

April 17, 1962

M. R. SCHROEDER
BAND COMPRESSION SYSTEM

3,030,450

Filed Nov. 17, 1958

3 Sheets-Sheet 1

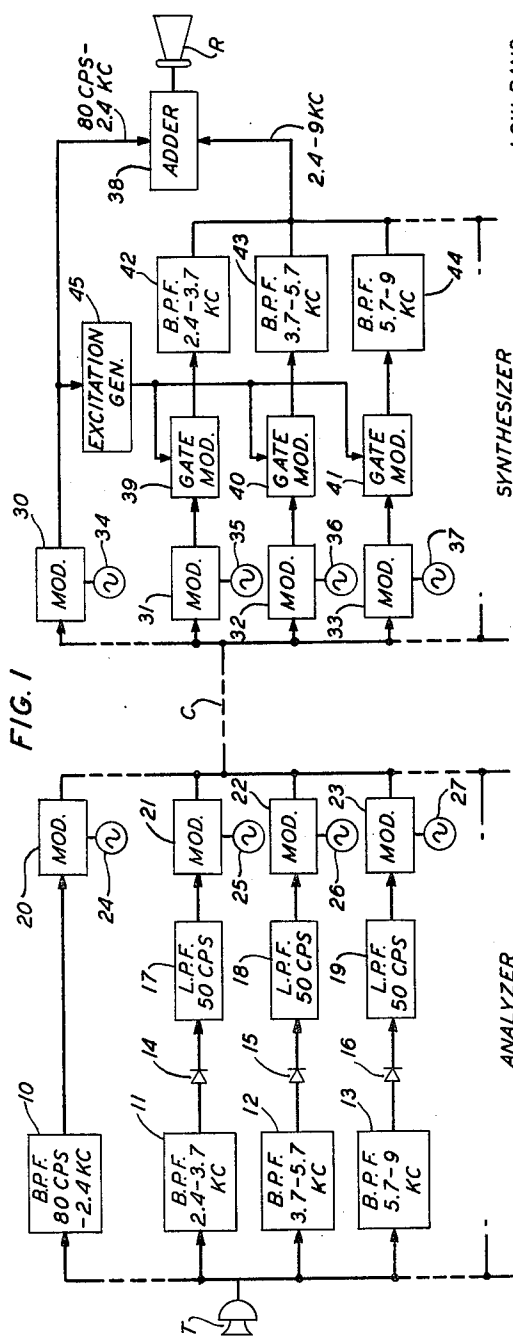


FIG. 1

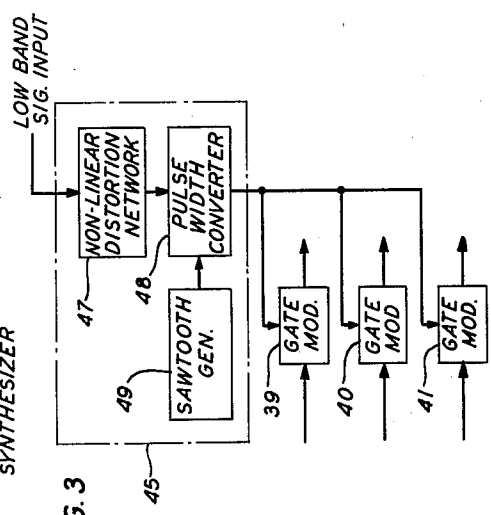


FIG. 3

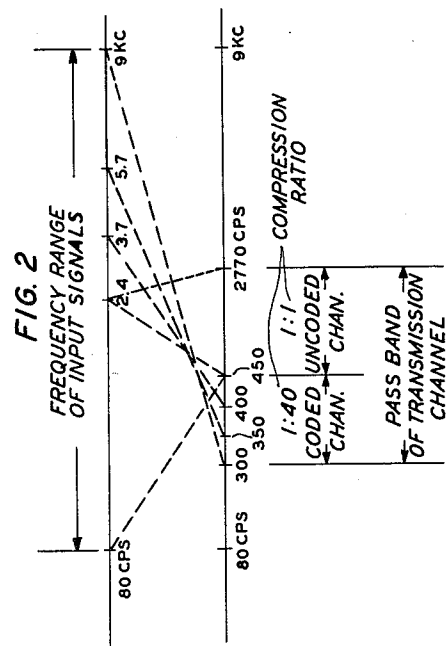


FIG. 2

INVENTOR
M. R. SCHROEDER
BY *A. E. Hirsch*
ATTORNEY

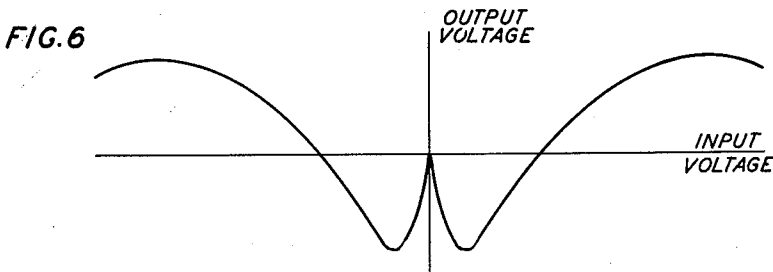
