



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
10.02.2010 Bulletin 2010/06

(51) Int Cl.:
G10L 21/02 (2006.01) G10L 19/00 (2006.01)

(21) Application number: **08014151.8**

(22) Date of filing: **07.08.2008**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT RO SE SI SK TR
 Designated Extension States:
AL BA MK RS

- **Krini, Mohamed**
89073 Ulm (DE)
- **Paranjpe, Shreyas**
Vancouver
British Columbia V6S 1L5 (CA)
- **Schmidt, Gerhard**
89077 Ulm (DE)

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

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

(72) Inventors:
 • **Haulick, Tim**
89143 Blaubeuren (DE)

(54) **Noise-reduction processing of speech signals**

(57) The present invention relates to a method for signal processing comprising the steps of providing a set of prototype spectral envelopes, providing a set of reference noise prototypes, wherein the reference noise prototypes are obtained from at least a subset of the provided set of prototype spectral envelopes, detecting a verbal

utterance by at least one microphone to obtain a microphone signal, processing the microphone signal for noise reduction based on the provided reference noise prototypes to obtain an enhanced signal and encoding the enhanced signal based on the provided prototype spectral envelopes to obtain an encoded enhanced signal.

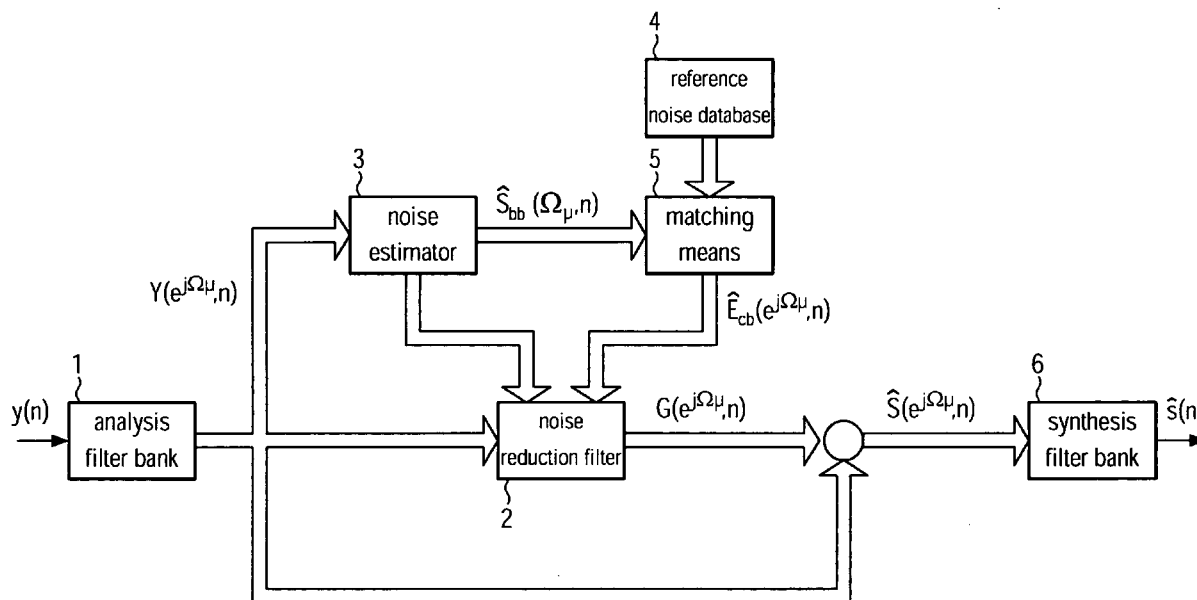


FIG. 1

Description**Field of Invention**

5 [0001] The present invention relates to the art of electronically mediated verbal communication, in particular, by means of hands-free sets that, for instance, are installed in vehicular cabins. The invention is particularly directed to the pre-processing of speech signals before speech codec processing.

Background of the invention

10 [0002] Two-way speech communication of two parties mutually transmitting and receiving audio signals, in particular, speech signals, often suffers from deterioration of the quality of the audio signals caused by background noise. Hands-free telephones provide comfortable and safe communication systems of particular use in motor vehicles. However, perturbations in noisy environments can severely affect the quality and intelligibility of voice conversation, e.g., by means of mobile phones or hands-free telephone sets that are installed in vehicle cabins, and can, in the worst case, lead to a complete breakdown of the communication.

15 [0003] Consequently, some noise reduction must be employed in order to improve the intelligibility of electronically mediated speech signals. In particular, in the case of hands-free telephones, it is mandatory to suppress noise in order to guarantee successful communication. In the art, noise reduction methods employing Wiener filters or spectral subtraction are well known. For instance, speech signals are divided into sub-bands by some sub-band filtering means and a noise reduction algorithm is applied to each of the frequency sub-bands.

20 [0004] However, the intelligibility of speech signals and quality of hands-free communication is still not improved sufficiently when perturbations, e.g., caused by driving and rolling noise of vehicles at high speeds, are relatively strong resulting in a relatively low signal-to-noise ratio. In particular, at transitions from verbal utterances (speech activity) to speech pauses after the encoding and decoding of speech employed in the transmission of speech from a near party to a remote party communication suffers from severe artifacts known as the gating effect. Thus, there is a need for an improved method and system for noise reduction in electronic speech communication, in particular, in the context of hands-free sets.

Description of the Invention

30 [0005] The above-mentioned problem is solved by the method for signal processing according to claim 1 comprising the steps of
 providing a set of prototype spectral envelopes;
 35 providing a set of reference noise prototypes, wherein the reference noise prototypes are obtained from at least a sub-set of the provided set of prototype spectral envelopes;
 detecting a verbal utterance by at least one microphone to obtain a microphone signal;
 processing the microphone signal for noise reduction based on the provided reference noise prototypes to obtain an enhanced signal; and

40 encoding the enhanced signal based on the provided prototype spectral envelopes to obtain an encoded enhanced signal.

