



(11) **EP 3 692 703 B9**

(12) **CORRECTED EUROPEAN PATENT SPECIFICATION**

(15) Correction information:

**Corrected version no 1 (W1 B1)**

**Corrections, see**

**Description Paragraph(s) 44, 46**

**Claims DE 9, 14**

**Claims EN 2, 9, 12, 14**

**Claims FR 9, 14**

(51) Int Cl.:

**H04M 9/08 (2006.01)**

(86) International application number:

**PCT/AT2017/060247**

(87) International publication number:

**WO 2019/068115 (11.04.2019 Gazette 2019/15)**

(48) Corrigendum issued on:

**17.11.2021 Bulletin 2021/46**

(45) Date of publication and mention  
of the grant of the patent:

**25.08.2021 Bulletin 2021/34**

(21) Application number: **17784831.4**

(22) Date of filing: **04.10.2017**

---

(54) **ECHO CANCELLER AND METHOD THEREFOR**

**ECHOKOMPENSATOR UND VERFAHREN DAFÜR**

**DISPOSITIF D'ANNULATION D'ÉCHO ET PROCÉDÉ À CET EFFET**

---

(84) Designated Contracting States:

**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**

(43) Date of publication of application:

**12.08.2020 Bulletin 2020/33**

(73) Proprietor: **proactivaudio GmbH**

**1100 Wien (AT)**

(72) Inventor: **WERUAGA, Luis**  
**1090 Wien (AT)**

(74) Representative: **Weiser & Voith**  
**Patentanwälte Partnerschaft**  
**Kopfgasse 7**  
**1130 Wien (AT)**

(56) References cited:

**EP-A1- 2 930 917 US-A1- 2006 018 460**

**EP 3 692 703 B9**

---

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

**Description****FIELD OF THE INVENTION**

5 **[0001]** The present invention relates to a computer-implemented method for updating at least one frequency-domain filter coefficient of an echo canceller having at least one channel and at least one segment per channel, the filter coefficients of the echo canceller being updatable in the frequency-domain at a time block. The invention further relates to an echo canceller configured to execute said method.

**BACKGROUND OF THE INVENTION**

10 **[0002]** Acoustic echo is a major impairment present in audio communications, such as in videoconferencing, in-car voice communications, voice inter-faces, and human-machine dialogue systems. Multi-channel acoustic echo (see "Acoustic multi-channel echo cancellation," EP 2438766 B1 and "Advances in digital speech transmission," Wiley, 2008) manifests when the sound from a plurality of loudspeakers connected to a terminal are captured by the microphone connected to said terminal. The terminal can refer to a desktop computer, or mobile phone, tablet, voice-commanded assistant, dedicated audioconferencing equipment, hand-free car telephony, et cetera.

15 **[0003]** As said terminal is usually placed within a room, the multiple sound reflections at the room walls arrive at the microphone at different time instants and intensities, hence creating a large number of acoustic echoes. The characteristics of said echo "fingerprint" is sensitive to the location of the microphone and loudspeaker(s), the room geometry and the objects and persons present therein.

20 **[0004]** The echo has a negative impact in the communication because participants in a videoconference hear their voices echoed as they speak, speech interfaces get "drowned" by the sounds (such as music) played by their loudspeaker system, and persons engaged in a mobile phone call to a car driver hear their own voices echoed inside the car chamber.

25 **[0005]** While mono systems still represent a good number of said terminals, multi-channel audio systems are increasingly appearing: car sound systems are at least stereo, dedicated videoconferencing equipment may use several loudspeakers for different participants, thus creating a more realistic meeting experience, and embedded voice interfaces are commonly integrated within stereo-sound equipment.

30 **[0006]** The present invention focuses on the problem scenario illustrated in Fig. 1, wherein a plurality of audio sources 10, e. g. audio channels, are played by a plurality of loudspeakers 20 in a room, and all the sounds in the room are captured by only one microphone 30. Its extension to more than one microphone (for instance to transmit stereo sound) is deemed straightforward, as the present invention can be applied on every micro-  
phone signal separately. Henceforth, "effective" loudspeaker refers to all electroacoustic device(s) that reproduce the same audio source or a linearly filtered version thereof. For instance if a stereo sound-reproduction system plays a single-channel (mono) audio source, the number of effective loudspeakers is one. Another example is a car stereo system, which may count several (more than two) physical loudspeakers, but if the car sound system is for instance reproducing stereo music, the number of effective loudspeakers, e. g. audio channels, is two.

35 **[0007]** The signal  $z(n)$  captured by the microphone 30 upon undergoing an analogue to digital conversion can be expressed as follows

40

$$z(n) = d(n) + v(n) + p(n) + \sum_{i=0}^{C-1} \sum_{l=0}^{L_H-1} h_i(l) s_i(n-l), \quad (1)$$

45

wherein  $n$  is discrete time,  $C$  is the number of far-end audio channels,  $s_i(n)$  is the far-end audio source at the  $i$ th-channel,  $h_i(l)$  is the equivalent discrete-time echo path between the  $i$ th effective loudspeaker and the microphone,  $L_H$  is the total length in samples of the echo path,  $d(n)$  is the main (or desired) near-end sound signal, and  $v(n)$  is the near-end background noise. The last term  $p(n)$  refers to the deviations from the linear echo model in (1), such as due to non-linear characteristics of the loudspeaker(s). The main near-end signal  $d(n)$  may result from several sources within the room. Usually  $d(n)$  is speech, music, or both, et cetera, of semantic value to the far-end listener, hence it is to be transmitted with minimum distortion. On the other hand, the near-end background noise  $u(n)$ , assumed to be nearly stationary, carries relevant contextual information of the near-end acoustic scene.

50 **[0008]** The objective of any acoustic echo reduction system is threefold:

55

1. to reduce the echo components in the microphone signal  $z(n)$  to the extent that they become inaudible,
2. to preserve the quality of the main near-end signal  $d(n)$  present in the microphone signal  $z(n)$  as much as possible,

and

3. to maintain the perceptual impression of the near-end background noise  $v(n)$  present in the microphone signal  $z(n)$ .

5 **[0009]** An acoustic echo canceller (AEC), a residual-echo non-linear processor (NLP), and a comfort-noise injector (CNI), all governed by an echo control apparatus, are meant to execute respectively said threefold objective.

**[0010]** In the following, a short overview of the technological background of the echo reduction system is given with reference to Figs. 1 and 2.

10 **[0011]** A basic multi-channel and multi-segment acoustic echo canceller 200 comprises a plurality of digital linear transversal filters 210, such that the response related to the  $i$ th channel is built as

$$u_i(n) = \sum_{j=0}^{S-1} \sum_{l=0}^{L-1} w_{i,j}(l) x_{i,j}(n-l), \quad (2)$$

wherein  $S$  is the number of segments per channel,  $L$  is the filter length in each segment,  $w_{i,j}(l)$  corresponds to the filter coefficients of the  $i$ th channel and  $j$ th segment, and  $x_{i,j}(n)$  is the input signal to filter  $w_{i,j}(l)$  defined as

$$x_{i,j}(n) = s_i(n-jL). \quad (3)$$

**[0012]** The concatenation of the filters from all  $S$  segments in the  $i$ th channel builds the impulse response of the echo canceller for said  $i$ th channel.

25 **[0013]** The synthetic echo model (2) is meant to deliver a signal that resembles the echoes in  $z(n)$  related to the  $i$ th channel. The global canceller output is built in 220 additively with the contributions from all  $C$  channels as follows

$$e(n) = z(n) - \sum_{i=0}^{C-1} u_i(n), \quad (4)$$

resulting in an output  $e(n)$  of reduced echo content.

35 **[0014]** A criterion to select the filter length  $L$  is namely to enclose large quasi-stationary intervals of the far-end signals  $s_i(n)$ . For instance, most speech phonemes are found within few tens of milliseconds, but, in longer ranges, speech usually contains at least two phonemes of different spectral characteristics. A stereo speech-based echo control for automotive spaces with echo path lengths in the order of 40-milliseconds could be designed as follows: the echo canceller would have two channels, e. g.  $C = 2$ , and at least two segments per channel, e. g.  $S \geq 2$ , with a filter length  $L$  equivalent to 20-milliseconds worth of signal samples, hence with four transversal filters,  $C \times S = 2 \times 2 = 4$ . Having different lengths for each filter could be possible too.

40 **[0015]** Because  $L$  is usually a large number, the time-domain convolution (2) turns out a costly operation. As alternative thereto, the response of the canceller can be efficiently computed in the frequency domain with  $N$ -point blocks, such that  $N$  is chosen as a power of two for better computational efficiency. Said frequency-domain block-based convolution is obtained as follows

$$\mathbf{u} = \mathbf{F}^{-1} \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathbf{W}_{i,j} \circ \mathbf{X}_{i,j}, \quad (5)$$

wherein  $\mathbf{F}$  is the  $N$ -point discrete Fourier transform (DFT) matrix ( $\mathbf{F}^{-1}$  corresponds thus to the inverse discrete Fourier transform), and the operation  $\circ$  denotes element-by-element multiplication. The  $N$ -point column vectors  $\mathbf{W}_{i,j}$  and  $\mathbf{X}_{i,j}$ , which refer respectively to the frequency-domain weights and far-end input block of the  $i$ th channel and  $j$ th segment, are built as follows

$$\mathbf{W}_{i,j} = \mathbf{F} [w_{i,j}(0) \cdots w_{i,j}(L-1) \overset{\leftarrow N-L \rightarrow}{0} \cdots 0]^T \quad (6)$$

$$\mathbf{X}_{i,j} = \mathbf{F} [x_{i,j}(n-N+1) \cdots x_{i,j}(n-1) x_{i,j}(n)]^T, \quad (7)$$

5 wherein  $T$  denotes transpose.

**[0016]** The operation (5) delivers exactly  $M = N - L + 1$  valid samples of the linear convolution, said samples being located in the last  $M$  rightmost elements of vector  $\mathbf{u}$ . The  $M$  valid errors, obtained with the difference between the microphone signal and the overall canceller response as per (4), are stored in vector  $\mathbf{e}$  for further processing

$$\mathbf{e} = [e(n-M+1) \cdots e(n)]^T. \quad (8)$$

15 **[0017]** In order to obtain a canceller response (5) closer to the actual echo in the microphone signal  $z(n)$ , the frequency-domain weights  $\mathbf{W}_{i,j}$  of the  $i$ th channel and  $j$ th segment are updated via the fast LMS algorithm (see "Acoustic echo devices and methods," US 2006/0018460 A1) as follows

$$\mathbf{W}_{i,j}^{\text{new}} = \mathbf{W}_{i,j} + \mu \mathbf{E} \circ \mathbf{X}_{i,j}^* / \left( \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} |\mathbf{X}_{i,j}|^2 + \epsilon \right), \quad (9)$$

25 wherein  $\mathbf{W}_{i,j}^{\text{new}}$  contains the updated weights,  $\mathbf{E}$  is the frequency-domain counterpart of the canceller output  $\mathbf{e}$ , obtained as follows

$$\mathbf{E} = \mathbf{F} [0 \cdots \overset{\leftarrow N-M \rightarrow}{0} \mathbf{e}^T]^T, \quad (10)$$

30 the operator  $*$  denotes complex conjugate, the vector operator  $/$  represents element-by-element division,  $|\mathbf{X}_{i,j}|^2$  is a vector built with the square magnitude of every complex-valued element in vector  $\mathbf{X}_{i,j}$ ,  $\epsilon > 0$  prevents division by zero or values near zero, and  $\mu$  is the step size.

35 **[0018]** Because the time-domain counterpart of the new weight vector  $\mathbf{W}_{i,j}^{\text{new}}$  should contain zeroes in its rightmost side (6), the new spectral weights (9) can be further updated by returning said weights to the time domain, resetting (to zero) the last  $N - L$  taps, and returning the result back to the frequency domain. Said time-constraint operation demands two additional DFT operations, hence increasing the computation complexity. If desired, this step can be avoided at the price of slower convergence (see "Unconstrained frequency-domain adaptive filter," IEEE Transactions on Acoustics, Speech and Signal Processing, October, 1982). As this operation is well known by any expert in the art, we use (9) as the generic representation of the weight update rule, which encompasses the option to perform (or not) said time-constraint operation.

40 **[0019]** It is known that the convergence of said adaptive system (9) is assured for a step size in the range  $0 < \mu < 2$ . However, the range  $0 < \mu \leq 1$  is considered in the practice, whose limits correspond to frozen update  $\mu = 0$  and fastest update  $\mu = 1$  in interference-free conditions.

45 **[0020]** The generic multi-channel and multi-segment adaptive system (9) has been matter of thorough study (see "Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication," Signal Processing 85, pp. 549-570, 2005, "Acoustic echo devices and methods," US 2006/0018460 A1, "Acoustic multi-channel echo cancellation," EP 2438766 B1, "Method and apparatus for multi-channel audio processing using single-channel components," US 8233632 B1, and "Subband echo cancellation method for multi-channel audio teleconference and echo canceller using the same," US 6246760 B1).

50 **[0021]** In order for the prior state of art to succeed in real scenarios, several challenges are yet to be effectively solved:

- 55 • The presence of non-cancellable terms, such as  $d(n)$  and  $v(n)$ , in error  $\mathbf{E}$  corrupts the update (9), which as result exhibits unstable behaviour. Double-talk detectors (DTT) aim to prevent this drawback by detecting when it is safe to train the adaptive system (see "Enhanced echo cancellation," US 8873740 B2). So far, DTTs have not proven accurate nor quick enough, especially under severe double-talk. Hence, robustness against any type of double-talk

is yet to be effectively solved.

- As the global error  $E$  is used for the update (9) of every filter, the misalignment of any filter is propagated to the remaining  $C \times S - 1$  filters, and vice versa. Despite being considered intrinsic in adaptive filtering, this fact harms the canceller's learning abilities. For instance, in web-based audioconferencing, user motion or change of the terminal orientation has larger impact on early echoes. It is thus desirable to count with a mechanism of true filter-independent update.

**[0022]** Upon the previous two drawbacks, attempting to adapt the filters with the canceller output  $E$  according to the state of art results in unstable and slow convergence, hence failing to reduce the echo quick enough to acceptable levels.

**[0023]** Upon the echo cancellation, the canceller output  $E$  may still contain audible echo because:

- the echo path length  $L_H$  is larger than the canceller length,  $L_H \gg S \times L$ , hence the late-arriving echoes cannot be cancelled,
- the adaptive canceller is yet to converge to the "steady state", and
- non-linear echo components, characterized by  $p(n)$  in (1), cannot be removed by the linear echo canceller (5).

**[0024]** The canceller output can be thus rewritten in the following additive model

$$\mathbf{E} = \mathbf{D} + \mathbf{V} + \mathbf{P} + \mathbf{Q} + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathbf{R}_{ij}, \quad (11)$$

wherein  $\mathbf{D}$  corresponds to the main near-end signal,  $\mathbf{V}$  is the near-end background noise,  $\mathbf{P}$  corresponds to the non-linear echo components,  $\mathbf{Q}$  represents the echo beyond the time scope of the canceller, and  $\mathbf{R}_{ij}$  is the residual echo related to the  $i$ th channel and  $j$ th segment.

**[0025]** In order to reduce effectively the residual echo  $\mathbf{P}$ ,  $\mathbf{Q}$ , and  $\mathbf{R}_{ij}$  to inaudible levels, a non-linear processing (NLP) stage is to act over the canceller output  $\mathbf{E}$ , delivering the echo-free output  $\mathbf{O}$ . Several strategies are possible thereto, such as filtering, spectral subtraction, or by following principles of psychoacoustics, this last one being the preferred option when dealing with music signals. Let us consider for the purpose of illustration a filtering-based approach

$$\mathbf{O} = \mathbf{E} \circ \mathbf{B}_{NLP}, \quad (12)$$

wherein  $0 \leq \mathbf{B}_{NLP} \leq 1$  is the frequency-response of the NLP filter.

**[0026]** It is important to remark that the non-linear processing (12), unlike echo cancellation (9), introduces distortion to the main near-end signal  $\mathbf{D}$  present in the canceller output  $\mathbf{E}$ , which creates a delicate tradeoff between residual echo reduction versus main near-end signal distortion. The stage of echo cancellation plays a decisive role in said tradeoff: if the echo canceller is notably misaligned, the NLP filtering must act aggressively over the residual echo at the risk of damaging the main near-end signal; conversely, if the canceller manages to remove efficiently most of the echo, the impact of the NLP filter over the main near-end signal may go unnoticed. Therefore, the accurate estimation of all residual echo terms,  $\mathbf{P}$ ,  $\mathbf{Q}$ , and  $\mathbf{R}_{ij}$  present in the canceller output  $\mathbf{E}$  is important for an effective NLP stage.

**[0027]** Residual echo suppression is concern of recent prior art, such as in "Acoustic echo suppression," US 2017/0118326 A1, wherein the activation of the echo suppressor depends on one or more transient or steady-state parameters. In another recent invention, "Robust acoustic echo cancellation for loosely paired devices based on semi-blind multichannel demixing," US 2016/0029120 A1, a semi-blind multichannel source separation is performed to decompose the audio signals into a near-end source signal and residual echoes based on independent component analysis, thus requiring more than one (multiple) microphones to operate.

**[0028]** The near-end background noise  $\mathbf{V}$  is distorted by the NLP operation (12) as well. This fact causes periods of silence and/or sudden changes in the background noise level, which can be distracting, if not annoying, for the listener. Comfort noise injection (see "Comfort noise generator for echo cancelers," US 5949888 A1) solves this drawback by filling up the gaps with synthetic noise, hence giving the illusion of a uniform background noise. The injected synthetic noise must resemble the actual near-end background noise  $v(n)$  in both spectral content and level. The final output  $\mathbf{Y}$  of the echo reduction system can be obtained as follows

$$\mathbf{Y} = \mathbf{O} + \hat{\mathbf{V}} \circ (1 - \mathbf{B}_{NLP}), \quad (13)$$

wherein  $\hat{\mathbf{V}}$  is the synthetic replica of the near-end background noise. Accurate spectral estimation of the near-end background noise  $\mathbf{V}$  turns out important to maintain a pleasant listening experience.

[0029] In the prior art "Method and apparatus for comfort noise generation in speech communication systems" US 7610197 B2, the background noise is obtained by accumulating and smoothing over time the spectral samples  $\mathbf{E}$ ; in the inventions "Comfort noise generation method and system," US 2011/0228946 A1, "Enhanced echo cancellation," US 8873740 B2, the estimation of the background noise is carried out during periods when both far-end and near-end are inactive. The state of the art may not always work: in case of audioconferencing in moderately reverberated rooms, the residual echo beyond the canceller scope  $\mathbf{Q}$  may fill up periods of silence in the near-end signal, thus acting de-facto as background noise; this problem is further stressed with gapless audio content such as music, in which the residual echo can be mistaken as noise in the background.

**SUMMARY OF THE INVENTION**

[0030] In view of the problems discussed above, is an object of the invention to provide an improved acoustic echo canceller that is more stable and yields a faster convergence.

