stress in voice for deception detection

V. Rodellar-Biarge, D. Palacios-Alonso, V. Nieto-Lluis and P. Gómez-Vilda
Neuromorphic Speech Processing Laboratory
Centro de Tecnología Biomédica and Escuela Técnica Superior de Ingenieros Informáticos
Universidad Politécnica de Madrid
Campus de Montegancedo - Pozuelo de Alarcón - 28223 Madrid - SPAIN
email: victoria@fi.upm.es

***Abstract* — The current work gives a method to estimate alterations in speech caused by stress. This estimation is used as an indicator to detect deceptive speech. Truthful or neutral speech and deceptive speech were elicited by forcing to answer "hot questions" related with politics and society. Data were processed by using log-likelihood ratios and Fisher's linear discriminant analysis. Independent results for male and female are presented. The classification results are around 100% for neutral speech in both genders, while the best classification rate (67%) for stressed speech is achieved for females. In a first approach we have seen, that subjects tend to be more deceptive when the questions are related with gender issues, giving a politically correct answer instead of their true opinion.**

Keywords—Voice production; biomechanical parameters; speaker biometry; emotional stress; deception detection

I. INTRODUCTION

Deception detection is a very difficult task, which has been always a subject of controversy, as there are persons with special ability in this sense, whereas that can be considered a gift not to the reach of everybody. Deception leaves marks in several physical parameters that are not under conscious control, as eye and facial activity, skin conductance, sweat, blush, altered heart and respiration rate, and muscular involuntary movements among others [1]. There are in the literature different techniques to detect deception based on the measurements of those physical parameters, but much more research must be done in order to get significant reliable results.

Deception, in most of the persons, causes psychological stress which has an evident reflect in face, voice, limbs and walking movements. Voice analysis techniques for deception detection have a lot of interest in the development of many applications for very different sectors, as banking and insurances, costumer services, forensic evidence analysis, etc. Police and fireman departments, hospitals and other emergency call centers could also benefit from this technology application because they receive a big amount of phone calls that must be carefully analyzed and filtered, in order to take quick decisions according to their trustfulness and relevance to prioritize actuations and optimize resources. But deception detection without visual cues becomes very difficult task.

The present approach to deception detection starts with the identification of presence of stress in speech supposedly caused by deception. Stress is signaled in the prosody, articulation and phonation correlates. Several works related with emotional stress detection in speech can be found in the literature. Most of them propose language independent systems by using just the speech sound and gender independent detection. The proposal of Lefter [2] separates neutral speech from emotional speech with a set of prosodic-based features at the utterance level and spectral features at the frame level. They use a database recorded with spontaneous emotions in a call centre. The proposal of Demeko [3] focused their attention on pitch variability and analyzed its behavior on a database provided by police with records containing spontaneous emergency calls. Other authors complete the information extracted from audio records with biological signals taken with biosensors, as skin conductance measurements [4], chemical analysis, or electrocardiograms for validation purposes. In [5] the relation between stress and fluctuations in physiological microtremor present in the vocal folds, which is projected to voice, is analyzed [6].

The objective of this work is to identify stress in voice from speech and detect deception. Voice characterization is carried out from a set of acoustical, glottal, biomechanical and tremor features. The neutral/stress states classification have been obtained using Fisher linear discriminant analysis, and a further qualitative analysis from them to identify deception. Data samples are from native Spanish speakers.

The rest of the paper is organized as follows: Section II analyzes the available databases, introduces the data corpus used in the presented results and the parameter selection criteria. Section III describes the pattern classification methodology used. Section IV presents and discusses results, with special emphasis on the interpretation of the complex scenario envisioned by deception detection. And finally, the conclusions and future work are presented in section V.

II. DATA TREATMENT

The validity of the studies presented in the literature depends heavily of the experimental data framework. In the majority of publications the results presented are difficult to reproduce due to the lack of public standardized databases which could serve as benchmarks.

