Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2300077 C 2007/09/04

(11)(21) 2 300 077

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 1999/06/08
(87) Date publication PCT/PCT Publication Date: 1999/12/16
(45) Date de délivrance/Issue Date: 2007/09/04
(85) Entrée phase nationale/National Entry: 2000/02/08
(86) N° demande PCT/PCT Application No.: JP 1999/003064
(87) N° publication PCT/PCT Publication No.: 1999/065017
(30) Priorités/Priorities: 1998/06/09 (JP10/160119);
1998/09/11 (JP10/258271)

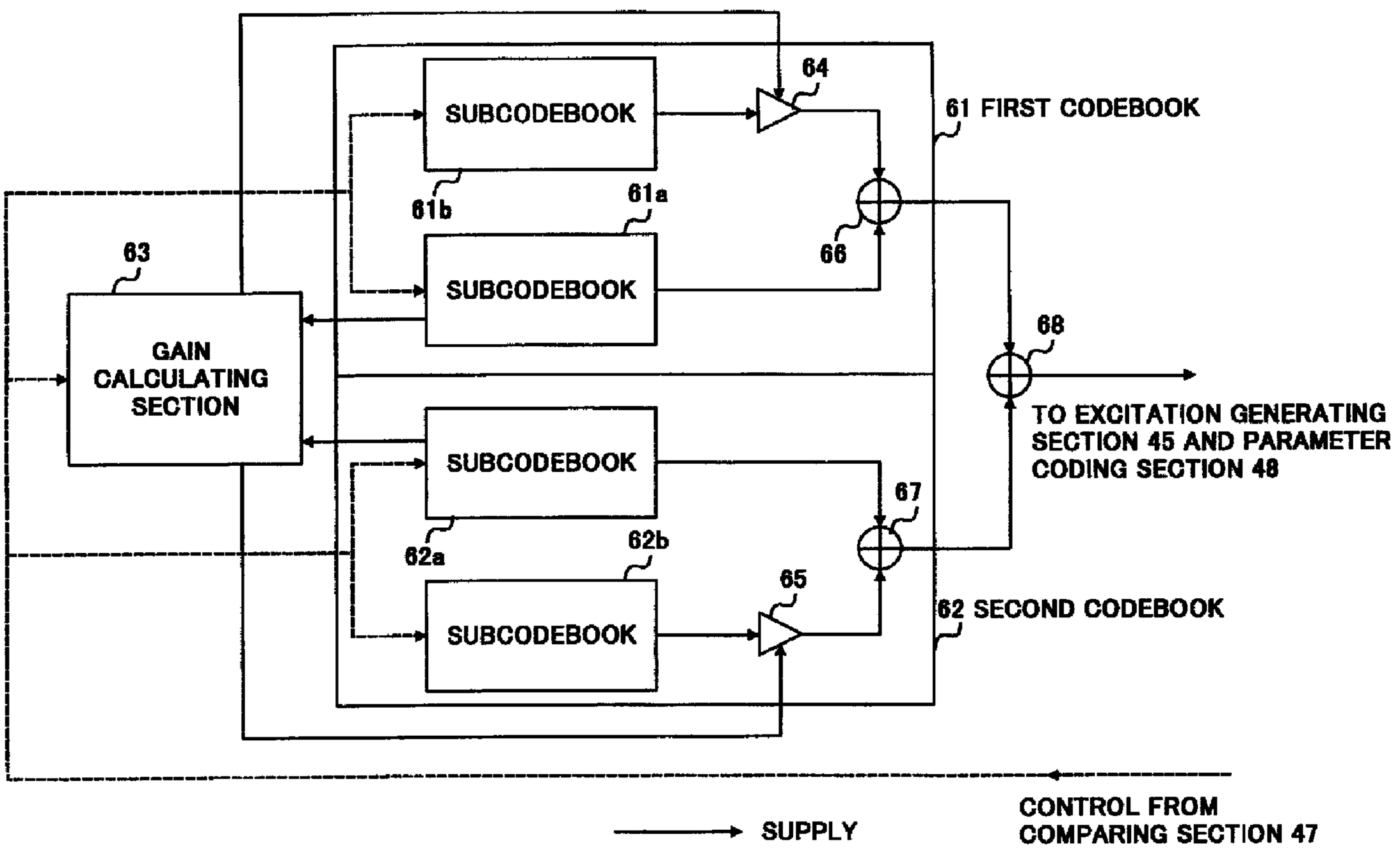
(51) Cl.Int./Int.Cl. *G10L 19/12* (2006.01)

(72) Inventeurs/Inventors:
MORII, TOSHIYUKI, JP;
YASUNAGA, KAZUTOSHI, JP

(73) Propriétaire/Owner:
MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD., JP

(74) Agent: OSLER, HOSKIN & HARCOURT LLP

(54) Titre : DISPOSITIF DE CODAGE ET DE DECODAGE DE LA PAROLE
(54) Title: SPEECH CODING APPARATUS AND SPEECH DECODING APPARATUS



(57) Abrégé/Abstract:
First codebook (61) and second codebook (62) respectively have two subcodebooks, and in respective codebooks, addition sections (66) and (67) obtain respective excitation vectors by adding sub-excitation vectors fetched from respective two

CA 2300077 C 2007/09/04

(11)(21) 2 300 077

(13) C

(57) Abrégé(suite)/Abstract(continued):
subcodebooks. Addition section (68) obtains an excitation sample by adding those excitation vectors. According to the aforementioned constitution, it is possible to store sub-excitation vectors with different characteristics in respective subcodebooks. Therefore, it is possible to correspond to input signals with various characteristics, and achieve excellent sound qualities at the time of decoding.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 9/14

A1

(11) International Publication Number: **WO 99/65017**

(43) International Publication Date: 16 December 1999 (16.12.99)

(21) International Application Number: PCT/JP99/03064

(22) International Filing Date: 8 June 1999 (08.06.99)

(30) Priority Data:
10/160119 9 June 1998 (09.06.98) JP
10/258271 11 September 1998 (11.09.98) JP

(71) Applicant *(for all designated States except US)*: MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD. [JP/JP]; 1006, Oaza Kadoma, Kadoma-shi, Osaka 571-8501 (JP).

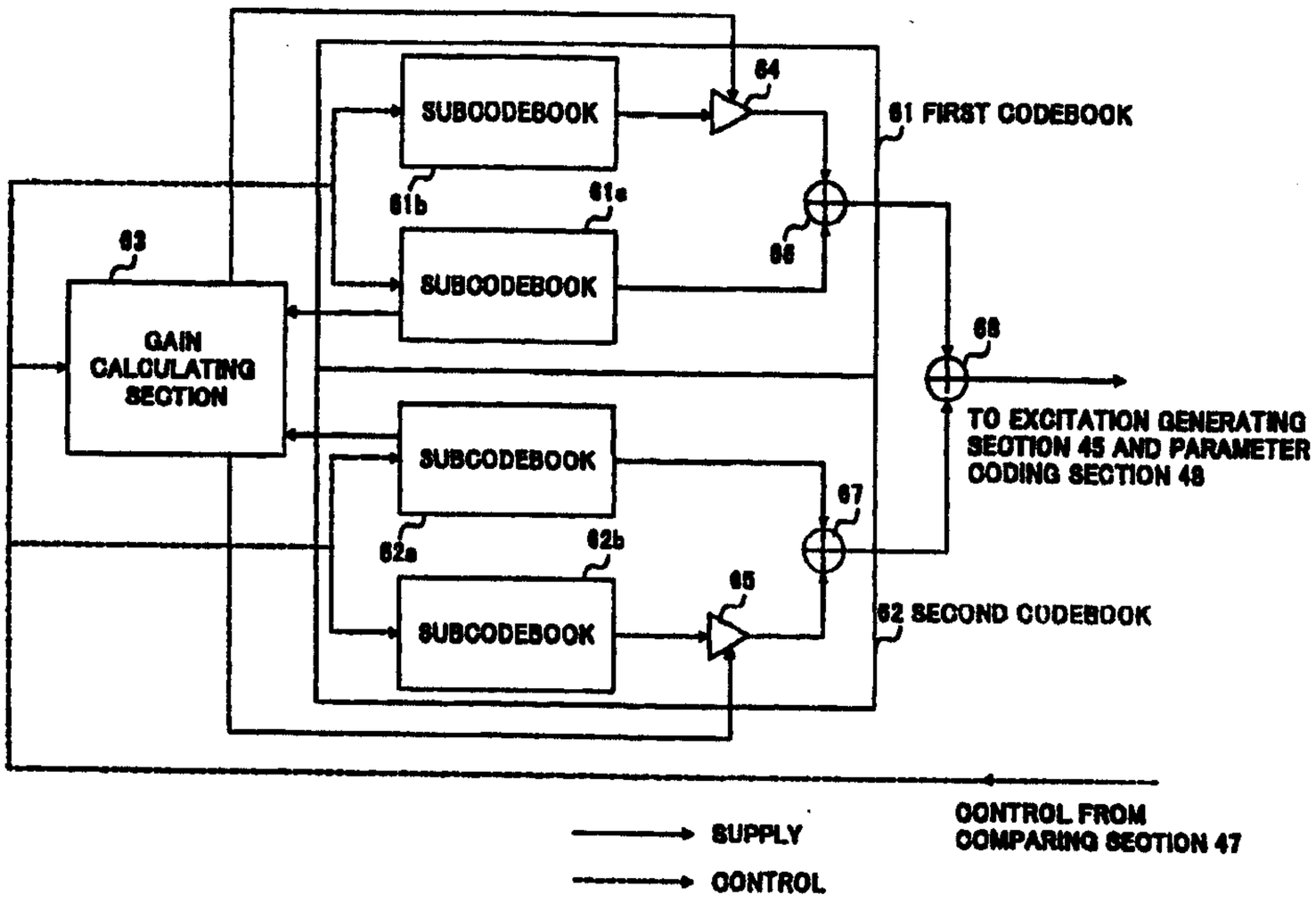
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MORII, Toshiyuki [JP/JP]; 2-3-7-501, Nijigaoka, Asao-ku, Kawasaki-shi, Kanagawa 215-0015 (JP). YASUNAGA, Kazutoshi [JP/JP]; 3-33-17-305, Sugao, Miyamae-ku, Kawasaki-shi, Kanagawa 216-0015 (JP).

(74) Agent: WASHIDA, Kimihito; Shintoshicenter Building, 5th floor, 24-1, Tsurumaki 1-chome, Tama-shi, Tokyo 206-0034 (JP).

(81) Designated States: CA, CN, JP, KR, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: SPEECH CODING APPARATUS AND SPEECH DECODING APPARATUS



(57) Abstract

First codebook (61) and second codebook (62) respectively have two subcodebooks, and in respective codebooks, addition sections (66) and (67) obtain respective excitation vectors by adding sub-excitation vectors fetched from respective two subcodebooks. Addition section (68) obtains an excitation sample by adding those excitation vectors. According to the aforementioned constitution, it is possible to store sub-excitation vectors with different characteristics in respective subcodebooks. Therefore, it is possible to correspond to input signals with various characteristics, and achieve excellent sound qualities at the time of decoding.

DESCRIPTION

SPEECH CODING APPARATUS AND SPEECH DECODING APPARATUS

Technical Field

The present invention relates to a speech coding apparatus and a speech decoding apparatus using speech coding algorithm at low bit rates, used in digital communications such as a portable telephone.

Background Art

Speech compression coding methods at low bit rates have been required in order to accept an increase of subscribers in digital mobile communications such as a portable telephone, and the researches and developments have been proceeded by many research institutions. In Japan, applied coding systems as a standard system in portable telephones are VSELP at a bit rate of 11.2kbps developed by Motorola and PSI-CELP at a bit rate of 5.6kbps developed by NTT Mobile Communications Network, INC., and portable telephones with these system are produced.

In addition, internationally, the ITU-T selected CS-ACELP, which was co-developed by Nippon Telegraph and Telephone Corporation and France Telecom, as an international standard speech coding system G.729 at 8kbps. The system is scheduled to be used in Japan as speech coding system for portable telephones.

The above-described systems are all achieved by modifying the CELP system (Code Excited Linear Prediction:

M.R.Schroeder "High Quality Speech at Low Bit Rates" described in Proc.ICASSP '85 pp.937-940). A feature of this system is to apply a method of dividing a speech into excitation information and vocal truct information, code the excitation information with indices of a plurality of excitation samples stored in a codebook, while coding the LPC (Linear Prediction Coefficients) with respect to the vocal truct information, and perform a comparison to an input speech considering of the vocal truct information in the excitation information coding (A-b-S: Analysis by Synthesis).

The basic algorithm of the CELP system will be described using FIG.1. FIG.1 is a block diagram illustrating a configuration of a speech coding apparatus in the CELP system. In the speech coding apparatus illustrated in FIG.1, LPC analyzing section 2 executes autocorrelation analysis and LPC analysis on input speech data 1 to obtain the LPC. LPC analyzing section 2 further codes the obtained LPC to obtain the coded LPC. LPC analyzing section 2 furthermore decodes the obtained coded LPC to obtain the decoded LPC.

Excitation generating section 5 fetches excitation samples stored in adaptive codebook 3 and stochastic codebook 4 (respectively referred to as an adaptive code vector (or adaptive excitation) and stochastic code vector (or stochastic excitation)) and provides respective excitation samples to LPC synthesis section 6. LPC synthesis section 6 executes filtering on two excitations obtained at excitation generating section 5 with the decoded LPC obtained at LPC analyzing section

2.

Comparing section 7 analyzes the relation of two synthesized speeches obtained at LPC synthesis section 6 and the input speech, obtains an optimal value (optimal gain) for two synthesized speeches, adds each synthesized speech respectively subjected to power adjustment with the optimal gain to obtain a total synthesized speech, and executes a distance calculation between the total synthesized speech and the input speech. Comparing section 7 further executes, with respect to all excitation samples in adaptive codebook 3 and stochastic codebook 4, the distance calculations between the input speech and each of other many synthesized speeches obtained by functioning excitation generating section 5 and LPC synthesis section 6, and obtains an index of the excitation sample whose distance is the smallest among the obtained distances. Then, comparing section 7 provides the obtained optimal gain, indices of excitation samples of respective codebooks and two excitation samples corresponding to respective index to parameter coding section 8.

Parameter coding section 8 executes coding on the optimal gain to obtain the coded gain and provides the coded gain, the coded LPC and the indices of excitation samples to transmission path 9. Further, parameter coding section 8 generates an actual excitation signal (synthesized excitation) using the coded gain and two excitations corresponding to the respective index and stores the excitation signal in adaptive codebook 3 while deleting old excitation samples.

In addition, it is general for the synthesis at LPC synthesis section 6 to use together Linear Prediction Coefficients and a high-frequency enhancement filter or a perceptual weighting filter with long-term prediction coefficients (which are obtained by the long-term prediction analysis of input speech). It is further general to execute the excitation search on the adaptive codebook and stochastic codebook at an interval (called subframe) obtained by further dividing an analysis interval.

The stochastic codebook will be described next.

The adaptive codebook is a codebook for an effective compression using a long-term correlation existing at intervals of human vocal cord vibrations, and stores previous synthesized excitations. On the contrary, the stochastic code book is a fixed codebook to reflect statistical characteristics of excitation signals. As excitation samples stored in the stochastic codebook, there are, for example, random number sequence, pulse sequence, random number sequence/pulse sequence obtained by statistic training with speech data, or pulse sequence with relatively small number of pulses generated algebraically (algebraic codebook). The algebraic codebook has been especially paid attention recently and known by that a good sound quality is obtained at bit rates such as 8kbps with small calculation amounts .

However, an application of algebraic codebook with a small number of pulses to coding at lower bit rates introduces a phenomenon that sound qualities greatly deteriorate mainly

on unvoiced consonants and background noises. On the other hand, an application of excitation with a large number of pulses such as random number sequence to coding at lower bit rates introduces a phenomenon that sound qualities greatly deteriorate mainly on voiced speeches. In order to improve the deterioration, a method with multi-codebook, in which a voiced/unvoiced judgement is performed, is examined. However, the method has the complicated processing and sometimes generates an allophone caused by a judgement error on a speech signal.

As described above, there has been no algebraic codebook which matches any effective coding on voiced speeches, unvoiced speeches and background noises. Therefore, it has been required to obtain a speech coding apparatus and a speech decoding apparatus capable of effectively coding any of voiced speeches, unvoiced speeches and background noises.

Disclosure of Invention

An object of the present invention is to provide a speech coding apparatus and a speech decoding apparatus capable of effectively coding any of voiced speeches, unvoiced speeches and background noises and obtaining speeches with excellent qualities with a small amount of information and a small amount of computations.