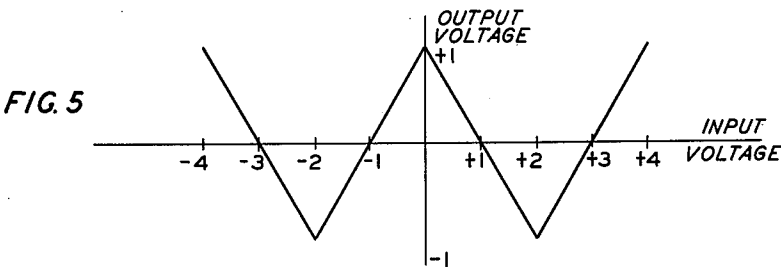
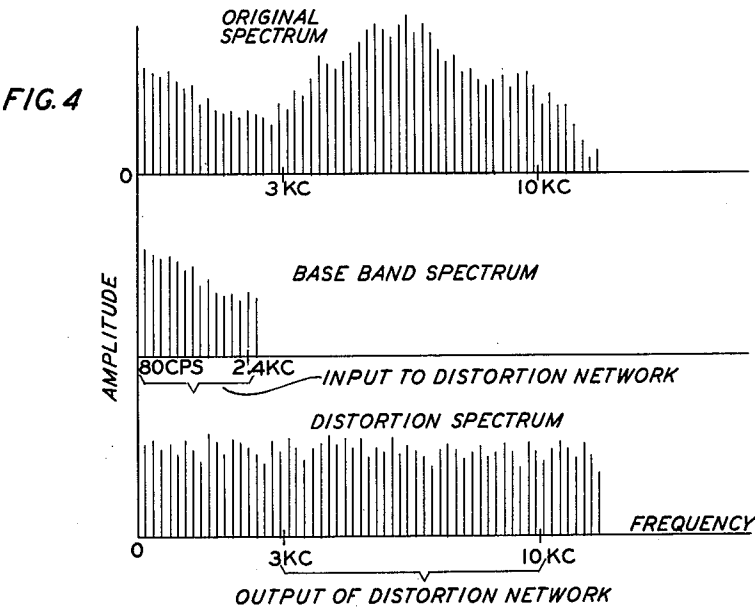
April 17, 1962

M. R. SCHROEDER
BAND COMPRESSION SYSTEM

3,030,450

Filed Nov. 17, 1958

3 Sheets--Sheet 2



INVENTOR
M. R. SCHROEDER
BY *G. E. Hirsch*
ATTORNEY

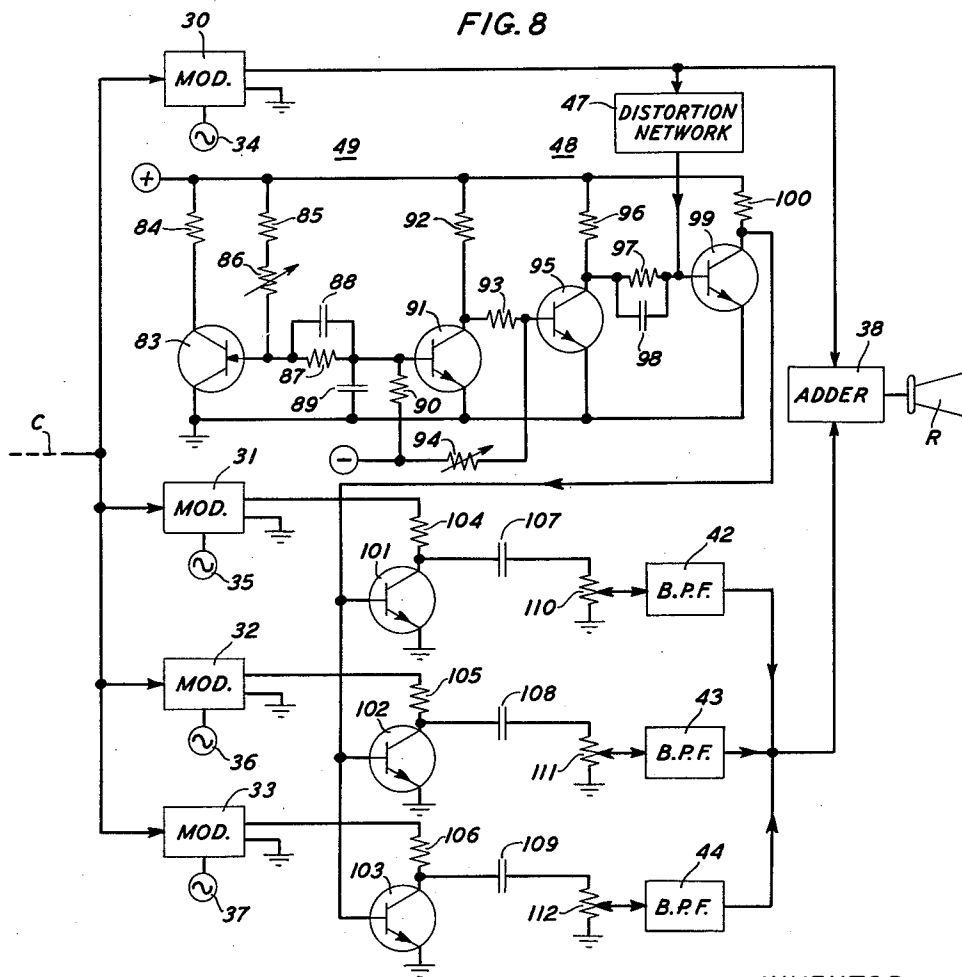
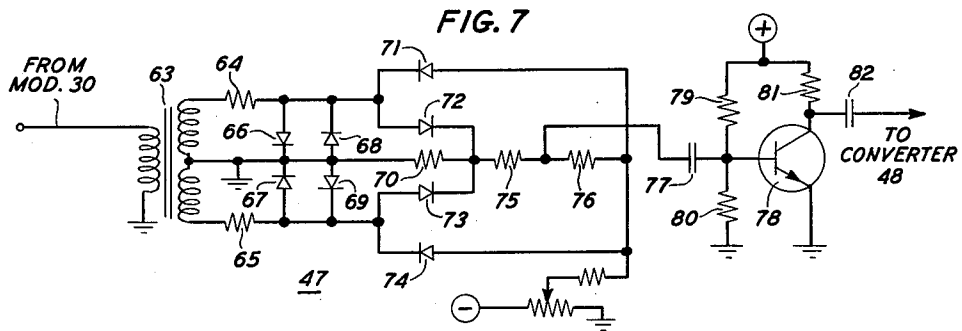
April 17, 1962