[0006] Spectral envelopes are commonly used in the art of speech signal processing, speech synthesis, speech recognition etc. (see, e.g., Y. Griffin and J.S. Lim, "Multi-Band Excitation Vocoder", IEEE Transactions Acoustical Speech Signal Processing, Vol. 36, No. 8, pages 1223-1235, 1988).

45 [0007] In the art, speech signals to be transmitted from a near party to a remote party, e.g., by hands-free telephony, are enhanced by noise reduction that does not consider the subsequent codec (encoding and decoding) processing of the noise-reduced signals which is performed in telephony communication. Contrary, in the present invention codec processing is taken into account and it is aimed to provide speech signals that show a significantly enhanced quality after both signal processing for noise reduction and codec processing.

50 [0008] This object is achieved by providing reference noise prototypes and noise-reduction of the processed speech signals based on the provided reference noise prototypes. The prototypes are predetermined such that subsequent codec processing does not severely affect the quality of the speech signals decoded and output at the end of some remote party that received the noise-reduced and encoded speech signals. This is particularly achieved by providing reference noise prototypes that are obtained from, e.g., chosen from, at least a sub-set of the provided set of prototype spectral envelopes. Thereby, artifacts that affect the intelligibility of speech signals after processing for noise reduction and encoding/decoding can be suppressed.

55 [0009] The reference noise prototypes can, in particular, be spectral envelopes modeled by an all-pole filter function. For instance, the reference noise prototypes may be chosen from the prototype spectral envelopes of a speech codec.

[0010] The provided set of prototype spectral envelopes may particularly be used for the encoding of the enhanced

signal in speech pauses detected in the microphone signal or when a signal-to-noise ratio of the microphone signal falls below a predetermined threshold (see also detailed discussion below). In particular, the disturbing so-called gating effect can efficiently be suppressed by the herein disclosed method for signal processing.

5 [0011] The speech encoding of the enhanced signal (and corresponding decoding on a receiver side) can be performed by any method known in the art, e.g., Enhanced Variable Rate Codec (EVRC) and Enhanced Full Rate Codec (EFRC) (see also detailed discussion below).

10 [0012] The above-described method according to an embodiment comprises transmitting the encoded enhanced signal to a remote party, receiving the transmitted encoded enhanced signal by the remote party and decoding the received signal by the remote party. The quality of the speech signal after decoding by the remote party is significantly enhanced as compared to the art, since the noise reduction of the microphone signal at the near side takes into account the subsequent encoding/decoding by the provided reference noise prototypes.

15 [0013] According to a further embodiment, the processing of the microphone signal for noise reduction comprises estimating the power density of a noise contribution in the microphone signal; matching the spectrum of the noise contribution obtained from the estimated power density of the noise contribution with the provided set of reference noise prototypes to find the best matching reference noise prototype; and using the best matching reference noise prototype for noise reduction of the microphone signal.

20 [0014] The best matching reference noise prototype is particularly used to determine maximum damping factors for a noise reduction characteristics of the noise reduction filtering means employed for noise reduction of the microphone signal. By this procedure it is achieved that noise reduction is based on the best matching reference noise prototype, i.e., the subsequent encoding is taken very suitably into account in the noise reduction process.

25 [0015] In general, the best matching reference noise prototype will change with time. In order to avoid associated abrupt changes in the maximum damping factors that might lead to disturbing artifacts, switching from one best matching reference noise prototype to another for determining the maximum damping factors might be performed in a smoothed manner. An example for a smooth transition from one reference noise prototype used for the noise reduction processing to another is described in the detailed description below.

30 [0016] In particular, the processing of the microphone signal for noise reduction can be performed by a Wiener-like filtering means comprising damping factors obtained based on the best matching reference noise prototype, the power density spectrum of sub-band signals obtained from the microphone signal and the estimated power density spectrum of the background noise. Employment of some Wiener characteristics allows for reliable noise reduction and fast convergence of standard algorithms for the determination of the filter coefficients (damping factors). The details for the determination of the damping factors are described in the detailed description below.

35 [0017] Moreover, it might be preferred that the spectrum of the noise contribution obtained from the estimated power density of the noise contribution is matched only with a sub-set of the provided reference noise prototypes within a predetermined frequency range, e.g., ranging from 300 - 700 Hz. This is advantageous, since the actual noise may differ largely from the provided reference spectra in low frequencies. Restricting the search for the best matching reference noise prototype to some predetermined frequency significantly accelerates the processing.

40 [0018] Furthermore, it is provided a method for speech communication with a hands-free set installed in a vehicle, particular, an automobile, comprising the method according to one of the preceding claims, wherein at least one of the provided reference noise prototypes on which the processing of the microphone signal for noise reduction to obtain an enhanced signal is based is determined from a sub-set of the provided set of reference noise prototypes that is selected according to a current (presently measured) traveling speed of the vehicle, in particular, the automobile; and/or

the reference noise prototypes are obtained from a sub-set of the provided set of prototype spectral envelopes selected according to the type of the vehicle, in particular, the automobile.

45 [0019] According to this example, the computation load is reduced as compared to the previous examples. For example, only a reduced number of reference noise prototypes has to be considered in finding the one that best matches the background noise spectrum depending on the type of the vehicle, in particular, the automobile, e.g., depending on the brand of an automobile or characteristics of the engine, etc. Further, depending on the travelling speed particular prototype spectral envelopes might be typically used for the speech codec processing and these envelopes are advantageously used for the noise reduction. Thus, other reference noise prototypes can be ignored thereby reducing the demand for computational resources.

50 [0020] The present invention, moreover, can be incorporated in a computer program product comprising at least one computer readable medium having computer-executable instructions for performing one or more steps of the method according to one of the above-described embodiments when run on a computer.

