社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－AVM－57（15）
2007／7／20

高品質 VoIP サービスを想定した超広帯域音声符号化技術

菊入 圭　仲 信彦　阿部 真也

（株）NTT ドコモ 総合研究所 〒239-8536 神奈川県横須賀市光の丘 3-5
E-mail: {kikuiri, nobuhiko.naka, abeshi}@nttdocomo.co.jp

あらまし 高速無線通信環境下での高品質な VoIP（Voice over Internet Protocol）サービスを想定して開発した超広帯域音声符号化技術について，概要および主観品質を述べる．本技術は，ビットレート 48k～64kbit/s，サンプリング周波数 22.05kHz 以上で動作し，音声信号に対して，既存の超広帯域音声符号化技術である ITU-T G.722.1 Annex C，および MPEG-4 AAC-LD（Advanced Audio Coding - Low Delay）と同等の主観品質である．本技術で符号化・復号された信号は，人間の音声信号に含まれるほぼ全ての周波数成分を含むため，従来の電話帯域音声符号化技術に比べ，格段に聞き取りやすく，臨場感のある通話サービスを実現する．このような特徴を生かし，電話会議，遠隔教育，常時接続型通話などのアプリケーションが考えられる．

キーワード 超広帯域音声符号化技術，VoIP

Super-wideband Speech Coding for High Quality VoIP Services

Kei KIKUIRI　Nobuhiko NAKA　and　Shinya ABE

Research Laboratories, NTT DoCoMo, Inc.　3-5 Hikari-no-oka, Yokosuka, Kanagawa, 239-8536 Japan
E-mail: {kikuiri, nobuhiko.naka, abeshi}@nttdocomo.co.jp

Abstract This paper describes the outline and the subjective quality of a super-wideband speech coding, which we have developed for high quality VoIP (Voice over Internet Protocol) services over a high speed radio communication. It operates at the sampling rate of 22.05 kHz and more, and the bit rate of 48k - 64kbit/s, and offers equivalent subjective quality to the international standard codings ITU-T G.722.1 Annex C and MPEG-4 AAC-LD (Advanced Audio Coding - Low Delay). It provides users with much more comfortable listening environment and more presence than conventional narrowband speech codings. These features can realize such applications as conference call, remote education, and always-online call.

Keyword Super-wideband speech coding, VoIP

1. はじめに

近年の音声符号化技術の研究開発は，より広い帯域の音声信号を 20k～64kbit/s 程度で符号化し，従来の電話網を上回る音質を提供することを目的とした技術に注目が集まっている．これは，ADSL や光ファイバによる IP（Internet Protocol）ベースの高速パケットアクセス回線と，その上で実現された VoIP（Voice over IP）による電話サービスが普及したことにより，伝送速度と符号化方式の制約が緩和されたことによる．通話に使用する音声信号を広帯域化できれば，電話帯域といわれる 300Hz から 3400Hz 以外の低域や高域の音声信号成分が使用できるようになり，本来の音声により近い自然な音声を再現できる．そのため，VoIP による通話サービスには，従来テレビ会議システムが主な用途であった広帯域（50Hz - 7kHz）音声符号化や，超広帯域（上限周波数 7kHz 以上）音声符号化を採用したものが存在する．また，電話帯域音声符号化である G.711[2]や G.729[3]をベースとしたエンベデッド型の広帯域符号化である，UEMCLIP（mU-law Embedded Coder for Low-delay IP communication）[4]や G.729.1[5]も開発されている．

一方，移動通信においても法人用途に W-LAN を利用した VoIP サービスが提供されており，3GPP の Long-Term Evolution といったリアルタイム伝送が可能な高速無線アクセス方式の技術検討が進められるなど，広帯域以上の音声による高音質な通話サービスが可能となる条件が整いつつある．

表 1　VoIP で用いられる音声符号化技術の仕様

	G.711	G.729 Annex A	G.722.1	G.722.1 Annex C	AAC-LD	超広帯域音声符号化
サンプリング周波数 (kHz)	8	8	16	32	22.05~48	22.05~32
ビットレート (kbit/s)	64	8	24, 32	24, 32, 48	12~160	48~64
フレーム長 (ms)	-	10	20	20	10~23.22	8~11.61
原理遅延 (ms)	0.125	15	40	40	20~46.44	16~23.22
対応する音声信号の帯域	電話帯域		広帯域	超広帯域		
符号化アルゴリズム	PCM	CELP	変換符号化			

