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

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

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

Vocal Folds Stroboscopic Image Processing for Otolaryngology

A. Méndez Zorrilla and B. García Zapirain

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/55343

1. Introduction

Nowadays it is very common to go to a specialist's surgery because of voice disorders. Voice pathologies are characterized by the abnormal production and/or absence of vocal quality, pitch, loudness or resonance. Approximately 28 million workers in the U.S. experience daily voice problems [1][2] and the statistics indicate that voice pathologies affect almost five percent of the population [3].

Daily life sometimes affects our voice and our vocal cords. Talking too much (in the case of occupational voice users, such as singers, teachers, lawyers, telephonists…), screaming, constantly clearing your throat or smoking can make you hoarse. For example, teachers have missed many workdays due to voice problems and are more likely to consider changing occupations because of their voice [4].

Other psychosocial factors, such as stress and anxiety involve voice problems as well [5]. All the causes mentioned above. All the aforementioned causes can also lead to pathologies such as nodules, polyps and sores on the vocal cords [6].

The vocal folds (or vocal cords) are composed of twin infoldings of mucous membrane stretched horizontally across the larynx. Their vibration produces each person's voice [7].

The most common benign pathologies are the following (all of them analysed in this study):

- Nodules. This disorder prevents the vocal folds from meeting in the midline, which produces an hourglass deformity on closure resulting in a harsh/rough/coarse…, breathy voice. Nodules are most common in children and females.

- Polyps. This pathology is a benign lesion of the larynx, occurring mostly in adult males, usually located on the phonating margin of the vocal folds and preventing the vocal folds from meeting in the midline.

- Cysts. A cyst is a firm mass of organic material contained within a membrane. Cysts can be located near the surface of the vocal fold or deeper, near the ligament of the vocal fold.

- Paralysis. Paralysis occurs when only one side is paralyzed in the paramedian position or has very limited movement. It is more common than bilateral involvement.

All previously mentioned morphological path morphological pathologies disturb the phonation process resulting into a hoarse voice signal, and the hoarse is the main reason To go to the specialist. Usually, patients can recover their voice with rehabilitation, and only sometimes the surgery is necessary.

The most widely used clinical voice disorder assessment tools for capturing vocal fold videos are digital videostroboscopies [8] and high speed recordings [9]. The vocal folds can be examined by inserting an endoscope through the nose or mouth. The examiner uses the endoscope light to view the folds and their movement patterns during phonation (producing sound) and when at rest. This test is invasive and very uncomfortable for the patient and in some cases the process has to be/ needs to be repeated in order to view the vocal folds correctly.

In this research, only low-speed recordings illuminated with a stroboscopic light are used. We employ low-speed recorded images since there are intensively used among othorhinolaryngologists and voice specialists [10].

These recordings are usually very problematic because of the different level… the different level of illumination inside the video, the patient's movements while the doctor is recording, or the zoom. Figure 1.a and 1.b shows two very poor-quality frames. Figure 1.a represents a low level of illumination and Figure 1.b illustrates patient movement because both frames are consecutive.

Due to recording difficulties… difficulties, segmentation without initialization (other related works with user interaction or initialization are supported in [11][12]) is more difficult to obtain than in high-speed recordings.

However, the problem lies in the fact that a specialist delivers a diagnosis in a subjective way and it depends on his experience in this area. The study of glottal space (illustrated in Figure 1) in a video sequence can be very useful and decisive when it comes to obtaining an accurate diagnosis. Figure 1.c shows a very good quality image of healthy vocal folds and in Figure 1.d vocal folds with one of the studied pathologies –polyps- can be seen.

The main goal is to find a methodology to parameterize vocal folds, obtaining glottal space segmentation without user interaction, a pre-diagnosis support based on a classification stage and some objective measurements to support the doctors' final diagnosis and it allows (the doctor) to do comparatives and control the patients' evolution after rehabilitation or surgery. This chapter has been drawn up taking into account specially the context of development in clinical medicine and telemedicine applications in the otolaryngology field.

In this process we consider that the following specific objectives have to be taken into account:

- To work and analyze the commercial database "Laryngeal Videostroboscopic Images (Dr. Wendy LeBorgne; Plural Publishing)" and some recordings given by the otolaryngologist Dr. Agustín Pérez Izquierdo from Basurto Hospital.

- To segment vocal fold glottal space correctly and reduce a significant amount of image data

- To classify the images to carry out a pre-diagnosis

- To apply block matching techniques to have information about vocal folds movement

