

## 8 kbit/s 音声符号化ITU-T標準の演算量削減版 (Annex A)勧告の経緯と品質評価

林 伸二・片岡 章俊<sup>+</sup>・栗原 祥子<sup>+</sup>

History of standardizing complexity-reduction algorithm (Annex A) to 8-kbit/s G.729  
in ITU-T and its Quality Assessment

HAYASHI Shinji, KATAOKA Akitoshi, and KURIHARA Sachiko

### Abstract

This paper presents the history of recommending Annex A to G.729 in ITU-T (International Telecommunication Union, Telecommunication Standardization Sector) and quality assessment of the coder. The requirements for this system were toll speech quality, low complexity, and short delay time at low bit-rate less than 12 kbit/s, and were not met by existing speech coding standards. Reduced complexity CS-ACELP was selected from among 5 candidates as the coding algorithm after extensive subjective selection tests and was recommended as Annex A to G.729. The subjective test in Japanese shows that this Annex is compatible with G.729 and meets all the requirements.

Key word : V.70 DSVD, V.34 MODEM, G.729 CS-ACELP, Middle and Low Bit-Rate Speech Coding

### [要約]

ITU-Tの8 kbit/s音声符号化方式G.729の演算量削減版がAnnex Aとして勧告された経緯とのアルゴリズムの特徴、及び、品質評価について述べる。Annex Aは、電話網上でモデムを介して、PC等の端末装置間でデジタルデータと音声とを同時に送受する方式DSVD (Digital Simultaneous Voice and Data)のITU-T勧告V.70に使われることを主目的として、G.729の演算量削減により開発された。DSVD用音声符号化式は、低ビットレート、商用電話品質、短遅延時間、低演算量が望まれ、また、伝送効率の要請からVAD (Voice Activity Detection) 機能の併用が必須である。G.729 CS-ACELPは、演算量を除いてほぼすべて要求を満たしていたが、Annex Aは演算量を大幅に削減し、DSVDの要求に合致させたものである。品質評価は、日本語ほか複数国語で行われ、G.729本体とAnnex Aが相互に接続可能であり、かつ、所定の品質条件を満たすことが明らかとなった。DSVD方式はその後、インターネット上のパケット音声通信にとって代わられるが、音声符号化標準は広くVoIP (Voice over Internet Protocol)などに使用されその用途が拡大している。

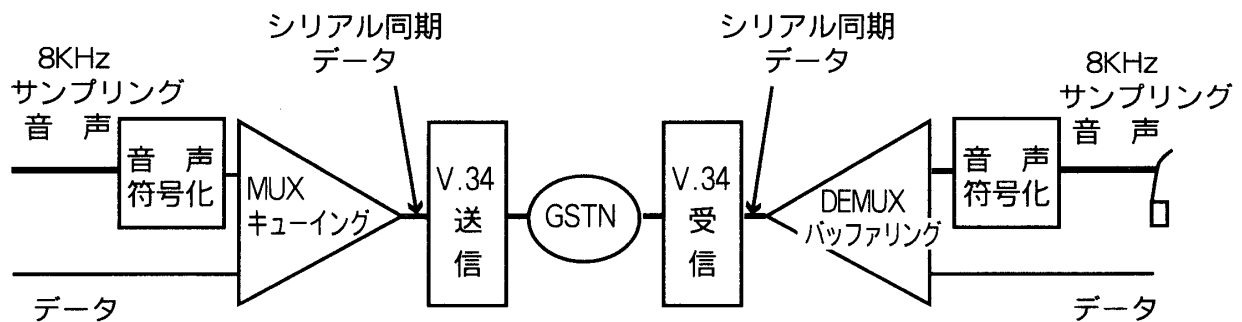
キーワード : V.70 DSVD, V.34 MODEM, G.729 CS-ACELP, 中低レート音声符号化方式

+ NTT研究所 東京都武蔵野市

## 1. まえがき

1996年にCS-ACELP(Conjugate Structure Algebraic Code Excited Linear Predictive、共約構造代数符号励振線型予測)符号化方式[1]がITU-T (International Telecommunication Union Telecommunication Standardization Sector: 国際電気通信連合標準化セクタ)において、8kbit/s音声符号化方式G.729[2]として勧告された。その主な用途は当初、次世代携帯電話(FPLMETS: Future Public Land Mobile Telecommunication System)とされていたが、FPLMETSのサービス開始は当初予定の1996年よりも大きく遅れることとなった。しかし、G.729は短い遅延時間と商用電話品質という特徴から、長距離、国際電話用の多重化装置や現行世代の携帯電話等に採用されていた。一方、パーソナルコンピュータ(PC)とモデムの急速な発達により、公衆網(PSTN: Public Switched Telephone Network)上のダイヤルアップで遠隔地とのデータ送受が容易に行われるようになった。データ送受の応用上は1回線をモデムに使いデータを送受しながら、他の1回線で会話してタスクを処理することが多いため、これを併せて1回線で済ませたいという要望が強かった。ITU-TのSG (Study Group: 研究部会)14で1996年に勧告されたV.70 DSVD (Digital Simultaneous Voice and Data)[3]は、この要望に応じて規定されたものであり、図1の構成をとりV.34などの高速モデムを用いた時の同期、伝送路誤り制御、フレーム構成、及び、プロトコル等が記述される。そこで使われる音声符号化方式は端末の能力に応じて複数種類の中から選ぶ余地があるが、音声品質や遅延時間、データのスループット等の面で所定の性能を発揮するために、G.729 Annex A[4]が勧告された後はこれが推奨されることとなった。本論文は、SG14からの要請に基づいてSG15 (信号処理、伝送システムを担当する部会)の音声符号化専門家会合で進められた、DSVD用音声符号化方式の要求条件の設定、候補案の提案受け付け、選択のための評価試験、および、検証のための品質試験と続いた標準化の経緯を述べ、また、選ばれた音声符号化アルゴリズムの構成、VAD(Voice Activity Detection)機能、品質評価法と評価結果について示す。

