



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 602 18 385 T2** 2007.06.14

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 271 472 B1**

(21) Deutsches Aktenzeichen: **602 18 385.5**

(96) Europäisches Aktenzeichen: **02 013 983.8**

(96) Europäischer Anmeldetag: **25.06.2002**

(97) Erstveröffentlichung durch das EPA: **02.01.2003**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **28.02.2007**

(47) Veröffentlichungstag im Patentblatt: **14.06.2007**

(51) Int Cl.⁸: **G10L 19/14** (2006.01)
G10L 21/02 (2006.01)

(30) Unionspriorität:

896062 **29.06.2001** **US**

(73) Patentinhaber:

Microsoft Corp., Redmond, Wash., US

(74) Vertreter:

**Grünecker, Kinkeldey, Stockmair &
Schwanhäusser, 80538 München**

(84) Benannte Vertragsstaaten:

**AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT,
LI, LU, MC, NL, PT, SE, TR**

(72) Erfinder:

**Wang, Hong, Bellevue, Washington 98004, US;
Cuperman, Vladiir, Goleta, California 93117, US;
Gersho, Allen, Santa Barbara, California 93111,
US; Khalil, Hosam A., Bellevue, Washington
98005, US**

(54) Bezeichnung: **Nachfilterung von kodierter Sprache im Frequenzbereich**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

TECHNISCHES GEBIET

[0001] Diese Erfindung bezieht sich im allgemeinen auf die Technik der Signalfilterung zum Verbessern der Qualität eines Signals und insbesondere auf ein Verfahren zum Nachfiltern eines synthetisierten Sprachsignals, um ein Sprachsignal verbesserter Qualität zu erzeugen.

HINTERGRUND DER ERFINDUNG

[0002] Die Erzeugung elektronischer Signale ist in sämtlichen Bereichen der elektronischen und elektrischen Technologie gegenwärtig. Wenn ein elektrisches Signal verwendet wird, um eine Menge realer Wörter zu reproduzieren, ist die Qualität des Signals von Bedeutung. Sprache wird beispielsweise häufig über ein Mikrofon oder einen anderen Klangwandler empfangen und in eine elektrische Darstellung oder ein Signal umgewandelt. Zusätzlich zum künstlichen Rauschen, das als Artefakt dieser Umwandlung einfließt, kann zusätzlich anderes künstliches Rauschen in das Signal während der Sendung und der Codierung und/oder Decodierung einfließen. Ein derartiges Rauschen kann in vielen Fällen von den Menschen gehört werden und tatsächlich ein wiedergegebenes Sprachsignal bis zum Punkt der Ablenkung oder Störung des Zuhörers beherrschen.

[0003] Sprachcodierer, und insbesondere jene, die mit niedrigen Bitraten arbeiten, neigen dazu, Quantisierungsrauschen einfließen zu lassen, das hörbar sein und dadurch die Qualität der wiederhergestellten Sprache beeinträchtigen kann. Im allgemeinen wird ein Nachfilter verwendet, um das Rauschen in codierten Sprachsignalen zu maskieren, indem die Formanten und der Feinaufbau derartiger Signale verbessert werden. Normalerweise ist das Rauschen in Bereichen starker Formanten eines Signals nicht hörbar, wohingegen das Rauschen in Talbereichen eines Signals zwischen zwei benachbarten Formanten wahrnehmbar ist, da der Signalrauschabstand (SNR) in Talbereichen niedrig ist. Der SNR im Talbereich kann im Zusammenhang mit einem Codec einer niedrigen Bitrate sogar noch geringer sein, da die herrschenden Linearprädiktions-(LP-) Verfahren die Spitzen präziser darstellen als die Täler und die verfügbaren Bits unzureichend sind, um das Signal in den Tälern geeignet darzustellen. Daher ist es erwünscht, dass ein Sprach-Nachfilter die Täler abschwächt, während es die Spitzen beibehält, um den hörbaren Rauschpegel zu verringern.

[0004] Techniken des Standes der Technik beinhalten einen adaptiven Nachfilter-Algorithmus, der aus einem Pol-Null-Langzeit-Nachfilter besteht, das mit einem Kurzzeit-Nachfilter kaskadiert ist. Das Kurzzeit-Nachfilter WIRD aus den Parametern des

LP-Modells derart abgeleitet, dass es das Rauschen in den Spektraltälern abschwächt. Diese Parameter werden im allgemeinen als lineare Prädiktiv-Codierkoeffizienten oder LPC-Koeffizienten oder LPC-Parameter bezeichnet. Darüber hinaus wurde ein frequenzdomänenadaptiver Nachfilteralgorithmus zum Unterdrücken von Rauschen in den Spektraltälern eingeführt. Die zuvor erwähnten Nachfilteralgorithmen unterdrücken Rauschen ohne eine wesentliche Spektralverzerrung hervorzurufen, sind jedoch beim Verringern des wahrnehmbaren Rauschens in flachen anstelle von tiefen Tälern zwischen Formanten insbesondere im Zusammenhang mit Codierern einer niedrigen Bitrate, wie etwa jenen, die unter 8 kbps arbeiten, nicht wirksam. Eine Haupterklärung für diesen Nachteil ist, dass das Frequenzansprechverhalten des Nachfilters an sich nicht in geeigneter Weise dem detaillierten Feinaufbau der Spektralhülle folgt, was zur Maskierung der flachen Täler zwischen dicht beabstandeten Formanten führt.

[0005] Eine Typische Frühzeitdomänen-LPC-Nachfilterarchitektur ist in [Fig. 1](#) dargestellt. Ein Eingangsbitstrom, der vielleicht von einem Codierer gesendet wird, wird an einem Decoder **100** empfangen. Ein Bitstromdecoder **110**, der dem Decoder **100** zugeordnet ist, decodiert den eintreffenden Bitstrom. Dieser Schritt führt zu einer Zerlegung des Bitstroms in seine logischen Bestandteile oder virtuellen Kanalhalte. Der Bitstromdecoder **110** trennt beispielsweise LPC-Koeffizienten von einem codierten Anregungssignal für linearprädiktionsbasierte Codecs. Die decodierten LPC-Koeffizienten werden zu einem Formantenfilter **131** gesendet, das die erste Stufe eines Zeitdomänen-Nachfilters **130** ist. Ein synthetisiertes Sprachsignal, das von einem Sprach-Synthesizer **120** erzeugt wird, wird in ein Formant-Filter **131** gefolgt von einem Tonhöhen-Filter **132** eingegeben, in dem der harmonische Tonhöhenaufbau des Signals verbessert wird. Mit dem Tonhöhen-Filter ist ein Tilt-Kompensationsmodul **133** kaskadiert, das den Hintergrund-Tilt des Formant-Filters entfernt, um eine unerwünschte Verzerrung des Nachfilters zu vermeiden. Schließlich wird eine Gewinnsteuerung am Signal in einem Gewinn-Controller **134** angewendet, um eine Diskontinuität der Signalleistung in benachbarten Frames zu eliminieren.

[0006] KABAL P et al.: "Adaptive postfiltering for enhancement of noisy speech in the frequency domain", SIGNAL IMAGE AND VIDEO PROCESSING. SINGAPORE, 11. bis 14 Juni, 1991, PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, NEW YORK, IEEE, US, vol. 1 SYMP. 24, 11. Juni 1991, Seite 312 bis 315 beschreibt ein frequenzdomänen-adaptives Nachfilter für die Verbesserung von Sprache mit Nebengeräuschen. Das Nachfilter ist durch seine DFT-Koeffizienten $H(k)$ dargestellt, die mit $P(k)$ multipliziert werden, das eine abgewandelte Form von $X(k)$ ist (den

DFT-Koeffizienten der eingegebenen geräuschvollen Sprache $x(n)$). Das Filtern der eingegebenen Sprache erfolgt in der Frequenzdomäne. Eine umgekehrte DFT ergibt das Nachfiltersignal $y(n)$. Eine Näherung des Sprachspektrums erhält man durch Berechnen des Logarithmusgrößenspektrums von $1/A_p(z)$. Als erstes werden die LPC-Koeffizienten a_i und somit das Filter $A_p(z)$ bestimmt. Das Logarithmusgrößenspektrum ist $R(k) = -20\log_{10}|A_p(k)|$. Dies wird zum Identifizieren der Formanten verwendet. Das Auffinden der Amplitude und des Ortes der Formanten ist ein wichtiger Schritt bei der Bestimmung der Nachfilterkoeffizienten $H(k)$. Das Logarithmusgrößenspektrum $R(k)$ wird abgeändert, um zu $S(k)$ zu werden, so dass in der nachgefilterten Sprache die Formantenspitzen geschärft werden, die Spektraltäler vertieft werden und kein unerwünschter Tiefpass-Tilt vorhanden ist. Zunächst wird $R(k)$ in Abschnitte unterteilt. Jeder Abschnitt wird individuell abgeändert. Die Nachfilterkoeffizienten $H(k)$ müssen aus dem abgeänderten Logarithmusgrößenspektrum $S(k) = -20\log_{10}|H(k)|$ bestimmt werden. Die Phase von $H(k)$ ist dieselbe wie die Phase von $1/A_p(k)$. Die Nachfilterkoeffizienten erhält man durch Abändern lediglich der Größe des LPC-Spektrums.

[0007] US-A-S 890 108 beschreibt ein Nachfilter, das verwendet wird, um das Rauschen zu modellieren und die wahrnehmbare Qualität synthetisierter Sprache zu verbessern. Da Sprachformanten für die Wahrnehmung von weitaus größerer Bedeutung sind, als die Formant-Nullen, besteht die Idee darin, die Formantinformationen zu erhalten, indem das Rauschen in den Formantenbereichen so niedrig wie möglich gehalten wird. Der erste Schritt bei der Entwicklung des Frequenzdomänen-Nachfilters besteht darin, die gemessene Spektralhülle $R_\omega(\omega) = H(\omega)W(\omega)$ zu gewichten, um den Spektral-Tilt zu entfernen und eine ebenes d.h. ein flacheres, Spektrum zu erzeugen. Bei dieser Gleichung ist $H(\omega)$ die gemessene Spektralhülle und $W(\omega)$ die Gewichtungsfunktion. Die gewichtete Spektralhülle R_ω wird anschließend normalisiert, um einen Einheitsgewinn zu erzeugen, und als Potenz von β genommen. Wenn R_{\max} der Maximalwert der gewichteten Spektralhülle ist, wird das Nachfilter als Quotient zwischen $R_\omega(\omega)$ und R_{\max} genommen und mit β potenziert, wobei β im Bereich zwischen 0 und 1 liegt. Das geschätzte Nachfilter-Frequenzansprechverhalten wird anschließend verwendet, um die ursprüngliche Sprachhülle zu gewichten, um $H(\omega) = P_f(\omega)H(\omega)$ zu ergeben.

