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

(11) **EP 2 158 588 B1**

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
13.10.2010 Patentblatt 2010/41

(21) Anmeldenummer: **08784249.8**

(22) Anmeldetag: **25.06.2008**

(51) Int Cl.:
G10L 21/00 *(2006.01)* ***G06T 5/00*** *(2006.01)*

(86) Internationale Anmeldenummer:
PCT/DE2008/001047

(87) Internationale Veröffentlichungsnummer:
WO 2009/000255 (31.12.2008 Gazette 2009/01)

(54) **SPEKTRALGLÄTTUNGSVERFAHREN VON VERRAUSCHTEN SIGNALEN**

SPECTRAL SMOOTHING METHOD FOR NOISY SIGNALS

PROCÉDÉ DE FILTRAGE SPECTRAL DE SIGNAUX PARASITÉS

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

(30) Priorität: **27.06.2007 DE 102007030209**

(43) Veröffentlichungstag der Anmeldung:
03.03.2010 Patentblatt 2010/09

(73) Patentinhaber:
- **Ruhr-Universität Bochum**
  44801 Bochum (DE)
- **Siemens Audiologische Technik GmbH**
  91058 Erlangen (DE)

(72) Erfinder:
- **MARTIN, Rainer**
  44803 Bochum (DE)
- **GERKMANN, Timo**
  44791 Bochum (DE)
- **BREITHAUPT, Colin**
  80538 München (DE)

(74) Vertreter: **Maier, Daniel Oliver et al**
Siemens AG
Postfach 22 16 34
80506 München (DE)

(56) Entgegenhaltungen:
WO-A-01/73761 US-A- 5 365 592

- **HARALD GUSTAFSSON ET AL: "Spectral Subtraction Using Reduced Delay Convolution and Adaptive Averaging" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, Bd. 9, Nr. 8, 1. November 2001 (2001-11-01), XP011054141 ISSN: 1063-6676 in der Anmeldung erwähnt**
- **BREITHAUPT C ET AL: "Cepstral Smoothing of Spectral Filter Gains for Speech Enhancement Without Musical Noise" IEEE SIGNAL PROCESSING LETTERS, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, Bd. 14, Nr. 12, 1. Dezember 2007 (2007-12-01), Seiten 1036-1039, XP011194896 ISSN: 1070-9908**

Anmerkung: Innerhalb von neun Monaten nach Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents im Europäischen Patentblatt kann jedermann nach Maßgabe der Ausführungsordnung beim Europäischen Patentamt gegen dieses Patent Einspruch einlegen. Der Einspruch gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist. (Art. 99(1) Europäisches Patentübereinkommen).

EP 2 158 588 B1

Beschreibung

[0001] Die Erfindung betrifft ein Glättungsverfahren zur Unterdrückung von fluktuierenden Artefakten bei der Störgeräuschreduktion.

[0002] In der digitalen Sprachsignalübertragung ist die Störgeräuschunterdrückung ein wichtiger Aspekt. Die mit einem Mikrofon erfassten und anschließend digitalisierten Audiosignale enthalten neben dem Nutzsignal (Figur 1) noch Umgebungsgeräusche, die dem Nutzsignal überlagert sind (Figur 2). So werden z.B. bei Freisprechanlagen in Fahrzeugen neben den Sprachsignalen noch Motoren- und Windgeräusche erfasst, bei Hörhilfen sind es ständig wechselnde Umgebungsgeräusche wie Verkehrsgeräusche oder im Hintergrund sprechende Personen wie etwa in einem Restaurant. Dadurch ist ein Verstehen des Sprachsignals nur mit erhöhter Anstrengung möglich. Die Störgeräuschreduktion zielt dementsprechend auf eine Erleichterung des Sprachverstehens ab. Daher darf eine Verringerung des Störgeräuschs auch das Sprachsignal nicht hörbar verzerren.

[0003] Für die Störgeräuschreduktion ist die Spektraldarstellung eine günstige Repräsentation des Signals. Hierbei wird das Signal in Frequenzen aufgeschlüsselt dargestellt. Eine praktische Realisierung der Spektraldarstellung sind Kurzzeitspektren, die durch eine Zerteilung des Signals in kurze Rahmen entstehen (Figur 3), die getrennt voneinander einer Spektraltransformation unterzogen werden (Figur 4). Ein Signalrahmen kann dabei bei einer Abtastrate von f_s = 8000 Hz beispielsweise M = 256 aufeinanderfolgende digitale Signalabtastwerte umfassen, was dann einer Dauer von 32 ms entspricht. Ein transformierter Rahmen besteht dann aus M sogenannten Frequenzbins. Der quadrierte Amplitudenwert eines Frequenzbins korrespondiert zur Energie, die das Signal in dem schmalen Frequenzausschnitt von ca. 31 Hz Bandbreite enthält, der vom jeweiligen Frequenzbin repräsentiert wird. Aufgrund der Symmetrieeigenschaften der Spektraltransformation sind von den M Frequenzbins nur M/2+1, also im vorangegangenen Beispiel 129 Bins relevant für die Signaldarstellung. Mit 129 relevanten Bins und 31 Hz Bandbreite pro Bin wird insgesamt ein spektrales Band von 0 Hz bis ca. 4000 Hz abgedeckt. Dies reicht aus, um viele Sprachlaute mit hinreichender spektraler Auflösung zu beschreiben. Eine andere gängige Bandbreite ist 8000 Hz, die durch eine höhere Abtastrate und somit mehr Frequenzbins bei gleicher Rahmendauer erreicht werden kann. In einem Kurzzeitspektrum sind die Frequenzbins mit μ indiziert. Der Index für Rahmen ist λ. Die Amplituden des Kurzzeitspektrums eines Rahmens λ werden hier allgemein als spektrale Größe $G_\mu(\lambda)$ notiert. Ein vollständiges Kurzzeitspektrum bestehend aus den M Frequenzbins eines Rahmens ergibt sich aus den Amplituden $G_\mu(\lambda)$ der Indizes μ = 0 bis μ = M-1, also μ = 0 ... M - 1. Für reelle Zeitsignale erfüllen Kurzzeitspektren die Symmetriebedingung $G_\mu(\lambda) = G_{M-\mu}(\lambda)$. Eine gängige Form der Präsentation der Kurzzeitspektren sind sogenannte Spektrogramme, die durch Aneinanderreihung zeitlich aufeinanderfolgender Kurzzeitspektren gebildet werden (vgl. beispielhaft Figuren 6 bis 9).

[0004] Vorteil der Spektraldarstellung ist, dass die wesentliche Sprachenergie in einer relativ geringen Anzahl von Frequenzbins konzentriert vorliegt (Figuren 4 und 6), während im Zeitsignal alle digitalen Abtastwerte gleich relevant sind (Figur 3). Die Signalenergie der Störung ist in den meisten Fällen auf eine größere Anzahl von Frequenzbins verteilt. Da die Frequenzbins unterschiedlich viel Sprachenergie enthalten, ist es möglich, das Rauschen in jenen Bins zu unterdrücken, die nur wenig Sprachenergie enthalten. Je schmalbandiger die Frequenzbins sind, desto besser gelingt diese Trennung.

[0005] Für die Störgeräuschreduktion wird eine spektrale Gewichtungsfunktion geschätzt, die nach unterschiedlichen Optimierungskriterien berechnet werden kann. Sie ergibt niedrige Werte oder Null in Frequenzbins, in denen hauptsächlich Störung vorliegt, und Werte nahe oder gleich Eins für Bins, in denen Sprachenergie dominiert (Figur 5). Die Gewichtungsfunktion wird im Allgemeinen für jeden Signalrahmen in jedem Frequenzbin neu geschätzt. Die Gesamtheit der Gewichtungswerte aller Frequenzbins eines Rahmens wird hier auch als "Kurzzeitspektrum der Gewichtungsfunktion" oder einfach als "Gewichtungsfunktion" bezeichnet.

[0006] Eine Multiplikation der Gewichtungsfunktion mit dem Kurzzeitspektrum des verrauschten Signals ergibt das gefilterte Spektrum, in dem die Amplituden der Frequenzbins, in denen Störung dominiert, stark verringert sind, während Sprachkomponenten nahezu unbeeinflusst bleiben (Figuren 8 und 9).

[0007] Schätzfehler bei der Berechnung der spektralen Gewichtungsfunktion, sogenannte Fluktuationen, ergeben gelegentlich zu hohe Gewichtungswerte für Frequenzbins, die hauptsächlich Störung enthalten (Figur 8). Dies geschieht unabhängig von spektral benachbarten oder zeitlich vorangegangenen Werten. Fluktuationen kommen auch schon in spektralen Zwischengrößen wie z.B. der Schätzung des Signal-zu-Rausch-Verhältnisses (signal-to-noise ratio, SNR) vor. Nach Multiplikation der schätzfehlerbehafteten Gewichtungsfunktion mit dem verrauschten Kurzzeitspektrum enthält das gefilterte Spektrum einzelne Frequenzbins, die hauptsächlich Störung enthalten und dennoch relativ hohe Amplituden aufweisen. Diese Bins heißen Ausreißer. Bei der Synthese eines Zeitsignals aus den gefilterten Kurzzeitspektren sind die vereinzelten Ausreißer als tonale Artefakte (musical noise) zu hören, die wegen ihrer Tonalität als besonders störend empfunden werden (Figuren 10 und 11). Ein einzelnes tonales Artefakt hat die Dauer eines Signalrahmens und seine Frequenz wird durch den Frequenzbin bestimmt, in dem der Ausreißer vorkam.

[0008] Zur Unterdrückung von Fluktuationen in der Gewichtungsfunktion oder in spektralen Zwischengrößen bzw. zur Unterdrückung von Ausreißern im gefilterten Spektrum können diese spektralen Größen durch ein Mittelungsverfahren

geglättet und somit von überhöhten Werten befreit werden. Spektrale Größen mehrerer spektral benachbarter oder zeitlich aufeinanderfolgender Frequenzbins werden dabei zu einem Mittelwert verrechnet, so dass die Amplitude einzelner Ausreißer relativiert wird. Eine Glättung ist über der Frequenz [1: Tim Fingscheidt, Christophe Beaugeant and Suhadi Suhadi. Overcoming the statistical independence assumption w.r.t. frequency in speech enhancement. Proceedings, IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 1:1081-1084, 2005], entlang der Zeit [2: Harald Gustafsson, Sven Erik Nordholm and Ingvar Claesson. Spectral subtraction using reduced delay convolution and adaptive averaging. IEEE Transactions on Speech and Audio Processing, 9(8):799-807, November 2001] oder als Kombination aus zeitlicher und spektraler Mittelung [3: Zenton Goh, Kah-Chye Tan and B.T.G. Tan. Postprocessing method for suppressing musical noise generated by spectral subtraction. IEEE Transactions on Speech and Audio Processing, 6 (3):287-292, May 1998] bekannt. Nachteil einer Glättung über der Frequenz ist, dass bei einer Verrechnung mehrerer Frequenzbins die spektrale Auflösung verringert wird, also die Unterscheidung zwischen Sprachbins und Rauschbins schwieriger wird. Eine zeitliche Glättung durch Zusammenfassung aufeinanderfolgender Werte eines Bins verringert die zeitliche Dynamik der spektralen Werte, also ihr Vermögen, schnellen zeitlichen Änderungen der Sprache zu folgen. Eine Verzerrung des Sprachsignals ist die Folge (clipping). Außerdem kann ein mit dem Sprachsignal korreliertes, irritierendes Restrauschen hörbar werden (noise shaping). Diese Glättungsverfahren im Spektralbereich müssen deshalb im Allgemeinen aufwändig an das Sprachsignal adaptiert werden.

