SOFTWARE FOR MEASURING AND IMPROVING ESOPHAGEAL VOICES

B. García, J. Vicente

"ESOIMPROVE" Group: Dpt. of Telecommunications
University of Deusto, Bilbao. Spain
mbgarcia@eside.deusto.es

ABSTRACT

The main aim of this paper is to describe a new software program for esophageal speech treatment developed at the University of Deusto. The software tool, named "ESOIMPROVE", allows both to characterize and to modify this speech, and provides the necessary framework to achieve a high quality and intelligible transformed esophageal speech by applying a complete range of sound effects and algorithms. In this field, this tool represents a considerable advance in the study of these voices. The final objective of the project is to obtain an esophageal speech with acceptable levels of quality and intelligibility, and some more works in this direction are being actually developed.

1. INTRODUCTION

A complete set of software tools have been developed for the here presented software package in order to allow the characterization of laryngectomees' esophageal voice through the measurement of its feature parameters: pitch, jitter, shimmer and harmonics-to-noise ratio (HNR), and the application of filters and specific algorithms for the transformation and enhancement of low-quality speech signals.

The main problem of using commercial software for the measuring and characterization of esophageal voices is that most of it is only designed to work with normal voices. Therefore, when used to assess the quality of esophageal speech, the values of the calculated feature parameters are often wrong and inconsistent. The main reason for their incorrect esophageal speech feature measurements is their inability to detect each cycle peak, which is a fundamental aspect of the method used in the application here presented.

2. METHODS

Matlab 6.5 [1] has been the chosen framework for the development and implementation of this software program. One of the latest characteristics of the application is that it can calculate esophageal speech feature parameters such as jitter, shimmer, pitch and HNR i.e. [2]. None of these parameters can be calculated using commercial applications, since the resulting values exceed the normal ranges.

In order to determine the characteristics, that the program should have, a complete study about esophageal speech was done, especially in two aspects: First of all, which features of this speech should be modified to improve quality. And secondly, why commercial programs were unable to measure and modify these parameters, with the obtained results "ESOIMPROVE" was designed, an example of the program's main screen can be appreciated in Figure 1.

2.1. Period Marking

The first step in the development of this software was to design a special algorithm to determine the marks [3] corresponding to each cycle peak, using the Discrete Time Fourier Transform in order to measure the harmonics and correctly identify the periodic cycles of the signal, which constitutes the base for all the algorithm included in the software.

The problem of detecting correctly the fundamental period's maximums is critical to the application. In order to achieve a correct detection, the sonority of small frames is calculated, and a threshold for it established, if the sonority is above this threshold, the considered frame contains a maximum. Once a small range for the position of the peak has been determined, it is possible to apply a conventional maximum detection algorithm in order to detect the exact location of it.

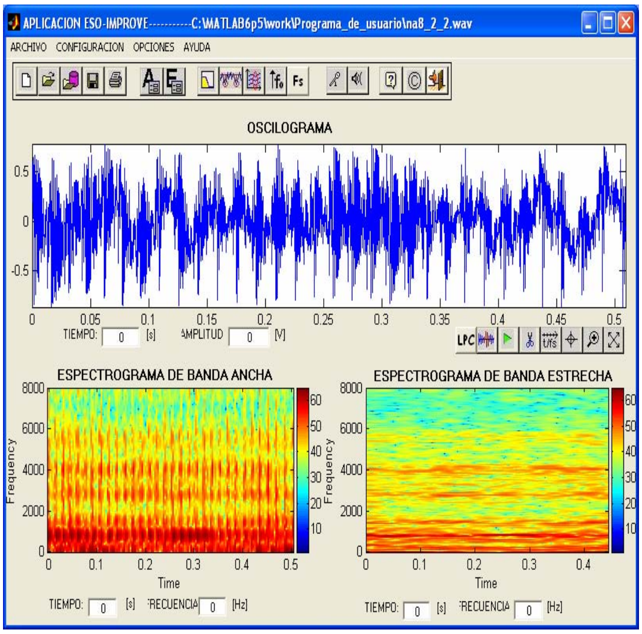

Figure 1: *Main screen*

2.2. Speech Processing

Apart from featuring the speech signal, the software includes some special algorithms designed for its treatment, concretely in two aspects: First, an algorithm that modifies the poles of the system that models the vocal tract making them more stable, an example of how this is done can be seen in Figure 2, and second an algorithm that implements a pitch scaling effect over the speech signal.

The first part called 'Formant evolution reconstruction' modifies the behavior of the main formants along the time, so that they don't suffer big power losses, due to the air pressure reduction in laryngectomized people's voice.