[0008] Gemäß WO 00 11655 A durchläuft synthetisierte, Sprache $\bar{s}(n)$ ein adaptives Nachfilter. Das adaptive Nachfilter ist eine Kaskade aus drei Filtern: ein Formant-Nachfilter und zwei Tilt-Kompensationsfilter. Das Formant-Nachfilter ist durch $H_f(z)$ gegeben, das gleich dem Verhältnis zwischen $\bar{A}(z/\gamma_n)$ und $\bar{A}(z/\gamma_d)$ ist, wobei $\bar{A}(z)$ das empfangene, quantisierte und interpolierte LPC-Umkehrfilter ist und γ_n sowie γ_d den

Umfang der Formant-Nachfilterung steuern. Das erste Tilt-Kompensationsfilter $H_{t1}(z)$ kompensiert den Tilt im Formant-Nachfilter $H_f(z)$ und ist durch $H_{t1}(z) = (1-\mu z^{-1})$ gegeben, wobei μ ein Tilt-Faktor ist. Der Nachfiltervorgang wird wie folgt ausgeführt. Zunächst wird die synthetisierte Sprache $\bar{s}(n)$ durch $\bar{A}(z/\gamma_n)$ umkehrgefiltert, um das Restsignal $r(n)$ zu erzeugen. Dieses Signal wird mit dem Synthesefilter $1/\bar{A}(z/\gamma_d)$ gefiltert und an das erste Tilt-Kompensationsfilter $H_{t1}(z)$ weitergeleitet, was zum nachgefilterten Sprachsignal $s_f(n)$ führt.

[0009] Das Frequenzansprechverhalten der Nachfilterarchitektur, die bei Sprachnachfiltersystemen des Standes der Technik dargestellt ist, folgt weder in geeigneter Weise dem feinen detaillierten Aufbau des Sprachspektrums, noch löst es immer die Spektralhüllen-Spitzen und -Täler in geeigneter Weise auf.

ÜBERSICHT ÜBER DIE ERFINDUNG

[0010] Das Ziel der vorliegenden Erfindung besteht darin, ein Verfahren und eine Vorrichtung zum Nachfiltern eines Sprachsignals anzugeben, mit denen es möglich ist, Änderungen der Kanaleigenschaften adaptiv zu berücksichtigen.

[0011] Dieses Ziel wird mit der Erfindung erreicht, wie sie in den unabhängigen Ansprüchen definiert ist.

[0012] Ausführungsformen sind in den abhängigen Ansprüchen beschrieben.

[0013] Eine Ausführungsform gibt ein Verfahren zum Nachfiltern in der Frequenzdomäne an, wobei das Nachfilter aus dem LPC-Spektrum abgeleitet wird. Um den Spektralaufbau zu verbessern, wird weiterhin eine nicht lineare Transformation des LPC-Spektrums angewendet, um das Nachfilter abzuleiten. Um eine ungleichmäßige Spektraldehnung infolge einer nicht linearen Transformation des Hintergrund-Spektral-Tilts zu vermeiden, werden eine Berechnung und eine Kompensation vorzugsweise vor der Anwendung des Formant-Nachfilters durchgeführt. Um schließlich ein Aliasing zu vermeiden, gibt die vorliegende Erfindung eine Anti-Aliasing-Prozedur in der Zeitdomäne an. Anfängliche Anwendungsergebnisse haben gezeigt, dass dieses Verfahren die Signalqualität insbesondere für jene Abschnitte des Signals deutlich verbessert, die Bereichen des Sprachspektrums mit geringerer Leistung zugeordnet werden können.

[0014] Im allgemeinen kann die Signalfilterung der Sprache und anderer Signale in der Zeitdomäne oder der Frequenzdomäne ausgeführt werden. In der Zeitdomäne ist die Filterverwendung äquivalent zur Ausführung einer Faltung, die einen Vektor, der repräsentativ für das Signal ist, und einen Vektor kombiniert, der jeweils für das Impulsansprechverhalten reprä-

sentativ ist, um einen dritten Vektor zu erzeugen, der dem gefilterten Signal entspricht. Im Gegensatz dazu ist in der Frequenzdomäne der Vorgang des Anwendens eines Filters auf das Signal äquivalent zu einer einfachen Multiplikation des Spektrums des Signals mit dem des Filters. Wenn das Spektrum des Filters das Spektrum des Signals im Detail beibehält, behält somit das Filtern des Signals den feinen Aufbau und die Formanten des Signals bei. Insbesondere wird ein Tal, das im Sprachspektrum vorhanden ist, weder niemals vollständig aus dem gefilterten Spektrum verschwinden, noch wird es in eine lokale Spitze anstelle eines Tals umgewandelt. Der Grund dafür ist, dass die Beschaffenheit des Nachfilters der Erfindung die Anordnung der Punkte im Spektrum beibehält; ein Spektralpunkt, der größer ist als sein Nachbar im vorgefilterten Spektrum bleibt im gefilterten Spektrum größer, wenngleich sich der Grad des Unterschiedes zwischen beiden infolge des Filters ändern kann.

[0015] Somit verwendet das Nachfilter, das hier beschrieben ist, ein Frequenzansprechverhalten, das den Spitzen und Tälern der Spektralhülle des Signals folgt, ohne dass ein gesamter Spektral-Tilt erzeugt wird. Ein derartiges Nachfilter kann vorteilhaft in einer Vielfalt technischer Anwendungen verwendet werden, die die Mobiltelefon-Sende- und Empfangstechnologie, die Internet-Medientechnologie und andere Speicher- oder Sende- oder Empfangsanwendungen umfassen, bei denen Codecs mit niedriger Bitrate beteiligt sind.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0016] [Fig. 1](#) ist eine schematische Darstellung, die eine typische Zeitdomänen-Nachfilter-Architektur des Standes der Technik darstellt;

[0017] [Fig. 2](#) ist eine Darstellung der Architektur netzwerkverknüpfter Codecs;

[0018] [Fig. 3](#) ist eine vereinfachte schematische Darstellungen des Aufbaus eines Frequenzdomänen-Nachfilters gemäß einer Ausführungsform der Erfindung;

[0019] [Fig. 4a](#), [Fig. 4b](#) und [Fig. 4c](#) sind schematische Darstellungen, die Bestandteile eines Frequenzdomänen-Formantfilters gemäß einer Ausführungsform der Erfindung zeigen;

[0020] [Fig. 5a](#) und [Fig. 5b](#) sind schematische Darstellungen des Aufbaus von Bestandteilen eines Frequenzdomänen-Formantfilters gemäß einer alternativen Ausführungsform der Erfindung;

[0021] [Fig. 6a](#) und [Fig. 6b](#) sind Flussdiagramme, die Schritte darstellen, die bei der Ausführung einer Nachfilterung gemäß einer Ausführungsform der Erfindung durchgeführt werden; und

[0022] [Fig. 7](#) ist eine vereinfachte schematische Darstellung der Architektur einer Berechnungsvorrichtung, die bei einer Berechnungsvorrichtung Verwendung findet, bei der eine Ausführungsform der Erfindung ausgeführt werden kann.

BESCHREIBUNG DER BEVORZUGTEN AUSFÜHRUNGSFORMEN

[0023] Die vorliegende Erfindung bezieht sich auf ein Verfahren und ein System zum Ausführen einer Nachfilterung zur Verbesserung der Sprachqualität, bei denen ein Nachfilter aus einer nicht linearen Transformation eines Satzes von LPC-Koeffizienten in der Frequenzdomäne abgeleitet wird. Das abgeleitete Nachfilter wird durch Multiplizieren des synthetisierten Sprachsignals mit Formantfiltergewinnen in der Frequenzdomäne angewendet. Bei einer Ausführungsform wird die Erfindung in einem Decoder zum Nachfiltern eines synthetisierten Sprachsignals verwendet. Gemäß alternativen Ausführungsformen der Erfindung können die LPC-Koeffizienten, die verwendet werden, um das Nachfilter abzuleiten, von einem Codierer gesendet werden oder können unabhängig aus der synthetisierten Sprache im Decoder abgeleitet werden.

[0024] Wenngleich dies nicht erforderlich ist, kann die vorliegende Erfindung mit Hilfe von Anweisungen, wie etwa Programmmodulen, angewendet werden, die von einem Computer ausgeführt werden. Im allgemeinen umfassen Programmmodule Routinen, Objekte, Komponenten, Datenstrukturen und dergleichen, die spezielle Aufgaben ausführen oder spezielle abstrakte Datentypen implementieren. Der Begriff "Programm" beinhaltet ein oder mehrere Programmmodule.

[0025] Die Erfindung kann bei einer Vielfalt von Gerätetypen, wie etwa Mobiltelefonen, PCs, Handgeräten, Multiprozessorsystemen, mikroprozessorbasierten programmierbaren Verbraucher-Elektronikgeräten, Netzwerk-PCs, Minicomputern, Großrechnern und dergleichen, eingesetzt werden. Die Erfindung kann zudem in einem verteilten System verwendet werden, bei dem Aufgaben von Komponenten ausgeführt werden, die über ein Kommunikationsnetzwerk verknüpft sind. In einem verteilten System können sich zusammenarbeitende Module sowohl an lokalen als auch an entfernten Orten befinden.

[0026] Ein beispielhaftes Telefonesystem, bei dem eine Ausführungsform der Erfindung verwendet werden kann, ist unter Bezugnahme auf [Fig. 2](#) beschrieben. Das Telefonesystem enthält Codecs **200**, **220**, die miteinander über ein Netzwerk **210** kommunizieren, das mit einer Wolke dargestellt ist. Das Netzwerk **210** kann hinlänglich bekannte Komponenten, wie etwa Router, Gateways, Hubs, etc., enthalten und es den Codecs **200** gestatten, über ein drahtgebunde-

nes und/oder ein drahtloses Medium zu kommunizieren. Jeder Codec **200**, **220** enthält im wesentlichen einen Codierer **201**, einen Decodierer **202** und ein Nachfilter **203**.

[0027] Die Codecs **200** und **220** enthalten vorzugsweise zudem eine Kommunikationsverbindung oder sind dieser zugeordnet, wobei diese es einer Host-Vorrichtung gestattet, mit anderen Vorrichtung zu kommunizieren. Eine Kommunikationsverbindung ist ein Beispiel eines Kommunikationsmediums. Kommunikationsmedien enthalten normalerweise computerlesbare Anweisungen, Datenstrukturen, Programmmodule oder andere Daten in einem modulierten Datensignal, wie etwa einer Trägerwelle oder einem anderen Transportmechanismus, und beinhalten beliebige Informationszustellungsmedien. Der Begriff computerlesbare Medien, wie er hier verwendet wird, beinhaltet sowohl Speichermedien als auch Kommunikationsmedien. Die Codec-Elemente, die hier beschrieben sind, können sich vollständig auf einem computerlesbaren Medium befinden. Die Codecs **200** und **220** können auch Eingabe- und Ausgabevorrichtungen zugeordnet sein, wie es allgemein später in dieser Beschreibung erläutert wird.