[0031] The problems are solved in a first aspect of the invention by a computer-implemented method for updating at least one frequency-domain filter coefficient  $W_{i,j}(k)$  of an echo canceller having at least one channel and at least one segment per channel, the filter coefficients of the echo canceller being updateable in the frequency domain at a time block  $m$  comprising:

determining a canceller output  $E_m(k)$  over the  $m$ th time block as the difference between  $Z_m(k)$  and a cancelling term based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m}(k); \tag{14}$$

determining a canceller error  $\varepsilon(k)$  over the  $m - \ell$ th time block as the difference between  $Z_{m-\ell}(k)$  and a cancelling term based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m-\ell}(k); \tag{15}$$

determining a look-backward error  $\overleftarrow{\varepsilon}(k)$  as

$$\overleftarrow{\varepsilon}(k) = \varepsilon(k) - \lambda \overleftarrow{\Delta}(k), \tag{16}$$

wherein  $\overleftarrow{\Delta}(k)$  is based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k); \tag{17}$$

determining a look-forward error  $\overrightarrow{\varepsilon}(k)$  as

$$\overrightarrow{\varepsilon}(k) = E_m(k) - \lambda \overrightarrow{\Delta}(k), \tag{18}$$

wherein  $\overrightarrow{\Delta}(k)$  is based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k); \quad (19)$$

5 determining an optimal update step size  $\mu_{i,j,m}(k)$  from said canceller output  $E_m(k)$  over the  $m$ th time block, from said canceller error  $\epsilon(k)$  over the  $m - \ell$ th time block, from said look-backward error  $\vec{\alpha}(k)$ , and from said look-forward error  $\vec{\alpha}(k)$ ; and  
 10 updating said at least one filter coefficient  $W_{i,j}(k)$  by using said optimal update step-size  $\mu_{i,j,m}(k)$ ; wherein

- C corresponds to the number of channels  $i = 0, \dots, C - 1$  of the echo canceller,
- 15 S corresponds to the number of segments  $j = 0, \dots, S - 1$  per channel of the echo canceller,
- $\ell$  is an integer different than zero,
- $k$  denotes a frequency bin index,
- 20  $\epsilon$  is a positive value,
- $Z_m(k)$  corresponds to the  $k$ th spectral bin of the microphone signal at the  $m$ th time block,
- 25  $Z_{m-\ell}(k)$  corresponds to the  $k$ th spectral bin of the microphone signal at the  $m - \ell$ th time block,
- $X_{i,j,m}(k)$  corresponds to the  $k$ th spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m$ th time block,
- 30  $X_{i,j,m-\ell}(k)$  corresponds to the  $k$ th spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m - \ell$ th time block,
- $\lambda$  is the overshoot factor, and
- $a, b$  are channel and segment indices, respectively.

35 **[0032]** The claimed method yields an optimal step size  $\mu_{i,j,m}(k)$  that is very accurate thanks to using the information from the canceller output  $E_m(k)$  at the  $m$ th block, the canceller error  $\epsilon(k)$  at the  $m - \ell$ th block, and the look-forward error  $\vec{\alpha}(k)$  in addition to the look-backward error  $\vec{\alpha}(k)$ . Because the integer  $\ell$  can take a positive value, the  $m - \ell$ th block is  
 40 already available, hence not incurring any delay in the update, that is, the optimal step size  $\mu_{i,j,m}(k)$  can be immediately obtained as soon as the  $m$ th time block is available. Said filter update with the optimal step size  $\mu_{i,j,m}(k)$  as computed with the inventive method yields a faster convergence and a more effective reduction of echo than echo cancellers of the state of the art.

45 **[0033]** The wording "based on", adopted when introducing  $E_m(k)$  and  $\epsilon(k)$  with (14) and (15), respectively, refers to additional standard signal processing, not explicitly included therein for the sake of simplicity in the notation as well as for being known to anyone skilled with the art, namely, discarding the invalid samples that result from the circular convolution inherent to the product of two frequency-domain sequences, such as in said (14) and (15). These arguments can be extended to the definition of  $\vec{\Delta}(k)$  and  $\vec{\Delta}(k)$  "based on" (17) and (19), respectively, as well as to terms computed in preferred embodiments.

50 **[0034]** Preferably, the optimal step-size  $\mu_{i,j,m}(k)$  is computed by either

$$\mu_{i,j,m}(k) = \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)} \quad (20)$$

55 or

$$\mu_{i,j,m}(k) = 1 - \sqrt{1 - \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}}, \quad (21)$$

wherein  $\Psi_m(k)$  is a power spectrum of the non-cancellable components at the  $m$ th time block and  $\Phi_{i,j}(k)$  is a power spectrum of the misalignment of each  $i$ th filter, wherein  $\Psi_m(k)$  and  $\Phi_{i,j}(k)$  are determined by solving the set of linear equations

$$\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |E_m(k)|^2 \quad (22)$$

$$Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\mathcal{E}(k)|^2 \quad (23)$$

$$Y(k) + 4 \overleftarrow{T}(k) \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}(k)|^2 \quad (24)$$

$$4 \overrightarrow{T}(k) Y(k) + \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}(k)|^2 \quad (25)$$

subject to inequality constraints

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0; \quad (26)$$

wherein

$\mathcal{X}_{i,j,m}(k)$  corresponds to the  $k$ th power spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m$ th time block, wherein said signal block is preferably windowed,

$\mathcal{X}_{i,j,m-\ell}(k)$  corresponds to the  $k$ th power spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m - \ell$ th time block, wherein said signal block is preferably windowed,

$Y(k)$  corresponds to a power spectrum of the non-cancellable components at the  $m - \ell$ th block,

$\overleftarrow{T}(k)$  is the normalized look-backward update, computed as

$$\overleftarrow{T}(k) = \overleftarrow{\Delta}(k) / |E_m(k)|^2, \quad (27)$$

and

$\overrightarrow{T}(k)$  is the normalized look-forward update, computed as

$$\overrightarrow{T}(k) = \overrightarrow{\Delta}(k) / |\mathcal{E}(k)|^2. \quad (28)$$

**[0035]** The set of linear equations with inequality constraints is one of the main foundations of the inventive method.

Because the unknowns  $\Upsilon(k)$  and  $\Psi_m(k)$  appear in said set of linear equations in a balanced way, the condition of the problem is excellent. Removing one of the four equations would make one of those unknowns less visible, thus deteriorating significantly the condition of the whole problem (as the overall condition is bound by the worst of its parts). Because the number of equations is equal to four, the inventive method is particularly suited to stereo systems  $C = 2$  with one segment per channel  $S = 1$ , as this case leads to four unknowns.

**[0036]** Moreover, the inventive method is especially useful for general multi-channel setups, e. g.  $C > 1$ , that are under frequent near-end activity since the power spectrum of the non-cancellable components  $\Psi_m(k)$  is obtained with high accuracy, and for the first time optimal continuous adaptation can be achieved regardless of the presence of near-end components, while in the state of the art a double-talk detector needs to assess whether there is a near-end signal to temporarily stop the update of all filter coefficients.

**[0037]** The first formula (20) for the optimal step size is simpler than the second one (21), as the latter involves the root square operation; however, formula (21) delivers better results. Therefore, if the second formula is chosen, its Taylor approximation can be used, delivering similar performance at much lower computational complexity.

**[0038]** Another problem that can arise is when at least the number of channels or the number of segments per channel is larger than one, e. g. when the canceller is composed of several transversal filters: as the global error  $E_m(k)$  is used for the update of every filter, the misalignment of a given filter is propagated to the remaining filters upon the update, and vice versa, and thus slowing down the convergence of the entire adaptive system. To overcome this undesired drawback, one at a time any chosen  $r$ ,sth filter can be considered not to be part of the canceller, e. g. its residual echo to be treated as non-cancellable component, and an additional look-backward error  $\overleftarrow{\varepsilon}_{r,s}(k)$  and an additional look-forward error  $\overrightarrow{\varepsilon}_{r,s}(k)$  involving the update off-line of all filters except the  $r$ , sth one are computed.

**[0039]** In order to employ the latter strategy, the invention provides a computer-implemented method for updating at least one frequency-domain filter coefficient  $W_{i,j}(k)$  of the echo canceller, wherein at least one of the number of channels or the number of segments per channel is larger than one; and

the optimal step size  $\mu_{i,j,m}$  is determined from said canceller output  $E_m(k)$  over the  $m$ th time block, from said canceller error  $\varepsilon(k)$  over the  $m - \ell$ th time block, from said look-backward error  $\overleftarrow{\varepsilon}(k)$ , from said look-forward error  $\overrightarrow{\varepsilon}(k)$  as well as from at least one additional look-backward error  $\overleftarrow{\varepsilon}_{r,s}(k)$  and from at least one additional look-forward error  $\overrightarrow{\varepsilon}_{r,s}(k)$  for at least one  $r$ ,sth filter,  $r$  and  $s$  being indices of the channel and the segment, respectively, wherein

$$\overleftarrow{\varepsilon}_{r,s}(k) = \mathcal{E}(k) - \lambda \overleftarrow{\Delta}_{r,s}(k), \quad (29)$$

wherein  $\overleftarrow{\Delta}_{r,s}(k)$  is based on

$$\sum_i \sum_j \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_a \sum_b |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k), \quad (30)$$

and

$$\overrightarrow{\varepsilon}_{r,s}(k) = E_m(k) - \lambda \overrightarrow{\Delta}_{r,s}(k), \quad (31)$$

wherein  $\overrightarrow{\Delta}_{r,s}(k)$  is based on

$$\sum_i \sum_j \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_a \sum_b |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k). \quad (32)$$

**[0040]** Including the information from the additional look-backward  $\overleftarrow{\varepsilon}_{r,s}(k)$  and look-forward  $\overrightarrow{\varepsilon}_{r,s}(k)$  errors allows the inventive method to explicitly accomplish a more accurate estimation of the optimal step size(s)  $\mu_{r,s,m}(k)$ , as well as implicitly improving the optimal step size for other  $i, j$ th filters, which translates in faster convergence.

**[0041]** In this embodiment, the optimal step-size  $\mu_{i,j,m}(k)$  is preferably computed as per either (20) or (21), wherein the power spectrum  $\Psi_m(k)$  of the non-cancellable components and the power spectrum  $\Phi_{i,j}(k)$  of the misalignment of each  $i, j$ th filter are determined by solving said set of linear equations (22), (23), (24), (25), with additional linear equations for the at least one  $r, s$  th filter

$$Y(k) + 4\overleftarrow{T}_{r,s}(k)\Psi_m(k) + 4\overleftarrow{T}_{r,s}(k)\mathcal{X}_{r,s,m}(k)\Phi_{r,s}(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-1}(k)\Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}_{r,s}(k)|^2 \quad (33)$$

$$\Psi_m(k) + 4\overrightarrow{T}_{r,s}(k)Y(k) + 4\overrightarrow{T}_{r,s}(k)\mathcal{X}_{r,s,m-1}(k)\Phi_{r,s}(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k)\Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}_{r,s}(k)|^2 \quad (34)$$

subject to inequality constraints

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0; \quad (35)$$

wherein

$\overleftarrow{T}_{r,s}(k)$  is the normalized look-backward update excluding the  $r, s$ th filter, computed as

$$\overleftarrow{T}_{r,s}(k) = \overleftarrow{\Delta}_{r,s}(k) / |E_m(k)|^2, \quad (36)$$

and

$\overrightarrow{T}_{r,s}(k)$  is the normalized look-forward update excluding the  $r, s$ th filter, computed as

$$\overrightarrow{T}_{r,s}(k) = \overrightarrow{\Delta}_{r,s}(k) / |\mathcal{E}(k)|^2. \quad (37)$$

**[0042]** These embodiments are especially useful not only in scenarios with frequent near-end activity, as discussed previously, but also for multi-channel and multi-segment setups, e. g.  $C > 1$  or  $S > 1$ , as the global learning convergence is substantially improved, in particular by improving explicitly the estimation of the filter misalignment  $\Phi_{r,s}(k)$ , hence the convergence for each  $r, s$ th filter(s) therewith, while in the state of art the global error is spread uniformly through all filters, which leads to slower convergence.

**[0043]** Until now, it was assumed that the overshoot factor  $\lambda$  needed to be equal to 1 as this led to a good trade-off between misalignment reduction and amplification of the non-cancellable components. However, it has been analytically found that even better results can be achieved if the overshoot factor  $\lambda$  is chosen to be greater than one, in the range  $1 < \lambda \leq 2$ , and particularly substantially when  $\lambda = 2$ . At said value the filter misalignment terms in the look-backward  $\overleftarrow{\varepsilon}(k)$  and look-forward  $\overrightarrow{\varepsilon}(k)$  errors have the same expected level as in the errors  $\varepsilon(k)$  and  $E_m(k)$ , but the non-cancellable components turn out exactly two times larger in amplitude, hence more "visible;" this fact translates into an estimation gain boost of the non-cancellable components on more than 6 dB. Likewise, an overshoot factor equal to 2 makes in particular the  $r, s$ th filter misalignment appear exactly two times larger in magnitude in both look-backward  $\overleftarrow{\varepsilon}_{r,s}(k)$  and look-forward  $\overrightarrow{\varepsilon}_{r,s}(k)$  errors, hence more "visible," which translates into an estimation gain boost of said filter misalignment

on more than 6 dB.

**[0044]** In a preferred embodiment of the invention, the method furthermore determines:

a power spectrum  $\mathcal{R}_m(k)$  of the residual echo due to filter misalignment at the  $m$ th block by means of

$$\mathcal{R}_m(k) = \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \Phi_{i,j}(k) \chi_{i,j,m}(k), \quad (38)$$

and/or

a power spectrum  $\mathcal{Q}_m(k)$  of the echo beyond the canceller's time scope at the  $m$ th block by means of the accumulative method

$$\mathcal{Q}_m(k) = \gamma \mathcal{Q}_{m-1}(k) + \gamma \sum_{i=0}^{C-1} |W_{i,S-1}(k)|^2 \chi_{i,S-1,m-1}(k), \quad (39)$$

wherein  $0 < \gamma < 1$ , and/or

a power spectrum  $\mathcal{P}_m(k)$  of the non-linear echo components at the  $m$ th block by means of

$$\mathcal{P}_m(k) \propto \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mu_{i,j,m}(k) \Phi_{i,j}(k) \chi_{i,j,m}(k), \quad (40)$$

wherein  $\propto$  denotes proportional,

and subsequently an improved canceller output  $O_m(k)$  based on the canceller output  $E_m(k)$  and on at least one of

the determined power spectra of residual echo due to the filter misalignment  $\mathcal{R}_m(k)$ , the echo beyond the canceller's time scope  $\mathcal{Q}_m(k)$ , and the non-linear echo  $\mathcal{P}_m(k)$ .

**[0045]** The filter misalignment terms  $\Phi_{i,j}(k)$  and the power spectra  $\chi_{i,j,m}(k)$  employed to evaluate  $\mathcal{R}_m(k)$  and  $\mathcal{P}_m(k)$  correspond to the same terms used to evaluate the optimal step size(s) (20) and (21). In consequence, those terms are re-used, hence keeping the computational complexity low while profiting from the high estimation accuracy of said

terms. On the other hand, the term  $\mathcal{Q}_m(k)$  involves the filter  $W_{i,S-1}(k)$ , which is quickly updated at every block by the optimal mechanism of the inventive method.

**[0046]** One way to obtain said improved canceller output  $O_m(k)$  with the canceller output  $E_m(k)$  while including the above mentioned terms (38), (39), and (40) is to use a Wiener filter or any other type known in the state of the art. If the preferred Wiener filter is to be implemented, the improved canceller output  $O_m(k)$  is determined as

$$O_m(k) = E_m(k) B_m(k) \quad (41)$$

$$B_m(k) = \left[ \frac{|E_m(k)|^2 - \Lambda_m(k)}{|E_m(k)|^2} \right]_{\varepsilon} \quad (42)$$

$$\Lambda_m(k) = \mathcal{R}_m(k) + \mathcal{P}_m(k) + \mathcal{Q}_m(k), \quad (43)$$

5 wherein

$\mathcal{L}_{\downarrow \varepsilon}$  denotes low clipping by  $\varepsilon > 0$ ,

10  $E_m(k)$  is the canceller output,

$R_m(k)$  is a power spectrum of the residual echo due to canceller filter misalignment at the  $m$ th block,

15  $Q_m(k)$  is a power spectrum of the echo beyond the canceller's time scope at the  $m$ th block, and

$P_m(k)$  is a power spectrum of the non-linear echo component(s) at the  $m$ th block.

[0047] The determination of the spectral power of the three residual echo terms (38), (39), and (40) are used for suppressing, e. g. filtering out (41), otherwise said residual echo components present in the canceller output  $E_m(k)$ . In some cases, it can be preferred to only compute one or two of the above-mentioned three terms because one or two terms might be significantly lower with respect to the other(s). In order to deliver a listening signal, the improved canceller output  $O_m(k)$  is to be transformed to the time domain. Standard overlap-add processing, well known to anybody familiar with the art, can be used in order to avoid block-effects in the resulting time-domain output.

25 [0048] Preferably, the at least one filter coefficient  $W_{i,j}(k)$  is updated based on

$$W_{i,j}^{new}(k) = W_{i,j}(k) + \mu_{i,j,m}(k) \frac{E_m(k) X_{i,j,m}^*(k)}{|X_{i,j,m}(k)|^2 + \varepsilon}. \quad (44)$$

30 [0049] Even though the inventive method governs a multi-channel/-segment echo canceller, the preferred update (44) resembles a single-filter adaptive system, as the normalization term involves only the  $i,j$ th far-end power spectrum  $|X_{i,j,m}(k)|^2$ , in contrast with an standard multi-filter canceller, whose normalization term involves the accumulation of all

35 far-end power spectra. This subtle detail, a real novelty in the art, implies ultimately that the  $i,j$ th input data  $X_{i,j,m}^*(k)$  is effectively normalized, and that the main responsibility of the update lies on the optimal step size  $\mu_{i,j,m}(k)$  determined by the inventive method.

40 [0050] In a second aspect, the invention provides for an echo canceller having at least one channel and at least one segment, the echo canceller having filter coefficients  $W_{i,j}(k)$  being updateable in the frequency domain at a time block  $m$ , wherein the echo canceller is configured to:

determine a canceller output  $E_m(k)$  over the  $m$ th time block as the difference between  $Z_m(k)$  and a cancelling term based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m}(k); \quad (45)$$

50 determine a canceller error  $e(k)$  over the  $m - \ell$ th time block as the difference between  $Z_{m-\ell}(k)$  and a cancelling term based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m-\ell}(k); \quad (46)$$

determine a look-backward error  $\overleftarrow{\varepsilon}(k)$  as

$$\overleftarrow{\varepsilon}(k) = \mathcal{E}(k) - \lambda \overleftarrow{\Delta}(k) \quad (47)$$

wherein  $\overleftarrow{\Delta}(k)$  is based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m}(k)|^2 + \varepsilon} X_{i,j,m-\ell}(k); \quad (48)$$

determine a look-forward error  $\overrightarrow{\varepsilon}(k)$  as

$$\overrightarrow{\varepsilon}(k) = E_m(k) - \lambda \overrightarrow{\Delta}(k) \quad (49)$$

wherein  $\overrightarrow{\Delta}(k)$  is based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m-\ell}(k)|^2 + \varepsilon} X_{i,j,m}(k); \quad (50)$$

determine an optimal update step-size  $\mu_{i,j,m}(k)$  from said canceller output  $E_m(k)$  over the  $m$ th time block, from said canceller error  $\varepsilon(k)$  over the  $m - \ell$ th time block, from said look-backward error  $\overleftarrow{\varepsilon}(k)$ , and from said look-forward error  $\overrightarrow{\varepsilon}(k)$ ; and

update said at least one filter coefficient  $W_{i,j}(k)$  by using said optimal update step-size  $\mu_{i,j,m}(k)$ ; wherein

- C corresponds to the number of channels  $i = 0, \dots, C - 1$  of the echo canceller,
- S corresponds to the number of segments  $j = 0, \dots, S - 1$  per channel of the echo canceller,
- $\ell$  is an integer different than zero,
- $k$  denotes a frequency bin,
- $\varepsilon$  is a positive value,
- $Z_m(k)$  corresponds to the  $k$ th spectral bin of the microphone signal at the  $m$ th time block,
- $Z_{m-\ell}(k)$  corresponds to the  $k$ th spectral bin of the microphone signal at the  $m - \ell$ th time block,
- $X_{i,j,m}(k)$  corresponds to the  $k$ th spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m$ th time block,
- $X_{i,j,m-\ell}(k)$  corresponds to the  $k$ th spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m - \ell$ th time block,
- $\lambda$  is the overshoot factor, and
- $a, b$  are channel and segment indices, respectively.

