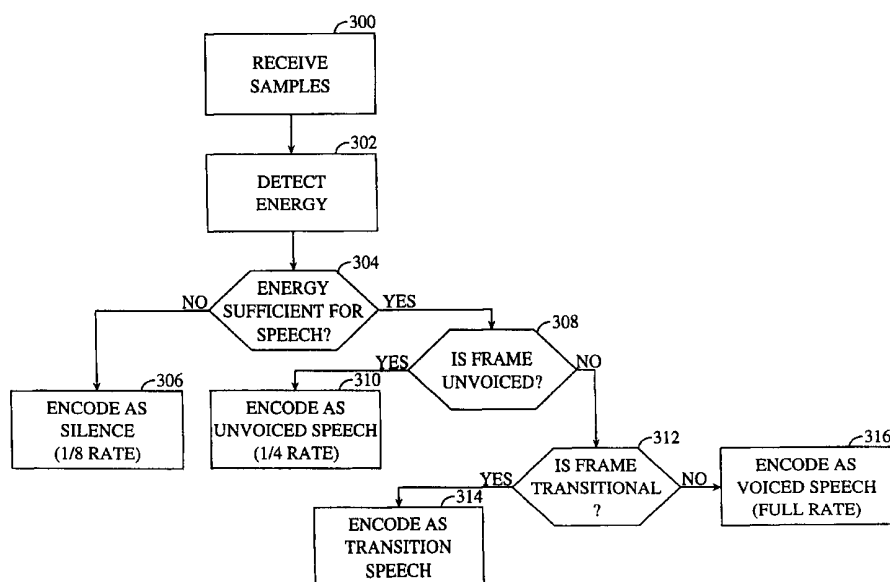




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : G10L 19/10, 19/14	A1	(11) International Publication Number: WO 00/68935 (43) International Publication Date: 16 November 2000 (16.11.00)
(21) International Application Number: PCT/US00/12656 (22) International Filing Date: 8 May 2000 (08.05.00) (30) Priority Data: 09/307,294 7 May 1999 (07.05.99) US (71) Applicant: QUALCOMM INCORPORATED [US/US]; 5775 Morehouse Drive, San Diego, CA 92121-1714 (US). (72) Inventors: DAS, Amitava; 11738 Ramsdell Court, San Diego, CA 92131 (US). MANJUNATH, Sharath; 7104 Schilling Avenue #5, San Diego, CA 92126 (US). (74) Agents: WADSWORTH, Philip, R. et al.; Qualcomm Incorporated, 5775 Morehouse Drive, San Diego, CA 92121-1714 (US).		(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MULTIPULSE INTERPOLATIVE CODING OF TRANSITION SPEECH FRAMES



(57) Abstract

A multipulse interpolative coder for transition speech frames includes an extractor configured to represent a first frame of transitional speech samples by a subset of the samples of the frame. The coder also includes an interpolator configured to interpolate the subset of samples and a subset of samples extracted from an earlier-received frame to synthesize other samples of the first frame that are not included in the subset. The subset of samples is further simplified by selecting a set of pulses from the subset and assigning zero values to unselected pulses. In the alternative, a portion of the unselected pulses may be quantized. The set of pulses may be the pulses having the greatest absolute amplitudes in the subset. In the alternative, the set of pulses may be the most perceptually significant pulses of the subset.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

MULTIPULSE INTERPOLATIVE CODING OF TRANSITION SPEECH FRAMES

BACKGROUND OF THE INVENTION

5 I. Field of the Invention

The present invention pertains generally to the field of speech processing, and more specifically to multipulse interpolative coding of transition speech frames.

II. Background

10 Transmission of voice by digital techniques has become widespread, particularly in long distance and digital radio telephone applications. This, in turn, has created interest in determining the least amount of information that can be sent over a channel while maintaining the perceived quality of the reconstructed speech. If speech is transmitted by simply sampling and
15 digitizing, a data rate on the order of sixty-four kilobits per second (kbps) is required to achieve a speech quality of conventional analog telephone. However, through the use of speech analysis, followed by the appropriate coding, transmission, and resynthesis at the receiver, a significant reduction in the data rate can be achieved.

20 Devices that employ techniques to compress speech by extracting parameters that relate to a model of human speech generation are called speech coders. A speech coder divides the incoming speech signal into blocks of time, or analysis frames. Speech coders typically comprise an encoder and a decoder. The encoder analyzes the incoming speech frame to
25 extract certain relevant parameters, and then quantizes the parameters into binary representation, i.e., to a set of bits or a binary data packet. The data packets are transmitted over the communication channel to a receiver and a decoder. The decoder processes the data packets, unquantizes them to produce the parameters, and resynthesizes the speech frames using the
30 unquantized parameters.

The function of the speech coder is to compress the digitized speech signal into a low-bit-rate signal by removing all of the natural redundancies inherent in speech. The digital compression is achieved by representing the

input speech frame with a set of parameters and employing quantization to represent the parameters with a set of bits. If the input speech frame has a number of bits N_i and the data packet produced by the speech coder has a number of bits N_o , the compression factor achieved by the speech coder is C_r ,
5 $= N_i/N_o$. The challenge is to retain high voice quality of the decoded speech while achieving the target compression factor. The performance of a speech coder depends on (1) how well the speech model, or the combination of the analysis and synthesis process described above, performs, and (2) how well the parameter quantization process is performed at the target bit rate of N_o .
10 bits per frame. The goal of the speech model is thus to capture the essence of the speech signal, or the target voice quality, with a small set of parameters for each frame.

Speech coders may be implemented as time-domain coders, which attempt to capture the time-domain speech waveform by employing high
15 time-resolution processing to encode small segments of speech (typically 5 millisecond (ms) subframes) at a time. For each subframe, a high-precision representative from a codebook space is found by means of various search algorithms known in the art. Alternatively, speech coders may be implemented as frequency-domain coders, which attempt to capture the
20 short-term speech spectrum of the input speech frame with a set of parameters (analysis) and employ a corresponding synthesis process to recreate the speech waveform from the spectral parameters. The parameter quantizer preserves the parameters by representing them with stored representations of code vectors in accordance with known quantization
25 techniques described in A. Gersho & R.M. Gray, *Vector Quantization and Signal Compression* (1992).

A well-known time-domain speech coder is the Code Excited Linear Predictive (CELP) coder described in L.B. Rabiner & R.W. Schafer, *Digital Processing of Speech Signals* 396-453 (1978), which is fully incorporated
30 herein by reference. In a CELP coder, the short term correlations, or redundancies, in the speech signal are removed by a linear prediction (LP) analysis, which finds the coefficients of a short-term formant filter. Applying the short-term prediction filter to the incoming speech frame

generates an LP residue signal, which is further modeled and quantized with long-term prediction filter parameters and a subsequent stochastic codebook. Thus, CELP coding divides the task of encoding the time-domain speech waveform into the separate tasks of encoding the LP short-term filter coefficients and encoding the LP residue. Time-domain coding can be performed at a fixed rate (i.e., using the same number of bits, N_0 , for each frame) or at a variable rate (in which different bit rates are used for different types of frame contents). Variable-rate coders attempt to use only the amount of bits needed to encode the codec parameters to a level adequate to obtain a target quality. An exemplary variable rate CELP coder is described in U.S. Patent No. 5,414,796, which is assigned to the assignee of the present invention and fully incorporated herein by reference.

Time-domain coders such as the CELP coder typically rely upon a high number of bits, N_0 , per frame to preserve the accuracy of the time-domain speech waveform. Such coders typically deliver excellent voice quality provided the number of bits, N_0 , per frame relatively large (e.g., 8 kbps or above). However, at low bit rates (4 kbps and below), time-domain coders fail to retain high quality and robust performance due to the limited number of available bits. At low bit rates, the limited codebook space clips the waveform-matching capability of conventional time-domain coders, which are so successfully deployed in higher-rate commercial applications.

There is presently a surge of research interest and strong commercial need to develop a high-quality speech coder operating at medium to low bit rates (i.e., in the range of 2.4 to 4 kbps and below). The application areas include wireless telephony, satellite communications, Internet telephony, various multimedia and voice-streaming applications, voice mail, and other voice storage systems. The driving forces are the need for high capacity and the demand for robust performance under packet loss situations. Various recent speech coding standardization efforts are another direct driving force propelling research and development of low-rate speech coding algorithms. A low-rate speech coder creates more channels, or users, per allowable application bandwidth, and a low-rate speech coder coupled with an additional layer of suitable channel coding can fit the overall bit-budget of

coder specifications and deliver a robust performance under channel error conditions.