[0028] Unter Bezugnahme auf [Fig. 3](#) ist ein exemplarisches Nachfilter **303** dargestellt, bei dem das System verwendet werden kann, das hier beschrieben ist. In seiner einfachsten Ausführung verwendet das Nachfilter **303** ein synthetisiertes Eingangssprachsignal $S(n)$ und LPC-Koeffizienten α in Verbindung mit einem Frequenzdomänen-Formantfilter **310**. Das Nachfilter kann zudem über zusätzliche Merkmale oder eine zusätzliche Funktionalität verfügen. Beispielsweise werden vorzugsweise ein Tönhöhenfilter **320** und ein Gewinncontroller **330** eingesetzt und verwendet, wie es im folgenden beschrieben wird.

[0029] Es ist bekannt, dass die Codierung und Decodierung eines Sprachsignals normalerweise unerwünschtes Rauschen im Signal hervorrufen wird. Im Signalfrequenzspektrum überlappt derartiges Rauschen das Sprachsignal und kann insbesondere in den Talbereichen zwischen aufeinanderfolgenden Formanten vom Menschen gehört werden. Ein in geeigneter Weise aufgebautes und angewendetes Nachfilter ist beim Entfernen dieses unerwünschten Rauschens hilfreich. Ein ideales Nachfilter ist derart beschaffen, dass es ein Frequenzansprechverhalten aufweist, das dem Frequenzspektrum des Signals von Interesse folgt. Die meisten Codecs basieren auf dem Prinzip der linearen Prädiktion, wobei die Koeffizienten der linearen Prädiktion dem Signalfrequenzspektrum folgen. Zusätzlich zu anderen innovativen Prozeduren, die zu beschreiben sind, nutzt die Erfindung diese Beziehung vorteilhaft aus, um ein Sprachnachfilter abzuleiten, wengleich die Erfindung darüber hinaus die unabhängige Erzeugung von LPC-Parametern gestattet.

[0030] Es gibt eine große Vielfalt von Möglichkeiten, mit denen die Frequenzdomänen-Nachfilterung in Übereinstimmung mit der Erfindung ausgeführt werden kann. Gemäß einer Ausführungsform wird die Frequenzdomänen-Nachfilterung sequentiell innerhalb des Nachfilters ausgeführt. Unter Bezugnahme auf [Fig. 4a](#) enthält das Frequenzdomänen-Formantfilter **410** ein Fouriertransformations-Modul **411**, ein Formant-Filtermodul **412** und ein Modul **413** für inverse Fouriertransformation. Das Fouriertransformations-Modul und das Modul für inverse Fouriertransformation sind für das Formant-Filtermodul **412** verfügbar, um Signale zwischen der Zeitdomäne und der Frequenzdomäne zu übertragen, wie es der Fachmann verstehen wird. Die Fouriertransformation und die inverse Fouriertransformation der Transformations-Module **411** und **413** werden vorzugsweise gemäß der herkömmlichen diskreten Fouriertransformation (DFT) ausgeführt.

[0031] Das Formant-Filtermodul **412** erzeugt Frequenzdomänen-Gewinne und filtert das synthetisierte Eingangssprachsignal durch Anwenden der erzeugten Gewinne bevor das entsprechende Signal in die Zeitdomäne zurücktransformiert wird. [Fig. 4b](#) zeigt weiterhin die Bestandteile des Formant-Filtermoduls **412**, das ein LPC-Tilt-Berechnungsmodul **415**, ein LPC-Tilt-Kompensationsmodul **420**, ein Gewinn-Berechnungsmodul **430** und ein Gewinn-Anwendungsmodul **440** enthält. Der Betrieb dieser Module wird im folgenden detaillierter unter Bezugnahme auf [Fig. 6](#) erläutert, hier jedoch ebenfalls kurz beschrieben.

[0032] Im allgemeinen hat ein codiertes LPC-Spektrum einen Hintergrund mit Tilt. Dieser Tilt kann zu einer inakzeptablen Signalverzerrung führen, wenn er verwendet wird, um das Nachfilter ohne Tilt-Kompensation zu berechnen. Insbesondere könnte dieser Hintergrund mit Tilt während der Nachfilterung unerwünscht verstärkt werden, wenn das Nachfilter eine nicht lineare Transformation beinhaltet, wie dies bei der vorliegenden Ausführungsform der Fall ist. Die Anwendung einer derartigen Transformation auf ein Spektrum mit Tilt würde die Wirkung einer nicht linearen Transformation auch des Tilt haben, wodurch es größere Schwierigkeiten bereitet, später ein geeignetes Spektrum ohne Tilt zu erhalten. Somit ist es vorzuziehen, den Hintergrund-Tilt des Spektrums von der nicht linearen Transformation zu entfernen. Gemäß dieser Erfindung entfernt das Tilt-Kompensationsmodul **420** in geeigneter Weise den Hintergrund mit Tilt gemäß dem Tilt der vom LPC-Spektrums-Tilt-Berechnungsmodul **415** geschätzt wird.

[0033] Das Gewinn-Berechnungsmodul **430** berechnet die Frequenzdomänen-Formantfiltergewinne einschließlich der Größe und des Phasenansprechverhaltens. Zu diesem Zeitpunkt wendet das Gewinn-Anwendungsmodul **440** die Gewinne multiplika-

tiv auf das Sprachsignal in der Frequenzdomäne an.

[0034] Unter Bezugnahme auf [Fig. 4c](#) enthält das Gewinn-Berechnungsmodul ein Zeitdomänen-LPC-Darstellungsmodul **431**, ein Modelliermodul **432** ein Modul **433** für nicht lineare LPC-Transformation, ein Phasenberechnungsmodul **434**, ein Gewinn-Kombinationsmodul **435** und ein Anti-Aliasing-Modul **436**.

[0035] Das LPC-Darstellungsmodul **431** erzeugt eine Zeitdomänen-Vektordarstellung des LPC-Spektrums, worauf der Vektor in die Frequenzdomäne zur weiteren Verarbeitung transformiert wird. Das Modelliermodul **432** modelliert den Frequenzdomänen-Vektor auf der Basis eines einer Vielzahl geeigneter Modelle, die dem Fachmann bekannt sind. Bei einer Ausführungsform wird die Inversion des LPC-Spektrums verwendet, um die Gewinne zu berechnen.

[0036] Das Modul **433** für die nicht lineare LPC-Transformation berechnet die Größe der Formantfiltergewinne durch Ausführen einer nicht linearen Transformation der Größe des inversen LPC-Spektrums. Gemäß einer Ausführungsform der Erfindung wird eine Skalierfunktion mit einem Skalierungsfaktor zwischen 0 und 1 als nicht lineare Transformationsfunktion verwendet, wie es im folgenden detaillierter beschrieben wird. Die Parameter in der Skalierfunktion sind gemäß dynamischer Umgebungen, wie etwa gemäß dem Typ des Eingangssprachsignals und der Codierate, einstellbar. Das Phasenberechnungsmodul **434** berechnet das Phasenansprechverhalten für die Formantfiltergewinne. Gemäß einer Ausführungsform berechnet das Phasenberechnungsmodul **434** das Phasenansprechverhalten über die Hilbert-Transformation, im besonderen über den Hilbert-Phasenschieber. Andere Phasenberechnungseinrichtungen, wie etwa die Anwendung der Kotangenten-Transformation der Hilbert-Transformation, können alternativ verwendet werden. Mit Hilfe der Größe und der Phase der Formantfiltergewinne, die vom Modul **433** für die nicht lineare LPC-Transformation und dem Phasenberechnungsmodul **434** bereitgestellt werden, erzeugt das Gewinn-Kombinationsmodul **435** die Gewinne in der Frequenzdomäne. Ein Anti-Aliasing-Modul **436** ist vorzugsweise vorgesehen, um ein Aliasing zu vermeiden, wenn das Signal nachgefiltert wird. Es wird bevorzugt, dass der Anti-Aliasing-Vorgang in der Zeitdomäne ausgeführt wird, wobei dies jedoch nicht von Bedeutung ist.

[0037] Gemäß der Erfindung wird das Frequenzdomänen-Nachfilter aus dem LPC-Spektrum abgeleitet und erzeugt beispielsweise die Frequenzdomänen-Formantgewinne, wobei die Ableitung eine Abfolge mathematischer Prozeduren beinhaltet. Es kann gewünscht sein, eine separate Berechnungsein-

heit vorzusehen, die für die gesamte oder einen Teil der mathematischen Verarbeitung verantwortlich ist. Bei einer weiteren Ausführungsform der Erfindung ist eine separate LPC-Bewertungseinheit vorgesehen, um die LPC-Koeffizienten abzuleiten, wie es in [Fig. 5](#) gezeigt ist.

[0038] Unter Bezugnahme auf [Fig. 5](#) enthält das Frequenzdomänen-Formantfilter **500** ein Fouriertransformations-Modul **511**, ein Modul **513** für die inverse Fouriertransformation, ein Gewinn-Anwendungsmodul **540** und eine LPC-Bewertungseinheit **521**. Das Fouriertransformations-Modul **511**, das Modul **513** für eine inverse Fouriertransformation und das Gewinn-Anwendungsmodul **540** können dieselben sein, wie die Module, auf die durch ähnliche Ziffern in [Fig. 4](#) Bezug genommen wurde. Gemäß der Erfindung enthält die LPC-Bewertungseinheit **521** ein LPC-Tilt-Berechnungsmodul **510**, ein LPC-Tilt-Kompensationsmodul **520** und ein Gewinn-Berechnungsmodul **530**, wobei diese Bestandteile dieselben sein können, wie die Komponenten, auf die durch ähnliche Ziffern in [Fig. 4](#) Bezug genommen wurde.

[0039] Was den Betrieb angeht, so weicht die alternative Ausführungsform, die in [Fig. 5](#) beschrieben ist, geringfügig von der Ausführungsform ab, die mit Hilfe von [Fig. 4](#) dargestellt ist. Insbesondere empfängt das Gewinn-Anwendungsmodul **540** als Eingabe ein synthetisiertes Sprachsignal und stellt als Ausgabe ein gefiltertes synthetisiertes Sprachsignal bereit. Das Fouriertransformations-Modul **511** und das Modul **513** für eine inverse Fouriertransformation sind für das Gewinn-Anwendungsmodul zur Transformation des vorgefilterten Sprachsignals in die Frequenzdomäne und zur Transformation des nachgefilterten Sprachsignals in die Zeitdomäne verfügbar. Die LPC-Bewertungseinheit **521** empfängt oder berechnet die LPC-Koeffizienten, greift auf die Transformations-Module **511** und **513** nach Erfordernis für die Transformation zwischen der Zeit- und der Frequenzdomäne zu und gibt berechnete Gewinne an das Gewinn-Anwendungsmodul **540** zurück.