The inventors of the present invention noticed that pulse positions are relatively near at a voiced sound segment of speech, while pulse positions are relatively far at segments

of unvoiced sound of speech and background noise, in the case of applying a pulse sequence to coding at low bit rates. In other words, the inventors noticed that energy-concentrated excitation samples, which are characteristics of human vocal cord wave, are needed in a voiced speech, and in this case, there is a tendency that a small number of pulses whose positions are near are selected, while an excitation having more random number characteristics is needed in a unvoiced speech and background noise, in this case, there is a tendency that a large number of energy-spread pulses are selected.

Based on the foregoing consideration, the inventors found out that the perception is improved by identifying a speech as voiced sound segment, or unvoiced sound segment and background noise segment by recognizing a distance of pulse positions, and based on the identification result, applying respective pulse sequences appropriate for the voiced sound segment, and the unvoiced and background noise segments, to achieve the present invention.

A feature of the present invention is to use a plurality of codebooks each having two subcodebook of which characteristics are different and add excitation vectors of each subcodebook to obtain excitation vectors. According to the algorithm, the characteristics as a small-number-pulse excitation appear in the case where pulse positions are near, which is caused by the relationships of positions of the excitation vectors with a small number of pulses, while the characteristics as a large-number-pulse excitation appear in

the case where pulse positions are far, which is suited to the characteristics of speech signals containing background noises.

Accordingly, without using particular voiced/unvoiced speech judgement algorithm, it is possible to automatically select an excitation most suitable for the localized characteristics in input signals, effectively code any of voiced speeches, unvoiced speeches and background noises, and obtain synthesized speeches with excellent sound qualities with a small amount of information and a small amount of computations.

Brief Description of Drawings

FIG.1 is a block diagram illustrating a configuration of a speech coding apparatus in a conventional CELP system;

FIG.2 is a block diagram illustrating a configuration of a radio communication apparatus having a speech coding apparatus and a speech decoding apparatus of the present invention;

FIG.3 is a block diagram illustrating a configuration of a speech coding apparatus in a CELP system according to a first embodiment to a third embodiment of the present invention;

FIG.4 is a block diagram illustrating a configuration of a speech decoding apparatus in the CELP system according to a first embodiment to a third embodiment of the present invention;

FIG.5 is a block diagram illustrating a stochastic codebook in a speech coding apparatus/speech decoding apparatus according to the first embodiment of the present invention;

FIG.6A and FIG.6B are concept diagrams of sub-excitation vectors stored in subcodebooks in the stochastic codebook;

FIG.s 7A to 7F are concept diagrams to explain a generation method of excitation sample;

FIG.8 is a block diagram illustrating a stochastic codebook in a speech coding apparatus/speech decoding apparatus according to the second embodiment of the present invention;

FIG.9 is a block diagram illustrating a stochastic codebook in a speech coding apparatus/speech decoding apparatus according to the third embodiment of the present invention;

FIG.10A and FIG.10B are concept diagrams of sub-excitation vectors stored in subcodebooks in the stochastic codebook;

FIG.s 11A to 11F are concept diagrams to explain a generation method of excitation sample; and

FIG.12 is a diagram illustrating a schematic configuration of a data medium storing a program for the speech coding apparatus/speech decoding apparatus of the present invention.

Best Mode for Carrying Out the Invention

Embodiments of the present invention will be described in detail with reference to accompanying drawings.

(First embodiment)

FIG.2 is a block diagram illustrating a configuration of a radio communication apparatus having a speech coding/decoding apparatus according to the first embodiment to the third embodiment of the present invention.

In this radio communication apparatus, at a transmitting side, a speech is converted into electric analogue signals at speech input device 21 such as a microphone and output to A/D converter 22. The analogue speech signals are converted into digital speech signals at A/D converter 22 and output to speech coding section 23. Speech coding section 23 executes speech coding processing on the digital speech signals and outputs the coded data to modulation/demodulation circuit 24. Modulation/demodulation circuit 24 executes digital modulation on the coded speech signals to output to radio transmission circuit 25. Radio transmission circuit 25 executes the predetermined radio transmission processing on the modulated signals. The signals are transmitted via antenna 26. In addition, processor 31 executes the processing properly using data stored in RAM 32 and ROM 33.

On the other hand, at a receiving side in the radio communication apparatus, received signals received at antenna 26 are subjected to the predetermined radio reception processing at radio reception circuit 27 and output to modulation/demodulation circuit 24.

Modulation/demodulation circuit 24 executes demodulation processing on the received signals and outputs the demodulated signals to speech decoding section 28. Speech decoding section 28 executes decoding processing on the demodulated signals to obtain digital decoded speech signals and output the digital decoded speech signals to D/A converter 29. D/A converter 29 converts the digital decoded speech signals output from speech decoding section 28 into analogue decoded speech signals to output to speech output device 30 such as a speaker. Finally, speech output device 30 converts electric analogue decoded speech signals into decoded speech to output.

Speech coding section 23 and speech decoding section 28 are operated by processor 31 such as DSP using codebooks stored in RAM 32 and ROM 33. The operation program is also stored in ROM 33.

FIG.3 is a block diagram illustrating a configuration of a speech coding apparatus in the CELP system according to the first embodiment to the third embodiment of the present invention. The speech coding apparatus is included in speech coding section 23 illustrated in FIG.2. In addition, adaptive codebook 43 illustrated in FIG.3 is stored in RAM 32 illustrated in FIG.2, and stochastic codebook 44 illustrated in FIG.3 is stored in ROM 33 illustrated in FIG.2.

In the speech coding apparatus (hereinafter, also referred to as coder) illustrated in FIG.3, LPC analyzing section 42 executes autocorrelation analysis and LPC analysis on input speech data 41 to obtain the LPC. LPC analyzing

section 42 further codes the obtained LPC to obtain the LPC code. LPC analyzing section 42 furthermore decodes the obtained LPC code to obtain the decoded LPC. In the coding, it is generally executed to convert into parameters having good interoperation characteristics such as LSP (Linear Spectrum Pair) then code by VQ (Vector Quantization).

Excitation generating section 45 fetches excitation samples stored in adaptive codebook 43 and stochastic codebook 44 (respectively referred to as adaptive code vector (or adaptive excitation) and stochastic code vector (or stochastic excitation)) and provides respective excitation samples to LPC synthesis section 46. The adaptive codebook is a codebook in which excitation signals previously synthesized are stored and an index represents which synthesized excitation is used among from excitations synthesized at different previous times, i.e., time lag.

LPC synthesis section 46 executes filtering on two excitations obtained at excitation generating section 45 with the decoded LPC obtained at LPC analyzing section 42.

Comparing section 47 analyzes the relation of two synthesized speeches obtained at LPC synthesis section 46 and the input speech, obtains an optimal value (optimal gain) for two synthesized speeches, adds each synthesized speech respectively subjected to power adjustment with the optimal gain to obtain a total synthesized speech, and executes a distance calculation between the total synthesized speech and the input speech. Comparing section 47 further executes, with

respect to all excitation samples in adaptive codebook 43 and stochastic codebook 44, the distance calculations between the input speech and each of other many synthesized speeches obtained by functioning excitation generating section 45 and LPC analyzing section 46, and obtains an index of the excitation sample whose distance is the smallest among the obtained distances. Then, comparing section 47 provides the obtained optimal gain, indices of excitation samples of respective codebooks and two excitation samples corresponding to respective index to parameter coding section 48.

Parameter coding section 48 executes coding on the optimal gain to obtain the gain code and provides the gain code, the LPC code and the indices of excitation samples to transmission path 49. Further, parameter coding section 48 generates an actual excitation signal (synthesized excitation) using the gain code and two excitations corresponding to the index and stores the excitation signal in adaptive codebook 43 while deleting old excitation samples.

In addition, it is general for the synthesis at LPC synthesis section 46 to use together Linear Prediction Coefficients and a high-frequency enhancement filter or a perceptual weighting filter with long-term prediction coefficients (which are obtained by the long-term prediction analysis of input speech). It is further general to execute the excitation search on the adaptive codebook and stochastic codebook at an interval (called subframe) obtained by further dividing an analysis interval.

FIG.4 is a block diagram illustrating a configuration of a speech decoding apparatus in the CELP system according to the first embodiment to the third embodiment of the present invention. The speech decoding apparatus is included in speech decoding section 28 illustrated in FIG.2. In addition, adaptive codebook 53 illustrated in FIG.4 is stored in RAM 32 illustrated in FIG.2, and stochastic codebook 54 illustrated in FIG.4 is stored in ROM 33 illustrated in FIG.2.

In the speech decoding apparatus illustrated in FIG.4, parameter decoding section 52 obtains coded speech signals from transmission path 51, while obtains respective coded excitation samples of excitation codebooks (adaptive codebook 53 and stochastic codebook 54), the coded LPC and coded gain . Parameter decoding section 52 then obtains the decoded LPC using the coded LPC and the decoded gain using the coded gain .

Excitation generating section 55 multiplies each excitation sample respectively by the decoded gain to obtain decoded excitation signals. At this stage, excitation generating section 55 stores the obtained decoded excitation signals in adaptive codebook 53 as excitation samples, while deletes old excitation samples. LPC synthesis section 56 executes filtering on the decoded excitation signals with the decoded LPC to obtain a synthesized speech.

In addition, these two excitation codebooks are the same as those included in the speech coding apparatus illustrated in FIG.3 (reference numerals 43 and 44 in FIG.3). Sample numbers to fetch excitation samples (codes to adaptive

codebook and codes to stochastic codebook) are both supplied from parameter decoding section 52 (which corresponds to the short dashes line in FIG.5 (control from comparing section 47) described later).

The following description is given to explain in detail about functions of stochastic codebooks 44 and 54 for storing excitation samples, in the speech coding apparatus and speech decoding apparatus with the above configurations, using FIG.5 FIG.5 is a block diagram illustrating a stochastic codebook in the speech coding apparatus and speech decoding apparatus according to the first embodiment of the present invention.

The stochastic codebook has first codebook 61 and second codebook 62, and first codebook 61 and second codebook 62 respectively have two subcodebooks 61a, 61b and 62a, 62b. The stochastic codebook further has gain calculating section 63 which calculates a gain for outputs from subcodebooks 61b and 62b using pulse positions in subcodebooks 61a and 62a.

Subcodebooks 61a and 62a are mainly used in the case where a speech is a voiced sound (pulse positions are relatively near), and formed by storing a plurality of sub-excitation vectors composed of a single pulse. Subcodebook 61b and 62b are mainly used in the case where a speech is an unvoiced sound or background noise (pulse positions are relatively far), and formed by storing a plurality of sub-excitation vectors composed of a sequence with a plurality of pulses in which power is spread. The excitation samples are generated in the stochastic codebooks formed as described above. In addition,

the near and far pulse positions will be described later.

In addition, subcodebooks 61a and 62b are formed by a method of arranging pulses algebraically, and subcodebooks 61b and 62b are formed by another method of dividing a vector length (subframe length) into some segment intervals and making a configuration so that a single pulse is always present at every segment interval (pulses are spread over a whole length).

These codebooks are formed in advance. In this embodiment, as illustrated in FIG.5, the number of codebooks is set at two and each codebook has two subcodebooks.

FIG.6A illustrates sub-excitation vectors stored in subcodebook 61a of first codebook 61. FIG.6B illustrates sub-excitation vectors stored in subcodebook 61b of first codebook 61. Similarly, subcodebooks 62a and 62b of second codebook 62 respectively have sub-excitation vectors illustrated in FIG.6A and FIG.6B.

In addition, positions and polarities of pluses of sub-excitation vectors in subcodebooks 61b and 62b are formed using random numbers. According to the configuration described above, it is possible to form sub-excitation vectors in which power is uniformly spread over a whole vector length even though some fluctuations are present. FIG.6B illustrates an example in the case where the number of segment intervals is four. In addition, in these two subcodebooks, respective sub-excitation vectors of the same index (number) are used at the same time.

The next description is given to explain speech coding

using the stochastic codebooks with the above-mentioned configuration.

Gain calculating section 63 calculates an excitation vector number (index) according to the code from comparing section 47 in the speech coding apparatus. The code provided from comparing section 47 corresponds to the excitation vector number, and therefore the excitation vector number is determined by the code. Gain calculating section 63 fetches sub-excitation vectors with a small number of pulses corresponding to the determined excitation vector number from subcodebooks 61a and 62a. Gain calculating section 63 further calculates an addition gain using pulse positions of the fetched sub-excitation vectors. The addition gain calculation is given by the following equation (1);

$$g = |P1 - P2| / L \quad \text{...equation (1)}$$

where g is an addition gain, P1 and P2 are respectively pulse positions in codebooks 61a and 62a, and L is a vector length (subframe length). Further, $|\ |$ represents an absolute value.

According to the above equation (1), the addition gain is smaller as the pulse positions are nearer (the pulse distance is shorter), while larger as pulse positions are further, and has a lower limit of 0 and an upper limit of 1. Accordingly, as the pulse positions are nearer, the gain for subcodebooks 61b and 62b is relatively smaller. As a result, an affect of subcodebooks 61a and 62b corresponding to voiced speech is larger. On the other hand, as the pulse positions are further (the pulse distance is longer), the gain for subcodebooks 61b

and 62b is relatively larger. As a result, an affect of subcodebooks 61b and 62b corresponding to unvoiced speech and background noise is relatively larger. Perceptually fine sounds are obtained by performing the gain control described above.

Next, gain calculating section 63 refers to the number of excitation vector provided from comparing section 47 and obtains two sub-excitation vectors from subcodebooks 61b and 62b with a large number of pulses. These two sub-excitation vectors from subcodebooks 61b and 62b are respectively provided to gain calculating sections 64 and 65 to be multiplied by the addition gain obtained at gain calculating section 63.

Further, excitation vector addition section 66 obtains a sub-excitation vector from subcodebook 61a with a small number of pulses by referring to the number of excitation vector provided from comparing section 47, and also obtains the sub-excitation vector, from subcodebook 61b, multiplied by the addition gain obtained at gain calculating section 63. Excitation vector addition section 66 then adds the obtained sub-excitation vectors to obtain an excitation vector. Similarly, excitation vector addition section 67 obtains a sub-excitation vector from subcodebook 62a with a small number of pulses by referring to the number of excitation vector provided from comparing section 47, and also obtains the sub-excitation vector, from subcodebook 62b, multiplied by the addition gain obtained at gain calculating section 63. Excitation vector addition section 67 then adds the obtained

sub-excitation vectors to obtain an excitation vector.

The excitation vectors respectively obtained by adding the sub-excitation vector are provided to excitation vector addition section 68 to be added. According to the foregoing processing, an excitation sample (stochastic code vector) is obtained. The excitation sample is provided to excitation generating section 45 and parameter coding section 48.

On the other hand, a decoding side prepares the same adaptive codebook and stochastic codebook as those in the coder in advance, and based on respective index, LPC code, and gain code of each codebook transmitted from the transmission path, multiplies respective excitation sample by the gain to add. Then the decoding side executes filtering on the added sample with the decoded LPC to decode the speech.

An example of excitation samples selected by the above-mentioned algorithm will be described next using FIG.7A to FIG.7F. Assume that an index of first codebook 61 is j, and an index of second codebook 62 is m or n.

As been understood from FIG.7A and FIG.7B, in the case of j+m, since the pulse positions of the sub-excitation vectors of subcodebooks 61a and 62a are relatively near, a small value of the addition gain is calculated using the equation (1) described previously. Accordingly, the addition gain for subcodebooks 61b and 62b is small. Because of it, as illustrated in FIG.7C , excitation vector addition section 68 obtains an excitation sample composed of a small number of pulses which reflects the characteristics of subcodebooks 61a

and 62a respectively illustrated in FIG.7A and FIG.7B. This excitation sample is effective on voiced speech.

Further, as been understood from FIG.7A and FIG.7B, in the case of j+n, since the pulse positions of the sub-excitation vectors of subcodebooks 61a and 62a are relatively far, a large value of the addition gain is calculated using the equation (1) described previously. Accordingly, the addition gain for subcodebooks 61b and 62b is large. Because of it, as illustrated in FIG.7F, excitation vector addition section 68 obtains an excitation sample with strong random characteristics with spread energy which reflects the characteristics of subcodebooks 61b and 62b respectively illustrated in FIG.7D and FIG.7E. This excitation sample is effective on unvoiced speech /background noise.

