(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 June 2004 (24.06.2004)

PCT

(10) International Publication Number
WO 2004/053838 A2

(51) International Patent Classification[7]: G10L 21/02

(21) International Application Number:
PCT/US2003/038657

(22) International Filing Date: 5 December 2003 (05.12.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/315,615 10 December 2002 (10.12.2002) US

(71) Applicant: **LIBERATO TECHNOLOGIES, INC.** [US/US]; 6 Colton Court, Durham, NC 27713 (US).

(72) Inventors: **ZANGI, Kambiz, C.**; 6 Colton Court, Durham, NC 27713 (US). **ISABELLE, Steven**; 11 David Road, Newton, MA 02459 (US).

(74) Agents: **ROBINSON, Kermit** et al.; Daly, Crowley & Mofford, LLP, Suite 101, 275 Turnpike Street, Canton, MA 02021 (US).

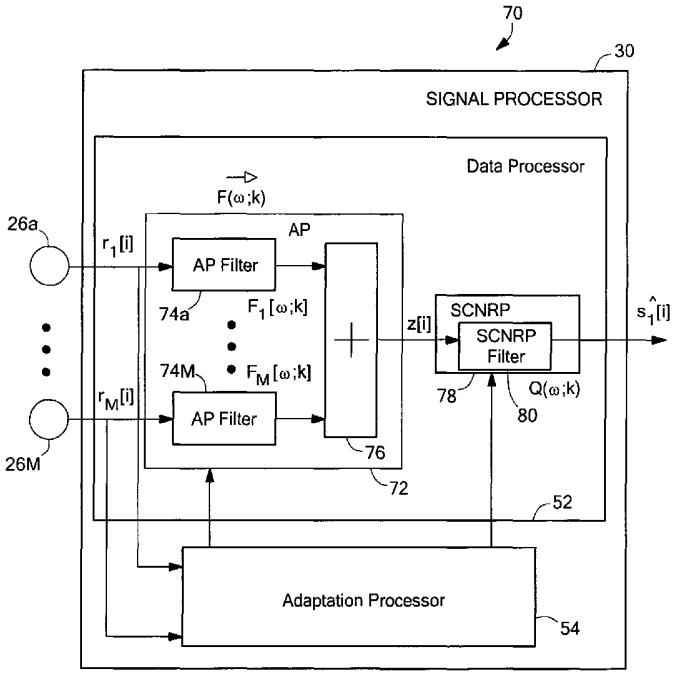
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND APPARATUS FOR NOISE REDUCTION



(57) Abstract: An apparatus and method for noise reduction is described. The method and apparatus can be used in a hands-free communication system to provide a hands-free a communication system having improved intelligibility. The apparatus includes a first and a second processor, each separately dynamically adapted to changing signals and noise, to improve a signal to noise ratio.

WO 2004/053838 A2

METHOD AND APPARATUS FOR NOISE REDUCTION

FIELD OF THE INVENTION

This invention relates generally to systems and methods for reducing noise in a communication, and more particularly to methods and systems for reducing the effect of acoustic noise in a hands-free telephone system.

BACKGROUND OF THE INVENTION

As is known in the art, a portable hand-held telephone can be arranged in an automobile or other vehicle so that a driver or other occupant of the vehicle can place and receive telephone calls from within the vehicle. Some portable telephone systems allow the driver of the automobile to have a telephone conversation without holding the portable telephone. Such systems are generally referred to as "hands-free" systems.

As is known, the hands-free system receives acoustic signals from various undesirable noise sources, which tend to degrade the intelligibility of a telephone call. The various noise sources can vary with time. For example, background wind, road, and mechanical noises in the interior of an automobile can change depending upon whether a window of an automobile is open or closed.

Furthermore, the various noise sources can be different in magnitude, spectral content, and direction for different types of automobiles, because different automobiles have different acoustic characteristics, including, but not limited to, different interior volumes, different surfaces, and different wind, road, and mechanical noise sources

It will be appreciated that an acoustic source such as a voice, for example, reflects around the interior of the automobile, becoming an acoustic source having multi-path acoustic propagation. In so reflecting, the direction from which the acoustic

source emanates can appear to change in direction from time to time and can even appear to come from more than one direction at the same time. A voice undergoing multi-path acoustic propagation is generally less intelligible than a voice having no multi-path acoustic propagation.

In order to reduce the effect of multi-path acoustic propagation as well as the effect of the various noise sources, some conventional hands-free systems are configured to place the speaker in proximity to the ear of the driver and the microphone in proximity to the mouth of the driver. These hands-free systems reduce the effect of the multi-path acoustic propagation and the effect of the various noise sources by reducing the distance of the driver's mouth to the microphone and the distance of the speaker to the driver's ear. Therefore, the signal to noise ratios and corresponding intelligibility of the telephone call are improved. However, such hands-free systems require the use of an apparatus worn on the head of the user.

Other hands-free systems place both the microphone and the speaker remotely from the driver, for example, on a dashboard of the automobile. This type of hands-free system has the advantage that it does not require an apparatus to be worn by the driver. However, such a hands-free system is fully susceptible to the effect of the multi-path acoustic propagation and also the effects of the various noise sources described above. This type of system, therefore, still has the problem of reduced intelligibility.

A plurality of microphones can be used in combination with some classical processing techniques to improve communication intelligibility in some applications. For example, the plurality of microphones can be coupled to a time-delay beam former arrangement that provides an acoustic receive beam pointing toward the driver.

However, it will be recognized that a time-delay beamformer provides desired acoustic receive beams only when associated with an acoustic source that generates planar sound waves. In general, only an acoustic source that is relatively far from the microphones generates acoustic energy that arrives at the microphones as a plane wave. Such is not the case for a hands-free system used in the interior of an automobile or in other relatively small areas.

Furthermore, multi-path acoustic propagation, such as that described above in the interior of an automobile, can provide acoustic energy arriving at the microphones from more than one direction. Therefore, in the presence of a multi-path acoustic propagation, there is no single pointing direction for the receive acoustic beam.

Also, the time-delay beamformer provides most signal to noise ratio improvement for noise that is incoherent between the microphones, for example, ambient noise in a room. In contrast, the dominant noise sources within an automobile are often directional and coherent.

Therefore, due to the non-planar sound waves that propagate in the interior of the automobile, the multi-path acoustic propagation, and also due to coherency of noise received by more than one microphone, the time-delay beamformer arrangement is not well suited to improve operation of a hands-free telephone system in an automobile. Other conventional techniques for processing the microphone signals have similar deficiencies.

It would, therefore, be desirable to provide a hands-free system configured for operation in a relatively small enclosure such as an automobile. It would be further desirable to provide a hands-free system that provides a high degree of intelligibility in the presence of the variety of noise sources in an automobile. It would be still further

desirable to provide a hands-free system that does not require the user to wear any portion of the system.

SUMMARY OF THE INVENTION

The present invention provides a noise reduction system having the ability to provide a communication having improved speech intelligibility.

In accordance with the present invention, the noise reduction system includes a first processor having one or more first processor filters configured to receive respective ones of one or more input signals from respective microphones. The first processor is configured to provide an intermediate output signal. The system also includes a second processor having a second processor filter configured to receive the intermediate output signal and provide a noise-reduced output signal. In operation, the one or more first processor filters are dynamically adapted and the second processor filter is separately dynamically adapted. In one particular embodiment, the first processor filters are adapted in accordance with a noise power spectrum at the microphones and the second processor filter is adapted in accordance with a power spectrum of the intermediate output signal.

Inherent in the above formulation is the assumption that the power spectrum of the noise and the power spectrum of the intermediate signal stay relatively constant, long enough so that good estimates of these power spectra can be obtained, and these estimates are then used to adapt the first processor filters and the second processor filter. The longer the period of time each of these power spectrum stays constant, the longer the longer the period of time over which it can be measured. Hence, the better the quality of the resulting estimate. Naturally, a higher quality estimate of the power

spectrum of the noise or a higher quality estimate of the power spectrum of the intermediate signal will lead to a better performance of the resulting noise reduction system. When the power spectrum of the noise changes at a significantly slower rate than the power spectrum of the intermediate signal, a slower time constant for estimating the power spectrum of the noise can be used, resulting in a more accurate estimate of the power spectrum of the noise. The more accurate estimate of the power spectrum of the noise can be used to adapt the first processor more accurately

With the above arrangement, because the noise power spectrum changes relatively slowly, the first processor filters can be adapted at a different rate than the second processor filter, therefore a more accurate estimate of the power spectrum of the noise can be obtained, and this more accurate estimate of the power spectrum of the noise leads to a more accurate adaptation of the first processor filters. The system provides a communication having a high degree of intelligibility. The system can be used to provide a hands-free system with which the user does not need to wear any part of the system.

In accordance with another aspect of the present invention, a method for processing one or more input signals includes receiving the one or more input signals with a first filter portion, the first filter portion providing an intermediate output signal. The method also includes receiving the intermediate output signal with a second filter portion, the second filter portion providing an output signal. The method also includes dynamically adapting a response of the first filter portion and a response of the second filter portion.

With this particular arrangement, the method provides a system that can dynamically adapt to varying signals and varying noises in a small enclosure, for

example in the interior of an automobile.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing features of the invention, as well as the invention itself may be more fully understood from the following detailed description of the drawings, in which:

FIG. 1 is a block diagram of an exemplary hands-free system in accordance with the present invention;

FIG. 2 is a block diagram of a portion of the hands-free system of FIG. 1, including an exemplary signal processor;

FIG. 3 is a block diagram showing greater detail of the exemplary signal processor of FIG. 2;

FIG. 4 is a block diagram showing greater detail of the exemplary signal processor of FIG. 3;

FIG. 5 is a block diagram showing greater detail of the exemplary signal processor of FIG. 4;

FIG. 6 is a block diagram showing an alternate embodiment of the exemplary signal processor of FIG. 5;

FIG. 7 is a block diagram of an exemplary echo canceling processor arrangement, which may be used in the exemplary signal processor of FIGS. 1-6;

FIG. 8 is a block diagram of an alternate echo canceling processor arrangement, which may be used in the exemplary signal processor of FIGS. 1-6;

FIG. 9 is a block diagram of yet another alternate echo canceling processor arrangement, which may be used in the exemplary signal processor of FIGS. 1-6;

FIG. 10 is a block diagram of a circuit for converting a signal from the time domain to the frequency domain which may be used in the exemplary signal processor of FIGS. 1-6; and

FIG. 11 is a block diagram of an alternate circuit for converting a signal from the time domain to the frequency domain, which may be used in the exemplary signal processor of FIGS. 1-6.

DETAILED DESCRIPTION OF THE INVENTION

Before describing the noise reduction system in accordance with the present invention, some introductory concepts and terminology are explained.

As used herein, the notation $x_m[i]$ indicates a scalar-valued sample "i" of a particular channel "m" of a time-domain signal "x". Similarly, the notation x[i] indicates a scalar-valued sample "i" of one channel of the time-domain signal "x". It is assumed that the signal x is band limited and sampled at a rate higher than the Nyquist rate. No distinction is made herein as to whether the sample $x_m[i]$ is an analog sample or a digital sample, as both are functionally equivalent.

As used herein, a Fourier transform, X(ω), of x[i] at frequency ω (where $0 \leq \omega \leq 2\pi$) is described by the equation:

$$X(\omega) = \sum_i x[i] e^{-j\omega i}$$

As used herein, an autocorrelation, $\rho_{xx}[t]$, of x[i] at lag t, is described by the

equation:

$$\rho_{xx}t] = E\{x\,[i]\,x^*\,[i+t]\},$$

where superscript "*" indicates a complex conjugate, and E{ } denotes expected value.

As used herein, a power spectrum, $P_{xx}(\omega)$, of x[i] at frequency ω (where $0 \leq \omega \leq 2\pi$) is described by the equation:

$$P_{xx}(\omega) = \sum_i \rho_{xx}[i]e^{-j\omega i}$$

A generic vector-valued time-domain signal, $\vec{x}[i]$, having M scalar-valued elements is denoted herein by:

$$\vec{x}[i] = [x_1[i] \ldots x_M[i]]^T$$

where the superscript T denotes a transpose of the vector. Therefore the vector $\vec{x}[i]$ is a column vector.

The Fourier Transform of $\vec{x}[i]$ at frequency ω (where $0 \leq \omega \leq 2\pi$) is an M x 1 vector $\vec{X}(\omega)$ whose m-th entry is the Fourier Transform of $x_m[i]$ at frequency ω.

The auto-correlation of $\vec{x}[i]$ at lag t is denoted herein by the M x M matrix $\rho_{\vec{x}\vec{x}}[t]$ defined as:

$$\rho_{\vec{x}\vec{x}}[t] = E\{\vec{x}[i]\vec{x}^H[i+t]\}$$

where the superscript H represents an Hermetian.

The power spectrum of the vector-valued signal $\vec{x}[i]$ at frequency ω (where $0 \leq \omega \leq 2\pi$) is denoted herein by $P_{\vec{x}\vec{x}}(\omega)$. The power spectrum $P_{\vec{x}\vec{x}}(\omega)$ is an M x M matrix whose (i, j) entry is the Fourier Transform of the (i, j) entry of the autocorrelation function $\rho_{\vec{x}\vec{x}}[m]$ at frequency ω.

Referring now to FIG. 1, an exemplary hands-free system 10 in accordance with the present invention includes one or more microphones 26a-26m coupled to a signal

processor 30. The signal processor 30 is coupled to a transmitter/receiver 32, which is coupled to an antenna 34. The one or more microphones 26a-26M are inside of an enclosure 28, which, in one particular arrangement, can be the interior of an automobile. The one or more microphones 26a-26M are configured to receive a local voice signal 14 generated by a person or other signal source 12 within the enclosure 28. The local voice signal 14 propagates to each of the one or more microphones 26a-26M as one or more "desired signals" $s_1[i]$ to $s_m[M]$, each arriving at a respective microphone 26a-26M on respective paths 15a-15M from the person 12 to the one or more microphones 26a-26M. The paths 15a-15M can have the same length or different lengths depending upon the position of the person 12 relative to each of the one or more microphones 26a-26M.

