

Stability and Performance Analysis of Pitch Filters in Speech Coders

RAVI P. RAMACHANDRAN AND PETER KABAL

Abstract—This paper analyzes the stability and performance of pitch filters in speech coding when pitch prediction is combined with formant prediction. A computationally simple stability test based on a sufficient condition is formulated for pitch synthesis filters. For typical orders of pitch filters, this sufficient test is very tight. Based on the test, a simple stabilization technique that minimizes the loss in prediction gain of the pitch predictor is employed to generate stable synthesis filters. Finally, it is observed that the quality of decoded speech improves significantly when stable synthesis filters are employed.

I. INTRODUCTION

IN the speech coders considered in this paper, two non-recursive prediction error filters are used to process the incoming speech signal. The first which removes near-sample redundancies is referred to here as the formant predictor. It is followed by the pitch predictor which removes distant-sample based redundancies. The resulting residual signal after both formant and pitch prediction is then coded for transmission. An adaptive predictive coder (APC) places these predictors in a feedback loop around the residual quantizer. An additional quantization noise shaping filter can also be employed to reduce the perceptual distortion in the decoded speech [1], [2]. An alternate description of an APC coder uses an open-loop predictor configuration and a noise feedback filter [3]. A block diagram of such a configuration is shown in Fig. 1. This type of open-loop arrangement is also used in code-excited linear prediction (CELP) [4]. In CELP, the coding is accomplished by selecting the candidate waveform (from a dictionary) that best represents the residual. Also, noise shaping is accomplished implicitly in the process of choosing a representational residual signal.

In both APC and CELP, the residual signal or the selected codeword (after scaling by the gain factor) is passed through a pitch synthesis and a formant synthesis filter to reproduce the decoded speech. The filtering in the synthesis phase can be viewed in the frequency domain as first inserting the fine pitch structure and then inserting the spectral envelope (formant structure). The synthesis

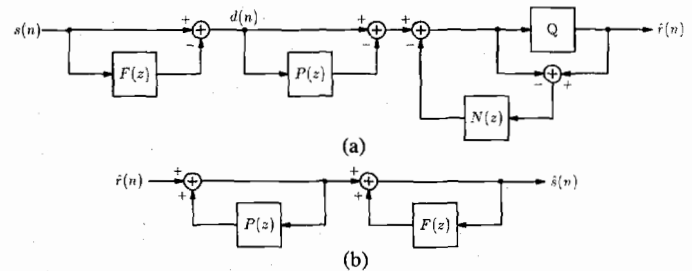


Fig. 1. Block diagram of an APC coder with noise feedback. (a) Analysis phase. (b) Synthesis phase.

filters are recursive and potentially may be unstable. For the formant filter, the autocorrelation [5], modified covariance [1], [6], or Burg [7] methods can be used to determine filter coefficients which ensure stability of the formant synthesis filter.

The procedure to determine a set of pitch predictor coefficients can result in an unstable pitch synthesis filter. This paper addresses the stability and performance issues of the pitch filter by formulating a computationally simple stability test based on a tight sufficient condition, introducing a stabilization technique, and evaluating the performance of the resulting suboptimum predictor. The effect of unstable pitch synthesis filters on decoded speech is also examined for a CELP system.

It should be noted that the filters in the speech coder are updated frame by frame and hence form a time varying system. Conventional notions of stability are in essence asymptotic properties of systems. In speech coding, an "unstable" filter may persist for a few frames (often corresponding to an interval with increasing energy—see the experimental results cited later), but eventually periods of stable filters are encountered. This means that, in practice, the output does not continue to increase in amplitude with time.

Consider the canonical case of an all-zero prediction error filter in cascade with a quantizer, followed by an all-pole synthesis filter. The quantizer can be modeled as adding noise (possibly correlated with the signal) to the residual signal. As long as the synthesis filter is the inverse to the prediction error filter and the filter coefficients are updated in step, the signal component emerges unaltered. For the signal component, stability is not a problem because of pole/zero cancellation. However, the quantization noise passes through only the synthesis filter. An "unstable" synthesis filter can cause the output noise to

Manuscript received June 13, 1986; revised January 12, 1987. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

R. P. Ramachandran is with the Department of Electrical Engineering, McGill University, Montreal, P.Q., Canada, H3A 2A7.

P. Kabal is with the Department of Electrical Engineering, McGill University, Montreal, P.Q., Canada, H3A 2A7, and INRS-Télécommunications, Université du Québec, Verdun, P.Q., Canada, H3E 1H6.

IEEE Log Number 8714483.

build up during the period of instability and can lead to degraded speech quality.

The effect of the quantization noise may be measured in a number of different ways. If the quantization noise is modeled as white noise, the output noise power can be expressed as the input noise power multiplied by the power gain of the filter. The power gain is the sum of the squares of the filter coefficients. The sufficient stability test introduced later results in a power gain for the pitch synthesis filter that is less than unity.

II. FORMANT AND PITCH PREDICTORS

The formant predictor has a transfer function

$$F(z) = \sum_{k=1}^{N_f} a_k z^{-k}. \quad (1)$$

The order N_f is typically between 8 and 16. The system function of the noise feedback filter is related to that of the formant predictor and is expressed as $N(z) = F(z/\gamma)$, where $0 < \gamma < 1$.

The pitch predictor has a small number of taps (typically 1-3) centered around a large delay M corresponding to the estimated pitch period in samples. The system function for the predictor is

$$P(z) = \begin{cases} \beta_1 z^{-M} & 1 \text{ tap} \\ \beta_1 z^{-M} + \beta_2 z^{-(M+1)} & 2 \text{ tap} \\ \beta_1 z^{-(M-1)} + \beta_2 z^{-M} + \beta_3 z^{-(M+1)} & 3 \text{ tap.} \end{cases} \quad (2)$$

At the receiver, the formant and pitch synthesis filters have transfer functions $H_F(z) = 1/(1 - F(z))$ and $H_P(z) = 1/(1 - P(z))$, respectively.

