(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 2 490 459 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
22.08.2012 Bulletin 2012/34

(51) Int Cl.:
H04R 3/00 *(2006.01)* *H04R 3/02* *(2006.01)*

(21) Application number: **11155021.6**

(22) Date of filing: **18.02.2011**

(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
Designated Extension States:
BA ME

(71) Applicant: **Svox AG**
8048 Zürich (CH)

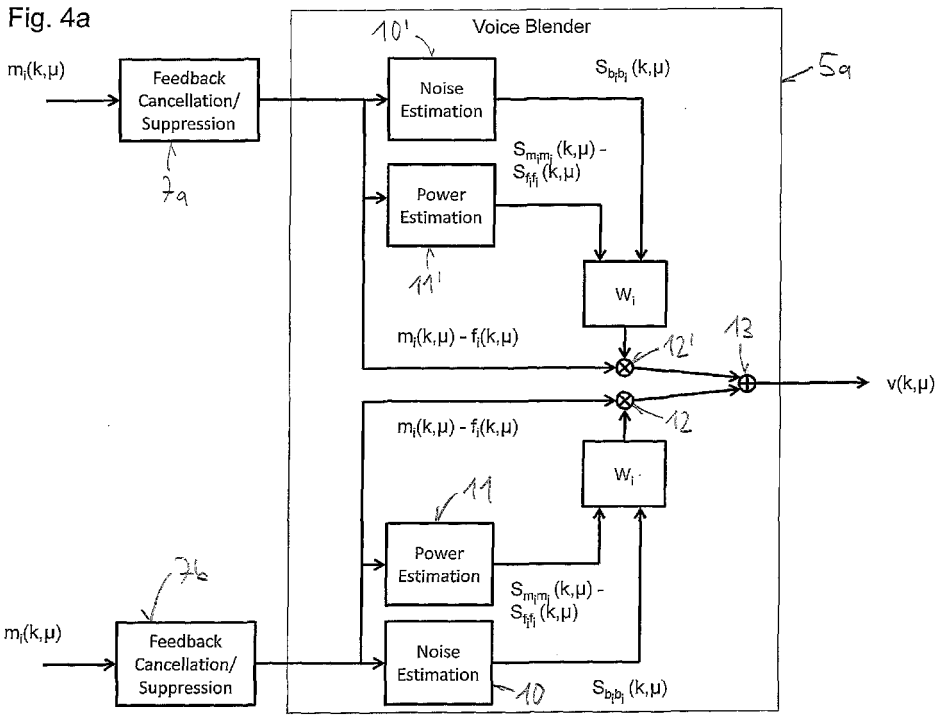
(72) Inventors:
- **Iser, Bernd**
890777 Ulm (DE)
- **Wolf, Arthur**
89073 Ulm (DE)
- **Hannon, Patrick**
89077 Ulm (DE)
- **Krini, Mohamed**
89075 Ulm (DE)

(74) Representative: **Stocker, Kurt**
Büchel, von Révy & Partner
Zedernpark
Bronschhoferstrasse 31
9500 Wil (CH)

(54) **Method for voice signal blending**

(57) In a method for voice signal blending in an indoor communication system, particularly in a vehicle, or in a hands-free telephony system or an automatic speech recognition system, there are at least two microphones, the voice signals of which should be blended to be delivered to at least one loudspeaker. Moreover, feedback suppression or compensation is provided. This feedback suppression or compensation is effected before blending the signals of the at least two microphones.

Fig. 4a



Printed by Jouve, 75001 PARIS (FR)

Description

Field of the invention

[0001] The invention refers to a method and system installed in a car for the communication of people sitting in remote locations. Therefore at least one loudspeaker is installed and at least two microphones, one assigned to each person. Alternatively, the invention relates to method applied to a hands-free telephony system or to an automatic speech recognition system, where conditions and requirements are quite similar. More specifically, the invention relates to a method according to the introductory clause of claim 1 as well as to a software product according to claim 8 and a system according to claim 9.

Background of the invention

[0002] EP1850640B1 shows and describes a typical setup for a basic intercom system with two microphones and one loudspeaker. The mixer in this setup decides depending on a criterion which of the microphone signals shall be switched to the next module or the output. This criterion can depend on the detected occupation of a seat. However, it has been also proposed in US006549629B2 such decision criterion can depend on the Signal plus Noise to Noise Ratio (SNNR).

[0003] The fundamental disadvantage of such a system is that the feedback is still contained in the signal and therefore SNNR is not a good approach to a Signal to Noise Ratio (SNR). Using SNNR as SNR causes a misinterpretation of the signal power when estimating the SNNR for each microphone. The result of such a misinterpretation is the preferred usage of the microphone containing the highest amount of feedback.

[0004] A further disadvantage of these systems is a possible change in background-noise level when switching from one speaking person to another (e.g. driver speaks to a passenger in the back while having his window open, and then the co-driver speaks to the passenger in the back having his window closed and driver as well as co-driver are each equipped with a microphone). This change in background-noise level of the signal played back over the loudspeaker might be experienced as unpleasant by a listener.

[0005] Another disadvantage that occurs is, when the play-back level between two speaking persons differs. This can occur if for example the driver is quite tall and therefore the distance to the microphone is little, which means significant speech level in the microphone signal. If for example the co-driver is a small person and therefore the distance to the microphone is large, which means low speech level within the microphone signal this basic setup results in a playback signal for the loudspeaker that differs in speech level for the driver and co-driver in an unpleasant manner concerning the listener impression.

