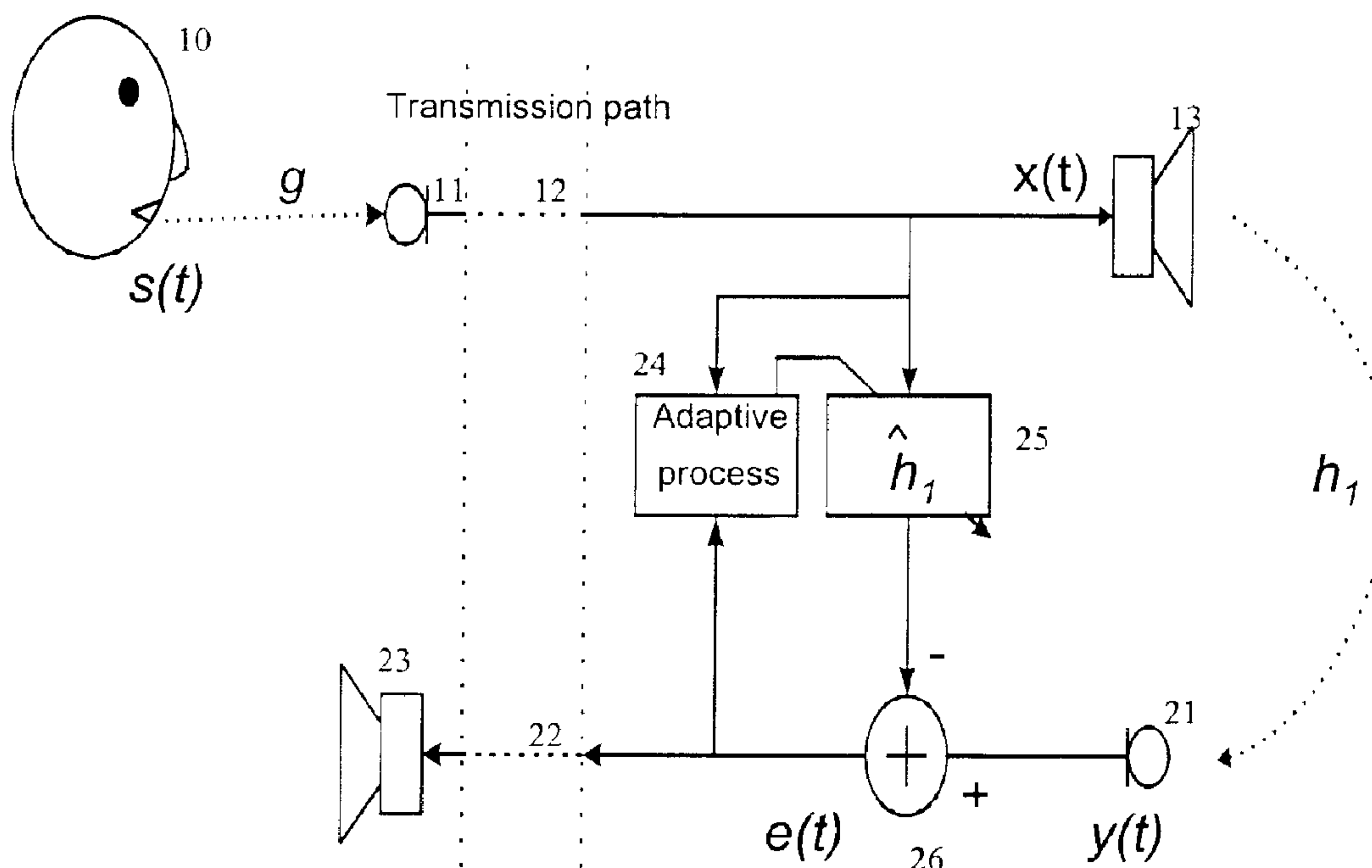




(86) Date de dépôt PCT/PCT Filing Date: 2000/05/18
(87) Date publication PCT/PCT Publication Date: 2000/11/30
(45) Date de délivrance/Issue Date: 2007/03/27
(85) Entrée phase nationale/National Entry: 2001/11/05
(86) N° demande PCT/PCT Application No.: GB 2000/001904
(87) N° publication PCT/PCT Publication No.: 2000/072567
(30) Priorité/Priority: 1999/05/25 (EP99304040.1)

(51) Cl.Int./Int.Cl. *H04M 9/08* (2006.01),
H04B 3/23 (2006.01)
(72) Inventeurs/Inventors:
REED, MARTIN, GB;
HAWKSFORD, MALCOLM JOHN, GB
(73) Propriétaire/Owner:
BRITISH TELECOMMUNICATIONS PUBLIC LIMITED
COMPANY, GB
(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : ANNULATION D'ECHO
(54) Title: ACOUSTIC ECHO CANCELLATION



(57) Abrégé/Abstract:

A multiple channel steered spatialised signal is generated from a signal input (12) modified according to respective spatialisation gain functions (g_1, g_2) to generate a plurality of audio channels ($x_1(t), x_2(t)$). An echo cancellation signal $e(t)$ is applied to a return path (22) using a combined spatialisation and echo path estimate (\hat{h}_1, \hat{h}_2). The estimate is derived from the gain functions applied to the respective channels 14L, 14R. When the gain functions applied in the respective channels are changed, for instance to represent a different apparent position of the sound source (10), a new estimate (\hat{h}_1, \hat{h}_2) of the echo paths (\hat{h}_1, \hat{h}_2) is generated, based on a previous estimate of the echo path and on the new gain functions (g_1, g_2).

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 November 2000 (30.11.2000)

PCT

(10) International Publication Number
WO 00/72567 A1

(51) International Patent Classification⁷: **H04M 9/08**

Engineering, Colchester, CO4 3SQ (GB). **HAWKSFORD, Malcolm, John** [GB/GB]; University of Essex, 40 Broome Grove, Wivenhoe, Colchester, Essex CO7 9QU (GB).

(21) International Application Number: PCT/GB00/01904

(22) International Filing Date: 18 May 2000 (18.05.2000)

(74) Agent: **LIDBETTER, Timothy, Guy, Edwin**; BT Group Legal Services, Intellectual Property Dept., Holborn Centre, 8th floor, 120 Holborn, London EC1N 2TE (GB).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
99304040.1 25 May 1999 (25.05.1999) EP

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, 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, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(71) Applicant (*for all designated States except US*): **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY** [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

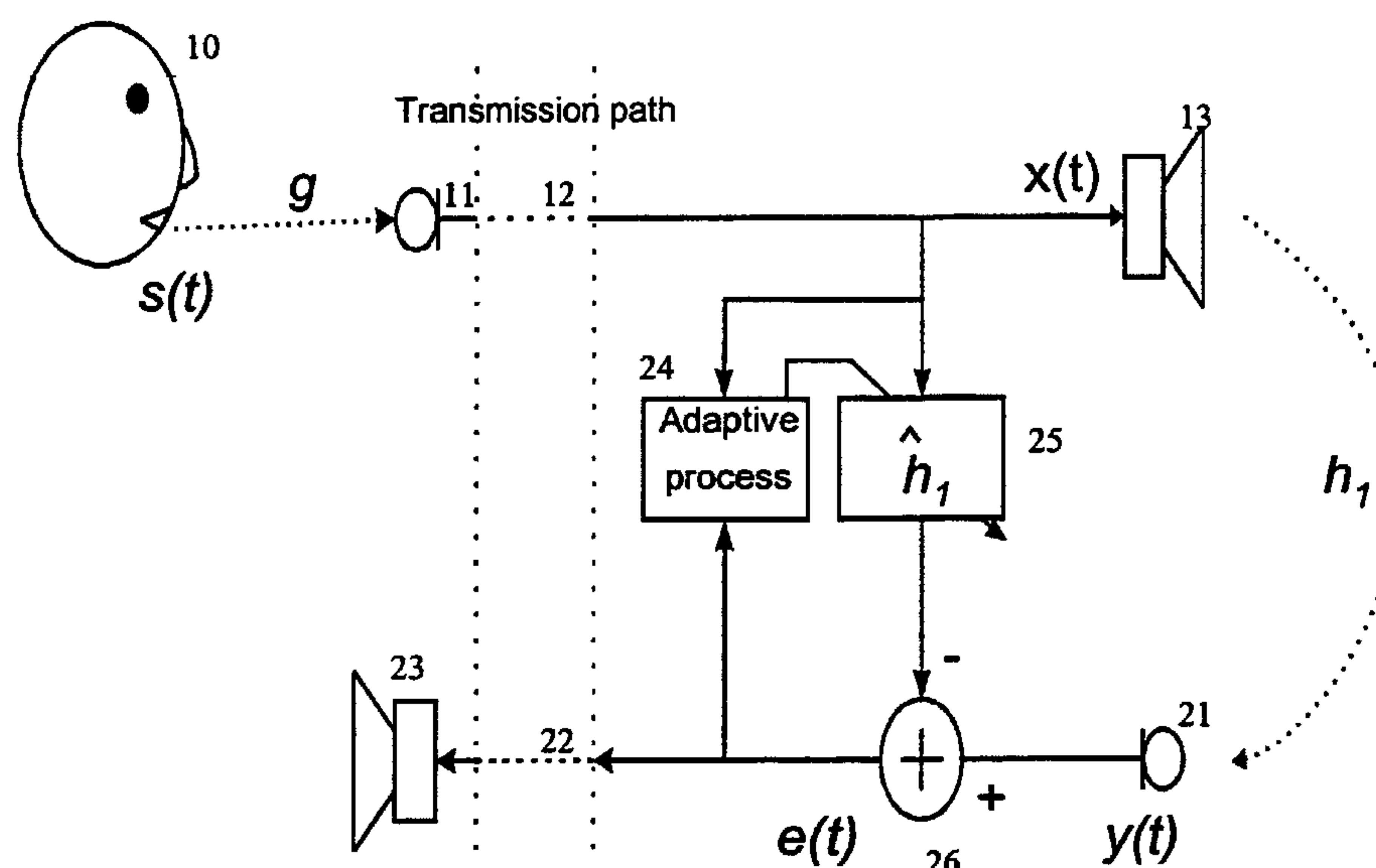
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR,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

