SPEECH PROCESSING APPARATUS

The present application is concerned with methods of, and apparatus for, the processing of speech signals, and finds particular application in coding or decoding or speech signals.

A common technique for speech coding is the so-called linear predictive coding (LPC) in which, at a coder, an input speech signal is divided into time intervals and each interval is analysed to determine the parameters of a synthesis filter whose response is representative of the frequency spectrum of the signal during that interval. The parameters are transmitted to a decoder where they periodically update the synthesis filter which, when fed with a suitable excitation signal, produces a synthetic speech output which approximates the original input.

Clearly the coder has also to transmit to the decoder information as to the nature of the excitation which is to be employed. A number of options have been proposed for achieving this, falling into two main categories, viz:

(i) Residual excited linear predictive coding (RELP) where the input signal is passed through a filter which is the inverse of the synthesis filter to produce a residual signal which can be quantised and sent (possibly after filtering) to be used as the excitation, or may be analysed, e.g. to obtain voicing and pitch parameters, for transmission to an excitation generator in the decoder.

(ii) Analysis by synthesis methods in which an excitation is derived such that, when passed through the synthesis filter, the difference between the output obtained and the input speech is minimised. In this category there are two distinct approaches: one is multipulse excitation (MP-LPC) in which a time frame corresponding to a number of speech samples contains a, somewhat smaller, limited number of excitation pulses whose amplitudes and positions are coded. The other approach is stochastic coding or code excited linear prediction (CELP). The coder and decoder each have a stored list of standard frames of excitations. For each frame of speech, that one

of the codebook entries which, when passed through the synthesis filter, produces synthetic speech closest to the actual speech is identified and a codeword assigned to it is sent to the decoder which can then retrieve the same entry from its stored list. Such codebooks may be compiled using random sequence generation; however another variant is the so-called "sparse vector" codebook in which a frame contains only a small number of pulses (e.g. 4 or 5 pulses out of 32 possible positions with a frame). A CELP coder maya for instance have a 1024-entry codebook.

The present invention can be applied in CELP techniques but may also be useful in other coding techniques for instance multipulse (MP) or regular pulse excitation (RPE) linear predictive coding.

Figure 1 shows a CELP decoder as described in our International patent application No. GB91/02291 (referred to below as "our earlier application") to illustrate the manner in which coded signals may be used upon receipt to synthesise a speech signal.

The basic structure involves the generation of an excitation signal, which is then filtered. The filter parameters are changed once every 20ms, a 20ms period of the excitation signal being referred to as a block. However, each block is assembled from shorter segments ("sub-blocks") of duration 5ms.

Every 5ms the decoder receives a codebook entry code k, and two gain values g_1, G_2 (though only one, or more than two, gain values may be used if desired). It has a stochastic codebook store 100 containing a number (for instance 128) of entries, each of which defines a 5ms period of excitation at a sampling rate of 8kHz. The excitation is a ternary signal (i.e. may take value +1, 0 or -1 at each 125μs sampling instant) and each entry contains 40 elements of three bits each, two of which define the amplitude value and the remaining one of which defines gain, as mentioned below.

If a sparse codebook (i.e. where each entry has a relatively small number of elements) is used a more compressed representation might be used.

The code k from an input register 101 is applied as an address to the store 100 to read out an entry into a 3-bit wide parallel-in-serial-out register 102. The output of this register (at 8k/samples per second) is then multiplied by one or other of the gains G_1, G_2 from a further input register 103 by multipliers 104, 105. Which gain is used for a given sample is determined by the third bit of the relevant stored element, as illustrated schematically by a changeover switch 106.

The filtering is performed in two stages, firstly by a long term predictor indicated generally by reference numeral 107, and then by an LPC filter 108. The LPC filter, of conventional construction, is updated at 20ms intervals with coefficients a_i from a third input register 109.

The long term predictor 107 is a "single tap" predictor having a variable delay (delay line 110) controlled by signals d from a fourth input register 111 and variable feedback gain (multiplier 112) controlled by a gain value G_3 from the fourth register 111. An adder 113 then forms the LPC filter input by summing the gain-multiplied codebook entry from the switch 106 and the delayed scaled signal from the multiplier 112.

