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

(11) EP 1 647 972 A2

(12) EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:
19.04.2006 Patentblatt 2006/16

(51) Int Cl.:
G10L 21/02 *(2006.01)*

(21) Anmeldenummer: **05019316.8**

(22) Anmeldetag: **06.09.2005**

(84) Benannte Vertragsstaaten:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR
Benannte Erstreckungsstaaten:
AL BA HR MK YU

(30) Priorität: **08.10.2004 DE 102004049347**

(71) Anmelder: **Micronas GmbH**
79108 Freiburg i. Br. (DE)

(72) Erfinder:
- **Vierthaler, Matthias**
79211 Denzlingen (DE)
- **Pfister, Florian**
79346 Endingen (DE)
- **Lücking, Dieter**
79110 Freiburg (DE)
- **Müller, Stefan**
79108 Freiburg (DE)

(74) Vertreter: **Patentanwälte**
Westphal, Mussgnug & Partner
Am Riettor 5
78048 Villingen-Schwenningen (DE)

(54) Verbesserung der Verständlichkeit von Sprache enthaltenden Audiosignalen

(57) Die Erfindung bezieht sich auf eine Schaltungsanordnung für eine Verbesserung der Verständlichkeit von ggf. Sprache (px) enthaltenden Audiosignalen (i) mit einem Eingang (I) zum Eingeben eines solchen Audiosignals (i). Vorteilhaft wird die Schaltungsanordnung durch einen Sprachdetektor (SD) zum Detektieren von Sprache (px) in dem eingegebenen Audiosignal (i) und zum Bereitstellen eines Steuersignals (s) zum Steuern einer Sprachverarbeitungseinrichtung (SV) und/oder eines Sprachverarbeitungsverfahrens zum Verarbeiten des Audiosignals (i). Vorteilhaft ist entsprechend ein Verfahren zur Verarbeitung von ggf. Sprache enthaltenden Audiosignalen (i), bei dem in einem Audiosignal (i) enthaltene Sprache bzw. Sprachanteile (px) detektiert werden und abhängig von dem Ergebnis der Detektion ein Steuersignal (s) für eine Sprachverarbeitungseinrichtung (SV) und/oder ein Sprachverarbeitungsverfahren für eine Sprachverbesserung erzeugt und bereitgestellt wird.

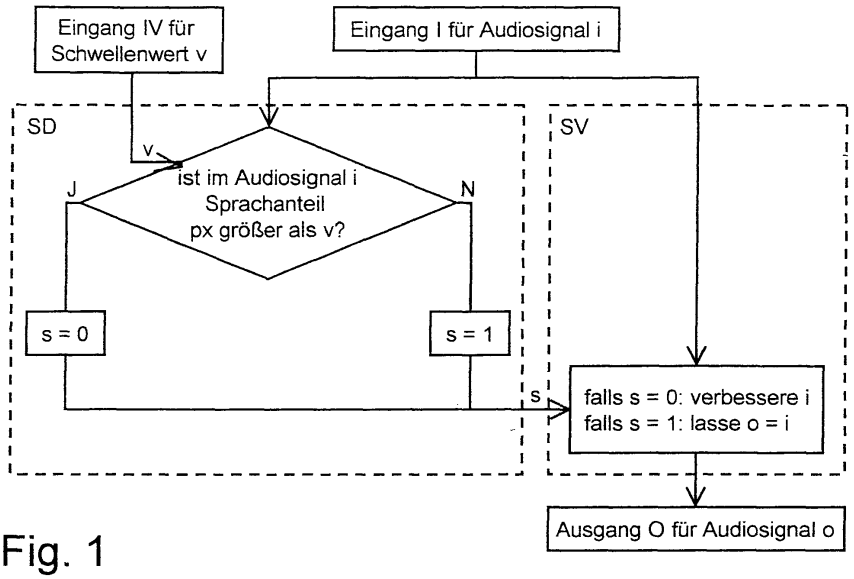

Fig. 1

EP 1 647 972 A2

Beschreibung

[0001] Die Erfindung bezieht sich auf eine Schaltungsanordnung für eine Verbesserung der Verständlichkeit von Sprache enthaltenden Audiosignalen mit den oberbegrifflichen Merkmalen des Patentanspruchs 1 bzw. auf ein Verfahren zur Verarbeitung von Sprache enthaltenden Audiosignalen.

[0002] Aus DE 101 24 699 C1 ist eine Schaltungsanordnung zur Verbesserung der Verständlichkeit von Sprache enthaltenden Audiosignalen bekannt, bei welcher Frequenz- und/oder Amplitudenanteile des Audiosignals nach vorgegebenen Parametern verändert werden. Dabei wird das Audiosignal in einer Verarbeitungsstrecke um einen vorgegebenen Faktor verstärkt sowie in einen Hochpass geführt, wobei eine Eckfrequenz des Hochpasses so regelbar ist, dass die Amplitude des Audiosignals nach der Verarbeitungsstrecke gleich oder proportional der Amplitude des Audiosignals vor der Verarbeitungsstrecke ist. Mit dieser Schaltungsanordnung soll die Grundwelle des Sprachsignals, welche relativ wenig zur Verständlichkeit der enthaltenen Sprachanteile beiträgt, aber die größte Energie besitzt, abgeschwächt werden, wobei das übrige Signalspektrum des Audiosignals entsprechend angehoben wird. Außerdem kann die Amplitude der Vokale, welche eine große Amplitude bei tiefer Frequenz aufweisen, im Übergangsbereich von einem Konsonanten, der eine kleine Amplitude bei großer Frequenz aufweist, zu einem Vokal abgesenkt werden, um ein sogenanntes "backward masking" zu verringern. Dazu wird das gesamte Signal um den Faktor angehoben. Letztendlich werden hochfrequente Anteile angehoben und die tieffrequente Grundwelle wird im gleichen Maße abgesenkt, so dass die Amplitude oder Energie des Audiosignals unverändert bleibt.

[0003] US 5,553,151 beschreibt ein "forward masking". Dabei werden schwache Konsonanten durch vorhergehende starke Vokale zeitlich überdeckt. Vorgeschlagen wird ein verhältnismäßig schneller Kompressor mit einer "attack time" von ca. 10 msec und einer "release time" von ca. 75 bis 150 msec.

[0004] Aus US 5,479,560 ist bekannt, ein Audiosignal in mehrere Frequenzbänder aufzuteilen und diejenigen Frequenzbänder mit großer Energie verhältnismäßig stark zu verstärken und die anderen abzusenken. Dies wird vorgeschlagen, weil Sprache aus einer Aneinanderreihung von Phonemen besteht. Phoneme bestehen aus einer Vielzahl von Frequenzen. Diese werden im Bereich der Resonanzfrequenzen des Mund- und Rachenraums besonders verstärkt. Ein Frequenzband mit solch einem spektralen Spitzenwert wird Formant genannt. Formants sind besonders wichtig zur Erkennung von Phonemen und somit Sprache. Ein Ansatz zur Verbesserung der Sprachverständlichkeit besteht darin, die Spitzenwerte bzw. Formants des Frequenzspektrums eines Audiosignals zu verstärken und die dazwischen liegenden Fehler abzuschwächen. Für einen erwachsenen Mann liegt die Grundfrequenz der Sprache bei etwa 60 bis 250 Hz. Die ersten vier zugeordneten Formants liegen bei 500 Hz, 1500 Hz, 2500 Hz und 3500 Hz.

