

System and Apparatus for Speech Communication
and Speech recognition

5 Field of the invention

The present invention relates to a system and apparatus for speech communication and speech recognition. It further relates to signal processing methods which can be implemented in the system.

10

Background of the Invention

The present applicant's PCT application PCT/SG99/00119, the disclosure of which is incorporated herein by reference in its entirety, proposes a method of processing signals in which signals received from an array of sensors are subject to a first adaptive filter arranged to enhance a target signal, followed by a second adaptive filter arranged to suppress unwanted signals. The output of the second filter is converted into the frequency domain, and further digital processing is performed in that domain.

20

The present invention seeks to provide a headset system performing improved signal processing of audio signals and suitable for speech communication.

25 The present invention further seeks to provide signal processing methods and apparatus suitable for use in a speech communication and/or speech recognition system.

Summary of the Invention

In general terms, a first aspect of the present invention proposes a headset system including a base unit and a headset unit to be worn by a user (e.g.
5 resting on the user's head or around the user's shoulders) and having a plurality of microphones, the headset unit and base unit being in mutual wireless communication, and at least one of the base unit and the headset unit having digital signal processing means arranged to perform signal processing in the time domain on audio signals generated by the
10 microphones, the signal processing means including at least one adaptive filter to enhance a wanted signal in the audio signals and at least one adaptive filter to reduce an unwanted signal in the audio signals.

Preferably the digital signal processing means are part of the headset unit.

15 The headset can be used for communication with the base unit, and optionally with other individuals, especially via the base unit. The headset system may comprise, or be in communication with, a speech recognition engine for recognizing speech of the user wearing the headset unit.

20 Although the signal processing may be as described in PCT/SG99/00119, more preferably, the signal processing is modified to distinguish between the noise and interference signals. Signals received from the microphones (array of sensors) are processed using a first adaptive filter to enhance a target
25 signal, and then divided and supplied to a second adaptive filter arranged to reduce interference signals and a third filter arranged to reduce noise. The outputs of the second and third filters are combined, and may be subject to further processing in the frequency domain.

In fact, this concept provides a second, independent aspect of the invention which is a method of processing signals received from an array of sensors comprising the steps of sampling and digitising the received signals and processing the digitally converted signals, the processing including:

- 5 filtering the digital signals using a first adaptive filter arranged to enhance a target signal in the digital signals,
 transmitting the output of the first adaptive filter to a second adaptive filter and to a third adaptive filter, the second filter being arranged to suppress unwanted interference signals, and the third filter being arranged to suppress
10 noise signals; and
 combining the outputs of the second and third filters.

The invention further provides signal processing apparatus for performing such a method.

15

Brief Description of the Drawings

An embodiment of the invention will now be described by way of example with reference to the accompanying drawings in which:

20

Fig.1 illustrates a general scenario in which an embodiment of the invention may operate.

Fig.2 is a schematic illustration of a general digital signal processing system which is an embodiment of present invention.

25 Fig.3 is a system level block diagram of the described embodiment of Fig.2.

Fig.4a-d is a flow chart illustrating the operation of the embodiment of Fig.3.

Fig.5 illustrates a typical plot of non-linear energy of a channel and the established thresholds.

Fig.6 (a) illustrates a wave front arriving from 40 degree off-boresight
30 direction.

Fig.6 (b) represents a time delay estimator using an adaptive filter.

Fig.6 (c) shows the impulse response of the filter indicates a wave front from the boresight direction.

Fig.7 shows the response of time delay estimator of the filter indicates an
5 interference signal together with a wave front from the boresight direction.

Fig.8 shows the schematic block diagram of the four channels Adaptive Spatial Filter.

Fig.9 is a response curve of S-shape transfer function (S function).

Fig.10 shows the schematic block diagram of the Adaptive Interference Filter.

10 Fig.11 shows the schematic block diagram of the Adaptive Ambient Noise Estimator.

Fig.12 is a block diagram of Adaptive Signal Multiplexer.

Fig.13 shows an input signal buffer.

Fig.14 shows the use of a Hanning Window on overlapping blocks of signals.

15 Fig.15 illustrates a sudden rise of noise level of the nonlinear energy plot.

Fig. 16 illustrates a specific embodiment of the invention schematically.

Fig. 17 illustrates a headset unit which is a component of the embodiment of Fig. 16.

Fig. 18, which is composed of Figs. 18(a) and 18(b), shows two ways of
20 wearing the headset unit of Fig. 17.

Detailed Description of the Embodiment of the Invention

Below, with reference to Figs. 16 and 17, we describe a specific embodiment
25 of the invention. Before that, we describe in detail a digital signal processing technique which may be employed by the invention.

FIG. 1 illustrates schematically the operating environment of a signal processing apparatus 5 of the described embodiment of the invention, shown
30 in a simplified example of a room. A target sound signal "s" emitted from a

source s' in a known direction impinging on a sensor array, such as a microphone array 10 of the apparatus 5, is coupled with other unwanted signals namely interference signals u_1 , u_2 from other sources A, B, reflections of these signals u_{1r} , u_{2r} and the target signal's own reflected signal s_r . These
5 unwanted signals cause interference and degrade the quality of the target signal "s" as received by the sensor array. The actual number of unwanted signals depends on the number of sources and room geometry but only three reflected (echo) paths and three direct paths are illustrated for simplicity of explanation. The sensor array 10 is connected to processing circuitry 20-60
10 and there will be a noise input q associated with the circuitry which further degrades the target signal.

An embodiment of signal processing apparatus 5 is shown in FIG.2. The apparatus observes the environment with an array of four sensors such as
15 microphones 10a-10d. Target and noise/interference sound signals are coupled when impinging on each of the sensors. The signal received by each of the sensors is amplified by an amplifier 20a-d and converted to a digital bitstream using an analogue to digital converter 30a-d. The bit streams are feed in parallel to the digital signal processor 40 to be processed digitally. The
20 processor provides an output signal to a digital to analogue converter 50 which is fed to a line amplifier 60 to provide the final analogue output.

FIG.3 shows the major functional blocks of the digital processor in more detail. The multiple input coupled signals are received by the four-channel
25 microphone array 10a-10d, each of which forms a signal channel, with channel 10a being the reference channel. The received signals are passed to a receiver front end which provides the functions of amplifiers 20 and analogue to digital converters 30 in a single custom chip. The four channel digitized output signals are fed in parallel to the digital signal processor 40.
30 The digital signal processor 40 comprises five sub-processors. They are (a) a

Preliminary Signal Parameters Estimator and Decision Processor 42, (b) a Signal Adaptive Filter 44, (c) an Adaptive Interference Filter 46, (d) an Adaptive Noise Estimation Filter 48, and (e) an Adaptive Interference and Noise Cancellation and Suppression Processor 50. The basic signal flow is from processor 42, to processor 44, to processor 46 and 48, to processor 50. The output of processor 42 is referred to as "stage 1" in this process; the output of processor 44 as "stage 2", and the output of processors 46, 48 as "stage 3". These connections being represented by thick arrows in FIG.3. The filtered signal S is output from processor 50. Decisions necessary for the operation of the processor 40 are generally made by processor 42 which receives information from processors 44-50, makes decisions on the basis of that information and sends instructions to processors 44-50, through connections represented by thin arrows in FIG.3. The outputs I, S of the processor 40 are transmitted to a Speech recognition engine, 52.

It will be appreciated that the splitting of the processor 40 into the five component parts 42, 44, 46, 48 and 50 is essentially notional and is made to assist understanding of the operation of the processor. The processor 40 would in reality be embodied as a single multi-function digital processor performing the functions described under control of a program with suitable memory and other peripherals. Furthermore, the operation of the speech recognition engine 52 also could in principle be incorporated into the operation of the processor 40.

A flowchart illustrating the operation of the processors is shown in FIG 4a-d and this will firstly be described generally. A more detailed explanation of aspects of the processor operation will then follow.

The front end 20,30 processes samples of the signals received from array 10 at a predetermined sampling frequency, for example 16kHz. The processor 42

includes an input buffer 43 that can hold N such samples for each of the four channels. Upon initialization, the apparatus collects a block of $N/2$ new signal samples for all the channels at step 500, so that the buffer holds a block of $N/2$ new samples and a block of $N/2$ previous samples. The processor 42
5 then removes any DC from the new samples and pre-emphasizes or whitens the samples at step 502.

Following this, the total non-linear energy of a stage 1 signal sample E_{r1} and a stage 2 signal sample E_{r3} is calculated at step 504. The samples from the
10 reference channel 10a are used for this purpose although any other channel could be used.

There then follows a short initialization period at step 506 in which the first 20 blocks of $N/2$ samples of signal after start-up are used to estimate a Bark
15 Scale system noise B_n at step 516 and a histogram P_b at step 518. During this short period, an assumption is made that no target signals are present. The updated P_b is then used with updated P_{bs} to estimate the environment noise energy E_n and two detection thresholds, a noise threshold T_{n1} and a larger signal threshold T_{n2} , are calculated by processor 42 from E_n using
20 scaling factors. The routine then moves to point B and point F.

