

[54] APPLICATION OF SIMULTANEOUS VOICE/UNVOICE EXCITATION IN A CHANNEL VOCODER

[75] Inventor: David C. Coulter, Vienna, Va.

[73] Assignee: The United States of America as represented by the Secretary of the Navy, Washington, D.C.

[22] Filed: Apr. 23, 1974

[21] Appl. No.: 463,339

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

[51] Int. Cl.<sup>2</sup> .... G10L 1/00

[58] Field of Search ..... 179/1 SA, 1 SG, 1 SM

[56] References Cited

UNITED STATES PATENTS

2,243,526	5/1941	Dudley.....	179/1 SA
3,078,345	2/1963	Campanella et al. ....	179/1 SA
3,190,963	6/1965	David.....	179/1 SA
3,431,362	3/1969	Miller .....	179/1 SA
3,649,765	3/1972	Rabiner et al. ....	179/1 SA
3,715,512	2/1973	Kelly.....	179/1 SA

Primary Examiner—Kathleen H. Claffy  
 Assistant Examiner—Thomas D'Amico  
 Attorney, Agent, or Firm—R. S. Sciascia; Arthur L. Branning; George A. Montayne

[57] ABSTRACT

An improved technique and apparatus for increasing the quality of voice transmission through vocoders, particularly channel vocoders. The vocoder is modified to permit the simultaneous transmission of both voiced and unvoiced portions of a speech sound by connecting the pitch source to the lowest channel at all times, and including an additional decision circuit to independently transmit the voiced/unvoiced decision. With this connection, it is possible to reproduce sounds such as the voiced fricatives which contain simultaneous voiced and unvoiced portions. In addition, the threshold for voicing can be adjusted for less sensitivity while providing for low level voicing such as nasal murmurs and voice bars associated with stop consonants, or plosive unvoicing.