One effective technique to encode speech efficiently at low bit rates is multimode coding. An exemplary multimode coding technique is described in Amitava Das et al., Multimode and Variable-Rate Coding of Speech, in *Speech Coding and Synthesis* ch. 7 (W.B. Kleijn & K.K. Paliwal eds., 1995). Conventional multimode coders apply different modes, or encoding-decoding algorithms, to different types of input speech frames. Each mode, or encoding-decoding process, is customized to optimally represent a certain type of speech segment, such as, e.g., voiced speech, unvoiced speech, transition speech (e.g., between voiced and unvoiced), and background noise (nonspeech) in the most efficient manner. An external, open-loop mode decision mechanism examines the input speech frame and makes a decision regarding which mode to apply to the frame. The open-loop mode decision is typically performed by extracting a number of parameters from the input frame, evaluating the parameters as to certain temporal and spectral characteristics, and basing a mode decision upon the evaluation. The mode decision is thus made without knowing in advance the exact condition of the output speech, i.e., how close the output speech will be to the input speech in terms of voice quality or other performance measures.

To retain high voice quality, it is critical to represent transition speech frames accurately. For a low-bit-rate speech coder that uses a limited number of bits per frame, this has traditionally proven to be difficult. Thus, there is a need for a speech coder that accurately represents transition speech frames coded at a low bit rate.

SUMMARY OF THE INVENTION

The present invention is directed to a speech coder that accurately represents transition speech frames coded at a low bit rate. Accordingly, in one aspect of the invention, a method of coding transitional speech frames advantageously includes the steps of representing a first frame of transitional speech samples by a first subset of the samples of the first frame; and interpolating the first subset of samples and a second subset of samples

extracted from a second, earlier-received frame of transitional speech samples to synthesize other samples of the first frame that are not included in the first subset.

5 In another aspect of the invention, a speech coder for coding transitional speech frames advantageously includes means for representing a first frame of transitional speech samples by a first subset of the samples of the first frame; and means for interpolating the first subset of samples and a second subset of samples extracted from a second, earlier-received frame of transitional speech samples to synthesize other samples of the first frame
10 that are not included in the first subset.

In another aspect of the invention, a speech coder for coding transitional frames of speech advantageously includes an extractor configured to represent a first frame of transitional speech samples by a first subset of the samples of the first frame; and an interpolator coupled to the
15 extractor and configured to interpolate the first subset of samples and a second subset of samples extracted from a second, earlier-received frame of transitional speech samples to synthesize other samples of the first frame that are not included in the first subset.

BRIEF DESCRIPTION OF THE DRAWINGS

20 FIG. 1 is a block diagram of a communication channel terminated at each end by speech coders.

FIG. 2 is a block diagram of an encoder.

FIG. 3 is a block diagram of a decoder.

FIG. 4 is a flow chart illustrating a speech coding decision process.

25 FIG. 5A is a graph speech signal amplitude versus time, and FIG. 5B is a graph of linear prediction (LP) residue amplitude versus time.

FIG. 6 is a flow chart illustrating a multipulse interpolative coding process for transition speech frames.

30 FIG. 7 is a block diagram of a system for filtering an LP-residue-domain signal to generate a speech domain signal, or for inverse filtering a speech-domain signal to generate an LP-residue-domain signal.

FIGS. 8A-D are graphs of signal amplitude versus time for, respectively, original transitional speech, uncoded residual, coded/quantized residual, and decoded/reconstructed speech.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In FIG. 1 a first encoder 10 receives digitized speech samples $s(n)$ and encodes the samples $s(n)$ for transmission on a transmission medium 12, or communication channel 12, to a first decoder 14. The decoder 14 decodes the encoded speech samples and synthesizes an output speech signal $s_{\text{SYNTH}}(n)$.
10 For transmission in the opposite direction, a second encoder 16 encodes digitized speech samples $s(n)$, which are transmitted on a communication channel 18. A second decoder 20 receives and decodes the encoded speech samples, generating a synthesized output speech signal $s_{\text{SYNTH}}(n)$.

The speech samples $s(n)$ represent speech signals that have been
15 digitized and quantized in accordance with any of various methods known in the art including, e.g., pulse code modulation (PCM), companded μ -law, or A-law. As known in the art, the speech samples $s(n)$ are organized into frames of input data wherein each frame comprises a predetermined number of digitized speech samples $s(n)$. In an exemplary embodiment, a
20 sampling rate of 8 kHz is employed, with each 20 ms frame comprising 160 samples. In the embodiments described below, the rate of data transmission may advantageously be varied on a frame-to-frame basis from 13.2 kbps (full rate) to 6.2 kbps (half rate) to 2.6 kbps (quarter rate) to 1 kbps (eighth rate). Varying the data transmission rate is advantageous because lower bit rates
25 may be selectively employed for frames containing relatively less speech information. As understood by those skilled in the art, other sampling rates, frame sizes, and data transmission rates may be used.

The first encoder 10 and the second decoder 20 together comprise a first speech coder, or speech codec. Similarly, the second encoder 16 and the
30 first decoder 14 together comprise a second speech coder. It is understood by those of skill in the art that speech coders may be implemented with a digital

signal processor (DSP), an application-specific integrated circuit (ASIC), discrete gate logic, firmware, or any conventional programmable software module and a microprocessor. The software module could reside in RAM memory, flash memory, registers, or any other form of writable storage medium known in the art. Alternatively, any conventional processor, controller, or state machine could be substituted for the microprocessor. Exemplary ASICs designed specifically for speech coding are described in U.S. Patent No. 5,727,123, assigned to the assignee of the present invention and fully incorporated herein by reference, and U.S. Application Serial No. 08/197,417, entitled VOCODER ASIC, filed February 16, 1994, assigned to the assignee of the present invention, and fully incorporated herein by reference.

In FIG. 2 an encoder 100 that may be used in a speech coder includes a mode decision module 102, a pitch estimation module 104, an LP analysis module 106, an LP analysis filter 108, an LP quantization module 110, and a residue quantization module 112. Input speech frames $s(n)$ are provided to the mode decision module 102, the pitch estimation module 104, the LP analysis module 106, and the LP analysis filter 108. The mode decision module 102 produces a mode index I_M and a mode M based upon the periodicity of each input speech frame $s(n)$. Various methods of classifying speech frames according to periodicity are described in U.S. Application Serial No. 08/815,354, entitled METHOD AND APPARATUS FOR PERFORMING REDUCED RATE VARIABLE RATE VOCODING, filed March 11, 1997, assigned to the assignee of the present invention, and fully incorporated herein by reference. Such methods are also incorporated into the Telecommunication Industry Association Industry Interim Standards TIA/EIA IS-127 and TIA/EIA IS-733.

The pitch estimation module 104 produces a pitch index I_p and a lag value P_0 based upon each input speech frame $s(n)$. The LP analysis module 106 performs linear predictive analysis on each input speech frame $s(n)$ to generate an LP parameter a . The LP parameter a is provided to the LP quantization module 110. The LP quantization module 110 also receives the mode M , thereby performing the quantization process in a mode-dependent

manner. The LP quantization module 110 produces an LP index I_{LP} and a quantized LP parameter \hat{a} . The LP analysis filter 108 receives the quantized LP parameter \hat{a} in addition to the input speech frame $s(n)$. The LP analysis filter 108 generates an LP residue signal $R[n]$, which represents the error
 5 between the input speech frames $s(n)$ and the reconstructed speech based on the quantized linear predicted parameters \hat{a} . The LP residue $R[n]$, the mode M , and the quantized LP parameter \hat{a} are provided to the residue quantization module 112. Based upon these values, the residue quantization module 112 produces a residue index I_R and a quantized residue
 10 signal $\hat{R}[n]$.

In FIG. 3 a decoder 200 that may be used in a speech coder includes an LP parameter decoding module 202, a residue decoding module 204, a mode decoding module 206, and an LP synthesis filter 208. The mode decoding module 206 receives and decodes a mode index I_M , generating therefrom a
 15 mode M . The LP parameter decoding module 202 receives the mode M and an LP index I_{LP} . The LP parameter decoding module 202 decodes the received values to produce a quantized LP parameter \hat{a} . The residue decoding module 204 receives a residue index I_R , a pitch index I_p , and the mode index I_M . The residue decoding module 204 decodes the received values to
 20 generate a quantized residue signal $\hat{R}[n]$. The quantized residue signal $\hat{R}[n]$ and the quantized LP parameter \hat{a} are provided to the LP synthesis filter 208, which synthesizes a decoded output speech signal $\hat{s}[n]$ therefrom.

Operation and implementation of the various modules of the encoder 100 of FIG. 2 and the decoder 200 of FIG. 3 are known in the art and described
 25 in the aforementioned U.S. Patent No. 5,414,796 and L.B. Rabiner & R.W. Schafer, *Digital Processing of Speech Signals* 396-453 (1978).

