Oct. 10, 1967 G. FANT 3,346,695

VOCODER SYSTEM

Filed April 27, 1964 3 Sheets-Sheet 1

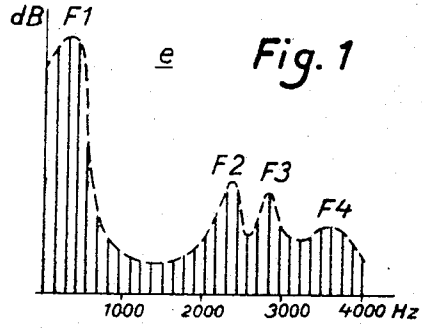


Fig. 1

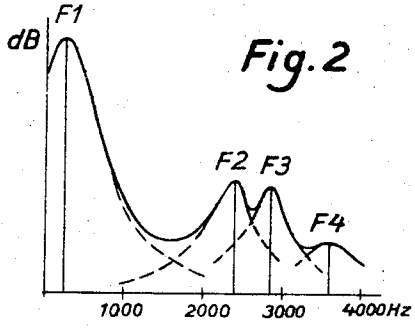


Fig. 2

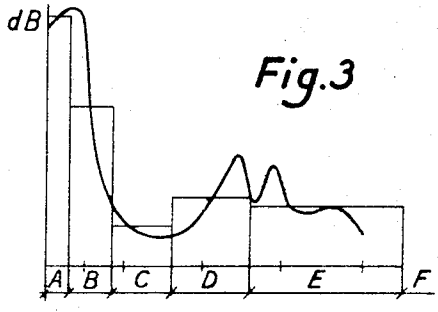


Fig. 3

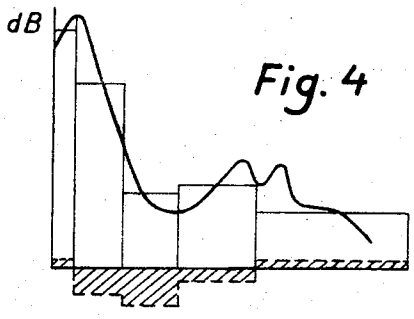


Fig. 4

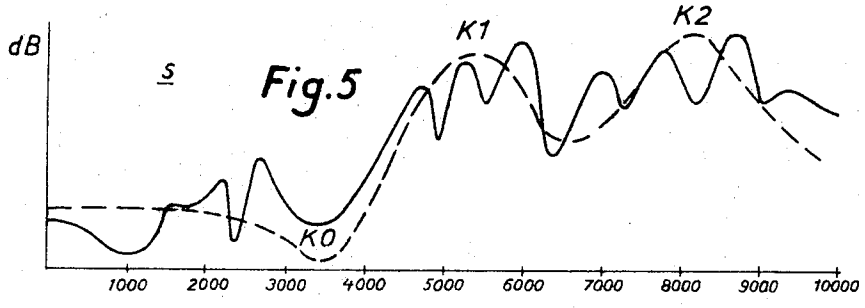


Fig. 5

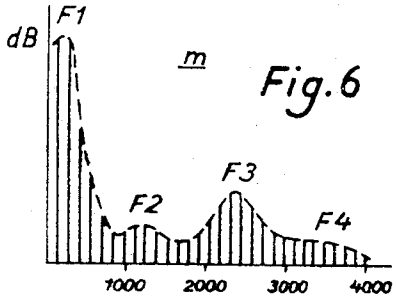


Fig. 6

INVENTOR.
GUNNAR FANT
BY Hane and Nyquist
ATTORNEYS

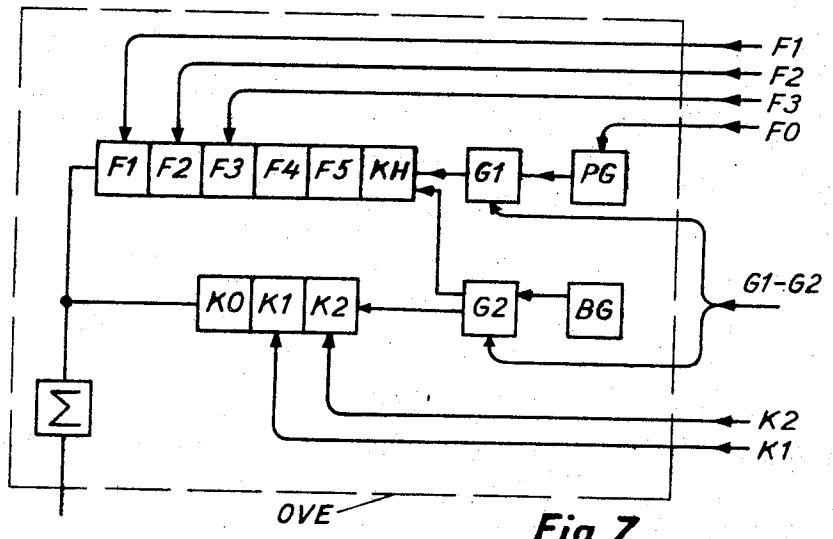


Fig. 7

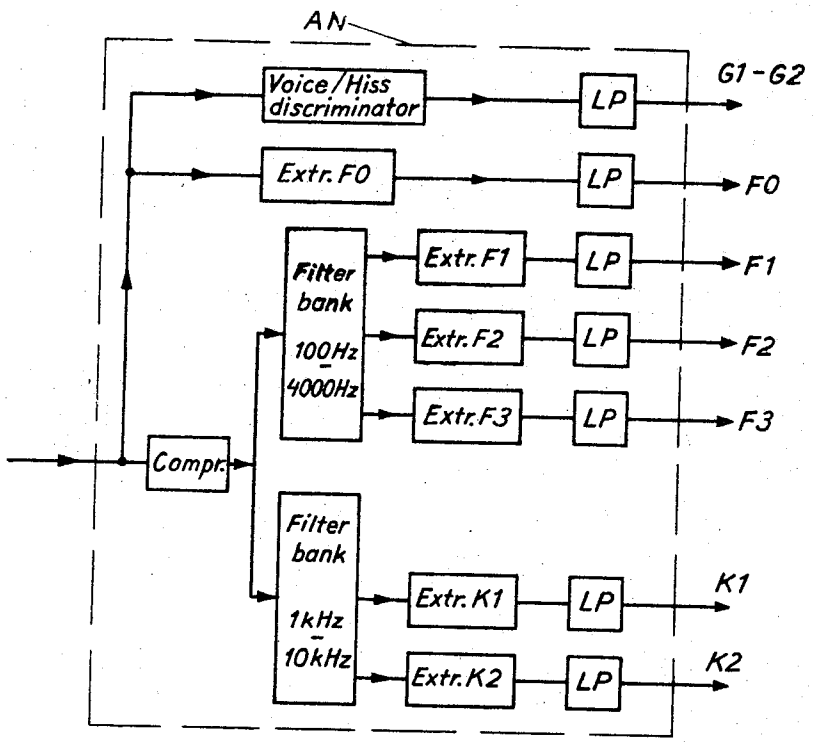


Fig. 8

INVENTOR.
GUNNAR FANT

BY Hane and Hydick
ATTORNEYS

Oct. 10, 1967 G. FANT 3,346,695

VOCODER SYSTEM

Filed April 27, 1964 3 Sheets-Sheet 3

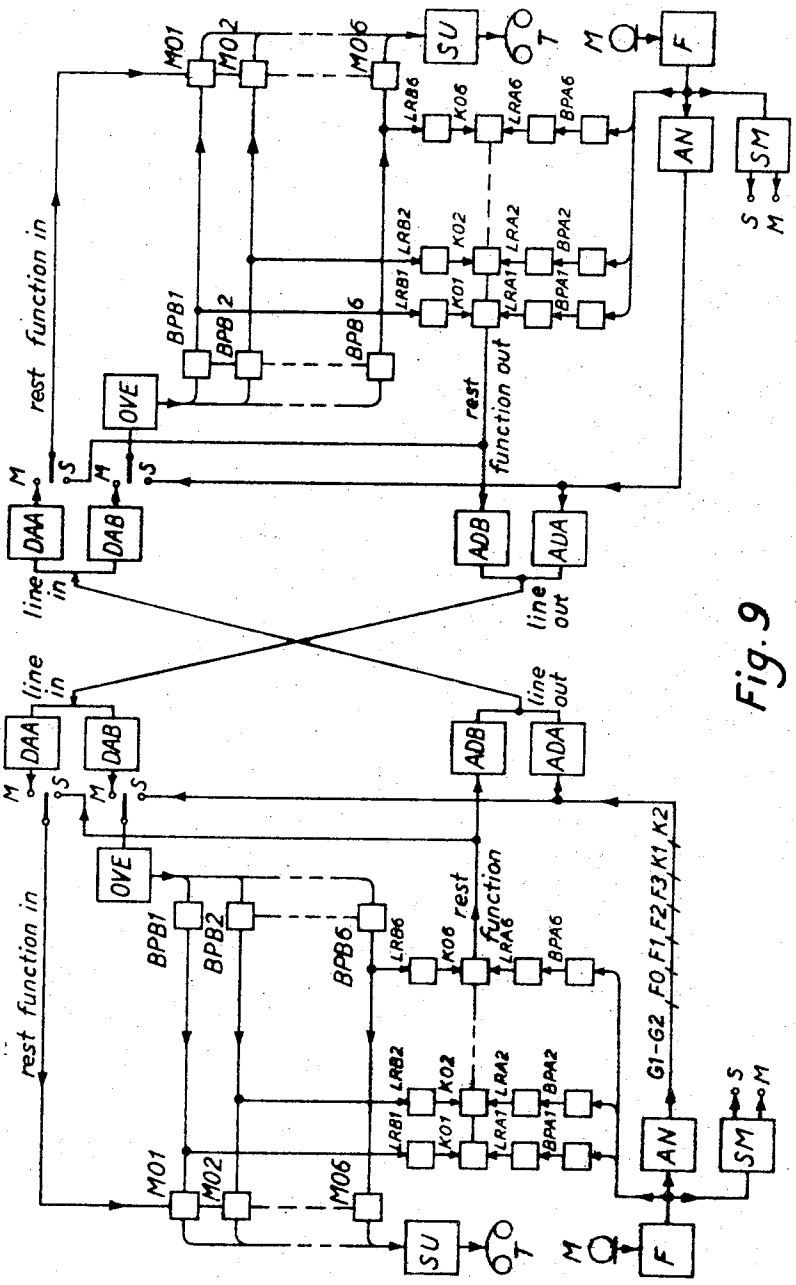


Fig. 9

