

(19)



(11)

EP 2 245 622 B1

(12)

EUROPÄISCHE PATENTSCHRIFT

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
13.07.2016 Patentblatt 2016/28

(51) Int Cl.:
G10L 19/012 (2013.01) G10L 19/24 (2013.01)

(21) Anmeldenummer: **09712583.5**

(86) Internationale Anmeldenummer:
PCT/EP2009/051120

(22) Anmeldetag: **02.02.2009**

(87) Internationale Veröffentlichungsnummer:
WO 2009/103609 (27.08.2009 Gazette 2009/35)

(54) VERFAHREN UND MITTEL ZUR DEKODIERUNG VON HINTERGRUNDRAUSCHINFORMATIONEN

METHOD AND MEANS FOR DECODING BACKGROUND NOISE INFORMATION

PROCÉDÉS ET MOYENS POUR DÉCODER DES 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

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

(30) Priorität: **19.02.2008 DE 102008009720**

(56) Entgegenhaltungen:
WO-A2-2007/064256 US-A1- 2006 293 885
US-A1- 2008 195 383

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

(73) Patentinhaber: **Unify GmbH & Co. KG**
81379 München (DE)

- **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, 1. Juni 2008 (2008-06-01), Seiten 1-36, XP002526623**
- **SOLLAUD A: "G.729.1 RTP Payload Format update: DTX support" INTERNET CITATION 8. Februar 2008 (2008-02-08), XP002526621**
Gefunden im Internet:
URL: <http://www.ietf.org.net:8080/proceedings/08mar/IDs/draft-ietf-avt-rfc4749-dtx-update-00.txt> [gefunden am 2008-02-08]

(72) Erfinder:

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

EP 2 245 622 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 Dekodierung 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 Bandbreiteabgrenzung des analogen Signals durchgeführt. Zur Kodierung und zur Decodierung kommt dabei ein Codec zum Einsatz, welcher aufgrund der beschriebenen Bandbreitenbeschränkung im Frequenzbereich zwischen 300 Hz und 3400 Hz in Folgenden auch als schmalbandiger Sprach-Codec (Narrow Band Speech Codec) bezeichnet wird. Unter dem Begriff Codec wird dabei sowohl die Kodiervorschrift zur digitale Kodierung von Audiosignale als auch die Dekodiertvorschrift zur Dekodierung von Daten mit dem Ziel einer Rekonstruktion des Audiosignals verstanden.

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

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

[0006] Üblicherweise sind Kodierungsverfahren für breitbandige Sprach-Codecs skalierbar gestaltet. Mit einer Skalierbarkeit ist hier gemeint, dass die übertragenen kodierten Daten verschiedene abgegrenzte Blöcke entfalten, 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 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 Höeindrucks möglichst ähnlich ist. Mit Hilfe diesem auch als »Analysis-by-Synthesis« bezeichneten Verfahren 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. 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 Signalpegel 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) abgewandt.

[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 - d.h. Dekodierung - des Komfortrauschens werden Daten mit einer weitaus niedrigeren Datenraten übertragen. Die hierbei übertragenen Daten werden in der Fachwelt auch als SID (Silence Insertion Description) bezeichnet.

[0012] Im Stand der Technik entstehen Probleme mit dem Verfahren zur diskontinuierlichen Übertragung bei einer Verwendung von breitbandigen Sprachcodecs, wie z.B. ITU-T G.729.1, G.72.2.2 oder 3GPP AMR-WB. Die genannten skalierbaren breitbandigen Sprachcodecs unterstützen üblicherweise unterschiedliche Datenübertragungsraten in einem Bandbreitenbereich von 50 bis 7000 Hz.

[0013] Mögliche Datenraten zur Enkodierung der Sprachinformationen sind z.B. 8, 12, 14, 16,..., 32 kbit/s, welche z.B. im Standard G.729.1 zum Einsatz kommen. Die Datenraten von 8 und 12 kbit/s werden auf schmalbandige Signale (50Hz bis 4kHz) angewandt. Datenraten von mehr als 12 kbit/s werden auf das obere Frequenz-

band von 4 bis 7 kHz angewandt.

[0014] Während einer Übertragung ist ein Wechsel zwischen den besagten Datenraten möglich. Ein plötzlicher Wechsel von einer schmalbandigen auf eine breitbandige Datenrate verursacht dabei bekanntermaßen einen störenden Effekt auf einen menschlichen Rezipienten. Ein solcher Übergang erfolgt beispielsweise in Folge einer Beschneidung des Datenstroms (Bitstream Truncation), welcher beispielsweise durch das Übertragungsnetzwerk zwischen Sender und Empfänger verursacht wird, beispielsweise in Folge einer Einrichtung weiterer zusätzlicher Verbindungen oder aufgrund eines Datenstaus (Congestion) im Übertragungsnetzwerk. Die genannte Beschneidung führt zu einer Änderung der Datenrate und schließlich zu einem Übergang von einer breitbandigen zu einer schmalbandigen Übertragung des Sprachsignals.

[0015] Werden im Encoder Verfahren zur diskontinuierlichen Übertragung - bzw. DTX-Verfahren - angewandt, ist eine Einsparung der Datenübertragungsrate für die Übertragung der jeweiligen Datenrahmen möglich. Das DTX-Verfahren kommt genau dann zum Einsatz, wenn ein entsprechender Rahmen als Sprechpause charakterisiert ist. In Anwendung des DTX-Verfahrens wird eine reduzierte Datenübertragungsrate an übertragenen Rahmen aufgrund zweier Faktoren erreicht. Erstens müssen seitens des Encoders nicht alle inaktiven Rahmen an den Dekoder gesandt werden. Zweitens belegt ein gesendeter SID-Rahmen bzw. inaktiver Rahmen viel weniger Bits als ein Sprachdatenrahmen.

