A method and apparatus for data processing by linear prediction. The second order prediction error filter (PEF) coefficients for the data are estimated in a PEF estimator (14) so as to approximate the pair of coefficients occurring in a binomial factor of the $z$ transform of a sequence of higher order PEF coefficients of the data. The binomial coefficient pair is associated with a formant in the spectrum of the data. The PEF coefficients are utilized to control filters (11, 12, 13) so as to modify the spectrum of the data according to changes in the data. The estimator (14) and filter arrangement (11, 12, 13) are effectively a data adaptive filter. A process of successive approximation is used to isolate the main formant and also to attenuate it in a data stream which is transmitted to a further similar cascaded module in the apparatus whereby the same process is performed to isolate and remove a further formant. Buffers (17-20) thus contain PEF coefficient pairs associated with each formant in the input data.
FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th>Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Austria</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
</tr>
<tr>
<td>DE</td>
<td>Germany, Federal Republic of</td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
</tr>
<tr>
<td>FI</td>
<td>Finland</td>
</tr>
<tr>
<td>FR</td>
<td>France</td>
</tr>
<tr>
<td>GA</td>
<td>Gabon</td>
</tr>
<tr>
<td>GB</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>HU</td>
<td>Hungary</td>
</tr>
<tr>
<td>JP</td>
<td>Japan</td>
</tr>
<tr>
<td>KP</td>
<td>Democratic People's Republic of Korea</td>
</tr>
<tr>
<td>LI</td>
<td>Liechtenstein</td>
</tr>
<tr>
<td>LU</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>MC</td>
<td>Monaco</td>
</tr>
<tr>
<td>MG</td>
<td>Madagascar</td>
</tr>
<tr>
<td>MW</td>
<td>Malawi</td>
</tr>
<tr>
<td>NL</td>
<td>Netherlands</td>
</tr>
<tr>
<td>NO</td>
<td>Norway</td>
</tr>
<tr>
<td>RO</td>
<td>Romania</td>
</tr>
<tr>
<td>SE</td>
<td>Sweden</td>
</tr>
<tr>
<td>SN</td>
<td>Senegal</td>
</tr>
<tr>
<td>SU</td>
<td>Soviet Union</td>
</tr>
<tr>
<td>TD</td>
<td>Chad</td>
</tr>
<tr>
<td>TG</td>
<td>Togo</td>
</tr>
<tr>
<td>US</td>
<td>United States of America</td>
</tr>
</tbody>
</table>
IMPROVEMENTS IN SIGNAL PROCESSING

The present invention relates to signal processing and more particularly to signal processing by means of linear prediction coding. The invention is primarily concerned with speech signals but is not limited thereto.

A fundamental problem in the analysis of speech, and in similar fields, concerns the identification of the broad features, or trend, of the power spectrum of a signal in the presence of fine structure brought about by the harmonics of low frequency components and by the presence of noise. The broad peaks in the envelope of the power spectrum of a signal are known as formants. The problem is a problem of formant identification.

Present techniques apply either in the frequency domain or in the time domain. In the former approach, the power spectrum of the signal is first approximated in some way such as by feeding the signal through a parallel array of filters or by computing the Fourier transform of a segment of the input data. Smoothing procedures are then applied and peaks in the smoothed frequency domain data are taken as estimates of the formant peaks of the spectrum.

The accurate determination of formant locations by this method is often confused by the presence of low frequency harmonics. These are manifested as large spikes occurring periodically along the spectrum which may vary in frequency independently of the formant locations.

Similarly, in a time domain approach, where zero crossings or peaks may be counted to estimate formant frequencies, the presence of a varying low frequency pitch can render such estimates unreliable.

A more sophisticated time domain approach, used in the analysis of speech and of which the invention is a special case, is the linear prediction method. Under this method, successive time segments of the signal are each assumed to be the outcome of a stationary autoregressive random process. The parameters of the process, known as prediction error filter (PEF) coefficients or "predictor coefficients" \( \gamma_i, \ i = 1, \ldots, n \), are estimated by finding those
values \( \hat{\gamma}_i \) of \( \gamma_i \) which minimize the quantity \( P_n \), where

\[
P_n = \hat{\gamma} \hat{R} \hat{\gamma}
\]

(1)

and \( \hat{\gamma} \) is the transpose of \( \gamma \).

The solution is

\[
\hat{R} \hat{\gamma} = \hat{P}_n \hat{J}
\]

(2)

where \( R \) is a \((n+1) \times (n+1)\) matrix of sample covariances, \( \hat{P}_n \) is a scalar known as the prediction error power estimate, and \( \hat{J} \) is a vector whose first element is unity and the rest zeros. Equation (2) is solved for the vector \( \hat{\gamma} \), which comprises an estimate of the population prediction error filter (PEF) coefficient vector, \( \gamma \). The integer, \( n \), is known as the "order" of the autoregressive model.

