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

(10) **DE 603 11 482 T2** 2007.10.25

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 1 543 503 B1**
(21) Deutsches Aktenzeichen: **603 11 482.2**
(86) PCT-Aktenzeichen: **PCT/IB03/03360**
(96) Europäisches Aktenzeichen: **03 797 392.2**
(87) PCT-Veröffentlichungs-Nr.: **WO 2004/027758**
(86) PCT-Anmeldetag: **05.08.2003**
(87) Veröffentlichungstag der PCT-Anmeldung: **01.04.2004**
(97) Erstveröffentlichung durch das EPA: **22.06.2005**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **24.01.2007**
(47) Veröffentlichungstag im Patentblatt: **25.10.2007**

(51) Int Cl.[8]: ***G10L 21/04*** (2006.01)
G10L 13/06 (2006.01)

(30) Unionspriorität:
02078847 17.09.2002 EP

(84) Benannte Vertragsstaaten:
AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LI, LU, MC, NL, PT, RO, SE, SI, SK, TR

(73) Patentinhaber:
Koninklijke Philips Electronics N.V., Eindhoven, NL

(74) Vertreter:
Volmer, G., Dipl.-Ing., Pat.-Anw., 52066 Aachen

(72) Erfinder:
GIGI, F., Ercan, NL-5656 AA Eindhoven, NL

(54) Bezeichnung: **VERFAHREN ZUR STEUERUNG DER DAUER BEI DER SPRACHSYNTHESE**

Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf das Gebiet der Sprachverarbeitung und im Besonderen ohne Einschränkung auf das Gebiet der Text/Sprache-Synthese.

[0002] Die Funktion eines Text/Sprache (TTS)-Synthesesystems besteht darin, Sprache von einem generischen Text in einer gegebenen Sprache zu synthetisieren. Heutzutage werden TTS-Systeme in vielen Anwendungsbereichen praktisch eingesetzt, beispielsweise für den Zugriff auf Datenbanken über das Telefonnetz oder als Hilfe für behinderte Personen. Ein Verfahren zum Synthetisieren von Sprache besteht darin, Elemente eines aufgezeichneten Satzes von Sprachteileinheiten wie Halbsilben oder Polyphone zu verketten. Die Mehrzahl erfolgreicher handelsüblicher Systeme verwendet die Verkettung von Polyphonen. Die Polyphone umfassen Gruppen von zwei (Diphone), drei (Triphone) oder mehr Phonen und können aus Unsinnwörtern ermittelt werden, indem die gewünschte Gruppierung von Phonen in stabilen spektralen Bereichen segmentiert wird. Bei einer Synthese auf der Basis der Verkettung ist die Erhaltung des Übergangs zwischen zwei benachbarten Phonen wesentlich für die Sicherstellung der Qualität der synthetisch erzeugten Sprache. Durch die Wahl der Polyphone als grundlegende Teileinheiten wird der Übergang zwischen zwei benachbarten Phonen in den aufgezeichneten Teileinheiten beibehalten, und die Verkettung erfolgt zwischen ähnlichen Phonen. Vor der Synthese muss jedoch die Dauer und die Tonhöhe der Phone verändert werden, damit die prosodischen Einschränkungen der neuen, derartige Phone enthaltenden Wörter erfüllt werden. Diese Verarbeitung ist erforderlich um zu vermeiden, dass die synthetisch erzeugte Sprache monoton klingt. In einem TTS-System wird diese Funktion durch ein prosodisches Modul ausgeführt. Damit die Dauer und die Tonhöhe in den aufgezeichneten Teileinheiten verändert werden kann, nutzen viele auf Verkettung basierende TTS-Systeme das TD-PSOLA-Synthesemodell (engl. time-domain pitch synchronous overlap-add, TD-PSOLA) (E. Moulines und F. Charpentier, „Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones", erschienen in Speech Commun., Band 9, S. 453–467, 1990). Bei dem TD-PSOLA-Modell wird das Sprachsignal zuerst einem die Tonhöhe kennzeichnenden Algorithmus unterzogen. Dieser Algorithmus ordnet den Spitzen des Signals in stimmhaften Segmenten und 10 ms entfernt in den stimmlosen Segmenten Marken zu. Die Synthese erfolgt durch Überlagerung von der Hanning-Fensterfunktion unterzogenen Segmenten, die an den Tonhöhenmarken zentriert sind und sich von der vorherigen Tonhöhenmarke bis zur nächsten erstrecken. Die Veränderung der Dauer erfolgt durch Löschen oder Replizieren einiger der gefensterten Segmente. Die Veränderung der Tonhöhenperiode erfolgt andererseits durch die Vergrößerung oder Reduzierung der Überlagerung zwischen den gefensterten Segmenten.

[0003] Trotz des in vielen handelsüblichen TTS-Systemen erzielten Erfolgs kann die unter Einsatz des TD-PSOLA-Synthesemodells erzeugte synthetische Sprache insbesondere bei starken prosodischen Schwankungen einige, im Folgenden dargelegte Nachteile aufweisen.