978-1-4673-7846-8/15/$31.00 ©2015 IEEE

A. Data corpus.

There are several spontaneous and simulated emotional speech databases [7] [8]. But concerning stress in speech, we have found very few databases and data collections. The SUSAS database (Speech Under Simulated and Actual Stress) is public and is the most referenced in the literature [9]. It contains 35 English words spontaneously spoken by aircraft pilots and other information from non-spontaneous speech. A speech corpus was recorded at MIT labs, eliciting stress in four car drivers by asking them to add two numbers summing less than 100 while driving a car [10]. Other initiative was a Czech database, recorded during final oral examinations at Brno University. The database contains speech and heart rate records of around 34 male speakers, mostly Czech natives [11]. A multilingual database in Mandarin and Cantonese and English for emotional stress is reported in [12]. It contains the answers to a questionnaire from university students during their examination periods. All answers were annotated by native speakers for each language database and divided by gender.

The objective of this work is to study stress in voice in Spanish native speakers. There are no publicly available databases to cover this objective; therefore a database had to be created from anew. Different strategies can be found in the literature to evoke stress in speech. In this work, the protocol followed is the one described in [13] consisting, in the first place, in evaluating the arousal and valence about personal opinions related with very controversial social topics. Then the two topics with the higher score were selected. Next the speakers are asked to defend their true opinion relative to one of the topics (self-agreement, supposedly non-stressed) and to defend a false opinion relative the other (self-disagreement, supposedly stressed). Each answer must be given in a time limit of 20 seconds. Self-agreed or true opinions are expressed fluently and can be considered as the unstressed or neutral state. On the contrary, self-disagreed or false opinions require preparing artificially the speech statements and this result in stress, which is shown as repetitions, stops, fillers, longer vowels, etc. This is the hypothetically ideal situation to be evaluated by the experimental setup. The words /de/ and /que/ and the vowel /e/, are the most common fillers used by Spanish speakers. The filler vowel /e/, has been used to obtain the results presented in this paper. The data collection is composed of four sets of samples (agreement/disagreement) from 18 women and 16 men who do not have any apparent alteration in their capability for speech production or perception. The records are from native Spanish speakers from the area of Madrid.

B. Extraction of features

The detection of stress is done by observing the variations in the set of features that characterize the neutral speech, compared with speech produced under stress. The optimal set of features that characterize an emotion is still an open issue. A large number of features may contain a lot of redundancy and it is computationally expensive but a small number may not have enough information to characterize an emotion.

The software tool BioMet®Phon [14] suitable for the evaluation of voice quality and biometry was used to parameterize voice segments from fillers of vowel /e/ [15]. This tool is able to extract up to 72 features, acoustical, glottal, biomechanical and tremor from voice. The list of features is described in TABLE I.

TABLE I. LIST OF FEATURES USED IN THE STUDY

Feature number	Description
1	Pitch
2	Jitter
3	Shimmer
4	Sharpness
5	Noise-Harmonics Ratio
6	Mucosal Wave energy average
7-20	Cepstral description of the glottal source
21-34	Spectral description of the glottal source power spectral density
35-46	Biomechanics of the Vocal Folds (Body and Cover)
47-58	Temporal description of the glottal source contact and open phases
59-62	Glottal gaps: contact, adduction, permanent
63-65	Indicators of neurological alteration
66-72	Physical, neurological and flutter tremor (amplitude and frequency)

The biomechanical features than can be extracted from the glottal source seem to be very meaningful, [15], [16]. They give a description of the phonation conditions of the larynx, and give also information on the conditions of the neurological pathways innervating the larynx, the hypothalamus relay state and the speech neuromotor cortex activity. These processing centers may be altered when the speakers are forced to fabricate an artificial opinion which is against their ideas and feelings. We calculate the set of 72 features for each speaker's phonation from different fragments of /e/ in time slots between 100 ms and 200 ms. One set was taken from neutral speech and the other from the fillers that appear in stressed speech. Then, the final value of each parameter was calculated as the mean of all parameters obtained from all time slots for the same vowel fragment.

