Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

A Robust Tool to Compare Pre- and Post-Surgical Voice Quality

Claudia Manfredi, *Member IEEE,* Riccardo Canalicchio, Giulio Cecconi, Giovanna Cantarella

Abstract - **Assessing voice quality by means of objective parameters is of great relevance for clinicians. A large number of indexes have been proposed in literature and in commercially available software tools. However, clinicians commonly resort to a small subset of such indexes, due to difficulties in managing set up options and understanding their meaning.**

In this paper, the analysis has been limited to few but effective indexes, devoting great effort to their robust and automatic evaluation. Specifically, fundamental frequency (F_0), along with its irregularity (Jitter (J) and Relative Average Perturbation (RAP)), noise and formant frequencies, are tracked on voiced parts of the signal only. Mean and std values are also displayed. The underlying high-resolution estimation procedure is further strengthened by an adaptive estimation of the optimal length of signal frames for analysis, linked to varying signal characteristics.

Moreover, the new tool allows for automatic analysis of any kind of signal, both as far as F_0 range and sampling frequency are concerned, no manual setting being required to the user. This makes the tool feasible for application by non-expert users, also thanks to its simple interface.

The proposed approach is applied here to patients suffering from cysts and polyps that underwent micro-laryngoscopic direct exeresis (MLSD).

1. INTRODUCTION

Assessing voice quality recovering by means of objective parameters is of great relevance for clinicians. It is however a difficult task, due to different analysis techniques and reference values to be used. A huge number of indexes have been proposed in literature, some of which implemented in commercially available software tools [1]. However, clinicians commonly resort to a small subset of such indexes, due to difficulties in understanding subtle differences among parameters, and to deal with rather technical options, especially concerning spectral analysis. Moreover, often commercial software suffers some limitation, linked to the implemented analysis techniques that sometimes prevent the analysis of highly degraded voices.

This paper aims at providing objective parameters and plots, easily understandable and usable by clinicians and logopaedicians, in order to assess voice quality recovering after vocal fold surgery or therapy. The proposed software tool performs pre- and post-surgical comparison of main voice characteristics (fundamental frequency, noise, formants) by means of robust analysis tools, specifically devoted to deal with highly degraded speech signals as those under study.

2. MATERIALS AND METHODS

The problem of tracking fundamental frequency F_0, noise level and formants during a voiced emission is considered, in order to assess voice quality recovering after vocal fold surgery. For pathological voices, usually such parameters considerably oscillate, due to the effort made by the dysphonic patient in speaking. The basic idea is that of implementing robust analysis techniques, capable to deal with highly irregular signals such as those coming from pathological voices. To this aim, the signal is divided into short frames, whose length adaptively varies according to varying signal characteristics: the higher the F_0 the shorter the frame length (kept fixed to 3 pitch periods). A voiced/unvoiced separation algorithm is implemented, to avoid F_0 estimation on signal frames that have no harmonic content.

F_0 tracking is achieved by means of a two-step procedure, based on well-established results: the AMDF approach is applied to a wavelet-smoothed SIFT estimation of F_0, with optimised and varying adaptive filter order [2], [3]. Among the huge number of available parameters for quantifying F_0 irregularities, Jitter (J) and Relative Average Perturbation (RAP) were recognised by the physicians of relevance in most applications and implemented here on previously obtained signal frames. Specifically, according to literature, and as in [1]:

1. Jitt - Jitter Percent /%/ - Relative evaluation of the period-to-period (very short-term) variability of F_0 within the analyzed voice sample. Cycle-to-cycle irregularity can be associated with the inability of the vocal cords to support a periodic vibration for a defined period. These types of variations are typically associated with hoarse voices.
2. RAP - Relative Average Perturbation /%/ - Relative evaluation of the period-to-period variability of F_0 within the analyzed voice sample with smoothing factor of 3 periods. The smoothing reduces the sensitivity of RAP to

Manuscript received April 2, 2007.

C.Manfredi is with the Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3, 50139 Firenze, Italy (Corresponding author. Phone: +39-055-4796410; fax: +39-055-494569; e-mail: manfredi@det.unifi.it).

R.Canalicchio is with the Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3, 50139 Firenze, Italy. (e-mail:manfredi@det.unifi.it).

G.Cecconi is with the Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3, 50139 Firenze, Italy. (e-mail: manfredi@det.unifi.it).

G.Cantarella is with the Otolaryngology Department, University of Milan, Ospedale Maggiore IRCCS, Via F. Sforza 35, 20122 Milano, Italy. (e-mail: giovanna.cantarella@policlinico.mi.it).