[0004] Beispiele für PSOLA-Verfahren sind in den Dokumenten EP-0363233, US-Patent Nr. 5.479.564 und EP-0706170 dargelegt. Ein spezielles Beispiel ist auch das MBR-PSOLA-Verfahren, wie es von T. Dutoit und H. Leich in Speech Communication, Elsevier Publisher, November 1993, veröffentlicht wurde. Das in der US-amerikanischen Patentschrift Nr. 5.479.564 beschriebene Verfahren schlägt Mittel vor zum Verändern der Frequenz eines Audiosignals mit konstanter Grundfrequenz durch die Überlappung und Addition von kurzzeitigen Signalen, die aus diesem Signal extrahiert werden. Die Breite der Gewichtungsfenster, die zur Erzielung der kurzzeitigen Signale eingesetzt werden, entspricht ungefähr der doppelten Periode des Audiosignals, und ihre Position innerhalb der Periode kann auf jeglichen Wert eingestellt werden (vorausgesetzt, dass die Zeitverschiebung zwischen aufeinander folgenden Fenstern der Periode des Audiosignals entspricht). In der US-amerikanischen Patentschrift Nr. 5.479.564 werden auch Mittel zum Interpolieren von Signalformen zwischen zu verkettenden Segmenten beschrieben, um Unstationärkeiten zu glätten. Derartige PSOLA-Verfahren ermöglichen es, die Dauer eines gegebenen Sprachsignals zu verändern. Dies kann durch Wiederholen oder Löschen von glockenförmigen Tonhöhenverläufen erfolgen, bevor ein Vorgang des Überlappens und Addierens für die Sprachsynthese durchgeführt wird. Die Informationen in einem glockenförmigen Tonhöhenverlauf sind nicht immer für eine Wiederholung geeignet, wie in einem Verschlusslaut. Ein geläufiger Nachteil der PSOLA-Verfahren nach dem Stand der Technik besteht darin, dass auf diese Weise Artefakte eingefügt werden. Diese Artefakte können zu einem metallischen Klang des synthetisch erzeugten Sprachsignals führen und sogar die Verständlichkeit des synthetisch erzeugten Signals erheblich beeinträchtigen oder verhindern.

[0005] In dem Dokument US-A-6.324.501 wird ein Verfahren zum Verändern eines eindimensionalen Eingangssignals dargelegt. Bei Sprachsignalen und ähnlichen eindimensionalen Signalen wird der Zeitmaßstab geändert, sie werden interpoliert und/oder falls erforderlich geglättet unter dem Einfluss eines Signals, das empfindlich für ein geringes stationäres Verhalten der Fenster des Signals ist, das verändert wird. Drei Maße für das stationäre Verhalten werden dargelegt: eines, das auf der Zeitbereichsanalyse ba-

siert, eines, das auf der Frequenzbereichsanalyse basiert, und eines, das sowohl auf der Zeit- als auch der Frequenzbereichsanalyse basiert.

[0006] In dem Dokument US-A-6.208.960 ist ein Verfahren zum Entfernen von Periodizität aus einem langen Audiosignal dargelegt. Ein Eingangsaudiosignal wird in eine Folge von sich überlappenden oder benachbarten Signalsegmenten unterteilt. Ein langes Signal wird synthetisiert, indem entsprechende Signalsegmente der Folge von Segmenten systematisch erhalten oder wiederholt werden. Durch die Wiederholung nicht periodischer Segmente, beispielsweise eines stimmlosen Teils eines Sprachsignals oder Rauschen in Musik, ergeben sich hörbare Artefakte. Die eingeführte Periodizität wird unterbrochen, indem eine Signalsektion, die von einem nicht periodischen Quellensignalsegment herrührt, in eine zweite Folge von Signalsegmenten unterteilt wird, wobei mindestens eines der Signalsegmente eine Dauer hat, die ungleich einer Dauer des Quellensignalsegmentes und ungleich einem Vielfachen der Dauer des Quellensignalsegmentes ist. Die Signalsegmente der zweiten Folge werden umgeordnet.

[0007] Der vorliegenden Erfindung liegt die Aufgabe zugrunde, ein verbessertes Verfahren zum Verarbeiten eines Sprachsignals zu schaffen. Die Erfindung ist durch die unabhängigen Ansprüche 1, 8 und 9 definiert. Abhängige Ansprüche beschreiben bevorzugte Ausführungsformen.

[0008] Die vorliegende Erfindung schafft ein Verfahren, ein Computerprogrammprodukt und ein Computersystem zum Verarbeiten eines Sprachsignals. Im Wesentlichen ermöglicht es die vorliegende Erfindung, ein natürlich klingendes synthetisiertes Sprachsignal mit verbesserter Verständlichkeit synthetisch zu erzeugen.

[0009] Diese Aufgabe wird gelöst durch das Klassifizieren gewisser in dem Originalsprachsignal enthaltener Intervalle. Gemäß einem bevorzugten Ausführungsbeispiel der Erfindung werden in dem Originalsprachsignal „stationäre" und „dynamische" Intervalle gekennzeichnet. Diese Klassifizierung braucht lediglich einmal durchgeführt zu werden.

[0010] Sie wird dazu verwendet, ein Sprachsignal basierend auf dem Originalsprachsignal mit einer geänderten Dauer zu synthetisieren.

[0011] Die vorliegende Erfindung basiert auf der Beobachtung, dass die Wiederholung von glockenverlaufsförmigen dynamischen Intervallen, wie es bei den PSOLA-Verfahren nach dem Stand der Technik erfolgt, eine unbeabsichtigte Periodizität einführt, die zu Artefakten, wie beispielsweise einem metallisch klingenden synthetisierten Signal, und dazu führt, dass es weniger oder gar nicht verständlich ist.

