COST 2103 – FIRST WORKSHOP ON VOICE FUNCTION ASSESSMENT

CLAUDIA MANFREDI

STATE-OF –THE-ART

Methods

Signal analysis

We have recently set up a software tool to perform pre- and post-surgical (or treatment) comparison of main voice characteristics (fundamental frequency (F_0), noise, formants) by means of robust analysis techniques, specifically devoted to deal with highly degraded voice signals as those under study.

The new tool, provided with a user-friendly interface, allows for automatic analysis of a large range of signal, no manual setting being required to the user. This makes the tool feasible for application by non-expert users. Specifically, the user just selects: data (.wav files, pre- and/or post surgical), voice type (ranging from high-pitched new-born and singers voices to adult male or female voices (the overall allowed F_0 range is $40Hz<F_0<1300Hz$), and kind of analysis (single audio file or two files, for pre- and post-surgical comparison).

The software tool automatically adjusts internal settings for optimal frame length, frequency range of analysis and plots.

At present, few but effective indexes are implemented, devoting great effort to their robust and automatic evaluation.

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 separation algorithm is implemented, to avoid parameter estimation on signal frames that have no harmonic content.

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. Jitter and RAP are also evaluated and tracked.

An adaptive noise estimation technique is implemented, that allows tracking varying noise level during phonation. It relies on a comb filtering approach (NNE), optimised in order to deal with data windows of varying length.

Finally, robust and 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 patient and signal characteristics, based on the relation $p = 2LF_s/c$, where: F_s = sampling frequency, L = vocal tract length (linked to patient's age), and c = sound speed.

Colour-coded spectrograms are also provided, with the tracking of the first three formants superimposed. Mean values and std are also shown. Comparing pre- and post-surgical spectrograms gives a first qualitative view of the residual noise present in the voice signal before and after surgery, as well as of harmonics intensity and stability.

PSD plots complete the set of pictures, allowing visual inspection of possible harmonic energy recovering. On the plot, PSD_{tot}, PSD_{low}, PSD_{high} quantify the signal global energy, the low-frequency (<2500Hz) and the high-energy one, respectively. This could further help the clinician in assessing voice quality recovering.

Plots are displayed and saved in printable format, for a visual comparison of results. Specifically, F_0, jitter, RAP, ANNE, spectrogram, formants and PSD are plotted, all in coloured map, along with their mean and std.

Image analysis

A digital image processing algorithm has been developed, optimized for real-time analysis of VKG recordings that require intensity adjustment and noise removal for exact glottis identification. Robust techniques for edge detection have been implemented, based on snake "surface" active contour, greatly reducing the presence of both noise and artefacts. The new tool has been provided with a user-friendly interface for managing patients' data and image analysis, according to a set of parameters that can be easily settled by the user, some of which devoted to refinement of the image analysis procedure, in case of strong noise.

The following parameters have been implemented:

R_amp = ratio between the amplitude of one vocal fold vibration and the opposite one. R_amp is an index of asymmetry.

R_per = ratio between the vibration wavelength of vocal folds. It is inversely related to frequency variation due to pathological structural alterations of the vocal fold.

R_oc = ratio between the measure of the length of the opening and closing phase within a single glottis cycle, related to the degree of glottis incompetence.

PSI, Phase Symmetry Index = ratio between the difference of the two vocal folds phases and the period T.

Moreover, in case of glottis incompetence, the mean distance between folds, M_op, has been evaluated.

Plots and pictures can be stored and retrieved, allowing the clinician an objective evaluation of results with reliable and reproducible measures.

APPLICATIONS

1. **New-born infant cry** (melody and formants) characterising human phonatory development. Basic voice parameters are also of importance as a non-invasive diagnostic aid for cerebellar pathologies and/or dysfunctions in new-born infants.

2. **Voice signal in dysphonic adult patients**. The aim is to give the clinician objective measures of parameters characterising vocal emissions that should allow evaluating surgical effectiveness, following post-surgical voice recovering and rehabilitation. The method has already been successfully applied in case of tyroplastic medialisation and to patients suffering from cysts and polyps that underwent micro-laryngoscopic direct exeresis (MLSD).

3. Objective measure of some key **singing voice parameter** (vibrato rate, vibrato extent, vocal intonation). High-resolution formant tracking, to find singer's formant bandwidth and intensity. First results have been obtained, as a help for non-professional singers and singing teachers to quantify possible improvements during and after training.

FUTURE TRENDS

Signal analysis **-** Assessing voice quality recovering by means of objective parameters is of great relevance for clinicians. 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.
As for the analysis tool, further details and parameters will be added according to the signal under study. When properly optimised, it could be implemented on a DSP board, as a mobile device useful for clinicians, logopaedicians and patients, also for rehabilitation purposes, after surgery or medical treatment. A prototype is under construction, for patient's voice recording, real-time analysis, and display of the results on the monitor. It should also allow for homogeneous data storing and transmission, to a server devoted to collecting and re-directing data to the National Health Service (family doctor and otorhinolaryngologist, by means of suitable data bases).

Image analysis **-** Based on first results, on literature and clinicians suggestions, future work will concern adding further parameters, as well as further simplifying the interface and refining the contour detection algorithm. Specifically, work is in progress as far as initial contour detection, starting interface, menu options and graphical display of parameters tracking are concerned.
New VKG enhanced devices will allow simplifying the contour detection procedure, as well as the parameters extraction techniques.
When applied to a large set of data, the new tool would allow defining reference values for normal and pathological cases, providing a valid support to diagnosis and surgical effectiveness evaluation.

Applications **-** Define and built reliable simulation models, in order to test the developed analysis techniques. Compare results to those obtained with most widely used commercial software.
Study of vocal emissions in patients affected by neurological dysfunctions (spasmodic dystonia, tremor, epilepsy, depression), under pharmacological treatment.
For new-borns, study of possible cerebellar pathologies and/or malformations of vocal and/or hearing system, as well as of the central nervous system (dysfunctions and malformations of the palate, neurological syndromes, vocal tract and glottis paralysis or malformations, and so on). Blood oxygenation vs. pain cry in preterm newborn infants is under investigation.

CO-OPERATION

Outside COST 2103:
Department of Physics, University of Firenze, Children Hospital E. Meyer, Firenze, University Chorus of Firenze, School of Music, Fiesole, Department of Information Engineering and CNR Clinical Physiology Institute, Pisa, IV Otolaryngology Clinic, Polyclinic Hospital of Milano, Univ. of Milano, Civil Hospital of Brescia, University of Catanzaro.
A PhD student from Madrid Polytechnic University (supervisor: J.I.Godino) is working within our Laboratory on VKG image analysis, as far as new contour detection algorithms are concerned.

Within COST 2103:
First encouraging contacts and common research interest have been recently set up with J.Svec and D. Howard, respectively concerning VKG and singing voice analysis. Possibly, other STSMs would be planned in the next future.