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

(10) **DE 600 16 532 T2** 2005.10.13

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 1 086 451 B1**
(21) Deutsches Aktenzeichen: **600 16 532.9**
(86) PCT-Aktenzeichen: **PCT/US00/10577**
(96) Europäisches Aktenzeichen: **00 923 528.4**
(87) PCT-Veröffentlichungs-Nr.: **WO 00/63884**
(86) PCT-Anmeldetag: **19.04.2000**
(87) Veröffentlichungstag der PCT-Anmeldung: **26.10.2000**
(97) Erstveröffentlichung durch das EPA: **28.03.2001**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **08.12.2004**
(47) Veröffentlichungstag im Patentblatt: **13.10.2005**

(51) Int Cl.[7]: **G10L 19/00**

(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 RAHMENAUSFALL**

**Beschreibung**

## HINTERGRUND DER ERFINDUNG

### 1. Erfindungsgebiet

**[0001]** Diese Erfindung betrifft Techniken zum Durchführen eines Paketverlusts oder einer Verdeckung des Rahmenausfalls (Frame Erasure Concealment, FEC).

### 2. Beschreibung des Standes der Technik

**[0002]** Die Algorithmen zur Verdeckung des Rahmenausfalls (FEC) verbergen Übertragungsverluste in einem Sprachkommunikationssystem, worin ein Eingabesprachsignal codiert und in einem Sender paketiert wird, über ein Netz (beliebiger Art) übertragen wird und von einem Empfänger empfangen wird, der das Paket decodiert und die Sprachausgabe abspielt. Viele der Standard-, auf CELP basierenden Sprachcodierer, wie beispielsweise der G.723.1, G.728 und G.729, haben eingebaute oder in ihren Standards vorgeschlagene FEC-Algorithmen.

**[0003]** Das Ziel der FEC ist es, ein synthetisches Sprachsignal zu erzeugen, um fehlende Daten in einem empfangenen Bitstrom zu verbergen. Im Idealfall wird das synthetisierte Signal dieselbe Tonqualität und dieselben Spektralmerkmale wie das fehlende Signal haben und keine unnatürlichen Artefakte erzeugen. Da Sprachsignale häufig lokal stationär sind, ist es möglich, die Vorgeschichte der Signale zu verwenden, um eine vernünftige Annäherung an das fehlende Segment zu erzeugen. Wenn die Ausfälle nicht zu lang sind und der Ausfall nicht in einen Bereich fällt, in dem sich das Signal schnell ändert, können die Ausfälle nach der Verdeckung unhörbar sein.

**[0004]** Frühere Systeme benutzten Tonhöhen-Wellenform-Wiederhol-Techniken, um Rahmenausfälle zu verdecken), wie z. B. D. J. Goodman u. a., Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications, Band 34, Nr. 6 IEEE Trans. on Acoustics, Speech, and Signal Processing 1440–48 (Dezember 1996), und O. J. Wasem u. a., The Effect of Waveform Substitution on the Quality of PCM Packet Communications, Band 36, Nr. 3 IEEE Transactions on Acoustics, Speech, and Signal Processing 342–48 (März 1988).

**[0005]** Obwohl die Tonhöhen-Wellenformwiederholungs- und Überlappungs-Additions-Techniken verwendet wurden, um Signale zu synthetisieren, damit verlorene Rahmen von Sprachdaten verdeckt werden, führen diese Techniken manchmal zu piepsenden Artefakten, die den Zuhörer nicht zufriedenstellen.

## ZUSAMMENFASSUNG DER ERFINDUNG

**[0006]** Die vorliegende Erfindung, wie im Anspruch 1 dargelegt, betrifft ein Verfahren zum Verdecken der Auswirkung fehlender Sprachinformation auf Sprache, die von einem Decoder eines Sprachcodierungssystems erzeugt wird. Speziell ist es der Gegenstand der Erfindung, wie Sprache, die synthetisiert wird, um ein nicht verfügbares Paket (z. B. eine fehlende Sprachinformation infolge eines Rahmenausfalls) zu verdecken, reibungslos mit Sprache kombiniert wird, die anschließend aus empfangener Sprachinformation decodiert wird, nachdem ein Ausfall vorüber ist. Diese Kombination wird mittels Verwendung einer Überlappungs-Additions-Technik durchgeführt, die das synthetisierte Signal mit dem decodierten Signal mischt. Anders als im Stand der Technik ist jedoch die genaue Art des Überlappungs-Additionsfensters eine Funktion der Ausfall-Länge. So werden z. B. für längere Ausfälle längere Überlappungs-Additionsfenster verwendet.

## KURZE BESCHREIBUNG DER ZEICHNUNGEN

**[0007]** Die Erfindung wird detailliert mit Bezug auf die folgenden Figuren beschrieben, wo llrin gleiche Zahlen auf die gleichen Elemente Bezug nehmen und worin:

**[0008]** Fig. 1 ein exemplarisches Audioübertragungssystem ist;

**[0009]** Fig. 2 ein exemplarisches Audioübertragungssystem mit einem G.711-Codierer und FEC-Modul ist;

**[0010]** Fig. 3 ein Ausgabeaudiosignal darstellt, das eine FEC-Technik verwendet;

**[0011]** Fig. 4 eine Überlappungs-Additions(overlap-add, OLA)-Operation am Ende eines Ausfalls darstellt;

**[0012]** Fig. 5 ein Flussdiagramm eines exemplarischen Verfahrens zum Durchführen des FEC mittels Verwendung eines G.711-Codierers ist;

**[0013]** Fig. 6 ein Graph ist, der das Aktualisierungsverfahren des Vorgeschichten-Puffers darstellt;

**[0014]** Fig. 7 ein Flussdiagramm eines exemplarischen Verfahrens ist, um den ersten Rahmen des Signals zu verdecken;

**[0015]** Fig. 8 den Tonhöhen-Schätzwert aus der Autokorrelation darstellt;

**[0016]** Fig. 9 feine vs. grobe Tonhöhen-Schätzwerte darstellt;

**[0017]** Fig. 10 Signale im Tonhöhen-Puffer und im Puffer des letzten Viertels darstellt;

**[0018]** Fig. 11 eine synthetische Signalerzeugung darstellt, die einen Einperioden-Tonhöhen-Puffer verwendet;

**[0019]** Fig. 12 ein Flussdiagramm eines exemplarischen Verfahrens ist, um den als zweiten oder später gelöschten Rahmen des Signals zu verdecken;

**[0020]** Fig. 13 synthetisierte Signale darstellt, die im als zweiten gelöschten Rahmen fortgesetzt werden;

**[0021]** Fig. 14 eine synthetische Signalerzeugung mittels Verwendung eines Zweiperioden-Tonhöhen-Puffers darstellt;

**[0022]** Fig. 15 eine OLA zu Beginn des als zweiten gelöschten Rahmens darstellt;

**[0023]** Fig. 16 ein Flussdiagramm eines exemplarischen Verfahrens zur Verarbeitung des ersten Rahmens nach dem Löschen ist;

**[0024]** Fig. 17 eine synthetische Signalerzeugung mittels Verwendung eines Dreiperioden-Tonhöhen-Puffers darstellt; und

**[0025]** Fig. 18 ein Blockdiagramm ist, das die Verwendung von FEC-Techniken mit anderen Sprachcodierern darstellt.

## DETAILLIERTE BESCHREIBUNG DER BEVORZUGTEN AUSFÜHRUNGSFORMEN

**[0026]** In jüngster Zeit gab es großes Interesse an der Verwendung des G.711 bei Paketnetzen ohne garantierte Dienstqualität zur Unterstützung des konventionellen Fernsprechdienstes (Plain-Old-Telephony Service, POTS). Wenn in diesen Netzen Rahmenausfälle (oder Paketverluste) auftreten, werden Verdeckungstechniken gebraucht, oder die Qualität des Anrufs verschlechtert sich ernsthaft. Eine hochqualitative, wenig komplexe Rahmenausfall-Verdeckungstechnik (Frame Erasure Concealment, FEC) wurde entwickelt und wird unten detailliert beschrieben.

**[0027]** Ein exemplarisches Blockdiagramm eines Audiosystems mit FEC wird in Fig. 1 gezeigt. In Fig. 1 empfängt ein Codierer **110** einen Eingabe-Audiorahmen und gibt einen codierten Bitstrom aus. Der Bitstrom wird vom Detektor für verlorene Rahmen **115** empfangen, der bestimmt, ob irgendwelche Rahmen verloren gingen. Wenn der Detektor für verlorene Rahmen **115** erkennt, dass Rahmen verloren gegangen sind, signalisiert der Detektor für verlorene Rahmen **115** dem FEC-Modul **130**, einen FEC-Algorithmus oder -Prozess anzuwenden, um die fehlenden Rahmen zu rekonstruieren.

**[0028]** Solchermaßen verbirgt das FEC-Verfahren die Übertragungsverluste in einem Audiosystem, worin das Eingabesignal codiert und an einem Sender paketiert wird, über ein Netz übertragen und an einem Detektor für verlorene Rahmen **115** empfangen wird, der ermittelt, dass ein Rahmen verloren ging. In Fig. 1 wird davon ausgegangen, dass der Detektor für verlorene Rahmen **115** eine Möglichkeit hat, zu ermitteln, ob ein erwarteter Rahmen nicht angekommen oder zu spät angekommen ist, um verwendet zu werden. In IP-Netzen wird dies normalerweise implementiert, indem zu den Daten im übertragenen Rahmen eine Folgenummer oder ein Zeit-

stempel hinzugefügt wird. Der Detektor für verlorene Rahmen **115** vergleicht die Folgenummern der ankommenden Rahmen mit den Folgenummern, die erwartet würden, wenn keine Rahmen verloren gingen. Wenn der Detektor für verlorene Rahmen **115** erfasst, dass ein Rahmen wie erwartet angekommen ist, wird er vom Decoder **120** decodiert, und dem Ausgabesystem wird der Audio-Ausgaberahmen gegeben. Wenn ein Rahmen verloren geht, wendet das FEC-Modul **130** ein Verfahren an, um den fehlenden Audiorahmen zu verbergen, indem stattdessen eine synthetische Audio-Entsprechung des Rahmens erzeugt wird.

**[0029]** Viele Sprachcodierer, denen der ITU-T-CELP-Standard zugrunde liegt, wie beispielsweise der G.723.1, G.728 und der G.729, modellieren die Sprachreproduktion in ihren Decodern. Somit verfügen die Decoder über genug Zustandsinformationen, um das FEC-Verfahren direkt im Decoder zu integrieren. Diese Sprachcodierer haben FEC-Algorithmen oder -Verfahren, die als Teil ihrer Standards festgelegt sind.

**[0030]** Der G.711, zum Vergleich, ist ein Abtastwert-für-Abtastwert-Codierungsschema, das keine Sprachreproduktion entwickelt. Es gibt im Codierer keine Zustandsinformation, um bei der FEC zu helfen. Folglich ist das FEC-Verfahren mit G.711 vom Codierer unabhängig.

**[0031]** Ein exemplarisches Blockdiagramm des Systems, wie es mit dem G.711-Codierer verwendet wird, wird in Fig. 2 gezeigt. Wie in Fig. 1 codiert und überträgt der G.711-Codierer **210** die Bitstromdaten an den Detektor für verlorene Rahmen **215**. Auch hier vergleicht der Detektor für verlorene Rahmen **215** die Folgenummern der ankommenden Rahmen mit den Folgenummern, die erwartet würden, wenn keine Rahmen verloren gingen. Wenn ein Rahmen wie erwartet ankommt, wird er weitergeleitet, um vom Decoder **220** decodiert zu werden, und dann an einen Vorgeschichten-Puffer **240** ausgegeben, der das Signal speichert. Wenn ein Rahmen verloren geht, teilt der Detektor für verlorene Rahmen **215** dies dem FEC-Modul **230** mit, das ein Verfahren anwendet, um den fehlenden Audiorahmen zu verbergen, indem stattdessen eine synthetische Audio-Entsprechung des Rahmens erzeugt wird.

**[0032]** Um jedoch die fehlenden Rahmen zu verbergen, wendet das FEC-Modul **230** ein G.711 FEC-Verfahren an, das die bisherige Entwicklung des vom Vorgeschichten-Puffer **240** bereitgestellten decodierten Ausgabesignals verwendet, um zu schätzen, was das Signal im fehlenden Rahmen sein sollte. Um einen reibungslosen Übergang zwischen gelöschten und nicht gelöschten Rahmen zu gewährleisten, verzögert zusätzlich ein Verzögerungsmodul **250** die Ausgabe des Systems um eine vorbestimmte Zeitspanne, z. B. 3,75 msec. Diese Verzögerung ermöglicht, dass das synthetische Löschsignal zu Beginn eines Löschvorgangs langsam mit dem echten Ausgabesignal vermischt wird.

**[0033]** Die Pfeile zwischen dem FEC-Modul **230** und jeweils den Ereignispuffer- **240** und den Verzögerungsmodul-Blöcken **250** bedeuten, dass die gespeicherte Vorgeschichte (History) vom FEC-Verfahren verwendet wird, um das synthetische Signal zu erzeugen. Zusätzlich wird die Ausgabe des FEC-Moduls **230** verwendet, um den Vorgeschichten-Puffer **240** während eines Löschvorgangs zu aktualisieren. Es soll angemerkt werden, dass, da das FEC-Verfahren nur von der decodierten Ausgabe des G.711 abhängt, das Verfahren genauso gut arbeiten wird, wenn kein Sprachcodierer vorhanden ist.

**[0034]** Ein graphisches Beispiel dafür, wie das Eingabesignal vom FEC-Verfahren im FEC-Modul **230** verarbeitet wird, wird in Fig. 3 gezeigt.

**[0035]** Die obere Wellenform in der Figur zeigt die Eingabe in das System, wenn ein Ausfall von 20 msec in einem Bereich der gesprochenen Sprache von einem männlichen Sprecher erfolgt. In der Wellenform darunter hat das FEC-Verfahren die fehlenden Segmente durch das Erzeugen synthetischer Sprache in der Lücke verdeckt. Für Vergleichszwecke wird auch das ursprüngliche Eingabesignal ohne einen Ausfall gezeigt. In einem idealen System klingt die verdeckte Sprache genau wie das Original. Wie in der Figur zu sehen, ähnelt die synthetische Wellenform in den fehlenden Segmenten stark dem Original. Wie die "verdeckte" Wellenform aus der "Eingabe"-Wellenform erzeugt wird, wird unten detailliert erörtert.

