

620384

FORM 1

SPRUSON & FERGUSON

COMMONWEALTH OF AUSTRALIA

PATENTS ACT 1952

APPLICATION FOR A STANDARD PATENT

NEC Corporation, incorporated in Japan, of 33-1, Shiba 5-chome, Minato-ku, Tokyo 108, JAPAN, hereby apply for the grant of a standard patent for an invention entitled:

Linear Predictive Speech Analysis-Synthesis Apparatus

which is described in the accompanying complete specification.

Details of basic application(s):-

<u>Basic Applic. No:</u>	<u>Country:</u>	<u>Application Date:</u>
75024/1988	JP	28 March 1988

The address for service is:-

Spruson & Ferguson
Patent Attorneys
Level 33 St Martins Tower
31 Market Street
Sydney New South Wales Australia

DATED this TWENTY EIGHTH day of MARCH 1989

NEC Corporation

By:

Registered Patent Attorney

TO: THE COMMISSIONER OF PATENTS
OUR REF: 90923
S&F CODE: 59060

6006388 28/03/89

5845/2

COMMONWEALTH OF AUSTRALIA

THE PATENTS ACT 1952

DECLARATION IN SUPPORT OF A
CONVENTION APPLICATION FOR A PATENTIn support of the Convention Application made for a
patent for an invention entitled:AUSTRALIA
CONVENTION
STANDARD
& PETTY PATENT
DECLARATION
SFP 4

Invention

LINEAR PREDICTIVE SPEECH ANALYSIS-SYNTHESIS APPARATUS

I/We Susumu Uchihara

Full name(s) and
address(es) of
Declarant(s)of c/o NEC Corporation of 33-1, Shiba 5-chome, Minato-ku,
Tokyo, Japan

do solemnly and sincerely declare as follows:—

Full name(s) of
Applicant(s)~~1. I am/We are the applicant(s) for the patent~~*(or, in the case of an application by a body corporate)*

1. I am/We are authorised by NEC Corporation

the applicant(s) for the patent to make this declaration on
its/their behalf.2. The basic application(s) as defined by Section 141 of the
Act was/were made

Basic Country(ies)

in Japan

Priority Date(s)

on 28th March, 1988

Basic Applicant(s)

by NEC Corporation

Full name(s) and
address(es) of
Inventor(s)~~3. I am/We are the actual inventor(s) of the invention referred
to in the basic application(s)~~*(or where a person other than the inventor is the applicant)*

3. Tetsu Taguchi

of c/o NEC Corporation of 33-1, Shiba 5-chome, Minato-ku,
Tokyo, Japan*(respectively)*is/are the actual inventor(s) of the invention and the facts upon
which the applicant(s) is/are entitled to make the application are
as follows:Set out how Applicant(s)
derive title from actual
inventor(s) e.g. The
Applicant(s) is/are the
assignee(s) of the
invention from the
inventor(s)

The Applicant is the assignee of the invention from the inventor.

4. The basic application(s) referred to in paragraph 2 of this
Declaration was/were the first application(s) made in a Convention
country in respect of the invention(s) the subject of the application.

Declared at Tokyo, Japan this 23rd day of March, 1989.

NEC CORPORATION

Signature of Declarant(s)
Susumu Uchihara
General Manager
Patents Division

To: The Commissioner of Patents

11/81

(12) PATENT ABRIDGMENT (11) Document No. AU-B-31754/89
(19) AUSTRALIAN PATENT OFFICE (10) Acceptance No. 620384

- (54) Title
LINEAR PREDICTIVE SPEECH ANALYSIS-SYNTHESIS APPARATUS
- (51)⁴ International Patent Classification(s)
G10L 009/14
- (21) Application No. : 31754/89 (22) Application Date : 28.03.89
- (30) Priority Data
- (31) Number (32) Date (33) Country
63.75024 28.03.88 JP JAPAN
- (43) Publication Date : 28.09.89
- (44) Publication Date of Accepted Application : 20.02.92
- (71) Applicant(s)
NEC CORPORATION
- (72) Inventor(s)
TETSU TAGUCHI
- (74) Attorney or Agent
SPRUSON & FERGUSON, GPO Box 3898, SYDNEY NSW 2001
- (56) Prior Art Documents
AU 520897 37017/78 G10L 1/10, H03H 7/28
EP 149724
- (57) Claim

