

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
27. August 2009 (27.08.2009)

(10) Internationale Veröffentlichungsnummer
WO 2009/103608 A1

(51) Internationale Patentklassifikation:
G10L 19/00 (2006.01)

(74) Anwalt: MAIER, Daniel; Postfach 22 16 34, 80506 München (DE).

(21) Internationales Aktenzeichen: PCT/EP2009/051118

(81) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) Internationales Anmeldedatum:
2. Februar 2009 (02.02.2009)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
10 2008 009 719.5
19. Februar 2008 (19.02.2008) DE

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): SIEMENS ENTERPRISE COMMUNICATIONS GMBH & CO. KG [DE/DE]; Hofmannstr. 51, 81379 München (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): TADDEI, Herve [FR/DE]; Poppelsdorfer Allee 90, 53115 Bonn (DE).
SCHANDL, Stefan [AT/AT]; Klimtgasse 6/5/4, A-1130 Wien (AT).
SETIAWAN, Panji [ID/DE]; Gustav-Heinemann-Ring 38, 81739 München (DE).

(84) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD AND MEANS FOR ENCODING BACKGROUND NOISE INFORMATION

(54) Bezeichnung: VERFAHREN UND MITTEL ZUR ENKODIERUNG VON HINTERGRUNDRÄUSCHINFORMATIONEN

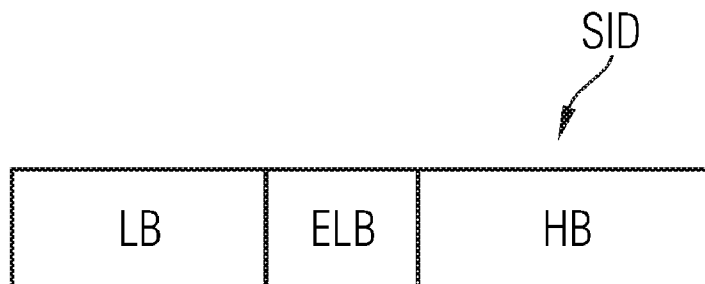


Fig. 1

(57) Abstract: The invention relates to a method and means for encoding background noise information during voice signal encoding methods. A basic idea of the invention is to provide the scalability known for transmitting voice information in a similar manner when forming an SID frame. The invention provides encoding of a narrowband first component and of a broadband second component of a piece of background noise information and formation of an SID frame which describes the background noise with separate areas for the first and second components.

(57) Zusammenfassung: Die Erfindung betrifft Verfahren und Mittel zur Enkodierung von Hintergrundrauschinformationen bei Sprachsignalkodierungsverfahren. Ein Grundgedanke der Erfindung besteht darin, die für die Übertragung von Sprachinformationen bekannte Skalierbarkeit analog bei der Bildung eines SID-Rahmens vorzusehen. Erfindungsgemäß ist eine Enkodierung eines schmalbandigen ersten Anteils und eines breitbandigen zweiten Anteils einer Hintergrundrauschinformation und eine Bildung eines das Hintergrundrauschen beschreibenden SID-Rahmens mit getrennten Bereichen für den ersten und den zweiten Anteil vorgesehen.



WO 2009/103608 A1



Veröffentlicht:

— mit internationalem Recherchenbericht (Artikel 21 Absatz 3)

— vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eingehen (Regel 48 Absatz 2 Buchstabe h)

Beschreibung

Verfahren und Mittel zur Enkodierung von Hintergrundrauschinformationen

5

Die Erfindung betrifft Verfahren und Mittel zur Enkodierung von Hintergrundrauschinformationen bei Sprachsignalkodierungsverfahren.

10 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.

15 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
20 Codec zum Einsatz, welcher aufgrund der beschriebenen Bandbreitenbeschränkung im Frequenzbereich zwischen 300 Hz und 3400 Hz im Folgenden auch als schmalbandiger Sprach-Codec (Narrow Band Speech Codec) bezeichnet wird. Unter dem Begriff Codec wird dabei sowohl die Kodiervorschrift zur digitalen
25 Kodierung von Audiosignalen als auch die Dekodiervorschrift zur Dekodierung von Daten mit dem Ziel einer Rekonstruktion des Audiosignals verstanden.

Ein schmalbandiger Sprach-Codec ist beispielsweise aus der
30 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.

Weiterhin sind sogenannte breitbandige Sprach-Codecs (Wide Band Speech Codec) bekannt, welche zur Verbesserung des Höreindrucks eine Kodierung eines in einem erweiterten Frequenzbereich vorsehen. Ein derart erweiterter Frequenzbereich
5 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.

Üblicherweise sind Kodierungsverfahren für breitbandige
10 Sprach-Codecs 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
15 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
20 übertragenen Datenrahmen vorzunehmen.

