A MULTI-PURPOSE USER-FRIENDLY VOICE ANALYSIS TOOL: APPLICATION TO LIPOFILLING TREATMENT

Claudia Manfredi, Giovanna Cantarella[1]

Department of Electronics and Telecommunications, Università degli Studi di Firenze, Firenze, Italy
[1] Department of Otolaryngology, Università degli Studi di Milano, Ospedale Maggiore IRCCS, Milano, Italy

Abstract - **A multi-purpose software tool (BioVoice) is presented, capable of performing automatic analysis of a large range of voice signal, no manual setting being required to the user. This makes the tool feasible for application by non-expert users in several fields, ranging from high-pitched new-born cries, to adult healthy singing vocalizations and to irregular, pathological voice signals. Main voice characteristics (fundamental frequency and formants) are evaluated and tracked by means of robust analysis techniques that can handle the above mentioned wide range of signals, as internal settings for optimal frame length, frequency range of analysis and plots are automatically adjusted.**
Specific parameters are evaluated according to the kind of signal under study, and displayed with suitable plots and tables.
In this paper, the method is applied to patient affected by laryngeal hemiplegia that underwent lipofilling treatment to recover phonatory capabilities.
Keywords: **multi-purpose voice analysis tool, robust parameter estimation, laryngeal hemiplegia.**

I. INTRODUCTION

Voice analysis is of great relevance in several fields, ranging from newborn infant cry to singing voice and to hoarse adult voices. Hence, paediatricians, surgeons, but also singing teachers, psychologists and logopedicians are involved with this field of research. Nowadays, several analysis techniques and reference values have been proposed in literature and are in use. A huge number of indexes is available, some of which implemented in free or commercially available software tools [1], [2]. However, users often 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 high-pitched and/or highly degraded voices.
The BioVoice tool proposed here aims at providing few objective parameters and plots, easily understandable and manageable by a wide range of users. The proposed software tool performs single or comparative analysis of main voice characteristics (fundamental frequency and formants) by means of robust analysis tools, specifically devoted to deal with a wide range of pitch values, and possibly highly degraded signals. At present, three main categories are considered with BioVoice: newborn infant cry, singing voice and adult hoarse voice.

II. METHOD

Basic voice characteristics (fundamental frequency (F_0) and formants) are evaluated and tracked by means of robust analysis techniques that can handle the mentioned wide range of signals. To this aim, automatic adjustment of internal settings for optimal frame length, frequency range of analysis and plots are implemented.
First, 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 (V/UV) separation algorithm is implemented, to avoid F_0 estimation on signal frames that have no harmonic content and could give misleading results.
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 [3], [4].
Robust and high-resolution formant (resonance frequencies) 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 relationship: $p=2LF_s/c$, where: F_s=sampling frequency, L=vocal tract length (linked to patient's age and sex), and c=sound speed [4]. Colour-coded spectrograms are also provided, with the tracking of formants F_i superimposed, whose number and frequency range depends on the signal under study. Mean values and std are also displayed.
Other ad hoc parameters are added to these basic features, for each category. They are summarised here.
Newborn infant cry - Newborn infant cry is characterised by high fundamental frequency F_0 (>300Hz), possibly with

Models and analysis of vocal emissions for biomedical applications: 5th international workshop: December 13-15, 2007: Firenze, Italy, ed. by C. Manfredi,
ISBN 978 88-8453-673-3 (print) ISBN 978-88-8453-674-7 (online)

abrupt changes and voiced/unvoiced features of very short duration within a single utterance. The frequency range is thus set up to 10 kHz. F_0, V/UV frames, spectrogram with the first 3 resonance frequencies superimposed, are plotted, all in coloured map. Some tables summarise mean, std, max, min values for F_0 and F_1-F_3, as well as cry length and the corresponding maximum energy. These parameters are in fact considered among the most meaningful in newborn cry analysis (see [5] and references therein).