As illustrated in the flow chart of FIG. 4, a speech coder in accordance with one embodiment follows a set of steps in processing speech samples for transmission. In step 300 the speech coder receives digital samples of a
 30 speech signal in successive frames. Upon receiving a given frame, the speech coder proceeds to step 302. In step 302 the speech coder detects the energy of the frame. The energy is a measure of the speech activity of the

frame. Speech detection is performed by summing the squares of the amplitudes of the digitized speech samples and comparing the resultant energy against a threshold value. In one embodiment the threshold value adapts based on the changing level of background noise. An exemplary
5 variable threshold speech activity detector is described in the aforementioned U.S. Patent No. 5,414,796. Some unvoiced speech sounds can be extremely low-energy samples that may be mistakenly encoded as background noise. To prevent this from occurring, the spectral tilt of low-energy samples may be used to distinguish the unvoiced speech from
10 background noise, as described in the aforementioned U.S. Patent No. 5,414,796.

After detecting the energy of the frame, the speech coder proceeds to step 304. In step 304 the speech coder determines whether the detected frame energy is sufficient to classify the frame as containing speech information. If
15 the detected frame energy falls below a predefined threshold level, the speech coder proceeds to step 306. In step 306 the speech coder encodes the frame as background noise (i.e., nonspeech, or silence). In one embodiment the background noise frame is encoded at 1/8 rate, or 1 kbps. If in step 304 the detected frame energy meets or exceeds the predefined threshold level,
20 the frame is classified as speech and the speech coder proceeds to step 308.

In step 308 the speech coder determines whether the frame is unvoiced speech, i.e., the speech coder examines the periodicity of the frame. Various known methods of periodicity determination include, e.g., the use of zero crossings and the use of normalized autocorrelation functions
25 (NACFs). In particular, using zero crossings and NACFs to detect periodicity is described in U.S. Application Serial No. 08/815,354, entitled METHOD AND APPARATUS FOR PERFORMING REDUCED RATE VARIABLE RATE VOCODING, filed March 11, 1997, assigned to the assignee of the present invention, and fully incorporated herein by reference. In addition,
30 the above methods used to distinguish voiced speech from unvoiced speech are incorporated into the Telecommunication Industry Association Interim Standards TIA/EIA IS-127 and TIA/EIA IS-733. If the frame is determined to be unvoiced speech in step 308, the speech coder proceeds to step 310. In step

310 the speech coder encodes the frame as unvoiced speech. In one embodiment unvoiced speech frames are encoded at quarter rate, or 2.6 kbps. If in step 308 the frame is not determined to be unvoiced speech, the speech coder proceeds to step 312.

5 In step 312 the speech coder determines whether the frame is transitional speech, using periodicity detection methods that are known in the art, as described in, e.g., the aforementioned U.S. Application Serial No. 08/815,354. If the frame is determined to be transitional speech, the speech coder proceeds to step 314. In step 314 the frame is encoded as transition
10 speech (i.e., transition from unvoiced speech to voiced speech). In one embodiment the transition speech frame is encoded in accordance with a multipulse interpolative coding method described below with reference to FIG. 6.

If in step 312 the speech coder determines that the frame is not
15 transitional speech, the speech coder proceeds to step 316. In step 316 the speech coder encodes the frame as voiced speech. In one embodiment voiced speech frames may be encoded at full rate, or 13.2 kbps.

Those of skill would appreciate that either the speech signal or the corresponding LP residue may be encoded by following the steps shown in
20 FIG. 4. The waveform characteristics of noise, unvoiced, transition, and voiced speech can be seen as a function of time in the graph of FIG. 5A. The waveform characteristics of noise, unvoiced, transition, and voiced LP residue can be seen as a function of time in the graph of FIG. 5B.

In one embodiment a speech coder uses a multipulse interpolative
25 coding algorithm to code transition speech frames in accordance with the method steps illustrated in the flow chart of FIG. 6. In step 400 the speech coder estimates the pitch period M of the current K -sample LP speech residue frame $S[n]$, where $n=1,2,...,K$, and the immediate future neighborhood of the frame $S[n]$. In one embodiment the LP speech residue
30 frame $S[n]$ comprises 160 samples (i.e., $K=160$). The pitch period M is a fundamental period that repeats within a given frame. The speech coder then proceeds to step 402. In step 402 the speech coder extracts a pitch prototype X having the last M samples of the current residue frame. The

pitch prototype X may advantageously be the final pitch period (M samples) of the frame $S[n]$. In the alternative, the pitch prototype X may be any pitch period M of the frame $S[n]$. The speech coder then proceeds to step 404.

In step 404 the speech coder selects N important samples, or pulses, having amplitudes Q_i and signs S_i , where $i=1,2,\dots,N$, from positions P_i from the M -sample, pitch prototype X . Thus, N "best" samples have been selected from the M -sample pitch prototype X , and $M-N$ unselected samples remain in the pitch prototype X . The speech coder then proceeds to step 406. In step 406 the speech coder encodes the positions of the pulses with B_p bits. The speech coder then proceeds to step 408. In step 408 the speech coder encodes the signs of the pulses with B_s bits. The speech coder then proceeds to step 410. In step 410 the speech coder encodes the amplitudes of the pulses with B_a bits. The quantized values of the N pulse amplitudes Q_i are denoted Z_i , for $i=1,2,\dots,N$. The speech coder then proceeds to step 412.

In step 412 the speech coder extracts the pulses. In one embodiment the pulse extraction step is performed by ordering all of the M pulses according to absolute (i.e., unsigned) amplitude, and then choosing the N highest pulses (i.e., the N pulses having the greatest absolute amplitudes). In an alternate embodiment the pulse extraction step selects the N "best" pulses from the standpoint of perceptual importance, in accordance with the following description.

As illustrated in FIG. 7, a speech signal may be converted from the LP residue domain to the speech domain by filtering. Conversely, the speech signal may be converted from the speech domain to the LP residue domain by inverse filtering. In accordance with one embodiment, as shown in FIG. 7, a pitch prototype X is input to a first LP synthesis filter 500, denoted $H(z)$. The first LP synthesis filter 500 produces a perceptually weighted speech-domain version of the pitch prototype X , denoted $S(n)$. A shape codebook 502 produces shape vector values, which are provided to a multiplier 504. A gain codebook 506 produces gain vector values, which are also provided to the multiplier 504. The multiplier 504 multiplies the shape vector values with the gain vector values, producing shape-gain product values. The shape-gain product values are provided to a first adder 508. A number, N , of

pulses (the number N , as described below, is the number of samples that minimizes the shape-gain error, E , between the pitch prototype X and a model prototype $e_mod[n]$) is also provided to the first adder 508. The first adder 508 adds the N pulses to the shape-gain product values, producing a
 5 model prototype $e_mod[n]$. The model prototype $e_mod[n]$ is provided to a second LP synthesis filter 510, also denoted $H(z)$. The second LP synthesis filter 510 produces a perceptually weighted speech-domain version of the model prototype $e_mod[n]$, denoted $Se(n)$. The speech-domain values $S(n)$ and $Se(n)$ are provided to a second adder 512. The second adder 512 subtracts
 10 $S(n)$ from $Se(n)$, providing difference values to a sum-of-squares calculator 514. The sum-of-squares calculator 514 computes the squares of the difference values, producing an energy, or error, value E .

In accordance with the alternate embodiment mentioned above with reference to FIG. 6, the impulse response for an LP synthesis filter $H(z)$ (not
 15 shown), or a perceptually weighted LP synthesis filter $H(z/\alpha)$, for the current transition speech frame is denoted $H(n)$. The model of the pitch prototype X is denoted $e_mod[n]$. A perceptually weighted speech domain error E may be defined in accordance with the following equation:

$$E = \sum_{n=1}^M (Se(n) - S(n))^2$$

20 where

$$Se(n) = H(n) * e_mod(n),$$

and

$$S(n) = H(n) * X,$$

25 where "*" denotes a suitable filtering or convolution operation, as known in the art, and $Se(n)$ and $S(n)$ denote perceptually weighted speech domain versions of the pitch prototypes $e_mod[n]$ and X , respectively. In the alternate embodiment described, the N best samples may be selected to form $e_mod[n]$ from the M samples of the pitch prototype X as follows: The N
 30 samples, which may be denoted the j -th set out of a possible MC_N combinations, are advantageously chosen to create the model $e_mod_j(n)$

such that the error E_j is minimized for all j belonging to $j=1,2,3,...,^MC_N$, where E_j is defined in accordance with the following equations:

$$E_j = \sum_{n=1}^M (Se_j(n) - S(n))^2$$

and

$$5 \quad Se_j(n) = H(n) * e_modj[n].$$

After extracting the pulses, the speech coder proceeds to step 414. In step 414 the remaining M-N samples of the pitch prototype X are represented in accordance with one of two possible methods associated with alternate embodiments. In one embodiment the remaining M-N samples of the pitch prototype X may be selected by replacing the M-N samples with zero values. In an alternate embodiment, the remaining M-N samples of the pitch prototype X may be selected by replacing the M-N samples with a shape vector using a codebook with Rs bits and a gain using a codebook with Rg bits. Accordingly, a gain g and a shape vector H represent the M-N samples.

