United States Patent

Manley et al.

[15] **3,681,530**

[45] **Aug. 1, 1972**

[54] **METHOD AND APPARATUS FOR SIGNAL BANDWIDTH COMPRESSION UTILIZING THE FOURIER TRANSFORM OF THE LOGARITHM OF THE FREQUENCY SPECTRUM MAGNITUDE**

[72] Inventors: **Harold J. Manley**, Sudbury; **Harry L. Shaffer**, Lynnfield, both of Mass.

[73] Assignee: **GTE Sylvania Incorporated**

[56]   **References Cited**

UNITED STATES PATENTS

3,448,216	6/1969	Kelly	179/1 SA
3,566,035	2/1971	Noll	179/1 SA
3,344,349	9/1967	Schroeder	179/1 SA
3,403,227	9/1968	Malm	179/1 SA
3,330,910	7/1967	Flanagan	179/1 SA
3,493,684	2/1970	Kelly	179/1 SA
3,471,648	10/1969	Miller	179/1 SA

OTHER PUBLICATIONS

Noll, Short–Time Spectrum and Cepstrum Techniques for Vocal Pitch Detection, J.A.S.A. 2/1964 p. 296–302.
Shively, A Digital Processor to Generate Spectra in Real Time, IEEE Trans. on Computers, 5/1968 p. 485–491.

Primary Examiner—Kathleen H. Claffy
Assistant Examiner—Jon Bradford Leaheey

Attorney—Norman J. O'Malley, Elmer J. Nealon and Robert T. Orner

[57]   **ABSTRACT**

A bandwidth compression system such as a digital vocoder including an analysis section employs a transducer to convert an input speech wave into an electrical signal which is then digitized by an analog to digital converter. The digitized signal is directed through a spectrum device where the magnitudes of the frequency spectrum of the input speech wave are obtained. These magnitudes are then directed to a logging circuit to obtain the logarithm of the frequency spectrum magnitudes of the input speech signal. The logged magnitudes of the frequency spectrum are then directed to a computer where the discrete Fourier transform of the logged spectrum magnitudes are obtained to form the Fourier transform of the logarithm of the frequency spectrum magnitude (FTLSM) of the input speech signal. An encoding unit selects and encodes certain ones of the FTLSM coefficients for transmission to a remote terminal for analysis. The encoded signals include pitch data and vocal tract impulse data, both of which are derived from the FTLSM signals.

The analysis section of a vocoder terminal employs a decoding device which decodes the received data and separates it into pitch data and vocal tract impulse data. Connected to the decoding device is a computing device for computing the logarithm of the spectrum envelope of the vocal tract impulse response function using the discrete Fourier transform. The logged spectrum is directed through a delogging device to a fast Fourier transform (FET) computer where the Fourier sine transform of the received spectrum signals (the impulse response) are obtained. A convolution unit then convolves the pitch data with the impulse response data to yield the desired synthesized speech signal.
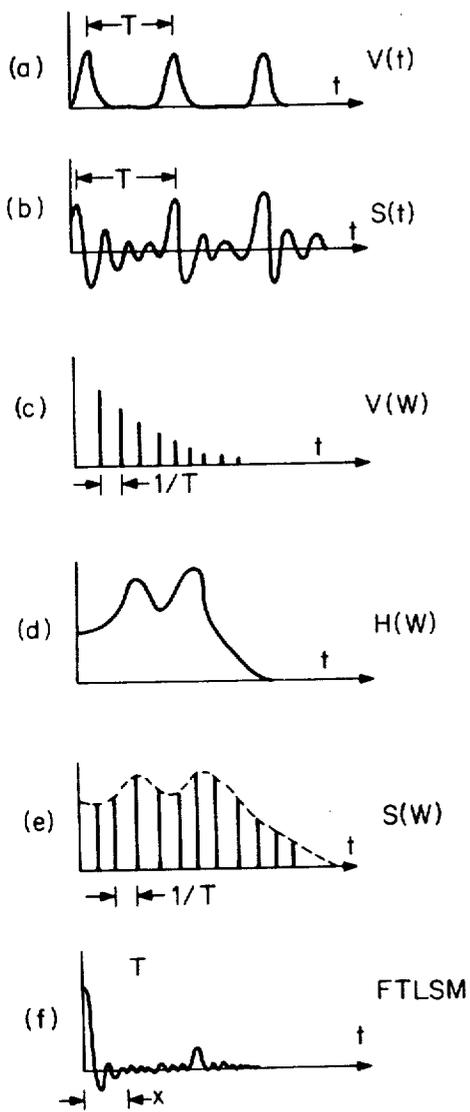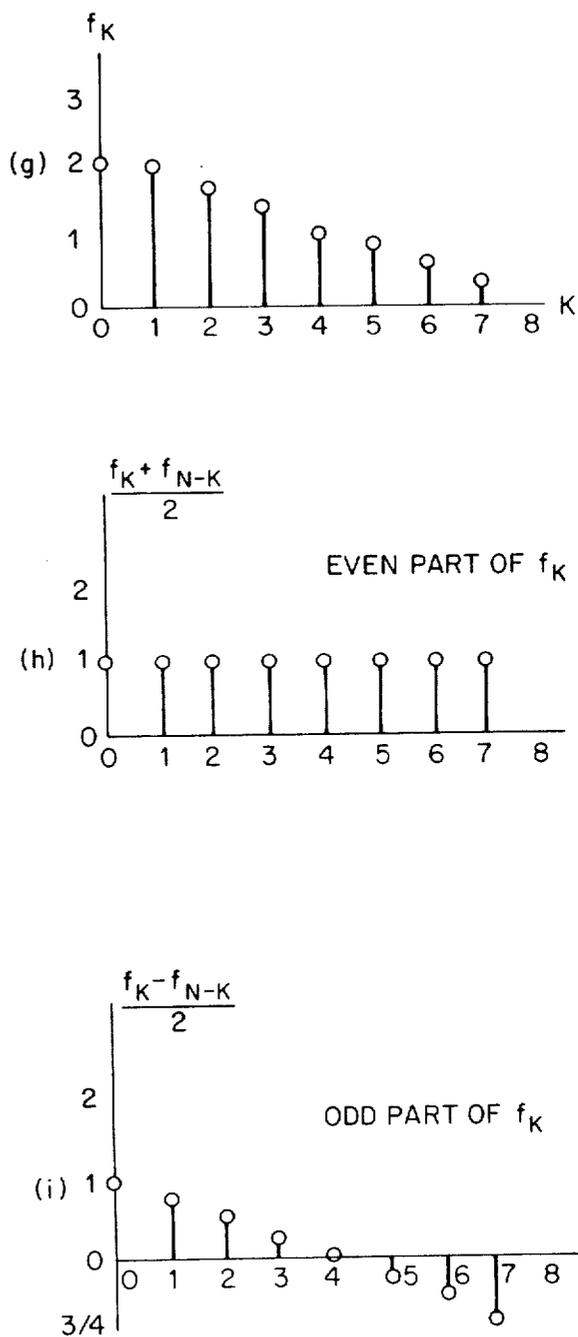
37 Claims, 19 Drawing Figures

Fig. 1A.

Fig. 1B.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER

BY

ATTORNEY

Fig. 2A.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER

BY

ATTORNEY

Fig. 2B.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T. Orner
ATTORNEY

Fig. 3.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T Osmer
ATTORNEY
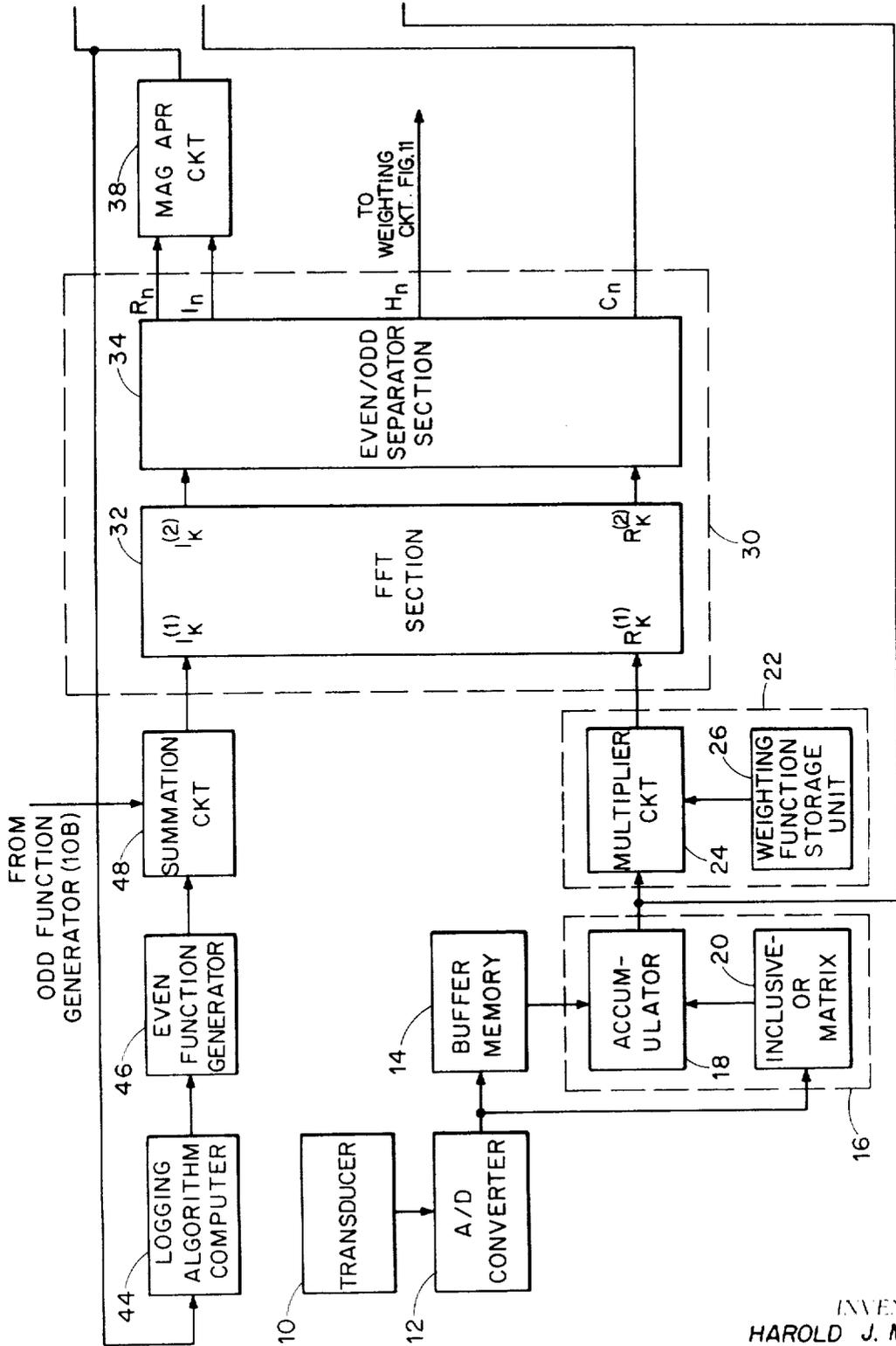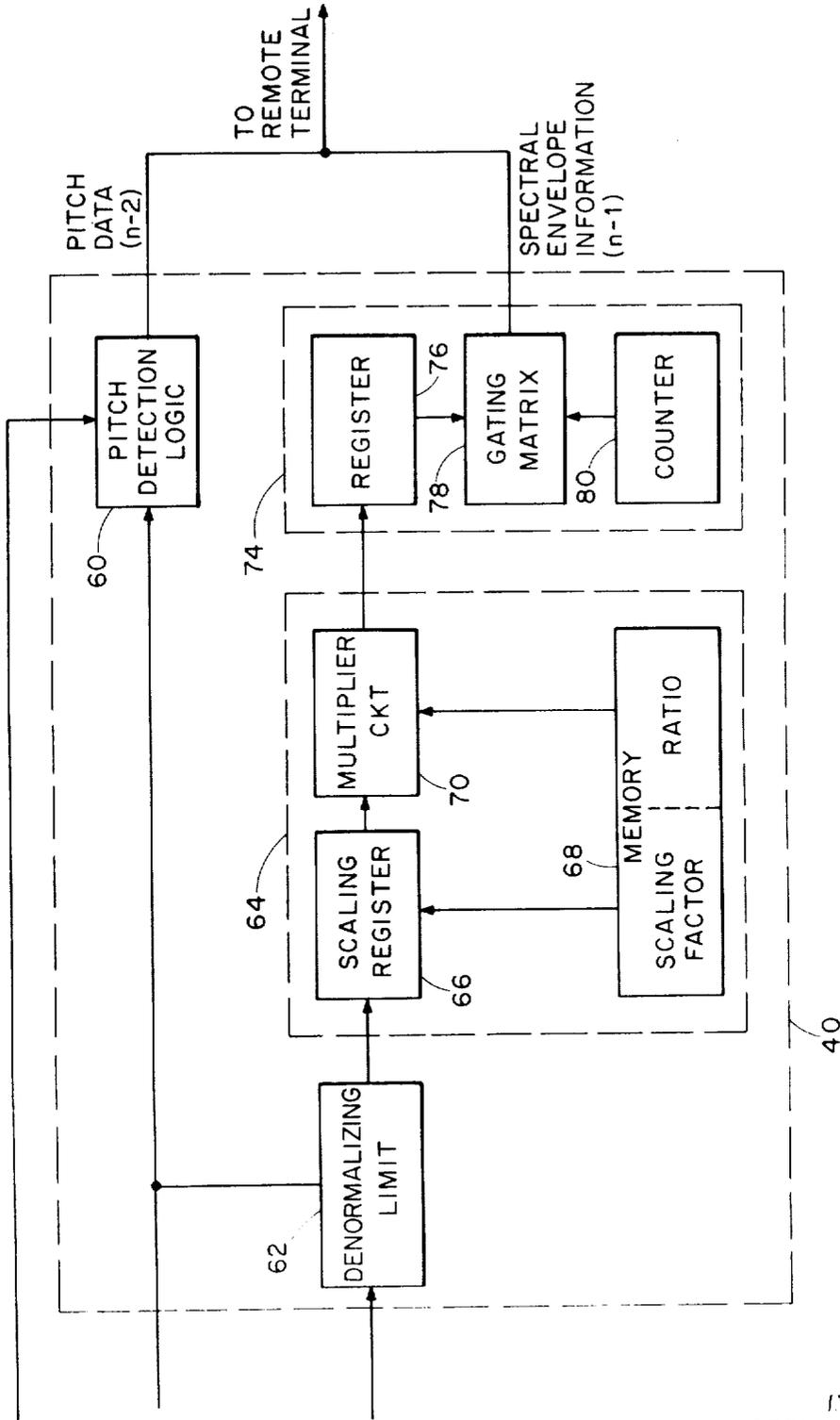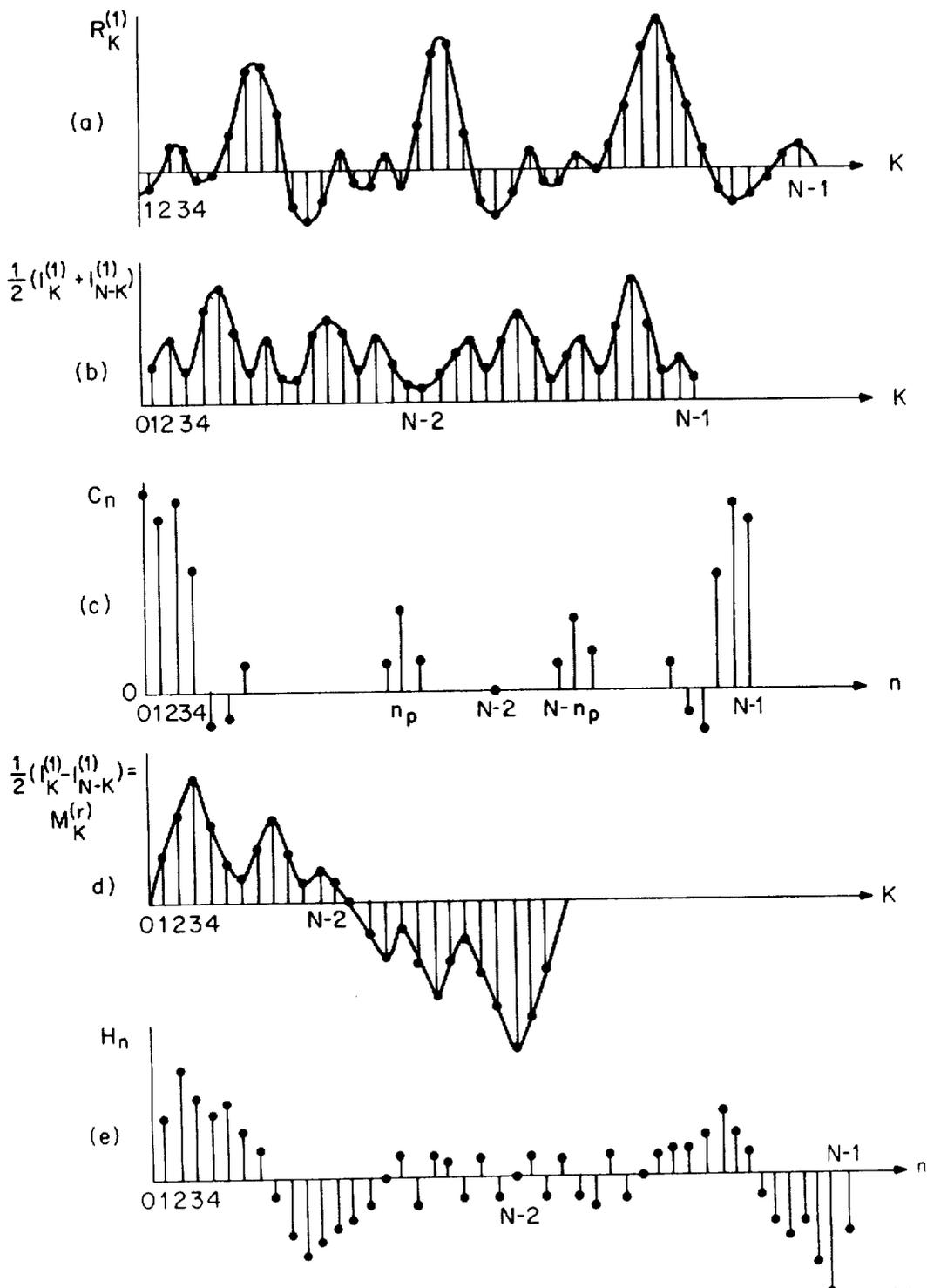
Fig. 4.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER

BY

ATTORNEY

Fig. 5.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T. Orner
ATTORNEY

Fig. 6.

DATA RETAINED FROM (n-1) AND (n-2) INTERVALS

$(E_L)_{n-1}$, $(E_H)_{n-1}$, $(E_H)_{n-2}$, $(E_L)_{n-2}$, AND $(T_1)_{n-2}$

UV SEND q_{n-2}

SET q_{n-1}=0

SET FLAG =0

CONDITION ϕ:
ARE $(E_L)_n$ AND $(E_L)_{n-1}$ AND $(E_L)_{n-2} \geq K_1$
AND $(E_H)_n$ AND $(E_H)_{n-1}$ AND $(E_H)_{n-2} \geq K_2$

NO

YES

TO POINT A
FIG. 7

WAS q_{n-3}=0

YES

NO

FIND $(P_1 \cdot T_1)_{n-1}$ IN SEARCH RANGE

COMPUTE $(E_L)_n$ AND $(E_H)_n$

n^{th} SPECTRUM AND $(n-1)^{st}$ CEPSTRUM

CONDITION ϕ':
IS $(E_L)_n$ OR $(E_L)_{n-1} \geq K_3$
AND $(E_L/E_H)_{n-1}^2 2$

NO

TO POINT B
FIG. 8

YES

IS FLAG =1

NO   TO POINT C
FIG. 9A.

YES

TO POINT D
FIG. 9B.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER

BY

Robert J. Osmer

ATTORNEY

Fig. 7.

FROM POINT A
FIG. 6

IS $(T_1)_{n-1}$ TRACKING WITHIN ±1MSEC OF $(T_1)_{n-2}$

YES

NO

SET FLAG=1

SET FLAG=0

$\forall$ SEND $q_{n-2} = (T_1)_{n-2}$

$\forall$ SEND $q_{n-2} = (T_1)_{n-1}$

SET $q_{n-1} = (T_1)_{n-1}$

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T. Crner
ATTORNEY

Fig. 8.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER

BY

ATTORNEY

Fig. 9A.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER

BY

Robert T. Cmier

ATTORNEY

Fig. 9B.

