

(21) Application No 9805682.3

(22) Date of Filing 16.03.1998

(30) Priority Data

(31) 08818137 (32) 14.03.1997 (33) US

(71) Applicant(s)

Digital Voice Systems Inc
(Incorporated in USA - Massachusetts)
One Van De Graaff Drive, Burlington, MA 01803,
United States of America

(72) Inventor(s)

John Clark Hardwick

(74) Agent and/or Address for Service

Lloyd Wise, Tregear & Co
Commonwealth House, 1-19 New Oxford Street,
LONDON, WC1A 1LW, United Kingdom

(51) INT CL⁶

G10L 9/14 7/06 , H03M 7/30 13/00

(52) UK CL (Edition P)

H4P PDCFP PEP

H4R RPV

U1S S2212

(56) Documents Cited

EP 0577488 A1 EP 0422232 A1 WO 94/12972 A1

WO 94/12932 A1 WO 92/10830 A1

(58) Field of Search

UK CL (Edition P) H4P PDCFP PDCFT PEP , H4R RPV

RPVA RPVS

INT CL⁶ G10L 7/06 9/04 9/14 9/18 , H03M 7/30 13/00 ,

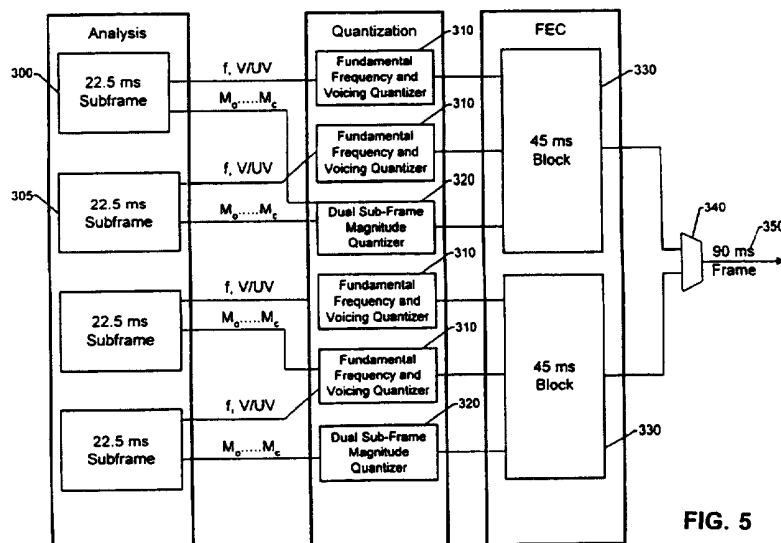
H04L 1/00

Online : WPI

(54) Abstract Title

Dual subframe quantisation of spectral magnitudes

(57) A speech signal is digitized into digital speech samples that are then divided into subframes 300,305. Model parameters that include a set of spectral magnitude parameters M_0, \dots, M_E that represent spectral information for the subframe are estimated for each subframe. Two consecutive subframes from the sequence of subframes are combined into a block and their spectral magnitude parameters are jointly quantized 320. The joint quantization includes forming predicted spectral magnitude parameters from the quantized spectral magnitude parameters from the previous block, computing the residual parameters as the difference between the spectral magnitude parameters and the predicted spectral magnitude parameters, combining the residual parameters from both of the subframes within the block, and using vector quantizers to quantize the combined residual parameters into a set of encoded spectral bits. Redundant error control bits are added to the encoded spectral bits from each block to protect the encoded spectral bits within the block from bit errors. The added redundant error control bits and encoded spectral bits from two consecutive blocks 330 are combined 340 into a 90 millisecond frame 350 of bits for transmission across a satellite communication channel.



The specification as filed includes Appendices A to O which are not reproduced here; they may be inspected in accordance with Section 118 of the Patents Act 1977.