[0005] Derartige Schaltungsanordnungen und Verfahrensweisen machen in einem Audiosignal enthaltene Sprache gegenüber weiteren im Audiosignal enthaltenen Komponenten verständlicher. Gleichzeitig werden aber auch nicht Sprache enthaltende Signalanteile verändert bzw. verfälscht. Nachteilhaft ist bei den Verfahren bzw. Schaltungsanordnungen auch, dass diese jeweils starr vorgegebene Sprachanteile, Frequenzanteile oder dergleichen kontinuierlich verbessern bzw. verarbeiten. Dadurch werden nicht Sprache enthaltende Signalanteile auch zu Zeiten verändert bzw. verfälscht, zu denen das Audiosignal keine Sprache bzw. Sprachanteile enthält.

[0006] Die Aufgabe der Erfindung besteht darin, eine Schaltungsanordnung bzw. ein Verfahren zur Verarbeitung von Sprache enthaltenden Audiosignalen zu verbessern.

[0007] Diese Aufgabe wird durch eine Schaltungsanordnung für eine Verbesserung der Verständlichkeit von ggf. Sprache enthaltenden Audiosignalen mit den Merkmalen des Patentanspruchs 1 bzw. durch ein Verfahren zur Verarbeitung von ggf. Sprache enthaltenden Audiosignalen mit den Merkmalen des Patentanspruchs 11 gelöst.

[0008] Vorteilhaft ist entsprechend eine Schaltungsanordnung für eine Verbesserung der Verständlichkeit von ggf. Sprache enthaltenden Audiosignalen mit einem Eingang zum Eingeben eines solchen Audiosignals. Vorteilhaft wird die Schaltungsanordnung durch einen Sprachdetektor zum Detektieren von Sprache in dem eingegebenen Audiosignal und zum Bereitstellen eines Steuersignals zum Steuern einer Sprachverarbeitungseinrichtung und/oder eines Sprachverarbeitungsverfahrens zum Verarbeiten des Audiosignals.

[0009] Vorteilhaft ist einVerfahren zur Verarbeitung von ggf. Sprache enthaltenden Audiosignalen, bei dem in einem Audiosignal enthaltene Sprache bzw. Sprachanteile detektiert werden und abhängig von dem Ergebnis der Detektion ein Steuersignal für eine Sprachverarbeitungseinrichtung und/oder ein Sprachverarbeitungsverfahren für eine Sprachverbesserung erzeugt und bereitgestellt wird.

[0010] Die Schaltungsanordnung bzw. das Verfahren sind somit als eine Vorstufe zu einer eigentlichen Signalverarbeitung zur Verbesserung der Verständlichkeit von Sprache enthaltenden Audiosignalen anzusehen. Das empfangene bzw. eingegebene Audiosignal wird demgemäß zuerst daraufhin untersucht, ob überhaupt Sprache bzw. Sprachanteile in dem Audiosignal enthalten sind. Abhängig von dem Ergebnis der Sprachdetektion wird dann ein Steuersignal ausgegeben, welches von einer eigentlichen Sprachverarbeitungseinrichtung bzw. einem eigentlichen Sprachverarbeitungsverfahren als Steuersignal verwendet wird. Dadurch wird ermöglicht, dass bei der Sprachverarbeitung zur Verbesserung der Sprachanteile im Audiosignal relativ zu anderen Signalanteilen im Audiosignal nur dann eine Verarbeitung bzw. Veränderung des Audiosignals durchgeführt wird, wenn auch tatsächlich Sprache oder Sprachanteile enthalten sind.

[0011] Entsprechend wird durch die Schaltungsanordnung bzw. durch das Verfahren ein Steuersignal bereitgestellt bzw. ausgegeben, welches für die eigentliche Sprachverbesserung z. B. als ein Triggersignal verwendet wird. Dadurch kann die Sprachverbesserung mittels Detektion bzw. Analyse eines vorherigen Audiosignals oder desgleichen, ggf. eines zeitverzögerten Audiosignals durchgeführt werden.

[0012] Die Schaltungsanordnung, welche das Steuersignal erzeugt und bereitstellt, kann als eigenständige bauliche Komponente bereitgestellt werden, kann aber auch Bestandteil einer einzigen baulichen Komponente mit der Sprachverarbeitungseinrichtung bzw. Sprachverbesserungseinrichtung sein. Insbesondere können die Schaltungsanordnung zur Detektion von Sprache und die Sprachverarbeitungseinrichtung zur Verbesserung der Sprachanteile des Audiosignals Bestandteil einer integrierten Schaltungsanordnung sein. Entsprechend können auch das Verfahren zum Detektieren von Sprache und das Sprachverarbeitungsverfahren zum Verbessern von Sprachkomponenten in dem Audiosignal getrennt voneinander durchgeführt werden. Besonders bevorzugt wird jedoch ein gemeinsames Verfahren, welches mittels technischer Komponenten einer Schaltungsanordnung oder mittels eines entsprechend ablaufenden Algorithmus in einer Berechnungseinrichtung durchgeführt wird.

[0013] Vorteilhafte Ausgestaltungen sind Gegenstand abhängiger Ansprüche.

[0014] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher der Sprachdetektor zum Detektieren von Sprachanteilen in dem Audiosignal ausgebildet und/oder gesteuert ist.

[0015] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher der Sprachdetektor eine Schwellenwert-Bestimmungseinrichtung zum Vergleichen eines Umfangs detektierter Sprachanteile mit einem Schwellenwert und zum Ausgeben des Steuersignals abhängig vom Vergleichsergebnis aufweist.

[0016] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher der Sprachdetektor einen Steuereingang zum Eingeben zumindest eines Parameters zum variablen Steuern des Detektierens hinsichtlich eines Umfangs der zu detektierenden Sprachanteile und/oder hinsichtlich eines Frequenzbereichs der zu detektierenden Sprachanteile aufweist.

[0017] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher der Sprachdetektor eine Korrelationseinrichtung zum Durchführen einer Kreuz- oder einer Autokorrelation des Audiosignals oder von Komponenten des Audiosignals aufweist.

[0018] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher der Sprachdetektor zum Verarbeiten eines mehrkomponentigen Audiosignals, insbesondere Stereo-Audiosignals oder Multikannal-Audiosignals, mit mehreren Audiosignal-Komponenten ausgebildet ist und als eine Verarbeitungseinrichtung zum Detektieren der Sprache anhand eines Vergleichs oder einer Verarbeitung der Komponenten untereinander ausgebildet oder gesteuert ist.

[0019] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher der Sprachdetektor eine Richtungsbestimmungseinrichtung zum Bestimmen einer Richtung gemeinsamer Signalanteile der verschiedenen Komponenten aufweist.

[0020] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher der Sprachdetektor einen Frequenz-Energie-Detektor zum Bestimmen einer Signalenergie in einem Sprachfrequenzbereich im Verhältnis zu einer sonstigen Signalenergie des Audiosignals aufweist.

[0021] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher der Sprachdetektor zum Ausgeben des Steuersignals abhängig von Ergebnissen sowohl des Frequenz-Energie-Detektors als auch der Korrelationseinrichtung, der Vergleichseinrichtung bzw. der Richtungsbestimmungseinrichtung ausgebildet und/oder gesteuert ist.

[0022] Bevorzugt wird insbesondere eine Schaltungsanordnung, bei welcher das Steuersignal zum Aktivieren oder Deaktivieren der Sprachverbesserungseinrichtung und/oder des Sprachverbesserungsverfahrens abhängig vom Sprachgehalt des Audiosignals ausgebildet und/oder gesteuert ist.