INVENTORS
HAROLD J. MANLEY
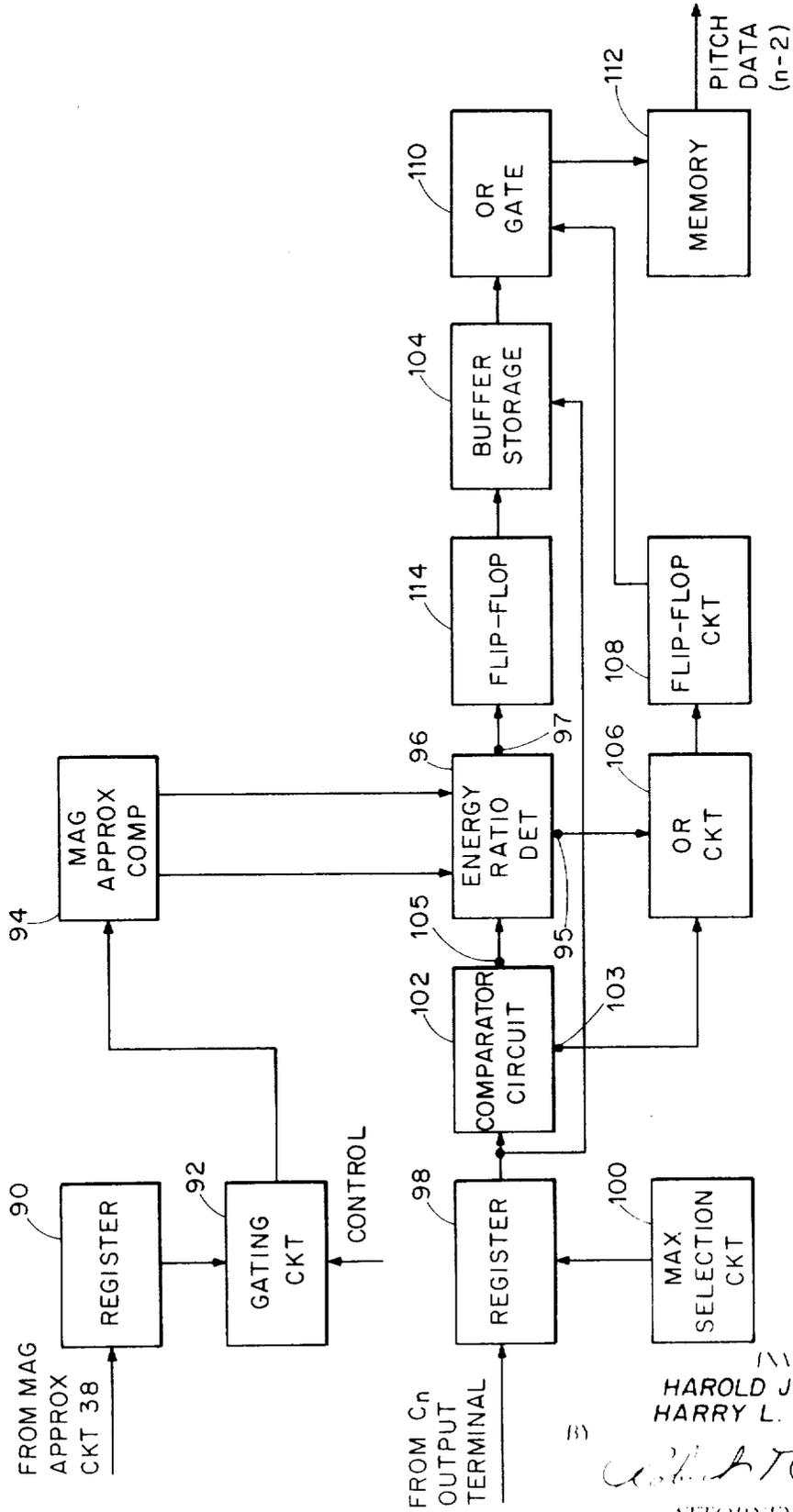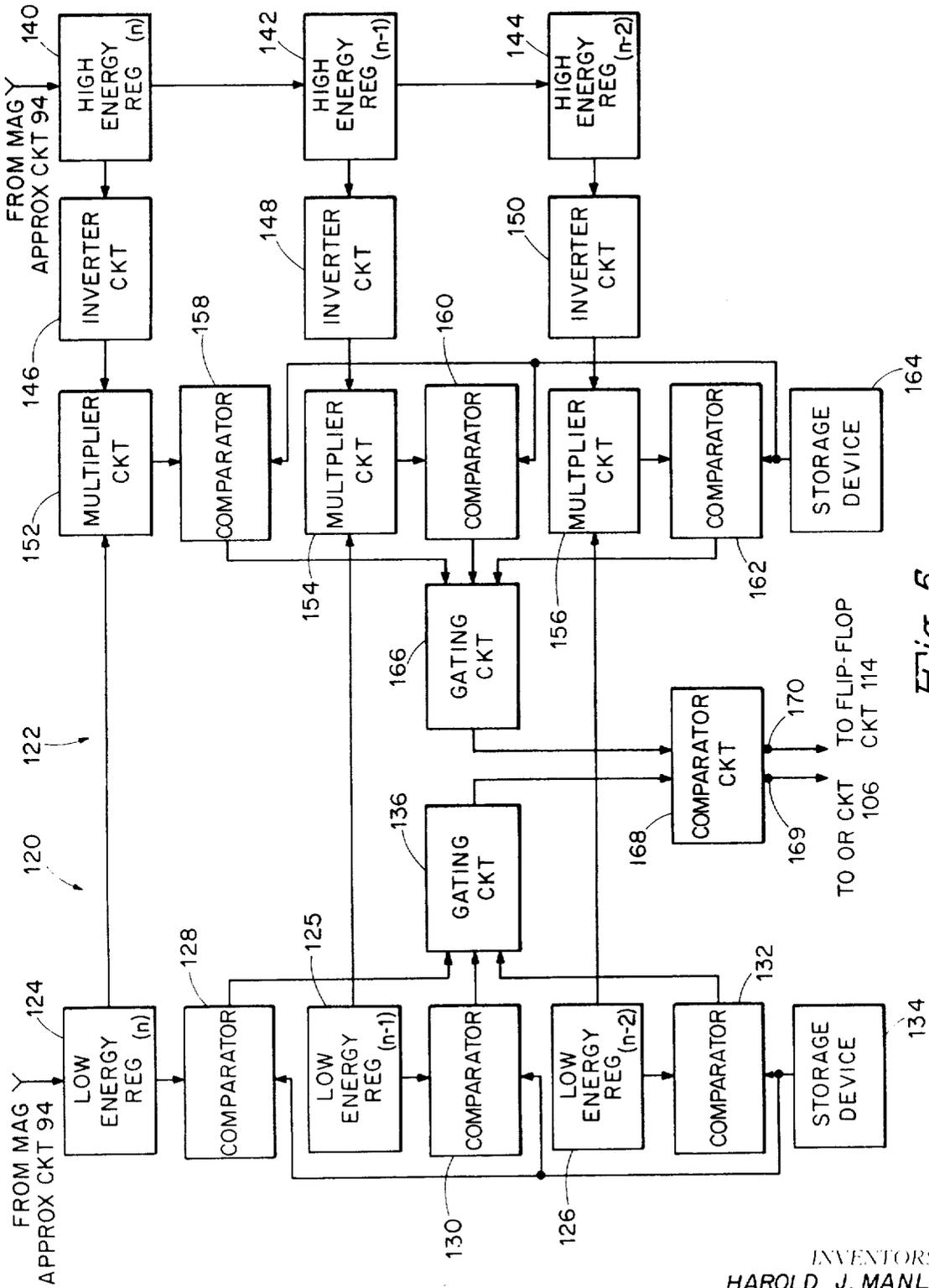HARRY L. SHAFFER
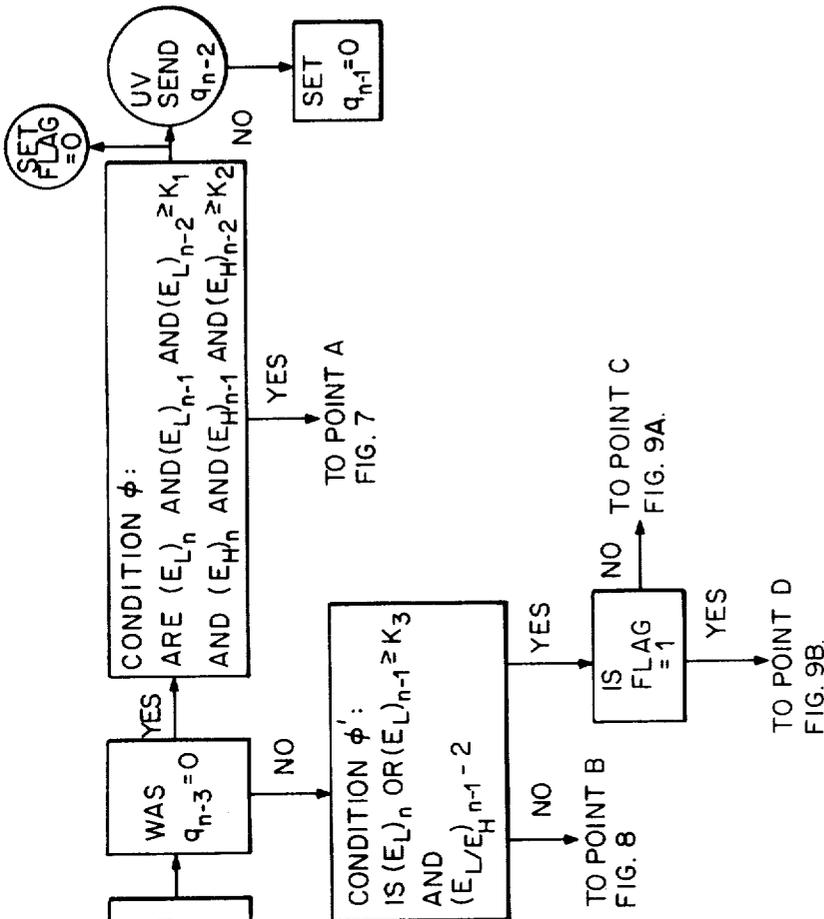BY
Robert T. Orner
ATTORNEY
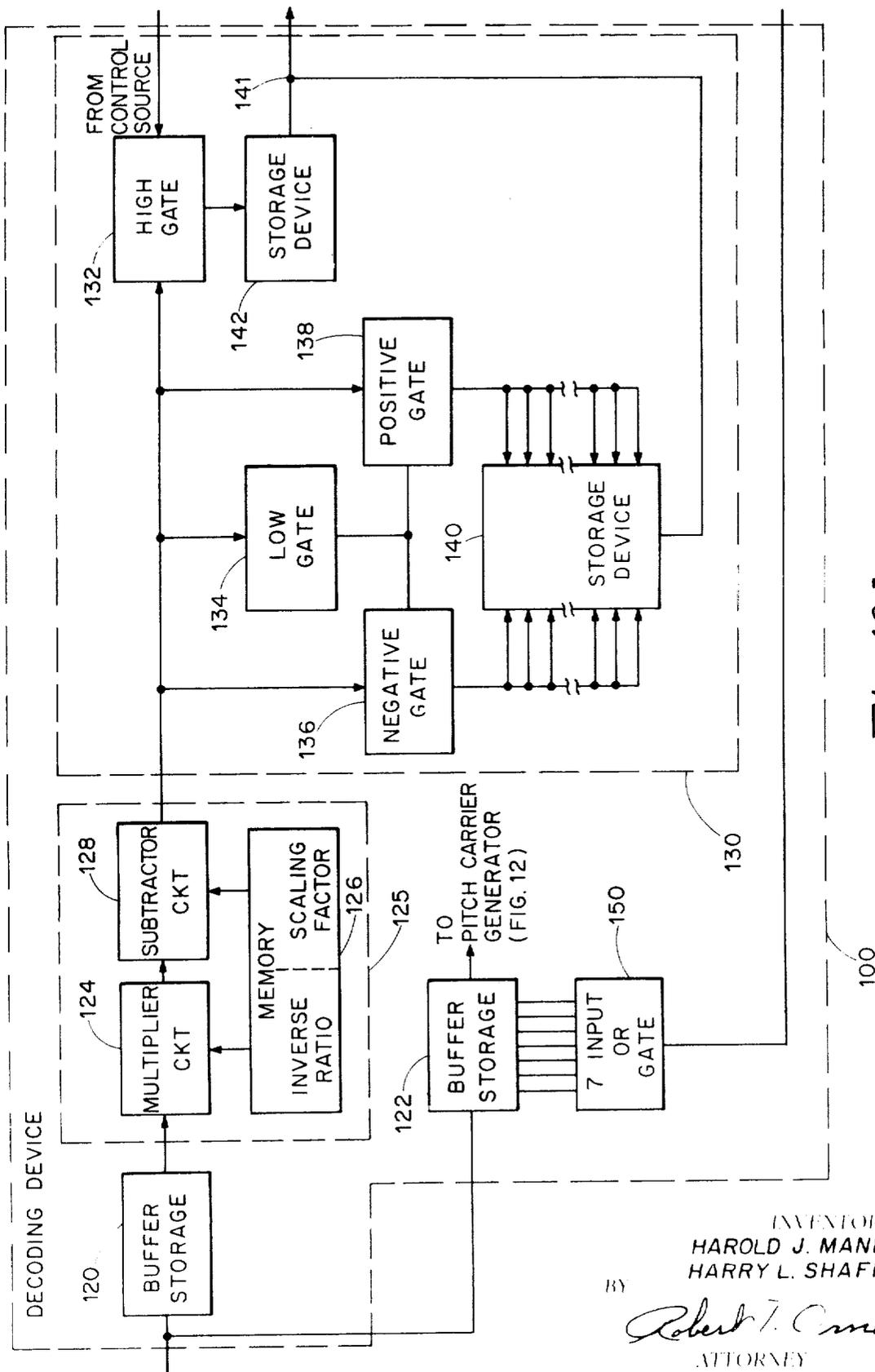
Fig. 10A.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER

BY

Robert T. Orner

ATTORNEY

TO
SUMMATION
CKT 48
(FIG. 2A)

ODD FUNCTION GENERATOR

108

184

255
252
248

NEGATING CKT   186

NEGATING CKT   186

GATE

GATE

GATE

GATE

182

182

SEQUENTIAL COUNTER

180

CONTROL

SPECTRUM DECODER   102

DELOGGING COMPUTER   104

DIGITAL MULTIPLIER   105

AND GATE   107

RANDOM SEQUENCE GENERATOR   109

106
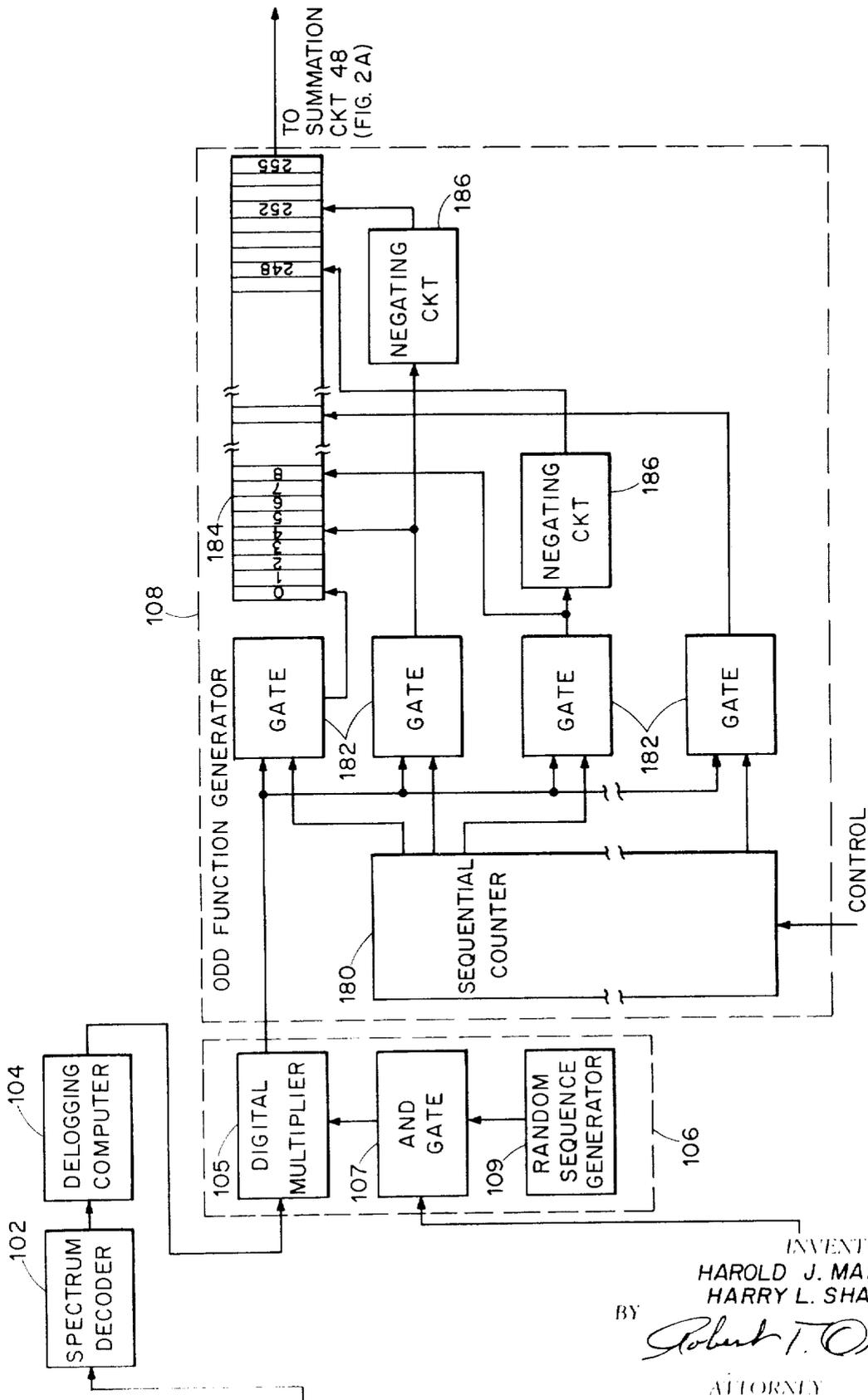
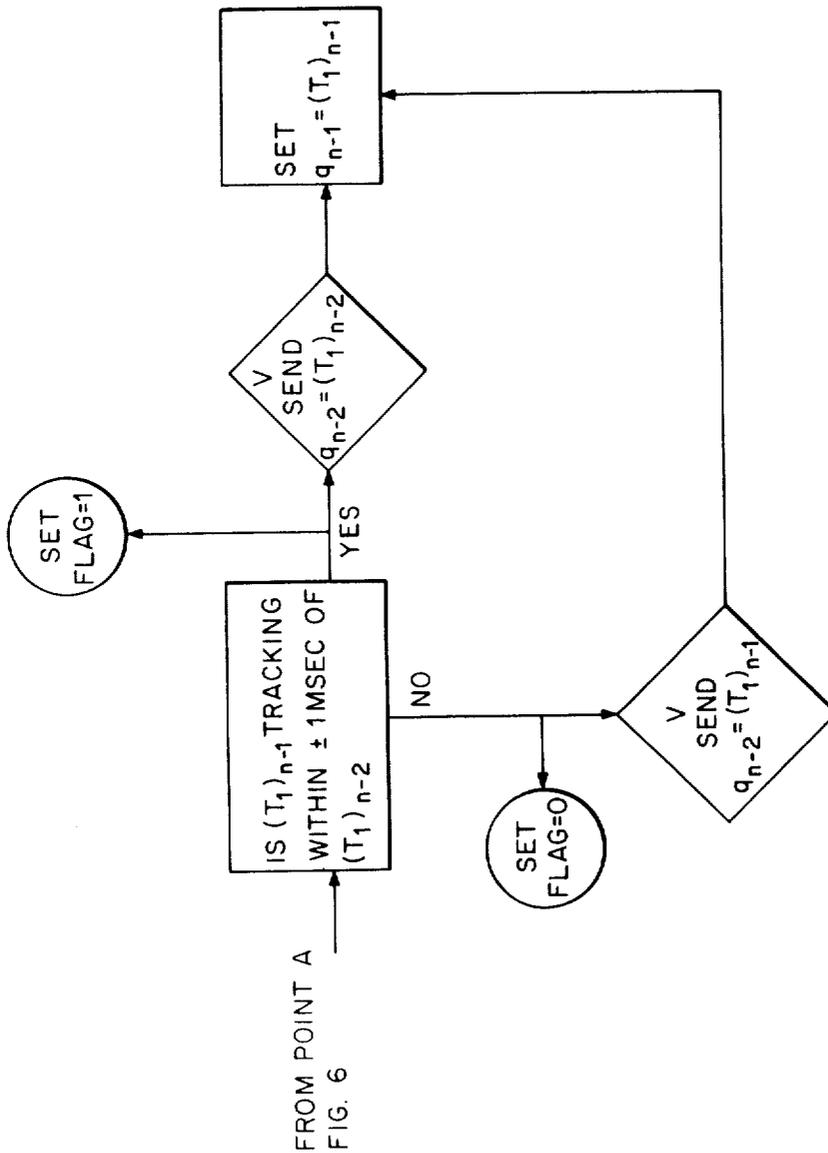Fig. 10B.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T. Orner
ATTORNEY

Fig. 11.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T. Orner
ATTORNEY

Fig. 12.

PITCH CARRIER GENERATOR 229

ENABLE 0 OCCURS EVERY 20 MSEC, T

ENABLE 1 OCCURS AT THE SAMPLING RATE $\frac{1}{T}$

TO AND GATE 320 (FIG.14)

TO STORAGE REGISTER 371 (FIG.14)

FROM BUFFER STORAGE DEVICE 122 (FIG. 10A)
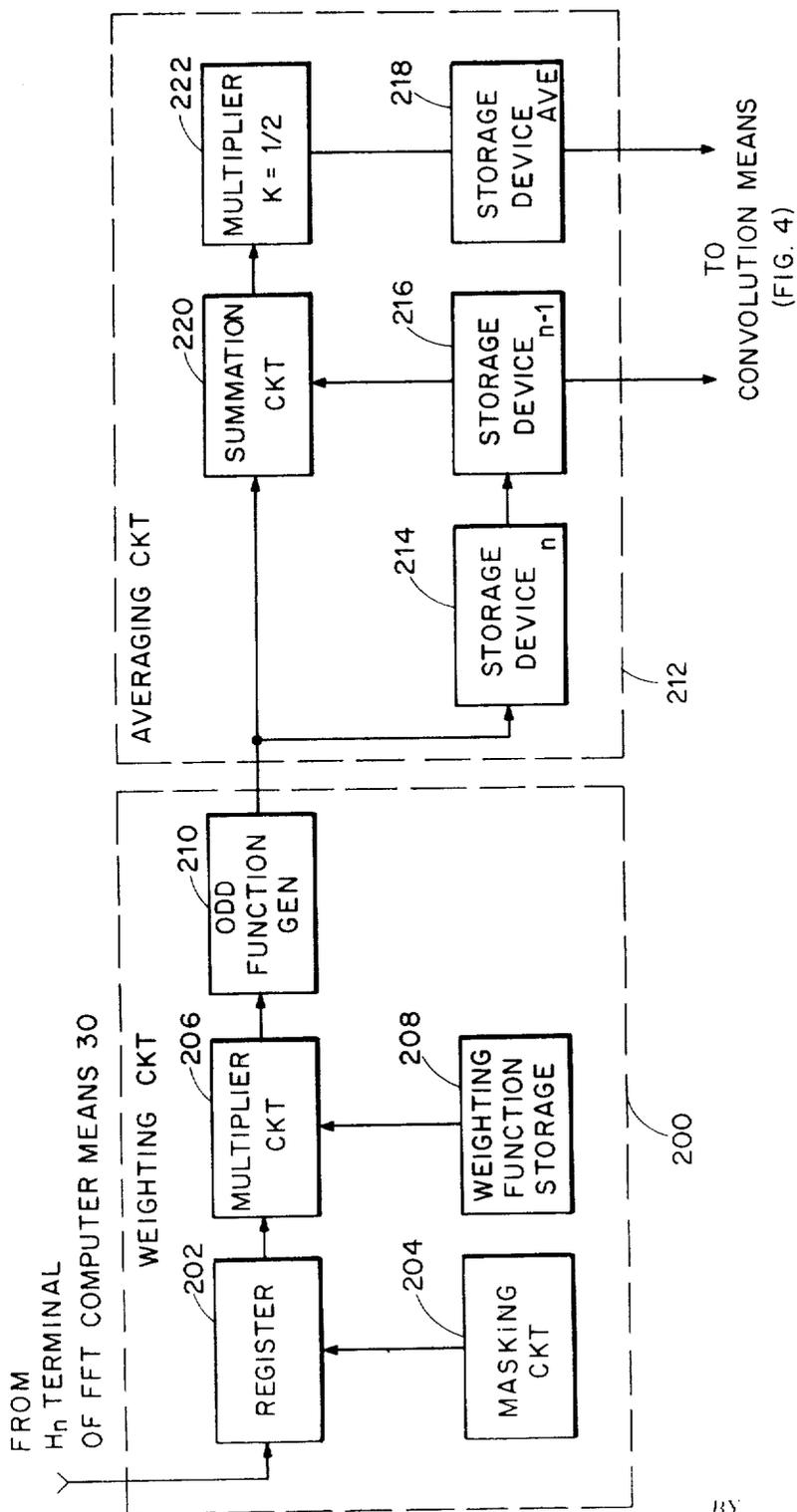
INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T. Ormer
ATTORNEY
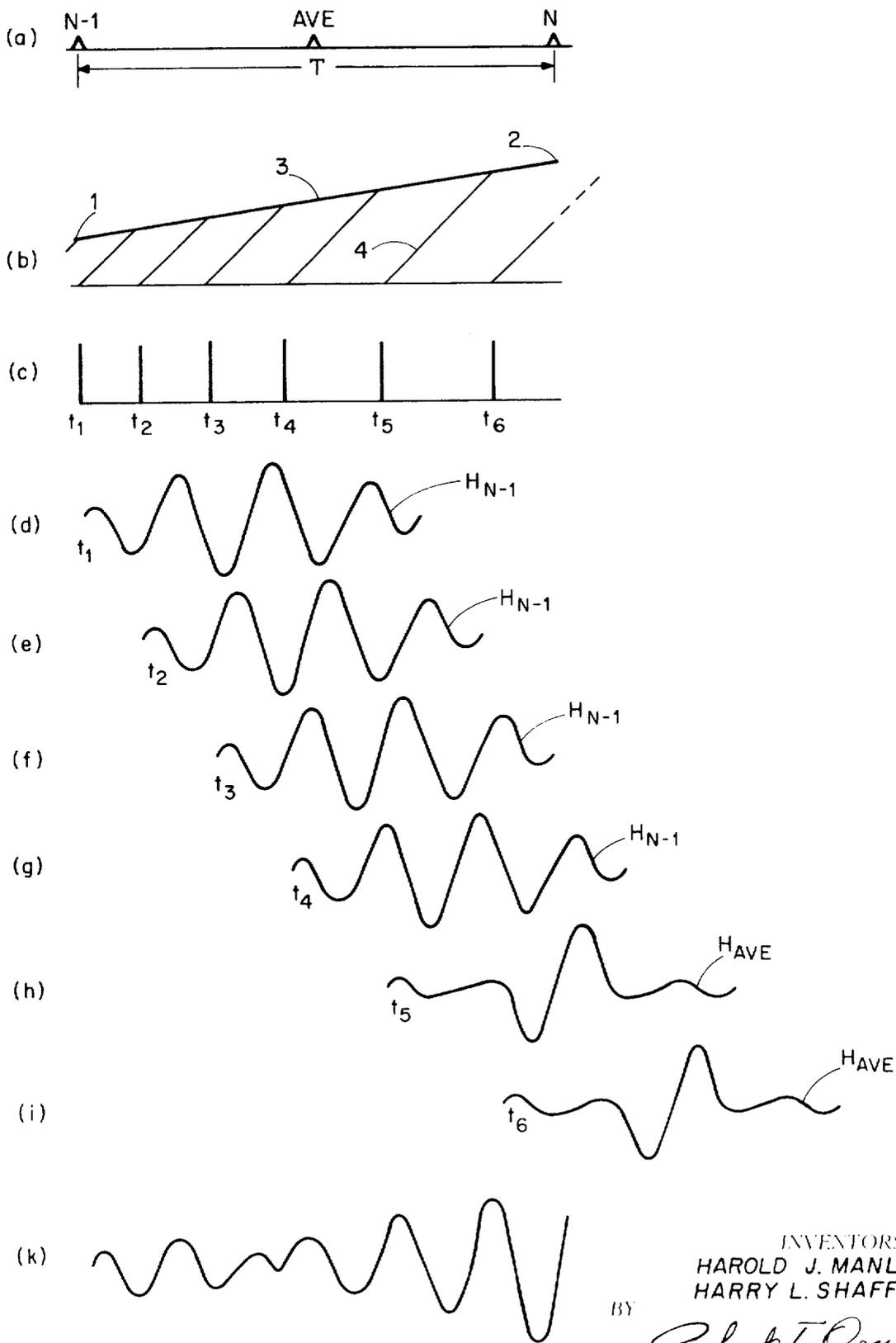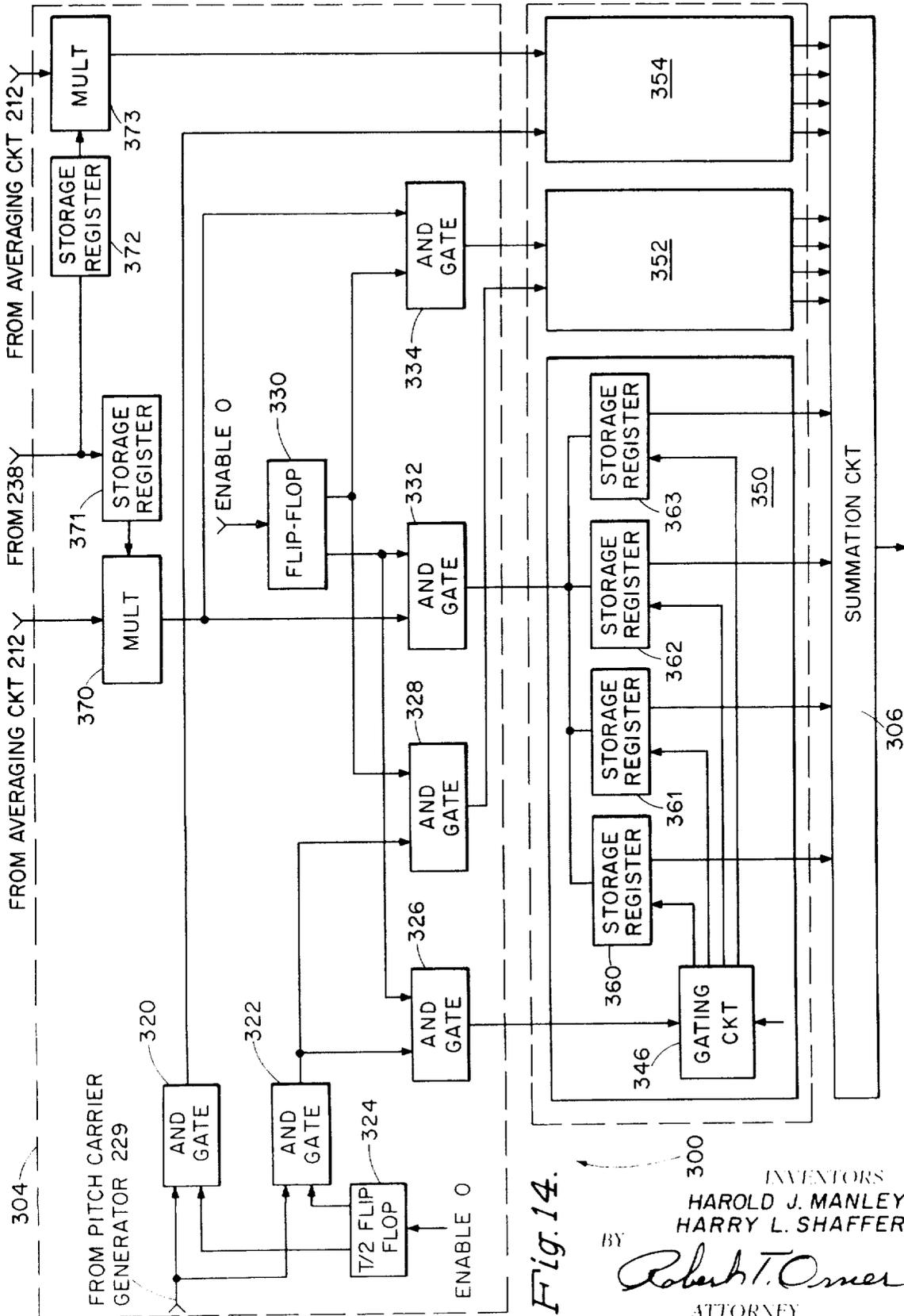
Fig. 13.

