(19) **Bundesrepublik Deutschland**
Deutsches Patent- und Markenamt

(10) **DE 602 26 200 T2** 2009.05.14

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 1 271 471 B1**
(21) Deutsches Aktenzeichen: **602 26 200.3**
(96) Europäisches Aktenzeichen: **02 014 365.7**
(96) Europäischer Anmeldetag: **27.06.2002**
(97) Erstveröffentlichung durch das EPA: **02.01.2003**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **23.04.2008**
(47) Veröffentlichungstag im Patentblatt: **14.05.2009**

(51) Int Cl.[8]: ***G10L 19/08*** (2006.01)

(30) Unionspriorität:
896272 **29.06.2001** **US**

(84) Benannte Vertragsstaaten:
AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE, TR

(73) Patentinhaber:
Microsoft Corp., Redmond, Wash., US

(72) Erfinder:
Rao, Ajit V., Redmond, Washington 98052, US

(74) Vertreter:
Grünecker, Kinkeldey, Stockmair & Schwanhäusser, 80802 München

(54) Bezeichnung: **Signaländerung mit Hilfe von kontinuierlicher Zeitverschiebung für CELP Kodierung mit niedriger Bitrate**

Beschreibung

TECHNISCHES FELD

[0001] Diese Erfindung bezieht sich generell auf Sprachkodierungstechniken und im Speziellen bezieht sie sich auf Techniken zum Modifizieren eines Signals, um beim Kodieren des Signals über eine Kodierungstechnik bei niedriger Bit-Rate, so wie einer durch ein Code-Buch hervorgerufenen linearen Vorhersage (codebook excited linear prediction – CELP) Kodierung, behilflich zu sein.

HINTERGRUND DER ERFINDUNG

[0002] In dem heutigen hochverbalen und hochinteraktiven technischen Klima ist es oft notwendig oder wünschenswert, menschliche Stimmen elektronisch von einem Punkt zu einem anderen zu übertragen, manchmal über große Entfernungen und oft über Kanäle mit begrenzter Bandbreite. Zum Beispiel sind Konversationen über Mobiltelefonverbindungen oder über das Internet oder andere digitale elektronische Netzwerke nun allgemein üblich. Gleichwohl ist es oft nützlich, menschliche Stimmen digital zu speichern, so wie auf der Festplatte eines Computers oder in dem flüchtigen oder nichtflüchtigen Speicher eines digitalen Aufzeichnungsgerätes. Zum Beispiel kann eine digital gespeicherte menschliche Stimme als ein Teil eines Telefonantwortprotokolls oder einer Audio-Präsentation wiedergegeben werden.

[0003] Kanäle und Medien, die für eine Übertragung und/oder Speicherung von digitalen Stimmen verwendbar sind, haben oft eine begrenzte Kapazität und wachsen dennoch jeden Tag weiter an. Zum Beispiel hat das Aufkommen von qualitativem Video zur Verwendung in Verbindung mit aufgezeichneten oder in Echtzeit verwendeten Stimmen ein Bedürfnis für Audio-/Video-Konferenzen über digitale Netzwerke sowie in Echtzeit als auch für nicht Echtzeit hochqualitative Audio-Videopräsentationen geschaffen, so wie solche, die in einem Streaming-Format empfangen werden können, und solche, die in ihrer Gesamtheit zum Speichern herunterladbar sind. Da Video-Inhalte Bandweiten- und Speicherkapazitäten in verschiedenen Übertragungskanälen und Speichermedien verdrängen, wird ein Bedarf, sowohl Stimmen als auch Video effizient und richtig zu komprimieren, unumgänglich. Andere Szenarien erschaffen auch einen Bedarf für extreme und effektive Kompression von Stimmen. Zum Beispiel müssen zunehmend überladene oder verstopfte Mobiltelefonverbindungen dazu fähig sein, eine größere Anzahl von Benutzern aufzunehmen, was oft über Kanäle geschieht, deren Kapazität sich nicht entsprechend der Anzahl der Nutzer verändert hat.

[0004] Was auch immer die Motivation ist, war und bleibt das Komprimieren von Stimmen ein wichtiger Bereich in der Kommunikationstechnologie. Zur Verfügung stehende digitale Stimmen-Kodierungstechniken überspannen ein Spektrum von ineffizienten Techniken, die keine Kompression anwenden, bis zu effizienten Techniken, die Kompressionsraten von vier oder mehr erreichen.

[0005] Generell können bestehende Kodierer als entweder Wellenformkodierer oder Stimmenkodierer klassifiziert werden. Wellenformkodierer versuchen eigentlich, die Schallwelle selber zu beschreiben, und erreichen üblicherweise keine hohen Kompressionsraten. Stimmenkodierer oder Vocoder ziehen die Quelle und die Eigenheiten der menschlichen Sprache eher in Betracht, als einfach zu versuchen, die sich ergebende Schall- bzw. Klangwelle abzubilden, und können dementsprechend viel höhere Kompressionsraten erreichen, obgleich dies auf Kosten einer erhöhten Rechenkomplexität erfolgt. Wellenformkodierer sind generell robuster bei eigenartigen menschlichen Stimmen, nicht der Sprache zuzuordnenden Klängen und hochgradigem Hintergrundrauschen.

[0006] Die meisten vorherrschenden Stimmenkodierer setzen Techniken ein, die auf einem linear vorhersagenden Kodieren (linear predictive coding) basieren. Die lineare vorhersagende bzw. prädiktive Kodierungstechnik nimmt an, dass für jeden Teil des Sprachsignals ein digitaler Filter existiert, der, wenn er durch ein bestimmtes Signal angeregt wird, ein Signal produziert, das dem Teil des originalen Sprachsignals sehr ähnlich ist. Insbesondere wird ein Kodierer, der eine lineare prädiktive Technik implementiert typischerweise zuerst eine Reihe von Koeffizienten ableiten, die eine spektrale Hüllkurve beschreiben oder Formanten bzw. Charakteristiken des Sprachsignals. Ein Filter, der diesen Koeffizienten entspricht, wird eingerichtet und dann dazu verwendet, das eingegebenen Sprachsignal auf ein vorhersagendes bzw. prädiktives Residuum bzw. Restsignal zu reduzieren. In allgemein üblichen Worten ist der oben beschriebene Filter ein inversiver bzw. umgekehrter Synthesefilter, so dass die Eingabe des Residuum-Signals in einen entsprechenden Synthesefilter ein Signal hervorrufen wird, das das originale Sprachsignal genau approximiert bzw. sich diesem annähert.

[0007] Typischerweise werden die Filterkoeffizienten und das Residuum für eine spätere und/oder entfernte Re-Synthese des Sprachsignals übertragen oder gespeichert. Während die Filterkoeffizienten wenig Platz zum Speichern oder eine geringe Bandbreite erfordern, z. B. 1,5 kbps zum Übertragen, ist das prädiktive Residuum ein Signal hoher Bandbreite und dem originalen Sprachsignal in seiner Komplexität ähnlich. Somit muss das prädiktive Residuum komprimiert werden, um das Sprachsignal effektiv zu komprimieren. Die Technik der Codebuch hervorge-

rufenen Linearprädiktionen (Codebook Excited Linear Prediction – CELP) wird dazu verwendet, diese Kompression zu erreichen. CELP verwendet einen oder mehrere Codebuchindizes, die zur Auswahl bestimmter Vektoren anwendbar sind, ein jeder aus einer Reihe von „Codebüchern" („codebooks"). Jedes Codebuch ist eine Sammlung von Vektoren. Die ausgewählten Vektoren werden so gewählt, dass sie, wenn sie skaliert und summiert werden, eine Rückmeldung bzw. Resonanz von dem Synthesefilter erzeugen, die am besten die Rückmeldung des Filters auf das Residuum selber approximiert. Der CELP-Dekodierer hat einen Zugriff auf die gleichen Codebücher, wie ihn der CELP-Kodierer hatte, und somit sind die einfachen Indizes dazu verwendbar, die gleichen Vektoren aus dem Kodier- und Dekodier-Codebuch zu identifizieren.

[0008] Wenn die verfügbare Kapazität oder Bandbreite hinreichend ist, ist es nicht schwierig ein Codebuch zu haben, das reichhaltig genug ist, um eine genaue Approximation bzw. Abschätzung des originalen Residuums zuzulassen, so komplex es auch sein mag. Allerdings nimmt die Reichhaltigkeit des CELP-Codebuches notwendigerweise ab, da die zur Verfügung stehende Kapazität oder Bandbreite abnimmt.

[0009] Ein Weg, um die Anzahl von Bits zu reduzieren, die dazu notwendig ist, um das Residuum-Signal bzw. Restsignal nachzuahmen, ist, die Periodizität zu erhöhen. Das heißt, dass die Redundanzen mit dem originalen Signal kompakter darstellbar sind als die nicht redundanten Merkmale. Eine Technik, welche dieses Prinzip zu ihrem Vorteil nutzt, ist ein Lockerungs-Codebuch-hervorgerufenes-linear-prädiktives-Kodieren (Relaxation Codebook Excited Linear Predicitve coding – RCELP). Ein Beispiel von dieser Technik wird dem Artikel „The RCELP Speech coding Algorithm", Eur. Trans. On Communications, Ausgabe 4, Nr. 5, Seiten 573–82 (1994), verfasst durch W. B. Kleijn et al., diskutiert. Insbesondere beschreibt dieser Artikel ein Verfahren des gleichförmigen Voranschreitens oder Verzögerns ganzer Segmente eines Residuum-Signals, so dass seine modifizierte Pitch-Periodenkontur bzw. Tonhöhen-Periodenkontur einer synthetischen Pitch-Periodenkontur gleicht. Probleme mit dieser Herangehensweise beinhalten den Fakt, dass als ein Artefakt der besonderen Warp-Verfahrensweise bestimmte Teile des originalen Signals ausgelassen oder wiederholt werden können. Insbesondere wenn zwei angrenzende Segmente des Signals eine kumulative kompressive Verschiebung erfahren, können Teile des originalen Signals nahe der Überlappung in dem modifizierten Signal ausgelassen werden. Ebenso wenn zwei angrenzende Segmente eine kumulative erweiternde Verschiebung erfahren, können Teile der Abschnitte des originalen Signals nahe der Überlappung in dem modifizierten Signal wiederholt werden. Diese Artefakte können hörbare Verzerrungen in der abschließend reproduzierten Sprache erzeugen. Anderer Stand der Technik hat eine ähnliche Herangehensweise vorgeschlagen. Siehe zum Beispiel der Artikel „Interpolation of the Pitch-Predictor parameters in Analysis-by-Synthesis Speech Coders", IEEE Transactions of Speech and Audio Processing, Ausgabe 2, Nr. 1, Teil I (Januar 1994), verfasst von W. B. Kleijn et al.

[0010] Alle Tonhöhen verzerrenden (pitch warping) Herangehensweisen, die in der Vergangenheit vorgeschlagen wurden, haben an ähnlichen Unzulänglichkeiten gelitten, inklusive einer Reduktion in der Qualität wegen dem Verschieben von Segmentkanten, was Auslassungen und Wiederholungen des originalen Signals erzeugt. Es ist wünschenswert, ein rahmenverzerrendes Verfahren (frame warping method) zur Verfügung zu stellen, um die Übertragungs-Bit-Rate für Sprachsignale zu reduzieren, während keine Signalwiederholungen oder -auslassungen eingebracht werden, und ohne die Komplexität oder Verzögerung der Codierungsberechnungen bis zu einem Punkt zu erhöhen, bei dem Echtzeit-Kommunikationen nicht möglich sind.

ZUSAMMENFASSUNG DER ERFINDUNG

[0011] Die Erfindung setzt eher eine kontinuierliche als einfach eine stückweise kontinuierliche Zeit-Verzerrungs-Kontur (time warp contour) ein, um das originale Residuum-Signal bzw. Restsignal so zu modifizieren, dass es an eine synthetische Kontur angepasst ist, wodurch Kantenverschiebungseffekte, die im Stand der Technik vorherrschen, vermieden werden. Insbesondere ist die Warp-Kontur bzw. Verzerrungskontur, die innerhalb der Erfindung verwendet wird, kontinuierlich, d. h., es fehlen ihr räumliche Sprünge oder Diskontinuitäten, und sie invertiert nicht die Positionen angrenzender Endpunkte in angrenzenden Rahmen oder weitet diese übermäßig aus.

[0012] Um die Komplexität des Kodierungsprozesses zu reduzieren, damit praktische und ökonomische Implementationen ermöglicht werden, wird die optimale lineare Verschiebung über eine quadratische oder eine andere Abschätzung abgeleitet. Insbesondere erfordert der Algorithmus, der innerhalb der Erfindung verwendet wird, um die ideale Warp-Kontur zu bestimmen, nicht, dass jede mögliche Warp-Kontur berechnet und verwendet wird, um das modifizierte Signal mit dem synthetischen Signal zu korrelieren. In einer Ausführungsform wird eine Untergruppe von möglichen Konturen quer aus einem Unterbereich von möglichen Konturen berechnet. Die relativen Korrelationsstärken von diesen Konturen werden dann als Punkte auf einer quadratischen Kurve oder einer anderen metrischen Funktionskurve modelliert. Die optimale Warp-Kontur, mög-

licherweise durch einen Punkt repräsentiert, der auf einem Ort zwischen den berechneten Abtastungs- bzw. Samplepunkten liegt, wird dann durch ein Maximieren der angemessenen bzw. dazugehörigen parametrischen Funktion berechnet. Andere Vereinfachungstechniken, so wie eine Zweiteilung (bisection) oder eine stückweise polynomische Modellierung können auch innerhalb der Erfindung verwendet werden.