[0023] Bevorzugt wird insbesondere ein Verfahren, bei welchem das Steuersignal abhängig vom Umfang detektierter Sprachanteile erzeugt wird.

[0024] Bevorzugt wird insbesondere ein Verfahren, bei welchem der Umfang der detektierten Sprachanteile mit einem Schwellenwert verglichen wird.

[0025] Bevorzugt wird insbesondere ein Verfahren, bei welchem das Detektieren hinsichtlich eines Umfangs der zu detektierenden Sprachanteile und/oder hinsichtlich eines Frequenzbereichs der zu detektierenden Sprachanteile mittels variabler Parameter einstellbar durchgeführt wird.

[0026] Bevorzugt wird insbesondere ein Verfahren, bei welchem eine Kreuz- oder Autokorrelation des Audiosignals oder von Komponenten des Audiosignals durchgeführt wird.

[0027] Bevorzugt wird insbesondere ein Verfahren, bei welchem von einem mehrkomponentigen Audiosignal mit mehreren Audiosignal-Komponenten die Audiosignal-Komponenten untereinander verglichen oder miteinander verarbeitet werden zum Detektieren der Sprache. Unter Komponenten sind dabei Signalanteile aus verschiedenen Entfernungen und Richtungen und/oder Signale verschiedener Kanäle zu verstehen.

[0028] Bevorzugt wird insbesondere ein Verfahren, bei welchem die Audiosignal-Komponenten hinsichtlich gemeinsamer Sprachanteile in den verschiedenen der Audiosignal-Komponenten verglichen bzw. verarbeitet werden, insbesondere zum Bestimmen einer Richtung der gemeinsamen Signalanteile verglichen bzw. verarbeitet werden. Anhand

unterschiedlicher Eintreffzeiten auf beispielsweise dem rechten und dem linken Kanal eines Stereosignals sowie anhand spezifischer Dämpfungen spezieller Frequenzen kann die Entfernung und Richtung des Sprachanteils bestimmt werden. Dadurch ist eine Anwendung der Sprachverbesserung insbesondere nur auf Sprachanteile anwendbar, welche als von einer Person, die dicht am Mikrophon steht, stammend erkannt werden. Signalanteile bzw. Sprachanteile von entfernteren Personen können dadurch ignoriert werden, so dass eine Sprachverbesserung nur dann aktiviert wird, wenn tatsächlich eine nahestehende Person spricht.

[0029] Bevorzugt wird insbesondere ein Verfahren, bei welchem eine Energie des Audiosignals in einem Sprachfrequenzbereich im Verhältnis zu einer sonstigen Signalenergie des Audiosignals bestimmt wird. Abgestellt wird dabei somit auf die Energie von Frequenzanteilen, welche für gesprochene Sprache typisch sind. Neben einer individuellen Abstimmung auf bedarfsweise beispielsweise eine männliche, eine weibliche oder eine kindliche Sprache als Kriterium für den zu wählenden Sprachfrequenzbereich wird der Vergleich der entsprechenden Energie vorzugsweise mit der Energie von den übrigen Signalanteilen des Audiosignals mit anderen Frequenzen oder mit dem Energiegehalt des gesamten Audiosignalanteils durchgeführt. Insbesondere Sprache von entfernt stehend sprechenden Personen, welche somit im Zweifelsfall für den Hörer nicht von Interesse ist, kann erkannt werden und zu einer Deaktivierung der Sprachverbesserung führen, wenn keine nahe stehende Person spricht.

[0030] Bevorzugt wird insbesondere ein Verfahren, bei welchem das Steuersignal zum Aktivieren oder Deaktivieren der Sprachverbesserungseinrichtung und/oder des Sprachverbesserungsverfahrens bereitgestellt wird.

[0031] Bevorzugt wird insbesondere eine Schaltungsanordnung und/oder ein Verfahren, wobei ein Frequenzgang mittels eines FIR- oder eines IIR-Filters (FIR: Finite-Impulse-Response, IIR: Infinite-Impulse-Response) bestimmt wird.

[0032] Bevorzugt wird insbesondere eine Schaltungsanordnung und/oder ein Verfahren, wobei Signalanteile des Audiosignals durch eine Matrix getrennt werden.

[0033] Bevorzugt wird insbesondere eine Schaltungsanordnung und/oder ein Verfahren, wobei Matrixkoeffizienten für eine Matrix über eine vom Sprachanteil abhängige Funktion bestimmt werden. Dabei ist die Funktion linear und stetig. Alternativ oder zusätzlich besitzt die Funktion eine Hysterese.

[0034] Die Signalanteile mit Sprachanteilen des Audiosignals können hinsichtlich verschiedener Kriterien analysiert und detektiert werden. Neben einer beispielsweise Mindestdauer, über welche Sprache als Sprachanteil erfasst wird, kann z.B. als Signalanteil auch auf die Frequenz erfassbarer Sprache und/oder die Richtung einer Sprachquelle erfasster Sprache abgestellt werden. Die Begriffe Signalanteile und Sprachanteile sind daher allgemein und nicht beschränkend auszulegen.

[0035] Die Erfindung wird nachfolgend anhand der Zeichnung näher erläutert. Es zeigen:

Fig. 1 schematisch Verfahrensschritte bzw. Komponenten eines Verfahrens bzw. einer Schaltungsanordnung zum Verarbeiten eines Audiosignals zur Detektion von darin enthaltener Sprache;

Fig. 2 eine beispielhafte Schaltungsanordnung gemäß einer ersten Ausführungsform zur Anwendung einer Korrelation auf Sprachanteile verschiedener Signalkomponenten;

Fig. 3 eine weiter beispielhafte Schaltungsanordnung zur Veranschaulichung einer Bestimmung von Energie in einem Sprachfrequenzbereich;

Fig. 4 eine beispielhafte Schaltungsanordnung zur Darstellung einer Matrixberechnung vor einer Durchführung einer Sprachverbesserung des Audiosignals; und

Fig. 5 ein Diagramm zur Veranschaulichung von Kriterien zur Festlegung eines Schwellenwerts.

[0036] Fig. 1 zeigt beispielhaft schematisch den Ablauf eines Verfahrens zum Detektieren von Sprache und/oder Sprachanteilen px in einem Audiosignal i zur optionalen nachfolgenden oder parallelen Sprachverbesserung der Sprache bzw. der Sprachanteile px, sofern solche detektiert werden, in dem Audiosignal i. Über einen Eingang I einer Schaltungsanordnung für eine Verbesserung der Verständlichkeit von ggf. Sprache oder Sprachanteilen px enthaltenden Audiosignalen i wird ein Audiosignal i eingegeben. Bei dem Audiosignal i kann es sich je nach Anwendungsfall um ein einkanaliges Monosignal handeln. Bevorzugt werden jedoch mehrkomponentige Audiosignale i einer Stereo-Audiosignalquelle oder dergleichen, d.h. ein Stereo-Audiosignal, ein 3D-Stereo-Audiosignal mit zusätzlicher Zentralkomponente oder ein Surround-Audiosignal mit derzeit üblicherweise fünf Komponenten für Audiosignal-Komponenten von rechts, links, der Mitte sowie von z. B. zwei entfernten Quellen rechts und links.

