VOICE MONITORING: TECHNICAL AND CLINICAL ASPECTS

Claudia Manfredi[1], Philippe H. Dejonckere[2]

[1]Department of Electronics and Telecommunications, Faculty of Engineering, Università degli Studi di Firenze, Via S.Marta 3, 50139 Firenze
[2]Catholic University of Leuven, Neurosciences, Exp. ORL, Belgium; Federal Institute of Occupational Diseases, Brussels, Belgium; Utrecht University UMC, AZU Heidelberglaan 100, F.02.504, NL 3584 CX Utrecht, The Netherlands

Abstract: **Occupational voice disorders are observed with increasing frequency in otolaryngological consultations. Devices have been developed that provide objective data on the way individuals use their voices throughout the day outside the clinic. However they do not provide information about the acoustic indexes of voice quality. A critical point is also the choice of the sensor.**
The device here proposed could be defined as a "portable laboratory" for voice analysis, its main advantage being the reliability of estimated parameters from both sustained vowels and running speech. A prototype has been set up on a DSP board and tested on short sustained vowels. Foreseen applications include: basic research, medico-legal, quantification of voice plasticity, vocal function exercises during rehabilitation, voice disorders, short term feedback in singing voice, etc.

Keywords : **Voice analysis, dosimeter, portable device, occupational voice disorders**

I. INTRODUCTION

The development of modern information telecommunication technology plays an increasingly important role in facilitating access to some diagnostic services, particularly in creating medical diagnostic applications small enough to fit into objects already in common use, such as cell phones.
Occupational voice disorders are observed with increasing frequency in otolaryngological consultations [1]. Speech therapists in voice clinical services rely on documenting information on therapy progress recording the examination/therapy session to diagnose the voice quality more precisely comparing the voice quality of the patient at the beginning, during and at the end of the therapy session and to review the evaluation later.
To this aim voice dosimeters and voice accumulators have been investigated, and suitable definitions of vocal load and dose have been given and applied to professional speakers and singers [2-9].
However few devices have been implemented, mostly based on a contact transducer (accelerometer) attached to the front part of the neck. A cable connects the accelerometer to the hardware module in a waist pack worn by patients. These devices provide data on the way how individuals use their voices throughout the day, outside the clinic, avoiding relying solely on subjective self-reports. In particular APM [10] records the total speaking time and sound level over a period of several hours. Quantitative measures of when, how long, how loud, and at what pitch the client vocalizes are obtained and a real-time feedback is provided, through a small vibrotactile unit. This information is very useful to identify those situations which might cause vocal fold damage. Other products implement similar voice quality parameters and indexes [11, 12]. Nevertheless they do not provide information about the acoustic indexes of voice quality. Another drawback with existing devices is the possible discomfort and embarrassment due to the contact transducer and the need of being returned to the clinic to download data into a PC for analysis using specific software. Moreover, a critical point is of course the correct wearing of the accelerometer that again could require clinical expertise.
The device here proposed differs from those above mentioned as is will be completely contact-less, the transducer being a small microphone. It could be defined as a "portable laboratory" for voice analysis, its main advantage being the reliability of estimated parameters of both sustained vowels and running speech and easy usage. Foreseen applications are:

- Research, to understand the early effects of fatigue on voice quality and/or the early mechanisms of vocal forcing.
- Medico-legal, by means of so-called "realistic provocation" test for patients that show a normal voice at the moment one examines them, but acknowledge a lot of voice symptoms during daily life.
- Quantification of voice plasticity in realistic conditions of use in voice professionals.
- Post-surgical monitoring and vocal function exercises during rehabilitation or for stuttering, dyslexia, psychogenic dysphonia, etc, where indirect interaction with the therapist could be more comfortable and effective.
- Occupational voice disorders (speakers, call center operators, teachers etc. with chronic voice over-use) to monitor how vocal folds react to the daily load and to

Claudia Manfredi (edited by), *Models and analysis of vocal emissions for biomedical applications : 7th international workshop : August 25-27, 2011*, ISBN 978-88-6655-009-9 (print) ISBN 978-88-6655-011-2 (online)

receive immediate feedback about possible risks. A long-term usage of the device could be foreseen over the days or weeks in subjects who are at risk for the development of voice pathologies, for easily available monitoring to be used by the voice clinician.
- Short term feedback in singing voice, while trying different vocal behaviours/voicing styles etc., e.g measuring the singer's formant and checking which one provides the best 'brilliance' to dominate the orchestra.

