Description

A common technique for speech coding is the so-called LPC coding 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 parameters of a 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. A modified form of Multipulse excitation is described in European patent application published as EP-0195487A, where the excitation pulses are constrained to lie on one of a plurality of regular grids of different phase; the grid position may be obtained by aligning it with the first pulse to be determined. 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 may typically have a 1024-entry codebook.

CELP coders are described in Proceedings of the ICASSP 87, International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, 6th - 9th April 1987, vol. 3, pages 1354-1357, IEEE, New York, US; D. LIN: "Speech coding using efficient pseudo-stochastic block codes" and also Proceedings of the ICASSP 86, International Conference on Acoustics Speech and Signal Processing, Tokyo, 7th - 11th April 1986, Vol. 1, pages 469-472, IEEE, New York, US; L.A. Hernandez-Gomez et al. "On the behaviour of reduced complexity code-excited linear prediction" (CELP). Lin describes a tree-structured search for choosing the desired codebook entry. The Hernandez-Gomez et al proposal involves searching the code-book using a first criterion to obtain a subset of the entries and then searching using a second criterion to identify the wanted entry within the subset.

According to the present invention there is provided a speech coder comprising:

means arranged in operation to generate, from successive time frame periods of input speech signals, filter information defining successive representations of a synthesis filter response, and to output the filter information;

means arranged in operation, for each of successive time frame periods of the speech, to receive the input speech signals and the respective filter information, and to generate excitation information comprising:

(a) a data store defining a plurality of excitation frames each consisting of a plurality of pulses

(b) means for selecting one excitation frame out of the said plurality of frames and rotationally shifted versions of the frames and for generating data identifying the store entry and the amount if any of its rotational shift;

in which the selecting means is arranged to:

(i) determine, out of a plurality of single-pulse frames each consisting of a single pulse at a different location, which frame meets the criterion that it would when applied to the input of a filter having the response defined by the filter information produce a frame of synthetic speech which most closely resembles the frame of input speech; and

(ii) determine which of the said plurality of stored frames, when rotationally shifted by an amount derived from the determined pulse location, meets the said criterion.

Other, optional, features of the invention are defined in the sub-claims appended.

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

- Figure 1 illustrates the rotational pulse shifting used in the invention;
- Figure 2 is a block diagram of one form of speech coder according to the invention; and
- Figure 3 is a block diagram of a suitable decoder.

It will be appreciated from the introduction that multipulse coders and sparse vectors CELP coders have in common the features that the excitation employed is in both cases a frame containing a number of pulses significantly smaller than the number of allowable positions within the frame.

The coder now to be described is similar to CELP in that it employs a sparse vector codebook which is, however much smaller than that conventionally used; perhaps 32 or 64 entries. Each entry represents one excitation from which can be derived other members of a set of excitations which differ from the one excitation - and from each other - only by a cyclic shift. Three such members of the set are shown in figures 1a, 1b and 1c for a 32 position frame with five pulses, where it is seen that 1b can be formed from 1a by cyclically shifting the entry to the left, and likewise 1c from 1a. The amount of shift is indicated in the figure by a double-headed arrow. Cyclic shifting means that pulses shifted out of the left-hand end wrap around and reenter from the the right. The entry representing the set is stored with the largest pulse in position 1, i.e. as shown in figure 1d. The magnitude of the largest pulse need not be stored if the others are normalised by it.

If the number of codebook entries is 32, then the excitation selected can be represented by a 5-bit codeword identifying the entry and a further 5 bits giving the number of shifts from the stored position (if all 32 possible shifts are allowed).

Figure 2 is a block diagram of a speech coder. Speech signals received at an input 1 are converted into samples by a sampler 2 and then into digital form in an analogue-to-digital converter 3. An analysis unit 4 computes, for each successive group of samples, the coefficients of a synthesis filter having a response corresponding to the spectral content of the speech. Derivation of LPC coefficients is well known and will not be described further here. The coefficients are supplied to an output multiplexer 5, and also to a local synthesis filter 6. The filter update rate may typically be once every 20 ms.

