

- [54] METHOD AND SYSTEM FOR SPEECH COMPRESSION
[76] Inventor: David C. Coulter, 9613 Pembroke Pl., Vienna, Va. 22180
[22] Filed: Feb. 28, 1973
[21] Appl. No.: 336,705
[52] U.S. Cl. 179/1 SA
[51] Int. Cl. G101 1/10
[58] Field of Search 179/1 SA, 15.55 R, 15.55 T, 179/15 BW

- [56] References Cited
UNITED STATES PATENTS
3,127,476 3/1964 David..... 179/1 SA
3,139,487 6/1964 Logan 179/15.55 R
3,431,362 3/1969 Miller..... 179/15.55 R
3,483,325 12/1969 Steward 179/1 SA

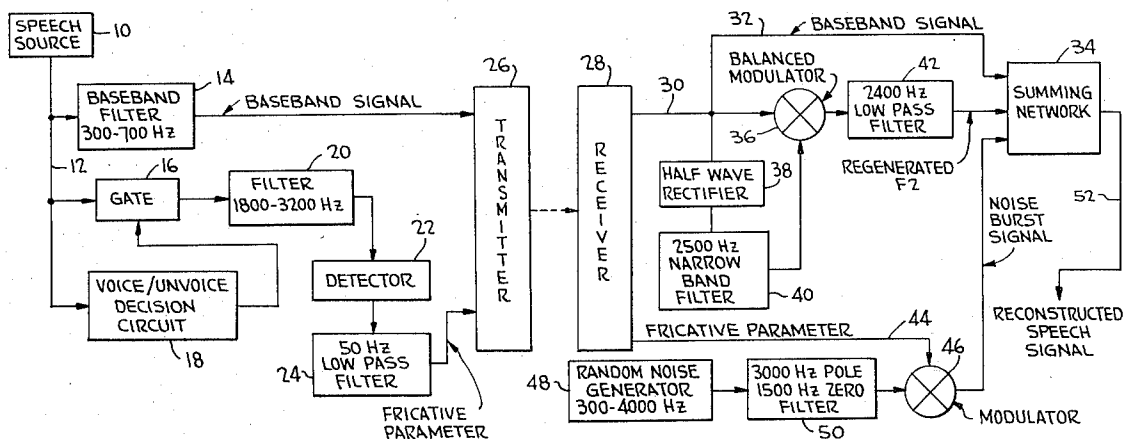
3,499,986 3/1970 Focht..... 179/1 SA

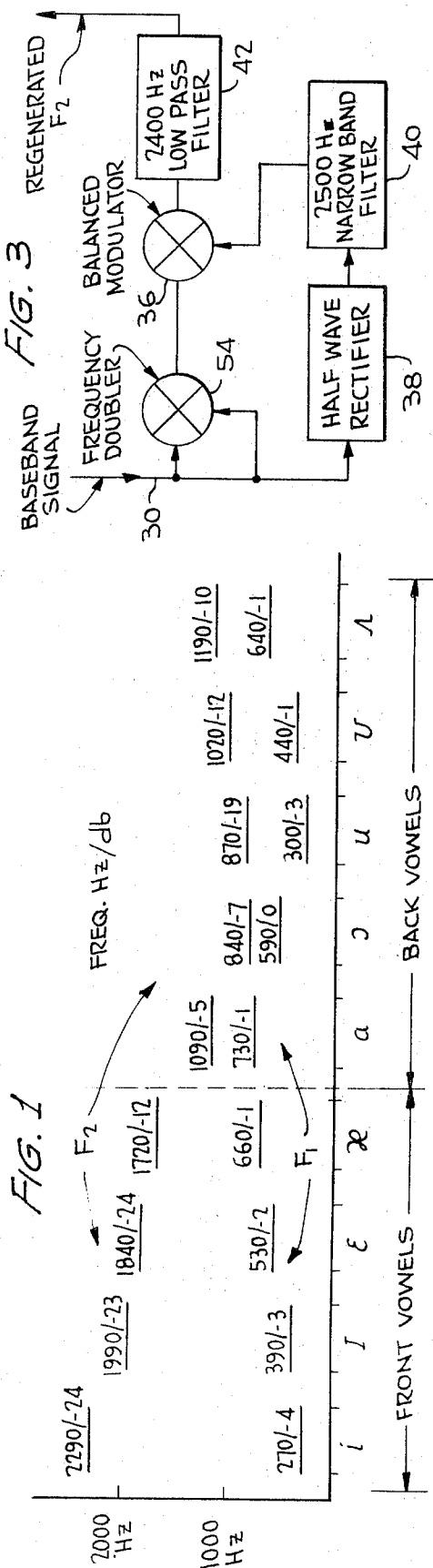
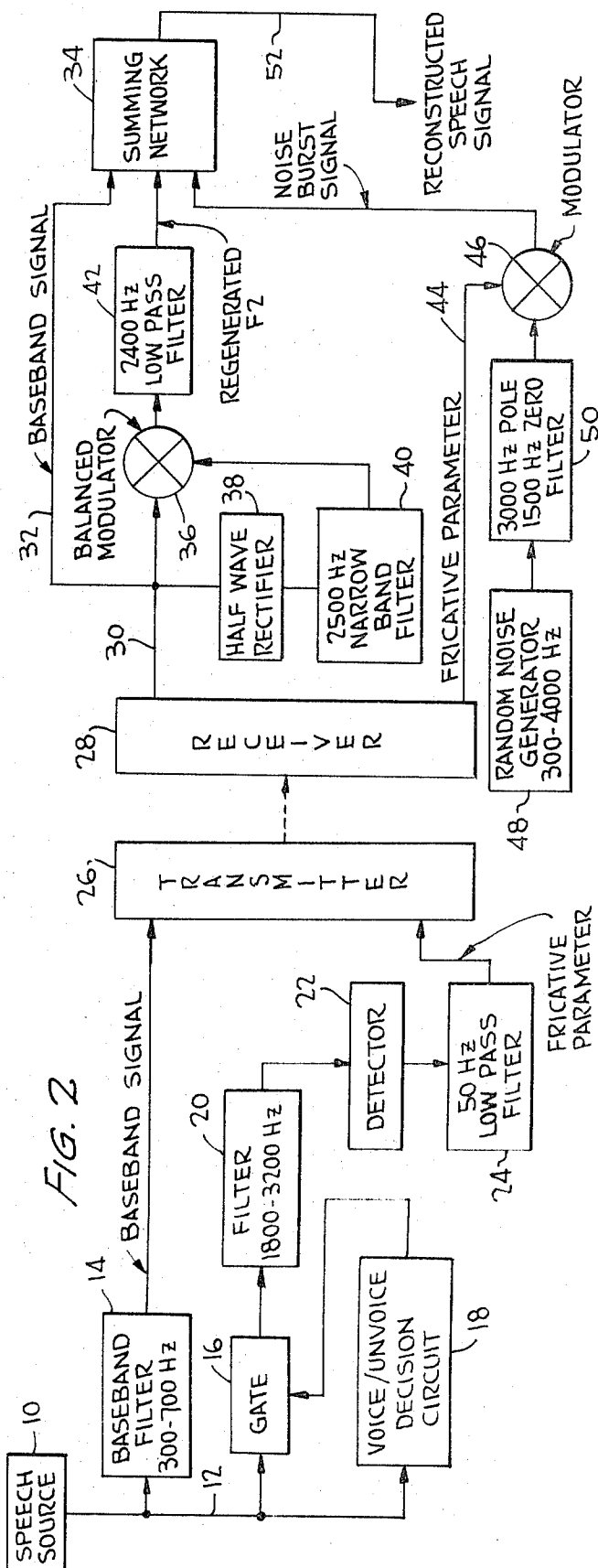
Primary Examiner—David L. Stewart
Attorney, Agent, or Firm—Robert H. Epstein

[57] ABSTRACT

A method and system of speech compression by baseband vocal regeneration wherein a baseband signal including the first formant, approximately between 300 and 700 Hz, is transmitted to a receiver and supplied to a signal combining network along with a second formant regenerated from the baseband signal to provide a reconstructed speech signal including the first and second formants. An unvoice or fricative parameter can be transmitted along with the baseband signal and utilized at the receiver to generate a synthetic noise burst signal for combining with the baseband signal and the regenerated second formant.