[0013] Andere Merkmale und Vorteile der Erfindung werden aus der folgenden detaillierten Beschreibung der veranschaulichenden Ausführungsformen offensichtlich gemacht, welche unter Bezug auf die begleitenden Zeichnungen voranschreitet.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0014] Während die angehängten Ansprüche die Merkmale der vorliegenden Erfindung mit Sorgfalt darlegen, kann die Erfindung zusammen mit ihren Zielen und Vorteilen am besten aus der nachfolgenden detaillierten Beschreibung verstanden werden, die in Verbindung mit den begleitenden Zeichnungen gebracht wird, von denen:

[0015] Fig. 1 ein architektonisches Diagramm eines beispielhaften Kodierers ist, in dem eine Ausführungsform der vorliegenden Erfindung implementiert werden kann;

[0016] Fig. 2 ein vereinfachtes Wellenformdiagramm ist, das eine Signal-Segmentation, Zeit-Verzerrung und Wiederherstellung in einer Ausführungsform der Erfindung darstellt;

[0017] Fig. 3a und Fig. 3b Flussdiagramme sind, die Schritte darstellen, welche vorgenommen werden, um die Signalmodifikation innerhalb einer Ausführungsform der vorliegenden Erfindung zu bewirken;

[0018] Fig. 4 ein Flussdiagramm ist, das die Schritte zum Berechnen einer optimalen Lag-Kontur innerhalb einer Ausführungsform der vorliegenden Erfindung darstellt;

[0019] Fig. 5 ein vereinfachter Graph ist, der das Nachzeichnen der Korrelationsstärke als eine Funktion der ersten Abtast-Lag-Werte zu berechnen, die in einer Ausführungsform der Erfindung verwendet werden, um einen optimalen letzten Abtast-Lag zu bestimmen;

[0020] Fig. 6 eine graphische Darstellung von Warp-Konturen gemäß dem Stand der Technik und gemäß einer Ausführungsform der vorliegenden Erfindung ist; und

[0021] Fig. 7 ein vereinfachtes schematisches Diagramm einer Computer-Vorrichtung ist, auf der eine Ausführungsform der vorliegenden Erfindung implementiert werden kann.

DETAILLIERTE BESCHREIBUNG DER ERFINDUNG

[0022] In der folgenden Beschreibung wird die Erfindung mit Bezug auf Handlungen und symbolische Darstellungen von Operationen beschrieben, die durch einen oder mehrere Computer durchgeführt werden, sofern dies nicht anders angezeigt wird. Als solches wird es verstanden werden, dass solche Handlungen und Operationen, auf welche bei Zeiten als durch einen Computer ausgeführt Bezug genommen wird, die Manipulation von elektrischen Signalen durch eine Prozessoreinheit des Computers beinhalten, die Daten in einer strukturierten Form repräsentieren. Die Manipulation formt die Daten um oder halt sie an Orten in dem Speichersystem des Computers, was den Betrieb des Computers rekonfiguriert oder anderweitig verändert, in einer Art und Weise, die von Fachmännern gut verstanden wird. Die Datenstrukturen, wo Daten gehalten werden, sind physikalische Orte des Speichers, die bestimmte Eigenschaften oder Werte haben, die durch das Format der Daten definiert werden. Während die Erfindung im vorangegangenen Kontext beschrieben wird, ist dieser allerdings nicht dazu gedacht, einschränkend zu sein, da Fachmänner anerkennen werden, dass verschiedene der Handlungen und Operationen, die hier im Folgenden beschrieben werden, auch als Hardware implementiert werden können.

[0023] Ein Sprachkodierer ist ein Software-Modul, das betriebsfähig ist, ein digitales Eingangsaudiosignal auf hoher Bit-Rate in ein Signal auf einer niederen Bit-Rate zu komprimieren, welches dann über einen digitalen Kanal, zum Beispiel das Internet, übertragen oder in einem digitalen Speichermodul, zum Beispiel einer Festplatte oder einer CD-R, gespeichert wird. Die übertragenen oder gespeicherten Bits werden durch einen Sprachdekodierer in dekodierte digitale Audiosignale umgewandelt bzw. konvertiert. Sprachkodierer und -dekodierer werden oft gemeinsam als ein Sprach-Codec bezeichnet. Sprach-Codecs sind dazu entworfen, bei dem Dekodierer die genauest mögliche Rekonstruktion eines eingegebenen Audiosignals zu erzeugen, insbesondere wenn das Eingangssignal menschliche Sprache ist. Das allgemein üblichste Paradigma, das beim Kodieren von Sprache verwendet wird, ist eine Code-Buch-hervorgerufene-Linearprädiktion (codebook excited linear prediction – CELP). CELP-Sprachkodierer basieren auf dem Prinzip einer kurzzeitigen Prädiktion bzw. Vorhersage und Code-Buch-Suche. Die Konzepte und Funktionen des CELP-Kodierens werden hierin diskutiert, um den Leser zu unterstützen. Diese Diskussion ist nicht dazu gedacht, das CELP-Kodieren auf einer anderen Art und Weise zu

definieren, als die im Stand der Technik bekannte.

[0024] Die Aufgabe eines jeglichen Sprach-Kodierers wird schwieriger und komplexer bei niedrigen Bit-Raten wegen den wenigen Bits, die zur Verfügung stehen, die komplexe und zeitvariable Natur der menschlichen Sprache einzufangen. Diese Erfindung stellt eine neue Methodik zum Modifizieren der eingegebenen digitalen Sprachdaten zur Verfügung, bevor diese durch einen Sprach-Kodierer kodiert werden, so dass weniger Bits für das Speichern oder Übertragen benötigt werden. Das Ziel der Signalmodifikation ist es, die Struktur der Wellenform des eingegebenen Sprachsignals zu vereinfachen, ohne die Wahrnehmungsqualität des rekonstruierten Signals ungünstig zu beeinflussen. Der Signalmodifikation folgend wird das modifizierte eingegebenen Sprachsignal zum Kodieren in den Sprach-Kodierer eingegeben. Wegen dieser vereinfachten Struktur der modifizierten Wellenform kann der Sprachkodierer richtiger und effizienter die Aufgabe des Kodierens des Signals durchführen. Wie zuvor erwähnt, ist die Signalmodifikation insbesondere bei niedrigen Bit-Raten vorteilhaft.

[0025] Die Signalmodifikationstechnik, die hierin beschrieben wird, basiert auf einem Modell einer kontinuierlichen Zeitverzerrung (continuous time warping). Ungleich der Signalmodifikationstechnik von RCELP, auf die oben Bezug genommen wurde, modifiziert die kontinuierliche Zeitverzerrung das Eingangssignal eher unter Verwendung einer kontinuierlichen Verzerrungskontur (continuous warping contour) als einfach einer stückweise kontinuierlichen Kontur. Das Ergebnis ist ein modifiziertes Sprachsignal, dessen Wellenform eine einfache Struktur hat und dessen Qualität der des originalen eingegebenen Signals geradezu gleich ist.

[0026] Um die Erfindung vollständig zu verstehen, ist es wichtig, die zugrundeliegenden Facetten der CELP-Familie von Codec-Techniken zu verstehen. Obwohl die verschiedenen CELP-Techniken dem Fachmann wohl bekannt sein werden, werden sie dennoch hierin für die Annehmlichkeit des Lesers beschrieben. Beim CELP-Kodieren wird das dekodierte Sprachsignal durch das Filtern eines Anregungssignals durch einen zeitvarianten Synthesefilter erzeugt. Der Kodierer sendet Informationen über das Anregungssignal und den Synthesefilter an den Dekodierer.

[0027] CELP ist ein Wellenform-Angleichungsverfahren; d. h., die Wahl des Anregungssignals wird über eine Korrelation eines vorgeschlagenen synthetischen Signals mit dem zu modellierenden Signal optimiert, z. B. das Residuum bzw. Restsignal. Folglich bewertet der Kodierer kurze Segmente des eingegebenen Sprachsignals und versucht die genaueste Replik für jedes Segment zu erzeugen. Insbesondere erzeugt der Kodierer zuerst einen Satz von Anregungssignalen durch ein Kombinieren bestimmter erlaubter Signale, die „Code-Vektoren" („code-vectors") genannt werden. Jedes Anregungssignal in dem Satz bzw. der Gruppe, die somit erzeugt werden, wird durch den Synthesefilter geschickt und das gefilterte Anregungssignal, das die genaueste Ähnlichkeit zu dem originalen Sprachsignal erzeugt oder zu anderen Signalen, die zu replizieren bzw. kopieren sind, wird ausgewählt. Dieser Prozedur folgend, überträgt der Kodierer Informationen über die Code-Vektoren, die zum Erzeugen des ausgewählten Anregungssignals ausgewählt wurden, und Informationen über den Synthesefilter an den Dekodierer. Üblicherweise werden die meisten der Bits dazu benötigt, Informationen über die Code-Vektoren zum Bilden des Synthesefilter-Anregungssignals zu übertragen, während die Synthesefilter-Parameter selber üblicherweise weniger als 1,5 kb/s benötigen. Folglich arbeitet CELP bei relativ hohen Bit-Raten gut, z. B. größer als 4 kbps, wobei es ausreichend Code-Vektoren gibt, um die komplexe Natur des eingegebenen Sprachsignals darzustellen. Bei niedrigen Bit-Raten sinkt die Qualität des reproduzierten Signals wegen der geringen Anzahl von Code-Vektoren, die zulässig sind, erheblich.

[0028] Die dominanten Charakteristiken des Restsignals für die zur Wahrnehmung wichtigen Stimmensegmente der Sprache sind eine Sequenz von kaum bzw. rau periodischen Spitzen (spikes). Obwohl diese Spitzen auf eine Art und Weise generell gleichförmig voneinander beabstandet sind, von einer Tonhöhenperiode bzw. Pitch-Periode getrennt, gibt es oft kleine Jitters in der Regelmäßigkeit der Orte dieser Spitzen. Diese Jitter nehmen, obwohl sie für die Wahrnehmung nicht wichtig sind, eine Mehrheit der zur Verfügung stehenden Bits in Wellenformkodierern niederer Bit-Rate ein.

[0029] Wie diskutiert wurde, versuchte RCELP diese Variation durch ein nichtkontinuierliches Verzerren des Restsignals zu eliminieren, um die Orte der Spitzen wieder anzupassen, so dass sie auf eine regelmäßige Art und Weise auftreten. Ein Modifizieren des Signals auf diese Weise erleichtert die Aufgabe eines Kodierers niederer Bit-Rate, da sehr wenige Bits dazu gebraucht werden, die Information über die Orte der Spitzen in dem modifizierten Signal zu versenden. Der Restwertmodifikation folgend, wird das modifizierte Restsignal auf die Sprachebenen zurücktransformiert, indem es durch eine Umkehr des Vorhersage- bzw. Prädiktionsfilters geschickt wird.

[0030] Allerdings ergibt sich aus der RCELP-basierten Signalmodifikation eine wahrnehmbare Abnahme der Stimmenqualität wegen den suboptimalen Eigenschaften der eingesetzten Verzerrungsfunktion. Insbesondere bei RCELP werden überlappende Abschnitte des originalen Restsignals, die jeweils eine

einzelne Spitze bzw. einen Zweig beinhalten, geschnitten und zusammengeschnürt, um das modifizierte Restsignal bzw. Restwertsignal zu erzeugen. Die geschnittenen Abschnitte können sich überlappen und tun das oft, woraus sich ergibt, dass einige Teile des Restsignals in dem modifizierten Restsignal zweifach auftauchen, während andere Teile überhaupt nicht auftauchen.

[0031] Die Erfindung überwindet diese ungewünschten Eigenschaften in der Modifikationsprozedur von RCELP, wie diskutiert, durch ein Verwenden eines kontinuierlichen Zeitverzerrungsalgorithmus, der in einer Ausführungsform der Erfindung mit einer verbesserten Verzerrungskontur- bzw. Warp-Kontur-Optimierungsverfahrensweise verbessert wird. Zusammenfassend identifiziert der erfindungsgemäße Algorithmus zuerst Stücke des originalen Restsignals, die eine einzelne Spitze beinhalten, so wie in RCELP. Allerdings überlappen sich diese Stücke im Unterschied zu RCELP nicht und decken den gesamten Rahmen bzw. Frame ab. Das bedeutet, wenn die geschnittenen Abschnitte verbunden werden würden, würde das originale Restsignal erhalten werden – kein Abschnitt des Restsignals würde zweifach erscheinen und kein Abschnitt würde weggelassen werden. Entweder beschleunigt oder verlangsamt der Algorithmus im Wesentlichen jedes Stück linear in einer kontinuierlichen und sich anpassenden Verzerrungsoperation, anstatt die Stücke einfach zu schneiden und zu bewegen bzw. zu verschieben, wie bei RCELP. Das Ziel beim Verzerren bzw. Warping von jedem Stuck ist es, sicherzustellen, dass die Spitzen in dem modifizierten Restsignal durch reguläre Intervalle separiert sind, wodurch die Bitrate reduziert wird, die dazu benötigt wird, die Positionen der Spitzen zu kodieren, wodurch das gleiche Ziel wie bei RCELP erreicht wird, ohne dessen Defizite. Wie diskutiert werden wird, ist der Grad der Beschleunigung oder der Verzögerung begrenzt, um eine Verschlechterung der Qualität der wiedergegebenen Sprache zu verhindern.