III. CLASSIFICATION METHOD

The definition of the problem is as follows: the male and female speaker sets (m: males, f: females) are designed as S_m and S_f, each one integrated by I_m=16 and I_f=18 subjects respectively. There are two observation matrices per gender, comprising features extracted from statements produced under supposedly self-agreement (a) and self-disagreement (d) conditions, these matrices being designed as $\mathbf{X}_{am}$, $\mathbf{X}_{dm}$, $\mathbf{X}_{af}$ and $\mathbf{X}_{df}$. Each matrix row contains an observation row vector ($\mathbf{x}_{ami}$, $\mathbf{x}_{dmi}$, $\mathbf{x}_{afi}$ and $\mathbf{x}_{dfi}$) from a given speaker i ($1\leq i\leq I_m$ or $1\leq i\leq I_f$), containing J=72 features as described in TABLE I: $\mathbf{x}_{ami}=\{x_{amij}\}$; $\mathbf{x}_{dmi}=\{x_{dmij}\}$; $\mathbf{x}_{afi}=\{x_{afij}\}$ and $\mathbf{x}_{dfi}=\{x_{dfij}\}$, with $1\leq i\leq J$. Therefore column vectors indexed by feature j will contain all the observations produced by each statement subset (am: male in agreement, af: female in agreement, dm: male in disagreement and df: female in disagreement). Averages and standard deviations for each column vector j from each observation matrix will be estimated as μ_{amj}, μ_{afj}, μ_{dmj}, and μ_{dfj}; and σ_{amj}, σ_{afj}, σ_{dmj}, and σ_{dfj}, respectively. These estimates will be used in evaluating the individual discrimination capability of each feature [17] for each male and female subset as

$$f_{mj}=\frac{\left(\mu_{maj}-\mu_{mdj}\right)^2}{\sigma^2_{maj}-\sigma^2_{mdj}}$$
$$f_{fj}=\frac{\left(\mu_{faj}-\mu_{fdj}\right)^2}{\sigma^2_{faj}-\sigma^2_{fdj}} \tag{1}$$

The means estimated for each speaker (i) will compose the mean row vectors $\boldsymbol{\mu}_{am}$, $\boldsymbol{\mu}_{af}$, $\boldsymbol{\mu}_{dm}$, and $\boldsymbol{\mu}_{df}$. Besides, the covariance matrices for each subset will be also estimated as: $\mathbf{C}_{am}=E\{\mathbf{X}_{am}\mathbf{X}'_{am}\}$, $\mathbf{C}_{af}=E\{\mathbf{X}_{af}\mathbf{X}'_{af}\}$, $\mathbf{C}_{dm}=E\{\mathbf{X}_{dm}\mathbf{X}'_{dm}\}$ and $\mathbf{C}_{df}=E\{\mathbf{X}_{df}\mathbf{X}'_{df}\}$, where E{.} refers to statistical expectation, and (') refers to matrix transposition. In what follows it will be assumed that observations are produced by Gaussian processes, therefore the Gaussian models $\Gamma_{am}=\{\boldsymbol{\mu}_{am}, \mathbf{C}_{am}\}$, $\Gamma_{af}=\{\boldsymbol{\mu}_{af}, \mathbf{C}_{af}\}$, $\Gamma_{dm}=\{\boldsymbol{\mu}_{dm}, \mathbf{C}_{dm}\}$ and $\Gamma_{df}=\{\boldsymbol{\mu}_{df}, \mathbf{C}_{df}\}$ are defined.

In the present approach, as a first evaluation pilot, each feature from individual vectors $\mathbf{x}_{mai}$, $\mathbf{x}_{mdi}$, $\mathbf{x}_{fai}$ or $\mathbf{x}_{fdi}$ will be tested against its corresponding simplified (naïve) Gaussian model using log-likelihood ratios defined as

$$\lambda_{amij}=\lg\frac{p\left(x_{amij}\middle|\Gamma_{sam}\right)}{p\left(x_{amij}\middle|\Gamma_{sdm}\right)}$$
$$\lambda_{dmij}=\lg\frac{p\left(x_{dmij}\middle|\Gamma_{sam}\right)}{p\left(x_{dmij}\middle|\Gamma_{sdm}\right)}$$
$$\lambda_{afij}=\lg\frac{p\left(x_{afij}\middle|\Gamma_{saf}\right)}{p\left(x_{afij}\middle|\Gamma_{sdf}\right)} \tag{2}$$
$$\lambda_{dfij}=\lg\frac{p\left(x_{dfij}\middle|\Gamma_{saf}\right)}{p\left(x_{dfij}\middle|\Gamma_{sdf}\right)}$$

