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

(11) EP 1 885 154 A1

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
06.02.2008 Bulletin 2008/06

(51) Int Cl.:
H04R 3/04 *(2006.01)* ***G10L 21/02*** *(2006.01)*

(21) Application number: **06016029.8**

(22) Date of filing: **01.08.2006**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR
Designated Extension States:
AL BA HR MK YU

(71) Applicant: **Harman Becker Automotive Systems GmbH**
76307 Karlsbad (DE)

(72) Inventors:
- **Buck, Markus**
  88400 Biberach (DE)
- **Schmidt, Gerhard Uwe**
  89081 Ulm (DE)
- **Haulick, Tim**
  89143 Blaubeuren (DE)

(74) Representative: **Grünecker, Kinkeldey, Stockmair & Schwanhäusser Anwaltssozietät**
Maximilianstrasse 58
80538 München (DE)

(54) **Dereverberation of microphone signals**

(57) The present invention relates to a method for dereverberation of a microphone signal, comprising the steps of dividing the microphone signal into frames or sub-bands providing at least one loudspeaker signal estimating the reverberation energy of at least some of the frames or sub-bands on the basis of the at least one loudspeaker signal and filtering the microphone signal on the basis of the estimated reverberation energy of the at least some of the frames or sub-bands. The invention also relates to a signal processing means, comprising at least one microphone configured to obtain a microphone signal, at least one loudspeaker configured to output a loudspeaker signal, a reverberation estimating means configured to estimate the reverberation energy of the reverberation portion in the microphone signal on the basis of the loudspeaker signal and a dereverberation filtering means configured to receive the microphone signal and to reduce a reverberation portion in the microphone signal on the basis of the estimated reverberation energy.

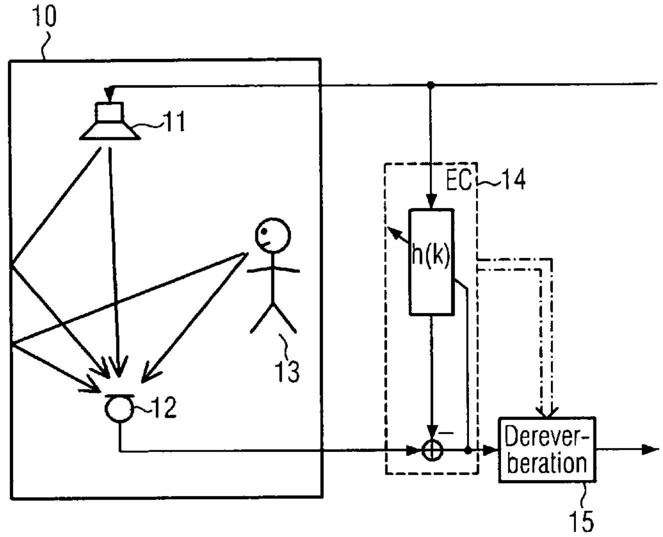

FIG. 2

Printed by Jouve, 75001 PARIS (FR)

Description

Field of Invention

[0001] The present invention relates to a system and a method for signal processing, in particular, speech signal processing, with dereverberation of a microphone signal. The invention particularly relates to dereverberation of a microphone signal detected in a loudspeaker-room-microphone system.

Background of the invention

[0002] The enhancement of the quality of audio and speech signals in a communication system is a central topic in acoustic and, in particular, speech signal processing. The communication between two parties is often carried out in a noisy background environment and noise reduction as well as echo compensation are necessary to guarantee intelligibility. Prominent examples are hands-free voice communication in vehicles and automatic speech recognition.

[0003] Of particular importance is the suppression of reverberation that can severely affect the quality of audio signals. A microphone used in an indoor communication system not only detects audio signals generated by an audio source directly but also detects reflections of the audio signals with some time delay due to the finite acoustic room, e.g. an office room or a vehicular cabin. The detected acoustic spectrum is, therefore, smeared over time. The intelligibility of speech signals can be significantly reduced by reverberant signal portions in microphone signals.

[0004] Thus, reverberation poses a severe problem, e.g., in automatic speech recognition and communication systems installed in vehicles, particular, since the reverberating characteristics of vehicular cabins as well as office rooms are rather complex and time-dependent.

[0005] Several methods for the dereverberation of microphone signals are known in the art. For example, it is attempted to reduce dereverberation by means of deconvolution, i.e. inverse filtering using an estimate for the acoustic channel. Deconvolution can be performed in the time domain or in the cepstral domain. However, this kind of signal processing suffers from the dependence on an accurate estimate of the acoustic channel which in practical applications is almost impossible.

[0006] According to an alternative approach, the direct-path speech signal is processed by pitch enhancement or by linear predictive coding (LPC) analysis. In a multi-channel approach averaging over multiple microphone signals is performed to obtain a reduction of the reverberation contribution to the processed signal. However, both approaches cannot guarantee a sufficiently high quality of the wanted signal. In addition, implementation of the multi-channel approach is rather expensive.

[0007] Despite the engineering process in recent years current dereverberation reduction is still not satisfying and reliable enough for practical applications.

[0008] It is therefore the problem underlying the present invention to overcome the above-mentioned drawbacks and to provide a system and a method for acoustic and speech signal processing exhibiting an improved dereverberation of microphone signals that is, in particular, suitable for hands-free telecommunication systems and automatic speech recognition systems.

Description of the Invention

[0009] The above-mentioned problem is solved by the method for dereverberation of a microphone signal according claim 1, comprising the steps of
dividing the microphone signal into frames or sub-bands;
providing at least one loudspeaker signal;
estimating the reverberation energy of at least some of the frames or sub-bands on the basis of the at least one loudspeaker signal; and
filtering the microphone signal on the basis of the estimated reverberation energy of the at least some of the frames or sub-bands.

[0010] Here and in the following the microphone signal is considered as a digital one, i.e. as an electric microphone signal that was already subject to A/D conversion. The microphone signal may preferably be transformed into the frequency domain and subsequently divided into the frames or each of the frames may separately be transformed into the frequency domain requiring multiple Fourier transform operation but each with less Fourier components to be calculated.