[0032] Nachdem die Erfindung oben im Generellen beschrieben wurde, werden die Details der bevorzugten Ausführungsform hiernach vollständiger beschrieben. Unter Bezug auf Fig. 1 ist eine beispielhafte Architektur zum Implementieren eines verbesserten Kodierers niederer Bitrate gemäß einer Ausführungsform der Erfindung dargestellt. Das System ist zusammengesetzt aus einem Digitalisierer **121**, einem Vorhersagefilter oder invertierten Synthesefilter **101**, einem linear kontinuierlichen Restwert- bzw. Restsignalmodifikationsmodul **103**, einem Synthesefilter **105** und einem Kodierer, so wie ein CELP-Kodierer **107**, die zusammen kaskadiert sind.

[0033] Der Vorhersagefilter **101** empfängt als Eingabe ein digitalisiertes Sprachsignal **109** von dem Digitalisierungsmodul **121**. Es gibt verschiedene Verfahren, die einem Fachmann bekannt sind, durch welche Sprache in ein digitales elektrisches Signal konvertiert werden kann, und dementsprechend werden solche Techniken hierin nicht groß im Detail diskutiert. Der Vorhersagefilter **101**, auf den sich manchmal auch als ein inverser bzw. invertierter oder umgekehrter Synthesefilter bezogen wird, ist einsetzbar bzw. betriebsbereit, um ein Restsignal **111** zu produzieren, das auf LPC-Koeffizienten und einem eingegebenen Signal bzw. Eingabesignal basiert. Fachmänner werden mit linear prädiktiven bzw. vorhersagenden Kodierungskonzepten, so wie dem invertierten Filter und dem Restsignal, vertraut sein. Das Restsignal **111** wird in das Restwertmodifikationsmodul **103** eingegeben, das das Signal in ein modifiziertes Restsignal **113** in einer Art und Weise konvertiert, die hiernach detaillierter diskutiert wird. Das modifizierte Restsignal **113** wird nachfolgend in einen Synthesefilter **105** eingegeben, um ein wiederhergestelltes bzw. wiedergegebenes Sprachsignal **115** zu erzeugen. Die Restsignalmodifikationstechnik, die durch das Restsignalmodifikationsmodul **103** implementiert wird, wird es dem modifizierten Sprachsignal **115** erlauben, sehr wie die originale Sprache **109** zu klingen, obwohl die Anregung oder das modifizierte Restsignal **113** sich von dem Restsignal **111** unterscheiden. Anschließend kodiert das CELP-Kodierungsmodul **107** das modifizierte Sprachsignal in einer Weise, die durch einen Fachmann gut verstanden wird, und gibt einen Strom (Stream) von kodierten Bits **117** zur Übertragung oder Speicherung aus.

[0034] Der Betrieb des Moduls, der in Fig. 1 dargestellt ist, wird nun in größerem Detail unter Bezug auf Fig. 2 in Verbindung mit Fig. 3a und Fig. 3b beschrieben. Im Einzelnen zeigt Fig. 2 vereinfachte Wellenformen **203**, **205**, **207**, **209**, **211**, die herausstechende Pitch- bzw. Tonhöhenspitzen **201** haben. Bemerke, dass die Spitzenverschiebungen, die in Fig. 2 dargestellt sind, zum Zwecke der Klarheit übertrieben sind. Eigentliche Verschiebungswerte sollten begrenzt werden, so wie hiernach diskutiert wird. Die Fig. 3a und Fig. 3b sind Flussdiagramme, welche die Schritte darstellen, die in einer Ausführungsform der Erfindung ausgeführt werden, um ein Sprachsignal zu kodieren. Bei einem Schritt **301** wird ein analoges Sprachsignal **119** von einem Digitalisierer **121** empfangen. In einem Schritt **303** tastet der Digitalisierer **121** das Signal bei einer Frequenz von 8 khz ab, um ein digital abgetastetes Audiosignal s(n) zu erhalten. Anschließend wird das Signal s(n) in einem Schritt **305** durch den Digitalisierer in nicht überlappende Rahmen bzw. Frames von 160 Abtastungen bzw. Samples gruppiert, die (20 ms) lang sind, wobei jedes von Ihnen in zwei sich nicht überlappende Unterrahmen bzw. Subframes unterteilt wird, die 80 Abtastungen (10 ms) lang sind. Somit wird das Signal in dem k-sten Rahmen durch s(160k) ... s(160k + 159) erhalten. Das gerahmte abgetastete Signal **109** wird von dem Digitalisierer **121** an den LPC-Ex-

trahierer **123** in einem Schritt **307** weitergereicht.

[0035] Der LPC-Extrahierer **123** arbeitet auf eine Art und Weise, die einen Fachmann wohlbekannt ist, um die dem eingegebenen Signal entsprechenden linear prädiktiven Koeffizienten zu berechnen. Im Einzelnen extrahiert der LPC-Extrahierer **123** in einem Schritt **309** einen Satz von prädiktiven Koeffizienten zehnter Ordnung für jeden Rahmen, indem er eine Korrelationsanalyse durchführt und den Levinson-Durbin-Algorithmus ausführt. Die optimalen linearen Prädiktionskoeffiiezenten in dem k-sten Rahmen $a_k(j)$, j = 1, ..., 10 werden in einem Schritt interpoliert, um einen Satz von LP-Koeffizienten $a_{ks}(j)$, j = 1, ..., 10 in jedem Unterrahmen zu erzeugen, worin s = 0,1 jeweils dem ersten und zweiten Unterrahmen entspricht bzw. zu diesem gehört. Die Interpolation kann durchgeführt werden, indem die LP-Koeffizienten in einen linienspektralen Frequenzbereich (Line Spectral Frequency – LSF – Domain) transformiert werden, indem LSF-Bereich linear interpoliert wird und die interpolierten Unterrahmen-LSF-Koeffizienten im LP-Koeffizienten zurücktransformiert werden. In einem Schritt **313** werden die Unterrahrnen-LP-Koeffizienten a_{ks} vom Vorhersagefilter **101** verwendet, um das Restsignal **111** in einer Weise zu reproduzieren, die einem Fachmann wohlbekannt ist. Das Restsignal **111** in dem k-sten Rahmen wird durch r(n), n = 160k ... 160k + 159 dargestellt.

[0036] Die dominanten Charakteristiken des Restsignals **111** können in der Wellenform **203** von Fig. 2 gesehen werden. Insbesondere für stimmhafte Abschnitte wird der Restwert **203** durch eine Abfolge von kaum periodischen, sondern unregelmäßig beabstandeten Spitzen oder Höchstwerten **201** dominiert. Diese Spitzen stellen typischerweise Stimmritzen- bzw. Glottallautimpulse dar, die den Sprachappart während des Vorgangs des Erzeugens von geäußerter Sprache anregen. Das Zeitintervall zwischen aneinanderangrenzenden Spitzen ist der Tonhöhen- bzw. Pitchperiode gleich. Menschliche Sprache hat typischerweise eine Pitch-Periode von zwischen etwa 2,5 ms und 18,5 ms. Das Intervall zwischen den Spitzen ist üblicherweise nicht konstant, sondern weist stattdessen kleinere Unregelmäßigkeiten oder Flimmern bzw. Zitterbewegungen auf.

[0037] Schritte **315** bis **333** werden den Betrieb des Restwertmodifikationsmoduls **103** beschreiben. in einem Schritt **315** empfängt das Restsignalmodifikationsmodul **103** das Restsignal **111** und bestimmt eine ganzzahlige Pitchperiode für den aktuellen Rahmen, den k-sten Rahmen. Die Pitch-Periode kann durch eine von einer Vielzahl von Techniken bestimmt werden, die im Stand der Technik bekannt sind. Eine Technik, die innerhalb dieser Ausführungsform anwendbar ist, ist der Einsatz einer Korrelationsanalyse in einer offenen Schleife. Was auch immer für ein Verfahren verwendet wird, hinreichende Sorgfalt sollte ausgeübt werden, um unerwünschte Artefakte, so wie eine Pitch-Verdopplung, zu vermeiden.

[0038] Bei einem Schritt **317** wird eine Interpolation der Pitchperiode des Rahmens durch eine lineare Interpolation von Abtastwert zu Abtastwert wie folgt ausgeführt:

$$c'(n) = p(k)\cdot((n - 160k)/160 + P(k - 1)\cdot(1 - (n - 160k)/160),\ n = 160k \ldots 160k + 159.$$

[0039] Die Funktion c'(n) kann als eine gerade Linie von p(k – 1) am Anfang des Rahmens bis p(k) am Ende des Rahmens dargestellt werden. Sie stellt eine sanft variierende Pitchperiode (floating point) für jedes Sample in dem aktuellen Rahmen dar.

[0040] In einem Schritt **319** wird eine Funktion c(n) durch ein Abrunden jedes Wertes von c'(n) zu dem nahesten Vielfachen von 0,125 gebildet. Effektiv ist c(n) ein Vielfaches von 1/8 und deswegen ist 8·c(n) eine ganzzahlige Pitchperiode in einem Bereich eines 8-fach überabgetasteten Signals. Hierin wird auf c(n) als die gewünschte Pitchkontur Bezug genommen. Die Wirkungen, die durch ein Modifizieren des Restsignals erzeugt werden, um diese idealisierte Kontur abzugleichen bzw. dieser zu entsprechen, sind signifikant. Zum Beispiel kann die Pitch-Periode eines Rahmens, der solch eine Kontur hat, übertragen werden, indem sehr wenige Bits verwendet werden und der Dekodierer kann den Pitch bzw. die Tonhöhe verwenden, um die Pitch-Kontur abzuleiten, und die Pitch-Kontur dann in Verbindung mit den Orten der Spitzen aus dem vorangegangenen Rahmen verwenden, um den Ort von Tonhöhen- bzw. Pitchspitzen für den aktuellen Rahmen abzuschätzen.

[0041] Der nächste Prozess ist dazu gedacht, den Dekodierer nachzuahmen und zu versuchen, die Orte der Spitzen in dem Restsignal des aktuellen Rahmens basierend auf der Pitch-Kontur und dem modifizierten Restsignal des vorangegangenen Rahmens zu rekonstruieren. Obwohl der eigentliche Kodierer typischerweise keinen Zugriff auf Informationen über das modifizierte Restsignal des vorangegangenen Rahmens haben wird, wird er Zugriff auf das Anregungssignal haben, das dazu verwendet wird, den vorangegangenen Rahmen zu rekonstruieren bzw. wiederherzustellen. Dementsprechend wird die Verwendung des vorangegangenen Anregungssignals durch den Dekodierer nicht mit der Verwendung des vorangegangenen modifizierten Restsignals in Konflikt geraten, da die Spitzen in dem Anregungssignal eines bestimmten Rahmens sich an die Spitzen in dem modifizierten Restsignal dieses Rahmens angleichen werden.

