(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 081 685 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
07.03.2001 Bulletin 2001/10

(51) Int. Cl.[7]: **G10L 21/02**

(21) Application number: **00118147.8**

(22) Date of filing: **29.08.2000**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **01.09.1999 US 388266**

(71) Applicant: **TRW Inc.**
Redondo Beach, California 90278 (US)

(72) Inventors:
- **Lambert, Russell H.**
Fountain Valley, CA 92708 (US)
- **Edmonds, Karina L.**
Pasadena, CA 91106 (US)
- **Hsu, Shi-Ping**
Pasadena, CA 91107 (US)

(74) Representative:
Schmidt, Steffen J., Dipl.-Ing.
Wuesthoff & Wuesthoff,
Patent- und Rechtsanwälte,
Schweigerstrasse 2
81541 München (DE)

(54) System and method for noise reduction using a single microphone

(57) A noise reduction technique for use with a single microphone channel. The technique provides a noise reduction framework that allows multiple parameters to be adjusted optimally for any given application, noise environment or automatic speech recognition (ASR) system. The system of the invention includes a fast Fourier transform (FFT) circuit (10) with a bandpass filter to remove known noise frequencies from a speech signal, a speech detector (14), a noise estimator (16) that updates a noise estimate only when speech is not detected, a spectrum subtraction circuit (18) to subtract the noise estimate from the speech and noise signal spectrum, and a speech emphasis circuit (20), which further emphasizes speech signal components with respect to any residual noise. The resulting noise-reduced signals in the frequency domain can be either input directly to an automatic speech recognition (ASR) system, or transformed back to the time domain for use in a voice communication system. A noise monitor (90) may be added to the system, to determine when noise reduction is appropriate, and to avoid unwanted signal distortion when noise reduction is not needed. For further improved performance, input signals are first processed into blocks that are each used twice in forming data blocks for the FFT circuit and subsequent processing, and a triangular weighting window is applied (44) at the FFT input.

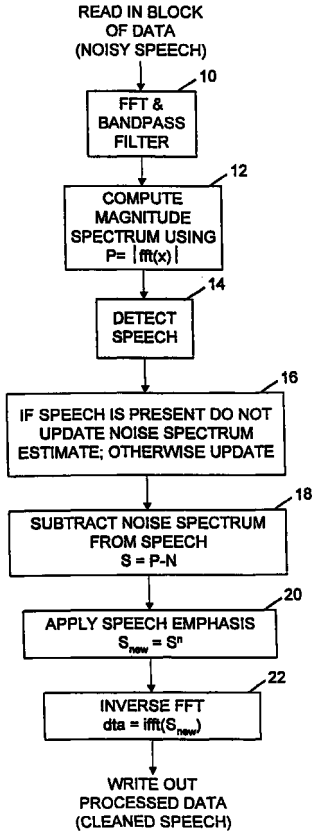

FIG. 1

EP 1 081 685 A2

Description

BACKGROUND OF THE INVENTION

[0001] This invention relates generally to techniques for reliable conversion of speech data from acoustic signals to electrical signals in an acoustically noisy and reverberant environment. There is a growing demand for "hands-free" cellular telephone communication from automobiles, using automatic speech recognition (ASR) for dialing and other functions. However, background noise from both inside and outside an automobile renders in-vehicle communication both difficult and stressful. Reverberation within the automobile combines with high noise levels to greatly degrade the speech signal received by a microphone in the automobile. The microphone receives not only the original speech signal but also distorted and delayed duplicates of the speech signal, generated by multiple echoes from walls, windows and objects in the automobile interior. These duplicate signals in general arrive at the microphone over different paths. Hence the term "multipath" is often applied to the environment. The quality of the speech signal is extremely degraded in such an environment, and the accuracy of any associated ASR systems is also degraded, perhaps to the point where they no longer operate. As an example, recognition accuracy of an ASR system as high as 96% in a quiet environment could drop to well below 50% in a moving automobile.

[0002] Another related technology affected by noise and reverberation is speech compression, which digitally encodes speech signals to achieve reductions in communication bandwidth and for other reasons. In the presence of noise, speech compression becomes increasingly difficult and unreliable.

[0003] There are a number of prior art systems that effect active noise cancellation in the acoustic field. The active noise reduction approaches cancel acoustic noise signals by generating an opposite signal, sometimes referred to as "anti-noise," through one or more transducers near the noise source, to cancel the unwanted noise signal. This technique often creates noise at some other location in the vicinity of the speaker, and is not a practical solution for canceling multiple unknown noise sources, especially in the presence of multipath effects.

[0004] Accordingly, there is still a significant need for reduction of the effects of noise in a reverberant environment, such as the interior of a moving automobile. As discussed in the following summary, the present invention addresses this need.

SUMMARY OF THE INVENTION

[0005] The present invention resides in a system and method for reducing noise in speech signals obtained from a single microphone in a noisy environment. The present invention is a general noise reduction framework that allows multiple parameters to be adjusted optimally for any given application, noise environment or automatic speech recognition (ASR) system. Briefly, and in general terms, the system of the invention comprises a fast Fourier transform (FFT) circuit for transforming blocks of input microphone data to a frequency domain representation; a bandpass filter to remove selected frequency bands in which noise is known to be present; a speech detector for sensing the presence of speech signals in microphone data; a noise spectrum estimator updated only for data blocks in which no speech signals are detected; a spectrum subtraction circuit, for subtracting the estimated noise spectrum from microphone signals containing noise and speech signal components; and a speech emphasis circuit, for emphasizing speech signal components with respect to any residual noise after operation of the spectrum subtraction circuit, to provide a noise-reduced speech signal in the frequency domain.

[0006] The system may further comprise means for reconstructing time-domain data from the noise-reduced speech signal in the frequency domain, including an inverse fast Fourier transform circuit for transforming blocks of data from the frequency domain back into the time domain, whereby the noise-reduced speech signals are more intelligible in voice communication systems. Alternatively, the system may further comprise an automatic speech recognition (ASR) system connected to receive the noise-reduced speech signals in the frequency domain, whereby the ASR system operates more reliably to generate selected control signals.

