



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) **DE 600 23 237 T2** 2006.07.13

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 088 301 B1**

(51) Int Cl.<sup>8</sup>: **G10L 19/00** (2006.01)

(21) Deutsches Aktenzeichen: **600 23 237.9**

(86) PCT-Aktenzeichen: **PCT/US00/10576**

(96) Europäisches Aktenzeichen: **00 926 156.1**

(87) PCT-Veröffentlichungs-Nr.: **WO 2000/063883**

(86) PCT-Anmeldetag: **19.04.2000**

(87) Veröffentlichungstag  
der PCT-Anmeldung: **26.10.2000**

(97) Erstveröffentlichung durch das EPA: **04.04.2001**

(97) Veröffentlichungstag  
der Patenterteilung beim EPA: **19.10.2005**

(47) Veröffentlichungstag im Patentblatt: **13.07.2006**

(30) Unionspriorität:  
**130016 P 19.04.1999 US**

(84) Benannte Vertragsstaaten:  
**DE, FR, GB**

(73) Patentinhaber:  
**AT & T Corp., New York, N.Y., US**

(72) Erfinder:  
**KAPILOW, A., David, Berkeley Heights, US**

(74) Vertreter:  
**Schieber und Kollegen, 80469 München**

(54) Bezeichnung: **VERFAHREN ZUR VERSCHLEIERUNG VON PAKETVERLUSTEN**

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

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

**Beschreibung**

## HINTERGRUND DER ERFINDUNG

## Gebiet der Erfindung

**[0001]** Diese Erfindung bezieht sich auf Techniken zum Ausführen der Paketverlust- oder Frame-Löschung-Verdeckung (FEC = frame erasure concealment).

## Beschreibung der zugehörigen Technik

**[0002]** Algorithmen der Frame-Löschung-Verdeckung (FEC) verstecken Sendungsverluste in einem Sprachkommunikationssystem, worin ein eingegebenes Sprachsignal verschlüsselt und an einem Sender verpackt, über ein Netzwerk (einer beliebigen Art) gesendet und an einem Empfänger empfangen wird, der das Paket entschlüsselt und die Sprachausgabe wiedergibt. Viele der CELP-basierten Standard-Sprachcodierer, wie das G.723.1, G.728 und G.729, verfügen über FEC-Algorithmen, die in ihren Standards eingebaut oder beabsichtigt sind.

**[0003]** Die Aufgabe des FEC besteht darin, ein synthetisches Sprachsignal zu erzeugen, um fehlende Daten in einem empfangenen Bit-Strom abzudecken. Im Idealfall wird das synthetisierte Signal die gleiche Klangfarbe und spektrale Charakteristika wie das fehlende Signal aufweisen und keine unnatürlichen Artefakte erzeugen. Weil Sprachsignale häufig örtlich stationär sind, ist es möglich, die vergangene History (Vorgeschichte) des Signals zu verwenden, um eine vernünftige Näherung zu dem fehlenden Abschnitt zu erzeugen. Wenn die Löschungen nicht zu lang sind und die Löschung nicht auf einen Bereich fällt, worin sich das Signal rasch verändert, dann können die Löschungen nach der Verdeckung unhörbar sein.

**[0004]** Frühere Systeme haben Vervielfältigungstechniken der Tonlagen-Wellenform verwendet, um Frame-Löschungen zu verdecken, wie zum Beispiel D. J. Goodman et al., "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications", Vol. 34, Nr. 6 IEEE Abhandlungen über Akustik, Sprache und Signalverarbeitung 1440-48 (Dezember 1996) und O.J. Wasem et al., "The Effect of Waveform Substitution on the Quality of PCM Packet Communications", Vol. 36, Nr. 3 IEEE Abhandlungen über Akustik, Sprache und Signalverarbeitung 342-48 (März 1988).

**[0005]** Obwohl das Vervielfältigen der Tonlagen-Wellenform und die Techniken der Überlappung-Addition verwendet worden sind, um Signale zu synthetisieren, damit verlorene Frames von Sprachdaten verdeckt werden, führen diese Techniken manchmal zu unnatürlichen Artefakten, die für den Zuhörer unbefriedigend sind.

**[0006]** Der Artikel "A new technique for audio packet loss concealment", Sanneck H. et al., Konferenz über Globale Telekommunikation (GLOBECOM), US, New York, IEEE, 18. Nov. 1996, Seiten 48-52, beschreibt einen Ansatz für die Fehler-Verdeckung für die Audio-Sendung basierend auf einer Zeitskala-Modifizierung von richtig empfangenen Paketen, worin die Zeitskala-Modifizierung aus einem passenden Algorithmus, dem Algorithmus der Überlappung-Addition mit Wellenform-Ähnlichkeit (WSOLA = waveform similarity overlap-add) resultiert. Die Zeitskala-Modifizierung, wie in dem Artikel beschrieben, verwendet eine gespreizte Version eines Sprachteils, die der Lücke unmittelbar folgt, die durch ein fehlendes Paket verursacht wurde.

## ZUSAMMENFASSUNG DER ERFINDUNG

**[0007]** Die vorliegende Erfindung ist, wie in dem Anspruch 1 dargelegt, auf eine Technik zum Verringern von nicht natürlichen Artefakten in Sprache gerichtet, die durch ein Sprachdecodierersystem erzeugt wurde, die aus der Anwendung einer FEC-Technik resultieren können. Die Technik bezieht sich auf die Erzeugung eines Sprachsignals durch einen Sprachdecodierer basierend auf empfangenen Paketen, die eine Sprachinformation darstellen, und, als Antwort auf eine Bestimmung, dass ein Sprachdaten enthaltendes Paket an dem Decodierer nicht verfügbar ist, um das Sprachsignal zu bilden, das Synthetisieren eines Teils von dem Sprachsignal entsprechend dem nicht verfügbaren Paket unter Verwendung eines Teils des zuvor gebildeten Sprachsignals. Wenn das zu erzeugende Sprachsignal eine Grundfrequenz oberhalb eines bestimmten Grenzwertes besitzt (z. B. eine Frequenz, die mit einem Kleinkind assoziiert ist), dann wird eine größere Anzahl von Tonlagen-Perioden des zuvor gebildeten Sprachsignals dazu verwendet, um die Sprache zu synthetisieren, als im Vergleich zu der Situation, in der sich die Grundfrequenz unterhalb des Grenzwertes befindet (z. B. eine Frequenz, die mit einem erwachsenen Mann assoziiert ist).

## KURZE BESCHREIBUNG DER ZEICHNUNGEN

- [0008] Die Erfindung wird im Detail mit Bezug auf die folgenden Figuren beschrieben, worin gleiche Bezugsziffern gleiche Elemente bezeichnen und worin:
- [0009] [Fig. 1](#) ein beispielhaftes Audio-Sendesystem ist;
- [0010] [Fig. 2](#) ein beispielhaftes Audio-Sendesystem mit einem G.711-Codierer und FEC-Modul ist;
- [0011] [Fig. 3](#) ein ausgegebenes Audio-Signal unter Verwendung einer FEC-Technik darstellt;
- [0012] [Fig. 4](#) einen Arbeitsablauf der Überlappung-Addition (OLA = overlap-add) am Ende einer Löschung darstellt;
- [0013] [Fig. 5](#) ein Flussdiagramm eines beispielhaften Prozesses zum Ausführen des FEC unter Verwendung eines G.711-Codierers ist;
- [0014] [Fig. 6](#) ein Graph ist, der den Aktualisierungsprozess des History-Puffers darstellt;
- [0015] [Fig. 7](#) ein Flussdiagramm eines beispielhaften Prozesses ist, um das erste Frame des Signals zu verdecken;
- [0016] [Fig. 8](#) die Tonlagen-Abschätzung aus einer Autokorrelation darstellt;
- [0017] [Fig. 9](#) feine gegenüber groben Tonlagen-Abschätzungen darstellt;
- [0018] [Fig. 10](#) Signale in dem Tonlagen- und letzten ¼-Puffer darstellt;
- [0019] [Fig. 11](#) eine synthetische Signalerzeugung unter Verwendung eines 1-Periode-Tonlagen-Puffers darstellt;
- [0020] [Fig. 12](#) ein Flussdiagramm eines beispielhaften Prozesses ist, um das zweite oder später gelöschte Frame des Signals zu verdecken;
- [0021] [Fig. 13](#) synthetisierte Signale darstellt, die in das zweite gelöschte Frame fortgesetzt werden;
- [0022] [Fig. 14](#) eine synthetische Signalerzeugung unter Verwendung eines 2-Perioden-Tonlagen-Puffers darstellt;
- [0023] [Fig. 15](#) ein OLA am Anfang des zweiten gelöschten Frames darstellt;
- [0024] [Fig. 16](#) ein Flussdiagramm eines beispielhaften Verfahrens zum Verarbeiten des ersten Frames nach der Löschung ist;
- [0025] [Fig. 17](#) eine synthetische Signalerzeugung unter Verwendung eines 3-Perioden-Tonlagen-Puffers darstellt; und
- [0026] [Fig. 18](#) ein Blockdiagramm ist, das die Verwendung von FEC-Techniken mit anderen Sprachcodierern darstellt.

## DETAILLIERTE BESCHREIBUNG VON BEVORZUGTEN AUSFÜHRUNGSFORMEN

[0027] Kürzlich hat es ein großes Interesse an der Verwendung des G.711 auf Paket-Netzwerken ohne eine garantierte Qualität des Dienstes gegeben, um die herkömmliche Telefonie (POTS = plain-old-telephony service) zu unterstützen. Wenn sich Frame-Löschungen (oder Paketverluste) auf diesen Netzwerken ereignen, dann sind Verdeckungstechniken erforderlich oder die Qualität des Gesprächs wird ernsthaft herabgesetzt. Eine Technik der Frame-Löschung-Verdeckung (FEC) mit hoher Qualität und niedriger Komplexität ist entwickelt worden und wird im Detail unten beschrieben.

[0028] Ein beispielhaftes Blockdiagramm eines Audio-Systems mit FEC ist in der [Fig. 1](#) gezeigt. In der [Fig. 1](#)

empfängt ein Codierer **110** ein eingegebenes Audio-Frame und gibt einen verschlüsselten Bit-Strom aus. Der Bit-Strom wird von dem Verlust-Frame-Detektor **115** empfangen, der bestimmt, ob irgendwelche Frames verloren worden sind. Wenn der Verlust-Frame-Detektor **115** bestimmt, dass Frames verloren worden sind, dann signalisiert der Verlust-Frame-Detektor **115** dem FEC-Modul **130**, einen FEC-Algorithmus oder einen Prozess anzuwenden, um die fehlenden Frames wieder herzustellen.

**[0029]** Somit versteckt der FEC-Prozess Sendungsverluste in einem Audio-System, worin das eingegebene Signal verschlüsselt und an einem Sender verpackt wird, über ein Netzwerk gesendet und an einem Verlust-Frame-Detektor **115** empfangen wird, der bestimmt, dass ein Frame verloren worden ist. Es wird in der [Fig. 1](#) angenommen, dass der Verlust-Frame-Detektor **115** über eine Art der Bestimmung darüber verfügt, ob ein erwartetes Frame nicht ankommt oder zu spät ankommt, um verwendet zu werden. Auf IP-Netzwerken wird das normalerweise durch Addieren einer Sequenznummer oder eines Zeitstempels zu den Daten in dem gesendeten Frame implementiert. Der Verlust-Frame-Detektor **115** vergleicht die Sequenz-Nummern der ankommenden Frames mit den Sequenz-Nummern, die erwartet würden, wenn keine Frames verloren worden wären. Wenn der Verlust-Frame-Detektor **115** erfasst, dass ein Frame erwartungsgemäß angekommen ist, dann wird es von dem Decodierer **120** entschlüsselt und der ausgegebene Audio-Frame wird an das Ausgabesystem gegeben. Wenn ein Frame verloren wurde, dann wendet das FEC-Modul **130** einen Prozess an, um das fehlende Audio-Frame zu verstecken, indem ein Wert eines synthetischen Frames anstatt dem Audio erzeugt wird.

**[0030]** Viele der ITU-T-CELP-basierten Standard-Sprachcodierer, wie das G.723.1, G.728 und G.729, modellieren die Sprachvielfältigkeit in ihren Decodierern. Somit verfügen die Decodierer über eine genügende Zustandsinformation, um den FEC-Prozess direkt in den Decodierer zu integrieren. Diese Sprachcodierer besitzen FEC-Algorithmen oder Prozesse, die als ein Teil ihrer Standards spezifiziert sind.

**[0031]** Im Vergleich dazu ist das G.711 ein Codierungsschema mit Abtastung-für-Abtastung, das die Sprachvielfältigkeit nicht modelliert. Es gibt keine Zustandsinformation in dem Codierer, um in dem FEC zu helfen. Als ein Ergebnis ist der FEC-Prozess mit dem G.711 von dem Codierer unabhängig.

**[0032]** Ein beispielhaftes Blockdiagramm des Systems, wie es mit dem G.711-Codierer verwendet wird, ist in der [Fig. 2](#) gezeigt. Wie in der [Fig. 1](#) verschlüsselt der G.711-Codierer **210** die Bit-Stromdaten und sendet sie zu dem Verlust-Frame-Detektor **215**. Wiederum vergleicht der Verlust-Frame-Detektor **215** die Sequenz-Nummern der ankommenden Frames mit den Sequenz-Nummern, die erwartet würden, wenn keine Frames verloren worden wären. Wenn ein Frame erwartungsgemäß ankommt, dann wird es zum Decodieren durch den Decodierer **220** weitergeleitet und dann an einen History-Puffer **240** ausgegeben, der das Signal speichert. Wenn ein Frame verloren wurde, dann informiert der Verlust-Frame-Detektor **215** das FEC-Modul **230**, das einen Prozess anwendet, um das fehlende Audio-Frame zu verstecken, indem ein Wert eines synthetischen Frames anstatt dem Audio erzeugt wird.