[0042] Um die Positionen der Spitzen in dem aktuellen Rahmen vorherzusagen, verwendet das Restsignalmodifikationsmodul **101** die Pitch-Kontur, um das

modifizierte Restsignal des vorangegangenen Rahmens in einem Schritt **321** zu verzögern, um ein Zielsignal für eine Modifikation $r_t(n)$ zu erzeugen. Eine beispielhafte Wellenform für $r_t(n)$ ist in Fig. 2 bei einem Element **211** gezeigt. Diese Zeitverzerrungsfunktion arbeitet im 8-fach überabgetasteten Bereich und verwendet einen Standard-Interpolationsfilter mit einer abgeschnitten bzw. verkürzten (truncated) sinc(x)-Impulsantwort und einem 90-prozentigen durchlass-band, da die Pitch-Kontur c(n) ein Vielfaches von 0,125 ist. Im Besonderen wird das 8-fache Überabtasten eingesetzt, um interpolierte Abtastwerte des modifizierten Restsignals r'(n) in dem vorangegangenen Rahmen zu erhalten, um wie folgt zu dem überabgetasteten Signal zu gelangen:

$$r''(n\cdot 0{,}125),\ n = 160\cdot 8\cdot(k-1) \ldots 160\cdot 8\cdot(k-1) + 1279.$$

[0043] Der Abtastwertindex von r'' ist ein Vielfaches von 0,125 und stellt die Überabtastungsbedingung bzw. den Überabtastungszustand dar. Nachfolgend wird eine Verzögerungslinienoperation ausgeführt, um das Zielsignal $r_t(n)$ wie folgt zu erhalten:

$$r_d(n\cdot 0{,}125) = r_d(n\cdot 0{,}125)\ n = 160\cdot 8\cdot(k-1) \ldots 160\cdot 8\cdot(k-1) + 1279$$

$$r_d(n\cdot 0{,}125) = r_d(n\cdot 0{,}125 - C(\mathrm{INT}(n\cdot 0{,}125))),\ n = 160\cdot 8\cdot k \ldots 160\cdot 8\cdot k + 1279$$

$$r_t(n) = r_d(n),\ n = 160\cdot k \ldots 160\cdot k + 159,$$

worin INT(x) den ganzzahligen Wert darstellt, der x am nächsten ist, eine Fließkommazahl (floating point number), und $r_d()$ ein dazwischenliegendes bzw. zwischengeschaltetes Signal ist. Beachte, dass der Kodierer eine identische Verzögerungslinienoperation bei dem Anregungssignal des vorangegangenen Rahmens ausführt. Nachdem die idealen Orte der Tonhöhenspitzen, die in dem Zielsignal **211** dargestellt sind, berechnet wurden, kann der Kodierer nun die Spitzen in dem eigentlichen Restsignal neu anordnen, um dieses in $r_t(n)$ anzugleichen. Anfänglich analysiert das Restsignalmodifikationsmodul **103** in einem Schritt **323** das unmodifizierte Restsignal **203**, um die verschiedenen Abschnitte des Signals zu identifizieren, die einen einzelnen hervorstechenden Höchstwert haben, der von einem Bereich niederer Energie umgeben ist. Eine beispielhaft resultierende Wellenform ist in Fig. 2 bei einem Element **205** dargestellt. Es gibt vorzugsweise keine Lücken zwischen Stücken von den Signalen, so wie sie unterteilt sind. In anderen Worten wird das Ergebnis des unmodifizierte Restsignals **203** sein, wenn die Stücke der Elemente **205** in diesem Stadium wieder zusammengesetzt werden würden. Vorzugsweise wird der Restwert bzw. das Restsignal **203** nur an Punkten geschnitten, die in der Wahrnehmung unsignifikant niedrige Energie haben. Anschließend assoziiert der Kodierer bei einem Schritt **325** einen Abschnitt des Zielsignals mit einem passenden Stück des unmodifizierten Restsignals.

[0044] Bei einem Schritt **327** berechnet das Restsignalmodifikationsmodul **103** eine optimale Verzerrungsfunktion für den identifizierten Abschnitt des unmodifizierten Restsignals, so dass eine Modifizierung über die optimale Verzerrungsfunktion die hervorstehenden Spitzen oder Maximalwerte in einem Segment des Restsignals **203** mit denen in dem assoziierten Abschnitt des Zielsignals **211** abgleichen wird. Die unternommenen Schritte, um eine optimale Verzerrungsfunktion für jeden der Abschnitte des Restsignals zu berechnen, werden unter Bezug auf die Fig. 4 dargestellt. Im Einzelnen stellt Fig. 4 die Ableitung einer Verzögerungs- bzw. Lag-Kontur l(n) dar, die die Verzögerung von Abtastwert zu Abtastwert zwischen dem Restsignal **203** und dem modifizierten Restsignal **209** repräsentiert. Die Menge l(n) ist ein Vielfaches von 0,125, so dass der modifizierte Restsignalabtastwert r'(m) dem Restsignalabtastwert in dem überabgetasteten Bereich entspricht, der durch l(m) verzögert ist. Das heißt:

$$r'(m) = r''(m - l(m)).$$

[0045] Das Problem des Findens der optimalen Verzerrungs- bzw. Warp-Kontur wird auf das Problem des Findens der optimalen Lag-Kontur l(n) reduziert.

[0046] Bei einem Schritt **401** wird die Verzögerung bzw. der Lag l_f für jeden ersten Abtastwert des aktuellen Abschnittes von Interesse gleich der Verzögerung für jeden letzten Abtastwert des vorangegangenen Abschnittes gesetzt und ein Satz von Kandidaten für die Verzögerung l, für den letzten Abtastwert des aktuellen Abschnittes wird identifiziert. Insbesondere wird ein Satz von 2K + 1 Kandidaten für die Verzögerung l_l des letzten Abtastwertes innerhalb eines Kandidatenbereiches identifiziert, so wie $\{l_f - K, l_f - K + 1, \ldots l_f + K\}$. Der Wert von K wird basierend auf Parametern, so wie der zur Verfügung stehenden Rechenleistung, ausgewählt, und die Periodizität jedes Sprachabtastwertes bzw. -samples und der Wert von l_f. Typischer Werte von K sind 0, 1, 2, 3 oder 4. Obwohl der Bereich von Kandidaten, der durch die obige Gleichung dargestellt ist, symmetrisch um l_f fällt, muss dies nicht der Fall sein.

[0047] Obwohl ein Verschieben der Abschnitte des Restsignals durch kleine Werte keinen negativen Effekt auf die wahrgenommene Qualität des reproduzierten Signals hat, können größere Verschiebungen wahrnehmbar negative Effekt haben. Folglich ist es wünschenswert, die Größe, durch welche ein Abtastwert verschoben werden kann, auf eine kleine Zahl zu reduzieren, so wie drei originale (nicht überabgetastete) Abtastwertinkremente inklusive jeglichen angehäuften Verschiebungen als ein Ergebnis des Verschiebens des vorangegangenen Abschnittes oder

Stückes. Somit sollte dann der letzte Abtastwert des aktuellen Stückes nicht zusätzlich mehr als das äquivalent einer Abtastwertposition verschoben werden, wenn der letzte Abtastwert in dem vorangegangenen Stück durch das Äquivalent von zwei Abtastwertpositionen verschoben wurde, oder er wird eine totale Verschiebung von mehr als drei Abtastwertpositionen von seinem ursprünglichen Ort erfahren. Die Lösung für dieses Problem ist es, den Wert für K zu begrenzen, so dass er keine Verschiebung über einen gewünschten Bereich hinaus zulässt oder einen asymmetrischen Bereich von Kandidaten zu verwenden. Folglich kann in dem obigen Beispiel eine Beschleunigung durch fünf Abtastwertpositionen zugelassen werden, obwohl eine Verzögerung durch mehr als einen Abtastwert unzulässig ist, wenn eine asymmetrische Verteilung von Kandidaten für Verzögerungs- bzw. Lag-Werte verwendet wird.

[0048] Beachte, dass weniger als die möglichen Verzögerungskandidaten in dem Satz von Kandidaten sind, weil die Rechenleistung, die zum Bewerten aller möglichen Verzögerungskandidaten notwendig ist, nicht zulässig wäre. Es wird eher nur eine Untergruppe von möglichen Verzögerungswerten für den letzten Abtastwert in einem aktuellen Abschnitt als Kandidaten verwendet. Verzögerungswerte außerhalb des Kandidatenbereiches werden nicht in den Satz bzw. die Gruppe mit einbezogen, noch werden die Werte mit einbezogen, die zwischen den Verzögerungswertkandidaten liegen. Somit kann der optimale Verzögerungswert für den letzten Abtastwert (eine sich ergebene Lag-Kontur) noch nicht einmal in dem Kandidatensatz selber beinhaltet sein, aber er ist vorzugsweise innerhalb des Kandidatenbereiches angeordnet.

[0049] Als nächstes führt der Kodierer in einem Schritt **403** eine lineare Interpolation zwischen den ersten und letzten Abtastwerten des aktuellen Abschnittes für jeden Verzögerungswertkandidaten durch, der in einem Schritt **401** identifiziert wurde, um einen Satz von 2K + 1 Verzögerungskonturkandidaten zu erzeugen. Ein Verzögerungskonturkandidat stellt eine lineare Funktion dar, durch die der erste und der letzte Wert jeweils l_f und l_l sind, worin l_l ein Kandidatenwert ist. In einem Schritt **405** wird jeder Verzögerungskonturkandidat auf das Restsignal angewendet, um einen Satz von 2K + 1 modifizierten Restsignalkandidaten zu erhalten, und die Korrelation zwischen dem Zielsignal $r_t(n)$ **211** und jedem modifizierten Restsignalkandidaten wird in einem Schritt **407** berechnet.

[0050] In einem Schritt **409** wird die Stärke der Korrelation automatisch quadratisch als eine Funktion des letzten Abtastwertverzögerungswertes modelliert und der optimale Verzögerungswert für den letzten Abtastwert wird erhalten. Im Einzelnen wird die Stärke der Korrelation für jeden modifizierten Restwertkandidaten als eine Funktion des assoziierten letzten Abtastwertverzögerungswertkandidaten gezeichnet, wie durch die Zeichenpunkte in dem Graphen von Fig. 5 dargestellt. Als nächstes werden die Zeichenpunkte in Sätze unterteilt, wobei jeder Satz aus drei Punkten besteht. Es gibt eine Überlappung von einem Punkt zwischen aneinander angrenzenden Sätzen. Die 2K + 1 Zeichenpunkte würden somit in K überlappende Sätze von jeweils drei Punkten unterteilt werden. Für sieben Punkte zum Beispiel würde es drei Sätze geben. Jeder Satz von drei aufeinanderfolgenden Zeichenpunkten wird gemäß einer quadratischen Funktion modelliert. In Fig. 5 zum Beispiel sind drei quadratische Modellierungsfunktionen als **501**, **503** und **505** dargestellt. Das Maximum von jeder quadratischen Funktion in einem Bereich von dem ersten bis zu dem letzten der assoziierten drei Punkte wird erhalten und ein Maximum des gesamten Abschnittes wird dann berechnet. Folglich wird für positive quadratische Funktionen, d. h. diejenigen, die konkav nach oben weisen, sowie für monotone Anordnungen von Punkten der maximale Korrelationswert an einem der Endpunkte liegen. Beachte, dass generell das Maximum für einen gegebenen Satz von drei Punkten nicht immer bei einem der drei Punkte liegen wird, aber oft irgendwo dazwischen liegen wird. Folglich könnte der optimale Verzögerungswert für den gesamten Abschnitt ein Wert sein, der nicht in dem Satz von Kandidaten für die Verzögerung bzw. den Lag l_l war.

[0051] Obwohl die graphische Darstellung von Fig. 5 hierin verwendet wird, um graphische Schritte gemäß einer Ausführungsform der Erfindung darzustellen, erfordern die Begriffe „Zeichen" oder „gezeichnet", so wie sie hierin verwendet werden, kein Erschaffen eines konkreten oder sichtbaren Graphen. Diese Begriffe implizieren einfach eher die Erschaffung einer Verknüpfung zwischen Größen, sei diese implizit, so als wären die Achsen, die verwendet werden, verschiedene Parameter, die sich auf in Fig. 5 gezeigte Mengen beziehen, oder explizit, und sei es tatsächlich, wie in einer graphischen Programmdatenstruktur, oder virtuell, wie in einem Satz von Zahlen in einem Speicher, von dem die passende Beziehung abgeleitet werden kann. Dementsprechend bezeichnen diese Begriffe einfach das Erschaffen einer Beziehung zwischen den angezeigten Mengen, wie auch immer so eine Beziehung errichtet wird.

[0052] Das Maximum von allen Quadratischen für die aktuelle Korrelationszeichnung wird einem Verzögerungswert für den letzten Abtastwert über die passende Quadratische assoziiert und dieser Wert ist der optimale letzte Verzögerungsabtastwert. Es ist nicht notwendig, dass eine quadratische Funktion verwendet wird, um den Satz von Punkten zu modellieren, oder dass es drei Punkte sind, die verwendet werden. Zum Beispiel könnte der Satz mehr als drei

Punkte beinhalten und die Modellierungsfunktion kann eine polynomische Funktion irgendeiner Ordnung sein, abhängig von dem akzeptablen Grad von Komplexität. Beachte auch, dass für monotone Abfolgen von Punkten es nicht notwendig ist, eine Abfolge als ein Polynom oder anderweitig zu modellieren, da der höchste Endpunkt einfach bestimmt wird und das Maximum der Abfolge bzw. Sequenz darstellt.

[0053] Nachdem die optimalen Verzögerungswerte für den letzten Abtastwert des aktuellen einen dominanten Maximalwert beinhaltenden Abschnittes oder Segments von Interesse bestimmt wurden, leitet das Restsignalmodifikationsmodul **103** in einem Schritt **411** eine dazugehörige Verzögerungs- bzw. Lag-Kontur ab, indem über den Abschnitt l_f bis zum optimalen l_l, was in dem Schritt **409** berechnet wird, linear interpoliert wird. Bei dem Schritt **329** in Fig. 3b wird die in dem Schritt **411** von Fig. 4 berechnete Verzögerungskontur auf das Restsignal wie oben beschrieben angewendet, das heißt:

$$r'(n) = r''(n - l(n)).$$

[0054] Schließlich wird bei einem Schritt **331** festgelegt, ob es irgendwelche weiteren Stücke in dem aktuellen Rahmen gibt, die zu analysieren und zu verschieben sind. Wenn es welche gibt, führt der Betriebsfluss zum Schritt **325** zurück. Andererseits endet der Prozess für den aktuellen Rahmen bei einem Schritt **333**. Eine Element **207** von Fig. 2 stellt verzerrte Abschnitte des modifizierten Restsignals **209** zwecks Klarheit separat dar. Das als eine Wellenform **209** dargestellte modifizierte Restsignal **113** wird schließlich als eine Eingabe für den Synthesefilter **105** zur Verfügung gestellt, um zu einer Wiedergabe bzw. Reproduktion des originalen Sprachsignals zu führen, wobei die Reproduktion eher reguläre als verschobenen Tonhöhenspitzen hat. Von diesem Punkt aus wird das Signal unter Verwendung einer Technik wie einer üblichen CELP verarbeitet. Allerdings ist die Bitrate, die nun dazu benötigt wird, um das Signal zu kodieren, stark gegenüber derjenigen reduziert, die dazu benötigt wird, um das unmodifizierte Signal zu kodieren, wegen der erhöhten Periodizität der Tonhöhenstruktur.