[0011] As a matter of fact, the overall processing may be performed in the frequency domain or after filtering the microphone signal by filter banks with sub-band signals according to the design properties and individual preferences. Moreover, the overall processing may be performed in the time domain with the microphone signal divided into frames.

[0012] The reverberation energy can be defined as the squared magnitude of the unwanted reverberant signal portion present in the microphone signal. According to the invention, in at least some frames or sub-bands the energy values of the reverberant signal portion are estimated. It is noted that herein the term "energy" is used for squared magnitudes of signals in the time as well as the frequency and sub-band domains. In effect, the filtering results in a subtraction of at least some part of the unwanted reverberant signal portion.

[0013] Processing for dereverberation is not necessary for all frames. In fact, the first frames almost exclusively exhibit the wanted signal (and, probably, some ambient noise perturbation), but rather no reverberation signal portion.

[0014] Different from the art not only the information contained in the microphone signal itself is used for the dereverberation but also the information of a loudspeaker signal. In the art, model parameters have to estimated for the dereverberation processing on the basis of the already reverberated microphone signals without any access to the original acoustic (speech) signals. According to the inventive method, dereverberation of a microphone can significantly be improved, since a loudspeaker signal can be considered as a reference signal by means of which the reverberation energy can be estimated.

[0015] In principle, there are several ways of using a loudspeaker signal for estimating the reverberation energy (see also discussion below). Given the loudspeaker signal on the one hand and the microphone signal in the other, the impulse response of the loudspeaker-room-microphone system can directly be determined. From the impulse response the reverberation energy can directly be derived, since the unwanted reverberant signal portion of a microphone signal can be represented as $r(n) = \sum_{l=D_t}^{\infty} x_c(n-l)h(l),$ with the discrete time index n, the loudspeaker signal $x_c(n)$ and the impulse response of the loudspeaker-room-microphone system h(n). The sum starts with some discrete time value D_t indicating the beginning of reflections of the detected acoustic signal causing the reverberation. In other word, for a time interval up to D_t the microphone signal is dominated by the energy of the wanted signal.

[0016] Thus, according to an embodiment the inventive method comprises estimating the impulse response of a loudspeaker-room-microphone system comprising the at least one loudspeaker providing the at least one loudspeaker signal and at least one microphone providing the microphone signal and wherein the reverberation energy of the at least some of the frames or sub-bands is estimated on the basis of the estimated impulse response of the loudspeaker-room-microphone system.

[0017] A particularly efficient way of estimating the impulse response is using an adaptive filtering means configured for automatic adaptation of the filter coefficients in order to model the impulse response. Therefore, the herein disclosed method may comprise filtering the microphone signal by means of an adaptive echo cancellation filtering means and wherein the impulse response of the loudspeaker-room-microphone system is determined from the adapted filter coefficients of the adaptive echo cancellation filtering means.

[0018] The echo cancellation filtering means can comprise linear or non-linear adaptive filter by which replica of acoustic feedback are synthesized and a compensation signal is obtained from the received signal of the loudspeakers. This compensation signal is subtracted from the sending signal of the microphone thereby generating a resulting echo reduced signal to be sent to the remote subscriber (see, e.g., Acoustic Echo and Noise Control, E. Hänsler and G. Schmidt, John Wiley & Sons, New York, 2004). An adaptive finite impulse response (FIR) filter can, e.g., be employed. The adaptation of the filter coefficients may be carried out by the normalized least mean square (NLMS) algorithm.

[0019] The actual filtering of the microphone signal on the basis of the estimated reverberation energy of the at least some of the frames or sub-bands can preferably be performed by a Wiener filter that is a well-approved robust filtering means. However, other filters as, e.g., a magnitude filter, may be employed instead. In the case of signal processing in the frequency domain the microphone signal has to be Fourier transformed to obtain Fourier transformed signals $Y_\mu(k)$, where k and μ denote the frame number and the index of the frequency bin, respectively. It is noted that, in general, the microphone signal may be Fourier transformed before the division into frames or it may be divided into frames followed by Fourier transformations of each frame.

[0020] At least some of the Fourier transformed signals $Y_\mu(k)$ can be filtered by a Wiener filter $W_\mu(k) = 1 - |\hat{R}_\mu(k)|^2 / |Y_\mu(k)|^2$ to obtain filtered signals $\hat{X}_\mu(k)$ according to $\hat{X}_\mu(k)= W_\mu(k)\, Y_\mu(k)$, where $|\hat{R}_\mu(k)|^2$ denotes the estimated reverberation energy of the at least some of the frames.

[0021] In an embodiment wherein the signal processing is performed for sub-band signals obtained from the microphone signal by means of a filter bank the microphone signal is filtered by a filter-bank to obtain sub-band signals $Y_\mu(k)$, where k and μ denote the time index of the subsampled microphone signal that is filtered by the filter-bank and the index of the sub-band, respectively; and

at least some of the sub-band signals $Y_\mu(k)$ are filtered by a Wiener filter $W_\mu(k) = 1 - |\hat{R}_\mu(k) / |Y_\mu(k)|^2$ to obtain filtered signals $\hat{X}_\mu(k)$ according to $\hat{X}_\mu(k) = W_\mu(k)\, Y_\mu(k)$, where $|\hat{R}_\mu(k)|^2$ denotes the estimated reverberation energy of the at least some of the sub-bands.

[0022] As usual the microphone signal is subsampled due to the frame- or block-based processing. In the inventive method, an estimate for the reverberation energie values of the considered frames (sub-bands) is required. It is desirable to obtain an estimate for the reverberation energy as accurate as possible in order to guarantee reliable dereverberation.

[0023] According to one example, the reverberation energy $|\hat{R}_{\mu}(k)|^2$ of the at least some of the frames or sub-blocks is estimated according to the following formula

$$|\hat{R}_{\mu}(k)|^2 = |Y_{\mu}(k-D)|^2 A_{\mu} \exp(-\gamma_{\mu} D) + |\hat{R}_{\mu}(k-1)|^2 \exp(-\gamma_{\mu})$$

where D is a predetermined delay, A_{μ} is an amplitude representing the ratio of direct-path energy to reverberation energy and y_{μ} is a parameter determined on the basis of the at least one loudspeaker signal and wherein k denotes the frame number or the time index of the subsampled microphone signal, respectively, and μ denotes the index of the frequency bin or the index of the sub-band, respectively. The predetermined delay takes into account that the initial part of the microphone signal is dominated by the direct acoustic path, i.e. a significant reverberant signal portion is present after some delay D, e.g. $D \approx 30$ ms. The predetermined amplitude A_{μ} can, in principle, be estimated when the position of a speaking person relative to a loudspeaker is known, e.g., from a beamforming of microphone signals obtained by a microphone array (see discussion below). However, A_{μ} can be chosen as a real value of the range from 0.1 to 0.5.