[0037] Das Audiosignal i wird einer ersten baulichen oder logischen Komponente, welche einen Sprachdetektor SD ausbildet, zugeführt. In dem Sprachdetektor SD wird untersucht, ob in dem Audiosignal i Sprache bzw. ein Sprachanteil px enthalten ist. Gemäß bevorzugter Ausführungsformen wird dabei geprüft, ob detektierte Sprache bzw. Sprachanteile px größer sind als ein entsprechend vorgegebener Schwellenwert v. Optional sind Detektionsparameter, insbesondere

der Schwellenwert v bedarfsweise anpassbar. Diesbezüglich weist die dargestellte Anordnung einen Eingang IV zum Eingeben des Schwellenwerts v auf.

[0038] Ergibt die Detektion, dass ein ausreichender Sprachanteil px in dem Audiosignal i enthalten ist, so wird ein Steuersignal beispielsweise auf den Wert 0 gesetzt. Andernfalls wird das Steuersignal auf beispielsweise den Wert 1 gesetzt. Das Steuersignal s wird von dem Sprachdetektor SD zur weiteren Verwendung durch eine Sprachverarbeitungseinrichtung bzw. ein Sprachverarbeitungsverfahren ausgegeben.

[0039] Falls das Steuersignal s einen Sprachanteil px signalisiert, d. h. falls im vorliegenden Fall s = 0 gilt, wird die Sprache bzw. Sprachanteile px verbessernde Sprachverarbeitung aktiviert. Das momentan in die Sprachverarbeitung eingegebene Audiosignal i wird entsprechend für sich bekannter Verfahren bzw. mit einer ansonsten für sich bekannten Schaltungsanordnung verbessert. An einem Ausgang O wird entsprechend ein hinsichtlich der Sprachanteile verbessertes Audiosignal o ausgegeben.

[0040] Falls bei dem Detektionsschritt kein ausreichender Sprachanteil px erfasst wird, d.h., falls s = 1 gilt, wird das in die Sprachverarbeitung SV eingegebene Audiosignal i belassen, d.h., unverändert als Audiosignal o ausgegeben.

[0041] Sofern durch die Sprachdetektion eine zeitliche Verzögerung des an der Sprachverarbeitung anliegenden Steuersignals s relativ zu dem momentan anliegenden Audiosignal i vorliegt, kann optional eine Verzögerung des in die Schaltungsanordnung bzw. das Verfahren eingegebenen Audiosignals i entsprechend der zeitlichen Verzögerung bei der Sprachdetektion vorgenommen werden.

[0042] Ermöglicht wird somit eine Schaltungsanordnung bzw. ein Verfahren oder Algorithmus, welche eine Sprachverbesserung nur auf Teile des Audiosignals anwenden lassen, welche tatsächlich Sprache enthalten oder welche tatsächlich einen bestimmten Sprachanteil im Audiosignal enthalten. Durch die Sprachdetektion wird somit Sprache detektiert bzw. vom restlichen Signal getrennt.

[0043] In der Realität wird sich Sprache von anderen Signalanteilen eines Audiosignals mathematisch nicht genau trennen lassen. Ziel ist somit, einen möglichst guten Schätzwert zu liefern. Sofern Algorithmen bzw. Schaltungsanordnungen nachfolgend aufgeführter Ausführungsformen sich durch entsprechende andere Signalanteile in die Irre führen lassen, wird gemäß erster Versuche trotzdem eine vorteilhafte Verbesserung eines ausgegebenen Audiosignals erzielt. Vorteilhaft ist dazu, darauf zu achten, dass das Audiosignal i auch bei einer Fehldetektion im Sprachdetektor SD nicht zu sehr verfälscht wird.

[0044] Fig. 2 zeigt eine erste Ausführungsvariante eines Sprachdetektors SD. Der Eingang besteht aus zwei individuellen Eingängen für jeweils eine Audiosignal-Komponente bzw. einen Audiosignal-Kanal L', R' eines Stereo-Audiosignals. Die beiden Audiosignal-Komponenten R', L' werden jeweils einem Bandpassfilter BP zur Bandbegrenzung zugeführt. Die Ausgangssignale der beiden Bandpassfilter BP werden einer Korrelationseinrichtung CR zum Durchführen einer Kreuzkorrelation zugeführt. Jedes der beiden von den Bandpassfiltern BP ausgegebenen Signale wird jeweils in einem Multiplikator M mit sich selber multipliziert, d. h. quadriert, und dann einem Additionsglied A zugeführt. Nach der Addition erfolgt optional in einem weiteren Multiplikator M* eine Multiplikation mit dem Faktor 0,5, um die Amplitude zu reduzieren. Das Ausgangssignal i der gegebenenfalls multiplizierten Additionswerte wird einem ersten bzw. zweiten Tiefpassfilter TP zugeführt.

[0045] Außerdem wird jedes der Ausgangssignale der beiden Bandpassfilter BP einer eigentlichen Schaltung zur Durchführung der Korrelation unter Einsatz insbesondere eines weiteren Multiplikators M zugeführt. Das davon ausgegebene Korrelationssignal L,* R' wird einem zweiten Tiefpassfilter TP zugeführt.

[0046] Die Ausgangssignale b, a des ersten Tiefpassfilters TP und des zweiten Tiefpassfilters TP werden einem Divisionsglied DIV zur Division des Ausgangssignals b des ersten Tiefpassfilters TP von dem Ausgangssignal a des zweiten Tiefpassfilters TP zugeführt. Das Divisionsergebnis des Divisionsglieds DIV wird als Steuersignals bzw. als Vorstufe D1 für das Steuersignal s bereitgestellt.

[0047] Mit einer solchen Schaltungsanordnung oder einem entsprechenden Verarbeitungsverfahren wird eine Kreuzkorrelation durchgeführt. Ein übliches Stereo-Audiosignal L', R' setzt sich als Audiosignal i in der Regel aus mehreren Audiosignal-Komponenten R, L, C, S zusammen. Im Fall eines Multikannal-Audiosignals können diese Komponenten auch separat bereitgestellt werden.

[0048] Im Fall eines Stereo-Audiosignals L', R' sind die beiden Audiosignal-Kanäle L', R' beschreibbar durch

$$a:\ L' = L + C + S \text{ bzw.}$$

$$b:\ R' = R + C - S,$$

wobei L für eine linke Signalkomponente steht, C für eine zentral von vorne kommende Signalkomponente steht, S für

eine Surround-Signalkomponente, d.h, ein rückwärtiges Signalund R für eine rechte Signalkomponente steht.

[0049] Sprache bzw. Sprachanteile px befinden sich hauptsächlich auf dem zentralen Kanal bzw. in der Zentralkomponente C. Diese Tatsache kann benutzt werden, um den Anteil von Sprache bzw. Sprachanteilen px zum restlichen Signalgehalt des Audiosignals i zu detektieren. Bestimmt werden kann die enthaltene Sprache bzw. der enthaltene Sprachanteil px im Verhältnis zu den restlichen Signalanteilen des Audiosignals i gemäß

$$px = 2*RMS(C) / ((RMS/L') + RMS(R'))$$

mit RMS als der zeitlich gemittelten Amplitude.

