ASSESSING THE EFFECTIVENESS OF BOTULINUM TREATMENT IN SPASMODIC DYSPHONIA

C. Manfredi, G. Cantarella, N. Migali*, A. Berlusconi*, B. Maraschi**

Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3, 50139 Firenze, Italy
* Otolaryngology Unit, Ospedale Maggiore di Milano IRCCS, Via F. Sforza 35, 20122 Milano, Italy

Introduction

Spasmodic dysphonia (SD) is a disabling voice motor disorder, characterised by unintentional and uncontrollable spasms of the laryngeal muscles during phonation. In extreme situations, it implies loss of speech intelligibility. Adductor SD is due to hyperadduction of the vocal folds. Videolaryngoscopy shows healthy laryngeal structures, as well as normal respiratory and sphincteric functions.
At present, botulinum toxin infiltration, under electromyographic guide, is the most common and effective treatment for this kind of pathology [1], [2]. It yields long lasting decrease of muscle strength, without systemic effects. After treatment, glottic closure is stabilised, with enhanced efficiency. However, SD was shown to be a heterogeneous pathology, both as far as its severity and ways of compensation are concerned. Hence, a personalized treatment is required.
Therefore, objective measures and tracking of voice parameters is of relevance, in order to assess the effectiveness of the adopted technique. Quality of speech is evaluated by the clinician by means of subjective indexes, which reflect how the signal is perceived by listeners. Moreover, objective speech quality measures are considered, as good predictors of subjective quality [3], [4]. The MDVP software tool (Kay Elemetrics) is recognised as a standard for reference, being the most widely used tool in the biomedical field devoted to voice analysis. It provides a number of indexes, to assess voice quality. However, in some cases, and especially when the analysis is performed on complete words rather than sustained vowels, tracking voice parameters, such as noise (roughness, hoarseness), rather than simply measuring their mean values, was proven to be of importance for diagnosis and treatment evaluation [5].

Materials and methods

This paper deals with the problem of quantifying voice quality, before and after drug treatment with botulinum toxin, in patients suffering from SD. First results are presented, obtained with both MDVP and some new voice analysis tools, based on robust estimators for fundamental frequency F_0 (Simple Inverse Filter Tracking and Wavelets), noise (Normalised Noise Energy) and formants (parametric Autoregressive (AR) PSD estimation) [5], [6], [7]. Specifically, a novel feature is the introduction of an adaptive noise estimation technique that allows tracking varying noise level during phonation. This could be of help for the physician, in order to evaluate the effort made by the patient in pronouncing complete words, besides sustained vowels only. The method, named ANNE (Adaptive Normalised Noise Energy), relies on a comb filtering approach [8], and has been proved effective in many applications [6], [7]. Large negative ANNE values correspond to good voice quality, while values close to zero reflect the presence of noise.
Moreover, a robust parametric formant estimation technique is proposed, obtained by peak picking in the PSD, evaluated on short time windows of varying length inversely proportional to local F_0 (see Fig. 1, middle plot, blue stars), and based on AR models of order equal to the signal sampling frequency F_s (in kHz). The varying window length allows following fast signal variations, with enhanced results as far as resolution is concerned [9], [10], [11]. Notice that choice of a model order equal to F_s, as suggested in the literature, prevents from spectral smoothing and consequently loss of spectral peaks.
Finally, new objective quality measures (PSD ratios, SNR) are defined, that easily allow assessing enhancement of voice and comparing pre and post-treatment results. A "harmonic range" is defined, given by frequencies below the threshold f_{th}=4 kHz, while the "noise range" refers to frequencies above f_{th}. The choice of f_{th} is based on the usual range for voiced sounds (first four formants) in adults, as well as on experimental results obtained from threshold tuning in a dataset of voiced and unvoiced sounds [6]. However, it can be changed to different values, if required.
Specifically, the following indexes are proposed, where the subscript "non-filt" refers to the original signal, while "filt" refers to the post-treatment signal:

$$PSD = 10\log_{10}\frac{PSD_{non-filt}}{PSD_{filt}} \qquad (1)$$

that represents the ratio of the PSDs, evaluated on the whole frequency range;

$$PSD_{low} = 10\log_{10}\frac{PSD_{non-filt}(f \le 4kHz)}{PSD_{filt}(f \le 4kHz)} \qquad (2)$$

that measures the ratio of the PSDs evaluated on the "harmonic range";

$$PSD_{high} = 10\log_{10}\frac{PSD_{non-filt}(f \ge 4kHz)}{PSD_{filt}(f \ge 4kHz)} \qquad (3)$$

