United States Patent [19]

Willimann

[11] **3,909,533**

[45] **Sept. 30, 1975**

[54] **METHOD AND APPARATUS FOR THE ANALYSIS AND SYNTHESIS OF SPEECH SIGNALS**

[75] Inventor: **Louis Sepp Willimann,** Eschenbach, Switzerland

[73] Assignee: **Gretag Aktiengesellschaft,** Switzerland

[22] Filed: **Oct. 8, 1974**

[21] Appl. No.: **513,160**

[30] **Foreign Application Priority Data**

July 22, 1974 Switzerland 10066/74

[52] **U.S. Cl.** ..... **179/1 SA;** 179/1 SA

[51] **Int. Cl.²** **G10L 1/00**

[58] **Field of Search** ..... 179/1 SA

[56] **References Cited**

UNITED STATES PATENTS

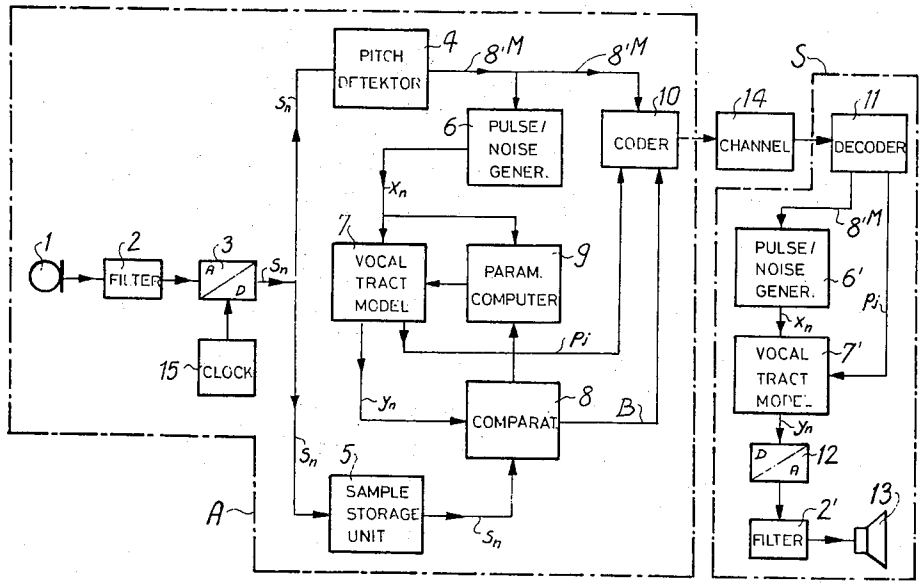
3,624,302	11/1971	Atal	179/1 SA
3,631,520	12/1971	Atal	179/1 SA

Primary Examiner—Kathleen H. Claffy
Assistant Examiner—E. S. Kemeny
Attorney, Agent, or Firm—Pierce, Scheffler & Parker

[57] **ABSTRACT**

Synthesized speech is produced by a vocal tract model corresponding functionally to the human vocal tract and constructed of a linear digital filter. The parameter of the synthesis vocal tract model are determined by an analysis operation on an original speech signal using an identical vocal tract model, which may be the same model as used for synthesis. The analysis vocal tract model has its parameters adjusted according to a comparison between the original speech signal and the output signal of the analysis model so as to minimize the deviation between these two signals. Those parameters for which the deviation falls below a threshold value are used directly as parameters of the synthesis vocal tract model. The adjustments of the parameters are determined by a parameter computer working on the results of the aforementioned comparison and which itself may contain the vocal tract model.

11 Claims, 6 Drawing Figures



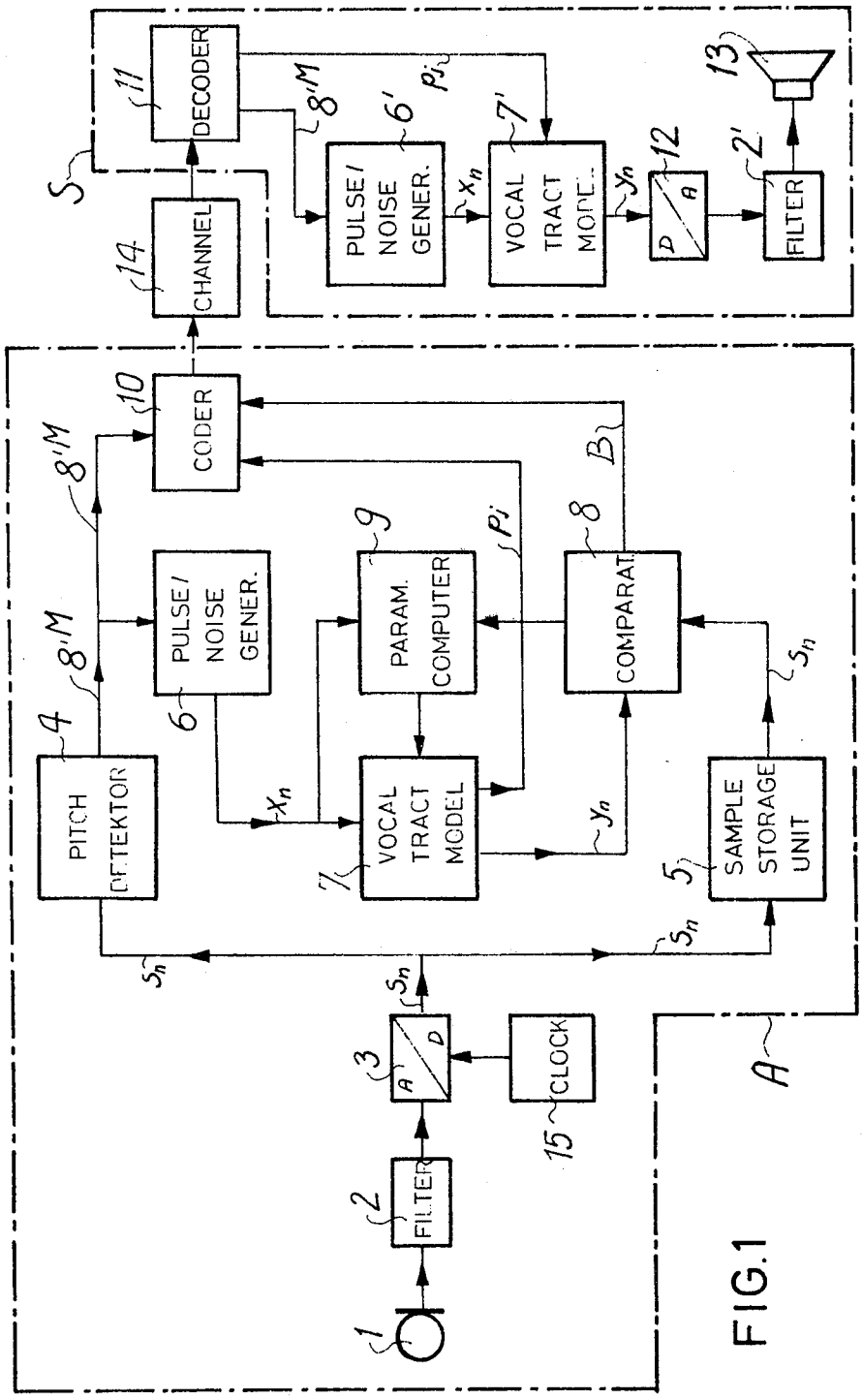

S
11
DECODER
8'M
6'
Pi
7'
13
PULSE/ NOISE GENER.
VOCAL TRACT MODEL
12
2'
FILTER
14
CHANNEL
10
CODER
B
9
8
PARAM. COMPUTER
COMPARAT.
4
PITCH DETEKTOR
6
PULSE/ NOISE GENER.
7
VOCAL TRACT MODEL
5
SAMPLE STORAGE UNIT
3
CLOCK
15
2
FILTER
1
A

