(19) DANMARK

(10) DK/EP 3252766 T3

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: *G 10 L 21/0264 (2013.01)*

(45) Oversættelsen bekendtgjort den: **2021-09-06**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2021-07-07**

(86) Europæisk ansøgning nr.: **16171986.9**

(86) Europæisk indleveringsdag: **2016-05-30**

(87) Den europæiske ansøgnings publiceringsdag: **2017-12-06**

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

(73) Patenthaver: **Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark**

(72) Opfinder: **Jensen, Jesper, c/o Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark**
KJEMS, Ulrik, Gammel Kongevej 163, 2. th, 1850 Frederiksberg C, Danmark
BERTELSEN, Andreas Thelander, c/o Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark
PEDERSEN, Michael Syskind, c/o Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark

(74) Fuldmægtig i Danmark: **Oticon A/S, Kongebakken 9, 2765 Smørum, Danmark**

(54) Benævnelse: **AUDIOBEHANDLINGSANORDNING OG FREMGANGSMÅDE TIL ESTIMERING AF SIGNAL-TIL-STØJ-FORHOLDET FOR ET LYDSIGNAL**

(56) Fremdragne publikationer:
WO-A1-2013/065010
WO-A1-2015/189261
EPHRAIM, Y.; MALAH, D.: "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator", IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol. 32, no. 6, December 1984 (1984-12), pages 1109-1121, XP002764005,
EPHRAIM, Y.; MALAH, D.: "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator", IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol. 33, no. 2, April 1985 (1985-04), pages 443-445, XP002764006,
V. HAMACHER ET AL: "Signal Processing in High-End Hearing Aids: State of the Art, Challenges, and Future Trends", EURASIP JOURNAL ON ADVANCES IN SIGNAL PROCESSING, vol. 2005, no. 18, 1 January 2005 (2005-01-01), pages 2915-2929, XP055318245, US ISSN: 1687-6172, DOI: 10.1155/ASP.2005.2915
ERKELENS J S ET AL: "Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients With Generalized Gamma Priors", IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, USA, vol. 15, no. 6, 1 August 2007 (2007-08-01) , pages 1741-1752, XP011187708, ISSN: 1558-7916, DOI: 10.1109/TASL.2007.899233

DESCRIPTION

SUMMARY

[0001] The present disclosure relates to a hearing aid, and a method of estimating an a priori signal to noise ratio of an electric input signal representing sound. The disclosure relates specifically to a scheme for obtaining an a priori signal-to-noise-ratio estimate by non-linear smoothing of an a posteriori signal-to-noise-ratio estimate.

[0002] In the present context 'an a posteriori signal to noise ratio', SNR_{post}, is taken to mean a ratio between the observed (available) noisy signal (target signal S plus noise *N*, *Y(t)=S(t)+N(t)*), e.g. a picked up by one or more microphones, such as the power of the noisy signal, and the noise *N(t)*, such as an estimate ($\hat{N}(t)$) of the noise, such as the power of the noise signal, at a given point in time *t*, i.e. $SNR_{post}(t) = Y(t)/\hat{N}(t)$. or $SNR_{post}(t) = Y(t)^2/\hat{N}(t)^2$. The 'a posteriori signal to noise ratio', SNR_{post}, may e.g. be defined in the time-frequency domain as a value for each frequency band (index k) and time frame (index n), i.e. $SNR_{post} = SNR_{post}(k,n)$, i.e. e.g. $SNR_{post}(k,n) = |Y(k,n)|^2/|\hat{N}(k,n)|^2$. Examples of the generation of an 'a posteriori' signal to noise ratio are illustrated in FIG. 1A and 1B for a one-microphone and a multi-microphone setup, respectively.

[0003] In the present context 'an a priori signal to noise ratio' SNR_{prio} is taken to mean a ratio of the target signal amplitude *S(t)* (or of the target signal power $S(t)^2$) to the noise signal amplitude *N(t)* (or to the noise signal power $N(t)^2$), respectively, such as a ratio between *estimates* of these signals at a given point in time *t*, e.g. $SNR_{prio} = SNR_{prio}(t) = \hat{S}(t)^2/\hat{N}(t)^2$. or $SNR_{prio} = SNR_{prio}(k,n)$, i.e. e.g. $SNR_{prio}(k,n) = |\hat{S}(k,n)|^2/|\hat{N}(k,n)|^2$.

[0004] Reference [1] deals with a class of speech enhancement systems which capitalize on the major importance of the short-time spectral amplitude (STSA) of the speech signal in its perception. A system which utilizes a minimum mean-square error (MMSE) STSA estimator is proposed and then compared with other widely used systems which are based on Wiener filtering and the· "spectral subtraction" algorithm.

[0005] WO2015189261A1 relates to a multi-band noise reduction system for digital audio signals producing a noise reduced digital audio output signal from a digital audio signal. The digital audio signal comprises a target signal and a noise signal, i.e. a noisy digital audio signal. The multi-band noise reduction system operates on a plurality of sub-band signals de-rived from the digital audio signal and comprises a second or adaptive signal-to-noise ratio estimator which is configured for filtering a plurality of first signal-to-noise ratio estimates of the plurality of sub-band signals with respective time-varying low-pass filters to produce respective second signal-to- noise ratio estimates of the plurality of sub-band signals. A low-pass cut-off

frequency of each of the time-varying low-pass filters is adaptable in accordance with a first signal-to-noise ratio estimate determined by a first signal-to-noise ratio estimator and/or the second signal-to-noise ratio estimate of the sub-band signal.

A hearing aid:

[0006] In an aspect of the present application, an audio processing device, as defined in claim 1 is provided. The audio processing device, a hearing aid, comprises the features of claim 1.

[0007] Thereby an improved noise reduction may be provided.

[0008] The noise signal components N(k,n) may e.g. originate from one or more other sources NS_i (i=1, ..., N_S) than the target sound source TS. In an embodiment, the noise signal components N(k,n) include late reverberations from the target signal (e.g. target signal components that arrive at the user more than 50 ms later than the dominant peak of the target signal component in question).

[0009] In other words, $\zeta(k,n) = F(\zeta(k,n-1), \gamma(k,n))$. Using the most recent frame power for the a posteriori SNR (SNR=Signal to Noise Ratio) in the determination of the a priori SNR may e.g. be beneficial for SNR estimation at speech onsets, where large increases to the SNR typically occur over a short time.

[0010] In an embodiment, the noise reduction system is configured to determine said a priori target signal to noise ratio estimate $\zeta(k,n)$ for the n^{th} time frame under the assumption that $\gamma(k,n)$ is larger than or equal to 1. In an embodiment, the a posteriori signal to noise ratio estimate $\gamma(k,n)$ of the electric input signal Y(k,n) is e.g. defined as the ratio between a signal power spectral density $|Y(k,n)|^2$ of the current value Y(k,n) of the electric input signal and an estimate $<\sigma^2>$ of the current noise power spectral density of the electric input signal Y(k,n), i.e. $\gamma(k,n) = |Y(k,n)|^2/<\sigma^2>$.

[0011] The noise reduction system is configured to determine said a priori target signal to noise ratio estimate $\zeta(k,n)$ for the n^{th} timeframe from said a priori target signal to noise ratio estimate $\zeta(k,n-1)$ for the $(n-1)^{th}$ timeframe, and from the maximum likelihood SNR estimator $\zeta^{ML}(k,n)$ of the a priori target signal to noise ratio estimate $\zeta(k,n)$ for the n^{th} timeframe.

[0012] The noise reduction system is configured to determine said maximum likelihood SNR estimator $\zeta^{ML}(k,n)$ as $MAX\{\zeta^{ML}_{min}(k,n); \gamma(k,m)-1\}$, where MAX is the maximum operator, and $\zeta^{ML}_{min}(k,n)$ is a minimum value of the maximum likelihood SNR estimator $\zeta^{ML}(k,n)$. In an embodiment, the minimum value $\zeta^{ML}_{min}(k,n)$ of the maximum likelihood SNR estimator $\zeta^{ML}(k,n)$ may e.g. be dependent of frequency band index. In an embodiment, the minimum value

$\zeta^{ML}_{min}(k,n)$ is independent. In an embodiment $\zeta^{ML}_{min}(k,n)$ is taken to be equal to '1' (i.e. = 0 dB on a logarithmic scale). This is e.g. the case when the target signal components S(k,m) are negligible; i.e. when only noise components N(k,m) are present in the input signal Y(k,m)).

[0013] In an embodiment, the noise reduction system is configured to determine said a priori target signal to noise ratio estimate ζ by non-linear smoothing of said a posteriori signal to noise ratio estimate γ, or a parameter derived therefrom. A parameter derived therefrom may e.g. be the maximum likelihood SNR estimator ζ^{ML}.

[0014] In an embodiment, the noise reduction system is configured to provide an SNR-dependent smoothing, allowing for more smoothing in low SNR conditions than for high SNR conditions. This may have the advantage of reducing musical noise. The terms 'low SNR conditions' and 'high SNR conditions' are intended to indicate first and second conditions where the true SNR is lower under the first conditions than under the second conditions. In an embodiment, 'low SNR conditions' and 'high SNR conditions' are taken to mean below and above 0 dB, respectively. Preferably, the dependence a time constant controlling the smoothing exhibit a gradual change in dependence of SNR. In an embodiment, the time constant(s) involved in smoothing are higher the lower the SNR. At 'low SNR conditions', the SNR estimate is generally relatively poorer than at 'high SNR conditions' (and hence less trustworthy at lower SNR; and hence a driver for more smoothing).

[0015] In an embodiment, the noise reduction system is configured to provide a negative bias compared to

ξ_n^{ML}

for low SNR conditions. This may have the advantage of reducing audibility of musical noise in noise-only periods. The term "bias" is in the present context used to reflect a difference between the expected value

$E(\xi_n^{\mathrm{ml}})$

of the maximum likelihood SNR estimator $\zeta^{ML}(k,n)$ and the expected value $E(\zeta_n)$ of the a priori signal to noise ratio ζ(k,n). In other words, for 'low SNR conditions' (e.g. for true SNR < 0 dB),

$E(\xi_n^{\mathrm{ml}}) - E(\zeta_n) < 0$

(as e.g. reflected in FIG. 3).

[0016] In an embodiment, the noise reduction system is configured to provide a recursive bias, allowing a configurable change from low-to-high and high-to-low SNR conditions.

[0017] In a logarithmic representation of the a priori signal to noise ratio for the n^{th} time frame may be expressed as $s_n=s(k,n)=10\log(\zeta(k,n))$ and correspondingly for the maximum likelihood SNR estimator for the n^{th} time frame: $s^{ML}_n=s^{ML}(k,n)=10\log(\zeta^{ML}(k,n))$.

[0018] The noise reduction system is configured to determine said a priori target signal to noise ratio estimate ζ(k,n) for the n^{th} timeframe from said a priori target signal to noise ratio

estimate $\zeta(k,n-1)$ for the $(n-1)^{th}$ timeframe, and from the maximum likelihood SNR estimator $\zeta^{ML}(k,n)$ of the a priori target signal to noise ratio estimate $\zeta(k,n)$ for the n^{th} time frame according to the following recursive algorithm:

$$s_n - s_{n-1} = (s_n^{ML} + \rho(s_{n-1}) - s_{n-1})\lambda(s_{n-1})$$

where $\rho(s_{n-1})$ represents a bias function or parameter and $\lambda(s_{n-1})$ represents a smoothing function or parameter of the $(n-1)^{th}$ time frame.

[0019] In an embodiment, $\rho(s_{n-1})$ is chosen as to be equal to the value of

$$10\log_{10}\frac{\xi}{\xi_{n-1}}$$

with ξ satisfying

$$10\log_{10}\Psi\left(\xi_{n-1},\frac{\xi}{\xi_{n-1}}\right) = 0\text{dB},$$

where

$$\Psi\left(\xi_{n-1},\frac{\xi}{\xi_{n-1}}\right)$$

is a non-linear function as defined in equation (8).

[0020] In an embodiment, the smoothing function $\lambda(s_{n-1})$ is chosen to be equal to the slope (w.r.t.

$$s_n^{ML}$$

) of the function

$$10\log_{10}\Psi\left(\xi_{n-1},\frac{\xi}{\xi_{n-1}}\right) = 0\text{dB}$$

(cf. curves in FIG. 3) at the location of the 0 dB crossing (i.e. when s_n^{ML}-s_n = $\rho(s_{n-1})$).

[0021] In an embodiment, the audio processing device comprises a filter bank comprising an analysis filter bank for providing said time-frequency representation Y(k,n) of said electric input signal. In an embodiment, the electric input signal is available as a number of frequency sub-band signals Y(k,n), k=1, 2, ..., K. In an embodiment, the a priori signal to noise ratio estimate $\zeta(k,n)$, depend on the a posteriori signal to noise ratio estimate $\gamma(k,n)$ in a neighboring frequency sub-band signal (e.g. on $\gamma(k-1,n)$ and/or $\gamma(k+1,n)$.

[0022] In an embodiment, the audio processing device is configured to provide that said analysis filter bank is oversampled. In an embodiment, the audio processing device is configured to provide that the analysis filter bank is a DFT-modulated analysis filter bank.

[0023] In an embodiment, the the recursive loop of the algorithm for determining said a priori target signal to noise ratio estimate $\zeta(k,n)$ for the n^{th} timeframe comprises a higher order delay element, e.g. a circular buffer. In an embodiment, the higher order delay element is configured to compensate for oversampling of the analysis filter bank.

[0024] In an embodiment, the noise reduction system is configured to adapt the algorithm for determining said a priori target signal to noise ratio estimate $\zeta(k,n)$ for the n^{th} timeframe

comprises to compensate for oversampling of the analysis filter bank. In an embodiment, the algorithm comprises a smoothing parameter (λ) and/or a bias parameter (ρ). In an embodiment, the smoothing parameter (λ) and/or a bias parameter (ρ) are adapted to compensate for a sampling rate, see e.g. FIG. 5. In an embodiment, different oversampling rates are compensated for by adapting the parameter α, cf. e.g. FIG. 8.

[0025] In an embodiment, the audio processing device comprises a hearing device, such as a hearing aid, a headset, an earphone, an ear protection device or a combination thereof.

[0026] In an embodiment, the audio processing device is adapted to provide a frequency dependent gain and/or a level dependent compression and/or a transposition (with or without frequency compression) of one or frequency ranges to one or more other frequency ranges, e.g. to compensate for a hearing impairment of a user and/or to compensate for challenging acoustic environment. In an embodiment, the audio processing device comprises a signal processing unit for enhancing the input signals and providing a processed output signal.

[0027] In an embodiment, the audio processing device comprises an output unit for providing a stimulus perceived by the user as an acoustic signal based on a processed electric signal. In an embodiment, the output unit comprises a number of electrodes of a cochlear implant or a vibrator of a bone conducting hearing device. In an embodiment, the output unit comprises an output transducer. In an embodiment, the output transducer comprises a receiver (loudspeaker) for providing the stimulus as an acoustic signal to the user. In an embodiment, the output transducer comprises a vibrator for providing the stimulus as mechanical vibration of a skull bone to the user (e.g. in a bone-attached or bone-anchored hearing device).

[0028] In an embodiment, the audio processing device comprises an input unit for providing an electric input signal representing sound. In an embodiment, the input unit comprises an input transducer, e.g. a microphone, for converting an input sound to an electric input signal. In an embodiment, the input unit comprises a wireless receiver for receiving a wireless signal comprising sound and for providing an electric input signal representing said sound.

[0029] In an embodiment, the audio processing device is portable device, e.g. a device comprising a local energy source, e.g. a battery, e.g. a rechargeable battery.

[0030] In an embodiment, an a priori SNR estimate of a given hearing aid that forms part of a binaural hearing aid system is based on a posteriori SNR estimates from both hearing aids of the binaural hearing aid system. In an embodiment, an a priori SNR estimate of a given hearing aid that forms part of a binaural hearing aid system is based on an a posteriori SNR estimate of the given hearing aid and an a priori SNR estimate of the other hearing aid of the binaural hearing aid system.

[0031] In an embodiment, the audio processing device comprises a forward (or signal) path between an input transducer (microphone system and/or direct electric input (e.g. a wireless receiver)) and an output transducer. In an embodiment, the signal processing unit is located in

the forward path. In an embodiment, the signal processing unit is adapted to provide a frequency dependent gain according to a user's particular needs. In an embodiment, the audio processing device comprises an analysis (or control) path comprising functional components for analyzing the input signal (e.g. determining a level, a modulation, a type of signal, an acoustic feedback estimate, etc.), and possibly controlling processing of the forward path. In an embodiment, some or all signal processing of the analysis path and/or the signal path is conducted in the frequency domain. In an embodiment, some or all signal processing of the analysis path and/or the signal path is conducted in the time domain.