[0006] US-A-2005/0265560 discloses a system, where a beamformer is used for microphone signal blending. The, thus, blended output signal is then subjected to feedback suppression. This has the disadvantage that the power of the signals to be blended is relative high and, moreover, the output signal of the beamformer is adulterated by the feedback component of the signal. The suppression, however, is merely the suppression of frequencies, which are just developing resonance oscillations, which means that feedback components remain, and only resonance oscillations are prevented.

Summary of the invention

[0007] It is an object of the present invention to improve the quality of the blended signal and to find a method that is more robust in relation to feedback components.

[0008] This object is achieved by the measure according to the characterizing clause of claim 1. The feedback suppression or compensation is effected before blending the signals of the at least two microphones. If in this connection the term "feedback suppression or compensation" is used, it is in any case a minimization of feedback components of a signal. A feedback compensation is for example described in EP 1 679 874 B1. Feedback suppression can be applied by filtering, for example with a notch-filter. Another known compensation method uses frequency shifting. These suppression methods reduce the development of feedback, but existing feedback components remain.

[0009] The method for voice signal blending is applied in a communication system, such as an indoor communication system, particularly in a vehicle, or in a hands-free telephony system or an automatic speech recognition system, comprising at least two microphones and at least one loudspeaker. The microphone signals are blended with respective weights to be delivered to the at least one loudspeaker. The feedback suppression or compensation is effected before blending the signals of the at least two microphones. This improves the quality of the blended signal and is more robust in relation to feedback components.

[0010] In preferred embodiments the feedback components are minimized by estimating the feedback signal or its energy level, preferably its power spectral density, and applying a Wiener filter which eliminates the estimated feedback signal from the signals to be blended.

[0011] According to a preferred embodiment, for at least a sub-band of the microphone signals the energies of the microphone signals are determined and at least the energies of the noise components are estimated, wherein for blending, a higher weight is given to at least the sub-band with the highest ratio of signal energy to noise energy. Sub-dividing the entire frequency band into sub-bands enables it to give individual sub-bands a different weight. Blending will be made for each sub-band with the respective weights.

[0012] It is favourable, if a signal level adjustment of the at least two microphone signals to a predetermined value is carried out particularly immediately before or immediately after blending, wherein the adjustment is depending on the energies of the noise components and minimizes the perceivable difference in the noise level when blending from one microphone to the other. In this way changes of noise levels are better avoided.

[0013] There can be a perceivable difference in the speech level when blending from one microphone to the other. To solve this problem, the signal levels of the microphone signals are adjusted to a predetermined value so that there is no perceivable difference in the speech level when blending from one microphone to the other.

[0014] The noise level of at least two microphones can be unpleasantly different. An adjustment can be achieved, if the characteristics of claim 5 are fulfilled.

[0015] The microphone signals comprise different components, each with a certain level. If a person close to a microphone is speaking, then the signal of this microphone comprises this voice with a corresponding voice level. In a preferred embodiment the voice levels of the at least two microphone signals are adjusted substantially to the same level by adjusting the microphone signals. Then the difference of the noise levels of the at least two adjusted microphone signals is determined. The noise suppressions in the microphone signals are controlled by adapting the parameters of the noise suppression characteristic in such a way, that the noise suppressed signals have substantially the same level of residual noise and the same level of speech signal in it for each microphone signal.

[0016] In a preferred embodiment the respective speaking and/or listening, non-speaking party is detected by one of the following measures:

a) by analyzing the signal of the at least two microphones during blending;
b) by sensing by means of vehicle sensors, e.g. for the seat occupancy.

[0017] In further preferred embodiments at least one of the following measures is taken:

a) the signal weight of the microphone signal where no speaking party is detected is reduced, when blending;

b) predetermined amplifier characteristics of the gain control, e.g. noise dependent gain control, for listeners and/or speakers are provided and that amplifier characteristic is chosen, which corresponds to the position of the respective listener and/or speaker;

c) equalizing is effected by means of an equalizer after voice signal blending and that the settings of the equalizer are chosen depending on the detection so as to enhance the perceived quality of the output for the listening and/or speaking party.

Brief description of the drawings

[0018] Further details and advantages will become apparent from the following description of embodiments with reference to the drawings, in which

Fig. 1 shows a block diagram of a typical setup for a basic intercom system with two microphones and one loudspeaker;

Fig. 2a is a block diagram of a first embodiment according to the invention, while

Fig. 2b shows a second embodiment according to the invention, and

Fig. 3 depicts a third embodiment according to the invention;

Fig. 4a illustrates a circuit for voice blending according to a first structure according to the invention;

Fig. 4b is a circuit for voice blending according to a second structure according to the invention; and

Fig. 4c shows a circuit for voice blending according to a third structure according to the invention;

Fig. 5 illustrates setting an equalizer after voice blending; and

Fig. 6 depicts a sample of a characteristic used for chosing a gain factor depending on a noise estimate..

Detailed description of the invention

[0019] Fig. 1 shows a typical setup for a basic intercom system with two microphones 1 and 2 and one loudspeaker 3. Between a pre-processing stage 4, directly connected to the microphones 1, 2 and a post-processing stage 6, leading to the loudspeaker 3 is a voice mixer 5 which receives the pre-processed signals of the microphones 1, 2 and decides depending on a criterion which of the microphone signals shall be switched to its output and to the next module 6.

[0020] In Fig. 2a, one embodiment of the invention is illustrated, where the signals of the microphones 1, 2, after an optional pre-processing stage 4a, are subjected to feedback suppression or compensation in a stage 7. Thus, the first component after some optional pre-processing is a feedback suppression or feedback compensation at 7. This module 7 suppresses or compensates for the portion of the loudspeaker signal that is coupling back into the microphone and therefore being an undesired signal component.

