

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 March 2011 (10.03.2011)

PCT

(10) International Publication Number
WO 2011/027005 A9

(51) International Patent Classification:
H04R 25/00 (2006.01)

(21) International Application Number:
PCT/EP2010/070282

(22) International Filing Date:
20 December 2010 (20.12.2010)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US):
PHONAK AG [—/CH]; Laubisrütistrasse 28, CH-8712
Stäfa (CH).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **MÜLDER, Hans**
[NL/CH]; Steinackerstrasse 21, CH-3184 Wünnewil
(CH).

(74) Agents: **SCHORER, Reinhard** et al.; Bauerstr. 22,
80796 München (DE).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,

KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG,
ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Published:

— upon request of the applicant, before the expiration of the
time limit referred to in Article 21(2)(a)

— without international search report and to be republished
upon receipt of that report (Rule 48.2(g))

(48) Date of publication of this corrected version:

6 October 2011

(15) Information about Correction:
see Notice of 6 October 2011

(54) Title: METHOD AND SYSTEM FOR SPEECH ENHANCEMENT IN A ROOM

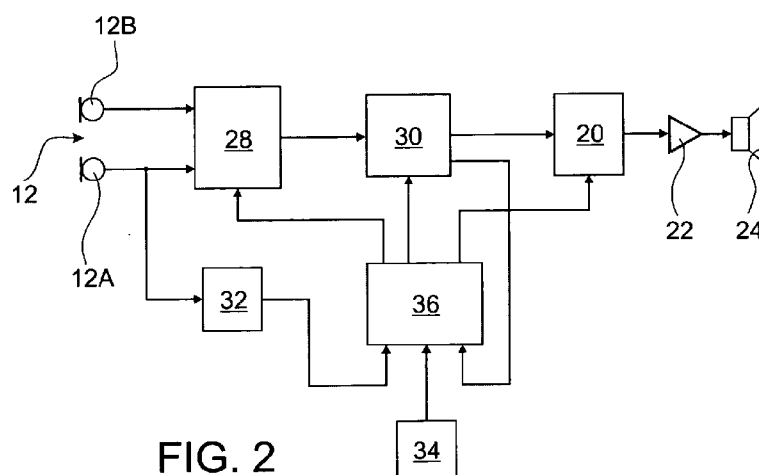


FIG. 2

(57) Abstract: The invention relates to a system for speech enhancement in a room (10), comprising: a microphone arrangement (12) comprising at least two spaced apart microphones (12 A, 12B) for capturing audio signals from a speaker's voice, an acoustic beamforming unit (28) for processing the captured audio signals in a manner so as to impart one of a plurality of different directional patterns to the microphone arrangement; a feedback cancellation unit (30) for applying a feedback cancellation algorithm to the processed audio signals and for providing a feedback status signal indicating how close an acoustic feedback loop (12, 17, 20, 22, 24, 28, 30, 40, 46) of the system is to feedback; means (20, 22) for amplifying the processed audio signals; a loudspeaker arrangement (24) to be located in the room for generating sound according to the amplified audio signals; and means (36) for selecting the directional pattern imparted to the microphone arrangement as a function of the feedback status signal.

Method and system for speech enhancement in a room

The present invention relates to a system for speech enhancement in a room comprising a microphone arrangement comprising at least two spaced-apart microphones for capturing audio signals from a speaker's voice, an acoustic beamformer unit for processing the captured
5 audio signals in a manner so as to impart a directional pattern to the microphone arrangement, a feedback cancellation unit for applying a feedback cancellation algorithm to the processed audio signals, means for amplifying the processed audio signals and a loudspeaker arrangement located in the room for generating sound according to the amplified audio signals.

10 Such a system is described, for example, in WO 2010/000878 A2.

US 2004/0170284 A1 relates to a speech enhancement system, wherein the microphones are provided with a microphone beamformer and a plurality of loudspeakers is provided with an adaptive loudspeaker beamformer, wherein the latter is able to create a beam pattern which is capable of creating a null in the direction of the speaker(s) using the microphones in order to
15 prevent feedback noise.