This algorithm analyses the energy of the formants in each instant and detects its sudden decrease, applying then the correction formula:

$$MIN_{\mathrm{mod}} = \frac{\frac{(MAX_1 + MAX_2)}{2} + MIN}{2}$$

After calculating the modified minimum, the points between the two maximums are adjusted to this new value with the help of two correcting straight lines: one for the left, and the other one for the right part of the minimum. With this modification we get the energy loss not to be so heavy.

Also some additional functions for speech transformation are included, such as applying various filters, taking signal pieces, and the possibility of watching the broad and narrowband spectrograms.

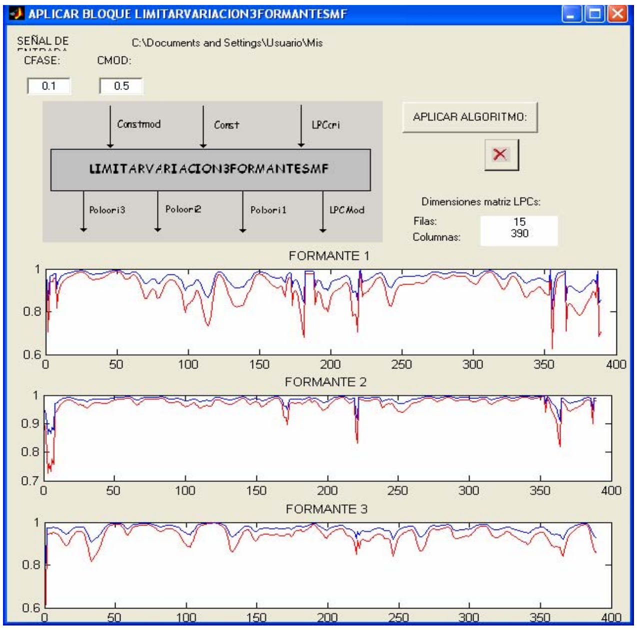

Figure 2: *Poles transformation algorithm screen.*

3. RESULTS

The feature parameters obtained after the application of the mark calculating algorithm match the results obtained with the manual analysis of the spectrograms, validating therefore the mark based measuring method.

The program allows the application of algorithms to transform the low quality esophageal speech signal [4] and improve its intelligibility. These algorithms can be executed step by step, in order to study the partial results at each stage and adjust the parameters that optimize the quality of the resulting speech signal. Figure 2 shows how the user can adjust the parameters the algorithm will apply, by typing the optimal values in the text-boxes.

It is important to remark the utility of the software here presented in the field of medicine, specialists could use it to measure the improvements of the patients after surgery (laryngectomy) and during esophageal speech learning process. It is also very useful for the digital speech processing due to the possibility of applying different effects to an esophageal speech.

4. DISCUSSION

Laryngectomees can't objectively evaluate their voices either after surgery or after the esophageal speech learning period, since there is no valid software application available. Algorithms applied by commercial programs mistake speech signal cycles due to esophageal speech signal's high noise level and to the fact that correct feature parameter values happen to be out of the normal ranges these programs work with.

The described mark calculation method solves these problems, in Figure 3 the parameter calculation screen can be seen, allowing the correct calculation of the esophageal speech feature parameters, in Figure 4 we can see how this application detect correctly the pitch [6] while the commercial programs are unable to measure esophageal voices.

5. CONCLUSIONS

The software application here presented allows the measurement and transformation of esophageal speech with the same feature parameters which define normal speech. Therefore, the objective evaluation of the improvements achieved either with new esophagus based speaking techniques or with new digital esophageal speech signal processing techniques is possible. In such a sense, any digital esophageal speech signal processing algorithm developed in Matlab can be integrated and evaluated against an esophageal speech data base, partial results can be studied through spectrograms and poles-zeros diagrams and the algorithm configuration parameters can be optimized in order to obtain a better speech quality.