[0012] Gemäß der vorliegenden Erfindung wird dieses Problem gelöst, indem die Verarbeitung von Glockenverläufen zum Zweck der Änderung der Dauer auf Glockenverläufe von stationären Intervallen des Originalsprachsignals beschränkt wird. Mit anderen Worten: Änderungen der Dauer werden nur an denjenigen Sprachintervallen vorgenommen, die eine unterschiedliche Dauer haben können. Dies gilt für die Mitte eines Vokals oder eines Konsonanten wie der Laut /s/. Es gibt jedoch Fälle, bei denen lokale Ereignisse auftreten, die kürzer als eine einzige Periode dauern. Dies sind plötzliche Veränderungen, wie beispielsweise der Beginn eines stimmlosen Verschlusslautes (/p/, /t/, /k/) oder die durch Zunge und Mund erzeugten Tick- und Schnalzlaute (/b/, /d/, /g/, /l/, /m/, /n, usw.). Perioden, die diese Ereignisse enthalten, sind wichtig für die Verständlichkeit und sollten bei der Bearbeitung nicht weggelassen werden. Ihre Wiederholung stellt auch ein Problem dar, da dadurch Artefakte eingefügt werden, die unnatürlich klingen. Auch die Perioden am Anfang eines Übergangs von einem stimmlosen Laut zu einem Vokal haben lokale Merkmale, die nicht verlängert oder verkürzt werden sollten. Zur Verhinderung von Artefakten werden alle Perioden mit einer speziellen Information zur Periodenklassenart gekennzeichnet. Diese Informationen werden dazu verwendet zu ermitteln, ob eine Periode wiederholt oder weggelassen werden kann. Somit werden Glockenverläufe, die durch Fensterung von dynamischen Intervallen des Originalsprachsignals erhalten werden, zur Änderung der Dauer nicht wiederholt. Glockenverläufe, die von Intervallen erzielt werden, die als dynamisch und wesentlich für die Verständlichkeit klassifiziert werden, werden in dem synthetisierten Signal beibehalten, um die Verständlichkeit aufrechtzuerhalten. Glockenverläufe, die durch Fensterung von Intervallen des Originalsprachsignals erhalten werden, die als dynamisch aber nicht wesentlich für die Verständlichkeit klassifiziert werden, können gelöscht werden oder nicht, bevor der Vorgang des Überlappens und Addierens durchgeführt wird, ohne dass die Qualität des resultierenden synthetisierten Sprachsignals erheblich beeinträchtigt wird.

[0013] Die vorliegende Erfindung findet bevorzugt Anwendung in Text/Sprache-Systemen, die eine große Anzahl von natürlichen Sprachaufzeichnungen speichern, die im Prozess der Text/Sprache-Synthese verändert werden.

[0014] Gemäß einer bevorzugten Ausführungsform der Erfindung wird ein angehobenes Kosinusfenster für die Fensterung des Sprachsignals eingesetzt. Vorzugsweise wird ein Sinusfenster für stationäre Intervalle eingesetzt, die stimmlose Sprache enthalten. Die für derartige stationäre Intervalle mit stimmloser Sprache erhaltenen Glockenverläufe werden randomisiert, um jegliche unbeabsichtigte Periodizität zu entfernen, die in dem Prozess der Änderung der Dau-

er eingefügt werden kann.

[0015] Bevorzugte Ausführungsbeispiele der Erfindung sind in den Zeichnungen dargestellt und werden im Folgenden näher beschrieben. Es zeigen:

[0016] Fig. 1 einen Ablaufplan eines bevorzugten Ausführungsbeispiels der vorliegenden Erfindung;

[0017] Fig. 2 die Synthese eines Sprachsignals basierend auf einem Originalsprachsignal gemäß einem Ausführungsbeispiel der vorliegenden Erfindung;

[0018] Fig. 3 ein Blockschaltbild eines Ausführungsbeispiels eines erfindungsgemäßen Computersystems.

[0019] Fig. 1 zeigt einen Ablaufplan zur Erläuterung eines bevorzugten Ausführungsbeispiels des erfindungsgemäßen Verfahrens. In Schritt **100** wird eine Aufzeichnung natürlicher Sprache geschaffen. In Schritt **102** werden Intervalle in der Aufzeichnung der natürlichen Sprache gekennzeichnet und klassifiziert. Für die Klassifizierung der Sprachintervalle wird in dem hier betrachteten Beispiel das folgende Klassifizierungssystem verwendet:

– Pause
. stimmlose Periode
v stimmhafte Periode
p wesentliche dynamische stimmlose Periode (sollte nur einmal verwendet werden)
b wesentliche dynamische stimmhafte Periode (sollte nur einma verwendet werden)
q dynamische stimmlose Periode (darf nur einmal verwendet werden)
c dynamische stimmhafte Periode (darf nur einmal verwendet werden).

[0020] Die beiden grundlegenden Kategorien von Sprachintervallen sind „stationäre" und „dynamische" Sprachintervalle. Ein Sprachintervall wird als „stationär" klassifiziert, wenn es eine im Wesentlichen konstante Signalkennlinie für eine aufeinander folgende Anzahl von mindestens zwei Perioden der Grundfrequenz des natürlichen Sprachsignals aufweist. Im Gegensatz dazu wird das Sprachintervall der Originalsprachaufzeichnung als „dynamisch" klassifiziert, wenn seine Signalkennlinie nur innerhalb einer Periode der Grundfrequenz auftritt.

[0021] In dem hier betrachteten Klassifizierungssystem sind die Perioden '.' und 'v' stationäre Perioden. Die Perioden 'p', 'b', 'q' und 'c' sind dynamische Perioden, die in der nachfolgenden Verarbeitung anders behandelt werden.

