IMPROVEMENT OF ESOPHAGEAL VOICES' PITCH

B. García, J. Vicente, I. Ruiz, A. Alonso, E. Loyo

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

ABSTRACT

In this paper it is described a new algorithm for esophageal speech regeneration, based on pitch and jitter modification. Traditional phase vocoder and resampling pitch scaling techniques have been used to develop a new adaptive method which scales the low esophageal speech pitch and applies a variable scaling factor significantly reducing its jitter. This method has shown to considerably improve esophageal speech quality, reducing its hoarseness and increasing its intelligibility. The presented algorithm pretends to be an important step forward in the regeneration of esophageal speech.

1. INTRODUCTION

Laryngectomees have to generate their speech by insufflation of the esophagus and vibration of the pharyngoesophageal segment in replacement of the vocal folds. These special conditions make esophageal speech's feature parameter differ from normal speech. This algorithm is part of a set of algorithms intended to correct these parameters.

Among esophageal speech irregularities there are many related to the pitch. The fundamental frequency of the resulting speech signal is lower than the pitch of normal voices, due to the different shape and size of the vibration organs. However, this is not the only problem. The unsustained air flow responsible for the vibration of the pharyngoesophageal segment causes the decrease of the fundamental frequency which does not remain constant over time. Therefore, esophageal speech has a very low pitch and a very high jitter [1].

With the aim of correcting these irregularities, the designed algorithm raises the fundamental frequency and lowers the jitter of esophageal speech. In this aspect the algorithm constitutes a total revolution for the esophageal speech treatment, and sets up many new lines of investigation.

2. METHODS

In order to raise the fundamental frequency of esophageal speech, standard pitch scaling techniques [2] have been modified to achieve an adaptive algorithm that responds to instantaneous pitch variations, modifying its pitch scaling factor and reducing the deviation from the medium pitch. A simplified example of how this is done is shown in Table 1. As it can be seen, if the original pitch is higher than the average one then the applied scaling factor is lower and vice-versa.

	Original pitch	**Applied PSF**	**Modified pitch**
Average	65	1.2	78
1st Period	55	1.418	78
2nd Period	75	1.04	78
3rd Period	65	1.2	78

Table 1: *Examples of the application of the variable PSF.*

In Figure 1 the whole process of the pitch scaling is shown: the first step consists of detecting the signal's peaks, with a special algorithm adapted for esophageal speech. Once the periods have been correctly determined, it is possible to measure the average speech's pitch. In the third and fourth steps the algorithm works period by period, first calculating the pitch scaling factor (PSF) appropriate for the period (the process is explained in Figure 2) and then applying the pitch scaling to it (pitch scaling can be applied in time or in frequency). Finally, the modified signal is reconstructed from the modified periods.

Calculating the variable PSF is not an easy task because if the duration of each period is set to be the same, the naturalness of the resulting speech is significantly reduced. Therefore, the recalculation of the scaling factor involves establishing a maximum deviation from its original value, in order to preserve speech naturalness. The developed algorithm involves a frame by frame calculation [3] of the average and instantaneous pitch, and the application of a different scaling factor according to the deviation from the average pitch in order to significantly reduce these pitch irregularities and, as a consequence, the speech signal's jitter.

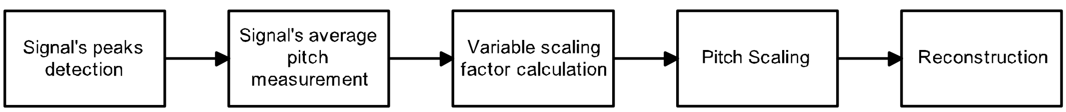

Figure 1: *Block diagram of the pitch scaling algorithm.*

As previously mentioned, the jitter cannot be reduced to zero because the speech would lose naturalness. In order to obtain a clear and natural signal, the variable scaling factor is calculated in the following way (Figure 2): first of all, the pitch of the period is calculated and compared with the average pitch. Secondly, as a result of this comparison, an initial variable factor is calculated; this factor would reduce jitter to zero. In the third step, the difference between the fixed and the variable factor is calculated. If this difference is larger than 0.5 and smaller than 1 the variable scaling factor is limited. This jitter reduction produces a significant enhancement of voice's quality, since speech's characteristics such as hoarseness and reverberance nearly disappear, making the speech much more intelligible. At the same time, the limiting of the scaling factor preserves the naturalness of the speech.