[0040] Unter Bezugnahme auf [Fig. 6a](#) und [Fig. 6b](#) sind beispielhafte Schritte dargestellt, die unternommen werden, um die Nachfilterung gemäß einer Ausführungsform der Erfindung auszuführen. Das synthetisierte Sprachsignal $\hat{S}(n)$ und die LPC-Koeffizienten α_i werden in Schritt **601** empfangen. Da ein codiertes LPC-Spektrum normalerweise einen Hintergrund mit Tilt hat, der bei der direkten Verwendung zur Berechnung des Formant-Nachfilters eine zusätzliche Verzerrung hervorruft, ist es vorzuziehen, einen Spektral-Tilt zuerst zu berechnen und zu korrigieren. Ein unkorrigierter Tilt kann während der Berechnung des Nachfilters unerwünscht verstärkt werden, insbesondere dann, wenn eine derartige Berechnung eine nicht lineare Transformation beinhaltet. Demzufolge wird bei Schritt **603** bzw. **605** der

LPC-Spektral-Tilt berechnet und das Spektrum dafür kompensiert. Beispielhafte mathematische Prozeduren, die geeignet sind, um diese Schritte auszuführen, sind wie folgt. Der Fachmann wird erkennen, dass die folgenden mathematischen Prozeduren hinsichtlich Anordnung und Detail abgeändert werden können und trotzdem zum selben Ergebnis führen. Für die LPC-Koeffizienten α_i ($i = 0, 1, \dots, P$ und $\alpha_0 = 1$), wobei P die Größenordnung der PLC-Polynomkoeffizienten ist, ist der Tilt μ des LPC-Spektrums definiert als:

$$\mu = \frac{R(1)}{R(0)}$$

wobei $R(1)$ und $R(0)$ Autokorrelationswerte der LPC-Parameter sind, die definiert sind durch

$$R(\tau) = \sum_{i=0}^{i=P-\tau} \alpha_i \alpha_{i+\tau} \quad \tau = 0, 1$$

[0041] Die LPC-Größenordnung P wird in Abhängigkeit der Abtastfrequenz gewählt, wie es dem Fachmann verständlich sein wird. Bei dieser Ausführungsform wird $P = 10$ für 8 kHz- und 11,025 kHz-Abtastraten verwendet, während $P = 16$ für 16 kHz- und 22,05 kHz-Abtastraten verwendet wird. Mit der Vorgabe des berechneten Tilt μ werden die LPC-Koeffizienten α_i wie folgt kompensiert:

$$\alpha_i = \begin{cases} \alpha_0 & i = 0 \\ \alpha_i - 0,7\mu\alpha_{i-1} & i = 1, \dots, p \\ -0,7\mu\alpha_p & i = p + 1 \end{cases}$$

[0042] Bei Schritt **607** erhält man eine Vektordarstellung des tilt-kompensierten LPC α_i , die mit A gekennzeichnet ist, in der Zeitdomäne durch Zero-Padding, um einen Vektor einer geeigneten Größe auszubilden. Eine beispielhafte Länge für einen derartigen Vektor ist **128**, wenngleich andere ähnliche oder stark unterschiedliche Vektorlängen in gleicher Weise verwendet werden können.

[0043] Bei den Schritten **609** bis **623** werden die Formant-Nachfiltergewinne einschließlich der Größe und des Phasenansprechverhaltens berechnet. Insbesondere wird bei Schritt **609** der Vektor A in einen Frequenzdomänen-Vektor $A'(k)$ über eine Fouriertransformation umgewandelt. Bei Schritt **613** wird der Frequenzdomänen-Vektor $A'(k)$ durch Invertieren der Größe von $A'(k)$ und Umwandeln in ein Logarithmusmaß (dB) abgeändert. Die Transferfunktion gemäß diesem Schritt ist mit $H(k)$ gekennzeichnet. Zur mathematischen Effizienz und Dienlichkeit wird $H(k)$ zuerst in Schritt **615** zu $\hat{H}(k)$ normalisiert, wie es im folgenden Beispiel gezeigt ist:

$$\hat{H}(k) = \frac{H(k) - H_{\min}(k)}{H_{\max}(k) - H_{\min}(k)} + 0,1$$

wobei $H_{\max}(k)$ und $H_{\min}(k)$ den Maximal- bzw. den Minimalwert von $H(k)$ darstellen.

[0044] In Schritt **615** wird die normalisierte Funktion $\hat{H}(k)$ durch eine Skalierfunktion nicht linear transformiert, wie es im folgenden dargestellt ist:

$$T(k) = g \left| \hat{H}(k) \right|^\gamma, \quad g = \frac{\ln 10}{20c} (H_{\max} - H_{\min})$$

wobei c eine Konstante ist. Ein beispielhafter Wert von c ist 1,47 für ein gesprochenes Signal und 1,3 für ein nicht gesprochenes Signal. Der Skalierfaktor γ kann in Übereinstimmung mit dynamischen Umgebungsbedingungen eingestellt werden. Beispielsweise können unterschiedliche Typen von Sprachcodierern und Codierraten im Optimalfall unterschiedliche Werte für diese Konstante verwenden. Ein beispielhafter Wert für den Skalierfaktor γ ist 0,25, wenngleich andere Skalierfaktoren akzeptable oder bessere Ergebnisse erbringen können. Obwohl die vorliegende Erfindung so beschrieben wurde, dass die obige Skalierfunktion für den Schritt der nicht linearen Transformation verwendet wird, können andere nicht lineare Transformationsfunktionen alternativ verwendet werden. Derartige Funktionen beinhalten geeignete Exponentialfunktionen und Polynomfunktionen.

[0045] Die Funktion $T(k)$, die man in Schritt **615** erhält, wird anschließend verwendet, um das Phasenansprechverhalten des Gewinns zu schätzen. Gemäß der Erfindung wenden die Schritte **617** bis **623** den Hilbert-Phasenschieber an, um das Phasenansprechverhalten $\theta(k)$ des Gewinns zu berechnen. Insbesondere wird bei Schritt **617** die Funktion $T(k)$ durch die Fouriertransformation in die Zeitdomäne transferiert, da der Hilbert-Phasenschieber in der Zeitdomäne ausgeführt wird. Bei Schritt **619** erhält man das Phasenansprechverhalten $\theta(n)$ durch Multiplizieren von $T(n)$ mit j , wobei j definiert ist als $j^2 = -1$. Bei Schritt **621** werden die berechneten Phasenansprechverhalten der Gewinne $\theta(n)$ in das Phasenansprechverhalten $\theta(k)$ der Frequenzdomäne zur weiteren Verarbeitung in der Frequenzdomäne transformiert.

[0046] Bei Schritt **623** erhält man den Frequenzdomänen-Formantfiltergewinn $F(k)$ durch Kombinieren der Größen- und der Phasenkomponenten wie folgt:

$$F(k) = L(k) e^{j\theta(k)}, \quad L(k) = 10^{\frac{g}{20} T(k)}$$

wobei q und g Konstanten sind, die wie folgt definiert sind:

$$q = \frac{H_{\max} - H_{\min}}{20c}, g = \frac{\ln 10}{20c} (H_{\max} - H_{\min})$$

wobei \ln der natürliche Logarithmus ist.

[0047] Die Schritte **625** bis **631** werden ausgeführt, um ein Anti-Aliasing in der Zeitdomäne durchzuführen. Insbesondere wird in Schritt **625** der Frequenzdomänen-Gewinn $F(k)$ in einen Zeitdomänen-Gewinn $f(n)$ durch die Ausführung einer inversen Fouriertransformation umgewandelt. Das heißt, die inverse Fouriertransformation von $F(k)$ gleicht $f(n)$. In Schritt **627** wird eine zweite Funktion $g(n)$ durch Nullsetzen der Koeffizienten von $f(n)$ gemäß der Länge N der Fouriertransformation und der Länge M des Eingangssprachsegments wie folgt definiert:

$$g(n) = \begin{cases} f(n) & n = 0, 1, \dots, N-M \\ 0 & n > N-M \end{cases}$$

[0048] Der Schritt **629** bedingt das Anwenden einer Standard-Normalisierungsprozedur auf $g(n)$ wie folgt:

$$g_n(n) = \frac{g(n)}{\sqrt{\sum_{n=0}^{N-M} g^2(n)}}$$

[0049] Schließlich erhält man den Frequenzdomänen-Gewinn $G(k)$ nach dem Anti-Aliasing durch Transferieren der Zeitdomänenfunktion $g_n(n)$ in die Frequenzdomäne durch eine Fouriertransformation in Schritt **631**. Das heißt, die Fouriertransformation von $g_n(n)$ gleicht $G(k)$.

[0050] Nach Berechnung des Frequenzdomänen-Formantgewinns $G(k)$, werden die Schritte **633** bis **637** ausgeführt, um die Filterung des synthetisierten Eingangssprachsignals $\hat{S}(n)$ zu bewirken. Insbesondere wird in Schritt **633** das Signal $\hat{S}(n)$ zuerst in ein Frequenzdomänensignal $\hat{S}(k)$ transferiert. Ruft man in sich in Erinnerung, dass die Nachfilterung in der Frequenzdomäne durch Multiplikation des Signals mit einem Gewinn für jede Frequenz angewendet wird, wird $\hat{S}(k)$ in Schritt **635** mit den Frequenzdomänen-Formantgewinnen $G(k)$ multipliziert, wodurch man das nachgefilterte Sprachsignal $\hat{S}'(k)$ erhält. Wird $\hat{S}'(k)$ anschließend bei Schritt **637** in die Zeitdomäne transformiert, erhält man ein nachgefiltertes Sprachsignal $\hat{S}'(n)$.

[0051] Unter Bezugnahme auf [Fig. 7](#) enthält ein beispielhaftes System zur Anwendung von Ausführungsformen der Erfindung eine Berechnungsvorrichtung, wie etwa eine Berechnungsvorrichtung **700**. In der einfachsten Ausführung enthält die Berechnungsvorrichtung **700** normalerweise wenigstens eine Verarbeitungseinheit **702** und einen Speicher **704**. In Abhängigkeit der exakten Konfiguration und des Typs der Berechnungsvorrichtung kann der Spei-

cher flüchtig (wie etwa ein RAM), nicht flüchtig (wie etwa ein ROM, ein Flash-Speicher, etc.) oder eine Kombination aus beiden sein. Diese einfachste Konfiguration ist in [Fig. 7](#) mit der Linie **706** dargestellt. Darüber hinaus kann die Vorrichtung **700** über zusätzliche Merkmale und Funktionalitäten verfügen. Beispielsweise kann die Vorrichtung **700** einen zusätzlichen Speicher (entnehmbar und/oder nicht entnehmbar) enthalten, der, ohne darauf beschränkt zu sein, über magnetische oder optische Platten oder ein Band verfügt. Ein derartiger zusätzlicher Speicher ist in [Fig. 7](#) durch einen entnehmbaren Speicher **708** und einen nicht entnehmbaren Speicher **710** dargestellt. Computerspeichermedien umfassen flüchtige und nicht flüchtige, entnehmbare und nicht entnehmbare Medien, die bei einem beliebigen Verfahren oder einer beliebigen Technologie zum Speichern von Informationen, wie etwa computerlesbaren Anweisungen, Datenstrukturen, Programmmodulen oder anderen Daten, eingesetzt werden. Der Speicher **704**, der entnehmbare Speicher **708** und der nicht entnehmbare Speicher **710** sind allesamt Beispiele von Computerspeichermedien. Computerspeichermedien umfassen, ohne darauf beschränkt zu sein, einen RAM, einen ROM, einen EEPROM, einen Flashspeicher oder eine andere Speichertechnologie, eine CD-ROM, eine DVD oder einen anderen optischen Speicher, Magnetkassetten, ein Magnetband, einen magnetischen Plattenspeicher oder andere Magnetspeichervorrichtungen oder ein beliebiges anderes Medium, das verwendet werden kann, um die gewünschten Informationen zu speichern, und auf das mit der Vorrichtung **700** zugegriffen werden kann. Ein beliebiges dieser Computerspeichermedien kann Teil der Vorrichtung **700** sein.