**[0036]** Das vom FEC-Modul **230** verwendete FEC-Verfahren verdeckt den fehlenden Rahmen durch die Erzeugung einer synthetischen Sprache, die ähnliche Eigenschaften wie die im Vorgeschichten-Puffer **240** gespeicherte Sprache hat. Die Grundidee ist wie folgt: Wenn das Signal über Sprache mitgeteilt wird, nehmen wir an, dass das Signal quasi-periodisch und lokal stationär ist. Wir werten die Tonhöhe aus und wiederholen ein paar mal die letzte Tonhöhen-Periode im Vorgeschichten-Puffer **240**. Wenn jedoch der Ausfall lang ist oder die Tonhöhe kurz ist (die Frequenz ist hoch), führt das zu häufige Wiederholen derselben Tonhöhen-Periode zu einer Ausgabe, die, verglichen mit der natürlichen Sprache, zu harmonisch ist. Um diese harmonischen Artefakte, die als Piepsen und Bongs hörbar sind, zu vermeiden, wird die Anzahl der aus dem Vorgeschich-

ten-Puffer **240** verwendeten Tonhöhen-Perioden erhöht, während die sich die Länge des Ausfalls erhöht. Kurze Ausfälle verwenden nur die letzte oder die letzten paar Tonhöhen-Perioden aus dem Vorgeschichten-Puffer **240**, um das synthetische Signal zu erzeugen. Lange Ausfälle verwenden auch Tonhöhen-Perioden, die im Vorgeschichten-Puffer **240** weiter zurückliegen. In Zusammenhang mit langen Ausfällen werden die Tonhöhen-Perioden aus dem Vorgeschichten-Puffer **240** nicht wieder in derselben Reihenfolge abgespielt, in der sie in der ursprünglichen Sprache auftraten. Jedoch wurde durch Tests festgestellt, dass das in langen Ausfällen erzeugte synthetische Sprachsignal dennoch einen natürlichen Klang erzeugt.

**[0037]** Je länger der Ausfall ist, desto wahrscheinlicher ist es, dass das synthetische Signal vom echten Signal abweichen wird. Um Artefakte zu vermeiden, die durch zu langes Halten bestimmter Klangtypen verursacht werden, wird das synthetische Signal gedämpft, wenn der Ausfall länger wird. Für Ausfälle mit einer Dauer von 10 msec oder weniger ist keine Dämpfung erforderlich. Für Ausfälle, die länger als 10 msec sind, wird das synthetische Signal mit einer Rate von 20% pro 10 zusätzliche Millisekunden gedämpft. Jenseits von 60 msec wird das synthetische Signal auf Null (Ruhe) gesetzt. Dies liegt daran, dass das synthetische Signal dem ursprünglichen Signal so wenig ähnelt, dass es im Durchschnitt eher schadet als nützt, mit dem Versuch fortzufahren, die fehlende Sprache nach 60 msec zu verdecken.

**[0038]** Wann immer ein Übergang zwischen Signalen aus unterschiedlichen Quellen durchgeführt wird, ist es wichtig, dass der Übergang keine Diskontinuitäten, die als Klicken hörbar sind, oder unnatürliche Artefakte in das Ausgabesignal einführt. Diese Übergänge erfolgen an mehreren Stellen:

1. Zu Beginn des Ausfalls am Grenzbereich zwischen dem Beginn des synthetischen Signals und dem Ende des letzten intakten Rahmens.
2. Am Ende des Ausfalls am Grenzbereich zwischen dem synthetischen Signal und dem Beginn des Signals im ersten intakten Rahmen nach dem Ausfall.
3. Immer dann, wenn die Anzahl der aus dem Vorgeschichten-Puffer **240** verwendeten Tonhöhen-Perioden geändert wird, um die Signalvariation zu erhöhen.
4. An den Grenzbereichen zwischen den wiederholten Abschnitten des Vorgeschichten-Puffers **240**. Um reibungslose Übergänge zu gewährleisten, werden an allen Signalgrenzbereichen Überlappungs-Additionen (OLA) durchgeführt.

**[0039]** OLAs sind ein Mittel zum reibungslosen Vereinigen zweier Signale, die sich an einem Rand überlagern. Im Bereich, wo sich die Signale überlagern, werden die Signale durch Fenster gewichtet und dann zusammengefügt (vermischt). Die Fenster sind so aufgebaut, dass die Summe der Wichtungen an jedem speziellen Abtastwert gleich 1 ist. Das bedeutet, dass auf die Gesamtsumme der Signale keine Verstärkung oder Dämpfung angelegt wird. Zusätzlich sind die Fenster so aufgebaut, dass das Signal links bei der Wichtung 1 einsetzt und graduell zur Wichtung 0 absinkt, während das Signal rechts an der Wichtung 0 einsetzt und graduell zur Wichtung 1 ansteigt. Somit liegt im Bereich links vom Überlappungsfenster nur das linke Signal vor, während im Bereich rechts vom Überlappungsfenster nur das rechte Signal vorliegt. Im Überlappungsbereich führt das Signal graduell einen Übergang vom linken Signal zum rechten durch. Im FEC-Verfahren werden dreieckige Fenster verwendet, um die Komplexität für die Berechung der Fenster mit veränderlicher Länge gering zu halten; es können stattdessen jedoch andere Fenster, wie beispielsweise Hanning-Fenster, verwendet werden.

**[0040]** Fig. 4 zeigt die synthetische Sprache am Ende eines 20-msec-Ausfalls, der mit der echten Sprache OLAed wird, die beginnt, nachdem der Ausfall vorüber ist. In diesem Beispiel ist das OLA-Wichtungsfenster ein dreieckiges Fenster von 5,75 msec. Das obere Signal ist das während des Ausfalls erzeugte synthetische Signal, und das überlappende Signal darunter ist die echte Sprache nach dem Ausfall. Die OLA-Wichtungsfenster werden unterhalb der Signale gezeigt. Infolge einer Tonhöhen-Änderung im echten Signal während des Ausfalls stimmen hier die Spitzen der synthetischen und echten Signale nicht überein, und die eingeführte Diskontinuität, – wenn wir versuchen, die Signale ohne eine OLA zu vereinigen – ist im Graph dargestellt, der mit "ohne OLA kombiniert" markiert ist. Der "ohne OLA kombiniert"-Graph wurde erstellt, indem das synthetische Signal bis zum Beginn des OLA-Fensters und das echte Signal für die Dauer kopiert wurden. Das Ergebnis der OLA-Operationen zeigt, wie die Diskontinuitäten an den Grenzbereichen geglättet werden.

**[0041]** Die obige Erörterung betrifft die Art, wie ein veranschaulichendes Verfahren mit einer stationären, über Stimme mitgeteilten Sprache arbeitet; wenn sich jedoch die Sprache schnell ändert oder nicht über Stimme mitgeteilt wird, kann es sein, dass die Sprache keine periodische Struktur hat. Jedoch werden diese Signale auf dieselbe Art und Weise verarbeitet, wie unten dargelegt.

**[0042]** Als erstes ist die kleinste Tonhöhen-Periode, die wir in der veranschaulichenden Ausführungsform in

der Tonhöhen-Schätzung zulassen, 5 msec, was der Frequenz von 200 Hz entspricht. Obwohl bekannt ist, dass manche Frauen und Kinder, die in Hochfrequenz sprechen, Grundfrequenzen von über 200 Hz haben, schränken wir sie auf 200 Hz ein, so dass die Fenster relativ groß bleiben. Auf diese Weise wird innerhalb eines ausgefallenen Rahmens von 10 msec die ausgewählte Tonhöhen-Periode maximal zweimal wiederholt. Bei Hochfrequenz-Sprechenden verschlechtert dies die Ausgabe nicht wirklich, da die Tonhöhen-Schätzfunktion ein Vielfaches der echten Tonhöhen-Periode zurückgibt. Und indem kein Sprechvorgang zu oft wiederholt wird, erzeugt das Verfahren keine synthetische, periodische Sprache aus nicht-periodischer Sprache. Da zweitens die Anzahl der zur Erzeugung der synthetischen Sprache verwendeten Tonhöhen-Perioden erhöht wird, wenn der Ausfall länger wird, wird dem Signal genügend Variation verliehen, so dass Periodizität für lange Ausfälle nicht eingeführt wird.

**[0043]** Es sollte angemerkt werden, dass das Wellenform-Ähnlichkeits-Überlappungs-Additions-(Waveform Similarity Overlap Add, WSOLA)-Verfahren für die Zeitmaßstabänderungen von Sprache ebenfalls große OLA-Fenster mit feststehender Größe verwendet, so dass dasselbe Verfahren sowohl für die Veränderung des Zeitmaßstabs periodischer als auch nicht-periodischer Sprachsignale angewandt werden kann.

**[0044]** Während oben ein Überblick über das veranschaulichende FEC-Verfahren gegeben wurde, werden die einzelnen Schritte unten detailliert erörtert.

**[0045]** Im Zusammenhang mit dieser Erörterung werden wir davon ausgehen, dass ein Rahmen 10 msec Sprache enthält und die Abtastfrequenz beispielsweie 8 kHz ist. Somit können Ausfälle in Inkrementen von 80 Abtastungen (8000 * 0,010 = 80) auftreten. Es ist anzumerken, dass das FEC-Verfahren leicht an andere Rahmengrößen und Abtastfrequenzen angepasst werden kann. Um die Abtastfrequenz zu ändern, müssen nur die in Millisekunden angegebenen Zeitperioden mit 0,001 und dann mit der Abtastfrequenz multipliziert werden, um die richtigen Puffergrößen zu erhalten. Zum Beispiel enthält der Vorgeschichten-Puffer **240** die letzten 48,75 msec Sprache. Bei 8 kHz würde dies implizieren, dass der Puffer (48,75 * 0,001 * 8000) = 390 Abtastungen lang ist. Bei einer 16 kHz-Abtastung würde er sich auf 780 Abtastungen verdoppeln.

**[0046]** Mehreren Puffergrößen liegt die niedrigste Frequenz zugrunde, die für das Verfahren erwartet werden kann. Zum Beispiel geht das veranschaulichende Verfahren davon aus, dass die geringste Frequenz, die bei der 8 kHz-Abtastung zu beobachten ist, 66 2/3 Hz ist. Dies führt zu einer maximalen Tonhöhen-Periode von 15 msec (1/(66 2/3) = 0,015). Die Länge des Vorgeschichten-Puffers **240** ist 3,25 mal die Periode der niedrigsten Frequenz. Somit ist der Vorgeschichten-Puffer **240** 15 * 3,25 = 48,75 msec. Wenn bei einer 16 kHz-Abtastung die Eingangsfilter Frequenzen gestatten, die nur 50 Hz (Periode von 20 Millisekunden) betragen, müßte der Ereignispuffer **240** auf 20 * 3,25 = 65 msec verlängert werden.

**[0047]** Die Rahmengröße kann ebenfalls verändert werden; 10 msec wurden als Vorgabe gewählt, da dies die Rahmengröße ist, die von mehreren Standard-Sprachcodierern wie beispielsweise dem G.729 verwendet wird und die auch in mehreren drahtlosen Systemen verwendet wird. Das Ändern der Rahmengröße ist unkompliziert. Wenn die gewünschte Rahmengröße ein Vielfaches von 10 msec ist, bleibt das Verfahren unverändert. Die Rahmengröße des Ausfallverfahrens sollte einfach bei 10 msec bleiben und mehrmals pro Rahmen aufgerufen werden. Wenn die gewünschte Paketrahmengröße ein Divisor von 10 msec ist, z. B. 5 msec, bleibt das FEC-Verfahren im wesentlichen unverändert. Jedoch wird die Rate, mit der die Anzahl von Perioden im Tonhöhen-Puffer erhöht wird, auf der Grundlage der Rahmenanzahl in 10 msec modifiziert werden müssen. Rahmengrößen, die nicht ein Vielfaches oder ein Divisor von 10 msec sind, z. B. 12 msec, können ebenfalls untergebracht werden. Das FEC-Verfahren toleriert in vernünftigem Maße die Änderung der Anstiegsrate in der Anzahl der vom Tonhöhen-Puffer verwendeten Tonhöhen-Perioden. Der Anstieg der Periodenanzahl alle 12 msec anstatt alle 10 msec wird keinen großen Unterschied ausmachen.

**[0048]** Fig. 5 ist ein Blockdiagramm des FEC-Verfahrens, das durch die veranschaulichende Ausführungsform in Fig. 2 durchgeführt wird. Die Unter-Schritte, die benötigt werden, um einige der Hauptoperationen zu implementieren, werden in den Fig. 7, Fig. 12 und Fig. 16 weiter detailliert dargelegt und unten erörtert. In der folgenden Erörterung werden mehrere Variablen verwendet, um Werte und Puffer zu halten. Diese Variablen sind unten zusammengefasst:

Tabelle 1. Variablen und ihre Inhalte

| Variable | Art | Beschreibung | Anmerkung |
|---|---|---|---|
| B | Array | Tonhöhen-Puffer | Bereich[-P*3,25:-1] |
| H | Array | Vorgeschichten-Puffer | Bereich[-390:-1] |
| L | Array | Puffer des letzten Viertels | Bereich[-P*.25:-1] |
| O | Skalar | Offset im Tonhöhen-Puffer | |
| P | Skalar | Tonhöhen-Schätzwert | 40<=P<120 |
| P4 | Skalar | ¼ Tonhöhen-Schätzwert | P4=P>>2 |
| S | Array | Synthetisierte Sprache | Bereich[0:79] |
| U | Skalar | Verwendete Wellenlängen | 1<=U<=3 |

**[0049]** Wie im Flussdiagramm in Fig. 5 zu sehen, beginnt das Verfahren, und im Schritt **505** wird der nächste Rahmen vom Detektor für verlorene Rahmen **215** empfangen. Im Schritt **510** bestimmt der Detektor für verlorene Rahmen **215**, ob der Rahmen ausgefallen ist. Wenn der Rahmen nicht ausgefallen ist, wird im Schritt **512** der Rahmen vom Decoder **220** decodiert. Dann wird im Schritt **515** der decodierte Rahmen zur Verwendung vom FEC-Modul 230 im Vorgeschichten-Puffer **240** gesichert.

**[0050]** Im Aktualisierungsschritt des Vorgeschichten-Puffers ist dieser Puffer **240** 3,25 mal so lang wie die erwartete längste Tonhöhen-Periode. Bei einer 8 kHz-Abtastung beträgt die längste Tonhöhen-Periode **15** msec oder 120 Abtastungen; somit beträgt die Länge des Vorgeschichten-Puffers **240** 48,75 msec oder 390 Abtastungen. Nachdem jeder Rahmen vom Decoder **220** decodiert wird, wird daher der Vorgeschichten-Puffer **240** aktualisiert, so dass er dri jüngste Sprach-Vorgeschichte enthält. Die Aktualisierung des Vorgeschickten-Puffers **240** wird in Fig. 6 gezeigt. Wie in dieser Figur dargestellt, enthält der Vorgeschichten-Puffer **240** rechts die jüngsten Sprach-Abtastwerte und links die ältesten Sprach-Abtastwerte. Wenn der neueste Rahmen der decodierten Sprache empfangen wird, wird er von rechts in den Puffer **240** geschoben, wobei die Abtastungen, die der am weitesten zurückliegenden Sprach-Abtrastwerte entsprechen, links aus dem Puffer geschoben werden (s. 6b).

