Restoration of hyperbaric speech by correction of the formants and the pitch

L. Charonnat, M. Guitton, J. Crestel, G. Allée

ENSSAT
6, rue de kérampont, BP 447
22305 Lannion
France

ABSTRACT

This paper describes an hyperbaric speech processing algorithm combining a restoration of the formants position and a correction of the pitch. The pitch is corrected using an algorithm of time-scale modification associated to an oversampling module. This operation does not only perform a shift of the fundamental frequency, but induces a shift of the other frequencies of the signal. This shift, as well as the formants shift due to the hyperbaric environment, is corrected by the formants restoration module, based on the linear speech production model.

1. INTRODUCTION

During deep water dives, professional divers are exposed to high pressures requiring them to breath a synthetic gas mixture. The physical properties of the breathing mixtures (sound velocity and density) induce a dramatic distortion of their speech called hyperbaric speech. The most important effect is an upward shift of the formant frequencies which is modelised by the Fant formula:

$$F_h^2 = c^2(F_a^2 + (\rho - 1)F_w^2)$$

F_h and F_a refer to one formant of a sound produced respectively in the hyperbaric environment and in the air. The coefficients c and ρ are the ratio of the sound velocity and the density of the two environments. F_w is the resonance frequency of the closed-lips vocal tract. The shift of the formants has been studied for many years and several correction algorithms have been realised. The intelligibility of the speech has been restored but the quality is far to be good enough because of the other alterations such as:

- a reduction of the pitch periods
- a broadening of the formants bandwidth
- a fading of the intensity of unvoiced sounds relative to the intensity of voiced sounds

The pitch correction would improve the speaker recognition task, especially when the speaker modifies it in a large rate. A contribution to a more natural voice could also be done using a formant correction module which allows a modification of the bandwidth and the central frequency of a formant at different ratio.

2. FORMANT MODIFICATION

The formant modification is based on the linear predictive analysis of the speech waveform. An p-order analysis identifies an all-poles system of p/2 poles of complex conjugate pairs to the vocal tract filter. Each pair P_i is linked to a formant by the expressions:

$$B_i = -(F_e/\pi)\ln|P_i|$$

$$F_i = (F_e/2\pi)\varphi_{P_i}$$

$|P_i|$ is the module of P_i and φ_i, its argument. B_i and F_i are respectively the bandwidth and the central frequency of the formant. F_e is the sampling rate. For a sharp formant, the module of the pole is closed to one, and its amplitude is written:

$$|H(Z)|_{Z=\exp(\varphi_{P_i})} = \frac{|r_i|}{1-|P_i|}$$

r_i is the residue of P_i.

This three expressions show that the three main parameters of a formant (central frequency, bandwidth and amplitude) can be independently modified by acting on the module, the angle or the residue of the corresponding pole.

2.1 Correcting the central frequency

The central frequencies of the formants of a sound produced in the hyperbaric environment are corrected by changing the angles of the poles of the transfer function:

$$\varphi_a = \sqrt{\frac{\varphi_h^2}{c^2} - \varphi_w^2 \cdot (\rho - 1)}$$

The suffix *a* and *h* refer to the air and hyperbaric environments. The angle φ_w, representing the contribution of the resonance frequency of the closed-lips vocal tract, is computed with:

$$\varphi_w = 2\pi\, F_w / F_e$$

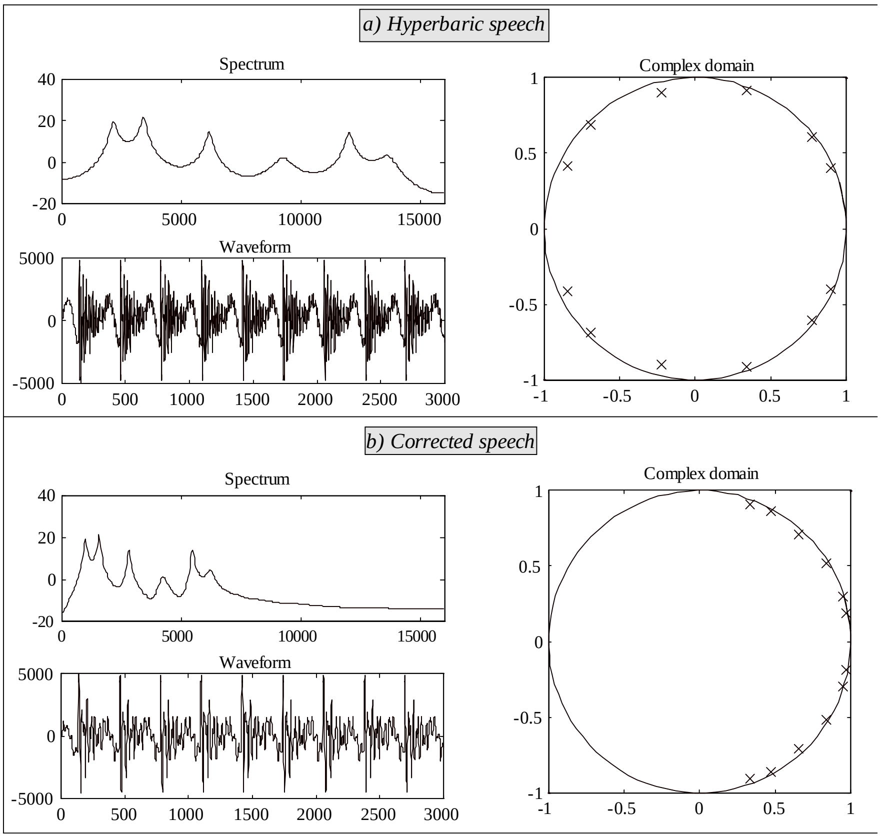


Figure 1: formants correction of a synthetic vowel

2.2 Correcting the bandwidth

Several assumptions have been made about the influence of the environment on the bandwidths ([1] and [2]). Some tests, realized on real signals, have shown that the best quality is obtained when the bandwidth formants is reduced by c [3]:

$$|P_a| = \frac{|P_h|}{c}$$

To ensure that the amplitude is unchanged, the residue must be modified by:

$$r_a = r_h \cdot \frac{1-|P_a|}{1-|P_h|}$$

The figure 1 shows the application of the formants correction method on a synthetic signal for a depth of -252 meters (c=2,2).

3. PITCH CORRECTION

During their underwater experiences, divers are changing their pitch. Most of them have a period reduction about 20%, but some divers modify it in a complete different way (some divers increase the period!) [4]. For their safety and for the quality of the communication it is important to be able to restore their original pitch.

Few methods of pitch modification are available, but the quality of their results is often obtained thanks to a high complexity [5]. A trivial way to change the pitch of a signal is to interpolate (or decimate) the temporal signal and then reduce (or increase) its duration. This method is not usually used because it also modifies the formants of the speech signal. In our case, the shift of the formants due to the pitch modification can be corrected with the formants shift produced by the hyperbaric environment.

Several methods have been tested and the choice has been fixed on an algorithm developed by Roucos and Wilgus [6]. Their method has been slightly modified and some measurements have been done on it to adapt the choice of the coefficients to the hyperbaric speech.

3.1 The time-scale modification algorithm

In 1985, Roucos and Wilgus have written a very efficient algorithm of time-scale modification called SOLA. The principle of this algorithm is to extract some blocks of signal at a specific rhythm (every *Sa* samples) and to add them at another rhythm (every *Ss* samples). If *Sa* is greater than *Ss* the signal duration is reduced by *Sa/Ss*, if Sa is smaller than Ss the signal duration is increased *Ss/Sa*.

However, it is necessary to adjust *Ss* at each step to ensure a synchronous addition between the analysis window and the synthesis window (figure 2). The shift of *Ss* (*ki*) is determined by the maximization of the crosscorrelation of the two windows.