10 The gain g and the shape vector H have component values g_j and H_k chosen from the codebooks by minimizing the distortion E_{jk} . The distortion E_{jk} is given by the following equations:

$$E_{jk} = \sum_{n=1}^M (Se_{jk}(n) - S(n))^2$$

and

$$20 \quad Se_{jk}(n) = H(n) * e_mod_{jk}[n],$$

where the model prototype $e_mod_{jk}[n]$ is formed with the M pulses described above and M-N samples represented by the j-th gain codeword g_j and the k-th shape code-word H_k . The selection may thus advantageously be performed in a jointly optimized way by selecting the combination $\{j,k\}$ that delivers the minimal value of E_{jk} . The speech coder then proceeds to step 416.

In step 416 the coded pitch prototype Y is computed. The coded pitch prototype Y models the original pitch prototype X by placing the N pulses back in the positions P_i , replacing the amplitudes Q_i with $S_i * Z_i$, and replacing the remaining M-N samples with either zeros (in one embodiment) or the samples from the chosen gain-shape representation,

30

g^*H , as described above (in an alternate embodiment). The coded pitch prototype Y corresponds to the sum of the reconstructed, or synthesized, N "best" samples plus the reconstructed, or synthesized, remaining $M-N$ samples. The speech coder then proceeds to step 418.

5 In step 418 the speech coder extracts an M -sample "past prototype" W from the past (i.e., immediately preceding) decoded residue frame. The past prototype W is extracted by taking the last M samples from the past decoded residue frame. Alternatively, the past prototype W could be constructed from another set of M samples of the past frame, provided the pitch
10 prototype X was taken from a corresponding set of M samples of the current frame. The speech coder then proceeds to step 420.

In step 420 the speech coder reconstructs the entire K samples of the decoded current frame of residue $S_{\text{SYNTH}}[n]$. The reconstruction is advantageously accomplished with any conventional interpolation method
15 in which the last M samples are formed with the reconstructed pitch prototype Y , and the first $K-M$ samples are formed by interpolating the past prototype W and the current coded pitch prototype Y . In one embodiment the interpolation may be performed in accordance with the following steps:

W and Y are first advantageously aligned to derive the optimal
20 relative positioning and the average pitch period to be used for interpolation. The alignment A^* is obtained as the rotation of the current pitch prototype Y that corresponds to the maximum cross-correlation of the rotated Y with W . The cross-correlations $C[A]$ at each possible alignment A , taking values from 0 to $M-1$ or a subset of the range 0 to $M-1$, may in turn be
25 computed in accordance with the following equation:

$$C[A] = \sum_{n=0}^{M-1} Y[(n+A)\%M]W$$

The average pitch period L_{av} is then computed in accordance with the following equation:

$$L_{av} = (160-M)M/(MNp-A^*),$$

30 where

$$Np = \text{round}\{A^*/M + (160-M)/M\}.$$

An interpolation is performed to compute the first K-M samples in accordance with the following equation:

$$S_{\text{SYNTH}} = \{(160-n-M)W[(n\alpha)\%M] + nY[(n\alpha+A^*)\%M]\}/(160-M),$$

where $\alpha = M/L_{av}$, and the sample at non-integral values for the indices n'

5 (which are equal to either $n\alpha$ or $n\alpha+A^*$) are computed using a conventional interpolation method depending upon the desired accuracy in the fractional value of n' . The round operation and the modulo operation (denoted by the % symbol) in the above equations are well known in the art. Graphs of original transitional speech, uncoded residue, coded/quantized residue, and
10 decoded/reconstructed speech with respect to time are depicted in, respectively, FIGS. 8A-D.

In one embodiment the encoded transition residue frame may be computed in accordance with a closed-loop technique. Accordingly, the encoded transition residue frame is computed, as described above. Then the
15 perceptual signal-to-noise ratio (PSNR) is computed for the entire frame. If the PSNR rises above a predefined threshold value, then a suitable high-rate, high-precision, waveform coding method such as CELP may be used to encode the frame. Such a technique is described in U.S. Application Serial No. 09/259,151, filed February 26, 1999, entitled CLOSED-LOOP MULTIMODE
20 MIXED-DOMAIN LINEAR PREDICTION (MDLP) SPEECH CODER, and assigned to the assignee of the present invention. By using the low-bit-rate speech coding method described above when possible, and substituting a high-rate CELP speech coding method when the low-bit-rate speech coding method fails to deliver a target value of the distortion measure, transition
25 speech frames can be coded with a relatively high quality (as determined by the threshold value or the distortion measure used) while using a low average coding rate.

Thus, a novel multipulse interpolative coder for transition speech frames has been described. Those of skill in the art would understand that
30 the various illustrative logical blocks and algorithm steps described in connection with the embodiments disclosed herein may be implemented or performed with a digital signal processor (DSP), an application specific

integrated circuit (ASIC), discrete gate or transistor logic, discrete hardware components such as, e.g., registers and FIFO, a processor executing a set of firmware instructions, or any conventional programmable software module and a processor. The processor may advantageously be a microprocessor, but
5 in the alternative, the processor may be any conventional processor, controller, microcontroller, or state machine. The software module could reside in RAM memory, flash memory, registers, or any other form of writable storage medium known in the art. Those of skill would further appreciate that the data, instructions, commands, information, signals, bits,
10 symbols, and chips that may be referenced throughout the above description are advantageously represented by voltages, currents, electromagnetic waves, magnetic fields or particles, optical fields or particles, or any combination thereof.

Preferred embodiments of the present invention have thus been
15 shown and described. It would be apparent to one of ordinary skill in the art, however, that numerous alterations may be made to the embodiments herein disclosed without departing from the spirit or scope of the invention. Therefore, the present invention is not to be limited except in accordance with the following claims.

20

We Claim:

CLAIMS

1. A method of coding transitional speech frames, comprising the
2 steps of:
representing a first frame of transitional speech samples by a
4 first subset of the samples of the first frame; and
interpolating the first subset of samples and a second subset of
6 samples extracted from a second, earlier-received frame of transitional
speech samples to synthesize other samples of the first frame that are not
8 included in the first subset.

2. The method of claim 1, further comprising the steps of
2 transmitting the first subset of samples after performing the representing
step, and receiving the first subset of samples before performing the
4 interpolating step.

3. The method of claim 1, further comprising the step of
2 simplifying the first subset of samples.

4. The method of claim 3, wherein the simplifying step comprises
2 the steps of selecting perceptually significant samples from the first subset of
samples, and assigning a zero value to all unselected samples.

5. The method of claim 3, wherein the simplifying step comprises
2 the steps of selecting samples with relatively high absolute amplitudes from
the first subset of samples, and assigning a zero value to all unselected
4 samples.

6. The method of claim 4, wherein the perceptually significant
2 samples are samples selected to minimize perceptually weighted speech-
domain error between the first frame of transitional speech samples and a
4 synthesized first frame of transitional speech samples.

7. The method of claim 3, wherein the simplifying step comprises
2 the steps of selecting perceptually significant samples from the first subset of
samples, and quantizing a portion of all unselected samples.

8. The method of claim 3, wherein the simplifying step comprises
2 the steps of selecting samples with relatively high absolute amplitudes from
the first subset of samples, and quantizing a portion of all unselected
4 samples.

9. The method of claim 7, wherein the perceptually significant
2 samples are samples selected to minimize gain and shape error between the
first frame of transitional speech samples and a synthesized first frame of
4 transitional speech samples.

10. A speech coder for coding transitional speech frames,
2 comprising:
means for representing a first frame of transitional speech
4 samples by a first subset of the samples of the first frame; and
means for interpolating the first subset of samples and a second
6 subset of samples extracted from a second, earlier-received frame of
transitional speech samples to synthesize other samples of the first frame
8 that are not included in the first subset.

11. The speech coder of claim 10, further comprising means for
2 simplifying the first subset of samples.

12. The speech coder of claim 11, wherein the means for
2 simplifying comprises means for selecting perceptually significant samples
from the first subset of samples, and means for assigning a zero value to all
4 unselected samples.

13. The speech coder of claim 11, wherein the means for
2 simplifying comprises means for selecting samples with relatively high

absolute amplitudes from the first subset of samples, and means for
4 assigning a zero value to all unselected samples.

