OESOPHAGEAL SPEECH ENHANCEMENT USING POLES STABILIZATION AND KALMAN FILTERING

B. García, I. Ruiz, A. Méndez

Department of Telecommunication, University of Deusto, Bilbao, Spain

ABSTRACT

This paper presents an oesophageal speech enhancement algorithm. Such an exceptionally special type of voice is due to the laryngectomy undergone by those persons with larynx cancer. An oesophageal voice has extremely low intelligibility. This work proposes a method to improve its quality, which consists of stabilizing the transfer function poles of the vocal tract model so as to improve a signal's formants. With this aim, the joint use of two techniques has been applied: on the one hand, an algorithm that transforms the modulus and phase of vocal tract's poles and, on the other hand, a Kalman filtering technique. When using this efficient speech enhancement procedure, the speech signal is usually modelled as an autoregressive (AR) process and represented in the state-space domain. The final speech improvement has been measured with the help of MDVP tools and the HNR parameter.

Index Terms— Oesophageal Voice, Acoustic *S*ignal Analysis, Speech Enhancement, Harmonics to Noise Ratio

1. INTRODUCTION

Patients who have undergone a laryngectomy as a result of larynx cancer have exceptionally low intelligibility. This is due to the removal of their vocal folds, which forces them to use the air flowing through the oesophagus: this is known as oesophageal speech. The characterization parameters for these kinds of oesophageal voices go beyond normal ranges, due to the low quality of the sound itself and its intelligibility.

Low intelligibility is the main problem in both oral and telephone communications with other people. In addition, the noise of this kind of speech signal is especially high. This fact has an extremely negative effect on the objective voice parameters, such as pitch, jitter, shimmer and HNR (Harmonic to Noise Ratio). Thus, it is necessary to process voice signal in order to increase intelligibility. The voice enhancement will be measured by those objective parameters. Therefore, the main aim of this work is to recover the normal range of those parameters, to facilitate the laryngectomized collective communication.

Although the Kalman filter is better known for tracking or linear prediction purposes, it is also recognised as an effective speech enhancement technique, and particularly as a noise suppression technique. Kalman filtering has been used since the late 1990s for voice enhancement purposes with good results [1], [2]. However, no algorithm testing its utility for oesophageal voices has been developed so far. The most important aim of this paper is to present a system that combines an algorithm designed for stabilizing the poles of the vocal tract model with Kalman filtering algorithms, in order to apply it to this type of pathological voice.

Naturally, this work is not only valid for oesophageal voices, but also for patients who suffer from vocal disorders, such as hoarseness. Oesophageal voices are just the most grievous among these pathologies.

Speech enhancement improves the quality and intelligibility of voice communication. Applications such as mobile phones, teleconference systems, hearing aids, voice coders and automatic speech recognition need standard quality signals to function correctly.

2. OBJECTIVES

The main goal is the improvement of oesophageal speech quality in order to make it more intelligible.

From this general aim, a series of specific objectives come into consideration:

- To stabilize the three main formants of the speech signal.
- To reduce the breathing noise inherent in oesophageal noise.
- To improve the HNR [3] parameter. This is related to intelligibility, and therefore, to speech quality.

3. METHODOLOGY

The general idea of the algorithm presented in this work is to filter the pathological speech signal to obtain a less noisily corrupted one. There are many algorithms that enhance speech, even using Kalman filters; but none of them have been applied to this type of severely pathological voices. To achieve the paper's main objective, several mathematical tools have been used:

3.1 Kalman Filter (KF)

Given the past and present observations, Kalman filtering is able to obtain the optimum estimate of the state, due to its recursive method. When using the KF, speech and noise are usually modelled as an autoregressive (AR) approach. Kalman filtering used for speech enhancement was first proposed in [1], extending it to the more realistic coloured noise case in [4]. More than a few Kalman filter implementations have been proposed for speech enhancement, some focusing on speech modelling aspects [5], [6] and others on parameter estimation [7],[8]. Figure 1 provides a general idea of the problem.

As the Kalman filter is a well known, efficient tool, only specific equations will be presented.