55 [0021] The above-mentioned problem is also solved by a signal processing means according to claim 9, comprising an encoding database comprising prototype spectral envelopes; a reference database comprising reference noise prototypes, wherein the reference noise prototypes are obtained from at least a sub-set of the provided set of prototype spectral envelopes; and

a noise reduction filtering means configured to process a microphone signal comprising background noise based on the reference noise prototypes to obtain an enhanced microphone signal; and

an encoder configured to encode the enhanced microphone signal based on the prototype spectral envelopes.

[0022] In particular, the reference noise prototypes may be a sub-set of the provided set of prototype spectral envelopes.

[0023] According to an embodiment, the signal processing means further comprises

a noise estimating means configured to estimate the power density of a background noise contribution of the microphone signal;

a matching means configured to match the spectrum of the noise contribution obtained from the estimated power density of the noise contribution with the set of reference noise prototypes comprised in the reference database to find the best matching reference noise prototype; and

the noise reduction filtering means is configured to use the best matching reference noise prototype for noise reduction of the microphone signal.

[0024] The noise reduction filtering means may be a Wiener-like filtering means comprising damping factors based on the best matching reference noise prototype, the power density spectrum of microphone sub-band signals obtained from the microphone signal and the estimated power density spectrum of the background noise present in the microphone signal.

[0025] In particular, the noise reduction filtering means may be configured to operate in the sub-band regime and to output noise-reduced microphone sub-band signals and the signal processing means may further comprise an analysis filter bank configured to process the microphone signal to obtain microphone sub-band signals and to provide the microphone sub-band signals to the noise reduction filtering means; and a synthesis filter bank configured to process the noise-reduced microphone sub-band signals to obtain a noise-reduced full-band microphone signal in the time domain.

[0026] The signal processing means may be installed in an automobile and the reference database may be derived from the encoding database dependent on type of the automobile.

[0027] According to another embodiment one of the above-mentioned examples for the signal processing means according to the present invention further comprises a control means configured to control determination of at least one of the reference noise prototypes used by the noise reduction filtering means to process the microphone signal to obtain the enhanced microphone signal based on a current traveling speed of the automobile.

[0028] The signal processing means is particularly useful for a hands-free telephony set. Thus, it is provided a hands-free (telephony) set, in particular, installed in a vehicle, e.g. an automobile, comprising at least one microphone, in particular, a number of microphone arrays, at least one loudspeaker and a signal processing means according to one of the above examples of the inventive signal processing means. Moreover, herein it is provided an automobile with such a hands-free set installed in the compartment of the automobile.

[0029] Additional features and advantages of the present invention will be described with reference to the drawing. 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.

[0030] Figure 1 illustrates an example of the processing of a microphone signal that is to be transmitted from a near party to a remote party according to the present invention including noise-reduction by means of reference noise prototypes.

[0031] Figure 2 illustrates an example of processing of a microphone signal according to the present invention including noise-reduction and encoding/decoding.

[0032] In the example shown in Figure 1 a microphone signal $y(n)$ comprising speech $s(n)$ and background noise $b(n)$ (n being a discrete time index) is processed by an analysis filter bank 1 to achieve sub-band signals $Y(e^{j\Omega_\mu}, n)$ where Ω_μ denotes the mid-frequency of the μ -th frequency sub-band. Whereas in the following processing in the sub-band domain is described, alternatively the microphone signal could be subject to a Discrete Fourier Transformation, e.g., of the order of 256, in order to perform processing in the frequency domain. In this context, it should be noted that processing employing Bark or Mel grouping of frequency nodes might be preferred.

[0033] As illustrated in Figure 1 the sub-band signals $Y(e^{j\Omega_\mu}, n)$ are input in a noise reduction filtering means 2 that applies damping factors (filter coefficients) $G(e^{j\Omega_\mu}, n)$ to each of the sub-band signals $Y(e^{j\Omega_\mu}, n)$ in order obtain enhanced sub-band signals, i.e., a noise reduced spectrum $\hat{S}(e^{j\Omega_\mu}, n) = Y(e^{j\Omega_\mu}, n) G(e^{j\Omega_\mu}, n)$. The realization of the noise reduction filtering means 2 represents the kernel of the present invention.

[0034] In the art the damping factors $G(e^{j\Omega_\mu}, n)$ of the noise reduction filtering means are determined depending on the present signal-to-noise ratio (SNR) and the noise reduction filtering means is realized by some Wiener filter or employs spectral subtraction, etc. Usually, the damping factors $G(e^{j\Omega_\mu}, n)$ are determined based on an estimate of the short-time power density of the microphone signal

$$\hat{S}_{yy}(\Omega_{\mu}, n) = \left| Y(e^{j\Omega_{\mu}}, n) \right|^2$$

5
and an estimate of the power density of the background noise. The power density of the background noise is determined during speech pauses and might be temporarily smoothed

$$\hat{S}_{bb}(\Omega_{\mu}, n) = \begin{cases} \lambda \hat{S}_{bb}(\Omega_{\mu}, n-1) + (1-\lambda) \left| Y(e^{j\Omega_{\mu}}, n) \right|^2 \\ \hat{S}_{bb}(\Omega_{\mu}, n-1) \text{ else.} \end{cases}$$

15
in speech pauses, wherein λ denotes the smoothing time constant $0 \leq \lambda < 1$.

[0035] However, in the art the processing of the microphone signal for noise reduction does not take into account subsequently performed codec processing. Codec processing is a mandatory component of signal processing in the context of telephony. Well-known codec methods comprise Enhanced Variable Rate Codec (EVRC) and Enhanced Full Rate Codec (EFRC). Present day speech codec algorithms are usually based on the source-filter model for speech generation wherein the excitation signal and the spectral envelope are determined (see, e.g., Y. Griffin and J.S. Lim, "Multi-Band Excitation Vocoder", IEEE Transactions Acoustical Speech Signal Processing, Vol. 36, No. 8, pages 1223-1235, 1988).