18 Claims, 3 Drawing Figures





METHOD AND SYSTEM FOR SPEECH COMPRESSION

BACKGROUND OF THE INVENTION

I. Field of the Invention

The present invention pertains to speech compression and, more particularly, to a method and system for speech compression to reduce transmission bandwidth by using baseband vocal regeneration.

2. Discussion of the Prior Art

Since the inception of practical speech compression coinciding with the channel vocoder concept of H. Dudley, a great amount of effort has been expended in attempts to improve the naturalness of reproduced speech transmitted by speech compression techniques. The channel vocoder method of speech compression is based on the recognition of the carrier nature of speech; that is, the carrier or pitch frequency of the larynx is modulated by the voice tract frequencies formed in the mouth. Channel vocoders, thus, include a pitch tracker for detecting and tracking the pitch frequency and supplying a signal through a frequency counter, low pass filter or the like representative of average frequency of the larynx as well as a plurality of bandpass filters covering the voice frequency range from about 300 to 3,000 KHz to measure the energy in each portion of the voice frequency and provide a stepped representation of the spectrum. The outputs from the filters are detected with low pass filters, and the pitch and voice tract signals are transmitted and used to regenerate speech at the receiver.

Channel vocoders, while providing acceptable results from an information transmission standpoint, have the disadvantage that the reconstructed or reproduced speech at the receiver lacks naturalness and has poor voice quality. That is, intelligibility, speaker recognition and emotional characteristics of the reconstructed speech have not been adequately reproduced with the result that the reconstructed speech has an irritating synthetic sound. One of the primary reasons for the unnaturalness of the reconstructed speech is due to the necessity of tracking pitch in that pitch trackers have not been developed sufficiently to provide accurate tracking.

In attempting to avoid the requirement of a pitch tracker in speech compression systems, it was suggested to band the first few harmonics of the pitch frequency together and to transmit this band of energy, which is within a general range of from 300 to 700 Hz, instead of transmitting the pitch parameter. At the receiver, the band of energy is spread to generate all possible harmonics up to 3,000 Hz instead of just the first few harmonics. This voice excited vocoder provides good voice quality; however, a 9600 bit/sec. rate is required to transmit the band of energy as compared with the 2400 bit/sec. rate required for channel vocoders due to the slowly varying nature of the pitch frequency. Another disadvantage of the voice excited vocoder is that commercial applications are limited due to the complexity of the equipment and the high cost thereof. That is, at the present, the cost of voice excited vocoder systems approaches the cost of extra speech channels; and, therefore, voice excited vocoders do not represent a viable method for reducing bandwidth channels for commercial voice transmission.

Other attempts to provide speech compression systems have included tracking the first three formants,

which have been considered the most important in transmitting speech intelligence. However, the frequency bands of the first three formants are not well defined in that the frequencies of the first three formants overlap; and, accordingly, this lack of isolation of formant frequency range has rendered tracking of the first three formants extremely difficult. Thus, formant tracking speech compression systems requiring the tracking of the first three formants have not been developed to the point of ultimate fruition.

Speech compression systems based on mathematical derivation and linear prediction have been proposed; however, after development and hardware implementation, such systems have been revealed to require pitch tracking or its equivalent and, thus, have the disadvantages associated with channel and formant tracking vocoders and improved variations and modifications thereof.

Another prior method of compressing the bandwidth required for speech transmission is based on transmitting certain key bands in the normal speech range, generally from three to six bands, whose total bandwidth is less than that required for normal speech transmission. This method has the disadvantage that when the bands are made sufficiently narrow to achieve a useful bandwidth compression of 4:1 or greater, the quality of the reconstructed speech is characterized by heavily filtered sound which is disturbing to the listener.