[0024] The parameter y_{μ} can be determined in the time domain or the frequency domain or the sub-band domain. When determined in the frequency domain or the sub-band domain they can be averaged over frequencies or sub-bands in order to obtain a parameter y_{μ} that does not any longer depend on the frequency. A motivation of this formula is given in the detailed description with reference to Figure 1 below.

[0025] If an adaptive echo cancellation filtering means is employed, the parameter Y_{μ} may be determined as follows. The reverberation time T_{60}, which is defined as the time the reverberation needs to decay by 60 dB, is estimated form the energy decay curve ($EDC_L(n)$) given by the filter coefficients $\hat{h}_L(n)$ of echo cancellation filtering means

$$EDC_L(n) = \frac{\sum_{i=n}^{L_L} \left|\hat{h}_L(i)\right|^2}{\sum_{i=0}^{L_L} \left|\hat{h}_L(i)\right|^2}$$

where L_L denotes the length of the echo cancellation filtering means. The slope of the $EDC_L(n)$ is related (inverse proportional) to the reverberation time T_{60}.

[0026] Then, the parameter y_{μ} can be calculated as a function of the reverberation time and a subsampling rate R_s by which the microphone signal is sub-sampled for frame or sub-block processing, in particular, according to the formula $y_{\mu} = 6 \ln 10\, R_s / T_{60} f_s$, where f_s denotes the sampling rate of the microphone signal in the time domain (see also description below).

[0027] In an alternative embodiment, the microphone signal is also filtered by means of a filtering means and wherein the filter coefficients of the filtering means are determined from the adapted filter coefficients of an adaptive echo cancellation filtering means in order to obtain an estimate for the reverberant signal portion.

[0028] In particular, according to one example of the herein disclosed method of dereverberation of a microphone signal, the reverberation energy is estimated more directly, namely by filtering the microphone signal by an adaptive filtering means having only filter coefficients that represent the impulse response of the loudspeaker-room-microphone system corresponding to the reverberation tail of the microphone signal to estimate a reverberation portion of the microphone signal. Then, the reverberation energy is estimated by means of the estimated reverberation portion of the microphone signal (see also detailed description with reference to Figure 3). By reverberation tail the frames or sub-blocks are referred to that include significant reverberation rather than the direct acoustic path from a sound or speech source to the detecting microphone(s).

[0029] The above-mentioned problem is also solved by the signal processing means according to claim 11, comprising at least one microphone configured to obtain a microphone signal that includes a reverberation signal portion;
at least one loudspeaker configured to output a loudspeaker signal; and
a reverberation estimating means configured to estimate the reverberation energy of the reverberation portion in the microphone signal on the basis of the loudspeaker signal; and
a dereverberation filtering means configured to receive the microphone signal and to reduce the reverberation portion in the microphone signal on the basis of the estimated reverberation energy.

[0030] The signal processing means may further comprise a processing means configured to divide the microphone signal into frames and a Fourier transformation means configured to Fourier transform the microphone signal before division into the frames or to Fourier transform the individual frames or further comprising a filter bank configured to divide the microphone signals into sub-band microphone signals.

[0031] The dereverberation filtering means of the signal processing means may comprise a spectral subtraction means, in particular, a Wiener filter, configured to filter the microphone signal for at least some of the frames or to filter at least some of the sub-band microphone signals on the basis of the estimated reverberation energy as explained above.

[0032] The reverberation estimating means may be configured to estimate the reverberation energy $|\hat{R}_\mu(k)|^2$ of the at least some of the frames or sub-blocks according to the following formula

$$|\hat{R}_\mu(k)|^2 = |Y_\mu(k-D)|^2 A_\mu \exp(-\gamma_\mu D) + |\hat{R}_\mu(k-1)|^2 \exp(-\gamma_\mu)$$

where D is a predetermined delay, A_μ is a predetermined amplitude and Y_μ is a parameter determined on the basis of the at least one loudspeaker signal and wherein k denotes the frame number or the time index of the subsampled microphone signal, respectively, and μ denotes the index of the frequency bin or the index of the sub-band, respectively.

[0033] In particular, the reverberation estimating means can be configured to estimate the reverberation time T_{60} and to calculate the parameter Y_μ as a function of the reverberation time and a subsampling rate R_s by which the microphone signal is sub-sampled for frame or sub-block processing, in particular, according to the formula $Y_\mu = 6 \ln 10\, R_s / T_{60} f_s$, where f_s denotes the sampling rate of the microphone signal in the time domain.

[0034] According to an embodiment, the signal processing means further comprises an adaptive echo cancellation filtering means configured to estimate the impulse response of the loudspeaker-room-microphone system and wherein the dereverberation filtering means is configured to reduce the reverberation portion in the microphone signal on the basis of the estimated impulse response of the loudspeaker-room-microphone system.

[0035] The signal processing means according to another example is configured to directly estimate the reverberation energy in at least some of the frames or sub-blocks by an adaptive filtering means having filter coefficients modeling only the reverberation tail included in the microphone signal. The reverberation estimating means is, thus, differently configured to the ones mentioned above. Rather, according to this example it is provided a signal processing means as recited in claims 10, 11 or 12 further comprising

a microphone array comprising at least two microphones,

a beamforming means configured to receive microphone signals from the microphones of the microphone array and to obtain a beamformed microphone signal; and

an adaptive filtering means configured to filter the beamformed microphone signals only by filter coefficients that represent the impulse response of the loudspeaker-room-microphone system corresponding to the reverberation tail of the beamformed microphone signal and to estimate a reverberation portion of the beamformed microphone signal. In this example, the reverberation energy is estimated by means of the estimated reverberation portion of the microphone signal.