[0016] Ein solches Verfahren erfordert eine Beteiligung einer Sprechpausenerkennung (VAD) auf Encoderseite. Mit Hilfe eines Sprechpausendetektors wird dem Encoder auf der Senderseite mitgeteilt, ob ein aktueller Abtastwerte enthaltender und zu enkodierender Rahmen ein Sprachsignal enthält oder eine Sprechpause mit Hintergrundrauschen. Mit Hilfe dieser Charakterisierung werden im Encoder Maßnahmen getroffen, welche die Wahrnehmungscharakteristika (Perceptual Characteristics) eines inaktiven Rahmens (Inactive Speech Frame) ermitteln. Zu solchen Wahrnehmungscharakteristika zählen beispielsweise die gemittelte Energie sowie spektrale und zeitliche Charakteristika.

[0017] Der Encoder sendet daraufhin einen speziell gekennzeichneten Rahmen, einen SID-Rahmen (Silence Insertion Descriptor) an den Dekoder. Der Dekoder synthetisiert auf Basis der in einem SID-Rahmen enthaltenen Informationen ein Komfortrauschen, wobei der Dekoder aufgrund des SID-Rahmens bestimmen kann, ob es sich bei der enthaltenen Rauschinformation um schmalbandige oder breitbandige Informationen handelt.

[0018] Ein Wechsel der Datenrate (»Bitrate Switching«) zwischen schmalbandiger und breitbandiger Information ist ein übliches Szenario für jeden skalierbaren breitbandigen Sprachcodec. Eine Behandlung eines Datenratenwechsels während einer normalen Sprachphase, d.h. in Abwesenheit von Sprechpausen, ist zwar hin-

reichend in der Literatur beschrieben, eine Behandlung bei Eintritt in eine DTX-Phase ist derzeit indes noch unbekannt.

[0019] Es ergibt sich daher ein dringendes Bedürfnis, ein Verfahren zum Datenratenwechsel während einer DTX-Phase und/oder bei Eintritt in eine DTX-Phase anzugeben, um auf einen Wechsel zwischen einer schmalbandigen und breitbandigen Datenrate vor oder während des Übergangs in die DTX-Phase optimal zu reagieren.

[0020] Während einer Sprechpause ist eine Datenratenbeschneidung unwahrscheinlich, da die Datenbelegung (Bitstream Elocation) eines SID-Rahmens ohnehin weniger Bits benötigt als ein aktiver Sprachdatenrahmen in einer »normalen« Codec-Operation, also eine Codec-Operation während einer ausschließlichen Sprechphase.

[0021] Dies führt zu einem möglichen Szenario, in dem die Datenrate während einer aktiven Sprechphase geändert wird, jedoch in Sprechpausen, also während der DTX-Phase in einem breitbandigen Modus verbleibt. Als sehr störend seitens eines menschlichen Empfängers auf Dekoderseite wird dabei der Fall empfunden, bei dem die aktiven Sprachrahmen schmalbandig dekodiert sind und das Hintergrundrauschen in Sprechpausen breitbandig wiedergegeben wird.

[0022] Dieser Fall tritt mit hoher Wahrscheinlichkeit z.B. in Situationen auf, bei denen die enkoderseitig gesendeten Sprachdatenrahmen durch das Übertragungsnetzwerk beschnitten werden, jedoch seitens des Übertragungsnetzwerks noch genügend Kapazität zur Übertragung der breitbandigen SID-Rahmen übrig bleibt.

[0023] Bislang sind keine Verfahren zum Wechsel der Datenrate von SID-Rahmen während einer Sprechpause bekannt. Die bestehenden Verfahren zum Wechsel der Datenrate beziehen sich lediglich auf den normalen Codec-Betrieb während einer aktiven Sprachphase.

[0024] Aufgabe der Erfindung ist es, ein Verfahren zum Wechsel einer Datenrate von SID-Rahmen während einer Sprechpause anzugeben, das eine verbesserte Qualität des am Dekoder synthetisierten Signals zur Folge hat.

[0025] Die Aufgabe wird durch den Gegenstand der unabhängigen Ansprüche gelöst.

[0026] Ein Grundgedanke der Erfindung besteht darin, eine Ermittlung von Informationen zum Verlauf der Bandbreitumschaltung (Bitrate Switching) während einer aktiven Sprachphase vorzunehmen. Die skalierbare Natur des erfindungsgemäß zum Einsatz kommenden Sprachsignalkodierungsverfahren bzw. Codecs drückt dabei die Möglichkeit des Codecs zur Bandbreitenumschaltung bereits aus.

[0027] Erfindungsgemäß werden während der Sprachphase auf Seiten des Dekoders Informationen zum prozentualen Anteil von breitbandigen aktiven Sprachrahmen im Vergleich zu schmalbandigen aktiven Sprachrahmen erhoben. Mit anderen Worten werden nicht erst zum Zeitpunkt eines Wechsels in eine Sprechpause Informationen zum Charakter des Hintergrundrauschens erhoben, wie aus dem Stand der Technik bislang

nahegelegt wurde. Ein höher prozentualer Anteil an breitbandigen aktiven Sprachrahmen zeigt dabei an, dass seitens des Codecs eine breitbandigen Verwendung bevorzugt wird und daher ein Bedarf besteht, während einer DTX-Phase Rauschinformationen breitbandig zu synthetisieren, d.h. zu dekodieren. Wird dagegen ein niedriger prozentualer Anteil ermittelt, wird auf Seiten des Dekoders bei Eintritt in eine DTX-Phase ein schmalbandiges Rauschens erzeugt, sogar dann, wenn die empfangenen SID-Rahmen eine Synthetisierung - d.h.. Dekodierung - eines breitbandigen Rauschens erlauben würden.

[0028] Mit dem hier vorgestellten erfindungsgemäßen Verfahren wird die Aufgabe der Erfindung, ein Verfahren zum Wechsel einer Datenrate von SID-Rahmen während einer Sprechpause anzugeben mehr als gelöst. Der zu lösende Wechsel zwischen Rauschinformationen mit verschiedener Datenrate gemäß der Aufgabe wird gemäß der hier vorgestellten erfinderischen Lösung in eine Bestimmung eines Anteils an Rauschinformationen mit verschiedenen Datenraten verfeinert. Der Anteil ist im Gegensatz zu einem Wechsel in beliebigem Verhältnis zwischen Rauschinformationen mit verschiedener Datenrate einstellbar.