where the conditional probabilities are defined as follows

$$p\left(x_{amij}\middle|\Gamma_{sam}\right)=\frac{1}{\sqrt{2\pi}\sigma_{amj}}e^{-\frac{\left(x_{amij}-\mu_{amj}\right)^2}{2\sigma^2_{amj}}}$$
$$p\left(x_{amij}\middle|\Gamma_{sdm}\right)=\frac{1}{\sqrt{2\pi}\sigma_{dmj}}e^{-\frac{\left(x_{amij}-\mu_{dmj}\right)^2}{2\sigma^2_{dmj}}}$$
$$p\left(x_{dmij}\middle|\Gamma_{sam}\right)=\frac{1}{\sqrt{2\pi}\sigma_{amj}}e^{-\frac{\left(x_{dmij}-\mu_{amj}\right)^2}{2\sigma^2_{amj}}} \tag{3}$$
$$p\left(x_{dmij}\middle|\Gamma_{sdm}\right)=\frac{1}{\sqrt{2\pi}\sigma_{dmj}}e^{-\frac{\left(x_{dmij}-\mu_{dmj}\right)^2}{2\sigma^2_{dmj}}}$$

$$p\left(x_{afij}\middle|\Gamma_{saf}\right)=\frac{1}{\sqrt{2\pi}\sigma_{afj}}e^{-\frac{\left(x_{afij}-\mu_{afj}\right)^2}{2\sigma^2_{afj}}}$$
$$p\left(x_{afij}\middle|\Gamma_{sdf}\right)=\frac{1}{\sqrt{2\pi}\sigma_{dfj}}e^{-\frac{\left(x_{afij}-\mu_{dfj}\right)^2}{2\sigma^2_{dfj}}}$$
$$p\left(x_{dfij}\middle|\Gamma_{saf}\right)=\frac{1}{\sqrt{2\pi}\sigma_{afj}}e^{-\frac{\left(x_{dfij}-\mu_{afj}\right)^2}{2\sigma^2_{afj}}}$$
$$p\left(x_{dfij}\middle|\Gamma_{sdf}\right)=\frac{1}{\sqrt{2\pi}\sigma_{dfj}}e^{-\frac{\left(x_{dfij}-\mu_{dfj}\right)^2}{2\sigma^2_{dfj}}}$$

The interpretation of the log-likelihood ratios defined in (2) is the following: each observations vector from the self-agreement statements is checked against the overall self-agreement distribution and against the overall self-disagreement distribution; if the self-agreement condition would be true, it should be expected that the observations vector membership with respect to the self-agreement distribution would be larger than with respect to the self-disagreement one, therefore the ratio of probabilities would be larger than one, and the log-likelihood would be positive. Otherwise it would render a negative value. Similarly, self-disagreement statement observations are checked against self-agreement distributions vs self-disagreement distributions. If self-disagreement statements were sincere it should be expected that the corresponding log-likelihood ratios had to be negative in this case, as they should show a larger membership with respect to the self-disagreement distribution. As the log-likelihood ratios have been estimated for each feature (naïve method) a method to fuse all the individual feature ratios for a given speaker is required. This is done combining the individual feature log-likelihood ratios weighted by the discriminant factors given in (1)

$$\vartheta_{ami}=\sum_1^J f_{mj}\lambda_{amij}$$
$$\vartheta_{dmi}=\sum_1^J f_{mj}\lambda_{dmij}$$
$$\vartheta_{afi}=\sum_1^J f_{fj}\lambda_{afij} \tag{4}$$
$$\vartheta_{dfi}=\sum_1^J f_{fj}\lambda_{dfij}$$

producing the speaker log-likelihood ratios given in TABLE II and TABLE III.

