



(11) **EP 2 031 583 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
06.01.2010 Bulletin 2010/01

(51) Int Cl.:
G10L 21/02^(2006.01) H04R 25/00^(2006.01)

(21) Application number: **07017134.3**

(22) Date of filing: **31.08.2007**

(54) **Fast estimation of spectral noise power density for speech signal enhancement**

Schnelle Schätzung der Spektraldichte der Rauschleistung zur Sprachsignalverbesserung

Estimation rapide de la densité spectrale de puissance de bruit pour l'amélioration d'un signal vocal

(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 MT NL PL PT RO SE SI SK TR

(43) Date of publication of application:
04.03.2009 Bulletin 2009/10

(73) Proprietor: **Harman Becker Automotive Systems GmbH**
76307 Karlsbad (DE)

(72) Inventors:
• **Schmidt, Gerhard Uwe**
89081 Ulm (DE)
• **Wolff, Tobias**
89073 Ulm (DE)
• **Buck, Markus**
88400 Biberach (DE)

(74) Representative: **Grünecker, Kinkeldey, Stockmair & Schwanhäusser**
Anwaltssozietät
Leopoldstrasse 4
80802 München (DE)

(56) References cited:
GB-A- 2 426 167 US-A1- 2003 128 851
US-A1- 2007 185 711

- **MARTIN ET AL: "Bias compensation methods for minimum statistics noise power spectral density estimation" SIGNAL PROCESSING, ELSEVIER SCIENCE PUBLISHERS B.V. AMSTERDAM, NL, vol. 86, no. 6, June 2006 (2006-06), pages 1215-1229, XP005365809 ISSN: 0165-1684**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 2 031 583 B1

Description

[0001] The invention is directed to a method and apparatus for providing an estimate of a spectral noise power density of an audio signal, in particular, a speech signal.

5 [0002] Acquiring the voice signal of a speaker by microphones often suffers from noise, which is due to a noisy environment and adds to the clean voice signal resulting in a disturbed acoustic signal. In case of hands free telephoning, for example, the voice signal may be interfered by noise such as background noise and echo components. In the case of a vehicle, the background noise may be composed of the noise of the engine, the windstream, and the rolling tires. In addition, unwanted signal components may be due to sound from loudspeakers, reproducing the output either of a

10 [0003] radio or of a hands-free telephony application, which may result in echoes. [0003] The performance of speech recognition software is diminished by such noise. In hands free telephoning applications, noise reduces communication quality and intelligibility.

[0004] Hence, there is a need to reduce noise in audio signals. To this end, noise reduction filters are being used. Usually, the audio signal is split into frequency bands by a filter bank. Noise reduction is then performed in each frequency band separately. The noise reduced signal is finally synthesized from the modified spectrum by a synthesizing filter bank, which transforms the signal back into the time domain.

15 [0005] A possible algorithm for noise reduction is based on estimates of the spectral power density of the distorted audio signal and that of the noise component. Depending on the ratio of both quantities, a weighting factor is applied in the distorted frequency band. The relation between the spectral signal power and the weighting factor is influenced by the filter characteristics.

[0006] The filters rely on a good estimate of the spectral noise power density. The estimate should be as close as possible to the actual or current noise power density. The quality of this estimate influences the overall performance of the filter.

20 [0007] GB 2 426 167 discloses a quantile based noise estimation in which a recursive function is applied to generate an estimated noise power spectrum.

[0008] Therefore, there is the problem to improve the known methods and apparatuses, which provide an estimate of the spectral noise power density, in such a way that they provide an estimate which more closely resembles the actual or current noise power density.

[0009] This problem is solved by the method according to claim 1 and the apparatus according to claim 12.

30 [0010] Accordingly, as set forth in independent claim 1, a method for providing an estimate of a spectral noise power density of an audio signal is provided, comprising:

providing a first estimate of a spectral noise power density of the audio signal,

35 determining a time dependent correction term,

40 summing the first estimate and the correction term to obtain a second estimate of the spectral noise power density of the audio signal, wherein the correction term is determined such that a spectral noise power density estimation error is reduced, and wherein the audio signal comprises a wanted signal component and a noise component and the correction term is based on the expectation value of the squared difference of the current spectral noise power density and the first estimate of the spectral noise power density of the audio signal and on the expectation value of the squared spectral power density of the wanted signal component.

45 [0011] The above-described method advantageously provides an estimate (the second estimate) of the spectral noise power density which resembles the current or actual noise power density much better than that of the prior art. The second estimate of the spectral noise power density according to the above-described method may be used in many applications and filters.

[0012] The audio signal is an electrical signal; it may be a digital or digitized signal. The audio signal may be based on an acoustic signal received by one or more microphones, and digitized by an Analog-to-Digital Converter (ADC). The step of providing the first estimate of a spectral noise power density of the audio signal may be preceded by one or more steps of filtering the signal. In the above-described method, the step of providing a first estimate of a spectral noise power density of the audio signal may be preceded by processing the audio signal by one or more filters or other processing units, like, e.g. a beam-former.

55 [0013] Some or all steps of the above-described methods may be performed in the frequency domain. In particular, signals may be transformed into the frequency domain by well-known techniques such as Discrete Fourier Transform (DFT), Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT) or wavelet transform.

[0014] In the above-described methods, the correction term comprises a spectral power density estimation error. In this way, the correction term may be small if the estimation error is small.

[0015] In the above-described methods, the correction term may comprise a product of a correction factor and the spectral power density estimation error. In particular, the second estimate of the spectral noise power density may take the form:

5

$$\hat{S}_{bb}(\Omega_{\mu}, n) = \tilde{S}_{bb}(\Omega_{\mu}, n) + K(\Omega_{\mu}, n) \cdot E_p(\Omega_{\mu}, n),$$

10

where $\tilde{S}_{bb}(\Omega_{\mu}, n)$ denotes the first estimate of the spectral noise power density, $\tilde{S}_{bb}(\Omega_{\mu}, n)$ denotes the second estimate of the spectral power density, $E_p(\Omega_{\mu}, n)$ denotes the spectral power density estimation error and $K(\Omega_{\mu}, n)$ denotes the correction factor. n is the time variable and Ω_{μ} is the frequency variable with frequency-index μ . In particular, the frequency variable may be frequency supporting points in the case of frequency bands. The frequency supporting points Ω_{μ} may be equally spaced or may be distributed non-uniformly.

15

[0016] This form of the correction term provides a way to adapt the correction term such that certain constraints are fulfilled like e.g. the constraint that a spectral noise power density estimation error is reduced.

20

[0017] In the above-described methods, the audio signal comprises a wanted signal component and a noise component. The correction term is based on the expectation value of the squared difference of the current spectral noise power density and the first estimate of the spectral noise power density of the audio signal and on the expectation value of the squared spectral power density of the wanted signal component. In particular, the correction term may have the form:

25

$$K(\Omega_{\mu}, n) = \frac{E\{E_n^2(\Omega_{\mu}, n)\}}{E\{E_p^2(\Omega_{\mu}, n)\}} \\ = \frac{E\{E_n^2(\Omega_{\mu}, n)\}}{E\{E_n^2(\Omega_{\mu}, n)\} + E\{S_{xx}^2(\Omega_{\mu}, n)\}},$$

30

where $E\{\cdot\}$ denotes the operation of taking the expectation value, $S_{xx}(\Omega_{\mu}, n)$ denotes the spectral power density of the wanted signal component and

35

$$E_n(\Omega_{\mu}, n) = S_{bb}(\Omega_{\mu}, n) - \tilde{S}_{bb}(\Omega_{\mu}, n).$$

40

[0018] In the above-described methods, the spectral noise power density estimation error may be based on the deviation of the second estimate of the spectral noise power density of the audio signal from the current spectral noise power density of the audio signal. The deviation may be based on a difference and/or a metric. The current spectral noise power density is the actual spectral noise power density and, therefore, the words "current" and "actual" may be used interchangeably in this context. In particular, the spectral noise power density estimation error may have the form:

45

$$E\{\hat{E}_n^2(\Omega_{\mu}, n)\},$$

50

with $\hat{E}_n(\Omega_{\mu}, n) = S_{bb}(\Omega_{\mu}, n) - \hat{S}_{bb}(\Omega_{\mu}, n)$. Thus, if this error is reduced, the second estimate of the spectral noise power density is closer to the current spectral noise power density.

