

Monitoring Neurological Disease in Phonation

Pedro Gómez-Vilda¹, Roberto Fernández-Baíllo¹,
José Manuel Ferrández-Vicente², Victoria Rodellar-Biarge¹,
Agustín Álvarez-Marquina¹, Luis Miguel Mazaira-Fernández¹,
Rafael Martínez-Olalla¹, and Cristina Muñoz-Mulas¹

¹Grupo de Informática Aplicada al Tratamiento de Señal e Imagen,
Facultad de Informática, Universidad Politécnica de Madrid,
Campus de Montegancedo, s/n, 28660 Madrid

²Dpto. Electrónica, Tecnología de Computadoras,
Univ. Politécnica de Cartagena,
30202, Cartagena
`pedro@pino.datsi.fi.upm.es`

Abstract. It is well known that many neurological diseases leave a fingerprint in voice and speech production. The dramatic impact of these pathologies in life quality is a growing concern. Many techniques have been designed for the detection, diagnosis and monitoring of the neurological disease. Most of them are costly or difficult to extend to primary services. The present paper shows that some neurological diseases can be traced at the level of voice production. The detection procedure would be based on a simple voice test. The availability of advanced tools and methodologies to monitor the organic pathology of voice would facilitate the implantation of these tests. The paper hypothesizes some of the underlying mechanisms affecting the production of voice and presents a general description of the methodological foundations for the voice analysis system which can estimate correlates to the neurological disease. A case of study is presented from spasmodic dysphonia to illustrate the possibilities of the methodology to monitor other neurological problems as well.

1 Introduction

During the last years a methodology to accurately detect voice pathology of organic origin has been developed [4]. This methodology supports early pathology detection and monitoring may be extended to other diseases of neurological nature if these leave alterations or correlates in voice production [2]. This objective could be of strategic interest towards research and innovation in e-health services, as according to the World Health Organization an estimate of 6.8 million people die yearly from diseases of neurological origin [10]. The highest incidence is due to Parkinson, epilepsy and Alzheimer, among others. The largest cause of disability and impairment among population above 65 is due to this group of diseases. Besides its enormous social impact the neurological diseases present also a strong economical burden, being responsible of at least 5 % of hospitalizations.

In the western world at least one out of 20 persons demanded a neurological disease consultation from the medical services during 2005. The resources devoted to their treatment were of around 139 billion euros in the Euro Zone during that same year. The evolution of the age pyramid and the increment in the number of elder people in the next years all over the first world will mean a larger impact of neurological diseases. It is important to remark that the diagnosis of these illnesses is rather complex because some of their symptoms may be related with other diseases as well. The attention to the patient with neurological disease will demand larger and more specialized sanitary resources to allow a more accurate and early diagnose. Early detection of these diseases is considered a key factor contributing to efficient attention and treatment. This early warning can be made possible from correlates of the neuromotor functions, as handwriting, eye movement, walking, as well as from electro-encephalography (EEG), functional Magnetic Resonance (fMR), etc. [6]. Voice alterations are good correlates due to the non-invasive character and simple methodology of voice inspection, which consists in obtaining samples involving common phonation styles. The rhythm, prosody, fluency and other related concepts have been largely used by neurologists to evaluate and grade the extent and severity of the neurological pathology [9]. In this sense the advances in the reconstruction of the glottal excitation and the biomechanical correlates found in this signal, including vocal fold tension patterns provided by the organic pathology detection methodology already developed [2] may be applied in the early detection of the neurological disease, as the voice production apparatus, and especially the larynx, is largely dependent on the muscular tone granted by complex neuromotor circuits and terminals. Tremor and spasm as correlates of dystonia are common symptoms appearing relatively often in a list of non-organic dysphonias. The paper is organized as follows: the relationship between the neurological disease and voice and speech production is reviewed in section 2. In section 3 the methodology used in voice pathology detection and grading is briefly presented. In section 4 the application of this methodology to the detection and grading of a kind of neurological alteration of voice known as spasmodic dysphonia which presents neuromuscular alterations to the control of vocal folds is presented, and some results are given from a study case, accompanied by a brief discussion. Conclusions and future lines are presented in section 5.

