

次世代移動通信向け音声／音響符号化技術 ITU-T G.718

ITU-T G.718 - Development of Speech/Audio Codec for Next-Generation Mobile Communication Systems

山 梨 智 史
Tomofumi Yamanashi

押 切 正 浩
Masahiro Oshikiri

森 井 利 幸
Toshiyuki Mori

佐 藤 薫
Kaoru Satoh

江 原 宏 幸
Hiroyuki Ehara

吉 田 幸 司
Koji Yoshida

要 旨

スケーラブル広帯域コーデックに対する低ビットレート高効率符号化方式の開発に取り組み、帯域選択型形状利得符号化方式 (BS-SGC: Band-Selective Shape-Gain Coding) をはじめとする要素技術を開発した。スケーラブル広帯域コーデックは、広い帯域 (~7 kHz) と伝送環境の変化に対する高い柔軟性を有しており、all-IP網での通信が想定される次世代通信システムに適したコーデックである。筆者らは、ITU-T (International Telecommunication Union-Telecommunication Standardization Sector) 新規国際標準G.718を次世代通信用コーデックと位置づけ、この標準化提案に参画した。BS-SGCをはじめとするパナソニックグループの技術は、主観評価試験において従来技術に対して明らかな性能向上が確認され、G.718の特長な処理の高性能化技術として採用された。G.718は、2008年6月にITU-Tで新規承認された。

Abstract

We worked toward development of a high-efficiency encoding algorithm for a scalable wide-band codec at low bit-rates, and many key technologies including BS-SGC (Band-Selective Shape-Gain Coding) were developed. The scalable codec supports wide-band (~7 kHz) signals and has flexibility to adapt to varying transport layer conditions; thus it is suitable for the next-generation communication systems in which all data are transported by the Internet Protocol (IP). We considered G.718, a new standard of ITU-T (International Telecommunication Union - Telecommunication Standardization Sector), the most promising codec for the next-generation communication systems and took part in its standardization activities. Several techniques developed by Panasonic were proven to outperform conventional technologies; and therefore, they contributed toward enhancing the performance of G.718 and were adopted as featured parts of G.718, which was approved in June 2008.

1. はじめに

近年、世界の携帯電話は第2世代 (2G) から第3世代 (3G) へのシフトが進んでおり、2010年には第3.9世代 (3.9G) のサービスが開始される見通しである。そして、現在、2015年頃のサービスをめざした第4世代 (4G) の標準化が進められている。

筆者らは、4G通信システムに最も適したコーデックは「パケットロス耐性の高いスケーラブル広帯域コーデック」であると考え、その有力候補となり得るITU-T G.718の標準化へ参画し、提案活動を推進した。

G.718¹⁾ は、VoIP (Voice over IP) に適した、パケットロス耐性が高く、スケーラビリティを有する広帯域音声／楽音信号符号化方式として、2008年6月にITU-Tで承認された国際標準である。標準化にあたり、既存の方式にはない特長である「低ビットレートでの高い楽音信号符号化性能」を実現することなどを要求条件に設定し、これらを達成するための技術開発に取り組んだ。今回、筆者らが開発したBS-SGC^{2), 3)} により、G.718は上記要求条件を達成した。このほか、低ビットレート音声符号化処理やフレーム消失隠蔽 (いんぺい) 処理など、G.718の特長

