We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

5

Oesophageal Speech's Formants Measurement Using Wavelet Transform

Begona García Zapirain,
Ibon Ruiz and Amaia Mendez
Deustotech Institute of Technology,
Deustotech-LIFE Unit, University of Deusto, Bilbao,
Spain

1. Introduction

One of the most important concerns for the specialists in otorrinolaringologists and the patients who have suffer a laringectomie is a complex process for their rehabilitation. At the present, it is no available any advanced technique either for the learning or the evaluation of this process.

Esophageal speech is characterized by its low intelligibility, which implies that its objective measurement parameters e.g. pitch, jitter, shimmer or HNR have values outside normal ranges [1]. One of the consequences of this fact is the impossibility of using speech recognizers, speech to text converters or any kind of automatic response device that requires a speech signal.

The here presented paper explains a work which is included in a research whose objective is to adapt speech controlled systems so that they can be used by people with vocal disorders. Esophageal voices are the most grievous among these pathologies.

Our research group has presented many works to the scientific community [2], [3], aimed to the improvement of esophageal speech quality by stabilizing the poles of the system which models the vocal tract with LPC. Nowadays the wavelet transform is being used in order to enhance the Harmonics to noise ratio. For this task, it is crucial to know accurately the frequency values of formants in vowels [7].

In this paper results of a new algorithm are presented, this algorithm uses Wavelets Transform as basis, but proposes a new technique to improve calculation accuracy. In order to evaluate this new technique a comparative between its results and the ones obtained with the LPC will be elaborated. As a reference for the comparative the results of analyzing the FFT transform will be taken [4].

The general objective of the chapter is the enhancement of esophageal speech quality in communications with humans and machines. This aim comes up of the low intelligibility of people who speak with esophageal voice after an operation called laryngectomy which is carry out like treatment of larynx cancer [6].

2. Methods

2.1 Wavelet transform

One of the most important techniques applied in the spectral analysis is the Fourier Transform (STFT), which will allow to recognize the spectral components of speech signal, so it makes possible to distinguish pathological voices and process them.

That transform has a resolution problem which is given by Heisenberg Uncertainty Principle. The Wavelet Transform (WT) was developed to overcome some resolution related problems of the STFT. It is possible to analyze any signal by using an alternative approach called the multiresolution analysis (MRA).

MRA, as implied by its name, analyzes the signal at different frequencies with different resolutions. MRA is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. The Continuous Wavelet Transform (CWT) is used for many different applications and it is defined as follows:

$$\Psi_x^\psi (\tau,s) = \frac{1}{\sqrt{|s|}} \int x(t) \cdot \psi * \left(\frac{t-\tau}{s} \right) dt \qquad (1)$$

As the here used signals are digital, it is more useful to use Semi-discrete Wavelet Transform (discretized by dyadic grid, described by $s = 2^j$ and $t = k \cdot 2^j$) or Discrete Wavelet Transform (DWT). The DWT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information [5].

The decomposition of the signal into different frequency bands is simply obtained by successive highpass and lowpass filtering of the time domain signal. The original signal x[n] is first passed through a halfband highpass filter g[n] and a lowpass filter h[n]. This constitutes one level of decomposition and can mathematically be expressed as follows:

$$y_{high}[k] = \sum_n x[n] \cdot g[2k-n] \qquad (2)$$

$$y_{low}[k] = \sum_n x[n] \cdot h[2k-n] \qquad (3)$$

where y_{high}[k] and y_{low}[k] are the outputs of the highpass and lowpass filters, respectively, after subsampling by 2. This decomposition halves the time resolution since only half the number of samples now characterizes the entire signal.

However, this operation doubles the frequency resolution, since the frequency band of the signal now spans only half the previous frequency band, effectively reducing the uncertainty in the frequency by half. The above procedure, which is also known as the subband coding, can be repeated for further decomposition.

The wavelet packet method is a generalization of wavelet decomposition that offers a richer signal analysis. Wavelet packet atoms are waveforms indexed by three naturally interpreted

parameters: position, scale (as in wavelet decomposition), and frequency. It will be then selected the most suitable decomposition of a given signal with respect to an entropy-based criterion.

2.2 Basis of speech analysis

At the present time, many otolaryngologists (ORLs) use the software tools they have available in order to corroborate the diagnosis of vocal cord pathologies by means of objective parameters. These parameters complete the information gathered by the specialist, which usually comprises: the images obtained from a stroboscope and several perceptual tests carried out on the patient.

Special attention needs to be paid to vocal cord cancer, that is to say, to its diagnosis, treatment, rehabilitation and monitoring, as this cancer can cause the death of the patient suffering from it. Once the cancer has been detected, the ORL specialist removes the patient's vocal cords. This means that the patient will no longer be able to produce what is called laryngeal voice and thus loses his/her speech.

After the operation, during rehabilitation, the patient begins the process of learning how to emit oesophageal voice: the voice produced by modulating air coming from the oesophagus. This enables the patient to communicate, albeit experiencing great difficulty to maintain fluent conversations, due to the poor quality of oesophageal voice. However, one of the major problems is that this type of oesophageal voice cannot be evaluated during the rehabilitation process as there is no application available on the market that can automatically obtain the previously mentioned acoustic parameters. The quality of oesophageal voice is so low that the algorithms obtaining the periodicity of the voice do not work properly, and thus measurements obtained by such software packs are not reliable.

Obviously, the accuracy of measurements made by the software pack presented in this work will also be applicable to less severe pathologies, such as polyps, nodules, hypo mobility of the vocal cords, etc. The deterioration of the voice in this type of pathology is also too high for the measurement of objective parameters to be precise. This means that these commercial software packs are not suitable for measuring these parameters in voices suffering from some kind of pathology. Being able to obtain accurate objective parameters is advantageous for the early detection of cancer in cases where the patient's laryngeal voice is of a very poor quality and has high noise levels [1].

The pitch, or fundamental frequency of the speech, is one of the properties of sound or musical tone perceived through frequency. Due to this natural pseudo-periodicity of the voiced voice, there are small variations in the peaks of the voice which change their fi frequency, so that the pitch can be defined as:

$$Pitch(Hz) = \frac{\sum_{i=1}^{N} f_i}{N}$$

(4)

N being the number of pitch periods.

Estimating fundamental frequency has been a recurring issue in the area of digital signal processing. This is due to the fact that obtaining the time instants that define voice cycles is a very complex task. These cycles are used to obtain the fi frequency instants. Furthermore, it is vitally important to calculate these instants in the acoustic parameterization, as this is the cornerstone of voice characterizations of this kind.

Jitter [2] is a parameter representing variation of fundamental frequency, that is, the variations of pitch in each voice cycle. On the other hand, specialists also usually employ the shimmer parameter [2], which represents variation in width of voice cycle peaks. The voice produced through larynx modulation is able to almost constantly maintain peak width of voice periods. Therefore, an increase in shimmer value can be a symptom of voice disorder. Tables 1 and 2 present the various mathematical definitions of the jitter and shimmer objective parameters.

