(12) **EUROPEAN PATENT APPLICATION**

(54) **Gain and spectral shape adjustment in audio signal processing**

(57) The present invention relates to a method for audio signal processing in a communication system comprising a near party and a remote party, comprising receiving an audio signal $(x(n))$ by the near party that is transmitted by the remote party; estimating local background noise by the near party; estimating local acoustic echo by the near party; and adjusting the gain and the spectral shape of the received audio signal $(x(n))$ based on the estimated local background noise and the estimated local acoustic echo.
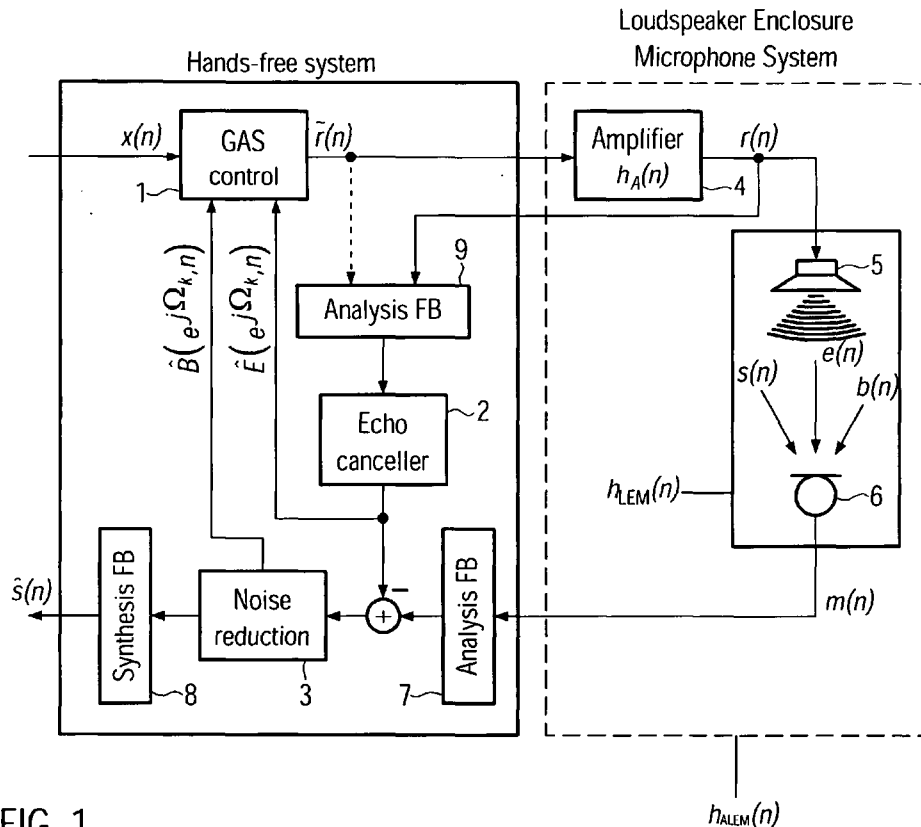
FIG. 1

EP 2 048 659 A1

Description

Field of Invention

5 [0001] The present invention relates to audio signal processing, in particular, to gain and spectral shape adjustment of a speech signal on in a hands-free set based on an estimated local noise and local echo. The hands-free set may be installed in a vehicle.

Background of the invention

10

[0002] Noise reduction and echo compensation are important issues in the art of audio signal processing in communication systems as, e.g., hands-free telephone sets. Two-way speech communication of two communication parties mutually transmitting and receiving audio signals, in particular, speech signals, often suffers from deterioration of the quality of the audio signals by background noise. Background noise in noisy environments can severely affect the quality
15 and intelligibility of voice conversation and can, in the worst case, lead to a complete breakdown of the communication.
[0003] In the case of a (telephone) hands-free set, e.g., it has to be avoided that signals received from a remote party and output by loudspeakers at the near end are fed again in the communication system by microphones at the near end and transmitted back to the remote party. Detection of signals by the microphones that are output by the loudspeakers can result in annoying acoustic echoes, if the acoustic echoes are not significantly attenuated or substantially removed.
20 [0004] Several methods for echo compensation have been proposed and implemented in communication systems in recent years. Adaptive filters are employed for echo compensation of acoustic signals (see, e.g., Acoustic Echo and Noise Control, E. Hänsler and G. Schmidt, John Wiley & Sons, New York, 2004) that are used to model the transfer function (impulse response) of the loudspeaker-room-microphone (LRM) - system by means of an adaptive finite impulse response (FIR) filter.
25 [0005] In the context of vehicle entertainment and communication system methods for adapting the output volume, e.g., of a radio, in response to estimated local noise are known. In hands-free communication in vehicles, the microphone signals suffer from a relatively low Signal-to-Noise Ratio (SNR). Consequently, some noise reduction must be employed in order to improve the speech signal quality.
[0006] A usual method to improve the signal quality in distant talking speech acquisition is the utilization of multi-
30 channel systems, i.e. microphone arrays, as described in "Microphone Arrays: Signal Processing Techniques and Applications", eds. Brandstein, M. and Ward, D., Springer, Berlin 2001.
[0007] However, a severe problem is caused by time-dependent perturbations resulting in a temporarily varying SNR and a significantly affected intelligibility of received speech signals, since fast and reliable adaptation of the employed noise reduction filters cannot be guaranteed, for instance. Moreover, the intelligibility is also affected, since not only the
35 gain of the wanted signal is insufficient but also the spectral shape of the wanted signal is deteriorated due to perturbations and acoustic echoes.
[0008] Therefore, there is a need for an improved method for signal processing in hands-free sets on the receiver side resulting in enhanced audio signals, in particular, speech signals showing less deterioration due to noise and/or echoes as compared to the art. The method must, in particular, be implemented in embedded systems with limited
40 computer resources and, thus, has to be designed and optimized in view of low processor loads and little memory demands.