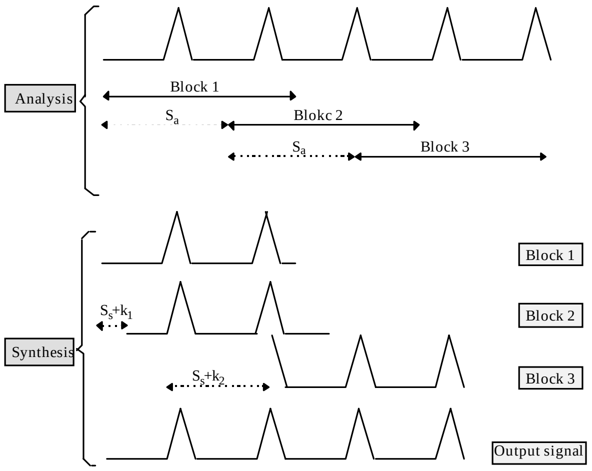


Figure 2: PSOLA method

Let *x(n)* be a speech signal and *h(n)* a window of length *N*. The block *m* is:

$$x_h(n) = h(n - mS_a)x(n)$$

The output signal is obtained by overlap and addition of the blocks:

$$y(n) = y(n) + h(n - mS_s - k)x_h(n - m(S_s - S_a) - k)$$

k is the position of the maximum of the expression:

$$\rho_{xx_h}(k) = \frac{\sum_{n=mS_s}^{mS_s+L} y(n)\ x_h(n - m(S_s - S_a) - k)}{\left[\sum_{n=mS_s}^{mS_s+L} y(n)^2 \sum_{n=mS_s}^{mS_s+L} (x_h(n - m(S_s - S_a) - k))^2\right]^{1/2}}$$

A normalize waveform is created to remove the windows effects:

$$c(n) = c(n) + h(n - mS_s - k)$$

The output signal is then divided by the normalize waveform.

3.2 Application to the hyperbaric speech

Before applying the algorithm to the hyperbaric case, we can notice that a simplification could be introduced to the algorithm: instead of adding the shift *ki* to *Ss*, it is better to subtract it to *Sa*; in this way, the normalize waveform is constant and the last step of the algorithm can be removed.

Considering the specific properties of the hyperbaric speech, it is necessary, to ensure an efficient implementation of the algorithm, to optimize the choice of the parameters for the hyperbaric case.

All the measurements have been done on a synthetic signal representing a sound produced at -252 meters sampled at 44,1 kHz. The expected reduction ratio is 0.8.

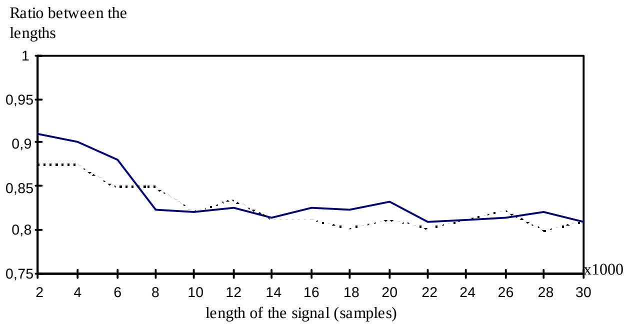


Figure 3: effect of the length of the signal

The first set of measurements (figure 3) shows that the signal must be long enough (10 000 samples, i.e. 250ms) to guarantee an error smaller than 5% on the ratio.

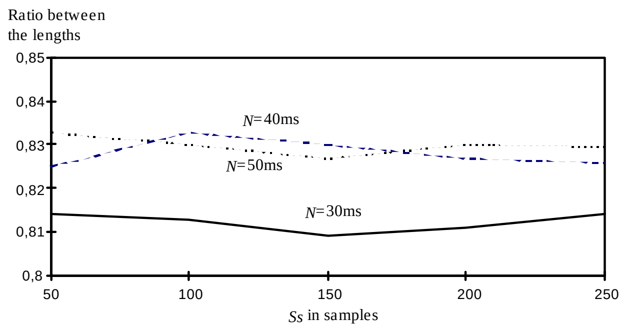


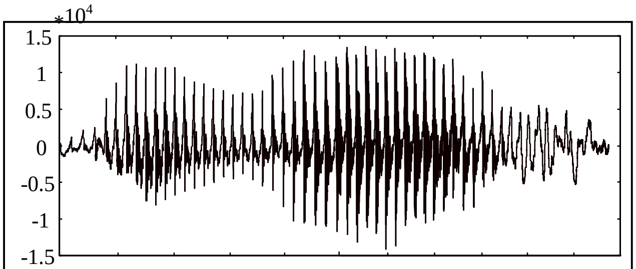
Figure 4: effect of the length of the window and effect of *Ss*

