TIME-SPECTRAL TECHNIQUE FOR ESOPHAGEAL VOICE REGENERATION

B.García, J.Vicente
Telecommunications Departament,
ESIDE, University of Deusto,
Avda. Universidades,24 48007,Bilbao.
Tel: +34 944139000 2826; fax: +34 944139101
e-mail: mbgarcia@eside.deusto.es
e-mail: jvicente@eside.deusto.es

ABSTRACT

The esophageal voice that people who have suffered a total laryngectomy use for their communication, can be characterized with the typical voice evaluation parameters such as the pitch, jitter, shimmer and SNR, that reflect its low quality. In the present abstract, an algorithm specifically designed for esophageal voices that increases its SNR from a minimum of 1,8dB to a maximum of 7,1dB is exposed.

1 Introduction

An important part of digital signal processing projects and investigations is voice signal processing. In this field there are two perfectly defined areas: voice synthesis/recognition and, on the other side, digitalized voice signal processing. The proposed algorithm is oriented to the last area, and so it allows to transform esophageal voice obtanied from the recording of people who have suffered a total laryngectomy. The processing that is made is based on the time-spectral study of this kind of voice signals, and it is a new proposal that proofs that digital processing of these signals allows to improve them, specially in terms of SNR [1].

2 Methods

Next, we detail the algorithm and the techniques that have been used for its design. We will take into account Figure 1, where it is shown how voice signals are filtered in the 300Hz-4000Hz band to remove part of the noise incorporated in the signal. The first processing of the algorithm has been called 'Instantaneous spectrum regeneration'. The aim of this block is to modify the spectrum of the signal, strengthening the formants to decrease its masking, due to the non-articulated air that laringectomized people produce speaking. This is achieved using:

- Lineal Prediction Thecniques (LPC coefficients) [2]
- Fast Fourier Transform (FFT)

The second block called 'Formant evolution reconstruction' modifies the behaviour of the principal formants towards time, so that they don't suffer big power losses, due to the air pressurereduction produced by laryngectomized people. This algorithm analyses the energy of the formants in each instant, and detects its suden decrease, applying in this situation the correction of the formula:

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

With this modification we get the energy loss not to be so drastic.

3 Description

Next, it is detailed the process made by each of the blocks mentioned in the algorithm of the figure 5.1. Firstly, it is going to be commented the "Instantaneous Spectrum Regeneration" block. In order to this it will be followed the organigram of the figure 3.1.

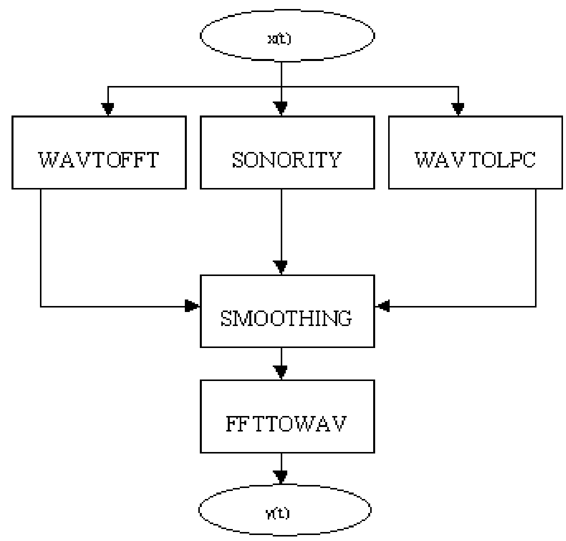


Figure 3.1 "Instantaneous Spectrum Regeneration"algorithm block diagram.

The functionality of each one of the blocks taking part to its realization are the following ones:

3.1 SONORIDAD

This block calculates the sonority of the entering voice signal, it is the same block as the one used in the "pitch" calculation function, implemented by a MATLAB function [3] called 'sonoridad'.

3.2 WAVTOFFT

This block divides the temporary signal in segments of 256 patterns length and 32 patterns displaced and it calculates the FFT of each segment, giving as a result a matrix containing in each of its lines the FFT [4] of a segment. This block is implemented by a MATLAB function called 'wavtofft'.