M. R. SCHROEDER
BAND COMPRESSION SYSTEM

3,030,450

Filed Nov. 17, 1958

3 Sheets-Sheet 3



INVENTOR
M. R. SCHROEDER
BY *A. E. Hirsch*
ATTORNEY

1

3,030,450

BAND COMPRESSION SYSTEM

Manfred R. Schroeder, Murray Hill, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York

Filed Nov. 17, 1958, Ser. No. 774,173

16 Claims. (Cl. 179—15.55)

This invention relates to the transmission of speech currents over narrow band media by vocoder techniques. One of its principal objects is to reduce the channel bandwidth required for such transmission. Another object is to simplify the analyzer and improve the synthesizer which form parts of vocoder apparatus. Another object is to improve the accuracy and realism with which speech sounds are artificially reconstructed.

Many proposals have been made in the past for reducing the frequency band required for the transmission of speech signals by modifying the voice currents in various ways. Among such proposals a notable one is the channel vocoder of H. W. Dudley Patent 2,151,091 which issued March 21, 1939. In this system an input speech wave is applied to a number of different filters connected in parallel to determine its fundamental frequency or pitch and the distribution of amplitudes among a number of frequency sub-bands into which the speech frequency range is divided. The result of this analysis is translated into a number of control currents each representative of the energy in one sub-band. In particular, one of these control signals represents the fundamental or pitch frequency of the voice. The control currents are transmitted to a receiver station and are there utilized to build up, from local energy sources in a speech synthesizer, an artificial speech wave having the characteristic pitch and amplitude-frequency distribution of the original impressed speech. The synthesizing apparatus at the receiver includes a "buzz" source and a "hiss" source to represent the source of voiced and unvoiced sounds, respectively. The incoming control signals derived at the transmitter station operate to switch the buzz source and the hiss source into action in alternation, as required, and to adjust the frequency of the buzz source, i.e., to tune it. This energy is applied to the synthesizer network which, in turn, is continuously adjusted by the control signals.

In another significant system, known as the resonance vocoder and described in H. W. Dudley Patent 2,243,527 which issued May 27, 1941, as well as elsewhere, a speech wave is divided into a small number, e.g. 3, of comparatively wide bands, each of which embraces a single group of harmonics or formants in which the speech energy tends to be concentrated. The resonance vocoder derives for each such band both a frequency control current and an amplitude control current. At a receiver station, these control currents which occupy much narrower spectrum bands than the voice currents from which they are derived, energize the buzz source or a hiss source in dependence on whether the sound being analyzed is voiced or unvoiced.

In the analyzer portion of both the channel vocoder and the resonance vocoder, an indication both as to the distribution of power among the frequency sub-bands (of either kind) and an indication as to whether a particular sound is voiced or unvoiced is generated for transmission. In the case of the former, the fundamental vocal cord frequency or pitch is also transmitted. In the case of many voice signals, however, particularly when they have been subjected to a certain amount of inevitable frequency distortion due to transmission through band limited apparatus, the fundamental component itself contains very little energy as compared with certain of its harmonics and may be entirely absent. Consequently,

2

the problem of determining the fundamental frequency assumes monumental proportions. The problem of identifying voiced segments of the sound is similarly a complex and trying one.

Thus, two significant problems associated with vocoder transmission remain: the problem associated with the derivation of the fundamental pitch frequency control signal, commonly called the pitch problem, and the problem of relating the character of the hiss and buzz sources utilized at the receiver synthesizer to the signal, which may conveniently be called the naturalness problem.

In a bandwidth compression system described in C. B. H. Feldman Patent 2,817,711, December 24, 1957, the pitch problem is effectively solved by transmitting an un-coded base band of the applied speech signal. The Feldman system employs vocoder techniques at a transmitter to derive narrow band control signals from the high frequency components when present, transmits them to the receiver station, and there employs them to control the high frequency synthesizing circuits. By virtue of the direct and unmodified transmission of the low frequency components of the signal, sufficient information is transmitted to the receiver station to satisfactorily identify the fundamental pitch of the speech signal without a separate pitch control signal. The frequency range of the base band is selected to be suitably broad to insure that sufficient identifiable frequency determining components are transmitted. At the synthesizer, a hiss source is employed to provide the necessary unvoiced sounds. A voiced-unvoiced identifying signal is ordinarily employed selectively to activate the hiss source. If, in the operation of the equipment, the hiss source is connected to the synthesizer at all times, satisfactory synthesis of unvoiced sounds and of many voiced sounds is achieved. However, during the occurrence of some of the more common and important vowel sounds, an audible background hiss is present which not only seriously distorts the reproduced voice signal but also detracts substantially from the naturalness with which the reconstructed speech wave represents the applied voice wave.