[0007] Preferably, the speech emphasis circuit raises signals in the frequency domain by a power N, where N is a positive quantity greater than one.

[0008] In the invention as disclosed, the input signals are presented to the noise reduction system in blocks of "A" samples each, and data blocks of size "2A" samples each are presented to the FFT circuit. The system further comprises means for combining input signal blocks of "A" samples in pairs to form data blocks. Moreover, the means for combining input signal blocks uses each input signal block twice, such that a currently input signal block is placed in a second half of a current data block and is then placed in a first half of a next data block. The system may further comprise means for applying a triangular weighting window to each data block; and the means for reconstructing time-domain data includes means for combining the first half of each reconstructed data block with the second half of a reconstructed data block saved from processing the previous data block, time-domain samples with a uniform envelope are reconstructed and unwanted artifacts of block processing are minimized.

[0009] In accordance with another aspect of the invention, the system further comprises a noise monitor to provide an indication of when use of noise reduction would be desirable; and means for selecting the noise-reduced signal when noise level detected in the noise monitor is detected as relatively high, and for selecting the original speech with noise signal when the detected noise level is relatively low.

[0010] The invention may also be defined in terms of a method for reducing noise in signals received by a single microphone in a noise environment. Briefly, and in general terms, the method comprises the steps of transforming blocks of input data from a single microphone from a time-domain representation to a frequency-domain representation; filtering out selected frequency bands to minimize the effect known noise sources; detecting the presence of speech in each block of data signals; estimating noise by updating a noise spectrum estimate when no speech is detected; subtracting the noise spectrum estimate from the input speech and noise signals; and emphasizing speech signal components with respect to noise signal components, by raising the result of the subtracting step to the Nth power, where N is a positive quantity greater than one, to provide frequency-domain speech signal data with a reduced noise content.

[0011] The method may also include the step of reconstructing time-domain data from the noise-reduced speech signal in the frequency domain, including transforming blocks of data from the frequency domain back into the time domain, whereby the noise-reduced speech signals are more intelligible in voice communication systems. Alternatively, the method includes the step of transmitting the noise-reduced speech signals in the frequency domain to an automatic speech recognition (ASR) system, whereby the ASR system operates more reliably to generate selected control signals.

[0012] Preferably the method step of emphasizing speech signal components includes raising signals in the frequency domain by a power N, where N is a positive quantity greater than one.

[0013] More specifically the method further includes the steps of presenting input signals to the noise reduction system in blocks of "A" samples each; presenting data blocks of size "2A" samples to the FFT circuit; combining input signal blocks of "A" samples in pairs to form data blocks, the combining step including using each input signal block twice, such that a currently input signal block is placed in a second half of a current data block and is then placed in a first half of a next data block; applying a triangular weighting window to each data block; and in the reconstructing step, combining the first half of each reconstructed data block with the second half of a reconstructed data block saved from processing the previous data block. Time-domain samples with a uniform envelope are reconstructed and unwanted artifacts of block processing are minimized with use of this method.

[0014] The method may further comprise the steps of continually monitoring the noise level with a noise monitor, to provide an indication of when use of noise reduction would be desirable; selecting the noise-reduced signal when the noise level detected by the noise monitor is detected as relatively high; and selecting the original speech and noise signal when the detected noise level is relatively low.

[0015] It will be appreciated from the foregoing summary that the present invention represents a significant advance in noise reduction techniques. The combination of features summarized above results in a speech signal that has noise greatly reduced, resulting in more intelligible speech when the signals are used in voice communication systems, and more reliable ASR system operation when the signals are used to operate for ASR and related systems. Other aspects and advantages of the invention will become apparent from the following more detailed description, taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016]

FIGURE 1 is a block diagram of a noise cancellation system in accordance with the present invention;

FIG. 2 is a more detailed block diagram of the noise cancellation system of the invention; and

FIG. 3 is a set of four related graphs, showing time domain correlation of a noise signal with itself, i.e., autocorrelation, and the time domain autocorrelation of a speech signal;

FIG. 4 is a block diagram depicting an alternative embodiment of the invention in which a noise detector is used to control operation of the noise cancellation system; and

FIG. 5 is a block diagram showing how the noise cancellation system of the invention may be integrated into an existing automatic speech recognition (ASR) system.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0017] As shown in the drawings, the present invention is concerned with a technique for significantly reducing the effects of noise in the detection of speech in a noisy and reverberant environment, such as the interior of a moving automobile. The quality of speech transmission from mobile telephones in automobiles has long been known to be poor

much of the time. Noise from within and outside the vehicle result in a relatively low signal-to-noise ratio and reverberation of sounds within the vehicle further degrades the speech signals. Available technologies for automatic speech recognition (ASR) and speech compression are at best degraded, and may not operate at all in the environment of the automobile.

[0018] In accordance with the present invention, and as shown in FIG. 1, a combination of processing steps, including spectral subtraction of noise, is performed to achieve a significant reduction in noise level. A noisy speech signal is converted to digital samples and is input a block of samples at a time for processing in a fast Fourier transform (FFT) circuit, as indicated in block 10. Upon conversion to the frequency domain by the fast Fourier transform, the signal is first bandpass filtered, as also indicated in block 10. Then the magnitude spectrum is computed, as indicated in block 12, as the absolute value of the FFT function. Then each block of data, still in the frequency domain, is analyzed to detect the presence or absence of speech, as indicated in block 14. An essential aspect of the invention is to reduce noise by spectral subtraction of noise spectrum estimate. Ideally, this estimate should be based on data obtained when speech is absent. As indicated in block 16, if speech is present, the noise spectrum estimate is not updated, but if speech is absent the noise estimate is updated.

