

**EUROPEAN PATENT APPLICATION**

Application number: 89302481.0

Int. Cl.<sup>4</sup>: G10L 9/14

Date of filing: 14.03.89

Priority: 16.03.88 GB 8806185

Date of publication of application:  
20.09.89 Bulletin 89/38

Designated Contracting States:  
AT BE CH DE ES FR GB GR IT LI LU NL SE

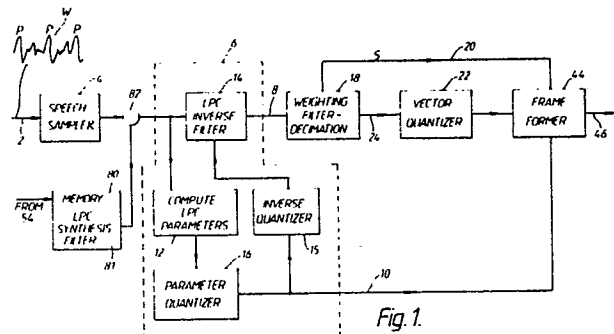
Applicant: **UNIVERSITY OF SURREY**  
**Guildford**  
**Surrey GU2 5XH(GB)**

Inventor: **Evans, Barry G. Dept. of Electronic & Elect. Eng.**  
**University of Surrey**  
**Guildford Surrey GU2 5XH(GB)**  
Inventor: **Kondoz, Ahmet M. Dept. of Electronic & Elect. Eng.**  
**University of Surrey**  
**Guildford Surrey GU2 5XH(GB)**

Representative: **Knott, Stephen Gilbert et al**  
**MATHISEN, MACARA & CO. The Coach**  
**House 6-8 Swakeleys Road**  
**Ickenham Uxbridge Middlesex UB10 8BZ(GB)**

**Speech coding.**

Speech for transmission over a low bit rate channel of a telecommunications link is digitally sampled and subjected to linear predictive coding in a coder (6). The LPC residual signal is passed through a weighting filter (18) and the output down sampled by decimation factor  $d$  resulting in  $d$  sequences of which the maximum energy sequence is selected for vector quantization. The data output from the vector quantizer (22) is formed into an output frame together with an index ( $s$ ) identifying the selected decimation sequence and the quantized LPC parameters. The frame transmitted over the telecommunications link (46) is decoded by means of an inverse vector quantizer (50) which restores the decimated sequence. An interpolator (52) interpolates zeros to restore an excitation signal for an LPC synthesis filter (54) from which the voice signal can be reconstructed.



**EP 0 333 425 A2**

## SPEECH CODING

The present invention relates to speech coders and, more particularly, to low bit rate speech coders. The invention also relates to a method of coding speech for transmission along a telecommunications link.

Speech is a complex analogue waveform. In order to transmit a speech signal along a digital telecommunications link, the information contained in the analogue signal must be reduced to information in digital form. This technique is known as speech coding. In a simple form of coding, the analogue speech signal is sampled and a digital representation of the amplitude of the signal at each sampling point is transmitted along the telecommunications link. This type of coding is pulse code modulation (PCM). The quality of sound reproduction using PCM depends on the sampling rate and also on the number of digits used to transmit each sample which determines the "quantization" or number of discrete amplitude levels that can be distinguished.

Currently, digital telephone networks use 64 Kb/s PCM or 32 Kb/s adaptive PCM (ADPCM). Speech coders requiring 16 Kb/s are proposed for use in the European mobile radio standard (GSM). For mobile telecommunications and telecommunications links involving satellites, for example satellite to aircraft or satellite to land mobile links, the bit rate required is crucial if a reasonable number of channels are to be available for use and if the system is to be economic.

There is therefore a significant technical problem represented by the need to achieve good speech quality while minimising the bit rate necessary for each channel in a telecommunications link.

One possible solution to this problem uses the technique known as code excited linear predictive coding (CELP). Algorithms for implementing coders using CELP for producing high quality speech at very low bit rates at around 7Kb/s have been described, for example in an article by M R Schroeder and B S Atal entitled "Code-Excited Linear Prediction (CELP): High Quality Speech at very Low Bit Rates" Proc. of ICASSP-85 pages 937 to 940. The algorithms which have been proposed to date require extremely complex processing of the input speech samples in order to produce the required bits for transmission. For telecommunications purposes it must be possible to carry out the encoding and decoding operations in real time. The existing algorithms require very large quantities of operations so that if they are to be carried out in real time, the amount of computing power that needs to be utilised is excessive for inclusion in, say, a mobile telephone subscriber's telephone

equipment, or even for use in a telephone exchange if the link between the subscriber and the exchange is able to have a greater capacity. Some reductions in bit rate are possible by concentrating on the base-band information in the speech signal as originally proposed by the inventors in a paper entitled "Low bit rate speech coding" IEE Symposium Digest Radio No. 1987/52 April 1987 pages 1 to 4.

The technical problems addressed by the present invention are therefore to provide a method of speech coding and a speech coder which are capable of operation in real time without requiring excessive amounts of processing power, and to provide improved speech quality relative to that available from existing coders below 7Kb/s.

One of the objects of the present invention is to produce a speech coder in which an encoder and a decoder can be implemented on a single commercially available digital signal processing chip. The encoders and decoders in accordance with the invention which will be described are each capable of implementation in real time using a DSP-32 floating point chip as manufactured by AT&T.

Such a single chip implementation is to be considered as a reasonable amount of processing power for practical commercial applications. It will be appreciated that alternative DSP chips may be used as they become available.

A second object of the present invention is to improve the digital speech quality significantly below 7Kb/s and to produce good quality at around 4Kb/s.

According to one aspect of the invention, there is provided a speech coder for encoding an input speech signal for transmission over a digital channel of a telecommunications link, comprising means for sampling the input speech signal to produce output digital samples,

means for dividing these digital samples into frames each consisting of a predetermined number of samples,

a linear predictive filter for inverse filtering each frame and producing an output LPC residual signal for said frame comprising a said predetermined number of digital samples, and LPC parameters for said frame, and

baseband extraction means, the speech coder being characterised by provision of

down-sampling means for extracting from the output of said base-band extraction means d interleaved sequences,

means for selecting one of said sequences which contains the maximum energy content and produc-

ing an output index representing the selected sequence,  
 means for deriving pitch period and pitch gain indices from the selected sequence,  
 means for removing long term correlation from the selected sequence to produce a remainder sequence,  
 means for comparing the remainder sequence with an identifiable reference sequence, and for deriving a scale factor from the compared sequences, the scale factor being defined by a scale factor index and being representative of the energy in the remainder sequence relative to that in the reference sequence,  
 wherein for each frame of the input, an output frame comprising data representing the LPC parameters, and for each block, the index of the selected sequence, and the scale factor, pitch period and pitch gain indices and an index representing said identifiable reference sequence is transmitted over the channel of the telecommunications link.

