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

(11) **EP 4 134 954 B1**

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
02.08.2023 Patentblatt 2023/31

(21) Anmeldenummer: **21190351.3**

(22) Anmeldetag: **09.08.2021**

(51) Internationale Patentklassifikation (IPC):
H04R 27/00 *(2006.01)* ***G10L 21/0364*** *(2013.01)*
G10L 25/78 *(2013.01)*

(52) Gemeinsame Patentklassifikation (CPC):
H04R 27/00; G10L 21/0364; G10L 25/78;
H04R 2225/43; H04R 2227/007; H04R 2227/009;
H04R 2430/01

(54) **VERFAHREN UND VORRICHTUNG ZUR AUDIOSIGNALVERBESSERUNG**

METHOD AND DEVICE FOR IMPROVING AN AUDIO SIGNAL

PROCÉDÉ ET DISPOSITIF D'AMÉLIORATION DU SIGNAL AUDIO

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

(43) Veröffentlichungstag der Anmeldung:
15.02.2023 Patentblatt 2023/07

(73) Patentinhaber: **OPTImic GmbH**
20357 Hamburg (DE)

(72) Erfinder:
- **Vieweg, Markus**
  20357 Hamburg (DE)
- **Schäfer, Dr. Bernd Dominik**
  34119 Kassel (DE)

(74) Vertreter: **VKK Patentanwälte PartG mbB**
An der Alster 84
20099 Hamburg (DE)

(56) Entgegenhaltungen:
EP-A2- 1 777 988
US-A1- 2006 247 922
US-A1- 2010 121 634
US-A1- 2016 019 905
US-A1- 2017 047 080
US-B1- 6 295 364

- **SCHEPKER H ET AL: "Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression", SPEECH IN LIFE SCIENCES AND HUMAN SOCIETIES : 14TH ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION (INTERSPEECH 2013) ; LYON, FRANCE, 25 - 29 AUGUST 2013, CURRAN, RED HOOK, NY, 25. August 2013 (2013-08-25), Seiten 3577-3581, XP002734731, ISBN: 978-1-62993-443-3**

EP 4 134 954 B1

Beschreibung

[0001] Die vorliegende Erfindung betrifft ein Verfahren zur Verbesserung eines Audiosignals. Das Verfahren wird vorzugsweise in Echtzeit ausgeführt, sodass es sich für eine im Wesentlichen gleichzeitige Aufnahme und Wiedergabe von Audiosignalen eignet.

[0002] Audiosignale werden in der Praxis häufig unter ungünstigen akustischen Bedingungen mithilfe von Mikrofonen aufgezeichnet. Beispielsweise ist ein gewünschter Sprachsignalanteil während der Aufzeichnung von einem unerwünschten Störgeräusch überlagert, welches die Qualität des Audiosignals beeinträchtigt, insbesondere im Hinblick auf die Sprachverständlichkeit. Darüber hinaus kann das Audiosignal aufgrund der räumlichen Gegebenheiten oder in Folge eines großen Abstandes zwischen dem Sprecher und dem Mikrofon verhallt sein, sodass der Sprachanteil des Audiosignals bei einer gleichzeitigen Wiedergabe über Lautsprecher trotz einer Verstärkung schwer zu verstehen ist. Der eigentliche Vorteil einer akustischen Verstärkung des Audiosignals ist aus diesem Grunde häufig nicht ausreichend, um für eine befriedigende Sprachsignalqualität und Sprachverständlichkeit zu sorgen.

[0003] Zur Reduzierung der genannten Probleme ist es grundsätzlich möglich, das Audiosignal nach der Aufzeichnung mittels eines Audiofilters zu verarbeiten, um unerwünschte Signalanteile zu reduzieren. Dies ist jedoch mit Schwierigkeiten verbunden, weil das Audiofilter auf das jeweilige Audiosignal abgestimmt sein muss. In der Praxis bedeutet dies, dass ein Audiofilter für ein bestimmtes Audiosignal, welches in einer bestimmten akustischen Umgebung mit einem bestimmten Mikrofon aufgezeichnet worden ist, gute Ergebnisse erzielen kann, für ein anderes Audiosignal, welches unter anderen Bedingungen aufgezeichnet worden ist, jedoch nicht.

[0004] Die vorstehend genannten Probleme sind insbesondere im Bereich der z.B. für Messen eingesetzten mobilen Tontechnik relevant, weil diese mit unterschiedlichsten akustische Umgebungen kompatibel sein muss und in aller Regel wenig Zeit zur Verfügung steht, um die Audioverarbeitungsgeräte optimal einzustellen. Darüber hinaus besteht häufig überhaupt keine Möglichkeit, die Audiogeräte auf einen jeweiligen Sprecher zu optimieren, beispielsweise im Hinblick auf den geeigneten Abstand zwischen dem Sprecher und dem Mikrofon. Darüber hinaus bereiten Unterschiede zwischen verschiedenen Sprechern Probleme. Beispielsweise können unterschiedliche Sprecher, die insbesondere aufgrund von Alters- und Geschlechterunterschieden unterschiedliche Stimmeigenschaften aufweisen (z.B. unterschiedliche Sprecherlautstärke und Frequenzzusammensetzung), mit denselben Audiogeräten bei konstanter Konfiguration nicht in der Weise behandelt werden, dass zuverlässig eine hohe Sprachsignalqualität erzielt wird.

[0005] Zwar ist es möglich, mithilfe eines Mischpults das Audiosignal zu filtern und die akustischen Filterparameter während der Aufnahme manuell einzustellen. Dies ist jedoch aufwendig und erfordert besonders geschultes Personal. Zudem sind die auf diese Weise erzielbaren Verbesserungen variabel. Probleme bestehen insbesondere bei stark wechselnden akustischen Aufnahmesituationen, die nicht mit ausreichender Geschwindigkeit und Zuverlässigkeit kompensiert werden können.

[0006] Verfahren zur Verbesserung der Sprachsignalqualität sind aus den Dokumenten US 2016 0 019 905 A1, US 2017 0 047 080 A1, US 6 295 364 B1, US 2010 012 1 634 A1, US 2006 024 7 922 A1 sowie Schepker et al., Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression, Interspeech 2013 bekannt.

[0007] Es ist eine Aufgabe der Erfindung, ein Verfahren zur Verbesserung von Audiosignalen bereitzustellen, welches für unterschiedliche Audiosignale geeignet ist und insbesondere eine zuverlässige automatische Verbesserung des Audiosignals in Echtzeit ermöglicht. Ferner ist es eine Aufgabe der Erfindung, eine Vorrichtung zur Verbesserung von Audiosignalen bereitzustellen, welches zur automatischen Verbesserung von unterschiedlichen Audiosignalen insbesondere in Echtzeit geeignet ist.

[0008] Die Aufgabe wird gemäß einem ersten Aspekt gelöst durch ein Verfahren mit den Merkmalen des Anspruchs 1.

[0009] Ein erfindungsgemäßes Verfahren zur Verbesserung eines Audiosignals weist zumindest folgende Schritte auf: Empfangen eines Audiosignals mit mehreren Amplitudenwerten, wobei das Audiosignal zumindest abschnittsweise Sprache aufweist; Detektieren von Sprachabschnitten des Audiosignals; Filtern des Audiosignals mit wenigstens einem Pegelfilter, um Signalpegelvariationen des Audiosignals in den detektierten Sprachabschnitten zu reduzieren; und Filtern des Audiosignals mit wenigstens einem Entzerrfilter, um spektrale Variationen des Audiosignals in den detektierten Sprachabschnitten zu reduzieren.

[0010] Das Verfahren umfasst ferner folgende Schritte: Bestimmen einer Rückkopplungsfrequenz, welche eine Rückkopplung des Audiosignals repräsentiert; Filtern des Audiosignals mit einem Rückkopplungsfilter auf der Grundlage der bestimmten Rückkopplungsfrequenz, um Rückkopplungen repräsentierende Spektralanteile des Audiosignals zu reduzieren.

[0011] Das Filtern mit dem wenigstens einen Entzerrfilter umfasst einen Schritt des Bestimmens von Grobspektralwerten auf der Grundlage von Feinspektralwerten des Audiosignals, wobei die Grobspektralwerte die Feinspektralwerte mit einer geringeren Spektralauflösung als die Feinspektralwerte repräsentieren. Ferner werden erste Entzerrgewichte bestimmt, die eine Abweichung der Grobspektralwerte von vorbestimmten Referenzspektralwerten repräsentieren. Das Audiosignal wird außerdem mit den ersten Entzerrgewichten gewichtet, um die Spektralwerte in Übereinstimmung mit den

Referenzspektralwerten zu bringen.

[0012] Das Bestimmen der Rückkopplungsfrequenz umfasst folgende Schritte: Bestimmen einer Untermenge von Spektralwerten des Audiosignals, die einen vorbestimmten Spektralschwellenwert verletzen; Bestimmen von mehreren ersten Spektralparameterwerten auf der Grundlage der Untermenge, wobei jeder der ersten Spektralparameterwerte eine vorbestimmte Relation zwischen einem zugeordneten Spektralwert der Untermenge und wenigstens einem zeitlich und/oder spektral benachbarten Spektralwert repräsentiert; und Bestimmen der Rückkopplungsfrequenz auf der Grundlage der mehreren ersten Spektralparameterwerte.

[0013] Es hat sich gezeigt, dass die Qualität von Audiosignalen besonders unter einer unzureichenden Verständlichkeit der enthaltenen Sprachanteile leidet, also insbesondere jenen Abschnitten des Audiosignals, welche gesprochene Sprache aufweisen. Vor diesem Hintergrund werden erfindungsgemäß Zeitabschnitte des Audiosignals detektiert, welche Sprache aufweisen und als Sprachabschnitte bezeichnet werden können. Auf der Grundlage der detektierten Abschnitte wird das Audiosignal sodann mit einem Pegelfilter und einem Entzerrfilter verarbeitet, um bestimmte Variationen des Audiosignals zu reduzieren. Hierbei können Variationen sowohl innerhalb eines Audiosignals, als auch zwischen verschiedenen Audiosignalen behandelt werden.

[0014] Das Pegelfilter dient zur Reduktion von Signalpegelvariationen, um den Pegel des Audiosignals zu vereinheitlichen. Beispielsweise werden abschnittsweise sehr laute und leise Sprachsignalanteile abgeschwächt bzw. verstärkt, sodass sich insgesamt ein einheitlicher Signalpegel einstellt. Unterschiedliche Signalpegel ergeben sich in der Praxis z.B. durch variable Abstände zwischen einem Sprecher und dem aufzeichnenden Mikrofon sowie durch die akustischen Eigenschaften des umgebenden Raums. Die hieraus resultierenden Pegelvariationen werden durch das Pegelfilter jedoch kompensiert, sodass sich die subjektive Signalqualität verbessert.

[0015] Zusätzlich oder alternativ kommt ein Entzerrfilter zum Einsatz, um spektrale Variationen des Audiosignals zu reduzieren. Spektrale Variationen treten einerseits durch unterschiedliche Sprecher auf, die mit ihren Stimmen dem Audiosignal eine jeweils eigene Spektralcharakteristik aufprägen. Hinzu kommt eine spektrale Färbung durch die akustische Umgebung während der Aufnahme sowie gegebenenfalls durch die verwendeten Tongeräte, insbesondere das Mikrofon und dessen Ausrichtung relativ zum Sprecher.

[0016] Für eine hohe Sprachverständlichkeit ist es von Bedeutung, dass die Spektralanteile in bestimmten Frequenzbereichen, die für die Sprachverständlichkeit relevant sind, möglichst nicht oder nur in geringem Umfang durch andere spektrale Anteile maskiert werden. Häufig führen die akustischen Umgebungsbedingungen jedoch dazu, dass die sprachrelevanten Anteile in denselben oder in benachbarten Frequenzbereichen von anderen Signalanteilen variabel überlagert werden, sodass die sprachrelevanten Anteile nicht immer gleich gut wahrgenommen werden können. Derartige Veränderungen des Signals sind anhand der spektralen Variationen über die Zeit feststellbar und können daher durch ein geeignetes Filter behandelt werden. Vor diesem Hintergrund wird das Entzerrfilter dazu eingesetzt, spektrale Variationen in den detektierten Sprachabschnitten zu reduzieren. Auf diese Weise kann das Audiosignal in spektraler Hinsicht vereinheitlicht werden, um die Signalqualität insbesondere im Hinblick auf eine gute Sprachverständlichkeit zu erhöhen.

[0017] Durch das Verfahren kann insbesondere eine vollautomatische Signalverbesserung erfolgen. Eine vorherige oder betriebsbegleitende manuelle Einstellung oder Nachregelung von Filterparametern ist somit nicht notwendig, d.h. die Parameter des Pegel- und/oder Entzerrfilters können bei bestimmungsgemäßer Ausführung des Verfahrens fest eingestellt sein oder werden durch eine Recheneinheit automatisch eingestellt. Darüber hinaus gewährleistet das Verfahren eine hervorragende Signalverbesserung für unterschiedlichste Audiosignale, auch in besonders schwierigen akustischen Umgebungen. Mit anderen Worten ist das Verfahren besonders robust gegenüber akustischen Variationen jeglicher Art und ist somit für den professionellen Einsatz in der Praxis besonders geeignet. Darüber hinaus kann das Verfahre in Echtzeit, d.h. mit einer Latenz von weniger als 20 ms, bevorzugt von weniger als 10 ms, insbesondere 6 ms.

[0018] Besonders vorteilhaft ist es, wenn sowohl das Pegelfilter, als auch das Entzerrfilter verwendet werden. Darüber hinaus können noch zusätzliche Filter vorgesehen sein, um das Audiosignal weiter zu verbessern, wie im Folgenden erläutert wird.

[0019] Es versteht sich, dass die Filterung des Audiosignals nicht notwendig auf die detektierten Sprachabschnitte beschränkt werden muss. Beispielsweise kann ein Entzerrfilter im Hinblick auf besondere Spektralanteile, die etwa durch Rückkopplungen verursacht werden, zusätzlich auch außerhalb von Sprachabschnitten wirksam sein. Das Audiosignal wird jedoch zumindest in den Sprachabschnitten gefiltert, weil diese für die Sprachverständlichkeit besonders bedeutsam sind. Zur Verbesserung der Effizienz des Verfahrens können bestimmte Aspekte der Filterung auf die Sprachabschnitte beschränkt werden.

[0020] Ausführungsformen sind in der Beschreibung, den Figuren und den abhängigen Ansprüchen offenbart.

[0021] Gemäß einer Ausführungsform umfasst das Verfahren einen Schritt des Bestimmens von mehreren Spektralwerten auf der Grundlage der Amplitudenwerte, wobei die Amplitudenwerte das Audiosignal in einem Zeitbereich repräsentieren und wobei die Spektralwerte das Audiosignal in einem Frequenzbereich repräsentieren. Das Detektieren der Sprachabschnitte, das Filtern mit dem wenigstens einen Pegelfilter und/oder das Filtern mit dem wenigstens einen Entzerrfilter erfolgt auf der Grundlage der Amplitudenwerte und/oder der Spek-

tralwerte. Die Filterung erfolgt somit auf der Grundlage von zwei unterschiedlichen Repräsentationen des Audiosignals, nämlich Zeitbereichs- und Frequenzbereichswerten des Audiosignals. Die Effizienz und Zuverlässigkeit des Verfahrens wird auf diese Weise gesteigert.

[0022] Die Spektralwerte können mittels bekannter Frequenzraumtransformationen, wie beispielsweise der schnellen FourierTransformation (Fast Fourier Transformation, FFT) auf der Grundlage der Zeitbereichsamplitudenwerte ermittelt werden. Die Spektralwerte sind vorzugsweise durch den Betrag der Frequenzkoeffizienten (Spektralamplitudenwerte) gebildet, die durch FFT auf der Grundlage der Zeitbereichsamplitudenwerte besonders effizient ermittelt werden können. Der vorteilhafte Einsatz der Amplitudenwerte und der Spektralwerte erfordert somit vergleichsweise wenig Rechnerressourcen.

[0023] Gemäß einer weiteren Ausführungsform umfasst das Detektieren der Sprachabschnitte zumindest folgende Schritte: Bestimmen wenigstens eines ersten Energieparameterwerts auf der Grundlage der Amplitudenwerte, wobei der erste Energieparameterwert eine mittlere Energie eines Abschnitts des Sprachsignals repräsentiert; Bestimmen wenigstens eines zweiten Spektralparameterwerts auf der Grundlage von Spektralwerten des Audiosignals, wobei der wenigstens eine zweite Spektralparameterwert eine harmonische Spektralstruktur des Abschnitts repräsentiert; und Detektieren des Abschnitts als Sprachabschnitt, wenn der wenigstens eine erste Energieparameterwert einen ersten Energieparameterschwellenwert und/oder der wenigstens eine zweite Spektralparameterwert einen Spektralparameterschwellenwert verletzt. Die Detektion von Sprachabschnitten auf der Grundlage von Zeitbereichsund Spektralparametern hat sich als besonders nützlich erwiesen, um sowohl rauschartige Abschnitte (z.B. bei Konsonanten), als auch tonale Abschnitte (z.B. bei Vokalen) zuverlässig zu erfassen und durch Schwellenwertvergleich zur Unterscheidung von Sprachabschnitten und Rauschabschnitten auszuwerten.

[0024] Die genannten Schwellenwerte (Energieparameterschwellenwert und Spektralparameterschwellenwert) können grundsätzlich fest eingestellt sein. Die Zuverlässigkeit der Detektion von Sprachabschnitten kann jedoch in besonderer Weise verbessert werden, indem der erste Energieparameterschwellenwert und/oder der erste Spektralparameterschwellenwert in Abhängigkeit von der Zeit angepasst wird. Beispielsweise kann der Signalpegel des Audiosignals zur Einstellung der Schwellenwerte herangezogen werden, um sicherzustellen, dass die Schwellenwerte auf das jeweils aktuelle Energieniveau abgestimmt sind.