IV. RESULTS

The use of naïve log-likelihood comparisons in feature by feature estimations is based in two facts: on one hand due to the strong differences in individual feature distributions, covariance matrix inversion renders rather unstable results in some cases, reducing the robustness of the log-likelihood ratio estimate; on the other hand keeping the comparisons restricted to individual features allows analyzing possible causes for agreement or discrepancy on the basis of new hypotheses. For instance, the discrimination power of feature 66 may be inferred from Fig. 1. It may be seen that the self-disagreement distribution (stress) is more disperse than that of the self-agreement (neutral) and that the tremor frequency average is also higher (around 4.7 Hz for the d-distribution vs 4.5 Hz for the a-distribution). The f_j for this parameter is 0.102.

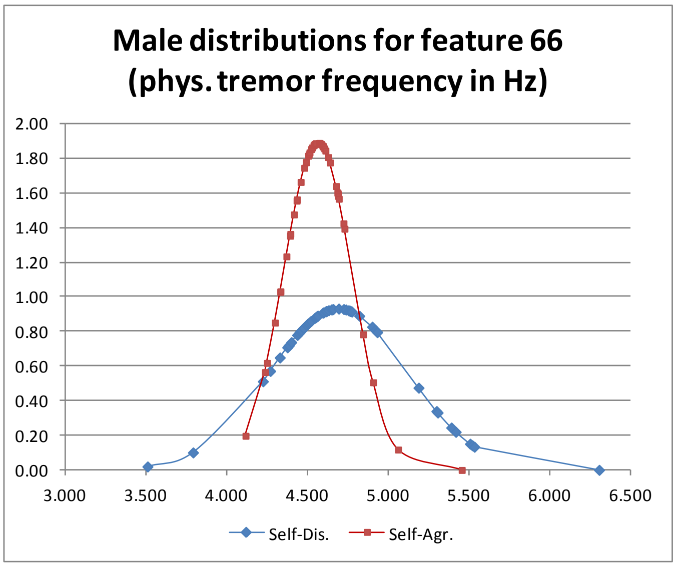

Fig. 1. Distributions of feature 66 (physiological tremor frequency) relative to self-agreement (diamonds) and self-disagreement (squares). The horizontal axis is in Hz, the vertical axis is in normalized counts. Both distributions can be relatively well represented by Gaussians.

The classification results from 16 males with the full set of 72 parameters extracted with BioMetroPhon [14] are shown in Table II. A positive score from equation (4) indicates a classification as neutral state while a negative score indicates stress state. The classification rate is 100% for speech unstressed/neutral while the rate drops to 56 % for stressed speech. In a first interpretation, the achieved results for stressed speech are very poor because the 44% of samples theoretically produced under stress are identified as neutrals. But the way in which the stress is elicited in the subjects could be taken into account for a second interpretation of the results. The speech samples have been obtained from two topics, that evocate on a person the highest valence and arousal. The topics have been chosen among a set of hot social topics, on which they may give the politically correct answer instead their true and sincere opinion. To analyze this issue, neutral and stress results for the same subject must be seen as a whole. Then, we will plot the numerical data associated to Table II to interpret the results with the last point of view. Neutral values are represented in x axis and stress in y axis.

The results for men are plotted on Figure 2. We can see individuals located in two well defined areas. The lower part includes the individuals that answer coherently with the data acquisition protocol. That is, they give their true opinion and their contrary opinion when they are asked to do so. We can qualify this area as the area of trustable people (samples 1, 2, 3, 5, 6, 9, 13, 15). The upper part of Figure 2, represents individuals that always present a neutral state and maybe their answers could be politically correct instead of coherent with their ideas. This area could be named as non-trustable (samples 4, 8, 10, 12, 14, 16). It is interesting to hear the content of its elicited stress sentences, because in most of the cases are related with gender issues giving answers to questions as "men at work, women at home", "men must be better paid that women for the same job". Samples M7 and M11 are located at both sides of the border line between both areas.