After this initialization period, P_{bs} and B_n are updated when an update condition is fulfilled.

25 At step 508, it is determined if the stage 3 signal energy E_{r3} is greater than the noise threshold T_{n1} . If not, the Bark Scale system noise B_n is updated at step 510. Then, it'll proceed to step 512. If so, the routine will skip step 510 and proceed to step 512. A test is made at step 512 to see if the signal energy E_{r1} is greater than the noise threshold T_{n1} . If so, P_b and P_{bs} are estimated at step
30 518 for computing E_n , T_{n1} and T_{n2} . The routine then moves to point B and

point F. If not, only P_{bs} will be updated and it's used with previous P_b to compute E_n , T_{n1} and T_{n2} at step 514. T_{n1} and T_{n2} will follow the environment noise level closely. The histogram is used to determine if the signal energy level shows a steady state increase which would indicate an increase in noise, since the speech target signal will show considerable variation over time and thus can be distinguished. This is illustrated in FIG.15 in which a signal noise level rises from an initial level to a new level which exceeds both thresholds.

10 A test is made at step 520 to see if the estimated energy E_{r1} in the reference channel 10a exceeds the second threshold T_{n2} . If so, a counter C_L is reset and a candidate target signal is deemed to be present. The apparatus only wishes to process candidate target signals that impinge on the array 10 from a known direction normal to the array, hereinafter referred to as the boresight direction, or from a limited angular departure there from, in this embodiment plus or minus 15 degrees. Therefore, the next stage is to check for any signal arriving from this direction.

At step 528, three coefficients are established, namely a correlation coefficient C_x , a correlation time delay T_d and a filter coefficient peak ratio P_k which together provide an indication of the direction from which the target signal arrived.

At step 530, three tests are conducted to determine if the candidate target signal is an actual target signal. First, the cross correlation coefficient C_x must exceed a predetermined threshold T_c , second, the size of the delay coefficient must be less than a value θ indicating that the signal has impinged on the array within the predetermined angular range and lastly the filter coefficient peak ratio P_k must exceed a predetermined threshold T_{Pk1} . If these conditions are not met, the signal is not regarded as a target signal and the routine

passes to step 534 (non-target signal filtering). If the conditions are met, the confirmed target signal is fed to step 532 (target signal filtering) of Signal Adaptive Spatial Filter 44.

- 5 If at step 520, the estimated energy E_{r1} in the reference channel 10a is found not to exceed the second threshold T_{n2} , the target signal is considered not to be present and the routine passes to step 534 via steps 522-526 in which the counter C_L is incremented. At step 524, C_L is checked against a threshold T_{CL} . If the threshold is reached, block leak compensation is performed on the filter
- 10 coefficient W_{td} and counter C_L is reset at step 526. This block leak compensation step improves the adaptation speed of the filter coefficient W_{td} to the direction of fast changing target sources and environment. If the threshold is not reached, the program moves to step 534 described below.
- 15 Following step 530, the confirmed target signal is fed to step 532 at the Signal Adaptive Spatial Filter 44. The filter is instructed to perform adaptive filtering at step 532 and 536, in which the filter coefficients W_{su} are adapted to provide a "target signal plus noise" signal in the reference channel and "noise only" signals in the remaining channels using the Least Mean Square (LMS)
- 20 algorithm. In order to prevent the filter coefficient updated wrongly, a running energy ratio R_{sd} is computed at every sample at step 532. This running energy ratio R_{sd} is used as a condition to test whether that the filter coefficient corresponding to that particular sample should be updated or not. The filter 44 output channel equivalent to the reference channel is for convenience referred
- 25 to as the Sum Channel and the filter 44 output from the other channels, Difference Channels. The signal so processed will be, for convenience, referred to as A' .

- 30 If the signal is considered to be a noise signal, the routine passes to step 534 in which the signals are passed through filter 44 without the filter coefficients

being adapted, to form the Sum and Difference channel signals. The signals so processed will be referred to for convenience as B'.

5 The effect of the filter 44 is to enhance the signal if this is identified as a target signal but not otherwise.

At step 538, a new filter coefficient peak ratio P_{k2} is calculated based on the filter coefficient W_{su} . At step 539, if the signal is not A' signals from step 532 the routine passes to step 548. Else, the peak ratio calculated at step 538 is
10 compared with a best peak ratio BP_k at step 540. If it is larger than best peak ratio, the value of best peak ratio is replaced by this new peak ratio P_{k2} and all the filter coefficients W_{su} are stored as the best filter coefficients at step 542. If it is not, the peak ratio P_{k2} is again compared with a threshold T_{Pk} at step 544. If the peak ratio is below the threshold, a wrong update on the filter
15 coefficients is deemed to be occurred and the filter coefficients are restored to the previous stored best filter coefficients at step 546. If it is above the threshold, the routine passes to step 548.

At step 548, an energy ratio R_{sd} and power ratio P_{rsd} between the Sum
20 Channel and the Difference Channels are estimated by processor 42. Besides these, two other coefficients are also established, namely an energy ratio factor R_{sdf} and a second stage non-linear signal energy E_{r2} . Following this, the adaptive noise power threshold T_{Prsd} is updated based on the calculated power ratio P_{rsd} .

25

At this point, the signal is divided into two parallel paths namely point C and point D. Following point C, the signal is subject to a further test at step 552 to determine if the noise or interference presence. First, if the signals are A' signals from step 532, the routine passes to step 556. Second, if the
30 estimated energy E_{r2} is found not to exceed the second threshold T_{n2} , the

signal is considered not to be present and the routine passes to step 556. Third, the filter coefficient peak ratio P_{k2} is compared to a threshold T_{Pk2} . If it is higher than threshold, this may indicate that there is a target signal and routine passes to step 556. Lastly, the R_{sd} and P_{rsd} are compared to threshold T_{rsd} and T_{Prsd} respectively. If the ratios are both lower than threshold, this indicates probable noise but if higher, this may indicate that there has been some leakage of the target signal into the Difference channel, indication the presence of a target signal after all. For such target signals, the routine also passes to step 556. For all other non-target signals, the routine passes to step 554.

At step 554-558, the signals are processed by the Adaptive Interference Filter 46, the purpose of which is to reduce the unwanted signals. The filter 46, at step 554 is instructed to perform adaptive filtering on the non-target signals with the intention of adapting the filter coefficients to reducing the unwanted signal in the Sum channel to some small error value e_{c1} . This computed e_{c1} is also fed back to step 554 to prevent signal cancellation cause by wrong updating of filter coefficients.

In the alternative, at step 556, the target signals are fed to the filter 46 but this time, no adaptive filtering takes place, so the Sum and Difference signals pass through the filter.

The output signals from processor 46 are thus the Sum channel signal S_{c1} and filtered Difference signal S_i .

Following point D, the signals will pass through few test conditions at step 560. First, if the signals are A' signals from step 532, the routine passes to step 564. Second, if the signals are classified as non-target signal by step 552 (C' signal), the routine passes to step 564. Third, the R_{sdf} and P_{rsd} are

compared to threshold T_{rsdf} and T_{Prsd} respectively. If the ratios are both lower than threshold, this indicates probable ambient noise signal but if higher, this may indicate that there has been some leakage of the target signal into the Difference channel, indication the presence of a target signal after all. Lastly,
 5 if the estimated energy E_{r2} is found exceeds the first threshold T_{n1} , signals are considered to be present. For such signals, the routine also passes to step 564. For all other ambient noise signals, the routine passes to step 562.

At step 562-566, the signals are processed by the Adaptive Ambient noise
 10 Estimation Filter 48, the purpose of which is to reduce the unwanted ambient noise. The filter 48, at step 562 is instructed to perform adaptive filtering on the ambient noise signals with the intention of adapting the filter coefficients to reducing the unwanted ambient noise in the Sum channel to some small error value e_{c2} .

15

In the alternative, at step 564, the signals are fed to the filter 48 but this time, no adaptive filtering takes place, so the Sum and Difference signals pass through the filter.

20 The output signals from processor 48 are thus the Sum channel signal S_{c2} and filtered Difference signal S_n .

At step 568, output signals from processor 46: S_{c1} and S_i and output signals from processor 48: S_{c2} and S_n are processed by an adaptive signal
 25 multiplexer. Here, those signals are multiplex and a weighted average error signal $e_s(t)$, a sum signal $S_c(t)$ and a weighted average interference signal $I_s(t)$ are produced. These signals are then collected for the new $N/2$ samples and the last $N/2$ samples from the previous block and a Hanning Window H_n is applied to the collected samples as shown in FIG.13 to form vectors S_h , I_h and
 30 E_h . This is an overlapping technique with overlapping vectors S_h , I_h and E_h

being formed from past and present blocks of $N/2$ samples continuously. This is illustrated in FIG.14. A Fast Fourier Transform is then performed on the vectors S_h , I_h and E_h to transform the vectors into frequency domain equivalents S_f , I_f and E_f at step 570.

