

Feb. 3, 1970

J. M. KELLY ET AL

3,493,684

VOCODER EMPLOYING COMPOSITE SPECTRUM-CHANNEL AND PITCH ANALYZER

Filed June 15, 1966

2 Sheets-Sheet 1

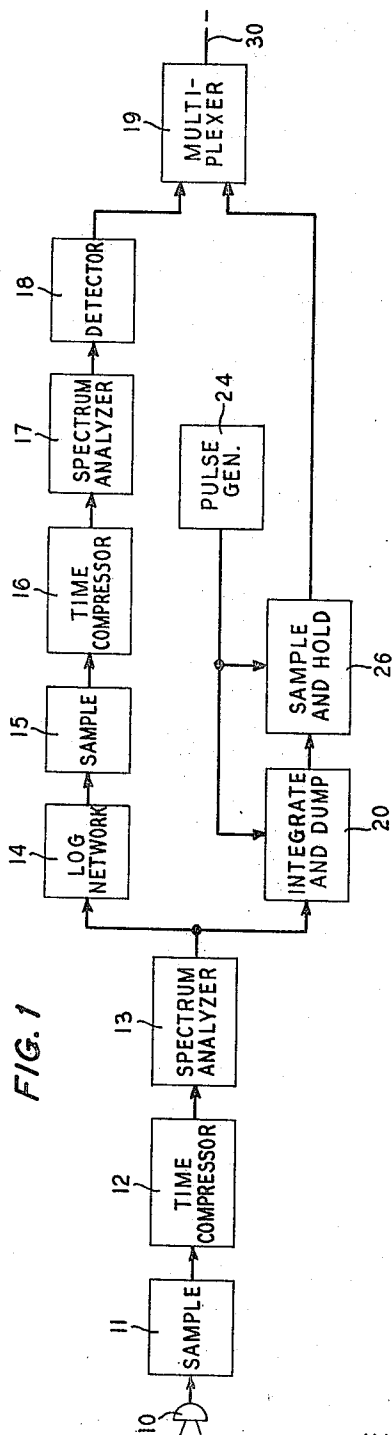


FIG. 1

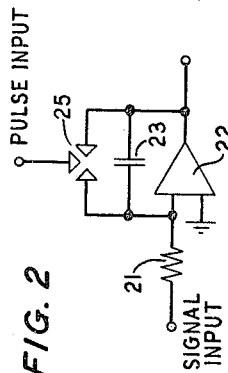


FIG. 2

INVENTORS
J. M. KELLY
A. M. NOLL
M. R. SCHROEDER
BY *G. E. Hirsch*
ATTORNEY

Feb. 3, 1970

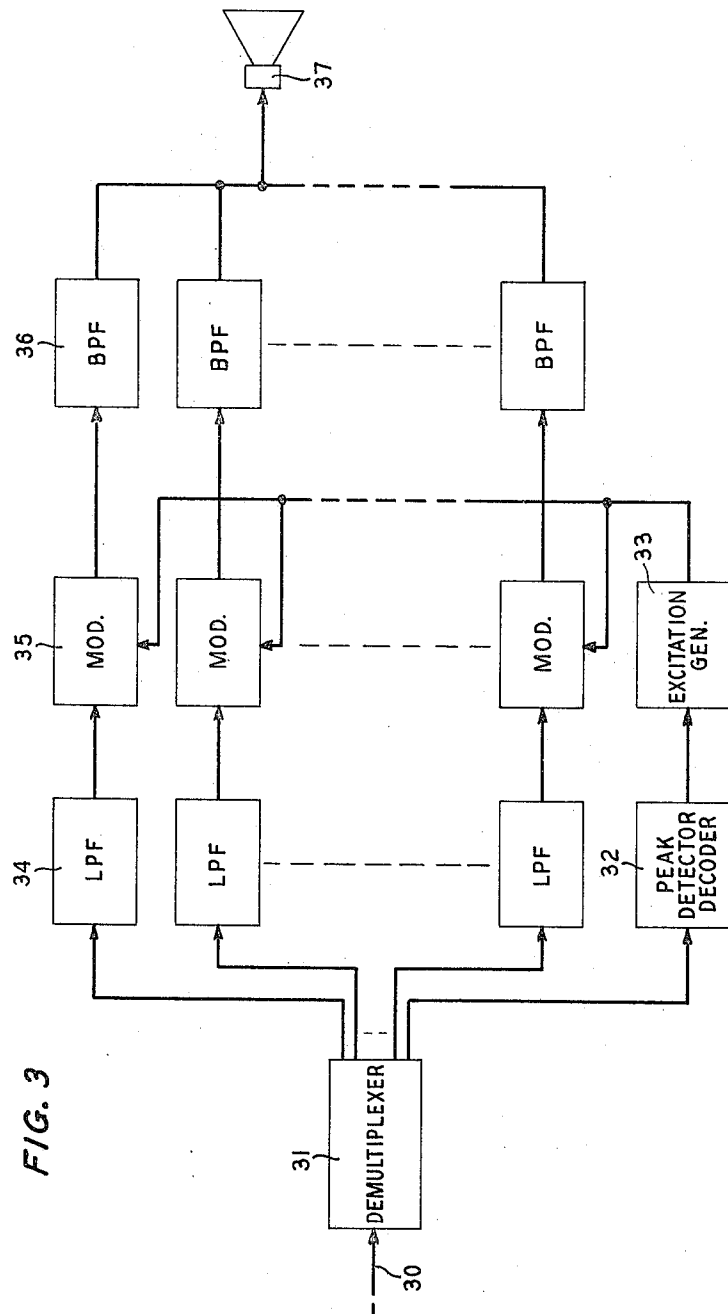
J. M. KELLY ET AL

3,493,684

VOCODER EMPLOYING COMPOSITE SPECTRUM-CHANNEL AND PITCH ANALYZER

Filed June 15, 1966

2 Sheets-Sheet 2



1

3,493,684

VOCODER EMPLOYING COMPOSITE SPECTRUM-CHANNEL AND PITCH ANALYZER

James M. Kelly, Morris Plains, A Michael Noll, Newark, and Manfred R. Schroeder, Gillette, N.J., assignors to Bell Telephone Laboratories, Incorporated, Murray Hill and Berkeley Heights, N.J., a corporation of New York

Filed June 15, 1966, Ser. No. 557,682

Int. Cl. H04b 1/66; H04m 1/00

U.S. Cl. 179-15.55

7 Claims

ABSTRACT OF THE DISCLOSURE

The construction and performance of a vocoder system are improved by performing a short-time analysis of applied speech signals, and using signals produced in the analysis both for developing a "cepstrum" pitch signal and spectrum envelope information for use in synthesizing the applied signal.

This invention relates to the transmission of speech signals over narrow band media by vocoder techniques. One of its principal objectives is to reduce the channel bandwidth required for such transmission. Another objective is to simplify and improve vocoder processing apparatus through which such transmission is carried out.