FIG.1

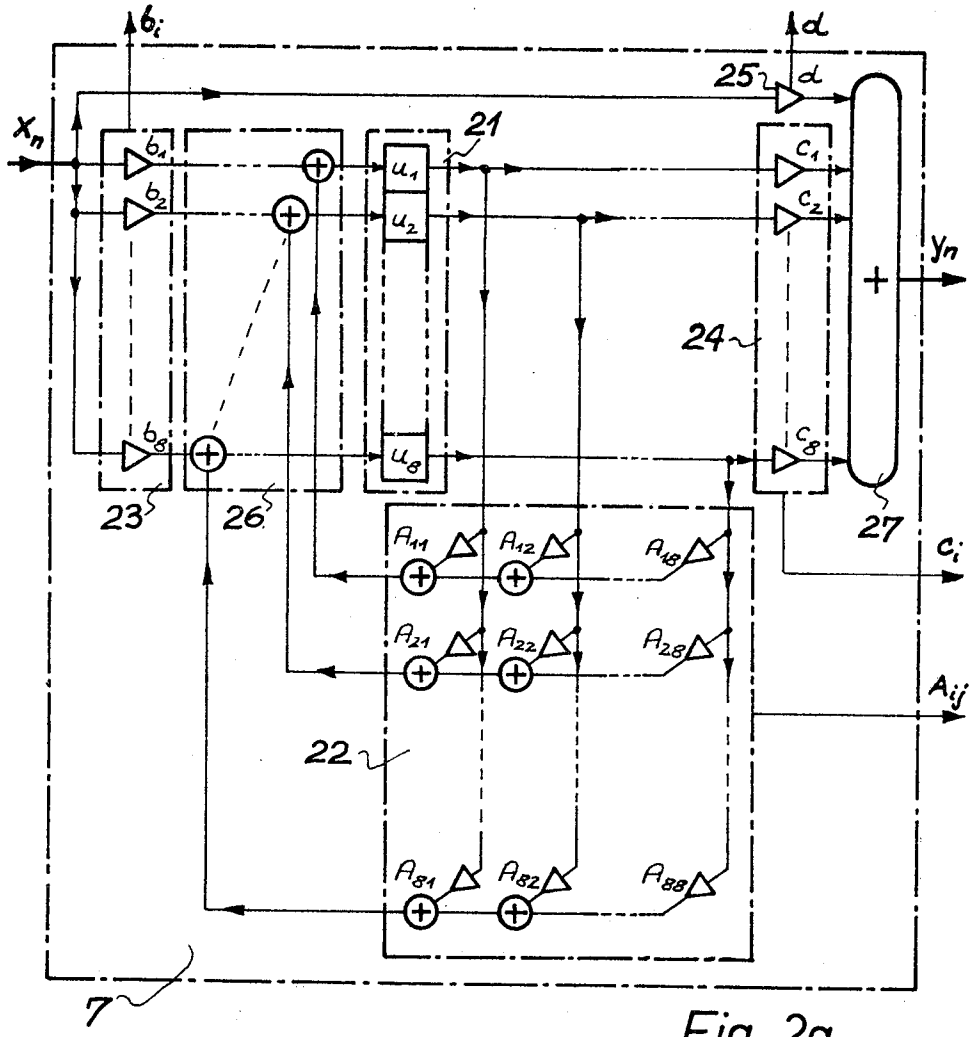

b_i
d
25
x_n
21
b_1
b_2
b_8
u_1
u_2
u_8
c_1
c_2
c_8
y_n
24
23
26
27
A_{11} A_{12} A_{18}
A_{21} A_{22} A_{28}
A_{81} A_{82} A_{88}
c_i
A_{ij}
22
7