[0052] Die Vorrichtung **700** kann zudem eine oder mehrere Kommunikationsverbindungen **712** enthalten, die es der Vorrichtung gestatten, mit anderen Vorrichtungen zu kommunizieren. Die Kommunikationsverbindungen **712** sind ein Beispiel von Kommunikationsmedien. Kommunikationsmedien beinhalten normalerweise computerlesbare Anweisungen, Datenstrukturen, Programmmodule oder andere Daten in einem modulierten Datensignal, wie etwa einer Trägerwelle oder einem anderen Transportmechanismus, und beinhalten beliebige Informationszustellungsmedien. Der Begriff "moduliertes Datensignal" bezeichnet ein Signal, bei dem eines oder mehrere seiner Charakteristika derart eingestellt oder verändert werden, dass Informationen im Signal codiert werden. Beispielsweise, und ohne dabei einschränkend zu wirken, enthalten Kommunikationsmedien drahtgebundene Medien, wie etwa ein drahtgebundenes Netzwerk oder eine direkt verdrahtete Verbindung, und drahtlose Medien etwa Akustik-, HF- und Infrarotmedien sowie andere Medien. Wie es oben erläutert wurde, umfasst der Begriff computerlesbare Medien, wie er hier verwendet wird, sowohl Speichermedien als auch Kommunikationsmedien.

[0053] Die Vorrichtung **700** kann zudem über eine oder mehrere Eingabevorrichtungen **714**, wie etwa eine Tastatur, eine Maus, einen Stift, eine Spracheingabevorrichtung, eine Tasteingabevorrichtung, etc., verfügen. Eine oder mehrere Ausgabevorrichtungen **716**, wie etwa eine Anzeigevorrichtung, Lautsprecher, ein Drucker, etc., können ebenfalls enthalten sein. Alle diese Vorrichtungen sind nach dem Stand der Technik hinlänglich bekannt und müssen hier nicht ausführlicher erläutert werden.

[0054] Der Fachmann wird verstehen, dass hier ein neuartiges und nützliches Verfahren sowie System zum Ausführen einer Nachfilterung beschrieben wurden. Angesichts der zahlreichen möglichen Ausführungsformen, bei denen die Prinzipien dieser Erfindung angewendet werden können, sollte jedoch erkannt werden, dass die Ausführungsformen, die hier im Bezug auf die Zeichnungen beschrieben sind, lediglich der Veranschaulichung dienen und nicht als Einschränkung des Geltungsbereiches der Erfindung angesehen werden sollten. Die Erfindung ist beispielsweise derart beschrieben, dass sie eine Skalierfunktion mit einem Skalierfaktor zwischen 0 und 1 für die nicht lineare Transformation verwendet. Es können jedoch andere Transformationsfunktionen und Faktoren ebenfalls verwendet werden. Beispielsweise können auch Exponential- und Polynomfunktionen innerhalb der Erfindung zur Anwendung kommen. Obwohl zudem weiterhin der Hilbert-Phasenschieber zum Berechnen des Phasenansprechverhaltens des Gewinns festgelegt ist, können andere Techniken zum Berechnen des Phasenansprechverhaltens einer Funktion verwendet werden, wie etwa die Kotangenten-Transformationstechnik. Bei der Durchführung der Transformation von der Zeitdomäne in die Frequenzdomäne schreibt diese Beschreibung die DFT vor, wobei jedoch andere Transformationstechniken in äquivalenter Weise Anwendung finden können, wie etwa die schnelle Fouriertransformation (FFT) oder selbst eine herkömmliche Fouriertransformation. Wengleich die Erfindung im Zusammenhang mit Softwaremodulen oder -komponenten beschrieben wurde, wird der Fachmann verstehen, dass diese durch Hardwarekomponenten ersetzt werden können. Daher berücksichtigt die Erfindung, wie sie hier beschrieben ist, sämtliche derartige Ausführungsformen, die in den Geltungsbereich der folgenden Ansprüche und deren Äquivalente fallen.

Patentansprüche

1. Verfahren zum Nachfiltern eines Sprachsignals unter Verwendung linearer Prädiktivkoeffizienten des Sprachsignals zum Verbessern von Qualität menschlicher Wahrnehmung des Sprachsignals, wobei das Verfahren die folgenden Schritte umfasst: Erzeugen (**607-631**) eines Nachfilters durch Durchführen (**615**) einer nichtlinearen Transformation in der Frequenzdomäne, wobei beim Schritt des Erzeugens

des Nachfilters die nicht-lineare Transformation an dem kompensierten Spektrum linearer Prädiktivkoeffizienten durchgeführt wird, und Anwenden (**635**) des erzeugten Nachfilters auf das synthetisierte Sprachsignal in der Frequenzdomäne, **dadurch gekennzeichnet**, dass vor dem Schritt des Erzeugens des Nachfilters das Verfahren des Weiteren die folgenden Schritte umfasst:

Durchführen von Tilt-Berechnung, um den Tilt (μ) des Spektrums linearer Prädiktivkoeffizienten in der Zeitdomäne zu berechnen (**603**); und Kompensieren (**605**) des Spektrums linearer Prädiktivkoeffizienten unter Verwendung des berechneten Tilt in der Zeitdomäne.

2. Verfahren nach Anspruch 1, das des Weiteren Transformieren (**637**) des gefilterten, frequenzdomänen-synthetisierten Sprachsignals in ein Sprachsignal in der Zeitdomäne umfasst.

3. Verfahren nach Anspruch 2, wobei der Schritt des Kompensierens des Weiteren Anwenden einer Zero-Padding-Technik umfasst.

4. Verfahren nach einem der Ansprüche 1 bis 3, wobei der Schritt des Erzeugens eines Nachfilters des Weiteren die folgenden Schritte umfasst: Darstellen (**607**) des Spektrums linearer Prädiktivkoeffizienten durch einen Zeitdomänen-Vektor; Transformieren (**609**) des Zeitdomänen-Vektors in einen Frequenzdomänen-Vektor durch eine Fourier-Transformation; Invertieren (**613**) des Frequenzdomänen-Vektors; und Berechnen (**615-623**) von Gewinnen entsprechend dem Betrag des Allpol-Modell-Vektors, wobei die Gewinne einen Betrag und einen Phasengang enthalten.

5. Verfahren nach Anspruch 4, wobei der Schritt des Berechnens der Gewinne des Weiteren die folgenden Schritte umfasst: Normalisieren (**615**) des Betrages des Allpol-Modell-Vektors; Durchführen (**615**) einer nicht-linearen Transformation für den normalisierten Betrag des Allpol-Modell-Vektors, um den Betrag der Gewinne zu ermitteln; Schätzen (**617-623**) des Phasengangs der Gewinne; und Ausbilden der Gewinne durch Kombinieren (**623**) des Betrages und des geschätzten Phasengangs der Gewinne.

6. Verfahren nach Anspruch 5, wobei der Schritt des Schätzens des Phasengangs des Weiteren Ausführen eines Phasenschiebers auf Basis einer schnellen Fourier-Transformation an den Gewinnen umfasst.

7. Verfahren nach einem der Ansprüche 1 bis 6, wobei der Schritt des Erzeugens eines Nachfilters des Weiteren Ausführens (**625-631**) einer Anti-Aliasing-Prozedur in der Zeitdomäne nach dem Schritt des Berechnens der Gewinne umfasst.

8. Verfahren nach einem der Ansprüche 4 bis 6, wobei das Allpol-Modell durch einen Logarithmus mit dem inversen Betrag des Frequenzdomänen-Vektors der linearen Prädiktivkoeffizienten dargestellt wird.

9. Verfahren nach Anspruch 5 oder 6, wobei die nicht-lineare Transformationsfunktion eine Skalierfunktion mit einem Skalierfaktor zwischen 0 und 1 umfasst.

10. Computerlesbares Medium (**704, 708, 710**), das computerlesbare Befehle zum Durchführen von Schritten zum Nachfiltern eines synthetisierten Sprachsignals unter Verwendung des Spektrums linearer Prädiktivkoeffizienten des Sprachsignals aufweist, die die folgenden Schritte umfassen:

Durchführen von Tilt-Berechnung, um den Tilt (μ) des Spektrums linearer Prädiktivkoeffizienten zu berechnen (**603**);

Kompensieren (**605**) des Spektrums linearer Prädiktivkoeffizienten unter Verwendung des berechneten Tilt;

Erzeugen (**607-631**) eines Nachfilters durch Ausführen (**615**) einer nicht-linearen Transformation des kompensierten Spektrums linearer Prädiktivkoeffizienten in der Frequenzdomäne; und

Anwenden (**635**) des erzeugten Nachfilters auf das synthetisierte Sprachsignal in der Frequenzdomäne.

11. Computerlesbares Medium nach Anspruch 10, wobei der Schritt des Erzeugens eines Nachfilters des Weiteren die folgenden Schritte umfasst:

Darstellen (**607**) der linearen Prädiktivkoeffizienten durch einen Zeitdomänen-Vektor;

Transformieren (**609**) des Zeitdomänen-Vektors in einen Frequenzdomänen-Vektor durch eine Fourier-Transformation;

Übertragen (**613**) des Frequenzdomänen-Vektors in einen Allpol-Modell-Vektors; und

Berechnen (**615-623**) von Gewinnen entsprechend dem Betrag des Allpol-Modell-Vektors, wobei die Gewinne einen Betrag und einen Phasengang enthalten.

12. Computerlesbares Medium nach Anspruch 11, wobei der Schritt des Berechnens der Gewinne des Weiteren die folgenden Schritte umfasst:

Normalisieren (**615**) des Betrages des Allpol-Modell-Vektors;

Durchführen (**615**) einer nicht-linearen Transformation für den normalisierten Betrag des Allpol-Modell-Vektors, um den Betrag der Gewinne zu ermitteln;

Schätzen (**617-623**) des Phasengangs der Gewinne;

und

Ausbilden der Gewinne durch Kombinieren (**623**) des Betrages und des geschätzten Phasengangs der Gewinne.