An estimate, \( \hat{S}_n(f) \), of \( S(f) \) the power spectral density function, or "spectrum", of the population at frequency, \( f \), is given by

\[
\hat{S}_n(f) = \frac{\hat{P}_n}{2f_N} |\hat{\gamma}(z)|^2
\]

(3)

for

\[
z = \exp \{-2\pi i \Delta t f\},
\]

where \( \Delta t \) is the sampling interval, \( f_N \) is the Nyquist frequency, and \( \hat{\gamma}(z) \) is the z transform of the PEF coefficients given by

\[
\hat{\gamma}(z) = \sum_{p=0}^{\infty} \hat{\gamma}_p z^p
\]

(4)

Obviously, for \( z \) constrained as it is to lie on the unit circle as in (3), \( \hat{\gamma}(z) \) becomes the discrete Fourier transform of the sequence \( \{\hat{\gamma}_i\} \).

Equations (3) and (2) form the basis of a technique of spectral estimation known as the "maximum entropy" method or "linear prediction" method. It has the advantage, over other methods, that the resolution of spectral peaks is
independent of the order or "maximum lag", n, chosen. Rather, the order determines the number of different spectral peaks which can be independently resolved in the interval (\(-f_n, f_n\)). If the order is chosen to be much smaller than the population value, the resulting spectral estimate forms a smooth best fit to the population spectrum. This can be seen as follows:

If the matrix, R, of sample covariances in (1), is considered as an approximation to the matrix of population covariances, it can easily be shown that

\[
P_n = \int_{-f_n}^{f_n} \frac{S(f)}{\hat{S}_n(f)} \, df
\]

and hence that

\[
P_n = \int_{-f_n}^{f_n} \exp\{\log S(f) - \log \hat{S}_n(f)\} \, df \quad (5)
\]

Now the estimated PEF coefficients \(\{\hat{\gamma}_j\}\) have been chosen specifically to minimize \(P_n\), and hence to minimize the right hand side of (5). Due to the exponentiation, difference areas above the \(\log \hat{S}_n(f)\) contour will contribute disproportionately more to the total integral than will those below. Hence, when the integral is minimized, the resulting locus of \(\log \hat{S}_n(f)\) will tend to follow the peaks in the log population spectrum.

Since \(\hat{S}_n(f)\) is proportional to the reciprocal of the squared modulus of \(\hat{\gamma}(z)\), as z moves around the unit circle, peaks will occur in \(S_n(f)\) when \(z\) passes close to a zero of \(\hat{\gamma}(z)\) in the \(z\) plane. That is, the roots of the polynomial equation

\[
\hat{\gamma}(z) = 0 \quad (6)
\]

determine the locations and widths of the peaks in the spectral estimate \(S_n(f)\). Since \(\gamma(z)\) occurs in the denominator of the right hand side of (3), the roots of (6) are often referred to as poles.
The linear prediction method of estimating spectral trends has several advantages over other methods, viz:

(i) only a small number of parameters are required to represent spectral trends,

(ii) the estimated spectrum at low orders represents a smooth approximation to the population spectrum and is completely unaffected by the presence of pitch harmonics; since these cannot affect the values of the covariances, for moderate values of the order, \( n \),

(iii) spectral resonances, in the form of formant peaks, are weighted most heavily in the error criterion and are thus represented most accurately, and

(iv) the PEF coefficients when used as the coefficients of a recursive filter acting on a suitable excitation function can be used to generate a sequence having a similar spectral character to the original sequence. This fact forms the basis of speech synthesis and vocoder applications of linear prediction coding.

However, there is a major disadvantage in the use of this method as currently practised. In order to accommodate the three or four formants occurring at frequencies below 4 KHz in normal speech, a PEF of order at least eight is required. Thus, a sample covariance matrix of dimension greater than or equal to eight must be compiled and inverted every twenty milliseconds or so, in real-time speech processing applications. Furthermore, in order to determine the formant peaks in terms of the resulting PEF coefficients, the Fourier transform of the coefficients must be computed as well, and the formant peaks selected from the resulting spectral approximation. These operations require a very extensive amount of computation to be performed at high speeds.

Although primitive speech recognition systems are currently viable, the complexity of the arithmetic manipulations required of them makes speech recognition for extensive
vocabularies difficult to achieve in real time. It is significant that modern electronic devices, which function on a time scale of microseconds or less, are unable to compete with animal nervous systems functioning on a time scale of milliseconds and with considerably less precision.

This fact implies that there must exist algorithms of less complexity than those discussed above, by means of which speech and similar naturally occurring signals can be broken down into more elementary units of information. The very large number of neural paths observed in physiological systems suggests that this end is achieved in Nature by means of a large number of elementary processing units acting on the data simultaneously.