Fig. 2a

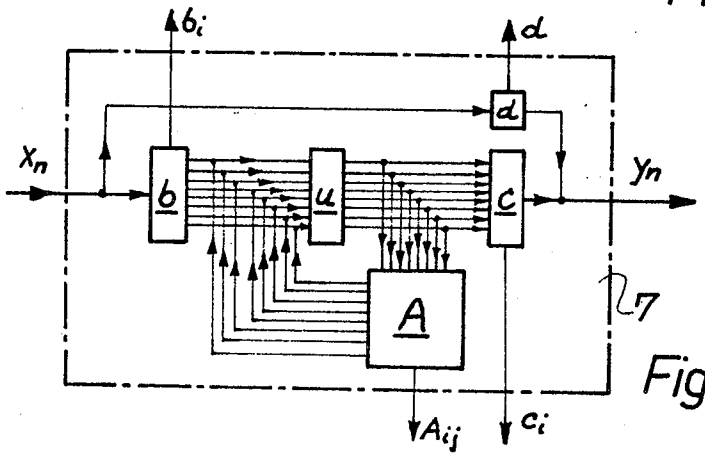

b_i
d
x_n
b
u
c
y_n
A
7
A_{ij}
c_i

Fig. 2b

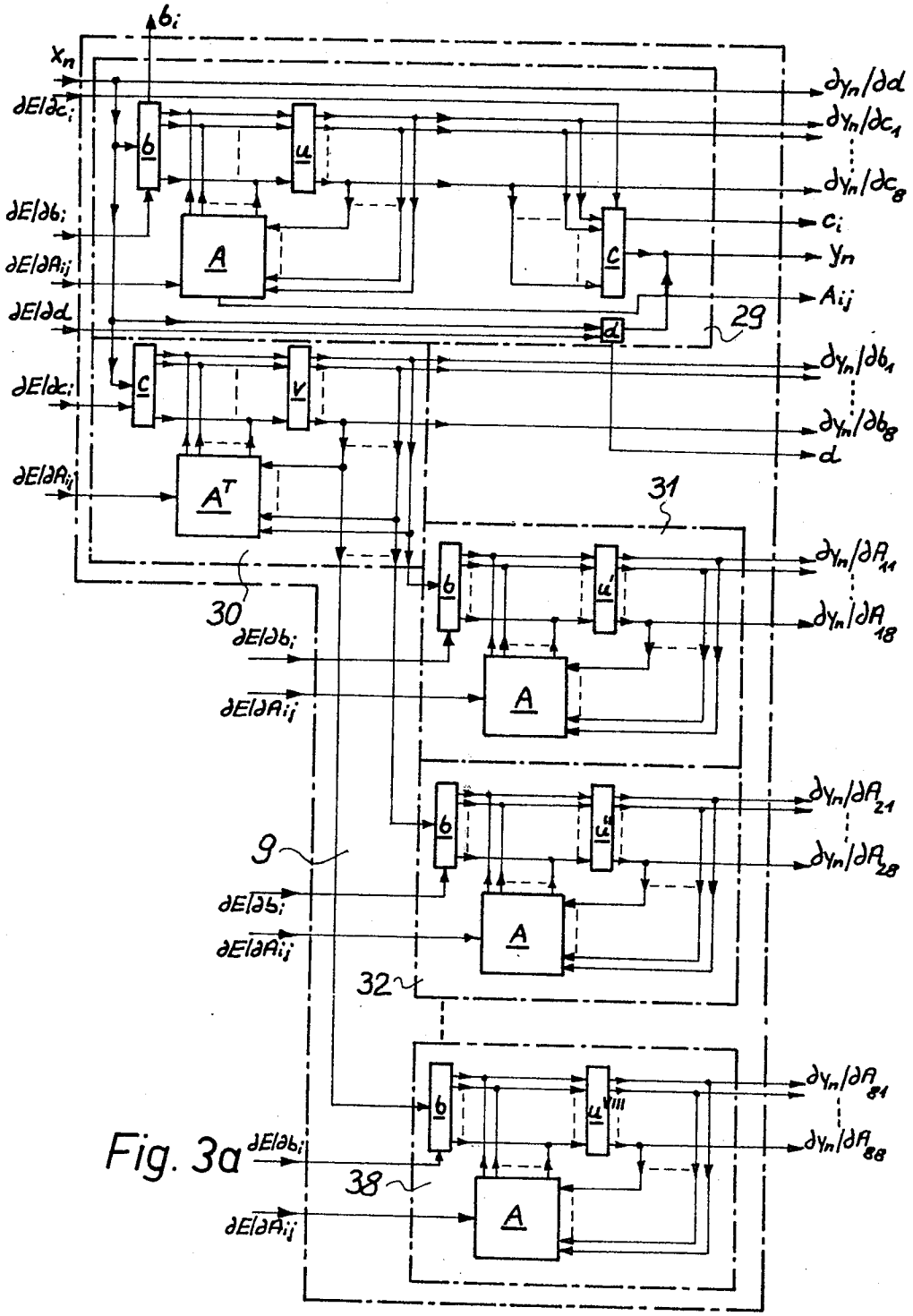

b_i
x_n
$\partial E/\partial c_i$
$\partial E/\partial b_i$
$\partial E/\partial A_{ij}$
$\partial E/\partial d$
$\partial E/\partial c_i$
$\partial E/\partial A_{ij}$
b
u
A
c
v
A^T
d
29
31
30
9
32
38
$\partial y_n/\partial d$
$\partial y_n/\partial c_1$
$\partial y_n/\partial c_8$
c_i
y_n
A_{ij}
$\partial y_n/\partial b_1$
$\partial y_n/\partial b_8$
$\partial y_n/\partial A_{11}$
$\partial y_n/\partial A_{18}$
$\partial y_n/\partial A_{21}$
$\partial y_n/\partial A_{28}$
$\partial y_n/\partial A_{81}$
$\partial y_n/\partial A_{88}$