14. The speech coder of claim 12, wherein the perceptually
2 significant samples are samples selected to minimize perceptually weighted
speech-domain error between the first frame of transitional speech samples
4 and a synthesized first frame of transitional speech samples.

15. The speech coder of claim 11, wherein the means for
2 simplifying comprises means for selecting perceptually significant samples
from the first subset of samples, and means for quantizing a portion of all
4 unselected samples.

16. The speech coder of claim 11, wherein the means for
2 simplifying comprises means for selecting samples with relatively high
absolute amplitudes from the first subset of samples, and means for
4 quantizing a portion of all unselected samples.

17. The speech coder of claim 15, wherein the perceptually
2 significant samples are samples selected to minimize gain and shape error
between the first frame of transitional speech samples and a synthesized first
4 frame of transitional speech samples.

18. A speech coder for coding transitional speech frames,
2 comprising:

an extractor configured to represent a first frame of transitional
4 speech samples by a first subset of the samples of the first frame; and

an interpolator coupled to the extractor and configured to
6 interpolate the first subset of samples and a second subset of samples
extracted from a second, earlier-received frame of transitional speech
8 samples to synthesize other samples of the first frame that are not included
in the first subset.

19. The speech coder of claim 18, further comprising a pulse
2 selector configured to select perceptually significant samples from the first
subset of samples, wherein a zero value is assigned to all unselected samples.

20. The speech coder of claim 18, further comprising a pulse
2 selector configured to select samples with relatively high absolute
amplitudes from the first subset of samples, wherein a zero value is assigned
4 to all unselected samples.

21. The speech coder of claim 19, wherein the perceptually
2 significant samples are samples selected to minimize perceptually weighted
speech-domain error between the first frame of transitional speech samples
4 and a synthesized first frame of transitional speech samples.

22. The speech coder of claim 18, further comprising a pulse
2 selector configured to select perceptually significant samples from the first
subset of samples, wherein a portion of all unselected samples is quantized.

23. The speech coder of claim 18, further comprising a pulse
2 selector configured to select samples with relatively high absolute
amplitudes from the first subset of samples, wherein a portion of all
4 unselected samples is quantized.

24. The speech coder of claim 22, wherein the perceptually
2 significant samples are samples selected to minimize gain and shape error
between the first frame of transitional speech samples and a synthesized first
4 frame of transitional speech samples.