**[0051]** Zusätzlich verzögert das Verzögerungsmodul **250** im Schritt **520** die Ausgabe der Sprache um ¼ der längsten Tonhöhen-Periode. Bei der 8 kHz-Abtastung sind dies 120 * ¼ = 30 Abtastungen oder 3,75 msec. Diese Verzögerung ermöglicht es dem FEC-Modul **230**, zu Beginn eines Ausfalls eine ¼-Wellenlängen-OLA durchzuführen, um einen reibungslosen Übergang zwischen dem echten Signal vor dem Ausfall und dem vom FEC-Modul **230** erzeugten synthetischen Signal zu gewährleisten. Die Ausgabe muss verzögert werden, da nach der Decodierung eines Rahmens nicht bekannt ist, ob der nächste Rahmen ausgefallen ist.

**[0052]** Im Schritt **525** wird das Audio ausgegeben, und im Schritt **530** bestimmt das Verfahren, ob es irgendwelche weiteren Rahmen gibt. Wenn es keine weiteren Rahmen gibt, endet das Verfahren. Wenn es mehr Rahmen gibt, kehrt das Verfahren zum Schritt **505** zurück, um den nächsten Rahmen zu holen.

**[0053]** Wenn jedoch im Schritt **510** der Detektor für verlorene Rahmen **215** ermittelt, dass der empfangene Rahmen ausgefallen ist, wird das Verfahren mit dem Schritt **535** fortgesetzt, wo das FEC-Modul **230** den ersten ausgefallenen Rahmen verdeckt, wobei der entsprechende Vorgang unten in Fig. 7 detailliert beschrieben wird. Nachdem der erste Rahmen verdeckt wird, holt der Detektor für verlorene Rahmen **215** im Schritt **540** den nächsten Rahmen. Im Schritt **545** bestimmt der Detektor für verlorene Rahmen **215**, ob der nächste Rahmen ausgefallen ist. Wenn der nächste Rahmen nicht ausgefallen ist, verarbeitet das FEC-Modul **230** im Schritt **555** den ersten Rahmen nach dem Ausfall, wobei dieser Vorgang unten in Fig. 16 detailliert beschrieben wird. Nachdem der erste Rahmen verarbeitet wird, kehrt das Verfahren zum Schritt **530** zurück, wo der Detektor für verlorene Rahmen **215** ermittelt, ob es weitere Rahmen gibt.

**[0054]** Wenn im Schritt **545** der Detektor für verlorene Rahmen **215** ermittelt, dass der nächste oder darauf folgende Rahmen ausgefallen sind, verdeckt das FEC-Modul **230** den zweiten und die darauf folgenden Rah-

men in Übereinstimmung mit einem Verfahren, das unten in Fig. 12 detailliert beschrieben wird.

**[0055]** Fig. 7 legt detailliert die Schritte dar, die durchgeführt werden, um die ersten 10 Millisekunden eines Ausfalls zu verdecken. Die Schritte werden unten detailliert untersucht.

**[0056]** Wie in Fig. 7 zu sehen ist, besteht die erste Operation zu Beginn eines Ausfalls darin, die Tonhöhe zu bestimmen. Hierzu wird am Signal des Vorgeschichten-Puffers **240** eine standardisierte Autokorrelation mit einem Fenster von 20 msec (160 Abtastungen) bei Abhörverzögerungen von 40 bis 120 Abtastungen durchgeführt. Bei einer 8 kHz-Abtastung entsprechen diese Verzögerungen Tonhöhen-Perioden von 5 bis 15 msec oder Grundfrequenzen von 200 bis 66 2/3 Hz. Der Abgriff am Spitzenwert der Autokorrelation ist der Tonhöhen-Schätzwert P. Angenommen, dass H diese Vorgeschichte enthält und von –1 (die Abtastung rechts vor dem Ausfall) bis –390 (die Abtastung 390 Abtastungen, bevor der Ausfall einsetzt) indiziert wird, 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 Spitzenwert der Autokorrelation bzw. der Tonhöhen-Schätzwert kann dann ausgedrückt werden als:

$$P = \{\max_j(Autocor(j)) \mid 40 \le j \le 120\}$$

**[0058]** Wie oben erwähnt, ist die kleinste gestattete Tonhöhen-Periode, 5 msec bzw. 40 Abtastungen, groß genug, dass eine einzige Tonhöhen-Periode in einem ausgefallenen Rahmen von 10 msec maximal zweimal wiederholt wird. Dadurch werden Artefakte in nicht über Stimme mitgeteilter Sprache und unnatürliche, harmonische Artefakte bei Sprechern mit hoher Tonhöhe vermieden.

**[0059]** Ein graphisches Beispiel für die Berechnung der standardisierten Autokorrelation für den Ausfall in Fig. 3 wird in Fig. 8 gezeigt.

**[0060]** Die Wellenform, die mit "Vorgeschichte" gekennzeichnet ist, stellt die Inhalte des Vorgeschichten-Puffers **240** unmittelbar vor dem Ausfall dar. Die gestrichelte horizontale Linie zeigt den Bezugsteil des Signals, den Vorgeschichten-Puffer 240 H[–1]: H[–160], der die 20-msec-Sprache unmittelbar vor dem Ausfall darstellt. Die durchgezogenen horizontalen Linien sind die Fenster von 20 msec, die an Abgriffen von 40 Abtastungen (die obere Linie, 5 msec Dauer, 200 Hz Frequenz) bis zu 120 Abtastungen (die untere Linie, 15 msec Dauer, 66,66 Hz Frequenz) verzögert werden. Die Ausgabe der Korrelation wird ebenfalls skizziert, in Ausrichtung mit den Positionen der Fenster. Die gestrichelte senkrechte Linie in der Korrelation ist der Scheitelpunkt der Kurve und stellt die ausgewertete Tonhöhe dar. Diese Linie liegt vom Beginn des Ausfalls an eine Periode zurück. In diesem Fall ist P gleich 56 Abtastungen, was einer Tonhöhen-Periode von 7 msec und einer Grundfrequenz von 142,9 Hz entspricht.

**[0061]** Um die Komplexität der Autokorrelation zu reduzieren, werden zwei besondere Verfahrensweisen angewandt. Diese Kurzverfahren verändern die Ausgabe zwar nicht wesentlich, haben jedoch einen großen Einfluss auf die Gesamt-Laufzeitkomplexität des Verfahrens. Der größte Teil der Komplexität im FEC-Verfahren liegt in der Autokorrelation.

**[0062]** Anstatt dass die Korrelation an jedem Abgriff berechnet wird, wird als erstes ein grober Schätzwert des Spitzenwerts bei einem stark abgeschwächten Signal bestimmt und dann in der Nähe des groben Spitzenwerts eine Feinsuche durchgeführt. Für den groben Schätzwert wird die Autocor-Funktion oben auf die neue Funktion abgeändert, die für ein 2:1 abgeschwächtes Signal gilt und nur jeden zweiten Abgriff untersucht:

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

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

**[0063]** Indem dann der grobe Schätzwert verwendet wird, wird das ursprüngliche Suchverfahren wiederholt, aber nur im Bereich $P_{rough} - 1 \le j \le P_{rough} + 1$. Es wird darauf geachtet sicherzustellen, dass j im ursprünglichen Bereich zwischen 40 und 120 Abtastungen bleibt. Es ist zu beachten, dass, wenn die Abtastfrequenz erhöht wird, auch der Dezimierungsfaktor erhöht werden sollte, so dass die Gesamtkomplexität des Verfahrens ungefähr konstant bleibt. Wir haben Tests mit Dezimierungsfaktoren von 8.1 bei Sprache durchgeführt, die bei 44,1 kHz abgetastet wurde, und gute Ergebnisse erzielt. Fig. 9 vergleicht den Graph von $Autocor_{rough}$ mit dem von Autocor. Wie aus der Figur ersichtlich, ist $Autocor_{rough}$ eine gute Annäherung an Autocor, und die Komplexität vermindert sich bei einer 8 kHz-Abtastung beinahe um einen Faktor von 4 (einen Faktor von 2, da nur jeder zweite Abgriff untersucht wird, und einen Faktor von 2, da bei einem bestimmten Abgriff nur jeder zweite Abtastwert untersucht wird).

**[0064]** Das zweite Verfahren wird durchgeführt, um die Komplexität der Energieberechnung bei Autocor und $Autocor_{rough}$ zu verringern. Anstatt einer Berechnung der Gesamtsumme bei jedem Schritt wird kontinuierlich eine laufende Summe für die Energie berechnet. Wenn also:

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

dann:

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

**[0065]** So sind nur 2 Vielfache und 2 Additionen erforderlich, um den Energieterm in jedem Schritt des FEC-Verfahrens zu aktualisieren, nachdem der erste Energieterm berechnet wurde.

**[0066]** Jetzt, da wir den Tonhöhen-Schätzwert, P, haben, beginnt die Erzeugung der Wellenform während des Ausfalls. Kehrt man zum Flussdiagramm in Fig. 7 zurück, werden im Schritt **710** die jüngsten 3,25 Wellenlängen (3,25 * P Abtastungen) aus dem Vorgeschichten-Puffer **240**, H, in den Tonhöhen-Puffer, B, kopiert. Die Inhalte des Tonhöhen-Puffers bleiben, mit Ausnahme der jüngsten Wellenlänge, für die Dauer des Ausfalls konstant. Der Vorgeschichten-Puffer **240** andererseits wird während des Ausfalls kontinuierlich mit der synthetischen Sprache aktualisiert.

**[0067]** Im Schritt **715** wird die neueste ¼ Wellenlänge (0,25 * P Abtastungen) aus dem Vorgeschichten-Puffer **240** im Puffer L des letzten Viertels gesichert. Diese ¼ Wellenlänge wird für mehrere OLA-Operationen benötigt. Aus praktischen Gründen werden wir dasselbe Negativ-Indexierungsschema verwenden, um auf B- und L-Puffer zuzugreifen, wie wir es für den Vorgeschichten-Puffer **240** taten. B[–1] ist die letzte Abtastung vor dem Einsetzen des Ausfalls, B[–2] die Abtastung davor, usw. Die synthetische Sprache wird in den synthetischen Puffer S gestellt, der von 0 an nach oben indiziert wird. So ist S[0] die erste synthetisierte Abtastung, S[1] die zweite, usw.

**[0068]** Die Inhalte des Tonhöhen-Puffers, B, und des Puffers des letzten Viertels, L, für den Ausfall in Fig. 3 werden in Fig. 10 gezeigt. Im vorherigen Abschnitt berechneten wir die Periode, P, als 56 Abtastungen. Der Tonhöhen-Puffer ist somit 3,25 * 56 = 182 Abtastungen lang. Der Puffer des letzten Viertels ist 0,25 56 = 14 Abtastungen lang. In der Figur wurden alle P Abtastungen vom Beginn des Ausfalls zurück senkrechte Linien gesetzt. Während der ersten 10 msec eines Ausfalls wird nur die letzte Tonhöhen-Periode vom Tonhöhen-Puffer verwendet; so ist im Schritt **720** U=1. Wenn das Sprachsignal wirklich periodisch und unser Tonhöhen-Schätzwert kein Schätzwert, sondern der genaue wirkliche Wert war, könnten wir einfach die Wellenform direkt aus dem Tonhöhen-Puffer B in den synthetischen Puffer S kopieren, und das synthetische Signal wäre

reibungslos und anhaltend. Das bedeutet: S[0]=B[–P], S[1]=B[–P+1], usw. Wenn die Tonhöhe kürzer ist als der 10-msec-Rahmen, d.h. P<80, wird die einzelne Tonhöhen-Periode im ausgefallenen Rahmen häufiger als einmal wiederholt. In unserem Beispiel ist P=56, also läuft das Kopieren bei S[56] über. Die Abtastung-für-Abtastung-Kopier-Sequenz in der Nähe der Abtastung 56 wäre:
S[54]=B[–2], S[55]=B[–1], S[56]=B[–56], S[57]=B[–55], usw.

**[0069]** In der praktischen Anwendung ist der Tonhöhen-Schätzwert nicht genau, und das Signal ist möglicherweise nicht wirklich periodisch. Um Diskontinuitäten (a) am Grenzbereich zwischen dem echten und dem synthetischen Signal und (b) am Grenzbereich, wo die Periode wiederholt wird, zu vermeiden, sind OLAs erforderlich. Für beide Grenzbereiche ist ein glatter Übergang vom Ende der echten Sprache B[–1] zur Sprache, die eine Periode zurückliegt, B[–P], wünschenswert. Im Schritt **725** kann dies daher durch die Überlappungs-Addition (OLA) der ¼ Wellenlänge vor B[–P] mit der letzten ¼ Wellenlänge des Vorgeschichten-Puffers **240** oder den Inhalten von L erreicht werden. Graphisch ist dies gleichbedeutend damit, die letzten 1¼ Wellenlängen im Tonhöhen-Puffer eine Wellenlänge nach rechts zu verschieben und eine OLA im ¼ Wellenlängen-Überlappungsbereich durchzuführen. Im Schritt **730** wird das Ergebnis der OLA zu der letzten Wellenlänge im Vorgeschichten-Puffer **240** kopiert. Um zusätzliche Perioden der synthetischen Wellenform zu erzeugen, wird der Tonhöhen-Puffer um zusätzliche Wellenlängen verschoben, und zusätzliche OLAs werden durchgeführt.

**[0070]** Fig. 11 zeigt die OLA-Operation für die ersten 2 Iterationen. In dieser Figur stellt die senkrechte Linie, die alle Wellenformen kreuzt, den Anfang des Ausfalls dar. Die kurzen senkrechten Linien sind Tonhöhenmarkierungen und werden P Abtastungen außerhalb des Ausfallgrenzbereichs gesetzt. Es sollte beachtet werden, dass der Überlappungsbereich zwischen den Wellenformen "Tonhöhen-Puffer" und "um P nach rechts verschoben" genau denselben Abtastungen wie jenen im Überlappungsbereich zwischen "um P nach rechts verschoben" und "um 2P nach rechts verschoben" entspricht. Daher muss die ¼ Wellenlängen-OLA nur einmal berechnet werden.