Für eine Reduzierung der Datenübertragungsrate durch einen Codec ist üblicherweise eine Komprimierung der zu übertragenen Daten vorgesehen. Eine Komprimierung wird beispielsweise
25 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
30 Sprachsignal hinsichtlich eines subjektiven Höreindrucks möglichst ähnlich ist. Mit Hilfe diesem auch als »Analysis-by-Synthesis« bezeichneten Verfahren werden nicht die ermittelten und digitalisierten Abtastwerte (Samples) selbst übertra-

gen, sondern ermittelte Parameter, die eine empfängerseitige Synthese des Sprachsignals ermöglichen.

Eine weitere Maßnahme zur Reduzierung der Datenübertragungs-
rate 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.

Hierzu kommt auf Seiten des Senders eine Sprechpausenerken-
nung (Voice Activity Detection, VAD) zum Einsatz, welche bei
Unterschreiten eines bestimmten Signalpegels auf eine Sprech-
pause erkennt. Üblicherweise wird vom Empfänger während einer
Sprechpause keine völlige Stille erwartet. Im Gegenteil würde
eine völlige Stille empfängerseitig zu Irritationen oder so-
gar zur Vermutung eines Verbindungsabbaus führen. Aus diesem
Grund werden Verfahren zur Erzeugung eines sogenannten Kom-
fortrauschen (Comfort Noise) angewandt.

Bei einem Komfortrauschen handelt es sich um Rauschen, wel-
ches zur Füllung von Stillephasen auf Seiten des Empfängers
synthetisiert wird. Das Komfortrauschen dient einem subjektivi-
ven Eindruck einer weiter bestehenden Verbindung, ohne die
für die Übertragung von Sprachsignalen vorgesehene Datenüber-
tragungsrate zu beanspruchen. Mit anderen Worten wird zur
senderseitigen Kodierung des Rauschens ein geringerer Aufwand
als zur Kodierung der Sprachdaten betrieben. Für eine empfän-
gerseitig 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.

Derzeit in der Entwicklung stehende Codecs konzentrieren sich auf eine skalierbare Enkodierung der Sprachinformation. Mit Hilfe einer skalierbaren Ansatzes wird erreicht, dass das Ergebnis des Enkodiervorgangs verschiedene Blöcke enthält, welche den schmalbandigen Anteil des ursprünglichen Sprachsignals enthalten, den breitbandigen Anteil oder auch die volle Bandbreite des Sprachsignals enthalten, also z.B. einen Frequenzbereich zwischen 50 und 7000 Hz.

In gegenwärtigen skalierbaren Kodierungsverfahren erfolgt die Enkodierung der Hintergrundrauschinformation entweder über die gesamte Bandbreite des Eingangsrauschsignals oder über einen Ausschnitt aus der Bandbreite des Eingangsrauschsignals. Das enkodierte Rauschsignal wird in Form von SID-Rahmen über das DTX-Verfahren übertragen und empfängerseitig rekonstruiert. Das rekonstruierte, d.h. synthetisierte Komfortrauschen weist also eventuell eine andere Qualität als die empfängerseitig synthetisierte Sprachinformation auf. Dies wirkt sich nachteilig auf die Rezeption des Empfängers aus.

Aufgabe der Erfindung ist es, eine verbesserte Implementierung des DTX-Verfahrens in skalierbaren Sprachcodecs anzugeben.

Die Aufgabe wird durch den Gegenstand der unabhängigen Ansprüche gelöst.

Ein Grundgedanke der Erfindung besteht darin, die für die Übertragung von Sprachinformationen bekannte Skalierbarkeit analog bei der Bildung eines SID-Rahmens vorzusehen.

Das erfindungsgemäße Verfahren zur Enkodierung eines SID-Rahmens für eine Übermittlung von Hintergrundrauschinformationen in Anwendung eines skalierbaren Sprachsignalkodierungs-

verfahren sieht eine Enkodierung eines schmalbandigen ersten und eines breitbandigen zweiten Anteils der Hintergrund-

rauschinformation vor. Die Enkodierung wird üblicherweise zeitgleich und auf verschiedene Art und Weise erfolgen. Die Enkodierung eines Anteils kann jedoch selbstverständlich auch zeitlich versetzt vor oder nach einer Enkodierung eines anderen Anteils erfolgen. Ebenso kann die Enkodierung der beiden Anteile optional auch in gleicher Weise erfolgen.

Nach der Enkodierung der beiden Anteile wird ein SID-Rahmen

gebildet mit getrennten Bereichen für den ersten und den zweiten Anteil. Dies bedeutet mit anderen Worten, dass im SID-Rahmen ein erster Datenbereich die Daten für den enkodierten ersten Anteil aufnimmt, während ein davon getrennter zweiter Datenbereich die Daten für den enkodierten zweiten Anteil aufnimmt.

