ACOUSTIC ASSESSMENT OF SPASMODIC DYSPHONIA USING A NEW MULTIPURPOSE VOICE ANALYSIS TOOL

A. Giordano[1], P.H. Dejonckere [2,3,4] , C. Manfredi[1]

[1] Department of Electronics and Telecommunications, Università degli Studi di Firenze, Firenze, Italy
2Utrecht University, Utrecht, The Netherlands
3Federal Institute of Occupational Diseases, Brussels, Belgium
4Catholic University of Leuven, Leuven, Belgium

Abstract: **A new multipurpose voice analysis tool named BioVoice2, suited for the analysis of strongly irregular signals and long sentences, is applied on voices of patients diagnosed with adductor spasmodic dysphonia before and after treatment with botulinum toxin injection. The speech material consists of 40 short German sentences phonetically selected to be constantly voiced. Nine acoustic parameters were taken into account from all those estimated with BioVoice2. Significant improvement of voice quality was estimated by a subset of these parameters related to increased voicing, improved regularity of vocal fold vibration, reduction of spasms and faster speech rate. BioVoice2 proves to be a useful tool for objectifying voice quality also in case of strong signal irregularity.**

Keywords: **Spasmodic dysphonia, voice analysis, acoustic parameters, botulinum.**

I. INTRODUCTION

Spasmodic dysphonia (SD) is a particular voice disorder characterized by involuntary movements of one or more muscles of the larynx during speech.

The most common form of this pathology is the adductor SD (ADSD) that is considered in this study. The ADSD is expressed with different severity from case to case, from mild cases that present only a slight tremor of the voice and occasional breaks to cases where severe spasms of the vocal cords make it impossible to speak, preventing airflow through the glottis. In such cases the patient's work and social life are compromised and this can also frequently lead to severe depression [1].

With (AD)SD, deviant acoustic events as aperiodicity, phonatory breaks and frequency shifts perturb fluency and intelligibility. These voices thus require specific acoustic parameters for an exhaustive analysis [2-4].

The present study is based on a new multipurpose voice analysis tool, named BioVoice2, developed under MatLab environment, capable to deal with highly irregular voice signals as those under study.

The aim of the present study is to test the ability of BioVoice2 to evaluate the improvement in patient's voice quality using objective parameters and, accordingly, the effectiveness of the medical treatment that consist of botulinum toxin injection in the vocalic muscles.

II. METHODS

Currently most of the software tools for voice analysis have limitations related to the level of irregularity in the voice and to their applicability to running speech instead of sustained vowels only [5]. To overcome these limitations, we designed a multipurpose program, BioVoice 2 that is applicable to the analysis of a wide range of voice signals, including the analysis of long sentences (several minutes of connected speech) other than short ones or sustained vowels only.

The main targets in designing BioVoice 2 were:

- Implementing a robust and reliable fundamental frequency (F0) estimation and a Voiced/Unvoiced (V/U) selection procedure, applicable to quasi-stationary/noisy signals such as highly hoarse, irregular voices and/or sentences;
- Allowing for the analysis of long sentences (several minutes) other than short ones or sustained vowels only;
- Giving the user a simple Graphic User Interface (GUI) that does not require any manual setting by the user, thus being well suited also for non-expert users.

BioVoice 2 performs the analysis of audio files resulting in objective parameters that are considered useful by clinicians in the diagnosis of voice disorders. It has been successfully tested on synthesized sustained vowels giving better results than most commonly used software tools when applied to strongly irregular and hoarse voice signals [9]. The main parameters of interest in the present work are:

F0: the fundamental frequency is estimated with a two steps procedure. First, Simple Inverse Filter Tracking is performed, obtaining a raw F0 estimation and its range of variation (F_L, F_H) where F_L= lowest F0 value and F_H =

Claudia Manfredi (edited by), *Models and analysis of vocal emissions for biomedical applications : 7th international workshop : August 25-27, 2011*, ISBN 978-88-6655-009-9 (print) ISBN 978-88-6655-011-2 (online)

highest F0 value. In the second step, F0 is estimated inside (F_L, F_H) with the Average Magnitude Difference Function (AMDF) approach [9]. The program provides F0 tracking and its mean, standard deviation, minimum and maximum values.

PVF: the ratio between the number of voiced frames and the total number of frames, that depends on the breaks which are present in the voice: the more the pauses, the less the PVF.

