

(12) **Österreichische Patentanmeldung**

(21) Anmeldenummer: A 207/2013  
(22) Anmeldetag: 15.03.2013  
(43) Veröffentlicht am: 15.12.2014

(51) Int. Cl.: **G10L 21/0216** (2013.01)

(56) Entgegenhaltungen:  
EP 2562752 A1  
Tachioka Y; Narita T; Iwasaki T: Direction of arrival estimation by cross-power spectrum phase analysis using prior distributions and voice activity detection information; Acoustical Science and Technology 2012; 06.09.2011  
Xiaoyu Ma; Kun Mu: Study on 2-dimensional microphone circular array speech enhancement algorithm based on the DOA; 2011 International Conference on Electronic & Mechanical Engineering and Information Technology (EMEIT 2011); 14.08.2011  
Bai M R; Pochang Ho: Using a steered-response power-phase transform to optimize speech pickup in reverberant environments; 2008 The Institution of Engineering and Technology; 04.2008  
Seon Man Kim; Hong Kook Kim; Sung Joo Lee; Yun Keun Lee: Adaptation mode control with residual noise estimation for beamformer-based multi-channel speech enhancement; 2012 IEEE International Conference on Acoustics, Speech and Signal Processing; 25.03.2012

(71) Patentanmelder:  
COMMEND INTERNATIONAL GMBH  
5020 Salzburg (AT)

(74) Vertreter:  
BAUMANN MARIUS MAG.  
6020 INNSBRUCK (AT)

(54) **Verfahren zur Erhöhung der Sprachverständlichkeit**

(57) Die Erfindung betrifft ein Verfahren zur Erhöhung der Sprachverständlichkeit eines Audiosignals (6), das aus einem Gemisch von Störgeräuschen (3) und Sprachsignalen (5) besteht, in digitalen elektroakustischen Kommunikationsanlagen (1). Die Sprachverständlichkeit wird durch Verminderung der Lautstärke der Störgeräusche (3) vom Ort eines Sprechers (4) erreicht. Zur Aufnahme des Audiosignals (6) werden zumindest zwei Mikrofonkanäle verwendet, welche das Audiosignal (6) an eine Auswerteeinheit (8) weiterleiten, in der das Störgeräusch (3) auf elektronischem Wege minimiert wird und das erzeugte Sprachsignal (5) von der Kommunikationsanlage (1) über eine Ausgabeeinheit (9) wiedergegeben wird. Der Betrieb der Kommunikationsanlage (1) und/oder der Auswerteeinheit (8) wird über eine Betriebssoftware, sichergestellt. Die Auswerteeinheit (8) umfasst

- ein erstes Modul (12), insbesondere ein DOA-Estimation-Modul (12), zur Bestimmung der Richtung des Sprachsignals (5),

- ein zweites Modul (13), insbesondere ein GSC-Beamforming-Modul (13), zur Feststellung und Trennung des Sprachsignals (5) von den Störgeräuschen (3) und
- ein drittes Modul (14), insbesondere ein Multichannel-Postfilter-Modul (14), zur Subtraktion des Störgeräusches (3) vom Sprachsignal (5).

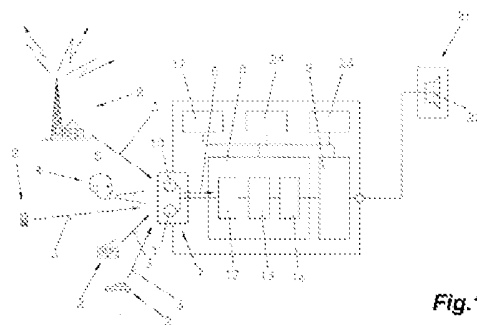
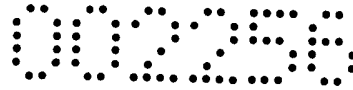


Fig. 1

**Zusammenfassung:**

Die Erfindung betrifft ein Verfahren zur Erhöhung der Sprachverständlichkeit eines  
5 Audiosignals (6), das aus einem Gemisch von Störgeräuschen (3) und  
Sprachsignalen (5) besteht, in digitalen elektroakustischen  
Kommunikationsanlagen (1). Die Sprachverständlichkeit wird durch Verminderung  
der Lautstärke der Störgeräusche (3) vom Ort eines Sprechers (4) erreicht. Zur  
Aufnahme des Audiosignals (6) werden zumindest zwei Mikrofonkanäle  
10 verwendet, welche das Audiosignal (6) an eine Auswerteeinheit (8) weiterleiten, in  
der das Störgeräusch (3) auf elektronischem Wege minimiert wird und das  
erzeugte Sprachsignal (5) von der Kommunikationsanlage (1) über eine  
Ausgabeeinheit (9) wiedergegeben wird. Der Betrieb der Kommunikationsanlage  
(1) und/oder der Auswerteeinheit (8) wird über eine Betriebssoftware,  
15 sichergestellt. Die Auswerteeinheit (8) umfasst

- ein erstes Modul (12), insbesondere ein DOA-Estimation-Modul (12), zur  
Bestimmung der Richtung des Sprachsignals (5),
- ein zweites Modul (13), insbesondere ein GSC-Beamforming-Modul (13),  
zur Feststellung und Trennung des Sprachsignals (5) von den  
20 Störgeräuschen (3) und
- ein drittes Modul (14), insbesondere ein Multichannel-Postfilter-Modul (14),  
zur Subtraktion des Störgeräusches (3) vom Sprachsignal (5).

Fig. 1

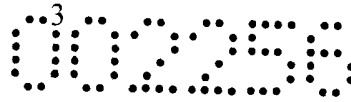
Die Erfindung beschreibt ein Verfahren zur Erhöhung der Sprachverständlichkeit eines Audiosignals, eines Schrankensystems mit einer Kommunikationsanlage,  
 5 einer Notrufanlage und einer Kommunikationseinheit, gemäß dem Oberbegriff der Ansprüche 1, 34, 36 und 39.

Es sind bereits Systeme bekannt, bei denen aus einem Gemisch von Störgeräuschen und Sprachsignalen zur besseren Verständlichkeit versucht wird,  
 10 die Störgeräusch auszufiltern. Hierzu erfolgt die Aufnahme über ein einziges Mikrofon, von dem aus der aufgenommene Schall, bzw. das Audiosignal an eine Auswerteeinheit übertragen wird. In der Auswerteeinheit wird anschließend bevorzugt über eine Software versucht die Störgeräusche mit den verschiedensten bekannten Algorithmen auszufiltern. Nachteilig ist hierbei, dass  
 15 durch die Verwendung eines einzelnen Mikrofones stets ein Kompromiss zwischen Störgeräusch-Reduktion und erzielbarer Sprachqualität zu treffen ist.

Weiters ist aus dem Stand der Technik bekannt, dass derartige Systeme dazu neigen, selbst störende Nebengeräusche zu erzeugen, welche unter dem Begriff  
 20 „musical artefacts“ bekannt sind.

Die Aufgabe der Erfindung liegt darin, ein Verfahren zur Erhöhung der Sprachverständlichkeit eines Audiosignals, eines Schrankensystems mit einer Kommunikationsanlage, einer Notrufanlage und einer Kommunikationseinheit zu  
 25 schaffen, bei dem die Wiedergabe-Qualität des Sprachsignals verbessert wird. Eine weitere Aufgabe der Erfindung liegt darin, die Aufnahmequalität des Sprachsignals zu verbessern. Darüber hinaus liegt die Aufgabe der Erfindung darin, die bekannten Nachteile aus dem Stand der Technik zu vermeiden und zu verbessern.

30 Eine weitere Aufgabe der Erfindung liegt darin, durch optimale Auswertung des Sprachsignals die bestmögliche Sprachqualität bei gleichzeitiger Vorgabe der



Störgeräusch-Unterdrückung zu erzielen. Diese Vorgabe erfolgt als numerischer Wert in dB (Dezibel).

Die Aufgabe der Erfindung wird dadurch gelöst, dass die Sprachverständlichkeit durch Verminderung der Lautstärke der Störgeräusche vom Ort des Sprechers erreicht wird, wobei zur Aufnahme des Audiosignals zumindest zwei Mikrofonkanäle verwendet werden, welche das Audiosignal an eine Auswerteeinheit weiterleiten, in der das Störgeräusch auf elektronischem Wege minimiert wird und das erzeugte Sprachsignal von der Kommunikationsanlage über eine Ausgabeeinheit wiedergegeben wird und dass der Betrieb der Kommunikationsanlage und/oder der Auswerteeinheit über eine Betriebssoftware, sichergestellt wird und dass die Auswerteeinheit

- ein erstes Modul, insbesondere ein DOA-Estimation-Modul, zur Bestimmung der Richtung des Sprachsignals,
- ein zweites Modul, insbesondere ein GSC-Beamforming-Modul, zur Feststellung und Trennung des Sprachsignals von den Störgeräuschen und
- ein drittes Modul, insbesondere ein Multichannel-Postfilter-Modul, zur Subtraktion des Störgeräusches vom Sprachsignal

umfasst.

Vorteilhaft ist dabei, dass durch die modularisierte Auswertung in einfache Art und Weise eine Erweiterung der Auswertemöglichkeiten geschaffen wird, sodass eine vielfältige Einsatzmöglichkeit geschaffen wird. Beispielsweise ist es dabei möglich, dass eine einseitige Spracherkennung durch Ergänzung eines weiteren Moduls auf eine beidseitige Sprachkommunikation ausgebaut werden kann. Ein wesentlicher Vorteil bei der erfindungsgemäßen Lösung ist auch, dass bereits aus dem Stand der Technik bekannte Analysen mit neuen Analysen und Technologien auf einfache Art und Weise ergänzt oder ausgetauscht werden können. Von Vorteil ist auch, dass je nach Anwendungsgebiet die unterschiedlichen Module optimal angepasst werden können, so dass die bestmögliche Spracherkennung und Sprachwiedergabe geschaffen werden kann.

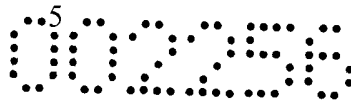
Darüber hinaus ist ein wesentlicher Vorteil gegenüber dem Stand der Technik, dass zuerst eine Richtungserkennung des Sprachsignals durchgeführt wird, sodass dadurch eine optimierte Aufnahme des Sprachsignals ungedämpft vorgenommen werden kann, wodurch die restlichen Bereich außerhalb der  
5 Aufnahmerichtung besser ausgefiltert, bzw. nur gedämpft aufgenommen werden. Vorteilhaft ist bei der erfindungsgemäßen Lösung aber auch, dass die Richtungserkennung auch bei hohen Störgeräuschpegeln robust genug arbeitet, um die Richtung und die Aktivität des Sprechers zuverlässig zu erkennen. Diese Information ist in vielen Kommunikationssystemen (Schrannen, etc.) nützlich um  
10 eine sprachgesteuerte Verbindung aufzubauen.

Es sind aber auch die Maßnahmen von Vorteil, bei dem im ersten Modul, dem DOA-Estimation-Modul, eine Richtwirkung der Mikrofonempfindlichkeit in Richtung des Sprechers erzeugt wird, wobei die Hauptkeule mit der Bewegung des  
15 Sprechers nachgeführt wird, damit wird der Aufnahmepegel für das Sprachsignal im Verhältnis zum Störgeräusch verbessert.

Bei der Maßnahme, bei der die Hauptkeule eine Breite von  $-20^\circ$  und  $+20^\circ$  um die aktuelle Position des Sprechers umfasst, wobei der nutzbare Frequenzbereich 20 bis 20 000 Hz umfasst, wird in vorteilhafter Weise erreicht, dass eine optimierte Aufnahme erfolgen kann und eine einfachere Ausblendung der umliegenden Geräusche möglich ist. Durch die Begrenzung der Hauptkeule und ev. einer Nachführung der Hauptkeule kann eine konzentriertere Auswertung erfolgen, sodass eine wesentlich höhere Sprachqualität erzielt wird, da eventuelle  
25 Überschneidungen mit den Störgeräuschen reduziert werden.

Eine vorteilhafte Maßnahme liegt darin, dass die Breite der Hauptkeule mit steigender Frequenz in der Bandbreite abnimmt, da dadurch wiederum eine konzentriertere Auswertung geschaffen wird.

30 Vorteilhaft ist auch eine Maßnahme, bei der die Hauptkeule derart ausgelegt wird, dass minimale Schwankungen, wie beispielsweise Kopfbewegungen des Sprechers, automatisch ausgeglichen werden, da dadurch immer eine optimale



Sprachaufnahme für die Auswertung ermöglicht wird. Auch kann damit eine Vereinfachung der Richtungsbestimmung des Sprachsignals, insbesondere des Standort des Sprechers, erzielt werden, da diese Daten für die Richtungsbestimmung herangezogen werden können.

5

Bei den Maßnahmen, bei dem im ersten Modul, insbesondere in dem DOA-Estimation Modul, folgender Algorithmus „Direct-to-Diffusion Ratio“ zur Schätzung bzw. Festlegung der Richtung des Sprachsignals läuft bzw. abgearbeitet wird, ist von Vorteil, dass durch die Verwendung eines speziellen Algorithmus eine qualitativ hohe Auswertung durchgeführt werden kann.

