OBJECTIVE ANALYSIS OF THE SINGING VOICE AS RELATED TO SINGER POSTURE

T. Sangiorgi, L. Mazzei[1], F. Felici[1], S. Lapi, G. Testi, C. Manfredi, P. Bruscaglioni[2]

Department of Electronics and Telecommunications, Università degli Studi di Firenze, Firenze, Italy
[1]Fiesole School of Music, Fiesole, Firenze, Italy
[2]Department of Physics, Università degli Studi di Firenze, Firenze, Italy

***Abstract*: This paper deals with objective analysis of the singing voice, and aims at giving non-professional singers both an aid to improve their voice capabilities and a criterion to prevent a wrong vocal attitude (head, neck, and body posture) that could even cause vocal pathologies.**

A new standalone application with a user-friendly interface is proposed, for robust and reliable analysis of singing voice characteristics. The tool performs tracking of fundamental frequency and formants, along with an objective measure of main singing voice parameters, such as vibrato rate (V_rate), vibrato extent (V_ext), and vocal intonation (V_int). A side-view camera allows displaying and recording the singer's posture.

Data are collected at the School of Music in Fiesole, Firenze, Italy, under the supervision of a voice teacher and a teacher of Alexander Technique (AT). First results are presented to compare some vocalizations coming both from professional and non-professional singers under different singer's postures.

***Keywords:* objective voice parameters estimation, singing voice, Alexander Technique.**

I. INTRODUCTION

At present, singing is learned basically by means of the perception and the psycho-physical control of the singer during his/her performance. Also, it is mainly up to the singing teacher to perceptually evaluate the quality of a performance. This makes it difficult defining standard procedures and reference values, also because few objective means to evaluate singer capability and improvements are currently available [1-4].

Singing voice results from complex activity of the larynx and of vocal tract articulators, and is characterized by possibly high-pitched, rapidly time-varying signals. In this preliminary study, the basic features of the singing voice are considered, i.e. the fundamental frequency F0 (linked to vocal folds oscillation), along with its modulation in time and frequency, and the formants Fi (resonance frequencies of the vocal tract), along with their energy. Time evolution, standard deviation (std) and maxima of such parameters over the whole vocalization are of importance for singers, being strictly related to correct vocal emission and hence to singer's performance. Moreover, V_rate, V_ext and V_int are of importance, in order to give the singer useful information on the degree of achieved professional level, possibly as compared to professional singers [4-6].

The AT postural technique has gained interest among singing teachers, for its possible advantages as a complement to vocal training. The AT is a method of re-education based on creating a dynamical, balanced relationship between head, neck and back, and, as a result, on one's whole body. Recent studies have shown that, after few months of AT application, singing voice becomes more resonant, and singing easier to perform [7-8]. Lessons are entirely practical. The teacher gives students helpful suggestions, also by means of a very skilful and subtle use of the hands. The student is taken through simple movements, like standing up, sitting down or walking, to understand the principles on which the dynamics of the whole body is based. The AT is taught in forty countries around the world. It is studied and taught since many years also at the School of Music in Fiesole, Italy.

II. METHODS

Singing voice signals are analysed by means of a multi-purpose, user-friendly tool, based on robust analysis techniques capable to deal also with high-pitched, quasi-stationary signals, that are among those under study.

To track fast signal variations, the signal is divided into short frames, whose length adaptively varies according to varying signal characteristics: the higher the F0 the shorter the frame length (kept fixed to 3 pitch periods). A voiced/unvoiced separation algorithm is implemented, to avoid parameter estimation on signal frames that have no harmonic content.

F0 tracking is achieved by means of a robust two-step procedure, based on well-established results [9]. 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 subject and signal characteristics, based on a

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)

simple relationship between p, F_s (sampling frequency), L (vocal tract length, linked to age and sex), and c (sound speed) [10]. Colour-coded spectrograms are provided, with formants tracking (F1-F5 for singers) superimposed. Mean values and std are also shown. PSD plots complete the set of pictures, allowing detailed inspection of harmonic energy characteristics.

A user-friendly interface (Fig. 1) allows selecting age, sex and type of vocal emission for each subject, 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 (new-born, singers, adults). The overall allowed F_0 range is $40Hz<F_0<1300Hz$;
– selecting the kind of analysis: single audio file or two files, for comparison purposes.

A moving bar shows the residual time during computations. For long files (>5s) and high sampling frequency (>40 kHz) the total time could approach 5min in total.

A number of plots is displayed and saved in printable format, for a visual comparison of results. Specifically, for singing voice, F0, V_rate, V_ext, V_int, spectrogram, formants and PSD are plotted, all in coloured map.