Ein wesentlicher Vorteil der Erfindung besteht darin, dass empfängerseitig bestimmt werden kann, ob ein Komforttausch 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 verringert wurde, dass nur noch schmalbandige Sprachinformationen übertragen

werden. Wird nämlich, wie im derzeitigen Stand der Technik, schmalbandige Sprachinformationen in Verbindung mit breitbandigen 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.

Vorteilhafte Weiterbildungen der Erfindung sind in den Unteransprüchen angegeben.

5 Gemäß einer ersten vorteilhaften Ausgestaltung der Erfindung ist vorgesehen ein dritter Anteil in der Definition des SID-Rahmens vorgesehen. Dieser enthält enkodierte Hintergrundrauschparameter, welche mit einer erhöhten Datenrate enkodiert sind, wenngleich der dritte Anteil immer noch schmalbandige Daten (erweiterte schmalbandige Daten bzw. »Enhanced
10 Low Band«) enthält. Der Vorteil einer Definition des SID-Rahmens mit diesem dritten Anteil besteht in einer Möglichkeit, ein Rauschsignal in einer im Vergleich zur herkömmlichen schmalbandigen Kodierungsweise gesteigerten Qualität wiederzugeben und dabei noch in Konformität zum Standard
15 G.729.B zu bleiben.

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

20

Dabei zeigt die einzige FIG eine Struktur eines erfindungsgemäßen SID-Rahmens.

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

25

In gegenwärtigen skalierbaren Kodierungsverfahren für breitbandige Sprach-Codecs implementierte Verfahren zur diskontinuierlichen Übertragung (DTX) unterstützen für die Übertragung der Hintergrundrauschinformation derzeit nicht den skalierbaren Charakter, welcher für die Übertragung der Sprachinformation vorgesehen ist.

30

Als derzeitige Umgehungslösung erfolgt eine Enkodierung entweder über die gesamte Bandbreite des Eingangssprauschsignals oder über einen Ausschnitt aus der Bandbreite des Eingangssprauschsignals. Aus diesem Grund besteht ein Bedarf für verbesserte Verfahren.

In der Vergangenheit wurden hauptsächlich zwei Typen von Sprachcodecs entwickelt, einerseits schmalbandige Sprachcodecs wie z.B. 3GPP AMR, ITU-T G.729 und andererseits breitbandige Sprachcodecs, wie z.B. 3GPP AMR-WB, ITU-T G.722. Ein schmalbandiger Sprachcodec enkodiert Sprachsignale mit einer Abtastfrequenz von 8 kHz mit einer Bandbreite welche üblicherweise im Frequenzbereich zwischen 300 und 3400 Hz liegt. Ein breitbandiger Sprachcodec enkodiert ein Sprachsignal mit einer Abtastfrequenz von 16 kHz bei einer Bandbreite in einem Frequenzbereich zwischen 50 und 7000 Hz.

Einige dieser Codecs verwenden DTX-Verfahren, also diskontinuierliche Übertragungsverfahren, um die Gesamtübertragungsrate im Kommunikationskanal zu reduzieren. Gemäß dem DTX-Verfahren werden SID-Rahmen gesendet, wobei die Bandbreite der SID-Rahmen mit der Bandbreite des Sprachsignals korrespondiert. In einem SID-Rahmen wird das Hintergrundrauschen während einer Sprechpause beschrieben.

Derzeit in der Entwicklung stehende Codecs konzentrieren sich auf eine skalierbare Kodierung. Mit Hilfe einer skalierbaren Ansatzes wird erreicht, dass das Ergebnis des Enkodiervorgangs verschiedene Blöcke enthält, welche den schmalbandigen Anteil des ursprünglichen Sprachsignals enthalten, den breitbandigen Anteil oder auch die volle Bandbreite des Sprachsignals enthalten, also z.B. einen Frequenzbereich zwischen 50 und 7000 Hz. Der breitbandige Anteil beginnt üblicherweise ab einer Frequenz von 4 kHz.

Die gegenwärtigen DTX-Verfahren unterstützen derzeit nicht den skalierbaren Charakter von Codecs. Stattdessen erfolgt eine Kodierung entweder über die gesamte Bandbreite des Eingangssprachsignals oder über einen Ausschnitt aus der Bandbreite des Eingangssignals. Aus diesem Grund besteht ein Bedarf für verbesserte Verfahren.

Zur Verdeutlichung wird im Folgenden das Enkodierverfahren gemäß ITU-T-Standards G.729.1 beschrieben. Bei diesem Codec G.729.1 handelt es sich um einen skalierbaren Sprachcodec, in welchem das DTX-Verfahren derzeit nicht skalierbar über die gesamte Bandbreite angewandt wird.