The baseband extraction means may comprise a weighting filter which amplifies a low frequency pitch component of the LPC residual signal and reduces the amplitude of higher frequency components of the LPC residual signal and said samples of the LPC residual signals are divided into blocks before being passed through the filter or alternatively the baseband extraction means may effect multipulse linear predictive analysis-by-synthesis whereby said baseband is obtained by minimizing an error between input speech signals and artificially reconstructed speech signals.

In a preferred embodiment of the invention the means for comparing may include a vector quantizer for matching the remainder sequence with the most closely resembling one of a plurality of vectors stored in a codebook, each stored vector being a random sequence having a Gaussian distribution and being identifiable by a unique index, and the output frame includes data representing the index of the selected vector.

By appropriate selection of the order of the LPC filter, which determines the number of parameters that need to be transmitted for each frame, the number of blocks in a frame, the block size, the decimation factor and the total number of vectors in the vector codebook and possibly other variables, the amount of data to be transmitted for each frame can be controlled in such a way as to select a bit rate in the range 2.4 Kb/s to 9.6 Kb/s which produces acceptable speech quality.

The technique of LPC filtering is already known and for a fuller description of the technique the reader is referred to: "Linear Prediction: A tutorial review" by J. Makhoul in Proc. IEEE, Vol-63, Pages 561-580, 1975. In most current implementations of

LPC designed for use at low bit rates the LPC parameters to be transmitted are usually scalar quantized and may be transformed into line spectral pairs (see: "Line spectrum pair (LSP) and speech data compression" by F.K. Soong and B.H. Juang in ICASSP-84 pages 1.10.1 to 1.10.4). In a preferred embodiment of the present invention a Gaussian codebook vector quantization technique is employed which allows even lower bit rates to be achieved without reduction in speech quality. This type of vector quantization may be used after the LPC parameters have been transformed into line spectral pairs.

Vector quantization is also a standard technique and reference may be made for example to: "Vector quantization" by R.M. Gray in IEEE ASSP Magazine, Vol-1 pp4-29, 1984. However, the embodiment of vector quantization using a Gaussian codebook and a scale factor as described more fully in the accompanying specific description is believed to be novel. The advantage of the proposed configuration is in the use of an analysis-by-synthesis procedure based around a pitch synthesis filter to select the optimum sequence from the Gaussian codebook and to compute its optimum scale factor.

The weighting filter is preferably a digital finite impulse response filter which has a gain-frequency characteristic which places emphasis on the large pulses in the signal represented by the input samples representing the LPC residual signal. These pulses occur periodically at a frequency corresponding to the underlying pitch of the voice signal. While the amplitude of these large pulses is relatively increased, the amplitude of the higher frequency components which contain proportionately less information is reduced. A typical filter characteristic is shown in Fig. 2. The purpose of the weighting filter is to produce near-optimal excitation pulses as in the Multi-Pulse LPC proposed by P. Kroon et al in "Regular pulse excitation - A novel approach to effective and efficient multi-pulse coding of speech" IEEE Trans, ASSP-34 pp 1054-1063, 1986.

Further improvements in the speech quality can be produced if a pitch filter is used to remove from the LPC residual signal a signal representing the pitch pulses. The parameters of such a pitch filter are preferably set by analysing the original LPC residual signal and a pitch filter memory. The pitch filter is placed in a feedback loop in which data for transmission over the telecommunications link is fed to a decoder which carries out the inverse operations of the described coding vector quantizer and decimation to reproduce an LPC excitation signal, which is fed via the pitch filter and subtracted from the actual LPC residual signal so as to enhance the effect of the weighting filter and

place more emphasis on the base band component of the speech signal. Using such a pitch filter in a feedback loop, the weighting filter, down-sampling and vector quantization steps effectively result in a minimisation of the difference between the output of the vector quantizer and the input to the weighting filter. When such a pitch filter is in use, the data to be transmitted along the telecommunications link includes pitch data relating to the pitch amplitude and period of the feedback pitch filter. Extra bits are required for the transmission of this information and if it is necessary to keep the transmission rate constant, the bit rate occupied by the data output from the vector quantizer operating on the selected down-sampled sequence can be commensurately reduced without any reduction in the speech quality because there is now less information in the signal subject to vector quantization. This pitch filter is able to operate in this manner because of the gain frequency characteristic of the weighting filter and the equivalent results would not be produced if a plain low-pass filter were used instead.

According to another aspect of the invention, there is provided a speech coder for decoding speech encoded with the encoder in accordance with said first aspect, the coder comprising means for separating from a received frame the scale factor index, and the data representing the LPC parameters,

means for outputting into a pitch synthesis filter a sequence corresponding to said identifiable sequence scaled by said scale factor, an interpolator for receiving said output sequence and said selected sequence index and interpolating zeros at appropriate positions in order to produce an LPC excitation signal, and an LPC synthesis filter for receiving said excitation signal and said data representing the LPC parameters and for restoring therefrom a sequence of digital samples representing the input speech signal.

In a preferred embodiment the outputting means may comprise an inverse vector quantizer including a codebook corresponding to the codebook in the vector quantizer of the encoder, and the inverse vector quantizer receives said unique index and the scale factor index and outputs in response thereto, as the output sequence, a corresponding sequence scaled by said scale factor.

Some embodiments of speech encoders and decoders in accordance with the present invention incorporating the novel algorithm will now be described, by way of example only, with reference to the accompanying diagrammatic drawings, in which:

Figure 1 is a block diagram of a first embodiment of a speech encoder;

Figure 2 is a frequency-gain characteristic of the weighting filter used in the encoder of Figure 1;

Figure 3 is a block diagram of a speech decoder for use with the encoder of Figure 1;

Figure 4 is a block diagram of the vector quantizer used in the embodiment of Figure 1 and Figure 5;

Figure 5 is a block diagram of a second embodiment of a speech encoder using a pitch filter;

Figure 6 is a block diagram of a decoder for use with the encoder of Figure 5;

Figure 7 is a block diagram of a vector quantizer for quantizing the LPC parameters for transmission in either of the embodiments; and

Figure 8 is a diagram illustrating the frame structure of information to be transmitted.

It will be appreciated that the encoders and decoders described hereinafter are implemented as software instructions carried out in a digital signal processor such as the DSP-32 chip referred to previously. The blocks shown in the drawings are intended merely to facilitate explanation of the functions of each of the processing steps carried out, rather than to indicate discrete components in the speech coder.

A speech channel of a telecommunications link using a speech coder requires an encoder at the voice signal input end and a decoder at the reception end. Therefore the speech coder associated with one end of the telecommunications link requires both an encoder and a decoder, which may be connected to separate channels in the case of a duplex link or the same channel in the case of a simplex link. In the first embodiment of the invention the encoder is diagrammatically illustrated in Figure 1 and the corresponding decoder is shown in Figure 3. Both the encoder and decoder may be implemented using the same digital signal processor.

Referring initially to Figure 1, the analogue speech signal input on line 2 of the encoder has a complex waveform (W) exhibiting, inter alia, relatively large amplitude pulses P, known as "pitch pulses", which are a characteristic of analogue speech signals.