Fig. 3a

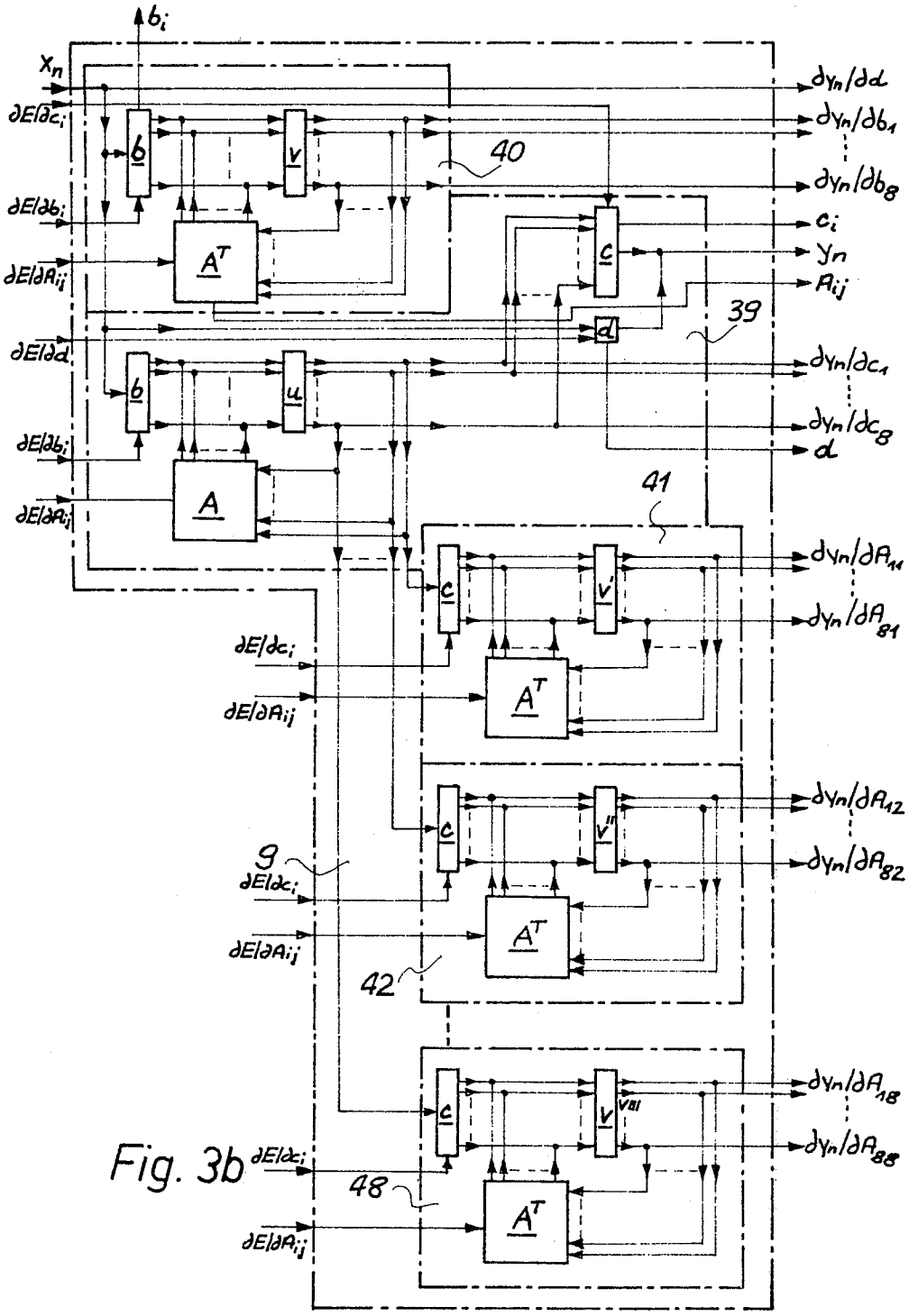

b_i
x_n
$\partial E/\partial c_i$
$\partial E/\partial b_i$
$\partial E/\partial A_{ij}$
$\partial E/\partial d$
$\partial E/\partial b_i$
$\partial E/\partial A_{ij}$
b
v
A^T
40
c
d
39
u
A
41
v'
v''
v^{VIII}
9
42
48
$\partial y_n/\partial d$
$\partial y_n/\partial b_1$
$\partial y_n/\partial b_8$
c_i
y_n
A_{ij}
$\partial y_n/\partial c_1$
$\partial y_n/\partial c_8$
d
$\partial y_n/\partial A_{11}$
$\partial y_n/\partial A_{81}$
$\partial y_n/\partial A_{12}$
$\partial y_n/\partial A_{82}$
$\partial y_n/\partial A_{18}$
$\partial y_n/\partial A_{88}$

Fig. 3b

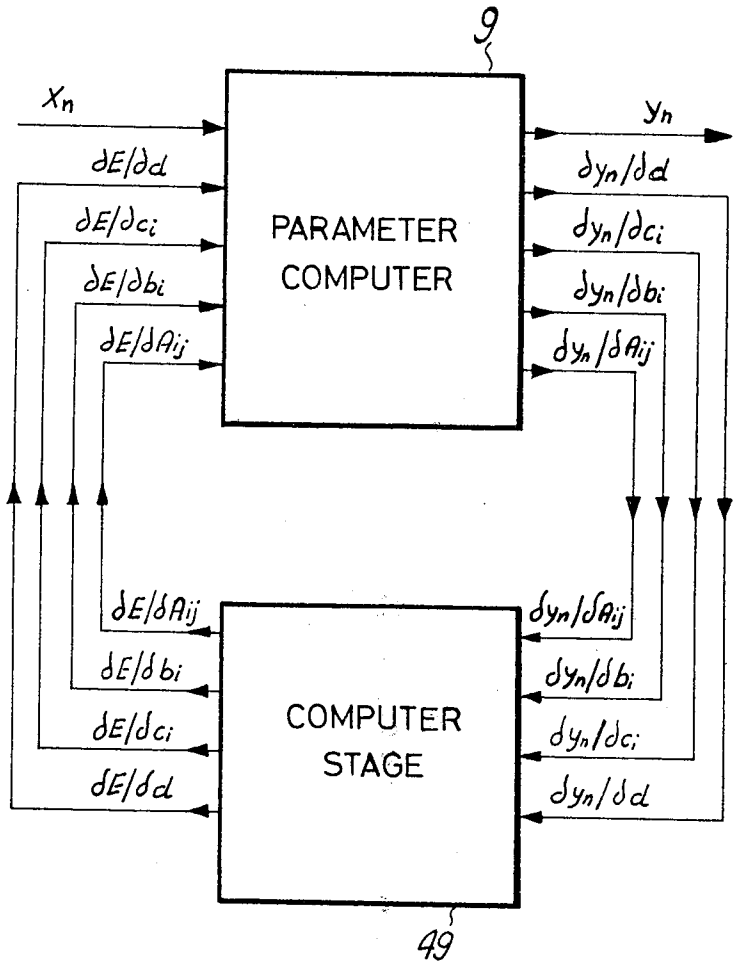

9
x_n
y_n
$\partial E/\partial d$
$\partial E/\partial c_i$
$\partial E/\partial b_i$
$\partial E/\partial A_{ij}$
PARAMETER COMPUTER
$\partial y_n/\partial d$
$\partial y_n/\partial c_i$
$\partial y_n/\partial b_i$
$\partial y_n/\partial A_{ij}$
$\partial E/\partial A_{ij}$
$\partial E/\partial b_i$
$\partial E/\partial c_i$
$\partial E/\partial d$
COMPUTER STAGE
$\partial y_n/\partial A_{ij}$
$\partial y_n/\partial b_i$
$\partial y_n/\partial c_i$
$\partial y_n/\partial d$
49

FIG. 4

METHOD AND APPARATUS FOR THE ANALYSIS AND SYNTHESIS OF SPEECH SIGNALS

FIELD OF THE INVENTION

This invention relates to a method and apparatus for the analysis and synthesis of speech signals.

A problem arising in the transmission of speech signals is to reduce the amount of speech information by eliminating speech redundancy, when the signals are in digital or pulse amplitude modulated form and are transmitted via limited bandwidth channels, or when the speech signals are stored in a store of limited capacity for example the memory of a computer.

PRIOR ART

Two methods have been proposed to solve this problem, one using apparatus known as a vocoder and the other a predictor.

The vocoder is based on the relationship between the spectral components of a sound and the redundancy reduction. This is possible because the voiced sounds, for example the vowels of a speech signal, have a quasi-periodic character. The associated frequency spectrum is accordingly linear, the space between the individual spectral lines being equivalent to a particular fundamental frequency, the pitch frequency. Unfortunately, the speech signal synthesized by the vocoder is of poor quality.

Redundancy reduction in the predictor is based on the statistical relationship between consecutive instantaneous values of the speech information as a function of time and only those instantaneous values which are substantially independent of one another and are situated outside a given tolerance interval, are transmitted. For this purpose, the transmission side determines, for each instantaneous value to be transmitted, whether it is substantially independent of the already transmitted preceding instantaneous values, while on the reception side the dependent instantaneous values which have not been transmitted are determined or interpolated. The predictor-synthesized speech signal has a very good quality but determination of the instantaneous value to be transmitted may in certain circumstances be expensive.

SUMMARY OF THE INVENTION

The present invention relates to a method of analysing and synthesizing speech, in which, for analysis, the original speech signal is sampled and three groups of signals representing each speech signal are derived for each sample. The first group of signals represents the parameters of a synthesis vocal tract model which functionally corresponds to the human vocal tract and which is contructed essentially from a discrete linear filter. The second and third group of signals respectively represent the fundamental frequency reciprocal (hereinafter referred to as the "pitch period") and the voiced/voiceless character of the original speech signal for the sample in question. For synthesis, the synthesis vocal tract model is adjusted by reference to the first group of signals. During voiced samples of the original speech signal the vocal tract model is excited by a train of pitch period spaced pulses, and during voiceless samples, is excited by white noise, a synthetic speech signal similar to the original speech signal thus being produced at the output of the synthesis vocal tract model.