Das Codierverfahren lässt sich während einer aktiven Sprachperiode – in Abgrenzung zu einer als »Silent Period« erkannten Sprechpause – wie folgt charakterisieren:

Das Sprachsignal wird in zwei Anteile, nämlich einen schmalbandigen (Lowband) Teil und einen breitbandigen (Highband) Anteil zerlegt. Beide Signale sind mit einer Abtastfrequenz von 8 kHz abgetastet. Die Aufteilung in einen schmalbandigen und einen breitbandigen Anteil erfolgt in einem speziellen Bandpassfilter, welcher auch als QMF (Quadrature Mirror Filter) bezeichnet wird.

Der schmalbandige Anteil des Sprachsignals wird mit einer Datenrate von 8 und 12 kbit/s enkodiert. Zur Enkodierung des Sprachsignals wird ein CELP-Verfahren (Code Excited Linear Prediction) angewandt. Für Datenraten oberhalb von 14 kbit/s wird der schmalbandige Anteil weiter unter Berücksichtigung des »Transform Codec«-Abschnitts von G.729.1 modifiziert. Der breitbandige Anteil des aktuellen Rahmens – wiederum unter der Voraussetzung, dass dieser Sprachsignale enthält – wird mit einer Datenrate von 14 kbit/s unter Anwendung des TDBWE-Verfahrens (Time Domain Bandwidth Extension) enkodiert. Für

Datenrate von über 14 kbit/s wird der »Transform Codec«-Abschnitt von G.729.1 angewandt.

5 Da der Standard G.729.1 keine Verfahren zur diskontinuierlichen Übertragung bereitstellt, wird in Sprechpausen bzw. »non active voice periods« eine Umgehungslösung angewandt, welche im Folgenden beschrieben wird.

10 Das Sprachsignal wird ebenfalls in einen schmalbandigen und einen breitbandigen Anteil zerlegt, wobei beide Anteile mit einer Frequenz von 8 kHz abgetastet werden. Die Zerlegung erfolgt ebenfalls über ein QMF-Filter.

15 Der schmalbandige Anteil wird unter Verwendung einer schmalbandigen SID-Information enkodiert. Diese schmalbandige SID-Information zu einem späteren Zeitpunkt in einem SID-Rahmen, welcher kompatibel zum Standard G.729 ist, an den Empfänger gesandt. Weitere wie oben beschriebene Maßnahmen können zu einer Verbesserung des schmalbandigen SID-Anteils beitragen.

20 Der breitbandige Anteil wird unter Anwendung eines modifizierten TDBWE-Verfahrens enkodiert. Während einer sog. Überhangperiode (Hangover Period) wird das Sprachsignal weiterhin mit einer Datenrate von 14 kbit/s enkodiert, während gleichzeitig das während der Sprechpause erkannte Hintergrundrauschen ausgewertet und entsprechende Parameter eingestellt werden. Die Auswertung des Hintergrundrauschens erfolgt hinsichtlich der Energie des Rauschsignals und hinsichtlich seiner Frequenzverteilung. Im Gegensatz zu dem vom Standard
30 G.729.1 vorgesehenen TDBWE-Verfahren wird jedoch die zeitliche Feinstruktur nicht ausgewertet, sondern lediglich ein Durchschnitt der Energie über den Rahmen gebildet.

35 Im Folgenden wird eine Ausführungsform des erfindungsgemäßen Verfahrens anhand der FIG erläutert.

Die FIG zeigt einen SID-Rahmen mit getrennten Bereichen für einen schmalbandigen ersten Anteil LB (»Low Band«), einen breitbandigen zweiten Anteil HB (»High Band«) und einen itermediären dritten Anteil ELB (»Enhanced Low Band«).

5

Der erste Anteil LB enthält dabei enkodierte Hintergrundrauschparameter, welche mit einer Datenrate von 8 kbit/s oder darunter enkodiert sind. Die Datenlänge des ersten Anteils LB beträgt beispielsweise 15 Bit.

10

Der zweite Anteil HB enthält enkodierte Hintergrundrauschparameter, welche mit einer Datenrate zwischen 14 kbit/s und 32 kbit/s enkodiert sind. Die Datenlänge des zweiten Anteils HB beträgt beispielsweise 19 Bit.

15

Der dritte Anteil ELB enthält enkodierte Hintergrundrauschparameter, welche mit einer Datenrate von größer als 8kbit/s also beispielsweise 12 kbit/s enkodiert sind. Die Datenlänge des dritten Anteils ELB beträgt beispielsweise 9 Bit. Der

20

Vorteil einer Definition des SID-Rahmens mit einem dritten Anteil ELB besteht in einer Möglichkeit, ein Rauschsignal in einer im Vergleich zur herkömmlichen schmalbandigen Kodierungsweise gesteigerten Qualität wiederzugeben und dabei noch in Konformität zum Standard G.729.B zu bleiben.