The analogue speech signal is input on line 2 to a speech sampler 4 which samples the analogue speech signal and produces a series of digital samples. When a DSP-32 chip is employed to implement the encoder, this has the capability for direct interfacing of an 8-bit encoder capable of sampling eight thousand times a second.

The output samples are divided into frames of, for example, 200 8-bit samples each, and the en-

coder is effective to translate the samples in each frame into a number of quantization indices which represent the input waveform but consist of relatively few bits, thereby facilitating a low bit rate. The frame size may be adjusted to suit the final bit rate required.

The digital samples in each frame are first input to a linear predictive filter (LPC) 6. Linear predictive filtering is a known technique and so the processing of the input samples will not be described in detail. However, in general terms, a linear predictive filter of order  $k$  will attempt to establish a linear relationship between each input sample and the  $k$  preceding samples. Therefore, if the  $i^{\text{th}}$  input sample is represented as  $\underline{a}_i$  and the LPC parameters as  $\underline{b}_j$ , then

$$\underline{a}_i = \sum_{j=1}^k \underline{b}_j \underline{a}_{i-j}$$

The LPC parameters  $\underline{b}_j$  are computed in a processing circuit 12 for each input digital sample  $\underline{a}_i$ . The LPC parameters  $\underline{b}_j$  derived for all the samples in the current frame are then fed to a parameter quantizer 16 which generates quantization indices therefrom, and these indices are routed on line 10 to a frame-forming circuit 44. The quantization indices are also routed to an inverse quantizer 15 which re-generates the parameters  $\underline{b}_j$ , though the original and the re-generated parameters  $\underline{b}_j$  will not be identical due to the effect of processing the signals in the quantizer 16 and the inverse quantizer 15.

The re-generated LPC parameters  $\underline{b}_j$  are passed to an LPC inverse filter 14 which generates a further sample  $\underline{c}_i$  representing the difference between the corresponding input sample  $\underline{a}_i$  and a predicted value thereof, evaluated using the re-generated parameters  $\underline{b}_j$ . Thus,

$$\underline{c}_i = \underline{a}_i - \sum_{j=1}^k \underline{b}_j \underline{a}_{i-j}$$

The samples  $\underline{c}_i$  constitute an LPC residual signal, there being as many samples  $\underline{c}_i$  as there are input samples  $\underline{a}_i$ .

The LPC residual signal generated at the output of linear predictive filter 6 is then subjected to further quantization, as will now be described.

Each frame of samples of the LPC residual

signal is divided into blocks. The frame represented in Figure 8 has been divided into four blocks which, in this example, would each contain 50 samples. In the filtering approach, as distinct from the multi-pulse method, each of these blocks is then fed separately into a weighting filter which is part of a processing circuit 18, shown in Figure 1. The weighting filter is a finite impulse response digital filter with, for example, 11 taps. The coefficients of the filter are such as to define a frequency-gain characteristic as shown in Figure 2, which is basically a low pass filter characteristic, but has important distinctions. As illustrated, low frequencies (below about 1 kHz) are subject to a positive gain which decays rapidly beyond 1 kHz. The purpose of this characteristic is to emphasise the relatively low frequency, periodic pulses of the voice signal which contain the most information and to diminish the significance of the higher frequency, intermediate parts of the signal much of which represents noise.

The blocks are each filtered separately. For an 11 tap filter to which the samples of successive blocks are fed continuously, the first five output samples and the last five output samples must be discarded. Therefore the number of output samples in the filtered block corresponds to the number of output samples in the input block.

The output samples from the filter for each block are then down-sampled by a decimation factor  $d$  in order to produce  $d$  decimated (interleaved) sequences. Typical values of  $d$  are 3 or 4 though higher values may be used for lower bit rate channels. The decimation factor is also partially determined by the block size, since each decimated sequence should be of equal length. Processing block 18 is also effective to select one of the decimated sequences of each block by comparing the total energy contents of the sequences and selecting the sequence having the maximum energy. The energy of a sequence is determined by summing the squares of each of its constituent samples. An index  $s$  identifies the selected sequence and this index  $s$  is also passed to the frame-forming circuit 44 on line 20.

The selected sequence is fed from processing block 18 to a vector quantizer 22 which is illustrated in more detail in Figure 4. The concept of vector quantization is not novel per se but the particular characteristics of the vector quantizer which will now be described are considered to be unique in the present combination.

The input on line 24 to the vector quantizer is a signal consisting of a series of the selected decimated sequences derived from successive blocks.

The input signal for the current block (i.e. the selected decimated sequence for that block), and the existing contents of the memory 27 of a pitch

synthesis filter 28 are fed to a control processor 26. As will be described hereinafter, memory 27 contains a data sequence derived from the selected sequences of the immediately preceding frame, and control processor 26 compares the sequences input thereto to obtain pitch indices  $\underline{p}$ ,  $\underline{h}$  representing the pitch period and the pitch gain respectively of the decimated sequence in the current block. The pitch period index  $\underline{p}$  represents the number of shifts (relative to a datum position) that have to be performed in order to reach the position of maximum correlation of the current decimated sequence input on line 24 and the sequence stored in memory 27, and this shift usually represents the time interval between neighbouring pitch pulses P (shown in the waveform W).

As many as five bits may be needed in order to adequately define the pitch period of the decimated sequence selected from the first block in each frame, this sequence being compared with the stored sequence at 32 (i.e.  $2^5$ ) different relative positions. The same number of bits could be used to define the pitch periods of the sequences selected from the remaining blocks in each frame. However, since the pitch period varies by only a small amount from block-to-block, fewer bits may be used to define the indices  $\underline{p}$  and  $\underline{h}$  for the remaining blocks.

In an example, the sequence selected from each remaining block in a frame could be correlated with the stored sequence at only eight (i.e.  $2^3$ ) different relative positions, distributed to either side of the pitch pulse already located by analysis of the sequence selected from the first block of the frame.

The pitch gain (represented by index  $\underline{h}$ ) is calculated as the ratio of the cross-correlation of the selected input sequence and the pitch filter memory (at the position of maximum correlation) normalised with respect to the block energy of the contents of the pitch filter memory.

The pitch indices  $\underline{p}, \underline{h}$  generated in this manner are output from the control processor 26 to the pitch synthesis filter 28 and also to the frame-forming circuit 44.

Having evaluated the pitch indices  $\underline{p}, \underline{h}$  for the current decimated sequence, this sequence is then subjected to vector quantization. This involves comparing the pattern of the current, selected sequence with the pattern of each of a number of reference sequences or vectors stored in a Gaussian codebook 38 in order to determine which of these reference sequences it most closely resembles, the selected reference sequence being represented by a unique index  $\underline{f}$ .

However, before vector quantization is carried out, the memory response of pitch synthesis filter 28 (which would detract from the effectiveness of

the matching procedure) is subtracted from the current sequence input on line 24. To that end, the contents of the memory 27 of the pitch synthesis filter 28 are transferred on line 25 to the memory 31 of an otherwise identical auxiliary pitch synthesis filter 29 which is used to compute the pitch synthesis filter memory response. The pitch synthesis filter 29, which is an infinite impulse response filter, is clocked with a zero input to find its memory response which is then output on line 33. This memory response is fed into a subtractor 35 together with the current input sequence on line 24, thereby to produce a difference or reference signal on line 37 at the output of the subtractor. This setting of the pitch synthesis filter 28 is carried out initially for each block to be processed. The zero input pitch filter response is subtracted from the input signal in order to reduce the mean-squared error during the subsequent matching operation which is designed to identify which one of a plurality of vectors stored in Gaussian codebook 38 most closely matches the input signal on line 37.

