



(11) **EP 2 245 620 B1**

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:  
**30.08.2017 Patentblatt 2017/35**

(21) Anmeldenummer: **09711709.7**

(22) Anmeldetag: **02.02.2009**

(51) Int Cl.:  
**G10L 19/012** <sup>(2013.01)</sup> *G10L 19/18* <sup>(2013.01)</sup>

(86) Internationale Anmeldenummer:  
**PCT/EP2009/051123**

(87) Internationale Veröffentlichungsnummer:  
**WO 2009/103610 (27.08.2009 Gazette 2009/35)**

(54) **VERFAHREN UND MITTEL ZUR ENKODIERUNG VON HINTERGRUNDRÄUSCHINFORMATIONEN**

METHOD AND MEANS FOR ENCODING BACKGROUND NOISE INFORMATION

PROCÉDÉ ET MOYENS D'ENCODAGE D'INFORMATIONS DE BRUIT DE FOND

(84) Benannte Vertragsstaaten:  
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO SE SI SK TR**

(30) Priorität: **19.02.2008 DE 102008009718**

(43) Veröffentlichungstag der Anmeldung:  
**03.11.2010 Patentblatt 2010/44**

(73) Patentinhaber: **Unify GmbH & Co. KG 80807 München (DE)**

(72) Erfinder:  
• **SCHANDL, Stefan A-1130 Wien (AT)**  
• **SETIAWAN, Panji 81739 München (DE)**  
• **TADDEI, Herve 2273 EG Voorburg (NL)**

(74) Vertreter: **Fritzsche, Thomas Fritzsche Patent Naupliastraße 110 81545 München (DE)**

(56) Entgegenhaltungen:  
**US-A1- 2008 027 716**

- **SOLLAUD A.: "G.729.1 RTP Payload Format update: DTX support"[Online] 8. Februar 2008 (2008-02-08), XP002526621 Gefunden im Internet: URL:<http://www.ietf.org/proceedings/08mar/IDs/draft-ietf-avt-rfc4749-dtx-update-00.t xt>**
- **SETIAWAN P. ET AL: "On the ITU-T G.729.1 Silence Compression Scheme" EUSIPCO-2008, 25. August 2008 (2008-08-25), - 29. August 2008 (2008-08-29) Seiten 1-5, XP002526622 Lausanne, Switzerland**
- **INTERNATIONAL TELECOMMUNICATION UNION ITU-T: "G.729.1 Amendment 4: New Annex C "DTX/CNG scheme"" SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS, Nr. G.729.1, Juni 2008 (2008-06), Seiten 1-36, XP002526623**

**EP 2 245 620 B1**

Anmerkung: Innerhalb von neun Monaten nach Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents im Europäischen Patentblatt kann jedermann nach Maßgabe der Ausführungsordnung beim Europäischen Patentamt gegen dieses Patent Einspruch einlegen. Der Einspruch gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist. (Art. 99(1) Europäisches Patentübereinkommen).

## Beschreibung

**[0001]** Die Erfindung betrifft Verfahren und Mittel zur Enkodierung von Hintergrundrauschinformationen bei Sprachsignalkodierungsverfahren.

**[0002]** Für Telefongespräche ist seit den Anfängen der Telekommunikation eine Bandbreitenbeschränkung für eine analoge Sprachübertragung vorgesehen. Die Sprachübertragung erfolgt auf einem eingeschränkten Frequenzbereich von 300 Hz bis 3400 Hz.

**[0003]** Ein solcher eingeschränkter Frequenzbereich ist auch bei vielen Sprachsignalkodierungsverfahren für die heutige digitale Telekommunikation vorgesehen. Vor einem Kodiervorgang wird hierzu eine Bandbreitenbegrenzung des analogen Signals durchgeführt. Zur Kodierung und zur Dekodierung kommt dabei ein Codec zum Einsatz, welcher aufgrund der beschriebenen Bandbreitenbeschränkung im Frequenzbereich zwischen 300 Hz und 3400 Hz im Folgenden auch als schmalbandiger Sprachcodec (Narrow Band Speech Codec) bezeichnet wird. Unter dem Begriff Codec wird dabei sowohl die Kodiervorschrift zur digitalen Kodierung von Audiosignalen als auch die Dekodiervorschrift zur Dekodierung von Daten mit dem Ziel einer Rekonstruktion des Audiosignals verstanden.

**[0004]** Ein schmalbandiger Sprachcodec ist beispielsweise aus der ITU-T-Empfehlung G.729 bekannt. Mittels der dort beschriebenen Kodiervorschrift ist eine Übertragung eines schmalbandigen Sprachsignals mit einer Datenrate von 8 kbit/s vorgesehen.

**[0005]** Weiterhin sind sogenannte breitbandige Sprachcodecs (Wide Band Speech Codec) bekannt, welche zur Verbesserung des Höreindrucks eine Kodierung eines Audiosignals in einem erweiterten Frequenzbereich vorsehen. Ein derart erweiterter Frequenzbereich liegt z.B. zwischen einer Frequenz von 50 Hz und 7000 Hz. Ein breitbandiger Sprachcodec ist beispielsweise aus der ITU-T-Empfehlung G.729.EV bekannt.

**[0006]** Üblicherweise sind Kodierungsverfahren für breitbandige Sprachcodecs skalierbar gestaltet. Mit einer Skalierbarkeit ist hier gemeint, dass die übertragenen kodierten Daten verschiedene abgegrenzte Blöcke enthalten, welche den schmalbandigen Anteil, den breitbandigen Anteil und/oder die volle Bandbreite des kodierten Sprachsignals enthalten. Eine solche skalierbare Gestaltung gestattet einerseits eine empfängerseitige Abwärtskompatibilität und andererseits bietet sie eine einfache Möglichkeit, im Falle von eingeschränkten Datenübertragungskapazitäten im Übertragungskanal eine sender- und empfängerseitige Anpassung der Datenrate und der Größe von übertragenen Datenrahmen vorzunehmen.