8 Claims, 2 Drawing Figures

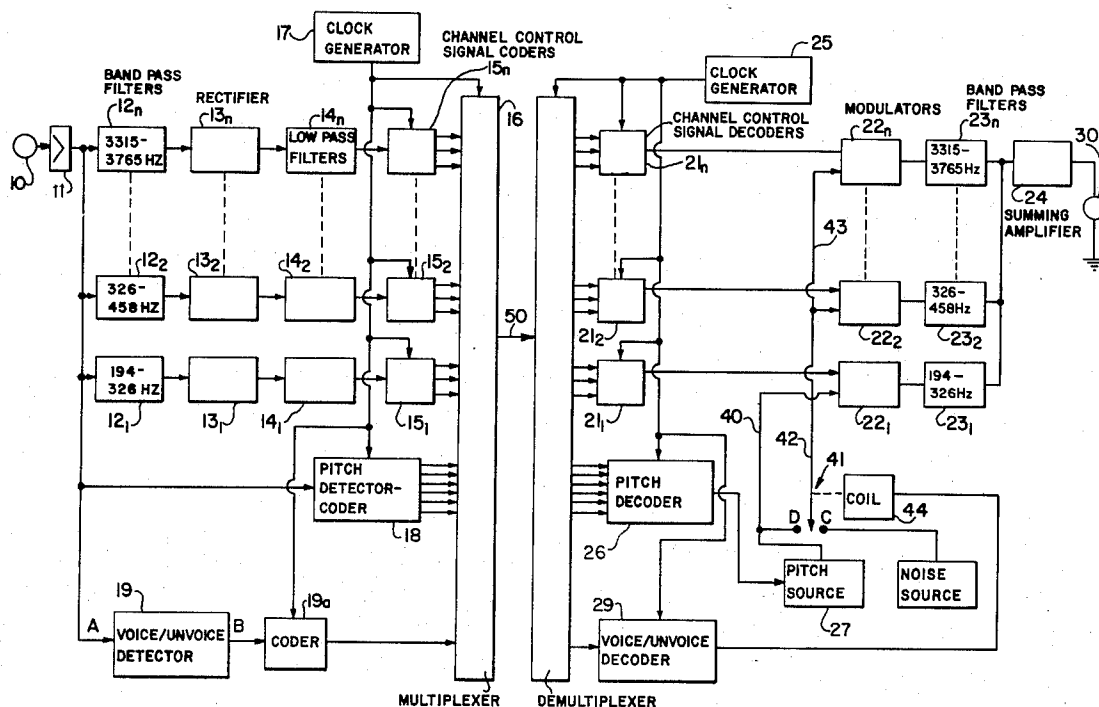


FIG. 1