**[0071]** Im Schritt **735** kann, indem zunächst die OLA berechnet und die Ergebnisse in die letzte ¼, Wellenlänge des Tonhöhen-Puffers gestellt werden, das Verfahren für ein wirklich periodisches Signal, das die synthetische Wellenform erzeugt, verwendet werden. Bei der Abtastung B(-P) beginnend, müssen einfach die Abtastungen aus dem Tonhöhen-Puffer in den synthetischen Puffer kopiert werden, und der Tonhöhen-Pufferzeiger muss zurück auf den Beginn der Tonhöhen-Periode rollen, wenn das Ende des Tonhöhen-Puffers erreicht ist. Mittels Anwendung dieser Technik kann eine synthetische Wellenform beliebiger Dauer erzeugt werden. Die Tonhöhen-Periode links vom Ausfallbeginn in der "mit OLAs kombiniert"-Wellenform in Fig. 11 entspricht den aktualisierten Inhalten des Tonhöhen-Puffers.

**[0072]** Die "mit OLAs kombiniert" Wellenform zeigt, dass der Einperioden-Tonhöhen-Puffer ein periodisches Signal mit der Periode P ohne Diskontinuitäten erzeugt. Die synthetische Sprache, die aus einer einzigen Wellenlänge im Vorgeschichten-Puffer **240** erzeugt wird, wird verwendet, um die ersten 10 msec eines Ausfalls zu verdecken. Die Auswirkung der OLA ist zu sehen, indem die ¼ Wellenlänge direkt vor Beginn des Ausfalls in den "Tonhöhen-Puffer"- und "mit OLAs kombiniert"-Wellenformen verglichen wird. Im Schritt **730** ersetzt diese ¼ Wellenlänge in der "mit OLAs kombiniert"-Wellenform auch die letzte ¼ Wellenlänge im Vorgeschichten-Puffer **240**.

**[0073]** Die OLA-Operation mit dreieckigen Fenstern kann auch mathematisch ausgedrückt werden. Als erstes definieren wir die Variable P4 als ¼ der Tonhöhen-Periode in den Abtastungen. Somit ist P4=P»2. In unserem Beispiel war P 56, also ist P4 14. Die OLA-Operation kann im Bereich 1 ≤ i ≤ P4 ausgedrückt werden als:

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

**[0074]** Das Ergebnis der OLA ersetzt die letzten ¼ Wellenlängen sowohl im Vorgeschichten-Puffer **240** als auch im Tonhöhen-Puffer. Indem der Vorgeschichten-Puffer **240** ersetzt wird, wird der ¼-Wellenlängen-OLA-Übergang ausgegeben, wenn der Vorgeschichten-Puffer **240** aktualisiert wird, da der Vorgeschichten-Puffer **240** auch die Ausgabe um 3,75 msec verzögert. Die Ausgabewellenform während der ersten 10 msec des Ausfalls ist im Bereich zwischen den ersten zwei gestrichelten Linien in der "verdeckten" Wellenform in Fig. 3 zu sehen.

**[0075]** Im Schritt **740** wird am Ende der Erzeugung der synthetischen Sprache für den Rahmen das aktuelle Offset als Variable 0 im Tonhöhen-Puffer gesichert. Dieses Offset ermöglicht es, dass die synthetische Wellen-

form im nächsten Rahmen zum Zweck einer OLA mit dem echten oder synthetischen Signal des nächsten Rahmens fortgesetzt wird. O ermöglicht es auch, dass die richtige synthetische Signalphase beibehalten wird, wenn sich der Ausfall über 10 msec hinaus erstreckt. In unserem Beispiel mit 80 Abtastrahmen und P=56 ist das Offset zu Beginn des Ausfalls –56. Nach **56** Abtastungen läuft es zurück auf –56. Nach zusätzlichen 80–56=24 Abtastungen beträgt das Offset –56+24=-32, also ist O am Ende des ersten Rahmens –32.

**[0076]** Im Schritt **745** wird, nachdem der Synthesepuffer von S[0] auf S[79] aufgefüllt wurde, S verwendet, um den Vorgeschichten-Puffer **240** zu aktualisieren. Im Schritt **750** fügt der Vorgeschichten-Puffer **240** auch die Verzögerung von 3,75 msec hinzu. Die Handhabung des Vorgeschichten-Puffers **240** ist während der ausgefallenen und nicht ausgefallenen Rahmen dieselbe. Zu diesem Zeitpunkt endet die erste Rahmen-Verdeckungsoperation im Schritt **535** in Fig. 5, und das Verfahren wird mit dem Schritt **540** in Fig. 5 fortgesetzt.

**[0077]** Die Details darüber, wie das FEC-Modul **230** arbeitet, um spätere Rahmen jenseits von 10 msec zu verdecken, wie im Schritt **550** in Fig. 5 gezeigt, werden in Fig. 12 detailliert gezeigt.

**[0078]** Die Technik, die verwendet wird, um das synthetische Signal während des zweiten und späterer ausgefallener Rahmen zu erzeugen, ähnelt stark dem ersten ausgefallenen Rahmen, obwohl ein bisschen zusätzlich gearbeitet werden muss, um dem Signal eine Änderung hinzuzufügen.

**[0079]** Im Schritt **1205** bestimmt der Ausfallcode, ob der zweite oder dritte Rahmen ausfällt. Während des zweiten und dritten ausgefallenen Rahmens wird die Anzahl der verwendeten Tonhöhen-Perioden aus dem Tonhöhen-Puffer erhöht. Dies fügt mehr Variation zum Signal hinzu und hält die synthetisierte Ausgabe davon ab, zu harmonisch zu klingen. Wie im Zusammenhang mit allen anderen Übergängen ist eine OLA erforderlich, um den Grenzbereich zu glätten, wenn die Anzahl der Tonhöhen-Perioden erhöht wird. Jenseits vom dritten Rahmen (Ausfall von 30 msec) wird der Tonhöhen-Puffer konstant bei einer Länge von 3 Wellenlängen gehalten. Diese drei Wellenlängen erzeugen die gesamte synthetische Sprache für die Dauer des Ausfalls. Somit wird der Zweig links in der Fig. 12 nur beim zweiten und dritten ausgefallenen Rahmen eingeschlagen.

**[0080]** Im Schritt **1210** erhöhen wir als nächstes die Anzahl der im Tonhöhen-Puffer verwendeten Wellenlängen. Das heißt, wir setzen U=U+1.

**[0081]** Zu Beginn des zweiten oder dritten ausgefallenen Rahmens wird im Schritt **1215** das synthetische Signal vom vorherigen Rahmen für eine zusätzliche ¼ Wellenlänge in den Beginn des aktuellen Rahmens fortgesetzt. Zum Beispiel erscheint zu Beginn des zweiten Rahmens das synthetisierte Signal in unserem Beispiel wie in Fig. 13 gezeigt. Diese ¼ Wellenlänge wird mit dem neuen synthetischen Signal, das ältere Wellenlängen aus dem Tonhöhen-Puffer verwendet, überlappt/addiert.

**[0082]** Zu Beginn des zweiten ausgefallenen Rahmens wird die Anzahl der Wellenlängen auf 2 erhöht, U=2. Wie der Tonhöhen-Puffer einer Wellenlänge muss am Grenzbereich, wo der Tonhöhen-Puffer der zwei Wellenlängen sich wiederholen kann, eine OLA durchgeführt werden. Diesmal wird die ¼ Wellenlänge, die U Wellenlängen hinter dem Ende des Tonhöhen-Puffers B endet, im Schritt **1220** mit den Inhalten des Puffers L des letzten Viertels überlappt/addiert. Dieser OLA-Operator kann im Bereich 1 ≤ i ≤ P4 ausgedrückt werden als:

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

**[0083]** Der alleinige Unterschied zur vorherigen Version dieser Gleichung ist der, dass die Konstante P, die verwendet wird, um B auf der rechten Seite zu indexieren, in PU umgewandelt wurde. Die Erzeugung des Tonhöhen-Puffers von zwei Wellenlängen wird graphisch in Fig. 14 gezeigt.

**[0084]** Wie in Fig. 11 sind der Bereich der "mit OLAs kombiniert"-Wellenform links vom Ausfallbeginn die aktualisierten Inhalte des Zweiperioden-Tonhöhen-Puffers. Die kurzen senkrechten Linien markieren die Tonhöhen-Periode. Die nähere Untersuchung der aufeinanderfolgenden Spitzen in der "mit OLAs kombiniert"-Wellenform zeigt, dass die Spitzen von den Spitzen, die ein und zwei Wellenlängen vor dem Beginn des Ausfalls zurückliegen, alternieren.

**[0085]** Zu Beginn der synthetischen Ausgabe im zweiten Rahmen muss das Signal vom neuen Tonhöhen-Puffer mit der in Fig. 13 erzeugten ¼ Wellenlänge verschmolzen werden. Es ist erwünscht, dass das synthetische Signal aus dem neuen Tonhöhen-Puffer aus dem ältesten Abschnitt des aktuell verwendeten Puffers kommt. Es muss jedoch darauf geachtet werden, dass der neue Teil aus einem ähnlichen Abschnitt der Wel-

lenform kommt, oder es werden hörbare Artefakte erzeugt, wenn sie gemischt werden. Mit anderen Worten, es soll die richtige Phase aufrechterhalten werden, oder die Wellenformen können störend interferieren, wenn sie gemischt werden.

**[0086]** Dies wird im Schritt **1225** (Fig. 12) erreicht, indem die Perioden P vom am Ende des vorherigen Rahmens gesicherten Offset O subtrahiert werden, bis er auf die älteste Wellenlänge im verwendeten Abschnitt des Tonhöhen-Puffers deutet.

**[0087]** Im ersten ausgefallenen Rahmen war z. B. der gültige Index für den Tonhöhen-Puffer B von –1 bis –P. Daher muss sich der gesicherte O aus dem ersten ausgefallenen Rahmen in diesem Bereich befinden. Im zweiten ausgefallenen Rahmen reicht der gültige Bereich von –1 bis –2P. Also muss man P von O subtrahieren, bis sich O im Bereich –2P<=O<–P befindet. Oder, um es allgemeiner auszudrücken, man muss P von O subtrahieren, bis es im Bereich –UP<=O<–(U–1)P liegt. In unserem Beispiel ist am Ende des ersten ausgefallenen Rahmens P=56 und O=–32. Wir subtrahieren 56 von –32, was –88 ergibt. Somit kommt die erste synthetische Abtastung im zweiten Rahmen von B[–88], die nächste von B[–87], usw.

**[0088]** Das OLA-Mischen der synthetischen Signale aus den Ein- und Zweiperioden-Tonhöhen-Puffern zu Beginn des zweiten ausgefallenen Rahmens wird in Fig. 15 gezeigt.

**[0089]** Es ist anzumerken, dass man durch das Subtrahieren von P von O die richtige Wellenformphase aufrechterhält und die Spitzen des Signals in den "1P Tonhöhen-Puffer"- und "2P Tonhöhen-Puffer"-Wellenformen ausgerichtet werden. Die "OLA kombiniert"-Wellenform zeigt ebenfalls einen reibungslosen Übergang zwischen den verschiedenen Tonhöhen-Puffern zu Beginn des zweiten ausgefallenen Rahmens. Eine weitere Operation ist nötig, bevor der zweite Rahmen in der "OLA kombiniert"-Wellenform in Fig. 15 ausgegeben werden kann.

**[0090]** Im Schritt **1230** (Fig. 12) wird das neue Offset verwendet, um die ¼ Wellenlänge vom Tonhöhen-Puffer in einen temporären Puffer zu kopieren. Im Schritt **1235** wird ¼ Wellenlänge zum Offset addiert. Dann wird im Schritt **1240** der temporäre Puffer mit dem Beginn des Ausgabepuffers überlappt/addiert und das Ergebnis in die erste ¼ Wellenlänge des Ausgabepuffers gesetzt.

**[0091]** Im Schritt **1245** wird dann das Offset verwendet, um den Rest des Signals im Ausgabepuffer zu erzeugen. Der Tonhöhen-Puffer wird für die Dauer des 10-msec-Rahmens in den Ausgabepuffer kopiert. Im Schritt **1250** wird das aktuelle Offset als Variable O im Tonhöhen-Puffer gesichert.

**[0092]** Während der als zweites und später ausgefallenen Rahmen wird das synthetische Signal im Schritt **1255** mit einer linearen Anstiegsfunktion gedämpft. Das synthetische Signal wird graduell ausgeblendet, bis es jenseits von 60 msec auf 0 bzw. Stille gesetzt wird. Wenn der Ausfall länger wird, ist es wahrscheinlicher, dass die verdeckte Sprache vom echten Signal abweicht. Das Zu-lange-Halten bestimmter Klangarten kann, selbst wenn der einzelne Klang für eine kurze Zeitspanne natürlich klingt, zu unnatürlichen, hörbaren Artefakten in der Ausgabe des Verdeckungsverfahrens führen. Um diese Artefakte im synthetischen Signal zu vermeiden, wird ein langsames Ausblenden verwendet. Eine ähnliche Operation wird in den Verdeckungsverfahren durchgeführt, die in allen Standard-Sprachcodierern, wie beispielsweise dem G.723.1, G.728 und G.729 zu finden sind.

**[0093]** Das FEC-Verfahren dämpft das Signal um 20% pro Rahmen von 10 msec, angefangen mit dem zweiten Rahmen. Wenn S, der Synthesepuffer, das synthetische Signal vor der Dämpfung enthält und F die Anzahl der aufeinanderfolgenden ausgefallenen Rahmen ist (F = 1 beim ersten ausgefallenen Rahmen, 2 beim zweiten ausgefallenen Rahmen), dann kann die Dämpfung ausgedrückt werden als:

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

**[0094]** Im Bereich 0 ≤ i ≤ 79 und 2 ≤ F ≤ 6. Zum Beispiel ist bei den Abtastungen zu Beginn des zweiten ausgefallenen Rahmens F = 2, daher F – 2 = 0 und .2/80 = .0025, so dass S'[0] = 1.S[0], S'[1] = 0.99755[1], S'[2] = 0.9955[2]und S'[79] = 0.80255[79]. Jenseits vom sechsten ausgefallenen Rahmen wird die Ausgabe F-einfach auf 0 gesetzt.

**[0095]** Nachdem das synthetische Signal im Schritt **1255** gedämpft wird, wird es im Schritt **1260** an den Vorgeschichten-Puffer **240** gegeben und die Ausgabe im Schritt **1265** um 3,75 msec verzögert. Der Offset-Zeiger

O wird am Ende des zweiten Rahmens ebenfalls auf seine Position im Tonhöhen-Puffer aktualisiert, so dass das synthetische Signal im nächsten Rahmen fortgesetzt werden kann. Das Verfahren kehrt dann zum Schritt **540** zurück, um den nächsten Rahmen zu holen.

