TRACKING AND QUANTIFYING POST-SURGICAL VOCAL FOLDS FUNCTIONAL RECOVERING

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

Department of Electronics and Telecommunications, Univ. degli Studi di Firenze, Firenze, ITALY
* Otolaryngology Department, University of Milan, Ospedale Maggiore IRCCS, Via F. Sforza 35, 20122 Milano, Italy

manfredi@det.unifi.it

Abstract: The problem of tracking fundamental frequency F_0, noise level and formants during a voiced emission is considered. For pathological voices, usually such parameters considerably oscillate, due to the effort made by the dysphonic patient in speaking. Hence, new robust procedures are implemented, capable to deal with almost non-stationary signals as those under study. New objective parameters and plots are also proposed, easily understandable and usable by clinicians and logopaedicians, to quantify pre and post-surgical voice quality, as well as functional recovery. The proposed approach is applied to patients suffering from cysts and polyps, that underwent micro-laryngoscopic direct exeresis (MLSD), showing good correlation with MDVP® indexes and GIRBAS scores, and is suited for integrating such features.

Keywords: Pitch, noise, formants, parameter tracking, voice quality, perceptual scale, objective indexes

Introduction

Functional evaluation after vocal fold surgery is commonly based on videolaryngostroboscopy (VLS), for morphological aspects evaluation, GIRBAS scale, relative to perceptive voice analysis, and Multi-Dimensional Voice Program (MDVP$^{®}$), for objective acoustic parameters [ ], [ ], [ ]. While GIRBAS has the drawback to entirely rely on perceptive evaluation of trained professionals, MDVP appears to be unsatisfactory at least for formant tracking, as default options should be manually adjusted to optimal values before analysis. This task is often too much involved for non-expert users, thus giving almost incorrect results. Also, tracking noise, strictly related to hoarseness, rather than merely measuring its mean value, was proven to be of importance for diagnosis and treatment evaluation [1]. Adaptive tracking of the most relevant voice parameters by means of new and robust techniques is proposed here, capable to deal with fundamental frequency, formants and noise also in extreme situations. Few but effective objective voice quality measures are added, based on PSDs and SNR values, and compared to scores obtained from the GIRBAS scale and a subset of MDVP parameters. Fine graphical display allows easily readable results for physicians.

Materials and methods

Perceptual GIRBAS Indexes

Among physicians, the most used perceptual scale is the GIRBAS one, which comprises six qualitative parameters: grade of dysphony (G), instability (I), roughness (R), breathiness (B), asthenicity (A), and strainess (S). For each parameter, a value in the range 0-3 is considered, where 0 corresponds to healthy voice, 1 to light disease, 2 to moderate and 3 to severe [4]. Notice that G is related to the degree of abnormality of the voice, B to the airflow size through the glottis. I and S refer to functional stability of voice and to hyperfunctional phonation, respectively, and A indicates low strength in voice. R reaches intermediate values, and represents the psycho-acoustic impression of the irregular vocal folds vibration, related to jitter and shimmer.

MDVP indexes

Among the huge number of MDVP parameters, the following ones: Jitt, RAP, PPQ, vFo, Shim, APQ, NHR, VTI, SPI, were recognised by the physicians as the most relevant for this application. The first four parameters (Jitt, RAP, PPQ, vF0) are relative to pitch variations within a single period or multiple periods. The 5th and the 6th parameters (Shim, APQ), reflect amplitude variations within a single or multiple periods. The last three parameters (NHR, VTI, SPI), take into account noise presence in the analysed signal.

Specifically, MDVP defines such parameters as follows:

1. Jitt - Jitter Percent /%/ - Relative evaluation of the period-to-period (very short-term) variability of the pitch. Pitch variations are typically associated with hoarse voices.
2. RAP - Relative Average Perturbation /%/ - Relative evaluation of the period-to-period variability of the pitch, with smoothing factor of 3 periods. Hoarse and/or breathy voices may have an increased RAP.
3. PPQ - Pitch Period Perturbation Quotient /%/ - Relative evaluation of the period-to-period variability of the pitch, with a smoothing factor of 5 periods. Hoarse and/or breathy voices may have an increased PPQ.