The invention is an implementation of a multi-processing approach to real timespectral analysis analogous to that which must occur in living organisms and is intended to exploit the cheapness and power of contemporary silicon chip technology. A linear prediction method is used but it is a method which avoids the time consuming aspects of conventional linear prediction methods. Essentially it involves a process by which the inversion of a single high order matrix is replaced by the simultaneous inversion of a number of second order matrices to yield a solution to equation (2) above. This solution although mathematically imprecise, is sufficiently accurate for practical purposes.

The spectrum, $S_n$, derived for a particular data sequence is completely controlled by the roots of the polynomial $\gamma(z)$ defined in equation (4). The polynomial may be factorized into a number of binomial factors having real coefficients, viz:

$$\gamma(z) = (1 + a_1z + a_2z^2)(1 + b_1z + b_2z^2)(\ldots\ldots\ldots) \quad (7)$$

Each binomial factor in (7) corresponds to a pair of roots in the complex number plane. Those root pairs which are close to the unit circle are related to peaks in the spectrum, $S_n$, which are called formants in the case
of speech data. Thus some of the binomial coefficient pairs 
\((a_1, a_2), (b_1, b_2)\) etc. are associated with particular 
formants while others which are of less practical importance 
merely control the spectral trend. Now suppose that the 
binomial coefficient pair \((a_1, a_2)\) associated with a 
particular formant is known precisely. They can be used 
as the coefficients of a non-recursive filter which acts 
on the data according to equation (32) below to yield a new 
data sequence whose linear prediction spectrum \(S^*\) is given 
by

\[
S^*(z) = |H(z)|^2 S(z) \tag{8}
\]

where \(H(z)\), the transfer function of the non-recursive 
filter, is given by

\[
H(z) = 1 + a_1 z + a_2 z^2 \tag{9}
\]

obviously

\[
S^*(z) = \frac{P_n}{|\hat{S}^*(z)|^2} \tag{10}
\]

where

\[
\hat{S}^*(z) = (1 + b_1 z + b_2 z^2)(\ldots) \tag{11}
\]

Thus the formant associated with the coefficient 
pair \((a_1, a_2)\) will have been removed from the spectrum and 
the spectrum itself will now be of order \(n-2\). Thus if the 
coefficient pair associated with a second formant could be 
found the process could be repeated and this formant removed 
and so on until no more peaks remained in the spectrum.

In practice of course, the coefficient pair 
associated with a given formant cannot be found precisely 
without resorting to a complete solution of equation (2). 
However, it has been found that if a data sequence having 
a number of peaks in its spectrum is treated as if it were
the outcome of a second order autoregressive process, and its second order PEF coefficients found, they will approximate the coefficients associated with the dominant peak among the peaks in the "true" spectrum. This occurs because of the peak following property discussed in the paragraph following equation (5). The coefficients so found will of course be biased or contaminated by the other peaks in the spectrum. Nevertheless if these coefficients are now convoluted with the original data the result will be a new data sequence in the spectrum of which the dominant peak is considerably attenuated.

If the second order PEF coefficients are found for this second data sequence they will summarize the trend in the spectrum after the partial removal of the dominant peak. Convolution of the original data sequence with this second set of coefficients will have the effect of reducing the residual peaks in this data sequence, leaving the dominant peak more isolated than before and the sequence will yield PEF coefficients less contaminated by the residual peaks.

In general, if this process of convoluting a pair of data sequences with the PEF coefficients derived from the alternate sequence is continued, one sequence of sequences will converge to a limit sequence in which the dominant peak or formant is present in isolation and the other will converge to a limit sequence which has the spectrum of the original sequence but with the dominant peak removed and which can itself be operated on in the same way so as to remove further peaks.

In practice when speech analysis is carried out in real time we are not dealing with finite data sequences with a constant spectral character but rather with streams of data whose spectral character is changing continually with time. Such data streams may not be as convenient to manipulate as the discussion in the preceding paragraphs suggests. Fortunately, a method does exist for computing covariances recursively (see equation (18) etc. below) for
such a data stream and from them computing, as frequently as desired, PEF coefficients which summarize the spectral features of the data stream in the immediate past. The problem is to implement the above algorithm for the isolation of peaks in the case of data streams.

According to an embodiment of the invention the data is divided into a plurality of data streams, for example, stream A and B, whereby the PEF coefficients computed from stream A are convoluted with the original data stream to yield stream B, while the PEF coefficients from stream B are convoluted with the original stream to yield stream A. However, this is not the entire solution. The PEF coefficients yielded by both streams are identical and are a poor approximation to the PEF coefficients of the dominant peak. Some asymmetry must be deliberately introduced into such a network.