With its memory 27 now set to zero, the pitch synthesis filter 28 is fed with different sequences from the Gaussian codebook 38. These sequences, together with the pitch data, are used to generate output signals which are routed to a further subtractor 32 on line 30, the other input to subtractor 32 being the difference, or reference signal on line 37. The output from subtractor 32 is therefore a difference signal representing the mismatch between the two inputs to the subtractor. This mismatch or "error" for each successive vector input to the pitch synthesis filter 28 is computed by summing the squares of the sample values in an error computing processor 34. The error for each successive signal is fed back to the control processor 26 and the error processor also produces an output signal on line 36 to indicate that the error computation is complete for that input signal.

The number of different pattern sequences or vectors stored in Gaussian codebook 38 determines the accuracy of quantization. The purpose of the vector quantizer is to determine which of the vectors stored in the codebook most closely resembles the pattern (but not necessarily the magnitude) of the selected decimated sequence which is input to the vector quantizer. Once the closest vector from the Gaussian codebook 38 has been identified by the vector quantizer, the entire decimated sequence can be represented by the index  $\underline{f}$  of this vector, the analysed pitch characteristics  $\underline{h}, \underline{p}$  and a scale factor  $\underline{g}$ , the derivation of which is now described.

Each of the vectors in Gaussian codebook 38 is a random sequence which has a zero mean Gaussian energy distribution and a normalized energy content. Because of this, each signal output

from the Gaussian codebook is multiplied, in a multiplication circuit 40 by the optimal scale factor  $g$  which is computed by control processor 26 from the energy contents of the signals on lines 37 and 30. The optimal scale factor  $g$  is given by the cross-correlation of the signals on lines 37 and 30 divided by the energy of the signal on line 30. The signal on line 30 is first computed with  $g$  set at 1. An aim of the scale factor calculation is to reduce the scale factor towards zero if the input sequence mainly contains noise and there is no significant correlation between the signal on line 37 and the signal on line 30.

In general, the energy in the input sequence on line 37 fluctuates over a relatively wide range, and so as many as 5 or 6 bits may be needed to define scale factor  $g$ . However, the number of bits can be reduced significantly by further normalising the scale factor (before coding) with respect to the energy in pitch filter memory 28, and this leads to a further reduction in the bit rate of the speech coder.

When the best vector has been selected from Gaussian codebook 38 and its optimum scale factor  $g$  computed, the memory 27 of the pitch synthesis filter 28 is updated as follows in readiness for processing the next block. The original contents of memory 27, which were transferred to memory 31, are returned to memory 27 and the selected vector from the Gaussian codebook 38 is scaled by the optimum scale factor  $g$  and input to the pitch synthesis filter 28. When the process of clocking the filter finishes (after clocking the required number of times depending on the number of elements in each vector in the Gaussian codebook) the resultant contents of the memory 27 are retained for processing the next block.

The output from the vector quantizer 22 is fed on line 42 to the frame forming circuit 44 which also receives inputs from the quantizer 16 of the linear predictive filter 6 and from the decimation processor 18. The frame forming circuit assembles this input data into a predetermined standard format for transmission over the channel 46.

As illustrated diagrammatically in Figure 8, it will be appreciated that for each input frame of digital samples, there is data from the parameter quantizer 16 of the linear predictive filter to be transmitted, and for each block into which the frame is divided, a sequence index  $s$  and output data from the vector quantizer, that is pitch data  $h$ ,  $p$ , index  $f$  and scale factor  $g$  must be transmitted. Bits for synchronisation purposes with the decoder and for identifying successive frames may also need to be added.

The index  $f$ , which represents the selected vector from the Gaussian codebook 38, may consist of as many as 8 or 9 bits, depending on the

accuracy of the quantization.

In an alternative embodiment of the invention, the reference pattern is derived from the memory response of the auxiliary pitch synthesis filter 29, and the corresponding index may then be defined using fewer bits, leading to a reduced bit rate.

The reference pattern is derived from the memory response of pitch filter 29 by means of an alternative circuit, shown generally at 47, which is used in place of the Gaussian codebook 38. The reference pattern is generated by suitably clipping the memory response using a clipping circuit 48.

The decoder illustrated in Figure 3 essentially carries out the inverse of each of the operations carried out in the described encoder working on the basis of the data transmitted over the channel 46. This data is first fed to a frame decoder 48 which extracts the various items of data transmitted. The data generated by the vector quantizer 22 in the encoder are fed to an inverse vector quantizer 50 which contains a memory storing identical vectors to those stored in the Gaussian codebook 38. The index  $f$  generated by the encoder determines which of these stored sequences is read out and multiplied by the scale factor  $g$ . If circuit 47 is used in the encoder instead of the Gaussian codebook 38, then the memory of the inverse quantizer 50 would contain a corresponding vector derived from the memory response of the pitch filter. The pitch parameters  $h$  and  $p$  are used to control a pitch synthesis filter corresponding to filter 28 of the vector quantizer in the encoder, and this adds in the pitch pulse components. Therefore the output of the inverse vector quantizer 50 is a representation of the decimated sequence which was fed to the vector quantizer in the encoder. Zeros must be interpolated into this decimated sequence in order to produce an LPC excitation signal which corresponds to a representation of the LPC residual signal. The frame decoder supplies to an interpolation processor 52 the sequence index  $s$  so that  $d - 1$  zeros may be interpolated between successive samples of the sequence and an appropriate number of zeros interpolated at the beginning and end in order to place the samples of the decimated sequence in their correct positions in the excitation signal. The output of the interpolation processor 52 is then fed to an LPC synthesis filter 54. This filter receives control inputs from the frame decoder representing the quantized parameters and, in a known manner, uses the excitation signal to produce a representation of the original digital samples. Since the LPC parameters have been quantized, an inverse quantization must first be carried out (as occurred in circuit 15 of the encoder in Figure 1), before the LPC synthesis filter can operate to restore the original speech samples. These samples are then fed out via a

digital-to-analogue converter 56 to reproduce an analogue signal corresponding to the voice signal originally input on line 2.

In practice the LPC synthesis filter 54 in the decoder does not have an "ideal" memory response, and this tends to detract from the quality of the processed signals. In order to alleviate this problem, the memory response of filter 54 may be subtracted from the digital samples at the output of sampler 4 of the encoder, before these samples are fed to the linear predictive inverse filter 6. To that end, an additional improvement may be obtained if the LPC synthesis filter 54 is clocked from time-to-time (once per frame, say) with a zero input, and the zero input memory response of the filter is passed to the memory 80 of an identical LPC synthesis filter 81, the output of which is connected to one input of a subtraction circuit 82 which interconnects the speech sampler 4 and the coder 6 of the encoder (Figure 1). Filter 81 is also clocked with a zero input and by this means the zero input memory response of filter 81, which is the same as that of filter 54 in the decoder, is subtracted by subtraction circuit 82 from the input digital samples produced by the speech sampler 4.