This embodiment describes about the case of using two codebooks (two channels). However, it is also preferable to apply the present invention to the case of using codebooks equal to or more than three (channels equal to or more than three). In this case, as a numerator of the equation in gain calculating section 63, i.e., equation (1), the minimum value among from intervals between two pulses or the averaged value of all pulse intervals is used. For example, in the case where the number of codebooks is three and the minimum pulse interval is used as a numerator of the equation (1), the calculation equation is given by the following equation (2);

$$g=\min(|P1-P2|,|P2-P3|,|P3-P1|)/L \quad \ldots \text{ equation (2)}$$

where g is an addition gain, P1, P2 and P3 are respective pulse

positions in those three codebooks, and L is a vector length (subframe length). In addition, | | represents an absolute value.

As described above, according to this embodiment, a plurality of codebooks have two subcodebooks each having respective sub-excitation vectors of which characteristics are different, and the excitation vector is obtained by adding each sub-excitation vector, thereby making it possible to correspond to input signals with various characteristics.

In addition, since the gain to be multiplied by the sub-excitation vector is varied corresponding to the characteristics of the sub-excitation vectors, it is possible to reflect both characteristics of excitation vectors stored in two codebooks in the speech by a gain adjustment, thereby making it possible to effectively execute coding and decoding most suitable for the characteristics of the input signals with various characteristics.

Specifically, since one of two subcodebooks stores a plurality of sub-excitation vectors composed of a small number of pulses, and another subcodebook stores a plurality of sub-excitation vectors composed of a large number of pulses, it is possible to achieve fine sound qualities in voiced speech by the excitation sample with the characteristics of a small number of pulses, and perform excitation generation most suitable to the characteristics of input signals with various characteristics.

In addition, since gain calculating section calculates

a gain using a distance of pulse positions of sub-excitation vectors composed of a small number of pulses, it is possible to achieve synthesized speeches with fine sound qualities in voiced speech by the small number of pulses of which distance is near, while achieve perceptually fine synthesized speeches in unvoiced speech/background noise by the large number of pulses with spread energy.

In the addition gain calculation described above, the processing is simplified by using a fixed value which is predetermined as an addition gain. In this case, it is not necessary to install gain calculating section 63. Even in this case, it is possible to achieve synthesized speeches matching the needs timely by varying the setting of the fixed value properly. For example, it is possible to achieve coding excellent for plosive speech such as low voice like male voice by setting the addition gain on a small scale, while to achieve coding excellent for random speeches such as background noise by setting the addition gain on a large scale.

In addition, it is also preferable to apply a method of calculating an addition gain adaptively using a level of input signal power, decoded LPC or adaptive codebook, besides the method of calculating the addition gain using pulse positions and another method of providing fixed coefficients to the addition gain. For example, it may be possible to achieve excellent coding adaptive for localized speech characteristics by preparing a function for determining voiced speech characteristics (such as vowel and standing wave) or

unvoiced speech characteristics (such as background noise and unvoiced consonant) and setting a small gain in the case of voiced speech characteristics, while setting a large gain in the case of unvoiced speech characteristics.

(Second embodiment)

This embodiment will describes about the case where a gain calculating section obtains decoded LPC from LPC analyzing section 42 and performs a voiced/unvoiced judgement using the obtained LPC.

FIG.8 is a block diagram illustrating a stochastic codebook in the speech coding apparatus/speech decoding apparatus according to the second embodiment of the present invention. The configurations of the speech coding apparatus and the speech decoding apparatus with the stochastic code book are the same as the first embodiment (FIG.3 and FIG.4).

The stochastic codebook has first codebook 71 and second codebook 72, and first codebook 71 and second codebook 72 respectively have two subcodebooks 71a, 71b and subcodebooks 72a, 72b. The stochastic codebook further has gain calculating section 73 which calculates a gain for outputs from subcodebooks 71b and 72b using pulse positions in subcodebooks 71a and 72a.

Subcodebooks 71a and 72a are mainly used in the case where a speech is a voiced sound (pulse positions are relatively near), and formed by storing a plurality of sub-excitation vectors composed of a single pulse. Subcodebook 71b and 72b are mainly used in the case where a speech is an unvoiced sound or

background noise (pulse positions are relatively far), and formed by storing a plurality of sub-excitation vectors composed of a sequence with a plurality of pulses in which power is spread. The excitation samples are generated in the stochastic codebooks formed as described above.

In addition, subcodebooks 71a and 72a are formed by a method of arranging pulses algebraically, and subcodebooks 71b and 72b are formed by another method of dividing a vector length (subframe length) into some segment intervals and making a configuration so that a single pulse is always present at every segment interval (pulses are spread over a whole length).

These codebooks are formed in advance. In this embodiment, as illustrated in FIG.8, the number of codebooks is set at two and each codebook has two subcodebooks. The number of codebooks and the number of subcodebooks are not limited.

FIG.6A illustrates sub-excitation vectors stored in subcodebook 71a of first codebook 71. FIG.6B illustrates sub-excitation vectors stored in subcodebook 71b of first codebook 71. Similarly, subcodebooks 72a and 72b of second codebook 72 respectively have sub-excitation vectors illustrated in FIG.6A and FIG.6B.

In addition, positions and polarities of pluses of sub-excitation vectors in subcodebooks 71b and 72b are formed using random numbers. According to the configuration described above, it is possible to form sub-excitation vectors in which power is uniformly spread over a whole vector length

even though some fluctuations are present. FIG.6B illustrates an example in the case where the number of segment intervals is four. In addition, in these two subcodebooks, respective sub-excitation vectors of the same index (number) are used at the same time.

The next description is given to explain speech coding using the stochastic codebooks with the above-mentioned configuration.

Gain calculating section 73 obtains decoded LPC from LPC analyzing section 42 and performs a voiced/unvoiced judgement using the obtained LPC. Specifically, gain calculating section 73 beforehand collects data corresponding to LPC, for example, obtained by converting the LPC into impulse response or LPC cepstrum, with respect to a lot of speech data, by relating to every mode, for example, voiced speech, unvoiced speech and background noise. Then the data are subjected to statistic processing and based on the result, a rule of judging voiced, unvoiced and background noise is generated. As an example of the rule, it is general to use linear determination function and Bayes judgment. Then, based on the judgment result obtained according to the rule, weighting coefficient R is obtained by a regulation of the following equation (3);

R=L :when judged as voiced speech

R=L×0.5 :when judged as unvoiced speech/ background noise ...equation (3)

where R is a weighting coefficient, and L is a vector length (subframe length).

Gain calculating section 73 next receives an instruction of the number of excitation vector (index number) from comparing section 47 in the speech coding apparatus, and according to the instruction, fetches sub-excitation vectors of the designated number respectively from subcodebooks 71a and 72a with a small number of pulses. Gain calculating section 73 calculates an addition gain using pulse positions of the fetched sub-excitation vectors. The calculation of the addition gain is executed according to the following equation (4);

$$g = |P1 - P2| / R \quad \text{...equation (4)}$$

where g is an addition gain, P1 and P2 are respectively pulse positions in codebooks 71a and 72a, and R is a weighting coefficient. Further, $|\ |$ represents an absolute value.

According to the above equations (3) and (4), the addition gain is smaller as the pulse positions are nearer, while larger as pulse positions are further, and has a lower limit of 0 and an upper limit of L/R. Accordingly, as the pulse positions are nearer, the gain for subcodebooks 71b and 72b is relatively smaller. As a result, an affect of subcodebooks 71a and 72a corresponding to voiced speech is larger. On the other hand, as the pulse positions are further, the gain for subcodebooks 71b and 72b is relatively larger. As a result, an affect of subcodebooks 71b and 72b corresponding to unvoiced speech and background noise is larger. Perceptually fine sounds are obtained by performing the gain calculation described above.

Further, excitation vector addition section 76 obtains a sub-excitation vector from subcodebook 61a with a small number of pulses by referring to the number of excitation vector provided from comparing section 47, and also obtains a sub-excitation vector, from subcodebook 71b, multiplied by the addition gain obtained at gain calculating section 73. Excitation vector addition section 76 then adds the obtained sub-excitation vectors to obtain an excitation vector. Similarly, excitation vector addition section 77 obtains a sub-excitation vector from subcodebook 72a with a small number of pulses by referring to the number of excitation vector provided from comparing section 47, and also obtains a sub-excitation vector, from subcodebook 72b, multiplied by the addition gain obtained at gain calculating section 73. Excitation vector addition section 77 then adds the obtained sub-excitation vectors to obtain an excitation vector.

The excitation vectors respectively obtained by adding the sub-excitation vector are provided to excitation vector addition section 68 to be added. According to the foregoing processing, an excitation sample (stochastic code vector) is obtained. The excitation sample is provided to excitation generating section 45 and parameter coding section 48.

On the other hand, a decoding side prepares the same adaptive codebook and stochastic codebook as those in the coder in advance, and based on respective index, LPC code, and gain code of each codebook transmitted from the transmission path, multiplies respective excitation sample by the gain to add.

Then the decoding side executes filtering on the added sample with the decoded LPC to decode the speech.

At this stage, it is necessary to provide the decoded LPC to the stochastic codebook in this embodiment, which differs from the first embodiment. Specifically, at this stage, parameter decoding section 52 provides the obtained LPC along with the sample number for the stochastic codebook to the stochastic codebook (which corresponds to that the signal line from parameter decoding section 52 to stochastic codebook 54 in FIG.4 includes the signal line from "LPC analyzing section 42" and the control line indicative of "control from comparing section 47").

The excitation samples selected by the above algorithm are the same as the first embodiment and illustrated in FIG.7A to FIG.7F.

As described above, according to this embodiment, gain calculating section 73 performs the voiced/unvoiced judgement using the decoded LPC, and calculates the addition gain using weighting coefficient R obtained according to equation (3), resulting in a small gain at the time of voiced speech and a large gain at the time of unvoiced speech and background noise. The obtained excitation samples are thereby a smaller number of pulses in voiced speech and a large number of pulses containing more noises in unvoiced speech and background noise. Accordingly, it is possible to further improve the effect by adaptive pulse positions described above, thereby enabling synthesized speech with more excellent sound

qualities to be achieved.

In addition, the speech coding in this embodiment also has the effect on transmission error. In the coding with a conventional voiced/unvoiced judgment, stochastic codebooks are switched generally by LPC. Because of it, when a transmission error introduces a wrong judgment, the decoding is sometimes executed with absolutely different excitation samples, resulting in a low transmission error resistance.

On the contrary, in the speech coding in this embodiment, if wrong LPC are used in the voiced/unvoiced judgment in decoding, only a value of addition gain varies a little, and the deterioration caused by the transmission error is little. Hence, according to this embodiment, it is possible to obtain synthesized speeches with excellent sound qualities without being affected by the transmission error of LPC code largely, while executing the adaptation by LPC.

This embodiment describes about the case of using two codebooks (two channels). However, it is also preferable to apply the present invention to the case of using codebooks equal to or more than three (channels equal to or more than three). In this case, as a numerator of the equation in gain calculating section 63, i.e., equation (4), the minimum value among from intervals between two pulses or the averaged value of all pulse intervals is used.

The first and second embodiments describe about the case of adjusting gains for outputs from subcodebooks 61b, 62b, 71b and 72b. However, it is also preferable to adjust outputs from

subcodebooks 61a, 62a, 71a and 72a or to adjust outputs from all subcodebooks, under the condition that a gain for outputs from subcodebooks is adjusted so that an affect by excitation vectors with a small number of pulses is large when pulse positions are near, while an affect by excitation vectors with a large number of pulses is large when pulse positions are far.

(Third embodiment)

This embodiment will describe about the case of switching an excitation vector to acquire from a subcodebook corresponding to a distance of pulse intervals.

FIG.9 is a block diagram illustrating a stochastic codebook in the speech coding apparatus/speech decoding apparatus according to the third embodiment of the present invention. The configurations of the speech coding apparatus and the speech decoding apparatus with the stochastic code book are the same as the first embodiment (FIG.3 and FIG.4).

The stochastic codebook has first codebook 91 and second codebook 92, and first codebook 91 and second codebook 92 respectively have two subcodebooks 91a, 91b and subcodebooks 92a, 92b. The stochastic codebook further has excitation switching instructing section 93 which executes switching between outputs from subcodebooks 91b and 92b corresponding to a pulse position in subcodebooks 91a and 92a.

Subcodebooks 91a and 92a are mainly used in the case where a speech is a voiced sound (pulse positions are relatively near), and formed by storing a plurality of sub-excitation vectors composed of a single pulse. Subcodebook 91b and 92b are mainly

used in the case where a speech is an unvoiced sound or background noise (pulse positions are relatively far), and formed by storing a plurality of sub-excitation vectors composed of a sequence with a plurality of pulses in which power is spread. The excitation samples are generated in the stochastic codebooks formed as described above.

In addition, subcodebooks 91a and 92a are formed by a method of arranging pulses algebraically, and subcodebooks 91b and 92b are formed by another method of dividing a vector length (subframe length) into some segment intervals and making a configuration so that a single pulse is always present at every segment interval (pulses are spread over a whole length).

These codebooks are formed in advance. In this embodiment, as illustrated in FIG.9, the number of codebooks is set at two and each codebook has two subcodebooks. The number of codebooks and the number of subcodebooks are not limited.

FIG.10A illustrates sub-excitation vectors stored in subcodebook 91a of first codebook 91. FIG.10B illustrates sub-excitation vectors stored in subcodebook 91b of first codebook 91. Similarly, subcodebooks 92a and 92b of second codebook 92 respectively have sub-excitation vectors illustrated in FIG.10A and FIG.10B.

In addition, positions and polarities of pluses of sub-excitation vectors in subcodebooks 91b and 92b are formed using random numbers. According to the configuration described above, it is possible to form sub-excitation vectors

in which power is uniformly spread over a whole vector length even though some fluctuations are present. FIG.10B illustrates an example in the case where the number of segment intervals is four. In addition, in these two subcodebooks, respective sub-excitation vectors of the same index (number) are not used at the same time.

The next description is given to explain speech coding using the stochastic codebooks with the above-mentioned configuration.

Excitation switching instructing section 93 calculates the excitation vector number (index) according to a code from comparing section 47 in the speech coding section. The code provided from comparing section 47 corresponds to the excitation vector number, and therefore the excitation vector number is determined by the code. Excitation switching instructing section 93 fetches sub-excitation vectors with a small number of pulses corresponding to the determined excitation vector number from subcodebooks 91a and 92a. Further, excitation switching instructing section 93 executes a judgment described as below, using pulse positions of the fetched sub-excitation vectors;

$|P1-P2| < Q$:using subcodebooks 91a and 92a

$|P1-P2| \geqq Q$:using subcodebooks 91b and 92b,

where P1 and P2 are respectively pulse positions in subcodebooks 91a and 92a, Q is a constant and $|\ |$ represents an absolute value.

In the above judgment, excitation vectors with a small

number of pulses are selected when pulse positions are near, while excitation vectors with a large number of pulses are selected when pulse positions are far. Performing the judgment and selection as described above enables perceptually fine sounds to be achieved. The constant Q is predetermined. It is possible to vary the ratio of the excitation with a small number of pulses and the excitation with a large number of pulses by varying the constant Q.

Excitation switching instructing section 93 fetches excitation vectors from subcodebooks 91a and 92a or subcodebooks 91b and 92b in codebooks 91 or 92 according to the switching information (switching signal) and the code of excitation (sample number). The switching is executed at first and second switches 94 and 95.

The obtained excitation vectors are provided to excitation vector addition section 96 to be added. The excitation sample (stochastic code vector) is thus obtained. The excitation sample is provided to excitation generating section 45 and parameter coding section 48. In addition, at a decoding side, the excitation sample is provided to excitation generating section 55.

An example of excitation samples selected by the above-mentioned algorithm will be described next using FIG.11A to FIG.11F. Assume that an index of first codebook 91 is j, and an index of second codebook 92 is m or n.

As been understood from FIG.11A and FIG.11B, in the case if j+m, since the pulse positions of the sub-excitation vectors

of subcodebooks 91a and 92a are relatively near, excitation switching instructing section 93 selects sub-excitation vectors with a small number of pulses according to the above judgment. Then, excitation vector addition section 96 adds two sub-excitation vectors selected respectively from subcodebooks 91a and 92a illustrated in FIG.11A and FIG.11B. and obtains an excitation sample with strong pulse characteristics as illustrated in FIG.11C. This excitation sample is effective on voiced speech.