A method of introducing an asymmetry is to filter stream A with a recursive filter (see equation (33) below) whose coefficients are equal to or derived from the PEF coefficients computed from stream A itself. This method is effective in isolating the dominant formant in stream A while stream B comprises a stream in which the dominant formant has been removed and which can be further processed in order to isolate remaining formants.

It should be appreciated that the invention is not restricted to the above described embodiment. A wide variety of networks is possible in which PEF coefficients estimated from various data streams are used to filter other data streams in order to locate spectral features in applications where conventional spectral methods may be too slow or otherwise inconvenient. The network selected to perform a particular task will depend upon the application; the type of spectral information required, the available hardware and so on. Even in the case of speech analysis different networks may be appropriate to each of three distinct applications, viz: phoneme recognition, speaker identification and speech compression for storage and transmission.
A cascaded two stream embodiment will yield, at any instant, several pairs of coefficients, one pair from each module associated with the formant isolated in the A stream by that module, plus a final coefficient pair associated with the B stream of the final module from which all the formants have been removed.

Each pair of PEF coefficients, $C_1$ and $C_2$, contain information about both the centre frequency and the bandwidth of the corresponding formant (see equation (34) below). This second dimension of information is extremely useful in practice since the second coefficient can be used directly as a criterion for accepting or rejecting a particular formant during phoneme recognition; small values of $C_2$, that is, less than some threshold value, correspond to peaks which are too broad and weak to be considered valid as formants. The value of the threshold chosen depends on the time constant $T$, used in the covariance computation and on the feedback factor $F$ (equations (12) and (13)) which has been used. A value of about .9 would be typical for $T = 10$ msec, $F = 0.5$ (sampling frequency 10KHz).

In order that the invention may be more readily understood, one specific embodiment in the form of a formant tracker for use with speech will now be described in detail with reference to the accompanying drawings. In the drawings:

Fig. 1 shows a circuit block diagram of an embodiment of the device as a formant tracker,

Fig. 2 shows a circuit block diagram of one of the second order prediction filter estimators depicted by circles in Fig. 1,

Fig. 3 is a graph showing the Fourier transform log power spectrum of a segment of the data sequence fed to the device at line 1.

Fig. 4 is a graph showing the Fourier transform log power spectrum of a segment of the data sequence fed to the device at line 4, after the first formant has been removed.
Fig. 5 is a graph showing the Fourier transform log power spectrum of a segment of the data sequence appearing at line 5. The second formant occurs in isolation.

Fig. 6 is a graph showing the Fourier transform log power spectrum of a segment of the data sequence appearing at line 6.

In Fig. 1 rectangles represent filters, either non-recursive (N) or recursive (R), circles represent prediction filter estimators, triangles represent prediction filter coefficient modifiers, lines with arrows represent the paths by which filter coefficients are passed from estimators to filters and lines without arrows represent paths by which data sequences or data streams are passed from filter to filter or from filters to estimators. Peripheral devices such as microphones, analogue filters and clocks are not shown.

Inspection of Fig. 1, reveals that it comprises three modules fed respectively by data streams 2, 4 and 6. The modules are identical except for the first and are arranged in the form of a hierarchy or cascade, each module except the last passing a data sequence to the next in line.

Consider the action of a single module for example the module fed by data sequence 4. The module comprises two non-recursive filters 11 and 12, one recursive filter 13, two prediction error filter (PEF) estimators 14 and 15, a coefficient modifier 16 and an output buffer 17. The PEF coefficients computed by estimator 15 are passed to the non-recursive filter 11 and the PEF coefficients computed by estimator 14 are passed to non-recursive filter 12, to recursive filter 13, via the coefficient modifier 16 and to the buffer 17.

When the device is switched on all the coefficients are set to zero with the result that the filters all act as identity filters and have no effect on the data sequence passing through them. The data sequence 4 arrives at both PEF estimators 14 and 15. Identical pairs of coefficients,
related to the dominant peak in the spectrum of data sequence 4, are therefore computed by 14 and 15 and passed to the various filters. Ignore, for a moment, the action of the modifier, 16, and assume that the coefficients are passed unchanged to filter 13. The filter 13, and the filter 11 will have opposite effects on the data sequence 4 which will appear initially unchanged at 5 while the filter 12 will have the effect of attenuating the dominant peak in the spectrum of data sequence 4 since this is one of the properties of prediction error filters.