Description of the Invention

45 [0009] The above-mentioned problem is solved by the method for audio signal processing in a communication system comprising a near party and a remote party according to claim 1. The method comprises the steps of
receiving an audio signal ($x(n)$) by the near party that is transmitted by the remote party;
estimating local background noise by the near party;
estimating local acoustic echo by the near party; and
50 adjusting the gain of the received audio signal ($x(n)$) based on the estimated local background noise and the estimated local acoustic echo.
[0010] The method operates on one side of a communication system, e.g., a telephone system. The remote party at a remote side transmits an acoustic signal, e.g., a speech signal, to the near party at a near side. Herein, by the expression "party" a near side or remote side communication device being part of the above-mentioned communication system is
55 meant. As it is usual in the art, the gain of the received signal is to be adjusted at the near side. It is known in the art to adjust the gain based on an estimate of the local noise at the near side (i.e., the receiving side in the present context).
[0011] However, in the present invention local acoustic echo is also estimated and the estimated echo is used for adjusting the gain of the received audio signal. The near party comprises a microphone and a loudspeaker. The local

echo is the echo component of a microphone signal obtained at the near side that is due to a loudspeaker output. At the near side the microphone detects this loudspeaker output, local background noise and a wanted signal, in particular, a speech signal of a near side user. The local echo can be estimated by means of an echo compensation filtering means as known in the art. By means of a linear (EP 0204718 B1) or non-linear (EP 1020068 B1; WO-A-96/26592) adaptive filter a replica of the acoustic feedback is synthesized and a compensation signal is obtained from the received signal (reference signal). This compensation signal is subtracted from the sending signal of the microphone thereby generating an enhanced resulting signal to be sent to the remote party. It is noted that the echo estimation in the present invention might be based on a reference signal obtained from the received signal before or after amplification (see detailed description below).

[0012] Since the gain of the received signal is adjusted depending on both the estimate of the local noise and the local echo, the signal-to-noise ratio (SNR) can almost be maintained constant (and at a high level) close to the ears of a local user/speaker at the near side. Thus, the intelligibility of a received speech signal can significantly be improved. The method is advantageously implemented in a hands-free set telephone set, e.g., a hands-free set installed in a vehicle. Such a hands-free set is usually provided with both some echo canceling means and some noise reduction means and, therefore, the inventive method can readily be implemented in present day hands-free sets. The method successfully operates for both highly time varying local background noise, e.g., caused by changing driving speeds and road pavements, and changing signal level (power) of the received signals from a remote communication partner whereby an automatic gain control is obtained.

[0013] According to a preferred embodiment the method may also comprise adjusting the spectral shape of the received audio signal (x(n)) based on the estimated local background noise and the estimated local acoustic echo. By adjusting the spectral shape, i.e. equalization, of the received audio signal (x(n)) the quality of the signal is further enhanced and the intelligibility of a received speech signal received by a hands-free set is significantly improved.

[0014] Equalization (modifying the frequency envelope of an audio signal) can be performed by means of either an Infinite Impulse Response (IIR) filter or an Finite Impulse Response (FIR) filter. Both kinds of filter have individual advantages. Whereas finite impulse response (FIR) filters are stable, since no feedback branch is provided, recursive infinite impulse response (IIR) filters typically meet a given set of specifications with a much lower filter order than a corresponding FIR filter. Efficient processing in terms of computational time may be achieved more readily by IIR filters, but these filters may suffer demand for permanent stability checks. In particular applications, a small gain peak may be modeled better by an IIR filter than by an FIR filter and a small attenuation peak may be modeled better by an FIR filter than by an IIR filter.

[0015] Thus, according to an embodiment a desired spectral shape is modeled by both an IIR filter and an FIR filter, the model results of both filters are compared with each other, e.g. for each frame separately, and (for each frame) the model result that better matches the desired spectral shape is used for adjusting the spectral shape of the received audio signal (x(n)).

[0016] The spectral shape of the received audio signal (x(n)) can be adjusted by means of an IIR filter based on the Inverse Discrete Fourier Transform of the squared magnitude of the desired spectral shape and/or an FIR filter based on the Inverse Discrete Fourier Transform of the inverse of the squared magnitude of the desired spectral shape.

[0017] Adjustment of the gain and/or the spectral shape of the received signal can efficiently by performed on the basis of an estimate of the average peak echo-to-noise ratio basically given by the summed up and averaged spectral estimate for the local echo and the summed up and averaged spectral estimate for the background noise (both at some time n) multiplied by the gain introduced by the adjustment of the gain (at time n - 1; see detailed description below. Advantageously, the estimate of the average peak echo-to-noise ratio is smoothed in frequency, e.g., by means for a first order IIR filter, in order to avoid the generation of artifacts due to outliers and/or annoying abrupt adjustment results.

[0018] The present invention also provides a computer program product, comprising one or more computer readable media having computer-executable instructions for performing the steps of an example of the herein disclosed method for audio signal processing in a communication system comprising a near party and a remote party.

[0019] The above-mentioned problem is also solved by a signal processing means for processing a received audio signal (x(n)), comprising

at least one loudspeaker configured to output a loudspeaker signal;

at least one microphone configured to generate a microphone signal;

a noise reduction filtering means (e.g., a Wiener like filter as known in the art) configured to filter the microphone signal and to output a signal representing an estimate of local background noise (present in the microphone signal);

an echo compensation filtering means configured to filter the microphone signal (subtract an estimate of a local echo from the microphone signal) and to output a signal representing an estimate of local echo present (that is

present in the microphone signal); and

a gain and spectral shape control unit connected with the noise reduction filtering means and the echo compensation filtering means and configured to adjust the gain and/or spectral shape of the received audio signal (x(n)) based on both the signal representing an estimate of local background noise and the signal representing an estimate of local echo.

[0020]   The signal processing means may also comprise an amplifier and the echo compensation filtering means may be configured to receive a reference signal necessary for adaptation of its filter coefficient either before or after amplification of the received audio signal processed by the gain and shape control unit.

[0021]   According to an example, the gain and shape control unit of the signal processing means comprises an FIR filter and/or an IIR filter configured to adjust the spectral shape of the received audio signal (x(n)). Moreover, the gain and shape control unit of the signal processing means may comprise a means configured to estimate the average peak echo-to-noise ratio.

[0022]   In one advantageous embodiment the signal processing means may further comprise a means for smoothing in frequency the estimated average peak echo-to-noise ratio.

[0023]   The signal processing means may comprise one or more microphones, in particular, a microphone array comprising one or more directional microphones and a beamforming means (see e.g., "Adaptive beamforming for audio signal acquisition", by Herbordt, W. and Kellermann, W., in "Adaptive signal processing: applications to real-world problems", p.155, Springer, Berlin 2003) configured to obtain a beamformed microphone signal. In this case the local echo and the background noise on the near side are obtained from the beamformed signal or from one or more signals obtained by the microphones of the microphone array before beamforming.

[0024]   Moreover, the herein disclosed signal processing means may further comprise a transmitting means for transmitting the microphone signal filtered by the echo compensation filtering means and the noise reduction filtering means to a remote party.