[0036] Unvoiced sound is synthesized by means of noise generators. Voiced parts of the microphone signal (speech signal) are synthesized by estimating the pitch and determining the corresponding signal of a provided excitation code book, extracting the spectral envelope (e.g., by Linear Prediction Analysis or cepstral analysis, see, Y. Griffin and J.S. Lim, "Multi-Band Excitation Vocoder", IEEE Transactions Acoustical Speech Signal Processing, Vol. 36, No. 8, pages 1223-1235, 1988) and determining the best matching spectral envelope of a provided spectral envelope code book.

[0037] Common codec processing usually employs several different code books from which entries are chosen and the number of different code books considered depends on the actual SNR. If the SNR is high, a large number of code books is used in order to model the excitation signal as well as the spectral envelope. If the SNR is low or during speech pauses, the speech encoding rate is low and a relatively small number of code books is used.

[0038] The codec processing may significantly affect the quality of the noise reduced microphone signals. In the case of hands-free telephony in automobiles the codec processing can result in poor intelligibility of the speech signals sent to and received by a remote communication party when the travelling speed is high. Thus, even when the noise reduction processing itself is successful, the quality of the transmitted/received speech signal can be relatively poor.

[0039] In view of this, according to the present invention the noise reduction filtering means 2 is operated taking into account subsequent codec processing. In particular, the noise reduction filtering means 2 is adapted based on a variety of predetermined reference noise spectra that can be processed by the subsequent codec without generating disturbing artifacts, particularly, at transitions from speech activity and speech pauses. It is particularly advantageous to choose spectral envelopes used by the codec processing for low SNR or during speech pauses for the reference noise spectra.

[0040] The spectral envelopes can be described by an all-pole filter as it is known in the art

$$E_{cb}(e^{j\Omega_{\mu}}, m) = \frac{1}{1 - \sum_{k=1}^P a_k(m) e^{-j\Omega_{\mu}k}}, \quad m \in \{0, \dots, L-1\}$$

45
where $a_k(m)$ denotes the predictor coefficients (LPCs) which are used for modeling a spectral envelope during the speech codec processing and L represents the number of different predetermined reference noise spectra provided in the present example of the inventive method.

[0041] A noise estimator 3 estimates the power density $\hat{S}_{bb}(\Omega_{\mu}, n)$ of the background noise that is present in the microphone sub-band signals $Y(e^{j\Omega_{\mu}}, n)$. As shown in Figure 1 a database 4 comprising reference noise spectra is

provided and by a matching means 5 the particular one of the predetermined reference noise spectra is determined that matches best the estimated spectrum of the background noise

5

$$\hat{B}(e^{j\Omega_\mu}, n) = \sqrt{\hat{S}_{bb}(\Omega_\mu, n)}.$$

10 **[0042]** Since the background noise may be highly temporally varying, smoothing in frequency in the positive direction

$$\bar{B}'(e^{j\Omega_\mu}, n) = \begin{cases} \hat{B}(e^{j\Omega_\mu}, n), & \text{for } \mu = 0, \\ \lambda_F \bar{B}'(e^{j\Omega_{\mu-1}}, n) + (1 - \lambda_F) \hat{B}(e^{j\Omega_\mu}, n), & \text{for } \mu \in \{1, \dots, M - 1\}, \end{cases}$$

20

followed by smoothing in the negative direction

25

$$\bar{B}(e^{j\Omega_\mu}, n) = \begin{cases} \bar{B}'(e^{j\Omega_\mu}, n), & \text{for } \mu = M - 1, \\ \lambda_F \bar{B}(e^{j\Omega_{\mu+1}}, n) + (1 - \lambda_F) \bar{B}'(e^{j\Omega_\mu}, n), & \text{for } \mu \in \{0, \dots, M - 2\}, \end{cases}$$

30

with a smoothing parameter λ_F smaller than 1, in particular, smaller than 0.5, e.g., $\lambda_F = 0.3$, might be performed.

35 **[0043]** According to the present example, both the smoothed estimated noise spectrum and the reference noise spectra are logarithmized

40

$$\bar{B}_{\log}(e^{j\Omega_\mu}, n) = 20 \log_{10} \{ \bar{B}(e^{j\Omega_\mu}, n) \}$$

and

45

$$E_{cb, \log}(e^{j\Omega_\mu}, m) = 20 \log_{10} \{ E_{cb}(e^{j\Omega_\mu}, m) \},$$

50

respectively.

[0044] Since the actual noise may differ significantly from the reference noise spectra at low frequencies, it might be preferred to restrict the search for the best matching reference noise spectrum stored in the database 4 to a middle frequency range. For instance, sub-band signals for frequencies below some predetermined threshold $\Omega_{\mu,0}$, e.g. below some hundred Hz, in particular, below 300 - 700 Hz, more particularly, below 500 Hz might be ignored for the search. In addition, sub-band signals for frequencies above some predetermined threshold $\Omega_{\mu,1}$, e.g., some thousand Hz, in particular, for frequencies above 3000 or 3500 Hz, might be ignored for good matching results depending on the actual application.