A loudspeaker 20, also within the enclosure 28, is coupled to the transmitter/receiver 32 for providing a remote voice signal 22 corresponding to a voice of a remote person (not shown) at any distance from the hands-free system 10. The remote person is in communication with the hands-free system by way of radio frequency signals (not shown) received by the antenna 34. For example, the communication can be a cellular telephone call provided over a cellular network (not shown) to the hands-free system 10. The remote voice signal 22 corresponds to a remote-voice-producing signal q[i] provided to the loudspeaker 20 by the transmitter/receiver 32.

The remote voice signal 22 propagates to the one or more microphones 26a-26M as one or more "remote voice signals" $e_1[i]$ to $e_M[i]$, each arriving at a respective microphone 26a-26M upon a respective path 23a-23M from the loudspeaker 20 to the one or more microphones 26a-26M. The paths 23a-23M can have the same length or

different lengths depending upon the position of the loudspeaker 20 relative to the one or more microphones 26a-26M.

One or more environmental noise sources generally denoted 16, which are undesirable, generate one or more environmental acoustic noise signals generally denoted 18, within the enclosure 28. The environmental acoustic noise signals 18 propagate to the one or more microphones 26a-26M as one or more "environmental signals" $v_1[i]$ to $v_M[i]$, each arriving at a respective microphone 26a-26M upon a respective path 19a-19M from the environmental noise sources 16 to the one or more microphones 26a-26M. The paths 19a-19M can have the same length or different lengths depending upon the position of the environmental noise sources 16 relative to the one or more microphones 26a-26M. Since there can be more than one environmental noise source 16, the environmental noise signals $v_1[i]$ to $v_M[i]$ from each such other noise source 16 can arrive at the microphones 26a-26M on different paths. The other noise sources 16 are shown to be collocated for clarity in FIG. 1, however, those of ordinary skill in the art will appreciate that in practice this typically will not be true.

Together, the remote voice signal 22 and the environmental acoustic noise signal 18 comprise noise sources 24 that interfere with reception of the local voice signal 14 by the one or more microphones 26a-26M.

It will be appreciated that the environmental noise signal 18, the remote voice signal 22, and the local voice signal 14 can each vary independently of each other. For example, the local voice signal 14 can vary in a variety of ways, including but not limited to, a volume change when the person 12 starts and stops talking, a volume and phase change when the person 12 moves, and a volume, phase, and spectral content

change when the person 12 is replaced by another person having a voice with different acoustic characteristics. For another example, the remote voice signal 22 can vary in the same way as the local voice signal 14. For another example, the environmental noise signal 18 can vary as the environmental noise sources 16 move, start, and stop.

Not only can the local voice signal 14 vary, but also the desired signals 15a-15M can vary irrespective of variations in the local voice signal 14. In this regard, taking the microphone 26a as representative of all microphones 26a-26M, it should be appreciated that, while the microphone 26a receives the desired signal $s_1[i]$ corresponding to the local voice signal 14 on the path 15a, the microphone 26a also receives the local voice signal 14 on other paths (not shown). The other paths correspond to reflections of the local voice signal 14 from the inner surface 28a of the enclosure 28. Therefore, while the local voice signal 14 is shown to propagate from the person 12 to the microphone 26a on a single path 15a, the local voice signal 14 can also propagate from the person 12 to the microphone 26a on one or more other paths or reflection paths (not shown). The propagation, therefore, can be a multi-path propagation. In FIG. 1, only the direct propagation paths 15a-15M are shown.

Similarly, the propagation paths 19a-19M and the propagation paths 23a-23M represent only direct propagation paths and the environmental noise signal 18 and the remote signal 22 both experience multi-path propagation in traversing from the environmental noise sources 16 and the loudspeaker 20 respectively, to the one or more microphones 26a-26M. Therefore, each of the local voice signal 14, the environmental noise signal 18, and the remote voice signal 22 arriving at the one or more microphones 26a-26M through multi-path propagation, are affected by the reflective characteristics and the shape, i.e., the acoustic characteristics, of the interior 28a of the enclosure 28.

In one particular embodiment, where the enclosure 28 is an interior of an automobile or other vehicle, not only can the acoustic characteristics of the interior of the automobile vary from automobile to automobile, but they can also vary depending upon the contents of the automobile, and in particular they can also vary depending upon whether one or more windows are up or down.

The multi-path propagation has a more dominant effect on the acoustic signals received by the microphones 26a-26M when the enclosure 28 is small and when the interior of the enclosure 28 is acoustically reflective. Therefore, a small enclosure corresponding to the interior of an automobile having glass windows, known to be acoustically reflective, is expected to have substantial multi-path acoustic propagation.

As shown below, equations can be used to describe aspects of the hands-free system of FIG. 1.

In accordance with the general notation $x_m[i]$ described above, the notation $s_1[i]$ corresponds to one sample of the local voice signal 14 traveling along the path 15a, the notation $e_1[i]$ corresponds to one sample of the echo signal 18 traveling along the path 23a, and the notation $v_1[i]$ corresponds to one sample of the environmental noise signal 18 traveling along the path 23a.

The i^{th} sample of the output of the m-th microphone is denoted $r_m[i]$. The i^{th} sample of the output of the m-th microphone may be computed as:

$$r_m[i] = s_m[i] + n_m[i], \qquad m = 1, \ldots, M$$

In the above equation, $s_m[i]$ corresponds to the local voice signal 14, and $n_m[i]$ corresponds to a combined noise signal described below.

The sampled signal $s_m[i]$ corresponds to a "desired signal portion" received by the m-th microphone. The signal $s_m[i]$ has an equivalent representation $s_m[i]$ at the

output of the m-th microphone within the signal $r_m[i]$. Therefore, it will be understood that the local voice signal 14 corresponds to each of the signals $s_1[i]$ to $s_M[i]$, which signals have corresponding desired signal portions $s_1[i]$ to $s_M[i]$ at the output of respective microphones.

Similarly, $n_m[i]$ corresponds to a "noise signal portion" received by the m-th microphone (from the loudspeaker 20 and the environmental noise sources 16) as represented at the output of the m-th microphone within the signal $r_m[i]$. Therefore, the output of the m-th microphone comprises desired contributions from the local voice signal 12, and undesired contributions from the noise 16, 20.

As described above, the noise $n_m[i]$ at the output of the m-th microphone has contributions from both the environmental noise signal 18 and the remote voice signal 22 and can, therefore, be described by the following equation:

$$n_m[i] = v_m[i] + e_m[i], \qquad m = 1, \ldots, M$$

In the above equation, $v_m[i]$ is the environmental noise signal 18 received by the m-th microphone, and $e_m[i]$ is the remote voice signal 22 received by the m-th microphone.

Both $v_m[i]$ and $e_m[i]$ have equivalent representations $v_m[i]$ and $e_m[i]$ at the output of the m-th microphone. Therefore, it will be understood that the remote voice signal 22 and the environmental noise signal 18 correspond to the signals $e_1[i]$ to $e_M[i]$ and $v_1[i]$ to $v_M[i]$ respectively, which signals both contribute to corresponding "noise signal portions" $n_1[i]$ to $n_M[i]$ at the output of respective microphones.

In operation, the signal processor 30 receives the microphone output signals $r_m[i]$ from the one or more microphones 26a-26M and estimates the local voice signal 14 therefrom by estimating the desired signal portion $s_m[i]$ of one of the signals $r_m[i]$ provided at the output of one of the microphones. In one particular embodiment, the

signal processor 30 receives the microphone output signals $r_m[i]$ and estimates the local voice signal 14 therefrom by estimating the desired signal portion $s_1[i]$ of the signal $r_1[i]$ provided at the output of the microphone 26a. However, it will be understood that the desired signal portion from any microphone can be used.

The hands-free system 10 has no direct access to the local voice signal 14, or to the desired signal portions $s_m[i]$ within the signals $r_m[i]$ to which the local voice signal 14 corresponds. Instead, the desired signal portions $s_m[i]$ only occur in combination with noise signals $n_m[i]$ within each of the signals $r_m[i]$ provided by each of the one or more microphones 26a-26M.

Each desired signal portion $s_m[i]$ provided by each microphone 26a-26M is related to the desired signal portion $s_1[i]$ provided by the first microphone through a linear convolution:

$$s_m[i] = s_1[i] * g_m[i], \qquad i = 1, \ldots, M$$

where the $g_m[i]$ are the transfer functions relating $s_1[i]$ provided by the first microphone 26a to $s_m[i]$ provided by the other microphones 26m. These transfer function are not necessarily causal. In one particular embodiment, the transfer functions $g_m[i]$ can be modeled as a simple time delays or time advances; however, these transfer functions can be any transfer function.

Similarly, each remote voice signal $e_m[i]$ provided by each microphone 26a-26M as part of the signals $r_m[i]$ is related to the remote voice-producing signal q[i] through a linear convolution:

$$e_m[i] = q[i] * k_m[i], \qquad m = 1, \ldots, M$$

In the above equation, $k_m[i]$ are the transfer functions relating q[i] to $e_m[i]$. The transfer functions $k_m[i]$ are strictly causal.

The above relationships have equivalent representations in the frequency domain. Lower case letters are used in the above equations to represent time domain signals. In contrast, upper case letters are used in the equations below to represent the same signals, but in the frequency domain. Furthermore, vector notations are used to represent the values among the one or more microphones 26a-26M. Therefore, similar to the above time-domain representations given above, in the frequency-domain:

$$\vec{R}(\omega) = \vec{S}(\omega) + \vec{N}(\omega)$$

$$= \vec{G}(\omega)S_1(\omega) + \vec{N}(\omega),$$

In the above equation, $\vec{R}(\omega)$ is a frequency-domain representation of a group of the time-sampled microphone output signals $r_m[i]$, $\vec{S}(\omega)$ is a frequency-domain representation of a group of the time-sampled desired signal portion signals $s_m[i]$, $\vec{N}(\omega)$ is a frequency-domain representation of a group of the time-sampled noise portion signals $n_m[i]$, $\vec{G}(\omega)$ is a frequency-domain representation of a group of the transfer functions $g_m[i]$, and $S_1(\omega)$ is a frequency-domain representation of a group of the time-sampled desired signal portion signals $s_1[i]$ provided by the first microphone 26a.

$\vec{G}(\omega)$ is a matrix of size M x 1 and $S_1(\omega)$ a scalar value is of size 1 x 1.

Similarly, in the frequency domain:

$$\vec{N}(\omega) = K(\omega)Q(\omega),$$

In the above equation, $\vec{N}(\omega)$) is a frequency-domain representation of a group of the time-sampled signals $n_m[i]$, $\vec{K}(\omega)$ is a frequency-domain representation of a group of the transfer functions $k_m[i]$, and $Q(\omega)$ is a frequency-domain representation of a group of the time-sampled signals q[i].

$\vec{K}(\omega)$ is a vector of size M x 1, and Q(ω) is a scalar value of size 1 x 1.

A mean-square error is a particular measurement that can be evaluated to characterize the performance of the hands-free system 10. The means square error can be represented as:

$$\mu[i] = s_1(i) - \hat{s}_1[i],$$

In the above equation. $\hat{s}_1[i]$ is an "estimate signal" corresponding to an estimate of the desired signal portion $s_1[i]$ of the signal $r_1[i]$ provided by the first microphone 26a. As described above, an estimate of any of the desired signal portions $s_m[i]$ could be used equivalently. In one particular embodiment, the estimate signal $\hat{s}_1[i]$ is the desired output of the hands-free system 10, providing a high quality, noise reduced signal to a remote person.

In one embodiment the signal processor 30 provides processing that comprises minimizing the variance of μ[i], where the variance of μ[i] can be expressed as:

$$\text{Var } \mu[i] = E\{|\mu[i]|^2\}.$$

or equivalently:

$$\text{Var } \{ s_1[i] - \hat{s}_1[i]\} = E\{| s_1[i] - \hat{s}_1[i] |^2\}$$

The above equations are used in conjunction with figures below to more fully describe the processing provided by the signal processor 30.

Referring now to FIG. 2, a portion 50 of an the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes the one or more microphones 26a-26M coupled to the signal processor 30. The signal processor 30 includes a data processor 52 and an adaptation processor 54 coupled to the data processor. The microphones 26a-26M provide the signals $r_m[i]$ to the data processor 52 and to the adaptation processor 54.

In operation, the data processor 52 receives the signal $r_m[i]$ from the one or more microphones 26a-26M and, by processing described more fully below, provides an estimate signal $\hat{s}_m[i]$ of a desired signal portion $s_m[i]$ corresponding to one of the microphones 26a-26M, for example an estimate signal $\hat{s}_1[m]$ of the desired signal portion $s_1[i]$ of the signal $r_1[i]$ provided by the microphone 26a. It will be recognized that the desired signal portion $s_1[i]$, corresponds to the local voice signal 14 (FIG. 1) and in particular to the local voice signal $s_1[i]$ (FIG. 1) provided by the person 12 (FIG. 1) along the path 15a (FIG. 1). However, in other embodiments, the desired signal portion $s_m[i]$ provided by any of the one or more microphones 26a-26M can be used equivalently in place of $s_1[i]$ above, and therefore, the estimate becomes $\hat{s}_m[i]$.

While in operation, the adaptation processor 54 dynamically adapts the processing provided by the data processor 52 by adjusting the response of the data processor 52. The adaptation is described in more detail below. The adaptation processor 54 thus dynamically adapts the processing performed by the data processor 52 to allow the data processor to provide an audio output as an estimate signal $\hat{s}_1[i]$ having a relatively high quality, and a relatively high signal to noise ratio in the presence of the varying local voice signal 14 (FIG. 1), the varying remote voice signal 22 (FIG. 1), and the varying environmental noise signal 18 (FIG. 1). The variation of these signals is described above in conjunction with FIG. 1.

Referring now to FIG. 3, a portion 70 of the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes the one or more microphones 26a-26M coupled to the signal processor 30. The signal processor 30 includes the data processor 52 and the adaptation processor 54 coupled to the data processor 52. The microphones 26a-26M provide the signals $r_m[i]$

to the data processor 52 and to the adaptation processor 54.

