INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

Thyroplastic Medialisation in Unilateral Vocal Fold Paralysis: Assessing Voice Quality Recovering

C. Manfredi, E. Dori, F. Iadanza, G. Peretti, L. Magnoni**

Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3, 50139 Firenze, Italy - manfredi@det.unifi.it
*Otolaryngology Clinic, Università degli Studi di Brescia, Spedali Civili di Brescia, Piazza Spedali Civili 1, 25123 Brescia, Italy - g.peretti@tin.it

10.21437/Interspeech.2004-89

Abstract

Medialization thyroplasty and endoscopic intracordal infusion of fat or heterologous materials are the treatments of choice for glottic incompetence, of both neurological and cicatricial origin. Functional evaluation after thyroplastic medialisation is often based on several approaches, in order to assess the effectiveness of the adopted technique. The most common analysis methods are: videolaryngostroboscopy (VLS), for morphological aspects evaluation, GRBAS scale and VHI (Voice Handicap Index), relative to perceptive and subjective voice analysis, and MDVP®, that provides objective acoustic parameters.
First results are presented here, obtained both with these approaches and a new voice analysis tool, based on robust estimators for tracking fundamental frequency F_0, noise and formants. New indexes are also proposed, to easily quantify voice quality recovering. The proposed approach was successfully applied to patients that underwent thyroplastic medialisation, and is suited for integrating the MDVP features.

1. Introduction

This paper deals with the problem of quantifying voice quality, before and after medialisation thyroplasty, for patients affected by glottic incompetence. Besides perceptive and subjective scales, such as GRBAS [1], [2] and VHI, the Multi-Dimensional Voice Program (MDVP) software tool (Kay Elemetrics) is well recognised as a standard for reference, being the most widely used tool in the biomedical field devoted to voice analysis. However, in some cases, tracking voice parameters, such as noise (roughness, hoarseness), rather than merely measuring their mean values, was proven to be of importance for diagnosis and treatment evaluation [3]. In this respect, new and robust objective measures, as well as tracking of the most relevant voice parameters are proposed here, capable to recover the fundamental frequency, track formants and the degree of hoarseness, to assess the patient's functional recovery. Fine graphical display enables easy readable results. The proposed approach is thus suited for integrating the MDVP features. When properly optimised, it could be implemented on a DSP board, as a mobile device useful both for clinicians and patients, also for rehabilitation purposes, after surgery or medical treatment.
The method is applied to sustained /a/ vowels, recorded from patients suffering from unilateral vocal cord paralysis, but can be successfully applied to complete words as well. Pre- and post-surgical parameters are evaluated, that allow the physician objectively quantifying the effectiveness of the Montgomery thyroplasty implant.

2. Materials and methods

In the proposed approach, F_0 is estimated by means of a two step procedure: 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]. In case of severe disease, that implies fast and abrupt F_0 changes, this procedure was shown to increase robustness in F_0 estimation in many cases, giving enhanced results with respect to standard ones [5], [6]. Another novel feature is the introduction of an adaptive noise estimation technique that allows tracking varying noise level during phonation. The method, named ANNE (Adaptive Normalised Noise Energy), relies on a comb filtering approach [7], optimised in order to deal with the varying window length previously obtained. Specifically, large negative ANNE values correspond to good voice quality, while values close to zero reflect the presence of noise.
Noise tracking 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. In fact, better evidence is obtained with complete words, as finer details relative to vowel transition can be highlighted [3].
Moreover, 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 before, 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].
Finally, new objective quality measures (PSD ratios, SNR) are defined, that easily allow assessing enhancement of voice and comparing pre and post-surgical 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 [9]. 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:

$$PSD = 10\log_{10}\frac{PSD_{pre}}{PSD_{post}} \quad (1)$$

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

$$PSD_{low} = 10\log_{10}\frac{PSD_{pre}(f \le 4kHz)}{PSD_{post}(f \le 4kHz)} \quad (2)$$

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

$$PSD_{high} = 10\log_{10}\frac{PSD_{pre}(f \ge 4kHz)}{PSD_{post}(f \ge 4kHz)} \quad (3)$$

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

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

where: $y_{pre}(n)$ = pre-surgical signal sample at time n, $y_{post}(n)$ = post-surgical 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.
Pre and post surgical plot overlap for PSD and fine graphical display for F_0, formants and ANNE, enables easy readable results. The proposed approach is thus suited for integrating the GRBAS and MDVP features.
The software is developed under MATLAB R12® (The Mathworks Inc.). When properly optimised, it could be implemented on a DSP board, as a mobile device useful for both clinicians and patients, also for rehabilitation purposes, after surgery or medical treatment. A prototype is under study.