2 Neurological Diseases and Voice Production

Voice pathologies are disorders of the phonatory system of different origins which manifest distinct symptoms. Classically two main groups are distinguished: organic and neurological. The first group is related with alterations found in the physiological structures of the larynx (nodules, polyps, edemae, sulci, cysts, granulomae, papilomae, carcinomae, etc). The second group does not show clear organic alterations in the larynx, and have to see more with neuromuscular diseases. Their origin has to be found in the central neural system (cortical

areas and centres related with language production) or within neural transmission (brain stem, vagus nerve, larynx muscle innervation, etc.). According to the symptoms shown many of them can be associated with certain neurodegenerative diseases as Parkinson, Alzheimer, Permanent Non-Fluent Apraxia, etc. Spasmodic dysphonia is one of them, which involves involuntary transient muscle spasms which interfere with voice production resulting in voice breaks either within vowels in the adductor form of the disorder or during vowel onsets after voiceless consonants in the abductor form. The adductor dystonia is due to an over-stressing spasm, which presses both vocal folds above the average, resulting in over-stressed glottal cycles during the episode (sharper and more abrupt). On the contrary, abductor dystonia is manifested as a decay of the muscular tone resulting in an opening of the passage between vocal folds, which eventually may stop vibrating. Spasmodic dysphonia may present different origins (see Fig.1), of which the neurological type differentiates from the psychologic and idiopathic variants in that it is not reversible by voice therapy [7]. This makes it a good study model as it may convey information on neurological underlying malfunction.

The common treatment for neurologic dysphonia is either by botulinum toxin infiltration in the muscle (temporal effects) or by surgical disruption of

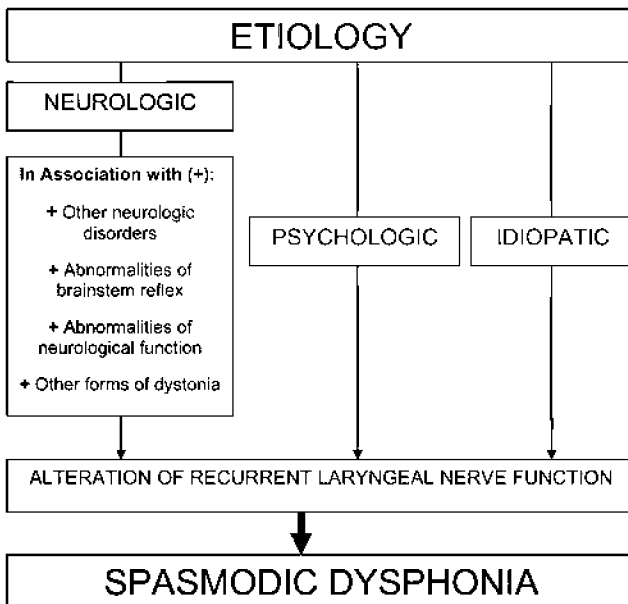


Fig. 1. Classical differentiation of spasmodic dysphonia attending to the underlying cause

innervations (permanent effects). Most of the times a clear diagnostic differentiating among the different variants is rather complex [7]. One of the objectives of the present study is to introduce possible inspection tools based on acoustic analysis of voicing speech by means of estimates of the biomechanical correlates of the vocal folds, which could be used in further discriminating the different origins of this set of dysphonias. Spasmodic dysphonia has been selected as a target because it shows clear biomechanical correlates as it will be presented in the sequel. This approach could open the study of other pathologies of neurological origin as well [3].