The coder has also a codebook store 7 containing the thirty-two codebook entries discussed above. The manner in which the entries are stored is not material to the present invention but it is assumed that each entry (for a five pulse excitation in a 32 sample period frame) contains the positions within the frame and the amplitudes of the four pulses after the first. This information, when read from the store is supplied to an excitation generator 8 which produces an actual excitation frame - i.e 32 values (of which 27 are zero, of course). Its output is supplied via a controllable shifting unit 9 to the input of the synthesis filter 6. The filter output is compared by a subtractor 10 with the input speech samples supplied via a buffer 11 (so that a number of comparisons can be made between one 32-sample speech frame and different filtered excitations).

In order to ascertain the appropriate shift value, certain techniques are borrowed from multipulse coding. In multipulse coding, a common method of deriving the pulse positions and amplitudes is an iterative one, in which one pulse is calculated which minimises the error between the synthetic and actual speech; a further pulse is then found which, in combination with the first, minimises the error and so on. Analysis of the statistics of MP-LPC pulses show that the first pulse to be derived usually has the largest amplitude.

This embodiment of the invention makes use of this by carrying out a multipulse search to find the location of this first pulse only. Any of the known methods for this may be employed, for example that described in B.S. Atal & J.R. Remde, 'A New Model of LPC Excitation for producing Natural Sounding Speech at Low Bit rates, Proc. IEEE Int. Conf. ASSP, Paris, 1982, p. 614.

A search unit 12 is shown in figure 2 for this purpose: its output feeds the shifter 9 to determine the rotational shift applied to the excitation generated by the generator 8. Effectively this selects, from 1024 excitations allowed by the codebook, a particular class of excitations, namely those with the largest pulse occupying the particular position determined by the search unit 13.

The output of the subtractor 10 feeds a control unit 13 which also supplies addresses to the store 7 and shift values to the shifting unit 9. The purpose of the control unit is to ascertain which of the 32 possible excitations represented by the selected class gives the smallest subtractor output (usually the mean square value of the differences, over a frame). The finally determined entry and shift are output in the form of a codeword C and shift value S to the output multiplexer 5.

The entry determination by the control unit for a given frame of speech available at the output of the buffer 11 is as follows:

(i) apply successive codewords (codebook addresses) to the store 7

(ii) apply to each codebook entry a shift such as to move the largest pulse to the position indicated by the 'multipulse' search.

(iii) monitor the output of the subtractor 10 for all 32 entries to ascertain which gives rise to the lowest mean square difference.

(iv) output the codeword and shift value to the multiplexer.

Compared with a conventional CELP coder using a 1024 entry codebook, there is a small reduction in the singal-to-noise ratio obtained due to the constraints placed on the excitations (i.e. that they fall into 32 mutually shiftable classes). However there is a reduction in the codebook size and hence the storage requirement for the store 7. Moreover, the amount of computation to be carried out by the control unit 13 is significantly reduced since only 32 tests rather than 1024 need to be carried out.

To allow for the sub-optimal selection, inherent in the 'multipulse search', the above process may also include excitations which are shifted a few positions before and after the position found by the search.

This could be achieved by the control unit adding/subtracting appropriate values from the shift value supplied to the shifting unit 9, as indicated by the dotted line connection. However, since the filtered output of a time shifted version of a given excitation is a time shifted version of the filter's response to the given excitation, these shifts could instead be performed by a second shifter 14 placed after the synthesis filter 6. Once wrap-around occurs, however, the result is no longer correct: this problem may be accommodated by (a) not performing shifts which cause wrap around (b) performing the shift but allowing pulses to be lost rather than wrapped around (and informing the decoder) or (c) permitting wraparound but performing a correction to account for the error.

The generation of the codebook remains to be mentioned. This can be generated by Gaussian noise techniques, in the manner already proposed in "Scholastic Coding of Speech Signals at very low Bit Rates", B.S. Atal & M.R. Schroeder, Proc IEEE Int Conf on Communications, 1984, pp1610-1613. A further advantage can be gained however by generating the codebook by statistical analysis of the results produced by a multipulse coder. This can remove the approximation involved in the assumption that the first pulse derived by the 'multipulse search' is the largest, since the codebook entries can then be stored with the first obtained pulse in a standard position, and shifted such that this pulse is brought to the position derived by the unit.

Although the various function elements shown in figure 2 are indicated separately, in practice some or all of them might be performed by the same hardware. One of the commercially available digital signal processing (DSP) integrated circuits, suitably programmed, might be employed, for example.