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **REED, Martin** [GB/GB]; University of Essex, Dept. Electronic Systems

[Continued on next page]

(54) Title: ACOUSTIC ECHO CANCELLATION



(57) Abstract: A multiple channel steered spatialised signal is generated from a signal input (12) modified according to respective spatialisation gain functions (g_1, g_2) to generate a plurality of audio channels ($x_1(t), x_2(t)$). An echo cancellation signal $e(t)$ is applied to a return path (22) using a combined spatialisation and echo path estimate (\hat{h}_1, \hat{h}_2). The estimate is derived from the gain functions applied to the respective channels 14L, 14R. When the gain functions applied in the respective channels are changed, for instance to represent a different apparent position of the sound source (10), a new estimate (\hat{h}_1, \hat{h}_2) of the echo paths (\hat{h}_1, \hat{h}_2) is generated, based on a previous estimate of the echo path and on the new gain functions (g_1, g_2).

WO 00/72567 A1

WO 00/72567 A1



Published:

— *With international search 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.

ACOUSTIC ECHO CANCELLATION

This invention relates to a method for multiple channel acoustic echo cancellation (AEC), applicable to systems that derive a multi-channel spatialised
5 signal from a monophonic signal, each channel of which is applied to a respective member of an array of loudspeakers at differing gains to give the percept or audible illusion of directionality. This class of spatialised signal will be termed here as steered mono. A steered mono system uses two or more gain elements to represent the spatialisation, which is mapped to a panning processor to generate corresponding
10 loudspeaker outputs. In the embodiments to be described, a two-channel stereophonic signal is used, with two loudspeakers – a system known as “stereo from steered mono” (SSM), but the principles of the invention can be applied to systems with more than two channels. The invention has application in teleconferencing systems where each talker’s voice is artificially given spatial
15 positioning for the benefit of the listener.

For comfortable speech communication in a teleconference system that uses a loudspeaker and microphone, as opposed to a headset, a method of acoustic echo cancellation (AEC) is required. For monophonic systems the topology shown in Figure 1 can be used with a number of different adaptive processes such as least mean
20 square (LMS), recursive least squares (RLS) or fast affine projection (FAP). However, for stereophonic and multiple channel systems, existing solutions are far less advanced with some major obstacles yet to be overcome. The example in Figure 2 shows that for a stereophonic system there are two echo paths, h_1 and h_2 (which include the microphone and loudspeaker impulse responses), compared to the single
25 path in the monophonic case. (This assumes a single microphone is used, which is generally the case when spatialisation is to be created artificially. More generally, the number of echo paths is the product of the number of loudspeakers with the number of microphones).

Existing solutions to the stereo acoustic echo cancellation problem generally
30 assume the system arrangement shown in Figure 2 where the talker-to-microphone path responses are unknown. The aim of the adaptive process in the echo canceller is to use the signals $x_1(t), x_2(t)$ and $e(t)$ to train the adaptive filters \hat{h}_1 and \hat{h}_2 such that

$$e(t) \rightarrow 0 \quad (1)$$

With existing adaptive filter processes it is not possible to achieve a convergent set of filters such that

$$h_1 = \hat{h}_1 \text{ and } h_2 = \hat{h}_2 \quad (2)$$

Instead, a convergent solution such as the following is obtained

$$h_1 * g_1 + h_2 * g_2 = \hat{h}_1 * g_1 + \hat{h}_2 * g_2 \quad (3)$$

where $*$ is the convolution operator. Note that Equation (3) satisfies Equation (1),

5 but that Equation (2) is not a unique solution for Equation (3), so the values for h_1 and h_2 cannot be derived from this result.

If the filters g_1 or g_2 change, possibly due to the talker moving, the equality in Equation (3) no longer holds (unless Equation (2) is also met). Thus, the echo canceller no longer produces a convergent solution and the echo heard by the talker
10 rises in level.

Various solutions to this problem have been proposed that either manipulate the loudspeaker signals, x_1 and x_2 , or use the properties of the signals x_1 and x_2 . The aim of these solutions is to make use of the cross-correlation properties of the two signals as it can be shown that a solution to Equation (2) exists when the two
15 signals are sufficiently decorrelated. However, as the signals x_1 and x_2 are inherently highly correlated in a teleconferencing system, techniques that exploit the small decorrelated features in the signals have poor performance in anything but ideal conditions.

It has been proposed to add a small amount of independent white noise to
20 the signals x_1 and x_2 . It is shown that this significantly aids the convergence of the solution to that in Equation (2) by introducing some signal de-correlation. However, although adding noise in this manner does improve the convergence, the noise has to be added at such a level that it is undesirably audible.

According to the invention, there is provided a method of acoustic echo
25 cancellation for a multiple channel steered spatialised signal, the steered spatialised signal being generated from a signal input modified according to respective spatialisation gain functions to generate a plurality of audio channels, the echo

cancellation process using a combined spatialisation and echo path estimate, the estimate being derived from the gain functions applied to the respective channels, whereby when the gain functions applied in the respective channels are changed, an estimate of the echo paths is generated, the estimate being based on a previous
5 estimate of the echo path and on the gain functions, the echo path estimates being used to generate an echo cancellation signal.

According to another aspect, there is provided apparatus for acoustic echo cancellation in a multiple channel steered spatialised audio system, the spatialised audio system comprising

10 signal input means for receiving an audio signal,
 a plurality of audio output means for generating acoustic signals derived from the audio signal;

 control means associated with the audio output means for generating gain control functions controlling the audio output means such that a spatialised version
15 of the audio signal is generated by the said plurality of audio output means;

 audio input means for detecting acoustic signals;

 signal output means for transmitting a signal derived from the acoustic signals detected by the audio input means;

 echo path estimation means comprising detection means for identifying
20 changes in the gain control functions in the respective control means, and estimation means for generating an estimate of the echo path between the acoustic output means and the acoustic input means, the estimate being based on a previous estimate of the echo path and on the gain control functions detected by the detection means,

25 echo cancellation signal generation means for generating an echo cancellation signal derived from the spatialised audio signals generated by the control means and the estimates derived by the echo path estimation means, and

 signal combination means for applying the echo cancellation signal to the signal generated by the audio input means.

30 This invention is an adaptation of the monophonic LMS process. The invention avoids multiple updates of two or more echo path estimates, such as \hat{h}_1 and \hat{h}_2 , and reduces the number of filter operations required when compared with

existing stereo echo cancellation processes, such as $\hat{h}_1 * x_1$ and $\hat{h}_2 * x_2$. Additionally, the invention uses the spatialisation parameters in the adaptive process, unlike existing stereo echo cancellation processes. The LMS update is modified to take into account the spatialisation parameters that are used to update the aggregated echo path estimate each time the spatialisation changes. After the N th spatialisation change, (where N is the number of channels in the system), the aggregated echo path estimate converges towards the aggregate echo path for future changes in spatialisation. Prior to the N th spatialisation change the process converges to a local solution for the aggregated echo path estimate so that some echo signal reduction is still given in the learning stage of the process.

The learning stage can be made part of a set-up phase prior to use of the system for live traffic. For example, the required number of spatialisation changes can be achieved by operating the monophonic LMS process for each channel in turn, by setting the gains of the other channels to zero.

An embodiment of the invention will now be described, by way of example, with reference to the Figures in which

Figure 1 illustrates the principles of monophonic echo cancellation

Figure 2 illustrates the basic principles of stereophonic echo cancellation (only one half of return path shown)

Figure 3 illustrates an embodiment of the invention

Figure 4 illustrates a theoretical equivalent system to that of Figure 3

Figure 5 is a graphical illustration of the averaged error for the system shown in Figure 3 comparing prior art systems with that of the invention

Figure 6 illustrates the decline over time in the misadjustment in the error path estimates \hat{h}_1 and \hat{h}_2 using the inventive process.

Figure 7 is a flow diagram showing the updating process of the preferred embodiment.