[0019] As indicated in block 18, the noise spectrum estimate is subtracted from the noisy speech signal spectrum, still in the frequency domain. Then, as indicated in block 20, speech is further emphasized over any residual noise by raising the speech signal (obtained after spectral subtraction of the noise) to the n^{th} power, where n is optimized to provide the most desirable result. Finally, as indicated in block 22, the blocks of data in the frequency domain are subjected to inverse transformation by an inverse FFT circuit, which outputs a "cleaned" speech signal in the time domain.

[0020] The functions depicted in FIG. 1 are depicted in more detail in FIG. 2. The general parameter set referred to in FIG. 2 is defined in the following table:

Parameter Name	Description	Range	Units
A	Block size (FFT size is 2A)	Real positive integer (usually a power of 2)	Samples
B	Input low cut-off point	0-parameter C	Frequency (Hz)
C	Input high cut-off point	Parameter B-sample rate/2	Frequency (Hz)
D	Spectral compression factor	Real positive (greater than 1)	Unitless
E	Speech location lower limit	0-parameter F	Frequency (Hz)
F	Speech location upper limit	Parameter E- sample rate/2	Frequency (Hz)
G	Running average energy update parameter	Real positive (between 0 and 1)	Unitless
H	Speech detect threshold parameter	Real positive	Unitless
I	Running average noise spectrum update parameter	Real positive (between 0 and 1)	Unitless
J	Speech enhancement parameter	Real positive (greater than 1)	Unitless

[0021] The functions shown in FIG. 2 may be implemented in any desired hardware or software configuration. In an experimental configuration, the noise cancellation system was implemented as software with code in a Microsoft Visual C++ compiler running on a personal computer in real time. Input speech signals are sampled and input in blocks of A samples each. Computation blocks for FFT processing are formed to contain 2A data samples each. Thus the FFT point size is 2A. For example, A may be 128 samples and 2A, 256 samples.

[0022] Rectangle 40 in FIG. 2 indicates the input of blocks of data. Rectangle 42 indicates that each data computation block of 2A samples is formed from the stream of A-sized blocks in overlapping fashion. More specifically, if the incoming stream of A-sized blocks are designated as block (a), block (b), block (c), block (d) and so forth, then the first data computation block is formed from blocks (a) and (b) together, the next data computation block is formed from blocks (b) and (c) together, the next from blocks (c) and (d) together, and so forth. The reason for overlapping the blocks in this way is to minimize sound artifacts that can be introduced by serially processing the blocks of data. Further, each data computation block, as indicated in rectangle 44, is subjected to "windowing" by a triangular weighting function having the profile of an isosceles triangle centered on the data computation block. Thus, a maximum weight is applied to a sample or samples at the center of the data computation block, and progressively less weight is applied to samples

towards the leading and trailing edges of the block. Because the data computation blocks derive data from overlapping A-sized blocks, these triangular windows also overlap. Moreover, when the signals are later converted to the frequency domain and back to the time domain, the contributions from each adjacent pair of overlapping data computation blocks combine to produce a set of samples having a relatively uniform amplitude envelope.

[0023] After each successive data block is formed and windowed, it is introduced to FFT processing, as indicated in rectangle 46, and then subjected to bandpass filtering between limits defined by parameters B and C, as indicated in rectangle 48. This filtering step eliminates noise at very low and very high frequencies, such as below 300 Hz and above 3,850 Hz. Next, as indicated in rectangle 50, a magnitude spectrum S is computed and placed in a compressed domain using parameter D. $S_{compressed} = S^{1/D}$.

[0024] As indicated in rectangle 52, the speech energy of the current data block is computed by summing the energy in the frequency range given by parameters E and F, such as 400 to 800 Hz, where speech is most likely to be dominant. The average speech energy in this range is kept in a running average estimator, as indicated in rectangle 54, using the computation:

$$SpeechEnergy_{avg} = (1\text{-}G)*SpeechEnergy_{avg} + G*SpeechEnergy_{current}$$

In decision block 56, the current speech energy is compared with H times the average speech energy E_{avg}, which provides a continually adapting speech detection threshold. If the current speech energy is greater that $H*E_{avg}$, then the noise spectrum is not updated, as indicated by path 58. If not, the noise spectrum is updated using parameter I, as indicated in rectangle 60, using the expression:

$$Spectrum_{avg} = (1\text{-}I)*Spectrum_{avg} + I*Spectrum_{current}.$$

The speech spectrum is then computed as the difference between the current spectrum and the noise spectrum estimate, as indicated in rectangle 62. Finally, there is an important speech enhancement step 64, in which the speech spectrum, together with any residual noise component, is raised to the power J, where J is selected to be greater than one. Raising the signal to a power greater than one further distinguishes speech components from noise components.

[0025] As an example of parameter optimization, the effects of various values of parameter J were observed (while holding all other parameters fixed), as indicated in the following table:

Speech Enhancement Parameter J	Accuracy from ASR
1.5	80%
1.7	81.4%
1.85	84%
1.9	**85.6%**
1.95	81.4%
2.0	80.7%
2.2	76.4%
2.5	67.1%

It will be observed that the best value of parameter J from the standpoint of automatic speech recognition is 1.9.

[0026] If the speech signals are to be transmitted to a human user of the system, they must next be transformed back to the time domain. Reconstruction of the time domain waveform is also performed on a block by block basis. An inverse FFT operation is performed on each data block, as indicated in rectangle 66. The triangularly windowed data samples that result must be added together in a manner that will produce a uniform data envelope for the reconstructed waveform. More specifically, the first half of a reconstructed data block is added to the second half of the previously converted block of data, as indicated in block 68. Because these two half-blocks were originally subject to triangular windowing, they now combine in a complementary way to produce a uniform signal envelope. The second half of the current block is saved for the next block iteration, as indicated in rectangle 70. The combined A samples from the current and previous blocks are output, as indicated in rectangle 72.