[0050] Durch eine Kreuzkorrelation lässt sich der Anteil der Zentralkomponente C bestimmen durch

$$L'*R' = 2*L*R + L*C + R*C - L*S + R*S + C*C - S*S.$$

[0051] Im zeitlichen Mittel werden für DC-freie Signale, d. h. für Signalkomponenten ohne einen Gleichspannungsanteil alle nicht korrelierten Produkte zu 0. Damit kann als Kriterium für das von dem Sprachdetektor SD ausgegebene Signal D1 gelten:

$$\begin{aligned} D1 &= 2*TP(L'*R') / (L'*L' + R'*R') \\ &= 2*TP(C*C - S*S) / TP (L'*L' + R'*R'). \end{aligned}$$

[0052] Damit ergibt sich für das Ausgangssignal D1, welches als Vorstufe zu dem Steuersignal s oder direkt als Steuersignal s verwendet werden kann, als Wert D1 = 1, falls das Audiosignal i ausschließlich aus einer Zentralkomponente C besteht. D1 = 0 ergibt sich, falls das Audiosignal i ausschließlich aus unkorrelierten rechten und linken Signalkomponenten L, R besteht. D = -1 ergibt sich, falls das Audiosignal i ausschließlich aus Surround-Komponenten S besteht. Bei einer Mischung der verschiedenen Komponenten, wie sie bei einem realen Signal gegeben ist, ergeben sich Werte für D1 zwischen -1 und +1. Je näher das Ausgangssignal bzw. der Ausgangswert D1 bei +1 liegt, desto zentral-lastiger ist das Audiosignal i bzw. L', R', so dass auf einen entsprechend großen Sprachanteil px geschlossen werden kann.

[0053] Die Zeitkonstante des Tiefpassfilters TP kann im Bereich von ca. 100 ms liegen, falls eine sehr schnelle Reaktion auf sich ändernde Signalkomponenten gewünscht ist. Die Zeitkonstante kann jedoch bis zu mehreren Minuten verlängert werden, falls eine sehr langsame Reaktion des Sprachdetektors SD gewünscht ist. Die Zeitkonstante des Tiefpassfilters ist daher ein vorteilhafterweise variabler Parameter. Vor der Durchführung eines Detektionsalgorithmus werden DC-Anteile zweckmäßigerweise mittels eines entsprechenden Filters, insbesondere DC-Kerbfilters (DC-Notch) herausgefiltert. Die weitere Bandbegrenzung ist optional.

[0054] Fig. 3 zeigt eine weitere beispielhafte Ausführungsform eines Sprachdetektors SD. Nachfolgend werden unter Bezug auf die Beschreibung zu Fig. 2 lediglich diejenigen Komponenten beschrieben, welche sich gegenüber der Schaltungsanordnung bzw. Verfahrensweise gemäß Fig. 2 unterscheiden.

[0055] Die beiden Ausgangssignale der beiden Bandpassfilter BP werden jeweils einer Energiebestimmungskomponente ABS eines Frequenz-Energie-Detektors Ef zur Bestimmung des Energiegehalts zugeführt. Sprache hat die größte Energie bei Frequenzen zwischen 100 Hz und 4 kHz. Zur Bestimmung des Sprachanteils px kann entsprechend der Anteil der Energie im Sprachfrequenzbereich f1...f2 im Verhältnis zur Gesamtenergie des Audiosignals i bzw. L', R' bestimmt werden.

[0056] Die Energiebestimmungskomponente ABS für die beiden Eingangssignale bzw. Eingangssignalkanäle ist im einfachsten Fall ein Glied, welches an seinem Ausgang den Betragswert eines am Eingang anliegenden Wertes ausgibt.

[0057] Entsprechend werden die Ausgangswerte der Energiebestimmungskomponenten ABS miteinander mittels eines Additionsglieds A addiert und wie im Fall der Ausführungsform von Fig. 2 einem ersten Tiefpassfilter TP zugeführt. Außerdem werden die beiden Ausgangssignale der Bandpassfilter BP, welche eine Bandbegrenzung durchgeführt haben, einem weiteren Additionsglied A zugeführt. Dessen Ausgangssignal wird einem Bandpassfilter BP* zugeführt, welcher entsprechend nur diejenigen Signalanteile durchlässt, welche im Sprachfrequenzbereich f1...f2 liegen. Dieses Bandpassgefilterte Signal wird einem zweiten Tiefpassfilter TP zugeführt. Abschließend erfolgt eine Division des Ausgangssignals b des ersten Tiefpassfilters TP durch das Ausgangssignal a des zweiten Tiefpassfilters TP in einem

Divisionsglied DIV, um als Ausgangswert bzw. Ausgangssignal D2 das Steuersignal oder eine Vorstufe für das Steuersignal bereitzustellen.

[0058] Berechnet werden kann das Ausgangssignal D2 durch

$$D2=2*RMS(BP(f1...f2)(L'+R'))/(RMS\ (L')+RMS(R').$$

Dabei gilt, je näher der Ausgangswert bzw. das Ausgangssignal D2 sich dem Wert 1 nähert, desto mehr Energie ist im Sprachfrequenzbereich vorhanden, so dass auf einen großen Sprachanteil px geschlossen werden kann. Die einleitende Bandbegrenzung des Eingangssignals L', R' ist wiederum optional.

[0059] Besonders bevorzugt wird als Steuersignal s oder als Vorstufe dazu ein Ausgangswert bzw. Ausgangssignal D3 verwendet, welches beide Verfahren bzw. Schaltungsanordnungen der beschriebenen Ausführungsformen gemäß Fig. 2 und Fig. 3 berücksichtig. Als Kriterium kann beispielsweise gelten

$$D3 = D1*D2.$$

[0060] Damit wird Sprache bzw. ein Sprachanteil px dann erkannt, wenn mehr Energie in der Zentralkomponente C des Audiosignals vorhanden ist und mehr Energie im Sprachfrequenzbereich vorhanden ist.

[0061] Optional kann den dargestellten Schaltungsanordnungen bzw. Verfahrensweisen zur Bereitstellung des Steuersignals s noch eine Stufe nachgeschaltet werden, in welcher ein Schwellenwert v festgelegt wird, der von dem Ausgangssignal D1, D2, D3 der beschriebenen Anordnungen bzw. Verfahren zu Überschreiten ist, um das Steuersignal s in einen aktiven Zustand zu schalten.

[0062] Bei einer parallelen oder nachfolgenden Sprachsignalverarbeitung des Audiosignals i besteht das Ziel darin, möglichst viele Signalanteile, die Sprache bzw. Sprachanteile px enthalten, durch einen Sprachverbesserungsalgorithmus zu leiten und die restlichen Signalanteile unverändert zu lassen, wie dies auch anhand Fig. 1 beschrieben ist. Dies wird vorteilhaft durch eine Matrix gelöst, wie dies anhand Fig. 4 skizziert ist.

[0063] Matrixkoeffizienten k1, k2,..., k6 werden abhängig von dem bestimmten Sprachanteil px bzw. abhängig von dem vom Sprachdetektor SD ausgegebenen Ausgangswert bzw. Ausgangssignal D1, D2 bestimmt bzw. werden als Funktion px = F(D1, D2) ermittelt.

Der eigentliche Sprachverbesserungsalgorithmus oder eine eigentliche Sprachverbesserungseinrichtung kann in für sich bekannter Art und Weise bereitgestellt werden. Beispielsweise kann eine in DE 101 24 699 C1, auf welche voll umfänglich Bezug genommen wird, beschriebene einfache Frequenzgangkorrektur durchgeführt werden. Einsetzbar sind aber auch beliebige andere Algorithmen und Einrichtungen zur Verbesserung der Sprachverständlichkeit.