図1 V.70 DSVD ブロックダイアグラム



## 2. 性能要求条件

DSVDシステムへの適用を考慮して音声符号化アルゴリズムを選定するための性能要求条件が1995年2月のSG15会合で設定された[5]。以下の3項目を考慮した主な要求条件を表1に示す。標準候補案の選択にあたっては、「要求」を満たすことが優先され、ついで、複数の候補案が要求を満たした場合、「目標」を満たすことが考慮される。

(1)ビットレート：モデムの能力を考慮すると、低い程好ましい。14.4, 9.6 kbit/sのモデム速

度では、オーバーヘッドビットを除くと音声符号化に割り振られるビットレートはそれぞれ、13.0, 8.0 kbit/s程度以下となる。

(2) 遅延時間：エコー、会話効率を考えると短い程良いが、モデムのフレーム構成とデータのスループットを考慮した現実的な条件は10 msないし15 msフレームが望ましい。  
 (3) 演算量：音声符号化アルゴリズムをモデム用LSIの一部としてインプリメントしたり、PCのMPUでソフト的に動作することを想定し、符号、復号同時動作で10 MIPS (Mega Instructions Per Second)が上限とされた。

また、電話網で使われる既存の勧告G.728[6], G.729, G.723.1[7]の拡張として極力これらと共通の構成とすることが望ましいことがこの要求条件の前文で謳われている。

### 3. 標準アルゴリズム候補案の提案と選択試験

#### 3.1 アルゴリズム候補案

設定された要求条件のもとで5件の候補案が提案された。各提案の概略を表2に示す。要求条件を満たすか否かを判定する主観評価試験を実施し、標準候補案を選ぶこととなった。

#### 3.2 選択試験手法と実施

要求条件への合否を限られた評価資源で効率よく判定するため、以下の試験手法をとることとした。品質試験は大きく3つの実験に分けられる。実験1は符号誤りや背景雑音等の妨害要因がないとき、基本性能を明らかにするためのACR (Absolute Category Rating、付録1参照)試験[8]である。実験2は背景雑音の影響を調べるCCR(Comparison Category Rating)試験[8]であり、実験3は伝送路符号誤りの影響を調べるACR試験である。背景雑音が混じった音声に対する品質が重視されたのは、当時普及しはじめた第2世代デジタル無線電話に使われた符号化方式に、背景雑音によるSwirling効果 (背景雑音が変調されて渦巻くような音に聞こえる作用) が問題になったためである。背景雑音として、オフィス騒音、バブル騒音、第2話者、音楽が用いられた。オフィス騒音は事務室内に響く空調装置やオフィス機器の発する音、バブル

表1 DSVD用音声符号化方式の要求条件

条件	要求
エラーなし音声品質	32kbit/sG.726以上
検出されたフレーム消失時の音声品質 (3%フレーム消失)	エラーなしの32kbit/sG.726からMOS劣化が0.75以内
音声品質の入力レベル依存性 (過負荷点から-16dbから-36dbの範囲)	32kbit/sG.726より悪くないこと
片方向遅延時間	アルゴリズム遅延20ms以下 全符号復号遅延40ms以下
音声帯域データ(モデム)伝送性能	不要
シグナリング信号伝送性能	DT/MF信号を伝送できること
音楽信号伝送性能	-
ビットレート	12kbit/s以下
タンデム性能	2段タンデムの品質が4段タンデムの32kbit/sG.726以上
複雑度	演算量:10MIPS以下、RAM:2K words以下、ROM:8K words以下
背景雑音時の音声品質、バブル騒音、妨害話者	32kbit/sG.726より悪くないこと