The second one (figure 4) shows that it is better to use a small window (but the window must be longer than one period) and that the choice of *Ss* has a very poor influence on the ratio.

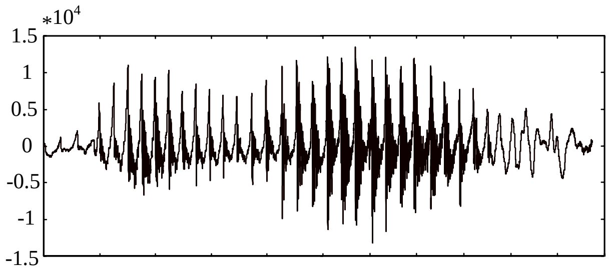
3.3 Application to the modification of the pitch

The modification of the pitch period is realized by a modification of the duration followed by an interpolation or a decimation. It results in a signal with the same duration as the original one but with a new pitch period.

The figure 5 gives an example of an application on a real signal recorded at -252 meters. The sampling rate is 44100 Hz. A reduction of 30% is performed using an Hanning window of 1024 samples.



a) input signal



b) pitch reduction of 30%

Figure 5: application to an hyperbaric speech signal

4. COMBINED CORRECTION OF THE FORMANTS AND THE PITCH PERIOD

The pitch period modification induces a shift of the formants which must be corrected as well as the shift due to the hyperbaric environment.

The Fant formula shows that the effect of the environment on the formants can be interpreted as the superposing of a linear shift of all the frequencies and an nonlinear shift due to the gas density, acting only on the lower frequencies. Regardless of the nonlinear shift, the Fant formula is reduced to $F_h = c_e F_a$ where c_e is the ratio of the sound velocity of the two the environments.

Let c_p be the correction coefficient of the pitch period:

$$P_a = c_p P_h$$

where P_a and P_h design the fundamental frequencies in air and in hyperbaric environment.

The contribution of the two shifts require to correct the formants by a new coefficient: $c=c_p\, c_f$.

5. CONCLUSION

Research on hyperbaric speech has given rise to a large selection of algorithms to correct the formants shift due to the environment, but they do not take in consideration the other distortions which alter, however, the sound quality. It is time now to focus on this small effects of the environment to ensure a better communication between all the people implied in deep water dives.

The difficulty is to improve the correction of the hyperbaric speech using simple enough methods to guarantee a real-time application. That is why we have chosen to apply a time-scale modification algorithm to modify the pitch of the signal using the formants restoration module to correct the distortions induced by this method. The result is a good quality pitch modification which allow a better recognition of the diver.

[1] Richards "Helium speech enhancement using the Short Time Fourier Transform" Ph. D. Georgia Institute of Technology, 1982.

[2] Belcher, Hatlestad "Formant frequencies, bandwiths and *Q*s in helium speech" JASA vol 74, N°2, pp 428-432, 1983.

[3] Charonnat, Guitton, Crestel "Un traitement de la parole hyperbare par corrections séparables des fréquences et des bandes passantes", 4eme Congrès francais d'accoustique, Marseille, avril 1997.

[4] Guitton, Crestel, Charonnat "Pitch and elocution rate of diver's speech", Eurospeech'95 pp 2035-2038.

[5] Allée "Correction du pitch dans le cadre de la parole hyperbare", rapport de DEA, Rennes 1996.

[6] Roucos, Wilgus "High quality time-scale modification for speech", pp 493-496, ICASSP 1985.