INVENTOR.
GUNNAR FANT
BY Hane and Nydick
ATTORNEYS

United States Patent Office 3,346,695 Patented Oct. 10, 1967

3,346,695

VOCODER SYSTEM

Gunnar Fant, Shirnervagen 1, Djursholm, Sweden

Filed Apr. 27, 1964, Ser. No. 362,771
Claims priority, application Sweden, May 8, 1963, 5,027/63
4 Claims. (Cl. 179—1)

This invention relates to a vocoder system, i.e. a speech compression system of the analysis-synthesis type, comprising a sender part which carries out a continuous analysis of the speech in respect to voice fundamental periodicity, formant frequencies and relative intensity levels within the spectrum, and a receiver part for the synthesis and the reproduction of the original speech.

As it is known, each speech sound is characterized by a sound spectrum, varying more or less in time, within which spectrum the energy generally is concentrated in certain characteristic frequency ranges, the so-called formants. The voiced sounds may be described by means of a line spectrum within the frequency range 50–4000 c.p.s., the origin of which is a voice fundamental frequency and its harmonics, generated by the vocal cords, while the frequency positions and relative intensity values of the formants are mainly dependent on the resonance effect when the sound passes through the throat, nose and mouth. The voiceless sounds have a continuous noise spectrum, the most important frequency range of which is 1000–8000 c.p.s. Also these sounds generally have a formant structure.

