Oesophageal voice acoustic parameterization by means of optimum shimmer calculation

B.GARCÍA, I.RUIZ, A.MÉNDEZ AND M.MENDEZONA
Department of Telecommunication
University of Deusto
Avenida de las Universidades 24, 48007, Bilbao
SPAIN
{mbgarcia, ibruiz, amendez}@eside.deusto.es, mmendezo@tecnologico.deusto.es
http://www.pas.deusto.es/

Abstract: - This paper presents an algorithm which works with oesophageal voice and aims to automatically calculate different measurements of shimmer parameter. The evaluation has required a novel processing for the detection of the periodicity cycles and instants in order to calculate amplitude perturbations of vocal signals. The designed algorithm performs the calculusof pitch in an automatic way with a mean maximum inaccuracy of 0.46% for oesophageal voices which is a novelty. The shimmer calculation process has been optimized in order to obtain a higher variation range in pathological cases. Thus, it will allow using such parameter as reference in diagnosis and rehabilitation processes, reaching a 51,741% of shimmer increment for oesophageal voices.

Key-Words: - Oesophageal voice, Acoustic signal analysis, Acoustic measurements, Amplitude perturbation

1 Introduction

Nowadays, there is a technological lack in the field of software tools that can be used by professionals working in treatment and rehabilitation therapies with laryngetomees. These people got their vocal folds removed due to a larynx cancer, so they have to learn how to speak with the oesophagus and must follow a learning process. Previous works (see [1]) in this area show that one of the objective parameters which reflect an improvement in voice quality is "shimmer" or the percentage variation of the peak-to-peak amplitude. Its measurement depends on an accurate detection of the voice periodicity instants and, with the algorithmic techniques available at present, those values cannot be calculated correctly in an automatic way. As stated by Chen and Kao in [2] few papers have focused in periodicity marking and although some works have been proposed in order to locate periodicity instants (for example [2] and [3]), none of them have been tested with oesophageal speech. In fact, the characterization of this kind of pathological voice is a field that has not been developed too much up to now.

In this paper a solution to this issue is presented, by means of an algorithm which obtains more accurate results than widely known applications. In the other hand, the algorithm could be also used by professionals to evaluate the degree of improvement in patients' volume control.

2 Objectives

The present research work is focused on two fields: medical and technical sciences. The main medical objective is:

- To help therapist in the education and learning stage of oesophageal voice.

On the other hand the following technical objectives can be defined:

- To optimize the accurate calculation of the shimmer by means of the exact detection of the periodicity cycles of voice signals.
- To optimize shimmer definition in order to obtain a higher variation range in abnormal voices.

To design an automatic algorithm in order to avoid therapists' manual intervention when performing objective measures.

3 Procedures and Methods

In this section some concepts will be presented. Some of them are used to characterize voices and measure accuracy while others represent the mathematical background employed in the design and development of the algorithm explained in this paper:

3.1 Shimmer generalities

Shimmer is the parameter which represents the amplitude perturbation of the voice signal. The voice produced in vocal folds is supposed to have the ability to maintain its amplitude almost constant, thus an increased value of shimmer may imply a symptom of a voice disorder. Possible shimmer definitions (see [4]) are:

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A^{(i+1)} - A^{(i)}\right|}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \qquad (1)$$

$$ShimdB = \frac{1}{N-1}\sum_{i=1}^{i=N-1}\left|20\log\left(A^{(i+1)}/A^{(i)}\right)\right| \quad (2)$$

$$sAPQ = \frac{\frac{1}{N-sf+1}\sum_{i=1}^{N-sf+1}\left|\frac{1}{sf}\sum_{r=1}^{sf-1}A^{(i+r)} - A^{(i+m)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \quad (3)$$

being A the peak-to-peak amplitude data, N the number of extracted period marks, sf the smoothing factor (usually odd) and m=(sf-1)/2.

3.2 Shimmer characterization procedure

The shimmer calculation procedure was not discovered easily because the peak-to-peak data (A) defined in equations (1),(2) and (3) is so ambiguous and the software used for result checking purposes showed confusing measures.

First of all, it was discovered that the definition used by the software used in the test stage corresponds with that referred in equation (2).

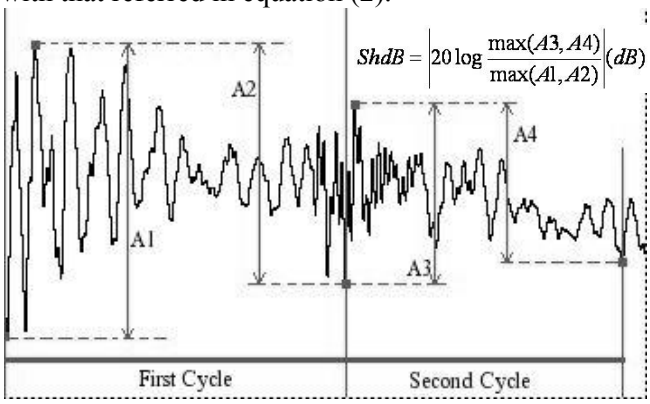

Fig.1: Shimmer calculation example

Finally, after intensive tests, it was deducted that the amplitude value could be defined as the maximum between the possible peak-to-peak values within a cycle as it can be seen in the previous figure so the equation (2) can be rewritten as:

$$ShdB = \sum\left|\log\left(\frac{\max(A_{i+1}|A_{i+1}')}{\max(A_i|A_i')}\right)\right| \quad (4)$$

As the minimum value can be taken from the period marks that define the beginning and end of each voice cycle, there are two possible values (i.e. A1 or A2 in the first cycle of the Fig.1 or A and A' in equation 5), and using the maximum lead to minimize the numerical value of the Shimmer.

