



(11) **EP 1 433 164 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
14.11.2007 Bulletin 2007/46

(21) Application number: **02757200.7**

(22) Date of filing: **19.08.2002**

(51) Int Cl.:
G10L 19/00 (2006.01)

(86) International application number:
PCT/US2002/026255

(87) International publication number:
WO 2003/023763 (20.03.2003 Gazette 2003/12)

(54) **IMPROVED FRAME ERASURE CONCEALMENT FOR PREDICTIVE SPEECH CODING BASED ON
EXTRAPOLATION OF SPEECH WAVEFORM**

VERBESSERTES VERBERGEN EINER RAHMENLÖSCHUNG FÜR DIE PRÄDIKTIVE
SPRACHCODIERUNG AUF DER BASIS EINER EXTRAPOLATION EINER SPRACHSIGNALFORM

MASQUAGE AMELIORE DE L'EFFACEMENT DES TRAMES DESTINE AU CODAGE PREDICTIF
DE LA PAROLE BASE SUR L'EXTRAPOLATION DE LA FORME D'ONDES DE LA PAROLE

(84) Designated Contracting States:
DE FR GB

(30) Priority: **17.08.2001 US 312789 P**
04.01.2002 US 344374 P
28.06.2002 US 183608
28.06.2002 US 183448
28.06.2002 US 183451

(43) Date of publication of application:
30.06.2004 Bulletin 2004/27

(73) Proprietor: **Broadcom Corporation**
Irvine, CA 92617 (US)

(72) Inventor: **CHEN, Juin-Hwey**
Irvine, CA 92612 (US)

(74) Representative: **Jehle, Volker Armin**
Patentanwälte
Bosch, Graf von Stosch, Jehle,
Flüggenstrasse 13
80639 München (DE)

(56) References cited:
EP-A- 0 747 882 WO-A-99/66494
US-A- 5 699 485 US-A- 6 085 158
US-B1- 6 188 980

- **WATKINS C R ET AL: "Improving 16 kb/s G.728 LD-CELP speech coder for frame erasure channels" ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1995. ICASSP-95., 1995 INTERNATIONAL CONFERENCE ON DETROIT, MI, USA 9-12 MAY 1995, NEW YORK, NY, USA, IEEE, US, vol. 1, 9 May 1995 (1995-05-09), pages 241-244, XP010625214 ISBN: 0-7803-2431-5**
- **ANONYMOUS: "Frame or packet loss concealment for the LD-CELP decoder" INTERNATIONAL TELECOMMUNICATION UNION. GENEVA, CH, May 1999 (1999-05), XP017400887**
- **JUIN-HWEY CHEN: "A high-fidelity speech and audio codec with low delay and low complexity" ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 2000. ICASSP '00. PROCEEDINGS. 2000 IEEE INTERNATIONAL CONFERENCE ON 5-9 JUNE 2000, PISCATAWAY, NJ, USA, IEEE, vol. 2, 5 June 2000 (2000-06-05), pages 1161-1164, XP010504934 ISBN: 0-7803-6293-4**

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 1 433 164 B1

Description

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to digital communications. More particularly, the present invention relates to the enhancement of speech quality when frames of a compressed bit stream representing a speech signal are lost within the context of a digital communications system.

Background Art

[0002] In speech coding, sometimes called voice compression, a coder encodes an input speech or audio signal into a digital bit stream for transmission. A decoder decodes the bit stream into an output speech signal. The combination of the coder and the decoder is called a codec. The transmitted bit stream is usually partitioned into frames. In wireless or packet networks, sometimes frames of transmitted bits are lost, erased, or corrupted. This condition is called frame erasure in wireless communications. The same condition of erased frames can happen in packet networks due to packet loss. When frame erasure happens, the decoder cannot perform normal decoding operations since there are no bits to decode in the lost frame. During erased frames, the decoder needs to perform frame erasure concealment (FEC) operations to try to conceal the quality-degrading effects of the frame erasure.

[0003] One of the earliest FEC techniques is waveform substitution based on pattern matching, as proposed by Goodman, et al. in "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications", IEEE Transaction on Acoustics, Speech and Signal Processing, December 1986, pp. 1440 - 1448. This scheme was applied to Pulse Code Modulation (PCM) speech codec that performs sample-by-sample instantaneous quantization of speech waveform directly. This FEC scheme uses a piece of decoded speech waveform immediately before the lost frame as the template, and slides this template back in time to find a suitable piece of decoded speech waveform that maximizes some sort of waveform similarity measure (or minimizes a waveform difference measure).

[0004] Goodman's FEC scheme then uses the section of waveform immediately following a best-matching waveform segment as the substitute waveform for the lost frame. To eliminate discontinuities at frame boundaries, the scheme also uses a raised cosine window to perform an overlap-add technique between the correctly decoded waveform and the substitute waveform. This overlap-add technique increases the coding delay. The delay occurs because at the end of each frame, there are many speech samples that need to be overlap-added to obtain the final values, and thus cannot be played out until the next frame of speech is decoded.

[0005] Based on the work of Goodman above, David Kapilow developed a more sophisticated version of an FEC scheme for G.711 PCM codec. This FEC scheme is described in Appendix 1 of the ITU-T Recommendation G.71 1. However, both the FEC of Goodman and the FEC scheme of Kapilow are limited to PCM codecs with instantaneous quantization.

[0006] For speech coding, the most popular type of speech codec is based on predictive coding. Perhaps the first publicized FEC scheme for a predictive codec is a "bad frame masking" scheme in the original TIA IS-54 VSELP standard for North American digital cellular radio (rescinded in September 1996). Here, upon detection of a bad frame, the scheme repeats the linear prediction parameters of the last frame. This scheme derives the speech energy parameter for the current frame by either repeating or attenuating the speech energy parameter of last frame, depending on how many consecutive bad frames have been counted. For the excitation signal (or quantized prediction residual), this scheme does not perform any special operation. It merely decodes the excitation bits, even though they might contain a large number of bit errors.

[0007] The first FEC scheme for a predictive codec that performs waveform substitution in the excitation domain is probably the FEC system developed by Chen for the ITU-T Recommendation G.728 Low-Delay Code Excited Linear Predictor (CELP) codec, as described in United States Patent No. 5,615,298 issued to Chen, titled "Excitation Signal Synthesis During Frame Erasure or Packet Loss." In this approach, during erased frames, the speech excitation signal is extrapolated depending on whether the last frame is a voiced or an unvoiced frame. If it is voiced, the excitation signal is extrapolated by periodic repetition. If it is unvoiced, the excitation signal is extrapolated by randomly repeating small segments of speech waveform in the previous frame, while ensuring the average speech power is roughly maintained.

[0008] What is needed therefore is an FEC technique that avoids the noted deficiencies associated with the conventional decoders. For example, what is needed is an FEC technique that avoids the increased delay created in the overlap-add operation of Goodman's approach. What is also needed is an FEC technique that can ensure the smooth reproduction of a speech or audio waveform when the next good frame is received.

[0009] EP 0 747 882 A2 describes a method for use in a speech decoder which fails to receive a portion of each of first and second consecutive frames of compressed speech information, wherein a pitch delay associated with the first

of the consecutive frames is incremented and the incremented value is used as a pitch delay for the second of the consecutive frames.

[0010] According to the invention, there are provided a method of and an apparatus for synthesizing a number of corrupted frames as defined by independent claims 1 and 6, respectively, and a computer readable medium as defined by independent claim 9.

[0011] Further advantageous features of the invention are defined in the dependent subclaims.

[0012] Consistent with the principles of the present invention as embodied and broadly described herein, an exemplary FEC technique includes a method of synthesizing a number of corrupted frames output from a decoder including one or more predictive filters. The corrupted frames are representative of one segment of a decoded signal (sq(n)) output from the decoder. The method comprises determining a first preliminary time lag (ppfe1) based upon examining a predetermined number (K) of samples of another segment of the decoded signal and determining a scaling factor (ptfe) associated with the examined number (K) of samples when the first preliminary time lag (ppfe1) is determined. The method also comprises extrapolating one or more replacement frames based upon the first preliminary time lag (ppfe1) and the scaling factor (ptfe).

[0013] Further embodiments, features, and advantages of the present invention, as well as the structure and operation of the various embodiments of the present invention, are described in detail below with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0014] The accompanying drawings, which are incorporated in and constitute a part of the specification, illustrate an embodiment of the invention and, together with the description, explain the purpose, advantages, and principles of the invention. In the drawings:

[0015] FIG. 1 is a block diagram illustration of a conventional predictive decoder;

[0016] FIG. 2 is a block diagram illustration of an exemplary decoder constructed and arranged in accordance with the present invention;

[0017] FIG. 3(a) is a plot of an exemplary unnormalized waveform attenuation window functioning in accordance with the present invention;

[0018] FIG. 3(b) is a plot of an exemplary normalized waveform attenuation window functioning in accordance with the present invention; and

[0019] FIG. 4 is a block diagram of an exemplary computer system on which the present invention can be practiced.

DETAILED DESCRIPTION OF INVENTION

[0020] The following detailed description of the present invention refers to the accompanying drawings that illustrate exemplary embodiments consistent with this invention. Other embodiments are possible, and modifications may be made to the embodiments within the scope of the present invention. Therefore, the following detailed description is not meant to limit the invention. Rather, the scope of the invention is defined by the appended claims.

[0021] It would be apparent to one of skill in the art that the present invention, as described below, may be implemented in many different embodiments of hardware, software, firmware, and/or the entities illustrated in the drawings. Any actual software code with specialized control hardware to implement the present invention is not limiting of the present invention. Thus, the operation and behavior of the present invention will be described with the understanding that modifications and variations of the embodiments are possible, given the level of detail presented herein. Before describing the invention in detail, it is helpful to describe an exemplary environment in which the invention may be implemented.

[0022] The present invention is particularly useful in the environment of the decoder of a predictive speech codec to conceal the quality-degrading effects of frame erasure or packet loss. FIG. 1 illustrates such an environment. The general principles of the invention can be used in any linear predictive codec, although the preferred embodiment described later is particularly well suited for a specific type of predictive decoder.

[0023] The invention is an FEC technique designed for predictive coding of speech. One characteristic that distinguishes it from the techniques mentioned above, is that it performs waveform substitution in the speech domain rather than the excitation domain. It also performs special operations to update the internal states, or memories, of predictors and filters inside the predictive decoder to ensure maximally smooth reproduction of speech waveform when the next good frame is received.

[0024] The present invention also avoids the additional delay associated with the overlap-add operation in Goodman's approach and in ITU-T G.711 Appendix I. This is achieved by performing overlap-add between extrapolated speech waveform and the ringing, or zero-input response of the synthesis filter. Other features include a special algorithm to minimize buzzing sounds during waveform extrapolation, and an efficient method to implement a linearly decreasing waveform envelope during extended frame erasure. Finally, the associated memories within the log-gain predictor are

updated.

[0025] The present invention is not restricted to a particular speech codec. Instead, it's generally applicable to predictive speech codecs, including, but not limited to, Adaptive Predictive Coding (APC), Multi-Pulse Linear Predictive Coding (MPLPC), CELP, and Noise Feedback Coding (NFC), etc.

[0026] Before discussing the principles of the invention, a description of a conventional decoder of a standard predictive codec is needed. FIG. 1 is a block diagram illustration of a conventional predictive decoder 100. The decoder 100 shown in FIG. 1 can be used to describe the decoders of APC, MPLPC, CELP, and NFC speech codecs. The more sophisticated versions of the codecs associated with predictive decoders typically use a short-term predictor to exploit the redundancy among adjacent speech samples and a long-term predictor to exploit the redundancy between distant samples due to pitch periodicity of, for example, voiced speech.