**[0033]** Um die fehlenden Frames zu verstecken, wendet jedoch das FEC-Modul **230** einen G.711-FEC-Prozess an, der die vergangene History (Vorgeschichte) von dem entschlüsselten Ausgangssignal verwendet, die von dem History-Puffer **240** bereitgestellt wird, um abzuschätzen, was das Signal in dem fehlenden Frame sein sollte. Außerdem verzögert ein Verzögerungsmodul **250** ferner die Ausgabe des Systems um eine vorbestimmte Zeitperiode, zum Beispiel um 3,75 ms, um einen glatten Übergang zwischen gelöschten und nicht gelöschten Frames sicherzustellen. Diese Verzögerung gestattet es dem synthetischen Löschungssignal, dass es mit dem realen Ausgangssignal am Anfang einer Löschung langsam eingemischt wird.

**[0034]** Die Pfeile zwischen dem FEC-Modul **230** und jedem der Blöcke des History-Puffers **240** und des Verzögerungsmoduls **250** zeigen an, dass die gespeicherte History von dem FEC-Prozess verwendet wird, um das synthetische Signal zu erzeugen. Außerdem wird die Ausgabe von dem FEC-Modul **230** verwendet, um den History-Puffer **240** während einer Löschung zu aktualisieren. Es sollte beachtet werden, dass der Prozess genauso gut ablaufen wird, wenn kein Sprachcodierer vorhanden ist, weil der FEC-Prozess ausschließlich von der entschlüsselten Ausgabe des G.711 abhängt.

**[0035]** Ein graphisches Beispiel dessen, wie das Eingangssignal von dem FEC-Prozess in dem FEC-Modul **230** verarbeitet wird, ist in der [Fig. 3](#) gezeigt.

**[0036]** Die obere Wellenform in der Figur zeigt die Eingabe an das System, wenn eine 20 ms-Löschung in einem Bereich von stimmhafter Sprache von einem männlichen Sprecher auftritt. In der Wellenform darunter hat der FEC-Prozess die fehlenden Abschnitte durch die Erzeugung von synthetischer Sprache in der Lücke

verdeckt. Für Vergleichszwecke ist ferner das ursprüngliche Eingangssignal ohne eine Löschung gezeigt. In einem idealen System klingt die verdeckte Sprache genauso wie die ursprüngliche Sprache. Wie aus der Figur ersichtlich ist, ähnelt die synthetische Wellenform eng der ursprünglichen Wellenform in den fehlenden Abschnitten. Wie die Wellenform "Verdeckt" aus der Wellenform "Eingabe" erzeugt wird, wird im Detail unten diskutiert.

**[0037]** Der von dem FEC-Modul **230** verwendete FEC-Prozess verdeckt das fehlende Frame durch die Erzeugung einer synthetischen Sprache, die ähnliche Charakteristika zu der Sprache aufweist, die in dem History-Puffer **240** gespeichert ist. Die grundlegende Idee ist wie folgt. Wenn das Signal stimmhaft ist, dann nehmen wir an, dass das Signal quasi-periodisch und örtlich stationär ist. Wir schätzen die Tonlage ab und wiederholen die letzte Tonlagen-Periode in dem History-Puffer **240** einige wenige Male. Wenn die Löschung lang ist oder die Tonlage kurz ist (die Frequenz ist hoch), dann führt jedoch das Wiederholen der gleichen Tonlagen-Periode zu viele Male zu einer Ausgabe, die verglichen mit der natürlichen Sprache zu harmonisch ist. Um diese harmonischen Artefakte zu vermeiden, die als Piep- und Gongtöne hörbar sind, wird die Anzahl von Tonlagen-Perioden, die aus dem History-Puffer **240** verwendet wurden, mit der voranschreitenden Länge der Löschung erhöht. Kurze Löschungen verwenden nur die letzte oder die letzten wenigen Tonlagen-Perioden aus dem History-Puffer **240**, um das synthetische Signal zu erzeugen. Lange Löschungen verwenden ferner Tonlagen-Perioden, die in dem History-Puffer **240** weiter zurückliegen. Mit langen Löschungen werden die Tonlagen-Perioden aus dem History-Puffer **240** nicht in der gleichen Reihenfolge wiedergegeben, in der sie in der ursprünglichen Sprache aufgetreten sind. Durch Testen wurde jedoch herausgefunden, dass das in langen Löschungen erzeugte synthetische Sprachsignal noch einen natürlichen Klang erzeugt. Je länger die Löschung ist, umso wahrscheinlicher ist es, dass das synthetische Signal von dem realen Signal abweicht. Um Artefakte zu vermeiden, die durch das zu lange Halten von bestimmten Klangtypen bewirkt wurden, wird das synthetische Signal abgeschwächt, wenn die Löschung länger wird. Für Löschungen mit einer Dauer von 10 ms oder weniger ist keine Abschwächung erforderlich. Für Löschungen, die länger als 10 ms sind, wird das synthetische Signal mit der Rate von 20% pro zusätzliche 10 ms abgeschwächt. Über 60 ms hinaus wird das synthetische Signal auf Null gesetzt (Geräuschlosigkeit). Das geschieht, weil das synthetische Signal derart unähnlich zu dem ursprünglichen Signal ist, dass es im Durchschnitt mehr schadet als nützt, um den Verdeckungsversuch der fehlenden Sprache nach 60 ms fortzusetzen.

**[0038]** Jedes Mal, wenn ein Übergang zwischen Signalen aus verschiedenen Quellen gemacht wird, ist es wichtig, dass der Übergang keine Unstetigkeiten, die als Klicken hörbar sind, oder unnatürliche Artefakte in das Ausgangssignal einführt. Diese Übergänge treten an verschiedenen Stellen auf:

1. Am Anfang der Löschung an der Grenze zwischen dem Anfang des synthetischen Signals und dem Ausläufer des letzten einwandfreien Frames.
2. Am Ende der Löschung an der Grenze zwischen dem synthetischen Signal und dem Anfang des Signals in dem ersten einwandfreien Frame nach der Löschung.
3. Jedes Mal, wenn die Anzahl von Tonlagen-Perioden verändert wird, die aus dem History-Puffer **240** verwendet wird, um die Signalvariation zu erhöhen.
4. An den Grenzen zwischen den wiederholten Teilen des History-Puffers **240**.

**[0039]** Um glatte Übergänge sicherzustellen, werden Überlappung-Additionen (OLA) an sämtlichen Signalgrenzen ausgeführt. OLAs stellen einen Weg von glattem Verbinden zweier Signale dar, die an einer Kante überlappen. In dem Bereich, in dem die Signale überlappen, werden die Signale durch Fenster gewichtet und dann zusammen addiert (vermischt). Die Fenster sind derart ausgestaltet, dass die Summe der Gewichtungen bei irgendeiner bestimmten Abtastung gleich 1 ist. Das heißt, dass keine Verstärkung oder Abschwächung auf die Gesamtsumme der Signale angewendet wird. Außerdem sind die Fenster derart ausgestaltet, dass das Signal zur Linken mit der Gewichtung 1 beginnt und stufenweise auf 0 ausklingt, während das Signal zur Rechten mit der Gewichtung 0 beginnt und stufenweise auf die Gewichtung 1 einklingt. Somit ist in dem Bereich zur Linken des Überlappungsfensters nur das linke Signal vorhanden, während in dem Bereich zur Rechten des Überlappungsfensters nur das rechte Signal vorhanden ist. In dem Überlappungsbereich macht das Signal stufenweise einen Übergang von dem Signal zur Linken zu dem Signal zur Rechten. In dem FEC-Prozess werden dreieckige Fenster verwendet, um die Komplexität der Berechnung der Fenster variabler Länge niedrig zu halten, aber andere Fenster, wie Hanning-Fenster, können an Stelle davon verwendet werden.

**[0040]** Die [Fig. 4](#) zeigt die synthetische Sprache am Ende einer 20 ms-Löschung, auf die das OLA mit der realen Sprache angewendet wurde, die beginnt, nachdem die Löschung zu Ende ist. In diesem Beispiel ist das OLA-Gewichtungsfenster ein dreieckiges 5,75 ms-Fenster. Das obere Signal ist das synthetische Signal, das während der Löschung erzeugt wurde, und das überlappende Signal darunter ist die reale Sprache nach der Löschung. Die OLA-Gewichtungsfenster sind unterhalb der Signale gezeigt. Hier stimmen aufgrund einer Ton-

lagen-Veränderung in dem realen Signal während der Löschung die Peaks von den synthetischen und realen Signalen nicht überein und die Unstetigkeit, die eingeführt wird, wenn wir versuchen, die Signale ohne ein OLA zu kombinieren, ist in dem Graphen mit der Bezeichnung "Kombiniert ohne OLA" gezeigt. Der Graph "Kombiniert ohne OLA" wurde durch Aufkopieren des synthetischen Signals bis zu dem Anfang des OLA-Fensters und des realen Signals für die Dauer erzeugt. Das Ergebnis der OLA-Arbeitsabläufe zeigt, wie die Unstetigkeiten an den Grenzen geglättet werden.

**[0041]** Die vorherige Diskussion betrifft, wie ein beispielhafter Prozess mit einer stationären stimmhaften Sprache arbeitet, aber wenn sich die Sprache rasch verändert oder stimmlos ist, dann kann die Sprache keine periodische Struktur besitzen. Diese Signale werden jedoch auf die gleiche Art und Weise verarbeitet, wie unten dargelegt wird.

**[0042]** Erstens beträgt die kleinste Tonlagen-Periode, die wir in der beispielhaften Ausführungsform zulassen, 5 ms in der Tonlagen-Abschätzung entsprechend einer Frequenz von 200 Hz. Während es bekannt ist, dass einige weibliche und kindliche Sprecher mit hoher Frequenz Grundfrequenzen über 200 Hz besitzen, beschränken wir es auf 200 Hz, so dass die Fenster relativ groß bleiben. Auf diese Art und Weise wird innerhalb eines gelöschten 10 ms-Frames die selektierte Tonlagen-Periode maximal zweimal wiederholt. Bei Sprechern mit hoher Frequenz setzt das die Ausgabe nicht wirklich herab, weil der Tonlagen-Abschätzer ein Mehrfaches von der realen Tonlagen-Periode zurückführt. Und dadurch, dass keine Sprache zu häufig wiederholt wird, erzeugt der Prozess keine synthetische periodische Sprache aus einer nicht periodischen Sprache. Zweitens wird eine genügende Variation zu dem Signal addiert, so dass eine Periodizität für lange Löschungen nicht eingeführt wird, weil die Anzahl von Tonlagen-Perioden, die verwendet wird, um die synthetische Sprache zu erzeugen, erhöht wird, während die Löschung länger wird.

**[0043]** Es sollte beachtet werden, dass der Prozess der Überlappung-Addition mit Wellenform-Ähnlichkeit (WSOLA = Waveform Similarity Overlapp Add) für die Zeitskalierung von Sprache ferner große OLA-Fenster mit fester Größe verwendet, so dass der gleiche Prozess verwendet werden kann, um sowohl periodische als auch nicht periodische Sprachsignale mit Zeitskalierung zu versehen.

**[0044]** Während ein Überblick von dem beispielhaften FEC-Prozess oben gegeben wurde, werden die individuellen Schritte im Detail unten diskutiert.

**[0045]** Zum Zwecke dieser Diskussion nehmen wir an, dass ein Frame **10** ms an Sprache enthält und dass zum Beispiel die Abtastungsrate 8 kHz beträgt. Somit können Löschungen in Inkrementen von 80 Abtastungen auftreten ( $8000 \cdot 0,010 = 80$ ). Es sollte beachtet werden, dass der FEC-Prozess auf weitere Frame-Größen und Abtastungsraten einfach anpassbar ist. Um die Abtastrate zu verändern, muss man nur die in Millisekunden angegebenen Zeitperioden mit 0,001 und dann mit der Abtastrate multiplizieren, um die passenden Puffergrößen zu erhalten. Zum Beispiel enthält der History-Puffer **240** die letzten 48,75 ms der Sprache. Bei 8kHz würde das implizieren, dass der Puffer ( $48,75 \cdot 0,001 \cdot 8000$ ) = 390 Abtastungen lang ist. Bei einer Abtastrate von 16 kHz würde das verdoppelt werden bzw. 780 Abtastungen bedeuten.

**[0046]** Mehrere der Puffergrößen basieren auf der niedrigsten Frequenz, die der Prozess erwartet. Zum Beispiel nimmt der beispielhafte Prozess an, dass die niedrigste Frequenz, die bei einer Abtastrate von 8 kHz zu sehen sein wird,  $66 \frac{2}{3}$  Hz beträgt. Dies führt zu einer maximalen Tonlagen-Periode von 15 ms ( $1/(66 \frac{2}{3}) = 0,015$ ). Die Länge des History-Puffers **240** beträgt das 3,25-fache der Periode der niedrigsten Frequenz. Folglich beträgt der History-Puffer **240** somit  $15 \cdot 3,25 = 48,75$  ms. Wenn bei einer Abtastrate von 16 kHz die Eingabefilter Frequenzen gestatten, die so niedrig sind wie 50 Hz (20 ms Periode), dann müsste der History-Puffer **240** auf  $20 \cdot 3,25 = 65$  ms verlängert werden.