As previously mentioned, a number of authors have written several works on the detection of voice cycles [3,4] and there are also many highly detailed techniques to be found in the corresponding literature, such as estimators in the temporary domain (ratio of crosses per zero [5]), estimators of fundamental frequency [6,7], self-correlation methods (Yin estimators [8]), representation of the phase space [9], Cepstrums [10] and statistical methods [11, 12, 13]. Some of these directly define voice cycles [3], whereas others use numerical approximations [8] in order to obtain fundamental frequency values. In that respect, another step must be taken if we are to clearly identify the instants that define voice cycles.

However, none of these works were tried out on oesophageal voices and, what is more, it can be stated without a shadow of a doubt that these algorithms are not suitable for voices of this kind. The software pack presented here is a tool designed for use by specialists in otolaryngology, and is specifically designed to obtain objective voice parameters with excellent precision. The tool contains a basic algorithm to calculate the acoustic parameters related to speech periodicity and serves as an aid for not only diagnosis and rehabilitation but also for monitoring the patient.

It can be concluded that the tool is user-friendly and that ORL specialists can use it for measuring such objective parameters as pitch, jitter and shimmer, as well as for keeping patient records on these parameters.

2.3 Software interface

Speech signal processing plays an important role within the digital processing projects and investigations. Within this field, the esophageal voices are being objective of analysis and transformation [2],[3] but these have the limitation of measuring their quality only with subjective criteria as hearing tests. This is because an evaluation based on the calculation of objective parameters like pitch, jitter, shimmer or the harmonic to noise ratio HNR demands a high precision in the definition of the beginnings and ends of cycle in the voice signal.

The oesophageal voice is generated using the air pass across the oesophagus but without the modulation possibility by the vocal fold because they have been removed due to,

generally, a larynx cancer. Because of this their time-spectral characteristics are atypical and include levels of noise, fundamental frequency asymmetry and formant unstructuration. This leads to wrong measures in commercial applications and therefore is impossible to assess the quality of oesophageal voices. The same is applicable to voices with severe pathologies.

In this sense, it is necessary to develop an algorithm for the exact calculation of the marks that correspond to each cycle of the signal of oesophageal or pathological voices so that the calculation of pitch is exact and, with it, the measures of jitter, shimmer or signal to noise ratio. This algorithm has been included in a software interface for allowing users to measure and to plot in a graph the results of the acoustic parameters of the speech signal. This is suitable for evaluating and comparing the results between original oesophageal speech signal and the processed one after applying the wavelet transform.

3. System design

The system design has been divided into two parts: the algorithm for improving the quality of oesophageal speech using wavelet transform and the user interface including the speech signal processing using that algorithm and the acoustic analysis of speech parameters.

3.1 Algorithm using wavelets

As it has been previously mentioned, wavelet packets will be used in order to detect formants location. The reason of the choosing of this technique is their ability to separate the speech signal in different subbands, allowing to separate the formants bands quite exactly.

The here proposed method makes use of a double analysis. Firstly, a general analysis is applied over the whole spectrum, in this step a band in which the formant is located is approximated, and secondly the exact formant location is determined more accurately, the formant location accuracy can be adjusted through introducing more analysis levels inside the formant approximation band.

The main advantage of this method is the possibility of achieving a great frequency resolution, without consuming excessive computational resources, which is crucial when implementing the algorithms in a real-time device, such as a DSP.

3.1.1 Step 1: Band approximation of formants location

The first step consists of a rough analysis of the signal's wavelet packets tree. In order to locate formants frequencies, the energy of each subband is analyzed. The maxima of this energy signal determine formants location. The scheme of the process is shown in Figure 1.

Firstly, the wavelet packets tree is calculated up to the desire level, the chosen level is calculated taking into account the sampling frequency and the resolution required.

After having obtained the wavelet packets, the energy of each last level node is calculated. The Energy is stored in an array and its envelope is estimated. This envelope smoothes the energy signal and thus, the maxima can be easily calculated.
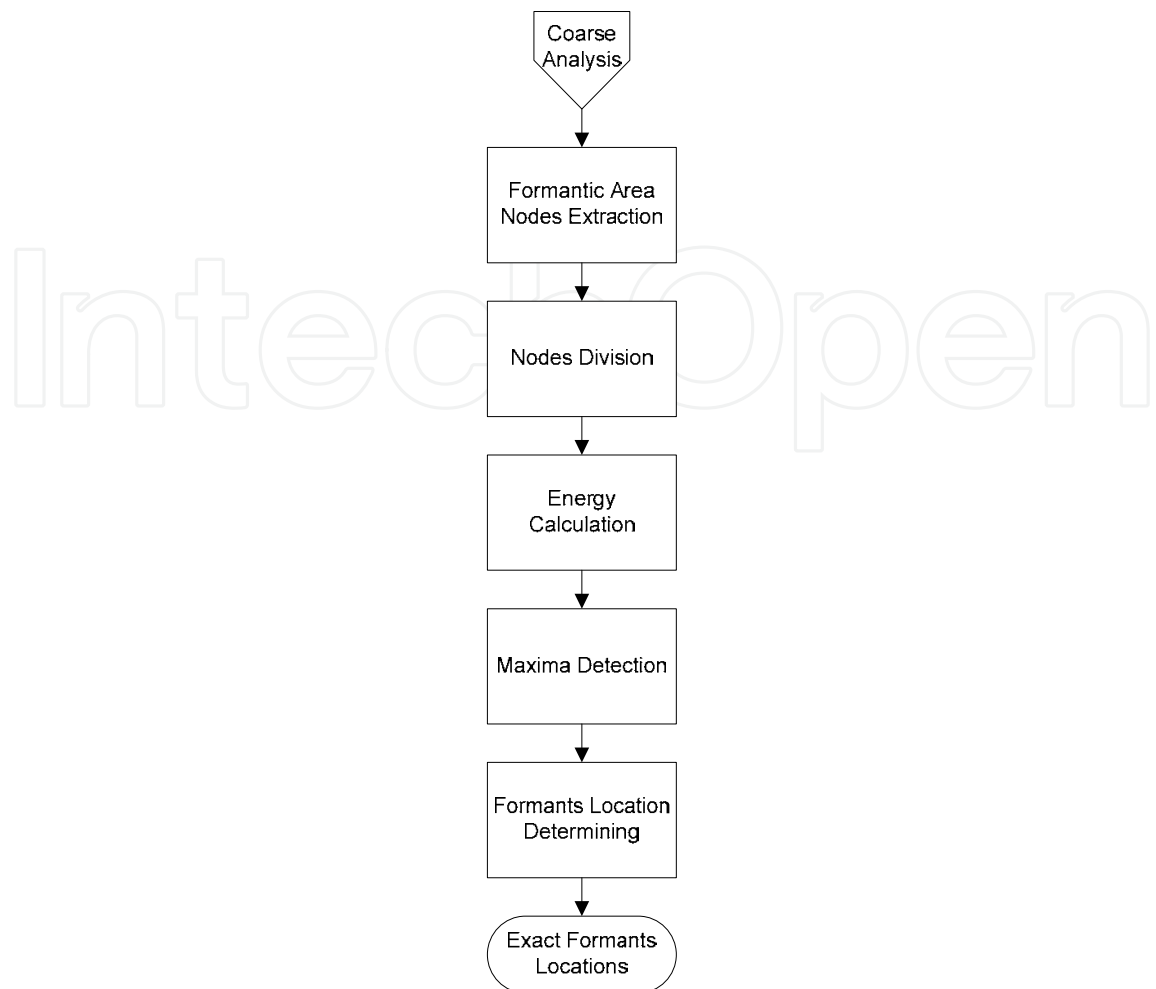
```
┌─────────────┐
│   Coarse    │
│  Analysis   │
└─────────────┘
       │
       ▼
┌─────────────┐
│Formantic Area│
│Nodes Extraction│
└─────────────┘
       │
       ▼
┌─────────────┐
│Nodes Division│
└─────────────┘
       │
       ▼
┌─────────────┐
│   Energy    │
│ Calculation │
└─────────────┘
       │
       ▼
┌─────────────┐
│  Maxima     │
│  Detection  │
└─────────────┘
       │
       ▼
┌─────────────┐
│Formants Location│
│ Determining │
└─────────────┘
       │
       ▼
┌─────────────┐
│Exact Formants│
│  Locations  │
└─────────────┘
```