The software tool is developed under Matlab® R2006b.

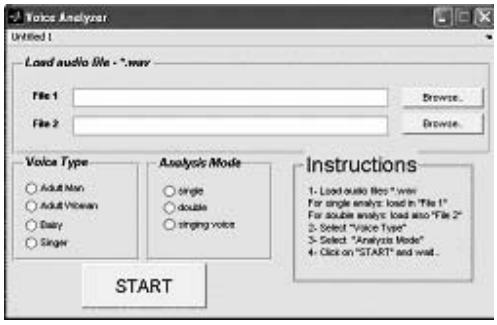


Figure 1 – The user interface for voice analysis

III. RESULTS

First results concerning one professional mezzo-soprano and 3 students (2 tenors, 1st and 2nd year of course, resp., and 1 soprano, first year course) were obtained. Data were recorded with a professional directional microphone (SHURE SM58) equipped with a A/D board TASCAM US144, and stored on a notebook. The sampling rate was F_s=44.1kHz, with 16 bit resolution.

After proper warming up, singers performed vocalizations based on the Italian sustained /a/ vowel, at different F0 values, under the supervision of both a singing teacher and an AT teacher. The choice of /a/ comes from its reasonably independence from other factors, mainly the tongue position. As for students, with no experience with the AT, vocalizations were performed with "natural" posture first (i.e. without conscious control of their body), and then with the posture suggested by the AT teacher.

Instead, the professional mezzosoprano, who is undertaking AT personal training since some years, performed "natural" /a/ vocalizations by intentionally keeping a non-balanced relationship between head, neck and back. Then, she applied the proper AT posture.

To compare results, some plots and parameters are reported here concerning two cases: the professional mezzosoprano and one non-professional soprano, both emitting a sustained /a/ with vibrato. Figs. 2 and 3 show, from top to bottom: signal amplitude, F0, V_int, and the PSD. On the left: without AT; on the right: with AT. V_rate and V_ext are reported below for both cases. According to literature [3-5], a good quality vibrato range should be approximately: $5 \le V\_rate \le 7.5$ cycles/s, and $V\_ext \le 2$ semitones (+/-1 semitone corresponds to a frequency swing of +/- 6 %, approximately).

As for the professional mezzo-soprano, without applying the AT, the following parameters were obtained: $F0_{mean}$=230.39Hz, V_rate=4.9 cycles/s, std=0.39cycles/s, V_ext=10.6Hz (2 semitones, ~28Hz), std=3.8Hz. After training with AT, the parameters were: $F0_{mean}$=231.87Hz, V_rate=5.0 cycles/s std=0.3cycles/s, V_ext=25Hz, std=3.9Hz. Fig. 2 shows some results as obtained with the proposed tool. Notice a more regular vibrato and higher energy for the 3th-5th formants with AT. Also V_int is remarkably more stable. Perceptual evaluation confirms better quality of the AT vocalization, that seems to enhance the singer's performance in this case.

Fig.3 shows the results obtained for the non-professional soprano. Without applying the AT, we found: $F0_{mean}$ = 439.7Hz, V_rate=5.2 cycles/s std=0.6cycles/s, V_ext=23.3Hz (2 semitones, ~53Hz), std=4.6Hz. After training with AT: $F0_{mean}$=439.7Hz, V_rate=4.9 cycles/s std=0.4cycles/s, V_ext=22.5 Hz, std=3.8Hz.

Notice that vibrato values are quite similar in both cases. However, different vocal strategies were applied, with different formants frequency and energy, especially above 3 kHz, as shown in the PSD plot. With AT, better perceptual results were obtained.

As for tenors, the analysis has shown no remarkable voice quality improvement with AT, in agreement with perceptual evaluation. This can be due to the following factors.

Professional singers make use of a precise control of both laryngeal and vocal tract functions, with several and continuous adjustments, that make up the basic tools for good singing. On the contrary, students did not yet developed a good auditory and self-receptive feedback,