**[0047]** Die Frame-Größe kann auch verändert werden; 10 ms wurden als die Voreinstellung gewählt, weil das die Frame-Größe ist, die von mehreren Standard-Sprachcodierern, wie das G.729, verwendet wird, und ferner in mehreren drahtlosen Systemen verwendet wird. Das Verändern der Frame-Größe ist einfach. Wenn die erwünschte Frame-Größe ein Mehrfaches von 10 ms ist, dann bleibt der Prozess unverändert. Man lässt einfach die Frame-Größe des Lösungsprozesses bei 10 ms und ruft es mehrere Male pro Frame auf. Wenn die erwünschte Frame-Größe des Pakets ein Teiler von 10 ms ist, wie 5 ms, dann bleibt der FEC-Prozess grundlegend unverändert. Die Rate, bei der die Anzahl von Perioden in dem Tonlagen-Puffer erhöht wird, wird jedoch basierend auf der Anzahl von Frames in 10 ms modifiziert werden müssen. Frame-Größen, die nicht Mehrfache oder Teiler von 10 ms sind, wie z. B. 12 ms, können ferner untergebracht werden. Der FEC-Prozess ist beim Verändern der Steigerungsrate der Anzahl von Tonlagen-Perioden vernünftig verzeihend, die aus dem Tonlagen-Puffer verwendet werden. Das Erhöhen der Anzahl von Perioden einmal alle 12 ms an Stelle von

einmal alle 10 ms macht keinen allzu großen Unterschied.

[0048] Die [Fig. 5](#) ist ein Blockdiagramm von dem FEC-Prozess, der von der beispielhaften Ausführungsform der [Fig. 2](#) ausgeführt wird. Die erforderlichen Teilschritte, um einige von den Hauptarbeitsabläufen zu implementieren, sind ferner in den [Fig. 7](#), [Fig. 12](#) und [Fig. 16](#) detailliert gezeigt und werden unten diskutiert. In der folgenden Diskussion werden mehrere Variablen verwendet, um Werte und Puffern zu halten. Diese Variablen sind unten zusammengefasst:

Tabelle 1: Variablen und ihre Inhalte

Variable	Typ	Beschreibung	Bemerkung
B	Feld	Tonlagen-Puffer	Bereich [-P*3, 35:-1]
H	Feld	History-Puffer	Bereich [-390:-1]
L	Feld	letztes $\frac{1}{4}$ -Puffer	Bereich [-P*0, 25:-1]
O	Skalar	Abweichung in dem Tonlagen-Puffer	
P	Skalar	Tonlagen-Abschätzung	$40 \leq P < 120$
P4	Skalar	$\frac{1}{4}$ - Tonlagen-Abschätzung	$P4 = P \gg 2$
S	Feld	synthetisierte Sprache	Bereich [0:79]
U	Skalar	verwendete Wellenlängen	$1 \leq U \leq 3$

[0049] Wie in dem Flussdiagramm in der [Fig. 5](#) gezeigt ist, beginnt der Prozess und bei dem Schritt 505 wird das nächste Frame von dem Verlust-Frame-Detektor 215 empfangen. In dem Schritt 510 bestimmt der Verlust-Frame-Detektor 215, ob das Frame gelöscht wurde. Wenn das Frame nicht gelöscht wurde, dann wird in dem Schritt 512 das Frame von dem Decoder 220 entschlüsselt. Dann wird in dem Schritt 515 das entschlüsselte Frame in dem History-Puffer 240 für die Verwendung durch das FEC-Modul 230 gespeichert.

[0050] In dem Aktualisierungsschritt des History-Puffers beträgt die Länge von diesem Puffer 240 das 3,25-fache der Länge von der längsten erwarteten Tonlagen-Periode. Bei einer Abtastrate von 8kHz beträgt die längste Tonlagen-Periode 15 ms, oder 120 Abtastungen, so dass die Länge des History-Puffers 240 48,75 ms, oder 390 Abtastungen, beträgt. Nachdem jedes Frame von dem Decoder 220 entschlüsselt wurde, wird deshalb der History-Puffer 240 aktualisiert, so dass er die neueste Sprach-History enthält. Die Aktualisierung des History-Puffers 240 ist in der [Fig. 6](#) gezeigt. Wie in dieser Figur gezeigt ist, enthält der History-Puffer 240 die neuesten Sprachabtastungen auf der rechten Seite und die ältesten Sprachabtastungen auf der linken Seite. Wenn das neueste Frame von der entschlüsselten Sprache empfangen wird, dann wird er in den Puffer 240 von der rechten Seite verschoben, während die Abtastungen, die der ältesten Sprache entsprechen, aus dem Puffer auf der linken Seite verschoben werden (siehe 6b).

[0051] Außerdem verzögert in dem Schritt 520 das Verzögerungsmodul 250 die Ausgabe der Sprache um  $\frac{1}{4}$  von der längsten Tonlagen-Periode. Bei einer Abtastrate von 8 kHz sind das  $120 \cdot \frac{1}{4} = 30$  Abtastungen, oder 3,75 ms. Diese Verzögerung gestattet es dem FEC-Modul 230, dass es eine  $\frac{1}{4}$ -Wellenlänge-OLA am Anfang von einer Löschung ausführt, um einen glatten Übergang zwischen dem realen Signal vor der Löschung und dem synthetischen Signal sicherzustellen, das von dem FEC-Modul 230 erzeugt wurde. Die Ausgabe muss

verzögert werden, weil nach dem Decodieren eines Frames es nicht bekannt ist, ob das nächste Frame gelöscht ist.

**[0052]** In dem Schritt **525** wird das Audio-Signal ausgegeben und in dem Schritt **530** bestimmt der Prozess, ob noch irgendwelche weitere Frames vorhanden sind. Wenn es keine weiteren Frames gibt, dann endet der Prozess. Wenn es weitere Frames gibt, dann geht der Prozess zu dem Schritt **505** zurück, um das nächste Frame zu erhalten.

**[0053]** Wenn in dem Schritt **510** der Verlust-Frame-Detektor **215** bestimmt, dass das empfangene Frame gelöscht wurde, dann geht jedoch der Prozess zu dem Schritt **535**, worin das FEC-Modul **230** das erste gelöschte Frame verdeckt, dessen Prozess im Detail unten in der [Fig. 7](#) beschrieben ist. Nachdem das erste Frame verdeckt wurde, erhält in dem Schritt **540** der Verlust-Frame-Detektor **215** das nächste Frame. In dem Schritt **545** bestimmt der Verlust-Frame-Detektor **215**, ob das nächste Frame gelöscht ist. Wenn das nächste Frame nicht gelöscht ist, dann verarbeitet in dem Schritt **555** das FEC-Modul **230** das erste Frame nach der Löschung, dessen Prozess im Detail unten in der [Fig. 16](#) beschrieben ist. Nachdem das erste Frame verarbeitet wurde, kehrt der Prozess zu dem Schritt **530** zurück, worin der Verlust-Frame-Detektor **215** bestimmt, ob irgendwelche weitere Frames vorhanden sind.

**[0054]** Wenn in dem Schritt **545** der Verlust-Frame-Detektor **215** bestimmt, dass der nächste oder dass die nachfolgenden Frames gelöscht sind, dann verdeckt das FEC-Modul **230** den zweiten und die nachfolgenden Frames gemäß einem Prozess, der im Detail unten in der [Fig. 12](#) beschrieben ist.

**[0055]** Die [Fig. 7](#) zeigt detailliert die Schritte, die unternommen werden, um die ersten 10 ms von einer Löschung zu verdecken. Die Schritte werden im Detail unten untersucht.

**[0056]** Wie in der [Fig. 7](#) gesehen werden kann, besteht in dem Schritt **705** der erste Arbeitsablauf am Anfang einer Löschung darin, die Tonlage abzuschätzen. Um das zu bewerkstelligen, wird eine normierte Autokorrelation an dem Signal des History-Puffers **240** mit einem Fenster von 20 ms (160 Abtastungen) bei Abgriff-Verzögerungen von 40 bis 120 Abtastungen ausgeführt. Bei einer Abtastrate von 8 kHz entsprechen diese Verzögerungen Tonlagen-Perioden von 5 bis 15 ms oder Grundfrequenzen von 200 bis 66 2/3 Hz. Der Abgriff an dem Peak der Autokorrelation ist die Tonlagen-Abschätzung P. Unter der Annahme, dass H diese History enthält und dass es von -1 (die Abtastung unmittelbar vor der Löschung) bis -390 (die Abtastung **390** Abtastungen, bevor die Löschung anfängt) indexiert ist, kann die Autokorrelation für den Abgriff j mathematisch ausgedrückt werden als:

$$Autocor(j) = \frac{\sum_{i=1}^{160} H[-i]H[-i-j]}{\sqrt{\sum_{k=1}^{160} H^2[-k-j]}}$$

**[0057]** Der Peak der Autokorrelation oder die Tonlagen-Abschätzung kann dann ausgedrückt werden als:

$$P = \{\max_j(Aulocor(j)) | 40 \leq j \leq 120\}$$

**[0058]** Wie oben erwähnt ist, ist die niedrigste gestattete Abstand-Periode, die 5 ms oder 40 Abtastungen beträgt, groß genug, dass eine einzelne Tonlagen-Periode maximal zweimal in einem gelöschten 10 ms-Frame wiederholt wird. Das vermeidet Artefakte in nicht stimmhafter Sprache und vermeidet ferner unnatürliche harmonische Artefakte in Lautsprechern mit einer hohen Tonlage.

**[0059]** Ein graphisches Beispiel der Berechnung der normierten Autokorrelation für die Löschung in der [Fig. 3](#) ist in der [Fig. 8](#) gezeigt.

**[0060]** Die mit "History" bezeichnete Wellenform ist der Inhalt von dem History-Puffer **240** unmittelbar vor der Löschung. Die gestrichelte waagrechte Linie zeigt den Referenzteil des Signals, den History-Puffer 240  $H[-1]:H[-160]$ , der die 20 ms an Sprache unmittelbar vor der Löschung sind. Die durchgezogenen waagrechten Linien sind die 20 ms-Fenster, die bei Abgriffen von 40 Abtastungen (die obere Linie, 5 ms-Periode, 200Hz-Frequenz) bis 120 Abtastungen (die untere Linie, 15 ms-Periode, 66,66 Hz-Frequenz) verzögert sind. Die Ausgabe von der Korrelation wird ferner zusammen mit den Positionen der Fenster aufgetragen. Die gepunktete senkrechte Linie in der Korrelation ist der Peak von der Kurve und stellt die abgeschätzte Tonlage

dar. Diese Linie ist eine Periode vor dem Anfang der Löschung. In diesem Fall ist P gleich 56 Abtastungen, entsprechend einer Tonlagen-Periode von 7 ms und einer Grundfrequenz von 142,9 Hz.

**[0061]** Um die Komplexität der Autokorrelation zu senken, werden zwei spezielle Prozeduren verwendet. Während diese Abkürzungen die Ausgabe nicht wesentlich verändern, verfügen sie über einen großen Einfluss auf die gesamte Laufzeit-Komplexität des Prozesses. Der Großteil der Komplexität in dem FEC-Prozess ist in der Autokorrelation angesiedelt.

**[0062]** Anstatt die Korrelation bei jedem Abgriff zu errechnen, wird zuerst eine grobe Abschätzung von dem Peak auf einem dezimierten Signal bestimmt und dann wird eine Feinsuche in der Nachbarschaft von dem groben Peak ausgeführt. Für die grobe Abschätzung modifizieren wir die obige Autocor-Funktion in die neue Funktion, die auf einem 2:1-dezimierten Signal arbeitet und nur jeden anderen Abgriff prüft:

$$Autocor_{rough}(j) = \frac{\sum_{i=1}^{80} H[-2i]H[-2i-j]}{\sqrt{\sum_{k=1}^{80} H^2[-2k-j]}}$$

$$P_{rough} = 2\{\max_j(Autocor_{rough}(2j)) \mid 20 \leq j \leq 60\}$$

**[0063]** Dann wird unter Verwendung der groben Abschätzung der ursprüngliche Suchprozess wiederholt, aber nur in dem Bereich  $P_{rough} - 1 \leq j \leq P_{rough} + 1$ . Es wird auf die Sicherstellung davon geachtet, dass j in dem ursprünglichen Bereich zwischen 40 und 120 Abtastungen verbleibt. Man beachte, dass der Dezimierungs-Faktor ebenfalls erhöht werden sollte, wenn die Abtastrate erhöht wird, so dass die gesamte Komplexität von dem Prozess annähernd konstant bleibt. Wir haben Untersuchungen mit Dezimierungs-Faktoren von 8:1 auf Sprache durchgeführt, die bei 44,1 kHz abgetastet wurde, und positive Ergebnisse erhalten. Die [Fig. 9](#) vergleicht den Graphen der Funktion  $Autocor_{rough}$  mit dem der Funktion  $Autocor$ . Wie in der Figur gesehen werden kann, stellt  $Autocor_{rough}$  eine gute Näherung zu  $Autocor$  dar und die Komplexität erniedrigt sich um beinahe einen Faktor von 4 bei einer Abtastrate von 8kHz – ein Faktor von 2, weil nur jeder andere Abgriff geprüft wird und ein Faktor von 2, weil bei einem gegebenen Abgriff nur jede andere Abtastung geprüft wird.