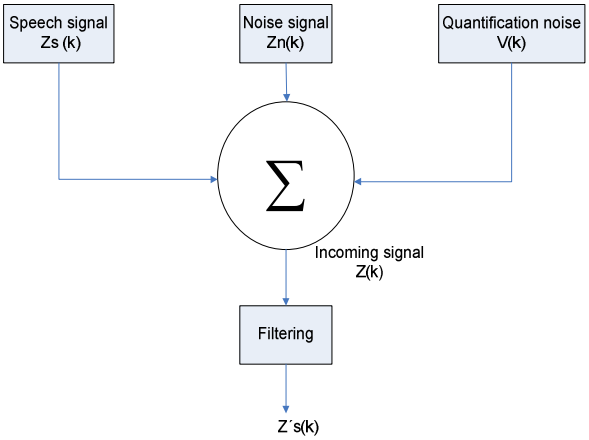

Figure 1 – General structure of the Kalman Filter algorithm

The state transition matrix A (k) is chosen to be diagonal, because, for uncorrelated speech and noise process processing, variations in one process should not influence variations in the other.

3.2 Autoregressive model (AR)

Voice must be characterized in order to obtain the speech parameters. This fundamental task is performed with the help of the autoregressive (AR) model approach. This model (similar to the LPC) is used to obtain the parameters of both signals: speech and additional noise. The model equation is described in (2), where v (t) is a unit-variance zero-mean coloured noise, the system poles a_i and the zero b:

$$y(t) = \sum_{i=1}^{n} a_i \cdot y(t-i) + b \cdot v(t), b \geq 0 \qquad (1)$$

The n represents the system order; in this case the algorithm is implemented with an order of 16, to achieve a compromise between complexity and efficiency.

3.3 Harmonic to Noise Ratio (HNR)

The HNR is a general evaluation of noise present in the analyzed signal. It is defined as (2), with $r_p(0)$ and $r_{ap}(0)$ being the respective energies of the periodic and aperiodic components:

$$HNR = \frac{r_p(0)}{r_{ap}(0)} \qquad (2)$$

The measures have been made with the help of MDVP from Kay Electronics [9], [10], important software that gives good estimations of a signal's parameters. However, this software does not specialize in pathological speech. To solve this problem, the voiced period marks are needed to calculate the pitch and then the HNR have been manually introduced, one by one, on each signal.

The MDVP calculates the HNR as the average ratio of the harmonic spectral energy in the 70-4500 Hz frequency range and the enharmonic spectral energy in the 1500-4500 Hz frequency range.

Several papers [3], [11] have proved the validity of this parameter when measuring speech quality. HNR is computed using a pitch-synchronous frequency-domain method. Low values of HNR are interpreted as increased spectral noise, which may be due to amplitude and frequency variations (i.e., shimmer and jitter), turbulent noise, sub-harmonic components and/or voice breaks.

4. DESIGN

What follows is a short description of the proposed oesophageal speech enhancement algorithm. As can be seen in Figure 2, the system consists of two blocks: a first algorithm which works directly with a Kalman Filter and a second stage based on stabilizing the system's poles.

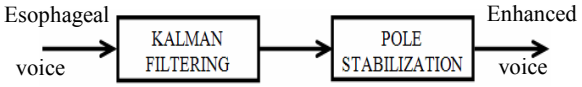

Figure 2– General Block Diagram

The second block in Figure 2 represents the algorithm responsible for analyzing and modifying the poles of the system modelized by the vocal tract. It works with an esophageal voice signal from which the excitation has been separated from the tract, and it calculates the evolution of modulus and phase of each formant of the vowel modifying such poles.

The stabilization of the first three formants is applied in those values of the vowel which is being enhanced by means of the modification of the first three poles, following these steps:

1) Calculation of the mean value of modulus and phase of each pole through the vocal signal

2) Calculation of the maximum deviations relative to the mean modulus and phase of the first three poles.
3) Whether the deviations exceed a certain threshold is analyzed, if so the modulus correction is applied:

ModulusModif=modulus+((1-modulus)*ConstMod);

and the phase correction:

AngModif=Angle-(ContPhase*(Angle+MeanPhase));

being the correction implanted by means of "ConstMod" and "ConstPhase" parameters which can be adjusted for each voice.
4) Reconstruction of the filter that modelizes the vocal tract with the new poles of the system corrected and stabilized.