Further, as been understood from FIG.11A and FIG.11B, in the case of j+n, since the pulse positions of the sub-excitation vectors of subcodebooks 91a and 92a are relatively far, excitation switching instructing section 93 selects sub-excitation vectors with a large number of pulses according to the above judgment. Then, excitation vector addition section 96 adds two sub-excitation vectors selected respectively from subcodebooks 91b and 92b illustrated in FIG.11D and FIG.11E. and obtains an excitation sample with strong random characteristics with spread energy as illustrated in FIG.11F. This excitation sample is effective on unvoiced speech/background noise.

As described above, according to this embodiment, an excitation sample is generated by switching excitation vectors in two subcodebooks which a plurality of codebooks each have to obtain, and using excitation vectors obtained from either of subcodebooks in each codebook. It is thus possible to correspond to input signals with various characteristics by

a fewer amount of computations.

Since one of two subcodebooks stores a plurality of excitation vectors with a small number of pulses while another one stores a plurality of excitation vectors with a large number of pulses in which power is spread, it is possible to use the excitation sample with a small number of pulses for voiced speech while use another excitation sample with a large number of pluses for unvoiced speech/background noise. It is thereby possible to obtain synthesized speeches with excellent sound qualities, and also to obtain excellent performances for input signals with various properties.

Further, since the excitation switching instructing section switches excitation vectors to acquire from a subcodebook corresponding to a distance between pulse positions, it is possible to achieve synthesized speeches with fine sound qualities in voiced speech by a small number of pulses of which distances are near, wile achieve perceptually fine synthesized speeches in unvoiced speech and background noise by a large number of pulses in which power is spread. Furthermore, since the excitation switching instructing section acquires excitation vectors from a subcodebook while switching, for example, it is not necessary to calculate a gain and multiple the gain by a vector in an stochastic codebook. Accordingly, in the speech coding according to this embodiment, a computation amount is much less than the case of calculating the gain.

That is, since the above-mentioned switching is executed

based on a relative distance between pulse positions of sub-excitation vectors composed of a small number of pulses, it is possible to achieve fine synthesized speeches in voiced speech by excitation samples with a small number of pulses of which distance are near, while achieve perceptually fine synthesized speeches in unvoiced speech/background noise by excitation samples with a large number of pulses with spread power.

This embodiment describes about the case of using two codebooks (two channels). However, it is also preferable to apply the present invention to the case of using codebooks equal to or more than three (channels equal to or more than three). In this case, as a judgment basis in excitation switching instructing section 93, the minimum value among from intervals between two pulses or the averaged value of all pulse intervals is used. For example, in the case of using three codebooks and the minimum value among from intervals between two pulses, the judgment basis is as follows;

$\min(|P1-P2|, |P2-P3|, |P3-P1|) < Q$: using subcodebooks a

$\min(|P1-P2|, |P2-P3|, |P3-P1|) \geqq Q$: using subcodebooks b

where P1, P2 and P3 are respectively pulse positions in respective codebooks, Q is a weighting coefficient, and $|$ represents an absolute value.

In the speech coding/decoding according to this embodiment, it may be possible to combine voiced/unvoiced judgment algorithm in the same way as the second embodiment. In other words, at a coding side, the excitation switching

instructing section obtains decoded LPC from the LPC analyzing section and executes the voiced/unvoiced judgment using the LPC, and at a decoding side, the decoded LPC is provided to the stochastic codebook. According to the aforementioned processing, it is possible to improve the effect by adapted pulse positions and achieve synthesized speeches with more excellent sound qualities.

The above constitution is achieved by providing voiced/unvoiced judgment sections separately at a coding side and a decoding side and corresponding to the judgment result, making Q variable as a threshold value for the judgment of excitation switching instructing section. In this case, Q is set at a large scale in the case of voiced speech while Q is set at a low scale in the case of unvoiced speech in order to enable varying the ratio of the number of excitations with a small number of pulses and the number of excitations with a large number of pulses corresponding to localized characteristics of speeches.

In addition, in the case where the voiced/unvoiced judgment is executed by backward (using other decoded parameters without transmitting as code), there is a possibility that a wrong judgment occurs by transmission error. According to the coding/decoding in this embodiment, since the voiced/unvoiced judgment is executed only by varying threshold Q, a wrong judgment affects only a difference of threshold Q between in the cases of voiced speech and unvoiced speech. Accordingly, the affects caused by the wrong judgment

is very small.

In addition, it may be possible to use a level of input signal power, decoded LPC and a method of calculating Q adaptively using an adaptive codebook. For example, prepare in advance a function for determining voiced characteristics (such as vowel and standing wave) or unvoiced characteristics (such as background noise and unvoiced consonant) using the above parameters, and set Q at a large scale at the time of the voiced characteristics, while set Q at a low scale at the time of the unvoiced characteristics. According to the aforementioned processing, it is possible to use an excitation sample composed of a small number of pulses in a voiced characteristics interval and another excitation sample composed of a large number of pulses in a unvoiced characteristics interval, thereby making it possible to obtain excellent coding performance adaptive for speech localized characteristics.

In addition, the speech coding/decoding according to the first to third embodiments are described as speech coding apparatus/speech decoding apparatus, however it may be possible to construct the speech coding/decoding as software. For example, it may be possible to store the program for the above-described speech coding/decoding in a ROM and operate by instructions of a CPU according to the program. Further, as illustrated in FIG.12, it may be possible to store program 101a, adaptive codebook 101b and algebraic codebook 101c in recording medium 101 which is readable by computer, write

program 101a of recording medium 101, adaptive codebook 101b and stochastic codebook 101c in a RAM of a computer and operate according to the program. These cases also achieve the same functions and effects as the first to third embodiments described above.

The first to third embodiments describe the case where the number of pulses is one as an excitation vector with a small number of pulses, it may be possible to use an excitation vector in which the number of pulses is equal to or more than two as an excitation vector with a small number of pulses. In this case, it is preferable to apply an interval of pulses whose positions are the nearest among from a plurality of pulses as the near-far judgment of pulse positions.

The first to third embodiments describe about the case of adapting the present invention to speech coding apparatus/speech decoding apparatus in the CELP system, however the present invention is applicable to any speech coding/decoding using "codebook" because the feature of the present invention is in an stochastic codebook. For example, the present invention is applicable to "RPE-LPT" that is a standard full rate codec by GSM and "MP-MLQ" that is an international standard codec "G.723.1" by ITU-T.

Industrial applicability

The speech coding apparatus and speech decoding apparatus according to the present invention are applicable to portable telephones and digital communications using speech coding algorithm at low bit rates.

CLAIMS

1. An apparatus for performing speech coding in a CELP system, said apparatus comprising;

an adaptive codebook in which previously synthesized excitation signals are stored;

a stochastic codebook in which a plurality of excitation vectors are stored, said stochastic codebook comprising a first subcodebook that stores excitation vectors composed of a smaller number of pulses than a predetermined value, and a second subcodebook that stores excitation vectors composed of a larger number of pulses than the predetermined value, and generating only one excitation vector for the output of said stochastic codebook by processing sub-excitation vectors from said first and second subcodebooks;

means for obtaining synthesized speech from excitation information acquired from said adaptive codebook and said stochastic codebook using a decoded LPC obtained by firstly performing LPC analysis on an input speech signal to obtain an LPC, and then by coding the obtained LPC to obtain an LPC code, and then further by decoding the obtained code;

means for obtaining gain information for said synthesized speech from a relationship between said synthesized speech and said input speech signal; and

means for transmitting said LPC, said excitation information and said gain information.

2. The apparatus according to claim 1, wherein said stochastic codebook further has control means for controlling a gain for respective excitation vectors in said first subcodebook and said second subcodebook in accordance with a distance between pulses of the excitation vectors in said first subcodebook, and computation means for obtaining

the excitation information from the gain controlled excitation vectors.

3. The apparatus according to claim 2, wherein said control means reducing the gain for the excitation vectors in said second subcodebook when the distance between pulses of excitation vectors in said first subcodebook decreases, while increasing the gain for the excitation vectors in said second subcodebook when the distance between pulses of excitation vectors in said first subcodebook increases.

4. The apparatus according to claim 3, wherein said control means calculates the gain according to a following equation 1,

$$g = | P1 - P2 | / L \qquad \text{...equation (1)}$$

wherein g is the gain, P1 and P2 are respectively excitation vector positions in first subcodebook, and L is a vector length.

5. The apparatus according to claim 1, said stochastic codebook further has instruction means for instructing an excitation vector to be acquired from said first subcodebook and said second subcodebook in accordance with a distance between pulses of the excitation vectors in said first subcodebook, and switching means for switching between outputs of the excitation vectors in said first subcodebook and said second subcodebook following the instruction by said instruction means.

6. An apparatus for performing speech coding in a CELP system, said apparatus comprising:

an adaptive codebook in which previously synthesized excitation signals are stored;

a stochastic codebook in which a plurality of excitation vectors are stored, said stochastic codebook comprising a first subcodebook that stores excitation vectors composed of a smaller number of pulses than a predetermined value, and a second subcodebook that stores excitation vectors composed of a larger number of pulses than the predetermined value, and generating only one excitation vector for the output of said stochastic codebook by processing sub-excitation vectors from said first and second subcodebooks;

means for obtaining a synthesized speech from excitation information acquired from said adaptive codebook and said stochastic codebook using a decoded LPC obtained by firstly performing LPC analysis on an input speech signal to obtain an LPC, and then by coding the obtained LPC to obtain an LPC code, and then further by decoding the obtained code;

means for executing a voiced/unvoiced judgment on said input speech signal using said LPC, and transmitting its output to said stochastic codebook;

means for obtaining gain information for said synthesized speech from a relationship between said synthesized speech and said input speech signal; and

means for transmitting said LPC, said excitation information and said gain information.

7. The apparatus according to claim 6, wherein said stochastic codebook further has control means for controlling a gain for respective excitation vectors in said first subcodebook and said second subcodebook in accordance with a distance between pulses of the excitation vector in said first subcodebook, and computation means for obtaining

the excitation information from the gain controlled excitation vectors.

8. The apparatus according to claim 7, wherein said control means reducing the gain for the excitation vector in said second subcodebook when the distance between pulses of excitation vectors in said first subcodebook decreases, while increasing the gain for the excitation vector in said second subcodebook when the distance between pulses of excitation vectors in said first subcodebook increases.

9. The apparatus according to claim 7, wherein said control means calculates the gain according to a following equation 2,

$$g = | P1 - P2 | / R \qquad \text{...equation (2)}$$

wherein g is the gain, P1 and P2 are respectively excitation vector positions in said first subcodebook, and R represents a weighting coefficient and is a vector length L in a case where a result of the voiced/unvoiced judgment indicates a voiced speech, and L X 0.5 in another case where the result of the voiced/unvoiced judgment indicates an unvoiced speech.

10. The apparatus according to claim 6, said stochastic codebook further has instruction means for instructing an excitation vector to be acquired from said first subcodebook and said second subcodebook in accordance with a distance between pulses of the excitation vectors of said first subcodebook, and switching means for switching between outputs of the excitation vectors in said first

subcodebook and said second subcodebook according to the instruction by said instruction means.

11. An apparatus for performing speech decoding in a CELP system, said apparatus comprising:

an adaptive codebook in which previously synthesized excitation signals are stored;

a stochastic codebook in which a plurality of excitation vectors are stored, said stochastic codebook comprising a first subcodebook that stores excitation vectors composed of a smaller number of pulses than a predetermined value, and a second subcodebook that stores excitation vectors composed of a larger number of pulses than the predetermined value, and generating only one excitation vector for the output of said stochastic codebook by processing sub-excitation vectors from said first and second subcodebooks;

means for receiving an LPC, excitation information and gain information transmitted from a coding side; and

means for decoding a speech using said LPC on excitation information multiplied by said gain information.

12. The apparatus according to claim 11, wherein said apparatus further comprises means for providing said LPC to said stochastic codebook.

13. A method for performing speech coding in a CELP system, said method comprising the steps of:

controlling a gain for respective excitation vectors in a first subcodebook and a second subcodebook corresponding to a distance between pulses of excitation vectors in said first subcodebook of a stochastic codebook, said stochastic codebook comprising a plurality of codebooks each having

said first subcodebook that stores excitation vectors composed of a smaller number of pulses than the number of pulses of excitation vectors in said second subcodebook and said second subcodebook that stores excitation vectors composed of a larger number of pulses than the number of pulses of excitation vectors in said first subcodebook;

obtaining excitation information from gain controlled excitation vectors;

obtaining a synthesized speech from excitation information acquired from an adaptive codebook and said stochastic codebook using a decoded LPC obtained by firstly performing LPC analysis on an input speech signal to obtain an LPC, and then by coding the obtained LPC to obtain an LPC code, and then further by decoding the obtained code; and

obtaining gain information for said synthesized speech from a relationship between said synthesized speech and said input speech signal.

14. The method according to claim 13, wherein said method further comprises the step of performing a voiced/unvoiced judgment on said input speech signal using said LPC.

15. A method for performing speech coding in a CELP system, said method comprising the steps of:

selecting one of a first subcodebook and a second subcodebook corresponding to a distance between pulses of excitation vectors in said first subcodebook of a stochastic codebook, said stochastic codebook comprising a plurality of codebooks each having said first subcodebook that stores excitation vectors composed of a smaller number of pulses than the number of pulses of excitation vectors in said second subcodebook, and said second subcodebook that stores

excitation vectors composed of a larger number of pulses than the number of pulses of excitation vectors in said first subcodebook;

obtaining excitation information from the selected excitation vector;

obtaining synthesized speech from excitation information acquired from an adaptive codebook and said stochastic codebook using a decoded LPC obtained by firstly performing LPC analysis on an input speech signal to obtain an LPC, and then by coding the obtained LPC to obtain an LPC code, and then further by decoding the obtained code; and

obtaining gain information for said synthesized speech from a relationship between said synthesized speech and said input speech signal.

16. The method according to claim 15, wherein said method further comprises the step of performing a voiced/unvoiced judgment on said input speech signal using said LPC.

17. A recording medium readable by a computer, said medium storing a speech coding program, an adaptive codebook in which previously synthesized excitation signals are stored, and a stochastic codebook in which a plurality of excitation vectors are stored, said stochastic codebook comprising a plurality of codebooks each comprising a first subcodebook that stores excitation vectors composed of a smaller number of pulses than a predetermined value, and a second subcodebook that stores excitation vectors composed of a larger number of pulses than the predetermined value, said program including the steps of:

controlling a gain for respective excitation vectors in said first subcodebook and said second subcodebook in

accordance with a distance between pulses of excitation vectors in said first subcodebook of said stochastic codebook;

obtaining excitation information from gain controlled excitation vectors;

obtaining a synthesized speech from excitation information acquired from said adaptive codebook and said stochastic codebook using an LPC obtained by performing LPC analysis on an input speech signal; and

obtaining gain information for said synthesized speech from a relationship between said synthesized speech and said input speech signal.