**[0064]** Die zweite Prozedur wird ausgeführt, um die Komplexität der Energieberechnung in  $Autocor$  und  $Autocor_{rough}$  zu senken. Anstatt die Gesamtsumme an jedem Schritt zu errechnen, wird eine laufende Summe der Energie aufrechterhalten. Das bedeutet, wenn:

$$Energy(j) = \sum_{k=1}^{160} H^2[-k-j]$$

ist, dann ist:

$$Energy(j+1) = \sum_{k=1}^{160} H^2[-k-j-1] = Energy(j) + H^2[-j-161] - H^2[-j-1]$$

**[0065]** Somit werden nur 2 Mehrfache und 2 Additionen benötigt, um das Energie-Glied an jedem Schritt von dem FEC-Prozess zu aktualisieren, nachdem das erste Energie-Glied errechnet wurde.

**[0066]** Nun nachdem uns die Tonlagen-Abschätzung P vorliegt, beginnt die Erzeugung der Wellenform während der Löschung. Zurückkehrend zu dem Flussdiagramm in der [Fig. 7](#) werden in dem Schritt **710** die neuesten 3,25 Wellenlängen (3,25-P Abtastungen) aus H, dem History-Puffer **240**, in den Tonlagen-Puffer B kopiert. Die Inhalte von dem Tonlagen-Puffer, ausgenommen die neueste ¼-Wellenlänge, verbleiben für die Dauer der Löschung konstant. Andererseits wird die Aktualisierung des History-Puffers **240** mit der synthetischen Sprache während der Löschung fortgesetzt.

**[0067]** In dem Schritt **715** wird die neueste (¼-Wellenlänge (0,25-P Abtastungen) aus dem History-Puffer **240** in dem letzten ¼-Puffer L gespeichert. Diese ¼-Wellenlänge wird für mehrere der OLA-Arbeitsabläufe benötigt. Aus Gründen der Bequemlichkeit werden wir das gleiche negative Indexierungsschema wie für den History-Puffer **240** verwenden, um auf die B- und L-Puffer zuzugreifen. B [-1] ist die letzte Abtastung, bevor die Löschung eintritt, B [-2] ist die Abtastung davor, usw. Die synthetische Sprache wird in dem synthetischen Puffer

S platziert, der von 0 aufwärts indexiert wird. Somit ist S[0] die erste synthetisierte Abtastung, S[1] ist die zweite, usw.

**[0068]** Der Inhalt des Tonlagen-Puffers B und des letzten  $\frac{1}{4}$ -Puffers L für die Löschung in der [Fig. 3](#) sind in der [Fig. 10](#) gezeigt. In dem vorherigen Abschnitt haben wir die Periode P zu 56 Abtastungen berechnet. Der Tonlagen-Puffer ist somit  $3,25 \cdot 56 = 182$  Abtastungen lang. Der letzte  $\frac{1}{4}$ -Puffer ist  $0,25 \cdot 56 = 14$  Abtastungen lang. In der Figur sind senkrechte Linien jede P Abtastung zurück von dem Anfang der Löschung platziert worden.

**[0069]** Während den ersten 10 ms einer Löschung wird nur die letzte Tonlagen-Periode aus dem Tonlagen-Puffer verwendet, also ist in dem Schritt **720**  $U = 1$ . Wenn das Sprachsignal tatsächlich periodisch war und unsere Tonlagen-Abschätzung keine Abschätzung war, sondern der genaue tatsächliche Wert, dann könnten wir einfach die Wellenform direkt aus dem Tonlagen-Puffer B in den synthetischen Puffer S kopieren und das synthetische Signal wäre glatt und kontinuierlich. Das bedeutet  $S[0] = B[-P]$ ,  $S[1] = B[-P + 1]$ , usw. Wenn der Tonlagen kürzer als das 10 ms-Frame ist, das heißt  $P < 80$ , dann wird die einzelne Tonlagen-Periode in dem gelöschten Frame mehr als einmal wiederholt. In unserem Beispiel mit  $P = 56$  kippt also das Kopieren bei S[56]. Die Kopier-Sequenz mit Abtastung-für-Abtastung nahe der Abtastung **56** würde sein:  $S[54] = B[-2]$ ,  $S[55] = B[-1]$ ,  $S[56] = B[-56]$ ,  $S[57] = B[-55]$ , usw.

**[0070]** In der Praxis ist die Tonlagen-Abschätzung nicht genau und das Signal kann tatsächlich nicht periodisch sein. Um Unstetigkeiten (a) an der Grenze zwischen dem realen und dem synthetischen Signal und (b) an der Grenze zu vermeiden, wo die Periode wiederholt wird, sind OLAs erforderlich. Für beide Grenzen erwünschen wir einen glatten Übergang von dem Ende der realen Sprache  $B[-1]$  zu der eine Periode zurück liegenden Sprache  $B[-P]$ . Deshalb kann das in dem Schritt **725** durch das Überlappung-Addieren (OLA) der  $\frac{1}{4}$ -Wellenlänge vor  $B[-P]$  mit der letzten  $\frac{1}{4}$ -Wellenlänge von dem History-Puffer **240** oder dem Inhalt von L ausgeführt werden. Graphisch ist das gleichbedeutend damit, die letzten  $\frac{1}{4}$ -Wellenlängen in dem Tonlagen-Puffer zu nehmen, ihn eine Wellenlänge nach rechts zu verschieben und eine OLA in dem überlappenden Bereich der  $\frac{1}{4}$ -Wellenlänge zu machen. In dem Schritt **730** wird das Ergebnis von dem OLA zu der letzten  $\frac{1}{4}$ -Wellenlänge in dem History-Puffer **240** kopiert. Um zusätzliche Perioden der synthetischen Wellenform zu erzeugen, wird der Tonlagen-Puffer um zusätzliche Wellenlängen verschoben und zusätzliche OLAs werden ausgeführt.

**[0071]** Die [Fig. 11](#) zeigt den OLA-Arbeitsablauf für die ersten 2 Iterationen. In dieser Figur ist die senkrechte Linie, die über sämtliche Wellenformen verläuft, der Anfang von der Löschung. Die kurzen senkrechten Linien sind Tonlagen-Marker und werden P Abtastungen von der Löschungsgrenze versetzt platziert. Es sollte beobachtet werden, dass der überlappende Bereich zwischen den Wellenformen "Tonlagen-Puffer" und "Um P nach rechts verschoben" genau den gleichen Abtastungen wie diejenigen in dem überlappenden Bereich zwischen "Um P nach rechts verschoben" und "Um 2P nach rechts verschoben" entsprechen. Deshalb muss die  $\frac{1}{4}$ -Wellenlänge-OLA nur einmal errechnet werden.

**[0072]** In dem Schritt **735** kann der Prozess für ein tatsächlich periodisches Signal, das die synthetische Wellenform erzeugt, durch Errechnen des OLA zuerst und Platzieren der Ergebnisse in der letzten  $\frac{1}{4}$ -Wellenlänge von dem Tonlagen-Puffer verwendet werden. Beginnend bei der Abtastung  $B[-P]$  kopiert man einfach die Abtastungen aus dem Tonlagen-Puffer in den synthetischen Puffer und dreht den Tonlagen-Puffer-Zeiger an den Anfang der Tonlagen-Periode zurück, wenn das Ende von dem Tonlagen-Puffer erreicht ist. Unter Verwendung dieser Technik kann eine synthetische Wellenform mit einer beliebigen Dauer erzeugt werden. Die Tonlagen-Periode zu der Linken des Löschanfangs in der Wellenform "Kombiniert mit OLAs" aus der [Fig. 11](#) entspricht dem aktualisierten Inhalt des Tonlagen-Puffers.

**[0073]** Die Wellenform "Kombiniert mit OLAs" zeigt, dass der 1-Periode-Tonlagen-Puffer ein periodisches Signal mit der Periode P ohne Unstetigkeiten erzeugt. Diese synthetische Sprache, die aus einer einzelnen Wellenlänge in dem History-Puffer **240** erzeugt wurde, wird dazu verwendet, um die ersten 10ms einer Löschung zu verdecken. Die Auswirkung des OLA kann durch Vergleichen der  $\frac{1}{4}$ -Wellenlänge unmittelbar vor dem Anfang der Löschung in den Wellenformen "Tonlagen-Puffer" und "Kombiniert mit OLAs" betrachtet werden. In dem Schritt **730** ersetzt diese  $\frac{1}{4}$ -Wellenlänge in der Wellenform "Kombiniert mit OLAs" ferner die letzte  $\frac{1}{4}$ -Wellenlänge in dem History-Puffer **240**.

**[0074]** Der OLA-Arbeitsablauf mit dreieckigen Fenstern kann auch mathematisch ausgedrückt werden. Zunächst bestimmen wir die Variable P4 zu  $\frac{1}{4}$  von der Tonlagen-Periode in den Abtastungen. Somit ist  $P4 = P \gg 2$ . In unserem Beispiel betrug P 56, also beträgt P4 14. Der OLA-Arbeitsablauf kann dann in dem Bereich  $1 \leq i \leq P4$  ausgedrückt werden als:

$$B[-i] = \frac{i}{P4} L[-i] + \left( \frac{P4 - i}{P4} \right) B[-i - P]$$

**[0075]** Das Ergebnis des OLA ersetzt sowohl die letzten  $\frac{1}{4}$ -Wellenlängen in dem History-Puffer **240** als auch in dem Tonlagen-Puffer. Durch Ersetzen des History-Puffers **240** wird der OLA-Übergang der  $\frac{1}{4}$ -Wellenlänge ausgegeben, wenn der History-Puffer **240** aktualisiert wurde, weil der History-Puffer **240** ferner die Ausgabe um 3,75 ms verzögert. Die ausgegebene Wellenform während der ersten 10 ms der Löschung kann in dem Bereich zwischen den ersten zwei gepunkteten Linien in der Wellenform "Verdeckt" aus der [Fig. 3](#) betrachtet werden.

**[0076]** In dem Schritt **740** am Ende der Erzeugung der synthetischen Sprache für das Frame wird die gegenwärtige Abweichung in den Tonlagen-Puffer als die Variable 0 gespeichert. Diese Abweichung gestattet es, dass die synthetische Wellenform in den nächsten Frame für ein OLA mit dem realen oder synthetischen Signal des nächsten Frames fortgesetzt wird. 0 gestattet es ferner, dass die passende synthetische Signalphase aufrechterhalten wird, wenn sich die Löschung über 10 ms hinaus erstreckt. In unserem Beispiel mit 80 Abtastung-Frames und  $P = 56$  beträgt am Anfang der Löschung die Abweichung  $-56$ . Nach 56 Abtastungen dreht sie sich zu  $-56$  zurück. Nach zusätzlichen  $80 - 56 = 24$  Abtastungen beträgt die Abweichung  $-56 + 24 = -32$ , also beträgt am Ende des ersten Frames  $0 - 32$ .

**[0077]** In dem Schritt **745**, nachdem der Synthese-Puffer von  $S[0]$  bis  $S[79]$  gefüllt worden ist, wird  $S$  dazu verwendet, um den History-Puffer **240** zu aktualisieren. In dem Schritt **750** addiert der History-Puffer **240** ferner die Verzögerung von 3,75 ms. Die Handhabung des History-Puffers **240** ist während gelöschten und nicht gelöschten Frames die gleiche. An diesem Punkt endet der Arbeitsablauf zur Verdeckung des ersten Frames in dem Schritt **535** aus der [Fig. 5](#) und der Prozess schreitet zu dem Schritt **540** in der [Fig. 5](#) fort.

**[0078]** Die Details darüber, wie das FEC-Modul **230** arbeitet, um später Frames von über 10 ms hinaus zu verdecken, wie in dem Schritt **550** der [Fig. 5](#) gezeigt ist, ist im Detail in der [Fig. 12](#) gezeigt. Die Technik, die verwendet wird, um das synthetische Signal während dem zweiten und später gelöschten Frames zu erzeugen, ist zu dem ersten gelöschten Frame relativ gleich, obwohl einige zusätzliche Arbeit erforderlich ist, um dem Signal irgendeine Variation zu addieren.

**[0079]** In dem Schritt **1205** bestimmt der Löschungscode, ob das zweite oder dritte Frame gelöscht ist. Während den zweiten und dritten gelöschten Frames wird die Anzahl von Tonlagen-Perioden erhöht, die aus dem Tonlagen-Puffer verwendet wird. Das führt mehr an Variation in dem Signal ein und hält die synthetisierte Ausgabe davon ab, zu harmonisch zu klingen. Wie im Fall von sämtlichen anderen Übergängen ist ein OLA erforderlich, um die Grenze zu glätten, wenn die Anzahl von Tonlagen-Perioden erhöht wird. Nach dem dritten Frame (30 ms an Löschung) wird der Tonlagen-Puffer konstant bei einer Länge von 3 Wellenlängen gehalten. Diese 3 Wellenlängen erzeugen die gesamte synthetische Sprache für die Dauer der Löschung. Somit wird die Verzweigung zur Linken der [Fig. 12](#) nur bei den zweiten und dritten gelöschten Frames eingenommen.

**[0080]** Als Nächstes erhöhen wir in dem Schritt **1210** die Anzahl von Wellenlängen, die in dem Tonlagen-Puffer verwendet wird. Das heißt, dass wir  $U = U + 1$  setzen.