3 Detection and Grading Dysphonic Voice

Once the glottal source has been reconstructed different parameters are obtained from its power spectral density envelope and timely evolution. These are estimated for a large database including normophonic and dysphonic male and female subjects [8]. The normophonic set is carefully modelled on the most sensitive parameters by Gaussian Mixtures as in Fig.3. This allows estimating a normophonic condition probability from sample \mathbf{y}_t from subject t with respect to the normophonic model $\Gamma_{Mm,f}$ (m: male, f: female populations) as:

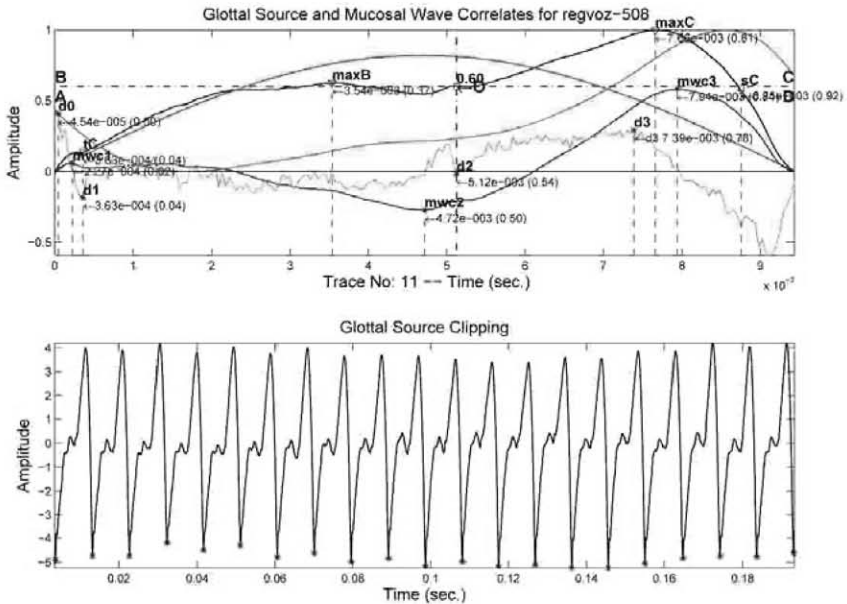


Fig. 2. Glottal Source Template (normophonic male voice). Top: Time Domain Parameterization. Bottom: Phonation Cycle Clipping. The classical Liljencrants-Fant (LF) pattern [5] may be clearly appreciated as a recovery phase (0-maxB), a closed resting phase (maxB-O), an opening phase (O-maxC) and a closing phase (maxC-0).

$$p(\mathbf{y}_t/\Gamma_{Mm,f}) = \sum_j w_j \left\{ \frac{1}{(2\pi)^{r_{m,f}/2} |C_{Mm,f}|^{1/2}} e^{-1/2(\mathbf{y}_t - \Psi_{Mm,f})^T C_{Mm,f}^{-1} (\mathbf{y}_t - \Psi_{Mm,f})} \right\}_j \quad (1)$$

where $\Psi_{Mm,f}$ and $C_{Mm,f}$ are the mean vector and covariance matrix of the population distribution. Once this probability has been calculated for all the gaussians j in the mixture weighted by the mixing vector w_j a likelihood ratio (A) may be defined the normophonic to dysphonic conditional probabilities:

$$A(\mathbf{y}_t) = \log[p(\mathbf{y}_t/\Gamma_{Mm,f})] - \log[1 - p(\mathbf{y}_t/\Gamma_{Nm,f})] \quad (2)$$

The log-likelihood ratio is contrasted against a threshold resulting in a decision of normophony vs pathology as:

$$L_t = \begin{cases} N; A(\mathbf{y}_t) \geq \vartheta \\ D; A(\mathbf{y}_t) < \vartheta \end{cases} \quad (3)$$

where N stands for normophonic and D for dysphonic. If a subject's voice is labelled as dysphonic the grade is of most importance to decide on how to proceed as most pathologies manifest as mild dysphonias which may go unnoticed to the general practitioner (GP).