55

[0019] In the above-described methods, the correction term may be based on the variance of a relative spectral noise power density estimation error, on the first estimate of the spectral noise power density of the audio signal and on the current spectral power density of the audio signal. In particular, the correction term may have the form:

$$K(\Omega_\mu, n) = \frac{\sigma_{E_{rel}}^2 \cdot \tilde{S}_{bb}^2(\Omega_\mu, n)}{(S_{yy}(\Omega_\mu, n) - \tilde{S}_{bb}(\Omega_\mu, n))^2},$$

where $\sigma_{E_{rel}}^2$ denotes the variance of the error E_{rel} , in relation to $\tilde{S}_{bb}(\Omega_\mu, n)$, i.e. $\sigma_{E_{rel}}^2 = \sigma_{E_n}^2 / \tilde{S}_{bb}(\Omega_\mu, n)$ and

$S_{yy}(\Omega_\mu, n)$ denotes the spectral power density of the audio signal. In this form, only the variance of the relative error has to be estimated, which fluctuates less. This form of the correction term results in a much better estimate of the spectral noise power density than the prior art, without requiring additional memory.

[0020] In the above-described methods, the relative spectral noise power density estimation error may be determined if no wanted signal component is detected in the audio signal. This is particularly simple. The step of detecting the wanted signal component may be performed with a voice activity detector, for example.

[0021] In the above-described methods, the first estimate of the spectral noise power density may be a mean noise power density. The mean noise power density may be for example a moving average. Computing means is comparatively simple and does not require much computing power.

[0022] In the above-described methods, the first estimate of the spectral noise power density may, in principle, be determined by any prior art method. In particular, it may be determined based on a minimum statistics method or a minimum tracking method. These methods are easy to implement.

[0023] Furthermore, the invention provides a method for reducing noise in an audio signal, comprising:

providing an estimate of the spectral noise power density according to the previously described methods,

filtering the audio signal based on the second estimate of the spectral noise power density.

[0024] This method advantageously reduces noise in an audio signal without suffering from the so called musical noise artifacts and without using additional memory.

[0025] In the previously described method, the step of filtering may be performed using a Wiener filter or a minimal subtraction filter having a filter characteristic based on the second estimate of the spectral noise power density of the audio signal. The resulting signal is an enhanced signal with reduced noise. Compared to prior art filters, the output of such a filter fluctuates less, if no wanted signal component is present, i.e. during speech pauses. In particular, the filter characteristics of the Wiener filter based on the second estimate of the spectral noise power density has the form:

$$H_{mod}(e^{j\Omega_\mu}, n) = 1 - \frac{\tilde{S}_{bb}(\Omega_\mu, n)}{S_{yy}(\Omega_\mu, n)} - \frac{\sigma_{E_{rel}}^2 \cdot \tilde{S}_{bb}^2(\Omega_\mu, n)}{S_{yy}^2(\Omega_\mu, n) - \tilde{S}_{bb}(\Omega_\mu, n) \cdot S_{yy}(\Omega_\mu, n)}$$

This Wiener filter characteristics may further be generalized by introducing frequency- and time-dependent weight factors, such that the characteristics looks like:

$$H_{mod}(e^{j\Omega_\mu}, n) = 1 - \alpha(\Omega_\mu, n) \frac{\tilde{S}_{bb}(\Omega_\mu, n)}{S_{yy}(\Omega_\mu, n)} - \beta(\Omega_\mu, n) \frac{\sigma_{E_{rel}}^2 \cdot \tilde{S}_{bb}^2(\Omega_\mu, n)}{S_{yy}^2(\Omega_\mu, n) - \tilde{S}_{bb}(\Omega_\mu, n) \cdot S_{yy}(\Omega_\mu, n)}$$

In the above filter characteristics, the coefficients α and β may depend on frequency or time, respectively, alone.

[0026] The steps of the above-described method may be preceded or followed by further filtering steps. For example, the audio signal may be the result of processing steps, performed by processing units such as, for example, a beamformer, one or more band-pass filters or an echo-cancellation component. The output of above-described method may further

be processed by processing units, such as, for example filters or a gain control component.

[0027] Furthermore, as set forth in independent claim 11, the invention provides a computer program product comprising one or more computer readable media having computer-executable instructions for performing the steps of the previously described methods when run on a computer.

5 [0028] In addition, the invention provides an apparatus for providing an estimate of a spectral noise power density of an audio signal as set forth in independent claim 12. Preferred embodiments of said apparatus are set forth in dependent claims 13-17.

[0029] The invention further provides a system for reducing noise in an audio signal, as set forth in independent claim 18. A preferred embodiment of said system is set forth in dependent claim 19.

10 [0030] Additional aspects of the invention will be described in the following with reference to the figures and illustrative examples.

Figure 1 illustrates schematically an example of a system for filtering noise in an audio signal;

15 Figure 2 illustrates schematically an example of the signal flow of a method for providing an estimate of the spectral power density of an audio signal according to the invention;

Figure 3 illustrates an example for the source of the "musical noise" artifact;

20 Figure 4 illustrates the effectiveness of the correction mechanism according to the present invention;

Figure 5 illustrates the effectiveness of the method according to the present invention in a noise reduction example.

25 [0031] An example of the structure and the corresponding signal flow in a noise reduction filter is illustrated in Figure 1. Such a noise reduction filter may be used in hands-free telephony applications, for example in a vehicle. The audio signal may be received by one or more microphones. In the case of a vehicle, the noise component may be composed of the noise of the engine, the windstream, and the rolling tires. In addition, unwanted signal components may be due to sound from loudspeakers, reproducing the output either of a radio or of a hands-free telephony application, which may result in echoes.

30 [0032] The disturbed audio signal $y(n)$ comprises the wanted signal component $x(n)$ such as the speech signal and a noise component $b(n)$, e.g. engine noise, echoes, etc.

[0033] Thus, the signal entering the short-term frequency analysis block 110 is the sum

35

$$y(n) = x(n) + b(n).$$

40 [0034] In the frequency analysis component 110 the signal is split into overlapping blocks of appropriate size. The block length may be for example 32 msec. Each block is transformed via a filter bank or a discrete frequency transformation (DFT) into the frequency domain. The frequency domain signal is then input into a spectral weighting component 120.

45 [0035] In order to remove the background noise components, each sub-band or frequency bin is weighted with an attenuation factor, which depends on the current signal to noise ratio. A possible filter for removing the noise is the Wiener filter (see for example, E. Hänsler, G. Schmidt: Audio Echo and Noise Control: A Practical Approach, Wiley IEEE Press, New York, NY (USA), 2004; E. Hänsler: Statistische Signale, Springer Verlag, Berlin (Germany), 2001; P. Vary, U. Heute, W. Hess: Digitale Sprachsignalverarbeitung, Teubner, Stuttgart, 1998). whose filter characteristic, in principle, looks like

50

$$H(e^{j\Omega_\mu}, n) = 1 - \frac{S_{bb}(\Omega_\mu, n)}{S_{yy}(\Omega_\mu, n)}.$$

55 [0036] Here, $S_{bb}(\Omega_\mu, n)$ denotes the spectral power density of the noise component $b(n)$, $S_{yy}(\Omega_\mu, n)$ the spectral power density of the distorted signal $y(n)=x(n)+b(n)$ and Ω_μ the frequency with frequency-index μ . The weighting factor computed according to the Wiener characteristics approaches 1, if the spectral power density of the distorted signal $y(n)$ is greater than the spectral power density of the background noise. In the absence of a wanted signal component $x(n)$, the spectral

noise power density equals the spectral power density of the distorted signal. In this case, $H(e^{j\Omega_\mu}, n) = 0$ and the filter is closed.