1. A linear predictive speech analysis-synthesis apparatus having an analysis part receiving an input speech signal and a synthesis part producing a synthesized speech signal,
said analysis part comprising:
means responsive to said input speech signal for extracting first parameters corresponding to linear predictive coefficients representing a spectral envelope of said input speech signal;
means responsive to said input speech signal for extracting second parameter corresponding to pitch information representing a fundamental frequency to said input speech signal;
means responsive to said input speech signal for extracting a third parameter corresponding to power information representing stress of said input speech signal; and
means for transmitting said first parameters, second parameter and third parameter,
said synthesis part comprising:
means for receiving said first parameters, second parameter and third parameter from said analysis part;
means responsive to said first parameters, second parameter and third parameter for generating an exciting source signal, said exciting source signal generating means having a first transfer function;
and

(11) AU-B-31754/89
(10) 620384

-2-

means responsive to said first parameters for synthesizing said synthesized speech signal by filtering said exciting source signal by a second transfer function, said second transfer function being defined by said first parameters and by a damping factor, wherein the product of said first and second transfer functions corresponds to a spectral envelope characteristic of said input speech signal.

620384

S & F Ref: 90923

FORM 10

COMMONWEALTH OF AUSTRALIA

PATENTS ACT 1952

COMPLETE SPECIFICATION

(ORIGINAL)

FOR OFFICE USE:

Class Int Class

Complete Specification Lodged:
Accepted:
Published:

Priority:

Related Art:

Name and Address
of Applicant: NEC Corporation
33-1, Shiba 5-chome
Minato-ku
Tokyo 108
JAPAN

Address for Service: Spruson & Ferguson, Patent Attorneys
Level 33 St Martins Tower, 31 Market Street
Sydney, New South Wales, 2000, Australia

Complete Specification for the invention entitled:

Linear Predictive Speech Analysis-Synthesis Apparatus

The following statement is a full description of this invention, including the best method of performing it known to me/us

LINEAR PREDICTIVE SPEECH ANALYSIS-SYNTHESIS APPARATUS

Background of Invention

The present invention relates to a linear predictive speech analysis-synthesis apparatus and, more particularly, to improvement of a synthesis side thereof.

5 In a conventional linear predictive speech analysis-synthesis apparatus, an impulse train having repetition frequency of a fundamental frequency of an input speech signal is used generally as an exciting source signal on the synthesis side in case the input speech signal is of
10 a voiced sound. An example of this type is disclosed in U.S.P. No. 4,301,329 bearing the title of "SPEECH ANALYSIS AND SYNTHESIS APPARATUS", assigned to this applicant.

In another conventional speech analysis-synthesis apparatus, a pulse train having a shape corresponding
15 to an envelope waveform which is repeated at a fundamental frequency is also used instead of the impulse train.

The above-mentioned conventional linear predictive speech analysis-synthesis apparatuses have the following shortcoming. In the former utilizing the impulse train
20 as the exciting source signal, energy concentrates on a pitch excitation point on the time axis and, thus, a synthesized output speech signal becomes unnatural. In the latter utilizing the shaped pulse train, the exciting source signal becomes colored while the concentration of

energy is avoided. Thus, a synthesized output speech signal become different from an input speech signal in a spectral structure, which results in unnaturalness.

SUMMARY OF THE INVENTION

5 An object of the present invention is, therefore, to furnish a linear predictive speech analysis-synthesis apparatus which is capable of synthesizing a speech signal having excellent sound quality while avoiding concentration of energy and securing the accordance of the spectral structure between an input speech signal and a synthesized
10 output speech signal.

According to one aspect of the present invention there is disclosed a linear predictive speech analysis-synthesis apparatus having an analysis part receiving an input speech signal and a synthesis part producing a synthesized speech signal,

15 said analysis part comprising:

means responsive to said input speech signal for extracting first parameters corresponding to linear predictive coefficients representing a spectral envelope of said input speech signal;

20 means responsive to said input speech signal for extracting second parameter corresponding to pitch information representing a fundamental frequency to said input speech signal;

means responsive to said input speech signal for extracting a third parameter corresponding to power information representing stress of said input speech signal; and

25 means for transmitting said first parameters, second parameter and third parameter,

said synthesis part comprising:

means for receiving said first parameters, second parameter and third parameter from said analysis part;

30 means responsive to said first parameters, second parameter and third parameter for generating an exciting source signal, said exciting source signal generating means having a first transfer function; and

means responsive to said first parameters for synthesizing said synthesized speech signal by filtering said exciting source signal by a
35 second transfer function, said second transfer function being defined by said first parameters and by a damping factor, wherein the product of said first and second transfer functions corresponds to a spectral

amg/0590y



envelope characteristic of said input speech signal.