INVENTORS
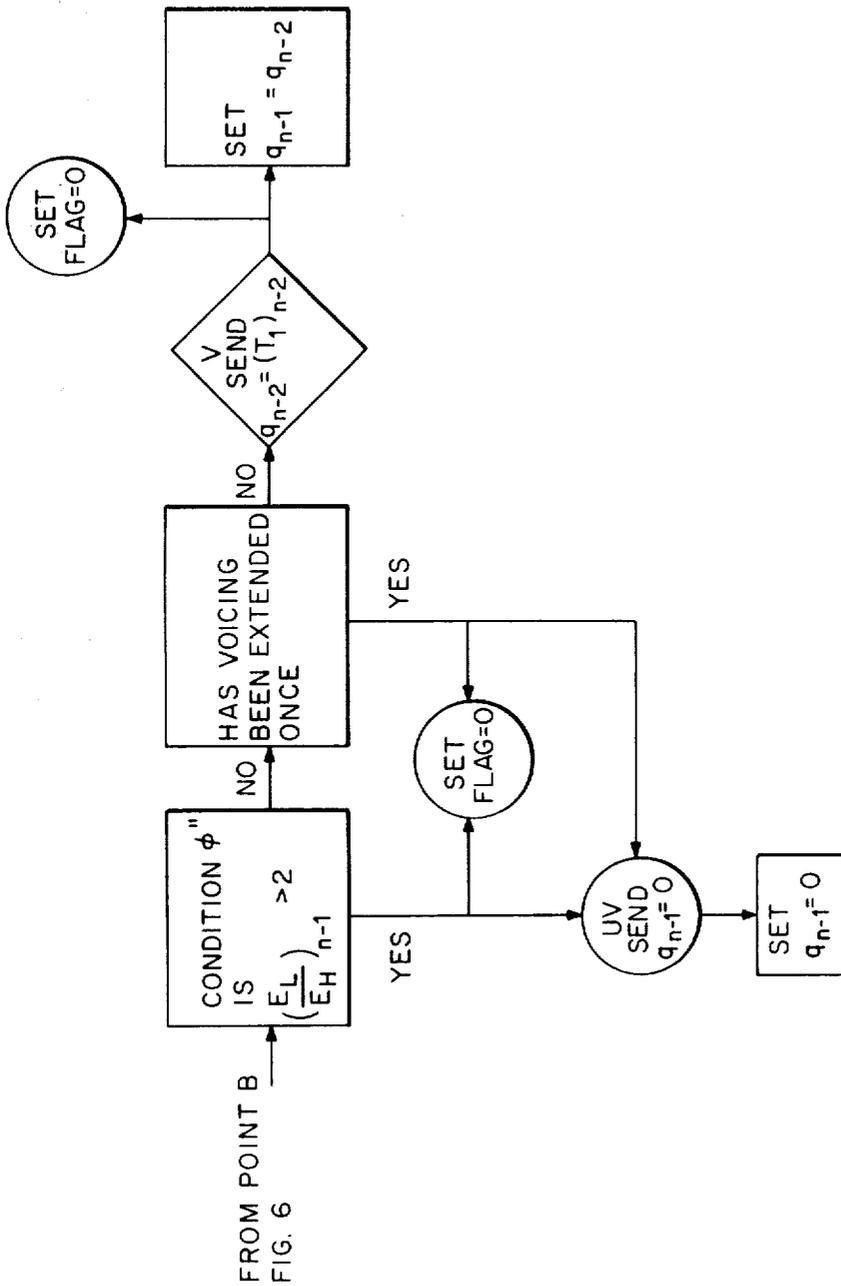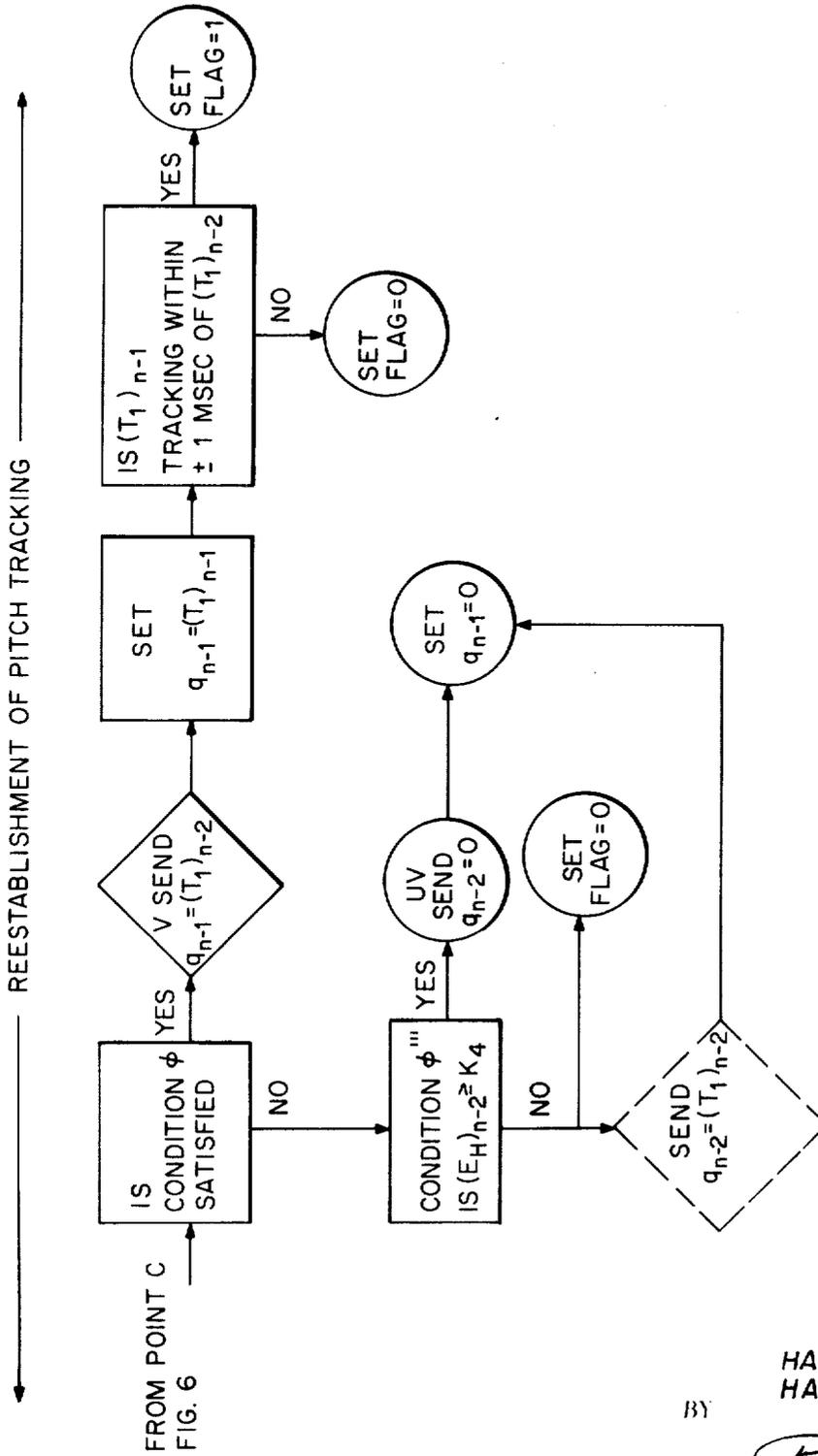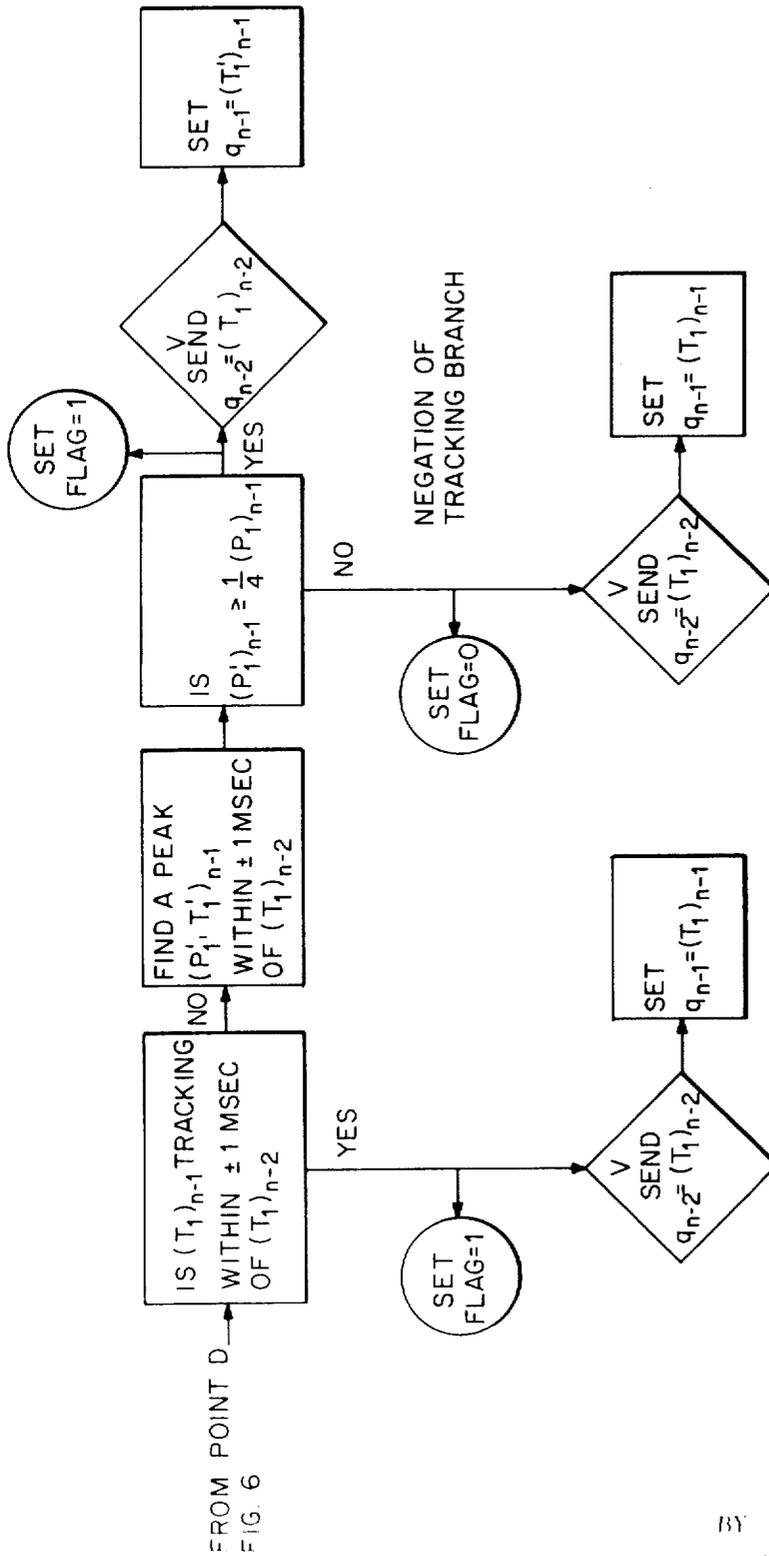HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T. Orner
ATTORNEY

Fig. 14.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER
BY
Robert T. Orner
ATTORNEY

TO STORAGE
REGISTER 360

TO STORAGE
REGISTER 361

TO STORAGE
REGISTER 362

TO
STORAGE
REGISTER 363

SHIFT
PULSES

AND
GATE

400

SHIFT
PULSES

396

AND
GATE

FLIP-FLOP
CKT

398

SHIFT
PULSES

392

AND
GATE

FLIP-FLOP
CKT

394

AND
GATE

386

FLIP-FLOP
CKT

388

SHIFT
PULSES

AND
GATE

FLIP-FLOP
CKT

390

384

AND
GATE

380

FLIP-FLOP
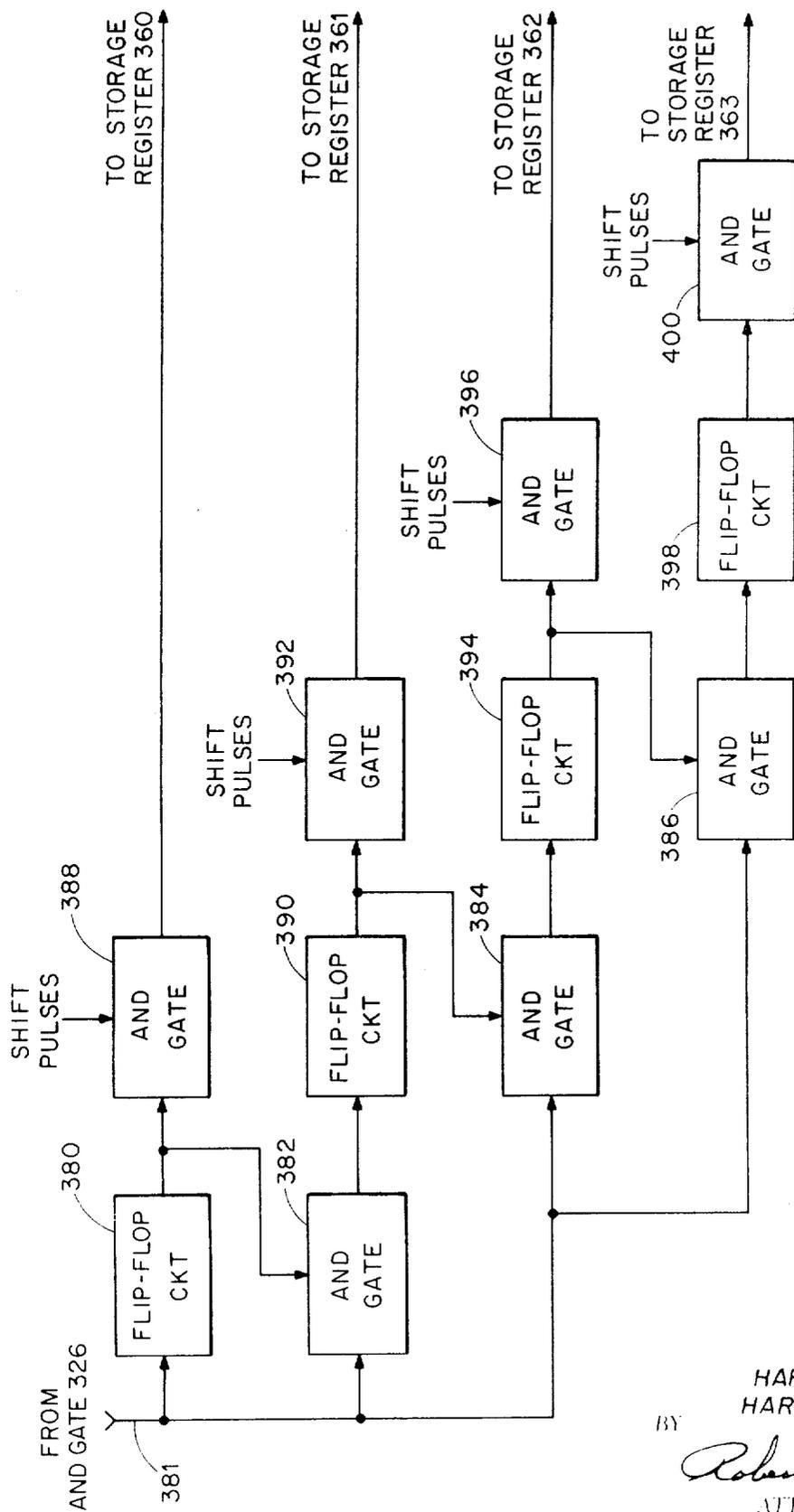CKT

382

AND
GATE

FROM
AND GATE 326

381

Fig. 15.

INVENTORS
HAROLD J. MANLEY
HARRY L. SHAFFER

BY

Robert J. Crimer

ATTORNEY

1

METHOD AND APPARATUS FOR SIGNAL BANDWIDTH COMPRESSION UTILIZING THE FOURIER TRANSFORM OF THE LOGARITHM OF THE FREQUENCY SPECTRUM MAGNITUDE

BACKGROUND OF THE INVENTION

This invention relates to speech compression systems and in particular to digital vocoder systems.

It is well-known that the vocal tract, consisting of throat, mouth, tongue, lips, teeth and nasal passages, forms a time varying linear filter in which the amplitude response versus frequency characteristics is responsible for practically all the information content in a speech signal. This filter is driven by energy sources, commonly known as "buzz" and "hiss" energy sources.

The term "buzz" is associated with the type of vocal source excitation function which exists when the vocal cords are oscillating at some quasi-periodic rate (called the pitch). Under this condition the chest cavity is supplying puffs of air to the vocal tract at the quasi-periodic rate at which the vocal cords are oscillating. The term "hiss" is associated with the type of vocal source excitation which exists when the vocal cords are not oscillating in a quasi-periodic manner but are always allowing air to pass through from the chest cavity and excite the vocal tract.

For voiced sounds, e.g., vowels, the excitation is from the buzz energy source. For unvoiced sounds, e.g., ss, sh, f and whispered speech, the excitation is from the hiss source. The information content is impressed upon the speech signal by the vocal tract acting essentially as a time varying distributed constant linear filter. Thus, to recreate speech which is both intelligible and natural sounding, it is necessary to use both the information describing the time varying spectral shape and the information describing the buzz and hiss energy sources. The latter information generally takes the form of measurements of the fundamental frequency of the buss sources as a function of time (pitch extraction). Information as to whether the excitation is buzz or hiss is used by the speech compression system. Combinations of buzz and hiss excitation are used to generate some sounds, but speech compression systems do not generally try to detect the combined excitation. A decision is usually made as to whether to use buzz or hiss excitation for this combined excitation in a speech compression system of this type.

Speech compression systems using spectral analysis are generally called vocoders. In existing speech compression systems, the spectrum data are transmitted by digitally encoding the logarithm of about 16 voltage spectrum amplitude which are derived from a filter bank spectrum analyzer. This method is known to be inefficient because of the high correlations among the various spectrum amplitudes. Various techniques are now used to remove these correlations and therefore reduce the required data rate for a given transmission fidelity. One approach which produces some improvements is the use of a delta pulse code modulation scheme in which only the decibel differences in level between adjacent frequency channels are transmitted. Another scheme is to form weighted sums of the logged, digitized spectrum amplitudes, the weighting being arranged so that cross-correlation of the speech wave against a waveform derived from the input speech are markedly reduced.

2

Another type of vocoder is called the autocorrelation vocoder which derives its name from the fact that in the first step of the analysis process the autocorrelation function of the speech input is measured in terms of orthonormal functions. Just as the power spectrum of the speech input varies with time (as a talker articulates various sounds), so does the autocorrelation function. There is a one-to-one correspondence between the power spectrum and the autocorrelation function of the speech signal so that measuring one is equivalent to measuring the other. Mathematically, the power spectrum and the autocorrelation function are Fourier transform pairs. Thus, autocorrelation is simply an alternative method of measuring the short time energy spectrum of the speech signal. In an autocorrelation vocoder, the input signal is first applied to the inputs of a set of orthogonal filters. The filter output signals are multiplied by the input speech signal, and the product signal is then directed through low pass filters. The output signals from the low pass filter are the coefficients in an expansion of the power spectrum.

The power spectrum $P(f)$ of a speech signal is the product of the power spectrum of a pitch excitation, $V(f)$, and the magnitude squares $|H(f)|^2$ of a vocal tract transfer function H(f).

$$P(f) = |H(f)|^2 \, V(f) \qquad (1)$$

As stated above, the autocorrelation function is the Fourier transform of $P(f)$ and is composed of the convolution of the transform of $|H(f)|^2$ and $V(f)$. Practically, this means that the autocorrelation function repeats itself at multiples of the pitch period, and it is necessary to represent the vocal tract out to fairly large delay values (near one-half of a pitch period) in order to represent the speech spectrum with any fidelity. The overlap of successive autocorrelation functions due to convolution properties raises some doubt as to the validity of the values of the autocorrelation function alone as a measure of the vocal tract shape. While the autocorrelation vocoder obtains nearly independent spectral measurements, it does not solve the problem caused by confounding the spectral envelope (vocal tract) data with the excitation spectrum data, which results in higher order transmitted coefficients. Furthermore, this type of vocoder is basically an analog device yielding an output consisting of voltage spectrum values which are subsequently digitized.

SUMMARY OF THE INVENTION

Briefly, a bandwidth compression system according to the present invention includes a means for generating an electrical signal representing the Fourier transform of the logarithm of the spectrum magnitudes (FTLSM) of an input signal having excitation and impulse response information included therein. A first detection means, coupled to the means for generating FTLSM electrical signal, is operative to separate out a first predetermined portion of the FTLSM electrical signal to represent the excitation information of the input signal. A second detection means, also coupled to the means for generating the FTLSM electrical signal, is operative to separate out a second predetermined portion of the FTLSM electrical signal to represent the impulse response information of the input signal. The bandwidth required to pass the combined first and

second predetermined portions is less than the bandwidth required to pass the input signal.

The bandwidth compression system further includes a synthesis section comprising am impulse response means coupled to the second detection means and being operative in response to the predetermined number of the first set of predetermined signals to generate an output signal corresponding to the impulse response information. An excitation means, coupled to the first detection means, is operative in response to the output signal from the first detection means to generate an excitation carrier signal. A convolution means having input connections from the excitation means and the impulse response means is operative to convolve the output signals from the impulse response means and the excitation means to thereby synthesize the speech signal.

A second embodiment of a bandwidth compression system according to the present invention is operative to simultaneously synthesize a first speech signal, for example, one received from a remote terminal, and analyze a second speech signal, for example, one to be transmitted at reduced bandwidth to a remote terminal. The system includes means for generating a pitch carrier signal from the first speech signal and means for obtaining the frequency spectrum magnitude of the first speech signal. Coupled to the means for obtaining the frequency spectrum magnitude of the first speech signal is a means for converting the frequency spectrum magnitudes into signals having a first predetermined symmetry.

A computing means, having first and second input ports and first, second, third and fourth output ports, is operative to compute simultaneously the Fourier transform of a set of first predetermined input signals at the first input port, the set of first predetermined input signals having a composite symmetry of the first and second predetermined symmetries, and the Fourier transform of a set of second predetermined input signals at said second input port, the set of second predetermined signals having the first and second predetermined symmetries. The computing means directs to the first, second, third and fourth output ports, respectively, a first set of output signals representing the Fourier transform of the portion of the first predetermined input signals having the second predetermined symmetry, a second set of output signals representing the Fourier transform of the portion of the set of first predetermined input signals having the first predetermined symmetry, a third set of output signals having the first predetermined symmetry and representing the Fourier transform of the portion of the set of second predetermined input signals at said second input port; and a fourth set of output signals representing the Fourier transform of the portion of the second set of predetermined input signals having the second predetermined symmetry.

A sampling means coupled to the first input port of the computing means is operable to sample the second speech signal over a first predetermined time interval. The first and second sets of output signals of the computing means then represents the spectrum of the second input speech signal. A magnitude means, coupled to the first and second input ports of the computing means, is operative to combine in a predetermined manner the first and second sets of output signals of the computing means to generate signals representing the frequency spectrum magnitudes of the second speech signal.

Coupled to the magnitude means is a means for generating output signals having the second predetermined symmetry and representing the logarithm of the frequency spectrum magnitudes of the second speech signal. A summing means, having input connections coupled to the means for converting and to the means for generating and an output connection coupled to the second input port of the computing means, is operative to sum the signals of the first predetermined symmetry with the signals having the second predetermined symmetry to form the set of second predetermined input signals. The third set of output signals of the computing means then represents the vocal tract information of the first speech signal and the fourth set of output signals of the computing means in the FTLSM representing the voicing and vocal tract data of the second speech input signal.

A pitch detection logic means, coupled to the fourth output port of the computing means, is operative to extract a pitch signal from the fourth set of output signals to represent the voicing information of the second input speech signal. Also coupled to the fourth output port of the computing means is a selecting means operative to select a predetermined number of the fourth set of output signals to represent the vocal tract information of the second input speech signal. The pitch signal and the output signals of the selecting means represent the analyzed second input speech signal having a substantially compressed bandwidth. The pitch carrier signal and the third set of output signals of the computing means when convolved in a convolution means represent the synthesized version of the first input speech signal.

A method of compressing the bandwidth of an input signal having an excitation portion and an impulse response portion comprises the steps of generating a time variant electrical signal representing the FTLSM of the input signal, separating out a first time interval signal of the time variant electrical signal to represent the impulse response portion of the input signal and separating out a second time interval signal of said time variant electrical signal to represent the excitation portion of the input signal. The first and second time interval signals of the time variant signal represent the reduced bandwidth input signal.