[0025] Nach einer weiteren Ausführungsform umfasst das Filtern des Audiosignals mit dem wenigstens einen Pegelfilter zumindest das Folgende: Bestimmen wenigstens eines Pegelparameterwerts auf der Grundlage der Amplitudenwerte, wobei der Pegelparameterwert einen mittleren Pegel des Audiosignals für einen detektierten Sprachabschnitt repräsentiert; Bestimmen von wenigstens einem Kompensationsgewicht auf der Grundlage des wenigstens einen Pegelparameterwerts; und Gewichten des Audiosignals mit dem wenigstens einen Kompensationsgewicht, um die Signalpegelvariationen des Audiosignals zu reduzieren.

[0026] Der wenigstens eine Pegelparameterwert kann allgemein mehrere Pegelparameterwerte umfassen, die den Pegel für detektierte Sprachabschnitte unterschiedlicher Länge angeben. Vorteilhaft können erste und zweite Pegelparameterwerte bestimmt werden, wobei die ersten Pegelparameterwerte den mittleren Pegel des Audiosignals mit einer ersten Zeitauflösung repräsentieren und wobei die zweiten Pegelparameterwerte den mittleren Pegel des Audiosignals mit einer zweiten Zeitauflösung repräsentieren. Die erste und zweite Zeitauflösung unterschieden sich voneinander. Auf diese Weise können kurzfristige und langfristige Effekte der auditorischen Wahrnehmung des Menschen vorteilhaft berücksichtigt werden. Insbesondere können kurzzeitige Pegelspitzen (Clipping) durch Pegelparameterwerte mit kurzer Zeitauflösung erfasst und zur Filterung herangezogen werden. Darüber hinaus können moderate Pegelvariationen, die erst ab einer Mindestdauer wahrnehmbar werden, durch Pegelparameterwerte mit größerer Zeitauflösung erfasst werden. Das Kompensationsgewicht für das Pegelfilter wird sodann auf der Grundlage der ersten und zweiten Pegelparameterwerte bestimmt.

[0027] Die ersten Pegelparameterwerte werden vorzugsweise auf der Grundlage von mehreren aufeinanderfolgenden Energiemittelwerten gebildet. Diese können geglättet werden, um erste Lautstärkewerte zu erhalten, die die ersten Pegelparameterwerte bilden. Die zweiten Pegelparameterwerte sind vorzugsweise durch zweite Lautstärkewerte gebildet. Diese können wiederum auf der Grundlage von mehreren aufeinanderfolgenden Energiemittelwerten gebildet werden, wobei abweichend von den ersten Lautstärkewerten eine größere Anzahl von Energiemittelwerten geglättet werden, sodass die zweiten Pegelparameterwerte den Pegel jeweils für eine größere Zeitdauer angeben als die ersten Pegelparameterwerte. Die zweite Zeitauflösung ist somit vorzugsweise größer als die erste Zeitauflösung.

[0028] Zur Bestimmung von Lautstärkewerten werden vorzugsweise zumindest einige der Amplitudenwerte in Zeitabschnitte des Audiosignals gruppiert. Sodann werden die Lautstärkewerte für zumindest einige der Zeitabschnitte auf der Grundlage der gruppierten Amplitudenwerte bestimmt, wobei jeder der Lautstärkewerte die Lautstärke eines der Zeitabschnitte des Audiosignals repräsentiert.

[0029] Die Begriffe "Energie" und "Pegel" repräsentieren jeweils eine Intensität oder Höhe der Amplitudenwerte. Pegelwerte können somit grundsätzlich als Energiewerte des Audiosignals angesehen werden und umgekehrt, wobei eine unterschiedliche Einheit für beide Werte möglich, jedoch nicht zwingend ist (z.B. kann für den Pegel im Gegensatz zur Energie die normierte log-

arithmische Einheit dB vorgesehen sein). Der Begriff "Pegel" stellt jedoch insbesondere einen funktionalen Bezug zum Pegelfilter her. Der Begriff "Lautstärke" repräsentiert die Intensität der Amplitudenwerte unter Berücksichtigung der auditorischen Wahrnehmbarkeit.

[0030] Gemäß einer weiteren vorteilhaften Ausführungsform werden erste Kompensationsgewichte und zweite Kompensationsgewichte bestimmt, wobei die ersten Kompensationsgewichte bestimmt werden, um Signalpegelvariationen mit wenigstens einem Pegel, der größer als ein vorbestimmter Pegelschwellenwert ist, zu reduzieren, wobei die zweiten Kompensationsgewichte bestimmt werden, um den Signalpegel des Audiosignals auf einen vorbestimmten Wert einzustellen. Auf diese Weise werden einerseits übermäßige Pegelwerte behandelt, die potentiell als qualitätsmindernde Verzerrung wahrnehmbar sind. Darüber hinaus wird das Audiosignal in den detektierten Sprachabschnitten auf einen Grundpegel eingestellt, sodass aus der Sicht des Hörers moderate Lautstärkeschwankungen kompensiert werden können. Vorzugsweise werden die ersten Kompensationsgewichte auf der Grundlage der ersten Pegelparameterwerte und die zweiten Kompensationsgewichte auf der Grundlage der zweiten Pegelparameterwerte ermittelt. Auf diese Weise kann die Filterung besonders gehörgerecht ausgeführt werden.

[0031] Wie oben erwähnt umfasst das Filtern mit dem wenigstens einen Entzerrfilter einen Schritt des Bestimmens von Grobspektralwerten auf der Grundlage von Feinspektralwerten des Audiosignals, wobei die Grobspektralwerte die Feinspektralwerte mit einer geringeren Spektralauflösung als die Feinspektralwerte repräsentieren. Ferner werden erste Entzerrgewichte bestimmt, die eine Abweichung der Grobspektralwerte von vorbestimmten Referenzspektralwerten repräsentieren. Das Audiosignal wird außerdem mit den ersten Entzerrgewichten gewichtet, um die Spektralwerte in Übereinstimmung mit den Referenzspektralwerten zu bringen. Die Feinspektralwerte sind vorzugsweise durch die oben genannten Spektralwerte gebildet, die insbesondere durch FFT effizient ermittelt werden können. Die Spektralauflösung dieser Spektralwerte ist bei einer Abtastrate von z.B. 48 kHz und einer Blocklänge von 1024 deutlich höher als die Auflösung, die durch das menschliche Gehör aufgelöst werden kann. Die Frequenzauflösung der Grobspektralwerte entspricht demgegenüber vorzugsweise der Auflösung des menschlichen Gehörs, sodass auf dieser Grundlage eine gehörgerechte Entzerrung ermöglicht wird. Die hierfür herangezogenen Referenzspektralwerte repräsentieren ein Referenzspektrum zur Erzielung einer hohen Sprachqualität von Audiosignalen. Die Grobspektralwerte können beispielsweise durch Oktavbandfilterung der Feinspektralwerte gewonnen werden.

[0032] Nach einer weiteren Ausführungsform umfasst das Filtern mit dem wenigstens einen Entzerrfilter ein Gewichten des Audiosignals mit zweiten Entzerrgewichten, wobei die zweiten Entzerrgewichte vorbestimmt sind. Es können somit zusätzlich oder alternativ zu den ersten Entzerrgewichten zweite Entzerrgewichte vorgesehen sein, die im Gegensatz zu den ersten Entzerrgewichten nicht dynamisch bestimmt werden, sondern im Vorfeld festgelegt sind. Durch die die zweiten Entzerrgewichte können beispielsweise Spektralanteile abgeschwächt werden, die für eine hohe Sprachqualität stets hinderlich sind und somit mit einem negativen Verstärkungsfaktor belegt werden können.

[0033] Nach einer weiteren Ausführungsform umfasst das Verfahren ein Filtern des Audiosignals mit wenigstens einem Kompressor, um einen Dynamikumfang des Audiosignals zu reduzieren. Für den wenigstens einen Kompressor können mehrere voneinander verschiedene Parametersätze vorgesehen sein, die in Abhängigkeit von einem Betrag des Audiosignals ausgewählt und der Filterung mit dem wenigstens einen Kompressor zugrunde gelegt werden. Vorteilhaft können sich die mehreren Parametersätze in einem Kompressionsgrad voneinander unterscheiden. Beispielsweise können die mehreren Parametersätze einen ersten Parametersatz umfassen, um den Dynamikumfang des Audiosignals mit einem ersten Kompressionsgrad zu reduzieren, wobei die mehreren Parametersätze einen zweiten Parametersatz umfassen, um den Dynamikumfang des Audiosignals mit einem zweiten Kompressionsgrad zu reduzieren, der stärker als der erste Kompressionsgrad ist. Für besonders gute Ergebnisse weisen die mehreren Parametersätze vorzugsweise einen dritten Parametersatz auf, um den Dynamikumfang des Audiosignals mit einem dritten Kompressionsgrad zu reduzieren, der geringer als der erste Kompressionsgrad ist. Auf diese Weise können qualitätsmindernde Verzerrungen, die durch eine starke Kompression hervorgerufen werden können, besonders effektiv vermieden werden. Der Kompressor kann als ein spezielles Pegelfilter angesehen werden, weil eine Reduktion des Dynamikumfangs mit einer Reduktion des Pegels und der Pegelvariationen einhergeht.

[0034] Wie oben erwähnt umfasst das Verfahren ferner folgende Schritte: Bestimmen einer Rückkopplungsfrequenz, welche eine Rückkopplung des Audiosignals repräsentiert; Filtern des Audiosignals mit einem Rückkopplungsfilter auf der Grundlage der bestimmten Rückkopplungsfrequenz, um Rückkopplungen repräsentierende Spektralanteile des Audiosignals zu reduzieren. Zum Bestimmen der Rückkopplungsfrequenz werden vorzugsweise die bereits vorliegenden Spektralwerte herangezogen, sodass diese hierfür nicht neu bestimmt werden müssen. Rückkopplungen entstehen, wenn wiedergegebene Signalanteile von dem Mikrofon nochmals aufgezeichnet und verstärkt werden, sodass sich ein instabiler Systemzustand einstellt, der akustisch durch eine starke Resonanz, z.B. durch Brummen oder einen schrillen Pfeifton, wahrnehmbar ist. Das Rückkopplungsfilter wirkt der Entstehung derartiger Kopplungseffekte entgegen, sodass die Signalqualität nicht beeinträchtigt wird. Das Rückkopplungsfilter kann als ein spezielles Entzerrfilter angesehen werden.

[0035] Wie weiter oben erwähnt umfasst das Bestim-

men der Rückkopplungsfrequenz umfasst vorzugsweise folgende Schritte: Bestimmen einer Untermenge von Spektralwerten des Audiosignals, die einen vorbestimmten Spektralschwellenwert verletzen; Bestimmen von mehreren ersten Spektralparameterwerten auf der Grundlage der Untermenge, wobei jeder der ersten Spektralparameterwerte eine vorbestimmte Relation zwischen einem zugeordneten Spektralwert der Untermenge und wenigstens einem zeitlich und/oder spektral benachbarten Spektralwert repräsentiert; und
Bestimmen der Rückkopplungsfrequenz auf der Grundlage der mehreren ersten Spektralparameterwerte. Der Rechenaufwand zur Bestimmung der Rückkopplungsfrequenz kann durch die schwellenwertbasierte Vorselektion von Spektralwerten stark reduziert werden, sodass die Echtzeitfähigkeit des Verfahrens begünstigt wird. Eine vorbestimmte Relation zwischen Spektralwerten kann insbesondere durch eine mathematische Verknüpfung der Spektralwerte gebildet werden, z.B. durch Verwendung von mathematischen Operatoren, wie Division oder Addition. Auf diese Weise können bestimmte Eigenschaften des Spektrums, die für eine Rückkopplungsfrequenz typisch sind, effizient erfasst werden.

[0036] Von besonderem Vorteil ist es, dass wenn die bestimmte Rückkopplungsfrequenz zwischen aufeinanderfolgenden Zeitabschnitten des Audiosignals verschwindet, die Wirksamkeit des Rückkopplungsfilters über mehrere Zeitabschnitte schrittweise reduziert wird. Auf diese Weise wird besonders zuverlässig gewährleistet, dass eine etwaige Rückkopplungsfrequenz wirksam aus dem Signal entfernt wird. Darüber hinaus werden eventuell wahrnehmbare Filterfluktuationen vermieden. Die schrittweise Reduktion des Rückkopplungsfilters erfolgt vorzugsweise nach dem Schema eines endlichen Automaten.

[0037] Nach einer weiteren Ausführungsform ist zur Filterung des Audiosignals ein Pausenfilter vorgesehen, um das Audiosignal in Bereichen außerhalb der detektierten Sprachabschnitte zu reduzieren. Hierdurch können z.B. zeitliche Maskierungseffekte durch Hintergrundstörgeräusche abgeschwächt werden.

[0038] Ferner kann das Audiosignal mit einem Rauschfilter gefiltert werden, um das Audiosignal in Bereichen mit Amplitudenwerten, die einen vorbestimmten Rauschschwellenwert verletzen, zu reduzieren. Insbesondere können auf diese Weise sehr kleine Amplitudenwerte, die unterhalb eines Schwellenwerts liegen und für eine gute Signalqualität irrelevant sind, im Wesentlichen vollständig entfernt werden. Das Entstehen von Rückkopplungen wird hierdurch entgegengewirkt. Vorzugsweise wird ein Noisegate-filter eingesetzt.

[0039] Nach einer weiteren Ausführungsform wird das Audiosignal mit einem Bandpassfilter gefiltert. Eine untere Grenzfrequenz des Bandpassfilters liegt vorzugsweise in einem Bereich von 50 bis 100 Hz. Eine obere Grenzfrequenz des Bandpassfilters liegt vorzugsweise in einem Bereich von 8000 bis 10000 Hz.

[0040] Die vorstehend beschriebenen Verfahrensaspekte können als Befehle in einem nicht-flüchtigen Speicher hinterlegt sein. Wenn die Befehle von einer Recheneinheit ausgeführt werden, wird die Recheneinheit durch die Befehle veranlasst, das beschriebene Verfahren gemäß einer Ausführungsform auszuführen. Allgemein kann das Verfahren somit teilweise oder vollständig durch einen Computer implementiert sein.

[0041] Die Aufgabe der Erfindung wird gemäß einem zweiten Aspekt gelöst durch eine Vorrichtung mit den Merkmalen des unabhängigen Vorrichtungsanspruchs.

[0042] Eine erfindungsgemäße Vorrichtung zur Verbesserung eines Audiosignals, welches Sprache aufweist, umfasst einerseits wenigstens eine Eingangsschnittstelle zum Erfassen eines Audiosignals. Die Eingangsschnittstelle weist einen Anschluss für ein Mikrofon auf, um das Audiosignal zu erfassen. Andererseits ist wenigstens eine Ausgangsschnittstelle zum Ausgeben des Audiosignals vorgesehen. Die Ausgangsschnittstelle weist einen Anschluss für ein Audiowiedergabegerät, z.B. eine Beschallungsanlage mit ein oder mehreren Schallwandlern auf. Die Vorrichtung weist außerdem eine Recheneinheit zum Ausführen eines Verfahrens zur Verbesserung des Audiosignals auf. Das Verfahren ist nach einem der vorhergehenden Ausführungsformen ausgebildet.

[0043] Die Vorrichtung ist vorzugsweise als ein kompaktes Audiogerät ausgebildet, sodass es sich insbesondere auch für den mobilen Einsatz besonders eignet.

[0044] Die Vorrichtung weist vorzugsweise einen nicht-flüchtigen Speicher auf, in dem Befehle zur Ausführung des Verfahrens hinterlegt sind. Der Speicher ist hierzu mit der Recheneinheit koppelbar.

[0045] Die Recheneinheit umfasst vorzugsweise einen Analog-zu-Digital-Umsetzer sowie einen Digital-zu-Analog-Umsetzer. Die Verbesserung des Audiosignals kann somit zumindest teilweise auf der Grundlage einer digitalen Version des Audiosignals erfolgen. Das Verfahren kann somit einerseits besonders effizient durchgeführt werden. Andererseits kann eine hohe Filterungsqualität gewährleistet werden.

[0046] Die Eingangs- oder Ausgangsschnittstelle kann jeweils als drahtgebundene Schnittstelle ausgeführt sein, um eine Kompatibilität mit anderen professionellen Tongeräten zu gewährleisten und Übertragungsverluste zu minimieren. Es ist jedoch auch denkbar, die Schnittstellen jeweils drahtlos auszubilden, wobei die Schnittstellen hierfür auch zu einer gemeinsamen Drahtlosschnittstelle zusammengefasst sein können.

[0047] Weitere Ausführungsformen der Vorrichtung sind in den abhängigen Ansprüchen, der nachfolgenden Beschreibung sowie den Zeichnungen beschrieben. Es versteht sich jedoch, dass auch beschriebene Verfahrensmerkmale in entsprechender Weise in der Vorrichtung verwirklicht sein können, insbesondere durch entsprechende Konfiguration der Recheneinheit. Umgekehrt können auch hier beschriebene Vorrichtungsmerkmale hinsichtlich ihrer Funktion als Verfahrensmerkmale einen Teil des Verfahrens bilden.

[0048] Gemäß einer Ausführungsform umfasst die Vorrichtung ferner einen Vorverstärker für das Audiosignal, der mit der Eingangsschnittstelle koppelbar ist. Auf diese Weise kann das Audiosignal vorteilhaft vor einer Abtastung auf einen vorbestimmten Pegelbereich verstärkt werden. Für den Vorverstärker können mehrere vorbestimmte Verstärkungswerte vorgesehen sein, wobei einer der Verstärkungswerte vorzugsweise automatisch oder durch einen Bediener der Vorrichtung ausgewählt und der Verstärkung zugrunde gelegt wird.

[0049] Die Vorrichtung verfügt vorzugsweise über eine elektrische Versorgung für die Eingangsschnittstelle. Somit wird im Sinne einer sogenannten Phantomspeisung eine elektrische Versorgung eines angeschlossenen Schallwandlers, z.B. eines Mikrofons, über die Eingangsschnittstelle ermöglicht.

[0050] Gemäß einer weiteren Ausführungsform weist die Vorrichtung ferner eine Schalteinrichtung auf, die mit der Eingangsschnittstelle, der Ausgangsschnittstelle und/oder der Recheneinheit koppelbar ist, um die Eingangsschnittstelle wahlweise über die Recheneinheit mit der Ausgangsschnittstelle zu verbinden. Mit anderen Worten kann die Recheneinheit überbrückt werden. Auf diese Weise kann eine Ausgabe des Audiosignals auch im Falle einer Fehlfunktion der Recheneinheit gewährleistet werden.