ISSN: 1727-1983 © 2005 IFMBE

4. vF0 - Coefficient of Fundamental Frequency Variation /%/ - Relative standard deviation of the fundamental frequency. It reflects the variation of Fo (short to long-term) within the analyzed voice sample.
5. Shim - Shimmer Percent /%/ - Relative evaluation of the period-to-period (very short term) variability of the peak-to-peak amplitude. Amplitude irregularity can be associated with the presence of turbulence noise in the voice signal, i.e. with hoarse and breathy voices.
6. APQ - Amplitude Perturbation Quotient /%/ - Relative evaluation of the period-to-period variability of the peak-to-peak amplitude, at smoothing of 11 periods. It can be associated with the inability of the cords to support a periodic vibration with a defined period and with the presence of turbulent noise in the voice signal. Breathy and hoarse voices usually have an increased APQ.
7. NHR - Noise-to-Harmonic Ratio - Average ratio of the inharmonic spectral energy in the frequency range 1500-4500 Hz to the harmonic spectral energy in the frequency range 70-4500 Hz. Increased values of NHR are interpreted as increased spectral noise. NHR measures the noise in the signal (includes contributions of jitter, shimmer and turbulent noise).
8. VTI - Voice Turbulence Index - Average ratio of the spectral inharmonic high-frequency energy in the range 2800-5800 Hz to the spectral harmonic energy in the range 70-4500 Hz. VTI mostly correlates with the turbulence caused by incomplete or loose adduction of the vocal folds, to extract an acoustic correlate to “breathiness”.
9. SPI - Soft Phonation Index - Average ratio of the lower-frequency harmonic energy in the range 70-1600 Hz to the higher-frequency harmonic energy in the range 1600-4500 Hz. Increased value of Soft Phonation Index is generally an indication of loosely or incompletely adducted vocal folds during phonation.

The new tool

Fundamental frequency tracking

F_0 tracking is achieved by means of a two-step procedure, based on well-established results, in order to enhance robustness to noise and signal quasi-stationarity [ ], [ ]. Simple Inverse Filter Tracking is applied first, on windows of fixed length ($3F_{0min}$, F_{0min}=lowest admissible F_0 value for the signal under consideration), followed by Wavelets and AMDF on short time windows of varying length, inversely proportional to previously estimated local F_0 (three pitch periods) [4]. The adaptive window length, tailored to the varying voice characteristics, allows a reliable pitch estimation and tracking. In case of severe disease, that implies fast and abrupt F_0 changes, this procedure was in fact shown to increase robustness in F_0 estimation in many cases, giving enhanced results with respect to standard ones [5], [6].

Noise estimation

Another novel feature is the introduction of an adaptive noise estimation technique 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 during an utterance by means of an adaptive version of the Normalised Noise Energy method [ ]. The method, named ANNE (Adaptive Normalised Noise Energy), is based on the NNE comb filtering approach [ ], optimised in order to deal with data windows of varying length.

The new method relies on choosing an arbitrary, but pre-specified, number of noise lines in spectral ‘dip’ regions (i.e. between two successive harmonics), thus resulting in accurate noise estimation, as it allows avoiding empty ‘dip’ regions along the frequency axis. The method has already been successfully applied to pathological voices, coming from vocal fold operated patients [ ].

The ANNE is thus suited to give the physician an objective tracking of voice hoarseness due to disease. Specifically, large negative ANNE values (in dB) correspond to good voice quality, while values close to zero reflect the presence of strong noise. Besides sustained vowels only, noise tracking could be of help also as far as evaluation of the effort made by the patient in pronouncing complete words is concerned. In fact, better evidence is obtained with complete words, as finer details relative to vowel and diphthongs transition can be highlighted.

Formants tracking

To recover formants' position, AutoRegressive Power Spectral Density is evaluated, thanks to its high-resolution properties. A robust parametric formant estimation technique is proposed, obtained by peak picking in the Power Spectral Density (PSD), evaluated on the same adaptive time windows as obtained in the previous step, and based on AR models of order equal to the signal sampling frequency F_s (in kHz). Notice that choice of a model order near or equal to F_s, as suggested in the literature [4], prevents from spectral smoothing and consequently loss of spectral peaks. This approach has already been proved effective in many applications, with enhanced results as far as resolution is concerned [8], [9], [10].