FIG.1
PRIOR ART

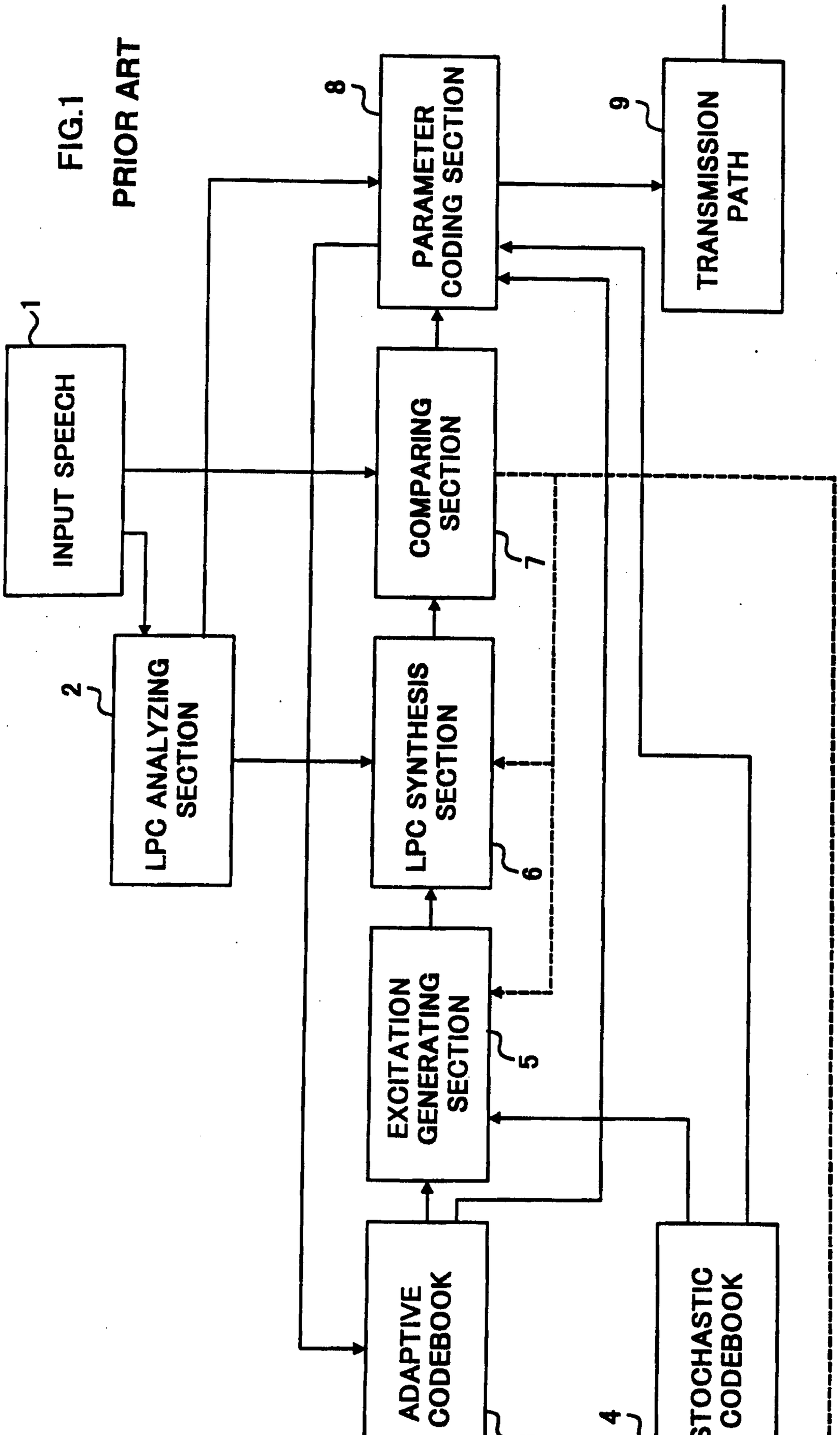


FIG.2

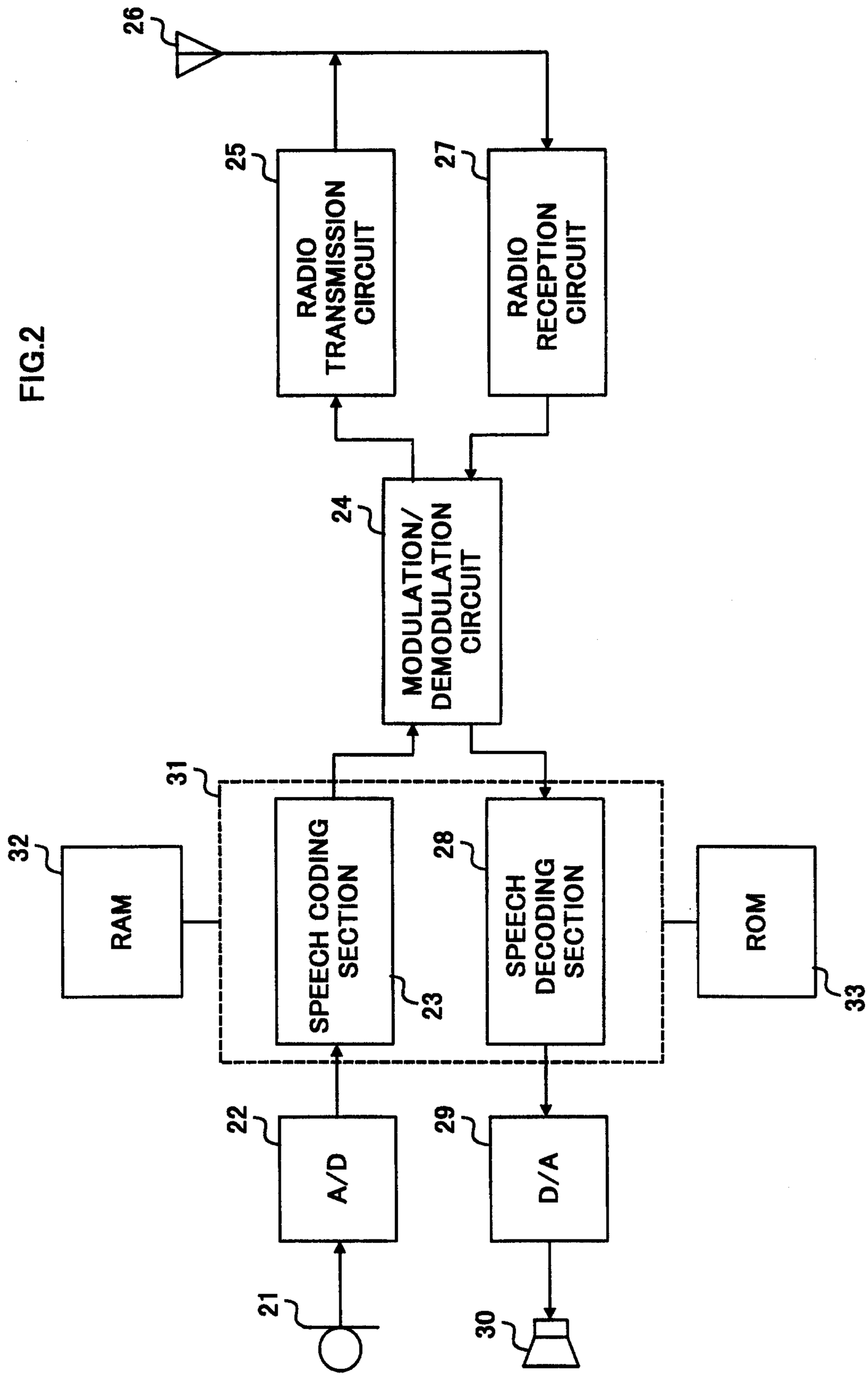

26
25
RADIO TRANSMISSION CIRCUIT
27
RADIO RECEPTION CIRCUIT
24
MODULATION/ DEMODULATION CIRCUIT
31
32
RAM
SPEECH CODING SECTION
23
28
SPEECH DECODING SECTION
ROM
33
22
A/D
29
D/A
21
30

FIG.3

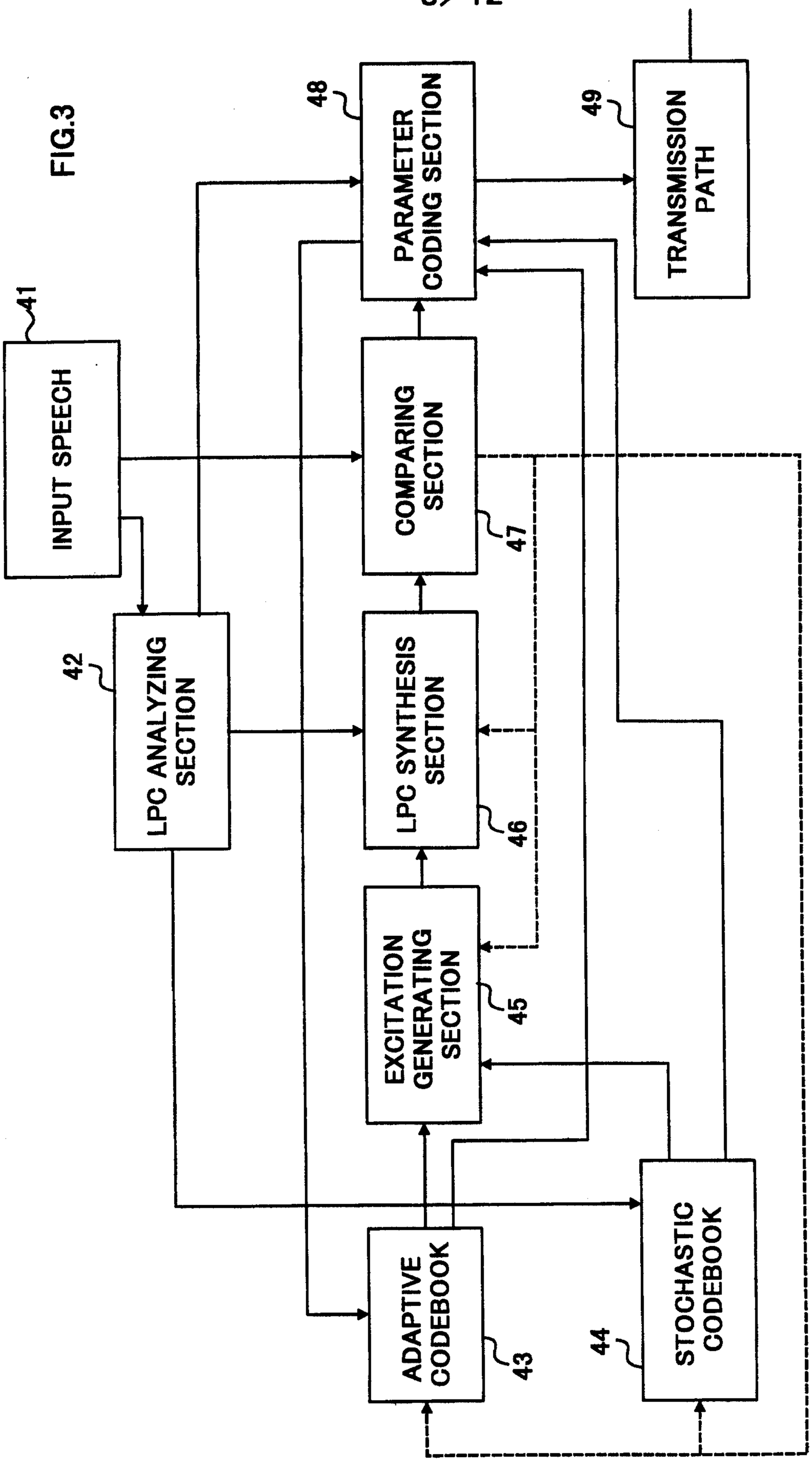

INPUT SPEECH
41
LPC ANALYZING SECTION
42
PARAMETER CODING SECTION
48
TRANSMISSION PATH
49
COMPARING SECTION
47
LPC SYNTHESIS SECTION
46
EXCITATION GENERATING SECTION
45
ADAPTIVE CODEBOOK
43
STOCHASTIC CODEBOOK
44
SUPPLY
CONTROL

FIG.4

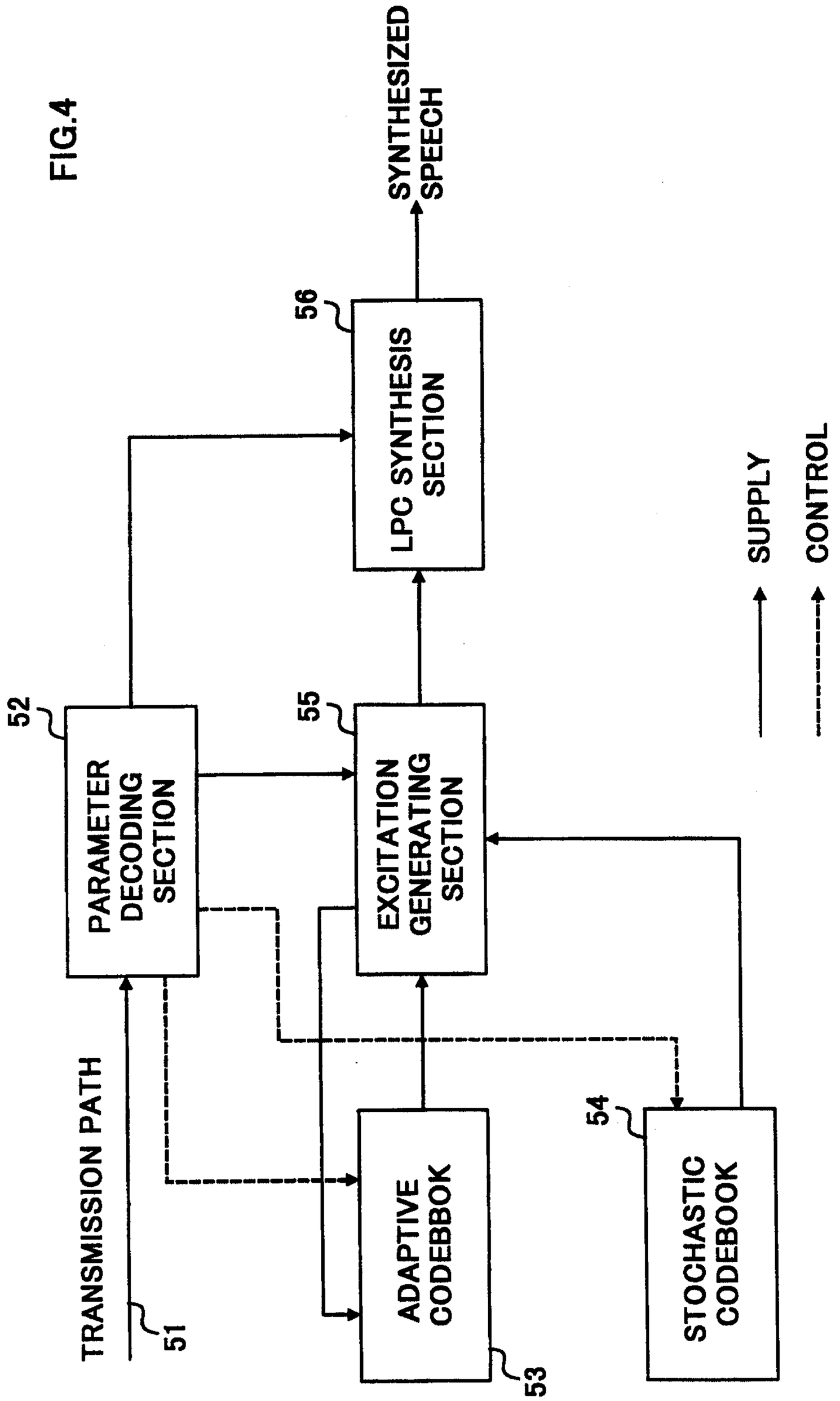
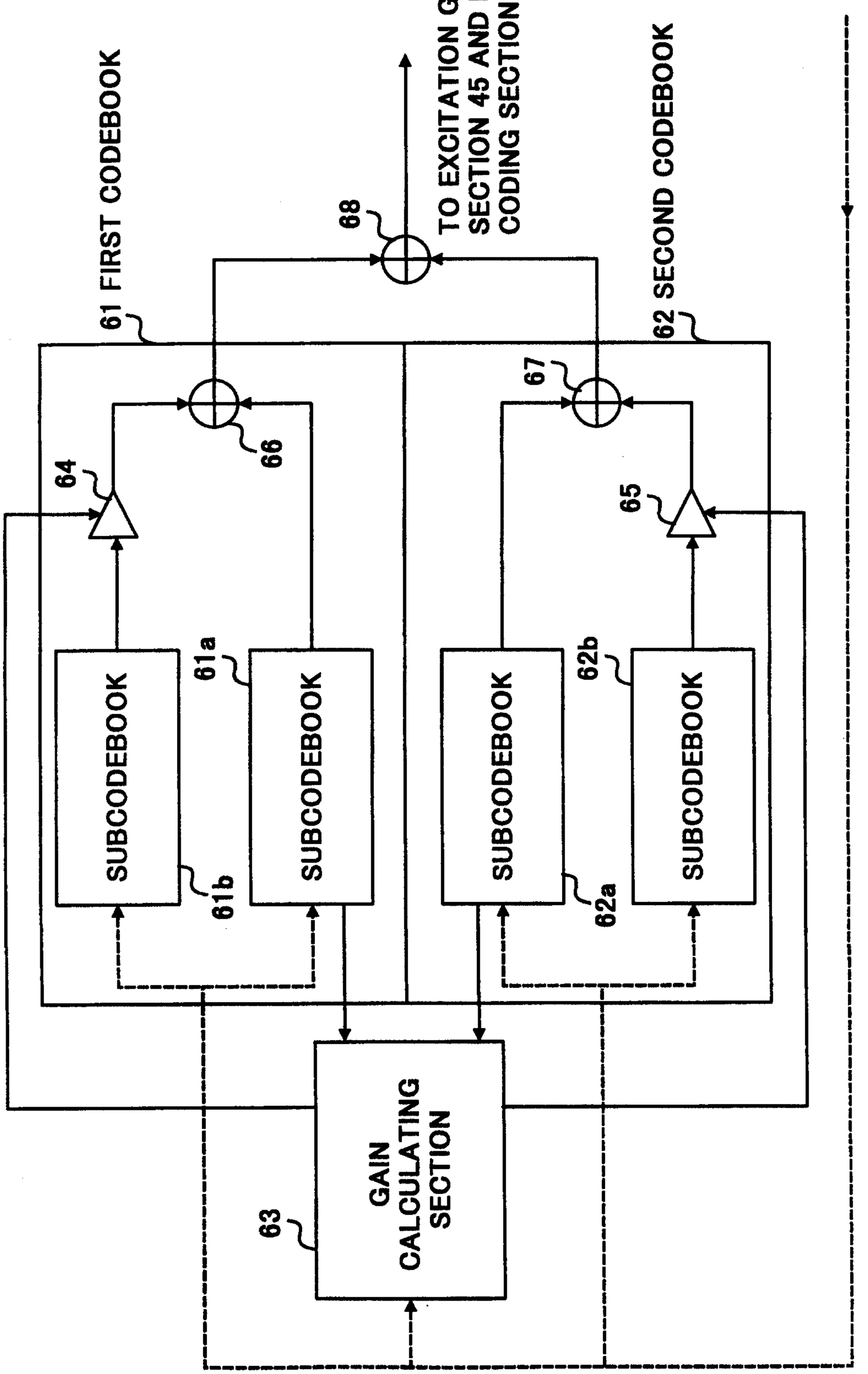