[0032] In an embodiment, the analysis (or control) path is operated in fewer channels (or frequency sub-bands) than the forward path. This can e.g. be done to save power in an audio processing device, such as a portable audio processing device, e.g. a hearing aid, where power consumption is an important parameter.

[0033] In an embodiment, an analogue electric signal representing an acoustic signal is converted to a digital audio signal in an analogue-to-digital (AD) conversion process, where the analogue signal is sampled with a predefined sampling frequency or rate f_S, f_S being e.g. in the range from 8 kHz to 48 kHz (adapted to the particular needs of the application) to provide digital samples x_n (or x[n]) at discrete points in time t_n (or n), each audio sample representing the value of the acoustic signal at t_n by a predefined number N_S of bits, N_S being e.g. in the range from 1 to 16 bits. A digital sample x has a length in time of $1/f_S$, e.g. 50 μs, for f_S = 20 kHz. In an embodiment, a number of audio samples are arranged in a time frame. In an embodiment, a time frame comprises 64 or 128 audio data samples. Other frame lengths may be used depending on the practical application. In an embodiment, a frame is shifted every ms or every 2 ms in case of oversampling (e.g. in case a critical sampling (no frame overlap) corresponds to a frame length of 3.2 ms (e.g. for f_S=20 kHz, and 64 samples per frame)). In other words the frames overlap, so that a only a certain fraction of samples are new from a given frame to the next, e.g. 25% or 50% or 75% of the samples.

[0034] In an embodiment, the audio processing devices comprise an analogue-to-digital (AD) converter to digitize an analogue input with a predefined sampling rate, e.g. 20 kHz. In an embodiment, the audio processing devices comprise a digital-to-analogue (DA) converter to convert a digital signal to an analogue output signal, e.g. for being presented to a user via an output transducer.

[0035] In an embodiment, the audio processing device, e.g. the microphone unit, and or the transceiver unit comprise(s) a TF-conversion unit for providing a time-frequency representation of an input signal. In an embodiment, the time-frequency representation comprises an array or map of corresponding complex or real values of the signal in question in a particular time and frequency range. In an embodiment, the TF conversion unit comprises a filter bank for filtering a (time varying) input signal and providing a number of (time varying) output signals each comprising a distinct frequency range of the input signal. In an embodiment, the TF conversion unit comprises a Fourier transformation unit for converting a time variant input signal to a (time

variant) signal in the frequency domain. In an embodiment, the frequency range considered by the audio processing device from a minimum frequency f_{min} to a maximum frequency f_{max} comprises a part of the typical human audible frequency range from 20 Hz to 20 kHz, e.g. a part of the range from 20 Hz to 12 kHz. In an embodiment, a signal of the forward and/or analysis path of the audio processing device is split into a number *NI* of frequency bands, where NI is e.g. larger than 5, such as larger than 10, such as larger than 50, such as larger than 100, such as larger than 500, at least some of which are processed individually. In an embodiment, the audio processing device is/are adapted to process a signal of the forward and/or analysis path in a number *NP* of different frequency channels ($NP \leq NI$). The frequency channels may be uniform or non-uniform in width (e.g. increasing in width with frequency), overlapping or non-overlapping.

[0036] In an embodiment, the audio processing device comprises a number of detectors configured to provide status signals relating to a current physical environment of the audio processing device (e.g. the current acoustic environment), and/or to a current state of the user wearing the audio processing device, and/or to a current state or mode of operation of the audio processing device. Alternatively or additionally, one or more detectors may form part of an *external* device in communication (e.g. wirelessly) with the audio processing device. An external device may e.g. comprise another hearing assistance device, a remote control, and audio delivery device, a telephone (e.g. a Smartphone), an external sensor, etc.

[0037] In an embodiment, one or more of the number of detectors operate(s) on the full band signal (time domain). In an embodiment, one or more of the number of detectors operate(s) on band split signals ((time-) frequency domain).

[0038] In an embodiment, the number of detectors comprises a level detector for estimating a current level of a signal of the forward path. In an embodiment, the predefined criterion comprises whether the current level of a signal of the forward path is above or below a given (L-)threshold value.

[0039] In a particular embodiment, the audio processing device comprises a voice detector (VD) for determining whether or not an input signal comprises a voice signal (at a given point in time). A voice signal is in the present context taken to include a speech signal from a human being. It may also include other forms of utterances generated by the human speech system (e.g. singing). In an embodiment, the voice detector unit is adapted to classify a current acoustic environment of the user as a VOICE or NO-VOICE environment. This has the advantage that time segments of the electric microphone signal comprising human utterances (e.g. speech) in the user's environment can be identified, and thus separated from time segments only comprising other sound sources (e.g. artificially generated noise). In an embodiment, the voice detector is adapted to detect as a VOICE also the user's own voice. Alternatively, the voice detector is adapted to exclude a user's own voice from the detection of a VOICE.

[0040] In an embodiment, the audio processing device comprises an own voice detector for

detecting whether a given input sound (e.g. a voice) originates from the voice of the user of the system. In an embodiment, the microphone system of the audio processing device is adapted to be able to differentiate between a user's own voice and another person's voice and possibly from NON-voice sounds.

[0041] In an embodiment, the hearing assistance device comprises a classification unit configured to classify the current situation based on input signals from (at least some of) the detectors, and possibly other inputs as well. In the present context 'a current situation' is taken to be defined by one or more of

1. a) the physical environment (e.g. including the current electromagnetic environment, e.g. the occurrence of electromagnetic signals (e.g. comprising audio and/or control signals) intended or not intended for reception by the audio processing device, or other properties of the current environment than acoustic;
2. b) the current acoustic situation (input level, feedback, etc.), and
3. c) the current mode or state of the user (movement, temperature, etc.);
4. d) the current mode or state of the hearing assistance device (program selected, time elapsed since last user interaction, etc.) and/or of another device in communication with the audio processing device.

[0042] In an embodiment, the audio processing device further comprises other relevant functionality for the application in question, e.g. compression, amplification, feedback reduction, etc.

[0043] In an embodiment, the audio processing device comprises a listening device, such as a hearing device, e.g. a hearing aid, e.g. a hearing instrument, e.g. a hearing instrument adapted for being located at the ear or fully or partially in the ear canal of a user, e.g. a headset, an earphone, an ear protection device or a combination thereof.

Use:

[0044] In an aspect, use of a audio processing device as described above, in the 'detailed description of embodiments' and in the claims, is moreover provided.

A method:

[0045] In an aspect, a method of estimating an a priori signal to noise ratio $\zeta(k,n)$ of a time-frequency representation $Y(k,n)$ of an electric input signal representing a time variant sound signal consisting of target speech components and noise components, where k and n are frequency band and time frame indices, respectively, as defined in claim 14, is furthermore

provided by the present application. The method comprises the features of claim 14.

[0046] It is intended that some or all of the structural features of the device described above, in the 'detailed description of embodiments' or in the claims can be combined with embodiments of the method, when appropriately substituted by a corresponding process and vice versa. Embodiments of the method have the same advantages as the corresponding devices.

[0047] The estimates of magnitudes A(k,n) of said target speech components are determined from said electric input signal Y(k,n) multiplied by a gain function G, where said gain function G is a function of said a posteriori signal to noise ratio estimate γ(k,n) and said a priori target signal to noise signal ratio estimate ζ(k,n).

[0048] In an embodiment, the method comprises providing an SNR-dependent smoothing, allowing for more smoothing in low SNR conditions than for high SNR conditions.

[0049] In an embodiment, the method comprises a smoothing parameter (λ) and/or a bias parameter (ρ) and/or a bypass parameter κ.

[0050] In an embodiment, the smoothing parameter (λ) and/or a bias parameter (ρ) depend on the a posteriori SNR γ, or on the spectral density of the electric input signal $|Y|^2$ and the noise spectral density $<\sigma^2>$. In an embodiment, the smoothing parameter (λ) and/or a bias parameter (ρ) and/or the parameter κ are selected depending on a user's hearing loss, cognitive skills or speech intelligibility score. In an embodiment, the smoothing parameter (λ) and/or a bias parameter (ρ) and/or the parameter κ are selected to provide more smoothing the poorer the hearing ability, cognitive skill or speech intelligibility skills are for the user in question.

[0051] In an embodiment, the method comprises adjusting the smoothing parameter (λ) in order to take a filter bank oversampling into account.

[0052] In an embodiment, the method comprises providing that the smoothing and/or the bias parameters depend on whether the input is increasing or decreasing.

[0053] In an embodiment, the method comprises providing that the smoothing parameter (λ) and/or a bias parameter (ρ) and/or the parameter κ are selectable from a user interface. In an embodiment, the user interface is implemented as an APP of a smartphone.

[0054] In an embodiment, the method comprises providing pre-smoothing of the maximum likelihood SNR estimator $\zeta^{ML}(k,n)$ of the a priori target signal to noise ratio estimate ζ(k,n) for the n^{th} time frame maximum likelihood by a selected minimum value
$\xi_{\min}^{ML}$.
This is used to cope with case
$\frac{|Y_n|^2}{\hat{\sigma}_n^2} < 1.$

[0055] In an embodiment, the recursive algorithm is configured to allow the maximum likelihood SNR estimate to bypass the a priori estimate of the previous frame in the calculation of the bias and smoothing parameters. In an embodiment, the recursive algorithm is configured to allow the current maximum likelihood SNR estimate s_n^{ML} to bypass the a priori estimate s_{n-1} of the previous frame, if the current maximum-likelihood SNR estimate s_n^{ML} minus a parameter κ is larger than the previous a priori SNR estimate s_{n-1} (cf. FIG. 4). In an embodiment, the value that is fed to the mapping unit MAP in FIG. 4 is $s_n^{ML} - \kappa$, as shown in FIG. 4, but in another embodiment, s_n^{ML} is directly fed to the mapping unit MAP (when the condition ($s_n^{ML} - \kappa > s_{n-1}$) is fulfilled). In an embodiment, the recursive algorithm comprises a maximum operator located in the recursive loop, allowing the maximum likelihood SNR estimate to bypass the a priori estimate of the previous frame in the calculation of the bias and smoothing parameters via a parameter κ. Thereby (large) SNR onsets can be immediately detected (and thus the risk over-attenuation of speech onsets can be reduced).

[0056] The a posteriori signal to noise ratio estimate $\gamma(k,n)$ of said electric input signal $Y(k,n)$ is provided as a combined a posteriori signal to noise ratio generated as a mixture of a first and an second a posteriori signal to noise ratio. Other combinations (than the a posteriori estimates) can be used (e.g. the noise variance estimate $<\sigma^2>$).

[0057] In an embodiment, the two a posteriori signal to noise ratios are generated from a single microphone configuration and from a multi-microphone configuration, respectively. In an embodiment, the first a posteriori signal to noise ratio is generated faster than the second a posteriori signal to noise ratio. In an embodiment, the combined a posteriori signal to noise ratio is generated as a weighted mixture of the first and the second a posteriori signal to noise ratios.

[0058] The first and second a posteriori signal to noise ratios that are combined to the a posteriori signal to noise ratio of an ipsi-lateral hearing aid originate from the ipsi-lateral and a contra-lateral hearing aid, respectively, of a binaural hearing aid system.

A data processing system:

[0059] In an aspect, a data processing system comprising a processor and program code means for causing the processor to perform at least some (such as a majority or all) of the steps of the method described above, in the 'detailed description of embodiments' and in the claims is furthermore provided by the present application.

A hearing system:

[0060] In a further aspect, a hearing system comprising an audio processing device as described above, in the 'detailed description of embodiments', and in the claims, AND an auxiliary device is moreover provided.

[0061] In an embodiment, the system is adapted to establish a communication link between the audio processing device and the auxiliary device to provide that information (e.g. control and status signals, possibly audio signals) can be exchanged or forwarded from one to the other.

[0062] The audio processing device is or comprises a hearing device, e.g. a hearing aid. In an embodiment, the audio processing device is or comprises a telephone.

[0063] In an embodiment, the auxiliary device is or comprises an audio gateway device adapted for receiving a multitude of audio signals (e.g. from an entertainment device, e.g. a TV or a music player, a telephone apparatus, e.g. a mobile telephone or a computer, e.g. a PC) and adapted for selecting and/or combining an appropriate one of the received audio signals (or combination of signals) for transmission to the audio processing device. In an embodiment, the auxiliary device is or comprises a remote control for controlling functionality and operation of the audio processing device or hearing device(s). In an embodiment, the function of a remote control is implemented in a SmartPhone, the SmartPhone possibly running an APP allowing to control the functionality of the audio processing device via the SmartPhone (the audio processing device(s) comprising an appropriate wireless interface to the SmartPhone, e.g. based on Bluetooth or some other standardized or proprietary scheme).

[0064] In an embodiment, the auxiliary device is another audio processing device, e.g. a hearing device, such as a hearing aid. In an embodiment, the hearing system comprises two hearing devices adapted to implement a binaural hearing system, e.g. a binaural hearing aid system.

Definitions:

[0065] In the present context, a 'hearing device' refers to a device, such as e.g. a hearing instrument or an active ear-protection device or other audio processing device, which is adapted to improve, augment and/or protect the hearing capability of a user by receiving acoustic signals from the user's surroundings, generating corresponding audio signals, possibly modifying the audio signals and providing the possibly modified audio signals as audible signals to at least one of the user's ears. A 'hearing device' further refers to a device such as an earphone or a headset adapted to receive audio signals electronically, possibly modifying the audio signals and providing the possibly modified audio signals as audible signals to at least one of the user's ears. Such audible signals may e.g. be provided in the form of acoustic signals radiated into the user's outer ears, acoustic signals transferred as

mechanical vibrations to the user's inner ears through the bone structure of the user's head and/or through parts of the middle ear as well as electric signals transferred directly or indirectly to the cochlear nerve of the user.

[0066] The hearing device may be configured to be worn in any known way, e.g. as a unit arranged behind the ear with a tube leading radiated acoustic signals into the ear canal or with a loudspeaker arranged close to or in the ear canal, as a unit entirely or partly arranged in the pinna and/or in the ear canal, as a unit attached to a fixture implanted into the skull bone, as an entirely or partly implanted unit, etc. The hearing device may comprise a single unit or several units communicating electronically with each other.

[0067] More generally, a hearing device comprises an input transducer for receiving an acoustic signal from a user's surroundings and providing a corresponding input audio signal and/or a receiver for electronically (i.e. wired or wirelessly) receiving an input audio signal, a (typically configurable) signal processing circuit for processing the input audio signal and an output means for providing an audible signal to the user in dependence on the processed audio signal. In some hearing devices, an amplifier may constitute the signal processing circuit. The signal processing circuit typically comprises one or more (integrated or separate) memory elements for executing programs and/or for storing parameters used (or potentially used) in the processing and/or for storing information relevant for the function of the hearing device and/or for storing information (e.g. processed information, e.g. provided by the signal processing circuit), e.g. for use in connection with an interface to a user and/or an interface to a programming device. In some hearing devices, the output means may comprise an output transducer, such as e.g. a loudspeaker for providing an air-borne acoustic signal or a vibrator for providing a structure-borne or liquid-borne acoustic signal. In some hearing devices, the output means may comprise one or more output electrodes for providing electric signals.

[0068] In some hearing devices, the vibrator may be adapted to provide a structure-borne acoustic signal transcutaneously or percutaneously to the skull bone. In some hearing devices, the vibrator may be implanted in the middle ear and/or in the inner ear. In some hearing devices, the vibrator may be adapted to provide a structure-borne acoustic signal to a middle-ear bone and/or to the cochlea. In some hearing devices, the vibrator may be adapted to provide a liquid-borne acoustic signal to the cochlear liquid, e.g. through the oval window. In some hearing devices, the output electrodes may be implanted in the cochlea or on the inside of the skull bone and may be adapted to provide the electric signals to the hair cells of the cochlea, to one or more hearing nerves, to the auditory cortex and/or to other parts of the cerebral cortex.

[0069] A 'hearing system' refers to a system comprising one or two hearing devices, and a 'binaural hearing system' refers to a system comprising two hearing devices and being adapted to cooperatively provide audible signals to both of the user's ears. Hearing systems or binaural hearing systems may further comprise one or more 'auxiliary devices', which communicate with the hearing device(s) and affect and/or benefit from the function of the hearing device(s). Auxiliary devices may be e.g. remote controls, audio gateway devices,

mobile phones (e.g. SmartPhones), public-address systems, car audio systems or music players. Hearing devices, hearing systems or binaural hearing systems may e.g. be used for compensating for a hearing-impaired person's loss of hearing capability, augmenting or protecting a normal-hearing person's hearing capability and/or conveying electronic audio signals to a person.