In a method of this kind disclosed in U.S. Pat. No. 3,624,302, the first group of signals, i.e., the predictor parameters, are computed arithmetically from the statistical relationship of 12 consecutive sampled values of the original speech signal. Since a linear equation system has to be solved for this purpose and the zeros of a 12th-degree polynomial have to be determined, the calculations are considerable and can be mastered only by a computer. Also, in this method, the energy of the original speech signal has to be determined for each sample.

This invention obviates these disadvantages and is characterised in that an analyser vocal tract model identical to the synthesizer vocal tract model is used on analysis in order to produce the signals representing the parameters of the synthesizer vocal tract model and, during voiced samples of the original speech signal, is excited by a pitch period spaced pulse train and, during voiceless samples, is excited by white noise, the output signal of the analyser vocal tract model is sampled for comparison with the original speech signal and the deviation between the two signals is minimized by changing the parameters of the analyser vocal tract model, and those parameters of the analyser vocal tract model for which the deviation falls below a predetermined threshold value are used directly as the first group of signals.

The invention also relates to apparatus for performing the method, in which the synthesizer comprises a synthesizer vocal tract model and a pulse/noise generator and the analyzer is provided with means for determining the parameters of the synthesizer vocal tract model, means for determining the pitch period and means for determining the voiced/voiceless character of the original speech signal.

This apparatus is characterized in that the means for determining the parameters of the synthesizer vocal tract model are formed by an analyser vocal tract model identical to the synthesizer vocal tract model, a pulse/noise generator identical to the synthesizer pulse/noise generator, a sample storage unit for storing samples of the original speech signal, a comparator for comparing the output signal of the analyser vocal tract model with the signal stored in the sample storage unit, and a parameter computer for minimizing the deviation between the two signals as determined in the comparator.

Since, therefore, in the apparatus according to the invention the essential components of the analyser and synthesizer are identical they can be used, for example, in the transmission of speech signals in alternate send-receive operation without appreciable additional expense. Another advantage over the apparatus operating in accordance with the known method is that the analyser and synthesizer vocal tract models are formed by any linear digital filter and it is therefore possible to use one having low quantization sensitivity. In the known apparatus, however, a quite specific recursive filter is used, i.e., the Frobenius form, in which the feedback consists of a transversal filter. It is known that the coefficients of this form are extremely quantization-sensitive.

BRIEF DESCRIPTION OF THE DRAWINGS

An embodiment of the invention will now be explained in detail with reference to the accompanying drawings in which:

FIG. 1 is a block schematic diagram of apparatus for speech analysis and synthesis;

FIG. 2*a* is a block schematic diagram of the vocal tract model forming part of the apparatus of FIG. 1;

FIG. 2*b* is a simplified block schematic diagram of the arrangement shown in FIG. 2*a*;

FIG. 3*a* is a block schematic diagram of the parameter computer shown in FIG. 1;

FIG. 3*b* is a variant of the circuit shown in FIG. 3*a*; and

FIG. 4 is a block schematic diagram of a computer stage which operates in conjunction with the parameter computer.

DETAILED DESCRIPTION OF EMBODIMENT OF THE INVENTION

A complete speech analyser and synthesizer is shown in FIG. 1 which comprises an analyser A and a synthesizer S. As shown in the drawing, a transmission or storage medium **14**, for example a digital transmission channel or a digital storage unit, is disposed between the output of the analyser and the input of the synthesizer.

Analyser A consists of a speech source **1**, a low-pass filter **2**, an analog-to-digital converter **3**, a clock **15** which energises the entire analyser A, a pitch detector **4**, a sample storage unit **5**, a pulse/noise generator **6**, an analyser vocal tract model **7**, a comparator **8**, a parameter computer **9**, and a coder **10**.

The synthesizer S consists of a decoder **11**, a pulse/noise generator **6'**, a synthesizer vocal tract model **7'**, a digital-to-analog converter **12**, a low-pass filter **2'** and a unit **13**, for example a loudspeaker. The low-pass filters **2**, **2'**, the pulse/noise generators **6**, **6'** and the vocal tract models **7**, **7'** are of identical construction in both the analyser A and the synthesizer S. Given appropriate facilities for changeover to analysis or synthesis, there need be only one of each of these three devices.

ANALYSER

The speech signal to be analysed is passed from the source **1**, for example a microphone or analog storage unit, to the low-pass filter **2** which has a specific cut-off frequency fg, for example 5 – 5 kHz. The output signal of the low-pass filter **2** is sampled and digitized in the analog-to-digital converter **3** at a sampling frequency of 2fg, for example, 6–10 kHz. The resulting sequence of sampled values s_n is then fed to the pitch detector **4** and to the sample storage unit **5**.

In the sample storage unit **5**, a short sample of the signal for s_n is stored for repeated call. The length of the sample is of the order of one to several pitch periods, i.e., for example about 10 to 30 msec. However, it need not be a complete multiple of a pitch period.

The pitch detector **4** determines by known methods whether the speech sample is or is not voiced as in conventional vocoders. If the sample is voiced, then the length and position of the pitch periods is determined at the same time, the term "pitch period" denoting the interval of time between two glottal pulses produced by the vocal chords in the case of voiced sounds. The pitch detector **4** passes its information in the form of a signal *g* representing the voiced/voiceless decision and period pitch signals M representing the length and position of the pitch periods in the case of voiced samples — to the coder **10** and to the pulse/noise generator **6**.

Under the control of the pitch detector **4**, the pulse/noise generator **6** delivers white noise during voiceless samples of the speech signal and pulse signals with the pitch period spacing during voiced samples of the speech signal. The white noise is generated by a pseudorandom generator of known construction and has a substantially constant power. In the simplest case, the pulses delivered by the pulse/noise generator **6** during voiced samples of the speech signal are ordinary unit pulses, but they may have some other form, for example a triangular form. The pulse sequence power is also substantially constant and is equal to that of the white noise.

The output signal of the pulse/noise generator **6** formed from the white noise or from pitch period spaced pulses forms the excitation signal for the analyser vocal tract model **7**.

By "vocal tract" is meant the system of tubes of variable cross-sectional area between the larynx and lips and between the palate and nostrils. This vocal tract is excited by periodic pulses, the pitch pulses, produced by the glottis during the vowels occurring in speech. In the case of consonants, the vocal tract is excited by substantially white noise. The latter is produced by a stream of air forced through a constriction in the vocal tract, for example the constriction between the upper teeth and lower lip in the case of the consonant *f*.

The model **7** of the human vocal tract is formed by a linear digital filter of any structure. Linear digital filters are described, for example, in H. W. Schüssler: "Digitale Systeme zur Signalverarbeitung," Springer 1973.