**[0007]** Für eine Reduzierung der Datenübertragungsrate durch einen Codec ist üblicherweise eine Komprimierung der zu übertragenden Daten vorgesehen. Eine Komprimierung wird beispielsweise durch Kodierungsverfahren erreicht, bei denen zur Kodierung der Sprachdaten Parameter für ein Anregungssignal und Filterparameter bestimmt werden. Die Filterparameter sowie das

Anregungssignal spezifizierende Parameter werden dann an den Empfänger übertragen. Dort wird mithilfe des Codecs ein synthetisches Sprachsignal synthetisiert, das dem ursprünglichen Sprachsignal hinsichtlich eines subjektiven Höreindrucks möglichst ähnlich ist. Mit Hilfe dieses auch als »Analysis-by-Synthesis« bezeichneten Verfahrens werden nicht die ermittelten und digitalisierten Abtastwerte (Samples) selbst übertragen, sondern ermittelte Parameter, die eine empfängerseitige Synthese des Sprachsignals ermöglichen.

**[0008]** Eine weitere Maßnahme zur Reduzierung der Datenübertragungsrate bietet ein Verfahren zur diskontinuierlichen Übertragung (Discontinuous Transmission), welches in der Fachwelt auch unter dem Begriff DTX geläufig ist. Dokument US 2008/0027716 A1 offenbart eine DTX Signalisierung. Das grundsätzliche Ziel von DTX ist eine Reduzierung der Datenübertragungsrate im Fall einer Sprechpause.

**[0009]** Hierzu kommt auf Seiten des Senders eine Sprechpausenerkennung (Voice Activity Detection, VAD) zum Einsatz, welche bei Unterschreiten eines bestimmten Signalpegels auf eine Sprechpause erkennt.

**[0010]** Üblicherweise wird vom Empfänger während einer Sprechpause keine völlige Stille erwartet. Im Gegenteil würde eine völlige Stille empfängerseitig zu Irritationen oder sogar zur Vermutung eines Verbindungsabbaus führen. Aus diesem Grund werden Verfahren zur Erzeugung eines sogenannten Komfortrauschens (Comfort Noise) angewandt.

**[0011]** Bei einem Komfortrauschen handelt es sich um Rauschen, welches zur Füllung von Stillephasen auf Seiten des Empfängers synthetisiert wird. Das Komfortrauschen dient einem subjektiven Eindruck einer weiter bestehenden Verbindung, ohne die für die Übertragung von Sprachsignalen vorgesehene Datenübertragungsrate zu beanspruchen. Mit anderen Worten wird zur senderseitigen Kodierung des Rauschens ein geringerer Aufwand als zur Kodierung der Sprachdaten betrieben. Für eine empfängerseitig noch als realistisch empfundene Synthetisierung des Komfortrauschens werden Daten mit einer weitaus niedrigeren Datenrate übertragen. Die hierbei übertragenen Daten werden in der Fachwelt auch als SID (Silence Insertion Description) bezeichnet.

**[0012]** Gegenwärtige skalierbare Kodierungsverfahren für breitbandige Sprachcodecs sehen derzeit keine Verfahren zur diskontinuierlichen Übertragung vor.

**[0013]** Im Stand der Technik bestehen Probleme mit einer Anwendung des diskontinuierlichen Übertragung (DTX) in Verbindung mit einem Komfortrauschgenerator auf Empfängerseite (CNG Comfort Noise Generator).

**[0014]** Derzeit bekannte Verfahren zur diskontinuierlichen Übertragung sehen eine Übermittlung SID-Rahmen mit aktualisierten Parametern zur Charakterisierung des Hintergrundrauschens nur dann vor, wenn seitens des Enkoders signifikante Änderungen in der Energie des Hintergrundrauschens während einer inaktiven Sprachperiode (Sprechpause) detektiert werden. Dies

betrifft sowohl schmalbandige (50Hz bis 4kHz) als auch breitbandige Sprachcodecs, welche Verfahren zur diskontinuierlichen Übertragung unterstützen. Üblicherweise wird bei der Entscheidung, einen SID-Rahmen mit aktualisierten Parametern zu übermitteln, ein im Dekoder spezifizierter Energiegrenzwert (Energy Threshold) herangezogen. Dies führt dazu, dass bei einer Nichtüberschreitung des definierten Energiegrenzwertes keine SID-Rahmen gesendet werden. Seitens des Übertragungsnetzwerks zwischen Empfänger und Sender wird jedoch ein derartiges Aussetzen des Sendens von SID-Rahmen als Ruhezustand bzw. »Idle Channel« angesehen. Zur Gewährleistung einer Aufrechterhaltung der Verbindung (»Connection Alive«) ist dann eventuell ein zusätzlicher Datenaustausch erforderlich, um anzuzeigen, dass die Verbindung aufrechterhalten werden soll.

**[0015]** Ein bekannter zusätzlich vorgesehener Datenaustausch erfolgt derzeit in der Weise, dass administrative Stellen im Netzwerkmanagement des Übertragungsnetzwerks den sendenden Knoten, d.h. den sendenden Encoder auffordern, den zuletzt übermittelten SID-Rahmen erneut zu übermitteln, falls die verstrichene Leerlaufzeit (»Idle Period«) zum letzten gesendeten SID-Rahmen als zu lang für die entsprechende Verbindung erachtet wird. Für eine solche erneute Übermittlung werden Parameter des erneut gesendeten SID-Rahmens nicht aktualisiert. Der Encoder führt also keine zusätzlichen Aktionen aus.