II. METHODS

A prototype has been set up on a DSP board that evaluates voice basic parameters and provides a LED/audio feedback that advices the patient for any abnormal vocal emission. The aim is to implement it on an object of common use, such as a cell phone, in order to overcome patient's distrust against medical devices. Data (audio files and parameters) could be saved on the device and possibly submitted to a PC for further analysis. This could be accomplished e.g. by means of MMS messages.

Voice quality indexes: On the prototype, voice quality analysis is based on the following indexes: fundamental frequency (F_0), along with its irregularities (Jitter, J, and Relative Average Perturbation, RAP), and hoarseness (Normalised Noise Energy, NNE). This is a subset of functions coming from a new user-friendly tool for voice analysis, named BioVoice [13] developed under Matlab R2009b, that can be easily extended to other relevant parameters related to vocal load.

Great attention is devoted to the selection of voiced/unvoiced frames as well as th F_0 estimation, as reliable estimates of other parameters depend upon it. The choice of the techniques adopted results from a detailed comparative analysis of F_0 extraction methods, with applications both to synthetic and real data in case of mild to severe dysphonia, showing enhanced performance against other approaches [14, 15].

The DSP prototype: The routines for F_0 and voice quality evaluation, developed under Matlab (release 2007b), were translated into C++ code (Microsoft Visual C++ 6.0) and optimised in order to run on the DSP board TMS320C6713, that is provided with a larger internal memory (192 kB) and faster clock (225 MHz) with respect to the one previously used [16]. The new DSP board allows for implementing many computations directly on a data buffer inside the internal memory. The buffer also allows for fewer transfers of data from external to internal memory. Moreover, floating point variables were implemented.

The board can work independently or connected to a laptop or PC (Fig. 1) for launching the debug and for showing on the monitor some plots as result of computations. The new DSP was also provided with the software required for audio signal recording through a microphone that must be connected to the MIC-IN input.

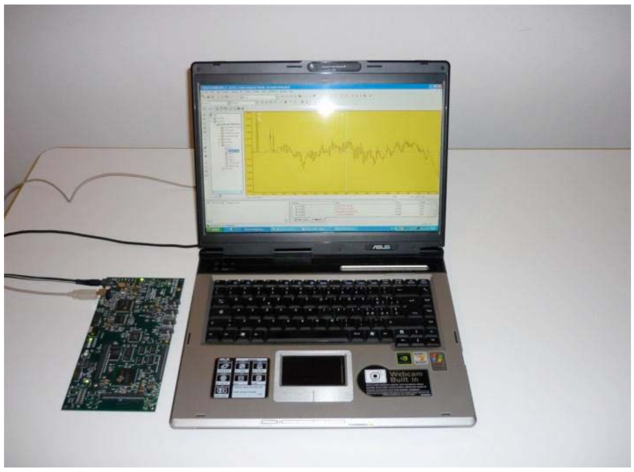

Figure 1: The DSP board connected to a laptop for further analysis and display of plots.

The microphone: A high quality voice recording should differentiate this system from existing ones, to allow for the evaluation of voice quality parameters and subtle differences induced by e.g. changing something in the voicing technique. Hence, the microphone plays a basic role in the device as it is required to be used in field. According to [17], if ambient noise or reverberation are a problem, as in this application, a head mounted omnidirectional microphone (that however reduces portability) or a directional microphone are suggested.

At present, the device works with a fixed distance of the mouth from a directional microphone. Recordings must be performed in a quiet environment. Users are carefully instructed and warned about these points, though a control has been implemented to test minimum amplitude, signal power and background noise requirements, in order to guarantee a satisfactory signal level while avoiding saturation. If such requirements are met, the signal analysis starts, according to the implemented algorithms. Otherwise, a devoted LED/audio alarm advices the patient that the recording must be repeated.

On the prototype the size of the data frame is limited to 2 s of recording, with F_s = 44 kHz sampling frequency, but of course longer frames are foreseen. More details can be found in [18].

As for any portable clinical device, before leaving the clinic to pursue her/his daily activities, the patient will receive instruction about how and when use the device. The clinician has to customize the device for each patient to elicit the audio/video alarm when a particular threshold, such as an irregularity value, is exceeded.