[0037] In almost all noise reduction methods, the problem is to estimate the portion of $S_{yy}(\Omega_\mu, n)$, which is due to noise. To this end, a slowly varying estimate $\tilde{S}_{bb}(\Omega_\mu, n)$ is generated which is the mean spectral power density of the noise component. Slowly varying in this context means, that the estimate $\tilde{S}_{bb}(\Omega_\mu, n)$ shows less fluctuations with respect to time than the spectral power density of the distorted signal $S_{yy}(\Omega_\mu, n)$. There are several methods known in the state of the art to estimate the spectral noise power density, such as the "minimum statistics" or "minimum tracking" methods (see, for example, E. Hänsler, G. Schmidt: Audio Echo and Noise Control: A Practical Approach, Wiley IEEE Press, New York, NY (USA), 2004).

[0038] The spectral power density of the distorted signal has to be estimated by a faster varying signal to account for the varying power of the speech signal. According to the prior art, this is achieved by slightly smoothing the squared moduli. The filter characteristics of the Wiener filter then takes the form

$$\tilde{H}(e^{j\Omega_\mu}, n) = 1 - \frac{\tilde{S}_{bb}(\Omega_\mu, n)}{S_{yy}(\Omega_\mu, n)}$$

[0039] Compared to the above characteristics, the spectral noise power density has been replaced by the estimated spectral noise power density.

[0040] According to the invention, the estimate of the spectral noise power density is replaced by an improved estimate, which resembles more closely the actual or current spectral noise power density. The method for providing this improved estimate will be outlined in greater detail below.

[0041] The output of the spectral weighting component 120, consisting of the weighted frequency components is then input into an optional post-processing unit 130. Further processing such as pitch adaptive filtering or automatic gain control can be applied in this post-processing unit 130.

[0042] Finally, the resulting frequency domain representation of the enhanced signal spectrum is transformed back into the time domain in the synthesis component 140. The output of this component is the enhanced signal.

[0043] Figure 1 depicts the general concept schematically and only contains the main steps of a noise reduction method. It may be that the output of any of the shown blocks is not directly input into the subsequent block, but that further processing is performed in between the blocks. For example, the signal $y(n)$ may be the result of processing steps, performed by processing units such as, for example, a beam-former, one or more band-pass filters or an echo-cancellation component. The enhanced signal output by the synthesis block 140 may further be processed by processing units, such as, for example, filters or a gain control component.

[0044] Often in the spectral weighting component 120, a Wiener filter is used. As already mentioned, the spectral noise power density $S_{bb}(\Omega_\mu, n)$ is estimated by a slowly varying estimate $\tilde{S}_{bb}(\Omega_\mu, n)$, whereas the estimate of the spectral power density of the disturbed signal $S_{yy}(\Omega_\mu, n)$ changes much faster. As a result, the sub-band attenuation factors are fluctuating randomly. Thus, the broadband background noise is transformed into a signal consisting of short-lasting tones if no wanted signal component is present, e.g. during speech pauses. This behavior is often called the "musical noise" or "musical tones" artifact.

[0045] The situation is depicted in Figure 3. The upper part of Figure 3 shows the slowly varying estimate $\hat{S}_{bb}(\Omega_\mu, n)$ and the spectral power density of the disturbed signal $S_{yy}(\Omega_\mu, n)$. In particular during speech pauses, $S_{yy}(\Omega_\mu, n)$ fluctuates much more than $\tilde{S}_{bb}(\Omega_\mu, n)$. As a result, the Wiener filter characteristic $\tilde{H}(e^{j\Omega_\mu}, n)$ fluctuates during speech pauses as shown in the lower part of the Figure. This statistic opening and closing of the filter produces the musical noise artifact.

[0046] A known method in the prior art to tackle this problem is to modify $\tilde{S}_{bb}(\Omega_\mu, n)$ with an overweighting factor $\beta(\Omega_\mu)$:

$$\bar{H}(e^{j\Omega_\mu}, n) = 1 - \beta(\Omega_\mu) \cdot \frac{\tilde{S}_{bb}(\Omega_\mu, n)}{S_{yy}(\Omega_\mu, n)}$$

[0047] If $\beta(\Omega_\mu)$ is chosen appropriately, the unwanted artifacts are reduced by this method. However, the filter does not open much during speech activity. Another approach is the adaptive adjustment of the overweighting factor. This method requires additional memory and is known as the recursive Wiener filter.

[0048] According to the present invention, the slowly varying estimate $\tilde{S}_{bb}(\Omega_\mu, n)$ is corrected to closer resemble the actual or current spectral noise power density, such that an underestimation in the absence of the wanted signal com-

ponent is avoided and in the presence of the wanted signal component, $\tilde{S}_{bb}(\Omega_\mu)$ is used without correction. Therefore, no global overestimation has to be used. Furthermore, no additional memory is required.

[0049] In the illustrated example according to Figure 2, the audio signal $y(n)$ enters the short-term frequency analysis block 210, which provides the spectral power density of the signal. A frequently used technique for providing the spectral power density of a signal is the fast Fourier transform (FFT). The FFT may be applied to overlapping signal segments. The segmentation can be described by extracting the last M samples of the input signal $y(n)$. Successive blocks may be overlapping by 50% or 75%. In addition, each segment may be multiplied by a windowing function. In the case of a short-time frequency analysis, the frequency-domain signal is composed of frequency bands characterized by frequency supporting points Ω_μ . The frequency supporting points Ω_μ may be chosen equidistantly over the normalized frequency range:

$$\Omega_\mu = \frac{2\pi}{M} \mu \text{ with } \mu \in \{0, \dots, M-1\}.$$

with $\mu \in \{0, \dots, M-1\}$.

[0050] The number M of frequency supporting points may be 256 for example. The frequency supporting points may, however, be chosen non-uniformly as well.

[0051] The audio signal $y(n)$ also enters the spectral noise power density estimation unit 220, which provides a first estimate of the spectral noise power density of the audio signal $\tilde{S}_{bb}(\Omega_\mu, n)$. The output of block 220 is a slowly varying estimate for the spectral noise power density, which represents the mean power of the background noise. To provide a first estimate of the spectral noise power density methods such as minimum statistics or minimum tracking may be used.

[0052] In the error variance estimation unit 230, the variance of the error $\sigma_{E_n}^2$ is estimated. This estimation may be performed when no wanted signal component is present, i.e., during speech pauses.

[0053] The output of block 220 is input to block 240, which estimates, based on the first estimate of the spectral noise power density $\tilde{S}_{bb}(\Omega_\mu, n)$, the variance of the relative error $\sigma_{E_{rel}}^2$ by computing $\sigma_{E_{rel}}^2 = \sigma_{E_n}^2 / \tilde{S}_{bb}(\Omega_\mu, n)$. $\sigma_{E_{rel}}^2$

may be estimated, when no wanted signal component is present, i.e. during speech pauses.