1-4244-0788-5/07/$20.00 ©2007 IEEE

pitch extraction errors. Hoarse and/or breathy voices may have an increased RAP.

J and RAP mean and standard deviation (std) over the whole signal are also evaluated and displayed.

An adaptive noise estimation technique is implemented, that allows tracking varying noise level during phonation. For pathological voices, spectral noise is in fact closely related to the degree of perceived hoarseness. In this paper, noise variations are tracked by means of an adaptive version of the Normalised Noise Energy method, named ANNE (Adaptive Normalised Noise Energy) [3], [4]. It relies on a comb filtering approach, optimised in order to deal with data windows of varying length. Large negative ANNE values correspond to good voice quality, while values close to zero reflect the presence of strong noise. The method has already been successfully applied to pathological voices, to compare pre- and post-surgical voice quality in case of tyroplastic medialisation [5].

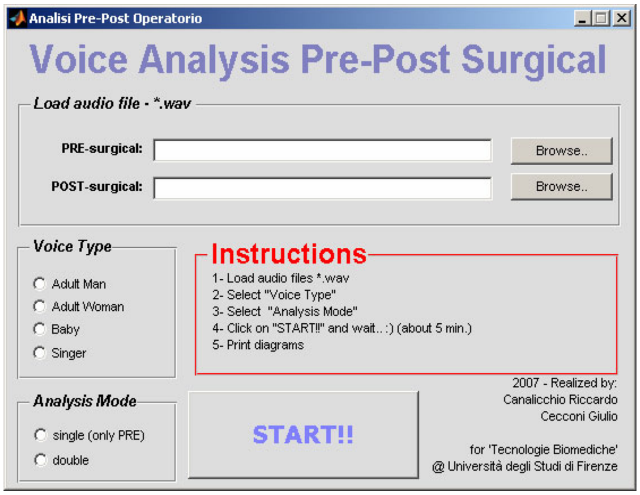
Finally, robust and high-resolution formant estimation is implemented, based on parametric AutoRegressive (AR) PSD evaluation. The AR model order p is automatically selected by the program according to patient and signal characteristics, based on the relation $p = 2LF_s/c$, where: F_s = sampling frequency, L = vocal tract length (linked to patient's age), and c = sound speed [6]. Colour-coded spectrogram is also provided, with the tracking of the first three formants superimposed. Mean values and std are also shown. Comparing pre- and post-surgical spectrograms gives a first qualitative view of the residual noise present in the voice signal before and after surgery, as well as of harmonics intensity and stability. PSD plots complete the set of pictures, allowing visual inspection of possible harmonic energy recovering. On the plot, PSD_{tot}, PSD_{low}, PSD_{high} quantify the signal global energy, the low-frequency and the high-energy one, respectively. SNR is also provided. These indexes could further help the clinician in assessing voice quality recovering.

A user-friendly interface (Fig. 1) allows selecting age, sex and type of vocal emission for each patient, performing computations without any other requirement. The software tool automatically adjusts internal settings for optimal frame length, frequency range of analysis and plots. Specifically, the interface allows for:

- selecting pre-post surgical data (.wav files);
- choosing the voice type, ranging from high-pitched new-born and singers voices to adult voices: the overall allowed F_0 range is $40Hz<F_0<1300Hz$;
- selecting the kind of analysis: single audio file or two files (pre- and post-surgical).

A notice is added concerning computer time required: for long files (> 5s) and high sampling frequency (>40 kHz) the total time could approach 5min in total. A moving bar shows the residual time during computations.

A number of plots is displayed and saved in printable format, for a visual comparison of results. Specifically, F_0, jitter, RAP, ANNE, spectrogram, formants and PSD are plotted, all in coloured map.


Analisi Pre-Post Operatorio

Voice Analysis Pre-Post Surgical

Load audio file - *.wav

PRE-surgical: [ ] Browse..

POST-surgical: [ ] Browse..

Voice Type

○ Adult Man
○ Adult Woman
○ Baby
○ Singer

Instructions

1- Load audio files *.wav
2- Select "Voice Type"
3- Select "Analysis Mode"
4- Click on "START!!" and wait..:) (about 5 min.)
5- Print diagrams

Analysis Mode

○ single (only PRE)
○ double

START!!

2007 - Realized by:
Canalicchio Riccardo
Cecconi Giulio

for 'Tecnologie Biomediche'
@ Università degli Studi di Firenze

Figure 1 – PrePost analysis tool: User interface

3. EXPERIMENTAL RESULTS

The proposed approach is applied to a set of 15 patients suffering from cysts and polyps that underwent micro-laryngoscopic direct exeresis (MLSD), showing good correlation with GIRBAS and MDVP indexes, thus being suited for integrating such features. Due to lack of space, only one example is reported here. More exhaustive results will be reported elsewhere.