The data processor 52 includes an array processor (AP) 72 coupled to a single channel noise reduction processor (SCNRP) 78. The AP 72 includes one or more AP filters 74a-74M, each coupled to a respective one of the one or more microphones 26a-26M. The outputs of the one or more AP filters 74a-74M are coupled to a combiner circuit 76. In one particular embodiment, the combiner circuit 72 performs a simple sum of the outputs of the one or more AP filters 74a-74M. In total, the AP 72 has one or more inputs and a single scalar-valued output comprising a time series of values.

The SCNRP 78 includes a single input, single output SCNRP filter. The input to the SCNRP filter 80 is an intermediate signal z[i] provided by the AP 72. The output of the SCNRP filter provides the estimate signal $\hat{s}_1[i]$ of the desired signal portion $s_1[i]$ of z[i] corresponding to the first microphone 26a. The estimate signal $\hat{s}_1[i]$, and alternate embodiments thereof, is described above in conjunction with FIG. 2.

In operation, the adaptation processor 54 dynamically adapts the response of each of the AP filters 74a-74M and the response of the SCNRP filter 80. The adaptation is described in greater detail below.

Referring now to FIG. 4, a portion 90 of an the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes the one or more microphones 26a-26M coupled to the signal processor 30. The signal processor 30 includes the data processor 52 and the adaptation processor 54 coupled to the data processor 52. The microphones 26a-26M provide the signals $r_m[i]$ to the data processor 52 and to the adaptation processor 54.

The data processor 52 includes the array processor (AP) 72 coupled to the single channel noise reduction processor (SCNRP) 78. The AP 72 includes the one or

more AP filters 74a-74M. The outputs of the one or more AP filters 74a-74M are coupled to the combiner circuit 76.

The adaptation processor 54 includes a first adaptation processor 92 coupled to the AP 72, and to each AP filter 74a-74M therein. The first adaptation processor 92 provides a dynamic adaptation of the one or more AP filters 74a-74M. However, it will be understood that the adaptation provided by the first adaptation processor 92 to any one of the one or more AP filters 74a-74M can be the same as or different from the adaptation provided to any other of the one or more AP filters 74a-74M.

The adaptation processor 54 also includes a second adaptation processor 94 coupled to the SCNRP 78 and to the SCNRP filter 80 therein. The second adaptation processor 94 provides an adaptation of the SCNRP filter 80.

In operation, the first adaptation processor 92 dynamically adapts the response of each of the AP filters 74a-74M in response to noise signals. The second adaptation processor 94 dynamically adapts the response of the SCNRP filter 80 in response to a combination of desired signals and noise signals. Because the signal processor 30 has both a first and a second adaptation processor 92, 94 respectively, each of the two adaptations can be different, for example, they can have different time constants. The adaptation is described in greater detail below.

Referring now to FIG. 5, a circuit portion 90 of an the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes the one or more microphones 26a-26M coupled to the signal processor 30. The signal processor 30 includes the data processor 52 and the adaptation processor 54 coupled to the data processor. The microphones 26a-26M provide the signals $r_m[i]$ to the data processor 52 and to the adaptation processor 54.

The variable 'k' in the notation below is used to denote that the various power spectra are computed upon a k-th frame of data. At a subsequent computation, the various power spectra are computed on a k+1-th frame of data, which may or may not overlap the k-th frame of data. The variable 'k' is omitted from some of the following equations. However, it will be understood that the various power spectra described below are computed upon a particular data frame 'k'.

Notation given above describes the power spectrum notation $P_{\bar{x}\bar{x}}(\omega)$ as an M x M matrix whose (i, j) entry is the Fourier Transform of the (i, j) entry of the autocorrelation function $\rho_{\bar{x}\bar{x}}[t]$ at frequency ω. The adaptation processor 54 can be described with similar notations.

The adaptation processor 54 includes the first adaptation processor 92 coupled to the AP 72, and to each AP filter 74a-74M therein. The first adaptation processor 92 includes a voice activity detector (VAD) 102. The VAD is coupled to an update processor 104 that computes a noise power spectrum $P_{\bar{n}\bar{n}}(\omega;k)$. The update processor 104 is coupled to an update processor 106 that receives the power spectrum and computes a noise power spectrum $P_{tt}(\omega;k)$ therefrom. The power spectrum $P_{tt}(\omega;k)$ is a power spectrum of the noise portion of the intermediate signal z[i]. In combination, the two update processors 104, 106 provide the noise power spectrums $P_{\bar{n}\bar{n}}(\omega;k)$ and $P_{tt}(\omega;k)$ in order to update the AP filters 74a-74M. The update of the AP filters 74a-74M is described in more detail below.

The adaptation processor 54 also includes the second adaptation processor 94 coupled to the SCNRP 78 and to the SCNRP filter 80 therein. The second adaptation processor 94 includes an update processor 106 that computes a power spectrum $P_{zz}(\omega;k)$. The power spectrum $P_{zz}(\omega;k)$ is a power spectrum of the entire

intermediate signal z[i]. The update processor 106 provides the power spectrum $P_{zz}(\omega;k)$ in order to update the SCNRP filter 80. The update of the SCNRP filter 80 is described in more detail below.

The one or more channels of time-domain input samples $r_1[i]$ to $r_M[i]$ provided to the AP 72 by the microphones 26a-26M can be considered equivalently to be a frequency domain vector-valued input signal $\vec{R}(\omega)$. Similarly, the single channel time domain output samples z[i] provided by the AP 72 can be considered equivalently to be a frequency domain scalar-valued output $Z(\omega)$. The AP 72 comprises an M-input, single-output linear filter having a response $\vec{F}(\omega)$ expressed in the frequency domain, where each element thereof corresponds to a response $F_m(\omega)$ of one of the AP filters 74a-74M. Therefore the output signal $Z(\omega)$ can be described by the following equation:

$$Z(\omega) = \sum_{m=1}^{M} F_m(\omega) R_m(\omega)$$

$$= \vec{F}^T(\omega)\vec{R}(\omega),$$

where

$$\vec{F}(\omega) = [F_1(\omega)\ F_2(\omega)\ \ldots\ F_M(\omega)]^T, \text{ and}$$

$$\vec{R}(\omega) = [R_1(\omega)\ R_2(\omega)\ \ldots\ R_M(\omega)]^T$$

As described above, the superscript T refers to the transpose of a vector, therefore $\vec{F}(\omega)$ and $\vec{R}(\omega)$ are column vectors having vector elements corresponding to each microphone 26a-26M. The asterisk symbol * corresponds to a complex conjugate.

In operation of the signal processor 54, the VAD 102 detects the presence or absence of a desired signal portion of the intermediate signal z[i]. The desired signal portion can be $s_1[i]$, corresponding to the voice signal provided by the first microphone 26a. One of ordinary skill in the art will understand that the VAD 102 can be

constructed in a variety of ways to detect the presence or absence of a desired signal portion. While the VAD is shown to be coupled to the intermediate signal z[i], in other embodiments, the VAD can be coupled to one or more of the microphone signals r_1 [i] to r_m[i], or to the output estimate signal $\hat{s}_1$[i].

In operation of the first adaptation processor 92, the response of the filters 74a-74m, $\vec{F}(\omega)$, is determined so that the output Z(ω) of the AP 72 is the maximum likelihood (ML) estimate of $S_1(\omega)$, where $S_1(\omega)$ is a frequency domain representation of the desired signal portion s_1[i] of the input signal r_1[i] provided by the first microphone 26a as described above. Therefore, it can be shown that the responses of the AP filters 74 can be described by vector elements in the equation:

$$\vec{F}^{T}(\omega) = \frac{1}{\vec{G}^{H}(\omega)P_{\vec{n}\vec{n}}^{-1}(\omega)\vec{G}(\omega)} \vec{G}^{H}(\omega)P_{\vec{n}\vec{n}}^{-1}(\omega)$$

In the above equation, $\vec{G}(\omega)$ is the frequency domain vector notation for the transfer function g_m[i] between the microphones as described above, $P_{\vec{n}\vec{n}}(\omega)$ corresponds to the power spectrum of the noise. The transfer function $\vec{F}(\omega)$ provides a maximum likelihood estimate of $S_1(\omega)$ based upon an input of $\vec{R}(\omega)$.

It will be understood that the m-th element of the vector $\vec{F}(\omega)$ is the transfer function of the m-th AP filter 74m. With the above vector transfer function, $\vec{F}(\omega)$, the sum, Z(ω), of the outputs of the AP filters 74a-74M includes the desired signal portion $S_1(\omega)$ associated with the first microphone, plus noise. Therefore, the desired signal portion $S_1(\omega)$ passes through the AP filters 74a-74M without distortion.

From the above equation, it can be seen that the response of the AP 72, $\vec{F}(\omega)$, does not depend on the power spectrum $P_{s_1s_1}(\omega)$ of the desired signal portion s_1[i]. Instead, it is only dependant upon $P_{\vec{n}\vec{n}}(\omega)$, the power spectrum of the noise signal

portions $n_m[i]$. This is as expected, since the AP filters are adapted in response to power spectra computed during times when the VAD 102 indicates the absence of the local voice signal (14, FIG. 1).

The desired signal portion $s_1[i]$ of the input signal $r_1[i]$, corresponding to the local voice signal 14 (FIG. 1), can vary rapidly with time. As seen from the above equation, the response of the AP 72, $\vec{F}(\omega)$, only depends upon the power spectrum $P_{\bar{n}\bar{n}}(\omega)$ of the noise signal portions $n_m[i]$ of the input signal $r_1[i]$, and also on the frequency domain vector $\vec{G}(\omega)$, corresponding to the time domain transfer functions $g_m[i]$ between the microphones described above. Therefore the transfer functions within the vector $\vec{F}(\omega)$ are adapted based only in proportion to the noise, irrespective of a local voice signal 14 (FIG. 1).

The transfer functions $\vec{F}(\omega)$, therefore, can be updated, i.e. have time constants, that vary more slowly than the desired signal portions corresponding to the local voice signal 14 (FIG. 1). As mentioned above, using a slower time constant for adaptation of the AP filters results in a more accurate adaptation of the AP filters. The AP filters are adapted based on estimates of the power spectrum of the noise, and using a slower time constant to estimate the power spectrum of the noise results in a more accurate estimate of the power spectrum of the noise; since, with a slower time constant, a longer measurement window can be used for estimating.

In order to compute the power spectrum $P_{\bar{n}\bar{n}}(\omega)$, and the inverse thereof, the VAD 102 provides to the update processor 104 an indication of when the local voice signal 14 (FIG. 1) is absent, i.e. when the person 12 (FIG. 1) is not talking. Therefore, the update processor 104 computes the power spectrum $P_{\bar{n}\bar{n}}(\omega)$ of the noise signal portions $n_m[i]$ of the input signal $r_m[i]$ during a time, and from time to time, when only

the noise signal portions $n_m[i]$ are present. When the person 12 (FIG. 1) is silent, $\vec{r}$ [i] = $\vec{n}$ [i] (since $\vec{s}$ [i] = 0), and on those frames of data, $\vec{r}$ [i] is used to update the inverse power-spectrum of the noise $P^{-1}_{\vec{n}\vec{n}}$ (ω; k), and therefore, to compute the transfer functions of the AP filters 74a-74M. Therefore, the responses of the AP filters 74a-74M, corresponding to the elements of the vector $\vec{F}$ (ω), are computed at a time when no desired signal portions $s_m[i]$ are present.

As seen in the above equations, the transfer function $\vec{F}$ (ω) contains terms for the inverse of the power spectrum of the noise. It will be recognized by one of ordinary skill in art that there are a variety of mathematical methods to directly calculate the inverse of a power spectrum, without actually performing a mathematical vector inverse operation may be used. One such method uses a recursive least squares (RLS) algorithm to directly compute the inverse of the power spectrum, resulting in improved processing time. However, other methods can also be used to provide the inverse of the power spectrum $P^{-1}_{\vec{n}\vec{n}}$ (ω).

The frequency domain representation Z(ω) of the scalar-valued intermediate output signal z[i] can be expressed as sum of two terms: a term $S_1(\omega)$ due to the desired signal $s_1[i]$ provided by the first microphone 26a, and a term T(ω) due to the noise t[i] provided by the one or more microphones 26a-26M. Therefore, it can be shown that:

$$Z(\omega) = S_1(\omega) + T(\omega)$$

where T(ω) has the following power spectrum:

$$P_{tt}(\omega) = \frac{1}{\vec{G}^H(\omega)P^{-1}_{\vec{n}\vec{n}}(\omega)\vec{G}(\omega)}$$

The scalar-valued Z(ω) is further processed by the SCNRP filter 80. The SCNRP filter 80 comprises a single-input, single-output linear filter with response:

$$Q(\omega) = \frac{P_{s1s1}(\omega)}{P_{zz}(\omega)}$$

Furthermore,

$$P_{zz}(\omega) = P_{s1s1}(\omega) - P_{tt}(\omega) \quad \text{or equivalently,}$$

$$P_{s1s1}(\omega) = P_{zz}(\omega) - P_{tt}(\omega)$$

In the above equations, $P_{s1s1}(\omega)$ is the power spectrum of the desired signal portion of the first microphone signal $r_1[i]$ within the intermediate output signal z[i], $P_{zz}(\omega)$ is the power spectrum of the intermediate output signal z[i], and $P_{tt}(\omega)$ is the power spectrum of the noise signal portion of the intermediate output signal z[i]. Therefore, $Q(\omega)$ can be equivalently expressed as:

$$Q(\omega) = 1 - \frac{P_{tt}(\omega)}{P_{zz}(\omega)}$$

Therefore, the transfer function $Q(\omega)$ of the SCNRP filter 80 can be expressed as a function of $P_{s1s1}(\omega)$ and $P_{zz}(\omega)$ or equivalently as a function of $P_{tt}(\omega)$ and $P_{zz}(\omega)$.

Therefore, the second adaptation processor 94, in the embodiment shown, receives the signal z[i], or equivalently the frequency domain signal $Z(\omega)$, and the update processor 108 computes the power spectrum $P_{zz}(\omega)$ corresponding thereto. The update processor 108 is also provided with the power spectrum $P_{tt}(\omega)$ computed by the update processor 106. Therefore, the second adaptation processor 94 can provide the SCNRP filter 80 with sufficient information to generate the desired transfer function $Q(\omega)$ described by the above equations.