10

Von Vorteil ist bei dem Vorgehen, bei dem sich das Audiosignal, insbesondere ein Mikrofonsignal, am Mikrofon durch das Störgeräusch und das Sprachsignal derart zusammen setzt und derart definiert wird

$$Z_m(j\Omega) = S(j\Omega)e^{-j\Omega f_s \Delta\tau_m} + N(j\Omega)$$

15

wobei  $Z_m(j)$  das Audiosignal, insbesondere des m-ten Mikrofonsignal,  $N(j)$  das Störsignal,  $S(j)$  das Sprachsignal und  $\Delta\tau_m$  die Richtung definieren,

da dadurch eine automatische Zuordnung der aufgenommen Signale geschaffen wird, sodass eine automatische Auswertung möglich ist.

20

Es ist aber auch eine Maßnahme von Vorteil, bei der im ersten Modul eine Spektraldichteschätzung über eine Zeitmittelung entsprechend dem folgenden Algorithmus

25

$$\mathbb{E}_{Z_m Z_m}(j\Omega) = \mathbb{E}\{|Z_m(j\Omega)|^2\} = \Phi_{SS}(j\Omega) + \Phi_{NN}(j\Omega).$$

berechnet wird.

Bei einem Vorgehen, bei dem aus der Spektraldichteschätzung eine komplexe räumlich Kohärenz zwischen dem ersten Audiosignal und dem weiteren Audiosignal, insbesondere Mikrofonsignalen, über den Algorithmus

30

$$\begin{aligned}\gamma_{Z_m Z_{m'}}(j\Omega) &= \frac{\Phi_{Z_m Z_{m'}}(j\Omega)}{\sqrt{\Phi_{Z_m Z_m}(j\Omega) \Phi_{Z_{m'} Z_{m'}}(j\Omega)}} \\ &= \frac{\Gamma_{DDR}(j\Omega) e^{-j\Omega f_s \Delta\tau_m} + \gamma_{N_m N_{m'}}(j\Omega)}{\Gamma_{DDR}(j\Omega) + 1}\end{aligned}$$

berechnet wird, wobei  $\Gamma_{DDR}(j\Omega)$  das Verhältnis aus den Energien des Direktschalles zum diffusen Schall beschreibt, wird in vorteilhafter Weise erreicht, dass eine optimale Einschätzung der Richtung des nützlichen Sprachsignals ermöglicht wird.

Von Vorteil ist dabei das Vorgehen, das Sprachsignal dem Direktschall und das Störsignal dem diffusen Schall zuzuordnen und ein derartiges Verhältnis wie folgt zu bilden:

$$\Gamma_{DDR}(j\Omega) = \frac{\Phi_{SS}(j\Omega)}{\Phi_{NN}(j\Omega)} = \frac{\gamma_{N_m N_{m'}}(j\Omega) - \gamma_{Z_m Z_{m'}}(j\Omega)}{\gamma_{Z_m Z_{m'}}(j\Omega) - e^{-j\Omega f_s \Delta\tau_{mm'}}}.$$

Vorteilhaft ist die Maßnahme, bei dem das Störgeräusch als ideal isotropisch betrachtet wird und derart definiert wird:

$$\gamma_{N_m N_{m'}}(j\Omega) = \frac{\sin(kd_{mm'})}{kd_{mm'}}$$

Durch die Maßnahme, bei der durch Integration über das komplexe Fourierspektrum entsprechend der Formel

$$J(\Delta\tau_{mm'}) = \int_{\Omega=0}^{2\pi} \frac{\gamma_{N_m N_{m'}}(j\Omega) - \gamma_{Z_m Z_{m'}}(j\Omega)}{\gamma_{Z_m Z_{m'}}(j\Omega) - e^{-j\Omega f_s \Delta\tau_m}} d\Omega$$

eine Kostenfunktion gebildet wird, wird erreicht, dass eine optimale Basis für die Richtungsbestimmung gebildet wird.

Von Vorteil sind die Maßnahmen, bei dem durch Vorgabe von mehreren, insbesondere fünfzig, möglichen Einfallswinkeln zwischen  $-90^\circ$  und  $+90^\circ$  man den wahrscheinlichen Einfallswinkel für das Maximum der Kostenfunktion gemäß der Formel

$$\Delta\hat{\tau}_m = \arg \max_{\Delta\tau_m} J(\Delta\tau_m)$$

erhält, da dadurch in einfacher Weise die Richtungsbestimmung vorgenommen werden kann.

- Bei einem Vorgehen, bei dem im Modul „RTF-Estimation“ der Betrag der Übertragungsfunktion zwischen dem m-ten und dem m'-ten Mikrofon gemäß der Formel

$$|\hat{A}_m(j\Omega)| \approx \sqrt{\frac{\Phi_{Z_m Z_m}(j\Omega)}{\Phi_{Z_{m'} Z_{m'}}(j\Omega)}}$$

- berechnet wird, wird erreicht, dass Exemplantoleranzen der Mikrofone ausgeglichen werden. Dadurch kann der teure und aufwändige Abgleich der verwendeten Mikrofone entfallen.

Die Maßnahmen, bei dem mit dem Vorhandensein eines Sprachsignals und einem ermittelten Einfallswinkel eine Transferfunktion gemäß der Formel

$$\hat{A}_m(j\Omega) = |\hat{A}_m(j\Omega)| e^{-j\Omega f_s \Delta \hat{\tau}_m}$$

- ermittelt wird, sind von Vorteil, da dadurch eine Weiterverarbeitung der Daten und/oder Signale vereinfacht wird.

- Bei einem Vorgehen, bei dem im zweiten Modul die Aufteilung der über die Mikrofonkanäle aufgenommenen Audiosignale in ein Sprachsignal und in ein Störgeräusch erfolgt, ist von Vorteil, dass dazu entsprechende Algorithmen verwendet werden können. Diese entsprechen dem Stand der Technik.

- Von Vorteil ist ein Vorgehen, bei dem die ermittelte Richtung des Sprachsignals und/oder die Daten hierfür vom ersten Modul an das zweite Modul übergeben werden, da dadurch eine qualitativ hochwertige Auswertung des Sprachsignals möglich ist.

- Durch die Maßnahmen, bei dem bei der Aufteilung des Audiosignales in das Sprachsignal und dem Störgeräusch die Richtung des Sprachsignals und/oder die Daten des ersten Moduls berücksichtigt werden, wird in vorteilhafter Weise erreicht, dass dadurch die in Richtung des Nutzers erzeugten Daten bzw. Signale

bevorzugt werden können und somit eine wesentlich bessere Auswertung durchgeführt werden kann.

Vorteilhaft ist auch eine Maßnahme, bei der im dritten Modul, im so genannten  
 5 Multichannel-Postfilter-Modul, ein mathematisches Verfahren basierend auf einer  
 räumlichen Autokorrelationsmatrix der Signale zum Einsatz kommt, welches das  
 Verhältnis von Sprachsignal-Energie zu Störsignal-Energie (SNR) im  
 Ausgabesignal des zweiten Modules ermittelt.

Auch ist ein Vorgehen von Vorteil, bei dem im dritten Modul unter Verwendung  
 10 einer Filterfunktion “ $\mathbf{W}(j\Omega)$ ” eines GSC-Beamformers die Schätzung des  
 Sprachsignals und des Störgeräusches wie folgt angegeben wird:

$$\begin{aligned}\hat{\mathbf{S}}(j\Omega) &= \hat{\mathbf{A}}(j\Omega) \mathbf{W}^H(j\Omega) \mathbf{Z}(j\Omega) \\ \hat{\mathbf{N}}(j\Omega) &= \mathbf{Z}(j\Omega) - \hat{\mathbf{S}}(j\Omega).\end{aligned}$$

dabei beschreibt der Vektor  $\mathbf{A}(\Omega_j)$  die Transferfunktionen vom Sprecher zu allen  
 Mikrofonen.

15

Von Vorteil ist ein Vorgehen, bei dem sich aus der Formel

$$\begin{aligned}\Phi_{\hat{\mathbf{S}}\hat{\mathbf{S}}} &= \sum_{m=1}^M |\hat{S}_m(j\Omega)|^2 \\ \Phi_{\hat{\mathbf{N}}\hat{\mathbf{N}}} &= \sum_{m=1}^M |\hat{N}_m(j\Omega)|^2.\end{aligned}$$

die Langzeit-Spektraldichteschätzung des Sprachsignals und respektive des  
 Störgeräusches ergibt, da dadurch eine bessere Trennung des Sprachsignals vom  
 20 Störgeräuschen ermöglicht wird.

Ein Vorteil ergibt sich durch die Maßnahmen, bei der eine effiziente  
 Implementation für die Langzeit-Spektraldichteschätzung gemäß der Formel

$$\begin{aligned}Tr(\mathbf{\Gamma}_{NN}) &= M \\ Tr(\mathbf{\Gamma}_{NN} \mathbf{W} \hat{\mathbf{A}}^H) &= Tr(\hat{\mathbf{A}} \mathbf{W}^H \mathbf{\Gamma}_{NN}) = \sum_{m=1}^M A_m (\mathbf{W}^H \mathbf{\Gamma}_{NN})_m \\ Tr(\hat{\mathbf{A}} \mathbf{W}^H \mathbf{\Gamma}_{NN} \mathbf{W} \hat{\mathbf{A}}^H) &= Tr(\hat{\mathbf{A}}^H \hat{\mathbf{A}} \mathbf{W}^H \mathbf{\Gamma}_{NN} \mathbf{W}) = \sum_{m=1}^M |A_m|^2 \mathbf{W}^H \mathbf{\Gamma}_{NN} \mathbf{W}\end{aligned}$$

erfolgt, da dadurch eine optimale Ausfilterung der Störgeräusche möglich ist.

Auch eine Verbesserung der Trennung der Signale, wird durch die Maßnahmen erreicht, bei der die Kurzzeit-Spektralschätzung durch direkte Berechnung gemäß der Formel

$$\begin{aligned}\tilde{\Phi}_{\hat{S}\hat{S}} &= \hat{A}W^H\Gamma_{NN}W\hat{A}^H \\ \tilde{\Phi}_{\hat{N}\hat{N}} &= \Gamma_{NN} - \Gamma_{NN}W\hat{A}^H - \hat{A}W^H\Gamma_{NN} + \hat{A}W^H\Gamma_{NN}W\hat{A}^H\end{aligned}$$

erhalten wird.

Durch die Maßnahmen, bei der das Verhältnis des Sprachsignals zum Störgeräusch gemäß dem Algorithmus

$$\xi(j\Omega) = \frac{\text{Tr}(\tilde{\Phi}_{\hat{S}\hat{S}})}{\text{Tr}(\tilde{\Phi}_{\hat{N}\hat{N}})} \frac{\Phi_{\hat{N}\hat{N}}}{\Phi_{\hat{S}\hat{S}}} - 1$$

berechnet wird, ist eine Ausfilterung der Störgeräusche effizient möglich.

Bei einem Vorgehen, bei dem im dritten Modul die vom zweiten Modul übergebenen Signale annähernd in Echtzeit, vorzugsweise mit 32 ms Verzögerung, verarbeitet werden, wird in vorteilhafter Weise erreicht, dass für den Nutzer so gut wie keine Zeitverzögerungen wahrnehmbar sind.

Das Vorgehen, bei dem vom dritten Modul eine Absenkung des Störgeräusches um 25 dB erreicht wird, hat zum Vorteil, dass dadurch bereits die Störgeräusche nur mehr gedämpft aufgenommen werden. Somit ist die Qualität des aufgenommen Sprachsignals wesentlich besser.

Von Vorteil ist, dass eine optimale Parametrierung der Auswerteeinheit über umfangreiche automatisierte Sprachtests, welche auf deutschen und englischen Sprachkorpora basieren, erfolgt, da dadurch eine optimale Anpassung der Anlage vor Ort unter Berücksichtigung der baulichen Gegebenheiten vorgenommen werden kann. Insbesondere kann dabei auf Echos an der Sprachstelle Rücksicht genommen werden.

Um eine umfangreiche Einsatzmöglichkeit zu schaffen, ist es von Vorteil, wenn eine zusätzliche weitere Software zur Steuerung angeschlossener Komponenten in der Kommunikationsanlage abgearbeitet wird und/oder Daten von der Kommunikationsanlage an angeschlossene Komponenten übergeben werden.

5

Durch die Maßnahmen, bei denen bei einem Einsatz in einer Gegensprechanlage ein weiteres Modul, insbesondere ein sogenanntes Echocanceller-Modul, zur Reduzierung von akustischen Echos, zusammen mit den Modulen eins bis drei in einer einzigen Betriebssoftware eingesetzt wird, wird in vorteilhafter Weise erreicht, dass dadurch Echos und Rückkopplungen für die Spracheingabe und Sprachausgabe verhindert werden können.

10

Auch ist ein Vorgehen von Vorteil, bei dem ein weiteres zusätzliches Modul, insbesondere ein sogenanntes RTF-Estimation-Modul, welches aus der Richtungsinformation des ersten Moduls die relativen akustischen Transferfunktionen zwischen den Mikrofonen berechnet, eingesetzt wird, da dadurch eine Verbesserung der Qualität erzielt werden kann.

