Abstract: Neurological degenerative diseases are becoming a growing concern in modern society. The successful treatment of these diseases depend greatly in early detection. Speech has been routinely used by specialists as a valuable correlate in the assessment of pathological disease. Specifically voicing can serve as a very introspective correlate for this practice. The present paper uses a methodology previously employed in organic pathology voice quality assessment to explore to what extent specific low-level correlates of neurological diseases may be established. The methodology uses voiced recordings of sustained vowels to estimate vocal fold visco-elastic parameters from inverse filtering. These parameters show to be clearly influenced by unstable neuronal spiking resulting in tremor which affects many phonation cycles. The possible modeling of tremor could be used as an index to neuro-motor problems in phonation and help in differential diagnose of the pathology at an early stage. The paper presents examples on parameter estimations from study cases of spasmodic dysphonia and Parkinson Disease. Further development of research lines on this estimation methodology is also addressed.

Keywords: Inverse Filtering, Vocal Fold Biomechanics, Parkinson Disease, Voice Quality Assessment, Tremor

I. INTRODUCTION

Classically Voice Quality Analysis has been focused to detect and establish the organic pathology in voice resulting from pathological alterations of larynx physiology. The study of other sources of dysphonic voice finding their ultimate reasons in the alterations of the neurological paths controlling phonation have been tagged as "functional" or "non-organic". Voice resulting from altered phonation due to neurological reasons may be a most valuable report of the etiology and progress of neural diseases affecting the production of voice, such as pathologies resulting in voice tremor [1]. These would include some kinds of spasmodic dysphonia, stammering and Parkinson. The possibility of early detection in the first stages of Parkinson's Disease (PD) may grant a better preventive treatment reducing the progress of the illness [2]. Monitoring treatment by objective methods is also an important goal, especially in modifying or defining new protocols. The deepest foundations of the methodology proposed in this paper are to be found in tracking the malfunctioning of neurological and neuromuscular paths involved in voice production (see Fig. 1).
vocal fold adduction and abduction. 5. Branch of the vagus nerve (phrenic) actuating on the diaphragmatic muscles. 6. Feedback loop in Basal Ganglia damping muscular tone.

These comprise links from the neuromotor linguistic cortex [3] to the subthalamic region [4] and through the laryngeal nerve and their associated pathways [5][6] to the muscles activating the thyro-arytenoid structure, responsible in the last term of vocal fold stretching, adduction and abduction (Superior Laryngeal Nerve, Internal and External Laryngeal Branches of the Inferior Laryngeal Nerve, Transverse and Oblique Arytenoid Muscles -TOAM-, and Cricothyroid Muscles -CM). Any alteration in the functionality of these pathways and in the associated muscles will result in temporary distortions of the parameters of tension and dynamic mass contribution of the vocal folds, both on the body and the cover biomechanics. Correlates of these alterations will be found in the pitch, and in long term jitter and shimmer, as the periodicity of these alterations may be of hundreds of milliseconds [7]. The aim of this paper is to give some phenomenological account in detecting and grading the neurological disease using biomechanical correlates obtained from the inverse filtering of voice. The technology has been tested in monitoring pre-post treatment of organic pathology, and due to its ubiquitous character can be applied as well to the neurological disease.

II. METHODOS

A database of voice recordings from neurological disease-affected patients is being recorded in Hospital del Henares of Madrid. This geographical area South East of Madrid is specially sensitive to PD. Being a heavy industrial area it is believed that some environmental factors may be responsible of the largest incidence of PD among the aging population compared to other regions of Madrid. For the preliminary and explorative character of the present study some specific cases are selected, these being strong spasmodic and PD voice samples, pathological voice of organic origin and voice from normophonic patients to serve as a contrast (all of them females). These voices are inverse filtered and some biometrical and biomechanical parameters are estimated, as the glottal closure sharpness, the noise/glottal ratio, dynamic mass and tension of the vocal fold body. It may be shown that these indices show a strong correlation with the spasmodic episodes both in their timely evolution and statistical dispersion. The methodology is based on the following steps:

1. Three emissions of the vowel /a/ are recorded at 44,100 Hz under normal phonation conditions.
2. For specific statistical comparison they are low-pass filtered and re-sampled to 22,050 Hz. High-pass filtering at 25Hz is also applied to eliminate low frequency flickering effects. Frames of 0.4 s long are used in the analysis.
3. Inverse Filtering is applied, and the glottal source is reconstructed [8]. Estimations of the glottal closure sharpness, noise/glottal ratio, dynamic mass and tension of the vocal fold body are derived following [8].