**[0096]** Wenn der Ausfall länger währt als zwei Rahmen, ist die Verarbeitung am dritten Rahmen genau wie im zweiten Rahmen, wenn man davon absieht, dass die Anzahl der Perioden im Tonhöhen-Puffer anstatt von 1 auf 2 von 2 auf 3 erhöht wird. Während unser Beispielausfall bei zwei Rahmen endet, wird der Dreiperioden-Tonhöhen-Puffer, der am dritten Rahmen und darüber hinaus verwendet würde, in Fig. 17 gezeigt. Jenseits vom dritten Rahmen bleibt die Anzahl der Perioden im Tonhöhen-Puffer bei 3 stehen, so dass nur der Weg auf der rechten Seite der Fig. 12 eingeschlagen wird. In diesem Fall wird der Offset-Zeiger O einfach verwendet, um den Tonhöhen-Puffer in die synthetische Ausgabe zu kopieren, und es werden keine Überlappungs/Additions-Operationen benötigt.

**[0097]** Die Operation des FEC-Moduls **230** am ersten intakten Rahmen nach einem Ausfall wird detailliert in Fig. 16 dargestellt. Am Ende eines Ausfalls wird zwischen der während des Ausfalls erzeugten synthetischen Sprache und der echten Sprache ein glatter Übergang benötigt. Wenn der Ausfall nur einen Rahmen lang war, wird im Schritt **1610** die synthetische Sprache für Wellenlänge fortgesetzt und eine Überlappung/Addition mit der echten Sprache durchgeführt.

**[0098]** Wenn das FEC-Modul **230** ermittelt, dass der Ausfall im Schritt **1620** länger war als 10 msec, sind Fehlanpassungen zwischen den synthetischen und den echten Signalen wahrscheinlicher; also wird im Schritt **1630** die synthetische Spracherzeugung fortgesetzt und das OLA-Fenster um zusätzliche 4 msec je ausgefallener Rahmen bis zu maximal 10 msec erhöht. Wenn der Schätzwert der Tonhöhe etwas daneben lag oder die Tonhöhe der echten Sprache während des Ausfalls verändert wurde, erhöht sich die Wahrscheinlichkeit einer Phasen-Fehlanpassung zwischen den synthetischen und echten Signalen mit der Länge des Ausfalls. Längere OLA-Fenster erzwingen das Ausblenden des synthetischen Signals und das langsamere Einblenden des echten Sprachsignals. Wenn der Ausfall länger war als 10 msec, ist es auch nötig, im Schritt **1640** die synthetische Sprache zu dämpfen, bevor eine OLA durchgeführt werden kann; daher stimmt sie mit dem Pegel des Signals im vorherigen Rahmen überein.

**[0099]** Im Schritt **1650** wird mit Beginn des neuen Eingaberahmens mit den Inhalten des Ausgabepuffers (synthetische Sprache) eine OLA durchgeführt. Der Beginn des Eingabepuffers wird durch das Ergebnis der OLA ersetzt. Die OLA am Ende des Ausfalls für das obige Beispiel ist in Fig. 4 zu sehen. Die vollständige Ausgabe des Verdeckungsverfahrens für das obige Beispiel kann in der "verdeckten" Wellenform in Fig. 3 eingesehen werden.

**[0100]** Im Schritt **1660** wird der Vorgeschichten-Puffer mit den Inhalten des Eingabepuffers aktualisiert. Im Schritt **1670** wird die Ausgabe der Sprache um 3,75 msec verzögert, und das Verfahren kehrt in Fig. 5 zum Schritt **530** zurück, um den nächsten Rahmen zu holen.

**[0101]** Mit einer kleinen Regulierung kann das FEC-Verfahren auf andere Sprachcodierer angewandt werden, die die Zustandsinformation zwischen Abtastungen oder Rahmen aufrechterhalten und keine Verdeckung bereitstellen, wie beispielsweise der G.726. Das FEC-Verfahren wird genau wie im vorherigen Abschnitt beschrieben verwendet, um die synthetische Wellenform während des Ausfalls zu erzeugen. Jedoch muss darauf geachtet werden sicherzustellen, dass die internen Zustandsvariablen des Codierers die vom FEC-Verfahren erzeugte synthetische Sprache nachvollziehen. Anderenfalls werden, nachdem der Ausfall vorüber ist, in der Ausgabe Artefakte und Diskontinuitäten erscheinen, wenn der Decoder mittels Verwendung seines fehlerhaften Zustands neu startet. Obwohl das OLA-Fenster am Ende eines Ausfalls hilft, muss mehr getan werden.

**[0102]** Bessere Ergebnisse können wie in Fig. 18 gezeigt gewonnen werden, indem der Decoder **1820** für die Dauer des Ausfalls in einen Codierer **1860** umgewandelt wird, indem die synthetisierte Ausgabe des FEC-Moduls **1830** als Eingabe des Codierers **1860** verwendet wird.

**[0103]** Auf diese Weise wird der Variablenstatus des Decoders **1820** die verdeckte Sprache nachvollziehen. Es ist anzumerken, dass der Codierer **1860**, anders als ein gewöhnlicher Codierer, nur läuft, um die Zustandsinformationen zu erhalten, und seine Ausgabe nicht verwendet wird. Daher können Kurzverfahren vorgenommen werden, um seine Laufzeitkomplexität erheblich zu verringern.

**[0104]** Wie oben dargelegt, gibt es viele Vorteile und Aspekte, die von der Erfindung bereitgestellt werden. Insbesondere wird, wenn ein Rahmenausfall fortschreitet, die Anzahl der aus der Signal-Vorgeschichte ver-

wendeten Tonhöhen-Perioden zum Erzeugen des synthetischen Signals abhängig von der Zeit erhöht. Dies reduziert erheblich harmonische Artefakte bei langen Ausfällen. Obwohl die Tonhöhen-Perioden nicht in ihrer ursprünglichen Reihenfolge abgespielt werden, klingt die Ausgabe dennoch natürlich.

**[0105]** Mit G.726 und anderen Codierern, die die Zustandsinformationen zwischen Abtastungen oder Rahmen verwalten, kann der Decoder als Codierer an der Ausgabe der synthetisierten Ausgabe des Verdeckungsverfahrens betrieben werden. Auf diese Weise werden die internen Zustandsvariablen des Decoders die Ausgabe nachverfolgen, wobei Diskontinuitäten aufgrund von fehlerhafter Zustandsinformation im Decoder, nachdem der Ausfall vorüber ist, vermieden oder zumindest reduziert werden. Da die Ausgabe vom Codierer nie verwendet wird (sein alleiniger Zweck ist der, die Zustandsinformation zu verwalten), kann eine Version des Codierers mit reduzierter Komplexität verwendet werden.

**[0106]** Die Mindest-Tonhöhen-Periode, die in den exemplarischen Ausführungsformen erlaubt ist (40 Abtastungen oder 200 Hz) ist größer als die Grundfrequenz, die für einige weibliche und kindliche Sprecher erwartet wird. Somit werden für Hochfrequenzsprecher selbst zu Beginn des Ausfalls mehrere Tonhöhen-Perioden verwendet, um die synthetische Sprache zu erzeugen. Bei Hoch-Grundfrequenzsprechern werden die Wellenformen häufiger wiederholt. Mehrere Tonhöhen-Perioden im synthetischen Signal machen harmonische Artefakte weniger wahrscheinlich. Diese Technik hilft auch dabei, das Signal während nicht über Stimme mitgeteilter Sprachsegmente sowie in Bereichen des schnellen Übergangs, wie beispielsweise einem Stopp, natürlich klingen zu lassen.

**[0107]** Das OLA-Fenster am Ende des ersten intakten Rahmens nach einem Ausfall wächst mit der Länge des Ausfalls. Im Zusammenhang mit längeren Ausfällen ist es wahrscheinlicher, dass Phasen-Anpassungen erfolgen, wenn der nächste intakte Rahmen eintrifft. Das Erweitern des OLA-Fensters in Abhängigkeit von der Ausfalllänge reduziert Überschwingimpulse, die von Phasen-Fehlanpassungen bei langen Ausfällen herrühren, ermöglicht aber dennoch die schnelle Rückgewinnung des Signals, wenn der Ausfall kurz ist.

**[0108]** Das FEC-Verfahren der Erfindung verwendet auch OLA-Fenster mit veränderlicher Länge, die nur einen Bruchteil der ausgewerteten Tonhöhe betragen, die ¼ Wellenlänge haben und nicht mit den Tonhöhen-Scheitelpunkten ausgerichtet sind.

**[0109]** Das FEC-Verfahren der Erfindung unterscheidet nicht zwischen einer gesprochenen und einer nicht über Stimme mitgeteilten Sprache. Stattdessen führt es die Reproduktion der nicht über Stimme mitgeteilten Sprache aufgrund zweier Attribute des Verfahrens gut durch: (A) Die Mindest-Fenstergröße ist durchaus so groß, dass selbst nicht über Stimme mitgeteilte Teile der Sprache angemessen variieren; und (B) Die Länge des Tonhöhen-Puffers wird erhöht, wenn das Verfahren fortschreitet, wiederum gewährleistend, dass keine harmonischen Artefakte eingeführt werden. Es ist anzumerken, dass die Verwendung großer Fenster zum Vermeiden der unterschiedlichen Handhabung von mit Stimme und nicht über Stimme mitgeteilter Sprache ebenfalls in der gut bekannten Zeitskalierungs-Technik WSOLA vorliegt.

**[0110]** Obwohl das Hinzufügen der Verzögerung zum Ermöglichen der OLA zu Beginn eines Ausfalls als ein unerwünschter Aspekt des Verfahrens der Erfindung betrachtet werden kann, ist es erforderlich, um einen reibungslosen Übergang zwischen echten und synthetischen Signalen zu Beginn des Ausfalls zu gewährleisten.

**[0111]** Obwohl diese Erfindung im Zusammenhang mit den spezifischen oben angeführten Ausführungsformen beschrieben wurde, ist es offenbar, dass für den Fachmann zahlreiche Alternativen, Modifikationen und Variationen offensichtlich sein werden. Dementsprechend sind die bevorzugten Ausführungsformen der Erfindung, wie oben ausgeführt, als erläuternd und nicht als einschränkend zu betrachten. Verschiedene Änderungen können durchgeführt werden, ohne vom Schutzumfang der Erfindung, wie im folgenden Anspruch definiert, abzuweichen.

## Patentansprüche

1. Ein Verfahren zum Verdecken der Auswirkung einer fehlenden Sprachinformation in einer erzeugten Sprache, wobei die Sprachinformation komprimiert und in Pakete an einen Empfänger gesendet wurde, der ein oder mehrere derartige Pakete nicht empfängt, das das Verdecken einer Diskontinuität zwischen synthetisierter Sprache und decodierter Sprache umfasst, wobei das Verfahren folgende Schritte umfasst:
das Synthetisieren eines Sprachsignals, das einem nicht verfügbaren Paket entspricht;
das Bestimmen eines Überlappung-Additionsfensters zur Verwendung beim Kombinieren eines Abschnitts des synthetisierten Sprachsignals mit einem anschließenden Sprachsignal, das aus einem vom Empfänger deco-

dierten empfangenen Paket herrührt, und
das Durchführen einer Überlappung-Additionsoperation an dem Abschnitt des synthetisierten Sprachsignals und des anschließenden Signals mittels Verwendung des Überlappung-Additionsfensters,
**dadurch gekennzeichnet**, dass die Größe des Überlappung-Additionsfensters auf der Grundlage der Dauer der Nicht-Verfügbarkeit der Pakete bestimmt wird.

Es folgen 14 Blatt Zeichnungen

**Anhängende Zeichnungen**

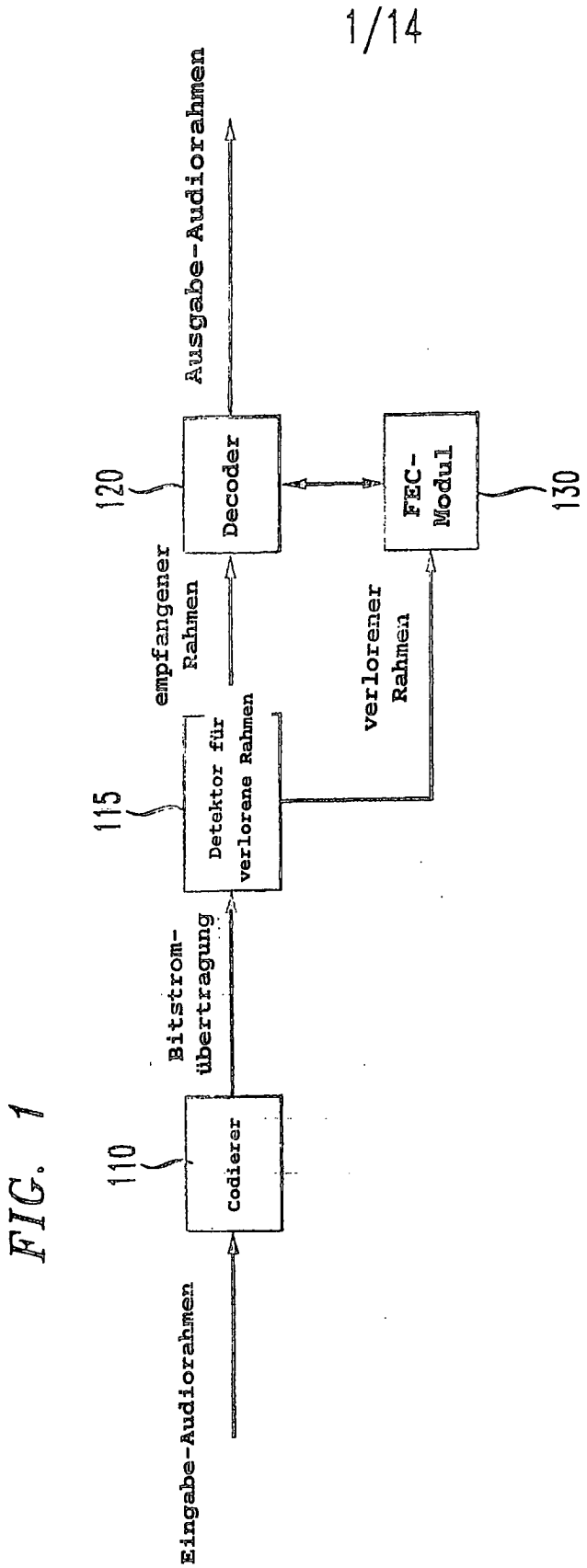



FIG. 1

FIG. 2

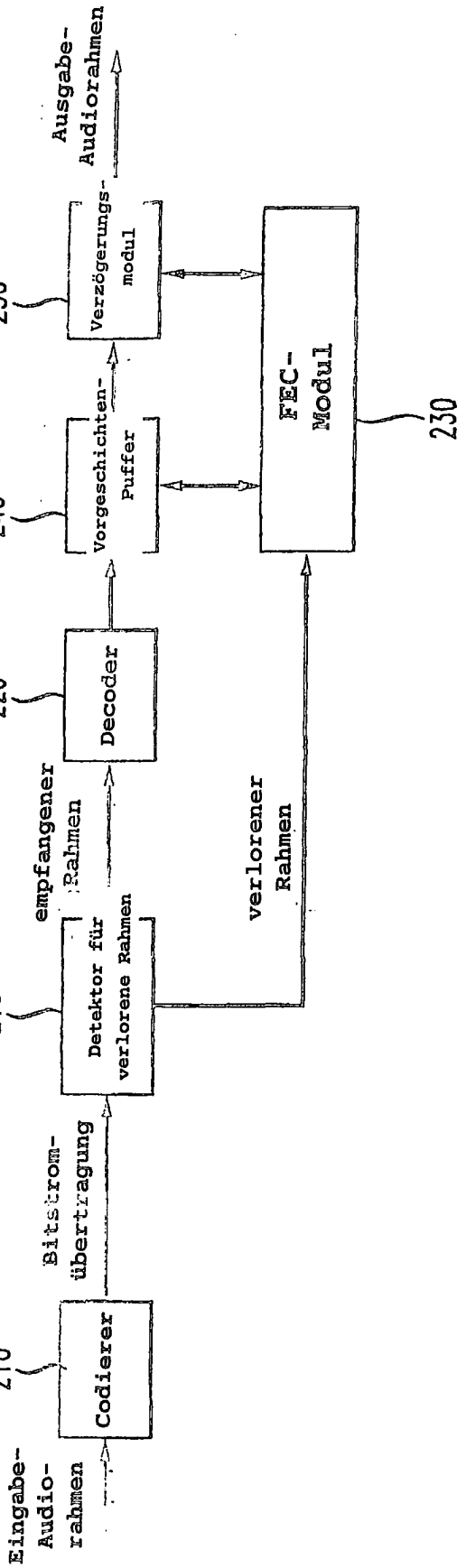

Eingabe-Audio-rahmen
210
Codierer
Bitstrom-übertragung
215
Detektor für verlorene Rahmen
empfangener Rahmen
220
Decoder
240
Vorgeschichten-Puffer
250
Verzögerungs-modul
Ausgabe-Audiorahmen
verlorener Rahmen
FEC-Modul
230