The cases marked by green circles are the normophonic set, distributed accordingly with the three most relevant biometrical parameters (22 and 23 are singularity marks on the glottal source power spectral density, whereas 45 is a biomechanical unbalance, for a complete explanation see [5]). Subject names are printed near each mark. The red diamonds are the dysphonic set. The blue stars are samples from the same subjects before (T1636 and T1637) and after treatment (T2636 and T2637). A prototype gaussian is tentatively placed on the normophonic distribution centroid. Mahalanobis distances from a given sample to the respective gaussian centroid are used to grade dysphonia. This grade is well correlated with subjective estimations of the grade in GRBAS scale [4]. The detection Receiver Operator Curves after pathology detection from cross-validation tests are shown in Fig.4.

The accuracy of the detection process is based on a trade-off between the sensitivity (ratio of detected dysphonic subjects to total number of dysphonic subjects) and the specificity (ratio of detected normophonics to the total number of normophonic subjects). This balance is given by the point in the DET curve where the percentage of false dysphonic and normophonic subjects become equal, which is around 3 %. The voice analysis methodology would then consist in obtaining a sample of voice from the subject, estimating the spectral and temporal correlates on the glottal source, contrasting them against the normophonic/pathologic database, estimating the dysphonic grade and deriving the case and its anamnesis to specialized services [2].

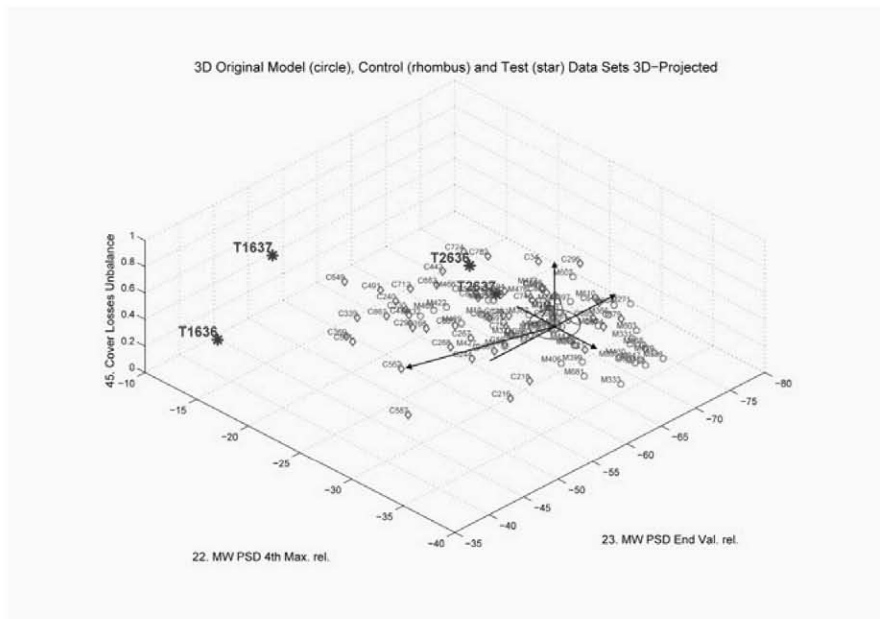


Fig. 3. Estimating the distance of a given subject sample to the normophonic model set as an index of normophonic vs dysphonic voice (female set). Green: samples from normophonic subjects. Red: samples from dysphonic subjects. Blue: Pre-post treatment cases: subjects before treatment (T1636 and T1637) and the same subjects after treatment (T2636 and T2637).