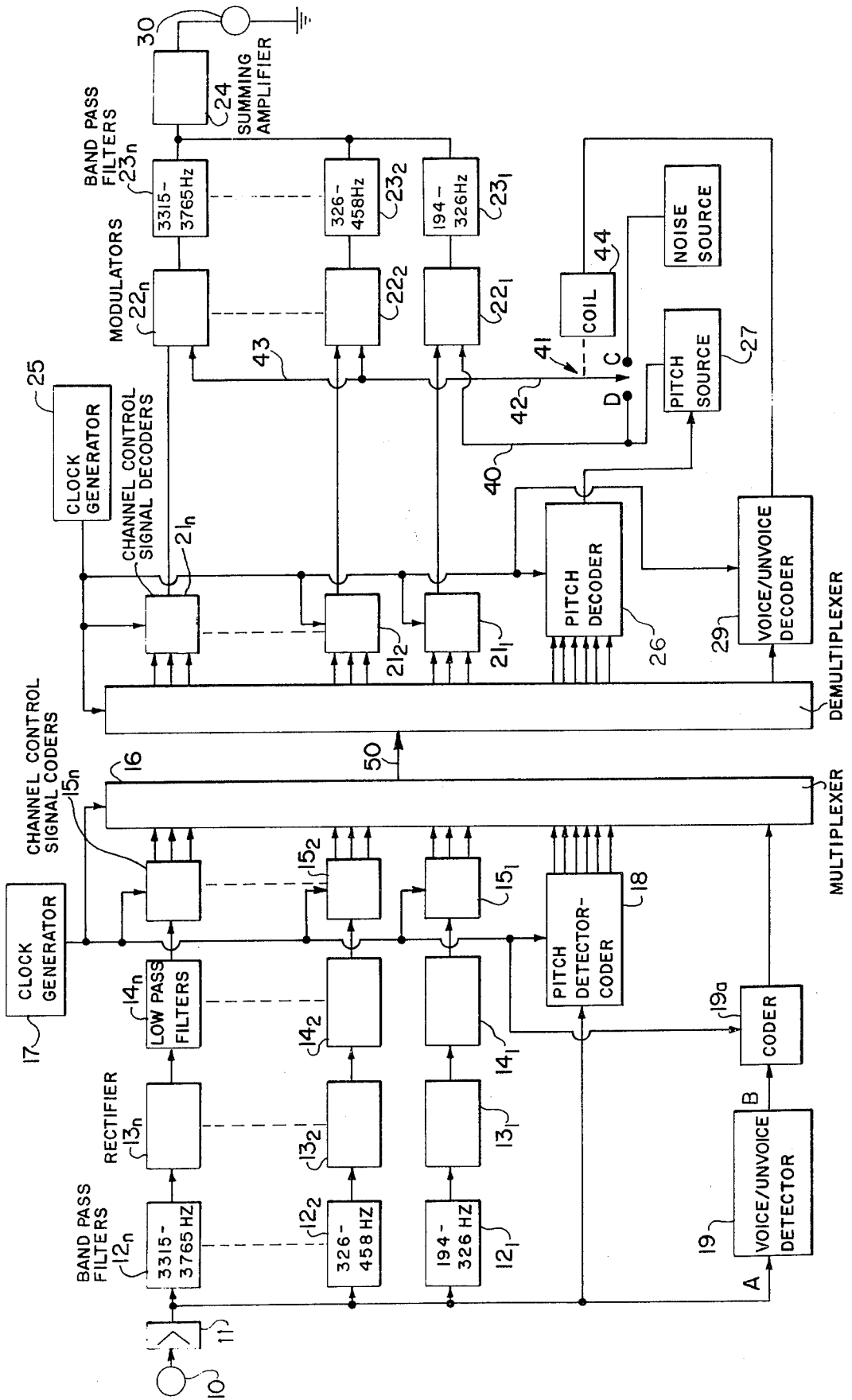
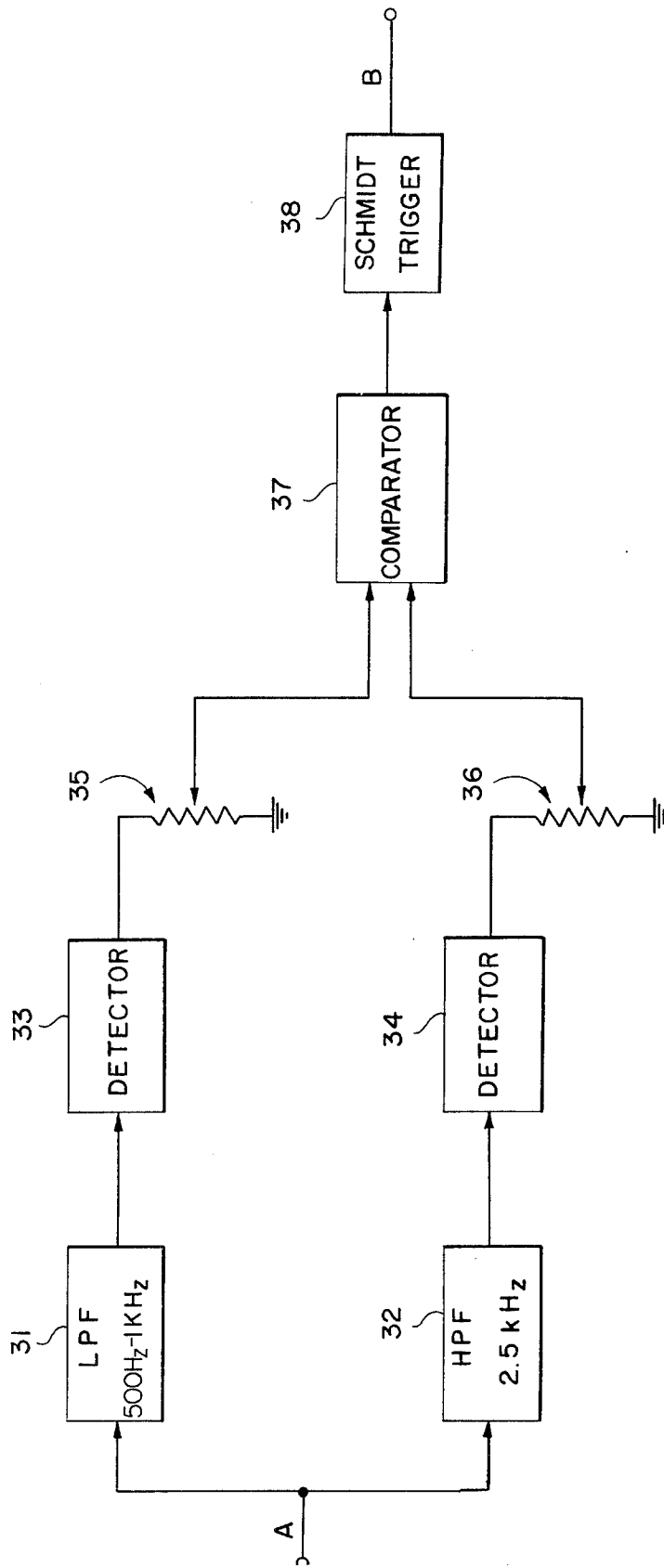