3. Experimental results

At the Otolaryngology Department of the University of Brescia, a number of patients with glottic incompetence were submitted to medialization thyroplasties, with Montgomery prostheses [11]. All patients underwent pre- and post-surgical diagnostic work-up, including: VLS, fundamental frequency, jitter, shimmer, and noise-to-harmonics ratio (NHR) by MDVP. Patients were asked to pronounce the Italian vowel /a/, sustained for at least 3 sec. Moreover, subjective analysis by the GRBAS scale, and self-evaluation of the voice by means of the Voice Handicap Index (VHI), was performed [12]. Finally, the proposed new tools were applied, providing new and useful details that successfully refine and integrate previous results. Tables 1-3 report the results obtained for 22 patients (M=male, F=female).
Table 1 concerns GRBAS scale. From reported scores, the perceptive quality of voice results enhanced in all cases.
Table 2 summarises the most relevant parameters obtained with MDVP (J=jitter, S=shimmer, NHR=Noise-to-Harmonics-ratio), before (pre) and after (post) medialisation. Lower post surgical jitter, shimmer and NHR values confirm voice quality recovering in most cases. Notice that, in some cases, pre surgical parameters are not available for comparison, as they were classified as "non voiced" with MDVP.

Table 1: *GRBAS scale for pre and post surgical signals*

	$G_{pre/post}$	$R_{pre/post}$	$B_{pre/post}$	$A_{pre/post}$	$S_{pre/post}$
1 - M	2/1	2/1	1/0	1/0	0/0
2 - F	3/2	1/1	2/1	2/2	3/2
3 - F	3/1	1/1	3/0	2/0	0/0
4 - F	2/1	1/1	2/1	1/1	2/1
5 - M	3/0	0/0	3/0	3/0	0/0
6 - F	2/0	2/0	1/0	1/0	0/0
7 - F	3/1	1/1	3/1	2/0	0/0
8 - F	3/0	3/0	2/0	1/0	2/1
9 - M	1/0	0/0	0/0	0/0	1/0
10 - F	3/0	1/0	3/0	3/0	1/0
11 - M	3/0	3/0	3/1	1/0	2/0
12 - F	3/0	0/0	3/0	3/0	0/0
13 - F	3/2	1/2	2/1	1/0	0/1
14 - M	3/2	2/2	2/2	1/0	3/2
15 - F	2/0	2/0	1/0	0/0	2/0
16 - M	3/0	2/0	3/0	2/1	1/0
17 - M	3/2	1/1	2/2	1/1	3/2
18 - F	2/0	1/0	1/0	0/0	2/0
19 - M	2/0	2/0	1/0	0/0	1/0
20 - M	2/0	1/0	2/0	2/0	0/0
21 - F	3/3	2/1	3/0	3/2	2/3
22 - F	3/0	2/0	3/0	3/0	2/1