Even though a 4000 c.p.s. band can be compressed into something substantially less than 4000 c.p.s. in all of the systems described above, thereby to improve the efficiency of transmission, the improvement is at the expense of signal fidelity. The reconstructed speech is, at its best, no better than ordinary telephone quality speech. Accordingly, it is desirable to improve the intelligibility, naturalness, and reproduction quality of the applied speech signals in a vocoder-like system using ordinary telephone transmission channels.

According to the present invention, both of the aforementioned vocoder problems are resolved so that high fidelity speech transmission over nominal telephone bandwidth channels becomes a reality. Specifically, the present invention permits a wide band speech signal of nominally 10,000 cycles per second to be compressed for transmission over an ordinary telephone channel, e.g., 4000 cycle channel or less, with a maximum preservation of fidelity and with a minimum of apparatus complexity. It achieves its object and attains this result by a direct and unmodified transmission of the low frequency components of voice message and by the transmission of the high frequency components by vocoder techniques as in the Feldman system. Unlike the Feldman system, however, excitation of the proper sort is continuously employed at the synthesizer to afford a faithful reconstruction of the input speech signals.

In essence, the aforementioned vocoder problems are resolved by generating a single excitation signal at the synthesizer which is highly correlated with the input voice signal itself. The single excitation signal assumes the role both of the hiss source and the buzz source of the

conventional vocoder synthesizer. Correlation is preserved by deriving the excitation signal from a base band of uncoded voice signals by means of a nonlinear distortion network which effectively spreads the spectrum of the base band to embrace the frequency range of the high band signals. Periodic wave portions of the base band signal applied to the input of the excitation generator, representative of voiced sounds, produce at the output of the generator a wide band of periodic signals of the same period. Similarly, aperiodic wave portions of the base band, representative of unvoiced sounds, produce at the generator output, a wide band of aperiodic signals. Periodicity of the excitation signal is thus automatically preserved together with irregular fluctuations of consecutive voice periods which contribute individuality to a voice whereas this periodicity and these characteristics are lost in the usual parametric representation. Irregular fluctuations at the onset and decay of voiced speech portions, particularly important insofar as voice individuality is concerned, are also preserved.

Thus, both hiss and buzz type sounds are synthesized by the proper type of excitation, i.e., hiss by aperiodic signals and buzz by correlated periodic signals. Accordingly, synthesis may be extended to relatively low frequencies and the required base band may be substantially reduced in width. Moreover, the voice excited hiss and buzz signals are produced without resort to measuring and decision elements so that the primary cause of synthesis failure is avoided and accuracy and reliability of operation are assured. Apparatus incorporating the features of the invention is thus eminently suitable for use in equipment provided for subscriber use as well as in that intended for laboratory work.

In conventional vocoder fashion, the energy of the excitation signal is continuously adjusted by the high frequency control signals to produce a replica of the high band voice signals. The reconstructed signals are combined with the unmodified base band signals to produce a composite signal which may be delivered to a reproducer.

The invention will be fully apprehended in the following detailed description of a preferred embodiment thereof taken in connection with the appended drawings in which:

FIG. 1 is a block schematic diagram showing speech transmission apparatus illustrating the invention;

FIG. 2 is a diagram illustrating the allocation on the frequency scale of sub-bands carrying coded and uncoded signals in accordance with the invention;

FIG. 3 is a block schematic diagram showing in more detail an excitation generator which may be used in the practice of the invention;

FIG. 4 is a diagram illustrating signal spectrograms helpful in explaining the invention;

FIG. 5 illustrates the transmission characteristic of one form of nonlinear distortion network which forms a part of the excitation generator shown in FIG. 3;

FIG. 6 illustrates a preferred distortion network characteristic;

FIG. 7 is a schematic diagram of a distortion network possessing the characteristic shown in FIG. 6; and

FIG. 8 is a schematic diagram, partially in block form, illustrating speech synthesizer apparatus in accordance with the invention.

Referring now to the drawings, FIG. 1 shows a speech analyzer at a transmitter station and a speech synthesizer at a receiver station interconnected by a transmission channel. At the analyzer, a voice current originating, for example, at a transmitter T is delivered in parallel to a number of band pass filters 10, 11, 12 and 13. In accordance with the invention, each of the several filters is constructed to pass contiguous portions of a band of frequencies embracing a speech band from approximately 80 cycles per second to 9000 cycles per second. Band pass filter 10, connected in the first of the parallel paths and proportioned to pass a base band of frequencies extending from approximately 80 c.p.s. to 2400 c.p.s., trans-

mits the lower portion of the voice current frequency range directly and without further modification to a modulator 20. Band pass filters 11, 12 and 13, connected in the other ones of the parallel paths and proportioned, for example, to pass energy in the sub-bands extending from 2400 to 3700, 3700 to 5700, and 5700 to 9000 c.p.s. respectively, transmit energy in these higher bands to signal processing apparatus associated with the respective channels. Thus, the output terminal of each of the band pass filters 11, 12 and 13 is connected to a full wave rectifier (14, 15 and 16) followed by a low pass filter (17, 18 and 19) whose output comprises a slowly varying control signal whose instantaneous magnitude represents the instantaneous magnitude of the energy in the frequency band with which it is associated. Additional sub-bands (not shown) may be employed, if desired, by providing additional parallel paths identical to those shown and proportioned to pass the desired frequency bands. In general, a geometrical subdivision is preferred in which the ratio of bandwidth to center frequency is maintained constant.