[0029] Durch die Einstellbarkeit bzw. Anpassbarkeit der Rauschsignalqualität auf die zuvor erhobene Sprachsignalqualität (schmalbandig/breitbandig) ergibt für das gesamte Signal, also Rausch- und Sprachsignal, auf Seiten des Empfänger insgesamt eine erheblich gesteigerte Signalqualität. Das erfindungsgemäße Verfahren löst somit die Aufgabe der Erfindung, eine verbesserte Qualität des am Dekoder synthetisierten Signals zu erzielen.

[0030] Ein solcher Ansatz gemäß dem erfindungsgemäßen Verfahren erweist sich als grundlegend für vorteilhafte Weiterbildungen der Erfindung, welche Gegenstand der Unteransprüche sind.

[0031] Wird gemäß des erfindungsgemäßen Verfahrens eine Entscheidung dahingehend getroffen, dass während einer Sprechpause ein Rauschsignal mit einer bestimmten Qualität (d.h. breitbandig oder schmalbandig) synthetisiert wird, kann ein Fall auftreten, bei dem in den letzten wenigen Rahmen während einer aktiven Sprachphase seitens des Netzwerks eine Beschneidung der aktiven Datenrahmen stattfand.

[0032] Zur Erklärung wird zunächst angenommen, dass der verwendete Codec eine breitbandigen Wiedergabeweise bevorzugt und eine breitbandige Übertragungsweise in der Vergangenheit durch das Übertragungsnetzwerk auch überwiegend gewährleistet war. Dies kann zu dem Fall führen, dass wenige aktive Sprachrahmen beim empfangenden Dekoder als schmalbandige Sprachrahmen eintreffen, bevor dort erste SID-Rahmen empfangen werden.

[0033] In diesem Fall würde ohne zusätzliche Maßnahmen ein abrupter Übergang von einem schmalbandigen Sprachsignal auf ein breitbandige Rauschsignal während der ersten wenigen SID-Rahmen erfolgen. So wichtig ein solcher Übergang zur Wiedereinstellung auf eine

breitbandige Empfangsbedingung im Allgemeines ist, wird dieser Übergang jedoch für den Empfänger als störend empfunden.

[0034] Gemäß einer Weiterbildung der Erfindung ist daher vorgesehen, dass bei Eintritt in die DTX-Phase zunächst eine vorwiegend schmalbandige Dekodierung der Hintergrundrauschinformationen erfolgt, welche nach einem einstellbaren Zeitraum in eine vorwiegend breitbandige Dekodierung übergeht. Ein solchen Übergang erfolgt also vorzugsweise quasi-stetig, wobei ein Übergang zu diskreten Zeitpunkten - daher »quasi«-stetig - auf einen bestimmten Anteilfaktor eingestellt wird.

[0035] Gemäß einer Weiterbildung der Erfindung wird ein Verfahren zum schnellen Umschalten (Fast Switching) vorgeschlagen, bei welchem innerhalb eines bestimmten Zeitrahmens von 100 ms ein quasi-stetiger Übergang von einer schmalbandigen (Anteilsfaktor = 0) auf eine breitbandige (Anteilsfaktor = 1) Rauschsignalqualität durchgeführt wird. Dieser Übergang wird seitens des Dekoders durchgeführt.

[0036] Als besonders vorteilhaft für das subjektive menschliche Hörempfinden haben sich gemäß einer Weiterbildung der Erfindung folgende Werte für den Anteilfaktor erwiesen:

zum Zeitpunkt des Eintritts in die DTX-Phase ein Anteilfaktor von 0, mithin ausschließlich schmalbandiges Rauschen;
zu einem Zeitpunkt von 20 ms nach Eintritt in die DTX-Phase ein Anteilfaktor von 0,09525986892242;
zu einem Zeitpunkt von 40 ms nach Eintritt in die DTX-Phase ein Anteilfaktor von 0,19753086419753;
zu einem Zeitpunkt von 60 ms nach Eintritt in die DTX-Phase ein Anteilfaktor von 0,36595031245237;
zu einem Zeitpunkt von 80 ms nach Eintritt in die DTX-Phase ein Anteilfaktor von 0,62429507696997; und;
zu einem Zeitpunkt von 100 ms nach Eintritt in die DTX-Phase ein Anteilfaktor von 1, mithin ausschließlich breitbandiges Rauschen.

[0037] Gemäß einer Weiterbildung der Erfindung wird angenommen, dass der verwendete Codec eine schmalbandige Wiedergabeweise bevorzugt und/oder eine breitbandige Übertragungsweise in der Vergangenheit durch das Übertragungsnetzwerk nicht gewährleistet war. Dies kann zu dem Fall führen, dass wenige aktive Sprachrahmen beim empfangenden Dekoder als breitbandige Sprachrahmen eintreffen, bevor dort erste SID-Rahmen empfangen werden.

[0038] Gemäß einer Weiterbildung der Erfindung ist vorgesehen, dass bei Eintritt in die DTX-Phase zunächst eine vorwiegend breitbandige Dekodierung der Hintergrundrauschinformationen erfolgt, welche nach einem einstellbaren Zeitraum in eine vorwiegend schmalbandi-

ge Dekodierung übergeht. Ein solcher Übergang erfolgt analog zum oben beschriebenen Weiterbildung vorzugsweise quasi-stetig, wobei ein Übergang zu diskreten Zeitpunkten auf einen bestimmten Anteilfaktor eingestellt wird.

[0039] Gemäß einer Weiterbildung der Erfindung wird ein Verfahren zum schnellem Umschalten (Fast Switching) vorgeschlagen, bei welchem innerhalb eines bestimmten Zeitrahmens von 100 ms ein quasi-stetiger Übergang von einer breitbandigen (Anteilfaktor = 1) auf eine schmalbandige (Anteilfaktor = 0) Rauschsignalqualität durchgeführt wird. Dieser Übergang wird seitens des Dekoders durchgeführt.