TABLE II LOG-LIKELIHOOD RESULTS FOR MALES

(A) AGREEMENT WITH PRIORS - (D): DISAGREEMENT WITH PRIORS

SUBJECT	SELF-AGREE (ϑ_{am})	SELF-DISAGREE (ϑ_{dm})
M1	0.327 (A)	-1.268 (A)
M2	0.081 (A)	-2.288 (A)
M3	0.343 (A)	-0.510 (A)
M4	0.595 (A)	0.290 (D)
M5	0.168 (A)	-1.002 (A)
M6	0.654 (A)	-7.323 (A)
M7	0.387 (A)	0.008 (D)
M8	0.808 (A)	0.621 (D)
M9	0.422 (A)	-5.965 (A)
M10	0.370 (A)	0.310 (D)
M11	2.094 (A)	-0.218 (A)
M12	0.335 (A)	0.322 (D)
M13	0.272 (A)	-0.823 (A)
M14	0.289 (A)	0.589 (D)
M15	1.118 (A)	-1.562 (A)
M16	0.536 (A)	0.461 (D)
CLASS. RATE	**100%**	**56%**

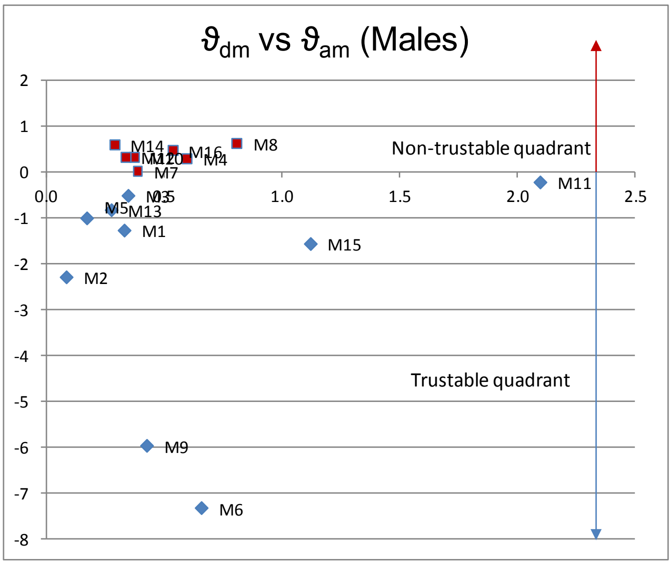

Fig. 2. Trustable (diamonds) and non-trustable (squares) male subjects.

We have done a similar study for females. The classification results for a group of 18 women using the 72 parameters obtained with BioMetroPhon are shown in TABLE III. The classification for neutral state fails in one case out of 18, giving a classification rate close to 100%, similar to males. While the classification for the stress state fails in six cases out of 18, achieving classification rate close to 67%, giving better results than males.

TABLE III LOG-LIKELIHOOD RESULTS FOR FEMALES

(A) AGREEMENT WITH PRIORS - (D): DISAGREEMENT WITH PRIORS

SUBJECT	SELF-AGREE (ϑ_{af})	SELF-DISAGREE (ϑ_{df})
F1	0.206 (A)	-0.222 (A)
F2	0.020 (A)	-1.435 (A)
F3	0.251 (A)	-0.860 (A)
F4	0.148 (A)	0.167 (D)
F5	0.227 (A)	-0.548 (A)
F6	0.172 (A)	-0.313 (A)
F7	0.042 (A)	-0.108 (A)
F8	0.200 (A)	0.027 (D)
F9	0.190 (A)	0.127 (D)
F10	0.145 (A)	-0.096 (A)
F11	0.210 (A)	-0.093 (A)
F12	0.066 (A)	-0.007 (A)
F13	-0.235 (D)	0.064 (D)
F14	0.374 (A)	-1.041 (A)
F15	0.172 (A)	0.105 (D)
F16	0.292 (A)	-0.121 (A)
F17	0.255 (A)	-0.215 (A)
F18	0.081 (A)	0.129 (D)
CLASS. RATE	**94%**	**67%**

The numerical results that generated Table III are plotted on Figure 3. Now we can see that the more populated area is the trustable part (cases 1, 2, 3, 5, 6, 7, 10, 11, 14, 16, 17). The non-trustable area has four clear cases: 4, 9, 15 and 18. And the border line contains cases 8 and 12. Case F13, behaves on contrary way than the persons classified in the trustable area, identifying neutral as stress and stress as neutral state. Maybe she has understood the interviewer directions in the opposite way or maybe she has been informed wrongly about how to answer.