BRIEF DESCRIPTION OF THE DRAWINGS

The construction and operation of the invention will be more fully understood from the following detailed description taken in conjunction with the accompanying drawings in which:

FIGS. 1A and 1B are a series of waveforms useful in explaining the concept of speech compression;

FIGS. 2A and 2B together represent a block diagram of an embodiment of analysis section of a speech compression system in accordance with the present invention;

FIG. 3 is a series of waveforms useful in explaining the operation of the embodiment of FIGS. 2A and 2B;

FIG. 4 is a block diagram of an embodiment of a pitch detection logic unit employed in the embodiment of FIGS. 2A and 2B;

5

6

FIG. 5 is a block diagram of an energy ratio detector employed in the pitch detection logic unit of FIG. 4;

FIGS. 6 through 9B are a series of flow charts useful in implementing the functions of the pitch detection logic unit of FIGS. 2A and 2B on a programable computer;

FIGS. 10A and 10B together form a block diagram of an embodiment of a decoding device employed in the synthesis section of the speech compression system according to the present invention;

FIG. 11 is a block diagram of an embodiment of a weighting and averaging circuit employed in the synthesis section of the speech compression system according to the present invention;

FIG. 12 is a block diagram of an embodiment of a pitch carrier generator employed in the synthesis section of the speech compression system according to the present invention;

FIG. 13 is a series of waveforms useful in explaining the operation of the pitch carrier and a convolution means both of which are employed in the synthesis section of the speech compression system according to the present invention;

FIG. 14 is a block diagram of an embodiment of a convolution means employed in the synthesis section of the speech compression system according to the present invention; and

FIG. 15 is a block diagram of an embodiment of a gating circuit employed in the convolution means of FIG. 14.

DETAILED DESCRIPTION OF THE INVENTION

Mathematical Preliminaries

The vocal track $h(\tau)$ (nasal and mouth cavities) is a time varying filter, and the vocal source $v(t)$ (chest cavity glottal source and pharynx cavity) is the oscillator function of the time varying filter. The output signal $s(t)$ then is the convolution of $v(t)$ and $h(\tau)$. It is well-known that convolution in the time domain corresponds to multiplication in the Fourier transform domain. The resultant signal $S(\omega)$ is equal to the product of $V(\omega)$ and $H(\omega)$ where $S(\omega)$, $T(\omega)$, and $H(\omega)$ are the Fourier transforms of $s(t)$, $v(t)$ and $H(\tau)$ respectively. Typical waveforms for $V(t)$, $s(t)$, $V(\omega)$, $H(\omega)$ and $S(\omega)$ are shown respectively in waveforms (a) through (e) of FIG. 1.

The mathematic basis for calculating the FTLSM of a speech signal will be discussed in conjunction with these waveforms. The vocal source output signal $v(t)$ is a quasi-periodic function with a period T for voiced sounds with an an output speech spectrum $S(\omega)$ being represented by harmonically related narrow bands of energy spaced 1/T apart. The spectrum envelope shape of waveform (e) is similar to the vocal tract transfer function $H(\omega)$ of waveform (d). Note that the spectrum $S(\omega)$ of the speech signal (waveform e) has a high frequency component represented by the narrow bands of harmonically spaced energy and a low frequency modulation in the form of a spectrum envelope shape.

A speech compression system which is based on obtaining vocal source information and vocal tract information could use the output speech spectrum if some convenient processing existed which would separate the envelope information from the fine structure information. One operation that can be used to separate the

product of two functions is a logging operation. The resulting function after logging the spectrum has a slowly varying envelope and a fast ripple (1/T) riding on the slow envelope. The The Fourier transform of the logged signal, log $| \ V(\omega) \ |$, has a spike, the position T of which on the time axis is related to the reciprocal of the periodic component and is shown in waveform (f) of FIG. 1A. The FTLSM separates the components into two distinct time regions. The lower time region, x, represents the transform of log$| \ V(\omega) \ |$ and the upper region has a peak which is related to the period T of the vocal source function.

In the discussion which follows, reference will be made many times to sampled functions defined at integer values of an independent variable (using K or n) ranging from 0 to a positive upper limit $N-1$. By way of example N will be 256 or 2^8. In any case, it is of great practical advantage to have N defined by a relation of the form $N = n^k$ where both n and k are integers because this makes possible the use of the most efficient methods of calculating discrete Fourier transforms of sampled functions where the samples are labeled consecutively over the interval 0 to $N-1$. These computationally advantageous methods are generally referred to by the term "Fast Fourier Transform," or briefly the "FFT."

The sampled functions to which reference is made can be portrayed by graphs of the kind shown in waveforms (g), (h) and (i) of FIG. 1B (for purposes of illustration $N=8$).

In waveform (g), a very simple sampled function f_K is shown. The numerical values of the samples (ordinates) are indicated by small circles. The function is defined only at discrete, integer values $K = 0, 1, 2, \ldots$, $N-1$ of the abscissa. Thus, the function f_K is an ordered set of N real numbers (an N-tuple) and such a function will often be referred to as a vector.

It will often be useful to break up or resolve functions like f_K into their even and odd components about the point $N/2$. The even and odd parts of f_K are thus defined as $(f_K + f_{N-K}/2)$ and $(f_K - f_{N-K}/2)$, respectively. f_{N-K} is obtained simply by reversing the order of the samples of f_K on the interval 0 to N. The even and odd parts of f_K are plotted in waveforms (h) and (i), respectively, of FIG. 1B. A function is obviously the sum of its even and odd parts, e.g.,

$$f_K = \left(\frac{f_K + f_{N-K}}{2} \right) + \left(\frac{f_K - f_{N-K}}{2} \right) \tag{2}$$

The operation of calculating the even and odd parts of a function will be called even-odd separation.

For mathematical convenience, complex sampled functions will also be defined only for integer values of an independent variable K or n. A sampled complex function is of the form:

$$Z_K = X_K \ 30 \ j Y_K \tag{3}$$

where, in general, the function X_K and Y_K have both an even and odd part which can be calculated in the same manner that the even and odd parts of the function f_K is calculated. The operator j is defined as

$$j = {}^+ \sqrt{-1} \tag{4}$$

and X_K and Y_K are real N-tuples similar to f_K. The functions X_K and Y_K are respectively referred to as the real and imaginary parts of Z_K. Thus, to store a complex N-tuple such as Z_K for $K = 0, 1, 2, \ldots, N-1$ in a digital machine requires N separate memory locations for the values of X_K and another set of N memory locations for the values of Y_K.

As part of the invention, it will be necessary to calculate the discrete Fourier cosine transform (DFCT) and the discrete Fourier sine transform (DFST) of sampled functions. The DFCT of a function is itself a sampled function defined only for integer values of some independent variable (say $n = 0, 1, 2, -, N-1$). The DFCT of the function Y_K is defined as

$$\sum_{K=0}^{N-1} Y_K \cos \frac{2\pi nK}{N} \qquad (5)$$

for $n = 0, 1, 2, \ldots, N-1$

Note that since the cosine is itself an even function the DFCT of any function Y_K depends only on the even part of Y_K, i.e., the DFCT of an odd function is zero. It is easy to show that the DFCT of Y_K is the same as the DFCT of its even part. Thus,

$$\sum_{K=0}^{N-1} Y_K \cos \left(\frac{2\pi nK}{N}\right) = \sum_{K=0}^{N-1} Y_{N-K} \left(\frac{\cos 2\pi nK}{N}\right) \qquad (6)$$

therefore,

$$\sum_{K=0}^{N-1} \left(\frac{Y_K + Y_{N-K}}{2}\right) \cos \left(\frac{2\pi nK}{N}\right) = \sum_{K=0}^{N-1} Y_K \cos \left(\frac{2\pi nK}{N}\right) \qquad (7)$$

and

$$\sum_{K=0}^{N-1} \left(\frac{Y_K - Y_{N-K}}{2}\right) \cos \left(\frac{2\pi nK}{N}\right) = 0 \qquad (8)$$

so that the DFCT of the odd part of any function is zero.

Similarly the DFST of Y_K is defined as

$$\sum_{K=0}^{N-1} Y_K \sin \frac{2\pi nK}{N} \qquad (9)$$

Since the sine is itself an odd function, the DFST of any function Y_K depends only on the odd part of Y_K, i.e., the DFST of an odd function is zero. It is easy to show that the DFST of Y_{N-K} is the negative of the DFST of Y_K. Thus,

$$\sum_{K=0}^{N-1} Y_K \sin \left(\frac{2\pi nK}{N}\right) = \sum_{K=0}^{N-1} Y_{N-K} \sin \left(\frac{2\pi nK}{N}\right) \qquad (10)$$

$$\sum_{K=0}^{N-1} \left(\frac{Y_K - Y_{N-K}}{2}\right) \sin \left(\frac{2\pi nK}{N}\right) = 0 \qquad (11)$$

and

$$\sum_{K=0}^{N-1} \left(\frac{Y_K - Y_{N-K}}{2}\right) \sin \frac{2\pi nK}{N} = \sum_{K=0}^{N-1} Y_K \sin \frac{2\pi nK}{N} \qquad (12)$$

so that the DFST of the even part of any function is zero.

The FFT box or computer that will be described actually calculates the discrete Fourier transform (DFT) of a complex input vector, say $Z_K^{(1)}$. The DFT is defined by Equation (13) with superscripts (1) and (2) denoting inputs and outputs respectively.

$$Z_n^{(2)} = \sum_{K=0}^{N-1} Z_K^{(1)} e^{-j2\pi \frac{nK}{N}} \qquad (13)$$

$Z_K^{(1)}$ is the complex input to the FFT and $Z_n^{(2)}$ is the complex output of the FFT. Using the well-known identity

$$e^{-j\theta} = \cos \theta - j \sin \theta \qquad (14)$$

yields

$$e^{-j2\pi \frac{nK}{N}} = \cos \left(\frac{2\pi nK}{N}\right) - j \sin \left(\frac{2\pi nK}{N}\right) \qquad (15)$$

Substituting Equation 15 and Equation 3

$$Z_K^{(1)} = X_K^{(1)} + jY_K^{(1)} \qquad (3)$$

into Equation 13 yields the complex output in terms of the DFCT's and DFST's of $X_K^{(1)}$ and $Y_K^{(1)}$:

$$Z_n^{(2)} = \sum_{K=0}^{N-1} \left\{ X_K^{(1)} \cos \left(\frac{2\pi nK}{N}\right) - jX_K^{(1)} \sin \left(\frac{2\pi nK}{N}\right) \right.$$
$$\left. + jY_K^{(1)} \cos \left(\frac{2\pi nK}{N}\right) + Y_K^{(1)} \sin \left(\frac{2\pi nK}{N}\right) \right\} \qquad (16)$$

Inspection of Equation 16 shows the output $Z_n^{(2)}$ of the FFT appears in two separate parts. Its real part $R_e[Z_n^{(2)}]$

$$R_e[Z_n^{(2)}] = \sum_{K=0}^{N-1} X_K^{(1)} \cos \left(\frac{2\pi nK}{N}\right) + \sum_{K=0}^{N-1} Y_K^{(1)} \sin \left(\frac{2\pi nK}{N}\right) \qquad (17)$$

and its imaginary part $I_m[Z_n^{(2)}]$:

$$I_m[Z_n^{(2)}] = \sum_{K=0}^{N-1} -X_K^{(1)} \sin \left(\frac{2\pi nK}{N}\right)$$
$$+ \sum_{K=0}^{N-1} Y_K^{(1)} \cos \left(\frac{2\pi nK}{N}\right) \qquad (18)$$

Using the facts developed above the even functions of K will have nonzero DFCT's and odd functions of K will have nonzero DFST's, we notice that $R_e[Z_n{}^{(2)}]$ is the sum of the DFCT of the even part of the real input $X_K{}^{(1)}$ plus the DFST of the odd part of the imaginary input $Y_K{}^{(1)}$, i.e.,

$$R_e[Z_n{}^{(2)}]=\sum_{K=0}^{N-1}\left(\frac{X_K{}^{(1)}+X^{(1)}{}_{N-K}}{2}\right)\cos\frac{2\pi nK}{N}$$
$$+\sum_{K=0}^{N-1}\left(\frac{Y_K{}^{(1)}-Y^{(1)}{}_{N-K}}{2}\right)\sin\frac{2\pi nK}{N}$$

(18-A)

Similarly $I_m[Z_n{}^{(2)}]$ is the DFCT of the even part of the imaginary input $Y_K{}^{(1)}$ minus the DFST of the odd part of the real input.

At this point it is interesting to note that the complex output function of the FFT can be considered to be made up of four functions. The real output $R_e[Z_n{}^{(2)}]$ is the sum of an even and odd function, and similarly the imaginary output also is a sum of an even and odd function.

If a procedure is defined to process one of the parts of the complex output (i.e., $R_e[Z_n{}^{(2)}]$ or $I_m[Z_n{}^{(2)}]$ in such a way that one of the summation terms in the expression for $R_e[Z_n{}^{(2)}]$ or $I_m[Z_n{}^{(2)}]$ would drop out of the final result, then this procedure would in effect separate each of the parts of the complex output to subparts which exhibited the property of being either even functions or odd functions of K. An appropriate name for such a process might be an EVEN/ODD separator.

The development of this process will begin by first calculating the expressions for the time reversed $R_e[Z_n{}^{(2)}]$, i.e., $R_e[Z^{(2)}{}_{N-n}]$. The reversed function is written as

$$R_e Z^{(2)}{}_{N-n}=\sum_{K=0}^{N-1}X_K{}^{(1)}\cos\frac{2\pi K(N-n)}{N}$$
$$+\sum_{K=0}^{N-1}Y_K{}^{(1)}\sin\frac{2\pi K(N-n)}{N}$$

(19)

Using familiar trigonometric identities to simplify the argument of the cosine function, it can be written

$$\cos\frac{2\pi K(N-n)}{N}$$

(20)

$=\cos(2\pi K)\cos(2\pi Kn/N+\sin(2\pi Kn/N$
$=(1)\cos(2\pi Kn/N)+(0)\sin(2\pi Kn/N)$
$=\cos(2\pi Kn/N)$

Similarly it can be shown that

$$\sin\frac{2\pi K(N-n)}{N}=-\sin\left(\frac{2\pi Kn}{N}\right)$$

(21)

Using Equations 20 and 21 in Equation 19 yields the following expression for the reversed real output:

$$R_e[Z^{(2)}{}_{N-n}]=\sum_{K=0}^{N-1}X_K{}^{(1)}\cos\left(\frac{2\pi nK}{N}\right)$$
$$-\sum_{K=0}^{N-1}Y_K{}^{(1)}\sin\left(\frac{2\pi nK}{N}\right)$$

(22)

Adding together the functions $R_e[Z_n{}^{(2)}]$ and $R_e[Z^{(2)}{}_{N-n}]$ and dividing by two yields the following equation:

$$\frac{R_e[Z_n{}^{(2)}]+R_e[Z^{(2)}{}_{N-n}]}{2}=\sum_{K=0}^{N-1}X_K{}^{(1)}\cos\left(\frac{2\pi nk}{N}\right)$$

DFCT of even part of $X_K{}^{(1)}$

(23)

which is seen to be the even part of $R_e[Z_n{}^{(2)}]$. The odd part of $R_e[Z_n{}^{(2)}]$ can be calculated by subtracting $R_e[Z_n{}^{(2)}]$

$$\frac{R_e[Z_n{}^{(2)}]-R_e[Z^{(2)}{}_{N-n}]}{2}=\sum_{K=0}^{N-1}Y_K{}^{(1)}\sin\left(\frac{2\pi nK}{N}\right)$$

DFST of odd part of $Y_K{}^{(1)}$

(24)

To calculate the even and odd parts of imaginary output $I_m[Z_n{}^{(2)}]$, the reversed quantity $I_m[Z^{(2)}{}_{N-n}]$ is also required. Using the identities in Equations 20 and 21 yields

$$I_m[Z^{(2)}{}_{N-n}]=\sum_{K=0}^{N-1}\left\{X_K{}^{(1)}\sin\left(\frac{2\pi nK}{N}\right)\right.$$
$$\left.+Y_K{}^{(1)}\cos\left(\frac{2\pi nK}{N}\right)\right\}$$

(25)

so that we may simply calculate the even part in a manner identical to the even part of the FFT

$$\frac{I_m[Z_n{}^{(2)}]+I_m[Z^{(2)}{}_{N-n}]}{2}=\sum_{K=0}^{N-1}Y_K{}^{(1)}\cos\left(\frac{2\pi nK}{N}\right)$$

DFCT of even part of $Y_K{}^{(1)}$

(26)

and the odd part of the imaginary output of the FFT is calculated as

$$\frac{I_m[Z_n{}^{(2)}]-I_m[Z^{(2)}{}_{N-n}]}{2}=\sum_{K=0}^{N-1}X_K{}^{(1)}\sin\left(\frac{2\pi nK}{N}\right)$$

DFST of odd part of $X_K{}^{(1)}$

(27)

Equations 23, 24, 26 and 27 show how two DFCT's and two DFST's are obtained by operations on the real and imaginary parts of the DFT output of the FFT box. In a latter section of the application, the various even and odd parts of the real and imaginary inputs and outputs of the DFT and the corresponding FFT computer with various time functions and discrete Fourier transforms to be dealt with in the speech processing system will be identified.

System Description — Part I

A digital vocoder system according to the present invention is shown in block diagram form in FIG. 2 and includes a transducer 10, such as a microphone, connected to an analog to digital (A/D) converter 12. The output of the A/D converter 12 is connected to a first buffer memory 14 and to a normalization unit 16 which includes an accumulator 18 and an inclusive OR matrix 20, such that blocks of samples from the A/D converter 12 are combined in the inclusive OR matrix 20. The inclusive OR of a block of samples equal in length to the updating number of samples used in the system is then processed to yield a normalizer gain factor for the samples in the buffer memory 14. While the inclusive OR matrix 20 is processing data, it also is supplying the normalization gain for the data being processed from the buffer memory 14. The data loaded into the accumulator 18 is shifted either right or left by the appropriate number of bits by the control signal defining the normalization gain derived from the inclusive OR matrix 20. Connected to the output of the normalization unit 16 is weighting function circuit 22 which includes a digital multiplier circuit 24 having input connections from the normalization unit 16 and from a weighting function storage unit 26.

The output connection of the weighting function circuit 22 is connected to a real input $R_K^{(1)}$ of an FFT computer means 30 such as the Sylvania Electric Products Inc. ACP-1 computer. The FFT computer means 30 includes an FFT section 32 and an even/odd (E/O) separator section 34, to be explained in detail hereinafter, and has first, second, third and fourth output terminals R_n, I_n, H_n and C_n, respectively. The first and second output terminals R_n and I_n are connected to a first magnitude approximation unit 38. An encoding unit 40 has input connections from the output of the normalization unit 16, the fourth output terminal C_n of the FFT computer means 30 and from the first magnitude approximation unit 38 and is operative to detect a pitch signal and to encode the pitch and spectral signals for transmission. Also connected to the first magnitude approximation circuit 38 is a logging algorithm computer 44, the output of which is connected to the input of an even function generator 46. The output side of the even function generator 46 is connected through a summation circuit 48 to the imaginary part input terminal $I_K^{(1)}$ of the FFT computer 30.

The system can be divided into two sections, analysis and synthesis sections. The units described thus far are employed in the analysis of an input speech waveform. It may be helpful at this point to describe, in conjunction with the waveform of FIG. 3, the operation of the analysis section of the vocoder system since many of the units employed in the synthesis section perform the reverse function of units in the analysis section.

An input sound wave is impressed on the transducer 10 and is converted to a continuous electrical signal shown as the solid envelope waveform (a) of FIG. 3. The continuous electrical signal is converted by the A/D converter 12 at the specified sampling rate. Blocks of 256 converted samples are stored in the buffer memory 14. The A/D converter 12 output samples are always processed by the normalizing unit 16, the purpose of which is to establish the appropriate normalizer gain so that the louder speech sounds have the N samples in their analysis intervals normalized to a fixed

dynamic range, for example 6 db. For weak sounds, the normalizer gain factor will shift the N samples of the analysis interval to make the N samples appear to have more amplitude. This is done to keep the input sample level to the weighting function circuit 22 and consequently the rear input $R_K^{(1)}$ of the FFT computer means 30 at a high signal input level.

A constant scaling factor is applied to the N samples by the normalizing unit 16 before being directed to the weighting function circuit 22 where the data is effectively multiplied by a smooth weighing function. Both the normalizing unit 16 and the weighting function circuit 22 will be discussed in detail hereinafter. The sampled, normalized and weighted data is then directed to the real part input $R_K^{(1)}$ of the FFT computer means 30.

The output signals of the FFT computer means 30 include three transforms: (1) the transform of the speech signal which includes the real and imaginary parts of the speech spectrum signal, (2) the transform of the received spectrum envelope which is the vocal tract impulse response H_n and (3) transform of the logarithm of magnitude spectrum which is the FTLSM function C_n. The transform of the received spectrum H_n will be discussed in connection with the synthesis section of the vocoder system.

The real and imaginary parts of the speech spectrum signal are directed to the magnitude approximation circuit 38 where they are combined to obtain the magnitude of $N/2$ samples of the spectrum. The magnitude of each of the samples may be obtained by taking the square root of the sum of squares of the real and imaginary parts of that sample. To reduce the number of calculations, a magnitude approximation circuit 38 is employed in lieu of taking the square root of the sums of the squares of the real and imaginary parts. The output signal of the magnitude approximation circuit 38 is directed to the logging algorithm computer 44 where the $N/2$ samples of the spectrum magnitude are logged and directed to the even function generator 46 which converts the $N/2$ samples to an even function signal having N samples. The logged magnitude spectrum signal (an even function) is combined with a received signal (an odd function) from the synthesis section to form the imaginary input $I_K^{(1)}$ of the FFT computer means 30.

The output signal from the magnitude approximation circuit 38 and the FTLSM signal C_n from the FFT computer means 30 are directed to the encoding unit 40 where they are combined to extract pitch data as well as spectral envelope information for transmission to a receiving unit (not shown).

COMPONENT DESCRIPTION

FFT Computer Means 30

In this section is described the physical significance of the inputs of the FFT section 32 and the E/O separator section 34.

The FFT section 32 takes an N sample, complex input vector $S_K^{(1)}$, and computes an N sample complex output vector $S_n^{(2)}$ in accordance with the discrete complex FFT relation

$$S_n^{(2)} = \sum_{K=0}^{N-1} S_K^{(1)} e^{-j2\pi \frac{nK}{N}}$$

(28)

where $K = 0, 1, 2, 3, \ldots, N{-}1$
and $n = 0, 1, 2, 3, \ldots, N{-}1$

In the present application, N is typically equal to 256 samples. Again, the inputs and outputs to the FFT section are denoted by superscripts, (1) for inputs and (2) for outputs.

The complex input vector $S_K^{(1)}$ has a real part $R_K^{(1)}$ and an imaginary part $I_K^{(1)}$ so that

$$S_K^{(1)} = R_K^{(1)} + jI_K^{(1)} \tag{29}$$

The real input vector, $R_K^{(1)}$, is contained in a set of N registers or storage locations (not shown).

The input signals $R_K^{(1)}$ are the samples of the input speech waveform presently to be analyzed and transmitted. It is thus the sum of the even and the odd part of the N samples of the input speech waveform to be analyzed. The $R_K^{(1)}$ signals will typically look like waveform (a) of FIG. 3 for a voiced speech signal waveform input and is, in general, neither a purely even nor an odd function of K.

Each sample of the $R_K^{(1)}$ signal is stored in one of the N registers or storage locations, the Kth sample $R_K^{(1)}$ being in the Kth location, $K = 0, 1, 2, 3, \ldots, N{-}1$. Since the $R_K^{(1)}$ signal has both an even and odd part, it will have both a nonzero discrete cosine transform R_n as well as a nonzero discrete sine transform I_n. The R_n and I_n are, respectively, the samples of the real and imaginary parts of the discrete FFT of the analyzed speech waveform $R_K^{(1)}$. R_n and I_n are two of the outputs to be obtained from the FFT computer means 30.

$I_K^{(1)}$, the imaginary input, is contained in another set of N registers or storage locations (not shown) with again the Kth sample $I_k^{(1)}$ being in the Kth location $K = 0, 1, 2, 3, \ldots, N{-}1$. $I_K^{(1)}$ is the sum of an even and odd function from the summation circuit 48. In this system the even part of $I_K^{(1)}$ is defined as the logarithm of the magnitude of the input speech signal to be transmitted and which was spectrum analyzed during immediately previous analysis operation of the FFT section 32.

The even part of $I_K^{(1)}$ is $\frac{1}{2}(I_K^{(1)} + I^{(1)}_{N-K})$ and, for a voiced speech signal input, will typically look like the sampled function of waveform (b) in FIG. 3. Since $\frac{1}{2}(I_K^{(1)} + I^{(1)}_{N-K})$ or the logged spectrum magnitude of the signal to be transmitted is a purely even function of K, centered around $K = N/2$, it will have a nonzero discrete cosine transform C_n and an identically zero discrete sine transform. Thus the Fourier cosine transform of the even part of $I_K^{(1)}$ is $C_n \cdot C_n$ is another of the outputs obtained from the operation of the FFT section 32 and E/O separator section 34. (A typical C_n function is shown in waveform (c) of FIG. 3.) The signal C_n includes the samples of the cosine transform of the logarithm of the magnitude spectrum of the input speech signal. The function C_n is therefore even and is referred to as the FTLSM of the input speech signal samples $R_K^{(1)}$ for the previous analysis interval. The samples of C_n for $n = 0, 1, 2, 3, \ldots, 20$ are, for example, used in the encoding unit 40 as the spectrum envelope information to be transmitted. C_0 is the average value of the logged spectrum magnitude and is always the largest C_n signal. For voiced speech signals, the signal C_n will have a noticeable peak at a value of $n = n_p$

approximately equal to the number of waveform samples in one pitch period of the voiced sounds. Thus, n_p is a measure of the pitch period of the signal to be transmitted and the value of n_p is therefore measured and transmitted so that the receiver may use this information in order to synthesize a speech signal with the correct pitch.

In this system, the odd part of the imaginary input $I_K^{(1)}$ is defined as equal to a received spectrum magnitude function $M_K^{(r)}$ (from the synthesis section) which has been arranged in reflected and inverted form so as to be an odd function of K centered about $K = N/2$, i.e.

$$M_K^{(r)} = \frac{1}{2}(I_K^{(1)} - I^{(1)}_{N-K}) \tag{30}$$

For a voiced signal being received to by synthesized, this received spectrum magnitude will look like the sampled function in waveform (d) of FIG. 3. In waveform (d) of FIG. 3 is shown the plot of $M_K^{(r)}$ where K represents frequency. The highest frequency in the synthesized speech signal corresponds to $K = N/2$ and the lowest to $K = 0$.

In terms of real frequency, the corresponding real frequencies involved are given by

$$f_K = rK/N \tag{31}$$

where r is the sampled rate in samples per second. Since $M_K^{(r)}$ is a purely odd function of K, it will have a nonzero discrete sine transform H_n and an identically zero discrete cosine transform. Thus, the discrete Fourier transform of $M_K^{(r)}$ is called H_n. A typical H_n is shown in waveform (c) of FIG. 3. H_n is another of the outputs obtained from the operation of the FFT computer means 30. Since the H_n are the samples of the discrete sine transform of the received spectrum magnitude, H_n is an odd function of n. H_n is the impulse response which is used in the synthesis of the received speech signal to be discussed hereinafter.