3. RESULTS

Table 2 and 3 show the results obtained after applying the algorithm to esophageal speech samples from a database of voiced patrons recorded from ten different laryngectomee patients. Scaling factors of 1.2 and 1.3 [4], have been used in order to increase the average esophageal speech pitch value from around 60 Hz – 80 Hz to 90 Hz. As it is shown in Table 2, pitch scaling has been achieved in all the samples and jitter has been significantly reduced in both cases.

As shown in the mentioned tables, the pitch of the voices rises according to the expected values. For example, in the first case, if the voice's original pitch was 61.475Hz, with a PSF of 1.2 it should have risen to 73,77, and in fact the real value (74.310) is very close to this value.

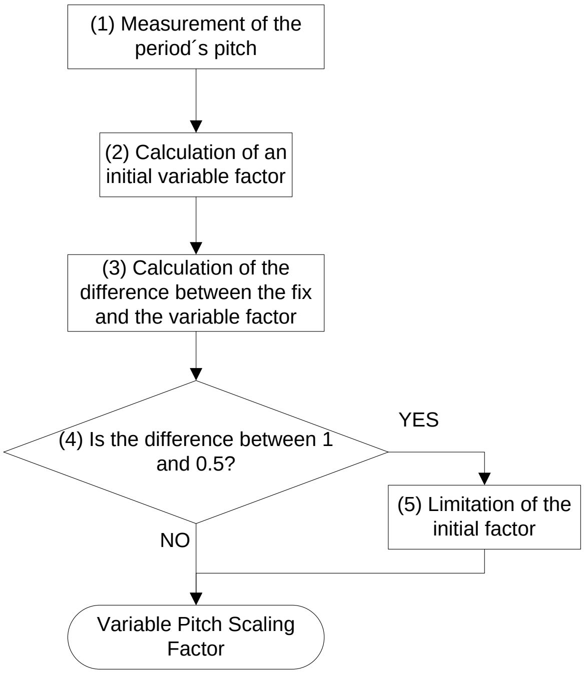

Figure 2: *Flowchart of the variable scaling factor calculation process*

Voices	Originals	PSF = 1.2	PSF = 1.3
Voice 1	61.475	74.310	80.850
Voice 2	104.112	122.137	134.595
Voice 3	69.921	86.184	94.598
Voice 4	58.085	71.577	78.730
Voice 5	60.556	80.301	81.901
Voice 6	84.409	101.124	109.871
Voice 7	70.494	85.383	92.510
Voice 8	62.626	75.959	83.627
Voice 9	61.485	75.553	82.748
Voice 10	58.754	67.408	74.240
Voice 11	58.284	69.534	75.128

Table 2: *Pitch (Hz) resulting of the application of the algorithm to ten samples of a data base*

Voices	Originals	PSF = 1.2	PSF = 1.3
Voice 1	10.333	2.505	2.690
Voice 2	16.349	9.051	6.864
Voice 3	18.339	4.519	6.169
Voice 4	17.960	8.012	9.977
Voice 5	20.302	12.348	10.782
Voice 6	15.635	5.620	5.126
Voice 7	9.037	2.810	4.547
Voice 8	12.312	4.338	3.462
Voice 9	10.747	4.717	3.625
Voice 10	7.020	3.957	4.442
Voice 11	3.536	4.373	2.697

Table 3: *Jitter (%) resulting of the application of the algorithm to ten samples of a data base*

With a scaling factor of 1.2, in 90% of the samples jitter is reduced to a half or even a quarter of its original value. For a scaling factor of 1.3, results are still better since all samples improve their jitter after the application of the algorithm. These results show that the proposed algorithm solves the esophageal speech characteristic low pitch and high jitter problem.

4. DISCUSSION

Improving esophageal speech is a difficult task, especially hard when such an enhancement involves the detection of the speech signal's fundamental periods. Therefore, the design of the adaptive pitch scaling algorithm has involved the development of a method to detect each speech cycle's peak first. The developed technique allows us to obtain the speech signal's average pitch and the pitch of each speech frame. Once these values have been calculated, the algorithm can be applied.