III. RESULTS

An episode of spasmodic dysphonia (SD) has been selected from the database to show the possibilities of the methodology, corresponding to a female voice (32 year old) manifesting about 2-3 spasms per second. The record is a segment of 0.4 s long from a sustained phonation of vowel /a/ (see Fig. 2 and Fig. 3).

Fig. 2. Episode of spasmodic dysphonia. Templates from top to bottom: Voice signal. Inverse filtering residual. Glottal source. Glottal flow.

Fig. 3. Left templates from top to bottom: Phonation cycle-synchronous estimates of the dynamic mass component of the vocal folds, friction losses and body stiffness. Right templates from top to bottom: Statistical distributions of the left templates given as boxplots.
The voice segment studied is a part of a recording of a sustained /a/ 0.4 s long where an episode of spasm is clearly recognizable by the amplitude decay. The reconstruction of the glottal source (template c) does not show such a strong decay in amplitude. Pitch ranges from 208-203 Hz in the sections out of the spasm to a minimum of 185 Hz during the spasm, following an almost regular fluctuation (tremor of about 2.5 Hz). It may be seen that the estimates of the dynamic mass of the vocal fold body, and especially the fold tension are highly correlated with the spasm, reporting changes of about 25% and 50% of variation respectively. Similar fluctuations are found in other distortion parameters, such as the sharpness of the closure spike in the glottal source, the noise/glottal energy ratio and some cepstral parameters of the glottal source spectral density.

Fig. 4. Phonation 0.4 s long from a patient affected from Parkinson Disease. Templates from top to bottom: Voice signal. Inverse filtering residual. Glottal source. Glottal flow.

A second example from a patient (72 year old) affected by Parkinson Disease (PD) corresponding also to female voice has been analyzed following the same methodology. The record is a segment of 0.4 s long from a sustained phonation of vowel /a/. The results of the analysis are reported in Fig. 4 and Fig. 5. In this case the changes in amplitude are not as relevant as in the spasmodic case. The reconstruction of the glottal source (c) does not show important changes in amplitude as well. Pitch ranges from 240-256 Hz following an irregular fluctuation (tremor) of about 5 Hz. The estimates of the vocal fold body dynamic mass and stiffness report changes of around 20% . To put the analysis into context at this point it would be worth to compare some overall results for these two cases against results from a normal female speaker and a pathological female speaker. The normal speaker (NF) is a 34 year old female, non-smoker not having reported any problem with voice, volunteering for the study. Normal condition was assessed by endoscopy and EGG. The case with organic pathology corresponded to a female 22 year old having been diagnosed from a left vocal fold cyst (LVFC) affecting the contralateral fold (contact lesion). The case was graded 2 (severe) in GRBAS scale. Endoscopy and EGG availed the diagnose. The acoustic processing of the four cases included the extraction of pitch, relative jitter and shimmer and the noise/glottal energy ratio (NGE). The stiffness of the vocal fold body was estimated as well. The results are given in Table 1 at the end of the paper.