While the second update processor updates the SCNRP filter 80 based upon $P_{tt}(\omega)$ and $P_{zz}(\omega)$, in another embodiment, an alternate second update processor updates the SCNRP filter 80 based upon $P_{s1s1}(\omega)$ and $P_{zz}(\omega)$. The above equations show these two alternatives to be equivalent.

In one particular embodiment, the SCNRP filter 80 is essentially a single-input single-output Weiner filter. The cascaded system of FIG. 5, consisting of the AP 72 followed by the SCNRP 78, is mathematically equivalent to an M-input/1-output Wiener filter for estimating $S_1(\omega)$ based on $\vec{R}(\omega)$, where the transfer function of the Wiener filter is described by the equation:

$$\vec{H}(w) = \vec{F}(\omega) \times Q(\omega).$$

Referring again to the above equation for $\vec{F}(\omega)$, that describes the transfer function of the AP filters 74a-74M, the hands-free system can also adapt the transfer function $\vec{G}(\omega)$ in addition to the dynamic adaptations to the AP filters 74 and the SCNRP filter 80. It is discussed above that $g_m[i]$ is the transfer function between the desired signal $s_1[i]$ and the other desired signals $s_m[i]$:

$$s_m[i] = g_m[i] * s_1[i]$$

or equivalently

$$S_m(\omega) = G_m(\omega) S_1(\omega)$$

Given samples of the desired signal portions $s_m[i]$, a variety of techniques known to one of ordinary skill in the art can be used to estimate $G_m(\omega)$. One such technique is described below.

To collect samples of the desired signal portions $s_m[i]$ at the output of the microphones 26a-26M, the person 12 (FIG. 1) must be talking and the noise $\vec{n}[i]$ corresponding to the environmental noise signals $v_m[i]$ and the remote voice signals $e_m[i]$ must he much smaller than the desired signal $\vec{s}[i]$, i.e. the SNR at the output of each microphone 26a-26M must be high. This high SNR occurs whenever the talker is talking in a quiet environment.

Whenever the SNR is determined to be high, the signal processor 30 can collect

the desired signal $s_1[i]$ ($s_1[i]= r_1[i]$ for high SNR) from the output of the first microphone, and the signal processor 30 can collect $s_m[i]$ ($s_m[i]= r_m[i]$ for high SNR) from the output of the m-th microphone. The signal processor 30 can then use these samples to estimate the cross power-spectrum between $s_1[i]$ and $s_m[i]$ (denoted herein as $P_{s1sm}(\omega)$). A well-known method for estimating $P_{s1sm}(\omega)$ from samples of $s_1[i]$ and $s_m[i]$ is the Welch method of spectral estimation. Recall that $P_{s1sm}(\omega)$ is the Fourier Transform of:

$$\rho_{s1sm}[t] = E\{s_1[i]s_m[i+t]\};$$

therefore $\rho_{s1sm}(w)$ can be estimated.

Once $P_{s1sm}(\omega)$ is estimated, the signal processor 30 can use $P_{s1sm}(\omega)/P_{s1s1}(\omega)$ as the final estimate of $G_m(\omega)$, where $P_{s1s1}(\omega)$ is the power spectrum of $s_1[i]$ obtained using a Welch method.

In one particular embodiment, the person 12 (FIG. 1) can explicitly initiate the estimation of $\vec{G}(\omega)$ by commanding the system to start estimating $\vec{G}(\omega)$ at a particular time (e.g. by pushing a button and starting to talk). With this particular arrangement, the person 12 (FIG. 1) commands the system to start estimating $G(\omega)$ only when they determine that the SNR is high (i.e. the noise is low). Generally, in the environment of an automobile, for example, $\vec{G}(\omega)$ changes little over time for a particular user and for a particular automobile. Therefore, $\vec{G}(\omega)$ can be estimated once at installation of the hands free system 10 (FIG. 1) into the automobile.

In some arrangements, the hands-free system 10 (FIG. 1) can be used as a front-end to a speech recognition system that requires training. Such speech recognition systems (SRS) require the user to train the SRS by uttering a few words/phrases in a quiet environment. The noise reduction system can use the same training period for

estimating $\vec{G}(\omega)$ since, the training of the SRS is done also in a quiet environment.

Alternatively, the signal processor 30 can determine when the SNR is high, and it can initiate the process for estimating $\vec{G}(\omega)$. For example, in one particular embodiment, to estimate the SNR at the output of the first microphone, the signal processor 30, during the time when the talker is silent (as determined by the VAD 102), measures the power of the noise at the output of the first microphone 26a. The signal processor 30, during the time when the talker is active (as determined by the VAD 102), measures the power of the speech plus noise signal. The signal processor 30 estimates the SNR at the output of the first microphone 26a as the ratio of the power of the speech plus noise signal to the noise power. The signal processor 30 compares the estimated SNR to a desired threshold, and if the computed SNR exceeds the threshold, the signal processor 30 identifies a quiet period and begins estimating elements of $\vec{G}(\omega)$.

In either arrangement, upon either identification of a quiet period by a user or by the signal processor 30, each element of $\vec{G}(\omega)$ is estimated by the signal processor 30 as the ratio of the cross power spectra $P_{s1sm}(\omega)$ to the power spectrum $P_{s1s1}(\omega)$

Therefore, having adapted the AP filters 74 with the transfer function $\vec{F}(\omega)$ above, the SCNRP filters with the transfer function $Q(\omega)$ above, and the transfer functions $\vec{G}(\omega)$ with the techniques above, the output of the hands-signal processor 30 is the estimate signal $\hat{s}_l[i]$, as desired.

The noise signal portions $n_m[i]$ and the desired signal portions $s_m[i]$ of the microphone signals $r_m[i]$ can vary at substantially different rates. Therefore, the structure of the signal processor 30, having the first and the second adaptation processors 92, 94 respectively, can provide different adaptation rates for the AP filters

74a-74M and for the SCNRP filter 80. As described above, having different adaptation rates results in a more accurate adaptation of the AP filters; therefore, this results in improved noise reduction.

Referring now to FIG. 6, a circuit portion 120 of an the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes a first adaptation processor 134. Unlike the first adaptation processor 92 of FIG. 5, the first adaptation processor 134 does not contain the VAD 102 (FIG. 5). Therefore, an update processor 130, must compute the noise power spectrum $P_{\bar{n}\bar{n}}(\omega)$ while both the noise portions $n_m[i]$ of the input signals $r_m[i]$ and the desired signal portions $s_m[i]$ of the input signals $r_m[i]$ are present, i.e. while the person 12 (FIG. 1) is talking.

In this particular embodiment, in order to accomplish calculation of $P_{\bar{n}\bar{n}}(\omega)$ while the person 12 (FIG. 1) is talking, it would be desirable to subtract the desired signal portions $s_m[i]$ from the input signals $r_m[i]$ before receiving them with the first adaptation processor 134. However, the desired signal portions $s_m[i]$ are not explicitly known by the signal processor 30. Therefore, signals representing the desired signal portions $s_m[i]$ are instead subtracted from input signals $r_m[i]$.

A good estimate of a particular desired signal portion from the first microphone appears as the estimate signal $\hat{s}_1[i]$ at the output of the SCNRP filter 80. Therefore, in one embodiment, the estimate signal $\hat{s}_1[i]$ is passed through subtraction processors 126a-126M, and the resulting signals are subtracted from the input signals $r_m[i]$ via subtraction circuits 122a-122M to provide subtracted signals 128a-128M to the update processor 130. The subtraction processors 126a-126M comprise filters that operate upon the estimate signal $\hat{s}_1[i]$. The subtracted signals 128a-128M are substantially

noise signals, corresponding substantially to the noise signal portions $n_m[i]$ of the input signals $r_m[i]$. Therefore, the update processor 130 can compute the noise power spectrum $P_{\vec{n}\vec{n}}(\omega)$ and the inverse thereof used in computation of the responses $\vec{F}(\omega)$ of the AP filters 74a-74M from the equations given above.

While this embodiment 120 couples the subtraction processors 126a-126M to the estimate signal $\hat{s}_1[i]$ at the output of the SCNRP filter 80, in other embodiments, the subtraction processors can be coupled to other points of the system. For example, the subtraction filters can be coupled to the intermediate signal z[i].

The subtraction processors 126a-126M have the transfer functions $G_m(\omega)$, which, as described above, relate the desired signal portion of the first microphone $S_1(\omega)$ to the desired signal portion of the m-th microphone $S_m(\omega)$, (i.e. $G_m(\omega) = S_m(\omega)/S_1(\omega)$).

Referring now to FIG. 7, a circuit portion 150 of an the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes a data processor 162. The data processor 162 is shown without the first and second adaptation processors 134, 94 respectively of FIG. 6. However, it will be understood that the data processor 162 is but part of a signal processor, for example the signal processor 30 of FIG 6, which includes first and second adaptation processors, for example the first and second adaptation processors 134, 94 of FIG. 6.

The data processor 162 includes an AP 156 and a SCNRP 160 that can correspond, for example to the AP 52 and the SCNRP 78 of FIG. 6. The remote-voice-producing signal q[i] that drives the loudspeaker 20 to produce the remote voice signal 22 (FIG. 1) is introduced to remote voice canceling processors 154a-154M. The remote voice canceling processors 154a-154M comprise filters that operate upon the

remote-voice-producing signal q[i]. The outputs of the remote voice canceling processors 154a-154M are subtracted via subtraction circuits 152a-152M from the signals $r_1[i]$ to $r_m[i]$ provided by the microphones 26a-26m. Therefore, noise attributed to the remote-voice-producing signal q[i] which forms a part of the signals $r_1[i]$ to $r_m[i]$ is subtracted from the signals $r_1[i]$ to $r_m[i]$ before the subsequent processing is performed by the AP 156 in conjunction with first and second adaptation processors (not shown).

Therefore, in this particular embodiment :

$$\vec{r}\,[i] = \vec{r}\,[i] - \vec{k}\,[i] * q[i]$$

In the above equation, k[i] is the impulse-response of the acoustic channel between q[i] and the intermediate signal z[i]. The transfer function of the m-th remote voice-canceling filter is $K_m(\omega)$, where $K_m(\omega)$ is an estimate of the transfer function with input q[i] and output $e_m[i]$, (i.e., $K_m(\omega) = E_m(\omega)/Q(\omega)$).

With this particular arrangement, the effect of the remote voice-producing signal q[i] on intelligibility of the estimate signal $\hat{s}_1[i]$ is reduced with the remote voice canceling processors 154a-154M.

Referring now to FIG. 8, a circuit portion 170 of an the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes a data processor 180. The data processor 180 is shown without the first and second adaptation processors 134, 94 respectively of FIG. 6. However, it will be understood that the data processor 180 is but part of a signal processor, for example the signal processor 30 of FIG 6, which includes first and second adaptation processors, for example the first and second adaptation processors 134, 94 of FIG. 6.

The data processor 180 includes an AP 172 and a SCNRP 174 that can

correspond, for example to the AP 52 and the SCNRP of FIG. 6. The remote-voice-producing signal q[i] that drives the loudspeaker 20 to produce the remote voice signal 22 (FIG. 1) is introduced to a remote voice canceling processor 178. The remote voice canceling processor 178 comprises a filter that operates upon the remote-voice-producing signal q[i]. The output of the remote voice canceling processor 178 is subtracted via subtraction circuit 176 from the estimate signal $\hat{s}_1[i]$, therefore providing an improved estimate signal $\hat{s}_1[i]'$. Therefore, noise attributed to the remote-voice-producing signal q[i] which forms a part of the signals $r_1[i]$ to $r_m[i]$ is subtracted from the final output of the data processor 180.

The response of the signal channel between q[i] and the output of the SCNRP 174 is:

$$\overline{P}(\omega) = \sum_{m=1}^{M} K_m(\omega)F_m(\omega)Q(\omega)$$

In the above equation, $K_m(\omega)$ is the transfer function of the acoustic channel with input q[i] and output $e_m[i]$, $F_m(\omega)$ is the transfer function of the m-th filter of the AP 172, and $Q(\omega)$ is the transfer function of the SCNRP 174.

With this particular arrangement, the effect of the remote-voice-producing signal q[i] on intelligibility of the improved estimate signal $\hat{s}_1[i]'$ is reduced with but one echo-canceling processor 178.

Referring now to FIG. 9, a circuit portion 190 of the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes a data processor 200. The data processor 200 is shown without the first and second adaptation processors 134, 94 respectively of FIG. 6. However, it will be understood that the data processor 200 is but part of a signal processor, for example the signal processor 30 of FIG 6, which includes first and second adaptation

processors, for example the first and second adaptation processors 134, 94 of FIG. 6.

The data processor 200 includes an AP 192 and a SCNRP 198 that can correspond, for example to the AP 52 and the SCNRP of FIG. 6. The remote-voice-producing signal q[i] that drives the loudspeaker 20 to produce the remote voice signal 22 (FIG. 1) is introduced to remote voice canceling processor 194. The remote voice canceling processor 194 comprises a filter that operates upon the remote-voice-producing signal q[i]. The output of the remote voice canceling processor 194 is subtracted via subtraction circuit 196 from the intermediate signal z[i], therefore providing an improved estimate signal z[i]'. Therefore, noise attributed to the remote-voice-producing signal q[i] which forms a part of the signals $r_1[i]$ to $r_m[i]$ is subtracted from the intermediate signal z[i].

The response of the signal channel between q[i] and the output of the AP 172 is:

$$\tilde{P}(\omega) = \sum_{m=1}^{M} K_m(\omega)F_m(\omega)$$

In the above equation, $K_m(\omega)$ is the transfer function of the acoustic channel with input q[i] and output $e_m[i]$, and $F_m(\omega)$ is the transfer function of the m-th AP filter within the AP 172 .

With this particular arrangement, the effect of the remote-voice-producing signal q[i] on intelligibility of the estimate signal $\hat{s}_1[i]$ is reduced with but one echo-canceling processor 194.