[0054] At block 250, the correction term is computed based on the variance of the relative spectral noise power density estimation error $\sigma_{E_{rel}}^2$ on the first estimate of the spectral noise power density of the audio signal $\tilde{S}_{bb}(\Omega_\mu, n)$, and on the current spectral signal power density of the audio signal $S_{yy}(\Omega_\mu, n)$. The correction term is computed according to the following formula:

$$K(\Omega_\mu, n) = \frac{\sigma_{E_{rel}}^2 \cdot \tilde{S}_{bb}^2(\Omega_\mu, n)}{(S_{yy}(\Omega_\mu, n) - \tilde{S}_{bb}(\Omega_\mu, n))^2}$$

[0055] An example of the resulting correction factor is shown in Figure 4. The middle part of Figure 4 shows the correction factor $K(\Omega_\mu, n)$. A correction takes place primarily in the absence of a wanted signal component, i.e. during speech pauses.

[0056] Finally, to obtain an estimate of the spectral noise power density of the audio signal such that the spectral noise power density estimation error is reduced, the correction term $K(\Omega_\mu, n)$ and the first estimate of the spectral noise power density are added at block 260. The new estimate of the spectral noise power density of the audio signal then takes the following form:

$$\hat{S}_{bb}(\Omega_\mu, n) = \tilde{S}_{bb}(\Omega_\mu, n) + \frac{\sigma_{E_{rel}}^2 \cdot \tilde{S}_{bb}^2(\Omega_\mu, n)}{S_{yy}(\Omega_\mu, n) - \tilde{S}_{bb}(\Omega_\mu, n)}.$$

[0057] This spectral noise power density estimate may be used instead of the first spectral noise power density estimate $\tilde{S}_{bb}(\Omega_\mu, n)$ in numerous methods and filter characteristics, respectively. The most important methods are power and amplitude SPS, Wiener filter and the methods according to Ephraim and Malah (see, for example, Y. Ephraim, D.

Malah: Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator, IEEE Transactions On Audios, Speech, And Signal Processing, Vol. ASSP-32, No. 6, 1984)

[0058] The upper part of Figure 4 shows $S_{yy}(\Omega_\mu, n)$, $\tilde{S}_{bb}(\Omega_\mu, n)$ and $\tilde{S}_{bb}(\Omega_\mu, n)$. As can be seen, in particular during speech pauses, $\tilde{S}_{bb}(\Omega_\mu, n)$ more closely follows $S_{yy}(\Omega_\mu, n)$, which consist of a noise component in the absence of a wanted signal component, than $\tilde{S}_{bb}(\Omega_\mu, n)$ does.

[0059] To demonstrate the effectiveness of the above-described method, an example is shown in the following, in which the enhanced estimate of the spectral noise power density is used in a Wiener filter to reduce noise in a speech signal.

[0060] The modified filter characteristics of the Wiener filter, based on the second estimate of the spectral noise power density according to the invention takes the form:

$$H_{\text{mod}}(e^{j\Omega_\mu}, n) = 1 - \frac{\tilde{S}_{bb}(\Omega_\mu, n)}{S_{yy}(\Omega_\mu, n)} - \frac{\sigma_{\text{Erel}}^2 \cdot \tilde{S}_{bb}^2(\Omega_\mu, n)}{S_{yy}^2(\Omega_\mu, n) - \tilde{S}_{bb}(\Omega_\mu, n) \cdot S_{yy}(\Omega_\mu, n)}$$

[0061] The last part of the sum is due to the correction term according to the invention.

[0062] The lower part of Figure 4 shows the modified Wiener filter characteristics $H_{\text{mod}}(\Omega_\mu, n)$. As can be seen, the filter is closed in the absence of a wanted signal component, i.e. during speech pauses.

[0063] The effectiveness of this modified Wiener filter is shown in Figure 5, which contains three spectrographs. The first one shows the time-frequency analysis of a distorted speech signal. The second spectrograph shows the noise-reduced speech signal without the application of a correction mechanism, i.e. a plain Wiener filter with characteristic $\tilde{H}(e^{j\Omega_\mu}, n)$. During speech pauses a residual noise component (musical noise) is still present. The third spectrograph shows the filtered speech signal processed by a modified Wiener filter according to the present invention. The musical noise during speech pauses is much reduced compared to the unmodified Wiener filter. In this example, the filter characteristic according to the above equation, i.e. $H_{\text{mod}}(e^{j\Omega_\mu}, n)$ has been used.

[0064] It is to be understood that the different parts and components of the methods and apparatuses described above can also be implemented independent of each other and be combined in different form. Furthermore, the above-described embodiments are to be construed as exemplary embodiments only.

Claims

1. Method for providing an estimate of the spectral noise power density of an audio signal, comprising:

providing a first estimate of the spectral noise power density of the audio signal,
determining a time dependent correction term.

summing the first estimate and the correction term to obtain a second estimate of the spectral noise power density of the audio signal,

wherein the correction term is determined such that a spectral noise power density estimation error is reduced,
and

wherein the audio signal comprises a wanted signal component and a noise component and the correction term is based on the expectation value of the squared difference of the current spectral noise power density and the first estimate of the spectral noise power density of the audio signal and on the expectation value of the squared spectral power density of the wanted signal component.

2. A method according to claim 1, wherein the correction term comprises a spectral power density estimation error.

3. Method according to claim 2, wherein the correction term comprises a product of a correction factor and the spectral power density estimation error.

4. Method according to one of the preceding claims, wherein the spectral noise power density estimation error is based on the deviation of the second estimate of the spectral noise power density of the audio signal from the current spectral noise power density of the audio signal.

5. Method according to one of the preceding claims, wherein the correction term is based on the variance of a relative

spectral noise power density estimation error, the first estimate of the spectral noise power density of the audio signal, and the current spectral signal power density of the audio signal.

- 5
6. Method according to claim 5, wherein the audio signal comprises a wanted signal component and a noise component and the relative spectral noise power density estimation error is determined if no wanted signal component is detected in the audio signal.
- 10
7. Method according to one of the preceding claims, wherein the first estimate of the spectral noise power density is a mean noise power density.
- 15
8. Method according to one of the preceding claims, wherein the first estimate of the spectral noise power density is determined based on a minimum statistics method or a minimum tracking method.
- 20
9. Method for reducing noise in an audio signal, comprising providing an estimate of the spectral noise power density according to the method of one of the claims 1 - 8 for the audio signal, filtering the audio signal based on the second estimate of the spectral noise power density.
- 25
10. Method according to claim 9, wherein the step of filtering is performed using a Wiener filter or a minimal subtraction filter having a filter characteristic based on the second estimate of the spectral noise power density of the audio signal.
- 30
11. Computer program product comprising one or more computer readable media having computer-executable instructions for performing the steps of the method of one of the preceding claims when run on a computer.
- 35
12. Apparatus for providing an estimate of the spectral noise power density of an audio signal, comprising:
- estimating means for providing a first estimate of the spectral noise power density of the audio signal, determining means for determining a time dependent correction term, summing means for summing the first estimate and the correction term to obtain a second estimate of the spectral noise power density of the audio signal, wherein the determining means is configured to determine the correction term such that a spectral noise power density estimation error is reduced, and wherein the audio signal comprises a wanted signal component and a noise component and the correction term is based on the expectation value of the squared difference of the current spectral noise power density and the first estimate of the spectral noise power density of the audio signal and on the expectation value of the squared spectral power density of the wanted signal component.
- 40
13. Apparatus according to claim 12, wherein the means for determining the correction term is configured to determine the correction term based on the variance of a relative spectral noise power density estimation error, on the first estimate of the spectral noise power density of the audio signal, and on the current spectral signal power density of the audio signal.
- 45
14. Apparatus according to claim 13, wherein the means for determining the time dependent correction term is configured to determine the relative spectral noise power density estimation error if no wanted signal component is detected in the audio signal .
- 50
15. Apparatus according to one of the claims 12 - 14, wherein the means for determining the correction term is configured to determine the relative spectral noise power density estimation error if no wanted signal component is detected in the audio signal.
- 55
16. Apparatus according to claim 15, further comprising a voice activity detector configured to detect whether a wanted signal component is present in the audio signal.
17. Apparatus according to one of the preceding claims, wherein the means for providing a first estimate of the spectral noise power density of the audio signal is configured to determine the first estimate of the spectral noise power density of the audio signal based on a minimum statistics method or minimum tracking method.
18. System for reducing noise in an audio signal, comprising:

an apparatus for providing an estimate of the spectral noise power density of an audio signal according to one of the claims 12 - 17,
filtering means for filtering the audio signal based on the second estimate of the spectral noise power density.