[0036] The beamformer combines multiple microphone input signals to one beamformed signal with an enhanced SNR. Adaptive and non-adaptive beamformers can be employed that are known in the art, see, e.g., "Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory" by H.L. van Trees, Wiley & Sons, New York 2002. The fixed beamformer improves the signals pre-processed, e.g., by a means for time delay compensation, using a fixed beam pattern. Adaptive processing methods are characterized by a permanent adaptation of processing parameters such as filter coefficients during operation of the system. In particular, beamformers can amplify a wanted signal according to the direction from which acoustic signals are detected. The position of a speaker relative to the microphones of the microphone array can be determined by a means of the beamformer (see also discussion below).

[0037] The signal processing means according to one of the above examples can advantageously be employed in a great variety of communication systems. It is particularly useful in a hands-free telephony system. In the art, hands-free telephony often suffers from strong reverberation. The intelligibility of speech signals detected and processed by the inventive signal processing means and subsequently transmitted to a remote communication party can be significantly enhanced.

[0038] The present invention also provides a speech recognition means comprising the signal processing means according to one of above examples as well as a speech dialog system and a voice control system comprising such a speech recognition means or the signal processing means according to one of above examples. Speech recognition often suffers from the degradation of the quality of speech signals due to reverberation. The reliability of the recognition result of a speech recognition means and the successful operation of a speech dialog system or speech control system are greatly enhanced by employment of one of the above examples of the inventive signal processing means.

[0039] 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.

[0040] Figure 1 illustrates basic steps of the herein disclosed method for dereverberation of a microphone signal comprising the steps of estimating the reverberation energy on the basis of the impulse response of a loudspeaker-room-microphone system and spectral subtraction.

[0041] Figure 2 illustrates an example of dereverberation of a microphone signal employing an echo cancellation filtering means.

[0042] Figure 3 illustrates an example of dereverberation of a microphone signal comprising the direct estimation of the reverberation energy used for spectral subtraction.

[0043] In the following example, it is assumed that signal processing is performed in the frequency domain (with the exception of a parameter describing the exponential decay in time of the reverberation energy, see below). Signal processing, however, could alternatively be performed in the sub-band regime.

[0044] With reference to Figure 1 an audio signal is detected by a microphone and A-D converted to obtain a digital microphone signal that is received 1 by a signal processing means and divided into multiple frames 2, e.g., of some 10 ms. The digital microphone signal is subject to a Short Time Fourier Transformation (STFT) for signal processing in the frequency domain 2. According to the present invention, at least one loudspeaker signal is provided by at least one loudspeaker that is used for the dereverberation of the microphone signal 3.

[0045] The impulse response of the loudspeaker-room-microphone system is estimated 4. According to the present example, an adaptive echo cancellation filtering means is employed. The filter coefficients of the echo cancellation filtering means are automatically adjusted to model the impulse response of the loudspeaker-room-microphone system. By means of the adapted filter coefficients the reverberation energy of the frames can be estimated 5. In detail, in this example this is achieved as follows.

[0046] Due to the effect of reverberation the detected acoustic spectrum is smeared over time. The smearing of the energy of the microphone signal can be modeled by

$$\left|Y_\mu(k)\right|^2 \approx \sum_{l=0}^{\infty}\left|X_{c,\mu}(k-l)\right|^2 G_\mu(l) \approx \left|X_\mu(k)\right|^2 + \left|R_\mu(k)\right|^2$$

where $X_{c,\mu}$ is the Fourier transformed signal emitted by the speaking person (clean speech signal) and G_μ models the energy decay of the impulse response of the loudspeaker-room-microphone system in the frequency domain and where it is assumed that the wanted signal $X_\mu(k)$ and the reverberation signal portion $R_\mu(k)$ are uncorrelated. The energy decay comprises a first part corresponding to a number of initial frames, say D frames, that exhibit no significant reverberation and a part contributing to the reverberation signal portion:

$$\left|R_\mu(k)\right|^2 \approx \sum_{l=D}^{\infty}\left|X_{c,\mu}(k-l)\right|^2 G_\mu(l).$$

[0047] The reverberation energy is to be determined for the spectral subtraction filtering of the microphone signal. Assuming an exponential decay of the reverberation energy for $k > 0$ (and $G_\mu(k) = 1$ for $k = 0$) the reverberation energy can be represented as follows:

$$|R_\mu(k)|^2 \approx \sum_{l=D}^{\infty} |X_{c,\mu}(k-l)|^2 A_\mu \exp(-\gamma_\mu l) = \sum_{m=-\infty}^{k-D} |X_{c,\mu}(m)|^2 A_\mu \exp(-\gamma_\mu (k-m))$$

$$= |X_{c,\mu}(k-D)|^2 A_\mu \exp(-\gamma_\mu D) + \sum_{m=-\infty}^{k-1-D} |X_{c,\mu}(m)|^2 A_\mu \exp(-\gamma_\mu (k-1-m)) \exp(-\gamma_\mu)$$

$$= |X_{c,\mu}(k-D)|^2 A_\mu \exp(-\gamma_\mu D) + |R_\mu(k-1)|^2 \exp(-\gamma_\mu)$$

where D is a fixed delay by D frames and Y_μ denotes the exponential decay parameter depending on room parameters as, e.g., the room size and absorption characteristics. The parameter A_μ denotes the ratio of the energy of the direct acoustic path from a source to the microphone to the reverberation energy that mainly depends on the position of a sound source or a speaking person relative to the microphone.

[0048] Approximating the speaking person's clean speech signal $|X_{c\mu}(k\text{-}D)|^2$ by the reverberated microphone signal $|Y_\mu(k\text{-}D)^2|$ an estimate for the reverberation energy can be calculated by the recursive formula

$$|\hat{R}_\mu(k)|^2 = |Y_\mu(k-D)|^2 A_\mu \exp(-\gamma_\mu D) + |\hat{R}_\mu(k-1)|^2 \exp(-\gamma_\mu).$$

[0049] The exponential decay parameter Y_μ according to the present example is determined in the time domain as follows (determination in the sub-band regime or the frequency domain would be possible correspondingly). Due to the determination in the time domain y_μ is the same for all μ.