i.e. the ratio of the PSDs, evaluated on the "noise range".
An effective treatment should give PSD and PSD_{low} values below zero (harmonic power enhancement after treatment), but PSD_{high} values above zero (loss of power due to noise). Finally, a measure of the denoising effectiveness due to treatment is defined as:

$$SNR = 10\log_{10}\frac{\sum_{n=1}^{M} y^2(n)}{\sum_{n=1}^{M}(y(n)-y_{filt}(n))^2} \qquad (4)$$

where: y(n) = pre treatment signal sample at time n, $y_{filt}(n)$ = post treatment signal sample at time n. SNR is thus the ratio between the noisy signal energy and that of removed noise. Negative SNR values correspond to voice quality enhancement.

Plot overlap and fine graphical display enables easy readable results. The proposed approach is thus suited for integrating the MDVP features. When properly optimised, these tools could be implemented on a DSP board, as a portable device useful for both clinicians and patients, also for rehabilitation purposes, after surgery or medical treatment. A prototype is under study.

Experimental results

Sustained vowel /a/ from 10 patients (age 36-77 years, mean 61.8, 7 females), affected by adductor SD, has been recorded and main voice features extracted by means of both MDVP and the proposed approach, before and after treatment with botulinum toxin. Moreover, for three patients, the Italian word /aiuole/ has been recorded and analysed.

The following plots and table show some results for one patient, pronouncing the word /aiuole/. Figure 1 shows the signal (black, top), F_0 tracking (red, middle) with the corresponding varying analysis window (blue, middle) and noise tracking (black, bottom), before (upper) and after (lower) treatment. Notice highly unstable and unreliable F_0 values before treatment. More stable F_0 values and lower noise level are clearly visible in the lower plots. Moreover, the noise plot highlights the huge, constant effort made by the patient in speaking before treatment (mean NNE=-4.64 dB), as compared to its strong reduction after treatment (mean NNE=-19.6 dB). However, the lower plot shows that some effort is still required after treatment, in the time intervals 0s-0.1s (at the beginning of the word) and 0.7s-0.8s (transition during /ol/), as highlighted by NNE values close to zero.

Fig. 2 depicts the signal spectrogram before (upper) and after (lower) the treatment. Strong noise, highly irregular behaviour and loss of harmonics are evident in the upper plot, with great enhancement in the lower plot. These plots are to be compared to those reported in Fig. 3, where formant tracking is shown. This figure clearly shows irregular and often unresolved formants before treatment, and their almost complete recovering after treatment. Finally, Fig. 4 compares pre (blue) and post treatment (red) PSDs, highlighting energy and harmonics recovering, as well as noise reduction after treatment. This is confirmed by the values of the indexes reported in the plot: as explained above, SNR=-11.26 dB, PSD_{tot}=-11.91 dB, PSD_{low}=-13.05 dB, being strictly negative, confirm enhancement in voice quality, while PSD_{high}=5.1 dB, being positive, quantifies the reduction of the noise component in the high frequency spectral region.

The following Table summarises some results obtained for this patient, both with MDVP (1-6) and with the proposed tool (7-11). (DVB=degree of voice breaks; DUV=degree of voiceless).

Table 1 - Results obtained for one patient, word /aiuole/. Black: MDVP; red: proposed approach

/aiuole/	1	2	3	4	5	6	7	8	9	10	11
	Jitter%	**Shimmer%**	**NHR**	**F_0 mean**	**DVB**	**DUV**	**NNE**	**SNR**	**PSD_{tot}**	**PSD_{low}**	**PSD_{high}**
pre	**11.5**	**25.13**	**1.33**	**558.17**	**56.09**	**87.72**	**-4.64**				
post	**2.98**	**5.4**	**0.162**	**155.35**	**8.2**	**38.63**	**-19.62**	**-11.26**	**-11.92**	**-13.05**	**5.1**

All parameters confirm increased voice stability (especially 1, 2, 5, 6) and enhanced voice quality (3, 7-11). Notice that F_0 before treatment was almost unrecoverable, and hence erroneously found by both methods. After treatment, both methods give quite similar results, although MDVP fails to recover F_0 in the time range 0.8s – 0.9s.

Final remarks

First results concerning patients suffering from SD are presented. The aim is that of assessing voice quality recovery after botulinum toxin treatment, by means of objective indexes and easily readable plots. Good results are obtained with MDVP, integrated with new indexes and plots, obtained with robust voice analysis techniques. Further work will concern refining the proposed technique and testing new parameters, possibly derived from non-linear analysis techniques.