- 5 19. System according to claim 18, wherein the filtering means comprises a Wiener filter or a minimal subtraction filter having a filter characteristic based on the second estimate of the spectral noise power density of the audio signal.

Patentansprüche

- 10 1. Verfahren zum Bereitstellen eines Schätzwertes der spektralen Rauschleistungsdichte eines Audiosignals, das umfasst:

15 Bereitstellen eines ersten Schätzwertes der spektralen Rauschleistungsdichte des Audiosignals,
Bestimmen eines zeitabhängigen Korrekturterms,
Addieren des ersten Schätzwertes und des Korrekturterms, um einen zweiten Schätzwert der spektralen Rauschleistungsdichte des Audiosignals zu ermitteln,

20 wobei der Korrekturterm so bestimmt wird, dass ein Schätzfehler der spektralen Rauschleistungsdichte reduziert wird, und
wobei das Audiosignal eine erwünschte Signalkomponente und eine Rauschkomponente umfasst und der Korrekturterm auf dem Erwartungswert der quadrierten Differenz der aktuellen spektralen Rauschleistungsdichte sowie dem ersten Schätzwert der spektralen Rauschleistungsdichte des Audiosignals und auf dem Erwartungswert der quadrierten spektralen Leistungsdichte der erwünschten Signalkomponente basiert.

- 25 2. Verfahren nach Anspruch 1, wobei der Korrekturterm einen Schätzfehler der spektralen Leistungsdichte umfasst.

- 30 3. Verfahren nach Anspruch 2, wobei der Korrekturterm ein Produkt aus einem Korrekturfaktor und dem Schätzfehler der spektralen Leistungsdichte umfasst.

4. Verfahren nach einem der vorangehenden Ansprüche, wobei der Schätzfehler der spektralen Rauschleistungsdichte auf der Abweichung des zweiten Schätzwertes der spektralen Rauschleistungsdichte des Audiosignals von der aktuellen spektralen Rauschleistungsdichte des Audiosignals basiert.

- 35 5. Verfahren nach einem der vorangehenden Ansprüche, wobei der Korrekturterm auf der Varianz des relativen Schätzfehlers der spektralen Rauschleistungsdichte, dem ersten Schätzwert der spektralen Rauschleistungsdichte des Audiosignals und der aktuellen spektralen Signalleistungsdichte des Audiosignals basiert.

- 40 6. Verfahren nach Anspruch 5, wobei das Audiosignal eine erwünschte Signalkomponente und eine Rauschkomponente umfasst und der relative Schätzfehler der spektralen Rauschleistungsdichte bestimmt wird, wenn keine erwünschte Signalkomponente in dem Audiosignal erfasst wird.

- 45 7. Verfahren nach einem der vorangehenden Ansprüche, wobei der erste Schätzwert der spektralen Rauschleistungsdichte eine mittlere Rauschleistungsdichte ist.

8. Verfahren nach einem der vorangehenden Ansprüche, wobei der erste Schätzwert der spektralen Rauschleistungsdichte auf Basis eines Minimum-Statistikverfahrens oder eines Minimum-Folgeverfahrens bestimmt wird.

- 50 9. Verfahren zum Reduzieren von Rauschen in einem Audiosignal, das Bereitstellen eines Schätzwertes der spektralen Rauschleistungsdichte gemäß dem Verfahren nach einem der Ansprüche 1-8 für das Audiosignal, und Filtern des Audiosignals auf Basis des zweiten Schätzwertes der spektralen Rauschleistungsdichte umfasst.

- 55 10. Verfahren nach Anspruch 9, wobei der Schritt des Filterns unter Verwendung eines Wiener-Filters oder eines Minimal-Subtraktionsfilters mit einer Filtercharakteristik durchgeführt wird, die auf dem zweiten Schätzwert der spektralen Rauschleistungsdichte des Audiosignals basiert.

11. Computerprogrammerzeugnis, das ein oder mehrere durch Computer lesbare/s Medium/Medien umfasst, das/die durch Computer ausführbare Befehle zum Durchführen der Schritte des Verfahrens nach einem der vorangehenden

Ansprüche bei Ausführung auf einem Computer aufweist/aufweisen.

- 5
12. Vorrichtung zum Bereitstellen eines Schätzwertes der spektralen Rauschleistungsdichte eines Audiosignals, die umfasst:

10 eine Schätzeinrichtung zum Bereitstellen eines ersten Schätzwertes der spektralen Rauschleistungsdichte des Audiosignals,
eine Bestimmungseinrichtung zum Bestimmen eines zeitabhängigen Korrekturterms, eine Addiereinrichtung zum Addieren des ersten Schätzwertes und des Korrekturterms, um einen zweiten Schätzwert der spektralen Rauschleistungsdichte des Audiosignals zu ermitteln,

15 wobei die Bestimmungseinrichtung so konfiguriert ist, dass sie den Korrekturterm so bestimmt, dass der Schätzfehler der spektralen Rauschleistungsdichte reduziert wird, und
wobei das Audiosignal eine erwünschte Signalkomponente und eine Rauschkomponente umfasst und der Korrekturterm auf dem Erwartungswert der quadrierten Differenz der aktuellen spektralen Rauschleistungsdichte sowie dem ersten Schätzwert der spektralen Rauschleistungsdichte des Audiosignals und auf dem Erwartungswert der quadrierten spektralen Leistungsdichte der erwünschten Signalkomponente basiert.

- 20 13. Vorrichtung nach Anspruch 12, wobei die Einrichtung zum Bestimmen des Korrekturterms so konfiguriert ist, dass sie den Korrekturterm auf Basis der Varianz eines relativen Schätzfehlers der spektralen Rauschleistungsdichte, des ersten Schätzwertes der spektralen Rauschleistungsdichte des Audiosignals und der aktuellen spektralen Signalleistungsdichte des Audiosignals bestimmt.

- 25 14. Vorrichtung nach Anspruch 13, wobei die Einrichtung zum Bestimmen des zeitabhängigen Korrekturterms so konfiguriert ist, dass sie den relativen Schätzfehler der spektralen Rauschleistungsdichte bestimmt, wenn keine erwünschte Signalkomponente in dem Audiosignal erfasst wird.

- 30 15. Vorrichtung nach einem der Ansprüche 12-14, wobei die Einrichtung zum Bestimmen des Korrekturterms so konfiguriert ist, dass sie den relativen Schätzfehler der spektralen Rauschleistungsdichte bestimmt, wenn keine erwünschte Signalkomponente in dem Audiosignal erfasst wird.

16. Vorrichtung nach Anspruch 15, die des Weiteren einen Voice-Activity-Detektor umfasst, der so konfiguriert ist, dass er erfasst, ob eine erwünschte Signalkomponente in dem Audiosignal vorhanden ist.