25

Während einer Sprechpause werden auf Seiten des Enkoders Charakteristika des Hintergrundrauschens angelernt. Die Charakteristika umfassen insbesondere die zeitliche Verteilung als auch die spektrale Form des Hintergrundrauschens. Für den Anlernvorgang wird ein Filterverfahren angewandt, welches zeitliche und spektrale Parameter des Hintergrundrauschens aus vorangegangenen Rahmen berücksichtigt. Ergeben sich signifikante Änderungen im Charakter oder in der Stärke des Hintergrundrauschens, wird eine Entscheidung auf Basis von Grenzwertparametern (Threshold Values) getroffen, ob ein Bedarf besteht, die angelernten Parameter zu aktualisieren.

Auf Seiten des Dekoders bzw. Empfängers wird folgendes Verfahren durchgeführt: Wenn ein »regulärer«, d.h. ein sprachsignalenthaltender Rahmen empfangen wird, wird die übliche Dekodierung ausgeführt. Die Datenrate für solche regulären Rahmen beträgt üblicherweise 8 kbit/s oder darüber. Wenn ein SID-Rahmen empfangen wird, wird Komfortrauschen synthetisiert, wobei im Falle eines breitbandigen SID ein breitbandiges Komfortrauschen synthetisiert und mit einem ausgelesenen Verstärkungsfaktor ausgegeben wird.

Im Folgenden wird das erfindungsgemäße Verfahren mit weiteren Ausgestaltungen der Erfindung beschrieben.

Die Ausgestaltungen betreffen weitere Details zur Einbeziehung des DTX-Verfahrens in breitbandige Codecs wie z.B. G.729.1 und weiterhin Verfahren zur Modifizierung des TDBWE-Verfahrens, welche eine Synthetisierung von Komfortrauschen während nicht-aktiver Rahmen (Non Active Frames), d.h. Rahmen ohne Sprachinformation, unterstützen.

Gemäß einer Ausgestaltung ist folgendes Vorgehen vorgesehen.

- Produzieren einer schmalbandigen SID-Information zur Erzeugung eines G.729- bzw. G.729.B- kompatiblen SID-Rahmens (erster Anteil LB des erfindungsgemäßen SID-Rahmens)
- Produzieren einer breitbandigen SID-Information unter Verwendung eines modifizierten TDBWE-Verfahrens (zweiter Anteil HB des erfindungsgemäßen SID-Rahmens)
- Optional werden Verbesserungen bezüglich der schmalbandigen und/oder der breitbandigen SID-Informationen vorgenommen.
- Das Hintergrundrauschen wird während einer Phase, welche einem Senden der ersten SID-Rahmen vorausgeht, bezüglich der Energie- und/oder Frequenzverteilung analysiert bzw. »angelernt«.
- SID-Rahmen werden gesendet, wenn eine signifikante Änderung des breitbandigen Anteils des Hintergrundrauschens detektiert wird oder wenn eine Aktualisierung der schmalbandigen SID-Informationen gesendet werden soll.

Eine Implementierung dieses Ausführungsbeispiels erfolgt in folgenden Phasen:

- 5 - Mit Hilfe eines VAD-Verfahrens wird eine aktive Sprachphase bzw. eine Sprechpause definiert.
- Wird durch das VAD-Verfahren ein Wechsel in eine Sprechpause angezeigt, wird eine Überhangperiode (Hang Over Period) gestartet. Während der Überhangperiode wird die Datenrate des Enkodierers auf 14kbit/s reduziert, wenn die
10 vorhergehende Datenrate einen höheren Wert aufgewiesen hat. Für den Fall dass die vorhergehende Datenrate des Enkodierers bereits Werte um 12 kbit/s betragen hat, wird die Datenrate auf einen Wert von 8 kbit/s reduziert.
- Während der Überhangperiode wird das Hintergrundrauschen
15 bezüglich des schmalbandigen Anteils in analoger Weise zum Vorgehen in Standard G.729 angelernt, jedoch unter Verwendung einer höheren Anzahl von Rahmen. Hierbei kann optional ein Filterverfahren angewandt werden, durch welches erreicht wird, dass aktuellen Rahmen eine höhere
20 Wichtigkeit zugeordnet wird als vorausgegangenen Rahmen.
- Während der Überhangperiode wird das Hintergrundrauschen darüber hinaus im breitbandigen Anteil angelernt. Optional wird für eine Vereinfachung der Implementierung, insbesondere zur Reduzierung des Speicherplatzbedarfs, ein
25 modifiziertes TDBWE-Verfahren eingesetzt, welche durch eine vereinfachte Enkodierung im Zeitbereich gekennzeichnet ist. Optional kann eine weitere Vereinfachung im modifizierten TDBWE-Verfahren dadurch erreicht werden, dass die Enkodierung im Zeitbereich nur mit der Energie des
30 Signals im Zeitbereich korrespondiert. Eine weitere optionale vereinfachte Enkodierung besteht darin, spektrale Glättungsverfahren anzuwenden, da die Energie im Zeitbereich und im Frequenzbereich als Folge des Parsevaltheorems gleich Werte liefert. Auch im breitbandigen Anteil
35 des Hintergrundrauschens können optional weitere Filterungsmaßnahmen angewandt werden, welche das Ziel haben, aktuellen Rahmen eine höhere Wichtigkeit als vorausgegangenen Rahmen zuzuordnen.