US 4,489,442 relates to a speech enhancement system comprising a plurality of microphone arrays, each comprising an unidirectional front microphone and an unidirectional rear microphone, which both have a cardioid sensitivity pattern and which are arranged at opposite ends of the array. The microphones also may have other sensitivity patterns such as
20 bidirectional or omni-directional. The system works as a voice activity detector, wherein that microphone array which receives speech is activated, while the other microphone arrays are deactivated.

US 2005/0254640 A1 relates to a speech enhancement system comprising a plurality of directional microphones and a signal processing block including an echo cancellation unit.

25 Hearing aids comprising acoustic beamforming are described, for example, in US 5,473,701, EP 1 005 783 B1, US 6,522,756, EP 1 391 138 A2, EP 1 320 281 A1 and WO 00/68703 A2.

Feedback noise is a major problem in speech enhancement systems, especially when a lapel microphone is used. Feedback limits the gain that can be applied and/or it limits the mobility of the user of the lapel microphone (which may be wireless); also, feedback may cause loud unpleasant whistling. It is known that feedback problems can be reduced, to some extent, by
5 applying a feedback cancellation algorithm in the feedback loop and by the use of directional microphones.

It is an object of the invention to provide for a speech enhancement system in a room, wherein feedback should be reduced while achieving a high signal-to-noise ratio. It is also an object of the invention to provide for a corresponding speed enhancement method.

10 According to the invention, these objects are achieved by a system as defined in claim 1 and by a method as defined in claim 22, respectively.

The invention is beneficial in that, by selecting the directional pattern imparted by the beamforming unit to the microphone arrangement as a function of a feedback status signal which is provided by the feedback cancellation unit and which indicates how close the system
15 is to an acoustic feedback condition, the signal-to-noise ratio may be optimized at low gain conditions, e.g. when the system is sufficiently far away from feedback, for example, by selecting a directional pattern which is optimized for capturing speech from the mouth of the speaker, while the sensitivity of the system to feedback can be reduced at high gain conditions, i.e. when the system is close to feedback, by selecting, for example, a directional
20 pattern which has low sensitivity in the direction of the loudspeaker arrangement.

The system is particularly useful for a lapel microphone arrangement, since lapel microphones are particularly prone to feedback.

The directional pattern selected at low gain conditions may be a cardioid pattern while the directional pattern selected at high gain conditions close to the feedback may be a
25 bidirectional pattern. A cardioid pattern, with the highest sensitivity facing upwards towards the mouth of the speaker and the lowest sensitivity facing downwards, has the advantage that head movements to the left and to the right do not deteriorate the level of the sound picked up by the microphone arrangement too much. The bidirectional pattern has the advantage that it

has a reduced sensitivity, compared to the cardioid pattern, in a horizontal plane; this is particularly useful if the microphone arrangement is in the near field of the loudspeaker arrangement, where most of the sound energy is propagating in a horizontal direction which will be the case if a loudspeaker arrangement is a line array positioned in an upright position at the height of the talker's mouth . However, the bidirectional pattern is more susceptible to changes in the level of the sound picked up by the microphone arrangement in case of head movements to the left or to the right, compared to the cardioid pattern; therefore, the bidirectional pattern should not be used at low gain conditions when the system is sufficiently far away from feedback.

- 10 Preferably, the audio signals captured by the microphone arrangement are transmitted via a wireless link in order to enable free movement of the speaker.

Hereinafter, preferred embodiments of the invention will be illustrated by reference to the attached drawings, wherein:

- 15 Fig. 1 is a schematic block diagram of a speech enhancement system according to the invention;

Fig. 2 is a more detailed block diagram of an example of a speech enhancement system according to the invention; and

Fig. 3 is a block diagram of an example of a wireless speech enhancement system according to the invention.