In the case of a pitch predictor, both the value of M (pitch lag) and the predictor coefficients have to be determined. The conventional strategy to determine the pitch lag is to search for the lag corresponding to the peak value of the correlation of the input signal [referred to as $d(n)$] to the pitch predictor [1]. The search range is often limited to those pitch values encountered in speech. After the value of M is determined, the coefficients of $P(z)$ are found by minimizing the mean-square value of the residual over a frame size of N samples [1], [6]. This covariance-type formulation results in a linear system of equations $\Phi\beta = \alpha$, where Φ is a matrix of correlation terms, β is the vector of predictor coefficients, and α is a vector of correlation terms. Specifically for a 3 tap predictor, the system of equations is

$$\begin{bmatrix} \phi(M-1, M-1) & \phi(M-1, M) & \phi(M-1, M+1) \\ \phi(M, M-1) & \phi(M, M) & \phi(M, M+1) \\ \phi(M+1, M-1) & \phi(M+1, M) & \phi(M+1, M+1) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \phi(0, M-1) \\ \phi(0, M) \\ \phi(0, M+1) \end{bmatrix} \quad (3)$$

where

$$\phi(i, j) = \sum_{n=0}^{N-1} d(n-i) d(n-j). \quad (4)$$

Given any vector of predictor coefficients β , the energy of the prediction residual is

$$\epsilon^2 = \phi(0, 0) - 2\beta^T \alpha + \beta^T \Phi \beta. \quad (5)$$

The residual energy for the optimum predictor is

$$\epsilon_{\min}^2 = \phi(0, 0) - \beta^T \alpha. \quad (6)$$

A normalized quantity related to ϵ^2 will be used in the sequel as the performance measure. The prediction gain is defined to be the ratio (in decibels) of the energy of the signal at the input to the predictor to the energy of the prediction residual.

III. STABILITY TEST FOR PITCH SYNTHESIS FILTERS

The covariance formulation does not guarantee that $H_P(z)$ is a stable function. To ensure stability, the denominator polynomial $D(z)$ of $H_P(z)$ must have all its zeros within the unit circle in the z -plane. The polynomial $D(z)$ is sparse in that it is of high order but has few non-zero coefficients. The Schur-Cohn procedure is a necessary and sufficient stability test [8]. Furthermore, an implementation can take into account the sparse nature of the characteristic polynomial of a pitch synthesis filter. Appendix A gives the general form of this test and shows how it can be applied to pitch synthesis filters. The Schur-Cohn test will be used later to evaluate the tightness of the test developed in this paper.

The Schur-Cohn test specialized for the case of pitch synthesis filters has a computational complexity which is proportional to the order (approximately equal to the pitch lag). The complexity of such a test is still large for pitch lags encountered in practice. In the following sections, a simple alternative test based on an asymptotically tight sufficient condition is derived. In addition, the new test will allow for the simple stabilization of unstable pitch synthesis filters.

A. Simple Sufficient Test

Two different sufficient tests will be developed. The first is a simple sufficient test, which will also serve to introduce the notation. The second is the final asymptotically tight sufficient test.

Consider a general denominator polynomial $D(z)$ of the

form

$$D(z) = z^n - B(z), \tag{7}$$

where

$$B(z) = \sum_{i=0}^{n-1} b_i z^i. \tag{8}$$

Then, $D(z) = z^n - B(z) = z^n(1 - z^{-n}B(z))$. The conditions for stability are that $1 - z^{-n}B(z) \neq 0$ or equivalently that $z^{-n}B(z) \neq 1$ on and outside the unit circle $z = e^{j\theta}$. By the maximum modulus theorem [9], $z^{-n}B(z)$ has its maximum modulus on the contour surrounding any region in which it is analytic. The expression $z^{-n}B(z)$ being a polynomial in z^{-1} is analytic on and outside the unit circle in the z -plane. Therefore, a sufficient condition for stability is that $|z^{-n}B(z)| < 1$ on the unit circle in the z -plane. This condition can be expressed as

$$|B(e^{j\theta})| < 1. \tag{9}$$

The left-hand side can be upper bounded,

$$|B(e^{j\theta})| \leq |b_0| + |b_1| + \dots + |b_{n-1}|. \tag{10}$$

Then, a simple sufficient condition for stability is that the sum of the moduli of the coefficients be less than one.

This simple test is well known (e.g., [10, p. 225]) and can be applied to any filter. For pitch filters, $B(z) = z^n P(z)$, where n is the highest power of z^{-1} in $P(z)$. The sufficient condition for stability becomes

$$\begin{aligned} |\beta_1| < 1 & \text{ 1 tap} \\ |\beta_1| + |\beta_2| < 1 & \text{ 2 tap} \\ |\beta_1| + |\beta_2| < |\beta_3| < 1 & \text{ 3 tap.} \end{aligned} \tag{11}$$

This test is both necessary and sufficient for a 1 tap filter. As will be shown later, this test applied to 2 tap filters also becomes asymptotically necessary and sufficient as n increases.

B. Tight Sufficient Test

A further examination of the expression for $|B(e^{j\theta})|$ will lead to a tight sufficient test for a 3 tap pitch filter. The 1 and 2 tap pitch filters will be special cases of the 3 tap filter.

For a 3 tap filter, $B(z) = \beta_1 z^2 + \beta_2 z + \beta_3$. For convenience, define

$$a = \beta_1 + \beta_3 \quad \text{and} \quad b = \beta_1 - \beta_3. \tag{12}$$

Then, $B(z)$ evaluated on the unit circle becomes

$$B(e^{j\theta}) = e^{j\theta} [\beta_2 + a \cos \theta + jb \sin \theta]. \tag{13}$$

The bracketed term in (13) defines an ellipse in the complex plane with center β_2 . The major axis is $|a|$ if β_1 and β_3 have the same signs, or $|b|$ if β_1 and β_3 have opposite signs. The two cases are illustrated in Fig. 2. A sufficient condition for stability is that the ellipse lie entirely within the unit circle. Since the cases $\beta_2 > 0$ and $\beta_2 < 0$ are symmetrical, the analysis proceeds by using $|\beta_2|$. Also,