According to another aspect of the present invention there is disclosed a linear predictive speech synthesis apparatus receiving a pitch parameter representing a fundamental frequency of a speech signal
5 and linear predictive coefficients representing a spectral envelope of said speech signal, comprising:

means for producing an exciting source signal in response to said pitch parameter and said linear predictive coefficients, and

means for filtering said exciting source signal in response to said
10 linear predictive coefficients to produce a synthesized speech signal, wherein a cascade frequency characteristic between spectral envelope frequency characteristics of said producing means and said filtering means is designated to correspond to a spectral envelope characteristic of said speech signal.

15

20

25

30

35

amg/0590y



Brief Description of the Drawings

Fig. 1 is a block diagram of an embodiment according to the present invention;

Fig. 2 is a block diagram of a loss-added synthesizing filter contained in Fig. 1;

Fig. 3 is a block diagram of an exciting source signal generator contained in Fig. 1;

Fig. 4 is a waveform diagram showing a spectral envelope characteristic of the loss-added synthesizing filter according to the present invention in comparison with that of a conventional synthesizing filter;

Fig. 5 is a waveform diagram showing an impulse response characteristic of the present loss-added synthesizing filter in comparison with that of the conventional synthesizing filter; and

Fig. 6 is a waveform diagram showing an output exciting source signal produced by the present invention in comparison with a conventional exciting source signal.

Description of The Preferred Embodiment

In Fig. 1 showing block diagram of one embodiment of the present invention, an analysis side of a linear predictive analysis - synthesis apparatus comprises window processors 1 and 2 receiving an input speech signal, a LPC analyzer 3 receiving an output signal of the window processor 1 and outputting K parameters

k_1 to k_p and a power parameter pw , a K quantizer 4 receiving the K parameters k_1 to k_p , a power quantizer 5 receiving the power parameter pw , a pitch extractor 6 receiving an output signal of the window processor 2 and outputting a pitch parameter pt , a pitch quantizer 7 receiving the pitch parameter pt , and a multiplexer circuit 8 receiving output signals of the K quantizer 4, the power quantizer 5 and the pitch quantizer 7.

Further, a synthesis side of Fig. 1 comprises a separator circuit 9 receiving an output signal of the multiplexer circuit 8 through a transmission channel CH, a K decoder 10, a power decoder 11, a pitch decoder 12, a K/ α converter 13 receiving the K parameters k_1 to k_p from the K decoder 10 and outputting parameters α_1 to α_p , a exciting source signal generator 14 receiving the power parameter pw from the power decoder 11, the pitch parameter pt from the pitch decoder 12 and the parameters α_1 to α_p from the K/ α converter 13, and a loss-added synthesizing filter 15 receiving an exciting output signal from the exciting source signal generator 14 and the α parameters α_1 to α_p from the K/ α converter 13 and outputting an output speech signal.

The feature of the present invention resides in the exciting source generator 14 which operates on the basis of the α parameters α_1 to α_p and in the loss-added synthesizing filter 15. In Fig. 1, the remaining blocks

except for the exciting source signal generator 14 and the loss-added synthesizing filter 15 are the same as those of the first conventional apparatus. Therefore, the exciting source signal generator 14 and the loss-added synthesizing filter 15 will be described, hereinafter, in detail.

First, a description will be made on the loss-added synthesizing filter 15. Fig. 2 is a block diagram of the loss-added synthesizing filter 15.