[0050] The above-mentioned echo cancellation filtering means is employed for filtering the microphone signal. The reverberation time T_{60}, which is defined as the time the reverberation needs to decay by 60 dB, is estimated form the energy decay curve ($EDC_L(n)$) given by the filter coefficients $\hat{h}_L(n)$ of the echo cancellation filtering means after convergence of the employed adaptation algorithm

$$EDC_L(n) = \frac{\sum_{i=n}^{L_L} |\hat{h}_L(i)|^2}{\sum_{i=0}^{L_L} |\hat{h}_L(i)|^2}$$

where L_L denotes the length of the echo cancellation filtering means. The $EDC_L(n)$ can be interpreted as representing the total amount of signal energy remaining in the reverberator impulse response at time n. The slope of the $EDC_L(n)$ is estimated as follows. An upper and a lower threshold for a range of values of the $EDC_L(n)$, e.g., E_{max}= -20 dB and E_{min} = -40 dB, are chosen. The discrete time indices n_1 and n_2 are determined for which the $EDC_L(n)$ exhibits values closest to E_{max} and E_{min}, respectively. The reverberation time T_{60} can, then, be determined by extrapolation of the slope of the $EDC_L(n)$:

$$T_{60} = \frac{1}{f_S} \frac{n_2 - n_1}{E_{max} - E_{min}} 60\,\text{dB}$$

where f_s denotes the sampling rate of the digital microphone signal in the time domain. Given an exponential decay of the reverberation energy $\exp(-y_\mu k)$ an energy decrease by 10^{-6} (after the reverberation time T_{60}) implies $\exp(-y_\mu\ T_{60}\ f_s$

$/ R_s) = 10^{-6}$, where R_s denotes a subsampling rate of the processed microphone signal (divided into frames or sub-blocks) due to the frame based or sub-block based overall processing. Accordingly, the exponential decay parameter is given by

$$\gamma_\mu = \frac{6 \ln 10}{T_{60}\, f_S} R_S.$$

[0051] It is obvious that this expression can also be used, if the impulse response is determined differently, i.e. without employment of an echo cancellation filtering means.

[0052] The thus estimated reverberation energy is used for a spectral subtraction 6 in order to obtain a dereverberated microphone signal. The spectral subtraction can be performed by a Wiener filter. The microphone signal in each frame $Y_\mu(k)$ is supposed to consist of two assumingly uncorrelated parts (other contributions as, e.g., ambient noise are neglected here for simplicity): the wanted signal $X_\mu(k)$ and the reverberant signal contribution $R_\mu(k)$ where k and μ denote the frame number and the index of the frequency bin.

[0053] According to the spectral subtraction employed to achieve the dereverberated microphone signal $\hat{X}_\mu(k)$ the amplitudes of the microphone signal in each frame $Y_\mu(k)$ are scaled with real valued coefficients $W_\mu(k)$: $\hat{X}_\mu(k) = W_\mu(k)\, Y_\mu(k)$. In this example, a Wiener filter is used for the coefficients $W_\mu(k)$:

$$\hat{X}_\mu(k) = \left(1 - \frac{\left|\hat{R}_\mu(k)\right|^2}{\left|Y_\mu(k)\right|^2}\right) Y_\mu(k).$$

with $\hat{R}_\mu(k)$ being determined by the above recursion formula.

[0054] The signal $\hat{X}_\mu(k)$ is eventually output 7. It may represent an enhanced microphone signal that is to be transmitted to a remote communication party in a hands-free telephony system. It can also represent an enhanced speech input for a speech recognition or voice control system.

[0055] Figure 2 illustrates parts of the inventive signal processing means. A loudspeaker-room-microphone system 10 is built by a loudspeaker 11, a microphone 11 and some closed surrounding as, e.g., a vehicular cabin or an office room. The microphone is used to detect a speaker's 13 utterances. The speech signals are detected in a direct path but also reflections from the surrounding are detected by the microphone 12.

[0056] The illustrated system also comprises an echo cancellation filtering means 14. The filter coefficients of the echo cancellation filtering means 14 are adapted to model the impulse response of the loudspeaker-room-microphone system. Thus, the information on the filter coefficients (after convergence) can be input in a dereverberation means 15 configured to perform dereverberation, e.g., in form of the spectral subtraction specified above. Note, that the impulse response of the speaker (speaking person) - room - microphone system, which is unknown, is approximated by the determined impulse response of the loudspeaker-room-microphone system. Thus, is desirable that the loudspeaker is as close to the speaking person as possible.

[0057] Figure 3 illustrates in some detail a signal processing means comprising loudspeakers 21, at least two microphones 22, a beamformer 23 and an adaptive filtering means 24 receiving a loudspeaker signal and filtering the beamformed microphone signal obtained by the beamformer 23 from the individual microphone signals of the microphones 22. The beamformer 23 may be a conventional delay-and-sum beamforming means.

[0058] A configuration as shown in Figure 3 can be used to directly estimate the reverberation energy of the microphone signal that is divided into frames without making use of the above recursion formula. The adaptive filtering means 24 comprises filter coefficients that model the impulse response of the loudspeaker-room-microphone system after convergence of the adaptation algorithm has been reached. For the direct estimate of the reverberation energy a subset $\tilde{h}$ of the filter coefficients $\hat{h}$ of the adaptive filtering means 24 matching the time section that correspond to the reverberation tail only is extracted and passed to another filtering means 25. For this filtering means 25, therefore, the direct acoustic path from the source of audio signals that are to be detected to the microphones 22 is not included.

[0059] By filtering of the beamformed microphone signal by the filtering means 25 an estimate for the reverberant signal portion of the beamfomed microphone signal and, thus, an estimate for the reverberation energy can be obtained. The output of the filtering means 25 and the beamformed microphone signal are input into a spectral subtraction section

26 that is configured to perform the processing described above, i.e. the beamformed microphone signal is subject to filtering by means of the above described Wiener filter on the basis of the estimated reverberation energy.