[0027] For best performance, a standard "star search" technique may be used, varying one parameter of the method described above while holding all others fixed. Ideally, this should be repeated for each type of speech and for different noise conditions. One of the most critical parameters is the speech emphasis term, J. This was varied from 1.5 to 2.5 while testing the recognition accuracy for each setting of J. The optimum parameter value indicated was for use of the invention in the presence of freeway road and vehicle noise and for spoken connected digits data.

[0028] As shown in FIG. 3, random noise, indicated by graph 80, has a distinctive 'spike' in its autocorrelation function 82, whereas a sine wave has a periodic auto-correlation function. A segment of speech 84 has strong components that are periodic sine waves. Therefore, the speech correlates strongly over several milliseconds, as indicated at 86. In contrast, the noise 80 correlates strongly only at the zero delay point, as indicated by the spike in its autocorrelation function 82. In the correlation domain, the spike due to noise can be easily zeroed out and this is the basis of the spectral subtraction approach used in the present invention.

[0029] The system of the invention has been tested under practical conditions in a moving vehicle, on a freeway with the windows closed and air-conditioning on, and also with the windows partly open. Two types of microphones were considered, omni-directional and unidirectional. Not unexpectedly, the unidirectional microphone led to significantly better recognition accuracy for all background noise levels. The highest recognition accuracy obtained was 86% from freeway driving with the windows up and air conditioning on using connected digits speech data.

[0030] The in-vehicle data were initially collected using a digital recorder and the microphone placement was selected to maximize signal-to-noise ratio (SNR). For both the omni-directional and the unidirectional microphone the position that yields the greatest signal was just above the driver's visor (i.e., directly in front of the source). All the tests were conducted using the passenger as the point source for speech. Since the car cabin is symmetric, the results for the driver's side are expected to be equivalent to those obtained from the passenger side. The speech recorded on the digital recorder in the automobile was sampled at 44.1 kHz and subsequently down-sampled to 8 kHz. In order to ensure the integrity of the audio files after down sampling, the files were tested with an automatic speech recognition (ASR) system. No degradation in ASR performance was observed for a file recorded at 44.1 kHz and down-sampled to 8 kHz.

[0031] In ASR systems, the recognition accuracy is calculated in terms of a digit error rate. The number of substitutions (S), deletions (D) and insertions (I) are divided by the total number of digits (N) tested:

$$Error = \frac{S+D+I}{N} \; x100$$

[0032] A software package designed by Lemout and Hauspie ASR1500 was utilized for testing since it allowed for connected digits and has a relatively short response time. The vocabulary tested consisted of eleven digits; 1-9, zero and oh. Connected digits were selected in order to account for the co-articulation factors in recognition process. In the test procedure, each digit is pronounced approximately fifteen times during a dialogue of a random series of connected digits.

[0033] The recognition accuracy for the digits is significantly improved after the removal of the background noise. With the windows up and air-conditioning on, recognition rates improved from 47% to 86% for a unidirectional microphone, and from 16% to 78% for an omni-directional microphone. With the windows partly open, recognition rates improved from 46% to 83% for a unidirectional microphone, and from less than 10% to 39% for the omni-directional microphone.

[0034] As shown in FIG. 4, background noise level monitoring system 90 may be incorporated into the standard noise cancellation system of the invention, which would then operate only when a specified level of background noise is present. This would eliminate speech degradation from the processing when there is no background noise. The decision need not be a "hard" (on or off) one. Rather the modified system would appropriately blend the processed and unprocessed speech in a continuously varying manner such that the effect of turning on the processing in high noise conditions would not be noticeable to the system user. By way of example, in this embodiment of the invention the monitored noise level is compared against an upper threshold, as indicated in decision block 92, and if the noise exceeds the threshold, the system selects processed (noise-reduced) speech as indicated in rectangle 94. If the monitored noise level is currently below the upper threshold, it is compared with a tower threshold, as indicated in decision block 96. If the noise is below the lower threshold, the original unprocessed speech is selected, as indicated in rectangle 98. If the monitored noise is between the upper and lower thresholds, the system selects a blend of inputs from the original speech and noise-reduced speech signals, as indicated in rectangle 100.

[0035] In another embodiment of the invention, the noise reduction system is incorporated into an automatic speech recognition (ASR) system 104 (FIG. 5). The noise reduction system is the same as the one illustrated in FIG. 1, but without the final inverse FFT process. This will eliminate some of the speech artifacts that are created when transforming back to the time domain waveform. Where the application calls for voice control of the ASR system only, there

is no need to reconstruct the time domain waveform. The inverse FFT function is eliminated from the noise cancellation system and the output of the noise cancellation system is coupled directly to frequency domain inputs 106 of the ASR system 104, which generates appropriate output control signals 108 in response to detection of input speech commands.

[0036] It will be appreciated from the foregoing that the present invention represents a significant advance in noise reduction for a single-microphone installed in noisy environment, such as a moving automobile. In particular, the invention provides a "cleaned" or noise-reduced speech signal that is more intelligible to the human ear and improves reliability of ASR systems. The system of the invention produces either time-domain output for transmission over voice communication systems, or frequency-domain output for direct connection to an ASR system. It will also be appreciated that, although a number of embodiments have been described in detail for purposes of illustration, various modifications may be made without departing from the spirit and scope of the invention. Accordingly, the invention should not be limited except as by the appended claims.

Claims

1. A noise reduction system for a single microphone in a noise environment, the system comprising:

a fast Fourier transform (FFT) circuit for transforming blocks of input microphone data to a frequency domain representation;
a bandpass filter to remove selected frequency bands in which noise is known to be present;
a speech detector for sensing the presence of speech signals in microphone data;
a noise spectrum estimator updated only for data blocks in which no speech signals are detected;
a spectrum subtraction circuit, for subtracting the estimated noise spectrum from microphone signals containing noise and speech signal components; and
a speech emphasis circuit, for the emphasizing speech signal components with respect to any residual noise after operation of the spectrum subtraction circuit, to provide a noise-reduced speech signal in the frequency domain.