The first block in Figure 2 represents the Kalman Filtering applied to the signal.

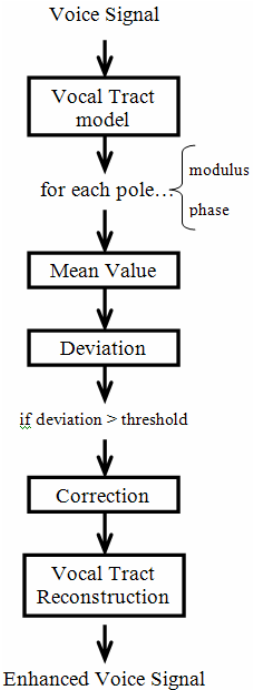

Figure 3– Pole Stabilization Block Diagram

As shown in figure 1, the incoming real signals are the speech and noise. The noise could be randomly created; but the results are much better using noise obtained from an oesophageal voice during periods of silence. Quantification noise is an unwanted but also unavoidable problem of every digital system. In this case, as both signals are quantified separately, this effect can be minimized without much computational load.

As the signals we are working with are of very low quality, it is recommended that a pre-processing step is taken to eliminate unwanted signal components. In this case, since the signals show a low frequency modulation, a high pass Chevychev filter would be a good choice.

The first important step of the algorithm is to obtain the AR parameters of both signals. This is probably the fundamental and most important step because those parameters form all Kalman matrixes. As can be observed in the next figure (2), the AR parameters are re-estimated to adapt the filter to the signal's necessities.

The main advantage of the KF is its properties of predicting and correcting, because it can work even though the initial information might be erroneous.

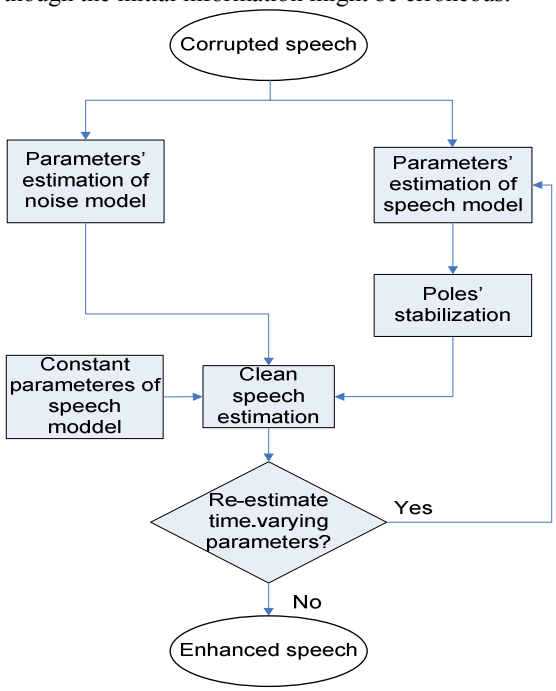

Figure 4– Estimating the AR parameters of the signals

So as to emphasize the correction, the Kalman filter is passed twice. The forward run already filters the signal, but with a backward run results are much better. This improvement is due to the correction on the first pass results for the same signal section.

5. RESULTS

Here the results are described. These results are obtained applying the algorithm to maintained oesophageal /a/ phoneme. The main reason for doing so is the fact that /a/ is the most commonly used phoneme in most languages. As a result, improving their quality will have an important impact on the intelligibility of the whole conversation.

In all cases an increase in harmonics to noise ratio has been achieved. For example, the value of "a1" signal is 5.001dB before processing and 1.701dB

afterwards. The increase in improved oesophageal signal, in this case, has been 3.3dB. The fourth column in table 1 shows the enhancement of HNR (dB) before and after processing. It can be appreciated that the improvement in HNR (dB) ranges from 0.916dB, for "a10" signal, to 4.462dB, for "a4". Taking into account all the database, the average HNR improvement (dB) is 2.941dB.