[0027] The main information transmitted by these codecs is the quantized version of the prediction residual signal after short-term and long-term prediction. This quantized residual signal is often called the excitation signal, because it is used in the decoder to excite the long-term and short-term synthesis filter to produce the output decoded speech. In addition to the excitation signal, several other speech parameters are also transmitted as side information frame-by-frame or subframe-by-subframe.

[0028] An exemplary range of lengths for each frame (called frame size) is 5 ms to 40 ms, with 10 ms and 20 ms as the two most popular frame sizes for speech codecs. Each frame usually contains a few equal-length subframes. The side information of these predictive codecs typically includes spectral envelope information (in the form of the short-term predictor parameters), pitch period, pitch predictor taps (both long-term predictor parameters), and excitation gain.

[0029] In FIG. 1, the conventional decoder 100 includes a bit de-multiplexer 105. The de-multiplexer 105 separates the bits in each received frame of bits into codes for the excitation signal and codes for short-term predictor, long-term predictor, and the excitation gain.

[0030] The short-term predictor parameters, often referred to as the linear predictive coding (LPC) parameters, are usually transmitted once a frame. There are many alternative parameter sets that can be used to represent the same spectral envelope information. The most popular of these is the line-spectrum pair (LSP) parameters, sometimes called line-spectrum frequency (LSF) parameters. In FIG. 1, *LSPI* represents the transmitted quantizer codebook index representing the LSP parameters in each frame. A short-term predictive parameter decoder 110 decodes *LSPI* into an LSP parameter set and then converts the LSP parameters to the coefficients for the short-term predictor. These short-term predictor coefficients are then used to control the coefficient update of a short-term predictor 120.

[0031] Pitch period is defined as the time period at which a voiced speech waveform appears to be repeating itself periodically at a given moment. It is usually measured in terms of a number of samples, is transmitted once a subframe, and is used as the bulk delay in long-term predictors. Pitch taps are the coefficients of the long-term predictor. The bit de-multiplexer 105 also separates out the pitch period index (*PPI*) and the pitch predictor tap index (*PPTI*), from the received bit stream. A long-term predictive parameter decoder 130 decodes *PPI* into the pitch period, and decodes the *PPTI* into the pitch predictor taps. The decoded pitch period and pitch predictor taps are then used to control the parameter update of a generalized long-term predictor 140.

[0032] In its simplest form, the long-term predictor 140 is just a finite impulse response (FIR) filter, typically first order or third order, with a bulk delay equal to the pitch period. However, in some variations of CELP and MPLPC codecs, the long-term predictor 140 has been generalized to an adaptive codebook, with the only difference being that when the pitch period is smaller than the subframe, some periodic repetition operations are performed. The generalized long-term predictor 140 can represent either a straightforward FIR filter, or an adaptive codebook, thus covering most of the predictive speech codecs presently in use.

[0033] The bit de-multiplexer 105 also separates out a gain index *GI* and an excitation index *CI* from the input bit stream. An excitation decoder 150 decodes the *CI* into an unscaled excitation signal, and also decodes the *GI* into the excitation gain. Then, it uses the excitation gain to scale the unscaled excitation signal to derive a scaled excitation gain signal $uq(n)$, which can be considered a quantized version of the long-term prediction residual. An adder 160 combines the output of the generalized long-term predictor 140 with the scaled excitation gain signal $uq(n)$ to obtain a quantized version of a short-term prediction residual signal $dq(n)$. An adder 170 combines the output of the short-term predictor 120 to $dq(n)$ to obtain an output decoded speech signal $sq(n)$.

[0034] A feedback loop is formed by the generalized long-term predictor 140 and the adder 160 and can be regarded as a single filter, called a long-term synthesis filter 180. Similarly, another feedback loop is formed by the short term predictor 120 and the adder 170. This other feedback loop can be considered a single filter called a short-term synthesis filter 190. The long-term synthesis filter 180 and the short-term synthesis filter 190 combine to form a synthesis filter module 195.

[0035] In summary, the conventional predictive decoder 100 depicted in FIG. 1 decodes the parameters of the short-term predictor 120 and the long-term predictor 140, the excitation gain, and the unscaled excitation signal. It then scales the unscaled excitation signal with the excitation gain, and passes the resulting scaled excitation signal $uq(n)$ through the long-term synthesis filter 180 and the short-term synthesis filter 190 to derive the output decoded speech signal $sg(n)$.

[0036] When a frame of input bits is erased due to fading in a wireless transmission or due to packet loss in packet networks, the decoder 100 in FIG. 1 unfortunately loses the indices *LSPI*, *PPI*, *PPTI*, *GI*, and *CI*, needed to decode the speech waveform in the current frame.

[0037] According to the principles of the present invention, the decoded speech waveform immediately before the current frame is stored and analyzed. A waveform-matching search, similar to the approach of Goodman is performed, and the time lag and scaling factor for repeating the previously decoded speech waveform in the current frame are identified.

[0038] Next, to avoid the occasional buzz sounds due to repeating a waveform at a small time lag when the speech is not highly periodic, the time lag and scaling factor are sometimes modified as follows. If the analysis indicates that the stored previous waveform is not likely to be a segment of highly periodic voiced speech, and if the time lag for waveform repetition is smaller than a predetermined threshold, another search is performed for a suitable time lag greater than the predetermined threshold. The scaling factor is also updated accordingly.

[0039] Once the time lag and scaling factor have been determined, the present invention copies the speech waveform one time lag earlier to fill the current frame, thus creating an extrapolated waveform. The extrapolated waveform is then scaled with the scaling factor. The present invention also calculates a number of samples of the ringing, or zero-input response, output from the synthesis filter module 195 from the beginning of the current frame. Due to the smoothing effect of the short-term synthesis filter 190, such a ringing signal will seem to flow smoothly from the decoded speech waveform at the end of the last frame. The present invention then overlap-adds this ringing signal and the extrapolated speech waveform with a suitable overlap-add window in order to smoothly merge these two pieces of waveform. This technique will smooth out waveform discontinuity at the beginning of the current frame. At the same time, it avoids the additional delays created by G.711 Appendix I or the approach of Goodman.

[0040] If the frame erasure has persisted for an extended period of time, the extrapolated speech signal is attenuated toward zero. Otherwise, it will create a tonal or buzzing sound. In the present invention, the waveform envelope is attenuated linearly toward zero if the length of the frame erasure exceeds a certain threshold. The present invention then uses a memory-efficient method to implement this linear attenuation toward zero.

[0041] After the waveform extrapolation is performed in the erased frame, the present invention properly updates all the internal memory states of the filters within the speech decoder. If updating is not performed, there would be a large discontinuity and an audible glitch at the beginning of the next good frame. In updating the filter memory after a frame erasure, the present invention works backward from the output speech waveform. The invention sets the filter memory contents to be what they would have been at the end of the current frame, if the filtering operations of the speech decoder were done normally. That is, the filtering operations are performed with a special excitation such that the resulting synthesized output speech waveform is exactly the same as the extrapolated waveform calculated above.

[0042] As an example, if the short-term predictor 120 is of an order *M*, then the memory of the short-term synthesis filter 190, after the FEC operation for the current frame, is simply the last *M* samples of the extrapolated speech signal for the current frame with the order reversed. This is because the short-term synthesis filter 190 in the conventional decoder 100 is an all-pole filter. The filter memory is simply the previous filter output signal samples in reverse order.

[0043] An example of updating the memory of the FIR long-term predictor 140 will be presented. In this example, the present invention performs short-term prediction error filtering of the extrapolated speech signal of the current frame, with initial memory of the short-term predictor 120 set to the last *M* samples (in reverse order) of the output speech signal in the last frame.

[0044] Similarly, if quantizers for side information (such as LSP and excitation gain) use inter-frame predictive coding, then the memories of those predictors are also updated based on the same principle to minimize the discontinuity of decoded speech parameters at the next good frame.

[0045] After the first received good frame following a frame erasure, the present invention also attempts to correct filter memories within the long-term synthesis filter 180 and the short-term synthesis 190 filter if certain conditions are met. Conceptually, the present invention first performs linear interpolation between the pitch period of the last good frame before the erasure and the pitch period of the first good frame after the erasure. Such linear interpolation of the pitch period is performed for each of the erased frames. Based on this linearly interpolated pitch contour, the present invention then re-extrapolates the long-term synthesis filter memory and re-calculates the short-term synthesis filter memory at the end of the last erased frame (i.e., before decoding the first good frame after the erasure).

[0046] The general principles of the present invention outlined above are applicable to almost any predictive speech decoder. What will be described in greater detail below is a particular implementation of those general principles, in a preferred embodiment of the present invention applied to the decoder of a two-stage noise feedback codec.

[0047] FIG. 2 is a block diagram illustration of an exemplary embodiment of the present invention. The decoder can be, for example, the decoder 100 shown in FIG. 1. Also included in the embodiment of FIG. 2 is an input frame erasure flag switch 200. If the input frame erasure flag 200 indicates that the current frame received is a good frame, the decoder 100 performs the normal decoding operations as described above. If, however, the frame is the first good frame after a frame erasure, the long-term and short-term synthesis filter memories can be corrected before starting the normal

decoding. When a good frame is received, the frame erasure flag switch 200 is in the upper position, and the decoded speech waveform $sq(n)$ is used as the output of the system. Furthermore, the current frame of decoded speech $sq(n)$ is also passed to a module 201, which stores the previously decoded speech waveform samples in a buffer. The current frame of decoded speech $sq(n)$ is used to update that buffer. The remaining modules in FIG. 2 are inactive during a good frame.

[0048] On the other hand, if the input frame erasure flag switch 200 indicates that the current frame was not received, was erased, or was corrupted, then the operation of the decoder 100 is halted and the frame erasure flag switch 200 is changes to the lower position. The remaining modules of FIG. 2 then perform frame erasure concealment operations to produce the output speech waveform $sq'(n)$ for the current frame, and also update the filter memories of the decoder 100 to prepare the decoder 100 for the normal decoding operations of the next received good frame. The remaining modules of FIG. 2 work in the following way.

[0049] A module 201 calculates L samples of "ringing," or zero-input response, of the synthesis filter in FIG. 1. A simpler approach is to use only the short-term synthesis filter 190, but a more effective approach (at least for voiced speech) is to use the ringing of the cascaded long-term and short-term synthesis filters 180 and 190. This is done in the following way. Starting with the memory of the synthesis filter (or filters) left in the delay line after the processing of the last frame, the filtering operations for L samples are performed while using a zero input signal to the filter. The resulting L samples of the long-term synthesis filter output signal form the desired "long-term filter ringing" signal, $ltr(n)$, $n = 1, 2, \dots, L$. The resulting L samples of the short-term synthesis filter output signal are simply called the "ringing" signal, $r(n)$, $n = 1, 2, \dots, L$. Both $ltr(n)$ and $r(n)$ are stored for later use.

[0050] A module 202 analyzes the previously decoded speech waveform samples stored in the module 201 to determine a first time lag $ppfe1$ and an associated scaling factor $ptfe1$ for waveform extrapolation in the current frame. This can be done in a number of ways. One way, for example, uses the approaches outlined by Goodman et al. And discussed above. If there are multiple consecutive frames erased, the module 202 is active only at the first erased frame. From the second erased frame on, the time lag and scaling factor found in the first erased frame are used.