的な処理の高性能化に貢献する技術を開発した。

本論文では、ITU-T国際標準G.718のアルゴリズム概要、およびG.718に採用されたBS-SGCについて述べる。

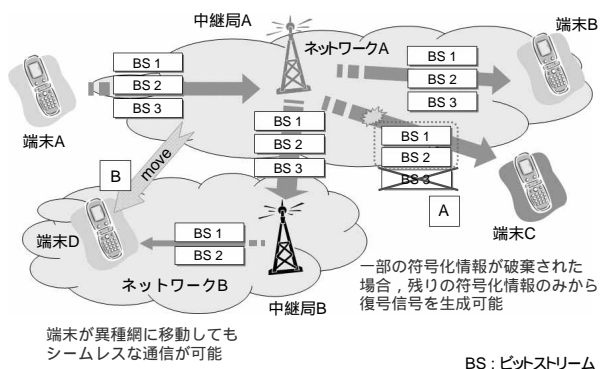
2. ITU-T G.718概要

2.1 G.718の特長

G.718は、符号化ビットレートが8/12/16/24/32 kbit/sであり、5階層 (レイヤ) から構成されるスケーラブル広帯域コーデックである。以下では、各レイヤを低位レイヤから、L1, L2, L3, L4, L5と呼ぶ。このG.718は、狭帯域 (~3.4 kHz) と広帯域 (~7 kHz) の信号に対応するとともに、5レイヤ構成により高いスケーラビリティを有し、さらに伝送中にフレーム消失が発生した場合にも高品質な隠蔽処理を行うことでパケットロス耐性を高めている。

ここで、スケーラブルコーデックとは、入力信号を階層的に符号化・復号するコーデックであり、「階層的に符号化された情報のうち一部の情報が失われても残りの情報から復号信号を得ることが可能である」という特長を有する。この特長により、第1図中Aに示すように、中継

局において任意の階層を破棄することができ輻輳制御が容易になる。また、第1図中Bに示すように、伝送路の速度が異なるネットワーク間を端末が移動した場合に対しても、中継局が一部の符号化情報を削除することにより、端末とのネゴシエーションなしに通信速度を決定することができる。



第1図 スケーラブルコーデックの特長
Fig. 1 Illustration of features for scalable codec

2.2 要求仕様

第1表および第2表に、G.718の標準化にあたり定めた要求仕様 (ToR: Terms of Reference) を示す。第1表、第2表はそれぞれ音声信号、楽音信号に対するToRであり、各ビットレートで達成すべき品質を要求品質 (Requirement) として定めている。要求品質として設定した既存標準コーデックはスケーラブル構成ではなく固定ビットレートで最適化されているため、たとえ同じビットレートであっても同等の品質を達成するには困難を伴う。また、最低ビットレート (L1) での音声信号に対しては、より高いビットレートの既存標準コーデックと同等の性能を達成しなければならない。さらに、第1表、第2表以外にも、高い伝送誤り率での高品質性、背景雑音条件下での高ロバスト性など多くの要求品質 (全80項目) が定められており、G.718はこれら全項目をスケーラブル構成で、かつ双方向通信に耐える遅延条件で達成する必要がある。

第1表 音声信号に対する要求性能

Table 1 Requirement for speech

レイヤ	要求品質
L1 (8 kbit/s)	G.729E@11.8 kbit/s G.722.2@8.85 kbit/s
L2 (12 kbit/s)	G.729@8 kbit/s 以上
L3 (16 kbit/s)	G.722.2@12.65 kbit/s
L4 (24 kbit/s)	G.722.2@15.85 kbit/s
L5 (32 kbit/s)	G.722.2@23.85 kbit/s

第2表 楽音信号に対する要求性能

Table 2 Requirement for music

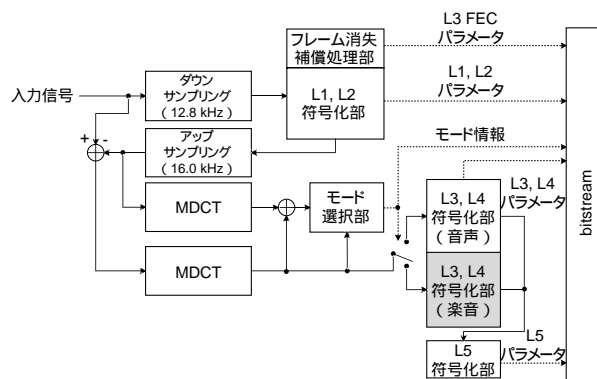
レイヤ	要求品質
L1 (8 kbit/s)	G.729@8 kbit/s
L2 (12 kbit/s)	G.729E@11.8 kbit/s
L3 (16 kbit/s)	G.722.2@12.65 kbit/s
L4 (24 kbit/s)	G.722@48 kbit/s
L5 (32 kbit/s)	G.722@56 kbit/s