5

At step 572, a modified spectrum is calculated for the transformed signals to provide "pseudo" spectrum values P_s and P_i .

10 In order to reduce signal distortion due to wrong estimation of the noise spectra, a frequency scanning is performed between P_s and P_i to look for the peaks in the same frequency components at step 574. Attenuation is then performed on those peaks in P_i to reduce the signal cancellation effect. P_s and P_i are then warped into the same Bark Frequency Scale to provide Bark Frequency scaled values B_s and B_i at step 576. At step 578, a voice unvoice
15 detection is performed on B_s and B_i to reduce the signal cancellation on the unvoice signal.

A weighted combination B_y of B_n (through path F) and B_i is then made at step 580 and this is combined with B_s to compute the Bark Scale non-linear gain
20 G_b at step 582.

G_b is then unwrapped to the normal frequency domain to provide a gain value G at step 584 and this is then used at step 586 to compute an output spectrum S_{out} using the signal spectrum S_f and E_f from step 570. This gain-
25 adjusted spectrum suppresses the interference signals, the ambient noise and system noise.

An inverse FFT is then performed on the spectrum S_{out} at step 588 and the output signal is then reconstructed from the overlapping signals using the
30 overlap add procedure at step 590.

Hence, besides providing the Speech Recognition Engine 52 with a processed signal S, the system also provides a set of useful information indicated as I on Fig. 3. This set of information may include any one or more

5 of:

1. The direction of speech signal, T_d (step 528).
2. Signal Energy, E_{r1} (step 504).
3. Noise threshold, T_{n1} & T_{n2} (step 514 and 518).
4. Estimated SINR (signal to interference noise ratio) and SNR
10 (signal to noise ratio), and R_{sd} (step 548).
5. Target speech signal presence, A' (steps 530 and 532)
6. Spectrum of processed speech signal, S_{out} (step 586).
7. Potential speech start and end point.
8. Interference signal spectrum, I_f (step 570).

15

Major steps in the above described flowchart will now be described in more detail.

Non-Linear Energy Estimation (STEPS 504.548)

20

At each stage of adaptive filter, the reference signal is taken at a delay half the tap-size. Thus, the end of two stages adaptive filter, the signal is delayed by $L_{su}/2$ and $L_{uq}/2$. In order for the decision-making mechanism for the different stages to accurately follow these delays, the signal energy
25 calculations are calculated at 3 junctions, resulting in 3 pairs of the signal energy.

The first signal energy is calculated at no delay and is used by the time delay estimation and stage1 Adaptive Spatial Filter.

30

$$E_{r1} = \frac{1}{J-2} \sum_{i=1}^{J-2} x(i)^2 - x(i+1)x(i-1)$$

A.1

The second signal energy is calculated at a delay of half of Adaptive Spatial Filter tap-size, $L_{su}/2$.

$$E_{r2} = \frac{1}{J-2} \sum_{i=-L_{su}/2}^{J-L_{su}/2-2} x(i)^2 - x(i+1)x(i-1) \quad A.2$$

The last signal energy is calculated at a delay of $L_{su}/2 + L_{uq}/2$ and is used by noise updating.

$$E_{r3} = \frac{1}{J-2} \sum_{i=-(L_{su}/2+L_{uq}/2)}^{J-(L_{su}/2+L_{uq}/2)-2} x(i)^2 - x(i+1)x(i-1) \quad A.3$$

These delays are implemented by means of buffering.

15 Threshold Estimation and Updating (STEPS 514.518)

The processor 42 estimates two thresholds T_{n1} and T_{n2} based on a statistical approach. Two sets of histogram, referred to as Pb and Pbs , are computed in the same way, except that Pbs is computed every block of $N/2$ samples and Pb is computed only on the first 20 blocks of $N/2$ samples or when $E_{r1} < T_{n1}$ which means that there is neither a target signal nor an interference signal is present. E_{r1} is used as the input sample of the histograms, and the length of the histograms is a number M (which may for example be 24). Each histogram is as found from the following equation:

25

$$H_i = \alpha H_i + (1-\alpha)\delta(i-D).I_{24 \times 1} \quad B.1$$

Where H_i stands for either of Pb and Pb, and has the form:

$$H_i = \begin{bmatrix} h(1) \\ h(2) \\ \vdots \\ h(i) \\ \vdots \\ h(24) \end{bmatrix} \quad \text{B.2}$$

$$I_{24 \times 1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{B.3}$$

5

$$\delta(i-D) = \begin{cases} 0, & i \neq D \\ 1, & i = D \end{cases} \quad \text{B.4}$$

Thus, α is a forgetting factor. For Pb, α is chosen empirically to be 0.9988 and for Pbs, α is equal to 0.9688.

10

The value of D which is used in Equation B1 is determined using table 1 below. Specifically, we find the value of Emax in table 1 which is lowest but which is above the input sample E_{r1} , and the corresponding D is used in Equation B.1. Thus, each D labels a corresponding band of values for E_{r1} . For example, if E_{r1} is 412, this the band up to Emax= 424, i.e. the range corresponding to D=13, and accordingly D=13 is used in Equation B.1. Thus, if E_{r1} continues to stay at a certain level, say in the band up to Emax(D), the weight of the corresponding D value in the histogram will start to build up to become the maximum. It indicates that the current running average noise level is approximately Emax(D).

20

D	E _{max} (D)	D	E _{max} (D)	D	E _{max} (D)
1	10	9	110	17	1624
2	11	10	115	18	2274
3	15	11	216	19	3181
4	21	12	303	20	4452
5	29	13	424	21	6232
6	40	14	593	22	8724
7	56	15	829	23	12199
8	79	16	1161	24	17686

Table 1

After computing Pb and Pbs, the peak values of Pb and Pbs are labelled pp
 5 and pps respectively. pp is reset to be equal to (pps – 5) if (pps – pp) > 5.

Below is the pseudo-C which uses pp to estimate T_{n1} and T_{n2} :

```

  Np = Emax[pp];
10  Rpp =  $E_{r1} / (E_{r1} + Np)$ ;
  gamma = sfun(Rpp, 0, 0.8);
   $E_p = \text{gamma} * E_p + (1 - \text{gamma}) * E_{r1}$ ;
  if ( $E_n \geq E_p$ )
     $E_n = 0.7 * E_n + 0.3 * E_p$ ;
15  else if ( $E_{r1} \leq E_{r\_old}$ )
    {
       $E_n = 0.9995 * E_n + 0.0005 * E_p$ ;
       $E_{r\_old} = E_{r1}$ ;
    }
20  else
     $E_n = 0.995 * E_n + 0.005 * E_p$ ;

```

The Emax values in table 1 were chosen experimentally based on a statistical
 method. Samples (in this case, E_{r1}) were collected under certain environments
 25 (office, car, super-market, etc) and a histogram was generated based on the

collected samples. From the histogram, a probability density function is computed and from there the Emax values were decided.

Similarly, all the factors in the first order recursive filters and the lower, upper
 5 limit of the s-function above are chosen empirically. Once the noise energy E_n is obtained, the two signal detection thresholds T_{n1} and T_{n2} are established as follows:

$$T_{n1} = \delta_1 E_n \quad \text{B.5}$$

$$10 \quad T_{n2} = \delta_2 E_n \quad \text{B.6}$$

δ_1 and δ_2 are scalar values that are used to select the thresholds so as to optimize signal detection and minimize false signal detection. As shown in FIG.5, T_{n1} should be above the system noise level, with T_{n2} sufficient to be
 15 generally breached by the potential target signal. These factors may be found by trial and error. In this embodiment, $\delta_1 = 1.375$ and $\delta_2 = 1.675$ have been found to give good results.

In comparison to the algorithms for setting T_{n1} and T_{n2} in PCT/SG99/00119,
 20 the noise level can be tracked more robustly yet faster. A further motivation for the above algorithm for finding the thresholds is to distinguish between signal and noise in all environments, especially noisy environments (car, supermarket, etc.). This means that the user can use the embodiment any where.

25

Time Delay Estimation (T_d) (STEP 528)

FIG 6A illustrates a single wave front impinging on the sensor array. The wave front impinges on sensor 10d first (A as shown) and at a later time impinges on
 30 sensor 10a (A' as shown), after a time delay t_d . This is because the signal

originates at an angle of 40 degrees from the boresight direction. If the signal originated from the boresight direction, the time delay t_d will have been zero ideally.