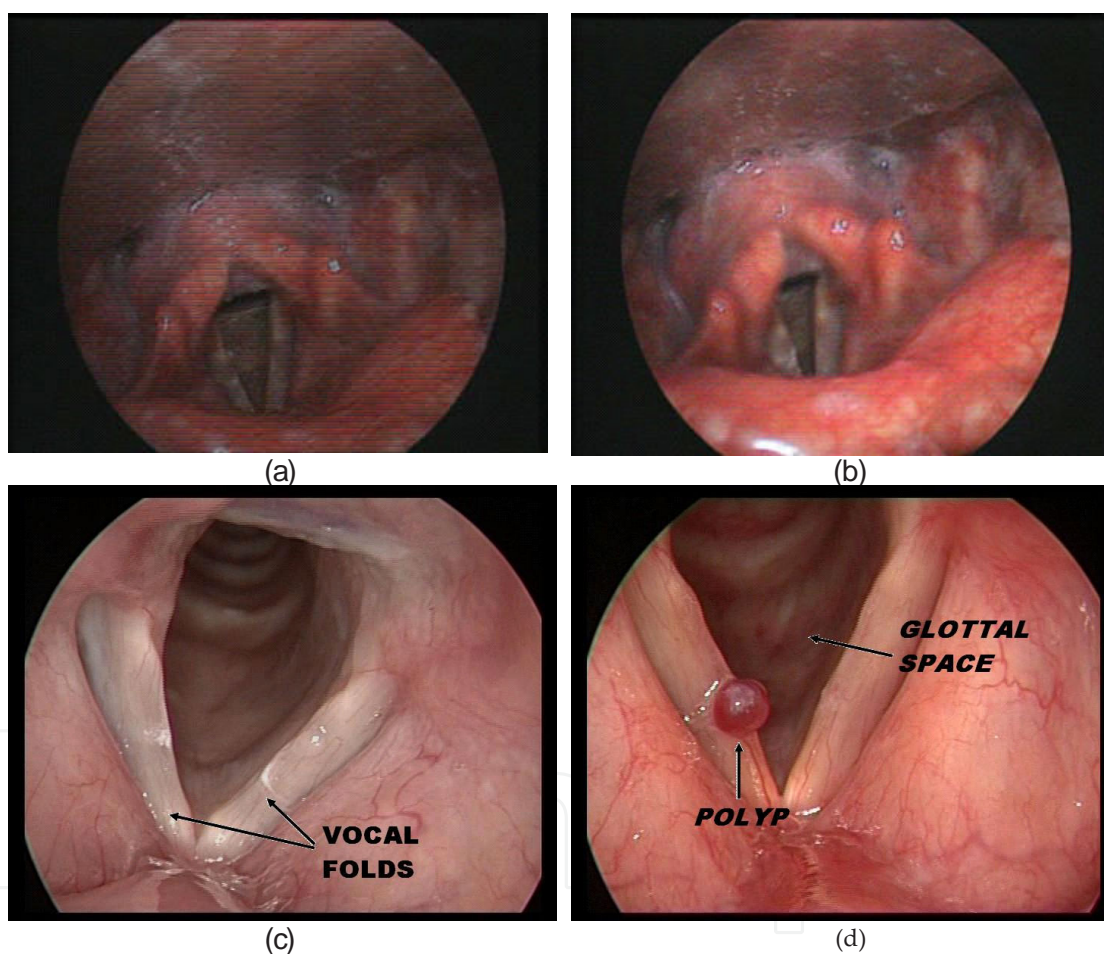- To define and measure objective parameters to help the diagnosis

Figure 1. Vocal folds Images. a) and b) Poor quality frames with different illumination. c) and d) Good quality vocal folds images. c) Healthy folds and b) Pathological folds.

2. State of the art in image processing for otolaryngology

This section is divides into 3 subsections to describe the state of the art in otolaryngology field.

2.1. Image capture techniques and software analysis

Given that this research is focused on the study of images (and videos) of vocal cords, It is also object of this study the analysis of different methods of image digital capture of them,.

Nowadays, the scientific and medical community accepts mainly two capture techniques whose results can help to determine the diagnosis of different vocal fold pathologies. On the one hand there is **videostroboscopy**, and, on the other hand, de **videoquimiography** or high speed image capture.

The way to visualize and study the vocal cords has been object of study for many centuries.

Manuel García is considered as the the first discoverer of the indirect laryngoscopy with the speculum. It was in 1854 [13]. At the beginning he was considered an intruder, since he was a composer, a tenor and a singing teacher, but not a doctor, as might have been expected. But his idea of placing a dentist speculum in the throat and going on illuminating the larynx with the sun light reflected in a mirror he holds in her hands, examining the vocal cords [14] was subsequently recognized by all the laryngological societies.

Later on, it was Johan Nepomuk Czermal from Budapest [15] who improved that technique using artificial light and speculums of different sizes and it is precisely him who achieved to introduce the indirect laryngoscopy as the main exploratory method.

In 1975, Stuckrad and Lakatos [16] developed this technique with amplification, that is to say, with a magnifier. But it is not until 1978 when Oertel [17] developed the laryngostroboscope, which permits the examination of vocal cords vibration.

This exploratory method allows the diagnosis by the observation of pathologies in initial stages, or else, those ones which do not affect the morphology of the vocal fold, but its movement.

In the nineties, the high speed digital videoquimiography of the larynx emerged [18]. This technique wishes to give a solution to the problems in the speed of the image capture in the videostroboscopy. The human eye is only capable of capturing 5 or 6 images per second, whereas the videostroboscopy captures between 25 and 30 frames per second [8] but it continues to be insufficient to observe the dynamic movements that take place in the larynx during the phonation. That is why the videoquimiography is currently being an appraised tool, mostly in the research field.

From the point of view of clinical examination, the digital videostroboscopy is the essential and routine method in the diagnosis of voice disorders, since it provides an extraordinary amount of information about the behavior of the vibratory cycle and its alterations.

And it is here where the next sections of this chapter are framed, and precisely in the diagnosis of the vocal cord alterations within the objective parameterization of stroboscopic videos.

2.2. Software analysis

Recently, international research groups, in collaboration with specialist medical staff are showing their interest in the development of the characteristics of the vocal fold images'

software analysis, though, habitually, they don't reach the market. Given that the VKG images are of better quality, the development of this application typology use to be developed for its analysis, as it can be observed in the literature [19-20].

Kay owns software named KIPS (Kay's Image Processing Software). KAY oriented to the vocal fold digital image processing, coming from stroboscopic captures or from high quality. It includes a multitude of tools for the image editing and processing, providing the specialist with a very valuable information in support of the diagnostic, but it neither issues a diagnosis nor it automates the process % 100.