[0009] Eine weitere bekannte Form der Glättung einzelner Kurzzeitspektren über der Frequenz ist ein als "Liftering" bekanntes Verfahren [4: Andrzej Czyzewski. Multitask noisy speech enhancement system. http://sound.eti.pg.gda.pl/denoise/main.html, 2004], [5: Francois Thibault. High-level control of singing voice timbre transformations. http://www.music.mcgill.ca/thibault/Thesis/node43.html, 2004]. Hierbei wird das Kurzzeitspektrum eines Rahmens λ zunächst in den sogenannten Cepstralbereich transformiert. Die cepstrale Repräsentation der spektralen Amplituden $G_\mu(\lambda)$ berechnet sich zu

$$G_{\mu'}^{\text{cepst}}(\lambda) = \text{IDFT}\left\{\log\left(G_\mu(\lambda)\right)\right\},\ \mu' = 0\ldots(M-1), \mu = 0\ldots(M-1) \qquad (1)$$

mit IDFT{·} der inversen diskreten Fourier-Transformation (DFT) einer Folge von Werten der Länge M. Diese Transformation resultiert in M Transformationskoeffizienten $G_{\mu'}^{\text{cepst}}(\lambda)$, den sogenannten cepstralen Bins mit Index μ'. Das Cepstrum besteht nach Gleichung (1) prinzipiell aus einer nicht-linearen Abbildung, nämlich der Logarithmierung, einer betragsmäßig vorliegenden spektralen Größe und einer anschließenden Transformation dieses logarithmierten Betragsspektrums mit einer Transformation. Der Vorteil einer cepstralen Repräsentation der Amplituden (Figur 14) ist, dass Sprache nicht mehr kammartig über die Frequenz verteilt ist (Figuren 4 und 6), sondern die wesentliche Information über das Sprachsignal in den cepstralen Bins mit kleinem Index repräsentiert ist. Außerdem wird wesentliche Sprachinformation noch in dem verhältnismäßig leicht zu detektierenden cepstralen Bin mit höherem Index repräsentiert, der die sogenannte Pitch-Frequenz (Sprachgrundfrequenz) des Sprechers repräsentiert.

[0010] Ein geglättetes Kurzzeitspektrum kann berechnet werden, indem cepstrale Bins mit relativ kleinen Beträgen zu Null gesetzt werden und anschließend das veränderte Cepstrum wieder in ein Kurzzeitspektrum rücktransformiert wird. Da allerdings starke Fluktuationen bzw. Ausreißer zu entsprechend hohen Amplituden im Cepstrum führen, können diese Artefakte durch diese Verfahren nicht detektiert und unterdrückt werden.

[0011] Alternativ zum Liftering gibt es noch das Verfahen nach [6: Petre Stoica and Niclas Sandgren. Smoothed nonparametric spectral estimation via cepstrum thresholding. IEEE Signal Processing Magazine, pages 34-45, November 2006]. Hier werden nach einem Kriterium ausgesuchte cepstrale Bins nicht zu Null gesetzt, sondern zu einem Wert, der für die Schätzung von Langzeitspektren stationärer Signale aus Kurzzeitspektren optimal ist. Diese Form der Schätzung von Signalspektren bringt für stark nicht-stationäre Signale wie Sprache generell keine Vorteile.

[0012] Das Dokument US 5,365,592 offenbart ein weiteres Liftering ähnlichen Verfahren.

[0013] Hiervon ausgehend liegt der Erfindung die Aufgabe zugrunde, für die Rauschreduktion ein Glättungsverfahren zur Unterdrückung von Fluktuationen in der Gewichtungsfunktion oder in spektralen Zwischengrößen bzw. von Ausreißern in gefilterten Kurzzeitspektren aufzuzeigen, das weder die Frequenzauflösung der Kurzzeitspektren verringert noch die zeitliche Dynamik des Sprachsignals beeinträchtigt.

[0014] Die Lösung dieser Aufgabe besteht in einem Glättungsverfahren mit den Maßnahmen von Patentanspruch 1. Vorteilhafte Weiterbildungen sind Gegenstand der Unteransprüche.

[0015] Das erfindungsgemäße Glättungsverfahren umfasst folgende Schritte:

- Bereitstellen von Kurzzeitspektren einer Folge von digitalen Abtastwerten umfassenden Signalrahmen,
- Transformieren jedes Kurzzeitspektrums durch eine Hintransformation, die das Kurzzeitspektrum durch Transformationskoeffizienten beschreibt, welche das Kurzzeitspektrum in seine groben und seine feinen Strukturen unterteilt

repräsentieren,

- Glätten der Transformationskoeffizienten jeweils gleicher Koeffizientenindizes durch Kombination von wenigstens zwei aufeinanderfolgenden transformierten Kurzzeitspektren und
- Transformieren der geglätteten Transformationskoeffizienten in geglättete Kurzzeitspektren durch eine Rücktransformation.

[0016] Das erfindungsgemäße Glättungsverfahren bedient sich einer Transformation wie des Cepstrums, um ein breitbandiges Sprachsignal mit möglichst wenig Transformationskoeffizienten in seiner wesentlichen Struktur zu beschreiben. Anders als in bekannten Verfahren werden die Transformationskoeffizienten aber nicht unabhängig voneinander zu Null gesetzt, wenn sie einen Schwellwert unterschreiten. Es werden stattdessen die Werte von Transformationskoeffizienten aus mindestens zwei aufeinanderfolgenden Rahmen durch eine Glättung über die Zeit miteinander verrechnet. Hierbei wird der Grad der Glättung davon abhängig gemacht, inwieweit die durch den Koeffizienten repräsentierte spektrale Struktur entscheidend für die Beschreibung des Nutzsignals ist. Der Grad der zeitlichen Glättung eines Koeffizienten hängt daher beispielsweise davon ab, ob ein Transformationskoeffizient viel Sprachenergie enthält oder wenig. Dies ist im Cepstrum oder ähnlichen Transformationen leichter zu bestimmen als im Kurzzeitspektrum. So kann beispielsweise angenommen werden, dass die ersten vier cepstralen Koeffizienten mit Indizes $\mu' = 0 \ldots 3$ und zusätzlich der Koeffizient mit maximalem Betrag und Index μ' größer 16 und kleiner 160 bei $f_s = 8000$ Hz (Pitch) Sprache repräsentieren. Koeffizienten mit viel Sprachinformation werden nur soweit geglättet, dass ihre zeitliche Dynamik nicht geringer wird als bei einem unverrauschten Sprachsignal. Gegebenenfalls werden diese Koeffizienten gar nicht geglättet. Sprachverzerrungen werden so verhindert. Da spektrale Fluktuationen und Ausreißer eine kurzzeitige Änderung in der Feinstruktur eines Kurzzeitspektrums darstellen, bilden sie sich im transformierten Kurzzeitspektrum als kurzzeitige Änderung derjenigen Transformationskoeffizienten ab, die die Feinstruktur des Kurzzeitspektrums repräsentieren. Da diese Transformationskoeffizienten bei unverrauschter Sprache eine relativ geringe zeitliche Änderungsrate haben, können eben diese Koeffizienten stärker geglättet werden. Eine verstärkte zeitliche Glättung wirkt somit der Ausbildung von Ausreißern entgegen, ohne die Struktur der Sprache zu beeinflussen. Das Glättungsverfahren resultiert somit nicht in einer verringerten spektralen Auflösung für Sprachlaute. Die Änderung der Feinstruktur des Kurzzeitspektrums bei aufeinanderfolgenden Rahmen ist derart verzögert, dass nur schmalbandige spektrale Änderungen mit Zeitkonstanten kleiner als derjenigen von unverrauschter Sprache unterbunden werden.

[0017] Aus der geglätteten Größe, notiert als $G^{cepst}_{\mu',smooth}(\lambda)$, kann durch eine Rücktransformation wieder eine spektrale Darstellung des geglätteten Kurzzeitspektrums gewonnen werden. Für eine cepstrale Repräsentation, wie unter (1) beschrieben, ist eine mögliche Rücktransformation:

$$G_{\mu,smooth}(\lambda) = \exp\left(\mathrm{DFT}\left\{G^{cepst}_{\mu',smooth}(\lambda)\right\}\right),\ \mu = 0\ldots(M-1), \mu' = 0\ldots(M-1) \quad (2)$$

mit DFT{} der diskreten Fourier-Transformation und exp() der Exponentialfunktion, die in (2) elementweise angewendet wird.

[0018] Die Vorteile, die sich aus der erfindungsgemäßen Glättung von kurzzeitspektren ergeben, sind:

- eine effektive Unterdrückung von Fluktuationen bzw. Ausreißern,
- Beibehaltung der spektralen Auflösung für Sprachsignale und
- keine hörbare Beeinflussung von Sprache.

[0019] Es ist wichtig anzumerken, dass die für das Cepstrum verwendete inverse DFT in (1) und die DFT für die Rücktransformation in (2) durch andere Transformationen ersetzt werden können, ohne dass dabei die prinzipiellen Eigenschaften der Transformationskoeffizienten bzgl. der kompakten Repräsentation von Sprache verloren gehen. Genauso verhält es sich mit der Logarithmierung in (1) und der entsprechenden Umkehrfunktion in (2), der Exponentialfunktion. Auch hier sind andere nicht-lineare Abbildungen und auch lineare Abbildungen denkbar.

[0020] Transformationen unterscheiden sich in ihren verwendeten Basisfunktionen. Der Vorgang der Transformation bedeutet, dass das Signal mit den verschiedenen Basisfunktionen korreliert wird. Der resultierende Grad der Korrelation zwischen dem Signal und einer Basisfunktion ist dann der zugehörige Transformationskoeffizient. Bei einer Transformation entstehen so viele Transformationskoeffizienten wie es Basisfunktionen gibt. Ihre Anzahl ist hier mit M bezeichnet. Für die Erfindung wichtige Transformationen sind solche, durch deren Basisfunktionen das zu transformierende Kurzzeitspektrum in seiner Grobstruktur und seiner Feinstruktur aufgeschlüsselt wird.

[0021] Ein Unterscheidungsmerkmal von Transformationen ist die Orthogonalität. Orthogonale Transformationsbasen enthalten nur Basisfunktionen, die unkorreliert sind. Für den Fall, dass das Signal mit einer der Basisfunktionen identisch ist, entstehen bei orthogonalen Transformationen Transformationskoeffizienten mit dem Wert Null, bis auf den einen Koeffizienten, der identisch zum Signal ist. Die Trennschärfe einer orthogonalen Transformation ist demnach hoch. Nicht-orthogonale Transformationen verwenden Funktionsbasen, die miteinander korreliert sind.

[0022] Ein weiteres Merkmal ist, dass die Basisfunktionen für den betrachteten Anwendungsfall diskret und endlich sind, da es sich bei den verarbeiteten Signalrahmen um diskrete Signale von der Länge eines Rahmens handelt.

[0023] Wichtiges Merkmal einer Transformation ist die Invertierbarkeit. Existiert zu einer Transformation (Hintransformation) eine inverse Transformation, so entsteht durch Transformation eines Signals in Transformationskoeffizienten und anschließender inverser Transformation (Rücktransformation) dieser Koeffizienten wieder das Ausgangssignal, falls die Transformationkoeffizienten nicht verändert wurden.

[0024] In der Signalverarbeitung, wie sie hier beschrieben wird, ist die diskrete Fourier-Transformation (DFT) eine bevorzugte Transformation. Ein dazugehöriger wichtiger Algorithmus in der diskreten Signalverarbeitung ist die "Fast-Fourier-Transformation" (FFT). Außerdem sind die diskrete Cosinus-Transformation (DCT) und die diskrete Sinus-Transformation (DST) häufig verwendete Transformationen. Diese Transformationen werden hier unter dem Begriff "Standardtransformationen" zusammengefasst. Eine für die Erfindung entscheidende bereits erwähnte Eigenschaft der Standardtransformationen ist, dass die Amplituden der verschiedenen Transformationskoeffizienten unterschiedliche Grade an Feinstruktur des transformierten Signals repräsentieren. So beschreiben Koeffizienten mit kleinen Indizes die groben Strukturen des transformierten Signals, weil die zugehörigen Basisfunktionen niederfrequente harmonische Funktionen sind. Je höher der Index eines Transformationskoeffizienten bis hin zu $\mu' = M/2$, desto feiner sind die Strukturen des transformierten Signals, die durch diesen Koeffizienten beschrieben werden. Für darüber hinausgehende Koeffizienten dreht sich diese Eigenschaft wegen der Symmetrie der Koeffizienten um. In der Regel werden bei der Signalverarbeitung nur die Koeffizienten mit Indizes $\mu' = 0$ bis $\mu' = M/2$ verarbeitet und die restlichen Werte durch spiegeln der Resultate ermittelt.

[0025] Die Invertierbarkeit der Transformationen macht es außerdem möglich, die Transformation und ihre Inverse bei der Hin- und Rücktransformation zu vertauschen. In (1) ist also auch beispielsweise die Verwendung der DFT aus (2) möglich, wenn in (2) die IDFT aus (1) verwendet wird.

[0026] Vorteilhaft werden die spektralen Koeffizienten der Kurzzeitspektren vor der Hintransformation nicht-linear abgebildet. Prinzipielle, für die Erfindung vorteilhafte Eigenschaft der nicht-linearen Abbildung ist eine Dynamik-Kompression relativ großer Amplituden und eine Dynamik-Expansion relativ kleiner Amplituden.