表2 各提案方式の諸元

アルゴリズム提案元(国)略号	レート(kbit/s)/フレーム長(ms)/先読み(ms), MIPS/RAM(Kw)/ROM(Kw)	特徴
SVD-CELP AT&T(USA) A	7.73/15/5, 7.6/1.0/5.8	G.729と同一の基本構成、LSPスカラー量子化、マルチパルス励振
MP-MLQ CELP DSPグループ(イスラエル)B	8.8/10/8.75, 8.5/1.8/6.5	G.723.1と同一の基本構成、フレーム長を短縮
Dual-pulse CELP NTT(日本) C	7.8/15/5, 10/1.96/8.0	G.729と同一の基本構成、デュアルパルス励振
Spike CELP Rokwell(USA) D	8.0/15/5, 9.7/2.0/8.062	Log-area-ratioスカラー量子化、2段階ピッチ探索、パルスと学習励振源を併用
Complexity-reduced CS-ACELP シェルブルーク大(カナダ)E	8.0/10/5, 11.48/2.114/9.032	G.729のランダム励振部分の探索を簡略化

騒音は大勢の人がレストランやホールで話す人声、第2話者は文字通り本来の通話者以外の音声である。主音声の背景に対するS/Nは利用実態を反映するため、第2話者では15 dB、オフィス騒音、バブル騒音、音楽では20 dBとした。試験言語は、英語と日本語が用いられ、それぞれを母国語とするものが受聴した。英語はCOMSAT（合衆国）が、日本語はNTT（日本）が試験を実施した。試験音声の処理はNTTが担当し、両試験実施機関から原音声を集め、各提案機関から集めた、PC上で動作する音声符号化実行プログラムにより試験音性を作成し、各試験機関に返した。

### 3.3 試験結果

日本語を用いた各実験の結果を図2, 3に示す。図2は適正入力レベル(-26 dBov、付録2参照)における評価、図3は、3%ランダムフレーム消失時の各候補のレファレンス（G.726 ADPCMのMOS-0.75）との比較である。各候補案はいずれもレファレンスよりも良いか、有意な差のない範囲に入っている。図4-1は、バブル背景雑音条件(SNR = 20 dB)下のCCR試験の平均スコア(MCCS)値であり、いずれの候補もレファレンス（同一条件下のG.726 ADPCM）よりも有意に劣る。図4-2は妨害となる第2話者(SNR = 15 dB)のいる場合で、A, Dの候補以外は要求条件を満たしている。結果の平均スコア間の有意差は、式(1)に示す95%信頼幅CI(Confidence Interval)が用いられた。図中の□は、平均値、信頼度幅の上下端の値を示す

$$CI = t(95\%, 2(m-1)) \sqrt{\frac{\sigma^2}{m}} \quad (1)$$

ここに、mは該刺激音に対する総投票数、 $t(95\%, 2(m-1))$ はスチューデントのT分布における95%境界、 $\sigma$ は刺激音の投票値の標準偏差を表す。背景雑音環境下の試験にCCR法を用いると、すべての要求条件を満たす候補案はなかった。次善の候補として、それ以外の条件ですべての要求を満たし、背景雑音環境下で最もスコアの高い2案B, Eが統計的に同水準にあるため、以後の検討に供するよう残された。

SG15の専門家会合での議論の結果、候補案Eが既存のITU-T勧告G.729と相互接続性があり、ビットレートもより低いことが決め手となり、最終案に選定された。

図2 適正入力レベル下のMOS

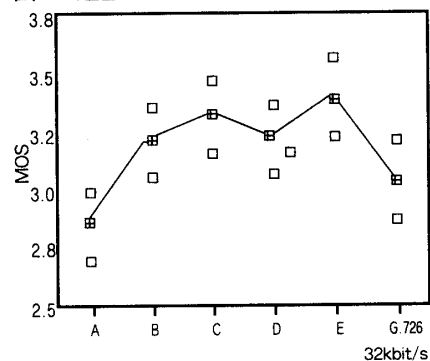


図3 3%ランダムフレーム消失下のMOS

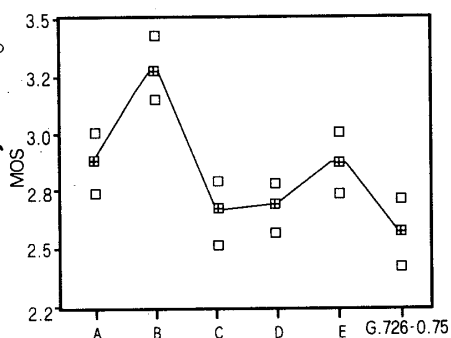
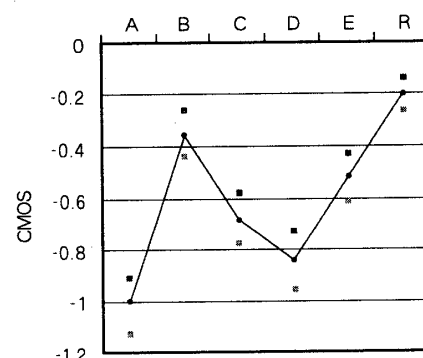


図4-1 バブル騒音下の平均比較スコア (CCR試験)