Although the 'multipulse search' option has been described in the context of shifted codebook entries, it can also be applied to other situations where the allowed excitations can be divided into classes within which all the excitations have the largest, or most significant, pulse in a particular position within the frame. The position of the derived pulse is then used to select the appropriate class and only the codebook entries in that class need to be tested.

Figure 3 shows a decoder for reproducing signals encoded by the apparatus of figure 2.

An input 30 supplies a demultiplexer 31 which (a) supplies filter coefficients to a synthesis filter 32; (b) supplies codewords to the address input of a codebook store 33; (c) supplies shift values to a shifter 34 which conveys the output of an excitation generator 35 connected to the store 33 to the input of the synthesis filter 32. Speech output from the filter 32 is supplied via a digital-to-analogue converter 36 to an output 37.

Claims

1. A speech coder comprising:

means arranged in operation to generate, from successive time frame periods of input speech signals, filter information defining successive representations of a synthesis filter response, and to output the filter information;

means arranged in operation, for each of successive time frame periods of the speech, to receive the input speech signals and the respective filter information, and to generate excitation information, comprising:

(a) a data store defining a plurality of excitation frames each consisting of a plurality of pulses;

(b) means for selecting one excitation frame out of the said plurality of frames and rotationally shifted versions of the frames and for generating data identifying the store entry and the amount if any of its rotational shift;

in which the selecting means is arranged to:

(i) determine, out of a plurality of single-pulse frames each consisting of a single pulse at a different location, which frame meets the criterion that it would when applied to the input of a filter having the response defined by the filter information produce a frame of synthetic speech which most closely resembles the frame of input speech; and

(ii) determine which of the said plurality of stored frames, when rotationally shifted by an amount derived from the determined pulse location, meets the said criterion.

(iv) output the codeword and shift value to the multiplexer.

Compared with a conventional CELP coder using a 1024 entry codebook, there is a small reduction in the singal-to-noise ratio obtained due to the constraints placed on the excitations (i.e. that they fall into 32 mutually shiftable classes). However there is a reduction in the codebook size and hence the storage requirement for the store 7. Moreover, the amount of computation to be carried out by the control unit 13 is significantly reduced since only 32 tests rather than 1024 need to be carried out.

To allow for the sub-optimal selection, inherent in the 'multipulse search', the above process may also include excitations which are shifted a few positions before and after the position found by the search.

This could be achieved by the control unit adding/subtracting appropriate values from the shift value supplied to the shifting unit 9, as indicated by the dotted line connection. However, since the filtered output of a time shifted version of a given excitation is a time shifted version of the filter's response to the given excitation, these shifts could instead be performed by a second shifter 14 placed after the synthesis filter 6. Once wrap-around occurs, however, the result is no longer correct: this problem may be accommodated by (a) not performing shifts which cause wrap around (b) performing the shift but allowing pulses to be lost rather than wrapped around (and informing the decoder) or (c) permitting wraparound but performing a correction to account for the error.

The generation of the codebook remains to be mentioned. This can be generated by Gaussian noise techniques, in the manner already proposed in "Scholastic Coding of Speech Signals at very low Bit Rates", B.S. Atal & M.R. Schroeder, Proc IEEE Int Conf on Communications, 1984, pp1610-1613. A further advantage can be gained however by generating the codebook by statistical analysis of the results produced by a multipulse coder. This can remove the approximation involved in the assumption that the first pulse derived by the 'multipulse search' is the largest, since the codebook entries can then be stored with the first obtained pulse in a standard position, and shifted such that this pulse is brought to the position derived by the unit.

Although the various function elements shown in figure 2 are indicated separately, in practice some or all of them might be performed by the same hardware. One of the commercially available digital signal processing (DSP) integrated circuits, suitably programmed, might be employed, for example.

Although the 'multipulse search' option has been described in the context of shifted codebook entries, it can also be applied to other situations where the allowed excitations can be divided into classes within which all the excitations have the largest, or most significant, pulse in a particular position within the frame. The position of the derived pulse is then used to select the appropriate class and only the codebook entries in that class need to be tested.

Figure 3 shows a decoder for reproducing signals encoded by the apparatus of figure 2.

An input 30 supplies a demultiplexer 31 which (a) supplies filter coefficients to a synthesis filter 32; (b) supplies codewords to the address input of a codebook store 33; (c) supplies shift values to a shifter 34 which conveys the output of an excitation generator 35 connected to the store 33 to the input of the synthesis filter 32. Speech output from the filter 32 is supplied via a digital-to-analogue converter 36 to an output 37.