Compression of the band between approximately 3000 and 10,000 c.p.s. into a few hundred c.p.s. using vocoder techniques is possible because the speech-hearing link has a relatively low information rate above approximately 3000 c.p.s. Accordingly, 15 to 25 c.p.s. control signals are conventionally employed to specify the energy distribution of the higher frequency sub-bands. However, it is in accordance with the present invention to increase the control signal bands to embrace bandwidths of from 30 to 50 c.p.s. This increase affords an increase in fidelity that far outweighs the price paid in bandwidth since the increase greatly improves the reproduction of fast attacks such as are found in the plosives "t" and "p," and in the affricate "ch." Also, the wider band holds the delay in the vocoder to less than 10 milliseconds. A differential delay distortion of this magnitude between the coded and uncoded bands is inaudible and thus there is no need for delay equalization.

The three (or more) control signal derived in this fashion, together with the low band signals transmitted by filter 10, are systematically arranged adjacent to each other on the frequency scale by conventional heterodyning techniques, the apparatus for which comprises modulators 20, 21, 22 and 23 and associated oscillators 24, 25, 26 and 27. The modulators may be alike, and indeed, all of the elements in the analyzer may be of well-known construction. The oscillators are adjusted to deliver oscillation frequencies which are suitably separated in the frequency scale by frequency differentials in order to place the individual signal components at any desired point within this frequency spectrum.

Preferably, the low band uncoded speech signal and the three narrow band control signals are contiguously arranged on the frequency side in the manner shown in FIG. 2. The speech band from 80 to 2400 c.p.s. is shifted upward by 370 c.p.s. to occupy the portion of the frequency spectrum between 450 and 2770 c.p.s. In a similar fashion, the control signals representative of the speech signal components occupying the 2.4-3.7, 3.7-5.7 and 5.7-9 kilocycles per second bands are shifted downward to occupy that portion of the spectrum between 300 and 450 c.p.s. By transposing the signals in this fashion, signals at both the low and high end of the input signal range are centered within the transmitted range so that noise disturbances and the like imparted to the signal during transmission will not seriously impair or distort these signals. A loss at the lower edge of the transmission band thus results only in a slight narrowing of the frequency range at the high end of the band.

While frequency reallocation of the sort illustrated in FIG. 2 affords substantial transmission advantages, it is, of course, possible to employ any other form of signal transmission according to methods well known in the communications art. For example, low index single-

sideband FM, single-sideband AM, or double-sideband AM carriers in quadrature may be employed for multiplexing the individual signal components for transmission. Separate channels may, of course, be provided if desired. By any of these techniques an over-all pass band of approximately 2500 c.p.s. is sufficient for transmitting the entire frequency range of the input signal. The actual compression for high frequencies in the illustrative example of FIG. 2 is approximately 6600 c.p.s. (9000 c.p.s. minus 2400 c.p.s.) to 150 c.p.s., or 44 to one. The effective compression for the entire transmitted band is 8920 c.p.s. to 2470 c.p.s. or 3.6 to one.

Returning now to FIG. 1, the low band speech signal and the control signals derived in the fashion described above may now be transmitted by way of channel C to a receiver station where they may be used to control artificial voice synthesizing apparatus.

At the synthesizer, shown in FIG. 1, the energy arriving by way of channel C is first separated into the corresponding components developed at the analyzer and these components are restored on the frequency scale to their original frequency allocations. Thus, the uncoded low band signal is shifted downward on the frequency scale to its original range of 80 c.p.s. to 2400 c.p.s. by modulator 30 and oscillator 34. The output of modulator 30 is delivered immediately and without further modification to an adder 38. It is also supplied as an input signal to an excitation generator 45 wherein a signal is generated which corresponds operationally to both the hiss and buzz signals used in conventional vocoder circuits.

At the synthesizer, the several control signals received from the analyzer are shifted upward on the frequency scale to their original allocations by means of modulators 31, 32 and 33 and oscillators 35, 36 and 37. The outputs of these modulators are applied, together with the excitation signal supplied by generator 45, to gate modulators 39, 40 and 41. In the modulators, the control signals serve to adjust the amplitude of the applied excitation signals. The outputs of the gates are passed through band pass filters 42, 43 and 44, which are identical to those at the analyzer, and are then combined additively to form a combined output signal occupying the band of frequencies from 2400 to 9000 c.p.s. This signal resembles very closely the high frequency signal applied to band pass filters 11, 12 and 13 at the transmitter station. It is then added to the low band signal occupying the frequency range from 80 to 2400 c.p.s. in adding circuit 38 to produce a replica of the voice current originating in transmitter T. This signal is applied to a suitable reproducer R.

The degree of naturalness and realism attained in the construction of artificial speech depends in large measure on the nature of the excitation signal employed in the synthesis process. Ideally, speech synthesis apparatus requires an excitation signal characterized by a flat spectrum of constant power density and of the proper type, i.e., discrete or continuous. In conventional vocoder apparatus of the sort referred to above, two separate excitation signals are employed; one, the buzz signal, is periodic in nature and the other, the hiss signal, is aperiodic in form. An auxiliary control signal operates to select as between the buzz energy and the hiss energy and a pitch signal adjusts the frequency of oscillation of the buzz signal. A decision element, with its inherent susceptibility to error, is required to make the selection. The need for these signals is avoided in the present invention by employing a single excitation signal that corresponds more generally to the ideal form of excitation signal than does a completely random source of noise, either hiss or buzz. Such an excitation signal, closely correlated with the speech currents delivered from transmitter T, imparts to the reconstructed speech signals much of the sound color of the speaker so that the artificial speech is highly intelligible. A considerable portion of the unnaturalness

of vocoder speech is thus eliminated. Since no decision of any sort need be made, the chance of error is eliminated and the synthesis apparatus is both improved in accuracy and simplified in structure.