SUMMARY OF THE INVENTION

Accordingly, it is a primary object of the present invention to overcome the disadvantages of the prior art by providing a speech compression method and system having more natural and recognizable voice quality and intelligence characteristics of reconstructed speech while being relatively simple in nature and exhibiting low cost relative to transmission bandwidth reduction.

The primary object of the present invention is accomplished by transmitting a baseband signal from approximately 300 to 700 Hz and utilizing the baseband signal at a receiver to generate an upper formant signal from known definable relationships between formants, particularly between the first and second formants in accordance with the Peterson-Barney estimates.

Another object of the present invention is to utilize the inverse relationship of the first and second formants for front vowels in regenerating the second formant at a receiver when only the first formant is transmitted due to the increased importance of the second formant in defining front vowels relative to the importance of the second formant in defining back vowels.

A further object of the present invention is to regenerate the second formant from the first formant by frequency translation with the use of a local frequency source and a balanced modulator to translate and invert the first formant to produce a signal harmonically related to the pitch frequency representing the second formant for front vowels.

The present invention has another object in the reconstruction of good quality, articulate speech from a small portion of the normal voice bandwidth with or without an additional signal of low bandwidth.

Yet an additional object of the present invention is to eliminate the requirement for devices which track either pitch or formant frequency in a speech compression system while reconstructing speech based on formant frequencies.

A further object of the present invention is to provide a speech compression method requiring only the transmission of a baseband signal including the first formant and a fricative parameter including noise burst, unvoice information with the second formant being regenerated from a portion of the spectrum containing the first formant and the unvoice noise burst being synthetically generated corresponding to the fricative parameter and combined with the regenerated second formant and the baseband signal to provide a speech output.

Still further, another object of the present invention is to regenerate the second formant from a baseband signal including the first formant by forming the regenerated second formant of frequencies only harmonically related to the pitch frequency of the baseband signal.

Some of the advantages of the present invention over the prior art are that no tracking devices are required, a speech signal can be processed to permit voice communications occupying a reduced bandwidth with the same signal-to-noise ratio or at a lower bit rate than the unprocessed speech, sufficient intelligibility and voice naturalness are maintained so that the speaker can be recognized from his voice and the voice quality does not sound synthetic, the system of the present invention can be used in tandem without appreciable reduction in voice quality or intelligence due to the synthetic production of noise burst signals and the regenerating of the second formant from the baseband signal and the circuitry required is simple thereby permitting the method and system of the present invention to be economically implemented and reliable in operation.

The present invention is generally characterized in a method of transmitting a speech signal in reduced bandwidth including the steps of transmitting a baseband signal having a frequency range including the first formant, receiving the baseband signal regenerating the second formant from the baseband signal, and combining the baseband signal with the regenerated second formant to reconstruct the speech signal.

The present invention is further generally characterized in a speech compression system including a speech source providing a speech signal having first and second formants, a filter for passing a baseband signal of the speech signal including the first formant, means for transmitting and receiving the baseband signal, a signal combining network for producing a reconstructed speech signal, means supplying the received baseband signal to the signal combining network, and means responsive to the received baseband signal for regenerating the second formant and supplying the regenerated second formant to the signal combining network whereby the reconstructed speech signal includes the first formant and the regenerated second formant.

Other objects and advantages of the present invention will become apparent from the following description of the preferred embodiments taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a plot illustrating the relationship between first and second formant frequencies for the most common vowel sounds.

FIG. 2 is a diagram of a speech compression system according to the present invention.