[0025]   Large parts of the processing can suitably be performed in the sub-band regime. Therefore, the signal processing means may comprise a first analysis filter bank configured to divide the microphone signal into microphone sub-band signals, a second analysis filter bank configured to divide a reference signal into sub-band signals to be input in the echo compensation filtering means and a synthesis filter bank configured to synthesize the microphone sub-band signals after they have been filtered by the echo compensation filtering means and the noise reduction filtering means.

[0026]   The present invention also provides a hands-free set, in particular, installed in a vehicle, e.g., an automobile, comprising a signal processing means according to one of the above examples of a signal processing means according to the present invention.

[0027]   Additional features and advantages of the present invention will be described with reference to the drawings. In the description, reference is made to the accompanying figures that are meant to illustrate preferred embodiments of the invention. It is understood that such embodiments do not represent the full scope of the invention.

[0028]   Figure 1 illustrates two alternative examples of a realization of the inventive hands-free system employing a gain and shape control unit based on estimated echo and noise.

[0029]   Figure 2 illustrates an example of the gain and shape control unit of Figure 1 comprising a gain control, a shape control and an FIR filter in some detail.

[0030]   Figure 3 illustrates the gain processing of the gain control of Figure 2 in detail.

[0031]   Figure 4 illustrates the adjustment of the gain when the shape control of Figure 2 is active.

[0032]   Figure 5 illustrates the adaptation of the FIR filter of Figure 2 by means of the Levinson Durbin algorithm.

[0033]   As shown in Figure 1 a realization of the herein disclosed audio signal processing in a hands-free set comprises a gain and shape control (GAS) unit 1 which is supplied with a speech signal x(n) that was transmitted by a remote party and received by the hands-free set. The GAS control unit 1 is provided for enhancing the quality of the received signal x(n) by an appropriate gain and shape processing and outputs an enhanced speech signal $\tilde{r}(n)$.

[0034]   The hands-free set comprises an echo compensation filtering means 2 as well as a noise reduction filtering means 3. Since filtering of a microphone signal obtained by a microphone 6 is performed in the sub-band regime (processing in the frequency domain would, of course, also be possible instead), the hands-free set comprises an analysis filter bank 7 for dividing a microphone signal obtained by the microphone 6 of the hands-free set into microphone sub-band signals. Accordingly, the filtered microphone sub-band signals are synthesized by a synthesis filter bank 8 to obtain an enhanced signal $\hat{s}(n)$.

[0035]   It is a basic feature of the present invention that the GAS control unit 1 operates on the basis of an estimated echo component and an estimated noise component present that are present in a microphone signal obtained by one or more microphones of the hands-free set. In the present example the echo compensation filtering means 2 outputs a frequency selected estimate of the echo component $\hat{E}(e^{j\Omega_k},n)$, where Ω_k and n denote the frequency sub-band and the discrete time index as known in the art, to the GAS control unit 1. The noise reduction filtering means 3 outputs the

square-root of the estimated short-term spectral power density $\hat{B}(e^{j\Omega k}, n)$ of the noise present in the microphone signal m(n) to the GAS control unit 1.

[0036] The enhanced speech signal $\tilde{r}(n)$ is transmitted from the GAS control unit 1 to an amplifier 4 that supplies an amplified signal r(n) to a loudspeaker that outputs the processed speech signal with an increased intelligibility due to the processing by the GAS control unit 1.

[0037] On the other hand, a microphone 6 of the hands-free set detects a speech signal s(n) of a local speaker together with an echo component e(n) produced by the loudspeaker and some background noise b(n) due to noise perturbations. The microphone signal m(n) = s(n) + e(n) + b(n) is filtered by the analysis filter bank 7 to obtain microphone sub-band signals that are echo compensated by the echo compensation filtering means 2 and noise reduced by the noise reduction filtering means 3. The thus enhanced microphone sub-band signals are synthesized by the synthesis filter bank 8 to obtain an enhanced signal $\hat{s}(n)$ that is to be transmitted to a remote party by a transmission means (not shown).

[0038] According to one embodiment the filter coefficients of the echo compensation filtering means 2 are adapted to model the impulse response $h_{LEM}(n)$ of the loudspeaker - enclosure - microphone system represented by the loudspeaker and the microphone of the hands-free set as well as the enclosure/room wherein the hands-free set is installed, e.g., a vehicular cabin. In this case, r(n) is input as a reference signal to the echo compensation filtering means 2 in order to adapt the filter coefficients of the echo compensation filtering means 2 (e(n) = r(n) * hLEM(n)).

[0039] In an alternative embodiment, the echo compensation filtering means 2 is supplied for adaptation of the filter coefficients by the signal $\tilde{r}(n)$ output by the GAS control unit 1 (before amplification by the amplifier 4). In this case, the echo compensation filtering means 2 has to perform two convolutions of time-dependent systems namely for the impulse (frequency) response $h_A(n)$ of the amplifier 4 and the impulse response $h_{LEM}(n)$ of the loudspeaker - enclosure - microphone system, i.e. it models a combined impulse response $h_{ALEM}(n)$, e(n) = $\tilde{r}(n)$ * $h_{ALEM}(n)$, see Figure 1. Either r(n) or $\tilde{r}(n)$ is input in an analysis filter bank 9 similar to the analysis filter bank 7 for the operation of the echo compensation filtering means 2 in the sub-band regime. Both analysis filter banks 7 and 9 as well as the synthesis filter bank 8 may comprise Hann or Hamming windows as known in the art.

[0040] In the present example, both the gain and the spectral shape of the received signal x(n) are changed to improve the quality of the signal transmitted by a remote party. Fig. 2 shows the GAS control unit 1 of Fig. 1 in some more detail. According to the present invention the gain is controlled on the basis of estimates of both noise and acoustic echo that are present in the microphone signal m(n). The spectral estimates $\hat{E}(e^{j\Omega k}, n)$ and $\hat{B}(e^{j\Omega k}, n)$ obtained by the echo compensation filtering means 2 and the noise reduction filtering means 3, respectively, are summed up and averaged (see also Figure 3 illustrating the operation of the gain control 10 in some more detail):

$$\overline{b}(n) \quad = \frac{1}{\frac{N_{FFT}}{2}+1} \sum_{k=0}^{N_{FFT}/2} \left| \hat{B}(e^{j\Omega_k},n) \right|$$

where N_{FFT} denotes the order of the FFT (number of interpolation points) and, if the shape control is not active,

$$\overline{e}(n) \quad = \frac{1}{\frac{N_{FFT}}{2}+1} \sum_{k=0}^{N_{FFT}/2} \left| \hat{E}(e^{j\Omega_k},n) \right|.$$

Whereas the present invention is particularly useful in the case of a combined operation of the gain and shape control (see reference signs 10 and 11 in Fig. 2) it can be restricted to the operation of the gain control 10. In this case the above equation determines the summed up and averaged spectral estimates for the echo. If, however, the shape control 11 is active, the above expression for the averaged estimate for the echo is to be replaced by

$$\overline{e}(n) \quad = \frac{1}{\frac{N_{FFT}}{2}+1} \sum_{k=0}^{N_{FFT}/2} \frac{\left| \hat{E}(e^{j\Omega_k},n) \right|}{G_{Des}(e^{j\Omega_k},n)}$$

where $G_{Des}(e^{j\Omega k}, n)$ is the desired gain that is output by the shape control 11 (see description below). Thereby, independent functionality of the gain control 10 and the shape control 11 can be guaranteed when both are active.