Singing voice - Singing voice results from complex, voluntary movements of the larynx and of vocal tract articulators, and is characterized by possibly high-pitched, rapidly time-varying signals. As we deal with adult singers, the frequency range is set up to 6 kHz. F_0, vibrato rate, vibrato extent, vocal intonation, spectrogram with the first 5 formants and PSD are plotted, along with formants maxima co-ordinates. These parameters are of importance for singers, being strictly related to correct vocal emission and hence to singer's performance (see [6] and references therein).

Adult hoarse voices - 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. 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. Within BioVoice, noise variations are tracked by means of an adaptive version of the Normalised Noise Energy method, named ANNE (Adaptive Normalised Noise Energy) [7], [8]. 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. Spectrograms and PSD plots complete the set of pictures, allowing visual inspection of possible harmonic energy recovering. On the PSD plot, PSD_{tot}, PSD_{low}, PSD_{high} are reported, 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.

III. THE INTERFACE

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 data (.wav files);
– choosing the voice type, ranging from high-pitched newborn and possibly 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 (for comparison).

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.

Plots and tables are displayed and saved in printable format, for a visual comparison of results, all in coloured map.

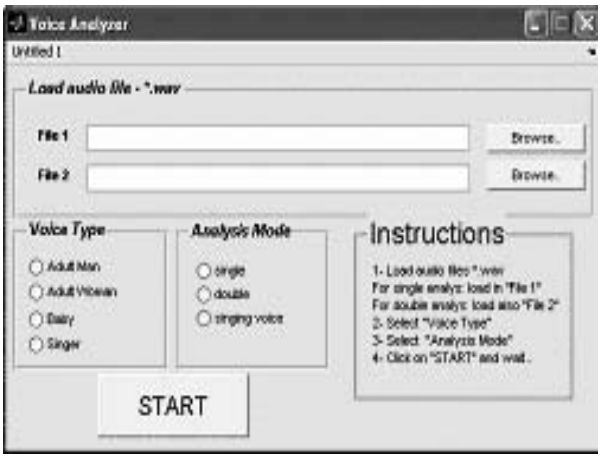

Figure 1 – BioVoice analysis tool: user interface

IV. EXPERIMENTAL RESULTS

BioVoice is applied here to nine patients (aged 18–74 years, mean 48) with breathy dysphonia, secondary to laryngeal hemiplegia or anatomical defects that underwent vocal fold lipoinjection. Lipostructure is a valuable technique for voice rehabilitation in glottis incompetence. Patients underwent pre- and post-treatment videolaryngostroboscopy, maximum phonation time (MPT) measurements, GRBAS perceptual evaluations, and Voice Handicap Index (VHI) self-assessments. Voice quality improved soon after surgery and remained stable over 3–26 months, as confirmed by GRBAS, MPT and VHI [9].

To show BioVoice features, one example is presented here, concerning a female patient. Before lipofilling, GRBAS scores were found as [3 3 2 2 0], denoting high level of dysphony, with a full recovering after the treatment (all GRBAS scores =0). Due to printing requests, figures are reported in a grey scale: (pre=light grey, post=black).

Fig. 2 shows pre-and post treatment F_0 tracking, along with its mean and std values. As pre- and post-treatment (PRT-POT) audio signals are usually of different length, the tool adjusts plots on the longer one. In this case, the PRT signal has a length of about 1.6s, while the POT one last about 3.6s.

Notice the long unvoiced period (above 2s) as found by the program for POT.

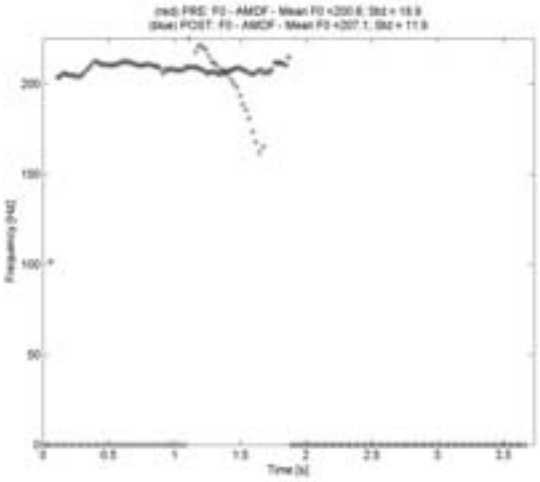

Figure 2 – Pre- and post surgical F_0 tracking, mean and values.