[0060] In this example, as well as in the examples described above, it must be assumed that the impulse response of the loudspeaker-to-microphone transfer is similar to the one from the speaking person (or, in general, the audio signal generating means generating audios signals that are to be detected) to the microphone which, in principle, is unknown. Thus, if a plurality of loudspeakers is present, it is important that the loudspeaker that is closest to the speaking person is chosen for an estimate of the impulse response of the loudspeaker-room-microphone system $\hat{h}$.

[0061] However, in order to obtain a reliable estimate for the reverberation energy the acoustic time delays have to be matched by the subset of filter coefficients $\tilde{h}$ chosen for estimating the reverberant portion of the beamformed microphone signal by the filtering means 25. Thus, the subset of filter coefficients $\tilde{h}$ is subject to a delay by D_h samples. In addition, the energy of the estimated reverberation energy can be adjusted by some factor b_h.

[0062] The parameters D_h and b_h that allow for a better estimate of the reverberation energy can be determined, if the actual position of the person uttering speech signals relative to the microphones 22 is known. In the present example, this position is determined by the beamformer 23 in which a source localization algorithm as known in the art is implemented.

[0063] 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 dereverberation of a microphone signal, comprising the steps of
dividing the microphone signal into frames or sub-bands;
providing at least one loudspeaker signal;
estimating the reverberation energy in at least some of the frames or sub-bands on the basis of the at least one loudspeaker signal; and
filtering the microphone signal on the basis of the estimated reverberation energy of the at least some of the frames or sub-bands.

2. Method according to claim 1, further comprising
estimating the impulse response of a loudspeaker-room-microphone system comprising the at least one loudspeaker providing the at least one loudspeaker signal and at least one microphone providing the microphone signal; and wherein
the reverberation energy of the at least some of the frames or sub-bands is estimated on the basis of the estimated impulse response of the loudspeaker-room-microphone system.

3. Method according to claim 2, further comprising filtering the microphone signal by means of an adaptive echo cancellation filtering means and wherein the impulse response of the loudspeaker-room-microphone system is determined from the adapted filter coefficients of the adaptive echo cancellation filtering means.

4. Method according to one of the preceding claims, wherein the microphone signal is transformed into the frequency domain and subsequently divided into the frames or wherein the microphone signal is divided into the frames that subsequently are transformed into the frequency domain.

5. Method according to claim 4, wherein
the microphone signal is Fourier transformed to obtain Fourier transformed signals $Y_\mu(k)$, where k and μ denote the frame number and the index of the frequency bin, respectively; and
at least some of the Fourier transformed signals $Y_\mu(k)$ are filtered by a Wiener filter $W_\mu(k) = 1 - |\hat{R}_\mu(k)|^2 / |Y_\mu(k)|^2$ to obtain filtered signals $\hat{X}_\mu(k)$ according to $\hat{X}_\mu(k) = W_\mu(k)\, Y_\mu(k)$, where $|\hat{R}_\mu(k)|^2$ denotes the estimated reverberation energy of the at least some of the frames.

6. Method according to claim 4, wherein
the microphone signal is filtered by a filter-bank to obtain sub-band signals $Y_\mu(k)$, where k and μ denote the time index of the subsampled microphone signal that is filtered by the filter-bank and the index of the sub-band, respectively; and
at least some of the sub-band signals $Y_\mu(k)$ are filtered by a Wiener filter $W_\mu(k) = 1 - |\hat{R}_\mu(k)|^2 / |Y_\mu(k)|^2$ to obtain filtered signals $\hat{X}_\mu(k)$ according to $\hat{X}_\mu(k) = W_\mu(k)\, Y_\mu(k)$, where $|\hat{R}_\mu(k)|^2$ denotes the estimated reverberation energy

of the at least some of the sub-bands.

7. Method according to one of the preceding claims, wherein the reverberation energy $|\hat{R}_{\mu}(k)|^2$ of the at least some of the frames or sub-blocks is recursively estimated according to the following formula

$$|\hat{R}_{\mu}(k)|^2 = |Y_{\mu}(k-D)|^2 A_{\mu} \exp(-\gamma_{\mu} D) + |\hat{R}_{\mu}(k-1)|^2 \exp(-\gamma_{\mu})$$

where D is a predetermined delay, A_{μ} is a predetermined amplitude and Y_{μ} is a parameter determined on the basis of the at least one loudspeaker signal and wherein k denotes the frame number or the time index of the subsampled microphone signal, respectively, and μ denotes the index of the frequency bin or the index of the sub-band, respectively.

8. Method according to claim 7, further comprising estimating reverberation time T_{60} and wherein the parameter y_{μ} is calculated as a function of the reverberation time and a subsampling rate R_s by which the microphone signal is subsampled for frame or sub-block processing, in particular, according to the formula $y_{\mu} = 6 \ln 10\ R_s / T_{60}\ f_s$, where f_s denotes the sampling rate of the microphone signal in the time domain.

9. Method according to one of the claims 1 - 4, further comprising filtering the microphone signal by means of a filtering means and wherein the filter coefficients are determined from the adapted filter coefficients of an adaptive echo cancellation filtering means.

10. Method according to one of the claims 1 - 4, further comprising
filtering the microphone signal by an adaptive filtering means having only filter coefficients that represent the impulse response of the loudspeaker-room-microphone system corresponding to the reverberation tail of the microphone signal to estimate a reverberation portion of the microphone signal; and wherein
the reverberation energy is estimated by means of the estimated reverberation portion of the microphone signal.

11. Signal processing means, comprising
at least one microphone configured to obtain a microphone signal;
at least one loudspeaker configured to output a loudspeaker signal;
a reverberation estimating means configured to estimate the reverberation energy of a reverberation portion in the microphone signal on the basis of the loudspeaker signal; and
a dereverberation filtering means configured to receive the microphone signal and to reduce the reverberation portion in the microphone signal on the basis of the estimated reverberation energy.

12. The signal processing means according to claim 11, further comprising a processing means configured to divide the microphone signal into frames and a Fourier transformation means configured to Fourier transform the microphone signal or further comprising a filter bank configured to divide the microphone signals into sub-band microphone signals.

13. The signal processing means according to claim 11 or 12, wherein the dereverberation filtering means comprises a spectral subtraction means, in particular, a Wiener filter, configured to filter the microphone signal for at least some of the frames or to filter at least some of the sub-band microphone signals on the basis of the estimated reverberation energy.