Narrow band speech transmission systems, such as the channel vocoder described by H. W. Dudley in Patent 2,151,091, Mar. 21, 1939, transmit the information content of wide band speech waves over a narrow band channel by analyzing an incoming speech wave to determine its significant characteristics, and by transmitting information regarding these characteristics, instead of the speech wave itself, to a distant receiver station. One important characteristic of speech extracted for coded transmission by a vocoder analyzer is the spectral energy distribution of the input signal; that is, the distribution signal energy in each of a number of relatively narrow, substantially contiguous, frequency subbands. Another, and perhaps the most important, of the speech characteristics is the so-called pitch characteristic. Complete specification of the pitch characteristic in a vocoder system requires information signifying whether the incoming speech wave at a particular instant represents a voiced or an unvoiced sound, and if the sound represented is voiced, information regarding either its fundamental frequency or the reciprocal, its fundamental period. In a typical vocoder system, representations of spectral energy information and pitch information are coded for transmission to a receiver station. At the receiver station natural sounding speech is synthesized from the information conveyed by the transmitted signals.

Difficulties encountered in determining the fundamental frequency, or pitch, of a speech signal, which is necessary for generating an excitation signal at the receiver synthesizer, have long prevented vocoder systems from gaining widespread use in speech communication systems. However, a pitch detector based on the identification of periodicities in the logarithmic power spectrum of a signal has recently been developed. Such a detector makes possible accurate, real time, pitch detection. The detection system is described in an application of A. M. Noll and M. R. Schroeder Ser. No. 420,365, filed Dec. 22, 1964.