IV. DISCUSSION

From the results in Table 1 the first consequence is that pathological data (except for PD) are clearly differentiated from normal data in the value of the dispersion (standard deviation) and in the stiffness of the vocal fold body. Mean values of the classical distortion estimates as jitter, shimmer or NGE do not show important differences among the pathological cases except in PD. This case shows distortion parameters which could be considered normal. The problem is that tremor in PD is observed as FM-modulations which do not leave clues in the jitter, whereas the SD case may be traced in shimmer. Going to the causes, it seems that the effects of FM-affected spiking producing tremor in SD may be observed on the specific muscles affecting vocal fold abduction and adduction (TOAM-CM) as well as in the muscles responsible for pressure build-up and sustenance in lungs during phonation (diaphragm). The influence of FM-affected spiking in the modulation of the vocal tract (naso-velar switch, glossomuscular and oro-labial complexes) could also introduce changes in the production of voice, interfering with vocal-fold induced tremor. These differences may affect the results observed, as in the two cases studied. In the spasmodic case the...
important changes in amplitude observed could be associated with some influence of the spasm on the diaphragm and other muscles inducing subglottal pressure, besides affecting strongly to the vocal fold stiffness as a result of the TOAM-CM action. The result during the spasm is a dystonic relaxation of the vocal folds (abduction) accompanied by a decay in subglottal pressure. The case of the PD patient may have to see only with the action of the TOAM-CM, resulting in a relatively cyclic dystonic behavior of the vocal fold but not in important changes of the subglottal pressure. It seems that parameters tracking amplitude changes as shimmer or APPQ measured directly on the glottal source, as well as the indirect estimates of vocal fold tension may serve as important marks to produce differential diagnose in tremor-affected dysphonias, and this line should be further studied. Other possible correlates could be the sharpness of the closure instant and the lowest cepstral coefficients of the glottal source spectral profile. This means that the study of tremor as a result of neurodegenerative diseases may require complex time-frequency analytical techniques. Chaotic modeling of tremor in stiffness and other correlates, and Wavelet Transform may be good candidates out these studies.

V. CONCLUSIONS

The first conclusion from this phenomenological description is that tremor appears as a mark in certain biomechanical estimates of vocal fold dynamics as body stiffness. Therefore the monitoring and modeling of tremor could be based on the study of these correlates. Indications that differential diagnose could also be based in combined amplitude-stiffness indices are plausible enough for the issue to deserve further study. The analysis of the mentioned correlates estimated directly from the glottal source obtained after vocal tract inversion instead of whole voice may be a beneficial methodology to unveil and quantize the extent or degree of the spasmodic or tremor illness. As the characterization of tremor in voice shows quasi-cyclic information, techniques to model this characteristic as chaotic attractors, wavelets, or ARMA coefficients may be of much higher resolution than the analysis of full voice. The monitoring of neurological diseases is of most importance in a world where the aging of general population will demand important resources for health care. The early detection and monitoring of these problems may help in devising more efficient treatment protocols. Routine voice tests may help in this task. The validation of this methodology for PD is in due course in cooperation with the ENT and Neurology Services at Hospital del Henares.

REFERENCES


Table 1. Parameter values for the four cases studied (standard deviations between parentheses)

<table>
<thead>
<tr>
<th>Subject/Parameter</th>
<th>Pitch Hz</th>
<th>Jitter %</th>
<th>Shim %</th>
<th>NGE %</th>
<th>Stiffness (g.s²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#346 (34y NF)</td>
<td>199 (1.15)</td>
<td>0.7 (0.6)</td>
<td>1.9 (1.3)</td>
<td>8.4 (0.5)</td>
<td>19542 (185)</td>
</tr>
<tr>
<td>#341 (22y LVFC)</td>
<td>215 (7.04)</td>
<td>4.2 (3.6)</td>
<td>3.8 (2.4)</td>
<td>6.6 (1.9)</td>
<td>24857 (4487)</td>
</tr>
<tr>
<td>#308 (45y SD)</td>
<td>199 (6.02)</td>
<td>1.5 (1.5)</td>
<td>6.5 (3.3)</td>
<td>8.9 (1.7)</td>
<td>22168 (2656)</td>
</tr>
<tr>
<td>#337523 (72y PD)</td>
<td>248 (3.87)</td>
<td>0.7 (0.5)</td>
<td>1.4 (1.0)</td>
<td>6.5 (1.1)</td>
<td>25138 (988)</td>
</tr>
</tbody>
</table>