TRANSMISSION PATH
51
52
PARAMETER DECODING SECTION
53
ADAPTIVE CODEBBOK
54
STOCHASTIC CODEBOOK
55
EXCITATION GENERATING SECTION
56
LPC SYNTHESIS SECTION
SYNTHESIZED SPEECH
SUPPLY
CONTROL

FIG.5


61 FIRST CODEBOOK
62 SECOND CODEBOOK
68
TO EXCITATION GENERATING SECTION 45 AND CODING SECTION 48
64
66
67
65
SUBCODEBOOK
61a
61b
62b
62a
63
GAIN CALCULATING SECTION
CONTROL FROM COMPARING SECTION 47
SUPPLY
CONTROL

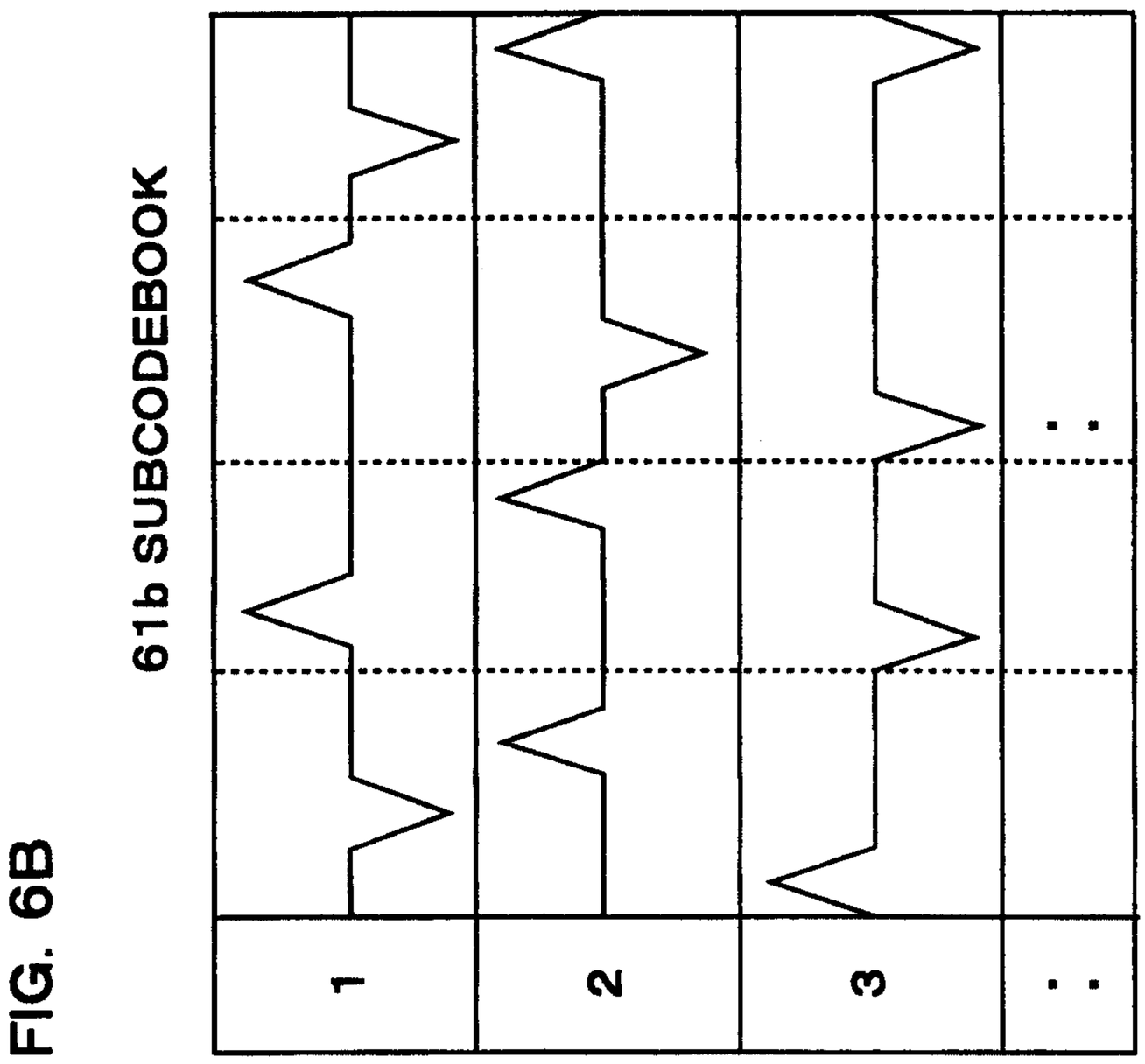


FIG. 6B

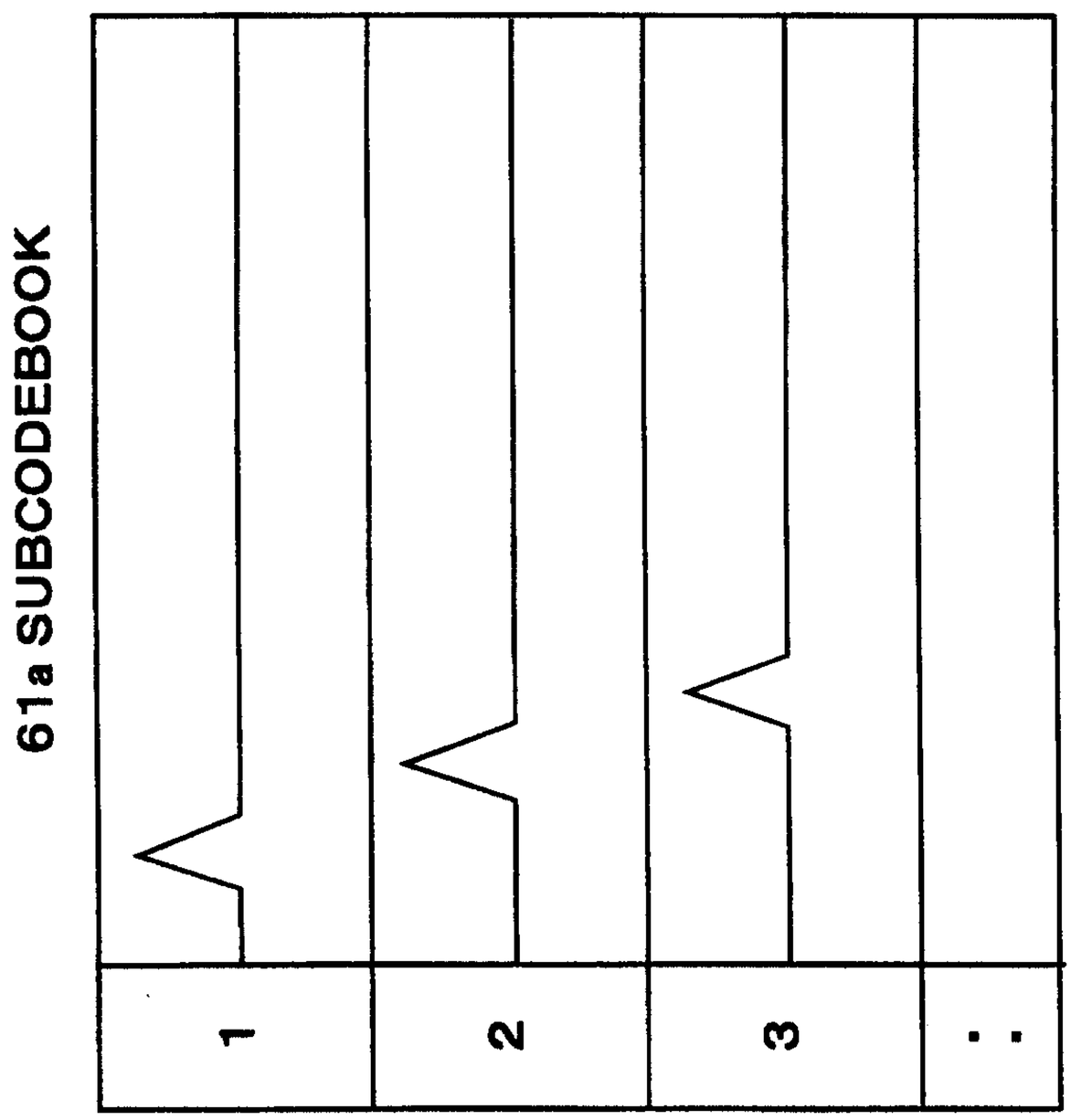


FIG. 6A

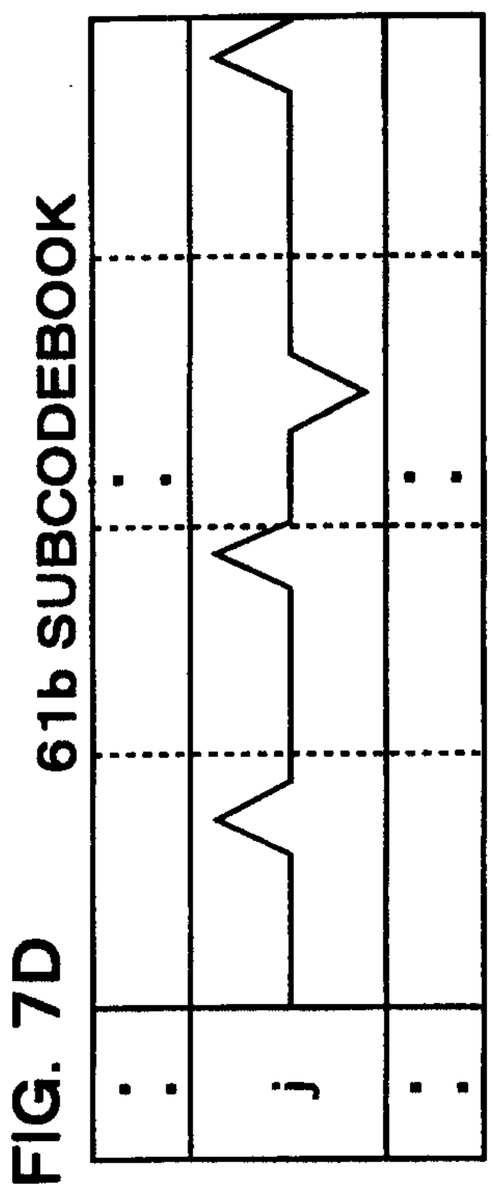


FIG. 7A

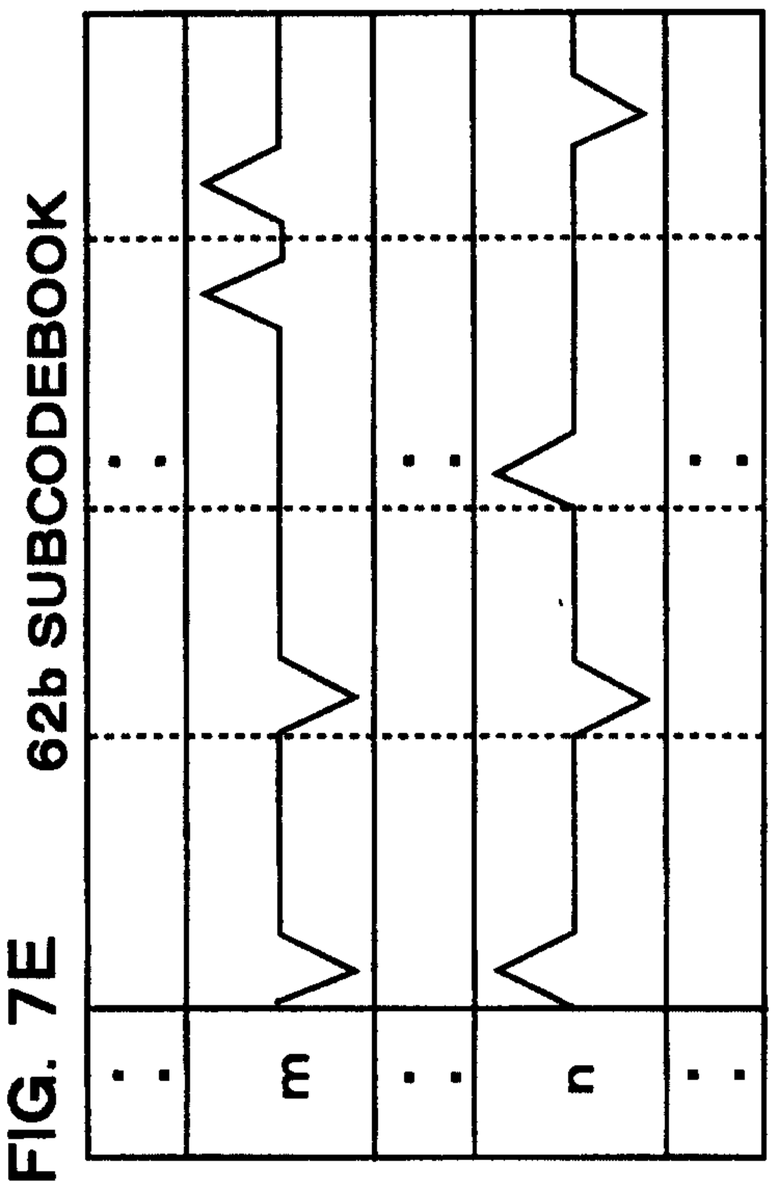


FIG. 7B

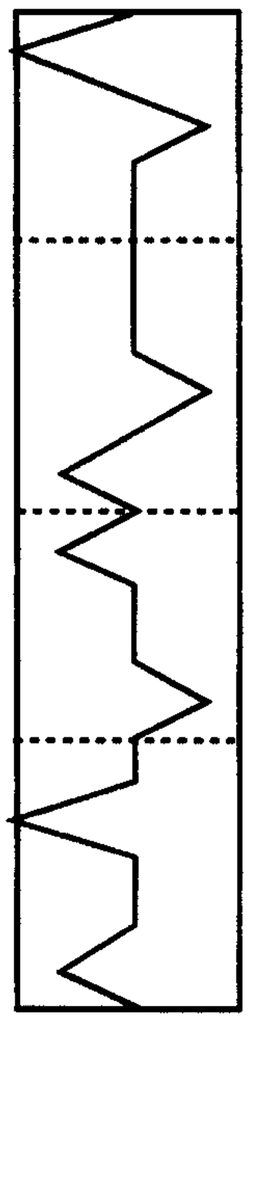


FIG. 7C

FIG. 7D

FIG. 7E

FIG. 7F

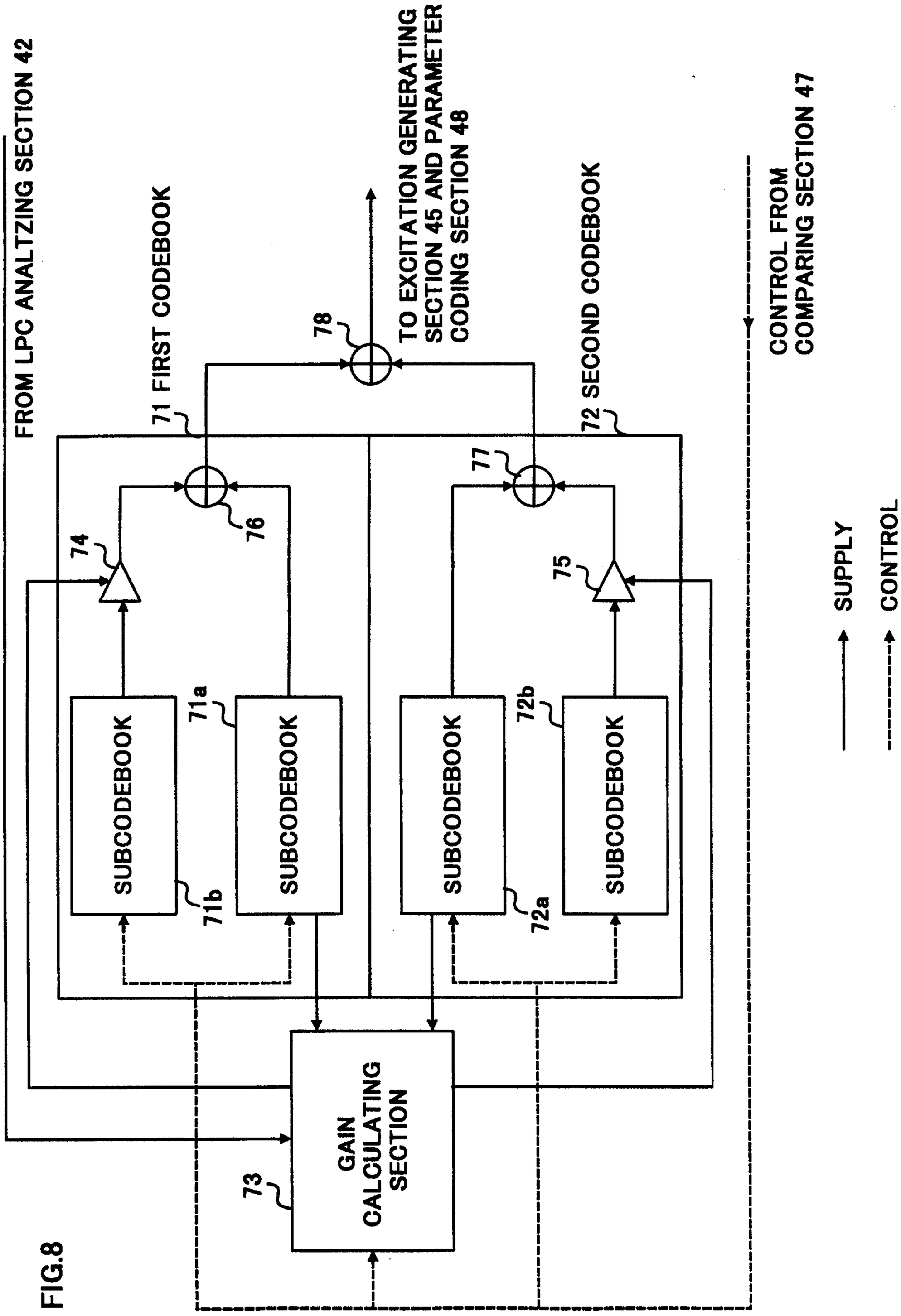
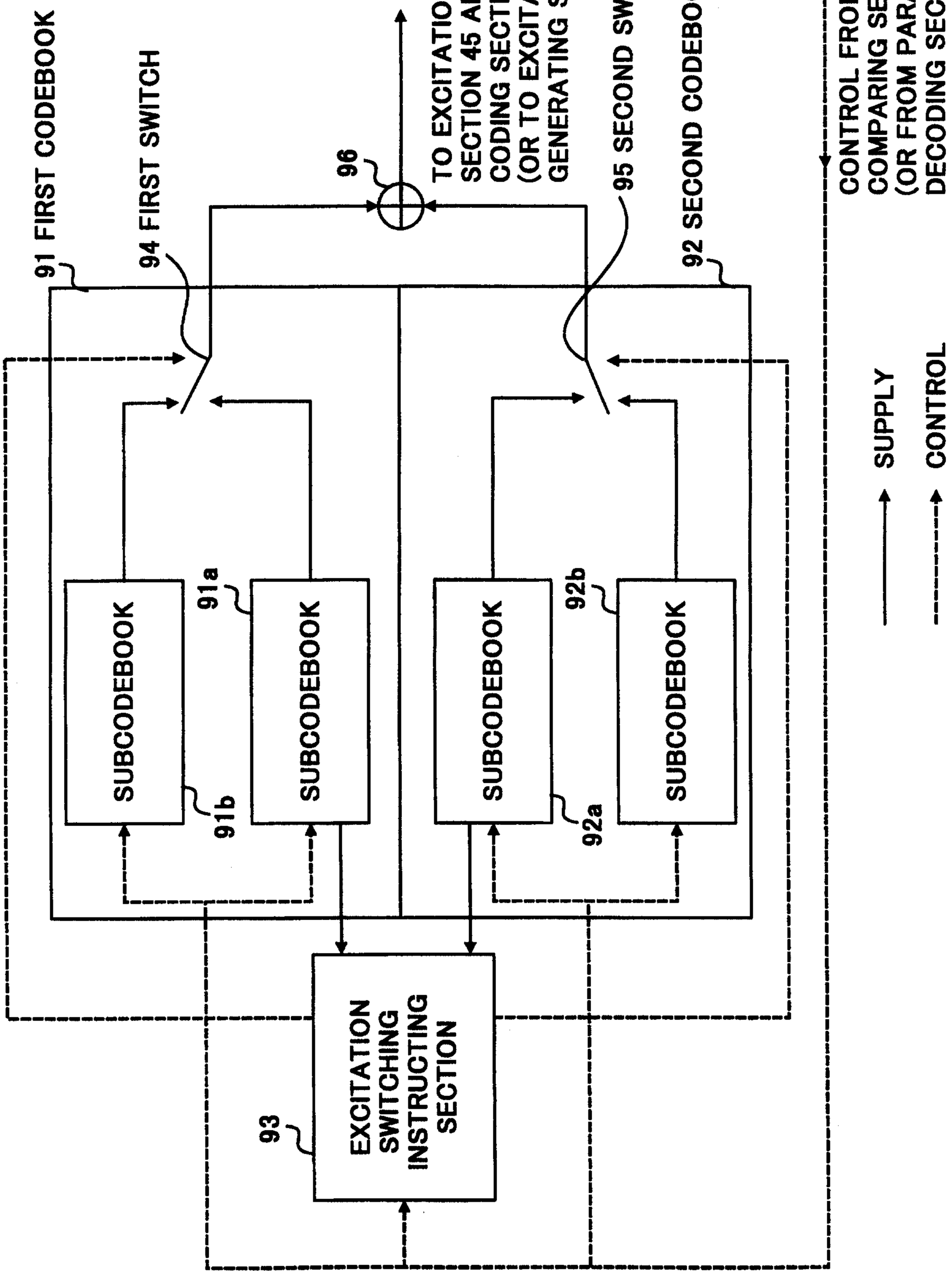

FROM LPC ANALTZING SECTION 42
71 FIRST CODEBOOK
72 SECOND CODEBOOK
78
TO EXCITATION GENERATING SECTION 45 AND PARAMETER CODING SECTION 48
CONTROL FROM COMPARING SECTION 47
SUPPLY
CONTROL
74
75
76
77
SUBCODEBOOK
71a
71b
72a
72b
73
GAIN CALCULATING SECTION

FIG.8

FIG.9


91 FIRST CODEBOOK
94 FIRST SWITCH
96
TO EXCITATION GENERATING SECTION 45 AND PARAMETER CODING SECTION 48 (OR TO EXCITATION GENERATING SECTION 45)
95 SECOND SWITCH
92 SECOND CODEBOOK
CONTROL FROM COMPARING SECTION 47 (OR FROM PARAMETER DECODING SECTION 52)
SUPPLY
CONTROL
SUBCODEBOOK
91a
91b
92a
92b
93
EXCITATION SWITCHING INSTRUCTING SECTION