[0051] Um eine zuverlässige Funktion der Vorrichtung auch im Dauerbetrieb zu ermöglichen, ist die Vorrichtung vorzugsweise mit einer Kühleinrichtung versehen. Sämtliche Komponenten der Vorrichtung einschließlich der Recheneinheit können somit in einem kompakten Gehäuse aufgenommen sein, wobei z.B. Abwärme der Recheneinheit dennoch wirksam durch die Kühleinrichtung abgeführt werden kann, um die Funktion der Recheneinheit nicht zu beeinträchtigen und die Lebensdauer aller Komponenten nicht zu verkürzen.

[0052] Die Recheneinheit kann vorteilhaft einen Einplatinenrechner aufweisen, sodass die Vorrichtung insgesamt besonders kompakt ausgebildet werden kann. Die Vorrichtung kann außerdem ein Gehäuse aufweisen, in das insbesondere alle elektrischen Komponenten der Vorrichtung aufgenommen sein können, um auf diese Weise vor äußeren Einflüssen geschützt zu werden. Die Recheneinheit kann einen oder mehrere Prozessoren sowie einen Speicher aufweisen, in dem Befehle zur Ausführung des Verfahrens gespeichert werden können.

[0053] Zur Konfiguration der Vorrichtung weist die Vorrichtung vorzugsweise wenigstens eine externe Kommunikationsschnittstelle auf. Beispielsweise kann die Vorrichtung mit einer Netzwerkschnittstelle, z.B. einer EthernetSchnittstelle, oder einer Bus-Schnittstelle ausgestattet sein, um über ein Netzwerk oder direkt mit einem Benutzerendgerät, beispielsweise einem PC oder einem mobilen Endgerät, wie etwa einem Laptop verbunden zu werden. Es kann auch eine Anbindung an drahtlose Endgeräte über das Internet erfolgen, um eine Anbindung an einen zentralen Server (Cloud) zu ermöglichen. Die Steuerungsschnittstelle kann auch als Drahtlosschnittstelle ausgebildet sein, sodass die Vorrichtung unmittelbar mit einem mobilen Endgerät verbunden werden kann (z.B. über Bluetooth oder ein lokales Drahtlosnetzwerk). Die Kommunikation mit der Vorrichtung, z.B. zum Zwecke der Konfiguration, kann somit besonders komfortabel erfolgen. Über die Kommunikationsschnittstelle können insbesondere Steuerungsdaten, z.B. Filterparameter zur Ausführung des beschriebenen Verfahrens zur Verbesserung eines Audiosignals eingestellt werden. Dies kann insbesondere aus der Ferne erfolgen, sodass eine Konfiguration durch den Endnutzer der Vorrichtung vollständig vermieden werden kann. Zusätzlich oder alternativ kann die Kommunikationsschnittstelle zur Übertragung des Audiosignals an ein mobiles Endgerät oder einen zentralen Server ausgebildet sein. Auf diese Weise kann das Audiosignal z.B. zu Dokumentationszwecken in dem Endgerät oder in einer Cloud gespeichert werden. Zur Übertragung an einen zentralen Server ist die Kommunikationsschnittstelle vorzugsweise als Ethernet-Schnittstelle ausgebildet, die auch eine Übertragung von Audiosignalen ermöglicht (z.B. unter Verwendung von Dante, Milan, AES (Advanced Encryption Standard).

[0054] Außerdem kann über eine Kommunikationsschnittstelle der Vorrichtung eine Firmware der Vorrichtung aktualisiert werden. Vorzugsweise ist eine Kommunikationsschnittstelle in Form einer separaten Busschnittstelle vorgesehen, die insbesondere zum Anschließen eines Speichermediums, z.B. eines Massenspeichers in Form eines USB-Sticks oder dergleichen dient. Auf dem Speichermedium können einerseits Konfigurationsund/oder Aktualisierungsdaten gespeichert sein, die an die Vorrichtung übertragen werden, um die lokal gespeicherten Daten zu aktualisieren. Darüber hinaus kann das Audiosignal zu Aufnahmezwecken an das Speichermedium ausgegeben und in dem Speichermedium gespeichert werden. Hierzu ist die Vorrichtung vorzugsweise mit einer Bedienschnittstelle ausgestattet, um die Aufnahme des Audiosignals unmittelbar an der Vorrichtung steuern zu können.

[0055] Ein nicht unter die Ansprüche fallender Aspekt der Offenbarung bezieht sich auf ein Verfahren zur selektiven Verbesserung eines ersten Audiosignals unter Verwendung eines Audioverarbeitungsmittels, wobei das erste Audiosignal zumindest abschnittsweise Sprache aufweist und das Verfahren zumindest folgende Schritte umfasst: Feststellen, ob das Audioverarbeitungsmittel einen vorbestimmten Tauglichkeitszustand aufweist; Wenn das Audioverarbeitungsmittel den vorbestimmten Tauglichkeitszustand aufweist, Ausführen eines Verfahrens zur Verbesserung des ersten Audiosignals unter Verwendung des Audioverarbeitungsmittels, um ein zweites Audiosignal bereitzustellen; Wenn das Audioverarbeitungsmittel den vorbestimmten Tauglichkeitszustand nicht aufweist, Bereitstellen des ersten Audiosignals. Das Verfahren ermöglicht somit eine selektive Verwendung des Audioverarbeitungsmittels in Abhängigkeit von seinem Tauglichkeitszustand. Fehlfunktionen des Audioverarbeitungsmittels führen somit nicht

dazu, dass kein Audiosignal mehr ausgegeben wird und die Nutzerzufriedenheit beeinträchtigt wird. Im Falle einer Fehlfunktion wird zumindest das erste Audiosignal bereitgestellt, sodass z.B. für Beschallungsanlagen ein brauchbares Audiosignal zur Verfügung steht und auf diese Weise eine Basisfunktionalität erhalten bleibt. Das Verfahren kann insbesondere durch eine Schalteinrichtung verwirklicht werden, welche in einer Vorrichtung z.B. als schaltbares Relais ausgeführt sein kann. Alternativ kann die Schaltfunktionalität auch durch die Recheneinheit selbst verwirklicht werden. Eine separate Schalteinrichtung besitzt jedoch den Vorteil eines Schutzes gegenüber einem vollständigen Ausfall der Recheneinheit, in dem keinerlei Durchleitung des Signals erfolgen kann.

[0056] Die hierin offenbarten Verfahren sind vorzugsweise mit der beschriebenen Vorrichtung ausführbar. Es ist jedoch auch möglich, die Verfahren ganz oder teilweise auf einem beliebigen Computer, insbesondere einem zentralen Server auszuführen. Beispielsweise kann das Audiosignal lokal erfasst und auf einen Server übertragen werden, wo die Signalverbesserung ausgeführt wird. Sodann kann das verbesserte Signal an einen lokalen Empfänger übermittelt werden, um es mit einem Schallwandler wiederzugeben.

[0057] Die Erfindung wird nachfolgend rein beispielhaft anhand der Zeichnungen weiter erläutert. Die Zeichnungen zeigen im Einzelnen:

Fig. 1 ein Blockdiagramm zur Illustration eines Verfahrens zur Verbesserung eines Audiosignals;

Fig. 2 ein Blockdiagramm zur Illustration eines Verfahrens zum Detektieren von Sprachabschnitten eines Audiosignals;

Fig. 3 Frequenzgänge von Oktavfiltern zur Bestimmung von Grobspektralwerten für ein Entzerrfilter für das Verfahren nach Fig. 1;

Fig. 4 ein Blockdiagramm zur Illustration eines Verfahrens zum Bestimmen einer Rückkopplungsfrequenz;

Fig. 5 ein Blockdiagramm zur Illustration eines schrittweisen Reduzierens eines Rückkopplungsfilters;

Fig. 6 eine schematische Darstellung eines Geräts zur Audiosignalverbesserung;

Fig. 7 eine Anordnung mit dem Gerät von Fig. 7.

[0058] In den Figuren sind gleiche oder sich entsprechende Elemente mit denselben Bezugszeichen gekennzeichnet.

[0059] Ein Verfahren zur Verbesserung eines Audiosignals wird nachfolgend mit Bezug auf Fig. 1 beschrieben.

[0060] Ein analoges Audiosignal wird mit einem nicht gezeigten Mikrofon erfasst (Schritt 10), wobei das Audiosignal mehrere Sprachabschnitte sowie mehrere Rauschabschnitte aufweist. Die Sprachabschnitte weisen Sprache auf und bilden einen Sprachsignalanteil. Die Rauschabschnitte sind durch alle übrigen Abschnitte gebildet, die keine Sprache aufweisen, insbesondere in Sprechpausen.

[0061] In Schritt 12 wird das Audiosignal vorverstärkt, d.h. als analoges Signal mit einem Verstärkungsfaktor elektronisch verstärkt. Für einen entsprechenden Vorverstärker (in Fig. 1 nicht gezeigt) kann eine feste Verstärkung eingestellt sein. Alternativ kann durch einen Benutzer einer von mehreren voreingestellten Verstärkungswerten in Abhängigkeit eines aufnahmebedingten Grundpegels ausgewählt werden, um ein nachfolgendes Pegelfilter zur Reduktion von Pegelvariationen zu entlasten.

[0062] Das vorverstärkte Audiosignal wird in Schritt 14 von einem analogen Signal zu einem digitalen Signal umgewandelt. Dies erfolgt vorzugsweise mittels eines Analog-zu-Digital-Umsetzer, welcher das Analogsignal mit einer vorbestimmten Abtastrate, z.B. 48.000 Hz abtastet. Der Schritt 14 kann alternativ auch nach dem Schritt 16 erfolgen, der im Folgenden erläutert wird.

[0063] Das Audiosignal wird in Schritt 16 mit einem Pegelfilter verarbeitet, um Variationen des Signalpegels auszugleichen. Das Pegelfilter wird hierzu in Abhängigkeit von ersten Filterdaten 44 betrieben, die auf der Grundlage des Audiosignals am Ausgang des Pegelfilters in Schritt 18 ermittelt werden. Sie umfassen erste Lautstärkewerte, detektierte Sprachabschnitte sowie detektierte Pegelspitzen. Pegelspitzen sind detektierte Signalpegel, die größer als ein vorbestimmter Pegelschwellenwert sind, in dem das Signal übersteuert (Clipping).

[0064] Die Lautstärkewerte werden für einzelne Blöcke des Audiosignals ermittelt, die vorzugsweise jeweils eine Länge von 64 Abtastwerten aufweisen. Für jeden Block wird ein erster Lautstärkewert ermittelt, indem die quadrierten Abtastwerte des Blocks aufsummiert werden und sodann die Quadratwurzel der Summe ermittelt wird. Es werden auf diese Weise sogenannte RMS-Werte (Root-Mean-Square) gebildet, die jeweils eine mittlere Energie des zugrundeliegenden Blocks von Abtastwerten repräsentieren.

[0065] Vorzugsweise werden für das Pegelfilter die RMS-Werte von mehreren Blöcken herangezogen. Zur Detektion von Pegelspitzen werden hierzu die RMS-Werte des aktuellen Blocks sowie des vorhergehenden Blocks gemeinsam ausgewertet, wobei eine Pegelspitze detektiert wird, wenn mindestens einer der beiden RMS-Werte einen vorbestimmten Schwellenwert überschreitet, zum Beispiel - 3 dB. Im Falle einer detektierten Pegelspitze wird diese Information als Teil der ersten Filterdaten 44 in Schritt 16 berücksichtigt. In Ansprechen auf eine detektierte Pegelspitze wird die Verstärkung des Pegelfilters in Schritt 16 stark und schnell vermindert,

zum Beispiel mit einer Rate von - 3 dB innerhalb von 200 ms. Auf diese Weise werden Pegelspitzen effektiv entfernt. Vorzugsweise werden Pegelspitzen unabhängig davon gefiltert, ob der betreffende Abschnitt des Audiosignals ein Sprachabschnitt ist oder nicht.

[0066] Das Pegelfilter von Schritt 16 ist ferner so konfiguriert, dass der Pegel des Audiosignals auf einen vorbestimmten Wert eingestellt wird. Hierzu werden die RMS-Werte des aktuellen Blocks sowie einer Vielzahl von mehreren vorhergehenden Blöcken, zum Beispiel 30 vorhergehenden Blöcken, herangezogen. Die RMS-Werte werden über die betrachteten Blöcke geglättet, sodass kurzzeitige Schwankungen entfernt werden, die für die menschliche Wahrnehmung (mit Ausnahme der separat behandelten Pegelspitzen) irrelevant sind. Vorzugsweise wird zur Glättung der Median der betrachteten RMS-Werte gebildet, um zweite Lautstärkewerte zu erhalten, die den aktuellen Signalpegel gehörgerecht angeben. Sodann wird ein Kompensationsgewicht bestimmt, der die Differenz zwischen einem vorbestimmten Referenzwert und dem aktuellen zweiten Lautstärkewert repräsentiert. Beispielsweise kann der aktuelle Lautstärkewert von einer Referenzlautstärke von - 20 dB subtrahiert werden, um ein Kompensationsgewicht zu bilden. Das Kompensationsgewicht wird sodann mit dem Audiosignal gewichtet, z.B. multipliziert, um die Lautstärke mit der Referenzlautstärke in Übereinstimmung zu bringen.

[0067] Vorzugsweise wird die maximale zeitliche Änderung des Kompensationsgewichts begrenzt, zum Beispiel auf 5 dB pro Sekunde. Auf diese Weise werden unnatürliche Fluktuationen in der Lautstärke des Audiosignals vermieden.

[0068] Darüber hinaus wird die Einstellung des Signalpegels mit Bezug auf die Referenzlautstärke vorzugsweise nur in solchen Abschnitten des Audiosignals durchgeführt, die als Sprachabschnitte detektiert worden sind. Die Information, welche Abschnitte als Sprachabschnitte detektiert worden sind, wird als Teil der Filterdaten 44 dem Pegelfilter von Schritt 16 bekanntgemacht.

[0069] Die Detektion von Sprachabschnitten erfolgt in Schritt 18 und wird im Folgenden anhand von Fig. 2 erläutert.

[0070] Die Detektion von Sprachabschnitten erfolgt auf der Grundlage von Amplitudenwerten 54 und Spektralwerten 56, wobei die Amplitudenwerte 54 das Audiosignal im Zeitbereich und die Spektralwerte 56 das Audiosignal im Frequenzbereich repräsentieren. Die Amplitudenwerte 54 sind durch die Abtastwerte des digitalen Audiosignals nach Schritt 14 gebildet. Die Spektralwerte 56 werden blockweise durch schnelle Fouriertransformationen (FTP) auf der Grundlage der Amplitudenwerte 54 ermittelt. Es können grundsätzlich jedoch auch andere Frequenztransformationen eingesetzt werden. Die Blocklänge zur Ermittlung der Spektralwerte 56 beträgt vorzugsweise 1024 Amplitudenwerte (Abtastwerte), wobei sich benachbarte Blöcke vorzugsweise um die Hälfte überlappen und die betreffenden Amplitudenwerte jedes Blocks vor der Transformation mit einem Hann-Fenster gewichtet werden, um unerwünschte Spektralanteile, die durch die Blockgrenzen verursacht werden, zu reduzieren. Ferner werden die Spektralwerte 56 mit einem vorbestimmten Faktor gewichtet, sodass die Spektralwerte 56 auf einen Bereich zwischen 0 und 1 normalisiert werden. Der Faktor hängt insbesondere von dem verwendeten Fenster ab. Im Fall des bevorzugten Hann-Fensters kann vorteilhaft ein Faktor von 0,00391 verwendet werden.

[0071] In Schritt 58 von Fig. 2 werden drei Parameter bestimmt und jeweils daraufhin geprüft, ob ein zugeordnetes Schwellenwertkriterium verletzt wird. Ein erster Parameterwert wird durch den oben beschriebenen RMS-Wert auf der Grundlage der Amplitudenwerte 54 gebildet. Der erste Parameterwert kann auch als Kurzzeitenergie (STE = Short Time Energy) bezeichnet werden, weil er die mittlere Energie über einen Block mit einer relativ kurzen Länge von 64 Amplitudenwerten repräsentiert. Sofern der erste Parameterwert einen zugeordneten Schwellenwert überschreitet (Schritt 62), zeigt der erste Parameterwert einen Sprachabschnitt an, andernfalls einen Rauschabschnitt (kein Sprachabschnitt). Hohe RMS-Werte können insbesondere durch Konsonanten hervorgerufen werden und deuten somit auf Sprache hin.

[0072] Ein zweiter Parameterwert wird auf der Grundlage der Spektralwerte 56 ermittelt und gibt die Ausprägung einer harmonischen Obertonstruktur des Frequenzspektrums an. Insbesondere stellt der zweite Parameterwert ein Maß für die spektrale Flachheit des Frequenzspektrums dar, das durch die Spektralwerte 56 repräsentiert wird (Spectral Flatness, SF). Der zweite Parameterwert wird vorzugsweise durch Division des geometrischen Mittelwerts der Spektralwerte 56 und des arithmetischen Mittelwerts der Spektralwerte 56 bestimmt. Der zweite Parameterwert wird sodann mit einem zugeordneten Schwellenwert verglichen (Schritt 62). Wenn der Schwellenwert unterschritten wird, zeigt der zweite Parameterwert einen Sprachabschnitt an, andernfalls einen Rauschabschnitt. Hohe Werte des zweiten Parameters deuten auf rauschartigen Blöcke hin, die untypisch für Sprache sind. Im Gegensatz zu dem ersten Parameter bezieht sich der zweite Parameter aufgrund der Spektralwerte auf eine deutlich längere Blocklänge von 1024, sodass die üblicherweise deutlich kürzeren Konsonanten gegenüber einer ansonsten tonalen Charakteristik nicht ins Gewicht fallen.