Consequently an asymmetry is immediately introduced into the module. The PEF estimator 15, will now compute PEF coefficients related to the residual peaks in data sequence 6 and pass these coefficients to filter 11. The residual peaks will then be attenuated in datastream 5 allowing PEF estimator 14 to compute coefficients relating to the dominant peak which are less contaminated by the residual peaks than was previously the case. This effect of isolating the dominant peak will be further reinforced by the action of filter 13. In practice the data sequence 5 rapidly converges to a data sequence whose spectrum contains only the dominant peak while data sequence 6 converges to a sequence whose spectrum contains only the residual information and is passed to the next module in order to isolate further peaks in the same way. The coefficients describing the dominant peak which is now isolated in data sequence 5, are passed to coefficient buffer 17. After a short time, in this way, coefficient buffer 17, 18, 19 and 20 contain PEF coefficient pairs associated with each of the peaks in the input sequence 1.

In practice the coefficients used in the filters 11, 12 and 13 are initially set to values close to the values they are expected to assume thus allowing more rapid convergence to take place.

It can be seen that the action of the non-recursive filters constitute a type of negative feedback since they
nullify an effect while the recursive filters constitute a form of positive feedback since they exaggerate an effect detected by the PEF estimators. If the filter 13 were not present there would be no asymmetry in the module and data streams 5 and 6 would remain identical. On the other hand the effect of the recursive filter as described above is too strong and although the action of the device in separating formants commences adequately, the device soon ceases to be responsive to changes in the incoming data stream. This effect can be overcome by lessening the degree of "positive feedback" i.e. by modifying the coefficients which are passed to the recursive filters. One way of doing this is to multiply \( C_2 \) by an attenuation factor \( F \) to yield a new coefficient \( C_2^* \). In order that this does not lead to a shift in the peak frequency associated with the coefficient pair, \( C_1 \) must be modified in such a way as to keep the peak frequency constant thus

\[
C_2^* = FC_2
\]  

(12)

and

\[
C_1^* = FC_1 \frac{1 + C_2}{(1 + C_2^*)}
\]

In practice, the simpler formula

\[
C_1^* = FC_1
\]  

(13)

is quite adequate.

These equations summarize the action of the coefficient modifier \( F \). The value of \( F \) is not critical. A value of 0.5 is used in this embodiment allowing the multiplication to be performed merely by right shifting the numbers.

It was found experimentally that the first formant is not completely removed by the action of a module such as the one described which left behind two small residual peaks not present in the original data spectrum. The first
formant behaves as a double resonance and the effect is easily overcome by the inclusion of two non-recursive filters (22 & 23) on one side of the first module.

Another anomaly occurs when the first formant is removed. The remaining peaks in the spectrum of data sequence 4 are frequently distorted in magnitude to the degree that the higher frequency peaks may predominate in the residual spectrum resulting in the fourth formant being removed second. This effect is overcome and the formants removed in the correct order by prior filtering of the data by a filter 24 with fixed coefficients. However, the order in which the formants are removed may not matter.

A block diagram of one specific embodiment of the second order PEF estimators 14 and 15 referred to above is depicted in Fig. 2. This description is given in digital terms although analogue embodiments are equally feasible. The data are presented one at a time by some external device such as a digitizer (not shown) to line 41 in the diagram. Delays 42 and 43 and lines 41, 44 and 45 constitute a three word shift register so that at time iΔt the quantities $x_i$, $x_{i-1}$ and $x_{i-2}$ appear at lines 41, 44 and 45 respectively.

These quantities are multiplied in pairs by multipliers 46, 48 and 50 and added to previously computed values of the covariances which have been multiplied by an attenuation factor, $\rho$, by multipliers 47, 49 and 51. Thus the current values of the covariances appear at the output of the adders 52, 53 and 54 once per clock cycle. Values of the variances and covariances computed in this way are passed via delays 55, 56, 57, 58 and 59 to lines 60, 61, 62, 63 and 64 where they are used to compute new variance/covariance values and to compute the PEF coefficients themselves.

The latter operation is commenced by multiplying the variances and covariances in pairs by multipliers 65, 66, 67, 68, 69 and 70 and passing the products to subtractors 71, 72 and 73 where their differences are found.
Finally the output from subtractors, 72 and 73 are divided by the output of subtractors 71 to yield the second order PEF coefficients $C_1$ and $C_2$ in the output buffers 74 and 75 in accordance with equations (30) and (31) below.

Some simplifications in the PEF estimators may be possible in practice. For example, the last step of division may be avoided and the output from the subtractors used themselves as the coefficients of a non-recursive filter since they are in the same proportion as the PEF coefficients. The attenuating factor $\rho$ will usually be very close to unity and the attenuation of previous covariance values may best be carried out less frequently than once every clock cycle, that is, previous values can be multiplied by $N \rho$ every $N$ clock cycles.

The PEF coefficients themselves do not change rapidly with time and they too need be computed less frequently than once every clock cycle.