[0070] Embodiments of the disclosure may e.g. be useful in applications such as hearing aids, headsets, ear phones, active ear protection systems, handsfree telephone systems, mobile telephones, etc.

BRIEF DESCRIPTION OF DRAWINGS

[0071] The aspects of the disclosure may be best understood from the following detailed description taken in conjunction with the accompanying figures. The figures are schematic and simplified for clarity, and they just show details to improve the understanding of the claims, while other details are left out. Throughout, the same reference numerals are used for identical or corresponding parts. The individual features of each aspect may each be combined with any or all features of the other aspects. These and other aspects, features and/or technical effect will be apparent from and elucidated with reference to the illustrations described hereinafter in which:

FIG. 1A illustrates a single-channel noise reduction unit, wherein a single microphone (M) obtains a mixture y(t) of target sound (x) and noise (v), and
FIG. 1B illustrates a multi-channel noise reduction unit, wherein a multitude of microphone(s) (M_1, M_2) obtain a mixture y(t) of target sound (x) and noise (v),

FIG. 2 shows <the mean value of the maximum likelihood estimator
ξ_n^{ML}
in [dB] as a function of the true SNR in [dB], illustrating the bias, which is introduced by the one-way rectification in the maximum likelihood a priori SNR estimate
$\xi_n^{ML} = \max(\xi_{min}^{ML}, \gamma_n - 1)$,

FIG. 3 shows <an input-output relationship ($\Delta_{output}=f(\Delta_{input})$) of the DD*-algorithm by numerical evaluation of Equation (7) for the STSA [1] gain function (with $\alpha = 0.98$),

FIG. 4 shows a diagram of an exemplary implementation of the proposed Directed Bias and Smoothing Algorithm (DBSA, implemented by unit Po2Pr),

FIG. 5 illustrates how ρ and λ may be derived from the parameters given by the decision directed approach,

FIG. 6A shows the slope λ of the function
$10\log_{10}\Psi\left(\xi_{n-1}, \frac{\xi_n^{ML}}{\xi_{n-1}}\right)$

for the STSA gain function, = 0.98, and
FIG. 6B shows the zero crossing ρ of the function

$$10\log_{10}\Psi\left(\xi_{n-1},\frac{\xi_n^{ML}}{\xi_{n-1}}\right)$$

for the STSA gain function, = 0.98,

FIG. 7 shows a comparison of the responses of the DBSA algorithm according to the present disclosure (crosses) and the DD-algorithm (lines) using the fitted functions in FIG. 6A, 6B, where the curves represent a priori SNR values ranging from -30 dB to +30 dB in 5 dB steps,

FIG. 8 illustrates a modification of the DBSA algorithm (shown in FIG. 4) to accommodate filter bank oversampling, where the purpose of inserting an additional D-frame delay in the recursive loop is to mimic the dynamic behavior of a system with less oversampling,

FIG. 9A shows an embodiment of an audio processing device, e.g. a hearing aid, according to the present disclosure, and
FIG. 9B shows an embodiment of a noise reduction system according to the present disclosure, e.g. for use in the exemplary audio processing device of FIG. 9A (for M=2),

FIG. 10 illustrates the generation of a combined a posteriori signal to noise ratio from two a posteriori signal to noise ratios, one being generated from a single microphone channel and the other from a multi-microphone configuration,

FIG. 11 shows an embodiment of a hearing aid according to the present disclosure comprising a BTE-part located behind an ear or a user and an ITE part located in an ear canal of the user.

[0072] The figures are schematic and simplified for clarity, and they just show details which are essential to the understanding of the disclosure, while other details are left out. Throughout, the same reference signs are used for identical or corresponding parts.

[0073] Further scope of applicability of the present disclosure will become apparent from the detailed description given hereinafter. However, it should be understood that the detailed description and specific examples, while indicating preferred embodiments of the disclosure, are given by way of illustration only. Other embodiments may become apparent to those skilled in the art from the following detailed description.

DETAILED DESCRIPTION OF EMBODIMENTS

[0074] The detailed description set forth below in connection with the appended drawings is intended as a description of various configurations. The detailed description includes specific details for the purpose of providing a thorough understanding of various concepts. However, it will be apparent to those skilled in the art that these concepts may be practised without these specific details. Several aspects of the apparatus and methods are described by various

blocks, functional units, modules, components, circuits, steps, processes, algorithms, etc. (collectively referred to as "elements"). Depending upon particular application, design constraints or other reasons, these elements may be implemented using electronic hardware, computer program, or any combination thereof.

[0075] The electronic hardware may include microprocessors, microcontrollers, digital signal processors (DSPs), field programmable gate arrays (FPGAs), programmable logic devices (PLDs), gated logic, discrete hardware circuits, and other suitable hardware configured to perform the various functionality described throughout this disclosure. Computer program shall be construed broadly to mean instructions, instruction sets, code, code segments, program code, programs, subprograms, software modules, applications, software applications, software packages, routines, subroutines, objects, executables, threads of execution, procedures, functions, etc., whether referred to as software, firmware, middleware, microcode, hardware description language, or otherwise.

[0076] The present application relates to the field of hearing devices, e.g. hearing aids.

[0077] Speech enhancement and noise reduction can be obtained by applying a fast-varying gain in the time-frequency domain. The objective of applying the fast-varying gain is to maintain time-frequency tiles dominated by speech unaltered while the time-frequency tiles dominated by noise is suppressed. Hereby, the resulting modulation of the enhanced signal increases, and will typically become similar to the modulation of the original speech signal, leading to a higher speech intelligibility.

[0078] Let us assume that the observed signal $y(t)$ is the sum of target speech signal $x(t)$ and noise $v(t)$, (e.g. picked up by a microphone or a number of microphones) processed in an analysis filter bank (FBA; FBA_1, FBA_2) to yield frequency sub-band signals Y_{kn} (Y(n,k)) corresponding to frequency k (the frequency index k is dropped from here on for simplicity of notation) and time frame n (cf. e.g. FIG. 1A, 1B). For example, Y_n may comprise (or consist of) complex coefficients obtained from a DFT filter bank. Spectral speech enhancement methods rely on estimating the amount of target signal (X) compared to the amount of noise (N) in each time-frequency tile, i.e. the signal-to-noise (SNR) ratio. In spectral noise reduction, SNR is typically described using two different terms: 1) the a posteriori SNR defined as

$$\gamma_n = \frac{|Y_n|^2}{\hat{\sigma}_n^2}$$

[0079] Where

$\hat{\sigma}_n^2$

is an estimate of the noise spectral density (noise spectral power variance) in the n^{th} time frame, and 2) the a priori SNR defined as

$$\xi_n = \frac{\langle |X_n|^2 \rangle}{\hat{\sigma}_n^2}.$$

[0080] Where $|X_n|^2$ is the target signal spectral density. The a posteriori SNR requires an estimate of the noise power spectral density

$\hat{\sigma}_n^2$,

while the a priori SNR requires access to both speech ($X_n|^2$) and noise power

$(\hat{\sigma}_n^2)$

spectral densities. If the a priori SNR is available, we can for each unit in time and frequency find an estimate of the target signal as

$\hat{X}_n = \frac{\xi_n}{\xi_n + 1} Y_n$,

which represents a Wiener gain approach. Other SNR to gain functions may be used, though. The terms 'a posteriori' and 'a priori' signal-to-noise-ratio are e.g. used in [4].

[0081] FIG. 1A shows a single-channel noise reduction unit, wherein a single microphone (M) receives a mixture y(t) of target sound (x) and noise (v), and FIG. 1B illustrates a multi-channel noise reduction unit, wherein a multitude of microphone(s) (M_1, M_2) receive a mixture y(t) of target sound (x) and noise (v).

[0082] In the present disclosure it is assumed that analogue to digital conversion units are applied as appropriate to provide digitized electric input signals from the microphones. Likewise, it is assumed that digital to analogue conversion unit(s) is/are applied to output signals, if appropriate (e.g. to signals that are to be converted to acoustic signals by a loudspeaker).

[0083] The mixture(s) is/are transformed into the frequency domain by respective analysis filter banks (denoted FBA (Analysis) and FBA_1 (Analysis), FBA_2 (Analysis) in FIG. 1A and 1B, respectively) and obtaining the signal Y(n,k) (denoted Y(n,k) and $Y(n,k)_1$, $Y(n,k)_2$ in FIG. 1A and 1B, respectively). In each case, the a posteriori SNR γ (A posteriori SNR, γ_n in FIG. 1A and 1B) is found as the ratio between the power spectral density $|Y_n|^2$ (provided by respective magnitude squared calculation units $|\cdot|^2$) containing the target signal and an estimate of noise power spectral density

$\hat{\sigma}_n^2$

(denoted $<\sigma^2>$ in FIG, 1A, 1B, and provided by respective noise estimation units NT) within the mixture (cf. combination unit '·/·' in FIG. 1A, 1B). In the case of more than one microphone (e.g. FIG. 1B), the noise within the mixture may be reduced by a linear combination of the microphone signals $Y(n,k) = w(k)_1 \cdot Y(n,k)_1 + w(k)_2 \cdot Y(n,k)_2$, and the remaining noise may be better estimated by using another linear combination (N(n,k)) of the microphone signals aiming at cancelling the target signal, $N(n,k) = w(k)_3 \cdot Y(n,k)_1 + w(k)_4 \cdot Y(n,k)_2$, as indicated by output signals from the beam former filtering unit BFU in FIG. 1B.

[0084] The a priori signal to noise ratio (A priori SNR, ζ_n in FIG. 1A, 1B) is determined by conversion unit Po2Pr implementing an algorithm according to the present disclosure, which is further described in the following. The a priori SNR may e.g. be converted to a gain in an

optional SNR to gain conversion unit SNR2G providing a resulting current noise reduction gain G_{NR} (e.g. based on a Wiener gain function), which may be applied to the signal Y(n,k) (input signal in FIG. 1A and spatially filtered signal in FIG. 1B) in combination unit 'X' to provide noise reduced signal $Y_{NR}(n,k)$.

[0085] Given that an estimate of the noise power density

$\hat{\sigma}_n^2$

(denoted <σ^2> in FIG, 1A, 1B) is available, we can find the a posteriori SNR directly (cf. combination (here division) unit '·/·' in FIG. 1A, 1B). As we typically do not have access to the target power spectral density (A_n^2), A_n being an estimate of the unknown target magnitude $|X_n|$, we do not have direct access to the a priori SNR. In order to estimate the a priori SNR, the decision directed (DD) algorithm has been proposed [1]:

$$\xi_n = \alpha \frac{\hat{A}_{n-1}^{\ 2}}{\hat{\sigma}_n^2} + (1-\alpha)\, max(0, \gamma_n - 1), \qquad (1)$$

where $\hat{A}_n$ is the an estimate of the target signal magnitude (in the n^{th} time frame),

$\hat{\sigma}_n^2$

is the noise spectral variance (power spectral density) at the frequency in question, and α is a weighting factor. The above expression is a linear combination of two estimates of the a priori SNR ξ_n : (because $\gamma-1=(|Y|^2/\sigma^2)-1==(|Y|^2-\sigma^2)/\sigma^2)\sim\zeta$) a recursive part

$\frac{\hat{A}_{n-1}^{\ 2}}{\hat{\sigma}_n^2}$

(since $\hat{A}_n$ generally depends on ξ_n) and 2) a non-recursive part max(0, γ_n - 1). The weighting parameter α is typically chosen in the interval 0.94 - 0.99, but obviously α may depend on the frame rate, and possibly other parameters. The noise estimate

$\hat{\sigma}_n^2$

is assumed to be available from a spectral noise estimator, e.g. a noise tracker (cf. e.g. [2] EP2701145A1

[3]), e.g. using a voice activity detector and a level estimator (estimating noise levels when no voice is detected; working in frequency sub-bands). The speech magnitude estimate $\hat{A}_n$ is obtained using a speech estimator, of which several are available. Generally, the speech estimator can be represented by the corresponding gain function G

$$\hat{A}_n = G(\xi_n, \gamma_n)|Y_n|. \qquad (2)$$

[0086] The gain function can be chosen depending on a cost function or objective to be minimized, and on the statistical assumptions w.r.t. the speech and noise processes. Well-known examples are the STSA gain function [1], LSA [4], MOSIE [5], Wiener, and spectral subtraction gain functions [5], [7]. While STSA (STSA=minimum-mean square error *Short-Time Spectral Amplitude estimator*), LSA, and MOSIE depend on both the (estimated) a priori SNR ξ_n and the a posteriori SNR

$\gamma_n = \frac{|Y_n|^2}{\hat{\sigma}_n^2}$,

the Wiener and spectral subtraction gain functions are one-dimensional and depend only on ξ_n.

As described in [5], $\hat{A}_n$ can be estimated using the following equation known as the MOSIE estimator:

$$\hat{A}_n = \sqrt{\frac{\xi_n}{\mu+\xi_n}\left[\frac{\Gamma\left(\mu+\frac{\beta}{2}\right)}{\Gamma(\mu)}\frac{\Phi\left(1-\mu-\frac{\beta}{2},1;-v_n\right)}{\Phi(1-\mu,1;-v_n)}\right]^{1/\beta}}\sqrt{\hat{\sigma}_n^2}, \qquad (3)$$

where $\Gamma(.)$ is the gamma function, $\Phi(a, b; x)$ is the confluent hypergeometric function and $v_n = \frac{\xi_n}{\mu+\xi_n}\gamma_n$.

Combining (2) and (3), we can write

$$G(\xi_n,\gamma_n) = \sqrt{\frac{\xi_n}{\mu+\xi_n}\left[\frac{\Gamma\left(\mu+\frac{\beta}{2}\right)}{\Gamma(\mu)}\frac{\Phi\left(1-\mu-\frac{\beta}{2},1;-v_n\right)}{\Phi(1-\mu,1;-v_n)}\right]^{1/\beta}}\frac{1}{\sqrt{\gamma_n}}.$$

The LSA estimator (cf. e.g. [4]) can be well approached having $\beta = 0.001$ and $\mu = 1$ (cf. e.g. [5]). The a priori SNR estimated by the decision-directed approach is thus a smoothed version of max(0, γ_n - 1) depending on the smoothing factor α as well as the chosen estimator for obtaining $\hat{A}_n$.

[0087] As mentioned above, α may depend on the frame rate. In an embodiment, the decision directed approach as originally proposed in [1] is designed with frames shifted every 8th millisecond (ms). In hearing instruments, the frames are typically updated with a much higher frame rate (e.g. every single millisecond). This higher oversampling factor of the filter bank allows the system to react much faster (e.g. in order to better maintain speech onsets). This advantage of a possible faster reaction time cannot fully be achieved just by adjusting α according to the higher frame rate. Instead we propose a method, which is better at taking advantage of a higher oversampling factor.

[0088] The DD-algorithm (1) can be reformulated as the recursive function

$$\xi_n = f(\xi_{n-1},\gamma_{n-1},\gamma_n) = \alpha G\left(\xi_{n-1},\frac{|Y_{n-1}|^2}{\hat{\sigma}_{n-1}^2}\right)^2\frac{|Y_{n-1}|^2}{\hat{\sigma}_n^2} + (1-\alpha)\max\left(0,\frac{|Y_n|^2}{\hat{\sigma}_n^2}-1\right) \qquad (4)$$

[0089] As a first simplification, we consider a slightly modified algorithm, which we will refer to as DD*. The recursion in DD* is changed to depend only on the current frame observations and on the previous a priori estimate:

$$\xi_n = F(\xi_{n-1},\gamma_n) = \alpha G(\xi_{n-1},\frac{|Y_n|^2}{\hat{\sigma}_n^2})^2\frac{|Y_n|^2}{\hat{\sigma}_n^2} + (1-\alpha)\max\left(0,\frac{|Y_n|^2}{\hat{\sigma}_n^2}-1\right) \qquad (5)$$

[0090] The effect on the a priori estimates by this modification can be quantified by numerical simulations (see later sections), where the effect is found to be generally small, albeit audible. In fact, using the most recent frame power for the a posteriori SNR in the gain function seems

beneficial for SNR estimation at speech onsets.