**[0016]** Aufgabe der Erfindung ist es, eine verbesserte Implementierung der diskontinuierlichen Übertragung in skalierbaren Sprachcodecs anzugeben.

**[0017]** Die Aufgabe wird durch den Gegenstand der unabhängigen Ansprüche gelöst.

**[0018]** Ein Grundgedanke der Erfindung besteht darin, den Encoder eines Sprachcodecs so auszugestalten, dass dieser nach einer zuvor ermittelten Leerlaufzeit (»Idle Period«) eine erneute Ermittlung bzw. Berechnung von Parameter über das Hintergrundrauschen, insbesondere der gemittelten Energie und der Autokorrelationsfunktion, vornimmt. Die besagte Ermittlung der Hintergrundrauschparameter entspricht mit anderen Worten einer Enkodierung des Rauschsignals. Administrative Stellen im Netzwerk informieren den Encoder dabei über die im Übertragungsnetzwerk eingestellte Leerlaufzeit. Der Encoder bestimmt also die Leerlaufzeit z.B. durch Anfrage administrativer Stellen im Übertragungsnetzwerk. Eine solche Anfrage ist nur einmal notwendig, wenn die ermittelte Leerlaufzeit seitens des Encoders gespeichert wird.

**[0019]** Eine Einstellung eines zeitlichen Abstands für zu sendende SID-Rahmen gestattet es administrativen Stellen im Übertragungsnetzwerk, den Encoder zu einem Senden eines aktualisierten Rahmens zu zwingen. Dies garantiert sowohl eine Aktualisierung zugunsten einer besseren Rekonstruktion des Hintergrundrauschens im CNG als auch ein zuverlässigeres Halten der Verbindung.

**[0020]** Ein Vorteil des erfindungsgemäßen Verfahrens besteht darin, dass zur Entscheidung, ob aktualisierte Hintergrundrauschparameter in Form eines aktualisierten SID-Rahmen gesendet werden sollen, kein Vergleich der Energie des Hintergrundrauschsignals mit einem Energiegrenzwert erforderlich ist. Das Verfahren spart somit Rechenressourcen gegenüber den bekannten Verfahren.

**[0021]** Ein weiterer Vorteil besteht darin, dass die eingestellte Zeitdauer zwischen zwei SID-Rahmen mit den Erfordernissen des jeweiligen Übertragungsnetzwerks übereinstimmt.

**[0022]** Vorteilhafte Weiterbildungen und Ausgestaltungen der Erfindung sind Gegenstand der Unteransprüche.

**[0023]** Eine vorteilhafte Ausgestaltung der Erfindung sieht eine SID-Struktur (SID Bitstream Structure) bei der der schmalbandige Anteil der Hintergrundrauschinformation vom breitbandigen Anteil der Hintergrundrauschinformation getrennt ist. Eine getrennte Behandlung von schmalbandiger und breitbandiger Hintergrundrauschinformation in einem SID-Rahmen ermöglicht eine getrennte Enkodierung des schmalbandigen und des breitbandigen Anteils des Hintergrundrauschens und macht die Bearbeitung transparent. Diese Ausgestaltung hat weiterhin den Vorteil, dass empfängerseitig bestimmt werden kann, ob ein Komfortauschen auf Basis des breitbandigen Anteils der übertragenen SID-Rahmen oder auf Basis des schmalbandigen Anteils erfolgen soll. Dies ist von besonderem Vorteil für die empfängerseitige akustische Rezeption in einer Situation, in der die Übertragungsrate für Sprachinformationsrahmen derart verringert wurde, dass nur noch schmalbandige Sprachinformationen übertragen werden. Wird nämlich, wie im derzeitigen Stand der Technik, schmalbandige Sprachinformationen in Verbindung mit breitbandigem Rauschen synthetisiert, ist dies für den Empfänger sehr irritierend. Die besagte Verringerung der Übertragungsrate für Sprachinformationsrahmen kann zum Beispiel durch eine hohe Auslastung (Congestion) des Netzwerks zwischen Sender und Empfänger verursacht sein. Die wesentlich kleineren SID-Rahmen sind von einem solchen Netzwerkengpass nicht betroffen. Für sie besteht also weder ein Zwang zur Reduzierung ihrer Datenübertragungsrate noch ihres Inhalts.

**[0024]** Eine vorteilhafte Ausgestaltung der Erfindung sieht vor, dass zur Bestimmung der Hintergrundrauschparameter des schmalbandigen ersten Anteils des Hintergrundrauschens Energie und Autokorrelationsfunktion des Hintergrundrauschens ermittelt werden. Im schmalbandigen Anteil ist eine Mittlung über einen relativ langen Zeitraum einer Sprechpause notwendig, in der Praxis über einen Zeitraum von z.B. 100 ms. Die verwendeten Berechnungsgrößen gemäß dieser Ausführungsform umfassen dabei die Energie (nicht die logarithmierte Energie) und die Autokorrelationsfunktion.

**[0025]** Zu Beginn eines Zeitabschnitts, welcher als inaktiv bzw. als Sprechpause klassifiziert ist, wird gemäß

einer weiteren vorteilhaften Ausgestaltung der Erfindung eine zusätzliche Überhangperiode (Hangover Period) eingeführt. Die neu eingeführte Überhangperiode, im Folgenden: DTX-Überhangperiode dient im Vergleich zur bisher bekannten VAD-Überhangperiode (Voice Activity Detection) einem weiteren bislang unbekanntem Zweck. Während beide Arten von Überhangperiode das Ziel verfolgen, mehrere Rahmen als aktive Sprachrahmen zu kennzeichnen und damit eine falsche Klassifizierung am Ende eines Sprachsignals zu vermeiden, hat die DTX-Überhangperiode den zusätzlichen Zweck, Informationen über das Hintergrundrauschen zu erheben.