2. A noise reduction system as defined in claim 1, and further comprising:

means for reconstructing time-domain data from the noise-reduced speech signal in the frequency domain, including an inverse fast Fourier transform circuit for transforming blocks of data from the frequency domain back into the time domain, whereby the noise-reduced speech signals are more intelligible in voice communication systems.

3. A noise reduction system as defined in claim 1, and further comprising:

an automatic speech recognition (ASR) system connected to receive the noise-reduced speech signals in the frequency domain, whereby the ASR system operates more reliably to generate selected control signals.

4. A noise reduction system as defined in claim 2, wherein:

input signals are presented to the noise reduction system in blocks of "A" samples each;
data blocks of size "2A" samples each are presented to the FFT circuit;
the system further comprises means for combining input signal blocks of "A" samples in pairs to form data blocks;
the means for combining input signal blocks uses each input signal block twice, such that a currently input signal block is place in a second half of a current data block and is then placed in a first half of a next data block;
the system further comprises means for applying a triangular weighting window to each data block; and
the means for reconstructing time-domain data includes means for combining the first half of each reconstructed data block with the second half of a reconstructed data block saved from processing the previous data block, time-domain samples with a uniform envelope are reconstructed and unwanted artifacts of block processing are minimized.

5. A method for reducing noise in signals generated by a single microphone in a noise environment, the method comprising the steps of:

transforming blocks of input data from a single microphone from a time-domain representation to a frequency-

domain representation;

filtering out selected frequency bands to minimize the effect known noise sources;
detecting the presence of speech in each block of data signals;
estimating noise by updating a noise spectrum estimate when no speech is detected;
subtracting the noise spectrum estimate from input speech and noise signals; and
empasizing speech signal components with respect to noise signal components, by raising the result of the subtracting step to the Nth power, where N is a positive quantity greater than one, to provide frequency-domain speech signal data with a reduced noise content.

6. A method as defined in claim 5, and further comprising:

reconstructing time-domain data from the noise-reduced speech signal in the frequency domain, including transforming blocks of data from the frequency domain back into the time domain, whereby the noise-reduced speech signals are more intelligible in voice communication systems.

7. A method as defined in claim 6, and further including the steps of:

presenting input signals to the noise reduction system in blocks of "A" samples each;
presenting data blocks of size "2A" samples each to the FFT circuit;
combining input signal blocks of "A" samples in pairs to form data blocks, the combining step including using each input signal block twice, such that a currently input signal block is placed in a second half of a current data block and is then placed in a first half of a next data block;
applying a triangular weighting window to each data block; and
in the reconstructing step, combining the first half of each reconstructed data block with the second half of a reconstructed data block saved from processing the previous data block, wherein time-domain samples with a uniform envelope are reconstructed and unwanted artifacts of block processing are minimized.

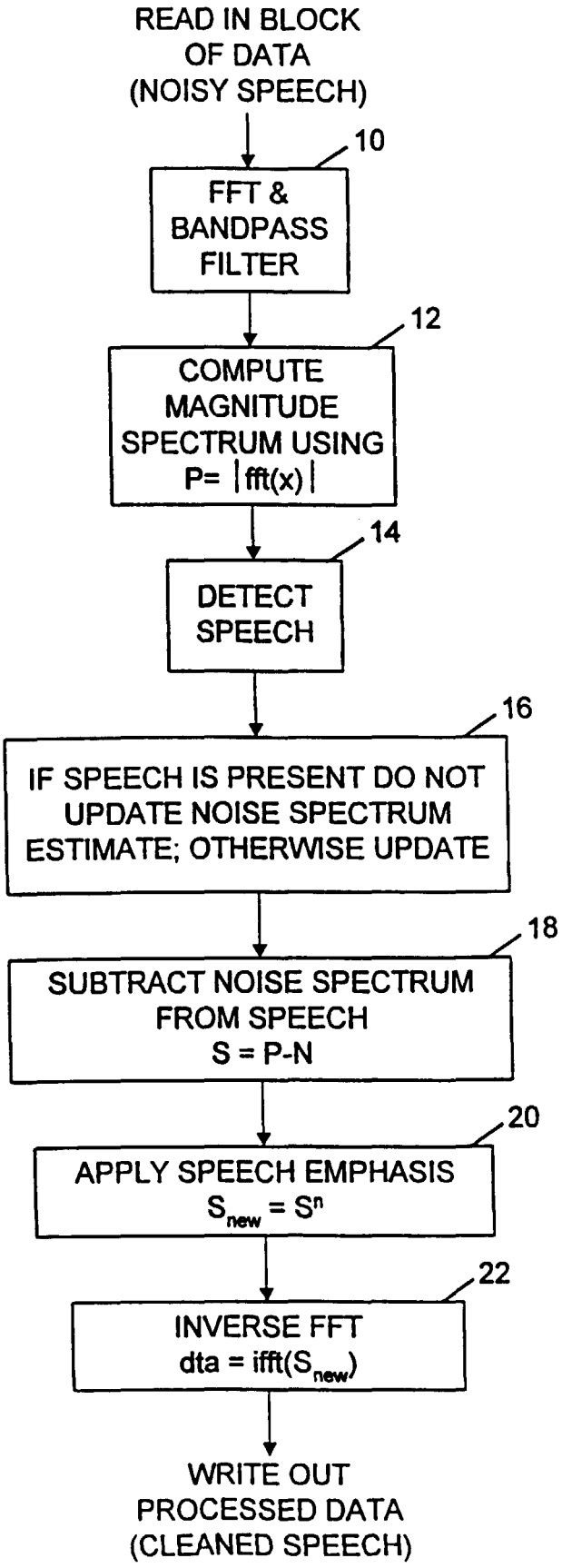
READ IN BLOCK OF DATA (NOISY SPEECH)
10
FFT & BANDPASS FILTER
12
COMPUTE MAGNITUDE SPECTRUM USING P= |fft(x)|
14
DETECT SPEECH
16
IF SPEECH IS PRESENT DO NOT UPDATE NOISE SPECTRUM ESTIMATE; OTHERWISE UPDATE
18
SUBTRACT NOISE SPECTRUM FROM SPEECH S = P-N
20
APPLY SPEECH EMPHASIS $S_{new} = S^n$
22
INVERSE FFT dta = ifft(S_{new})
WRITE OUT PROCESSED DATA (CLEANED SPEECH)