[0041]   Depending on the actual application it is sufficient to use only some portion of the entire bandwidth and, thus, the summation may be shortened ($< N_{FFT} / 2$). The normalization factor before the summations is not obligatory.

[0042]   By the gain control 10 an average peak echo level $\bar{e}_{Pk}(n)$ is determined by smoothing whenever a predetermined echo-to-noise (ENR) threshold ($t_{ENR, 1}$) is exceeded (see also Figure 3)

$$\bar{e}_{Pk}(n) = \begin{cases} \tau\bar{e}(n) + (1-\tau)\bar{e}_{Pk}(n-1), & \text{if } \dfrac{\bar{e}(n)}{\bar{b}(n)} > t_{ENR,1} \\ \bar{e}_{Pk}(n-1), & \text{otherwise} \end{cases}$$

where the smoothing constant is chosen as $0 < T << 1$. Other ways of smoothing can be used, e.g., by multiplication by two different time constants for falling and increasing echo power, respectively. For $t_{ENR,1}$ values out of the interval [2, 10] have proven to be advantageous.

The average peak ENR can, thus, be determined by

[0043]

$$p_{ENR}(n) = \frac{\bar{e}_{Pk}(n)}{\bar{b}(n)\, g_{Real,lim}(n-1) + \varepsilon}$$

where $p_{ENR}(n)$ is compensated for the gain introduced by the gain control 10 $g_{Real,lim}(n-1)$ and the small constant $\varepsilon << 1$ is added in the denominator to avoid division by zero.

[0044]   A desired gain $g_{Des}(n)$ for establishing a constant ENR level ($t_{ENR,2}$) in order to improve the intelligibility of the received speech signal $x(n)$ is determined by

$$g_{Des}(n) = \begin{cases} \dfrac{t_{ENR,2}}{P_{ENR}(n)}, & \text{if } P_{ENR}(n) < t_{ENR,2} \\ 1, & \text{otherwise} \end{cases}$$

[0045]   For $t_{ENR,2}$ values out of the interval [4, 30] have proven to be advantageous.

A preliminary gain can thus be determined by the gain control 10 as

[0046]

$$g_{Real}(n) = \begin{cases} g_{Real}(n-1)\, \tau_{g,rise,} \\ \quad \text{for } (1-\alpha_g)g_{Real}^2(n-1) + \alpha_g\, g_{Real}(n-1) \le g_{Des}(n) \text{ and } \dfrac{\overline{e}(n)}{\overline{b}(n)} > t_{ENR,1}, \\ g_{real}(n-1)\, \tau_{g,fall,} \\ \quad \text{for } (1-\alpha_g)g_{Real}^2(n-1) + \alpha_g\, g_{Real}(n-1) > g_{Des}(n) \text{ and } \dfrac{\overline{e}(n)}{\overline{b}(n)} > t_{ENR,1}, \\ g_{Real}(n-1), \text{ otherwise} \end{cases}$$

where the characteristic can be adjusted by the positive real parameter $\alpha_g < 1$. The increment and decrement parameters $\tau_{g,rise}$ and $\tau_{g,fall}$ satisfy

$$0 \ll \tau_{g,fall} < 1 < \tau_{g,rise} \ll \infty .$$

[0047] The preliminary gain, thus, is adjusted ranging from merely compensating the current signal-to-noise ratio up to the predetermined limit of $t_{ENR,2}$ and can be limited to a maximum allowable gain g_{max} (see Figure 3):

$$g_{Real,lim}(n) = \min\{g_{max,}\, g_{Real}(n)\}.$$

[0048] If the shape control is not active, an enhanced signal $\tilde{r}(n) = \tilde{r}_{Gain}(n) = g_{Real,lim}(n)\, x(n)$ is obtained. By the processing described above the gain increases as time-dependent noise increases and also when the background noise is almost stationary but the speech power of a remote speaker decreases (thereby realizing an automatic gain control).
[0049] In the example shown in Fig. 2 shaping of x(n) is performed by means of a Finite Impulse Response (FIR) Filter 12 controlled by the shape control 11. It is noted that an Infinite Impulse Response (IIR) Filter could alternatively be employed. Figure 4 illustrates the operation of the shape control in detail. An average peak echo level is determined by the shape control 11 by smoothing the estimated echo spectrum provided by the echo compensation filtering means 2 of Figure 1 each time a predetermined ENR threshold ($t_{ENR,1}$) is exceeded similar to the above-described determination by the gain control 10 but for each frequency bin

$$\hat{E}_{Sm}(e^{j\Omega_k}, n) = \begin{cases} \beta\big|\hat{E}(e^{j\Omega_k}, n)\big| + (1-\beta)\hat{E}_{Sm}(e^{j\Omega_k}, n-1), & \text{if } \dfrac{\overline{e}(n)}{\overline{b}(n)} > t_{ENR,1} \\ \hat{E}_{Sm}(e^{j\Omega_k}, n-1) \end{cases}$$

with $0 < \beta \ll 1$. Thus, a spectral average peak ENR is obtained

$$P_{ENR}(e^{j\Omega_k}, n) = \frac{\hat{E}_{Sm}(e^{j\Omega_k}, n)}{\big|\hat{B}(e^{j\Omega_k}, n)\big| G_{Des}(e^{j\Omega_k}, n) + \varepsilon}$$

that advantageously is smoothed in the positive and negative frequency directions to obtain a smoothed spectral average peak ENR, i.e. $P_{ENR,Sm}(e^{j\Omega k}, n)$. The smoothing can, e.g., be performed by a first order IIR filter. The desired gain vector

is obtained by (see Figure 4)

$$\widetilde{G}_{Des}(e^{j\Omega_k}, n) = \begin{cases} G_{Des}(e^{j\Omega_k}, n-1)\tau_g, \\ \quad \text{if } ((1-\alpha_s)G^2_{Des}(e^{j\Omega_k}, n-1) + \alpha_s G_{Des}(e^{j\Omega_k}, n-1)) P_{ENR,Sm}(e^{j\Omega_k}, n) > t_{ENR,2} \\ G_{Des}(e^{j\Omega_k}, n-1)\tau_G, \text{ otherwise.} \end{cases}$$

with $0 < \alpha_s < 1$ and $0 \ll \tau_g < 1 < \tau_G \ll \infty$.