Although referred to as "single tap", the delay line actually has two outputs one sample period delay apart, with a linear interpolator 114 to form (when required) the average of the two values, thereby providing an effective delay resolution of ½ the sample period.

The parameters k, G_1, G_2, d, G_3 and a_i are derived from a multiplexed input signal by means of a demultiplexer 115; however the gains G_1, G_2, and G_3 are identified by a single codeword G, which is used to look up a gain combination from a gain codebook store 116 containing 128 such entries.

The task of a coder for use with the above decoder is to generate, from input speech, the parameters referred to above. Suitable coders are known and further description is not therefore given herein. For instance, reference may be made to our earlier application, in which a suitable coder is described.

It has now been found that it is possible to improve the manner in which delay and gain information is conveyed to the long term predictor.

According to a first aspect of the present invention, there is provided a speech coder comprising means for receiving input speech signals and determining:

(a) the parameters of a synthesis filter having a frequency response resembling the frequency spectrum of the input speech signals;

(b) information defining a second filter whose output is the sum of a delayed version of its output multiplied by a gain factor and its input; and

(c) information defining an excitation signal which at a decoder may be filtered by the filters to produce a speech signal resembling the input speech signal;

wherein the information defining the second filter defines a delay value and one of only two gain values one of which is substantially zero.

In another aspect the invention provides a speech decoding apparatus comprising:

(a) an excitation signal generator;

(b) a recursive filter having controllable feedback delay to filter the excitation signal; and

(c) a synthesis filter;

the generator and filters being controllable by received parameters;

wherein the recursive filter (b) is responsive to the parameters to set the feedback gain to one of only two values, one of which is substantially zero.

In an example, one or two selected delay words from all possible delay words might be used to set the gain to zero, other delay words leaving the gain fixed at an alternative value. Effectively then, the delay word has at least a dual function, that is to set the delay in the long term predictor, and, by setting the gain to zero, to switch off the predictor under certain circumstances. Except when switched off by a delay word, the predictor then has a fixed gain.

According to a further aspect of the present invention, there is provided speech decoding apparatus, for use with or in coding apparatus in which input speech is analysed to determine parameters of a decoding synthesis filter and to select at least one excitation component for the filter from a plurality of possible components, said decoding apparatus comprising a long term predictor incorporating gain and delay controls, the gain control setting the gain in the predictor to have a selected one of only two values one of which is zero.

Preferably, these two values comprise zero and slightly less than one, such as 0.95. Although it is more usual to control gain to lie somewhere between -1 and 4 in the long term predictor, by keeping gain to less than 1 errors tend to die away, and are lost when the gain value is zero.

The delay word or words setting the gain in the long term predictor to zero might for instance be words otherwise setting an extremely low delay, such as, in a

delay word having seven bits, the words wherein the least significant bit is zero or one, all other bits being zero.

Some embodiment of the present invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 2 shows a decoder according to an embodiment of the present invention;

Figure 3 shows a flow diagram for use with a long term predictor according to an embodiment of the present invention embodied at least partially in software; and

Figure 4 is a block diagram of a speech coder according to an embodiment of the present invention.

It should be noted that the same reference numerals are used on each of the figures to indicate components or assemblies which perform the same or equivalent functions.

Referring to Figure 2, the arrangement of a decoder according to an embodiment of the invention can be treated as equivalent to the decoder described with reference to Figure 1 in that the parameters of the decoder are controlled by incoming information from a coder, via a demultiplexer. In this case, in contrast to the decoder of Figure 1, there is a single control input 212 to the long term predictor 107, carrying seven-bit delay words and no separate gain information. All but two of the delay words available control the delay set by the delay element 110, for instance to be equal to the value of the most significant six bits of the delay word plus a minimum setting of 20ms, if a minimum delay is to be set. (A typical delay range might be 20 to 80ms).