Fig. 1. Adjustable Resolution Analysis Schema

3.1.2 Step 2: Adjustable resolution analysis

In the above explained step, an approximation to the formantic frequencies was obtained. As it will explained in next head, the resolution obtained with this approximation, though it is better than the one obtained with conventional methods, may not be enough for some environments.

In order to achieve a finer resolution, an adjustable resolution analysis was designed. The scheme of this analysis is shown in Figure 1. The core idea of the designed technique is to obtain a higher resolution in the previously detected bands by dividing the selected nodes and their adjacent ones.

The main reason of using narrower bands is that energy in wavelets packet spreads among various adjacent nodes, the solution to this problem is to divide the spectrum in such narrow bands that the energy of the formant locates in only one node.

As it can be seen in Figure 1 the first step of the algorithm consists on splitting the approximated formantic bands and their adjacent as many times as necessary. Secondly, the energy of each node is calculated again and the maximum value located, this value indicates formant location.

The main advantage of this method is that it is possible to save a lot of computational load but preserving a high accuracy level at the same time. For example, if an 8 level basic tree is to be taken and its formantic nodes are expanded two levels, it is possible to obtain a 10 level resolution by consuming an 8 level computational load.

3.2 User interface

Using the advantages of the previously described algorithms, authors have developed a tool called "PAS Voice". The welcome screen will then be displayed:

Fig. 2. Welcome Screen

Once the application has been started up, the main screen will be displayed:
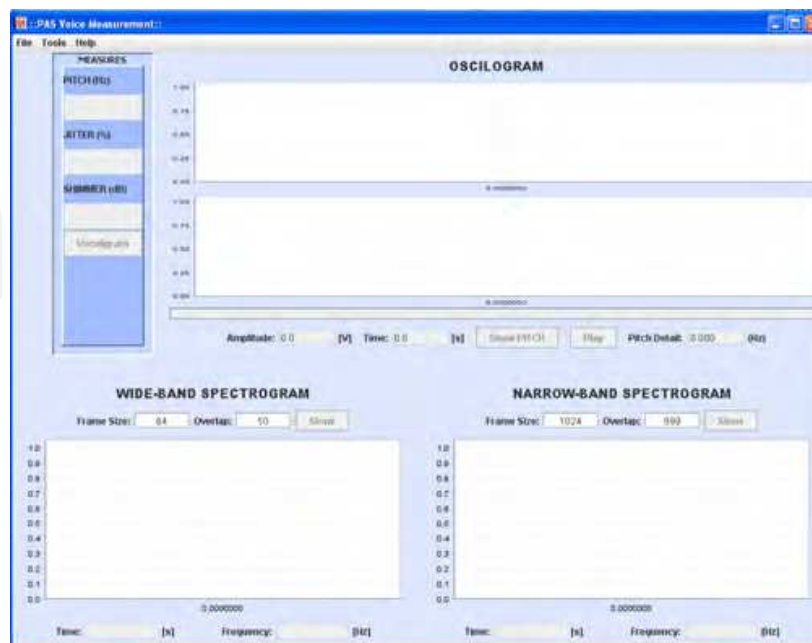
Fig. 3. Main Screen

The following areas can be observed on this screen:

1.   Menu: The program's general option menus can be identified in this area:

File.- Menu with the "Open file" option, which allows you to open a voice signal in order to process it. The signal has to be in .WAV, .AU or .AIFF format. Voice processing begins automatically once the file to be analyzed has been chosen.

Save Results – This enables you to save the signal processing results; results from several sessions can be added for the same person or a new profile can be created. Once the results have been correctly saved, a graph will be displayed showing the evolution of the parameters throughout all sessions of analysis. When this graph is closed, an informative message on development since the previous session will be displayed.

Tools.- Tool menu for application configuration.

Language.- This allows the language to be chosen for the program (initially English and Spanish, although personalized translations can be applied). If the language is changed, the application will have to be rebooted.

Octave Path.- The octave.exe file, essential for the running of this program, can be specified using this option.

Help.- By clicking on this, help is provided for running the program.

2.   Measurement area: In this area, once the a voice signal has been processed (through the File/Open file option), the numerical measurements of Pitch, Jitter and Shimmer are displayed. If one wishes to observe the measurements in graphic form with the normality threshold, the "Vocaligram" button can be clicked on; this will only be enabled once the processing has been performed to obtain the measurements needed to create the vocaligram. The vocaligram is a graphic representation of a measurement in each axis (in blue) superimposed over the threshold values for each parameter. The measurements are scaled so that abnormal values are always greater than the threshold (a value above the threshold implies that it is abnormal).
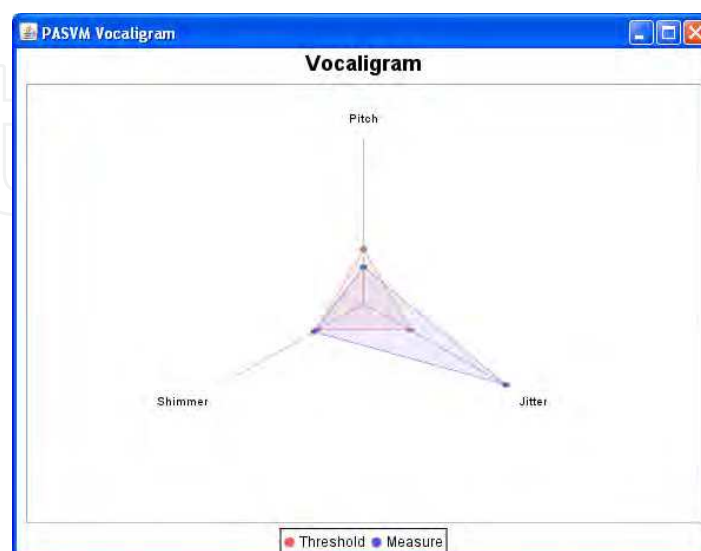