本稿では，高速な無線通信環境下における VoIP サービスを想定し，肉声に近い自然な音声品質を目指して開発した，超広帯域音声符号化技術の概要および主観品質を報告する．また，超広帯域音声符号化技術を利用した，高速無線通信環境下の高品質 VoIP のアプリケーションについても述べる．

2. VoIP で用いられる音声符号化技術

現在の VoIP アプリケーションで利用されている音声符号化技術の仕様を表 1 に示す．

G.711 は，固定網との互換性も考慮され，多くの VoIP アプリケーションで利用されている．PCM (Pulse-Code Modulation)アルゴリズムを用いた符号化であるため，原理遅延はわずか 0.125ms である．さらに，パケット損失などのエラーにも強いパケットロスコンシールメントが併せて勧告に記載されている．

G.729 Annex A (G.729A) [3]は，G.729 の低演算量版であり，人間の音声発生機構をモデル化した CELP (Code Excited Linear Prediction) 符号化方式を採用し，8kbit/s で G.711 に近い音質を実現している．

G.722.1 [6]は，主に会議システム用途に標準化された広帯域音声符号化であり，VoIP アプリケーションでも利用されつつある．その拡張版である G.722.1 Annex C (G.722.1C) [6]は，14kHz 帯域の音声信号に対応し，後述の AAC-LD (Advanced Audio Coding-Low Delay) [7]に比べて，低ビットレート，低演算量であるが，原理遅延は 40ms と長い．

AAC-LD は，音楽配信などで利用されている MPEG-4 AAC [7]を，双方向通話が可能なように低遅延化したものである．音響符号化用の聴覚心理モデルを用いるために，多くの演算量を必要とする．

G.722.1 や AAC-LD は，時間領域の信号を周波数領域に変換して符号化するアルゴリズム（変換符号化方式）を採用している．これらは発声モデルを使わないため，音声以外の信号に対しても柔軟に対応できる．

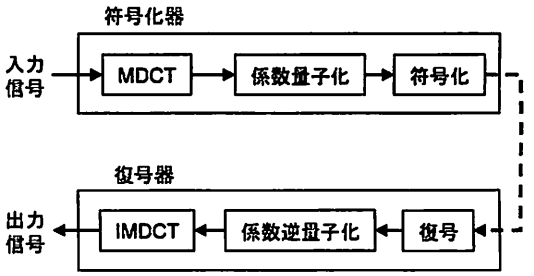


図 1　超広帯域音声符号化の構成

3. 超広帯域音声符号化技術

3.1. 仕様と構成

表 1 には，開発した超広帯域音声符号化の仕様も示されている．本符号化は，22.05kHz 以上のサンプリング周波数で動作することで符号化の対象となる音声の上限周波数を 10～16kHz へ拡大し，人間の音声に含まれるほぼ全ての成分をカバーしている．また，ビットレートは 48k～64kbit/s で任意に設定可能であり，ネットワークやアプリケーションに応じて柔軟に対応可能である．原理遅延は，既存の超広帯域音声符号化よりも短く，従来の電話帯域の音声符号化と比較してもほぼ同等であり，音声通話に適している．

本符号化は，G.722.1 や AAC-LD と同様に，変換符号化方式を採用している．変換符号化方式は，周波数領域における人間の聴覚特性を利用した処理を施しやすいという利点がある．また前述の通り，音声以外の信号に対しても柔軟に対応可能であるが，音声信号のみに対する符号化効率としては CELP 符号化方式に劣る．

図 1 に本符号化の構成を示す．符号化器において，時間領域の入力信号を MDCT(Modified Discrete Cosine Transform, 修正離散コサイン変換)により周波数領域に変換し，変換係数を量子化，符号化して復号器へ伝送される．復号器では，受信データを復号，逆量子化して得られた変換係数を，IMDCT(Inverse Modified Discrete Cosine Transform, 逆修正離散コサイン変換)により時間信号に変換し出力する．また本

表 3　主観評価試験条件

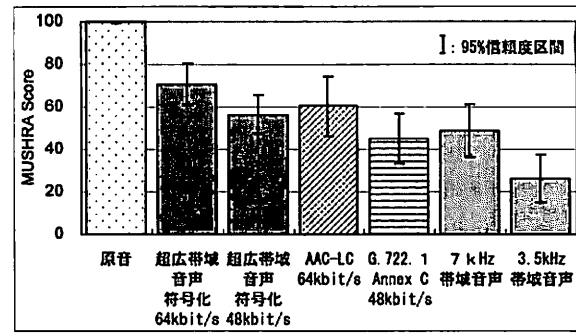
試験方法	MUSHRA 法
被験者数	19 名
参照音声 (サンプリング周波数)	原音 (男声, 女声, 音声+BGM) (32 kHz)
符号化音声 (ビットレート, サンプリング周波数)	超広帯域音声符号化 (48, 64 kbit/s, 22.05 kHz) AAC-LC (64 kbit/s, 48 kHz) G.722.1 Annex C (48 kbit/s, 32 kHz)
帯域制限音声	7 kHz 帯域 3.5 kHz 帯域
受聴方法	ヘッドホン両耳受聴