- Nach Beendigung der Überhangperiode wird ein erster SID-Rahmen gesendet, welche eine grobe Repräsentierung des Hintergrundrauschens enthält. Die grobe Beschreibung des Hintergrundrauschens wurde während der Überhangperiode angelernt.
- Solange durch die VAD keine aktive Phase (sprechen) detektiert wurde, wird auf Seiten des Dekoders bzw. Empfängers ein Komfortrauschen auf Basis der empfangenen SID-Rahmen synthetisiert.
- Änderungen des Hintergrundrauschens werden im schmalbandigen Anteil des SID-Rahmens detektiert, wobei ein ähnliches Verfahren zu G.729 verfolgt wird, wenngleich verschiedene Parameter berücksichtigt werden.
- Im breitbandigen Anteil werden gefilterte Energieparameter zur Beschreibung des Hintergrundrauschens benutzt. Diese umfassen z.B. Parameter von Einhüllkurven im Zeitbereich $tenv_fidx$ und/oder Parameter von Einhüllkurven im Frequenzbereich $fenv_fidx[i]$, wobei ein jeweiliger Index idx einen jeweiligen Rahmen identifiziert und wobei die Einhüllkurve im Frequenzbereich von einer geeigneten Anzahl von Frequenzwerten $i=\{1,..., NB-SUBBANDS\}$ zur Beschreibung der spektralen Eigenschaften des Hintergrundrauschens gebildet wird. Die gefilterten Energieparameter werden von den in G.729.1 definierten TDBWE-Parameter abgeleitet unter Verwendung geeigneter Tiefpassfilter:

$$tenv_f_{idx} = \alpha_{tenv} \cdot tenv_{f_{idx}} + (1 - \alpha_{tenv}) \cdot tenv_f_{idx-1}$$

$$fenv_f_{idx}[i] = \alpha_{tenv} \cdot fenv_{f_{idx}}[i] + (1 - \alpha_{tenv}) \cdot fenv_f_{idx-1}[i]$$

- Welche auf die Einhüllparameter im Frequenz- und im Zeitbereich entsprechend angewandt werden.
- Änderungen im breitbandigen Anteil der Energieparameter werden überwacht und detektiert, indem die gefilterten Energieparameter des gegenwärtigen Rauschsignals verglichen werden mit zwei Sätzen aus Vergleichswerten dieser Parameter, wobei ein Satz von Vergleichswerten die Para-

meter aus dem vorangegangenen Rahmen mit dem Index $idx-1$ ist.

$$temp_d = 20 \cdot \frac{\log(2)}{\log(10)} \cdot |tenv_f_{idx} - tenv_f_{idx-1}|$$

5

$$spec_d = 20 \cdot \frac{\log(2)}{\log(10)} \cdot \frac{1}{NB_SUBBANDS} \cdot \sum_{i=1}^{NB_SUBBANDS} |fenv_f_{idx}[i] - fenv_f_{idx-1}[i]|$$

10

Und wobei der andere Satz aus Parametern des zuletzt übertragenen Rahmens mit dem Index $last_tx$ besteht. Wenn einer der Parameterunterschiede ($temp_d$, $spec_d$, $temp_ch$, $spec_ch$) einen geeignet gewählten Grenzwert überschreitet:

$$temp_ch = 20 \cdot \frac{\log(2)}{\log(10)} \cdot |tenv_f_{idx} - tenv_f_{last_tx}|$$

15

$$spec_ch = 20 \cdot \frac{\log(2)}{\log(10)} \cdot \frac{1}{NB_SUBBANDS} \cdot \sum_{i=1}^{NB_SUBBANDS} |fenv_f_{idx}[i] - fenv_f_{last_tx}[i]|$$

muss ein neuer SID-Update-Rahmen gesendet werden.