[0027] Entsprechend können die spektralen Koeffizienten der geglätteten Kurzzeitspektren nach der Rücktransformation nicht-linear abgebildet werden, wobei die nicht-lineare Abbildung nach der Rücktransformation die Umkehrung der nicht-linearen Abbildung vor der Hintransformation ist.

[0028] Zweckmäßigerweise werden die spektralen Koeffizienten vor der Hintransformation durch Logarithmierung nicht-linear abgebildet.

[0029] Eine Form der zeitlichen Glättung kann durch ein rekursives System vorzugsweise erster Ordnung erreicht werden:

$$G^{cepst}_{\mu',smooth}(\lambda) = \beta_{\mu'}\, G^{cepst}_{\mu',smooth}(\lambda-1) + (1-\beta_{\mu'})\, G^{cepst}_{\mu'}(\lambda)\ . \qquad (3)$$

[0030] Mögliche Werte für die Glättungskonstanten für Koeffizienten der Standardtransformationen im Falle von Sprachsignalen sind $\beta_{\mu'} = 0$ für $\mu' = 0 \ldots 3$, $\beta_{\mu'} = 0{,}8$ für $\mu' = 4 \ldots M/2$ mit Ausnahme der Transformationskoeffizienten durch die die Pitch-Frequenz eines Sprechers repräsentiert wird, und $\beta_{\mu'} = 0{,}4$ für Transformationskoeffizienten, die die Pitch-Frequenz repräsentieren. Verfahren zur Bestimmung des Pitch-Koeffizienten sind zahlreich in der Literatur verfügbar. Beispielsweise kann zur Bestimmung des Koeffizienten der Pitch derjenige Koeffizient gewählt werden, dessen Index zwischen $\mu' = 16$ und $\mu' = 160$ liegt und der die maximale Amplitude aller Koeffizienten in diesem Indexbereich aufweist. Für die verbleibenden Transformationskoeffizienten mit Indizes $\mu' = M/2 + 1 \ldots M - 1$ gilt die Symmetriebedingung $\beta_{M-\mu'} = \beta_{\mu'}$. Die Werte sind für die Standardtransformationen sowie Kurzzeitspektren, die aus Signalen mit $f_s = 8000$ Hz entstanden sind, geeignet. Sie sind durch verhältnismäßige Umrechnung an andere Systeme anpassbar. Die Wahl $\beta_{\mu'} = 0$ bedeutet, dass die betreffenden Koeffizienten nicht geglättet werden. Es ist eine entscheidende Eigenschaft der Erfindung, dass Koeffizienten, die den groben Verlauf des Kurzzeitspektrums beschreiben, möglichst wenig geglättet werden, wenn es sich um die Entrauschung von Sprachsignalen handelt. So werden die groben Strukturen des breitbandigen Sprachspektrums vor Glättungseffekten geschützt. Die feinen Strukturen von Fluktuationen bzw. spektralen Ausreißern bilden sich bei den Standardtransformationen in den Transformationskoeffizienten zwischen $\mu' = 4$ und $\mu' = M/2$ ab, weshalb diese bis auf den Pitch der Sprache stark geglättet werden.

[0031] Vorteilhafterweise wird das Glättungsverfahren auf den Betrag oder eine Potenz des Betrags der Kurzzeitspektren angewendet.

[0032] Besonders vorteilhaft ist es, wenn zum Glätten der jeweiligen Transformationskoeffizienten unterschiedliche Zeitkonstanten verwendet werden. Die Zeitkonstanten können so gewählt werden, dass die Transformationskoeffizenten, die vornehmlich Sprache repräsentieren, wenig geglättet werden. Zweckmäßigerweise können die Transformationskoeffizenten, die hauptsächlich fluktuierende Hintergrundgeräusche und Artefakte der Geräuschreduktionsalgorithmen beschreiben, stark geglättet werden.

[0033] Als Kurzzeitspektrum kann die spektrale Gewichtungsfunktion eines Geräuschreduktionsalgorithmus bereitgestellt werden. Vorteilhaft kann als Kurzzeitspektrum auch die spektrale Gewichtungsfunktion eines Postfilters für mehrkanalige Verfahren zur Geräuschreduktion verwendet werden. Zweckmäßigerweise ergibt sich die spektrale Gewichtungsfunktion hierbei aus der Minimierung eines Fehlerkriteriums.

[0034] Als Kurzzeitspektrum kann auch ein gefiltertes Kurzzeitspektrum bereitgestellt werden.

[0035] Nach einer anderen Weiterbildung des Verfahrens, wird als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines mehrkanaligen Verfahrens zur Geräuschreduktion bereitgestellt.

[0036] Als Kurzzeitspektrum kann auch eine geschätzte Kohärenz oder eine geschätzte "Magnitude Squared Coherence" zwischen wenigstens zwei Mikrofonkanälen bereitgestellt werden.

[0037] Vorteilhaft wird als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines mehrkanaligen Verfahrens zur Sprecher- oder Quellentrennung bereitgestellt.

[0038] Weiterhin ist vorgesehen, dass als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines mehrkanaligen Verfahrens zur Sprechertrennung auf Basis von Phasenunterschieden von Signalen in den verschiedenen Kanälen (Phase Transform - PHAT) bereitgestellt wird.

[0039] Ferner ist es möglich, als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines mehrkanaligen Verfahrens auf Basis einer "Generalized Cross-Correlation" (GCC) zu verwenden.

[0040] Als Kurzzeitspektrum können auch spektrale Größen, die sowohl Sprach- als auch Störanteile enthalten, bereitgestellt werden.

[0041] So kann als Kurzzeitspektrum auch eine Schätzung des Signal-zu-Rausch-Verhältnisses in den einzelnen Frequenzbins bereitgestellt werden. Ferner kann als Kurzzeitspektrum eine Schätzung der Rauschleistung verwendet werden.

[0042] Das Problem von Fluktuationen in Kurzzeitspektren ist nicht nur in der Audiosignalverarbeitung bekannt. Weitere vorteilhafte Anwendungsgebiete sind die Bild- und die medizinische Signalverarbeitung.

[0043] In der Bildverarbeitung kann z.B. die Zeile eines Bildes als Signalrahmen interpretiert werden, der in den Spektralbereich transformiert werden kann. Die entstehenden Frequenzbins werden hier Ortsfrequenzbins genannt. Bei der Verarbeitung von Bildern im Ortsfrequenzbereich werden Algorithmen verwendet, die denen in der Audiosignalverarbeitung äquivalent sind. Mögliche Fluktuationen, die diese Algorithmen im Ortsfrequenzbereich erzeugen, resultieren im verarbeiteten Bild in optischen Artefakten. Diese sind äquivalent zum tonalen Rauschen in der Audioverarbeitung.

[0044] In der medizinischen Signalverarbeitung werden vom menschlichen Körper Signale abgeleitet, die wie akustische Signale verrauscht sein können. Das verrauschte Signal kann entsprechend rahmenweise in den Spektralbereich transformiert werden. Die entstehenden Spektrogramme lassen sich wie Audiospektren verarbeiten.

[0045] Das Glättungsverfahren kann in einem Telekommunikationsnetzwerk und/oder bei einer Rundfunkübertragung zur Verbesserung der Sprach- und/oder Bildqualität sowie zur Unterdrückung von Artefakten eingesetzt werden. In der mobilen Sprachkommunikation treten Verzerrungen des Sprachsignals auf, die zum einen durch die eingesetzten Sprachcodierverfahren (redundanzvermindernde Sprachkompression) und das damit verbundene Quantisierungsrauschen und zum anderen durch die vom Übertragungskanal hervorgerufenen Störungen bedingt sind. Letztere sind wiederum stark zeitlich und spektral fluktuierend und führen zu einer deutlich wahrnehmbaren Verschlechterung der Sprachqualität. Auch hier muss die empfängerseitig oder im Netzwerk eingesetzte Signalverarbeitung sicherstellen, dass die quasi zufälligen Artefakte reduziert werden. Zur Qualitätsverbesserung werden bisher sogenannte Postfilter und Fehlerverdeckungsverfahren eingesetzt. Während das Postfilter überwiegend die Reduktion von Quantisierungsrauschen zur Aufgabe hat, werden Fehlerverdeckungsverfahren zur Unterdrückung von übertragungsbedingten Kanalstörungen eingesetzt. In beiden Anwendungen können Verbesserungen erzielt werden, wenn in das Postfilter oder das Verdeckungsverfahren das erfindungsgemäße Glättungsverfahren integriert wird. Das Glättungsverfahren kann somit als Postfilter, in einem Postfilter, in Kombination mit einem Postfilter, im Rahmen eines Fehlerverdeckungsverfahrens oder in Zusammenhang mit einem Verfahren zur Sprach- und/oder Bildcodierung (Dekompressionsverfahren bzw. Dekodierungsverfahren) insbesondere empfängerseitig eingesetzt werden. Mit der Verwendung des Verfahrens als Postfilter ist dabei gemeint, dass das Verfahren zum Postfiltern eingesetzt wird, dass also mit einem das Verfahren umsetzenden Algorithmus die in den Anwendungen entstehenden Daten prozessiert werden. Weiterhin ist es möglich, die Qualität des Sprachsignals im Telekommunikationsnetzwerk zu verbessern, indem das Sprachsignalspektrum oder eine davon abgeleitete Größe mit dem erfindungsgemäßen Glättungsverfahren geglättet wird.

[0046] Die Erfindung wird nachfolgend anhand von in den Figuren dargestellten Abbildungen näher erläutert. Es zeigen:

Figur 1 ein unverrauschtes Zeitsignal;

Figur 2 ein verrauschtes Zeitsignal;

Figur 3 einen einzelnen Signalrahmen im Zeitbereich;

Figur 4 einen einzelnen Signalrahmen im Spektralbereich;

Figur 5 eine Gewichtungsfunktion für einen einzelnen Rahmen;

Figur 6 das Spektrogramm eines unverrauschten Signals;

Figur 7 das Spektrogramm eines verrauschten Signals;

Figur 8 das Spektrogramm eines mit der ungeglätteten Gewichtungsfunktion gefilterten Signals;

Figur 9 das Spektrogramm eines mit einer erfindungsgemäß geglätteten Gewichtungsfunktion gefilterten Signals;

Figur 10 ein gefiltertes Zeitsignal mit tonalen Artefakten;

Figur 11 ein gemäß der Erfindung gefiltertes Zeitsignal;

Figur 12 das Spektrogramm einer ungeglätteten Gewichtungsfunktion;

Figur 13 das Spektrogramm einer erfindungsgemäß geglätteten Gewichtungsfunktion;

Figur 14 den Betrag des Cepstrums eines unverrauschten Sprachsignals und

Figur 15 den Signalflussgraphen gemäß einer bevorzugten Ausführungsform der Erfindung.

[0047] In Figur 1 ist ein unverrauschtes Signal in Form der Amplitude über die Zeit dargestellt. Die Dauer des Signals ist 4 Sekunden, die Amplituden reichen von ca. -0,18 bis ca. 0,18. In Figur 2 ist das Signal in verrauschter Form dargestellt. Man erkennt ein zufälliges Grundrauschen über dem gesamten Zeitverlauf.

[0048] In Figur 3 ist das Signal eines einzelnen Signalrahmens λ dargestellt. Der Signalrahmen hat eine Segmentdauer von 32 Millisekunden. Die Amplitude beider Graphen bewegt sich zwischen -0,1 und 0,1. Die einzelnen Abtastwerte der digitalen Signale sind zu Graphen verbunden. Der verrauschte Graph repräsentiert das Eingangssignal, in dem das unverrauschte Signal enthalten ist. Eine Trennung von Signal und Rauschen im verrauschten Signal ist in dieser Repräsentation des Signals kaum möglich.

[0049] Figur 4 ist eine Darstellung desselben Signalrahmens nach der Transformation in den Frequenzbereich. Die einzelnen Frequenzbins μ sind zu Graphen verbunden. Auch in dieser Figur sind die Frequenzbins verrauscht und unverrauscht dargestellt, wobei wieder das unverrauschte Signal das im verrauschten Signal enthaltene Sprachsignal ist. Über der Abszisse sind die Frequenzbins μ von 0 bis 128 eingezeichnet. Sie haben Amplituden von ca. -40 Dezibel (dB) bis ca. 10 dB. Aus dem Vergleich der Graphen ist ersichtlich, dass die Energie des Sprachsignals in einigen Frequenzbins in einer kammartigen Struktur konzentriert ist, während das Rauschen auch in den dazwischenliegenden Bins vorhanden ist.