Table 2: *MDVP pre and post surgical parameters*

	$J_{pre/post}$	$S_{pre/post}$	$NHR_{pre/post}$
1 - M	3.27/6.33	8.8/22.5	0.25/0.39
2 - F	1.9/1.88	3.04/4.6	0.1/0.14
3 - F	--/1.55	--/8.53	--/0.11
4 - F	4.04/3.05	6.99/7.13	0.16/0.14
5 - M	--/2.8	--/2.06	--/0.1
6 - F	6.47/2.75	10.47/4.8	0.16/0.16
7 - F	22.12/3.4	24.9/7.7	1.3/0.17
8 - F	3.7/0.96	7.2/5.3	0.15/0.11
9 - M	0.64/0.42	3.76/2.4	0.14/0.13
10 - F	7.2/0.45	10.6/4.03	0.28/0.13
11 - M	6.67/3.67	15.8/9.3	0.4/0.2
12 - F	--/0.47	--/2.75	--/0.13
13 - F	14.8/3.8	14.2/6.4	1.7/0.15
14 - M	3.7/3.02	15.7/14.1	0.43/0.26
15 - F	3.5/2.1	11.3/7.2	0.23/0.12
16 - M	5.3/3.7	13.5/8.3	0.3/0.27
17 - M	5.7/3.3	17/12	0.57/0.27
18 - F	4.4/1.2	7.7/4.8	0.33/0.11
19 - M	--/0.63	--/3.86	--/0.14
20 - M	2.65/0.3	8.8/1.8	0.23/0.12
21 - F	6.7/5.6	11.3/10.4	0.26/0.22
22 - F	6.8/2.13	8.6/4.9	0.25/0.15

Finally, Table 3 reports the new indexes (1)-(4). Notice that SNR <0 (voice quality enhancement) in all cases, while PSD_{tot} and PSD_{low} <0 in all but two cases (due to a slight decrease of harmonics power after surgery). As for PSD_{high} significant hoarseness reduction is pointed out by high positive values, while values around or below zero concern patients with voice recovering, but with residual noise. Thanks to robust estimators, all signals could be analysed. Comparison of these results to the GRBAS and MDVP ones, shows good correlation in most cases. A deeper insight is under study.

Table 3: *Pre and post surgical voice quality indexes*

	SNR	PSD_{tot}	PSD_{low}	PSD_{high}
1 - M	-2.48	1.10	1.04	7.21
2 - F	-5.05	-3.43	-3.45	-1.02
3 - F	-7.20	-6.33	-7.56	13.88
4 - F	-3.90	-1.47	-1.47	-1.95
5 - M	-10.96	-11.3	-12.9	13.01
6 - F	-4.04	-1.69	-1.69	0.60
7 - F	-8.79	-8.24	-8.50	-1.80
8 - F	-3.37	-0.83	-0.84	1.24
9 - M	-3.84	-1.47	-1.46	-5.66
10 - F	-2.79	-0.08	-0.35	7.58
11 - M	-2.31	1.54	1.02	8.52
12 - F	-8.73	-8.22	-8.40	7.20
13 - F	-3.11	-0.04	0.001	-2.26
14 - M	-3.87	-1.62	-1.56	-4.80
15 - F	-3.94	-1.66	-1.94	3.01
16 - M	-4.58	-2.62	-2.79	3.22
17 - M	-2.57	-2.67	-2.71	-0.28
18 - F	-3.19	-0.38	-0.42	4.58
19 - M	-6.69	-4.75	-6.14	19.71
20 - M	-6.91	-5.83	-5.96	7.27
21 - F	-3.68	-0.96	-1.02	1.41
22 - F	-3.52	-0.97	-1.10	9.66

With the new tool, simultaneous inspection of the plots relative to each patient allows deeper understanding of the surgical effectiveness. As an example, Figures 1-4 show results concerning patient n.3.

Specifically, Fig.1 plots F_0, estimated by means of the two step procedure cited in sect.2: black crosses correspond to the first step, while red ones to the refinement obtained in step 2, on varying time windows (oscillating in the range 20ms-80ms before surgery, and almost constant, around 15.4ms, after) (blue stars). F_0 mean is evaluated around 215 Hz before surgery, with huge std, due to almost unvoiced intervals (0s-0.8s and 2s-2.6s). This value reduces to 28.8 if the analysis is restricted to the interval (0.8s-2s). Notice that MDVP classifies the whole signal as unvoiced. After surgery, F_0 stabilises around 278Hz. The lowest plots highlight possible noise changes during utterance: notice high and oscillating ANNE values, corresponding to F_0 instability and large effort, before surgery, but an almost constant and low level (-28.5dB) after it, assessing good voice quality recovering. Fig. 2 plots the signal spectrogram, on a self-optimised time-frequency scale, automatically linked to the actual sampling frequency and time window length. This possibly implies different bandwidth in the pre and post-surgical plot, with possible slight difficulty in comparison, recoverable in formats and PSD plots. Notice high noise energy before surgery also at high frequencies, typical of hoarseness. Figure 3 depicts formant trajectory before (blue,+) and after (red, *) surgery. The two lowest formants are recovered after surgery, but the third one has almost disappeared, while the fourth and fifth are settled on lower frequencies. Changes in F_0 and timbre are in fact rather common after prosthesis implantation. Figure 4 reports the PSD plot before (blue) and after (red) surgery, along with the new indexes. The energy of each spectral component is clearly shown.