3.3 WAVTOLPC

This block divides the temporary signal in segments of 256 patterns length and 32 patterns displaced and it calculates the linear prediction coeficients of each segment, giving as a result a matrix containing in each of its lines the linear prediction coeficients [4] of a segment. This block is implemented by a MATLAB function called 'wavtolpc'.

3.4 SUAVIZAR

This block receiving as entering parameters the FFT matriz, the LPC matris and the sonority vector, softens each one of the entering signal instantaneous spectrums using the linear prediction coeficients. Obtaining as a result a matrix containing in each lines the modified FFT. This blocked is implemented by a MATLAB function called 'suavizar'

3.5 FFTTOWAV

This block, receiving a FFT matriz as entering parameter, developes the frequency to time conversion, generaiting a temporary voice signal. This block is implemented by a MATLAB function called 'ffttowav'.

On the other hand, once this block is applied wich is centered in the process on the frecuency axis, the processed esofagic signals are passed through the second block in the algorithm esposed in the figure 3.2.

In this figure the "Formant Evolution Reconstruction" algorithm diagram block can be seen divided in functional blocks. In following explanations it will be described the functionality of each block.

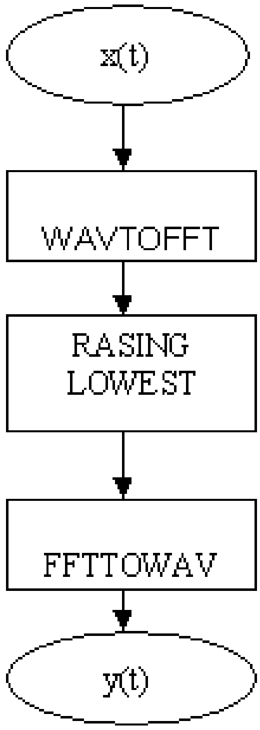


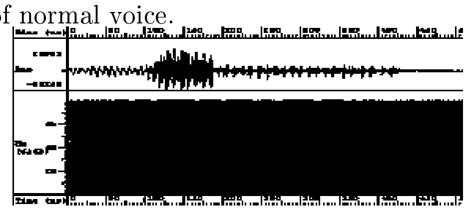
Figura 3.2 Formant Evolution Reconstruction diagram block.

3.6 ELEVA MÍNIMOS

This blocks mades the softening of the formants evolution receiving as entering parameter a matrix containing in each line the FFT of a voice segment. This function takes each of the matrix columns, checks the evolution trough the time of each one of the frecuencies and softens the transitions betwen relative maximums and minimums of each one of the evolutions. Giving back a matrix containing in each line a segment modified FFT.This block is implemented by a MATLAB function called 'elevaminimos'.

4 Results

The final algorithm modifies an input voice signal, transforming it into a voice signal similar to the larynged voice one. This can be checked out with the study of the original and resulting voice signals. In the study that has been made, we have taken into account the following analisys: Spectrograms, oscilograms, and the calculation of SNR. From the results obtained, we check that we can get a voice signal whose parameters are between the ranks of larynged voice, modifying the original voice whose parameters were not in the ranks of normal voice.



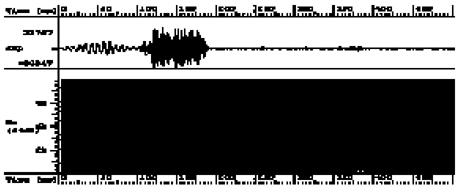


Figure 1.- Spectrograms before and after the processing.

On top of figure 1 we can see the narrow band spectrogram of an esophageal voice, where we apreciate that the phonems can't be easily identified. And in the bottom, we can see the narrow band spectrogram of the modified voice signal, where we can clearly see the apperacance of the characteristic voice formants.

5 Discussion

The algorithm that has been developed for the regeneration of the esophageal voice using digital signal processing techniques, corresponds to the organigram of Figure 2.