Quality indexes

Finally, quantitative PSD and SNR measures are defined, to objectively evaluate functional recovery, along with easily readable plots.

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 (male and female), as well as on experimental results obtained from threshold tuning in a dataset of voiced and unvoiced sounds. However, it can be changed to different values, if required.

Specifically, the following indexes are proposed (in dB), where the subscripts “pre” and “post” refer to the pre and post surgical signal, respectively:

$$\mathrm{PSD} = 10\log_{10}\frac{\mathrm{PSD}_{\mathrm{pre}}}{\mathrm{PSD}_{\mathrm{post}}} \qquad (1)$$

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

$$\mathrm{PSD}_{\mathrm{low}} = 10\log_{10}\frac{\mathrm{PSD}_{\mathrm{pre}}(f \le 4\mathrm{kHz})}{\mathrm{PSD}_{\mathrm{post}}(f \le 4\mathrm{kHz})} \qquad (2)$$

that measures the ratio of the PSDs evaluated on the “harmonic range”;

$$\mathrm{PSD}_{\mathrm{high}} = 10\log_{10}\frac{\mathrm{PSD}_{\mathrm{pre}}(f \ge 4\mathrm{kHz})}{\mathrm{PSD}_{\mathrm{post}}(f \ge 4\mathrm{kHz})} \qquad (3)$$

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

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_{n=1}^{M} y_{\mathrm{pre}}{}^{2}(n)}{\sum_{n=1}^{M}(y_{\mathrm{pre}}(n) - y_{\mathrm{post}}(n))^{2}} \qquad \mathbf{(4)}$$

SNR is thus the ratio between the noisy signal energy and that of removed noise. Negative SNR values correspond to voice quality enhancement.

Results

19 patients (6 males, 13 females, age 29-74, mean 52) underwent microlaryngoscopy, to remove vocal fold lesions (cysts and polyps). Pre- and post-surgical sustained vowel /a/ was analysed for all of them with GIRBAS perceptive scores. Moreover, both /a/ and the Italian word /aiuole/ were analysed for 16 patients, by means of MDVP and the new tool. 6 healthy subjects (2 males, 4 females) were analysed as reference set, all of them pronouncing both /a/ and /aiuole/, with almost uniform low values for quality indexes.
New global indexes are also defined, that allow comparing the new indexes to perceptual GIRBAS scores and MDVP ones [2]. Specifically, a Score Ratio, SR, is defined over GIRBAS scores, according to the following criterion: For each subject and each quality index, pre and post-surgical scores are subtracted, all the results added, and the final sum divided by the total number of scores, excluding those that were equal to zero both before and after surgery. This in fact means that the voice was evaluated as healthy under that respect, both before and after surgery, which thus had no enhancing effect on it. With MDVP indexes, the ratio between pre and post-surgical values was evaluated, and a global score MR (Mean Ratio) for each patient was obtained, defined as the mean value among the nine MDVP ratios. Notice that, in 2 cases, pre-surgical parameters were not available, as the signal amplitude was classified as “too low” with MDVP. This fact, which is not so uncommon with MDVP, prevents from pre- and post-surgical effectiveness comparison.
As quality indexes (2)-(5) are already pre/post-surgical ratios, a Total Score (TS) is obtained, adding SNR, $\mathrm{PSD}_{\mathrm{tot}}$, $\mathrm{PSD}_{\mathrm{low}}$, with reversed sign to $\mathrm{PSD}_{\mathrm{high}}$. This amounts to introducing penalty terms whenever some index does not reflect an increase in voice quality after surgery, according to eqns.(1)-(4) and their relative meaning.
Finally, for each patient, a number of plots is available, that allow a visual comparison of pre and post-surgical results. Specifically, fundamental frequency, ANNE, spectrogram, formants tracking and PSD plots, all in coloured map, can be obtained, to help the physican in voice quality evaluation.
By comparing pre and post surgical F0, ANNE and spectrograms, one can have a first qualitative view of the residual noise present in the voice signal after surgery, as well as of harmonics intensity and stability. Formant tracking and PSD plots complete the picture, with specific details about possible formant recovering along with their intensity.
Figure 1 reports some results obtained with the new tool, concerning one female (cyst), along with new objective indexes: $\mathrm{PSD}_{\mathrm{low}}$ and SNR<0, $\mathrm{PSD}_{\mathrm{high}}$>0 and low ANNE denote good functional recovering.