符号化では，VoIP での利用を想定したオプションとして，パケットロスコンシールメントと 1 パケット内に閉じたフレーム間予測がある．フレーム間予測は，複数フレームを 1 パケットに多重化した際のみ有効となり，予測範囲をパケット内に閉じることで，パケットロス等に対する耐性を保ちつつ，符号化効率を向上できる．

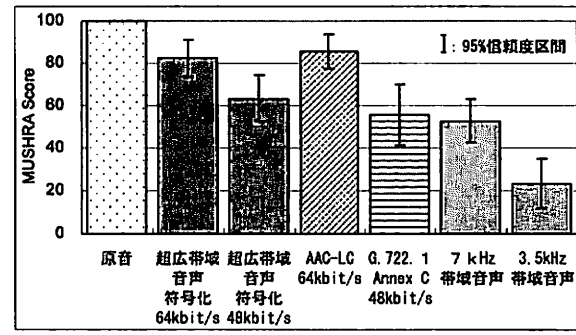
3.2. 主観品質

本符号化技術の性能を評価するため，音声信号を用いた主観評価試験を行った．試験条件を表 2 に示す．試験方法は MUSHRA (MUlti Stimulus test with Hidden Reference and Anchor)法[8]を用いた．比較する符号化技術は，G.722.1 C および AAC-LC (Low-Complexity)[7]とした．AAC-LC は AAC の低演算量版であり，AAC-LD とほぼ同程度の主観品質であることが報告されている[9]．また，3.5 kHz 帯域制限音声は電話帯域音声符号化の上限品質と，7 kHz 帯域制限音声は広帯域音声符号化の上限品質とほぼ同等である考えられる．

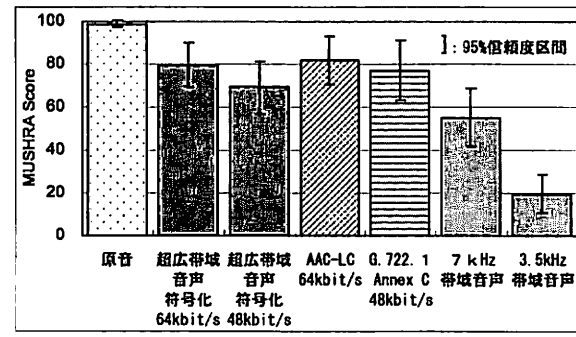
試験結果を図 2 に示す．これらより，超広帯域音声符号化は，64 kbit/s において AAC-LC と同等であり，7kHz 帯域制限音声より有意によいことがわかる．一方，48 kbit/s において，G.722.1 C と同等であり，3.5 kHz 帯域制限音声より有意によいことがわかる．すなわち，超広帯域音声符号化の主観品質は，従来の電話帯域音声符号化，および広帯域音声符号化よりも有意によく，G.722.1 C および AAC-LD と同等であると言える．



(a)女声



(b)男声



(c)音声+BGM

図 2　主観評価試験結果

4. 高品質 VoIP のアプリケーション

4.1. アプリケーション例

超広帯域音声符号化による高品質 VoIP では，従来の VoIP で提供されていたアプリケーションの高度化に加えて，高品質であるからこそ従来の電話帯域音声では想定されなかったアプリケーションも考えられる．このような高品質 VoIP を高速無線通信環境下において利用することで，想定されるアプリケーションは広がる．以下にアプリケーション例を紹介する．

A)　電話会議

電話会議は，従来利用されてきたアプリケーションであるが，超広帯域音声符号化を用いた高品質 VoIP により，音声品質・受聴了解度が飛躍的に向上する．また超広帯域音声符号化では，日常，直接話をしている際に感じている個人の音声信号の特徴の一部や息遣いのような話し方の特徴がより含まれていることで，話者の識別も容易になる．さらに，