[0050]   After limitation of the desired gain in the directions of the maximum gain and maximum attenuation, respectively, one gets

$$G_{Des}(e^{j\Omega_k}, n) = \min\left\{ \tau_{G,max}(n), \max\left\{ \tau_{G,min}(n), \widetilde{G}_{Des}(e^{j\Omega_k}, n) \right\} \right\}$$

where $T_{G,max}(n)$ and $T_{G,min}(n)$ depend on the gain $g_{Real,lim}(n)$ computed by the gain control 10. If only a small gain is introduced by the gain control 10 (or no gain at all), only a small gain is introduced by the shape control 11 (or no gain at all). A significant gain introduced by the gain control 10, on the other hand, will result in a significant gain introduced by the shape control 11 (see also Figure 4).

[0051]   In the example shown in Figure 2 the signal $\tilde{r}_{Gain}(n)$ obtained by processing the received speech signal x(n) by the gain control 10 is shaped. In the shaping process both Discrete Fourier Transformations (DFT) / Inverse Discrete Fourier Transformations (IDFT) and N_{DFT} multiplications for each frequency bin in the frequency domain are avoided in view of computational costs. For this purpose, a low order FIR filter is 12 is designed. It is noted that the delay introduced by the FIR filter is significantly lower than the one that would be introduced by DFT/IDFT processing of $\tilde{r}_{Gain}(n)$.

[0052]   According to the present example, the inverse of the squared magnitude of the desired spectral shape correction is transformed into the time domain

$$\mathbf{a}_{Des}(n) \quad = \quad IDFT\left\{ \frac{1}{\left| G_{Des}(e^{j\Omega_k}, n) \right|^2} \right\}$$

with the vector containing the auto correlation coefficients $\mathbf{a}_{DES,i}(n)$

$$\mathbf{a}_{Des}(n) \quad = \quad \left[a_{Des,0}(n), a_{Des,1}(n), \dots, a_{Des,N_{DFT}-1}(n) \right]^T$$

where N_{DFT} denotes the order of the DFT. This vector is shortened to the order of the desired FIR filter (N_{FIR}) + 1:

$$\mathbf{a}_{Des,mod}(n) \quad = \quad \left[a_{Des,mod,0}(n), a_{Des,mod,1}(n), \dots, a_{Des,mod,N_{FIR}}(n) \right]^T$$

by

$$\mathbf{a}_{Des,mod}(n) \quad = \quad \mathbf{W}_{cut} \; \mathbf{a}_{Des}(n)$$

with

$$\mathbf{W}_{cut} \quad = \quad \begin{bmatrix} w_{0,0} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & w_{1,1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & w_{N_{FIR},N_{FIR}} & 0 & \cdots & 0 \end{bmatrix}.$$

[0053]   The elements of the matrix $\mathbf{W}_{cut}$ are given by

$$w_{i,i} \quad = \quad 1 \quad \text{for} \quad i \in \{0,...,N_{FIR}\}.$$

[0054]   The vector of the filter coefficients a(n) of the FIR filter 12

$$\mathbf{a}(n) \quad = \quad \left[a_0(n), a_1(n), ..., a_{N_{FIR}-1} \right]^T$$

are determined by the shape control 11 to reproduce the shape $G_{des}(e^{j\Omega k}k, n)$. For this purpose, the following vector equation has to be solved

$$\mathbf{a}(n) \quad = \quad \mathbf{A}_{Des,mod}^{-1}(n) \, \widetilde{\mathbf{a}}_{Des,mod}(n)$$

with

$$\mathbf{A}_{Des,mod}(n) \quad = \quad \begin{bmatrix} a_{Des,mod,0}(n) & a_{Des,mod,1}(n) & \cdots & a_{Des,mod,N_{FIR}-1}(n) \\ a_{Des,mod,1}(n) & a_{Des,mod,0}(n) & \cdots & a_{Des,mod,N_{FIR}-2}(n) \\ \vdots & \vdots & \ddots & \vdots \\ a_{Des,mod,N_{FIR}-1}(n) & a_{Des,mod,N_{FIR}-2}(n) & \cdots & a_{Des,mod,0}(n) \end{bmatrix}$$

and

$$\widetilde{\mathbf{a}}_{DES,mod}(n) \quad = \quad \left[a_{Des,mod,1}(n), a_{Des,mod,2}(n), ..., a_{Des,mod,N_{FIR}} \right]^T.$$

[0055] The above vector equation, e.g., can efficiently be solved by means of the well-known Levinson-Durbin recursion algorithm (see, e.g., Acoustic Echo and Noise Control, E. Hänsler and G. Schmidt, John Wiley & Sons, New York, 2004). The output of the FIR filter 12 has to be enhanced by the correct gain $g_{cor}(n)$ provided by the shape control 11 (see Figures 2 and 5). This can be achieved by using the energy of the residual signal $e_{LD}(n)$ obtained by the Levinson-Durbin recursion algorithm

$$g_{cor}(n) = \sqrt{\frac{N_{DFT}}{e_{LD}(n)}}.$$

As already mentioned above an IIR filter may be employed rather than the FIR filter 12 of Figure 2. In this case, the inverse of the frequency response used for the FIR filter design has to be used

$$\mathbf{a}_{Des,IIR}(n) = IDFT\left\{\left|G_{Des}(e^{j\Omega_k},n)\right|^2\right\}.$$

[0056] After carrying out the same computations described above but using the inverse frequency response $\mathbf{a}_{Des,IIR}(n)$ an all-pole IIR filter results that can be used for shaping the received audio signal $x(n)$.

[0057] Employment of either an FIR filter or an IIR filter for the equalization of the received audio signal $x(n)$ exhibits different advantages. For example, the IIR filter may be superior (inferior) to the FIR filter in modeling small gain peaks (attenuation peaks). Therefore, it might be preferred to model the desired spectral shape by both the IIR and the FIR filters and to compare the model results with each other and choose the better one for shaping.