FIG. 3 is a diagram of a modification of the speech compression system of FIG. 2.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Speech is characterized by the production of larynx impulses at a repetitive frequency referred to as pitch frequency. These larynx pulses are rich in harmonics due to their impulse nature, and these harmonics along with the pitch fundamental frequency are transmitted through the voice tract to the outside via the lips. According to the shape and size of the voice tract, as determined by the tongue, lower jaw and lips, certain frequency ranges are selectively emphasized such that three principle bands of harmonics called formants are emphasized. The frequency positions and relative motions of these bands of harmonics determine the transmission of speech intelligence for voiced sounds. The other major class of sounds bearing speech intelligence are unvoice or fricative sounds, generated without vibrating the larynx by blowing air through a constricted point somewhere along the voice tract. The combination of these voiced and unvoiced sounds comprises the generation of human speech; and, accordingly, speech signals are formed of the pitch frequency and the formants which are harmonics of the pitch frequency and which form the voice portion of speech and the fricative sounds which form the unvoice portion of speech.

The relationship of the first and second formants F_1 and F_2 are illustrated in FIG. 1 for the most common vowel sounds, and from FIG. 1, it will be appreciated that a practically linear inverse relationship exists between F_1 and F_2 for the front vowels, whereas F_1 and F_2 have a substantially direct relationship for the back vowels. While the frequency and amplitude of the first formant F_1 are important to human perception of front and back vowel sounds, it has been found that the second formant F_2 is relatively significant to the perception of front vowel sounds only in that perception of back vowel sounds is strongly linked to the F_1 frequency only. Thus, in accordance with the method and system of the present invention, only a small band of energy including F_1 is transmitted, and F_2 is regenerated at the receiver to maintain sufficient intelligibility and voice naturalness such that the speaker can be recognized from the reconstructed speech and the voice quality does not have synthetic characteristics.

One example of a system according to the present invention is illustrated in FIG. 2 and includes a speech source 10 providing a speech signal on an output 12 to a baseband filter 14 passing frequencies within the band from 300 to 700 Hz and to a gate 16 and a voice/unvoice decision circuit 18. Decision circuit 18 is operative to enable gate 16 when an unvoice signal is received to pass the unvoice signal to a bandpass filter 20 passing the band of frequencies between 1,800 and 3,200 Hz, the output from filter 20 being supplied through a detector 22 to a low pass filter 24 to provide a DC to 50 Hz output. Thus, a 300 to 700 Hz band analog signal is supplied from baseband filter 14 and a DC to 50 Hz signal is supplied from low pass filter 24, the signals from filters 14 and 24 providing a baseband signal including F_1 and a noise burst or fricative parameter, respectively, and being supplied to a suitable transmitter 26 for transmitting the signals to a receiver 28.

The baseband signal and the fricative parameter can be transmitted in any suitable manner, for instance, by

combining the baseband signal and the fricative parameter by summation, by modulating the fricative parameter on a subcarrier of approximately 200 Hz by frequency modulation or by amplitude modulation.

At the receiver, the baseband signal is provided on an output 30 and supplied directly to a summing network 34 on a lead 32 and to a balanced modulator 36 and through a half-wave rectifier 38 to a 2,500 Hz ringing circuit or narrow-band filter 40. The output from the narrowband filter 40 is supplied to the balanced modulator 36 which has an output supplied through a low pass filter 42 having an upper frequency cutoff skirt at approximately 2,400 Hz to summing network 34. The fricative parameter is supplied on an output 44 to a modulator 46 which also receives as an input noise from a 300 to 4,000 Hz, broadband random noise generator 48 through a filter 50 having a 3,000 Hz pole and a 1,500 Hz zero. The output from modulator 46 is supplied to summing network 34. Thus, summing network 34 receives a baseband signal input on lead 32 including F_1 , and input from filter 42 corresponding to regenerated F_2 and a noise burst input from modulator 46 and combines the signals to provide a reconstructed speech signal on an output 52 for reproduction by a suitable transducer.