*FIG. 3*

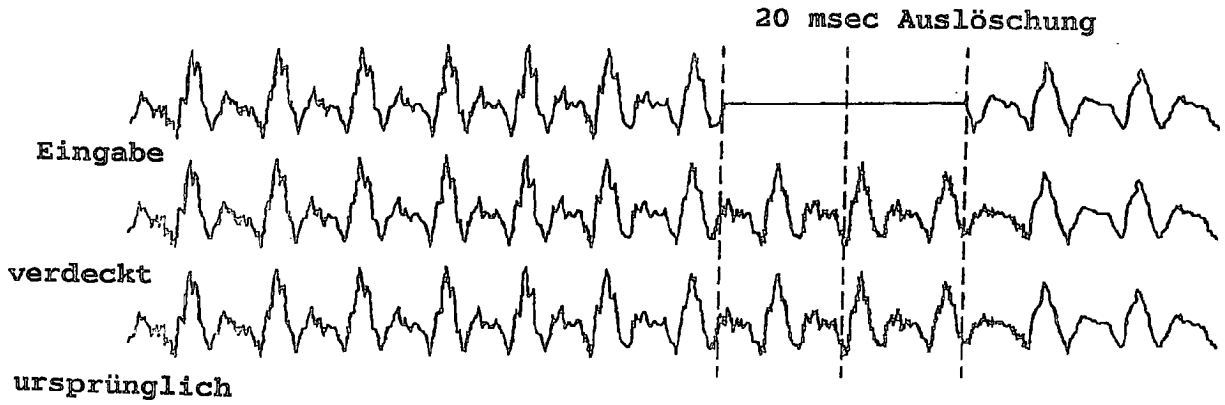

20 msec Auslöschung
Eingabe
verdeckt
ursprünglich

*FIG. 4*

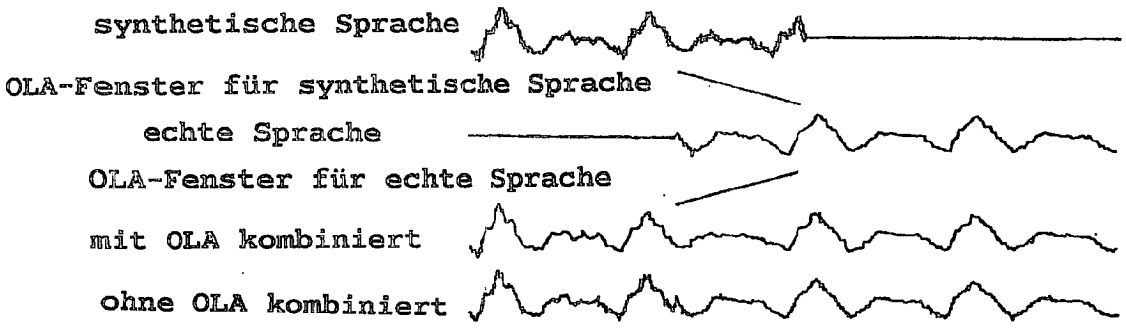

synthetische Sprache
OLA-Fenster für synthetische Sprache
echte Sprache
OLA-Fenster für echte Sprache
mit OLA kombiniert
ohne OLA kombiniert

*FIG. 5*

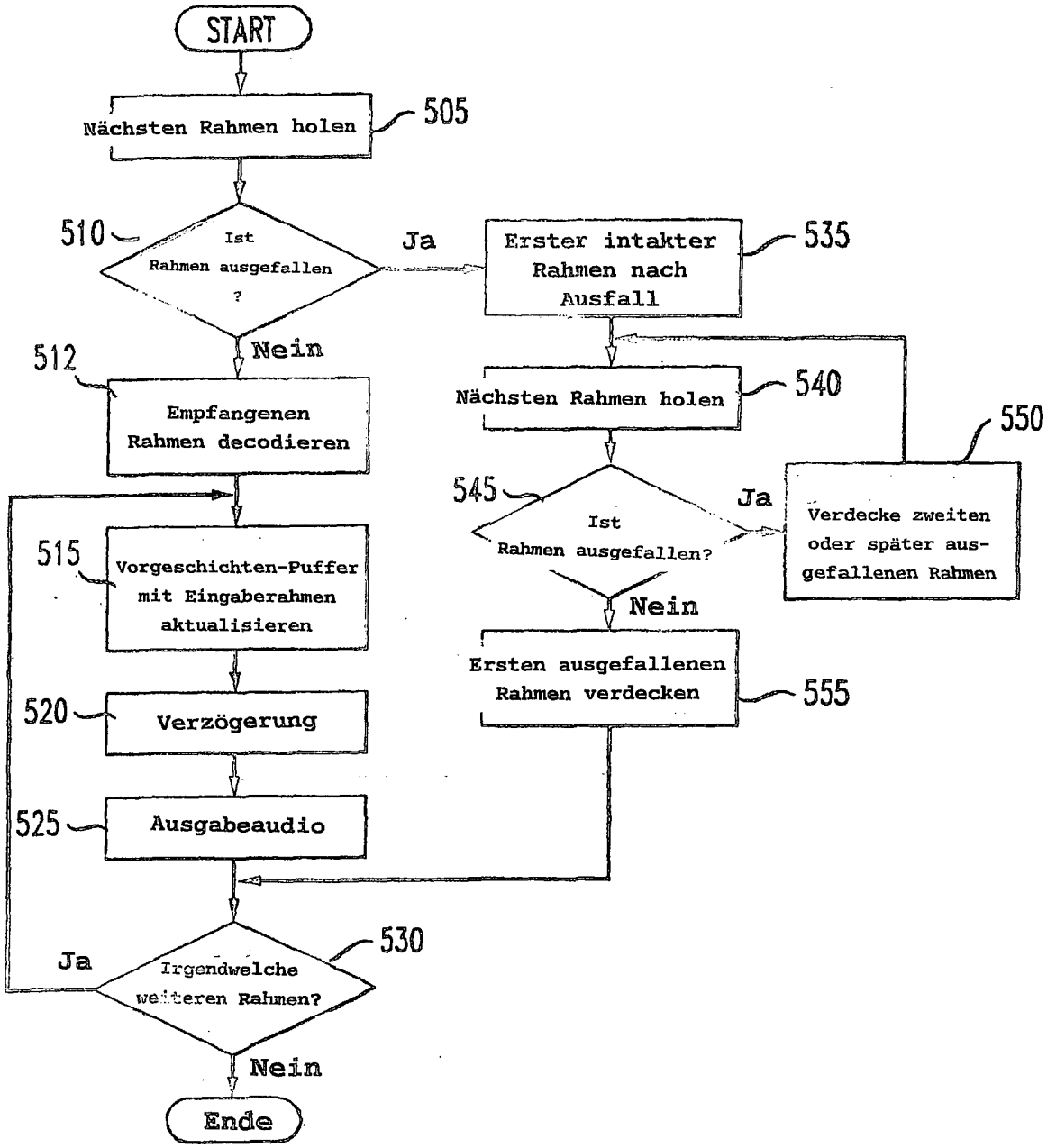

START
Nächsten Rahmen holen
505
510
Ist Rahmen ausgefallen ?
Ja
Erster intakter Rahmen nach Ausfall
535
Nein
512
Empfangenen Rahmen decodieren
Nächsten Rahmen holen
540
550
545
Ist Rahmen ausgefallen?
Ja
Verdecke zweiten oder später ausgefallenen Rahmen
Nein
515
Vorgeschichten-Puffer mit Eingaberahmen aktualisieren
Ersten ausgefallenen Rahmen verdecken
555
520
Verzögerung
525
Ausgabeaudio
530
Ja
Irgendwelche weiteren Rahmen?
Nein
Ende

*FIG. 6*

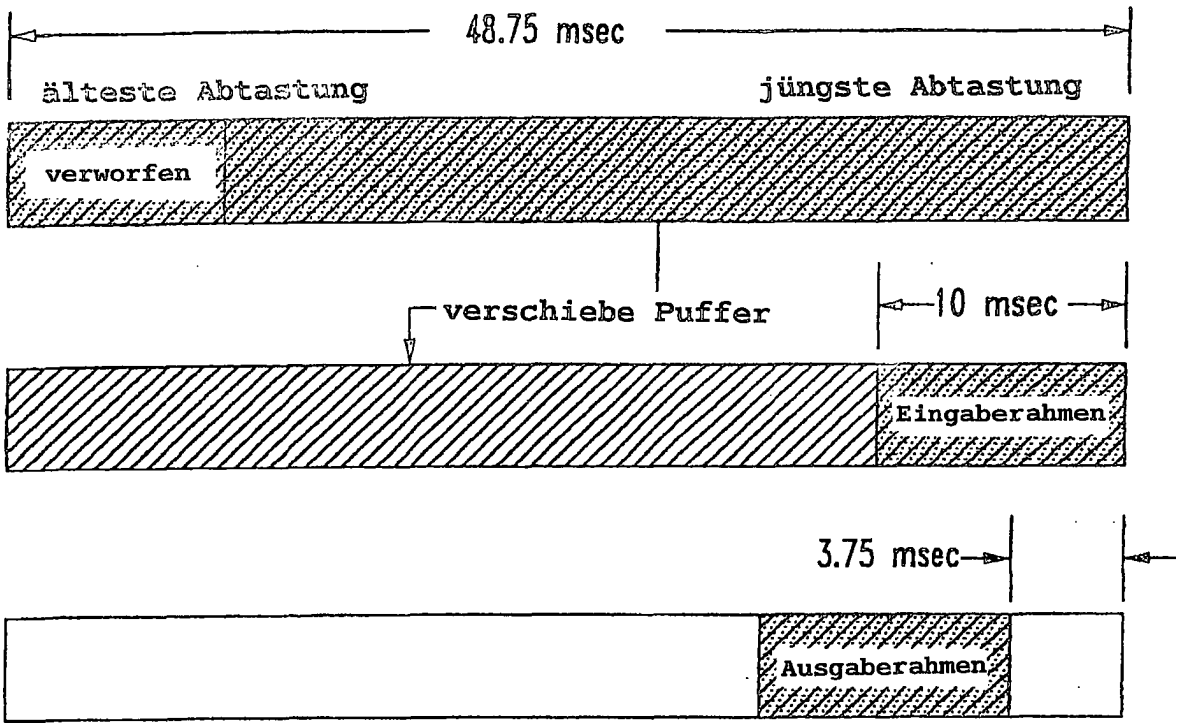

48.75 msec
älteste Abtastung
jüngste Abtastung
verworfen
verschiebe Puffer
10 msec
Eingaberahmen
3.75 msec
Ausgaberahmen