FIG. 1

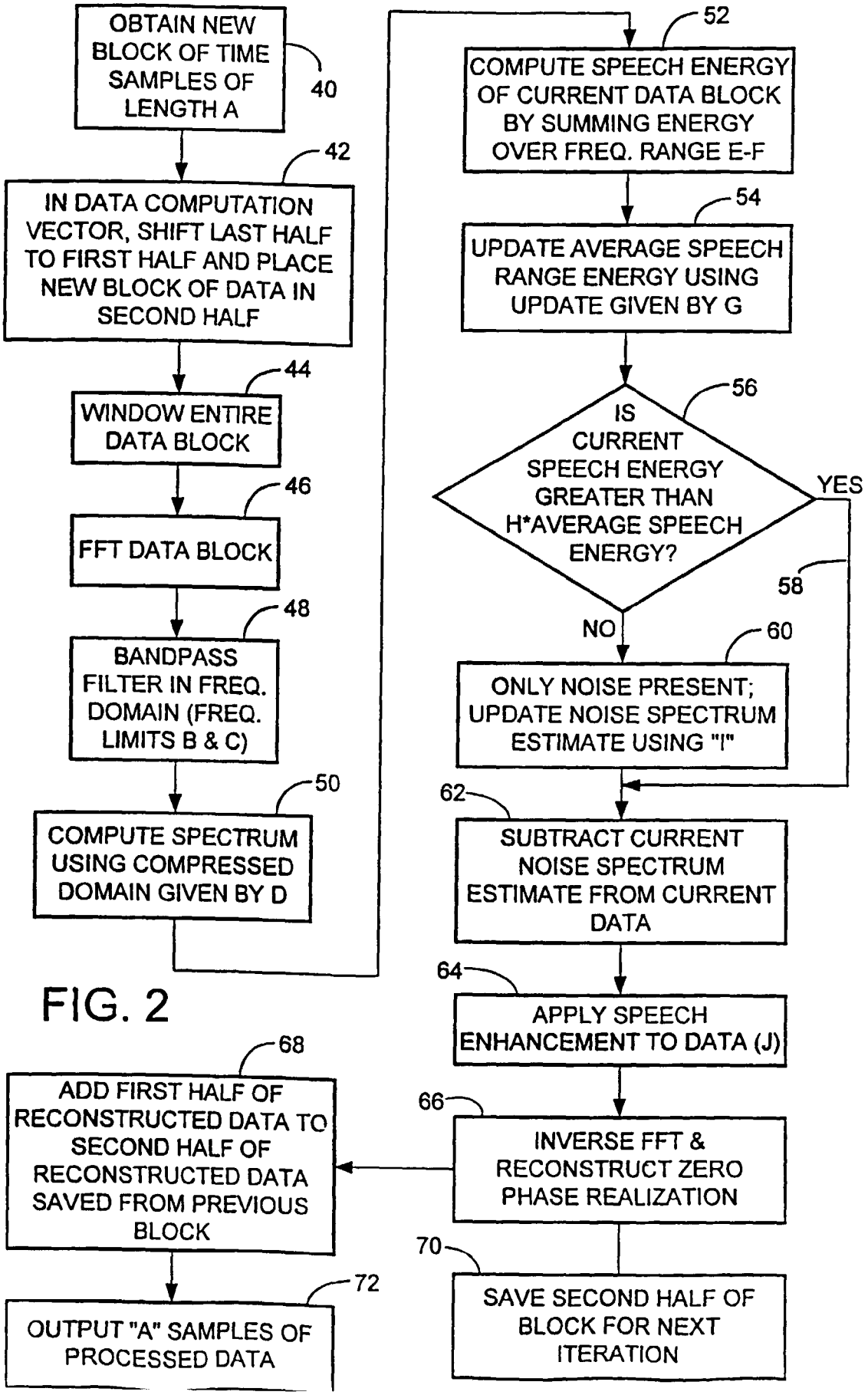
OBTAIN NEW BLOCK OF TIME SAMPLES OF LENGTH A
40
IN DATA COMPUTATION VECTOR, SHIFT LAST HALF TO FIRST HALF AND PLACE NEW BLOCK OF DATA IN SECOND HALF
42
WINDOW ENTIRE DATA BLOCK
44
FFT DATA BLOCK
46
BANDPASS FILTER IN FREQ. DOMAIN (FREQ. LIMITS B & C)
48
COMPUTE SPECTRUM USING COMPRESSED DOMAIN GIVEN BY D
50
COMPUTE SPEECH ENERGY OF CURRENT DATA BLOCK BY SUMMING ENERGY OVER FREQ. RANGE E-F
52
UPDATE AVERAGE SPEECH RANGE ENERGY USING UPDATE GIVEN BY G
54
IS CURRENT SPEECH ENERGY GREATER THAN H*AVERAGE SPEECH ENERGY?
56
YES
58
NO
ONLY NOISE PRESENT; UPDATE NOISE SPECTRUM ESTIMATE USING "I"
60
SUBTRACT CURRENT NOISE SPECTRUM ESTIMATE FROM CURRENT DATA
62
APPLY SPEECH ENHANCEMENT TO DATA (J)
64
INVERSE FFT & RECONSTRUCT ZERO PHASE REALIZATION
66
SAVE SECOND HALF OF BLOCK FOR NEXT ITERATION
70
ADD FIRST HALF OF RECONSTRUCTED DATA TO SECOND HALF OF RECONSTRUCTED DATA SAVED FROM PREVIOUS BLOCK
68
OUTPUT "A" SAMPLES OF PROCESSED DATA
72