In operation, baseband filter 14 passes a band of frequencies from 200 to 1,000 Hz, preferably from 300 to 700 Hz, such that the transmitted signal includes a baseband signal including the first formant F_1 . The voice/unvoice decision circuit 18 detects whether the speech signal is voiced, that is pitch excited and periodic, or unvoiced, that is, random noise-like, as is conventional and described in detail on pages 95 - 97 of a report entitled "Formant Tracking Vocoder System" prepared for the U.S. Army under Contract Number DA36-039-AMC-00006(E), Oct. 1965. When an unvoiced sound is detected, gate 16 is enabled to permit the speech signal to be supplied to the filter 20, the speech signal supplied to filter 20 thus corresponding to a noise burst or fricative parameter representing the S, T and K sounds, for example. Filter 20 is centered in the band of greatest noise energy, that is, between 1,800 and 3,200 Hz; and, the energy in the noise band is detected by detector 22 and passed by low pass filter 24 to provide the DC to 50 Hz fricative parameter. Thus, a baseband signal and a fricative parameter are transmitted representative of the speech signal.

At the receiver, the noise bursts from the DC to 50 Hz fricative parameter are reproduced by controlling the modulator 46 by the fricative parameter to pass noise bursts from the combination of the random noise generator 48 and the filter 50 of amplitude proportional to the fricative parameter voltage. The filter 50 is centered around 3,000 Hz and, preferably, has a pole at 3,000 Hz and a zero at 1,500 Hz which represents a median tuning in order to pass a sound somewhere between S and SH. Thus, when an unvoice signal is detected by the voice/unvoice decision circuit 18, the fricative parameter transmitted corresponding thereto is supplied to the modulator 46 such that the modulator 46 provides the product of the fricative parameter and the noise from filter 50 to produce a variable amplitude noise burst signal proportional to the original unvoice sound to the summing network 34.

In the circuit for regenerating F_2 , the half-wave rectifier 38 distorts the baseband signal to generate a spectrum including all the harmonics of the pitch frequency

at high frequencies which are in the pass band of the 2,500 Hz filter 40. The narrowband filter 40 has a center frequency of 2,500 Hz and a band wide enough to pass at least one harmonic of the basic pitch frequency, preferably a band 150 Hz wide to provide plus or minus 75 Hz at the 3dB points with broad skirt selectivity. While filter 40 is described as having a center frequency of 2,500 Hz, the center frequency can be within the range of from 2,200 to 2,800 Hz and preferably from 2,400 to 2,600 Hz. Basically, the filter 40 should have a bandwidth such that at least one pitch harmonic of the highest larynx frequency anticipated is always captured by the filter. The balanced modulator 36 combines or processes the baseband signal with the frequency from filter 40 to produce a double sideband of plus and minus frequencies relative to the pitch harmonic near 2,500 Hz captured in the filter 40, and only the lower sideband of the signal is passed by low pass filter 42 such that the output from filter 42 represents a folded over or inverted F_1 constituting a regenerated F_2 signal for the front vowels. That is, the low pass filter 42 separates the lower sideband which contains the difference between 2,500 Hz and the F_1 frequency which constitutes the regenerated second formant F_2 . While a narrowband filter 40 is disclosed for isolating a harmonic of the pitch frequency, this function can be provided by any suitable circuitry, for example, by a phase-locked loop or a combination of a narrow-band filter, a clipper and another narrowband filter in order to assure that only a single harmonic is modulated with the baseband signal.

The regenerating circuit, thus, assures that the regenerated F_2 is harmonically related to the pitch frequency in that the frequency of the signal from the narrowband filter 40 is a harmonic of the pitch frequency of the baseband signal and, since the baseband signal is formed of harmonics of the pitch frequency, modulator 36 operates to add and subtract one harmonic frequency from another. Accordingly, after passing through low pass filter 42, the regenerated second formant F_2 is formed of a harmonic frequency or frequencies obtained by subtracting the harmonic or harmonics of the baseband signal from the harmonic from narrowband filter 40. If a local oscillator, such as a 2,500 Hz sine wave were utilized in place of the harmonic generating circuitry, new frequencies, not harmonics of the fundamental pitch frequency, would be produced if the local oscillator was not an exact multiple of pitch frequency; and, thus, the regenerated frequencies of F_2 would not be harmonics of pitch frequency and, therefore, would produce "beats" in the human ear. The use of a ringing circuit or narrowband filter instead of a local oscillator is extremely important to the voice quality obtained with the present invention in that the use of such circuitry assures that no new frequencies which are not harmonics of fundamental pitch frequency would be produced, it being appreciated that in normal human speech, all frequencies in voiced sounds are harmonics of the pitch frequency.