PVS: the percentage of voiced speech frames, which means the ratio of voiced frames over the frames that have been classified as speech in a previous step of analysis. Speech frames are those in which the zero crossing rate is less than 3000 zero-crossings per milliseconds and the energy exceeds a threshold value that depends on the signal characteristics. Hence PVS should be higher or equal to PVF. On a sustained vowel and for a healthy voice, PVS is ideally 100%. As a general rule the better the voice, the higher both PVF and PVS.

PFU: the percentage of frames that have an unreliable F0 among the total number of frames. This parameter is therefore a measure of the fundamental frequency F0 instability. Frequency variations make F0 unstable. In a frame F0 is evaluated as unreliable if it has a deviation of more than 25% compared to the average F0 value over all voiced frames. In this case the better the voice, the lower the PFU percentage.

VL90: the 90th percentile of voicing length distribution, defined as the maximum number of consecutive voiced frames found. The sharp breaks featuring the voice of patients with SD reduce this parameter.

Duration: the total time required to the patient for pronouncing sentences. As a general rule a healthy voice, that is more fluent, will have a shorter duration than a pathological one.

In addition BioVoice 2 evaluates the time duration of the voiced and unvoiced part of the signal, and the average length of voiced frames (mean duration of voicing, **MDV**).

Jitter: a measure of the degree of variability of the period length. It gives a measure of the aperiodicity of the signal measuring the changes in fundamental period T0=1/F0 from period to period. Of course, good voices have low jitter. Jitter J is evaluated here according to Eq.1:

$$J = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i+1}\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \qquad (1)$$

Where N is the number of frames and T_i is the i-th period length.

Corrected jitter: the correction means that only frames with reliable F0 are taken in account. F0 is reliable if it has less than 25% deviance from the mean value of F0 of all voiced frames. The formula for Corrected jitter is the same as for the jitter.

NNE: Normalized Noise Energy is a noise estimation method that relies on a comb filtering approach: it is the ratio of the energy between the harmonics and the whole signal energy [8].

Moreover, BioVoice2 allows for the estimation of the signal spectrogram, formants, Power Spectral Density and other parameters related to the kind of voice signal under analysis (adult male, adult female, newborn cry and singing voice) that are not described here. Plots and tables can be displayed, printed and saved in an easy way. Details can be found in [5].

To test the capability of BioVoice2 of analyzing long sentences in a reliable way, 24 audio files (12 pre- and 12 post-treatment) from 12 German patients diagnosed with ADSD are considered here. Each patient read a standardized list of 40 German sentences for a total duration of about 2’30”. These sentences are phonetically selected by clinicians for being constantly voiced. This is in fact supposed to increase the sensitivity for detecting interruptions of vocal fold vibrations induced by SD.

Audio files are provided in uncompressed audio wave format with sampling frequency Fs= 44.100 Hz and 16 bit of resolution. All the recordings were made in a quiet room by one of the authors of this work.

III. EXPERIMENTAL RESULTS

Table 1 reports the mean value of the parameters previously described, obtained from pre and post-treatment recordings. From Table 1 a clear trend towards better voice quality is shown (post-treatment values higher or lower than pre-treatment ones, according to the specific parameter).

Table 1 – Mean value of the acoustic parameter computed by BioVoice2.

Parameter		PRE	POST
PVF %	Mean	51.80	69.45
PVS %	Mean	52.82	69.54
PFU %	Mean	50.29	44.81
Jitter %	Mean	14.63	13.76
Corrected Jitter %	Mean	6.12	6.24
VL90 [s]	Mean	0.0008	0.0160
Duration [s]	Mean	142.07	137.68
MDV [s]	Mean	0.277	4.03
NNE [dB]	Mean	-17.49	-17.58
Mean F0 [Hz]	Mean	180.65	188.62
Std F0 [Hz]	Mean	46.05	50.05

These results show that the parameters PVF, PVS, VL90 as well as MDV are strongly indicative of voice improvement, while for the other parameters the pre-post difference seems less significant.

Figure 1 shows the difference between pre- and post-treatment values of the most relevant acoustic parameters that are PVF, PVS, VL90 and MDV.