[0051] All advantages of the embodiments and variants of the method as described above also hold for the echo canceller.

**BRIEF DESCRIPTION OF THE DRAWINGS**

5 [0052] The invention shall not be explained in more detail below on the basis of the preferred exemplary embodiments thereof with reference to the accompanying drawings, in which:

- 10 Fig. 1 illustrates the problem and scenario of multi-channel acoustic echo reduction.
- Fig. 2 illustrates the proposed computer-based system to control and update the multi-channel acoustic echo cancellation, the residual echo suppression, and the comfort noise injection at every time block.
- Fig. 3 illustrates the proposed computer-based apparatus to decompose the echo canceller output into the non-cancellable components and the residual echo for each channel and segment.
- Fig. 4 discloses a computer-based method to compute the echo canceller errors at two different time blocks.
- 15 Fig. 5 illustrates the impact of the "overshoot" factor in the frequency-domain adaptive filter.

**DETAILED DESCRIPTION OF THE INVENTION**

20 [0053] In several embodiments of the present invention, a method and apparatus for multi-channel acoustic echo canceller 50 is disclosed, which governs a first stage of multi-filter adaptive echo cancellation 200 robust to near-end interferences and with improved update for each transversal filter, followed by a second stage of residual echo suppression 300 that reduces the residual echo present in the canceller output to inaudible levels while preserving the quality of the main near-end signal, and followed by a third and last stage of comfort noise injection 400 that replenishes the suppressed near-end background noise with synthetic noise that resembles in level and spectrum said actual near-end background noise.

**Acoustic Echo Cancellation**

30 [0054] The output (4) of the multi-channel and multi-segment acoustic echo canceller 200 object of the present invention can be either obtained directly in the time domain (2) or with the frequency-domain mechanism (5). Regardless of the choice, the canceller 50 is trained in the frequency domain.

[0055] In the present invention, the frequency-domain coefficients of the *i*, *j*th filter, wherein *i* and *j* are channel and segment indices, respectively, are updated in 230 according to the rule

35 
$$W_{i,j}^{new} = W_{i,j} + \mathcal{T}_1 \left\{ \mu_{i,j,m} \circ E_m \circ X_{i,j,m}^* / (|X_{i,j,m}|^2 + \epsilon) \right\}, \quad (51)$$

40 wherein index *m* denotes the *m*th time block, that is, vector *E<sub>m</sub>* contains the frequency-domain conversion of *M* time-domain output samples

45 
$$E_m = F [0 \dots 0 \ e((m-1)M+1) \ \dots \ e(mM)]^T, \quad (52)$$

while *X<sub>i,j,m</sub>* is built as follows

50 
$$X_{i,j,m} = F [x_{i,j}(mM-N+1) \ \dots \ x_{i,j}(mM)]^T, \quad (53)$$

the operation  $\mathcal{T}_1 \{ \}$  corresponds to the standard time-domain windowing to disregard invalid weight taps that result from the inherent circular convolution therein, see explanations after (10),  $\epsilon > 0$  prevents division by zero or values near zero, and  $\mu_{i,j,m}$  is the frequency-selective step size for the *i*th channel, *j*th segment, and *m*th time block. The appropriate selection of the step size(s)  $\mu_{i,j,m}$  for *i* = 0, ..., *C* - 1, and *j* = 0, ..., *S* - 1, is important to guarantee stable and fast convergence of the adaptive echo canceller.

[0056] The canceller output at the  $m$ th time block can be represented as

$$\mathbf{E}_m = \mathbf{A}_m + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathbf{R}_{i,j,m}, \quad (54)$$

wherein the first term  $\mathbf{A}_m$  comprises all non-cancellable components - see (11) for definitions - and the second term corresponds to the residual echoes  $\mathbf{R}_{i,j,m}$  yet to be cancelled.

[0057] A theoretical framework to combat interferences present in the canceller training error is to selectively update the frequency-domain filter coefficients based on the magnitude of the interference signals. For instance, an optimal frequency-domain (FD) step size for a single filter was deduced in "A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain, Signal Processing 80, September 2000.

[0058] A novel application thereof in the context of the present invention is

$$\mu_{i,j,m} = E \left\{ |\mathbf{R}_{i,j,m}|^2 \right\} / E \left\{ |\mathbf{E}_m|^2 \right\}, \quad (55)$$

wherein  $E\{\}$  denotes statistical expectation. Another expression for an optimal FD step size found in "A frequency domain step-size control method for LMS algorithms," IEEE Signal Processing Letters 17, February 2010. A novel application thereof in the context of the present invention is

$$\mu_{i,j,m} = 1 - \sqrt{1 - E \left\{ |\mathbf{R}_{i,j,m}|^2 \right\} / E \left\{ |\mathbf{E}_m|^2 \right\}}. \quad (56)$$

[0059] Before proceeding further, it is worth mentioning that the standard update (9) is a particular case of the proposed update rule (51) when the frequency-domain step size is chosen as

$$\mu_{i,j,m} = \mu |\mathbf{X}_{i,j,m}|^2 / \left( \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |\mathbf{X}_{a,b,m}|^2 \right). \quad (57)$$

[0060] Said standard step size (57) neither acknowledges different misalignment levels on each transversal filter  $\mathbf{R}_{i,j,m}$  nor the presence of the interfering non-cancellable term  $\mathbf{A}_m$ . As result, the standard update exhibits slow and unstable convergence in most practical scenarios.

[0061] Computing the optimal frequency-domain step size as per (55) or (56) can be thus revamped into estimating the power spectral density of the interferences and the residual echoes. However, none of those  $C \times S + 1$  unknown terms are directly measurable. The common belief for an expert in the art is that the effective and accurate estimation of the square magnitude of said terms  $\mathbf{A}_m$  and  $\mathbf{R}_{i,j,m}$ , required in the evaluation of the optimal step size (55) or (56), is near to impossible.

[0062] We introduce hereby the following notation:  $A_m(k)$  represents the  $k$ th frequency bin of the frequency-domain vector  $\mathbf{A}_m$ ,  $E_m(k)$  represents the  $k$ th frequency bin of the frequency-domain vector  $\mathbf{E}_m$ , and so on. The  $i,j$ th residual echo can be written as

$$R_{i,j,m}(k) = G_{i,j}(k) X_{i,j,m}(k), \quad (58)$$

wherein  $G_{i,j}(k)$  denotes the misalignment in the  $i, j$ th filter. We can assume the following statistical relations

$$E \left\{ G_{i,j}^*(k) A_m(k) \right\} = 0, \quad (59)$$

$$E\{G_{i,j}^*(k)G_{a,b}(k)\} = \begin{cases} E\{|G_{i,j}(k)|^2\}, & a = i, b = j \\ 0, & \text{otherwise} \end{cases} \quad (60)$$

**[0063]** Based on the previous assumptions, the frequency-domain step as per (55) can be recast as follows

$$\mu_{i,j,m}(k) = \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}, \quad (61)$$

wherein  $\Phi_{i,j}(k)$  and  $\Psi_m(k)$  are the power spectral density of the misalignment of each  $i, j$ th filter and that of the non-cancellable term respectively

$$\Phi_{i,j}(k) = E\{|G_{i,j}(k)|^2\} \quad (62)$$

$$\Psi_m(k) = E\{|A_m(k)|^2\}, \quad (63)$$

and  $\mathcal{X}_{i,j,m}(k)$  is the expected power spectrum of the  $i, j$ th far-end signal at the  $m$ th block

$$\mathcal{X}_{i,j,m}(k) = E\{|X_{i,j,m}(k)|^2\}. \quad (64)$$

**[0064]** Said expected power spectrum (64) is preferably obtained from the power spectrum of a windowed signal block.

**[0065]** The method to accurately estimate said terms (62) and (63) disclosed in the present invention is carried out in 110 according to the reasoning and arguments exposed in what follows. It is known in the field of machine learning that the objective assessment of a learning machine, such as the set of adaptive filters of concern, must be performed on testing data statistically independent from the training data. This important axiom implies hereby that the weight update performed with the data from the  $m$ th time block can only be assessed with data from another time block, and never with the very  $m$ th block.

**[0066]** The immediately-preceding  $m - 1$ th time block can act as such a "testing" block, hence not incurring any delay because said block is already available. Moreover, because both testing and updating time blocks are consecutive in time, the actual echo path at each one is considered to be the same, which is an important requirement for the inventive method to perform best. Other blocks from different times, i. e., an  $m - \ell$ th time block, wherein  $\ell$  is an integer different than zero, could be used as long as the latter requirement is met. For ease of explanation, the  $m - 1$ th time block, e. g.  $\ell = 1$ , is used as an example in the following description.

**[0067]** Supported on the embodiments from Fig. 3 and Fig. 4, in a first stage two errors are computed, namely, the canceller output (e. g. error) over the  $m$ th time-block  $E_m$  in

$$E_m = Z_m - \mathcal{T}_2 \left\{ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j} \circ X_{i,j,m} \right\}, \quad (65)$$

wherein  $Z_m$  corresponds to the spectral samples of the microphone signal  $z(n)$  at the  $m$ th block, the operator  $\mathcal{T}_2$  { } corresponds to the time-domain windowing upon the filter operation to disregard invalid samples that result from the inherent circular convolution therein, see (8), (10) and (52) for explanations; and the "look-backward" error  $\bar{\varepsilon}_m$  in 120B, obtained as follows: the canceller weights are updated in 130 according to the classical frequency-domain adaptive filtering rule (9) with said  $m$ th output  $E_m$ , and the impact thereof is assessed

over the  $m - 1$ th block, that is,

$$\overleftarrow{\mathcal{E}} = \mathbf{Z}_{m-1} - \mathcal{T}_2 \left\{ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \left( \mathbf{W}_{i,j} + \lambda \overleftarrow{\Omega}_{i,j} \right) \circ \mathbf{X}_{i,j,m-1} \right\}, \quad (66)$$

wherein  $\mathbf{Z}_{m-1}$  corresponds to the spectral samples of the microphone signal  $z(n)$  at the  $m - 1$ th block, and  $\overleftarrow{\Omega}_{i,j}$  is the  $i, j$ th look-backward weight update

$$\overleftarrow{\Omega}_{i,j} = \mathcal{T}_1 \left\{ \mathbf{E}_m \circ \mathbf{X}_{i,j,m}^* / \left( \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |\mathbf{X}_{a,b,m}|^2 + \epsilon \right) \right\}. \quad (67)$$

**[0068]** The constant  $\lambda$  in (66) is the "overshoot" factor, to be addressed later.

**[0069]** A positive outcome from the previous operations is namely that the non-cancellable term  $\mathbf{A}_m$  can not only be observed in the  $m$ th error (65), as outlined in (54), but also in the look-backward error (66). However, the non-cancellable term  $\mathbf{A}_{m-1}$  at the  $m - 1$ th time block unfortunately appears in the look-backward error (66) as new unknown in the problem. This situation invites us to compute two additional errors, this time by swapping the training and testing time blocks. Hence, the  $m - 1$ th error  $\mathcal{E}$  is computed as

$$\mathcal{E} = \mathbf{Z}_{m-1} - \mathcal{T}_2 \left\{ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathbf{W}_{i,j} \circ \mathbf{X}_{i,j,m-1} \right\}, \quad (68)$$

and the "look-forward" error  $\overrightarrow{\mathcal{E}}$  as

$$\overrightarrow{\mathcal{E}} = \mathbf{Z}_m - \mathcal{T}_2 \left\{ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \left( \mathbf{W}_{i,j} + \lambda \overrightarrow{\Omega}_{i,j} \right) \circ \mathbf{X}_{i,j,m} \right\}, \quad (69)$$

wherein  $\overrightarrow{\Omega}_{i,j}$  is the  $i, j$ th look-forward weight update

$$\overrightarrow{\Omega}_{i,j} = \mathcal{T}_1 \left\{ \mathcal{E} \circ \mathbf{X}_{i,j,m-1}^* / \left( \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |\mathbf{X}_{a,b,m-1}|^2 + \epsilon \right) \right\}. \quad (70)$$

**[0070]** The look-backward error (66) and look-forward error (69) can be also obtained in a compact fashion as

$$\overleftarrow{\mathcal{E}} = \mathcal{E} - \lambda \overleftarrow{\Delta} \quad (71)$$

$$\overrightarrow{\mathcal{E}} = \mathbf{E}_m - \lambda \overrightarrow{\Delta}, \quad (72)$$

wherein  $\overleftarrow{\Delta}$  and  $\overrightarrow{\Delta}$  are the impact of the look-backward and look-forward update respectively

$$\overleftarrow{\Delta} = \mathcal{T}_2 \left\{ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \overleftarrow{\Omega}_{i,j} \circ \mathbf{X}_{i,j,m-1} \right\} \quad (73)$$

$$\overrightarrow{\Delta} = \mathcal{T}_2 \left\{ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \overrightarrow{\Omega}_{i,j} \circ \mathbf{X}_{i,j,m} \right\}. \quad (74)$$

**[0071]** In the prior art "Method and apparatus for updating filter coefficients of an adaptive echo canceller," EP 2930917 B1, disclosed by the author of the present invention, the evaluation of only three errors was thought sufficient to estimate the terms (62) and (63) for a single-channel single-segment echo canceller. However, such a strategy leads often to ill-conditioned problems because the presence of the unknown terms in the problem model is unbalanced. Including both the look-backward (77) and the look-forward (78) errors yields a balanced well-conditioned estimation problem.

**[0072]** In order to get the equations that relate said four errors  $E_m$ ,  $\varepsilon$ ,  $\overleftarrow{\varepsilon}$ , and  $\overrightarrow{\varepsilon}$  with the unknown terms required to evaluate the optimal step size(s), said spectral errors (65) and (66) can be recast in term of the filter misalignment  $G_{i,j}(k)$  as follows

$$E_m(k) \equiv A_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} G_{i,j}(k) X_{i,j,m}(k), \quad (75)$$

$$\mathcal{E}(k) \equiv A_{m-1}(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} G_{i,j}(k) X_{i,j,m-1}(k), \quad (76)$$

$$\overleftarrow{\mathcal{E}}(k) \equiv A_{m-1}(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} G_{i,j}(k) X_{i,j,m-1}(k) - \lambda \overleftarrow{\Delta}(k), \quad (77)$$

$$\overrightarrow{\mathcal{E}}(k) \equiv A_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} G_{i,j}(k) X_{i,j,m}(k) - \lambda \overrightarrow{\Delta}(k), \quad (78)$$

wherein  $\equiv$  denotes equivalency.

**[0073]** Translating the previous equation to power spectral densities finally leads to one of the main advantages of the invention. Supported on the assumptions of uncorrelation (59) and (60) between the terms involved, it can be concluded

$$E \left\{ |E_m(k)|^2 \right\} = \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j,m}(k) \quad (79)$$

$$E \left\{ |\mathcal{E}(k)|^2 \right\} = Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-1}(k) \Phi_{i,j,m}(k), \quad (80)$$

wherein term  $Y(k) = E \{|A_{m-1}(k)|^2\}$  is unknown and not used in the evaluation of the optimal step size(s) (61).

**[0074]** In order to translate the look-backward (77) and look-forward (78) errors to power spectral densities, the effect of the overshoot factor  $\lambda$  needs to be first clarified. It can be deduced that the reduction of the filter misalignment (62)

is ruled by the power factor

$$f(\lambda) = \exp(-2\lambda + \lambda^2). \quad (81)$$

**[0075]** On the other hand, the update equations (77) and (78) yield the amplification of the non-cancellable terms  $\mathbf{A}_m$  and  $\mathbf{A}_{m-1}$  proportionally to  $A$ , that is, its power (63) gets amplified by

$$g(\lambda) = \lambda^2. \quad (82)$$

**[0076]** The illustration of this new tradeoff is brought in Fig. 5, wherein the solid line corresponds to (81), and the dashed line to (82). An attractive situation appears when  $\lambda$  is set to two ( $\lambda = 2$ ): in this "overshoot" case, the expected misalignment level remains the same,  $f(2) = 1$ , but the power of the non-cancellable term is amplified by a large factor, e. g.  $g(2) = 4$ .

**[0077]** Such a value,  $\lambda = 2$ , has never been used as step size in adaptive filtering because the resulting adaptive system would not converge. However, in the scope of this invention, said overshoot off-line update used in (77) and (78) makes the interference term  $\mathbf{A}_m$ , as well as  $\mathbf{A}_{m-1}$ , appear 6 dB larger in the look-backward error and look-forward error respectively. Henceforth, the value  $\lambda = 2$  is adopted in 130, but other  $\lambda$  values can also be used as long as rules (81) and (82) are observed.

**[0078]** The power spectral densities of the set of spectral errors (77) and (78) can be thus written as

$$E\left\{|\overleftarrow{\mathcal{E}}(k)|^2\right\} = Y(k) + 4\overleftarrow{T}(k)\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \Phi_{i,j}(k) \mathcal{X}_{i,j,m-1}(k) \quad (83)$$

$$E\left\{|\overrightarrow{\mathcal{E}}(k)|^2\right\} = \Psi_m(k) + 4\overrightarrow{T}(k)Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k), \quad (84)$$

wherein  $\overleftarrow{T}(k)$  and  $\overrightarrow{T}(k)$  are the normalized update in the look-backward and look-forward errors respectively

$$\overleftarrow{T}(k) = |\overleftarrow{\Delta}(k)|^2 / |E_m(k)|^2 \quad (85)$$

$$\overrightarrow{T}(k) = |\overrightarrow{\Delta}(k)|^2 / |\mathcal{E}(k)|^2. \quad (86)$$

**[0079]** Said four equations (79), (80), (83), and (84) are used to obtain the terms (62) and (63), which are required in the evaluation of the optimal frequency-domain step size(s). Said equations can be written as a system of linear equations subject to inequality constraints. Without loss of generality, we exemplify the solution system over a stereo scenario,  $C = 2$ , with one segment per channel,  $S = 1$ , in what follows

$$\begin{bmatrix} 1 & 0 & \mathcal{X}_{0,0,m-1}(k) & \mathcal{X}_{1,0,m-1}(k) \\ 0 & 1 & \mathcal{X}_{0,0,m}(k) & \mathcal{X}_{1,0,m}(k) \\ 1 & 4\overleftarrow{T}(k) & \mathcal{X}_{0,0,m-1}(k) & \mathcal{X}_{1,0,m-1}(k) \\ 4\overrightarrow{T}(k) & 1 & \mathcal{X}_{0,0,m}(k) & \mathcal{X}_{1,0,m}(k) \end{bmatrix} \begin{bmatrix} Y(k) \\ \Psi_m(k) \\ \Phi_{0,0}(k) \\ \Phi_{1,0}(k) \end{bmatrix} = \begin{bmatrix} |\mathcal{E}(k)|^2 \\ |E_m(k)|^2 \\ |\overleftarrow{\mathcal{E}}(k)|^2 \\ |\overrightarrow{\mathcal{E}}(k)|^2 \end{bmatrix}$$

subject to

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \Phi_{1,0}(k) > 0. \quad (87)$$

5 **[0080]** Said inverse linear problem with inequality constraints (87) requires an iterative methodology, such as linear programming. In order to speed up the process (to reduce the number of iterations), a good initial guess is welcome. The very solution obtained at the previous  $m - 1$ th time block can serve as said initialization. The solution to (87) is computed in 150.