Good recovering is shown, with stable POT F_0, ar 207Hz, as compared to highly varying PRT that cou evaluated for less that 1s. Fig.3 reports Jitter, RAP and tracking (with mean and std values), both for PRT and POT signals. From the figure, it is clearly shown that lipofilling greatly enhances voice quality under all these parameters. Again, notice non-voiced regions, where parameters could not be computed.

Fig.4 shows PRT and POT spectrograms with formant tracking superimposed (black dots), along with mean and std values: after treatment, harmonics and formants are almost recovered and show a more regular behaviour and higher energy level (dark black) with respect to PRT ones.

To quantify such results, the PSD plot is displayed in fig. 5, where the almost unvoiced and noisy PRT frequency content of the signal is evidenced (light grey line). On the contrary, POT PSD is characterized by a rather well-structured high-energy harmonic shape in the frequency range typical of voiced emission in adults (≤2500Hz), and a low-energy one above this range, mainly related to noise (black line).

Good recovering was found for almost all cases, and results were found in agreement with GRBAS scores. Due to the small number of available cases, statistical tests to assess reliability were not applied.

V. FINAL REMARKS

A new tool for voice analysis has been developed, based on robust adaptive techniques, capable to deal with a wide range of voice sounds. It is provided with a user-friendly interface that requires few basic options to be made by the user. The method has already been successfully applied to pathological voices, to compare pre- and post-surgical voice quality in case of tyroplastic medialisation and cyst/nodule exeresis [10], [11].

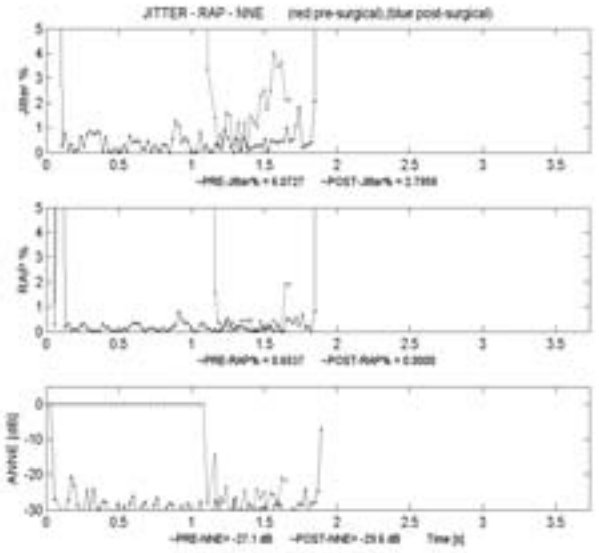

Figure 3 – Jitter, RAP and NNE tracking.

As far as reliability of results is concerned, the method has been compared to one of the most used commercial software tools, i.e. MultiDimensional Voice Program (MDVP®, KayPentax Corp.), where NHR has been considered instead of NNE [11]. First results have shown that BioVoice performs more reliable analysis than MDVP. This could be due to more robust F_0 estimation with BioVoice, and to the different analysis windows used: fixed for MDVP and adaptively tailored to varying F_0 for BioVoice.

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.

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. When properly optimised, the tool could be implemented on a mobile device, as an aid for clinicians, logopaedicians and patients, also for rehabilitation purposes, after surgery or medical treatment.

ACKNOWLEDGMENTS

This work has been partially supported by “Ente Cassa di Risparmio di Firenze”, under the project: n. 2006.1517 "Analisi di segnali ed immagini vocali per applicazioni

biomediche", 2007, and COST Action 2103: "Voice Function Assessment".