[0021] Before reaching the voice blender 5a (Fig. 2a) or after the blender 5a (Fig. 2b), noise suppression in a noise suppression module 8 can be effected. This module 8 suppresses the background noise components of the microphone signals resulting from the background noise being present in the car cabin and being picked up by the microphones. A practical embodiment with a noise suppression postponed to the voice blender 5a is shown in Fig. 4b.

[0022] The signal m_i of microphone i is made up of the signal from the speaker in the cabin s_i, the feedback of the system from the loudspeaker into the microphone f_i and the background noise of the cabin recorded by the microphone b_i.

[0023] This is suitably depicted by the following

Equation 1

$$m_i(k,\mu) = s_i(k,\mu) + f_i(k,\mu) + b_i(k,\mu)$$

[0024] Where i corresponds to the microphone index, k to the time interval and μ to the frequency band.

[0025] In the following the noise suppression is located before the voice blender 5a which is not necessarily required (without limitation of generality, see e.g. Fig. 4b).

[0026] In any case (vide Figs. 2a, 2b), the blended and noise suppressed signal goes suitably to a NDGC or noise dependent gain control 9.

Voice decision

[0027] The voice blender module blends the at least two feedback suppressed or compensated and noise suppressed signals according to Fig. 2a together to one output signal following below described criterion. According to Fig. 2b the blended signals are only feedback suppressed or compensated.

[0028] The blending is made by giving weights to the microphones. The blending or weighting criterion can either be evaluated in a non-frequency selective manner or in a frequency selective manner resulting in a non-frequency selective weighting or in a frequency selective weighting of the microphone signals ${}^{m}i$. The criterion used in this invention for the weighting according to Fig. 4a, 4b, 4c is a function of the ratio of the energy of every microphone signal S_{mimi} subtracted by the feedback f_i (respectively S_{fifi} for the energy of the feedback component at every microphone) due to the coupling of the loudspeaker signal back into the microphone signal and further subtracted by the noise b_i (respectively S_{bibi} for the energy of the noise component at every microphone) within the microphone signal resulting from the noise present in the car cabin and being picked up by the microphone (for the setup according to Fig. 2b each microphone signal is only subtracted by the feedback) to the estimated noise present in each microphone signal. Double indices are used for the energy of the signal, and for power spectral densities.

Equation 2

$$w_i(\mathrm{k},\mu) = f\left(S_{mm}(k,\mu), S_{ff}(k,\mu), S_{bb}(k,\mu)\right)$$

[0029] For k being the time block index and μ being the sub-band index. The output of the voice blender 5a is denoted with:

$$v(k,\mu) = \sum_{i=0}^{N-1} w_i(\mathrm{k},\mu)\left(m_i(k,\mu) - f_i(k,\mu) - b_i(k,\mu)\right)$$

[0030] With N being the number of microphones.

[0031] For estimating the power of the microphone signal a first order IIR filter for smoothing can be applied according to

Equation 3

$$S_{m_i m_i}(k,\mu) = \beta_m |m_i(k,\mu)|^2 + (1-\beta_m) S_{m_i m_i}(k-1,\mu)$$

[0032] Where $ß_m$ is a smoothing constant that has to be chosen between zero and one. This means the current short-term power is weighted with $ß_m$ and the estimate of the previous time frame is weighted with (1-$ß_m$).

[0033] The feedback component $S_{fifi}(k,\mu)$ is estimated in the feedback compensation or feedback suppression module and is therefore given.

[0034] The noise component can be estimated using a minimum tracker according to

Equation 4

$$S_{b_i b_i}(k,\mu) = \min\left\{S_{m_i m_i}(k,\mu), S_{b_i b_i}(k-1,\mu)\right\}(1+\varepsilon)$$

[0035] Where ε is a small number depending on the sampling rate. A good choice for a sampling rate of 44.1 kHz is e.g. epsilon=0.00001. This means that the minimum of the current time frame and the previous one is weighted by (1+epsilon) resulting in a signal that tracks local minima of the microphone signal power for estimating the noise signal power.

[0036] One possible implementation of $w_i(\mathrm{k},\mu)$ is:

Equation 5

$$w_i(\mathrm{k},\mu) = \frac{S_{m_i m_i}(k,\mu) - S_{f_i f_i}(k,\mu) - S_{b_i b_i}(k,\mu)}{S_{b_i b_i}(k,\mu)\sum_i \frac{S_{m_i m_i}(k,\mu) - S_{f_i f_i}(k,\mu) - S_{b_i b_i}(k,\mu)}{S_{b_i b_i}(k,\mu)}}$$

[0037] This means that the modified SNR, after subtracting the power of the feedback signal and of the background

noise from the microphone signal power, is set into relation with the sum over all microphones of the modified SNR. The resulting weight is more robust against misinterpretations of feedback power as desired speech power as would be the case using an SNNR as described in US006549629B2.

Perception enhancement

[0038] Further criteria that can be included in the above weights are:

- Adjusting the signal before or after the blending in terms of level to a common background noise level so that there is no perceivable difference for the listening party considering background-noise level when switching from one microphone to the other, e.g. using

Equation 6

$$w_{i,noise}(\mathrm{k},\mu) = w_i(\mathrm{k},\mu)\frac{S_{bb,desired}(k,\mu)}{\frac{1}{M}\sum_{\mu=0}^{M-1} S_{b_i b_i}(k,\mu)}$$

Where M is the amount of sub-bands μ and $S_{bb,desired}(k,\mu)$ is the desired noise level. This means applying a gain to the signal with lower noise level than the desired one to reach the desired noise level.