[0051] The present invention will typically usually just search for a "pitch period" in the general sense, as in a pitch-prediction-based speech codec. If the decoder 100 has a decoded pitch period of the last frame, and if it is deemed reliable, then the embodiment of FIG. 2 will simply search around the neighborhood of this pitch period pp to find a suitable time lag. If the decoder 100 does not provide a decoded pitch period, or if this pitch period is deemed unreliable, then the embodiment of FIG. 2 will perform a full-scale pitch estimation to get the desired time lag. In FIG. 2, it is assumed that such a decoded pitch period pp is indeed available and reliable. In this case, the embodiment of FIG. 2 operates as follows.

[0052] Let $pplast$ denote the pitch period of the last good frame before the frame erasure. If $pplast$ is smaller than 10 ms (80 samples and 160 samples for 8 kHz and 16 kHz sampling rates, respectively), the module 202 uses it as the analysis window size K . If $pplast$ is greater than 10 ms, the module 202 uses 10 ms as the analysis window size K .

[0053] The module 202 then determines the pitch search range as follows. It subtracts 0.5 ms (4 samples and 8 samples for 8 kHz and 16 kHz sampling, respectively) from $pplast$, compares the result with the minimum allowed pitch period in the codec, and chooses the larger of the two as the lower bound of the search range, lb . It then adds 0.5 ms to $pplast$, compares the result with the maximum allowed pitch period in the codec, and chooses the smaller of the two as the upper bound of the search range, ub .

[0054] The $sq(n)$ buffer in the module 201 stores $N + N_f$ samples of speech, where the samples $sq(n)$, $n = 1, 2, \dots, N$ correspond to the decoder output speech for previous frames, with $sq(N)$ being the last sample of decoded speech in the last frame. N_f is the number of samples in a frame. The storage spaces $sq(n)$, $n = N + 1, N + 2, \dots, N + N_f$ are unpopulated at the beginning of a bad frame, but will be filled with extrapolated speech waveform samples once the operations of the modules 202 through 210 are completed.

[0055] For time lags $j = lb, lb+1, lb+2, \dots, ub-1, ub$, the module 202 calculates the correlation value

$$c(j) = \sum_{n=N-K+1}^N sq(n)sq(n-j)$$

for $j \in [lb, ub]$. Among those time lags that give a positive correlation $c(j)$, module 202 finds the time lag j that maximizes

$$nc(j) = \frac{\left(\sum_{n=N-K+1}^N sq(n)sq(n-j) \right)^2}{\sum_{n=N-K+1}^N sq^2(n-j)}.$$

[0056] The division operation above can be avoided by the "cross-multiply" method. The time lag j that maximizes $nc(j)$ is also the time lag within the search range that maximizes the pitch prediction gain for a single-tap pitch predictor. This is called the optimal time lag $ppfe1$, which stands for *pitch period for frame erasure*, 1st version. In the extremely rare case where no $c(j)$ in the search range is positive, $ppfe1$ is set to lb in this degenerate case.

[0057] Once $ppfe1$ is identified, the associated scaling factor $ptfe1$ is calculated as follows.

$$ptfe1 = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|}$$

[0058] Such a calculated scaling factor $ptfe1$ is then clipped to 1 if it is greater than 1 and clipped to -1 if it is less than -1. Also, in the degenerate case when the denominator on the right-hand side of the above equation is zero, $ptfe1$ is set to 0.

[0059] Although the module 202 performs the above calculation only for the first erased frame when there are multiple consecutive erased frames, it also attempts to modify the first time lag $ppfe1$ at the second consecutively erased frame, depending on the pitch period contour at the good frames immediately before the erasure. Starting from the last good frame before the erasure, and going backward frame-by-frame for up to 4 frames, the module 202 compares the transmitted pitch period until there is a change in the transmitted pitch period. If there is no change in pitch period found during these 4 good frames before the erasure, then the first time lag $ppfe1$ found above at the first erased frame is also used for the second consecutively erased frame. Otherwise, the first pitch change identified in the backward search above is examined to see if the change is relatively small. If the change is within 5%, then, depending on how many good frames back the pitch change is found, the amount of pitch period change per frame is calculated and is rounded to the nearest integer. The module 202 then adds this rounded pitch period change per frame, whether positive or negative, to the $ppfe1$ found above at the first erased frame. The resulting value is used as the first time lag $ppfe1$ for the second and subsequent consecutively erased frames. This modification of the first time lag after the second erased frame improves the speech quality on average.

[0060] It has been discovered that if this time lag is used directly as the time lag for periodic repetition in waveform extrapolation of the current frame, buzz sounds occur when a small time lag is used in a segment of speech that does not have a high degree of periodicity. To combat this problem, the present invention uses a module 203 to distinguish between highly periodic voiced speech segments and other types of speech segments. If the module 203 determines that the decoded speech is in a highly periodic voiced speech region, it sets the periodic waveform extrapolation flag $pwef$ to 1; otherwise, $pwef$ is set to 0. If $pwef$ is 0, and if the first time lag $ppfe1$ is less than a threshold of 10 ms, then a module 204 will find a second, larger time lag $ppfe2$ greater than 10 ms to reduce or eliminate the buzz sound.

[0061] Using $ppfe1$ as its input, the module 203 performs further analysis of the previously decoded speech $sq(n)$ to determine the periodic waveform extrapolation flag $pwef$. Again, this can be done in many possible ways. One exemplary method of determining the periodic waveform flag $pwef$ is described below.

[0062] The module 203 calculates three signal features: signal gain relative to long-term average of input signal level, pitch prediction gain, and the first normalized autocorrelation coefficient. It then calculates a weighted sum of these three signal features, and compares the resulting figure of merit with a pre-determined threshold. If the threshold is exceeded, $pwef$ is set to 1, otherwise it is set to 0. The module 203 then performs special handling for extreme cases.

[0063] The three signal features are calculated as follows. First, the module 203 calculates the speech energy in the analysis window:

$$E = \sum_{n=N-K+1}^N sq^2(n)$$

[0064] If $E > 0$, the base-2 logarithmic gain is calculated as $lg = \log_2 E$; otherwise, lg is set to 0. Let lv be the long-term average logarithmic gain of the active portion of the speech signal (that is, not counting the silence). A separate estimator for input signal level can be employed to calculate lv . An exemplary signal level estimator is disclosed in U.S. Provisional Application No. 60/312,794, filed August 17, 2001, entitled "Bit Error Concealment Methods for Speech Coding," and U.S. Provisional Application No. 60/344,378, filed January 4, 2002, entitled "Improved Bit Error Concealment Methods for Speech Coding." The normalized logarithmic gain (i.e., signal gain relative to long-term average input signal level) is then calculated as $nlg = lg - lv$.

[0065] The module 203 further calculates the first normalized autocorrelation coefficient

$$\rho_1 = \frac{\sum_{n=N-K+2}^N sq(n)sq(n-1)}{E}$$

[0066] In the degenerate case when $E = 0$, ρ_1 is set to 0 as well. The module 203 also calculates the pitch prediction gain as

$$ppg = 10 \log_{10} \left(\frac{E}{R} \right),$$

where

$$R = E - \frac{c^2(ppfe1)}{\sum_{n=N-K+1}^N sq^2(n - ppfe1)}$$

is the pitch prediction residual energy. In the degenerate case when $R = 0$, ppg is set to 20.

[0067] The three signal features are combined into a single figure of merit:

$$fom = nlg + 1.25 ppg + 16 \rho_1$$

If $fom > 16$, $pwef$ is set to 1, otherwise it is set to 0. Afterward, the flag $pwef$ may be overwritten in the following extreme cases:

- If $nlg < -1$, $pwef$ is set to 0.
- If $ppg > 12$, $pwef$ is set to 1.
- If $\rho_1 < -0.3$, $pwef$ is set to 0.

[0068] If $pwef = 0$ and $ppfe1 < T_0$, where T_0 is the number of samples corresponding to 10 ms, there is a high likelihood for periodic waveform extrapolation to produce a buzz sound. To avoid the potential buzz sound, the present invention searches for a second time lag $ppfe2 \geq T_0$. Two waveforms, one extrapolated using the first time lag $ppfe1$, and the other extrapolated using the second time lag $ppfe2$, are added together and properly scaled, and the resulting waveform is

used as the output speech of the current frame. By requiring the second time lag $ppfe2$ to be large enough, the likelihood of a buzz sound is greatly reduced. To minimize the potential quality degradation caused by a misclassification of a periodic voiced speech segment into something that is not, the present invention searches in the neighborhood of the first integer multiple of $ppfe1$ that is no smaller than T_0 . Thus, even if the flag $pwef$ should have been 1 and is misclassified as 0, there is a good chance that an integer multiple of the true pitch period will be chosen as the second time lag $ppfe2$ for periodic waveform extrapolation.

[0069] A module 204 determines the second time lag $ppfe2$ in the following way if $pwef = 0$ and $ppfe1 < T_0$. First, it finds the smallest integer m that satisfies

$$m \times ppfe1 \geq T_0 .$$

[0070] Next, the module 204 sets m_1 , the lower bound of the time lag search range, to $m \times ppfe1 - 3$ or T_0 , whichever is larger. The upper bound of the search range is set to $m_2 = m_1 + N_s - 1$, where N_s is the number of possible time lags in the search range. Next, for each time lag j in the search range of $[m_1, m_2]$, the module 204 calculates

$$cor(j) = \sum_{n=N-N_f+1}^N sq(n)sq(n-j)$$

and then selects the time lag $j \in [m_1, m_2]$ that maximizes $cor(j)$. The corresponding scaling factor is set to 1.

[0071] The module 205 extrapolates speech waveform for the current erased frame based on the first time lag $ppfe1$. It first extrapolates the first L samples of speech in the current frame using the first time lag $ppfe1$ and the corresponding scaling factor $ptfe1$. A suitable value of L is 8 samples. The extrapolation of the first L samples of the current frame can be expressed as

$$sq(n) = ptfe1 \times sq(n - ppfe1), \text{ for } n = N+1, N+2, \dots, N+L.$$

[0072] For the first L samples of the current frame, the module 205 smoothly merges the $sq(n)$ signal extrapolated above with $r(n)$, the ringing of the synthesis filter calculated in module 206, using the overlap-add method below.

$$sq(N+n) \leftarrow w_u(n)sq(N+n) + w_d(n)r(n), \text{ for } n = 1, 2, \dots, L.$$

[0073] In the equation above, the sign " $\leftarrow$ " means the quantity on its right-hand side overwrites the variable values on its left-hand side. The window function $w_u(n)$ represents the overlap-add window that is ramping up, while $w_d(n)$ represents the overlap-add window that is ramping down. These overlap-add windows satisfy the constraint:

$$w_u(n) + w_d(n) = 1$$

[0074] A number of different overlap-add windows can be used. For example, the raised cosine window mentioned in the paper by Goodman et al. is one exemplary method. Alternatively, simpler triangular windows can also be used.