Claims

1. A speech coder comprising:

means arranged in operation to generate, from successive time frame periods of input speech signals, filter information defining successive representations of a synthesis filter response, and to output the filter information;

means arranged in operation, for each of successive time frame periods of the speech, to receive the input speech signals and the respective filter information, and to generate excitation information, comprising:

(a) a data store defining a plurality of excitation frames each consisting of a plurality of pulses;

(b) means for selecting one excitation frame out of the said plurality of frames and rotationally shifted versions of the frames and for generating data identifying the store entry and the amount if any of its rotational shift;

in which the selecting means is arranged to:

(i) determine, out of a plurality of single-pulse frames each consisting of a single pulse at a different location, which frame meets the criterion that it would when applied to the input of a filter having the response defined by the filter information produce a frame of synthetic speech which most closely resembles the frame of input speech; and

(ii) determine which of the said plurality of stored frames, when rotationally shifted by an amount derived from the determined pulse location, meets the said criterion.

2. A speech coder according to claim 1 in which the said rotationally shifted versions consist of the stored frames each shifted by an amount corresponding to the determined pulse location.

3. A speech coder according to claim 1 in which the said rotationally shifted versions comprise the stored frames each shifted by an amount corresponding to the determined pulse location, and those frames subjected to additional shifts which are small relative to the frame size.

4. A speech coder according to claim 2 or 3 in which the said amount of shift corresponding to the determined pulse location is that shift which brings the largest pulse of the excitation frame into the same location within the frame as the determined single pulse.

5. A speech coder according to claim 3 or 4 in which each of the said plurality of stored excitation frames has been generated by a training sequence comprising identification of the location, within a single-pulse frame which meets the said criterion of a single, first, pulse followed by determination of further pulses to be included in the excitation frame, and in which the said amount of shift corresponding to the determined location is that shift which brings the said first pulse of the excitation frame into the same location within the frame as has the single pulse of the single pulse frame determined by the selecting means.

Revendications

1. Un codeur de parole comprenant :

des moyens agencés en fonctionnement pour engendrer, à partir de périodes successives de base de temps de signaux d'entrée de parole, une information de filtre définissant des représentations successives d'une réponse de filtre de synthèse et pour envoyer l'information de filtre;

des moyens agencés en fonctionnement, pour chacune des périodes successives de base de temps de la parole, pour recevoir des signaux d'entrée de parole et l'information respective de filtre, et pour engendrer une information d'excitation comprenant :

(a) une mémoire de donnée définissant une pluralité de bases d'excitation consistant chacune en une pluralité d'impulsions;

(b) des moyens pour choisir une base d'excitation, parmi ladite pluralité de bases et de versions décalées par rotation des bases, et pour engendrer des données identifiant l'entrée de mémoire et l'amplitude éventuelle de son décalage en rotation;

dans lequel le moyen de choix est agencé pour:

(i) déterminer, parmi une pluralité de bases à impulsion unique consistant chacune en une impulsion unique à un emplacement différent, quelle base répond au critère consistant en ce qu'elle produirait, si elle était appliqué à l'entrée d'un filtre présentant la réponse définie par l'information de filtre, un base de parole synthétique qui ressemble de la façon la plus étroite à la base de la parole d'entrée; et

(ii) déterminer quelle base parmi la pluralité des bases mémorisées répond audit critère lorsqu'elle est décalée par rotation d'une ampleur dérivée de l'emplacement déterminé d'impulsion.

2. Un codeur de parole selon la revendication 1 dans lequel lesdites versions décalées par rotation sont constituées par les bases mémorisées décalées chacune d'une ampleur correspondant à l'emplacement déterminé d'impulsion.

3. Un codeur de parole selon la revendication 1 dans lequel lesdites versions décalées par rotation comprennent les bases mémorisées décalées chacune d'une ampleur correspondant à l'emplacement déterminé d'impulsion, et les bases soumises à des décalages additionnels qui sont petits par rapport à la dimension de base.

4. Un codeur de parole selon la revendication 2 ou 3, dans lequel ladite ampleur de décalage correspond à l'emplacement déterminé d'impulsion est le décalage qui amène la plus grande impulsion de la base d'excitation au même emplacement que l'impulsion unique déterminée à l'intérieur de la base.