[0058] For instance, when the Levinson-Durbin recursion algorithm is employed, the prediction error power of the residual signal is automatically obtained and depending on the prediction error power on a frame by frame basis the result of either the FIR filter or the IIR filter can be used for the subsequent processing, i.e. shaping. If the Levinson-Durbin recursion algorithm is not used, the respective prediction error powers of the residual signals can be calculated by

$$e_{LD,FIR}(n) = a_{Des,0}(n) - \sum_{i=0}^{N_{FIR}-1} a_i(n)\, a_{Des,i+1}(n)$$

and

$$e_{LD,IIR}(n) = a_{Des,IIR,0}(n) - \sum_{i=0}^{N_{IIR}-1} a_{HR,i}(n)\, a_{Des,IIR,i+1}(n)$$

that can subsequently be compared with each other for determining the best choice for each frame.

[0059] All previously discussed embodiments are not intended as limitations but serve as examples illustrating features and advantages of the invention. It is to be understood that some or all of the above described features can also be combined in different ways.

Claims

1. Method for audio signal processing in a communication system comprising a near party and a remote party, comprising

    receiving an audio signal ($x(n)$) by the near party that is transmitted by the remote party;

estimating local background noise by the near party;
estimating local acoustic echo by the near party; and
adjusting the gain of the received audio signal (x(n)) based on the estimated local background noise and the estimated local acoustic echo.

2. The method according to claim 1, further comprising adjusting the spectral shape of the received audio signal (x(n)) based on both the estimated local background noise and the estimated local acoustic echo.

3. The method according to claim 2, wherein a desired spectral shape is modeled by both an IIR filter and an FIR filter, the model results of both filters are compared with each other for each frame, and for each frame the model result that better matches the desired spectral shape is used for the adjusting of the spectral shape of the received audio signal (x(n)).

4. The method according to claim 2 or 3, wherein the spectral shape of the received audio signal (x(n)) is adjusted by means of an IIR filter based on the Inverse Discrete Fourier Transform of the squared magnitude of the desired spectral shape and/or an FIR filter based on the Inverse Discrete Fourier Transform of the inverse of the squared magnitude of the desired spectral shape.

5. The method according to one of the claims 2 to 4, wherein the gain and the spectral shape of the received audio signal (x(n)) are adjusted based on an average peak echo-to-noise ratio.

6. The method according to claim 5, wherein the average peak echo-to-noise ratio is smoothed in frequency by an IIR filter.

7. Computer program product, comprising one or more computer readable media having computer-executable instructions for performing the steps of the method according to one of the claims 1 to 6.

8. Signal processing means for processing a received audio signal (x(n)), comprising

    at least one loudspeaker (5) configured to output a loudspeaker signal;
    at least one microphone (6) configured to generate a microphone signal;
    a noise reduction filtering means (3) configured to filter the microphone signal (m(n)) and to output a signal representing an estimate of local background noise;
    an echo compensation filtering means (2) configured to filter the microphone signal m(n) and to output a signal representing an estimate of local echo; and
    a gain and shape control unit (1) connected with the noise reduction filtering means (3) and the echo compensation filtering means (2) and configured to adjust gain and/or spectral shape of the received audio signal (x(n)) based on both the signal representing an estimate of local background noise and the signal representing an estimate of local echo.

9. The signal processing means according to claim 8, further comprising an amplifier (4) and wherein the echo compensation filtering means (2) is configured to receive a reference signal necessary for adaptation of its filter coefficient either before or after amplification of the received audio signal processed by the gain and shape control unit (1).

10. The signal processing means according to claim 8 or 9, wherein the gain and shape control unit (1) comprises an FIR filter (12) and/or an IIR filter configured to adjust the spectral shape of the received audio signal (x(n)).

11. The signal processing means according to one of the claims 8 to 10, wherein the gain and shape control unit (1) comprises a means configured to estimate the average peak echo-to-noise ratio.

12. The signal processing means according to claim 11, further comprising a means for smoothing in frequency the estimated average peak echo-to-noise ratio.

13. The signal processing means according to one of the claims 8 to 12, further comprising a microphone array comprising one or more directional microphones and a beamforming means configured to obtain a beamformed microphone signal.

14. The signal processing means according to one of the claims 8 to 13, further comprising a first analysis filter bank

(7) configured to divide the microphone signal (m(n)) into microphone sub-band signals, a second analysis filter bank (9) configured to divide a reference signal into sub-band signals to be input in the echo compensation filtering means and a synthesis filter bank (9) configured to synthesize microphone sub-band signals filtered by the echo compensation filtering means (2) and the noise reduction filtering means (3).