- Adjusting the signal before or after the blending in terms of level to a common speech level so that there is no perceivable difference for the listening party considering speech level when switching from one microphone to the other.

Equation 7

$$S_{s_i s_i}(k,\mu) = S_{m_i m_i}(k,\mu) - S_{f_i f_i}(k,\mu) - S_{b_i b_i}(k,\mu)$$

Equation 8

$$w_{i,speech}(\mathrm{k},\mu) = w_i(\mathrm{k},\mu)\frac{S_{ss,desired}(k,\mu)}{\frac{1}{M}\sum_{\mu=0}^{M-1} S_{s_i s_i}(k,\mu)}$$

Where $S_{ss,desired}(k,\mu)$ is the desired speech level.

[0039] A combination of the above mentioned criteria with some trade-offs is permitted as well. A possibility of combining both criteria at the same time without having to trade off is described in the following.

Speaker individual noise suppression parameterization

[0040] For obtaining the same noise level for each speaker, the voice blender module 5a determines the difference in noise levels between the different signals and causes noise suppression to adapt the parameters of the noise suppression characteristic to result in a noise suppressed signal having the same level of residual noise in it for each speaker.

[0041] When having in mind that we as well want to have a balanced speech level between the different speakers we have to evaluate the differences in noise level after having balanced the signals for their speech level and use the resulting differences in noise level to adjust the parameters of the noise suppression module for each speaker.

[0042] This allows compensating for different background-noise levels and different speech levels at the same time. No trade off between the both optimization criteria is necessary. This means applying Equation 8 and adjusting the filter

coefficients of the noise suppression (e.g. a Wiener filter) to (see Fig. 4b):

Equation 9

$$H(k,\mu) = \min\left\{\beta, 1 - \sum_i w_{i,speech}(\mathrm{k},\mu)\frac{S_{b_i b_i}(k,\mu)}{S_{m_i m_i}(k,\mu)}\right\}$$

With

Equation 10

$$\beta = \beta_{NR}\frac{\frac{1}{M}\sum_{\mu=0}^{M-1}\sum_i w_{i,speech}(\mathrm{k},\mu)S_{s_i s_i}(k,\mu)}{S_{ss,desired}(k,\mu)}$$

And β_{NR} being the regular spectral floor or maximum attenuation of the Wiener filter (typically β_{NR} = —10dB).

Joint information usage

[0043] Fig. 4a shows, how a voice blender for carrying out the method according to the invention may be structured. It is clearly visible that feedback suppression or feedback compensation is done for each of the signals stemming from the microphones 1 and 2 in modules 7a and 7b. The output of these modules 7a, 7b is delivered to the voice blender 5a, i.e. to noise estimation modules 10, 10' (see equations 1-4), to power estimation modules 11, 11' (see equation 8) and to multiplicators 12 and 12' for weighting the incoming signals.

[0044] To this end, the multiplicators 12, 12' are each controlled by a module Wi which determines the blender weights for each signal of the microphones 1, 2. At the output of blender 5a is a summing point 13.

[0045] It is clear, that the representation as distinct blocks 10, 11, 12 does not mean that these blocks have to be realized as distinctive devices. In practice, all these activities or at least part of them will be handled by software.

[0046] As has been mentioned above, the output signal v(μ,k) of the voice blender 5a can be used to influence noise suppression in module 8. This is shown in Fig. 4b, where the weighting signals $W_{i,speech}(k,\mu)$ are supplied to module 8. Noise suppression down to zero gives mostly a bad feeling to the listener, for which reason it is preferred to adjust noise suppression to a predetermined level to a predetermined minimum level. This level typically equals 0.316 meaning a maximum suppression of -10 dB.

[0047] Alternatively or in combination with the above-mentioned measures (see also Fig. 3 which shows the information exchange between the modules in dotted lines), a configuration according to Fig. 4c is possible within the scope of the present invention taking equation 8 in mind. According to this figure, the weighting signals $W_{i,speech}(k,\mu)$ are supplied to module 9, the noise dependent gain control, which, to have an information of noise present in the signal, receives also the output signal $S_{bibi}(k,\mu)$ of the noise estimation modules 10 and 10'.

[0048] In a further modification according to Fig. 5, the weighting signal W_i of the voice blender 5a could be used to control an equalizer 6a which forms part of the post-processor 6 (Fig. 1 to Fig. 2b).

[0049] To figures 3, 4c and 5, the following details are now given:

(i) Noise dependent gain control

[0050] The decision of the voice blender module is used for the noise dependent gain control module. The noise dependent gain control module is responsible for increasing the level of the output signal by applying a gain $g_{\ell,NDGC}(S_{b\ell b\ell}(k,\mu))$,depending on the level of noise perceived by the listening party. This increase of the output signal level depends on a characteristic which maps the level of background noise in the car cabin to a gain applied to the output signal. The characteristic chosen for this mapping depends on which speaker is active. This information is delivered by the voice blender module in the form of the microphone index of the most active microphone $\ell(k)$.