15

Für die bestmögliche Sprachaufnahme ist es von Vorteil, wenn die Anordnung der Mikrofone während des Betriebes ermittelt wird, wobei lediglich ein Bereich für den Abstand der Mikrofone vorgegeben wird.

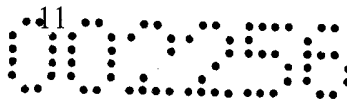
20

Auch ist ein Vorgehen möglich, bei dem das resultierende Sprachsignal entweder direkt ausgegeben wird und/oder, vorzugsweise permanent, in einem Sprachspeicher abgelegt wird, wobei der Signaltransport über ein beliebiges Medium und/oder Protokoll erfolgt.

25

Vorteilhaft ist auch ein Vorgehen, bei dem von der Kommunikationsanlage aufgrund von hinterlegten Sprachbefehlen angeschlossene Komponenten aktiviert und deaktiviert werden, da dadurch keine Aktivierungselemente benötigt werden, sondern alles per Sprache ansteuerbar ist. Somit sind theoretisch beliebig viele Steuermöglichkeiten vorhanden.

30



Weiters werden die Aufgaben der Erfindung auch durch ein Schrankensystem mit einer Kommunikationsanlage gelöst, bei dem die Kommunikationsanlage und/oder die Zentraleinheit zur Durchführung des Verfahrens nach einem oder mehreren der Ansprüche 1 bis 18 ausgebildet ist. Vorteilhaft ist dabei, dass dadurch ein Einsatzgebiet geschaffen wird, bei dem mit sehr vielen umliegenden Störgeräuschen, insbesondere von vorbeifahrenden Fahrzeugen, zu rechnen ist.

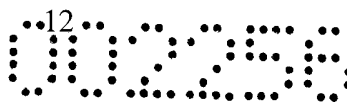
Dabei ist ein Vorgehen von Vorteil, bei dem die Wiedergabe des Sprachsignals über die Ausgabeeinheit in der Meldezentrale erfolgt, dass dadurch bei Problemen an der Schrankenanlage oder bei Fragen eines Kunden bzw. Nutzers diesem auch über die erfindungsgemäße Kommunikationsanlage eine Sprachausgabe und somit eine Rückmeldung ermöglicht wird.

Auch werden die Aufgaben der Erfindung durch eine Notrufanlage gelöst, bei der die Zentraleinheit und/oder Kommunikationsanlage zur Durchführung des Verfahrens nach einem oder mehreren der Ansprüche 1 bis 18 ausgebildet ist. Von Vorteil ist hierbei, dass dadurch der Einsatz mehrerer Kommunikationsanlagen in einem Gesamtsystem möglich ist.

Von Vorteil ist eine Anordnung, bei der die Übertragungseinheit durch Datenleitungen, Telefonleitungen, Funksignale und/oder Internetverbindungen gebildet ist, da dadurch die optimale Verbindungsart für die Datenübertragung ausgewählt werden kann. Somit wird eine sichere und bestmögliche Übertragung erreicht.

Auch ist eine Anordnung von Vorteil, bei der die Ausgabeeinheit für das Sprachsignal mit der Übertragungseinheit für die externe Meldestelle verbunden ist, da dadurch eine direkte Datenübertragung, insbesondere des Sprachsignals, möglich ist.

Schließlich werden die Aufgaben der Erfindung auch durch eine Kommunikationseinheit gelöst, bei der die Kommunikationsanlage in Verkehrstunneln, Industrieparks, Hotelanlagen, usw. einsetzbar ist.



Die Erfindung wird nachstehend mit mehreren Ausführungsbeispielen beschrieben.

Es zeigen:

- 5 Fig.1 eine schematische Darstellung einer Kommunikationsanlage;
  - Fig.2 eine symbolhafte Darstellung eines Richtdiagramms für eine Schallquelle zu Mikrofonen der Kommunikationsanlage;
  - Fig.3 eine schaubildliche Darstellung für einen Einfallswinkel eines Sprachsignals eines Nutzers;
  - 10 Fig.4 eine Draufsicht auf eine Schrankenanlage, in vereinfachter, schematischer Darstellung;
  - Fig.5 eine weitere Draufsicht eines Ausführungsbeispiels für ein Notrufsystem, in vereinfachter, schematischer Darstellung.
- 15 Einführend sei festgehalten, dass in den beschriebenen Ausführungsformen gleiche Teile mit gleichen Bezugszeichen bzw. gleichen Bauteilbezeichnungen versehen werden, wobei die in der gesamten Beschreibung enthaltenen Offenbarungen sinngemäß auf gleiche Teile mit gleichen Bezugszeichen bzw. gleichen Bauteilbezeichnungen übertragen werden können. Auch sind die in der
- 20 Beschreibung gewählten Lageangaben, wie z.B. oben, unten, seitlich usw. auf die unmittelbar beschriebene sowie dargestellte Figur bezogen und sind bei einer Lageänderung sinngemäß auf die neue Lage zu übertragen. Weiters können auch Einzelmerkmale oder Merkmalskombinationen aus den gezeigten und beschriebenen unterschiedlichen Ausführungsbeispielen für sich eigenständige,
- 25 erfinderische oder erfindungsgemäße Lösungen darstellen.

In den Figuren 1 bis 3 ist eine bevorzugt digitale elektroakustische Kommunikationsanlage 1 mit unterschiedlich symbolhaft dargestellten Schallquellen 2 zum Erzeugen eines Störgeräusche 3, wie beispielsweise

30 Industrielärm, Autolärm, Musiklärm, Handylärm, usw., gezeigt. Weiters ist schaubildlich ein Nutzer 4 zum Erzeugen eines Sprachsignals 5 dargestellt. Die beiden Schallquellen 2, insbesondere die Störgeräusche 3 und das Sprachsignal

5, bilden zusammen ein Audiosignal 6, welches von einem Aufnahmemittel 7 aufgenommen wird.

Die Kommunikationsanlage 1 ist zur Erhöhen der Sprachverständlichkeit des Audiosignals 6, insbesondere des Sprachsignals 5, ausgebildet und umfasst bei dem dargestellten Ausführungsbeispiel beispielsweise ein Auswerteeinheit 8 und eine Ausgabeeinheit 9, wobei die einzelnen Komponenten bevorzugt modular aufgebaut sind. Dabei wird die Sprachverständlichkeit durch Verminderung der Lautstärke der Störgeräusche 3 vom Ort eines Sprechers bzw. Nutzers 4 erreicht. Hierzu ist mit der Auswerteeinheit 8 das Aufnahmemittel 7 verbunden, wobei zur Aufnahme des Audiosignals 6 das Aufnahmemittel 7 aus zumindest zwei Mikrofonen 10, 11 gebildet ist. Grundsätzlich ist es möglich, eine beliebige Anzahl an Mikrofonen 10, 11 für die Aufnahme des Audiosignals 6 zu verwenden, wobei es sich jedoch gezeigt hat, dass bis zu vier Mikrofone 10, 11 sich als geeignet herausgestellt haben. Dabei ist nämlich zu beachten, je mehr Mikrofone 10, 11 angeschlossen werden, desto höher ist der Aufwand für die Auswertung des Sprachsignals 5, wodurch sich längere Laufzeiten bzw. Verzögerungen ergeben können. D.h., dass das von den hier gezeigten zwei Mikrofonen 10, 11 aufgenommene Audiosignal 6, welches sich aus dem Sprachsignal 5 und Störgeräuschen 3 zusammen setzt, an die Kommunikationsanlage 1, insbesondere der Auswerteeinheit 8, weitergeleitet wird, in der das Störgeräusch 3 auf elektronischem Wege minimiert wird und das erzeugte Sprachsignal 5 von der Kommunikationsanlage über die Ausgabeeinheit 9 wiedergegeben wird.

Dabei ist die Kommunikationsanlage 1, insbesondere die Auswerteeinheit 8 und die Ausgabeeinheit 9, derart ausgebildet, dass der Betrieb über eine Betriebssoftware sichergestellt ist, d.h., dass in den einzelnen Modulen eine bzw. mehrere Programme laufen und darin entsprechende Algorithmen, Verfahrensabläufe, usw. abgearbeitet werden. Die Auswerteeinheit 8, wird dabei aus drei Modulen 12, 13 und 14 mit unterschiedlichen Funktionen gebildet, wobei das erste Modul 12, insbesondere ein DOA-Estimation-Modul 12, zur Bestimmung der Richtung des Sprachsignals 5, das zweite Modul 13, insbesondere ein GSC-Beamforming-Modul 13, zur Feststellung und Trennung des Sprachsignals 5 von

den Störgeräuschen 3 und das dritte Modul 14, insbesondere ein Multichannel-Postfilter-Modul 14, zur Subtraktion des Störgeräusches 3 vom Sprachsignal 5 umfasst.

- 5 Bei der erfindungsgemäßen Lösung wird ein sogenanntes „Dual Channal Speech Enhancement“, abgekürzt DCSE-System, durchgeführt, wobei das DCSE-System ein technisch-mathematisches Verfahren beschreibt, welches von der Kommunikationsanlage 1, insbesondere der einzelnen Module, umgesetzt wird. Der Vorteil des DCSE liegt darin, dass das Sprachsignal 5 klar verständlich und  
10 mit hoher Qualität an der Gegenstelle wiedergegeben wird, wobei dazu eine optimale Minimierung bzw. Ausfilterung der Störgeräusche, wie beispielsweise Echos und Umgebungslärm, durchgeführt wird.

- Wie bereits zuvor kurz erwähnt, wird das Sprachsignal 5 des Nutzers 4 über die  
15 Mikrofone 10, 11 aufgenommen, wobei die Mikrofon 10, 11 auch die Nebengeräusche um den Nutzer 4 in Form von den Störgeräuschen 3 aufnimmt. Dieses aufgenommen Signal wird als Audiosignal 6 an die Auswerteeinheit 8 übergeben, von der nunmehr die Aufbereitung des Audiosignals 8 erfolgt.

- 20 Vom ersten Modul 12 wird zuerst die Richtung des Sprachsignals 5 ermittelt, wozu eine Richtwirkung der Mikrofonempfindlichkeit in Richtung des Sprechers bzw. Nutzers 4 erzeugt wird, wobei eine Hauptkeule 15, wie aus Fig. 2 ersichtlich, mit der Bewegung eines Sprechers 4 nachgeführt wird. Dabei ist in Fig. 2 ein Diagramm für die Mikrofonempfindlichkeit für zwei Mikrofone 10, 11 dargestellt,  
25 wobei hierbei die Einfallsrichtung des Schalles gegen die Frequenz aufgetragen ist. Der schraffierte Bereich in der Mitte wird dabei unbedämpft übertragen und bildet die Hauptkeule 15, wogegen die restlichen Bereiche mit -20dB stark unterdrückt werden. Der Bereich der größten Empfindlichkeit (Hauptkeule 15) ist zwischen  $-20^\circ$  und  $+20^\circ$ , entsprechend dem schraffierten Bereich zu finden. Diese  
30 Hauptkeule 15 wird mit den Bewegungen des Sprechers nachgeführt, und ist auch ausreichend breit, um minimale Schwankungen wie etwa Kopfbewegungen auszugleichen. D.h., dass der Nutzer 4 oftmals beim Sprechen den Kopf leicht bewegt, wobei dies erkannt wird und durch Nachführen der Hauptkeule 15



wiederum die bestmögliche Aufnahme erzielt wird. Die weiteren Bereiche oberhalb von 6000Hz bei  $+90^\circ$  sind die sog. Nebenkeulen 16 (Sidelobes). Aufgrund der hohen Frequenz stellen diese jedoch keine Qualitätsbeeinträchtigung dar.

- 5 Die Hauptkeule 15 umfasst also eine Breite von  $-20^\circ$  und  $+20^\circ$  um die aktuelle Position des Sprechers 4 umfasst, wobei der nutzbare Frequenzbereich 20 bis 20 000 Hz umfasst. Gleichzeitig nimmt die Breite der Hauptkeule 15 mit steigender Frequenz in der Bandbreite ab, wobei die Hauptkeule 15 derart ausgelegt wird, dass minimale Schwankungen, wie beispielsweise Kopfbewegungen des
- 10 Sprechers, automatisch ausgeglichen werden.