[0091] Now, consider the maximum likelihood SNR estimator, which expresses the SNR value with highest likelihood; we make here the standard assumptions that the noise and speech processes are uncorrelated Gaussian processes, and that the spectral coefficients are independent across time and frequency [1]. Then, the maximum likelihood SNR estimator
ξ_n^{ML}
is given by:

$$\xi_n^{ML} = \max\left(\xi_{min}^{ML}, \frac{|Y_n|^2}{\sigma_n^2} - 1\right). \qquad (6)$$

[0092] Note that the maximum likelihood estimator is not a central estimator because its mean differs from the true value. In this case an example of a central estimator is
$\frac{|Y_n|^2}{\sigma_n^2} - 1,$
which can take negative values.

[0093] FIG. 2 shows <the mean value of the maximum likelihood estimator
ξ_n^{ML}
in [dB] as a function of the true SNR in [dB], illustrating the bias, which is introduced by the one-way rectification in the maximum likelihood a priori SNR estimate
$\xi_n^{ML} = \max\left(\xi_{min}^{ML}, \gamma_n - 1\right).$
The target signal is assumed Gaussian. For noise-only input, the estimated SNR equals ξ^{ML} = $e^{-1} \approx$ -4.3 dB (assuming that
$\xi_{\min}^{\mathrm{ML}} = 0,$
cf. also [5]), cf. *Bias* in FIG. 2. One effect of the DD approach is to provide a compensation for this bias.

Input-output Relationship

[0094] In the following a functional approximation of the DD* algorithm in Equation (5) is proposed. For mathematical convenience, we assume in the following that
$\frac{|Y_n|^2}{\hat{\sigma}_n^2} \geq 1,$

[0095] and derive such an approximation. This assumption simplifies the non-recursive part because ξ_n = max(0, γ_n - 1) simplifies to $\xi_n = \gamma_n - 1$ and $\gamma_n = \xi_n + 1$. It can be shown that the impact (on results) of this assumption is indeed minor. Thus, ignoring the cases, where
$\frac{|Y_n|^2}{\hat{\sigma}_n^2} < 1,$
the DD* algorithm in Equation (5) can be described as the following function of
ξ_n^{ML}

$$\xi_n = \Psi\left(\xi_{n-1}, \frac{\xi_n^{ML}}{\xi_{n-1}}\right)\xi_{n-1}. \qquad (7)$$

[0096] As

$$\frac{\xi_n}{\xi_{n-1}} = \Psi\left(\xi_{n-1}, \frac{\xi_n^{ML}}{\xi_{n-1}}\right),$$

the function Ψ maps out the relative change in the a priori estimate as function of the ratio between the current

ξ_n^{ML}

and the previous a priori SNR estimate ξ_{n-1}. We thus have

$$\Psi\left(\xi_{n-1}, \frac{\xi_n^{ML}}{\xi_{n-1}}\right) = \alpha G(\xi_{n-1}, \xi_n^{ML}+1)^2(\xi_n^{ML}+1) + (1-\alpha)\max(0, \xi_n^{ML}). \qquad (8)$$

[0097] By representing the SNR ratios on a logarithmic (dB) scale, the above relationship expresses the non-linear input-output relationship represented by the DD*-algorithm.

[0098] FIG. 3 shows <an input-output relationship (Δoutput=f(Δinput)) of the DD*-algorithm by numerical evaluation of Equation (7) for the STSA [1] gain function (with α = 0.98). At low a priori SNR estimates (e.g., curve labeled -30 dB), smoothing is in effect since small changes in output for moderate input changes. Furthermore, a bias is introduced - seen by the non-zero abscissa zero crossings - resulting in the average estimated a priori SNR being lower than the average maximum likelihood SNR estimate. Although the term "bias" is often used to reflect a difference between an expected value E() and a "true" reference value, the term is here used to reflect the difference between expected values

$E(\xi_n^{\mathrm{ml}})$

and $E(\xi_n)$. FIG. 3 gives a graphical relationship allowing the determination of a difference between (or ratio of) the current a priori SNR estimate ζ_n and the previous a priori SNR estimate ζ_{n-1} (*output*) from knowledge of the difference between (or ratio of) the current maximum-likelihood estimate

ξ_n^{ML}

and the previous a priori SNR estimate ζ_{n-1} (and the absolute value of the previous a priori SNR estimate ζ_{n-1}) (*input*).

[0099] FIG. 3 shows this relationship, revealing two noticeable effects: For low a priori SNR values (e.g. the curve labelled ζ_{n-1}=-30 dB), the output changes are smaller than the input changes, effectively implementing a low-pass filtering/smoothing of the maximum likelihood SNR estimates

ξ_n^{ML}.

For high a priori SNR values (ζ_{n-1}=+30 dB), the DD* a priori SNR estimate ξ_n varies as much as

the change in
ξ_n^{ML},
resulting in a very small amount of smoothing. Secondly, the zero crossings of the curves for low a priori SNR values are shifted to positive dB values of
$\frac{\xi_n^{ML}}{\xi_{n-1}}$,
up to about 10 dB. This means that for low SNR regions, the a priori SNR estimate ξ_n should settle at values approximately 10 dB below the average value of ξ^{ML}.

[0100] FIG. 3 gives a graphical relationship allowing the determination of a difference between (or ratio of) the current a priori SNR estimate ζ_n and the previous a priori SNR estimate ζ_{n-1} (*output*) from knowledge of the difference between (or ratio of) the current maximum-likelihood estimate
ξ_n^{ML}
and the previous a priori SNR estimate ζ_{n-1} (and the absolute value of the previous a priori SNR estimate ζ_{n-1}) (*input*).

[0101] Values of a smoothing parameter (λ_{DD}) and a bias parameter (ρ) to be discussed below can be read from the graphs as indicated in FIG. 3 for graphs relating to a priory SNR ζ_{n-1}=-30 dB, ζ_{n-1}= 0 dB, and ζ_{n-1}=+30 dB. The bias parameter ρ is found as the zero-crossing of the graph with the horizontal axis. The smoothing parameter λ_{DD} is found as the slope indicated as $\alpha(\cdot)$ of the graph in question at the zero crossing. These values are e.g. extracted and stored in a table for relevant values of the a priori SNR, cf. e.g. mapping unit MAP in FIG. 4.

[0102] FIG. 4 shows a diagram of an exemplary implementation of the proposed Directed Bias and Smoothing Algorithm (DBSA) implemented in the conversion unit Po2Pr.

The Directed Bias and Smoothing Algorithm (DBSA)

[0103] FIG. 4 shows a diagram of the proposed Directed Bias and Smoothing Algorithm (DBSA, implemented by unit Po2Pr), the aim of which is to provide a configurable alternative implementation of the DD approach, encompassing three main effects of DD

1. 1. An SNR-dependent smoothing, allowing for more smoothing in low SNR conditions reducing musical noise.
2. 2. A negative bias compared to
   ξ_n^{ML}
   for low SNR conditions, reducing audibility of musical noise in noise-only periods.
3. 3. A recursive bias, allowing fast switching from low-to-high and high-to-low SNR conditions.

[0104] The DBSA algorithm operates with SNR estimates in the dB domain; thus, introduce
$s_n^{ml} = 10\log_{10}(\xi_n^{ml})$,
and
$s_n = 10\log_{10}(\xi_n)$.
The central part of the embodiment of the proposed algorithm is a 1st order IIR low pass filter with unit DC-gain, and an adaptive time constant. The two functions $\lambda(s_n)$ and $\rho(s_n)$ control the amount of smoothing and the amount of SNR bias, as a recursive function of the estimated SNR.

[0105] In the following we will derive the controlling functions so as to mimic the input-output relationship of the DD system described above. Let s_n and
s_n^{ML}
be the a priori and maximum likelihood SNR expressed in dB, and ignoring the max-operation (let $\kappa \to \infty$ for now) the DBSA input-output relationship is defined by

$$s_n - s_{n-1} = (s_n^{ML} + \rho(s_{n-1}) - s_{n-1})\lambda(s_{n-1}) \qquad (9)$$

Thus, equating the DBSA to the DD* approach is equivalent to the approximation

$$s_n - s_{n-1} = (s_n^{ML} + \rho(s_{n-1}) - s_{n-1})\lambda(s_{n-1}) \approx 10\log_{10}\Psi\left(\xi_{n-1}, \frac{\xi_n^{ML}}{\xi_{n-1}}\right) \qquad (10)$$

In order to fully specify the DBSA in (10), the bias function $\rho(s_n)$ and the smoothing function $\lambda(s_n)$ must be specified. Since our goal is to mimic the behavior of the DD* approach, we could e.g. measure the zero-crossing location and the slope at this location of the function
$10\log_{10}\Psi\left(\xi_{n-1}, \frac{\xi_n^{ML}}{\xi_{n-1}}\right)$,
(evaluated as a function of
ξ_n^{ML}
), and choose the functions $\rho(s_n)$ and $\lambda(s_n)$ to have the same values. Thus, for the bias function $\rho(s_n)$ we choose it to be equal to the value of
$10\log_{10}\frac{\xi}{\xi_{n-1}}$
with ξ satisfying
$10\log_{10}\Psi\left(\xi_{n-1}, \frac{\xi}{\xi_{n-1}}\right) = 0\text{dB}$.
Likewise, the smoothing function $\lambda(s_{n-1})$ can be set to equal the slope (w.r.t.
s_n^{ML}
) of the curves in FIG. 3 at the location of its 0 dB crossing (i.e. when
$s_n^{ML} - s_n = -\rho(s_{n-1})$
).

[0106] FIG. 4 shows an implementation of the Directed Bias and Smoothing Algorithm (DBSA) alternative to the DD-approach. The dashed box in the upper right part of FIG. 4 represents a 1st order IIR low-pass filter with unit DC-gain and variable smoothing coefficient λ(λ_{n-1} in FIG. 4). This part together with combination unit '+' (providing signal
s_n^{ml}-ρ_{n-1}

) and mapping unit MAP (providing smoothing and bias parameters λ, ρ, respectively) providing the inputs to the 1st order IIR low-pass filter implements equation 10 below (cf. indication 'From Eq. (10)' in FIG. 4). The two mapping functions $\lambda(s)$ and $\rho(s)$ (cf. mapping unit MAP) control the amount of smoothing (λ) and bias (ρ), respectively, as a recursive function of the estimated a priori SNR (s_{n-1} (ζ_{n-1}) in FIG. 4). The left part of FIG. 4 providing the maximum likelihood value

ζ_n^{ml}

of the a priori signal to noise ratio of the n^{th} time frame implements equation (6) above (cf. indication 'From Eq. (6)' in FIG. 4). The maximum likelihood value

ζ_n^{ml}

of the a priori signal to noise ratio is converted to the logarithmic domain by the 'dB' unit. The mapping unit MAP is e.g. implemented as a memory comprising a lookup table with values of smoothing and bias parameters λ and ρ extracted from FIG. 3 (or equivalent data material) (cf. indication 'From FIG. 3' in FIG. 4) for relevant values of the a priori SNR ζ (e.g. for a larger range of ζ and/or for a larger number of values, e.g. one curve for every 5 dB, or one for every dB). An implementation of an algorithm for (off-line) calculation of the relevant smoothing and bias parameters λ and ρ for storage in an memory of the mapping unit MAP is illustrated in FIG. 5. The embodiment of FIG. 4 additionally comprises the bypass branch for larger values of the current maximum-likelihood value

$s_n^{ml}(\zeta_n^{ml})$

of the a priori SNR, implemented by unit BPS. The bypass unit BPS comprises combination unit '+' and maximum operator unit 'max'. The combination unit '+' takes parameter κ as input. The value of κ is subtracted from the current maximum-likelihood value

s_n^{ml}

and the resulting value

s_n^{ml} - κ

is fed to the maximum unit max together with the previous value s_{n-1} of the a priori SNR. Thereby relatively large values (larger than s_{n-1} + κ) of the current maximum-likelihood value

$s_n^{ml}(\zeta_n^{ml})$

of the a priori SNR are allowed to have immediate impact on the input to the mapping unit MAP.

[0107] FIG. 5 shows how bias parameter ρ and smoothing parameter λ may be derived from the parameters of the decision directed approach (cf. equation 5). FIG. 5 shows an embodiment of an algorithm for generating relevant data to the mapping unit MAP in FIG. 4. The algorithm determines bias parameter ρ and smoothing parameter λ from the current maximum-likelihood value

$s_n^{ml}(\zeta_n^{ml})$

of the a priori SNR, and the previous a priori SNR value s_{n-1}. Contrary to having a single mapping of ρ and λ, we may choose to have different sets of ρ and λ depending on whether the input is increasing or decreasing. That corresponds to having different attack and release values for ρ and λ. Such sets of parameters could be derived from different values of α

corresponding to different attack and release times (and subsequently stored in the mapping unit MAP). As mentioned later, a compensation of the smoothing parameter to take account of a frame rate (or frame length) different from the one used in the LSA approach [4] is preferably implemented (so that the values of the smoothing parameter λ stored in the mapping unit are directly applicable). This is further discussed below, e.g. in relation to FIG. 8.

[0108] FIG. 6A shows the slope λ and FIG. 6B shows the zero crossing ρ of the function

$$10\log_{10}\Psi\left(\xi_{n-1},\frac{\xi_n^{ML}}{\xi_{n-1}}\right)$$

for the STSA gain function [1], using α = 0.98 in both cases. FIG. 7 shows a comparison of the responses of the DBSA algorithm according to the present disclosure (crosses) and the DD-algorithm (lines) using the fitted functions in FIG. 6A and 6B, where the curves represent a priori SNR values ranging from -30 dB to +30 dB in 5 dB steps.

[0109] FIG. 6 shows the results from numerical evaluation, and FIG. 7 shows a comparison between the input-output responses of the DD*-algorithm and the DBSA algorithm. The difference is seen to be quite small in most cases, as shown in the simulations in a later section.

The Case of Low Observed SNR

[0110] Now, consider the case

$$\frac{|Y_n|^2}{\hat{\sigma}_n^2} < 1.$$

In DBSA, this case is caught by the minimum value

$$\xi_{\min}^{ML},$$

which limits the influence. Recalling Equation (2),

$$\hat{A}_n = G(\xi_n, \frac{|Y_n|^2}{\hat{\sigma}_n^2})\,|Y_n|,$$

we note that the class of gain functions that can be expressed as a power of the Wiener gain function generally have that $\hat{A}_n \to 0$ when

$$\frac{|Y_n|^2}{\hat{\sigma}_n^2} \to 0.$$

This property makes the DD-algorithm bias quite large and negative, which can be mimicked in DBSA with a relatively low value of

$$\xi_{\min}^{ML}.$$

[0111] On the other hand for the STSA, LSA and MOSIE gain functions, a gain larger than 0dB occurs when

$$\frac{|Y_n|^2}{\hat{\sigma}_n^2} \to 0$$

resulting in a non-zero $\hat{A}_n$ in the limit. This effect can to some extent by handled by a larger

$$\xi_{\min}^{ML}.$$

In practice the remaining difference between the DD* approach and DBSA can be made to be negligible.

Numerical Issues

[0112] It should be noted that in some cases (typically for low a priori SNR values) the function $10\log_{10}\Psi\left(\xi_{n-1},\frac{\xi_n^{ML}}{\xi_{n-1}}\right)$ does not have a zero crossing. This reflects a limitation in the range of actual a priori SNR values that the system can produce. One particular example occurs when the gain function $G(\xi_n,\frac{|Y_n|^2}{\hat{\sigma}_n^2})$ is limited by some minimum gain value $G_{\min}$. Inserting this minimum value into Equation (5) it can easily be shown that $\Psi\left(\xi_{n-1},\frac{\xi_n^{ML}}{\xi_{n-1}}\right)\geq\frac{\alpha G_{\min}}{\xi_{n-1}}$. So when ξ_{n-1} is sufficiently low, the function Ψ will be greater than 1, which again means no zero crossing for the function $10\log_{10}\Psi$. A numerical implementation will need to detect this situation and specify some reasonable lookup table values for $\rho(s_n)$ and $\lambda(s_n)$ all the same. The exact values used will not matter in reality since they most likely will only be sampled during convergence from an initial state.

The Maximum Operator

[0113] In FIG. 4, a maximum operator is located in the recursive loop, allowing the maximum likelihood SNR estimate to bypass the a priori estimate of the previous frame in the calculation of the bias and smoothing parameters (via parameter κ). The reason for this element is to aid the detection of SNR onsets and thus reducing the risk over-attenuation of speech onsets. In the DD approach Equation (1), the term (1 - α) allows for large onsets in the current frame to reduce the negative bias quickly, the maximum mimics this behavior as controlled by the parameter κ. We thus have the ability to bypass the smoothing using the factor κ. By increasing κ we may better maintain the speech onsets. On the other hand, an increased κ may also raise the noise floor. An increased noised noise floor will however only have influence when we apply a high amount of attenuation. Thus the selected value of κ depends on the chosen maximum attenuation.