Another practical simplification which may be advantageous in some circumstances is the removal of PEF estimator 15 and non-recursive filter 11 from the circuit shown in Fig. 1 (and likewise the corresponding elements in the other modules in Fig. 1). Recursive filter 13, PEF estimator 14 and coefficient modifier 16 will act alone to isolate the formant in the data, while non-recursive filter 12 will act to remove this formant from stream 6 in the same way as in the original circuit. However, the resulting embodiment is not quite as effective in following formants as they change with time as is the originally described embodiment. Nevertheless and notwithstanding the description given above the combination of elements 13, 14 and 16 can be seen as comprising the basic circuit for formant isolation with elements 11, 12 and 15 comprising a refinement for better operation of this basic circuit.

Fig. 3 shows the input signal on line 1 of Fig. 1 wherein it can be seen that there are four formants $F_1$ to $F_4$ present in the spectrum. This diagram comprises a 1024
point Fourier transform log power spectrum of the utterance "i" in the word "television".

Fig. 4 shows the signal on line 4 of Fig. 1 wherein it can be seen that the first formant \( F_1 \) has been removed and the remaining formants are more pronounced.

Similarly Fig. 6 shows the signal on line 6 of Fig. 1 wherein the first two formants have been removed.

Fig. 5 shows the second formant in isolation, that is, the spectrum of the data stream on line 5 of Fig. 1. The coefficient pair summarizing this spectrum appears in buffer 17 of Fig. 1.

The operation can be summarized mathematically as follows:

In the case where the data sequence comprises a discrete set of quantities, \( \{ x_i \} \), one definition of the elements, \( r_{ij} \), of the variances and covariances, \( r_{ij} \),
referred to is as follows:

\[ R_{ij} = \sum_{p=3}^{N} x_{p-i} \cdot x_{p-j} \]  \hspace{1cm} (14)

In the analogue case where the data comprises a function \( x(t) \) defined over a domain \((0,N \Delta t)\) of \( t \), one definition of the variances and covariances is as follows:

\[ R_{ij} = \int_{3 \Delta t}^{N \Delta t} x(t-i \Delta t) \cdot x(t-j \Delta t) \, dt \]  \hspace{1cm} (15)

where \( \Delta t \) is a constant lag or separation of the data in the domain.

The above definitions assume that the data sequence has zero mean. This would be achieved in practice by prior filtering of the data. These definitions differ from the usual definitions in that there is no division by the sequence length, \( N \). This scaling factor is not required as the PEF coefficients are scale free.

In some applications it may be convenient to assume that the variances and covariances at each lag are equal viz: that

\[ R_{00} = R_{11} = R_{22} \]  \hspace{1cm} (16)

and

\[ R_{01} = R_{12} \]  \hspace{1cm} (17)

This approximation leads to some degradation in accuracy and reliability in the case of speech data.

Another simplification is to compute each variance or covariance recursively, viz:

\[ r_{22}(t) = x_t^2 + \rho r_{22}(t-1) \]  \hspace{1cm} (18)

\[ r_{11}(t) = r_{22}(t-1) \]  \hspace{1cm} (19)
\[ r_{00}(t) = r_{11}(t-1) \]  
\[ r_{12}(t) = x_t x_{t-1} + \rho r_{12}(t-1) \]  
\[ r_{01}(t) = r_{12}(t-1) \]  
and \[ r_{02}(t) = x_t x_{t-2} + \rho r_{02}(t-1) \]

where \( \rho \) is a positive constant less than unity which causes the values of \( r_{ij}(t) \) computed in this way to be bounded. It can be shown that this method of computing \( r_{ij} \) is equivalent to using a tapered window on the data, that is \( r_{ij}(t) \) is in fact the variance/covariance of a data sequence \( \{ y_i(t) \} \) defined in terms of the original data sequence at time \( t \) by

\[ y_i(t) = a^{p} x_{t+p}, \text{ for } p \leq 0, \]  
where \( a = \rho^{-\frac{1}{2}} \)

Thus past values of the data sequence are weighted with an exponential decay. The time constant \( T \) of the decay, where

\[ T = 1/\log a \]  
or \[ T = \Delta t/\log a \]

takes the place of the frame length \( N \) or \( N \Delta t \) which occurred in the original definitions.

The prediction error filter coefficients \( C_0, C_1 \) and \( C_2 \) are found in terms of the covariances by solving the equations

\[ C_0 = 1 \]  
\[ r_{01} + C_1 r_{11} + C_2 r_{12} = 0 \]
\[ r_{02} + C_1 r_{12} + C_2 r_{22} = 0 \]  
(29)

The solutions are

\[ C_1 = \frac{(r_{12} r_{02} - r_{01} r_{22})}{(r_{11} r_{22} - r_{12}^2)} \]  
(30)

\[ C_2 = \frac{(r_{12} r_{01} - r_{02} r_{11})}{(r_{11} r_{22} - r_{12}^2)} \]  
(31)

These coefficients may be used as the coefficients of a non-recursive (or "finite impulse response") filter which acts on a data sequence \( \{ x_i \} \) to yield a new data sequence \( \{ y_i \} \), viz:

\[ y_i = x_i + C_1 x_{i-1} + C_2 x_{i-2} \]  
(32)

This operation is also referred to as the "convolution" of the data sequence \( \{ x_i \} \) with the PEF coefficients \( \{ 1, C_1, C_2 \} \) to yield a new data sequence \( \{ y_i \} \).