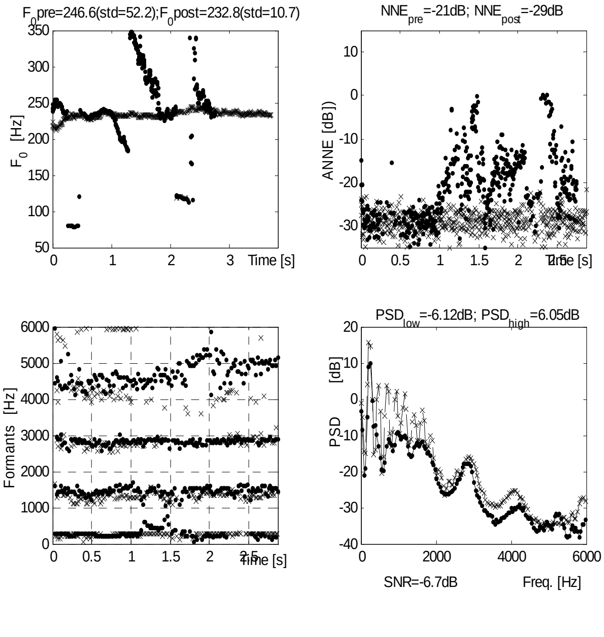

Figure 1: F_0, ANNE, formants and PSD pre (•) and post-surgical (x).

This is also confirmed by the spectrogram and formant tracking, before and after surgery (figs. 2 and 3, respectively): harmonics are recovered in the low-frequency spectral region, the first formant is almost stabilised and lower residual noise appears in the high-frequency spectral region.

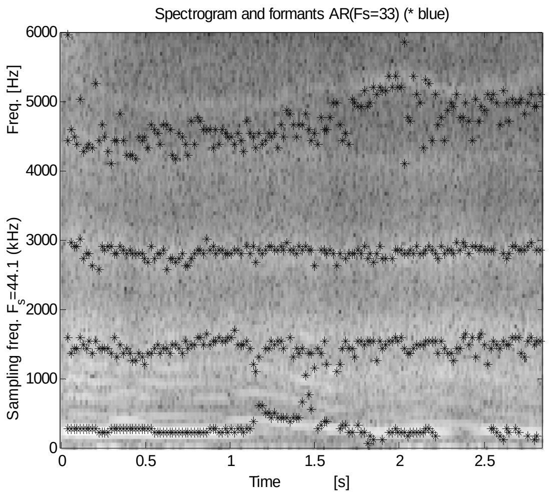

Figure 2: Pre-surgical spectrogram and formants

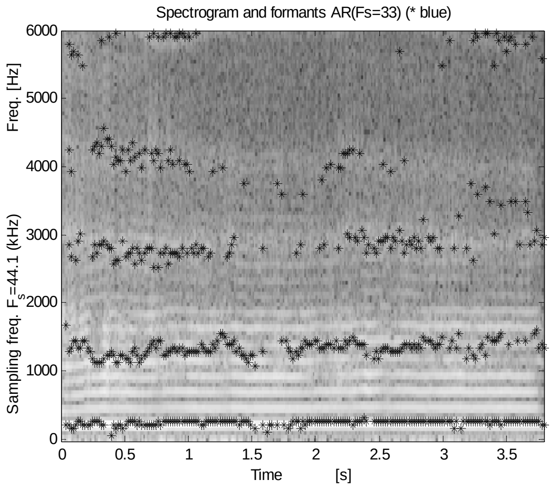

Figure 3: Post-surgical spectrogram and formants.