[0050] In Figur 5 ist eine Gewichtungsfunktion für den verrauschten Rahmen aus Figur 4 dargestellt. Für jeden Frequenzbin μ ergibt sich in Abhängigkeit vom Verhältnis aus Sprach- und Rauschenergie ein Faktor zwischen 0 und 1. Die einzelnen Gewichtungsfaktoren sind zu einem Graphen verbunden. Man erkennt die kammartige Struktur des Sprachspektrums wieder.

[0051] In den Figuren 6 und 7 sind Spektrogramme aus einer Folge von unverrauschten bzw. verrauschten Kurzzeitspektren (Figur 4) dargestellt. Auf der Abszisse ist der Rahmenindex λ aufgetragen, über der Ordinate der Frequenzbinindex μ. Die Amplituden der einzelnen Frequenzbins sind als Grauwerte dargestellt. Im Vergleich von Figur 6 und 7 wird deutlich, wie Sprache in wenigen Frequenzbins konzentriert ist. Sie bildet zudem regelmäßige Strukturen aus. Das Rauschen ist dagegen über alle Frequenzbins verteilt.

[0052] In Figur 8 ist das Spektrogramm eines gefilterten Signals dargestellt. Die Achsen entsprechen denen aus den Figuren 6 und 7. Aus einem Vergleich mit Figur 6 ist erkennbar, dass durch Schätzfehler in der Gewichtungsfunktion hohe Amplituden in Frequenzbins verbleiben, die keine Sprache enthalten. Diese Ausreißer zu unterdrücken ist Ziel des erfindungsgemäßen Verfahrens.

[0053] In Figur 9 ist das Spektrogramm eines Signals dargestellt, das gemäß einer bevorzugten Weiterbildung des erfindungsgemäßen Verfahrens mit einer geglätteten Gewichtungsfunktion gefiltert wurde. Die Achsen entsprechen denen der vorangegangenen Spektrogramme. Im Vergleich mit Figur 8 sind die Ausreißer stark vermindert. Die Sprachanteile im Spektrogramm sind dagegen in ihrer wesentlichen Form erhalten.

[0054] In den Figuren 10 und 11 sind die Zeitsignale dargestellt, die sich jeweils aus den gefilterten Spektren der Figuren 8 und 9 ergeben. Aufgetragen ist die Amplitude über der Zeit. Die Signale sind 4 Sekunden lang und haben Amplituden zwischen ca. -0,18 und 0,18. Die Ausreißer im Spektrogramm aus Figur 8 ergeben im zugehörigen Zeitsignal in Figur 10 deutlich sichtbare tonale Artefakte, die im unverrauschten Signal aus Figur 1 nicht vorhanden sind. Das Zeitsignal in Figur 11 weist einen deutlich ruhigeren Verlauf des Restrauschens auf. Dieses Zeitsignal ergibt sich aus dem Spektrogramm von Figur 9, das durch Filterung mit der geglätteten Gewichtungsfunktion erzeugt wurde.

[0055] In Figur 12 ist die ungeglättete Gewichtungsfunktion für alle Rahmen dargestellt. Zu jedem Rahmen λ sind entlang der Ordinate Frequenzbins μ aufgetragen. Die Werte der Gewichtungsfunktion sind als Grauton dargestellt. Die Fluktuationen, die aus Schätzfehlern resultieren, sind als unregelmäßige Flecken erkennbar.

[0056] In Figur 13 ist die geglättete Gewichtungsfunktion für alle Rahmen dargestellt. Die Achsen entsprechen denen aus Figur 12. Durch die Glättung werden die Fluktuationen verschmiert und im Wert stark vermindert. Die Struktur der Sprachfrequenzbins bleibt dagegen deutlich erkennbar.

[0057] In Figur 14 ist der Betrag des Cepstrums eines unverrauschten Signals über alle Rahmen dargestellt. Zu jedem Rahmen λ sind entlang der Ordinate die cepstralen Bins μ' aufgetragen. Die Werte der Beträge der cepstralen Koeffizienten $G_{\mu'}^{\text{cepst}}(\lambda)$ sind als Grautöne dargestellt. Ein Vergleich mit Figur 6 zeigt, dass Sprache im Cepstrum auf eine noch geringere Anzahl von Koeffizienten konzentriert ist. Außerdem sind diese Koeffizienten in ihrer Position weniger variabel. Deutlich erkennbar ist auch der Verlauf des cepstralen Koeffizienten, der die Pitch-Frequenz repräsentiert.

[0058] In Figur 15 ist ein Signalflussgraph gemäß einer bevorzugten Ausführungsform der Erfindung dargestellt. Ein verrauschtes Eingangssignal wird in eine Folge von Kurzzeitspektren transformiert und daraus über spektrale Zwischengrößen anschließend eine Gewichtungsfunktion zur Filterung geschätzt. Es wird jeweils ein Rahmen zur Zeit bearbeitet. Zunächst werden die Kurzzeitspektren der Gewichtungsfunktion einer nicht-linearen, logarithmischen Abbildung unterworfen. Es folgt eine Hintransformation in den cepstralen Bereich. Die so transformierten Kurzzeitspektren werden damit durch Transformationskoeffizienten der Basisfunktionen repräsentiert. Die auf diesem Wege berechneten Transformationskoeffizienten werden getrennt voneinander unter Verwendung von unterschiedlichen Zeitkonstanten geglättet. Der rekursive Charakter der Glättung ist durch die Rückführung der Ausgabe der Glättung zu ihrem Eingang angedeutet. Von den Signalpfaden der insgesamt M Transformationskoeffizienten sind nur 3 dargestellt, die restlichen sind durch drei Punkte "..." ersetzt. Nach der Glättung erfolgen eine Rücktransformation und danach die nicht-lineare Umkehrabbildung. Auf diese Weise erhält man als Ergebnis eine Folge von geglätteten Kurzzeitspektren der Gewichtungsfunktion. Diese geglätteten Kurzzeitspektren der Gewichtungsfunktion können mit den verrauschten Kurzzeitspektren multipliziert werden, wodurch gefilterte Kurzzeitspektren mit wenigen Ausreißern entstehen. Diese können dann in ein Zeitsignal mit verringertem Rauschpegel umgerechnet werden. Der Teil des Signalflussgraphen, der die erfindungsgemäße Glättung beschreibt, ist gestrichelt umrandet.

Patentansprüche

1. Glättungsverfahren zur Unterdrückung von fluktuierenden Artefakten bei der Störgeräuschreduktion mit folgenden Schritten:

   - Bereitstellen von Kurzzeitspektren einer Folge von digitalen Abtastwerten umfassenden Signalrahmen,
   - Transformieren jedes Kurzzeitspektrums durch eine Hintransformation, die das Kurzzeitspektrum durch Transformationskoeffizienten beschreibt, welche das Kurzzeitspektrum in seine groben und seine feinen Strukturen unterteilt repräsentieren,
   - Glätten der Transformationskoeffizienten jeweils gleicher Koeffizientenindizes durch Kombination von wenigstens zwei aufeinanderfolgenden transformierten Kurzzeitspektren und
   - Transformieren der geglätteten Transformationskoeffizienten in geglättete Kurzzeitspektren durch eine Rücktransformation.

2. Glättungsverfahren nach dem vorhergehenden Anspruch, **dadurch gekennzeichnet, dass** für die Rücktransformation die Inverse zur Hintransformation verwendet wird.

3. Glättungsverfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** eine Transformation mit orthogonaler Basis verwendet wird.

4. Glättungsverfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** eine Transformation mit nicht-orthogonaler Basis verwendet wird.

5. Glättungsverfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** für die Transformationen die diskrete Fourier-Transformation und ihre Inverse verwendet werden.

6. Glättungsverfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** für die Transformationen die Fast-Fourier-Transformation und ihre Inverse verwendet werden.

7. Glättungsverfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** für die Transformationen die diskrete Cosinus-Transformation und ihre Inverse verwendet werden.

8. Glättungsverfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** für die Transformationen die diskrete Sinus-Transformation und ihre Inverse verwendet werden.

9. Glättungsverfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** die Kurzzeitspektren vor der Hintransformation nicht-linear abgebildet werden.

10. Glättungsverfahren nach dem vorhergehenden Anspruch, **dadurch gekennzeichnet, dass** die geglätteten Kurzzeitspektren nach der Rücktransformation nicht-linear abgebildet werden, wobei die nicht-lineare Abbildung der Rücktransformation die Umkehrung der nicht-linearen Abbildung der Hintransformation ist.

11. Glättungsverfahren nach einem der beiden vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** die Kurzzeitspektren vor der Hintransformation durch Logarithmierung nicht-linear abgebildet werden.

12. Glättungsverfahren nach einem der Ansprüche 1 bis 11, **dadurch gekennzeichnet, dass** zum Glätten der Transformationskoeffizienten eine rekursive Glättung verwendet wird.

13. Glättungsverfahren nach einem der Ansprüche 1 bis 11, **dadurch gekennzeichnet, dass** zum Glätten der Transformationskoeffizienten eine nicht-rekursive Glättung verwendet wird.

14. Glättungsverfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** die Glättung auf den Betrag oder eine Potenz des Betrags der Kurzzeitspektren angewendet wird.

15. Glättungsverfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** zum Glätten der jeweiligen Transformationskoeffizienten unterschiedliche Zeitkonstanten verwendet werden.

16. Glättungsverfahren nach dem vorhergehenden Anspruch, **dadurch gekennzeichnet, dass** Zeitkonstanten so gewählt werden, dass die Transformationskoeffizienten, die typischerweise spektrale Strukturen von Sprache beschreiben, wenig geglättet werden.

17. Glättungsverfahren nach einem der beiden vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** Zeitkonstanten so gewählt werden, dass die Transformationskoeffizienten, die spektrale Strukturen von fluktuierenden spektralen Größen und von Artefakten von Geräuschreduktionsalgorithmen beschreiben, stark geglättet werden.

18. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines Geräuschreduktionsalgorithmus bereitgestellt wird.

19. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines Postfilters für mehrkanalige Verfahren zur Geräuschreduktion verwendet wird.

20. Glättungsverfahren nach einem der beiden vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** sich die spektrale Gewichtungsfunktion aus der Minimierung eines Fehlerkriteriums ergibt.

21. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum ein gefiltertes Kurzzeitspektrum bereitgestellt wird.

22. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines mehrkanaligen Verfahrens zur Geräuschreduktion bereitgestellt wird.

23. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine geschätzte Kohärenz oder eine geschätzte "Magnitude Squared Coherence" zwischen wenigstens zwei Mikrofonkanälen bereitgestellt wird.

24. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines mehrkanaligen Verfahrens zur Sprecher- oder Quellentrennung bereitgestellt wird.

25. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines mehrkanaligen Verfahrens zur Sprechertrennung auf Basis von Phasenunterschieden von Signalen in den verschiedenen Kanälen (Phase Transform - PHAT) bereitgestellt wird.

26. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine spektrale Gewichtungsfunktion eines mehrkanaligen Verfahrens zur Geräuschreduktion auf Basis einer "Generalized Cross-Correlation" (GCC) bereitgestellt wird.

27. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum spektrale Größen, die sowohl Sprach- als auch Störanteile enthalten, bereitgestellt werden.

28. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine Schätzung des Signal-zu-Rausch-Verhältnisses bereitgestellt wird.

29. Glättungsverfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum eine Schätzung der Rauschleistung bereitgestellt wird.

30. Glättungsverfahren nach einem der Ansprüche 1 bis 15, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum transformierte Signalrahmen eines Bildsignals bereitgestellt werden und die zeilen- oder spaltenweise oder 2-dimensional berechneten Koeffizienten des transformierten Bildsignals einer räumlichen Glättung mit unterschiedlichen Glättungsparametern unterworfen werden.

31. Glättungsverfahren nach dem vorhergehenden Anspruch, **dadurch gekennzeichnet, dass** das Bildsignal ein Videosignal ist.

32. Glättungsverfahren nach einem der Ansprüche 1 bis 15, **dadurch gekennzeichnet, dass** als Kurzzeitspektrum ein transformiertes, vom menschlichen Körper abgeleitetes, medizinisches Signal verwendet wird.

33. Glättungsverfahren nach einem der Ansprüche 1 bis 32, **dadurch gekennzeichnet dass** das Glättungsverfahren in einem Postfilter, in Kombination mit einem Postfilter, im Rahmen eines Fehlerverdeckungsverfahrens oder in Zusammenhang mit einem Verfahren zur Sprach- und/oder Bildcodierung insbesondere empfängerseitig eingesetzt wird.