**[0026]** Eine vorteilhafte Ausgestaltung der Erfindung sieht vor, dass der breitbandige zweite Anteil gedämpft wird. Die Dämpfung des breitbandigen Anteils spielt eine Rolle bei der Dämpfung des gesamten Energieanteils im breitbandigen Anteil. Diese Maßnahme ist aufgrund der Tatsache notwendig, dass der Generator zur Erzeugung (synthetisieren) des Komfortrauschens im Dekoder nicht in der Lage ist, dieselben Rauscheigenschaften zu produzieren als das Originalhintergrundrauschen im Enkoder.

**[0027]** Eine vorteilhafte Ausgestaltung der Erfindung sieht vor, dass auf das gesamte Hintergrundrauschsignal, also der Kombination aus breitbandigen und schmalbandigen Anteil, eine nachgeschaltete Betonungsminderungsfilterung (»De-emphasis Post Filter«) angewandt wird. Das »De-Emphasis Post Filter« führt zu einer Betonungsminderung (De-Emphasis) der Energie und der höheren Frequenzkomponenten. Da die Mittelung in bestimmter Weise die spektrale Einhüllende deformiert, kann diese Dämpfung in vorteilhafter Weise dazu beitragen, den störenden Effekt eines gestörten breitbandigen Rauschens auf einen menschlichen Empfänger zu reduzieren.

**[0028]** Ein Ausführungsbeispiel mit weiteren Vorteilen und Ausgestaltungen der Erfindung wird im Folgenden anhand der Zeichnung näher erläutert.

**[0029]** Dabei zeigt die einzige FIG eine zeitliche Darstellung eines Übergangs von einem als Sprache klassifizierten zu einem als Hintergrundrauschen klassifizierten Eingangssignal an einem Dekoder.

**[0030]** Im Folgenden wird der der Erfindung zugrundeliegende technische Hintergrund, zunächst ohne Bezugnahme auf die Zeichnung, näher beschrieben.

**[0031]** Im Stand der Technik bestehen Probleme mit einer Anwendung des diskontinuierlichen Übertragung (DTX) in Verbindung mit einem Komfortrauschgenerator auf Empfängerseite (CNG Comfort Noise Generator). Während der DTX/CNG-Operation müssen folgende Überlegungen berücksichtigt werden:

1. Seitens des CNG ist eine geeignete Erzeugung des Hintergrundrauschens bzw. Komfortrauschens erforderlich, welche von einem Zuhörer auf Empfängerseite als realistisch aufgefasst werden soll. Im Falle von breitbandigen Sprachcodecs, also beispielsweise Sprachcodecs mit einer Bandbreite zwi-

schen Frequenzen von 50 Hz und 7 kHz, wird eine Generierung eines breitbandigen Rauschens als eine Verschlechterung angesehen. Darüber hinaus ist der Charakter bzw. »die Farbe« des Hintergrundrauschens auf Dekoder- und Enkoderseite nicht immer gleich, sodass gegenwärtige Lösungen, welche eine Mittelwertbildung der Energie und der spektralen Einhüllenden vorsehen, eine Verfälschung der ursprünglichen Hintergrundrauschinformation bewirken.

2. Das DTX-Verfahren übermittelt nur dann aktualisierte SID-Rahmen wenn seitens des Enkoders signifikante Änderungen in der Energie des Hintergrundrauschens während einer inaktiven Sprachperiode (Sprechpause) detektiert werden. Dies betrifft sowohl schmalbandige (50Hz bis 4kHz) als auch breitbandige Sprachcodecs, welche das DTX/CNG-Verfahren unterstützen. Üblicherweise spielt dabei ein Energiegrenzwert (Energy Threshold) eine zentrale Rolle. Dies führt dazu, dass bei einer Nichtüberschreitung eines definierten Energiegrenzwertes keine SID-Rahmen gesendet werden. Seitens des Übertragungsnetzwerks zwischen Empfänger und Sender wird jedoch ein derartiges Aussetzen des Sendens von SID-Rahmen als Ruhezustand bzw. »Idle Channel« angesehen. Zur Gewährleistung einer Aufrechterhaltung der Verbindung (»Connection Alive«) ist dann eventuell ein zusätzlicher Datenaustausch erforderlich, um anzuzeigen, dass die Verbindung aufrechterhalten werden soll.

**[0032]** Derzeit wird mit dem oben genannten Problemen wie folgt umgegangen:

Zu 1.: Die den breitbandigen Anteil betreffende Information wird in dem SID-Rahmen enkodiert. Dabei werden die gemittelte logarithmische Energie und die gemittelte Immitance Spectral Frequency (ISF) zur Beschreibung des breitbandigen Hintergrundrauschens herangezogen, z.B. in den Sprachcodecs G.722.2 und AMR-WB. Es ist dabei keine getrennte Behandlung eines unteren Teils und eines oberen Teils des breitbandigen Hintergrundrauschens vorgesehen. Der schmalbandige Sprachcode G.729 verwendet eine gemittelte logarithmische Energie und eine gemittelte Autokorrelationsfunktion. Die Mittelungsperiode für die Energie und die Mittelungsperiode für die Autokorrelationsfunktion stimmen dabei nicht überein.