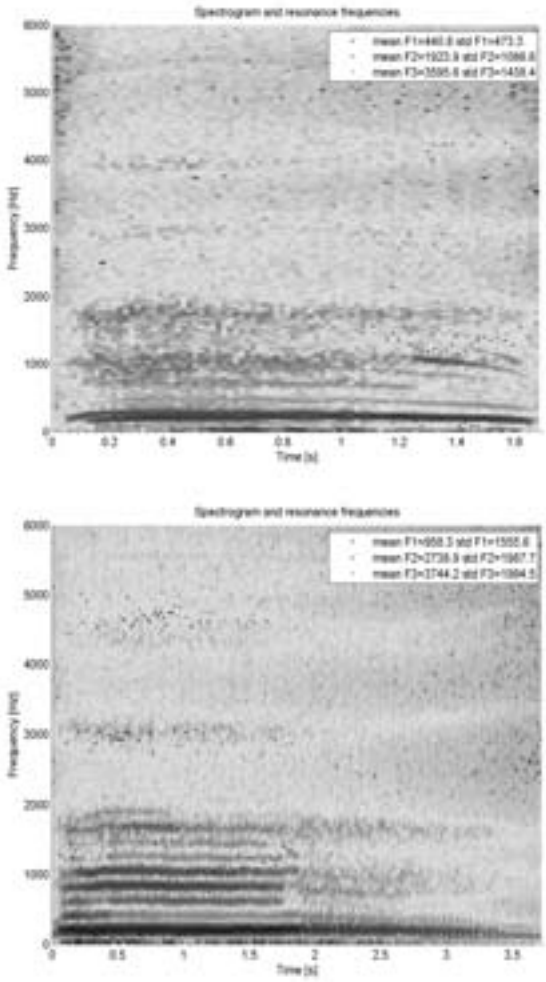

Figure 4 – Pre- (upper) and post-surgical (lower) spectrogram and formants tracking. Mean and std values are displayed.

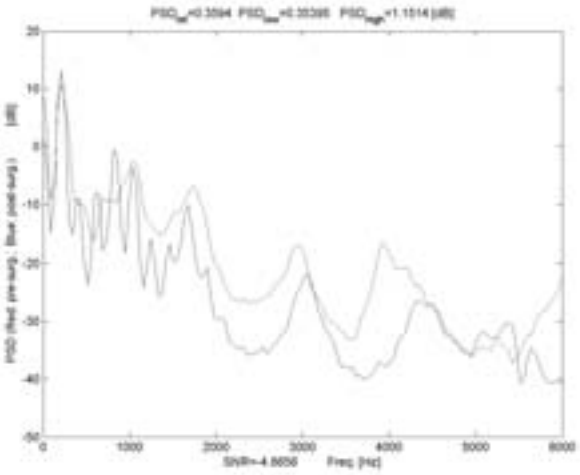

Figure 5 – Pre- and post-surgical PSD plot. Global, low- and high-frequency PSD values are also reported.

VI. REFERENCES

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

[2] D. M. Howard, G. F. Welch, J. Brereton, E. Himonides, M. DeCosta, J. Williams, A. W. Howard, "WinSingad: A real-time display for the singing studio", *Logopedics Phoniatrics Vocology* vol. 29, num 3, p. 135-144, 2004.

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

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

[5] K. Wermke, W. Mende, C. Manfredi, P. Bruscaglioni, "Developmental Aspects of infant's Cry melody and Formants", *Medical Engineering and Physics*, vol.24, n.7-8, pp..501-514, 2002.

[6] T.Sangiorgi, C.Manfredi, P.Bruscaglioni, "Objective analysis of the singing voice as a training aid", *Logopedics Phoniatrics Vocology*, vol.30, n.3-4, p.136-146, 2005.

[7] 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.

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

[9] G.Cantarella, R.F.Mazzola, E.Domenichini, F.Arnone, B.Maraschi, "Vocal fold augmentation by autologous fat injection with lipostructure procedure", Otolaryngology - Head and Neck Surgery**,** vol. 132, n. 2, p. 239-243, 2005.

[10] 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.

[11] C.Manfredi, R.Canalicchio, G.Cecconi, G.Cantarella, "A robust tool to compare pre- and post-surgical voice quality", *EMBS Conf., Lyon, FR*, 23-26 Aug. 2007.