Habitually, the softwares used by the otolaryngologists/ otolaryngologists, or any voice specialist have only into account the voice analysis, and take part of a unit or voice laboratory [21].

It is important to remember that the measurement of the acoustic parameters does not issue the diagnosis of the injury, but it may indicate the level of alteration of the dysphonia. The objective parameters accepted by the scientific community in the acoustic analysis are: pitch, jitter, shimmer and HNR (and its variants) and are indicative of a good or low vocal quality [22].

2.3. Image processing techniques to process vocal fold images

In this subsection, the authors describe come techniques which could be applied to this kind of images.

2.3.1. Active contours

These contours [23] model the boundaries between the object, the background/the bottom, and the remaining objects of the image. They allow to extract the contours of the object of interest based on models or forms that use previous information about the form of the objects. The active contours, or, also named Snakes [24] are much more robust against the presence of noise and other elements, and allow to segment much more complex images, like the medical images which are the object of this study, for instance.

The segmentation results, provided by this technique, applied to the vocal cords images are fairly positive, as it can be seen, but no previous initialization is required.

2.3.2. Wavelet transform

The Wavelet transforms [25] and other multiscale analysis functions are widely used in the digital image processing for the applications of noise elimination/reduction, compression and feature extraction.

Its application in bioengineering is very extended [26], and more specifically in applications of medical images like: ultrasound images, tomography images or magnetic resonance images. The capture methods of these images and, therefore, its characteristics are very different to the vocal cord images that are object of study in this chapter.

2.3.3. Kalman filter

Within the digital image processing field, the Kalman filter is a recursive algorithm that is used to estimate the position of a moving point or characteristic, and the measurement uncertainty in the following image. It is about to reach the feature (point, edge, corner, region, and so on), in a particular area of the following image around the aforesaid position, in which we are sure to find the feature within a certain degree of confidence.

The aim of the filter is the acquiring of an optimal estimator of the state variables of a dynamic system, based on noisy observations and on an uncertainty model of the dynamics of the system [27-28].

It is more appropriate to conduct the study of the vocal cord movements through high-speed captured images, and in this study stroboscopic images were used. Thus, in this case, the Kalman filter has not been applied

3. Parameterization system proposal

The developed algorithms, belonging to the Middleware layer, normalize process, analyze and extract the features of different vocal fold pathologies. Mainly those pathologies which affect to the morphology of the vocal fold, and which, at least lead to dysphonia, and, in other cases, pathologies caused by an abnormal function in the movement and vibration of the vocal folds, as it can be the paralysis of the vocal fold.

The application layer is responsible for the representation of the data to the user in a friendly graphic interface.

Thus, a sequential order for the implementation of the study and the application of the algorithms, it is considered 0 stage, to the capture of the image by the specialist. Once the images are available, we focus in the extraction of all the parameters/features of the sequences of images that will give us some information about the pathology suffering by the patient or the absence of it. Subsequently, we try to classify/identify the image comparing withe the database (with previously classified and diagnosed images) that is available. Finally, with all the information a diagnosis to the specialist can be offered, supported by all the objective parameters extracted during the process.

The methodology followed during the whole design process has been constant, defining in each case the entrance and output requirements and variables of each of the blocks of the defined stages, as it can be described in the following point.

3.1. Design

The high level design establishes the form and substance of the system considering it as a whole, as a set of functions that constitute the structure or architecture of the system without going into detail of each one, since this will be made in the low level detailed Design.

In all the applications, except by the smallest ones, the first step to design a complete system consist on dividing it in a small number of components, blocks or stages. Each of the main blocks of a system covers some aspects of the system that share any common property. Each one of the blocks or stages contains/comprises a package of functions and interconnected events that share a common purpose, which have a well-defined interface with the remaining blocks or stages (and they habitually can be reused in various systems or applications).