- 20 Fig. 1 is a schematic representation of a system for enhancement of speech in a room 10. The system comprises a microphone arrangement 12 for capturing audio signals from the voice of a speaker 14. The microphone arrangement 12 comprises at least two spaced apart acoustic sensors/microphones 12A, 12B (see Fig. 2) for achieving a directional pattern of the acoustic sensitivity. The audio signals are supplied to a unit 16 which may provide for pre-amplification of the audio signals and which, in case of a wireless microphone arrangement, includes a transmitter or transceiver for establishing a wireless audio link 17, such as an analog FM link or, preferably, a digital link. The audio signals are supplied, either by wire or, in case of a wireless microphone arrangement, via an audio signal receiver 18 to an audio
- 25

signal processing unit 20 for processing the audio signals, in particular in order to apply a spectral filtering and gain control to the audio signals (alternatively, such audio signal processing, or at least part thereof, could take in the unit 16). The processed audio signals are supplied to a power amplifier 22 operating at constant gain or at an adaptive gain (preferably
5 dependent on the ambient noise level) in order to supply amplified audio signals to a loudspeaker arrangement 24 in order to generate amplified sound according to the processed audio signals, which sound is perceived by listeners 26.

A more detailed example of such a system is shown in Fig. 2, wherein the microphone arrangement 12 consists of two spaced apart microphones 12A and 12B which capture audio
10 signals which are supplied to a beamformer unit 28 for processing the audio signals in a manner so as to impart a certain directional pattern to the microphone arrangement 12. According to the invention, the beamforming unit 28 is adapted to provide for at least two different directional patterns, wherein the presently applied directional pattern is selected according to the feedback status of the system, as will be explained later in more detail. The
15 audio signals as processed by the beamformer unit 28 are supplied to a feedback cancellation unit 30 for applying a feedback cancellation algorithm to the audio signals in order to reduce feedback noise. The feedback cancellation unit 30 also provides for a feedback status signal which indicates how close the system to an acoustic feedback condition. The audio signals as processed by the feedback cancellation unit 30 are supplied to an audio signal processing unit
20 20 which applies spectral filtering and gain control to the audio signals. The audio signals as processed by the audio signal processing unit 20 are supplied to a power amplifier 22 and from there to a loudspeaker arrangement 24.

According to the invention, the selection of the directional pattern imparted by the beamformer unit 28 is controlled by the feedback status signal provided by the feedback
25 cancellation unit 30. The feedback status signal also may serve to select a specific feedback cancellation algorithm in the feedback cancellation unit 30 according to the presently prevailing feedback status of the system.

Typically, the system also comprises an audio signal analyzer unit 32 for analyzing the audio signals as captured by the microphone arrangement 12. Such analyzer unit 32 may comprise a
30 voice activity detector (VAD) for determining whether the user of the microphone

arrangement 12 is presently speaking and an ambient noise level estimator for estimating the ambient voice level. The output signals of the analyzer unit 32 may be used for controlling the audio signal processing in the audio signal processing unit 20, for example by adjusting the gain and/or the spectral filtering according to the information provided by the analyzer unit 32. Typically, the system also comprises a user interface 34 for allowing the users of the system to provide for individual adjustment of the system, such as for adjustment of the desired gain.

Typically, the system comprises a controller 36 for controlling operation of the system. In particular, the controller 36 may receive the output signal of the analyzer unit 32, the user interface 34 and the feedback status signal from the feedback cancellation unit 30 in order to control operation of the beam former unit 28, the feedback cancellation unit 30 and the audio signal processing unit 20 accordingly.

Preferably, the microphone arrangement is a lapel microphone arrangement, and the microphones 12A, 12B preferably are of an omni-directional type. Typically the microphone arrangement 12 will be arranged in such a manner that the imaginary line connecting the two microphones 12A, 12B is oriented substantially vertically, i.e. the microphone arrangement 12 is fixed at the speaker's cloth accordingly. The feedback status signal may be provided by the feedback cancellation unit 30 by estimating the gain of the feedback loop (which in the example of Fig. 2 is formed by the microphone arrangement 12, the electronic component processing the audio signals, such as the units 28, 30, 20, the power amplifier 22 and the loudspeaker arrangement 24).