The definition used in eqn. (4) led initially to results that were considered satisfactory as it can be seen in results section. Then further tests were to be performed as the aim was to increase somehow the shimmer ranges of pathological and especially oesophageal voices in order to use shimmer as reference in medical diagnosis.

The hypothesis was that eqn. 4 tends to obtain a lower shimmer value as the difference between the numerator and denominator becomes minimal. If the numerator in (4) is changed, taking the minimum instead of the maximum, the difference grows, increasing the shimmer value. In this sense, the shimmer increment would be more noticeable in the cases of pathological and oesophageal voice since healthy voices have less amplitude perturbation so the introduced change wouldn't affect its value in a significant way:

$$ShdB = \sum\left|\log\left(\frac{\min(A_{i+1}|A_{i+1}')}{\max(A_i|A_i')}\right)\right| \quad (5)$$

Apart from the defined hypothesis, the last combination with both minimal numerator and denominator is also considered in order to analyze all possible combinations:

$$ShdB = \sum\left|\log\left(\frac{\min(A_{i+1}|A_{i+1}')}{\min(A_i|A_i')}\right)\right| \quad (6)$$

Thus, in the example of figure 1 the value of the shimmer as in Eqn. (4) would be A3/A1, as Eqn. (5) A4/A1 and as Eqn. (6) A4/A2. The discussion about the hypothesis will be presented later, in the results section of this article.

As it can be seen, in the present procedure for characterize the amplitude perturbation, it is of capital importance the accuracy of the cycle determination algorithm. Such algorithm suitable to work with pathological voices is also presented in the next section of the present article.

3.3 Sonority

It is defined as a kind of measure of the energy contained within a windowed part of the signal and is used as the reference to find the pitch instants:

$$Sonority(k) = \sum_{n=1}^{N}|B(k,n)| \quad (7)$$

being B(k,n) a matrix containing in each column the FFTs of N points of the k windows of the signal.

3.4 Cycle detection algorithm

This subsection details the algorithm developed to characterize the shimmer and the procedure to define

accurately voice cycles in speech signals including those that are damaged due to laryngeal diseases.

3.4.1 General Overview of the algorithm

The algorithm is an iterative process based in a negative peak-picking procedure. The core of the algorithm is a function that extracts the pitch instants (or marks) of the voice signal in order to calculate each pitch component. The iterations stand for an accurate parameterization of the function according to the type of the voice and the likely pitch range in which it is located which is estimated by means of the cepstral representation of the signal.

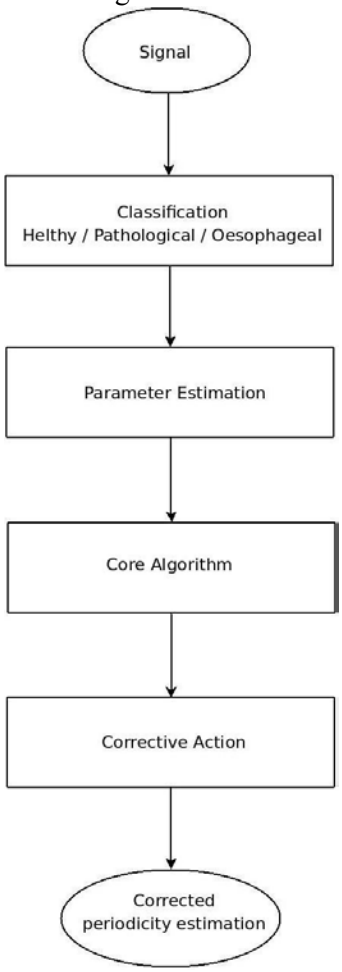

Fig.2: Block-Diagram of the algorithm

With the core explained before the algorithm shown in the previous diagram was developed.

3.4.2 Core Algorithm

The core algorithm (Fig.1, block 3) has been designed for a suitable pitch marks (or voice cycles) determination.

The key algorithm is based on the sonority measurement. Sonority is defined as a kind of measure of the energy contained within a windowed part of the signal (see (7) in the subsection 3.3) and it is used as the reference to find the pitch instants.

The core algorithm works as it is shown in Fig 3. Thus, the pitch extraction is done by means of a time domain algorithm which takes into account the energy (the sonority in this case) within each windowed frame. The higher the sonority within one frame, the higher the probability of containing a pitch instant on it.

In the Sonority analysis block of the Fig 3 the first step is to take only the negative side of the signal because it has been shown that the typical behaviour of the oesophageal voice is to fall to negative values with more energy.

Once this has been done, the signal is windowed and the Fast Fourier Transform (fft) of each frame calculated. After this, the sonority is calculated as in (10). The sonority is then filtered forward and backward using the following transfer function (given in the 'z' domain)

$$H(z) = \frac{1 + z^{-1} + z^{-2} + z^{-3} + z^{-4}}{5} \qquad (8)$$

for smoothing purposes.
Once the sonority is obtained and smoothed, its maximum values are obtained.