[0073] Außerdem wird ein dritter Parameterwert bestimmt, der angibt, ob ein Maximum der Spektralwerte 56 in einem vorbestimmten Frequenzbereich liegt. Hierzu wird vorzugsweise ermittelt, ob der Spektralwert, dessen Betrag ein Maximum gegenüber den übrigen Spektralwerten 56 eines Blocks bildet (Schritt 58), in einem Frequenzbereich zwischen 70 und 250 Hz liegt, d.h. es wird geprüft, ob der maximale Spektralwert eine Frequenz repräsentiert, die größer als ein unterer Frequenzschwellenwert und kleiner als ein oberer Frequenz-

schwellenwert ist (Schritt 62). Zutreffendenfalls zeigt der dritte Parameterwert einen Sprachabschnitt an, andernfalls einen Rauschabschnitt. Die Grundfrequenz von Sprache liegt in der Regel im Bereich zwischen 70 und 250 Hz, sodass ein Maximum der Spektralwerte 56 in diesem Bereich auf Sprache hinweist.

[0074] Für die ersten und zweiten Parameterwerte sind vorzugsweise adaptive Schwellenwerte vorgesehen, um variable Distanzen zwischen einem jeweiligen Sprecher und dem aufzeichnenden Mikrofon zu kompensieren. Der Schwellenwert wird für einen betreffenden Block adaptiv auf der Grundlage der Parameterwerte von mehreren vorhergehenden Blöcken bestimmt (Schritt 60), wobei die vorhergehenden Blöcke vorzugsweise detektierte Sprachabschnitte und Rauschabschnitte umfassen. Beispielsweise werden zur Bestimmung des Schwellenwerts für den ersten Parameterwert die ersten Parameterwerte von 30 vorhergehenden als Sprachabschnitt klassifizierten Blöcken und die ersten Parameterwerte von dreißig vorhergehenden als Rauschabschnitt klassifizierten Blöcken herangezogen. Die ersten Parameterwerte werden für jeden Abschnittstyp aufsummiert und die erhaltenen Summen voneinander subtrahiert. Das Ergebnis wird mit einem Gewichtungsfaktor gewichtet, um den zugeordneten Schwellenwert für den ersten Parameterwert des aktuellen Blocks zu erhalten. Auf diese Weise wird gewährleistet, dass der Schwellenwert an das aktuelle Betragsniveau des ersten Parameterwerts angepasst wird, um Falschklassifikationen zu vermeiden. Der Gewichtungsfaktor wird vorzugsweise zwischen 0 und 1 eingestellt und steuert die Empfindlichkeit der Detektion.

[0075] Nach dem Prinzip des Schwellenwerts für den ersten Parameter wird vorzugsweise auch der Schwellenwert für den zweiten Parameter ermittelt. Hierbei wird die Berechnungsvorschrift jedoch invertiert, da der zweite Parameter mit abnehmendem Betrag Sprache indiziert und somit im Vergleich zum ersten Parameter umgekehrt mit Sprache korreliert ist. Folglich wird die Summe der zweiten Parameterwerte für Sprachabschnitte von der Summe der zweiten Parameterwerte für Rauschabschnitte subtrahiert und mit einem Gewichtungsfaktor beaufschlagt, der vorzugsweise zwischen 0 und 1 liegt und die Empfindlichkeit der Detektion steuert.

[0076] In Schritt 64 werden die drei Parameter gemeinsam ausgewertet und festgestellt, ob die Parameterwerte jeweils das zugeordnete Schwellenwertkriterium verletzen oder nicht. Wenn zwei der drei Parameterwerte einen Sprachabschnitt anzeigen, d.h. dass jeweils zugeordnete Schwellenwertkriterium verletzen, wird der betreffende Block vorläufig als Sprachabschnitt detektiert.

[0077] Um stark fluktuierende Detektionsergebnisse zu vermeiden, insbesondere nicht plausible alternierende Wechsel zwischen Sprachabschnitten und Rauschabschnitten, wird ein Wechsel zwischen einem Sprachabschnitt und einem Rauschabschnitt und umgekehrt nur dann zugelassen, wenn eine vorbestimmte Anzahl von aufeinanderfolgenden Blöcken als Sprachabschnitt oder Rauschabschnitt klassifiziert worden sind (Schritt 66 und 68). Beispielsweise müssen nach einem als Rauschabschnitt detektierten Block fünf unmittelbar aufeinanderfolgende Blöcke vorläufig als Sprachabschnitt detektiert werden, um diese Blöcke final als Sprachabschnitt zu detektieren (Schritt 70). Andernfalls werden die Blöcke weiterhin als Rauschabschnitte detektiert (Schritt 72). Umgekehrt müssen nach einem als Sprachabschnitt detektierten Block z.B. acht unmittelbar aufeinanderfolgende Blöcke vorläufig als Rauschabschnitt detektiert werden, um diese Blöcke final als Rauschabschnitte zu detektieren (Schritt 72). Andernfalls werden die Blöcke weiterhin als Sprachabschnitte detektiert (Schritt 70).

[0078] Im Folgenden werden weitere Schritte des Verfahrens von Fig. 1 erläutert. In Schritt 20 wird das Audiosignal mit einem festen Verstärkungsfaktor gewichtet, um Pegelverluste durch nachfolgende Filter vorab zu kompensieren. Beispielsweise kann das Signal um 3 bis 6 dB verstärkt werden.

[0079] In Schritt 22 wird das Audiosignal mit einem Rauschfilter gefiltert, welches dazu angepasst ist, sehr leise Abschnitte des Audiosignals zu reduzieren. Hierbei wird davon ausgegangen, dass sehr leise Signalabschnitte keine relevante Information beinhalten und die empfundene Sprachqualität insoweit allenfalls negativ beeinträchtigen können. Insbesondere wird durch eine Reduktion des Signalpegels in sehr leisen Signalabschnitten das Risiko von Rückkopplungen reduziert. Als Rauschfilter kann insbesondere ein sogenanntes Noise-Gate verwendet werden, welches dazu angepasst ist, leise Signalabschnitte zu unterdrücken. Als Kriterium zur Erkennung von leisen Signalabschnitten wird ein Schwellenwert zugrunde gelegt, der mit dem aktuellen Signalpegel verglichen wird. Sofern der aktuelle Signalpegel den Schwellenwert unterschreitet, wird das Rauschfilter aktiviert. Der Schwellenwert liegt vorzugsweise deutlich unterhalb der in Schritt 16 eingestellten Referenzlautstärke. Beispielsweise kann der Schwellenwert bei -55 dB liegen. Bei Unterschreiten des Schwellenwerts wird das Audiosignal mit einem Ratio im Bereich von 5 bis 10 abgesenkt. Als Anstiegszeit (attack time) und Ausklingzeit (release time) werden vorzugsweise Werte im Bereich von 10 ms bzw. 100 ms verwendet.

[0080] In Schritt 24 werden zweite Filterparameter 46 bestimmt, welche für die nachfolgenden Schritte 32, 34 und 36 herangezogen werden. Die zweiten Filterparameter 46 umfassen einerseits die bereits in Schritt 18 detektierten Sprachabschnitte 52. Außerdem werden Oktavspektralwerte 48 bestimmt, die im Vergleich zu den Spektralwerten 56 eine gröbere Spektralauflösung aufweisen, die der auditorischen Wahrnehmung des Menschen nachgebildet ist. Hierzu werden die z.B. mittels FFT bestimmten Spektralwerte 56 mit einer Oktavfilterbank gefiltert. Die Oktavfilterbank umfasst insgesamt acht sich im Spektralbereich überlappende Filter, die in Fig. 3 beispielhaft durch Betragsfrequenzgänge 37 über die Frequenz F und den Betrag G dargestellt sind. Die

Frequenzgänge 37 weisen ihr jeweiliges Maximum bei einer filtereigenen Mittenfrequenz fc auf und fallen zu kleineren und größeren Frequenzwerten hin ab. Die Mittenfrequenzen fc betragen vorzugsweise 63, 125, 250, 500, 1000, 2000, 4000 und 8000 Hz. Die Grenzfrequenzen (Betragsfrequenzgang von - 3 dB) können auf der Grundlage der jeweiligen Mittenfrequenz fc generisch berechnet werden. Die untere Grenzfrequenz beträgt 32fc/45 und die obere Grenzfrequenz beträgt 45fc/32. Zur Filterung werden die in ein jeweiliges Filter fallenden Spektralwerte gewichtet aufsummiert, wobei die Gewichte jeweils den Betragsfrequenzgang bei der Frequenz des betreffenden Spektralwerts repräsentieren.

[0081] In Schritte 24 werden ferner Rückkopplungsfrequenzen 50 bestimmt, die als Teil der Filterdaten 46 für ein Rückkopplungsfilter verwendet werden, welches in Schritt 34 zum Einsatz kommt. Die Bestimmung der Rückkopplungsfrequenzen wird nachfolgend anhand von Fig. 4 näher erläutert.

[0082] Aus den Spektralwerten 56 werden mittels einer Maximalwertanalyse mehrere Kandidaten selektiert, die mögliche Rückkopplungsfrequenzen repräsentieren. Beispielsweise können als Kandidaten aus den Spektralwerten 56 diejenigen Spektralwerte herausgesucht werden, die jeweils den höchsten Betrag aller Spektralwerte eines Blocks aufweisen und von Spektralwerten mit ähnlichem Betrag benachbart sind. Die Kandidaten repräsentieren somit die Maxima von ausgeprägten Extrema des Spektrums. Für jeden Kandidaten werden drei Parameterwerte bestimmt (Schritt 74) und mit einem jeweiligen Schwellenwert verglichen (Schritt 78). Die Schwellenwerte sind für jeden Parameter vorzugsweise fest eingestellt, weil die Parameter in der Regel unempfindlich gegen eine im Vergleich zum Hintergrundrauschen geringe Sprachsignallautstärke sind.

[0083] Ein erster Parameter repräsentiert das Verhältnis zwischen dem Betrag des Kandidaten und den zugehörigen Harmonischen (Peak-to-Harmonic Ratio, PHPR). Vorzugsweise werden die ersten beiden Harmonischen herangezogen, d.h. die Spektralwerte, die im Vergleich zum Kandidaten die doppelte und dreifache Frequenz repräsentieren. Hohe PHPR-Werte deuten auf eine Rückkopplungsfrequenz (Feedbackfrequenz) hin, weil Sprache in der Regel eine klare Obertonstruktur mit Harmonischen aufweist.

[0084] Ein zweiter Parameter repräsentiert das Verhältnis zwischen dem Betrag des Kandidaten und dem Betrag von unmittelbar benachbarten Spektralwerten (Peak-to-Neighbouring Ratio, PNPR). Vorzugsweise werden die ersten drei benachbarten Spektralwerte in jeder Frequenzrichtung herangezogen. Hohe PNPR-Werte deuten auf eine Rückkopplungsfrequenz hin, weil Sprache in der Regel weniger steile Frequenzmaxima aufweist.

[0085] Ein dritter Parameter repräsentiert den zeitlichen Verlauf des Betrags des Kandidaten (Interframe Magnitude Slope Deviation, IMSD). Vorzugsweise wird der mittlere Anstieg des Betrags des Kandidaten sowie mehrerer benachbarter Spektralwerte über fünf vorhergehende Blöcke ermittelt. Positive IMSD-Werte von z.B. 0,5 dB deuten typischerweise auf eine Rückkopplungsfrequenz hin, weil der Betrag der Grundfrequenz von Sprache über mehrere Blöcke hinweg in der Regel nicht ansteigt.

[0086] Für weiteren Informationen zur Berechnung der Parameter wird auf die Veröffentlichung, T.V. Waterschoot, M. Moonen, "Comparative Evaluation of Howeling Detection Criteria in Notch-Filter-Based Howling Supression", Journal of the Audio Engineering Society, Vol. 58, pp. 923-940, 2010, verwiesen.

[0087] Wenn alle drei Parameter zur Bestimmung der Rückkopplungsfrequenz für einen betreffenden Kandidaten das zugeordnete Schwellenwertkriterium verletzen, wird die Rückkopplungsfrequenz vorzugsweise als ein Maximums des Spektrums im Bereich des betreffenden Kandidaten ermittelt. Hierzu wird das Spektrum auf der Grundlage des Kandidaten und der benachbarten Spektralwerte mit einer Interpolationsfunktion (z.B. durch parabolische Interpolation) interpoliert und sodann das Maximum der Interpolationsfunktion gebildet. Dieses Maximum kann insbesondere zwischen zwei Spektralwerten liegen, sodass das interpolierte Maximum genauer ist. Die auf diese Weise bestimmte Rückkopplungsfrequenz wird als Teil der Filterdaten 50 dem Rückkopplungsfilter zugrunde gelegt (Schritt 34).

[0088] Zur Entlastung der Rechnerressourcen ist es bevorzugt, für einen vorbestimmten Zeitraum nach einer erfolgreich bestimmten Rückkopplungsfrequenz den zugrundeliegenden Kandidaten nicht erneut der Parameteranalyse zu unterziehen, wenn der Kandidat erneut als solcher identifiziert wird. Beispielsweise werden dieselben Kandidaten innerhalb eines Zeitfensters von 1 Sekunde nicht erneut daraufhin überprüft, ob Sie eine Rückkopplungsfrequenz repräsentieren oder nicht. Stattdessen wird die für den zeitlich vorherigen Kandidaten bestimmte Rückkopplungsfrequenz für den nachfolgenden, selben Kandidaten übernommen, weil eine hohe Wahrscheinlichkeit dafür besteht, dass dieselbe Rückkopplungsfrequenz auch für den nachfolgenden Kandidaten bestimmt werden würde. Erst nach Ablauf der vorbestimmten Zeit wird ein betreffender Kandidat erneut überprüft.

[0089] Für jede bestimmte Rückkopplungsfrequenz ist in dem Rückkopplungsfilter ein sogenanntes Glocken-Filter (Peak-Filter) vorgesehen, dessen Mittenfrequenz auf die bestimmte Rückkopplungsfrequenz eingestellt wird. Der Q-Wert der Filter wird vorzugsweise auf einen festen Wert eingestellt. Außerdem wird die Verstärkung des Filters vorzugsweise adaptiv eingestellt, wie nachfolgend anhand von Fig. 5 erläutert wird.

[0090] Der in Fig. 5 dargestellte Algorithmus verwirklicht einen endlichen Automaten (Finite-State Machine, FSM), der sich zunächst in einem inaktiven Zustand 90 befindet, d.h. das Glockenfilter hat eine Verstärkung von 0 dB und beeinflusst das Audiosignal nicht. Bei einer neu bestimmten Rückkopplungsfrequenz wird in einen akti-

ven Zustand 92 gewechselt, in dem das Glockenfilter mit voller (negativer) Verstärkung betrieben wird. Nach Ablauf einer ersten vorbestimmten Zeit X wird in einen ersten Reduktionszustand 94 gewechselt, wenn bis dahin die Rückkopplungsfrequenz nicht erneut bestimmt worden ist und der aktive Zustand deswegen beibehalten wird (Rückführung 96). Im ersten Reduktionszustand hat das Glockenfilter eine reduzierte Verstärkung, beispielsweise 2/3 der vollen Verstärkung. Das Rückkopplungsfilter wird somit mit abgeschwächter Wirksamkeit betrieben. Nach Ablauf einer zweiten vorbestimmten Zeit Y wird in einen zweiten Reduktionszustand 98 gewechselt, wenn bis dahin die Rückkopplungsfrequenz nicht erneut bestimmt worden ist und der aktive Zustand beibehalten wird (Rückführung 96).

[0091] Nach erneutem Ablauf der zweiten vorbestimmten Zeit Y wird in dritten Reduktionszustand 100 gewechselt, in dem das Glockenfilter für den Wechsel in den inaktiven Zustand beim nächsten Filterdurchlauf vorgemerkt ist.

[0092] Die zeitabhängige Adaption des Rückkopplungsfilters ist aus mehreren Gründen vorteilhaft. Einerseits wird sichergestellt, dass eine bestimmte Rückkopplungsfrequenz ausreichend lange gefiltert wird. Rückkopplungen halten in der Regel für mindestens einige 100 ms an, sodass eine ausreichend lange Filterung erforderlich ist, um die Rückkopplung wirksam zu unterdrücken. Darüber hinaus werden aufgrund der stufenweisen Reduktion des Rückkopplungsfilters hörbare Verzerrungen des Audiosignals reduziert.

[0093] In Schritt 26 wird das Audiosignal mit einem zweistufigen Kompressor gefiltert, um Pegelspitzen zu entfernen, die zu hörbaren Verzerrungen führen können. Eine erste Kompressorstufe wird bei einem Signalpegel oberhalb eines ersten Schwellenwerts aktiviert und filtert das Audiosignal mit einem ersten Filter, welches moderate Pegelspitzen mit einem geringen Kompressionsgrad reduziert (z.B. Ratio 20, Anstiegszeit 10 ms, Ausklingzeit 100 ms). Die zweite Kompressorstufe wird bei einem Signalpegel oberhalb eines zweiten Schwellenwerts aktiviert, welcher größer als der erste Schwellenwert ist. Das Audiosignal wird dann mit einem zweiten Filter gefiltert, um extreme Pegelspitzen besonders wirksam zu beseitigen. Hierzu wird ein stärkerer Kompressionsgrad gewählt (z.B. Ratio 1000, Anstiegszeit 0,1 ms, Ausklingzeit 5 ms). Die zweite Kompressorstufe stellt ein Notfallfilter dar, um zu gewährleisten, dass alle Amplitudenwerte unterhalb eines kritischen Maximalwerts liegen

[0094] In Schritt 28 wird das Audiosignal mit einem Bandpass gefiltert, um potentielle Störsignale zu entfernen. Hierzu werden vorzugsweise alle Spektralanteile, die zumindest überwiegen keine Sprache repräsentieren aufweisen, reduziert. Sprachsignalanteile sind überwiegend auf den Frequenzbereich zwischen 70 und 8000 Hz begrenzt, sodass Spektralanteile außerhalb dieses Frequenzbereichs gefiltert werden können. Als Bandpassfilter wird vorzugsweise ein doppelt kaskadierter Hochpass zweiter Ordnung mit einem ebenfalls doppelt kaskadierten Tiefpass zweiter Ordnung kombiniert. Der Hochpass und der Tiefpass weisen vorzugsweise jeweils eine Flankensteilheit von 24 dB pro Oktave auf. Die Grenzfrequenzen liegen vorzugsweise im Bereich zwischen 60 und 80 Hz (untere Grenzfrequenz) und zwischen 8000 und 10000 Hz (obere Grenzfrequenz). Ferner sollten sich die Q-Werte der Filter über eine Oktave erstrecken und z.B. Werte im Bereich von 1,4 aufweisen.