**[0081]** Am Anfang von dem zweiten oder dritten gelöschten Frame wird in dem Schritt **1215** das synthetische Signal aus dem vorherigen Frame für eine zusätzliche  $\frac{1}{4}$ -Wellenlänge in den Anfang des gegenwärtigen Frames fortgesetzt. Zum Beispiel erscheint am Anfang des zweiten Frames das synthetisierte Signal in unserem Beispiel, wie in der [Fig. 13](#) gezeigt ist. Diese  $\frac{1}{4}$ -Wellenlänge wird einer Überlappung-Addition mit dem neuen synthetischen Signal unterzogen, das ältere Wellenlängen aus dem Tonlagen-Puffer verwendet.

**[0082]** Zu Beginn des zweiten gelöschten Frames wird die Anzahl von Wellenlängen auf 2 erhöht, womit  $U = 2$  ist. Wie der 1-Wellenlänge-Tonlagen-Puffer muss eine OLA an der Grenze ausgeführt werden, wo sich der 2-Wellenlängen-Tonlagen-Puffer wiederholen kann. Dieses Mal wird die  $\frac{1}{4}$ -Wellenlänge, die  $U$  Wellenlängen zurück von dem Ausläufer des Tonlagen-Puffers  $B$  endet, einer Überlappung-Addition mit dem Inhalt von dem letzten  $\frac{1}{4}$ -Puffer  $L$  in dem Schritt **1220** unterzogen. Dieser OLA-Operator kann in dem Bereich  $1 \leq i \leq P4$  ausgedrückt werden als:

$$B[-i] = \frac{i}{P4} L[-i] + \left( \frac{P4 - i}{P4} \right) B[-i - PU]$$

**[0083]** Der einzige Unterschied zu der vorherigen Version dieser Gleichung besteht darin, dass die Konstante

P, die dazu verwendet wurde, um B auf der rechten Seite zu indexieren, zu PU umgeformt worden ist. Die Erstellung des 2-Wellenlängen-Tonlagen-Puffers ist graphisch in der [Fig. 14](#) gezeigt.

**[0084]** Wie in der [Fig. 11](#) stellt der Bereich der Wellenform "Kombiniert mit OLAs" zur Linken von dem Lösungsbeginn den aktualisierten Inhalt des 2-Perioden-Tonlagen-Puffers dar. Die kurzen senkrechten Linien markieren die Tonlagen-Periode. Eine nähere Überprüfung der nachfolgenden Peaks in der Wellenform "Kombiniert mit OLAs" zeigt, dass sich die Peaks von den Peaks ein und zwei Wellenlängen zurück vor dem Beginn der Löschung abwechseln.

**[0085]** Am Anfang der synthetischen Ausgabe in dem zweiten Frame müssen wir das Signal aus dem neuen Tonlagen-Puffer mit der  $\frac{1}{4}$ -Wellenlänge mischen, die in der [Fig. 13](#) erzeugt wurde. Wir erwünschen, dass das synthetische Signal aus dem neuen Tonlagen-Puffer aus dem ältesten Teil des in Verwendung befindlichen Puffers kommen sollte. Aber wir müssen darauf achten, dass der neue Teil aus einem ähnlichen Teil der Wellenform stammt, oder wenn wir sie vermischen, hörbare Artefakte erstellt werden. Mit anderen Worten wollen wir die richtige Phase aufrechterhalten oder die Wellenformen können auslöschend interferieren, wenn wir sie vermischen.

**[0086]** Das wird in dem Schritt **1225** ([Fig. 12](#)) durch Subtrahieren von Perioden P von der Abweichung ausgeführt, die am Ende des vorherigen Frames 0 gespeichert wurde, solange bis sie zu der ältesten Wellenlänge in dem verwendeten Teil des Tonlagen-Puffers zeigt.

**[0087]** Zum Beispiel war in dem ersten gelöschten Frame der gültige Index für den Tonlagen-Puffer B von  $-1$  bis  $-P$ . Also muss das gespeicherte 0 aus dem ersten gelöschten Frame in diesem Bereich sein. In dem zweiten gelöschten Frame ist der gültige Bereich von  $-1$  bis  $-2P$ . Also ziehen wir P von 0 ab, solange sich 0 in dem Bereich von  $-2P \leq 0 < -P$  befindet. Oder, um allgemeiner zu sein, ziehen wir P von 0 ab, solange es sich in dem Bereich von  $-UP \leq 0 < -(U - 1)P$  befindet. In unserem Beispiel sind am Ende von dem ersten gelöschten Frame  $P = 56$  und  $0 = -32$ . Wir ziehen 56 von  $-32$  ab und erhalten  $-88$ . Somit stammt die erste Synthese-Abtastung in dem zweiten Frame aus  $B[-88]$ , die nächste aus  $B[-87]$ , usw.

**[0088]** Das OLA-Mischen der synthetischen Signale aus den 1- und 2-Perioden-Tonlagen-Puffern zu Beginn des zweiten gelöschten Frames ist in der [Fig. 15](#) gezeigt.

**[0089]** Es sollte beachtet werden, dass durch das Subtrahieren von P von 0 die richtige Wellenform-Phase aufrechterhalten wird und dass die Peaks des Signals in den Wellenformen "1P-Tonlagen-Puffer" und "2P-Tonlagen-Puffer" ausgerichtet werden. Die Wellenform "OLR-kombiniert" zeigt ferner einen glatten Übergang zwischen den verschiedenen Tonlagen-Puffern zu Beginn des zweiten gelöschten Frames. Ein weiterer Arbeitsablauf ist erforderlich, bevor das zweite Frame in der Wellenform "OLA-kombiniert" der [Fig. 15](#) ausgegeben werden kann.

**[0090]** In dem Schritt **1230** ([Fig. 12](#)) wird die neue Abweichung dazu verwendet, um die  $\frac{1}{4}$ -Wellenlänge aus dem Tonlagen-Puffer in einen temporären Puffer zu kopieren. In dem Schritt **1235** wird die  $\frac{1}{4}$ -Wellenlänge zu der Abweichung addiert. Dann wird in dem Schritt **1240** der temporäre Puffer einer Überlappung-Addition mit dem Anfang des Ausgabepuffers unterworfen, und das Ergebnis wird in der ersten  $\frac{1}{4}$ -Wellenlänge des Ausgabepuffers platziert.

**[0091]** In dem Schritt **1245** wird die Abweichung dann dazu verwendet, um den Rest des Signals in dem Ausgabepuffer zu erzeugen. Der Tonlagen-Puffer wird in den Ausgabepuffer für die Dauer des 10 ms-Frames kopiert. In dem Schritt **1250** wird die gegenwärtige Abweichung in den Tonlagen-Puffer als die Variable O gespeichert.

**[0092]** Während den zweiten oder später gelöschten Frames wird das synthetische Signal in dem Schritt **1255** mit einer linearen Ansteigung abgeschwächt. Man lässt das synthetische Signal stufenweise ausklingen, bis es bei über 60 ms auf 0 oder Geräuschlosigkeit gesetzt wird. Während die Löschung länger wird, wird es wahrscheinlicher, dass die verdeckte Sprache von dem tatsächlichen Signal abweicht. Das zu lange Halten von bestimmten Klangtypen kann zu unnatürlichen hörbaren Artefakten in der Ausgabe des Verdeckungsprozesses führen, sogar wenn sich der Klang in der Absonderung für eine kurze Zeitperiode natürlich anhört. Um diese Artefakte in dem synthetischen Signal zu vermeiden, wird ein langsames Ausklingen verwendet. Ein ähnlicher Arbeitsablauf wird in den Verdeckungsprozessen ausgeführt, die in sämtlichen standard Sprachcodierern zu finden sind, wie das G.723.1, G.728 und G.729.

**[0093]** Der FEC-Prozess schwächt das Signal mit 20 % pro 10 ms-Frame ab, wobei an dem zweiten Frame begonnen wird. Wenn der Synthese-Puffer S das synthetische Signal vor der Abschwächung enthält und F die Anzahl von nachfolgend gelöschten Frames (F = 1 für das erste gelöschte Frame, 2 für das zweite gelöschte Frame) ist, dann kann die Abschwächung in den Bereichen von  $0 \leq i \leq 79$  und von  $2 \leq F \leq 6$  ausgedrückt werden als:

$$S'[i] = [1 - .2(F - 2) - \frac{.2i}{80}]S[i]$$

**[0094]** Zum Beispiel bei den Abtastungen am Anfang von dem zweiten gelöschten Frame F = 2, sind also F - 2 = 0 und  $0,2/80 = 0,0025$ , sind also  $S'[0] = 1,0S[0]$ ,  $S'[1] = 0,99755[1]$ ,  $S'[2] = 0,9955[2]$  und  $S'[79] = 0,8025S[79]$ . Nach dem sechsten gelöschten Frame wird die Ausgabe einfach auf 0 gesetzt.

**[0095]** Nachdem das synthetische Signal in dem Schritt **1255** abgeschwächt wurde, wird es an den History-Puffer **240** in dem Schritt **1260** gegeben und die Ausgabe wird in dem Schritt **1265** um 3,75 ms verzögert. Der Abweichungs-Zeiger O wird ferner auf dessen Standort in dem Tonlagen-Puffer am Ende von dem zweiten Frame aktualisiert, so dass das synthetische Signal in dem nächsten Frame fortgesetzt werden kann. Der Prozess geht dann zu dem Schritt **540** zurück, um das nächste Frame zu erhalten.

**[0096]** Wenn die Löschung über zwei Frames hinaus andauert, dann ist die Verarbeitung an dem dritten Frame genau wie an dem zweiten Frame, mit der Ausnahme, dass die Anzahl von Perioden in dem Tonlagen-Puffer von 2 auf 3 anstatt von 1 auf 2 erhöht wird. Während unsere Beispiellöschung bei zwei Frames endet, ist der 3-Perioden-Tonlagen-Puffer, der an dem dritten Frame und darüber hinaus verwendet würde, in der [Fig. 17](#) gezeigt. Nach dem dritten Frame verbleibt die Anzahl von Perioden in dem Tonlagen-Puffer fest bei drei, so dass nur der Weg auf der rechten Seite der [Fig. 12](#) eingenommen wird. In diesem Fall wird der Abweichungs-Zeiger O einfach dazu verwendet, um den Tonlagen-Puffer in die synthetische Ausgabe zu kopieren, und Arbeitsabläufe der Überlappung-Additionen sind nicht erforderlich.

**[0097]** Der Arbeitsablauf von dem FEC-Modul **230** an dem ersten einwandfreien Frame nach einer Löschung ist in der [Fig. 16](#) detailliert gezeigt. Am Ende einer Löschung ist ein glatter Übergang zwischen der synthetischen Sprache, die während der Löschung erzeugt wurde, und der realen Sprache erforderlich. Wenn die Löschung nur einen Frame lang war, dann wird in dem Schritt **1610** die synthetische Sprache für eine  $\frac{1}{4}$ -Wellenlänge fortgesetzt und eine Überlappung-Addition wird mit der realen Sprache ausgeführt.

**[0098]** Wenn das FEC-Modul **230** bestimmt, dass die Löschung in dem Schritt **1620** länger als 10 ms war, dann sind Fehlanpassungen zwischen den synthetischen und den realen Signalen wahrscheinlicher, so dass in dem Schritt **1630** die Erzeugung der synthetischen Sprache fortgesetzt wird, und das OLA-Fenster wird um zusätzliche 4 ms pro gelöschtes Frame bis zu einem Maximum von 10 ms vergrößert. Wenn die Abschätzung von der Tonlage ein wenig daneben war oder wenn sich die Tonlage der realen Sprache während der Löschung verändert hat, dann erhöht sich mit der Länge der Löschung die Wahrscheinlichkeit einer Fehlanpassung der Phase zwischen den synthetischen und den realen Signalen. Längere OLA-Fenster zwingen das synthetische Signal zum Ausklingen und das reale Sprachsignal zum langsameren Einklingen. Wenn die Löschung länger als 10 ms war, dann ist es ferner erforderlich, die synthetische Sprache in dem Schritt **1640** abzuschwächen, bevor ein OLA ausgeführt werden kann, so dass es mit dem Niveau des Signals in dem vorherigen Frame übereinstimmt.

**[0099]** In dem Schritt **1650** wird ein OLA an dem Inhalt des Ausgabepuffers (synthetische Sprache) mit dem Beginn von dem neuen eingegebenen Frame ausgeführt. Der Beginn des Eingabe-Puffers wird mit dem Ergebnis des OLA ersetzt. Das OLA am Ende der Löschung kann für das obige Beispiel in der [Fig. 4](#) betrachtet werden. Die vollständige Ausgabe des Prozesses der Verdeckung für das obige Beispiel kann in der Wellenform "Verdeckt" der [Fig. 3](#) betrachtet werden.

**[0100]** In dem Schritt **1660** wird der History-Puffer mit dem Inhalt des Eingabe-Puffers aktualisiert. In dem Schritt **1670** wird die Ausgabe der Sprache um 3,75 ms verzögert und der Prozess kehrt zu dem Schritt **530** in der [Fig. 5](#) zurück, um das nächste Frame zu erhalten.

**[0101]** Mit einer kleinen Anpassung kann der FEC-Prozess auf andere Sprachcodierer angewendet werden, die Zustandsinformation zwischen Abtastungen oder Frames aufrechterhalten und keine Verdeckung bereitstellen, wie das G.726. Der FEC-Prozess wird genauso verwendet, wie er in dem vorherigen Abschnitt beschrieben wurde, um die synthetische Wellenform während der Löschung zu erzeugen. Man muss jedoch dar-

auf achten, dass die Verfolgung der synthetischen Sprache, die von dem FEC-Prozess erzeugt wurde, durch die internen Zustandsvariablen des Codierers sichergestellt wird. Ansonsten erscheinen Artefakte und Unstetigkeiten in der Ausgabe während sich der Decodierer unter Verwendung von dessen fehlerhaftem Zustand neu startet, nachdem die Löschung vorbei ist. Während das OLA-Fenster am Ende einer Löschung hilft, muss mehr unternommen werden.