FIG. 2



## APPLICATION OF SIMULTANEOUS VOICE/UNVOICE EXCITATION IN A CHANNEL VOCODER

### BACKGROUND OF THE INVENTION

The present invention relates to improved techniques and apparatus for reproducing speech with vocoder devices and more particularly to new and improved techniques for providing more natural speech sounds through channel vocoders.

Channel vocoders presently known in the art employ circuitry that only allows the production of either voiced or unvoiced conditions in the synthesizer as exemplified by U.S. Pat. Nos. 2,151,091 and 3,622,704. Generally, as shown in the U.S. Pat. No. 3,622,704 binary voice/unvoice (V/UV) decision circuitry in the analyzer will combine the voicing detection and pitch frequency transmission in a single fixed-bit description where the "unvoiced" condition is indicated by all zeros, and the "voiced" condition by the presence of any "ones" in the binary representation of pitch frequency. The result of this type of transmission is to exclude the transmission of simultaneous voiced and unvoiced states since the unvoiced state precludes the transmission of a binary value for pitch.

Because of the above limitations on present channel vocoder systems, certain speech sounds which contain both voiced and unvoiced portions cannot be reproduced with good quality. For example, none of the prior systems allow for the transmission of voiced fricatives such as the /v/ and /z/ sounds. Instead, they must be reproduced as all voiced or all unvoiced, neither of which is correct.

Further, present systems suffer from the inability to provide correct adjustment to the V/UV detector to prevent incorrect decisions and errors with sounds such as unvoiced plosives (/t/ and /k/ bursts) which may be decided as voiced, and with low level voice sounds such as nasal murmurs and voice bars associated with stop consonants. The sensitivity of such decision circuits to different speakers and sounds makes it substantially impossible to provide a compromise adjustment to handle all combinations of speakers and sounds while still generating a natural and faithful reproduction of speech.

While attempts have been made to minimize the effects of accidental voicing, these techniques have had to sacrifice optimum conditions for a variety of other sounds. In addition, such techniques have failed to solve the problem of reproducing the voiced fricatives requiring simultaneous voiced and unvoiced states.

Accordingly, the present invention has been developed to overcome the specific shortcomings of the above known and similar techniques and to provide a technique for improving the performance of channel vocoders.

### SUMMARY OF THE INVENTION

The general purpose of this invention is to provide an improved vocoder, particularly a channel vocoder, that has all the advantages of similar apparatus and none of the disadvantages.

Accordingly, it is an object of the present invention to provide an improved vocoder for reproducing speech sounds more naturally and with higher quality.

Another object of the invention is to provide for the transmission and reproduction of sounds having both

voiced and unvoiced portions, particularly the voiced fricatives.

A further object of the invention is to provide voice/unvoice detection and transmission separate from pitch detection and transmission.

Still another object of the invention is to increase the adjustment capability of the voice/unvoice detection circuitry to render the same less sensitive while improving the naturalness of sounds.

A still further object of the invention is to provide a direct pitch connection for low level voicing with an unvoiced decision.

These and other objects are accomplished by modifying known vocoder systems in a relatively simple and inexpensive manner. According to the present invention, the voice/unvoice (V/UV) detector is separated from the pitch detector in the analyzer to produce a separate bit of binary information indicating the V/UV condition. This separate bit is transmitted along with a separate binary value of pitch frequency from the pitch detector. In the synthesizer of the vocoder the separate bits are received and decoded to produce separate V/UV switch control and pitch source control. In addition, the pitch frequency source is connected to the lowest channel of the vocoder to provide voicing excitation to the lowest channel at all times. The V/UV switch is then constructed to control the switching of the remaining channels from the pitch source during voice detection to a noise source during unvoice detection. In this manner reproduction of sounds containing simultaneous voiced and unvoiced portions can be realized along with a corresponding decrease in the sensitivity of threshold adjustment of the V/UV detection circuitry.

Other objects, advantages and novel features of the invention will become apparent from the following detailed description of the invention when considered with the accompanying drawings wherein:

### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic diagram of a preferred embodiment showing a channel vocoder modified according to the present invention.

FIG. 2 is a schematic diagram of a voice/unvoice detector for use in the embodiment of FIG. 1.

### DESCRIPTION OF PREFERRED EMBODIMENT

Referring now to FIG. 1, a channel vocoder modified according to the present invention is shown having an analyzer portion 1 and a synthesizer portion 2. The analyzer portion 1 is designed to convert a speech sound to a coded signal which can be transmitted over a reduced capacity transmission medium to the synthesizer portion 2 where the coded signal is decoded and reconstructed to produce the original speech as the output. The basic construction of a channel vocoder is well known and may be similar to that shown in U.S. Pat. No. 3,622,704, to which reference is now made.