- 5 Time delay estimation of performed using a tapped delay line time delay estimator included in the processor 42 which is shown in Fig. 6B. The filter has a delay element 600, having a delay $Z^{-L/2}$, connected to the reference channel 10a and a tapped delay line filter 610 having a filter coefficient W_{td} connected to channel 10d. Delay element 600 provides a delay equal to half of that of the
- 10 tapped delay line filter 610. The outputs from the delay element is $d(k)$ and from filter 610 is $d'(k)$. The Difference of these outputs is taken at element 620 providing an error signal $e(k)$ (where k is a time index used for ease of illustration). The error is fed back to the filter 610. The Least Mean Squares

$$W_{td}(k+1) = W_{td}(k) + 2 \mu_{td} S_{10d}(k) e(k) \quad \dots B.1$$

(LMS) algorithm is used to adapt the filter coefficient W_{td} as follows:

15

$$W_{td}(k+1) = \begin{bmatrix} W_{td}^0(k+1) \\ W_{td}^1(k+1) \\ \vdots \\ W_{td}^{L_d}(k+1) \end{bmatrix} \quad \dots B.2$$

$$S_{10d}(k) = \begin{bmatrix} S_{10d}^0(k) \\ S_{10d}^1(k) \\ \vdots \\ S_{10d}^{L_d}(k) \end{bmatrix} \quad \dots B.3$$

$$\mu_{td} = \frac{\beta_{td}}{\|S_{td}(k)\|} \dots B.6$$

- 5 where β_{td} is a user selected convergence factor $0 < \beta_{td} \leq 2$, $\| \cdot \|$ denoted the norm of a vector, k is a time index, L_o is the filter length.

$$e(k) = d(k) - d'(k) \dots B.4$$

$$d'(k) = W_{td}(k)^T \cdot S_{td}(k) \dots B.5$$

- The impulse response of the tapped delay line filter 620 at the end of the adaptation is shown in Fig. 6c. The impulse response is measured and the
 10 position of the peak or the maximum value of the impulse response relative to origin O gives the time delay T_d between the two sensors which is also the angle of arrival of the signal. In the case shown, the peak lies at the centre indicating that the signal comes from the boresight direction ($T_d=0$). The threshold θ at
 15 step 506 is selected depending upon the assumed possible degree of departure from the boresight direction from which the target signal might come. In this embodiment, θ is equivalent to $\pm 15^\circ$.

Normalized Cross Correlation Estimation C_x (STEP 528)

The normalized crosscorrelation between the reference channel 10a and the most distant channel 10d is calculated as follows:

5

Samples of the signals from the reference channel 10a and channel 10d are buffered into shift registers X and Y where X is of length J samples and Y is of length K samples, where $J > K$, to form two independent vectors X_r and Y_r :

10

$$X_r = \begin{bmatrix} x_r \\ x_r(2) \\ \cdot \\ \cdot \\ \cdot \\ x_r(J) \end{bmatrix} \quad \dots C.1$$

$$Y_r = \begin{bmatrix} y_r \\ y_r(2) \\ \cdot \\ \cdot \\ \cdot \\ y_r(K) \end{bmatrix} \quad \dots C.2$$

15

A time delay between the signals is assumed, and to capture this Difference, J is made greater than K. The Difference is selected based on angle of interest. The normalized cross-correlation is then calculated as follows:

$$C_x(l) = \frac{Y_r^T * X_{rl}}{\|Y_r\| * \|X_{rl}\|} \dots C.3$$

$$\text{Where } X_{rl} = \begin{bmatrix} X_r \\ X_r(l+1) \\ \vdots \\ x_r(K+l-1) \end{bmatrix} \dots C.4$$

Where T represents the transpose of the vector and $\| \cdot \|$ represent the norm of the vector and l is the correlation lag. l is selected to span the delay of interest.

- 5 For a sampling frequency of 16kHz and a spacing between sensors 10a, 10d of 18cm, the lag l is selected to be five samples for an angle of interest of 15°.

The threshold T_c is determined empirically. $T_c = 0.65$ is used in this embodiment.

10

Block Leak compensation LMS for Time Delay Estimation (STEP 526)

In the time delay estimation LMS algorithm, a modified leak compensation form is used. This is simply implemented by:

15

$$W_{ld} = \alpha W_{ld} \quad (\text{where } \alpha = \text{forgetting_factor} \sim 0.98)$$

This leak compensation form has the property of adapting faster to the direction of fast changing sources and environment.

Filter Coefficient Peak Ratio, P_k (STEP 528)

The impulse response of the tapped delay line filter with filter coefficients W_{td} at the end of the adaptation with the present of both signal and interference sources is shown in FIG.7. The filter coefficient W_{td} is as follows:

$$W_{td}(k) = \begin{bmatrix} W_{td}^0(k) \\ W_{td}^1(k) \\ \cdot \\ \cdot \\ W_{td}^{L0}(k) \end{bmatrix}$$

With the present of both signal and interference sources, there will be more than one peak at the tapped delay line filter coefficient. The P_k ratio is calculated as follows:

$$A = \text{Max} W_{td}^n$$

$$\text{where } \frac{L0}{2} - \Delta \leq n \leq \frac{L0}{2} + \Delta$$

$$B = \text{Max} W_{td}^n$$

$$\text{where } 0 \leq n < \frac{L0}{2} - \Delta, \frac{L0}{2} + \Delta < n$$

$$P_k = \frac{A}{A+B}$$

Δ is calculated base on the threshold θ at step 530. In this embodiment, with θ equal to $\pm 15^\circ$, Δ is equivalent to 2. A low P_k ratio indicates the present of strong interference signals over the target signal and a high P_k ratio shows high target signal to interference ratio.

Adaptive Spatial Filter 44 (STEPS 532-536)

FIG.8 shows a block diagram of the Adaptive Linear Spatial Filter 44. The function of the filter is to separate the coupled target interference and noise signals into two types. The first, in a single output channel termed the Sum Channel, is an enhanced target signal having weakened interference and noise i.e. signals not from the target signal direction. The second, in the remaining channels termed Difference Channels, which in the four channel case comprise three separate outputs, aims to comprise interference and noise signals alone.

The objective is to adopt the filter coefficients of filter 44 in such a way so as to enhanced the target signal and output it in the Sum Channel and at the same time eliminate the target signal from the coupled signals and output them into the Difference Channels.

The adaptive filter elements in filter 44 acts as linear spatial prediction filters that predict the signal in the reference channel whenever the target signal is present. The filter stops adapting when the signal is deemed to be absent.

The filter coefficients are updated whenever the conditions of steps are met, namely:

- i. The adaptive threshold detector detects the presence of signal;
- ii. The peak ratio exceeds a certain threshold;
- iii. The running R_{sd} exceeds a certain threshold;

As illustrate in FIG.8, the digitized coupled signal X_0 from sensor 10a is fed through a digital delay element 710 of delay $Z^{-L_{su}/2}$. Digitized coupled signals X_1, X_2, X_3 from sensors 10b, 10c, 10d are fed to respective filter elements

712,4,6. The outputs from elements 710,2,4,6 are summed at Summing element 718, the output from the Summing element 718 being divided by four at the divider element 719 to form the Sum channel output signal. The output from delay element 710 is also subtracted from the outputs of the filters 712,4,6 at respective Difference elements 720,2,4, the output from each Difference element forming a respective Difference channel output signal, which is also fed back to the respective filter 712,4,6. The function of the delay element 710 is to time align the signal from the reference channel 10a with the output from the filters 712,4,6.

10

The filter elements 712,4,6 adapt in parallel using the normalized LMS algorithm given by Equations E.1...E.8 below, the output of the Sum Channel being given by equation E.1 and the output from each Difference Channel being given by equation E.6:

15

$$\hat{S}_c(k) = \frac{\bar{S}(k) + \bar{X}_0(k)}{4} \quad \text{E.1}$$

$$\text{Where:} \quad \bar{S}(k) = \sum_{m=1}^{M-1} \bar{S}_m(k) \quad \text{E.2}$$

20

$$\bar{S}_m(k) = (W_{su}^m(k))^T X_m(k) \quad \text{E.3}$$

Where m is 0,1,2...M-1, the number of channels, in this case 0...3 and ^T denotes the transpose of a vector.

$$X_m(k) = \begin{bmatrix} X_{1m}(k) \\ X_{2m}(k) \\ \vdots \\ X_{LSUm}(k) \end{bmatrix} \quad \text{E.4}$$

$$W_{su}^m(k) = \begin{bmatrix} W_{su1}^m(k) \\ W_{su2}^m(k) \\ \vdots \\ W_{suLSU}^m(k) \end{bmatrix} \quad \text{E.5}$$

5

Where $X_m(k)$ and $W_{su}^m(k)$ are column vectors of dimension $(L_{su} \times 1)$.

The weight $X_m(k)$ is updated using the normalized LMS algorithm as follows:

$$\partial_{cm}(k) = \bar{X}_0(k) - \bar{S}_m(k) \quad \text{E.6}$$

$$W_{su}^m(k+1) = W_{su}^m(k) + 2\mu_{su}^m X_m(k) \partial_{cm}(k) \quad \text{E.7}$$

$$\text{Where: } \mu_{su}^m = \frac{\beta_{su}}{\|X_m(k)\|} \quad \text{E.8}$$

15

and where β_{su} is a user selected convergence factor $0 < \beta_{su} \leq 2$, $\| \cdot \|$ denoted the norm of a vector and k is a time index.