2.3 アルゴリズム概要

G.718符号化部は、下位レイヤ (L1, L2) にはACELP (Algebraic Code-Excited Linear Prediction) と呼ばれる低ビットレートでの音声符号化に適した符号化方式を用い、上位レイヤ (L3-L5) にはMDCT (Modified Discrete Cosine Transform) による変換符号化方式を用いる (第2図)。まず、サンプリング周波数が16 kHzである入力信号を12.8 kHzにダウンサンプリングし、L1, L2で符号化を行う。次に、16 kHzにアップサンプリングしたL1, L2の復号信号と入力信号との誤差信号のMDCT係数 (ターゲットMDCT係数) をL3-L5で符号化する。

L3-L5では、まず入力信号が音声信号であるか楽音信号であるかを判定する。音声信号であると判定された場合には、L3, L4において、音声信号の符号化に適したAVQ (Algebraic Vector Quantization) によりターゲットMDCT係数を符号化する。一方、楽音信号であると判定された場合、L3ではパナソニックグループの技術であるBS-SGCを用い、L4では高ビットレートにおいて高音質を実現するFPC (Factorial Pulse Coding) を用いて符号化する。L5では入力信号の種類にかかわらずFPCにより符号化する。第3表に各レイヤのビットレートと符号化技術を、第4表にG.718のアルゴリズム遅延、演算量などを示す。

G.718には、BS-SGCのほか、代数的符号帳の高音質直交化探索技術、固定音源符号帳の低演算探索技術 (L1), 高性能フレーム消失隠蔽処理 (L1, L2) など、多くのパ



第2図 G.718符号化部の構成

Fig. 2 Block diagram of G.718 encoder

ナソニックグループの技術が採用されている。本稿では、3章において、特に性能向上に大きく寄与したL3におけるBS-SGC（第2図中のハイライト部）について説明する。

第3表 G.718のレイヤ構成

Table 3 Layer structure information of G.718

レイヤ	ビットレート	技術	サンプリングレート
L1	8 kbit/s	マルチモード コアレイヤ(ACELP)	12.8 kHz
L2	+4 kbit/s	ACELP拡張レイヤ	12.8 kHz
L3	+4 kbit/s	FEC AVQ/BS-SGC	12.8 kHz 16 kHz
L4	+8 kbit/s	AVQ/FPC	16 kHz
L5	+8 kbit/s	FPC	16 kHz

第4表 G.718の遅延・演算量

Table 4 Delay and complexity of G.718

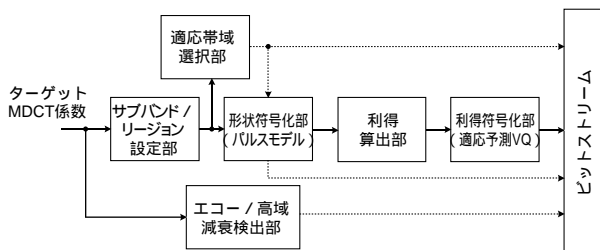
アルゴリズム遅延 (ms)	演算量 (WMOPS)	RAM (kwords)	ROM (kwords)
42.875 (WB input)	70.0 (WB input)	25.7	31.8
43.875 (NB input)	63.0 (NB input)		

WMOPS : Weighting Million Operating Per Second 1 word = 2 Byte
 NB : Narrow Band (-3.4 kHz) WB : Wide Band (-7.0 kHz)

3. 帯域選択型形状利得符号化方式 (BS-SGC)