Zu 2.: Administrative Stellen im Netzwerkmanagement fordern den sendenden Knoten, d.h. den sendenden Enkoder auf, den zuletzt übermittelten SID-Rahmen erneut zu übermitteln, falls die »Idle Period« als zu lang für die zugehörige Verbindung erachtet wird. Der erneut gesendete SID-Rahmen und die darin enthaltene Information wird daher nicht

aktualisiert. Der Enkoder führt also keine zusätzlichen Aktionen aus.

**[0033]** Das erfinderische Verfahren sieht vor, den Enkoder so auszugestalten, dass dieser nach einer bestimmten gegebenen Zeit eine erneute Berechnung der gemittelten Energie und der Autokorrelationsfunktion vornimmt. Administrative Stellen im Netzwerk informieren den Enkoder dabei über die benötigte Idle Time.

**[0034]** Im Folgenden werden weitere Ausführungsformen zur Generierung des SID-Rahmens beschrieben.

**[0035]** Eine SID-Struktur (SID Bitstream Structure) bei der der schmalbandige Anteil der Hintergrundrauschinformation vom breitbandigen Anteil der Hintergrundrauschinformation getrennt ist, wird erzeugt. Eine getrennte Behandlung von schmalbandiger und breitbandiger Hintergrundrauschinformation in einem SID-Rahmen ermöglicht eine getrennte Einkodierung des schmalbandigen und des breitbandigen Anteils des Hintergrundrauschens und macht die Bearbeitung transparent.

**[0036]** Im schmalbandigen Anteil ist eine Mittelung über einen relativ langen Zeitraum einer Sprechpause notwendig, in der Praxis über einen Zeitraum von z.B. 100 ms. Die verwendeten Berechnungsgrößen umfassen dabei die Energie (nicht die logarithmierte Energie) und die Autokorrelationsfunktion. Die Autokorrelationsfunktion wird benutzt für eine spektrale Hüllkurvenpräsentation. Ein Gesamtverstärkungsfaktor kann dabei durch eine Kombination aller Verstärkungs- und Mittelungsmethoden kompensiert werden. Die Werte für die Autokorrelationsfunktion werden durch Summierung oder Mittelwertbildung jeweils normiert (Equally Weighted). Dies betrifft alle SID-Rahmen. Eine relativ lange Mittelung (Averaging) des schmalbandigen Anteils führt zu einer Glättung der schmalbandigen Energie und der spektralen Einhüllenden, sodass eine plötzliche Energieänderung keinen merklichen Einfluss auf die Synthetisierung des Komfortrauschens im Empfänger zeitigt. Dieselbe Mittelungsperiode wird sowohl für die Energie als auch zur Mittelung der spektralen Einhüllenden verwendet, nachdem ein erster SID-Rahmen nach einem Einsetzen eines Sprachsignals (Speak Burst) erzeugt wird. Diese Maßnahme sichert eine konsistentere Abschätzung des schmalbandigen Hintergrundrauschens während einem Übergang von einer Sprachperiode in eine Sprechpause.

**[0037]** Im Folgenden wird auf die FIG Bezug genommen. Die FIG zeigt ein Sprachsignal (Speech Burst), dass zu einem bestimmten Zeitpunkt  $t$  einen bestimmten Signalpegel, Threshold, in der Zeichnung als strichlierte Linie dargestellt, unterschreitet. Die Ordinate ist als Pegel oder als Energiewert des Signals zu verstehen. Hierzu kommt auf Seiten des Senders eine Sprechpausenerkennung (Voice Activity Detection, VAD) zum Einsatz, welche bei Unterschreiten des Thresholds auf eine Sprechpause erkennt. Das VAD-Verfahren sieht eine bekannte Überhangperiode VAD-HO vor, in welcher weiterhin aktive Sprachrahmen gesendet werden und erst nach üblicherweise zwei Rahmenlängen in einen Modus

überwechselt, der eine Generierung von SID-Rahmen vorsieht.

**[0038]** Gemäß der hier beschriebenen Ausführungsform der Erfindung wird eine zusätzliche Überhangperiode DTX-HO eingeführt. Die neue Überhangperiode DTX-HO schließt sich an die bislang bekannte Überhangperiode VAD-HO an, welche als »Black Box« verwendet wird. Während dieser Überhangperiode DTX-HO wird das im Enkoder bearbeitete Signal immer noch als Sprachsignal klassifiziert, während parallel dazu bereits eine Bestimmung von Hintergrundrauschparametern beginnt. Die Datenrate der Sprachkodierung ist bereits reduziert, da zu Beginn einer Sprechpause keine hochqualitative Einkodierung benötigt wird. Weiterhin wird für den schmalbandigen Anteil ein Teil der Überhangperiode benutzt für die Mittelwertbildung des ersten SID-Rahmens. Die vorgenannten Ausführungen beziehen sich vorzugsweise auf die letzten Rahmen FRAMES innerhalb einer Überhangperiode DTX-HO, VAD-HO. Die Information der ersten Rahmen der Überhangperiode wird dagegen vorzugsweise nicht benutzt.

**[0039]** Die neu eingeführte Überhangperiode DTX-HO dient im Vergleich zur bisher von Bedürfnissen der Voice Activity Detection motivierten bekannten Überhangperiode VAD-HO einem weiteren bislang nicht beachteten Zweck. Während beide Arten von Überhangperioden DTX-HO, VAD-HO das Ziel verfolgen, mehrere Rahmen als aktive Sprachrahmen zu kennzeichnen und damit eine falsche Klassifizierung am Ende eines Sprachsignals zu vermeiden, hat die DTX-Überhangperiode DTX-HO den zusätzlichen Zweck, Informationen über das Hintergrundrauschen zu erheben.