The loss-added synthesizing filter 15 comprises a subtracter 31, p multipliers 32 which receive a constant (damping factor) r of $0 < r < 1$ as an input from one input end respectively, p delay circuits 33 which give a delay equal to the sampling period in the window processors 1 and 2, p multipliers 34 which receive the α parameter α_i ($i=1, \dots, p$) and the respective outputs of the delay circuits 33 as an input, and an adder 35. In Fig. 2, the combination of the multiplier 32 and the delay circuit 33 is serially connected as p sets. The output of the i -th delay circuit 33 is also supplied to the other input of the multiplier 34 to which the parameter α_i is inputted. The adder 35 adds up multiplication outputs of all the multipliers 34. The subtracter 31 subtracts the addition output of the adder 35 from an inputted exciting source signal. The subtraction output of the subtracter 31

is also delivered as an output synthesized speech signal. In the loss-added synthesizing filter 15, when the constant γ is set to be 1, in other words, when all multipliers 32 are removed, this synthesizing filter 15 becomes the same as a well known conventional LPC synthesizing filter.

The loss-added synthesizing filter 15 has a construction wherein the loss set by the constant γ is given to each stage of the LPC synthesizing filter, and the waveform response thereof is one obtained by damping a waveform response of the conventional LPC synthesizing filter as shown in Fig. 4 and Fig. 5.

The transfer function $H_1(Z)$ of the loss-added synthesizing filter 15 is expressed by

$$H_1(Z) = \frac{1}{1 - \sum_{i=1}^p \alpha_i \gamma^i Z^{-i}} \quad \dots\dots\dots (1)$$

Besides, the transfer function $H(Z)$ of the conventional LPC synthesizing filter employed for a conventional linear predictive speech analysis-synthesis apparatus is expressed generally by

$$H(Z) = \frac{1}{1 - \sum_{i=1}^p \alpha_i Z^{-i}} \quad \dots\dots\dots (2)$$

Examples of frequency transmission characteristics (spectral envelope characteristics) of $H(Z)$ and $H_1(Z)$

are shown in Fig. 4, and examples of impulse responses thereof are shown in Fig. 5. $H_1(Z)$ in Figs. 4 and 5 is one obtained when $r = 0.8$. When this coefficient r is set at 1.0, $H_1(Z)$ is equal to $H(Z)$. When $r = \text{Zero}$, the frequency transmission characteristic of $H_1(Z)$ is leveled completely, and the impulse response is turned to be a unit pulse.

A loss-added synthesizing filter having the same transfer function as the loss-added synthesizing filter 15 can be constructed as well when all the multipliers 32 are removed while a value $\alpha_i r^i$ is inputted, instead of the α parameter α_i , to the multiplier 34.

Next, a description will be made on the exciting source signal generator 14.

Fig. 3 is a block diagram of the exciting source signal generator 14, which comprises a clock generator 20, a pulse generator 21, a standard type digital filter 22 which receives output signals of the clock generator 20, and the pulse generator 21, and the α parameters α_1 to α_p as inputs, delay circuits 23 in a plurality (the number thereof will be mentioned later) which are connected in cascade to the output of the digital filter 22 and receive the clock of the clock generator 20, a pulse train generator 24 which receives the pitch parameter p_t , a noise generator 25, a switching unit 26 which selects the output of either the pulse train generator 24 or

the noise generator 25 under the control of the pitch parameter pt , a plurality of delay circuits 27 which give a delay equal to the sampling period in the window processors 1 and 2, respectively, and which are connected
5 in cascade to the output of the switching unit 26 and numbering less than the delay circuits 23 by one, a plurality of multipliers 28 which receive the set of the outputs of the delay circuits 23 and 27 arranged in the same sequence with each other from the last ones,
10 a multiplier 28, which receives the output of the delay circuit 23 disposed at the first stage and the input to the delay circuit 27 disposed at the first stage, an adder 29 which adds up the multiplication outputs of all of the multipliers 28 and 28', and a multiplier 30
15 which multiplies the power parameter pw by the addition output of the adder 29 and delivers the multiplication output as an exciting source signal. According to a conventional exciting source signal generator, the output of the switching unit 26 is delivered as an
20 output exciting source signal after multiplication by the power parameter pw .

The pulse train generator 24 generates a impulse train at a repetition frequency corresponding to a pitch period in the pitch parameter pt . The noise
25 generator 25 outputs white noise of M sequences or the like. The switching unit 26 selects the output

impulse train from the pulse generator 24 in the case of a voiced sound or selects the noise from the noise generator 25 in the case of an unvoiced sound, corresponding to the result of determination of the pitch parameter p_t , and delivers the selected output as an exciting pulse.