[0095] In Schritt 30 wird das Audiosignal mit einem zweiten Kompressor gefiltert, um den Dynamikumfang des Audiosignals zu reduzieren. Hierdurch wird die subjektive Lautstärke einheitlicher und die Sprachverständlichkeit wird verbessert. Als Kompressor dient ein Filter mit relativ mildem Kompressionsgrad, der insbesondere geringer ist, als die Kompressionsgrade des ersten Kompressors von Schritt 28. Beispielsweise kann ein niedriges Ratio gewählt werden, welches den Wert von drei nicht übersteigen sollte. Außerdem sind vorzugsweise längere Anstiegs- bzw. Ausklingzeiten im Bereich von 0,5 und 1 Sekunden vorgesehen.

[0096] In Schritt 32 wird das Audiosignal mit einem Entzerrer gefiltert, um spektrale Variationen zu reduzieren. Der Entzerrer wird hierzu mit acht Glockenfiltern betrieben, deren Mittenfrequenzen denjenigen der Oktavbandfilter von Fig. 3 entsprechen, die zur Bestimmung der Oktavspektralwerte dienen. Die Q-Werte der Glockenfilter sind vorzugsweise so eingestellt, dass sie jeweils etwa eine Oktave abdecken. Für jedes Glockenfilter ist ein eigener Verstärkungsfaktor vorgesehen, der in Abhängigkeit von den Oktavspektralwerten 48 und vordefinierten Referenzspektralwerten bestimmt wird. Die Referenzspektralwerte korrespondieren in ihrer Spektralauflösung mit den Oktavspektralwerten, sodass jedem Oktavspektralwert ein Referenzspektralwert zugeordnet ist.

[0097] Die Referenzspektralwerte bilden zusammen eine Referenzspektralkurve, deren Form mit einer hohen Sprachverständlichkeit korreliert ist und beispielsweise durch spektrale Auswertung einer Vielzahl von ungestörten Sprachsignalen, z.B. auf der Grundlage eines Mittelwerts des oktavgefilterten Spektrums ermittelt werden kann. Jeder Oktavspektralwert wird mit einem zugeordneten Referenzspektralwert verglichen, um einen Verstärkungsfaktor zu ermitteln, welcher die Abweichung zwischen dem Oktavspektralwert und dem zugeordneten Referenzspektralwert repräsentiert. Wenn ein betreffender Oktavspektralwert beispielsweise einen Betrag unterhalb des zugeordneten Referenzspektralwerts aufweist, wird ein Verstärkungsfaktor für das Glockenfilter dieses Spektralbereichs derart bestimmt, dass eine Gewichtung des Oktavspektralwerts mit dem Gewichtungsfaktor den Referenzspektralwert zumindest näherungsweise ergibt. Die Verstärkungsfaktoren sind auf diese Weise dazu angepasst, das Frequenzspektrum des Audiosignals in Übereinstimmung mit der Referenzspektralkurve zu bringen und somit spektrale Variationen innerhalb des Audiosignals und zwischen verschiedenen Audiosignalen zu reduzieren. Beispielsweise werden Eigenschaften unterschiedlicher Sprecher oder spektrale

Einflüsse durch unterschiedliche Mikrofonpositionen zugunsten einer hohen Sprachverständlichkeit ausgeglichen.

[0098] Zur Vermeidung von Verzerrungen werden die Verstärkungsfaktoren vorzugsweise nach oben und unten begrenzt. Darüber hinaus wird auch die zeitliche Änderung der Verstärkungsfaktoren begrenzt.

[0099] Die Glockenfilter zur Filterung des Audiosignals in Schritt 32 werden vorzugsweise nur zur Filterung von Blöcken verwendet, die als Sprachabschnitt detektiert worden sind. Somit wird die Anpassung des Spektrums an die Referenzspektralkurve auf Sprachabschnitte begrenzt. Etwaige Verzerrungen sowie eine ineffiziente Nutzung der Rechenressourcen werden somit vermieden.

[0100] Die Filterung mit dem Entzerrer bzw. den Glockenfiltern in Schritt 32 kann unerwünschte Variationen des Signalpegels verursachen. Um derartige Variationen zu kompensieren, wird das Audiosignal vorzugsweise mit einem Korrekturfaktor gewichtet, welcher als Mittelwert der vorzeicheninvertierten Gewichtungsfaktoren bestimmt wird.

[0101] In Schritt 36 wird das Audiosignal mit einem Pausenfilter gefiltert, um den Signalpegel in Bereichen außerhalb der detektierten Sprachabschnitte, d.h. in Sprachpausen, zu reduzieren und auf diese Weise Störgeräusche zu reduzieren. Hierzu werden die in Schritt 18 bzw. 24 detektierten Sprachabschnitte als Filterdaten 52 herangezogen. Diejenigen Abschnitte des Audiosignals, die nicht als Sprachabschnitte detektiert worden sind, bilden Rauschabschnitte, die durch das Pausenfilter gefiltert werden. Das Audiosignal wird in den detektierten Rauschabschnitten vorzugsweise mit einem festen negativen Verstärkungsfaktor von z.B. -3 dB gewichtet.

[0102] In Schritt 38 wird das Audiosignal mit einem weiteren Entzerrer gefiltert, um die Effekte der verschiedenen Filterungen auszugleichen. Hierzu wird vorzugsweise eine Filterbank bestehend aus 23 Glockenfiltern zwischen 50 Hz und 10 kHz eingesetzt. Die Filter erstrecken sich vorzugsweise jeweils über eine Dritteloktave, wobei der Q-Wert auf 4,3 eingestellt werden kann. Für jedes Glockenfilter ist vorzugsweise ein fester negativer Verstärkungsfaktor vorgesehen.

[0103] In Schritt 40 kann das Audiosignal zu Testzwecken während einer Entwicklungsphase analysiert werden. Diese Möglichkeit ist rein optional und für eine spätere Anwendung des Verfahrens im Praxisbetrieb nicht notwendig.

[0104] In Schritt 42 wird das nunmehr verbesserte Audiosignal zunächst mittels eines Digital-Analog-Wandlers in ein analoges Signal transformiert und sodann über eine Ausgabeschnittstelle bereitgestellt. Von dort kann das Audiosignal für eine Wiedergabe über ein Beschallungssystem abgegriffen werden. Denkbar ist auch die Ausgabe des digitalen Audiosignals anstelle einer analogen Fassung, sofern das Beschallungssystem einen digitalen Signaleingang für das Audiosignal aufweist.

[0105] Mit Bezug auf Fig. 6 wird nachfolgend ein Audiogerät 102 beschrieben, welches dazu eingerichtet ist, das Verfahren von Fig. 1 auszuführen. Das Audiogerät 102 weist ein schematisch angedeutetes Gehäuse 104 auf. Die Außenmaße des Gehäuses 104 sind vorzugsweise nicht größer als wenige Zentimeter, beispielsweise maximal 10 Zentimeter, sodass das Gehäuse 104 insgesamt kompakt und insbesondere auch für mobile Anwendungen geeignet ist.

[0106] Das Audiogerät 102 weist eine Eingangsschnittstelle 112 zum Empfangen eines analogen Audiosignals sowie eine Ausgangsschnittstelle zum Ausgeben des verbesserten Audiosignals aus. Ferner weist die Vorrichtung eine USB-C-Schnittstelle 110 sowie eine Ethernetschnittstelle 108 auf. Die USB-C-Schnittstelle 110 kann allgemein als eine Energieversorgungsschnittstelle zum Anschließen an eine externe Energieversorgung ausgebildet sein. Sie muss nicht zwingend gemäß dem USB-C-Standard ausgebildet sein.

[0107] Zusätzlich oder alternativ können ein oder mehrere Drahtlosschnittstellen vorgesehen sein, um Audiosignale und/oder Steuerungssignale und/oder elektrische Energie auf drahtlosem Wege von außen zu empfangen und/oder zu einem nicht gezeigten Empfänger zu übertragen.

[0108] Die Eingangsschnittstelle 112 und die Ausgangsschnittstelle 106 sind vorzugsweise jeweils als XLR-Schnittstellen ausgebildet, sodass herkömmliche Schallwandler über XLR-Steckverbinder direkt mit dem Audiogerät 102 verbunden werden können.

[0109] Das Audiogerät 102 kann somit insbesondere in einer in Fig. 7 gezeigten Anordnung betrieben werden, in der die Eingangsschnittstelle 112 mit einem Mikrofon 134 zum Erfassen eines Audiosignals von einem nicht gezeigten Sprecher verbunden ist. Ferner ist die Ausgangsschnittstelle 106 über einen Verstärker 130 mit einem Lautsprecher 132 oder einem Beschallungssystem mit mehreren Lautsprechern verbunden, um das mittels des Audiogeräts 102 verbesserte Audiosignal wiederzugeben. Der Lautsprecher 132 und das Mikrofon 134 befinden sich in demselben Raum, beispielsweise einem Konferenzraum oder dergleichen. Die Signalverbesserung erfolgt in Echtzeit, sodass das mit dem Mikrofon 134 aufgenommene Audiosignal im Wesentlichen gleichzeitig über den Lautsprecher 132 wiedergegeben werden kann und somit für eine akustisch vorteilhafte Verstärkung des Audiosignals sorgt.

[0110] Das Audiogerät 102 weist ferner eine manuelle Schnittstelle 128 auf, die in Fig. 6 lediglich schematisch angedeutet ist und allgemein dazu eingerichtet ist, Steuerungsdaten für das Audiogerät 102 durch manuelle Eingabe eines Benutzers unmittelbar an dem Audiogerät 102 zu empfangen.

[0111] Das Audiosignal wird zunächst mit dem Mikrofon 134 erfasst und über die Eingangsschnittstelle 112 einem Vorverstärker 116 zugeführt. Sodann gelangt das Audiosignal in Abhängigkeit von einer Stellung einer Schalteinrichtung 118 entweder über eine Recheneinheit

114 oder direkt zu der Ausgangsschnittstelle 106. Die Schalterstellung der Schalteinrichtung 118 wird über die Recheneinheit 114 gesteuert. Hierzu kann die Recheneinheit 114 von extern über die Schnittstellen 108, 110 und/oder 128 eine Vorgabe empfangen, die festlegt, ob das Audiosignal durch die Recheneinheit 114 geführt und durch diese verbessert werden soll oder nicht. Alternativ oder zusätzlich kann die Recheneinheit 114 im Wege einer Selbstdiagnose ihre Funktionstüchtigkeit zur Ausführung des Verfahrens zu Verbesserung des Audiosignals feststellen und in Abhängigkeit von der Prüfung die Schalterstellung der Schalteinrichtung 118 einstellen. Beispielsweise kann die Schalteinrichtung 118 in einer Grundeinstellung die Eingangsschnittstelle 112 über den Vorverstärker 116 direkt mit der Ausgangsschnittstelle 106 verbinden, wobei die Schalteinrichtung 118 lediglich im Falle der vollen Funktionstüchtigkeit der Recheneinheit 114 einschließlich der notwendigen Energieversorgung umgeschaltet wird, um die Eingangsschnittstelle 112 mit der Recheneinheit 114 zu verbinden. Auf diese Weise wird gewährleistet, dass das Audiosignal von der Ausgangsschnittstelle 106 unabhängig von einer etwaigen Fehlfunktion der Recheneinheit 114 und eines Ausfalls der Energieversorgung abgegriffen werden kann. Das Audiogerät 102 ist somit für den professionellen Einsatz besonders gut geeignet.

[0112] Der Vorverstärker 116 kann mit variabler Verstärkung betrieben werden. Hierzu kann von der Recheneinheit 114 ein jeweiliger Verstärkungswert eingestellt werden. Dieser kann beispielsweise mittels der Schnittstelle 128 aus einer vorbestimmten Menge an unterschiedlichen Verstärkungswerten, z.B. drei Verstärkungswerten, unmittelbar an der Vorrichtung 102 ausgewählt werden. Die Auswahl des Verstärkungswerts kann dem Bediener durch eine Leuchtanzeige, z.B. durch mehrere LED-Dioden, am Audiogerät 102 visuell vermittelt werden. Durch geeignete Einstellung der Vorverstärkung können große Pegelvariationen vorzugsweise bereits im analogen Signal kompensiert werden, sodass digitales Rauschen aufgrund hoher Verstärkungen des Digitalsignals vermieden werden kann.

[0113] Zur Energieversorgung des Audiogeräts 102 ist einerseits die Schnittstelle 110 vorgesehen, die mittels zugeordnetem Versorgungskabel mit einer Netzquelle verbunden werden kann, um das Audiogerät 102 im Netzbetrieb zu betreiben. Alternativ kann das Audiogerät 102 über einen in dem Gehäuse 104 integrierten Energiespeicher, beispielsweise einen elektrischen Akku 126, versorgt werden. Der Akku 126 ist mit der Schnittstelle 110 gekoppelt und kann über diese geladen werden. Anstelle der USB-C-Schnittstelle 110 kann auch ein anderer Schnittstellentyp zur Energieversorgung vorgesehen sein.

[0114] Zum Schutz vor Überspannung oder Falschpolung ist die Vorrichtung 102 vorzugsweise mit einer elektrischen Schutzeinrichtung 120 ausgestattet, welche die elektrischen Verbraucher des Audiogeräts 102 vor Spannungsschäden schützt. Hierzu zählen insbesondere die Recheneinheit 114, ein Lüfter 124 zum Kühlen der Recheneinheit 114 und eine Phantomspeisungseinrichtung 122, die mit der Eingangsschnittstelle 112 gekoppelt ist. Die Phantomspeisungseinrichtung 122 dient zur elektrischen Versorgung des an die Eingangsschnittstelle 112 angeschlossenen Mikrofons 134, beispielsweise mit einer Mikrofonversorgungsspannung von 48 Volt. Die Phantomspeisungseinrichtung 122 weist einen nicht näher gezeigten Spannungswandler auf, um die Versorgungsspannung des Audiogeräts 102, die über die USB-C-Schnittstelle 110 bereitgestellt wird, beispielsweise 5 Volt, in die Mikrofonversorgungsspannung zu wandeln.

[0115] Die Recheneinheit 114 ist vorzugsweise als ein Einplatinenrechner ausgebildet, sodass das Audiogerät 102 unter diesem Aspekt kompakt ausgebildet und außerdem kostengünstig hergestellt werden kann. Die Recheneinheit 114 wird insbesondere über eine Busschnittstelle 107 konfiguriert, die vorzugsweise vom Typ USB-A ist. Die Schnittstelle 107 wird hierzu mit einem Server oder direkt mit einem mobilen Endgerät verbunden (nicht gezeigt), um von außen auf die Recheneinheit 114 zugreifen und wahlweise ein oder mehrere Konfigurationsparameter für das Verfahren von Fig. 1 (z.B. Schwellenwerte, Anstiegs- und Ausklingzeiten) einstellen zu können. Denkbar ist auch eine Konfiguration über die USB-C-Schnittstelle 110.

[0116] Alternativ ist es möglich, einen USB-Stick oder dergleichen an die Schnittstelle 107 anzuschließen, wobei die gewünschten Konfigurationsdaten oder eine neue Firmware in dem USB-Stick gespeichert sind. Die Daten werden sodann automatisch oder nach Initiierung durch einen Bediener über die Schnittstelle 107 an die Recheneinheit 114 übertragen, um die Konfigurationsparameter oder die Firmware entsprechend zu aktualisieren. Dieser Vorgang kann durch einen Endbenutzer der Vorrichtung durchgeführt werden.

[0117] Vorzugsweise ist eine detaillierte Konfiguration von Filterparametern durch den Endbenutzer jedoch nicht erforderlich. In einem internen Speicher der Recheneinheit (nicht gezeigt) sind bereits alle notwendigen Konfigurationsparameterwerte hinterlegt, sodass das Verfahren bei nahezu allen üblichen akustischen Umgebungsbedingungen vollautomatisch gute Ergebnisse gewährleistet. Für besondere akustische Umgebungen kann der Konfigurationsparametersatz beispielsweise durch einen geschulten Fachmann aus der Ferne oder lokal über die Schnittstelle 107 angepasst werden. Für den Endbenutzer fällt somit kein Einrichtungsaufwand an. Zur Inbetriebnahme im Anwendungsfall von Fig. 7 ist es lediglich erforderlich, das Audiogerät 102 über die vorgesehenen Schnittstellen 112 und 106 mit dem Mikrofon 134 und dem Lautsprecher 132 zu verbinden. Sodann kann das Audiogerät 102 direkt im Sinne einer plug-and-play-Funktionalität verwendet werden. Sofern kein Akkubetrieb gewünscht ist, wird das Audiogerät 102 über die USB-C-Schnittstelle 110 mit einer Netzquelle (nicht gezeigt) verbunden, um das Audiogerät 102 elektrisch zu versorgen.

[0118] Das Audiogerät 102 weist ferner eine manuelle Bedienschnittstelle 113 (z.B. mit einer manuell betätigbaren Taste) sowie eine optische Anzeigeeinrichtung 109 auf (z.B. eine LED). Über die Bedienschnittstelle 113 kann ein Benutzer des Audiogeräts 102 eine Aufzeichnung des an der Ausgangsschnittstelle 106 bereitgestellten Audiosignals steuern. Beispielsweise schließt der Benutzer zunächst einen USB-Stick oder dergleichen an die Schnittstelle 107. Dwe USB-Stick wird durch die Recheneinheit 114 detektiert und es wird dem Benutzer an der Anzeigeeinrichtung 109 durch Aktivierung eines ersten Anzeigemodus angezeigt, dass das Audiogerät 102 aufnahmebereit ist. Um das Audiosignal (in seiner digitalen Form) in dem USB-Stick abzuspeichern, wird sodann die Bedienschnittstelle 113 betätigt. Die Anzeigeeinrichtung 109 zeigt den erfolgreichen Start der Aufnahme durch Aktivierung eines zweiten Anzeigemodus an (z.B. blinkende LED). Das Audiosignal wird sodann fortlaufend in einer Datei auf dem USB-Stick abgelegt. Wenn die Speicherkapazität erschöpft ist, wird die Aufnahme automatisch beendet. Dem Benutzer wird dies durch Aktivierung eines dritten Anzeigemodus an der Anzeigeeinrichtung 109 angezeigt. Die Aufnahme kann wahlweise vorzeitig durch nochmalige Betätigung der Bedienschnittstelle 107 beendet werden.