34. Glättungsverfahren nach einem der Ansprüche 1 bis 33, **dadurch gekennzeichnet, dass** das Glättungsverfahren in einem Telekommunikationsnetzwerk und/oder bei einer Rundfunkübertragung zur Verbesserung der Sprach- und/oder Bildqualität sowie zur Unterdrückung von Artefakten eingesetzt wird.

Claims

1. Smoothing method for suppressing fluctuating artifacts during noise reduction, having the following steps:

   • short-term spectra for a series of signal frames comprising digital samples are provided,
   • each short-term spectrum is transformed by forward transformation, which describes the short-term spectrum using transformation coefficients which represent the short-term spectrum divided into its coarse and its fine structures,
   • the transformation coefficients with the same coefficient indices in each case are smoothed by combining at

least two successive transformed short-term spectra, and
• the smoothed transformation coefficients are transformed into smoothed short-term spectra by backward transformation.

2. Smoothing method according to the preceding claim, **characterized in that** the inverse of the forward transformation is used for the backward transformation.

3. Smoothing method according to Claim 1 or 2, **characterized in that** transformation with an orthogonal base is used.

4. Smoothing method according to Claim 1 or 2, **characterized in that** transformation with a nonorthogonal base is used.

5. Smoothing method according to Claim 1 or 2, **characterized in that** the discrete Fourier transformation and the inverse thereof are used for the transformations.

6. Smoothing method according to Claim 1 or 2, **characterized in that** the fast Fourier transformation and the inverse thereof are used for the transformations.

7. Smoothing method according to Claim 1 or 2, **characterized in that** the discrete cosine transformation and the inverse thereof are used for the transformations.

8. Smoothing method according to Claim 1 or 2, **characterized in that** the discrete sine transformation and the inverse thereof are used for the transformations.

9. Smoothing method according to one of the preceding claims, **characterized in that** the short-term spectra are mapped nonlinearly before the forward transformation.

10. Smoothing method according to the preceding claim, **characterized in that** the smoothed short-term spectra are mapped nonlinearly after the backward transformation, wherein the nonlinear mapping of the backward transformation is the reversal of the nonlinear mapping of the forward transformation.

11. Smoothing method according to one of the two preceding claims, **characterized in that** the short-term spectra are mapped nonlinearly before the forward transformation by logarithmization.

12. Smoothing method according to one of Claims 1 to 11, **characterized in that** recursive smoothing is used for smoothing the transformation coefficients.

13. Smoothing method according to one of Claims 1 to 11, **characterized in that** nonrecursive smoothing is used for smoothing the transformation coefficients.

14. Smoothing method according to one of the preceding claims, **characterized in that** the smoothing is applied to the absolute value or to a power of the absolute value of the short-term spectra.

15. Smoothing method according to one of the preceding claims, **characterized in that** different time constants are used for smoothing the respective transformation coefficients.

16. Smoothing method according to the preceding claim, **characterized in that** time constants are chosen such that the transformation coefficients which typically describe spectral structures of voice are smoothed little.

17. Smoothing method according to one of the two preceding claims, **characterized in that** time constants are chosen such that the transformation coefficients which describe spectral structures of fluctuating spectral magnitudes and of artifacts of noise reduction algorithms are smoothed much.

18. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is a spectral weighting function of a noise reduction algorithm.

19. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum used is a spectral weighting function of a post filter for multichannel methods for noise reduction.

20. Smoothing method according to one of the two preceding claims, **characterized in that** the spectral weighting function results from the minimization of an error criterion.

21. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is a filtered short-term spectrum.

22. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is a spectral weighting function of a multichannel method for noise reduction.

23. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is an estimated coherence or an estimated "Magnitude Squared Coherence" between at least two microphone channels.

24. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is a spectral weighting function of a multichannel method for speaker or source separation.

25. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is a spectral weighting function of a multichannel method for speaker separation on the basis of phase differences for signals in the different channels (Phase Transform - PHAT).

26. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is a spectral weighting function of a multichannel method for noise reduction on the basis of a "Generalized Cross-Correlation" (GCC).

27. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is spectral magnitudes which contain both voice and noise components.

28. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is an estimate of the signal-to-noise ratio.

29. Smoothing method according to one of Claims 1 to 17, **characterized in that** the short-term spectrum provided is an estimate of the noise power.

30. Smoothing method according to one of Claims 1 to 15, **characterized in that** the short-term spectrum provided is transformed signal frames of an image signal, and the coefficients of the transformed image signal which are calculated row by row or column by column or two-dimensionally are subjected to spatial smoothing with different smoothing parameters.

31. Smoothing method according to the preceding claim, **characterized in that** the image signal is a video signal.

32. Smoothing method according to one of Claims 1 to 15, **characterized in that** the short-term spectrum used is a transformed medical signal derived from the human body.

33. Smoothing method according to one of Claims 1 to 32, **characterized in that** the smoothing method is used in a post filter, in combination with a post filter, as part of an error masking method or in connection with a method for voice and/or image coding, particularly at the receiver end.

34. Smoothing method according to one of Claims 1 to 33, **characterized in that** the smoothing method is used in a telecommunication network and/or during a broadcast transmission to improve the voice and/or image quality and to suppress artifacts.

Revendications

1. Procédé de lissage pour supprimer des artefacts fluctuants lors de la réduction du bruit parasite, comprenant les stades suivants :

    • on se procure des spectres de temps court d'une succession de trames de signal comprenant des valeurs

numériques d'échantillonnage,
• on transforme chaque spectre de temps court par une transformation, qui décrit le spectre de temps court par des coefficients de transformation, qui représentent le spectre de temps court subdivisé en ses structures grossières et ses structures fines,
• on lisse les coefficients de transformation respectivement de même indice de coefficient par combinaison d'au moins deux spectres de temps court successifs transformés, et
• on transforme par une retransformation les coefficients de transformation lissés en des spectres de temps court lissés.

2. Procédé de lissage suivant la revendication précédente, **caractérisé en ce qu'**on utilise l'inverse de la transformation pour la retransformation.

3. Procédé de lissage suivant la revendication 1 ou 2, **caractérisé en ce qu'**on utilise une transformation de base orthogonale.

4. Procédé de lissage suivant la revendication 1 ou 2, **caractérisé en ce qu'**on utilise une transformation de base non orthogonale.

5. Procédé de lissage suivant la revendication 1 ou 2, **caractérisé en ce qu'**on utilise pour la transformation la transformation de Fourier discrète et son inverse.

6. Procédé de lissage suivant la revendication 1 ou 2, **caractérisé en ce qu'**on utilise pour la transformation la transformation de Fourier rapide et son inverse.

7. Procédé de lissage suivant la revendication 1 ou 2, **caractérisé en ce qu'**on utilise pour la transformation la transformation cosinus discrète et son inverse.

8. Procédé de lissage suivant la revendication 1 ou 2, **caractérisé en ce qu'**on utilise pour la transformation la transformation sinus discrète et son inverse.

9. Procédé de lissage suivant l'une des revendications précédentes, **caractérisé en ce qu'**on reproduit de manière non linéaire les spectres de temps court avant la transformation.

10. Procédé de lissage suivant la revendication précédente, **caractérisé en ce qu'**on reproduit de manière non linéaire, après la retransformation, les spectres de temps court lissés, la reproduction non linéaire de la retransformation étant l'inverse de la reproduction non linéaire de la transformation.

11. Procédé de lissage suivant l'une des deux revendications précédentes, **caractérisé en ce qu'**on reproduit de manière non linéaire en prenant le logarithme les spectres de temps court avant la transformation.

12. Procédé de lissage suivant l'une des revendications 1 à 11, **caractérisé en ce qu'**on utilise un lissage récursif pour le lissage des coefficients de transformation.

13. Procédé de lissage suivant l'une des revendications 1 à 11, **caractérisé en ce qu'**on utilise un lissage non récursif pour le lissage des coefficients de transformation.

14. Procédé de lissage suivant l'une des revendications précédentes, **caractérisé en ce qu'**on applique le lissage à la valeur absolue ou à une puissance de la valeur absolue des spectres de temps court.

15. Procédé de lissage suivant l'une des revendications précédentes, **caractérisé en ce qu'**on utilise pour le lissage des coefficients de transformation respectifs des constantes de temps différentes.

16. Procédé de lissage suivant la revendication précédente, **caractérisé en ce qu'**on choisit des constantes de temps de manière à lisser peu les coefficients de transformation, qui décrivent typiquement des structures spectrales de parole.

17. Procédé de lissage suivant l'une des revendications précédentes, **caractérisé en ce qu'**on choisit des constantes de temps de manière à lisser beaucoup les coefficients de transformation, qui décrivent des structures spectrales

de grandeurs spectrales fluctuantes et d'artéfacts d'algorithmes de réduction de bruit.

18. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une fonction spectrale de pondération d'un algorithme de réduction de bruit.

19. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une fonction spectrale de pondération d'un post-filtre pour des procédés à plusieurs canaux de réduction du bruit.

20. Procédé de lissage suivant l'une deux revendications précédentes, **caractérisé en ce qu'**on obtient la fonction spectrale de pondération à partir de la minimisation d'un critère d'erreur.

21. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court un spectre de temps court filtré.

22. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une fonction spectrale de pondération d'un procédé à plusieurs canaux de réduction du bruit.

23. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une cohérence estimée ou une "Magnitude Squared Coherence" estimée entre au moins deux microcanaux radio.

24. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une fonction spectrale d'un procédé à plusieurs canaux pour la séparation de locuteurs ou de sources.

25. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une fonction spectrale d'un procédé à plusieurs canaux pour la séparation de locuteurs sur la base de différences de phase de signaux dans les divers canaux (phase Transform - PHAT).

26. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une fonction spectrale de pondération d'un procédé à plusieurs canaux pour la réduction du bruit sur la base d'une "Generalized Cross-Correlation" (GCC).

27. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court des grandeurs spectrales, qui contiennent à la fois des parties vocales et des parties parasites.

28. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une évaluation du rapport du signal au bruit.

29. Procédé de lissage suivant l'une des revendications 1 à 17, **caractérisé en ce qu'**on se procure comme spectre de temps court une évaluation de la puissance du bruit.

30. Procédé de lissage suivant l'une des revendications 1 à 15, **caractérisé en ce qu'**on se procure comme spectre de temps court des trames transformées d'un signal d'image et on soumet les coefficients ligne par ligne ou colonne par colonne ou calculés en deux dimensions du signal d'image transformé à un lissage dans l'espace par des paramètres de lissage différents.

31. Procédé de lissage suivant la revendication précédente, **caractérisé en ce que** le signal d'image est un signal vidéo.

32. Procédé de lissage suivant l'une des revendications 1 à 15, **caractérisé en ce qu'**on utilise comme spectre de temps court un signal médical transformé dérivé du corps humain.

33. Procédé de lissage suivant l'une des revendications 1 à 32, **caractérisé en ce qu'**on utilise le procédé de lissage dans un post-filtre, en combinaison avec un post-filtre, dans le cas d'un procédé de découverte d'erreur ou en liaison avec un procédé de codage de la voix et/ou de l'image, notamment du côté récepteur.

34. Procédé de lissage suivant l'une des revendications 1 à 33, **caractérisé en ce qu'**on utilise le procédé de lissage dans un réseau de télécommunication et/ou dans une radiodiffusion pour améliorer la qualité de la voix et/ou de

l'image, ainsi que pour supprimer des artefacts.