[0040] Zum quasi-stetigen Übergang von der breitbandigen auf die schmalbandige Rauschsignalqualität wird der Anteilfaktor mit Werten wie oben, allerdings in umgekehrter Reihenfolge eingestellt.

[0041] Ein Ausführungsbeispiel mit weiteren Vorteilen und Ausgestaltungen der Erfindung wird im Folgenden anhand der Zeichnung näher erläutert.

[0042] Dabei zeigen:

Fig. 1 eine zeitliche Darstellung einer Datenrate zwischen einem Sender und einem Empfänger mit mehreren Bandbreitenumschaltungen und einem Eintritt in eine Sprechpause, wobei SID-Rahmen gesendet werden;

Fig. 2A eine schematische Darstellung eines ersten Szenarios einer Bandbreitenumschaltung;

Fig. 2B eine schematische Darstellung eines zweiten Szenarios einer Bandbreitenumschaltung; und;

Fig. 3 Ein dekodierseitig ausgeführter Umschaltvorgang mit einem quasi-stetigen Übergang von einer schmalbandigen auf eine breitbandige Rauschsignalqualität.

[0043] In Fig. 1 ist eine zeitliche Übertragung von Sprachdatenrahmen mit einer jeweiligen Datenrate DR sowie, ab einer dritten Zeitpunkt t3, eine Übertragung von SID-Rahmen dargestellt.

[0044] Vor einem ersten Zeitpunkt t1 erfolgt eine Übertragung von breitbandigen aktiven Sprachrahmen mit einer Datenrate von 32 kbit/s. Ab dem Zeitpunkt t1 erfolgt eine Umschaltung auf eine Datenrate von 22 kbit/s und ab einem zweiten Zeitpunkt t2 auf eine Datenrate von 12 kbit/s. Eine Datenrate von 12 kbit/s entspricht bereits einem schmalbandigen Sprachrahmen.

[0045] Zu einem dritten Zeitpunkt t3 wird angenommen, dass aufgrund einer Sprechpause seitens des Senders ein Eintritt in eine DTX-Phase erfolgt. Ab dem dritten Zeitpunkt t3 werden somit SID-Rahmen SID in einer bestimmten zeitlichen Periode gesendet.

[0046] Ab dem dritten Zeitpunkt t3 tritt nun die zuvor erläuterte Situation ein, dass in der Vergangenheit - wäh-

rend der Zeitphase zwischen dem zweiten Zeitpunkt t2, und dem dritten Zeitpunkt t3 - ein schmalbandiges Sprachsignal übermittelt wurde, wobei ab dem dritten Zeitpunkt t3 nunmehr ein breitbandiges Rauschsignal durch entsprechende SID-Rahmen zur Verfügung gestellt wird. Die Datenrate der SID-Rahmen entspricht 43 bit/20ms = 2,15 kbit/s bei einer Länge von 43 Bit pro SID-Rahmen und einer Periode von 20 ms pro gesendetem SID-Rahmen.

[0047] In dieser Situation tritt der Fall ein, dass seitens des Dekoders ein unmittelbarer, d.h. unstetiger, Übergang von einem schmalbandigen Sprachsignal auf ein breitbandiges Rauschsignal erfolgen würde. Ein solcher abrupter Übergang wird für einen menschlichen Rezipienten als äußerst störend empfunden.

[0048] Figur 2A und Figur 2B zeigen zwei mögliche Szenarien für einen Verlauf der Datenrate DR über die Zeit t.

[0049] In Figur 2A ist aufgrund von Einschränkungen des Netzwerks oder aufgrund anderer Umstände eine Übertragung weitgehend schmalbandig, im Beispiel der Fig. 2A mit 8. kbit/s, während zu wenigen Zeitpunkten, zwischen einem ersten Zeitpunkt t1 und einem zweiten Zeitpunkt t2, ausnahmsweise eine breitbandige Übertragung mit 32 kbit/s erfolgt.

[0050] In Figur 2B wiederum ist eine umgekehrte Situation zu verzeichnen, nämlich eine überwiegend breitbandige Übertragungsweise mit 32 kbit/s und eine ausnahmsweise kurze, zwischen einem vierten Zeitpunkt t4 und einem fünften Zeitpunkt t5 erfolgende, schmalbandige Übertragungsweise.

[0051] Im Folgenden wird zu einem Zeitpunkt t3 für das Beispiel der Fig. 2A sowie zu einem Zeitpunkt t6 für das Beispiel der Fig. 2B angenommen, dass ein Eintritt in eine DTX-Phase erfolgt.

[0052] Gemäß dem erfindungsgemäßen Vorgehen werden während der Sprachphase auf Seiten des Dekoders Informationen zum Anteil von breitbandigen aktiven Sprachrahmen im Vergleich zu schmalbandigen aktiven Sprachrahmen erhoben.

[0053] Für das Beispiel der Fig. 2A ist dabei der prozentuale Anteil von breitbandigen aktiven Sprachrahmen als sehr niedrig zu bezeichnen, während im Beispiel der Fig. 2B ein hoher prozentualer Anteil von breitbandigen aktiven Sprachrahmen vorliegt.

[0054] Bei Eintritt in eine DTX-Phase zum Zeitpunkt t3 im Beispiel der Fig. 2A wird nun, unter Anwendung des erfindungsgemäßen Verfahrens ein schmalbandiges Rauschen erzeugt, obwohl die ab dem Zeitpunkt t3 empfangenen - nicht dargestellten - SID-Rahmen eine Synthesisierung eines breitbandigen Rauschens erlauben würden.

[0055] Im Beispiel der Fig. 2B dagegen werden zu einem Zeitpunkt t6 mit der dort beginnenden DTX-Phase eine breitbandige Synthesisierung der Rauschinformationen bevorzugt.