15. The signal processing means according to one of the claims 8 to 14, further comprising a transmitting means for transmitting the microphone signal filtered by the echo compensation filtering means (2) and the noise reduction filtering means (3) to a remote party.
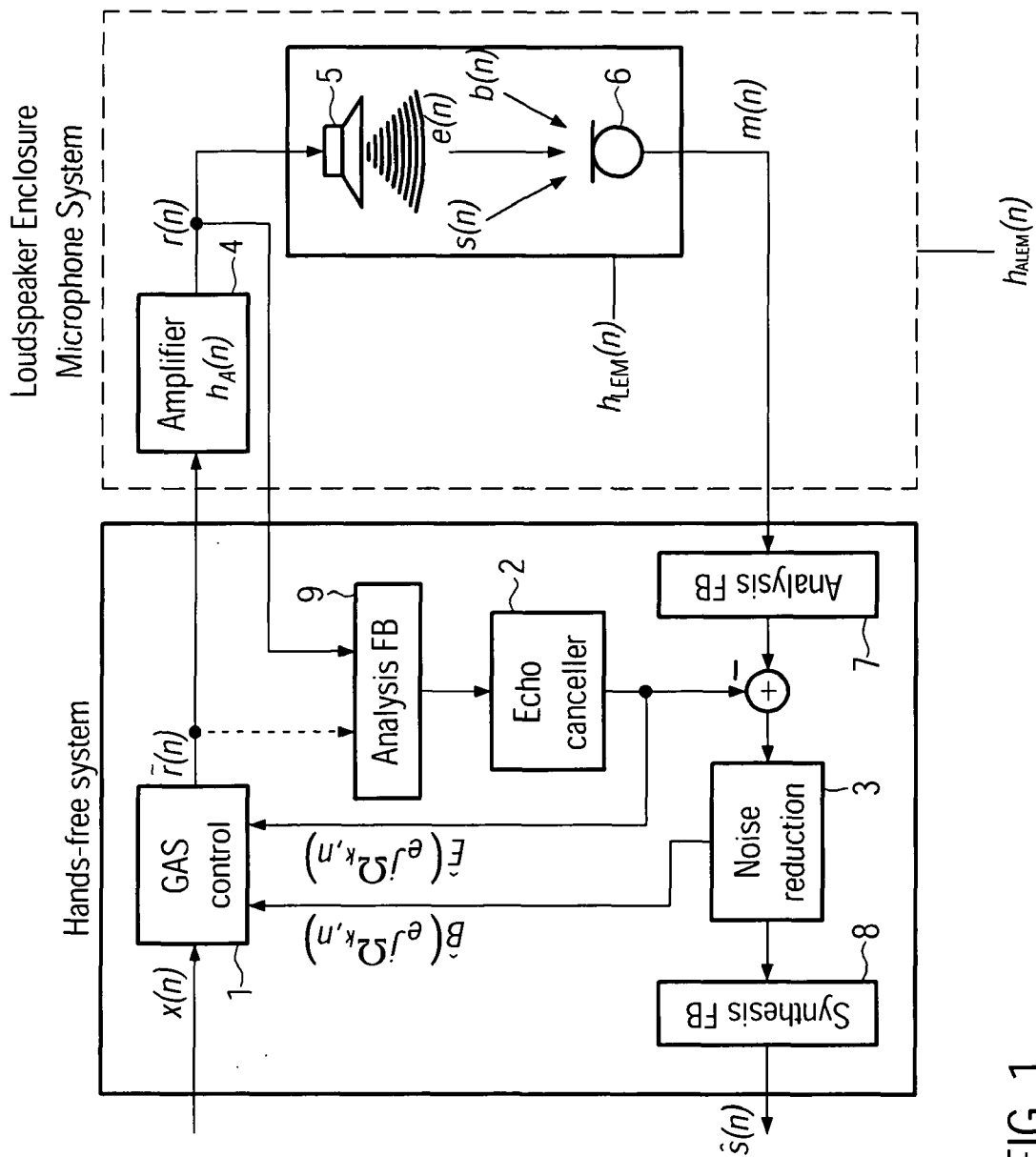
16. Hands-free set comprising a signal processing means according to claim 15.

FIG. 1

FIG. 2

FIG. 3

FIG. 4
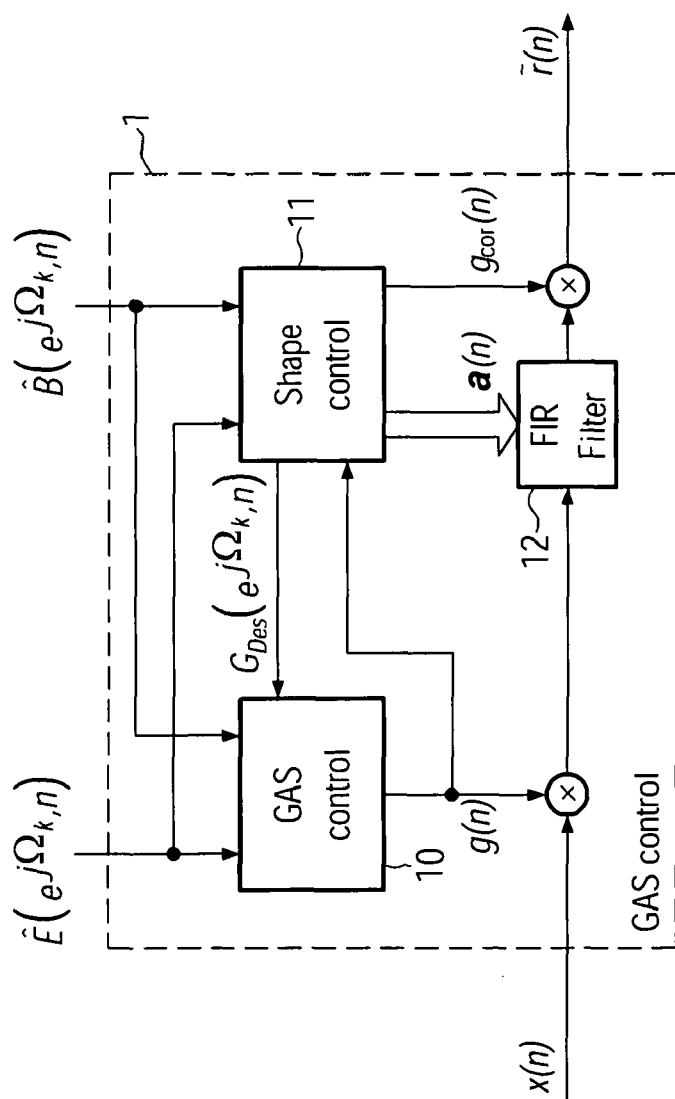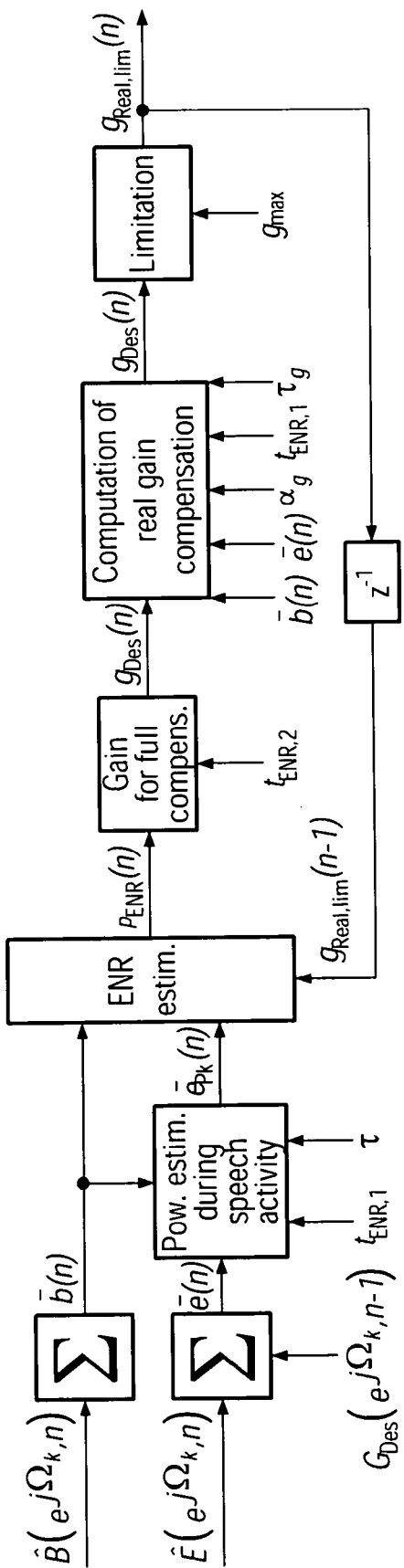
$G_{Des}\left(e^{j\Omega_{k},n}\right)$ →