14. The signal processing means according to one of the claims 11 - 13, wherein the reverberation estimating means configured to recursively estimate the reverberation energy $|\hat{R}_{\mu}(k)|^2$ of the at least some of the frames or sub-blocks according to the following formula

$$|\hat{R}_{\mu}(k)|^2 = |Y_{\mu}(k-D)|^2 A_{\mu} \exp(-\gamma_{\mu} D) + |\hat{R}_{\mu}(k-1)|^2 \exp(-\gamma_{\mu})$$

where D is a predetermined delay, A_{μ} is a predetermined amplitude and Y_{μ} is a parameter determined on the basis of the at least one loudspeaker signal and wherein k denotes the frame number or the time index of the subsampled microphone signal, respectively, and μ denotes the index of the frequency bin or the index of the sub-band, respectively.

15. The signal processing means according to claim 14, wherein the reverberation estimating means is configured to estimate the reverberation time T_{60} and to calculate the parameter y_μ as a function of the reverberation time and a subsampling rate R_s by which the microphone signal is sub-sampled for frame or sub-block processing, in particular, according to the formula $y_\mu = 6 \ln 10\, R_s / T_{60} f_s$, where f_s denotes the sampling rate of the microphone signal in the time domain.

16. The signal processing means according to one of the claims 11 - 15, further comprising an adaptive echo cancellation filtering means configured to estimate the impulse response of the loudspeaker-room-microphone system and wherein the dereverberation filtering means is configured to reduce the reverberation portion in the microphone signal on the basis of the estimated impulse response of the loudspeaker-room-microphone system.

17. The signal processing means according to one of the claims 11 - 13, further comprising
a microphone array comprising at least two microphones,
a beamforming means configured to receive microphone signals from the microphones of the microphone array and to obtain a beamformed microphone signal; and
an adaptive filtering means configured to filter the beamformed microphone signals only by filter coefficients that represent the impulse response of the loudspeaker-room-microphone system corresponding to the reverberation tail of the beamformed microphone signal and to estimate a reverberation portion of the beamformed microphone signal; and wherein
the reverberation estimating means configured to estimate the reverberation energy of the reverberation portion in the microphone signal by means of the estimated reverberation portion of the microphone signal.

18. Hands-free telephony system, comprising the signal processing means according to one of the claims 11 - 17.

19. Speech recognition means comprising the signal processing means according to one of the claims 11 - 17.

20. Speech dialog system or voice control system comprising the speech recognition means according to claim 19 or the signal processing means according to one of the claims 11 - 17.

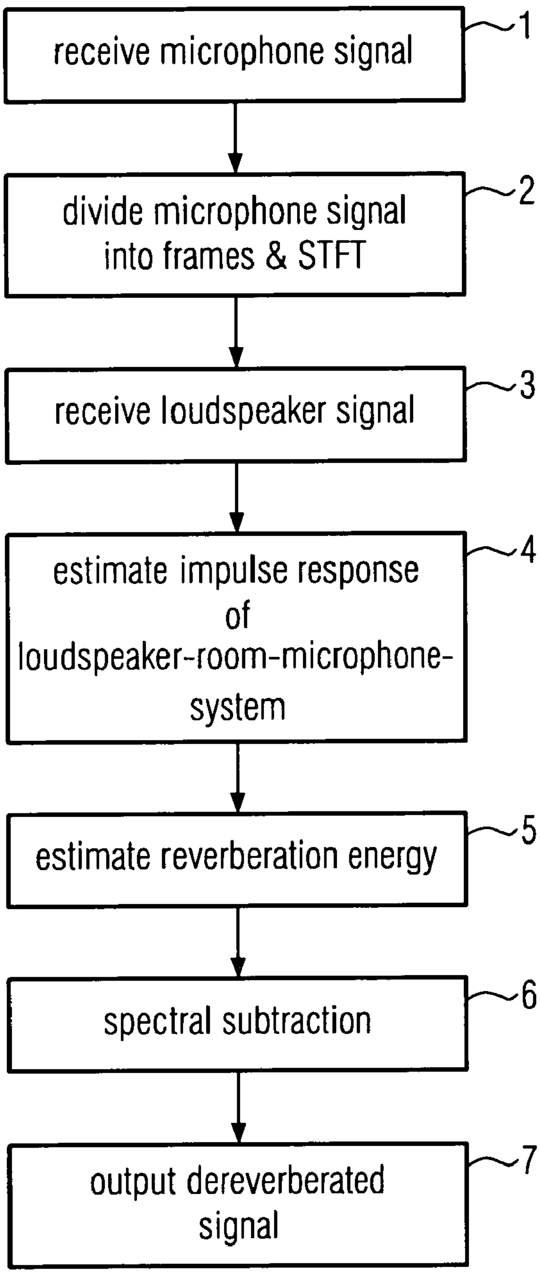
receive microphone signal
1
divide microphone signal into frames & STFT
2
receive loudspeaker signal
3
estimate impulse response of loudspeaker-room-microphone-system
4
estimate reverberation energy
5
spectral subtraction
6
output dereverberated signal
7