Fig. 4. Sample Vocaligram

3. Graphic Representation Area: In this area, the voice signal graphics and the evolution of pitch over the time are displayed once the voice has been analyzed. Underneath is a progress bar indicating the approximate percentage of analysis completed.
4. Graphic Representation Options Area: Once the voice and pitch evolution have been represented, this area is enabled so as to be able to check other data in greater detail:

Amplitude/Time/Pitch Detail.- When a particular point in the graphics above are clicked on, these frames fill up with information corresponding to the point that has been selected. The Amplitude/Time values (clicking on the upper one) or Pitch Detail values (clicking on the lower) will be displayed in accordance with the graphic function selected. Show Pitch.- Once the pitch of the voice signal has been calculated, when this button is pressed the marks situated in the signal over the relevant points indicating periodicity will be shown.

Play.- By pressing this button, the voice signal will be reproduced through the computer loudspeakers (if applicable).
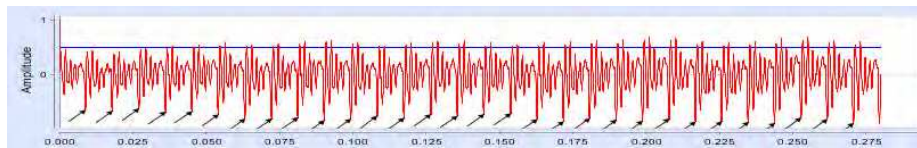
Fig. 5. Example of oscilogram with "Show Pitch" option activated

5. Spectrogram Options Area: By default, the spectrograms are not calculated during the analysis process. If it is wished to do so, this should be done through the following area:

Frame Size/Overlap.- These are the parameters composing the spectrogram. The parameters indicated by default are typical ones for the representation of broad-band and narrow-band spectrograms respectively. Beware! It is not recommended to touch these parameters … A poor configuration may considerably increase spectrogram calculation time.

Show.- This displays the spectrogram in its area corresponding to the indicated parameters.
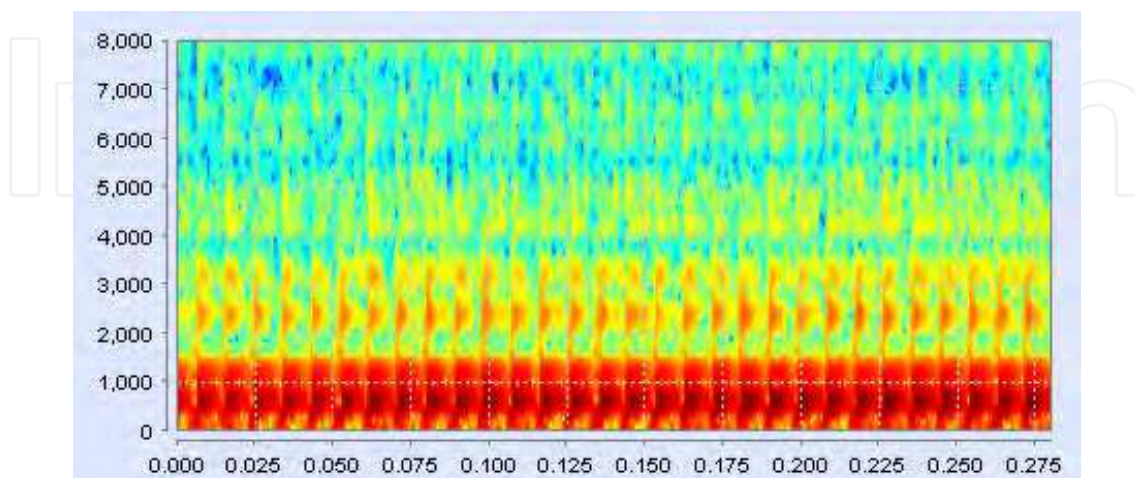

Fig. 6. Example of Broad-Band Spectrogram

6. Spectrogram Area: This shows the spectrograms when the "Show" button is pressed.

4. Results

Tables 1 and 2 show the measurements of the first formant location for healthy (left) and esophageal (right) voices. Tables 3 and 4 show the measurement errors absolutely and relatively, the relative value is calculated comparing the obtained error with the average formant value. As it can be seen in those tables, conventional methods obtain very poor results, achieving an average deviation of about 70 Hz, approximately the value of the pitch in esophageal speech. These deviations could be inappropriate for some applications which require great accuracy, thus a new measurement method is necessary.

A simple wavelet algorithm with approximation to the formant band improves considerably this results reducing the deviation about a 30%. This represents quite an improvement comparing with LPC, but it is possible to obtain higher resolution without increasing substantially computational costs. The results of the adjustable resolution algorithm show that it is possible to reduce the average deviations up to a 50%.

The obtained values prove that it is feasible to locate formants position with minimum errors and effective algorithms. This fact constitutes a fundamental advance in esophageal speech regeneration, because formant location has great importance in many speech processing algorithms. Taking as an example previous works of the research group, for example for such as an algorithm as the one presented in [2], much better results would be obtained with more accurate formant location estimations.

It is important to highlight the great relevance that this results may have in some other speech technologies fields such as speech recognition, etc. So the applications of this analysis is not restricted to esophageal speech processing but can be implemented with many others purposes.

Speech Signal	Original Values (Hz)	F. with LPC (Hz)	F. with B.A (Hz)	F. with R.A. (Hz)
He. 1	851	842	883	848
He. 2	776	633	711	756
He. 3	938	893	969	950
He. 4	960	929	926	966

Table 1. 1st Formant location for **healthy voices** calculated with different methods: LPC, Band Approximation (B.A.) and Resolution Adjustement (R.A.).

Speech Signal	Original Values (Hz)	F. with LPC (Hz)	F. with B.A (Hz)	F. with R.A. (Hz)
Es. 1	894	698	883	890
Es. 2	830	762	754	778
Es. 3	808	774	754	805
Es. 4	776	744	754	756

Table 2. 1st Formant location for **esophageal voices** calculated with different methods: LPC, Band Approximation (B.A.) and Resolution Adjustement (R.A.).

Speech Signal	Deviations obtained with LPC (Hz)			Deviations obtained only wit band approximation (Hz)			Deviations obtained with resolution adjustment (Hz)		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
Healthy 1	10	31	185	31	9	159	4	4	17
Healthy 2	143	12	78	65	31	41	20	9	10
Healthy 3	45	68	157	31	106	30	12	34	16
Healthy 4	31	15	10	34	12	35	6	12	5
Esophageal 1	196	77	72	11	33	35	4	26	75
Esophageal 2	50	193	101	76	77	2	42	48	11
Esophageal 3	34	88	51	54	33	2	3	15	10
Esophageal 4	32	21	60	22	0	13	20	9	21
Average Deviation	65	61	89	41	36	40	18	23	21

Table 3. Deviations obtained in formants location values with different methods.