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

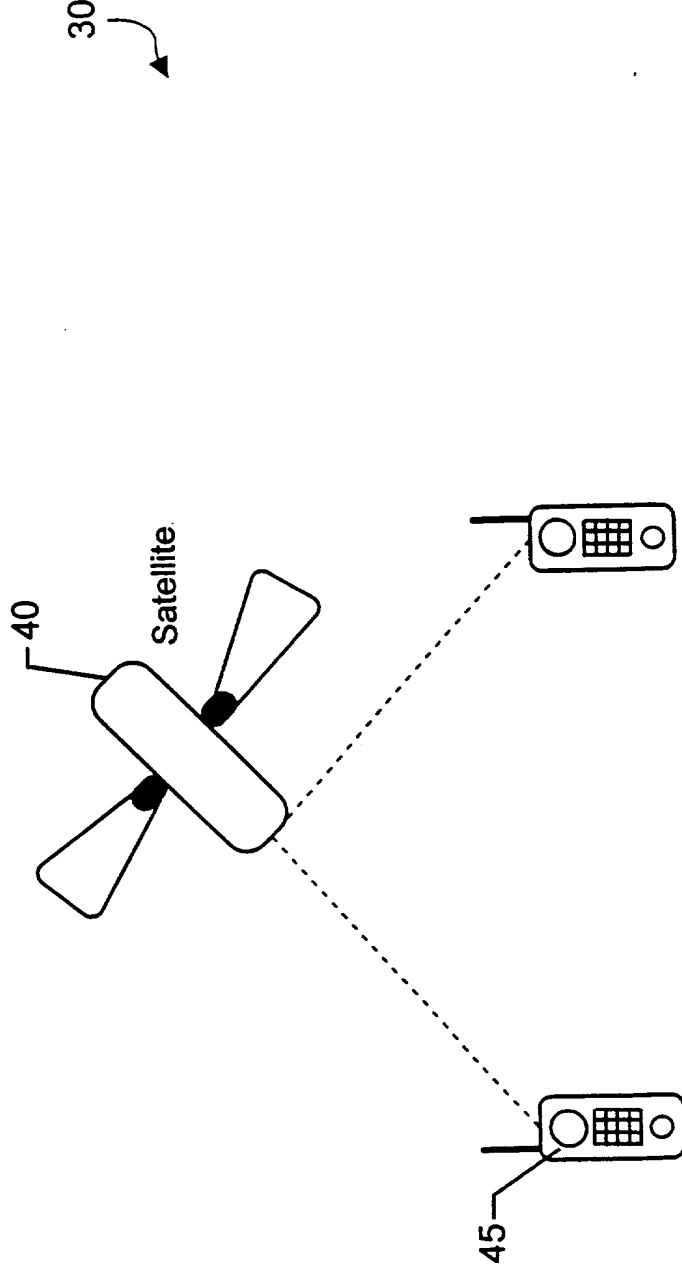


FIG. 1

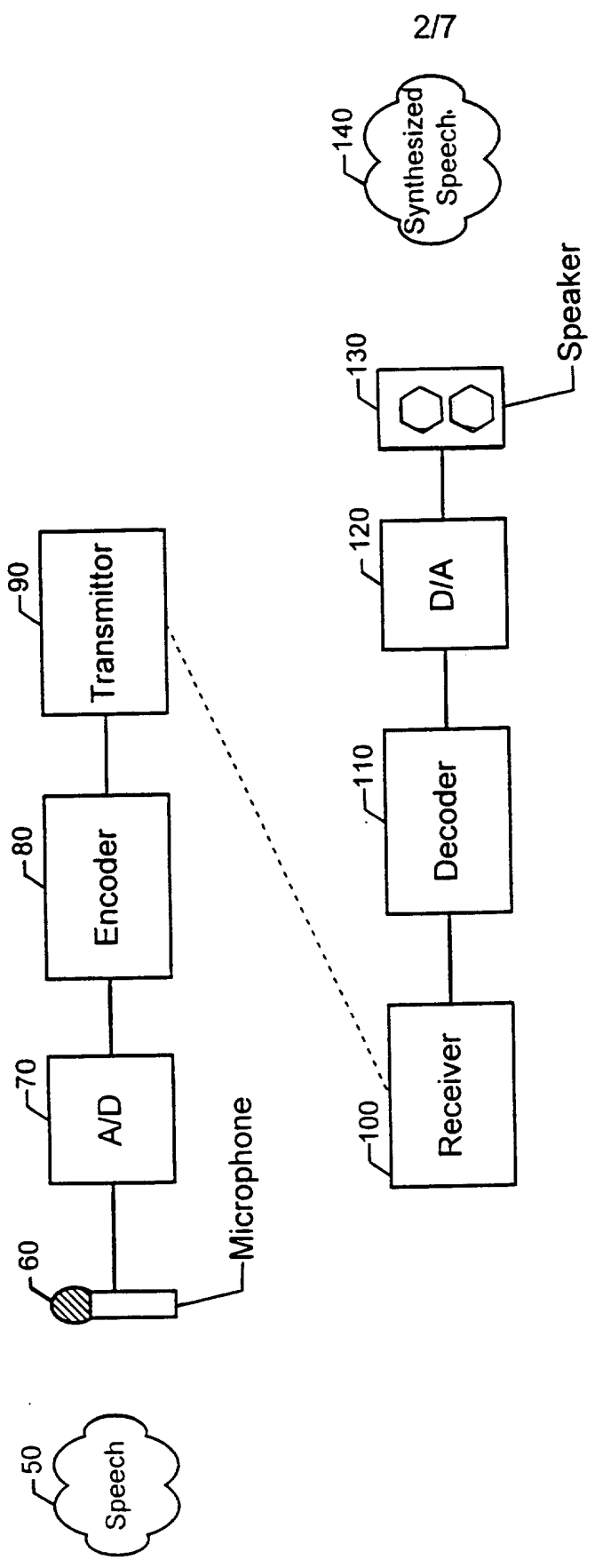


FIG. 2

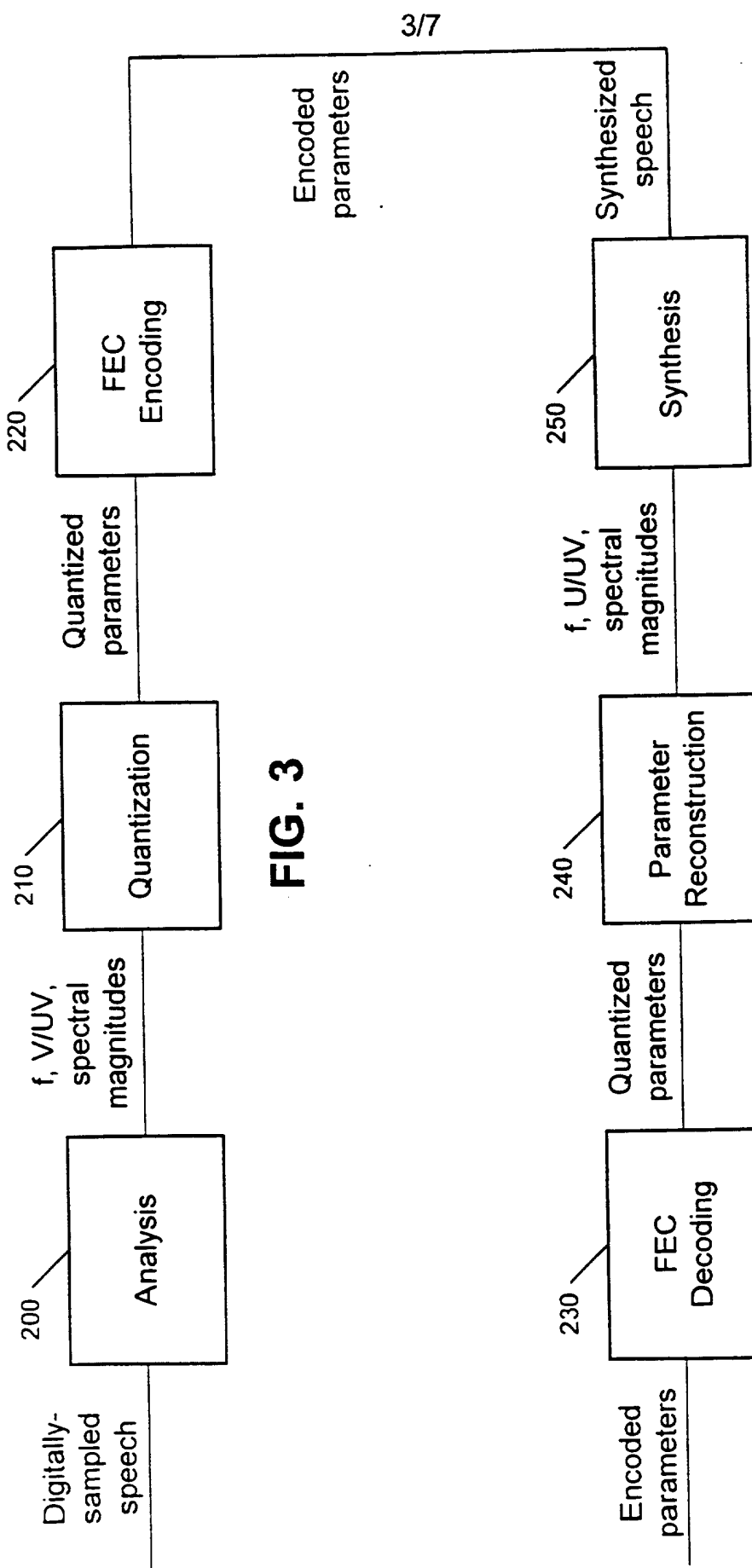


FIG. 3

FIG. 4

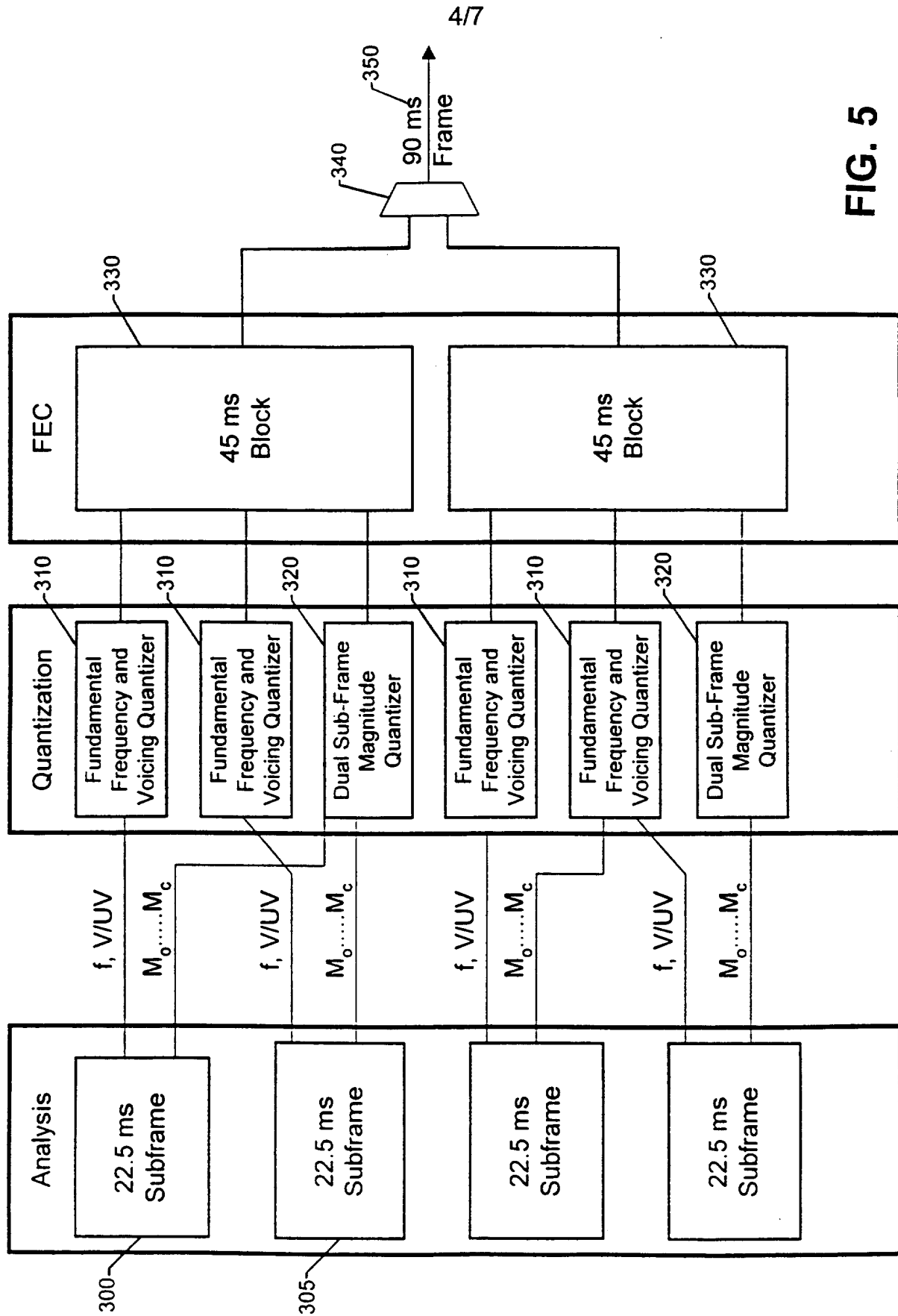


FIG. 5

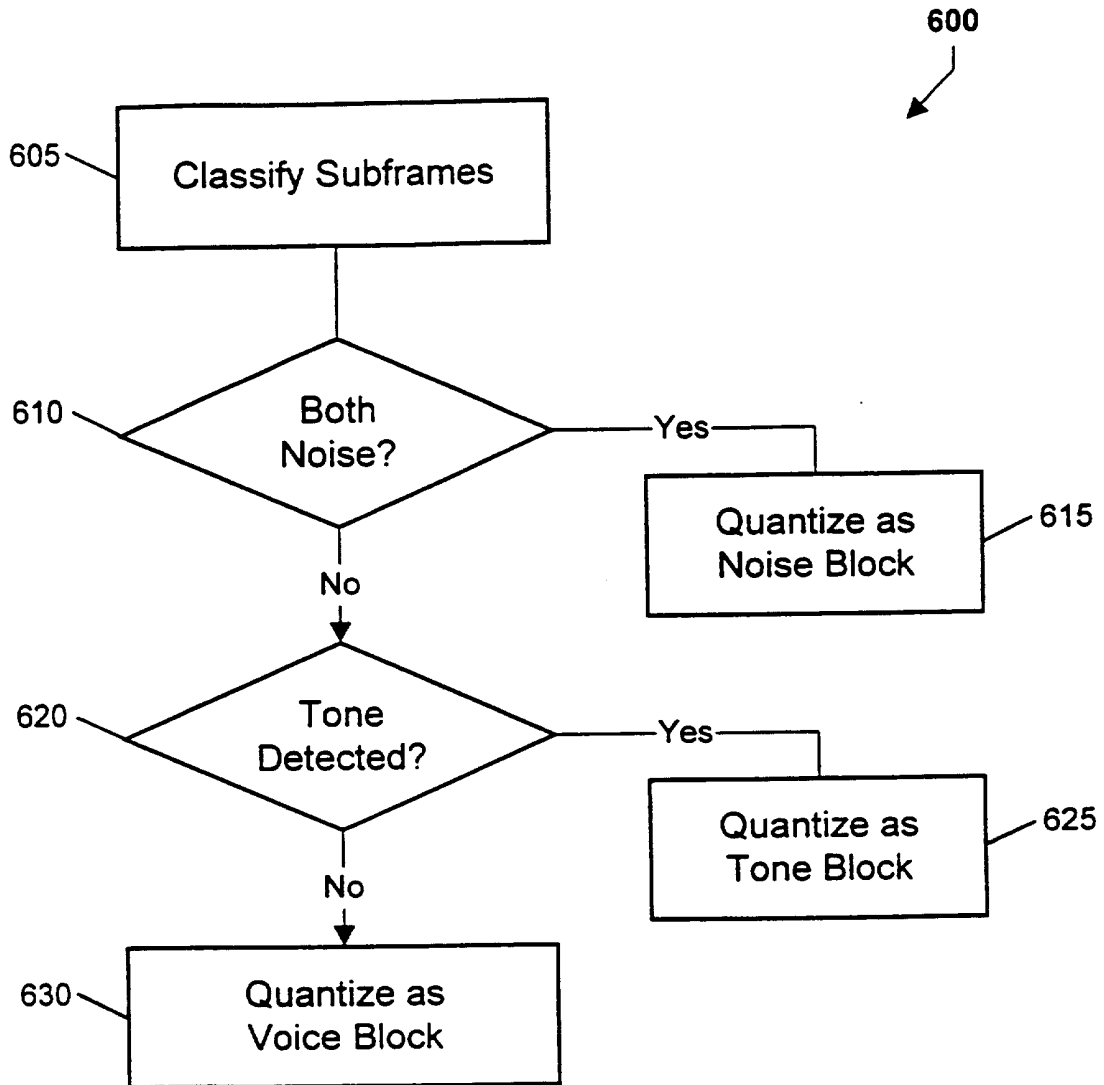


FIG. 6

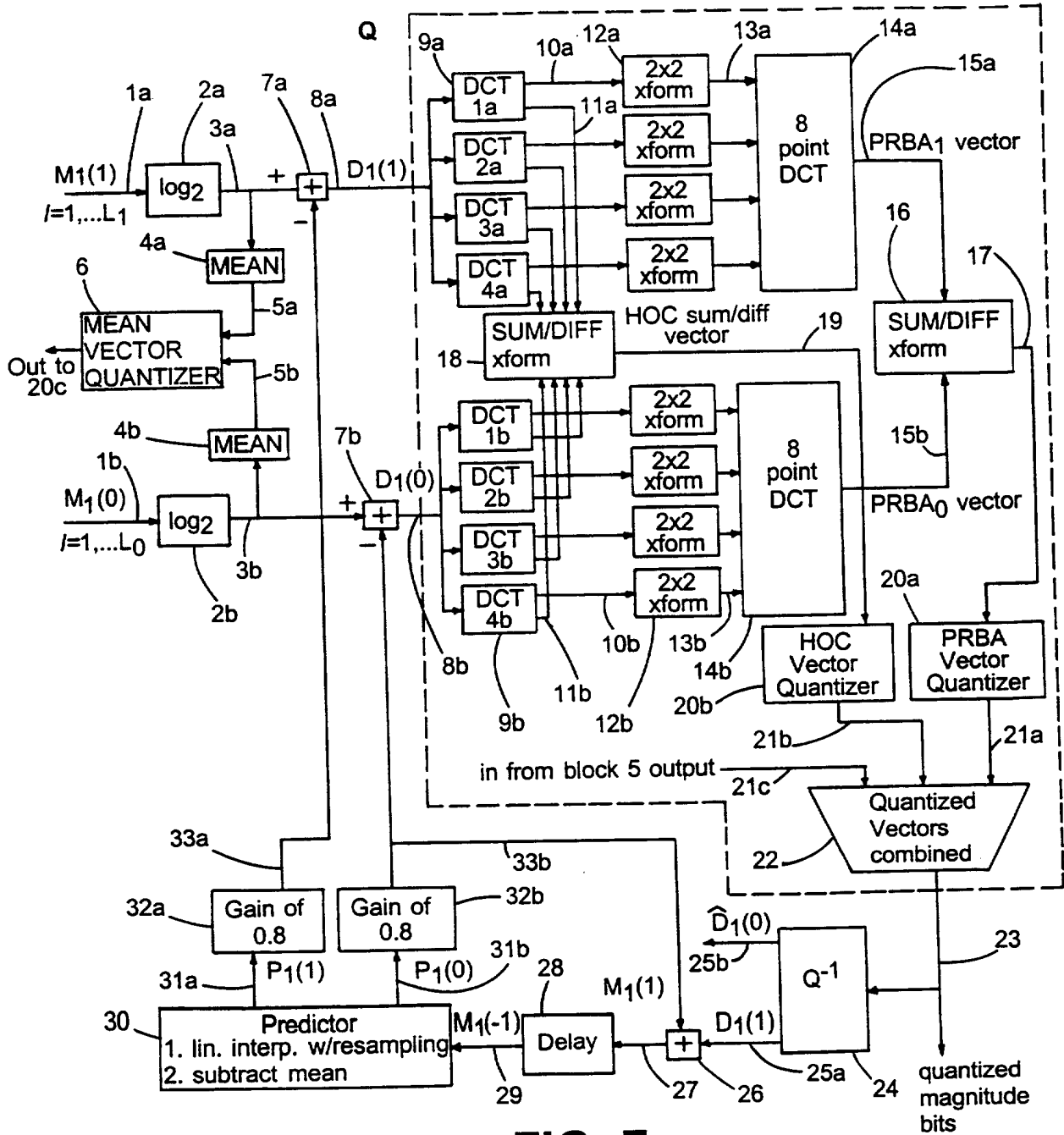


FIG. 7

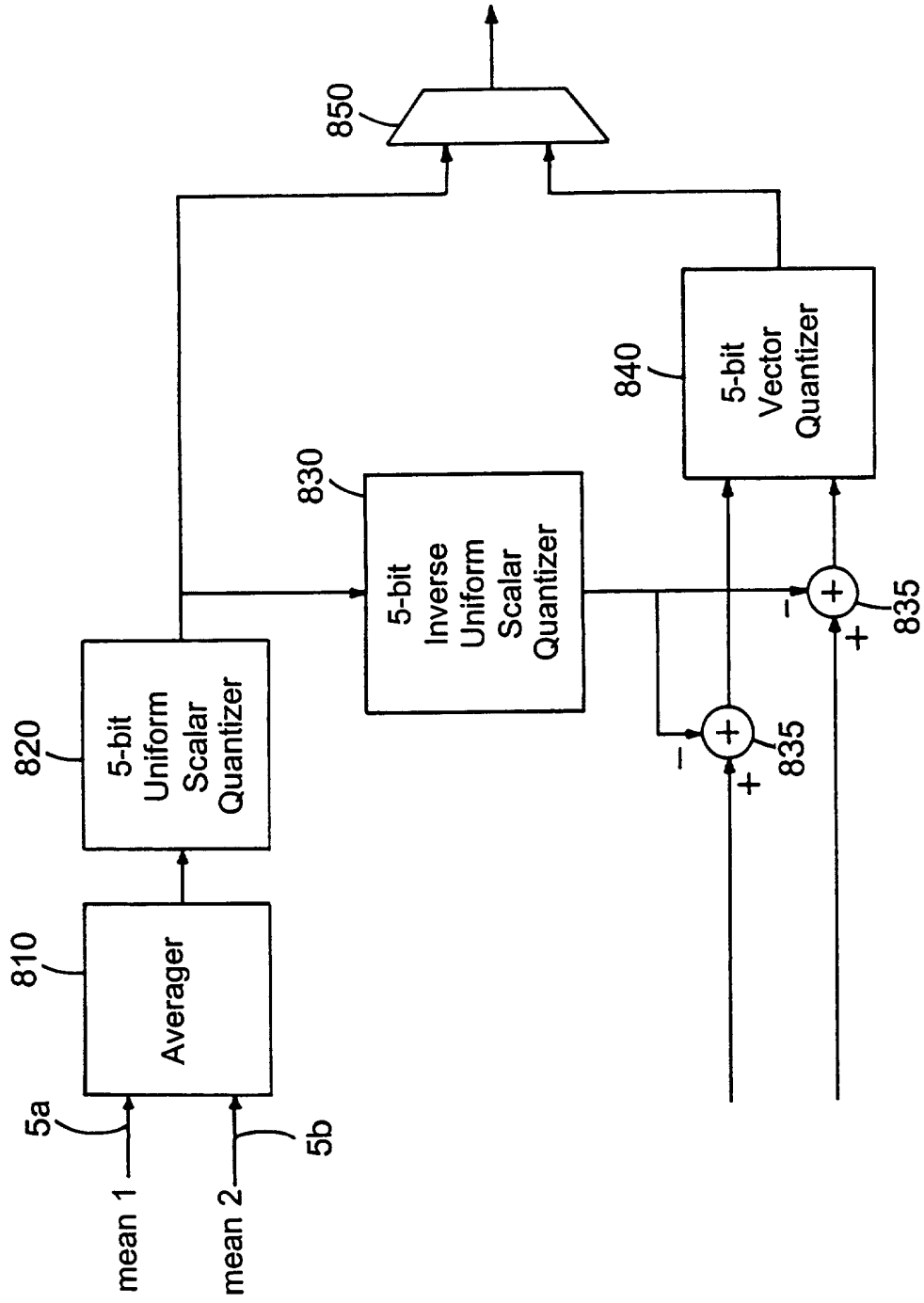


FIG. 8

DUAL SUBFRAME QUANTIZATION OF SPECTRAL MAGNITUDES

The invention is directed to encoding and decoding speech.

Speech encoding and decoding have a large number of applications and have been studied extensively. In general, one type of speech coding, referred to as speech compression, seeks to reduce the data rate needed to represent a speech signal without substantially reducing the quality or intelligibility of the speech. Speech compression techniques may be implemented by a speech coder.

A speech coder is generally viewed as including an encoder and a decoder. The encoder produces a compressed stream of bits from a digital representation of speech, such as may be generated by converting an analog signal produced by a microphone using an analog-to-digital converter. The decoder converts the compressed bit stream into a digital representation of speech that is suitable for playback through a digital-to-analog converter and a speaker. In many applications, the encoder and decoder are physically separated, and the bit stream is transmitted between them using a communication channel.

A key parameter of a speech coder is the amount of compression the coder achieves, which is measured by the bit rate of the stream of bits produced by the encoder. The bit rate of the encoder is generally a function of the desired fidelity (i.e., speech quality) and the type of speech coder employed. Different types of speech coders have been designed to operate at high rates (greater than 8 kbs), mid-rates (3 - 8 kbs) and low rates (less than 3 kbs). Recently, mid-rate and low-rate speech coders have received attention with respect to a wide range of mobile communication applications (e.g., cellular telephony,

satellite telephony, land mobile radio, and in-flight telephony). These applications typically require high quality speech and robustness to artifacts caused by acoustic noise and channel noise (e.g., bit errors).

5 Vocoders are a class of speech coders that have been shown to be highly applicable to mobile communications. A vocoder models speech as the response of a system to excitation over short time intervals. Examples of vocoder systems include linear prediction vocoders, homomorphic
10 vocoders, channel vocoders, sinusoidal transform coders ("STC"), multiband excitation ("MBE") vocoders, and improved multiband excitation ("IMBE™") vocoders. In these vocoders, speech is divided into short segments (typically 10-40 ms) with each segment being characterized by a set of model
15 parameters. These parameters typically represent a few basic elements of each speech segment, such as the segment's pitch, voicing state, and spectral envelope. A vocoder may use one of a number of known representations for each of these parameters. For example the pitch may be represented
20 as a pitch period, a fundamental frequency, or a long-term prediction delay. Similarly the voicing state may be represented by one or more voiced/unvoiced decisions, by a voicing probability measure, or by a ratio of periodic to stochastic energy. The spectral envelope is often
25 represented by an all-pole filter response, but also may be represented by a set of spectral magnitudes or other spectral measurements.

 Since they permit a speech segment to be represented using only a small number of parameters, model-based speech
30 coders, such as vocoders, typically are able to operate at medium to low data rates. However, the quality of a model-based system is dependent on the accuracy of the underlying model. Accordingly, a high fidelity model must be used if

these speech coders are to achieve high speech quality.

One speech model which has been shown to provide high quality speech and to work well at medium to low bit rates is the Multi-Band Excitation (MBE) speech model developed by Griffin and Lim. This model uses a flexible voicing structure that allows it to produce more natural sounding speech, and which makes it more robust to the presence of acoustic background noise. These properties have caused the MBE speech model to be employed in a number of commercial mobile communication applications.