[0045] In order to avoid that the search is affected by different gains/volumes of the noise, the logarithmic mean is subtracted from the smoothed estimated noise spectrum

5

$$\bar{B}_{\log,u}(e^{j\Omega_\mu}, n) = \bar{B}_{\log}(e^{j\Omega_\mu}, n) - \bar{B}_{\log,m}(n)$$

10 with

15

$$\bar{B}_{\log,m}(n) = \frac{1}{\mu_1 - \mu_0 + 1} \sum_{\mu=\mu_0}^{\mu_1} \bar{B}_{\log}(e^{j\Omega_\mu}, n).$$

20

[0046] Moreover, the logarithmic mean value of the reference noise spectra for the chosen frequency range is subtracted from the reference noise spectra

25

$$E_{cb,\log,u}(e^{j\Omega_\mu}, m) = E_{cb,\log}(e^{j\Omega_\mu}, m) - E_{cb,\log,m}(m)$$

with

30

$$E_{cb,\log,m}(m) = \frac{1}{\mu_1 - \mu_0 + 1} \sum_{\mu=\mu_0}^{\mu_1} E_{cb,\log}(e^{j\Omega_\mu}, m).$$

35

[0047] The search for the best matching one of the reference noise spectra can, e.g., be performed based on a logarithmic distance norm

40

$$m_{opt}(n) = \arg \min_m \sum_{\mu=\mu_0}^{\mu_1} \left(\bar{B}_{\log,u}(e^{j\Omega_\mu}, n) - E_{cb,\log,u}(e^{j\Omega_\mu}, m) \right)^2.$$

45

[0048] Other cost functions based, for instance, on the cepstral or LPC distance norm, might be employed for the search for the best matching reference noise spectrum that is carried out by the matching means 5.

[0049] After the best matching reference noise spectrum has been determined, the power is adjusted. After linearization one obtains

50

$$\hat{E}_{cb}(e^{j\Omega_\mu}, n) = 10^{(E_{cb,\log,u}(e^{j\Omega_\mu}, m_{opt}(n)) + \bar{B}_{\log,m}(n)) / 20}.$$

55

[0050] This spectrum is input in the noise reduction filtering means 2 by the matching means 5. It is noted that in the case of time-varying background noise, e.g., due to different driving situations in the context of a hands-free telephony

set installed in an automobile, the matching results differ in time. Hard switching from one best matching reference noise spectrum to another shall be avoided in order not to generate disturbing artifacts. For instance, recursive smoothing may advantageously be employed

5

$$\hat{E}_{cb,sm}(e^{j\Omega_\mu}, n) = \gamma_z \hat{E}_{cb,sm}(e^{j\Omega_\mu}, n-1) + (1 - \gamma_z) \hat{E}_{cb}(e^{j\Omega_\mu}, n)$$

10

with a time smoothing constant $0 \leq \gamma_z < 1$.

[0051] In the noise reduction filtering means 2 the modified best matching reference noise spectrum input by the matching means 5 is adapted with respect to the total power density according to

15

$$\tilde{E}_{cb}(e^{j\Omega_\mu}, n) = G_{cor}(n) \hat{E}_{cb,sm}(e^{j\Omega_\mu}, n)$$

20

with

25

$$G_{cor}(n) = \begin{cases} \Delta_{inc} G_{cor}(n-1), & \text{if } \sum_{\mu=\mu_2}^{\mu_3} \tilde{E}_{cb}^2(e^{j\Omega_\mu}, n-1) < \tilde{G}_{min}^2 \sum_{\mu=\mu_2}^{\mu_3} \hat{S}_{bb}(\Omega_\mu, n) \\ \Delta_{dec} G_{cor}(n-1), & \text{else,} \end{cases}$$

30

wherein \tilde{G}_{min} is a predetermined damping value for a predetermined frequency sub-band range $[\Omega_{\mu_2}, \Omega_{\mu_3}]$ by which the reference noise shall fall below the actual background noise and wherein Δ_{inc} and Δ_{dec} are multiplicative correcting constants that satisfy the relation

35

$$0 \ll \Delta_{dec} < 1 < \Delta_{inc} \ll \infty.$$

40

[0052] Experiments have proven that suitable choices for Ω_{μ_2} and Ω_{μ_3} are $\Omega_{\mu_2} = 500$ Hz and $\Omega_{\mu_3} = 700$ Hz, respectively. Maximum damping factors depending on time and frequency can be determined based on the adapted reference noise spectrum according to

45

$$G_{min}(e^{j\Omega_\mu}, n) = \min \left\{ G_0, \frac{\tilde{E}_{cb}(e^{j\Omega_\mu}, n)}{|Y(e^{j\Omega_\mu}, n)|} \right\}$$

50

with the predetermined minimum damping G_0 . A suitable choice for the minimum damping is $0.3 < G_0 < 0.7$, in particular, $G_0 = 0.5$. The thus obtained time and frequency selective maximum damping factors are used for determining the filter characteristics of the noise reduction filtering means 2. For instance, a recursive Wiener characteristics may be employed according to

55

$$G(e^{j\Omega_\mu}, n) = \max \left\{ G_{\min}(e^{j\Omega_\mu}, n), 1 - \beta(e^{j\Omega_\mu}, n) \frac{\hat{S}_{bb}(\Omega_\mu, n)}{\hat{S}_{yy}(\Omega_\mu, n)} \right\}$$

with real coefficients $\beta(e^{j\Omega_\mu}, n)$.

[0053] The microphone sub-band signals $Y(e^{j\Omega_\mu}, n)$ are filtered by the noise reduction filtering means 2 in order to obtain the noise reduced spectrum $\hat{S}(e^{j\Omega_\mu}, n) = Y(e^{j\Omega_\mu}, n) G(e^{j\Omega_\mu}, n)$. The noise reduced spectrum $\hat{S}(e^{j\Omega_\mu}, n)$ (noise reduced microphone sub-band signals) is input in a synthesis filter bank 6 to obtain the noise reduced total band signal $\hat{S}(n)$ in the time domain. Since this signal is obtained by means of the best matching reference noise spectrum of predetermined reference noise spectra that are also used for codec processing of the noise-reduced signal $s(n)$, the overall quality of a speech signal (microphone signal) transmitted to a remote party is significantly enhanced as compared to the art. In particular, artifacts at transitions of speech activity to speech pauses (gating effect) are reduced.

[0054] It is to be understood that the noise reduction filtering means 2, the noise estimator 3 and the matching means 5 of Figure 1 may or may not be realized in separate physical/processing units.