Equation 11

$$\ell(k) = arg\left\{\max_i\left\{\sum_{\mu=0}^{M-1} w_{i,speech}(\mathrm{k},\mu)\right\}\right\}$$

Equation 12

$$g_{\ell,NDGC}\left(S_{b_\ell b_\ell}(k,\mu)\right) = \begin{cases} 0, for\ S_{b_\ell b_\ell}(k,\mu) \le S_{bb,\ell,low}(k,\mu) \\ a_\ell S_{b_\ell b_\ell}(k,\mu), for\ S_{bb,\ell,low}(k,\mu) < S_{b_\ell b_\ell}(k,\mu) < S_{bb,\ell,high}(k,\mu) \\ g_{\ell,NDGC,max}\left(S_{b_\ell b_\ell}(k,\mu)\right), for\ S_{b_\ell b_\ell}(k,\mu) \ge S_{bb,\ell,high}(k,\mu) \end{cases}$$

Equation 13

$$a_\ell = \frac{g_{\ell,NDGC,max}\left(S_{b_\ell b_\ell}(k,\mu)\right)}{S_{bb,\ell,high}(k,\mu) - S_{bb,\ell,low}(k,\mu)}$$

Where $g_{\ell,NDGC,max}\,(S_{b\ell b\ell}(k,\mu))$ represents the maximum gain of the NDGC (e.g. 10dB). So all properties of the different NDGC characteristics are described by $g_{\ell,NDGC,max}\,(S_{b\ell b\ell}(k,\mu))$, $S_{bb,\ell,high}(k,\mu)$ and $S_{bb,\ell,low}(k,\mu)$.

[0051] In Equation 12, $S_{b\ell b\ell}(k,\mu)$ represents the noise component that is present in the microphone signal of the active speaker. Some sample characteristics are depicted in Fig. 6.

(ii) Equalizer

[0052] As has been mentioned above, part of post processing 6 of an intercom system is the equalizer 6a. This equalizer 6a can as well be implemented as a multi-channel equalizer individual for each channel. The settings of the equalizer 6a are chosen depending on the voice blender decision of which person is speaking to enhance the perceived quality of the output for the listening party (see Fig.3 and Fig. 5).

(iii) Further optional modifications

[0053]

- A multi-microphone array can be used for beamforming. This allows for localization of the speaker in the car cabin (similar to the decision $\ell(\mathrm{k})$ of the voice blender). This information is used by the voice blender to decide which speaker is active.
- Detection of position of listening party (e.g. by weight sensors in the seats) to further enhance joint information usage.

    o Voice blender knows where no speaker is and can set weights accordingly.

[0054] Furthermore by knowing who speaks and who listens, NDGC mapping characteristic and equalizer setting can be chosen in an even more specific manner.

[0055] Numerous modifications can be made within the scope of the invention; for example instead of voice blending, beamforming could be performed, as has been indicated above, so that the expression "blending" has to be understood in a broader sense.

Claims

1. Method for voice signal blending in a communication system, such as an indoor communication system, particularly in a vehicle, or in a hands-free telephony system or an automatic speech recognition system, comprising at least two microphones and at least one loudspeaker, wherein the microphone signals are blended with respective weights to be delivered to the at least one loudspeaker and feedback suppression or compensation is provided, **characterized in that** feedback suppression or compensation is effected before blending the signals of the at least two microphones.

2. Method according to claim 1, **characterized in that** for at least a sub-band of the microphone signals the energies of the microphone signals are determined and at least the energies of the noise components are estimated,, wherein for blending, a higher weight is given to at least the sub-band with the highest ratio of signal energy to noise energy.

3. Method according to claim 1 or 2, **characterized in that** a signal level adjustment of the at least two microphone signals to a predetermined value is carried out particularly immediately before or immediately after blending, wherein the adjustment is depending on the energies of the noise components and minimizes the perceivable difference in the noise level when blending from one microphone to the other.

4. Method according to any of the preceding claims, **characterized in that** signal levels of the microphone signals are adjusted to a predetermined value so that there is no perceivable difference in the signal level when blending from one microphone to the other.

5. Method according to claim 4, **characterized in that** voice levels of the at least two microphone signals are adjusted substantially to the same level, then the difference of the noise levels of the at least two adjusted microphone signals is determined and the noise suppression is controlled by adapting the parameters of the noise suppression characteristic, to generate a noise suppressed signal having substantially the same level of residual noise and the same level of speech signal in it for each microphone signal.

6. Method according to any of the preceding claims, **characterized in that** the respective speaking and/or listening, non-speaking party is detected by one of the following measures:

    a) by analyzing the signal of the at least two microphones during blending;
    b) by sensing by means of vehicle sensors, e.g. for the seat occupancy.

7. Method according to claim 6, **characterized in that** at least one of the following measures is taken after detection:

    a) the signal weight of the microphone signal where no speaking party is detected is reduced, when blending;
    b) predetermined amplifier characteristics of the gain control, e.g. noise dependent gain control, for listeners and/or speakers are provided and that amplifier characteristic is chosen, which corresponds to the position of the respective listener and/or speaker;
    c) equalizing is effected by means of an equalizer after voice signal blending and that the settings of the equalizer are chosen depending on the detection so as to enhance the perceived quality of the output for the listening and/or speaking party.

8. Software product which executes the method according to any of the preceding claims.

9. Communication system, such as an in-door communication system, a hands-free telephony system or an automatic speech recognition system, comprising at least one loudspeaker and at least two microphones, as well as a signal treatment device, which carries out a method according to any of claims 1 to 7.