The MBE speech model represents segments of speech using a fundamental frequency, a set of binary voiced/unvoiced (V/UV) metrics, and a set of spectral magnitudes. A primary advantage of the MBE model over more traditional models is in the voicing representation. The MBE model generalizes the traditional single V/UV decision per segment into a set of decisions, each representing the voicing state within a particular frequency band. This added flexibility in the voicing model allows the MBE model to better accommodate mixed voicing sounds, such as some voiced fricatives. In addition this added flexibility allows a more accurate representation of speech that has been corrupted by acoustic background noise. Extensive testing has shown that this generalization results in improved voice quality and intelligibility.

The encoder of an MBE-based speech coder estimates the set of model parameters for each speech segment. The MBE model parameters include a fundamental frequency (the reciprocal of the pitch period); a set of V/UV metrics or decisions that characterize the voicing state; and a set of spectral magnitudes that characterize the spectral envelope. After estimating the MBE model parameters for each segment, the encoder quantizes the parameters to produce a frame of

bits. The encoder optionally may protect these bits with error correction/detection codes before interleaving and transmitting the resulting bit stream to a corresponding decoder.

5 The decoder converts the received bit stream back into individual frames. As part of this conversion, the decoder may perform deinterleaving and error control decoding to correct or detect bit errors. The decoder then uses the frames of bits to reconstruct the MBE model
10 parameters, which the decoder uses to synthesize a speech signal that perceptually resembles the original speech to a high degree. The decoder may synthesize separate voiced and unvoiced components, and then may add the voiced and unvoiced components to produce the final speech signal.

15 In MBE-based systems, the encoder uses a spectral magnitude to represent the spectral envelope at each harmonic of the estimated fundamental frequency. Typically each harmonic is labeled as being either voiced or unvoiced, depending upon whether the frequency band containing the
20 corresponding harmonic has been declared voiced or unvoiced. The encoder then estimates a spectral magnitude for each harmonic frequency. When a harmonic frequency has been labeled as being voiced, the encoder may use a magnitude estimator that differs from the magnitude estimator used
25 when a harmonic frequency has been labeled as being unvoiced. At the decoder, the voiced and unvoiced harmonics are identified, and separate voiced and unvoiced components are synthesized using different procedures. The unvoiced component may be synthesized using a weighted overlap-add
30 method to filter a white noise signal. The filter is set to zero all frequency regions declared voiced while otherwise matching the spectral magnitudes labeled unvoiced. The voiced component is synthesized using a tuned oscillator

bank, with one oscillator assigned to each harmonic that has been labeled as being voiced. The instantaneous amplitude, frequency and phase are interpolated to match the corresponding parameters at neighboring segments.

5 MBE-based speech coders include the IMBE™ speech coder and the AMBE® speech coder. The AMBE® speech coder was developed as an improvement on earlier MBE-based techniques. It includes a more robust method of estimating the excitation parameters (fundamental frequency and V/UV
10 decisions) which is better able to track the variations and noise found in actual speech. The AMBE® speech coder uses a filterbank that typically includes sixteen channels and a non-linearity to produce a set of channel outputs from which the excitation parameters can be reliably estimated. The
15 channel outputs are combined and processed to estimate the fundamental frequency and then the channels within each of several (e.g., eight) voicing bands are processed to estimate a V/UV decision (or other voicing metric) for each voicing band.

20 The AMBE® speech coder also may estimate the spectral magnitudes independently of the voicing decisions. To do this, the speech coder computes a fast Fourier transform ("FFT") for each windowed subframe of speech and then averages the energy over frequency regions that are
25 multiples of the estimated fundamental frequency. This approach may further include compensation to remove from the estimated spectral magnitudes artifacts introduced by the FFT sampling grid.