Referring now to FIG. 10, a circuit portion 210 of an the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes the microphones 26a-26M each coupled to a respective serial-to-parallel converter 212a-212M. The serial to parallel converters store data samples from the signals $r_1[i]$-$r_m[i]$ into data groups. The serial to parallel converters 212a-212M

provide the data groups to N1-point discrete Fourier transform (DFT) processors 214a-214M. The DFT processors 212a-212M are each coupled to a data processor 216 and an adaptation processor 218 which can be similar to the data processor 52 and adaptation processor 54 described above in conjunction with FIG. 6.

In operation, the DFT processors convert the time-domain samples $r_m[i]$ into frequency domain samples, which are provided to the data processor 216 and to the adaptation processor 218. Therefore, frequency domain samples are provided to both the data processor 216 and the adaptation processor 218. Filtering performed by AP filters (not shown) within the data processor 216 and power spectrum calculations provided by the adaptation processor 218 can be done in the frequency domain as is described above.

Referring now to FIG. 11, a circuit portion 230 of an the exemplary hands-free system 10 of FIG. 1, in which like elements of FIG. 1 are shown having like reference designations, includes the microphones 26a-26M each coupled to respective serial-to-parallel converter 232a-232M and respective serial-to parallel converters 234a-234M. The serial to parallel converters store data samples from the signals $r_1[i]$ to $r_m[i]$ into data groups and provide the data groups to N1-point discrete Fourier transform (DFT) processors 236a-236M. The serial to parallel converters 234a-234M provide the data groups to window processors 238a-238M and thereafter to N2-point discrete Fourier transform (DFT) processors 238a-238M. The DFT processors 236a-236M are each coupled to a data processor 242. The DFT processors 240a-240M are each coupled to an adaptation processor 244. The data processor 242 and the adaptation processor 244 can be the type of data processor 52 and adaptation processor 54 of FIG. 6.

In operation, the DFT processors convert the time-domain data groups into

frequency domain samples, which are provided to the data processor 242 and to the adaptation processor 244. Therefore, frequency domain samples are provided to both the data processor 242 and the adaptation processor 244. Therefore, filtering provided by AP filters (not shown) in the data processor 242 and power spectrum calculations provided by the adaptation processor 244 can be done in the frequency domain as is described above.

It is known in the art that the accuracy of estimating the noise power spectrum $P_{\bar{n}\bar{n}}(\omega)$ and the inverse thereof $P^{-1}_{\bar{n}\bar{n}}(\omega)$ can be improved by applying a windowing function, such as that provided by the windowing processors 238a-238M. Therefore, the windowing processors 238a-238M provide the adaptation processor 244 with an improved ability to accurately determine the noise power spectrum and therefore to update the AP filters (not shown) within the data processor 242. However, it is also known that the use of windowing on signals that are used to provide an audio output in the data processor 216 results in distorted audio and a less intelligible output signal. Therefore, while is it desirable to provide the windowing processors 238a-238M for the signals to the adaptation processor 244, it is not desirable to provide windowing processors for the signals to the data processor 242.

With the particular arrangement shown in the circuit portion 230, the N1-point DFT processors 236a-236M and the N2-point DFT processors 240a-240M can compute using a number of time domain data samples N1 different from a number of time domain data samples N2.

All references cited herein are hereby incorporated herein by reference in their entirety.

Having described preferred embodiments of the invention, it will now become

apparent to one of ordinary skill in the art that other embodiments incorporating their concepts may be used. It is felt therefore that these embodiments should not be limited to disclosed embodiments, but rather should be limited only by the spirit and scope of the appended claims.

5

What is claimed is:

CLAIMS

1. A system for processing one or more input signals, the system comprising:

a first processor having one or more channels, each channel comprising a respective first processor filter, each channel configured to receive a respective one of the one or more input signals, wherein the first processor is configured to provide an intermediate output signal;

a second processor comprising a second processor filter configured to receive the intermediate output signal and provide a noise-reduced output signal;

a first adaptation processor coupled to the first processor; and

a second adaptation processor coupled to the second processor.

2. The system of Claim 1, wherein a noise signal portion of each respective one of the one or more input signals comprises a representation of acoustic noise, and a desired signal portion of each respective one of the one or more input signals comprises a representation of a voice.

3. The system of Claim 1, wherein the first adaptation processor adapts the first processor filter in each of the one or more channels in response to a variation of a power spectral density (PSD) of a noise signal portion of respective ones of the one or more input signals.

4. The system of Claim 3, wherein the first adaptation processor does not respond to variations of the power spectral density of a desired signal portion of respective ones of the one or more input signals.

5. The system of Claim 3, wherein the first adaptation processor includes a power spectral density inversion processor that directly provides the inverse of the power spectral density (PSD) of the noise signal portion of respective ones of the one or more input signals.

6. The system of Claim 1, wherein the second adaptation processor adapts the second processor filter in response to variations of the power spectral density (PSD) of a desired signal portion of the intermediate output signal.

7. The system of Claim 1, wherein the second adaptation processor adapts the second processor filter in response to variations of the power spectral density (PSD) of the intermediate output signal and to variations of the power spectral density (PSD) of a noise portion of the intermediate output signal.

8. The system of Claim 1, wherein the first adaptation processor includes a voice activity detection (VAD) processor coupled to the intermediate output signal, the VAD processor having a VAD processor output for indicating when a desired signal portion of the intermediate output signal is absent.

9. The system of Claim 8, wherein the first adaptation processor adapts the first processor filter in each of the one or more channels in response to the VAD processor output.

10. The system of Claim 9, wherein the first adaptation processor adapts the first processor filter in each of the one or more channels in response to a noise portion of respective ones the one or more input signals, in response to the VAD processor output.

11. The system of Claim 1, wherein the first adaptation processor includes a voice activity detection (VAD) processor coupled to at least one of the one or more input signals, the VAD processor having a VAD processor output for indicating when a desired signal portion of the at least one of the one or more input signals is absent.

12. The system of Claim 11, wherein the first adaptation processor adapts the first processor filter in each of the one or more channels in response to the VAD processor output.

13. The system of Claim 12, wherein the first adaptation processor adapts the first

processor filter in each of the one or more channels in response to the noise portion of respective one of the one or more input signals, in response to the VAD processor output.

14. The system of Claim 1, wherein the first adaptation processor includes a subtraction processor for subtracting a filtered version of an estimate of a desired signal portion from each of the one or more input signals to provide one or more respective subtracted signals.

15. The system of Claim 14, wherein the first adaptation processor adapts the first processor filter in each of the one or more channels in response to a variation of a power spectral density (PSD) of the one or more subtracted signals.

16. The system of Claim 14, wherein the first adaptation processor includes a subtraction processor for subtracting a filtered version of the intermediate signal or a filtered version of the noise-reduced output signal from each of the one or more input signals to provide one or more respective subtracted signals.

17. The system of Claim 16, wherein the first adaptation processor adapts the first processor filter in each of the one or more channels in response to a variation of a power spectral density (PSD) of the one or more subtracted signals.

18. The system of Claim 1, wherein the first adaptation processor adapts the first processor filters in each of the one or more channels so that the intermediate output is a maximum-likelihood estimate of a desired signal portion of the one or more input signals.

19. The system of Claim 1, wherein the second processor filter comprises a single-input single-output Weiner filter.

20. The system of Claim 1, wherein the first adaptation processor adapts the first processor filters in each of the one or more channels so that the intermediate output is a

maximum-likelihood estimate of a desired signal portion of the one or more input signals, and the second processor filter comprises a single-input single-output Weiner filter.

21. The system of Claim 1, wherein the first processor includes an un-windowed discrete Fourier transform (DFT) processor.

22. The system of Claim 1, wherein the first adaptation processor includes a windowed discrete Fourier transform (DFT) processor.

23. The system of Claim 1, further including a remote voice canceling processor for subtracting a remote-voice-producing signal from each of the one or more input signals.

24. The system of Claim 1, further including a remote voice canceling processor for subtracting a remote-voice-producing signal from the intermediate output signal.

25. The system of Claim 1, further including a remote voice canceling processor for subtracting a remote-voice-producing signal from the noise-reduced output signal.

26. A system, comprising:

a first filter portion configured to receive one or more input signals and to provide a single intermediate output signal;

a second filter portion configured to receive the single intermediate output signal and to provide a single output signal; and

a control circuit configured to receive at least a portion of each of the one or more input signals and at least a portion of the single intermediate output signal and to provide information to adapt filter characteristics of the first and second filter portions.

27. The system of Claim 26, wherein the control circuit comprises a first adaptation processor for providing first information to adapt the filter characteristics of the first filter portion and a second adaptation processor for providing second information to adapt the filter characteristics of the second filter portion.

28. The system of Claim 27, wherein the first information corresponds to a noise power spectral density of the one or more input signals and the second information corresponds to one or more of: a power spectral density of: a noise portion of the intermediate output signal, a power spectral density of a desired signal portion of the intermediate output signal, and a power spectral density of the intermediate output signal.

29. The system of Claim 26, further including an echo canceling processor coupled to receive the single output signal, for reducing an echo signal portion of the output signal by subtracting a remote-voice-producing signal from at least one of: the one or more input signals, the single intermediate output signal, and the single output signal.

30. A method for processing one or more input signals, comprising:

receiving the one or more input signals with a first filter portion, the first filter portion providing an intermediate output signal;

receiving the intermediate output signal with a second filter portion, the second filter portion providing an output signal;

dynamically adapting a response of the first filter portion and a response of the second filter portion.

31. The method of Claim 30, wherein the dynamically adapting comprises adapting a response of the first filter portion in response to a noise portion of the one or more input signals and adapting a response of the second filter portion in response to a power spectral density of at least one of: a noise portion of the intermediate output signal, a desired signal portion of the intermediate output signal, and characteristics of the intermediate output signal.

32. The method of Claim 31, wherein the receiving with a first filter portion comprises receiving with a maximum-likelihood filter having multiple inputs and a single output, and the receiving with a second filter portion comprises receiving with a

single-input single-output Weiner filter.

33. The method of Claim 30, further including:

reducing a remote voice signal portion of the output signal by subtracting a remote-voice-producing signal from at least one of: the one or more input signals, the intermediate output signal, and the output signal.

34. The method of Claim 30, further including:

estimating a transfer function between respective ones of the one or more input signals in a training period during which a person determines that the one or more input signals have a high signal to noise ratio.

35. The method of Claim 30, further including:

estimating a transfer function between respective ones of the one or more input signals in a training period during which a signal processor determines that the one or more input signals have a high signal to noise ratio.

36. The method of Claim 35, wherein the estimating the transfer function in the training period comprises estimating the transfer function in the training period corresponding to the training period associated with a voice recognition system.

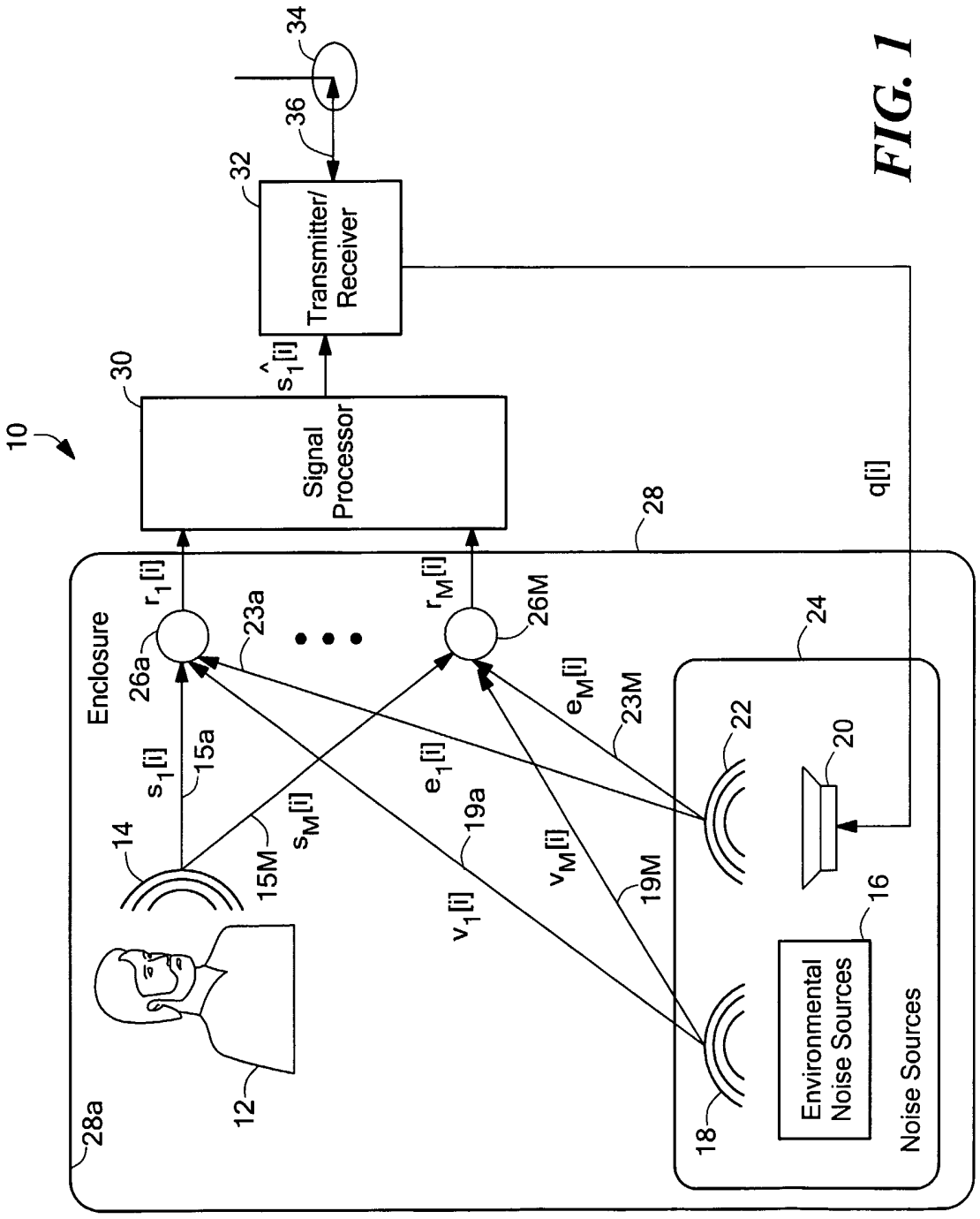

10
34
36
32
Transmitter/
Receiver
30
Signal
Processor
$\hat{s}_1[i]$
28
q[i]
Enclosure
$r_1[i]$
26a
23a
$r_M[i]$
26M
24
$e_M[i]$
23M
22
20
$s_1[i]$
15a
14
$s_M[i]$
15M
$e_1[i]$
19a
$v_M[i]$
19M
$v_1[i]$
16
12
18
Environmental
Noise Sources
Noise Sources
28a