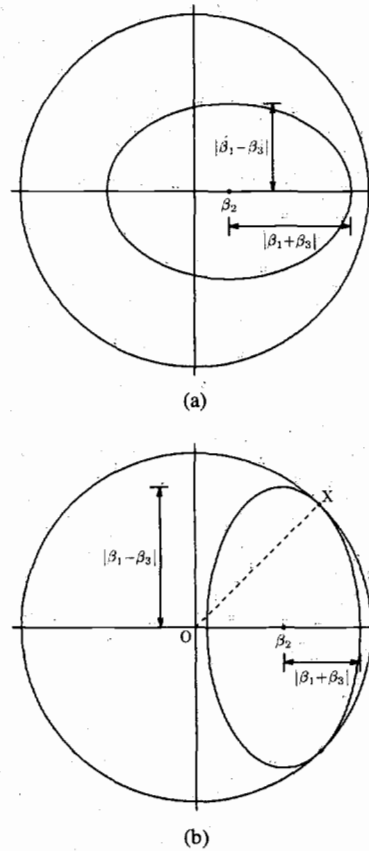


Fig. 2. Illustration of the stability ellipse for a 3 tap filter. (a) The horizontal axis is the major axis. (b) Tangency when the vertical axis is the major axis.

the cases for $a > 0$ and $b > 0$ are symmetrical with the cases $a < 0$ and $b < 0$, respectively.

If β_1 and β_3 have the same signs ($|a| > |b|$), the ellipse lies entirely within the unit circle if

$$|\beta_2| + |a| < 1, \tag{14}$$

or equivalently if

$$|\beta_1| + |\beta_2| + |\beta_3| < 1. \tag{15}$$

If β_1 and β_3 have opposite signs ($|a| < |b|$), the analysis is more complicated. The condition $|\beta_2| + |a| < 1$ ensures that no point on the minor axis lies outside the circle. This is a necessary condition for the ellipse to lie within the unit circle and will be assumed to be satisfied for the following discussion.

The aim is to establish the critical conditions on $|a|$, $|b|$, and $|\beta_2|$ which cause the ellipse to be tangent to the circle. Such a case is shown in Fig. 2(b). Let θ_c be the value of the angle θ which gives tangency. Since the case for $|\beta_2|$ is being considered, it suffices to consider $0 \leq \theta_c \leq \pi/2$. A point of tangency occurs at X when the length of OX is equal to unity,

$$(|\beta_2| + |a| \cos \theta_c)^2 + (b \sin \theta_c)^2 = 1. \tag{16}$$

In addition, tangency requires that X be the point on the ellipse that is furthest away from the origin, which in turn requires that the derivative of the left-hand side of (16) be

zero. This condition gives

$$|a| \sin \theta_c (|\beta_2| + |a| \cos \theta_c) = b^2 \sin \theta_c \cos \theta_c. \quad (17)$$

Solving this equation for $\cos \theta_c$ gives

$$\cos \theta_c = \frac{|a\beta_2|}{b^2 - a^2}, \quad (18)$$

where $\sin \theta_c$ is assumed to be nonzero. Tangency can also occur if $\theta_c = 0$ ($\sin \theta_c = 0$), giving $|\beta_2| + |a| = 1$. However, this case is precluded by the assumption that the minor axis lies within the unit circle ($|\beta_2| + |a| < 1$).

If tangency occurs for $\theta_c \neq 0$, from (16) and (18), it can be shown that the critical values of $|a|$, $|b|$, and $|\beta_2|$ are those which satisfy $f(b^2, a^2, \beta_2^2) = 0$ where

$$f(b^2, a^2, \beta_2^2) = b^2\beta_2^2 - (1 - b^2)(b^2 - a^2). \quad (19)$$

The analysis proceeds by assuming that two of the three parameters a , b , and β_2 are given, and then finding the critical value of the third. Consider a and b given. This determines the shape factor for the ellipse. Varying β_2 slides the ellipse along the horizontal axis. For $|a| < 1$ and $|b| < 1$, if $|\beta_2|$ is less than a critical value β_{2c} , where $0 \leq \beta_{2c} < 1$, the ellipse lies entirely within the unit circle. The critical value is that which results in tangency of the ellipse with the unit circle. Depending on the relative values of a and b , this can occur in one of two ways.

For $b^2 \leq |a|$, the point of tangency occurs for $\theta_c = 0$, giving $|a| + \beta_{2c} = 1$. Since $|\beta_2| < \beta_{2c}$, this condition is equivalent to having the minor axis within the unit circle. For $b^2 > |a|$, the point of tangency occurs for $0 < \theta_c \leq \pi/2$. Then the equation $f(b^2, a^2, \beta_{2c}^2) = 0$ can be solved for the critical value β_{2c} . Since this function is monotonic in β_{2c}^2 , the check that $|\beta_2| < \beta_{2c}$ is equivalent to the check that $f(b^2, a^2, \beta_2^2) < 0$.

For $|a| \geq b^2$, having the horizontal axis of the ellipse lie inside the unit circle is sufficient for stability. Otherwise, the function $f(b^2, a^2, \beta_2^2)$ must be tested. It can be shown that this test along with the minor axis requirement is sufficient for stability. The stability test for a 3 tap pitch synthesis filter is summarized below.

Stability Test: Let $a = \beta_1 + \beta_3$ and $b = \beta_1 - \beta_3$.

1) If $|a| \geq |b|$, the following is sufficient for stability:

a) $|\beta_1| + |\beta_2| + |\beta_3| < 1$.

2) If $|a| < |b|$, the satisfaction of the two following conditions is sufficient for stability:

a) $|\beta_2| + |a| < 1$

b) i) $b^2 \leq |a|$ or

ii) $b^2\beta_2^2 - (1 - b^2)(b^2 - a^2) < 0$.

Part 2 of this stability test is tighter than the simple sufficient test given earlier. This part of the test is invoked when $|a| < |b|$ or equivalently when β_1 and β_2 have opposite signs. Experiments show that in voiced speech,¹ β_2 is greater than zero in about 90 percent of the frames.

¹A frame was considered to be voiced when the pitch prediction gain was greater than 1 dB.

Given that $\beta_2 > 0$, the number of voiced frames in which β_1 and β_3 have opposite signs is about 3.7 times the number of voiced frames in which they have the same signs. Therefore, the presence of a tighter test when $|a| < |b|$ is important for speech coders.

The test for 3 tap filters subsumes the test for 2 tap filters. By setting $\beta_1 = 0$, a 3 tap filter becomes a 2 tap filter. Then $|a| = |b|$ and the test involves checking that the sum of the moduli of the coefficients is less than 1. This is equivalent to the simple sufficient test given earlier.