References

[1] Blitzer A, Brin MF, Stewart CF., "Botulinum Toxin Management of Spasmodic Dysphonia (Laryngeal Dystonia): a 12-year experience in more than 900 patients", *Laryngoscope* Vol. 108, p. 1435-1441, 1998.

[2] Cantarella G, Galardi G., "Il Trattamento Botulinico in Fonochirurgia". In: Casolino D, Ricci Maccarini A. Fonochirurgia endolaringea. Pisa: Pacini Ed., p.142-154, 1998.

[3] Hirano M., "Psycho-Acoustic Evaluation of Voice", In Hirano M. (Ed.): "Clinical Examination of Voice", (Springer-Verlag, New York), 1981.

[4] Dejonckere P.H., Remacle M, Fresnel-Elbaz F., Woisard V., Crevier-Buchman L., Millet B., "Differentiated Perceptual Evaluation of Pathological Voice Quality: Reliability and Correlations with Acoustic Measurements", *Rev. Laryngol. Otol. Rhinol*., 117, n.3, p.219-224, 1996.

[5] Manfredi C. , "Adaptive Noise Energy Estimation in Pathological Speech Signals", *IEEE Trans. Biomed. Eng.*, 47, p.1538-1542, 2000.

[6] Manfredi C., Peretti G., "Robust Techniques for Pre- and Post-Surgical Voice Analysis", Proc. Eurospeech Conf., August 29-31, 2003, Geneve, CH.

[7] Manfredi C., Dori F., Iadanza E. , "Improvement in Hoarse Voice Denoising for Real-Time DSP Implementation", Proc. Voqual'03, August 27-29, 2003, Geneve, CH, p.63-68.

[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., D'Aniello M., Bruscaglioni P., Ismaelli A., "A Comparative Analysis of Fundamental Frequency Estimation Methods with Application to Pathological Voices", *Medical Engineering And Physics*, Vol.22, n.2, p.135-147, 2000.

[10] Fort A., Ismaelli A., Manfredi C., Bruscaglioni P., "Parametric and non Parametric Estimation of Speech Formants: Application to Infant Cry", *Medical Engineering And Physics*, Vol.18, n.8, p.677-691, 1996.

[11] Manfredi C., D'Aniello M., Bruscaglioni P., "A Simple Subspace Approach for Speech Denoising", *Logopedics Phoniatrics Vocology*, 26, p.179-192, 2002.