The analyzer portion 1 contains a transducer 10 for converting speech into an electrical signal. The transducer may be a microphone or any other well-known device. The transducer is connected to an amplifier 11 which in turn feeds the amplified signals to a plurality of bandpass filters 12<sub>1</sub> to 12<sub>n</sub> which separate the signal into *n* separate channels. In the present example *n*=16 and the frequency divisions for each channel in Hz, starting with the lowest channel, is as follows: 194-326,

326-458, 458-592, 592-725, 725-858, 858-992, 992-1142, 1142-1307, 1307-1490, 1490-1705, 1705-1950, 1950-2230, 2230-2550, 2550-2917, 2917-3315, and 3315-3765. As can be seen, the frequency range generally extends from 200-3700 Hz and represents a range permitting high intelligibility and good quality speech reproduction.

The output of each filter is connected by way of full wave rectifiers 13<sub>1</sub> to 13<sub>n</sub>, to low-pass filters 14<sub>1</sub> to 14<sub>n</sub>, respectively. Each output from the low-pass filter represents the time varying average signal amplitude of the particular frequency band and forms a channel control signal. When taken together the control signals represent the envelope of the short-time spectrum of the speech signal.

The channel control signals are coded into pulse code modulation (PCM) signals by signal coders 15<sub>1</sub> to 15<sub>n</sub>, respectively which may be any well-known device for converting an analog signal to a binary representation (A/D converter). The coders may be of any bit capacity such as 3-bit coders as shown in the present example. From the coders the bits are fed to a well known multiplexer circuit 16 prior to being transmitted over a reduced capacity transmission medium. Connected to the signal coders 15 and the multiplexer 16 is a clock generator 17 which acts in a well-known manner to correlate the operation of the devices for synchronizing transmission.

In addition to the above structure, the analyzer of a channel vocoder contains a pitch detector as well as a voice/unvoice detector for detecting the value of pitch and detecting the V/UV state respectively. In prior art systems, as exemplified by U.S. Pat. No. 3,622,704, the pitch detector and V/UV detector were combined to provide all information in one binary coded output. Normally, the combined detector would force the binary output to an all zero representation upon detection of the unvoiced state, and provide a binary output for pitch upon detection of the voiced state. The detection of any "ones" in the binary output (which would necessarily occur during pitch transmission) would therefore represent voicing while the all zero representation, unvoicing. As can be seen, this operation precludes transmission of sounds having both voiced and unvoiced portions since no pitch would be transmitted when the detector indicated the unvoiced state.

Contrary to the above prior art systems, the present invention separates the pitch detector from the V/UV detector to provide separate binary outputs. The pitch detector-coder 18 is connected to the amplifier 11 in parallel with the bandpass filters 12 and provides a 6-bit coded output for pitch which in turn is fed to the multiplexer 16 for transmission. The pitch detector-coder can be constructed as any well-known pitch detector providing an output for binary coding in any conventional manner.

Connected in parallel with the pitch detector-coder is the V/UV detector 19. The detector 19 can be constructed in any manner, such as the structure shown in FIG. 2, to provide an indication of voice or unvoice in the speech sounds. According to FIG. 2, the detector is constructed to provide an input at A to a low-pass filter (LPF) 31 having a frequency setting in the range of 500 Hz-1 KHz, and to a high-pass filter (HPF) 32 having a frequency setting at approximately 2.5 KHz. The presence or absence of voicing is determined by comparing the energy in the frequencies below the setting

on the LPF with the energy in the frequencies above the setting on the HPF. The energy from each is detected by detectors 33 and 34 and fed to comparator 37. The comparator 37 provides a difference output between the two detected values which is fed to a Schmidt trigger 38. When the difference exceeds a threshold level set by the trigger 38 an output pulse will be generated. The presence of the pulse, therefore, indicates voicing, while the absence indicates unvoicing. The output B from the detector is fed to a well-known signal coder 19<sub>n</sub>, which provides a single bit binary representation of the V/UV state, and then to the multiplexer 16 for transmission. The clock generator is connected to control the signal flow from 19<sub>n</sub> as well as from pitch detector-coder 18 in the same well-known manner as required for coders 15.