20

Running R_{sd} within Adaptive Spatial Filter (STEP 532)

To prevent filter coefficients being updated wrongly, conditions for updating a block of $N/2$ samples is insufficient. Running R_{sd} is computed every $N/2$ samples and it's being used with other conditions to test whether that particular sample should update or not.

Running R_{sd} is calculated as follows:

$$\overline{rsd} = \frac{\overline{E_{sum}}}{\overline{E_{sum}} + \overline{E_{err}}} \quad \text{F.9}$$

10

Where:

$$\overline{E_{sum}} = 0.98\overline{E_{sum}} + 0.02(abs[(\hat{S}_c(k+1))^2 - \hat{S}_c(k)\hat{S}_c(k+2)]) \quad \text{F.10}$$

15

$$\overline{E_{err}} = 0.98\overline{E_{err}} + 0.02(abs[(\partial_{cm}(k+1))^2 - \partial_{cm}(k)\partial_{cm}(k+2)]) \quad \text{F.11}$$

Adaptive Spatial Filter Coefficient Restoration (STEPS 540-546)

In the events of wrong updating, the coefficients of the filter could adapt to the wrong direction or sources. To reduce the effect, a set of 'best coefficients' is kept and copied to the beam-former coefficients when it is detected to be pointing to a wrong direction, after an update.

Two mechanisms are used for these:

- 25 A set of 'best weight' includes all of the three filter coefficients ($W_{su}^1 - W_{su}^3$). They are saved based on the following conditions:

When there is an update on filter coefficients W_{su} , the calculated P_{k2} ratio is compared with the previous stored B_{Pk} , if it is above the B_{Pk} , this new set of filter coefficients shall become the new set of 'best weight' and current P_{k2} ratio is saved as the new B_{Pk} .

5

A second mechanism is used to decide when the filter coefficients should be restored with the saved set of 'best weights'. This is done when filter coefficients are updated and the calculated P_{k2} ratio is below B_{Pk} and threshold T_{Pk} . In this embodiment, the value of T_{Pk} is equal to 0.65.

10

Calculation of Energy Ratio R_{sd} (STEP 548)

This is performed as follows:

$$\hat{S}_c = \begin{bmatrix} \hat{S}_c(0) \\ \hat{S}_c(1) \\ \cdot \\ \cdot \\ \cdot \\ \hat{S}_c(J-1) \end{bmatrix} \quad \dots F.1$$

$$\hat{D}_c = \begin{bmatrix} \hat{d}_c(0) \\ \hat{d}_c(1) \\ \cdot \\ \cdot \\ \cdot \\ \hat{d}_c(J-1) \end{bmatrix} = \begin{bmatrix} \hat{d}_{c1}(0) \\ \hat{d}_{c1}(1) \\ \cdot \\ \cdot \\ \cdot \\ \hat{d}_{c1}(J-1) \end{bmatrix} + \begin{bmatrix} \hat{d}_{c2}(0) \\ \hat{d}_{c2}(1) \\ \cdot \\ \cdot \\ \cdot \\ \hat{d}_{c2}(J-1) \end{bmatrix} + \begin{bmatrix} \hat{d}_{c3}(0) \\ \hat{d}_{c3}(1) \\ \cdot \\ \cdot \\ \cdot \\ \hat{d}_{c3}(J-1) \end{bmatrix} \quad \dots F.2$$

$J=N/2$, the number of samples, in this embodiment 256.

Where E_{SUM} is the sum channel energy and E_{DIF} is the difference channel

$$E_{SUM} = \frac{1}{J-2} \sum_{j=1}^{J-2} \hat{S}_c(j)^2 - \hat{S}_c(j-1) \hat{S}_c(j+1) \dots F.3$$

energy.

$$E_{DIF} = \frac{1}{3(J-2)} \sum_{j=1}^{J-2} \hat{d}_c(j)^2 = \frac{E_{SUM}}{E_{DIF}} \cdot \hat{d}_c(j-1) \dots F.4$$

- 5 The energy ratio between the Sum Channel and Difference Channel (R_{sd}) must not exceed a predetermined threshold. In the four channel case illustrated here the threshold is determined to be about 1.5.

Calculation of Power Ratio P_{rsd} (STEP 548)

10

This is performed as follows:

15

$$\hat{S}_c = \begin{bmatrix} \hat{S}_c(0) \\ \hat{S}_c(1) \\ \vdots \\ \hat{S}_c(J-1) \end{bmatrix}$$

20

$$\hat{\hat{d}}_c = \begin{bmatrix} \hat{\hat{d}}_c(0) \\ \hat{\hat{d}}_c(1) \\ \vdots \\ \hat{\hat{d}}_c(J-1) \end{bmatrix} = \begin{bmatrix} \hat{\hat{d}}_{c1}(0) \\ \hat{\hat{d}}_{c1}(1) \\ \vdots \\ \hat{\hat{d}}_{c1}(J-1) \end{bmatrix} + \begin{bmatrix} \hat{\hat{d}}_{c2}(0) \\ \hat{\hat{d}}_{c2}(1) \\ \vdots \\ \hat{\hat{d}}_{c2}(J-1) \end{bmatrix} + \begin{bmatrix} \hat{\hat{d}}_{c3}(0) \\ \hat{\hat{d}}_{c3}(1) \\ \vdots \\ \hat{\hat{d}}_{c3}(J-1) \end{bmatrix}$$

$J = N/2$, the number of samples, in this embodiment 128.

Where P_{SUM} is the sum channel power and P_{DIF} is the difference channel
 5 power.

$$P_{SUM} = \frac{1}{J} \sum_{j=0}^{J-1} \hat{S}_c(j)^2$$

$$10 \quad P_{DIF} = \frac{1}{3(J)} \sum_{j=0}^{J-1} \hat{\delta}_c(j)^2$$

$$P_{rsd} = \frac{P_{SUM}}{P_{DIF}}$$

The power ratio between the Sum Channel and Difference Channel must not
 15 exceed a dynamic threshold, T_{Prsd} .

Calculation of Energy Ratio Factor R_{sdf} (STEP 548)

This Energy Ratio Factor R_{sdf} is obtained by passing the R_{sd} to a non-linear S-
 20 shape transfer function as shown in FIG. 9. Certain range of the R_{sd} value can
 be boosted up or suppressed by changing the shape of the transfer function
 using different sets of threshold level, S_L and S_H .

Dynamic Noise Power Threshold Updating T_{Prsd} (STEP 550)

25

This dynamic noise power threshold, T_{Prsd} is updated base on the following
 conditions:

If the reference channel signal energy is more than 700 and power ratio is
 less than 0.45 for 64 consecutive processing blocks,

30

$$T_{Prsd} = \alpha_1 * T_{Prsd} + (1-\alpha_1) * P_{rsd}$$

Else if the reference channel signal energy is less than 700, then

$$T_{Prsd} = \alpha_2 * T_{Prsd} + (1-\alpha_2) * \text{Max_Prsd}$$

In this embodiment, $\alpha_1 = 0.67$, $\alpha_2 = 0.98$ and $\text{Max_Prsd} = 1.3$ have been found to give good results.

5

Adaptive Interference Filter 46 (STEPS 554-558)

FIG.10 shows a schematic block diagram of the Adaptive Interference Filter 46. This filter adapts to interference signal and subtracts it from the Sum Channel so as to derive an output with reduced interference noise.

10

The filter 46 takes outputs from the Sum and Difference Channels of the filter 44 and feeds the Difference Channel Signals in parallel to another set of adaptive filter elements 750,2,4 and feed the Sum Channel signal to a corresponding delay element 756. The outputs from the three filter elements 750,2,4 are subtracted from the output from delay element 756 at Difference element 758 to form an error output e_{c1} , which is fed back to the filter elements 750,2,4. The output from filter 44 is also passed to an Adaptive Signal Multiplexer to mix with filter output from filter 48 and subtract it from the Sum Channel.

15

20

Again, the Least Mean Square algorithm (LMS) is used to adapt the filter coefficients W_{uq} as follows:

$$e_{c1}(k) = \hat{S}_e(k) - S_i(k) \quad (1.1)$$

25

$$\text{Where } S_i(k) = \sum_{m=1}^{M-1} \partial_{cm}(k) \quad \text{and} \quad \partial_{cm}(k) = W_{uq}^m(k)^T \cdot Y^m(k) \quad (1.2)$$

32

$$Y^m(k) = \begin{bmatrix} \partial_{c1m}(k) \\ \partial_{c2m}(k) \\ \vdots \\ \partial_{cL_{uqm}}(k) \end{bmatrix} \quad (1.3)$$

5

$$W_{uq}^m(k+1) = W_{uq}^m(k) + 2\mu_{uq}^m Y^m(k) e_{c1}(k) \quad (1.4)$$

$$\mu_{no}^m = \frac{\beta_{no}}{\|Y^m\| + \|e_{c1}\|} \quad (1.5)$$

- 10 and where β_{uq} is a user select factor $0 < \beta_{uq} \leq 2$ and where m is $0, 1, 2, \dots, M-1$, the number of channels, in this case 0..3.