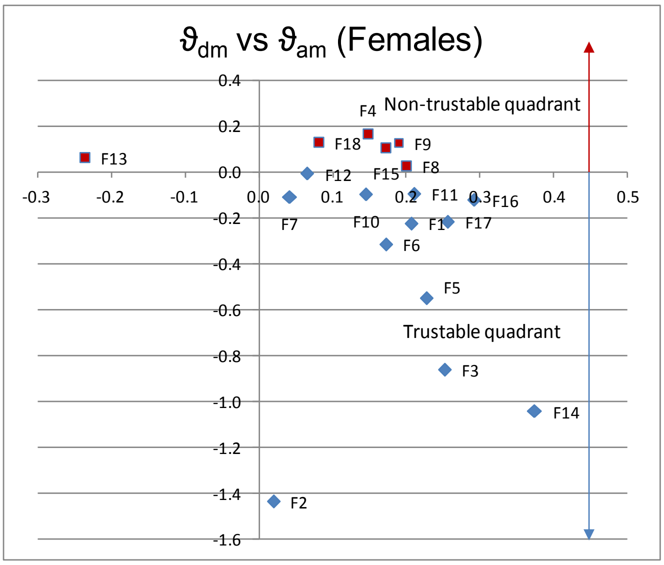

Fig. 3. Trustable (diamonds) and non-trustable (squares) female subjects.

IV. CONCLUSIONS AND FUTURE WORK

We have presented a method for evaluating an indicator in detecting self-agreement or contradiction in speech. The major problem to make an advance in this field is to use a technique or method that allows estimating reliable data to work with. The neurological processing centers are altered when a person is forced to fabricate an artificial opinion which is against their ideas and feeling, but it seems that data samples should not be taken two times from the same speaker because the second time they are taken, the spontaneity is lost and maybe the data could be corrupted. The lack of a clear reference to estimate the ground truth is another problem in this kind of studies. As there is no way to know the degree of coherence in the answers of the speakers related with certain embarrassing or delicate questions, it seems highly plausible that subjects may try to answer "what is expected" or "politically correct", therefore altering somewhat their self-agreement or disagreement coherence. The proposed methodology is devised to compensate the skew induced by such behavior as much as possible using cross-check. Therefore each subject responses are evaluated against the sample distributions of the agreement and disagreement answers produced by other speakers. After examining the results of TABLE II and TABLE III also plotted in Fig. 2 and Fig.3 the interpretation of the results becomes clearer. The coherence and incoherence of the answers point to trustable and non-trustable speaker-answer pairs.