For example, the feedback status signal may have a first value when the estimated gain of the feedback loop is at or above a predetermined total gain threshold value, with this first value indicating that the system is close to feedback, and a second value when the estimated gain of the feedback loop is below said total gain threshold value, with this second value indicating that the system is not critically close to feedback (feedback is reached when the gain of the feedback loop is unity ("Larsen condition")). The gain of the feedback loop depends on the specific circumstances under which the system is used, such as the manual gain adjustment, the acoustic conditions in the room in which the system is used, the position and orientation of the microphone arrangement 12 and the loudspeaker arrangement 24, etc. A first directional

pattern is selected for audio signal processing the beamformer unit 28 when the feedback status signal has the first value and a second directional pattern is selected when the feedback status signal has the second value.

Typically, the loudspeaker arrangement 24 is formed by an array of loudspeakers which is placed at or close to a wall of the room. Preferably the ratio of the sensitivity for sound impinging in a horizontal plane onto the microphone arrangement 12 and the sensitivity for sound impinging in a vertical direction onto the microphone arrangement 12 is lower for the second direction pattern than for the first directional pattern. in order to reduce pick-up of sound from the loudspeaker arrangement 24 when the system is close to feedback.

Preferably, the first directional pattern is a cardioid pattern, wherein the direction of the highest sensitivity is oriented substantially towards the mouth of the speaker 14. The second directional pattern preferably is a bidirectional pattern (also called "figure 8 pattern"), wherein the direction of the highest sensitivity is oriented substantially towards the mouth of the speaker. Thus, the second directional pattern, which is selected when the system is close to feedback, has a reduced sensitivity to the sound generated by the loudspeaker arrangement 24 (which sound typically has a directional pattern with high contributions in a horizontal plane), whereby the total gain in the feedback loop is reduced, since the microphone arrangement 12, when operated with a bidirectional pattern, picks up less sound from the loudspeaker arrangement 24, thereby enhancing stability against feedback.

It is to be understood that, depending on the specific design of the system, other directional patterns may be utilized.

As already mentioned above, the respective directional pattern of the microphone arrangement 12 is created by the beamformer unit 28 by accordingly processing the audio signal input from the microphones 12A, 12B. For example, a cardioid pattern may be created by a simply delay-and-sum design of the beam former unit 28 (i.e. one of the two microphone signals is delayed before the two signals are combined). A bidirectional pattern may be created, for example, by simply subtracting the signals of the two microphones 12A, 12B (i.e. by adding after multiplying one of the signals by -1 , with no delay being applied). More advanced techniques for beamforming may involve, for example, spatial frequency or the concept of virtual

microphones, see e.g. "Robust phase shift estimation in noise for microphone arrays with virtual sensors" by M. Arcienega, A. Drygajlo, and J. Maisano, at <http://www.eurasip.org/Proceedings/Eusipco/Eusipco2000/sessions/ThuAm/PO1/cr1355.pdf>.

Overviews regarding acoustic beamforming concepts are found, for example, in "Microphone
5 Arrays: A Tutorial" by I. McCowan, at <http://www.idiap.ch/~mccowan/arrays/tutorial.pdf>,
(see also "Robust Speech Recognition using Microphone Arrays", by I. McCowan, PhD
Thesis, Queensland University of Technology, Australia 2001) and "Microphone Arrays", M.
Brandstein and D. Ward (Eds.), Springer, 2001.

Preferably, the selection of the directional pattern imparted by the beam former unit 28
10 employs some hysteresis, i.e. the value of the feedback status signal at which the beam former
unit 28 switches from one directional pattern to the other depends on the direction of the
switching, i.e. the threshold values may be different depending on whether the system
switches from the first pattern to the second pattern (i.e. when the feedback loop is found to
increase) or whether the beam forming unit 28 switches from the second pattern to the first
15 pattern (i.e. when the gain in the feedback loop is found to decrease).