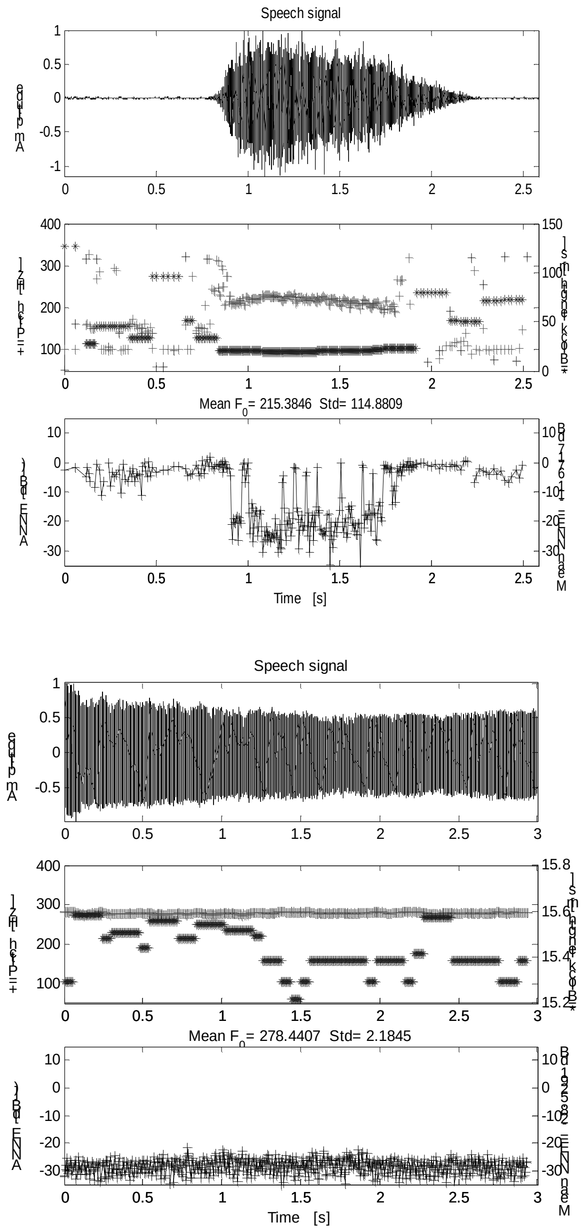

Figure 1*: Pre (top) and post (bottom) surgical F_0 and ANNE tracking with varying window length for analysis.*

By comparison among Figs. 2, 3 and 4, one can easily evaluate changes in noise level, formant position and intensity that has occurred after prostheses implantation. These new features, along with perceptive GRBAS scores and MDVP indexes, allow setting up a complete clinical picture.

4. Final remarks

Montgomery prosthesis is particularly well suited to glottic incompetence of neurological origin. A data set of 22 patients with unilateral vocal fold paralysis is considered, that underwent thyroplastic medialisation. Both perceptual and objective indexes, some of them new, were evaluated, in order to set up a complete clinical picture. Specifically, robust tracking for F_0, noise and formants is proposed, along with easily readable plots and quality indexes. A deeper insight into voice recovering and its possible changes after surgery can thus be obtained and evaluated by the clinician. Results confirm the effectiveness of medialization thyroplasty in optimizing glottic closure and phonation. Further work will concern finding correlations among new indexes and GRBAS or MDVP ones, as well as exploiting and adding new possibly helpful indexes and plots.