In a conventional formant vocoder the sound spectrum is determined by a number of parameter values representing average values during time intervals of about 20 ms. An example of spectrum determining parameter values which have been used, is the voice fundamental frequency, the frequencies of the first three and most important formants, the intensity values (amplitudes) of these formants and information as to whether the speech sound interval in question belongs to the voiced or the voiceless type. In the conventional formant vocoder the purpose of the sender is to produce an analysis of the spectrum in accordance with the parameters while the receiver contains a synthesis arrangement in which the speech is continuously reconstructed on the basis of incoming parameter values. The "raw material" for the synthesis comes from two generators, one for voiced sounds and one for voiceless sounds while the formant structure in the spectrum is formed by means of formant circuits connected parallelly to the generators, and resonance frequency and amplification with each circuit are varied so as to obtain an intended variation of the formant frequency and the formant intensity. The output signals from these formant circuits are mixed in a common mixer.

The principle for a vocoder according to the parallel system has been proved among others by Stead and Jones from SRDE in England (the system has been described in "Proceedings of Seminar on Speech Compression and Processing," Air Force Cambridge Research Center, Bedford, Mass., U.S.A., September 1959, AFCR TR–59–198), and it allows as a principle a correct reproducing of the formants in view of frequency and amplitude. The disadvantages consist therein that ranges in the spectrum between the formants may obtain intensity levels completely different from the original value, which results in a decreased naturalness, furthermore that signalling of direct formant intensity values implies a waste of information. An essential part of the intensity values of the formants may namely be derived from the total information concerning the frequency positions of all formants within a range of the sound.

In another type of the formant vocoder, called the series system, which is known by Flanagan's works

(Flanagan and House: "Development and Testing of a Formant-Coding Speech Compression System," J. Acoust. Soc. Am., vol. 31, 1956, pages 1099–1106, and Flanagan: "Note on the Design of Terminal-Analog-Speech Synthetizers," J. Acoust. Soc. Am., vol. 29, 1957, pages 306–310), the synthesis is carried out with the formant circuits arranged in cascade, in which case the regulation of the intensity levels of the formants is not necessary. The intensity values dependent on the formant frequencies will however be introduced automatically to permit an approvable vowel reproduction but a less good consonant reproduction because the formant levels of the consonants are not predictable in the same way as in the case of vowel sounds, i.e. the synthesis model having resonance circuits connected in cascade has only a limited validity for the consonant sounds. According to Flanagan's and House's report (the reference above) the mixing up of consonants will be very apparent.

An object of this invention is to eliminate the above-described disadvantages in parallel systems as well as in series systems. As compared with earlier known series systems a throughout individual variation of the formant level is introduced and the level information is defined in such a way that the variation depending on the formant frequency pattern does not need to be signalled. A more correct voice reproduction is obtained, and an appreciably improved consonant reproduction. As compared with the parallel system a more natural voice reproduction will be obtained and the special definition of the formant level implies a saving of the data capacity required.

The vocoder system according to the invention is substantially characterized by the fact that the sender part includes, besides means for continuous analysis of the speech, band pass filters each having a series-connected rectifier in order to determine by integration the intensity values in the original speech in a number of subsequent frequency ranges, a synthesis device of the same type which is included in the receiving part and which carries out a synthesis in the same manner as the synthesis device on the receiver side, and further band pass filters each having a series-connected rectifier in order to determine by integration the intensity values in the speech obtained by synthesis in said subsequent frequency ranges, comparator means for comparing being arranged in the intensity values in the frequency ranges in the original and in the synthetic speech and to produce a difference or residual function signal which depends on the comparison result and modulates the intensity of the synthetical speech in the respective frequency ranges on the receiver side and possibly on the sender side.

The invention will be described herebelow with reference to the enclosed drawing in which FIG. 1 shows a line spectrum of a vowel. FIG. 2 shows an envelope for a first approximation of the spectrum according to FIG. 1, FIG. 3 shows the result of a spectral level measurement of the original sound within a number of frequency ranges; FIG. 4 shows a corresponding measurement of the synthesis approximation and makes clear the meaning of the so-called residual function; FIG. 5 shows a spectrum for a voiceless sound; FIG. 6 shows a sound spectrum of a nasal sound; FIGS. 7 and 8 show details of FIG. 9; and FIG. 9 shows a block diagram of a vocoder system according to the invention.