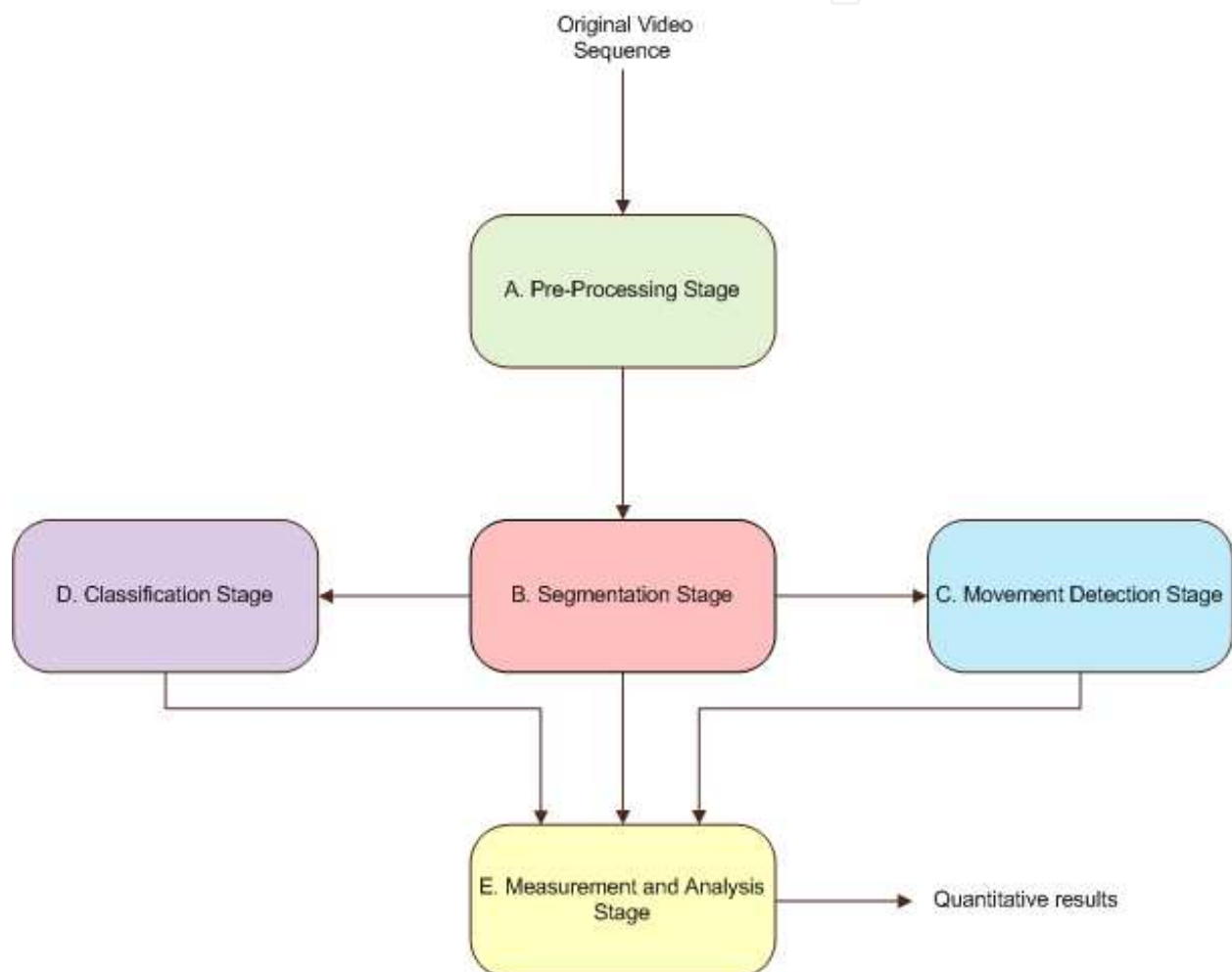The general bock diagram proposed to reach the established objectives can be observed in Figure 2.

Figure 2. High Level Diagram

The entrance of the systems are the frames of each of the sequences in original format (In colour, and with the quality and resolution acquired by the camera). Each one of the blocks apply the necessary transformations to obtain as a final result a report with those objective data which may help and support the otolaryngologist with the diagnoses, and to evaluate rehabilitation processes or the evolution after a surgical intervention.

Next, the five main blocks which compose/contain the low level design are described (Each one identified with a colour coding which will be maintained throughout the chapter (As it can be seen in Figure 2).

3.1.1. Pre-processing stage

This first block carries out the necessary functions like the unification, normalization and standardization of the characteristics of the stroboscopic videos of the vocal folds, or vocal fold images, so they can be treated more easily and efficiently in the subsequent blocks.

The *input* of this block is the original recording provided by the otolaryngologist. An isolated image can also be analyzed, but the results in some of the stages have a low reliability. In some stages the comparison/study of certain parameter between frames within the same sequence is necessary. It has to be taken into account the variability of the intraframe recording. Thus, the comparison is a utility that provides much information,

The *Output* of this block is a sequence of frames in JPG format normalized in characteristics and transformed into a grayscale whose gradient has been calculated (to facilitate the application of the filters of the subsequent stage).

3.1.2. Segmentation stage

Due to its level of criticality this is the main stage of the system. The validity of the system and the error rate obtained in subsequent stages depend, in part of its success. The three subsequent stages are based (even are used as Input) in the result of the segmentation stage.

During this stage the necessary transformations are used to segment and isolate, in this case, the region of interest. In this system the ROI it is going to be working with from this moment on is the glottal space.

The *input* of this block are the gradient images result of the normalization carried out in the A block.

The *output* of this block is the sequence of frames where the glottal space has been segmented. From the block B, the attention is paid mainly to this part of the image and to the study of the features of the glottal space perimeter.

3.1.3. Movement detection stage

In this stage Block Matching algorithms are applied, habitually used for the image compression MPEG standards, and widely extended in a multitude of applications [29]. The aim is to detect and study the interframe movement of the vocal folds. Always assuming that the stroboscopic images are not the most adequate ones to evaluate neither the movement nor the vibrations of the folds, since the frame rate is between 25 and 30 frames per second, far lower the vocal fold vibration speed.

For this reason, the results of this stage are relative and have to be complemented with the E stage measures, or even with other studies.

The *input* of this block is the frame sequence whose glottal space has been segmented in the block B.

The *Output* of this block are all the motion vectors calculated as a result of the application of the Block Matching algorithms.

3.1.4. Classification stage

With the stages described hitherto it has not been obtained any result which can guide us in the diagnosis. It only has been extracted the region of interest and movement measures. It is in this stage where a pre-diagnosis of the data obtained in the stages B and C is carried out. A classification of the entrance images is made through various algorithms, to discern between morphological and non-morphological pathologies, comparing the entrance of this stage with a previously classified and studied images database. The results of this stage, partially, depend on the choice and the size of that database to compare the inputs of this stage.

The D block has various inputs unlike the previous blocks. For the execution of this blocks three inputs are needed: the original frame sequence, the sequence of frames with the segmented glottal space (result of the B block), and the motion vectors calculated in Block C (Output of Block C).

The output of this block is a pre diagnosis/classification, according to the identification/classification algorithms applied.