- 35 17. Vorrichtung nach einem der vorangehenden Ansprüche, wobei die Einrichtung zum Bereitstellen eines ersten Schätzwertes der spektralen Rauschleistungsdichte des Audiosignals so konfiguriert ist, dass sie den ersten Schätzwert der spektralen Rauschleistungsdichte des Audiosignals auf Basis eines Minimum-Statistikverfahrens oder eines Minimum-Folgeverfahrens bestimmt.

- 40 18. System zum Reduzieren von Rauschen in einem Audiosignal, das umfasst:

45 eine Vorrichtung zum Bereitstellen eines Schätzwertes der spektralen Rauschleistungsdichte eines Audiosignals nach einem der Ansprüche 12-17,
eine Filtereinrichtung zum Filtern des Audiosignals auf Basis des zweiten Schätzwertes der spektralen Rauschleistungsdichte.

- 50 19. System nach Anspruch 18, wobei die Filtereinrichtung ein Wiener-Filter oder ein Minimal-Subtraktionsfilter mit einer Filtercharakteristik umfasst, die auf dem zweiten Schätzwert der spektralen Rauschleistungsdichte des Audiosignals basiert.

Revendications

- 55 1. Procédé pour la mise à disposition d'une estimation de la densité de puissance de bruit spectrale d'un signal audio, comprenant :

la mise à disposition d'une première estimation de la densité de puissance de bruit spectrale du signal audio, la détermination d'un terme de correction dépendant du temps,

la somme de la première estimation et du terme de correction pour l'obtention d'une seconde estimation de la densité de puissance de bruit spectrale du signal audio, dans lequel le terme de correction est déterminé de manière telle qu'une erreur d'estimation de densité de puissance de bruit spectrale est réduite, et

où le signal audio comprend un composant de signal voulu et un composant de bruit et le terme de correction est basé sur l'espérance mathématique de la différence élevée au carré de la densité de puissance de bruit spectrale actuelle et la première estimation de la densité de puissance de bruit spectrale du signal audio et sur l'espérance mathématique de la densité de puissance spectrale élevée au carré du composant de signal voulu.

2. Une méthode selon la revendication 1, dans laquelle le terme de correction comprend une erreur d'estimation de densité de puissance spectrale.

3. Procédé selon la revendication 2, dans lequel le terme de correction comprend un produit d'un facteur de correction et de l'erreur d'estimation de densité de puissance spectrale.

4. Procédé selon l'une des revendications précédentes, dans lequel l'erreur d'estimation de densité de puissance de bruit spectrale est basée sur la déviation de la seconde estimation de densité de puissance de bruit spectrale du signal audio de la densité de puissance de bruit spectrale courante du signal audio.

5. Procédé selon l'une des revendications précédentes, dans lequel le terme de correction est basé sur la variance d'une erreur d'estimation de densité de puissance de bruit spectrale relative, la première estimation de la densité de puissance de bruit spectrale du signal audio, et la densité de puissance de signal spectrale courante du signal audio.

6. Procédé selon la revendication 5, dans lequel le signal audio comprend un composant de signal voulu et un composant de bruit et l'erreur d'estimation de densité de puissance de bruit spectrale relative est déterminée si aucun composant de signal voulu n'est détecté dans le signal audio.

7. Procédé selon l'une des revendications précédentes, dans lequel la première estimation de la densité de puissance de bruit spectrale est une densité de puissance de bruit moyenne.

8. Procédé selon l'une des revendications précédentes, dans lequel la première estimation de densité de puissance de bruit spectrale est déterminée sur la base d'un procédé de statistiques minimum ou sur un procédé de traçage minimum.

9. Procédé pour la réduction du bruit dans un signal audio, comprenant la mise à disposition d'une estimation de la densité de bruit spectrale selon le procédé de l'une des revendications 1 à 8 pour le signal audio, le filtrage du signal audio basé sur la seconde estimation de la densité de puissance de bruit spectrale.

10. Procédé selon la revendication 9, dans lequel l'étape de filtrage est exécutée par l'utilisation d'un filtre de Wiener ou d'un filtre de soustraction minimale ayant une caractéristique de filtre basée sur la seconde estimation de la densité de puissance de bruit spectrale du signal audio.

11. Produit de programme informatique comprenant un ou plusieurs médias lisibles par l'ordinateur ayant des instructions exécutables par l'ordinateur pour l'exécution des étapes du procédé de l'une des revendications précédentes lors de la mise en marche sur un ordinateur.

12. Appareil pour la mise à disposition d'une estimation de la densité de puissance de bruit spectrale d'un signal audio, comprenant :

un moyen d'estimation pour la mise à disposition d'une première estimation de la densité de puissance de bruit spectrale du signal audio,

un moyen de détermination pour la détermination d'un terme de correction dépendant du temps,

un moyen d'addition pour l'addition de la première estimation et du terme de correction pour l'obtention d'une seconde estimation de la densité de puissance de bruit spectrale du signal audio,

dans lequel le moyen de détermination est configuré pour la détermination du terme de correction de manière à ce qu'une erreur d'estimation de densité de puissance de bruit spectrale soit réduite, et

où le signal audio comprend un composant de signal voulu et un composant de bruit et le terme de correction est basé sur l'espérance mathématique de la différence élevée au carré de la densité de puissance de bruit spectrale courante et la première estimation de la densité de puissance de bruit spectrale du signal audio et sur l'espérance mathématique de la densité de puissance spectrale élevée au carré du composant de signal voulu.

- 5
13. Appareil selon la revendication 12, dans lequel le moyen de détermination du terme de correction est configuré pour déterminer le terme de correction basé sur la variance d'une erreur d'estimation de densité de puissance de bruit spectrale relative, sur la première estimation de la densité de puissance de bruit spectrale du signal audio, et sur la densité de puissance de signal spectrale courante du signal audio.
- 10
14. Appareil selon la revendication 13, dans lequel le moyen de détermination du terme de correction dépendant du temps est configuré pour la détermination de l'erreur d'estimation de densité de puissance de bruit spectrale relative si aucun composant de signal voulu n'est détecté dans le signal audio.
- 15
15. Appareil selon l'une des revendications 12 à 14, dans lequel le moyen de détermination du terme de correction est configuré pour la détermination de l'erreur d'estimation de densité de puissance de bruit spectrale relative si aucun composant de signal voulu n'est détecté dans le signal audio.
- 20
16. Appareil selon la revendication 15, comprenant en outre un détecteur d'activité vocale configuré pour la détection de la présence ou non d'un composant de signal voulu dans le signal audio.
- 25
17. Appareil selon l'une des revendications précédentes, dans lequel le moyen de mise à disposition d'une première estimation de la densité de puissance de bruit spectrale du signal audio est configurée pour la détermination de la première estimation de la densité de puissance de bruit spectrale du signal audio basé sur un procédé de statistiques minimum ou un procédé de traçage minimum.
- 30
18. Système pour la réduction du bruit dans un signal audio, comprenant :
- un appareil pour la mise à disposition d'une estimation de la densité de puissance de bruit spectrale d'un signal audio selon l'une des revendications 12 à 17,
- un moyen de filtrage pour le filtrage du signal audio basé sur la seconde estimation de la densité de puissance de bruit spectrale.
- 35
19. Système selon la revendication 18, dans lequel le moyen de filtrage comprend un filtre de Wiener ou un filtre de soustraction minimale ayant une caractéristique de filtre basée sur la seconde estimation de la densité de puissance de bruit spectrale du signal audio.
- 40
- 45
- 50
- 55