#### 4. 選定アルゴリズムの概要

##### 4.1 G.729の構成の概略勧告

G.729 CS-ACELP(本体)の構成を図5に示す。CS-ACELP符号化方式のフレーム長は10 ms (80サンプル)である。スペクトラム包絡の分析にはさらに前方に5 msはみ出した分析窓(先読み)を用いるため、アルゴリズム遅延時間は15ms、送受の遅延時間(伝送遅延時間を除く)は通常35 msとなる。スペクトル包絡合成フィルタの更新周期は10 msであるが、それ以外の処理は、時間的分解能の要求から半分の5 msサブフレーム(40サンプル)で行われる。励振信号源は、適応符号帳ベクトルと、ランダム励振代数符号帳ベクトルとを併用する。適応符号帳は、過去の音声波形の基本周期成分から生成され、高い基本周波数では、標準化周期の3倍の時間的精度を有する。また、ランダム励振も基本周期の繰返しと同期する。代数励振符号方式は代数的に励振符号を発生するため、符号帳に要するメモリが不要であり、最適符号探索時の演算量を節約できる。スペクトラム包絡には、LSP (Line Spectrum Pair: 線スペクトル対)のベクトル量子化が用いられている。振幅情報は、共約構造ベ

トル量子化手法を用いることにより、符号誤り耐性の強化と、符号帳に用いるメモリ量の削減がなされている。復号側では、各指標から励振ベクトル、振幅ベクトル、LSPパラメータが再現され合成音が作られる。また、ポストフィルタリングによりスペクトル特性の補正とFrame Erasure Concealment (消失フレーム復元)が行われる。この処理により、フレーム消失直前の基本周期、スペクトル包絡情報をもとにフレーム情報の外挿と振幅の制御がおこなわれ、聴感上の劣化が軽減される。

##### 4.2 G.729 と候補案の違い

候補案は、符号器について以下の(1)から(5)、復号器で(6)の点で簡略化を行い演算量を下げた。最も大きく演算量削減に寄与しているのは(5)であり、検索される励振パルスの可能性が狭まるものの、演算回数が2/3程度まで低下する。これらのアルゴリズムはそのまま勧告に採用されたため詳細の記述は勧告書に譲る。

(1)聴覚重みづけフィルタの係数を固定, (2)開ループピッチ分析において、重み付き音声の相関演算に間引きを使用, (3)重み付き合成フィルタ $W(z)/A(z)$ を $1/A(z/\gamma)$ により近似, (4)適応符号帳の検索で、相関のみを用い、振幅は無視, (5)代数符号帳の検索の簡略化。励振パルス位置

図4-2 妨害第2話者下の平均比較スコア (CCR試験)

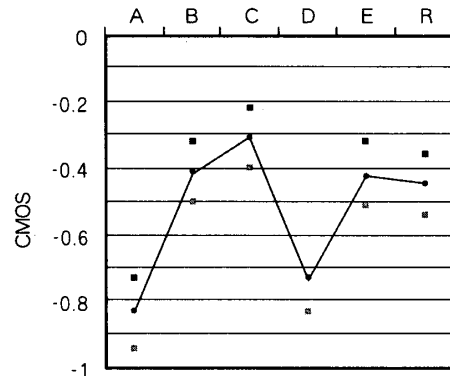
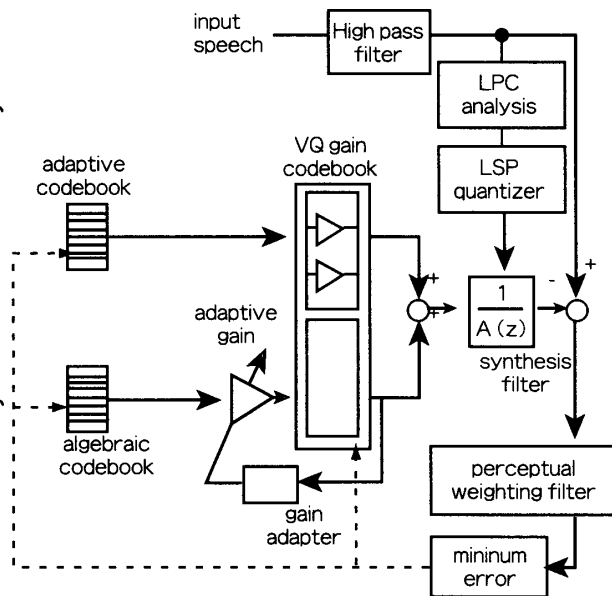


図5 CS-ACELP方式の構成



の検索で、G.729本体で使われているnested-loop 検索法から、繰り返しdepth-first tree 検索法を使用、(6)調波構造ポストフィルタにおいて整数精度のみを使用

#### 4.3 付随するVAD/CNGアルゴリズム

DSVD用途では、タスク遂行上会話以外の動作の占める比率が大きいため、VADアルゴリズムの併用が効果的である。G.729に適用するVADアルゴリズムは、音声符号化時のパラメータを利用する方式が提案された。単に無音区間の伝送を停止するのみでは語頭や背景雑音の切断が起こるため、雑音区間の特性変化を伝送し雑音の不自然さを避けるCNG (Comfort Noise Generation)アルゴリズムが併用された。同アルゴリズムは後にG.729 のAnnex B[9]として勧告された。