Filter Bank Oversampling

[0114] The filter bank parameters have a large influence on the result of the DD approach. Oversampling is the major parameter to consider, since it has a direct effect on the effect of the smoothing and amount of bias introduced into the a priori SNR estimate.

[0115] How to correct for filter bank oversampling in the DD approach has not been well described in the literature. In the original formulation [1], a 256-point FFT was used with a Hanning window, with 192 samples overlap corresponding to four-fold oversampling, and a sample rate of 8 kHz. In general, two-fold oversampling (50% frame overlap) is usual, see [1] and the references therein. In hearing aids and other low-latency applications, however, oversampling by a factor of 16 or higher is not unrealistic.

[0116] All things equal, oversampling reduces the recursive effects of the DD-approach, as well as of the DBSA method. In the limit of "infinite" oversampling, the recursive bias is replaced with the asymptotic bias function.

[0117] One possible approach for oversampling compensation is to downsample the DD/DBSA estimation by a factor proportional to the oversampling, keeping the priori estimate constant over a number of frames. A drawback of this approach may be that gain jumps are introduced, which may reduce sound quality when used in combination with an oversampled filter bank. With oversampling, the equivalent synthesis filters are shorter and may be insufficient for attenuation of the convolutive noise introduced by the gain jumps.

[0118] With the DBSA method, the temporal behavior (i.e. smoothing of SNR estimates and responsiveness to onsets) is controlled by the combination of the directed recursive smoothing, the directed recursive bias. A more computationally demanding but in theory more precise way of handling filter bank oversampling is by means of a higher order delay element (circular buffer) in the recursive loop, as shown In FIG. 8.

[0119] FIG. 8 illustrates a modification of the DBSA algorithm (shown in FIG. 4) to accommodate filter bank oversampling, where the purpose of inserting an additional D-frame delay in the recursive loop is to mimic the dynamic behavior of a system with less oversampling.

Advantages of the proposed implementation

[0120] The proposed implementation has the following advantages over the decision directed approach:

- We can adjust the smoothing parameter in order to take the filter bank oversampling into account, which is important for implementation in low-latency applications such as hearing instruments.
- Rather than having the smoothing and bias depending on the selected gain function, the smoothing $\lambda(s)$ and bias $\rho(s)$ is directly controlled by the parameterization of the two mapping functions. This enables tuning of each of the mapping functions separately for a desired tradeoff between noise reduction and sound quality. E.g. target energy may be better maintained by over-emphasizing the bias. Also, the parameters can be set in

order to address a certain range of SNR which is of interest. Such sets of parameters may be chosen different for individual users, as some users mainly benefit from noise reduction (in terms of a fluctuating gain) in low-SNR regions and do not need noise reduction as higher signal to noise ratios. On the other hand, other users may require noise reduction at a higher signal to noise ratio region, and a constant attenuation at low signal to noise ratios.

- As an extension to the proposed system, the smoothing and bias parameters may depend on whether the input is increasing or decreasing. I.e. we may use different attack and release values of the two parameters.
- The change of the decision directed approach to only depend on the current frame observations and on the previous a priori estimate seems beneficial for the SNR estimation at speech onsets.
- Likewise, the maximum operator controlled by the parameter κ can be used to reduce the risk of over-attenuating speech onsets. The selected value may depend on the chosen maximum attenuation
- Pre-smoothing of the
  ξ_{n}^{ML}
  by a selected minimum value
  $\xi_{\min}^{ML}$
  is used to cope with case
  $\frac{|Y_n|^2}{\hat{\sigma}_n^2} < 1$
- The noise estimator may rely on multichannel as well as single channel inputs, or on both, and/or on binaural inputs, cf. e.g. FIG. 10. The DBSA parameters may be adjusted differently depending on whether the noise estimator relies on a single channel input or multi-channel inputs.

[0121] FIG. 9A shows an embodiment of an audio processing device APD, e.g. a hearing aid, according to the present disclosure. A time variant input sound s(t) is assumed to comprise a mixture of a target signal component x(t) and a noise signal component v(t) is picked up by the audio processing device processed and provided in a processed for to a user as an audible signal. The audio processing device - here a hearing aid - of FIG. 9A comprises a multitude of input units IU_j, j=1, ..., M, each providing an electric input signal S_i representative of sound s(t) in a time-frequency representation (k,m). In the embodiment of FIG. 9A, each input unit IU_i comprises an input transducer IT_i for converting input sound s_i from the environment (as received at input unit IU_i) to an electric time-domain signal s'_i, i=1, ..., M. The input unit IU_i further comprises an analysis filter bank FBA_i for converting the electric time-domain signal s'_i to a number of frequency sub-band signals (k=1, ..., K), thereby providing the electric inputs signal in a time-frequency representation $S_i(k,m)$. The hearing aid further comprises a multi-input noise reduction system NRS, providing a noise reduced signal Y_{NR} based on the multitude of electric input signals S_i, i=1, ..., M. The multi-input noise reduction system NRS

comprises a multi-input beam former filtering unit BFU, a post filter unit PSTF, and a control unit CONT. The multi-input beam former filtering unit BFU (and the control unit CONT) receives the multitude of electric input signals S_i, i=1, ..., M, and provides signals Y and N. The control unit CONT comprises a memory MEM wherein complex weights W_{ij} are stored. The complex weights W_{ij} define possible pre-defined fixed beam formers of the beam former filtering unit BFU (fed to BFU via signal W_{ij}), cf. e.g. FIG. 9B. The control unit CONT further comprises one or more voice activity detectors VAD for estimating whether or not a given input signal (e.g. a given time-frequency unit of the input signal) comprises (or is dominated by) a voice. Respective control signals V-N1 and VN-2 are fed to the beam former filtering unit BFU and to the post filtering unit PSTF, respectively. The control unit CONT receives the multitude of electric input signals S_i, i=1, ..., M, from input units IU_i and the signal Y from the beam former filtering unit BFU. The signal Y comprises an estimate of the target signal component, and the signal N comprises an estimate of the noise signal component. The (single channel) post filtering unit PSTF receives (spatially filtered) target signal estimate Y and (spatially filtered) noise signal estimate N, and provides a (further) noise reduced target signal estimate Y_{NR} based on knowledge of the noise extracted from the noise signal estimate N. The hearing aid further comprises a signal processing unit SPU for (further) processing the noise reduced signal and providing a processed signal ES. The signal processing unit SPU may be configured to apply a level and frequency dependent shaping of the noise reduced signal Y_{NR}, e.g. to compensate for a user's hearing impairment. The hearing aid further comprises a synthesis filter bank FBS for converting the processed frequency sub-band signal ES to a time domain signal es, which is fed to an output unit OT for providing stimuli es(t) to a user as a signal perceivable as sound. In the embodiment of FIG. 9A, the output unit comprises a loudspeaker for presenting the processed signal es to the user as sound. The forward path from the input unit to the output unit of the hearing aid is here operated in the time-frequency domain (processed in a number of frequency sub-bands FB_k, k=1, ..., K). In another embodiment, the forward path from the input unit to the output unit of the hearing aid may be operated in the time domain. The hearing aid may further comprise a user interface and one or more detectors allowing user inputs and detector inputs to be received by the noise reduction system NRS, e.g. the beam former filtering unit BFU. An adaptive functionality of the beam former filtering unit BFU may be provided.

[0122] FIG. 9B shows a block diagram of an embodiment of a noise reduction system NRS, e.g. for use in the exemplary audio processing device of FIG. 9A (for M=2), e.g. a hearing aid, according to the present disclosure. An exemplary embodiment of the noise reduction system of FIG. 9A is further detailed out in FIG. 9B. FIG. 9B shows an embodiment of an adaptive beam former filtering unit (BFU) according to the present disclosure. The beam former filtering unit comprises first (omni-directional) and second (target cancelling) beam formers (denoted Fixed BF O and Fixed BF C in FIG. 9B and symbolized by corresponding beam patterns). The first and second fixed beam formers provide beam formed signals O and C, respectively, as linear combinations of first and second electric input signals S_1 and S_2, where first and second sets of complex weighting constants ($W_{o1}(k)^*$, $W_{o2}(k)^*$) and ($W_{c1}(k)^*$, $W_{c2}(k)^*$) representative of the respective beam patterns are stored in memory unit (MEM) (cf. memory unit MEM in

control unit CONT of FIG. 9A and signal W_{ij}). * indicates complex conjugation. The beam former filtering unit (BFU) further comprises an adaptive beam former (Adaptive BF, ADBF) providing adaptation constant $\beta_{ada}(k)$ representative of an adaptively determined beam pattern. By combining the fixed and adaptive beam formers of the beam former filtering unit BFU, a resulting (adaptive) estimate of the target signal Y is provided as Y=O- β_{ada}C. The beam former filtering unit (BFU) further comprises voice activity detector VAD1 providing control signal V-N1 (e.g. based on signal O or one of the input signals S_i) indicative of whether or not (or with what probability) the input signal (here O or one of S_i) comprises voice content (e.g. speech) that allows the adaptive beam former to update a noise estimate $<\sigma_c^2>$ (here based on the target cancelling beam former C) during time segments where no (or a low probability of) voice/speech is indicated by the voice activity detector VAD1.

[0123] The resulting (spatially filtered or beam formed) target signal estimate Y from the beam former filtering unit can thus be expressed as

$$Y(k) = O(k) - \beta_{ada}(k) \cdot C(k)$$

$$Y(k) = (W_{o1}^* \cdot S_1 + W_{o2}^* \cdot S_2) - \beta_{ada}(k) \cdot (W_{c1}^* \cdot S_1 + W_{c2}^* \cdot S_2)$$

It may, however, be computationally advantageous just to calculate the actual resulting weights applied to each microphone signal rather than calculating the different beam formers used to achieve the resulting signal.

[0124] The embodiment of a post filtering unit PSTF in FIG. 9B receives input signals Y (spatially filtered target signal estimate) and $<\sigma_c^2>$ (noise power spectrum estimate) and provides output signal Y_{BF} (noise reduced target signal estimate) based thereon. The post filtering unit PSTF comprises noise reduction correction unit N-COR for improving the noise power spectrum estimate $<\sigma_c^2>$ received from the beam former filtering unit and providing an improved noise power spectrum estimate $<\sigma^2>$. The improvement results from the use of voice activity detector VAD2 to indicate the presence of no-voice time-frequency units in the spatially filtered target signal estimate Y (cf. signal V-N2). The post filtering unit PSTF further comprises magnitude square ($|\cdot|^2$) and divide ($\cdot/\cdot$) processing units for providing the target signal power spectrum estimate $|Y|^2$ and a posteriori signal to noise ratio $\gamma=|Y|^2/<\sigma^2>$ respectively. The post filtering unit PSTF further comprises a conversion unit Po2Pr for converting the a posteriori signal to noise ratio estimate γ to an a priori signal to noise ratio estimate ζ implementing an algorithm according to the present disclosure. The post filtering unit PSTF further comprises a conversion unit SNR2G configured to convert the a priori signal to noise ratio estimate ζ to a corresponding gain G_{NR} to be applied to the spatially filtered target signal estimate (here by multiplication unit 'X') to provide the resulting noise reduced target signal estimate Y_{BF}. Frequency and time indices k and n are not shown in FIG. 9B for simplicity. But it is assumed that corresponding time frames are available for the processed signals, e.g. $|Y_n|^2$, $<\sigma_n^2>$, γ_n, ζ_n, $G_{NR,n}$, etc.).

[0125] The multi-input noise reduction system comprising a multi-input beam former filtering unit BFU and a single channel post filtering unit PSTF may e.g. be implemented as discussed in [2] with the modifications proposed in the present disclosure.

[0126] The noise power spectrum $<\sigma^2>$ is in the embodiment of FIG. 9B based on the two microphone beam former (the target cancelling beam former C), but may instead be based on a single-channel noise estimate, e.g. based on an analysis of modulation (e.g. a voice activity detector).

[0127] FIG. 10 illustrates an input stage (e.g. of a hearing aid) comprising microphones M_1 and M_2 electrically connected to respective analysis filter banks FBA_1 and FBA_2 and providing respective mixed electric input frequency sub-band signals $Y(n,k)_1$, $Y(n,k)_2$, as described in connection with FIG. 1B. The electric input signals $Y(n,k)_1$, $Y(n,k)_2$, based on the first and second microphone signals, are fed to a multi-input (here 2) a posteriori signal to noise calculation unit (APSNR-M) for providing multi-input a posteriori SNR $\gamma_{n,m}$ (for the n^{th} time frame), e.g. as discussed in connection with FIG. 1B above. One of the two electric input signals $Y(n,k)_1$, $Y(n,k)_2$, or a third different electric input signal (e.g. beamformed signal or a signal based on a third microphone, e.g. a microphone of a contra-lateral hearing aid or of a separate microphone) is fed to a single-input a posteriori signal to noise calculation unit (APSNR-S) for providing single-input a posteriori SNR $\gamma_{n,s}$ (for the n^{th} time frame), e.g. as discussed in connection with FIG. 1A above. The two a posteriori SNRs $\gamma_{n,m}$ and $\gamma_{n,s}$ are fed to mixing unit MIX for generation of a combined (resulting) a posteriori signal to noise ratio $\gamma_{n,res}$ from the two a posteriori signal to noise ratios. The combination of two independent a posteriori estimates will typically provide a better estimate than each of the estimates alone. As the multichannel estimate $\gamma_{n,m}$ typically is more reliable than the single channel estimate $\gamma_{n,s}$, the multichannel estimate will require less smoothing compared to the single input channel noise estimate. Thus, different sets of the smoothing parameters ρ (bias), λ (smoothing), and κ (bias) (cf. FIG. 3, 4) are required for smoothing of the multi microphone a posteriori SNR estimate $\gamma_{n,m}$ and the single microphone a posteriori SNR estimate $\gamma_{n,s}$. The mixing of the two estimates to provide the resulting a posteriori SNR estimate $\gamma_{n,res}$ could e.g. be provide as a weighted sum of the two estimates $\gamma_{n,m}$, $\gamma_{n,s}$.

[0128] In an embodiment of a binaural hearing aid system, either the a posteriori SNR, the a priori SNR, or the noise estimate or the gain from the hearing instrument on the contralateral side is transmitted to and used in the hearing instrument on the ips-ilateral side.

[0129] Besides the a posteriori estimate from the ipsi-lateral hearing instrument, the a priori estimate may also depend on the a posteriori estimate, the a priori, or the noise estimate (or gain estimate) from the contra-lateral hearing instrument. Again, an improved a priori SNR estimate can be achieved by combining different independent SNR estimates.

[0130] FIG. 11 shows an embodiment of a hearing aid according to the present disclosure comprising a BTE-part located behind an ear or a user and an ITE part located in an ear canal of the user.

[0131] FIG. 11 illustrates an exemplary hearing aid (HD) formed as a receiver in the ear (RITE) type hearing aid comprising a BTE-part (BTE) adapted for being located behind pinna and a part (ITE) comprising an output transducer (e.g. a loudspeaker/receiver, SPK) adapted for being located in an ear canal (Ear canal) of the user (e.g. exemplifying a hearing aid (HD) as shown in FIG. 9A). The BTE-part (BTE) and the ITE-part (ITE) are connected (e.g. electrically connected) by a connecting element (IC). In the embodiment of a hearing aid of FIG. 11, the BTE part (BTE) comprises two input transducers (here microphones) (M_{BTE1}, M_{BTE2}) each for providing an electric input audio signal representative of an input sound signal (S_{BTE}) from the environment (in the scenario of FIG. 11, from sound source S). The hearing aid of FIG. 11 further comprises two wireless receivers (WLR_1, WLR_2) for providing respective directly received auxiliary audio and/or information signals. The hearing aid (HD) further comprises a substrate (SUB) whereon a number of electronic components are mounted, functionally partitioned according to the application in question (analogue, digital, passive components, etc.), but including a configurable signal processing unit (SPU), a beam former filtering unit (BFU), and a memory unit (MEM) coupled to each other and to input and output units via electrical conductors Wx. The mentioned functional units (as well as other components) may be partitioned in circuits and components according to the application in question (e.g. with a view to size, power consumption, analogue vs digital processing, etc.), e.g. integrated in one or more integrated circuits, or as a combination of one or more integrated circuits and one or more separate electronic components (e.g. inductor, capacitor, etc.). The configurable signal processing unit (SPU) provides an enhanced audio signal (cf. signal ES in FIG. 9A), which is intended to be presented to a user. In the embodiment of a hearing aid device in FIG. 11, the ITE part (ITE) comprises an output unit in the form of a loudspeaker (receiver) (SPK) for converting the electric signal (es in FIG. 9A) to an acoustic signal (providing, or contributing to, acoustic signal S_{ED} at the ear drum (Ear drum). In an embodiment, the ITE-part further comprises an input unit comprising an input transducer (e.g. a microphone) (M_{ITE}) for providing an electric input audio signal representative of an input sound signal S_{ITE} from the environment at or in the ear canal. In another embodiment, the hearing aid may comprise only the BTE-microphones (M_{BTE1}, M_{BTE2}). In yet another embodiment, the hearing aid may comprise an input unit (IT_3) located elsewhere than at the ear canal in combination with one or more input units located in the BTE-part and/or the TTE-part. The TTE-part further comprises a guiding element, e.g. a dome, (DO) for guiding and positioning the ITE-part in the ear canal of the user.