Alternatively the PEF coefficients may be used as the coefficients of a recursive (or "infinite impulse response") filter which acts on a data sequence \( \{ u_i \} \) to yield a new data sequence \( \{ \nu_i \} \), viz:

\[ \nu_i = -C_1 \nu_{i-1} - C_2 \nu_{i-2} + u_i \]  
(33)

The coefficients \( C_1 \) and \( C_2 \) summarize the gross features of the power spectrum (i.e. the power spectral density function) of the data sequence from which they were derived. In the case where the spectrum has a single dominant peak as in Fig. 5 the frequency of the peak, \( f_o \), is given by

\[ \cos (2 \pi \Delta t f_o) = \frac{C_1 (1 + C_2)}{4C_2} \]  
(34)

where \( \Delta t \) is the sampling interval in the discrete case.

The coefficient \( C_2 \) is controlled by the half power width
of the peak. It is close to unity when the peak is narrow
and is closer to zero when the peak is broad or where more
than one peak is present in the spectrum. Thus $C_2$ can be
used in some threshold criterion to decide whether a peak
is sufficiently narrow to be classified as a single
formant and the quantity, $f_o$, can be used to determine
the frequency of that formant. In practice $C_1$ and $C_2$
themselves or simple functions of them can be checked
against population ranges in order to classify a formant.

As a further variation with reference to Fig. 1,
a network for the recognition of different phonemes in
speech data could dispense with the coefficient modifiers
such as 16 and replace them with switches so that full
positive feedback is maintained for a short time causing
rapid convergence in the isolation of the formants. Once
a particular group of formants has been isolated and
"recognized", the recursive filters such as 13 can be switched
out of the network for the remainder of the duration of
the phoneme. Statistically significant changes in the
coefficient values appearing in the buffers 17, 18, 19
and 20 can be used to detect the onset of a new phoneme and
so cause the recursive filters to again become operative.
In this way a speech recognition device can be constructed
which is phoneme synchronous, thus avoiding the need for
time axis normalization.

The simplest embodiment of the invention comprises
the PEF estimator of Fig. 2. This device alone is unsuited
to the analysis of spectrally complex signals such as
speech but it can be used for the recognition of simpler
signals such as the signalling tones used in telephone
switching systems. In this way a single device can be
used to distinguish between a wide variety of tones when
the input (line 1 in Fig. 2) is fed from an appropriate
line in the telephone system. The coefficients $C_1$ and $C_2$
generated by the device are then compared with predefined
values in order to classify the incoming signal and to
cause the remainder of the telephone system to take appropriate action. An analogue embodiment of the device utilizing fixed delays in place of shift registers would be more suited to this application.
1. A method of data processing by linear prediction characterized in that it includes the step of estimating the second order prediction error filter (PEF) coefficients for said data in order to approximate the pair of coefficients occurring in a binomial factor of the $z$ transform of a sequence of higher order PEF coefficients of the said data, said binomial coefficient pair being associated with a formant in the spectrum of said data.

2. A method according to claim 1 further characterized in that said second order PEF coefficients are utilized to control filter means so as to modify the spectrum of said data according to changes in said data.

3. A method according to claim 2 further characterized in that a process of successive approximation is used to isolate said formant.

4. A method according to claim 2 further characterized in that said formant becomes attenuated in the said spectrum.

5. A method according to claim 3 further characterized in that said filter means comprises a first and a second filter and said method involves feeding said data to each respective filter to obtain a first derived data stream and a second derived data stream and said step of determining the PEF coefficients comprises determining the said coefficients for the said first derived data stream and the said step of utilizing said coefficients to control said filter means comprises using the said coefficients from said first derived data stream to control said first filter and said second filter whereby said formant becomes isolated in said first data stream and attenuated in said second data stream.

6. A method according to claim 5 characterized in that said first filter comprises a recursive filter, the values of the coefficients of which are controlled by the values of
the said second order PEF coefficients and said second filter comprises a non-recursive filter the values of the coefficients of which are also controlled by the values of the said second order PEF coefficients.

7. A method according to claim 5 characterized in that said second data stream is further processed in a similar manner to the processing of said data, in order to isolate and remove a further formant.