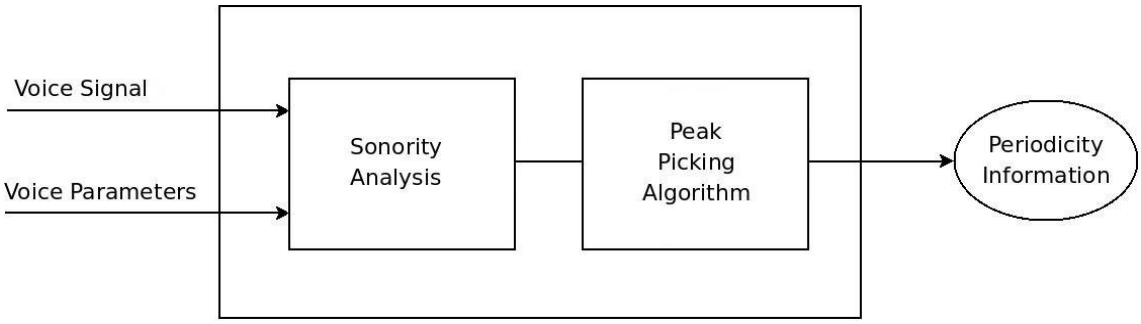

Fig.3: Core Algorithm's Block detail

With these maximum values the second part (Fig 3, block 2) of the algorithm is applied, the peak picking algorithm. Thus, once the maximums of the sonority are obtained, length and amplitude values of the array are adjusted in order to match them to original signal's features. Finally the signal obtained is flipped to put the maximums in the negative side.

The next step is to adjust each minimum to its real position on the original signal. To perform this a frame of the original signal is taken, each frame is centred on the position of each minimum obtained previously then, the absolute minimum of that frame is obtained and the resulting array is built inserting the value of each absolute minimum of each frame in its position and leaving the rest of values as zero.

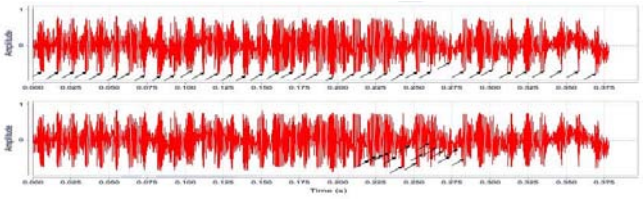

Fig.4:
a) Correct Voice Marks represented with arrows (up)
b) Erroneous calculation of voice marks imported from benchmark application (down)

The result of the present core algorithm is a vector with the information of each detected voice cycle (see Figure 4.a) from which the variation of the instantaneous frequency and pitch can be easily obtained.

3.4.3 Classification Block and Pitch Range Estimation Block

As it's shown in the general block-diagram (Figure 2) the core algorithm (with some base parameters, first block of the diagram) is applied to obtain pre-pitch which is a previous calculation on the signal. It provides a classification (List 1) of the voice signals in two categories: oesophageal or severely damaged voices (those with lower pitch and noisy voices) and healthy/pathological voices.

The pre-pitch is obtained using the core algorithm with experimentally estimated parameters (base parameters) which provide this classification feature.

```
if (prepitch < 110)
        Type="oesophageal"
else
        Type="healthy" OR "pathological"
end
```

List 1: Classification block (pseudo-code)

The limit has been fixed to 110Hz because it has been proved as the best classifier. In one hand, this is because that value is close to the real upper limit for oesophageal voices' pitch (which is always below 110Hz) and, in the other hand it works properly with healthy voice with such a low pitch. Moreover, that value avoids wrong measurements that can mismatch the estimation, which can happen when the range of frequential analysis is so wide and the second harmonic confuses the measure.

Once the voice is classified, there is a parameter estimation performed with healthy and pathological voices which is detailed in Figure 5. First of all, previous pitch range estimation is performed by means of the cepstrum of the signal. The cepstrum is obtained as follows:

$$C(q) = \left| fft(\log_{10}(\left| fft(signal) \right|)) \right| \qquad (9)$$

where *quefrencies* scale is directly related with the frequency scale with

$$q_i = \frac{fs}{f_i} \qquad (10)$$

This is logical because the FFT of a voice signal is a periodic signal due to the fundamental frequency and its harmonics.

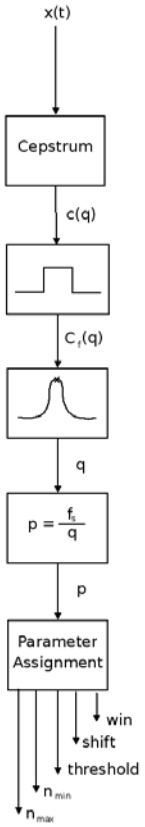

Fig.5: Pitch Range Estimation Block Detail

As it can be seen in the previous figure, the pitch estimation is obtained as the absolute maximum value of the C(q) in the range corresponding to 110 to 250Hz (qi=fs/250 and qf=fs/110 because the *quefrency domain* is inverted).

This value is not the exact pitch the algorithm is looking for, but it approaches a frequency range where the real pitch is probably going to be. At this point must be explained that in the other category this is not possible because in oesophageal voices the quefrency range analyzed is very noisy and the estimated range would be wrong otherwise.