[0075] After the first L samples of the current frame are extrapolated and overlap-added, the module 205 then extrapolates the remaining samples of the current frame. If $ppfe1 \geq N_f$, the extrapolation is performed as

$$sq(n) = ptfe1 \times sq(n - ppfe1), \text{ for } n = N+L+1, N+L+2, \dots, N+N_f.$$

If $ppfe1 < N_f$, then the extrapolation is performed as

$$sq(n) = ptfel \times sq(n - ppfel), \text{ for } n = N + L + 1, N + L + 2, \dots, N + ppfel,$$

5 and then

$$sq(n) = sq(n - ppfel), \text{ for } n = N + ppfel + 1, N + ppfel + 2, \dots, N + N_f.$$

10

[0076] A module 207 extrapolates speech waveform for the current erased frame based on the second time lag $ppfe2$. Its output extrapolated speech waveform $sq_2(n)$ is given by

15

$$sq_2(n) = \begin{cases} sq(n - ppfe2), & \text{for } n = N + 1, \dots, N + N_f, \text{ if } pwef = 0 \text{ and } ppfel < T_0 \\ 0, & \text{for } n = N + 1, \dots, N + N_f, \text{ if } pwef \neq 0 \text{ or } ppfel \geq T_0 \end{cases}$$

20

[0077] If the $sq_2(n)$ waveform is zero, a module 208 directly passes the output of the module 205 to a module 209. Otherwise, the module 208 adds the output waveforms of the modules 205 and 207, and scales the result appropriately. Specifically, it calculates the sums of signal sample magnitudes for the outputs of the modules 205 and 207:

25

$$sum1 = \sum_{n=N+1}^{N+N_f} |sq(n)|$$

30

$$sum2 = \sum_{n=N+1}^{N+N_f} |sq_2(n)|$$

35

It then adds the two waveforms and assign the result to $sq(n)$ again:

40

$$sq(n) \leftarrow sq(n) + sq_2(n), \text{ for } n = N + 1, N + 2, \dots, N + N_f.$$

Next, it calculates the sum of signal sample magnitudes for the summed waveform:

45

$$sum3 = \sum_{n=N+1}^{N+N_f} |sq(n)|$$

50

[0078] If $sum3$ is zero, the scaling factor is set to 1; otherwise, the scaling factor is set to $sfac = \min(sum1, sum2) / sum3$, where $\min(sum1, sum2)$ means the smaller of $sum1$ and $sum2$. After this scaling factor is calculated, the module 208 replaces $sq(n)$ with a scaled version: $sq(n) \leftarrow sfac \times sq(n)$, for $n = N + 1, N + 2, \dots, N + N_f$, and the module 209 performs overlap-add of $sq(n)$ and the ringing of the synthesis filter for the first L samples of the frame. The resulting waveform is passed to the module 210.

55

[0079] If the frame erasure lasts for an extended time period, the frame erasure concealment scheme should not continue the periodic extrapolation indefinitely, otherwise the extrapolated speech sounds like some sort of steady tone signal. In the preferred embodiment of the present invention, the module 210 starts waveform attenuation at the instant

when the frame erasure has lasted for 20 ms. From there, the envelope of the extrapolated waveform is attenuated linearly toward zero and the waveform magnitude reaches zero at 60 ms into the erasure of consecutive frames. After 60 ms, the output is completely muted. See FIG. 3 (a) for a waveform attenuation window that implements this attenuation strategy.

[0080] The preferred embodiment of the present invention is used with a noise feedback codec that has a frame size of 5 ms. In this case, the time interval between each adjacent pair of vertical lines in FIG. 3 (a) represent a frame.

[0081] Suppose a frame erasure lasts for 12 consecutive frames ($5 \times 12 = 60$ ms) or more, the easiest way to implement this waveform attenuation is to extrapolate speech for the first 12 erased frames, store the resulting 60 ms of waveform, and then apply the attenuation window in FIG. 3 (a). However, this simple approach requires extra delay to buffer up 60 ms of extrapolated speech.

[0082] To avoid any additional delay, the module 210 applies the waveform attenuation window frame-by-frame without any additional buffering. However, starting from the sixth consecutive erased frame (from 25 ms on in FIG. 3), the module 210 cannot directly apply the corresponding section of the window for that frame in FIG. 3 (a). A waveform discontinuity will occur at the frame boundary, because the corresponding section of the attenuation window starts from a value less than unity (7/8, 6/8, 5/8, etc.). This will cause a sudden decrease of waveform sample value at the beginning of the frame, and thus an audible waveform discontinuity.

[0083] To eliminate this problem, the present invention normalizes each 5 ms section of the attenuation window in FIG. 3 (a) by its starting value at the left edge. For example, for the sixth frame (25 ms to 30 ms), the window is from 7/8 to 6/8, and normalizing this section by 7/8 will give a window from 1 to $(6/8)/(7/8) = 6/7$. Similarly, for the seventh frame (30 ms to 35 ms), the window is from 6/8 to 5/8, and normalizing this section by 6/8 will give a window from 1 to $(5/8)/(6/8) = 5/6$. Such normalized attenuation window for each frame is shown in FIG. 3 (b).

[0084] Rather than storing every sample in the normalized attenuation window in FIG. 3 (b), the present invention simply stores the decrement between adjacent samples of the window for each of the eight window sections for fifth to twelfth frame. This decrement is the amount of total decline of the window function in each frame (1/8 for the fifth erased frame, 1/7 for the sixth erased frame, and so on), divided by N_f , the number of speech samples in a frame.

[0085] If the frame erasure has lasted for only 20 ms or less, the module 210 does not need to perform any waveform attenuation operation. If the frame erasure has lasted for more than 20 ms, then the module 210 applies the appropriate section of the normalized waveform attenuation window in FIG. 3 (b), depending on how many consecutive frames have been erased so far. For example, if the current frame is the sixth consecutive frame that is erased, then the module 210 applies the section of the window from 25 ms to 30 ms (with window function from 1 to 6/7). Since the normalized waveform attenuation window for each frame always starts with unity, the windowing operation will not cause any waveform discontinuity at the beginning of the frame.

[0086] The normalized window function is not stored; instead, it is calculated on the fly. Starting with a value of 1, the module 210 multiplies the first waveform sample of the current frame by 1, and then reduces the window function value by the decrement value calculated and stored beforehand, as mentioned above. It then multiplies the second waveform sample by the resulting decremented window function value. The window function value is again reduced by the decrement value, and the result is used to scale the third waveform sample of the frame. This process is repeated for all samples of the extrapolated waveform in the current frame.

[0087] The output of the module 210, that is, $sq'(n)$ for the current erased frame, is passed through the switch 200 and becomes the final output speech for the current erased frame. The current frame of $sq'(n)$ is passed to the module 201 to update the current frame portion of the $sq(n)$ speech buffer stored there. Let $sq'(n)$, $n = 1, 2, \dots, N_f$ be the output of the module 210 for the current erased frame, then the $sq(n)$ buffer of the module 201 is updated as:

$$sq(N + n) = sq'(n), n = 1, 2, \dots, N_f.$$

[0088] This signal is also passed to a module 211 to update the memory, or internal states, of the filters inside the decoder 100. Such a filter memory update is performed in order to ensure that the filter memory is consistent with the extrapolated speech waveform in the current erased frame. This is necessary for a smooth transition of speech waveform at the beginning of the next frame, if the next frame turns out to be a good frame. If the filter memory were frozen without such proper update, then generally there would be audible glitch or disturbance at the beginning of the next good frame.

[0089] If the short-term predictor is of order M , then the updated memory is simply the last M samples of the extrapolated speech signal for the current erased frame, but with the order reversed. Let $stsm(k)$ be the k -th memory value of the short-term synthesis filter, or the value stored in the delay line corresponding to the k -th short-term predictor coefficient a_k . Then, the memory of the short-term synthesis filter 190 is updated as

$$stsm(k) = sq(N + N_f + 1 - k), k = 1, 2, \dots, M.$$

[0090] To update the memory of the long-term synthesis filter 180 of FIG. 1, the module 211 extrapolates the long-term synthesis filter memory based on the first time lag $ppfe1$, using procedures similar to speech waveform extrapolation performed at the module 205. Let $ltsm(n)$ be the memory (or delay line) of the long-term synthesis filter 180, where the indexing convention is the same as that of $sq(n)$. The module 211 first extrapolates the first L samples of $ltsm(n)$ in the current frame

$$ltsm(n) = ptfe1 \times ltsm(n - ppfe1), \text{ for } n = N + 1, N + 2, \dots, N + L.$$

[0091] Next, this extrapolated filter memory is overlap-added with the ringing of the long-term synthesis filter calculated in the module 201:

$$ltsm(N + n) \leftarrow w_u(n)ltsm(N + n) + w_d(n)ltr(n), \text{ for } n = 1, 2, \dots, L.$$

[0092] After the first L samples of the current frame are extrapolated and overlap-added, the module 211 then extrapolates the remaining samples of the current frame. If $ppfe1 \geq N_f$, the extrapolation is performed as

$$ltsm(n) = ptfe1 \times ltsm(n - ppfe1), \text{ for } n = N + L + 1, N + L + 2, \dots, N + N_f.$$

If $ppfe1 < N_f$, then the extrapolation is performed as

$$ltsm(n) = ptfe1 \times ltsm(n - ppfe1), \text{ for } n = N + L + 1, N + L + 2, \dots, N + ppfe1,$$

and

$$ltsm(n) = ltsm(n - ppfe1), \text{ for } n = N + ppfe1 + 1, N + ppfe1 + 2, \dots, N + N_f.$$

[0093] If none of the side information speech parameters (LPC, pitch period, pitch taps, and excitation gain) is quantized using predictive coding, the operations of the module 211 are completed. If, on the other hand, predictive coding is used for side information, then the module 211 also needs to update the memory of the involved predictors to minimize the discontinuity of decoded speech parameters at the next good frame.

[0094] In the noise feedback codec that the preferred embodiment of the present invention is used in, moving-average (MA) predictive coding is used to quantize both the Line-Spectrum Pair (LSP) parameters and the excitation gain. The predictive coding schemes for these parameters work as follows. For each parameter, the long-term mean value of that parameter is calculated off-line and subtracted from the unquantized parameter value. The predicted value of the mean-removed parameter is then subtracted from this mean-removed parameter value. A quantizer quantizes the resulting prediction error. The output of the quantizer is used as the input to the MA predictor. The predicted parameter value and the long-term mean value are both added back to the quantizer output value to reconstruct a final quantized parameter value.

[0095] In an embodiment of the present invention, the modules 202 through 210 produce the extrapolated speech for the current erased frame. Theoretically, for the current frame, there is no need to extrapolate the side information speech parameters since the output speech waveform has already been generated. However, to ensure that the LSP and gain decoding operations will go smoothly at the next good frame, it is helpful to assume that these parameters are extrapolated from the last frame by simply copying the parameter values from the last frame, and then work "backward" from these extrapolated parameter values to update the predictor memory of the predictive quantizers for these parameters.

[0096] Using the principle outlined in the last paragraph, we can update the predictor memory in the predictive LSP quantizer, can be updated as follows. The predicted value for the k -th LSP parameter is calculated as the inner product of the predictor coefficient array and the predictor memory array for the k -th LSP parameter. This predicted value and the long-term mean value of the k -th LSP are subtracted from the k -th LSP parameter value at the last frame. The resulting value is used to update the newest memory location for the predictor of the k -th LSP parameter (after the original set of predictor memory is shifted by one memory location, as is well-known in the art). This procedure is repeated for all the LSP parameters (there are M of them).

[0097] If the frame erasure lasts only 20 ms or less, no waveform attenuation window is applied, and an assumption is made that the excitation gain of the current erased frame is the same as the excitation gain of the last frame. In this case, the memory update for the gain predictor is essentially the same as the memory update for the LSP predictors described above. Basically, the predicted value of log-gain is calculated (by calculating the inner product of the predictor coefficient array and the predictor memory array for the log-gain). This predicted log-gain and the long-term mean value of the log-gain are then subtracted from the log-gain value of the last frame. The resulting value is used to update the newest memory location for the log-gain predictor (after the original set of predictor memory is shifted by one memory location, as is well-known in the art).

[0098] If the frame erasure lasts more than 60 ms, the output speech is zeroed out, and the base-2 log-gain is assumed to be at an artificially set default silence level of 0. Again, the predicted log-gain and the long-term mean value of log-gain are subtracted from this default level of 0, and the resulting value is used to update the newest memory location for the log-gain predictor.