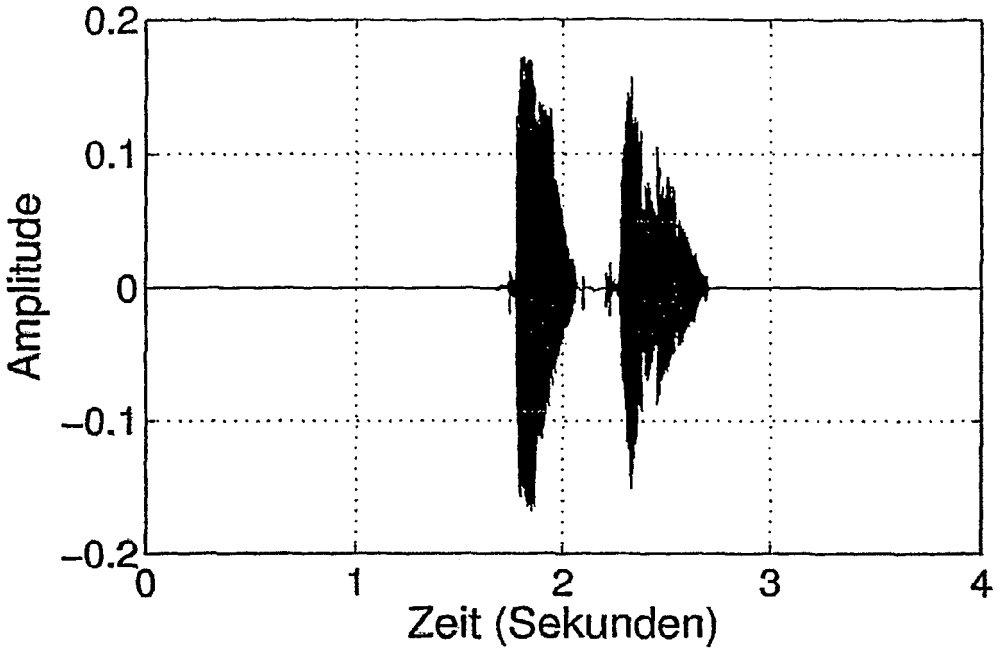

0.2
0.1
0
−0.1
−0.2
Amplitude
0
1
2
3
4
Zeit (Sekunden)

Fig. 1

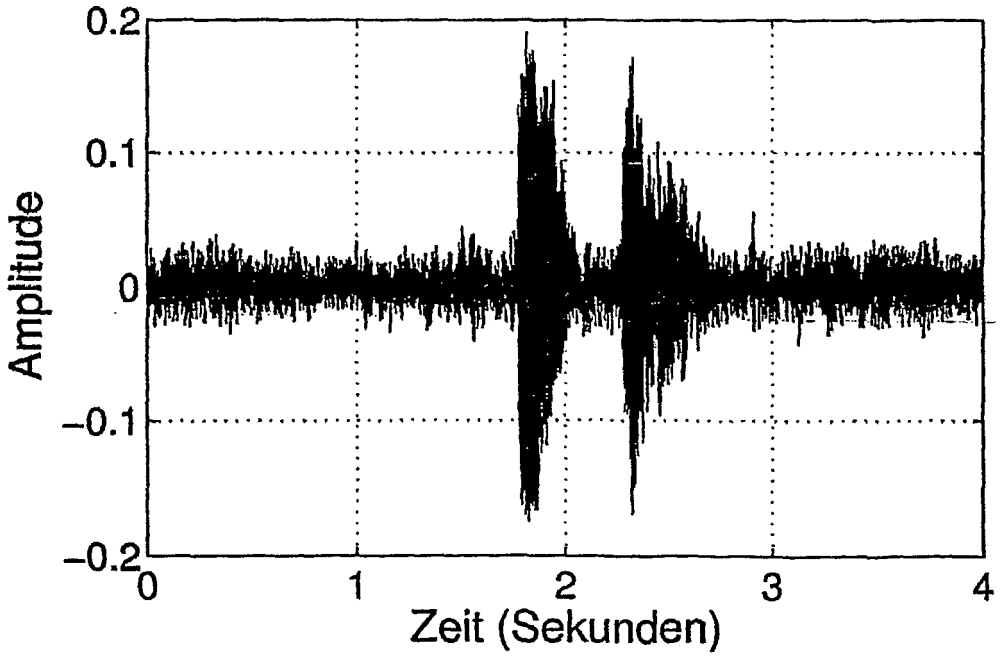

0.2
0.1
0
−0.1
−0.2
Amplitude
0
1
2
3
4
Zeit (Sekunden)

Fig. 2

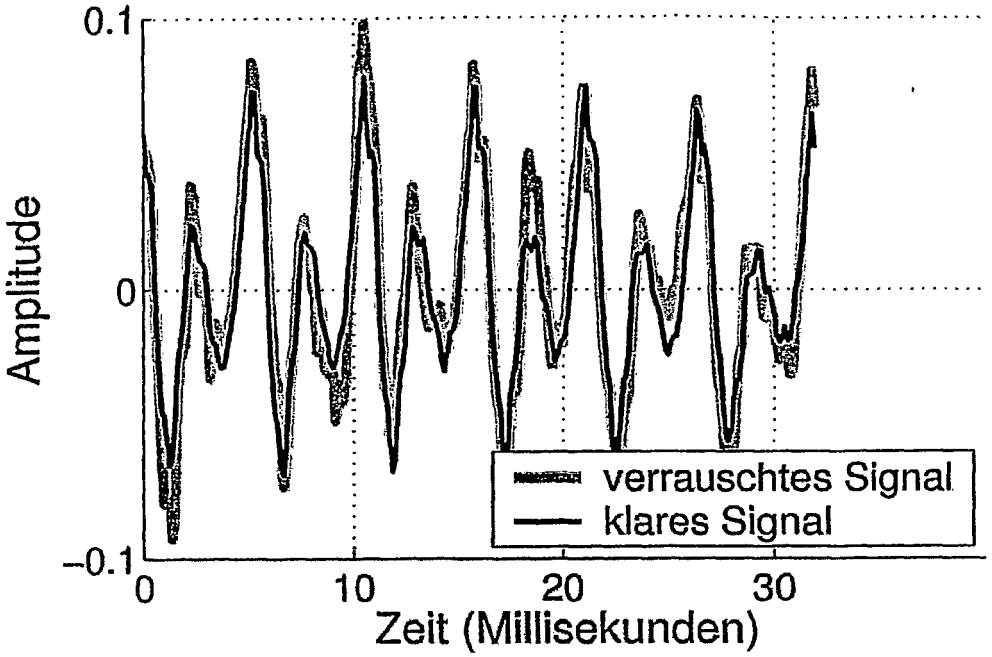

0.1
0
−0.1
Amplitude
0
10
20
30
Zeit (Millisekunden)
verrauschtes Signal
klares Signal

Fig. 3

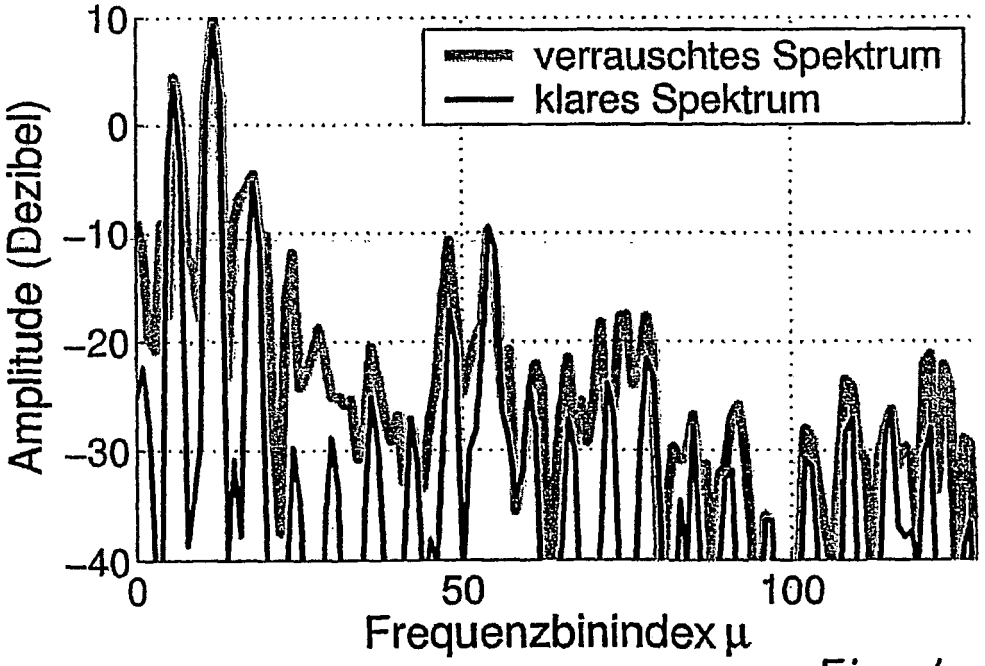

10
0
−10
−20
−30
−40
Amplitude (Dezibel)
0
50
100
Frequenzbinindex μ
verrauschtes Spektrum
klares Spektrum

Fig. 4

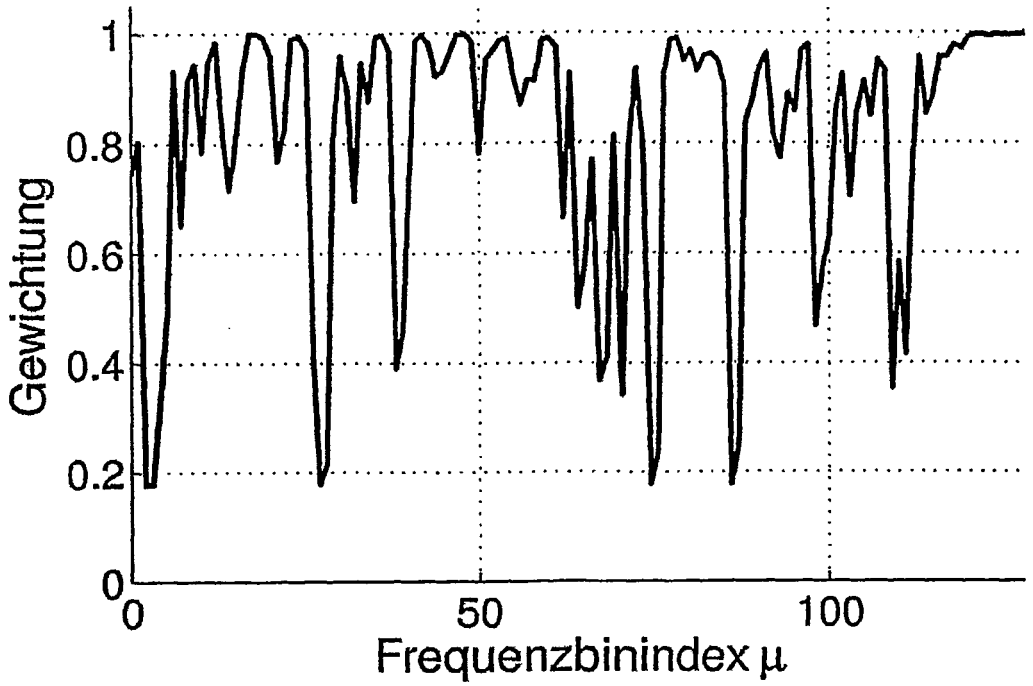


Fig. 5

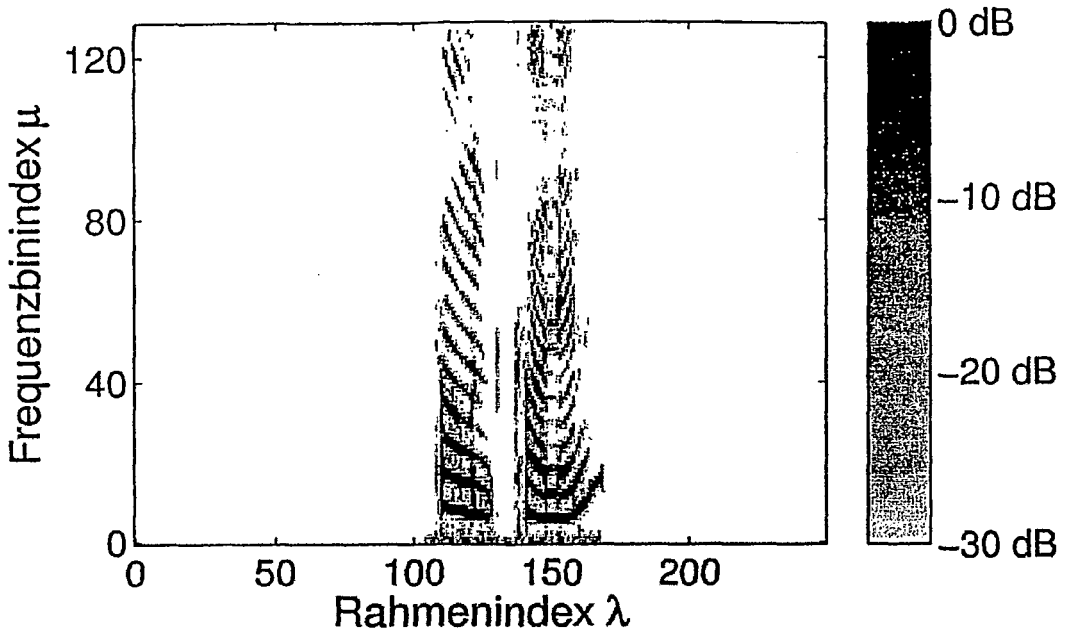

Frequenzbinindex μ
120
80
40
0
0
50
100
150
200
Rahmenindex λ
0 dB
−10 dB
−20 dB
−30 dB