[0022] In Schritt **104** wird ein natürliches Sprachsignal gefenstert, um Glockenverläufe zu erzielen. Die Fensterung wird vorzugsweise mit Hilfe eines angehobenen Kosinusfensters oder mit einem Sinusfenster für die Perioden '.' durchgeführt.

[0023] In Schritt **106** werden die aus Perioden, die als 'stationär' klassifiziert werden, erhaltenen Glockenverläufe verarbeitet, um die Dauer des Sprachsignals zu verändern. Dies kann durch Wiederholen oder Löschen von Glockenverläufen erfolgen, um die ursprüngliche Dauer zu verlängern bzw. zu verkürzen. Aus Perioden, die als 'dynamisch' klassifiziert werden, erhaltene Glockenverläufe werden nicht wiederholt, um das Einfügen von Artefakten zu verhindern. Aus Perioden, die als 'p' oder 'b' klassifiziert werden, erhaltene Glockenverläufe können nicht gelöscht werden, damit die Verständlichkeit der Originalsignals erhalten bleibt. Aus Perioden, die als 'q' oder 'c' klassifiziert werden, erhaltene Glockenverläufe werden ebenfalls nicht wiederholt, können jedoch gelöscht werden, ohne dass die Verständlichkeit des resultierenden synthetischen Signals wesentlich beeinflusst wird.

[0024] Vorzugsweise werden Glockenverläufe aus Perioden, die als '.' klassifiziert werden, randomisiert, um die Einführung von Periodizität zu verhindern. Dies wird außerdem unterstützt durch den Einsatz eines Sinusfensters für die Fensterung derartiger Perioden.

[0025] In Schritt **108** werden die verarbeiteten Glockenverläufe zur Erzeugung des synthetischen Signals überlappt und addiert.

[0026] Fig. 2 zeigt ein Beispiel für die Verarbeitung eines natürlichen Sprachsignals **200**. Das natürliche Sprachsignal **200** weist dynamische Intervalle **202**, **204**, **206**, **208**, **210** und **212** auf. Das dynamische Intervall **202** enthält Perioden, die als 'b', 'c' klassifiziert werden. Das dynamische Intervall **204** enthält Perioden, die als 'c', 'q' klassifiziert werden. Das dynamische Intervall **206** enthält Perioden, die als 'q' klassifiziert werden. Das dynamische Intervall **208** enthält Perioden, die als 'q', 'c' und 'b' klassifiziert werden. Das dynamische Intervall **210** enthält Perioden, die als 'c', 'b' klassifiziert werden. Schließlich enthält das dynamische Intervall **212** Perioden, die als 'c' und 'b' klassifiziert werden. Ferner enthält das natürliche Sprachsignal **200** stationäre Intervalle **214**, **216**, **218**, **220**, **222** und **224**. Das stationäre Intervall **214** enthält Perioden, die als 'v' klassifiziert werden, das stationäre Intervall **216** enthält Perioden, die als '.' klassifiziert werden, das stationäre Intervall **218** enthält Perioden, die als '.' klassifiziert werden, das stationäre Intervall **220** enthält Perioden, die als 'v' klassifiziert werden, das stationäre Intervall **222** enthält Perioden, die als 'v' klassifiziert werden, und das stationäre Intervall **224** enthält Perioden, die als 'v' klassifiziert werden. Diese Klassifizierung kann entweder manuell oder automatisch mittels eines geeigneten Signal-

analyseprogramms durchgeführt werden. Vorzugsweise wird eine automatische Analyse mit Hilfe eines derartigen Programms durchgeführt, das dann von einem Fachmann gesteuert und, falls erforderlich, manuell korrigiert wird. Es ist anzumerken, dass diese Klassifizierung nur einmal durchgeführt zu werden braucht, um eine unbegrenzte Anzahl von Signalsynthesen zu ermöglichen.

[0027] Bei dem hier betrachteten Beispiel ist ein Signal auf der Grundlage des natürlichen Sprachsignals **200** zu synthetisieren, das eine längere Dauer im Vergleich zu dem Originalsprachsignal **200** aufweist. Zu diesem Zweck wird das natürliche Sprachsignal **200** mit Hilfe eines Fensters gefenstert, das synchron zur Grundfrequenz des natürlichen Sprachsignals **200** positioniert wird, wie es nach dem Stand der Technik bekannt ist und in Verfahren des PSOLA-Typs eingesetzt wird.