The encoder shown in Figure 5 is similar to that illustrated in Figure 1. However, in this embodiment an additional feedback loop comprising a decoder and interpolation processor 60 and a pitch filter 62 is included. The output of the pitch filter 62 is fed to a subtractor 64 connected to receive the LPC residual signal from the inverse 14 filter of the linear predictor 6. The decoder and interpolation processor generates from the output frame for transmission a representation of the excitation signal which in the decoder proper would be fed to the LPC synthesis filter. However, in this feedback loop it is fed to the pitch filter 62 which removes from it the pitch pulses so that the output of subtractor 64 which is fed to the weighting filter and decimation processor 18 contains a less significant contribution from the pitch pulses. The parameters of the pitch filter 62, that is the pitch gain  $\underline{q}$  and pitch period  $\underline{r}$  are determined by analysing the LPC residual signal output from filter 14 in a processor not shown. The contents of a memory of the pitch filter may also be used in the analysis. This pitch data must also be transmitted over the channel 46 and is therefore also supplied to the frame former 44. The pitch data  $\underline{q}$ ,  $\underline{r}$  may be updated for each block, each frame or even less frequently depending on the bit rate constraints. However, because there is less information now present in the signal which is fed to the vector quantizer, the number of bits required for transmitting its output data may be reduced. In particular, the size of the Gaussian codebook may be restricted.

The decoder for use in conjunction with the encoder of Figure 5 is illustrated diagrammatically in Figure 6. It is essentially identical to the decoder described with reference to Figure 3 except that the output from the interpolator 52 is not fed directly to the LPC synthesis filter but is first fed through a pitch synthesis filter 68 which is controlled by the pitch gain  $\underline{q}$  and the pitch period  $\underline{r}$  transmitted over the channel 46. This pitch synthesis filter restores to the excitation signal a series of pitch pulses corresponding to those identified in the originally encoded LPC residual signal. In this way the interpolated zeros are to some extent overwritten by contributions from the pitch synthesis filter. This results in much smoother quality, because the high frequency distortion created by the spectral folding of the interpolation is largely eliminated. The effectiveness of this pitch filter feedback loop in the encoder is dependent upon the presence of the particular frequency-gain characteristic in the weighting filter of the encoder as described previously with reference to Figure 1. It will be noted that separate sets of pitch filter parameters together with their filter memories are calculated in this embodiment both directly from the LPC residual signal, which is outside the baseband, and from the baseband output of the weighting filter and decimation processor 18. This results in much better prediction performance than an ordinary pitch filter and higher levels of stability at all times.

In the foregoing, the method of quantizing the LPC parameters in the quantizer 16 has not been discussed in detail. It is possible to use any scalar quantizer or vector quantizer already proposed for this purpose but the novel quantizer illustrated in Figure 7 is found to be particularly effective for low bit rate applications.

The input on line 68 to this quantizer is the parameters after having been transformed in a processor (not shown) into the line spectral pairs domain using transforms as described in the literature. A transformation into some other domain which will enable correlation or matching of similarities between the transformed LPC parameter vectors and vectors stored in a codebook may also be used. The input is a sequence of  $\underline{k}$  values, where  $\underline{k}$  is the order of the linear predictive filter, for example  $\underline{k} = 10$  is a typical value. A new input sequence is generated for each frame of input digital samples. Each input sequence is fed to a control processor 70 and to a first codebook 72. The control processor 70 analyses an error signal on line 74 produced after matching in the first codebook and with every sequence or vector of a second codebook 76 to compute a scale factor  $\underline{w}$ . This scale factor  $\underline{w}$  performs the same function as the scale factor  $\underline{g}$  used in the previously described

vector quantizer. The codebook 72 is a first in first out (FIFO) memory store. The vectors generated for storage in this codebook are derived from previously received sequences as will be described in more detail later.

The input sequence is matched with each of the stored vectors in turn and, by a process of least squares minimisation conducted by the control processor, the index  $t$  of the vector which most closely matches the input sequence is generated. The vector  $V_t$  so identified is output from the codebook 72 to a subtractor 74 which also receives the actual input sequence. The output from the subtractor 74 is therefore an error signal representing the difference between the input sequence and the most closely matched vector of the codebook 72. This error signal is fed to the control processor and to a second Gaussian codebook 76. The vectors stored in the codebook 76 are each normalised random sequences with a Gaussian distribution and zero mean. This codebook 76 is therefore of the same type as the codebook 38 used in the previously described vector quantizer of Figure 4. Under the control of the control processor 70, the index  $u$  of the vector in the codebook 76 which most closely matches the input error signal is determined by a least squares minimisation technique. The selected vector  $V_u$  is then output to a multiplier 78 where it is multiplied with the optimal scale factor  $w$  producing a sequence  $wV_u$  which is added in an adder 79 to the vector  $V_t$  selected from the first codebook 72. The output of the adder 79 is then placed into the codebook 72 displacing the oldest previously stored vector. The contents of the codebook 72 are therefore being continuously updated so that the effectiveness of this codebook continuously increases while the input voice signal has characteristics of the same speaker. With this method of parameter quantization, the outputs for transmission over the channel 46 link are the two indices  $t$ ,  $u$  and the optimal scale factor  $w$ .

With normal scaler quantization of the parameters of a tenth order linear predictive coder, 40 bits per frame would normally be required whereas with this method of vector quantization, equivalent or improved speech quality can be produced with 20 bits per frame with 13 bits allocated to the indices  $t(5)$  and  $u(8)$  and 7 bits allocated to the scale factor  $w$ . These values are given as a typical example only and may be varied in dependence upon other constraints in the system.

## Claims

1. A speech coder for encoding an input speech signal for transmission over a digital channel of a telecommunications link, comprising

means for sampling the input speech signal to produce output digital samples,  
means for dividing these digital samples into frames each consisting of a predetermined number of samples,

5 a linear predictive filter for inverse filtering each frame and producing an output LPC residual signal for said frame comprising a said predetermined number of digital samples, and LPC parameters for said frame, and

10 baseband extraction means, the speech coder being characterised by provision of down-sampling means for extracting from the output of said baseband extraction means  $d$  interleaved sequences,

15 means for selecting one of said sequences which contains the maximum energy content and producing an output index representing the selected sequence,

20 means for deriving pitch period and pitch gain indices from the selected sequence,

means for removing long term correlation from the selected sequence to produce a remainder sequence,

25 means for comparing the remainder sequence with an identifiable reference sequence, and for deriving a scale factor from the compared sequences, the scale factor being defined by a scale factor index and being representative of the energy in the remainder sequence relative to that in the reference sequence,

30 wherein for each frame of the input, an output frame comprising data representing the LPC parameters, and for each block, the index of the selected sequence, and the scale factor, pitch period and pitch gain indices and an index representing said identifiable reference sequence is transmitted over the channel of the telecommunications link.