### 5. 検証試験

検証主観評価試験は、勧告候補案がG.729と相互接続可能であり、かつ、品質要求条件を満たすことを確認、検証するための試験である。試験はSG15の要請を受けて、SQEG (Speech Quality Experts Group)/SG12が実施法の詳細を詰めた[10]。試験は3つの実験に分かれており、実験1はG.729と候補符号化方式がbit stream compatibleであることを確認するため、それぞれの符号器、復号器を対向してACR法により評価する。実験2は、候補案の環境雑音や背景音楽下の相互接続特性と単独特性をCCR法により明らかにする。実験3は伝送チャネルがランダム符号誤りやフレーム消失を起こしたときの品質をACR法により明確化する。実験1は日本語(NTT)、実験2は日本語(NTT)、仏語(FT)、実験3は日本語(NTT)、伊語(FUB: Fondazione Ugo Bordoni)により行われた。音声処理はNTTが担当した。

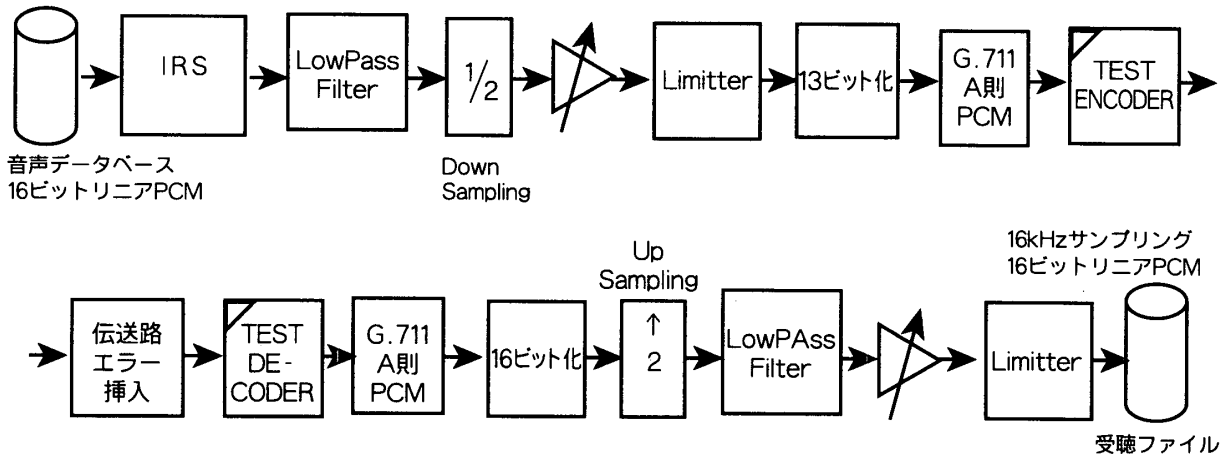
#### 5.1 音声処理

候補案の音声符号化方式は、1ビット精度で演算語長を規定する固定小数点演算を表すANSI-Cソースコードから生成されたPC上の実行形式で実現された。これは、DSPハードウェア上にアセンブラ言語で実現することと等価な処理であり、莫大となるDSPコード開発費用をかけることなく、実現可能性を検証し、また、正確に演算量を算定することができるという利益があり、以後の標準化において性能比較評価をするとき、この手法が踏襲されるものとなった。基本的な処理系統を図6に示す。同図のフィルタ、IRS (Intermediate Reference System) 特性[11]等化器、サンプリング周波数変換等のツールはITU-T勧告G.191[12]を使用した。音声符号化の前後にG.711 PCM符号化処理を含めることがSQEG から指定されたが、これは必ずしも妥当でない。なぜなら、DSVDシステムでは音声符号化は端末に置かれ、ネットワークからのトランスコーディングは稀と考えられること、今日のA/D、D/Aコンバータが13ビットリニアPCM以上の精度を十分有しているからである。G.711 PCMを除き、13ビット以上の精度のリニアPCMインタフェースを用いれば、品質の向上が望めることを念頭において、最悪品質を確かめる意味でこれが含まれた。

#### 5.2 評価手法

評価には基本的にACR法を用い、背景雑音・音楽条件下の試験はCCR法を用いる。これは、

図6 音声処理系統



低レートの符号化で背景雑音がSwirling Effectとして問題にされつつあることを踏まえ、いわば分析的に性能を解明する必要があるためであり、要求条件への合否判断はACR評価で行われた。また、CCR法は従来符号化品質の評価に使われた実績がないため、その分析精度を知るとい趣旨も込められている。CCR法では、一方の刺激は常に対象とする劣化要因を迂回した系（レファレンス）であり、レファレンスに対する相対判断がスコアとして得られる。表3に示す指示が与えられ、1対の刺激が対称に扱われ、かつ、被験者はいずれがレファレンスか知りえないため、先入観による偏りが避けられ、また、符号化の前処理、後処理等で背景雑音時の性能に改善感があれば結果に反映される性質がある。ただし、基準に対する劣化の比較の程度が曖昧であり、意味のあるカテゴリたとえば、DCR(Degradation Category Rating)法で使われる「差は検知できるが、劣化として気にならない」などの指示を与えられない難点がある。本実験の結果等も踏まえ、CCR法はITU-TにおいてP.800[8]に符号化音声の評価法として勧告された。