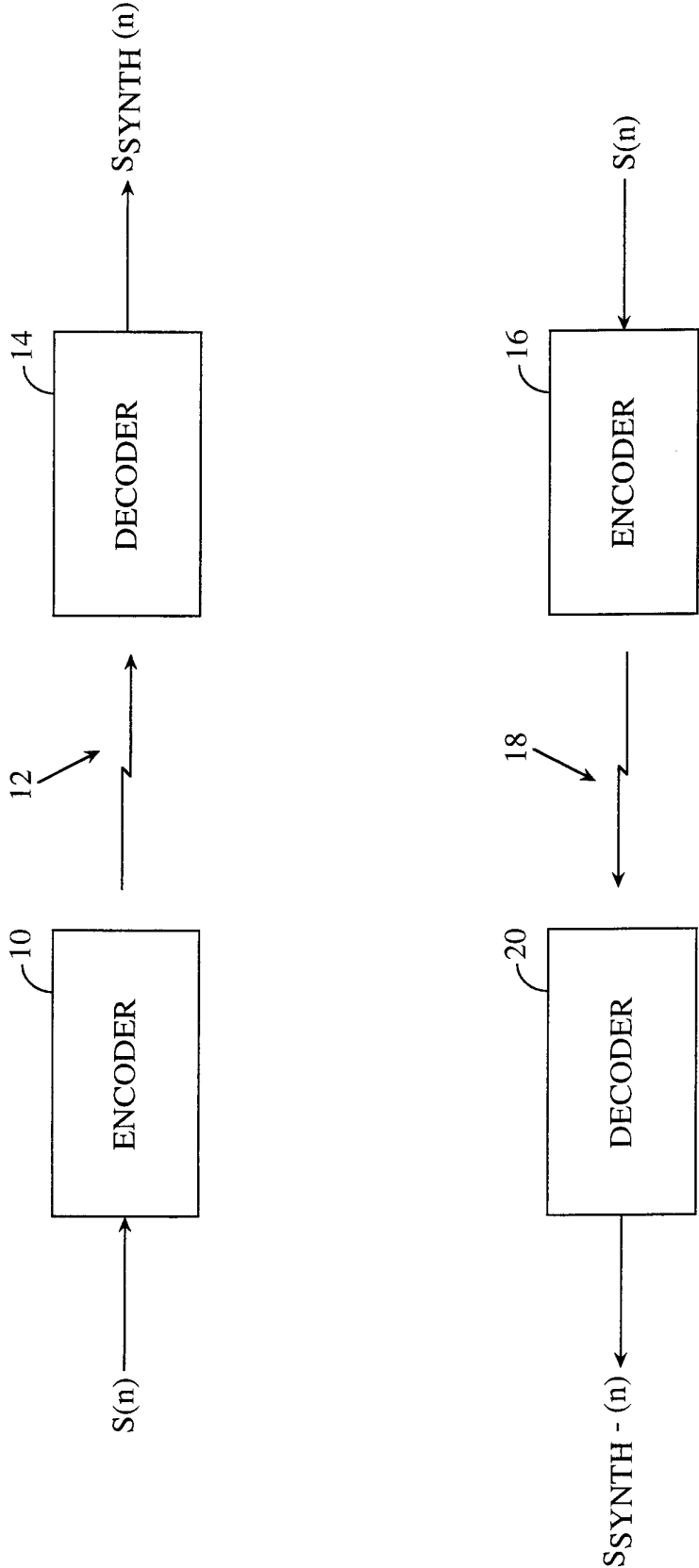


FIG. 1

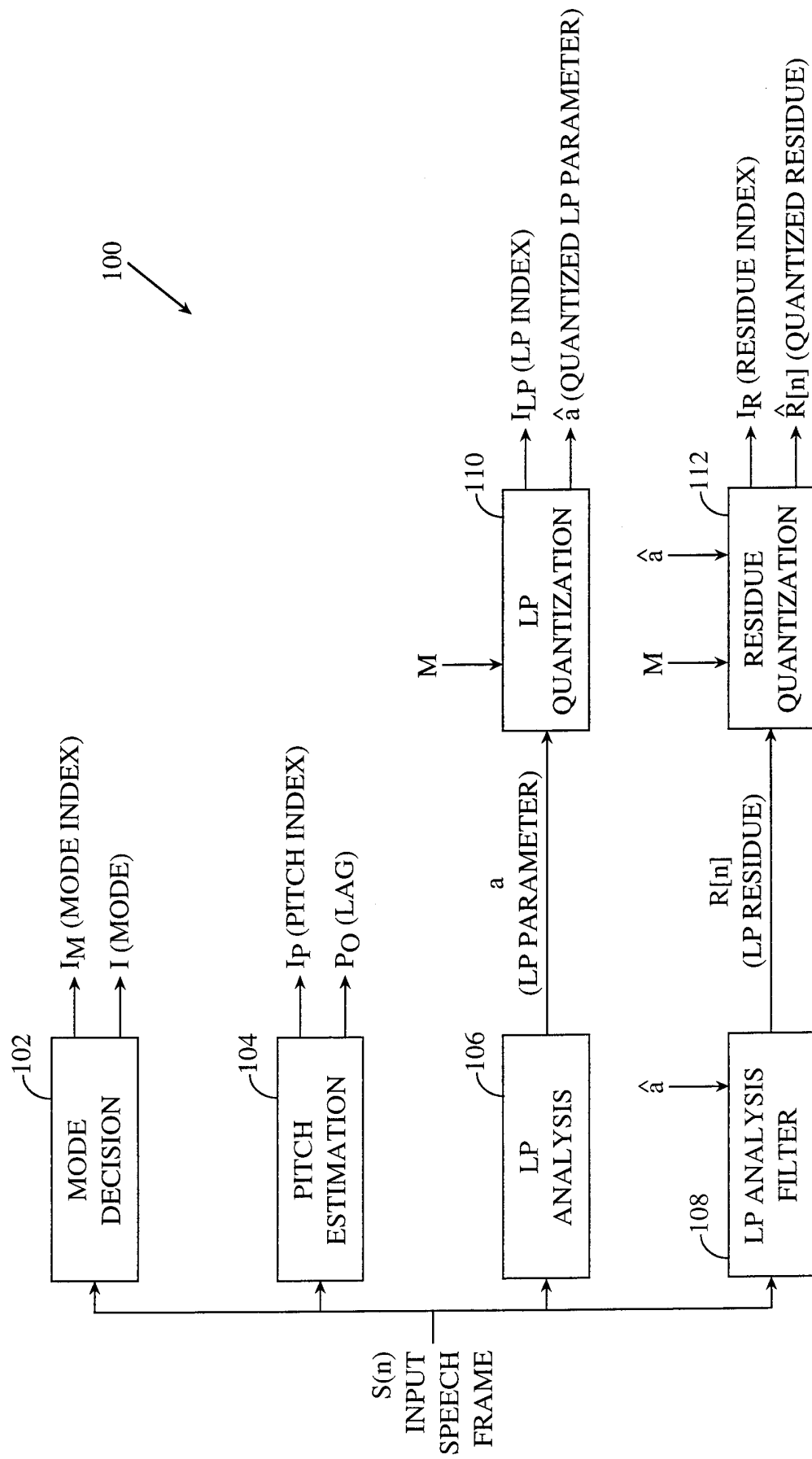


FIG. 2

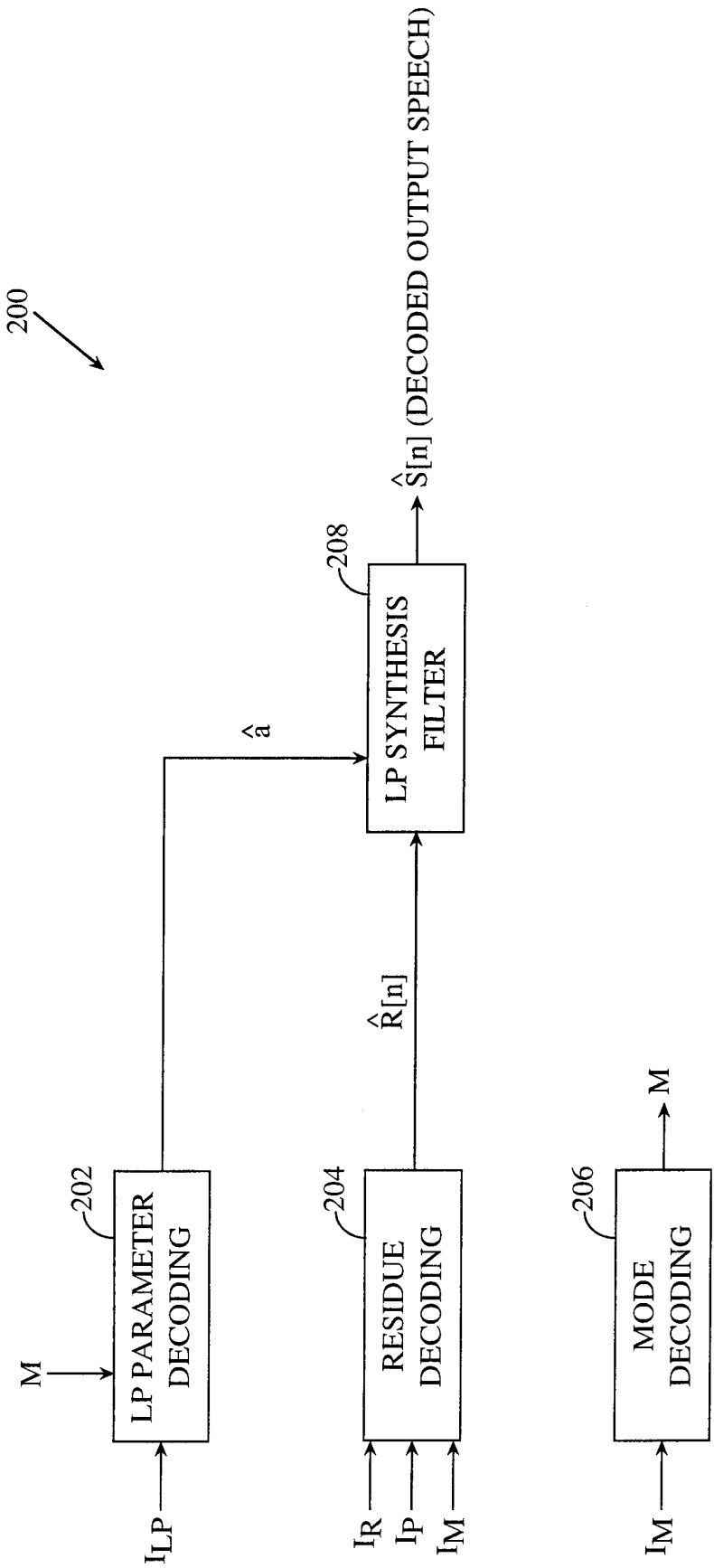


FIG. 3

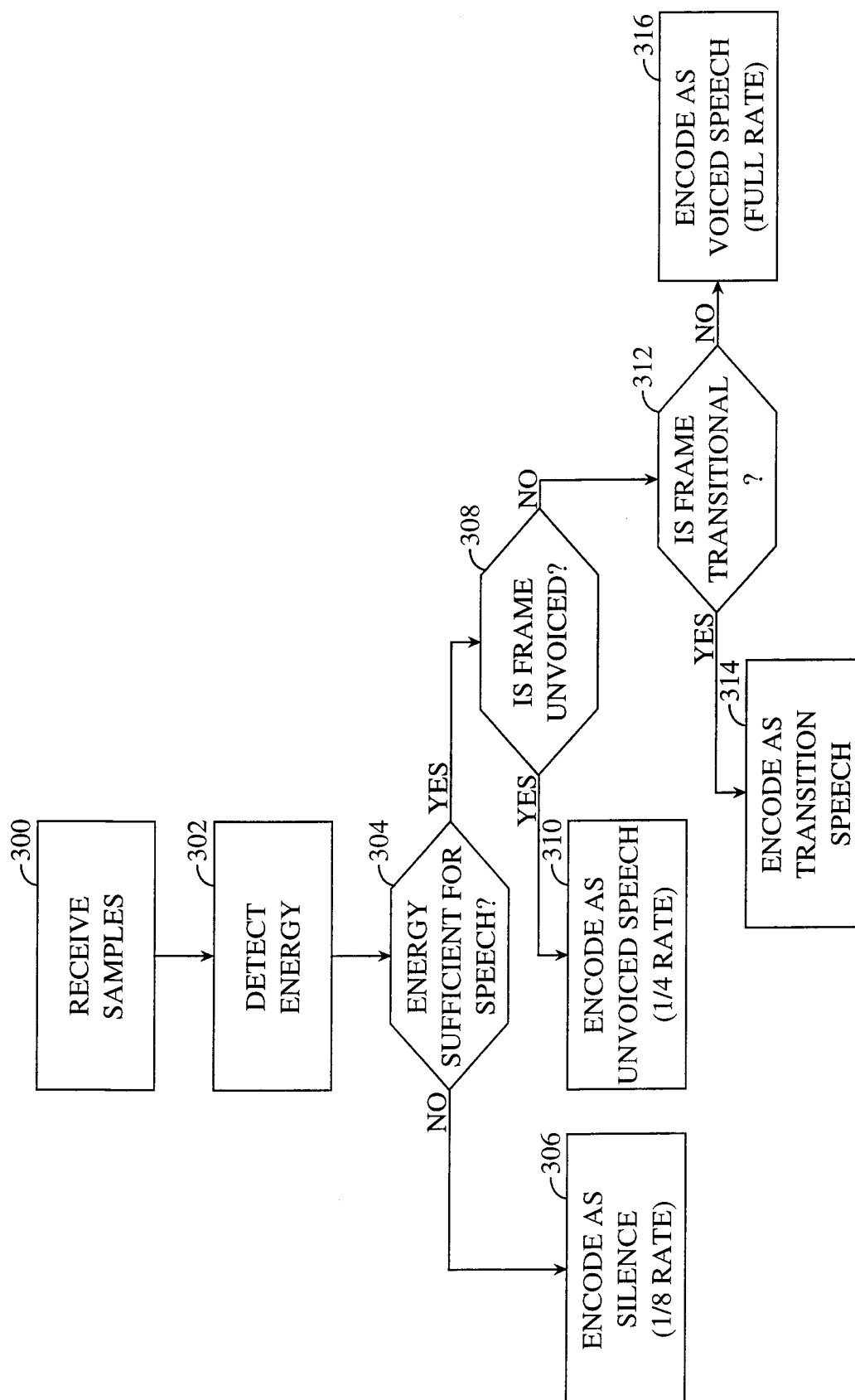


FIG. 4

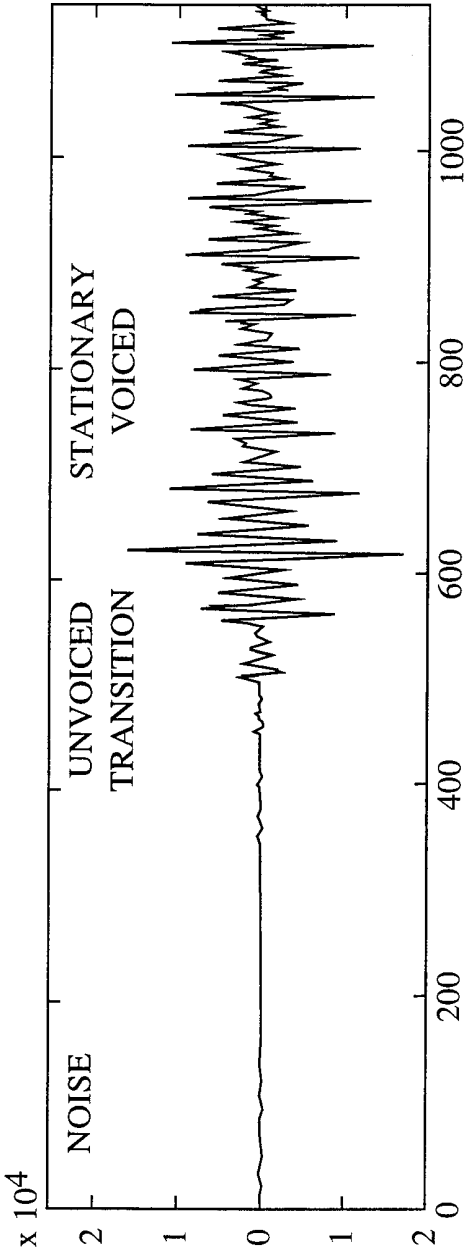


FIG. 5A

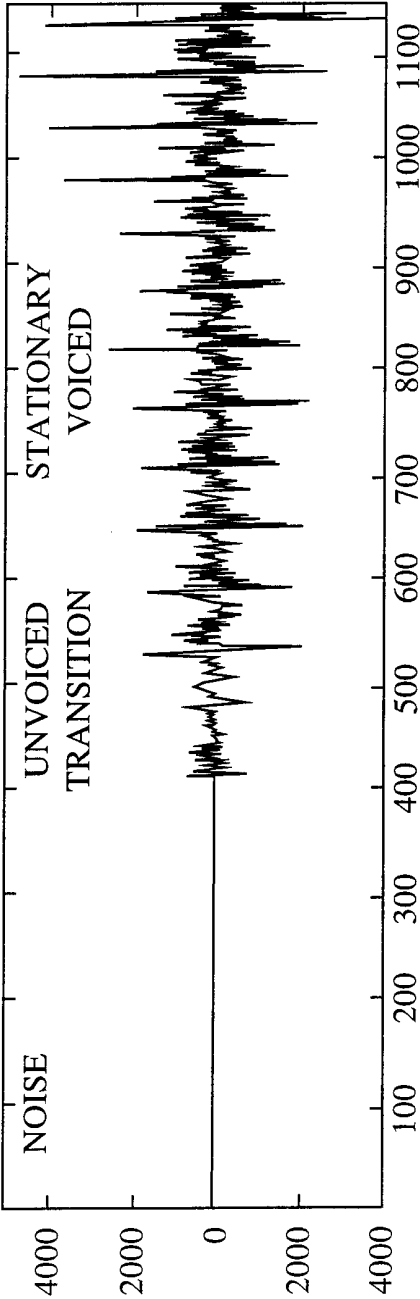


FIG. 5B

6/8

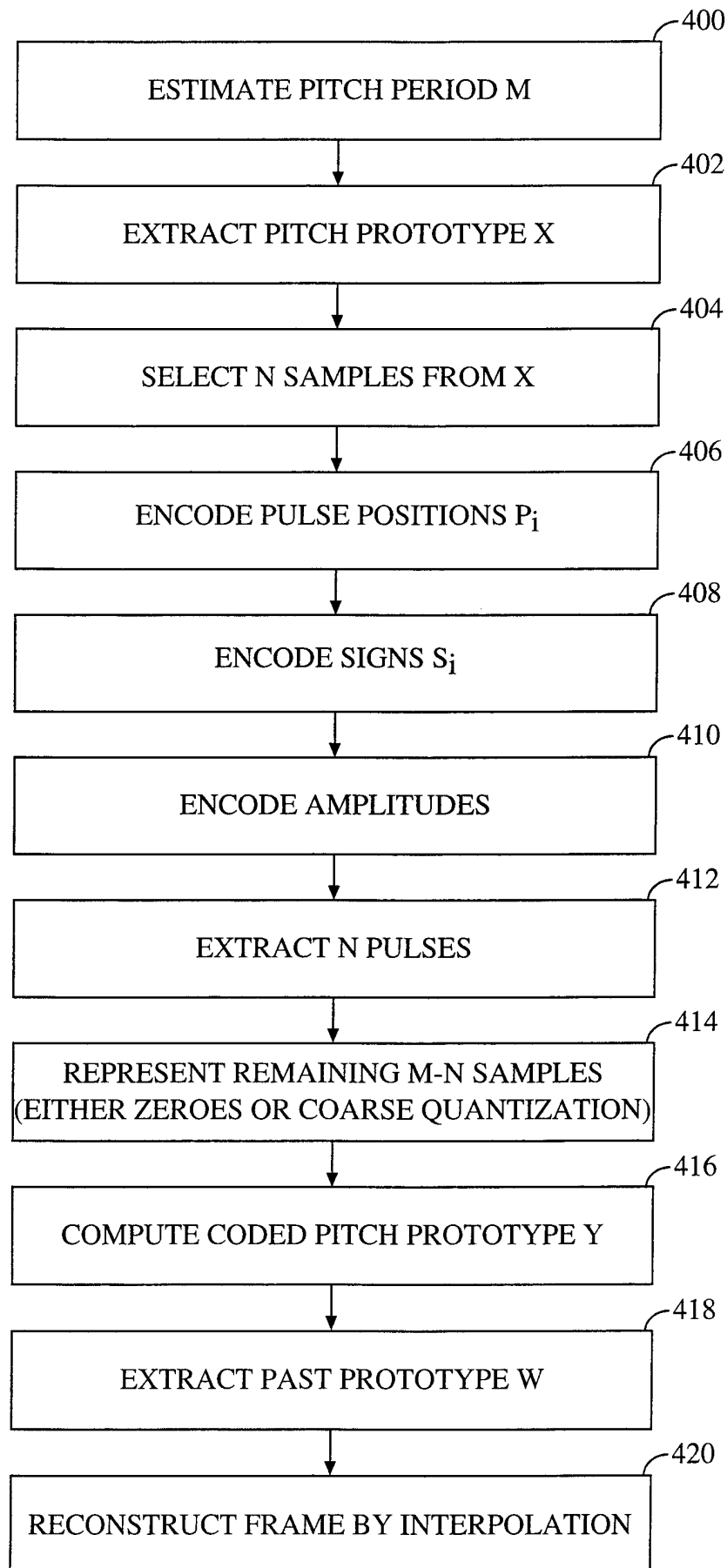


FIG. 6

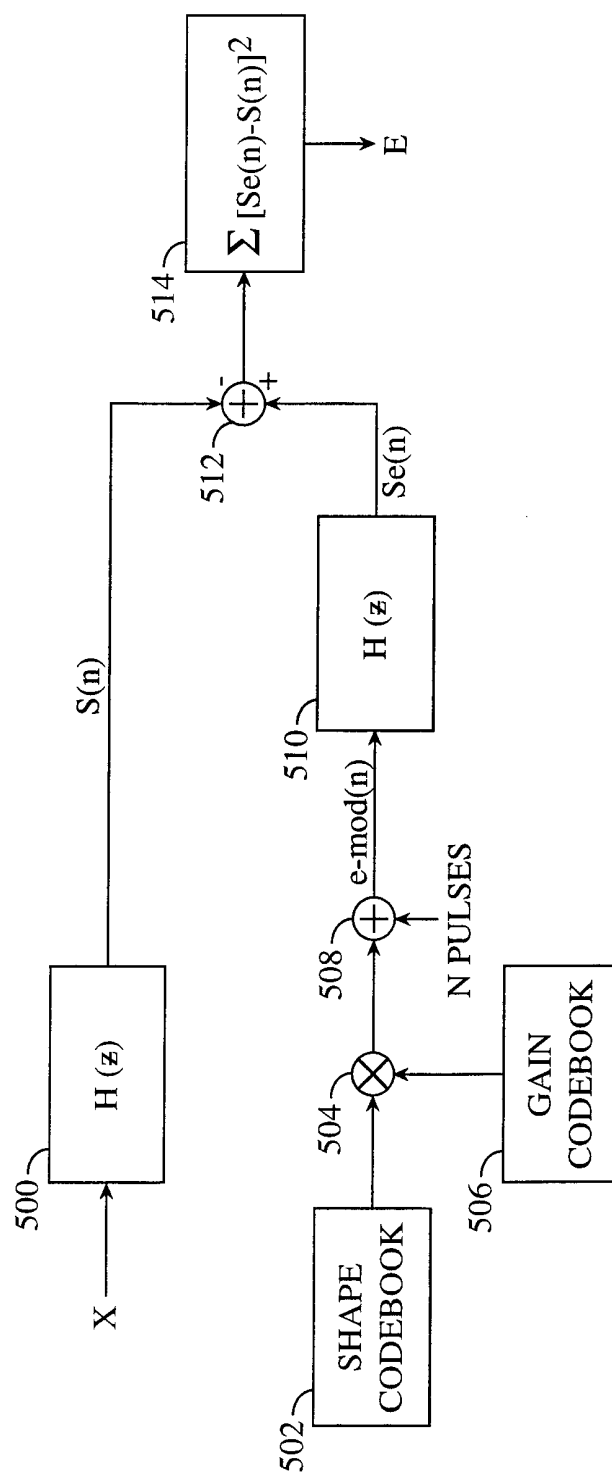


FIG. 7

8/8

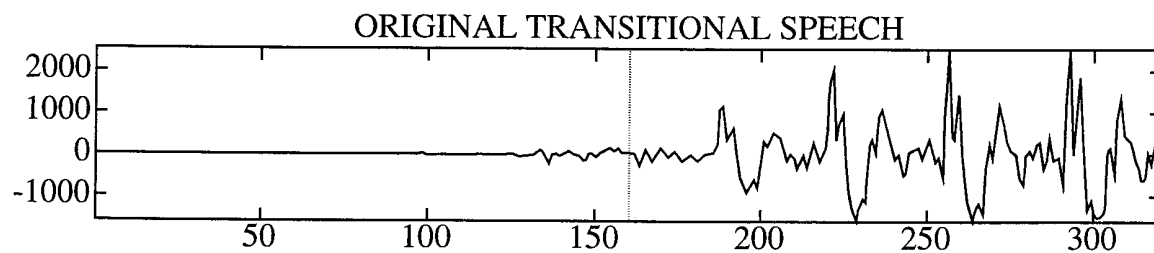


FIG. 8A

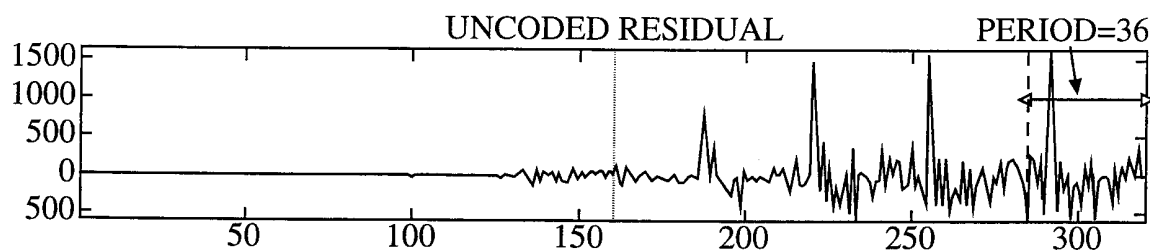


FIG. 8B

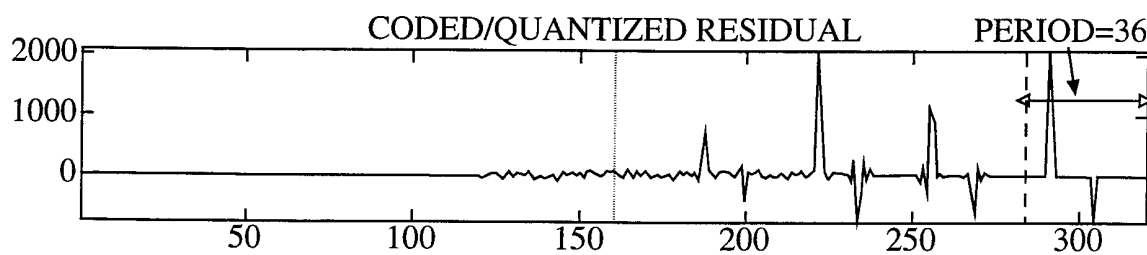


FIG. 8C

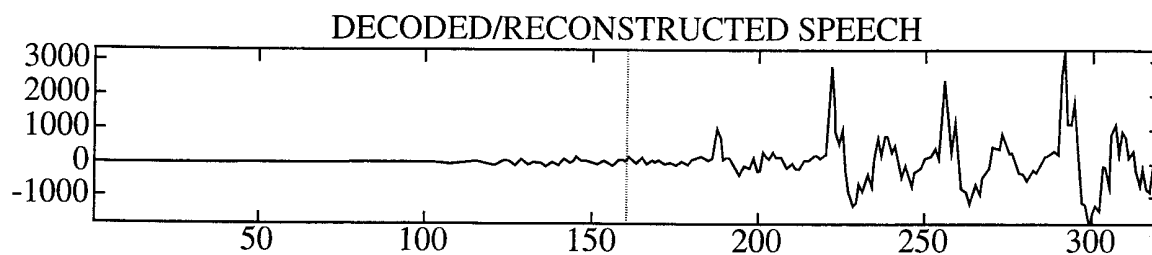


FIG. 8D

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/12656

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G10L19/10 G10L19/14

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4 821 324 A (OZAWA KAZUNORI ET AL) 11 April 1989 (1989-04-11) abstract column 3, line 10 - line 14 column 7, line 6 - line 23 column 7, line 52 - line 65 column 9, line 39 - line 68 column 11, line 55 - column 12, line 6 column 14, line 29 - line 39 ---	1, 10, 18