13. Computerlesbares Medium nach Anspruch 12, wobei der Schritt des Schätzens des Phasengangs des Weiteren Ausführens eines Phasenschiebers auf Basis einer schnellen Fourier-Transformation umfasst.

14. Computerlesbares Medium nach einem der Ansprüche 10 bis 13, wobei der Schritt des Erzeugens eines Nachfilters des Weiteren Ausführens (**625-631**) einer Anti-Aliasing-Prozedur in der Zeitdomäne umfasst.

15. Computerlesbares Medium nach einem der Ansprüche 11 bis 13, wobei das Allpol-Modell durch einen Logarithmus mit dem inversen Betrag des Frequenzdomänen-Vektors dargestellt wird.

16. Computerlesbares Medium nach Anspruch 12 oder 13, wobei die nicht-lineare Transformationsfunktion eine Skalierfunktion mit einem Skalierfaktor zwischen 0 und 1 umfasst.

17. Vorrichtung (**310, 410, 412, 521**) zum Einsatz mit einem Nachfilter (**303**) zum Verarbeiten linearer Prädiktivkoeffizienten eines Signals und zum Bereitstellen von Gewinnen für ein Frequenzdomänen-Formant-Filter (**310, 410, 500**), wobei die Vorrichtung umfasst:

ein Modul (**415, 510**) für Berechnung eines Tilt linearer Prädiktivkoeffizienten, das Tilt-Berechnung durchführt, um den Tilt (μ) der linearen Prädiktivkoeffizienten zu berechnen (**603**);

ein Modul (**420, 520**) für Kompensation des Tilt linearer Prädiktivkoeffizienten, das das Spektrum linearer Prädiktivkoeffizienten entsprechend dem berechneten Tilt des Spektrums linearer Prädiktivkoeffizienten kompensiert (**605**); und

ein Modul (**430, 530**) für Berechnung des Gewinns eines Formant-Filters, das die Gewinne des Frequenzdomänen-Formant-Filters entsprechend den kompensierten linearen Prädiktivkoeffizienten berechnet (**607-631**), wobei die Gewinne einen Betrag und einen Phasengang enthalten.

18. Vorrichtung nach Anspruch 17, die des Weiteren zum Nachfiltern eines Sprachsignals unter Verwendung einer Vielzahl linearer Prädiktivkoeffizienten des Sprachsignals zum Verbessern von Qualität menschlicher Wahrnehmung des Sprachsignals dient, wobei die Vorrichtung des Weiteren umfasst:

ein Fourier-Transformations-Modul (**411, 511**), das zum Durchführen einer Fourier-Transformation betrieben werden kann;

ein Modul (**513, 513**) für inverse Fourier-Transformation, das zum Durchführen einer inversen Fou-

rier-Transformation betrieben werden kann; und ein Formant-Filter (**412**), das die Gewinne des Frequenzdomänen-Formant-Filters umfasst, wobei die Gewinne in der Frequenzdomäne berechnet werden, indem eine nicht-lineare Transformation der linearen Prädiktivkoeffizienten durchgeführt wird.

19. Vorrichtung nach Anspruch 18, wobei das Formant-Filter des Weiteren umfasst: das Modul für Berechnung eines Tilt der linearen Prädiktivkoeffizienten, das den Tilt des Spektrums linearer Prädiktivkoeffizienten berechnet; das Modul für Kompensation des Tilt linearer Prädiktivkoeffizienten, das die linearen Prädiktivkoeffizienten entsprechend dem berechneten Tilt des Spektrums linearer Prädiktivkoeffizienten kompensiert; das Modul für Berechnung des Formant-Gewinns, das die Gewinne des Formant-Filters in der Frequenzdomäne durch Durchführen einer nicht-linearen Transformation der linearen Prädiktivkoeffizienten nach Tilt-Kompensation durchführt, wobei die Gewinne einen Betrag und einen Phasengang enthalten; und ein Modul (**440**) zur Anwendung von Gewinnen des Formant-Filters auf ein Sprachsignal anwendet (**635**), indem es die Gewinne und das Sprachsignal in der Frequenzdomäne multipliziert.

20. Vorrichtung nach Anspruch 19, wobei das Modul für Berechnung des Formant-Gewinns des Weiteren umfasst: ein Modul (**431**) zur Darstellung linearer Prädiktivkoeffizienten, das die linearen Prädiktivkoeffizienten durch einen Zeitdomänen-Vektor darstellt (**607**); ein Modelliermodul (**432**), das einen Frequenzdomänenvektor entsprechend einem vordefinierten Modell zum Erzeugen eines Betrags modelliert (**609**), wobei der Frequenzdomänen-Vektor aus dem Zeitdomänen-Vektor transformiert wird, der die LPC-Koeffizienten darstellt; ein Modul (**433**) für nicht-lineare Transformation der linearen Prädiktivkoeffizienten, das eine nicht-lineare Transformation an dem Betrag durchführt (**615**) und dem Betrag der Gewinne des Formant-Filters erzeugt; ein Phasenberechnungsmodul (**434**), das einen Phasengangs der Formant-Filter-Gewinne entsprechend dem Betrag des Modells nach nicht-linearer Transformation berechnet (**617-623**); ein Modul zum Kombinieren des Gewinns des Formant-Filters (**435**), das den Betrag und den Phasengang des Formant-Filter-Gewinns kombiniert (**635-631**); und ein Anti-Aliasing-Modul (**436**), das Aliasing verhindert (**635-631**), das durch Anwendung des Formant-Filters verursacht wird; für

21. Verfahren nach Anspruch 20, wobei das Modul zur Darstellung der linearen Prädiktivkoeffizienten so eingerichtet ist, dass es die linearen Prädiktivkoeff-

fizienten mit einer Zero-Padding-Technik darstellt.

22. Vorrichtung nach Anspruch 20 oder 21, wobei das Modul für nicht-lineare Transformation linearer Prädiktivkoeffizienten des Weiteren eine Skalierfunktion mit einem Skalierfaktor zwischen 0 und 1 umfasst.

23. Vorrichtung nach einem der Ansprüche 20 bis 22, wobei das Phasenberechnungsmodul des Weiteren einen Hilbert-Phasenschieber in der Zeitdomäne umfasst.