[0099] If the frame erasure lasts more than 20 ms but does not exceed 60 ms, then updating the predictor memory for the predictive gain quantizer may be challenging, because the extrapolated speech waveform is attenuated using the waveform attenuation window of FIG. 3. The log-gain predictor memory is updated based on the log-gain value of the waveform attenuation window in each frame.

[0100] To minimize the code size, for each of the frames from the fifth to the twelfth frames into frame erasure, a correction factor is calculated from the log-gain of the last frame based on the attenuation window of FIG. 3, and the correction factor is stored. The following algorithm calculates these 8 correction factors, or log-gain attenuation factors.

1. Initialize $lastlg = 0$. ($lastlg$ = last log-gain = log-gain of the last frame)

2. Initialize $j = 1$.

3. Calculate the normalized attenuation window array

$$w(n) = 1 - \frac{n-1}{(9-j)N_f}, n = 1, 2, \dots, N_f.$$

4. Calculate $lg = 2 \log_2 \left[\frac{1}{N_f} \sum_{n=1}^{N_f} w^2(n) \right]$

5. Calculate $lga(j) = lastlg - lg$

6. If $j < 8$, then set $lastlg = lg - 2 \log_2 \left(\frac{8-j}{9-j} \right)$

7. If $j = 8$, stop; otherwise, increment j by 1 (i.e. $j \leftarrow j + 1$), then go back to step 3.

[0101] Basically, the above algorithm calculates the base-2 log-gain value of the waveform attenuation window for a given frame, and then determines the difference between this value and a similarly calculated log-gain for the window

of the previous frame, compensated for the normalization of the start of the window to unity for each frame. The output of this algorithm is the array of log-gain attenuation factors $lga(j)$ for $j = 1, 2, \dots, 8$. Note that $lga(j)$ corresponds to the $(4+j)$ -th frame into frame erasure.

[0102] Once the $lga(j)$ array has been pre-calculated and stored, then the log-gain predictor memory update for frame erasure lasting 20 ms to 60 ms becomes straightforward. If the current erased frame is the j -th frame into frame erasure ($4 < j \leq 12$), $lga(j-4)$ is subtracted from the log-gain value of the last frame. From the result of this subtraction, the predicted log-gain and the long-term mean value of log-gain are further subtracted, and the resulting value is used to update the newest memory location for the log-gain predictor.

[0103] After the module 211 calculates all the updated filter memory values, the decoder 100 uses these values to update the memories and of its short-term synthesis filter 190, long-term synthesis filter 180, LSP predictor, and gain predictor, in preparation for the decoding of the next frame, assuming the next frame will be received intact.

[0104] The frame erasure concealment scheme described above can be used as is, and it will provide significant speech quality improvement compared with applying no concealment. So far, essentially all the frame erasure concealment operations are performed during erased frames. The present invention has an optional feature that improves speech quality by performing "filter memory correction" at the first received good frame after the erasure.

[0105] The short-term synthesis filter memory and the long-term synthesis filter memory are updated in the module 211 based on waveform extrapolation. After the frame erasure is over, at the first received good frame after the erasure, there will be a mismatch between such filter memory in the decoder and the corresponding filter memory in the encoder. Very often the difference is mainly due to the difference between the pitch period (or time lag) used for waveform extrapolation during erased frame and the pitch period transmitted by the encoder. Such filter memory mismatch often causes audible distortion even after the frame erasure is over.

[0106] During erased frames, the pitch period is typically held constant or nearly constant. If the pitch period is instantaneously quantized (i.e. without using inter-frame predictive coding), and if the frame erasure occurs in a voiced speech segment with a smooth pitch contour, then, linearly interpolating between the transmitted pitch periods of the last good frame before erasure and the first good frame after erasure often provides a better approximation of the transmitted pitch period contour than holding the pitch period constant during erased frames. Therefore, if the synthesis filter memory is re-calculated or corrected at the first good frame after erasure, based on linearly interpolated pitch period over the erased frames, better speech quality can often be obtained.

[0107] The long-term synthesis filter memory is corrected in the following way at the first good frame after the erasure. First, the received pitch period at the first good frame and the received pitch period at the last good frame before the erasure are used to perform linear interpolation of the pitch period over the erased frames. If an interpolated pitch period is not an integer, it is rounded off to the nearest integer. Next, starting with the first erased frame, and then going through all erased frames in sequence, the long-term synthesis filter memory is "re-extrapolated" frame-by-frame based on the linearly interpolated pitch period in each erased frame, until the end of the last erased frame is reached. For simplicity, a scaling factor of 1 may be used for the extrapolation of the long-term synthesis filter. After such re-extrapolation, the long-term synthesis filter memory is corrected.

[0108] The short-term synthesis filter memory may be corrected in a similar way, by re-extrapolating the speech waveform frame-by-frame, until the end of the last erased frame is reached. Then, the last M samples of the re-extrapolated speech waveform at the last erased frame, with the order reversed, will be the corrected short-term synthesis filter memory.

[0109] Another simpler way to correct the short-term synthesis filter memory is to estimate the waveform offset between the original extrapolated waveform and the re-extrapolated waveform, without doing the re-extrapolation. This method is described below. First, "project" the last speech sample of the last erased frame backward by $ppfe1$ samples, where $ppfe1$ is the original time lag used for extrapolation at that frame, and depending on which frame the newly projected sample lands, it is backward projected by the $ppfe1$ of that frame again. This process is continued until a newly projected sample lands on a good frame before the erasure. Then, using the linearly interpolated pitch period, a similar backward projection operation is performed until a newly projected sample lands on a good frame before the erasure. The distance between the two landing points in the good frame, obtained by the two backward projection operations above, is the waveform offset at the end of the last erased frame.

[0110] If this waveform offset indicates that the re-extrapolated speech waveform based on interpolated pitch period is delayed by X samples relative to the original extrapolated speech waveform at the end of the last erased frame, then the short-term synthesis filter memory can be corrected by taking the M consecutive samples of the original extrapolated speech waveform that are X samples away from the end of the last erased frame, and then reversing the order. If, on the other hand, the waveform offset calculated above indicates that the original extrapolated speech waveform is delayed by X samples relative to the re-extrapolated speech waveform (if such re-extrapolation were ever done), then the short-term synthesis filter memory correction would need to use certain speech samples that are not extrapolated yet. In this case, the original extrapolation for X more samples can be extended, and the last M samples can be taken with their order reversed. Alternatively, the system can move back one pitch cycle, and use the M consecutive samples (with order reversed) of the original extrapolated speech waveform that are $(ppfe1 - X)$ samples away from the end of the last erased

frame, where $ppfe1$ is the time lag used for original extrapolation of the last erased frame, and assuming $ppfe1 > X$.

[0111] One potential issue of such filter memory correction is that the re-extrapolation generally results in a waveform shift or offset; therefore, upon decoding the first good frame after an erasure, there may be a waveform discontinuity at the beginning of the frame. This problem can be eliminated if an output buffer is maintained, and the waveform shift is not immediately played out, but is slowly introduced, possibly over many frames, by inserting or deleting only one sample at a time over a long period of time. Another possibility is to use time scaling techniques, which are well known in the art, to speed up or slow down the speech slightly, until the waveform offset is eliminated. Either way, the waveform discontinuity can be avoided by smoothing out waveform offset over time.

[0112] If the additional complexity of such time scaling or gradual elimination of waveform offset is undesirable, a simpler, but sub-optimal approach is outlined below. First, before the synthesis filter memory is corrected, the first good frame after an erasure is decoded normally for the first Y samples of the speech. Next, the synthesis filter memory is corrected, and the entire first good frame after the erasure is decoded again. Overlap-add over the first Y samples is then used to provide a smooth transition between these two decoded speech signals in the current frame. This simple approach will smooth out waveform discontinuity if the waveform offset mentioned above is not excessive. If the waveform offset is too large, then this overlap-add method may not be able to eliminate the audible glitch at the beginning of the first good frame after an erasure. In this case, it is better not to correct the synthesis filter memory unless the time scaling or gradual elimination of waveform offset mentioned above can be used.

[0113] The following description of a general purpose computer system is provided for completeness. As stated above, the present invention can be implemented in hardware, or as a combination of software and hardware. Consequently, the invention may be implemented in the environment of a computer system or other processing system. An example of such a computer system 400 is shown in FIG. 4. In the present invention, all of the elements depicted in FIGs. 1 and 2, for example, can execute on one or more distinct computer systems 400, to implement the various methods of the present invention.

[0114] The computer system 400 includes one or more processors, such as a processor 404. The processor 404 can be a special purpose or a general purpose digital signal processor and it's connected to a communication infrastructure 406 (for example, a bus or network). Various software implementations are described in terms of this exemplary computer system. After reading this description, it will become apparent to a person skilled in the relevant art how to implement the invention using other computer systems and/or computer architectures.

[0115] The computer system 400 also includes a main memory 408, preferably random access memory (RAM), and may also include a secondary memory 410. The secondary memory 410 may include, for example, a hard disk drive 412 and/or a removable storage drive 414, representing a floppy disk drive, a magnetic tape drive, an optical disk drive, etc. The removable storage drive 414 reads from and/or writes to a removable storage unit 418 in a well known manner. The removable storage unit 418, represents a floppy disk, magnetic tape, optical disk, etc. which is read by and written to by removable storage drive 414. As will be appreciated, the removable storage unit 418 includes a computer usable storage medium having stored therein computer software and/or data.

[0116] In alternative implementations, the secondary memory 410 may include other similar means for allowing computer programs or other instructions to be loaded into the computer system 400. Such means may include, for example, a removable storage unit 422 and an interface 420. Examples of such means may include a program cartridge and cartridge interface (such as that found in video game devices), a removable memory chip (such as an EPROM, or PROM) and associated socket, and the other removable storage units 422 and the interfaces 420 which allow software and data to be transferred from the removable storage unit 422 to the computer system 400.

[0117] The computer system 400 may also include a communications interface 424. The communications interface 424 allows software and data to be transferred between the computer system 400 and external devices. Examples of the communications interface 424 may include a modem, a network interface (such as an Ethernet card), a communications port, a PCMCIA slot and card, etc. Software and data transferred via the communications interface 424 are in the form of signals 428 which may be electronic, electromagnetic, optical or other signals capable of being received by the communications interface 424. These signals 428 are provided to the communications interface 424 via a communications path 426. The communications path 426 carries the signals 428 and may be implemented using wire or cable, fiber optics, a phone line, a cellular phone link, an RF link and other communications channels.

[0118] In the present application, the terms "computer readable medium" and "computer usable medium" are used to generally refer to media such as the removable storage drive 414, a hard disk installed in the hard disk drive 412, and the signals 428. These computer program products are means for providing software to the computer system 400.

[0119] Computer programs (also called computer control logic) are stored in the main memory 408 and/or the secondary memory 410. Computer programs may also be received via the communications interface 424. Such computer programs, when executed, enable the computer system 400 to implement the present invention as discussed herein.

[0120] In particular, the computer programs, when executed, enable the processor 404 to implement the processes of the present invention. Accordingly, such computer programs represent controllers of the computer system 400. By way of example, in the embodiments of the invention, the processes/methods performed by signal processing blocks of

encoders and/or decoders can be performed by computer control logic. Where the invention is implemented using software, the software may be stored in a computer program product and loaded into the computer system 400 using the removable storage drive 414, the hard drive 412 or the communications interface 424.

[0121] In another embodiment, features of the invention are implemented primarily in hardware using, for example, hardware components such as Application Specific Integrated Circuits (ASICs) and gate arrays. Implementation of a hardware state machine so as to perform the functions described herein will also be apparent to persons skilled in the relevant art(s).