[0055] Nachdem ein Rahmen verarbeitet wurde, beginnt der Prozess bei einem nachfolgenden Rahmen. Im Falle eines ungesprochenen Segmentes gibt es typischerweise keine Tonhöhenspitzen und es muss die hierin beschriebenen Verfahrensweise nicht angewendet werden. Während des ungesprochenen Intervalls werden alle Werte in dem Algorithmus zurückgesetzt. Zum Beispiel wird die Anzeige von angehäuften Verschiebungen zu Null zurückgesetzt. Wenn eine geäußerte Sprache wiederaufgenommen wird, wird der erste gesprochene Rahmen k als ein Spezialfall behandelt, weil der Tonhöhenwert des vorangehenden Rahmens p(k – 1) nicht in diesem Rahmen bekannt ist. Die Tonhöhen- bzw. Pitch-Kontur wird in diesem speziellen Rahmen k zu einer konstanten Funktion gesetzt, die dem Tonhöhen- bzw. Pitch-Wert des Rahmens p(k) gleich ist. Der Rest der Prozedur ist identisch zu der von regulären Rahmen.

[0056] Beachte, dass andere Techniken als ein polynomisches Modellieren innerhalb der Erfindung verwendet werden können, um einen optimalen Verzögerungswert l_f und eine assoziierte Verzögerungskontur für einen gegebenen Abschnitt oder ein Stück eines Sprachsignals innerhalb des aktuellen Rahmens verwendet werden können. Es ist nur von Bedeutung für die Erfindung, dass eine wesentliche Untergruppe bzw. ein Untersatz von möglichen Verzögerungswerten, zum Beispiel die Hälfte von allen möglichen Verzögerungswerten, zum Erschaffen von Korrelationswerten verwendet werden, da dies einen erheblichen Rechnungsaufwand zum Finden der optimalen Verzögerungskontur reduziert. Folglich können alternative Techniken, so wie Zweiteilung verwendet werden, um die optimalen Verzögerungswerte ohne alle oder sogar die meisten möglichen Verzögerungswerte auszuprobieren. Die Zweiteilungstechnik zieht ein Identifizieren von zwei Verzögerungskandidatenwerten nach sich und deren assoziierte Korrelationsstärken. Die Verzögerungskandidaten mit höherer Korrelation und ein neuer Verzögerungskandidat, der zwischen den beiden Verzögerungskandidaten liegt, werden als Endpunkte verwendet, um den Zweiteilungsprozess zu wiederholen. Dieser Prozess kann nach einer vorbestimmten Anzahl von Iterationen abgeschlossen werden oder wenn ein Verzögerungswert, der zu einer Korrelationsstärke über einen vorbestimmten Schwellwert liegt, gefunden wird.

[0057] Eine kontinuierliche lineare Verzögerungskontur, die sich aus der hierin beschriebenen Verfahrensweise ergibt, ist in Fig. 6 dargestellt. Im Einzelnen ist die kontinuierliche lineare Verzögerungskontur **601** als eine durchgehende schwarze Linie gezeigt, während die diskontinuierliche Kontur **603**, die in der RCELP-Technik im Stand der Technik verwendet wird, als eine gestrichelte Linie dargestellt ist. Beide Konturen repräsentieren Linien, die durch die Sätze von Punkten für Signalabtastwerte verlaufen, die als eine Funktion der ursprünglichen Zeit (Vorverzögerung bzw. pre-warp) gegenüber der modifizierten Zeit (Nachverzögerung bzw. post-warp) gezeichnet sind. Folglich repräsentiert jeder gerade Abschnitt in der Kontur **601** und jedes separate Stück der Kontur **603** einen Abschnitt des originalen bzw. ursprünglichen Restsignals, das gemäß der jeweiligen Technik verzerrt wurde. Es kann gesehen werden, dass sich aus der RCELP-Technik oft fehlende oder überlappende Abschnitte ergeben, während die kontinuierliche lineare Verzerrungskontur der vorliegenden Erfindung Überlappungen oder Weglassungen nicht zulässt. Obwohl die kontinuierliche lineare Verzer-

rungskontur **601** Unterbrechungen in ihrer Steigung haben kann, ist sie eher kontinuierlich als einfach nur stückweise kontinuierlich in ihrer Position. Insbesondere eine Region **605** wird durch zwei Stücke von Verzerrungskonturen **603** besetzt, während ein Abschnitt **607** frei von Daten ist, die der gleichen Kontur folgen. Andererseits wird der gesamte Signalraum ohne Überlappungen oder Weglassungen durch eine Kontur **601** gemäß der vorliegenden Erfindung eingenommen.

[0058] Beachte, dass die Verzerrungskontur **601** für aneinanderangrenzende Abschnitte bzw. Segmente die gleiche Steigung oder verschiedene Steigungen haben kann, abhängig von der Beschleunigung oder Verzögerung, die für jedes Segment benötigt wird. Im Gegensatz ist die Steigung jedes Abschnittes der RCELP-Kontur **603** ungleichförmig. Dies resultiert daraus, dass RCELP Abschnitte des Signals verschiebt, aber nicht die Zeitlinie bzw. Zeiteinheit innerhalb der Abschnitte ändert. Folglich kann beobachtet werden, dass das Verfahren gemäß der Erfindung die Zeitlinie innerhalb jedes Abschnittes einer linearkontinuierlichen Art und Weise verzerrt, so dass die Spitzen jedes Abschnittes zu dem gewünschten Ort verschoben werden, ohne dass ungewünschte Zeitlinienbrechungen an den Kanten der Abschnitte erzeugt werden.

[0059] Obwohl es nicht erforderlich ist, kann die vorliegende Erfindung unter Verwendung von Befehlen, so wie Programm-„Modulen", implementiert werden, die von einem Computer ausgeführt werden. Generell beinhalten Programmmodule Routinen, Objekte, Komponenten, Datenstrukturen und Ähnliches, die bestimmte Aufgaben ausführen oder bestimmte abstrakte Datentypen implementieren. Ein Programm kann ein oder mehrere Programmmodule beinhalten.

[0060] Die Erfindung kann als eine Vielzahl von Typen von Maschinen implementiert werden, inklusive Mobiltelefone, Personalcomputer (PCs), handgehaltene Geräte, Multiprozessorsysteme, mikroprozessorbasierte programmierbare Verbraucherelektronik, Netzwerk-PCs, Minicomputer, Großrechner und Ähnliches oder eine andere Maschine, die zum Kodieren oder Dekodieren von Audiosignalen, wie hierin beschrieben ist, verwendet werden kann oder die Signale speichern, abrufen, übertragen oder empfangen kann. Die Erfindung kann an einem verteilten Computersystem eingesetzt werden, worin Aufgaben durch voneinander fernliegende Komponenten durchgeführt werden, die durch ein Kommunikationsnetzwerk miteinander verbunden sind.

[0061] Unter Bezug auf Fig. 7 beinhaltet ein beispielhaftes System zum Implementieren von Ausführungsformen der vorliegenden Erfindung eine Rechen- bzw. Computervorrichtung, so wie ein Rechengerät bzw. eine Computervorrichtung **700**. In dieser einfachsten bzw. grundlegendsten Konfiguration beinhaltet die Computervorrichtung **700** typischerweise wenigstens eine Prozessoreinheit **702** und einen Speicher **704**. Abhängig von der genauen Konfiguration und der Art der Computervorrichtung kann ein Speicher **704** flüchtig sein (so wie ein RAM), nicht flüchtig sein (so wie ein ROM, flash memory etc.) oder eine Kombination der beiden. Diese einfachste Konfiguration ist in Fig. 7 innerhalb einer Linie **706** dargestellt. Zusätzlich kann das Gerät **700** auch zusätzliche Merkmale oder Funktionen haben. Zum Beispiel kann die Vorrichtung **700** auch einen zusätzlichen Speicher (entfernbar und/oder nicht entfernbar) beinhalten, der nicht auf magnetische oder optische Scheiben oder Bänder beschränkt ist. Ein zusätzlicher Speicher ist in Fig. 7 durch einen entfernbaren Speicher **708** und einen nicht entfernbaren Speicher **710** dargestellt. Computerspeichermedien beinhalten flüchtige und nicht flüchtige, entnehmbare und nicht nicht entnehmbare Medien, die nach irgendeinem Verfahren oder einer Technologie zum Speichern von Informationen implementiert werden, so wie computerlesbare Instruktionen, Datenstrukturen, Programmmodule oder andere Daten. Der Speicher **704**, der entnehmbare Speicher **708** und der nicht entnehmbare Speicher **710** sind alles Beispiele von Computerspeichermedien. Ein Computerspeichermedium beinhaltet, aber ist nicht auf RAM, ROM, EEPROM, flash memory oder andere Speichertechnologien, CDROM, digital versatile disc (DVD) oder andere optische Speicher, magnetische Kassetten, magnetische Bänder, magnetische Scheibenspeicher oder andere magnetische Speichervorrichtungen oder irgendein anderes Medium, das dazu verwendet werden kann, die gewünschten Informationen zu speichern, und auf das von der Vorrichtung **700** zugegriffen werden kann. Alle solchen Computerspeichermedien können ein Teil der Vorrichtung **700** sein.

[0062] Die Vorrichtung **700** kann auch einen oder mehrere Kommunikationsverbindungen **712** beinhalten, die es der Vorrichtung erlauben, mit anderen Geräten zu kommunizieren. Kommunikationsverbindungen **712** sind ein Beispiel von Kommunikationsmedien. Kommunikationsmedien verkörpern typischerweise computerlesbare Instruktionen, Datenstrukturen, Programmmodule oder andere Daten in einem modulierten Datensignal, so wie einer Trägerwelle oder anderen Transportmechanismen, und beinhalten jegliche informationsliefernde Medien. Der Begriff „moduliertes Datensignal" bedeutet ein Signal, das eine oder mehrere Charakteristiken hat, die in einer Weise gesetzt oder geändert werden, das sie die Information in dem Signal kodieren. Zum Zwecke eines Beispiels, aber nicht einschränkend beinhalten Kommunikationsmedien verkabelte Medien, so wie ein verkabeltes Netzwerk oder eine direkte Kabelverbindung, und kabellose Medien, so wie Akustik, RF, Infrarot und andere kabellose Medien. Wie oben disku-

tiert wurde, beinhaltet der Begriff computerlesbares Medium, so wie er hierin verwendet wird, sowohl Speichermedien als auch Kommunikationsmedien.

[0063] Die Vorrichtung **700** kann auch ein oder mehrere Eingabegeräte **714**, so wie eine Tastatur, eine Maus, einen Stift, ein Stimmeneingabegerät, ein berührungsempfindliches Gerät usw., aufweisen. Eines oder mehrere Ausgabegeräte **716**, so wie einen Bildschirm, Lautsprecher, einen Drucker usw., können auch enthalten sein. Alle diese Geräte sind im Stand der Technik wohlbekannt und brauchen hier nicht ausführlich diskutiert zu werden.

[0064] In Anbetracht der vielen möglichen Ausführungsformen, auf die die Prinzipien dieser Erfindung angewendet werden können, sollte es erkannt werden, dass die hierin unter Bezug auf die Zeichnungsfiguren beschriebenen Ausführungsformen nur dazu gedacht sind, darstellend zu sein, und nicht dazu verwendet werden sollten, den Anwendungsbereich der vorliegenden Erfindung zu begrenzen. Zum Beispiel werden es Fachmänner erkennen, dass die Elemente der dargestellten Ausführungsformen, die als Software gezeigt sind, als Hardware implementiert werden können und umgekehrt oder dass die dargestellten Ausführungsformen in ihrer Anordnung und im Detail modifiziert werden können. Folglich betrachtet die Erfindung, so wie sie hierin beschrieben ist, alle solche Ausführungsformen, die da kommen mögen, innerhalb des Anwendungsbereichs der folgenden Ansprüche.