40 2. A speech coder as claimed in claim 1, wherein the baseband extraction means comprises a weighting filter which amplifies a low frequency pitch component of the LPC residual signal and reduces the amplitude of higher frequency components of the LPC residual signal and said samples of the LPC residual signals are divided into blocks before being passed through the filter.

45 3. A speech coder as claimed in claim 1, wherein the baseband extraction means effects multipulse linear predictive analysis-by-synthesis whereby said baseband is obtained by minimizing an error between input speech signals and artificially reconstructed speech signals.

50 4. A speech coder as claimed in any one of claims 1 to 3, wherein the means for comparing includes a vector quantizer for matching the remainder sequence with the most closely resembling one of a plurality of vectors stored in a

codebook, each stored vector being a random sequence having a Gaussian distribution and being identifiable by a unique index, and the output frame includes data representing the index of the selected vector.

5. A speech coder as claimed in any one of claims 1 to 3, wherein the means for comparing includes a pitch synthesis filter, and the reference sequence is derived from the memory response of the pitch synthesis filter.

6. A speech coder as claimed in any one of claims 1 to 5, further comprising means for processing the LPC residual signal in order to extract pitch data relating to the period and amplitude of the pitch pulses,

a pitch filter for receiving said pitch data, said pitch filter having an input and an output and being operative to remove from any input signal, pitch pulses characterised by the pitch data,

a subtractor connected to receive the LPC residual signal and an output from said pitch filter and produce an output representing the difference signal for extraction of a baseband signal by said baseband extraction means,

a decoder for deriving from the output of the encoder a decoded LPC excitation signal and applying said signal to said input of the pitch filter, the pitch data being transmitted over the channel.

7. A speech coder for decoding speech encoded with the coder according to any one of claims 1 to 6, comprising means for separating from a received frame the scale factor index, and the data representing the LPC parameters,

means for outputting to a pitch synthesis filter a sequence corresponding to said identifiable sequence scaled by said scale factor,

an interpolator for receiving said output sequence and said selected sequence index and interpolating zeros at appropriate positions in order to produce an LPC excitation signal, and

an LPC synthesis filter for receiving said excitation signal and said data representing the LPC parameters and for restoring therefrom a sequence of digital samples representing the input speech signal.

8. A speech coder as claimed in claim 5 for decoding speech encoded with the encoder according to claim 4, wherein the outputting means comprises an inverse vector quantiser including a codebook corresponding to the codebook in the vector quantizer of the encoder, and the inverse vector quantizer receives said unique index and the scale factor and outputs in response thereto, as the output sequence, a corresponding sequence scaled by said scale factor.

9. A coder as claimed in claim 7 or claim 8 for use with an encoder as claimed in claim 6, wherein said separating means extracts from a received

frame the pitch data and further comprises a pitch synthesis filter connected between the output of said interpolator and the input to the LPC synthesis filter, said pitch synthesis filter receiving said pitch data and restoring into said output of the interpolator pitch pulses having the period and amplitude represented by said pitch data.

10. A speech encoder as claimed in any one of claims 7 to 9 including means for subtracting the memory response of the LPC synthesis filter in the decoder from the digital samples produced at the output of the sampling means of the encoder.

11. A speech coder comprising an encoder as claimed in any one of claims 1 to 6, and further comprising a vector quantizer for quantizing the LPC parameters prior to transmission over the channel.

12. A baseband code-excited linear predictive coder in which the LPC residual signal is divided into blocks which are separately passed through a baseband extraction means, down-sampled and vector quantized.

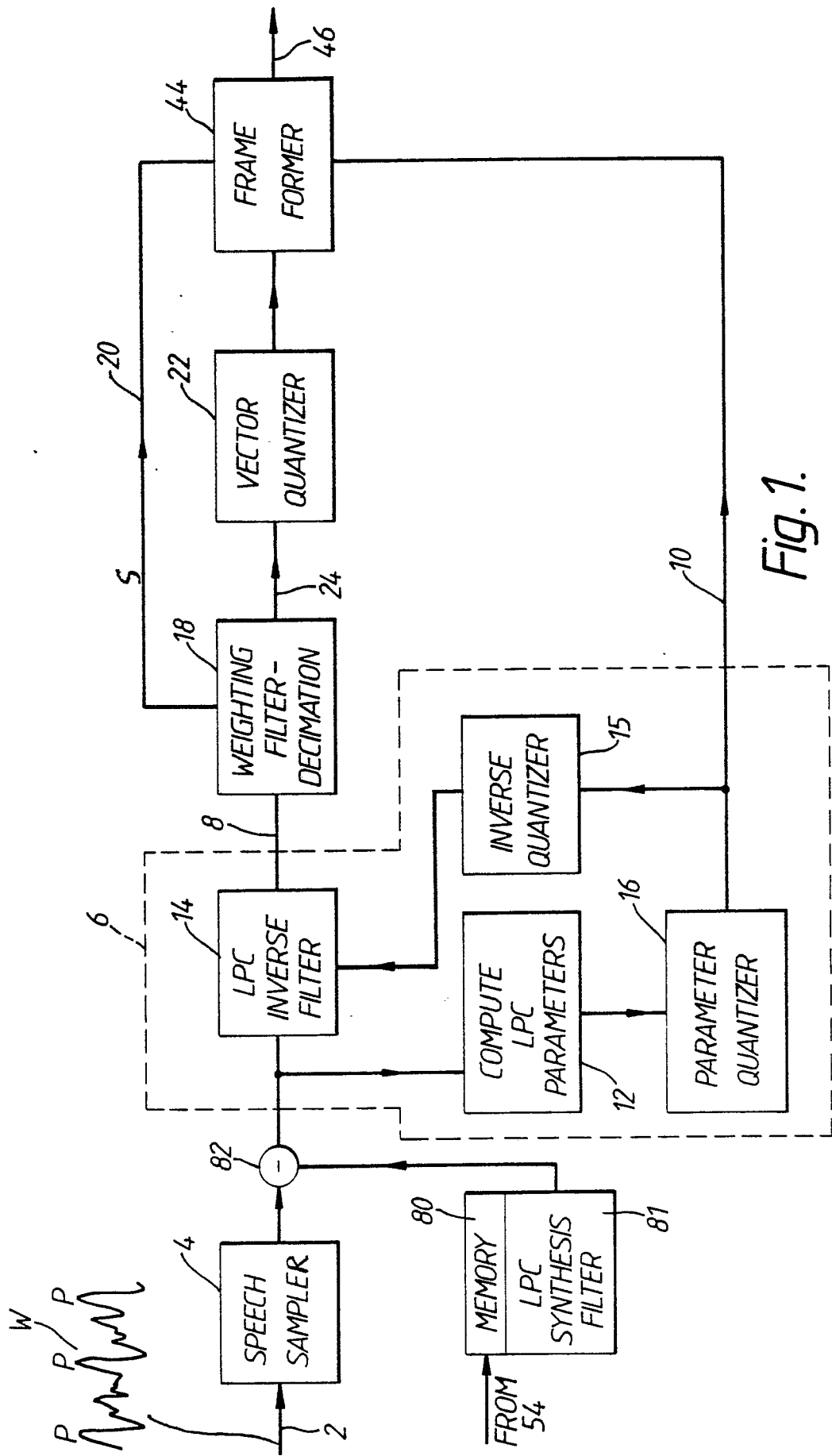


Fig. 1.