[0028] Als Fenster wird vorzugsweise ein angehobenes Kosinusfenster eingesetzt. Für Perioden, die als '.' klassifiziert werden, wird ein Sinusfenster eingesetzt, um eine unbeabsichtigte Periodizität zu reduzieren, die eventuell eingeführt wird, wenn Glockenverläufe des verrauschten Signalanteils wiederholt werden. Als weitere Maßnahme gegen eine unbeabsichtigte Periodizität werden die Glockenverläufe für die als '.' klassifizierten Perioden randomisiert erfasst. Bei dem hier betrachteten Beispiel wird das zu synthetisierende Signal folgendermaßen im Bereich der Zeitachse **226** zusammengesetzt:
Das erste Intervall **228** des zu synthetisierenden Sprachsignals enthält die Glockenverläufe von dem dynamischen Intervall **202**. Diese Glockenverläufe werden für das Intervall **228** ohne Veränderung verwendet, was impliziert, dass die Dauer des Intervalls **228** in Hinblick auf das dynamische Intervall **202** unverändert ist. Die Dauer des Intervalls **230** ist ungefähr das Doppelte der Dauer des entsprechenden stationären Intervalls **214**. Dies wird durch Wiederholen jedes der für das stationäre Intervall **214** erfassten Glockenverläufe erreicht. Das Intervall **232** enthält die Glockenverläufe von dem dynamischen Intervall **204**. Die Dauer von **232** ist unverändert im Vergleich zu dem dynamischen Intervall **204**. Das Intervall **234** besteht aus Glockenverläufen, die von dem stationären Intervall **216** erfasst wurden. Wiederum wird jeder der in dem stationären Intervall **216** enthaltenen Glockenverläufe wiederholt, um die Dauer dieses Intervalls zu verdoppeln. In gleicher Weise werden die folgenden Intervalle **236**, **238**, **240**, **242**,... aus den Intervallen **206**, **218**, **208**, **220**, **210**, **222**, **212**, **242** erzielt. Danach werden die Glockenverläufe im Bereich der Zeitachse **226** überlappt, um das resultierende synthetisierte Signal zu erhalten. Als Alternative können die aus den als 'q' oder 'c' klassifizierten Perioden des natürlichen Sprachsignals **200** erzielten Glockenverläufe gelöscht werden. Auf keinen Fall werden die Glockenverläufe, die aus als 'dynamisch' klassifizierten Perioden des natürlichen Sprachsignals **200** erzielt wurden, wiederholt. Auf diese Weise kann eine Änderung der Dauer durchgeführt werden, ohne dass Artefakte eingefügt werden, die sonst einen erheblichen Einfluss auf die Qualität und Verständlichkeit des synthetisierten Signals hätten. Bei dem hier betrachteten Beispiel wird 'p' verwendet, um lokale (stimmlose) Ereignisse zu markieren, die wesentlich für die Verständlichkeit der gesprochenen Äußerung sind. Normalerweise gehört der Rauschburst nach dem Ablassen von Luft durch den Mund oder die Zunge zu diesem Typ. Die Phoneme /p/, /t/ und /k/ weisen mindestens eine derartige Periode auf. Mit 'p' markierte Perioden sollten unabhängig von der endgültigen Dauer der Phoneme nur einmal in der synthetischen Sprache auftauchen. Einige lokale (stimmlose) Ereignisse sind für die Verständlichkeit nicht wesentlich, jedoch so dynamisch, dass ihre Wiederholung eine Folge von unnatürlich klingenden Perioden einfügen würde. Diese Perioden werden mit dem Buchstaben 'q' markiert. Sie dürfen nur einmal verwendet werden, können jedoch auch weggelassen werden, ohne dass eine wesentliche Verschlechterung der Qualität oder der Verständlichkeit die Folge wäre. Die stimmhaften Gegenstücke zu 'p' und 'q' sind die mit 'b' und 'c' gekennzeichneten Arten. Die stimmhaften Verschlusslaute /b/, /d/ und /g/ weisen normalerweise mindestens eine mit 'b' markierte Periode auf. Auch die Zunge kann Tick- und Schnalzlaute erzeugen, wenn sie andere Teile des Mundes trifft oder sich von ihnen löst. Das Phonem /l/ ist ein Beispiel, bei dem dies auftritt. Der Übergang von einer Pause zu Vokalen oder von stimmlosen Konsonanten zu Vokalen kann ebenfalls Perioden mit lokalen Ereignissen aufweisen. Die Perioden in der Mitte eines Vokals können zwar viele Male wiederholt werden, ohne dass die Natürlichkeit beeinträchtigt wird, die Perioden, die genau in die Mitte des Übergangs fallen, sind jedoch zu dynamisch für eine Wiederholung.

[0029] Fig. 3 zeigt ein Blockschaltbild eines Ausführungsbeispiels eines erfindungsgemäßen Computersystems. Das Computersystem ist vorzugsweise ein Text/Sprache-System, das die Prinzipien der vorliegenden Erfindung verkörpert. Das Computersystem **300** umfasst ein Modul **302**, das zum Speichern natürlicher Sprachsignale dient. Das Modul **304** dient dazu, automatisch, manuell oder interaktiv Perioden der in dem Modul **302** gespeicherten natürlichen Sprachsignale zu klassifizieren. Das Modul **306** dient dazu, die Fensterung eines in dem Modul **302** gespeicherten natürlichen Sprachsignals durchzuführen. Auf diese Weise wird eine Anzahl von Glockenverläufen erzielt. Das Modul **308** dient zur Verarbeitung der Glockenverläufe. Die Verarbeitung von Glockenverläufen zur Änderung der Dauer wird nur an Glockenverläufen vorgenommen, die aus Intervallen erzielt werden, die als stationär klassifiziert werden. Zusätzlich können Glockenverläufe aus dynami-

schen Intervallen, die als nicht wesentlich für die Verständlichkeit klassifiziert wurden, durch das Modul **308** gelöscht werden, so dass sie in dem synthetisierten Signal nicht auftreten. Das Modul **310** dient dazu, einen Vorgang des Überlappens und Addierens an den resultierenden Glockenverläufen vorzunehmen, um das synthetische Signal zu erzeugen. Die gewünschte Änderung der Dauer des im Modul **302** gespeicherten natürlichen Originalsprachsignals wird in das Computersystem **300** eingegeben. Das resultierende synthetische Signal wird vom Computersystem **300** auf einer Trägerwelle oder als Datendatei ausgegeben.