Figure 4 shows a bar plot of the global indexes MR, TS and SR for all 19 patients, with reference to the sustained /a/ vowel. Due to different ranges, they were normalised for comparison. The higher the value of the index, the better the voice quality recovering. Although an almost common trend could be observed for the three indexes, especially for patients n. 4, 6, 8, 11 (high voice quality recovering), and 1, 2, 7 (low voice quality recovering) , until now no reliable correlation was found among such data. This point needs further study, being of relevance for physicians. Perhaps a reduced set of MDVP indexes, as well as more meaningful new ones, could provide better results.

Conclusions

Adaptive and automatic robust tracking of F_0, noise level and formants during a voiced emission is proposed, to assess voice quality recovering after vocal fold surgery. An adaptive comb filtering approach is proposed, which allows tracking the signal noise component on subsequent short time windows of varying length. Robust and high-resolution formant estimation is implemented, based on parametric PSD evaluation. F_0, noise, formants, spectrogram and PSD plots allow easy interpretation of results.

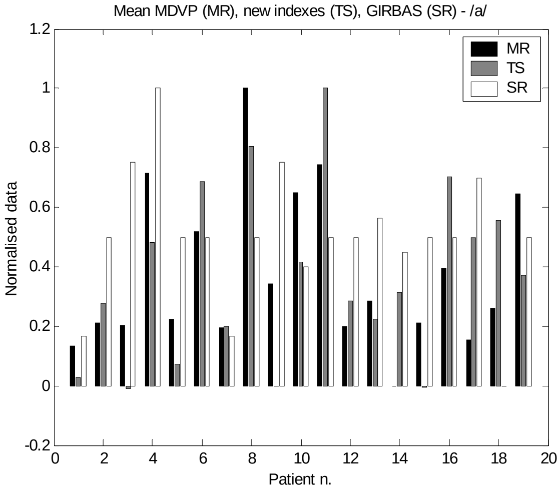

Figure 4: Normalised global indexes for MDVP, the new tool and GIRBAS

Global indexes are defined, to correlate the new ones with a subset of MDVP indexes and the GIRBAS scores. Simultaneous inspection of plots and indexes relative to each patient allows setting up a complete clinical picture, giving the physician easier and deeper understanding of surgical effectiveness. If properly optimised, and with the aid of a user-friendly interface, the new tool could be implemented on a DSP board, as a mobile device useful for clinicians, logopedicians and patients, also for rehabilitation purposes, after surgery or medical treatment.

Further work will concern finding more strict correlations among new indexes and GIRBAS or MDVP ones, as well as exploiting and adding new possibly helpful indexes and plots.

References

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

[2] MANFREDI C., PERETTI G., “A new insight into post-surgical objective voice quality evaluation. Application to thyroplastic medialisation”, *IEEE Trans. Biomed. Eng.*, (in print) 2005.

[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] DELLER J.R., PROAKIS J. G., HANSEN J.H.L.”Discrete-time processing of speech signals”, Macmillan Pub. Co., N.Y., 1993.

[6] MANFREDI C., D'ANIELLO M., BRUSCAGLIONI P., ISMAELLI A., "A Comparative Analysis of Fundamental Frequency Estimation Methods with Application to Pathological Voices", *Med.Eng. Phys.*, vol.22, n.2, p.135-147, 2000.

[7] MANFREDI C., PERETTI G., “Robust Techniques for Pre- and Post-Surgical Voice Analysis”, Proc. Eurospeech Conf., vol.3, p.2365-2368, Sept. 1-4, 2003, Geneve, Switzerland.

[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] FORT A., ISMAELLI A., MANFREDI C., BRUSCAGLIONI P., "Parametric and non Parametric Estimation of Speech Formants: Application to Infant Cry", *Med. Eng. Phys.*, vol.18, n.8, p.677-691, 1996.

[10] 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.

[11] MANFREDI C., PERETTI G., BOCCHI L., BRUSCAGLIONI P., "Tracking disphonic voice parameters: application to unilateral vocal cord paralysis", Proc. Irish Signals and System Conf., p.142-147, Limerick, Ireland June 30-July 2, 2003.