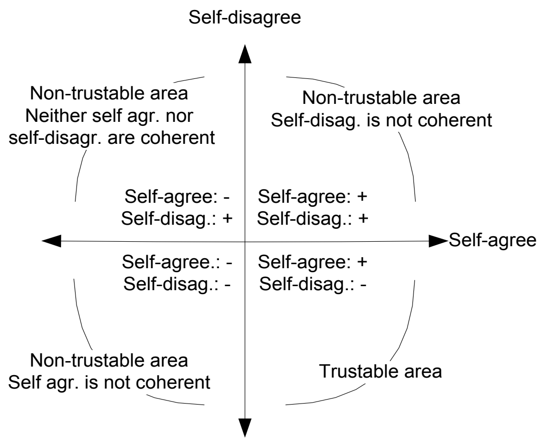

Fig. 4. Specific areas for deception detection

This conclusion can be generalized considering Fig. 4, where all possible pairs may be plotted. The trustable area corresponds to the fourth quadrant, where self-agreement and self-disagreement ratios are congruent with prior expected behavior. The first quadrant corresponds to subjects who are coherent with self-agreement prior expectation, but are not coherent with self-disagreement prior expectation. The third quadrant would show subjects in which self-agreement answers would be in contradiction with prior expectation, but coherent with self-disagreement prior expectation. The meaning of contradiction in giving an expected self-agreement opinion may be skewed by fear of judgment in expressing a socially impolite or incorrect opinion. The second quadrant is even more conflictive, as it is assumed that the subject is contradicting expectations both in expressing self-agreement as well as self-disagreement, which should correspond to the perfect "counter-stream swimmer". One of these cases is F13, although the self-disagreement score is close zero, therefore weakening the hypothesis of strong self-disagreement. In any case, the study is still open to completion by adding new subjects to the study. This task is not that simple, as from what has been said, a subject should be discarded from future interview once they have been interviewed once.

ACKNOWLEDGMENT

This work is being funded by grants TEC2012-38630-C04-01 and TEC2012-38630-C04-04 from Plan Nacional de I+D+i, Ministry of Economic Affairs and Competitiveness of Spain.

REFERENCES

[1] B. M. DePaulo et al. "Cues to Deception", Psychology Bulletin, Vol. 129, No.1, pp.74-128, 2003.

[2] I. Lefter, L. J. M. Rothkrantz, D. A. van Leeuwen and P. Wiggers, "Automatic stress detection in emergency (telephone) calls" Int. J. Intelligent Defence Support Systems, Vol. 4, No. 2, pp. 148-168, 2011.

[3] G. Demenko and M. Jastrzebska, "Analysis of voice stress in call centers conversations", Proceedings of Speech Prosody, pp. 183-186, 2012.

[4] J. Hernandez, R. R. Morris and R. W. Picard, "Call center stress recognition with person-specific models". LNCS 6964, S. D'Mello Ed., pp. 125-134, 2011.

[5] N. Mbitiru, P. Tay, J. Z. Zhang and R. D. Adams, "Analysis of Stress in Speech Using Empirical Mode Decomposition". Proceedings of the 2008, IAJC-IJME International Conference, pp. 140-146, 2008.

[6] O. Lippold, "Physiological Microtremor", Scientific American, 224(3), pp. 65-73, 1971.

[7] D. Ververidis and C. Kotropoulos, "A Review of Emotional Speech Databases". Proc. 9th Panhellenic Conference on informatics (PCI), pp. 560-574, Thessaloniki, Greece, 2003.

[8] M. El Ayadi, M. S. Kamel and F. Karray, "Survey on speech emotion recognition: Features, classification schemes and databases", Pattern Recognition 44, pp. 572-587, 2011.

[9] J. H. L. Hansen and S. E. Bou-Ghazale, "Getting started with SUSAS: A speech under simulated and actual stress database," Proceedings of the International Conference on Speech Communication and Technology (Eurospeech) Rhodes, Greece, pp. 1743–1746, 1997.

[10] R. Fernández and R. W. Pickard, "Modeling drivers' speech under stress", Speech Communication, Vol. 40 (1-2), pp.145-159, 2003.

[11] M. Sigmud, " Introducing the Database ExamStress for Speech under Stress". Proc. of Signal Processing Symposium, Reykajavic, pp. 1162-118, 2006.

[12] X. Zuo, T. Li and P. Fung, "A multilingual Database of Natural Stress Emotion". Proc. of the Eight International Conference on Language Resources and Evaluation, pp. 1174-1178, 2012.

[13] J. Arciuli, G. Villar and D. Mallard, "Lies, lies and more lies", 31st Annual Meeting of the Cognitive Science Society (CogSci 2009), pp. 2329-2334, 2009.

[14] BioMet®Phon, http://www.glottex.com/

[15] P. Gómez, et al, "Glottal Source Biometrical Signature for Voice Pathology Detection", Speech Comm., Vol. 51, pp. 759-781, 2009.

[16] V. Rodellar, D. Palacios, E. Bartolomé and P. Gómez, "Vocal fold stiffness estimates for emotion description in speech", *Proc. of the International Conference on Bio-inspired Systems and Signal Processing,* BIOSIGNALS'2013, pp. 112-119, 2013.

[17] Venables, W. N.; Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). Springer Verlag.