Es folgen 8 Blatt Zeichnungen

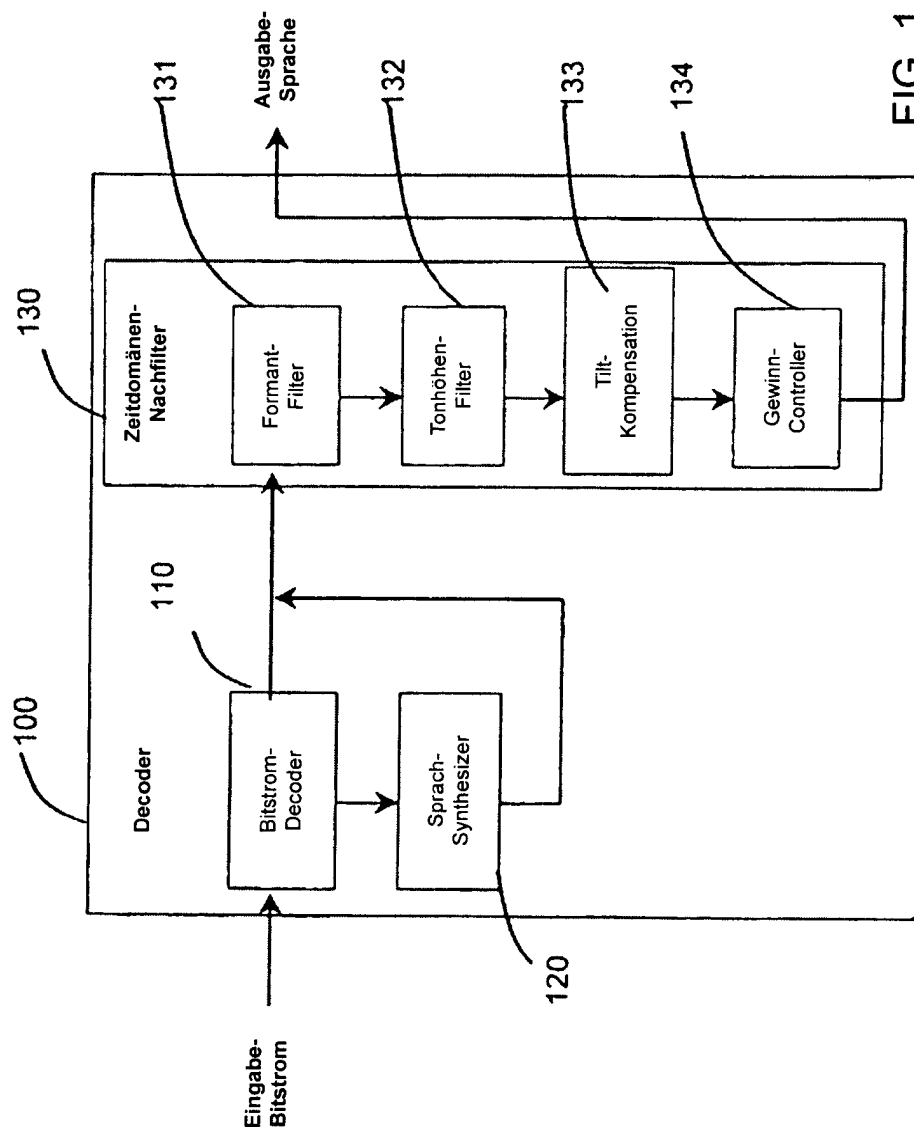


FIG. 1 Stand der Technik

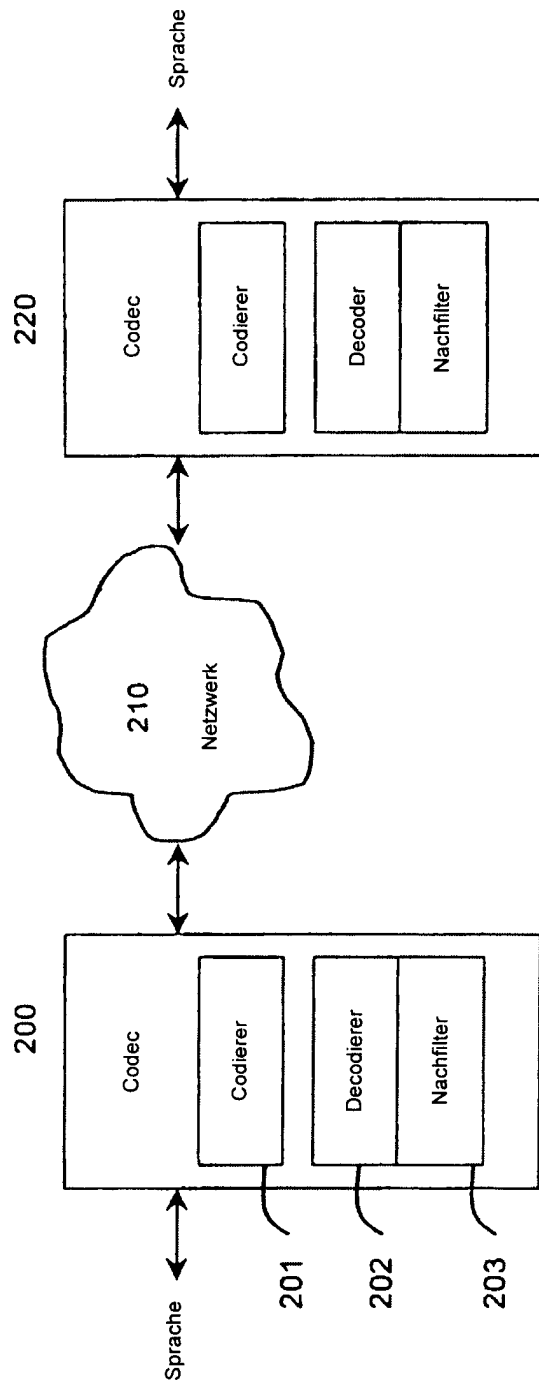


FIG. 2

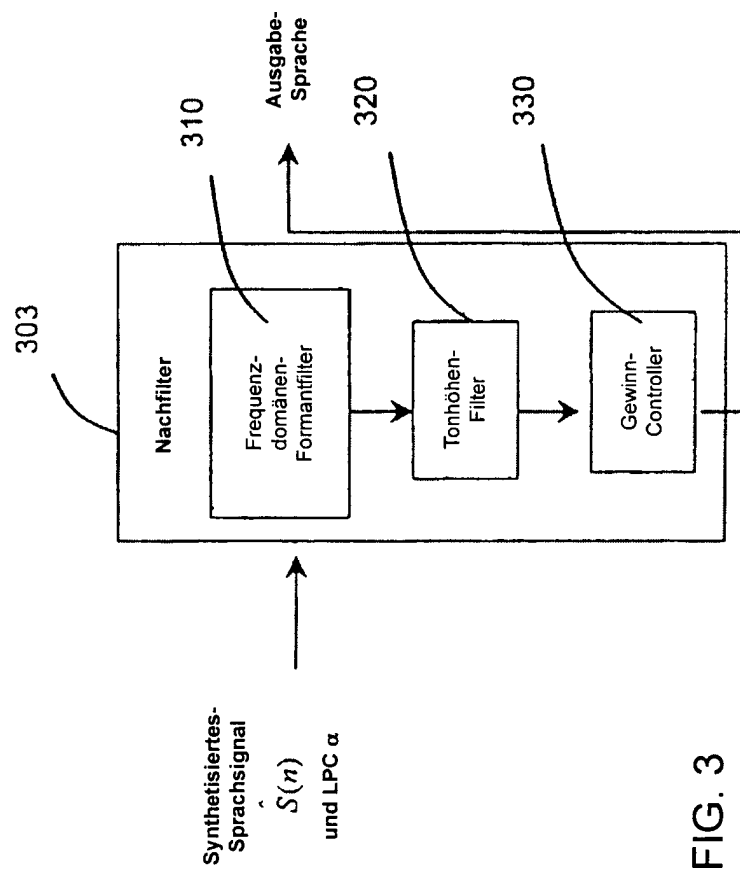


FIG. 3

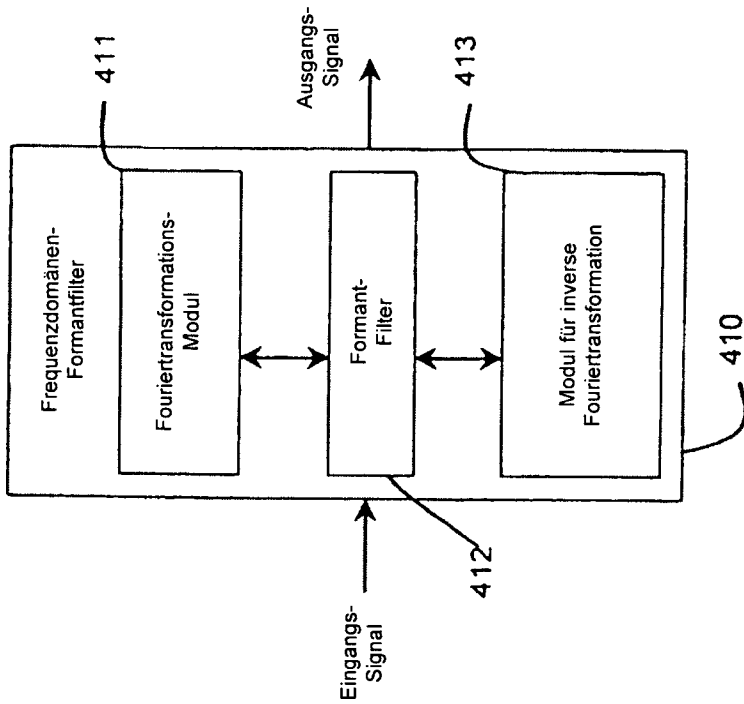


FIG. 4a

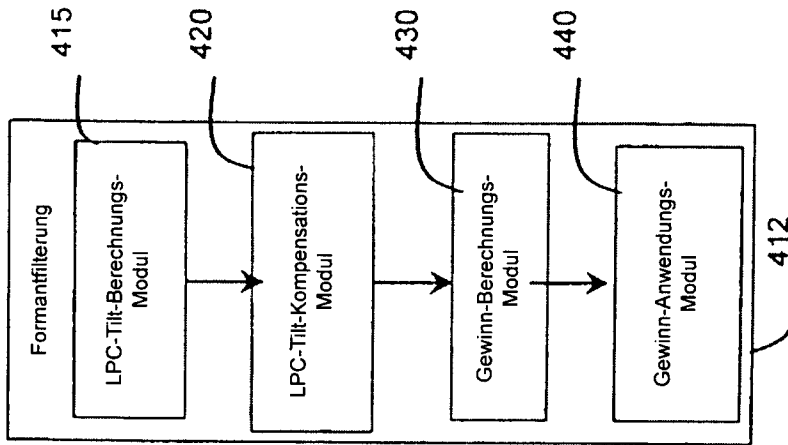