With this cepstral approximation to the pitch, we can divide the whole pitch range in frequency bands of 20/30Hz on which different parameters will be set in order to be used in the core functions to identify correctly the pitch marks. Those parameters (as they appear in Figure 5 and the next table) are:

- ***win***.- *Window length in samples used in the analysis of the Core Algorithm*
- ***shift.-*** *Number of samples the window is shifted*
- ***threshold.-*** *the relative value used to remove the less energetic peaks from the voice cycle vector in the Core Algorithm*
- ***nmin*** and ***nmax***.- correction parameters, see next subsection

Case	win	shift	thres	nmin	nmax
Prepitch					
Always	64	10	0.2	-	-
Prepitch<110					
Always	72	10	0.5	fs/250	fs/40
Prepitch>110					
100<p<110	50	10	0.2	fs/150	fs/40
110<p<130	50	10	0.2	fs/(p+5)	fs/(p-5)
130<p<160	50	10	0.2	fs/(p+15)	fs/(p-16.5)
160<p<180	50	10	0.2	fs/(p+15)	fs/(p-15)
180<p<210	20	5	0.2	fs/(p+15)	fs/(p-25)
210<p<230	20	5	0.2	fs/(p+25)	fs/(p-15)
p>230	10	5	0.2	fs/270	fs/220

Table 1: Parameter assignation depending on the frequency range (fs=16kHz)

where 'p' is the value of the cepstral estimation of the pitch and 'fs' is the sample frequency of the signal.

3.4.4 Corrective Actions

Apart from the known parameters, two new parameters were introduced for correction purposes. These parameters represent the minimum and maximum samples allowed between consecutive peaks of the base respectively. In the most general sense the human voice range is from 40Hz to 250Hz in normal voicing including the oesophageal voices but with these parameters the accuracy could be improved.

The core algorithm has two drawbacks. The first of all is related to the threshold parameter: if the signal level falls outside the threshold value, there are some peaks missing and this affect to the pitch. In this case it can be said that the algorithm has lost the tracking of the voice for a while. This produces an increment in the distance (in samples) between consecutive peaks, which lead to the algorithm to obtain an instantaneous pitch values which are so low, but this doesn't mean that voice has low pitch component(s), and must be corrected (Missing Peaks Corrections). So if the distance in samples between two consecutive peaks is greater than a given value (nmax), that distance is not taken in account.

In the other hand, the windowing parameterization to obtain the pitch sometimes get wrong consecutive peaks (normally due to a small window values) and take two peaks instead of one for one region. This happens when the absolute maximum of a window is in the end of that window and the relative maximum of the next consecutive window is in the beginning of it. This situation produces the opposite effect to the previous, when this happens the distance between two consecutive peaks is too short and it produces high pitch components so one of both peaks (normally the smaller one) must be removed from the base (Consecutive Peaks Correction) if the number of samples between them is smaller than another value (nmin). Apart from correcting these drawbacks, this block can increase the accuracy of the algorithm by adjusting also such parameters to the band in which the pitch is going to be.

This correction is a kind of pitch component filtering so if it is expected to obtain a pitch in i.e. the range of 200Hz (with the cepstral pitch estimation) it can be set up to only allow the instantaneous pitch components near that one, improving the accuracy that way.

In the other category, the one with a pre-pitch lower than 110 of oesophageal voice, the estimation is quite straight, the parameters for oesophageal and low pitch voices are used directly in the core algorithm and the pitch is directly extracted. In this category there is only a slight correction to check that the instantaneous pitch components are within the human voice range (40-250Hz).

Once the *"mark vector"* has been calculated voice related measures can be directly measured.

4 Experimental Results and Discussion

In this section, the obtained results are presented. First of all an introduction to speech databases is presented which is focused in the difficulties appeared in oesophageal voice samples gathering. Then the results are widely detailed and commented taking a benchmark tool as reference to finish with a discussion about the results and other possible alternatives.

4.1. Oesophageal and Healthy Voice Database

There are some known databases widely used in this field, however, none of them includes oesophageal voice samples. One of them is Paul Bagshaw's database which is used in the evaluation of fundamental frequency determination algorithms [7] and the other one is the "Disordered Voice Database" from Kay [8].

As these databases were not the best choice for the aimed purpose, another approach was thought in the database gathering: real patients' voices from local institutions were to be recorded in order to use them in the group's research work.

A corpus was built with the help of a local ORL and the local association of laryngetomees, who gently provided several utterances of patients and oesophageal voice.

The final corpus was composed by 316 utterances: From these utterances, 119 of them were of patients with slight pathologies, 108 were of esophageal voices (or severe pathologies), and 89 of healthy people.

The recording process was performed using a MZ-R700PC portable MiniDisc (MD) device with an integrated microphone. The recording room selected was an ordinary empty meeting room in silence which was lent by the local association of laryngetomees. As the MD uses analogical information, the recording files were digitized in 44100Hz by means of a computer and its corresponding recording software.

This corpus was used to check the accuracy and results of the designed algorithms.

4.2 Cycle bounds detection and mean period

The achieved results are so good for oesophageal voices where typical marking techniques such as autocorrelation related ones ([5] and [6]), fail.

To illustrate the results, Table 2 is presented which compares the mean inaccuracy (the percentage of the difference between the obtained value, shown in Table 3, and the real value measured manually, shown in the fourth column in the same table, relative to the real value) of: a) the proposed algorithm, b) autocorrelation of sign coding and clipping (ASCC, [4]) and c) autocorrelation with window effect correction (AWEC, [5]).