[0122] The foregoing description of the preferred embodiments provide an illustration and description, but is not intended to be exhaustive or to limit the invention to the precise form disclosed. Modifications and variations are possible consistent with the above teachings, or may be acquired from practice of the invention. The scope of the invention is defined by the appended claims.

Claims

1. A method of synthesizing a number of corrupted frames output from a decoder (100) including one or more predictive filters (180, 190), the corrupted frames being representative of one segment of a decoded signal, $sq(n)$, output from the decoder (100), the method comprising:

determining (202) a preliminary time lag, $ppfe1$, based upon examining a predetermined number, K , of samples of another segment of the decoded signal;
wherein the examined number, K , of samples is selected from within a number, N , of stored samples;
determining (202) a scaling factor, $ptfe1$, associated with the examined number, K , of samples when the preliminary time lag, $ppfe1$, is determined; and
extrapolating (205) one or more replacement frames based upon the preliminary time lag, $ppfe1$, and the scaling factor, $ptfe1$;

the method being **characterized in that**

the scaling factor, $ptfe1$, is determined in accordance with the expression:

$$ptfe1 = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|} ;$$

wherein $c(ppfe1)$ is a correlation value dependent on $ppfe1$.

2. The method of claim 1, further comprising updating internal states of the filters based upon the extrapolating.

3. The method of claim 2, wherein correlation values, $c(j)$, associated with candidate preliminary time lags, j , are determined in accordance with the expression:

$$c(j) = \sum_{n=N-K+1}^N sq(n)sq(n-j) ;$$

and

wherein the preliminary time lag, $ppfe1$, is chosen from within the candidate preliminary time lags, j , and maximizes the expression:

$$nc(j) = \frac{\left(\sum_{n=N-K+1}^N sq(n)sq(n-j) \right)^2}{\sum_{n=N-K+1}^N sq^2(n-j)}$$

and the correlation value $c(ppfe1)$ is equal to the expression $c(j)$ taken at $j=ppfe1$.

4. The method of claim 1, further comprising:

correcting internal states of the filters when a first good frame is received, the first good frame being received after the number of corrupted frames.

5. The method of claim 4, wherein the updating includes updating short-term and long-term synthesis filters associated with the one or more predictive filters (180, 190).

6. An apparatus for synthesizing a number of corrupted frames output from a decoder (100) including one or more predictive filters (180, 190), the corrupted frames being representative of one segment of a decoded signal, $sq(n)$, output from the decoder (100), the apparatus comprising:

means for determining a preliminary time lag, $ppfe1$, based upon examining a pre-determined number, K , of samples of another segment of the decoded signal;

wherein the examined number, K , of samples is selected from within a number, N , of stored samples;

means for determining a scaling factor, $ptfe1$, associated with the examined number, K , of samples when the preliminary time lag, $ppfe1$, is determined; and

means for extrapolating one or more replacement frames based upon the preliminary time lag, $ppfe1$, and the scaling factor, $ptfe1$;

the apparatus being **characterized in that**

the means for determining the scaling factor, $ptfe1$, are configured to determine the scaling factor, $ptfe1$, in accordance with the expression:

$$ptfe1 = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|} ;$$

wherein $c(ppfe1)$ is a correlation value dependent on $ppfe1$.

7. The apparatus of claim 6, further comprising means for updating internal states of the filters based upon the extrapolating.

8. The apparatus of claim 6, further comprising:

means for correcting internal states of the filters when a first good frame is received, the first good frame being received after the number of corrupted frames.

9. A computer readable medium carrying one or more sequences of one or more instructions for execution by one or more processors to perform a method of synthesizing a number of corrupted frames output from a decoder (100) including one or more predictive filters (180, 190), the corrupted frames being representative of one segment of a decoded signal, $sq(n)$, output from the decoder (100), the instructions when executed by the one or more processors, cause the one or more processors to perform all the steps of:

determining (202) a preliminary time lag, ppfe1, based upon examining a predetermined number, K, of samples of another segment of the decoded signal;
 wherein the examined number, K, of samples is selected from within a number, N, of stored samples;
 determining (202) a scaling factor, ptfe1, associated with the examined number, K, of samples when the preliminary time lag, ppfe1, is determined; and
 extrapolating (205) one or more replacement frames based upon the preliminary time lag, ppfe1, and the scaling factor, ptfe1;

the computer readable medium being **characterized in that**
 the scaling factor, ptfe1, is determined in accordance with the expression:

$$ptfe1 = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|} ;$$

wherein c(ppfe1) is a correlation value dependent on ppfe1.

10. The computer readable medium of claim 9, wherein the one or more instructions, when executed by the one or more processors, cause the one or more processors to further perform the step of:

correcting internal states of the filters when a first good frame is received, the first good frame being received after the number of corrupted frames.

Patentansprüche

1. Verfahren zum Synthetisieren einer Anzahl von beschädigten Rahmen, die von einem Decodierer (100) ausgegeben werden, der ein oder mehrere Prädiktionsfilter (180, 190) umfasst, wobei die beschädigten Rahmen repräsentativ für ein Segment eines decodierten Signals sq(n) sind, das von dem Decodierer (100) ausgegeben wird, wobei das Verfahren Folgendes umfasst:

Bestimmen (202) einer vorläufigen Verzögerungszeit ppfe1 auf der Grundlage der Überprüfung einer vorbestimmten Anzahl K von Abtastwerten eines anderen Segments des decodierten Signals;
 wobei die überprüfte Anzahl K von Abtastwerten aus einer Anzahl N von gespeicherten Abtastwerten ausgewählt wird;

Bestimmen (202) eines Skalierungsfaktors ptfe1, der mit der überprüften Anzahl K von Abtastwerten assoziiert ist, wenn die vorläufige Verzögerungszeit ppfe1 bestimmt wird; und
 Extrapolieren (205) eines oder mehrerer Ersatzrahmen auf der Grundlage der vorläufigen Verzögerungszeit ppfe1 und des Skalierungsfaktors ptfe1;

wobei das Verfahren **dadurch gekennzeichnet ist, dass**
 der Skalierungsfaktor ptfe1 gemäß dem folgenden Ausdruck bestimmt wird:

$$ptfe1 = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|} ;$$

wobei c(ppfe1) ein Korrelationswert ist, der von ppfe1 abhängt.

2. Verfahren nach Anspruch 1, das des Weiteren das Aktualisieren von internen Zuständen der Filter auf der Grundlage der Extrapolation umfasst.

3. Verfahren nach Anspruch 2, wobei die Korrelationswerte $c(j)$, die mit den vorläufigen Verzögerungszeit-Kandidaten j assoziiert sind, gemäß dem folgenden Ausdruck bestimmt werden:

$$c(j) = \sum_{n=N-K+1}^N sq(n)sq(n-j) ;$$

und

wobei die vorläufige Verzögerungszeit $ppfe1$ aus den vorläufigen Verzögerungszeit-Kandidaten j ausgewählt wird und den folgenden Ausdruck maximiert:

$$nc(j) = \frac{\left(\sum_{n=N-K+1}^N sq(n)sq(n-j) \right)^2}{\sum_{n=N-K+1}^N sq^2(n-j)}$$

und der Korrelationswert $c(ppfe1)$ gleich dem Ausdruck $c(j)$ ist, der bei $j = ppfe1$ genommen wird.

4. Verfahren nach Anspruch 1, das des Weiteren Folgendes umfasst:

Korrigieren interner Zustände der Filter, wenn ein erster guter Rahmen empfangen wird, wobei der erste gute Rahmen nach der Anzahl von beschädigten Rahmen empfangen wird.

5. Verfahren nach Anspruch 4, wobei das Aktualisieren das Aktualisieren von Kurzzeit- und Langzeit-Synthesefiltern umfasst, die mit dem einen oder den mehreren Prädiktionsfiltern (180, 190) assoziiert sind.

6. Vorrichtung zum Synthetisieren einer Anzahl von beschädigten Rahmen, die von einem Decodierer (100) ausgegeben werden, der ein oder mehrere Prädiktionsfilter (180, 190) umfasst, wobei die beschädigten Rahmen repräsentativ für ein Segment eines decodierten Signals $sq(n)$ sind, das von dem Decodierer (100) ausgegeben wird, wobei die Vorrichtung Folgendes umfasst:

Einrichtungen zum Bestimmen einer vorläufigen Verzögerungszeit $ppfe1$ auf der Grundlage der Überprüfung einer vorbestimmten Anzahl K von Abtastwerten eines anderen Segments des decodierten Signals; wobei die überprüfte Anzahl K von Abtastwerten aus einer Anzahl N von gespeicherten Abtastwerten ausgewählt wird;

Einrichtungen zum Bestimmen eines Skalierungsfaktors $ptfe1$, der mit der überprüften Anzahl K von Abtastwerten assoziiert ist, wenn die vorläufige Verzögerungszeit $ppfe1$ bestimmt wird; und

Einrichtungen zum Extrapolieren eines oder mehrerer Ersatzrahmen auf der Grundlage der vorläufigen Verzögerungszeit $ppfe1$ und des Skalierungsfaktors $ptfe1$;

wobei die Vorrichtung **dadurch gekennzeichnet ist, dass**

die Einrichtungen zum Bestimmen des Skalierungsfaktors $ptfe1$ so konfiguriert sind, dass sie den Skalierungsfaktor $ptfe1$ gemäß folgendem Ausdruck bestimmen:

$$ptfe1 = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|}$$

wobei $c(ppfe1)$ ein Korrelationswert ist, der von $ppfe1$ abhängt.

7. Vorrichtung nach Anspruch 6, die des Weiteren Einrichtungen zum Aktualisieren von internen Zuständen der Filter auf der Grundlage der Extrapolation umfasst.

8. Vorrichtung nach Anspruch 6, die des Weiteren Folgendes umfasst:

Einrichtungen zum Korrigieren interner Zustände der Filter, wenn ein erster guter Rahmen empfangen wird, wobei der erste gute Rahmen nach der Anzahl von beschädigten Rahmen empfangen wird.

9. Computerlesbares Medium, das eine oder mehrere Sequenzen einer oder mehrerer Anweisungen zur Ausführung durch einen oder mehrere Prozessoren enthält, um ein Verfahren zum Synthetisieren einer Anzahl von beschädigten Rahmen durchzuführen, die von einem Decodierer (100) ausgegeben werden, der ein oder mehrere Prädiktionsfilter (180, 190) umfasst, wobei die beschädigten Rahmen repräsentativ für ein Segment eines decodierten Signals $sq(n)$ sind, das von dem Decodierer (100) ausgegeben wird, wobei die Anweisungen dann, wenn sie von dem einen oder den mehreren Prozessoren ausgeführt werden, bewirken, dass der eine oder die mehreren Prozessoren alle folgenden Schritte durchführen:

Bestimmen (202) einer vorläufigen Verzögerungszeit $ppfe1$ auf der Grundlage der Überprüfung einer vorbestimmten Anzahl K von Abtastwerten eines anderen Segments des decodierten Signals; wobei die überprüfte Anzahl K von Abtastwerten aus einer Anzahl N von gespeicherten Abtastwerten ausgewählt wird;

Bestimmen (202) eines Skalierungsfaktors $ptfe1$, der mit der überprüften Anzahl K von Abtastwerten assoziiert ist, wenn die vorläufige Verzögerungszeit $ppfe1$ bestimmt wird; und

Extrapolieren (205) eines oder mehrerer Ersatzrahmen auf der Grundlage der vorläufigen Verzögerungszeit $ppfe1$ und des Skalierungsfaktors $ptfe1$;

wobei das computerlesbare Medium **dadurch gekennzeichnet ist, dass** der Skalierungsfaktor $ptfe1$ gemäß dem folgenden Ausdruck bestimmt wird:

$$ptfe1 = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|};$$

wobei $c(ppfe1)$ ein Korrelationswert ist, der von $ppfe1$ abhängt.