[0064] Bei der in Fig. 4 dargestellten Matrixberechnung werden die Eingangskomponenten bzw. Eingangskanäle L', R' des Audiosignals i jeweils mit drei Faktoren k1, k3, k5 bzw. k2, k4, k6 multipliziert und Additionsgliedern zugeführt. Dem ersten Additionsglied A wird das Signal des ersten Kanals L' multipliziert mit dem ersten Koeffizienten k1 und das Signal des zweiten Kanals R' multipliziert mit dem zweiten Koeffizienten k2 zur Addition angelegt. Dem zweiten Additionsglied A werden das Signal des ersten Kanals L' multipliziert mit dem dritten Koeffizienten k3 und das Signal des zweiten Kanals R' multipliziert mit dem vierten Koeffizienten k4 zur Addition angelegt. Dem dritten Additionsglied A werden das Signal des ersten Kanals L' multipliziert mit dem fünften Koeffizienten k5 und das Signal des zweiten Kanals R' multipliziert mit dem sechsten Koeffizienten k6 zur Addition angelegt. Der Ausgangswert des zweiten Additionsglieds A wird einer Sprachverbesserungsschaltung VS oder einem Sprachverbesserungsverfahren bzw. Algorithmus zugeführt. Dessen Ausgangsergebnis wird mittels weiterer Additionsglieder A dem Ausgangswert bzw. Ausgangssignal des ersten Additionsglieds A zur Bereitstellung eines ersten Ausgangskanals LE und einem Ausgangswert bzw. Ausgangssignal des dritten Additionsglieds A mittels eines weiteren Additionsglied A zum Bereitstellen eines zweiten Ausgangskanals RE aufaddiert.

[0065] Für die Bestimmung der Koeffizienten wird beispielsweise berücksichtigt, dass der Sprachanteil px durch die beschriebenen Verfahren durch einen Wertebereich von insbesondere $0 \leq P \leq 1$ und als Funktion der Bestimmten Sprachanteile mit px = F(D1,D2,D3) bestimmbar ist. Gemäß einer einfachen Variante können die Koeffizienten festgelegt werden gemäß

$$k1 = k6 = 1 - px/2,$$

$$k2 = K5 = -px/2$$

und

$$k3 = k4 = px/2.$$

[0066] Die beiden letztendlich ausgegebenen Signalkanäle bzw. Komponenten LE, RE entsprechen den verarbeiteten Signalen, welche dem Ausgang O für das verarbeitete Audiosignal o zugeführt werden.

[0067] Fig. 5 stellt beispielhaft Funktion F(D1, D2=0, D3=0) dar. Im Fall der ersten dargestellten Funktion F = F1(D1) reagiert die Schaltungsanordnung schon auf einen geringen detektierten Sprachanteil. Die Wahrscheinlichkeit einer Fehldetektion ist für kleine Werte von D1 relativ hoch. Allerdings ist durch den stetigen Verlauf der ersten Funktion F1 (D1) die Auswirkung des Sprachalgorithmus bei kleinem D1 auf das Audiosignal relativ gering, so dass eine Beeinträchtigung des Audiosignals kaum wahrgenommen wird.

[0068] Im Fall einer zweiten Funktion F2(D1) bleibt das Audiosignal vollkommen unbeeinträchtigt bis zu einem Schwellenwert v = Ps2. Danach sind die Auswirkungen auf das Audiosignal bei Änderungen des Werts von P1 umso größer.

[0069] Im Fall einer dritten Funktion F = F3(D1) wird der Algorithmus beim Überschreiten eines bestimmten Schwellenwerts v = Ps31 eingeschaltet und beim Unterschreiten eines anderen, niedrigeren Schwellenwerts v=Ps32 ausgeschaltet. Durch den Einbau einer solchen Hysterese wird ein ständiges Umschalten im Übergangsbereich verhindert.

Patentansprüche

1. Schaltungsanordnung für eine Verbesserung der Verständlichkeit von ggf. Sprache (px) enthaltenden Audiosignalen (i) mit

   - einem Eingang (I) zum Eingeben eines solchen Audiosignals (i),

   gekennzeichnet durch

   - einen Sprachdetektor (SD) zum Detektieren von Sprache (px) in dem eingegebenen Audiosignal (i) und zum Bereitstellen eines Steuersignals (s) zum Steuern einer Sprachverarbeitungseinrichtung (SV) und/oder eines Sprachverarbeitungsverfahrens zum Verarbeiten des Audiosignals (i).

2. Schaltungsanordnung nach Anspruch 1, bei welcher der Sprachdetektor (SD) zum Detektieren von Sprachanteilen (px) in dem Audiosignal (i) ausgebildet und/oder gesteuert ist.

3. Schaltungsanordnung nach Anspruch 1 oder 2, bei welcher der Sprachdetektor (SD) eine Schwellenwert-Bestimmungseinrichtung zum Vergleichen eines Umfangs detektierter Sprachanteile mit einem Schwellenwert (v) und zum Ausgeben des Steuersignals (s) abhängig vom Vergleichsergebnis aufweist.

4. Schaltungsanordnung nach Anspruch 3, bei welcher der Sprachdetektor (SD) einen Steuereingang (IV) zum Eingeben zumindest eines Parameters (v) zum variablen Steuern des Detektierens hinsichtlich eines Umfangs der zu detektierenden Sprachanteile (px) und/oder hinsichtlich eines Frequenzbereichs der zu detektierenden Sprachanteile (px) aufweist.

5. Schaltungsanordnung nach einem vorstehenden Anspruch, bei welcher der Sprachdetektor (SD) eine Korrelationseinrichtung (CR) zum Durchführen einer Kreuz- oder einer Autokorrelation des Audiosignals oder von Komponenten des Audiosignals aufweist.

6. Schaltungsanordnung nach einem vorstehenden Anspruch, bei welcher der Sprachdetektor (SD)

   - zum Verarbeiten eines mehrkomponentigen Audiosignals (i), insbesondere Stereo-Audiosignals (L', R'), 3D-Stereo-Audiosignals (L, R, C) und/oder Surround-Audiosignals (L, R, C, S), mit mehreren Audiosignal-Komponenten (L, R, C, S) ausgebildet ist und
   - eine Verarbeitungseinrichtung (CR) zum Detektieren der Sprache anhand eines Vergleichs oder einer Verarbeitung der Komponenten (L, R, C, S) untereinander aufweist.

7. Schaltungsanordnung nach Anspruch 6, bei welcher der Sprachdetektor (SD) eine Richtungs- und/oder Entfernungsbestimmungseinrichtung zum Bestimmen einer Richtung und/oder Entfernung gemeinsamer Signalanteile der verschiedenen Komponenten (L, R, C, S) aufweist.

8. Schaltungsanordnung nach einem vorstehenden Anspruch, bei welcher der Sprachdetektor (SD) einen Frequenz-Energie-Detektor (Ef) zum Bestimmen einer Signalenergie in einem Sprachfrequenzbereich im Verhältnis zu einer sonstigen Signalenergie des Audiosignals (i) aufweist.

9. Schaltungsanordnung nach Anspruch 8 und einem der Ansprüche 5 bis 7, bei welcher de Sprachdetektor (SD) zum Ausgeben des Steuersignals (s) abhängig von Ergebnissen sowohl des Frequenz-Energie-Detektors (Ef) als auch der Korrelationseinrichtung (CR), der Vergleichseinrichtung bzw. der Richtungs- und/oder Entfernungsbestimmungseinrichtung ausgebildet und/oder gesteuert ist.