When only target signal is present and the Interference filter is updated wrongly, the error signal in equation 1.1 will be very large and the norm of Y^m will be very small. Hence, by including norm of error signal $\|e_{c1}\|$ into weight updating μ calculation (equation 1.5), the μ will become very small whenever there is a wrong updating of Interference filter occur. This step help to prevent a wrong updating of weight coefficients of Interference filter and hence reduce the effect of signal cancellation.

20

Adaptive Ambient Noise Estimation Filter 48 (STEPS 562-566)

FIG.11 shows a schematic block diagram of the Adaptive Ambient Noise Estimation Filter 48. This filter adapts to the environment noise and subtracts it from the Sum Channel so as to derive an output with reduced noise.

25

The filter 48 takes outputs from the Sum and Difference Channels of the filter 44 and feeds the Difference Channel Signals in parallel to another set of adaptive filter elements 760,2,4 and feed the Sum Channel signal to a

corresponding delay element 766. The outputs from the three filter elements 760,2,4 are subtracted from the output from delay element 766 at Difference element 768 to form an error output e_{c2} , which is fed back to the filter elements 760,2,4. The output from filter 48 also passed to an Adaptive Signal
 5 Multiplexer to mix with filter output from filter 46 and subtract it from the Sum Channel.

Again, the Least Mean Square algorithm (LMS) is used to adapt the filter
 10 coefficients W_{no} as follows:

$$e_{c2}(k) = \hat{S}_c(k) - S_n(k)$$

$$\text{Where: } S_n(k) = \sum_{m=1}^{M-1} \partial_{cm}(k) \text{ and } \partial_{cm}(k) = W_{no}^m(k)^T \cdot Y^m(k)$$

$$15 \quad Y^m(k) = \begin{bmatrix} \partial_{c1m}(k) \\ \partial_{c2m}(k) \\ \cdot \\ \cdot \\ \cdot \\ \partial_{cLnom}(k) \end{bmatrix}$$

$$20 \quad W_{no}^m(k+1) = W_{no}^m(k) + 2\mu_{no}^m Y^m(k) e_{c2}(k)$$

$$\mu_{no}^m = \frac{\beta_{no}}{\|Y^m\|}$$

25

and where β_{no} is a user select factor $0 < \beta_{no} \leq 2$ and where m is $0, 1, 2, \dots, M-1$, the number of channels, in this case $0 \dots 3$.

Adaptive Signal Multiplexer (STEP 568)

FIG.12 shows a schematic block diagram of the Adaptive Signal Multiplexer. This multiplexer adaptively multiplex the output from interference filter 46 S_i and ambient noise filter 48 S_n to produce two interference signals I_c and I_s as follows:

10

$$I_c(t) = W_{e1}S_i(t) + W_{e2}S_n(t)$$

$$I_s(t) = W_{n1}S_i(t) + W_{n2}S_n(t)$$

The weights (W_{e1} , W_{e2}) and (W_{n1} , W_{n2}) can be changed base on different input signal environment conditions to minimize signal cancellation or improve unwanted signal suppression. In this embodiment, the weights are determined base on the following conditions:

If target signal is detected and updating condition for filter 46 (552) and filter 48 (560) are false then $W_{e1} = 0$, $W_{e2} = 1.0$, $W_{n1} = 0.8$ and $W_{n2} = 1.0$.

Else if no target signal is detected and updating condition for filter 46 (552) is true then $W_{e1} = 1.0$, $W_{e2} = 1.0$, $W_{n1} = 1.0$ and $W_{n2} = 1.0$.

Else if no target signal is detected and updating condition for filter 46 (552) is false and updating condition for filter 48 (560) is true then $W_{e1} = 0$, $W_{e2} = 1.0$, $W_{n1} = 1.0$ and $W_{n2} = 1.0$.

I_c is subtracted from the Sum Channel $\hat{S}_c$ so as to derive an output e_s with reduced noise and interference. This output e_s is almost interference and noise free in an ideal situation. However, in a realistic situation, this cannot be achieved. This will cause signal cancellation that degrades the target signal quality or noise or interference will feed through and this will lead to degradation of the output signal to noise and interference ratio. The signal

cancellation problem is reduced in the described embodiment by use of the Adaptive Spatial Filter 44, which reduces the target signal leakage into the Difference Channel. However, in cases where the signal to noise and interference is very high, some target signal may still leak into these channels.

5

To further reduce the target signal cancellation problem and unwanted signal feed through to the output, the other output signal from Adaptive Signal Multiplexer I_s is fed into the Adaptive Non-Linear Interference and Noise Suppression Processor 50.

10

Adaptive Non-Linear Interference and Noise Suppression Processor 50 (STEPS 570-590)

This processor processes input signals in the frequency domain coupled with the well-known overlap add block-processing technique.

15

$S_c(t)$, $e_s(t)$ and $I_s(t)$ is buffered into a memory as illustrated in FIG.13. The buffer consists of $N/2$ of new samples and $N/2$ of old samples from the previous block.

20

A Hanning Window is then applied to the N samples buffered signals as illustrated in FIG.14 expressed mathematically as follows:

$$S_h = \begin{bmatrix} S_c(t+1) \\ S_c(t+2) \\ \vdots \\ S_c(t+N) \end{bmatrix} \bullet H_n \quad (H.3)$$

25

$$E_h = \begin{bmatrix} e_s(t+1) \\ e_s(t+2) \\ \vdots \\ e_s(t+N) \end{bmatrix} \bullet H_n \quad (\text{H.4})$$

$$I_h = \begin{bmatrix} I_s(t+1) \\ I_s(t+2) \\ \vdots \\ I_s(t+N) \end{bmatrix} \bullet H_n \quad (\text{H.5})$$

- 5 Where (H_n) is a Hanning Window of dimension N , N being the dimension of the buffer. The "dot" denotes point-by-point multiplication of the vectors. T is a time index.

The resultant vectors [S_h], [E_h] and [I_h] are transformed into the frequency domain using Fast Fourier Transform algorithm as illustrated in equation H.6,

- 10 H.7 and H.8 below:

$$S_f = FFT(S_h) \quad (\text{H.6})$$

$$E_f = FFT(E_h) \quad (\text{H.7})$$

- 15

$$I_f = FFT(I_h) \quad (\text{H.8})$$

A modified spectrum is then calculated, which is illustrated in Equations H.9 and H.10:

- 20

$$P_s = |\text{Re}(S_f)| + |\text{Im}(S_f)| + F(S_f) * r_s \quad (\text{H.9})$$

$$P_I = |\text{Re}(I_f)| + |\text{Im}(I_f)| + F(I_f) * r_i \quad (\text{H.10})$$

Where "Re" and "Im" refer to taking the absolute values of the real and imaginary parts, r_s and r_i are scalars and $F(S_f)$ and $F(I_f)$ denotes a function of S_f and I_f respectively.

5

One preferred function F using a power function is shown below in equation H.11 and H.12 where "Conj" denotes the complex conjugate:

$$P_s = |\text{Re}(S_f)| + |\text{Im}(S_f)| + (S_f * \text{conj}(S_f)) * r_s \quad (\text{H.11})$$

10

$$P_i = |\text{Re}(I_f)| + |\text{Im}(I_f)| + (I_f * \text{conj}(I_f)) * r_i \quad (\text{H.12})$$

A second preferred function F using a multiplication function is shown below in equations H.13 and H.14:

15

$$P_s = |\text{Re}(S_f)| + |\text{Im}(S_f)| + |\text{Re}(S_f)| * |\text{Im}(S_f)| * r_s \quad (\text{H.13})$$

$$P_i = |\text{Re}(I_f)| + |\text{Im}(I_f)| + |\text{Re}(I_f)| * |\text{Im}(I_f)| * r_i \quad (\text{H.14})$$

- 20 The values of the scalars (r_s and r_i) control the tradeoff between unwanted signal suppression and signal distortion and may be determined empirically. (r_s and r_i) are calculated as $1/(2^{v_s})$ and $1/(2^{v_i})$ where v_s and v_i are scalars. In this embodiment, $v_s=v_i$ is chosen as 8 giving $r_s = r_i = 1/256$. As v_s, v_i reduce, the amount of suppression will increase.

25

Frequency Scan for similar peak between P_s and P_i .

P_i may contain some of the frequency components of P_s due to the wrong estimation of P_i . Therefore, frequency scanning is applied to both P_s and P_i to

look for the peaks in the same frequency components. For those peaks in P_i is then multiplied by an attenuation factor which is chosen to be 0.1 in this case.

The Spectra (P_s) and (P_i) are warped into (Nb) critical bands using the Bark
 5 Frequency Scale [See Lawrence Rabiner and Bing Hwang Juang, Fundamental of Speech Recognition, Prentice Hall 1993]. The number of Bark critical bands depends on the sampling frequency used. For a sampling of 16kHz, there will be $Nb = 22$ critical bands. The warped Bark Spectrum of (P_s) and (P_i) are denoted as (B_s) and (B_i).