Returning to the FFT Equation 28, it can be shown exactly how each of the inputs and outputs discussed above are obtained from the transform. As in Equations 13 and 16, the following identity is substituted:

$$e^{-j\frac{2\pi nK}{N}} = \cos\left(\frac{2\pi nK}{N}\right) - j\sin\left(\frac{2\pi nK}{N}\right) \tag{32}$$

also, Equation 29 is substituted into Equation (28) obtaining:

$$S_n^{(2)} = \sum_{K=0}^{N-1} \left\{ \begin{array}{l} R_K^{(1)} \cos\left(\frac{2\pi nK}{N}\right) - jR_K^{(1)} \sin\left(\frac{2\pi nK}{N}\right) \\ + jI_K^{(1)} \cos\left(\frac{2\pi nK}{N}\right) + I_K^{(1)} \sin\left(\frac{2\pi nK}{N}\right) \end{array} \right\} \tag{33}$$

At the end of the DFT or FFT operation, the output complex vector $S_n^{(2)}$ appears in two sets of N registers or memory locations each. One set of N registers contains the real part of $S_n^{(2)}$ given by

$$R_n^{(2)} = \sum_{K=0}^{N-1} \left\{ R_K^{(1)} \cos\frac{2\pi nK}{N} + I_K^{(1)} \sin\left(\frac{2\pi nK}{N}\right) \right\} \tag{34}$$

The other set of N registers contains the imaginary part of $S_n^{(2)}$ given by

$$I_n^{(2)} = \sum_{K=0}^{N-1} \left\{ - R_K^{(1)} \sin\left(\frac{2\pi nK}{N}\right) + I_K^{(1)} \cos\left(\frac{2\pi nK}{N}\right)\right\}$$

(35)

In each case the two sets of registers are numbered $n = 0, 1, 2, 3, \ldots, N-1$.

The even and odd separator section operates on $R_n^{(2)}$ to produce the even part of $R_n^{(2)}$

$$\frac{R_n^{(2)} + R^{(2)}_{-N1}}{2} = \sum_{K=0}^{N-1} R_K^{(1)} \cos\left(\frac{2\pi nK}{N}\right)$$

(36)

$= \mathcal{R}_n$ the real part of the spectrum of $R_K^{(1)}$ the analyzed speech waveform

and the odd part of $R_n^{(2)}$

$$\frac{R_n^{(2)} - R^{(2)}_{N-n}}{2} = \sum_{K=0}^{N-1} I_K^{(1)} \sin\left(\frac{2\pi nK}{N}\right)$$

(37)

H_n the impulse response to be synthesized

The even the odd separator section 34 operates similarly on $I_n^{(2)}$ to produce the even part of $I_n^{(2)}$

$$\frac{I_n^{(2)} + I^{(2)}_{N-n}}{2} = \sum_{K=0}^{N-1} I_K^{(1)} \cos\left(\frac{2\pi nK}{N}\right) = C_n \text{ the FTLSM}$$

(38)

and the odd part of $I_n^{(2)}$

$$\frac{I_n^{(2)} - I^{(2)}_{N-n}}{2} = \sum_{K=0}^{N-1} R_K^{(1)} \sin\left(\frac{2\pi nK}{N}\right)$$

(39)

$= \mathcal{G}_n$ the imaginary part of the spectrum of $R_K^{(1)}$ the analyzed speech waveform

Thus, there are basically four parts of the FFT computer means 30 output as given by Equations 36, 37, 38 and 39. Equations 36 and 39 give, respectively, the real and imaginary parts of the Fourier transform of the speech input waveform $R_K^{(1)}$ which is to be further processed, and information describing its spectrum magnitude envelope is to be transmitted. Equation 36 defines the transform H_n of the received signal to be used by a synthesizer as the vocal tract impulse response, and Equation 37 represents the C_n function (the FTLSM signal).

16

Logging Algorithm Computer

As stated hereinabove, the logging algorithm computer 44 must take the log of the frequency spectrum magnitudes from the magnitude approximation circuit 38. The logging algorithm computer can be any computer capable of solving the algorithm discussed hereinbelow (for example, the Sylvania Electric Products Inc. ACP-1 computer can be used). The log function can be approximated in several ways, one of which is by a set of $n-1$ inscribed straight lines, where n equals the number of bits in the binary integer word which describes the number to be logged. Thus if (as in the Sylvania Electric Products Inc. ACP-1 computer) one has an 11-bit magnitude to be logged, the log is approximated by ten inscribed straight lines.

The easiest way to explain the method is by way of an example. Suppose the integer to be logged is the number y, e.g.,

$$\begin{array}{ccccc ccccc cc}
v = 2^{10} & 2^9 & 2^8 & & 2^7 \, 2^6 \, 2^5 & & 2^4 & 2^3 & 2^2 & 2^1 & 2^0 \\
y = 0 & 0 & \boxed{1} & & \boxed{1 \ 0 \ 1} & & 0 & 1 & 1 & 1 & 0
\end{array}$$

Furthermore, suppose, for example, that a 7-bit logarithm is desired. First, the position of the most significant unit is found by a shift and test operation. Here this is in the exponent $= 8$ position. The 4-bit binary code for this exponent is generated and shifted to the left three binary places. In the three empty binary places, the next three most significant bits of y are simply inserted. The resulting logarithm is $1\ 0\ 0\ 0\ 1\ 0\ 1$, i.e.,

$$\underbrace{1\ 0\ 0\ 0}_{\text{binary code for 8}} \qquad \underbrace{\boxed{1\ 0\ 1}}_{\text{next 3 bits of } y}$$

The rationale behind the method is simple. The number to be logged in the example was

$$y = 2^8 [1 + x]$$

where

$$0 \le x \le 1$$

Therefore

$$\log_2 y = 8 + \log_2 [1 + x]$$

The method simply replaces $\log_2 [1 + x]$ by x and codes the result in binary. In general, then one forms the approximation

$$\log_2 2^v [1 + x] \approx v + x$$

where v is the exponent of the position where the most significant "one" appears. In most engineering applications, there will be no point in taking more than about three bits for x since this gives $1 + x$ to within a factor of $1 + \frac{1}{8}$ at worst and this corresponds to an error of 0.5 db.

The error obtained in the three least significant bits due to inscribed straight line approximation obtained by replacing $\log_2 [1 + x]$ by x may be seen in Table I.

65

TABLE I

LOGGING ALGORITHM VS. TRUE LOG

X	Code Used= Code for X	$\log_e [1+x]$	correct code	
0	0 0 0	0	0 0 0	
0.125	0 0 1	0.17	0 0 1	
0.250	0 1 0	0.32	0 1 1	
0.375	0 1 1	0.46	1 0 0	codes in
0.500	1 0 0	0.58	1 0 1	error by
0.625	1 0 1	0.70	1 1 0	one count
0.750	1 1 0	0.81	1 1 0	
0.875	1 1 1	0.91	1 1 1	

It can be seen from the table that the largest error made is just one count. Since the largest 7-bit logarithm generated for an 11-bit magnitude input is

$$\underline{1\ 0\ 1\ 0}1\ 1\ 1 = 87$$
$$[\text{code for }\overline{10}]$$

the 60 db dynamic range of this system is divided up into 60/87 = 0.7 db steps. Therefore an error of one count, as shown in the table, corresponds to an output error of 0.7 db. This error will occur on the average half the time and is additive to the worst case error of 0.5 db which could occur as a result of ignoring all bits beyond the third bit to the right of the most significant "one" in y. On the average, the error will be less than 1 db.

Plots of the function FTLSM when the log of the spectrum magnitude was generated by the above algorithm were essentially indistinguishable from those generated by a full accuracy log routine.

Encoding Means 40

The encoding means 40 combines information derived from the pitch detection logic unit 60 with the FTLSM signal C_n from the FFT computer means 30 to generate and encode the pitch and spectral information for transmission to some receiving terminal. The encoding unit 40 has two sections, the first of which is the pitch detection logic unit 60 which employs the spectral magnitude signal from the magnitude approximation circuit 38 and a second section section which uses the FTLSM signal C_n from the FFT computer means 30 to extract the pitch information, if any, from the speech signal. The remaining units of the encoding means 40 extract and encode the spectral envelope information from the FTLSM signal C_n for transmission to a receiver. (As was stated hereinabove, a speech signal can be represented by its excitation function (pitch signal) and the vocal tract transfer function (spectral envelope information).

The spectral information section of the encoding means 40 includes a denormalizing unit 62 having input connections from the C_n terminal of the FFT computer means 30, and the normalizing unit 16, and having an output connection to a scaling unit 64. The number of C_n terms processed by the encoding means 40 is some number K where K is less than the maximum value of n (i.e., 127). Typically K is some value in the range of 10 to 30 and is a function of the system sampling rate once the amount of real time that C_n should define for some optimum representation of the spectrum magnitude en-

velope has been determined.

The scaling unit 64 includes a plurality of scaling registers 66, such as an accumulator, connected to a scaling storage device, such as a memory 68, and a digital multiplier circuit 70. The multiplier circuit 70 also has a connection from the scaling memory 68. The multiplier circuit 70 has an output connection to a gating means 74. The gating means 74 includes a storage device, such as the register 16, connected to a gating matrix 78 which has a second input from a counter 80. The output signal of the gating matrix 78 is the coded spectral envelope to be transmitted.

The coded spectral envelope signal is obtained from the sampled C_n signal as follows. Assume the FTLSM data, as shown in waveform (c) of FIG. 3, is transferred to a storage register (not shown in the denormalizing unit 62). In the instant invention, only a predetermined number of the N samples of the FTLSM signal C_n are selected to characterize the spectrum of the analyzed speech signal. For example, at an input sampling rate of 6.4 KHz, the first 21 (K=20) C_n coefficients are used to describe the magnitude of the spectrum envelope. The amount of time spanned by K coefficients of C_n, at a given sampling rate (SR), is equal to the ratio K/SR.

To preserve information about the original input speech level in the transmitted data, the normalization gain factor supplied by the normalization unit 16 must be removed from C_0, which is the $n = 0$ sample of the FTLSM function (C_n). This is done by the denormalization unit 62. Since the application of the normalization gain affects all the speech samples in an analysis interval, this normalization gain only affects the spectrum amplitude and not the spectrum envelope shape. An increase in the input samples therefore only affects the average value of the log of the spectrum magnitude signal. This average value change is only reflected in the C_n function as a corresponding change in amplitude of the DC component in the function C_n or the C_0 term. The C_0 value ($\overline{C}_0$) that would have been calculated if no normalization gain was included can be computed from the following relation

$$\overline{C}_0 = C_0 - 16 \sqrt{N} \log_2 (G_N) \tag{40}$$

where G_n is the value of normalization gain supplied by the normalization unit 16. The denormalizing unit 62 can be a small computer capable of solving Equation 40. In practice the FFT computer means 30 could perform the function of the denormalizing unit and transmit to the encoding means 40 a corrected value of the C_0 coefficient of the C_n signal.

To achieve as much accuracy as possible in the transmission of the lower 21 FTLSM coefficients C_0 through C_{20}, it is necessary to determine the peak-to-peak range of variation of each coefficient. Computer studies were performed to determine experimentally the range of variation for each of the selected FTLSM coefficients C_n using fifteen test sentences as input speech signals. It was found during these tests that some of the 21 FTLSM coefficients do not vary over symmetric ranges. This fact implies that an average shape exists for the FTLSM. As a result, each coefficient can be defined as having a peak-to-peak range about some average level. Therefore, by adding an experimentally determined bias level (where the bias for each channel is fixed) to each coefficient, the range of variation can

be made to exist primarily for positive values only. The value of the bias constant determines the probability with which each one of the C_n samples to be quantized exceeds the allowable range for the particular C_n coefficient. In the case of a C_n coefficient which falls outside the peak-to-peak range allowable, the system will truncate that C_n coefficient to the closest allowable value (i.e., the maximum positive or negative value).

The total positive range of variation of each coefficient (also a known experimental quantity) is employed in conjunction with the bias level to scale and quantize each FTLSM coefficient. In this regard, the quantized value of the jth FTLSM coefficient, $\tilde{C}_j$, is computed by the relationship

$$\tilde{C}_j = \frac{C_j + (\text{Bias})_j}{Q_j} \quad j = 0, 1, 2, \ldots, 20 \quad (41)$$

where Q_j is the size of the quantum step in the jth channel. For a peak-to-peak range for the jth channel given by P_j and assuming b_j bits are assigned to the jth channel, Q_j is given by

$$Q_j = \frac{P_j}{(2_j^{b} - 1)} \quad (42)$$

Table II presents typical values of the bias levels and peak-to-peak ranges for 21 FTLSM coefficients. Additional data in the table (discussed subsequently) is the bit assignment across the 21 FTLSM coefficients and the interleaved coefficients that are updated every other frame.

TABLE II

MAXIMUM VARIATION RANGE, BIAS INSERTION LEVEL AND CHANNEL BIT ASSIGNMENTS FOR TRANSMITTED FTLSM COEFFICIENTS

channel number J	peak-to-peak range	insertion bias (BIAS)$_j$	bit assignments b_j	
0	1500	0	5	First five
1	820	400	4	coeffi-
2	510	200	4	cients.
3	575	200	4	Updated every 20 msec
4	530	260	4	↓
5	430	185	3	
6	390	225	3	
7	390	225	3	
8	450	330	3	
9	320	165	3	Interleaved
10	300	185	3	channels.
11	225	120	2	Each set updated
12	245	140	2	every
13	225	120	2	40 msec
14	245	120	2	
15	225	120	2	
16	205	100	2	
17	205	100	2	
18	185	100	2	
19	205	100	2	
20	205	100	2	

notes		
Total number of channel bits		40
Bits used to specify interleaved set		1
Bits used for pitch word		7
Total number of bits/frame		48

Twenty-one FTLSM C_n coefficients (including the denormalized C_0 coefficient) are transferred to the scaling registers 66 where the scaling factors, Bias$_j$, from the memory 68 are added to their respective C_j coefficients to adjust the range of the numbers being quantized such that they go from zero to some positive maximum. The scaled C_n coefficients, C_0 through C_{20}, are then directed to the multiplier 70 where each coefficient is multiplied by a separate predetermined ratio from the memory 68. (The combination of the scaling registers 66, memory 68 and multiplier 70 comprise in effect a quantizer circuit.) The predetermined ratios stored in the memory 68 are given by Equation 43.

$$Q_{(j)} = (2^{b_j} - 1)/P \quad (43)$$

which is seen to be the reciprocal of Equation 42. For example, the ratio of the 0 channel of Table II is equal to $(2^5 - 1)/1500$. By multiplying each FTLSM coefficient C_0 through C_{20} by its respective ratio, the quantizer guarantees that the quantized value of each FTLSM coefficient never exceeds the highest number that its particular bit assignment can represent. For example, if a value of 150 were calculated for the C_2 coefficient, the quantized value would be calculated in accordance with the equation

$$C_2 = (C_2 + BIAS_2) \cdot Q(2) \quad (43\text{-A})$$

$$= (150 + 200)(2^4 - 1/510) = 350(15/510) = 10$$

The value 10 can be represented by a four bit binary number and therefore, since $b_2 = 4$ the amplitude value of $C_2(150)$, falls within the quantizer range and is representable as the integer value 10.

The denormalized, scaled and quantized $\tilde{C}_j$ coefficients are then directed to the storage register 76 of the gating means 74. It is appreciated that there are many well-known species of gating means 74 that could be employed, for example, those employed in telemetry systems for varying the number of times a particular channel is sampled. Still another way of implementing the gating means would be to have the data shifted out of register 76 under program control. Another technique (the one chosen for illustration in FIG. 2B) is to having a gating matrix 78, such as the type employed in time division telemetry systems, connected to the registers 76. A second input to the gating matrix 78 would be provided by the counter 80 which would supply a gating pulse to the appropriate gate in the matrix to thereby gate through the $\tilde{C}_j$ data stored in register 76 in accordince with Table II (i.e., the interleaving operation on the upper $\tilde{C}_j$ values would be performed).

Pitch Detection Logic

One embodiment of the pitch detection logic is shown in block diagram form in FIG. 4 and includes a first storage register 90 having input connections from the first magnitude approximation circuit 38 and a gating circuit 92. A second magnitude approximation circuit 94 has an input connection from the first storage register 90 and first and second output connections to an energy ratio detector 96 (to be discussed in detail hereinafter). A second storage register 98 has input connections from the C_n terminal of the FFT computer means 30 and a maximum selection circuit 100 and

output connection to a comparator circuit 102, well-known in the art, and a buffer storage register 104. The comparator circuit 102 has first and second output connections, respectively, to the energy ratio detector 96 and to a third gating circuit, such as the OR circuit 106, the output of which is connected to a first flip-flop circuit 108.

A fourth gating circuit, such as the OR gate 110, has first and second inputs from the AND gate 104 and the flip-flop circuit 108, respectively, and an output connection to a third storage device, such as a second buffer memory 112. Connected between a second output of the energy ratio detector 96 is a second flip-flop circuit 114, the output of which is connected to the buffer storage register 104.

The pitch detection logic has as its inputs the nth interval of spectrum magnitudes (128 samples) and the $n-1$ interval FTLSM coefficients, each of which is stored in its respective registers 90 and 98. The low-band and high-band energy values $(E_L)_n$ and $(E_H)_n$ are computed directly from the spectrum magnitude samples by the second magnitude approximation computer 94. The gating circuit 92, in response to a control signal from a source not shown, gates two sets of a predetermined number of the spectrum magnitude samples into the magnitude approximation computer 94. To determine the low-band energy, samples 6 through 32 (covering approximately a total bandwidth of 170 to 900 Hz at 7,200 Hz sampling rate) are gated out of register 90 and are directed to the magnitude approximation computer 94. Similarly, to determine the high-band energy, samples 99 through 125 (covering approximately a total bandwidth of 2,800 to 3,500 Hz) are employed. The second magnitude approximation computer 94 can be any well-known computer capable of solving the following relation

$$E_{(L\ or\ H)} = \tfrac{3}{4} \max\ (Si) + \tfrac{1}{4}\Sigma Si \qquad (44)$$

where i equals 6 through 32 for E_L and 99 through 125 for E_H, and the values Si are obtained as the output spectrum magnitude samples from the first magnitude approximation means 38. The results of solving the Equation 44 give two numerical values, one for low-band energy and one for high-band energy, both of which are directed to the energy ratio detector 96, to be discussed in detail hereinafter.

The $N/2$ C_n coefficients stored in the second storage device 98 are scanned by the maximum selection circuit 100 to determine the magnitude and position of the largest C_n coefficient in a predetermined range (i.e., $20 \leq n \leq 115$). The maximum selection circuit 100 can also be a two-stage comparator which initially stores the magnitude and position of the C_{20} value and thereafter compares the C_{20} magnitude sequentially with the remaining C_n coefficients (in the range of n) until it finds a larger magnitude. When a larger magnitude is found, the new magnitude and position become the reference to test C_n magnitudes against. This type of operation can also be programmed on a general purpose computer with well-known techniques; in fact, the FFT computer can perform the function.

The magnitude of the largest C_n coefficient found by the maximum selection circuit 100 is directed to the comparator circuit 102 where it is compared with a predetermined low threshold (LOWTHR) value (for

example, 30) which can be determined emperically. If the magnitude of the largest of the C_n coefficients is less than the predetermined LOWTHR value, an output signal is directed from the terminal 103 of the comparator circuit 102 through the first OR circuit 106 to cause the first flip-flop circuit 108 to change its state. The output signal from the first flip-flop circuit 108 is directed to the second OR circuit 110 to a particular memory location in the second buffer memory 112. If the peak C_n sample in the search range for an analysis interval is less than the LOWTHR value, an unvoiced (UV) condition exists. (An unvoiced condition is the absence of a pitch excitation signal.)

If the magnitude of one of the C_n coefficients is greater than the LOWTHR value for some analysis intervals, then an output signal from the 105 terminal is directed to the energy ratio detector 96 where a second test is made to determine voice (v) and unvoiced sounds. To minimize voicing errors during certain types of sounds that yield a sufficiently large C_n peak to pass the comparator test 102 but do not have the normal energy distribution of low-band and high-band energy or sufficient absolute low-band energy, the energy ratio detector 96 (to be discussed in detail hereinafter) checks the energy ratio of low-band energy to high-band energy and the absolute value of the low-band energies for three consecutive analysis intervals. If the energy ratio detector tests are not satisfied, then the unvoiced output signal appears at the output terminal 95 and is directed through the OR circuit 106 indicating an unvoiced sound.

If on the other hand the threshold values of energy ratio detector 96 are satisfied, an output signal is directed to the second flip-flop circuit 114 from terminal 97. The second flip-flop circuit 114 changes its state causing the signal representing magnitude and position of the largest C_n coefficient to be passed by the AND gate 104 through the OR gate 110 to the second buffer memory 112. This particular sequence indicates a voiced condition, with the voiced pitch signal indicated by the position of the C_n coefficient stored in the second buffer memory 112.

Shown in FIG. 5 is a block diagram of one embodiment of an energy ratio detector that can be employed in the pitch detection logic of FIG. 4. In the normal operation of the pitch detection logic, a delay equivalent to two analysis intervals exists before the pitch detection logic will start transmitting nonzero pitch values assuming data put into the pitch detection logic was a voiced sound. This delay is due to the low-band, high-band energy ratio detector which requires three consecutive low-band, high-band energy values that must satisfy the threshold requirements before voicing can occur.

The energy ratio detector of FIG. 5 includes two channels, a low-band channel 120 and a high-band channel 122. The low-band channel 120 includes three storage means, for example, registers 124, 125 and 126, which store the magnitude of the low frequency energy for three successive analysis intervals, n, $n-1$ and $n-2$ respectively. (While three analysis intervals were chosen, it is obvious that more or less than three can be employed.) Connected to one output of respective low energy resistors 124, 125 and 126 is a first set of comparators 128, 130 and 132, each of which has as

23

a second input a connection from a first storage device 134. The output of each of the comparators 128, 130 and 132 is connected to the input side of a first gating means, for example, the AND gate 136.

The high-band energy channel 122 includes three registers 140, 142 and 144 which store the magnitude of the high-band energy for the successive analysis intervals n, $n-1$ and $n-2$ respectively. Three inverter circuits 146, 148 and 150 have input connections from respective registers 140, 142 and 144 and output connections to three multiplier circuits 152, 154 and 156, respectively. A second input connection to the multiplier circuits 152, 154 and 156 originates at respective low-band energy registers 124, 125 and 126. A second set of comparators 158, 160 and 162 has a first input connection from respective multiplier circuits 152, 154 and 156 and a second input connection from a second storage device 164. (While two storage devices are shown, it is obvious that only one may be employed.) A second gating means 166 has input connections from each of the second set of comparators 158, 160 and 162. A third comparator means 168, for example a two input AND gate, has input connections from each of the AND circuits 136 and 166 and first and second output connections to the OR circuit 106 and the flip-flop circuit 114, respectively, of the pitch detection logic.

In operation, the magnitudes of the low-band energy signal and the high-band energy signal for the nth analysis interval are directed to respective registers 124 and 140. It is to be appreciated that after each analysis interval n, the data in register 124 is shifted sequentially through the registers 125 and 126 and similarly the data in the high-band energy register is shifted sequentially through registers 142 and 144. For example, after the nth analysis, the data in register 124 is shifted into register 125, and the data in register 125 is shifted into register 126. (Note, for simplicity the connection lines between these registers are not shown.) The comparators 128, 130 and 132 compare the low-band energy value in their respective registers 124, 125 and 126 with a predetermined value 30 from the first storage device 134. When the magnitude of the signals from each register 124, 125 and 126 exceeds the predetermined value stored in the storage device 134 and a signal is received from the first comparator 102 of FIG. 4, the gating circuit 36 produces a first output signal, for example, a positive pulse. If one or all of the signals stored in the low energy resistors 124, 125 and 126 is less than the predetermined reference (from the storage device 134), then the second predetermined output signal, for example, a negative pulse, is generated.

The object of the second channel 122 of the energy ratio detector 96 is to compare the ratio of the high-band and low-band energies with a predetermined ratio stored in the storage device 164. It has been found experimentally that for a proper varied sound the ratio of the low-band and high-band energy exceeds a certain value, for example, four. To obtain the ratio, the magnitude of the high-band energy signals for the n, $n-1$ and $n-2$ intervals stored in registers 140, 142 and 144 are directed through respective inverters 146, 148 and 150 to one input of respective multipliers 152, 154 and 156. A second input signal to the multipliers 152, 154

24

and 156 originates from the respective low-band energy registers 124, 125 and 126, and the resultant products are directed to respective comparators 158, 160 and 162 where they are compared to the predetermined constant from the storage device 164.

If the products (ratios) from all the multiplier circuits exceed the predetermined number stored in the storage device 164, then the gating circuit 166 delivers a predetermined signal, for example, a positive pulse to the comparator circuit 168. If one or all of the products (ratios) is less than the predetermined value, then the gating circuit 166 delivers a negative pulse to the comparator 168. The comparator circuit 168 produces an output signal at terminal 169 only when both signals at its input terminals are positive and produce a signal at its output terminal 167 under all other conditions. The gating circuits 136 and 166 and the comparator circuit 168 are simple logic circuits and can be assembled by any person having ordinary skill in the art of designing logic circuits. (Another technique for performing the logic functions specified hereinabove is by programming a general purpose computer, in accordance with well-known techniques.)

In addition to implementing the pitch detection logic 60 of FIG. 2B by the embodiment of FIGS. 4 and 5, a special purpose computer, for example, the Sylvania Electric Products Inc. ACP-1 computer, can be programmed in accordance with the flow charts of FIGS. 6, 7, 8, 9A and 9B.

The block labeled "PITCH DETECTION LOGIC" in FIG. 2B includes the low-band, high-band unvoiced to voiced (UV-V) and voiced to unvoiced (V-UV) detectors and pitch logic. A block diagram of the overall pitch detection function is shown in FIGS. 6 through 9A. The symbols used inside the blocks shown in these figures defined in Table III. Numerical values for the threshold levels K_1, K_2, K_3 and K_4 used in the UV-V and V-UV detections operations are given in Table IV.