FIG. 1

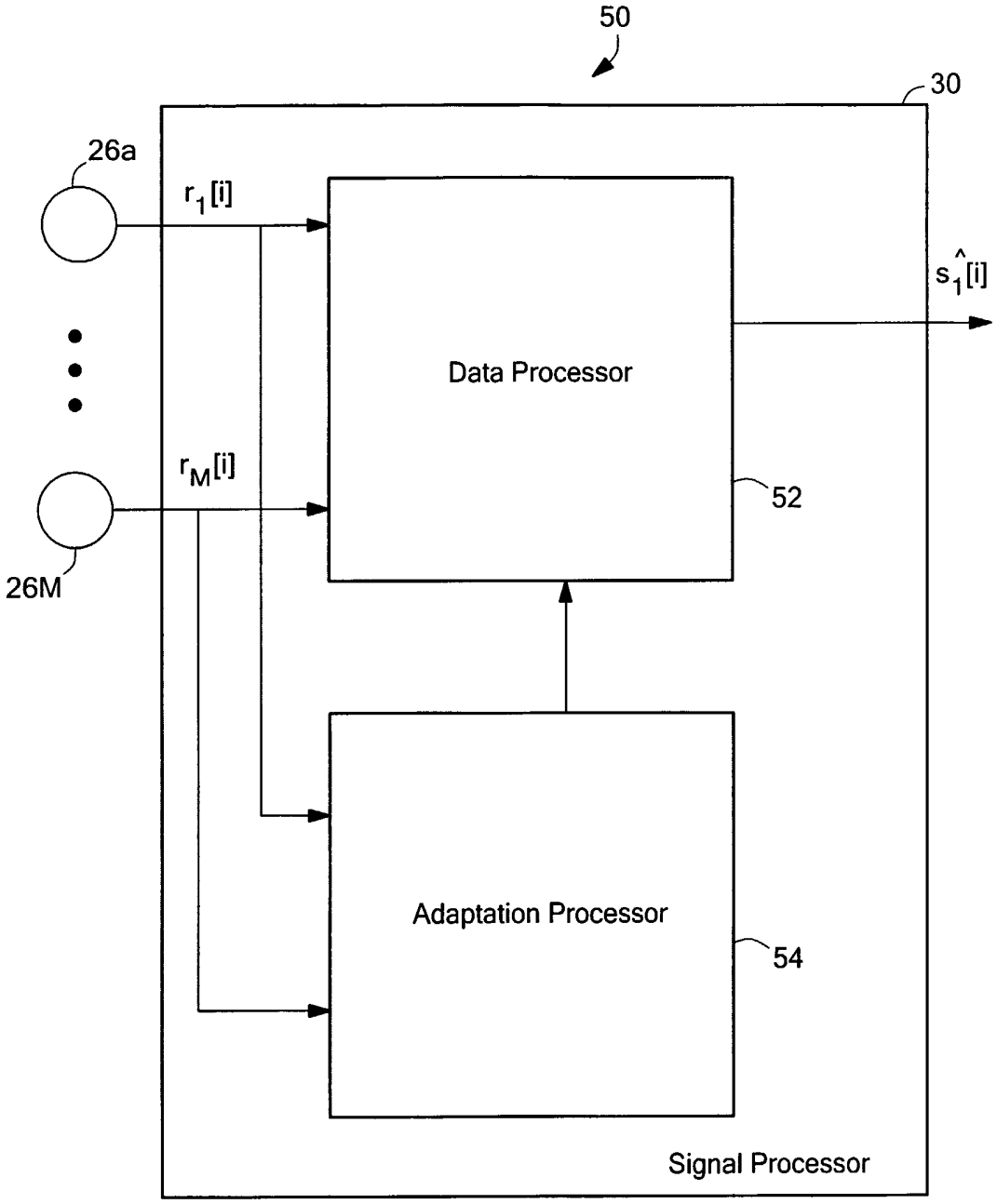

50
30
26a
$r_1[i]$
$\hat{s_1}[i]$
Data Processor
$r_M[i]$
52
26M
Adaptation Processor
54
Signal Processor

FIG. 2

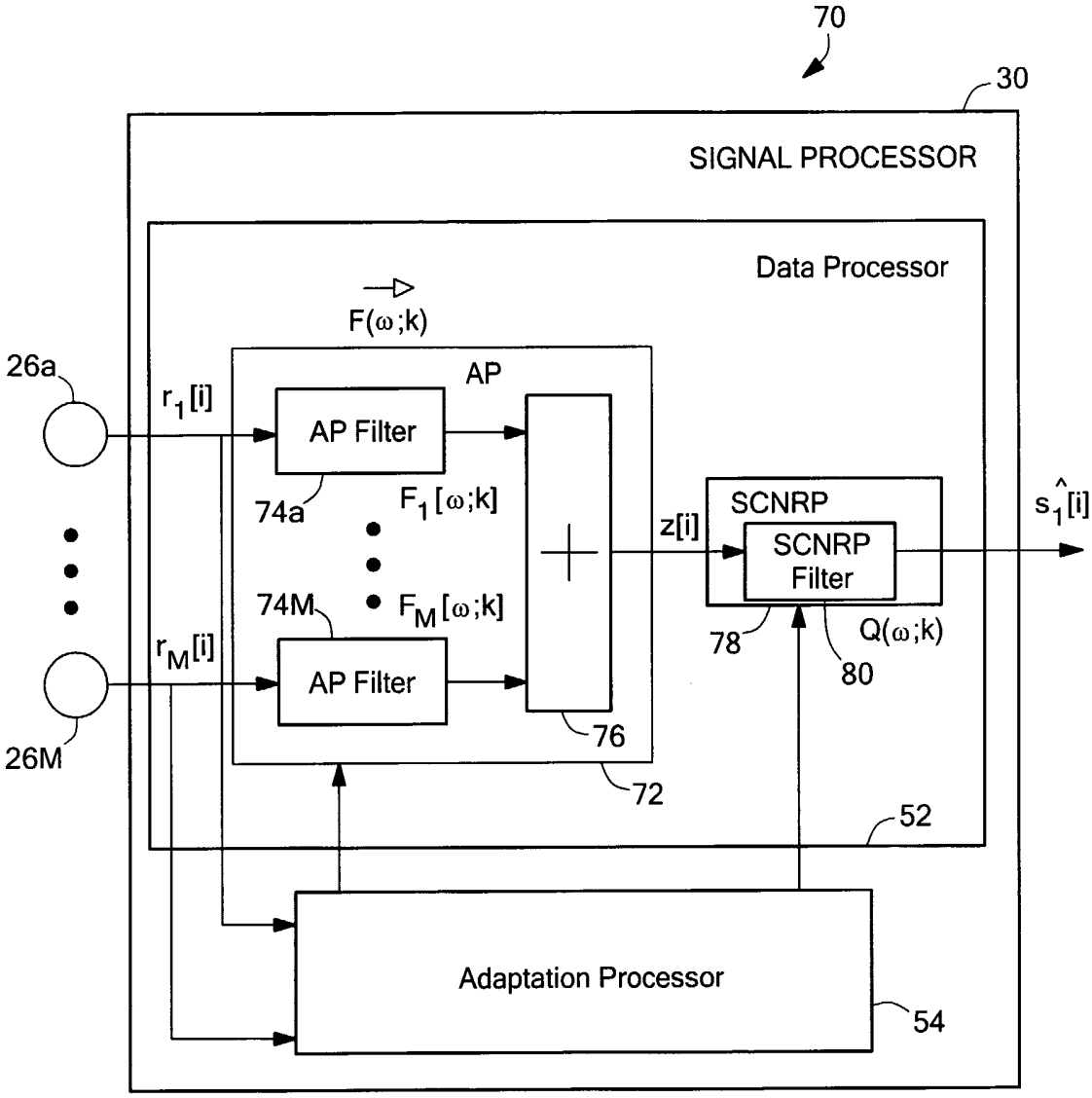

70
30
SIGNAL PROCESSOR
Data Processor
F(ω;k)
AP
26a
$r_1[i]$
AP Filter
74a
$F_1[\omega;k]$
$F_M[\omega;k]$
74M
$r_M[i]$
AP Filter
26M
76
72
z[i]
SCNRP
SCNRP Filter
78
80
Q(ω;k)
$\hat{s}_1[i]$
52
Adaptation Processor
54

FIG. 3

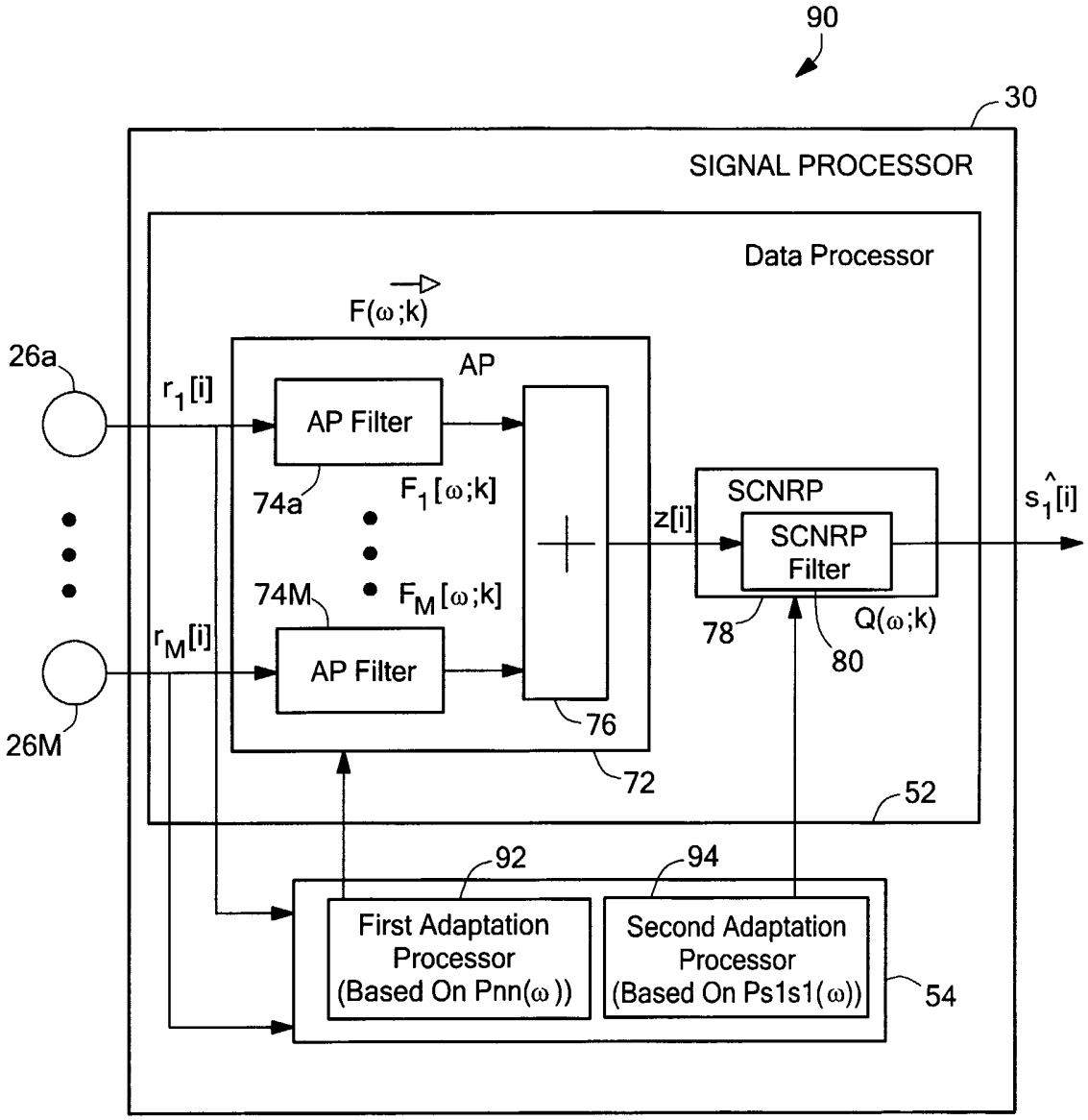

90
30
SIGNAL PROCESSOR
Data Processor
F(ω;k)
AP
26a
r1[i]
AP Filter
74a
F1[ω;k]
74M
FM[ω;k]
rM[i]
AP Filter
26M
76
72
z[i]
SCNRP
SCNRP Filter
78
80
Q(ω;k)
ŝ1[i]
52
92
94
First Adaptation Processor (Based On Pnn(ω))
Second Adaptation Processor (Based On Ps1s1(ω))
54