10

Voice Unvoiced Detection and Amplification

This is used to detect voice or unvoiced signal from the Bark critical bands of sum signal and hence reduce the effect of signal cancellation on the unvoiced signal. It is performed as follows:

15

$$B_s = \begin{bmatrix} B_s(0) \\ B_s(1) \\ \vdots \\ B_s(Nb) \end{bmatrix}$$

20

$$V_{sum} = \sum_{n=0}^k B_s(n)$$

where k is the voice band upper cutoff

$$U_{sum} = \sum_{n=l}^{Nb} B_s(n)$$

where l is the unvoiced band lower cutoff

25

$$Unvoice_Ratio = \frac{U_{sum}}{V_{sum}}$$

If $Unvoice_Ratio > Unvoice_Th$

$$B_s(n) = B_s(n) \times A$$

5

where $l \leq n \leq Nb$

In this embodiment, the value of voice band upper cutoff k , unvoiced band lower cutoff l , unvoiced threshold $Unvoice_Th$ and amplification factor A is equal to 16, 18, 10 and 8 respectively.

10 A Bark Spectrum of the system noise and environment noise is similarly computed and is denoted as (B_n) . B_n is first established during system initialization as $B_n = B_s$ and continues to be updated when no target signal is detected (step) by the system i.e. any silence period. B_n is updated as follows:

15 if $((E_{r3} < T_{n1}) \parallel (\text{loop_cnt} < 20))$
 {

if $(E_{r3} < nl1)$

$\alpha = 0.98;$

else

20 $\alpha = 0.90;$

$nl1 = \alpha * nl1 + (1 - \alpha) * E_{r1};$

$B_n = \alpha * B_n + (1 - \alpha) * B_s;$

}

25

Using $(B_s, B_i$ and $B_n)$ a non-linear technique is used to estimate a gain (G_b) as follows:

First the unwanted signal Bark Spectrum is combined with the system noise
 30 Bark Spectrum by using as appropriate weighting function as illustrate in Equation J.1.

$$B_y = \Omega_1 B_i + \Omega_2 B_n \quad (\text{J.1})$$

Ω_1 and Ω_2 are weights whose can be chosen empirically so as to maximize
 5 unwanted signals and noise suppression with minimized signal distortion. In
 this embodiment, $\Omega_1 = 1.0$ and $\Omega_2 = 0.25$.

Following that a post signal to noise ratio is calculated using Equation J.2 and
 J.3 below:

10

$$R_{po} = \frac{B_s}{B_y} \quad (\text{J.2})$$

$$R_{pp} = R_{po} - I_{Nb \times 1} \quad (\text{J.3})$$

The division in equation J.2 means element-by-element division and not
 15 vector division. R_{po} and R_{pp} are column vectors of dimension $(Nb \times 1)$, Nb
 being the dimension of the Bark Scale Critical Frequency Band and $I_{Nb \times 1}$ is a
 column unity vector of dimension $(Nb \times 1)$ as shown below:

$$R_{po} = \begin{bmatrix} r_{po}(1) \\ r_{po}(2) \\ \vdots \\ r_{po}(Nb) \end{bmatrix} \quad (\text{J.4})$$

20

$$R_{pp} = \begin{bmatrix} r_{pp}(1) \\ r_{pp}(2) \\ \vdots \\ r_{pp}(Nb) \end{bmatrix} \quad (\text{J.5})$$

$$I_{Nb \times 1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (J.6)$$

If any of the r_{pp} elements of R_{pp} are less than zero, they are set equal to zero.

- 5 Using the Decision Direct Approach [see Y. Ephraim and D. Malah: Speech Enhancement Using Optimal Non-Linear Spectrum Amplitude Estimation; Proc. IEEE International Conference Acoustics Speech and Signal Processing (Boston) 1983, pp1118-1121.], the a-priori signal to noise ratio R_{pr} is calculated as follows:

10

$$R_{pr} = (1 - \beta_i) * R_{pp} + \beta_i * \frac{B_o}{B_y} \quad (J.7)$$

- The division in Equation J.7 means element-by-element division. B_o is a column vector of dimension $(Nb \times 1)$ and denotes the output signal Bark Scale
 15 Bark Spectrum from the previous block $B_o = G_b \times B_s$ (See Equation J.15) (B_o initially is zero). R_{pr} is also a column vector of dimension $(Nb \times 1)$. The value of β_i is given in Table 2 below:

i	1	2	3	4	5
β_i	0.01625	0.1225	0.245	0.49	0.98

Table 2

20

The value i is set equal to 1 on the onset of a signal and β_i value is therefore equal to 0.01625. Then the i value will count from 1 to 5 on each new block of

N/2 samples processed and stay at 5 until the signal is off. The i will start from 1 again at the next signal onset and the β_i is taken accordingly.

Instead of β_i being constant, in this embodiment β_i is made variable and starts at a small value at the onset of the signal to prevent suppression of the target signal and increases, preferably exponentially, to smooth R_{pr} .

From this, R_{rr} is calculated as follows:

$$R_{rr} = \frac{R_{pr}}{I_{Nb \times 1} + R_{pr}} \quad (J.8)$$

The division in Equation J.8 is again element-by-element. R_{rr} is a column vector of dimension $(Nb \times 1)$.

From this, L_x is calculated:

$$L_x = R_{rr} \bullet R_{po} \quad (J.9)$$

The value L_x of is limited to π (≈ 3.14). The multiplication is Equation J.9 means element-by-element multiplication. L_x is a column vector of dimension $(Nb \times 1)$ as shown below:

$$L_x = \begin{bmatrix} l_x(1) \\ l_x(2) \\ \vdots \\ l_x(nb) \\ \vdots \\ l_x(Nb) \end{bmatrix} \quad (J.10)$$

A vector L_y of dimension $(Nb \times 1)$ is then defined as:

$$L_y = \begin{bmatrix} l_y(1) \\ l_y(2) \\ \vdots \\ l_y(nb) \\ \vdots \\ l_y(Nb) \end{bmatrix} \quad (J.11)$$

5 Where $nb = 1, 2, \dots, Nb$. Then L_y is given as:

$$l_y(nb) = \exp\left(\frac{E(nb)}{2}\right) \quad (J.12)$$

and

$$10 \quad E(nb) = -0.57722 - \log(l_x(nb)) + l_x(nb) - \frac{(l_x(nb))^2}{4} + \frac{(l_x(nb))^3}{8} - \frac{(l_x(nb))^4}{96} \dots \quad (J.13)$$

$E(nb)$ is truncated to the desired accuracy. L_y can be obtained using a look-up table approach to reduce computational load.

15

Finally, the Gain G_b is calculated as follows:

$$G_b = R_{rr} \bullet L_y \quad (J.14)$$

20 The "dot" again implies element-by-element multiplication. G_b is a column vector of dimension $(Nb \times 1)$ as shown:

$$G_b = \begin{bmatrix} g(1) \\ g(2) \\ \vdots \\ g(nb) \\ \vdots \\ g(Nb) \end{bmatrix} \quad (J.15)$$

As G_b is still in the Bark Frequency Scale, it is then unwrapped back to the normal linear frequency scale of N dimensions. The unwrapped G_b is denoted as G.

The output spectrum with unwanted signal suppression is given as:

$$\bar{S}_f = (1 - R_{sdf}).G \bullet S_f + R_{sdf}.E_f \quad (J.16)$$

10

The “•” again implies element-by-element multiplication. In eqn J.16 if R_{sdf} is high (implying high signal energy to interference energy) the output signal spectrum is weighted more from E_f than the Noise suppression part ($G \bullet S_f$) to prevent signal cancellation caused by the noise suppression part.

15

The recovered time domain signal is given by:

$$\bar{S}_t = \text{Re}(\text{IFFT}(\bar{S}_f)) \quad (J.17)$$

IFFT denotes an Inverse Fast Fourier Transform, with only the Real part of the inverse transform being taken.

Finally, the output time domain signal is obtained by overlap add with the previous block of output signal:

25

$$\hat{S}_i = \begin{bmatrix} \bar{S}_i(1) \\ \bar{S}_i(2) \\ \vdots \\ \bar{S}_i(N/2) \end{bmatrix} + \begin{bmatrix} Z_i(1) \\ Z_i(2) \\ \vdots \\ Z_i(N/2) \end{bmatrix} \quad (\text{J.18})$$

Where:

$$Z_i = \begin{bmatrix} \bar{S}_{i-1}(1+N/2) \\ \bar{S}_{i-1}(2+N/2) \\ \vdots \\ \bar{S}_{i-1}(N) \end{bmatrix} \quad (\text{J.19})$$

The embodiment described is not to be construed as limitative. For example, there can be any number of channels from two upwards. Furthermore, as will be apparent to one skilled in the art, many steps of the method employed are essentially discrete and may be employed independently of the other steps or in combination with some but not all of the other steps. For example, the adaptive filtering and the frequency domain processing may be performed independently of each other and the frequency domain processing steps such as the use of the modified spectrum, warping into the Bark scale and use of the scaling factor β_i can be viewed as a series of independent tools which need not all be used together.