Fig. 2 shows pre-and post surgical F_0 tracking, along with its mean and std values, for a female patient (cysts and nodule MLSD).

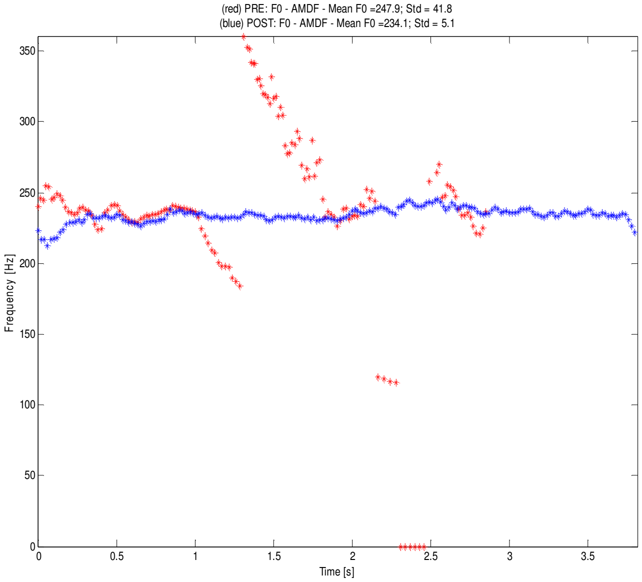


Figure 2 – Pre- and post surgical F_0 tracking, mean and std values.

Good recovering is evident, with regular post-surgical F_0, quantified by reported mean and std values. Notice that, in most cases, and especially with highly degraded voices, F_0 tracking obtained with autocorrelation analysis provides too smoothed plots that make results less reliable. Fig.3 reports Jitter, RAP and noise tracking (with mean and std values), both for pre- and post-surgical signals. From the figure, it is clearly shown that surgery greatly enhances voice quality under all these parameters. The proposed visual comparison also shows non-voiced regions for pre-surgical sig (around 2.2-2.5s), where parameters could not be compute Fig.4 shows pre- and post-surgical spectrograms w formant tracking superimposed (dots), along with mean a std values, for the same patient as before: after surge harmonics and formants are recovered and show a mc regular behaviour and higher energy level (dark red).

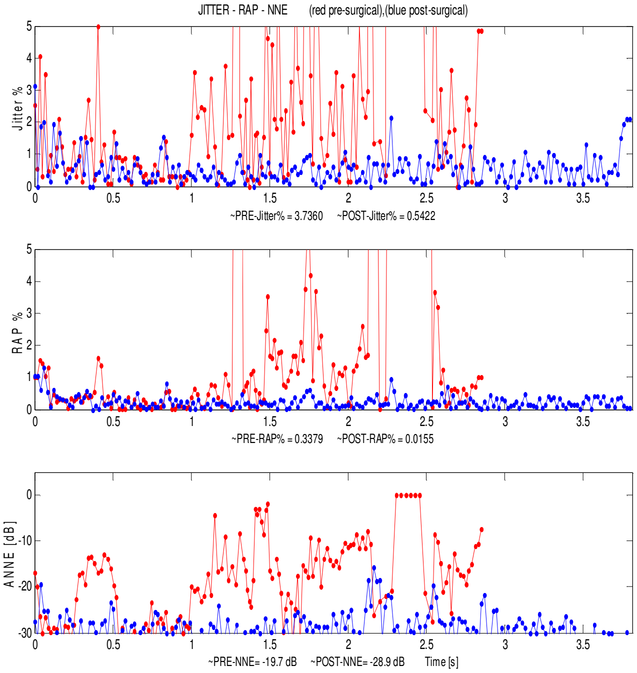


Figure 3 – Jitter, RAP and NNE tracking.

To quantify such results, the PSD plot is displayed in fig. where the low and almost unvoiced pre-surgical frequen content of the signal is evidenced (red line). On the contrary, post-surgical PSD is characterized by a well-structured high-energy harmonic shape in the frequency range typical of voiced emission in adults (≤2500Hz), and a low-energy one above (blue line).