As already mentioned above, the feedback cancellation unit 30 may apply different feedback
cancellation algorithms as a function of the estimated gain in the feedback loop. For example,
a time domain feedback cancellation algorithm may be selected when the gain in feedback
loop is below a certain threshold value, and a frequency domain feedback cancellation
20 algorithm may be selected when the gain in the feedback loop is at or above that threshold
value. The advantage of time domain feedback cancellation is that there is no delay of the
audio signals due to the signal processing in the frequency domain. However, since frequency
domain feedback cancellation tends to be more efficient, frequency domain feedback
cancellation is preferably employed at high gain in the feedback loop. Typically, the selection-
25 switching will employ some kind of hysteresis, for example 3 dB with regard to the estimated
gain in the feedback loop. The frequency domain feedback cancellation algorithm may apply a
Wiener filter to the audio signals. In particular, the frequency domain feedback cancellation
algorithm may estimate the transfer function of the feedback loop and apply a filter
corresponding to the inverse estimated transfer function to the audio signals in order to

eliminate the signal parts caused by feedback. Of course, also other feedback cancellation algorithms may be used, as it is known in the art.

In Fig. 3 a block diagram of a wireless speech enhancement system is shown, wherein the microphone arrangement 12 is connected to a transmission unit 16 comprising a beamformer unit 28, an audio signal analyzer unit 32, a user interface 34 and a controller 36, with these elements having the same functionality as in the system shown in Fig. 2. In addition, the transmission unit 16 comprises a gain model unit 38, a digital transceiver 40 and an antenna 42. The audio signals as processed by the beamformer unit 28 are supplied to the gain model unit 38 in order to apply a suitable gain model to the audio signals (typically a gain model wherein the gain is reduced at low and high input levels relative to medium input levels). The output of the gain model unit 38 is supplied to the digital transceiver for sending the audio signals via the digital link 17 to a receiver unit 18. In addition, also the output of the analyzer unit 32 and an output of the controller 36 concerning commands/data received from the user interface 34 are supplied to the transceiver 40 in order to transmit corresponding data/commands to the receiver unit 18.

The receiver unit 18 comprises an antenna 44 and a transceiver 46 for receiving the audio signals and other data and commands transmitted from the transmission unit 16 via the link 17. The received audio signals are supplied to a feedback cancellation unit 30 which corresponds to the feedback cancellation unit 30 of the system of Fig. 2. The feedback status signal provided by the feedback cancellation unit 30 is supplied to the transceiver 46 in order to transmit the feedback status signal to the transceiver 40 of the transmission unit 16, from where the feedback status signal is supplied to the controller 36 in order to use it for controlling the beamformer unit 28. The output of the feedback cancellation unit 30 is supplied to an audio signal processing unit 20, from where the audio signals are supplied via a power amplifier 22 to a loudspeaker arrangement 24, with these elements corresponding to the respective elements of the system of Fig. 2. The data and commands resulting from the analyzer unit 32 and controller 36 as received via the link 17 are supplied from the transceiver 46 to the feedback cancellation unit 30 and the audio signal processing unit 20.

Claims

1. A system for speech enhancement in a room (10), comprising:

a microphone arrangement (12) comprising at least two spaced apart microphones (12A, 12B) for capturing audio signals from a speaker's voice,

an acoustic beamforming unit (28) for processing the captured audio signals in a manner so as to impart one of a plurality of different directional patterns to the microphone arrangement;

a feedback cancellation unit (30) for applying a feedback cancellation algorithm to the processed audio signals and for providing a feedback status signal indicating how close an acoustic feedback loop (12, 17, 20, 22, 24, 28, 30, 40, 46) of the system is to feedback;

means (20, 22) for amplifying the processed audio signals;

a loudspeaker arrangement (24) to be located in the room for generating sound according to the amplified audio signals; and

means (36) for selecting the directional pattern imparted to the microphone arrangement as a function of the feedback status signal.
2. The system of claim 1, wherein the microphone arrangement (12) is a lapel microphone arrangement.
3. The system of claim 1 or 2, wherein the microphones (12A, 12B) of the microphone arrangement (12) are of an omnidirectional type.
4. The system of one of the preceding claims, wherein the beamforming unit (28) is of a delay-and-sum-type.
5. The system of one of the preceding claims, wherein the microphone arrangement (12) forms part of or is connected to a transmission unit (16) for transmitting audio signals via a wireless link (17) to a receiver unit (18) forming part of or being connected to the loudspeaker arrangement (24).
6. The system of claim 5, wherein the beamformer unit (28) forms part of the transmission unit (16) and the feedback cancellation unit (30) forms part of the receiver unit (18),

wherein the receiver unit comprises means (44, 46) for transmitting the feedback status signal via the wireless link (17) to the transmission unit.