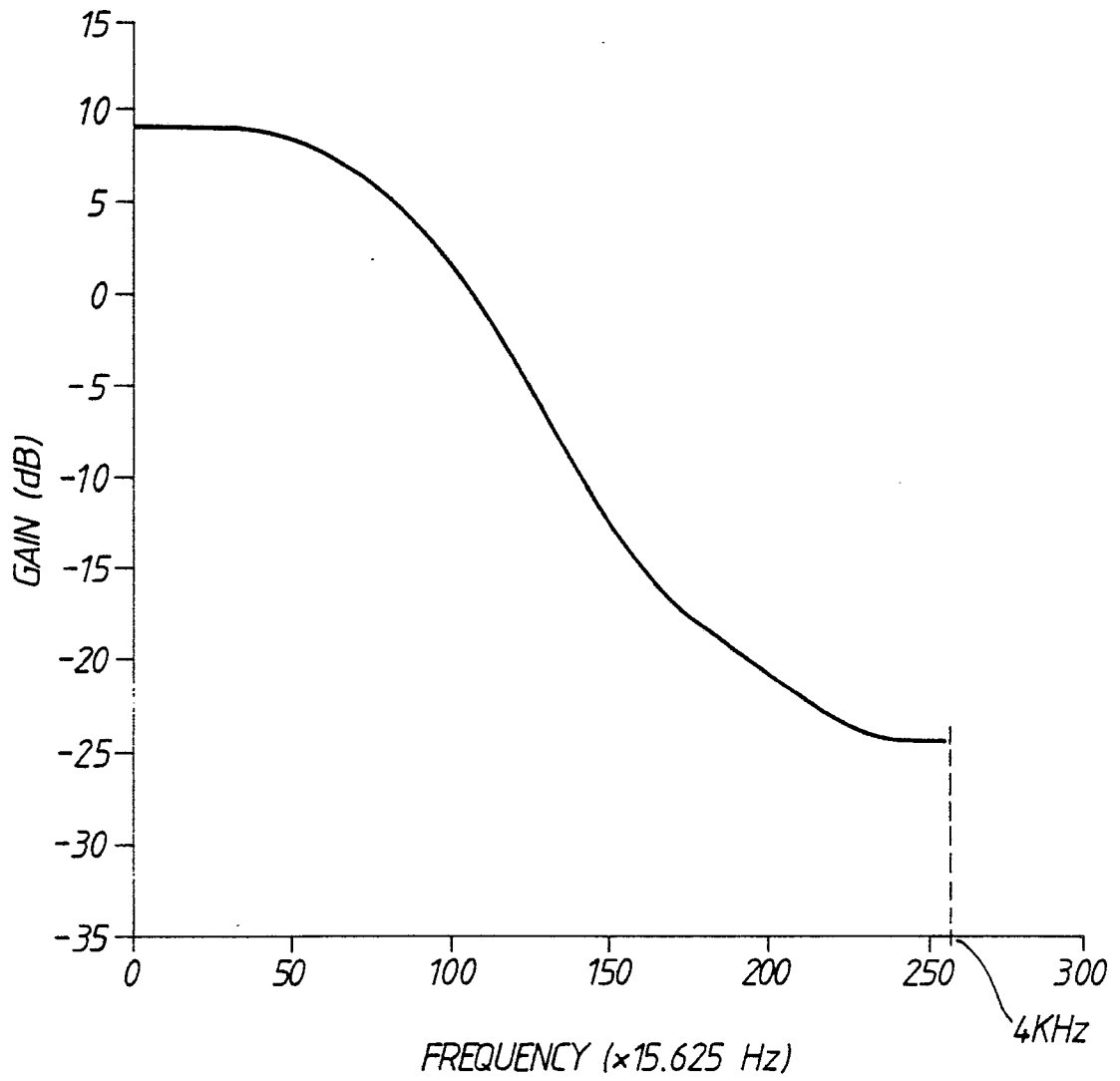


Fig.2



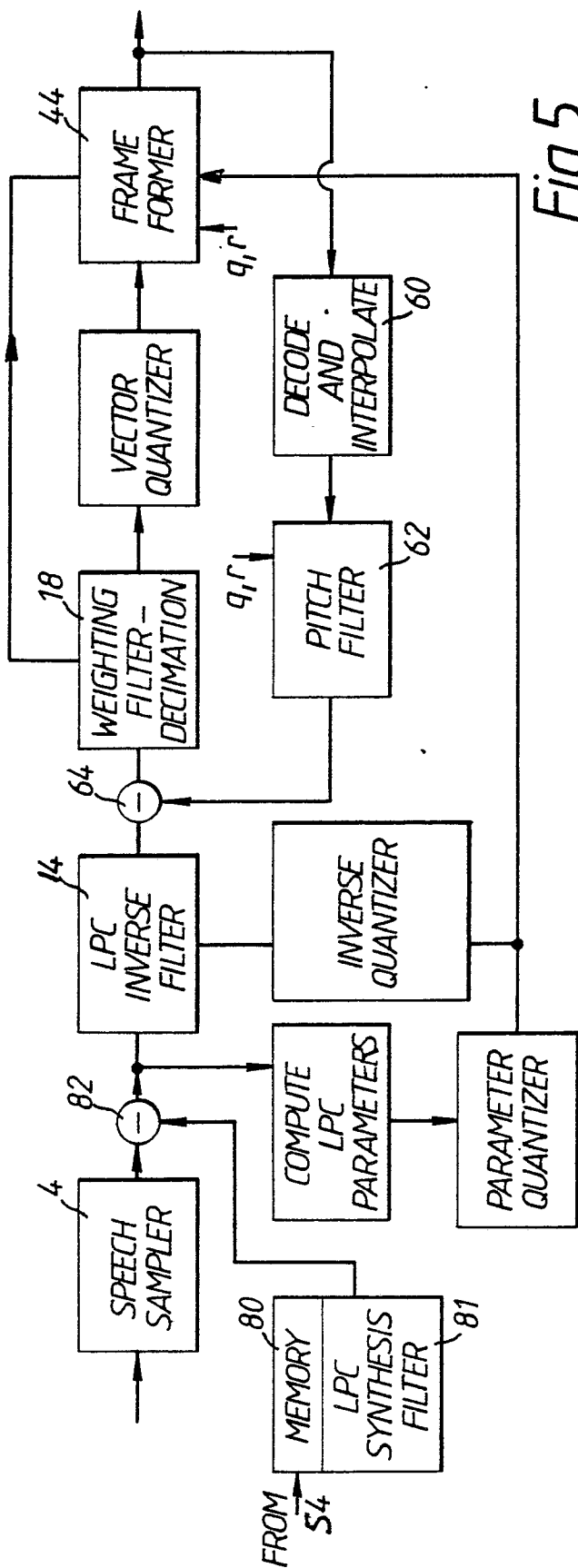


Fig. 5.