Phoneme	*Original phoneme HNR (dB)*	*Kalman+ Stabilization HNR (dB)*	*HNR (dB) increase*
a1	-5.001	-1.701	**3.300**
a2	0.549	1.656	**1.107**
a3	-3.684	-2.219	**1.465**
a4	-4.901	-0.668	**4.462**
a5	-6.375	-2.631	**3.744**
a6	-6.803	-3.159	**3.644**
a7	-6.389	-4.451	**1.938**
a8	-8.724	-5.615	**3.109**
a9	-3.737	-0.040	**3.697**
a10	0.930	1.846	**0.916**
Average			**2.941**

Table 1: HNR measures with the /a/ phonemes.

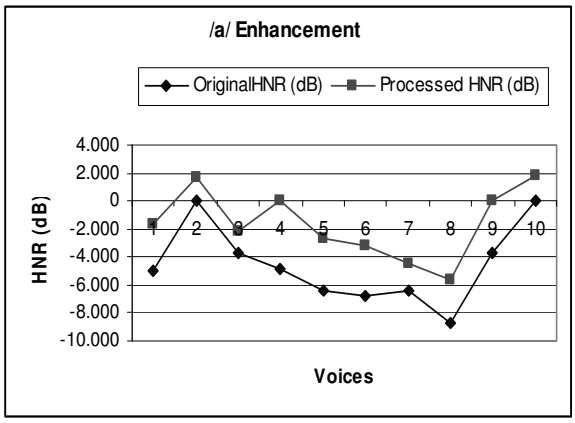

Figure 5: HNR before and after processing the /a/ phoneme.

As shown in figure 5, both curves represent the value of HNR (dB), before and after applying the algorithm presented in this paper, for each 10 signals of database. The values of these graphics are taken from the two central columns in table 1. As shown in figure 5, the values of the processed curve are always above the original curve values. Therefore, the results are satisfactory in 100% of cases.

6. CONCLUSIONS

It can be concluded that the aimed objectives have been achieved because of the fact that the algorithm is very suitable. This statement is based on the improvement of every case studied. The processed oesophageal signal has been improved 2.941dB on average from the original.

Therefore, both pole stabilization improvement and kalman filtering are suitable techniques in the speech enhancement context.

Due to the favourable results achieved, new tests are being carried out in a real context in order to apply this algorithm in a DSP board for real time applications improving oesophageal signal in telephone communications.

7. ACKNOWLEDGEMENTS

This research was partially carried out under grant TEC2006-12887-C02-02 from the Ministry of Science and Technology of Spain. The authors also wish to thank the support of INRIA's EUROMED 2007 project.

8. REFERENCES

[1] K. K. Paliwal and A. Basu, "A Speech Enhancement Method Based on Kalman Filtering," in *Proc. ICASSP'87*, pp. 177–180

[2] M. Gabrea, "Robust Adaptive Kalman Filtering-based Speech Enhancement Algorithm," in *Proc. ICASSP'04*, pp. 301–304.

[3] F. Severin, B. Bozkurt, T. Dutoit, "HNR extraction in voiced speech, oriented towards voice quality analysis", *Proc. EUSIPCO'05*, Antalya,Turkey.

[4] J. Gibson, B. Koo, and S. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1732–1742, Aug. 1991.

[5] Z. Goh, K.-C. Tan, and B. Tan, "Kalman-filtering speech enhancement method based on a voiced-unvoiced speech model," *IEEE Trans. Speech, Audio Process.*, vol. 7, no. 5, pp. 510–524, Sep. 1999.

[6] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and sequential Kalman filter-based speech enhancement algorithms," *IEEE Trans. Speech, Audio Process.*, vol. 6, no. 4, pp. 373–385, Jul. 1998.

[7] P. Sorqvist, P. Handel, and B. Ottersten, "Kalman filtering for low distortion speech enhancement in mobile communication," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, Munich, Germany, Apr. 1997, pp. 1219–1222.

[8] Biddle A, Watson L, Hooper C, et al. "Criteria for Determining Disability in Speech-Language Disorders". Evidence Report/Technology Assessment No. 52 AHRQ Publication No. 02-E010. Rockville, MD: Agency for Healthcare Research and Quality. January 2002.

[9] Zelcer S, Henri C, Tewfi k TL, Mazer B, "Multidimensional voice program analysis (MDVP) and the diagnosis of pediatric vocal cord dysfunction". Ann Allergy Asthma Immunol, 2002; 88: 601–8.