Fig. 6

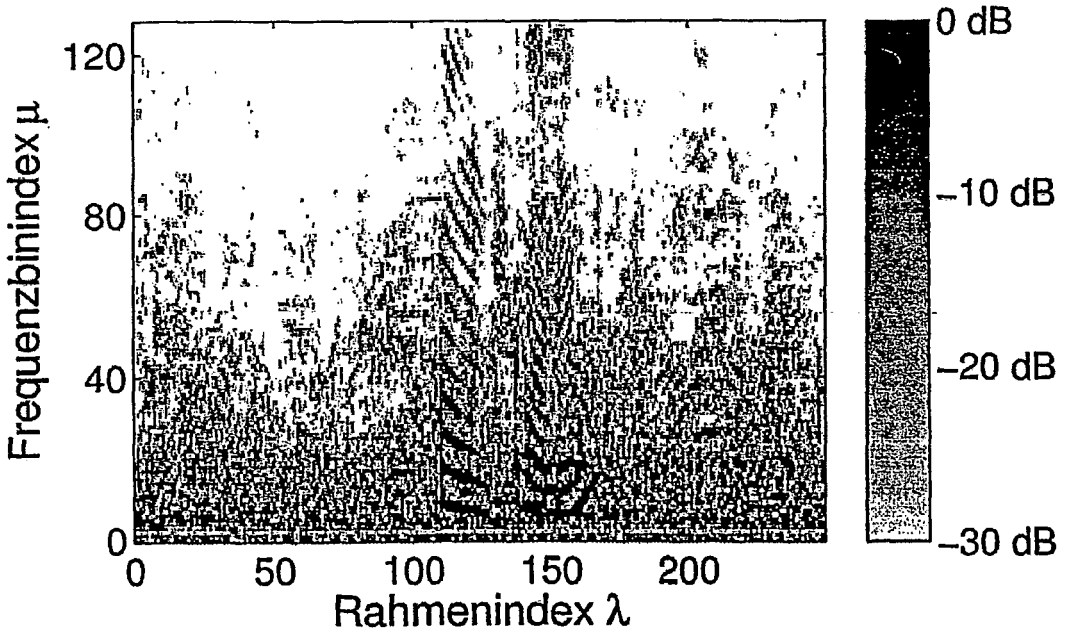

Frequenzbinindex μ
120
80
40
0
0
50
100
150
200
Rahmenindex λ
0 dB
−10 dB
−20 dB
−30 dB

Fig. 7

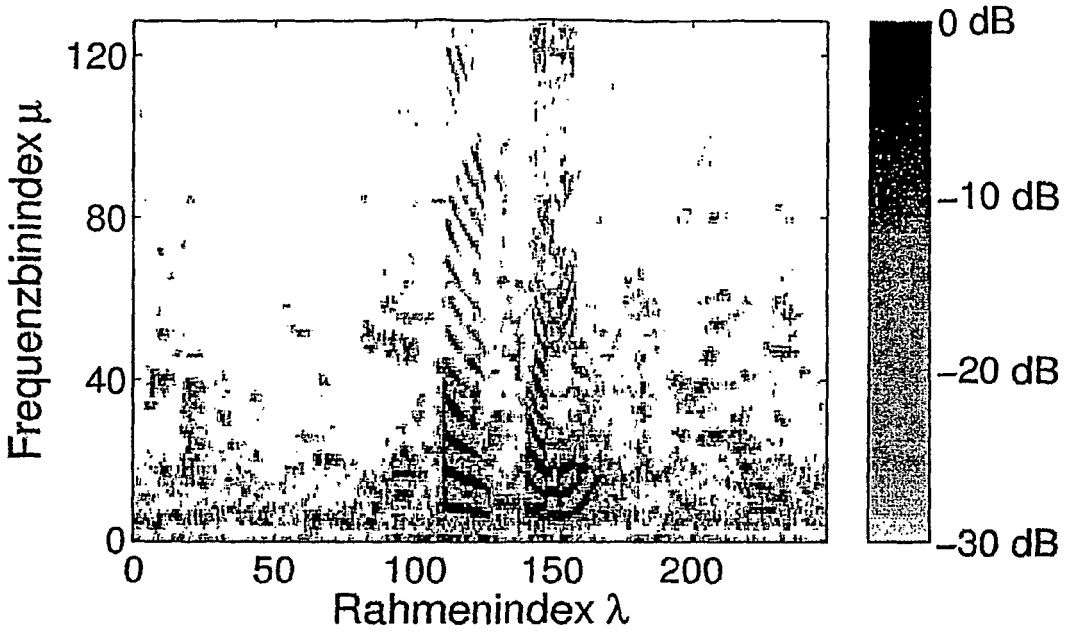

Frequenzbinindex μ
120
80
40
0
0
50
100
150
200
Rahmenindex λ
0 dB
−10 dB
−20 dB
−30 dB

Fig. 8

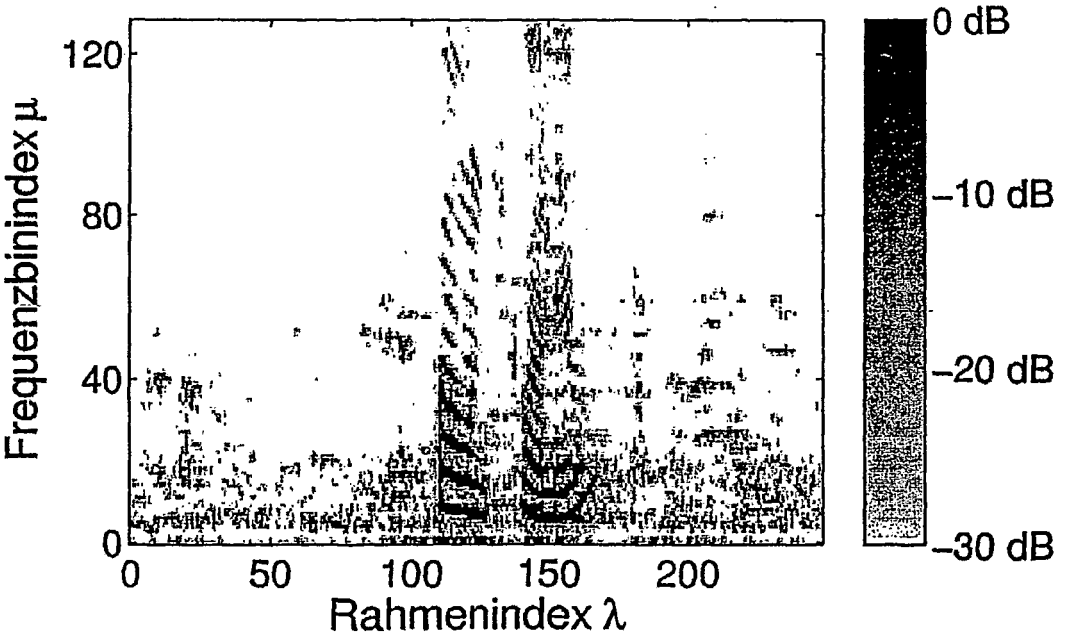

Frequenzbinindex μ
120
80
40
0
0
50
100
150
200
Rahmenindex λ
0 dB
−10 dB
−20 dB
−30 dB

Fig. 9

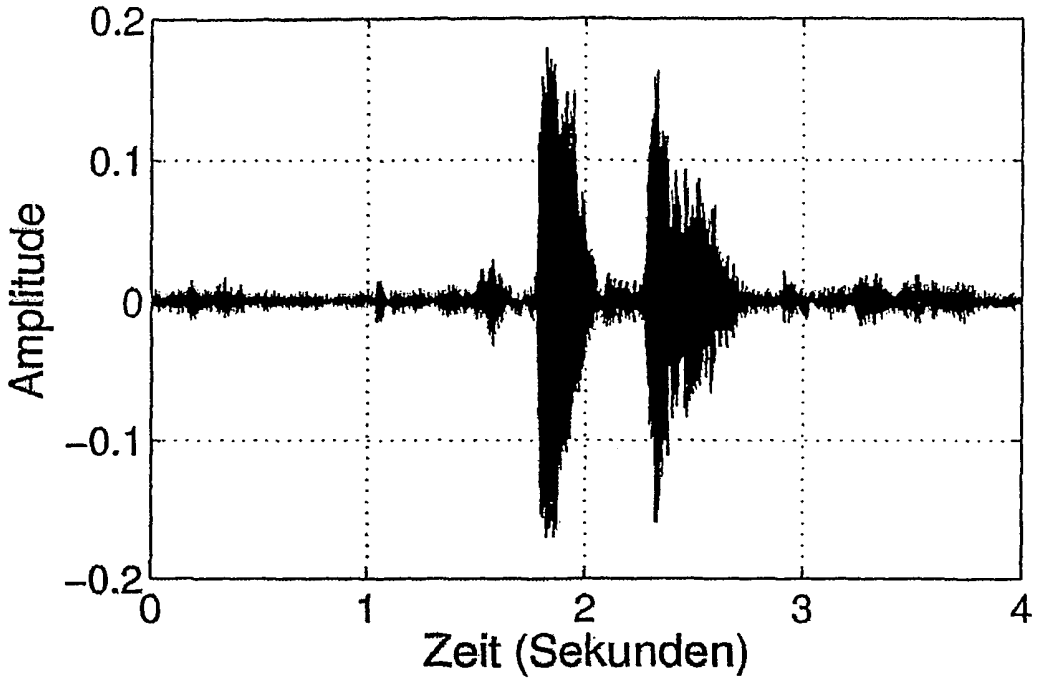

0.2
0.1
0
−0.1
−0.2
Amplitude
0
1
2
3
4
Zeit (Sekunden)

Fig. 10

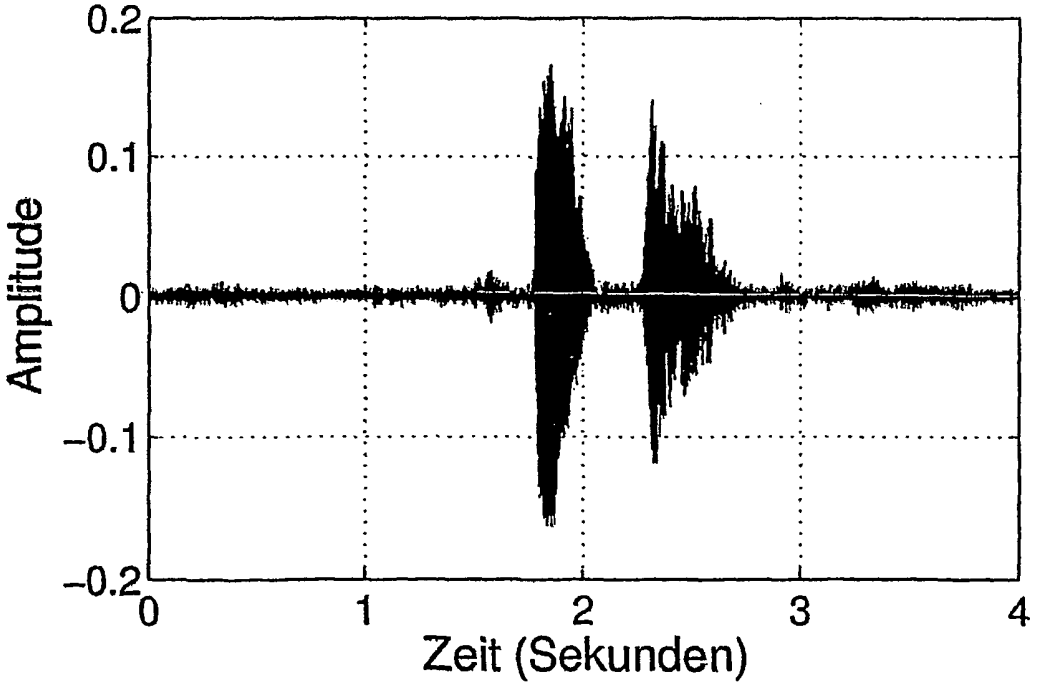

0.2
0.1
0
−0.1
−0.2
Amplitude
0
1
2
3
4
Zeit (Sekunden)