	Mean Inaccuracy (%)		
	Healthy	*Pathological*	*Oesophageal*
Proposed Algorithm	**0,13**	**0,02**	**0,46**
ASCC	0,31	0,1	12,82
AWEC	3,58	0,06	5,08

Table 2: Algorithm's Mean Inaccuracies

The results obtained for healthy and pathological voices are better with a mean inaccuracy of 0,13% and 0,02% (Table 2) comparing with the 0,31% and 0,1% of ASCC, which is the best among the other possibilities: see for example #2 of healthy and #7 of pathological in Table 3 which compares the results with the exact value measured manually:

	#	**Mean Period (ms)**			
		AWEC	**ASCC**	**Proposed Algorithm**	**Exact Value**
Oesophageal	1	NA	10,860	*9,562*	**9,561**
	2	14,714	12,195	*14,366*	**14,366**
	3	NA	15,727	*16,404*	**17,022**
	4	NA	16,560	*16,250*	**16,251**
	5	9,992	11,185	*11,196*	**11,197**
	6	17,326	11,728	*15,920*	**15,919**
	7	16,345	11,301	*16,830*	**16,829**
	8	16,873	15,940	*16,795*	**16,793**
Pathological	1	7,374	7,378	*7,373*	**7,374**
	2	4,922	4,943	*4,926*	**4,928**
	3	5,425	5,430	*5,429*	**5,431**
	4	5,527	5,517	*5,518*	**5,519**
	5	6,264	6,257	*6,261*	**6,262**
	6	4,163	4,153	*4,161*	**4,162**
	7	6,700	6,700	*6,699*	**6,699**
	8	6,343	6,351	*6,348*	**6,347**
Healthy	1	9,307	9,695	*9,655*	**9,645**
	2	4,640	4,657	*4,643*	**4,642**
	3	5,194	5,194	*5,184*	**5,188**
	4	9,040	9,050	*9,049*	**9,038**
	5	8,938	10,100	*10,037*	**10,014**
	6	5,178	5,177	*5,171*	**5,178**
	7	8,343	9,693	*9,734*	**9,725**
	8	8,831	8,818	*8,857*	**8,830**

Table 3: Comparative Results

They are still similar but comparing the results for oesophageal voices, the measure improves significantly: the reached mean inaccuracy is of 0,46% against the 12,82% of ASCC (as example see #2 of oesophageal in Table 3). It must be pointed out that there are cases where measurement is impossible (marked as Not Available, NA, in Table 3).

4.3 Amplitude perturbation characterization

As introduced before, this article is focused in shimmer measurement and the algorithm described in the previous section is suitable enough to obtain better results than the chosen benchmark tool: MDVP from Kay Elemetrics, which is considered a *de facto* reference in acoustic measurements applied to voice pathologies.

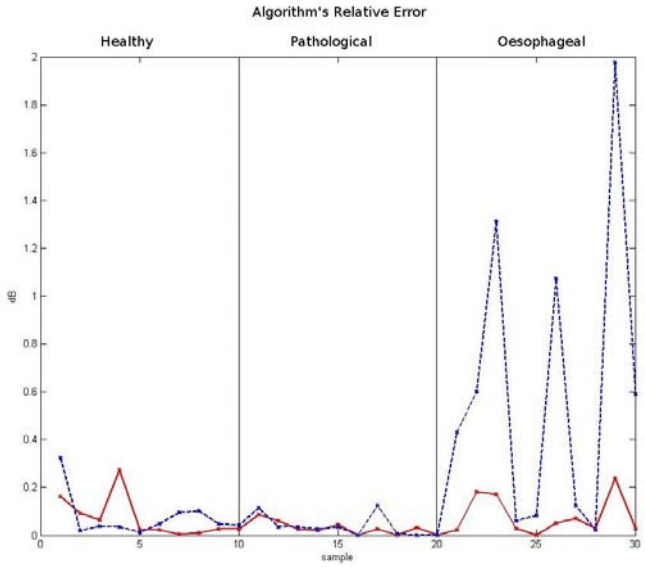

Fig. 6: Shimmer Results by Voice Type

The previous figure shows the relative error in dBs of the proposed algorithm (continuous line) and the benchmark tool (BM, dotted line). To obtain the error, the same measurements were performed in the same samples using two alternatives previously mentioned. The error was defined as the distance to the real value (measured manually) in decibels. As it can be seen, all the results are accurate enough for evaluate the three kinds of voices within a limited threshold error.

Apart from that two further deductions have been done from figure 6. In one hand, the results for healthy and pathological voices are very similar in both alternatives. Furthermore, all the results are below the same threshold for both alternatives.

In the other hand the results of the benchmark tool are not good for oesophageal voices. This is mainly because such kind of commercial tools were not designed in the beginning to work with oesophageal voice, thus, their results with that kind of voices must be wrong. This can be clearly seen in Figure 6 where almost all the results for oesophageal voices measured with the benchmark tool have errors that cannot be admitted.

Focusing in the accuracy of the algorithm and comparing it with the benchmark tool several tests and measurements have been performed in order to check the shimmer characterization. For that, 10 samples for each kind of voice (healthy, pathological and oesophageal) have been taken and their perturbation have been calculated in a cycle by cycle basis.

Next, a full-detail analysis is presented in Tables 4, 5 and 6. First of all the voice cycles have been determined using the previously specified algorithm and then the shimmer have been measured taking into account up to 15 cycles using the same cycle bounds in both the proposed definition (4) and the benchmark tool.

The table distribution is as follows, in the left side appears the number (#) of cycles analyzed and the each row represents each measure of shimmer (in dB) using both alternatives for each voice sample. BM stands for Benchmark Tool and the final row (e) represents the deviation of each measure against the other considering 15 cycles.