10. Schaltungsanordnung nach einem vorstehenden Anspruch, bei welcher das Steuersignal (s) zum Aktivieren oder Deaktivieren der Sprachverbesserungseinrichtung (SV) und/oder des Sprachverbesserungsverfahrens abhängig vom Sprachgehalt des Audiosignals (i) ausgebildet und/oder gesteuert ist.

11. Verfahren zur Verarbeitung von ggf. Sprache enthaltenden Audiosignalen (i), bei dem

    - in einem Audiosignal (i) enthaltene Sprache bzw. Sprachanteile (px) detektiert werden und
    - abhängig von dem Ergebnis der Detektion ein Steuersignal (s) für eine Sprachverarbeitungseinrichtung (SV) und/oder ein Sprachverarbeitungsverfahren für eine Sprachverbesserung erzeugt und bereitgestellt wird.

12. Verfahren nach Anspruch 11, bei welchem das Steuersignal (s) abhängig vom Umfang detektierter Sprachanteile (px) erzeugt wird.

13. Verfahren nach Anspruch 12, bei welchem der Umfang der detektierten Sprachanteile (px) mit einem Schwellenwert (v) verglichen wird.

14. Verfahren nach einem der Ansprüche 11 bis 13, bei welchem das Detektieren hinsichtlich eines Umfangs der zu detektierenden Sprachanteile und/oder hinsichtlich eines Frequenzbereichs der zu detektierenden Sprachanteile (px) mittels variabler Parameter (v) einstellbar durchgeführt wird.

15. Verfahren nach einem der Ansprüche 11 bis 14, bei welchem eine Kreuz- oder Autokorrelation des Audiosignals (i) oder von Komponenten (R, L, C, S) des Audiosignals (i) durchgeführt wird.

16. Verfahren nach einem der Ansprüche 11 bis 15, bei welchem von einem mehrkomponentigen Audiosignal mit mehreren Audiosignal-Komponenten (R, L, C, S) die Audiosignal-Komponenten untereinander verglichen oder miteinander verarbeitet werden zum Detektieren der Sprache.

17. Verfahren nach Anspruch 16, bei welchem die Audiosignal-Komponenten (R, L, C, S) hinsichtlich gemeinsamer Sprachanteile in den verschiedenen der Audiosignal-Komponenten verglichen bzw. verarbeitet werden, insbesondere zum Bestimmen einer Richtung und/oder Entfernung der gemeinsamen Signalanteile verglichen bzw. verarbeitet werden.

18. Verfahren nach einem der Ansprüche 11 bis 17, bei welchem eine Energie des Audiosignals (i) in einem Sprachfrequenzbereich (f1, ..., f2) im Verhältnis zu einer sonstigen Signalenergie des Audiosignals (i) bestimmt wird.

19. Verfahren nach einem der Ansprüche 11 bis 18, bei welchem das Steuersignal (s) zum Aktivieren oder Deaktivieren der Sprachverbesserungseinrichtung (SV) und/oder des Sprachverbesserungsverfahrens bereitgestellt wird.

20. Schaltungsanordnung nach einem der Ansprüche 1 bis 10 und/oder Verfahren nach einem der Ansprüche 11 bis 19, wobei ein Frequenzgang mittels eines FIR- oder eines IIR-Filters (FIR: Finite-Impulse-Response, IIR: Infinite-Impulse-Response) bestimmt wird.

21. Schaltungsanordnung nach einem der Ansprüche 1 bis 10 und/oder Verfahren nach einem der Ansprüche 11 bis 19, wobei Signalanteile des Audiosignals durch eine Matrix getrennt werden.

22. Schaltungsanordnung nach einem der Ansprüche 1 bis 10 und/oder Verfahren nach einem der Ansprüche 11 bis 19, wobei Matrixkoeffizienten für eine Matrix (MX) über eine vom Sprachanteil (px) abhängige Funktion (P = F(px)) bestimmt werden.

23. Schaltungsanordnung und/oder Verfahren nach Anspruch 22, wobei die Funktion (P = F(px)) linear und stetig ist.

24. Schaltungsanordnung und/oder Verfahren nach Anspruch 22, wobei die Funktion (P = F(px)) eine Hysterese besitzt.

25. Sprachverbesserungs-Schaltungsanordnung oder -verfahren mit einer Schaltungsanordnung und/oder einem Verfahren nach einem der vorstehenden Ansprüche.

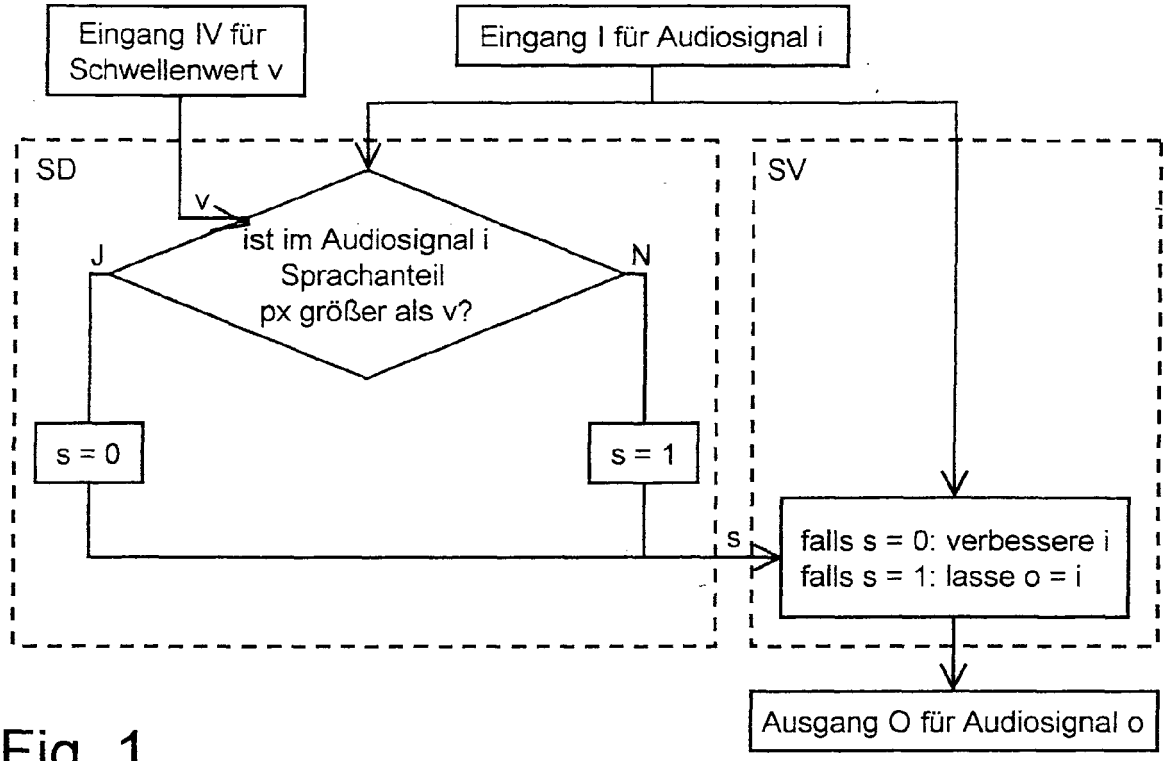
Eingang IV für Schwellenwert v
Eingang I für Audiosignal i
SD
SV
v
ist im Audiosignal i Sprachanteil px größer als v?
J
N
s = 0
s = 1
s
falls s = 0: verbessere i
falls s = 1: lasse o = i
Ausgang O für Audiosignal o