Patentansprüche

1. Verfahren zum Vorbereiten eines Rahmens eines digitalen Sprachsignals für Kompression, das die folgenden Schritte umfasst:
Erzeugen (**313**) eines Restsignals (**203**) linearer Prädiktion für den Rahmen, wobei das Restsignal linearer Prädiktion unregelmäßig beabstandete dominante Spitzen (**201**) aufweist;
Teilen (**323**) des Restsignals in eine Reihe zusammenhängender, nicht überlappender Abschnitte, wobei jeder Abschnitt nicht mehr als eine dominante Spitze enthält;
Herleiten (**321**) eines idealisierten Signals (**211**), das eine Reihe regelmäßig beabstandeter dominanter Spitzen aufweist, die in einer Reihe sequenzieller Abschnitte angeordnet sind;
Verknüpfen (**325**) jedes Abschnitts des Restsignals mit einem entsprechenden Abschnitt des idealisierten Signals;
Berechnen (**327**) einer linearen kontinuierlichen Warp-Kontur auf Basis einer Teilgruppe möglicher Lag-Werte für den letzten Abtastwert in dem jeweiligen Restsignal-Abschnitt innerhalb eines Teilbereiches möglicher Lag-Werte für diesen letzten Abtastwert für jeden Restsignal-Abschnitt; und
Modifizieren des Restsignals durch Anwenden (**329**) der berechneten Warp-Kontur auf die Abschnitte des Restsignals, um den Abschnitt des Restsignals kontinuierlichem Warping zu unterziehen, ohne irgendeinen Teil eines Abschnitts des Restsignals wegzulassen oder zu wiederholen, so dass jede dominante Spitze in jedem Abschnitt des Restsignals auf die dominante Spitze in dem entsprechenden Abschnitt des idealisierten Signals ausgerichtet ist und dominante Pitch-Spitzen in dem modifizierten Restsignal regelmäßig beabstandet sind.

2. Verfahren nach Anspruch 1, wobei der Schritt des Erzeugens eines Restsignals linearer Prädiktion für den Rahmen des Weiteren die folgenden Schritte umfasst:
Extrahieren von Koeffizienten linearer Prädiktion für den Rahmen;
Interpolieren der Koeffizienten linearer Prädiktion für den Rahmen, um Koeffizienten linearer Prädiktion für eine Vielzahl von Teil-Rahmen des Rahmens zu schaffen; und Erzeugen eines Prädiktions-Restsignals für jeden Teil-Rahmen, wobei das Prädiktions-Restsignal für den Rahmen eine Gruppe von Prädiktions-Restsignalen des Teilrahmens umfasst.

3. Verfahren nach Anspruch 1, wobei der Schritt des Teilens des Restsignals in eine Reihe zusammenhängender, nicht überlappender Abschnitte des Weiteren die Schritte des Analysierens des Rahmens zum Identifizieren einer ganzzahligen Pitch-Periode umfasst.

4. Verfahren nach Anspruch 3, wobei der Schritt des Analysierens des Rahmens zum Identifizieren einer ganzzahligen Pitch-Periode des Weiteren den Schritt des Verwendens von Korrelations-Analyse in der offenen Schleife umfasst.

5. Verfahren nach Anspruch 1, wobei der Schritt des Berechnens einer linearen kontinuierlichen Warp-Kontur für jeden Restsignal-Abschnitt des Weiteren die folgenden Schritte umfasst:
Einrichten eines ersten Abtastwert-Lag für den ersten Abtastwert des Restsignal-Abschnitts;
Identifizieren einer Gruppe von Kandidaten für den letzten Abtastwert-Lag für den letzten Abtastwert des Restsignal-Abschnitts, wobei die Gruppe von Kandidaten aus einer Teilgruppe aller möglichen Lag-Werte für den letzten Abtastwert des jeweiligen Restsignal-Abschnitts innerhalb eines Teilbereiches aller möglichen Lag-Werte für diesen letzten Abtastwert besteht;
Durchführen einer linearen Interpolation zwischen dem ersten und dem letzten Abtastwert des Restsignal-Abschnitts für jeden Kandidaten für den letzten Abtastwert-Lag, um eine Gruppe von Kandidaten für Lag-Konturen zu schaffen;
Anwenden jedes Kandidaten für die Lag-Kontur auf den Restsignal-Abschnitt, um eine Gruppe von Kandidaten für modifizierte Restsignale zu gewinnen;

Berechnen einer Korrelationsstärke zwischen jedem Kandidaten für das modifizierte Restsignal und dem entsprechenden Abschnitt des idealisierten Signals, um eine Gruppe von Korrelationsstärken zu schaffen;
Herleiten eines optimalen letzten Abtastwert-Lag für den Restsignal-Abschnitt auf Basis der Gruppe von Korrelationsstärken; und
Herleiten einer linearen kontinuierlichen Warp-Kontur durch lineares Interpolieren über den Abschnitt von dem ersten Abtastwert-Lag zu dem hergeleiteten optimalen letzten Abtastwert-Lag für den Restsignal-Abschnitt.

6. Verfahren nach Anspruch 5, wobei der Schritt des Herleitens eines optimalen letzten Abtastwert-Lag für den Restsignal-Abschnitt auf Basis der Gruppe von Korrelationsstärken des Weiteren die folgenden Schritte umfasst:
Trennen der Gruppe von Korrelationsstärken in überlappende Teilabschnitte als eine Funktion der zum Herleiten der Stärken verwendeten letzten Abtastwert-Lags;
Darstellen jedes Teilabschnitts als eine Kurve;
Berechnen des Maximalwertes jeder Kurve, wobei der Maximalwert aus der Gruppe ausgewählt werden kann, die aus allen möglichen Lag-Werten innerhalb eines Bereiches möglicher Lag-Werte besteht, der die zum Herleiten der Stärken in dem Teilabschnitt verwendeten letzten Abtastwert-Lags einschließt; und
Berechnen der maximalen Korrelationsstärke für den Abschnitt auf Basis der Maximalwerte für die Kurven der Teilabschnitte.

7. Verfahren nach Anspruch 6, wobei die Kurve ein Polynom ist.

8. Verfahren nach Anspruch 7, wobei das Polynom eine quadratische Funktion ist.

9. Verfahren nach Anspruch 1, wobei der Teilbereich möglicher Lag-Werte für den letzten Abtastwert für jeden Restsignal-Abschnitt so ausgewählt wird, dass die größte kumulative Verschiebung für jeden beliebigen Abtastwert in dem Abschnitt bei Anwendung der berechneten Warp-Kontur geringer ist als vier Abtastwertpositionen.

10. Vorrichtung zum Modifizieren eines Sprachsignals vor Kodieren des Sprachsignals, wobei die Vorrichtung Einrichtungen umfasst, die zum Ausführen aller Schritte des Verfahrens nach Anspruch 1 eingerichtet sind.

11. Vorrichtung nach Anspruch 10, das des Weiteren ein CLP(codebook excited linear prediction)-Kodiermodul zum Empfangen des modifizierten digitalen Sprachsignals und zum Erzeugen eines komprimierten Sprachsignals umfasst.

12. Computerlesbares Medium, das durch Computer lesbare Befehle zum Durchführen eines Verfahrens zum Vorbereiten eines Rahmens eines digitalen Sprachsignals für Kompression aufweist, das die folgenden Schritte umfasst:
Erzeugen (**313**) eines Restsignals (**203**) linearer Prädiktion für den Rahmen, wobei das Restsignal linearer Prädiktion unregelmäßig beabstandete dominante Spitzen (**201**) aufweist;
Teilen (**323**) des Restsignals in eine Reihe zusammenhängender, nicht überlappender Abschnitte, wobei jeder Abschnitt nicht mehr als eine dominante Spitze enthält;
Herleiten (**321**) eines idealisierten Signals (**211**), das eine Reihe regelmäßig beabstandeter dominanter Spitzen aufweist, die in einer Reihe sequenzieller Abschnitte angeordnet sind;
Verknüpfen (**325**) jedes Abschnitts des Restsignals mit einem entsprechenden Abschnitt des idealisierten Signals;
Berechnen (**327**) einer linearen kontinuierlichen Warp-Kontur auf Basis einer Teilgruppe möglicher Lag-Werte für den letzten Abtastwert in dem jeweiligen Restsignal-Abschnitt innerhalb eines Teilbereiches möglicher Lag-Werte für diesen letzten Abtastwert für jeden Restsignal-Abschnitt; und
Modifizieren des Restsignals durch Anwenden (**329**) der berechneten Warp-Kontur auf die Abschnitte des Restsignals, um den Abschnitt des Restsignals kontinuierlichem Warping zu unterziehen, ohne irgendeinen Teil eines Abschnitts des Restsignals wegzulassen oder zu wiederholen, so dass jede dominante Spitze in jedem Abschnitt des Restsignals auf die dominante Spitze in dem entsprechenden Abschnitt des idealisierten Signals ausgerichtet ist und dominante Pitch-Spitzen in dem modifizierten Restsignal regelmäßig beabstandet sind.

13. Computerlesbares Medium nach Anspruch 12, wobei der Schritt des Erzeugens eines Restsignals linearer Prädiktion für den Rahmen des Weiteren die folgenden Schritte umfasst:
Extrahieren von Koeffizienten linearer Prädiktion für den Rahmen;
Interpolieren der Koeffizienten linearer Prädiktion für den Rahmen, um Koeffizienten linearer Prädiktion für eine Vielzahl von Teil-Rahmen des Rahmens zu schaffen; und
Erzeugen eines Prädiktions-Restsignals für jeden Teil-Rahmen, wobei das Prädiktions-Restsignal für den Rahmen eine Gruppe von Prädiktions-Restsignalen des Teilrahmens umfasst.

14. Computerlesbares Medium nach Anspruch 12, wobei der Schritt des Teilens des Restsignals in eine Reihe zusammenhängender, nicht überlappender Abschnitte des Weiteren die Schritte des Analysierens des Rahmens zum Identifizieren einer ganzzahligen Pitch-Periode umfasst.

15. Computerlesbares Medium nach Anspruch 14, wobei der Schritt des Analysierens des Rahmens zum Identifizieren einer ganzzahligen Pitch-Periode des Weiteren den Schritt des Verwendens von Korrelations-Analyse in der offenen Schleife umfasst.

16. Computerlesbares Medium nach Anspruch 12, wobei der Schritt des Berechnens einer linearen kontinuierlichen Warp-Kontur für jeden Restsignal-Abschnitt des Weiteren die folgenden Schritte umfasst:
Einrichten eines ersten Abtastwert-Lag für den ersten Abtastwert des Restsignal-Abschnitts;
Identifizieren einer Gruppe von Kandidaten für den letzten Abtastwert-Lag für den letzten Abtastwert des Restsignal-Abschnitts, wobei die Gruppe von Kandidaten aus einer Teilgruppe aller möglichen Lag-Werte für den letzten Abtastwert des jeweiligen Restsignal-Abschnitts innerhalb eines Teilbereiches aller möglichen Lag-Werte für diesen letzten Abtastwert besteht;
Durchführen einer linearen Interpolation zwischen dem ersten und dem letzten Abtastwert des Restsignal-Abschnitts für jeden Kandidaten für den letzten Abtastwert-Lag, um eine Gruppe von Kandidaten für Lag-Konturen zu schaffen;
Anwenden jedes Kandidaten für die Lag-Kontur auf den Restsignal-Abschnitt, um eine Gruppe von Kandidaten für modifizierte Restsignale zu gewinnen;
Berechnen einer Korrelationsstärke zwischen jedem Kandidaten für das modifizierte Restsignal und dem entsprechenden Abschnitt des idealisierten Signals, um eine Gruppe von Korrelationsstärken zu schaffen;
Herleiten eines optimalen letzten Abtastwert-Lag für den Restsignal-Abschnitt auf Basis der Gruppe von Korrelationsstärken; und
Herleiten einer linearen kontinuierlichen Warp-Kontur durch lineares Interpolieren über den Abschnitt von dem ersten Abtastwert-Lag zu dem hergeleiteten optimalen letzten Abtastwert-Lag für den Restsignal-Abschnitt.

17. Computerlesbares Medium nach Anspruch 16, wobei der Schritt des Herleitens eines optimalen letzten Abtastwert-Lag für den Restsignal-Abschnitt auf Basis der Gruppe von Korrelationsstärken des Weiteren die folgenden Schritte umfasst:
Trennen der Gruppe von Korrelationsstärken in überlappende Teilabschnitte als eine Funktion der zum Herleiten der Stärken verwendeten letzten Abtastwert-Lags;
Darstellen jedes Teilabschnitts als eine Kurve;
Berechnen des Maximalwertes jeder Kurve, wobei der Maximalwert aus der Gruppe ausgewählt werden kann, die aus allen möglichen Lag-Werten innerhalb eines Bereiches möglicher Lag-Werte besteht, der die zum Herleiten der Stärken in dem Teilabschnitt verwendeten letzten Abtastwert-Lags einschließt; und
Berechnen der maximalen Korrelationsstärke für den Abschnitt auf Basis der Maximalwerte für die Kurven der Teilabschnitte.

18. Computerlesbares Medium nach Anspruch 17, wobei die Kurve ein Polynom ist.

19. Computerlesbares Medium nach Anspruch 18, wobei das Polynom eine quadratische Funktion ist.

20. Computerlesbares Medium nach Anspruch 12, wobei der Teilbereich möglicher Lag-Werte für den letzten Abtastwert für jeden Restsignal-Abschnitt so ausgewählt wird, dass die größte kumulative Verschiebung für jeden beliebigen Abtastwert in dem Abschnitt bei Anwendung der berechneten Warp-Kontur geringer ist als vier Abtastwertpositionen.

21. Computerlesbares Medium nach Anspruch 12, wobei das computerlesbare Medium ein magnetisch lesbares Plattenmedium umfasst.

22. Computerlesbares Medium nach Anspruch 12, wobei das computerlesbare Medium ein optisch lesbares Plattenmedium umfasst.