8. Apparatus for data processing by linear prediction characterized in that it includes a second order prediction error filter (PEF) estimator (14) adapted to estimate the second order PEF coefficients for said data (4) in order to approximate the pair of coefficients occurring in a binomial factor of the z transform of a sequence of higher order PEF coefficients of the said data (4), said binomial coefficient pair being associated with a formant (F₂) in the spectrum of said data.

9. Apparatus according to claim 8 characterized in that it includes filter means (12, 13) adapted to receive said data (4) and said second order PEF coefficients whereby said coefficients control said filter means (12, 13) so as to modify the spectrum of said data (4) according to changes in said data (4).

10. Apparatus according to claim 9 further characterized in that said filter means provides derived data from said data, and said PEF estimator estimates said second order PEF coefficients for said derived data whereby said formant becomes isolated in said derived data.

11. Apparatus according to claim 10 characterized in that said modifying of the spectrum causes said formant (F₂) to become attenuated in the said spectrum.
12. Apparatus according to claim 11 characterized in that said filter means (12, 13) comprises a first filter (13) and a second filter (12) for receiving said data (4) and providing first (5) and second (6) derived data streams, respectively, and said PEF estimator (14) is adapted to estimate the said coefficients for said first derived data stream (5) and provide said coefficients to said first filter (13) and said second filter (12) to cause said formant to become isolated in said first derived data stream (5) and attenuated in said second derived data stream (6).

13. Apparatus according to claim 12 characterized in that said first filter (13) comprises a recursive filter the values of the coefficients of which are controlled by the values of the said second order PEF coefficients and said second filter (12) comprises a non-recursive filter the values of the coefficients of which are also controlled by the values of the said second order PEF coefficients.

14. Apparatus according to claim 13 characterized in that said recursive filter (13) includes a non-recursive filter part (11) and a second PEF estimator (15) is arranged to receive said second derived data stream (4, 6) and estimate said coefficients therein, the said coefficients estimated in said second PEF estimator (15) being provided to control said non-recursive filter part (11).

15. Apparatus according to claim 14 characterized in that it includes a plurality of similar cascaded modules each adapted to receive said second derived data stream (4, 6) from the preceding module and process the said second derived data stream in a manner to isolate a further formant (F₃) and attenuate said further formant from the derived data stream to the next cascaded module.
16. Apparatus according to claim 8 characterized in that said PEF estimator (14) comprises correlators which cause the variants and first and second covariances of the data to be computed, multipliers (65 – 70) which multiply said variants and said covariances in pairs to yield products, and subtractors (71 – 73) which subtract said products to yield differences; whereby said differences are in proportion to the sample second order PEF coefficients of the said data.
**INTERNATIONAL SEARCH REPORT**

**International Application No.** PCT/AU 81/00060

---

### I. CLASSIFICATION OF SUBJECT MATTER

According to International Patent Classification (IPC) or to both National Classification and IPC:

- **G10L 1/02, 1/04, H03H 17/00, 17/02**

---

### II. FIELDS SEARCHED

#### Classification System

<table>
<thead>
<tr>
<th>Classification System</th>
<th>Classification Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPC</td>
<td>H03H 7/28, 15/00, 17/00, 17/02, 17/04, 17/06, G10L 1/02</td>
</tr>
<tr>
<td>US. Cl.</td>
<td>179/15A, 179/15C, 324/77R</td>
</tr>
</tbody>
</table>

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched:

- **AU: IPC as above; Australian Classification 05.7; 05.52; 05.2**

---

### III. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of Document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to Claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>US, A, 3296374, published 1967, January 3, Clapper</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>US, A, 3327057, published 1967, June 20, Coker</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>US, A, 3335225, published 1967, August 8, Campanella et al.</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>US, A, 3369076, published 1968, February 13, Bakis</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>US, A, 4032711, published 1977, June 28, Sambur</td>
<td></td>
</tr>
</tbody>
</table>

* Special categories of cited documents:
  - "A" document defining the general state of the art
  - "E" earlier document but published on or after the international filing date
  - "F" document cited for special reason other than those referred to in the other categories
  - "O" document referring to an oral disclosure, use, exhibition or other means
  - "P" document published prior to the international filing date but on or after the priority date claimed
  - "T" later document published on or after the international filing date or priority date and not in conflict with the application, but cited to understand the principle or theory underlying the invention
  - "X" document of particular relevance

---

### IV. CERTIFICATION

- **Date of the Actual Completion of the International Search:** 09 July 1981 (09.07.81)
- **Date of Mailing of this International Search Report:** 13 July 1981 (13.07.81)

**International Searching Authority:** Australian Patent Office

**Signature of Authorized Officer:**

A.S. Moore. A.M. Moore

Form PCT/ISA/210 (second sheet) (October 1977)