- Um eine Richtungsbestimmung durchführen zu können, wird im ersten Modul 12, insbesondere in dem DOA-Estimation Modul 12, folgender Algorithmus „Direct-to-Diffusion Ratio“ zur Schätzung bzw. Festlegung der Richtung des Sprachsignals
- 15 5 abgearbeitet. Dabei wird das Audiosignal 6, insbesondere ein Mikrofonsignal, am Mikrofon 10, 11 durch das Störgeräusch 3 und das Sprachsignal 5 derart zusammen gesetzt und derart definiert:

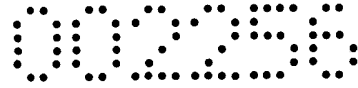
$$Z_m(j\Omega) = S(j\Omega)e^{-j\Omega f_s \Delta\tau_m} + N(j\Omega). \quad (1.1)$$

- 20  $Z_m(j)$  definiert dabei das Audiosignal 6, insbesondere des m-ten Mikrofonsignal,  $N(j)$  das Störsignal 3,  $S(j)$  das Sprachsignal 5 und  $\Delta\tau_m$  die Richtung definieren. Daraus wird nunmehr eine Spektraldichteschätzung über eine Zeitmittelung entsprechend dem folgenden Algorithmus

25 
$$\Phi_{Z_m Z_m}(j\Omega) = \mathbb{E}\{|Z_m(j\Omega)|^2\} = \Phi_{SS}(j\Omega) + \Phi_{NN}(j\Omega). \quad (1.2)$$

berechnet. Aus der Spektraldichteschätzung wird die komplexe räumliche Kohärenz zwischen dem m-ten und dem m'-ten Mikrofon 10, 11 über folgenden Zusammenhang ermittelt:

30



$$\begin{aligned}\gamma_{Z_m Z_{m'}}(j\Omega) &= \frac{\Phi_{Z_m Z_{m'}}(j\Omega)}{\sqrt{\Phi_{Z_m Z_m}(j\Omega) \Phi_{Z_{m'} Z_{m'}}(j\Omega)}} \\ &= \frac{\Gamma_{DDR}(j\Omega) e^{-j\Omega f_s \Delta\tau_m} + \gamma_{N_m N_{m'}}(j\Omega)}{\Gamma_{DDR}(j\Omega) + 1},\end{aligned}\quad (1.3)$$

wobei  $\Gamma_{DDR}(j\Omega)$  das Verhältnis aus den Energien des Direktschalles zum diffusen Schall beschreibt. Eine wesentliche Grundvoraussetzung ist dabei die

- 5 Beobachtung, dass der Direktschall hauptsächlich aus dem Sprachsignal 5, und der diffuse Schall hauptsächlich aus dem Störgeräusch 3 besteht:

$$\Gamma_{DDR}(j\Omega) = \frac{\Phi_{SS}(j\Omega)}{\Phi_{NN}(j\Omega)} = \frac{\gamma_{N_m N_{m'}}(j\Omega) - \gamma_{Z_m Z_{m'}}(j\Omega)}{\gamma_{Z_m Z_{m'}}(j\Omega) - e^{-j\Omega f_s \Delta\tau_{mm'}}}. \quad (1.4)$$

Dabei wird das Schallfeld des Störgeräusches 3 als ideal isotropisch betrachtet:

10

$$\gamma_{N_m N_{m'}}(j\Omega) = \frac{\sin(kd_{mm'})}{kd_{mm'}}. \quad (1.5)$$

Durch Integration von Gleichung 1.4 über das komplexe Fourierspektrum erhält man eine Kostenfunktion, welche ein eindeutiges Maximum beim Einfallswinkel des Sprachsignals 5 aufweist:

15

$$J(\Delta\tau_{mm'}) = \int_{\Omega=0}^{2\pi} \frac{\gamma_{N_m N_{m'}}(j\Omega) - \gamma_{Z_m Z_{m'}}(j\Omega)}{\gamma_{Z_m Z_{m'}}(j\Omega) - e^{-j\Omega f_s \Delta\tau_m}} d\Omega \quad (1.6)$$

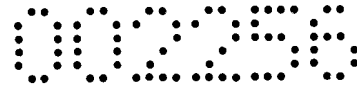
Durch Vorgabe von 50 möglichen Einfallswinkeln  $m$  zwischen  $-90^\circ$  und  $+90^\circ$  und jeweiliges Lösen von Gleichung 1.6 erhält man den wahrscheinlichsten

- 20 Einfallswinkel als Maximum dieser Kostenfunktion:

$$\Delta\hat{\tau}_m = \arg \max_{\Delta\tau_m} J(\Delta\tau_m). \quad (1.7)$$

Dabei erhält man eine globale Bedingung für das Vorhandensein eines

- 25 Sprachsignales 5 durch  $J(\Delta\tau_m) > 1$ . Diese Bedingung wird verwendet um ein



Zusatzmodul 17 (RTF-estimation Modul 17), welches später noch im Detail beschrieben wird und in Fig. 1 dargestellt ist, aufzurufen. In diesem Zusatzmodul 17 wird der Absolutbetrag der relativen Übertragungsfunktion zwischen dem m-ten und dem m'-ten Mikrofon 10, 11 geschätzt. Dieser Betrag wird aus den Kurzzeit-  
 5 Leistungsdichtespektren der Mikrofonsignale berechnet:

$$|\hat{A}_m(j\Omega)| \approx \sqrt{\frac{\Phi_{Z_m Z_m}(j\Omega)}{\Phi_{Z_{m'} Z_{m'}}(j\Omega)}} \quad (1.8)$$

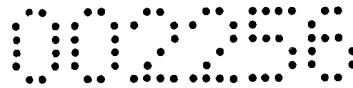
Diese Schätzung ist nur bei Vorhandensein eines Sprachsignales 5 gültig. Damit  
 10 und mit dem Einfallswinkel  $\Delta\tau_m$  ergibt sich die relative Transferfunktion  $\hat{A}(\Omega_j)$  zu

$$\hat{A}_m(j\Omega) = |\hat{A}_m(j\Omega)| e^{-j\Omega f_s \Delta\tau_m}. \quad (1.9)$$

Die korrekte Funktion des GSC-Beamformer-Modules 13, also des zweiten Moduls 13, hängt unmittelbar von dieser Transferfunktion ab. Sie wird für jedes  
 15 der m Mikrofone 10, 11 berechnet.

Zusammenfassend kann also gesagt werden, dass im ersten Modul 12 die Schätzung der Richtung, aus welcher der Nutzschaall auf das Mikrofonarray trifft (also der Ort, wo der Sprecher 4 vor dem Gerät steht) beschrieben wird. Der  
 20 Störschaall 3 kann dabei jegliche Art von Lärm sein, aber auch das Echo von einer ggf. vorhandenen Gegensprechanlage, wozu das Zusatzmodul 17 vorhanden ist. Dem Störgeräusch 3 ist dabei nicht zwingend ein akustisches Zentrum zugeordnet. Das erste Modul 12 umfasst also die Bestimmung der Einfallsrichtung des Sprachsignales 5, welche für die nachfolgenden Module 13 und 14 benötigt  
 25 wird. Nachstehende Skizze in Fig. 3 veranschaulicht das Zustandekommen der Einfallsrichtung eines Sprachsignales 5  $s(k)$  am Mikrofonarray.

In Fig. 3 ist eine schaubildliche Darstellung zum Veranschaulichen vom Zustandekommen der Einfallsrichtung des Sprachsignals 5 am Mikrofon 10, 11  
 30 gezeigt. Hierzu wurden jedoch eine Anordnung mit drei Mikrofonen 10 und 11 als Mikrofonarray gezeigt, um eine bessere Wirkungsweise erkennen zu können.



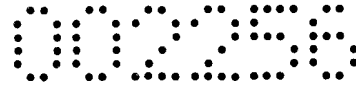
Daraus ist ersichtlich, dass sich aufgrund der unterschiedlichen Positionen der Mikrofone 10, 11 unterschiedliche Abstände 18 bis 20 zu den Mikrofonen 10, 11 ergeben. Somit kommen auch unterschiedlich Aufnahmezeiten zu Stande, sodass aufgrund der unterschiedlichen Aufnahmezeiten auch die Richtung der  
 5 Schallquelle 2, insbesondere des Sprachsignals 5, grob herleitbar ist.

Zurückkommend zu Fig.1 wird die ermittelte Richtung des Sprachsignals 5 und/oder die Daten hierfür vom ersten Modul 12 an das zweite Modul 13 übergeben. Im zweiten Modul 13 erfolgt die Aufteilung der über die Mikrofonkanäle  
 10 aufgenommenen Audiosignale 6 in ein Sprachsignal 5 und in ein Störgeräusch 3. Dabei werden bei der Aufteilung des Audiosignales 6 in das Sprachsignal 5 und das Störgeräusch 3 die Richtung des Sprachsignals 5 und/oder die Daten des ersten Moduls 12 berücksichtigt.

15 Dieses Modul 13 liefert also eine Schätzung des Nutz- und Störschalles getrennt voneinander. Diese Schätzung ist umso genauer, je exakter die Richtung des Sprechers 4 im vorher genannten Modul 12 ermittelt werden kann. Dem Sprachsignal 5 wird dabei ein gerichtetes Schallfeld zu Grunde gelegt, und den Störgeräuschen 3 ein diffuses bzw. isotropes Schallfeld.

20 Dabei sind schon einige aus dem Stand der Technik bekannte Verfahren zum Trennen des Audiosignals 6 in das Sprachsignal 5 und den Störgeräuschen 3 bekannt, die entsprechend eingesetzt werden, wozu hierzu nicht mehr näher auf den Ablauf des Trennverfahrens eingegangen wird.

25 Wesentlich ist nur, dass zuerst die Richtungsbestimmung mit dem gesamten aufgenommen Schallspektrum, also den Störschall 3 und dem Sprachsignal 5, durchgeführt wird und erst anschließend die Auftrennung der Signale vorgenommen wird. Damit können auch für die Richtungsbestimmung die  
 30 Störsignale 3 herangezogen werden. Würde nämlich zuerst eine Aufteilung in die Störgeräusche 3 und das Sprachsignal 5 erfolgen und für die Richtungsbestimmung nur das Sprachsignal 5 herangezogen werden, so würde die Daten von den Störsignalen 3 nicht mehr zur Verfügung stehen und die



Richtungsbestimmung nicht so exakt erfolgen können. Grundsätzlich wäre es möglich, dass die Daten für die Störsignale 3 gespeichert werden und für die Richtungsbestimmung mit einfließen können, sodass die beiden Module getauscht werden könnten. Hierbei ist jedoch der Rechenaufwand höher und es müssten die  
5 Algorithmen entsprechend angepasst werden.

Nachdem eine Auftrennung des Audiosignals 6 im zweiten Modul 13 vollbracht ist, werden die Daten bzw. die Signale, insbesondere das Sprachsignal 5 und die Störgeräusche 3, an das dritte Modul 14 übergeben. Im dritten Modul 14, im so  
10 genannten Multichannel-Postfilter-Modul, kommt ein mathematisches Verfahren basierend auf einer räumlichen Autokorrelationsmatrix der Signale zum Einsatz, welches das Verhältnis von Sprachsignal-Energie zu Störsignal-Energie (SNR) im Ausgabesignal des zweiten Modules 13 ermittelt.

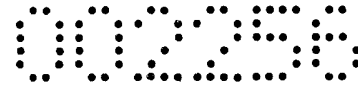
15 Unter Verwendung der Filterfunktion  $W(j\Omega)$  des GSC-Beamformers (drittes Modul 14) kann die Schätzung des Sprachsignales 5 und des Störgeräusches 3 am GSC-Modul 13 wie folgt angegeben werden:

$$\begin{aligned}\hat{S}(j\Omega) &= \hat{A}(j\Omega)W^H(j\Omega)Z(j\Omega) \\ \hat{N}(j\Omega) &= Z(j\Omega) - \hat{S}(j\Omega).\end{aligned}\tag{2.1}$$

20

Dabei beschreibt der Vektor  $Z(j\Omega)$  das Audiosignal 6 an allen  $M$  Mikrofonen 10, 11 nach Gleichung 1.1. Der Vektor  $\hat{A}(j\Omega)$  beschreibt die relativen Transferfunktionen zwischen den Mikrofonen 10, 11 aus Gleichung 1.9, welche mit dem DOA Algorithmus ermittelt wurden. Die Kurzzeit-Spektralschätzungen des  
25 Sprachsignales 5 und des Störgeräusches 3 erhält man durch direkte Berechnung der jeweiligen Signalenergien:

$$\begin{aligned}\Phi_{\hat{S}\hat{S}} &= \sum_{m=1}^M |\hat{S}_m(j\Omega)|^2 \\ \Phi_{\hat{N}\hat{N}} &= \sum_{m=1}^M |\hat{N}_m(j\Omega)|^2.\end{aligned}\tag{2.2}$$



Unter der Annahme eines idealen isotropischen Störgeräusch-Schallfeldes lässt sich die Matrix  $\Gamma_{NN}$  definieren. Sie beschreibt die räumliche Kohärenz zwischen allen Kombinationen aus allen Mikrofonsignalen. Jedes Element dieser Matrix ist durch Gleichung 1.5 definiert. Damit ergeben sich die langzeit-

- 5 Spektraldichteschätzungen des Sprachsignales 5 und des Störgeräusches 3:

$$\begin{aligned}\tilde{\Phi}_{\hat{S}\hat{S}} &= \hat{A}W^H\Gamma_{NN}W\hat{A}^H \\ \tilde{\Phi}_{\hat{N}\hat{N}} &= \Gamma_{NN} - \Gamma_{NN}W\hat{A}^H - \hat{A}W^H\Gamma_{NN} + \hat{A}W^H\Gamma_{NN}W\hat{A}^H.\end{aligned}\quad (2.3)$$

Hierbei wurde zur besseren Lesbarkeit das komplexe Argument  $j\Omega$  weggelassen.