10 **[0081]** At this point it is important to remark that the  $m - \ell$ th time block precedes the  $m$ th one, e. g.  $\ell > 0$  and preferably  $\ell = 1$ . Adopting a negative value,  $\ell < 0$ , would incur an undesired processing delay, as the  $m - \ell$ th time block would be yet to come after the  $m$ th one. However, that option does not actually differ in essence from said preferred case because both time blocks serve as training and testing data on any of the look-forward/- backward errors: because of this symmetry in the solution system (87), the option  $\ell < 0$  becomes moot.

15 **[0082]** It is important to point out that any sub-matrix in (87) involving only  $\Psi_m(k)$  and  $Y(k)$  has full rank, e. g. the solution obtained for the non-cancellable terms is accurate. This proves theoretically that the resulting step size provides "instantaneous" robustness against double-talk, which was one of the major challenges yet to be effectively solved by the state of the art.

20 **[0083]** On the other hand, the estimation of the filter misalignment becomes ill-conditioned as the number of filters increases. In the practice, the solution accomplished with a normal initialization, such as that from the  $m - 1$ th block, outperforms the state of art. Nonetheless, in order to improve the condition of this problem when the number of channels or the number of segments per channel is large, a second embodiment of the invention is disclosed in what follows.

25 **[0084]** Improving the estimation accuracy for the misalignment power requires evaluating at most two additional errors, and including the resulting linear equations into the original system. The strategy behind this methodology is namely not to treat said  $r$ , sth filter as part of the adaptive system, e. g. to consider its misalignment as a non-cancellable "interference" term, and to update off-line all canceller filters but the  $r$ , sth one, resulting in the additional look-backward error

$$\overleftarrow{\varepsilon}_{r,s} = \varepsilon - \lambda \overleftarrow{\Delta}_{r,s} \quad (88)$$

wherein  $\overleftarrow{\Delta}_{r,s}$  is the impact of a look-backward filter update except for the  $(r,s)$ th filter

$$\overleftarrow{\Delta}_{r,s} = \mathcal{T}_2 \left\{ \sum_i^{i \neq r} \sum_j^{j \neq s} \mathcal{T}_1 \left\{ \mathbf{E}_m \circ \mathbf{X}_{i,j,m}^* / \left( \sum_a^{a \neq r} \sum_b^{b \neq s} |\mathbf{X}_{a,b,m}|^2 \right) \right\} \circ \mathbf{X}_{i,j,m-1} \right\}, \quad (89)$$

and the additional look-forward error

$$\overrightarrow{\varepsilon}_{r,s} = \varepsilon - \lambda \overrightarrow{\Delta}_{r,s} \quad (90)$$

45 wherein  $\overrightarrow{\Delta}_{r,s}$  is the impact of a look-forward filter update except for the  $(r,s)$ th filter

$$\overrightarrow{\Delta}_{r,s} = \mathcal{T}_2 \left\{ \sum_i^{i \neq r} \sum_j^{j \neq s} \mathcal{T}_1 \left\{ \varepsilon \circ \mathbf{X}_{i,j,m-1}^* / \left( \sum_a^{a \neq r} \sum_b^{b \neq s} |\mathbf{X}_{a,b,m-1}|^2 \right) \right\} \circ \mathbf{X}_{i,j,m} \right\}. \quad (91)$$

**[0085]** The additional look-backward error  $\overleftarrow{\varepsilon}_{r,s}$  and look-forward error  $\overrightarrow{\varepsilon}_{r,s}$  for  $\lambda = 2$ , bring two important additional relations

55

$$\begin{aligned}
 E\{|\overleftarrow{\mathcal{E}}_{r,s}(k)|^2\} &= \Psi_{m-1}(k) + 4\overleftarrow{T}_{r,s}(k)\Psi_m(k) + 4\overleftarrow{T}_{r,s}(k)\Phi_{r,s}(k)\mathcal{X}_{r,s,m}(k) \\
 &+ \sum_{i=0}^{C-1S-1} \sum_{j=0}^{C-1S-1} \Phi_{i,j}(k)\mathcal{X}_{i,j,m-1}(k)
 \end{aligned} \tag{92}$$

$$+ \sum_{i=0}^{C-1S-1} \sum_{j=0}^{C-1S-1} \Phi_{i,j}(k)\mathcal{X}_{i,j,m}(k), \tag{93}$$

wherein

$$\overleftarrow{T}_{r,s}(k) = |\overleftarrow{\Delta}_{r,s}(k)|^2 / |E_m(k)|^2 \tag{94}$$

$$\overrightarrow{T}_{r,s}(k) = |\overrightarrow{\Delta}_{r,s}(k)|^2 / |\mathcal{E}(k)|^2. \tag{95}$$

**[0086]** Said four equations (79), (80), (83), (84), in addition to (92) and (93), embody the main algorithmic foundations to obtain the terms (62) and (63). For the sake of simplicity let us consider a one-channel system,  $C = 1$ , composed of two segments,  $S = 2$ . Since the filter  $W_{0,0}$  of the first segment takes care of the early echoes, which are prone to change and have larger energy than the late ones, it is desired to improve the tracking performance for said filter. For the sake of simplicity, we drop the frequency bin index  $k$  whenever convenient, that is  $\mathcal{X}_{i,j,m}$  denotes  $\mathcal{X}_{i,j,m}(k)$ , and so on, the final solution system results as follows

$$\begin{bmatrix}
 1 & 0 & \mathcal{X}_{0,0,m-1} & \mathcal{X}_{0,1,m-1} \\
 0 & 1 & \mathcal{X}_{0,0,m} & \mathcal{X}_{0,1,m} \\
 1 & 4\overleftarrow{T} & \mathcal{X}_{0,0,m-1} & \mathcal{X}_{0,1,m-1} \\
 4\overrightarrow{T} & 1 & \mathcal{X}_{0,0,m} & \mathcal{X}_{0,1,m} \\
 1 & 4\overleftarrow{T}_{0,0} & 4\overleftarrow{T}_{0,0}\mathcal{X}_{0,0,m} + \mathcal{X}_{0,0,m-1} & \mathcal{X}_{0,1,m-1} \\
 4\overrightarrow{T}_{0,0} & 1 & 4\overrightarrow{T}_{0,0}\mathcal{X}_{0,0}^{m-1} + \mathcal{X}_{0,0,m} & \mathcal{X}_{0,1,m}
 \end{bmatrix}
 \begin{bmatrix}
 Y \\
 \Psi_m \\
 \Phi_{0,0} \\
 \Phi_{0,1}
 \end{bmatrix}
 =
 \begin{bmatrix}
 |\mathcal{E}|^2 \\
 |E_m|^2 \\
 |\overleftarrow{\mathcal{E}}|^2 \\
 |\overrightarrow{\mathcal{E}}|^2 \\
 |\overleftarrow{\mathcal{E}}_{0,0}|^2 \\
 |\overrightarrow{\mathcal{E}}_{0,0}|^2
 \end{bmatrix}$$

subject to

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \Phi_{0,1}(k) > 0. \tag{96}$$

**[0087]** The misalignment  $\Phi_{0,0}(k)$  is observed in the look-backward/-forward errors  $\overleftarrow{\varepsilon}_{0,0}(k)$  and  $\overrightarrow{\varepsilon}_{0,0}(k)$  with a gain of 6 dB. It is evident for anyone familiar with the art that the inclusion of said two additional equations, that is, the last two rows in (96), improves the condition of the problem.

**[0088]** The solution to (96) is computed in 150. Upon said solution, the optimal FD step size as per (56) is evaluated in 180 as follows

$$\mu_{i,j,m}(k) = 1 - \sqrt{1 - \frac{\Phi_{i,j,m}(k)\mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b,m}(k)\mathcal{X}_{a,b,m}(k)}}}. \quad (97)$$

### Residual Echo Suppression

**[0089]** Due to very long echo paths, to an adaptive system in early convergence, or to non-linear phenomena, echo components in the canceller output (11) may still be audible. The non-linear processing (NLP) 300 aims to suppress the residual echo components still present in the canceller output to inaudible levels while preserving the integrity of the main near-end sounds as much as possible. In order to perform an effective NLP, accurate estimation of the residual echo in level and spectrum is important.

**[0090]** The residual echo present in the canceller output (11) is composed of the terms  $R_m(k)$ ,  $Q_m(k)$  and  $P_m(k)$ . As said terms are uncorrelated with each other, the estimated power spectrum of the residual echo is built additively in 330 as follows

$$\Lambda_m(k) = \mathcal{R}_m(k) + \mathcal{Q}_m(k) + \mathcal{P}_m(k), \quad (98)$$

wherein  $\mathcal{R}_m(k)$ ,  $\mathcal{Q}_m(k)$ , and  $\mathcal{P}_m(k)$  are the estimated power spectrum of every term in the residual echo respectively.

**[0091]** The power spectrum of the residual echo due to filter misalignment is readily given by

$$\mathcal{R}_m(k) = E\{|R_m(k)|^2\} = \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \Phi_{i,j}(k)\mathcal{X}_{i,j,m}(k), \quad (99)$$

wherein  $\Phi_{i,j}(k)$  is outcome of the generic decomposition problem (96).

**[0092]** The power spectrum of the residual echo due to long acoustic echo paths ( $L_H \gg S \times L$ ) can be defined as follows

$$\mathcal{Q}_m(k) = E\{|Q_m(k)|^2\} = \sum_{i=0}^{C-1} \sum_{j=S}^{\infty} |W_{i,j}(k)|^2 \mathcal{X}_{i,j,m}(k), \quad (100)$$

wherein  $W_{i,j}(k)$  for  $j = S, \dots, \infty$  represent filters beyond the time scope of the canceller, hence unknown "virtual" filters. Said terms in (100) can be approximated according to the reasoning that follows.

**[0093]** Echo paths present typically an exponential decay, especially in the last part of its impulse response. The time-domain filter  $w_{i,S-1}(n)$ , regardless of the channel, is thus assumed to follow the model

$$E\{\log|w_{i,S-1}(n)|^2\} = -\tau n + a, \quad (101)$$

wherein  $\tau$  is the time constant related to said filter, and the level  $a$  is not relevant henceforth; the time constant  $\tau$  can be simply obtained with linear regression over said square filter weights in logarithm scale, as pointed out in (101). The power spectrum of said "virtual" filters in (100) can be thus approximated as

$$|W_{i,S-1+\alpha}(k)|^2 \simeq \gamma^\alpha |W_{i,S-1}(k)|^2, \quad (102)$$

wherein  $W_{i,S-1}(k)$  is an existing filter,  $\alpha \geq 1$  is an integer, and  $\gamma = \exp(-\tau L)$ . By replacing (102) into (100), and because

$X_{i,j,m}(k) = X_{i,j-1,m-1}(k)$ , the power spectrum of the residual echo beyond the canceller scope is obtained with the following accumulative formula

$$Q_m(k) = \gamma Q_{m-1}(k) + \gamma \sum_{i=0}^{C-1} |W_{i,S-1}(k)|^2 X_{i,S-1,m-1}(k), \quad (103)$$

wherein all elements involved in said formula are available.

[0094] Finally, the non-linear echo components are correlated to the input signals  $x_i(n)$  somewhat in an unpredictable way. This means that the adaptive filter usually jitters or fluctuates when non-linear echo is present in the training error. Based on the misalignment  $\Phi_{i,j}(k)$  and step size  $\mu_{i,j,m}(k)$  obtained from the decomposition problem (96), the non-linear residual echo is readily obtained as follows

$$P_m(k) \propto \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mu_{i,j,m}(k) \Phi_{i,j}(k) X_{i,j,m}(k), \quad (104)$$

wherein  $\propto$  denotes proportional to.

[0095] Upon the estimated overall residual echo  $\Lambda_m(k)$ , several ways to carry out the NLP operation by 300 are available, such as with the Wiener filter, minimum mean-square error, perceptual enhancement, et cetera. As example, the Wiener filter can be approximated as follows

$$B_m(k) = \left[ \frac{|E_m(k)|^2 - \Lambda_m(k)}{|E_m(k)|^2} \right]_{\varepsilon}, \quad (105)$$

wherein  $[\ ]_{\varepsilon}$  denotes low clipping by  $\varepsilon > 0$ . Finally, the frequency-domain output of the acoustic echo control system is obtained by

$$O_m(k) = E_m(k) B_m(k). \quad (106)$$

[0096] In order to deliver a listening signal, the frequency-domain output  $O_m(k)$  is transformed to the time domain. Standard overlap-add processing must be considered to avoid block-effects in the resulting output. Said processing is well known to any person familiar with the art.

### Comfort Noise Injection

[0097] The echo cancellation 200 along with the NLP operation 300 delivers an echo-free output (first objective in echo control), that is

$$O_m(k) \simeq D_m(k) B_m(k) + V_m(k) B_m(k). \quad (107)$$

[0098] Since the magnitude of the residual echo is significantly lower than that of the main near-end signal, hence  $D_m(k) B_m(k) \simeq D_m(k)$ , the main near-end signal goes through nearly undistorted upon the NLP (second objective in echo control). We can thus approximate the previous output (107) as

$$O_m(k) \simeq D_m(k) + V_m(k) B_m(k). \quad (108)$$

[0099] However, the near-end background noise in (108), e. g.  $V_m(k) B_m(k)$ , suffers fluctuations in level and spectrum. In order to replenish the background noise in 400 as per (13) or equivalent technique and deliver the same perceptual impression (third and last objective in echo control), accurate estimation thereof in level and spectrum is important.

**[0100]** In case of far-end inactivity (an event that is straightforward to detect), hence  $B_m(k) = 1$ , the background noise can be estimated from the power spectrum  $|O_m(k)|^2$  during (short) periods of silence. However, the injection of comfort noise is actually required in the opposite case, that is, during far-end activity, hence  $0 \leq B_m(k) < 1$ . Therefore, being able to track the background noise on any situation, hence from the original canceller output  $|E_m(k)|^2$ , represents the main challenge to address hereby.

**[0101]** The usual approach to background noise estimation is to obtain the power spectral floor from the most recent  $K$  canceller outputs

$$\Gamma_m(k) = \left\{ |E_m(k)|^2, |E_{m-1}(k)|^2, \dots, |E_{m-K+1}(k)|^2 \right\} \quad (109)$$

either by averaging selected items within  $\Gamma_m$ , or with minimum statistics over the set  $\Gamma_m(k)$ , et cetera. Let  $\mathcal{N}\{\Gamma_m(k)\}$  denote a generic method to obtain said spectral floor from the sets  $\Gamma_m(k)$ .

**[0102]** The present invention proposes to estimate the power spectrum of the background noise in 430 according to

$$\Omega_m(k) = \left[ \mathcal{N}\{\Gamma_m(k)\} - \mathcal{N}\{\mathcal{L}_m(k)\} \right]_{\delta(k)} \quad (110)$$

wherein  $\delta(k)$  represents the minimum comfort noise for the  $k$ th frequency bin, and  $\mathcal{L}_m(k)$  contains the  $K$  most recent time spectral snapshots of the residual echo power, obtained with (98) during the NLP phase

$$\mathcal{L}_m(k) = \left\{ \Lambda_m(k), \Lambda_{m-1}(k), \dots, \Lambda_{m-K+1}(k) \right\}. \quad (111)$$

**[0103]** The proposed rule (110) involves the computation of a spectral floor "two times", once for the set  $\Gamma_m(k)$  and then for  $\mathcal{L}_m(k)$ . Although it demands larger computational complexity, it delivers accurate results. Evaluating only one spectral floor over the difference set  $\Gamma_m(k) - \mathcal{L}_m(k)$  does not lead in general to the same result, that is

$$\left[ \mathcal{N}\{\Gamma_m(k)\} - \mathcal{N}\{\mathcal{L}_m(k)\} \right]_{\delta(k)} \neq \mathcal{N}\left\{ \left[ \Gamma_m(k) - \mathcal{L}_m(k) \right]_{\delta(k)} \right\}. \quad (112)$$

**[0104]** The invention is not restricted to the specific embodiments described in detail herein, but encompasses all variants, combinations and modifications thereof that fall within the framework of the appended claims.

### Claims

1. A computer-implemented method for updating at least one frequency-domain filter coefficient  $W_{i,j}(k)$  of an echo canceller (50) having at least one channel and at least one segment per channel, the filter coefficients of the echo canceller (50) being updateable in the frequency domain at a time block  $m$ , **characterized by:**

determining a canceller output  $E_m(k)$  over the  $m$ th time block as the difference between  $Z_m(k)$  and a cancelling term based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m}(k);$$

determining a canceller error  $\varepsilon(k)$  over the  $m - \ell$ th time block as the difference between  $Z_{m-\ell}(k)$  and a cancelling

term based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m-\ell}(k);$$

determining a look-backward error  $\overleftarrow{\varepsilon}(k)$  as

$$\overleftarrow{\mathcal{E}}(k) = \mathcal{E}(k) - \lambda \overleftarrow{\Delta}(k),$$

wherein  $\overleftarrow{\Delta}(k)$  is based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k);$$

determining a look-forward error  $\overrightarrow{\varepsilon}(k)$  as

$$\overrightarrow{\mathcal{E}}(k) = E_m(k) - \lambda \overrightarrow{\Delta}(k),$$

wherein  $\overrightarrow{\Delta}(k)$  is based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k);$$

determining an optimal update step size  $\mu_{i,j,m}(k)$  from said canceller output  $E_m(k)$  over the  $m$ th time block, from said canceller error  $\varepsilon(k)$  over the  $m - \ell$ th time block, from said look-backward error  $\overleftarrow{\varepsilon}(k)$ , and from said look-forward error  $\overrightarrow{\varepsilon}(k)$ ; and

updating said at least one filter coefficient  $W_{i,j}(k)$  by using said optimal update step-size  $\mu_{i,j,m}(k)$ ; wherein

$C$  corresponds to the number of channels  $i = 0, \dots, C - 1$  of the echo canceller (50),

$S$  corresponds to the number of segments  $j = 0, \dots, S - 1$  per channel of the echo canceller (50),

$\ell$  is an integer different than zero,

$k$  denotes a frequency bin index,

$\epsilon$  is a positive value,

$Z_m(k)$  corresponds to the  $k$ th spectral bin of a microphone signal at the  $m$ th time block,

$Z_{m-\ell}(k)$  corresponds to the  $k$ th spectral bin of the microphone signal at the  $m - \ell$ th time block,

$X_{i,j,m}(k)$  corresponds to the  $k$ th spectral bin of a far-end signal at the  $i$ th channel,  $j$ th segment, and  $m$ th time block,

$X_{i,j,m-\ell}(k)$  corresponds to the  $k$ th spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m - \ell$ th time block,

$\lambda$  is an overshoot factor, and

$a, b$  are channel and segment indices, respectively.