30 The AMBE® speech coder also may include a phase synthesis component that regenerates the phase information used in the synthesis of voiced speech without explicitly transmitting the phase information from the encoder to the decoder. Random phase synthesis based upon the V/UV

decisions may be applied, as in the case of the IMBE™ speech coder. Alternatively, the decoder may apply a smoothing kernel to the reconstructed spectral magnitudes to produce phase information that may be perceptually closer to that of the original speech than is the randomly-produced phase information.

The techniques noted above are described, for example, in Flanagan, Speech Analysis, Synthesis and Perception, Springer-Verlag, 1972, pages 378-386 (describing a frequency-based speech analysis-synthesis system); Jayant et al., Digital Coding of Waveforms, Prentice-Hall, 1984 (describing speech coding in general); U.S. Patent No. 4,885,790 (describing a sinusoidal processing method); U.S. Patent No. 5,054,072 (describing a sinusoidal coding method); Almeida et al., "Nonstationary Modeling of Voiced Speech", IEEE TASSP, Vol. ASSP-31, No. 3, June 1983, pages 664-677 (describing harmonic modeling and an associated coder); Almeida et al., "Variable-Frequency Synthesis: An Improved Harmonic Coding Scheme", IEEE Proc. ICASSP 84, pages 27.5.1-27.5.4 (describing a polynomial voiced synthesis method); Quatieri et al., "Speech Transformations Based on a Sinusoidal Representation", IEEE TASSP, Vol. ASSP34, No. 6, Dec. 1986, pages 1449-1986 (describing an analysis-synthesis technique based on a sinusoidal representation); McAulay et al., "Mid-Rate Coding Based on a Sinusoidal Representation of Speech", Proc. ICASSP 85, pages 945-948, Tampa, FL, March 26-29, 1985 (describing a sinusoidal transform speech coder); Griffin, "Multiband Excitation Vocoder", Ph.D. Thesis, M.I.T, 1987 (describing the Multi-Band Excitation (MBE) speech model and an 8000 bps MBE speech coder); Hardwick, "A 4.8 kbps Multi-Band Excitation Speech Coder", SM. Thesis, M.I.T, May 1988 (describing a 4800 bps Multi-Band Excitation speech coder);

Telecommunications Industry Association (TIA), "APCO Project 25 Vocoder Description", Version 1.3, July 15, 1993, IS102BABA (describing a 7.2 kbps IMBE™ speech coder for APCO Project 25 standard); U.S. Patent No. 5,081,681 (describing IMBE™ random phase synthesis); U.S. Patent No. 5,247,579 (describing a channel error mitigation method and formant enhancement method for MBE-based speech coders); U.S. Patent No. 5,226,084 (describing quantization and error mitigation methods for MBE-based speech coders); U.S. Patent No. 5,517,511 (describing bit prioritization and FEC error control methods for MBE-based speech coders).

In accordance with a first aspect of the present invention, there is provided a method of encoding speech into a 90 millisecond frame of bits for transmission across a satellite communication channel, the method comprising the steps of:

digitizing a speech signal into a sequence of digital speech samples;

dividing the digital speech samples into a sequence of subframes, each of the subframes comprising a plurality of the digital speech samples;

estimating a set of model parameters for each of the subframes; wherein the model parameters comprise a set of spectral magnitude parameters that represent spectral information for the subframe;

combining two consecutive subframes from the sequence of subframes into a block;

jointly quantizing the spectral magnitude parameters from both of the subframes within the block, wherein the joint quantization includes forming predicted spectral magnitude parameters from the quantized spectral magnitude parameters from a previous block, computing residual parameters as the difference between the spectral magnitude parameters and the predicted spectral magnitude parameters, combining the residual parameters from both of the subframes

within the block, and using a plurality of vector quantizers to quantize the combined residual parameters into a set of encoded spectral bits;

adding redundant error control bits to the encoded spectral bits from each block to protect at least some of the encoded spectral bits within the block from bit errors; and

combining the added redundant error control bits and encoded spectral bits from two consecutive blocks into a 90 millisecond frame of bits for transmission across a

satellite communication channel.

According to a second and alternative aspect of this invention, we provide a method of decoding speech from a 90

millisecond frame of bits received across a satellite communication channel, the method comprising the steps of:

dividing the frame of bits into two blocks of bits, wherein each block of bits represents two subframes of speech;

applying error control decoding to each block of bits using redundant error control bits included within the block to produce error decoded bits which are at least in part protected from bit errors;

using the error decoded bits to jointly reconstruct spectral magnitude parameters for both of the subframes within a block, wherein the joint reconstruction includes using a plurality of vector quantizer codebooks to reconstruct a set of combined residual parameters from which separate residual parameters for both of the subframes are computed, forming predicted spectral magnitude parameters from the reconstructed spectral magnitude parameters from a previous block, and adding the separate residual parameters

to the predicted spectral magnitude parameters to form the reconstructed spectral magnitude parameters for each subframe within the block; and

synthesizing a plurality of digital speech samples for each subframe using the reconstructed spectral magnitude parameters for the subframe.

We provide, in a third alternative aspect of this invention, an encoder for encoding speech into a 90

millisecond frame of bits for transmission across a satellite communication channel, the system including:

a digitizer configured to convert a speech signal into a sequence of digital speech samples;

a subframe generator configured to divide the digital speech samples into a sequence of subframes, each of the subframes comprising a plurality of the digital speech samples;

a model parameter estimator configured to estimate a set of model parameters for each of the subframes, wherein the model parameters comprise a set of spectral magnitude parameters that represent spectral information for the subframe;

a combiner configured to combine two consecutive subframes from the sequence of subframes into a block;

a dual-frame spectral magnitude quantizer configured to jointly quantize parameters from both of the subframes within the block, wherein the joint quantization includes forming predicted spectral magnitude parameters from the quantized spectral magnitude parameters from a previous block, computing residual parameters as the difference between the spectral magnitude parameters and the predicted spectral magnitude parameters, combining the residual parameters from both of the subframes within the block, and using a plurality of vector quantizers to quantize the combined residual parameters into a set of encoded spectral bits;

an error code encoder configured to add redundant error control bits to the encoded spectral bits from each block to protect at least some of the encoded spectral bits within the block from bit errors; and

a combiner configured to combine the added redundant error control bits and encoded spectral bits from two consecutive blocks into a 90 millisecond frame of bits for transmission across a satellite communication channel.

The invention provides, in a further alternative aspect thereof, a decoder for decoding speech from a 90 millisecond frame of bits received across a satellite communication channel, the decoder including:

a divider configured to divide the frame of bits into two blocks of bits, wherein each block of bits represents two subframes of speech;

an error control decoder configured to error decode each block of bits using redundant error control bits included within the block to produce error decoded bits which are at least in part protected from bit errors;

a dual-frame spectral magnitude reconstructor configured to jointly reconstruct spectral magnitude parameters for both of the subframes within a block, wherein the joint reconstruction includes using a plurality of vector quantizer codebooks to reconstruct a set of combined residual parameters from which separate residual parameters for both of the subframes are computed, forming predicted spectral magnitude parameters from the reconstructed spectral magnitude parameters from a previous block, and adding the separate residual parameters to the predicted spectral magnitude parameters to form the reconstructed spectral magnitude parameters for each subframe within the block; and

a synthesizer configured to synthesize a plurality of digital speech samples for each subframe using the

reconstructed spectral magnitude parameters for the subframe.

We describe below a new AMBE® speech coder for use in a satellite communication system to produce high quality speech from a bit stream transmitted across a mobile satellite channel at a low data rate. The speech coder combines low data rate, high voice quality, and robustness to background noise and channel errors. This promises to advance the state of the art in speech coding for mobile satellite communications. The new speech coder achieves high performance through a new dual-subframe spectral magnitude quantizer that jointly quantizes the spectral magnitudes estimated from two consecutive subframes. This quantizer achieves fidelity comparable to prior art systems while using fewer bits to quantize the spectral magnitude parameters. AMBE® speech coders are described generally in

U.S. Patent 5 715 365; U.S. Patent 5 701 390; and U.S. Application No: 08392,099, filed February 22, 1995 and entitled "SYNTHESIS OF SPEECH USING REGENERATED PHASE INFORMATION", all of which are incorporated by reference.

In our method described in detail below, speech is encoded into a 90 millisecond frame of bits for transmission across a satellite communication channel. A speech signal is digitized into a sequence of digital speech samples, the digital speech samples are divided into a sequence of subframes nominally occurring at intervals of 22.5 milliseconds, and a set of model parameters is estimated for each of the subframes. The model parameters for a subframe include a set of spectral magnitude parameters that represent the spectral information for the subframe. Two consecutive subframes from the sequence of subframes are combined into a block and the spectral magnitude parameters from both of the subframes within the block are jointly quantized. The joint quantization includes forming predicted spectral magnitude

parameters from the quantized spectral magnitude parameters from the previous block, computing residual parameters as the difference between the spectral magnitude parameters and the predicted spectral magnitude parameters for the block, combining the residual parameters from both of the subframes within the block, and using vector quantizers to quantize the combined residual parameters into a set of encoded spectral bits. Redundant error control bits then are added to the encoded spectral bits from each block to protect the encoded spectral bits within the block from bit errors. The added redundant error control bits and encoded spectral bits from two consecutive blocks are then combined into a 90 millisecond frame of bits for transmission across a satellite communication channel.

Embodiments of the invention may include one or more of the following features. The combining of the residual

parameters from both of the subframes within the block may include dividing the residual parameters from each of the subframes into frequency blocks, performing a linear transformation on the residual parameters within each of the frequency blocks to produce a set of transformed residual coefficients for each of the subframes, grouping a minority of the transformed residual coefficients from all of the frequency blocks into a PRBA vector and grouping the remaining transformed residual coefficients for each of the frequency blocks into a HOC vector for the frequency block. The PRBA vectors for each subframe may be transformed to produce transformed PRBA vectors and the vector sum and difference for the transformed PRBA vectors for the subframes of a block may be computed to combine the transferred PRBA vectors. Similarly, the vector sum and difference for each frequency block may be computed to combine the two HOC vectors from the two subframes for that frequency block.

The spectral magnitude parameters may represent the log spectral magnitudes estimated for the Multi-Band Excitation ("MBE") speech model. The spectral magnitude parameters may be estimated from a computed spectrum independently of the voicing state. The predicted spectral magnitude parameters may be formed by applying a gain of less than unity to the linear interpolation of the quantized spectral magnitudes from the last subframe in the previous block.

The error control bits for each block may be formed using block codes including Golay codes and Hamming codes. For example, the codes may include one [24,12] extended Golay code, three [23,12] Golay codes, and two [15,11] Hamming codes.

The transformed residual coefficients may be

computed for each of the frequency blocks using a Discrete Cosine Transform ("DCT") followed by a linear 2 by 2 transform on the two lowest order DCT coefficients. Four frequency blocks may be used for this computation and the length of each the frequency block may be approximately proportional to the number of spectral magnitude parameters within the subframe.

The vector quantizers may include a three way split vector quantizer using 8 bits plus 6 bits plus 7 bits applied to the PRBA vector sum and a two way split vector quantizer using 8 bits plus 6 bits applied to the PRBA vector difference. The frame of bits may include additional bits representing the error in the transformed residual coefficients which is introduced by the vector quantizers.