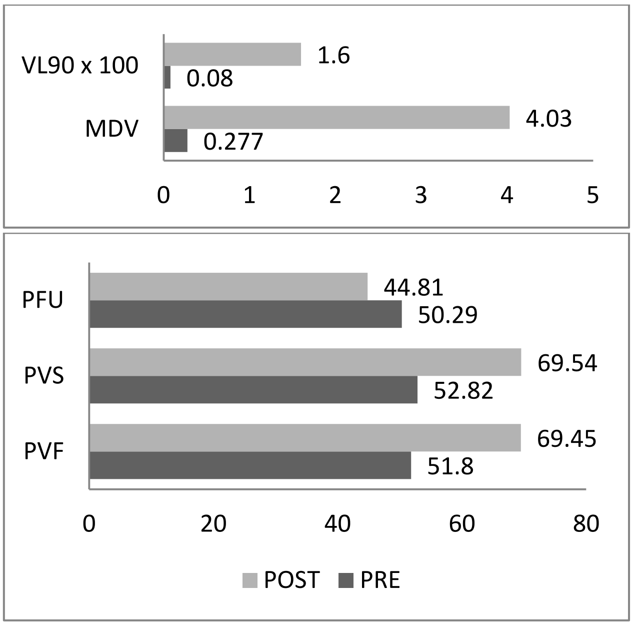

Fig. 1: Upper: mean value of VL90 and MDV (seconds). Lower: mean value of PVF, PVS, PFU (%) for pre-post treatment data.

Moreover the Wilcoxon test was applied on each parameter separately. Results are reported in Table 2.

Table 2 – Results of the Wilcoxon test for all the acoustic parameters.

Wilcoxon's test	
Parameter	***P***
meanF0	0.5186
Std F0	0.3804
PVF	0.0269
PVS	0.0161
PFU	0.3394
Jitter	0.8501
Corrected jitter	0.8984
NNE	0.9697
VL90	0.0342
Duration	0.7334
MDV	0.0425

As expected, only few of the acoustic parameters reveal a significant post- vs. pre- improvement. Specifically these parameters are: PVF, PVS, VL90 and MDV. In particular Jitter, Corrected jitter, PFU, NNE and even duration show no statistically significant differences and are thus not suited for evaluating the improvement of voice quality with the present data. As these parameters have different measurements units and ranges a standardization step was performed according to the following equation [7]:

$$Z_i = \frac{x_i - \bar{x}}{\sigma} \qquad (2)$$

Where x_i is the variable to be standardized, $\bar{x}$ is its mean value and σ is its standard deviation.

Figure 2 shows the boxplot of pre- post-treatment data for all the acoustic parameters considered here. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles respectively, and the whiskers extend to the most extreme data points that are not considered outliers. The small circles are the outliers. Data are standardized according to Eq.2.

The plot confirms the best results obtained with parameters PVF, PVS, VL90 and MDV.

IV. Discussion

Results in Table 2 show that only four of the whole acoustic parameters considered here are capable to point out a significant post- vs. pre-treatment improvement in voice quality. These parameters are all related to the increased voicing capability of the patient after medical treatment.

Hence only the acoustic parameters that are in some way related to the selection of voiced/unvoiced parts of the signal are successful in the analysis of a long sentence, while other, and also jitter, seem to have less relevance. This result suggests that a different analysis should be performed on fluent speech other than that usually made on sustained vowels or short sentences.

Results are in agreement with previous studies made on the same speech material where a different analysis program was used [7]. However, differently from [7], with BioVoice2 more parameters, such as a noise measure, F0 and its standard deviation can be included in the analysis.

Moreover, a new parameter was introduced here for the first time, namely the mean duration of voiced frames, MDV. From the preliminary results presented here (Table 1 and Table 2), this parameter seems indeed to be very promising in evaluating the quality of voice in long sentences.

V. Conclusion

A new multipurpose voice analysis tool is presented here and its performance is evaluated on fluent speech coming from patients affected by adductor spasmodic dysphonia. Even if the data set consists of a limited number of patients, significant changes in the value of acoustic parameters were found comparing the pre- and post-treatment recordings, pointing out the improvement of voice quality after botulinum toxin treatment. Some parameters such as jitter, already proved valid in the analysis of short sentence or sustained vowels, seem to lose meaningfulness when evaluated on long sentences. Even the PFU parameter, that is a measure of the