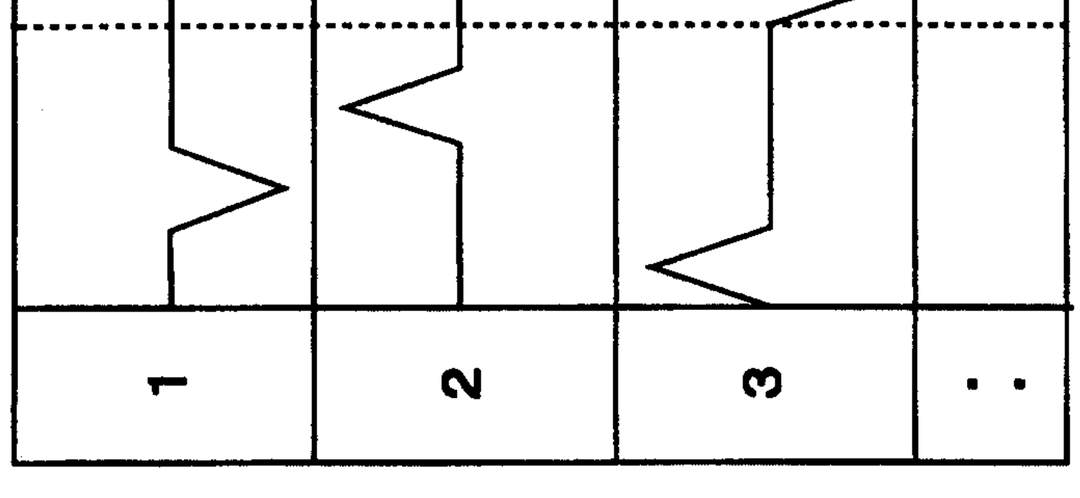


FIG. 10B

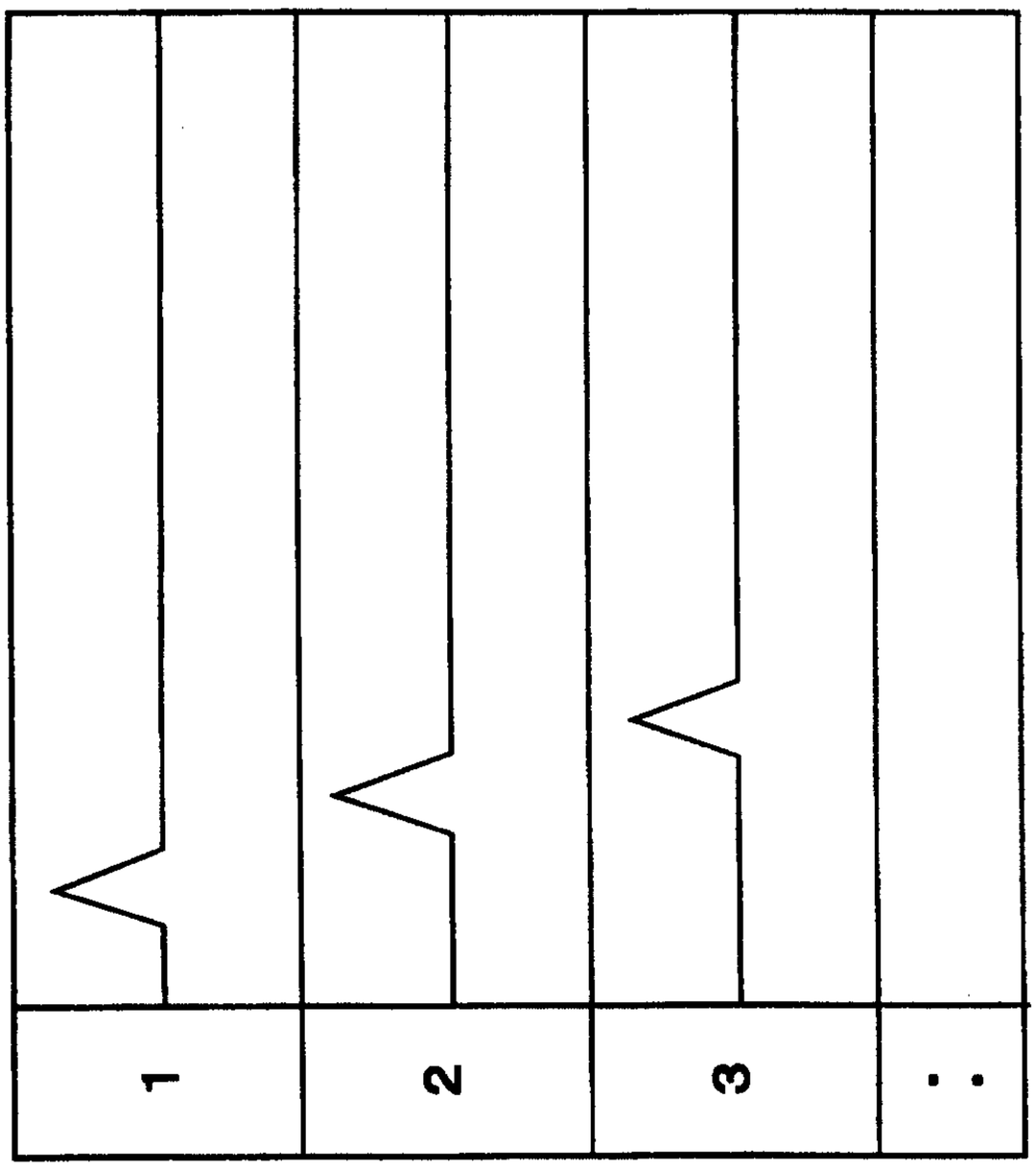


FIG. 10A

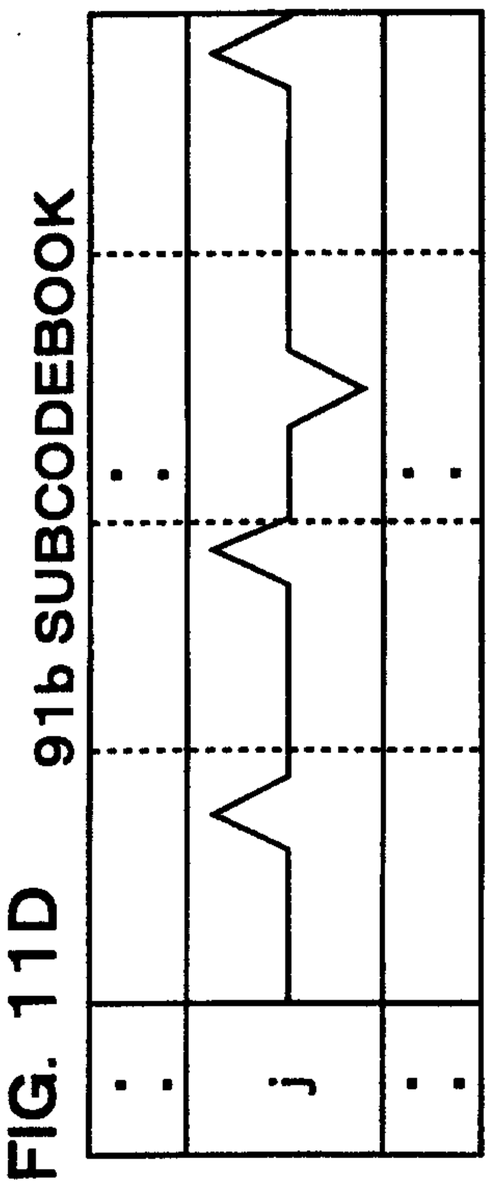


FIG. 11D

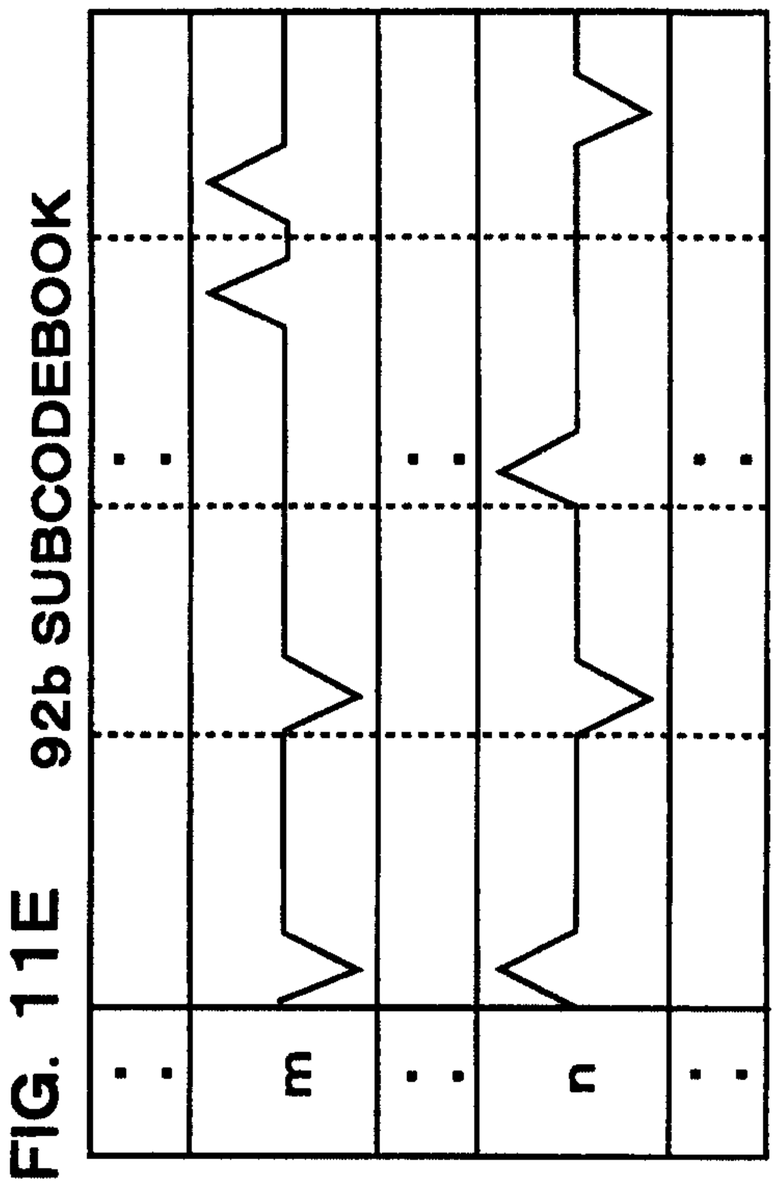


FIG. 11E

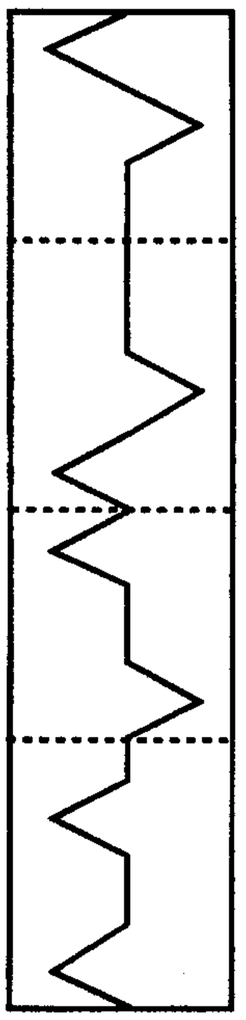


FIG. 11F IN THE CASE OF j+n

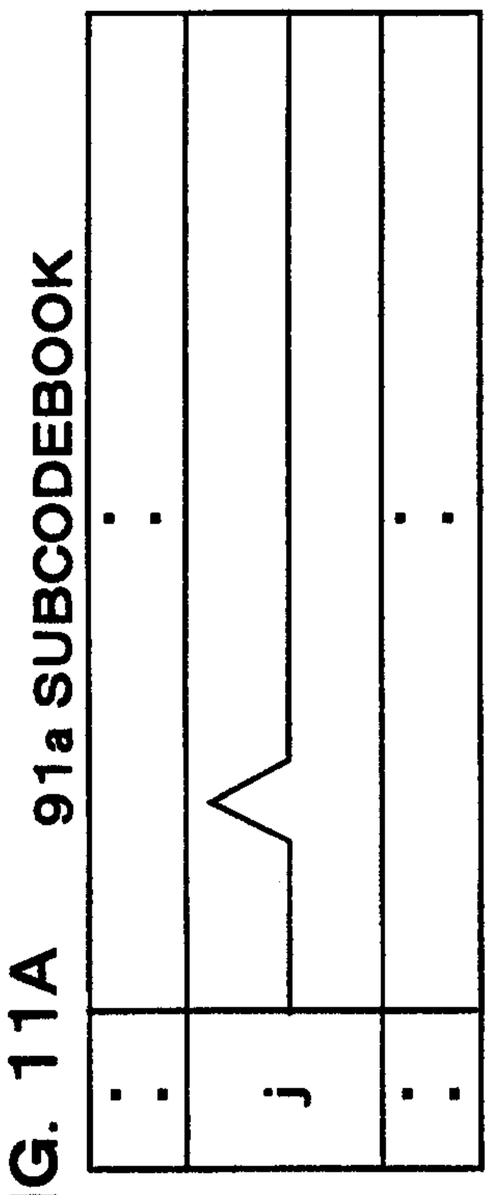


FIG. 11A

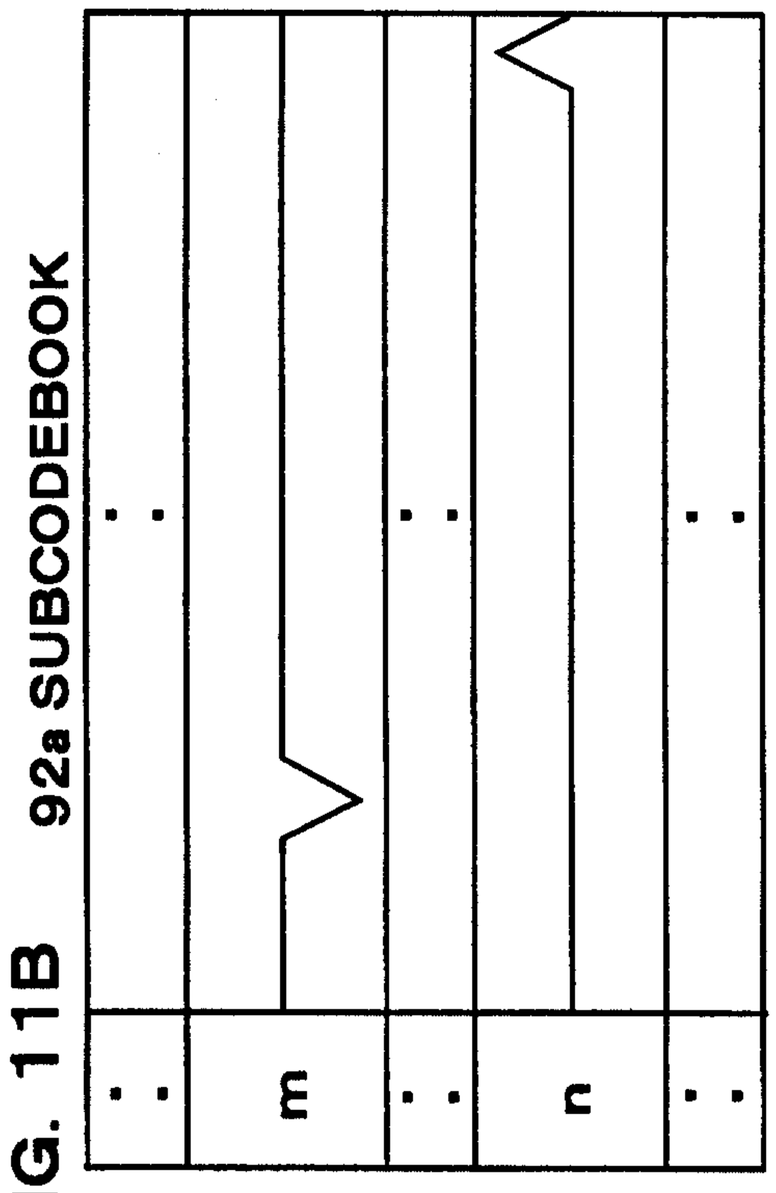


FIG. 11B

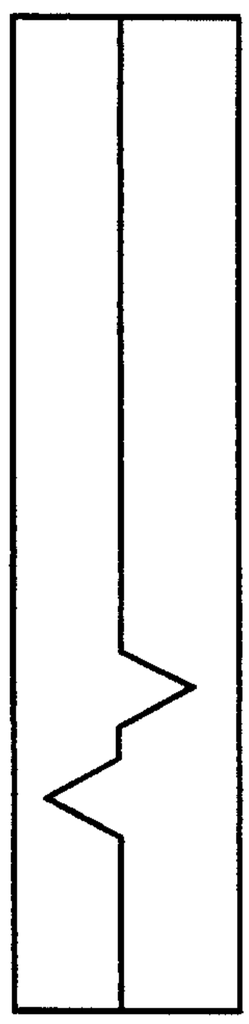