**[0040]** Bezüglich des verfolgten Ziels, eine falsche Klassifizierung am Ende eines Sprachsignals zu vermeiden, stellt die neue Überhangperiode DTX-HO eine zusätzliche Versicherung dar, dass nach Ablauf der Überhangperiode DTX-HO definitiv ein Hintergrundrauschen und keine Sprachsignale am Eingang des Dekoders anliegen. Bei einer bisherigen Verwendung der bekannten Überhangperiode VAD-HO konnte nicht ausgeschlossen werden, dass es sich bei dem anliegenden Signal exklusiv nur um Hintergrundrauschen handelte. In der Praxis konnten während dieser bekannten Überhangperiode VAD-HO noch Sprachanteile (Speech Bursts) auftreten. Im Übrigen dient die neue Überhangperiode DTX-HO ausschließlich zur Anlernung des Hintergrundrauschens.

**[0041]** Bezüglich der Wahl der Zeitdauer dieser Überhangperioden DTX-HO, VAD-HO und damit der Wahl der Anzahl an Rahmen FRAMES ist eine vorteilhafte Einstellung z.B. so zu wählen, dass eine Zeitdauer von zwei Rahmen - vgl. gestrichelte Achse FRAMES - für die bekannte Überhangperiode VAD-HO und eine Zeitdauer von fünf Rahmen für die neue Überhangperiode DTX-HO vorgesehen ist.

**[0042]** Im breitbandigen Anteil wird eine Energiedämpfung ausgeführt. Die Dämpfung des breitbandigen Anteils spielt eine Rolle bei der Dämpfung des gesamten

Energieanteils im breitbandigen Anteil. Diese Maßnahme ist aufgrund der Tatsache notwendig, dass der Generator zur Erzeugung (synthetisieren) des Komfortrauschens im Dekoder nicht in der Lage ist, dieselben Rauscheigenschaften zu produzieren als das Originalhintergrundrauschen im Enkoder.

**[0043]** Auf das ausgegebene Breitbandsprachsignal, also der Kombination aus breitbandigen und schmalbandigen Anteil, wird eine nachgeschaltete Betonungsminderungsfilterung (»De-emphasis Post Filter«) angewandt. Diese Filterung dämpft hauptsächlich höhere Frequenzkomponenten. Das »De-Emphasis Post Filter« führt weiterhin zu einer Betonungsminderung (De-Emphasis) der Energie und der höheren Frequenzkomponenten. Da die Mittelung in bestimmter Weise die spektrale Einhüllende deformiert, kann diese Dämpfung dazu beitragen, den störenden Effekt eines gestörten breitbandigen Rauschens auf einen menschlichen Empfänger zu reduzieren.

#### Patentansprüche

1. Verfahren zur Generierung von SID-Rahmen für eine diskontinuierliche Übertragung von Hintergrundrauschparametern über ein Übertragungsnetzwerk, mit dem in der Telekommunikation Audiosignale von einem Sender an einen Empfänger übertragen werden, wobei bei dem Verfahren eine periodenweise Ermittlung von Hintergrundrauschparametern und eine auf Basis der ermittelten Hintergrundrauschparameter erfolgende Generierung und Übersendung von SID-Rahmen vorgesehen ist, wobei die Periode einer ermittelten, im Übertragungsnetzwerk eingestellten Leerlaufzeit (Idle Period) des Übertragungsnetzwerks entspricht.
2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** als Audiosignal ein Sprachsignal verwendet wird.
3. Verfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** zur Entscheidung, ob die Hintergrundrauschparameter in Form eines der SID-Rahmen übersandt werden sollen, anstelle eines Vergleichs der Energie eines Hintergrundrauschsignals mit einem Energiegrenzwert eine Einstellung eines zeitlichen Abstandes zwischen zwei zu übersendenden SID-Rahmen erfolgt.
4. Verfahren nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet, dass** Hintergrundrauschparameter eines schmalbandigen ersten Anteils und eines breitbandigen zweiten Anteils ermittelt werden und dass die Generierung des SID-Rahmens mit getrennten Bereichen für den ersten und den zweiten Anteil erfolgt.
5. Verfahren nach Anspruch 4, **dadurch gekennzeichnet, dass** seitens des Empfängers bestimmt wird, ob ein Komfortrauschen auf Basis des breitbandigen zweiten Anteils der übertragenen SID-Rahmen oder auf Basis des schmalbandigen ersten Anteils erfolgen soll.
6. Verfahren nach Anspruch 4 oder 5, **dadurch gekennzeichnet, dass** zur Bestimmung der Hintergrundrauschparameter des schmalbandigen ersten Anteils des Hintergrundrauschens Energie und Autokorrelationsfunktion des Hintergrundrauschens ermittelt werden.
7. Verfahren nach Anspruch 6, **dadurch gekennzeichnet, dass** die Hintergrundrauschparameter des schmalbandigen ersten Anteils über einen Zeitraum um 100 Millisekunden gemittelt werden.
8. Verfahren nach Anspruch 6 oder 7, **dadurch gekennzeichnet, dass** anstelle einer logarithmierten Energie des Hintergrundrauschens die Energie des Hintergrundrauschens ermittelt wird.
9. Verfahren nach einem der vorgenannten Ansprüche, **dadurch gekennzeichnet, dass** beim Übergang von einem als Sprache kategorisierten Signal zu einem als Hintergrundrauschen kategorisierten Signal eine zusätzliche Überhangperiode (Hangover Period, DTX-HO) vorgesehen ist, während der eine Bestimmung von Hintergrundrauschparametern erfolgt.
10. Verfahren nach Anspruch 9, **dadurch gekennzeichnet, dass** sich die zusätzliche Überhangperiode (DTX-HO) an eine bekannte Überhangperiode (VAD-HO) anschließt, in welcher weiterhin aktive Sprachrahmen gesendet werden.
11. Verfahren nach Anspruch 10, **dadurch gekennzeichnet, dass** die zusätzliche Überhangperiode (DTX-HO) und die bekannte Überhangperiode (VAD-HO) jeweils Rahmen (FRAMES) aufweisen und die Anzahl der Rahmen (FRAMES) derart gewählt wird, dass eine Zeitdauer von zwei Rahmen für die bekannte Überhangperiode (VAD-HO) und eine Zeitdauer von fünf Rahmen für die zusätzliche Überhangperiode (DTX-HO) vorgesehen wird.
12. Verfahren nach einem der Ansprüche 4 bis 9, **dadurch gekennzeichnet, dass** der breitbandige zweite Anteil gedämpft wird.
13. Verfahren nach einem der vorgenannten Ansprüche, **dadurch gekennzeichnet, dass** auf das gesamte Hintergrundrauschsignal eine nachgeschaltete Betonungsminderungsfilterung (De-Emphasis Post Filter) angewandt wird.