$(\ )^{-2}$

IDFT

$\mathbf{a}_{DES}(n)$

Levinson Durbin Recursion

$\mathbf{a}(n)$

$e_{LD}(n)$

N_{DFT}

$g_{cor}(n)$

FIG. 5

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 07 01 9652
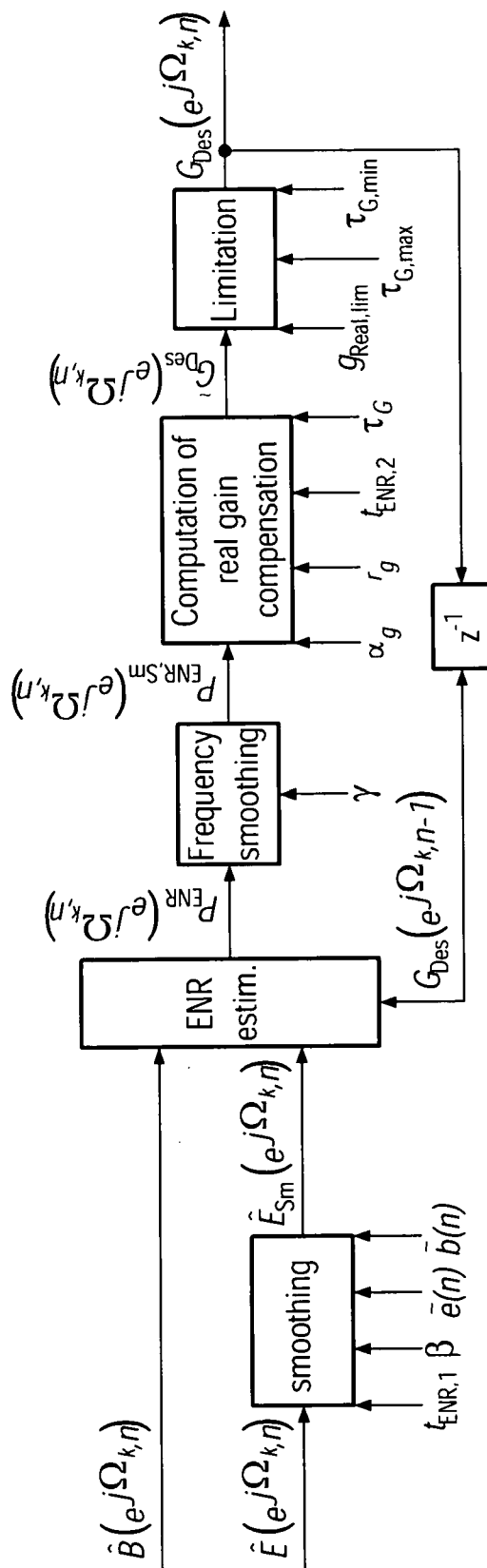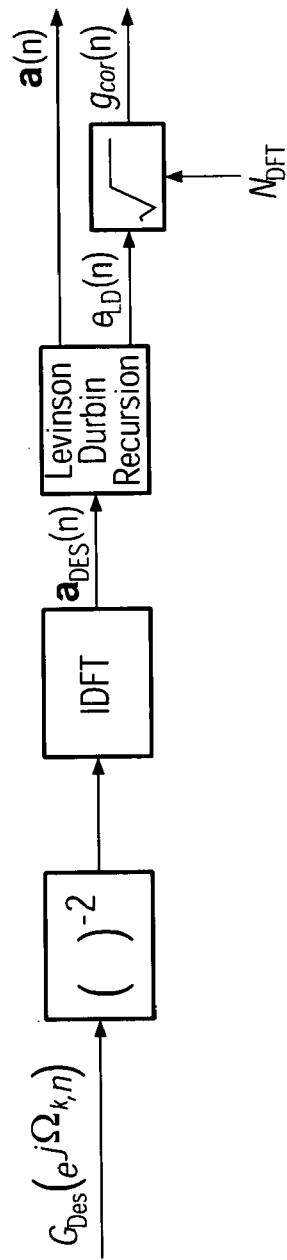
	DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	DE 101 41 493 A1 (INFINEON TECHNOLOGIES AG [DE]) 20 March 2003 (2003-03-20) * paragraphs [0011], [0016], [0020] - [0031] *	1,7	INV. G10L21/02 H04M1/60 H04M9/08
Y	* claims 1,5,7-10 *	2,5,6,8, 9,11-16	ADD. H04R3/00
Y	JP 2000 190785 A (FUJITSU TEN LTD) 11 July 2000 (2000-07-11) * paragraphs [0008], [0014], [0016] *	2,5,6	
Y	WOLFGANG TÄGER: "Etudes en traitement d'antenne pour la prise de son"[Online] 28 October 1998 (1998-10-28), XP002463250 Univ. Rennes I Retrieved from the Internet: URL:http://www.wolfgang-taeger.de/these04. pdf> [retrieved on 2007-12-20]	8,9, 11-16	
A	* pages 79-88 *	3,10	
A	EP 1 677 431 A (CIT ALCATEL [FR]) 5 July 2006 (2006-07-05) * abstract *	3,10	TECHNICAL FIELDS SEARCHED (IPC) H04M G10L

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	21 December 2007	Täger, Wolfgang

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

2

ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.

EP 07 01 9652

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

21-12-2007

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
DE 10141493	A1	20-03-2003	NONE		
JP 2000190785	A	11-07-2000	NONE		
EP 1677431	A	05-07-2006	CN	1797545 A	05-07-2006
			US	2006147030 A1	06-07-2006

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- EP 0204718 B1 **[0011]**
- EP 1020068 B1 **[0011]**

- WO 9626592 A **[0011]**

Non-patent literature cited in the description

- **E. HÄNSLER ; G. SCHMIDT.** Acoustic Echo and Noise Control. John Wiley & Sons, 2004 **[0004]** **[0055]**
- Microphone Arrays: Signal Processing Techniques and Applications. Springer, 2001 **[0006]**

- Adaptive beamforming for audio signal acquisition. **HERBORDT, W. ; KELLERMANN, W.** Adaptive signal processing: applications to real-world problems. Springer, 2003, 155 **[0023]**