FIG. 4

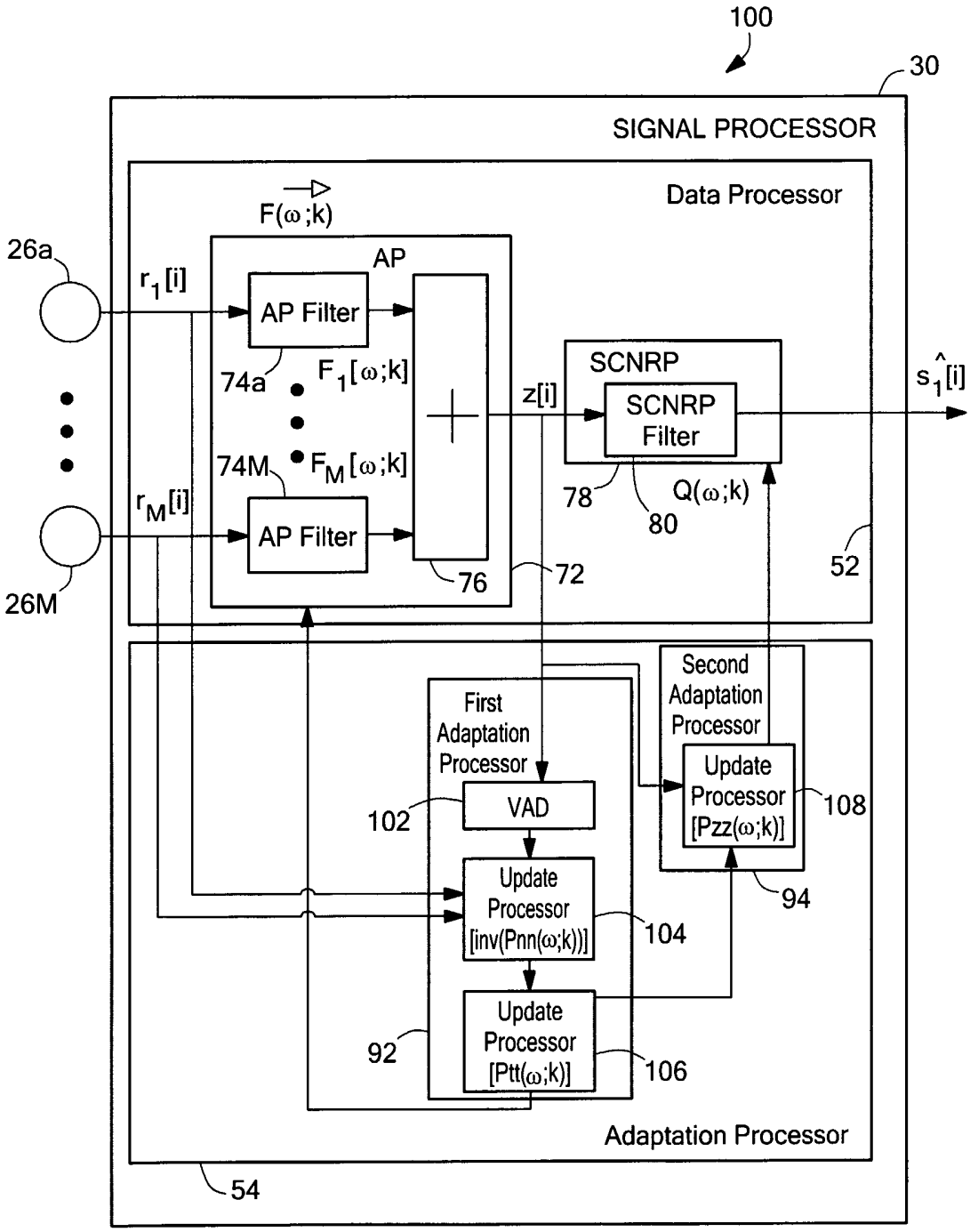

100
30
SIGNAL PROCESSOR
Data Processor
F(ω;k)
AP
26a
r1[i]
AP Filter
74a
F1[ω;k]
74M
FM[ω;k]
rM[i]
AP Filter
26M
76
72
z[i]
SCNRP
SCNRP Filter
78
Q(ω;k)
80
ŝ1[i]
52
First Adaptation Processor
VAD
102
Update Processor [inv(Pnn(ω;k))]
104
Update Processor [Ptt(ω;k)]
92
106
Second Adaptation Processor
Update Processor [Pzz(ω;k)]
108
94
Adaptation Processor
54

FIG. 5

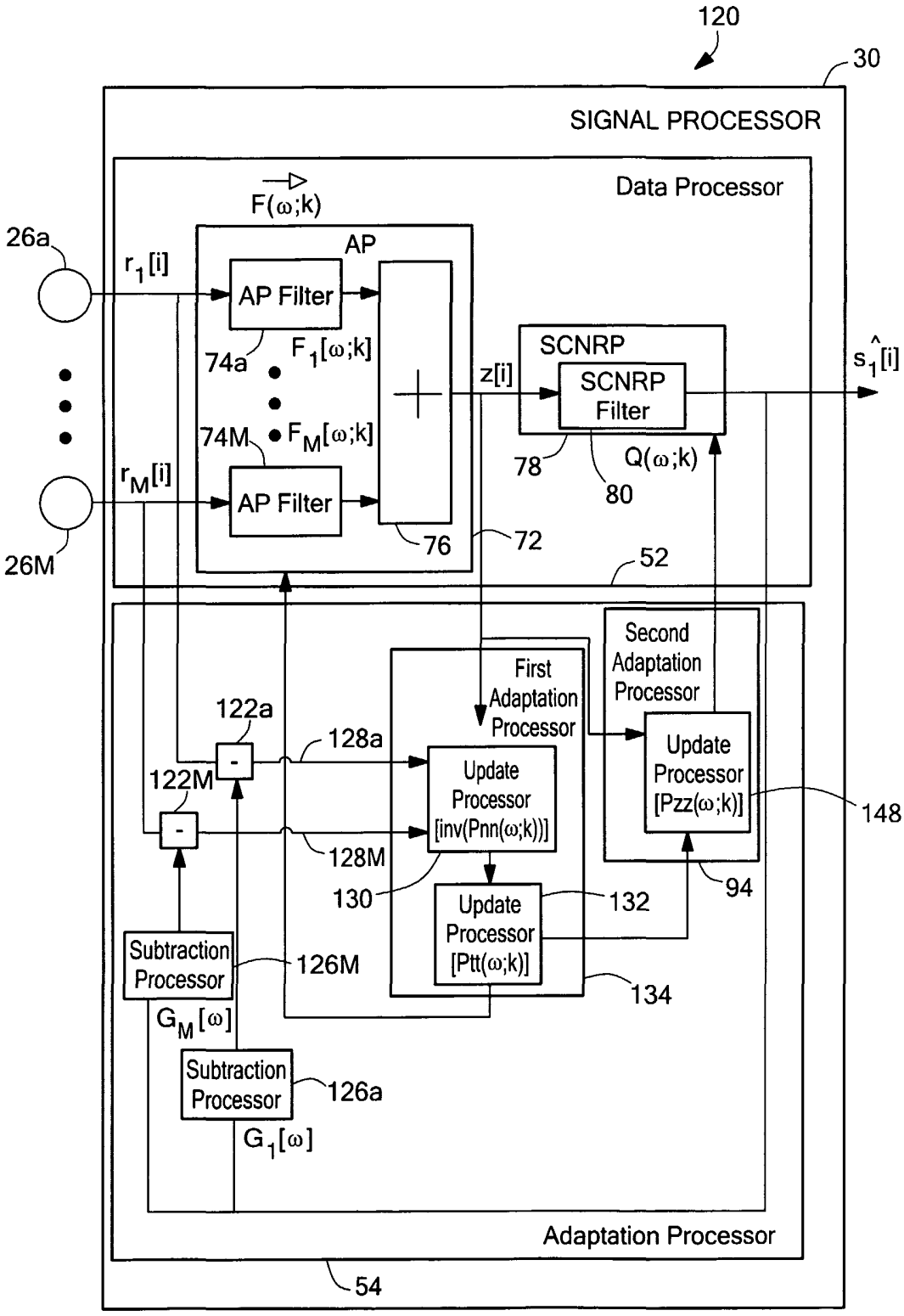

120
30
SIGNAL PROCESSOR
Data Processor
F(ω;k)
26a
AP
r1[i]
AP Filter
74a
F1[ω;k]
74M
FM[ω;k]
rM[i]
AP Filter
26M
76
72
z[i]
SCNRP
SCNRP Filter
78
80
Q(ω;k)
ŝ1[i]
52
Second Adaptation Processor
First Adaptation Processor
122a
128a
122M
128M
Update Processor [inv(Pnn(ω;k))]
Update Processor [Pzz(ω;k)]
148
130
Update Processor [Ptt(ω;k)]
132
94
Subtraction Processor
126M
GM[ω]
134
Subtraction Processor
126a
G1[ω]
Adaptation Processor
54

FIG. 6

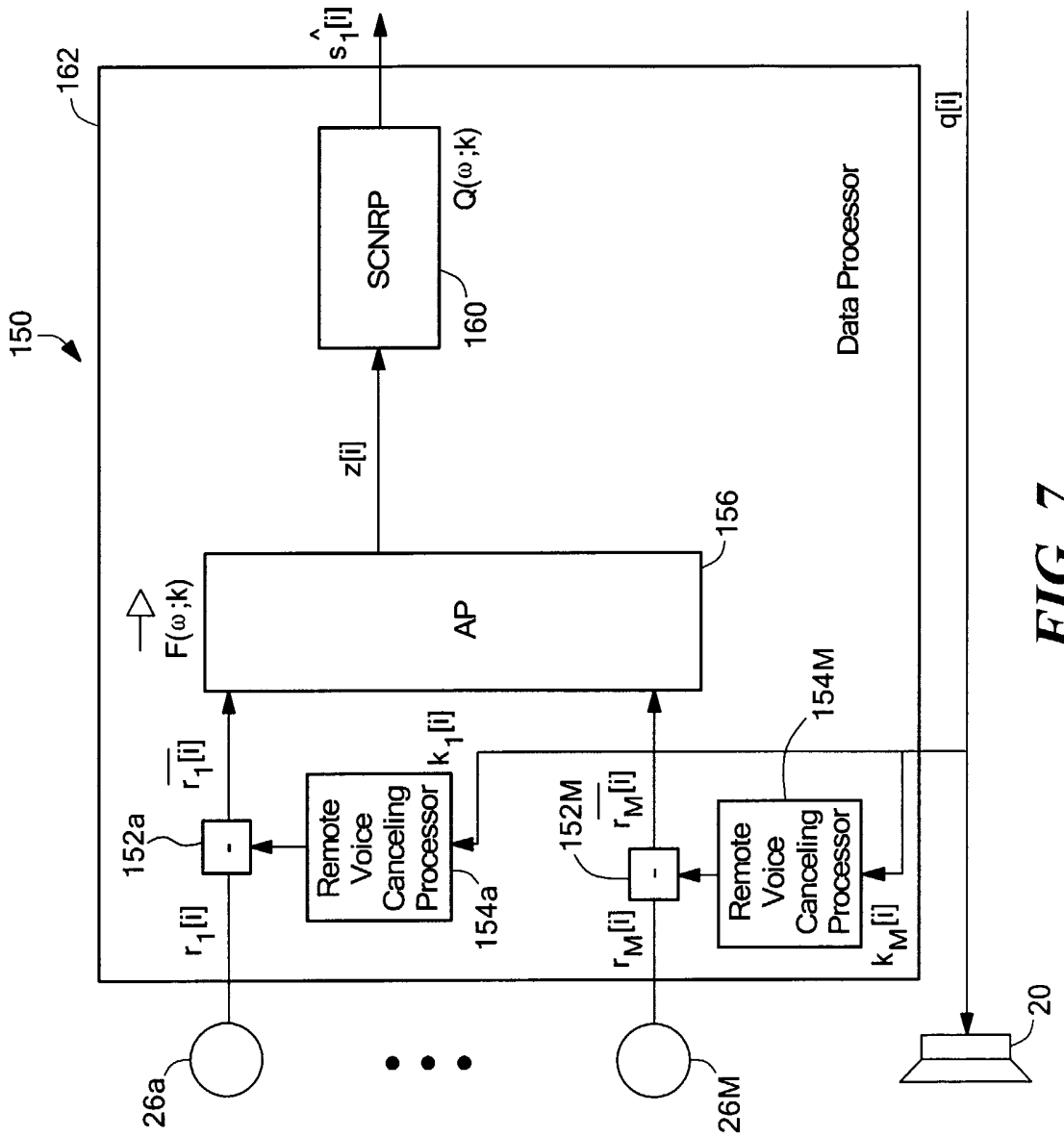

150
162
$\hat{s}_1[i]$
SCNRP
$Q(\omega;k)$
160
Data Processor
q[i]
z[i]
156
AP
$F(\omega;k)$
$\overline{r_1}[i]$
152a
$k_1[i]$
Remote Voice Canceling Processor
154a
154M
152M
$\overline{r_M}[i]$
$r_1[i]$
$r_M[i]$
Remote Voice Canceling Processor
$k_M[i]$
20
26a
26M

FIG. 7

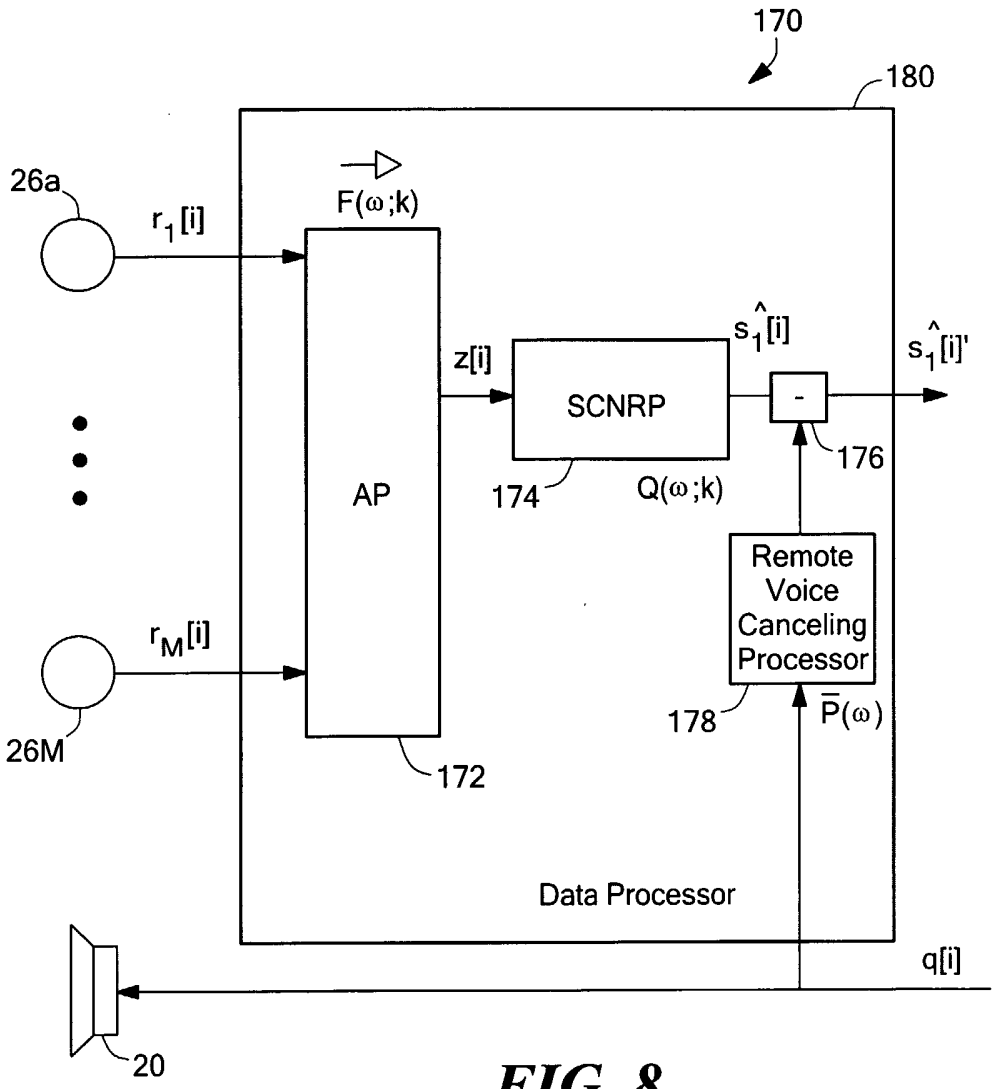

170
180
26a
$r_1[i]$
F(ω;k)
AP
z[i]
SCNRP
$\hat{s}_1[i]$
$\hat{s}_1[i]'$
-
176
174
Q(ω;k)
Remote Voice Canceling Processor
178
$\bar{P}(\omega)$
$r_M[i]$
26M
172
Data Processor
q[i]
20