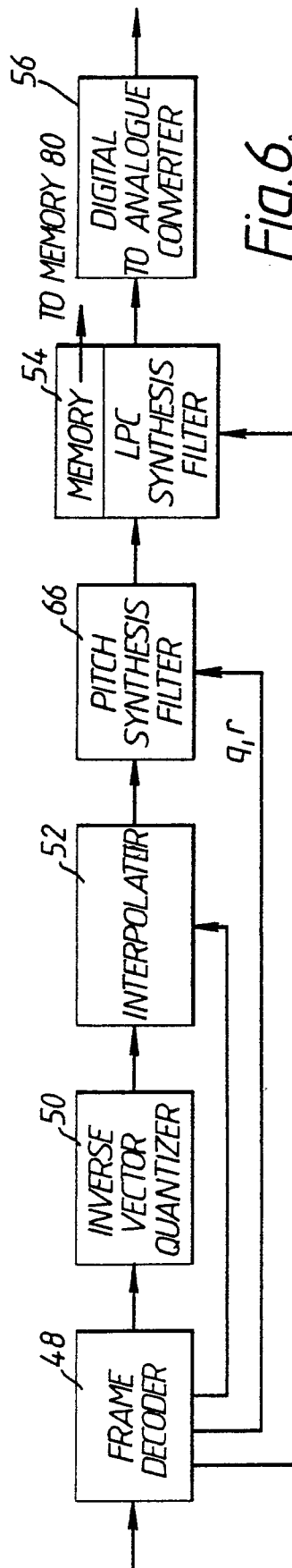


Fig. 6.

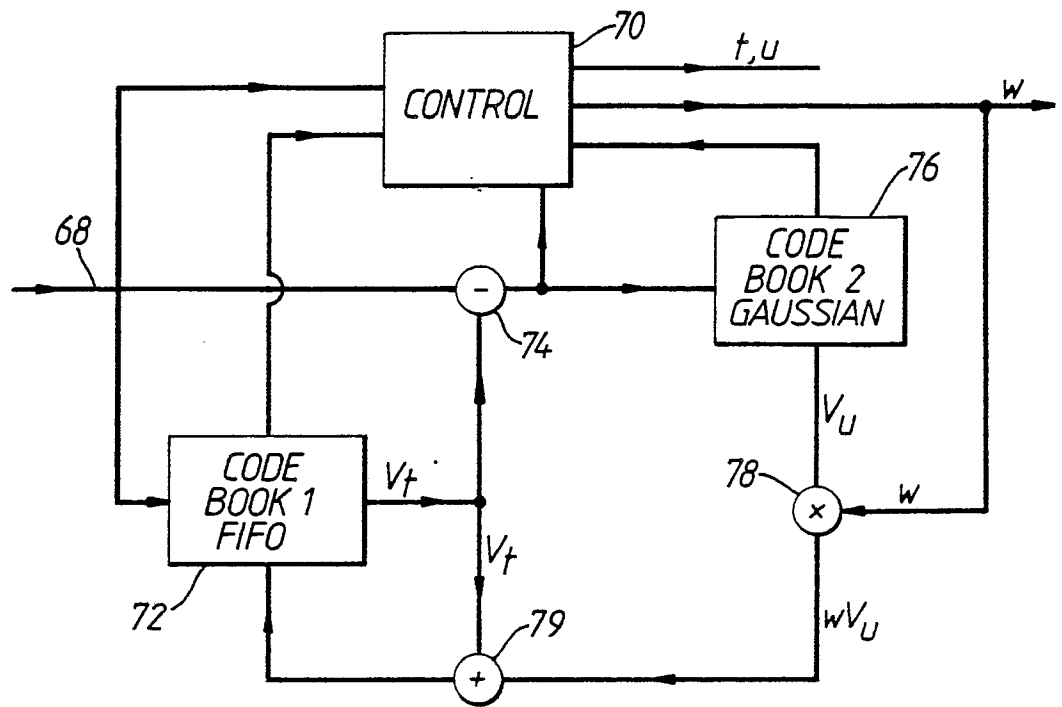


Fig. 7.

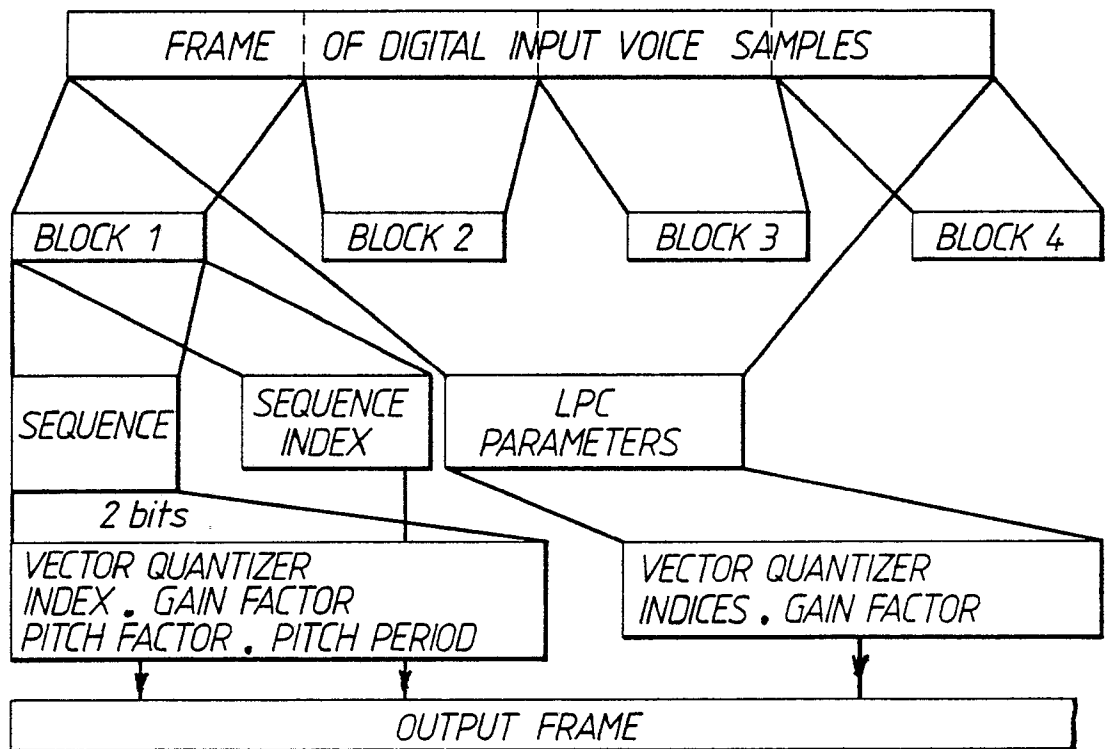


Fig. 8.