This algorithm is an important step forward in esophageal speech regeneration, since two characteristic problems of esophageal speech, low pitch and high jitter, are solved at the same time.

Future work should be focused on the development of a shimmer correction algorithm, attempting to improve esophageal speech's particularly high shimmer value. The effect of applying this set of algorithms to the excitation signal of voiced phonemes rather than to the speech signal should also be investigated.

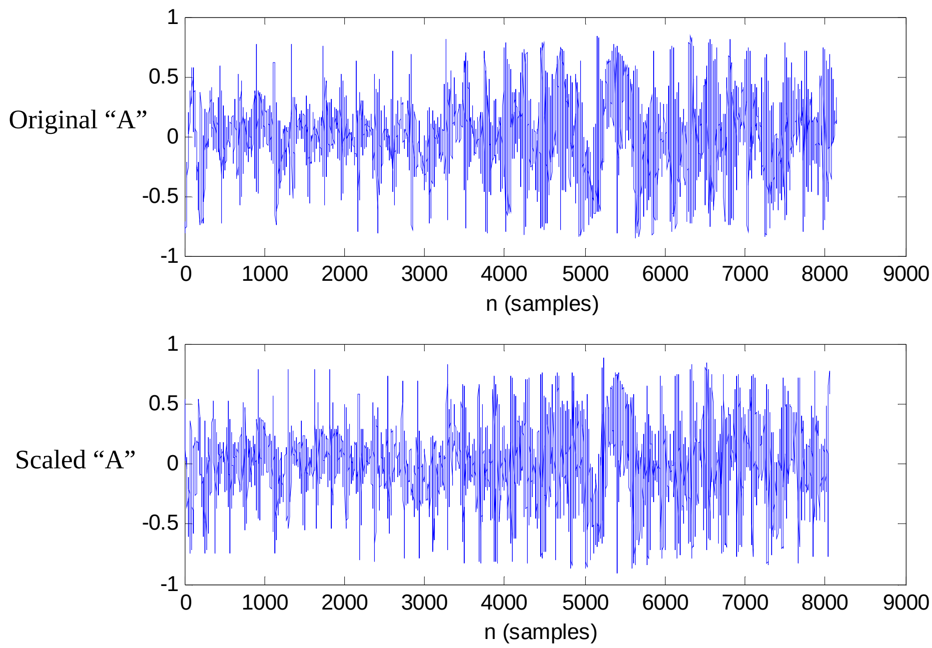

Figure 3: *Original and modified signal*

5. CONCLUSIONS

The presented algorithm has achieved the goals of raising the fundamental frequency of esophageal speech to normal speech's average values and of reducing its jitter level. These two improvements make esophageal speech sound more natural and intelligible. The obtained results show, as it can be seen in figures 3 and 4, that esophageal speech can be enhanced and regenerated to normal values. These results support our research and encourage us to develop new algorithms that correct other esophageal speech irregularities, in order to help laryngectomees recover the voice they lost after their illness.
It has been demonstrated once more, that engineering can make easier the life of those who suffer some kind of "disability," helping them to integrate and build a fairest society.

6. ACKNOWLEDGEMENT

The authors wish to acknowledge the University of Deusto which kindly lent infrastructures and material for this investigation. They would also like to thank all the scholarships that so enthusiastically have collaborated with this project. And especially it cannot be forgotten the help of the members of the "Asociación Vizcaína de Laringectomizados." whose voices constituted the data base for the investigation; without their help it would not be possible to carry out this project.

7. REFERENCES

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

[2] J. Garas, P. C. W. Sommen, "Time/pitch scaling using the Constant-Q Phase Vocoder," in *CSSP Proceedings*, ISBN: 90-73461-15-4, 1998.

[3] 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.

[4] F. Hammer, *Time-scale Modification using the Phase Vocoder. An approach based on deterministic/stochastic component separation in frequency domain*, Diploma Thesis, Institute for Electronic Music and Acoustics (IEM), Graz University of Music and Dramatic Arts, Austria, 2001.