**[0102]** Bessere Ergebnisse können, wie in der [Fig. 18](#) gezeigt ist, durch das Konvertieren des Decodierers **1820** in einen Codierer **1860** für die Dauer der Löschung erzielt werden, wobei die synthetisierte Ausgabe von dem FEC-Modul **1830** als die Eingabe des Codierers **1860** verwendet wird.

**[0103]** Auf diese Art und Weise verfolgen die Zustandsvariablen des Decodierers **1820** die verdeckte Sprache. Es sollte beachtet werden, dass einem typischen Codierer ungleich, der Codierer **1860** nur betrieben wird, um Zustandsinformation aufrechtzuerhalten, und dass dessen Ausgabe nicht verwendet wird. Somit können Abkürzungen genommen werden, um dessen Laufzeit-Komplexität bedeutend zu senken.

**[0104]** Wie oben ausgeführt wurde, gibt es viele Vorteile und Aspekte, die von der Erfindung bereitgestellt werden. Insbesondere wird, während eine Frame-Löschung fortschreitet, die Anzahl von Tonlagen-Perioden, die aus der Signal-History verwendet werden, um das synthetische Signal zu erzeugen, als eine Funktion der Zeit erhöht. Das verringert harmonische Artefakte auf langen Löschungen bedeutend. Obwohl die Tonlagen-Perioden nicht in ihrer ursprünglichen Reihenfolge wiedergegeben werden, klingt die Ausgabe noch natürlich.

**[0105]** Mit dem G.726 und anderen Codierern, die Zustandsinformation zwischen Abtastungen oder Frames aufrechterhalten, kann der Decodierer als ein Codierer auf der Ausgabe von der synthetisierten Ausgabe des Verdeckungs-Prozesses betrieben werden. Auf diese Art und Weise verfolgen die internen Zustandsvariablen des Decodierers die Ausgabe, wobei, nachdem die Löschung vorbei ist, Unstetigkeiten vermieden – oder zumindest verringert – werden, die durch fehlerhafte Zustandsinformation in dem Decodierer verursacht wurden. Weil die Ausgabe aus dem Codierer niemals verwendet wird (dessen einziger Zweck besteht darin, Zustandsinformation aufrechtzuerhalten), kann eine vereinfachte Version des Codierers mit niedriger Komplexität verwendet werden.

**[0106]** Die minimale Tonlagen-Periode, die in den beispielhaften Ausführungsformen (40 Abtastungen oder 200 Hz) gestattet ist, ist größer als die, welche wir als Grundfrequenz für manche weibliche oder kindliche Sprecher erwarten. Somit wird für Sprecher mit hoher Frequenz sogar zu Beginn der Löschung mehr als eine Tonlagen-Periode verwendet, um die synthetische Sprache zu erzeugen. Bei Sprechern mit hoher Grundfrequenz werden die Wellenformen öfter wiederholt. Die mehrfachen Tonlagen-Perioden in dem synthetischen Signal machen harmonische Artefakte unwahrscheinlicher. Diese Technik hilft ferner dabei, das Signal während stimmlosen Abschnitten von Sprache sowie in Bereichen mit einem schnellen Übergang, wie eine Unterbrechung, natürlich klingend zu halten.

**[0107]** Das OLA-Fenster am Ende des ersten einwandfreien Frames nach einer Löschung wächst mit der Länge der Löschung. Bei längeren Löschungen treten Phasenübereinstimmungen wahrscheinlicher auf, wenn das nächste einwandfreie Frame eintrifft. Das Strecken der OLA-Fenster als eine Funktion von der Löschungslänge verringert Störspitzen, die durch Fehlanpassungen der Phase auf einer langen Löschung hervorgerufen wurden, gestattet es aber noch, dass sich das Signal wieder schnell aufbaut, wenn die Löschung kurz ist.

**[0108]** Der FEC-Prozess der Erfindung verwendet ferner OLA-Fenster variabler Länge, die einen kleinen Bruchteil von der abgeschätzten Tonlage darstellen, eine  $\frac{1}{4}$ -Wellenlänge aufweisen und nicht mit den Tonlagen-Peaks ausgerichtet sind.

**[0109]** Der FEC-Prozess der Erfindung unterscheidet nicht zwischen stimmhafter und stimmloser Sprache. Anstatt dessen erbringt er gute Leistungen bei der Vervielfältigung von stimmloser Sprache aufgrund der folgenden zwei Attribute des Prozesses: (A) die minimale Fenster-Größe ist vernünftig groß gewählt, so dass sogar stimmlose Bereiche von Sprache über eine vernünftige Variation verfügen, und (B) die Länge von dem Tonlagen-Puffer wird gesteigert, während der Prozess fortschreitet, was wiederum sicherstellt, dass harmonische Artefakte nicht eingeführt werden. Es sollte beachtet werden, dass die Verwendung von großen Fenstern, um die unterschiedliche Handhabung von stimmhafter und stimmloser Sprache zu vermeiden, ferner in der wohl-bekannteren Zeitskalierungs-Technik WSOLA vorhanden ist.

**[0110]** Während das Addieren von der Verzögerung des Gestattens des OLA zu Beginn einer Löschung als

ein nicht erwünschter Aspekt von dem Prozess der Erfindung betrachtet werden kann, ist es erforderlich, dass ein glatter Übergang zwischen realen und synthetischen Signalen am Anfang der Löschung sichergestellt wird.

**[0111]** Während diese Erfindung in Verbindung mit den oben skizzierten spezifischen Ausführungsformen beschrieben worden ist, ist es ersichtlich, dass dem Fachmann viele Alternativen, Modifikationen und Veränderungen ersichtlich sein werden. Dementsprechend ist es vorgesehen, dass die bevorzugten Ausführungsformen der Erfindung, wie sie oben dargelegt wurden, beispielhaft und nicht beschränkend sein sollen. Verschiedene Veränderungen können vorgenommen werden, ohne von dem Schutzzumfang der Erfindung abzuweichen, wie er in dem folgenden Anspruch bestimmt wird.

### **Patentansprüche**

1. Ein Verfahren zum Verdecken der Auswirkung einer fehlenden Sprachinformation in einer erzeugten Sprache, wobei die Sprachinformation komprimiert und in Paketen an einen Empfänger gesendet worden ist, der ein oder mehrere derartige Pakete nicht empfängt, wobei das Verfahren die folgenden Schritte umfasst: das Bilden eines Sprachsignals basierend auf zuvor empfangenen Paketen, die eine Sprachinformation darstellen;

als Antwort auf eine Bestimmung, dass ein Paket bei dem Empfänger nicht verfügbar ist, um das Sprachsignal zu bilden, das Synthetisieren eines Teils des Sprachsignals entsprechend dem nicht verfügbaren Paket unter Verwendung eines Teils des zuvor gebildeten Sprachsignals durch Abschätzen der Tonlage an der Basis der neuesten Sprach-Vorgeschichte und das Wiederholen eines Signalteils entsprechend einer Anzahl von Tonlagen-Perioden des zuvor gebildeten Sprachsignals und das Ausführen von Überlappung-Additionen an den Grenzen des Signalteils,

worin die Anzahl von Tonlagen-Perioden des zuvor gebildeten Teils, die in einer derartigen Synthese verwendet wurde, für Sprache einer Grundfrequenz über einem Grenzwert größer als für Sprache einer Grundfrequenz unter dem Grenzwert ist.