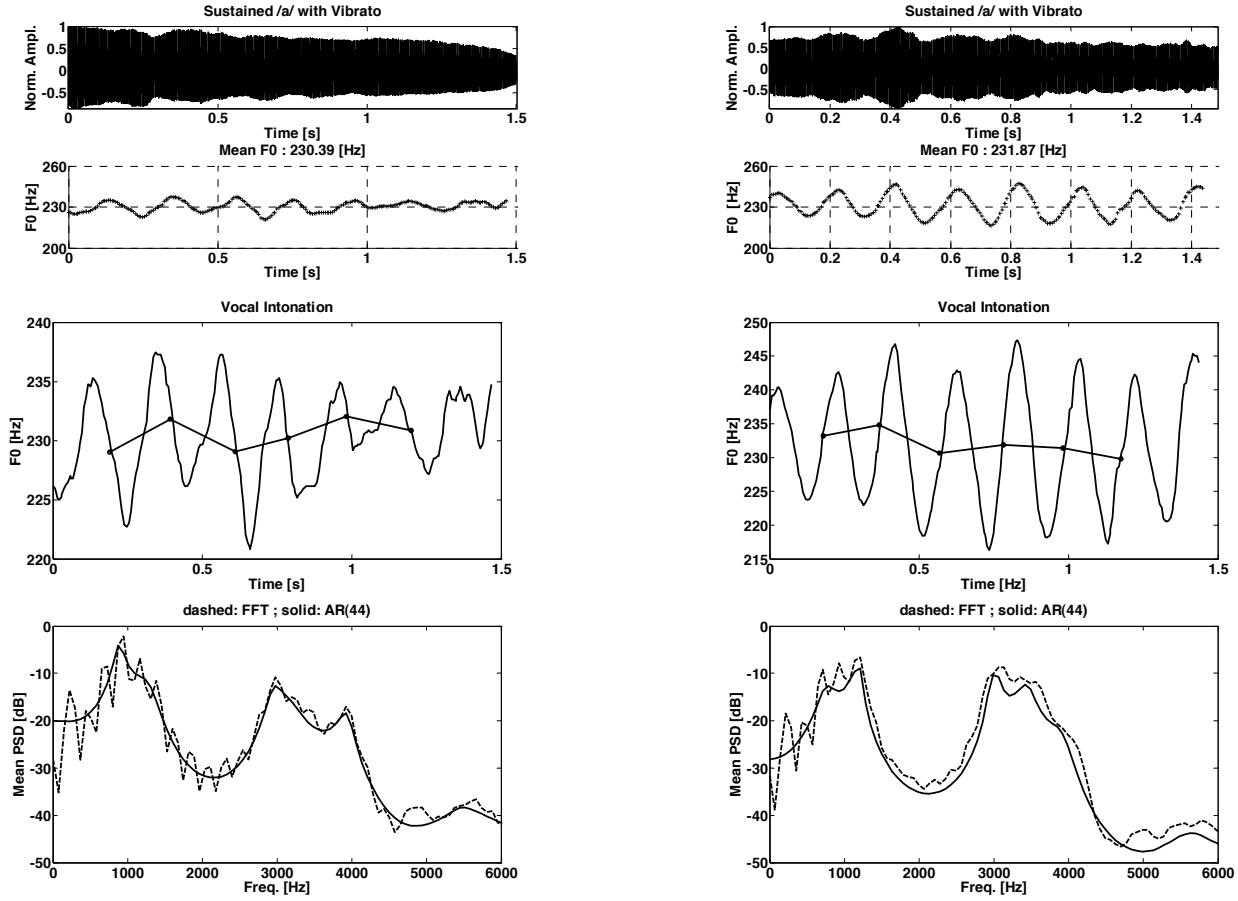


Figure 2 – Professional mezzo-soprano. Left: without AT; Right: with AT

nor a reliable muscular training. This often causes a modified, para-physiological vocal emission, that involves the whole neck and shoulders, and hence an altered global posture. The AT makes unstable the attitudes that are deep-rooted in students' postural habits and needs great psycho-physical concentration. This fact could even have made unstable the knowledge and compensatory skills used by students in vocal emissions before the AT teacher suggestion. Finally, the students involved in this experiment being non-professionals, were not accustomed to sing with an audience. The presence of microphone and camera could have further influenced their performance.

Notice that, from a perceptive point of view, the AT has softened the onset of the vocalization in most cases.