第3表に示すように、L3で量子化に利用できるビットはわずか4 kbit/s（内、0.8 kbit/sはFEC（Frame Erasure Concealment）に使用）である。それにもかかわらず、L3ではL1、L2よりも帯域の広い信号を量子化する最初のレイヤであるため、量子化時のビットが大きく不足する。またL1、L2では音声信号の符号化に適したモデルをベースとした符号化処理を行っているため、楽音信号に対する符号化性能はそれほど高くない。そのため、L3では特に楽音信号に対して要求品質を達成することが大きな課題となっていた。

この課題に対し、筆者らは楽音信号に対して低ビットレート（+3.2 kbit/s）で高い符号化性能を達成するBS-SGCを開発した。BS-SGCは、まずターゲットMDCT係数を複数のサブバンドに分割し、各フレームで量子化する帯域を適応的に選択する。次に、選択された帯域のターゲット



第3図 BS-SGC符号化部の構成

Fig. 3 Block diagram of BS-SGC encoder

MDCT係数を形状利得符号化方式で量子化する。第3図にBS-SGCの構成、第5表にBS-SGCのビット配分を示す。

第5表 BS-SGC符号化部のビット配分

Table 5 Bit allocation of BS-SGC encoder

パラメータ	ビット数	
モード情報（音声 / 楽音）	1	
リージョン情報	2	
エコー検出情報	1	
高域減衰情報	1	
メインパルス	位置	20
	符号	5
サブパルス	位置	21
	符号	4
利得情報	9	
合計	64	

3.1 適応量子化スペクトル帯域選択方式

〔1〕楽音信号の符号化に対する課題

先述したように、L3においては量子化に利用できるビットレートは4 kbit/sと低く、全帯域（0~7 kHz）のターゲットMDCT係数を量子化することは難しい。このため、量子化する帯域をフレームごとに部分的に選択する方式を検討した。ここで、単純にエネルギーの大きいサブバンドから優先的に量子化する構成では、エネルギーが比較的少ない傾向にある高域サブバンドが常に選択されず、復号信号に帯域感が不足しこもった音に聴こえてしまうという問題がある。一方、高域サブバンドを優先的に量子化した場合には、エネルギーの大きい低中域サブバンドが量子化されず、低中域の量子化歪（ひずみ）が増加し、復号信号の音質が劣化してしまう。

〔2〕帯域感向上と量子化歪削減の両立

BS-SGCでは、エネルギーの大きい低中域サブバンドと帯域感の向上には不可欠である高域サブバンドを組み合わせ、量子化する帯域をフレーム単位に適応的に選択する。この方式により、高帯域感と低歪を両立した復号信号を生成することができる。

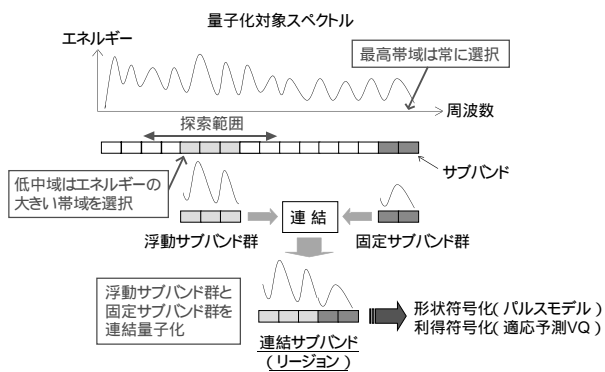
具体的には、まずターゲットMDCT係数を17のサブバンドに分割し、量子化対象とする5つのサブバンドを選択する。この時、選択されるサブバンドは固定サブバンド群と浮動サブバンド群とに分類される。固定サブバンド群は最も高域の2サブバンドであり、浮動サブバンド群は低中域の中で最もエネルギーの大きいサブバンド群である。前者は全フレームを通して固定であり、後者はフレームごとに適応的に全4候補の中から選択される（(1)式）。浮動サブバンド群の選択候補は、予備実験の結果、効率的に量子化できることが確認された4候補に限定している。これにより、わずかな情報（2ビット）で選択帯域を符号