TABLE III

GLOSSARY OF TERMS USED IN PITCH DETECTOR BLOCK DIAGRAMS (FIGS. 6 THROUGH 9A)

$(E_L)_{n-2}$ $(E_L)_{n-1}$ $(E_L)_n$	= Low-band energy value for $(n-2)$, $(n-1)$ and n analysis intervals
$(E_H)_{n-2}$ $(E_H)_{n-1}$	= High-band energy value for $(n-2)$, $(n-1)$ and n analysis intervals
q_{n-3}	= Value of pitch actually transmitted three analysis intervals ago
q_{n-2}	= Value of pitch that is transmitted during during n^{th} analysis interval. (The pitch value $(\tau_1)_{n-1}$ for the $(n-1)^{st}$ interval is checked for tracking relative to q_{n-2} before deciding a value for q_{n-1}.)
q_{n-1}	= Value of pitch for $(n-1)^{st}$ FTLSM. This value is being determined during the nth analysis interval and is transmitted during the $(n+1)$ analysis interim.
$(\tau_1)_{n-1}$	= Position of primary FTLSM peak in search range of $(n-1)^{st}$ FTLSM
$(\tau'_1)_{n-1}$	= Position of peak that is tracking within ±1 msec of previous pitch value $(\tau_{1\ n-2})$
$(P_1)_{n-1}$	= Value of FTLSM peak at $(\tau_{1\ n-1})$
$(P'_1)_{n-1}$	= Value of FTLSM peak at $(\tau'_1)_{n-1}$
FLAG	= Computation variable

TABLE IV

THRESHOLD VALUES USED IN UV-V AND V-UV DETECTION OPERATIONS IN PITCH LOGIC OF FIGURES 6 THROUGH 9A

Threshold	K value using simple sum of low-band or high-band samples
K_1	3000
K_2	1000
K_3	1000
K_4	1000

A complete flow chart of the logical functions performed by the pitch detector in extracting fundamental pitch is presented in FIGS. 6 through 9B. An outline of the logical operations performed in each of the figures is given below.

FIG. 6: UV-V detection (Condition ϕ) and continue voicing condition (Condition ϕ')

FIG. 7: Initial pitch tracking following a UV-V boundary (Condition ϕ satisfied)

FIG. 8: Extension of voicing one analysis interval (Conditions ϕ' and ϕ'' not satisfied)

FIG. 9A: Test for voicing mode following failure to pitch track and reestablishment of pitch tracking

FIG. 9B: Pitch tracking logic during steady voicing (including negation of tracking branch)

The initial operations performed by the pitch logic are diagrammed in FIG. 6 and include the nth spectrum and $(n-1)$ FTLSM functions (which appear simultaneously at the FFT output) directed to the pitch detector logic. As a first step in the process, the low-band and high-band energy values $(E_L)_n$ and $(E_H)_n$ are computed directly from spectrum magnitude samples that correspond to the nth analysis interval. The relations used to compute E_L and E_H are given in Equation 44. This relation gives values of E_L and E_H that are proportional to the square root of the energy contained in the low-band and high-band regions of the speech spectrum respectively. Following this operation, the $(n-1)$ C_n function is scanned to find the amplitude and position of the largest peak in the search range. This peak and its position are designated as $(P_1, \tau_1)_{n-1}$.

Energy values computed for the previous two intervals, namely $(E_L)_{n-1}$, $(E_H)_{n-1}$, $(E_L)_{n-2}$, and $(E_H)_{n-2}$, are retained in storage together with the position of the primary peak found for the $n-2^{nd}$ C_n function. This value is designated as $(\tau_1)_{n-2}$. The only additional data that may be necessary occurs during the pitch tracking mode (FIG. 9B) and involves a search through the $(n-1)C_n$ function for a peak amplitude and its position in the vicinity of $(\tau_1)_{n-2}$. This secondary peak and its position are designated as $(P'_1, \tau'_1)_{n-1}$ in FIG. 9B.

As an aid to the following discussion, it is convenient to summarize the initial operations and subsequent branch functions performed by the logic of FIG. 6 in terms of the following five conditions and the branch functions that are performed corresponding to each condition:

Condition		Branch Function Performed
1. $q_{n-3} = 0$;	Condition ϕ not satisfied	Continue unvoiced (FIG. 6)
2. $q_{n-3} = 0$;	Condition ϕ	Initiate initial tracking
3. $q_{n-3} \ne 0$;	Condition ϕ' not satisfied satisfied	mode (FIG. 7) Extend voicing mode (FIG. 8)
4. $q_{n-3} \ne 0$;	Condition ϕ' satisfied	Test for continued voicing (FIG. 9A)
5. $q_{n-3} \ne 0$; Flag $= 0$	Condition ϕ' satisfied	Steady voicing. Test for pitch tracking. (FIG. 9B)
Flag $= 1$		

Conditions ϕ and ϕ' are threshold tests involving the low-band and high-band energy measures and are defined in the block diagram of FIG. 6. In the present system, during the nth analysis interval, the following data is available:

1. Whether q_{n-3} is zero or not
2. $q_{n-2} = (\tau_1)_{n-2}$ or $(\tau'_1)_{n-2}$ and $(\tau_1)_{n-1}$
3. $(E_L)_{n-2}$, $(E_H)_{n-2}$, $(E_L)_{n-1}$, $(E_H)_{n-1}$, $(E_L)_n$, $(E_H)_n$

With reference to the nth analysis interval, q_{n-3} was the previous value of pitch actually transmitted, q_{n-2} is the value of pitch to br transmitted and $(\tau_1)_{n-1}$ is the position of the largest peak in the $(n-1)$ C_n function. The value q_{n-1}, which is transmitted during the $(n+1)$ analysis interval, may or may not be equal to $(\tau_1)_{n-1}$. This depends primarily upon whether or not $(\tau_1)_{n-1}$ tracks q_{n-2} and is discussed subsequently with reference to Condition 5 listed above.

It should be noted that in the present system $q_{n-3} = 0$ does not always imply that the $(n-3)$ analysis interval was called unvoiced and a zero value for pitch was actually transmitted. A nonzero value could actually have been transmitted but conditions were such that the $(n-3)$ interval was tagged as unvoiced within one of the branch operations. This situation can occur in the branch operation shown in FIG. 9A when conditions ϕ and ϕ''' are not satisfied.

Referring to Conditions 1 through 5 listed above, it is seen that a different course of action is taken depending upon whether q_{n-3} was zero or not. In particular, Condition 1 results in continued unvoicing since the previous interval was unvoiced and the low-band and high-band energies, E_L and E_H, are too small and too large respectively to satisfy condition ϕ.

Condition 2 generally occurs at a UV-V boundary where the previous interval was tagged as unvoiced (i.e., $q_{n-3} = 0$) but now condition ϕ is satisfied. When this occurs, an initial FTLSM peak tracking mode is initiated during the n analysis interval which is detailed in the block diagram of FIG. 7. The operations performed in this figure are self-explanatory.

Condition 3 generally occurs during the trailing off of voicing where the low-band energies are starting to diminish. In this regard, condition ϕ', which involves only two energy measures $[(E_L)_n$ and $(E_L)_{n-1}]$ and one energy ratio $[(E_L/E_H)_{n-1}]$, is clearly an easier condition to satisfy during trailing off of voicing than is condition ϕ. This less stringent condition is incorporated to ensure that pitch tracking can continue well into the trailing off of voicing. When condition 3 is present, the logic branch shown in FIG. 8 is actuated. Voicing may or may not be extended one analysis interval at this point, depending upon whether or not a secondary condition (condition ϕ'') involving the energy ratio $(E_L/E_H)_{n-1}$ is satisfied (see FIG. 8).

Condition 4 occurs during the voicing mode whenever pitch tracking of the primary FTLSM peak $(p_1, \tau_1)_{n1}$ fails to occur and the amplitude P'_1 of the tracking peak $(P'_1, \tau'_1)_{n-1}$ is less than one-fourth of the

primary peak P_1. Under this condition, Flag is set to zero and Condition 4 results in actuating the logical branch shown in FIG. 9A. The most stringent condition ϕ is applied at this point to determine if the vocoder is in a steady voicing mode. Pitch tracking can be reestablished within this branch as indicated in FIG. 9A if both condition ϕ is satisfied and the FTLSM peak again tracks the peak that failed to track one analysis interval ago. Occasionally, two or more analysis intervals are required to reestablish the pitch tracking mode. When condition ϕ is not satisfied, assuming condition 4 has occurred, a fourth energy condition test is performed (condition ϕ''' of FIG. 9A) that involves only the high-band energy measure for the $n-2$ analysis interval, $(E_H)_{n-2}$. If this value is found to be larger than the preset threshold level, K_4, in this case, the decision is made to call the $(n-2)$ interval unvoiced and then to transmit $q_{n-2} = 0$. On the other hand, if the high-band energy measure $(E_H)_{n-2}$ is less than K_4, the decision is made that q_{n-2} was a bona fide pitch value and it is transmitted. In either event, q_{n-1} is set to zero whether or not condition ϕ''' was satisfied since condition ϕ itself was not satisfied. Setting $q_{n-1} = 0$ in this case, in effect, is to assume that a UV-V boundary has occurred due to the failure to satisfy condition ϕ.

Condition 5 is referred to as the steady voicing mode. The branch that is actuated during this mode is shown in FIG. 9B. In normal, steady-state voicing, pitch tracking generally occurs normally, and the logical operations performed by the pitch detector are confined for the most part to this branch. The logic is straightforward with the possible exception of the negation of the pitch tracking subbranch. Negation of pitch tracking occurs when the largest peak $(P_1, \tau_1)_{n-1}$ fails to track $(\tau_1)_{n-2}$ and the amplitude of the peak in the vicinity of ± 1 msec of $(\tau_1)_{n-2}$, P'_1, fails to exceed one-fourth of the largest peak P_1. When this occurs, Flag is set to zero and the value q_{n-2} is transmitted. Then q_{n-1} is set to equal to $(\tau_1)_{n-1}$ which is the position value of the largest peak in the data. Tracking of $(\tau_1)_{n-1}$ relative to $(\tau_1)_{n-2}$ has thus been negated at this point. Tracking can again be reestablished, however, during the next or a later analysis interval by means of the logic branch shown in FIG. 9A.

Synthesis Section of Vocoder

The synthesis section of the vocoder system has basically three functions to perform: (1) to obtain the vocal tract response function from the FTLSM data, (2) to obtain the voicing data from the FTLSM and (3) to convolve the vocal tract response with voicing data to obtain the desired synthesized speech signal. One embodiment of a device for performing the above-recited functions is given in respective FIGS. 10A, 10B, 11, 12 and 14.

Referring to FIGS. 10A and 10B, the embodiment of the device to obtain the vocal tract response function from the received FTLSM data includes a decoding device 100 (to be discussed in detail hereinafter) which dequantizes, descales and interlaces one set of 13 received FTLSM coefficients C_n with a set of eight interlaced values previously sent to form a set of 21 C_n coefficients to be processed. Connected to the output of the decoding means 100 is a spectrum decoder means 102, for example, a DFT computer such as the

Sylvania Electric Products Inc. ACP-1 computer. The output of the spectrum decoder means 102 is connected to a delogging computer 104, the output of which is connected to an unvoiced modifying unit 106. The unvoiced modifying unit 106 has a second input connection originating at the decoding device 100 and an output connection connected to an odd function generator 108. (The second input to the summation circuit 48 of FIG. 2A originates at the odd function generator 108.)

The input data to the decoding device 100 is in the same form as the output data of the processor unit 40 of FIG. 2B as calculated per Equation 41. The spectral envelope data is contained in the scaled and quantized C_n coefficients received at the input to decoding device 100. The particular functions performed by the decoding device are then to dequantize, descale and to interleave two sets of 13 received coefficients into sets of 21 coefficients. (The details of the decoding device 100 will be discussed in detail hereinafter.) The C_n coefficients (received FTLSM signals) thus obtained are applied to the spectrum data decoder 102 where they are Fourier transformed to obtain the spectral magnitude of the analyzed speech signal. Any well-known special purpose computer can be employed as the spectrum data decoder 102, such as the Sylvania Electric Products Inc. ACP-1 computer referred to hereinabove. The spectrum data decoder 102 must solve the equation

$$\log|V_n| = \sum_{K=0}^{20} C_k \cos\left(\frac{2\pi n K}{N}\right) \tag{45}$$

where

$V_n = n^{th}$ frequency sample
$C_k = k^{th}$ sample of the C_n function
$N =$ effective size of the transform

For example, n may run from 4 to 128 in steps of four, i.e., 32 steps of data points. Therefore the log $|V_n|$ will be evaluated at 32 discrete frequencies.

The logged samples are then directed to a delogging computer 104, such as the Sylvania Electric Products Inc. ACP-1 computer, which in essence performs the inverse of the algorithm solved by the logging algorithm computer 44 of FIG. 2B and described in detail hereinabove. The output signal from delogging computer 104 includes 32 samples of spectral magnitudes of the received signal which are directed through the unvoiced modifying unit 106. The data will pass through the unvoiced modifying unit 106 unaltered if the data is a result of a voiced sound and will be modified in a random manner if the data results from an unvoiced sound. To be compatible with the imaginary input, the 32 delogged samples are directed through the odd function generator 108 where they are converted into 256 samples. The odd function generator inserts the 32 samples into every fourth position of a 256 word storage means 184. The original 32 samples when placed into every fourth storage word of storage means 184 occupies the lower 128 storage locations. At the same time the samples are being inserted into every fourth position, the negative of the sample values

obtained by negating sample values using the negating circuit **186** are inserted in corresponding positions which form an odd symmetric function about $N/2$ (128).

Decoding Device

The decoding device **100** includes first and second buffer storage devices **120** and **122** respectively. A multiplier circuit **124** has input connections from the buffer storage device **120** and from one section of a two section memory **126** and an output connection to a subtractor circuit **128** which is well-known in the art. A second input to the subtractor circuit **128** originates at a second section of two section memory **126**, and the output is connected to an interlacing means **130** which converts the incoming 13 $\bar{C}_n$ samples to 21 C_n samples.

The interlacing means **130** includes four input gates, such as a high gate **132**, a low gate **134**, a negative gate **136** and a positive gate **138**, connected to the output of the subtractor circuit **128**. A second input connection to the negative gate **136** and the positive gate **138** is the output connection of the low gate **134**, and the outputs from the respective gates are connected to a plurality of separate locations in a first storage device **140**. The high gate **132** has a second input connection from a central source (not shown) and an output connection to a second storage device **142**. The first and second storage devices **140** and **142** are connected together and represent the output connection of interlacing means **130** and the decoding device **100**.

The buffer storage **122** (which stores the seven bit pitch information) is connected to the input of a seven input OR gate **150**. The output of the OR gate **150** provides a second input to the unvoiced modifying unit **106**. If there is a "one" in any of the bit locations of the buffer storage **122** (indicating a voiced sound), then a signal is directed to the unvoiced modifying unit **106** which passes the spectral data from the delogging computer **104** to the odd function generator **108** unmodified. However, if there is no "one" in any of the seven bit positions (indicating an unvoiced sound) of the buffer storage **122**, then the unvoiced modifying unit **106** modifies the spectral data in accordance with a manner to be discussed hereinafter.

Recalling that the output signal of the processor unit **40** is a 48 bit data word wherein seven bits represent pitch information (voiced or unvoiced information), 40 bits represent the vocal tract information (spectral information represented by the $\bar{C}_n$ coefficients) and one bit indicating whether the even or odd set of the $\bar{C}_n$ values is encoded in the present frame. The functions of the decoding device **100** are then to dequantize, descale and generate 21 coefficients from the 41 bits of the vocal tract information.

A descaling unit **125** includes the combination of the multiplier circuit **124**, subtractor circuit **128** and the two section memory **126** and performs the reverse functions of the scaling unit **64** of FIG. 2B. For example, each received coefficient is multiplied in the multiplier circuit **124** by a memory location in the memory **126**. This predetermined ratio is the inverse of the ratio as defined by Equation 43. The output signals of the multiplier circuit **124** are the dequantized 13 C_n coefficients. Recalling that the C_n coefficients prior to transmission were scaled in the positive direction to insure

that no negative values were transmitted, to reconstruct the true coefficient the analyzer section of the vocoder must remove the scaling factor. The substractor circuit **128** subtracts the appropriate scaling factor (which is stored in the memory **126**) from each dequantized coefficient received from the multiplier circuit **124**. The details of the scaling factor are discussed hereinabove.

The coefficients generated in the analyzer section of the vocoder were quantized independently in accordance with Table II such that sets of 13 coefficients C_n were transmitted during each data interval as stated hereinabove. The first five coefficients C_0 through C_4 are transmitted every data interval while the higher even and odd numbered coefficients are sent during alternate data frames. For example, in one time interval, coefficients C_0 through C_4 and C_6, C_8 . . . C_{20} are sent and the next data interval C_0 through C_4 and C_5 . . . C_{19} are sent. It has been emperically determined that approximately 3 msec of the C_n function are required to adequately specify the spectrum envelope information. Therefore at the input to the spectrum decoder means **102**, the even and odd sets of C_n coefficients from two adjacent analysis intervals are used to specify the C_n values for $n = 5, 6, . . . 20$, in conjunction with the C_0 –C_4 samples of the most recent frame.

The interlacing means **130** accomplishes this interlacing in the following way. The first five coefficients C_0 through C_4 represented by 21 bits are directed through the high gate **132** to the storage device **142**. The control pulse from a source (while the control source is not shown, it is to be appreciated that control circuits to perform the functions recited herein are within the knowledge of one skilled in the art) is removed thereby closing the high gate **132**, and a control pulse is applied to the low gate **134** to pass a single bit, the polarity (one or zero) of which activates the appropriate gate, the negative gate **136** or the positive gate **138**. (The polarity of one of the 41 bits indicates whether the even numbered or odd numbered coefficients are being received in the present frame as indicated in Table II.)

For example, a one in the particular bit location is directed through the low gate **134** and opens the positive gate **138**. The remaining 19 bits representing, for example, the even numbered C_n coefficients are then written into predetermined memory locations of storage device **140**. Assuming there are 16 C_n coefficients in the storage device **140** from the pervious data intervals, the new even numbered C_n coefficients write over the old numbered coefficients previously stored. The five lower C_n coefficients C_0 through C_4 stored in the storage device **142** are then combined with the 16 coefficients in the storage device **140** at point **141** to yield the desired 21 descaled, dequantized coefficients.

Recalling that the C_n function is the FTLSM of the input speech and that the log operation separates the source information from the vocal tract information, the function of the spectrum data decoder **120** is to generate the received logged spectrum envelope from the first 21 C_n coefficients. The process of selecting the low delay values of the FTLSM is the equivalent of low-pass filtering the logged spectrum magnitude. The spectrum data decoder **102** implements Equation 45 which yields the log of the spectrum envelope. THe

logged data is then directed through a delogging computer which solves the logging algorithm (discussed hereinabove) in an inverse manner. Both Equation 45 and the delogging operation can be implemented on a computer such as the Sylvania Electric Products Inc. ACP-1 computer using well-known programming techniques.

Unvoiced Modifying Unit

During the synthesis of unvoiced sounds, the excitation function should ideally be the digital equivalent of a random noise impulse carrier in which the impulse areas are ±1 with equal probability and the frequency of the noise carrier is equal to the output sampling rate.

A convolution implemented system with a high data noise carrier would preserve the vocal tract impulse response spectrum shape; however, the processing time would be prohibitive in a real time digital processor. The prohibitive processing time is due to the fact that, for voicing, the maximum excitation rate is approximately 3.3 milliseconds at a typical sampling rate of 7,200 Hz. The voiced evaluation has pulses spaced approximately every 22 samples.

For the method of unvoiced carrier generation just described, there would be carrier pulses each sampling time or an increase of 22 times the amount of computation required for unvoiced synthesis vs. voiced synthesis. An alternative method of preserving the impulse response spectral magnitude shape would be to implement the high data rate time domain convolution with a frequency domain multiplication. In effect, the high data rate digital noise carrier pulses are replaced with, for example, 32 multiplications in the frequency domain.

This multiplication in the frequency domain is implemented by the unvoiced modifying unit 106. The output signal of the delogging computer means 104 is supplied as a first input signal to a digital multiplier 105. A second input to the digital multiplier 105 is the output connection of a gate 107, such as an AND gate which has a first input connection coupled to the seven input OR gate 150 of FIG. 10A and a random generator means 109, for example, the lower bit of the A/D converter 12 of FIG. 2A.

When the output signal of the OR gate 150 is comprised of all zeros (indicating an unvoiced sound) then the gate 107 is held open by the application of a logical "1" to its first input. Under this condition, the random sequence generator output signal is passed through the gate 107 and becomes the second input signal to the digital multiplier 105 which multiplies the spectral data from the delogging computer 104. For the case in which the pitch word is voiced, the application by the OR gate 150 of a logical "0" inhibits the gate 107 causing the gate 107 to supply a constant +1 value as to the digital multiplier 105.

Odd Function Generator/Even Function Generator

The odd function generator 108 and the even function generator 46 (FIG. 2A) operate in a similar manner. Therefore, only the odd function generator 108 will be discussed in detail with the appropriate circuit modifications to obtain an even function generator noted.

The output signal from the unvoiced modifying unit 106 includes 32 samples of data. To be compatible with the other data being processed in the FFT computer means 30 of FIG. 2A, the output signals must be converted to 256 samples having odd symmetry about the 128th sample. (The even function generator 46 must generate an even function about the 128th sample.)

The odd function generator 108 includes any well-known sequential counter 180 having a plurality of output terminals which are sequentially activated. Each of the plurality of output terminals is connected to a separate one of a plurality of gates 182, for example, an AND gate. In the instant example, there are 32 terminals and gates. A second input connection to each of the gates 182 originates at the output of the unvoiced modifying unit 106. The output of each gate 182 is connected directly to a first predetermined location in a storage means such as the memory 184. The OUTPUT OF each gate 182, except the first and last gates, is connected to a second predetermined location in the memory 184 through a plurality of negating circuits 186. The negating circuits 186, which change the sign of the data and leave the magnitude unaltered, are omitted in the case of the even function generator 46. The reason will become apparent as the operation of the odd function generator 108 is explained.

The sequential counter 180 provides a gating pulse which sequentially activates each of the gates 182. Each of 32 delogged spectrum samples are gated through its associated gate 182 to a predetermined location in the memory 184. For example, the first sample is stored in the 0th location and the 32 sample is stored in the 128th location. The remaining 30 samples are distributed in steps of four between the 0th and the 128th memory location. To generate the odd function, each of the remaining 30 samples is also directed through its respective negating circuits 186 to a location in the memory 184 symmetrically related to its first storage location with location 128 as a reference. For example, the sample stored in the memory location number 3 would be negated and stored in location number 252, the sample stored in memory location 7 would be negated and stored in memory location number 249, etc. Thus, the desired 256 sampled odd symmetrical function is generated and directed to the imaginary input terminal $I_K^{(1)}$ of the FFT computer means 30 via the summation circuit 48 (FIG. 2A) and is processed by the FFT computer means 30 in the manner described hereinabove. The FFT computer means 30 generates the discrete sine transform of the received spectrum magnitude which appears at the output terminal H_n of the FFT computer means 30 as the impulse response of the received speech signal.

The vocal tract response data (as represented by the impulse response H_n) is weighted, for example, Hanning weighted, to both limit its duration is time by a precisely controlled amount and to remove the effects of discontinuities at the tails of the response function. The effect of the Hanning weighting in the frequency domain is that of filtering the samples of the spectral envelope function to obtain a smooth spectral envelope. The weighting operation is performed by the weighting circuit 200 of FIG. 11 which includes a storage device such as the register 202 having a first input connection from the H_n terminal of the FFT com-

3,681,530

puter means 30 and a second input connection from a masking circuit 204 well-known in the art. A multiplier circuit 206 has a first input connection from the register 202, a second input from a weighting function storage device 208 and an output connection connected to an odd function generator 210.

In operation, 128 samples representing the impulse response are received at the register 202 from the H_n terminal of the FFT computer means 30. The masking circuit 204 masks out a predetermined number of samples, for example, the last 78 allowing only the first 50 to be transferred to the multiplier circuit 206. (This in effect limits the time of the impulse response.) For each of the remaining 50 samples, there is stored in the weighting function storage device 208 a predetermined value by which the sample is to be multiplied by the multiplier 206. The predetermined values stored in the weighting function storage device 208 is in accordance with the well-known Hanning weighting function.

The 50 weighted samples are then directed to the odd function generator 210 which operates in the same manner as the odd function generator 108 described hereinabove except that point of symmetry is the 50th sample in the case of the odd function generator 210. The output signal of the weighting circuit then is 100 samples of Hanning weighted vocal tract data having an odd symmetry.

The vocal tract data could at this point be convolved with the pitch information to yield the desired speech information. However, the synthesized voice would contain a certain harshness. By obtaining the average impulse response for two data analysis intervals and convolving both the average impulse response and the FFT calculated impulse response with the pitch data during appropriate parts of the analysis interval, the transition in vocal tract information between analysis intervals is reduced thereby reducing the harshness of the synthesized voice. This averaging is particularly useful when the impulse responses generated by the system for adjacent analysis intervals are very different. The occurs at transitions between sounds when the spectrum magnitude of the speech signal is rapidly changing.

Averaging Circuit

An averaging circuit 212 of FIG. 11 for obtaining the average spectral data includes first, second and third storage devices 214, 216 and 218 respectively. While three storage devices are shown, it is to be appreciated that one storage device having different storage addresses could be used. The input connection to the averaging circuit 212 is connected to the first storage device 214 and to a summation circuit 220, the output of which is connected to a multiplier circuit 222. (Both digital summation circuits and multiplier circuits are well-known.) A second input connection to the summation circuit 220 originates at the second storage device 216 which has an input connection from the first storage device 214. The third storage device 218 has an input connection from the multiplier circuit 222.

In operation, the 100 weighted samples from the weighting circuit 200 are directed simultaneously to the summation circuit 220 and to the first storage device 214. The storage device 214 holds the 100 samples for one data analysis interval before shifting its

contents into the second storage device 216 for use in the next interval. The present 100 samples (for the nth analysis interval) are added to the 100 samples from the previous or $n-1$ analysis interval stored in the second storage device 216. The summed output samples are directed through the multiplier circuit 222 where the samples are multiplied by a constant, for example one-half, to obtain the average vocal tract data for two data analysis intervals. This 100 samples of average vocal tract data is stored in the third storage device 218. The 100 samples of data stored in each of the second and third storage devices 216 and 218, respectively, represents the spectral data which is to be convolved with the excitation function data derived from the seven bits of pitch information during any analysis interval, as will be explained hereinbelow.