[0132] The hearing aid (HD) exemplified in FIG. 11 is a portable device and further comprises a battery (BAT) for energizing electronic components of the BTE- and ITE-parts.

[0133] The hearing aid (*HD*) comprises a directional microphone system (beam former filtering unit (*BFU*)) adapted to enhance a target acoustic source among a multitude of acoustic

sources in the local environment of the user wearing the hearing aid device. In an embodiment, the directional system is adapted to detect (such as adaptively detect) from which direction a particular part of the microphone signal (e.g. a target part and/or a noise part) originates and/or to receive inputs from a user interface (e.g. a remote control or a smartphone) regarding the present target direction. The memory unit (*MEM*) comprises predefined (or adaptively determined) complex, frequency dependent constants defining predefined or fixed (or adaptively determined 'fixed') beam patterns according to the present disclosure, together defining the beamformed signal Y (cf. e.g. FIG. 9A, 9B)

[0134] The hearing aid of FIG. 11 may constitute or form part of a hearing aid and/or a binaural hearing aid system according to the present disclosure.

[0135] The hearing aid (HD) according to the present disclosure may comprise a user interface UI, e.g. as shown in FIG. 11 implemented in an auxiliary device (AUX), e.g. a remote control, e.g. implemented as an APP in a smartphone or other portable (or stationary) electronic device. In the embodiment of FIG. 11, the screen of the user interface (UI) illustrates a *Smooth beamforming APP.* Parameters that govern or influence the current smoothing of signal to noise ratios of a beamforming noise reduction system, here parameters ρ (bias), λ (smoothing), (cf. discussion in connection with FIG. 3, 4) can be controlled via the *Smooth beamforming APP* (with the subtitle: 'Directionality. Configure smoothing parameters'). The bias parameter ρ can be set via a slider to a value between a minimum value (e.g. 0) and a maximum value .e.g. 10 dB. The currently set value (here 5 dB) is shown on the screen at the location of the slider on the (grey shaded) bar that span the configurable range of values. Likewise, the smoothing parameter λ can be set via a slider to a value between a minimum value (e.g. 0) and a maximum value .e.g. 1. The currently set value (here 0.6) is shown on the screen at the location of the slider on the (grey shaded) bar that span the configurable range of values. The arrows at the bottom of the screen allow changes to a preceding and a proceeding screen of the APP, and a tab on the circular dot between the two arrows brings up a menu that allows the selection of other APPs or features of the device. The parameters ρ and λ related to smoothing may not necessarily be visible to the user. The sets of ρ, λ could be derived from a third parameter (e.g. a calm to aggressive noise reduction bar or set via an environment detector).

[0136] The auxiliary device and the hearing aid are adapted to allow communication of data representative of the currently selected smoothing parameters to the hearing aid via a, e.g. wireless, communication link (cf. dashed arrow WL2 in FIG. 11). The communication link WL2 may e.g. be based on far field communication, e.g. Bluetooth or Bluetooth Low Energy (or similar technology), implemented by appropriate antenna and transceiver circuitry in the hearing aid (HD) and the auxiliary device (AUX), indicated by transceiver unit WLR_2 in the hearing aid. The communication link may be configured to provide one-way (e.g. APP to hearing instrument) or two way communication (e.g. audio and/or control or information signals).

[0137] It is intended that the structural features of the devices described above, either in the

detailed description and/or in the claims, may be combined with steps of the method, when appropriately substituted by a corresponding process.

[0138] As used, the singular forms "a," "an," and "the" are intended to include the plural forms as well (i.e. to have the meaning "at least one"), unless expressly stated otherwise. It will be further understood that the terms "includes," "comprises," "including," and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. It will also be understood that when an element is referred to as being "connected" or "coupled" to another element, it can be directly connected or coupled to the other element but an intervening elements may also be present, unless expressly stated otherwise. Furthermore, "connected" or "coupled" as used herein may include wirelessly connected or coupled. As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed items. The steps of any disclosed method is not limited to the exact order stated herein, unless expressly stated otherwise, within the scope defined by the claims.

[0139] It should be appreciated that reference throughout this specification to "one embodiment" or "an embodiment" or "an aspect" or features included as "may" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the disclosure. Furthermore, the particular features, structures or characteristics may be combined as suitable in one or more embodiments of the disclosure. The previous description is provided to enable any person skilled in the art to practice the various aspects described herein. Various modifications to these aspects will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other aspects, within the scope defined by the claims.

[0140] The claims are not intended to be limited to the aspects shown herein, but is to be accorded the full scope consistent with the language of the claims, wherein reference to an element in the singular is not intended to mean "one and only one" unless specifically so stated, but rather "one or more." Unless specifically stated otherwise, the term "some" refers to one or more.

[0141] Accordingly, the scope should be judged in terms of the claims that follow.

REFERENCES

[0142]

1. [1] Ephraim, Y.; Malah, D., "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator", IEEE Transactions on Acoustics, Speech and Signal Processing, vol.32, no.6, pp. 1109- 1121, Dec 1984 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1164453&isnumber=26187

2. [2] EP2701145A1
3. [3] Martin, R., "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics", IEEE Transactions on Speech and Audio Processing, vol.9, no.5, pp. 504-512, Apr 2001
4. [4] Ephraim, Y.; Malah, D., "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator", IEEE Transactions on Acoustics, Speech and Signal Processing, vol.33, no.2, pp. 443- 445, Apr 1985 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1164550&isnumber=26190
5. [5] Breithaupt, C.; Martin, R., "Analysis of the Decision-Directed SNR Estimator for Speech Enhancement With Respect to Low-SNR and Transient Conditions", IEEE Transactions on Audio, Speech, and Language Processing, vol.19, no.2, pp. 277-289, Feb. 2011 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5444986&isnumber=56092 32
6. [6] Cappe, O., "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," Speech and Audio Processing, IEEE Transactions on , vol.2, no.2, pp. 345-349, Apr 1994 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=279283&isnumber=6926
7. [7] Loizou, P. (2007). Speech Enhancement: Theory and Practice, CRC Press, Boca Raton: FL

REFERENCES CITED IN THE DESCRIPTION

Cited references

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

- WO2015189261A1 [0005]
- EP2701145A1 [0085] [0142]

Non-patent literature cited in the description

- **EPHRAIM, Y.MALAH, D.**Speech enhancement using a minimum-mean square error short-time spectral amplitude estimatorIEEE Transactions on Acoustics, Speech and Signal Processing, 1984, vol. 32, 61109-1121 [0142]
- **MARTIN, R.**Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum StatisticsIEEE Transactions on Speech and Audio Processing, 2001, vol. 9, 5504-512 [0142]
- **EPHRAIM, Y.MALAH, D.**Speech enhancement using a minimum mean-square error log-spectral amplitude estimatorIEEE Transactions on Acoustics, Speech and Signal Processing, 1985, vol. 33, 2443-445 [0142]
- **BREITHAUPT, C.MARTIN, R.**Analysis of the Decision-Directed SNR Estimator for Speech Enhancement With Respect to Low-SNR and Transient ConditionsIEEE Transactions on Audio, Speech, and Language Processing, 2011, vol. 19, 2277-289 [0142]
- **CAPPE, O.**Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressorSpeech and Audio Processing, IEEE Transactions on, 1994, vol. 2, 2345-349 [0142]
- **LOIZOU, P.**Speech Enhancement: Theory and PracticeCRC Press20070000 [0142]

Patentkrav

1. Høreapparat, der omfatter

- i det mindste en indgangsenhed til tilvejebringelse af en tidsfrekvensrepræsentation Y(k,n) af et elektrisk indgangssignal, der repræsenterer et tidsvariant lydsignal bestående af måltalesignalkomponenter S(k,n) fra en mållydkilde TS og støjsignalkomponenter N(k,n) fra andre kilder end mållydkilden, hvor k og n er henholdsvis frekvensbånd og tidsrammeindeks,
- et støjreduktionssystem, der er konfigureret til
  - at bestemme et a-posteriori-signal-til-støj-forhold-estimat γ(k,n) af det elektriske indgangssignal, hvor a-posteriori-signal-til-støj-forholdet er et forhold mellem det observerede støjsignal Y, der omfatter målsignal S plus støj N, og støjen N, eller et estimat af støjen N, på et givet tidspunkt, og til
  - at bestemme et a-priori-mål-signal-til-støj-forhold-estimat ζ(k,n) af det elektriske indgangssignal fra a-posteriori-signal-til-støj-forhold-estimatet γ(k,n), der er baseret på en rekursiv algoritme, hvor a-priori-signal-til-støj-forholdet er et forhold for målsignalet S til støjsignalet N på et givet tidspunkt, og til
- at bestemme a-priori-mål-signal-til-støj-forhold-estimatet ζ(k,n) for den n^{te} tidsramme fra
  - a-priori-mål-signal-til-støj-forhold-estimatet ζ(k,n-1) for den $(n-1)^{te}$ tidsramme, og fra
  - a-posteriori-signal-til-støj-forhold-estimatet γ(k,n) for den n^{te} tidsramme,
- hvor støjreduktionssystemet er konfigureret til at bestemme a-priori-mål-signal-til-støj-forhold-estimatet ζ(k,n) for den n^{te} tidsramme fra a-priori-mål-signal-til-støj-forhold-estimatet ζ(k,n-1) for den $(n-1)^{te}$ tidsramme, og fra en maksimal sandsynligheds-SNR-estimator $\zeta^{ML}(k,n)$ af a-priori-mål-signal-til-støj-forhold-estimatet ζ(k,n) for den n^{te} tidsramme i henhold til den følgende rekursive algoritme:

$$s_n - s_{n-1} = (s_n^{ML} + \rho(s_{n-1}) - s_{n-1})\lambda(s_{n-1})$$

hvor $\rho(s_{n-1})$ repræsenterer en biasfunktion eller -parameter, og $\lambda(s_{n-1})$ repræsenterer en udjævningsfunktion eller -parameter for den $(n-1)^{te}$ tidsramme, og hvor s er en logaritmisk repræsentation af ζ, og hvor støjreduktionssystemet er konfigureret til at bestemme den maksimale sandsynligheds-SNR-estimator $\zeta^{ML}(k,n)$ som $MAX\{\zeta^{ML}_{min}(k,n); \gamma(k,n)-1\}$, hvor MAX er den maksimale operatør, og $\zeta^{ML}_{min}(k,n)$ er en minimumsværdi af den maksimale sandsynligheds-SNR-estimator $\zeta^{ML}(k,n)$;

- hvor støjreduktionssystemet yderligere omfatter en konverteringsenhed til tilvejebringelse af en forstærkning af støjreduktion G_{NR} i afhængighed af a-priori-mål-signal-til-støj-forhold-estimatet ζ(k,n) og en kombinationsenhed til at anvende forstærkningen af støjreduktionen G_{NR} til det elektriske indgangssignal eller et signal afledt derfra, og KENDETEGNET VED, AT

a-posteriori-signal-til-støj-forhold-estimatet γ(k,n) af det elektriske indgangssignal Y(k,n) er tilvejebragt som et kombineret a-posteriori-signal-til-

støj-forhold, der er genereret som en blanding af et første og andet a-posteriori-signal-til-støj-forhold fra henholdsvis høreapparatet og fra et kontralateralt høreapparat af et binauralt høreapparatsystem.

2. Høreapparat ifølge krav 1, hvor støjreduktionssystemet er konfigureret til at bestemme a-priori-mål-signal-til-støj-forhold-estimatet $\zeta(k,n)$ for den n^{te} tidsramme under antagelsen af, at $\gamma(k,n)$ er større end eller lig med 1.

3. Høreapparat ifølge krav 1 eller 2, hvor minimumsværdi $\zeta^{ML}_{min}(k,n)$ af den maksimale sandsynligheds-SNR-estimator $\zeta^{ML}(k,n)$ er lig med 1, det vil sige lig med 0 dB på en logaritmisk skala.

4. Høreapparat ifølge et hvilket som helst af kravene 1-3, hvor støjreduktionssystemet er konfigureret til at bestemme a-priori-mål-signal-til-støj-forhold-estimatet ζ ved ikke-lineær udjævning af a-posteriori-signal-til-støj-forhold-estimatet γ, ved at støjreduktionssystemet er konfigureret til at tilvejebringe, at udjævningsfunktionen eller -parameteren er SNR-afhængig, hvilket tillader mere udjævning under lave SNR-forhold end ved høje SNR-forhold.

5. Høreapparat ifølge krav 4, hvor tidskonstanten(erne), der er involveret i udjævning, er højere jo lavere SNR.

6. Høreapparat ifølge et hvilket som helst af kravene 1-5, hvor støjreduktionssystemet er konfigureret til at tilvejebringe en negativ bias til værdien af den maksimale sandsynligheds-SNR-estimator under lave SNR-forhold, hvor bias'en er en forskel mellem den forventede værdi af den maksimale sandsynligheds-SNR-estimator og den forventede værdi af a-priori-signal-til-støj-forholdet.

7. Høreapparat ifølge et hvilket som helst af kravene 1-6, der omfatter en filterbank, der omfatter en analysefilterbank til tilvejebringelse af tidsfrekvensrepræsentationen $Y(k,n)$ af det elektriske indgangssignal.

8. Høreapparat ifølge krav 7, der er konfigureret til at tilvejebringe, at analysefilterbanken er oversamplet.

9. Høreapparat ifølge krav 8, hvor den rekursive sløjfe af algoritmen til bestemmelse af a-priori-mål-signal-til-støj-forhold-estimatet $\zeta(k,n)$ for den n^{te} tidsramme omfatter et forsinkelseselement af højere orden, f.eks. en cirkulær buffer, for at kompensere for oversampling af analysefilterbanken.

10. Høreapparat ifølge et hvilket som helst af kravene 1-9, hvor det første og andet a-posteriori-signal-til-støj-forhold genereres fra henholdsvis en enkelt mikrofonkonfiguration og fra en multimikrofonkonfiguration.

11. Høreapparat ifølge et hvilket som helst af kravene 1-10, hvor det første a-posteriori-signal-til-støj-forhold genereres hurtigere end det andet a-posteriori-

signal-til-støj-forhold.

12. Høreapparat ifølge et hvilket som helst af kravene 1-11, hvor det kombinerede a-posteriori-signal-til-støj-forhold genereres som en vægtet blanding af det første og det andet a-posteriori-signal-til-støj-forhold.

13. Høreapparat ifølge et hvilket som helst af kravene 1-12, hvor et a-priori-SNR-estimat af høreapparatet, der er en del af et binauralt høreapparatsystem, er baseret på a-posteriori-SNR-estimater fra begge høreapparater af det binaurale høreapparatsystem.