Linear digital filters enable an output sequence y_n to be produced from an input sequence x_n in accordance with the following law:

$$u_{n+1} = A \cdot u_n + b \cdot x_n \tag{1a}$$

$$y_n = c^T \cdot u_n + d \cdot x_n \tag{1b}$$

where u_n is the n^{th} state vector of the dimension N. u_o is predetermined and in most cases is the zero vector. The model is completely described by the NxN-matrix A, as the two N-dimensional vectors b and c and by the scalar quantity d.

As already stated, the input sequence x_n is formed by a sequence of pitch period spaced pulses during voiced samples of the speech signal and by white noise during voiceless samples of the speech signal.

On excitation in the manner described, the analyser vocal tract model **7**, explained in detail with reference to FIGS. 2*a* and 2*b* delivers a still untreated speech signal y_n to the comparator **8**, in which this approximation signal is compared with the sample of the original speech signal s_n stored in the sample storage unit **5**.

Any desired criterion which constitutes a mathematical dimension of the deviation between the two sequences y_n and s_n and which in respect of evaluation should be as close as possible to the physiological perception of the human ear may be selected for the comparison. A dimension which is particularly preferred because of its analytical simplicity is the quadratic deviation:

$$E = \sum_{n=o}^{L-1} (y_n - s_n)^2 \qquad (2)$$

where L denotes the length of the speech sample.

As a result of this comparison, the parameter computer determines the changes necessary in the output of the analyser vocal tract model **7** in such a manner that on the next comparison the deviation in accordance with equation (2) between the synthetic signal y_n and the original speech signal s_n is smaller.

For this purpose, the parameter computer **9** determines the gradient of the error dimension in respect of the parameters of the analyser vocal tract model **7.** The parameters of the analyser vocal tract model **7** represent that group of all the components of said model on which the said changes are carried out, i.e., the variable components. Non-variable components, i.e., for example, fixed electrical connection are unchanged and are accordingly disregarded in determining the gradient of the error dimension. The gradient is a vector pointing in the direction of the steepest increase of the error and its absolute value indicates the local slope in that direction. Calculation of the gradient is explained in detail hereinafter with reference to FIGS. **3***a* and **3***b*.

After the gradient has been calculated, the new parameters for the analyser vocal tract model **7** are so determined as to give a small step in the opposite direction to the direction of the gradient. The error naturally decreases most in that direction. If p_k is the vector of all the parameters of the analyser vocal tract model **7** with respect to the k^{th} iteration, then on the next iteration the parameters are defined in accordance with the following formula:

$$P_{k+1} = P_{-k} - \Delta_k \cdot \text{grad}_k (E) \qquad (3)$$

Δ_k represents a small positive step width, which is either fixed or redefined each time.

In the iteration method according to equation (3), the error decreases at each step. As soon as the comparator **8** determines that the error has dropped below a predetermined threshold, i.e., has become acceptable, it delivers a command signal B to the coder **10** to accept the current parameters p_j of the analyser vocal tract model **7** and, together with the information of the pitch detector **4,** i.e., voiced/voiceless signal g and possibly pitch period signals M, to prepare for binary transmission or storage. From that moment on the analyser is ready to analyse the next speech sample.

Referring to FIG. **2***a*, which is a block schematic of the analyser vocal tract model **7** for the order $N = 8$, the vocal tract model includes a storage unit **21** having eight storage locations, a feedback matrix **22,** a stage **23** comprising **8** first multipliers, a stage **24** having **8** second multipliers, a multiplier **25,** a stage **26** having **8** adder networks and a summing network **27.** The feedback matrix **22** is constructed from adder networks and multipliers.

An additional storage unit (not shown) is allocated to each of the stages **23** and **24,** the multiplier **25** and the feedback matrix **22** and stores in each case the current parameters of said stages, i.e., their variable components b_i, c_i, d and a_{ij}, which together form the parameter set p_j (FIG. **1**). The parameters p_j stored in this way can readily be read out of the vocal tract model **7** and fed into the coder **10** by the command signal B from comparator **8** (FIG. **1**).

As already stated, the vocal tract model is a linear digital filter which obeys the recursive vector equations (**1***a*) and (**1***b*).

$$u_{n+1} = A \cdot u_n + b \cdot x_n \qquad (1a)$$

$$y_n = c^T \cdot u_n + d \cdot x_n \qquad (1b)$$

written in component form, equations (**1**a) and (**1**b) read as follows:

$$u_{n+1}^{(i)} = \sum_{j=1}^{N} A_{ij} u_n^{(j)} + b_i x_n \text{ for all values of } i \text{ where } 1 \leq i \leq N \qquad (1a')$$

$$y_n = \sum_{i=1}^{N} c_i u_n^{(i)} + d \cdot x_n$$

The content of the **8** storage locations of the storage unit **21** forms the state vector u_n of the model on the n^{th} cycle. **8** linear combinations are formed from these 8 storage values u_1 to u_8 by means of the feedback matrix **22.** This corresponds in each case to the first term of the right-hand side of equation (1*a*) or (1*a'*). The n^{th} sample of the excitation sequence x_n multiplied by a component of the input vector b is added to each of these linear combinations $A_{11} \ldots A_{18}$ to $A_{81} \ldots A_{88}$ in the adder stage **26.** Multiplication of the sampled values of the excitation sequence x_n by the components b_1 to b_8 of the input vector b is effected by means of the first multipliers of stage **23.** Addition of the linear combinations $A_{11} \ldots A_{18}$ to $A_{81} \ldots A_{88}$ to the product of the sampled value of the sequence x_n and the component of the input vector b is in each case equivalent to the second term of the right-hand side of equation (1*a*) or (1*a'*).

The sums resulting from the said addition form the new storage values which are accepted on the next, i.e., the $(n+1)^{th}$ cycle in the state storage unit **21.**

The n^{th} answer sample y_n is calculated as a linear combination of the storage values in the storage unit **21.** The coefficients used form the output vector c, by whose components c_1 to c_8 the output signals of the individual storage locations of the storage unit **21** are multiplied by means of the second multipliers of stage **24.** The linear combination of the output signals of the second multipliers of stage **24,** which also includes the input signal x_n multiplied by the transit coefficient d in the multiplier stage **25,** is effected in the summing network **27.**

The components of matrix A and of vectors b and c and possibly scalar quantity d can be divided up into three groups. The components of the first group are predetermined. They usually have simple values such as 0, i.e., the corresponding connection does not exist at all, or 1, i.e., the corresponding signal is included in the linear combination purely additively without additional multiplication, or −1, i.e., pure subtraction. The componants of this group are therefore not influenced by the optimization process. The second group comprises those components which are changed on each

optimization step. Finally, the components of the third group are linear combinations of variable and invariable partial components. For example, the matrix A may have a component of the form $A_{ij} = 1+p_k$. In this case, p_k would be changed on each optimization step and 1 would denote fixed wiring. The signal path which couples the i^{th} component of the n^{th} state vector u_n back to the j^{th} component of u_{n+1} would therefore consist of a fixed path and a variable path.