The voice excitation signal used in the reconstruction of speech according to the present invention is derived directly from the portion of the original speech signal contained in the uncoded band, i.e., in the illustrative example described above, the voice excitation signal is derived from the uncoded band in the range of approximately 80 c.p.s. to 2400 c.p.s.. This uncoded band is available at the output terminal of modulator 30.

Before considering in detail the instrumentation of a suitable excitation generator, it is helpful to review briefly the nature of the spectra of the various signals available at the receiver station. The uncoded speech embraced in the frequency band from 80 to 2400 c.p.s. has a short time-power spectrum either continuous (unvoiced) or of the quasi-discrete type (voiced), shaped by the talker to impart his intended sound color or phonetic value. On the other hand, the spectra of speech sounds above 3000 c.p.s., especially those of fricatives, affricates, and stops, which are the predominant variety of sounds above 3000 c.p.s., are rather broad. The ear is not very sensitive to spectral modifications of these sounds. For example, it is known that a broad resonance around 3000 c.p.s. excited by noise produces an acceptable "sh" sound in spite of the fact that this is a rather drastic simplification of the spectrum of a human-made "sh." Furthermore, the subjective "sh" percept does not depend on the width of this resonance within wide limits. Hence, the "sh" can be adequately specified by a single parameter. The same is true for the "s" which can be synthesized from a band of noise around 7000 c.p.s. Other fricatives require somewhat more complex synthesis but are equally susceptible to considerable alteration to their spectra.

In the present invention, an appropriate combination of nonlinear distortion and fast automatic gain control is employed effectively to spread or transform the spectrum of the shaped and band limited spectrum of the uncoded speech band into a substantially flat wide-band spectrum of constant power density suitable for the synthesis of speech in the frequency range of approximately 3000 to 10,000 c.p.s. Generator 45 thus transforms the uncoded base band into a wave which has at all times the proper "fine" structure, i.e., it is continuous for aperiodic input signals and is quasi-discrete for periodic input signals, and has a sufficiently invariant envelope over the high band frequency range so that satisfactory synthesis can be carried out. Since the signal is automatically of the proper form at all times (by virtue of its correlation with the original speech signal) it may continuously be applied to the synthesizer modulator. No decision elements of any sort are needed.

FIG. 3 illustrates in block schematic form an arrangement of components suitable for generating the required excitation signal. In the figure, excitation generator 45 comprises a nonlinear distortion network 47 supplied with the uncoded base band speech signal derived from modulator 30. Since gate modulators are used in the synthesizer, binary amplitude pulses of varying widths are required to energize them. Accordingly, the excitation function emerging from the nonlinear distortion network is transformed by means of a converter 48 into a pulse-width modulated signal. Operation of the pulse-width converter requires the application of periodic sawtooth pulses which may be derived, for example, in generator 49. A sawtooth frequency of 30 kc.p.s. is suitable. The transformed distortion signal is supplied continuously to each of the gate modulators 39, 40 and 41, which correspond to those illustrated in the synthesizer of FIG. 1. Thus, the gates are opened in accordance with the duration of the applied excitation signal pulses. The control signals derived from modulators 31, 32 and 33 are applied to the control terminals of the gates 39, 40 and

41 to adjust the amplitude of the excitation signal passed by the gates.

Before turning to a detailed description of the elements of the excitation generator 45, it is helpful to consider the spectrograms illustrated in FIG. 4. In line A of FIG. 4, the spectrogram of a typical speech sound is shown in the range extending beyond 10 kc.p.s. It is seen to be rather broad. The base-band of the spectrum extending from approximately 80 c.p.s. to 2.4 kc.p.s., illustrated in line B, comprises the input signal of the excitation generator 45. The output of the excitation generator is illustrated in line C. It is a substantially flat spectrum of continuous power density and is satisfactory for synthesizing a wide variety of speech sounds. The portion between 3-10 kc.p.s. is preferably extracted for use in speech synthesis.

The distortion network used to produce a flat spectrum of the sort illustrated in FIG. 4C may be realized in a number of ways. For example, the uncoded speech signal may be rectified or clipped to produce a broader power spectrum. However, the most powerful distortion device in this context increases the number of zero crossings per second of the signal applied to the input terminal. Accordingly, generation of the excitation signal need only comprise suitable means for increasing the mean rate of sign change thereby effectively to spread the spectrum to wide limits. This increase may be accomplished conveniently by clipping different versions of the speech wave, i.e., after integrating, differentiating the speech wave or the like, and multiplying together four or more square waves obtained in this manner. In principle, this multiplication can be done by triggering a flip-flop circuit with sign changes from all the square waves. However, because of recovery time in the flip-flop, sign changes may occasionally be missed. Clearly, the multiplication must be done in an error-free, time-independent manner.

One form of sign change multiplier comprises a piecewise network with an input-output characteristic of straight line segments. An illustration of such a characteristic is shown in FIG. 5. For this characteristic, the output is +1 for inputs zero and ± 4 and -1 for inputs ± 2 . Thus, if 4 square waves of amplitude ± 1 are added and applied to the input of a network possessing this characteristic, the output is the product of these square waves. If a sine wave of amplitude greater than three is applied to the network, four zeros at its output are obtained for every zero at the input. This form of sign multiplier may be employed to generate a great many axis crossings so that distortion components up to 10,000 c.p.s. and beyond may be produced. This increase is achieved without resort to infinite clipping or the like. Moreover, the multiplier is quite independent of the specific form of its input. In general, the network characteristic need not be linear. Satisfactory multiplication is produced so long as it contains a sufficient number of slope inversions.