20

- Sobald durch die VAD eine Sprachperiode erkannt wird, wird das Sprachsignal mit der benötigten Übertragungsrate übertragen und die Synthetisierung von Komfortrauschen auf der Dekoderseite beendet. Somit stellt sich ein regulärer Dekodierungsbetrieb ein wie in G.729.1.

25

Patentansprüche

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

Enkodierung eines schmalbandigen ersten Anteils (LB) und eines breitbandigen zweiten Anteils (HB) der Hintergrundrauschinformation;

10 Bildung des SID-Rahmens (SID) mit getrennten Bereichen für den ersten (LB) und den zweiten (HB) Anteil.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass ein erweiterter schmalbandiger dritter Anteil (ELB) enkodiert
15 wird und dass die Bildung des SID-Rahmens mit einem zusätzlichen getrennten Bereich für den dritten Anteil (ELB) gebildet wird.

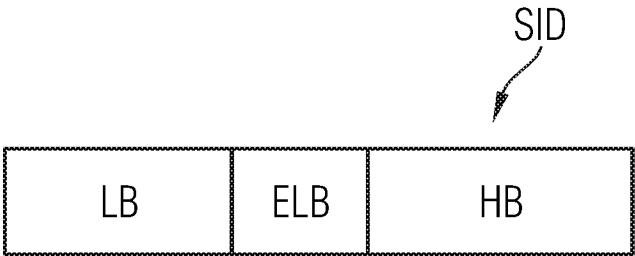
3. Verfahren nach einem der vorgenannten Ansprüche, dadurch gekennzeichnet, dass der erste Anteil (LB) der Hintergrundrauschinformation gemäß Kodierungsrichtlinien des an sich be-
20 kannten Standards G.729.B enkodiert werden.

4. Verfahren nach einem der vorgenannten Ansprüche, dadurch gekennzeichnet, dass der zweite Anteil (HB) der Hintergrundrauschinformation gemäß eines modifizierten TDBWE-Verfahrens enkodiert wird.
25

5. Verfahren nach einem der vorgenannten Ansprüche, dadurch gekennzeichnet, dass Während einer Überhangperiode Filterverfahren zur Zuordnung einer höheren Wichtigkeit eines aktuellen Rahmens als vorausgegangenen Rahmen angewandt werden.
30

6. Codec mit Mitteln zur Durchführung des Verfahrens nach einem der Ansprüche 1 bis 5.

7. Codec nach Anspruch 6, gekennzeichnet durch eine Implementierung im an sich bekannten ITU-T Standard G.729.1.



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/051118

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>SETIAWAN P ET AL: "On the ITU-T G.729.1 Silence Compression Scheme" EUSIPCO-2008, [Online] 25 August 2008 (2008-08-25), pages 1-5, XP002526622 LAUSANNE, SWITZERLAND Retrieved from the Internet: URL: http://www.eurasip.org/Proceedings/Eusipco/Eusipco2008/papers/1569104920.pdf [retrieved on 2009-05-29] paragraph [0005]</p> <p style="text-align: center;">----- -/--</p>	1-7

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- * & * document member of the same patent family

Date of the actual completion of the international search

29 Mai 2009

Date of mailing of the international search report

01/07/2009

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Van Doremalen, Hans

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2009/051118

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SOLLAUD A: "G.729.1 RTP Payload Format update: DTX support" INTERNET CITATION, [Online] 8 February 2008 (2008-02-08), - 8 February 2008 (2008-02-08) XP002526621 Retrieved from the Internet: URL: http://www.ietf.org.nyud.net:8080/proceedings/08mar/IDs/draft-ietf-avt-rfc4749-dtx-update-00.txt [retrieved on 2008-02-08]	1-3,6,7
Y	paragraph [0002]	4,5
Y	----- BERND GEISER ET AL: "Bandwidth Extension for Hierarchical Speech and Audio Coding in ITU-T Rec. G.729.1" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, vol. 15, no. 8, 1 November 2007 (2007-11-01), pages 2496-2509, XP011192970 IEEE SERVICE CENTER, NEW YORK, NY, US ISSN: 1558-7916 abstract paragraph [0III]	4
Y	----- BRUHN S ET AL: "Continuous and discontinuous power reduced transmission of speech inactivity for the GSM system" IEEE GLOBECOM 1998, vol. 4, 8 November 1998 (1998-11-08), - 12 November 1998 (1998-11-12) pages 2091-2096, XP002529976 SYDNEY, NSW, AUSTRALIA ISBN: 0-7803-4984-9 the whole document	5
A	----- SERIZAWA M ET AL: "A silence compression algorithm for multi-rate/dual-bandwidth MPEG-4 CELP standard" IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, ICASSP '00, vol. 2, 5 June 2000 (2000-06-05), - 9 June 2000 (2000-06-09) pages 1173-1176, XP010504937 PISCATAWAY, NJ, USA abstract; figure 3; table 1	1
A	----- EP 1 808 852 A1 (NOKIA CORP [FI]) 18 July 2007 (2007-07-18) abstract; claims 1-3 ----- -/--	1