BEZUGSZEICHENLISTE

[0119]

10 Erfassen eines Audiosignals mit einem Mikrofon
12 Vorverstärkung
14 Erfassen des Audiosignals an einer Eingangsschnittstelle
16 Elektronischer Verstärker (Erstes Pegelfilter)
18 Eingangsanalyse
20 Softwareverstärker (Zweites Pegelfilter)
22 Rauschfilter
24 Zwischenanalyse
26 Bandpass
28 Erster Kompressor
30 Zweiter Kompressor
32 Erstes Entzerrfilter und drittes Pegelfilter
34 Rückkopplungsfilter
36 Pausenfilter
37 Betragsfrequenzgang
38 Zweites Entzerrfilter
40 Ausgangsanalyse
42 Bereitstellen des Audiosignals an einer Ausgabeschnittstelle
44 Erste Filterdaten
46 Zweite Filterdaten
48 Oktavlautstärken
50 Rückkopplungsfrequenzen
52 Detektierte Sprachabschnitte
54 Amplitudenwerte
56 Spektralwerte
58 Parameterberechnung
60 Schwellenwertberechnung
62 Vergleichen mit Schwellenwerten
64 Bestimmen ob Schwellenwerte verletzt
66 Bestimmen Anzahl aufeinanderfolgender Abschnitte
68 Vergleichen mit Mindestanzahl
70 Detektion als Sprachabschnitt
72 Detektion als Rauschabschnitt
74 Suche Frequenzkandidaten
76 Parameterberechnung
78 Vergleich mit Schwellenwerten
80 Verzweigung
82 Interpolation
84 Speicherung Rückkopplungsfrequenz
86 Löschen der Rückkopplungsfrequenz
88 Ende
90 Inaktiver Zustand
92 Aktiver Zustand
94 Erster Reduktionszustand
96 Rückführung
98 Zweiter Reduktionszustand
100 Dritter Reduktionszustand
102 Audiogerät
104 Gehäuse
106 Ausgangsschnittstelle
107 USB-A-Schnittstelle
108 Ethernetschnittstelle
109 Anzeigeeinrichtung
110 USB-C-Schnittstelle
112 Eingangsschnittstelle
113 Manuelle Bedienschnittstelle
114 Recheneinheit
116 Vorverstärker
118 Schalteinrichtung
120 Schutzeinrichtung
122 Phantomspeisung
124 Lüfter
126 Energiespeicher
128 Manuelle Bedienschnittstelle
130 Verstärker
132 Lautsprecher
134 Mikrofon

F Frequenz
G Betrag
fc Mittenfrequenz

Patentansprüche

1. Verfahren zur Verbesserung eines Audiosignals, insbesondere in Echtzeit, wobei das Verfahren zumindest folgende Schritte umfasst:

   - Empfangen eines Audiosignals mit mehreren Amplitudenwerten, wobei das Audiosignal zumindest abschnittsweise Sprache aufweist;
   - Detektieren von Sprachabschnitten des Audi-

osignals (18, 24);
- Filtern des Audiosignals mit wenigstens einem Pegelfilter (16), um Signalpegelvariationen des Audiosignals in den detektierten Sprachabschnitten zu reduzieren;
- Bestimmen einer Rückkopplungsfrequenz (50), welche eine Rückkopplung des Audiosignals repräsentiert;
- Filtern des Audiosignals mit einem Rückkopplungsfilter (34) auf der Grundlage der bestimmten Rückkopplungsfrequenz (50), um Rückkopplungen repräsentierende Spektralanteile des Audiosignals zu reduzieren; und
- Filtern des Audiosignals mit wenigstens einem Entzerrfilter (32), um spektrale Variationen des Audiosignals in den detektierten Sprachabschnitten zu reduzieren, wobei das Filtern mit dem wenigstens einen Entzerrfilter (32) umfasst:

  - Bestimmen von Grobspektralwerten (48) auf der Grundlage von Feinspektralwerten (56) des Audiosignals, wobei die Grobspektralwerte (48) die Feinspektralwerte (56) mit einer geringeren Spektralauflösung als die Feinspektralwerte (56) repräsentieren;
  - Bestimmen von ersten Entzerrgewichten, die eine Abweichung der Grobspektralwerte (48) von vorbestimmten Referenzspektralwerten repräsentieren;
  - Gewichten des Audiosignals mit den ersten Entzerrgewichten, um Spektralwerte des Audiosignals in Übereinstimmung mit den Referenzspektralwerten zu bringen;

wobei das Bestimmen der Rückkopplungsfrequenz (50) umfasst:

- Bestimmen einer Untermenge von Spektralwerten des Audiosignals, die einen vorbestimmten Spektralschwellenwert verletzen (74);
- Bestimmen von mehreren ersten Spektralparameterwerten auf der Grundlage der Untermenge, wobei jeder der ersten Spektralparameterwerte eine vorbestimmte Relation zwischen einem zugeordneten Spektralwert der Untermenge und wenigstens einem zeitlich und/oder spektral benachbarten Spektralwert repräsentiert (76); und
- Bestimmen der Rückkopplungsfrequenz (50) auf der Grundlage der mehreren ersten Spektralparameterwerte (78, 80, 82, 84).

2. Verfahren nach Anspruch 1,
ferner umfassend Bestimmen von mehreren Spektralwerten (56) auf der Grundlage der Amplitudenwerte (54), wobei die Amplitudenwerte (54) das Audiosignal in einem Zeitbereich repräsentieren und wobei die Spektralwerte (56) das Audiosignal in einem Frequenzbereich repräsentieren, und wobei das Detektieren der Sprachabschnitte (18, 24), das Filtern mit dem wenigstens einen Pegelfilter (16) und/oder das Filtern mit dem wenigstens einen Entzerrfilter (32) auf der Grundlage der Amplitudenwerte (54) und/oder der Spektralwerte (56) erfolgt.

3. Verfahren nach Anspruch 1 oder 2,
wobei das Detektieren der Sprachabschnitte (18, 24) umfasst:

- Bestimmen wenigstens eines ersten Energieparameterwerts auf der Grundlage der Amplitudenwerte (54), wobei der erste Energieparameterwert eine mittlere Energie des Audiosignals für mehrere der Amplitudenwerte (54) repräsentiert;
- Bestimmen wenigstens eines zweiten Spektralparameterwerts auf der Grundlage von Spektralwerten (56) des Audiosignals, wobei der wenigstens eine zweite Spektralparameterwert eine harmonische Spektralstruktur des Audiosignals für mehrere der Spektralwerte (56) repräsentiert; und
- Detektieren eines Abschnitts des Audiosignals als Sprachabschnitt, wenn der wenigstens eine erste Energieparameterwert einen ersten Energieparameterschwellenwert und/oder der wenigstens eine zweite Spektralparameterwert einen Spektralparameterschwellenwert verletzt (62, 64), insbesondere wobei der Energieparameterschwellenwert und/oder der Spektralparameterschwellenwert in Abhängigkeit von der Zeit angepasst wird.

4. Verfahren nach einem der vorhergehenden Ansprüche, wobei das Filtern des Audiosignals mit dem wenigstens einen Pegelfilter (16) umfasst:

- Bestimmen wenigstens eines Pegelparameterwerts auf der Grundlage der Amplitudenwerte (54), wobei der Pegelparameterwert einen mittleren Pegel des Audiosignals für einen detektierten Sprachabschnitt repräsentiert;
- Bestimmen von wenigstens einem Kompensationsgewicht auf der Grundlage des wenigstens einen Pegelparameterwerts;
- Gewichten des Audiosignals mit dem wenigstens einen Kompensationsgewicht, um die Signalpegelvariationen des Audiosignals zu reduzieren.

5. Verfahren nach Anspruch 4,

wobei der wenigstens eine Pegelparameterwert erste und zweite Pegelparameterwerte für mehrere detektierten Sprachabschnitte umfasst,

wobei die ersten Pegelparameterwerte den mittleren Pegel des Audiosignals mit einer ersten Zeitauflösung repräsentieren, wobei die zweiten Pegelparameterwerte den mittleren Pegel des Audiosignals mit einer zweiten Zeitauflösung repräsentieren, wobei die zweite Zeitauflösung größer als die erste Zeitauflösung ist, und wobei das wenigstens eine Kompensationsgewicht auf der Grundlage der ersten und zweiten Pegelparameterwerte ermittelt wird,
insbesondere wobei die ersten Pegelparameterwerte durch Energiemittelwerte und/oder erste Lautstärkewerte und die zweiten Pegelparameterwerte durch zweite Lautstärkewerte gebildet sind.

6. Verfahren nach Anspruch 4 oder 5,

wobei das wenigstens eine Kompensationsgewicht erste Kompensationsgewichte und zweite Kompensationsgewichte umfasst, wobei die ersten Kompensationsgewichte bestimmt werden, um Signalpegelvariationen mit wenigstens einem Pegel, der größer als ein vorbestimmter Pegelschwellenwert ist, zu reduzieren,
wobei die zweiten Kompensationsgewichte bestimmt werden, um den Signalpegel des Audiosignals auf einen vorbestimmten Wert einzustellen.

7. Verfahren nach einem der vorhergehenden Ansprüche, wobei das Filtern mit dem wenigstens einen Entzerrfilter (32) umfasst:
Gewichten des Audiosignals mit zweiten Entzerrgewichten (38), wobei die zweiten Entzerrgewichte vorbestimmt sind.

8. Verfahren nach einem der vorhergehenden Ansprüche, ferner umfassend:
Filtern des Audiosignals mit wenigstens einem Kompressor (28, 30), um einen Dynamikumfang des Audiosignals zu reduzieren, insbesondere wobei für den wenigstens einen Kompressor (28, 30) mehrere voneinander verschiedene Parametersätze vorgesehen sind, die in Abhängigkeit von einem Betrag des Audiosignals ausgewählt und der Filterung mit dem wenigstens einen Kompressor zugrunde gelegt werden, wobei sich die mehreren Parametersätze in einem Kompressionsgrad voneinander unterscheiden.

9. Verfahren nach einem der vorhergehenden Ansprüche,
wobei, wenn die bestimmte Rückkopplungsfrequenz (50) zwischen aufeinanderfolgenden Zeitabschnitten des Audiosignals verschwindet, die Wirksamkeit des Rückkopplungsfilters (34) über mehrere Zeitabschnitte schrittweise reduziert wird (94, 98, 100).

10. Verfahren nach einem der vorhergehenden Ansprüche, ferner umfassend:

- Filtern des Audiosignals mit einem Pausenfilter (36), um das Audiosignal in Bereichen außerhalb der detektierten Sprachabschnitte zu reduzieren; und/oder
- Filtern des Audiosignals mit einem Rauschfilter (22), um das Audiosignal in Bereichen mit Amplitudenwerten, die einen vorbestimmten Rauschschwellenwert verletzen, zu reduzieren und/oder
- Filtern des Audiosignals mit einem Bandpassfilter (26), wobei eine untere Grenzfrequenz des Bandpassfilters vorzugsweise in einem Bereich von 50 bis 100 Hz liegt, und wobei eine obere Grenzfrequenz des Bandpassfilters vorzugsweise in einem Bereich von 8000 bis 10000 Hz liegt.

11. Vorrichtung zur Verbesserung eines Audiosignals, insbesondere in Echtzeit, wobei das Audiosignal Sprache aufweist, wobei die Vorrichtung (102) umfasst:

- wenigstens eine Eingangsschnittstelle (112) zum Erfassen eines Audiosignals, wobei die Eingangsschnittstelle (112) einen Anschluss für ein Mikrofon (134) aufweist;
- wenigstens eine Ausgangsschnittstelle (106) zum Ausgeben des Audiosignals, wobei die Ausgangsschnittstelle (106) einen Anschluss für ein Audiowiedergabegerät (130, 132) aufweist; und
- eine Recheneinheit (114), die zum Ausführen eines Verfahrens zur Verbesserung des Audiosignals nach einem der vorhergehenden Ansprüche eingerichtet ist.

12. Vorrichtung nach Anspruch 11,
ferner umfassend:

- einen Vorverstärker (116) für das Audiosignal, wobei der Vorverstärker (116) mit der Eingangsschnittstelle (112) koppelbar ist; und/oder
- eine elektrische Versorgung (122) für die Eingangsschnittstelle (112); und/oder
- eine Schalteinrichtung (118), die mit der Eingangsschnittstelle (112), der Ausgangsschnittstelle (106) und der Recheneinheit (114) koppelbar ist; und/oder
- eine Kühleinrichtung (124);

und/oder wobei die Recheneinheit (114) einen Einplatinenrechner aufweist; und/oder
wobei die Vorrichtung (102) ein Gehäuse (104) und/oder wenigstens eine externe Kommunikationsschnittstelle (108, 110) aufweist.

Claims

1. A method for enhancing an audio signal, in particular in real time, the method comprising at least the following steps:

    - Receiving an audio signal having multiple amplitude values, the audio signal having speech at least in sections;
    - Detecting speech sections of the audio signal (18, 24);
    - filtering the audio signal with at least one level filter (16) to reduce signal level variations of the audio signal in the detected speech portions;
    - determining a feedback frequency (50) representing a feedback of the audio signal;
    - filtering the audio signal with a feedback filter (34) based on the determined feedback frequency (50) to reduce spectral components of the audio signal representing feedback; and
    - filtering the audio signal with at least one equalization filter (32) to reduce spectral variations of the audio signal in the detected speech portions, the filtering with the at least one equalization filter (32) comprising:

        - determining coarse spectral values (48) based on fine spectral values (56) of the audio signal, wherein the coarse spectral values (48) represent the fine spectral values (56) at a lower spectral resolution than the fine spectral values (56);
        - determining first equalization weights representing a deviation of the coarse spectral values (48) from predetermined reference spectral values;
        - weighting the audio signal with the first equalization weights to bring spectral values of the audio signal into agreement with the reference spectral values;

    wherein determining the feedback frequency (50) comprises:

    - Determining a subset of spectral values of the audio signal that violate a predetermined spectral threshold (74);
    - determining a plurality of first spectral parameter values based on the subset, each of the first spectral parameter values representing a predetermined relationship between an associated spectral value of the subset and at least one temporally and/or spectrally adjacent spectral value (76); and
    - determining the feedback frequency (50) based on the plurality of first spectral parameter values (78, 80, 82, 84).

2. The method of claim 1,
further comprising determining a plurality of spectral values (56) based on the amplitude values (54), wherein the amplitude values (54) represent the audio signal in a time domain and wherein the spectral values (56) represent the audio signal in a frequency domain, and wherein detecting the speech portions (18, 24), filtering with the at least one level filter (16) and/or filtering with the at least one equalization filter (32) is based on the amplitude values (54) and/or the spectral values (56).

3. Method according to claim 1 or 2,

    wherein detecting the speech segments (18, 24) comprises:

    - Determining at least a first energy parameter value based on the amplitude values (54), wherein the first energy parameter value represents an average energy of the audio signal for a plurality of the amplitude values (54);
    - determining at least one second spectral parameter value based on spectral values (56) of the audio signal, the at least one second spectral parameter value representing a harmonic spectral structure of the audio signal for a plurality of the spectral values (56); and
    - detecting a portion of the audio signal as a speech portion when the at least one first energy parameter value violates a first energy parameter threshold value and/or the at least one second spectral parameter value violates a spectral parameter threshold value (62, 64),

    in particular wherein the energy parameter threshold value and/or the spectral parameter threshold value is adjusted as a function of time.

4. A method according to any one of the preceding claims,
wherein filtering the audio signal with the at least one level filter (16) comprises:

    - Determining at least one level parameter value based on the amplitude values (54), wherein the level parameter value represents an average level of the audio signal for a detected speech portion;
    - determining at least one compensation weight based on the at least one level parameter value;
    - weighting the audio signal with the at least one compensation weight to reduce signal level variations of the audio signal.

5. A method according to claim 4,

wherein the at least one level parameter value comprises first and second level parameter values for a plurality of detected speech segments, wherein the first level parameter values represent the average level of the audio signal at a first time resolution, wherein the second level parameter values represent the average level of the audio signal at a second time resolution, wherein the second time resolution is greater than the first time resolution, and wherein the at least one compensation weight is determined based on the first and second level parameter values,
in particular wherein the first level parameter values are formed by average energy values and/or first loudness values and the second level parameter values are formed by second loudness values.

6. Method according to claim 4 or 5,

wherein said at least one compensation weight comprises first compensation weights and second compensation weights, said first compensation weights being determined to reduce signal level variations having at least one level greater than a predetermined level threshold value,
wherein the second compensation weights are determined to adjust the signal level of the audio signal to a predetermined value.

7. A method according to any one of the preceding claims,
wherein filtering with the at least one equalization filter (32) comprises:

weighting the audio signal with second equalization weights (38),
wherein the second equalization weights are predetermined.

8. A method according to any one of the preceding claims, further comprising:

filtering the audio signal with at least one compressor (28, 30) to reduce a dynamic range of the audio signal,
in particular wherein a plurality of mutually different parameter sets are provided for the at least one compressor (28, 30), which parameter sets are selected in dependence on a magnitude of the audio signal and are used as a basis for the filtering with the at least one compressor, wherein the plurality of parameter sets differ from one another in a degree of compression.

9. Method according to any one of the preceding claims,
wherein, when the determined feedback frequency (50) disappears between successive time periods of the audio signal, the effectiveness of the feedback filter (34) is gradually reduced (94, 98, 100) over a plurality of time periods.

10. Method according to any one of the preceding claims,
further comprising:

- Filtering the audio signal with a pause filter (36) to reduce the audio signal in regions outside of the detected speech segments; and/or
- filtering the audio signal with a noise filter (22) to reduce the audio signal in regions having amplitude values that violate a predetermined noise threshold; and/or
- filtering the audio signal with a bandpass filter (26), wherein a lower cut-off frequency of the bandpass filter preferably lies in a range of 50 to 100 Hz, and wherein an upper cut-off frequency of the bandpass filter preferably lies in a range of 8000 to 10000 Hz.