Speech Signal	Deviations obtained with LPC (%)			Deviations obtained only wit band approximation (%)			Deviations obtained with resolution adjustment (%)		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
Healthy 1	1.135	2.292	7.000	3.518	0.665	6.016	0.454	0.296	0.643
Healthy 2	16.227	0.887	2.951	7.376	2.292	1.551	2.270	0.665	0.378
Healthy 3	5.106	5.028	5.941	3.518	7.837	1.135	1.362	2.514	0.605
Healthy 4	3.518	1.109	0.378	3.858	0.887	1.324	0.681	0.887	0.189
Esophageal 1	23.700	5.487	2.580	1.330	2.352	1.254	0.484	1.853	2.687
Esophageal 2	8.222	13.754	3.619	9.190	5.487	0.072	6.288	3.421	0.394
Esophageal 3	4.111	6.271	1.827	6.530	2.352	0.072	0.363	1.069	0.358
Esophageal 4	3.869	1.497	2.150	2.660	0.000	0.466	2.418	0.641	0.752
Average Deviation	7.964	4.359	3.306	4.747	2.632	1.486	1.790	1.781	0.751

Table 4. Percentual deviations obtained in formants location values with different methods.

After having applied the wavelet transform to the oespphageal speech signal, we can measure the final value of the acoustic parameters. Below is an example describing the basic operation of the software that authors have develop named "PASVoice software pack". It

analyses a speech signal in order to obtain objective parameters and graphic representation of values for helping doctors to understand the patient's stage,

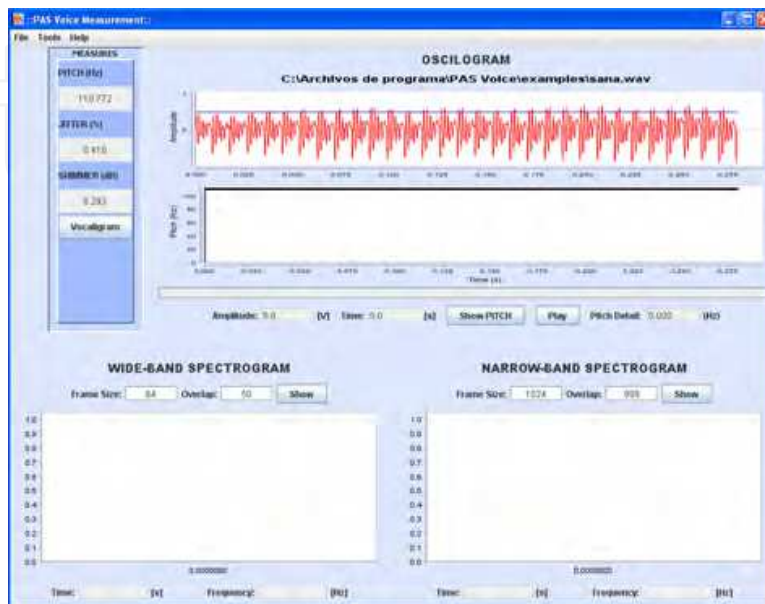When appliying over a healthy voice the results can be as follows:

Fig. 7. 'sana.wav' Results

If 'Show Pitch' is pressed/selected, we can observe the marks that have been located as a reference for measuring pitch:
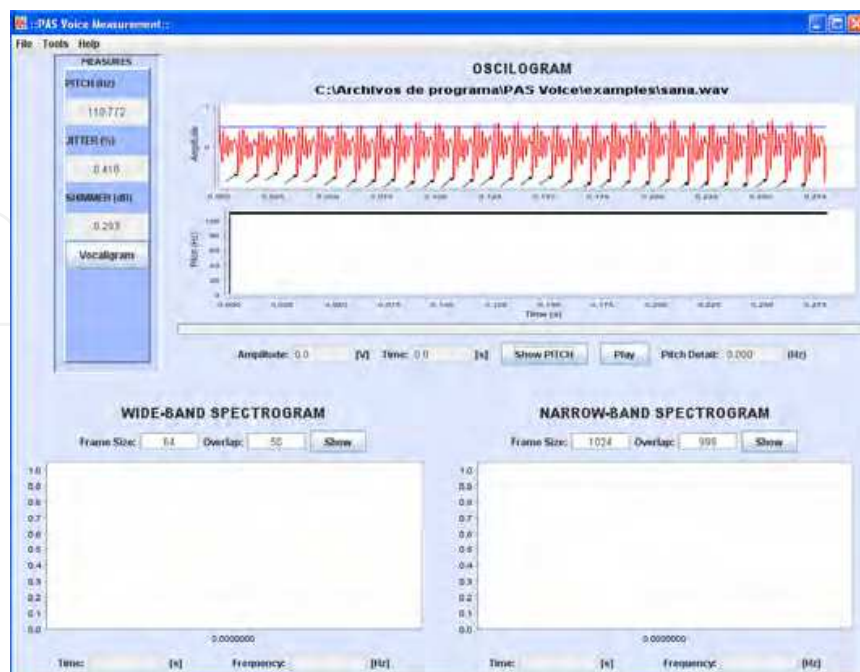

Fig. 8. Details of pitch marks

When the 'Vocaligram' button is clicked on, we can see the same (in this example the results are below the threshold for each parameter, as expected):
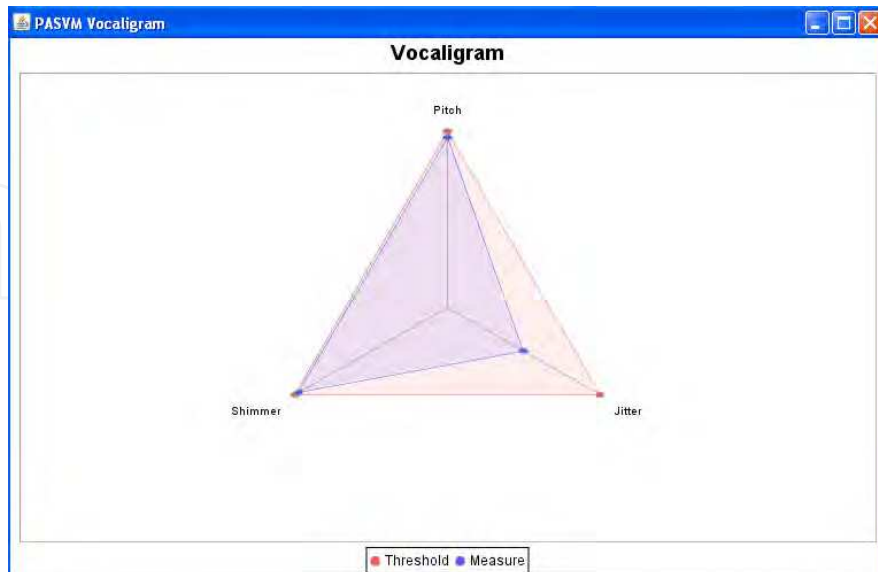
Fig. 9. 'sana.wav' Vocaligram