FIG. 2

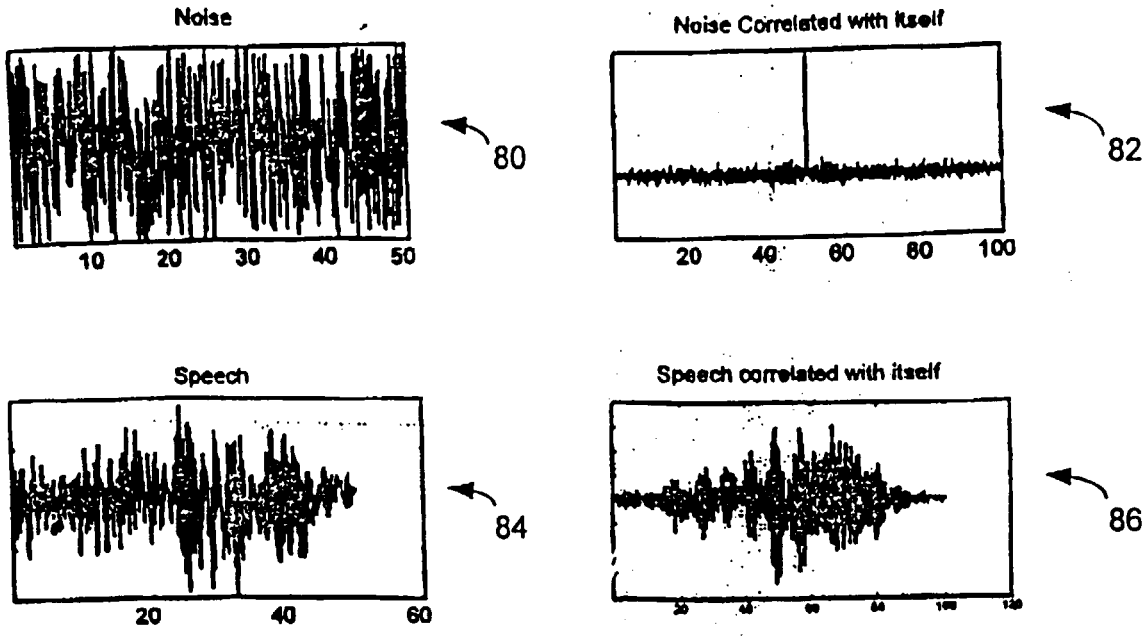
Noise
10
20
30
40
50
80
Noise Correlated with itself
20
40
60
80
100
82
Speech
20
40
60
84
Speech correlated with itself
86

FIG. 3

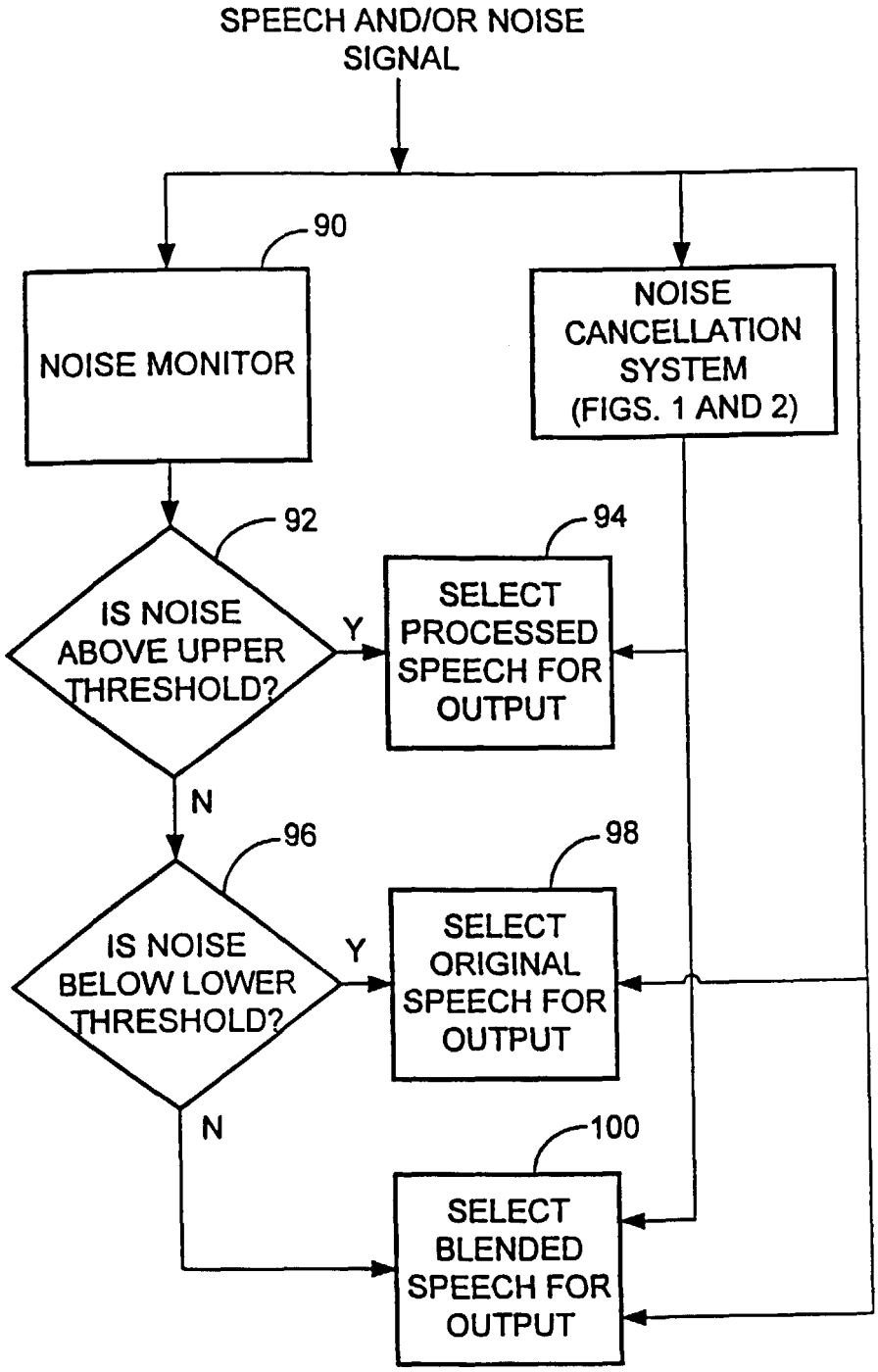
SPEECH AND/OR NOISE SIGNAL
90
NOISE MONITOR
NOISE CANCELLATION SYSTEM (FIGS. 1 AND 2)
92
IS NOISE ABOVE UPPER THRESHOLD?
Y
94
SELECT PROCESSED SPEECH FOR OUTPUT
N
96
IS NOISE BELOW LOWER THRESHOLD?
Y
98
SELECT ORIGINAL SPEECH FOR OUTPUT
N
100
SELECT BLENDED SPEECH FOR OUTPUT