Fig. 11

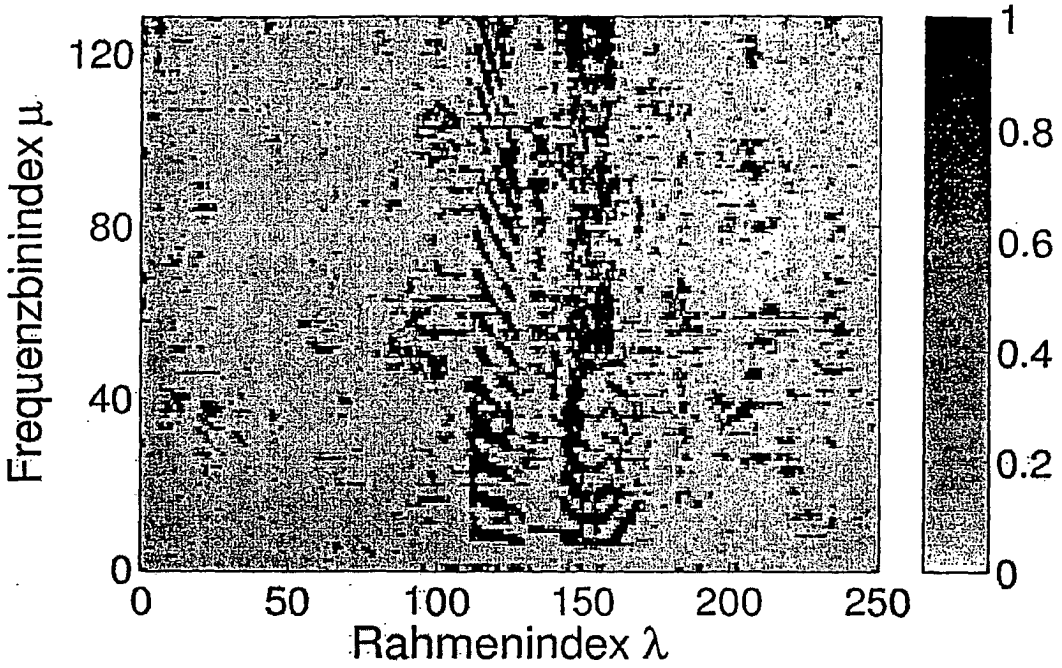

Frequenzbinindex μ
120
80
40
0
0
50
100
150
200
250
Rahmenindex λ
1
0.8
0.6
0.4
0.2
0

Fig. 12

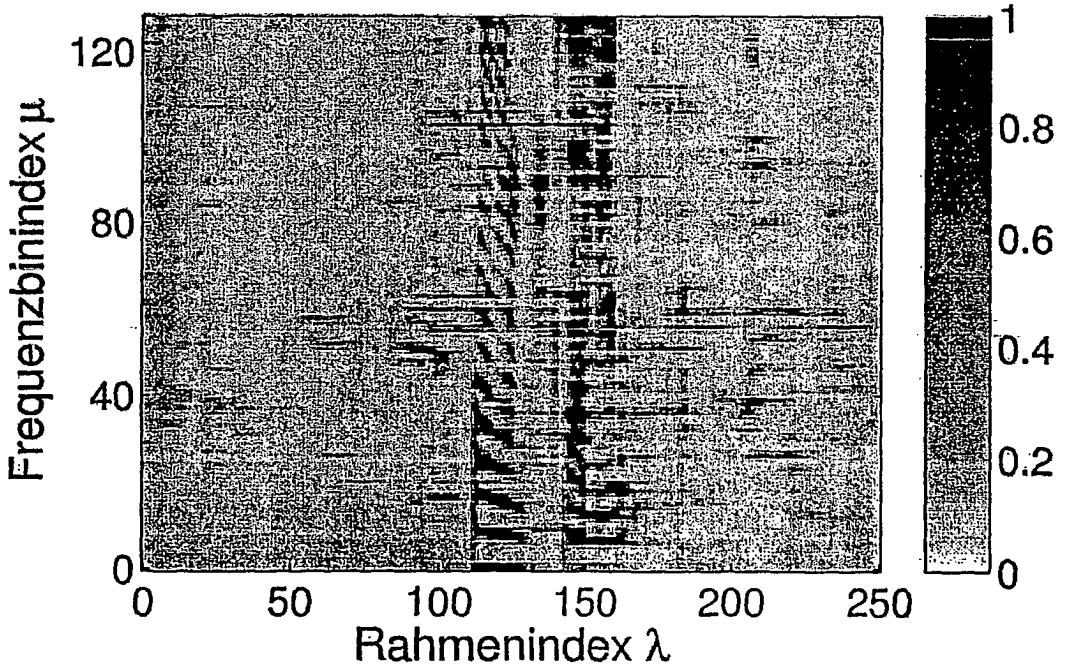

Frequenzbinindex μ
120
80
40
0
0
50
100
150
200
250
Rahmenindex λ
1
0.8
0.6
0.4
0.2
0

Fig. 13

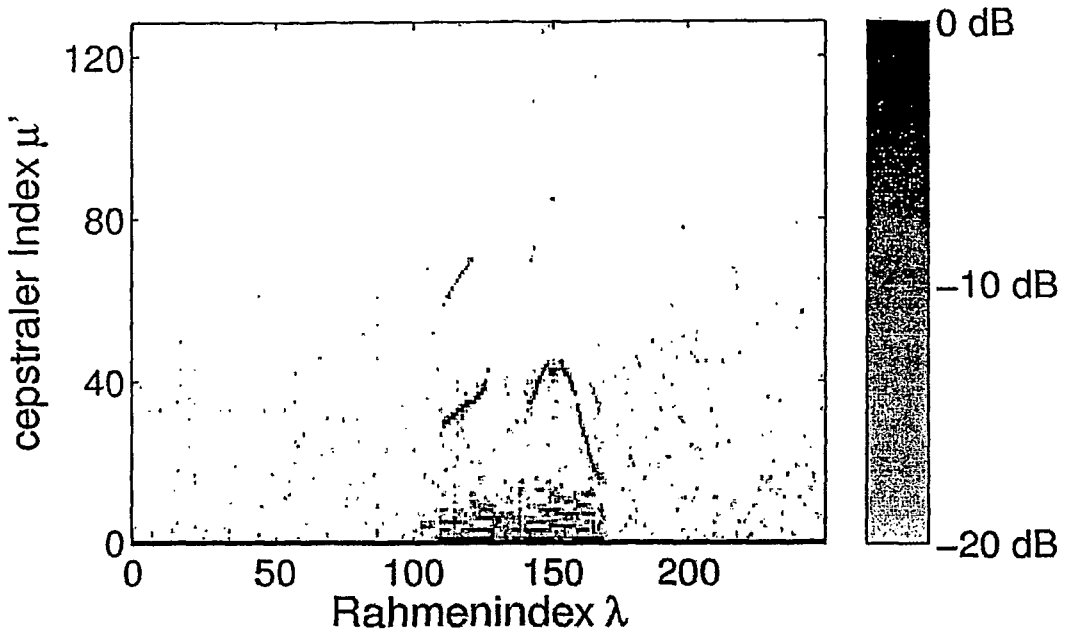

0 dB
−10 dB
−20 dB
120
80
40
0
0
50
100
150
200
cepstraler Index μ'
Rahmenindex λ

Fig. 14

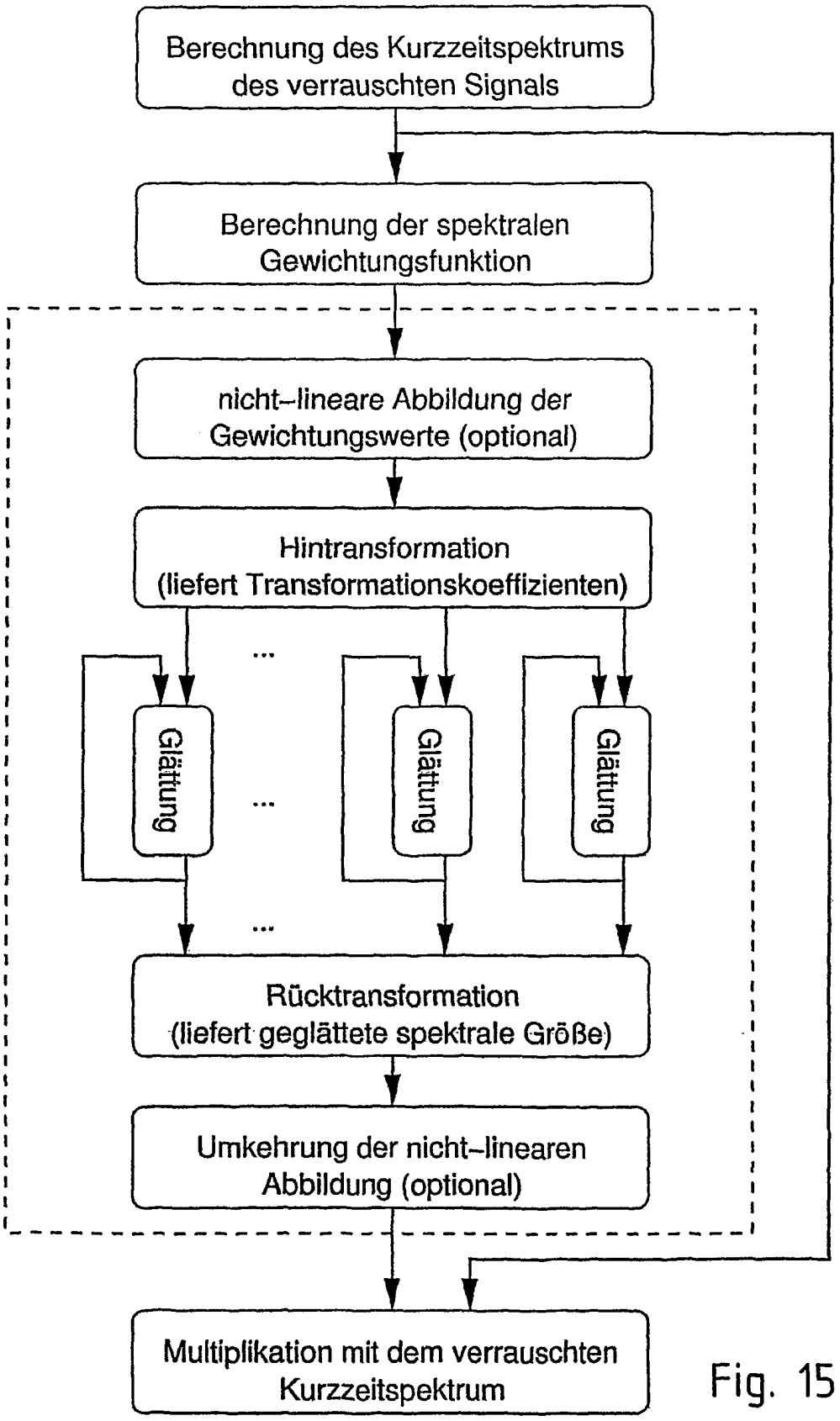

Berechnung des Kurzzeitspektrums des verrauschten Signals
Berechnung der spektralen Gewichtungsfunktion
nicht–lineare Abbildung der Gewichtungswerte (optional)
Hintransformation (liefert Transformationskoeffizienten)
Glättung
Glättung
Glättung
Rücktransformation (liefert geglättete spektrale Größe)
Umkehrung der nicht–linearen Abbildung (optional)
Multiplikation mit dem verrauschten Kurzzeitspektrum

Fig. 15

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 5365592 A **[0012]**

In der Beschreibung aufgeführte Nicht-Patentliteratur

- **Tim Fingscheidt ; Christophe Beaugeant ; Suhadi Suhadi.** Overcoming the statistical independence assumption w.r.t. frequency in speech enhancement. *Proceedings, IEEE Int. Conf. Acoustics, Speech, Signal Processing,* 2005, vol. 1, 1081-1084 **[0008]**
- **Harald Gustafsson ; Sven Erik Nordholm ; Ingvar Claesson.** Spectral subtraction using reduced delay convolution and adaptive averaging. *IEEE Transactions on Speech and Audio Processing,* November 2001, vol. 9 (8), 799-807 **[0008]**
- **Zenton Goh ; Kah-Chye Tan ; B.T.G. Tan.** Postprocessing method for suppressing musical noise generated by spectral subtraction. *IEEE Transactions on Speech and Audio Processing,* Mai 1998, vol. 6 (3), 287-292 **[0008]**
- **Andrzej Czyzewski.** *Multitask noisy speech enhancement system,* 2004 **[0009]**
- **Francois Thibault.** *High-level control of singing voice timbre transformations,* 2004 **[0009]**
- **Petre Stoica ; Niclas Sandgren.** Smoothed nonparametric spectral estimation via cepstrum thresholding. *IEEE Signal Processing Magazine,* November 2006, 34-45 **[0011]**