If any of the spectrogram 'Show' buttons are selected, the corresponding spectrograms are automatically calculated and visualized:
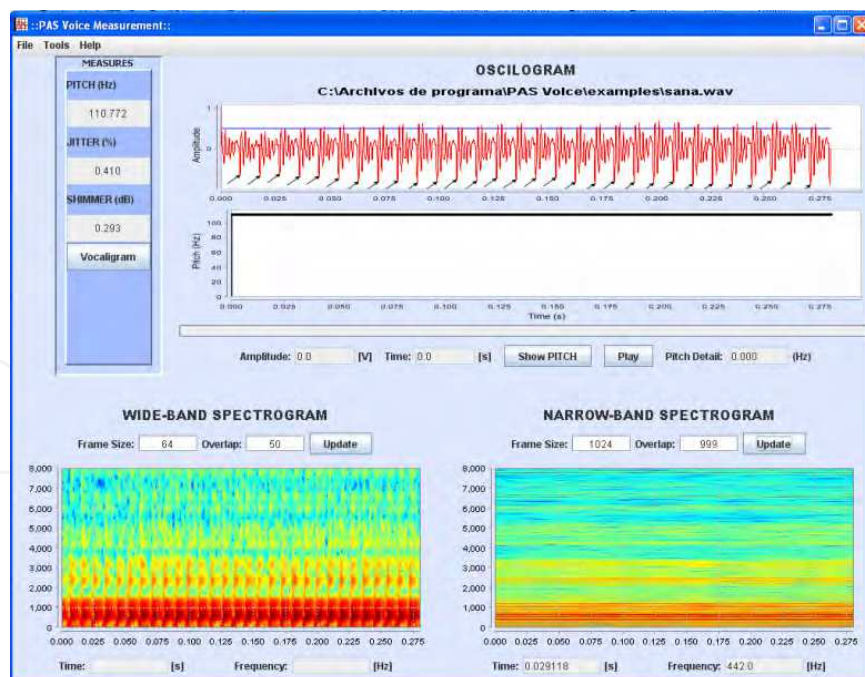
Fig. 10. Details of the application with both spectrograms calculated

Finally, if it is wished to save the numerical results, the File/Save Results option can be chosen, after which the following dialogue appears. As can be seen, it contains data corresponding to other patients:
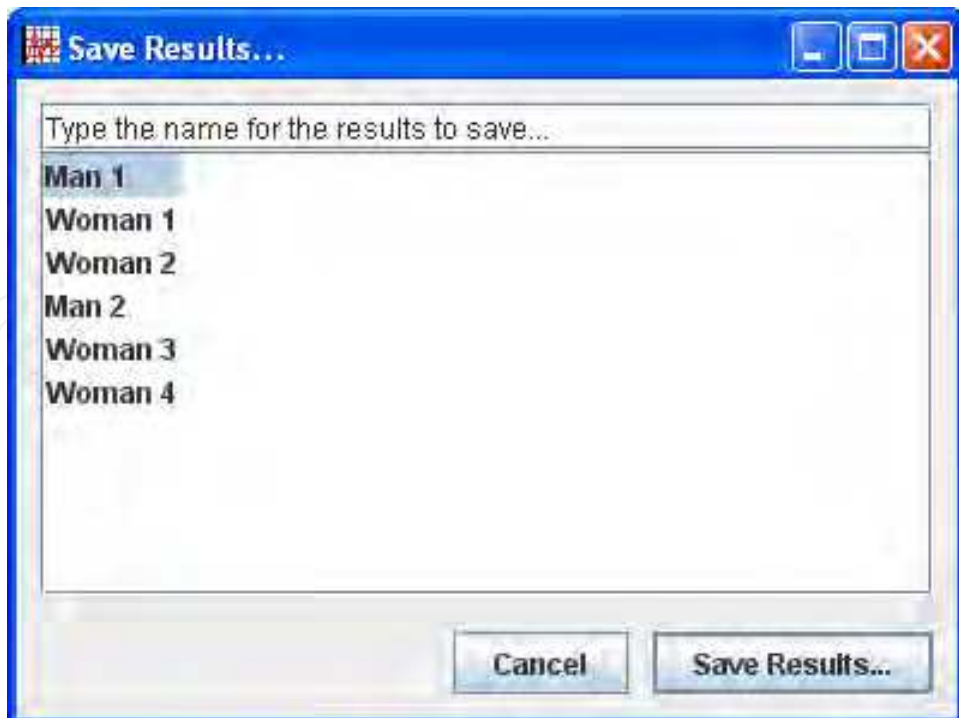
Fig. 11. Dialogue box used to save results

If this is not the first session for the person we are dealing with, his/her name can be searched for by typing the first letters of the name in the box at the top. All concurrences, if there are any, will then be displayed in the main box below.
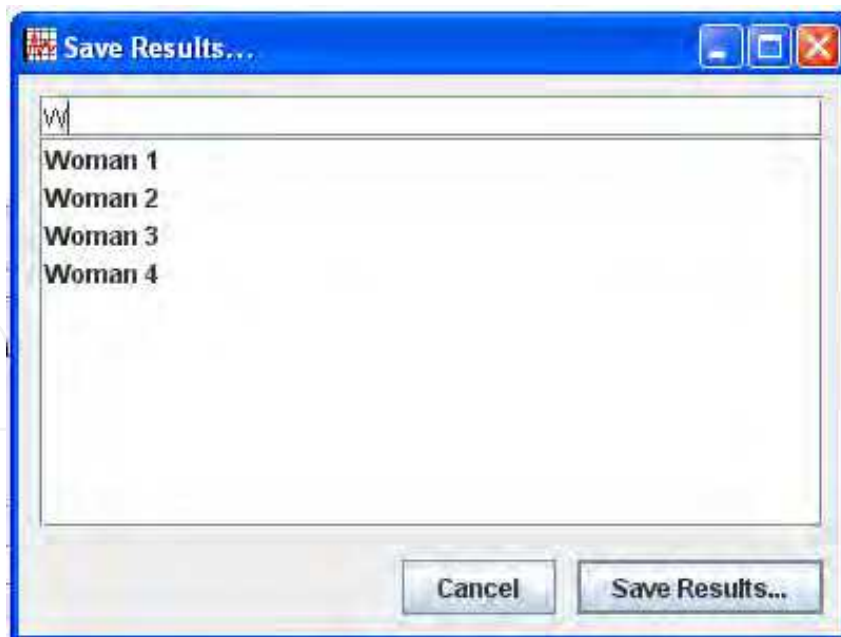
Fig. 12. Search for people whose names begin with 'w'

In our case we are going to create a new profile. As the name "Example" does not exist, by typing it out completely and clicking on "Save Results…", the new name will be created and the data saved. No results will appear as this is the patient's first session:

Fig. 13. Message displayed when saving results
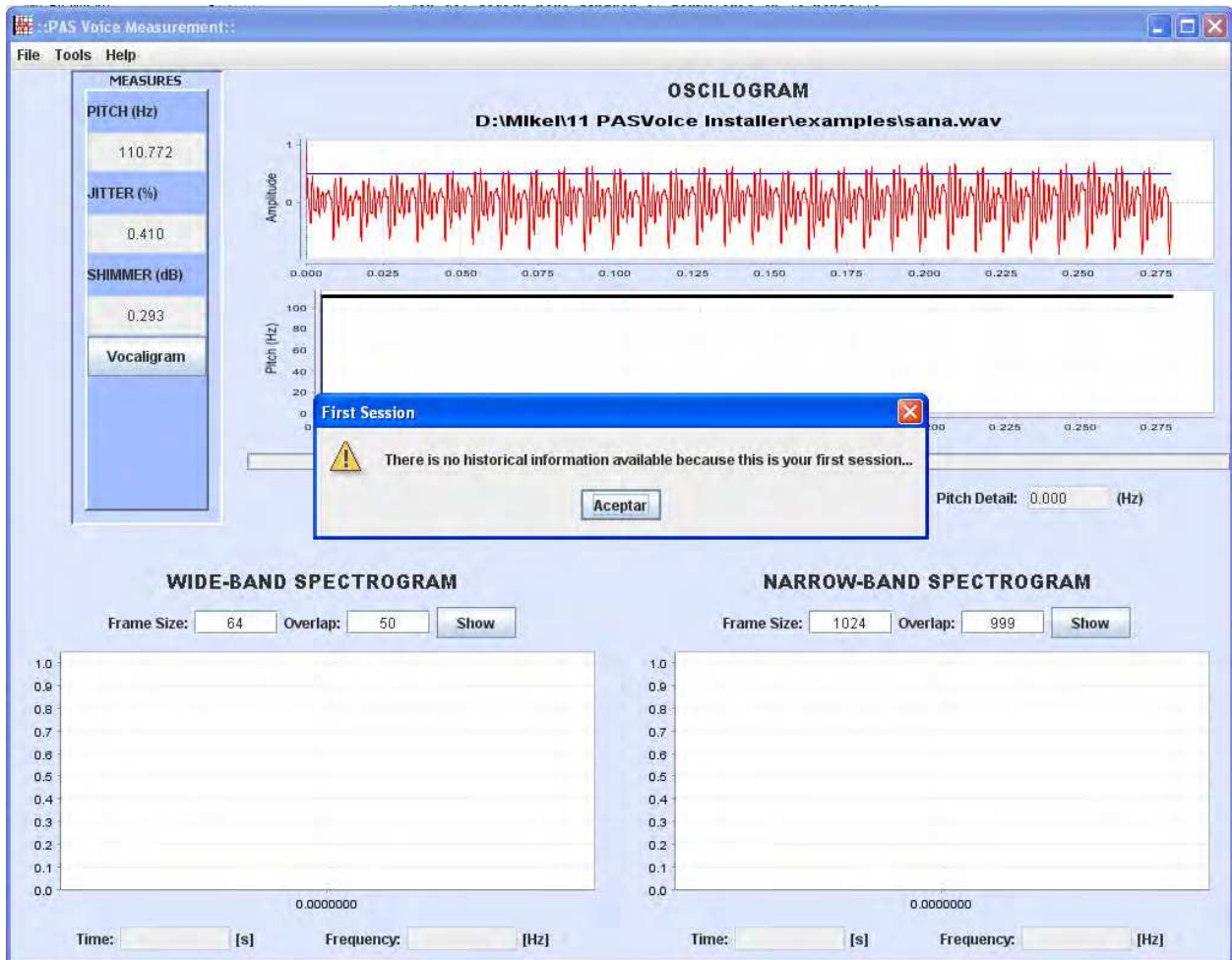
Fig. 14. Message displayed at initial session

We could also have added results as if they were for a patient not coming for the first time. We choose an already existing patient, "Man 1", by choosing from the list and clicking on "Save Results…". A graph showing all the results saved to date from previous sessions is provided (pitch information is separated from that on jitter and shimmer as they are different units):
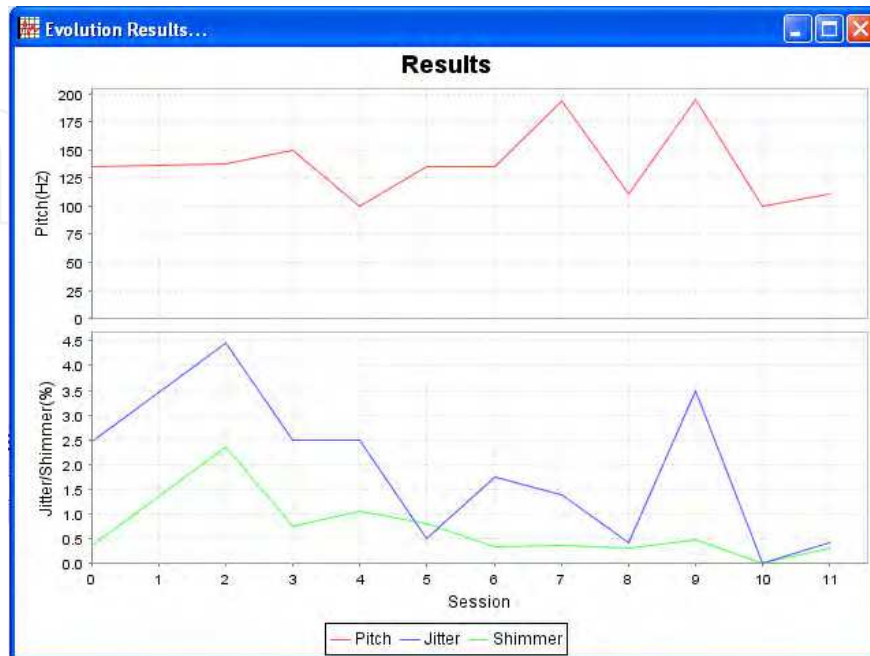
Fig. 15. Evolution of results by session