We describe herein

a system for encoding speech into a 90 millisecond frame of bits for transmission across a satellite communication channel. The system includes a digitizer that converts a speech signal into a sequence of digital speech samples, a subframe generator that divides the digital speech samples into a sequence of subframes that each include multiple digital speech samples. A model parameter estimator estimates a set of model parameters that include a set of spectral magnitude parameters for each of the subframes. A combiner combines two consecutive subframes from the sequence of subframes into a block. A dual-frame spectral magnitude quantizer jointly quantizes parameters from both of the subframes within the block. The joint quantization includes forming predicted spectral magnitude parameters from the quantized spectral magnitude parameters from a previous block, computing residual parameters as the difference between the spectral magnitude parameters and the predicted spectral magnitude parameters, combining the

residual parameters from both of the subframes within the block, and using vector quantizers to quantize the combined residual parameters into a set of encoded spectral bits. The system also includes an error code encoder that adds redundant error control bits to the encoded spectral bits from each block to protect at least some of the encoded spectral bits within the block from bit errors, and a combiner that combines the added redundant error control bits and encoded spectral bits from two consecutive blocks into a 90 millisecond frame of bits for transmission across a satellite communication channel.

We also describe

decoding speech from a 90 millisecond frame that has been encoded as described above. The decoding includes dividing the frame of bits into two blocks of bits, wherein each block of bits represents two subframes of speech. Error control decoding is applied to each block of bits using redundant error control bits included within the block to produce error decoded bits which are at least in part protected from bit errors. The error decoded bits are used to jointly reconstruct spectral magnitude parameters for both of the subframes within a block. The joint reconstruction includes using vector quantizer codebooks to reconstruct a set of combined residual parameters from which separate residual parameters for both of the subframes are computed, forming predicted spectral magnitude parameters from the reconstructed spectral magnitude parameters from a previous block, and adding the separate residual parameters to the predicted spectral magnitude parameters to form the reconstructed spectral magnitude parameters for each subframe within the block. Digital speech samples are then synthesized for each subframe using the reconstructed spectral magnitude parameters for the subframe.

We also describe herein a decoder for decoding speech from a 90 millisecond frame of bits received across a satellite communication channel. The decoder includes a divider that divides the frame of bits into two blocks of bits. Each block of bits represents two subframes of speech. An error control decoder error decodes each block of bits using redundant error control bits included within the block to produce error decoded bits which are at least in part protected from bit errors. A dual-frame spectral magnitude reconstructor jointly reconstructs spectral magnitude parameters for both of the subframes within a block, wherein the joint reconstruction includes using vector quantizer codebooks to reconstruct a set of combined residual parameters from which separate residual parameters for both of the subframes are computed, forming predicted spectral magnitude parameters from the reconstructed spectral magnitude parameters from a previous block, and adding the separate residual parameters to the predicted spectral magnitude parameters to form the reconstructed spectral magnitude parameters for each subframe within the block. A synthesizer synthesizes digital speech samples for each subframe using the reconstructed spectral magnitude parameters for the subframe.

Other features and advantages of the invention will be apparent from the following description, including the drawings, in which:

Fig. 1 is a simplified block diagram of a satellite system.

Fig. 2 is a block diagram of a communication link of the system of Fig. 1.

Figs. 3 and 4 are block diagrams of an encoder and a decoder of the system of Fig. 1.

Fig. 5 is a general block diagram of components of the encoder of Fig. 3.

5 Fig. 6 is a flow chart of the voice and tone detection functions of the encoder.

Fig. 7 is a block diagram of a dual subframe magnitude quantizer of the encoder of Fig. 5.

10 Fig. 8 is a block diagram of a mean vector quantizer of the magnitude quantizer of Fig. 7.

An embodiment of the invention is described in the context of a new AMBE speech coder, or vocoder, for use in the IRIDIUM[®] mobile satellite communication system 30, as shown in Fig. 1. IRIDIUM[®] is a global mobile satellite communication system consisting of sixty-six satellites 40 in low earth orbit. IRIDIUM[®] provides voice communications through handheld or vehicle based user terminals 45 (i.e., mobile phones).

20 Referring to Fig. 2, the user terminal at the transmitting end achieves voice communication by digitizing speech 50 received through a microphone 60 using an analog-to-digital (A/D) converter 70 that samples the speech at a frequency of 8 kHz. The digitized speech signal passes through a speech encoder 80, where it is processed as described below. The signal is then transmitted across the communication link by a transmitter 90. At the other end of the communication link, a receiver 100 receives the signal and passes it to a decoder 110. The decoder converts the signal into a synthetic digital speech signal. A digital-to-analog (D/A) converter 120 then converts the synthetic digital speech signal into an analog speech signal that is

converted into audible speech 140 by a speaker 130.

The communications link uses burst-transmission time-division-multiple-access (TDMA) with a 90 ms frame. Two different data rates for voice are supported: a half-
5 rate mode of 3467 bps (312 bits per 90 ms frame) and a full-
rate mode of 6933 bps (624 bits per 90 ms frame). The bits of each frame are divided between speech coding and forward error correction ("FEC") coding to lower the probability of bit errors that normally occur across a satellite
10 communication channel.

Referring to Fig. 3, the speech coder in each terminal includes an encoder 80 and a decoder 110. The encoder includes three main functional blocks: speech analysis 200, parameter quantization 210, and error
15 correction encoding 220. Similarly, as shown in Fig. 4, the decoder is divided into functional blocks for error correction decoding 230, parameter reconstruction 240 (i.e., inverse quantization) and speech synthesis 250.

The speech coder may operate at two distinct data
20 rates: a full-rate of 4933 bps and a half-rate of 2289 bps. These data rates represent voice or source bits and exclude FEC bits. The FEC bits raise the data rate of the full-rate and half-rate vocoders to 6933 bps and 3467 bps,
respectively, as noted above. The system uses a voice frame
25 size of 90 ms which is divided into four 22.5 ms subframes. Speech analysis and synthesis are performed on a subframe basis, while quantization and FEC coding are performed on a 45 ms quantization block that includes two subframes. The
30 use of 45 ms blocks for quantization and FEC coding results in 103 voice bits plus 53 FEC bits per block in the half-
rate system, and 222 voice bits plus 90 FEC bits per block in the full-rate system. Alternatively, the number of voice bits and FEC bits can be adjusted within a range with only

gradual effect on performance. In the half-rate system, adjustment of the voice bits in the range of 80 to 120 bits with the corresponding adjustment in the FEC bits in the range of 76 to 36 bits can be accomplished. Similarly, in the full-rate system, the voice bits can be adjusted over the range of 180 to 260 bits with the corresponding adjustment in the FEC bits spanning from 132 to 52 bits. The voice and FEC bits for the quantization blocks are combined to form a 90 ms frame.

The encoder 80 first performs speech analysis 200. The first step in speech analysis is filterbank processing on each subframe followed by estimation of the MBE model parameters for each subframe. This involves dividing the input signal into overlapping 22.5 ms subframes using an analysis window. For each 22.5 ms subframe, a MBE subframe parameter estimator estimates a set of model parameters that include a fundamental frequency (inverse of the pitch period), a set of voiced/unvoiced (V/UV) decisions and a set of spectral magnitudes. These parameters are generated using AMBE techniques. AMBE® speech coders are described generally in U.S. Patent 5 715 365; U.S. Patent 5 701 390 and U.S. Application No: 08/392,099, filed February 22, 1995 and entitled "SYNTHESIS OF SPEECH USING REGENERATED PHASE INFORMATION", all of which are incorporated by reference.

In addition, the full-rate vocoder includes a time-slot ID that helps to identify out-of-order arrival of TDMA packets at the receiver, which can use this information to place the information in the correct order prior to decoding. The speech parameters fully describe the speech

signal and are passed to the encoder's quantization 210 block for further processing.

Referring to Fig. 5, once the subframe model parameters 300 and 305 are estimated for two consecutive 5 22.5 ms subframes within a frame, the fundamental frequency and voicing quantizer 310 encodes the fundamental frequencies estimated for both subframes into a sequence of fundamental frequency bits, and further encodes the voiced/unvoiced (V/UV) decisions (or other voicing metrics) 10 into a sequence of voicing bits.

In the described embodiment, ten bits are used to quantize and encode the two fundamental frequencies. Typically, the fundamental frequencies are limited by the fundamental estimate to a range of approximately [0.008, 15 0.05] where 1.0 is the Nyquist frequency (8 kHz), and the fundamental quantizer is limited to a similar range. Since the inverse of the quantized fundamental frequency for a given subframe is generally proportional to L, the number of spectral magnitudes for that subframe (L = bandwidth/ 20 fundamental frequency), the most significant bits of the fundamental are typically sensitive to bit errors and consequently are given high priority in FEC encoding.

The described embodiment uses eight bits in half-rate and sixteen bits in full-rate to encode the voicing 25 information for both subframes. The voicing quantizer uses the allocated bits to encode the binary voicing state (i.e., 1 = voiced and 0 = unvoiced) in each of the preferred eight voicing bands, where the voicing state is determined by voicing metrics estimated during speech analysis. These 30 voicing bits have moderate sensitivity to bit errors and hence are given medium priority in FEC encoding.

The fundamental frequency bits and voicing bits are combined in the combiner 330 with the quantized spectral

magnitude bits from the dual subframe magnitude quantizer 320, and forward error correction (FEC) coding is performed for that 45 ms block. The 90 ms frame is then formed in a combiner 340 that combines two consecutive 45 ms quantized
5 blocks into a single frame 350.

The encoder incorporates an adaptive Voice Activity Detector (VAD) which classifies each 22.5 ms subframe as either voice, background noise, or a tone according to a procedure 600. As shown in Fig. 6, the VAD algorithm uses
10 local information to distinguish voice subframes from background noise (step 605). If both subframes within each 45 ms block are classified as noise (step 610), then the encoder quantizes the background noise that is present as a special noise block (step 615). When the two 45 ms block
15 comprising a 90 ms frame are both classified as noise, then the system may choose not to transmit this frame to the decoder and the decoder will use previously received noise data in place of the missing frame. This voice activated transmission technique increases performance of the system
20 by only requiring voice frames and occasional noise frames to be transmitted.

The encoder also may feature tone detection and transmission in support of DTMF, call progress (e.g., dial, busy and ringback) and single tones. The encoder checks
25 each 22.5 ms subframe to determine whether the current subframe contains a valid tone signal. If a tone is detected in either of the two subframes of a 45 ms block (step 620), then the encoder quantizes the detected tone parameters (magnitude and index) in a special tone block as
30 shown in Table 1 (step 625) and applies FEC coding prior to transmitting the block to the decoder for subsequent synthesis. If a tone is not detected, then a standard voice block is quantized as described below (step 630).

Half-Rate		Full-Rate	
b [] element #	Value	b [] element #	Value
0 - 3	15	0 - 7	212
4 - 9	16	8 - 15	212
10-12	3 MSB's of Amplitude	16-18	3 MSB's of Amplitude
13-14	0	19-20	0
15-19	5 LSB's of Amplitude	21-25	5 LSB's of Amplitude
20-27	Detected Tone Index	26-33	Detected Tone Index
28-35	Detected Tone Index	34-41	Detected Tone Index
36-43	Detected Tone Index	42-49	Detected Tone Index
.	.	.	.
84-91	Detected Tone Index	194-201	Detected Tone Index
92-99	Detected Tone Index	202-209	Detected Tone Index
100-102	0	210-221	0

Table 1: Tone Block Bit Representation

The vocoder includes VAD and Tone detection to classify each 45 ms block as either a standard Voice block, a special Tone block or a special noise block. In the event a 45 ms block is not classified as a special tone block, then the voice or noise information (as determined by the VAD) is quantized for the pair of subframes comprising that block. The available bits (156 for half-rate, 312 for full-rate) are allocated over the model parameters and FEC coding as shown in Table 2, where the Slot ID is a special parameter used by the full-rate receiver to identify the correct ordering of frames that may arrive out of order. After reserving bits for the excitation parameters (fundamental frequency and voicing metrics), FEC coding and the Slot ID, there are 85 bits available for the spectral

magnitudes in the half-rate system and 183 bits available for the spectral magnitudes in the full-rate system. To support the full-rate system with a minimum amount of additional complexity, the full-rate magnitude quantizer uses the same quantizer as the half-rate system plus an error quantizer that uses scalar quantization to encode the difference between the unquantized spectral magnitudes and the quantized output of the half-rate spectral magnitude quantizer.