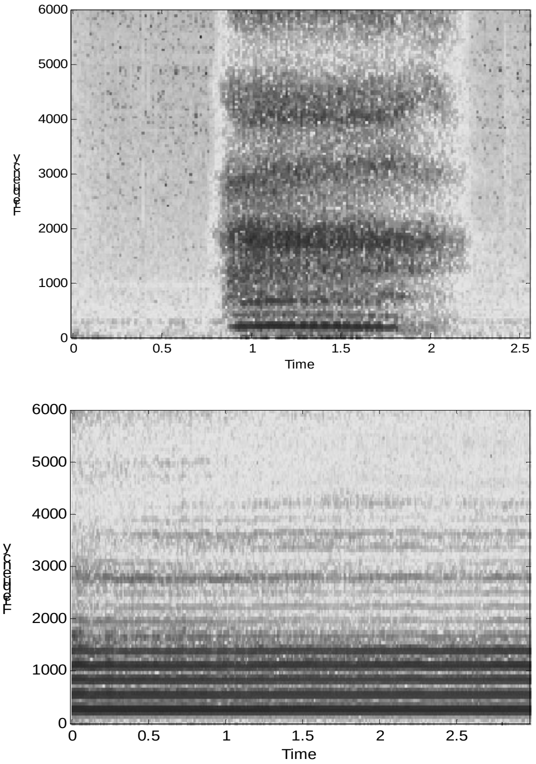

Figure 2: *Pre(top) and post (bottom) surgical signal spectrogram*

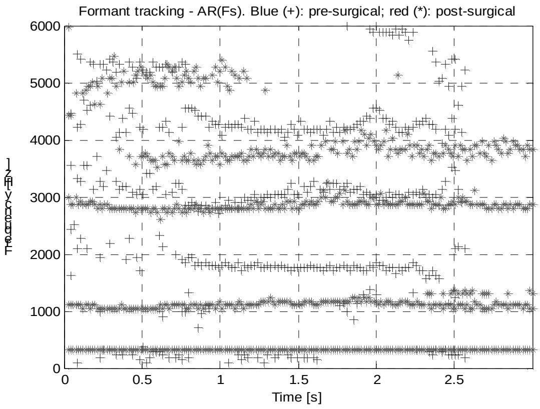

Figure 3: *Pre (+) and post(*) surgical formant tracking*

5. References

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

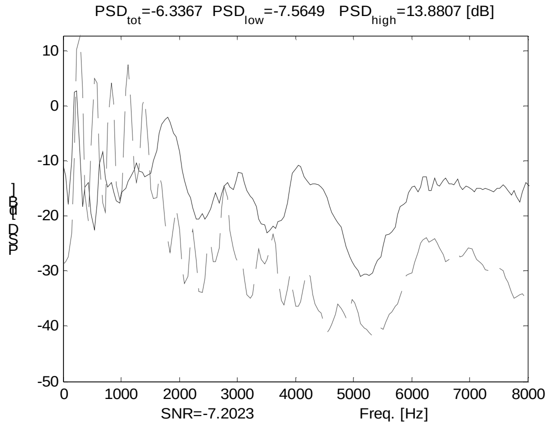

Figure 4: *Pre(blue, solid) and post(red, dashed) surgical PSD and new indexes.*

[2] 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.

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

[4] Deller J.R., Proakis J. G., Hansen J.H.L."Discrete-time processing of speech signals", Macmillan Pub. Co., N.Y., 1993.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] C. Manfredi, G. Peretti, L. Bocchi, P. Bruscaglioni, "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.

[11] Montgomery W.W., Blaugrund S.M., Varvares M.A., "Thyroplasty: a new approach", Ann.Otol. Rhinol. Laryngol., 102, 571, 1993.

[12] Peretti G., Provenzano L., Piazza C., Giudice M., Antonelli A.R., **"**Functional Results After Type I Thyroplasty with Montgomery Prosthesis", Acta Otorhinolaryngol Ital 21(3), p.156-162, 2001.