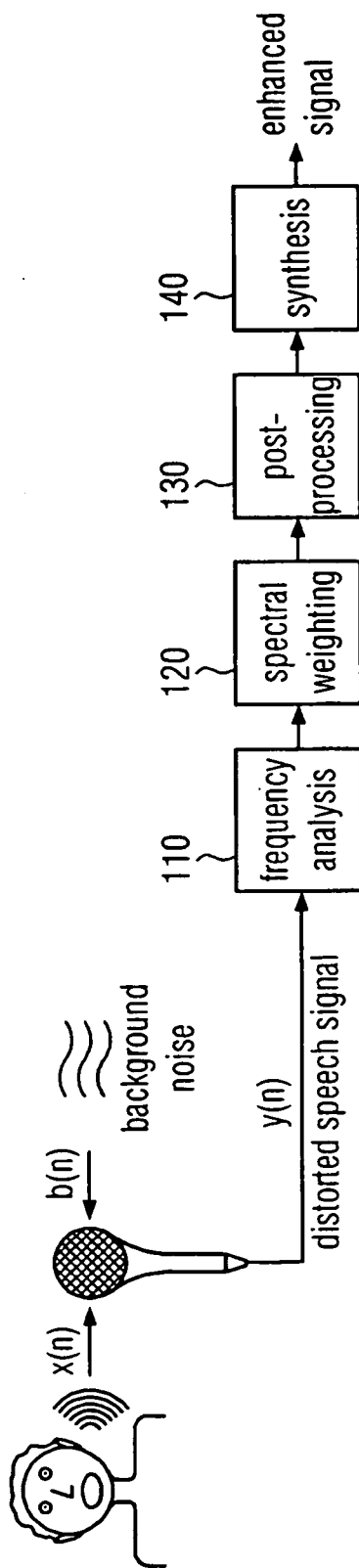


FIG. 1

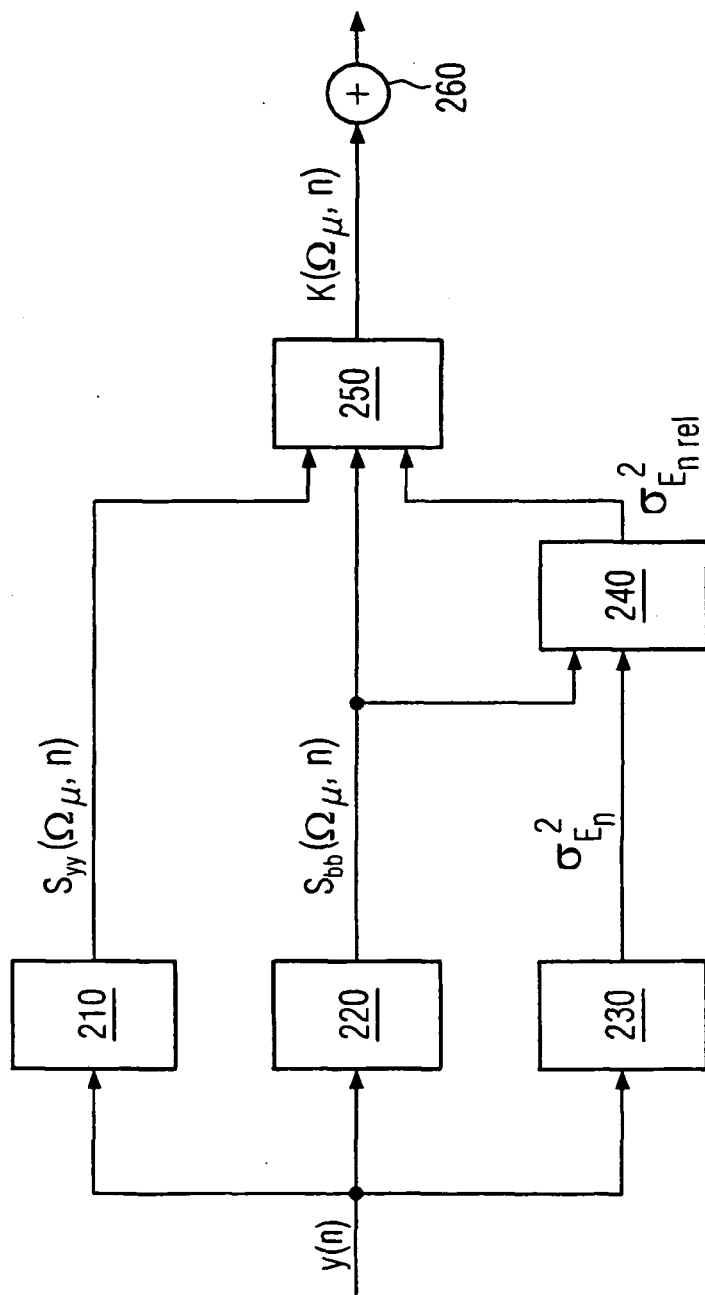


FIG. 2

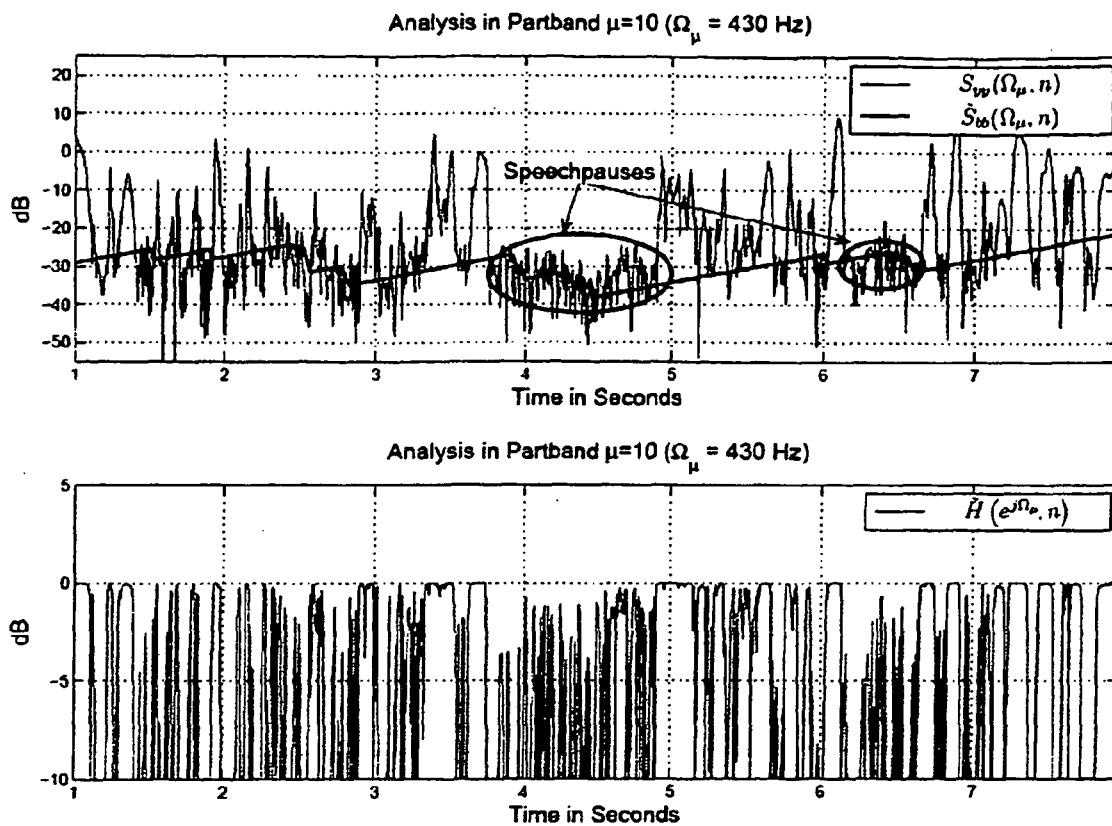


FIG. 3

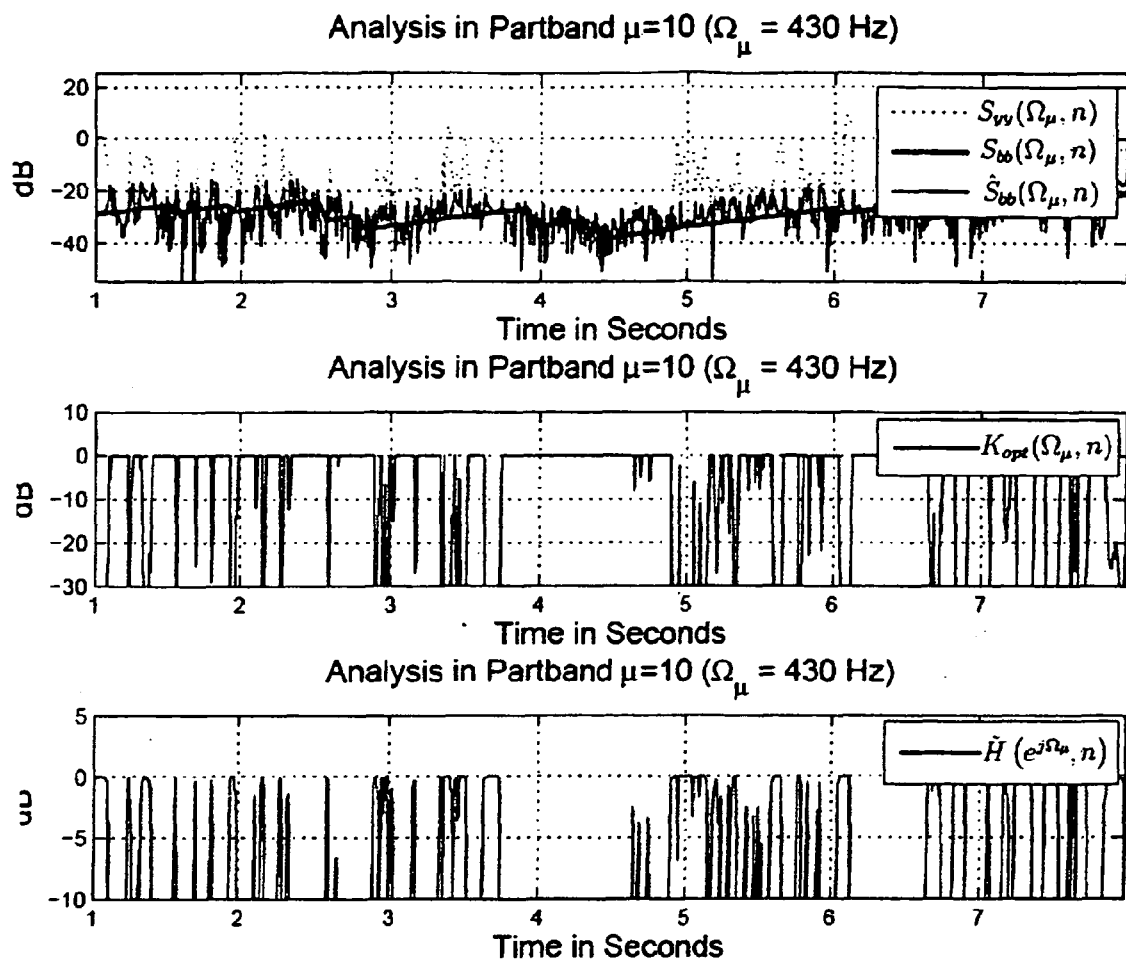


FIG. 4

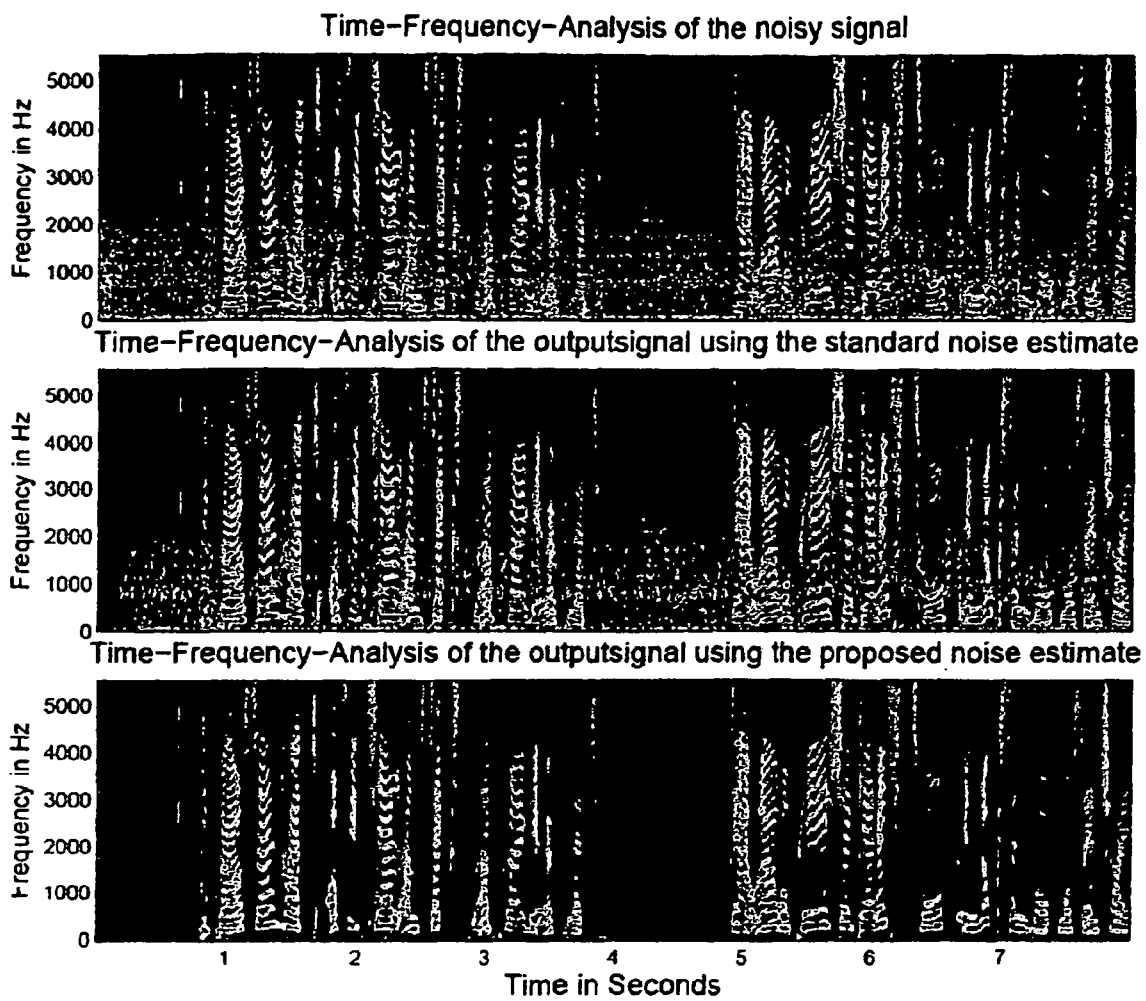


FIG. 5

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

- GB 2426167 A [0007]

Non-patent literature cited in the description

- **E. Hänslér ; G. Schmidt.** Audio Echo and Noise Control: A Practical Approach. Wiley IEEE Press, 2004 [0035] [0037]
- **E. Hänslér.** Statistische Signale. Springer Verlag, 2001 [0035]
- **P. Vary ; U. Heute ; W. Hess.** Digitale Sprachsignalverarbeitung. 1998 [0035]
- **Y. EPHRAIM ; D. MALAH.** Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. IEEE Transactions On Audios [0057]
- *Signal Processing*, 1984, vol. ASSP-32 (6 [0057]