化している。(1)式において、 $w(k)$ は聴覚重みであり、 k は選択用インデックス j で指定される各浮動サブバンド群に含まれる周波数インデックスである。また、(1)式において、 $P(j)$ を最大とする j が示す帯域を量子化対象帯域とする。

$$P(j) = \sum_{k=Start(j)}^{End(j)} (w(k) \cdot S(k))^2 \quad (j=0,1,2,3) \dots \dots \dots (1)$$

ここで、(1)式中の $Start(j)$ 、 $End(j)$ は、それぞれ浮動サブバンド群の最低周波数および最高周波数を表す。

次に、選択された浮動サブバンド群と固定サブバンド群を連結し(以下、連結されたサブバンドをリージョンと記す)、リージョン単位に形状利得量子化を行う。このように、フレーム単位に適応的に量子化する帯域を選択し、選択した低中域サブバンドと高域サブバンドを組み合わせ効率的にベクトル量子化する方式により、帯域感の向上と低中域成分に対する量子化歪の減少が両立でき、復号信号の品質を大きく改善できる。選択した量子化対象帯域のMDCT係数は、後述する形状利得符号化方式により効率的に量子化される。第4図に、上述したスペクトル帯域選択方式の処理概念図を示す。



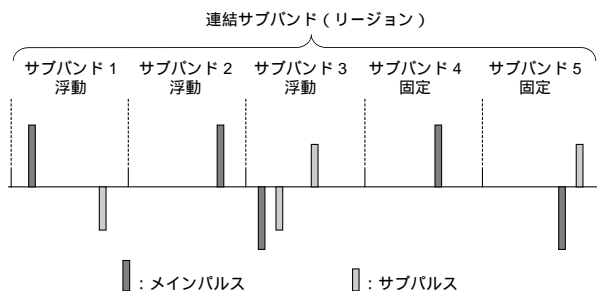
第4図 サブバンド/リージョン構成概念図
Fig. 4 Illustration of region formation

3.2 パルスモデル形状符号化方式

BS-SGCにおけるスペクトルの形状符号化では、予備検討で得られた知見である「広帯域音声において、人間の聴覚はスペクトルパルスの振幅よりもスペクトルパルスの周波数に敏感である」という特性を生かし、スペクトルの形状を大きさのそろった少数のパルスで表す符号化モデルを導入した。この方式により、パルスの振幅と比較して聴感に与える影響が大きいパルスの位置(周波数)により多くのビットを配分することができ、高品質な音声(low-bit-rate)で実現できる。

具体的には、まず5つのサブバンドごとに大きさが固定

であるパルス(メインパルス)を1本探索し量子化する。次にリージョン全体に対し、メインパルスより若干大きさが小さいパルス(サブパルス)を4本探索し量子化する。このように、探索範囲の異なる2種類のパルスを用いることにより、リージョン内の5つのサブバンド間でバランスよくパルスを量子化することができる。なお、本方式では各パルスの振幅情報は符号化せず、位置情報および符号情報を符号化している。本方式により量子化されたスペクトルの形状の一例を、第5図に示す。



第5図 形状符号化部の処理概念図
Fig. 5 Illustration of shape coding

3.3 適応予測利得ベクトル量子化方式

少ないビットでスペクトルの利得成分を量子化する手法としては、予測ベクトル量子化方法が挙げられる。しかし、3.1節で述べた方式により選択される帯域はフレームごとに異なるため単純な予測ベクトル量子化を適用することはできない。そこで、BS-SGCでは、予測ベクトル量子化の適用をフレームごとに切り替える構成を採用した。

(1) 予測ベクトル量子化方式

時間的に1つ前のフレームにおいて量子化されたリージョンと、現フレームのリージョンとのいずれにおいても選択されたサブバンド(共通サブバンド)の数が閾値(しきいち)以上である場合に対してのみ、ベクトルの全要素に対してフレーム間で予測ベクトル量子化(PVQ: Predictive Vector Quantization)を行う。この時、PVQでは(2)式の二乗誤差 $D(i)$ を最小にするコードインデックス i を算出する。ここで、(2)式中の G_{ideal} 、 G_{state} 、 G_{code} 、 α_1 はそれぞれ、量子化対象の理想利得、過去のフレームの量子化利得値、現フレームの量子化利得候補値、予測係数を示す。

$$D(i) = \sum_{j=0}^4 \left\{ G_{ideal}(j) - \sum_{t=1}^3 (\alpha_1(t,j) \cdot G_{state}(t,j)) - \alpha_1(0,j) \cdot G_{code}(i,j) \right\}^2 \dots (2)$$