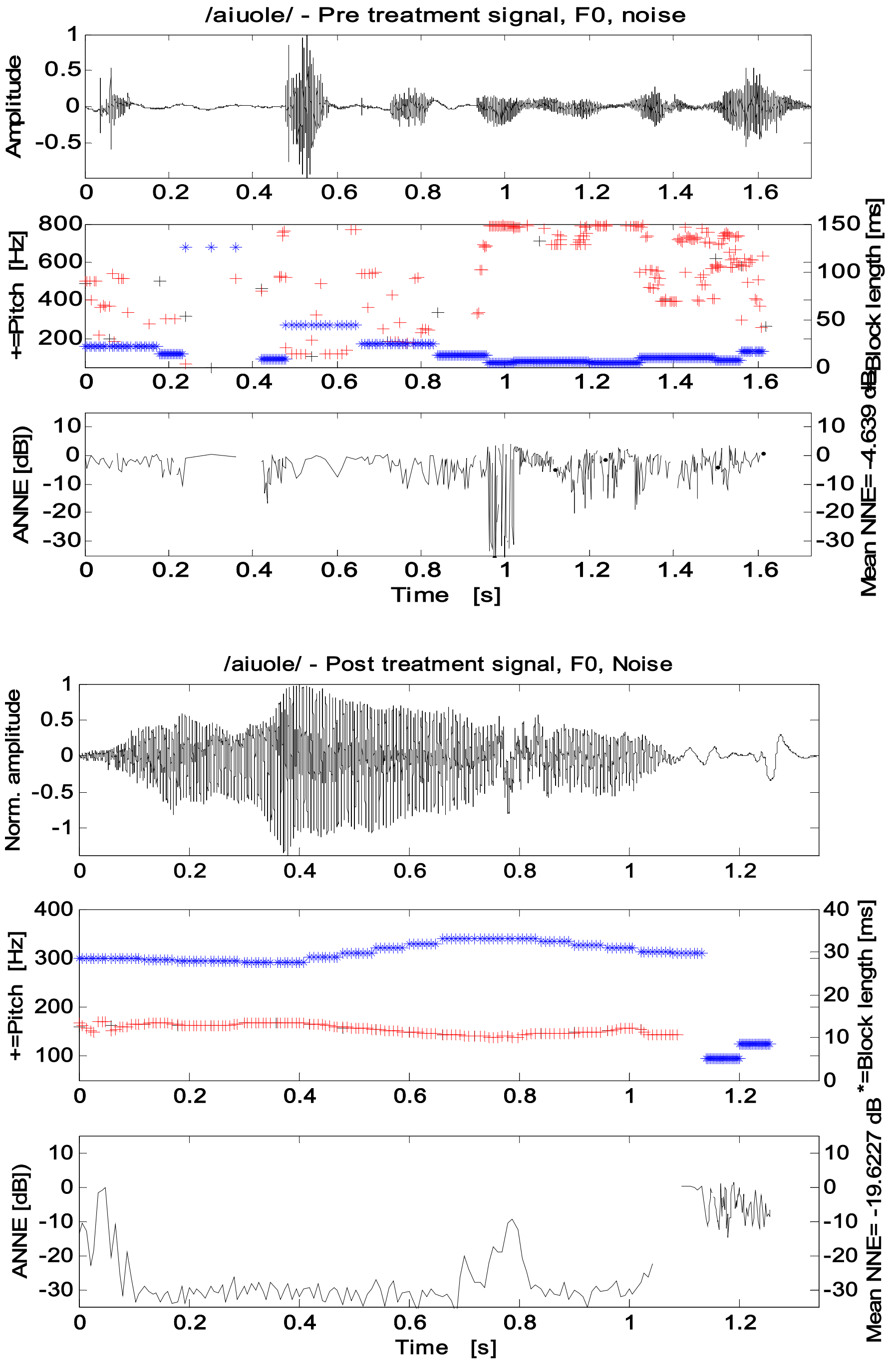

Figure 1 – Plot of the signal, fundamental frequency and noise before (upper) and after treatment (lower).

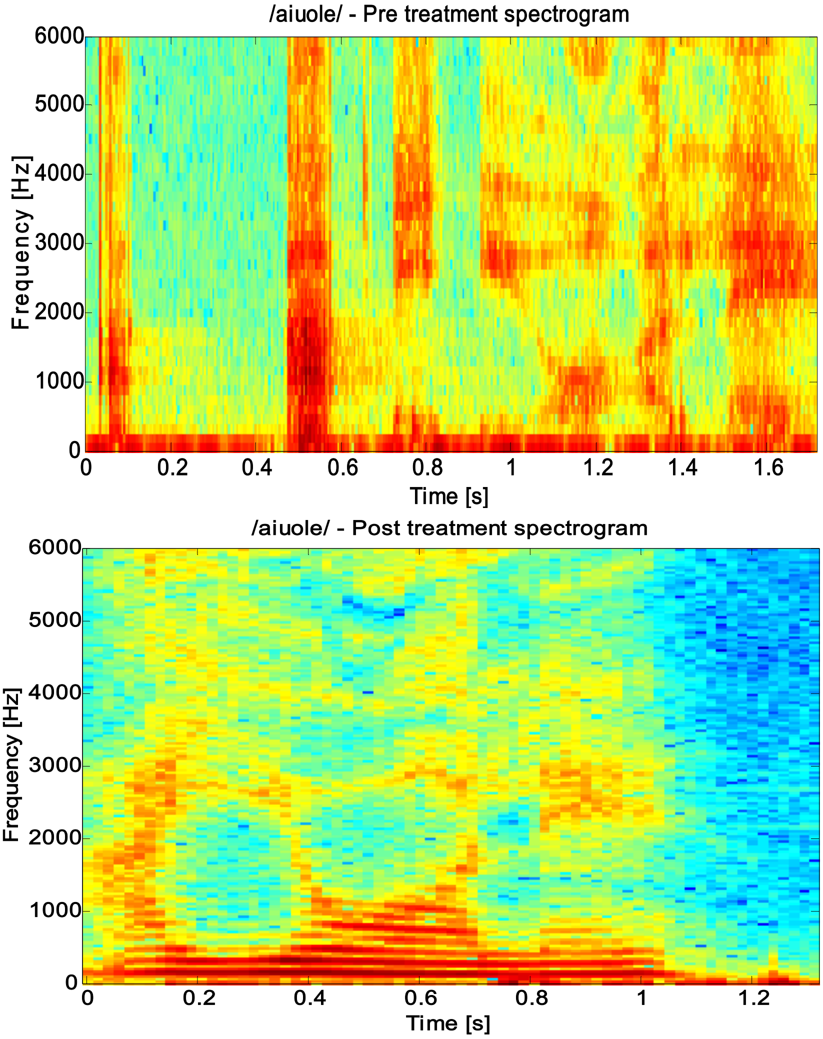

Figure 2 – Signal spectrogram before (left) and after (right) the treatment with botulinum toxin.