表3 CCR評価表

評点	評価
+3	非常に良い
+2	良い
+1	やや良い
0	ほとんど同じ
-1	やや悪い
-2	悪い
-3	非常に悪い

### 5.3 実験条件

試験音源として、男女各2名の送話する短い文章を用いた。ACR法では、1刺激音はただ1度提示され、CCR法では提示順を変えて2度提示された。被験者は健聴者24名であり、4人ずつ6グループに分け、刺激提示順序の影響を避けるため、グループ毎に異なるランダムな提示順を用いた。入力音声はすべてIRS等化器を通し、処理後の音声の受聴は平坦な特性を有する密閉型受話器を用いた。受聴時は快適な音量（外耳道入り口で74 dB）となるよう受話回路を設定し、低レベルの擾乱の影響を避けるため、約30 dB(A特性)となるHoth特性(付録3参照)の雑音を加えた。

表4 実験1の条件

条件番号	条件
1/2	G.729/G729A
3/4	E(G.729) ->D(G.729A)/E(G.729A) ->D(G.729)
5/6/7	G.729,3 tandem/G.729A,3 tandem/G.729,2 tandem
8	G.729->E(G.729)->D(G.729A)
9	E(G.729)->D(G.729A)->E(G.729A)->D(G.729)
10	E(G.729)->D(G.729A)->G.729A
11	E(G.729A)->D(G.729)->G.729
12	G.729A->E(G.729)->D(G.729A)
13	G.729A->E(G.729A)->D(G.729)
14	G.729A,2 tandem
15/16	G.729A,low input(-10 dB)/high input(+10 dB)
17/18	G.726/G726,4 tandem
19/20	G.726,low input(-10 dB)/high input(+10 dB)
21/.../28	MINRU,SNR=5/10/15/20/25/30/35/50 dB

### 5.3 評価結果

#### 5.3.1 実験1

実験1の条件を表4に示し、評価結果を図7に示す。表中、E(G.729) -> D(G.729A)は、G.729

図7 実験1の評価結果 (MOS)

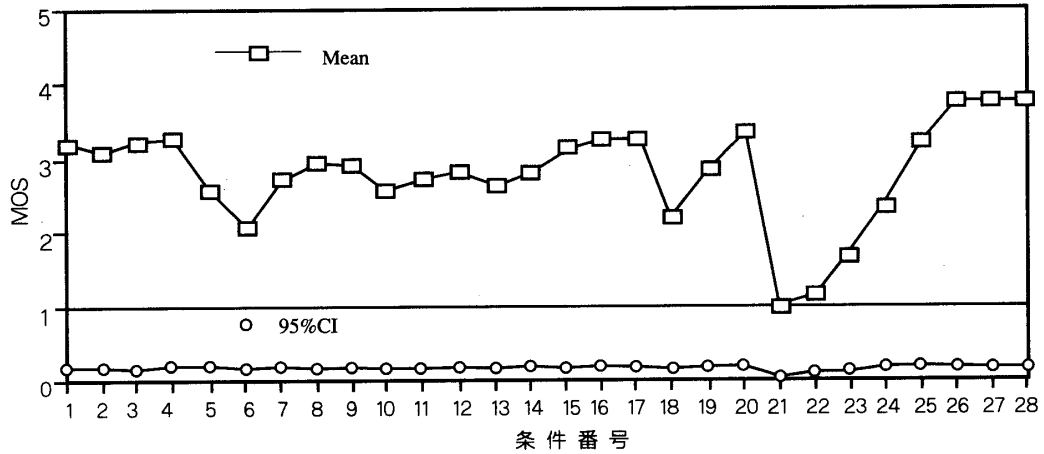


表5 実験2の条件

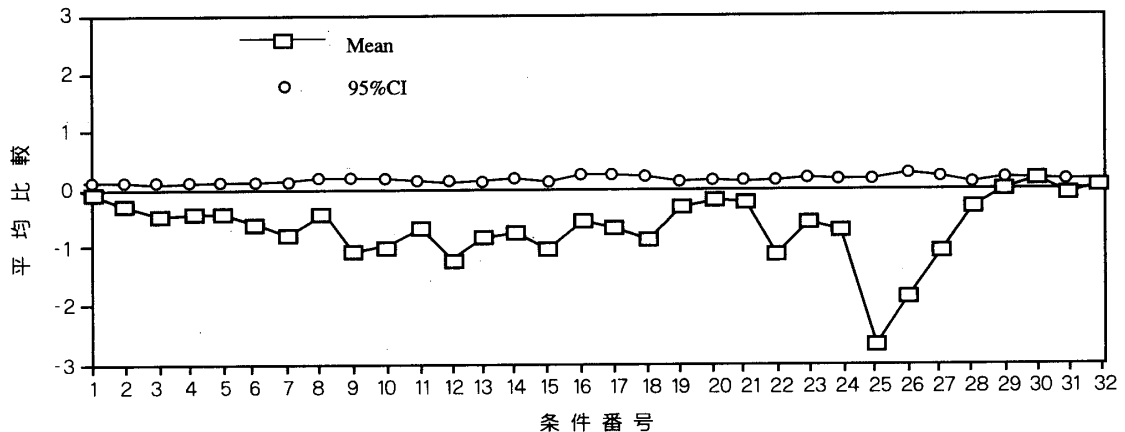
の符号器出力をG.729Aの復号器入力とする条件を表す。条件1, 2, 3, 4のスコアが示すように本体, 候補案とその相互接続に有意な差は無かった。しかし、3段のタンデム接続では候補案が有意に劣る。条件7,...,14の様々な組み合わせの2段タンデム接続のスコアは有意な差がない。候補案の高・低入力レベルのスコアも要求条件を満たしている。