Text in den Figuren

Figur 3

Modification of duration	Änderung der Dauer
Synthesized signal	Synthetisiertes Signal

Bezugszeichenliste

100	Aufzeichnung natürlicher Sprache schaffen
102	Intervall klassifizieren
104	Tonhöhenperioden ermitteln
106	Dauer der stationären Tonhöhenperioden verändern
108	für Synthese überlappen und addieren
200	natürliches Sprachsignal
202	dynamisches Intervall
204	dynamisches Intervall
206	dynamisches Intervall
208	dynamisches Intervall
210	dynamisches Intervall
212	dynamisches Intervall
214	stationäres Intervall
216	stationäres Intervall
218	stationäres Intervall
220	stationäres Intervall
222	stationäres Intervall
224	stationäres Intervall
226	Zeitachsenintervall
230	Intervall
232	Intervall
234	Intervall
236	Intervall
238	Intervall
240	Intervall
242	Intervall
300	Computersystem
302	Modul
304	Modul
306	Modul
308	Modul
310	Modul

Patentansprüche

1. Verfahren zum Synthetisieren eines Sprachsignals, das Folgendes umfasst:
– Zuordnen eines ersten Identifikators zu stationären Intervallen eines Originalsprachsignals,
– Zuordnen eines zweiten Identifikators zu dynamischen Intervallen des Originalsprachsignals,
– Kennzeichnen dynamischer stimmloser Perioden (q) und dynamischer stimmhafter Perioden (c),
– Fenstern des Originalsprachsignals zum Erzeugen einer Anzahl von Tonhöhenperioden, gekennzeichnet durch
– Löschen der Tonhöhenperioden, die dynamischen stimmlosen Perioden (q) und dynamischen stimmhaften Perioden (c) entsprechen,
– Verarbeiten der Tonhöhenperioden mit dem ihnen zugeordneten ersten Identifikator zum Verändern einer Dauer des Sprachsignals,
– Durchführen eines Vorgangs des Überlappens und Addierens an den verarbeiteten Tonhöhenperioden.

2. Verfahren nach Anspruch 1, wobei ein erster Code oder ein zweiter Code als erster Identifikator verwendet wird, wobei der erste Code eine stimmlose Periode und der zweite Code eine stimmhafte Periode kennzeichnet.

3. Verfahren nach einem der vorherigen Ansprüche, wobei ein dritter Code, ein vierter Code, ein fünfter Code oder ein sechster Code als zweiter Identifikator verwendet wird, wobei der dritte Code eine stimmlose Periode kennzeichnet, die wesentlich für die Verständlichkeit des Sprachsignals ist, der vierte Code eine stimmhafte Periode kennzeichnet, die wesentlich für die Verständlichkeit des Sprachsignals ist, und der fünfte Code eine stimmlose Periode kennzeichnet, die nicht wesentlich für die Verständlichkeit des Sprachsignals ist, und der sechste Code eine stimmhafte Periode kennzeichnet, die nicht wesentlich für die Verständlichkeit des Sprachsignals ist.

4. Verfahren nach einem der vorhergehenden Ansprüche, wobei eine angehobene Kosinusfunktion für die Fensterung des Sprachsignals verwendet wird.

5. Verfahren nach einem der vorhergehenden Ansprüche, wobei ein Sinusfenster für die Fensterung stationärer stimmloser Intervalle des Sprachsignals verwendet wird.

6. Verfahren nach einem der vorhergehenden Ansprüche, das ferner das Randomisieren der Tonhöhenperioden von stationären, stimmlosen Perioden umfasst, bevor der Vorgang des Überlappens und Addierens durchgeführt wird.

7. Verfahren nach einem der vorhergehenden

Ansprüche, wobei die Fensterung mit Hilfe eines Fensters durchgeführt wird, das synchron mit einer Grundfrequenz des Sprachsignals positioniert wird.

8. Computerprogrammprodukt, das Programmcodemittel umfasst, die einen Computer veranlassen, alle Schritte des Verfahrens nach Anspruch 1 auszuführen, wenn das genannte Programm auf einem Computer läuft.

9. Computersystem, im Besonderen Text/Sprache-System, das Folgendes umfasst:
– Mittel (**2302**) zum Speichern eines Sprachsignals,
– Mittel (**304**) zum Speichern erster Identifikatoren, die stationären Intervallen eines Originalsprachsignals zugeordnet sind, und zum Speichern zweiter Identifikatoren, die dynamischen Intervallen des Originalsprachsignals zugeordnet sind,
– Mittel zum Kennzeichnen dynamischer stimmloser Perioden (q) und dynamischer stimmhafter Perioden (c),
– Mittel (**306**) zum Fenstern des Sprachsignals zum Erzeugen einer Anzahl von Tonhöhenperioden, dadurch gekennzeichnet, das sie Folgendes umfassen:
– Mittel zum Löschen der Tonhöhenperioden, die dynamischen stimmlosen Perioden (q) und dynamischen stimmhaften Perioden (c) entsprechen,
– Mittel (**308**) zum Verarbeiten der Tonhöhenperioden mit dem ihnen zugeordneten ersten Identifikator, um die Dauer des Sprachsignals zu verändern, und
– Mittel (**310**) zum Durchführen eines Vorgangs des Überlappens und Addierens an den verarbeiteten Tonhöhenperioden.