Vocoder Parameter	Bits (Half-Rate)	Bits (Full-Rate)
Fund. Freq.	10	16
Voicing Metrics	8	16
Gain	$5+5=10$	$5+5+2*2=14$
PRBA Vector	$8+6+7+8+6=35$	$8+6+7+8+6+2*12=59$
HOC Vector	$4*(7+3)=40$	$4*(7+3)+2*(9+9+9+8)=110$
Slot ID	0	7
FEC	$12+3*11+2*4=53$	$2*12+6*11=90$
Total	156	312

Table 2: Bit Allocation for 45 ms Voice or Noise block

A dual-subframe quantizer is used to quantize the spectral magnitudes. The quantizer combines logarithmic companding, spectral prediction, discrete cosine transforms (DCTs) and vector and scalar quantization to achieve high efficiency, measured in terms of fidelity per bit, with reasonable complexity. The quantizer can be viewed as a two dimensional predictive transform coder.

Fig. 7 illustrates the dual subframe magnitude quantizer that receives inputs 1a and 1b from the MBE parameter estimators for two consecutive 22.5 ms subframes.

Input 1a represents the spectral magnitudes for odd numbered 22.5 ms subframes and is given an index of 1. The number of magnitudes for subframe number 1 is designated by L_1 . Input 1b represents the spectral magnitudes for the even numbered 22.5 ms subframes and is given the index of 0. The number of magnitudes for subframe number 0 is designated by L_0 .

Input 1a passes through a logarithmic compander 2a, which performs a log base 2 operation on each of the L_1 magnitudes contained in input 1a and generates another vector with L_1 elements in the following manner:

$$y[i] = \log_2(x[i]) \quad \text{for } i = 1, 2, \dots, L_1,$$

where $y[i]$ represents signal 3a. Compander 2b performs the log base 2 operation on each of the L_0 magnitudes contained in input 1b and generates another vector with L_0 elements in a similar manner:

$$y[i] = \log_2(x[i]) \quad \text{for } i = 1, 2, \dots, L_0,$$

where $y[i]$ represents signal 3b.

Mean calculators 4a and 4b following the companders 2a and 2b calculate means 5a and 5b for each subframe. The mean, or gain value, represents the average speech level for the subframe. Within each frame, two gain values 5a, 5b are determined by computing the mean of the log spectral magnitudes for each of the two subframes and then adding an offset dependent on the number of harmonics within the subframe.

The mean computation of the log spectral magnitudes 3a is calculated as:

$$y = \frac{1}{L_1} \sum_{i=1}^{L_1} x[i] + 0.5 \log_2(L_1)$$

where the output, y , represents the mean signal 5a.

The mean computation 4b of the log spectral

magnitudes 3b is calculated in a similar manner:

$$y = \frac{1}{L_0} \sum_{i=1}^{L_0} x[i] + 0.5 \log_2(L_0)$$

where the output, y , represents the mean signal 5b.

The mean signals 5a and 5b are quantized by a quantizer 6 that is further illustrated in Fig. 8, where the mean signals 5a and 5b are referenced, respectively, as mean1 and mean2. First, an averager 810 averages the mean signals. The output of the averager is $0.5 * (\text{mean1} + \text{mean2})$. The average is then quantized by a five-bit uniform scalar quantizer 820. The output of the quantizer 820 forms the first five bits of the output of the quantizer 6. The quantizer output bits are then inverse-quantized by a five-bit uniform inverse scalar quantizer 830. Subtracters 835 then subtract the output of the inverse quantizer 830 from the input values mean1 and mean2 to produce inputs to a five-bit vector quantizer 840. The two inputs constitute a two-dimensional vector ($z1$ and $z2$) to be quantized. The vector is compared to each two-dimensional vector (consisting of $x1(n)$ and $x2(n)$) in the table contained in Appendix A ("Gain VQ Codebook (5-bit)"). The comparison is based on the square distance, e , which is calculated as follows:

$$e(n) = [x1(n) - z1]^2 + [x2(n) - z2]^2,$$

for $n = 0, 1, \dots, 31$. The vector from Appendix A that

minimizes the square distance, e , is selected to produce the last five bits of the output of block 6. The five bits from the output of the vector quantizer 840 are combined with the five bits from the output of the five-bit uniform scalar
5 quantizer 820 by a combiner 850. The output of the combiner 850 is ten bits constituting the output of block 6 which is labeled 21c and is used as an input to the combiner 22 in Fig. 7.

10 Referring further to the main signal path of the quantizer, the log companded input signals 3a and 3b pass through combiners 7a and 7b that subtract predictor values 33a and 33b from the feedback portion of the quantizer to produce a $D_1(1)$ signal 8a and a $D_1(0)$ signal 8b.

15 Next, the signals 8a and 8b are divided into four frequency blocks using the look-up table in Appendix O. The table provides the number of magnitudes to be allocated to each of the four frequency blocks based on the total number of magnitudes for the subframe being divided. Since the number of magnitudes contained in any subframe ranges from a
20 minimum of 9 to a maximum of 56, the table contains values for this same range. The length of each frequency block is adjusted such that they are approximately in a ratio of 0.2:0.225:0.275:0.3 to each other and the sum of the lengths equals the number of spectral magnitudes in the current
25 subframe.

Each frequency block is then passed through a discrete cosine transform (DCT) 9a or 9b to efficiently decorrelate the data within each frequency block. The first two DCT coefficients 10a or 10b from each frequency block
30 are then separated out and passed through a 2×2 rotation operation 12a or 12b to produce transformed coefficients 13a or 13b. An eight-point DCT 14a or 14b is then performed on the transformed coefficients 13a or 13b to produce a PRBA

vector 15a or 15b. The remaining DCT coefficients 11a and 11b from each frequency block form a set of four variable length higher order coefficient (HOC) vectors.

As described above, following the frequency division, each block is processed by the discrete cosine transform blocks 9a or 9b. The DCT blocks use the number of input bins, W , and the values for each of the bins, $x(0)$, $x(1)$, ... , $x(W-1)$ in the following manner:

$$y(k) = \frac{1}{W} \sum_{i=0}^{W-1} x(i) \cos \frac{(2i+1)k\pi}{2W} \quad \text{for } 0 \leq k \leq (W-1)$$

The values $y(0)$ and $y(1)$ (identified as 10a) are separated from the other outputs $y(2)$ through $y(W-1)$ (identified as 11a).

A 2x2 rotation operation 12a and 12b is then performed to transform the 2-element input vector 10a and 10b, $(x(0), x(1))$, into a 2-element output vector 13a and 13b, $(y(0), y(1))$ by the following rotation procedure:

$$\begin{aligned} y(0) &= x(0) + \text{sqrt}(2) * x(1), \text{ and} \\ y(1) &= x(0) - \text{sqrt}(2) * x(1). \end{aligned}$$

An 8-point DCT is then performed on the four, 2-element vectors, $(x(0), x(1), \dots, x(7))$ from 13a or 13b according to the following equation:

$$y(k) = \frac{1}{8} \sum_{i=0}^7 x(i) \cos \frac{(2i+1)k\pi}{16} \quad \text{for } 0 \leq k \leq 7$$

The output, $y(k)$, is an 8-element PRBA vector 15a or 15b.

Once the prediction and DCT transformation of the individual subframe magnitudes have been completed, both PRBA vectors are quantized. The two eight-element vectors

are first combined using a sum-difference transformation 16 into a sum vector and a difference vector. In particular, sum/difference operation 16 is performed on the two 8-element PRBA vectors 15a and 15b, which are represented by x and y respectively, to produce a 16-element vector 17, represented by z, in the following manner:

$$z(i) = x(i) + y(i), \text{ and}$$
$$z(8+i) = x(i) - y(i),$$

for $i = 0, 1, \dots, 7$.

10 These vectors are then quantized using a split vector quantizer 20a where 8, 6, and 7 bits are used for elements 1-2, 3-4, and 5-7 of the sum vector, respectively, and 8 and 6 bits are used for elements 1-3 and 4-7 of the difference vector, respectively. Element 0 of each vector
15 is ignored since it is functionally equivalent to the gain value that is quantized separately.

The quantization of the PRBA sum and difference vectors 17 is performed by the PRBA split-vector quantizer 20a to produce a quantized vector 21a. The two elements
20 $z(1)$ and $z(2)$ constitute a two-dimensional vector to be quantized. The vector is compared to each two-dimensional vector (consisting of $x_1(n)$ and $x_2(n)$ in the table contained in Appendix B ("PRBA Sum[1,2] VQ Codebook (8-bit)"). The comparison is based on the square distance, e , which is
25 calculated as follows:

$$e(n) = [x_1(n) - z(1)]^2 + [x_2(n) - z(2)]^2,$$
$$\text{for } n = 0, 1, \dots, 255.$$

The vector from Appendix B that minimizes the square

distance, e , is selected to produce the first 8 bits of the output vector 21a.

Next, the two elements $z(3)$ and $z(4)$ constitute a two-dimensional vector to be quantized. The vector is compared to each two-dimensional vector (consisting of $x1(n)$ and $x2(n)$ in the table contained in Appendix C ("PRBA Sum[3,4] VQ Codebook (6-bit)"). The comparison is based on the square distance, e , which is calculated as follows:

$$e(n) = [x1(n) - z(3)]^2 + [x2(n) - z(4)]^2,$$

10 for $n = 0, 1, \dots, 63$.

The vector from Appendix C which minimizes the square distance, e , is selected to produce the next 6 bits of the output vector 21a.

Next, the three elements $z(5)$, $z(6)$ and $z(7)$ constitute a three-dimensional vector to be quantized. The vector is compared to each three-dimensional vector (consisting of $x1(n)$, $x2(n)$ and $x3(n)$ in the table contained in Appendix D ("PRBA Sum[5,7] VQ Codebook (7bit)"). The comparison is based on the square distance, e , which is calculated as follows:

$$e(n) = [x1(n) - z(5)]^2 + [x2(n) - z(6)]^2 + [x3(n) - z(7)]^2 \text{ for } n = 0, 1, \dots, 127.$$

The vector from Appendix D which minimizes the square distance, e , is selected to produce the next 7 bits of the output vector 21a.

Next, the three elements $z(9)$, $z(10)$ and $z(11)$ constitute a three-dimensional vector to be quantized. The vector is compared to each three-dimensional vector (consisting of $x1(n)$, $x2(n)$ and $x3(n)$ in the table contained

in Appendix E ("PRBA Dif[1,3] VQ Codebook (8-bit)"). The comparison is based on the square distance, e , which is calculated as follows:

$$e(n) = [x1(n) - z(9)]^2 + [x2(n) - z(10)]^2 + [x3(n) - z(11)]^2 \text{ for } n = 0, 1, \dots, 255.$$

The vector from Appendix E which minimizes the square distance, e , is selected to produce the next 8 bits of the output vector 21a.

Finally, the four elements $z(12)$, $z(13)$, $z(14)$ and $z(15)$ constitute a four-dimensional vector to be quantized. The vector is compared to each four-dimensional vector (consisting of $x1(n)$, $x2(n)$, $x3(n)$ and $x4(n)$ in the table contained in Appendix F ("PRBA Dif[4,7] VQ Codebook (6-bit)"). The comparison is based on the square distance, e , which is calculated as follows:

$$e(n) = [x1(n) - z(12)]^2 + [x2(n) - z(13)]^2 + [x3(n) - z(14)]^2 + [x4(n) - z(15)]^2 \text{ for } n = 0, 1, \dots, 63.$$

The vector from Appendix F which minimizes the square distance, e , is selected to produce the last 6 bits of the output vector 21a.