FIG. 1

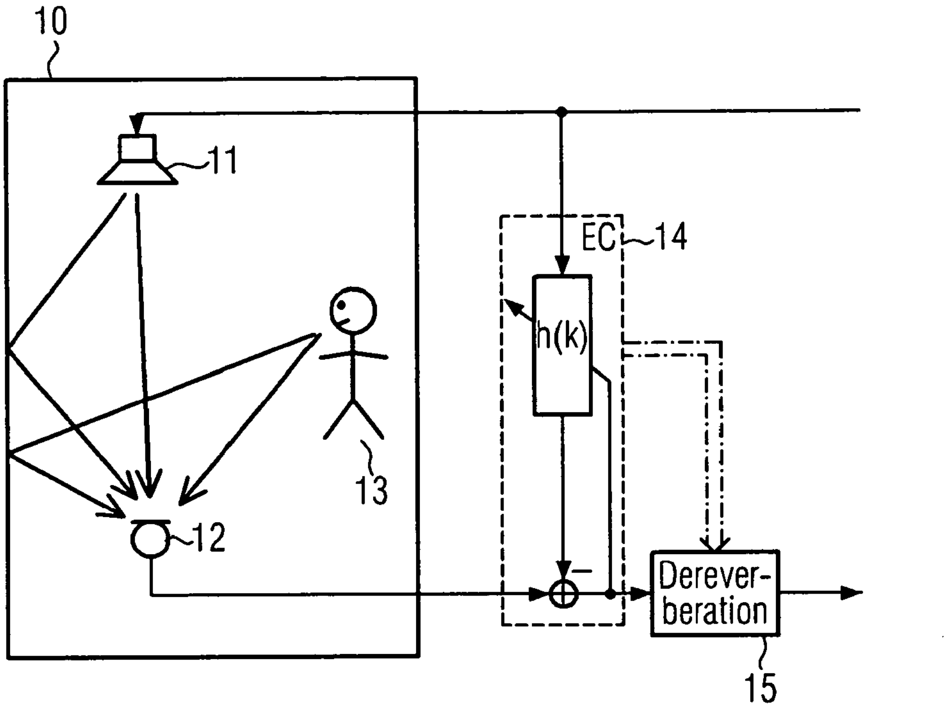
10
11
EC
14
h(k)
13
12
Dereverberation
15

FIG. 2

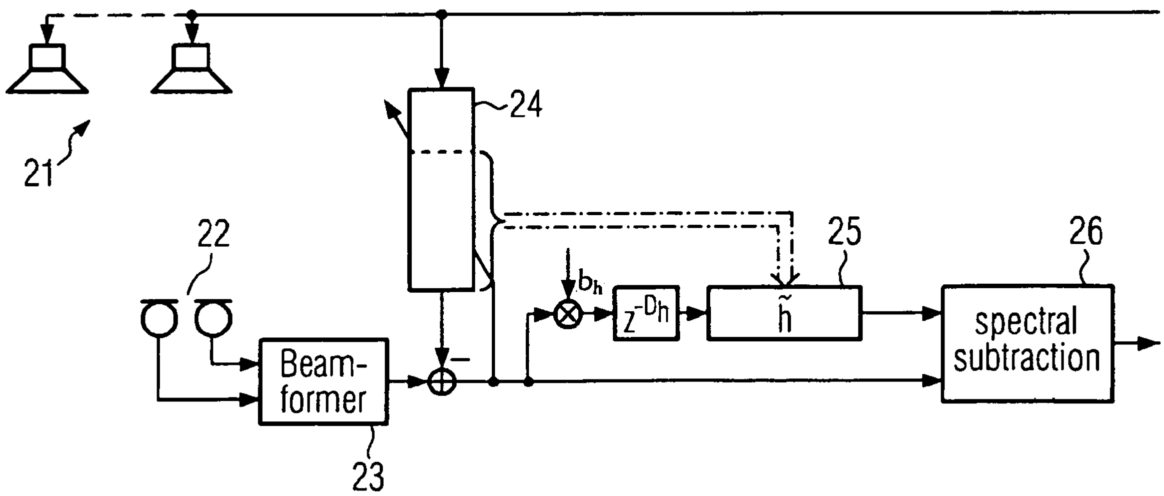
24
21
22
25
26
b_h
z^{-D_h}
$\tilde{h}$
spectral subtraction
Beamformer
23

FIG. 3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 06 01 6029

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 A	EP 1 521 240 A (SIEMENS AG [DE]) 6 April 2005 (2005-04-06) * figures 1,4 * * paragraph [0006] - paragraph [0010] * * paragraph [0022] - paragraph [0025] * * paragraph [0037] - paragraph [0039] * * paragraph [0045] - paragraph [0046] * * paragraph [0050] * * paragraph [0082] - paragraph [0091] * * claim 1 * -----	1-6,9, 11-13, 16,18-20 7,8,10, 14,15,17	INV. H04R3/04 G10L21/02
X A	LEBART K ET AL: "A NEW METHOD BASED ON SPECTRAL SUBTRACTION FOR SPEECH DEREVERBERATION" May 2001 (2001-05), ACUSTICA, S. HIRZEL VERLAG, STUTTGART, DE, PAGE(S) 359-366 , XP009053193 ISSN: 0001-7884 chapters 1,2 subchapters 3.1, 3.3 -----	1-6,9, 11-13, 16,18-20 7,8,10, 14,15,17	TECHNICAL FIELDS SEARCHED (IPC) G10L H04R
X A	WO 2006/011104 A (KONINKL PHILIPS ELECTRONICS NV [NL]; DERKX RENE M M [NL]; JANSE CORNEL) 2 February 2006 (2006-02-02) * page 1, line 1 - page 2, line 22 * * page 4, line 16 - page 9, line 14 * ----- -/--	1-6,9, 11-13, 16,18-20 7,8,10, 14,15,17	

The present search report has been drawn up for all claims

2

Place of search	Date of completion of the search	Examiner
The Hague	13 April 2007	Moscu, Viorel

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

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 06 01 6029

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	M. R. SCHROEDER: "New Method of Measuring Reverberation Time" THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, [Online] vol. 37, no. 3, 14 December 1964 (1964-12-14), - March 1965 (1965-03) pages 409-412, XP002429121 Retrieved from the Internet: URL:http://scitation.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=JASMAN000037000003000409000001&idtype=cvips&prog=normal> [retrieved on 2007-04-12] * the whole document * -----	1-20	
A	US 2005/244023 A1 (ROECK HANS-UELI [CH] ET AL) 3 November 2005 (2005-11-03) * the whole document * -----	1-20	
			TECHNICAL FIELDS SEARCHED (IPC)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
The Hague	13 April 2007	Moscu, Viorel

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

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 06 01 6029

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.

13-04-2007

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1521240 A	06-04-2005	WO 2005031707 A1	07-04-2005
WO 2006011104 A	02-02-2006	NONE	
US 2005244023 A1	03-11-2005	NONE	

EPO FORM P0459

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.

Non-patent literature cited in the description

- **E. HÄNSLER ; G. SCHMIDT.** Acoustic Echo and Noise Control. John Wiley & Sons, 2004 **[0018]**
- **H.L. VAN TREES.** Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory. Wiley & Sons, 2002 **[0036]**