Es folgen 14 Blatt Zeichnungen

Anhängende Zeichnungen

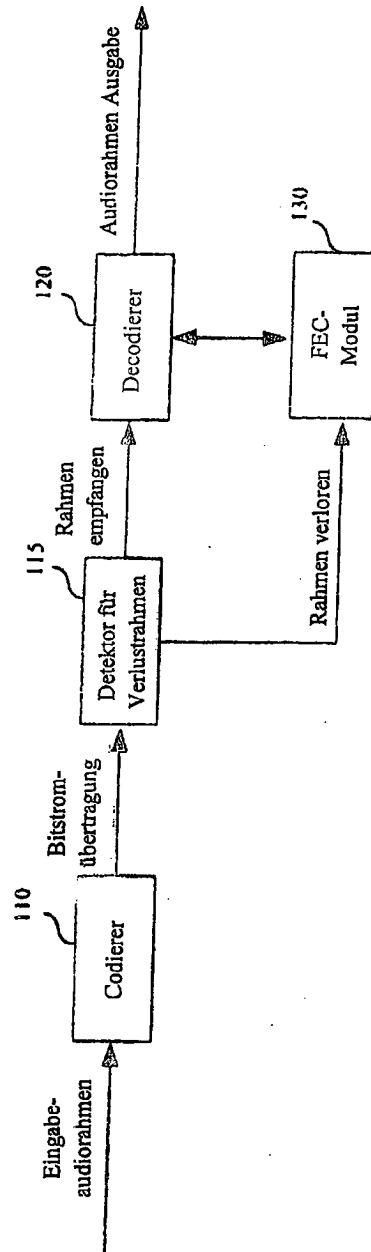


Fig. 1

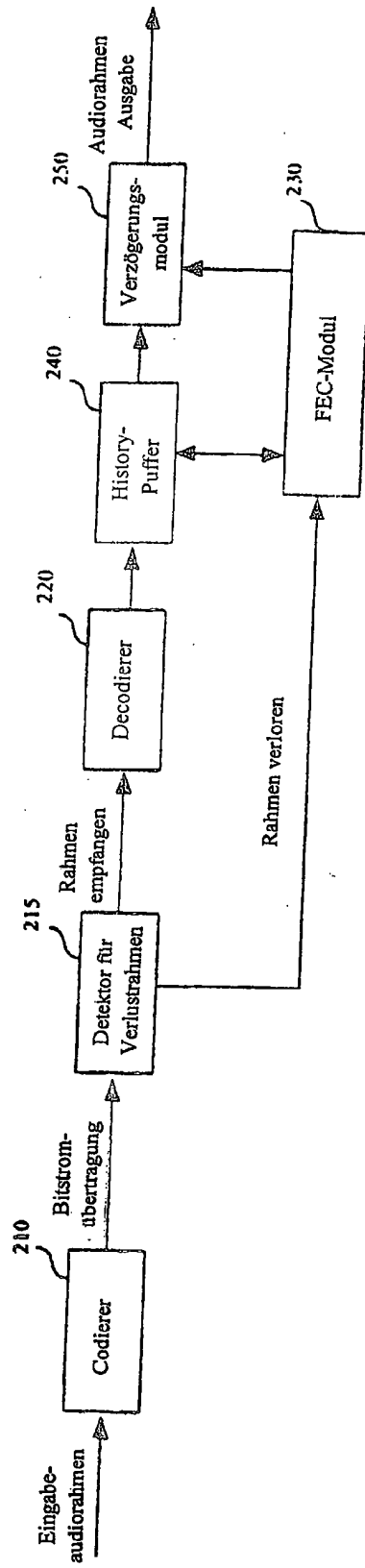


Fig. 2

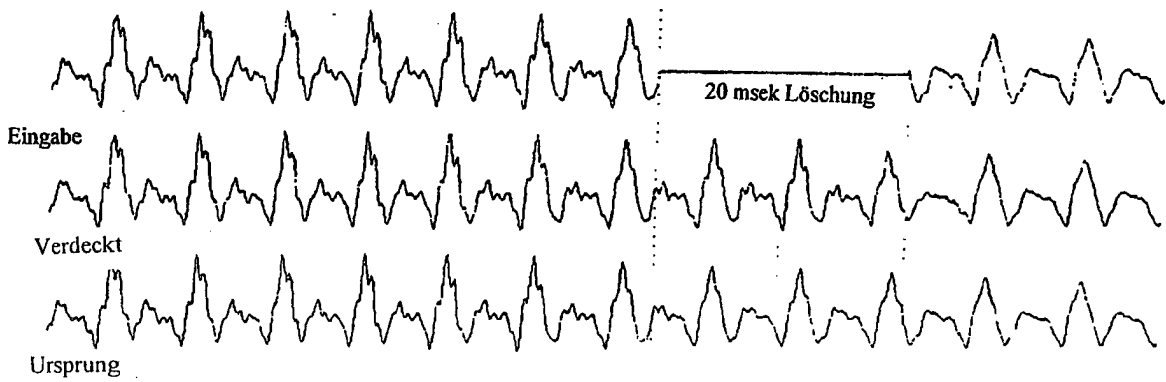


Fig. 3

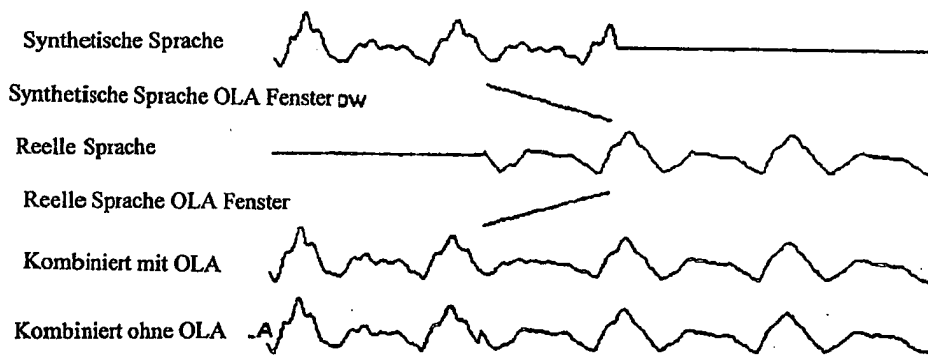


Fig. 4

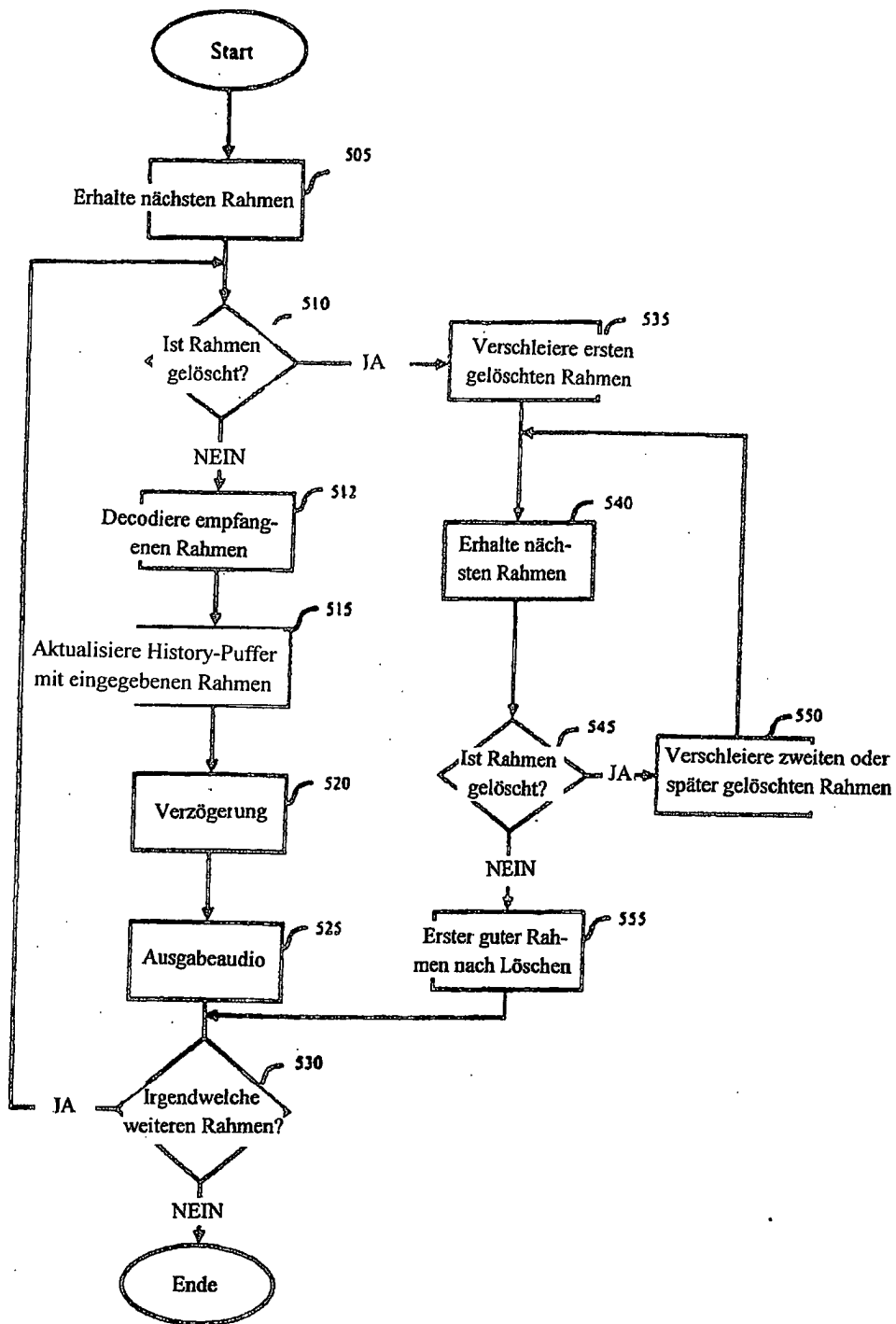


Fig. 5

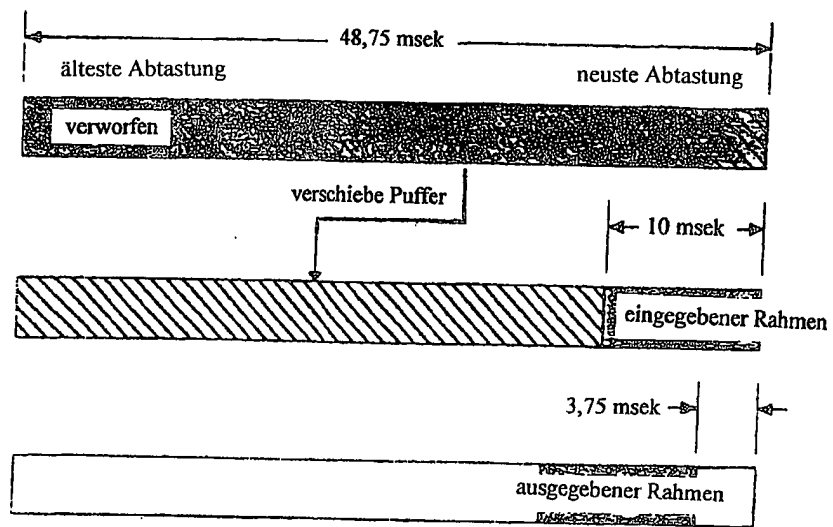


Fig. 6

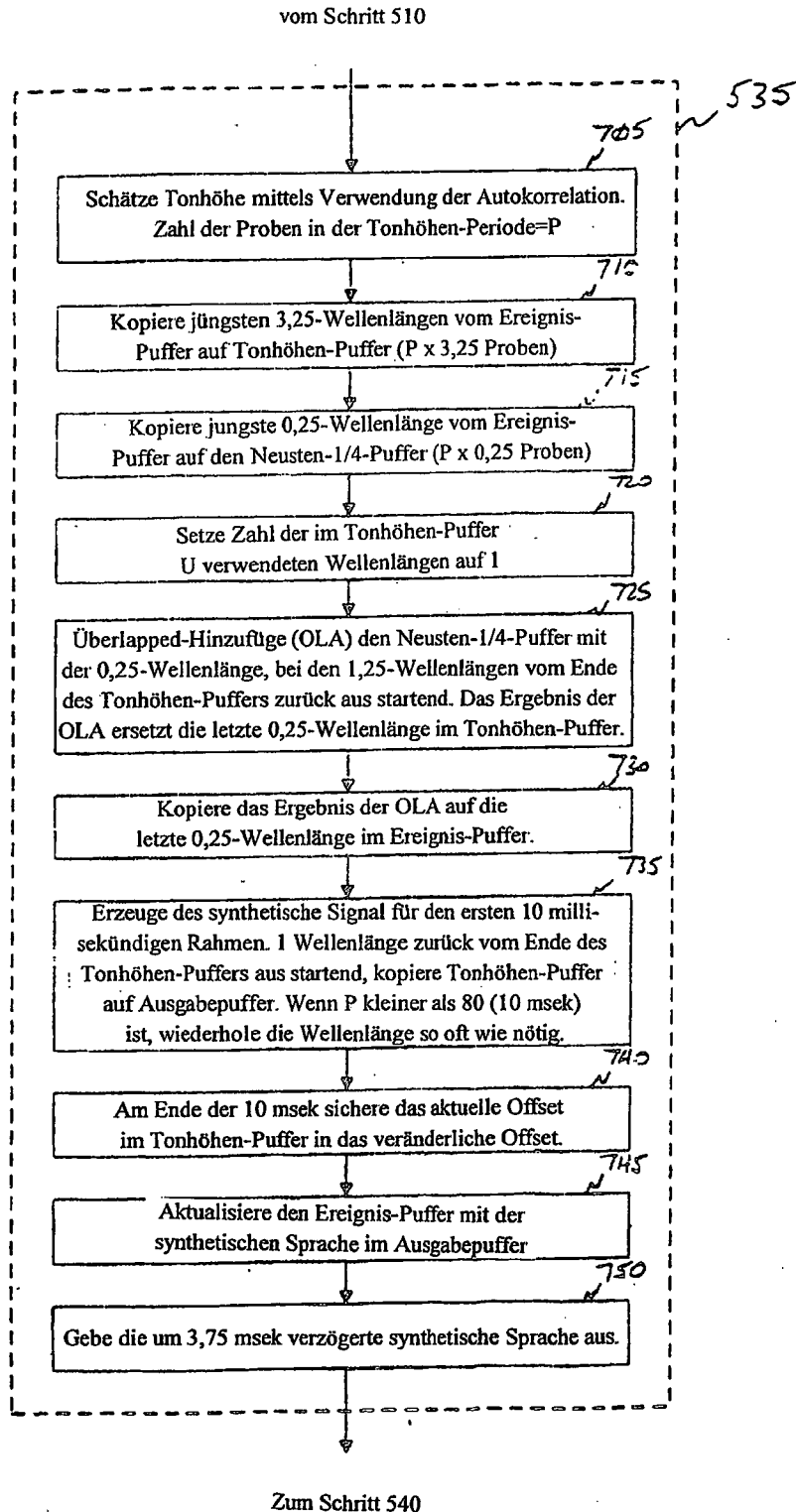


Fig. 7

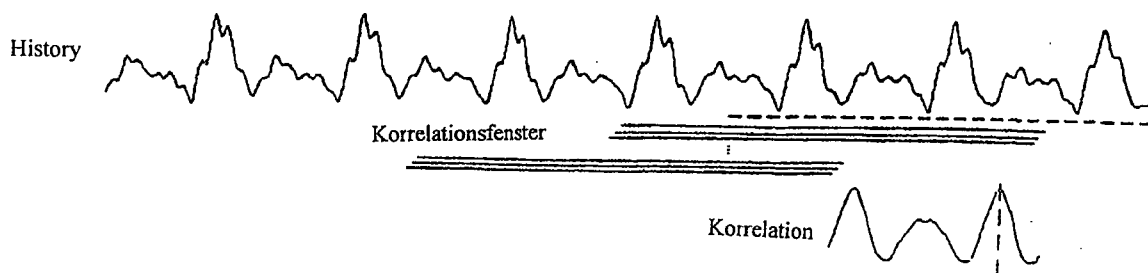


Fig. 8