In essence, the periodicity and aperiodicity of a complex wave are determined, in accordance with the Noll-Schroeder invention, with a high degree of accuracy by performing two successive spectral analyses. The first analysis is performed upon a selected segment of a complex wave to obtain a first so-called short-time spectrum,

2

while the second analysis is performed upon a waveform representing the logarithm of the first short-time spectrum to obtain a second short-time spectrum. The second short-time spectrum obtained in this manner is also referred to as a "cepstrum," which, as described by A. M. Noll in "Short-Time Spectrum and 'Cepstrum' Techniques for Vocal-Pitch Detection," Vol. 36 Journal of the Acoustical Society of America, page 296 (1964), is simply an abbreviation for the phrase "short-time spectrum of the logarithm of a short-time spectrum."

Voiced and unvoiced intervals are respectively detected by examining each second short-time spectrum for the presence or absence of a single large peak exceeding a predetermined threshold. When a voice sound interval is indicated, by the presence of a single large peak, the fundamental period of the sound is obtained by measuring the time of occurrence of the single large peak in each second short-time spectrum. Apparatus for carrying out the necessary examination of the second short-time spectrum is described in an application of A. M. Noll Ser. No. 508,726, filed Nov. 19, 1965, now Patent 3,420,955, granted Jan. 7, 1969.

In a typical vocoder application, a cepstrum detector is substituted directly for the pitch determining apparatus previously employed. The spectral energy characteristic of the signal is obtained for the modified vocoder system in the conventional manner. For example, it is obtained as shown in the Dudley patent by passing the incoming signal through a bank of filters with contiguous pass bands, and individually detecting the energy level of each subfrequency channel.

It is in accordance with the present invention to simplify the construction and improve the performance of a vocoder analyzer employing a cepstrum detector channel by eliminating the channel bank of bandpass filters and energy detectors commonly used for developing the spectral energy distribution of a signal.

According to the invention, a short-time analysis of applied speech signals is performed, and signals defining the analysis are employed both in the development of a cepstrum pitch signal and in the development of spectrum envelope information needed in the synthesis process at the receiver. In essence, the short-time analyzer serves both as the first analyzer of the cepstrum detector and as a replacement for the usual bank of bandpass filters employed for developing channel information. Thus, the short-time spectrum signal is delivered to a logarithmic network and then to a second short-time spectrum analyzer to develop a cepstrum signal. It is also delivered to apparatus which averages the spectrum over a number of periods, each of which corresponds substantially to the bandwidth of one filter in an analyzer filter bank, to develop the channel signals. The latter operation may be accomplished in a number of ways, e.g., by sampling the short-time spectrum and averaging selected groups of signals, or preferably by means of a so-called integrate-and-dump circuit.

The invention will be more fully understood from the following detailed description of an illustrative embodiment thereof taken in connection with the appended drawings in which:

FIG. 1 is a schematic block diagram of a vocoder analyzer embodying the principles of the invention;

FIG. 2 is a schematic diagram of a typical integrate-and-dump circuit useful in the practice of the invention; and

FIG. 3 is a schematic block diagram of a vocoder synthesizer which may be employed to reconstruct speech signals from coded information supplied from the analyzer illustrated in FIG. 1.

A transmitter station employing a vocoder analyzer embodying the general principles of the present invention

is illustrated in FIG. 1. Speech signals are converted by transducer 10, for example, a conventional microphone, into time-varying waves which are thereupon sampled, as desired, in sampler 11 to produce a sequence of relatively short pulses whose amplitudes indicate the amplitude of the corresponding speech signal. In a typical system supplied with speech signals band limited to approximately 3 kHz., sampling once every 125 microseconds, i.e., at an 8 kHz. rate, is satisfactory.

Since a first spectrum analysis must provide adequate resolution for the pitch extraction process, and is preferably performed in real time to follow the variations in pitch frequency occurring in normal speech, the required frequency resolution of the first spectrum analyzer is less than or equal to one half the lowest pitch frequency. Such resolution requires a time interval of speech for analysis greater than or equal to twice the highest pitch period of the signal. Therefore, the spectrum should be specified at N_s points, where N_s is equal to the bandwidth of the input signal divided by the desired resolution. Specification with this degree of resolution is conveniently accomplished by compressing the sampled input signal in time by a factor N_s prior to analysis. Time compressor apparatus 12 is employed for this purpose. It may take any desired form; digital delay line time compressor apparatus is suitable. In a typical system supplied with a 3 kHz. signal sampled at an 8 kHz. rate, regardless of the spectrum analyzer technique employed, it is desirable that a complete spectrum be produced approximately every 12.5 milliseconds.