Fig. 1

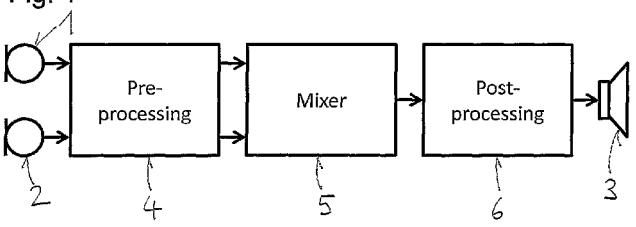


Fig. 2a

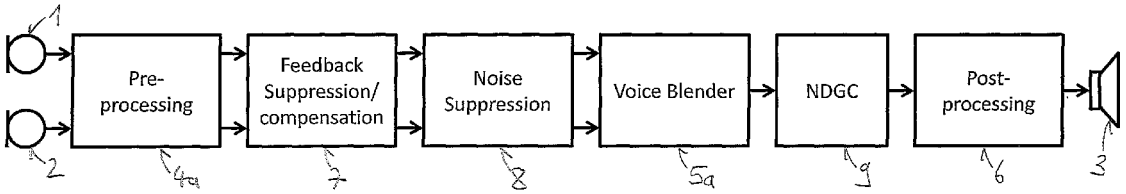


Fig. 2b

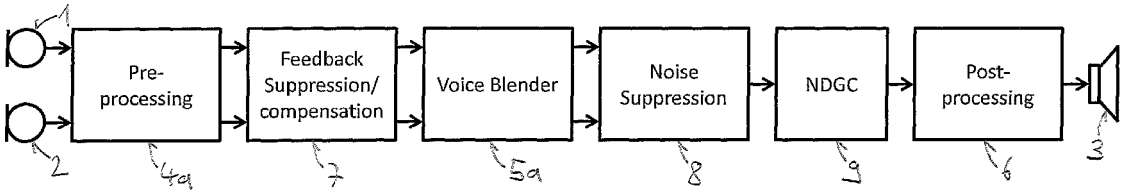


Fig. 3

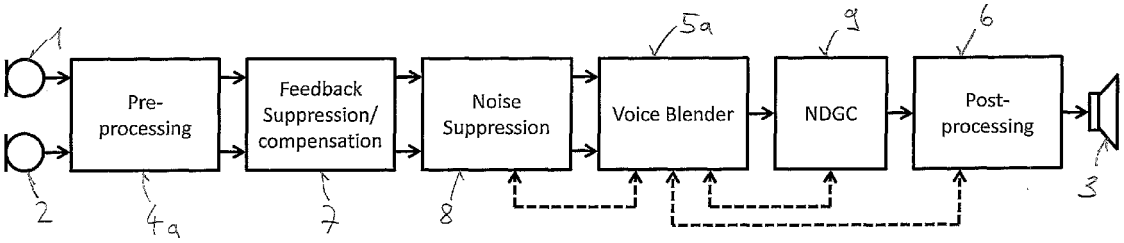

1
2
Pre-processing
4a
Feedback Suppression/ compensation
7
Noise Suppression
8
5a
Voice Blender
9
NDGC
6
Post-processing
3

Fig. 5

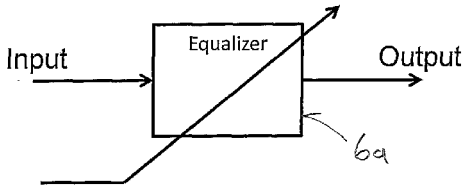

Input
Equalizer
Output
6a
w_i
from voice blender

Fig. 4a

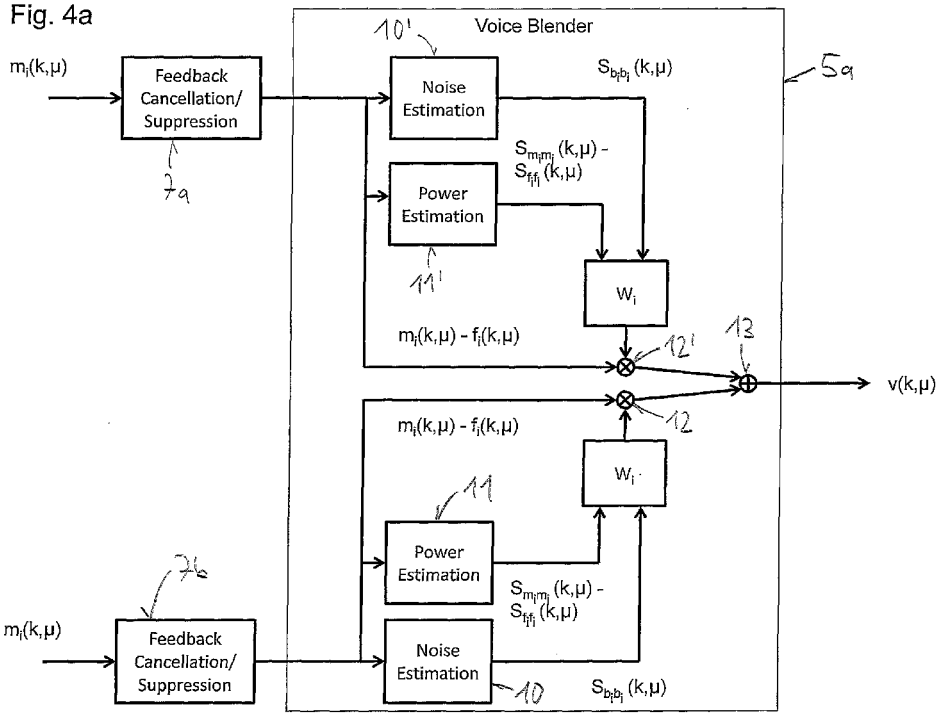

Voice Blender
10'
5a
$m_i(k,\mu)$
Feedback Cancellation/ Suppression
7a
Noise Estimation
$S_{b_ib_i}(k,\mu)$
$S_{m_im_i}(k,\mu) - S_{f_if_i}(k,\mu)$
Power Estimation
11'
W_i
$m_i(k,\mu) - f_i(k,\mu)$
12'
13
$v(k,\mu)$
12
W_i
11
Power Estimation
7b
$m_i(k,\mu)$
Feedback Cancellation/ Suppression
Noise Estimation
10
$S_{b_ib_i}(k,\mu)$