Pitch Carrier Generator

The purpose of the pitch carrier generator is to generate from the seven bits of pitch data the proper excitation function. If the pitch data represents an unvoiced sound, then the pitch carrier generator generates an excitation function occurring at a fixed rate, for example, every 4.0 milliseconds. On the other hand, if the pitch information is voiced, the pitch carrier generator generates an excitation function, the instantaneous period of which is related to the value of the pitch signal represented by the seven bit pitch word.

An embodiment of a pitch carrier generator 229 according to the present invention is shown in FIG. 12 and includes a means 230 for obtaining voiced data samples based on the slope of a line as determined by two successive pitch words received during two successive data analysis intervals. Also included is a means 232 for generating a predetermined ramp function, for example, a 45° ramp function. A comparator means 234 has input connections from the means 230 for obtaining voiced data samples and from the means 232 for generating a predetermined ramp function and has an output connection to a gating means such as the OR gate 236. A means 238 for generating an unvoiced pitch carrier provides a second input connection to the OR gate 236, the output of which is connected to a convolution circuit, to be discussed in detail hereinafter.

The waveforms (a), (b) and (c) of FIG. 13 will be useful in explaining the operation of the pitch carrier generator 229. Waveform (a) is a series of time marks where the time T between the timing marks n and $n-1$ is equivalent to one data analysis interval. In waveform (b) the end points 1 and 2 indicate the value of the pitch words at the $n-1$ interval and n interval, respectively, and are employed to calculate the slope of the line 3. The lines 4 of waveform (b) indicate the predetermined ramp function. The waveform (c) of FIG. 13 indicates the pitch carrier signals generated by the pitch carrier generator 229. While the waveforms are shown as solid lines, it is to be appreciated that there are in the instant example 100 discrete samples occurring in the time interval T.

In operation, the pitch word (assuming a voiced sound) for the nth data analysis is received at the input to the means 230 for obtaining voiced data samples from the buffer storage 122 of the decoding device

35

100. The means 230 determines the slope of the line connecting the magnitude of the pitch signal received at the nth (or present) data analysis interval with the magnitude of the pitch signal received in the previous data interval. The calculated slope value is sequentially added to each of the 100 samples and directed to the comparator means 234. Simultaneously, the means 232 for generating a 45° ramp signal is supplying the comparator means 234 with samples of the ramp function signal. When the magnitude of the ramp function signal is equal to or greater than the magnitude of the output signal from the means 230, the comparator means 234 generates a fixed amplitude signal as shown in waveform (c) of FIG. 13. If the pitch word indicates an unvoiced sound (the seven bits of the pitch word are all zeros), the means for generating an unvoiced pitch carrier genrates a fixed amplitude signal at a fixed rate, for example, one pitch carrier every 4.0 millseconds.

Means for Obtaining Voiced Data Samples

The input connection from the buffer storage device 122 of FIG. 10A is connected to a gating means such as the AND gate 240 to a slope calculator 242 and to a first pitch storage means such as a pitch storage register 244. A storage counter 246 has a first input connection from the AND gate 240 and an output connection to a gating means such as an AND gate 248, the output of which is connected to a summation circuit 250. The slope calculator 242 has a second input connection from a second pitch storage register 252 and an output connection to a third gating means such as the AND gate 254, the output of which is connected to a second input connection of the summation circuit 250. The output of the summation circuit 250 is connected simultaneously to the comparator means 234 and as a second input to the storage counter 246.

Originating at the system timing unit (not shown) are second input connections to the AND gates 240, 248 and 254. The timing signals, called enabling signals, include an enabling 0 signal occurring once every T seconds where T corresponds to the length (seconds) of a data analysis interval, and an enabling 1 signal occurring once every T̄ seconds where T̄ corresponds to the sampling rate of the A/D converter 12 of FIG. 2A. In the system described herein, T is approximately 20 milliseconds and T̄ is approximately 140 microseconds.

The means for obtaining voiced data samples solves the equation

$$Y_t = Y_{t-1} + m \qquad (46)$$

where Y_t is a calculated value of a particular sample of the pitch signal, Y_{t-1} is the value of a previously calculated pitch signal and m is the slope of a line joining the two pitch signals received during the nth and $n-1$ data analysis intervals.

The operation of means 230 for obtaining voiced data samples is initiated by the arrival of the seven bit pitch word for the nth data analysis interval. The pitch word is directed simultaneously to the pitch storage register 244, the slope calculator 242 and through the AND gate 240 to the storage counter 246. The pitch word will be held in the pitch storage register 244 for one time period T (corresponding to one data analysis interval) before being transferred to the pitch storage register 252.

36

The slope calculator 242 takes two pitch words, one for the nth analysis interval and one for the $n-1$ data analysis interval received from the pitch storage register $n-1$, and calculates the slope of a line connecting these two pitch words. The slope calculator can be any well-known computing means for solving the simple calculation

$$m = \frac{T_{p(n-1)} - T_{pn}}{T} \qquad (47)$$

where $T_{p(n-1)}$ is the value of the pitch word for the $n-1$ data analysis interval and T_{pn} is the value of the pitch word for the nth data analysis interval and T is the length of a data analysis interval. The output signal of the slope calculator is a constant over the interval T for which it was calculated.

Once every sampling interval T̄, an enabling signal, enable 1, activates the AND gates 254 and 248 to thereby direct the slope signal m and the previously calculated pitch signal (stored in the storage counter 246) to the summation circuit where the two signals are added in accordance with Equation 46. The summation output signal, Y_t, is simultaneously directed to the storage counter 246 to become the Y_{t-1} signal for the next calculation and to the comparator means 234 for further processing.

Means for Generating a Predetermined Ramp Function

The means 232 for generating a predetermined ramp function includes a storage counter 260 connected through a gating means such as an AND gate 262 to a summation circuit 264. A fixed slope source 266, for example, a 45° slope source, is connected through a gating means such as the AND gate 268 to a second input connection of the summation circuit 264. The summation circuit 264 has an output connection to the storage counter 260 and to the comparator means 234.

The means 232 for generating a predetermined ramp function solves the equation

$$X_t = X_{t-1} + m \qquad (48)$$

where X_t is the calculated value of the ramp function at a particular time, X_{t-1} is the value of the ramp function calculated during the previous sample interval T̄ and m is the slope value which is unity for a 45° ramp. The means 232 recursively solves Equation 48 in the following way. Initially (when enable 0 occurs) the storage counter 260 starts counting from zero and for every sampling interval T̄ the Equation 48 is solved placing a new ramp value in the storage counter 260 via the line from the summation circuit 264 to the storage counter 260. At any sample time T̄, an enable 1 signal activates the AND gates 202 and 208 thereby directing a signal having the previously calculated value X_{t-1} of the ramp function from the storage counter 260 to the summation circuit 264 and directing the slope signal from the fixed slope course 266 to the summation circuit 264. The two signals are added together in accordance with Equation 48 to yield the desired solution.

Comparator Means

The comparator means 234 includes a digital comparator circuit 270, well-known in the art, having first and second input connections, respectively from the

summation circuit 250 and the summation circuit 264 and first and second output connections, respectively, to a pulsing means such as a one shot multivibrator circuit 272 and the OR gate 236. The one shot multivibrator circuit 272 is connected to the storage counter 260 and to the common juncture of the comparator circuit 270 and the OR gate 236.

The input signal Y_t from the summation circuit 250 corresponds to line 3 of waveform (b) of FIG. 13 taken over the interval T, and similarly the input signal X_t corresponds to the line 4 in the same waveform and over the same time interval. (The data analysis interval T changes once every 20 milliseconds and the sampling interval $\bar{T}$ occurs once every 140 microseconds.) It is to be appreciated that a value Y_t corresponding to one increment of line 3 and a value X_t corresponding to one increment of one of the lines 4 and then values Y_t and X_t are calculated once every 140 microseconds $\bar{T}$ over the interval T. Therefore, the comparator 234 makes a sample by sample comparison.

If the signal Y_t is greater than the signal X_t, then the comparator circuit 270 increments a zero from the second output connection to the OR circuit 236. If the signal X_t is greater than or equal to the signal Y_t, then the comparator circuit directs an output connection to the one shot multivibrator circuit 272 which generates and supplies a pulse output signal (see waveform (c) of FIG. 13) to the OR circuit 236. As can be seen in waveform (b) of FIG. 13, $X_t \geq Y_t$ at the intersections of the lines 4 with the line 3 and therefore a pitch carrier signal, indicating a voiced sound, will be generated whenever the intersections occur. Thus the spacing between voiced pitch carrier signals is a function of the slope of line 3. For example, the greater the slope of line 3, the less frequent are the occurrences of the varied pitch carrier pulses as seen by the spacing between times t_5 and t_6 as compared with the spacing between times t_2 and t_3.

It can be seen from waveform (b) of FIG. 13 that each time an intersection occurs ($X_t \geq Y_t$) then the ramp function for the next X_t starts at zero. The reset to zero is implemented by the connection of one shot multivibrator circuit 272 to the storage counter 260. Upon receipt of a signal from the one shot multivibrator circuit 272, the storage counter 260 is reset to zero thus setting the initial value of the ramp function each time the signal $X_t \geq$ the signal Y_t.

Means for Generating an Unvoiced Pitch Carrier

The means for generating an unvoiced pitch carrier signal 238 includes a gating means such as an AND gate 280 having an output connection to the OR circuit 236 and a first input connection from a means operative to sense unvoiced pitch words such as the series combination of a seven bit OR gate 282 and a complement circuit 281 and a second input from a pulse generation circuit such as a free-running multivibrator circuit 284. It is to be appreciated that the number of inputs to the OR gate 282 corresponds to the number of bits in the pitch word and that the input connection to the OR gate 282 can originate anywhere in the vocoder system where the pitch word for the analysis interval of interest is stored, for example, from the pitch storage register 252 of means 230.

In operation the free-running multivibrator circuit 284 generates a constant amplitude pulse at a predeter-

mined rate, for example, once every 4.0 milliseconds. The second input signal to the AND gate 280 from the complement circuit 281 maintains the AND gate 280 in an open condition in the absence of an input signal to the seven input OR gate 282. For example, in the unvoiced case, the seven bits in the pitch word (stored in pitch storage register 252) are all zeros, and the AND gate 280 directs the output signal from the free-running multivibrator circuit 284 to the OR gate 236. It should also be mentioned that for the unvoiced condition, the second input to the OR gate 236, which normally would come from the voiced pitch carrier generation portion of the pitch carrier generation unit 229, is inhibited. When a voiced pitch word is shifted into the pitch storage register 252 (a voiced pitch signal is characterized by a "1" in any of the seven bit positions), a signal is directed through one of the seven inputs of the seven input OR gate 282 to the complement circuit 281 where it is complemented and inhibits the AND gate 280 from passing any pulse from the free-running multivibrator 284.

Thus the output signal of the OR gate 236 is a fixed amplitude signal having either one of two repetition rate sequences. In the unvoiced case, the repetition rate is fixed, and in the voiced case, the repetition rate varies as a function of the slope of a line connecting two successive pitch words.

Convolution Unit

The synthesis of voiced sounds is accomplished by convolving the odd symmetric and discretely sampled time varying vocal tract response function with the digital equivalent of a unit area impulse carrier at a rate calculated in accordance with the pitch carrier generator unit 229 output signal. A convolution unit 300 according to the present invention is shown in FIG. 14 and includes a convolution means 302 for storing a predetermined vocal tract response signal having input connections from a logic means 304 and output connections to a summation circuit 306. Input connections to the logic means 304 include input connections from the pitch carrier generator 229 of FIG. 12 and the averaging means 212 of FIG. 11. (An input connection may come directly from the H_n terminal of the FFT computer means 30 if the weighting circuit means 200 and/or the averaging circuit 212 are not employed.)

In operation the logic means 304 directs the vocal tract response signals from the averaging circuit 212 to predetermined storage locations in the convolution means 302. In response to each pitch carrier signal received from the pitch carrier generator 229, the logic means 304 selects a predetermined block of storage locations within the convolution means 302 from which a complete set of data samples, 100 in the instant example, are sequentially directed to the summation circuit 306 during the next 100 sample intervals. For each pitch carrier received, the convolution means 302 will supply sequentially 100 samples of the vocal tract response signal. If, for example, four pitch carrier signals are received at predetermined time intervals, the convolution means 302 will supply sequentially 100 samples of vocal tract data, each of the 100 samples starting at a time corresponding to the receipt of one of the pitch carrier signals.

This operation is shown in FIG. 13 by the waveforms

(d), (e), (f), (g), (h) and (i). At each time t_j ($j = 1, 2, \ldots$
$\ldots 6$) a new scan of the appropriate impulse response is begun. The output signal corresponding to the output of summation circuit **306** is shown in waveform (k) of FIG. 13. For any specific time instant, only the discrete samples of the impulse response that are in the process of being scanned out are processed by summation circuit **306**. For example, at time t_4, the summation circuit would be summing together four samples. These samples would be the first sample of the impulse response starting to be scanned at time t_4 and one sample each from the impulse response which began at times t_1, t_2 and t_3. The 100 samples will therefore have periods of overlap and will be added together at the summation circuit **306** to form a composite signal which represents the desired synthesized speech signal. The composite signal may be directed through any well-known digital to analog converter and speaker, not shown, for listening.

The logic means **304** includes first and second gating means such as first and second AND gates **320** and **322**, respectively, each having a first input connection from pitch carrier generator **229**. A second input connection to each of the first and second AND gates **320** and **322** originates at first and second output connections, respectively, of a timing means such as a flip-flop **324**. A third and fourth gating means, such as the AND gates **326** and **328**, have a first input connection from the AND gate **322**. A second input connection to the AND gates **326** and **328** originates at the first and second output connections, respectively, of a signaling means such as the flip-flop **330** which is also connected to the first input connections, of a fifth and sixth gating means such as the AND gates **332** and **334**. The second input connection to the fifth and sixth AND gates **332** and **334** originates at the output connection of the storage device **216** of the averaging circuit **212** shown in FIG. 11.

The convolution means **302** includes a plurality, for example, three, of storage means **350**, **352** and **354**. The first storage means **350** has input connections from the third and fifth AND gates **326** and **332**, the second storage means **325** has input connections from the fourth and sixth AND gates **328** and **334**, and the third storage means **354** has input connections from the first AND gate **320** and the storage device **218** of the averaging circuit **212** shown in FIG. 11.

Each of the storage means **350**, **352** and **354** may be similar; therefore, only the details of the storage means **350** are shown in FIG. 14. Each storage means includes a plurality of storage registers, the number of which depends upon the maximum pitch carrier rate. For example, at a frame updating time of 20 msec and with averaging of impulse responses occurring every one-half frame (10 msec), any specific impulse response is used to represent the vocal tract for only one 10 msec interval. Since the maximum pitch rate for the system is approximately a 3 msec rate, then at most four pitch pulses can be generated by the pitch carrier generator during any 10 msec interval. On this basis, at most four storage areas per impulse response are required to satisfy the maximum pitch carrier rate. These four storage blocks are shown for a typical storage means **350** as storage registers **360**, **361**, **362** and **363** of FIG. 14, and each of the four storage registers has a common

input connection from the AND gate **332** and a separate input connection from a gating circuit **346**, to be discussed in detail hereinafter. The output connection from each of the four storage registers **360** through **364** is connected to the summation circuit **306**.

The operation of the convolution circuit will be explained in conjunction with the waveforms of FIG. 13. Assume that at time t_1 an enable 0 signal is directed to the flip-flop **324**, the output signal of which opens the AND gate **322** and blocks the AND gate **320**. Simultaneously the enable 0 signal sets the flip-flop **330** such that the output signal of the flip-flop **330** opens the AND gates **326** and **332** and blocks the AND gates **328** and **334**. The impulse response H_n (vocal tract response function data) is then passed by AND gate **332** from the storage device **216** of the averaging circuit **212** to the storage registers **360** through **363**.

A pitch carrier signal received from the pitch carrier generator **229**, for example at time t_1, will be directed through the AND gates **322** and **326** (the AND gate **320** is closed to the gating circuit **346**. The gating circuit allows shift pulses at the sampling rate $\overline{T}$ to be directed to the storage register **360**. One sample of the impulse response function H_n, stored in register **360**, is shifted into the summation circuit with the receipt of each shift pulse ($\overline{T}$ rate). While the waveforms of FIG. 13 are shown as continuous waveforms, it is to be appreciated that they are composed of discrete samples, but because of the relative time periods of T and $\overline{T}$ (20 milliseconds versus 140 microseconds), the signals appear as continuous waveforms.

At time t_2, another pitch carrier signal, received from the pitch carrier generator **229**, is directed through the AND gates **322** and **326** to the gating circuit **346** which opens a gate to the storage register **361** allowing the shift pulses occurring at the $\overline{T}$ rate to also access the H_n data stored in the storage register **361**. The summation circuit **306**, receiving data samples from the two storage registers **360** and **361** every $\overline{T}$ seconds generates a composite signal, waveform (k) of FIG. 13, at its output terminal. As stated hereinabove, the composite signal generated at the output of the summation circuit **306** is the desired synthesized speech signal in digital form. Note that the four waveforms (d), (e), (f) and (g) are identical but start at times corresponding to the receipt of pitch carrier signals at t_1, t_2, t_3 and t_4.

At the time $T/2$, corresponding to the time of one-half the data analysis interval, the flip-flop circuit **324** changes state such that its output signal opens the AND gate **320** and closes the AND gate **322**. Therefore, any pitch carrier signal occurring after the time $T/2$ is directed to the equivalent of the gating circuit **346** in the third storage means **354**. For example, at times t_5 and t_6, pitch carrier signals received from the pitch carrier generator **229** are now directed through the gate **320** to the gating circuit of the third storage means **354** which transfers the shift pulses to the first and second storage registers, respectively, of the third storage means **354**. The average impulse response data have waveforms (h) and (i) and are transferred to the summation circuit **306** where they are added to the impulse responses that have not been completely scanned as represented by the waveforms (d) through (g) to supply a digital composite signal, waveform (k),

representing the synthesized speech.

Note that shortly after time t_5 the impulse response that began to be scanned out of storage means 350 at time t_1 is complete and storage register 360 no longer supplies data to the summation circuit 306. Similarly by time t_8 the data in storage register 361 which was being supplied to the summation circuit 306 due to the pitch carrier generator 229 output at time t_2 has also been completely scanned. Therefore, the convolution is actually a process of gating in and gating out appropriate storage register contents as a function of the pitch carrier generator output. The synthesis for unvoicing is almost identical to the voiced synthesis just described. The differences between the two modes include the following:

1. the pitch carrier generator 229 instead of supplying a pitch carrier which is a function of the speaker's pitch (i.e., during unvoiced sounds there does not exist a periodic excitation) arbitrarily supplies a pitch carrier at some fixed rate (4 msec for this example).

2. The spectrum of H_n must previously have been multiplied by a random sequence of ±1 amplitude. This multiplication in the frequency domain is equivalent to convolving the impulse response H_n with a noise carrier in the time domain.

3. The impulse response used in the convolution means 300 during the unvoiced excitation pitch carrier is randomly multiplied by ±1 before storage in its respective storage means 350, 352 or 354.

To perform the random multiply by ±1 of the impulse response, a control signal from the unvoiced pitch carrier signal means 238 is supplied to storage registers 371 and 372 of FIG. 14. These storage means contain the constants +1 and −1. During normal voicing the multiplier supplied by storage registers 371 and 372 to the multiplier circuits 370 and 373 are always +1. During the unvoiced synthesis intervals the storage registers 371 and 372 randomly supply either +1 or −1 as the second input to the multiplier. In this manner the H_n generated for an unvoiced sound is convolved with a fixed rate sequence whose individual impulse values may be +1 or −1.

One embodiment of a gating circuit 346 employed in the first, second and third storage means 350, 352 and 354 is shown in FIG. 15 and includes an input terminal 381 connected to a first flip-flop circuit 380 and to first, second and third gating means such as the AND gates 382, 384 and 386. The output connection of the first flip-flop 380 is connected to a fourth gating means such as the AND gate 388 and to a second input connection of the first AND gate 382. The output connection of the fourth AND gate 388 is connected to the first storage register 360 of the first storage means 350. A second flip-flop circuit 390 has an input connection from the first AND gate 382 and an output connection to the second AND gate 384 and to a fifth gating means such as the fifth AND gate 392, the output connection of which is connected to the second storage register 361.

A third flip-flop circuit 394 has an input connection from the second AND gate 384 and an output connection to the third AND gate 386 and to a sixth gating means such as the sixth AND gate 396, the output of which is connected to the third storage register 362. A

fourth flip-flop circuit 398 has an input connection from the third AND gate 386 and an output connection from the third AND gate 386 and an output connection to a seventh gating means such as the seventh AND gate 400, the output of which is connected to the fourth, storage register 363. A second input connection to each of the fourth, fifth, sixth and seventh AND gates 388, 392, 396 and 400, respectively, originates at the source of shift pulses (not shown).

The operation of the gating circuit of FIG. 15 will be explained in conjunction with the waveforms of FIG. 13. At time t_1, a carrier pulse, waveform (c), is received at terminal 381 which changes the state of the first flip-flop circuit 380 such that the first and fourth AND gates 382 and 388 which are normally closed are opened The fourth AND gate 388 immediately directs shift pulses to the first storage register 360 of FIG. 14 which in turn starts reading out the data corresponding to waveform (d). At time t_2, a second pitch carrier signal is received at the input terminal 381 and is directed simultaneously to the first flip-flop circuit 380 and through the now open first AND gate 382 to the second flip-flop circuit 390 which in turn opens the normally closed second and fifth AND gates 384 and 392. The second pitch carrier signal has no effect on the first flip-flop circuit 380 as its state was changed by the first pitch carrier signal received at time t_1. Both the fourth and fifth AND gates 388 and 392 are simultaneously directing shift pulses to the respective first and second storage registers 360 and 361 to generate the waveforms (d) and (e).

Similarly at times t_3 and t_4 the respective pitch carrier signals are directed to the associated second and third AND gates 384 and 386 to open the sixth and seventh AND gates 396 and 400 via the third and fourth flip-flop circuits 394 and 398 whereby the shift pulses are transferred to the respective third and fourth storage registers 362 and 363 of FIG. 14. Thus the gating circuit 346 not only provides shift pulses to an additional storage register upon the receipt of each carrier signal but also continues to provide the shift pulses to each activated storage register until the 100 samples of vocal tract data are passed to the summation circuit 306.

Synthesizer Summary

In summary, the synthesizer includes a decoding device 100 (FIG. 10A) which dequantizes, descales and converts 13 received $\bar{C}_n$ coefficients into 21 C_n coefficients. These 21 C_n coefficients are then directed through a spectrum decoder 102 (FIG. 10B) where the logarithm of the spectrum envelope of the received vocal tract impulse response function is computed using the discrete Fourier transform (DFT). The resultant 32 samples of a logged signal are then de-logged in the delogging computer 104. The 32 de-logged samples are spaced along the frequency axis in every fourth frequency position in the lower 128 locations of a 256 word data area and converted into a signal having 256 samples of odd symmetry by the odd function generator 108. The function is made odd symmetric to use the remaining odd symmetric input of the imaginary input $I_k{}^{(1)}$ of the FFT computer means 30 (FIG. 2A). This odd part of $I_k{}^{(1)}$ is transformed by the FFT computer means 30 and is available at the output terminal H_n of the FFT computer means 30. Since the

43

256 samples are the discrete sine transform of the received spectrum magnitude, H_n is an odd function (see waveform (c) of FIG. 3) and is the impulse response which is to be used in synthesis of the received speech signal.

The 256 samples of impulse response data are then directed through a weighting circuit 200 (FIG. 11) where the data is Hanning weighted and the number of samples is reduced to 100. These 100 samples are chosen such that the remaining H_n function is still odd symmetric. The 100 samples of weighted data are directed to an averaging circuit 212 where they are both stored for use in conjunction with the previous 100 samples to form 100 samples of the average impulse response. The previous 100 samples and the 100 average samples are directed to the convolution means 300 (FIG. 14).

The seven bits of received pitch data are directed to a pitch carrier generator 229 (FIG. 12) where the pitch data is converted into a fixed amplitude pitch carrier signal having a fixed rate in the case of unvoiced sounds and having a rate related to the slope of a line connecting to successive pitch words for the voiced sounds. The pitch carrier signals are then directed to the convolution means 300 (FIG. 14) where they are convolved with the 100 samples of the appropriate impulse response data to generate the desired synthesized speech.

What has been shown and described herein is considered a preferred embodiment of the present invention. It will be obvious to those skilled in the art that various modifications and changes may be made without departing from the invention as defined by the appended claims.

What is claimed is:

1. A bandwidth compression system including an analysis section comprising:

means for generating electrical signals representing the Fourier transform of the logarithm of the magnitudes of the spectrum of an input signal, said input signal having excitation and impulse response information included therein;

first detection means coupled to said means for generating electrical signals and being operative to provide from said electrical signals an output signal representing the excitation information of said input signal; and

second detection means coupled to said means for generating electrical signals and being operative to separate out a predetermined portion of said electrical signals, said predetermined portion representing the impulse response information of said input signal.

2. A processor according to claim 1 including a synthesis section comprising:

impulse response means coupled to said second detection means and being operative in response to the predetermined portion of said electrical signals to generate an output signal corresponding to the impulse response information;

excitation means coupled to said first detection means and being operative in response to the output signal from said first detection means to generate an excitation carrier signal; and

convolution means having input connections from said impulse response means and from said excita-

44

tion means and being operative to convolve the output signals from said impulse response means and from said excitation means to thereby synthesize the input signal.

3. A digital vocoder including an analysis section comprising:

means for obtaining spectrum magnitude signals of an input speech signal having voicing and vocal tract information;

logging means coupled to said means for obtaining spectrum magnitude signals and being operative to generate output signals representing the logarithm of the spectrum magnitude of the input speech signal;

first Fourier transform means coupled to said logging means and being operative to generate output signals having magnitude and positions and representing the Fourier transform of the logarithm of spectrum magnitudes of the input speech signal;

pitch detection logic means coupled to said Fourier transform means and being operative to extract a pitch signal from the output signal of said first Fourier transform means, said pitch signal having a magnitude representing the voicing information of the input speech signal; and

selecting means coupled to said first Fourier transform means and being operative to select a predetermined number of the output signals of said first Fourier transform means, said predetermined number of output signals representing the vocal tract information of the input speech signal.