A network having a characteristic of the form shown in FIG. 5 has exactly the desired properties: it has a continuous input-output characteristic and allows for any desired spectral spreading by using a sufficient number of slope inversions. In addition, this form of nonlinear network exhibits a certain threshold above which its peak-to-peak output voltage is constant. However, while the peak-to-peak output voltage of the network is constant, and the same is approximately true for the R.M.S. output voltage, the spectral power density is somewhat dependent on the input amplitude. It may, in fact, decrease with increased input amplitude. For use in speech synthesis, however, the power density must be maintained constant. For this purpose, logarithmic compression is employed in addition to the zig-zag network. The characteristic of a suitable composite distortion network is shown in FIG. 6. This is, of course, only one example since the network may have many more slope inversions than shown.

FIG. 7 shows a distortion network which exhibits the characteristic illustrated in FIG. 6 for a finite range of input voltages. It comprises a diode network supplied by means of transformer 63 with the low band speech signal derived from modulator 30. The diodes 71, 72, 73 and 74 together produce the W-shaped response, or approximately two complete oscillations of a sinusoid. Diodes 66, 67, 68 and 69 implement the instantaneous logarithmic compression function. The two distortion signals produced by these diode arrays are developed across resistors 75 and 76, respectively, and combined to produce a composite distortion signal which may be used as the synthesizer excitation signal. It is coupled by way of capacitor 77 to the base of transistor 78 for amplification to a usable value. The amplified distortion signal is transferred by capacitor 82 to the pulse-width converter 48.

FIG. 8 illustrates schematically the synthesizer employed at the receiver station of a system outlined, for example, in FIG. 1. Block diagram portions of the synthesizer of FIG. 6, which correspond to those shown in FIG. 1, are identified by like numerals. Sawtooth oscillator 49 employs a uni-junction diode 83 which operates much in the fashion of a gas tube, and a triode transistor 91 connected in a conventional circuit. The output of the oscillator, suitably adjusted in frequency by potentiometer 86 and in amplitude by adjustment of the bias potentiometer 94, is applied to the base of transistor 95 which forms a part of the pulse-width converter 48. This converter comprises transistor 95 and transistor 99 coupled by a parallel RC network 97-98. The time-constant of the circuit is adjusted to produce a binary amplitude output signal.

The distortion signals derived from distortion network 47, which may take the form of the network shown in FIG. 7, are applied to the base of transistor 99. Transistor 99 is driven between cut-off and saturation so that the wave appearing at its emitter is a 30 kc.p.s. square wave with the position of one of its edges modulated by the output of the distortion network 47. This wave is applied to the bases of the transistors 101, 102 and 103 which comprise the gating modulators 39, 40 and 41 of FIG. 1.

The respective gates are opened for intervals corresponding to the variable width square waves thereby to pass during those intervals the applied distortion signal. The several high band control signals derived from modulators 31, 32 and 33 and representative of the energy content of the respective sub-bands are applied to the collectors of the respective gates and control or modulate the amplitude of the distortion signal applied to the bases of these transistors. Modulated output signals are developed across potentiometers 110, 111 and 112 and are adjusted to provide suitable interchannel equalization. The equalized signals are passed through band-pass filters 42, 43 and 44 and combined additively to produce a combined output signal which resembles very closely the high frequency signal applied to the control signal channels at the analyzer. It is thus representative of the speech band from approximately 2400 to 9000 c.p.s. It is combined in adder 38 with the low band signal supplied from modulator 30. The resultant output signal constitutes a replica of the voice signal generated in transmitter T and is supplied to reproducer R.

Although the invention has been described as relating to specific embodiments, the invention should not be deemed limited to the embodiments illustrated, since various modifications and other embodiments will readily occur to one skilled in the art.

What is claimed is:

1. In a speech producing system, a source of voice waves representative of a first selected portion of a speech signal, a source of narrow band control waves representative of predominant speech energy in a second selected portion of said speech signal, means supplied with said voice waves for generating a relatively wide band of

waves, and means under the influence of said control waves for converting said wide band of waves into an artificial second portion of said speech signal.

2. In a speech producing system, a source of voice waves representative of low frequency components of a speech signal, a source of narrow band control waves representative of predominant speech energy in different parts of the high frequency portion of said speech signal, means for generating from said voice waves a band of waves from which high frequency speech waves can be synthesized, and means under the influence of said control waves for converting said band of waves into high frequency artificial speech waves.

3. Apparatus for synthesizing a voice wave from a base band signal representative of low frequency components of said voice wave and compressed high frequency components of said voice wave which comprises, means for effectively spreading the spectrum of said base band signal to produce an auxiliary signal whose spectrum encompasses the original spectrum of said high frequency signals, adding means, means responsive to said compressed signal components for selectively connecting said auxiliary signal to said adding means, means for connecting said base band signal to said adding means, and speech reproducing means supplied with the composite signal derived from said adding means.

4. Apparatus for synthesizing a voice wave as defined in claim 3 wherein said means for spreading the spectrum of said base band signal comprises an axis crossing multiplier supplied with said base band signals.

5. Apparatus for synthesizing a voice wave as defined in claim 3 wherein said means for spreading the spectrum of said base band signal comprises a network with an input-output characteristic having a plurality of slope inversions.

6. Apparatus for synthesizing a voice wave as defined in claim 3 wherein said means for spreading the spectrum of said base band signal comprises a first diode network supplied with said base band signal, said first diode network exhibiting a W shaped response approximating two complete oscillations of a sinusoid, a second diode network supplied with said base band signal, said second diode network exhibiting a logarithmic compression response, and means for combining the signals produced at the outputs of said first and said second networks to produce a composite wide-band distortion signal.