14. Fremgangsmåde til estimering af et a-priori-signal-til-støj-forhold $\zeta(k,n)$ af en tidsfrekvensrepræsentation $Y(k,n)$ af et elektrisk indgangssignal, der repræsenterer et tidsvariant lydsignal bestående af måltalekomponenter og støjkomponenter, hvor k og n er henholdsvis frekvensbånd og tidsrammeindeks, hvor fremgangsmåden omfatter

- bestemmelse af et a-posteriori-signal-til-støj-forhold-estimat $\gamma(k,n)$ af det elektriske indgangssignal $Y(k,n)$, hvor a-posteriori-signal-til-støj-forholdet er et forhold mellem det observerede støjsignal Y, der omfatter målsignal S plus støj N, og støjen N, eller et estimat af støjen N, på et givet tidspunkt;
- bestemmelse af et a-priori-mål-signal-til-støj-signal-forhold-estimat $\zeta(k,n)$ af det elektriske indgangssignal fra a-posteriori-signal-til-støj-forhold-estimatet $\gamma(k,n)$, der er baseret på en rekursiv algoritme, hvor a-priori-signal-til-støj-forholdet er et forhold for målsignalet S til støjsignalet N på et givet tidspunkt,
- bestemmelse af a-priori-mål-signal-til-støj-signal-forhold-estimatet $\zeta(k,n)$ for den n^{te} tidsramme fra a-priori-mål-signal-til-støj-signal-forhold-estimatet $\zeta(k,n-1)$ for den $(n-1)^{te}$ tidsramme og a-posteriori-signal-til-støj-forhold-estimatet $\gamma(k,n)$ for den n^{te} tidsramme;
- bestemmelse af a-priori-mål-signal-til-støj-forhold-estimatet $\zeta(k,n)$ for den n^{te} tidsramme fra a-priori-mål-signal-til-støj-forhold-estimatet $\zeta(k,n-1)$ for den $(n-1)^{te}$ tidsramme, og fra en maksimal sandsynligheds-SNR-estimator $\zeta^{ML}(k,n)$ af a-priori-mål-signal-til-støj-forhold-estimatet $\zeta(k,n)$ for den n^{te} tidsramme i henhold til den følgende rekursive algoritme:

$$s_n - s_{n-1} = (s_n^{ML} + \rho(s_{n-1}) - s_{n-1})\lambda(s_{n-1})$$

hvor $\rho(s_{n-1})$ repræsenterer en biasfunktion eller -parameter, og $\lambda(s_{n-1})$ repræsenterer en udjævningsfunktion eller -parameter for den $(n-1)^{te}$ tidsramme, og hvor s er en logaritmisk repræsentation af ζ, og hvor støjreduktionssystemet er konfigureret til at bestemme den maksimale sandsynligheds-SNR-estimator $\zeta^{ML}(k,n)$ som $MAX\{\zeta^{ML}_{min}(k,n); \gamma(k,n)-1\}$, hvor MAX er den maksimale operatør, og $\zeta^{ML}_{min}(k,n)$ er en minimumsværdi af den maksimale sandsynligheds-SNR-estimator $\zeta^{ML}(k,n)$; og
- hvor estimater af størrelser $\hat{A}(k,n)$ af måltalekomponenterne bestemmes fra det elektriske indgangssignal $Y(k,n)$ ganget med en forstærkningsfunktion G, hvor forstærkningsfunktionen G er en funktion af a-posteriori-signal-til-støj-forhold-estimatet $\gamma(k,n)$ og a-priori-mål-

signal-til-støj-signal-forhold-estimatet $\zeta(k,n)$; og KENDETEGNET VED, AT

• a-posteriori-signal-til-støj-forhold-estimatet $\gamma(k,n)$ af det elektriske indgangssignal $Y(k,n)$ er tilvejebragt som et kombineret a-posteriori-signal-til-støj-forhold, der er genereret som en blanding af et første og andet a-posteriori-signal-til-støj-forhold fra henholdsvis et ipsi-lateralt og et kontra-lateralt høreapparat af et binauralt høreapparatsystem.

15. Fremgangsmåde ifølge krav 14, der omfatter tilvejebringelse af en SNR-afhængig udjævning af udjævningsfunktionen eller -parameteren, hvilket muliggør mere udjævning under lave SNR-forhold end ved høje SNR-forhold.

16. Fremgangsmåde ifølge et hvilket som helst af kravene 13 eller 14, hvor udjævningsfunktionen eller -parameteren (λ) og/eller biasfunktionen eller -parameteren (ρ) afhænger af den a-posteriori-SNR γ, eller af den spektrale densitet $|Y|^2$ og støjspektraldensiteten $<\sigma^2>$ af det elektriske indgangssignal.

17. Fremgangsmåde ifølge et hvilket som helst af kravene 14-16, der omfatter tilvejebringelse af, at udjævningen og/eller biasfunktionen eller -parametrene afhænger af, hvorvidt inputtet er stigende eller faldende.

18. Fremgangsmåde ifølge et hvilket som helst af kravene 14-17, der omfatter tilvejebringelse af, at udjævningsfunktionen eller -parameteren (λ) og/eller biasfunktionen eller -parameteren (ρ) kan vælges fra en brugergrænseflade.

19. Anvendelse af et høreapparat ifølge et hvilket som helst af kravene 1-13 i et binauralt høreapparatsystem.

20. Binauralt høreapparatsystem, der omfatter første og andet høreapparater ifølge et hvilket som helst af kravene 1-13.

21. Binauralt høreapparatsystem ifølge krav 20, hvor et a-priori-SNR-estimat for et givet høreapparat, der er en del af et binauralt høreapparatsystem, er baseret på a-posteriori-SNR-estimater fra begge høreapparater af det binaurale høreapparatsystem.

22. Binauralt høreapparatsystem ifølge krav 20, hvor et a-priori-SNR-estimat for et givet høreapparat, der er en del af et binauralt høreapparatsystem, er baseret på et a-posteriori-SNR-estimat for det givne høreapparat og et a-priori-SNR-estimat af det andet høreapparat af det binaurale høreapparatsystem.

23. Databehandlingssystem, der omfatter en processor og programkodemiddel til at få processoren til at udføre fremgangsmåden ifølge et hvilket som helst af kravene 14-18.

DRAWINGS

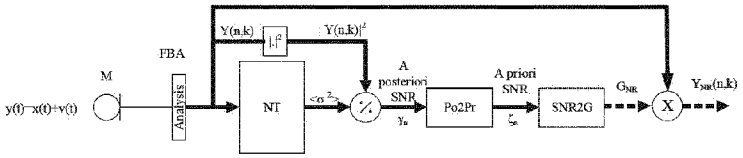


FIG. 1A

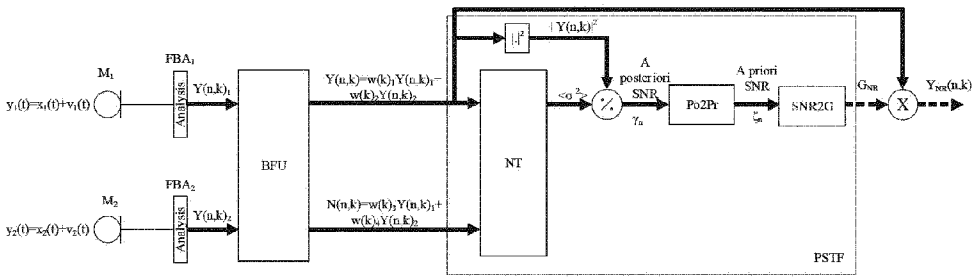


FIG. 1B

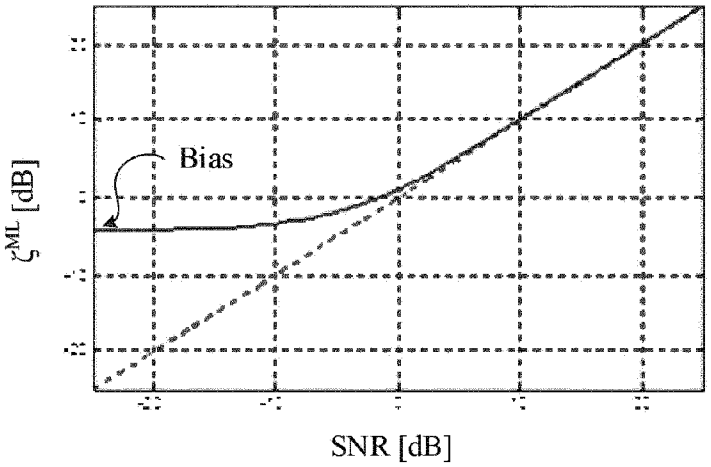


FIG. 2

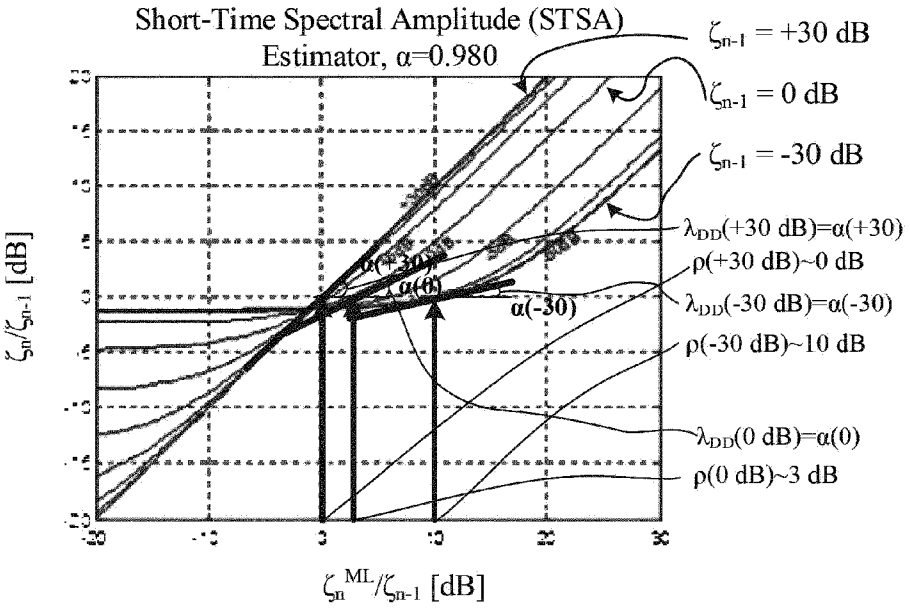

Short-Time Spectral Amplitude (STSA)
Estimator, α=0.980
ζn-1 = +30 dB
ζn-1 = 0 dB
ζn-1 = -30 dB
λDD(+30 dB)=α(+30)
ρ(+30 dB)~0 dB
λDD(-30 dB)=α(-30)
ρ(-30 dB)~10 dB
λDD(0 dB)=α(0)
ρ(0 dB)~3 dB
α(+30)
α(0)
α(-30)
ζn/ζn-1 [dB]
ζn ML/ζn-1 [dB]

FIG. 3

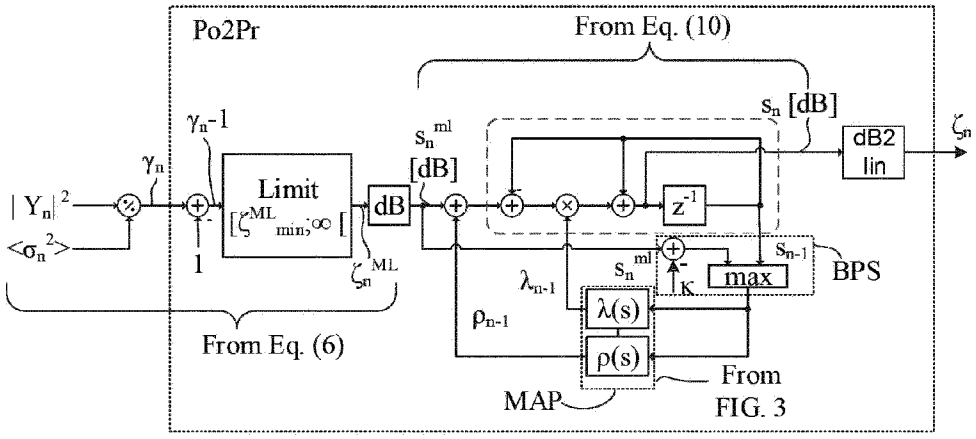

Po2Pr
From Eq. (10)
sn [dB]
dB2 lin
ζn
γn-1
γn
sn ml [dB]
| Yn|2
<σn2>
Limit
[ζML min;∞ [
dB
1
ζn ML
z-1
sn-1
BPS
λn-1
sn ml
K
max
ρn-1
λ(s)
ρ(s)
From Eq. (6)
MAP
From FIG. 3

FIG. 4

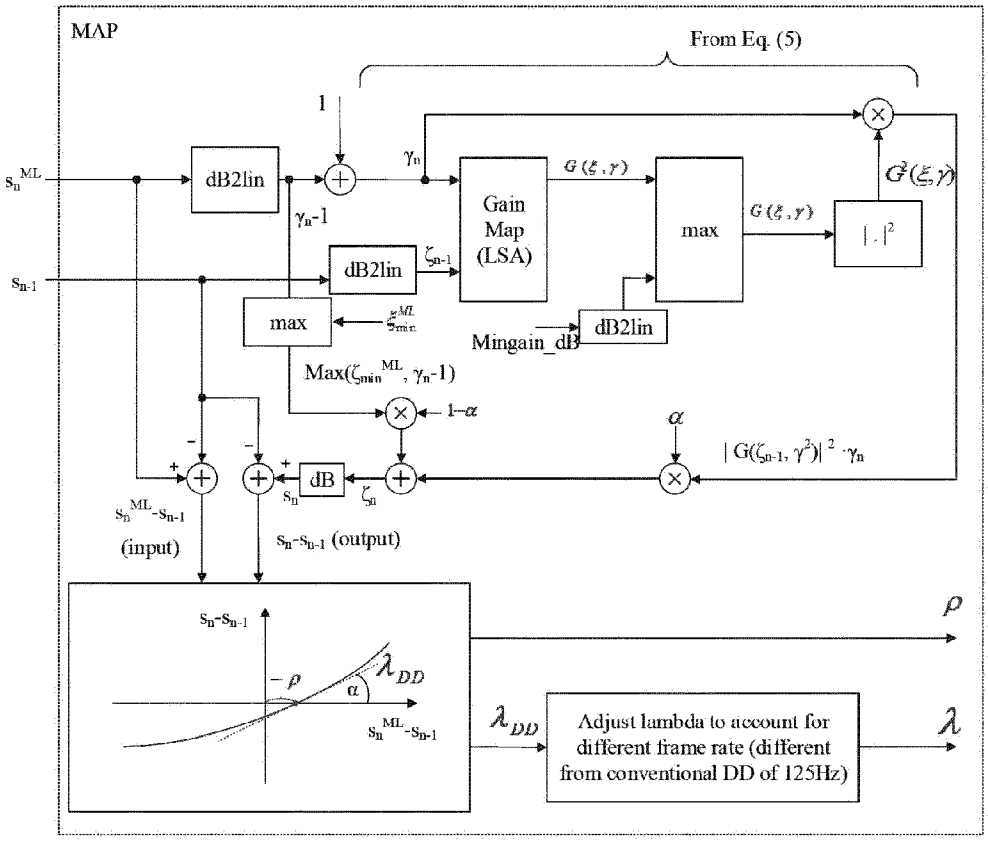

MAP
From Eq. (5)
1
dB2lin
Gain Map (LSA)
max
dB2lin
dB2lin
Mingain_dB
Max(ζmin ML, γn-1)
1−α
α
dB
(input)
sn-sn-1 (output)
sn-sn-1
−ρ
α
λDD
ρ
λDD
Adjust lambda to account for different frame rate (different from conventional DD of 125Hz)
λ