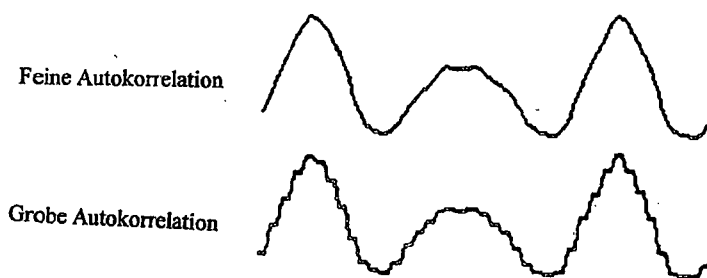


Fig. 9



Fig. 10

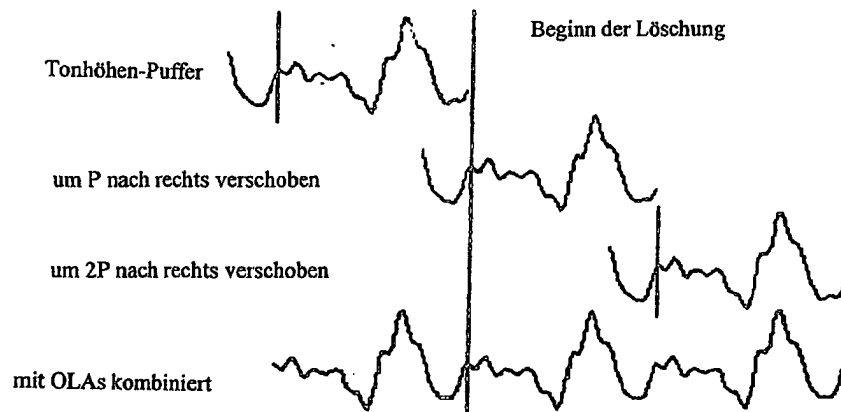


Fig. 11

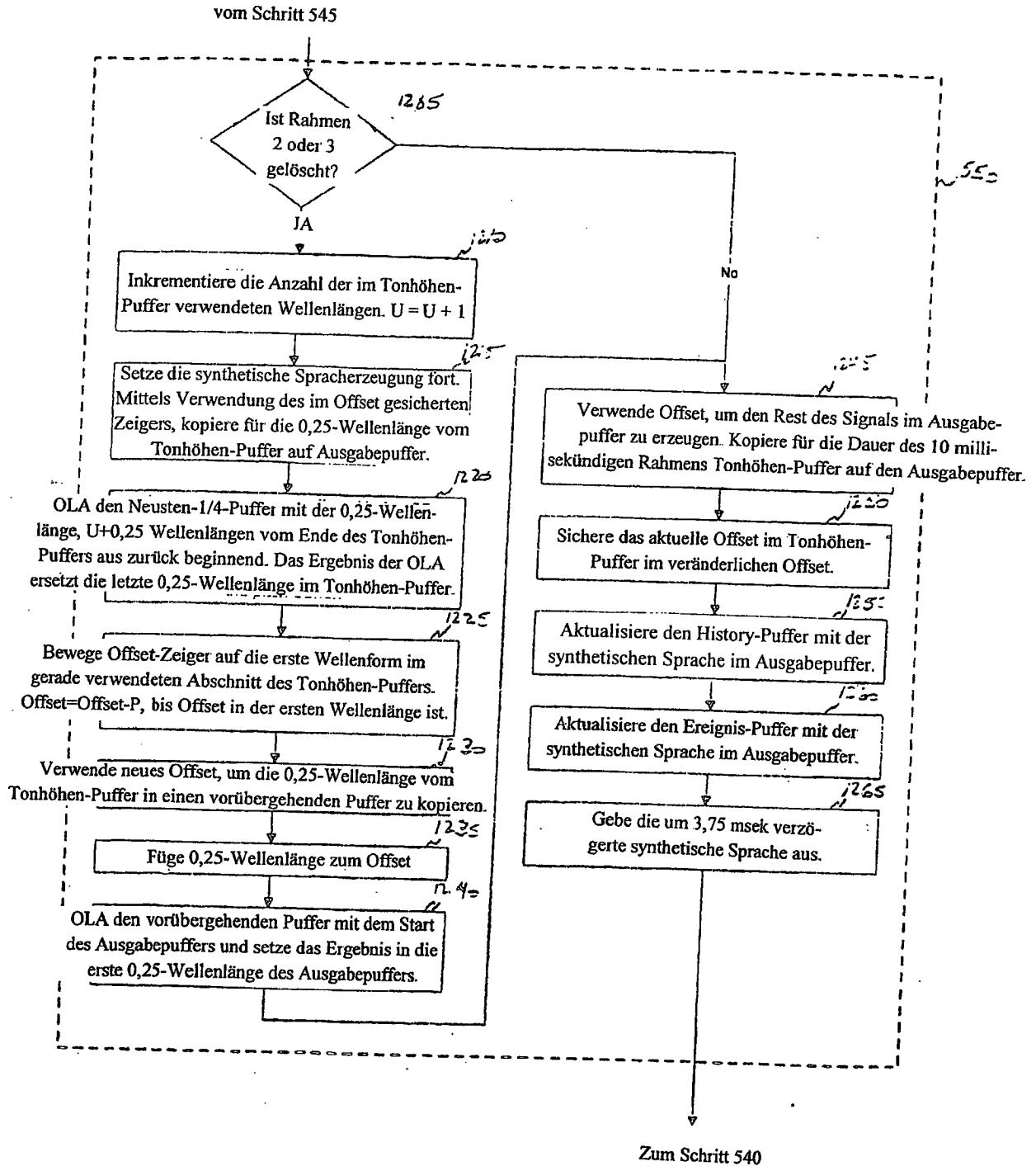


Fig. 12

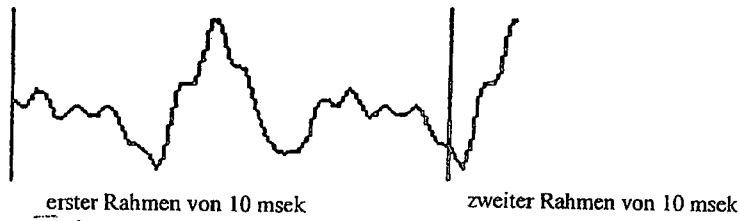


Fig. 13

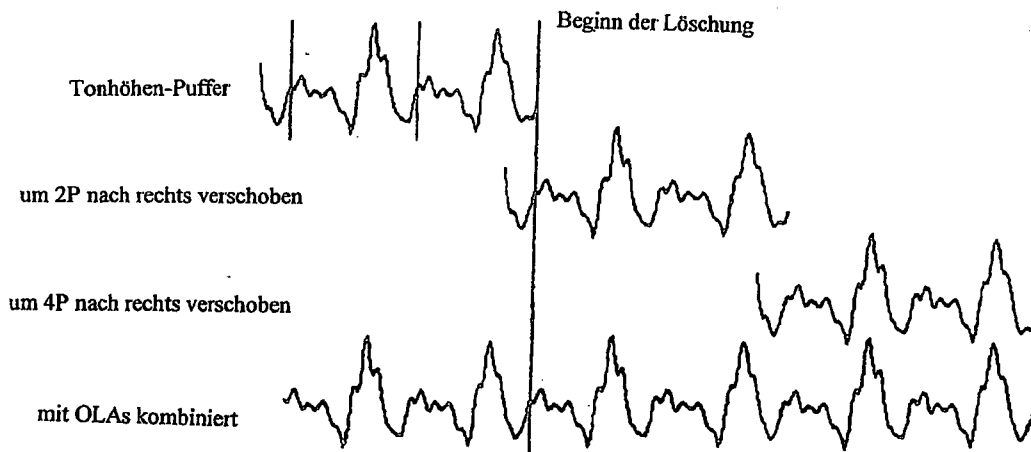


Fig. 14

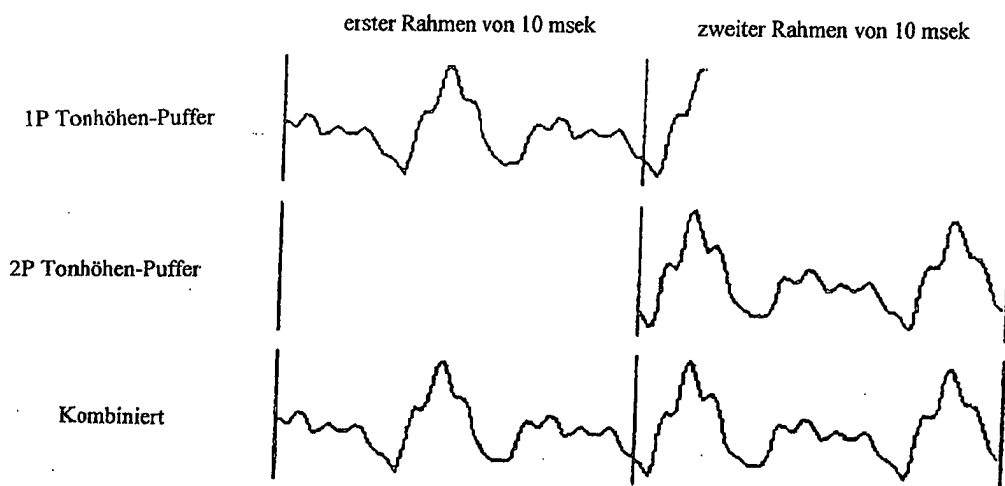


Fig. 15

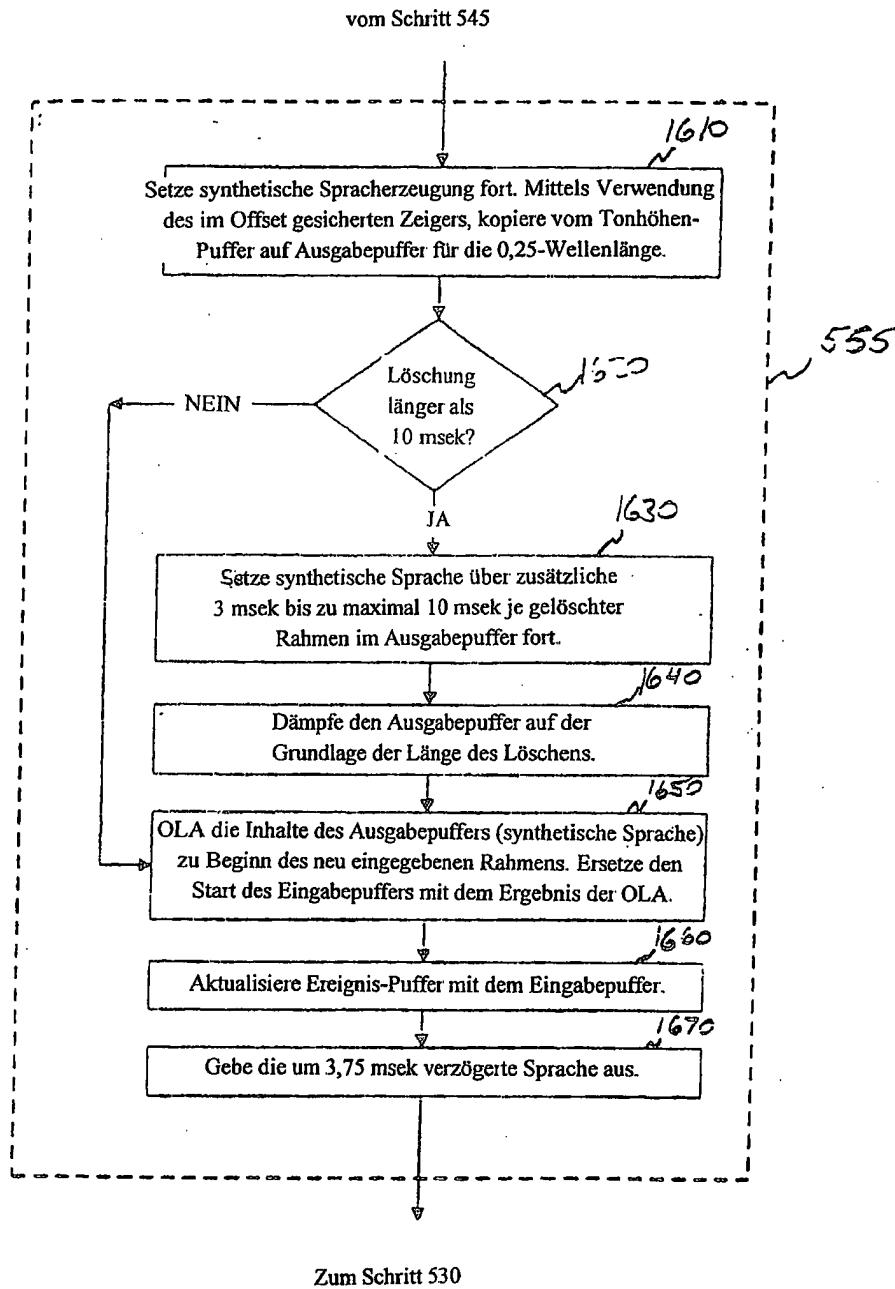


Fig. 16

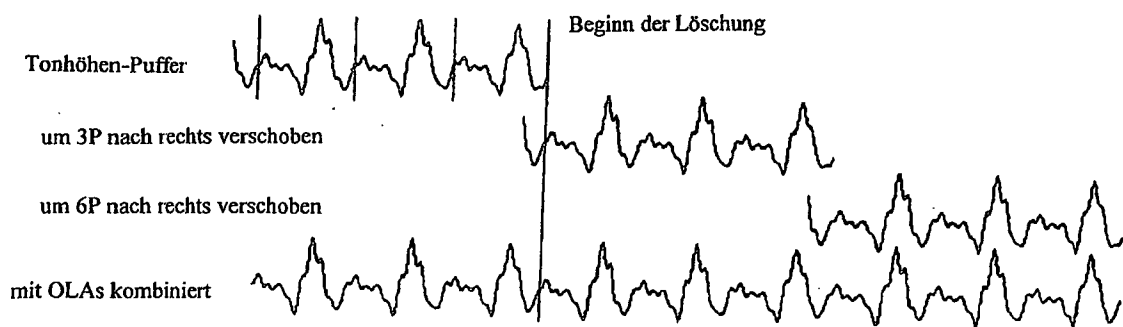


Fig. 17

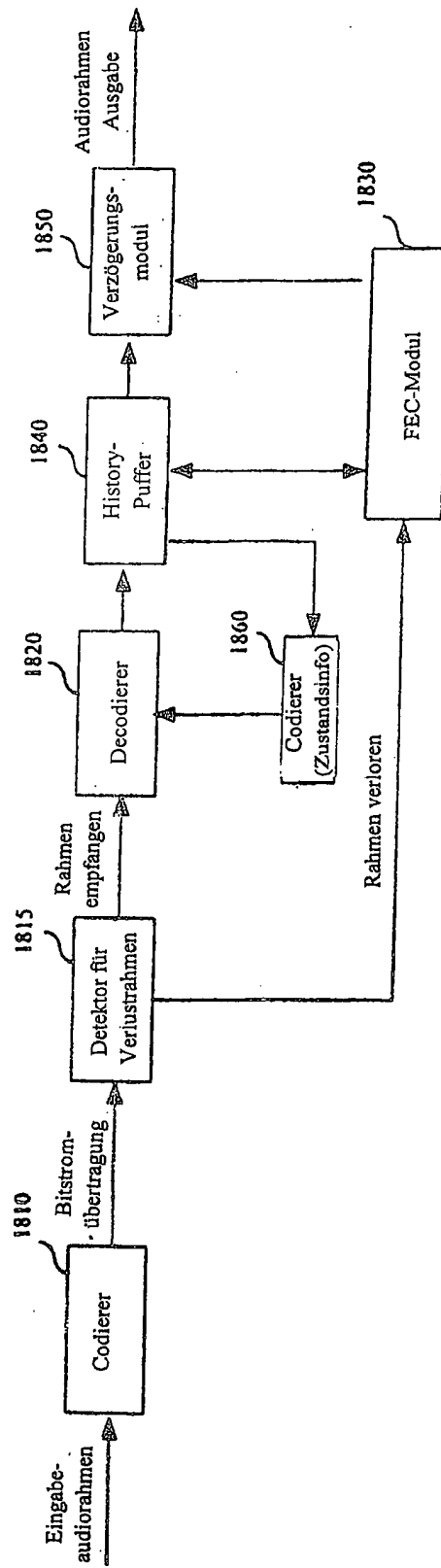


Fig. 18