5. Un codeur de parole selon la revendication 3 ou 4 dans lequel chaque base de ladite pluralité de bases d'excitation mémorisées a été engendrée par une séquence d'apprentissage comprenant une identification de l'emplacement, à l'intérieur d'une base à impulsion unique qui répond audit critère d'une première impulsion, unique, suivie par la détermination d'autres impulsions à inclure dans la base d'excitation, et dans lequel ladite ampleur de décalage correspondant à l'emplacement déterminé est le décalage qui amène ladite première impulsion de la base d'excitation au

même emplacement à l'intérieur de la base que celui de l'impulsion unique de la base à impulsion unique déterminée par le moyen de choix.

Patentansprüche

1. Sprachkodierer, der aufweist:
eine Einrichtung, die im Betrieb ausgelegt ist, aus aufeinanderfolgenden Zeit-Block-Perioden von Eingabesprachsignalen eine Filterinformation zu erzeugen, die aufeinanderfolgende Darstellungen einer Synthese-Filterantwort definiert, und die Filterinformation auszugeben;
eine Einrichtung, die im Betrieb für jede der aufeinanderfolgenden Zeit-Block-Perioden der Sprache ausgelegt ist, die Eingabesprachsignale und die jeweilige Filterinformation zu empfangen und eine Anregungsinformation zu erzeugen, wobei die Einrichtung aufweist:
(a) einen Datenspeicher, der eine Vielzahl von Anregungsblöcken definiert, wobei jeder aus einer Vielzahl von Impulsen besteht;
(b) eine Einrichtung zum Auswählen eines Anregungsblöcken aus der Vielzahl von Blöcken und rotationsmäßig verschobenen Versionen der Blöcke und zum Erzeugen von Daten, die den Speichereingang und den Betrag irgendeiner ihrer Rotationsverschiebung identifizieren;
wobei die Auswahleinrichtung angeordnet ist zum:
(i) Festlegen aus einer Vielzahl von Einzelimpulsblöcken, von denen jeder aus einem einzelnen Impuls an einem unterschiedlichen Ort besteht, welcher Block das Kriterium erfüllt, das er, wenn er an den Eingang eines Filters mit der Antwort angelegt wird, durch die Filterinformation definiert, einen Block einer synthetischen Sprache erzeugen würde, der dem Block der Eingabesprache am meisten ähnelt; und
(ii) Bestimmen, welcher der Vielzahl gespeicherter Blöcke, wenn er um einen Betrag rotationsmäßig verschoben ist, der sich von dem bestimmten Impulsort ableitet, das Kriterium erfüllt.

2. Sprachkodierer nach Anspruch 1, wobei die rotationsmäßig verschobenen Versionen aus den gespeicherten Blöcke bestehen, von denen jeder um einen Betrag entsprechend dem bestimmten Impulsort verschoben ist.

3. Sprachkodierer nach Anspruch 1, wobei die rotationsmäßig verschobenen Versionen die gespeicherten Blöcke umfassen, von denen jeder um einen Betrag entsprechend dem bestimmten Impulsort verschoben ist, und jenen Block, die zusätzlichen Verschiebungen ausgesetzt sind, die relativ zu der Blockgröße klein sind.

4. Sprachkodierer nach Anspruch 2 oder 3, wobei der Betrag der Verschiebung entsprechend dem bestimmten Impulsort jene Verschiebung ist, die den größten Impuls des Anregungsblockes in den gleichen Ort innerhalb des Blockes bringt, wie der bestimmte einzelne Impuls.

5. Sprachkodierer nach Anspruch 3 oder 4, wobei jeder der Vielzahl gespeicherter Anregungsblöcke durch eine Übungsfolge erzeugt worden ist, die eine Erkennung des Ortes umfaßt, und zwar innerhalb eines Einzelimpulsblockes, der das Kriterium eines einzelnen, ersten Impulses erfüllt, gefolgt durch eine Bestimmung weiterer Impulse, die in dem Anregungsblock enthalten sein sollen, und wobei der Verschiebungsbetrag, der dem bestimmten Ort entspricht, jene Verschiebung ist, die den ersten Impuls des Anregungsblockes in den gleichen Ort innerhalb des Rahmens bringt, wie ihn der Einzelimpuls des Einzelimpulsblockes hat, der durch die Auswahleinrichtung bestimmt ist.