FIG. 8

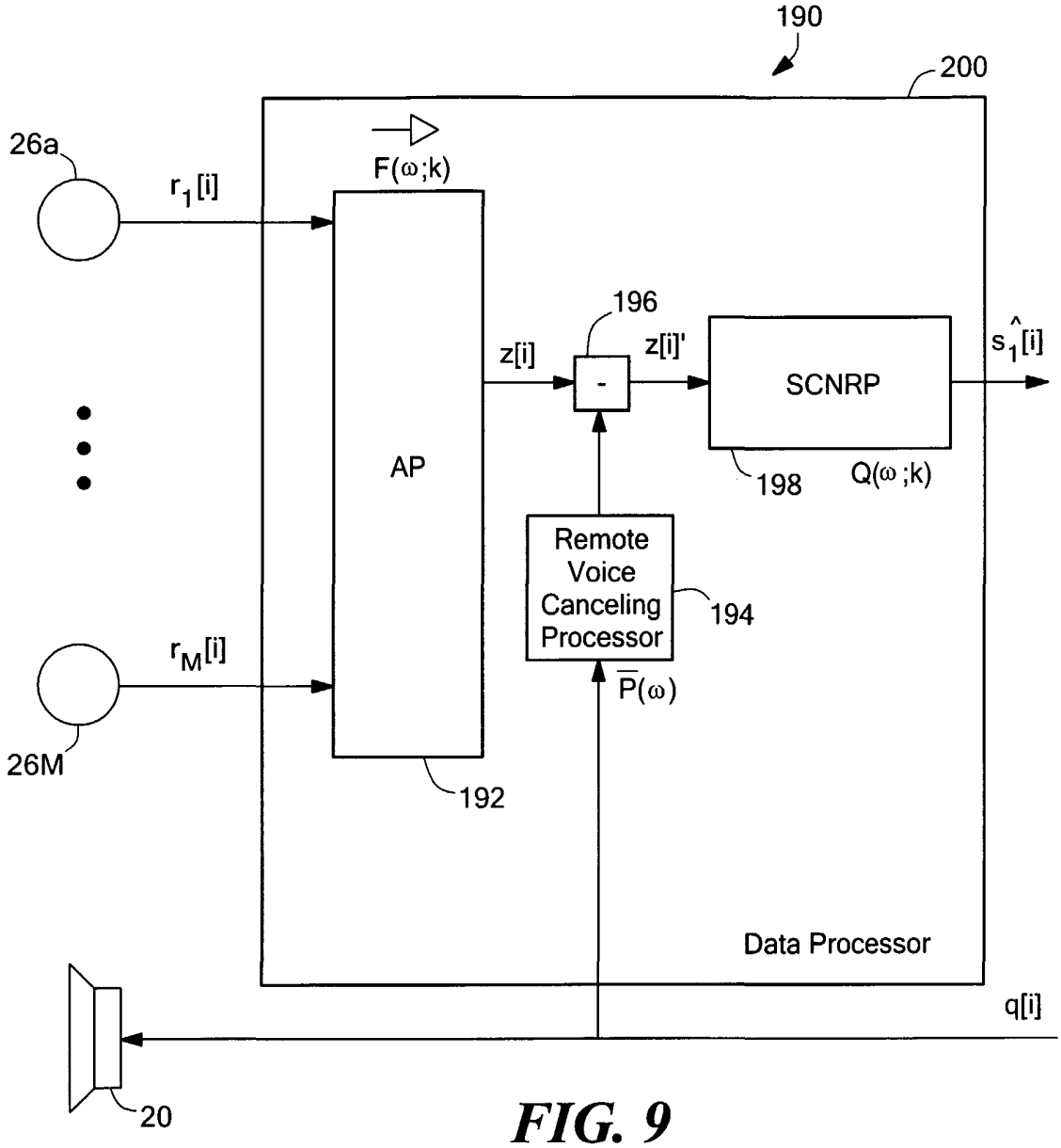

190
200
26a
$r_1[i]$
F(ω;k)
196
z[i]
z[i]'
$\hat{s}_1[i]$
SCNRP
AP
198
Q(ω;k)
Remote Voice Canceling Processor
194
$r_M[i]$
$\overline{P}(\omega)$
26M
192
Data Processor
q[i]
20

FIG. 9

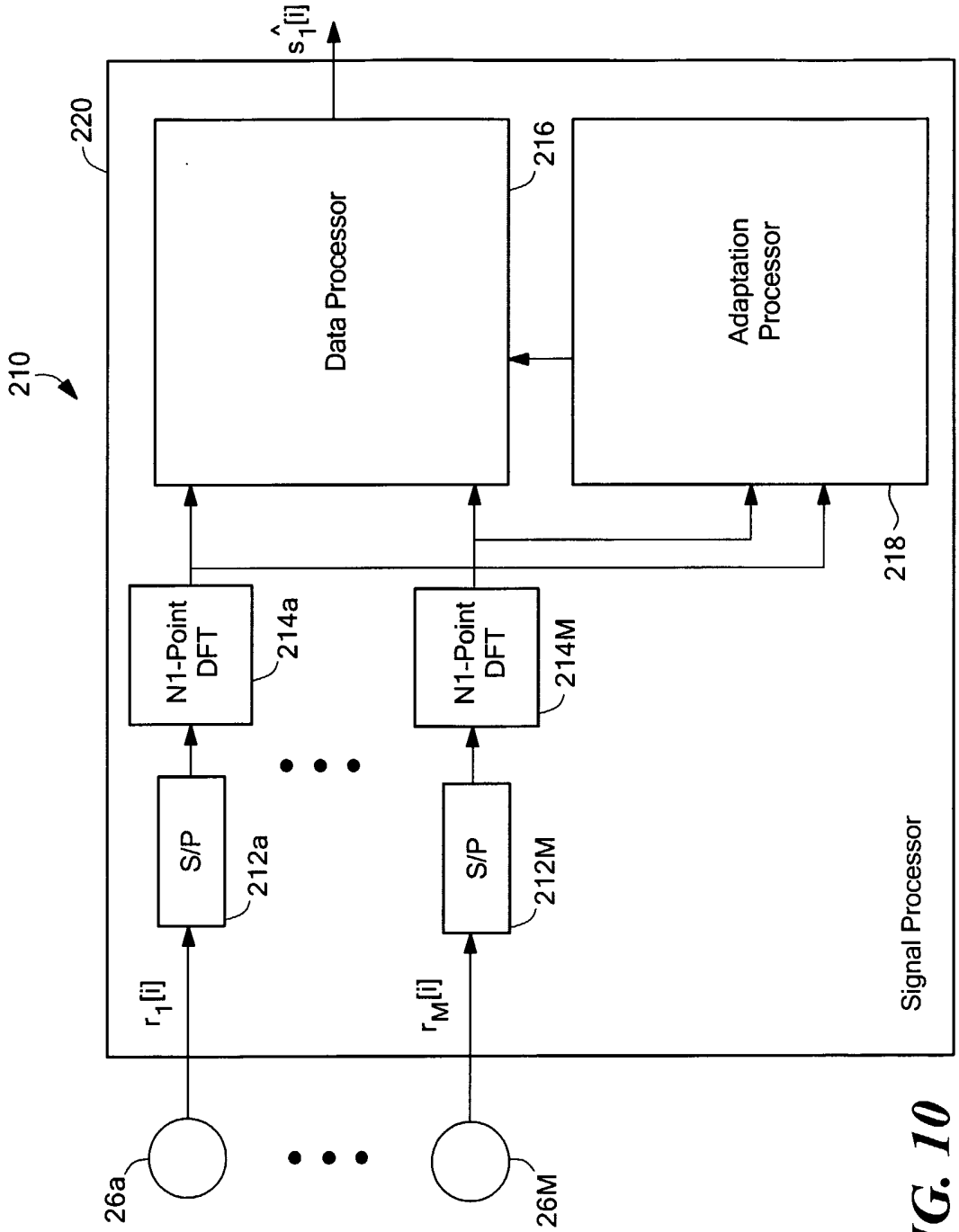

210
220
$\hat{s}_1[i]$
Data Processor
216
Adaptation Processor
218
N1-Point DFT
214a
N1-Point DFT
214M
S/P
212a
S/P
212M
$r_1[i]$
$r_M[i]$
Signal Processor
26a
26M

FIG. 10

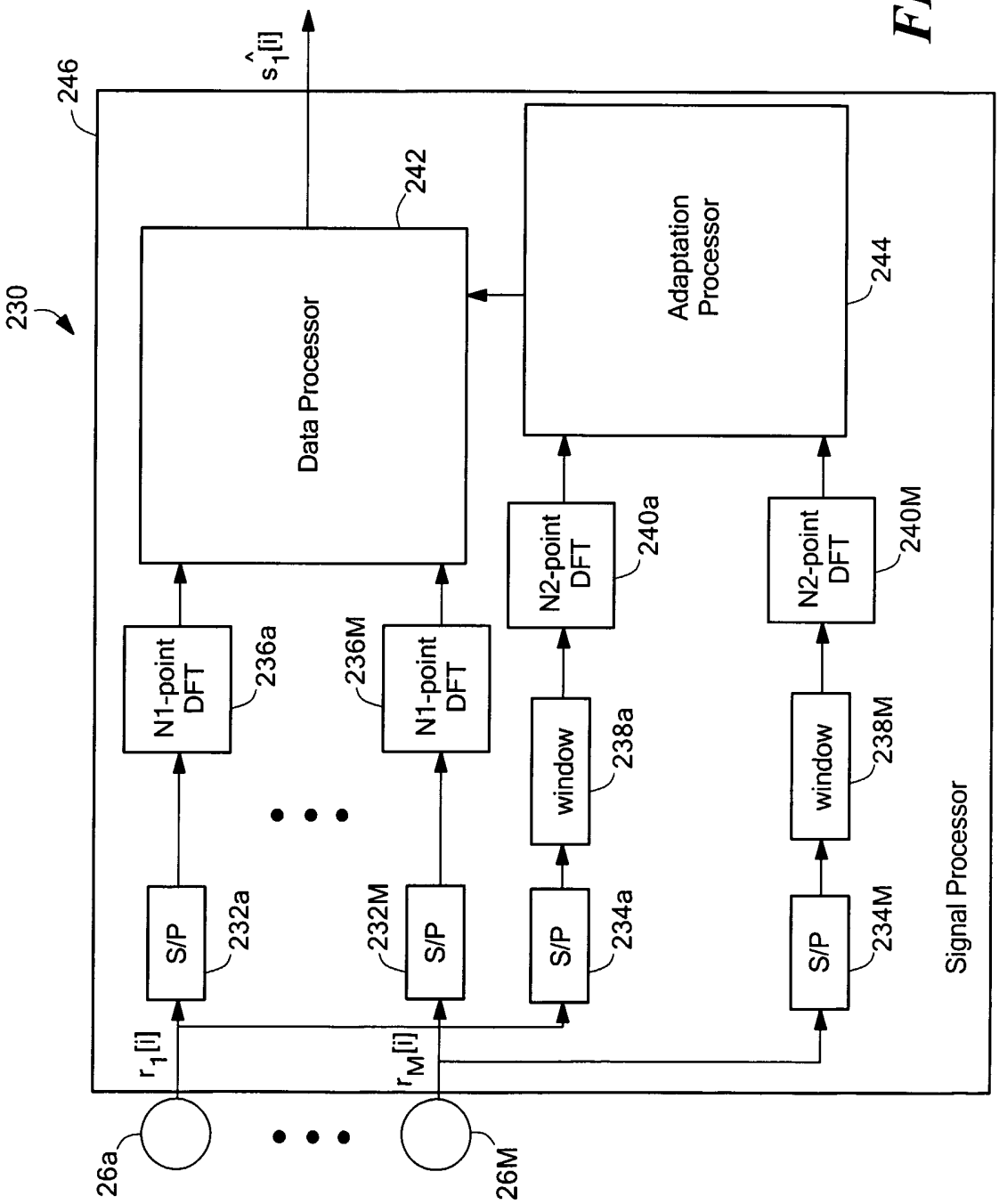

230
246
$\hat{s}_1[i]$
Data Processor
242
Adaptation Processor
244
N1-point DFT
236a
N1-point DFT
236M
N2-point DFT
240a
N2-point DFT
240M
window
238a
window
238M
S/P
232a
S/P
232M
S/P
234a
S/P
234M
$r_1[i]$
$r_M[i]$
26a
26M
Signal Processor

FIG. 11