2. The method according to claim 1, wherein the optimal step size  $\mu_{i,j,m}(k)$  is computed by either

$$\mu_{i,j,m}(k) = \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}$$

or

$$\mu_{i,j,m}(k) = 1 - \sqrt{1 - \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}}$$

wherein  $\Psi_m(k)$  is a power spectrum of the non-cancellable components at the  $m$ th time block and  $\Phi_{i,j}(k)$  is a power spectrum of the misalignment of each  $i, j$ th filter, wherein  $\Psi_m(k)$  and  $\Phi_{i,j}(k)$  are determined by solving the set of linear equations

$$\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |E_m(k)|^2$$

$$Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\mathcal{E}(k)|^2$$

$$Y(k) + 4\overleftarrow{T}(k)\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}(k)|^2$$

$$4\overrightarrow{T}(k)Y(k) + \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}(k)|^2$$

subject to inequality constraints

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0;$$

wherein

$\mathcal{X}_{i,j,m}(k)$  corresponds to the  $k$ th power spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m$ th time block, wherein said signal block is preferably windowed,

$\chi_{i,j,m-\ell}(k)$  corresponds to the  $k$ th power spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m - \ell$ th time block, wherein said signal block is preferably windowed,

$Y(k)$  corresponds to the  $k$ th power spectral bin of the non-cancellable components at the  $m - \ell$ th block,

$\overleftarrow{T}(k)$  is the normalized look-backward update, computed as

$$\overleftarrow{T}(k) = \overleftarrow{\Delta}(k) / |E_m(k)|^2,$$

and

$\vec{T}(k)$  is the normalized look-forward update, computed as

$$\vec{T}(k) = \vec{\Delta}(k) / |\mathcal{E}(k)|^2.$$

3. The method according to claim 1, wherein at least one of the number of channels and the number of segments per channel is larger than one; and the optimal step size  $\mu_{i,j,m}(k)$  is determined from said canceller output  $E_m(k)$  over the  $m$ th time block, from said canceller error  $\mathcal{E}(k)$  over the  $m - \ell$ th time block, from said look-backward error  $\overleftarrow{\mathcal{E}}(k)$ , from said look-forward error  $\overrightarrow{\mathcal{E}}(k)$  as well as from at least one additional look-backward error  $\overleftarrow{\mathcal{E}}_{r,s}(k)$  and from at least one additional look-forward error  $\overrightarrow{\mathcal{E}}_{r,s}(k)$  for at least one  $r, s$ th filter,  $r$  and  $s$  being indices of the channel and the segment, respectively, wherein

$$\overleftarrow{\mathcal{E}}_{r,s}(k) = \mathcal{E}(k) - \lambda \overleftarrow{\Delta}_{r,s}(k),$$

wherein  $\overleftarrow{\Delta}_{r,s}(k)$  is based on

$$\sum_i^{i \neq r} \sum_j^{j \neq s} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_a^{a \neq r} \sum_b^{b \neq s} |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k),$$

and

$$\overrightarrow{\mathcal{E}}_{r,s}(k) = E_m(k) - \lambda \overrightarrow{\Delta}_{r,s}(k),$$

wherein  $\overrightarrow{\Delta}_{r,s}(k)$  is based on

$$\sum_i^{i \neq r} \sum_j^{j \neq s} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_a^{a \neq r} \sum_b^{b \neq s} |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k).$$

4. The method according to claims 2 and 3, wherein the power spectrum  $\Psi_m(k)$  of the non-cancellable components and the power spectrum  $\Phi_{i,j}(k)$  of the misalignment of each  $i, j$ th filter are determined by solving said set of linear equations with additional linear equations for the at least one  $r, s$ th filter

$$\begin{aligned} Y(k) + 4 \overleftarrow{T}_{r,s}(k) \Psi_m(k) + 4 \overleftarrow{T}_{r,s}(k) \mathcal{X}_{r,s,m}(k) \Phi_{r,s}(k) + \\ + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-1}(k) \Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}_{r,s}(k)|^2 \end{aligned}$$

$$\Psi_m(k) + 4\overrightarrow{T}_{r,s}(k)Y(k) + 4\overrightarrow{T}_{r,s}(k)\mathcal{X}_{r,s,m-1}(k)\Phi_{r,s}(k) +$$

$$+ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k)\Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}_{r,s}(k)|^2$$

subject to inequality constraints

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0;$$

wherein

$\overleftarrow{T}_{r,s}(k)$  is the normalized look-backward update excluding the  $r$ ,  $s$ th filter, computed as

$$\overleftarrow{T}_{r,s}(k) = \overleftarrow{\Delta}_{r,s}(k) / |E_m(k)|^2,$$

and

$\overrightarrow{T}_{r,s}(k)$  is the normalized look-forward update excluding the  $r$ ,  $s$ th filter, computed as

$$\overrightarrow{T}_{r,s}(k) = \overrightarrow{\Delta}_{r,s}(k) / |\mathcal{E}(k)|^2.$$

5. The method according to any one of the claims 1 to 4, wherein the overshoot factor  $\lambda$  is in the range  $1 < \lambda \leq 2$ , and preferably equals 2.
6. The method according to any one of the claims 1 to 5, **characterized by** determining a power spectrum  $\mathcal{R}_m(k)$  of the residual echo due to filter misalignment at the  $m$ th block by means of

$$\mathcal{R}_m(k) = \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \Phi_{i,j}(k)\mathcal{X}_{i,j,m}(k),$$

and subsequently an improved canceller output  $O_m(k)$  based on the canceller output  $E_m(k)$  and on said determined power spectrum  $\mathcal{R}_m(k)$  of the residual echo due to the filter misalignment.

7. The method according to any one of the claims 1 to 6, **characterized by** determining a power spectrum  $\mathcal{Q}_m(k)$  of the echo beyond the canceller's time scope at the  $m$ th block by means of the accumulative method

$$\mathcal{Q}_m(k) = \gamma \mathcal{Q}_{m-1}(k) + \gamma \sum_{i=0}^{C-1} |W_{i,S-1}(k)|^2 \mathcal{X}_{i,S-1,m-1}(k),$$

wherein  $0 < \gamma < 1$ ,

and subsequently an improved canceller output  $O_m(k)$  based on the canceller output  $E_m(k)$  and on said determined power spectrum  $\mathcal{Q}_m(k)$  of the echo beyond the canceller's time scope.

8. The method according to any one of the claims 1 to 7, **characterized by** determining a power spectrum  $\mathcal{P}_m(k)$  of the non-linear echo components at the  $m$ th block by means of

$$\mathcal{P}_m(k) \propto \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mu_{i,j,m}(k) \Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k),$$

5 wherein  $\propto$  denotes proportional,  
and subsequently an improved canceller output  $O_m(k)$  based on the canceller output  $E_m(k)$  and on said determined  
power spectrum  $\mathcal{P}_m(k)$  of the non-linear echo components.

- 10 **9.** The method according to any one of the claims 1 to 8, wherein the method determines an improved canceller output  $O_m(k)$  as

$$O_m(k) = E_m(k) B_m(k)$$

$$B_m(k) = \left[ \frac{|E_m(k)|^2 - \Lambda_m(k)}{|E_m(k)|^2} \right]_{\varepsilon}$$

$$\Lambda_m(k) = \mathcal{R}_m(k) + \mathcal{Q}_m(k) + \mathcal{P}_m(k),$$

25 wherein

$\lfloor \cdot \rfloor_{\varepsilon}$  denotes low clipping by  $\varepsilon > 0$ ,

$E_m(k)$  is the canceller output,

$\mathcal{R}_m(k)$  is a power spectrum of the residual echo due to filter mis-alignment at the  $m$ th block,

$\mathcal{Q}_m(k)$  is a power spectrum of the echo beyond the canceller's time scope at the  $m$ th block, and

$\mathcal{P}_m(k)$  is a power spectrum of the non-linear echo components at the  $m$ th block.

- 35 **10.** The method according to any one of claims 1 to 9, wherein the at least one filter coefficient  $W_{i,j}(k)$  is updated based on

$$W_{i,j}^{new}(k) = W_{i,j}(k) + \mu_{i,j,m}(k) \frac{E_m(k) X_{i,j,m}^*(k)}{|X_{i,j,m}(k)|^2 + \varepsilon}.$$

- 40 **11.** An echo canceller (50) having at least one channel and at least one segment per channel, the echo canceller (50) having filter coefficients  $W_{i,j}(k)$  being updateable in the frequency domain at a time block  $m$ ,  
45 **characterised in that** the echo canceller is configured to:

determine a canceller output  $E_m(k)$  over the  $m$ th time block as the difference between  $Z_m(k)$  and a cancelling term based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m}(k);$$

55 determine a canceller error  $\varepsilon(k)$  over the  $m - \ell$ th time block as the difference between  $Z_{m-\ell}(k)$  and a cancelling term based on

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m-\ell}(k);$$

5 determine a look-backward error  $\overleftarrow{\varepsilon}(k)$  as

$$10 \quad \overleftarrow{\mathcal{E}}(k) = \mathcal{E}(k) - \lambda \overleftarrow{\Delta}(k)$$

wherein  $\overleftarrow{\Delta}(k)$  is based on

$$15 \quad \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k);$$

20 determine a look-forward error  $\overrightarrow{\varepsilon}(k)$  as

$$25 \quad \overrightarrow{\mathcal{E}}(k) = E_m(k) - \lambda \overrightarrow{\Delta}(k)$$

wherein  $\overrightarrow{\Delta}(k)$  is based on

$$30 \quad \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k);$$

35 determine an optimal update step-size  $\mu_{i,j,m}(k)$  from said canceller output  $E_m(k)$  over the  $m$ th time block, from said canceller error  $\varepsilon(k)$  over the  $m - \ell$ th time block, from said look-backward error  $\overleftarrow{\varepsilon}(k)$ , and from said look-forward error  $\overrightarrow{\varepsilon}(k)$ ; and  
 update said at least one filter coefficient  $W_{i,j}(k)$  by using said optimal update step-size  $\mu_{i,j,m}(k)$ ;  
 wherein

- 40 C corresponds to the number of channels  $i = 0, \dots, C - 1$  of the echo canceller (50),  
 S corresponds to the number of segments  $j = 0, \dots, S - 1$  per channel of the echo canceller (50),  
 $\ell$  is an integer different than zero,  
 $k$  denotes a frequency bin,  
 $\epsilon$  is a positive value,  
 45  $Z_m(k)$  corresponds to the  $k$ th spectral bin of a microphone signal at the  $m$ th time block,  
 $Z_{m-\ell}(k)$  corresponds to the  $k$ th spectral bin of the microphone signal at the  $m - \ell$ th time block,  
 $X_{i,j,m}(k)$  corresponds to the  $k$ th spectral bin of a far-end signal at the  $i$ th channel,  $j$ th segment, and  $m$ th time block,  
 $X_{i,j,m-\ell}(k)$  corresponds to the  $k$ th spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m - \ell$ th time block,  
 50  $\lambda$  is an overshoot factor, and  
 $a, b$  are channel and segment indices, respectively.

55 **12.** The echo canceller (50) according to claim 11, wherein the echo canceller (50) is further configured to compute the optimal step size  $\mu_{i,j,m}(k)$  by either

$$\mu_{i,j,m}(k) = \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}$$

or

$$\mu_{i,j,m}(k) = 1 - \sqrt{1 - \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}}$$

wherein  $\Psi_m(k)$  is a power spectrum of the non-cancellable components at the  $m$ th time block and  $\Phi_{i,j}(k)$  is a power spectrum of the misalignment of each  $i, j$ th filter, wherein the echo canceller is configured to determine  $\Psi_m(k)$  and  $\Phi_{i,j}(k)$  by solving the set of linear equations

$$\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |E_m(k)|^2$$

$$Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\mathcal{E}(k)|^2$$

$$Y(k) + 4\overleftarrow{T}(k)\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}(k)|^2$$

$$4\overrightarrow{T}(k)Y(k) + \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}(k)|^2$$

subject to inequality constraints

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0;$$

wherein

$\mathcal{X}_{i,j,m}(k)$  corresponds to the  $k$ th power spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m$ th time block, wherein said signal block is preferably windowed,

$\mathcal{X}_{i,j,m-\ell}(k)$  corresponds to the  $k$ th power spectral bin of the far-end signal at the  $i$ th channel,  $j$ th segment, and  $m - \ell$ th time block, wherein said signal block is preferably windowed,

$Y(k)$  corresponds to the  $k$ th power spectral bin of a power spectrum of the non-cancellable components at the  $m - \ell$ th block,

$\overleftarrow{T}(k)$  is the  $k$ th frequency bin of the normalized look-backward up-date, computed as

$$\overleftarrow{T}(k) = \overleftarrow{\Delta}(k) / |E_m(k)|^2,$$

and

$\vec{T}(k)$  is the  $k$ th frequency bin of the normalized look-forward up-date, computed as

$$\vec{T}(k) = \vec{\Delta}(k) / |\mathcal{E}(k)|^2.$$

13. The echo canceller (50) according to claim 11 or 12, wherein the overshoot factor  $\lambda$  is in the range  $1 < \lambda \leq 2$ , and preferably equals 2.

14. The echo canceller (50) according to any one of the claims 11 to 13, wherein the echo canceller (50) is configured to determine an improved canceller output  $O_m(k)$  as

$$O_m(k) = E_m(k) B_m(k)$$

$$B_m(k) = \left[ \frac{|E_m(k)|^2 - \Lambda_m(k)}{|E_m(k)|^2} \right]_{\epsilon}$$

$$\Lambda_m(k) = \mathcal{R}_m(k) + \mathcal{Q}_m(k) + \mathcal{P}_m(k),$$

wherein

$\lfloor \cdot \rfloor_{\epsilon}$  denotes low clipping by  $\epsilon > 0$ ,

$E_m(k)$  is the canceller output,

$\mathcal{R}_m(k)$  is a power spectrum of the residual echo due to filter mis-alignment at the  $m$ th block,

$\mathcal{Q}_m(k)$  is a power spectrum of the echo beyond the canceller's time scope at the  $m$ th block, and

$\mathcal{P}_m(k)$  is a power spectrum of the non-linear echo components at the  $m$ th block.

15. The echo canceller (50) according to any one of claims 11 to 14, wherein the echo canceller (50) is configured to update the at least one filter coefficient  $W_{i,j}(k)$  based on

$$W_{i,j}^{new}(k) = W_{i,j}(k) + \mu_{i,j,m}(k) \frac{E_m(k) X_{i,j,m}^*(k)}{|X_{i,j,m}(k)|^2 + \epsilon}.$$

### Patentansprüche

1. Computer-implementiertes Verfahren zum Aktualisieren zumindest eines Filterkoeffizienten  $W_{i,j}(k)$  eines Echokompensators (50) im Frequenzbereich, wobei der Echokompensator zumindest einen Kanal und zumindest einen Abschnitt pro Kanal hat, und wobei die Filterkoeffizienten des Echokompensators (50) im Frequenzbereich in einem Zeitblock  $m$  aktualisiert werden können, **gekennzeichnet durch:**

Ermitteln einer Kompensatorausgabe  $E_m(k)$  über den  $m$ -ten Zeitblock als Differenz zwischen  $Z_m(k)$  und einem Kompensationsterm basierend auf

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m}(k);$$

Ermitteln eines Kompensatorfehlers  $\varepsilon(k)$  über den  $(m - \ell)$ -ten Zeitblock als Differenz zwischen  $Z_{m-\ell}(k)$  und einem Kompensationsterm basierend auf

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m-\ell}(k);$$

Ermitteln eines zurückschauenden Fehlers  $\overleftarrow{\varepsilon}(k)$  als

$$\overleftarrow{\varepsilon}(k) = \varepsilon(k) - \lambda \overleftarrow{\Delta}(k),$$

wobei  $\overleftarrow{\Delta}(k)$  basiert auf

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k);$$

Ermitteln eines vorausschauenden Fehlers  $\overrightarrow{\varepsilon}(k)$  als

$$\overrightarrow{\varepsilon}(k) = E_m(k) - \lambda \overrightarrow{\Delta}(k),$$

wobei  $\overrightarrow{\Delta}(k)$  basiert auf

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{\varepsilon(k) X_{i,j,m-\ell}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k);$$

Ermitteln einer optimalen Aktualisierungsschrittweite  $\mu_{i,j,m}(k)$  aus der genannten Kompensatorausgabe  $E_m(k)$  über den  $m$ -ten Zeitblock, aus dem genannten Kompensatorfehler  $\varepsilon(k)$  über den  $(m - \ell)$ -ten Zeitblock, aus dem genannten zurückschauenden Fehler  $\overleftarrow{\varepsilon}(k)$  und aus dem genannten vorausschauenden Fehler  $\overrightarrow{\varepsilon}(k)$ ; und Aktualisieren des zumindest einen Filterkoeffizienten  $W_{i,j}(k)$  unter Verwendung dieser optimalen Aktualisierungsschrittweite  $\mu_{i,j,m}(k)$ ;

wobei

$C$  der Anzahl an Kanälen  $i = 0, \dots, C - 1$  des Echokompensators (50) entspricht,

$S$  der Anzahl an Abschnitten  $j = 0, \dots, S - 1$  pro Kanal des Echokompensators (50) entspricht,

$\ell$  eine ganze Zahl ungleich 0 ist,

$k$  einen Frequenzklassenindex bezeichnet,

$\varepsilon$  ein positiver Wert ist,

$Z_m(k)$  der  $k$ -ten Spektralklasse eines Mikrofonsignals im  $m$ -ten Zeitblock entspricht,

$Z_{m-\ell}(k)$  der  $k$ -ten Spektralklasse des Mikrofonsignals im  $(m - \ell)$ -ten Zeitblock entspricht,

$X_{i,j,m}(k)$  der  $k$ -ten Spektralklasse eines entfernten Signals im  $i$ -ten Kanal,  $j$ -ten Abschnitt und  $m$ -ten Zeitblock entspricht,

$X_{i,j,m-\ell}(k)$  der  $k$ -ten Spektralklasse des entfernten Signals im  $i$ -ten Kanal,  $j$ -ten Abschnitt, und  $(m - \ell)$ -ten Zeitblock entspricht,

$\lambda$  ein Überschreitungs faktor ist, und

$\alpha, b$  Kanal- bzw. Abschnittsindizes sind.