Fig. 4b

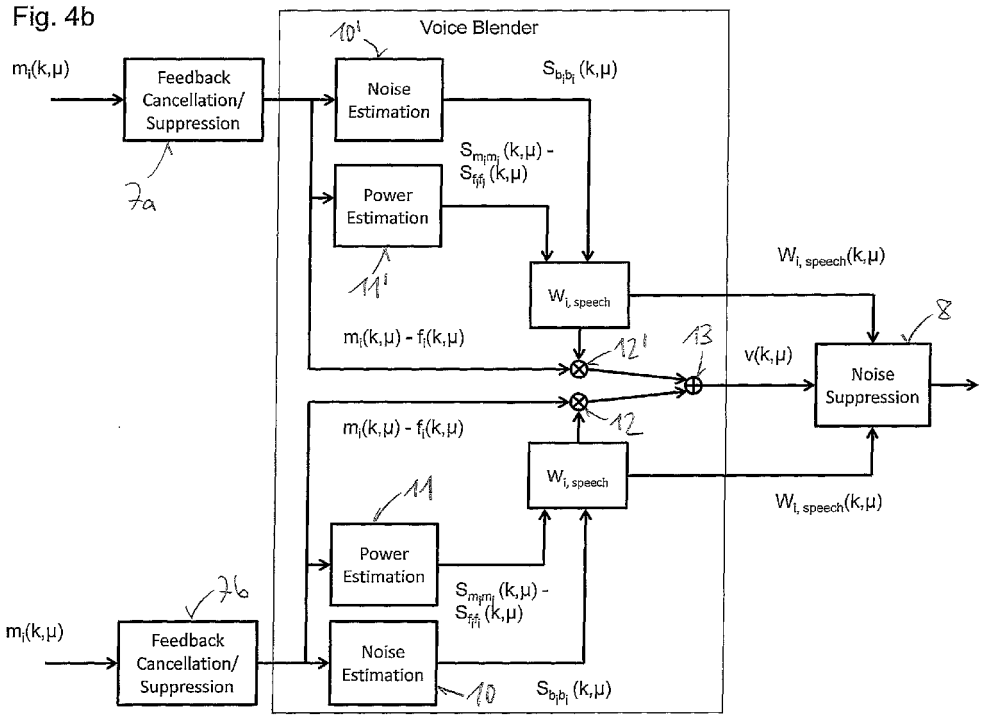

Voice Blender
10'
$m_i(k,\mu)$
Feedback Cancellation/ Suppression
7a
Noise Estimation
$S_{b_ib_i}(k,\mu)$
$S_{m_im_i}(k,\mu)$ - $S_{f_if_i}(k,\mu)$
Power Estimation
11'
$W_{i,speech}$
$W_{i,speech}(k,\mu)$
8
$m_i(k,\mu) - f_i(k,\mu)$
12'
13
$v(k,\mu)$
Noise Suppression
12
$m_i(k,\mu) - f_i(k,\mu)$
$W_{i,speech}$
$W_{i,speech}(k,\mu)$
11
Power Estimation
$S_{m_im_i}(k,\mu)$ - $S_{f_if_i}(k,\mu)$
7b
$m_i(k,\mu)$
Feedback Cancellation/ Suppression
Noise Estimation
10
$S_{b_ib_i}(k,\mu)$

Fig. 4c

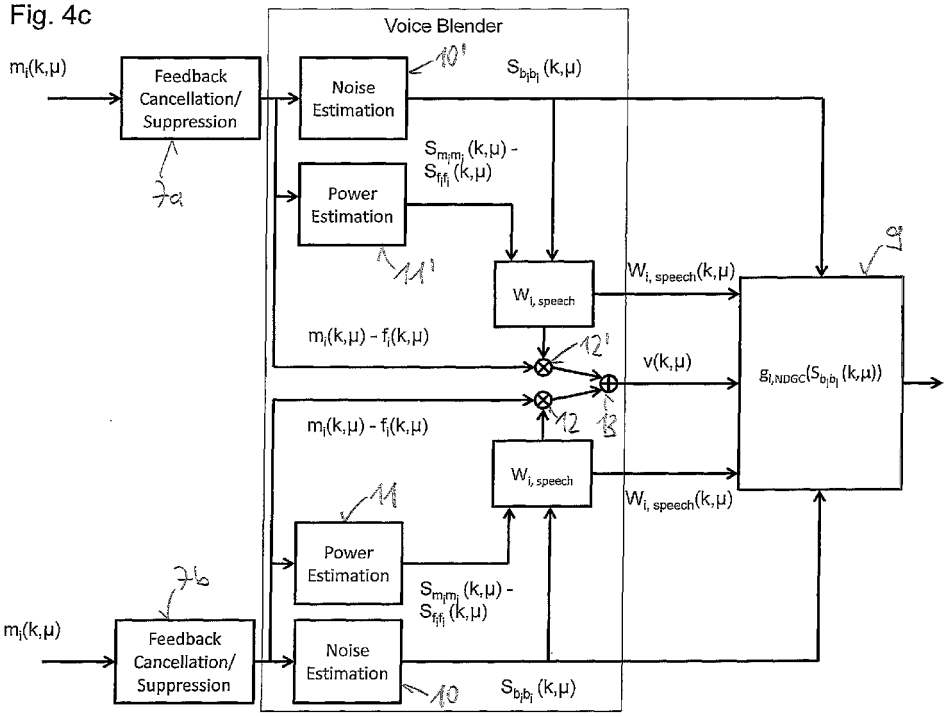
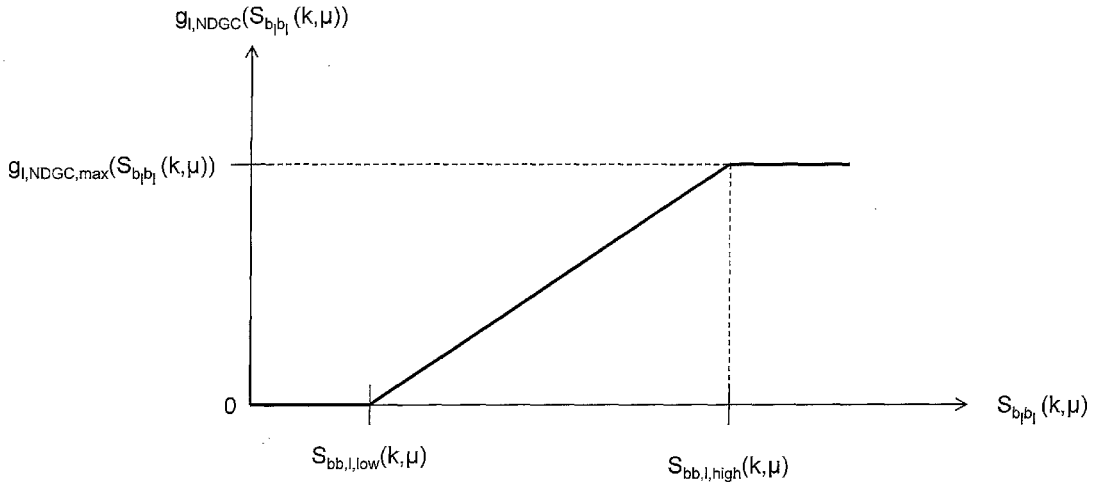