23. Computerlesbares Medium nach Anspruch 12, wobei das computerlesbare Medium ein moduliertes Datensignal umfasst.

24. Computerlesbares Medium nach Anspruch 12, wobei das computerlesbare Medium flüchtigen, computerlesbaren Speicher umfasst.

Es folgen 8 Blatt Zeichnungen

Anhängende Zeichnungen

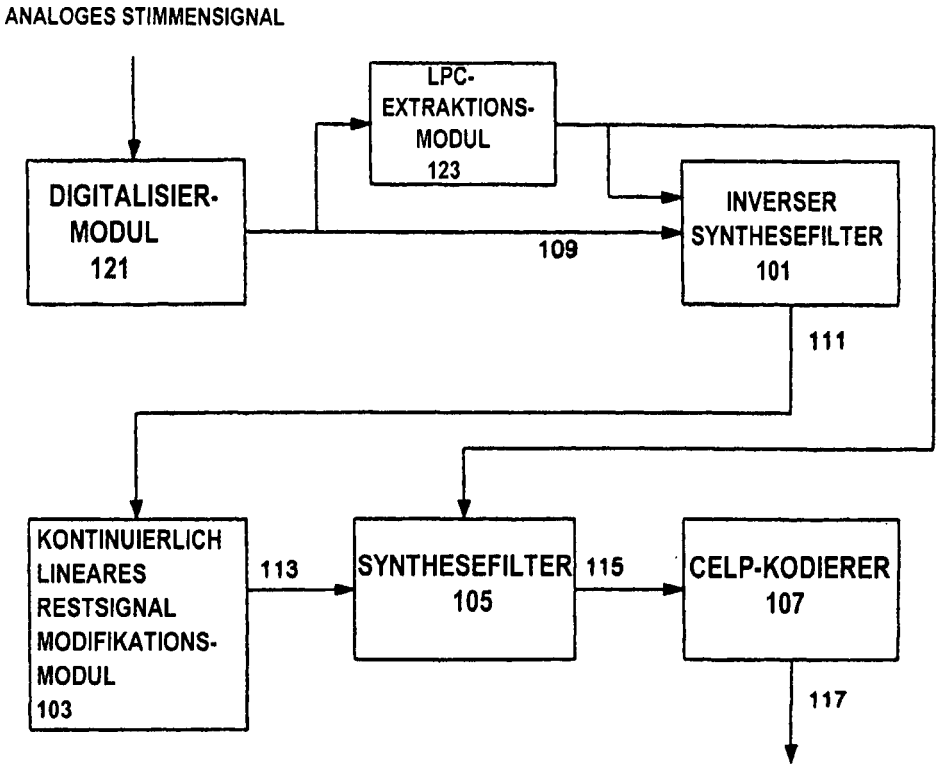

FIG. 1

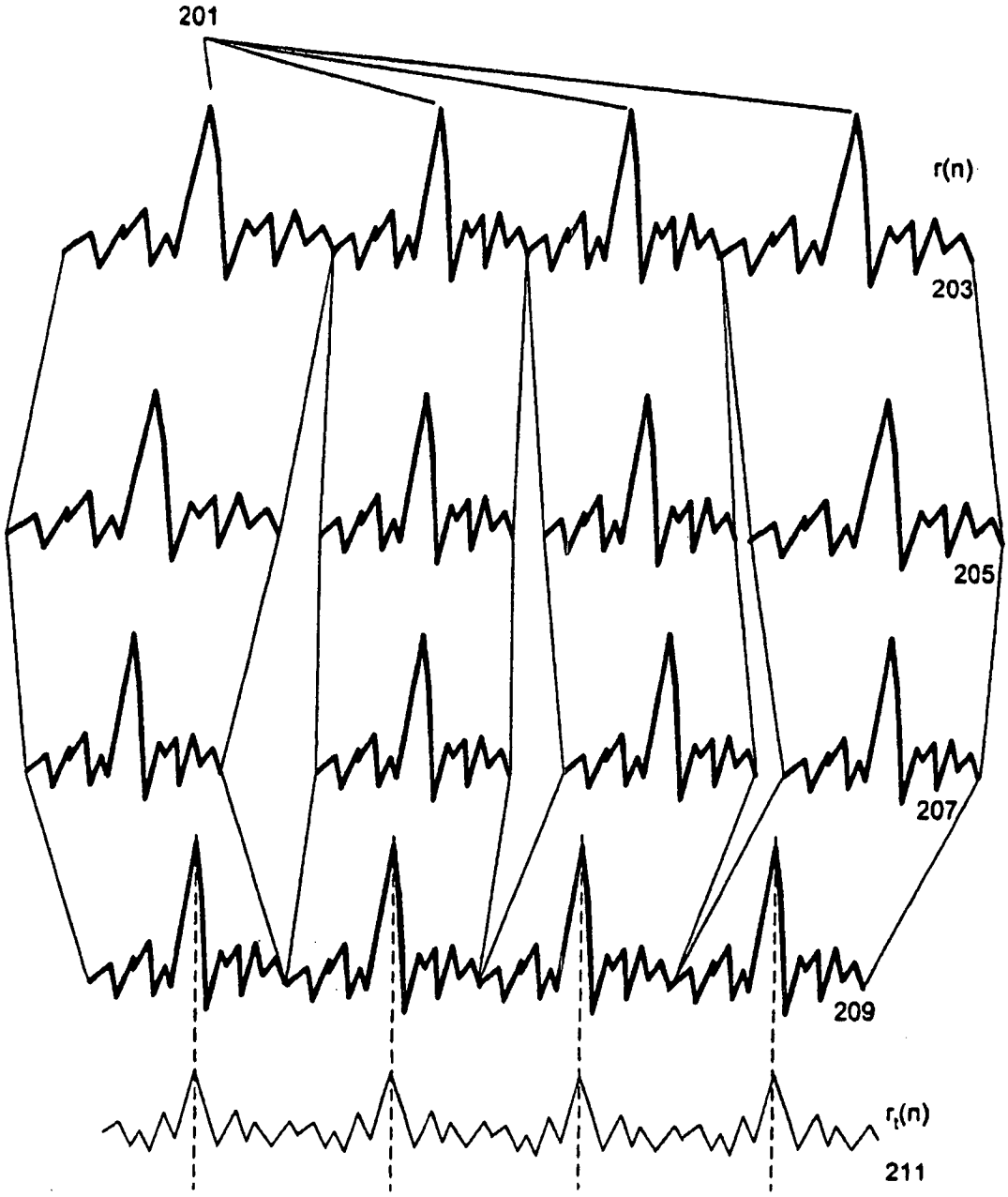

FIG.2

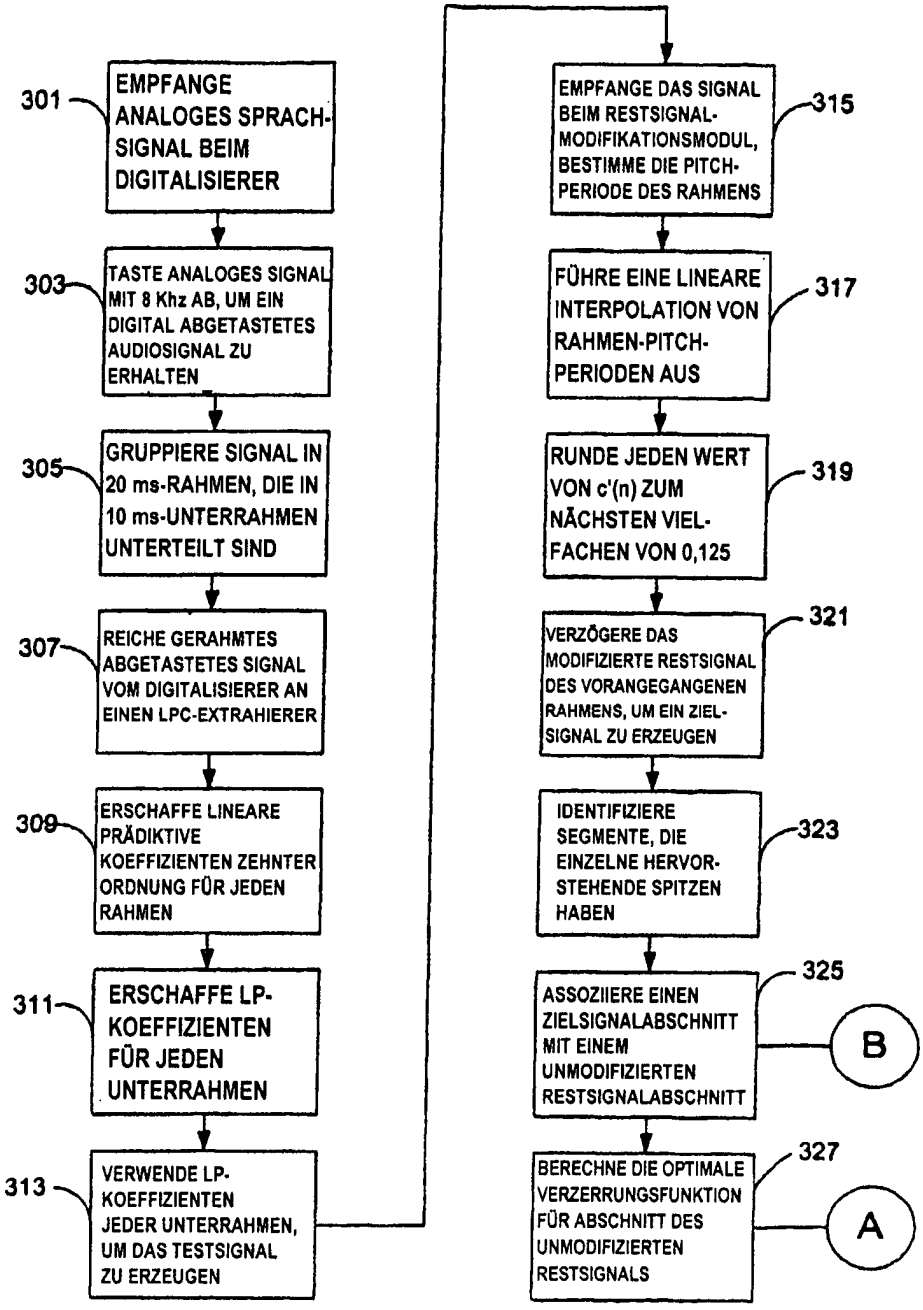
301
EMPFANGE ANALOGES SPRACH-SIGNAL BEIM DIGITALISIERER
303
TASTE ANALOGES SIGNAL MIT 8 Khz AB, UM EIN DIGITAL ABGETASTETES AUDIOSIGNAL ZU ERHALTEN
305
GRUPPIERE SIGNAL IN 20 ms-RAHMEN, DIE IN 10 ms-UNTERRAHMEN UNTERTEILT SIND
307
REICHE GERAHMTES ABGETASTETES SIGNAL VOM DIGITALISIERER AN EINEN LPC-EXTRAHIERER
309
ERSCHAFFE LINEARE PRÄDIKTIVE KOEFFIZIENTEN ZEHNTER ORDNUNG FÜR JEDEN RAHMEN
311
ERSCHAFFE LP-KOEFFIZIENTEN FÜR JEDEN UNTERRAHMEN
313
VERWENDE LP-KOEFFIZIENTEN JEDER UNTERRAHMEN, UM DAS TESTSIGNAL ZU ERZEUGEN
315
EMPFANGE DAS SIGNAL BEIM RESTSIGNAL-MODIFIKATIONSMODUL, BESTIMME DIE PITCH-PERIODE DES RAHMENS
317
FÜHRE EINE LINEARE INTERPOLATION VON RAHMEN-PITCH-PERIODEN AUS
319
RUNDE JEDEN WERT VON c'(n) ZUM NÄCHSTEN VIEL-FACHEN VON 0,125
321
VERZÖGERE DAS MODIFIZIERTE RESTSIGNAL DES VORANGEGANGENEN RAHMENS, UM EIN ZIEL-SIGNAL ZU ERZEUGEN
323
IDENTIFIZIERE SEGMENTE, DIE EINZELNE HERVOR-STEHENDE SPITZEN HABEN
325
ASSOZIIERE EINEN ZIELSIGNALABSCHNITT MIT EINEM UNMODIFIZIERTEN RESTSIGNALABSCHNITT
B
327
BERECHNE DIE OPTIMALE VERZERRUNGSFUNKTION FÜR ABSCHNITT DES UNMODIFIZIERTEN RESTSIGNALS
A