The monophonic system illustrated in Figure 1 comprises an input path 12 connected to the source (microphone 11 and speaker 10) of the original signal $s(t)$. This signal $s(t)$ is modified by the transmission path 12 to generate a loudspeaker signal $x(t)$ which is fed to the loudspeaker 13. The return path consists of a microphone 21, return transmission path 22 and loudspeaker 23. An acoustic path h_1

exists between the loudspeaker 13 of one path and the microphone 21 of the other path. Hence, some sound originating with the speaker 10 will be returned to the loudspeaker 23 and will be heard by the speaker 10 as echo. This effect can be intrusive, especially as the transmission paths 12, 22 can introduce delays, so an
 5 echo cancellation processor 24, 25, 26 is installed to eliminate this signal. An adaptive processor 24 compares the signal $e(t)$ to be transmitted over the return path 22 with the signal $x(t)$ arriving over the input path 12 and identifies correlations between them. This is used to control an adaptive filter bank 25, to generate a vector $\hat{\mathbf{h}}_1$. This vector is an estimate of the acoustic path h_1 . This vector is applied to the
 10 input signal $x(t)$ by the filter and the result subtracted from the input $y(t)$ from the microphone 21 in a combiner 26 to generate a return signal $e(t)$. If the adaptive filter 25 generates an accurate estimate $\hat{\mathbf{h}}_1$ of the function h_1 , the echo signal $y(t)$ should therefore be cancelled out by the output from the adaptive filter 25 and $e(t)$ should therefore be zero. Therefore the only signals transmitted over the return path 22 will
 15 be sounds generated near the microphone 21 and not received from the loudspeaker 13.

Figure 2 illustrates a generalised stereophonic system. The situation is similar to that of Figure 1, except that there are now two loudspeakers and two microphones (indicated by the suffixes "L" and "R" in Figure 2) at each end. Only one
 20 channel 22L of the return path is shown. For each return path 22L, (22R) there are two echo path estimates $\hat{\mathbf{h}}_{1L}, \hat{\mathbf{h}}_{2L}$, one for each outward path 12L, 12R, because each microphone 21L, (21R) is capable of receiving acoustic feedback from both loudspeakers 13L, 13R. Because the two outward paths 12L, 12R are carrying signals from the same source 10, there is significant correlation between the two estimates
 25 $\hat{\mathbf{h}}_{1L}, \hat{\mathbf{h}}_{2L}$ and the processes generating the adaptive filter values 25LL, 25RL are therefore not independent.

In the general case shown in Figure 2, the original functions g_1, g_2 are unknown at the receiving end, being dependant on the relative positions of the microphones 11L, 11R and the speaker 10. It is therefore not possible to derive a
 30 unique solution for $\hat{\mathbf{h}}_{1L}, \hat{\mathbf{h}}_{2L}$. For a given spatialisation the adaptive processor 24L will generate two suitable functions $\hat{\mathbf{h}}_{1L}, \hat{\mathbf{h}}_{2L}$, for which equation (3) above holds:

$$h_1 * g_1 + h_2 * g_2 = \hat{h}_1 * g_1 + \hat{h}_2 * g_2 \quad (3)$$

but this does not necessarily imply that Equation (2) also holds:

$$h_1 = \hat{h}_1 \text{ and } h_2 = \hat{h}_2 \quad (2)$$

If the spatialisation, and hence the functions g_1, g_2 are changed, it will be seen that the adaptive filters must be reset to correspond to the new spatialisation.

In a steered system such as that illustrated in Figures 3 and 4, a monophonic source 11 is used and the functions g_1, g_2 become scalar values g_1, g_2 representing gain. Consequently, there is only one transmit path 12, 22 in each direction. Control information, either generated at the receive end or transmitted with the monophonic signal $s(t)$, is used to control variable gain amplifiers 14L, 14R to vary the gain functions g_1, g_2 defining the spatialisation at the receive end, to emulate a stereophonic source. The fact that the gain functions g_1, g_2 are known at the receive end allows them to be used in the adaptive processor 24, 25.

The operation of the adaptive processor 24, 25 will now be described, with reference to Figure 7. The underlying mathematical theory will then be described, with reference to Figures 3 and 4.

Following an initialisation step 101 in which notional values for the gain functions g_1, g_2 are set, the process runs on an iterative loop for each sampling period n as follows.

Firstly, values k_1, k_2, k_3 are set (step 102). These identify the last three sampling periods at which the spatialisation values g_1, g_2 changed. If the spatialisation gain values g_1, g_2 have not changed since the previous sample $n-1$, the values of k_1, k_2, k_3 are the same as for the previous sample. However, if the values have changed, then k_3 is set to the previous value of k_2 , k_2 is set to the previous value of k_1 , and k_1 is set to $n-1$.

The estimated gain function is then determined (step 103). This is the matrix $\begin{bmatrix} g_1(k_3) & g_2(k_3) \\ g_1(k_2) & g_2(k_2) \end{bmatrix} = \mathbf{G}(k_2, k_3)$. If the spatialisation values g_1, g_2 are unchanged, this matrix is also unchanged and does not need to be recalculated. The inverse of this matrix is then determined.

$$\begin{bmatrix} y_{00} & y_{01} \\ y_{10} & y_{11} \end{bmatrix} = \mathbf{G}^{-1}(k_2, k_3)$$

Again, if the spatialisation values, g_1, g_2 are unchanged, this matrix is also
 5 unchanged and does not need to be recalculated.

Next, (step 104), if $r = n - k_1 + 1$ is less than the number of terms L in the estimated
 echo path vector $\hat{\mathbf{h}}$ (in other words, if the number of samples r elapsed since the
 last spatialisation change is less than L), one term in the estimated echo path vector
 $\hat{\mathbf{h}}_{n-1}$ is amended as follows

10

$$\begin{aligned} \hat{h}_{n-1}(r) = & \left(y_{00} \hat{h}_{k3}(r) + y_{01} \hat{h}_{k2}(r) \right) g_1 \\ & + \left(y_{10} \hat{h}_{k3}(r) + y_{11} \hat{h}_{k2}(r) \right) g_2 \end{aligned}$$

($\hat{\mathbf{h}}_{n-1}$ is the specific instance of the estimated echo path function $\hat{\mathbf{h}}$ from the
 previous iteration). All other terms $\hat{h}_{n-1}(0) \dots \hat{h}_{n-1}(r-1)$ and $\hat{h}_{n-1}(r+1) \dots \hat{h}_{n-1}(L-1)$
 15 remain unchanged

The error cancellation signal $\mathbf{s}_n^T \hat{\mathbf{h}}_{n-1}$ (where \mathbf{s}_n is the vector representing the last L
 samples of the input signal $s(n)$) is then generated using the revised estimate echo
 path vector $\hat{\mathbf{h}}_{n-1}$ (step 105) and subtracted from the signal $y(n)$ to generate the
 20 output signal $e(n)$.

The estimated echo path vector $\hat{\mathbf{h}}_{n-1}$ is then adapted in response to the echo signal
 $e(n)$ (step 106) for use in the next iteration.

$$\hat{\mathbf{h}}_n = \hat{\mathbf{h}}_{n-1} + \mu \mathbf{s}_n \varepsilon(n)$$

where $\varepsilon(n) = \mathbf{s}_n^T \mathbf{s}_n e(n)$ and μ is the step size.

The process is stable provided that the spatialisation changes on a longer timescale than the period L , and that $0 \leq \mu < 2$.

The computational complexity of steps 105-106 in the above process is the same as the normalised LMS process which is of the order $2L$. The number of
 5 computations is of the order of two multiplications and one division for the matrix inversion used in step 103. As this is only performed once after each change in spatialisation it adds little to the complexity of the process for large L . Step 104 is only calculated in the first L samples after a spatialisation change and is insignificant for large L . Thus, when the process shown above is used for acoustic echo
 10 cancellation with a steered mono system, for which it is likely that $L > 100$, the process has a complexity of approximately $2L$.

A mathematical description follows. This will start from the system shown in Figure 4 and then proceed to show that it is equivalent to the system according to the invention shown in Figure 3 which utilises a single aggregated echo path
 15 estimate.

As shown in Figure 4, the two microphone impulse response paths 12L, 12R (Figure 2) are replaced by two artificially generated impulse responses g_1 and g_2 that act upon a single microphone signal $s(t)$ received over the monophonic path 12 to generate the percept or illusion of spatialisation. For the solution presented here the
 20 functions g_1 and g_2 are further simplified by representing them as simple gain functions rather than impulse responses. This system works in a completely different manner from the prior art system of Figure 2 as, rather than trying to decorrelate the signals x_1 and x_2 to help the convergence, the method presented here relies upon the inherent highly correlated nature of the signals x_1 and x_2 . The method uses the
 25 knowledge of the functions g_1 and g_2 (which are unknown in the prior art systems) to achieve a convergent echo cancellation. Before describing an adaptive filter solution to the echo problem it will be demonstrated that a solution to the problem exists.

Let the input to the spatialisation block at sample time n be represented by a column vector $\mathbf{s}_n = [s(n) \ s(n-1) \dots \ s(n-(L-1))]^T$, the input to the listener end microphone by $\mathbf{y}_n = [y(n) \ y(n-1) \dots \ y(n-(L-1))]^T$ and the two loudspeaker-to-microphone echo paths be length L column vectors \mathbf{h}_1 and \mathbf{h}_2 (which incorporate
5 the loudspeaker and microphone impulse responses), then

$$\mathbf{y}_n = g_1(n)\mathbf{S}_n\mathbf{h}_1 + g_2(n)\mathbf{S}_n\mathbf{h}_2 \quad (4)$$

where the spatialisation is represented as the gain values $g_1(n)$ and $g_2(n)$ which are constant over the sample periods $n-(L-1) \dots n$, and $\mathbf{S}_n = [\mathbf{s}_n \ \dots \ \mathbf{s}_{n-(L-1)}]^T$ (This is a “Toeplitz” matrix, that is, a symmetrical matrix of order $L \times L$, having the terms of \mathbf{s}_n in the first row and the first column, the terms of \mathbf{s}_{n-1} in the second row and column,
10 and so on). It can be shown that \mathbf{h}_1 and \mathbf{h}_2 cannot be solved from Equation (4).