10. Computerlesbares Medium nach Anspruch 9, wobei die eine oder mehreren Anweisungen dann, wenn sie von dem einen oder den mehreren Prozessoren ausgeführt werden, bewirken, dass der eine oder die mehreren Prozessoren des Weiteren den folgenden Schritt durchführen:

Korrigieren interner Zustände der Filter, wenn ein erster guter Rahmen empfangen wird, wobei der erste gute Rahmen nach der Anzahl von beschädigten Rahmen empfangen wird.

Revendications

1. Procédé de synthèse d'un nombre de trames altérées sorties d'un décodeur (100) comprenant un ou plusieurs filtres prédictifs (180, 190), les trames altérées étant représentatives d'un segment de signal décodé $sq(n)$, sorti du décodeur (100), le procédé comprenant :

la détermination (202) d'un retard temporel préliminaire, $ppfe1$, sur la base de l'examen d'un nombre prédéterminé, K , d'échantillons d'un autre segment du signal décodé ;

dans lequel où le nombre, K , d'échantillons examinés est choisi parmi un nombre, N , d'échantillons stockés ; la détermination (202) d'un facteur d'échelle, $ptfe1$, associé au nombre, K , d'échantillons examinés lorsque le retard temporel préliminaire, $ppfe1$, est déterminé ; et

l'extrapolation (205) d'une ou plusieurs trames de remplacement sur la base du retard temporel préliminaire, $ppfe1$, et du facteur d'échelle, $ptfe1$;

le procédé étant **caractérisé en ce que**

le facteur d'échelle, $ptfe1$, est déterminé selon l'expression :

$$ptfe1 = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|} ;$$

dans laquelle $c(ppfe1)$ est une valeur de corrélation qui est fonction de $ppfe1$.

2. Procédé selon la revendication 1, comprenant en outre des états internes de mise à jour des filtres sur la base de l'extrapolation.
3. Procédé selon la revendication 2, dans lequel les valeurs de corrélation $c(j)$, associées avec des retards temporels préliminaires candidats, j , sont déterminées selon l'expression :

$$c(j) = \sum_{n=N-K+1}^N sq(n)sq(n-j) ;$$

et

dans laquelle le retard temporel préliminaire, $ppfe1$, est choisi parmi les retards temporels préliminaires candidats, j , et maximise l'expression :

$$nc(j) = \frac{\left(\sum_{n=N-K+1}^N sq(n)sq(n-j) \right)^2}{\sum_{n=N-K+1}^N sq^2(n-j)}$$

et la valeur de corrélation $c(ppfe1)$ est égale à l'expression $c(j)$ prise à $j = ppfe1$.

4. Procédé selon la revendication 1, comprenant en outre :

la correction des états internes des filtres lorsqu'une première bonne trame est reçue, la première bonne trame étant reçue après le nombre de trames altérées.

5. Procédé selon la revendication 4, dans lequel la mise à jour comprend la mise à jour de filtres de synthèse à court terme et à long terme associés avec le ou les filtres prédictifs (180, 190).
6. Appareil de synthèse d'un nombre de trames altérées sorties d'un décodeur (100) comprenant un ou plusieurs filtres prédictifs (180, 190), les trames altérées étant représentatives d'un segment de signal décodé $sq(n)$, sorti par le décodeur (100), l'appareil comprenant :

un moyen de détermination d'un retard temporel préliminaire, $ppfe1$, basé sur l'examen d'un nombre prédéterminé, K , d'échantillons d'un autre segment du signal décodé ;
 dans lequel le nombre, K , d'échantillons examinés est choisi parmi un nombre, N , d'échantillons stockés ;
 un moyen de détermination d'un facteur d'échelle, $ptfe1$, associé au nombre, K , d'échantillons examinés lorsque le retard temporel préliminaire, $ppfe1$, est déterminé ; et
 un moyen d'extrapolation d'une ou plusieurs trames de remplacement basé sur le retard temporel préliminaire, $ppfe1$, et le facteur d'échelle, $ptfe1$;
 l'appareil étant **caractérisé en ce que**
 le moyen de détermination du facteur d'échelle, $ptfe1$, est configuré pour déterminer le facteur d'échelle, $ptfe$,

selon l'expression :

$$ptfel = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|} ;$$

dans laquelle $c(ppfe1)$ est une valeur de corrélation qui est fonction de $ppfe1$.

7. Appareil selon la revendication 6, comprenant en outre un moyen pour mettre à jour les états internes des filtres sur la base de l'extrapolation.

8. Appareil selon la revendication 6, comprenant en outre :

un moyen de correction des états internes des filtres lorsqu'une première bonne trame est reçue, la première bonne trame étant reçue après le nombre de trames altérées.

9. Support lisible par ordinateur portant une ou plusieurs séquences d'une ou plusieurs instructions pour l'exécution par un ou plusieurs processeurs afin d'exécuter un procédé de synthétisation d'un nombre de trames altérées sorties d'un décodeur (100) comprenant un ou plusieurs filtres prédictifs (180, 190), les trames altérées étant représentatives d'un segment de signal décodé $sq(n)$, sorti par le décodeur (100), les instructions lorsqu'elles sont exécutées par le ou les processeurs amènent le ou les processeurs à exécuter toutes les étapes suivantes :

déterminer (202) un retard temporel préliminaire, $ppfe1$, basé sur l'examen d'un nombre prédéterminé, K , d'échantillons d'un autre segment du signal décodé ;

dans lequel le nombre, K , d'échantillons examinés est choisi parmi un nombre, N , d'échantillons stockés ;

déterminer (202) un facteur d'échelle, $ptfe1$, associé au nombre, K , d'échantillons examinés lorsque le retard temporel préliminaire, $ppfe1$, est déterminé ; et

extrapoler (205) une ou plusieurs trames de remplacement sur la base du retard temporel préliminaire, $ppfe1$, et du facteur d'échelle, $ptfe1$;

le support lisible par ordinateur étant **caractérisé en ce que**

le facteur d'échelle, $ptfe1$, est déterminé selon l'expression :

$$ptfel = \text{sign}[c(ppfe1)] \times \frac{\sum_{n=N-K+1}^N |sq(n)|}{\sum_{n=N-K+1}^N |sq(n - ppfe1)|} ;$$

dans laquelle $c(ppfe1)$ est une valeur de corrélation qui est fonction de $ppfe1$.

10. Support lisible par ordinateur selon la revendication 9, dans lequel la ou les instructions, lorsqu'elles sont exécutées par le ou les processeurs, amènent le ou les processeurs à exécuter en outre l'étape suivante :

corriger les états internes des filtres lorsqu'une première bonne trame est reçue, la première bonne trame étant reçue après le nombre de trames altérées.

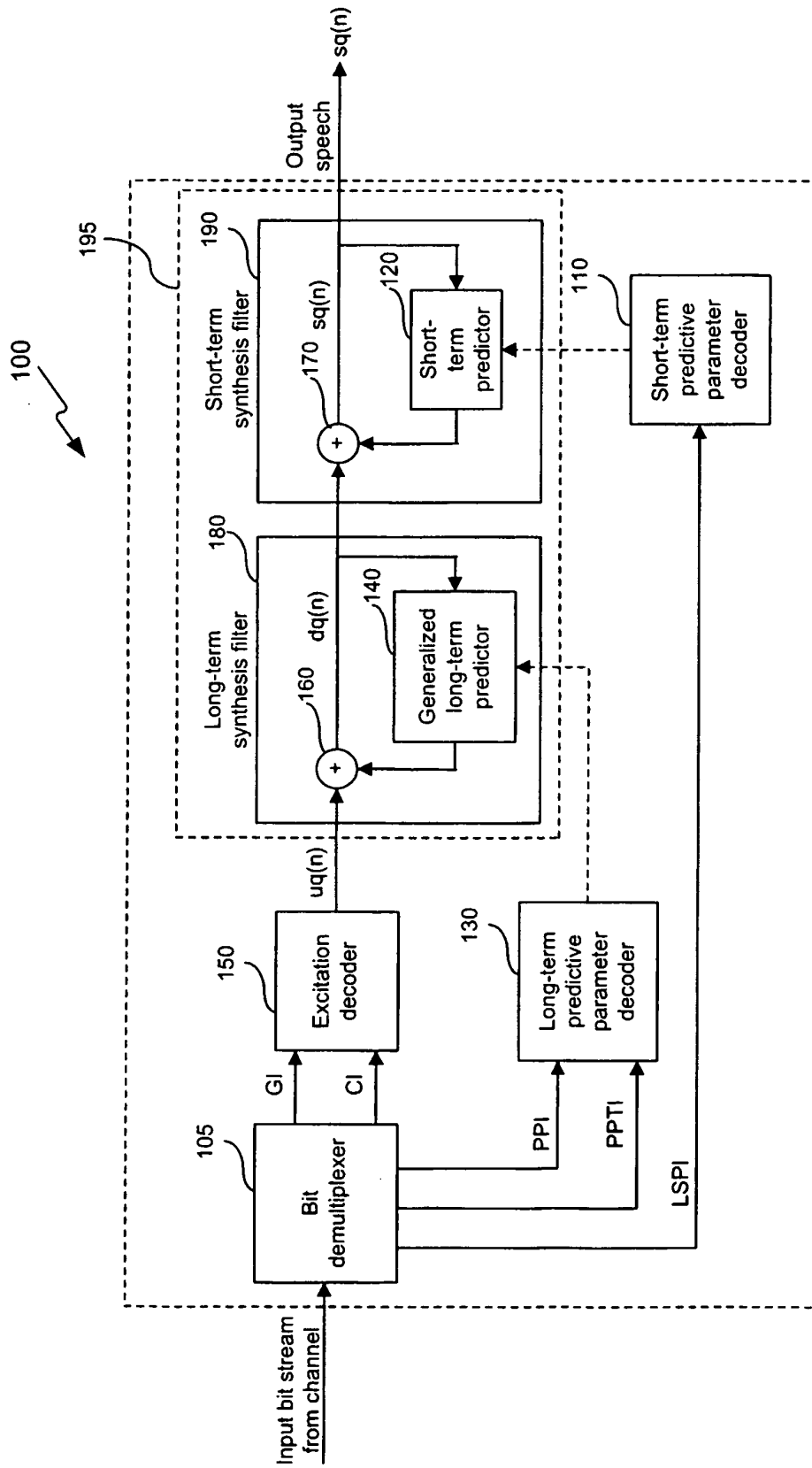


FIG. 1
(conventional)