However, two of the delay words available, these being 0000000 and 0000001, act on the gain of the long term predictor, determined at the multiplier 112, to set the gain to zero. Effectively, this switches off the long term predictor. For all other delay words, the gain in the long term predictor is set to 0.95.

The delay mechanism 112 by which delay is implemented in the long term predictor 107 is that described above in relation to Figure 1: a delay line with two outputs which are one sample period delay apart. Again there is a linear interpolator 114 which forms, when required, the average of two taps with a 50-50 weighting, effectively providing a delay resolution of ½ the sample period. This type of delay mechanism is described in our earlier application referenced above. It might be noted that the least significant bit in the 7-bit delay word is used not only to carry the gain information for the long term predictor but also to control use of this "half-tap" delay. That is, seven bits are already used for other reasons in delay words in this context. Therefore, in embodiments of the present invention, a 7-bit delay word can be used to carry the gain information without adding extra bits to the delay words which have to be transmitted form the coder.

The long term predictor 107 might simply comprise a delay line 110 and a multiplier 112, both controlled by a common delay word input "d" to feed back a received signal via an adder 113. The multiplier 112 however is controlled by the delay word input d by means of a comparator 300 and a switch 301. The comparator 300 will check an incoming delay word against for instance stored values and, in the case that the delay word matches the stored values, bring the switch 301 to the zero gain position, thereby switching off the long term predictor 107. On receipt of a delay word input d having at least one "1" amongst the six most significant bits, the comparator 300 will bring the switch 301 back to the positive gain position, reactivating the long term predictor 107.

As discussed above, because gain control is less flexible in the long term predictor 107 in embodiments of the present invention, it is preferable to compensate by increasing the emphasis on the fixed codebook 100, and thus to increase its size. Typically, in embodiments of the

present invention, the fixed codebook 100 might be a 1024 entry codebook (10 bits) where previously a 32 entry codebook (5 bits) might have been used.

In the above description, the elements of the speech coding/decoding equipment are described substantially in hardware terms. It will be clear to anyone skilled in this technical area that it may be preferable in practice to use a software equivalent approach, for instance based on a suitably programmed digital signal processing device. Referring to Figure 3, a simple flowchart for controlling the gain and delay in a long term predictor 107, based on a 7-bit delay word "d", may comprise relatively few steps. As shown, these may be for instance a first decision point 400 at which a check is made that the gain indicated by an incoming delay word "d" is not to be set to zero. If it is to be set to zero, there is no benefit is assessing the delay information since the long term predictor will be switched off. If it is not to be set to zero but left at whatever fixed value is to be used, such as 0.95, the delay word d goes forward to a further decision point 401 to test for the presence of a half-tap condition, and to the delay line 110 to determine the main part of the delay in accordance with the six most significant bits.

Clearly, a modified decoder is needed to produce signals that may be decoded by the decoder described above. A suitable coder is shown in Figure 4.

The input speech is analysed by an LPC analysis unit 200 to derive the coefficients a_i of an LPC filter (impulse response H) having a spectral response similar to that of each 20ms block of input speech. Such analysis is conventional and will not be described further.

The remainder of the processing is performed on a sub-block by sub-block basis.

The input speech sub-block and the LPC coefficients for that sub-block are then processed to evaluate the other parameters. First, however, because the decoder LPC

filter, due to the length of its impulse responses, will produce for a given sub-block an output in the absence on any input to the filter. This output - the filter memory $\underline{M}$ - is subtracted from the input speech in a subtractor 202 to produce a target speech signal $\underline{y}$. Note that this adjustment does not include any memory contribution from the long term predictor as its new delay is not yet known.

Secondly, this target signal $\underline{y}$ and the LPC coefficients a_i are used in a first analysis unit 203 to find that LTP delay d which produces in a local decoder with optimal LTP gain G_3 and zero excitation a speech signal with minimum difference from the target.

Thirdly, the target signal, coefficients a_i and delay d are used by a second analysis unit 204 to select an entry form a codebook store 205 having the same contents as the decoder store 100.