However, now consider using a second set of input and output observations at sample time $n+a$ where $a \geq L$ and

$$\begin{aligned} g_1(n+a) &= g_1(n+a-1) = \dots = g_1(n+1) \neq g_1(n) \\ g_2(n+a) &= g_2(n+a-1) = \dots = g_2(n+1) \neq g_2(n) \end{aligned} \quad (5)$$

in other words the functions g_1 and g_2 have changed between sample time n and sample time $n+1$, but then remained unchanged between time $n+1$ and sample time $n+a$.

Hence,

$$\begin{bmatrix} \mathbf{y}_n \\ \mathbf{y}_{n+a} \end{bmatrix} = (\mathbf{G}_{n,n+a} \otimes \mathbf{I}_L) \begin{bmatrix} \mathbf{S}_n \\ \mathbf{S}_{n+a} \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} \quad (6)$$

where \mathbf{I}_L is the $L \times L$ identity matrix,

$$\mathbf{G}_{n,n+a} = \begin{bmatrix} g_1(n) & g_2(n) \\ g_1(n+a) & g_2(n+a) \end{bmatrix} \quad (7)$$

and \otimes is the Kronecker product. (The Kronecker product of two matrices A and B is given by multiplying matrix B separately by each individual term in matrix A and forming a new matrix, (whose order is the product of the original two matrices) with the resulting terms).

The solution to Equation (6) is
$$\begin{bmatrix} \mathbf{S}_n \\ \mathbf{S}_{n+a} \end{bmatrix}^{-1} (\mathbf{G}_{n,n+a} \otimes \mathbf{I}_L)^{-1} \begin{bmatrix} \mathbf{y}_n \\ \mathbf{y}_{n+a} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} \quad (8)$$

and using Kronecker product identities

$$\begin{bmatrix} \mathbf{S}_n \\ \mathbf{S}_{n+a} \end{bmatrix}^{-1} (\mathbf{G}_{n,n+a}^{-1} \otimes \mathbf{I}_L) \begin{bmatrix} \mathbf{y}_n \\ \mathbf{y}_{n+a} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} \quad (9)$$

Thus, a solution for \mathbf{h}_1 and \mathbf{h}_2 exists if the signal s is persistently exciting (*i.e.* it has a full spectral content) and the matrix $\mathbf{G}_{n,n+a}$ is non-singular, that is, it has an inverse
5 matrix. The non-singular condition for $\mathbf{G}_{n,n+a}$ is met if the spatialisation values at sample times n and $n+a$ are different and not scalar multiples of each other (*i.e.* $g_1(n)/g_2(n) \neq g_1(n+a)/g_2(n+a)$). Ideally the values should be sufficiently different such that the solution of Equation (9) is well conditioned.

Having established that a solution exists the adaptive process for the solution
10 is now derived from the LMS process. The normalised LMS (NLMS) process is used to perform monophonic echo cancellation as discussed with reference to Figure 1 using the following updates

$$e(n) = y(n) - \mathbf{s}_n^T \hat{\mathbf{h}}_{n-1} \quad (10)$$

$$\varepsilon(n) = \mathbf{s}_n^T \mathbf{s}_n e(n) \quad (11)$$

$$\hat{\mathbf{h}}_n = \hat{\mathbf{h}}_{n-1} + \mu \mathbf{s}_n \varepsilon(n) \quad (12)$$

where $e(n)$ is the echo signal, μ is the step size parameter and $\hat{\mathbf{h}}_n$ is the echo path
15 estimate at the n th sample instance. The single channel normalised LMS equations above can be modified for the steered mono case by using a single aggregate echo path estimate and redefining $\hat{\mathbf{h}}_{n-1}$ as

$$\hat{\mathbf{h}}_{n-1}^T = \left[\hat{h}_1(0)g_1(n) + \hat{h}_2(0)g_2(n), \dots, \right. \\ \left. \hat{h}_1(L-1)g_1(n-(L-1)) + \hat{h}_2(L-1)g_2(n-(L-1)) \right] \quad (13)$$

where $\hat{h}_1(t)$ and $\hat{h}_2(t)$ are functions representing the two echo path estimates at
sample interval n . Likewise define \mathbf{h} as the combination of $h_1(t)$ and $h_2(t)$ in a form
20 equivalent to that shown in Equation (13)

$$\mathbf{h}^T = [h_1(0)g_1(n) + h_2(0)g_2(n), \dots, h_1(L-1)g_1(n-(L-1)) + h_2(L-1)g_2(n-(L-1))] \quad (14)$$

The task is then to use and update $\hat{\mathbf{h}}$ such that the normalised LMS updates of equations 10, 11 and 12 are used for the echo cancellation rather than using two echo path estimates explicitly. If the values of $g_1(n)$ and $g_2(n)$ are constant for all n then the updates in Equation 10, 11 and 12 can be used unchanged to determine an estimate of \mathbf{h} , as \mathbf{h} is constant over time. However, if $g_1(n)$ and $g_2(n)$ change over time then this solution can not be used as a change in \mathbf{h} is not taken into account in the LMS updates of Equations 10, 11 and 12.

Consider three sample epochs $i, i-a$ and $i-b$ where

$$b \gg L > a \quad (15)$$

and

$$\begin{aligned} g_1(i) &= g_1(i-1) = \dots = g_1(i-a) \neq g_1(i-a-1) = g_1(i-a-2) = \dots = g_1(i-b) \\ g_1(i-b) &\neq g_1(i-b-1) = g_1(i-b-2) = \dots = g_1(i-\infty) \end{aligned} \quad (16)$$

and likewise for $g_2(n)$, i.e. values of $g_1(n)$ and $g_2(n)$ change only on the epochs $i-a$ and $i-b$.

Consider the value of the j th coefficient in the combined echo path at the epochs $i-a-1$ and $i-b-1$ (i.e. just prior to the spatialisation changes) which from Equations 14 and 16 is given by

$$h_{i-b-1}(j) = h_1(j)g_1(i-b-1) + h_2(j)g_2(i-b-1) \quad (17)$$

$$h_{i-a-1}(j) = h_1(j)g_1(i-a-1) + h_2(j)g_2(i-a-1) \quad (18)$$

15

Equations 17 and 18 can be expressed as

$$\begin{bmatrix} h_{i-b-1}(j) \\ h_{i-a-1}(j) \end{bmatrix} = \begin{bmatrix} g_1(i-b-1) & g_2(i-b-1) \\ g_1(i-a-1) & g_2(i-a-1) \end{bmatrix} \begin{bmatrix} h_1(j) \\ h_2(j) \end{bmatrix} \quad (19)$$

and thus

$$\mathbf{G}_{i-b-1, i-a-1}^{-1} \begin{bmatrix} h_{i-b-1}(j) \\ h_{i-a-1}(j) \end{bmatrix} = \begin{bmatrix} h_1(j) \\ h_2(j) \end{bmatrix} \quad (20)$$

using the definition of \mathbf{G} from (7).

Further consider the value of the j th coefficient in the combined echo path at the epoch i which, from (14) and (16) is given by

$$h_i(j) = \begin{cases} h_{i-1}(j) & j < i-a \\ h_1(j)g_1(i) + h_2(j)g_2(i) & j = i-a \\ h_{i-1}(j) & j > i-a \end{cases} \quad (21)$$

If the elements of \mathbf{G}^{-1} are defined by a variable γ such that

$$\mathbf{G}_i^{-1} = \begin{bmatrix} \gamma_{00} & \gamma_{01} \\ \gamma_{10} & \gamma_{11} \end{bmatrix} \quad (22)$$

then from (19), (21) and (22)

$$h_i(j) = (\gamma_{00} h_{i-b-1}(j) + \gamma_{01} h_{i-a-1}(j))g_1(i) + (\gamma_{10} h_{i-b-1}(j) + \gamma_{11} h_{i-a-1}(j))g_2(i) \quad \text{(Equation 23)}$$

5

This equation is the additional update required for the normalised LMS update of Equations (10), (11) and (12). Note that from (21) only one coefficient in \mathbf{h} need be updated in each sample period to take account of a spatialisation change.

The process can be extended to a system that has more than two channels, by making a small modification to the process. Specifically for an N - channel system the previous $N+1$ changes in the spatialisation position are recorded in variables k_{N+1}, \dots, k_1 from the least to most recent respectively. The matrix \mathbf{G} is generalised as

$$\mathbf{G} = \begin{bmatrix} g_1(k_{N+1}) & \dots & g_N(k_{N+1}) \\ \vdots & \ddots & \vdots \\ g_1(k_2) & \dots & g_N(k_2) \end{bmatrix} \quad (24)$$

The step 104 may be generalised as

$$\hat{h}(r) = [g_1(n) \quad \dots \quad g_N(n)] \mathbf{G}^{-1} \begin{bmatrix} \hat{h}(r) \\ \vdots \\ \hat{h}(r) \end{bmatrix} \quad (25)$$

The steps 105 and 106 remain unchanged.