番号	CODEC	背景 SNR= /20/20 dB
1/2/3	G.729	Clean/Office/Music
4/5/6	G.729A	C/O/M
7/8/9	G.729, 2 tandem	C/O/M
10/11/12	G.729A, 2 tandem	C/O/M
13/14/15	E(G.729) -> G.729A -> D(G.729)	C/O/M
16/17/18	E(G.729A) -> G.729 -> D(G.729A)	C/O/M
19/20/21	G.726	C/O/M
22/23/24	G.726, 4 tandem	C/O/M
25/26/27 28/29	MINRU SNR= 6/12/18/24/30 dB	Clean
30/31/32	Direct	C/O/M

#### 5.3.2 実験2

実験2の条件を表5に示し、評価結果を図8に示す。背景雑音なし、オフィス雑音、および、背景音楽のいずれの場合も候補案はG.729本体よりも一貫してスコアが低い。しかし、劣化の程度はいずれの条件でもほぼ、(-1: やや悪い)の範囲に入っている。G.726は波形符号化の特徴を反映し、背景雑音や背景音楽に強く、粒状雑音が背景によりマスクされる

図8 実験2 (背景雑音) の評価結果 (MCCS)





ことから、むしろスコアが高まる傾向がある。試験手法としては、CCR試験はレファレンスとの微小な差を分離出来る精度があることが明らかになった。

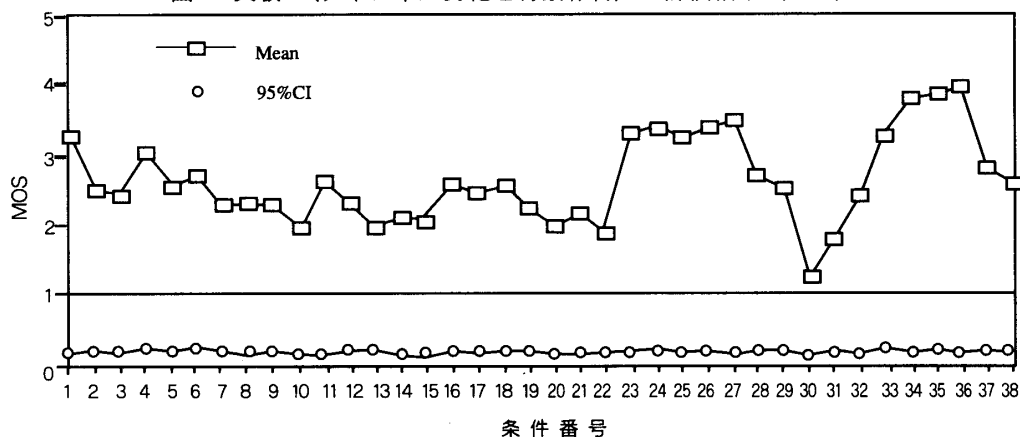
同種の実験が仏語を用いて行われ、その試験結果[13]は、日本語と同一の傾向を示したが、候補案の品質の低下がよりきわだったものとなった。これは、仏語の試験では入力音声に平坦な周波数特性が用いられたため、誤差評価関数の適応化が省略された影響が出たためと推定される。

### 5.3.3 実験3

実験3の条件を表6に示し、評価結果を図9に示す。背景雑音のない条件下でランダム符号誤りとフレーム消失のある場合、および、バブル雑音、オフィス雑音が背景にあるとき、フレーム消失が起きた場合のG.729本体と Annex Aの各接続条件でACR試験が行われた。様々な条件下の各接続条件による有意な差は見つからず、候補案はビット誤り、フレーム消失に対しても問題がないことが明らかとなった。

同種の実験が伊語を用いても行われ、その試験結果[13]は日本語と同様の傾向を示した。

図9 実験3 (チャンネル劣化と背景雑音) の評価結果 (MOS)