As an added vernier to the V/UV detector of FIG. 2, potentiometers 35 and 36 can be added to the output of detectors 33 and 34. This allows an additional adjustment to V/UV triggering by controlling the relative signal magnitudes into the comparator.

Turning again to FIG. 1, the coded signals from analyzer 1 are transmitted from multiplexer 16 over a reduced transmission medium 50 to the synthesizer 2. The synthesizer 2 conventionally contains a demultiplexer 20 which passes the coded signals to channel control signal decoders 21 which may be of any well-known construction for converting the 3-bit signal to analog form (D/A converter). From the decoders 21 the analog control signals are fed to modulators 22<sub>1</sub> to 22<sub>n</sub>, and bandpass filters 23<sub>1</sub> to 23<sub>n</sub>, respectively, having the same frequency breakdown as filters 12. The outputs from the filters 23 are then fed through a summing amplifier 24 to an output device, such as a loudspeaker, for transmitting the reconstructed speech.

As was the case in the analyzer portion 1, the synthesizer portion 2 also contains a clock generator 25 synchronized with the generator 17 for controlling the information flow through demultiplexer 20 and decoders 21 in a well-known manner.

In addition to receiving the channel control signals from the analyzer, the vocoders of prior art systems used the combination pitch-V/UV information to connect either a pitch source or a noise source to excite the modulators 22. An all zero binary signal would represent the unvoiced state thereby connecting a noise source to the modulators, while a binary signal containing a one would indicate the voiced state thereby connecting a pitch source to the modulators, controlled in value by the binary signal. In either case, only sounds containing either voiced or unvoiced energy would be transmitted accurately while sounds containing both would be transmitted in error.

According to the present invention, however, a channel pitch decoder 26 is connected to receive the pitch signal from 18 separate from the V/UV signal received at 29 from the coder 19<sub>n</sub>. In addition, a pitch source 27, controlled by the decoded signal from 26, is directly connected to the lowest channel modulator 22 through line 40 to provide pitch excitation to the modulator at all times regardless of the indication from the V/UV detector. The remaining channel modulators are then connected through line 43 to the switch arm 42 of a two position switch. The switch is then operated such that when a voice state is indicated by decoder 29, the pitch source is connected to the remaining channel modulators, and when an unvoice state is indicated, the

noise source is connected to all but the lowest channel modulator. In practice the V/UV decoder could be a Schmidt trigger or flip-flop controlling a magnetic coil 44 which in turn controls the switch arm. In the unvoiced state the coil could be deenergized and the switch arm normally biased to contact pole C of switch 41, while in the voice state, the coil could be energized to move the arm 42 to contact pole D. In this manner the lowest channel modulator would be energized at all times with the current value of pitch regardless of an unvoice decision that would in prior art systems normally connect all channel modulators to the noise source.

As was noted in regard to decoders 21, the pitch decoder 26 and V/UV decoder 29 are controlled by clock generator 25 to provide synchronous operation in a well-known manner.

Turning now to the operation of the vocoder system, the speech is received at 10, amplified, and divided into various frequency bands through filters 12, rectifiers 13, and low pass filters 14 to provide channel control signals for each range of frequency. At the same time, the value of pitch for the speech sound is detected at 18 along with the sound state as voiced or unvoiced at 19. Each of the signals are then coded by their respective coders 15, 18, or 19, and transmitted by multiplexer 16 under the control of clock generator 17. Upon reception of the coded signals, the demultiplexer reconverts the coded signals to corresponding analog signals through respective decoders 21, 26 and 29 under the control of clock generator 25. The decoded channel control signals are then used to excite the channel modulators along with the corresponding excitation from the pitch or noise source as determined by the switch 41. The output from the modulators controls the amplitude signal input to the bandpass filters 23 which are combined at 24 to provide a reconstruction of the original speech through output 30.