To demonstrate the described process both the stereo normalised least mean square process according to the invention and the normalised least mean square process
 5 were simulated using the configuration shown in Figure 3 and with echo paths taken from actual measured microphone/speaker paths. The spatialisation position was changed five times during the simulation to test the ability of the process to adapt to changing spatialisation. Figure 5 shows the ensemble-averaged error in the echo cancellation for a sampled Gaussian white noise (GWN) input. For the simulations the
 10 step-size parameter μ was 0.1, the filter length $L = 200$, the sample rate was 8 kHz and a noise signal was added 40 dB below the input signal. Figure 5 shows that initially the Normalised LMS process (Figure 5a) and the Stereo Normalised LMS process according to the invention (Figure 5b) have identical performance up until sample 40000. Note that the error in both processes falls initially as the adaptive
 15 filters converge to the static value of the aggregate path vector \mathbf{h} . However, at sample period 19000 the spatialisation changes, such that \mathbf{h} changes and consequently the error rises suddenly in both processes. The normalised LMS process (Figure 5a) does not yet have enough information to form the correct aggregate path vector \mathbf{h} as it requires two spatialisation positions to determine the individual path
 20 vectors \mathbf{h}_1 and \mathbf{h}_2 . However, after the second spatialisation position the process according to the invention can make an estimate of the value of the changed \mathbf{h} . This can be observed in Figure 5 as the error does not rise after sample 40000 for the error curve in b) unlike that for the normalised LMS process in a) which demonstrates increases in the error each time the spatialisation changes.

25 The performance of the process according to the invention can also be observed for speech signals in Figure 6 which shows the misadjustment in the two echo path estimate vectors $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$. These vectors are not actually calculated explicitly by the process: however, it does use a solution for them as part of the

update. Equation 20 was used to form the graphs shown in Figure 6. Note that the misadjustment falls in one of the filters and rises in the other over the course of the first spatialisation position, which is between samples 0 and 19000. This is caused by the fact that $g_1(0)=0.2$ and $g_2(0)=0.8$ which means that most of the echo
 5 signal energy is derived from the second filter path such that the aggregate filter estimate $\hat{\mathbf{h}}$ tends towards the second path filter \mathbf{h}_2 . At this point there is no *a priori* information which allows the true solution of \mathbf{h}_1 and \mathbf{h}_2 . However, as soon as the second spatialisation position starts at sample 19000 the misadjustment in both $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$ starts to fall as the Stereo Normalised LMS process allows a solution for \mathbf{h}_1
 10 and \mathbf{h}_2 to be formed.

The process described uses the normalised least mean square adaptive filter to form the update of the combined echo path estimate. However, any current or future adaptive process that updates an estimate of an unknown filter on a sample by sample basis can be used in place of the described normalised least mean square
 15 algorithm. The only modification required is to replace the process step 106 with another filter update. Suitable existing examples are fast affine projection, least mean squares or recursive least mean squares adaptive filters.

CLAIMS

1. A method of acoustic echo cancellation for a multiple channel steered spatialised signal, the steered spatialised signal being generated from a signal input
5 12 modified according to respective spatialisation gain functions (g_1, g_2) to generate a plurality of audio channels ($x_1(t), x_2(t)$), the echo cancellation process using a combined spatialisation and echo path estimate ($\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2$), the estimate being derived from the gain functions applied to the respective channels, whereby when the gain functions applied in the respective channels are changed, an estimate ($\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2$) of the
10 echo paths ($\mathbf{h}_1, \mathbf{h}_2$) is generated, the estimate being based on a previous estimate of the echo path and on the gain functions (g_1, g_2), the echo path estimates ($\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2$) being used to generate an echo cancellation signal $e(t)$.
2. A method according to claim 1, in which the gain functions applied to the
15 respective channels are scalar gain factors
3. A method according to claim 1 or 2 in which a normalised least mean square adaptive filter is used to form the update of the combined echo path estimate.
- 20 4. A method according to claim 1 or 2 in which a fast affine projection adaptive filter is used to form the update of the combined echo path estimate.
5. A method according to claim 1 or 2 in which a recursive least mean squares adaptive filter is used to form the update of the combined echo path estimate.
- 25 6. Apparatus for acoustic echo cancellation in a multiple channel steered spatialised audio system, the spatialised audio system comprising
signal input means (12) for receiving an audio signal $s(t)$,
a plurality of audio output means (13L, 13R) for generating acoustic signals
30 ($x_1(t), x_2(t)$) derived from the audio signal $s(t)$;

control means (14L, 14R) associated with the audio output means (13L, 13R) for generating gain control inputs g_1 , g_2 controlling the audio output means such that a spatialised version h_1 , h_2 of the audio signal $s(t)$ is generated by the said plurality of audio output means (13L, 13R);

5 audio input means (21) for detecting acoustic signals;

signal output means (22) for transmitting a signal $y(t)$ derived from the acoustic signals detected by the audio input means (21);

echo path estimation means (25) comprising detection means for identifying changes in the gain functions g_1 , g_2 in the respective control means (14L, 14R), and
10 estimation means for generating an estimate of the echo path between the acoustic output means 13L, 13R and the acoustic input means (21), the estimate being based on a previous estimate of the echo path and on the gain functions g_1 , g_2 detected by the detection means,

echo cancellation signal generation means (24L, 24R) for generating an echo
15 cancellation signal derived from the spatialised audio signals generated by the control means (14L, 14R) and the estimates derived by the echo path estimation means (25), and

signal combination means (26) for applying the echo cancellation signal to the signal generated by the audio input means.

20

7. A method according to Claim 6, in which the gain functions applied to the respective channels are scalar gain factors

8. A method according to claim 6 or 7 in which a normalised least mean square
25 adaptive filter is used to form the update of the combined echo path estimate.

9. A method according to claim 6 or 7 in which a fast affine projection adaptive filter is used to form the update of the combined echo path estimate.

30 10. A method according to claim 6 or 7 in which a recursive least mean squares adaptive filter is used to form the update of the combined echo path estimate.