Es folgen 3 Blatt Zeichnungen

Anhängende Zeichnungen

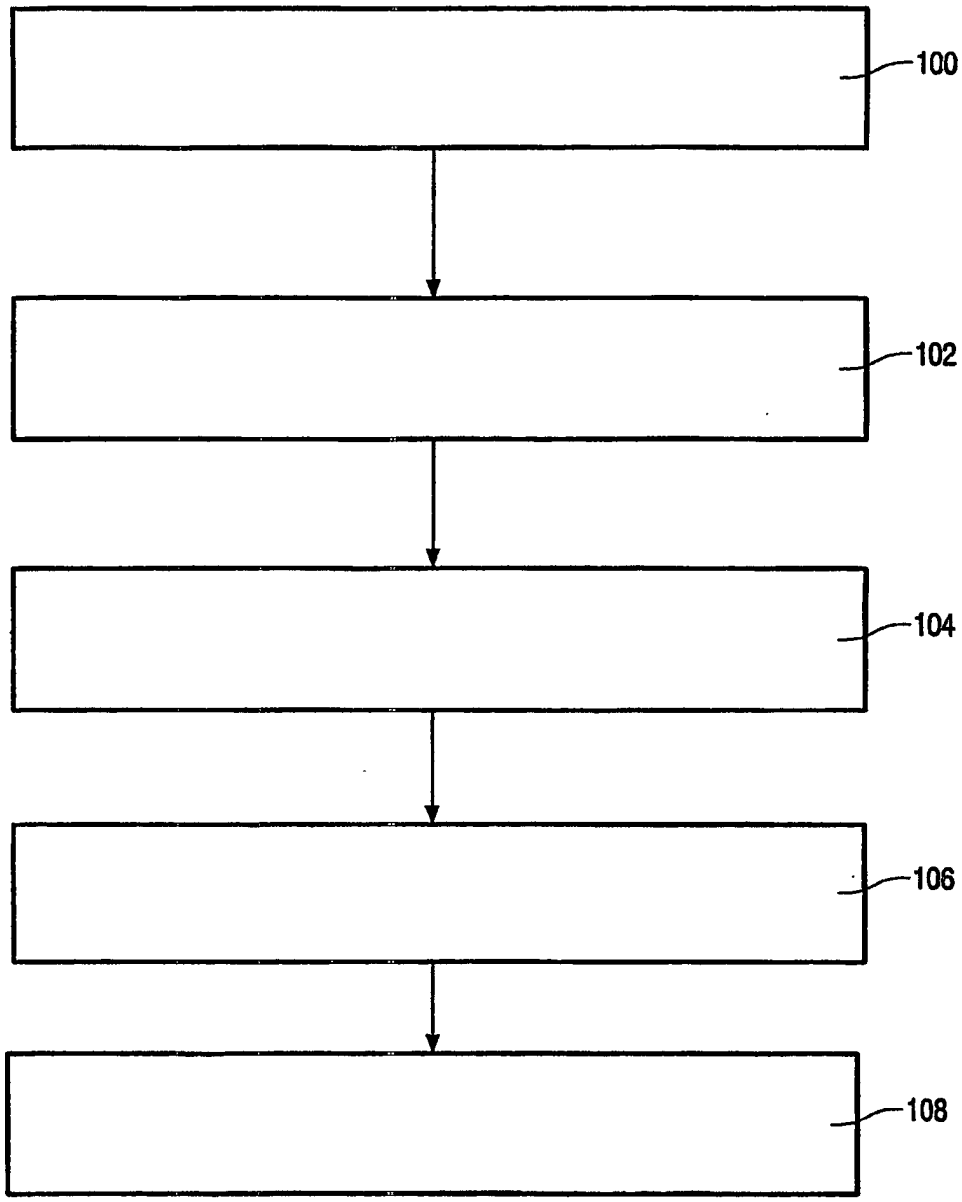

FIG. 1

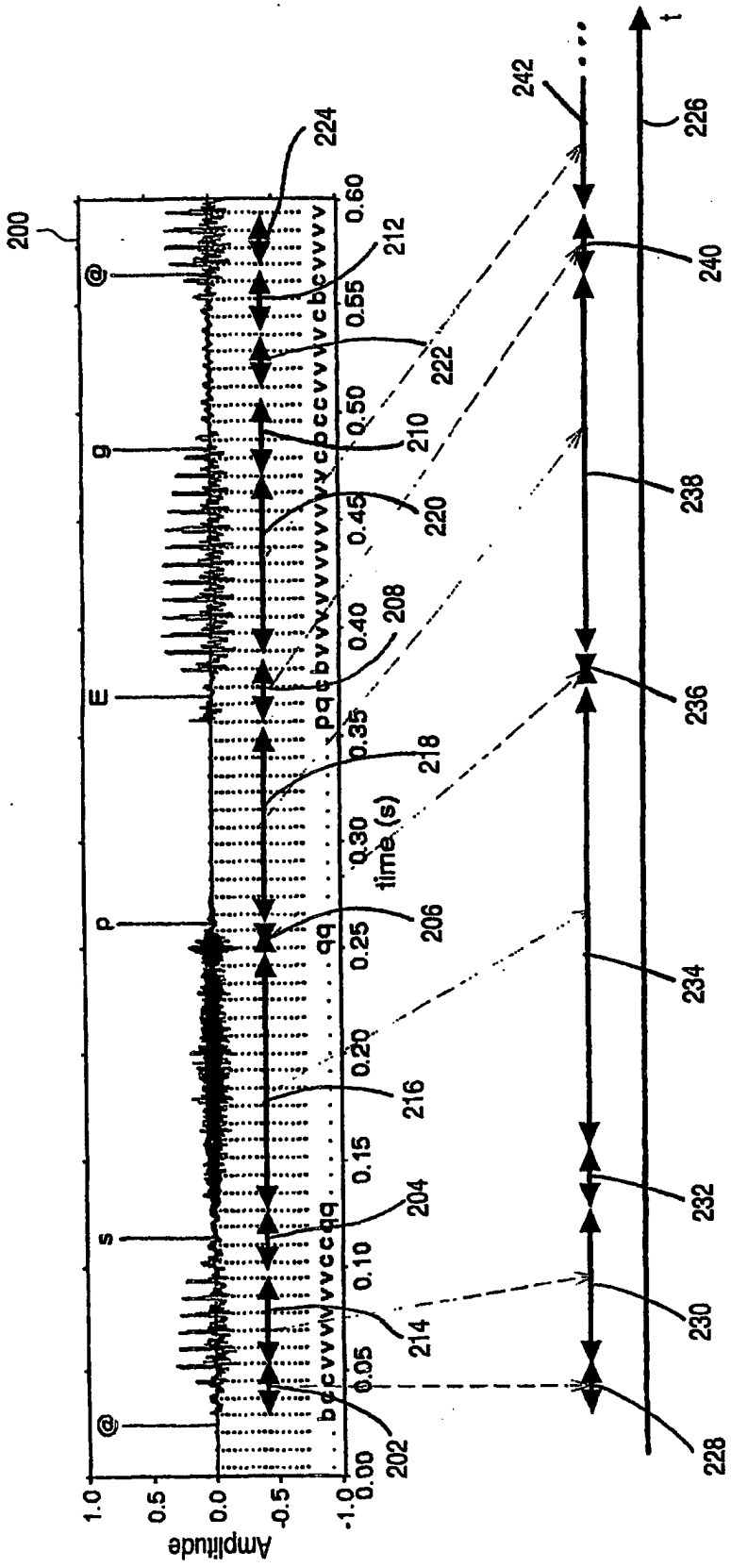
200
224
212
222
210
220
208
218
206
216
204
214
202
242
240
238
236
234
232
230