FIG. 1 shows a line spectrum of a voiced sound, the vowel [e], where each line corresponds to a harmonic of the voice fundamental and F1, F2, F3 and F4 represent the first four formants. FIG. 2 shows a first synthesis approximation consisting of a spectrum envelope which can be considered as the sum of single resonance curves, one for each formant. The envelope is unambiguously determined by the formant frequencies of the sound. As it appears this first approximation which is based on the

formant frequencies without specific amplitude information gives according to an earlier known method (see Fant: "Automatic Extraction of Formant Frequencies From Continuous Speech," Ericsson Technics, vol. 15, No. 1, 1959, pages 110–118), only an incomplete reconstruction of the original spectrum envelope.

According to the invention the sound spectrum is divided into frequency ranges and within each range a comparison is carried out between the energy of the first synthesis approximation according to FIG. 2 and the original sound spectrum according to FIG. 1. This is indicated in FIGS. 3 and 4. According to the example the sound spectrum is divided into six frequency ranges: A=0–350, B=350–900, C=900–1600, D=1600–2600, E=2600–4500 and F=4500–9000 c.p.s. The difference value between the synthesis approximation and the original sound spectrum that indicates the so-called spectral residual function, is indicated in FIG. 4 in the form of staple diagrams. This residual function is used according to the invention to correct the deviation between the direct synthesis approximation and the natural speech, so that the final product of the synthesis obtains the same intensity levels as the natural speech within each of the six frequency ranges. According to the invention a synthesis approximation is carried out both in the sender and in the receiver and the residual function is sent to the receiver to correct the synthesis approximation made in the receiver. The same correction is carried out also in the sender so that the person speaking can hear a synthetic version of his own speech and can compensate technical deficiencies with a better voice technique. The principle will be explained more in detail with the description of a vocoder system according to the invention.

Similar principles as for voiced sounds are also valid for voiceless sounds. FIG. 5 shows a spectrum of a voiceless sound(s) and an approximation for this spectrum consisting of a hollow or anti-resonance K0 and two formants K1 and K2. Such an approximation has turned out to be completely sufficient to produce a satisfactory sound reproduction. The anti-resonance can be unambiguously determined by K1 and K2 and thus it does not need to be determined by the analysis FIG. 6 shows a sound spectrum of a nasal sound (m) and as it appears this sound spectrum differs from the sound spectrum of a vowel substantially by the relative intensity of the formants. The residual function which is used for the correction according to the invention makes possible that the nasal sound can be well distinguished from a vowel.

FIG. 8 shows an embodiment of a sound analyzer which substantially contains means known in connection with a formant vocoder (see for example G. Fant and K. N. Stevens: Fortschritte der Hochfrequenztechnik, Band 5, 1960, page 225), one means for distinguishing voiced sounds G1 from voiceless sounds G2, one for determining the voice fundamental frequency F0 of voide sounds and one for determining the formant frequencies F1, F2, F3. Furthermore means are added for determining the frequency positions of the first and second consonant formants K1 and K2.

An example of a block diagram for a synthesis arrangement is shown in FIG. 7. It constitutes a simplification of an earlier published synthesis arrangement (G. Fant: The Acoustics of Speech, Proceedings of the Third International Congress on Acoustics, Stuttgart, 1959, page 200), in which are included cascade-connected filters for the first five formants F1, F2, F3, F4, F5 and also a correction network of highpass character KH being active over the whole frequency range. Incoming control signals for varying the frequency positions of F1, F2 and F3 also determine the positions of F4, F5 and KH. Parallelly to this so-called F-system, a synthesis device is arranged for the spectrum of the voiceless sounds in the frequency range above 3500 c.p.s. This system consists of an anti-resonance circuit K0 connected in cascade with two resonant circuits K1 and K2. A pulse generator PG is arranged to supply the synthesis system F with "raw material" of voiced sounds having the voice fundamental frequency F0 while another generator BG gives a noise signal which supplies voltage to the K-system. The switch G1 is intended to open and close the connection between the generator PG and the F-system while the switch G2 opens and closes the connection between the generator BG and both the F- and the K-system. In voiceless intervals there is no need to vary the F1-unit of the F-system. The data capacity which is released hereby, is utilized for the frequency variation of K1 and K2 and of K0 which is unambiguously determined by K1 and K2.