	HEALTHY									
	a_health_1		a_health_2		a_health_3		a_health_4		a_health_5	
#	Algorithm	BM	Algorithm	BM	Algorithm	BM	Algorithm	BM	Algorithm	BM
2	0,179	0,121	0,134	0,183	0,255	0,255	0,090	0,097	0,514	0,514
3	0,309	0,325	0,097	0,107	0,168	0,168	0,140	0,050	0,367	0,395
4	0,284	0,293	0,091	0,098	0,143	0,127	0,378	0,382	0,289	0,273
5	0,319	0,315	0,122	0,127	0,182	0,189	0,297	0,331	0,346	0,360
6	0,267	0,255	0,126	0,130	0,240	0,234	0,272	0,274	0,415	0,420
7	0,246	0,236	0,125	0,129	0,208	0,216	0,229	0,231	0,457	0,456
8	0,251	0,238	0,129	0,132	0,212	0,208	0,207	0,228	0,521	0,518
9	0,242	0,226	0,115	0,117	0,216	0,209	0,207	0,242	0,495	0,494
10	0,218	0,204	0,108	0,109	0,241	0,219	0,197	0,229	0,444	0,442
11	0,201	0,192	0,100	0,099	0,281	0,270	0,205	0,234	0,416	0,408
12	0,190	0,181	0,091	0,092	0,315	0,312	0,191	0,217	0,428	0,420
13	0,190	0,185	0,093	0,094	0,312	0,309	0,202	0,215	0,414	0,413
14	0,216	0,212	0,090	0,091	0,304	0,301	0,204	0,202	0,408	0,411
15	0,223	0,220	0,089	0,090	0,288	0,290	0,203	0,206	0,467	0,468
e	*0,003*		*0,001*		*0,002*		*0,003*		*0,001*	

Table 4: Details of the cycle by cycle analysis (healthy)

	PATHOLOGICAL									
	a_pat_1		a_pat_2		a_pat_3		a_pat_4		a_pat_5	
#	Algorithm	MDVP	Algorithm	MDVP	Algorithm	MDVP	Algorithm	MDVP	Algorithm	MDVP
2	0,374	0,409	0,188	0,116	0,041	0,045	0,059	0,116	0,112	0,104
3	0,226	0,302	0,106	0,113	0,056	0,061	0,079	0,077	0,202	0,198
4	0,188	0,256	0,167	0,161	0,046	0,047	0,154	0,077	0,182	0,182
5	0,161	0,203	0,210	0,157	0,086	0,086	0,142	0,066	0,221	0,222
6	0,140	0,173	0,190	0,130	0,118	0,117	0,114	0,059	0,200	0,200
7	0,118	0,154	0,177	0,111	0,098	0,098	0,105	0,058	0,200	0,205
8	0,158	0,168	0,227	0,189	0,096	0,096	0,098	0,063	0,174	0,210
9	0,148	0,163	0,283	0,261	0,087	0,087	0,089	0,065	0,168	0,175
10	0,141	0,146	0,253	0,241	0,091	0,091	0,081	0,066	0,191	0,197
11	0,131	0,138	0,238	0,219	0,085	0,085	0,090	0,082	0,199	0,204
12	0,131	0,127	0,221	0,216	0,077	0,078	0,086	0,076	0,184	0,189
13	0,128	0,125	0,219	0,229	0,073	0,074	0,084	0,071	0,191	0,196
14	0,120	0,124	0,244	0,255	0,071	0,071	0,089	0,073	0,198	0,202
15	0,124	0,125	0,235	0,244	0,070	0,070	0,087	0,074	0,205	0,207
e	*0,001*		*0,009*		*0,000*		*0,013*		*0,002*	

Table 5: Details of the cycle by cycle analysis (pathological)

	OESOPHAGEAL									
	a_oeso_1		a_oeso_2		a_oeso_3		a_oeso_4		a_oeso_5	
#	Algorithm	MP	Algorithm	MP	Algorithm	MP	Algorithm	MP	Algorithm	MP
2	0,740	0,275	1,614	0,184	0,103	0,472	0,277	0,091	0,929	1,032
3	0,380	0,138	0,826	0,096	0,493	0,487	0,560	0,184	0,708	1,033
4	0,558	0,120	0,959	0,067	0,412	0,434	0,403	0,235	0,604	0,965
5	0,423	0,315	0,744	0,057	0,516	0,329	0,325	0,356	1,037	0,815
6	0,931	0,493	0,817	0,200	0,639	0,343	0,289	0,401	1,458	0,680
7	0,960	0,573	0,719	0,308	0,813	0,387	0,296	0,421	1,744	1,450
8	1,210	0,516	0,664	0,362	0,839	0,385	0,313	0,382	2,091	1,681
9	1,059	0,465	0,998	0,333	0,763	0,357	0,700	0,518	1,891	1,725
10	0,986	0,496	0,917	0,330	0,761	0,411	0,803	0,518	1,884	1,565
11	0,898	0,480	1,089	0,317	0,850	0,404	1,216	0,518	1,761	1,512
12	0,831	0,471	1,074	0,315	1,006	0,388	1,703	0,975	1,634	1,432
13	1,135	0,441	1,029	0,329	0,976	0,369	1,634	0,882	1,532	1,323
14	1,063	0,433	0,976	0,333	0,998	0,348	1,612	0,820	1,505	1,323
15	1,235	0,425	0,925	0,342	1,006	0,345	1,711	0,929	1,448	1,323
e	*0,810*		*0,583*		*0,661*		*0,782*		*0,125*	

Table 6: Details of the cycle by cycle analysis (oesophageal)

The results for healthy voices are very good with a mean deviation of 0,002dB. It can be seen how the shimmer measurement is better as the number of the cycles taken into account rises and finally the deviation is reduced to the minimum.