FIG. 4

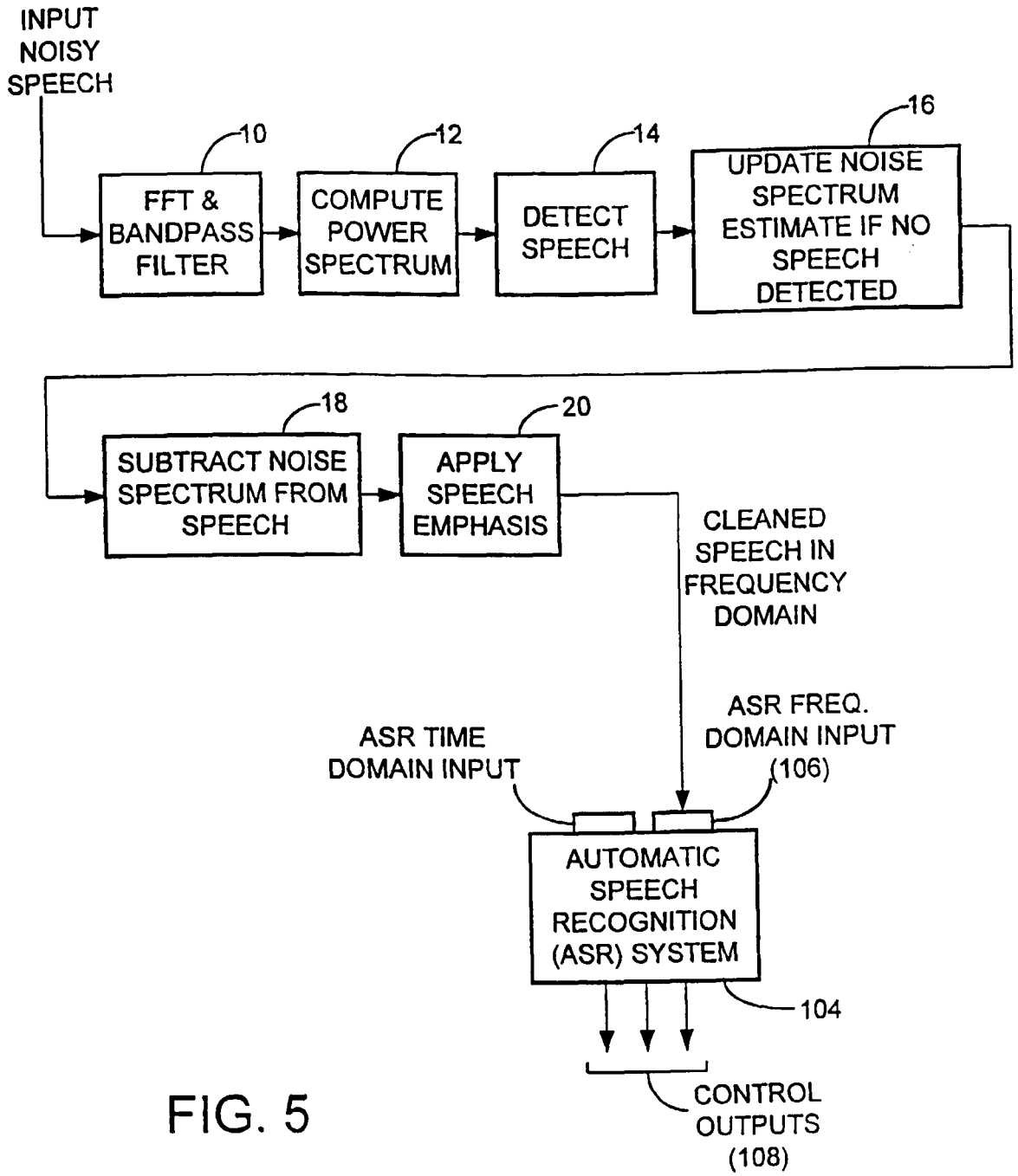
INPUT NOISY SPEECH
10
FFT & BANDPASS FILTER
12
COMPUTE POWER SPECTRUM
14
DETECT SPEECH
16
UPDATE NOISE SPECTRUM ESTIMATE IF NO SPEECH DETECTED
18
SUBTRACT NOISE SPECTRUM FROM SPEECH
20
APPLY SPEECH EMPHASIS
CLEANED SPEECH IN FREQUENCY DOMAIN
ASR TIME DOMAIN INPUT
ASR FREQ. DOMAIN INPUT (106)
AUTOMATIC SPEECH RECOGNITION (ASR) SYSTEM
104
CONTROL OUTPUTS (108)

FIG. 5