C. Further Examination of the Sufficient Condition

The sufficient test defines a stability region for 3 tap filters that is independent of the order n . Consider first the simple sufficient test ($|\beta_1| + |\beta_2| + |\beta_3| < 1$). The stability region can be viewed in $(\beta_1, \beta_2, \beta_3)$ space as a region bounded by 8 flat surfaces. The volume enclosed is $4/3$ units.

The stability region described by the tight sufficient test has two types of surfaces. Four of the surfaces are flat and coincide with the flat surfaces of the simple sufficient test. The other four surfaces bulge out significantly beyond the flat surfaces. The volume enclosed can be determined in closed form and is calculated to be $16/9$ units. Fig. 3 shows a contour plot of the stability region defined by the tight sufficient test for $\beta_3 \geq 0$. A plot for $\beta_3 \leq 0$ is a mirror image reflected about the vertical axis. It can be shown that this stability region is enclosed by a unit sphere and hence, the sum of the squares of the coefficients (power gain) is less than unity.

Both the magnitude and phase of $B(e^{j\theta})$ determine the necessary and sufficient conditions. If the stability ellipse lies entirely inside the circle of unit radius, all of the roots of $D(z)$ are inside the unit circle in the z -plane. As some combination of the parameters $|a|$, $|b|$, or $|\beta_2|$ is increased, the stability ellipse will emerge outside the circle of unit radius and the roots of $D(z)$ will eventually cross the unit circle. The critical combination of parameters which cause roots to lie on the unit circle can be determined from

$$B(e^{j\theta}) = e^{jn\theta}. \quad (20)$$

The points at which the ellipse crosses the unit circle correspond to points which satisfy the above equation in magnitude. For a given n , these intersection points correspond to phase angles which may or may not satisfy the above equation. However, as n increases, the phase angles which satisfy the above equation become increasingly dense. In the limit of large n , as soon as the stability ellipse crosses the unit circle, at least one root of $D(z)$ crosses the unit circle. This indicates that the stability ellipse must lie entirely within the unit circle. The stability test given earlier becomes both necessary and sufficient in the limit of large n .

The necessary and sufficient conditions as determined by the Schur-Cohn test define a region which depends on the order n . This region is bounded by four types of sur-

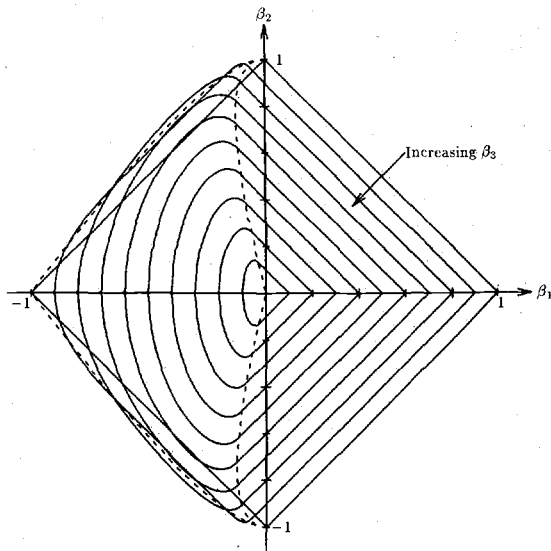


Fig. 3. Region described by the tight sufficient test. Equal value contours are shown for $\beta_3 = 0, 0.1, \dots, 0.9$. The dotted lines represent lines on the stability surface with $b^2 = |a|$.

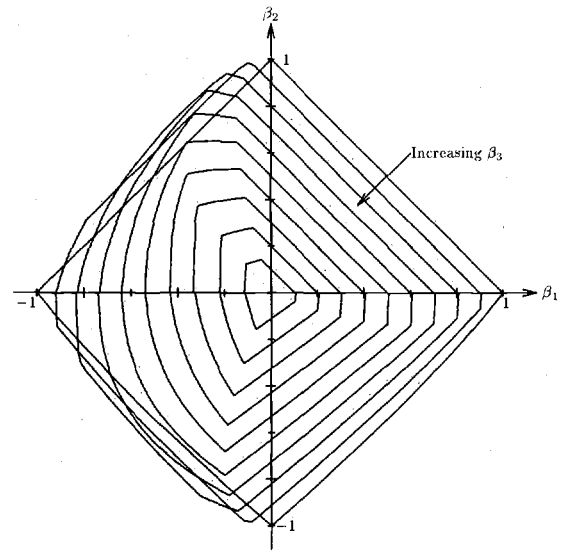


Fig. 4. Necessary and sufficient stability region ($n = 7$). Equal value contours are shown for $\beta_3 = 0, 0.1, \dots, 0.9$.

faces. Two surfaces are flat and coincide with the flat surfaces for the regions described above. Two other surfaces bulge slightly but become flat in the limit of large n . Two other pairs of surfaces bulge significantly and coincide with the bulging surfaces of the tight sufficient test in the limit of large n . The symmetries of the surfaces switch depending on whether n is even or odd. Fig. 4 shows the stability region determined by the Schur-Cohn test for $n = 7$ and $\beta_3 \geq 0$. This rather small value of n is used to accentuate the differences between this region and that determined by the sufficient test. Note also that the contour with $\beta_3 = 0$ is the stability region for the 2 tap filter with $n = 6$.

It remains to ascertain how tight the sufficient test is for finite n . The area and volume enclosed by the true stability regions for 2 and 3 tap filters were computed using numerical integration techniques for various values of n . The boundaries of the true stability regions were computed using the simplified Schur-Cohn procedure described in Appendix A. The region defined by the sufficient test is contained in that defined by the Schur-Cohn test. Fig. 5 shows the percent differences between the areas and volumes enclosed by the sufficient test and the Schur-Cohn test for 2 and 3 tap filters. It is observed that the percent difference decreases rapidly as n increases. The lowest order of a pitch filter is typically around 20. Even at this low order, the difference in volume is below 1 percent. For higher orders, the sufficient test is very tight and involves much less computation than the Schur-Cohn test.