1/4

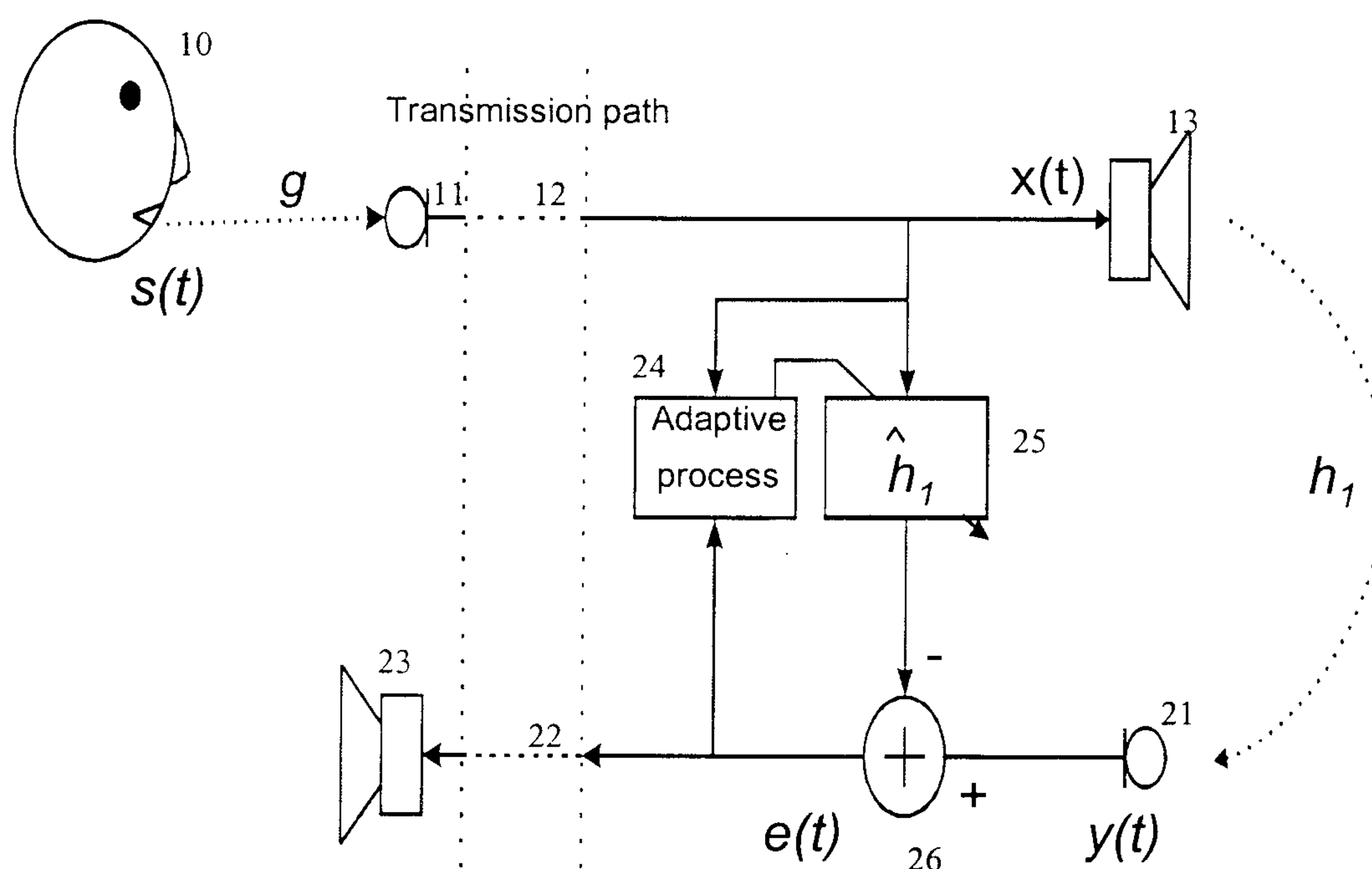


Figure 1

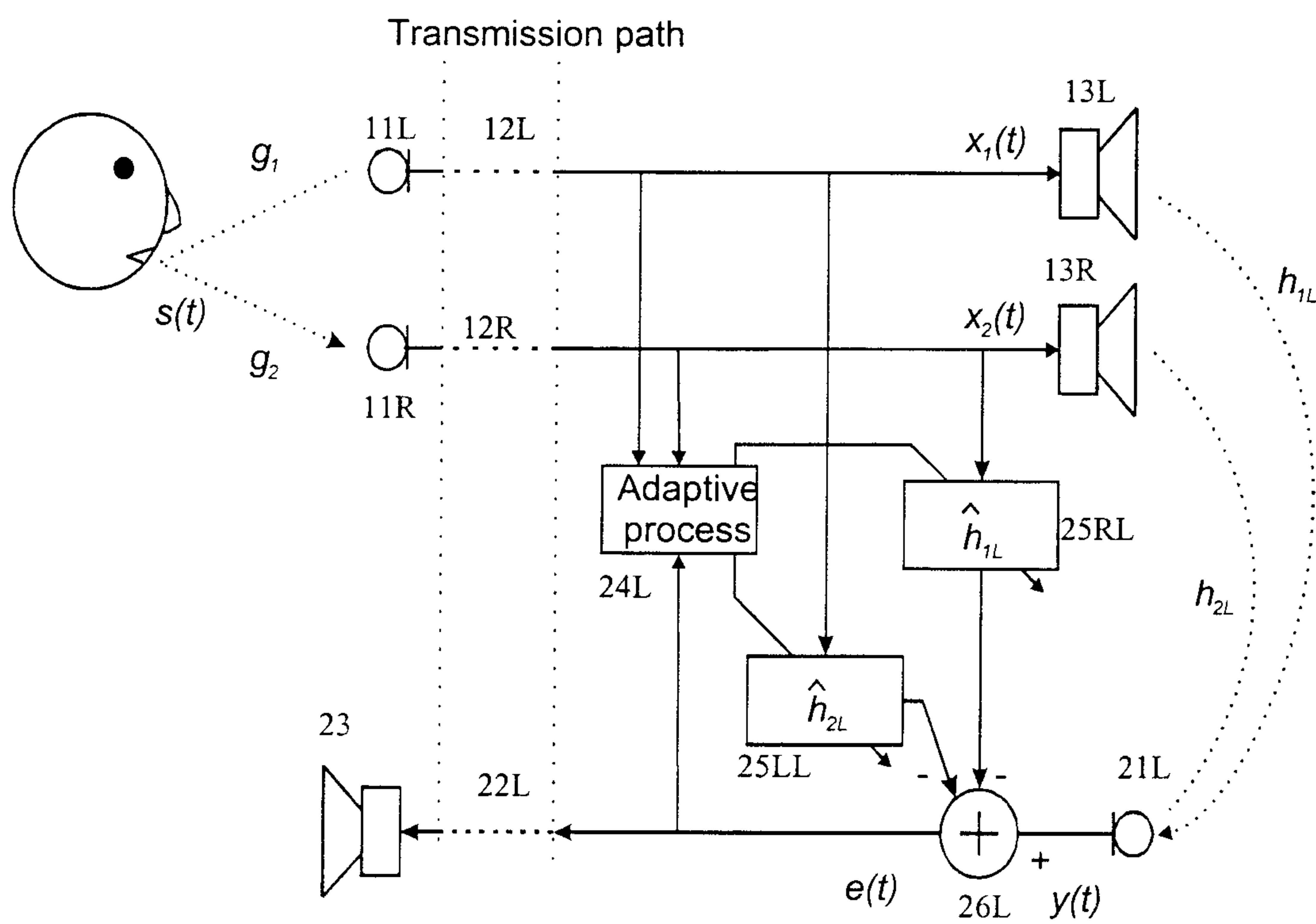


Figure 2

2/4

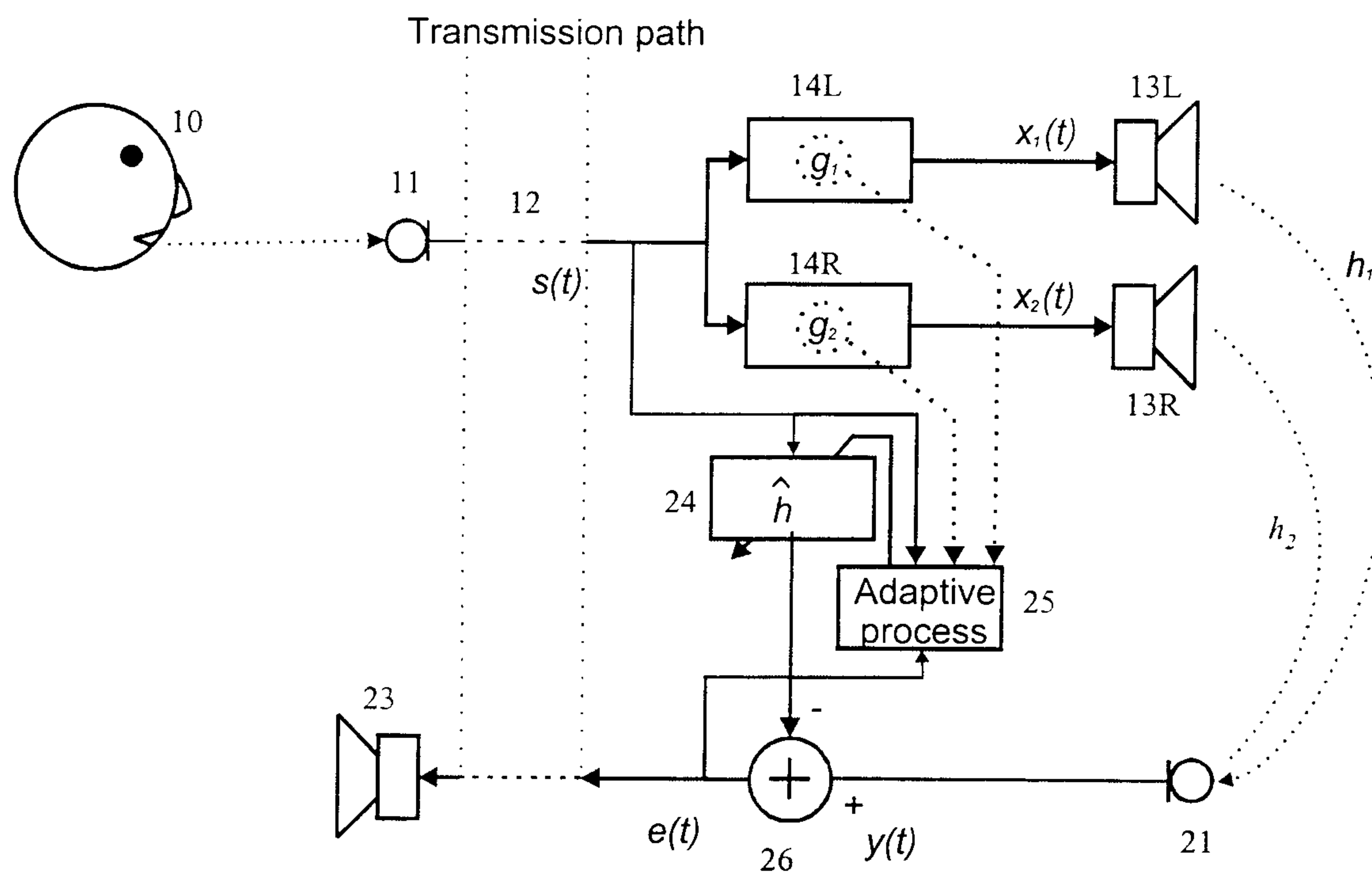