In Fig. 3, components other than the pulse train generator 24, the noise generator 25 and the switching unit 26 are excited by the exciting pulse from the switching unit 26 and the exciting source signal to be outputted is produced in the following.

In relation to the transfer function $H(z)$ (set by the α parameters α_1 to α_p) of the LPC synthesizing filter and the transfer function $H_1(z)$ (set by the parameters α_1 to α_p) of the loss-added synthesizing 15, which are described previously, the standard type digital filter 22 is so constructed that its transfer function is

$$H_2(z) = \frac{H(z)}{H_1(z)} = \frac{1 - \sum_{i=1}^p \alpha_i r^i z^{-i}}{1 - \sum_{i=1}^p \alpha_i z^{-i}} \quad \dots\dots (3)$$

The clock generator 20 outputs the clock in the number corresponding to a required impulse response length of the standard type digital filter 22 for every analysis frame. The repetition frequency of the clock is set to be shorter enough than the sampling frequency in the

window processors 1 and 2. The pulse generator 21 outputs one impulse for each analysis frame. Each delay circuit 23 is constructed by D-type flip-flops each using the clock outputted from the clock generator 20 as an operating pulse. Particularly, the flip-flops are combined in parallel for the required number of bits. The number of the delay circuits 23 is made to be equal to the number of generated clock pulses of the clock generator 20 during the analysis frame.

10 In each analysis frame, the α parameters α_1 to α_p are inputted so that the transfer function $H_2(z)$ of the digital filter 22 is set. Subsequently, the impulse is inputted from the pulse generator 21, and the digital filter 22 is made to operate by the clock from the clock generator 20. When a plurality of clocks are outputted for the entire frame, a signal representing the impulse response of the standard type digital filter 22 is obtained in the output of each delay circuit 23, and it is held until a subsequent analysis frame comes.

20 In Fig. 3, a combination of the delay circuits 27, the multipliers 28 and the adder 29 composes a transversal filter having an impulse response which corresponds to the inversion of the impulse response of the digital filter 22 on a time basis. Namely, 25 in this configuration, each tap coefficient is obtained from each delay circuit 23 and each circuit 23 and

each multiplier 28 are connected as shown in the drawing. The exciting pulse from the switching unit 26 is applied to this transversal filter, and the output of this filter is made to correspond to the power of the input speech signal by the multiplier 30. Thus, the result is delivered as the exciting source signal to the loss-added synthesizing filter 15. In this case, it is possible that the multiplier 30 is inserted just behind the switching unit 26 instead of just behind the adder 29.

10 The spectral structure of the exciting source signal from the exciting source signal generator 14 is equal to the spectral structure of the output obtained by that the digital filter having the transfer function $H_2(z)$ is excited by the exciting pulse from the switching unit 26. Since this exciting source signal is outputted through the loss-added synthesizing filter 15 having the transfer function $H_1(z)$, the spectral structure of the synthesized output speech signal accords with a spectral structure which is obtained by exciting the

15 LPC synthesizing filter having the transfer function $H(z)$ ($= H_1(z) \times H_2(z)$) by the exciting pulse and, consequently, the synthesized output speech signal accords with the spectral structure of the input speech signal.

20 In addition, according to the present invention, since the impulse response of the transversal filter,

which produces the exciting source signal from the exciting pulse, is formed as the time-inversed impulse response as compared with that of the digital filter having the transfer function $H_2(z)$, phase relationship in the process, wherein the synthesized output speech signal is formed from the exciting pulse, is made to be different from phase relationship in processing of the LPC synthesizing filter having the transfer function $H(z)$. Thus the energy in the synthesized output speech signal does not concentrate on a pitch excitation point even when the impulse train is applied as the exciting pulse.

With regard to the constant γ applied to the loss-added synthesizing filter 15 and the digital filter 22 in the exciting source signal generator 14, its value is determined through computer simulation or through experimentation. In practice, one preferable value is about 0.8 to derive a good result.