D. Extension to Circles of Arbitrary Radius

The sufficient condition can be extended in order to determine whether or not all the roots of $D(z) = z^n - B(z)$ [$B(z)$ defined in (8)] are within a circle of radius r centered at the origin in the z -plane. Just as before, the max-

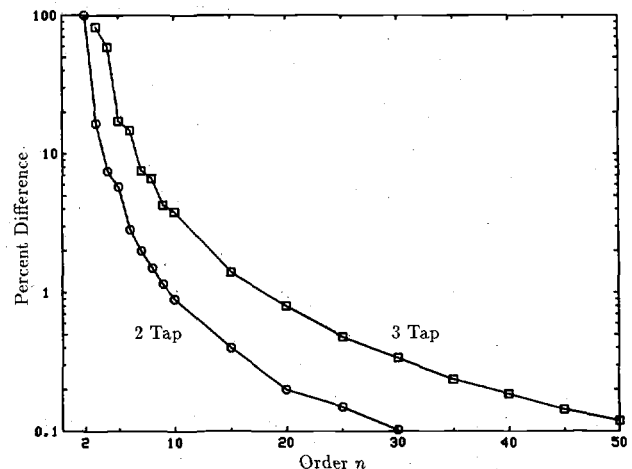


Fig. 5. Percent differences between areas and volumes enclosed by the sufficient test and the Schur-Cohn test.

imum modulus theorem is used to derive the condition

$$|B(re^{j\theta})| < r^n. \quad (21)$$

Expanding $|B(re^{j\theta})|$ yields

$$\begin{aligned} |B(re^{j\theta})| &= |b_0 + b_1 re^{j\theta} + \dots + b_{n-1} r^{n-1} e^{j(n-1)\theta}| \\ &\leq |b_0| + |b_1| r + \dots + |b_{n-1}| r^{n-1}. \end{aligned} \quad (22)$$

A simple sufficient condition that ensures that all the roots of $D(z)$ are within the circle $|z| = r$ is

$$|b_0| + |b_1| r + \dots + |b_{n-1}| r^{n-1} < r^n. \quad (23)$$

The condition $|\beta_1| < r^n$ is necessary and sufficient for 1 tap filters. For 3 tap filters, a more detailed examination of $|B(re^{j\theta})|$ is done in the same fashion as for $|B(e^{j\theta})|$. The conditions are given below. The 3 tap case subsumes the 2 tap case. Hence, the condition for 2 tap filters is merely $|\beta_1| r + |\beta_2| < r^n$.

Conditions for 3 Tap Filters: Let $a = \beta_1 r^2 + \beta_3$ and $b = \beta_1 r^2 - \beta_3$.

- 1) If $|a| \geq |b|$,
 - a) $|\beta_1| r^2 + |\beta_2| r + |\beta_3| < r^n$.
- 2) If $|a| < |b|$,
 - a) $|\beta_2| r + |a| < r^n$
 - b) i) $b^2 \leq |a| r^n$ or
ii) $b^2 \beta_2^2 r^2 - (r^{2n} - b^2)(b^2 - a^2) < 0$.

IV. STABILIZATION PROCEDURE

In each frame of speech, the predictor coefficients are calculated by using the covariance formulation. Then, the stability test is used to determine whether the corresponding synthesis filter is stable. If the filter is found to be stable, no modification of the coefficients is made. However, if the filter is unstable, a stabilization procedure is used to find a new set of coefficients. It is assumed that the pitch estimate M remains unaltered. Note that the new coefficients are used for both the pitch predictor and the pitch synthesis filter. Therefore, the prediction gain is reduced at the analysis stage. A major concern in performing stabilization is to keep the loss in prediction gain to a minimum. Consider three different stabilization strategies which demonstrate different possibilities.

A. Scaled Coefficients

The first stabilization method is based on a feedback control viewpoint. The predictor is used in the feedback path of the synthesis filter. Stabilization is accomplished by varying the feedback gain. This action is equivalent to multiplying each predictor coefficient by the same factor c . As c varies, the poles of the synthesis filter move along the root loci. The stability test given above will be used to determine the factor c so as to guarantee that all the poles lie within the unit circle.

Consider a vector of predictor coefficients β that minimizes the mean-square prediction residual energy. After scaling by a factor c , the vector of predictor coefficients is $\beta' = c\beta = \beta + \delta$. This results in a suboptimum predictor for which the energy of the residual is $\epsilon^2 = \epsilon_{\min}^2 + \delta^T \Phi \delta$, where ϵ_{\min}^2 is the minimum residual energy for an optimum predictor [see (6)]. The quantity $\delta^T \Phi \delta$ represents the excess residual energy resulting from the use of a suboptimum predictor. Since $\delta = (c - 1)\beta$, the excess residual energy is $(c - 1)^2 \beta^T \Phi \beta$. It is observed that as c deviates from one, the loss in prediction gain increases. Note that the largest value of c which stabilizes an unstable filter lies in the range $0 < c < 1$. Therefore, in order to minimize the loss in prediction gain, c must be as close to one as possible and at the same time give a stable pitch synthesis filter. This procedure is referred to as the common scaling factor method.

B. Radial Scaling

A second stabilization strategy is to move the poles radially inward, each by the same proportion. The motivation behind such a method is to produce a stable filter

which preserves the frequency at which spectral features occur. Radial scaling is equivalent to using the transformation $z' = rz$.

This stabilization procedure is related to the extended stability test for circles of arbitrary radius. The difference is that in a stability test, the value of r is given, whereas in the stabilization process, the value of r which results in marginal stability must be determined. For filters with more than 1 tap, the value of r must be found iteratively. Use of the simple sufficient test [condition 1a) in the test], can simplify the determination of r .

Radial scaling involves multiplying the coefficient of z^{-i} by r^{-i} . For a 1 tap pitch filter, radial scaling is the same as the previous method. For a 3 tap filter, the new coefficients become

$$\begin{aligned}\beta_1' &= r^{-(M-1)}\beta_1, \\ \beta_2' &= r^{-M}\beta_2, \\ \beta_3' &= r^{-(M+1)}\beta_3.\end{aligned}\quad (24)$$

For typical values of M and for filters that originally have their roots only slightly beyond the unit circle, the factors multiplying the coefficients are very nearly equal. This argues for the case that in practice radial scaling is essentially the same as using the common scaling factor method. From another viewpoint, moving along the root loci using a single scaling factor is very nearly the same as radial scaling.