Figure 3

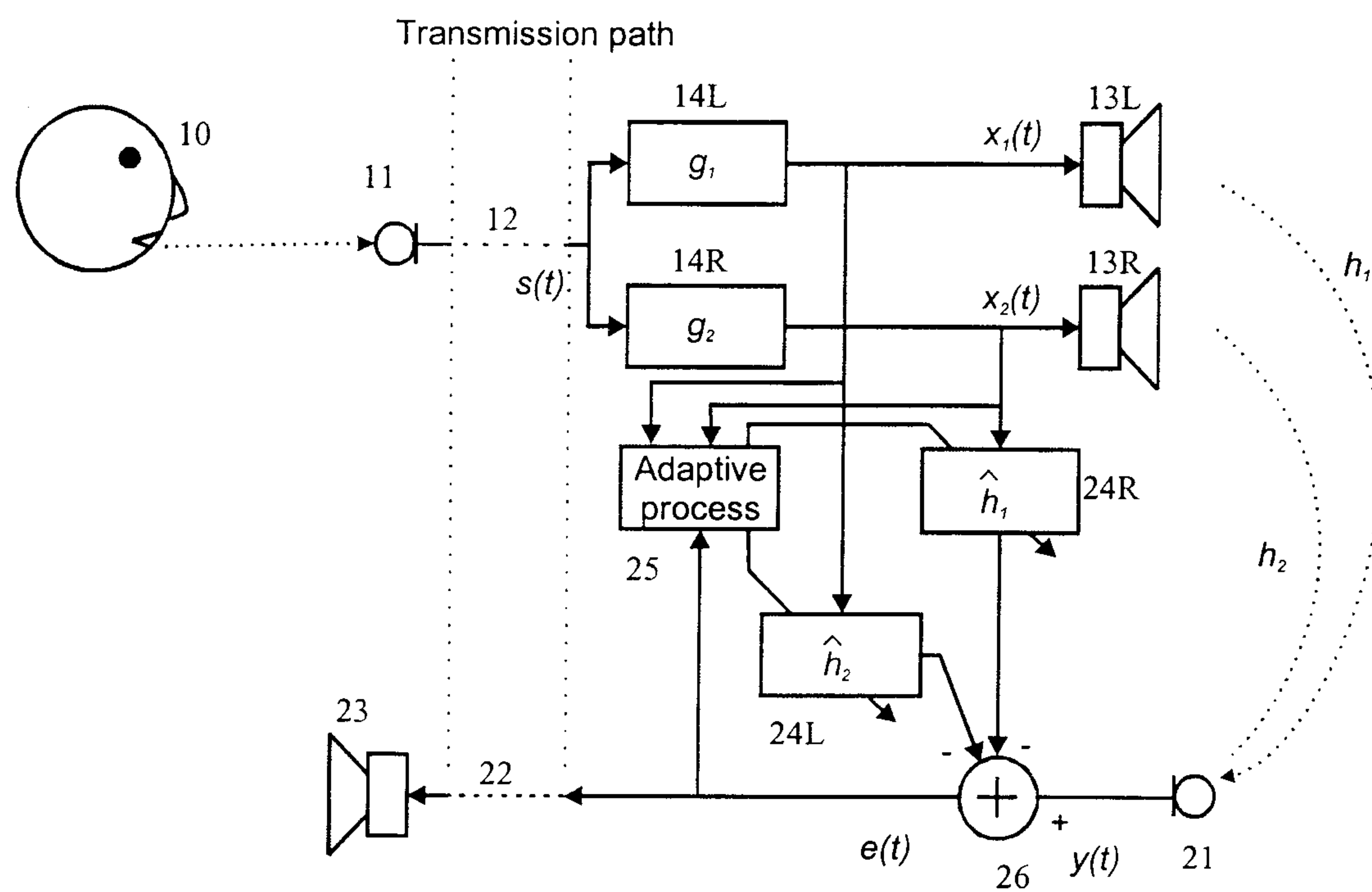


Figure 4

3/4

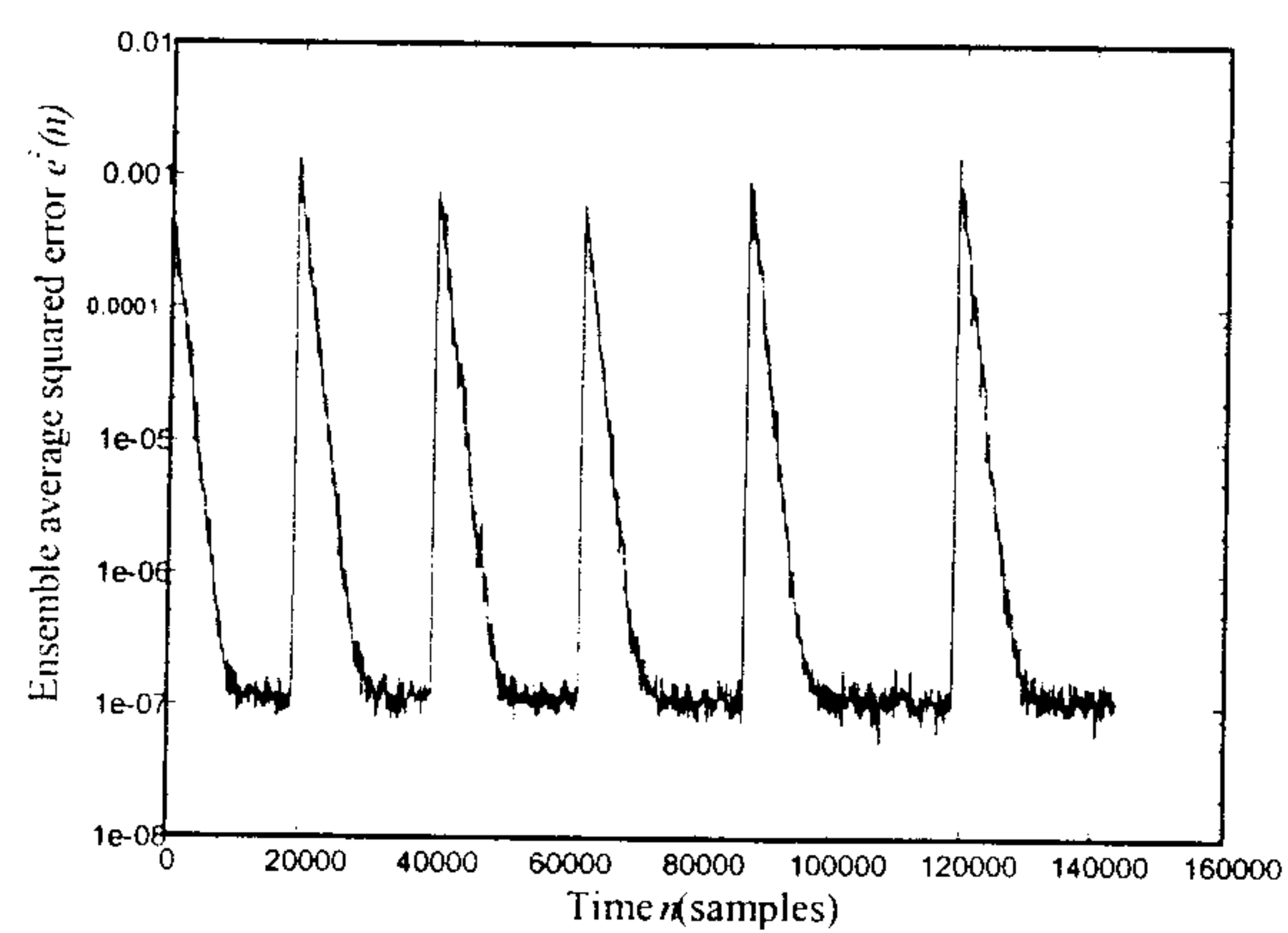


Figure 5a)