2. Verfahren nach Anspruch 1, wobei die optimale Schrittweite  $\mu_{i,j,m}(k)$  entweder als

$$\mu_{i,j,m}(k) = \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}$$

oder als

$$\mu_{i,j,m}(k) = 1 - \sqrt{1 - \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}}$$

berechnet wird,

wobei  $\Psi_m(k)$  ein Leistungsspektrum der nicht-kompensierbaren Komponenten im  $m$ -ten Zeitblock und  $\Phi_{i,j}(k)$  ein Leistungsspektrum der Fehlausrichtung jedes  $i,j$ -ten Filters ist, wobei  $\Psi_m(k)$  und  $\Phi_{i,j}(k)$  durch Lösen des Satzes von linearen Gleichungen

$$\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |E_m(k)|^2$$

$$Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\mathcal{E}(k)|^2$$

$$Y(k) + \overleftarrow{4T}(k) \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}(k)|^2$$

$$\overrightarrow{4T}(k) Y(k) + \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}(k)|^2$$

unter den Ungleichungs-Nebenbedingungen

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0$$

ermittelt werden; wobei

$\mathcal{X}_{i,j,m}(k)$  der  $k$ -ten Leistungsspektralklasse des entfernten Signals im  $i$ -ten Kanal,  $j$ -ten Abschnitt und  $m$ -ten Zeitblock entspricht, wobei der genannte Signal-block bevorzugt gefenstert ist,

$\mathcal{X}_{i,j,m-\ell}(k)$  der  $k$ -ten Leistungsspektralklasse des entfernten Signals im  $i$ -ten Kanal,  $j$ -ten Abschnitt und  $(m - \ell)$ -ten

Zeitblock entspricht, wobei der genannte Signalblock bevorzugt gefenstert ist,  
 $Y(k)$  der  $k$ -ten Leistungsspektralklasse der nicht-kompensierbaren Komponenten im  $(m - \ell)$ -ten Block entspricht,  
 $\overleftarrow{T}(k)$  die normierte zurückschauende Aktualisierung ist, die als

5

$$\overleftarrow{T}(k) = \overleftarrow{\Delta}(k) / |E_m(k)|^2,$$

berechnet wird, und

10

$\overrightarrow{T}(k)$  die normierte vorausschauende Aktualisierung ist, die als

$$\overrightarrow{T}(k) = \overrightarrow{\Delta}(k) / |\mathcal{E}(k)|^2$$

15

berechnet wird.

3. Verfahren nach Anspruch 1, wobei zumindest eine Anzahl der Anzahl an Kanälen und Anzahl an Abschnitten pro Kanal größer als 1 ist; und  
 die optimale Schrittweite  $\mu_{i,j,m}(k)$  aus der genannten Kompensatorausgabe  $E_m(k)$  über den  $m$ -ten Zeitblock; aus dem genannten Kompensatorfehler  $\varepsilon(k)$  über den  $(m - \ell)$ -ten Zeitblock, aus dem genannten zurückschauenden Fehler  $\overleftarrow{\varepsilon}(k)$ , aus dem genannten vorausschauenden Fehler  $\overrightarrow{\varepsilon}(k)$  sowie aus zumindest einem zusätzlichen zurückschauenden Fehler  $\overleftarrow{\varepsilon}_{r,s}(k)$  und aus zumindest einem zusätzlichen vorausschauenden Fehler  $\overrightarrow{\varepsilon}_{r,s}(k)$  für zumindest ein  $r,s$ -tes Filter ermittelt wird, wobei  $r$  und  $s$  die Indizes des Kanals bzw. des Abschnitts sind, wobei

25

$$\overleftarrow{\varepsilon}_{r,s}(k) = \mathcal{E}(k) - \lambda \overleftarrow{\Delta}_{r,s}(k),$$

30

wobei  $\overleftarrow{\Delta}_{r,s}(k)$  basiert auf

35

$$\sum_i \sum_{j \neq r \& j \neq s} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_a \sum_b \substack{a \neq r \& b \neq s \\ |X_{a,b,m}(k)|^2 + \epsilon}} X_{i,j,m-\ell}(k),$$

40

und

$$\overrightarrow{\varepsilon}_{r,s}(k) = E_m(k) - \lambda \overrightarrow{\Delta}_{r,s}(k),$$

45

wobei  $\overrightarrow{\Delta}_{r,s}(k)$  basiert auf

50

$$\sum_i \sum_j \substack{i \neq r \& j \neq s \\ a \neq r \& b \neq s} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_a \sum_b |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k).$$

55

4. Verfahren nach den Ansprüchen 2 und 3, wobei das Leistungsspektrum  $\Psi_m(k)$  der nicht-kompensierbaren Komponenten und das Leistungsspektrum  $\Phi_{i,j}(k)$  der Fehlausrichtung jedes  $i,j$ -ten Filters durch Lösen des genannten Satzes von linearen Gleichungen ermittelt werden mit den zusätzlichen linearen Gleichungen für das zumindest eine  $r, s$ -te Filter

$$Y(k) + 4\overleftarrow{T}_{r,s}(k)\Psi_m(k) + 4\overleftarrow{T}_{r,s}(k)\mathcal{X}_{r,s,m}(k)\Phi_{r,s}(k) +$$

$$+ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-1}(k)\Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}_{r,s}(k)|^2$$

$$\Psi_m(k) + 4\overrightarrow{T}_{r,s}(k)Y(k) + 4\overrightarrow{T}_{r,s}(k)\mathcal{X}_{r,s,m-1}(k)\Phi_{r,s}(k) +$$

$$+ \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k)\Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}_{r,s}(k)|^2$$

unter den Ungleichungs-Nebenbedingungen

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0;$$

wobei

$\overleftarrow{T}_{s,r}(k)$  die normierte zurückschauende Aktualisierung unter Ausschluss des  $r,s$ -ten Filters ist, die berechnet wird als

$$\overleftarrow{T}_{r,s}(k) = \overleftarrow{\Delta}_{r,s}(k) / |E_m(k)|^2,$$

und  $\overrightarrow{T}_{r,s}(k)$  die normierte vorausschauende Aktualisierung unter Ausschluss des  $r,s$ -ten Filters ist, die berechnet wird als

$$\overrightarrow{T}_{r,s}(k) = \overrightarrow{\Delta}_{r,s}(k) / |\mathcal{E}(k)|^2.$$

5. Verfahren nach einem der Ansprüche 1 bis 4, wobei der Überschreitungsfaktor  $\lambda$  im Bereich  $1 < \lambda \leq 2$  liegt und bevorzugt gleich 2 ist.
6. Verfahren nach einem der Ansprüche 1 bis 5, **gekennzeichnet durch** Ermitteln eines Leistungsspektrums  $\mathcal{R}_m(k)$  des Restechos aufgrund der Filter-Fehlausrichtung im  $m$ -ten Block mittels

$$\mathcal{R}_m(k) = \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \Phi_{i,j}(k)\mathcal{X}_{i,j,m}(k),$$

und danach einer verbesserten Kompensatorausgabe  $O_m(k)$  basierend auf der Kompensatorausgabe  $E_m(k)$  und auf dem genannten ermittelten Leistungsspektrum  $\mathcal{R}_m(k)$  des Restechos aufgrund der Filter-Fehlausrichtung.

7. Verfahren nach einem der Ansprüche 1 bis 6, **gekennzeichnet durch** Ermitteln eines Leistungsspektrums  $\mathcal{Q}_m(k)$  des Echos jenseits des Zeiterfassungsbereichs des Kompensators im  $m$ -ten Block mittels des akkumulierenden Verfahrens

$$Q_m(k) = \gamma Q_{m-1}(k) + \gamma \sum_{i=0}^{C-1} |W_{i,S-1}(k)|^2 \mathcal{X}_{i,S-1,m-1}(k),$$

5

wobei  $0 < \gamma < 1$ ,  
und danach einer verbesserten Kompensatorausgabe  $O_m(k)$  basierend auf der Kompensatorausgabe  $E_m(k)$  und  
auf dem genannten ermittelten Leistungsspektrum  $Q_m(k)$  des Echos jenseits des Zeiterfassungsbereiches des  
Kompensators.

10

8. Verfahren nach einem der Ansprüche 1 bis 7, **gekennzeichnet durch** Ermitteln eines Leistungsspektrums  $P_m(k)$  der nicht-linearen Echokomponenten im  $m$ -ten Block mittels

15

$$P_m(k) \propto \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mu_{i,j,m}(k) \Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k),$$

20

wobei  $\alpha$  eine Proportionalität bezeichnet,  
und danach einer verbesserten Kompensatorausgabe  $O_m(k)$  basierend auf der Kompensatorausgabe  $E_m(k)$  und  
auf dem genannten ermittelten Leistungsspektrum  $P_m(k)$  der nicht-linearen Echokomponenten.

25

9. Verfahren nach einem der Ansprüche 1 bis 8, wobei das Verfahren eine verbesserte Kompensatorausgabe  $O_m(k)$  ermittelt als

30

$$O_m(k) = E_m(k) B_m(k)$$

$$B_m(k) = \left[ \frac{|E_m(k)|^2 - \Lambda_m(k)}{|E_m(k)|^2} \right]_{\varepsilon}$$

35

$$\Lambda_m(k) = \mathcal{R}_m(k) + Q_m(k) + P_m(k),$$

40

wobei

$\lfloor \cdot \rfloor_{\varepsilon}$  ein unteres Begrenzen durch  $\varepsilon > 0$  bezeichnet,

45

$E_m(k)$  die Kompensatorausgabe ist,

$\mathcal{R}_m(k)$  ein Leistungsspektrum des Restechos aufgrund der Filter-Fehlausrichtung im  $m$ -ten Block ist,

$Q_m(k)$  ein Leistungsspektrum des Echos jenseits des Zeiterfassungsbereiches des Kompensators im  $m$ -ten Block ist, und

$P_m(k)$  ein Leistungsspektrum der nicht-linearen Echokomponenten im  $m$ -ten Block ist.

50

10. Verfahren nach einem der Ansprüche 1 bis 9, wobei der zumindest eine Filterkoeffizient  $W_{i,j}(k)$  aktualisiert wird basierend auf

55

$$W_{i,j}^{new}(k) = W_{i,j}(k) + \mu_{i,j,m}(k) \frac{E_m(k) X_{i,j,m}^*(k)}{|X_{i,j,m}(k)|^2 + \varepsilon}.$$

11. Echokompensator (50), der zumindest einen Kanal und zumindest einen Abschnitt pro Kanal hat, wobei der Echokompensator (50) Filterkoeffizienten hat, die im Frequenzbereich in einem Zeitblock  $m$  aktualisierbar sind, **dadurch gekennzeichnet, dass** der Echokompensator ausgebildet ist zum:

5 Ermitteln einer Kompensatorausgabe  $E_m(k)$  über den  $m$ -ten Zeitblock als Differenz zwischen  $Z_m(k)$  und einem Kompensationsterm basierend auf

$$10 \quad \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m}(k);$$

15 Ermitteln eines Kompensatorfehlers  $\varepsilon(k)$  über den  $(m - \ell)$ -ten Zeitblock als Differenz zwischen  $Z_{m-\ell}(k)$  und einem Kompensationsterm basierend auf

$$20 \quad \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m-\ell}(k);$$

Ermitteln eines zurückschauenden Fehlers  $\overleftarrow{\varepsilon}(k)$  als

$$25 \quad \overleftarrow{\varepsilon}(k) = \varepsilon(k) - \lambda \overleftarrow{\Delta}(k)$$

wobei  $\overleftarrow{\Delta}(k)$  basiert auf

$$30 \quad \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k);$$

Ermitteln eines vorausschauenden Fehlers  $\overrightarrow{\varepsilon}(k)$  als

$$40 \quad \overrightarrow{\varepsilon}(k) = E_m(k) - \lambda \overrightarrow{\Delta}(k)$$

wobei  $\overrightarrow{\Delta}(k)$  basiert auf

$$45 \quad \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{\varepsilon(k) X_{i,j,m-\ell}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k);$$

55 Ermitteln einer optimalen Aktualisierungsschrittweite  $\mu_{i,j,m}(k)$  aus der genannten Kompensatorausgabe  $E_m(k)$  über den  $m$ -ten Zeitblock, aus dem genannten Kompensatorfehler  $\varepsilon(k)$  über den  $(m - \ell)$ -ten Zeitblock, aus dem genannten zurückschauenden Fehler  $\overleftarrow{\varepsilon}(k)$  und aus dem genannten vorausschauenden Fehler  $\overrightarrow{\varepsilon}(k)$ ; und Aktualisieren des zumindest einen Filterkoeffizienten  $W_{i,j}(k)$  unter Verwendung dieser optimalen Aktualisierungsschrittweite  $\mu_{i,j,m}(k)$ ;

wobei

C der Anzahl an Kanälen  $i = 0, \dots, C - 1$  des Echokompensators (50) entspricht,  
 S der Anzahl an Abschnitten  $j = 0, \dots, S - 1$  pro Kanal des Echokompensators (50) entspricht,  
 $\ell$  eine ganze Zahl ungleich 0 ist,  
 $k$  einen Frequenzklassenindex bezeichnet,  
 $\varepsilon$  ein positiver Wert ist,  
 $Z_m(k)$  der  $k$ -ten Spektralklasse eines Mikrofonsignals im  $m$ -ten Zeitblock entspricht,  
 $Z_{m-\ell}(k)$  der  $k$ -ten Spektralklasse des Mikrofonsignals im  $(m - \ell)$ -ten Zeitblock entspricht,  
 $X_{i,j,m}(k)$  der  $k$ -ten Spektralklasse eines entfernten Signals im  $i$ -ten Kanal,  $j$ -ten Abschnitt und  $m$ -ten Zeitblock entspricht,  
 $X_{i,j,m-\ell}(k)$  der  $k$ -ten Spektralklasse des entfernten Signals im  $i$ -ten Kanal,  $j$ -ten Abschnitt und  $(m - \ell)$ -ten Zeitblock entspricht,  
 $\lambda$  ein Überschreitungsfaktor ist, und  
 $\alpha, b$  Kanal- bzw. Abschnittsindizes sind.

12. Echokompensator (50) nach Anspruch 11, wobei der Echokompensator (50) ferner dazu ausgebildet ist, die optimale Schrittweise  $\mu_{i,j,m}(k)$  entweder als

$$\mu_{i,j,m}(k) = \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}$$

oder als

$$\mu_{i,j,m}(k) = 1 - \sqrt{1 - \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}}$$

zu berechnen,  
 wobei  $\Psi_m(k)$  ein Leistungsspektrum der nicht-kompensierbaren Komponenten im  $m$ -ten Zeitblock und  $\Phi_{i,j}(k)$  ein Leistungsspektrum einer Fehlausrichtung jedes  $i,j$ -ten Filters ist, wobei der Echokompensator dazu ausgebildet ist,  $\Psi_m(k)$  und  $\Phi_{i,j}(k)$  durch Lösen des Satzes von linearen Gleichungen

$$\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |E_m(k)|^2$$

$$Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\mathcal{E}(k)|^2$$

$$Y(k) + 4 \overleftarrow{T}(k) \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}(k)|^2$$

$$4 \vec{T}(k) Y(k) + \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |\vec{\mathcal{E}}(k)|^2$$

unter den Ungleichungs-Nebenbedingungen

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0$$

zu ermitteln, wobei

$X_{i,j,m}(k)$  der  $k$ -ten Leistungsspektralklasse des entfernten Signals im  $i$ -ten Kanal,  $j$ -ten Abschnitt und  $m$ -ten Zeitblock entspricht, wobei der genannte Signalblock bevorzugt gefenstert ist,  
 $X_{i,j,m-\ell}(k)$  der  $k$ -ten Leistungsspektralklasse des entfernten Signals im  $i$ -ten Kanal,  $j$ -ten Abschnitt und  $(m - \ell)$ -ten Zeitblock entspricht, wobei der genannte Signalblock bevorzugt gefenstert ist,  
 $Y(k)$  der  $k$ -ten Leistungsspektralklasse der nicht-kompensierbaren Komponenten im  $(m - \ell)$ -ten Block entspricht,  
 $\overleftarrow{T}(k)$  die  $k$ -te Frequenzklasse der normierten zurückschauenden Aktualisierung ist, die als

$$\overleftarrow{T}(k) = \overleftarrow{\Delta}(k) / |E_m(k)|^2,$$

berechnet wird, und

$\overrightarrow{T}(k)$  die  $k$ -te Frequenzklasse der normierten vorausschauenden Aktualisierung ist die als

$$\overrightarrow{T}(k) = \overrightarrow{\Delta}(k) / |\mathcal{E}(k)|^2.$$

berechnet wird.

**13.** Echokompensator (50) nach Anspruch 11 oder 12, wobei der Überschreitungsfaktor  $\lambda$  im Bereich  $1 < \lambda \leq 2$  liegt und bevorzugt gleich 2 ist.

**14.** Echokompensator (50) nach einem der Ansprüche 11 bis 13, wobei der Echokompensator (50) dazu ausgebildet ist, eine verbesserte Kompensatorausgabe  $O_m(k)$  als

$$O_m(k) = E_m(k) B_m(k)$$

$$B_m(k) = \left\lfloor \frac{|E_m(k)|^2 - \Lambda_m(k)}{|E_m(k)|^2} \right\rfloor_{\varepsilon}$$

$$\Lambda_m(k) = \mathcal{R}_m(k) + \mathcal{Q}_m(k) + \mathcal{P}_m(k),$$

zu ermitteln,

wobei

$\lfloor \cdot \rfloor_{\varepsilon}$  ein unteres Begrenzen durch  $\varepsilon > 0$  bezeichnet,  
 $E_m(k)$  die Kompensatorausgabe ist,.

$R_m(k)$  ein Leistungsspektrum des Restechos aufgrund von Filterfehlausrichtung im  $m$ -ten Block ist,  
 $Q_m(k)$  ein Leistungsspektrum des Echos jenseits des Zeiterfassungsbereichs des Kompensators im  $m$ -ten Block ist, und  
 $P_m(k)$  ein Leistungsspektrum der nicht-linearen Echokomponenten im  $m$ -ten Block ist.

15. Echokompensator (50) nach einem der Ansprüche 11 bis 14, wobei der Echokompensator (50) dazu ausgebildet ist, den zumindest einen Filterkoeffizienten  $W_{i,j}(k)$  zu aktualisieren basierend auf

$$W_{i,j}^{new}(k) = W_{i,j}(k) + \mu_{i,j,m}(k) \frac{E_m(k) X_{i,j,m}^*(k)}{|X_{i,j,m}(k)|^2 + \epsilon}.$$

### Revendications

1. Procédé mis en œuvre par ordinateur pour actualiser au moins un coefficient de filtre  $W_{i,j}(k)$  d'un dispositif d'annulation d'écho (50) dans le domaine de fréquence, le dispositif d'annulation d'écho (50) ayant au moins un canal et au moins un segment par canal, les coefficients de filtre du dispositif d'annulation d'écho (50) pouvant être actualisés dans le domaine de fréquence à un bloc temporel  $m$ , **caractérisé par** :

la détermination d'une sortie du dispositif d'annulation  $E_m(k)$  sur le  $m$ <sup>ième</sup> bloc temporel sous forme de la différence entre  $Z_m(k)$  et un terme d'annulation basé sur

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m}(k);$$

la détermination d'une erreur du dispositif d'annulation  $\varepsilon(k)$  sur le  $m - \ell$ <sup>ième</sup> bloc temporel sous la forme de la différence entre  $Z_{m-\ell}(k)$  et un terme d'annulation basé sur

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m-\ell}(k);$$

la détermination d'une erreur vers l'arrière  $\overleftarrow{\varepsilon}(k)$  sous la forme

$$\overleftarrow{\varepsilon}(k) = \varepsilon(k) - \lambda \overleftarrow{\Delta}(k),$$

où  $\overleftarrow{\Delta}(k)$  est basé sur

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k);$$

la détermination d'une erreur vers l'avant  $\overrightarrow{\varepsilon}(k)$  sous la forme

$$\overrightarrow{\varepsilon}(k) = E_m(k) - \lambda \overrightarrow{\Delta}(k),$$

où  $\vec{\Delta}(k)$  est basé sur

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k);$$

la détermination d'une largeur de pas d'actualisation optimale  $\mu_{i,j,m}(k)$  à partir de ladite sortie du dispositif d'annulation  $E_m(k)$  sur le  $m^{\text{ième}}$  bloc temporel, à partir de ladite erreur de dispositif d'annulation  $\varepsilon(k)$  sur le  $m - \ell^{\text{ième}}$  bloc temporel, à partir de ladite erreur vers l'arrière  $\overleftarrow{\varepsilon}(k)$  et à partir de ladite erreur vers l'avant  $\overrightarrow{\varepsilon}(k)$ ; et l'actualisation dudit au moins un coefficient de filtre  $W_{i,j}(k)$  en utilisant ladite largeur de pas d'actualisation optimale  $\mu_{i,j,m}(k)$ ;

où

$C$  correspond au nombre de canaux  $i = 0, \dots, C-1$  du dispositif d'annulation d'écho (50),

$S$  correspond au nombre de segments  $j = 0, \dots, S-1$  par canal du dispositif d'annulation d'écho (50),

$\ell$  est un nombre entier différent de 0,

$k$  désigne un indice de classe de fréquence,

$\varepsilon$  est une valeur positive,

$Z_m(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale d'un signal de microphone au  $m^{\text{ième}}$  bloc temporel,

$Z_{m-\ell}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale du signal de microphone au  $m - \ell^{\text{ième}}$  bloc temporel,

$X_{i,j,m}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale d'un signal lointain au  $i^{\text{ième}}$  canal,  $j^{\text{ième}}$  segment et  $m^{\text{ième}}$  bloc temporel,

$X_{i,j,m-\ell}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale d'un signal lointain au  $i^{\text{ième}}$  canal,  $j^{\text{ième}}$  segment et  $m - \ell^{\text{ième}}$  bloc temporel,

$\lambda$  est un facteur de dépassement, et

$\alpha, b$  sont des indices de classe et de segment, respectivement.