11. Device for enhancing an audio signal, in particular in real time, said audio signal comprising speech, the device (102) comprising:

- at least one input interface (112) for detecting an audio signal, the input interface (112) having a connector for a microphone (134);
- at least one output interface (106) for outputting the audio signal, the output interface (106) having a connector for an audio playback device (130, 132); and
- a computing unit (114) for performing a method for enhancing the audio signal according to any of the preceding claims.

12. Device according to claim 11,

further comprising:

- a preamplifier (116) for the audio signal, the preamplifier (116) being couplable to the input interface (112); and/or
- an electrical supply (122) for the input interface (112); and/or
- a switching means (118) couplable to the input interface (112), the output interface (106) and the computing unit (114); and/or
- a cooling device (124);

and/or wherein the computing unit (114) comprises a single board computer; and/or
wherein the device (102) comprises a housing

(104); and/or
at least one external communication interface (108, 110).

Revendications

1. Procédé d'amélioration d'un signal audio, en particulier en temps réel, le procédé comprenant au moins les étapes suivantes :

- Réception d'un signal audio ayant plusieurs valeurs d'amplitude, le signal audio comprenant au moins des portions de parole ;
- Détection de portions de parole du signal audio (18, 24) ;
- filtrer le signal audio avec au moins un filtre de niveau (16) pour réduire les variations de niveau de signal du signal audio dans les portions de parole détectées ;
- déterminer une fréquence de rétroaction (50) représentant une rétroaction du signal audio ;
- filtrer le signal audio avec un filtre de Larsen (34) sur la base de la fréquence de Larsen déterminée (50) afin de réduire les composantes spectrales du signal audio représentant le Larsen ; et
- filtrer le signal audio avec au moins un filtre d'égalisation (32) pour réduire les variations spectrales du signal audio dans les portions de parole détectées, le filtrage avec ledit au moins un filtre d'égalisation (32) comprenant:

  - déterminer des valeurs spectrales approximatives (48) sur la base de valeurs spectrales fines (56) du signal audio, les valeurs spectrales approximatives (48) représentant les valeurs spectrales fines (56) avec une résolution spectrale plus faible que les valeurs spectrales fines (56) ;
  - déterminer des premiers poids d'égalisation représentant un écart des valeurs spectrales grossières (48) par rapport à des valeurs spectrales de référence prédéterminées ;
  - pondérer le signal audio avec les premiers poids d'égalisation afin d'amener les valeurs spectrales du signal audio en conformité avec les valeurs spectrales de référence ;

dans lequel la détermination de la fréquence de rétroaction (50) comprend :

- déterminer un sous-ensemble de valeurs spectrales du signal audio qui violent un seuil spectral prédéterminé (74) ;
- déterminer une pluralité de premières valeurs de paramètres spectraux sur la base du sous-ensemble, chacune des premières valeurs de paramètres spectraux représentant une relation prédéterminée entre une valeur spectrale associée du sous-ensemble et au moins une valeur spectrale adjacente temporellement et/ou spectralement (76) ; et
- déterminer la fréquence de rétroaction (50) sur la base de la pluralité de premières valeurs de paramètres spectraux (78, 80, 82, 84).

2. Procédé selon la revendication 1,
comprenant en outre la détermination de plusieurs valeurs spectrales (56) sur la base des valeurs d'amplitude (54), les valeurs d'amplitude (54) représentant le signal audio dans un domaine temporel et les valeurs spectrales (56) représentant le signal audio dans un domaine fréquentiel, et la détection des parties de parole (18, 24), le filtrage avec le au moins un filtre de niveau (16) et/ou le filtrage avec le au moins un filtre d'égalisation (32) étant effectués sur la base des valeurs d'amplitude (54) et/ou des valeurs spectrales (56).

3. Procédé selon la revendication 1 ou 2,
dans lequel la détection des portions de parole (18, 24) comprend :

- la détermination d'au moins une première valeur de paramètre d'énergie sur la base des valeurs d'amplitude (54), la première valeur de paramètre d'énergie représentant une énergie moyenne du signal audio pour plusieurs des valeurs d'amplitude (54) ;
- déterminer au moins une deuxième valeur de paramètre spectral sur la base de valeurs spectrales (56) du signal audio, ladite au moins une deuxième valeur de paramètre spectral représentant une structure spectrale harmonique du signal audio pour une pluralité desdites valeurs spectrales (56) ; et
- détecter une partie du signal audio en tant que partie vocale lorsque la au moins une première valeur de paramètre d'énergie viole une première valeur de seuil de paramètre d'énergie et/ou la au moins une deuxième valeur de paramètre spectral viole une valeur de seuil de paramètre spectral (62, 64), en particulier dans lequel la valeur seuil de paramètre d'énergie et/ou la valeur seuil de paramètre spectral sont adaptées en fonction du temps.

4. Procédé selon l'une quelconque des revendications précédentes,
dans lequel le filtrage du signal audio avec le au moins un filtre de niveau (16) comprend :

- déterminer au moins une valeur de paramètre

de niveau sur la base des valeurs d'amplitude (54), la valeur de paramètre de niveau représentant un niveau moyen du signal audio pour une section de parole détectée ;
- déterminer au moins un poids de compensation sur la base de la au moins une valeur de paramètre de niveau ;
- pondérer le signal audio avec l'au moins un poids de compensation afin de réduire les variations de niveau de signal du signal audio.

5. Procédé selon la revendication 4,

dans lequel ladite au moins une valeur de paramètre de niveau comprend des première et deuxième valeurs de paramètre de niveau pour une pluralité de sections de parole détectées, dans lequel lesdites premières valeurs de paramètre de niveau représentent le niveau moyen du signal audio avec une première résolution temporelle, dans lequel lesdites deuxièmes valeurs de paramètre de niveau représentent le niveau moyen du signal audio avec une deuxième résolution temporelle, dans lequel ladite deuxième résolution temporelle est supérieure à ladite première résolution temporelle, et dans lequel ladite au moins une pondération de compensation est déterminée sur la base desdites première et deuxième valeurs de paramètre de niveau,
en particulier dans lequel les premières valeurs de paramètre de niveau sont formées par des valeurs moyennes d'énergie et/ou des premières valeurs de volume et les secondes valeurs de paramètre de niveau sont formées par des secondes valeurs de volume.

6. Procédé selon la revendication 4 ou 5,

dans lequel ledit au moins un poids de compensation comprend des premiers poids de compensation et des seconds poids de compensation, lesdits premiers poids de compensation étant déterminés pour réduire les variations de niveau de signal ayant au moins un niveau supérieur à un seuil de niveau prédéterminé,
dans lequel les deuxièmes poids de compensation sont déterminés pour ajuster le niveau de signal du signal audio à une valeur prédéterminée.

7. Procédé selon l'une quelconque des revendications précédentes,
dans lequel le filtrage avec ledit au moins un filtre d'égalisation (32) comprend: pondérer le signal audio avec des deuxièmes poids d'égalisation (38), les deuxièmes poids d'égalisation étant prédéterminés.

8. Procédé selon l'une quelconque des revendications précédentes, comprenant en outre :

le filtrage du signal audio avec au moins un compresseur (28, 30) afin de réduire une plage dynamique du signal audio,
en particulier dans lequel il est prévu pour l'au moins un compresseur (28, 30) plusieurs jeux de paramètres différents les uns des autres, qui sont sélectionnés en fonction d'une amplitude du signal audio et qui servent de base au filtrage avec l'au moins un compresseur, les plusieurs jeux de paramètres se différenciant les uns des autres par un degré de compression.

9. Procédé selon l'une quelconque des revendications précédentes,
dans lequel, lorsque la fréquence de rétroaction déterminée (50) disparaît entre des périodes de temps successives du signal audio, l'efficacité du filtre de rétroaction (34) est réduite progressivement (94, 98, 100) sur plusieurs pério des de temps.

10. Procédé selon l'une quelconque des revendications précédentes, comprenant en outre :

- filtrer le signal audio avec un filtre de pause (36) pour réduire le signal audio dans des zones situées à l'extérieur des sections de parole détectées ; et/ou
- filtrer le signal audio avec un filtre de bruit (22) pour réduire le signal audio dans des zones ayant des valeurs d'amplitude qui violent un seuil de bruit prédéterminé ; et/ou
- filtrer le signal audio avec un filtre passe-bande (26), une fréquence de coupure inférieure du filtre passe-bande étant de préférence comprise dans une plage de 50 à 100 Hz, et une fréquence de coupure supérieure du filtre passe-bande étant de préférence comprise dans une plage de 8000 à 10 000 Hz.

11. Dispositif pour améliorer un signal audio, en particulier en temps réel, le signal audio comprenant de la parole,
dans lequel le dispositif (102) comprend :

- au moins une interface d'entrée (112) pour détecter un signal audio, l'interface d'entrée (112) comprenant une connexion pour un microphone (134) ;
- au moins une interface de sortie (106) pour émettre le signal audio, l'interface de sortie (106) ayant une connexion pour un appareil de reproduction audio (130, 132) ; et
- une unité de calcul (114) pour mettre en oeuvre un procédé d'amélioration du signal audio selon l'une quelconque des revendications précéden-

tes.

12. Dispositif selon la revendication 11,

comprenant en outre :

- un préamplificateur (116) pour le signal audio, le préamplificateur (116) pouvant être couplé à l'interface d'entrée (112) ; et/ou
- une alimentation électrique (122) pour l'interface d'entrée (112) ; et/ou
- un dispositif de commutation (118) qui peut être couplé à l'interface d'entrée (112), à l'interface de sortie (106) et à l'unité de calcul (114) ; et/ou
- un dispositif de refroidissement (124) ;

et/ou dans lequel l'unité de calcul (114) comprend un ordinateur monocarte; et/ou
le dispositif (102) comprenant un boîtier (104) et/ou
au moins une interface de communication externe (108, 110).

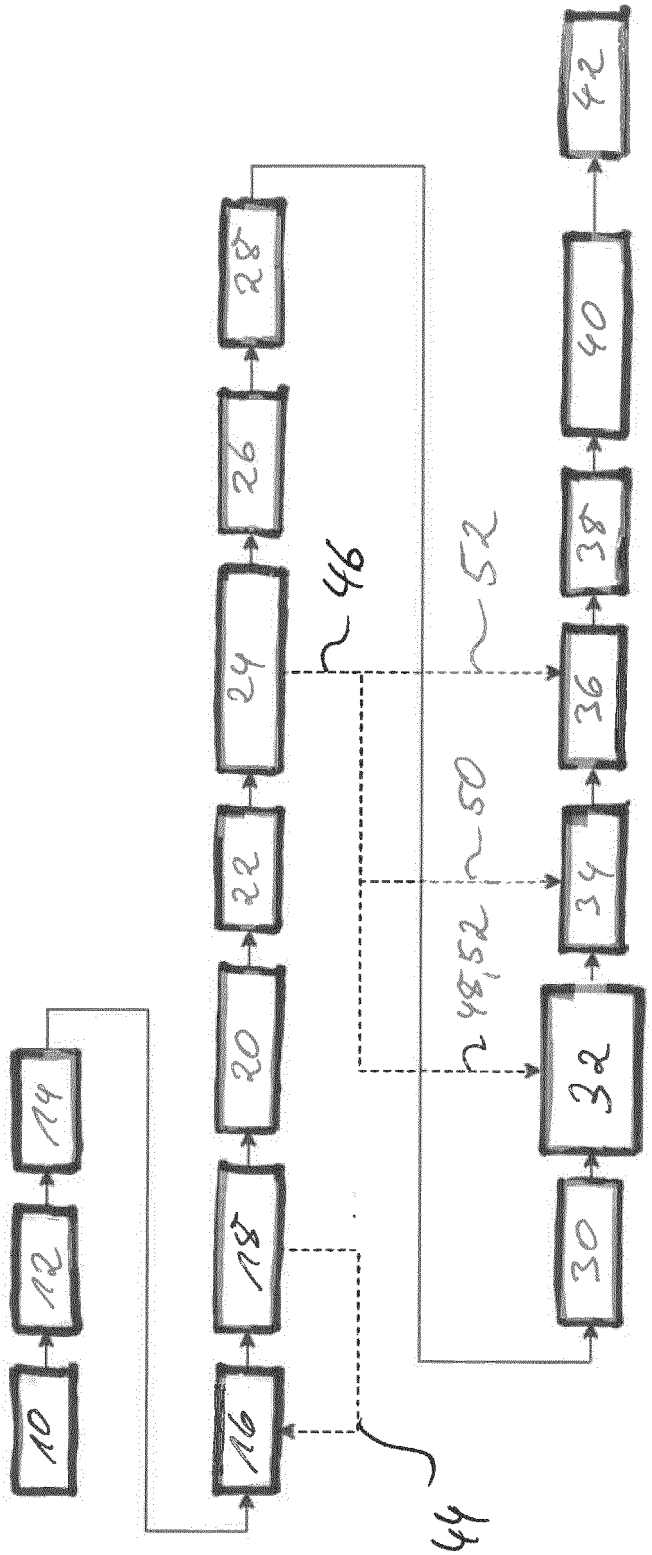

Fig. 1

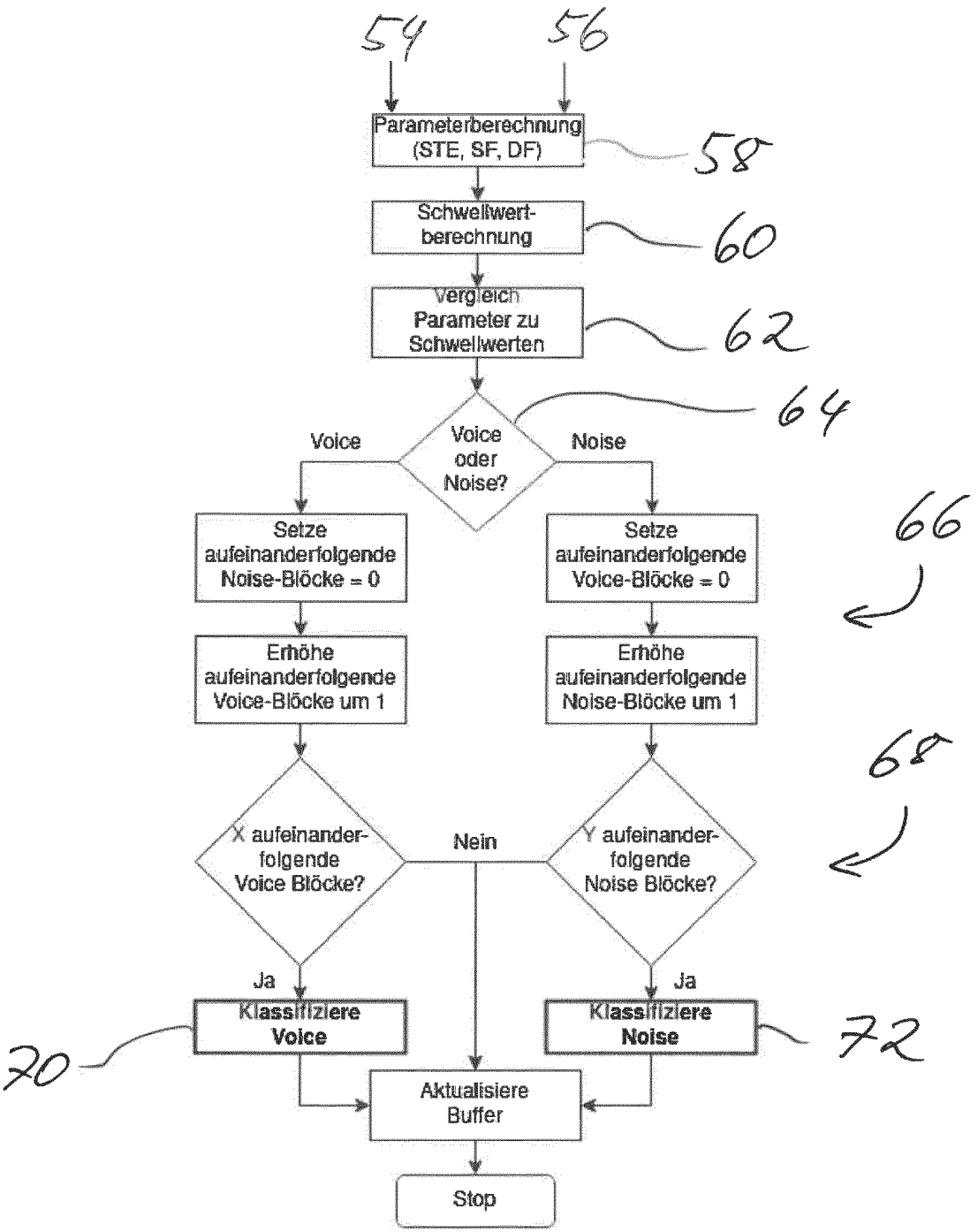
54
56
Parameterberechnung (STE, SF, DF)
58
Schwellwert-berechnung
60
Vergleich Parameter zu Schwellwerten
62
Voice oder Noise?
64
Voice
Noise
Setze aufeinanderfolgende Noise-Blöcke = 0
Setze aufeinanderfolgende Voice-Blöcke = 0
66
Erhöhe aufeinanderfolgende Voice-Blöcke um 1
Erhöhe aufeinanderfolgende Noise-Blöcke um 1
X aufeinander-folgende Voice Blöcke?
Nein
Y aufeinander-folgende Noise Blöcke?
68
Ja
Ja
Klassifiziere Voice
Klassifiziere Noise
70
72
Aktualisiere Buffer
Stop