If the principle of the invention is applied to such a system not only the vowel sounds will be reproduced correctly but also other patterns differing from the pure vowel patterns may be reproduced. A typical example is a relative attenuation of F1 and an increase of the spectral level at frequencies below F1, characteristic for nasalized vowels. Also the second and the third formant may be attenuated greatly in order to obtain an F2 that is very weakened in relation to vowel sounds (compare FIG. 6). Without the correction effect of the residual function according to the invention, a nasal would thus be reproduced as a vowel. The residual function gives furthermore a greater naturalness and correspondency in the spectrum than is obtained in a pure series system which implies a better reproducing of individual voice characteristics and of such occasional displacements of the relative levels within different frequency ranges which are associated with accentuation, rhythm and phrasing.

FIG. 9 shows a block diagram of a vocoder system built up according to the principle of the invention. The sound is supplied from a microphone M through an amplifier F to an analyzer AN which corresponds to the arrangement shown in FIG. 8, and to six band pass filters BPA1–BPA6. The band pass filters correspond to the six frequency ranges A–E mentioned as examples hereabove. Each of them is intended to supply through rectifiers LRA1–LRA6 signals corresponding to the instantaneous energy values within the associated frequency ranges, to comparators KO1–KO6 each belonging to one of the filters. The analyzer AN sends the parameters F1, F2, F3, K0, K1, K2, G1, G2, G3, F0 through an analog-digital converter ADA to the receiver where they are supplied through a digital-analog converter DAB to a synthesis arrangement OVE according to FIG. 7, reconstructing the sound spectrum with the deficiencies discussed above. According to the invention a similar synthesis arrangement OVE is, however, arranged in the sender and it participates actively in the calculation of those measuring values of the analysis which are to be sent. The output of the synthesis arrangement is connected to six band pass filters BPB1–BPB6 which correspond to the six frequency ranges mentioned above and which each through its rectifier LRB1–LRB6 supplies signals corresponding to the instantaneous energy values within the respective range, to the comparators KO1–KO6. The comparators thus carry out a comparison between the synthesis approximation obtained from the synthesis arrangement OVE and the original speech in the six frequency ranges, obtained from the microphone amplifier and these six values, or the residual function, are sent through an analog-digital converter ADB to the receiver where they through a digital-analog converter DAA are fed to six modulators MO1–MO6. These modulators also receive the synthesis approximation from arrangement OVE in the receiver, so that a sound spectrum corrected by the residual function is obtained in the telephone receiver T through the summing means SU. The final product of the synthesis thus obtains the same intensity levels as the natural speech in each of the six frequency ranges. This correction is carried out not only in the receiving unit but also in the sender so that the person speaking can hear a synthetic version of his own speech as mentioned above.

For this purpose the sender contains in the same way as the receiver. Six modulators which obtain the residual function from the comparators KO1–KO6, and the synthesis approximation from the synthesis arrangement OVE of the sender, so that the same level correction is obtained as in the receiver. The synthesis system is switched between sending and receiving by means of a switch SM which is controlled by the microphone current.

As it appears from FIG. 9 analog-digital converters are included in the system for signals going out from a sender, and digital-analog converters for incoming signals in accordance with general practice for speech compression systems in which binarily coded signals are used for the transmission. In connection with these converters also circuits known per se have to be included in order to convert space-divided information to time-divided information and vice versa, as is easy to understand. With a suitable scheme for quantizing the data capacity of the system might be held within 600–1200 bits/s. which is sufficiently low for PCM-modulated data transmission in telephone lines. The invention is applicable not only to systems where the formant circuits for synthesis are located in series but also to parallel systems and channel vocoder systems, i.e., parallel systems having many channels with unvariable frequencies.