3.1.5. Analysis and measurement stage

It is in this stage where finally, the necessary calculations are made to achieve some objective results which allow us to carry out an evaluation and a possible diagnosis, and even to discern which is the morphological pathology of the vocal folds the patient suffers. As far as the pathologies related to the vocal fold movements are concerned, some parameters which may guide the specialist to a more deep study will be provided, but bearing in mind that the results provided may not be 100% consistent, due to the frame rate of the available capture.

Once the necessary transformations are made, to achieve those results, some objective measures over the vocal fold images are made, and it is on their representation in which the evaluation and/or the diagnosis and the value of this contribution will be based.

The *input* of this block is multiple, as in Block D. To calculate the objective parameters which will be the final result of this thesis it is necessary to make calculations and transformations over the results in blocks B, C and D.

The *output* of this block will be the set of objective measures supporting the final diagnosis proposed to the specialist, which may become part of the final report given to the patient.

The novelty, complexity and main feature of the proposed system is that it does not **require** any type of neither **initialization** nor interaction with the user during its execution for the achieving of the diagnosis, evaluation and/or the measurement of the effectiveness of the treatment (computer-aided pre-diagnosis).

4. Results

As it has been reflected in the "Design" section, to illustrate de experimentation carried out, it would be taken into account that the system has 5 differentiated stages, and the results for each stage will be showed.

Special emphasis will be placed in the results of the pathologies related to the morphology of the vocal folds against those related to the movement, because of the type of the images studied and because of the sample available in the database, and finally the results obtained in different formats will be presented.

4.1. Segmentation stage tests

The segmentation stage is the first and the most important one in the whole process, since from its results it depends on the success in the subsequent stages. The obtaining of objective results for the characterization of the vocal folds and the emission of the diagnosis depend also on this stage.

In this stage we can observe graphic and numerical results. Those later ones are used to express the success percentage in the segmentation within the sequence.

If we analyze the results grouping the sequences into pathologies they can be further improved:

- *Healthy vocal fold Sequences*. The average frames in these sequences is 22,1, having obtained a correctly segmented frames average of 95,95%.

- *Sequences of Vocal Folds with Nodules*. The average frames in these sequences is 26, having obtained an average of correctly segmented frames average of 95,34%. Although the mean number of frames in the sequences is slightly higher, and this does not affect considerably to the percentage value of the correctly segmented frames.

- *Sequences of Vocal Folds with polyps*. The average frames in these sequences is 33,44, having obtained a correctly segmented frames average of 94,96%.

- *Sequences of Vocal Folds with Oedema*. The average frames in these 4 sequences is 31,75, having obtained a correctly segmented frames average of 89,82%.

- *Sequences of Vocal Folds with Cyst*. The average frames in these 5 sequences is 28,6, having obtained a correctly segmented frames average of 94,96%.

- *Sequences of Vocal Fold with Paralysis*. For the segmentation and the subsequent classification, these images are not differentiated each other in an independent way, from the frames of the healthy vocal folds. The average frames in these 7 sequences is 30, having obtained a correctly segmented frames average of 96,88%.

4.2. Classification stage tests

The classification stage is the third stage in the system, and it is applied to provide the doctor with a reference pre-diagnosis. It is also used as an entrance in subsequent stages, and,
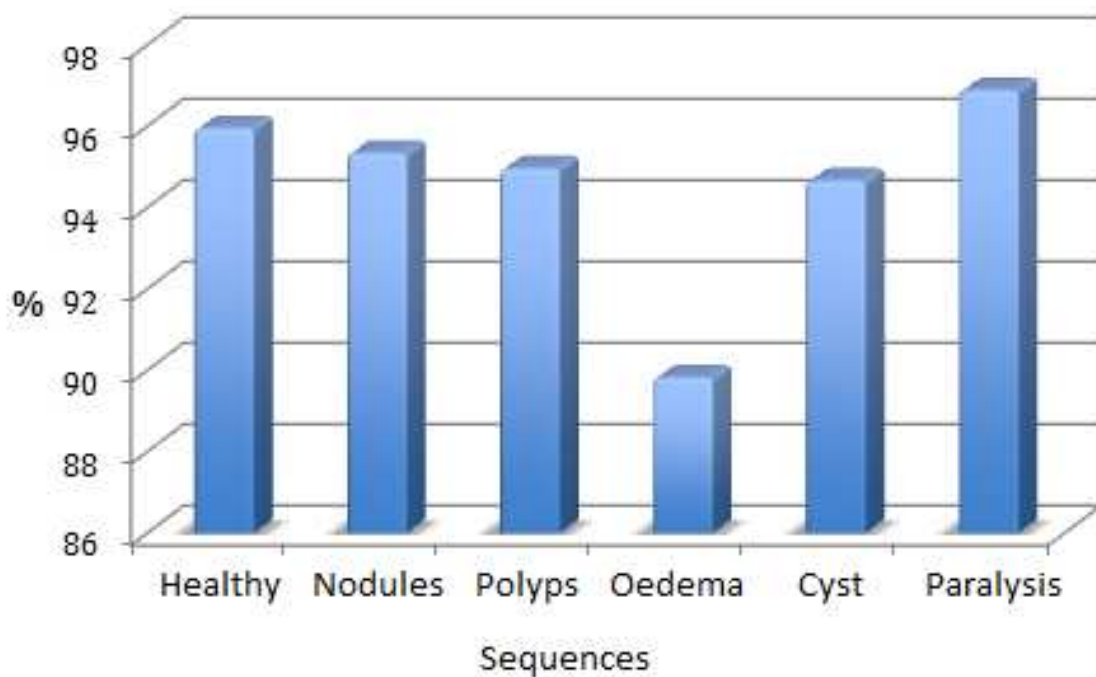
Figure 3. Graphic representation of the segmentation results

depending on the results in this stage, either one or other transformations are applied in the "Analysis and measurement" stage.