[0055] The signal processing described with reference to Figure 1 can be part of a method for electronically mediated verbal communication between two or more communication parties. In particular, it can be realized in hands-free telephony, e.g., by means of a hands-free set installed in an automobile. As already discussed audio signal processing in the context of telephony not only comprises noise reduction of signals detected by microphones but also codec processing.

[0056] Figure 2 illustrates an example of a method of processing a microphone signal $y(n)$ in order to obtain an encoded/decoded speech signal that is provided to a remote communication party. Consider a situation in that a near communication party makes use of a hands-free set installed in a vehicular cabin. The hands-free set comprises one or more microphones that detect the utterance of a user, i.e. a driver or other passenger sitting in the vehicular cabin. A microphone signal $y(n)$ corresponding to the utterance but also including some background noise is obtained by means of the at least one microphone.

[0057] This microphone signal $y(n)$ is processed as described with reference to Figure 1 in order to obtain an enhanced microphone signal (speech signal) $\hat{S}(n)$. The reference sign 10 in Figure 2 denotes a signal processing means comprising the analysis filter bank 1, noise reduction filtering means 2, noise estimator 3, reference noise database 4, matching means 5 and synthesis filter bank 6 of Figure 1. The enhanced signal $\hat{S}(n)$ is transmitted from the near party to a remote party by codec processing, e.g., EVRC or EFRC. Since the sampling rate of the speech encoding according to the present example is different from the sampling rate of the enhanced signal $\hat{S}(n)$ a first means for sampling rate conversion 11 adapts the sampling rate of $\hat{S}(n)$ to the one of the speech encoding performed by a speech encoder 12.

[0058] The encoded signal is wirelessly transmitted via some transmission channel 13 to a remote communication party. At the remote side a speech decoder 14 decodes the coded signal as known in the art and synthesizes a speech signal to be output by a loudspeaker. The decoded signal is subject to sampling rate conversion by a second means for sampling rate conversion 15 located at the remote site. The second means for sampling rate conversion 15 can, e.g., process the transmitted and decoded signal for bandwidth extension. Eventually, the re-sampled decoded signal $\hat{S}_{\text{cod}}(n)$ is output to a remote user.

[0059] Since noise-reduction of the microphone signal $y(n)$ by the means 10 of Figure 2 is carried out based on reference noise spectra that are also used for the codec processing, the quality of the output signal $\hat{S}_{\text{cod}}(n)$ is significantly enhanced as compared to conventional noise reduction and codec processing of a speech signal to be transmitted from a near communication party to a remote communication party.

Claims

1. Method for signal processing comprising the steps of
 - providing a set of prototype spectral envelopes;
 - providing a set of reference noise prototypes, wherein the reference noise prototypes are obtained from at least a sub-set of the provided set of prototype spectral envelopes;
 - detecting a verbal utterance by at least one microphone to obtain a microphone signal;
 - processing the microphone signal for noise reduction based on the provided reference noise prototypes to obtain an enhanced signal; and
 - encoding the enhanced signal based on the provided prototype spectral envelopes to obtain an encoded enhanced signal.

2. The method according to claim 1, further comprising transmitting the encoded enhanced signal to a remote party; receiving the transmitted encoded enhanced signal by the remote party; and decoding the received signal by the remote party.
3. The method according to claim 1 or 2, wherein the provided set of prototype spectral envelopes is used for encoding the enhanced signal in speech pauses detected in the microphone signal or when a signal-to-noise ratio of the microphone signal falls below a predetermined threshold.
4. The method according to one of the preceding claims, wherein the reference noise prototypes are spectral envelopes modeled by an all-pole filter function.
5. The method according to one of the preceding claims, wherein the processing of the microphone signal for noise reduction comprises estimating the power density of a noise contribution in the microphone signal; matching the spectrum of the noise contribution obtained from the estimated power density of the noise contribution with the provided set of reference noise prototypes to find the best matching reference noise prototype; and using the best matching reference noise prototype to determine maximum damping factors for noise reduction of the microphone signal.
6. The method according to claim 5, wherein the processing of the microphone signal for noise reduction is performed by a Wiener-like filtering means comprising damping factors obtained based on the best matching reference noise prototype, the power density spectrum of sub-band signals obtained from the microphone signal and the estimated power density spectrum of the background noise.
7. The method according to claim 5 or 6, wherein the spectrum of the noise contribution obtained from the estimated power density of the noise contribution is matched only with a subset of the provided reference noise prototypes within a predetermined frequency range.
8. Method for speech communication with a hands-free set installed in a vehicle, particular, an automobile, comprising the method according to one of the preceding claims, wherein at least one of the provided reference noise prototypes on which the processing of the microphone signal for noise reduction to obtain an enhanced signal is based is determined from a sub-set of the provided set of reference noise prototypes that is selected according to a current traveling speed of the vehicle, in particular, the automobile; and/or the reference noise prototypes are obtained from a sub-set of the provided set of prototype spectral envelopes selected according to the type of the vehicle, in particular, the automobile.
9. Computer program product comprising at least one computer readable medium having computer-executable instructions for performing one or more steps of the method of one of the preceding claims when run on a computer.
10. Signal processing means, comprising an encoding database comprising prototype spectral envelopes; a reference database comprising reference noise prototypes, wherein the reference noise prototypes are obtained from at least a sub-set of the provided set of prototype spectral envelopes; and a noise reduction filtering means configured to process a microphone signal comprising background noise based on the reference noise prototypes to obtain an enhanced microphone signal; and an encoder configured to encode the enhanced microphone signal based on the prototype spectral envelopes.
11. The signal processing means according to claim 10, further comprising a noise estimating means configured to estimate the power density of a background noise contribution of the microphone signal; a matching means configured to match the spectrum of the noise contribution obtained from the estimated power density of the noise contribution with the set of reference noise prototypes comprised in the reference database to find the best matching reference noise prototype; and wherein the noise reduction filtering means is configured to use the best matching reference noise prototype for noise reduction of the microphone signal.
12. The signal processing means according to claim 11, wherein the noise reduction filtering means is a Wiener-like

filtering means comprising damping factors obtained based on the best matching reference noise prototype, the power density spectrum of microphone sub-band signals obtained from the microphone signal and the estimated power density spectrum of the background noise.