The fixed components, i.e., those of the first group and the fixed parts of the third group, determine the structure of the vocal tract model. The variable components, i.e., those of the second group and the variable parts of the third group, form the vocal tract model parameters p_j (FIG. 1) which are to be transmitted via channel **14.**

FIG. 2*b* shows the vocal tract model of FIG. 2*a* in simplified form, the individual stages of the circuit being denoted only by the corresponding signal or signal components.

FIGS. 3*a* and 3*b* are each a block schematic of the parameter computer **9** (FIG. 1).

As already stated, the parameter computer **9** has to compute a set of new parameters p_{k+1} in accordance with formula (3) on each optimization step:

$$P_{k+1} = P_k - \Delta_k \cdot \text{grad}_k (E) \tag{3}$$

where p_k is the vector of the old parameters, Δ_k is a small positive step width. This can be selected to be identical on each step, i.e., $\Delta_k = \Delta$ for all values of k, or alternatively it can be re-defined for each optimization step.

The article by L. S. Willimann: "Computation of the Response-Error Gradient of Linear Discrete Filters," IEEE Transactions, Vol. ASSP-22, No. 1, Feb. 1974," also shows that the computation of $\text{grad}_k (E)$ falls into two stages. The first stage is very simple and mathematically elementary and depends only on the nature of the error dimension E, and not on the choice of the structure of the vocal tract model. The second stage depends only on the structure of the vocal tract model but not on the error dimension.

The publication by L. S. Willimann also shows, by means of a duality theorem, that the parameter computer **9** can simultaneously carry out the function of the filter and hence of the vocal tract model **7** (FIG. **1**).

FIG. 3*a* shows a first version of a combined parameter computer **9** and vocal tract model **7** according to FIGS. 2*a* and 2*b* respectively, the order N again being equal to **8.**

Referring to FIG. 3*a*, the parameter computer **9** includes a first primary model **29,** a unit **30,** and N=8 additional primary part-models 31 to 38. The first primary model 29 is identical to the vocal tract model shown in FIG. 2*b* as will be apparent from comparison of FIGS. 2*b* and 3*a*.

The first primary model **29** is excited by the pulse/noise generator **6** (FIG. **1**) and in addition to the synthetic speech signal y_n yields the partial derivatives $\delta y_n/\delta c_1 \ldots \delta y_n/\delta c_8$ and $\delta y_n/\delta d$. The derivative $\delta y_n/\delta c_i$ is precisely equal to the i^{th} component of the state vector u (equation 1*a*). The mathematical grounds for this and the following relationships are given in the aforementioned article. The derivative (sensitivity) $\delta y_n/\delta d$ of the model output y_n is also equal to the corresponding term of the excitation sequence x_n in respect of the transit coefficient d.

The unit **30** which is also excited by the pulse/noise generator **6** (FIG. **1**) is a part of the model which is a dual model with respect to the first primary model **29** and hence to the vocal tract model **7,** for it can be shown that there is an equation system (4*a*) and (4*b*) which is equivalent to the equations (1*a*) and (1*b*) and which, for an identical excitation sequence x_n, yields the same answer sequence y_n as the primary model:

$$v_{n+1} = A^T \cdot v_n + c \cdot x_n \tag{4a}$$

$$y_n = b^T \cdot v_n + d \cdot x_n \tag{4b}$$

The feedback matrix of the dual model is the transpose A^T of the feedback matrix A of the primary model. The primary output vector c becomes the input vector in the dual model and the primary input vector b becomes the output vector. The transit coefficient d is the same in both models.

Unit **30** represents the equation (4*a*). The components of the state vector v of this dual model are the partial derivatives $\delta y_n/\delta b_1 \ldots \delta y_n/\delta b_8$ of the current term y_n of the output sequence with respect to the components of the input vector $b_1 \ldots b_8$.

The components of the state vector v of the dual part-model **30** each excite a primary part-model **31** to **38.** The state vectors $u' \ldots u$ VIII of this primary part-model yield the partial derivatives of the term y_n of the output sequence with respect to the elements A_{ij} of the feedback matrix A in the manner indicated.

A second equivalent arrangement is shown in FIG. 3*b*. Here again the input sequence x_n excites a complete primary model **39** and a dual part-model **40.** Contrary to FIG. 3*a*, however, the components of the state vector u of the primary model are used in this case to excite N=8 additional dual part-models **41** to **48.** The model answer y_n and the required partial derivatives with respect to the model parameters $\delta y_n/\delta A_{ij}$, $\delta y_n/\delta b_i$, $\delta y_n/\delta c_j$ and $\delta y_n/\delta d$ are as shown in the Figure.

As shown in FIG. **4,** the partial derivatives $\delta y_n/\delta d$, $\delta y_n/\delta c_i$, $\delta y_n/\delta b_i$, and $\delta y_n/\delta A_{ij}$ obtainable at the output of the parameter computer **9** are fed to a computer stage **49** in which they are subjected to a computing operation dependent upon the selected error dimension E. The partial derivatives $\delta E/\delta d$, $\delta E/\delta c_i$, $\delta E/\delta b_i$ and $\delta E/\delta A_{ij}$ changed in this way are fed back from the output of the computer stage **49** as shown in FIGS. 3*a*, 3*b* and **4** to the corresponding multipliers d, c_i, b_i and A_{ij} of the parameter computer **9** and hence also of the vocal tract model **7** and change their coefficient on each optimization step in dependence on the deviation between the sequences s_n and y_n as determined in the comparator **8** (FIG. **1**).

If the error dimension selected is the quadratic deviation according to formula (2), and if the partial derivatives at the output of the parameter computer **9** are designated $\delta y_n/\delta P_j$, then the following formula applies to the computing operation in stage **49**:

$$\frac{\delta E}{\delta P_j} = 2 \cdot \sum_{n=0}^{L-1} (y_n - s_n); \quad \frac{\delta y_n}{\delta P_j} \qquad (5)$$

In this connection reference should be made to the parameter definition given hereinbefore. The parameters of course represent only a part of all the components d, c_i, b_i and A_{ij} of the parameter computer. It is self-evident that in the optimization process only those components are changed which really represent parameters. Consequently, only those partial derivatives which are associated with real parameters need be fed to stage **49** and the parameter computer **9**. In practice this means that **15** parameters are sufficient, given a suitable model structure, instead of the possible **81** model parameters (one parameter d + 8 parameters c_i + 8 parameters b_i + 8×8 parameters A_{ij}).

It should be repeated that the parameter computer contains a complete vocal tract model, as shown in FIGS. **3***a* and **3***b*. In the practical construction of the analyser and synthesizer described, the vocal tract model **7** is contained in the parameter computer **9** (FIG. **1**). The separate representation of the two elements in FIG. **1** has been given solely to simplify the description.

SYNTHESIZER

The decoder **11** (FIG. **1**) samples its input signal to obtain the appropriate signals from which it is built up, i.e., it obtains the model parameters p_j, the voiced or voiceless information signal g, and, if present, the pitch period information M, from the channel signal or the stored digital signals.