14. Codec mit Mitteln zur Durchführung des Verfahrens nach einem der Ansprüche 1 bis 13.
15. Codec nach Anspruch 14, **gekennzeichnet durch** eine Implementierung im an sich bekannten ITU-T Standard G.729.1 (05/06) .

### Claims

1. A method for the generation of SID frames for a discontinuous transmission of background noise parameters via a transmission network with which audio signals are transmitted by a sender to a recipient during telecommunication, said method involving a periodic determination of background noise parameters and a generation and transmission of SID frames based on the determined background noise parameters, wherein the period corresponds to a determined idle period of the transmission network that is adjusted in the transmission network.
2. The method according to claim 1, **characterized in that** a speech signal is used as the audio signal.
3. The method according to claim 1 or 2, **characterized in that**, in the decision as to whether the background noise parameters should be sent in the form of an SID frame, an interval of time between two SID frames to be sent is adjusted, rather than comparing the energy of a background noise signal with an energy threshold.
4. The method according to any of claims 1 to 3, **characterized in that** background noise parameters of a narrow band first portion and of a wide band second portion are determined, and **in that** the SID frame is generated with separate areas for the first portion and the second portion.
5. The method according to claim 4, **characterized in that** the recipient determines whether a comfort noise based on the wide band second portion of the transmitted SID frame or based on the narrow band first portion should occur.
6. The method according to claim 4 or 5, **characterized in that** the background noise parameters of the narrow band first portion of the background noise are determined by determining an energy and an auto-correlation function of the background noise.
7. The method according to claim 6, **characterized in that** the background noise parameters of the narrow band first portion are averaged over a period in 100 millisecond increments.
8. The method according to claim 6 or 7, **characterized**

**in that** the energy of the background noise, rather than a logarithmized energy of the background noise, is determined.

9. The method according to any of the preceding claims, **characterized in that** during the transition from a signal categorized as speech to a signal categorized as background noise, an additional hangover period (DTX-HO) is provided, during which background noise parameters are determined.
10. The method according to claim 9, **characterized in that** the additional hangover period (DTX-HO) follows a known hangover period (VAD-HO), in which active speech frames continue to be sent.
11. The method according to claim 10, **characterized in that** the additional hangover period (DTX-HO) and the known hangover period (VAD-HO) each include frames, and the number of frames is selected in such a way that a duration of two frames is provided for the known hangover period (VAD-HO) and a duration of five frames is provided for the additional hangover period (DTX-HO).
12. The method according to any of claims 4 to 9, **characterized in that** the wide band second portion is attenuated.
13. The method according to any of the preceding claims, **characterized in that** a subsequent de-emphasis post filter is applied to the entire background noise signal.
14. A codec having means for carrying out the method according to any of claims 1 to 13.
15. The codec according to claim 14, **characterized by** an implementation in the known ITU-T Standard G.729.1(05/06).

### Revendications

1. Procédé servant à générer des trames SID en vue d'une transmission discontinue de paramètres de bruit de fond par l'intermédiaire d'un réseau de transmission, à l'aide duquel des signaux audio sont transmis d'un émetteur à un récepteur lors de la télécommunication, dans lequel une détermination périodique de paramètres de bruit de fond et une génération effectuée sur la base des paramètres de bruit de fond déterminés et un envoi de trames SID sont prévus dans le procédé, dans lequel la période correspond à un temps d'inactivité (Idle Period), réglé dans le réseau de transmission, du réseau de transmission.