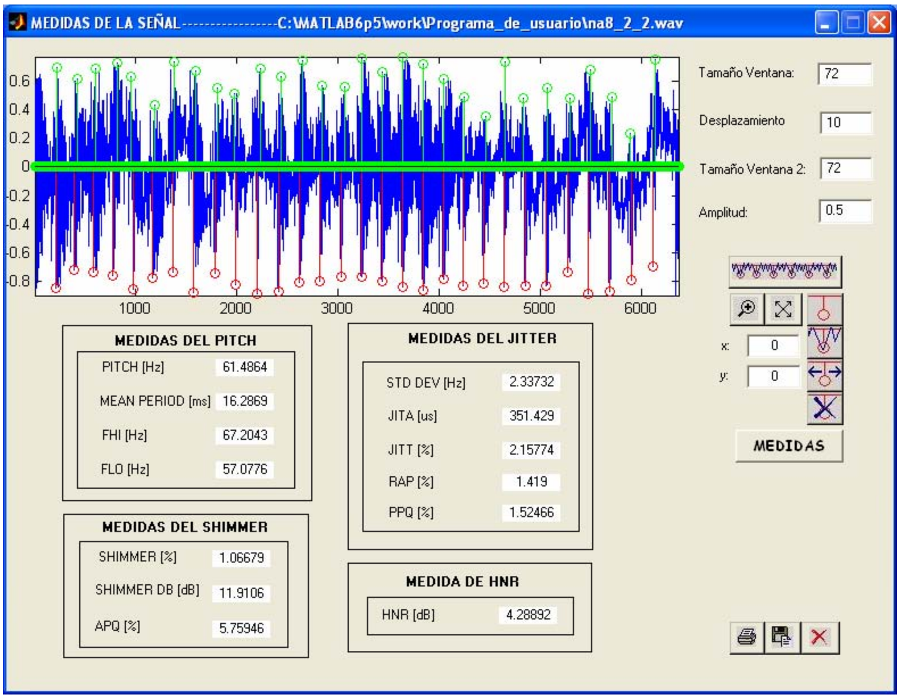

Figure 3: *Parameter calculation screen*

Time (ms)	Original pitch	Pitch measured with a commercial software	Pitch measured with "ESOIMPROVE"
0-20	61Hz	101Hz	59Hz
20-40	70Hz	111Hz	73Hz
40-60	68Hz	125Hz	69Hz
60-80	74Hz	94Hz	74Hz
80-100	55Hz	107Hz	59Hz
100-120	63Hz	98Hz	62Hz

Table 1: *Pitch evolution measured with a commercial program and with "ESOIMPROVE" software*

Voices	Original pitch	Pitch measurement error with a commercial software	Pitch measurement error with "ESOIMPROVE"
Voice 1	61.475Hz	65.34%	3.33%
Voice 2	69.921Hz	68.23%	3.45%
Voice 3	84.409Hz	55.76%	2.76%
Voice 4	58.085Hz	73.83%	1.35%
Voice 5	70.494Hz	64.32%	2.27%
Average error	**-**	**65.498%**	**2.62%**

Table 2: *Percentage error in average pitch measurement for five different voices*

Tables 1 and 2 represent the comparison of final results of pitch calculation between a commercial program and "ESOIMPROVE", as it can easily be seen a commercial program is completely unable to estimate signal's pitch and clearly measures a much higher pitch than real. On the other hand, "ESOIMPROVE" measures correctly pitch's value, with a very small error (2.62% in average). This proofs that specific algorithms work correctly to measure esophageal speech's features. These excellent results can also be extrapolated to other feature parameters such as jitter, shimmer or HNR.

Such methods of measuring constitute the base for effective transformation algorithms, like the ones which have been here explained, because if it is possible to detect speech's irregularities, it won't be difficult to find a way of correcting them. In this sense, as it has been said, two algorithms have been developed, one correcting the characteristic instability of the poles and another one that is able to modify esophageal speech's low pitch. With these two algorithms we are able to enhance substantially alaryngeal speech.

These algorithms can then be used in the fabrication of devices such as telephone adapters, in order to enhance the esophageal speech signal before sending it through the telephone line, so that the receiver has a better intelligibility during the conversation. But also it is possible to think in other type of traditional applications for example: special sound effects designed to be applied in different fields, could also be modified to work with esophageal voices, in this case this effects would not be oriented to regeneration but to modification of speech.

6. ACKNOWLEDGEMENT

The authors wish to thank the support from the University of Deusto during the development of this work and also the valuable contributions of all the scholarships who have taken part in the project. Equally, they sincerely appreciate the help of the "Asociación de Laringectomizados de Bizkaia".

7. REFERENCES

[1] D. G. Childers, *Speech Processing and Synthesis Toolboxes.* John Wiley & Sons, Inc, ISBN: 0-471-34959-3, 2000.

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

[3] J. H. Chen, Y. A. Kao, "Pitch Marking Based on an Adaptable Filter and Peak-Valley Estimation Method," *Computational Linguistics and Chinese Language Processing,* vol. 6, no. 2, pp. 1–12, February 2001.

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

[5] J. C. Brown, M. S. Puckette, "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.