III. RESULTS

The new board has been tested on two sets of voice signals (sustained /a/ vowel). The first set consists of 40

pre-post surgical recordings with different degrees of hoarseness due to different pathologies (polyps, oedemas, cysts, tyroplastic prosthesis etc). The second set comes from healthy subjects recorded in non-protected environment, to test robustness against environmental noise.
The mean values of the parameters F_0, ANNE, J, and RAP were considered. The same computations were performed on both the DSP and the Matlab program running on a standard PC.
Table 1 shows some results from a subset of pathological signals (lines 1-8) and from male healthy subjects (lines 9-12). Only the results obtained from the DSP are reported, as they coincide up to the last digit with those obtained with Matlab. The computational time was 30-50 s on DSP, and 13 s on PC. Notice that the computational time could be greatly reduced if a reliable initial range of variation for F_0 is available that avoids a first step for F_0 estimation.

IV. Discussion

At present the proposed device is at a first stage of development, both as far as the implemented parameters and the hardware requirements are concerned. Adding other parameters of clinical relevance poses relatively simple problems, mainly concerning computational time, that could be solved with dedicated hardware or other techniques such as sending data to a server connected to a PC for visualization and further analysis with devoted tools (e.g. BioVoice). Parameters could include formants, spectrogram and PSD, as well as statistical results and plots on the whole recording period.
Sending data could be done through the GPRS/UMTS network (e.g. MMS messages) that should warrant for privacy, as the user could be identified through the telephone number available on a SIM card obtained only presenting a personal document and signing a legal document. The device could be provided with a HDD or memory card to store data and a USB connection to download data on a PC.
Moreover, an optimised version of the software could be developed to be downloaded as an application for mobile phones or i-phones.
The voice quality enhancement problem against environmental noise and/or simultaneous presence of other speakers rises more difficulties that could be partially solved applying e.g. blind source separation techniques or neural network algorithms to teach the phone to recognise the voice of the user against other voices or sounds. Another possibility could be that of applying spectral subtraction techniques, that would require having two microphones installed on the device. The first microphone should be very close to the mouth, as in usual mobile phones, while the second one should be mounted at a certain distance.
Also sound pressure level (SPL) should be taken into account. Some commercially available SPL meters do not need a calibration and will be investigated in future work
Another characteristics could be adding the possibility for the subject to indicate (by means of a button) relevant moments: the patient pushes on a button when he/she starts perceiving fatigue or burning throat, or other symptoms. Comparing voice quality before-during-after the button is pushed could give useful information in real time so that the patient could immediately react.
To keep the device user-friendly, at the output the user will be advised for abnormal phonation with intuitive audio/visual messages only. In-depth analysis is deferred to the complete analysis tool, available on a laptop or PC.

V. Conclusion

A DSP prototype is proposed for a contact-less portable device to be used by a patient in order to extract important parameters of vocal behaviour when pursuing normal activities. In addition to the important objective data the device provides, a real-time audio/video alarm is implemented, as a feedback tool to help patients to remind abusive vocal behaviours during routine daily activity and help the patient to learn how to modify vocal behaviour and achieve desired vocal function as defined by the clinician. This feature may enhance therapy carryover and expedite the patients rehabilitation process.
The proposed device could be useful for clinicians to monitor results of phonosurgery and to obtain objective acoustic data for statistical and scientific purposes avoiding expenses and time consumption to the patient under study.
Clinicians as well as speech therapists and psychiatrists could have benefits to obtain objective acoustic data for statistical and scientific purposes, avoiding a waste of money and time to the patient under study.
The possibility of making use of a simple and reliable self-monitoring tool, for non-expert users, with no restrictions on accessibility and logistics, will allow sensitising people on a still underestimated subject, such as the prevention of vocal apparatus pathologies

ACKNOWLEDGEMENTS

This work was supported by Ente Cassa di Risparmio di Firenze, under the project: “Interdisciplinary Laboratory for Biomedical Acoustics - LIAB”, n.2007.0754

REFERENCES

[1] P.H. Dejonckere (Editor), Occupational voice: care and cure, Kugler Pub., The Hague, The Netherlands, 2001.

Table 1 - F_0, Jitter, RAP, ANNE mean values and computational time with the DSP board. Slightly longer time (5”-10”) is observed for female patients with respect to males, due to higher number of loops performed that is proportional to F_0 (usually higher in females).