A	US 5 119 424 A (ASAKAWA YOSHIAKI ET AL) 2 June 1992 (1992-06-02) abstract column 2, line 27 - line 39 column 3, line 64 - line 67 column 5, line 7 - line 23 column 6, line 51 - line 55 --- -/--	1, 10, 18

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

30 August 2000

Date of mailing of the international search report

08/09/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Ramos Sánchez, U

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/12656

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>US 4 945 565 A (ARASEKI TAKASHI ET AL) 31 July 1990 (1990-07-31) column 2, line 55 - line 58 column 7, line 14 - line 60 column 8, line 62 - line 66 column 9, line 52 - column 10, line 8 column 11, line 36 - line 61 column 13, line 22 - line 59 column 14, line 38 - line 43</p> <p style="text-align: center;">---</p>	1,10,18
A	<p>PATENT ABSTRACTS OF JAPAN vol. 013, no. 511 (P-961), 16 November 1989 (1989-11-16) -& JP 01 207800 A (NEC CORP), 21 August 1989 (1989-08-21) abstract</p> <p style="text-align: center;">-----</p>	1,10,18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/12656

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 4821324 A	11-04-1989	JP 2615548 B	28-05-1997
		JP 62038500 A	19-02-1987
		JP 2844590 B	06-01-1999