- 10 Das gesuchte Verhältnis der Energien des Sprachsignales 5 und der Störgeräusche 3 am Ausgang des GSC-Beamformer-Modules 13 ergibt sich dann zu:

$$\xi(j\Omega) = \frac{\text{Tr}(\tilde{\Phi}_{\hat{S}\hat{S}})}{\text{Tr}(\tilde{\Phi}_{\hat{N}\hat{N}})} \frac{\Phi_{\hat{N}\hat{N}}}{\Phi_{\hat{S}\hat{S}}} - 1. \quad (2.4)$$

15

Zur effizienten Implementation von Gleichung 2.3 werden folgende Identitäten verwendet:

$$\begin{aligned}\text{Tr}(\Gamma_{NN}) &= M \\ \text{Tr}(\Gamma_{NN}W\hat{A}^H) &= \text{Tr}(\hat{A}W^H\Gamma_{NN}) = \sum_{m=1}^M A_m (W^H\Gamma_{NN})_m \\ \text{Tr}(\hat{A}W^H\Gamma_{NN}W\hat{A}^H) &= \text{Tr}(\hat{A}^H\hat{A}W^H\Gamma_{NN}W) = \sum_{m=1}^M |A_m|^2 W^H\Gamma_{NN}W.\end{aligned}\quad (2.5)$$

20

Nachdem die Auswertung durchgeführt wurde, wird das ermittelte Sprachsignal 5 an die Ausgabeeinheit 9 übergeben. Im dritten Modul 14 werden die vom zweiten Modul 13 übergebenen Signale annähernd in Echtzeit, vorzugsweise mit 32 ms Verzögerung, verarbeitet werden, sodass so gut wie keine wahrnehmbare Zeitverzögerung auftritt. Weiters wird vom dritten Modul 14 eine Absenkung des Störgeräusches 3 um 25 dB erreicht.

25



Weiters ist aus Fig. 1 ersichtlich, dass an die Ausgabeeinheit 9 ein Ausgabemittel 21, insbesondere ein Lautsprecher 22, verbunden ist. Dabei ist das Ausgabemittel 21 nicht direkt in der Kommunikationsanlage 1 integriert, sondern über Leitungen extern angeordnet. Selbstverständlich ist es möglich, dass Ausgabemittel 21 direkt  
5 in die Kommunikationsanlage 1 zu integrieren, um beispielsweise einen kompakten Aufbau zu erhalten. Es ist aber auch möglich, die Kommunikationsanlage 1 dezentral zur Spracheingabeposition, insbesondere dem Ort des Nutzers 4, zu positionieren, sodass dadurch das Aufnahmemittel 7 vor Ort beim Nutzer 4 angeordnet ist und dieses über eine oder mehrere Leitungen mit  
10 der Kommunikationsanlage 1, insbesondere der Auswerteeinheit 8, verbunden wird.

Damit bei der Installation der Kommunikationsanlage 1 eine optimale Einstellung vorgenommen werden kann, sind in der Kommunikationsanlage 1  
15 unterschiedlichsten Tools und/oder Test-Software, wie beispielsweise ein Installationstool, ein Spracherkennungstool, Sprachtest, usw., integriert. Beispielsweise kann eine optimale Parametrierung der Auswerteeinheit 8 über umfangreiche automatisierte Sprachtests, welche auf deutsche und englische Sprachkorpora basieren, erfolgen.

20 Auch ist es möglich, dass die Kommunikationsanlage 1 eine oder mehrere Schnittstellen 23 aufweist, mit der Daten mit externen angeschlossenen bzw. drahtlos verbundenen Komponenten ausgetauscht und/oder gesteuert werden. Dabei wird beispielsweise eine zusätzliche weitere Software zur Steuerung  
25 angeschlossener Komponenten in der Kommunikationsanlage 1 abgearbeitet und/oder Daten von der Kommunikationsanlage 1 an angeschlossene bzw. verbundenen Komponenten übergeben.

Wie zuvor bereits erwähnt, kann in der Kommunikationsanlage 1 ein weiteres  
30 zusätzliches Modul 17, insbesondere das sogenannte RTF-Estimation-Modul 17, welches aus der Richtungsinformation des ersten Moduls 12 die relativen akustischen Transferfunktionen zwischen den Mikrofonen 10, 11 berechnet, eingesetzt werden. Bei einem Einsatz der Kommunikationsanlage 1 in einer

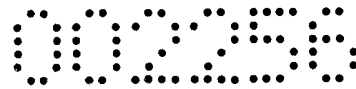


- Gegensprechanlage wird ein weiteres Modul 24, insbesondere ein sogenanntes Echochancellor-Modul 24, zur Reduzierung von akustischen Echos, zusammen mit den Modulen 12 bis 14 eins bis drei in einer einzigen Betriebssoftware eingesetzt, wie dies schematisch dargestellt ist. Darüber hinaus ist es möglich,
- 5 dass die Anordnung der Mikrofone 10, 11 während des Betriebes ermittelt wird, wobei lediglich ein Bereich für den Abstand der Mikrofone 10, 11 vorgegeben wird, d.h., dass Vorort eine optimale Einstellung im Bezug auf die mögliche Sprachposition des Nutzers 4 vorgenommen werden kann.
- 10 Aufgrund der optimale Sprachaufbereitung ist es auch möglich, dass das resultierende Sprachsignal 5 entweder direkt ausgegeben wird und/oder, vorzugsweise permanent, in einem Sprachspeicher abgelegt wird, wobei der Signaltransport über ein beliebiges Medium und/oder Protokoll erfolgen kann. Somit können von der Kommunikationsanlage aufgrund von hinterlegten
- 15 Sprachbefehlen angeschlossene Komponenten aktiviert und deaktiviert werden.

- Man kann also sagen, dass die Sprachverständlichkeit erhöht wird, indem bei gleichbleibender Sprachqualität die Lautheit der Störgeräusche 3 vermindert wird. Dies erfolgt auf elektronischem Wege als Teil der Betriebssoftware der
- 20 Kommunikationsanlagen 1. Das Verfahren nutzt dabei die Information mehrerer, insbesondere zweier Mikrofonkanäle, um zwischen Stör- und Nutzschall 3 und 5 zu unterscheiden. Grundsätzlich wird weiters darauf hingewiesen, dass anstelle eines modularen Aufbaus jede andere bekannte Aufbaumöglichkeit gewählt werden kann, insbesondere durch eine Mikroprozessorsteuerung. Hierbei ist es
- 25 lediglich wichtig, dass bei einer Implementierung in andere Systeme bzw. Aufbauten die entsprechenden Algorithmen verwendet werden, um eine verbesserte Sprachausgabe zu erreichen.

- In Fig. 4 ist eine Schrankenanlage 25 beispielsweise für ein Parkhaus u/o
- 30 Tiefgarage, usw., schematisch und symbolhaft dargestellt.

Wie schematisch dargestellt, befindet sich die Schrankenanlage 25 an einer Zufahrt 26. Üblicherweise weisen derartige Anlagen dezentrale Meldezentralen 27

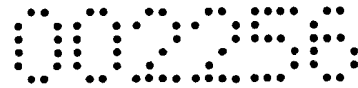


auf, in der eine Zentraleinheit 28 zur Steuerung aller Komponenten angeordnet ist, d.h., dass an der Zentraleinheit 28 die Kommunikationsanlage 1, die bevorzugt ebenfalls dezentral positioniert ist, eine Steuereinheit 29 der Schrankenanlage 25 mit dem Schranken 30 angeschlossen bzw. verbunden sind, wobei hierzu

5 bevorzugt ein serielles oder paralleles Daten-/Steuer-Netzwerk 31 verwendet wird. Hierbei ist es möglich, dass beispielsweise nunmehr die Auswertung des Sprachsignals 5 in der Zentraleinheit 28 erfolgt, sodass an jenem Ort, wo sich der Nutzer 4 befindet, also im Ein- oder Ausgangsbereich, die Kommunikationsanlage 1 durch das Aufnahmemittel 7 gebildet wird. Die Wiedergabe des Sprachsignals 5  
10 über die Ausgabeeinheit 9, insbesondere über das Ausgabemittel 21, erfolgt in der Meldezentrale 27. Da bei derartigen Anlagen jedoch ein gegenseitiger Sprechverkehr notwendig ist, ist sowohl in der Meldezentrale 27 als auch an der Kommunikationsanlage 1 Vorort beim Nutzer ein Aufnahmemittel 7 und ein Ausgabemittel 21, gemäß Beschreibung zu Fig. 1, vorhanden. Dadurch kann der  
15 Nutzer 4 an der Schrankenanlage 21 mit einem Nutzer 4 in der Meldezentrale (nicht dargestellt) kommunizieren. Selbstverständlich ist es auch möglich, dass in der Kommunikationsanlage 1 die Auswertung des Sprachsignals 5 stattfinden kann, d.h., also, dass die Zentraleinheit und/oder Kommunikationsanlage zur Durchführung des Verfahrens, wie zuvor beschrieben, ausgebildet ist.

20

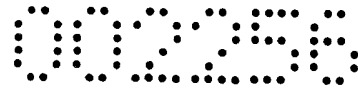
In Fig. 5 ist ein weiteres Ausführungsbeispiel für die Anwendung einer Kommunikationsanlage 1 gezeigt. Hierbei wird die Kommunikationsanlage 1 in einer Notrufanlage 32 beispielsweise für Verkehrsstraßen, insbesondere Autobahnen 33, Schnellstraßen, Highways, usw., eingesetzt, wobei hierzu eine  
25 Vielzahl von Meldestellen 34 installiert bzw. eingesetzt werden, d.h., dass bei diesem Ausführungsbeispiel mehrere Kommunikationsanlagen 1 zusammen wirken, wobei jede Kommunikationsanlage 1 für sich eigenständig das DCSE-System zum Aufbereiten des Sprachsignals 5 durchführt oder dies zentral in einer Zentraleinheit 28 für alle Meldestellen 34 mit der Kommunikationsanlage 1  
30 durchführt, d.h., dass die Zentraleinheit 28 und/oder Kommunikationsanlage 1 zur Durchführung des Verfahrens, wie zuvor beschrieben, ausgebildet ist.



Bei dem Ausführungsbeispiel weist jede Meldestelle 34 und die Zentraleinheit 28 eine Übertragungseinheit 35 auf. Die Übertragungseinheit 35 ist durch Datenleitungen, Telefonleitungen, Funksignale und/oder Internetverbindungen gebildet. Dabei ist es auch möglich, dass die unterschiedlichen Systeme miteinander kombiniert werden, d.h., dass beispielsweise alle Meldestellen 34 über Datenleitungen 36 miteinander verbunden sind und an einer Zentrallen Stelle 37 eine Übersetzung auf ein Funksignal oder eine Internetverbindung 38 für das Internat 39 stattfindet, sodass eine weit entfernte Zentraleinheit 28 über das Funksignal oder die Internetverbindung 38 verbunden wird, wie dies schematisch beim Ausführungsbeispiel in Fig. 5 dargestellt ist. Damit ein unverwechselbarer Datenaustausch zwischen den Meldestellen 34 untereinander und der Zentraleinheit 28 reibungslos funktioniert, ist es von Vorteil, wenn jede Meldestelle 34 und somit die Kommunikationseinheit 1 und die Zentraleinheit 28 codiert sind, sodass die einzelnen Meldestellen 34 einzelnen angesprochen werden können. Dabei ist es auch möglich, dass die Ausgabeeinheit 9 für das Sprachsignal 5 mit der Übertragungseinheit 35 für die externe Meldestelle 34 verbunden ist, sodass von der Zentraleinheit 28 direkt die Ausgabeeinheit 9 aktiviert werden kann, bzw. ein Sprachsignal 5 vom Nutzer 4 in der Zentrale ausgegeben werden kann.

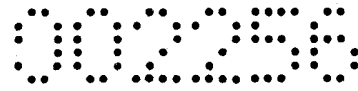
Grundsätzlich wird darauf hingewiesen, dass die Kommunikationseinheit 1 auch in anderen Gebieten eingesetzt werden kann, wie beispielsweise in Verkehrstunnel, Industrieparks, Hotelanlagen, usw. Auch ist es möglich, dass die Kommunikationseinheit 1 noch weitere Module, wie beispielsweise ein Display zum Anzeigen von Informationen, ein Eingabeblock zum Eingeben von Codes, Nummern, Namen, usw., aufweist. Weiters ist es auch möglich, dass zum Aktivieren der Kommunikationsanlage 1 diese mit einem Aktivierungsmittel verbunden sind. Insbesondere wenn die Kommunikationsanlage 1 als Gegensprechanlage eingesetzt wird, kann diese mit einem Aktivierungsknopf bzw. Schalter verbunden sein.

Der Ordnung halber sei abschließend darauf hingewiesen, dass zum besseren Verständnis des Aufbaus dieser bzw. dessen Bestandteile teilweise unmaßstäblich und/oder vergrößert und/oder verkleinert dargestellt wurden.



Die den eigenständigen erfinderischen Lösungen zugrunde liegende Aufgabe kann der Beschreibung entnommen werden.

- 5 Vor allem können die einzelnen in den Fig. 1 bis 3, 4, 5 gezeigten Ausführungen den Gegenstand von eigenständigen, erfindungsgemäßen Lösungen bilden. Die diesbezüglichen, erfindungsgemäßen Aufgaben und Lösungen sind den Detailbeschreibungen dieser Figuren zu entnehmen.