- 5 **13.** The signal processing means according to one of the claims 10 to 12,
 wherein the noise reduction filtering means is configured to operate in the sub-band regime and to output noise-
 reduced microphone sub-band signals;
 and further comprising
 an analysis filter bank configured to process the microphone signal to obtain microphone sub-band signals and to
10 provide the microphone sub-band signals to the noise reduction filtering means; and
 a synthesis filter bank configured to process the noise-reduced microphone sub-band signals to obtain a noise-
 reduced full-band microphone signal in the time domain.
- 15 **14.** The signal processing means according to one of the claims 10 to 13, wherein the signal processing means is
 installed in an automobile and the reference database is derived from the encoding database dependent on type
 of the automobile.
- 20 **15.** The signal processing means according to one of the claims 10 to 14, further comprising a control means configured
 to control determination of at least one of the reference noise prototypes used by the noise reduction filtering means
 to process the microphone signal to obtain the enhanced microphone signal based on a current traveling speed of
 the automobile.

25

30

35

40

45

50

55

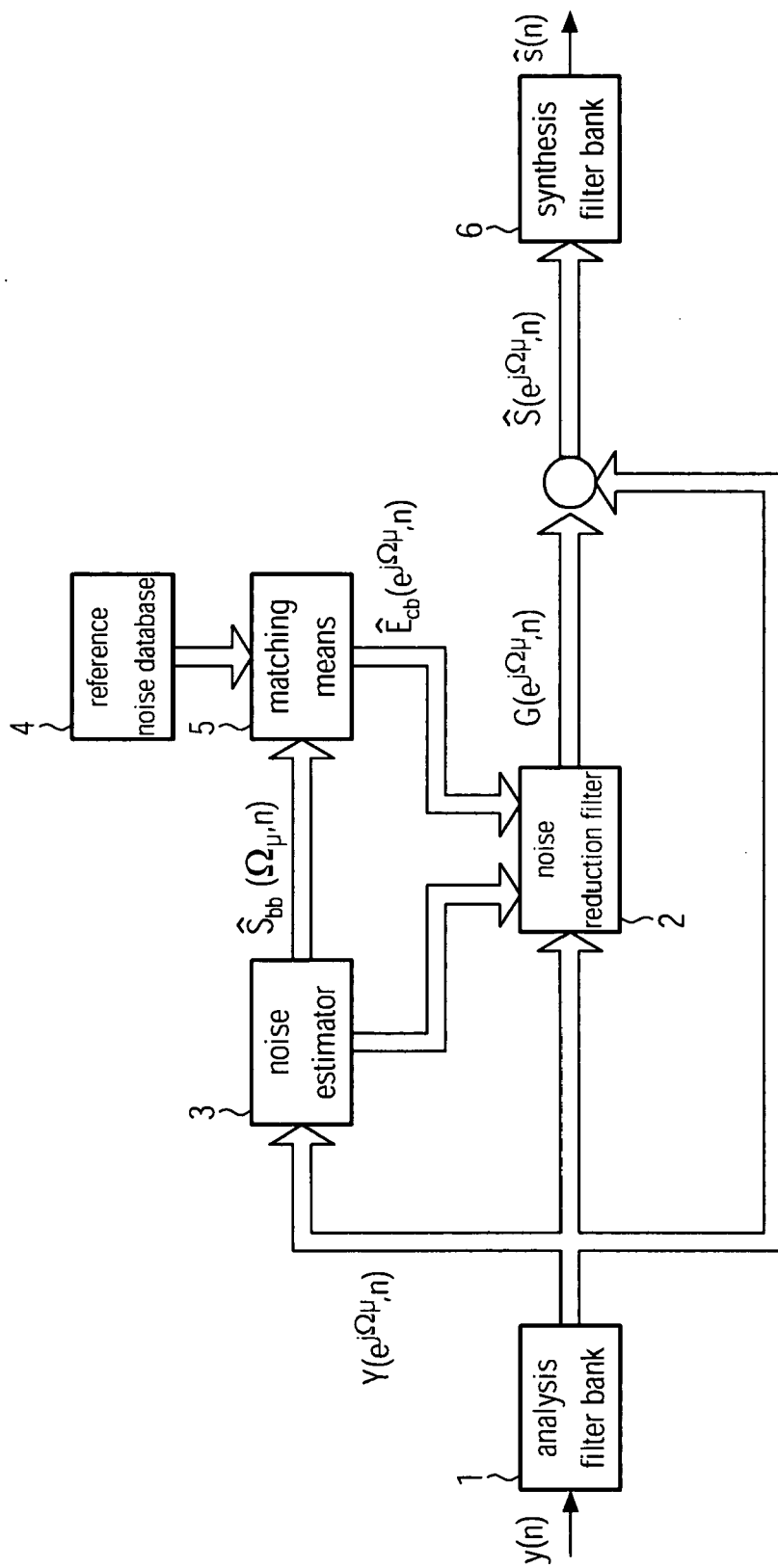


FIG. 1

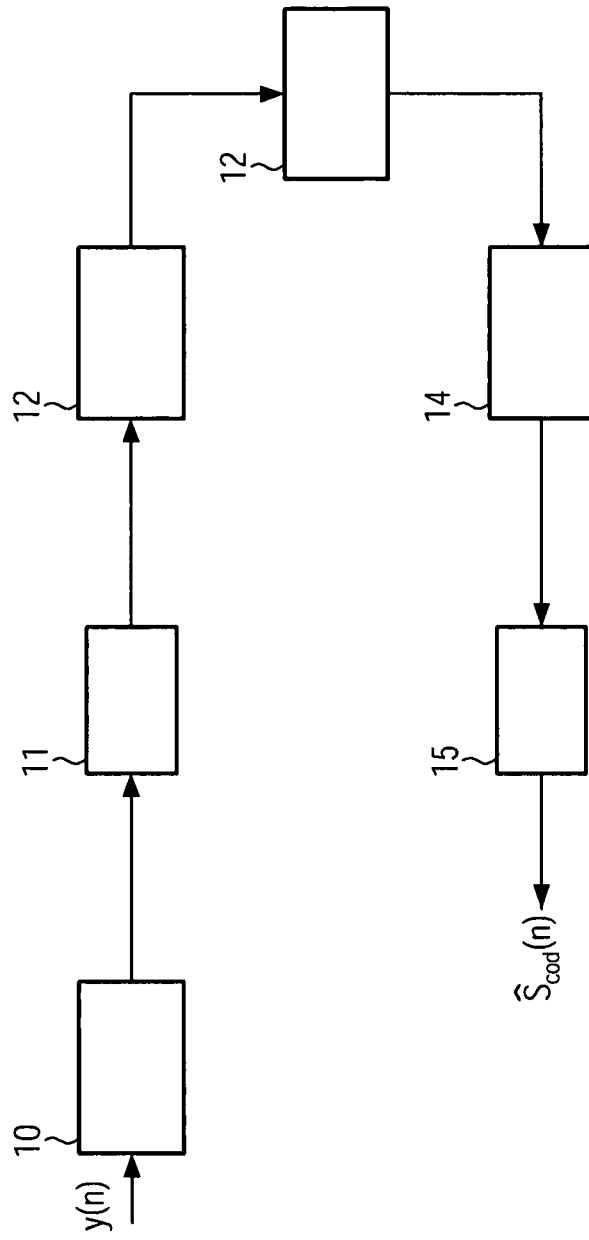


FIG. 2



EUROPEAN SEARCH REPORT

Application Number
EP 08 01 4151

| 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 | US 2002/035471 A1 (BRETON PIERRE-ALBERT [FR]) 21 March 2002 (2002-03-21) * paragraphs [0004], [0005], [0070], [0072] - [0074], [0077], [0082], [0083] * | 1-4,9, 10,13 6 | INV. G10L21/02 G10L19/00 |
| X | ----- YOSHITAKA NISHIMURA ET AL: "Speech Recognition for a Humanoid with Motor Noise Utilizing Missing Feature Theory" HUMANOID ROBOTS, 2006 6TH IEEE-RAS INTERNATIONAL CONFERENCE ON, IEEE, PI, 1 December 2006 (2006-12-01), pages 26-33, XP031052994 ISBN: 978-1-4244-0199-4 | 1,5,6, 9-12 | |
| Y | * page 27, right-hand column, line 12 * * page 29, right-hand column, line 9 - page 30, left-hand column, line 12 * | 7 | |