Turning now to Figs. 16 and 17, an embodiment of the invention is shown which is a headset system. As shown schematically in Fig. 16, the system has two units, namely a base station 71 and a mobile unit 72.

The base unit provides connection to any host system 73 (such as a PC) through a USB (universal serial bus). It acts as a router for streaming audio information between the host system and the mobile unit 72. It is formed with a cradle (not shown) for receiving and holding the mobile unit 72. The cradle is preferably provided with a charging unit co-operating with a rechargeable

power source which is part of the mobile unit 72. The charging unit charges the power source while the mobile unit 72 is held by the cradle.

5 The base unit 71 includes at least one aerial 74 for two-way wireless communication with at least one aerial 75 of the mobile unit 72. The mobile unit includes a loudspeaker 76 (shown physically connected to the mobile unit 72 by a wire, though as explained below, this is not necessary), and at least two microphones (audio sensors) 77. The wireless link between mobile unit 72 and base station 71 is a highly secure RF Bluetooth link.

10

Fig. 17 shows the mobile unit 72 in more detail. It has a structure defining an open loop 78 to be placed around the head or neck of a user, for example so as to be supported on the user's shoulders. At the two ends of the loop are multiple microphones 77 (normally 2 or 4 in total), to be placed in proximity of the user's mouth for receiving voice input. One of more batteries 79 may be provided near the microphones 76. In this case there are two antennas 75 embedded in the structure. Away from the antennas, the loop 78 is covered with RF absorbing material. A rear portion 80 of the loop is a flex-circuit containing digital signal processing and RF circuitry.

20

The system further includes an ear speaker (not shown) magnetically coupled to the mobile unit 72 by components (not shown) provided on the mobile unit 72. The user wears the ear speaker in one of his ears, and it allows audio output from the host system 73. This enables two-way communication applications, such as internet telephony and other speech and audio applications.

25

Preferably, the system includes digital circuitry carrying out a method according to the invention on audio signals received by the multiple

microphones 76. Some or all of the circuitry can be within the circuitry 80 and/or within the base unit 71.

Figures 18(a) and 18(b) show two ways in which a user can wear the mobile unit 72 having the shape illustrated in Fig. 17. In Fig. 18(a) the user wears the mobile unit 72 resting on the top of his head with the microphones close to his mouth. In Fig. 18(b) the user has chosen to wear the mobile unit 72 supported by his shoulders and with the two arms of the loop embracing his neck, again with the microphone close to his mouth.

10

Use of first, second etc. in the claims should only be construed as a means of identification of the integers of the claims, not of process step order. Any novel feature or combination of features disclosed is to be taken as forming an independent invention whether or not specifically claimed in the appendant claims of this application as initially filed.

15

Claims

1. A headset system including a base unit and a headset unit to be worn by a user and having a plurality of microphones, the headset unit and base unit being in mutual wireless communication, and at least one of the base unit and the headset unit having digital signal processing means arranged to perform signal processing in the time domain on audio signals generated by the microphones, the digital signal processing means including at least one adaptive filter to enhance a wanted signal in the audio signals and at least one adaptive filter to reduce an unwanted signal in the audio signals.
2. A headset system according to claim 1 in which the base unit includes a cradle for holding the headset unit.
3. A headset system according to claim 1 or claim 2 in which the headset unit is associated with a loudspeaker operable by the headset unit for generating audio signals to the user.
4. A headset system according to claim 1 in which the digital signal processing means includes:
 - a first adaptive filter arranged to enhance a target signal in the digital signals, and
 - a second adaptive filter and to a third adaptive filter each receiving the output of the first adaptive filter,
 - the second filter being arranged to suppress unwanted interference signals, and the third filter being arranged to suppress noise signals.
5. A headset system according to claim 4 in which the digital processing means is adapted to combine the outputs of the second and third adaptive

filters, convert to the frequency domain and perform further processing in the frequency domain.

6. A headset system according to claim 5 in which an output $S_i(t)$ of the second filter and an output $S_n(t)$ of the third filter are linearly combined using weighting factors to derive two interference signals, a first of the interference signals I_c being subtracted from the output of the first filter, and a second of the interference signals I_s being converted into the frequency domain.
7. A headset system according to any of claims 4 to 6 in which the second and third filter are not adapted if it is determined that a target signal is present.
8. A headset system according to any of claims 4 to 7 in which the second filter is not updated if it is determined that an interference signal is not present.
9. A headset system according to claim 7 or claim 8 further comprising the step of at intervals determining signal energy, and deriving at least one noise threshold from a plurality of values of the signal energy, said determination including determining whether a further signal energy is above the noise threshold.
10. A headset system according to claim 9 in which the derivation of said noise threshold includes using the plurality of signal energy to derive a histogram representing the statistical frequencies of signal energy values in each of a number of bands, and deriving the noise threshold from a signal energy value E_{max} associated with the band having the highest histogram value.

11. A headset system according to any of claims 4 to 10 in which the digital signal processing means comprises a fourth adaptive filter for determining the direction of arrival of the target signal.
- 5 12. A headset system according to claim 11 in which the weights of the fourth adaptive filter are updated including repeatedly performing an update process which attenuates each existing weight value by a forgetting factor α .
- 10 13. A headset system according to claim 11 or 12, in which the digital signal processing means is adapted to determine a ratio P_k indicating the ratio of the highest central weight value A of the fourth adaptive filter to the sum of A and the highest peripheral weight value B, the digital signal processing means only adapting the first filter if the ratio P_k is above a given value T_{Pk1} .
- 15 14. A headset system according to claim 13 in which, following an adaptation of the first filter, the digital signal processing means calculates a new value P_{k2} of the ratio, determines whether the value of P_{k2} is below the previous maximum value of P_{k2} and below a threshold T_{Pk} , and if so restores at least one of the first, second and third filters to its previous state.
- 20 15. A headset system according to claim 13 or claim 14 when dependent on claim 8 in which the determination that an interference signal is not present includes a determination that the value of said ratio is below a threshold T_{Pk2} .
- 25 16. A headset system according to any of claims 4 to 15 in which the weights of the second filter are adapted by a weight updating factor μ which varies inversely with an error output e_{c1} of the second filter.
- 30 17. A headset system according to claim 5 in which the combined signals are transformed into two frequency domain signals which are a desired signal

S_f and an interference signal I_f , S_f and I_f are transformed into respective modified spectra P_s and P_i , the modified spectra are warped into respective Bark spectra B_s and B_i .

- 5 18. A headset system according to claim 17 in which, prior to said warping, frequency scanning is applied to the modified spectra P_s and P_i , and peaks which are found to be common to both are attenuated in P_i .

- 10 19. A headset system according to claim 17 or claim 18 in which a ratio is derived of the sum of the values of B_s over the Bark critical bands up to the voice band upper cutoff, and the sum of the values of B_s over the Bark critical bands at and above the unvoiced band lower cutoff.

- 15 20. A headset system according to claim 16 in which the ratio is above a given threshold, the values of B_s above the unvoiced band lower threshold are amplified.

- 20 21. A headset system according to any preceding claim further including a speech recognition engine receiving the output of the digital signal processing means.

22. A headset system according to claim 18 in which the speech recognition engine receives from the digital signal processing means information indicating any one or more of:

- 25 a) a direction of a target signal T_d ,
 b) a signal Energy E_{r1} ,
 c) a noise threshold used by the digital signal processing means,
 d) an estimated SINR (target signal to interference ratio) and SNR (target signal to noise ratio),
30 e) a signal A' indicating the presence of target speech,

- f) a spectrum of processed speech signal S_{out} ,
- g) potential speech start and end points, and
- h) an interference signal spectrum, I_f (step 570).

5 23. A headset system according to any preceding claim in which the headset unit comprises two arms for location proximate the mouth of the user and for positioning to either side of the user's head.

10 24. A headset system according to claim 23 in which the headset is suitable for positioning supported by the user's shoulders with the arms embracing the user's neck.

25. A headset system according to claim 23 or claim 24 in which at least one microphone is provided on a free end of each of the arms.

15

26. A headset unit for use in the headset system of any preceding claim.

27. A method of processing signals received from an array of sensors comprising the steps of sampling and digitising the received signals and
20 processing the digitally converted signals, the processing including:

filtering the digital signals using a first adaptive filter arranged to enhance a target signal in the digital signals,

transmitting the output of the first adaptive filter to a second adaptive filter and to a third adaptive filter, the second filter being arranged to suppress
25 unwanted interference signals, and the third filter being arranged to suppress noise signals; and

combining the outputs of the second and third filters.

28. Signal processing apparatus arranged to carry out a method according
30 to claim 27.

29. A microphone headset comprising first and second microphones disposed at respective ends of a support, the support being adapted to be worn around the neck or head of a user.