## Bezugszeichenliste:

|    |    |                      |
|----|----|----------------------|
|    | 1  | Kommunikationsanlage |
|    | 2  | Schallquelle         |
| 5  | 3  | Störgeräusche        |
|    | 4  | Nutzer               |
|    | 5  | Sprachsignal         |
|    | 6  | Audiosignal          |
|    | 7  | Aufnahmemittel       |
| 10 | 8  | Auswerteeinheit      |
|    | 9  | Ausgabeeinheit       |
|    | 10 | Mikrofon             |
|    | 11 | Mikrofon             |
|    | 12 | Modul                |
| 15 | 13 | Modul                |
|    | 14 | Modul                |
|    | 15 | Hauptkeule           |
|    | 16 | Nebenkeule           |
|    | 17 | Zusatzmodul          |
| 20 | 18 | Abstand              |
|    | 19 | Abstand              |
|    | 20 | Abstand              |
|    | 21 | Ausgabemittel        |
|    | 22 | Lautsprecher         |
| 25 | 23 | Schnittstelle        |
|    | 24 | Modul                |
|    | 25 | Schrankenanlage      |
|    | 26 | Zufahrt              |
|    | 27 | Meldezentrale        |
| 30 | 28 | Zentraleinheit       |
|    | 29 | Steuereinheit        |
|    | 30 | Schranken            |
|    | 31 | Netzwerk             |



- 32 Notrufanlage
- 33 Autobahn
- 34 Meldestelle
- 35 Übertragungseinheit
- 5 36 Datenleitung
- 37 Stelle
- 38 Internetverbindung
- 39 Internet

10

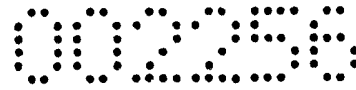


### **Patentansprüche:**

1. Verfahren zur Erhöhung der Sprachverständlichkeit eines Audiosignals (6), das  
 5 aus einem Gemisch von Störgeräuschen (3) und Sprachsignalen (5) besteht, in  
 digitalen elektroakustischen Kommunikationsanlagen (1), dadurch  
 gekennzeichnet, dass die Sprachverständlichkeit durch Verminderung der  
 Lautstärke der Störgeräusche (3) vom Ort eines Sprechers (4) erreicht wird, wobei  
 zur Aufnahme des Audiosignals (6) zumindest zwei Mikrofonkanäle verwendet  
 10 werden, welche das Audiosignal (6) an eine Auswerteeinheit (8) weiterleiten, in  
 der das Störgeräusch (3) auf elektronischen Wege minimiert wird und das  
 erzeugte Sprachsignal (5) von der Kommunikationsanlage (1) über eine  
 Ausgabeeinheit (9) wiedergegeben wird und dass der Betrieb der  
 Kommunikationsanlage (1) und/oder der Auswerteeinheit (8) über eine  
 15 Betriebssoftware, sichergestellt wird und dass die Auswerteeinheit (8)

  - ein erstes Modul (12), insbesondere ein DOA-Estimation-Modul (12), zur  
 Bestimmung der Richtung des Sprachsignals (5),
  - ein zweites Modul (13), insbesondere ein GSC-Beamforming-Modul (13),  
 zur Feststellung und Trennung des Sprachsignals (5) von den  
 20 Störgeräuschen (3) und
  - ein drittes Modul (14), insbesondere ein Multichannel-Postfilter-Modul (14),  
 zur Subtraktion des Störgeräusches (3) vom Sprachsignal (5)

umfasst.
- 25 2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass im ersten Modul  
 (12), dem DOA-Estimation-Modul (12), eine Richtwirkung der  
 Mikrofonempfindlichkeit in Richtung des Sprechers (4) erzeugt wird, wobei eine  
 Hauptkeule (15) mit der Bewegung des Sprechers (4) nachgeführt wird.
- 30 3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, dass die Hauptkeule (15)  
 eine Breite von  $-20^\circ$  und  $+20^\circ$  um die aktuelle Position des Sprechers (4) umfasst,  
 wobei der nutzbare Frequenzbereich 20 bis 20 000 Hz umfasst.



4. Verfahren nach mindestens einem der vorhergehenden Ansprüche 2 oder 3, dadurch gekennzeichnet, dass die Breite der Hauptkeule (15) mit steigender Frequenz in der Bandbreite abnimmt.

5. Verfahren nach mindestens einem der vorhergehenden Ansprüche 2 bis 4, dadurch gekennzeichnet, dass die Hauptkeule (15) derart ausgelegt wird, dass minimale Schwankungen, wie beispielsweise Kopfbewegungen des Sprechers (4), automatisch ausgeglichen werden.

10 6. Verfahren nach mindestens einem der vorhergehenden Ansprüche 1 bis 5, dadurch gekennzeichnet, dass im ersten Modul (12), insbesondere in dem DOA-Estimation Modul (12), folgender Algorithmus „Direct-to-Diffusion Ratio“ zur Schätzung bzw. Festlegung der Richtung des Sprachsignals (5) läuft bzw. abgearbeitet wird.

15

7. Verfahren nach mindestens einem der vorhergehenden Ansprüche 1 bis 6 dadurch gekennzeichnet, dass das Audiosignal (6), insbesondere ein Mikrofonsignal, am Mikrofon (10, 11) durch das Störgeräusch (3) und das Sprachsignal (5) derart zusammen setzt und derart definiert wird

20 
$$Z_m(j\Omega) = S(j\Omega)e^{-j\Omega f_s \Delta\tau_m} + N(j\Omega)$$
,

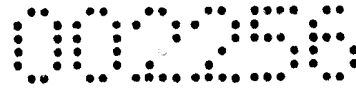
wobei  $Z_m(j)$  das Audiosignal (6), insbesondere des m-ten Mikrofonsignal,  $N(j)$  das Störsignal (3),  $S(j)$  das Sprachsignal (5) und  $\Delta\tau_m$  die Richtung definieren.

25

8. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 7, dadurch gekennzeichnet, dass im ersten Modul (12) eine Spektraldichteschätzung über eine Zeitmittelung entsprechend dem folgenden Algorithmus

30 
$$\mathbb{P}_{Z_m Z_m}(j\Omega) = \mathbb{E}\{|Z_m(j\Omega)|^2\} = \Phi_{SS}(j\Omega) + \Phi_{NN}(j\Omega).$$

berechnet wird.



9. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 8, dadurch gekennzeichnet, dass aus der Spektraldichteschätzung eine komplexe räumlich Kohärenz zwischen dem ersten Audiosignal (6) und dem weiteren Audiosignal (6), insbesondere Mikrofonsignalen, über den Algorithmus

$$\begin{aligned}\gamma_{Z_m Z_{m'}}(j\Omega) &= \frac{\Phi_{Z_m Z_{m'}}(j\Omega)}{\sqrt{\Phi_{Z_m Z_m}(j\Omega) \Phi_{Z_{m'} Z_{m'}}(j\Omega)}} \\ &= \frac{\Gamma_{DDR}(j\Omega) e^{-j\Omega f_s \Delta\tau_m} + \gamma_{N_m N_{m'}}(j\Omega)}{\Gamma_{DDR}(j\Omega) + 1}\end{aligned}$$

berechnet wird, wobei  $\Gamma_{DDR}(j\Omega) + 1$  das Verhältnis aus den Energien des Direktschalles zum diffusen Schall beschreibt.

10. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 9, dadurch gekennzeichnet, dass aus dem Sprachsignal (5) der Direktschall und aus dem Störsignal (3) der diffuse Schall gebildet wird und ein derartiges Verhältnis wie folgt bilden:

$$\Gamma_{DDR}(j\Omega) = \frac{\Phi_{SS}(j\Omega)}{\Phi_{NN}(j\Omega)} = \frac{\gamma_{N_m N_{m'}}(j\Omega) - \gamma_{Z_m Z_{m'}}(j\Omega)}{\gamma_{Z_m Z_{m'}}(j\Omega) - e^{-j\Omega f_s \Delta\tau_{mm'}}}.$$

11. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 10, dadurch gekennzeichnet, dass das Störgeräusch (3) als ideal isotropisch betrachtet wird und derart definiert wird:

$$\gamma_{N_m N_{m'}}(j\Omega) = \frac{\sin(kd_{mm'})}{kd_{mm'}}$$

12. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 11, dadurch gekennzeichnet, dass durch Integration über das komplexe Fourierspektrum entsprechend der Formel

$$J(\Delta\tau_{mm'}) = \int_{\Omega=0}^{2\pi} \frac{\gamma_{N_m N_{m'}}(j\Omega) - \gamma_{Z_m Z_{m'}}(j\Omega)}{\gamma_{Z_m Z_{m'}}(j\Omega) - e^{-j\Omega f_s \Delta\tau_m}} d\Omega$$

- eine Kostenfunktion gebildet wird.



13. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 12, dadurch gekennzeichnet, dass durch Vorgabe von mehreren, insbesondere fünfzig, möglichen Einfallswinkeln zwischen  $-90^\circ$  und  $+90^\circ$  man den wahrscheinlichen Einfallswinkel für das Maximum der Kostenfunktion gemäß der
- 5 Formel

$$\Delta \hat{\tau}_m = \arg \max_{\Delta \tau_m} J(\Delta \tau_m)$$

erhält.

14. Verfahren nach mindestens einem oder mehreren der vorhergehenden
- 10 Ansprüche 1 bis 13, dadurch gekennzeichnet, dass im Zusatzmodul (17) „RTF-Estimation“ ein Betrag der Übertragungsfunktion zwischen dem m-ten und dem m'-ten Mikrofon gemäß der Formel

$$|\hat{A}_m(j\Omega)| \approx \sqrt{\frac{\Phi_{Z_m Z_m}(j\Omega)}{\Phi_{Z_{m'} Z_{m'}}(j\Omega)}}$$

berechnet wird.

15

15. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 14, dadurch gekennzeichnet, dass mit dem Vorhandensein eines Sprachsignals und einem ermittelten Einfallswinkel eine Transferfunktion gemäß der Formel

$$\hat{A}_m(j\Omega) = |\hat{A}_m(j\Omega)| e^{-j\Omega f_s \Delta \hat{\tau}_m}$$

20

ermittelt wird.

16. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 14, dadurch gekennzeichnet, dass im zweiten Modul (13) die
- 25 Aufteilung der über die Mikrofonkanäle aufgenommenen Audiosignale (6) in ein Sprachsignal (5) und in ein Störgeräusch (3) erfolgt.

17. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 16, dadurch gekennzeichnet, dass die ermittelte Richtung des



Sprachsignals (5) und/oder die Daten hierfür vom ersten Modul (12) an das zweite Modul (13) übergeben werden.

18. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 17, dadurch gekennzeichnet, dass bei der Aufteilung des Audiosignales (6) in das Sprachsignal (5) und dem Störgeräusch (3) die Richtung des Sprachsignals (5) und/oder die Daten des ersten Moduls (12) berücksichtigt werden.
19. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 18, dadurch gekennzeichnet, dass im dritten Modul (14), im so genannten Multichannel-Postfilter-Modul (14), ein mathematisches Verfahren basierend auf einer räumlichen Autokorrelationsmatrix der Signale zum Einsatz kommt, welches das Verhältnis von Sprachsignal-Energie zu Störsignal-Energie (SNR) im Ausgabesignal des zweiten Modules (13) ermittelt.
20. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 19, dadurch gekennzeichnet, dass im dritten Modul (14) unter Verwendung einer Filterfunktion "W(jΩ)" eines GSC-Beamformers die Schätzung des Sprachsignals (5) und des Störgeräusches (3) wie folgt angegeben wird:

$$\begin{aligned}\hat{S}(j\Omega) &= \hat{A}(j\Omega)W^H(j\Omega)Z(j\Omega) \\ \hat{N}(j\Omega) &= Z(j\Omega) - \hat{S}(j\Omega).\end{aligned}$$

dabei beschreibt der Vektor A(Ωj) das Audiosignal (6) an allen Mikrofonen (10, 11).

21. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 20, dadurch gekennzeichnet, dass sich aus der Formel

$$\begin{aligned}\Phi_{\hat{S}\hat{S}} &= \sum_{m=1}^M |\hat{S}_m(j\Omega)|^2 \\ \Phi_{\hat{N}\hat{N}} &= \sum_{m=1}^M |\hat{N}_m(j\Omega)|^2.\end{aligned}$$



die Langzeit-Spektraldichteschätzung des Sprachsignals (5) und respektive des Störgeräusches (3) ergibt.

22. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 21, dadurch gekennzeichnet, dass eine effizienten Implementation für die Langzeit-Spektraldichteschätzung gemäß der Formel

$$\begin{aligned} \text{Tr}(\Gamma_{NN}) &= M \\ \text{Tr}(\Gamma_{NN} \mathbf{W} \hat{\mathbf{A}}^H) &= \text{Tr}(\hat{\mathbf{A}} \mathbf{W}^H \Gamma_{NN}) = \sum_{m=1}^M A_m (\mathbf{W}^H \Gamma_{NN})_m \\ \text{Tr}(\hat{\mathbf{A}} \mathbf{W}^H \Gamma_{NN} \mathbf{W} \hat{\mathbf{A}}^H) &= \text{Tr}(\hat{\mathbf{A}}^H \hat{\mathbf{A}} \mathbf{W}^H \Gamma_{NN} \mathbf{W}) = \sum_{m=1}^M |A_m|^2 \mathbf{W}^H \Gamma_{NN} \mathbf{W} \end{aligned}$$

erfolgt.

23. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 22, dadurch gekennzeichnet, dass die Kurzzeit-Spektralschätzung durch direkte Berechnung gemäß der Formel

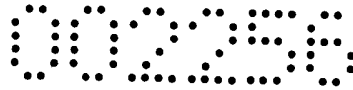
$$\begin{aligned} \tilde{\Phi}_{\hat{S}\hat{S}} &= \hat{\mathbf{A}} \mathbf{W}^H \Gamma_{NN} \mathbf{W} \hat{\mathbf{A}}^H \\ \tilde{\Phi}_{\hat{N}\hat{N}} &= \Gamma_{NN} - \Gamma_{NN} \mathbf{W} \hat{\mathbf{A}}^H - \hat{\mathbf{A}} \mathbf{W}^H \Gamma_{NN} + \hat{\mathbf{A}} \mathbf{W}^H \Gamma_{NN} \mathbf{W} \hat{\mathbf{A}}^H \end{aligned}$$

erhalten wird.

24. Verfahren nach mindestens einem oder mehreren der vorhergehenden Ansprüche 1 bis 23, dadurch gekennzeichnet, dass das Verhältnis des Sprachsignals (5) zum Störgeräusch (3) gemäß dem Algorithmus

$$\xi(j\Omega) = \frac{\text{Tr}(\tilde{\Phi}_{\hat{S}\hat{S}})}{\text{Tr}(\tilde{\Phi}_{\hat{N}\hat{N}})} \frac{\Phi_{\hat{N}\hat{N}}}{\Phi_{\hat{S}\hat{S}}} - 1$$

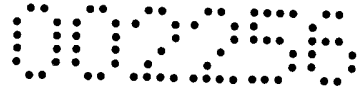
- berechnet wird.
25. Verfahren nach einem oder mehreren der vorhergehenden Ansprüche 1 bis 24, dadurch gekennzeichnet, dass im dritten Modul (14) die vom zweiten Modul (13) übergebenen Signale annähernd in Echtzeit, vorzugsweise mit 32 ms Verzögerung, verarbeitet werden.



26. Verfahren nach einem oder mehreren der vorhergehenden Ansprüche 1 bis 25, dadurch gekennzeichnet, dass vom dritten Modul (14) eine Absenkung des Störgeräusches (3) um 25 dB erreicht wird.
- 5 27. Verfahren nach einem oder mehreren der vorhergehenden Ansprüche 1 bis 26, dadurch gekennzeichnet, dass eine optimale Parametrierung der Auswerteeinheit (8) über umfangreiche automatisierte Sprachtests, welche auf deutsche und englische Sprachkorpora basieren, erfolgt.
- 10 28. Verfahren nach einem oder mehreren der vorhergehenden Ansprüchen 1 bis 27, dadurch gekennzeichnet, dass eine zusätzliche weitere Software zur Steuerung angeschlossener Komponenten in der Kommunikationsanlage (1) abgearbeitet wird und/oder Daten von der Kommunikationsanlage (1) an angeschlossene Komponenten übergeben werden.
- 15 29. Verfahren nach einem oder mehreren der vorhergehenden Ansprüche 1 bis 28, dadurch gekennzeichnet, dass bei einem Einsatz in einer Gegensprechanlage ein weiteres Modul (24), insbesondere ein sogenanntes Echochancellor-Modul (24), zur Reduzierung von akustischen Echos, zusammen mit den Modulen
- 20 (10,11,12) eins bis drei in einer einzigen Betriebssoftware eingesetzt wird.
30. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass ein weiteres zusätzliches Modul (17), insbesondere ein sogenanntes RTF-Estimation-Modul (17), welches aus der Richtungsinformation des ersten Moduls (12) die relativen
- 25 akustischen Transferfunktionen zwischen den Mikrofonen (10, 11) berechnet, eingesetzt wird.
31. Verfahren nach einem oder mehreren der vorhergehenden Ansprüche 1 bis 30, dadurch gekennzeichnet, dass die Anordnung der Mikrofone (10, 11) während
- 30 des Betriebes ermittelt wird, wobei lediglich ein Bereich für den Abstand der Mikrofone (10, 11) vorgegeben wird.



32. Verfahren nach einem oder mehreren der vorhergehenden Ansprüche 1 bis 31, dadurch gekennzeichnet, dass das resultierende Sprachsignal (5) entweder direkt ausgegeben wird und/oder, vorzugsweise permanent, in einem sprachspeicher abgelegt wird, wobei der Signaltransport über ein beliebiges Medium und/oder Protokoll erfolgt.
33. Verfahren nach einem oder mehreren der vorhergehenden Ansprüche 1 bis 32, dadurch gekennzeichnet, dass von der Kommunikationsanlage (1) aufgrund von hinterlegten Sprachbefehlen angeschlossene Komponenten aktiviert und deaktiviert werden.
34. Schrankensystem (25) mit einer Kommunikationsanlage (1), die eine Zentraleinheit aufweist, an der die Kommunikationsanlage (1), die Schrankenanlagen (25), eine Meldezentrale (27) und weitere Komponenten angeschlossen sind, dadurch gekennzeichnet, dass die Kommunikationsanlage (1) und/oder die Zentraleinheit (28) zur Durchführung des Verfahrens nach einem oder mehreren der Ansprüche 1 bis 18 ausgebildet ist.
35. Schrankensystem (25) nach Anspruch 18, dadurch gekennzeichnet, dass die Wiedergabe des Sprachsignals (5) über die Ausgabeeinheit (9) in der Meldezentrale (27) erfolgt.
36. Notrufanlage (32), umfassend eine Zentraleinheit (28) an der eine Kommunikationsanlage (1) angeschlossen ist, wobei die Zentraleinheit (28) bevorzugt über Übertragungseinheit (35) mit einer externen Meldestelle (34) verbunden ist, dadurch gekennzeichnet, dass die Zentraleinheit (28) und/oder Kommunikationsanlage (1) zur Durchführung des Verfahrens nach einem oder mehreren der Ansprüche 1 bis 18 ausgebildet ist.
37. Notrufanlage (32) nach Anspruch 36, dadurch gekennzeichnet, dass die Übertragungseinheit (35) durch Datenleitungen (36), Telefonleitungen, Funksignale und/oder Internetverbindungen (38) gebildet ist



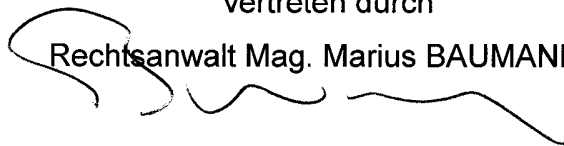
38. Notrufanlage (32) nach Anspruch 36 oder 37, dadurch gekennzeichnet, dass die Ausgabeeinheit (9) für das Sprachsignal (5) mit der Übertragungseinheit (35) für die externe Meldestelle (34) verbunden ist.

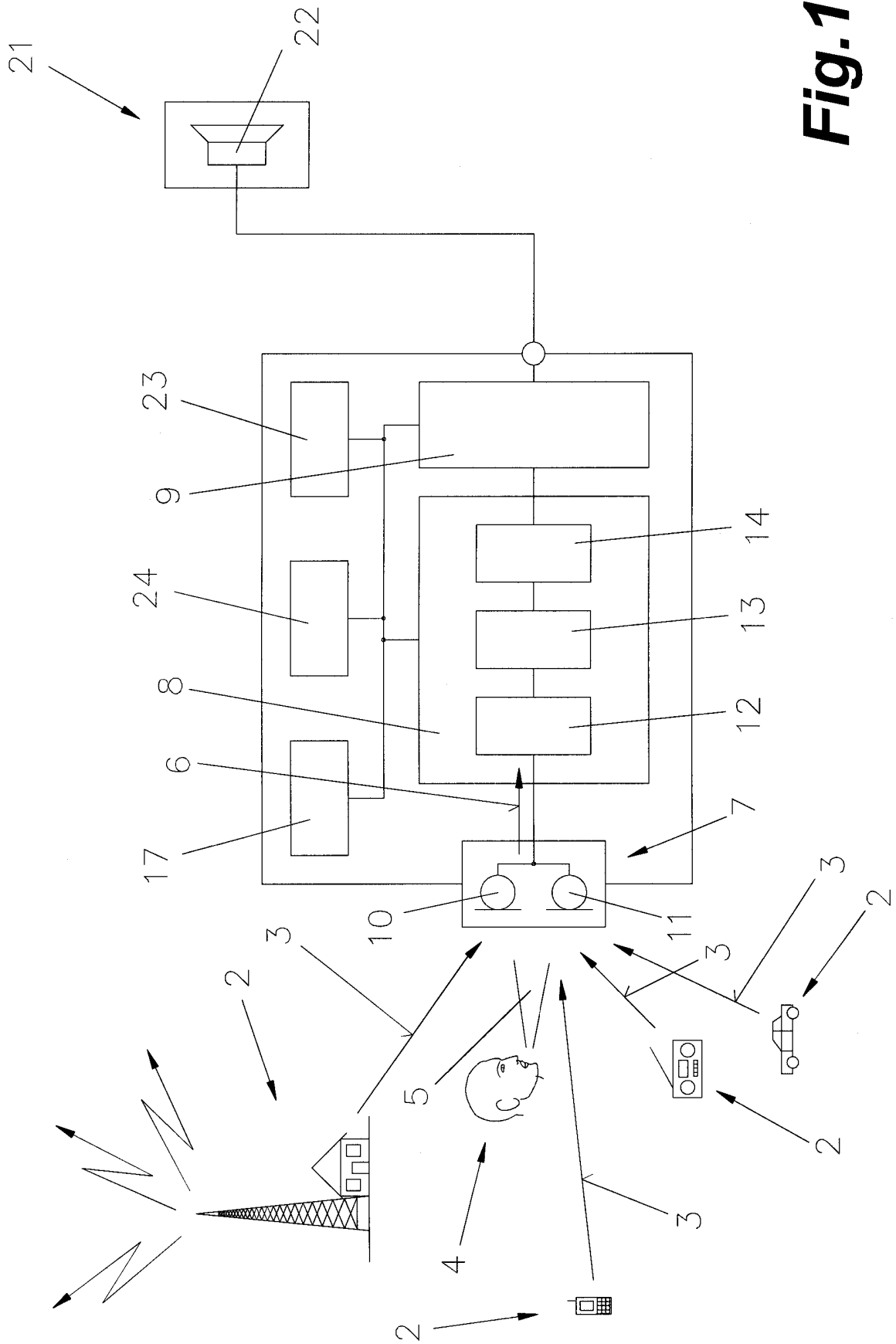
- 5 39. Kommunikationseinheit (1), die zur Durchführung des Verfahrens nach einem oder mehreren der Ansprüche 1 bis 18 geeignet ist, dadurch gekennzeichnet, dass diese in Verkehrstunnel, Industrieparks, Hotelanlagen, usw. einsetzbar ist.

