A PROBABILISTIC PITCH ESTIMATION METHOD WITH COMBINATION OF SEVERAL TECHNIQUES

G.Bruno, M.G.Di Benedetto, A.Gilio
Universita' di Roma

Abstract

In this paper, a bayesian approach to pitch estimation is presented. The pitch frequency is assumed to be a random variable. At step k, some detections of the pitch F_k, obtained by different classical procedures, are used to modify the a priori pdf of F_k. In this way, the estimation of F_k is made by choosing the expected value of the a posteriori pdf of F_k. In addition, the previous measurements are used to update the a priori pdf of F_k for the next step. The results obtained by the application of this algorithm, considering the combination of two fast techniques of pitch detection, are very satisfactory and better than those obtained by applying each method separately.

1- Model

With reference to fig.1, let f_k indicate the true value of the pitch frequency corresponding to the k^{th} segment of the speech signal, which is supposed to be voiced; in addition, let f_k be obtained by the relation:

$$f_k = f + e_k \tag{1}$$

where f is the expected value of the pitch frequency of the speaker and e_k is an error. In expression (1), the quantities f, e_k and therefore f_k, will be treated as determinations of random variables, which will be indicated by capital letters. In particular, one assumes:

1- that the sequence $[E_k]$ be stationary and memoryless;
2- that F ed E_k be stochastically independent.

The estimation of the pitch frequency $\hat{f}_k$ is made by combining the results of the two different measurement procedures which will be described in section 4 and indicated by a_k and b_k in fig.1, with the a priori information available concerning F_k in position k. In particular, the quantities measured a_k e b_k are also considered to be determinations of the random variables obtained from f_k by means of the relations:

$$a_k = f_k + q_k \quad ; \quad b_k = f_k + r_k \tag{2}$$

in which $[Q_k]$ and $[R_k]$ are supposed to be stationary, stochastically indepedent reciprocally and of $[E_k]$,$[F_k]$ and F. In addition, the quantities a_k and b_k are used to update the pdf of F, necessary for the estimation of f_{k+1}. For the sake of simplicity, all the pdf updated on the basis of the measurements made in the previous positions will be indicated by the index k, without explicetly mentioning this condition.

As regards the estimation of f_k, the Bayes criterion is used with a quadratic cost function for the estimation error. Consequently, one has:

$$\hat{f}_k = [F_k/a_k,b_k] = \int_{-\infty}^{+\infty} f_k \cdot p_k(f_k/a_k,b_k) \cdot df_k \tag{3}$$

where from (2) one gets:

$$p_k(f_k/a_k,b_k) = \frac{p(a_k,b_k/f_k) \cdot p_k(f_k)}{\int_{-\infty}^{+\infty} p(a_k,b_k/f_k) \cdot p_k(f_k) \cdot df_k} = \frac{p(a_k/f_k) \cdot p(b_k/f_k) \cdot p_k(f_k)}{\int_{-\infty}^{+\infty} p(a_k,b_k/f_k) \cdot p_k(f_k) \cdot df_k} \tag{4}$$

$$p_k(f_k) = \int_{-\infty}^{+\infty} p_k(f_k,f) \cdot df = \int_{-\infty}^{+\infty} p_k(f_k/f) \cdot p_k(f) \cdot df = \int_{-\infty}^{+\infty} p(f_k/f) \cdot p_k(f) \cdot df \tag{5}$$

as, due to the hypothesis made, the pdf of E_k is not required to be updated and therefore, conditioned to the event F=f, the random variable F_k is independent of the random variables $A_1,B_1, \ldots, A_{k-1},B_{k-1}$. As regards updating, from equations (4) and (5) it is evident that the only quantity which must be updated is $p_k(f)$. Using for this purpose the measuraments a_k and b_k, one gets:

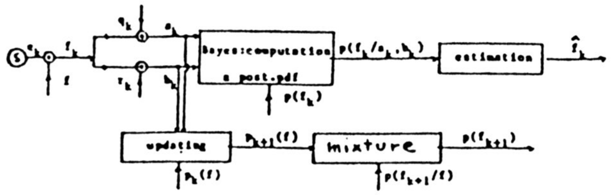

Fig.1 Algorithm proposed

$$p_{k+1}(f) \triangleq p_k(f/a_k,b_k) = \frac{p(a_k,b_k/f).p_k(f)}{\int_{-\infty}^{+\infty} p(a_k,b_k/f).p_k(f).df} \tag{6}$$

and, under the hypothesis that:

$$p(a_k,b_k/f,f_k) = p(a_k,b_k/f_k) \tag{7}$$

the following relation results:

$$p(a_k,b_k/f) \int_{-\infty}^{+\infty} p(a_k,b_k,f_k/f).df_k \int_{-\infty}^{+\infty} p(a_k,b_k/f,f_k).p(f_k/f).df_k =$$
$$\int_{-\infty}^{+\infty} p(a_k/f_k).p(b_k/f_k).p(f_k/f).df_k \int_{-\infty}^{+\infty} p(a_k/f_k).p(b_k/f_k)p_E(f_k-f).df_k \tag{8}$$

Finally, the estimation of f_k is made, when a_k, b_k and the a priori pdf $p_k(f)$ are known, through relation (4), by using relation (5). In addition, the updating of $p_k(f)$ is made from relation (6), by using relation (8).

2- Initial distributions

It_is assumed that:

$$p_{E_k}(e) \triangleq p_E(e) = g(e;0,s^2) \tag{9}$$
$$p(a_k/f_k) = g(a_k;f_k,u^2) \tag{10}$$
$$p(b_k/f_k) = g(b_k;f_k,v^2) \tag{11}$$
$$p_1(f) = g(f;m_1,s^2_1) \tag{12}$$

in which a gaussian pdf with expected value m and variance s^2 is indicated by $g(x;m,s^2)$. The initial quantities u^2 and v^2 are calculated in a preliminary training phase. The quantities s^2, m_1 and s_1^2 have been assumed to be quite arbitrary equal to: $m_1=210$, $s^2=6400$, $s_1^2=14700$.

3- Developments of the algorithm

In this paragraph, the related developments of the algorithm for updating and estimation are reported in detail, on the basis of the assumptions made.

In the first place, it is possible to demonstrate by induction on the basis of relations (6) and (12), that for each k, one gets:

$$p_k(f) = g(f;\ m_k,s^2_k) \tag{13}$$

where:

$$m_k = v^2.s^2_{k-1}.a_{k-1}/D_{k-1} + u^2s^2_{k-1}.b_{k-1}/D_{k-1} + C.m_{k-1}/D_{k-1} \tag{14}$$
$$s^2_k = C.s^2_{k-1}/D_{k-1}$$

with:

$$C = v^2.s^2 + u^2.s^2 + u^2.v^2 \ ;\ D_{k-1} = C + (v^2 + u^2).s^2_{k-1} \tag{15}$$

i.e., by setting $Q=(u^2 + v^2)/C$, in particular one obtains:

$$s^2_k = s^2_1/[1 + (k-1).Q.s^2_1] \tag{16}$$

Therefore, the variance s^2_k tends to zero with increasing k, with the evident physical significance that when k increases the expected value of the pitch of the speaker becomes known. Relations (14) and (16) constitute the updating relations required.

Finally, as regards the estimation relation, from equations (1) and (3) one gets:

$$p(f_k/f) = p_E(f_k-f) = g(f_k;f,s^2) \tag{17}$$

and by taking into account equations (5),(13),(17):

$$p_k(f_k) = g(f_k;m_k,s^2_k + s^2) \tag{18}$$

From the latter and taking into account relations (4),(10) and (11), it is inferred that also $p_k(f_k/a_k,b_k)$ belongs to a gaussian distribution with parameters given by:

$$\mu_k \triangleq \hat{f}_k = (s^2_k + s^2).(v^2.a_k + u^2.b_k)/D_k + u^2.v^2.m_k/D_k \tag{19}$$
$$\sigma^2_k = u^2.v^2.(s^2_k + s^2)/D_k$$

Relation (19) constitutes, the relation of estimation required.

4- Experimentation

This algorithm has been experimented at the "Centre Scientifique IBM France" in Paris, on an IBM Series/1 computer.The speech signal is first filtered at 4.8 kHz by means of a Butterworth 8-pole low-pass filter and then sampled at 10 kHz. The samples so obtained are stocked on a disk and analyzed, considering 128 samples for each frame, which correspond to windows of 12.8 ms duration.The two methods which have been combined have the characteristic of being very simple and implementable in real-time.Both methods make the computation by calculating the mean value of the pitch, window by window (128 samples). Let's examine the principle of these two algorithms.