FIG. 3a

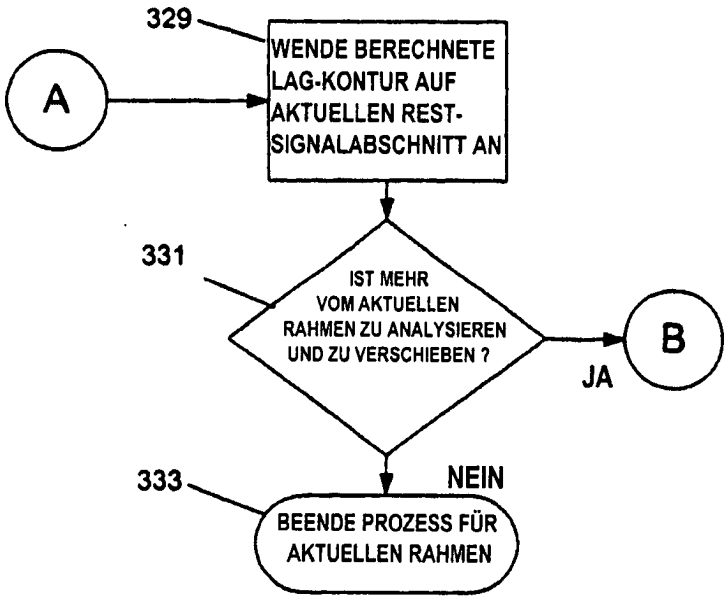
329
A
WENDE BERECHNETE LAG-KONTUR AUF AKTUELLEN REST-SIGNALABSCHNITT AN
331
IST MEHR VOM AKTUELLEN RAHMEN ZU ANALYSIEREN UND ZU VERSCHIEBEN ?
B
JA
NEIN
333
BEENDE PROZESS FÜR AKTUELLEN RAHMEN

FIG.3b

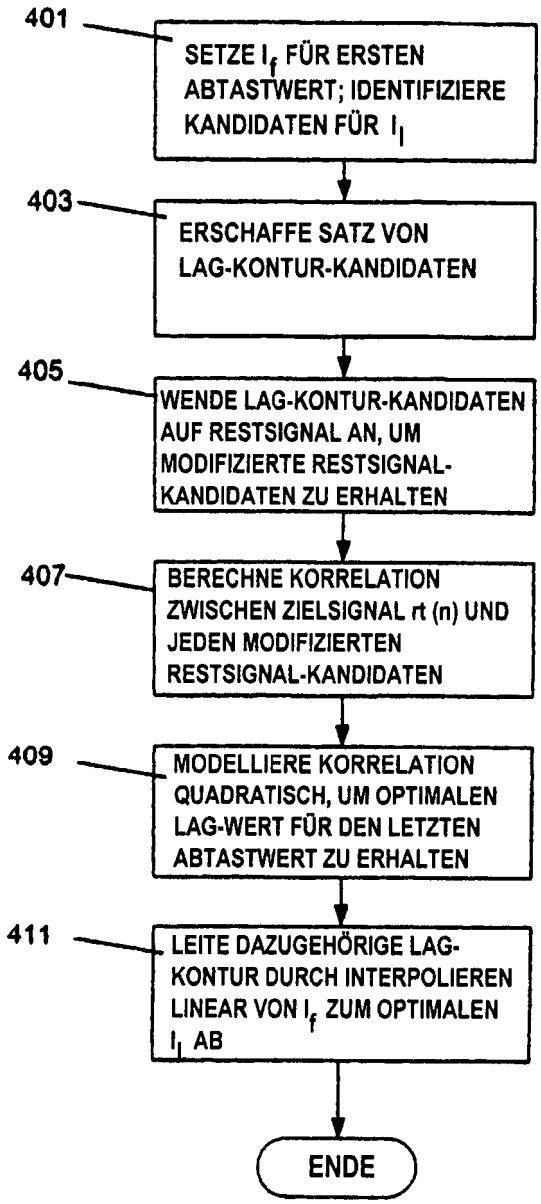
401
SETZE I_f FÜR ERSTEN ABTASTWERT; IDENTIFIZIERE KANDIDATEN FÜR I_l
403
ERSCHAFFE SATZ VON LAG-KONTUR-KANDIDATEN
405
WENDE LAG-KONTUR-KANDIDATEN AUF RESTSIGNAL AN, UM MODIFIZIERTE RESTSIGNAL-KANDIDATEN ZU ERHALTEN
407
BERECHNE KORRELATION ZWISCHEN ZIELSIGNAL rt (n) UND JEDEN MODIFIZIERTEN RESTSIGNAL-KANDIDATEN
409
MODELLIERE KORRELATION QUADRATISCH, UM OPTIMALEN LAG-WERT FÜR DEN LETZTEN ABTASTWERT ZU ERHALTEN
411
LEITE DAZUGEHÖRIGE LAG-KONTUR DURCH INTERPOLIEREN LINEAR VON I_f ZUM OPTIMALEN I_l AB
ENDE

FIG. 4

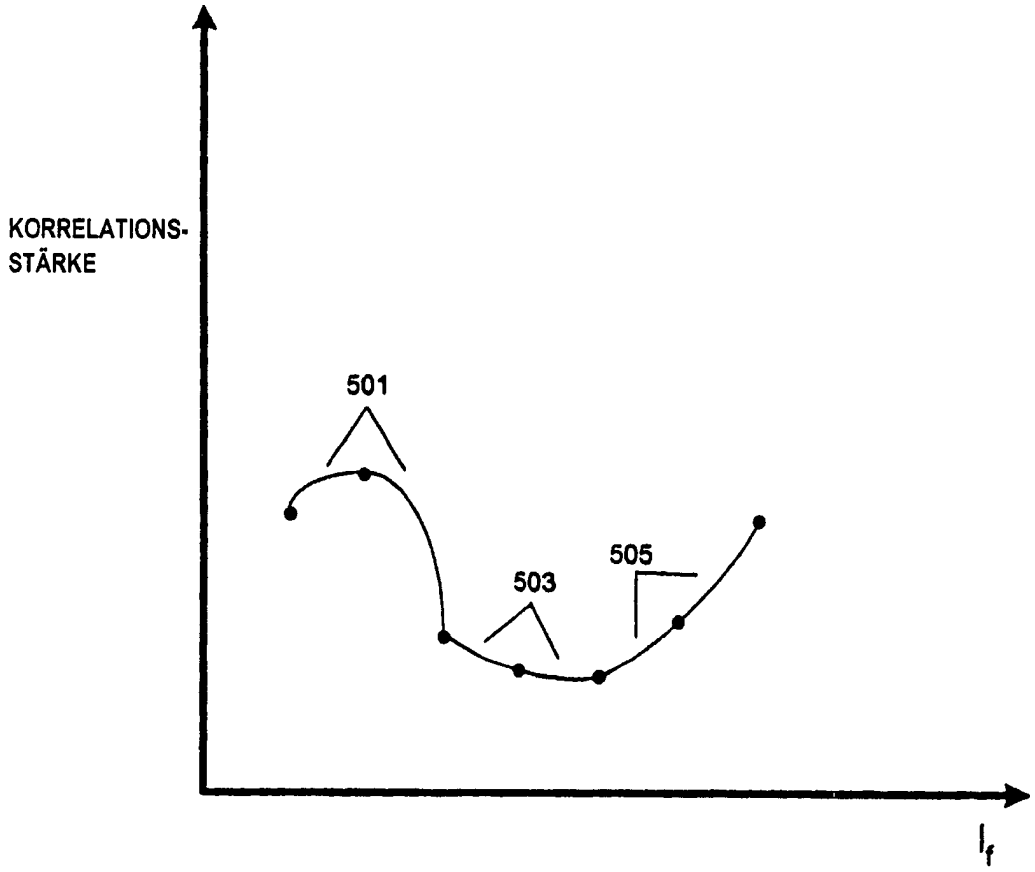

FIG. 5

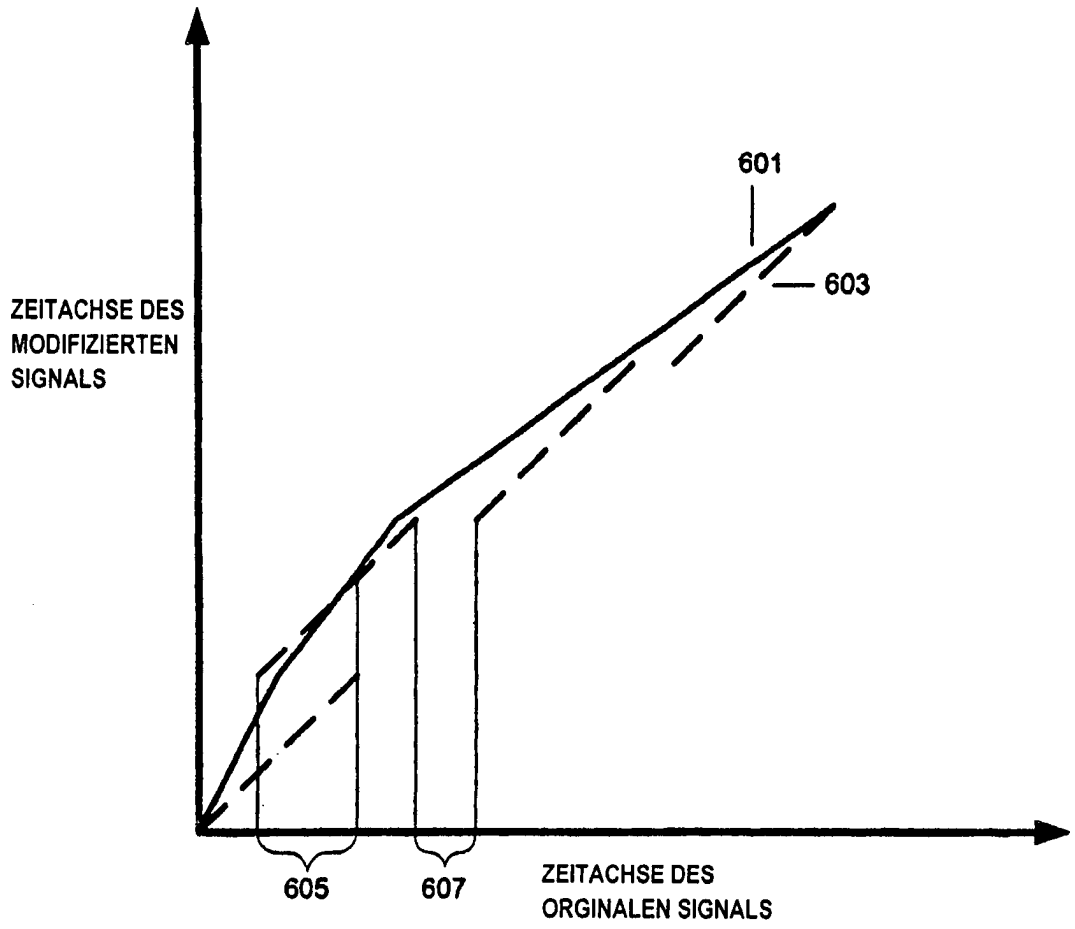
601
603
ZEITACHSE DES
MODIFIZIERTEN
SIGNALS
605
607
ZEITACHSE DES
ORGINALEN SIGNALS

FIG. 6

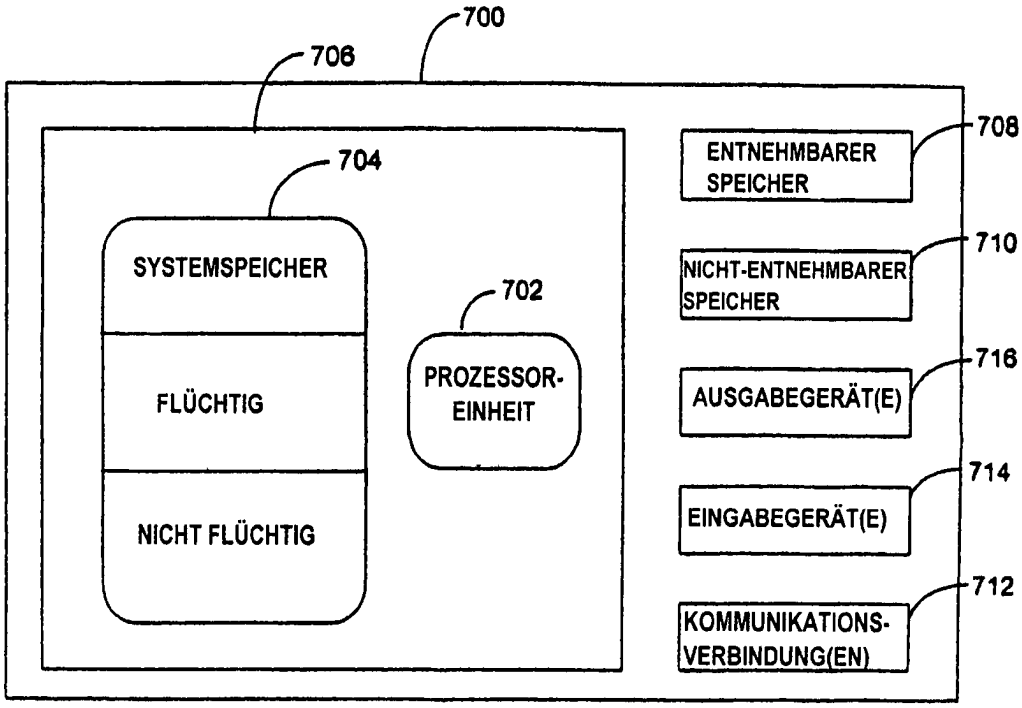
700
706
704
SYSTEMSPEICHER
FLÜCHTIG
NICHT FLÜCHTIG
702
PROZESSOR-EINHEIT
ENTNEHMBARER SPEICHER
708
NICHT-ENTNEHMBARER SPEICHER
710
AUSGABEGERÄT(E)
716
EINGABEGERÄT(E)
714
KOMMUNIKATIONS-VERBINDUNG(EN)
712

FIG. 7