The HOC vectors are quantized similarly to the PRBA vectors. First, for each of the four frequency blocks, the corresponding pair of HOC vectors from the two subframes are combined using a sum-difference transformation 18 that produces a sum and difference vector 19 for each frequency block.

The sum/difference operation is performed separately for each frequency block on the two HOC vectors 11a and 11b, referred to as x and y respectively, to produce a vector,

z_m :

$$J = \max(B_{m0}, B_{m1}) - 2$$
$$K = \min(B_{m0}, B_{m1}) - 2$$

$$z_m(i) = 0.5[x(i) + y(i)] \quad \text{for } 1 \leq i \leq K$$

$$z_m(i) = \begin{cases} y(i) & \text{if } L_0 > L_1 \\ x(i) & \text{otherwise} \end{cases} \quad \text{for } K < i \leq J$$

$$z_m(J+i) = 0.5[x(i) - y(i)] \quad \text{for } 0 \leq i < K.$$

where B_{m0} and B_{m1} are the lengths of the m th frequency block for, respectively, subframes zero and one, as set forth in Appendix O, and z is determined for each frequency block (i.e., m equals 0 to 3). The $J+K$ element sum and difference vectors z_m are combined for all four frequency blocks (m equals 0 to 3) to form the HOC sum/difference vector 19.

Due to the variable size of each HOC vector, the sum and difference vectors also have variable, and possibly different, lengths. This is handled in the vector quantization step by ignoring any elements beyond the first four elements of each vector. The remaining elements are vector quantized using seven bits for the sum vector and three bits for the difference vector. After vector quantization is performed, the original sum-difference transformation is reversed on the quantized sum and difference vectors. Since this process is applied to all four frequency blocks a total of forty ($4 * (7+3)$) bits are used to vector quantize the HOC vectors corresponding to both subframes.

The quantization of the HOC sum and difference vectors 19 is performed separately on all four frequency blocks by the HOC split-vector quantizer 20b. First, the

vector z_m representing the m th frequency block is separated and compared against each candidate vector in the corresponding sum and difference codebooks contained in the Appendices. A codebook is identified based on the frequency block to which it corresponds and whether it is a sum or difference code. Thus, the "HOC Sum0 VQ Codebook (7-bit)" of Appendix G represents the sum codebook for frequency block 0. The other codebooks are Appendix H ("HOC Dif0 VQ Codebook (3-bit)"), Appendix I ("HOC Sum1 VQ Codebook (7-bit)"), Appendix J ("HOC Dif1 VQ Codebook (3-bit)"), Appendix K ("HOC Sum2 VQ Codebook (7-bit)"), Appendix L ("HOC Dif2 VQ Codebook (3-bit)"), Appendix M ("HOC Sum2 VQ Codebook (7-bit)"), and Appendix N ("HOC Dif3 VQ Codebook (3-bit)"). The comparison of the vector z_m for each frequency block with each candidate vector from the corresponding sum codebooks is based upon the square distance, $e1_n$ for each candidate sum vector (consisting of $x1(n)$, $x2(n)$, $x3(n)$ and $x4(n)$) which is calculated as:

$$e1_n = \sum_{i=1}^{\min(J,4)} [z(i) - xi(n)]^2 \quad 0 \leq n < 128,$$

and the square distance $e2_m$ for each candidate difference vector (consisting of $x1(n)$, $x2(n)$, $x3(n)$ and $x4(n)$), which is calculated as:

$$e2_m = \sum_{i=1}^{\min(K,4)} [z(J+i) - xi(m)]^2 \quad 0 \leq m < 8,$$

where J and K are computed as described above.

The index n of the candidate sum vector from the corresponding sum notebook which minimizes the square distance $e1_n$ is represented with seven bits and the index m of the candidate difference vector which minimizes the

square distance e_{2m} is represented with three bits. These ten bits are combined from all four frequency blocks to form the 40 HOC output bits 21b.

5 Block 22 multiplexes the quantized PRBA vectors 21a, the quantized mean 21b, and the quantized mean 21c to produce output bits 23. These bits 23 are the final output bits of the dual-subframe magnitude quantizer and are also supplied to the feedback portion of the quantizer.

10 Block 24 of the feedback portion of the dual-subframe quantizer represents the inverse of the functions performed in the superblock labeled Q in the drawing. Block 24 produces estimated values 25a and 25b of $D_1(1)$ and $D_1(0)$ (8a and 8b) in response to the quantized bits 23. These estimates would equal $D_1(1)$ and $D_1(0)$ in the absence of
15 quantization error in the superblock labeled Q.

Block 26 adds a scaled prediction value 33a, which equals $0.8 * P_1(1)$, to the estimate of $D_1(1)$ 25a to produce an estimate $M_1(1)$ 27. Block 28 time-delays the estimate $M_1(1)$ 27 by one frame (40 ms) to produce the estimate $M_1(-1)$
20 29.

A predictor block 30 then interpolates the estimated magnitudes and resamples them to produce L_1 estimated magnitudes after which the mean value of the estimated magnitudes is subtracted from each of the L_1 estimated
25 magnitudes to produce the $P_1(1)$ output 31a. Next, the input estimated magnitudes are interpolated and resampled to produce L_0 estimated magnitudes after which the mean value of the estimated magnitudes is subtracted from each of the L_0 estimated magnitudes to produce the $P_1(0)$ output 31b.

30 Block 32a multiplies each magnitude in $P_1(1)$ 31a by 0.8 to produce the output vector 33a which is used in the feedback element combiner block 7a. Likewise, block 32b multiplies each magnitude in $P_1(1)$ 31b by 0.8 to produce the

output vector 33b which is used in the feedback element combiner block 7b. The output of this process is the quantized magnitude output vector 23, which is then combined with the output vector of two other subframes as described above.

Once the encoder has quantized the model parameters for each 45 ms block, the quantized bits are prioritized, FEC encoded and interleaved prior to transmission. The quantized bits are first prioritized in order of their approximate sensitivity to bit errors. Experimentation has shown that the PRBA and HOC sum vectors are typically more sensitive to bits errors than corresponding difference vectors. In addition, the PRBA sum vector is typically more sensitive than the HOC sum vector. These relative sensitivities are employed in a prioritization scheme which generally gives the highest priority to the average fundamental frequency and average gain bits, followed by the PRBA sum bits and the HOC sum bits, followed by the PRBA difference bits and the HOC difference bits, followed by any remaining bits.

A mix of [24,12] extended Golay codes, [23,12] Golay codes and [15,11] Hamming codes are then employed to add higher levels of redundancy to the more sensitive bits while adding less or no redundancy to the less sensitive bits. The half-rate system applies one [24,12] Golay code, followed by three [23,12] Golay codes, followed by two [15,11] Hamming codes, with the remaining 33 bits unprotected. The full-rate system applies two [24,12] Golay codes, followed by six [23,12] Golay codes with the remaining 126 bits unprotected. This allocation was designed to make efficient use of limited number of bits available for FEC. The final step is to interleave the FEC encoded bits within each 45 ms block to spread the effect of

any short error bursts. The interleaved bits from two consecutive 45 ms blocks are then combined into a 90 ms frame which forms the encoder output bit stream.

The corresponding decoder is designed to reproduce high quality speech from the encoded bit stream after it is transmitted and received across the channel. The decoder first separates each 90 ms frame into two 45 ms quantization blocks. The decoder then deinterleaves each block and performs error correction decoding to correct and/or detect certain likely bit error patterns. To achieve adequate performance over the mobile satellite channel, all error correction codes are typically decoded up to their full error correction capability. Next, the FEC decoded bits are used by the decoder to reassemble the quantization bits for that block from which the model parameters representing the two subframes within that block are reconstructed.

The AMBE[®] decoder uses the reconstructed log spectral magnitudes to synthesize a set of phases which are used by the voiced synthesizer to produce natural sounding speech. The use of synthesized phase information significantly lowers the transmitted data rate, relative to a system which directly transmits this information or its equivalent between the encoder and decoder. The decoder then applies spectral enhancement to the reconstructed spectral magnitudes in order to improve the perceived quality of the speech signal. The decoder further checks for bit errors and smoothes the reconstructed parameters if the local estimated channel conditions indicate the presence of possible uncorrectable bit errors. The enhanced and smoothed model parameters (fundamental frequency, V/UV decisions, spectral magnitudes and synthesized phases) are used in speech synthesis.

The reconstructed parameters form the input to the

decoder's speech synthesis algorithm which interpolates successive frames of model parameters into smooth 22.5 ms segments of speech. The synthesis algorithm uses a set of harmonic oscillators (or an FFT equivalent at high
5 frequencies) to synthesize the voiced speech. This is added to the output of a weighted overlap-add algorithm to synthesize the unvoiced speech. The sums form the synthesized speech signal which is output to a D-to-A converter for playback over a speaker. While this
10 synthesized speech signal may not be close to the original on a sample-by-sample basis, it is perceived as the same by a human listener.

Other embodiments are feasible.

CLAIMS

1 1. A method of encoding speech into a 90
2 millisecond frame of bits for transmission across a
3 satellite communication channel, the method comprising the
4 steps of:
5 digitizing a speech signal into a sequence of
6 digital speech samples;
7 dividing the digital speech samples into a sequence
8 of subframes, each of the subframes comprising a plurality
9 of the digital speech samples;
10 estimating a set of model parameters for each of the
11 subframes; wherein the model parameters comprise a set of
12 spectral magnitude parameters that represent spectral
13 information for the subframe;
14 combining two consecutive subframes from the
15 sequence of subframes into a block;
16 jointly quantizing the spectral magnitude parameters
17 from both of the subframes within the block, wherein the
18 joint quantization includes forming predicted spectral
19 magnitude parameters from the quantized spectral magnitude
20 parameters from a previous block, computing residual
21 parameters as the difference between the spectral magnitude
22 parameters and the predicted spectral magnitude parameters,
23 combining the residual parameters from both of the subframes
24 within the block, and using a plurality of vector quantizers
25 to quantize the combined residual parameters into a set of
26 encoded spectral bits;
27 adding redundant error control bits to the encoded
28 spectral bits from each block to protect at least some of
29 the encoded spectral bits within the block from bit errors;
30 and
31 combining the added redundant error control bits and
32 encoded spectral bits from two consecutive blocks into a 90
33 millisecond frame of bits for transmission across a

34 satellite communication channel.

1 2. The method of claim 1 wherein the combining of
2 the residual parameters from both of the subframes within
3 the block further comprises:

4 dividing the residual parameters from each of the
5 subframes into a plurality of frequency blocks;

6 performing a linear transformation on the residual
7 parameters within each of the frequency blocks to produce a
8 set of transformed residual coefficients for each of the
9 subframes;

10 grouping a minority of the transformed residual
11 coefficients from all of the frequency blocks into a PRBA
12 vector and grouping the remaining transformed residual
13 coefficients for each of the frequency blocks into a HOC
14 vector for the frequency block;

15 transforming the PRBA vector to produce a
16 transformed PRBA vector and computing the vector sum and
17 difference to combine the two transformed PRBA vectors from
18 both of the subframes; and

19 computing the vector sum and difference for each
20 frequency block to combine the two HOC vectors from both of
21 the subframes for that frequency block.

1 3. The method of claim 1 or 2, wherein the spectral
2 magnitude parameters represent log spectral magnitudes
3 estimated for a Multi-Band Excitation (MBE) speech model.

1 4. The method of claim 3, wherein the spectral
2 magnitude parameters are estimated from a computed spectrum
3 independently of a voicing state.

1 5. The method of claim 1 or 2, wherein the

2 predicted spectral magnitude parameters are formed by
3 applying a gain of less than unity to a linear interpolation
4 of the quantized spectral magnitudes from the last subframe
5 in the previous block.

1 6. The method of claim 1 or 2, wherein the
2 redundant error control bits for each block are formed by a
3 plurality of block codes including Golay codes and Hamming
4 codes.