現在の電話会議システムは固定の電話網や IP 網に接続して利用されているが，高速無線通信環境下であれば場所を選ばず利用可能となる．

B) 遠隔教育

固定電話および携帯電話や専用の電話端末を用いた遠隔教育は，特に語学学習を中心として行われている．最も多くの人が学習していると思われる英語は，母音中心の日本語と異なり，子音中心の言語である．したがって，電話帯域音声および広帯域音声では制限されていた高周波数成分を多く含んでおり，これが学習の妨げになり得る．一方，音声信号のほぼ全ての周波数成分を含む超広帯域符号化であれば，このような制限はない．さらに高速無線通信環境下で利用することで，どこにいても学習することができ，より学習効果の高い遠隔教育が提供できる．

C) 常時接続型通話

従来の通話，とくに携帯電話による通話は，目的志向が強く，用件を伝えるために必要な時間だけ通話するという形態が中心であるといわれている．これは当然ながら利用料金によるところもあるが，通話音声の品質の観点から考えると，従来の電話帯域音声は要件を伝えるには十分な品質であるが，長時間通話するためには向いていないためと考えられる．電話帯域音声による通話では，非常に集中して聞く必要がある特殊な状況であり，ユーザはストレスを感じる．一方，超広帯域音声符号化の品質であれば，通常の会話と同様にストレス無く相手の話を聞くことができ，長時間通話も可能である．高速無線通信環境下であれば，常に通話相手と接続しておき，どこにいても，話したいときに話したいだけ通話するといった利用方法が可能となる．

4.2. 高品質 VoIP プロトタイプソフトウェア

Windows Mobile 5.0 上にてリアルタイムの符号化・復号を可能にする超広帯域音声符号化モジュールを実装した．携帯電話上での高品質 VoIP サービスを想定して，Windows Mobile 5.0 を搭載したスマートフォンで動作する本モジュールを搭載した高品質 VoIP プロトタイプソフトウェア（図 3）を作成した．本ソフトウェアでは，音声データの伝送には RTP (Real-time Transport Protocol)を，呼制御には SIP (Session Initiation Protocol)を採用している．

5. まとめ

本稿では，高品質 VoIP サービスを想定して開発した超広帯域音声符号化技術について，概要と主観品質を述べた．本技術は，変換符号化方式をベースとしており，同程度の周波数帯域の音声信号を対象としている既存の符号化技術に対して，同ビットレートで同等の主観品質であることを確認した．また本技術による，高速無線通信環境下における高品質 VoIP アプリケーションの例を紹介した．

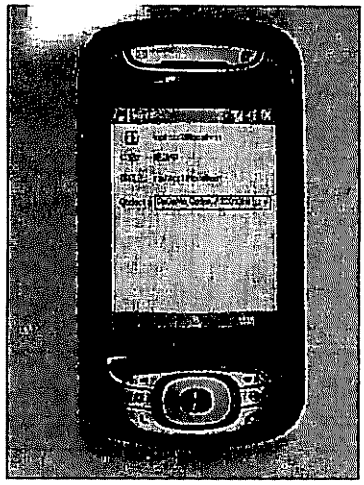


図 3 高品質 VoIP プロトタイプソフトウェア

今後は，本技術の更なる高品質化，機能拡張に向け，開発を進める．

文 献

[1] 3GPP TS26.090, “Adaptive Multi-Rate (AMR) speech codec; Transcoding functions,” 1999.

[2] ITU-T Recommendation G.711, “Pulse Code Modulation (PCM) of Voice Frequencies”, 1988.

[3] ITU-T Recommendation G.729, “Coding of Speech at 8 kbit/s using conjugate structure algebraic-code-excited linear prediction (CS-ACELP),” 2007.

[4] Y. Hiwasaki, H Ohmuro, T. Mori, S Kurihara, and A. Kataoka, “A G.711 Embedded Wideband Speech Coding for VoIP Conferences,” IEICE Trans. Inf. And Syst, Vol. 89-D, No. 9, Sept. 2006.

[5] ITU-T Recommendation G.729.1, “G.729 based Embedded Variable bit-rate coder: An 8 – 32 kbit/s scalable wideband coder bitstream interoperable with G.729,” 2006.

[6] ITU-T Recommendation G.722.1, “Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss,” 2005.

[7] ISO/IEC 14496-3: 2001, “Information technology — Coding of audio-visual objects — Part 3: Audio,” 2001.

[8] ITU-R Recommendation BS.1534-1, “Method for the subjective assessment of intermediate quality level of coding systems,” 2003.

[9] E. Allamanche, R. Geiger, J. Herre, and T. Sporer, “MPEG-4 Low Delay Audio Coding Based on the AAC Codec,” 106th AES Convention, Munich, Germany, May 1999.