The critic element in the validation and the success of this stage is the choice of the Training Set and to establish which ones have to be those images are the most important decisions that was to be taken in this stage. Indeed, the success rate and the processing time in this stage, which it can be considerable, also depend on that decision. This whole system is a post processing in itself, and it is assumed that it is not carried out in real time.

During the experimentation some tests have been made with different Training Sets of 16,18, 25 and 100 images, being the definitive/final one that Set composed by 100 images.

The results obtained can be seen in Table 1. The percentage of correctly classified images with the exact pathology (nodules, polyp, oedema, cyst..) is 71,8%, whereas if we simply focus on the correctly classified ones as healthy or pathological the percentage is 92,1%. The latter is the result that matters to us, since the success of the final parameterization depends, in part, on it.

It is demonstrated that to change the Training Set makes the results to be completely different, see the Training Set of 25 images tested with 330 images, where the percentage of the images correctly classified with the exact pathology varies from 67,23% to 71,20%. 4% of difference, which, depending on the images of the Training Set can be wider.

In Table 1 is demonstrated that as we go by raising the number of images of the Training Set the results improve, from 85% of a Training Set with 16 images to 92,09 % of the Training Set of 25 Images.

Number of images in Training Set	Number of test images	% of images classified with the correct pathology	% of images correctly classified (healthy or pathological)
25	330	67,23	92,09
25	330	71,20	87,65
18	330	69,98	86,34
16	330	68,51	85,00

Table 1. Classification Experiment. Eigenfolds results in %

Up to now, the tests were made with frames containing both Vocal Folds, but, in reality, each Vocal fold separately can be classified as healthy or pathological.

In the following tests each of the vocal folds were treated independently though the same Training Sets were maintained but, this time with the double of images, as Table 2 shows.

Number of images in Training Set	Number of test images	% of images classified with the correct pathology	% of images correctly classified (healthy or pathological)
25 *2	330 *2	65,10%	95,04%
18*2	330*2	64,23%	90,56%
16*2	330*2	37,41%	80,21%

Table 2. Eigenfolds Results using a Vocal Fold Training Set treated the right Fold and the Left Fold independently

Observing Table 2, It can be seen how, analyzing each fold separately, the results improve up to obtaining 95,04%of success differentiating between those folds that suffer any pathology related to the morphology or not.

Number of images in Training Set	Number of test images	% of images classified with the correct pathology	% of images correctly classified (healthy or pathological)
100*2	900*2	71,80%	92,10%

Table 3. Eigenfolds Results Treating Independently the Right Fold and the Left Fold

Finally, in Table 3 the result considered as final is shown, using a Training Set of 100 images (200 since each vocal fold is studied separately) tested with 900 test images (1800 in reality). The final result of this stage is that 92,1% of the images are correctly classified. Hence, the Analysis and Measurement stage will follow the process correctly.

The Movement detection stage is applied only to those images or sequences of images previously classified as "Morphologically healthy" and which are really healthy or have a vocal Fold paralysis. At this point in the system, the movement in some points of the lower third of each of the vocal fold is measured.

Statistical measures are made, mainly mean, variance, standard deviation, paying special attention to the measure of the variance.
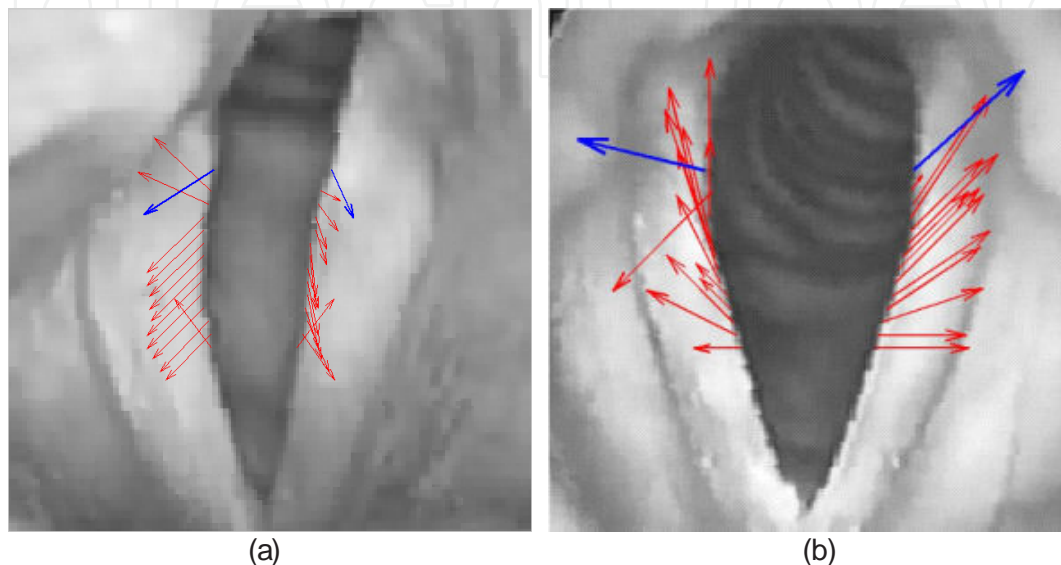

(a)   (b)

Figure 4. Motion vectors in a) a sequence of vocal Folds with paralysis b) a sequence of Healthy Vocal Folds

In Figure 4 two examples with studied frames are showed, with and without paralysis. In red appear all the studied vectors in the Vocal Fold contour and calculated through block matching techniques and in blue the mean vector for all of them.