The time compressed signal developed in apparatus 12 is applied to the input terminal of spectrum analyzer 13 where, in effect, the frequency of the time compressed signal is swept across a bandpass filter with a suitable time response. The envelope of the filter output as a function of frequency represents the spectrum of the time compressed interval of speech. Circuit parameters may be selected to allow processing in real time to follow pitch variations. Accordingly, apparatus 13 provides at its output a first short-time spectrum signal representative of the short-time power spectrum of the applied signal.

Spectrum analyzer apparatus 13 may take any desired form. A suitable analyzer is described in volume 18 of the Journal of the Acoustical Society of America at page 19 (1947), as well as in the aforementioned Noll-Schroeder application.

The short-time spectrum output signal of analyzer 13 is delivered to two parallel channels; the first for developing the spectrum of the logarithm of the first spectrum, and the second for developing the power spectrum of the input speech signal. Considering the cepstrum channel first, the spectrum representation of pitch is obtained, in the fashion described in the Noll-Schroeder application, by seeking peaks in the short-time spectrum analysis of the logarithm of the first short-time spectrum of the signal applied from source 10. Accordingly, the short-time spectrum signal from analyzer 13 is delivered to logarithmic network 14, of any well known design, to develop a wave that represents the logarithm of the short-time spectrum signal. From network 14 the logarithmic wave is supplied by way of sample network 15 and time compressor 16 to spectrum analyzer 17. Networks 15, 16, and 17 may be identical in all respects to networks 11, 12 and 13, discussed above. The second spectrum signal, available at the output of spectrum analyzer 17, contains a well defined peak when the first spectrum contains a periodic component. The location of this peak in the second spectrum is a direct measure of the pitch period. Absence of such a peak indicates the absence of a periodic component in the speech spectrum and, therefore, the absence of a pitch frequency in the speech signal. This denotes an unvoiced sound. Accordingly, the detection and measurement of the pitch period of the applied signal is accomplished by peak detector apparatus 18, which may be of the type described in the Noll application cited

above. In general, voiced and unvoiced intervals are detected by examining each second short-time spectrum for the presence or absence of a single large peak exceeding a predetermined threshold. Further, when a voiced sound interval is indicated by the presence of a single large peak, the fundamental period of the sound is obtained by measuring the time of occurrence of the single large peak in each second short-time spectrum. The output of detector 18 thus identifies voiced and unvoiced intervals at the fundamental period of voiced intervals. Signals from detector 18 are employed as one component of the analyzer specification of the signal. This composite signal may, for example, be delivered to multiplex apparatus 19 wherein it is combined with other signal specifications for delivery by way of a communications channel to a receiver station.

The output of first spectrum analyzer 13 is also employed to produce the spectrum envelope, or channel information, required for synthesizing speech at a receiver station. In accordance with this invention, this information is obtained by averaging the spectrum of the applied signal available at the output of analyzer 13 over a period corresponding to the bandwidth of one filter in an analyzer filter bank. In a conventional vocoder employing such a filter bank, a sampling rate of approximately 40 samples per second is sufficient to describe the speech spectrum envelope variation. Since a complete spectrum is obtained in a typical system each 12.5 milliseconds, or 80 times per second, every other spectrum signal from analyzer 13 is employed to develop channel information.

In practice, spectrum signals from analyzer 13 are applied to integrate-and-dump network 20 wherein successive values, each corresponding to one sample of the output of one of N spectrum channel envelope detectors, are produced. A suitable form of integrate-and-dump network is illustrated in FIG. 2. In essence, incoming signals are delivered by way of resistor 21 to operational amplifier 22 connected in an integrating configuration. Input signal values are integrated over a period determined by the time constant of resistor 21 and capacitor 23. Periodically, the stored value available at the output of the network is "dumped" by discharging capacitor 23. This may be done by applying a pulse from generator 24 (FIG. 1) to normally-open switch 25 which bridges capacitor 23 and amplifier 22. With one speech spectrum applied from analyzer 13 every 12.5 ms., a vocoder analyzer with an equivalent of 15 channels requires the integrated spectrum to be dumped at a 1200 Hz. rate (15 channels $\times$ 1/12.5 ms.). Pulse generator 24 thus emits pulses at a 1200 Hz. rate.