Fig. 2

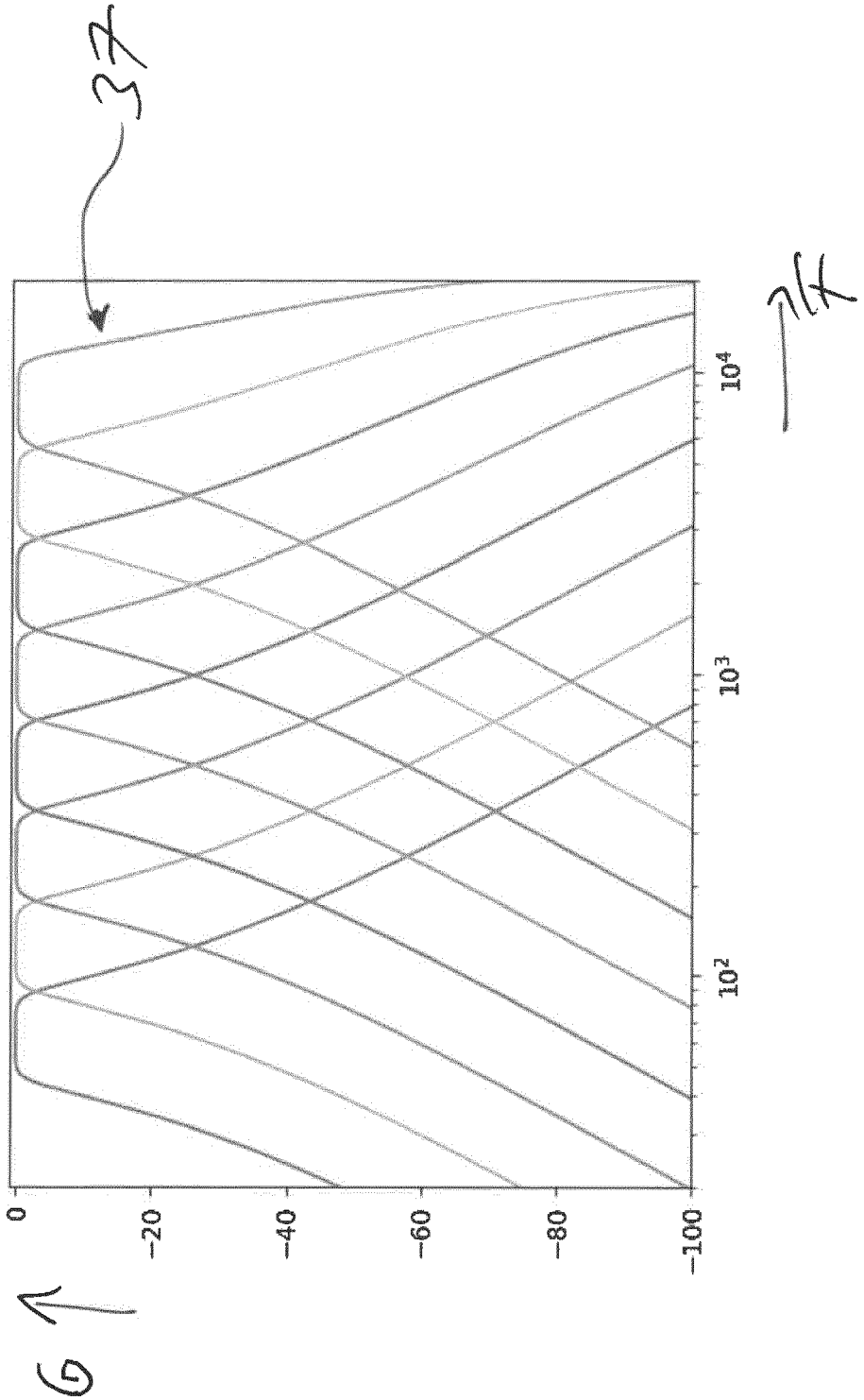

Fig. 3

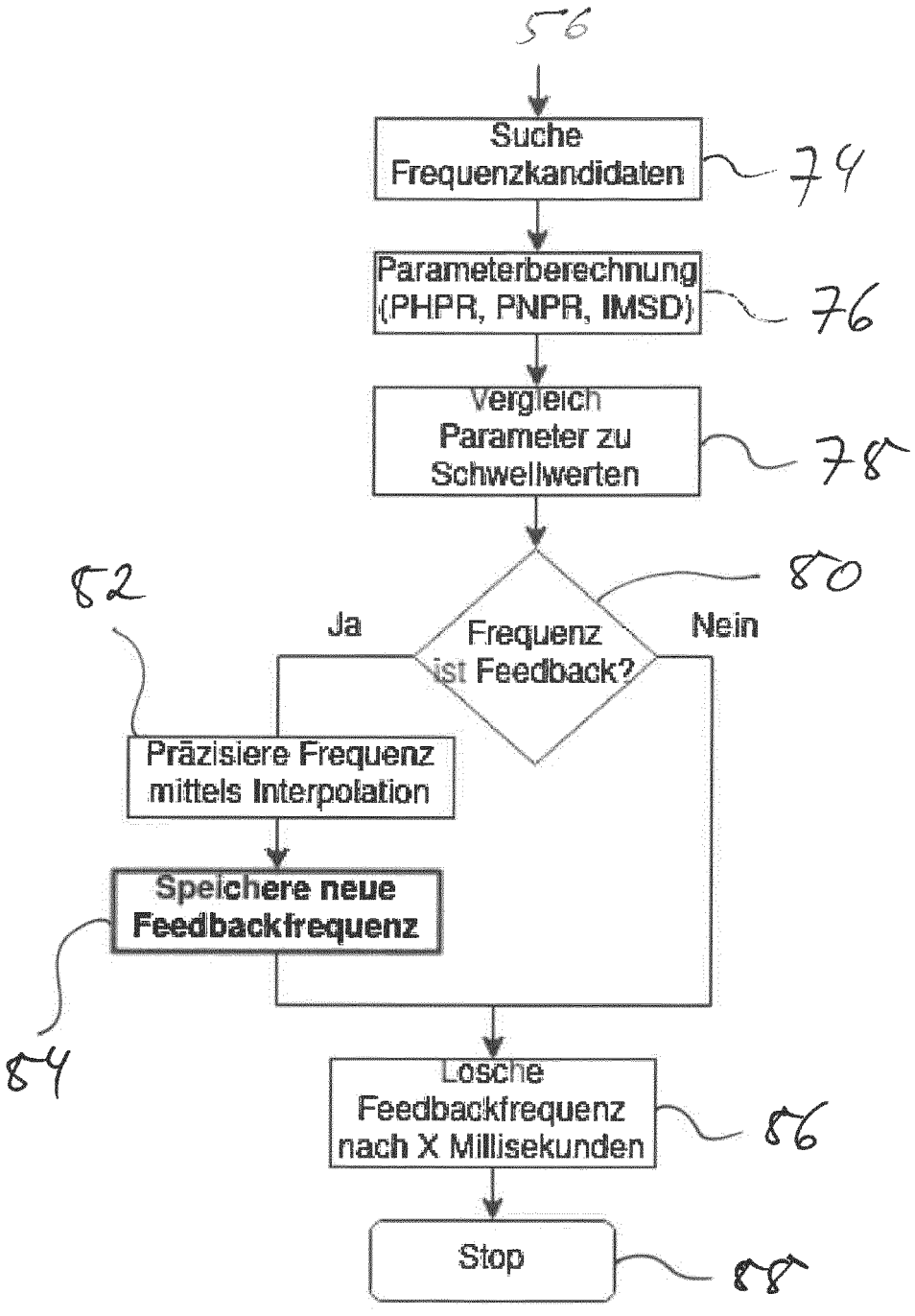
56
Suche Frequenzkandidaten
74
Parameterberechnung (PHPR, PNPR, IMSD)
76
Vergleich Parameter zu Schwellwerten
78
80
Frequenz ist Feedback?
Ja
Nein
82
Präzisiere Frequenz mittels Interpolation
Speichere neue Feedbackfrequenz
84
Losche Feedbackfrequenz nach X Millisekunden
86
Stop
88

Fig. 4

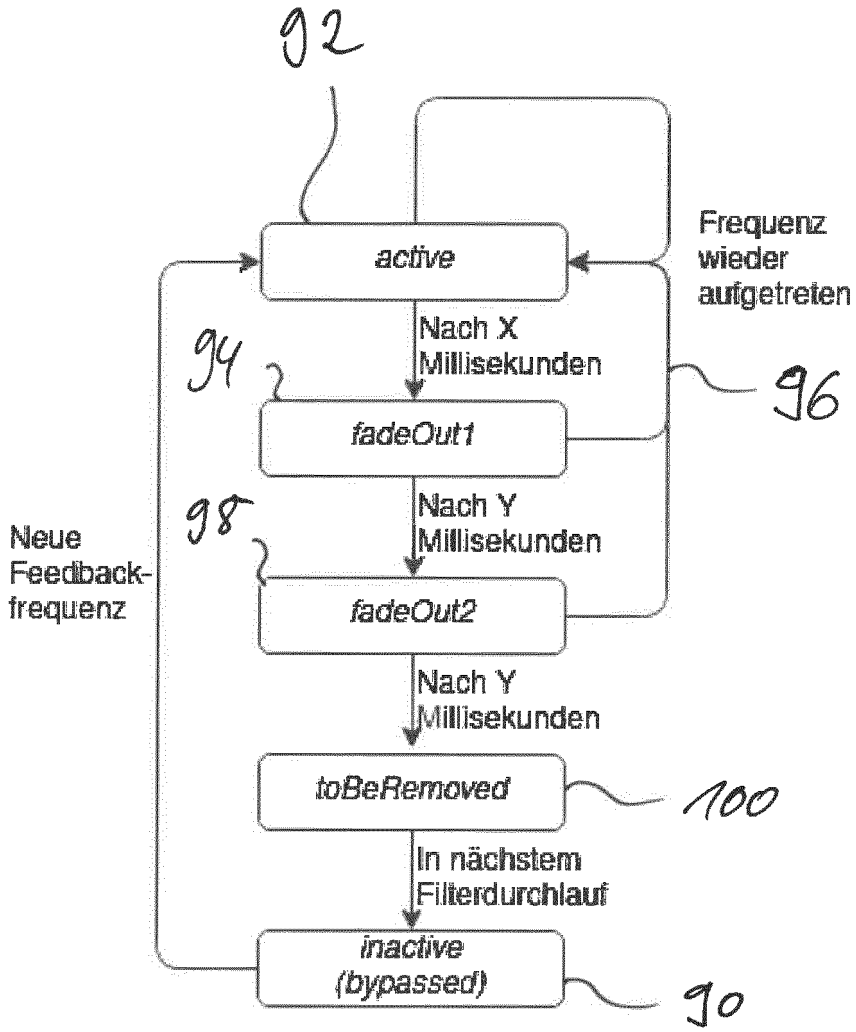
92
active
Frequenz wieder aufgetreten
Nach X Millisekunden
94
96
fadeOut1
Nach Y Millisekunden
98
Neue Feedback-frequenz
fadeOut2
Nach Y Millisekunden
toBeRemoved
100
In nächstem Filterdurchlauf
inactive (bypassed)
90

Fig. 5

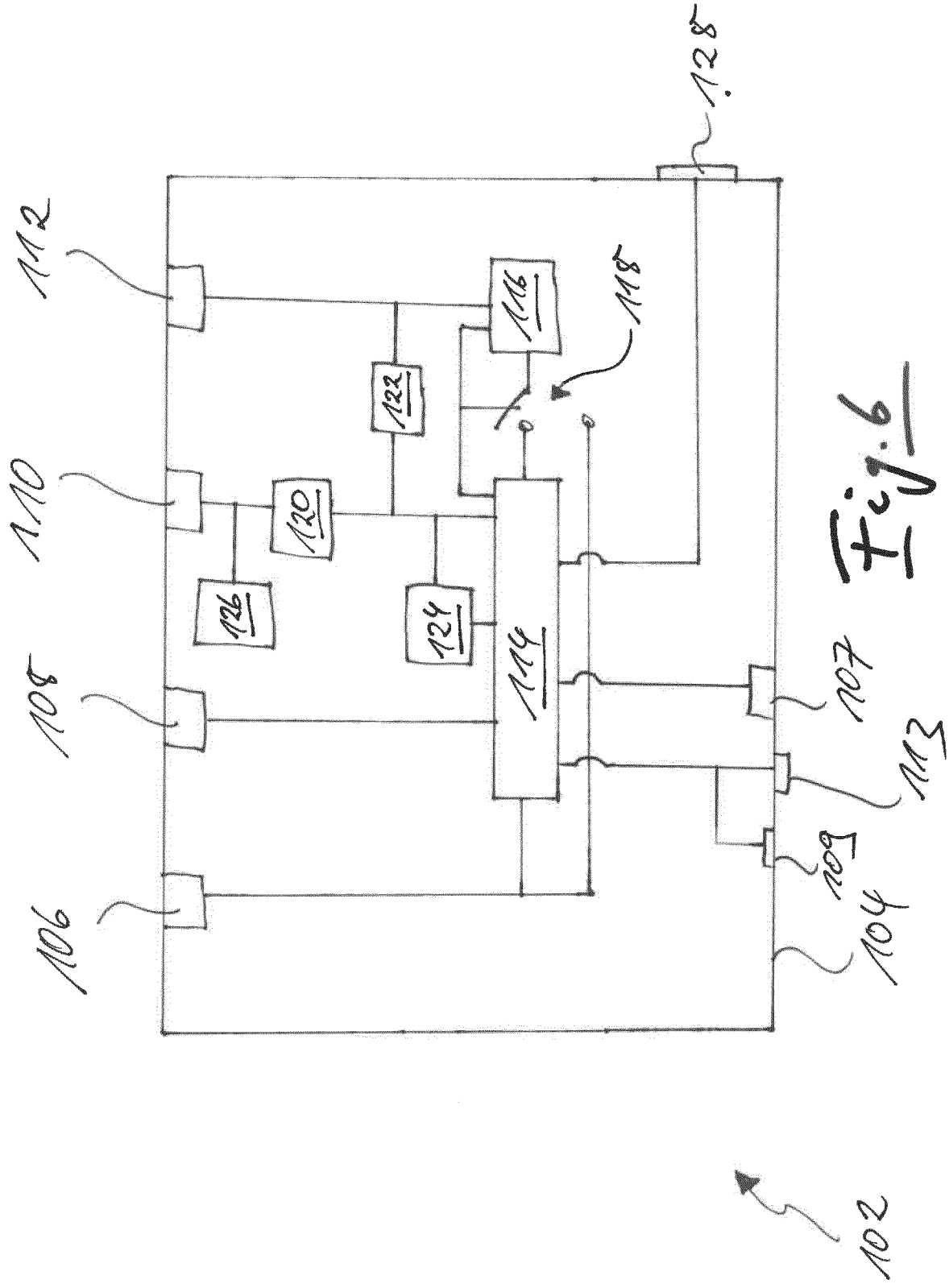
128
112
116
118
122
110
120
126
124
114
108
107
113
109
104
106
102

Fig. 6

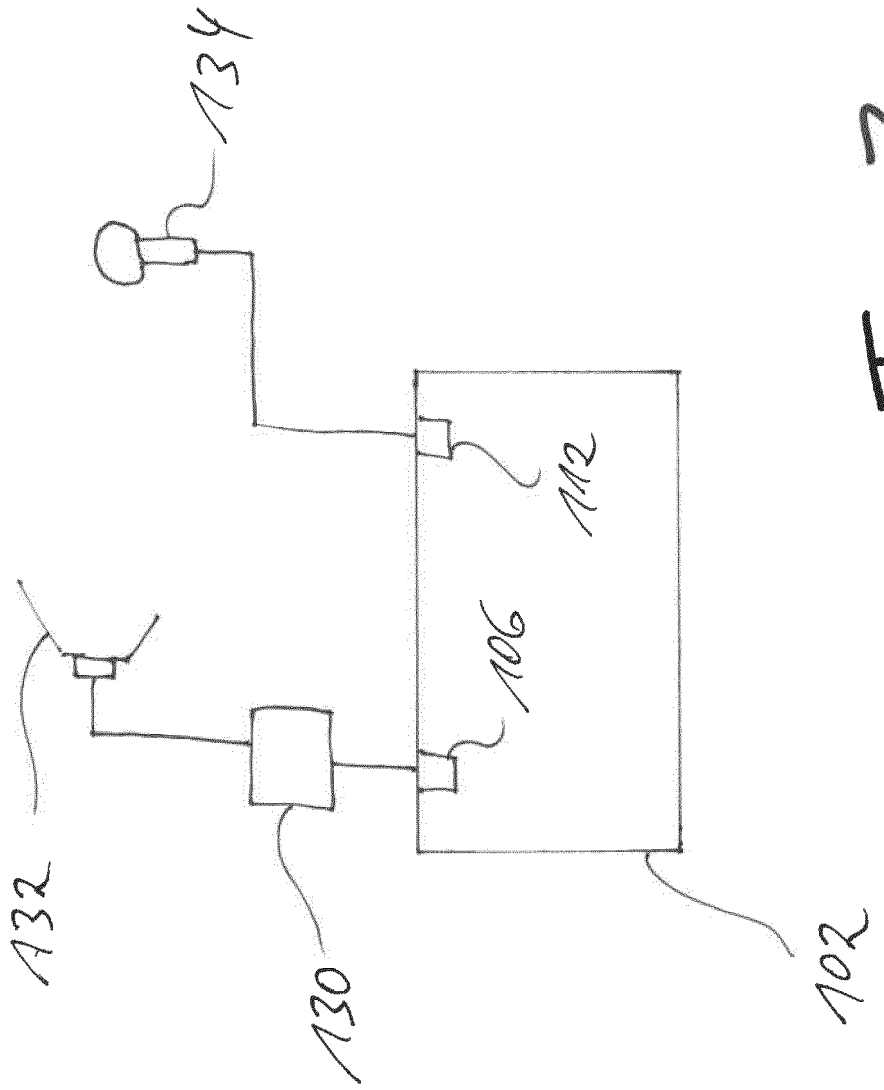

Fig. 1

IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

Diese Liste der vom Anmelder aufgeführten Dokumente wurde ausschließlich zur Information des Lesers aufgenommen und ist nicht Bestandteil des europäischen Patentdokumentes. Sie wurde mit größter Sorgfalt zusammengestellt; das EPA übernimmt jedoch keinerlei Haftung für etwaige Fehler oder Auslassungen.

In der Beschreibung aufgeführte Patentdokumente

- US 20160019905 A1 **[0006]**
- US 20170047080 A1 **[0006]**
- US 6295364 B1 **[0006]**
- US 20100121634 A1 **[0006]**
- US 20060247922 A1, Schepker **[0006]**

In der Beschreibung aufgeführte Nicht-Patentliteratur

- **VERÖFFENTLICHUNG ; T.V. WATERSCHOOT ; M. MOONEN.** Comparative Evaluation of Howeling Detection Criteria in Notch-Filter-Based Howling Supression. *Journal of the Audio Engineering Society,* 2010, vol. 58, 923-940 **[0086]**