FIG. 7

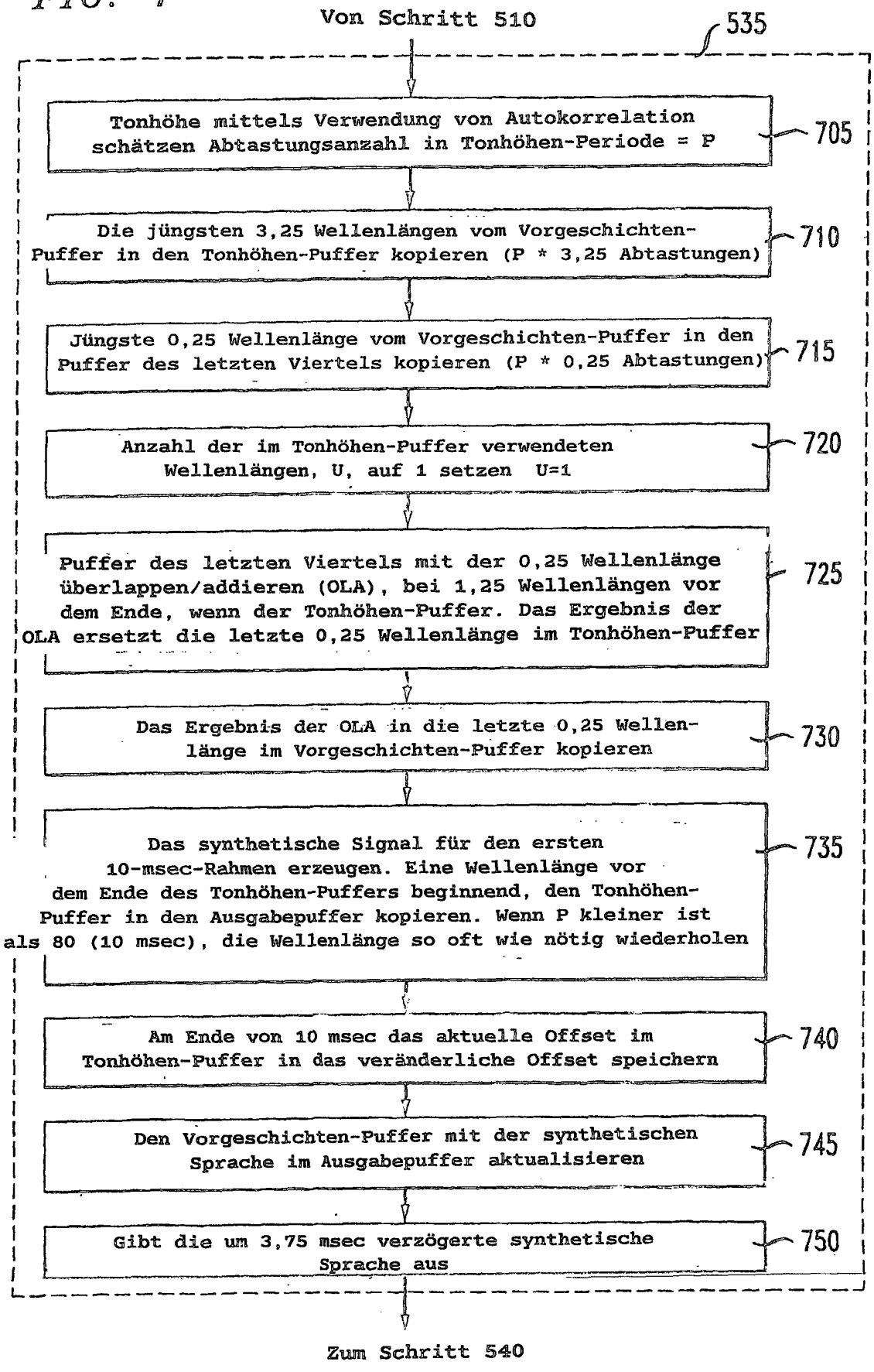

Von Schritt 510
535
Tonhöhe mittels Verwendung von Autokorrelation schätzen Abtastungsanzahl in Tonhöhen-Periode = P
705
Die jüngsten 3,25 Wellenlängen vom Vorgeschichten-Puffer in den Tonhöhen-Puffer kopieren (P * 3,25 Abtastungen)
710
Jüngste 0,25 Wellenlänge vom Vorgeschichten-Puffer in den Puffer des letzten Viertels kopieren (P * 0,25 Abtastungen)
715
Anzahl der im Tonhöhen-Puffer verwendeten Wellenlängen, U, auf 1 setzen U=1
720
Puffer des letzten Viertels mit der 0,25 Wellenlänge überlappen/addieren (OLA), bei 1,25 Wellenlängen vor dem Ende, wenn der Tonhöhen-Puffer. Das Ergebnis der OLA ersetzt die letzte 0,25 Wellenlänge im Tonhöhen-Puffer
725
Das Ergebnis der OLA in die letzte 0,25 Wellenlänge im Vorgeschichten-Puffer kopieren
730
Das synthetische Signal für den ersten 10-msec-Rahmen erzeugen. Eine Wellenlänge vor dem Ende des Tonhöhen-Puffers beginnend, den Tonhöhen-Puffer in den Ausgabepuffer kopieren. Wenn P kleiner ist als 80 (10 msec), die Wellenlänge so oft wie nötig wiederholen
735
Am Ende von 10 msec das aktuelle Offset im Tonhöhen-Puffer in das veränderliche Offset speichern
740
Den Vorgeschichten-Puffer mit der synthetischen Sprache im Ausgabepuffer aktualisieren
745
Gibt die um 3,75 msec verzögerte synthetische Sprache aus
750
Zum Schritt 540

*FIG. 8*

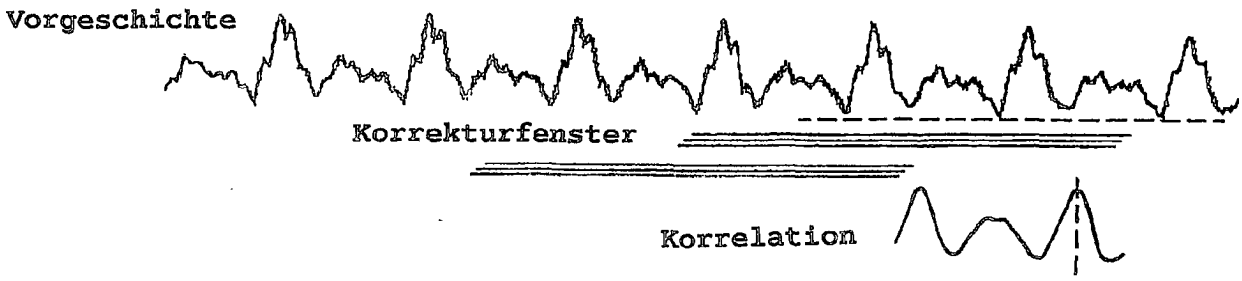

Vorgeschichte
Korrekturfenster
Korrelation

*FIG. 9*

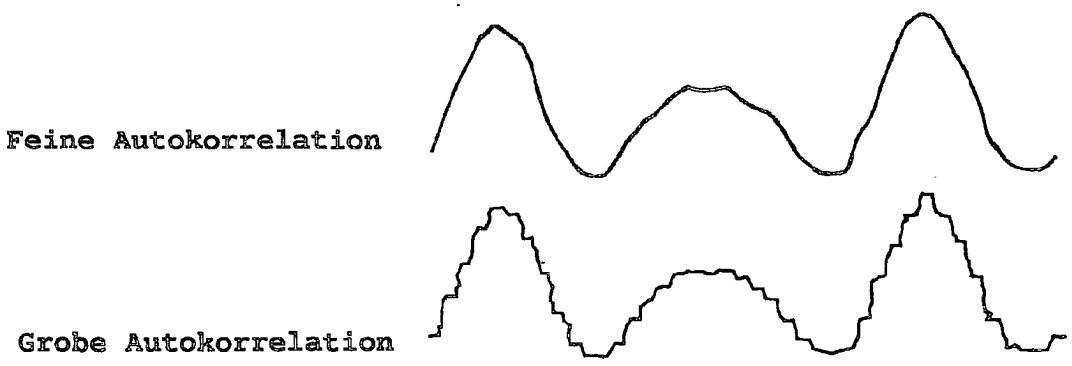

Feine Autokorrelation
Grobe Autokorrelation

*FIG. 10*

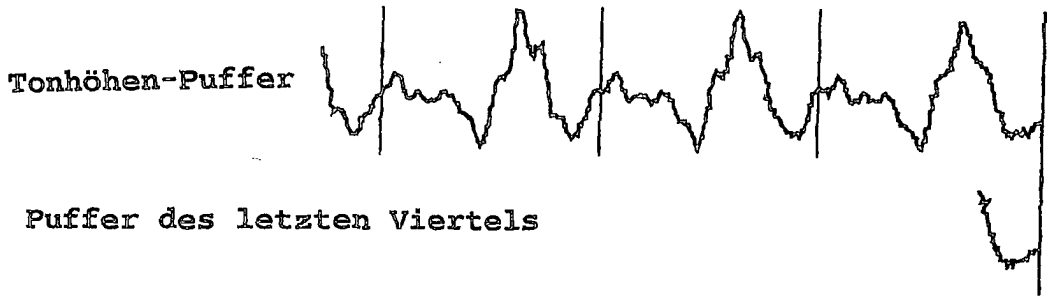

Tonhöhen-Puffer
Puffer des letzten Viertels

*FIG. 11*

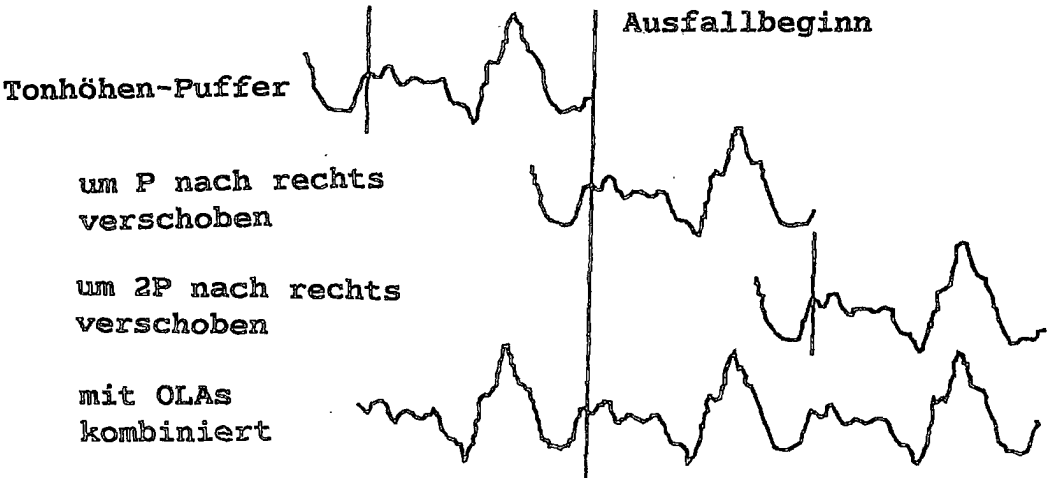

Ausfallbeginn
Tonhöhen-Puffer
um P nach rechts verschoben
um 2P nach rechts verschoben
mit OLAs kombiniert

*FIG. 12*

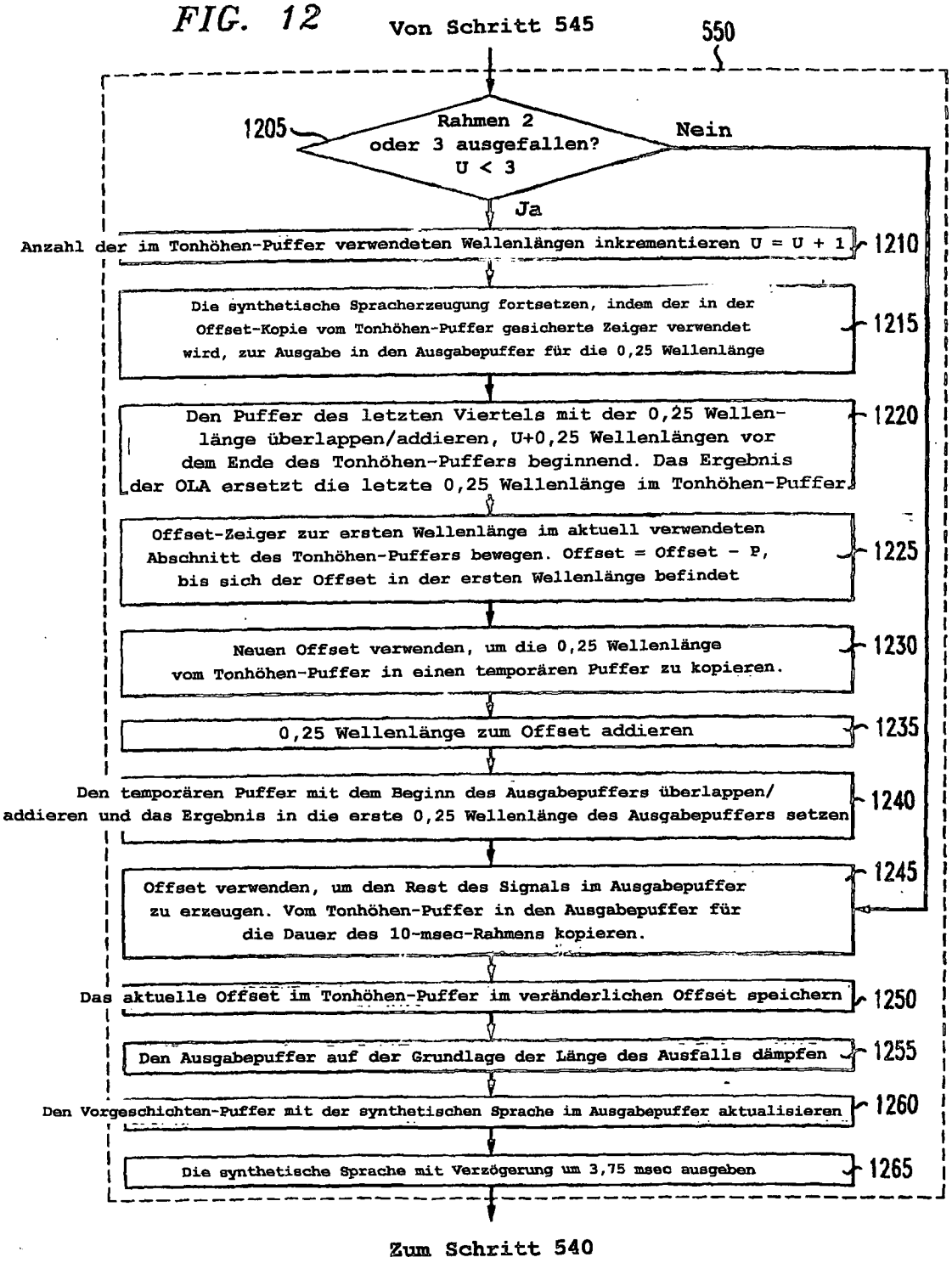

Von Schritt 545
550
1205
Rahmen 2 oder 3 ausgefallen? U < 3
Nein
Ja
Anzahl der im Tonhöhen-Puffer verwendeten Wellenlängen inkrementieren U = U + 1
1210
Die synthetische Spracherzeugung fortsetzen, indem der in der Offset-Kopie vom Tonhöhen-Puffer gesicherte Zeiger verwendet wird, zur Ausgabe in den Ausgabepuffer für die 0,25 Wellenlänge
1215
Den Puffer des letzten Viertels mit der 0,25 Wellenlänge überlappen/addieren, U+0,25 Wellenlängen vor dem Ende des Tonhöhen-Puffers beginnend. Das Ergebnis der OLA ersetzt die letzte 0,25 Wellenlänge im Tonhöhen-Puffer
1220
Offset-Zeiger zur ersten Wellenlänge im aktuell verwendeten Abschnitt des Tonhöhen-Puffers bewegen. Offset = Offset - P, bis sich der Offset in der ersten Wellenlänge befindet
1225
Neuen Offset verwenden, um die 0,25 Wellenlänge vom Tonhöhen-Puffer in einen temporären Puffer zu kopieren.
1230
0,25 Wellenlänge zum Offset addieren
1235
Den temporären Puffer mit dem Beginn des Ausgabepuffers überlappen/addieren und das Ergebnis in die erste 0,25 Wellenlänge des Ausgabepuffers setzen
1240
Offset verwenden, um den Rest des Signals im Ausgabepuffer zu erzeugen. Vom Tonhöhen-Puffer in den Ausgabepuffer für die Dauer des 10-msec-Rahmens kopieren.
1245
Das aktuelle Offset im Tonhöhen-Puffer im veränderlichen Offset speichern
1250
Den Ausgabepuffer auf der Grundlage der Länge des Ausfalls dämpfen
1255
Den Vorgeschichten-Puffer mit der synthetischen Sprache im Ausgabepuffer aktualisieren
1260
Die synthetische Sprache mit Verzögerung um 3,75 msec ausgeben
1265
Zum Schritt 540