File name	F_0 [Hz]	Jitter %	RAP %	ANNE [dB]	Comp. Time [s]
Pathol_male_1	104.4127	0.7395	0.3151	-17.0376	32.23
Pathol_ male_2	113.7185	0.4804	0.1961	-29.6496	33.49
Pathol_ male_3	156.7025	1.3709	0.4563	-27.6346	39.20
Pathol_ male_4	136.9295	0.6144	0.1039	-16.8766	36.31
Pathol_ female_1	202.0758	0.9030	0.0601	-28.6182	44.83
Pathol_ female_2	186.3836	0.3716	0.0833	-28.6481	43.42
Pathol_ female_3	184.5098	0.8682	0.0513	-29.2505	42.66
Pathol_ female_4	233.0304	0.5477	0.0584	-24.9314	50.07
FileTest1_a	116.5368	0.5320	0.0656	-23.7415	33.63
FileTest2_a	146.1851	0.2697	0.0000	-26.2802	37.49
FileTest3_a	114.4827	0.4110	0.0970	-24.7037	33.36
FileTest4_a	93.7563	0.7054	0.0942	-19.4262	31.45

[2] Zicker, J.E., Tompkins, W.J. Rubow, R.T. Abbs, J.H. (1980), A portable microprocessor-based biofeedback Training device, IEEE Trans. Biomed. Eng., 27, 509-515.

[3] Szabo, A., Hammarberg, B., Håkansson, A., Södersten, M., (2001), A voice accumulator device: evaluation based on studio and field recordings. Logoped. Phoniatr. Vocol. 26, 102-117.

[4] Cheyne, H.A., Hanson, H.M. , Genereux, R.P., Stevens, K.N. , Hillman, R.E., (2003), Development and testing of a portable vocal accumulator, J. Speech Lang. Hear. Res. 46(6),1457-67.

[5] R.E. Hillman , J.T. Heaton , A. Masaki , S.M. Zeitels , H.A. Cheyne . (2006), Ambulatory monitoring of disordered voices, Annals of Otology, Rhinology, and Laryngology 115, 795-801.

[6] Titze, I. R., Švec, J. G., Popolo, P. S., (2003), Vocal Dose Measures: Quantifying Accumulated Vibration Exposure in Vocal Fold Tissues, J Speech Lang Hear Res, 46, 919 – 932

[7] Carroll, T., Nix, J., Hunter, E., Emerich, K., Titze, I., Abaza, M., (2006), Objective measurement of vocal fatigue in classical singers: A vocal dosimetry pilot study, Otolaryngol. - Head Neck Surg. **135**, 595-602.

[8] Kelchner, L. N., Toner, M. M., Lee, L., (2006), Effects of Prolonged Loud Reading on Normal Adolescent Male Voices, Lang Speech Hear Serv Sch, 37, 96 - 103.

[9] J.Nix, J. Svec, A. Laukkanen, I. Titze. Protocol Challenges for On-the-Job Voice Dosimetry of teachers in the United States and Finland. Journal of Voice 21 (2007): 385-396.

[10] www.kayelemetrics.com/Product20Info/3200/3200.htm

[11] www.wevosys.com

[12] www.sonvox.com

[13] Manfredi, C., Bocchi, L., Cantarella, G., (2009), A multipurpose user-friendly tool for voice analysis: application to pathological adult voices, Biomed. Signal Proc. and Control, **4**, 212-220 .

[14] Manfredi, C., D’Aniello, M., Bruscaglioni, P., Ismaelli, A., (2000), A comparative analysis of fundamental frequency estimation methods with application to pathological voices, Med. Eng. Phys., **22**, 135-147.

[15] Manfredi, C., Giordano, A., Schoentgen, J., Fraj, S., Bocchi, L., Dejonckere, P.H., (2011), Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools, Biom. Signal Proc. and Control (in print).

[16] Manfredi, C., Bruschi, T., Dallai, A., Ferri, A., Tortoli, P., Calisti, M., (2008), Voice Quality Monitoring: a Portable Device Prototype”, Proc. 30th EMBS Conf., Vancouver, CA, 997-1000.

[17] Granqvist, S., Švec, J., (2009), Microphones and Room Acoustics and Their Influence on Voice Signals, 8th Europ. Voice Conf., PEVOC8, Dresden, Germany.

[18] C.Manfredi, G.Cantarella, “A contact-less portable device for voice quality monitoring”, Riv. Ital. Acust., vol. 35, n.1, p.53-57, 2011.