Finally, the gains g_1, g_2 are jointly selected by a gain analysis unit 206 and jointly quantised by reference to a gain codebook 223 to minimise the difference between a local decoder output and the speech input.

The operation of the parts may be as described in our earlier patent application, where a detailed description of the functioning of the analysis units may be found.

The arrangement therefore described differs from the earlier one in that the long term predictor gain G_3 is not included in the analysis performed by the gain analysis unit 206. Instead, the "optimum" gain G_3 is quantised by a quantizer 210 to produce a single-bit output which is 0 if the gain is below a threshold value (typically 0.6) and 1 otherwise. The delay value d passes via a gate arrangement 211 in which all but its least significant bit are forced to zero whenever the gain bit from the quantizer 210 is zero: otherwise they pass unchanged.

In the described embodiments because gain control in the long term predictor is simplified, more incoming bits to the decoder can for a given bit rate be supplied to the

codebook because fewer are required for the long term predictor gain. This increased emphasis on the codebook is preferable for robustness, because (a) entries in a gain codebook store (116) for a long term predictor (or an equivalent thereof) are stored in a memory which extra step (relative to entries in the codebook 100 which do not go into an equivalent memory) can introduce errors, (b) errors are perpetuated by the feedback in the long term predictor, being present for only single frames if generated in relation to the stochastic codebook, and (c) control over the gain level in the long term predictor may simply be exercised by a delay word, already used to control the delay in the long term predictor.

* C L M S *

CLAIMS

1. A speech coder comprising means for receiving input speech signals and determining:

(a) the parameters of a synthesis filter having a frequency response resembling the frequency spectrum of the input speech signals;

(b) information defining a second filter whose output is the sum of a delayed version of its output multiplied by a gain factor and its input; and

(c) information defining an excitation signal which at a decoder may be filtered by the filters to produce a speech signal resembling the input speech signal;

wherein the information defining the second filter defines a delay value and one of only two gain values one of which is substantially zero.

2. A speech coder according to Claim 1 wherein the information defining the second filter comprises a digital word capable of taking one of a plurality of values, one or more predetermined ones of the values representing a gain value of zero and the remaining values representing a delay value.

3. A speech coder according to Claim 2 including means for determining a second filter delay value and a single bit gain value and means for forcing the delay value to a predetermined value (or values) wherever the gain value bit has a value representing zero gain.

4. A speech coder according to Claim 3 in which the determining means comprises means for determining a delay value and a multilevel gain value and quantising means for quantising the gain value to a single bit.

5. A speech decoding apparatus comprising:

* C L M S *

(a) an excitation signal generator;

(b) a recursive filter having controllable feedback delay to filter the excitation signal; and

(c) a synthesis filter;

the generator and filters being controllable by received parameters;

wherein the recursive filter (b) is responsive to the parameters to set the feedback gain to one of only two values, one of which is substantially zero.

6. A decoder according to Claim 5 in which the received parameters include a delay/gain control signal and the recursive filter (b) includes recognition means operable in the event that the delay/gain control signal assumes a, or one of a plurality of, predetermined value(s) to set the feedback gain to zero and permit other values of the signal to control the delay of the filter.

7. Speech decoding apparatus, for use with or in coding apparatus in which input speech is analysed to determine parameters of a decoding synthesis filter and to select at least one excitation component for the filter from a plurality of possible components, said decoding apparatus comprising a long term predictor incorporating gain and delay controls, the gain control setting the gain in the predictor to have a selected one of only two values one of which is zero.

8. Speech decoding apparatus according to Claim 7 wherein the other of said two values is close to but less than one.

9. Speech decoding apparatus according to Claim 7 or 8 wherein said gain and delay controls are both controlled by a common incoming data word or series of words.

* C L M S *

10
12. Speech decoding apparatus according to Claim 9, wherein the incoming data word or words are in binary code and only data words in which all bits except the least significant bit are zero, said significant bit being either one or zero, operate on the gain control to set the gain in the long term predictor to zero.