The pulse/noise generator **6'** is excited by the voiced/voiceless information and the length of the pitch period and this generator is identical to the pulse/noise generator **6** of the analyser. The pulse/noise generator **6'** delivers the excitation sequence for the synthesizer vocal tract model **7'**, which is identical to the analyser vocal tract model **7**. Since the model **7'** has the same structure as the model **7**, being adjusted on the basis of the same parameters and also excited by the same excitation sequence x_n, it yields the same answer sequence y_n. As a result of the optimization algorithm used in the analyser, this answer sequence y_n deviates only insignificantly, i.e., barely perceptibly to the ear, from the original sampled speech signal s_n.

The output sequence y_n of the synthesizer vocal tract model **7'** is converted in the digital-to-analog converter **12** into an analog signal which is demodulated in the following low-pass filter **2'**. The demodulation filter **2'** is of the same design as the analyser input filter **2**. The speech signal synthesized in this way is fed to the unit **13**, which is generally a loudspeaker or an analog store.

The essential elements of the synthesizer, i.e., the pulse/noise generator **6'**, the vocal tract model **7'** and the filter **2'** are thus contained in identical form in the analyser. Since analog/digital converters of conventional construction usually have a digital/analog converter in their feedback circuit, the digital/analog converter **12** is also already present in the analyser. These circumstances enable the apparatus to be used very easily in half-duplex operation.

Practical tests have shown that the variables requiring to be transmitted or stored, i.e., voiced/voiceless information, pitch period and model parameters, have to be re-defined about 30 times per second to obtain an acceptable synthetic speech quality. It has also been found that a model order of $N = 8$ is sufficient with a sampling frequency of 6 kHz. Also, given a suitable model structure, 15 model parameters per 8 bits are sufficient. Bearing in mind that the voice/voiceless information requires 1 bit and taking the pitch period as 10 bits, a transmission rate of 30 (15.8+10+1) bits/sec = 4000 bits/sec is obtained.

In comparison with conventional PCM transmission, the channel capacity required is thus reduced by about 90%. The transmission rate can probably be reduced still further by a suitable choice of the structure of the vocal tract model.

What is claimed is:

1. In a method of analysing and synthesizing speech in which speech synthesis is effected by use of a synthesis vocal tract model which functionally corresponds to the human vocal tract and which is constructed essentially from a discrete linear filter, comprising:

A. an analysis operation including the steps of

1. sampling an original speech signal and deriving therefrom

1a. a first group of signals representing parameters of said synthesis vocal tract model;

1b. a second group of signals representing the fundamental frequency reciprocal (hereinafter referred to as the "pitch period"); and

1c. a third group of signals representing the voiced/voiceless character of each sample of the original speech signal; and

B. A synthesis operation including the steps of

1. adjusting said synthesis vocal tract model by reference to said first group of signals, and

2a. during voiced samples of the original speech signal, exciting said synthesis vocal tract model by a train of pitch period spaced pulses,

2b. during voiceless samples of the original speech signal, exciting said synthesis vocal tract model by white noise,

whereby a synthetic speech signal similar to said original speech signal is produced,

The improvement wherein:

C. said analysis operation (A) is performed by use of an analysis vocal tract model identical to said synthesis vocal tract model;

D1. during voiced samples of the original speech signal said analysis vocal tract model is excited by a train of pitch period spaced pulses;

D2. during voiceless samples of the original speech signal said analysis vocal tract model is excited by white noise;

E. an output signal from the analysis vocal tract model is sampled and compared with the original speech signal;

F. The parameters of the analysis vocal tract model are modified as a result of step (E) to minimize the deviation between the output signal and the original speech signal; and

G. those parameters of said analysis vocal tract model for which the deviation falls below a predetermined threshold value are used as said first group of signals in step (A1*a*).

2. A method as claimed in claim **1**, in which step (F) includes determining the gradient of the error dimension representing the deviation with respect to the parameters of the analysis vocal tract model and modifying the parameters of the analysis vocal tract model in

the opposite direction to the direction of the gradient.

3. A method as claimed in claim 2, in which following each determination of the error dimension representing the deviation between the original speech signal and the output signal of the analysis vocal tract model the parameters of the analysis vocal tract model are modified in a small step.

4. A method according to claim 3, in which the width of the step in the change of the parameters of the analysis vocal tract model is selected to have a fixed value.

5. A method as claimed in claim 1, in which in performing steps (D1) and (D2) said pulse train and said white noise, respectively, have a power which is substantially constant and substantially the same.

6. A method as claimed in claim 5, in which in performing step (D1) said pulse train comprises unit pulses.

7. Apparatus for performing speech analysis and synthesis comprising a synthesizer which includes a synthesis vocal tract model which functionally corresponds to the human vocal tract, and generator selectively operable to provide pulses or white noise; and an analyser including first means for determining parameters of said synthesis vocal tract model, second means for determining the pitch period of an original speech signal, and third means for determining the voiced/-voiceless character of the original speech signal; wherein

said first means comprises an analysis vocal tract model identical to the synthesis vocal tract model; a generator selectively operable to provide pulses or white noise identical to said generator of the synthesizer; a sample storage unit for storing samples of the original speech signal; a comparator for comparing an output signal of said analysis vocal tract model with the signal stored in the sample storage unit; and a parameter computer for minimizing the deviation between the two signals compared by said comparator.

8. Apparatus as claimed in claim 7, wherein said analysis vocal tract model and said synthesis vocal tract model are each comprised by a linear digital filter.

9. Apparatus as claimed in claim 8, wherein the parameter computer is adapted to be excited by the signal from said synthesizer generator and to provide an output signal corresponding to the gradient of the error dimension representing the deviation determined by said comparator.

10. Apparatus according to claim 9, wherein said parameter computer and said analyser vocal tract model are constituted by parts of a common unit which comprises:

a primary model identical to said vocal tract model, a part of a model which is a dual model with respect to said primary model, and a number of additional part-models of the primary model corresponding to the number of components of the state vector of the primary model and of the dual part-model respectively;

and wherein an input of said primary model and an input of said dual part-model are connected to the output of said synthesizer generator, and each of the additional primary part-models is connected by its input to each of those outputs of the dual part-model which yield the components of the state-vector of this dual part-model.

11. Apparatus as claimed in claim 9, wherein said parameter computer and said analyser vocal tract model are constituted by parts of a common unit which comprises:

a primary model identical to said vocal tract model, a part of a model which is a first dual model with respect to said primary model, and a number of additional dual part-models corresponding to the number of components of the state vector of the primary model and of the dual part-model respectively;

and wherein an input of said primary model and an input of said first dual part-model are connected to the output of said synthesizer generator and each of the other dual part-models is connected by its input to one of those outputs of the primary model which yield the components of the state vector of that primary part-model.

* * * * *