Experiments show that stabilization using radial scaling performs nearly the same as the common scaling factor method. Given the need for an iterative solution, radial scaling seems unnecessarily complex.

C. Reciprocal Poles

An often-suggested procedure to stabilize a filter is to replace each pole outside the unit circle by its reciprocal. This will preserve the frequency response of the filter, but in the case of a pitch synthesis filter involves factoring the high degree denominator polynomial. This may be impractical for filters with more than a single coefficient.

The loss in prediction gain associated with this technique can be more than when scaling the coefficients by a factor c . This can be seen by considering a 1 tap filter. Suppose the filter is unstable with $|\beta_1| > 1$. Replacing each unstable pole by its reciprocal is equivalent to scaling the coefficient β_1 by a factor which is smaller than if the poles had been brought just inside the unit circle. This implies a larger loss in prediction gain for the reciprocal pole strategy. Since the 2 tap and 3 tap cases subsume the 1 tap case, the reciprocal method can perform worse than simple scaling of the coefficients.

D. Scaling Procedure

In light of the foregoing discussion, stabilization will be accomplished by using the common scaling factor method. The value of the scaling factor c for the 1 and 2 tap cases is easily determined since the stability test in-

volves only a single condition. The procedure for determining c needs some elaboration for the 3 tap case.

Again define, $a = \beta_1 + \beta_3$ and $b = \beta_1 - \beta_3$. When $|a| \geq |b|$, the factor c must force $|\beta_1| + |\beta_2| + |\beta_3|$ to be at most equal to 1. The value of c which gives marginal stability is

$$c = \frac{1}{|\beta_1| + |\beta_2| + |\beta_3|}. \quad (25)$$

Corresponding to the simple sufficient test, this same value of c can be used for any a and b . However, using the tighter test will result in a larger value of c and hence will result in a smaller loss in prediction gain. If $b^2 \leq |a|$, marginal stability is achieved when the stability test ellipse for the scaled coefficients is tangent to the circle with $\theta_c = 0$. The appropriate scaling factor is

$$c = \frac{1}{|a| + |b_2|}. \quad (26)$$

After applying this scaling factor, it can be shown that condition 2b)i) remains satisfied.

If $b^2 > |a|$, there is a point of tangency for the scaled ellipse (for $0 < c < 1$) for which $0 < \theta_c \leq \pi/2$. For that value of c , the function $f(b^2, a^2, \beta_2^2)$ with scaled coefficients is equal to zero. Solving for the scaling factor gives

$$c = \sqrt{\frac{b^2 - a^2}{b^4 + b^2\beta_2^2 - b^2a^2}}. \quad (27)$$

With this value of c , it can be shown that all points along the minor axis of the scaled ellipse are within the unit circle [condition 2a) is satisfied].

V. EXPERIMENTAL RESULTS

The experimental results were derived by implementing a CELP coder. In CELP, the representational waveform is chosen from a set of entries in a codebook constructed of Gaussian random numbers with unit variance. Conceptually, each entry in the codebook (see Fig. 6) is scaled by the gain factor, filtered by $H_p(z)$ and $H_f(z)$, and subtracted from the original speech to form a difference signal. This signal is passed through a weighting filter $W(z) = (1 - F(z))/(1 - N(z))$. The error is formed by squaring and averaging the filtered difference signal. The entry in the codebook that gives the smallest error is used to represent the residual and its index is transmitted.

The experimental conditions involved the use of a 10th-order formant predictor (coefficients determined by using the modified covariance method) and a 3 tap pitch predictor. The speech was sampled at 8 kHz and divided into frames of length 80 samples (these short frame lengths tend to exacerbate the stability problem). Forty sample blocks of the residual were compared to a codebook of $2^{10} = 1024$ waveforms. The parameter $\gamma = 0.8$ was used in implementing the weighting filter $W(z)$. The performance of the suboptimum predictor that results from the appli-

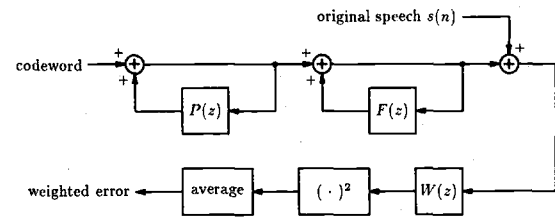


Fig. 6. Calculating the weighted error.

cation of the common scaling factor method and the effect of unstable pitch synthesis filters on decoded speech were investigated. The database consisted of six sentences, three of which were spoken by males and three by females.

A. Performance of the Suboptimum Predictor

The common scaling factor method ensures at least marginal stability since the scaled stability ellipse lies entirely within the circle except for a point of tangency. Complete stability is achieved by subtracting a small quantity from the calculated value of c . For the coder under study, the average prediction gain achieved by the covariance formulation is 4.18, 5.32, and 5.71 dB for 1, 2, and 3 tap filters, respectively. The average loss in prediction gain associated with stabilization is 0.03, 0.26, and 0.21 dB for 1, 2, and 3 tap filters, respectively.

A strategy employed in previous studies [11], [12] involves the use of a 1 tap filter whenever a 3 tap filter is found to be unstable. Experiments reveal that this approach diminishes the prediction gain by an average of 1.07 dB when an unstabilized 1 tap filter is used, and 1.11 dB when a stabilized 1 tap filter is used. This is a significantly larger loss than the 0.21 dB that results when using a stabilized 3 tap filter.

In an APC system, output speech quality is degraded when the quantizer clips high-amplitude portions of the residual signal [13]. The 3 tap pitch filters outperform 1 tap filters since they tend to further reduce the peak values of the residual signal [14]. In view of this, stabilization of 3 tap filters is preferred to a scheme which backs off to a 1 tap filter during frames of instability. This is also true for a CELP coder since the codebook entries (which are constructed from Gaussian random numbers) more closely resemble a residual which is free of high-amplitude pitch pulses.