In operation, the above described configuration of the channel vocoder according to the present invention provides significant improvements in speech reproduction over vocoders known to the prior art. In low level voicing sounds such as the nasal murmurs phase of /m/ and /n/, voice bars preceeding or following stop consonants /b/, /d/ and /g/, and voiced fricatives such as /t/ as in then, /v/ as in vent, /z/ as in zoo and /z/ as in azure, the voicing detector circuitry of prior art vocoders would not necessarily be sensed as voiced, due in-part to the critical and sensitive adjustment necessary for deciding other sounds in the combined pitch-V/UV detector, and in-part to the low level nature of the sounds themselves. This resulted in objectionable audible effects causing a decrease in speech quality.

In the present invention, by feeding the pitch source to the lowest channel modulator at all times, the low level energy is always transmitted by the pitch detector to the lowest channel resulting in a more natural reproduction of the above sounds. The voiced fricatives, for example, contain all the voicing (periodic pitch) energy in a band substantially corresponding to the band of the lowest channel, and all the fricative (unvoiced) energy in the bands above the one corresponding to the lowest channel. With the connection shown (and the voiced fricatives detected as unvoiced), the pitch source is supplying voiced energy over line 40 to channel modulator 22, at the same time the noise source is supplying the fricative energy to the remaining bands.

All the energy necessary for faithful reproduction of the voiced fricative is therefore being delivered to the modulators. In a like manner, the voicing energy of the low level voice sounds, also substantially contained in the frequencies of the band corresponding to the lowest channel, will be transmitted and reproduced at all times regardless of the V/UV detection.

Even more important than the ability to produce good voiced fricatives is the resultant adjustment capability of the V/UV detector to make the voicing detection less critical and sensitive to different speakers and amplitude levels, while increasing the naturalness of many sounds. By way of example, erroneous voicing on plosive bursts such as /p/, /t/, and /k/, characteristic of most prior art voicing detection circuitry, can be substantially eliminated while still improving the sound. Since the direct connection of the pitch source provides low level voicing energy at all times, it is no longer necessary to set the threshold in the voice detector to pick up the weak sounds (such as voiced fricatives, nasal murmurs, voice bars, etc.), as they will be reproduced anyway by the direct pitch connection. The threshold of the detection circuitry of FIG. 2 can therefore be adjusted to a higher setting resulting in fewer voicing decisions and thus a less critical determination of voicing and unvoicing, particularly in the presence of plosive bursts.

As can be seen from the above description, the present invention provides a channel vocoder capable of providing improved and more natural reproduction of speech sounds with only a simple and inexpensive modification to present vocoders. The modifications reduce the complexity of the circuitry normally necessary to provide good speech reproduction, by providing separate binary representations of pitch and the V/UV state. The separated binary signals are then used to control the value of the pitch source and the connection of the pitch or noise source to the channel modulators. By connecting the pitch source directly to the lowest channel modulator, low level voicing energy can be transmitted at all times while allowing a decrease in the sensitivity of the threshold level of the V/UV detector circuitry.

While the present invention has been described with particular reference to a channel vocoder, it is to be understood that the same principles are applicable to increasing the quality of speech sounds in other types of vocoders, such as formant vocoders. Additionally, while the voice/unvoice detector of the present invention was described as generating a separate bit position in the multiplexer, it is possible to transmit the separate bit without having to add a bit position to the existing vocoder systems. For example, one of the least significant bit positions of one of the channel control signals from coders 15 could be used to handle the binary digit from the coder of 19. While the channel control signal of the chosen coder would then only transmit a signal coded by two bits, the decrease in accuracy might well be offset by less costly modifications required in the vocoder.

Obviously many modifications and variations of the present invention are possible in light of the above teachings. It is therefore to be understood that within the scope of the appended claims the invention may be practiced otherwise than as specifically described.