[0056] In Fig. 3 ist eine Rauschsignalqualität HB-SHARE über eine Zeit TIME, welche in ms angegeben

ist, aufgetragen. Die Fig. 3 stellt dabei eine Gestaltung des Rauschsignals in Anschluss an ein Szenario gemäß der vorangegangenen Figur 2B dar, bei der aufgrund des dekodierseitig ermittelten prozentualen Anteils von breitbandigen aktiven Sprachrahmen ein Bedarf ermittelt wurde, während der DTX-Phase Rauschinformationen breitbandig zu synthetisieren.

[0057] Der Übergang in die DTX-Phase erfolgt in der Darstellung der Fig. 3 zum dargestellten Zeitpunkt TIME von 0 ms. Um diesen Übergang von einem schmalbandigen Sprachsignal auf ein breitbandiges Rauschsignal quais-stetig zu gestalten, was sich als günstigste Ausgestaltung für ein subjektive Hörempfinden eines menschlichen Rezipienten erwiesen hat, wird zu diesem Zeitpunkt TIME mit einem ausschließlich schmalbandigen Signal begonnen, d.h. mit einem Anteil HB-SHARE des breitbandigen Rauschens von 0. Zu einem Zeitpunkt von 100 ms ist der breitbandige Rauschanteil 1 bzw. 140 %. Zum quasi-stetigen Übergang von einem ausschließlich schmalbandigen Rauschsignal zum Zeitpunkt TIME = 0 ms auf einem ausschließlich breitbandigen Rauschsignal zum Zeitpunkt TIME = 100 ms haben sich in der Praxis folgende weitere Werte des Anteils HB-SHARE zu diskreten Zeitpunkten TIME bewahrt:

Zum Zeitpunkt TIME = 20 ms ein Anteil HB-SHARE von 0,09525986892242;

Zum Zeitpunkt TIME = 40 ms ein Anteil HB-SHARE von 0,19753086419753;

Zum Zeitpunkt TIME = 60 ms ein Anteil HB-SHARE von 0,36595031245237; und;

Zum Zeitpunkt TIME = 80 ms ein Anteil HB-SHARE von 0,62429507696997.

[0058] Eine weitere Ausführungsform der Erfindung sieht in analoger Weise einen Übergang von einem breitbandigen Sprachsignal auf ein schmalbandiges Rauschsignal vor.

[0059] Hierzu wird zunächst in Bezugnahme auf Fig 2A ein leicht abgeändertes Szenario angenommen, bei dem in Abweichung zu dem in Fig. 2A dargestellten Szenario kurz vor dem Zeitpunkt t3 noch eine nicht dargestellte - Änderung auf eine breitbandige Übertragung mit 32 kbit/s erfolgt. Trotz dieses »Peaks« bleibt der prozentuale Anteil von breitbandigen aktiven Sprachrahmen sehr niedrig, sodass nunmehr bei Übergang in die DTX-Phase ein Rauschsignal zu synthetisieren ist, dass breitbandig beginnt, jedoch - aufgrund der überwiegend schmalbandigen Übertragungshistorie und der damit für die Zukunft zu erwartenden Fortsetzung des schmalbandigen Übertragungscharakters - in ein schmalbandiges Rauschsignal überzuführen ist. Um diesen Übergang von einem breitbandigen Sprachsignal auf ein schmalbandiges Rauschsignal quais-stetig zu gestalten, wird zum Eintritt in die DTX-Phase mit einem ausschließlich

breitbandigen Signal begonnen, d.h. mit einem Anteil HB-SHARE des breitbandigen Rauschens von 1. Zu einem Zeitpunkt von 100 ms ist der schmalbandige Rauschanteil 0. Zum quasi-stetigen Übergang von einem ausschließlich breitbandigen Rauschsignal zum Zeitpunkt des Eintritts in die DTX-Phase auf einem ausschließlich schmalbandigen Rauschsignal zu einem Zeitpunkt nach 100 ms werden in vorteilhafter Weise die oben vorgeschlagenen Werte in einer umgekehrten Reihe eingestellt. Die entspräche einer an der Ordinate HB-SHARE gespiegelten Kurve gemäß Fig. 3.

Patentansprüche

1. Verfahren zur Dekodierung eines SID-Rahmens (SID) für eine Übermittlung von Hintergrundrauschinformationen in Anwendung eines skalierbaren Sprachsignalkodierungsverfahren mit folgenden Schritten:

Bestimmung eines Anteils an empfangenen breitbandigen Sprachrahmen im Verhältnis zu empfangenen schmalbandigen Sprachrahmen während einer Sprachphase, Dekodierung der in einem SID-Rahmen enthaltenen Hintergrundrauschinformationen bei Eintritt in eine DTX-Phase, wobei die Dekodierung nach Maßgabe des ermittelten Anteils vorwiegend breitbandig oder schmalbandig erfolgt.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** im Falle einer Bestimmung eines hohen Anteils an empfangenen breitbandigen Sprachrahmen bei Eintritt in die DTX-Phase eine vorwiegend breitbandige Dekodierung der Hintergrundrauschinformationen erfolgt.

3. Verfahren nach Anspruch 2, **dadurch gekennzeichnet, dass** bei Eintritt in die DTX-Phase zunächst eine vorwiegend schmalbandige Dekodierung der Hintergrundrauschinformationen erfolgt, welche nach einem einstellbaren Zeitraum in eine vorwiegend breitbandige Dekodierung übergeht.

4. Verfahren nach Anspruch 3, **dadurch gekennzeichnet, dass** der Übergang in die vorwiegend breitbandige Dekodierung mit einem Anteilsfaktor (HB-SHARE) einstellbar ist, welcher ein Verhältnis zwischen einer breitbandigen und einer schmalbandigen Rauschsignalqualität ausdrückt.

5. Verfahren nach Anspruch 4, **dadurch gekennzeichnet, dass** der Anteilsfaktor zum Zeitpunkt des Eintritts in die DTX-Phase zu Null dimensioniert wird.