B. Effect of Instability on Decoded Speech

Decoded waveforms for unstabilized and stabilized pitch filters are shown in Fig. 7 for the sentence "cats and dogs each hate the other" (spoken by a male). Frames having unstable pitch synthesis filters are marked by a nonzero indicator function. Both decoded waveforms are distorted due to quantization noise filtered by $H_p(z)$ and $H_f(z)$. If $H_p(z)$ is unstable, the energy of the noise is enhanced and causes further degradation as observed in waveform (2). Although both decoded signals were intelligible, waveform (2) suffers from pops or clicks that can

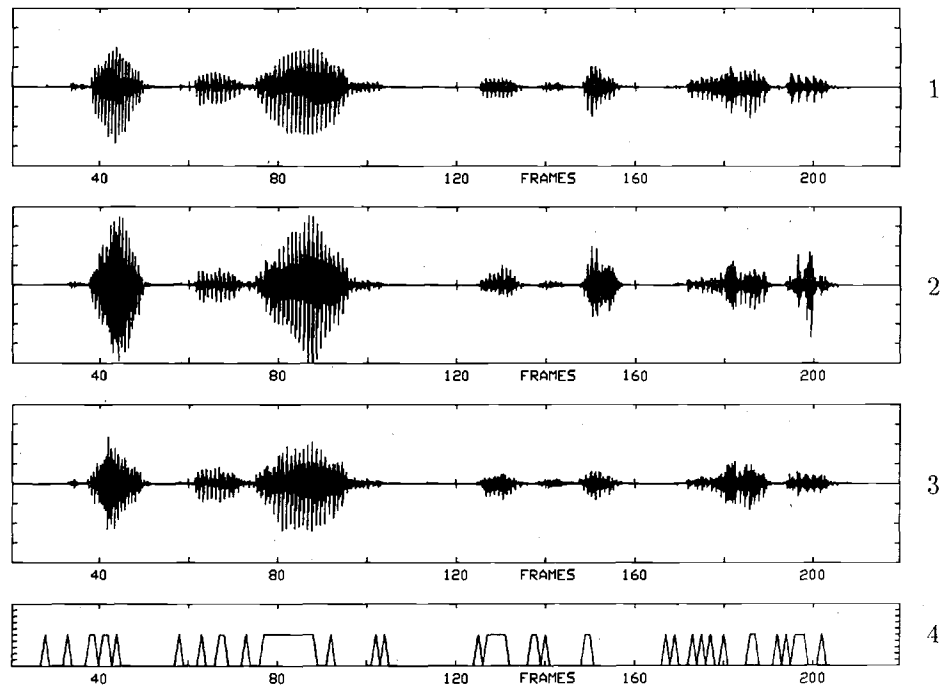


Fig. 7. Original and decoded signals (1) original speech, (2) decoded signal for the covariance pitch formulation, (3) decoded signal for stabilized pitch synthesis filters, and (4) frames having unstable pitch synthesis filters.

be annoying. Also, background noise is more dominant in portions of waveform (2) than in waveform (3).

Degradations in the output speech are perceptible if a sequence of consecutive frames of high input energy have unstable filters or if the numerical value of $\sum_i |\beta_i|$ is large. Frames 77–88 consist of high-energy voiced speech and have unstable pitch synthesis filters. The quantization noise continues to build up causing the energy of the output signal to keep rising. This increased noise is perceptible. Although the average value of $\sum_i |\beta_i|$ is just 1.3, distortion is noticeable since this segment has a high input energy.

Certain isolated frames of high input energy having unstable filters also demonstrate a rising output energy (frames 38–50). Frames 38 and 39 have filters whose value of $\sum_i |\beta_i|$ equals 2.36 and 2.46, respectively. These large values are responsible for considerable distortion. This distortion continues since frames 41, 42, and 44 have unstable filters where $\sum_i |\beta_i|$ equals 1.45, 1.57, and 1.54, respectively.

The degradation is not serious when unstable filters with small values of $\sum_i |\beta_i|$ are present in frames of relatively low energy. This situation occurs during frames 127, 128, and 129 in which $\sum_i |\beta_i|$ equals 1.11, 1.53, and 1.35. However, if an unstable filter with a very large value of $\sum_i |\beta_i|$ occurs even in a frame of low input energy, an impulse-type distortion that is heard as a pop or click is present. Frames 149 and 150 with values of $\sum_i |\beta_i|$ equal to 4.43 and 2.70 clearly depict this. Another example of this phenomenon is during frames 196–198. Here, $\sum_i |\beta_i|$ equals 2.37, 4.02, and 2.23. In both cases, a clear pop sound is heard. When pitch synthesis filters are stabilized,

the undesirable pops, clicks, and enhanced background noise disappear. Similar degradations in the output speech have been reported for an APC coder when unstable pitch synthesis filters are employed [11].

VI. SUMMARY AND CONCLUSIONS

Formant and pitch filters are essential components of both the transmitter and receiver in low bit rate speech coders. Algorithms that ensure the stability of the formant synthesis filter are available. However, the covariance formulation can result in an unstable pitch synthesis filter. This paper provides a computationally simple but tight sufficient test for pitch synthesis filters that is independent of the order n . A closer examination reveals that this test is both necessary and sufficient in the limit of large n and involves much less computation than the Schur–Cohn test. For typical orders of pitch filters, the sufficient test is very tight. This sufficient test has also been extended to check if the denominator polynomial of the system function has all its roots within a circle of arbitrary radius r centered at the origin.

From the sufficient stability test, three different stabilization techniques are suggested. The first technique involves scaling the predictor coefficients by a common factor. A second technique involves scaling each of the poles radially inward. This is equivalent to scaling the predictor coefficients by different factors and requires the solution of a nonlinear equation. The third method of replacing each pole outside the unit circle by its reciprocal is not practical since it involves the factoring of a high degree polynomial. The first technique is judged to be the most practical and the common factor is chosen to minimize the

loss in prediction gain at the analysis stage. The actual average loss in prediction gain is negligible. The simple stabilization procedure derived from an efficient stability test can be easily implemented as part of a coding system in a real-time environment.

It is finally observed that decoded speech generated by a CELP coder improves in quality when stable pitch synthesis filters are used. Degradations in the output speech manifest themselves either as background noise or as pops when unstable pitch synthesis filters are used. These undesirable sounds are clearly audible when a set of consecutive frames of high input energy have unstable filters or when the value of $\sum_i |\beta_i|$ is large in an isolated frame of high input energy.