4. A digital vocoder according to claim 3 including an encoding means coupled to said selecting means and being operative to quantize at a predetermined rate and scale by a predetermined factor each of the predetermined number of output signals of said Fourier transform means selected by said selecting means.

5. A digital vododer according to claim 3 including a synthesis section comprising:

second Fourier transform means being operative in response to the selected output signals of said first Fourier transform means to generate output signals representing the Fourier transform of said selected output signals of said first Fourier transform means;

delogging means coupled to said second Fourier transform means and being operative to generate output signals representing the antilogarithm of the output signals of said second Fourier transform means;

third Fourier transform means coupled to said delogging means and being operative to generate output signals representing the vocal tract information of the input speech signal;

pitch carrier generator coupled to said pitch detection logic means and being operative in response to said pitch signal to generate pitch carrier signals having predetermined rates; and

convolution unit coupled to said third Fourier transform means and to said pitch carrier generator and being operative to combine the output signals of said third Fourier transform means and the pitch carrier signals from said pitch carrier generator to thereby generate the synthesized version of the

input speech signal.

6. A digital vocoder according to claim 3 wherein said means for obtaining the spectrum magnitude signals of an input speech signal includes:

transducer means being operative to convert said input signal into an electrical input speech signal;

an analog to digital converter connected to said transducer means and being operative to convert said electrical input speech signal into a digital speech signal;

computer means coupled to said analog to digital converter and being operative to generate real and imaginary signals representing the spectrum of the digital speech signal; and

a magnitude computation circuit connected to said computer means and being operative to combine in a predetermined manner said real and imaginary signals to generate the spectrum magnitude signals of said input speech signal.

7. A digital vocoder according to claim 6 further including a normalization unit connected between said analog to digital converter means and said computer means and being operative to change the level of the input signals a predetermined factor to maintain the peak value of the digital speech signal to said computer means within a predetermined dynamic range.

8. A digital vocoder according to claim 6 further including a weighting function circuit connected between said analog to digital converter means and said computer means and being operative to weight the digital speech signal to obtain a smooth spectral signal from said computer means.

9. A digital vocoder according to claim 3 wherein said pitch detection logic means includes:

selection means having an input connection from said first Fourier transform means and being operative to select the output signal of said first Fourier transform means having the largest magnitude;

first comparator means having an input connection from said selection means and a first and second output connection, said first comparator means being operative to compare the magnitude of the selected output signal of said selection means to a predetermined threshold level and to generate an output signal at said first output connection if the magnitude of said selected output signal exceeds the predetermined threshold level and to generate a predetermined output signal at said second output connection if the magnitude of said selected · output signal is less than the predetermined threshold level; and

buffer storage means having a first input connection connected to the common juncture of said selection means and said first comparator means, a second input connection connected to the first output connection of said first comparator means and an output terminal and being operative to store the output signal from said selection means and to shift the stored signal to the output terminal upon the receipt of a signal from said first comparator, means,

whereby an unvoiced speech signal is indicated when said first comparator means has an output signal at said second output connection and a voiced

speech signal is indicated when the output signal of said first Fourier transform means is shifted to the output of said buffer storage means.

10. A digital vocoder according to claim 9 further including means for determining voicing information having input connections connected to said means for obtaining spectrum magnitude signals and the first output connection of said first comparator means, a first output connection connected to the second input connection of said buffer storage means and a second output connection and being operative in response to the spectrum magnitude signals to provide an output at said first output connection when said spectrum magnitude signals include a voiced signal and to provide an output signal at said second output connection when said spectrum magnitude signals include an unvoiced signal.

11. A digital vocoder according to claim 10 wherein said means for determining voicing information comprises:

means connected to said means for obtaining spectrum magnitude signals for computing a first output signal representing the low-band energy of the spectrum magnitude signals and a second output signal representing the high-band energy of the spectrum magnitude signals;

means for combining said first output signal representing the low-band energy with said second output signal representing the high-band energy to form a composite signal representing the ratio of said first and second output signals;

second comparator means having an input connection coupled to said means for computing, an output connection, and a predetermined threshold level and being operative to generate an output signal at its input connection when the output signal representing the low-band energy is greater than its predetermined threshold level;

third comparator means having an input connection coupled to said means for combining, an output connection and a predetermined threshold level and being operative to generate an output signal at its output connection when said composite signal representing the ratio of said first and second output signals is greater than its predetermined threshold level; and

fourth comparator means having a first input connection coupled to the output connection of said second comparator means, a second input connection coupled to the output connection of said third comparator means and a first output connection coupled to said buffer storage means and a second output connection and being operative to generate a signal at its first output connection when two predetermined signals are received at its first and second input connections, respectively, and to generate a signal at its second output connection when only one predetermined signal is received at either its first or second input connection.

12. A digital vocoder according to claim 7 further including a denormalizing unit coupled to said normalization unit and to said first Fourier transform means and being operative to alter the magnitude of the output signal of said first Fourier transform means in a predetermined manner related to the predetermined

factor of said normalization unit.

13. A digital vocoder according to claim 12 wherein said denormalizing unit is a computer capable of solving the equation

$$C_o = C'_o - 16 \quad \sqrt{N} \log_2 (G_N) \qquad 5$$

where C_o is the altered magnitude, C'_o is the unaltered magnitude, N is the selected predetermined number of output signals from said first Fourier transform means and G_N is the predetermined factor of said normalization unit.

14. A digital vocoder according to claim 4 wherein said encoding means comprises:

scaling factor storage means operative to store a predetermined scaling factor for each of the predetermined number of output signals of said first Fourier transform means;

scaling means coupled to said scaling factor storage means and to selecting means and being operative to add each of the predetermined scaling factors to a separate one of the predetermined number of output signals of said first Fourier transform means to eliminate negative values in said predetermined number of output signals;

ratio storage means operative to store a predetermined ratio signal for each of the predetermined number of output signals of said first Fourier transform means; and

multiplier means coupled to said scaling means and said ratio storage means and being operative to multiply each of the scaled output signals of said scaling means by a corresponding ratio signal stored in said ratio storage means to thereby quantize each of the predetermined numbers of output signals of said first Fourier transform means.

15. A digital vocoder according to claim 14 further including gating means coupled to said multiplier means and being operable to gate certain ones of said predetermined number of output signals of said first Fourier transform means at a first predetermined rate and to gate the remainder of the output signals of said first Fourier transform means at a second predetermined rate.

16. A digital vocoder according to claim 5 wherein said second Fourier transform means is a Fourier transform computer means operable to solve the expression

$$\log |V_n| = \sum_{k=0}^{R} C_k \cos \left(\frac{2\pi nK}{N} \right)$$

where V_n is the nth frequency sample of the selected output signals of said first Fourier transform means, C_k is the kth sample of the selected output signals of said first Fourier transform means and K and R are the limits of summation.

17. A digital vocoder according to claim 5 wherein said pitch carrier generator includes:

first means responsive to said pitch signal from said pitch detection logic means for generating a first predetermined pitch carrier signal when the magnitude of the pitch signal indicates a voiced signal;

second means responsive to said pitch signal from said pitch detection logic means for generating a second predetermined pitch carrier signal when the magnitude of the pitch signal indicates an unvoiced signal; and

gating means coupled to said first and second means

for generating and being operative to gate a first predetermined pitch carrier signal to said convolution means when the magnitude of the pitch signal is less than a predetermined magnitude and to gate a second predetermined pitch carrier signal to said convolution means when the magnitude of the pitch signal is greater than a predetermined magnitude.

18. A digital vocoder according to claim 17 wherein said first means for generating includes:

third means for generating signals, the magnitudes of which describe a predetermined function;

fourth means for generating signals, the magnitudes of which describe the slope of a line connecting the magnitudes of two successive pitch signals from said pitch detection means; and

first comparator means having input connections coupled to said third and fourth means for generating and an output connection coupled to said gating means and being operative to generate a first predetermined pulse when the signals from said fourth means for generating are equal to or greater than the magnitude of the signal from said third means for generating.

19. A digital vocoder according to claim 18 including an inhibiting means responsive to said pitch signal from said pitch detection logic means to inhibit the second predetermined pitch carrier signal of said second means for generating.

20. A digital vocoder according to claim 18 wherein said third means for generating signals includes:

first storage counter means having a first input connection and an output connection and being operative to store a first predetermined signal, to add to said first predetermined signal a second predetermined signal appearing at said first input connection and to supply the resultant signal to said output connection;

slope means for generating a third predetermined signal; and

first summation means having a first input connection coupled to the output connection of said first storage counter means, a second input connection coupled to said slope means and an output connection coupled to said first input connection of said storage counter means and to said gating means of said pitch carrier generators, said first summation means being operative to add the resultant signal of said first storage counter means to the third predetermined signal from said slope means to form said second predetermined signal and to direct said second predetermined signal simultaneously to said gating means of said pitch carrier generator and to said first storage counter means to update said first predetermined signal stored therein.

21. A digital vocoder according to claim 18 wherein said fourth means for generating signals includes:

means for computing a slope signal m wherein $m = T_p(n-1) - T_{pn}/T$, where T_p is a first pitch signal received from said pitch detection logic at a first predetermined time, $T_p(n-1)$ is a second pitch signal received from said pitch detection logic at a second predetermined time and T is the elapsed time between said first and second predetermined

times;

second storage counter means having a first input connection and an output connection and being operative to store a first predetermined signal, to add to said first predetermined signal a second predetermined signal appearing at said first input connection and to supply the resultant signal to said output connection; and

second summation means having a first input connection coupled to the output connection of said second storage counter means, a second input connection coupled to said means for computing a slope signal and an output connection coupled to said first input connection of said second storage counter means and to said gating means of said pitch carrier generator, said second summation means being operative to add the resultant signal of said second storage counter means to the slope signal from said means for computing a slope signal to form said second predetermined signal and to direct said second predetermined signal to said gating means of said pitch carrier generator and to said second storage means to update said first predetermined signal stored therein.

22. A digital vocoder according to claim 5 including a weighting circuit having an input connection coupled to said third Fourier transform means and an output connection coupled to said convolution means and being operative to apply weighting function signals to the output signals of said third Fourier transform means to thereby improve the quality of the synthesized version of the input speech signal.

23. A digital vocoder according to claim 22 wherein the weighting circuit includes:

a masking circuit having an input connection coupled to said third Fourier transform means and being operative to select a predetermined number of the output signals of said third Fourier transform means;

weighting function storage means being operative to store a predetermined number of signals corresponding to the predetermined number of output signals selected by said masking circuit; and

multiplier means having input connections coupled to said masking circuit and to said weighting function storage means and an output connection coupled to said convolution means and being operative to multiply each of the predetermined number of output signals selected by said masking circuit by a different one of the predetermined number of signals stored in said weighting function storage means to thereby weight the vocal tract data being directed to said convolution means.

24. A digital vocoder according to claim 5 wherein said convolution unit includes:

logic means having a first input connection coupled to said pitch carrier generator, a second input connection coupled to said third Fourier transform means, and first, second, third and fourth output connections, said logic means being operative in response to a first predetermined time period to provide a data path from said first and second input connections to said first and second output connections, respectively, and being operative in response to a second predetermined time period to

provide a data path from said first and second input connections to said third and fourth output connections respectively;

first storage means having first and second input connections coupled respectively to said first and second output connections of said logic means and a plurality of output connections, said first storage means being operative to store the output signals representing the vocal tract information received from the third Fourier transform means via the data path established by said logic means during said first predetermined time period and to gate from a different one of said plurality of output connections a complete set of vocal tract signals upon the receipt of each signal from said pitch carrier generator during said first predetermined time period;

second storage means having first and second input connections coupled respectively to said third and fourth output connections of said logic means and a plurality of output connections, said second storage means being operative to store the output signals representing the vocal tract information received from the third Fourier transform means via the data path established by said logic means during said second predetermined time period and to gate from a different one of said plurality of output connections a complete set of vocal tract signals upon receipt of each signal from said pitch carrier generator during said second predetermined time period; and

summing means having a plurality of input connections each coupled to one of said plurality of output connections of said first and second storage means and being operative to add the vocal tract signals from said first and second storage means whereby a synthesized version of the input speech signal is obtained.

25. A vocoder system for synthesizing a first speech signal and analyzing a second speech signal simultaneously, said first and second speech signals including voicing and vocal tract information, said digital vocoder comprising:

means for generating a pitch carrier signal from said first speech signal;

means for obtaining the frequency spectrum magnitude signals of said first speech signal;

means coupled to said means for obtaining the frequency spectrum magnitudes of said first speech signal for converting the frequency spectrum magnitudes into signals having a first predetermined symmetry;

means for obtaining the frequency spectrum magnitudes of a second speech signal;

means coupled to said means for obtaining the frequency spectrum magnitudes of said second speech signal for generating signals having a second predetermined symmetry and representing the logarithm of the frequency spectrum magnitudes of said second speech signal;

summing means coupled to said means for converting and to said means for generating and being operative to sum said signals having a first predetermined symmetry and said signals having a second predetermined symmetry to form a com-

posite signal;

computing means having an input connection coupled to said summing means and first and second output connections, said computing means being operative to compute a first and second set of 5 signals representing the complex Fourier transform of said composite signal, said first set of signals having said first predetermined symmetry and being directed to said first output connection and said second set of signals having said second 10 predetermined symmetry and being directed to said second output connection;

convolution means coupled to said means for generating a pitch carrier signal and to said first output connection of said computing means and 15 being operative to combine in a predetermined manner said pitch carrier signal and said first set of signals having said first predetermined symmetry to thereby generate a synthesized version of said 20 first speech signal;

pitch detection means coupled to said second output connection of said computing means and being operative to extract the voicing information of said second speech signal from said second set of 25 signals having said second predetermined symmetry; and

selection means coupled to said second output connection of said computing means and being operative to select a predetermined number of said set 30 of signals having said second predetermined symmetry, said selected signals representing the vocal tract information of said second speech signal.

26. A vocoder system for synthesizing a first speech signal and analyzing a second speech signal simultane- 35 ously with said first and second speech signals including voicing and vocal tract information, said digital vocoder system comprising:

means for generating a pitch carrier signal from said first speech signal; 40

means for obtaining the frequency spectrum magnitudes of said first speech signal;

means coupled to said means for obtaining the frequency spectrum magnitude of said first speech signal for converting the frequency spectrum mag- 45 nitudes into signals having a first predetermined symmetry;

computing means having first and second input ports and first, second, third and fourth output ports and 50 being operable to compute simultaneously the Fourier transform of a set of first predetermined input signals at said first input ports, said set of first predetermined signals having a composite symmetry of said first and second predetermined sym- 55 metries and the Fourier transform of a set of second predetermined input signals at said second input port, said set of second predetermined signals having first and second predetermined symmetries and operable to direct to said first, second, 60 third and fourth output ports respectively a first set of output signals representing the Fourier transform of the portion of the set of first predetermined input signals having the second predetermined symmetry, a second set of output signals 65 representing the Fourier transform of the portion of the set of first predetermined input signals hav-

ing the first predetermined symmetry, a third set of output signals having the first predetermined symmetry and representing the Fourier transform of the portion of the set of second predetermined input signals at said second input port and a fourth set of output signals representing the Fourier transform of the portion of the set of second predetermined input signals having the second predetermined symmetry;

sampling means having an output connection coupled to said first input port of said computing means and being operable to sample said second speech signal over a first predetermined time interval, said first and second sets of output signals of said computing means representing the spectrum of said sampled second input speech signal;

magnitude means coupled to said first and second output ports of said computing means and being operative to combine in a predetermined manner said first and second sets of output signals of said computing means to generate signals representing the frequency spectrum magnitudes of said second speech signal;

means coupled to said magnitude means for generating output signals having a second predetermined symmetry and representing the logarithm of the frequency spectrum magnitudes of said second speech signal;

summing means having input connections coupled to said means for converting and to said means for generating and an output connection coupled to said second input port of said computing means and being operative to sum said signals having a first predetermined symmetry with said signals having said second predetermined symmetry to form said set of second predetermined input signals,

whereby said third set of output signals of said computing means represents the vocal tract information of said first speech signal and said fourth set of output signals of said computing means is the Fourier transform of the logarithm of the spectrum magnitudes representing the voicing and vocal tract data of said second speech input signal;

pitch detection logic means coupled to the fourth output port of said computing means and being operative to extract a pitch signal from the fourth set of output signals of said computing means to thereby represent the voicing information of said second input speech signal;

selecting means coupled to the fourth output port of said computing means and being operative to select a predetermined number of the fourth set of output signals to represent the vocal tract information of said second input speech signal; and

convolution means coupled to said means for generating a pitch carrier signal from said first speech signal and to the third output port of said computing means and being operative to combine in a predetermined manner the pitch carrier signals with the third set of output signals of said computing means to synthesize the first speech signal.

27. A vocoder system according to claim 26 wherein said means for generating a pitch carrier signal in-

cludes:

first means responsive to said first speech signal for
generating a first predetermined pitch carrier
signal when the magnitude of the pitch signal in-
dicates a voiced signal;

second means responsive to said pitch signal from
said pitch detection logic means for generating a
second predetermined pitch carrier signal when
the magnitude of the pitch signal indicates an un-
voiced signal; and

gating means coupled to said first and second means
for generating and being operative to gate a first
predetermined pitch carrier signal to said convolu-
tion means when the magnitude of the pitch signal
is less than a predetermined magnitude and to gate
a second predetermined pitch carrier signal to said
convolution means when the magnitude of the
pitch signal is greater than a predetermined mag-
nitude.

28. A vocoder system according to claim 27 wherein
said first means for generating includes:

third means for generating signals, the magnitudes of
which describe a predetermined function;

fourth means for generating signals, the magnitudes
of which describe the slope of a line connecting
the magnitudes of the voiced information of two
successive first input signals; and

first comparator means having input connections
coupled to said third and fourth means for generat-
ing and an output connection coupled to said gat-
ing means of said means for generating a pitch car-
rier signal and being operative to generate a first
predetermined pulse when the signals from said
fourth means for generating are equal to or greater
than the magnitude of the signal from said third
means for generating.

29. A vocoder system according to claim 28 includ-
ing an inhibiting means responsive to said voicing infor-
mation of said first input speech signal to inhibit the
second predetermined pitch carrier signal of said
second means for generating when the voicing informa-
tion exceeds a predetermined magnitude.

30. A vocoder system according to claim 29 wherein
said third means for generating signals includes:

first storage counter means having a first input con-
nection and an output connection and being
operative to store a first predetermined signal, to
add to said first predetermined signal a second
predetermined signal appearing at said first input
connection and to supply the resultant signal to
said output connection;

slope means for generating a third predetermined
signal; and

first summation means having a first input connec-
tion coupled to the output connection of said first
storage counter means, a second input connection
coupled to said slope means and an output connec-
tion coupled to said first input connection of said
storage counter means and to said gating means of
said means for generating a carrier generator, said
first summation means being operative to add the
resultant signal of said first storage counter means
to the third predetermined signal from said slope
means to form said second predetermined signal
and to direct said second predetermined signal

simultaneously to said gating means of said means
for generating a pitch carrier and to said first
storage counter means to update said first
predetermined signal stored therein.

31. A vocoder system according to claim 30 wherein
said fourth means for generating signals includes:

means for computing a slope signal m wherein $m = T_p(n-1) - T_{pn}/T$, where T_p is a first voicing signal
received from said first input speech signal at a
first predetermined time, $T_p(n-1)$ is a voicing
signal received from said first input speech signal
at a second predetermined time and T is the
elapsed time between said first and second
predetermined times;

second storage counter means having a first input
connection and an output connection and being
operative to store a first predetermined signal, to
add to said first predetermined signal a second
predetermined signal appearing at said first input
connection and to supply the resultant signal to
said output connection; and

second summation means having a first input con-
nection coupled to the output connection of said
second storage counter means, a second input con-
nection coupled to said means for computing a
slope signal and an output connection coupled to
said first input connection of said second storage
counter means and to said gating means of said
means for generating a pitch carrier signal, said
second summation means being operative to add
the resultant signal of said second storage counter
means to the slope signal from said means for com-
puting a slope signal to form said second predeter-
mined signal and to direct said second predeter-
mined signal to said gating means of said means for
generating a pitch carrier signal and to said second
storage counter means to update said first
predetermined signal stored therein.

32. A vocoder system according to claim 25 wherein
said means for obtaining the frequency spectrum mag-
nitude of said first speech signal includes:

Fourier transform computer means operable to solve
the expression

$$\log |V_n| = \sum_{k=0}^{R} C_k \cos \left(\frac{2\pi nK}{N} \right)$$

where V_n is the nth frequency sample of said first speech
signal, C_k is the kth sample of said first speech signal and
k and R are predetermined limits of summation; and

delogging computer means operative to obtain the
antilogarithm of said expression to yield the
frequency spectrum magnitude of said first speech
signal.

33. A vocoder system according to claim 26 wherein:

said first and second input ports of said computing
means are real and imaginary input ports respec-
tively;

said first and second predetermined symmetries of
said set of first predetermined input signals are
even and odd symmetries respectively;

said set of first predetermined input signals includes
256 samples of said input speech signal;

said first set of output signals at said first output port
of said computing means includes 128 samples
having even symmetry and representing the Fouri-

er transform of the even portion of the 256 samples at the real input port of said computing means;

said second set of output signals at said second output port of said computing means includes 128 samples representing the Fourier transform of the portion of the 256 input samples at said real input port having odd symmetry, said first and second sets of output signals representing, respectively, the real and imaginary parts of the frequency spectrum of the 256 samples of the second input speech signal at the real input port of said computing means;

said set of second predetermined input signals at said imaginary input port of said computing means includes 256 samples having even and odd symmetry associated therewith, said even symmetry portion representing the logarithm of the spectrum magnitudes of the second input speech signal and said odd symmetry portion representing the frequency spectrum of the first input speech signal;

said third set of output signals at the third output port of said computing means includes 128 samples having odd symmetry and representing the Fourier transform of the odd symmetry portion of 256 samples at said imaginary input port of said computing means, said 128 samples at said third output port of said computing means represents the vocal tract information of said first speech signal; and

said fourth set of output signals at the fourth output port of said computing means includes 128 samples having even symmetry and representing the Fourier transform of the logarithm of the spectrum magnitudes from which the vocal tract and the voicing information of the second input speech signal are derived.

34. A vocoder system according to claim 33 wherein said pitch detection logic means includes:

selection means having an input connection coupled to said fourth output port of said computing means and being operative to select the sample of said fourth set of output signals having the largest magnitude;

first comparator means having an input connection coupled to said selection means and a first and second output connection, said first comparator means being operative to compare the magnitude of the selected output signal of said selection means to a predetermined threshold level and to generate an output signal at said first output connection if the magnitude of said selected sample exceeds the predetermined threshold level and to generate a predetermined output signal at said second output connection if the magnitude of said selected sample is less than the predetermined threshold level; and

buffer storage means having a first input connection connected to the common juncture of said selection means and said first comparator means, a second input connection connected to the first output connection of said first comparator means and an output terminal and being operative to store the selected sample from said selection means and to shift the stored sample to said output terminal upon receipt of a signal from said first comparator means,

whereby an unvoiced second speech signal is indicated when said first comparator means has an output signal at said second output connection and a voiced signal is indicated when the fourth output signal of said computing means is shifted to the output of said buffer storage means.

35. A vocoder system according to claim 33 wherein said convolution unit includes:

logic means having a first input connection coupled to said means for generating a pitch carrier signal, a second input connection coupled to the third output port of said computing means and first, second, third and fourth output connections, said logic means being operative in response to a first predetermined time period to provide a data path from said first and second input connections to said first and second output connections, respectively, and being operative in response to a second predetermined time period to provide a data path from said first and second input connections to said third and fourth output connections respectively;

first storage means having first and second input connections coupled respectively to said first and second output connections of said logic means and a plurality of output connections, said first storage means being operative to store the output signals representing the vocal tract information received from the third output port of said computing means via the data path established by said logic means during said first predetermined time period and to gate from a different one of said plurality of output connections a complete set of vocal tract signals upon the receipt of each signal from said means for generating a pitch carrier signal during said first predetermined time period;

second storage means having first and second input connections coupled respectively to said third and fourth output connections of said logic means and a plurality of output connections, said second storage means being operative to store the output signals representing the vocal tract information received from the third output port of said computing means via the data path established by said logic means during said second predetermined time period and to gate from a different one of said plurality of output connections a complete set of vocal tract signals upon receipt of each signal from said pitch carrier generator during said second predetermined time period; and

summing means having a plurality of input connections each coupled to one of said plurality of output connections of said first and second storage means and being operative to add the vocal tract signals from said first and second storage means whereby a synthesized version of the first input speech signal is obtained.

36. A method of compressing the bandwidth of an input signal having an excitation portion and an impulse response portion comprising the steps of:

generating a time variant electrical signal representing the Fourier transform of the logarithm of the spectrum magnitude of the input signal;

separating out a first time interval signal of said time variant electrical signal to represent the impulse

response portion of the input signal; and

separating out a second time interval signal of said time variant electrical signal to represent the excitation portion of the input signal, said first and second time interval signals of said time variant electrical signal having a reduced bandwidth.

37. A method of simultaneously synthesizing a first speech signal and analyzing a second speech signal, said first and second speech signals including voicing and vocal tract data, said method comprising the steps of:

generating a pitch carrier signal from said first speech signal;

generating the frequency spectrum magnitude signals of said first speech signal;

converting the frequency spectrum magnitude signals into signals having a first predetermined symmetry;

generating the frequency spectrum magnitude signals of the second speech signal;

converting the frequency spectrum magnitude signals of the second speech signal into a series of signals having a second predetermined symmetry and representing the logarithm of the frequency

spectrum magnitudes of said second speech signal;

combining the signals having the first predetermined symmetry with the series of signals having the second predetermined symmetry to generate a series of composite signals;

generating from said series of composite signals first and second sets of signals representing the complex Fourier transform of the composite signal, said first set of signals having said first predetermined symmetry and said second set of signals having said second predetermined symmetry;

combining the pitch carrier signal from said first speech signal and said first set of signals having said first predetermined symmetry to thereby generate a synthesized version of the first speech signal;

selecting a predetermined number of said second set of signals to represent the vocal tract data of said second input speech signal; and

selecting a predetermined number of the remaining signals of said second set of signals to represent the voicing information of said second input speech signal.

* * * * *