10

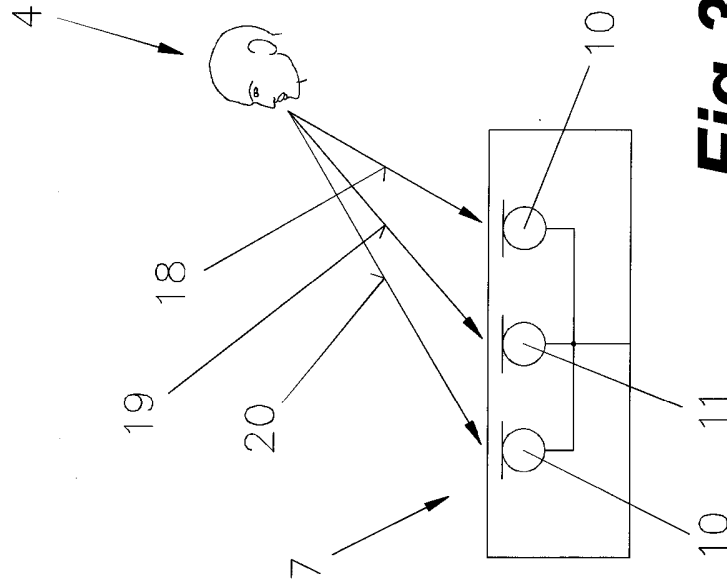
Commend International GmbH

vertreten durch

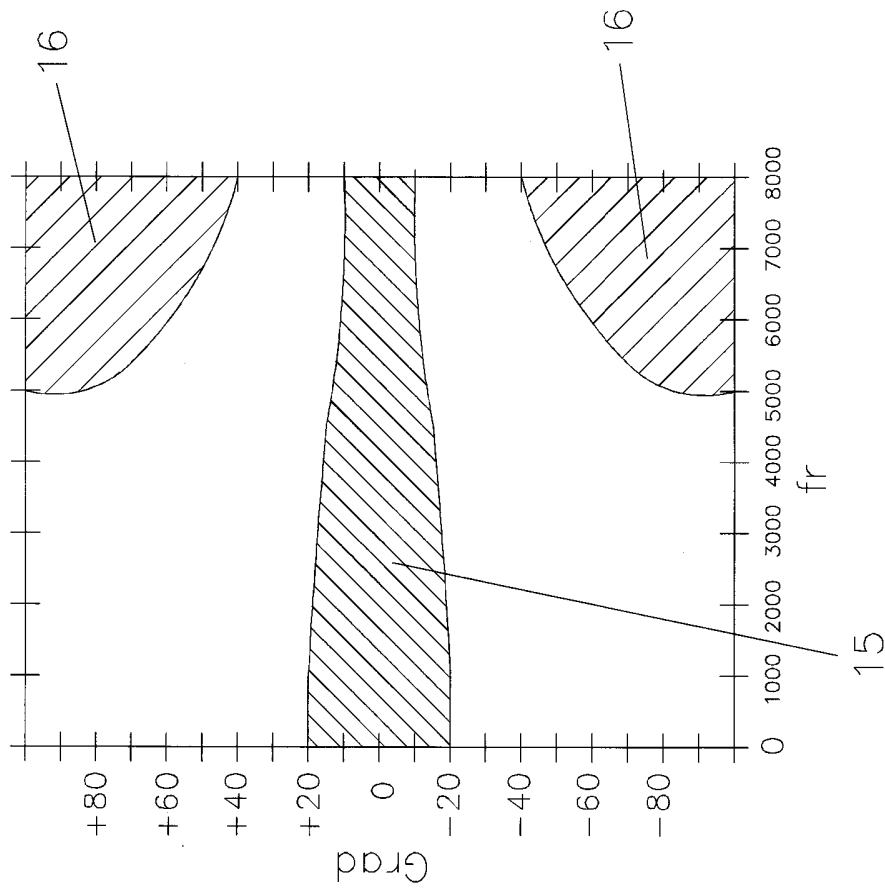
  
Rechtsanwalt Mag. Marius BAUMANN



**Fig.1**



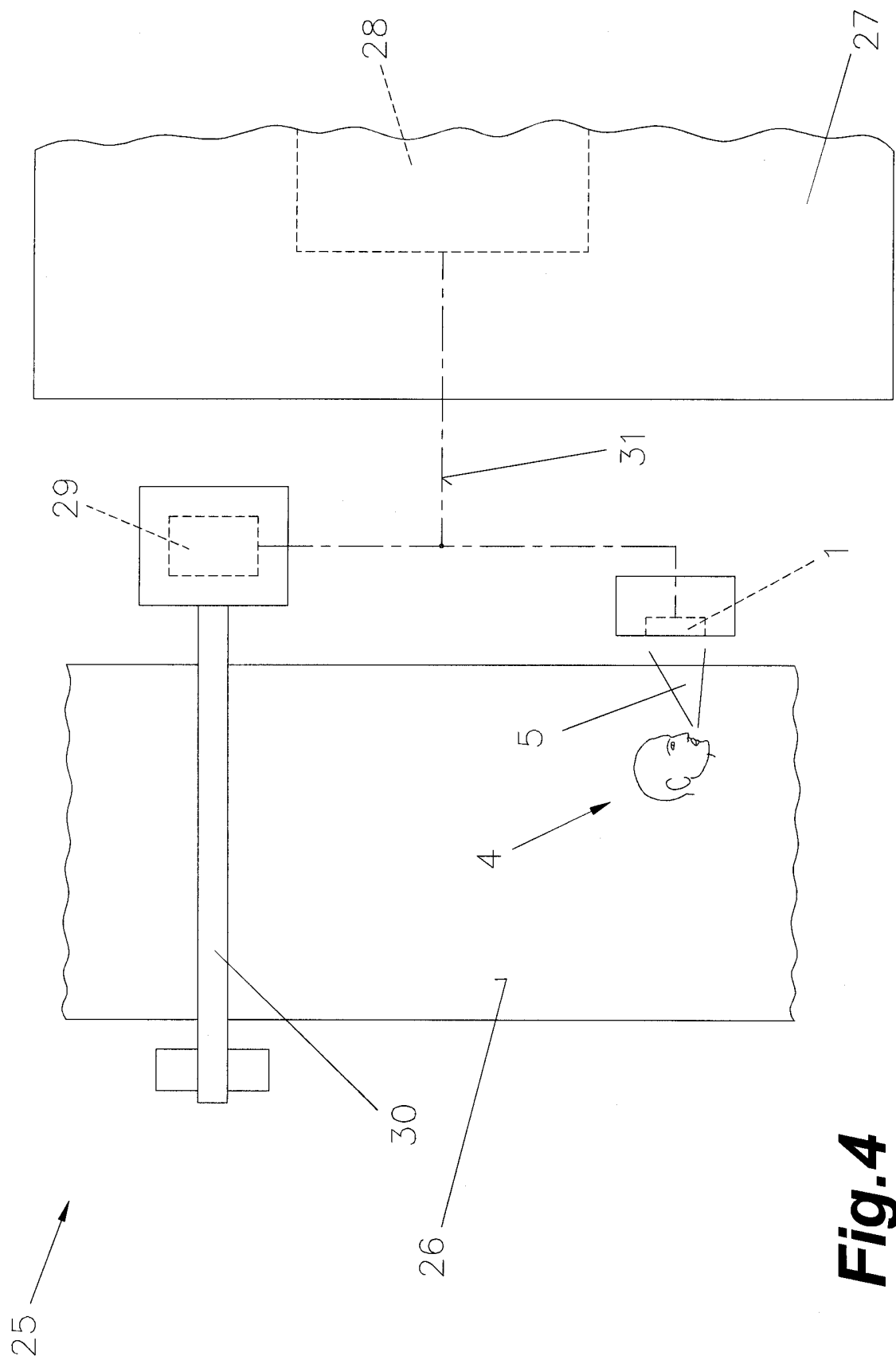
**Fig. 3**



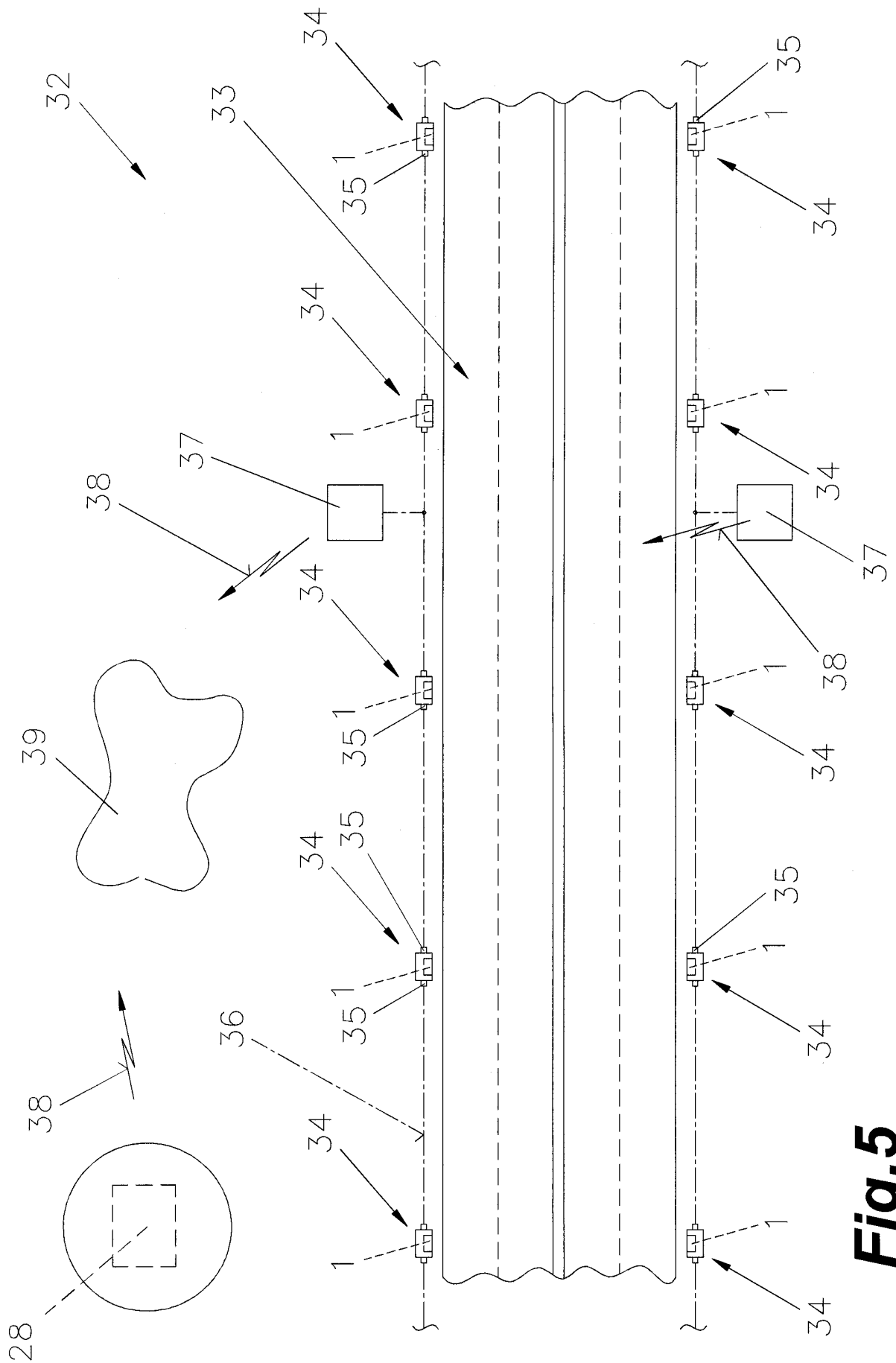
**Fig. 2**

00258

3/4



**Fig. 4**



**Fig.5**

Klassifikation des Anmeldungsgegenstands gemäß IPC:  
**G10L 21/0216** (2013.01)

Klassifikation des Anmeldungsgegenstands gemäß CPC:  
**G10L 21/0216** (2013.01); **G10L 2021/02166** (2013.01)

Recherchierter Prüfstoff (Klassifikation):  
G10L

Konsultierte Online-Datenbank:  
epodoc wpi ieee

Dieser Recherchenbericht wurde zu den am **15.03.2013** eingereichten Ansprüchen **1-39** erstellt.

| Kategorie <sup>*)</sup> | Bezeichnung der Veröffentlichung:<br>Ländercode, Veröffentlichungsnummer, Dokumentart (Anmelder),<br>Veröffentlichungsdatum, Textstelle oder Figur soweit erforderlich                                                                                                                           | Betreffend<br>Anspruch |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| X                       | EP 2562752 A1 (ASAHI CHEMICAL IND) 27. Februar 2013<br>(27.02.2013)<br>Fig 2 und Beschreibung; Absatz 0116                                                                                                                                                                                       | 1-39                   |
| X                       | Tachioka Y; Narita T; Iwasaki T: Direction of arrival<br>estimation by cross-power spectrum phase analysis using prior<br>distributions and voice activity detection information;<br>Acoustical Science and Technology 2012; 06.09.2011<br>Seite 68 - 71                                         | 1-39                   |
| X                       | Xiaoyu Ma; Kun Mu: Study on 2-dimensional microphone circular<br>array speech enhancement algorithm based on the DOA; 2011<br>International Conference on Electronic & Mechanical<br>Engineering and Information Technology (EMEIT 2011);<br>14.08.2011<br>Seite 2505 - 2508                     | 1-39                   |
| X                       | Bai M R; Pochang Ho: Using a steered-response power-phase<br>transform to optimize speech pickup in reverberant<br>environments; 2008 The Institution of Engineering and<br>Technology; 04.2008<br>Seite 280 - 291                                                                               | 1-39                   |
| X                       | Seon Man Kim; Hong Kook Kim; Sung Joo Lee; Yun Keun Lee:<br>Adaptation mode control with residual noise estimation for<br>beamformer-based multi-channel speech enhancement; 2012 IEEE<br>International Conference on Acoustics, Speech and Signal<br>Processing; 25.03.2012<br>Seiten 301 - 304 | 1-39                   |

Datum der Beendigung der Recherche:  
19.11.2013

Seite 1 von 1

Prüfer(in):  
SCHLECHTER Burkhard

<sup>\*)</sup> Kategorien der angeführten Dokumente:

- X** Veröffentlichung **von besonderer Bedeutung**: der Anmel-  
dungsgegenstand kann allein aufgrund dieser Druckschrift nicht als neu bzw. auf  
erfinderischer Tätigkeit beruhend betrachtet werden.
- Y** Veröffentlichung **von Bedeutung**: der Anmel-  
dungsgegenstand kann nicht  
als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die  
Veröffentlichung mit einer oder mehreren weiteren Veröffentlichungen  
dieser Kategorie in Verbindung gebracht wird und diese **Verbindung für  
einen Fachmann naheliegend** ist.

- A** Veröffentlichung, die den allgemeinen **Stand der Technik** definiert.
- P** Dokument, das von **Bedeutung** ist (Kategorien **X** oder **Y**), jedoch **nach  
dem Prioritätstag** der Anmeldung veröffentlicht wurde.
- E** Dokument, das **von besonderer Bedeutung** ist (Kategorie **X**), aus dem  
ein „**älteres Recht**“ hervorgehen könnte (früheres Anmeldedatum, jedoch  
nachveröffentlicht, Schutz ist in Österreich möglich, würde Neuheit in Frage  
stellen).
- &** Veröffentlichung, die Mitglied der selben **Patentfamilie** ist.