I claim:

1. A vocoder system comprising a sender part and a receiver part, means in said sender part for continuously analyzing a speech signal with respect to values of voice fundamental periodicity, formant frequencies and voice/hiss relationship and for producing a characteristic signal for each of said analyzed values, means in said sender part for transmitting said characteristic signals, means in said receiver part for receiving said characteristic signals, synthesis means in said receiver part for reconstructing the voice fundamental frequency and the formant frequencies by means of said characteristic signals so as to reconstruct the original speech signal, said synthesis means including output means for transmitting said speech signal, said sender part including furthermore a first group of bandpass filters each belonging to a definite frequency range among a number of sequential frequency ranges and each having a series connected rectifier so as to produce intensity values by integration in said definite frequency ranges of the original speech signal, further synthesis means in said sender part and of the same type as in said receiver part for reconstructing the voice fundamental frequency and the formant frequencies by means of the characteristic signals produced by said analyzing means in said sender part so as to reconstruct the original speech, said further synthesis means including output means for transmitting said speech signal, a second group of bandpass filters in said sender part belonging to the same frequency ranges to which the bandpass filters in said first group belong and each being connected to said output means of said further synthesis means and each having a series connected rectifier so as to produce intensity values by integration in said definite frequency ranges of the synthesized speech signal, comparator means at least in said sender part for each of said frequency ranges in the original speech signal and the synthesized speech signal to produce a residual function signal for each of the ranges in dependence on the difference between said two intensity values, means for transmitting said residual function signals each corresponding to the residual function signal in one of said frequency ranges, and modulating means at least in said receiver part for receiving a residual signal corresponding to one of said

frequency ranges and for modulating the output of one of said bandpass filters in said second group of filters so as to compensate for differences of intensity in the respective frequency range between the original and the synthesized speech signal, and a summation circuit arranged at least in said receiver part for summing the output signals of said modulators.

2. A vocoder system according to claim 1 comprising means in said sender part for converting said characteristic signals to digital signals before transmitting them, means in said receiver part for converting said digital signals back to said characteristic signals, means in said sender part for converting said residual function signals to digital signals before transmitting them, and means in said receiver part for converting said digital signals back to residual function signals after they have been received.

3. A vocoder system according to claim 1 wherein said sender part further comprises modulating means connected to the output of said further synthesis means in said sender part, and a summation circuit supplied by outputs of said modulating means so as to produce in said sender part a corrected reconstruction of the speech signal.

4. A vocoder system comprising a sender part and a receiver part: said sender part including input means for receiving speech signals, means for analyzing said speech signals for frequency characteristics and generating a characteristic signal for each of the frequency characteristics, synthesis means for receiving the characteristic signals and for reconstructing therefrom the speech signals, a first plurality of frequency selective signal generator means receiving the speech signals, each of said signal generator means generating a first intensity signal having an amplitude functionally related to the amplitude of signals in said speech signals having frequencies within a definite frequency range among a plurality of sequential frequency ranges respectively, a second plurality of frequency selective signal generator means receiving the reconstructed speech signals, each of said signal generator means generating a second intensity signal having an amplitude functionally related to the amplitude of signals in the reconstructed speech signals having frequencies within one of said definite frequency ranges respectively, comparator means responsive to said first and second pluralities of frequency selective signal generator means for generating a residual function signal for each of said definite frequency ranges dependent upon the difference between the associated first and second intensity signals; said receiver part including a further synthesis means for receiving the characteristic signals generated by said analyzing means for reconstructing therefrom the speech signals, a plurality of bandpass filter means responsive to said further synthesis means, each of said bandpass filter means transmitting the portion of the reconstructed speech signals having frequencies within one of said definite frequency ranges respectively, a plurality of modulating means, each of said modulating means receiving the residual function signal and said portion of the reconstructed speech signals associated with one of said definite frequency ranges respectively for generating a compensated reconstructed portion of the speech signals, and means for combining the compensated reconstructed portions of the speech signals of all of said modulating means to generate compensated reconstructed speech signals.

No references cited.

KATHLEEN H. CLAFFY, *Primary Examiner.*

R. MURRAY, *Assistant Examiner.*