2. Procédé selon la revendication 1, **caractérisé en ce qu'un signal vocal est utilisé en tant que signal audio.**
3. Procédé selon la revendication 1 ou 2, **caractérisé en ce que** pour décider si les paramètres de bruit de fond doivent être envoyés sous la forme d'une des trames SID, un réglage d'un intervalle de temps entre deux trames SID à envoyer est effectué en lieu et place d'une comparaison de l'énergie d'un signal de bruit de fond à une valeur limite d'énergie. 5
4. Procédé selon l'une quelconque des revendications 1 à 3, **caractérisé en ce que** des paramètres de bruit de fond d'une première partie à bande étroite et d'une deuxième partie à bande large sont déterminés, et **en ce que** la génération de la trame SID est effectuée avec des zones séparées pour la première et la deuxième partie. 10
5. Procédé selon la revendication 4, **caractérisé en ce qu'il est défini du côté du récepteur si un bruit de confort doit être effectué sur la base de la deuxième partie à bande large des trames SID transmises ou sur la base de la première partie à bande étroite.** 15
6. Procédé selon la revendication 4 ou 5, **caractérisé en ce que** pour définir les paramètres de bruit de fond de la première partie à bande étroite du bruit de fond, une énergie et une fonction d'autocorrélation du bruit de fond sont déterminées. 20
7. Procédé selon la revendication 6, **caractérisé en ce qu'une moyenne des paramètres de bruit de fond de la première partie à bande étroite est établie sur une période de 100 millisecondes.** 25
8. Procédé selon la revendication 6 ou 7, **caractérisé en ce que** l'énergie du bruit de fond est déterminée en lieu et place d'une énergie logarithmique du bruit de fond. 30
9. Procédé selon l'une quelconque des revendications précédentes, **caractérisé en ce qu'une période résiduelle supplémentaire (Hangover Period, DTX-HO) est prévue lors du passage d'un signal classé comme voix à un signal classé comme bruit de fond, au cours de laquelle période résiduelle une définition de paramètres de bruit de fond est effectuée.** 35
10. Procédé selon la revendication 9, **caractérisé en ce que** la période résiduelle supplémentaire (DTX-HO) se situe dans le prolongement d'une période résiduelle connue (VAD-HO), dans laquelle par ailleurs des trames vocales actives sont envoyées. 40
11. Procédé selon la revendication 10, **caractérisé en ce que** la période résiduelle supplémentaire (DTX-HO) et la période résiduelle connue (VAD-HO) pré- 45
- sentent respectivement des trames (FRAMES), et **en ce que** le nombre des trames (FRAMES) est choisi de telle manière que sont prévues une durée de deux trames pour la période résiduelle connue (VAD-HO) et une durée de cinq trames pour la période résiduelle supplémentaire (DTX-HO).
12. Procédé selon l'une quelconque des revendications 4 à 9, **caractérisé en ce que** la deuxième partie à bande large est atténuée. 50
13. Procédé selon l'une quelconque des revendications précédentes, **caractérisé en ce qu'un filtre de désaccentuation (De-Emphasis Post Filter) installé en aval est appliqué sur l'ensemble du signal de bruit de fond.** 55
14. Codec comprenant des moyens servant à mettre en oeuvre le procédé selon l'une quelconque des revendications 1 à 13.
15. Codec selon la revendication 14, **caractérisé par** une implémentation dans la recommandation UIT-T connue en soi G.729.1 (05/06).

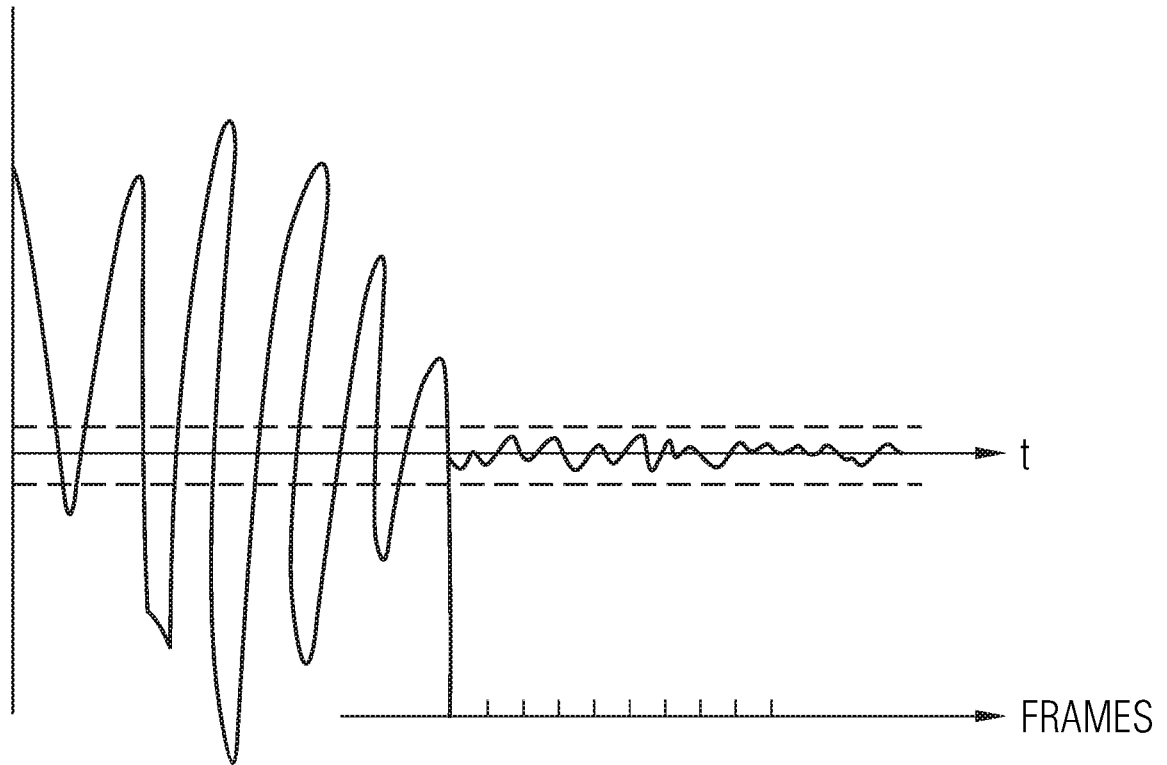


FIG. 1

**IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE**

*Diese Liste der vom Anmelder aufgeführten Dokumente wurde ausschließlich zur Information des Lesers aufgenommen und ist nicht Bestandteil des europäischen Patentdokumentes. Sie wurde mit größter Sorgfalt zusammengestellt; das EPA übernimmt jedoch keinerlei Haftung für etwaige Fehler oder Auslassungen.*

**In der Beschreibung aufgeführte Patentdokumente**

- US 20080027716 A1 [0008]