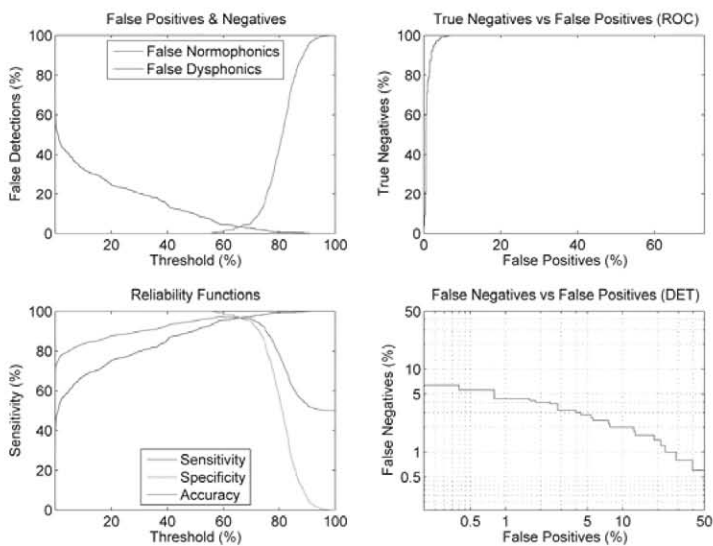


Fig. 4. ROC and DET curves from cross-validation of GLOTTEX for pathology detection in a female data set of normophonic and dysphonic cases. Top left: False detection cases as a function of the threshold. Top right: Associate Receiver Operator Characteristic (ROC) curve. Bottom left: Sensitivity, Specificity and Accuracy as a function of the threshold. Bottom right: Detection-Error Trade-off (DET) curve. The equal tax point is around 3 %.

4 Study Case: Abductor Spasmodic Dysphonia

Spasmodic dysphonia is a neurological disease characterized by the inability to maintain constant the phonation tension in vowels and other voiced sounds. This is due to strong involuntary changes in the neuromuscular activity of the larynx, resulting in fluctuations of the vocal fold tenseness. Pitch does not change much, but the amplitude of voice suffers large changes as seen in Fig.5.

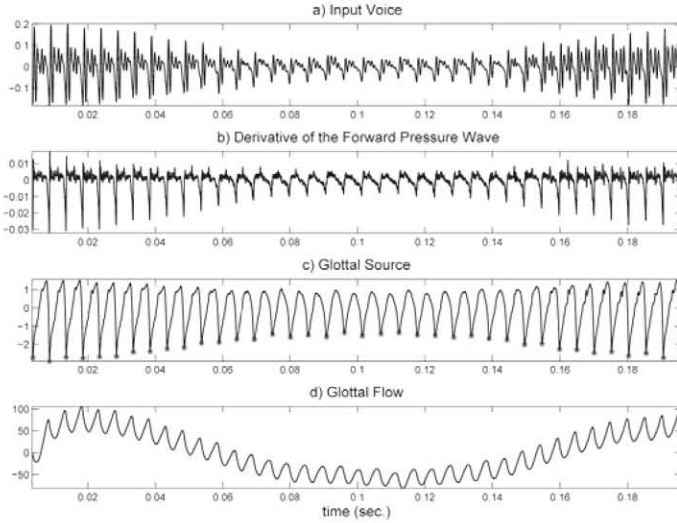


Fig. 5. Episode of spasmodic dysphonia of abductor type. a) Voice trace. b) Residual after vocal tract removal. c) Glottal Source (estimate of supraglottal pressure wave). d) Glottal flow (the fluctuation due to integration effects). All magnitudes are in relative units.

This tendency would be reflected in a smooth decaying and recovering of the shimmer (distortion parameter measuring the relative change in amplitude between neighbour phonation cycles). This case was graded as pathologic level 1 in subjective assessment, and 2 in objective assessment [4]. The detailed results of the detection process may be seen in Fig.6. The upper template gives the glottal source (black line), which is related to the supraglottal pressure, raising from a minimum to a maximum following a glottal cycle. The opening point is given by (O). It may be seen that the glottal source in this case is rather different than the characteristic LF pattern in Fig.2, as the plateau observable in it before the opening does not exist. The closed phase (AO) is shorter than in a normal

phonation cycle (34 % of the total phonation cycle), the open interval (OD) is much larger than expected (66 %). The lower template shows the evolution of the glottal arches at the beginning of the dystonic episode, where the LF plateau is still perceptible as a hunchback, and at the middle of the episode, where that mark is not present.