During the integrating period, the output of network 20 is applied to sample-and-hold network 26 which, under control of pulses from generator 24, samples the applied value and holds it over the sampling interval. The output of sample-and-hold network 26 constitutes N ($N=15$ in the example given above) analog signals representative of the energy contained in N spectrum channels. This information is applied, for example, to multiplexer 19 for transmission to a receiver station.

Alternatively, the channel spectrum information may be developed by sampling the analyzer signals delivered by analyzer 13 at a periodic rate, and by averaging selected groups of signals. Each average group represents one channel signal.

Control signals developed in the analyzer illustrated in FIG. 1 typically are transmitted over transmission channel 30, generally with a reduced bandwidth, to a receiver station which includes a speech synthesizer. Such a station is illustrated in FIG. 3 and is in all respects conventional. At the receiver, signals are separated in demultiplexer apparatus 31. The coded pitch control signal is delivered by way of peak detector decoder 32 to excitation generator 33. These units cooperate to generate a suitable excitation signal for synthesis. Both decoder 32 and generator 33 may be of the form described in

McDonald Patent 3,109,142, issued Oct. 29, 1963. Channel control signals delivered from demultiplexer 31 are delivered in parallel by way of low-pass filters 34 to modulators 35. For the example of practice cited above, i.e., spectrum sampling at a rate of 40 per second, each of filters 34 has a passband of approximately 20 Hz. Excitation signals from generator 33 are similarly delivered to modulators 35 so that, in conventional fashion, the modulators deliver reconstructed channels of the input speech to bandpass filters 36. These channel signals are combined and delivered to reproducer 37, which may be a conventional loudspeaker. It converts the replica speech wave into audible sound.

It is apparent that the principles of the invention may be embodied in apparatus which differs in construction from that illustrated herein. For example, all of the operations may be carried out entirely on an analog basis, entirely on a digital basis, or on a mixed basis of analog and digital operations. In either case, a considerable saving in equipment and control is afforded since the first spectrum analyzer serves both as a source of excitation for the cepstrum pitch detector channel and for the spectrum envelope channels.

Optical techniques may, if desired, be employed for carrying out the short-time spectrum and cepstrum analyses. Such techniques are based upon the Fourier transform properties of a lens but differ in implementation in dependence upon the nature of illumination employed—i.e., coherent or noncoherent. For example, a coherent light source, e.g., a laser, may be employed for illuminating a thermoplastic tape on which the signal to be analyzed has been written. A noncoherent light source, such as a cathode ray beam, may be used to display the signal to be analyzed on the face of a cathode ray tube. Conventional amplitude, width, or intensity modulation of the light source, may be used for displaying the signal for either method.

In any event, it is to be understood that the above described embodiments of the invention are merely illustrative of the numerous arrangements that may be devised for the principles of the invention by those skilled in the art without, however, departing from the spirit and scope of the invention.

What is claimed is:

1. A vocoder analyzer which comprises:

means supplied with speech signals for deriving from each of a selected number of segments thereof a first short-time spectrum waveform;
means responsive to said first short-time waveform for developing a control signal that represents the pitch of said speech signal segment,
said means for developing a pitch representative control signal comprising,
means for converting said first spectrum waveform into a signal representing the logarithm of said first short-time spectrum waveform,
means supplied with said logarithmic signal for deriving therefrom a short-time spectrum waveform of the logarithm of said first short-time spectrum waveform,
means for deriving a control signal for each spectrum waveform of the logarithm of said first spectrum waveform which exhibits a peak that exceeds a predetermined threshold level, and
means for developing, from each spectrum waveform of the logarithm of said first spectrum waveform which exhibits a peak that exceeds said threshold level, a control signal that represents the time of occurrence of the largest peak that exceeds said threshold level;
means responsive to said first short-time spectrum waveform for developing a plurality of control signals representative of the spectral energy characteristic of said speech signal segment; and
means for utilizing all of said control signals developed

for each of said selected segments together as a representative of said speech signal.