With respect to the back vowels, a front-back vowel detector may be utilized to permit regeneration of F_2 for the back vowel, such as that described on pages 81-83 of the above-mentioned report entitled "Formant Tracking Vocoder System" prepared for the U.S. Army under Contract Number DA36-039-AMC-00006(E), Oct. 1965; however, it has been found that for some of the back vowels, F_2 is transmitted within

the baseband signal and, further, that folding over or inverting F_1 to regenerate F_2 for the back vowels did not adversely affect intelligibility or voice quality probably because sufficient F_2 is passed by the baseband to permit the ear to pick up the back vowel. While the reason for the good quality obtained even with the inverted regeneration of F_2 for back vowels is not completely understood, the use of the F_2 regenerating circuitry of FIG. 2 is preferable due to the advantages of simplified circuitry and economy.

A modification of the second formant regenerating circuitry is illustrated in FIG. 3 with the primary difference relative to the regenerating circuitry of FIG. 2 being that the frequency of the baseband signal is doubled by a frequency doubler 54 prior to being supplied to balanced modulator 36. Frequency doubler 54 is essentially a squaring modulator receiving the baseband signal at both inputs. By utilizing the frequency doubler, the frequency subtracted in modulator 36 from the harmonic frequency isolated by filter 40 is doubled to produce a regenerated F_2 closer to the F_2 shown in FIG. 1 for each front vowel. Of course, the baseband signal could be altered in any manner, such as tripling, while maintaining its harmonic relationship with the pitch frequency in order to obtain a regenerated F_2 closer to the original F_2 .

The method and system of speech compression according to the present invention, thus, permit transmission bandwidth to be reduced to 700 Hz with an open channel between the fricative parameter and the baseband signal or further to 450 Hz. By transmitting a baseband of 300 to 700 Hz, the most important portion of the speech signal is utilized to reconstruct the speech signal in that this baseband essentially contains the personality of the speaker as well as pitch information to obviate pitch tracking and the first formant F_1 which contains excellent intelligibility. The combination of the regenerated F_2 with the baseband signal provides a reconstructed speech signal having good voice quality and intelligibility, and these reconstructed speech characteristics are further improved when the synthetic noise burst signal is combined with the baseband signal and the regenerated F_2 .

Inasmuch as the present invention is subject to many variations, modifications and changes in detail, it is intended that all subject matter described above or shown in the accompanying drawings be interpreted as illustrative and not in a limiting sense.

I claim:

1. A method of transmitting a speech signal in reduced bandwidth comprising the steps of transmitting a baseband signal having a frequency range including the first formant;
receiving the baseband signal;
regenerating the second formant from the baseband signal by producing harmonics of the pitch frequency of the baseband signal, isolating at least one of the harmonics and processing the at least one harmonic with the baseband signal; and
combining the baseband signal with the regenerated second formant to reconstruct the speech signal.
2. A method of transmitting a speech signal in reduced bandwidth comprising the steps of transmitting a baseband signal having a frequency range including the first formant;
regenerating the second formant from the baseband signal of frequencies harmonically related to the

pitch frequency of the baseband signal including generating harmonics of the pitch frequency of the baseband signal and modulating one of the generated harmonics with the baseband signal; and
combining the baseband signal with the regenerated second formant to reconstruct the speech signal.

3. A method of transmitting a speech signal as recited in claim 2 wherein said step of regenerating the second formant includes modulating the baseband signal to provide a double sideband signal and passing only the lower sideband signal to be combined with the baseband signal.

4. A method of transmitting a speech signal as recited in claim 3 and further comprising the steps of transmitting a fricative parameter of the speech signal, generating a synthetic noise burst signal in accordance with the fricative parameter and combining the noise burst signal with the baseband parameter and the regenerated second formant to produce the speech signal.

5. A method of transmitting a speech signal as recited in claim 3 wherein said baseband signal represents the band of frequencies from 200 to 1,000 Hz.