Fig. 6 shows waveforms of the exciting source signal according to the present invention as compared with a conventional exciting source signal. In this figure, S_1 indicates the conventional exciting source signal, i.e., the impulse train. S_2 indicates the exciting source signal in case of $\gamma = 1$ and S_3 indicates the exciting source signal in case of $\gamma = 0.8$. When $\gamma = 1$, the loss-added synthesizing filter 15 becomes equal to

the conventional LPC synthesizing filter as described above. However, in the exciting source signal generator 14, a certain effect can be obtained even when $\gamma = 1$.

As described above, according to the present invention, by providing the loss-added synthesizing filter having the function $H_1(z)$ and the exciting source signal generator which forms the exciting source signal from the exciting pulse by using the filter having the function $H_2(z)$ ($= \frac{H(z)}{H_1(z)}$) and the transversal filter having the time-inverted impulse response, the linear predictive speech analysis-synthesis apparatus, which is capable of producing the synthesized output speech signal wherein no energy concentrates on a pitch excitation point and the accordance is established in the spectral structure between the input speech signal and the output speech signal, thus resulting in excellent sound quality, is obtained.

The claims defining the invention are as follows:

1. A linear predictive speech analysis-synthesis apparatus having an analysis part receiving an input speech signal and a synthesis part producing a synthesized speech signal,

5 said analysis part comprising:

 means responsive to said input speech signal for extracting first parameters corresponding to linear predictive coefficients representing a spectral envelope of said input speech signal;

 means responsive to said input speech signal for extracting
10 second parameter corresponding to pitch information representing a fundamental frequency to said input speech signal;

 means responsive to said input speech signal for extracting a third parameter corresponding to power information representing stress of said input speech signal; and

15 means for transmitting said first parameters, second parameter and third parameter,

 said synthesis part comprising:

 means for receiving said first parameters, second parameter and third parameter from said analysis part;

20 means responsive to said first parameters, second parameter and third parameter for generating an exciting source signal, said exciting source signal generating means having a first transfer function; and

 means responsive to said first parameters for synthesizing
25 said synthesized speech signal by filtering said exciting source signal by a second transfer function, said second transfer function being defined by said first parameters and by a damping factor, wherein the product of said first and second transfer functions corresponds to a spectral envelope characteristic of said input speech signal.

30 2. A linear predictive speech analysis-synthesis apparatus as claimed in claim 1, wherein said exciting source signal generating means includes:

 a impulse generator for generating a impulse for each analysis frame period;

35 filter means responsive to said first parameters for filtering said impulse from said impulse generator, said filter means having a function corresponding to said first function;

amg/0590y



first delay array means for sequentially delaying the output of said filter means to deliver a plurality of first delay outputs each having a different delay time;

exciting pulse generating means responsive to said second
5 parameter for generating an exciting pulse;

transversal filter means for filtering said exciting pulse from said exciting pulse generating means to produce said exciting source signal, said transversal filter means receiving said plurality of first delay outputs as a plurality of coefficients; and

10 means for controlling the level of said exciting source signal in response to said third parameter.

3. A linear predictive speech analysis-synthesis apparatus as claimed in claim 1, wherein said first function is defined by

$$15 \quad \frac{1 - \sum_{i=1}^P \alpha_i \gamma^i z^{-i}}{1 - \sum_{i=1}^P \alpha_i z^{-i}}$$

where: α_i corresponds to said first parameters and,
20 γ corresponds to said damping factor,
said second function is defined by

$$25 \quad \frac{1}{1 - \sum_{i=1}^P \alpha_i \gamma^i z^{-i}}$$

4. A linear predictive speech synthesis apparatus receiving a pitch parameter representing a fundamental frequency of a speech signal and linear predictive coefficients representing a spectral envelope of
30 said speech signal, comprising:

means for producing an exciting source signal in response to said pitch parameter and said linear predictive coefficients, and

means for filtering said exciting source signal in response to said linear predictive coefficients to produce a synthesized speech
35 signal, wherein a cascade frequency characteristic between spectral envelope frequency characteristics of said producing means and said filtering means is designated to correspond to a spectral envelope

4

-16-

characteristic of said speech signal.

5. A linear predictive speech analysis-synthesis apparatus substantially as described with reference to the accompanying drawings.