Fig. 1

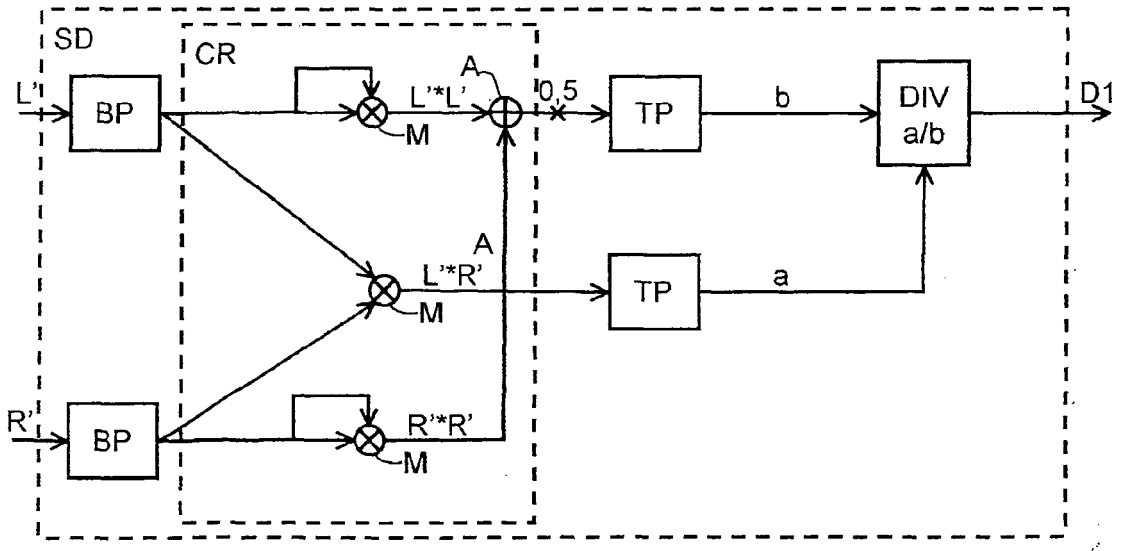
SD
CR
L'
BP
L'*L'
M
A
0,5
TP
b
DIV a/b
D1
L'*R'
M
A
TP
a
R'
BP
R'*R'
M

Fig. 2

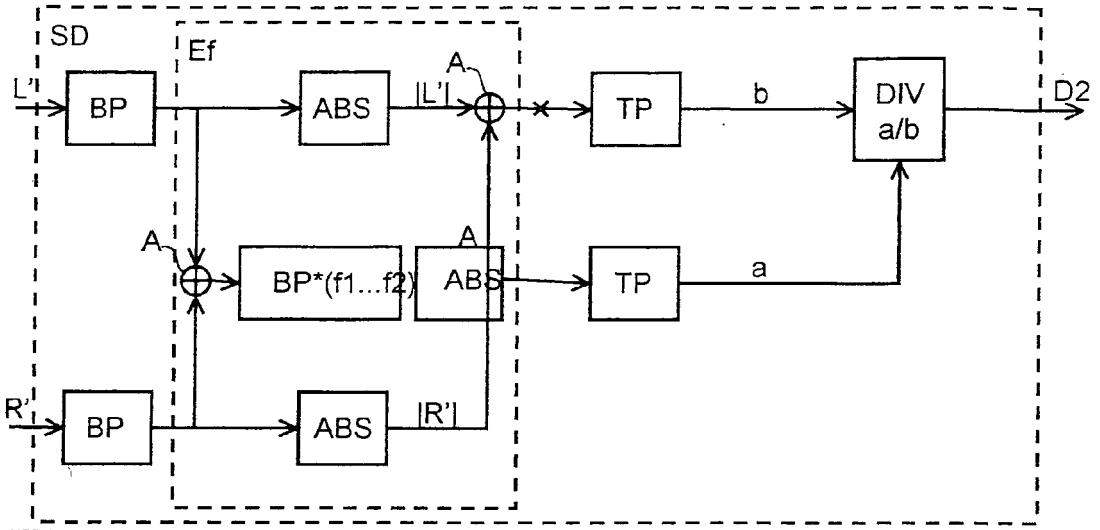

Fig. 3

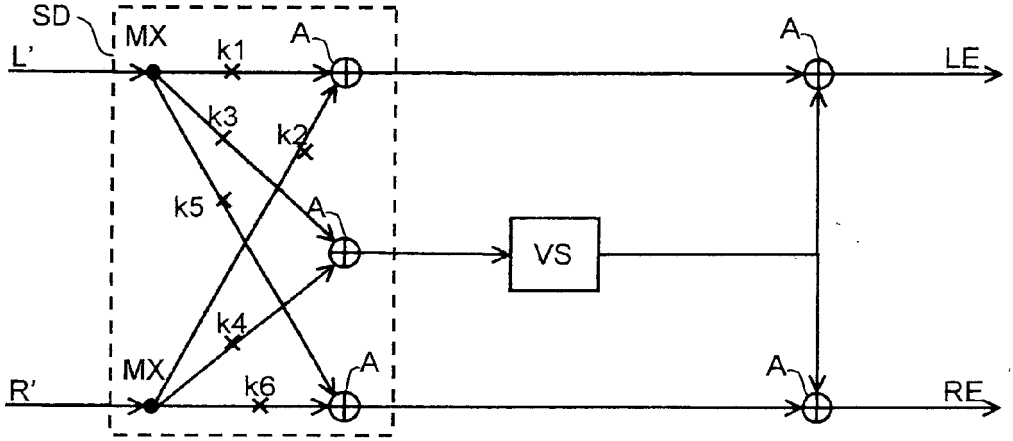

Fig. 4

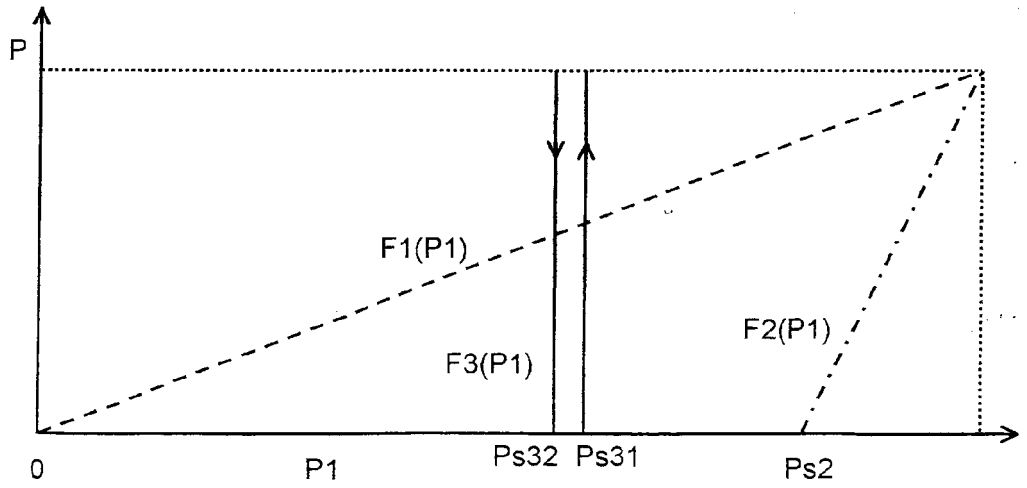

Fig. 5