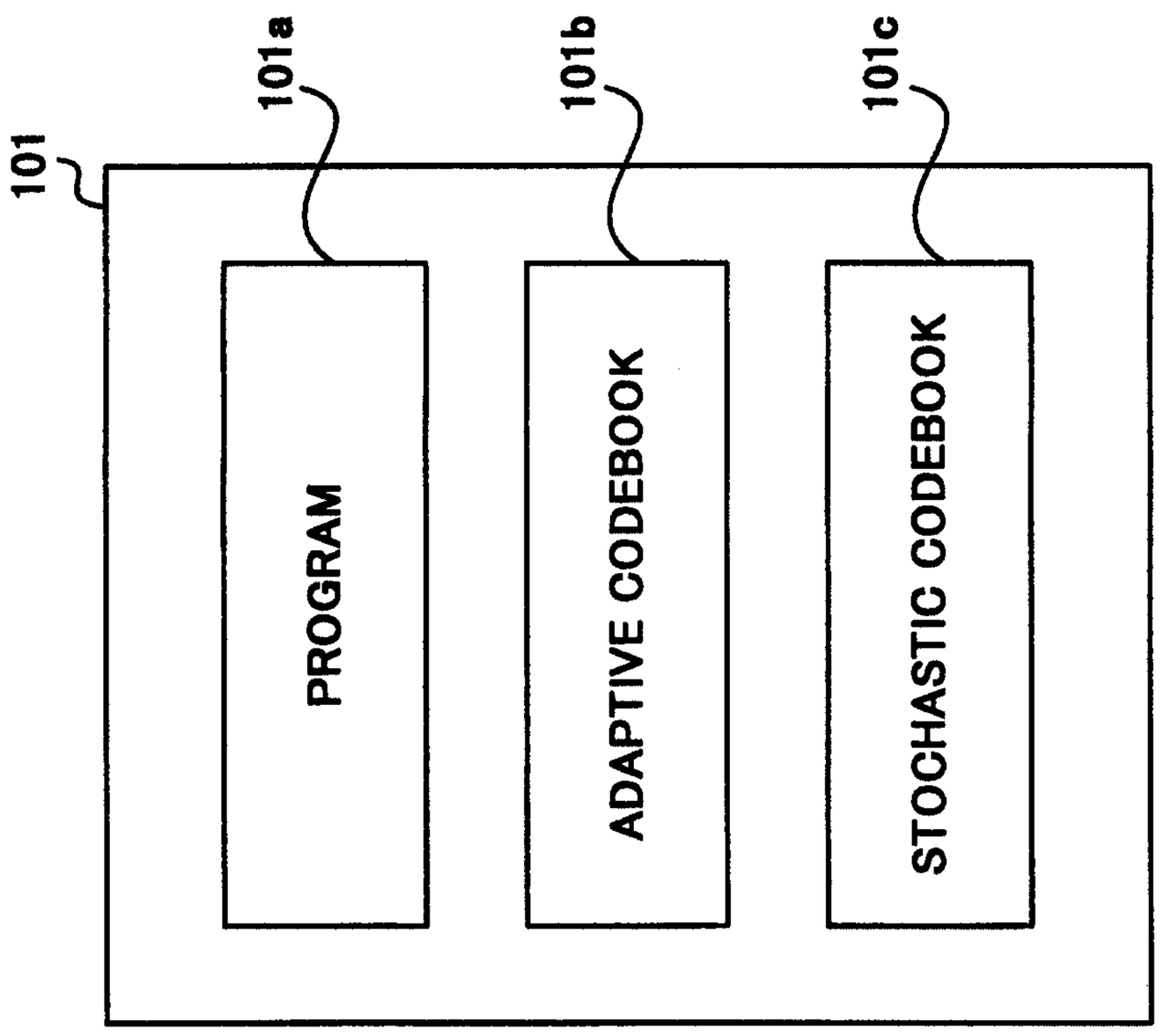
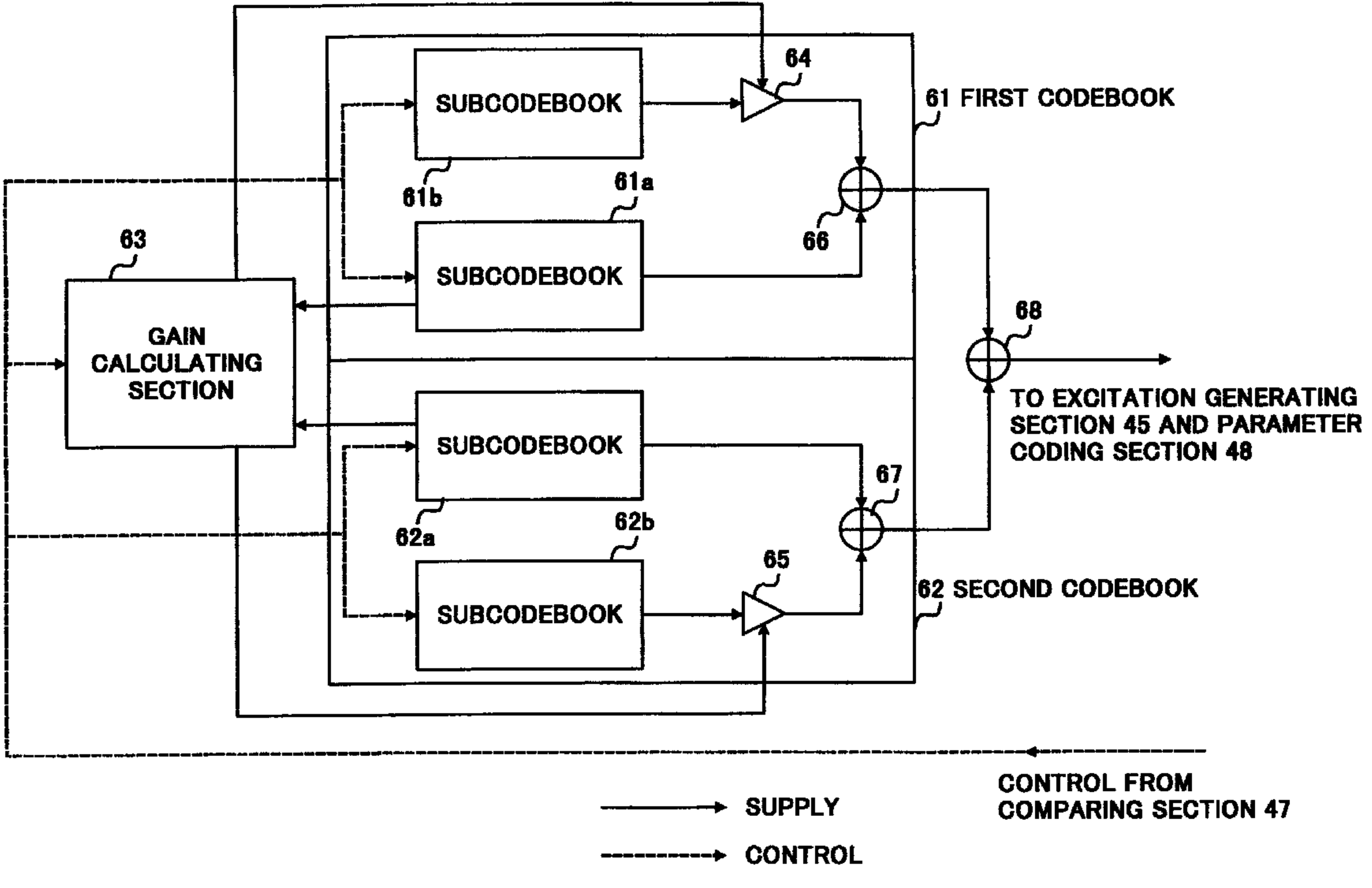


FIG. 11C IN THE CASE OF j+m

FIG.12


101
101a
101b
101c
PROGRAM
ADAPTIVE CODEBOOK
STOCHASTIC CODEBOOK

*rB


SUBCODEBOOK
64
61 FIRST CODEBOOK
61b
61a
SUBCODEBOOK
66
63
GAIN CALCULATING SECTION
68
TO EXCITATION GENERATING SECTION 45 AND PARAMETER CODING SECTION 48
SUBCODEBOOK
67
62a
62b
65
SUBCODEBOOK
62 SECOND CODEBOOK
CONTROL FROM COMPARING SECTION 47
SUPPLY
CONTROL