7. A method of speech enhancement in a room (10), comprising:

capturing audio signals from a speaker's voice by a microphone arrangement (12) comprising at least two spaced apart microphones (12A, 12B);

applying acoustic beamforming processing to the captured audio signals in a manner so as to impart one of a plurality of different directional patterns to the microphone arrangement;

applying a feedback cancellation algorithm to the processed audio signals;

generating sound according to the processed audio signals by a loudspeaker arrangement (24);

providing a feedback status signal indicating how close an acoustic feedback loop (12, 17, 20, 22, 24, 28, 30, 40, 46) of the system to feedback; and

selecting the directional pattern imparted to the microphone arrangement as a function of the feedback status signal.
8. The system of one of the preceding claims, wherein the selection of the directional pattern to be imparted to the microphone arrangement (12) employs a hysteresis.
9. The system of one of the preceding claims, wherein the feedback status signal has a first value indicating that the system is critically close to the feedback condition and a second value indicating that the system is not critically close to the feedback condition, wherein a first directional pattern is selected when the feedback status signal has the first value and a second directional pattern is selected when the feedback status signal has the second value.
10. The system of one of the preceding claims, wherein the feedback status signal is provided by estimating the gain of the feedback loop (12, 17, 20, 22, 24, 28, 30, 40, 46) of the system.
11. The system of claims 9 and 10, wherein the feedback status signal has the first value when the estimated gain of the feedback loop (12, 17, 20, 22, 24, 28, 30, 40, 46) of the

system is at or above a predetermined gain threshold value and wherein the feedback status has the second value when the estimated gain of the feedback loop of the system is below said gain threshold value.

12. The system of one of claims 9 and 11, wherein the ratio of the sensitivity for sound impinging in a horizontal plane onto the microphone arrangement (12) and the sensitivity for sound impinging in a vertical direction onto the microphone arrangement is lower for the second directional pattern than for the first directional pattern.
13. The system of claim 12, wherein the first directional pattern is a cardioid pattern.
14. The system of claim 13, wherein the direction of the highest sensitivity of the cardioid pattern is oriented substantially towards the mouth of the speaker (14).
15. The system of one of claims 12 to 14, wherein the second directional pattern is a bidirectional pattern.
16. The system of claim 15, wherein the direction of the highest sensitivity of the bidirectional pattern is oriented substantially towards the mouth of the speaker (14).
17. The system of one of the preceding claims, wherein the microphone arrangement (12) is arranged such that an imaginary line connecting the two microphones (12A, 12B) of the microphone arrangement is oriented substantially vertically.
18. The system of one of the preceding claims, wherein a feedback cancellation algorithm is selected from a plurality of feedback cancellation algorithms as a function of the feedback status signal.
19. The system of claim 18, wherein a time domain feedback cancellation algorithm is selected when the estimated gain of the feedback loop (12, 17, 20, 22, 24, 28, 30, 40, 46) of the system is below a threshold value and a frequency domain feedback cancellation algorithm is selected when the estimated gain of the feedback loop of the system is at or above said threshold value.
20. The system of claim 19, wherein the frequency domain feedback cancellation algorithm applies a Wiener filter to the audio signals.

21. The system of one of claims 19 and 20, wherein the frequency domain feedback cancellation algorithm estimates the transfer function of the feedback loop (12, 17, 20, 22, 24, 28, 30, 40, 46) of the system and applies a filter corresponding to the inverse estimated transfer function to the audio signals in order to eliminate signal parts caused by feedback.
22. The system of one of claims 18 to 21, wherein the selection of the feedback cancellation algorithm employs a hysteresis.
23. The system of one of the preceding claims, wherein the loudspeaker arrangement (24) is positioned at or close to a wall of the room (10).