| X | ----- SHINGO KUROIWA ET AL: "Wind noise reduction method for speech recording using multiple noise templates and observed spectrum fine structure" COMMUNICATION TECHNOLOGY, . 2006. ICCT '06. INTERNATIONAL CONFERENCE ON, IEEE, PI, 1 November 2006 (2006-11-01), pages 1-5, XP031071933 ISBN: 978-1-4244-0800-9 * abstract * * chapter III * | 1,9,10 | TECHNICAL FIELDS SEARCHED (IPC) G10L |
| Y | | 7 | |
| X | ----- DE 10 2004 012209 A1 (SIEMENS AG [DE]) 6 October 2005 (2005-10-06) * paragraphs [0001], [0009] - [0015], [0029] - [0033] * ----- -/-- | 1,8-10, 14,15 | |
| 2 The present search report has been drawn up for all claims | | | |
| Place of search Munich | | Date of completion of the search 15 December 2008 | Examiner Geißler, Christian |
| 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 | | | |

EPO FORM 1503 03.82 (P04001)



EUROPEAN SEARCH REPORT

Application Number
EP 08 01 4151

| DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|---|---|---|
| Category | Citation of document with indication, where appropriate, of relevant passages | Relevant to claim |
| A | EP 1 258 715 A (MATSUSHITA COMM IND CO LTD [JP]; JAPAN GOVERNMENT [JP] PANASONIC MOBIL) 20 November 2002 (2002-11-20) * paragraphs [0011], [0017], [0018], [0027], [0042], [0057] * ----- | 3 |
| | | CLASSIFICATION OF THE APPLICATION (IPC) |
| | | 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 |
| Munich | 15 December 2008 | Geißler, Christian |
| 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 03.02 (P/04C01)

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

EP 08 01 4151

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.

15-12-2008

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|--|------------------|-------------------------|------------------|
| US 2002035471 A1 | 21-03-2002 | DE 60123161 T2 | 06-09-2007 |
| | | EP 1154405 A1 | 14-11-2001 |
| | | FR 2808917 A1 | 16-11-2001 |
| ----- | | | |
| DE 102004012209 A1 | 06-10-2005 | NONE | |
| ----- | | | |
| EP 1258715 A | 20-11-2002 | AU 8261601 A | 22-03-2002 |
| | | WO 0221091 A1 | 14-03-2002 |
| | | JP 3670217 B2 | 13-07-2005 |
| | | JP 2002156999 A | 31-05-2002 |
| | | US 2002165681 A1 | 07-11-2002 |
| ----- | | | |

EPC 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

- **Y. Griffin ; J.S. Lim.** Multi-Band Excitation Vocoder.
IEEE Transactions Acoustical Speech Signal Processing, 1988, vol. 36 (8), 1223-1235 [0006]
[0035] [0036]