1 7. The method of claim 6, wherein the plurality of
2 block codes consists of one [24,12] extended Golay code,
3 three [23,12] Golay codes, and two [15,11] Hamming codes.

1 8. The method of claim 2 wherein the transformed
2 residual coefficients are computed for each of the frequency
3 blocks using a Discrete Cosine Transform (DCT) followed by a
4 linear 2 by 2 transform on the two lowest order DCT
5 coefficients.

1 9. The method of claim 8 wherein four frequency
2 blocks are used and wherein the length of each frequency
3 block is approximately proportional to a number of spectral
4 magnitude parameters within the subframe.

1 10. The method of claim 2, wherein the plurality of
2 vector quantizers includes a three way split vector
3 quantizer using 8 bits plus 6 bits plus 7 bits applied to
4 the PRBA vector sum and a two way split vector quantizer
5 using 8 bits plus 6 bits applied to the PRBA vector
6 difference.

1 11. The method of claim 10 wherein the frame of

2 bits includes additional bits representing the error in the
3 transformed residual coefficients which is introduced by the
4 vector quantizers.

1 12. The method of claim 1 or 2, wherein the
2 sequence of subframes nominally occurs at an interval of
3 22.5 milliseconds per subframe.

1 13. The method of claim 12, wherein the frame of
2 bits consists of 312 bits in half-rate mode or 624 bits in
3 full-rate mode.

1 14. A method of decoding speech from a 90
2 millisecond frame of bits received across a satellite
3 communication channel, the method comprising the steps of:
4 dividing the frame of bits into two blocks of bits,
5 wherein each block of bits represents two subframes of
6 speech;
7 applying error control decoding to each block of
8 bits using redundant error control bits included within the
9 block to produce error decoded bits which are at least in
10 part protected from bit errors;
11 using the error decoded bits to jointly reconstruct
12 spectral magnitude parameters for both of the subframes
13 within a block, wherein the joint reconstruction includes
14 using a plurality of vector quantizer codebooks to
15 reconstruct a set of combined residual parameters from which
16 separate residual parameters for both of the subframes are
17 computed, forming predicted spectral magnitude parameters
18 from the reconstructed spectral magnitude parameters from a
19 previous block, and adding the separate residual parameters
20 to the predicted spectral magnitude parameters to form the
21 reconstructed spectral magnitude parameters for each

22 subframe within the block; and
23 synthesizing a plurality of digital speech samples
24 for each subframe using the reconstructed spectral magnitude
25 parameters for the subframe.

1 15. The method of claim 14 wherein the computing of
2 the separate residual parameters for both of the subframes
3 from the combined residual parameters for the block
4 comprises the further steps of:
5 dividing the combined residual parameters from the
6 block into a plurality of frequency blocks;
7 forming a transformed PRBA sum and difference vector
8 for the block;
9 forming a HOC sum and difference vector for each of
10 the frequency blocks from the combined residual parameters;
11 applying an inverse sum and difference operation and
12 an inverse transformation to the transformed PRBA sum and
13 difference vectors to form the PRBA vectors for both of the
14 subframes; and
15 applying an inverse sum and difference operation to
16 the HOC sum and difference vectors to form HOC vectors for
17 both of the subframes for each of the frequency blocks; and
18 combining the PRBA vector and the HOC vectors for
19 each of the frequency blocks for each of the subframes to
20 form the separate residual parameters for both of the
21 subframes within the block.

1 16. The method of claim 14 or 15, wherein the
2 reconstructed spectral magnitude parameters represent the
3 log spectral magnitudes used in a Multi-Band Excitation
4 (MBE) speech model.

1 17. The method of claim 14 or 15, further

2 comprising a decoder synthesizing a set of phase parameters
3 using the reconstructed spectral magnitude parameters.

1 18. The method of claim 14 or 15, wherein the
2 predicted spectral magnitude parameters are formed by
3 applying a gain of less than unity to the linear
4 interpolation of the quantized spectral magnitudes from the
5 last subframe in the previous block.

1 19. The method of claim 14 or 15, wherein the error
2 control bits for each block are formed by a plurality of
3 block codes including Golay codes and Hamming codes.

1 20. The method of claim 19, wherein the plurality
2 of block codes consists of one [24,12] extended Golay code,
3 three [23,12] Golay codes, and two [15,11] Hamming codes.

1 21. The method of claim 15, wherein the transformed
2 residual coefficients are computed for each of the frequency
3 blocks using a Discrete Cosine Transform ("DCT") followed by
4 a linear 2 by 2 transform on the two lowest order DCT
5 coefficients.

1 22. The method of claim 21, wherein four frequency
2 blocks are used and wherein the length of each frequency
3 block is approximately proportional to the number of
4 spectral magnitude parameters within the subframe.

1 23. The method of claim 15, wherein the plurality
2 of vector quantizer codebooks includes a three way split
3 vector quantizer codebook using 8 bits plus 6 bits plus 7
4 bits applied to the PRBA sum vector and a two way split
5 vector quantizer codebook using 8 bits plus 6 bits applied

6 to the PRBA difference vector.

1 24. The method of claim 23, wherein the frame of
2 bits includes additional bits representing the error in the
3 transformed residual coefficients which is introduced by the
4 vector quantizer codebooks.

1 25. The method of claim 14 or 15, wherein the
2 subframes have a nominal duration of 22.5 milliseconds.

1 26. The method of claim 25, wherein the frame of
2 bits consists of 312 bits in half-rate mode or 624 bits in
3 full-rate mode.

1 27. An encoder for encoding speech into a 90
2 millisecond frame of bits for transmission across a
3 satellite communication channel, the system including:
4 a digitizer configured to convert a speech signal
5 into a sequence of digital speech samples;
6 a subframe generator configured to divide the
7 digital speech samples into a sequence of subframes, each of
8 the subframes comprising a plurality of the digital speech
9 samples;
10 a model parameter estimator configured to estimate a
11 set of model parameters for each of the subframes, wherein
12 the model parameters comprise a set of spectral magnitude
13 parameters that represent spectral information for the
14 subframe;
15 a combiner configured to combine two consecutive
16 subframes from the sequence of subframes into a block;
17 a dual-frame spectral magnitude quantizer configured
18 to jointly quantize parameters from both of the subframes
19 within the block, wherein the joint quantization includes

20 forming predicted spectral magnitude parameters from the
21 quantized spectral magnitude parameters from a previous
22 block, computing residual parameters as the difference
23 between the spectral magnitude parameters and the predicted
24 spectral magnitude parameters, combining the residual
25 parameters from both of the subframes within the block, and
26 using a plurality of vector quantizers to quantize the
27 combined residual parameters into a set of encoded spectral
28 bits;

29 an error code encoder configured to add redundant
30 error control bits to the encoded spectral bits from each
31 block to protect at least some of the encoded spectral bits
32 within the block from bit errors; and

33 a combiner configured to combine the added redundant
34 error control bits and encoded spectral bits from two
35 consecutive blocks into a 90 millisecond frame of bits for
36 transmission across a satellite communication channel.

1 28. The encoder of claim 27, wherein the dual-frame
2 spectral magnitude quantizer is configured to combine the
3 residual parameters from both of the subframes within the
4 block by:

5 dividing the residual parameters from each of the
6 subframes into a plurality of frequency blocks;

7 performing a linear transformation on the residual
8 parameters within each of the frequency blocks to produce a
9 set of transformed residual coefficients for each of the
10 subframes;

11 grouping a minority of the transformed residual
12 coefficients from all of the frequency blocks into a PRBA
13 vector and grouping the remaining transformed residual
14 coefficients for each of the frequency blocks into a HOC
15 vector for the frequency block;

16 transforming the PRBA vector to produce a
17 transformed PRBA vector and computing the vector sum and
18 difference to combine the two transformed PRBA vectors from
19 both of the subframes; and
20 computing the vector sum and difference for each
21 frequency block to combine the two HOC vectors from both of
22 the subframes for that frequency block.

1 29. A decoder for decoding speech from a 90
2 millisecond frame of bits received across a satellite
3 communication channel, the decoder including:
4 a divider configured to divide the frame of bits
5 into two blocks of bits, wherein each block of bits
6 represents two subframes of speech;
7 an error control decoder configured to error decode
8 each block of bits using redundant error control bits
9 included within the block to produce error decoded bits
10 which are at least in part protected from bit errors;
11 a dual-frame spectral magnitude reconstructor
12 configured to jointly reconstruct spectral magnitude
13 parameters for both of the subframes within a block, wherein
14 the joint reconstruction includes using a plurality of
15 vector quantizer codebooks to reconstruct a set of combined
16 residual parameters from which separate residual parameters
17 for both of the subframes are computed, forming predicted
18 spectral magnitude parameters from the reconstructed
19 spectral magnitude parameters from a previous block, and
20 adding the separate residual parameters to the predicted
21 spectral magnitude parameters to form the reconstructed
22 spectral magnitude parameters for each subframe within the
23 block; and
24 a synthesizer configured to synthesize a plurality
25 of digital speech samples for each subframe using the

26 reconstructed spectral magnitude parameters for the
27 subframe.

1 30. The decoder of claim 29, wherein the dual-frame
2 spectral magnitude quantizer is configured to compute the
3 separate residual parameters for both of the subframes from
4 the combined residual parameters for the block by:
5 dividing the combined residual parameters from the
6 block into a plurality of frequency blocks;
7 forming a transformed PRBA sum and difference vector
8 for the block;
9 forming a HOC sum and difference vector for each of
10 the frequency blocks from the combined residual parameters;
11 applying an inverse sum and difference operation and
12 an inverse transformation to the transformed PRBA sum and
13 difference vectors to form the PRBA vectors for both of the
14 subframes; and
15 applying an inverse sum and difference operation to
16 the HOC sum and difference vectors to form HOC vectors for
17 both of the subframes for each of the frequency blocks; and
18 combining the PRBA vector and the HOC vectors for
19 each of the frequency blocks for each of the subframes to
20 form the separate residual parameters for both of the
21 subframes within the block.

31. A method, substantially as hereinbefore described with
reference to the accompanying drawings, for encoding speech into
a 90 millisecond frame of bits for transmission across a
satellite communication channel.

32. A method, substantially as hereinbefore described with
reference to the accompanying drawings, for decoding speech from
a 90 millisecond frame of bits received across a satellite
communication channel.

33. An encoder for encoding speech into a 90 millisecond frame of bits for transmission across a satellite communication channel, substantially as hereinbefore described with reference to and as shown in the accompanying drawings.

34. A decoder for decoding speech from a 90 millisecond frame of bits received across a satellite communication channel, substantially as hereinbefore described with reference to and as shown in the accompanying drawings.



Application No: GB 9805682.3
Claims searched: 1-34

Examiner: Keith Williams
Date of search: 21 August 1998

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:
UK Cl (Ed.P): H4P (PDCFP, PDCFT, PEP); H4R (RPV, RPVA, RPVS)
Int Cl (Ed.6): G10L 7/06, 9/04, 9/14, 9/18, 19/00, 19/02, 19/04, 19/06, 19/10, 19/14;
H03M 7/30, 13/00; H04L 1/00
Other: Online WPI

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	EP 0577488 A1 Nippon Telegraph - see abstract (and see EP 0751496)	1,14,27,29
A	EP 0422232 A1 K K Toshiba - see abstract (and WO 90/13112, US 5265167)	1,14,27,29
A	WO 94/12972 A1 Digital Voice Systems - see abstract (and US 5630011)	1,14,27,29
A	WO 94/12932 A1 Digital Voice Systems - see page 7, line 17 onwards (and US 5517511)	1,14,27,29
A	WO 92/10830 A1 Digital Voice Systems - see abstract (and US 5226084, US 5247579, EP 0560931, US 5491772)	1,14,27,29

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.