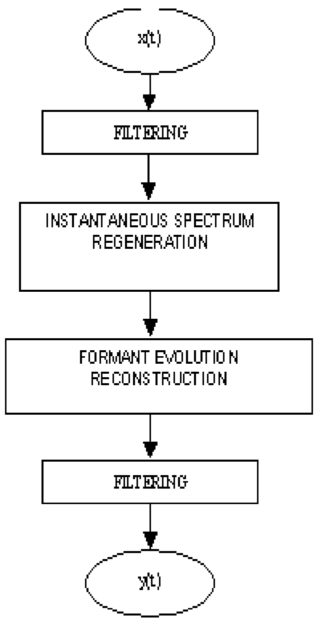


Figure 2.- General organigram of the esophageal voice regeneration algorithm.

As you can see, the voice signal is filtered in the 300-4000 Hz band, before and after the application of the algorithm, so that the possible noise localized outside this band is removed.

Next, the ' Instantaneous Spectrum Regeneration' algorithm is applied. We can numerically evaluate its results in the next table:

PATTERNS		SIGNAL TO NOISE RATIO (dB)		
		AVERAGE	MINIMUM	MAXIMUM
BO1	INICIAL	1.3	-4.1	6.1
	PASO 1	8.4	3.0	12.0
	PASO 2	8.5	3.5	12.5
DE1	INICIAL	0.9	-4.2	-4.6
	PASO 1	2.7	0.4	5.4
	PASO 2	2.8	0.8	5.8
GA1	INICIAL	0.7	-3.5	6.5
	PASO 1	3.4	-1.4	8.0
	PASO 2	3.8	-0.7	8.7
BO2	INICIAL	1.2	-3.5	-10.3
	PASO 1	6.5	-0.9	11.6
	PASO 2	6.5	0	11.2
DE2	INICIAL	2.6	-1.5	13.2
	PASO 1	5.4	-0.6	11.6
	PASO 2	5.5	0.1	11.8
GA2	INICIAL	0.9	-2.6	5.8
	PASO 1	3.4	-2.2	12.9
	PASO 2	3.5	-2.2	13

Table 1.- SNR results before and after the application of the algorithm. The rows corresponding to STEP 1 contain the measures, made with the Praat v 3.9.10 program, on the modified voices and we can clearly appreciate a big improvement in the SNR of the output of the algorithm, with a variation from a minimum of 1.8dB to a maximum of 7.1dB, which reverts on the quality of the obtained voice.

Finally, a second algorithm applied on the previous signal that works on the time axis called 'Formant Evolution Reconstruction' is proposed. The results of the application of this processing are gathered in the rows corresponding to STEP 2, where we also check out how the parameter SNR improves in relation to STEP 1 in five of the six voices studied. However, the 'breathlessness sensation' noticed when you listen to the result is reduced a lot in the six voices.

Everything looks to point that the proposed algorithms can improve the quality of esophageal voice, which will revert not only into its inteligilibility but also into laryngectomized people's mind.

6 Conclusions

Due to the special characteristics of esophageal voice in relation to larynged voice, it is sometimes quite complicated even the measure of the parameters that are used for its characterization such as the "jitter" and "shimmer" [3]. However, the SNR choosed for the evaluation of the proposed algorithm does not cause any problems when we measure it, but the results that are obtained before applying any algorithm are very low, in relation to larynged voice. In this sense, the proposed algorithm cleans the noisy spectral components and enhaces the formants that are quite minimized. In spite of the notable improvement that suposes the application of the algorithm, the possibilities to work on new digital signal processing algorithms are very varied and could be centered to improve the values of other parameters such as the "jitter" or the "shimmer".

References

[1] Fernández Liesa, R., Damborenea Tajada, D. *Acoustic analisis of the normal voice in non-smokers adults*, Acta Otorrinolaring ESP 1999;50(2):1344-141.

[2] J.Bernal, J.Bobadilla, P.Gómez *Reconocimiento de voz y fonética acústica*, 2000.

[3] P.J. BAKEN, Robert f. Orlikoff *Clinical Measurement of Speech and Voice*, 2º Ed. Singular Thonson Learning, 2000.

[4] M. Faundez *Tratamiento Digital de voz e imagen y aplicaciones a la multimedia*, Ed. Marcombo, 2000.

[5] A. Oppenheim *Digital Signal Processing*, Ed. Prentice Hall, 1999.