*FIG. 13*


erster 10-msec-Rahmen
zweiter 10-msec-Rahmen

*FIG. 14*

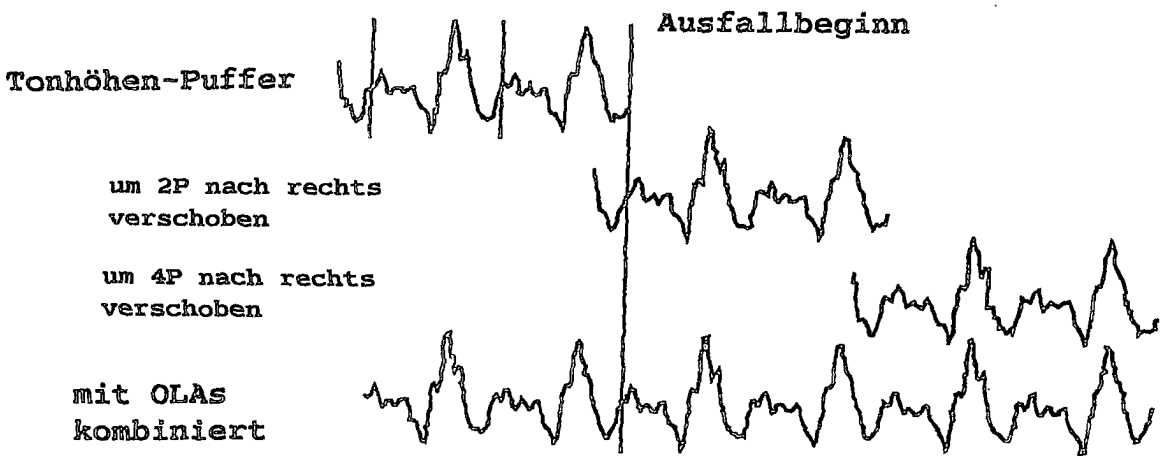

Ausfallbeginn
Tonhöhen-Puffer
um 2P nach rechts verschoben
um 4P nach rechts verschoben
mit OLAs kombiniert

FIG. 15

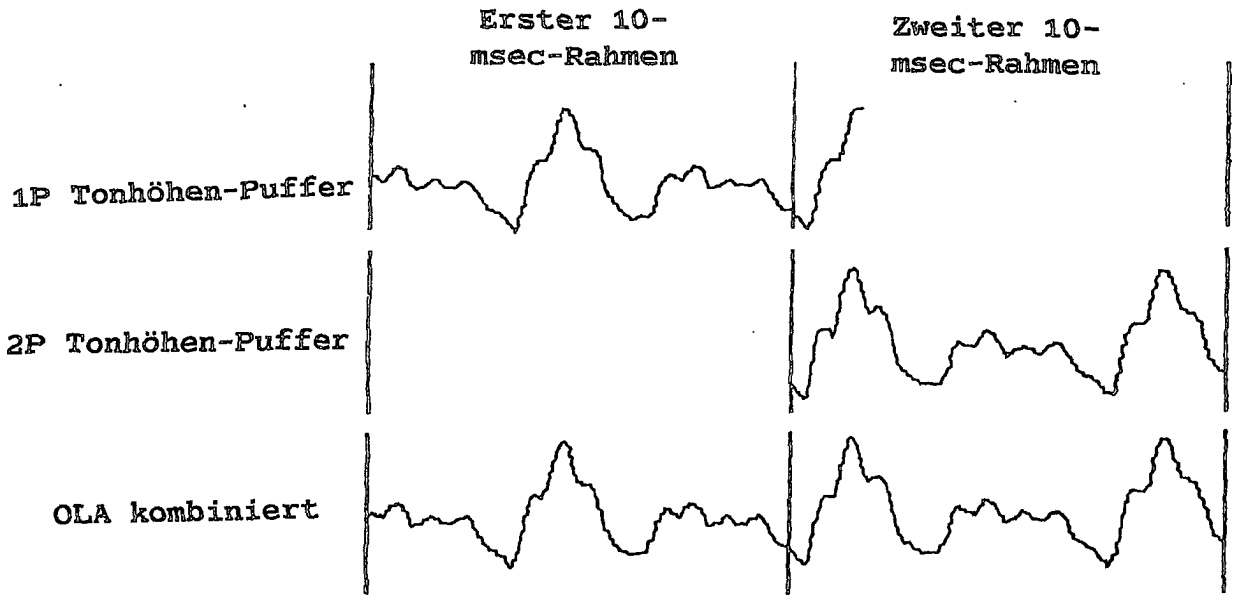

Erster 10-msec-Rahmen
Zweiter 10-msec-Rahmen
1P Tonhöhen-Puffer
2P Tonhöhen-Puffer
OLA kombiniert

FIG. 16

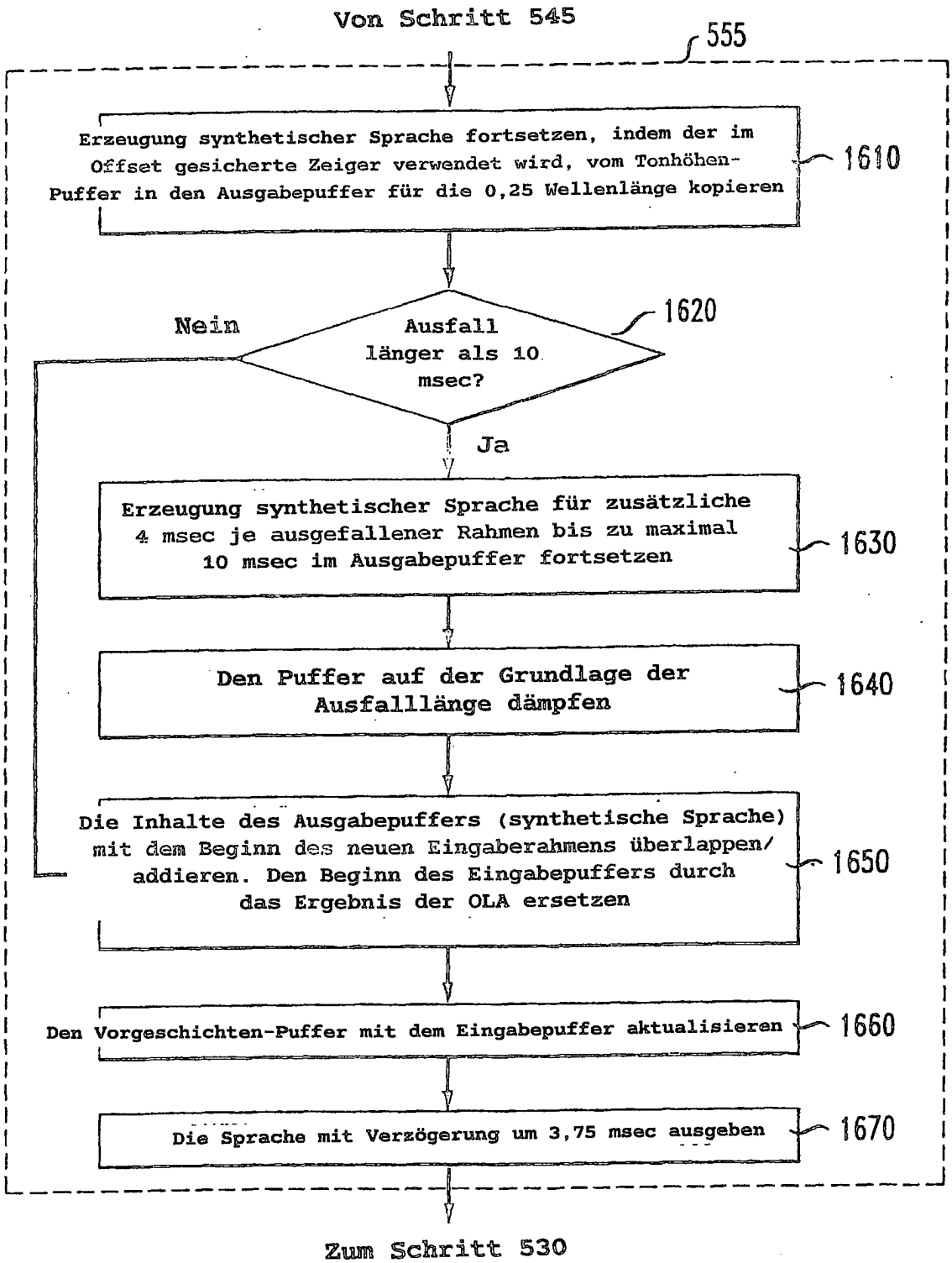

Von Schritt 545
555
Erzeugung synthetischer Sprache fortsetzen, indem der im Offset gesicherte Zeiger verwendet wird, vom Tonhöhen-Puffer in den Ausgabepuffer für die 0,25 Wellenlänge kopieren
1610
Nein
Ausfall länger als 10 msec?
1620
Ja
Erzeugung synthetischer Sprache für zusätzliche 4 msec je ausgefallener Rahmen bis zu maximal 10 msec im Ausgabepuffer fortsetzen
1630
Den Puffer auf der Grundlage der Ausfalllänge dämpfen
1640
Die Inhalte des Ausgabepuffers (synthetische Sprache) mit dem Beginn des neuen Eingaberahmens überlappen/addieren. Den Beginn des Eingabepuffers durch das Ergebnis der OLA ersetzen
1650
Den Vorgeschichten-Puffer mit dem Eingabepuffer aktualisieren
1660
Die Sprache mit Verzögerung um 3,75 msec ausgeben
1670
Zum Schritt 530

*FIG. 17*

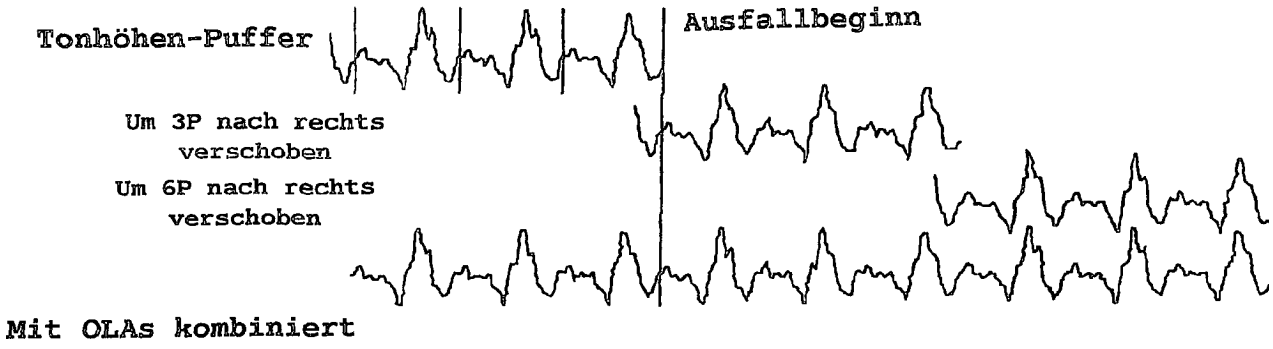
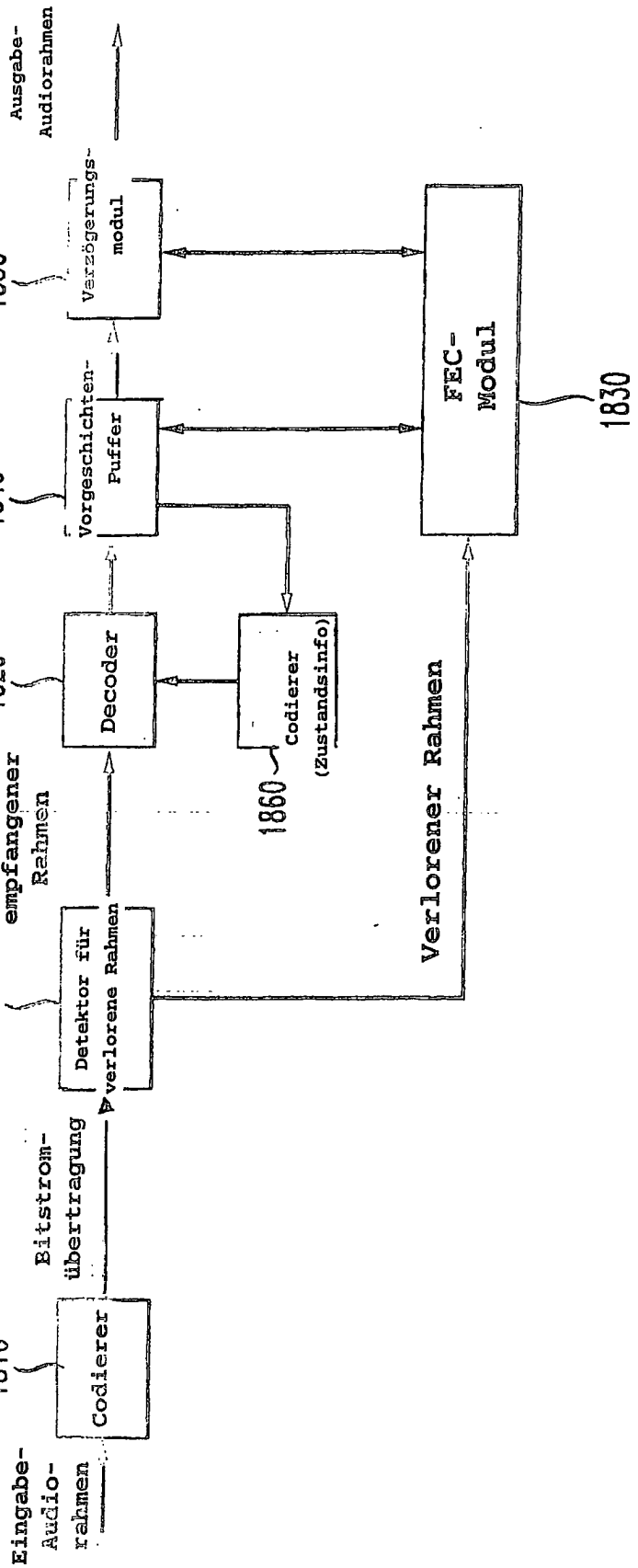

Tonhöhen-Puffer
Ausfallbeginn
Um 3P nach rechts verschoben
Um 6P nach rechts verschoben
Mit OLAs kombiniert

FIG. 18


Eingabe-Audio-rahmen
1810
Codierer
Bitstrom-übertragung
1815
Detektor für verlorene Rahmen
empfangener Rahmen
1820
Decoder
1860
Codierer (Zustandsinfo)
1840
Vorgeschichten-Puffer
1850
Verzögerungs-modul
Ausgabe-Audiorahmen
FEC-Modul
1830
Verlorener Rahmen