APPENDIX A SCHUR-COHN TEST

The Schur-Cohn test [8] can be used to determine whether or not the roots of a polynomial $D(z) = a_0 + a_1z + \dots + a_nz^n$ are within the unit circle. This test has been shown to be equivalent to the procedure to convert a set of predictor coefficients to reflection coefficients and then checking that their magnitudes are less than unity [15].

A sequence of polynomials of decreasing order $D_0(z) = D(z)$, $D_1(z)$, \dots , $D_{n-1}(z)$ are defined such that

$$D_{j+1}(z) = a_0^{(j)} D_j(z) - a_{n-j}^{(j)} z^{n-j} D_j(z^{-1}),$$

for $j = 0$ to $n - 1$. (A.1)

In this recursion, $a_i^{(0)} = a_i$, the original coefficients of $D(z)$. The coefficients of $D_{j+1}(z)$ are derived from those of $D_j(z)$ by the relationship

$$a_k^{(j+1)} = a_0^{(j)} a_k^{(j)} - a_{n-j}^{(j)} a_{n-j-k}^{(j)},$$

for $k = 0$ to $n - j - 1$. (A.2)

A set of necessary and sufficient conditions for the roots of $D(z)$ to lie within the unit circle is

$$\begin{aligned} |a_0^{(0)}| &< |a_n^{(0)}| \\ |a_0^{(1)}| &> |a_{n-1}^{(1)}| \\ &\vdots \\ |a_0^{(n-1)}| &> |a_1^{(n-1)}|. \end{aligned} \quad (A.3)$$

Application to Pitch Synthesis Filters

For pitch synthesis filters, it is noted that many of the intermediate coefficients are zero. For an l tap filter, a maximum of $2l - 1$ coefficients must be calculated at each stage in the recursion. Specifically, the denominator polynomial of a 3 tap pitch synthesis filter is

$$\begin{aligned} D(z) &= z^{M+1} - \beta_1 z^2 - \beta_2 z^2 - \beta_3 \\ &= a_{M+1}^{(0)} z^{M+1} + a_2^{(0)} z^2 + a_1^{(0)} z + a_0^{(0)}. \end{aligned} \quad (A.4)$$

For the simplified procedure, there are at most five non-zero coefficients at any stage. These coefficients are $a_0^{(j+1)}$, $a_1^{(j+1)}$, $a_2^{(j+1)}$, $a_{M-j-1}^{(j+1)}$, and $a_{M-j}^{(j+1)}$. When $j = M - 4$, the computed coefficients are naturally ordered from $a_0^{(M-3)}$ to $a_4^{(M-3)}$. Then, the number of coefficients is reduced by one at each successive stage and the test terminates when the last condition $|a_0^{(M)}| > |a_1^{(M)}|$ is examined.

REFERENCES

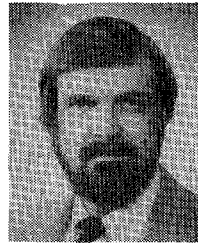
- [1] B. S. Atal and M. R. Schroeder, "Predictive coding of speech and subjective error criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 247-254, June 1979.
- [2] J. Makhoul and M. Berouti, "Adaptive noise spectral shaping and entropy coding of speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 63-73, Feb. 1979.
- [3] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [4] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High-quality speech at very low bit rates," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Tampa, FL, Mar. 1985, pp. 25.1.1-25.1.4.
- [5] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [6] B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Commun.*, vol. COM-30, pp. 600-614, Apr. 1982.
- [7] J. Makhoul, "Stable and efficient lattice methods for linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 423-428, Oct. 1977.
- [8] E. I. Jury, *Theory and Application of the z-Transform Method*. New York: Wiley, 1964.
- [9] R. A. Silverman, *Introductory Complex Analysis*. New York: Dover, 1967.
- [10] D. S. Mitrinovic, *Analytic Inequalities*. New York: Springer-Verlag, 1970.
- [11] R. Viswanathan, W. Russell, A. Higgins, M. Berouti, and J. Makhoul, "Speech-quality optimization of 16 kb/s adaptive predictive coders," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Denver, CO, Apr. 1980, pp. 520-525.
- [12] M. D. Cowing and J. W. Fussell, "16 kbps APC with hybrid quantization," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, San Diego, CA, Mar. 1984, pp. 27.2.1-27.2.4.
- [13] B. S. Atal and M. R. Schroeder, "Improved quantizer for adaptive predictive coding of speech signals at low bit rates," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Denver, CO, Apr. 1980, pp. 535-538.
- [14] J. Makhoul and M. Berouti, "Predictive and residual encoding of speech," *J. Acoust. Soc. Amer.*, pp. 1633-1641, Dec. 1979.
- [15] S. Haykin, *Adaptive Filter Theory*. Englewood, Cliffs, NJ: Prentice-Hall, 1986.



Ravi P. Ramachandran was born in Bangalore, India, on July 12, 1963. He received the B.Eng. degree (with great distinction) from Concordia University, Montreal, P.Q., Canada, in 1984 and the M.Eng. degree (on the Dean's Honour List) from McGill University, Montreal, in 1986.

During the period of his undergraduate studies, he held a Concordia University Undergraduate Fellowship. When studying for the M.Eng. degree he held a Natural Sciences and Engineering Research Council of Canada (NSERC) Postgraduate Fellowship. He is currently a doctoral student at McGill University and holds another NSERC Postgraduate Fellowship for the Ph.D. degree. His main research interests are in speech coding, data communications, and digital signal processing.

Mr. Ramachandran received the Order of Engineers of Quebec Student Award in 1984, the John H. Chapman Award (presented by Spar Aerospace for scholarship in communications), the Electrical Engineering Medal as the Most Outstanding Student in Electrical Engineering, and the Morris Chait Medal as the Highest Ranking Student in the B.Eng. program.



Peter Kabal received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, Ont., Canada.

He is an Associate Professor in the Department of Electrical Engineering at McGill University, Montreal, P.Q., Canada, and a Visiting Professor at INRS-Telecommunications (a research institute affiliated with the Université du Québec), Verdun, P.Q., Canada. His current research interests focus on digital signal processing as applied to speech coding, adaptive filtering, and data transmission.