Observing the vectors in Figure 4a it can be seen, how, one of the folds has less movement, and, what is more, the quantity of motion represented by the vector is unequal. However in Figure 4b the motion represented is more synchronous between both vocal folds.

4.3. Analysis and measurement stage tests

This is the stage where the last tests of the system are made and it might be suggested a diagnosis supported by the objective results obtained over the sequence of images captured by the doctor.

The objective measurements made of the first 15 sequences studied can be appreciated in Table 4. The measurements carried out are intended to identify the morphological pathologies and quantify the size of the previously classified polyp, cyst, nodule or edema.

The measurements are made in pixels and per vocal cord, as that is the only way of doing it. The "0" means that nothing significant was found to measure and therefore confirms that there is no morphological pathology. When intervals appear, this means that there is some kind of

irregularity on the inside edge of the vocal cord, and the area comprising this pathology is measured.

Interval is the term used for the measured area because the vibration of the cords and the image captured by the camera mean that the measurement can vary from one frame to another. The final column in Table 4 provides the algorithm's final decision; that is, when the image analyzed has some kind of morphological pathology and when it does not. There were sequences where quantifying the measurement was scrapped due to its being practically imperceptible: areas below 10 pixels were not taken into account.

5. Conclusions

In this section the compliance of the foreseen objectives will be checked, and the lines that open as a continuation of this research work, whose contributions were demonstrated in the "Proposed System" Chapter will be proposed.

As a summary of the above exposed, a series of conclusions can be obtained, from which a very important one is underscored. It is possible to provide with help to the diagnosis of vocal fold pathologies through the vocal fold digital image processing, and the extracting of objective measures. Furthermore, it has been demonstrated that the interaction with the user it is not necessary (in this case, the doctor specialist in otolaryngology) during the analysis process and the vocal fold digital image processing. An algorithm has been developed and it is divided mainly in two parts: One to carry out the segmentation of the glottal space of the vocal folds and other one to discern between healthy vocal folds, with morphological pathologies, as could be those ones related to the movement.

In connection with the segmentation topic, in chapter for it can be observed how, the 95,07% of the images containing the database were correctly segmented, after the applying of the proposed algorithm. This ratio was calculated without taking into account the small segmentation errors which not affect neither to the subsequent stages of the algorithm nor to the diagnosis (taking into account that the location of this small deviations in the segmentation is the higher part of the glottal space). The results of the segmentation by sequences grouped by pathologies were presented, and in all of the 95% ratio is exceeded, except for those sequences of vocal folds with oedema, where this ratio is below 90%.To differentiate and optimize the processing concerning the detected pathology, the statistic PCA algorithm was developed/applied, to carry out a first classification between vocal folds with a morphological pathology and with the absence of it. It has been demonstrated that the Principal Component Analysis gives good results using a Training Set of significant images of each pathology. In this block, results of 92,1% were obtained with a Training set of 100 images (200 in reality, since the right fold and the left fold were separated).The tests with greater Training Sets are not included, since the processing time increased exponentially, reaching to be of various minutes.

The objective measures carried out allow the specialist to quantify the size of the pathology which is describing/visualizing, being able, this way to provide the patient with more infor-

Sequence number	Otolaryngologist's Diagnosis	Pathology size of right-hand cord (in pixels) Average Size	Standard Deviation	Pathology size of left-hand cord (in pixels) Average Size	Standard Deviation	Final Decision
1	Healthy	0,2	0,0	2,4	1,1	Morphologically Healthy
2	Healthy	0,1	0,0	0,0	0,0	Morphologically Healthy
3	Healthy	0,1	0,0	0,0	0,0	Morphologically Healthy
4	Healthy	5,2	2,9	0,0	0,0	Morphologically Healthy
5	Healthy	0,0	0,0	0,0	0,0	Morphologically Healthy
6	Healthy	0,0	0,0	0,0	0,0	Morphologically Healthy
7	Nodules	95,5	0,8	120	0,4	Nodules
8	Nodules	32,5	1,8	36,3	0,5	Nodules
9	Healthy	0,0	0,0	0,0	0,0	Morphologically Healthy
10	Polyps	97,9	0,8	0,0	0,0	Pathological Right cord
11	Nodules	157	1,1	156	0,7	Nodules
12	Paralysis	0,0	0,0	0,0	0,0	Morphologically Healthy
13	Paralysis	0,0	0,0	0,0	0,0	Morphologically Healthy
14	Edema	0,0	0,0	925	0,6	Pathological Left cord
15	Cyst	0,0	0,0	90,9	1,1	Pathological Left cord

Table 4. Measurements results

mation, to **refine the treatment** or even **measure the evolution** in post-operative processes or in **vocal rehabilitation processes**. The system, without foregoing the doctor, can suggest a diagnosis supported by the results obtained.

From these results we can conclude that the designed algorithm operates properly, and, which is more important, avoiding to the maximum the interaction with the user.

And it is precisely in this interaction with the user where the few commercial softwares which are in the market about the subject-matter before us.

Attending to clinical benefits, the proposed system is based mainly on providing the otolaryngologist with the most sophisticated technological means for the diagnosis, quantification of the extent of the pathology, on measuring the efficiency of the treatment of a surgery, and in a collateral way, on reducing the health costs in this service.

Author details

A. Méndez Zorrilla[*] and B. García Zapirain

*Address all correspondence to: amaia.mendez@deusto.es, mbgarciazapi@deusto.es

DeustoTech Life Unit, DeustoTech Institute of Techology, University of Deusto, Bilbao, Spain

References

[1] Roy, N, Merrill, R. N, Gray, S. D, & Smith, E. M. Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *Laryngoscope, 1151119881995*November, (2005).