## 6.むすび

本論文は、ITU-T勧告G.729 CS-ACELPの演算量削減版がITU-T勧告V.70 DSVD用の音声符号化方式として採用され、勧告G.729 Annex Aとなった経緯と品質試験の結果を述べた。G.729 Annex Aは、G.729本体に比べわずかに品質劣化が認められるものの、10 MIPS以下の演算量で符号・復号の全2重処理が可能であり、また、要求される品質条件も満足していることが明らかとなった。V.70はモデムを対象としているが、DSVDはより一般的にLAN等を介したコンピュータテレフォニに広く使われるシステムであり、後のインターネット上のVoIPに発展する端緒となった。G.729をはじめとするITU-Tの標準符号化方式はその後、有線公衆網に限らず、無線電話網、LAN上のパケット電話、インターネット上のVoIP等に広く利用されることとなった。しかし、ITU-T標準にも複数の方式があり、また、各種地域標準等も並立してい

表6 実験3の条件

番号	CODEC	背景雑音	フレーム消失率 (%)
1/2/3	G.729A	Clean	0/3/5
4/9/14	E(G.729A) -> D(G.729)	Clean/Babble/Office	3
5/10/15	E(G.729) -> D(G.729A)	C/B/O	5
6/7/8	G.729A	Babble	0/3/5
11/12/13	G.729	Babble	0/3/5
16	G.729A	Office	0
17	E(G.729) -> D(G.729A)	Office	0
18/19/20	G.729	Office	0/3/5
21/22	G.729A	Office	3/5
23/24	G.729/G.729A	Clean	Error = 10 <sup>-3</sup>
25/26	E(G.729) / D(G.729A) / E(G.729A) -> D(G.729)	Clean	Error = 10 <sup>-3</sup>
27/28/29	G.726	C/B/O	0
30/31/32	MNRU	Clean	-
33/34/35	10/15/20/25/30/50dB		
36/37/38	Direct	C/B/O	-

る状況から、あらゆる網、サービス間で変換なしに音声通信が行われ品質や遅延時間等の面で利点を享受するまでには至らず、音声符号化技術以外の多方面で依然解決すべき問題が多い。

文献

- [ 1 ] R. Salami et al., "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder", IEEE Trans. Speech and Audio Processing, vol. 6, no. 2, pp. 116-130, 1998.
- [ 2 ] ITU-T Recommendation G.729, Coding of speech at 8 kbit/s using Conjugate Structure-Algebraic Code Excited Linear Prediction (CS-ACELP).
- [ 3 ] ITU-T Recommendation V.70 " Procedures for the simultaneous transmission of data and digitally encoded voice signals over the GSTN, or over 2-wire leased point-to-point telephone type circuits," 1996
- [ 4 ] ITU-T Recommendation G.729 Annex A, "A Reduced complexity 8 kbit/s CS-ACELP speech codec," Nov. 1996
- [ 5 ] ITU-T Report COM 15-R31-E, Annex IX "Provisional Terms of Reference for 13-kbit/s G.DSVD Voice Codec," March 1995
- [ 6 ] ITU-T Recommendation G.728 "Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction," 1992
- [ 7 ] ITU-T Recommendation G.723.1 "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s," 1996
- [ 8 ] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality," 1996
- [ 9 ] ITU-T Recommendation G.729 Annex B, "A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70," Nov. 1996
- [10] ITU-T Study Group 12 Speech Quality Experts Group, "Subjective Test Plan for Characterization of an 8 kbit/s Speech Codec for DSVD Applications," SQ-35.96, March 1996
- [11] ITU-T Recommendation P.48, "Specification for Intermediate Reference System," 1976
- [12] ITU-T Recommendation G.191, "Software Tools for the Qualification Test of a Codec at 8 kbit/s," May 1992
- [13] ITU-T SG15 TD46-E(WP2/15), "Characterization Phase of Testing G.729A: Global Analysis of Results," May 1996

付録1

ACR試験：電話音声の品質評価に用いられる絶対カテゴリ評価試験であり、オピニオン試験とも呼ばれる。被験者は提示された試験音が次の5つのカテゴリのいずれに相当するか、自己の内在する絶対評価尺度により評価する。

カテゴリ5：非常に良い、カテゴリ4：良い、カテゴリ3：まあ良い、カテゴリ2：悪い、カテゴリ1：非常に悪い  
被験者は、格別評価行為に慣れていない通常の電話利用者から、100人程度をランダムサンプリングして選ぶことが望ましい。評価結果はカテゴリ番号に同じ評点を全被験者について平均した平均オピニオン値(MOS)が用いられる。MOS値はわかりやすい絶対評価値として用いられるが、試験の枠組み（試験で提示される音声の品質の偏りなど）に幾分影響されることが知られている。

付録2

dBov：デジタル音声のレベルを飽和レベルを基準とする相対値（dB）で表現したもの。16ビットで量子化し

た場合、音声信号レベルは-32768 から+32767の整数値で表される。音声データの実効値(r.m.s.)値が32767となる場合、0dBovと称する。通常20 dB程度のピークファクタと6 dBの余地を見込んで-26 dBovを標準レベルとする。

付録3

**Hoth特性**：室内騒音を模擬した雑音特性であり、ほぼ-5dB/octのスペクトルを有する。但し、50Hz以下の周波数は緩やかに遮断されている。