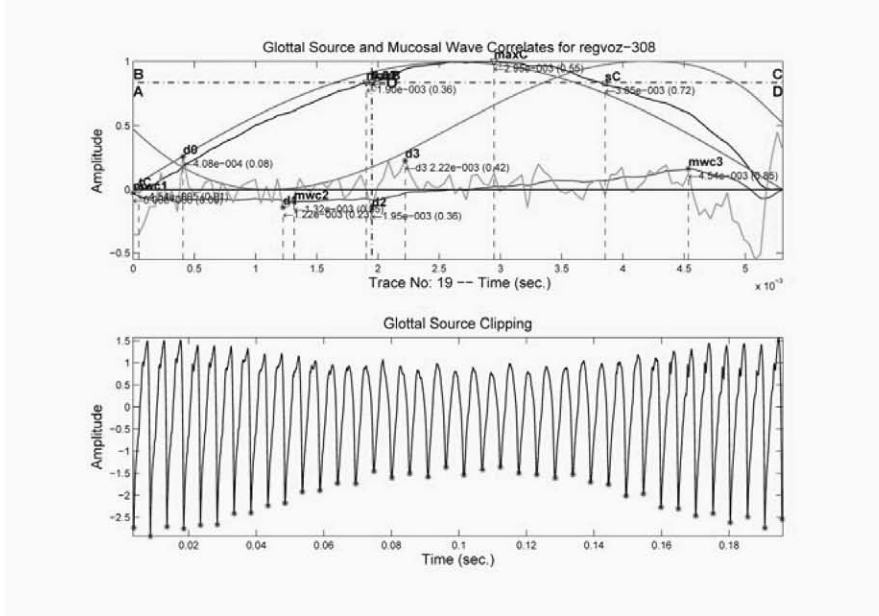


Fig. 6. Close-up view of the glottal source in the middle of the dystonic episode (upper template, black line), corresponding to a cycle in the middle of the lower template

This means that the extreme glottal cycles are almost normal, and the middle ones are under-stressed. The most important characteristic of the methodology used (see [5]) is the possibility of estimating the biomechanical parameters of the vocal fold given as the dynamic mass, losses and stiffness correlates of the fold body and cover. The templates in Fig.7 show the estimates of the dynamic mass of the vocal fold body for each phonation cycle (upper left). This parameter experiences a 13 % decay from normal to dystonic phonation. The boxplot in the upper right shows the median, first and third quartiles illustrating the resulting statistical spread. The median dynamic mass is around 13.5 mg. The middle left and right templates show similar estimates for the losses (due to viscosity and turbulence), which experience an increment of about 12 % during the episode. The bottom left and right give the corresponding estimates for the vocal fold body stiffness, measuring the strain experienced by the average vocal fold structure along its extension, and directly related with muscular tension.

It may be seen that the tension decays in around 30 % during the episode, to recover immediately after. The question which now arises is how biomechanical parameter estimations, and especially tension, may help in monitoring the neurological disease. The answer must be modulated depending on the specific task to fulfil. The classical ones in organic pathology are detection, grading, early warning, pre- and post-treatment monitoring, among others.

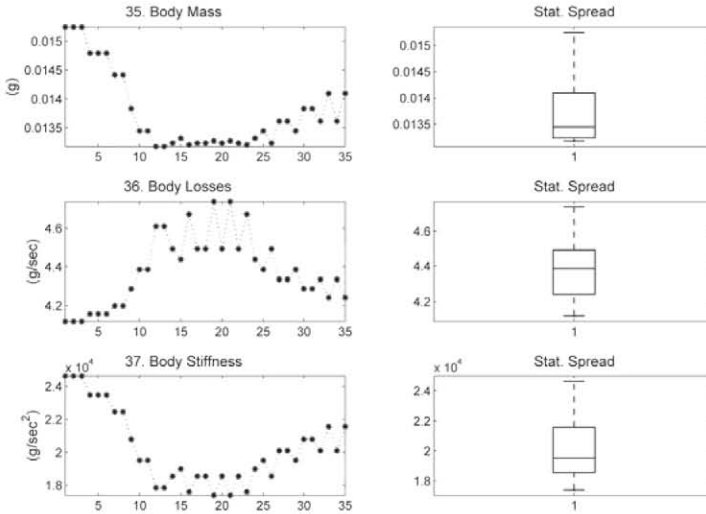


Fig. 7. Monitoring an episode of spasmodic dysphonia with GLOTTEX (female voice). Top two templates: Glottal Source in detail and clipping (compare them with those from normal phonation in Fig.2). Bottom three templates (right): Estimates of vocal fold body mass, losses and stiffness. Left: statistical dispersion boxplots. The dystonic correlate of the disease is evident in the decay appreciated in vocal fold tension (stiffness) during the episode.