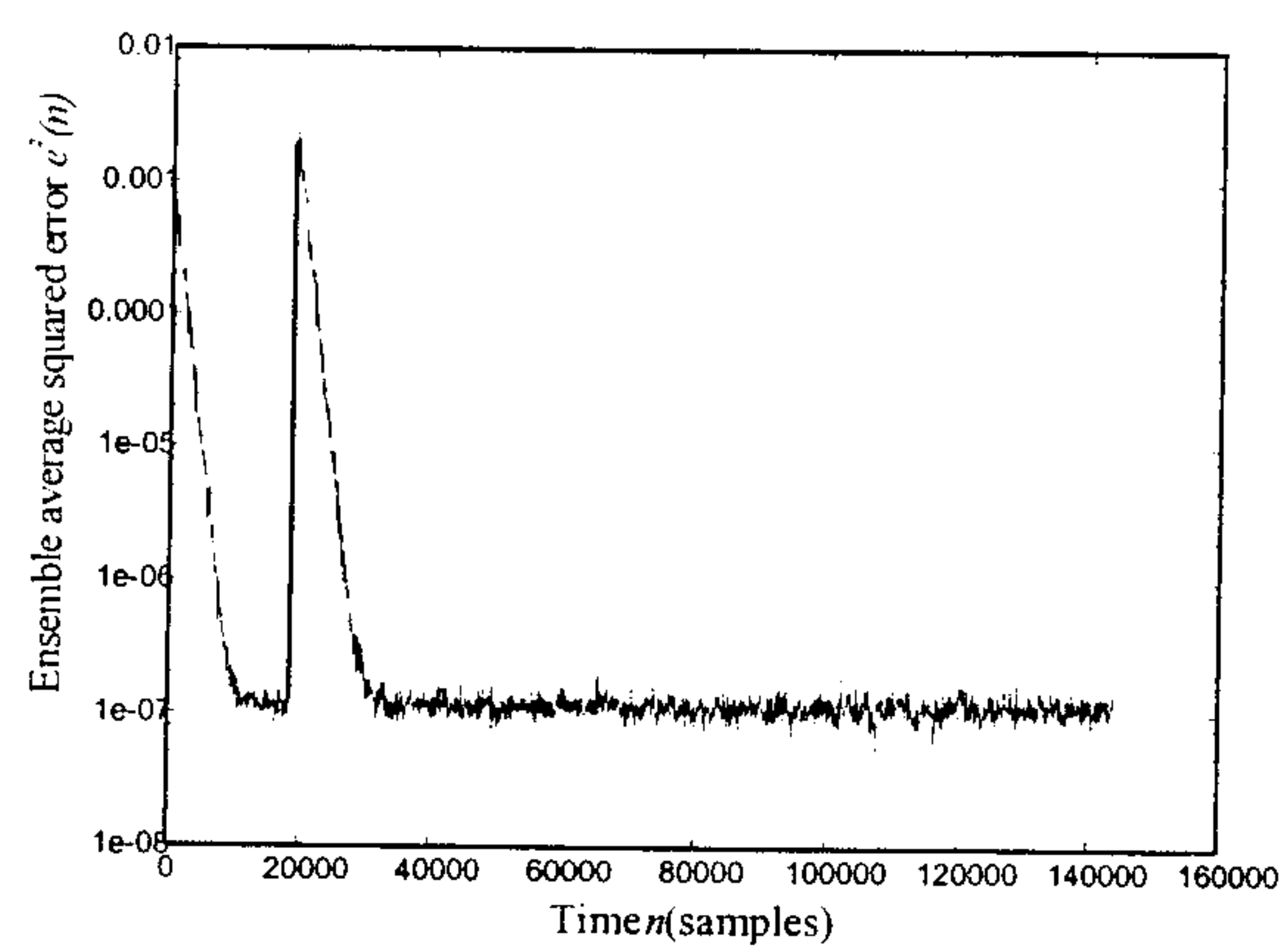


Figure 5b

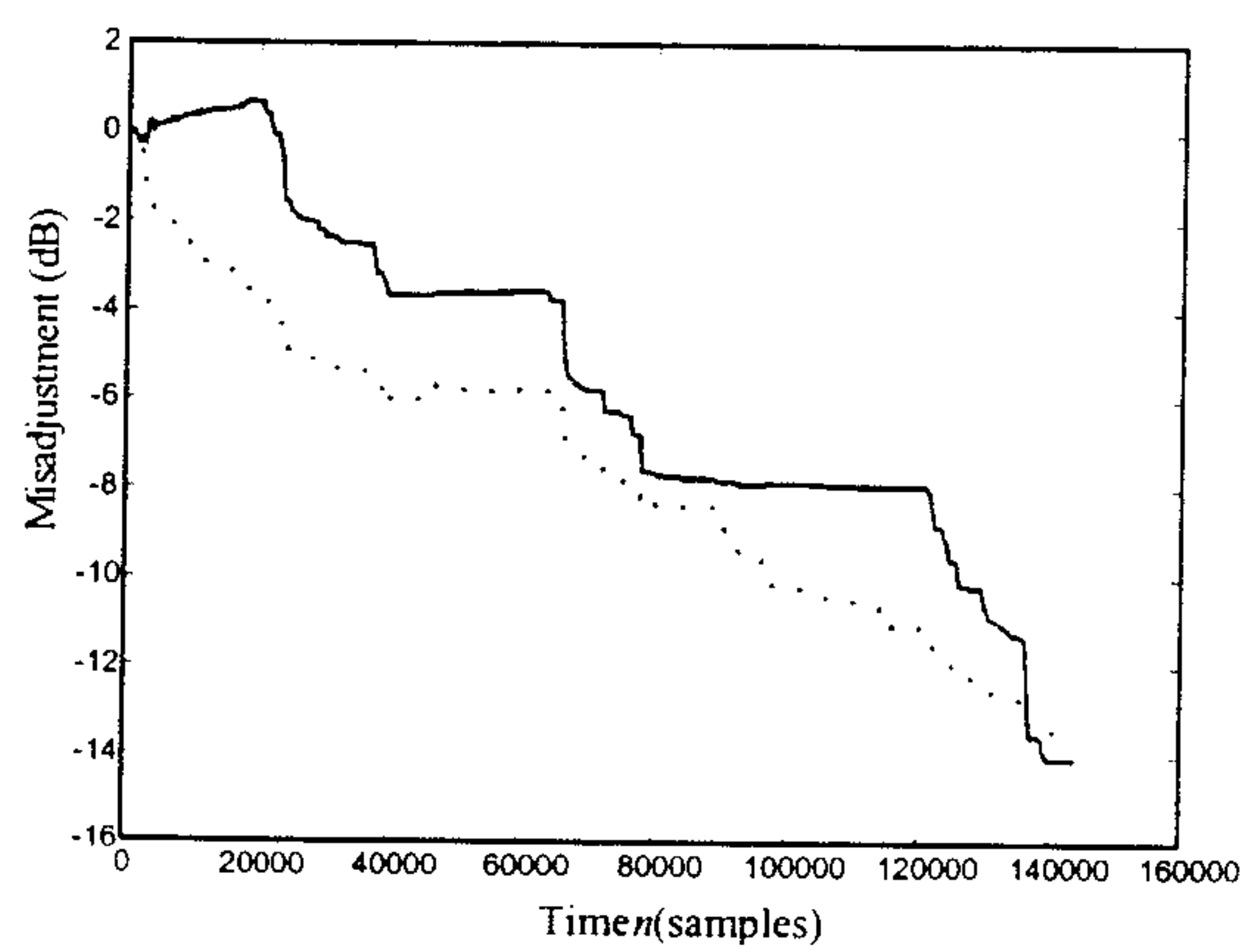


Figure 6

4/4

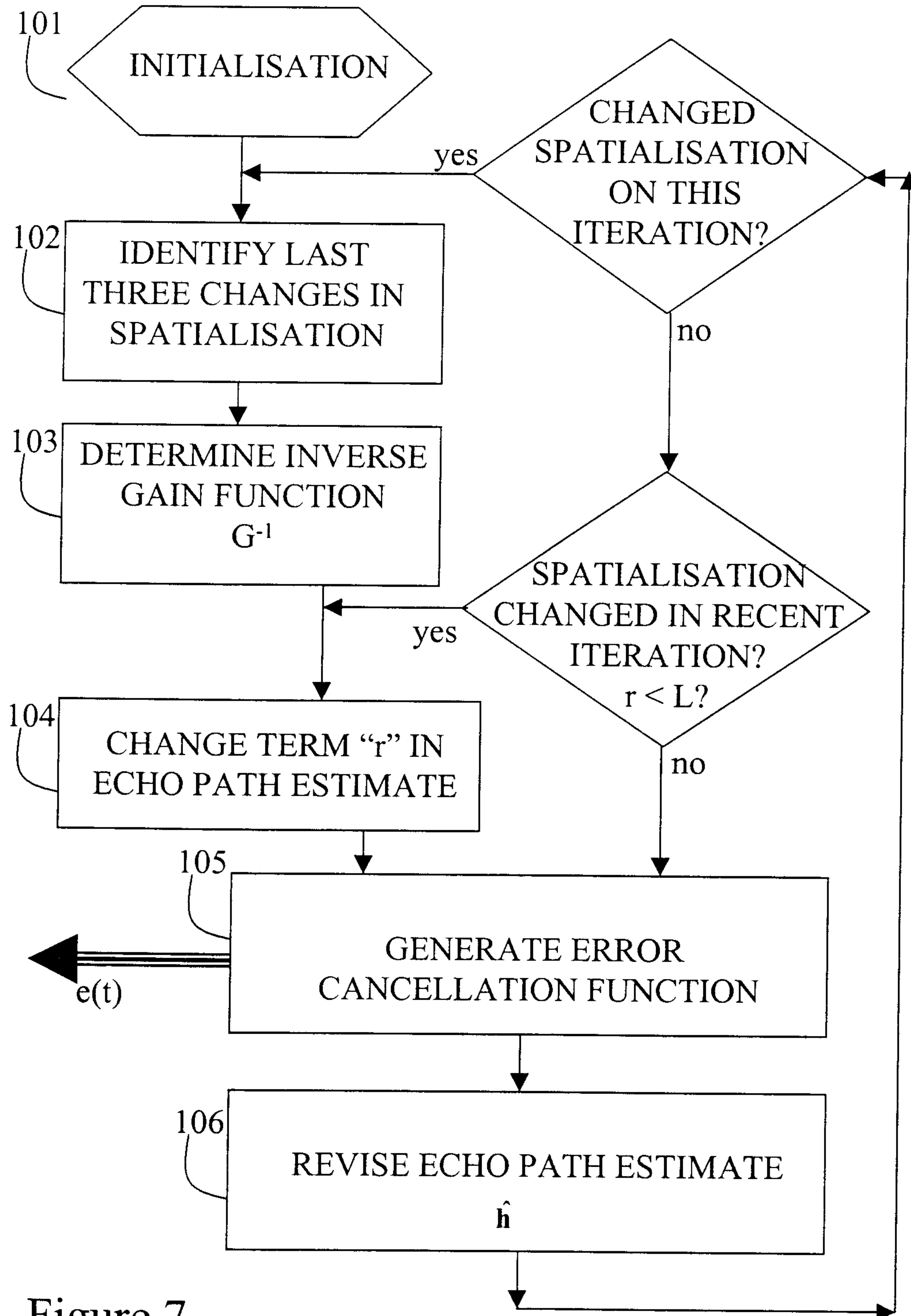


Figure 7