5 DATED this TWENTY FIFTH day of NOVEMBER 1991
NEC Corporation

10 Patent Attorneys for the Applicant
SPRUSON & FERGUSON

15

20

25

30

35

amg/0590y



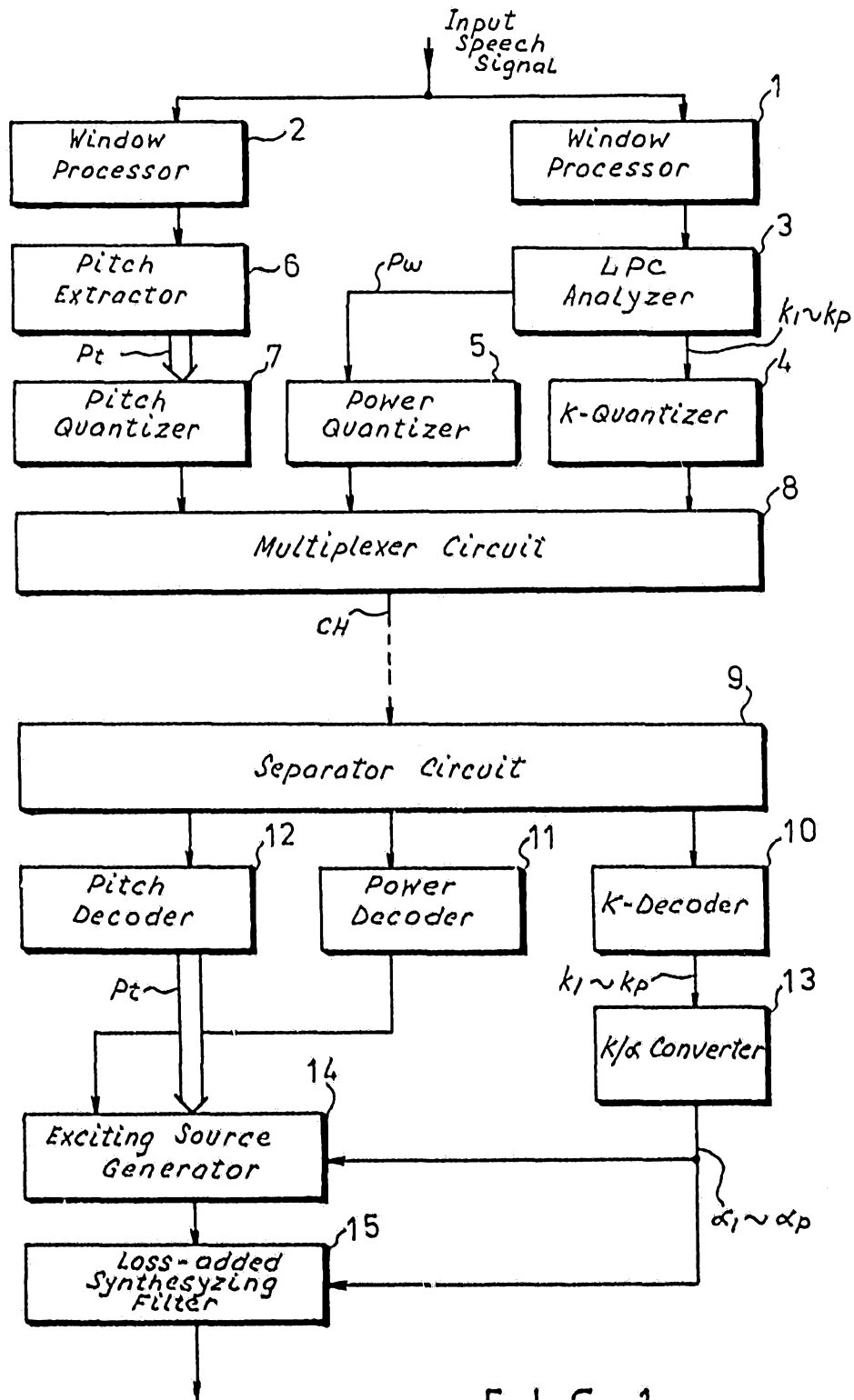


FIG. 1

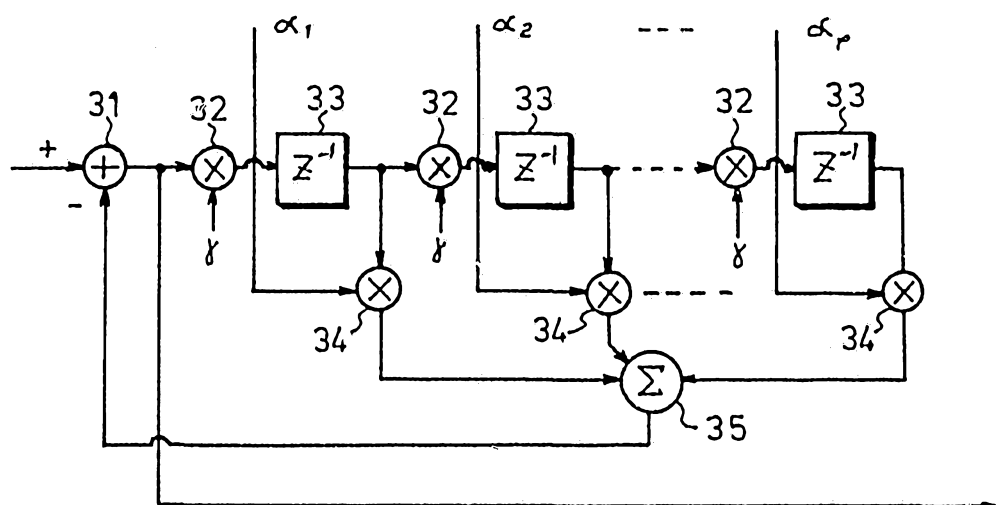


FIG. 2

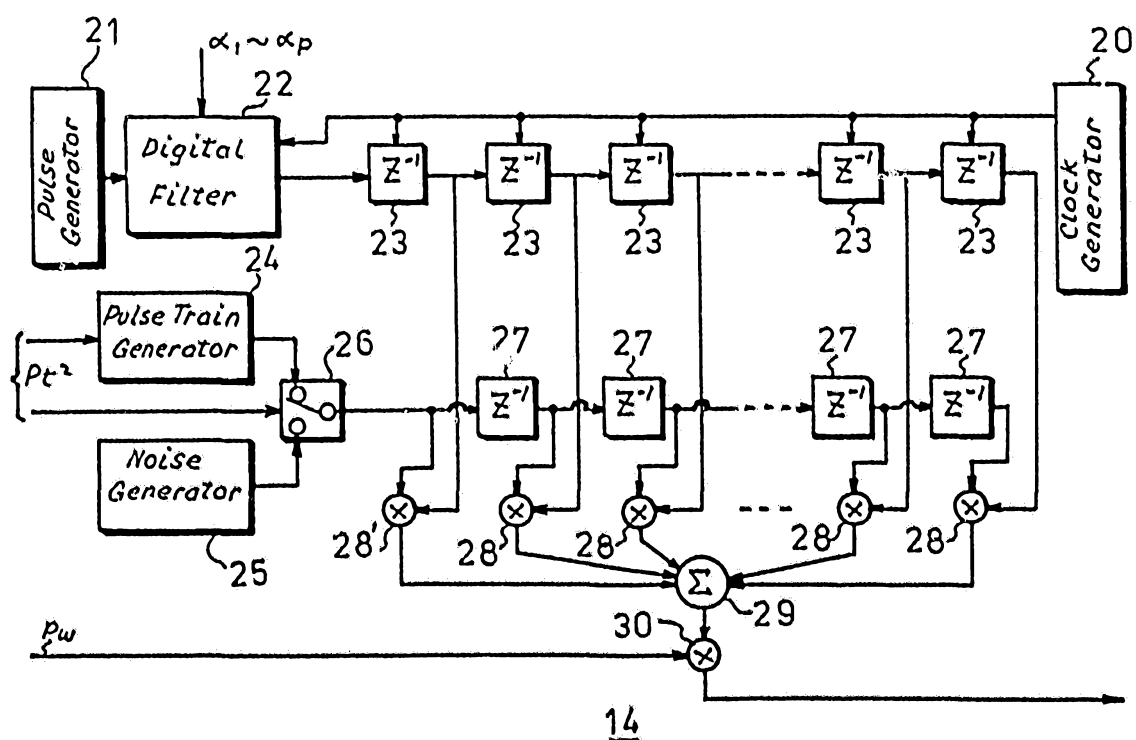


FIG. 3