Voice Blender
$m_i(k,\mu)$
Feedback Cancellation/ Suppression
7a
Noise Estimation
10'
$S_{b_ib_i}(k,\mu)$
$S_{m_im_i}(k,\mu)$ - $S_{f_if_i}(k,\mu)$
Power Estimation
11'
$W_{i,\,speech}$
$W_{i,\,speech}(k,\mu)$
9
$m_i(k,\mu) - f_i(k,\mu)$
12'
$v(k,\mu)$
$g_{i,NDGC}(S_{b_ib_i}(k,\mu))$
12
13
$m_i(k,\mu) - f_i(k,\mu)$
$W_{i,\,speech}$
$W_{i,\,speech}(k,\mu)$
11
Power Estimation
$S_{m_im_i}(k,\mu)$ - $S_{f_if_i}(k,\mu)$
7b
$m_i(k,\mu)$
Feedback Cancellation/ Suppression
Noise Estimation
10
$S_{b_ib_i}(k,\mu)$

Fig. 6


$g_{l,NDGC}(S_{b_l b_l}(k,\mu))$
$g_{l,NDGC,max}(S_{b_l b_l}(k,\mu))$
0
$S_{b_l b_l}(k,\mu)$
$S_{bb,l,low}(k,\mu)$
$S_{bb,l,high}(k,\mu)$

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 11 15 5021

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X Y A	US 2009/316923 A1 (TASHEV IVAN JELEV [US] ET AL) 24 December 2009 (2009-12-24) * figure 5 * * paragraph [0001] * * paragraph [0025] - paragraph [0027] * * paragraph [0050] - paragraph [0054] * * paragraph [0057] - paragraph [0058] *	1,6,8,9 2,7 3-5	INV. H04R3/00 ADD. H04R3/02
X Y A	US 2003/026437 A1 (JANSE CORNELIS PIETER [NL] ET AL) 6 February 2003 (2003-02-06) * figure 1 * * paragraph [0019] * * paragraph [0029] - paragraph [0034] * * paragraph [0063] *	1,8,9 7 2-6	
Y	US 5 602 962 A (KELLERMANN WALTER [DE]) 11 February 1997 (1997-02-11) * figures 1,2 * * column 2, line 7 - line 12 * * column 3, line 3 - line 6 * * column 3, line 48 - column 4, line 55 *	2	TECHNICAL FIELDS SEARCHED (IPC)
A,D	US 6 549 629 B2 (FINN BRIAN M [US] ET AL) 15 April 2003 (2003-04-15) * column 9, line 41 - line 65 *	6	H04R
A,D	EP 1 850 640 A1 (HARMAN BECKER AUTOMOTIVE SYS [DE]) 31 October 2007 (2007-10-31) * paragraph [0007] *	6	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	15 June 2011	Guillaume, Mathieu

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
.....
& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 11 15 5021

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

15-06-2011

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009316923 A1	24-12-2009	NONE	
US 2003026437 A1	06-02-2003	EP 1413167 A2	28-04-2004
		WO 03010995 A2	06-02-2003
		JP 2004537232 T	09-12-2004
US 5602962 A	11-02-1997	DE 4330243 A1	09-03-1995
		EP 0642290 A2	08-03-1995
		JP 3373306 B2	04-02-2003
		JP 7240992 A	12-09-1995
US 6549629 B2	15-04-2003	WO 02069517 A1	06-09-2002
		US 2002141601 A1	03-10-2002
EP 1850640 A1	31-10-2007	AT 434353 T	15-07-2009
		CA 2581774 A1	25-10-2007
		CN 101064975 A	31-10-2007
		JP 2007290691 A	08-11-2007
		KR 20070105260 A	30-10-2007
		US 2007280486 A1	06-12-2007

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

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

Patent documents cited in the description

- EP 1850640 B1 **[0002]**
- US 006549629 B2 **[0002] [0037]**
- US 20050265560 A **[0006]**
- EP 1679874 B1 **[0008]**