228
226
t
Amplitude
time (s)
1.0
0.5
0.0
-0.5
-1.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60

FIG. 2

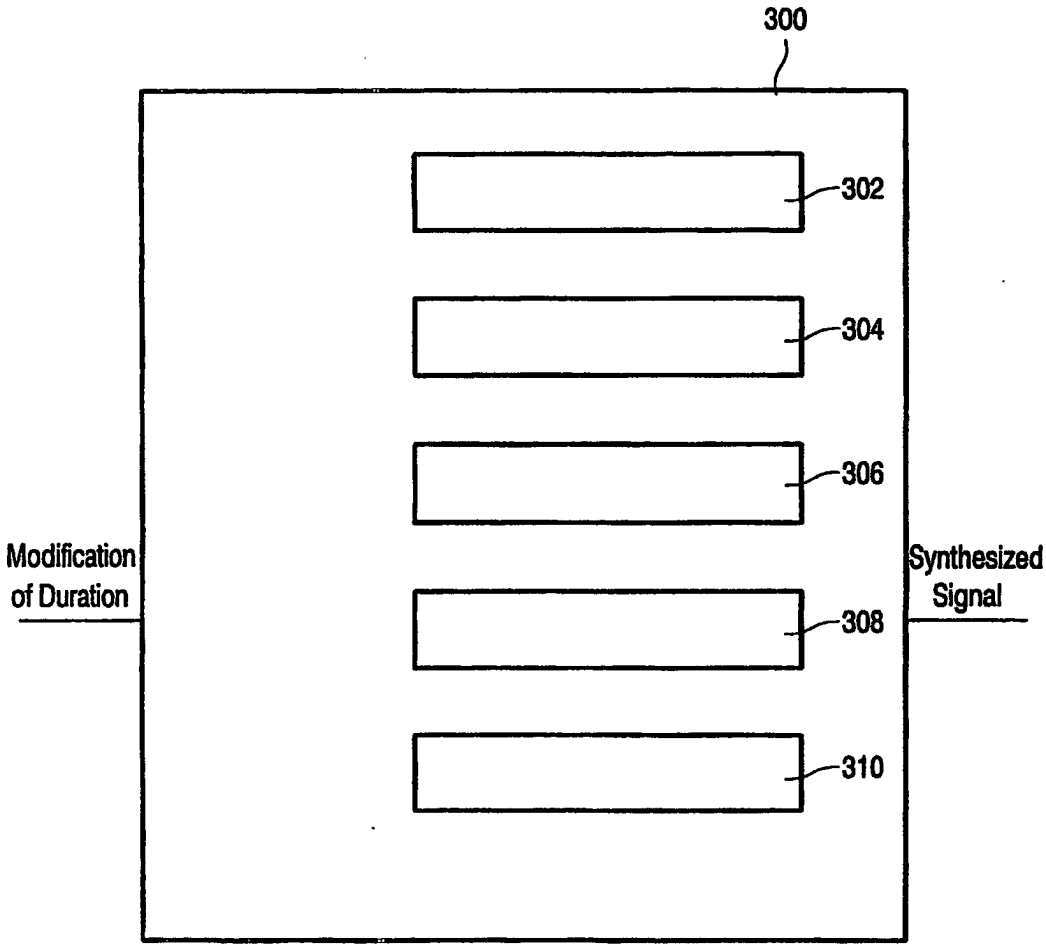
300
302
304
306
308
310
Modification of Duration
Synthesized Signal

FIG. 3