What is claimed and desired to be secured by Letters Patent of the United States is:

1. A vocoder apparatus for increasing the quality of voice transmission comprising:  
 transducer means for converting a speech sound into an electrical signal;  
 channel means coupled to said transducer means for separating said electrical signal into a plurality of frequency bands each forming a channel control signal;  
 pitch means coupled to said transducer means for detecting a value of pitch from said electrical signal and providing a control signal representative thereof;  
 first detector means coupled to said transducer means for adjustably detecting a speech sound as voiced or unvoiced from said electrical signal and providing a control signal indicative thereof;  
 transmitting means coupled to said channel means, pitch means, and first detector means for transmitting their respective control signals over a reduced capacity transmission medium;  
 a pitch source means coupled to receive said transmitted pitch control signal for providing a corresponding value of pitch output;  
 a noise source providing noise output;  
 a plurality of frequency band signal control means, each having first and second inputs, for providing a plurality of electrical outputs for reconstructing said electrical signal in response to said first and second inputs, said first input of each signal control means being coupled to receive a different one of the transmitted frequency band channel control signals, said second input of the lowest frequency band signal control means being directly coupled at all times to receive a value of pitch from said pitch output;  
 second detector means, responsive to said transmitted first detector means control signal, for coupling said pitch output to each second input of the remaining signal control means only when said speech sound is voiced, and coupling said noise output to each second input of the remaining signal control means only when said speech sound is unvoiced; and  
 means coupled to receive said electrical outputs from said signal control means for reconstructing speech sounds therefrom.

2. The apparatus of claim 1 wherein said first detector means comprises; a low pass filter coupled to receive said electrical signal from said transducer means and provide an output, a high pass filter coupled to receive said electrical signal from said transducer means, and provide an output, first and second detectors coupled to the outputs of said low and high pass filters, respectively, providing detected signals therefrom, comparator means for comparing the detected signals and providing a difference output, and trigger means coupled to said comparator means for providing a voicing code as said first detector means control signal when said difference output exceeds a threshold level and an unvoiced code as said first detector means output when said difference signal is less than said threshold level.

3. The apparatus of claim 2 wherein said first detector means further includes means for individually adjusting the magnitude of the signal outputs from said first and second detectors.

4. The apparatus of claim 3 wherein said channel

means comprises; a plurality of band pass filters coupled to receive said electrical signal and provide individual outputs representative of said frequency bands, each of said filters having sequential pass bands such that the highest frequency of one filter band constitutes the lowest frequency of the next sequential filter band, a rectifier means coupled to rectify each filter output, a low pass filter coupled to each rectifier means for filtering each rectified output, and a coding means coupled to each low pass filter for converting each low pass filtered output to a coded signal forming one of said channel control signals.

5. The apparatus of claim 4 wherein each of said signal control means comprises; a decoding means for decoding a separate one of said transmitted channel control signals to form one of said first inputs, modulator means for combining said one of said first inputs and one of said second inputs to produce a modulation signal, and a band pass filter coupled to filter the modulation signal to form one of said electrical outputs, each of said band pass filters in said signal control means having a band pass corresponding to the band pass of the channel control signal coupled to the first input of its respective modulator means.

6. The apparatus of claim 5 wherein the low pass filter of said first detector means is constructed to have a frequency setting of about 500 Hz-1 kHz and the high pass filter of said first detector means is constructed to have a frequency setting of about 2.5 kHz, and further wherein the band pass filters of said channel means and said signal control means have frequency bands within the range of about 194-3765 Hz.

7. A method of improving speech reproduction with a vocoder comprising:  
 converting a speech sound into a plurality of different frequency bands each represented by a control signal;  
 deriving a signal representing the pitch of said speech sound and a state signal representing said speech sound as voiced or unvoiced;  
 transmitting each of said signals over a reduced capacity transmission medium;  
 generating a value of pitch output in response to said transmitted pitch signal;  
 providing a source of noise output;  
 connecting said pitch output to a single output in response to said state signal indicating voiced speech and connecting said noise source to said single output in response to said state signal indicating unvoiced speech;  
 combining the transmitted control signal of the lowest frequency band with said pitch output to form a channel output representing a part of said speech sound;  
 combining each of the remaining transmitted control signals with said single output to form channel outputs each representing a part of said speech sound; and  
 summing each of said channel outputs to reconstruct said speech signal.

8. The method of claim 7 wherein the step of converting includes the step of forming said plural frequency bands as sequential frequency bands wherein the highest frequency of one band forms the lowest frequency of the next sequential band.

\* \* \* \* \*