The results for pathological voices (Table 5) are very similar to the healthy ones. It looks like the accuracy is not as good as in healthy ones but even though they are very similar.

Finally the results for oesophageal voices (Table 6) are not similar as it was observed in the analysis of figure 6. The differences are noticeable enough to determine the lack of accuracy of the benchmark tool in the shimmer estimation for this kind of voices. See for example, samples a_oeso_1 to a_oeso_4 where the deviation is greater than 0,5dB.

4.4 Discussion: Alternative shimmer definitions

In this section, the three different shimmer definitions presented in section 3.2 will be discussed. The main idea was to show how equation (5) could lead to increase the value of the shimmer for oesophageal voices which could be useful in medical diagnosis of pathologies. For that the results will be compared with the ones obtained with equation (4) and (6).

The definition in (4) is the one which has been used in the previous analysis so it can be used as reference in this case to establish de validity of the hypothesis while (6) has been also chosen to complete all the possibilities.

The absolute values obtained with (4), (5) and (6) are presented below:

	Shimmer(dB)		
	Eqn. (4)	Eqn. (5)	Eqn. (6)
Healthy			
1	0,223	0,221	0,213
2	0,089	0,096	0,081
3	0,288	0,329	0,280
4	0,203	0,263	0,231
5	0,467	0,517	0,447
Pathological			
1	0,124	0,137	0,116
2	0,235	0,231	0,284
3	0,070	0,080	0,100
4	0,087	0,106	0,068
5	0,205	0,258	0,193
Oesophageal			
1	1,235	1,874	0,925
2	0,925	1,253	0,582
3	1,006	1,224	0,388
4	1,711	1,872	1,032
5	1,448	1,864	1,239

Table 7: Alternative definition's results

It can be seen how especially oesophageal values fulfil completely the hypothesis defined previously. Furthemore, from that it can be verified that (6) is useless for the defined aim as the value is generally reduced.

In fact, the equation (5) fulfils the objective of increasing the shimmer's value for all cases but two, and in those two cases the reduction of the value is of less than 0.004dB.

Finally, the increment in shimmer value is generally higher for oesophageal voices than in the other two kinds of voice which could be very useful to establish shimmer ranges of pathology.

Figure 7 illustrates the variation of the shimmer value comparing the results of (5) and (6) with (4) by means of the next expression:

$$\Delta Shim^{(i)} = 100 \cdot \frac{Shim^{(i)} - Shim^{(4)}}{Shim^{(4)}} \qquad (11)$$

Where 'i' is referencing the equation that is being compared with (4) which has been taken as reference in previous subsection. As it can be seen (11) is a directional variation parameter given in percentage: negative values indicate a reduction of the value while positive ones are increments with respect to results of (4).

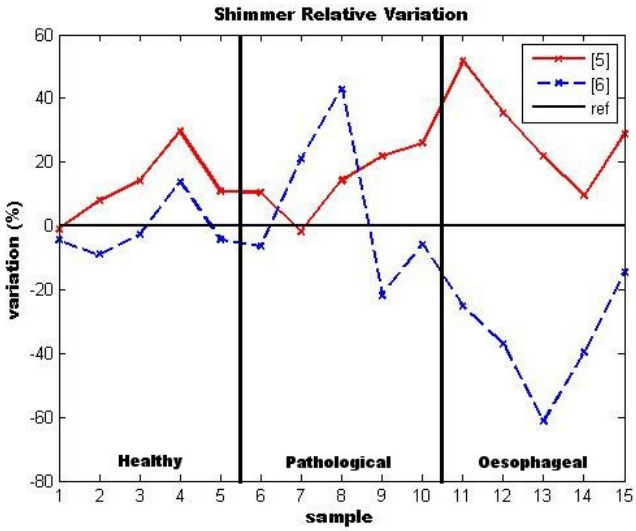

Fig. 7: Variation of Shimmer Values

As it can be seen in figure 7, the dashed line corresponding to de variation of (6) is not stable: the variation is not maintained nor over neither below the reference so it has been discarded.

Regarding the continuous line representing the variation of equation (5) it can be seen that almost all the values are over the reference which was the original idea so, in the end, it has been proven that with the definition (5) the shimmer value is increased specially for oesophageal voices.

Moreover, percentage variations of healthy voices suppose a smaller variations in absolute value because they have a lower shimmer value i.e. voice "healthy 4" of figure 7 has a variation of 29,56% which means a variation of 0,06dB while voice "oesophageal 5" with a similar variation (28,73%) have a variation of 0,416dB.

5 Conclusions

The conclusions to be pointed out for the present article are divided in two fields. In one hand, the algorithmic field which is related to the voice processing applied to acoustical parameterization. In the other hand it is also related to the medical field which includes the application of the present research in the improvement of laryngetomee diagnosis and rehabilitation.

Regarding the processing, which in this particular case has been applied in voice periodicity characterization, the pitch has been measured and its estimation has been significantly improved. Comparing the obtained results with those calculated with other techniques, it

has been reached a maximum error of 0,46% for the worst case: oesophageal voices. The algorithm also allows Shimmer calculation whose value represents the amplitude instability of the analyzed voice. In the case of oesophageal voices, is not possible to calculate this value accurately with other techniques but with the proposed algorithm, the automatically obtained values are so close the real ones obtained manually.