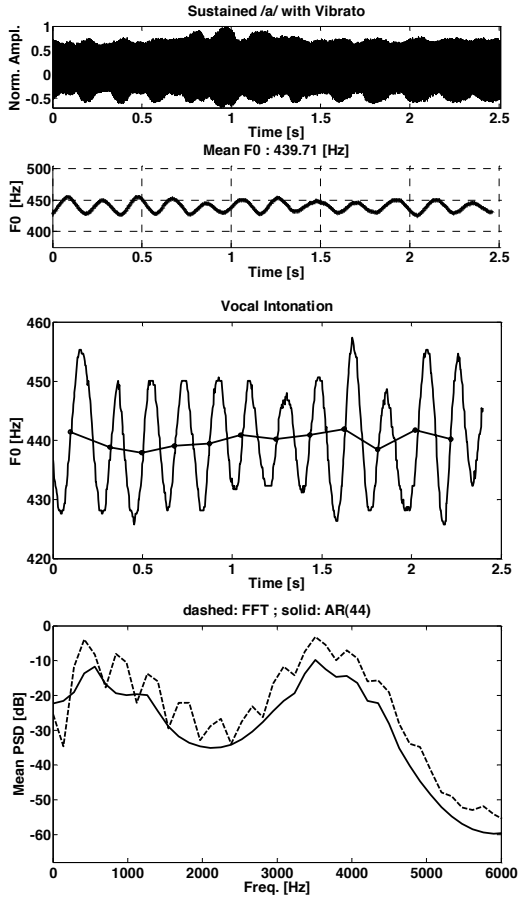
Though very limited in number and in time, first results allow for supposing that, if non-professionals add AT to their vocal training, they could found easier to enter the main functions related to a proper sound emission, and could be facilitated in overcoming limitations, such as limited vocal extension, voice breaks, an improper use of vocal registers, and often vocal fatigue. As functions are carried out on posture, a good postural balance allows for the cheapest usage of a function, and makes it possible to perform even subtle adjustments.

Features of the professional singer were found in agreement to those proposed in literature. If a larger set of data will be available, a reference set for non-professionals could be set up.

IV. FINAL REMARKS

A user-friendly, robust tool for voice analysis has been presented. It allows for the analysis of voice recordings, in a wide range of F0 values, that makes the tool a multi-purpose one. At present, the new tool works off-line. If properly implemented, it would allow for real-time analysis of voice signals.

As for singing voice, preliminary results show different F0 and formant strategies, as related to singing technique and/or posture. Hence, it could be of help in giving non-professional singers and singing teachers reliable objective measures of possible improvements during and after training with any teaching technique.



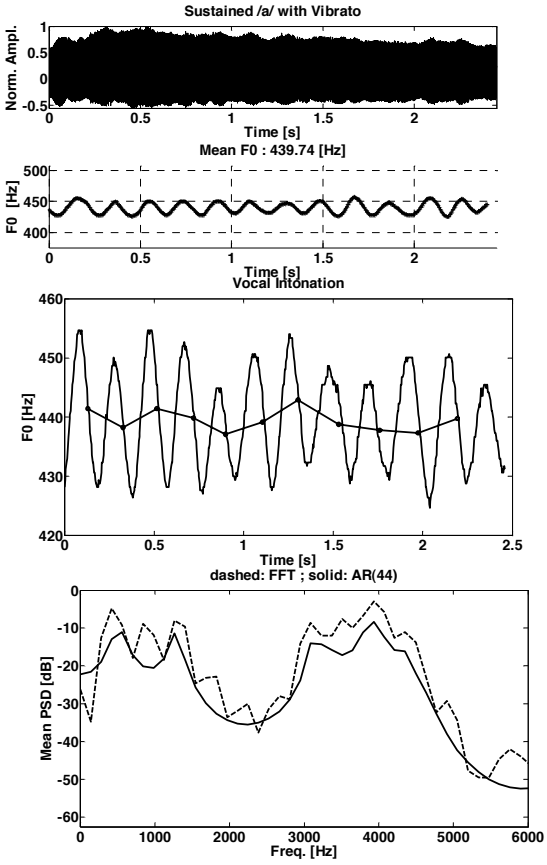


Figure 3 – Non-professional soprano. Left: without AT; Right: with AT

Collecting and analyzing several audio/video files (not reported here due to space limitations) is going on, in order to give more reliable results.
Finally, further studies are needed to investigate in detail the influence of a proper posture in singing voice production.

ACKNOWLEDGMENTS

This paper 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.

V. REFERENCES

[1] 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.

[2] P.E. Garner, D. M. Howard, "Real time display of voice source characteristics", *Logopedics Phoniatrics Vocology* vol. 24, 19-25, 1999.

[3] 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.

[4] T. Shipp, R. Leanderson, J. Sundberg, "Some acoustic characteristic of vocal vibrato", *J. Research in Singing*, 4, pp.18-25, 1980.

[5] J. Sundberg, “The Science of Singing Voice", *North. Illinois Univ. Press, Dekalb, Illinois*, 1987.

[6] F. Fussi, S. Magnani, "L’Arte Vocale", *Omega*, 1994.

[7] F. P. Jones, “Freedom to Change”, London: Mouritz, 1997. Mornum Time Press, San Francisco, CA.

[8] C. Stevens, “The Alexander Technique”, Rutland, Vermont: Charles E. Tuttle Co., Inc., 1987, 1994.

[9] 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.

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