Finally, when this window is closed (top-right x), we are provided with a message informing us on evolution since the previous session:
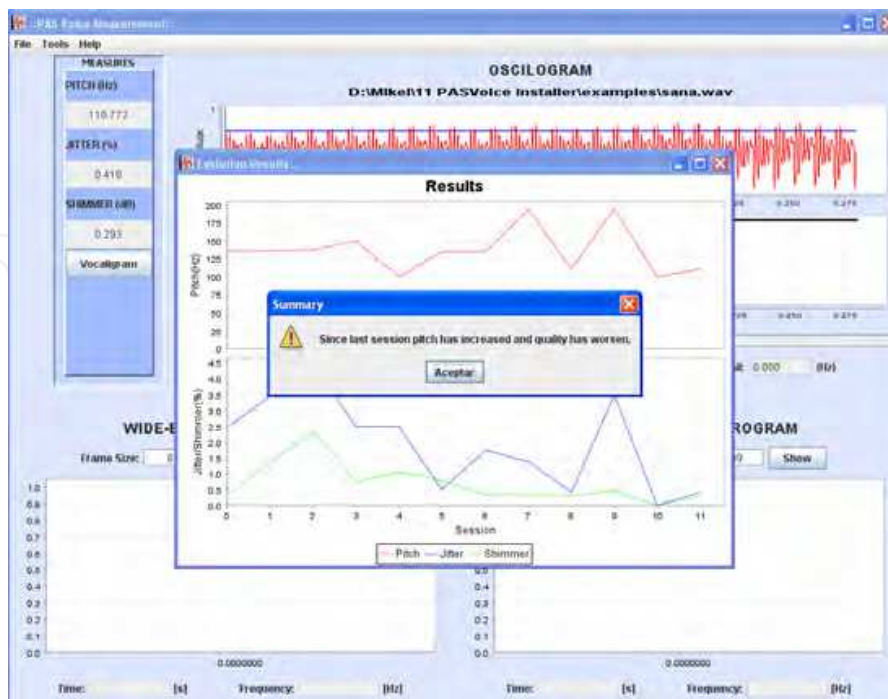
Fig. 16. Message providing information on evolution since last session

Recent scientific progress has made it possible to take great steps forward in such fields of major interest as biomedical engineering. In this area, the application of new technologies becomes essential in order to improve techniques in the diagnosis, treatment and rehabilitation of certain medical pathologies. However, there are also collectives suffering from an illness or treatments that only affect a minority of people. This is a characteristic which usually implies that the level of technological development corresponding to the resources having to be used for these pathologies is way behind that for other more common disorders.

The laryngectomized are people who, for various reasons, have had to undergo surgery to remove their larynx, vocal cords, epiglottis and the cartilages surrounding the larynx. These elements are of vital importance for the generation of speech as they form part of the phoning apparatus. Therefore, the removal of these seriously affects the quality of their speech.

The issue of treating a barely intelligible voice is also of great use from the point of view of the patient's psychology. We have noticed that a high proportion of the laryngectomized feel embarrassed when using this voice, particularly women, who would rather not speak than do so with oesophageal voice, as they consider it unfeminine.

The results obtained from this research work have been useful mainly due to the IT contribution involving the design, development and implementation of a software application specifically intended for the assessment of laryngectomized voices, with a view to performing a correct medical monitoring that will make it possible to measure evolution and prevent relapses. In order to verify improvement in the quality of oesophageal voices, a database containing several phonemes of all kinds of voices was worked with; these voices, both pathological and healthy, were recorded with the help of members from the Asociación Vizcaína de Laringectomizados.

Future work deriving from this research includes, most importantly, the incorporation of functionalities for vocal recognition and synthesis of phrases, as well as implementing the digital signal processing algorithms developed in systems based on cell phones and PDAs; all this with the goal of improving the laryngectomized's quality of life.

5. Conclusions

Due to the great relevance of Wavelet Transform for the analysis and processing of esophageal speech, and assuming that the final goal will be the implementation in a hardware DSP based device, with very strict real-time requirements, a significant computing resources optimization has been achieved, and consequently, a reduction of the code length in order to minimize computational load. Also it is important to highlight that the obtained wavelet calculi can be used in later processing.

These advantages are achieved through a preprocessing algorithm, which, although Wavelets based, includes some improvements. Firstly, an approximation to the formant subband. And secondly, an adjustable resolution applied over the bands among which the formant energy is shared.

On the other hand, the here proposed algorithm allows to optimize previous research works concerning the treatment of the poles of the system which models esophageal speech, according to LPC. Taking into account the obtained accuracy, it is logical to assume an improvement in results if this technique is used as a first stage of the whole algorithm.

6. References

[1] García, B., Vicente, J. & Aramendi, E. "Time-Spectral Technique for Esophageal Speech Regeneration" Biosignal '02, 2002.

[2] García, B., Vicente, J., Ruiz, I., Alonso, A. & Loyo, E. "Esophageal Voices: Glottal flow Restoration" ICASSP 2005.

[3] Baken, R. & Orlikoff, R. "Clinical measurement of speech and voice" Second Edition. San Diego, CA: Singular Publishing Group, ISBN:1565938690, 2000.

[4] Brown, J.C. & Puckette, M.S. "A high resolution fundamental frequency determination based on phase changes of the Fourier transform" J. Acoust. Soc. Amer., vol.94, pp. 662-667, August 1993.

[5] Mallat, S "A Wavelet Tour of Signal Processing" Second Edition.. Academic Press, ISBN: 0-12-466606-X.

[6] Hooper, C.R. "Using evidence-based research in speech-language pathology: a project that changed my thinking"American – Speech – Language – Hearing -Association, January 2003.

[7] Cnockaert, L.; Grenez, F.; Schoentgen, J. "Fundamental Frequency Estimation And Vocal Tremor Analysis By Means Of Morlet wavelet Transforms", ICASSP 2005.

[8] J. K. Maccallum, L. Cai, L. Zhou, Y. Zhang, and J. J. Jiang, "Acoustic analysis of aperiodic voice: perturbation and nonlinear dynamic properties in esophageal phonation.," Journal of Voice, vol. 23, no. 3, pp. 283-90, May 2009.

[9] M. Carello and M. Magnano, "A first comparative study of oesophageal and voice prosthesis speech production," EURASIP Journal on Advances in Signal Processing, vol. 2009, no. 821304, pp. 1-7, 2009.

[10] D. Rudoy and T. Quatieri, "Time-varying autoregressions in speech: detection theory and applications," IEEE Trans. Audio, Speech, and Language Process., vol. 19, no. 4, pp. 977-989, 2011.

[11] S. Aviyente and A. Yener Mutlu, "A time–frequency based approach to phase and phase synchrony estimation," IEEE Transactions on Signal Processing, vol. 59, no. 7, pp. 3086-3098, 2011.
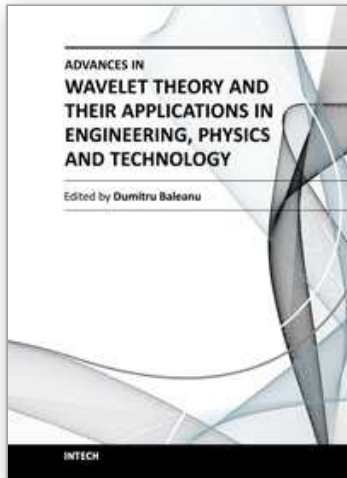
Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology

Edited by Dr. Dumitru Baleanu

The use of the wavelet transform to analyze the behaviour of the complex systems from various fields started to be widely recognized and applied successfully during the last few decades. In this book some advances in wavelet theory and their applications in engineering, physics and technology are presented. The applications were carefully selected and grouped in five main sections - Signal Processing, Electrical Systems, Fault Diagnosis and Monitoring, Image Processing and Applications in Engineering. One of the key features of this book is that the wavelet concepts have been described from a point of view that is familiar to researchers from various branches of science and engineering. The content of the book is accessible to a large number of readers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH

open science | open minds