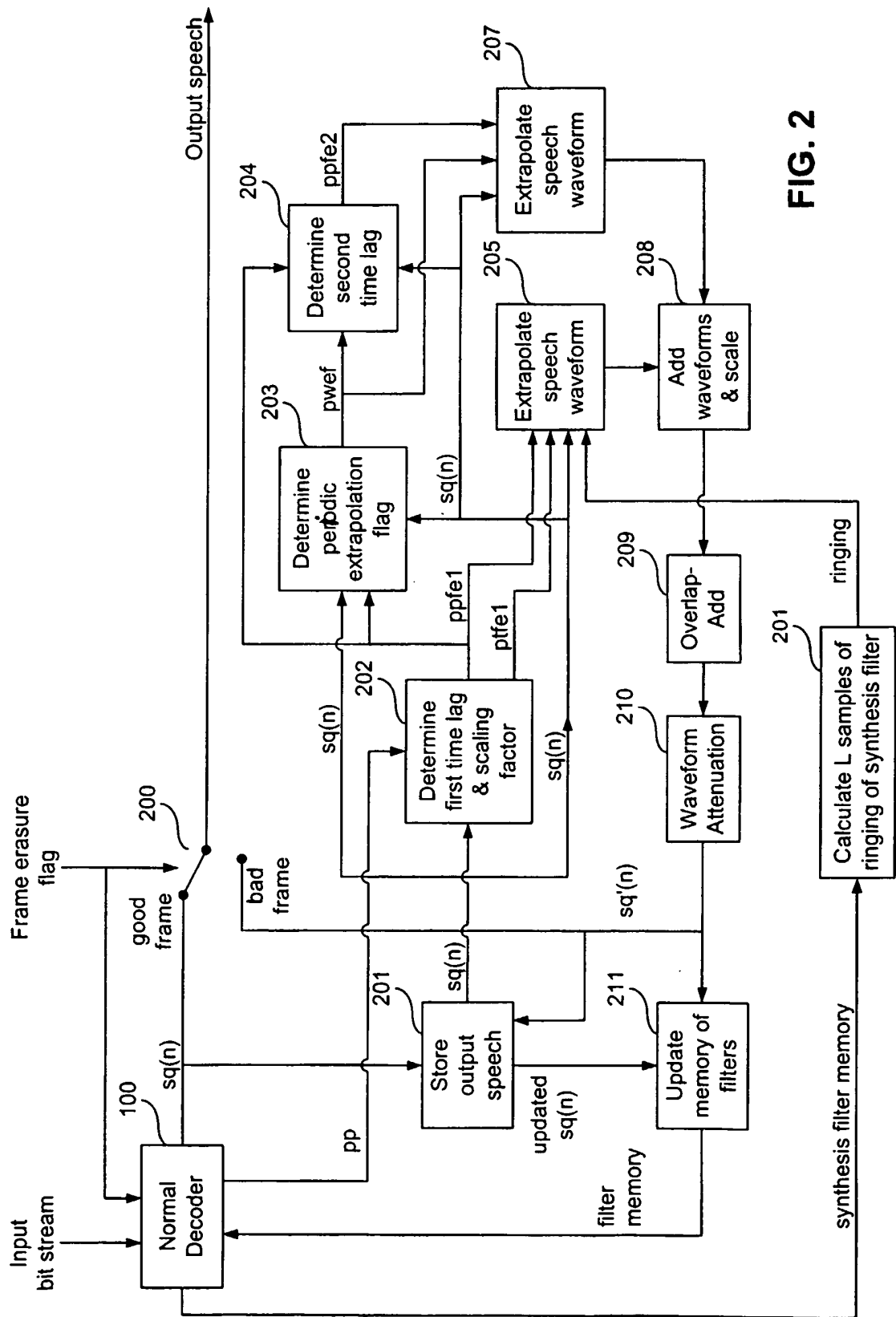


FIG. 2

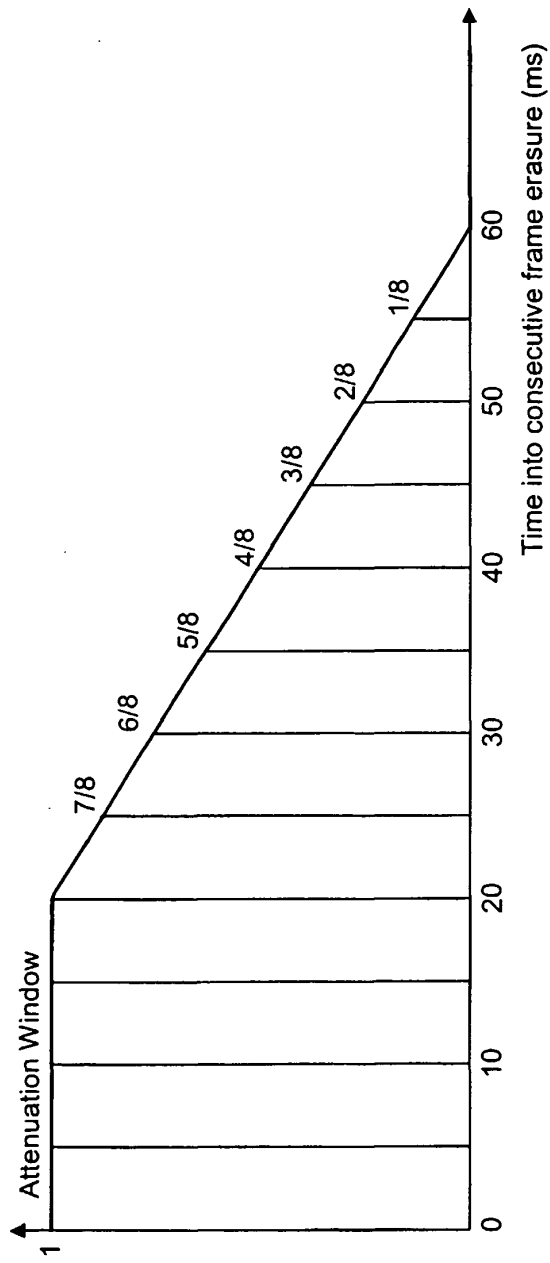


FIG. 3A

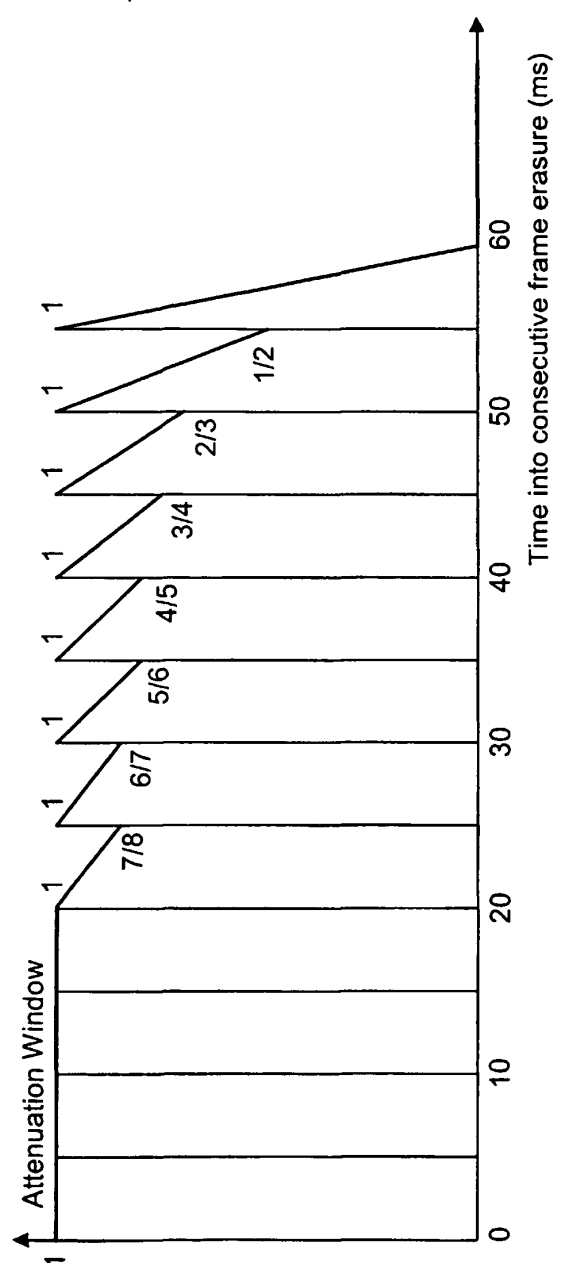


FIG. 3B

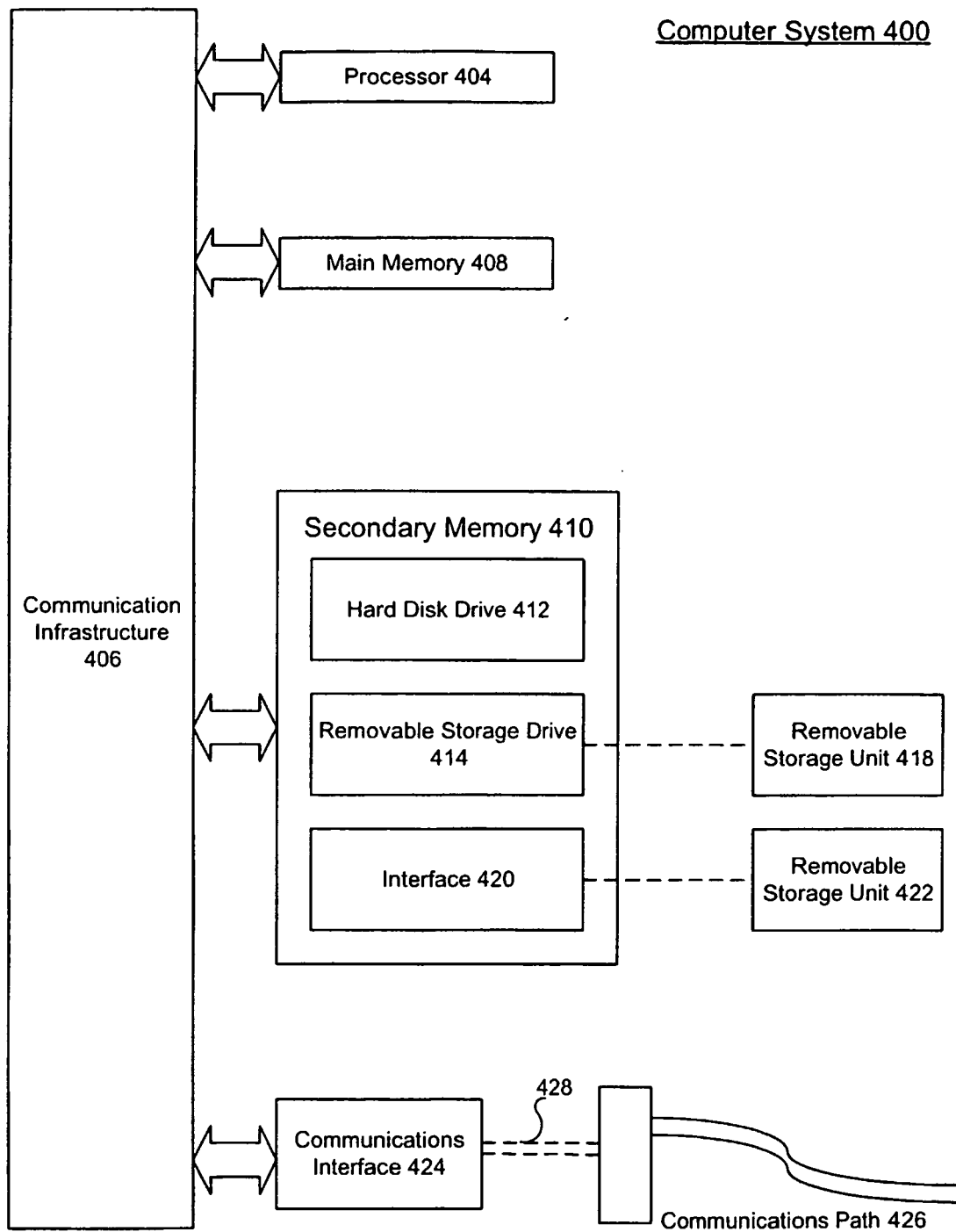


FIG. 4

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 5615298 A [0007]
- EP 0747882 A2 [0009]
- US 31279401 P [0064]
- US 34437802 P [0064]

Non-patent literature cited in the description

- **GOODMAN et al.** Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications. *IEEE Transaction on Acoustics, Speech and Signal Processing*, December 1986, 1440-1448 [0003]