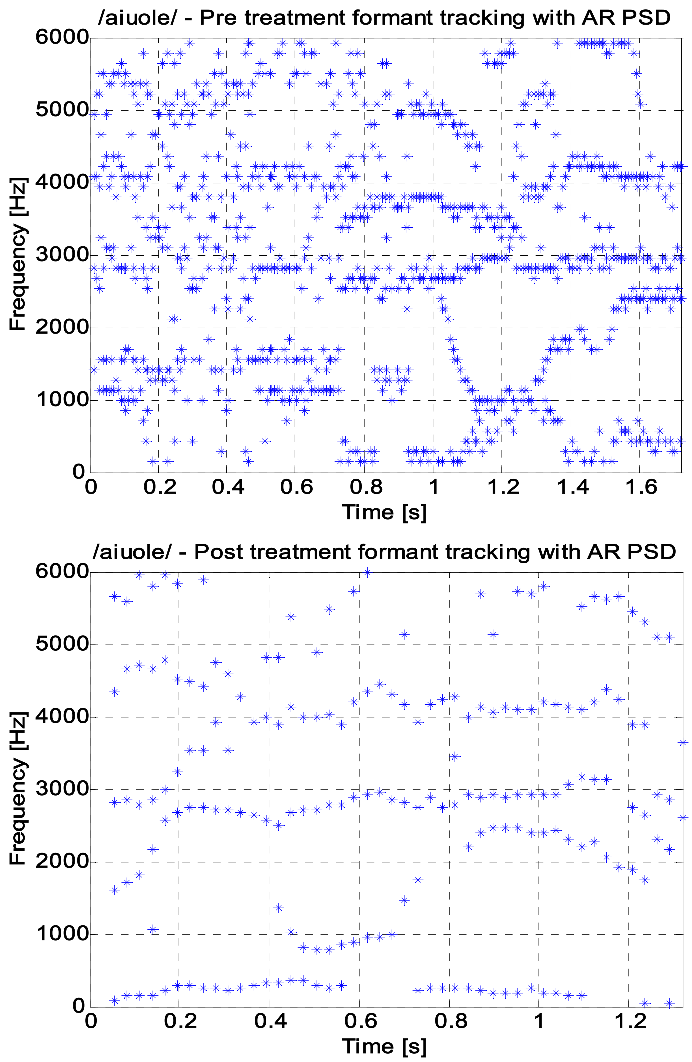

Figure 3 – Formant tracking obtained with the parametric AR PSD estimation. The order of the AR model is set equal to the signal sampling frequency (in kHz).

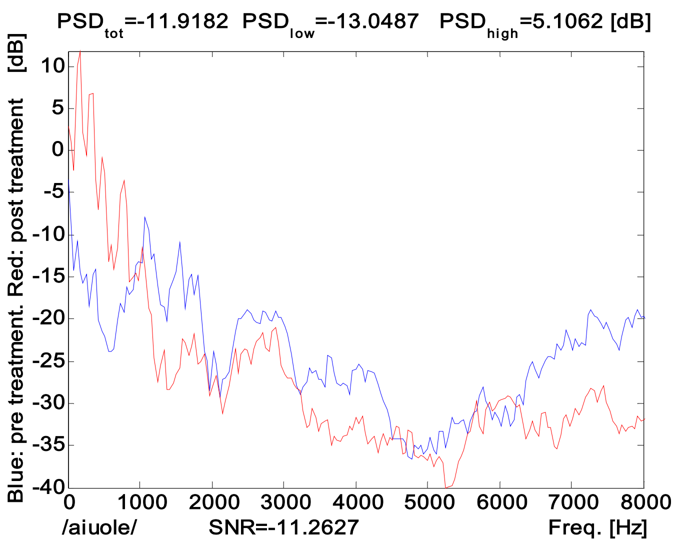

Figure 4 – Plot of PSD before and after treatment. Main parameters are also reported.