7. In a speech producing system, a source of voice waves representative of low frequency components of a speech signal, a source of narrow band control waves representative of predominant speech energy in different parts of the high frequency portion of said speech signal, means for generating from said voice waves a band of waves from which high frequency speech waves can be synthesized, said band of waves having a substantially invariant envelope and a fine structure correlated with said low frequency signals, and means under the influence of said control waves for converting said band of waves into artificial speech waves.

8. Apparatus for transmitting a voice signal from point-to-point which comprises means for transmitting low frequency components of said signal directly to a receiver station, means for deriving control signals representative of voice component currents from high frequency components of said signal, means for transmitting said control signals to said receiver station and, at said receiver station, means for deriving from said directly transmitted low frequency components a speech excitation signal, means controlled by said transmitted control signals for selectively employing said excitation signal to generate artificial speech currents representative of said high frequency components, and means for combining said artificial speech currents with said directly transmitted lower frequency components to produce a reconstruction of said voice signal.

9. Apparatus for transmitting a voice signal from point-to-point which comprises means for transmitting low frequency components of said signal directly to a receiver station, means for deriving narrow band control signals representative of the energy distribution of high frequency components of said signal, means for transmitting said control signals to said receiver station, and at said receiver station, means for utilizing said low frequency components of said signal to generate a speech excitation signal, means for utilizing said control signals and said speech excitation signal to synthesize speech currents representative of said high frequency components, and means for combining said synthesized high frequency components with said directly transmitted low frequency components to produce a reconstruction of said voice signal.

10. Apparatus for transmitting a voice signal from a transmitter station to a receiver station which comprises at said transmitter station means for deriving from said voice signal a base band of frequencies representative of the low frequency components of said signal, means for deriving narrow band control signals representative respectively of the energy distribution of individual high frequency components of said signal, means for transmitting said base band signal and said control signals to said receiver station, and which comprises at said receiver station, means for deriving from said base band signal a broad band excitation signal whose periodicity from instant-to-instant is closely related to that of said voice signal, a plurality of gating means each having an input terminal and an output terminal for selectively passing signals applied to said input terminals to said corresponding output terminals, means for supplying said excitation signal to all of the input terminals of said gating means, means responsive to individual ones of said control signals for altering the magnitude of signals passed by individual ones of said gating means, and means for combining the signals passed by said gating means with said base band signal to produce a reconstruction of said voice signal.

11. Apparatus as defined in claim 10 wherein said means for deriving from said base band signal a broad band excitation signal comprises a nonlinear distortion network supplied with said base band signal for producing a relatively wide distortion signal, a source of periodic high frequency waves, and means supplied with said distortion signal and with said high frequency waves for producing a sequence of square waves, the positions of whose edges are indicative of the instantaneous amplitude of said distortion signal.

12. Vocoder apparatus for transmitting an information signal which occupies a relatively wide frequency range to a receiver station over a transmission channel which has a relatively narrow frequency range which comprises means for dividing the frequency range of said information signal into a lower range and an upper range signal, means for transmitting component currents in said lower range directly and continuously to said receiver station, means for deriving control signals representative of signal component currents in said upper range, means for transmitting said control signals to said receiver station, and at said receiver station, means for transforming the component currents to said transmitted lower frequency range into a signal of substantially constant power density whose frequency range extends substantially over said upper range and whose periodicity corresponds at every instant substantially to the periodicity of said information signal, means for utilizing said control signals to control the synthesis of artificial upper range component currents from said transformed component currents, means for combining said artificially synthesized upper range component currents with said directly transmitted lower range component currents, and means for reproducing said combined currents as a facsimile of said information signal.

11

13. In combination with apparatus as defined in claim 12, means at said transmitter station for shifting the component currents in said lower range upward on the frequency scale by a pre-established frequency band, means for shifting each of said control signals downward on the frequency scale to occupy frequency ranges adjacent to one another and contiguous to the lower extreme of the shifted range of said lower range signal, and means at said receiver station for restoring said component currents of said low range and of said control signals to their original ranges on the frequency scale.

14. Apparatus as defined in claim 13 wherein each shifting means comprises a modulator and associated oscillator.

15. In a speech producing system, a source of voice waves representative of selected components of a speech signal, a source of narrow band control waves representative of predominant speech energy in different selected components of said speech signal, means supplied with said voice waves for generating a relatively wide band of waves, wherein said generating means comprises means for infinitely clipping the received base band signal, means for differentiating said clipped signal, and means for rectifying said differentiated signal to retain positive peaks only, and means under the influence of said control

12

waves for converting said wide band of waves into artificial speech waves.

16. Apparatus for transmitting a voice signal from point-to-point which comprises means for transmitting low frequency components of said signal directly to a receiver station, means for deriving control signals representative of frequency regions of principal resonance of individual sounds of said voice signal, means for transmitting said control signals to said receiver station and, at said receiver station, means for deriving from said directly transmitted low frequency components a speech excitation signal, means controlled by said transmitted control signals for selectively employing said excitation signal to generate artificial speech signals representative of said frequency regions of principal resonance, and means for combining said artificial speech signals with said directly transmitted low frequency components to produce a reconstruction of said voice signal.

References Cited in the file of this patent

UNITED STATES PATENTS

2,151,091	Dudley	Mar. 21, 1939
2,640,880	Aigrain et al.	June 2, 1953
2,817,711	Feldman	Dec. 24, 1957