FIG. 5

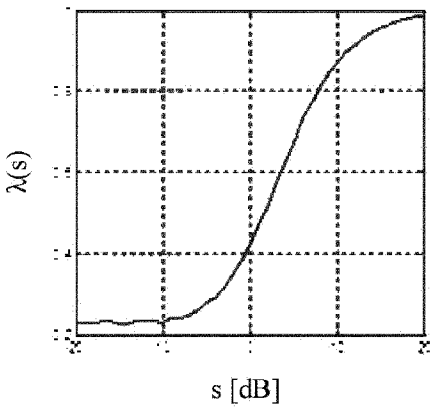


FIG. 6A

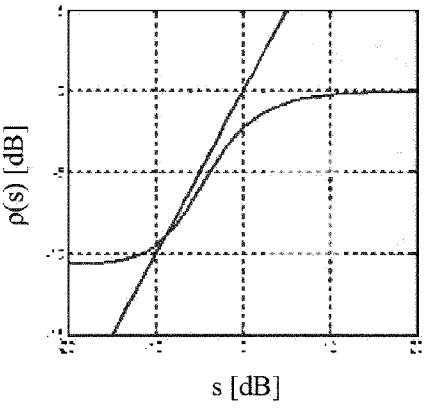


FIG. 6B

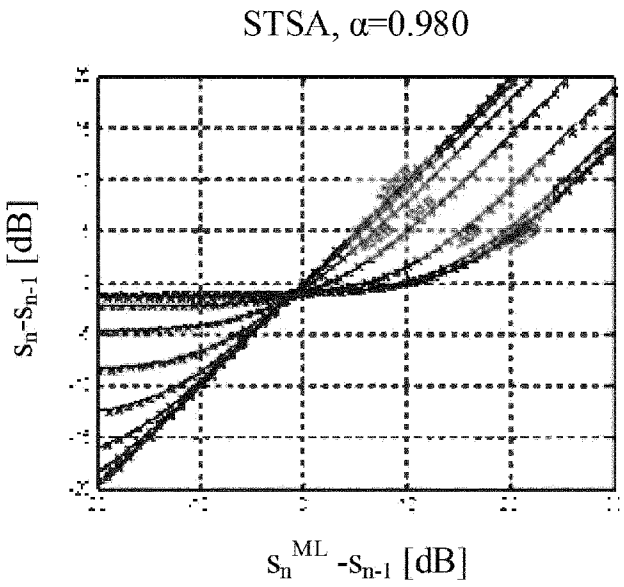


FIG. 7

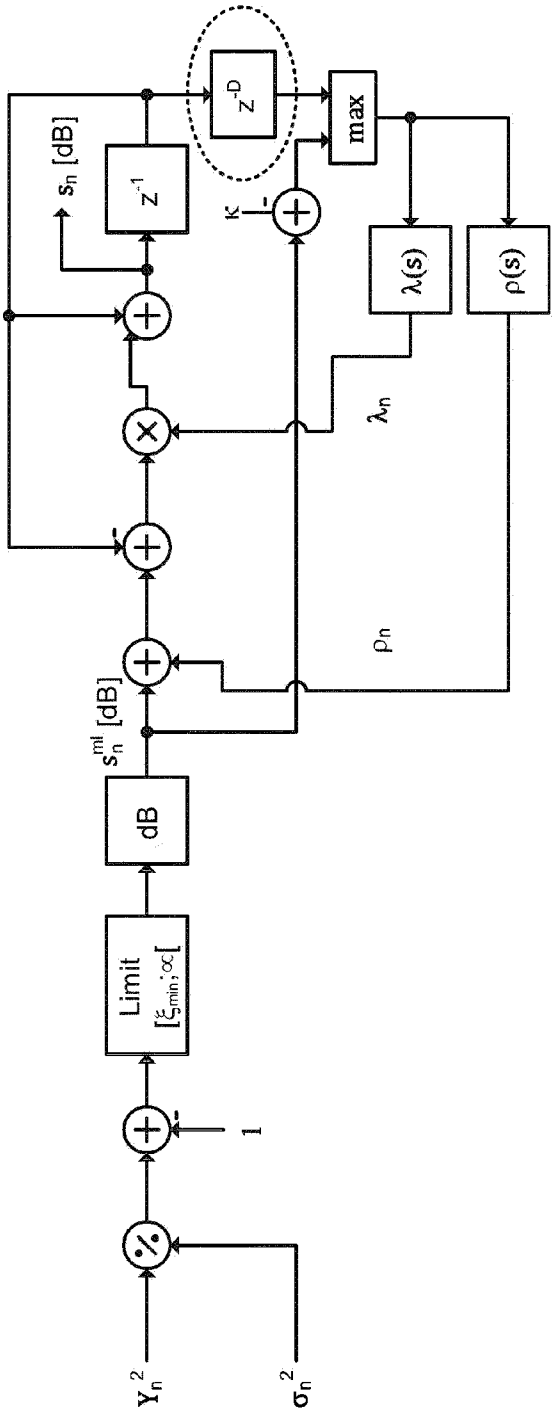

Y_n^2
σ_n^2
1
Limit
$[\xi_{min};\infty[$
dB
s_n^{ml} [dB]
ρ_n
λ_n
s_n [dB]
z^{-1}
z^{-D}
max
$\lambda(s)$
$\rho(s)$

FIG. 8

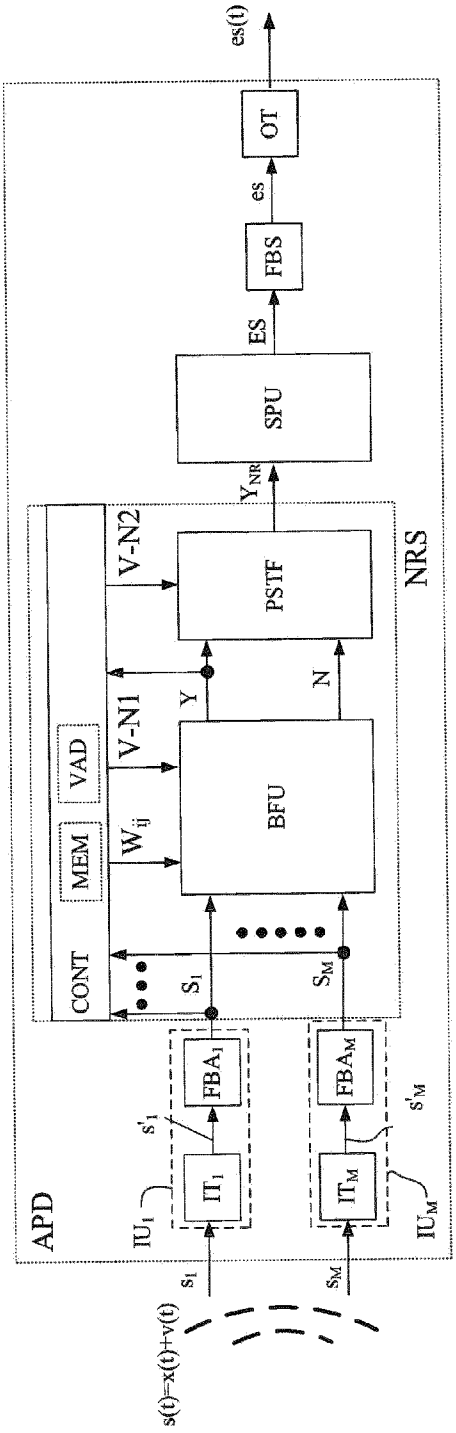

es(t)
OT
es
FBS
ES
SPU
Y_{NR}
PSTF
V-N2
NRS
Y
N
V-N1
VAD
MEM
W_{ij}
BFU
CONT
S_1
S_M
FBA_1
FBA_M
s'_1
s'_M
IT_1
IT_M
IU_1
IU_M
APD
s_1
s_M
s(t)=x(t)+v(t)

FIG. 9A

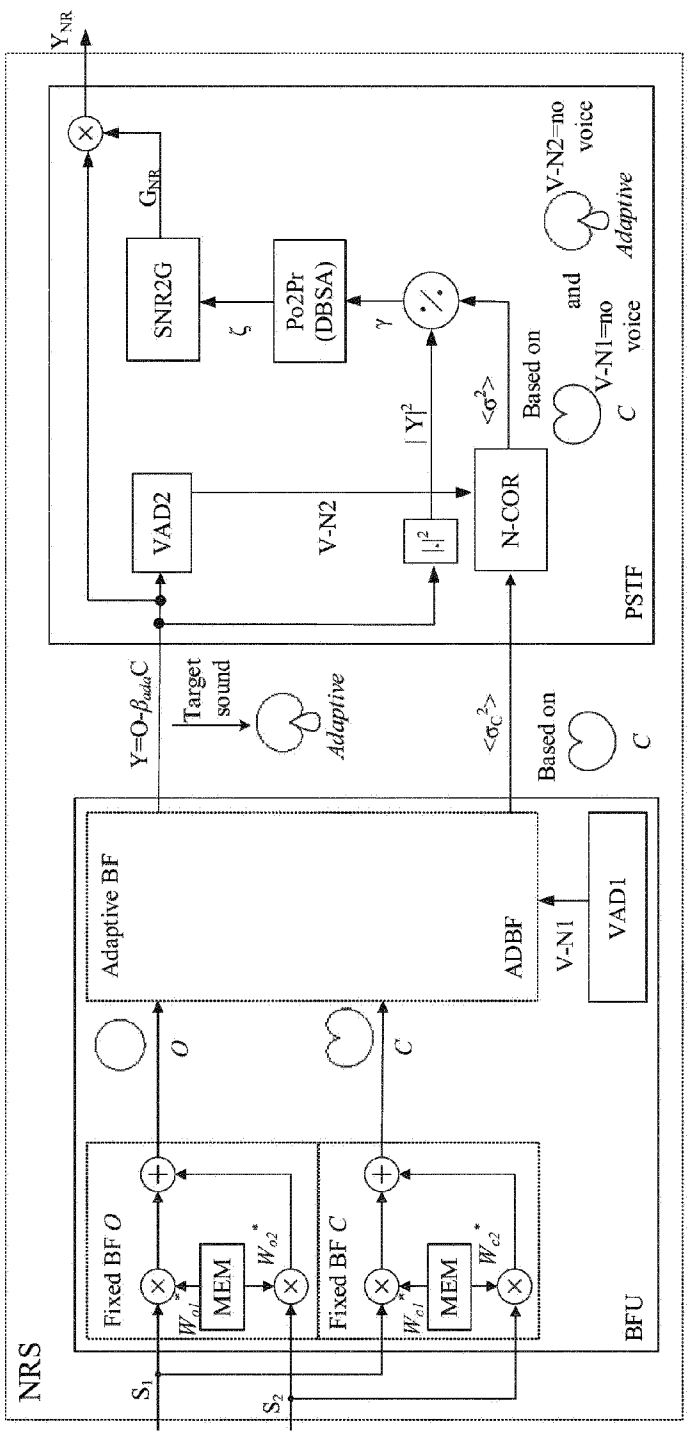

NRS
Y_{NR}
G_{NR}
SNR2G
ζ
Po2Pr (DBSA)
γ
V-N2=no voice
Adaptive
and
V-N1=no voice
C
Based on
$|Y|^2$
$<\sigma^2>$
VAD2
V-N2
$|.|^2$
N-COR
PSTF
$Y=O-\beta_{ada}C$
Target sound
Adaptive
$<\sigma_C^2>$
Based on
C
Adaptive BF
ADBF
V-N1
VAD1
O
C
Fixed BF O
Fixed BF C
MEM
MEM
W_{o1}
W_{o2}
W_{c1}
W_{c2}
S_1
S_2
BFU

FIG. 9B

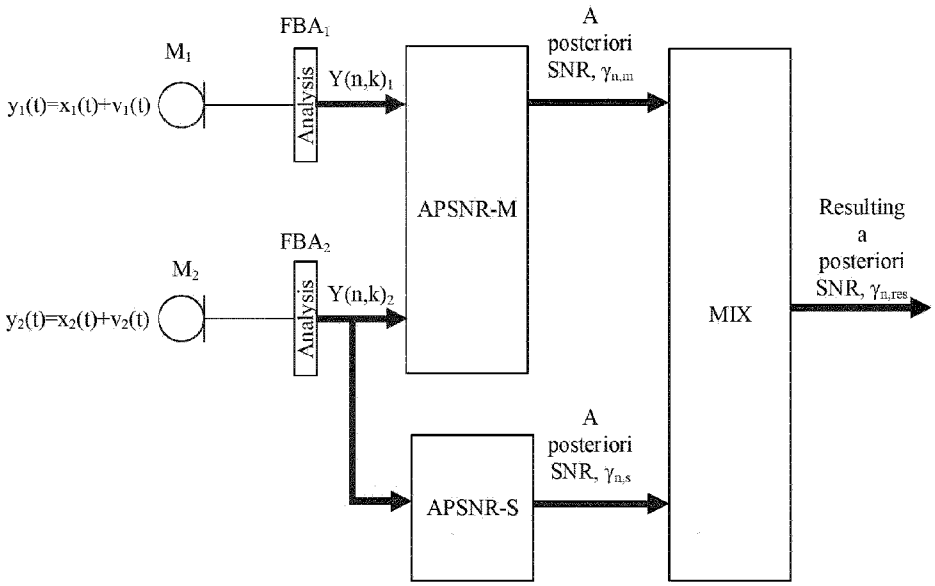

M1
FBA1
Analysis
Y(n,k)1
y1(t)=x1(t)+v1(t)
APSNR-M
A postcriori SNR, γn,m
M2
FBA2
Analysis
Y(n,k)2
y2(t)=x2(t)+v2(t)
MIX
Resulting a posteriori SNR, γn,res
APSNR-S
A posteriori SNR, γn,s

FIG. 10

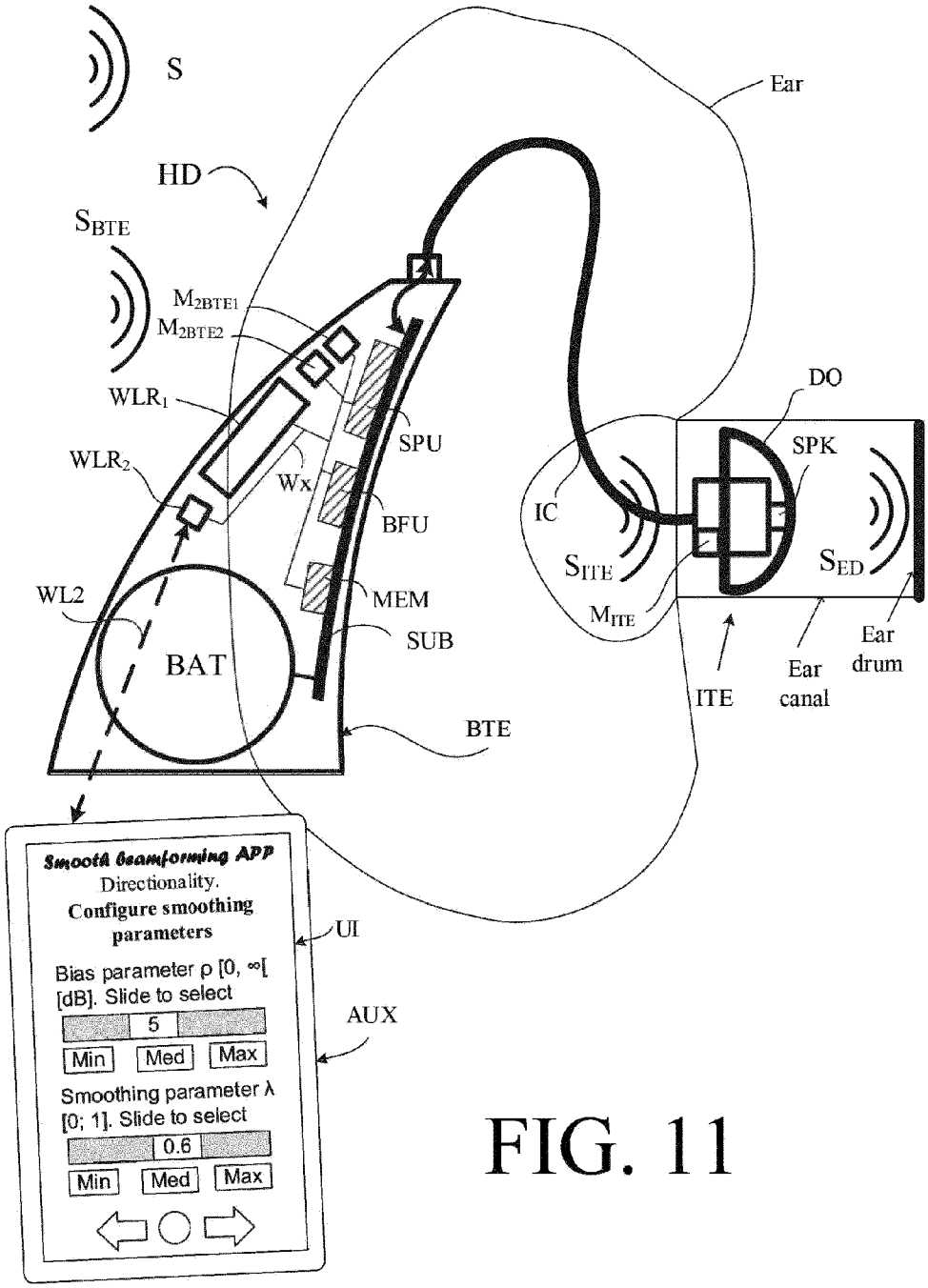

S
HD
S_{BTE}
M_{2BTE1}
M_{2BTE2}
WLR_1
WLR_2
Wx
SPU
BFU
MEM
SUB
WL2
BAT
BTE
Ear
DO
SPK
IC
S_{ITE}
M_{ITE}
S_{ED}
ITE
Ear canal
Ear drum
Smooth beamforming APP
Directionality.
Configure smoothing parameters
Bias parameter ρ [0, ∞[[dB]. Slide to select
5
Min
Med
Max
Smoothing parameter λ [0; 1]. Slide to select
0.6
Min
Med
Max
UI
AUX

FIG. 11