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/051118

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>US 2008/027716 A1 (RAJENDRAN VIVEK [US] ET AL) 31 January 2008 (2008-01-31) abstract claims 10,11 paragraphs [0005] - [0007] -----</p>	1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2009/051118

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1808852	A1	18-07-2007	NONE
US 2008027716	A1	31-01-2008	CA 2657420 A1 07-02-2008
			EP 2047457 A2 15-04-2009
			KR 20090033461 A 03-04-2009
			WO 2008016942 A2 07-02-2008

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen
PCT/EP2009/051118

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
INV. G10L19/00

Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
G10L

Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data, INSPEC

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
P,X	<p>SETIAWAN P ET AL: "On the ITU-T G.729.1 Silence Compression Scheme" EUSIPCO-2008, [Online] 25. August 2008 (2008-08-25), Seiten 1-5, XP002526622 LAUSANNE, SWITZERLAND Gefunden im Internet: URL: http://www.eurasip.org/Proceedings/Eusipco/Eusipco2008/papers/1569104920.pdf [gefunden am 2009-05-29] Absatz [0005]</p> <p style="text-align: center;">----- -/--</p>	1-7

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen ☒ Siehe Anhang Patentfamilie

- | | |
|--|---|
| <p>* Besondere Kategorien von angegebenen Veröffentlichungen</p> <p>*A* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist</p> <p>*E* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist</p> <p>*L* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)</p> <p>*O* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht</p> <p>*P* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist</p> | <p>*T* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist</p> <p>*X* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden</p> <p>*Y* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist</p> <p>*G* Veröffentlichung, die Mitglied derselben Patentfamilie ist</p> |
|--|---|

Datum des Abschlusses der internationalen Recherche

29. Mai 2009

Absendedatum des internationalen Recherchenberichts

01/07/2009

Name und Postanschrift der Internationalen Recherchenbehörde
Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Van Doremalen, Hans

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	SOLLAUD A: "G.729.1 RTP Payload Format update: DTX support" INTERNET CITATION, [Online] 8. Februar 2008 (2008-02-08), - 8. Februar 2008 (2008-02-08) XP002526621 Gefunden im Internet: URL: http://www.ietf.org.nyud.net:8080/proceedings/08mar/IDs/draft-ietf-avt-rfc4749-dtx-update-00.txt [gefunden am 2008-02-08] Absatz [0002]	1-3,6,7
Y	----- BERND GEISER ET AL: "Bandwidth Extension for Hierarchical Speech and Audio Coding in ITU-T Rec. G.729.1" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, Bd. 15, Nr. 8, 1. November 2007 (2007-11-01), Seiten 2496-2509, XP011192970 IEEE SERVICE CENTER, NEW YORK, NY, US ISSN: 1558-7916 Zusammenfassung Absatz [0111]	4,5
Y	----- BRUHN S ET AL: "Continuous and discontinuous power reduced transmission of speech inactivity for the GSM system" IEEE GLOBECOM 1998, Bd. 4, 8. November 1998 (1998-11-08), - 12. November 1998 (1998-11-12) Seiten 2091-2096, XP002529976 SYDNEY, NSW, AUSTRALIA ISBN: 0-7803-4984-9 das ganze Dokument	5
A	----- SERIZAWA M ET AL: "A silence compression algorithm for multi-rate/dual-bandwidth MPEG-4 CELP standard" IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, ICASSP '00, Bd. 2, 5. Juni 2000 (2000-06-05), - 9. Juni 2000 (2000-06-09) Seiten 1173-1176, XP010504937 PISCATAWAY, NJ, USA Zusammenfassung; Abbildung 3; Tabelle 1	1
A	----- EP 1 808 852 A1 (NOKIA CORP [FI]) 18. Juli 2007 (2007-07-18) Zusammenfassung; Ansprüche 1-3	1
A	----- US 2008/027716 A1 (RAJENDRAN VIVEK [US] ET AL) 31. Januar 2008 (2008-01-31) Zusammenfassung Ansprüche 10,11 Absätze [0005] - [0007]	1

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP2009/051118

Im Recherchenbericht angeführtes Patentedokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 1808852	A1	18-07-2007	KEINE
US 2008027716	A1	31-01-2008	CA 2657420 A1 07-02-2008
		EP 2047457 A2 15-04-2009	
		KR 20090033461 A 03-04-2009	
		WO 2008016942 A2 07-02-2008	