2. Procédé selon la revendication 1, dans lequel la largeur de pas optimale  $\mu_{i,j,m}(k)$  est calculée soit par

$$\mu_{i,j,m}(k) = \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}$$

soit par

$$\mu_{i,j,m}(k) = 1 - \sqrt{1 - \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}},$$

où  $\Psi_m(k)$  est un spectre de puissance des composants non annulables au  $m^{\text{ième}}$  bloc temporel et  $\Phi_{i,j}(k)$  est un spectre de puissance du désalignement de chaque  $i,j^{\text{ième}}$  filtre, où  $\Psi_m(k)$  et  $\Phi_{i,j}(k)$  sont déterminés en résolvant l'ensemble des équations linéaires

$$\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |E_m(k)|^2$$

$$Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} X_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\mathcal{E}(k)|^2$$

$$Y(k) + 4\overleftarrow{T}(k)\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} X_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}(k)|^2$$

$$4\overrightarrow{T}(k)Y(k) + \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} X_{i,j,m}(k) \Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}(k)|^2$$

soumises à des contraintes d'inégalité

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0;$$

où

$X_{i,j,m}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale de puissance d'un signal lointain au  $i^{\text{ième}}$  canal,  $j^{\text{ième}}$  segment et  $m^{\text{ième}}$  bloc temporel, où ledit bloc de signal est de préférence soumis à une fenêtre,

$X_{i,j,m-\ell}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale de puissance d'un signal lointain au  $i^{\text{ième}}$  canal,  $j^{\text{ième}}$  segment et  $m - \ell^{\text{ième}}$  bloc temporel, où ledit bloc de signal est de préférence soumis à une fenêtre,

$Y(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale de puissance des composants non annulables au  $m - \ell^{\text{ième}}$  bloc,  $\overleftarrow{T}(k)$  est l'actualisation en arrière normalisée, calculée sous la forme

$$\overleftarrow{T}(k) = \overleftarrow{\Delta}(k) / |E_m(k)|^2,$$

et

$\overrightarrow{T}(k)$  est l'actualisation en avant normalisée, calculée sous la forme

$$\overrightarrow{T}(k) = \overrightarrow{\Delta}(k) / |\mathcal{E}(k)|^2.$$

3. Procédé selon la revendication 1, dans lequel au moins un parmi le nombre de canaux et le nombre de segments par canal est plus grand que un ; et la largeur de pas optimale  $\mu_{i,j,m}(k)$  est déterminée à partir de ladite sortie du dispositif d'annulation  $E_m(k)$  sur le  $m^{\text{ième}}$  bloc temporel, à partir de ladite erreur du dispositif d'annulation  $\varepsilon(k)$  sur le  $m - \ell^{\text{ième}}$  bloc temporel, à partir de ladite erreur vers l'arrière  $\overleftarrow{\varepsilon}(k)$  et à partir de ladite erreur vers l'avant  $\overrightarrow{\varepsilon}(k)$  ainsi qu'à partir d'au moins une erreur vers l'arrière supplémentaire  $\overleftarrow{\varepsilon}_{r,s}(k)$  et à partir d'au moins une erreur vers l'avant supplémentaire  $\overrightarrow{\varepsilon}_{r,s}(k)$  pour au moins un  $r,s^{\text{ième}}$  filtre,  $r$  et  $s$  étant des indices du canal et du segment, respectivement, où

$$\overleftarrow{\varepsilon}_{r,s}(k) = \mathcal{E}(k) - \lambda \overleftarrow{\Delta}_{r,s}(k),$$

où  $\overleftarrow{\Delta}_{r,s}(k)$  est basé sur

$$\sum_i \sum_{j \neq s} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_a \sum_b |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k),$$

et

$$\overrightarrow{\mathcal{E}}_{r,s}(k) = E_m(k) - \lambda \overrightarrow{\Delta}_{r,s}(k),$$

où  $\overrightarrow{\Delta}_{r,s}(k)$  est basé sur

$$\sum_i \sum_{j \neq s} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_a \sum_b |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k).$$

4. Procédé selon les revendications 2 et 3, dans lequel le spectre de puissance  $\Psi_m(k)$  des composants non annulables et le spectre de puissance  $\Phi_{i,j}(k)$  de désalignement de chaque  $i$  <sup>j</sup>ème filtre sont déterminés en résolvant ledit ensemble d'équations linéaires avec des équations linéaires supplémentaires pour au moins un  $r,s$ ème filtre

$$Y(k) + 4\overleftarrow{T}_{r,s}(k)\Psi_m(k) + 4\overleftarrow{T}_{r,s}(k)\mathcal{X}_{r,s,m}(k)\Phi_{r,s}(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-1}(k)\Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}_{r,s}(k)|^2$$

$$\Psi_m(k) + 4\overrightarrow{T}_{r,s}(k)Y(k) + 4\overrightarrow{T}_{r,s}(k)\mathcal{X}_{r,s,m-1}(k)\Phi_{r,s}(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k)\Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}_{r,s}(k)|^2$$

soumises à des contraintes d'inégalité

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0;$$

où

$\overleftarrow{T}_{r,s}(k)$  est l'actualisation vers l'arrière normalisée excluant le  $r,s$ ème filtre, calculée sous la forme

$$\overleftarrow{T}_{r,s}(k) = \overleftarrow{\Delta}_{r,s}(k) / |E_m(k)|^2,$$

et  $\overrightarrow{T}_{r,s}(k)$  est l'actualisation vers l'avant normalisée excluant le  $r,s$ ème filtre, calculée sous la forme

$$\vec{T}_{r,s}(k) = \vec{\Delta}_{r,s}(k) / |\mathcal{E}(k)|^2.$$

5 5. Procédé selon l'une quelconque des revendications 1 à 4, dans lequel le facteur de dépassement  $\lambda$  est dans la plage  $1 < \lambda \leq 2$ , et est de préférence égal à 2.

6. Procédé selon l'une quelconque des revendications 1 à 5, **caractérisé par** la détermination d'un spectre de puissance  $\mathcal{R}_m(k)$  de l'écho résiduel dû à un désalignement de filtre au  $m^{\text{ième}}$  bloc au moyen de

$$15 \quad \mathcal{R}_m(k) = \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k),$$

et par la suite d'une sortie du dispositif d'annulation améliorée  $O_m(k)$  basée sur la sortie du dispositif d'annulation  $E_m(k)$  et sur ledit spectre de puissance  $\mathcal{R}_m(k)$  déterminé de l'écho résiduel dû au désalignement de filtre.

7. Procédé selon l'une quelconque des revendications 1 à 6, **caractérisé par** la détermination d'un spectre de puissance  $\mathcal{Q}_m(k)$  de l'écho au-delà de la portée temporelle du dispositif d'annulation au  $m^{\text{ième}}$  bloc au moyen d'un procédé accumulatif

$$30 \quad \mathcal{Q}_m(k) = \gamma \mathcal{Q}_{m-1}(k) + \gamma \sum_{i=0}^{C-1} |W_{i,S-1}(k)|^2 \mathcal{X}_{i,S-1,m-1}(k),$$

où  $0 < \gamma < 1$ ,  
et par la suite d'une sortie du dispositif d'annulation améliorée  $O_m(k)$  basée sur la sortie du dispositif d'annulation  $E_m(k)$  et sur ledit spectre de puissance  $\mathcal{Q}_m(k)$  déterminé de l'écho au-delà de la portée temporelle du dispositif d'annulation.

8. Procédé selon l'une quelconque des revendications 1 à 7, **caractérisé par** la détermination d'un spectre de puissance  $\mathcal{P}_m(k)$  des composants d'écho non linéaires au  $m^{\text{ième}}$  bloc au moyen de

$$40 \quad \mathcal{P}_m(k) \propto \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mu_{i,j,m}(k) \Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k),$$

où  $\alpha$  désigne une proportionnalité,  
et par la suite d'une sortie du dispositif d'annulation améliorée  $O_m(k)$  basée sur la sortie du dispositif d'annulation  $E_m(k)$  et sur ledit spectre de puissance  $\mathcal{P}_m(k)$  déterminé des composants d'écho non linéaires.

9. Procédé selon l'une quelconque des revendications 1 à 8, dans lequel le procédé détermine une sortie de dispositif d'annulation améliorée  $O_m(k)$  sous la forme

$$55 \quad O_m(k) = E_m(k) B_m(k)$$

$$B_m(k) = \left[ \frac{|E_m(k)|^2 - \Lambda_m(k)}{|E_m(k)|^2} \right]_\varepsilon$$

$$\Lambda_m(k) = \mathcal{R}_m(k) + \mathcal{Q}_m(k) + \mathcal{P}_m(k),$$

où

$\lfloor \cdot \rfloor_\varepsilon$  désigne un écrêtage à bas par  $\varepsilon > 0$ ,

$E_m(k)$  est la sortie du dispositif d'annulation,

$\mathcal{R}_m(k)$  est un spectre de puissance de l'écho résiduel dû au désalignement de filtre au  $m^{\text{ième}}$  bloc,

$\mathcal{Q}_m(k)$  est un spectre de puissance de l'écho au-delà de la portée temporelle du dispositif d'annulation au  $m^{\text{ième}}$  bloc, et

$\mathcal{P}_m(k)$  est un spectre de puissance des composants d'écho non linéaires au  $m^{\text{ième}}$  bloc.

10. Procédé selon l'une quelconque des revendications 1 à 9, dans lequel l'au moins un coefficient de filtre  $W_{i,j}(k)$  est actualisé sur la base de

$$W_{i,j}^{\text{new}}(k) = W_{i,j}(k) + \mu_{i,j,m}(k) \frac{E_m(k) X_{i,j,m}^*(k)}{|X_{i,j,m}(k)|^2 + \varepsilon}.$$

11. Dispositif d'annulation d'écho (50) ayant au moins un canal et au moins un segment par canal, le dispositif d'annulation d'écho (50) ayant des coefficients de filtre  $W_{i,j}(k)$  pouvant être actualisés dans le domaine de fréquence à un bloc temporel  $m$ , **caractérisé en ce que** le dispositif d'annulation d'écho est conçu pour :

la détermination d'une sortie de dispositif d'annulation  $E_m(k)$  sur le  $m^{\text{ième}}$  bloc temporel lorsque la différence entre  $Z_m(k)$  et un terme d'annulation basé sur

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m}(k);$$

la détermination d'une erreur du dispositif d'annulation  $\varepsilon(k)$  sur le  $m - \ell^{\text{ième}}$  bloc temporel lorsque la différence entre  $Z_{m-\ell}(k)$  et un terme d'annulation basé sur

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} W_{i,j}(k) X_{i,j,m-\ell}(k);$$

la détermination d'une erreur vers l'arrière  $\overleftarrow{\varepsilon}(k)$  sous la forme

$$\overleftarrow{\varepsilon}(k) = \varepsilon(k) - \lambda \overleftarrow{\Delta}(k),$$

où  $\overleftarrow{\Delta}(k)$  est basé sur

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{E_m(k) X_{i,j,m}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m}(k)|^2 + \epsilon} X_{i,j,m-\ell}(k);$$

la détermination d'une erreur vers l'avant  $\vec{\mathcal{E}}(k)$  sous la forme

$$\vec{\mathcal{E}}(k) = E_m(k) - \lambda \vec{\Delta}(k),$$

où  $\vec{\Delta}(k)$  est basé sur

$$\sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \frac{\mathcal{E}(k) X_{i,j,m-\ell}^*(k)}{\sum_{a=0}^{C-1} \sum_{b=0}^{S-1} |X_{a,b,m-\ell}(k)|^2 + \epsilon} X_{i,j,m}(k);$$

la détermination d'une largeur de pas d'actualisation optimale  $\mu_{i,j,m}(k)$  à partir de ladite sortie du dispositif d'annulation  $E_m(k)$  sur le  $m^{\text{ième}}$  bloc temporel, à partir de ladite erreur de dispositif d'annulation  $\mathcal{E}(k)$  sur le  $m - \ell^{\text{ième}}$  bloc temporel, à partir de ladite erreur vers l'arrière  $\overleftarrow{\mathcal{E}}(k)$  et à partir de ladite erreur vers l'avant  $\vec{\mathcal{E}}(k)$ ; et l'actualisation dudit au moins un coefficient de filtre  $W_{i,j}(k)$  en utilisant ladite largeur de pas d'actualisation optimale  $\mu_{i,j,m}(k)$  ;  
où

- C correspond au nombre de canaux  $i = 0, \dots, C - 1$  du dispositif d'annulation d'écho (50),
- S correspond au nombre de segments  $j = 0, \dots, S - 1$  par canal du dispositif d'annulation d'écho (50),
- $\ell$  est un nombre entier différent de 0,
- $k$  désigne une classe de fréquence,
- $\epsilon$  est une valeur positive,
- $Z_m(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale d'un signal de microphone au  $m^{\text{ième}}$  bloc temporel,
- $Z_{m-\ell}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale du signal de microphone au  $m - \ell^{\text{ième}}$  bloc temporel,
- $X_{i,j,m}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale d'un signal lointain au  $i^{\text{ième}}$  canal,  $j^{\text{ième}}$  segment et  $m^{\text{ième}}$  bloc temporel,
- $X_{i,j,m-\ell}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale d'un signal lointain au  $i^{\text{ième}}$  canal,  $j^{\text{ième}}$  segment et  $m - \ell^{\text{ième}}$  bloc temporel,
- $\lambda$  est un facteur de dépassement, et
- $a, b$  sont des indices de canal et de segment, respectivement.

12. Dispositif d'annulation d'écho (50) selon la revendication 11, dans lequel le dispositif d'annulation de l'écho (50) est en outre conçu pour calculer la largeur de pas optimale  $\mu_{i,j,m}(k)$  soit par

$$\mu_{i,j,m}(k) = \frac{\Phi_{i,j}(k) \mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k) \mathcal{X}_{a,b,m}(k)}$$

soit par

5

$$\mu_{i,j,m}(k) = 1 - \sqrt{1 - \frac{\Phi_{i,j}(k)\mathcal{X}_{i,j,m}(k)}{\Psi_m(k) + \sum_{a=0}^{C-1} \sum_{b=0}^{S-1} \Phi_{a,b}(k)\mathcal{X}_{a,b,m}(k)}}$$

10

où  $\Psi_m(k)$  est un spectre de puissance des composants non annulables au  $m^{\text{ième}}$  bloc temporel et  $\Phi_{i,j}(k)$  est un spectre de puissance du désalignement de chaque  $i,j^{\text{ième}}$  filtre, où le dispositif d'annulation de l'écho (50) est conçu pour déterminer  $\Psi_m(k)$  et  $\Phi_{i,j}(k)$  en résolvant l'ensemble des équations linéaires

15

$$\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |E_m(k)|^2$$

20

$$Y(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\mathcal{E}(k)|^2$$

25

$$Y(k) + 4\overleftarrow{T}(k)\Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m-\ell}(k) \Phi_{i,j}(k) = |\overleftarrow{\mathcal{E}}(k)|^2$$

30

$$4\overrightarrow{T}(k)Y(k) + \Psi_m(k) + \sum_{i=0}^{C-1} \sum_{j=0}^{S-1} \mathcal{X}_{i,j,m}(k) \Phi_{i,j}(k) = |\overrightarrow{\mathcal{E}}(k)|^2$$

35

soumises à des contraintes d'inégalité

40

$$Y(k) > 0, \Psi_m(k) > 0, \Phi_{0,0}(k) > 0, \dots, \Phi_{C-1,S-1}(k) > 0;$$

45

où

$\mathcal{X}_{i,j,m}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale de puissance d'un signal lointain au  $i^{\text{ième}}$  canal,  $j^{\text{ième}}$  segment et  $m^{\text{ième}}$  bloc temporel, où ledit bloc de signal est de préférence soumis à une fenêtre,

50

$\mathcal{X}_{i,j,m-\ell}(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale de puissance d'un signal lointain au  $i^{\text{ième}}$  canal,  $j^{\text{ième}}$  segment et  $m - \ell^{\text{ième}}$  bloc temporel, où ledit bloc de signal est de préférence soumis à une fenêtre,

$Y(k)$  correspond à la  $k^{\text{ième}}$  classe spectrale de puissance des composants non annulables au  $m - \ell^{\text{ième}}$  bloc,  $\overleftarrow{T}(k)$  est la  $k^{\text{ième}}$  classe de fréquence de l'actualisation en arrière normalisée, calculée sous la forme

55

$$\overleftarrow{T}(k) = \overleftarrow{\Delta}(k) / |E_m(k)|^2,$$

et

$\vec{T}(k)$  est la  $k^{\text{ième}}$  classe de fréquence de l'actualisation en avant normalisée, calculée sous la forme

$$\vec{T}(k) = \vec{\Delta}(k) / |\mathcal{E}(k)|^2.$$

13. Dispositif d'annulation d'écho (50) selon la revendication 11 ou 12, dans lequel le facteur de dépassement  $\lambda$  est dans la plage  $1 < \lambda \leq 2$ , et est de préférence égal à 2.