6. Verfahren nach einem der Ansprüche 4 bis 5, **dadurch gekennzeichnet, dass** der Anteilsfaktor zu

einem Zeitpunkt von 100 ms nach Eintritt in die DTX-Phase zu Eins dimensioniert wird.

7. Verfahren nach einem der Ansprüche 4 bis 6, **dadurch gekennzeichnet, dass** der Anteilsfaktor

- zu einem Zeitpunkt von 20 ms nach Eintritt in die DTX-Phase zu 0,09525986892242 dimensioniert wird;
 - zu einem Zeitpunkt von 40 ms nach Eintritt in die DTX-Phase zu 0,19753086419753 dimensioniert wird;
 - zu einem Zeitpunkt von 60 ms nach Eintritt in die DTX-Phase zu 0,36595031245237 dimensioniert wird; und;
 - zu einem Zeitpunkt von 80 ms nach Eintritt in die DTX-Phase zu 0,62429507696997 dimensioniert wird.

8. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** im Falle einer Bestimmung eines niedrigen Anteils an empfangenen breitbandigen Sprachrahmen bei Eintritt in die DTX-Phase eine vorwiegend schmalbandige Dekodierung der Hintergrundrauschinformationen erfolgt.

9. Verfahren nach Anspruch 8, **dadurch gekennzeichnet, dass** bei Eintritt in die DTX-Phase zunächst eine vorwiegend breitbandige Dekodierung der Hintergrundrauschinformationen erfolgt, welche nach einem einstellbaren Zeitraum in eine vorwiegend schmalbandige Dekodierung übergeht.

10. Verfahren nach Anspruch 9, **dadurch gekennzeichnet, dass** der Übergang in die vorwiegend schmalbandige Dekodierung mit einem Anteilsfaktor (HB-SHARE) einstellbar ist, welcher ein Verhältnis zwischen einer breitbandigen und einer schmalbandigen Rauschsignalqualität ausdrückt.

11. Verfahren nach Anspruch 10 **dadurch gekennzeichnet, dass** der Anteilsfaktor zum Zeitpunkt des Eintritts in die DTX-Phase zu Eins dimensioniert wird.

12. Verfahren nach einem der Ansprüche 10 bis 11, **dadurch gekennzeichnet, dass** der Anteilsfaktor zu einem Zeitpunkt von 100 ms nach Eintritt in die DTX-Phase zu Null dimensioniert wird.

13. Verfahren nach einem der Ansprüche 10 bis 12, **dadurch gekennzeichnet, dass** der Anteilsfaktor

- zu einem Zeitpunkt von 20 ms nach Eintritt in die DTX-Phase zu 0,62429507696997 dimensioniert wird;
 - zu einem Zeitpunkt von 40 ms nach Eintritt in die DTX-Phase zu 0,36595031245237 dimensioniert wird;

sioniert wird;

- zu einem Zeitpunkt von 60 ms nach Eintritt in die DTX-Phase zu 0,19753086419753 dimensioniert wird; und;

- zu einem Zeitpunkt von 80 ms nach Eintritt in die DTX-Phase zu 0,09525986892242 dimensioniert 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.

Claims

1. A method for decoding an SID frame (SID) for transmitting background noise information using a scalable speech signal coding method with the following steps:

Determining an amount of received broadband speech frames relative to received narrowband speech frames during a speech phase,
 Decoding the background noise information contained in an SID frame upon entrance into a DTX phase, wherein the decoding takes place according to the determined amount of primarily broadband or narrowband.

2. The method according to Claim 1, **characterized in that** in the case of a determination of a high amount of received broadband speech frames, upon entrance into the DTX phase a primarily broadband decoding of the background noise information takes place.

3. The method according to Claim 2, **characterized in that** upon entrance into the DTX phase at first a primarily narrowband decoding of the background noise information takes place which merges after an adjustable time period into a primarily broadband decoding.

4. The method according to Claim 3, **characterized in that** the transition into the primarily broadband decoding can be adjusted with an amount factor (HB-SHARE) which expresses a relationship between a broadband and a narrowband noise signal quality.

5. The method according to Claim 4, **characterized in that** the amount factor is dimensioned to zero at the time of the entrance into the DTX phase.

6. The method according to one of Claims 4 to 5, **characterized in that** the amount factor is dimensioned

to one at a time of 100 ms after entrance into the DTX phase.

7. The method according to one of Claims 4 to 6, **characterized in that** the amount factor

- is dimensioned to 0.09525986892242 at a time of 20 ms after entrance into the DTX phase;
 - is dimensioned to 0.19753086419753 at a time of 40 ms after entrance into the DTX phase;
 - is dimensioned to 0.36595031245237 at a time of 60 ms after entrance into the DTX phase; and
 - is dimensioned to 0.62429507696997 at a time of 80 ms after entrance into the DTX phase.

8. The method according to Claim 1, **characterized in that** in the case of a determination of a low amount of received broadband speech frames upon entrance into the DTX phase a primarily narrowband decoding of the background noise information takes place.

9. The method according to Claim 8, **characterized in that** upon entrance into the DTX phase at first a primarily broadband decoding of the background noise information takes place which merges after an adjustable time period into a primarily narrowband decoding.

10. The method according to Claim 9, **characterized in that** the transition into the primarily narrowband decoding can be adjusted with the amount factor (HB-SHARE), that expresses a relationship between a broadband and a narrowband noise signal quality.

11. The method according to Claim 10, **characterized in that** the amount factor is dimensioned to one at the time of the entrance into the DTX phase.

12. The method according to one of Claims 10 to 11, **characterized in that** the amount factor is dimensioned to zero at a time period of 100 ms after entrance into the DTX phase.

13. The method according to one of Claims 10 to 12, **characterized in that** the amount factor

- is dimensioned to 0.62429507696997 at a time of 20 ms after entrance into the DTX phase;
 - is dimensioned to 0.36595031245237 at a time of 40 ms after entrance into the DTX phase;
 - is dimensioned to 0.19753086419753 at a time of 60 milliseconds after entrance into the DTX phase; and
 - is dimensioned to 0.9525986892242 at a time of 80 ms after entrance into the DTX phase.