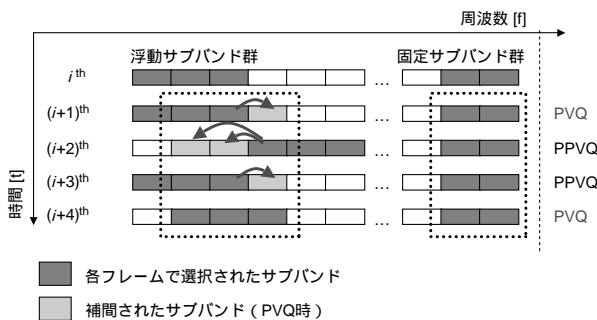
〔2〕部分的予測ベクトル量子化方式

一方、共通サブバンドの数が閾値未満である場合は、固定サブバンド群に対してのみPVQを適用する、部分的予測ベクトル量子化 (PPVQ : Partial PVQ) を行う。この時、PPVQは(3)式中の二乗誤差 $D(i)$ を最小にするコードインデックス i を算出する。ここで、(3)式中の α_2 は予測係数であり、それ以外の変数は(2)式と同様である。

$$D(i) = \sum_{j=0}^2 \{G_{ideal}(j) - G_{code}(i, j)\}^2 + \sum_{j=3}^4 \left\{ G_{ideal}(j) - \sum_{l=1}^3 (\alpha_2(t, j) \cdot G_{state}(t, j)) - \alpha_2(0, j) \cdot G_{code}(i, j) \right\}^2 \quad (3)$$

〔3〕PVQ/PPVQの適応切り替え方式

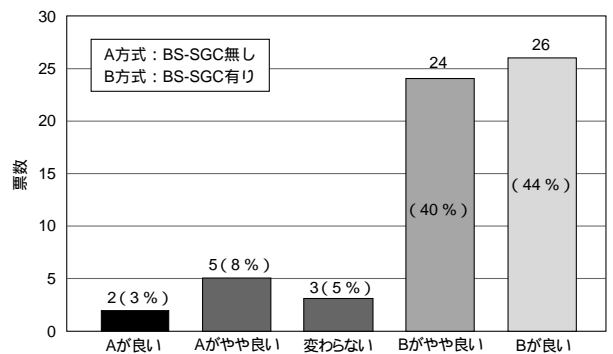
このように、フレームごとにPVQの適用可能性を判定しPVQとPPVQを切り替える構成により、フレーム間で量子化する帯域が変化する場合に対しても効率的な予測量子化が実現できる。第6図は、PVQとPPVQとの切り替えについて図解したものである。第 $i+2$ 、第 $i+3$ フレームでは、共通サブバンドが固定サブバンド群のみであるためPPVQを行う。一方、第 $i+1$ 、第 $i+4$ フレームでは共通サブバンドの数が一定数 (=3) 以上であるため、PVQを行う。なお、図中破線で囲んだサブバンドは第 $i+4$ フレームにおいて過去の3フレームの量子化値を利用する4次の線形予測によるPVQを行う場合に用いられるサブバンドであることを表す。ここで、過去3フレームで量子化していないサブバンドについては隣接するサブバンド(過去3フレームで量子化したサブバンド)から補間し、補間した値をPVQに用いる。この方式によりベクトルの全要素に対して効率的に予測量子化を適用できる。