6. A method of transmitting a speech signal as recited in claim 3 wherein said baseband signal represents the band of frequencies from 300 to 700 Hz.

7. A method of transmitting a speech signal as recited in claim 3 wherein the one harmonic is within a range of from 2,200 to 2,800 Hz.

8. A method of transmitting a speech signal as recited in claim 3 wherein the one harmonic is within a range of from 2,400 to 2,600 Hz.

9. A speech compression system comprising a speech source providing a speech signal having first and second formants;

filter means for passing a baseband signal of said speech signal including said first formant;
means for transmitting said baseband signal;
means for receiving said baseband signal;
signal combining means for producing a reconstructed speech signal;

means supplying said baseband signal from said receiving means to said signal combining means; and
means receiving said baseband signal from said receiving means for generating at least one harmonic of the pitch frequency and processing said at least one harmonic with said baseband signal to regenerate said second formant and supply said regenerated second formant to said signal combining means whereby said reconstructed speech signal includes said first formant and said regenerated second formant.

10. A speech compression system comprising a speech source providing a speech signal having first and second formants;

filter means for passing a baseband signal of said speech signal including said first formant;
means for transmitting said baseband signal;
means for receiving said baseband signal;
signal combining means for producing a reconstructed speech signal;

means supplying said baseband signal from said receiving means to said signal combining means; and
means receiving said baseband signal from said receiving means for regenerating said second formant from said baseband signal and supplying said regenerated second formant to said signal combining means including means for processing said

baseband signal to produce harmonics of the pitch frequency of said speech signal, means for isolating at least one of said harmonics, and means for modulating said at least one harmonic with said baseband signal to regenerate said second formant harmonically related to said baseband signal whereby said reconstructed speech signal includes said first formant and said regenerated second formant.

11. A speech compression system as recited in claim 10 wherein said isolating means includes a narrowband filter having a center frequency within the range of from 2,200 to 2,800 Hz.

12. A speech compression system as recited in claim 10 wherein said isolating means includes a narrowband filter having a center frequency substantially at 2,500 Hz.

13. A speech compression system as recited in claim 12 wherein said processing means is a half wave rectifier, said modulating means is a balanced modulator providing a double sideband signal and said means for regenerating said second formant further includes a low pass filter for passing the lower sideband signal from said balanced modulator to said signal combining means.

14. A speech compression system as recited in claim 13 wherein said speech signal includes unvoice sounds and further comprising means for detecting said unvoice sounds and generating a fricative parameter corresponding thereto and means responsive to the received fricative parameter for synthetically generating a noise burst signal and supplying said noise burst signal to said signal combining means whereby said reconstructed speech signal includes said first formant, said regenerated second formant and said noise burst signal.

15. A speech compression system as recited in claim 10 wherein said speech signal includes unvoice sounds

and further comprising means for detecting said unvoice sounds and generating a fricative parameter corresponding thereto and means responsive to the received fricative parameter for synthetically generating a noise burst signal and supplying said noise burst signal to said signal combining means whereby said reconstructed speech signal includes said first formant, said regenerated second formant and said noise burst signal.

16. A receiver for reconstructing a speech signal represented by a frequency band signal comprising means for receiving said frequency band signal; means receiving said frequency band signal from said receiving means for processing said frequency band signal to produce harmonics of said frequency band signal; means for isolating at least one of said harmonics; means for modulating said at least one harmonic with said frequency band signal to produce a modulated signal; and signal combining means receiving said frequency band signal from said receiving means and said modulated signal from said modulating means for combining said frequency band signal and said modulated signal to produce a reconstructed speech signal.

17. A receiver as recited in claim 16 wherein said modulating means includes a balanced modulator providing a double sideband signal and low pass filter means for passing the lower sideband signal to said signal combining means as said modulated signal.

18. A receiver as recited in claim 17 wherein said isolating means includes a narrowband filter having a center frequency within the range of from 2,200 to 2,800 Hz.

* * * * *

40

45

50

55

60

65