14. Dispositif d'annulation d'écho (50) selon l'une quelconque des revendications 11 à 13, dans lequel le dispositif d'annulation d'écho (50) est conçu pour déterminer une sortie du dispositif d'annulation améliorée  $O_m(k)$  sous la forme

$$O_m(k) = E_m(k) B_m(k)$$

$$B_m(k) = \left\lfloor \frac{|E_m(k)|^2 - \Lambda_m(k)}{|E_m(k)|^2} \right\rfloor_{\varepsilon}$$

$$\Lambda_m(k) = \mathcal{R}_m(k) + \mathcal{Q}_m(k) + \mathcal{P}_m(k),$$

où

$\lfloor \cdot \rfloor_{\varepsilon}$  désigne un écrêtage à bas par  $\varepsilon > 0$ ,

$E_m(k)$  est la sortie de dispositif d'annulation,

$\mathcal{R}_m(k)$  est un spectre de puissance de l'écho résiduel dû au désalignement de filtre au  $m^{\text{ième}}$  bloc,

$\mathcal{Q}_m(k)$  est un spectre de puissance de l'écho au-delà de la portée temporelle du dispositif d'annulation au  $m^{\text{ième}}$  bloc, et

$\mathcal{P}_m(k)$  est un spectre de puissance des composants d'écho non linéaires au  $m^{\text{ième}}$  bloc.

15. Dispositif d'annulation d'écho (50) selon l'une quelconque des revendications 11 à 14, dans lequel le dispositif d'annulation d'écho (50) est conçu pour actualiser l'au moins un coefficient de filtre  $W_{i,j}(k)$  sur la base de

$$W_{i,j}^{new}(k) = W_{i,j}(k) + \mu_{i,j,m}(k) \frac{E_m(k) X_{i,j,m}^*(k)}{|X_{i,j,m}(k)|^2 + \epsilon}.$$

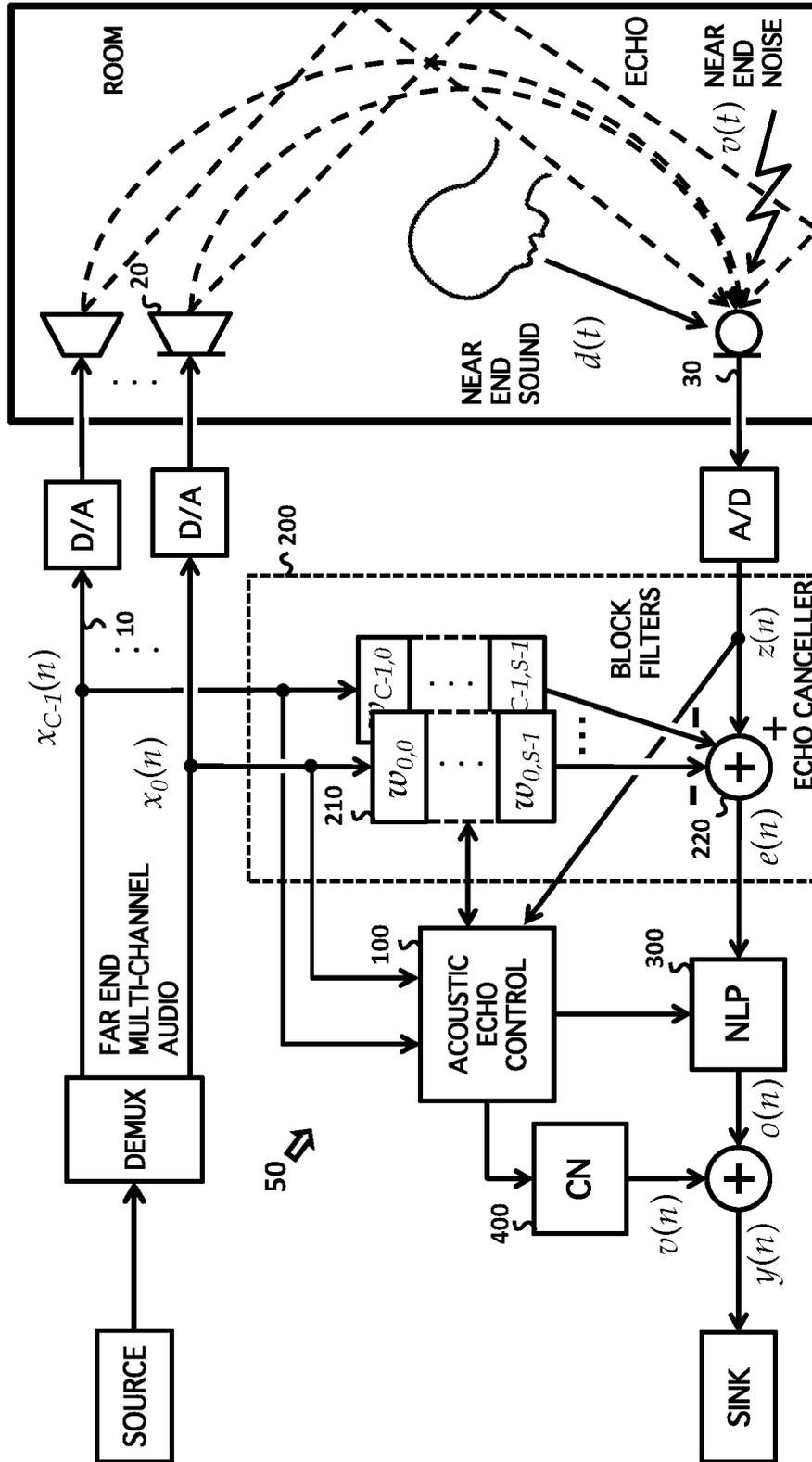


Figure 1

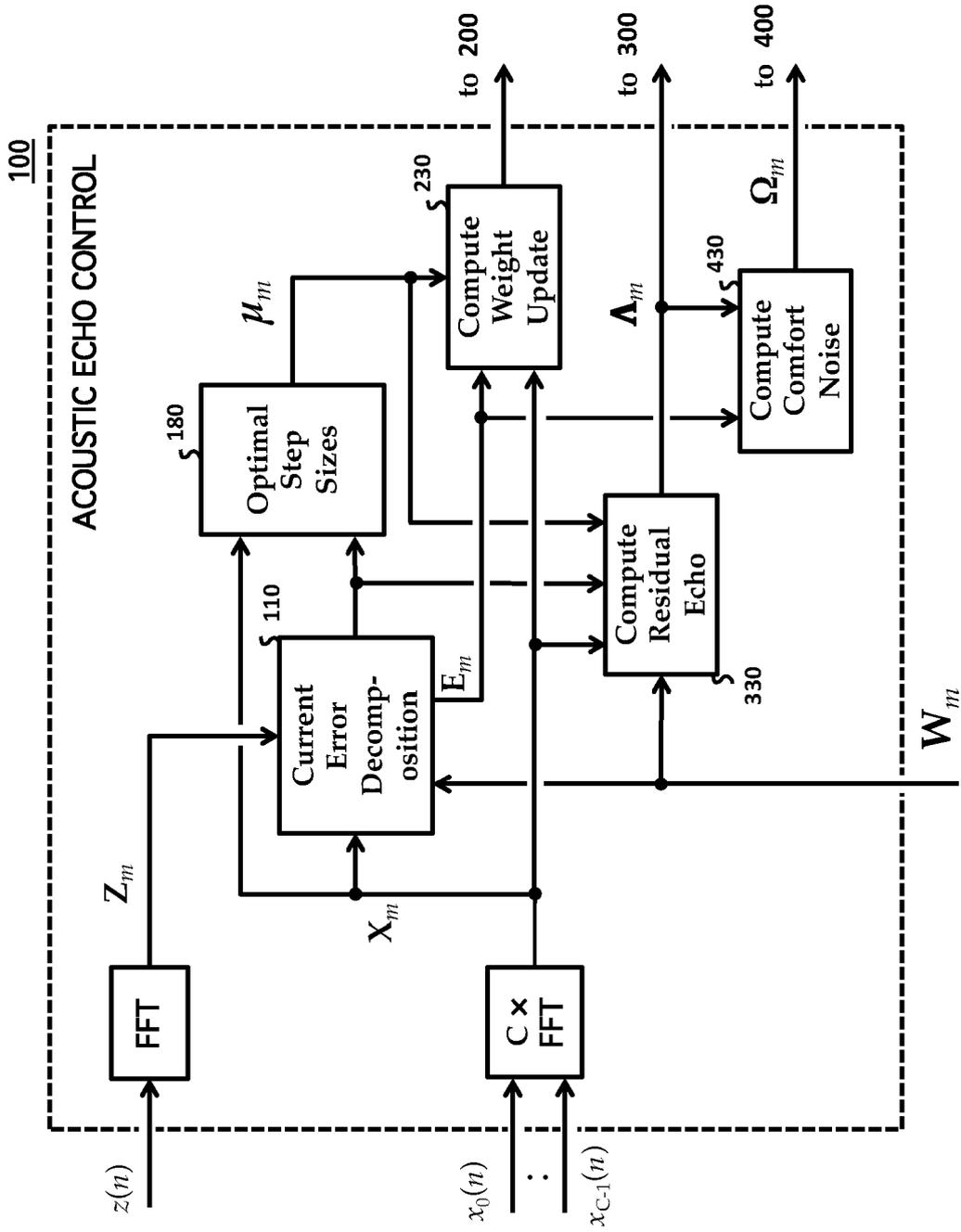


Figure 2

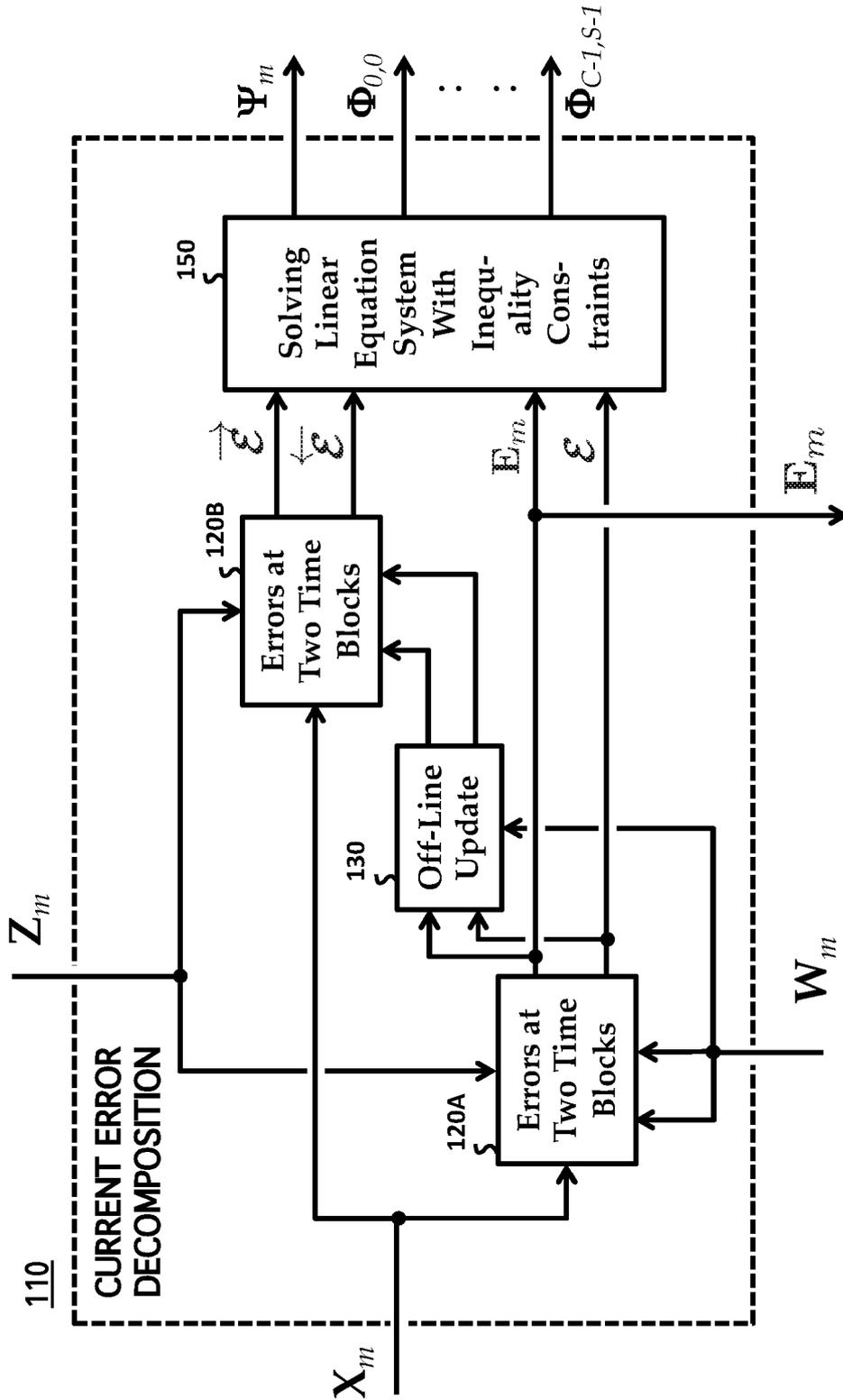


Figure 3

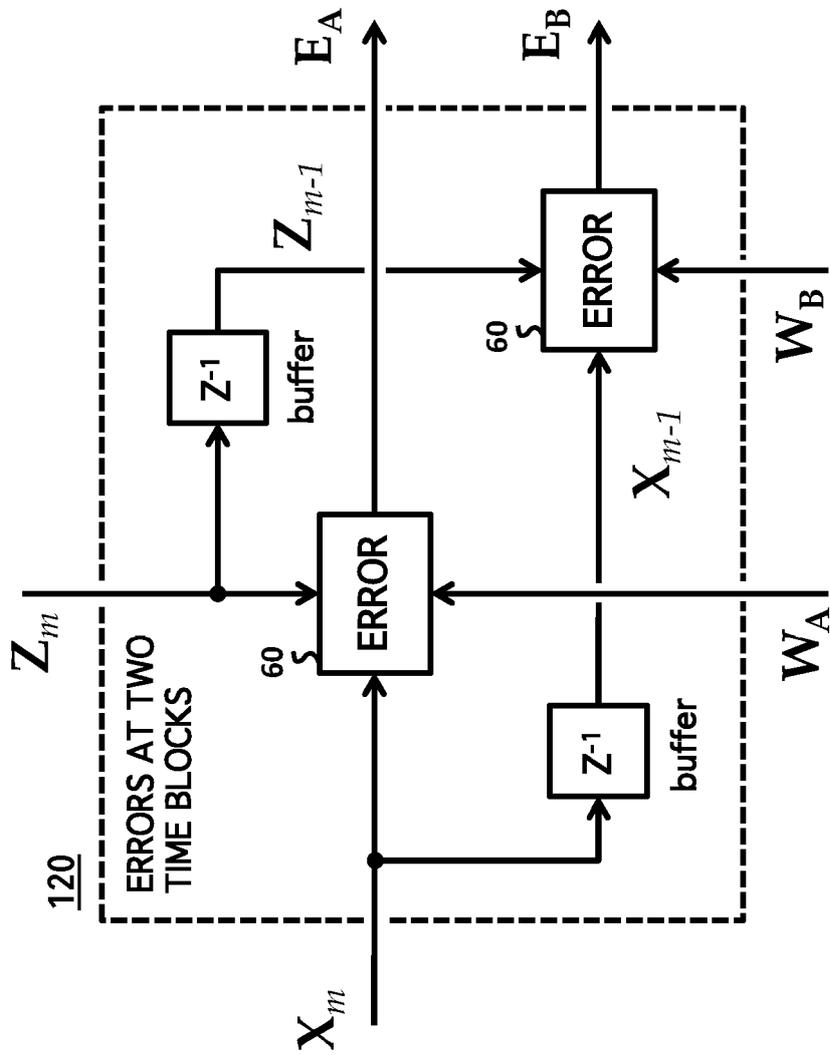


Figure 4

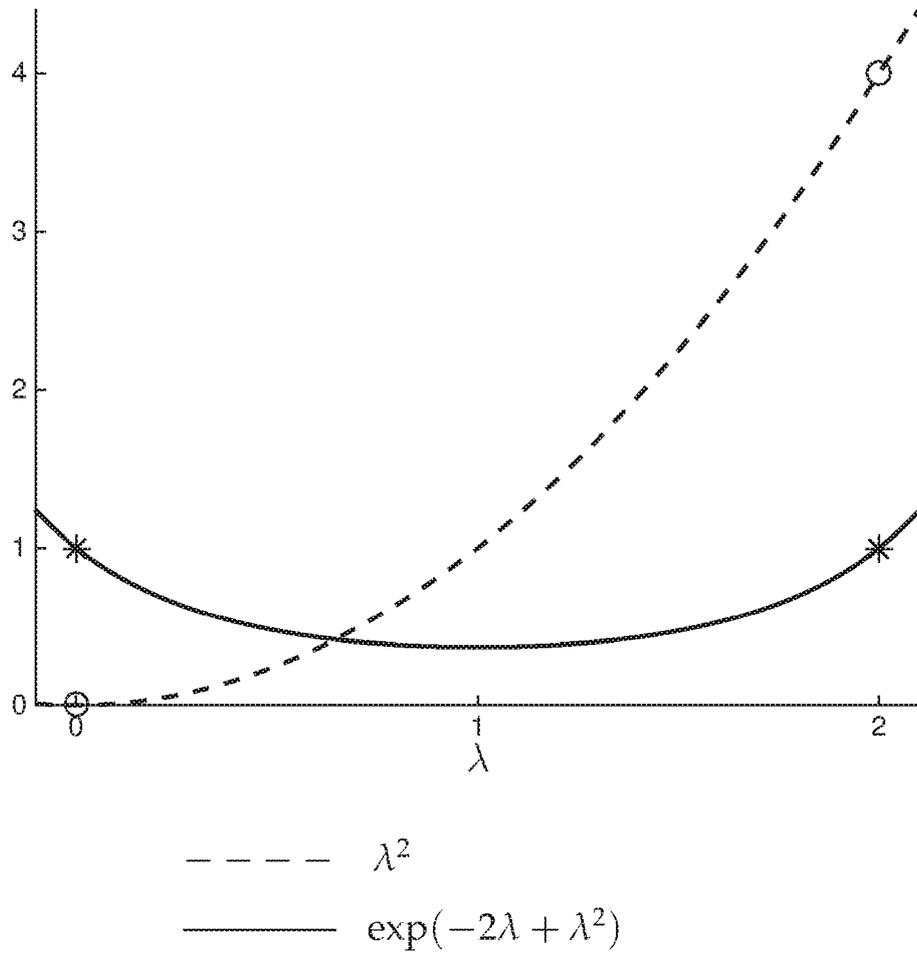


Figure 5

**REFERENCES CITED IN THE DESCRIPTION**

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

**Patent documents cited in the description**

- EP 2438766 B1 [0002] [0020]
- US 20060018460 A1 [0017] [0020]
- US 8233632 B1 [0020]
- US 6246760 B1 [0020]
- US 8873740 B2 [0021] [0029]
- US 20170118326 A1 [0027]
- US 20160029120 A1 [0027]
- US 5949888 A1 [0028]
- US 7610197 B2 [0029]
- US 20110228946 A1 [0029]
- EP 2930917 B1 [0071]

**Non-patent literature cited in the description**

- Advances in digital speech transmission. Wiley, 2008 [0002]
- Unconstrained frequency-domain adaptive filter. IEEE Transactions on Acoustics, Speech and Signal Processing. October 1982 [0018]
- Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication. *Signal Processing*, 2005, vol. 85, 549-570 [0020]
- A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain. *Signal Processing*, September 2000, vol. 80 [0057]
- A frequency domain step-size control method for LMS algorithms. *IEEE Signal Processing Letters*, 17 February 2010 [0058]