FIG. 4b

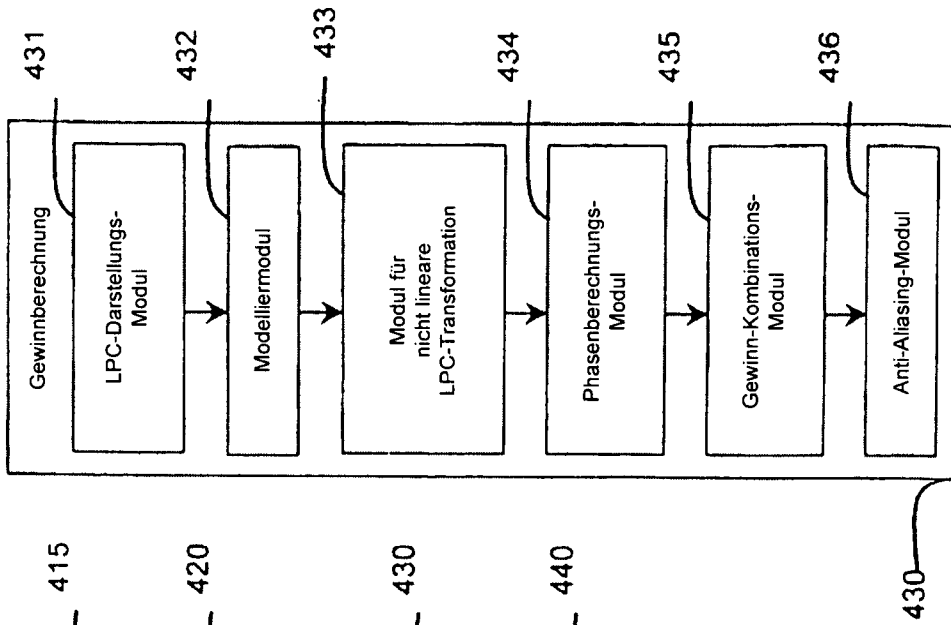


FIG. 4c

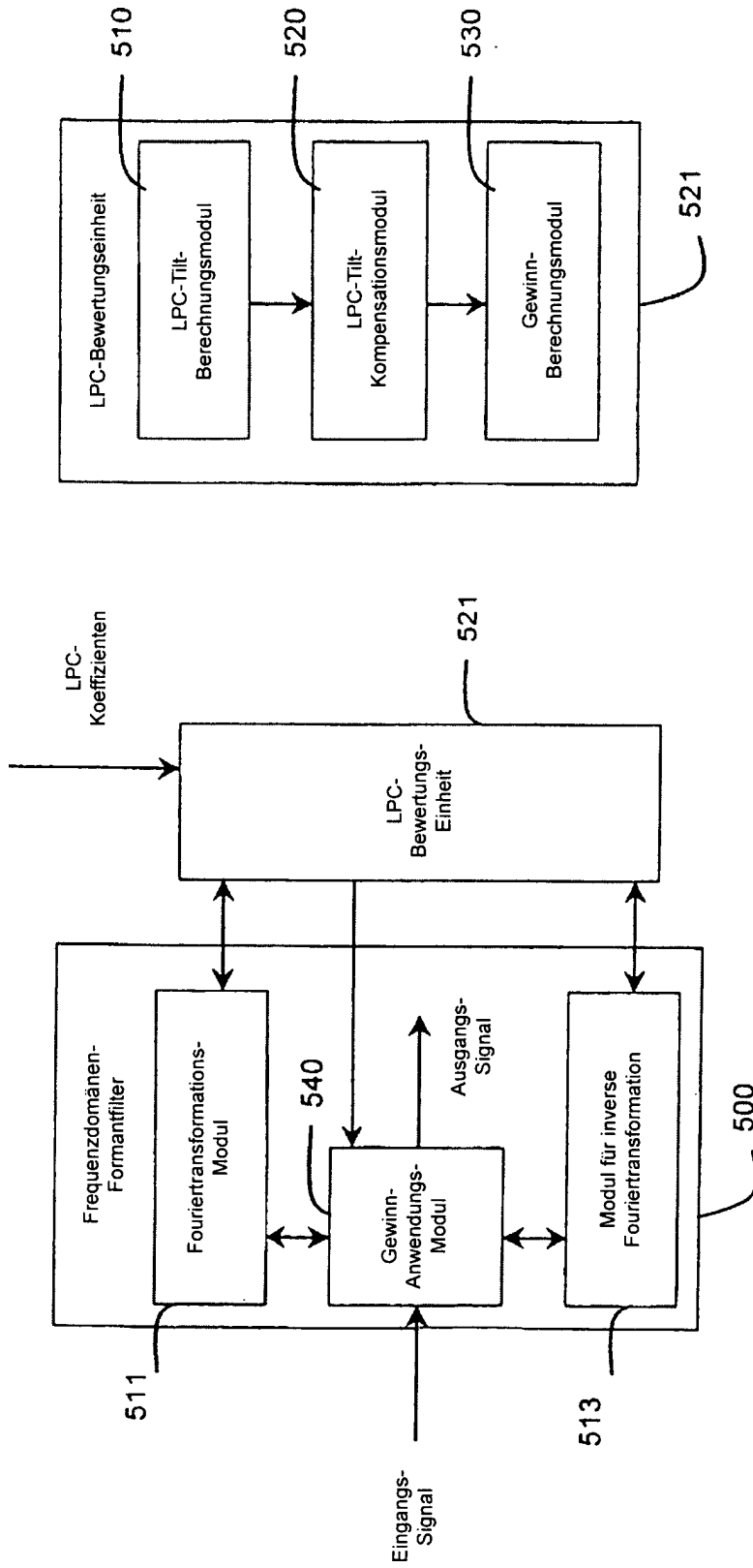


FIG. 5b

FIG. 5a

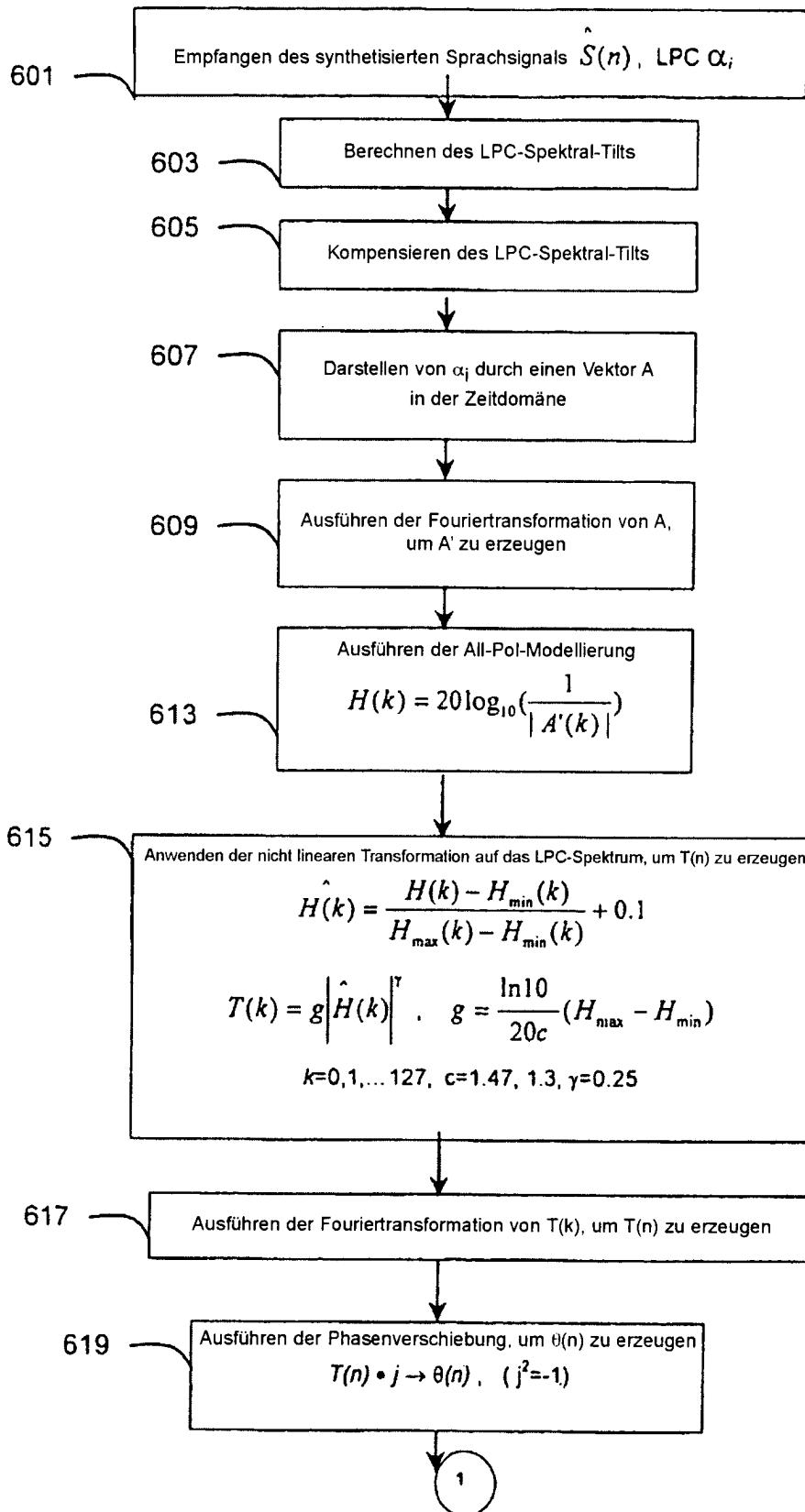


FIG. 6a

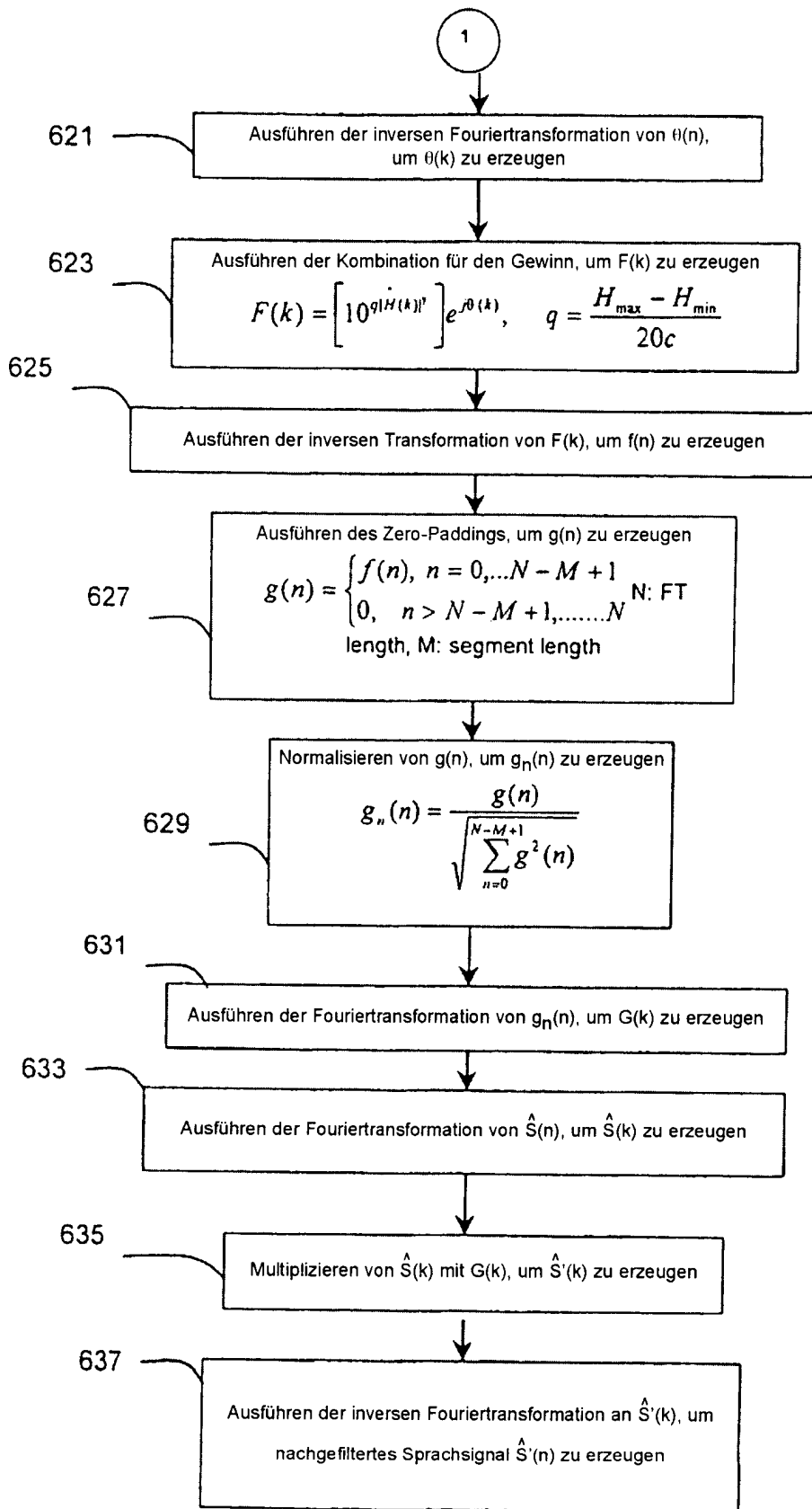


FIG. 6B

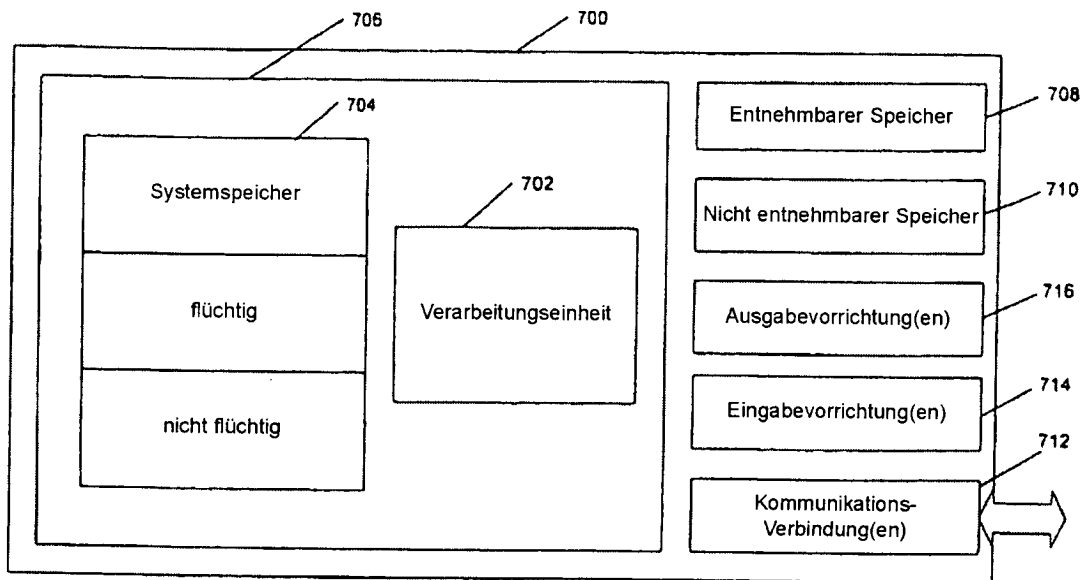


FIG. 7