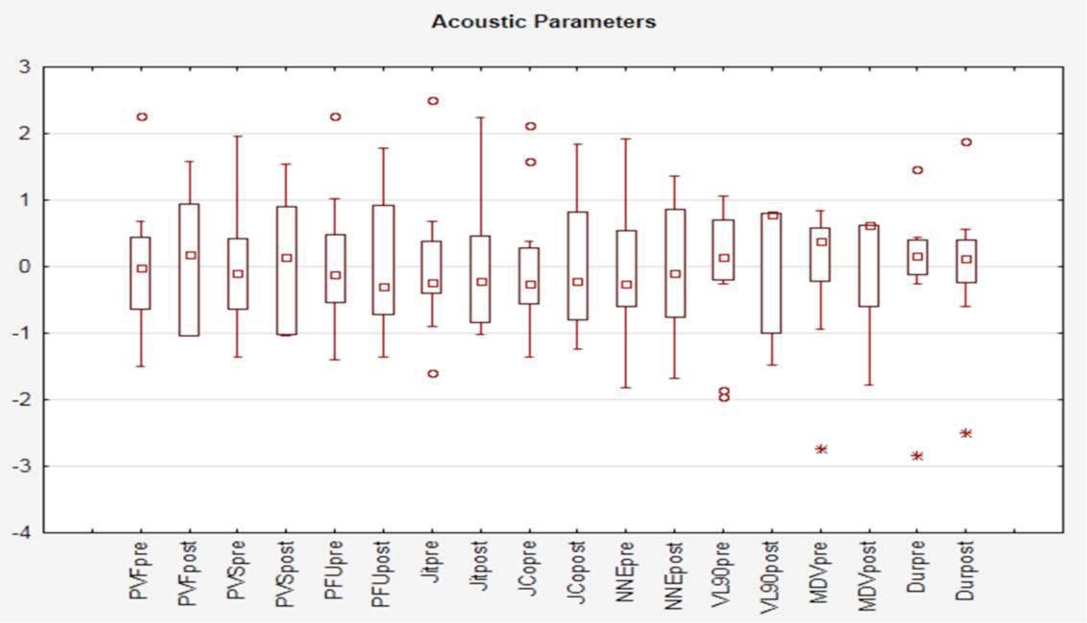

Fig. 2: Boxplot of pre-post treatment acoustic parameters.

fundamental frequency instability, seem to lose its capability in evaluating the voice signal in long sentences.

However, the proposed tool is successful in objectifying the increased voicing and the improved regularity of vocal fold vibration after treatment. One newly defined parameter, the mean value of voiced frame duration, seems very promising in evaluating voice quality improvement when applied to long sentences. Future work will be devoted to refining the tool in order to reduce the computational time while preserving its high resolution capabilities and robustness against noise. The tool will be also tested on a new corpus of synthetic signals with varying F0 and formants that should mimic fluent speech.

REFERENCES

[1] Baylor CR, Yorkston KM, Eadie TL. *"The consequences of spasmodic dysphonia on communication related quality of life: A qualitative study of the insider's experiences."* J. Comm. Disorders. 2005; 38:395–419.

[2] Dejonckere P.H., Bradley P., Clemente P., Cornut G., Crevier-Buchman L, Friedrich G, Van De Heyning P, Remacle M., Woisard V.,*"A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessments techniques"*, Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS), *Eur. Arch. Otorhinolaryngol.* 258, pp.77-82, 2001.

[3] Dejonckere P.H., *"Critères acoustiques de fluence pour l'évaluation des dysphonies spasmodiques. In : Voix parlée et chantée. "* C. Klein – Dallant, Ed. Paris. 63 – 73. ISBN 978-2-9528061, 2007.

[4] Sapienza CM, Cannito MP, Murry T, Branski R, Woodson G : "*Acoustic variations in reading produced by speakers with spasmodic dysphonia pre-Botox injection and within early stages of post-Botox injection*." J Speech Language Hearing Res 45: 830 – 843, 2002.

[5] Manfredi C, Giordano A, Schoentgen J, Fraj S,Bocchi L, Dejonckere PH, "*Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools*" Biomedical Signal Processing and Control, 2011 (in print).

[6] Siemons-Luhring DI, Moerman M, Martens JP, Deuster D, Muller F, Dejonckere PH, *"Spasmodic dysphonia, perceptual and acoustic analysis: presenting new diagnostic tools"* European Archives of Oto-Rhino-Laryngology, Vol. 266, n. 12, 2009.

[7] Dejonckere PH, Moermann KJ, Merman MBJ, Martens JP, *"Perceptual and acoustic assessment of adductor spasmodic dysphonia pre- and post-treatment with botulinum toxin"*, Proc. 3rd AVFA International Worshop, 18th-20th May 2009, Madrid (Spain).

[8] Kasuya H, Ogawa S, Mashima K, Ebihara S, *"Normalised Noise Energy as an Acoustic Measure to Evaluate Pathologic Voice"*, J. Acoust. Soc. Am., vol. 80, n.5, p.1329-1334, 1986.

[9] Manfredi C, Giordano A, Schoentgen J, Fraj S, Bocchi L, Dejonckere PH *"Reliability of voice analysis software tools for highly irregular signals Part II: the effect of noise"* Logopedics Phoniatrics Vocology, 2011 (in print).