Pre- post-monitoring is by far the most complex task to track the grades obtained before and after treatment (surgical, pharmaceutical or voice therapy), relative to a historical discourse. The key point is to establish intra- and inter-subject distributions of the most discriminating parameters. As the historical account is very important for pre-post monitoring intra-subject distributions may have more relevance than in organic pathological studies. This is the case of spasmodic dysphonia, showing important intra-speaker distribution changes, indicating that biomechanical parameter unbalance may play a very special role. These questions are subjects of further study.

5 Conclusions

Through the present work several important facts have been presented, the relevance of which is to be emphasized. On one hand the importance of the neurological disease has been presented. Among different diseases of this kind spasmodic dysphonia has been selected as a study target because, having in most cases a neurologic origin it might show correlates with vocal fold biomechanics under the working hypothesis. The classical methodology to detect the dysphonic condition was then proposed as a possible candidate to characterize spasmodic dysphonia using the biomechanical parameters of the vocal fold. This was confirmed by the estimation of the mass, losses and stiffness of the fold body. Especially this last one showed a clear correspondence with the abductor-type of pathology monitored. Of course, this case of study will need a confirmation against a wide set of cases including adduction and both male and female subjects. Besides, the neurologic origin of each case studied has to be carefully asserted, and the dysphonic grade must be evaluated both subjectively as by the GRBAS index as well as by the dysphonic index proposed in [4]. This would allow to give this study the statistical validation required to consolidate the conclusions commented, defining the next steps in the study. The most interesting fact is that this methodology may open ways to characterize and quantify the degree of spasmodic dysphonia, and by extension to other neurological diseases, helping in early warning and monitoring.

Acknowledgements

This work is being funded by grants TEC2009-14123-C04-03 from Plan Nacional de I+D+i, Ministry of Science and Technology of Spain and CCG06-UPM/TIC-0028 from CAM/UPM. Special thanks are due to the support and invaluable help of the doctors at the ORL service in HU "Gregorio Marañón" of Madrid, especially to Drs. Scola and Ramírez without whose encouragement and push forward this project would not have been possible.

References

1. Bobadilla, J., Gómez, P., Godino, J.I.: Mapaci: A Real Time e-Health Application to Assist Throat Complaint Patients. In: Proc. of ICIW 2007 (2007)
2. Colton, R.H., Kasper, J.K., Leonard, R.: Understanding voice problems, pp. 107–151. Williams and Wilkins, Baltimore (1990)
3. Das, R.: A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications* 37, 1568–1572 (2010)
4. Gómez, P., Fernández, R., Rodellar, V., Godino, J.I.: Evaluating the grade of voice pathology from the glottal-source spectral profile. In: Proc of AVFA 2009, Madrid, pp. 161–164 (April 2009)
5. Gómez, P., Fernández-Bafllo, R., Rodellar, V., Nieto, V., Álvarez, A., Mazaira, L.M., Martínez, R., Godino, J.I.: Glottal Source Biometrical Signature for Voice Pathology Detection. *Speech Communication* 51, 759–781 (2009)

6. Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.J.: Automatic classification of MR scans in Alzheimer's disease. *Brain* 131(3), 681–689 (2008)
7. Ludlow, C. L. Spasmodic Dysphonia is a Neurological Disorder. National Spasmodic Dysphonia Association, www.dysphonia.org (retrieved 2.1.2011)
8. Project MAPACI, <http://www.mapaci.com>
9. Rosenfeld, D.B.: Neurolaryngology. *Journal of Ear Nose and Throat* 66(8), 323–326 (1987)
10. World Health Organization, <http://www.who.int/healthinfo>