In spite of the accuracy obtained in shimmer calculation, it has been analyzed the possibility of reformulating the measure. Several alternatives have been defined and as conclusions of the present study, authors have checked that a concrete alternative allows maximizing the measurements of the variation of shimmer. This fact is very interesting in order to use this value as automatic classifier of pathological and oesophageal voices because the calculated values maximize the cases of low intelligibility and quality.

The applicability of the present work fits several fields:

- Biomedical engineering and, in particular, in the specialty of otorhinolaryngology where Dr. Agustin Perez Izquierdo in the Hospital of Basurto is using a software which integrates the presented algorithms, using them in evaluation and rehabilitation of laryngectomee patients. Thanks to this tool, he is able to obtain a set of quantifiable objective parameters.
- Design of new software tools in order to support patients easily at home in their oesophageal voice learning stage after the extirpation of their vocal folds due to a severe disease..
- Design of new software tools in order to support teachers of oesophageal voice learning classes with the aim of having quantifiable references of their students evolution
- In the research field the calculus of those parameters will allow the evaluation in an objective way the performance and results of several signal processing algorithms which are applied on oesophageal voices in order to improve their quality, e.g. in telephonic communications or VoIP.

Finally and even having outstanding results, the research line remains open in order to, in one hand, develop measurement algorithms which detect even more accurately the periodicity instants. In the other hand, it will be very significant to explore new parameter for oesophageal voice characterization in order to improve the model of their characteristics and make their analysis, process and evaluation easier.

6 Acknowledgements

This research was partially carried out under grant TEC2006-12887-C02-02 from the Ministry of Science and Technology of Spain. Authors also wish to thank the support of INRIA's EUROMED 2007 project.
The authors wish to thank to Andrés Gola his support in the algorithms testing stage. In particular, PhD. Agustín Pérez Izquierdo, Doctor otorhinolaryngologist of the Basurto Hospital in Bilbao, who helps our research with voice recordings of his patients.

References:

[1] B.García, J.Vicente, I.Ruiz, A.Alonso and E.Loyo, "Esophageal Voices: Glottal Flow Regeneration", *in Proc. ICASSP 2005*. Philadelphia, USA, March 2005.
[2] J.Chen and Y.Kao, " Pitch Marking Based on an Adaptable Filter and a Peak-Valley Estimation Method", in *Computational Linguistics and Chinese Language Processing*, vol. 6, no. 2, 2001, pp.1-12.
[3] M.Kobayashi,M.Sakamoto,T.Saito,Y.Hashimoto, M.Nishimura and K.Suzuki, "Wavelet analysis used in text-to-speech synthesis," *IEEE Transactions on Circuits and Systems-II, Analog and Digital Signal Processing*, 45(8), 1998, pp. 1125-1129.
[4] P.J. Baken, R.F. Orlikoff, Clinical Measurement of Speech and Voice. 2nd Edition, CA:Singular Publishing Group, San Diego, 2000.
[5] D.D.Deliyski, "MDVP Acoustic Model and Evaluation of Pathological Voice Production", *Eurospeech*, Berlin, Germany, September 1993.
[6] P.Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound", *IFA Proceedings*, nº 17, 1993, pp97-110.
[7] P. Bagshaw, S. Hiller, and M. Jack, "Enhanced pitch tracking and the processing of F0 contours for computer aided intonation teaching", European Conference on Speech Communication and Technology, *Eurospeech*, Berlin, Germany, September 22-25, 1993
[8] "Disordered Voice Database", Version 1.03, Kay. Elemetrics Corp, 1994

Begoña García Zapirain *was born in San Sebastian (Spain) in 1970. She graduated in Telecommunication Engineering speciality in Telematics for the Basque Country University in 1994. In 2003 defended her doctoral thesis in pathological speech digital processing field. After many years working in ZIV Company, in 1997 she incorporates to University of Deusto faculty as teacher in signal theory and*

electronics area. She is leading since 2002 the Telecommunication Department of University of Deusto. In 2001, creates with Javier Vicente Sáez the researching group Advanced Signal Processing (PAS) in the same university, playing the role of main researcher. Member of IEEE and EURASIP, she is nowadays part of the ISIVC organization committee.

Ibon Ruiz Oleagordia *was born in Bilbao (Spain) in 1975. He graduated in Physical Science at the Basque Country University in 1999 and got the degree of Electronic Engineering at the same university in 2001. He has his doctoral thesis registered. Since 2000 works as teacher and is part of the Computer Architecture, Electronic, Automation and Telecommunication department at the University of Deusto. In 2002, he became part of the researching group Advanced Signal Processing (PAS) in Deusto University.*

Amaia Méndez Zorrilla *was born in Barakaldo (Spain) in 1978. She graduated in Technical Industrial Engineering, speciality Electronics in 1999 at University of Deusto (UD). Later, in 2001, she graduated in Telecommunication Engineering at the same university. She has her thesis registered in Biomedical engineering. She is teacher of the UD since 2003 and is part of the Telecommunication department in UD. She becomes part of the researching group Advanced Signal Processing (PAS) in UD in 2005.*

Mikel Mendezona Goyarzu *was born in Bilbao (Spain) in 1982. He graduated in Technical Industrial Engineering, speciality Electronics in 2003 at University of Deusto (UD). Later, in 2005, he graduated in Telecommunication Engineering at the same university. In 2006 he becomes part of the Advanced Signal Processing (PAS) researching group at Deusto University as researcher. Nowadays is realizing the doctorate courses in Information Systems at the same university.*