2. A vocoder analyzer which comprises:

means supplied with speech signals for deriving from each of a selected number of segments thereof a first short-time spectrum waveform;
means responsive to said first short-time waveform for developing a control signal that represents the pitch of said speech signal segment;
means responsive to said first short-time spectrum waveform for developing a plurality of control signals representative of the spectral energy characteristic of said speech signal segment,
said means for developing spectral energy control signals comprising,
means for developing samples of said first short-time spectrum waveform, and
means for averaging selected ones of said samples to develop a plurality of channel control signal; and
means for utilizing all of said control signals developed for each of said selected segments together as a representative of said speech signal.

3. A vocoder analyzer which comprises,

means supplied with speech signals for deriving from each of a selected number of segments thereof a first short-time spectrum,
means for converting said first spectrum into a signal representing the logarithm of said first short-time spectrum,
means supplied with said logarithmic signal for deriving therefrom a short-time spectrum of the logarithm of said first short-time spectrum,
means for deriving a control signal for each spectrum of the logarithm of said first spectrum which exhibits a peak that exceeds a predetermined threshold level,
means for developing, for each spectrum waveform of the logarithm of said first spectrum waveform which exhibits a peak that exceeds said threshold level, a control signal that represents the time of occurrence of the largest peak that exceeds said threshold level,
means for averaging selected portions of said first short-time spectrum over selected periods to develop a plurality of control signals, and
means for utilizing all of said control signals together as a representation of said speech signals.

4. A vocoder analyzer as defined in claim 3 wherein,

said means for averaging selected portions of said first short-time spectrum over selected periods to develop a plurality of control signals comprises,
a signal generator,
means responsive to signals from said generator for developing and holding brief samples of said first short-time spectrum, and
means for developing a plurality of control signals, each representing the average of a selected plurality of said samples.

5. A vocoder analyzer as defined in claim 3 wherein,

said means for averaging selected portions of said first short-time spectrum over selected periods to develop a plurality of control signals comprises,
a pulse generator,
means responsive to selected pulses from said generator for continuously integrating said first short-time spectrum over a first selected time interval,
means responsive to selected pulses from said generator for sampling the value of the integrated spectrum at the end of each of said time intervals, and
means for utilizing all of the sampled values developed in a second selected time interval as a control signal representative of the spectral energy characteristic of said speech signal.

6. A vocoder analyzer which comprises,

means supplied with speech signals for deriving from each of a selected number of segments thereof a first

7

plurality of samples representing selected values of a first short-time spectrum of said segment,
 means for converting said first plurality of samples into a signal representing the logarithm of said first short-time spectrum,
 means supplied with said logarithmic signal for deriving therefrom a second plurality of samples representing selected values of a short-time spectrum of the logarithm of said first short-time spectrum,
 means for deriving a first control signal for each spectrum of the logarithm of said first spectrum which exhibits a peak that exceeds a predetermined threshold level.
 means for developing, for each spectrum waveform of the logarithm of said first spectrum waveform which exhibits a peak that exceeds said threshold level, a second control signal that represents the time of occurrence of the largest peak that exceeds said threshold level,
 means for integrating said first short-time spectrum over selected periods,
 means for selectively sampling said integrated spectrum to develop a third control signal, and
 means for developing from said first, said second and said third control signals a coded representation of said speech signals.
 7. In a vocoder analyzer in which the pitch characteristic of an applied signal is obtained by apparatus which includes,
 means for deriving a first short-time spectrum of selected segments of an applied signal,
 means for converting said first spectrum into a signal

8

representing the logarithm of said first short-time spectrum, and
 means for deriving a short-time spectrum from said logarithmic signal,
 the improvement which comprises,
 channel information means responsive to said first short-time spectrum for developing spectrum envelope signals representative of the energy in contiguous subbands of said applied signal,
 said channel information means comprising,
 means for developing samples of said first short-time spectrum at a selected rate, and
 means for averaging selected samples to develop a plurality of channel control signals.

References Cited

UNITED STATES PATENTS

3,071,652	1/1963	Schroeder	179—15.55
3,109,070	10/1963	David et al.	179—15.55
3,361,877	1/1968	Kreer	179—15.55

OTHER REFERENCES

Noll: "Short-Time Spectrum and 'Cepstrum' Techniques for Vocal-Pitch Detection," February 1964, 36 Journ. of Acous., Soc. of Amer. 296-302.

JOHN W. CALDWELL, Primary Examiner

J. A. BRODSKY, Assistant Examiner

U.S. Cl. X.R.

179—1