4. DISCUSSION

The method, named here PrePost, has been exploited as far as reliability of results is concerned. A first consideration is relative to the percentage of success (voice quality recovering) obtained with the three parameters Jitter (J), RAP and NNE: the less reliable parameter seems to be RAP (53%), while the other two parameters have a percentage >80%, especially NNE, with 93% of success. This result was compared to "successful" results as described by GIRBAS scores, where a 100% of positive results were obtained. Moreover, a comparison has been made between the results obtained with PrePost and with one of the most used commercial software tools, i.e. MultiDimensional Voice Program (MDVP®), by Kay Elemetrics Corp., where NHR has been considered for comparison with NNE.

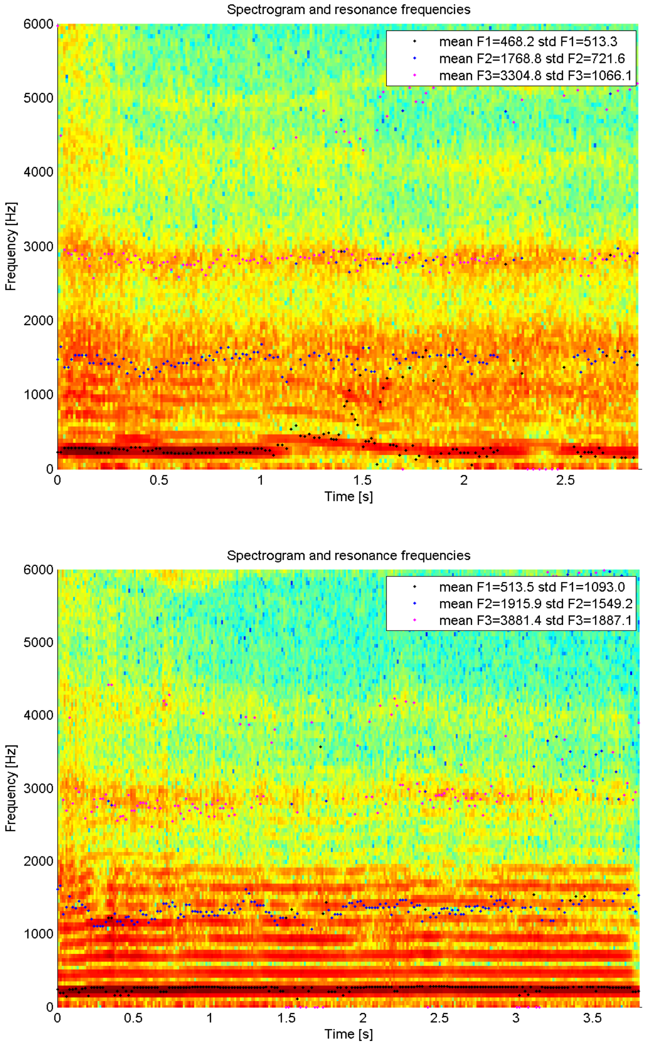


Figure 4 – Pre- (upper) and post-surgical (lower) spectrogram and formants tracking. Mean and std values are displayed.

Table 1 summarizes mean values for both approaches (pre, post), while fig. 6 shows the box-plot of Jitter as measured with MDVP and PrePost, pointing out similar results with Jitter, while RAP reaches lower values with PrePost. Box-plots for RAP and noise indexes (NNE and HNR) are not reported, due to less significance for RAP and different scales for noise.

As for GIRBAS scores, taking into account only "enhanced" indexes, i.e. those corresponding to post-surgical voice quality enhancement, high correlation was found between

Table 1 – Mean values for J, RAP and noise with PrePost and MDVP. First value = pre, second value = post

	Jitter	RAP	NNE , NHR
PrePost	2.8%, 1.8%	1.35%, 0.63%	-22.8%, -26.8
MDVP	2.9%, 1.7%	1.8%,1.02%	0.2%, 0.16%

PrePost (J and NNE) and GIRBAS (G, I and B), with more than 80% of correspondence. Lower correlation was found for R, A and S.

Moreover, both PrePost and MDVP approaches have been compared to GIRBAS scores by means of statistical analysis (t-test for correlated samples). To perform the test, GIRBAS scores >0 have been considered and summed up, both before and after surgery, for each patient.