14. A codec with means for carrying out the method ac-

cording to one 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.

Revendications

1. Procédé de décodage d'une trame SID (SID) pour une transmission d'informations de bruit de fond en application d'un procédé de codage de signal vocal modulable avec les étapes suivantes :

la détermination d'une proportion de trames vocales à large bande reçues par rapport à des trames vocales à bande étroite reçues pendant une phase vocale,
 le décodage des informations de bruit de fond contenues dans une trame SID lors de l'entrée dans une phase DTX, dans lequel le décodage est effectué conformément à la proportion déterminée principalement à large bande ou bande étroite.

2. Procédé selon la revendication 1, **caractérisé en ce que**, dans le cas d'une détermination d'une proportion élevée de trames vocales à large bande reçues lors de l'entrée dans la phase DTX, un décodage principalement à large bande des informations de bruit de fond est effectué.

3. Procédé selon la revendication 2, **caractérisé en ce que** lors de l'entrée dans la phase DTX, un décodage principalement à bande étroite des informations de bruit de fond est, tout d'abord, effectué, lequel décodage passe, après une période réglable à un décodage principalement à large bande.

4. Procédé selon la revendication 3, **caractérisé en ce que** le passage au décodage principalement à large bande est réglable avec un facteur de proportion (HB-SHARE) qui exprime un rapport entre une qualité de signal de bruit à large bande et une qualité de signal de bruit à bande étroite.

5. Procédé selon la revendication 4, **caractérisé en ce que** le facteur de proportion est dimensionné à zéro au moment de l'entrée dans la phase DTX.

6. Procédé selon l'une des revendications 4 à 5, **caractérisé en ce que** le facteur de proportion est dimensionné à un, à un moment de 100 ms après l'entrée dans la phase DTX.

7. Procédé selon l'une des revendications 4 à 6, **caractérisé en ce que** le facteur de proportion

- est dimensionné à un moment de 20 ms après l'entrée dans la phase DTX à 0,09525986892242 ;
 - est dimensionné à un moment de 40 ms après l'entrée dans la phase DTX à 0,19753086419753 ;
 - est dimensionné à un moment de 60 ms après l'entrée dans la phase DTX à 0,36595031245237 ;
 - est dimensionné à un moment de 80 ms après l'entrée dans la phase DTX à 0,62429507696997.
8. Procédé selon la revendication 1, **caractérisé en ce que**, dans le cas d'une détermination d'une proportion faible de trames vocales à large bande reçues lors de l'entrée dans la phase DTX, un décodage principalement à bande étroite des informations de bruit de fond est effectué.
9. Procédé selon la revendication 8, **caractérisé en ce que**, lors de l'entrée dans la phase DTX, un décodage principalement à large bande des informations de bruit de fond est tout d'abord effectué, lequel décodage passe, après une période réglable à un décodage principalement à bande étroite.
10. Procédé selon la revendication 9, **caractérisé en ce que** le passage au décodage principalement à bande étroite est réglable avec un facteur de proportion (HB-SHARE) qui exprime un rapport entre une qualité de signal de bruit à large bande et une qualité de signal de bruit à bande étroite.
11. Procédé selon la revendication 10, **caractérisé en ce que** le facteur de proportion est dimensionné à un au moment de l'entrée dans la phase DTX.
12. Procédé selon l'une des revendications 10 à 11, **caractérisé en ce que** le facteur de proportion est dimensionné à un moment de 100 ms après l'entrée dans la phase DTX à zéro.
13. Procédé selon l'une des revendications 10 à 12, **caractérisé en ce que** le facteur de proportion
- est dimensionné à un moment de 20 ms après l'entrée dans la phase DTX à 0,62429507696997 ;
 - est dimensionné à un moment de 40 ms après l'entrée dans la phase DTX à 0,36595031245237 ;
 - est dimensionné à un moment de 60 ms après l'entrée dans la phase DTX à 0,19753086419753 ; et ;
 - est dimensionné à un moment de 80 ms après l'entrée dans la phase DTX à 0,09525986892242.
14. Codec avec des moyens pour la réalisation du procédé selon l'une des revendications 1 à 13.
15. Codec selon la revendication 14, **caractérisé par** une implémentation dans la norme UIT-T G.729.1 connue en elle-même.