[2] Verdolini, K, & Ramig, L. O. Review: occupational risks for voice problems. *Logopedics, Phoniatrics, Vocology, 2613746* (2001).

[3] Becker, W, Naumann, H. H, & Faltz, C. R. Ear, Nose and Throat Diseases. Thieme Medical Publishers, 2nd Edition. (1994).

[4] Roy, N, Merrill, R. N, Thibeault, S, Gray, S. D, & Smith, E. M. Voice disorders in teachers and the general population: effects on work performance, attendance, and future career choices. *Journal of Speech, Language, Hearing Research, 473542551*June, (2004).

[5] Goldman, S, Hargrave, J, Hillman, R, & Holmberg, E. Stress, Anxiety, Somatic Complaints, and Voice Use in Women With Vocal Nodules. *American Journal of Speech-Language Pathology* February (1996). , 5

[6] Cogwell AndersonR., Rusch, M., Pitt, S., Stacy, S., Franke, K. Observed Similarities in Four Adolescents with Paradoxical Vocal Fold Disorder. *The Internet Journal of Pulmonary Medicine*. (2005). , 5(1)

[7] Titze, I. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.* 83, 1536, DOI:10.1121/1.395910. (1998).

[8] Poburka, B. A New Stroboscopy Rating Form. *Journal of Voice*, (1999). , 13(3), 403-413.

[9] Švec, J, & Schutte, H. Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, (1996). (2), 201-205.

[10] Hess, M. M, Ludwigs, M, Kobler, J. B, & Schade, G. Imaging of the larynx-extending the use of stroboscopy-related techniques. *Journal of Logopedics Phoniatrics Vocology*, May (2002). , 27, 50-58.

[11] Allin, S, Galeotti, J, Stetten, G. D, & Dailey, S. H. Enhanced snake based segmentation of vocal folds. *Proc. of IEEE International Symposium on Biomedical Imaging: Macro to Nano*, April, (2004). , 1, 812-815.

[12] Manfredi, C, Bocchi, L, Bianchi, S, Migali, N, & Cantarella, G. Objective vocal fold vibration assessment from videokymographic images. *Biomedical Signal Processing and Control*. April (2006). , 1(2), 129-136.

[13] Fernández GonzálezS., Vázquez de la Iglesia, F., Marqués Girbau, M., & García-Tapia Urrutia, R. Manuel P. García. *Revista Medica Univ. Navarra*, *50* (3), 14-18. (2006).

[14] García, M. *Traité complet du chant.* Paris. (1847).

[15] Nepomuk CzermakJ. *Du laryngoscope.* Paris. (1860).

[16] Stuckrad, H, & Lakatos, I. A new magnifying laryngoscope (epipharyngoscope). *Laryngol Rhinol Otol*, *54* (4), 336-40. (1975).

[17] Oertel, M. Das laryngokospiche untersuchung. *Arch Laryngol Rhinol, 3116* (1985).

[18] Švec, J, & Schutte, H. Videokymography: High-speed line scanning. *Journal of Voice*, 10 (2), 201-205. (1996).

[19] Manfredi, C, Bocchi, L, Cantarella, G, Peretti, G, Guidi, G, & Mezzatesta, C. Objective parameters from videokymographic images: a user-friendly interface. *Proceedings of INTERSPEECH 2007200712221225*

[20] Manfredi, C, Bocchi, L, Cantarella, G, & Peretti, G. Videokymographic image processing: Objective parameters and user-friendly interface. *Biomedical Signal Processing and Control*, (2012). , 7, 192-201.

[21] González, J, Cervera, T, & Miralles, J. L. Análisis Acústico de la voz: Fiabilidad de un conjunto de parámetros multidimensionales. *Acta Otorrinolaringol Esp,53256268* (2002).

[22] Baken, P, & Orlikoff, R. *Clinical Measurement of Speech and Voice* (2 nd ed.). San Diego: Singular Publishing Group. (2000).

[23] Kass, M, Witkin, A, & Terzopoulos, D. Snakes: Active contour models. *International Journal of Computer Vision*. (1988). , 1(4)

[24] Allin, S, Galeotti, J, Stetten, G, & Dailey, S. H. Enhanced snake based segmentation of vocal folds. *In proc. ISBI 2004*. 812- 815 (2004). , 1

[25] Cavalcanti, N, Silva, S, Bresolin, A, Bezerra, H, & Guerreiro, A. Comparative analysis between wavelets for the identification of pathological voices. *Proceedings of the 15th Iberoamerican congress conference on Progress in pattern recognition, image analysis, computer vision, and applications*. Sao Paulo. Brasil. (2010).

[26] Rees, J. M, Regunath, G, Whiteside, S. P, Wadnerkar, M. B, Cowell, P. E, et al. Adaptation of Wavelet Transform analysis to the investigation of biological variations in speech signals. *Medical Engineering & Physics*, 30 (7), 865-871. (2008).

[27] Ertürk, S. Real-Time Digital Image Stabilization Using Kalman Filters. *Real-Time Imaging*, 8 (4), 317-328. (2002).

[28] Di, M, Joo, E. M, & Beng, L. H. A comprehensive study of Kalman filter and extended Kalman filter for target tracking in Wireless Sensor Networks. *Proceedings of International Conference on System, Man and Cibernetics*. (2008). , 2792-2797.

[29] Goffredo, M, Schmid, M, Conforto, S, & Alessio, D. T. A markerless sub-pixel motion estimation technique to reconstruct kinematics and estimate the centre of mass in posturography". *Medical Engineering & Physics*, 28 (7), 719-726. (2006).