		JP 61150000 A	08-07-1986
		CA 1252568 A	11-04-1989
US 5119424 A	02-06-1992	JP 1155400 A	19-06-1989
		JP 2707564 B	28-01-1998
US 4945565 A	31-07-1990	JP 1903425 C	08-02-1995
		JP 6032035 B	27-04-1994
		JP 61019000 A	27-01-1986
		JP 1976594 C	17-10-1995
		JP 7007275 B	30-01-1995
		JP 61020997 A	29-01-1986
		CA 1255802 A	13-06-1989
JP 01207800 A	21-08-1989	NONE	

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/US 00/12704

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
DE 19604084 A	02-10-1996	AT 187835 T AU 5268796 A CZ 9703004 A WO 9629688 A DE 59603925 D EP 0815547 A ES 2142053 T HU 9801653 A PL 324636 A US 6012012 A DE 19604083 A	15-01-2000 08-10-1996 13-05-1998 26-09-1996 20-01-2000 07-01-1998 01-04-2000 28-10-1998 08-06-1998 04-01-2000 24-10-1996
WO 9607110 A	07-03-1996	AU 684228 B AU 3393195 A AU 695816 B AU 5538998 A CA 2198886 A DE 69506563 D DE 69506563 T EP 0777863 A EP 0837341 A ES 2126931 T FI 970848 A JP 10505420 T NO 970940 A NZ 292210 A NZ 329889 A	04-12-1997 22-03-1996 20-08-1998 23-04-1998 07-03-1996 21-01-1999 06-05-1999 11-06-1997 22-04-1998 01-04-1999 28-02-1997 26-05-1998 28-02-1997 28-07-1998 28-05-1999