FIG 1

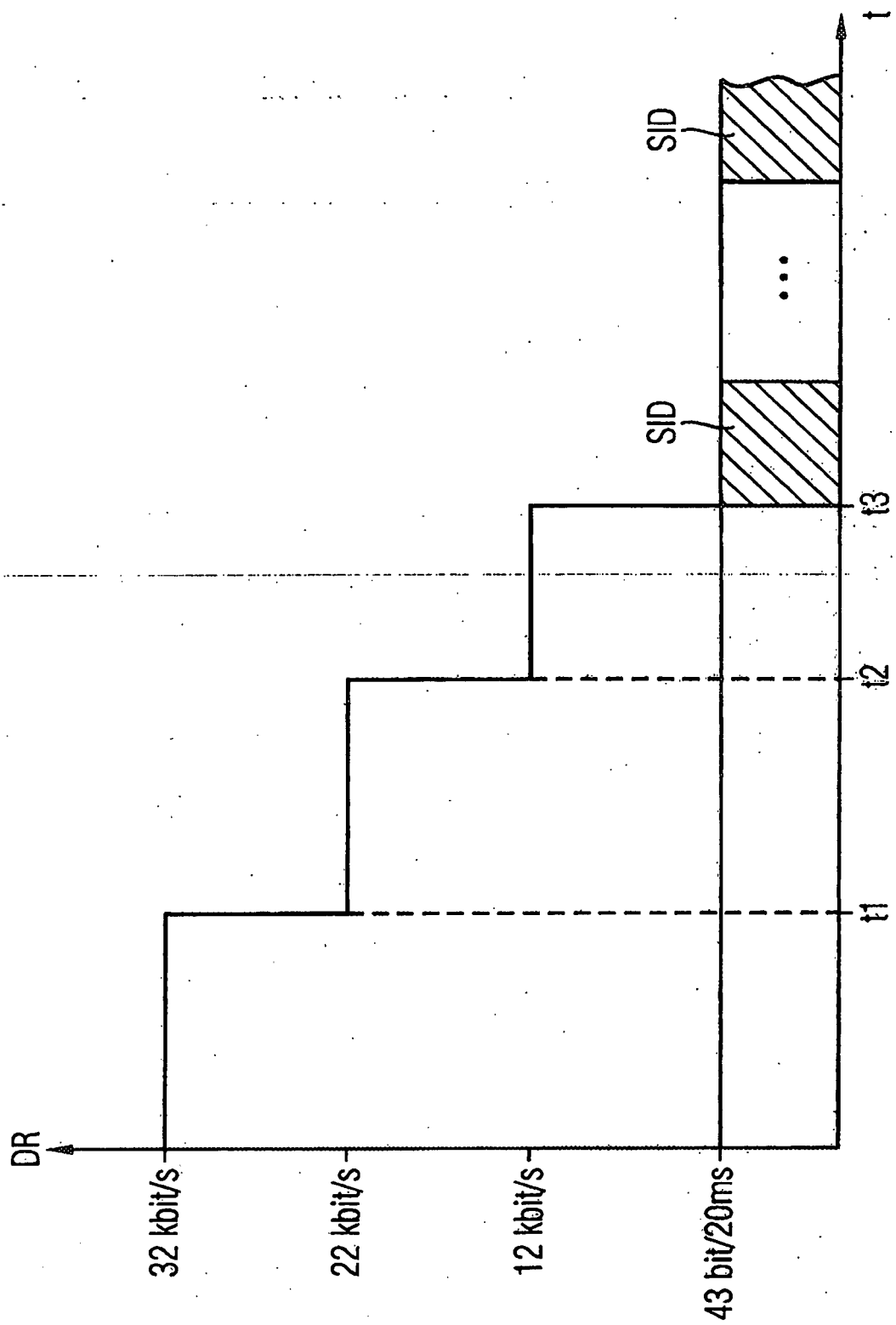


FIG 2A

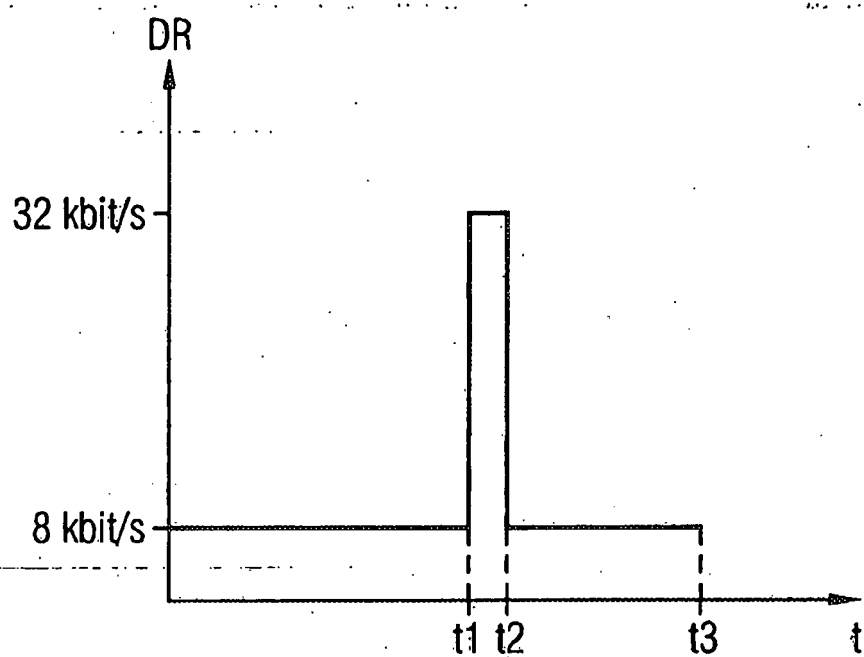


FIG 2B

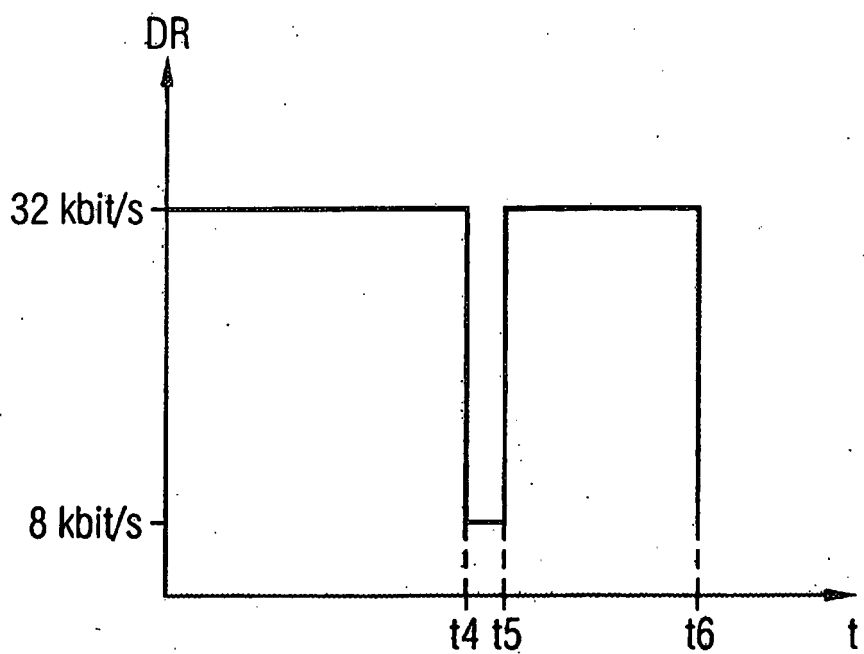


FIG 3