第6図 適応予測ベクトル量子化の処理概念図
Fig. 6 Illustration of adaptive predictive VQ

4. 性能評価

L3 (16 kbit/s) にBS-SGCを適用することによる効果を確認するための主観評価試験を行った(第7図)。試験方法には片耳受聴の一对比較法を用い、AとBの2方式で処理したデータを対にして試聴し、5段階(Aが良い、Aがやや良い、変わらない、Bがやや良い、Bが良い)で評価した。被験者はエキスパート5名であり、評価データには合計約54秒の6種類の楽音信号(クラシック2、オペラ2、ポップス1、ジャズ1)を用いた。第7図において、A方式は入力信号の種類にかかわらず常にL3にAVQ方式を用いる構成であり、B方式は入力信号が楽音信号と判定された場合にはL3にBS-SGCを用いる構成である。試験の結果、L3にBS-SGCを適用することにより、楽音信号に対する符号化性能の大幅な改善を確認した(84%がBS-SGCの適用を支持)。



第7図 BS-SGCに対する主観評価試験結果
Fig. 7 Subjective test result for BS-SGC

5. まとめ

2008年6月に承認された国際標準ITU-T新規勧告G.718の概要、および筆者らが開発した低ビットレート楽音信号符号化方式BS-SGCについて述べた。BS-SGCについては、主観評価試験により本方式の有効性を実証した。BS-SGCを含めた多くのパナソニックグループの技術は、他社によるクロスチェックにおいても有効性が確認され、その結果G.718に採用された。これらパナソニックグループの技術は、G.718の特長的な処理の高性能化に大きく貢献している。

G.718は、IP網での高品質VoIPサービスを実現する上で有用な数々の特長を有しており、次世代移動通信用コーデックの有力候補と目されている。現在、更なる広帯域化 (~14 kHz) と多チャンネル化の機能を追加する拡張標準を策定中であり、次世代のコーデックにふさわしい魅

力ある機能を実現するため、今後も追加機能提案および実用化に向けた要素技術開発に取り組む。

参考文献

- 1) ITU-T : G.718;Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s. ITU-T Recommendation G.718 (2008).
- 2) M. Oshikiri, et al. : An 8-32 kbit/s scalable wideband coder extended with MDCT-based bandwidth extension on top of a 6.8 kbit/s narrowband CELP coder. Proc. Interspeech, pp.1701-1704 (2007).
- 3) 江原宏幸 他 : 高効率CELP符号化と帯域選択型MDCT符号化を用いた32kbit/sスケーラブル広帯域音声・オーディオ符号化方式の開発 日本音響学会誌 64, No.4, pp.196-207 (2008).
- 4) M. Xie, et al. : Embedded algebraic vector quantization (EAVQ) with application to wideband audio coding. Proc. IEEE ICASP 1, pp.240-243 (1996).
- 5) U. Mittal, et al. : Low complexity factorial pulse coding of MDCT coefficients using approximation of combinatorial functions. Proc. IEEE ICASSP 1, pp.289-292 (2007).

著者紹介



山梨智史 Tomofumi Yamanashi
東京R&Dセンター 次世代モバイル開発センター
Next-Generation Mobile Communications
Development Center, Tokyo R&D Center



押切正浩 Masahiro Oshikiri
東京R&Dセンター 次世代モバイル開発センター
Next-Generation Mobile Communications
Development Center, Tokyo R&D Center
工学博士



森井利幸 Toshiyuki Morii
東京R&Dセンター 次世代モバイル開発センター
Next-Generation Mobile Communications
Development Center, Tokyo R&D Center



佐藤 薫 Kaoru Satoh
東京R&Dセンター 次世代モバイル開発センター
Next-Generation Mobile Communications
Development Center, Tokyo R&D Center



江原宏幸 Hiroyuki Ehara
東京R&Dセンター 次世代モバイル開発センター
Next-Generation Mobile Communications
Development Center, Tokyo R&D Center
工学博士



吉田幸司 Koji Yoshida
東京R&Dセンター 次世代モバイル開発センター
Next-Generation Mobile Communications
Development Center, Tokyo R&D Center