Maxima:The signal is first filtered by means of a low-pass filter (Tchebychev filter). This filtering is made by software. Then, after a test using a threshold, the distance between two significant maxima is computed. This algorithm has been used on an Intel 8086 and works in real-time. A flow chart of this algorithm is shown in Fig.2

Autocorrelation: On windows of 256 samples with an overlap of 128 samples, only the signs of the signal are conserved. The signal is not filtered so that the algorithm can work in real-time on an Intel 8086. After the computation of the pseudo-autocorrelation function, the value for which the autocorrelation is maximum over a threshold is chosen.This algorithm is shown in Fig.3

One should note that there is no V/UV classification system before the pitch detector. Both methods make the classification. It is obvious why two errors will be considered: an error for the pitch detection and one for the V/UV classification.

Experimental results:This algorithm has been tested on 28 words spoken in english by a professional speaker. For these words, the glottal signal has been obtained by means of an accelerometer.During a training phase the values u^2 and v^2 have been evaluated and it has been found that: $u^2=60$ (maxima), $v^2=50$

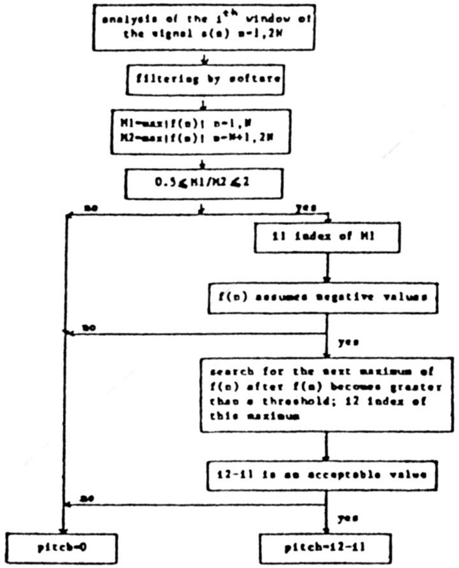

Fig.2 Pitch extraction by the maxima algorithm

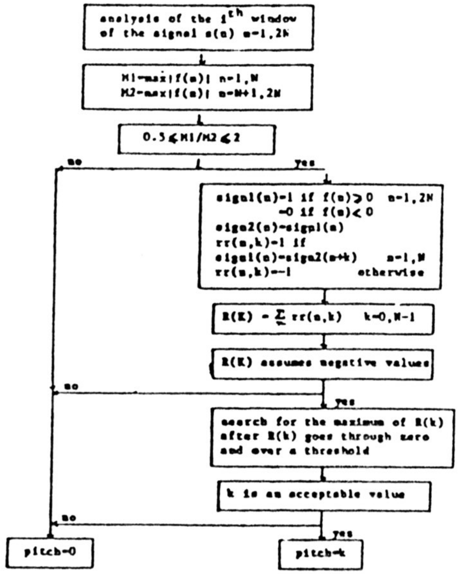

Fig.3 Pitch extraction by the autocorrelation algorithm

(autocorrelation). As mentioned previously, two errors must be considered:

1- the classification error defined as:

$$CLE = \frac{VUV + UVV}{NTOT}$$

where VUV is the number of voiced windows classified as unvoiced, UVV the number of unvoiced windows classified as voiced and NTOT the total number of windows considered.

2- the pitch error:

$$PE = \frac{\sum_{i=1}^{NTOTV} | PG(i)-P(i) |}{NTOTV}$$

where PG(i) is the real value of the pitch and P(i) the estimated value for the i^{th} voiced window. PE gives the mean value of the absolute error in the evaluation of the fundamental period expressed in number of samples.

The results obtained are the following:

	CLE	PE
Maxima	0.07	1.38
Autocorrelation	0.02	1.59
New method	0.04	1.07

For example, if PG=25 (400 Hz at 10 kHz), one has the following variation for the estimated pitch:

$$25-PE \leq P \leq 25+PE$$

As one can see, the results obtained by the combination of the two methods are better than the ones obtained by applying each method separately.