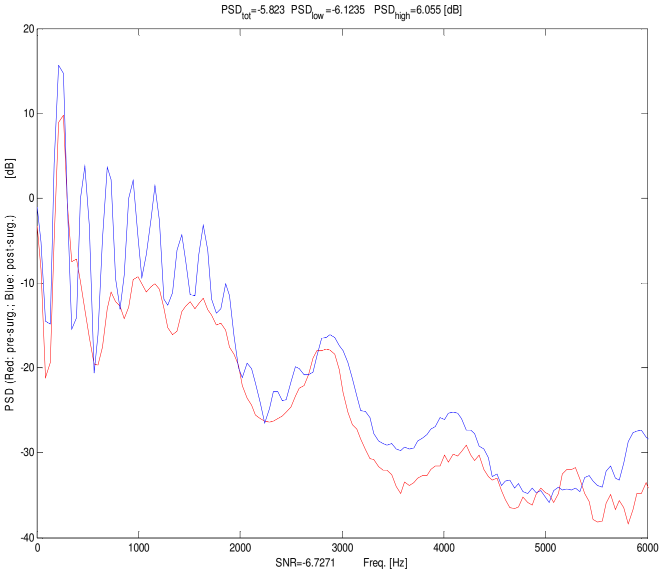


Figure 5 – Pre- and post-surgical PSD plot. Global, low- and high-frequency PSDs are also reported.

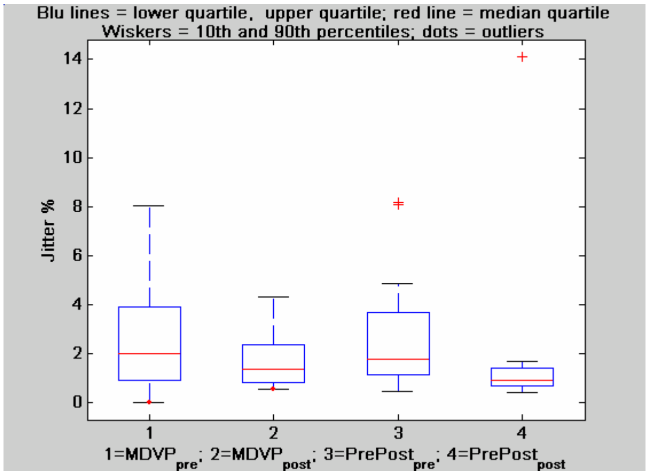


Figure 6 – Jitter box-plot for MDVP and PrePost

Pre-post mean values for GIRBAS were found as 5.07 and 1.07 respectively. Results of the t-test are reported in Table2.

From the Table, it results that all PrePost values have highly significant statistical difference (p<0.01), both pre- and post-surgical. As for MDVP, highly significant statistical difference (p<0.01) was found for pre-surgical values, while significant difference (p<0.05) was found for J and NHR post-surgical values, and no statistically relevant difference for RAP. Summing up, it seems that PrePost performs more reliable analysis than MDVP. This could be due to more robust F_0 estimation with PrePost, and to the different analysis windows used: fixed for MDVP and adaptively tailored to varying F_0 for PrePost.

Table 2 – t-test between GIRBAS, PrePost and MDVP

t-test	Pre-surgical	Post-surgical
$J_{PrePost}$	2,88E-06	2,74E-06
$RAP_{PrePost}$	0,0066	0,006
$NNE_{PrePost}$	1,015E-10	6,67E-14
J_{MDVP}	0,0048	0,043
RAP_{MDVP}	0,0001	0,88
NHR_{MDVP}	1,266E-05	0,016

The tool was developed under Matlab 7.3 and requires few minutes to perform complete pre-post analysis. If properly optimised and implemented under C++ environment, it could perform computations in almost real time.

5. FINAL REMARKS

A new tool for pre-post voice analysis has been developed, based on robust adaptive techniques, capable to deal with highly irregular and hoarse voices. It is provided with a user-friendly interface that requires few basic options to be made by the user. The new tool can be applied both to low- and high-pitched voices (such as newborn infant cry and singing voices) without any manual setting requirement.

Further work will concern finding more strict correlations among objective indexes and perceptive ones, as well as exploiting and adding new possibly helpful indexes and plots. A mobile device implementing the whole procedure is under construction. When properly optimised, the tool could be implemented on a DSP board, as a mobile device useful for clinicians, logopaedicians and patients, also for rehabilitation purposes, after surgery or medical treatment.

REFERENCES

[1] Kay Elemetrics Corp., "Multi Dimensional Voice Program (MDVP). Operations manual", 1994.

[2] S.L. Marple, *Digital spectral analysis with applications*, Englewood Cliffs, NJ, U.S.A.: Prentice Hall, 1987.

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

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

[5] C.Manfredi, G.Peretti, "A new insight into post-surgical objective voice quality evaluation. Application to thyroplastic medialisation", *IEEE Trans. on Biomedical Engineering*, vol.53, n.3, p.442-451, 2006.

[6] J.D. Markel, A.H. Gray, *Linear prediction of speech*, Berlin, DE: Spriger-Verlag, 1982.