

# AMR 音声コーデックにおけるピッチディレイパラメータへの データハイディング

西村 明<sup>†</sup>

<sup>†</sup> 東京情報大学 総合情報学部 情報文化学科

〒 265-8501 千葉県若葉区御成台 4-1

TEL 043-236-4658, akira@rsch.tuis.ac.jp

あらまし CELP 系音声符号化方式のひとつである, AMR 音声符号化におけるピッチディレイパラメータにデータを埋め込む新しい手法を提案する. 従来法と新しい手法との比較を, 埋め込み情報量と, 情報埋め込み済み符号化音声の復号化した音声の客観的品質劣化度合について PESQ を用いて行った. その結果, 5.9kbps 以上の音声符号化モードにおいて 100~150 bps 程度の埋め込みを行えば, MOS-LQO 値における劣化度合が平均 0.1 程度に抑えられることが分かった. また, 250 bps 程度の情報埋め込みであれば, 劣化度合は平均 0.2 程度であることが分かった. 埋め込み情報量と品質劣化の双方から評価して, 新しい2つの手法が最適な埋め込み手法であることが分かった. それらは, ピッチディレイの値が大きいほど, 埋め込み情報と置き換えるピッチディレイデータの LSB 幅を広くとる手法と, これと急激なピッチ変化点のピッチディレイデータを埋め込み情報と置き換える手法を併用する手法であった.

キーワード ハイブリッド音声符号化, 電話通信, 客観的音声品質評価, ステガノグラフィ

## Data hiding in pitch delay parameter of AMR speech codec

Akira NISHIMURA<sup>†</sup>

<sup>†</sup> Department of Media and Cultural Studies, Faculty of Informatics,

Tokyo University of Information Sciences

4-1, Onaridai, Wakaba-ku, Chiba-city, Chiba 265-8501, Japan

TEL +81-43-236-4658, akira@rsch.tuis.ac.jp

**Abstract** New methods of data hiding into the pitch delay parameter of AMR speech codec, which is one of the CELP based speech codec, are proposed. Comparisons between the new and the previous methods were made in terms of embedding bit rate and objective speech quality obtained by the PESQ algorithm. The results showed that embedding from 100 to 150 bps above the 5.9 kbps mode achieved the mean degradation of MOS-LQO value of below 0.1. Embedding data around 250 bps showed quality degradation of 0.2 in mean MOS-LQO value. Considering embedding bit rate and quality degradation, the new embedding methods, which replace hidden data into adaptive width of LSBs and into the abrupt change in pitch delay, were superior to the previous methods.

**Key words** Hybrid speech codecs, Telephony, Objective speech quality measurement, Steganography

### 1. はじめに

携帯電話や IP 電話では, 音声信号を高度に情報圧縮符号化して伝送し, 復号化する仕組みを用いている. その基本的な圧縮符号化アルゴリズムは, CELP(Code Excited Linear Prediction) [1] とよばれ, 音声信号を周期的な波形が繰り返す有声音部分とそれ以外の部分(子音に代表される無声音部分であったり, 有声音部分に含まれる非周期的な波形部分)に分けて効率良く符号化する仕組みをとっている.

この音声符号化アルゴリズムにおける符号化時 [2]~ [5] ある

いは符号化後のデジタルデータ [6] に, 情報を埋め込むハイディング手法が, 従来いくつか考案されてきた. この音声符号化情報ハイディングの特徴は, 図 1 に示したように, 埋め込み, 伝送, 検出のいずれもデジタル領域で行われる点である. よって検出処理は受信したビットストリームから埋め込みに用いたルールに基づいて埋め込み情報ビットを選び出す処理となる. また, 音声信号の復号器と埋め込みデータの検出器は, 同じビットストリームを受け取って処理することから, 埋め込み後の音声データを復号化した際の音声品質の劣化が少ないことが重要である.

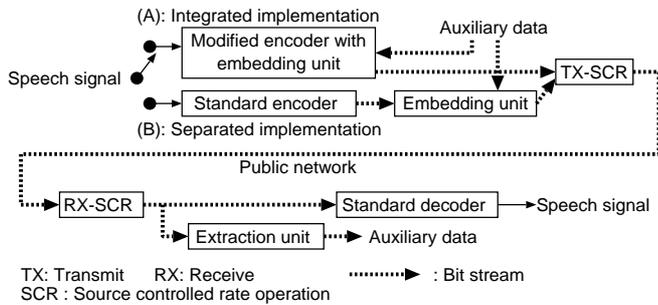


図1 デジタル電話ネットワーク上での音声符号化情報ハイディングの模式図

音声符号化情報ハイディングの用途としては、単に文字情報を埋め込む利用 [3] や、音声の内容に関連した情報 (タイムスタンプ, 要約, メタデータ等) を埋め込むことにより、記録検索や情報サービスに生かしたり、音声以外のメディアデータを埋め込むことにより音声通信サービスに付加価値を持たせるもの [6]、従来の符号化方式と互換性を保ちながら、伝送する音声の周波数帯域を拡張するための情報を埋め込むもの [5]、などが挙げられる。

本稿では、CELP 音声符号化方式の流れをくみ、第三代携帯電話 (いわゆる 3G 携帯) において採用されている AMR (Advanced Multi Rate) 音声符号化方式における情報ハイディング手法として、新たな手法を提案すると共に、従来法とのデータ埋め込み量および客観的指標による音声品質劣化度合の比較を行うことを目的とする。

## 2. CELP 系音声符号化における情報ハイディング

### 2.1 AMR 音声符号化方式

CELP 系音声符号化方式である AMR 符号化方式は、おおまかにいって図 2 のような流れで、音声信号のパラメータを量子化して符号化する仕組みを持つ。符号化音声データのビットレートは、4.75 kbps から 12.2 kbps までの 8 つのモードをもち、回線状況に応じて切替える機能 (Source controlled rate operation) が持たされている。簡略化してその符号化アルゴリズムを説明すると、次のとおりである [7]。

まず、入力される音声信号は、サンプリング周波数 8kHz、直線量子化ビット 13bit であり、これを 20ms 毎のフレームに分割して処理を行う。フレーム化された音声信号は最初にハイパスフィルタに通され、線形予測に基づきスペクトルエンベロープの特徴を表現する LP (Linear Prediction) 係数が求められる。LP 係数は量子化と補間に適した LSP (Line Spectral Pair) 係数に変換され、分割ベクトル量子化により符号化データに変換される。

次に 20 ms のフレームは 5ms ごとの 4 つのサブフレームに分割され、音声パラメータを求める処理が行われる。音声信号における周期的な成分は、適応コードブック検索部において、最初にオープンループでおおまかに、次にクローズドループで細かくピッチ周期 (ピッチディレイ: Pitch Delay) が求められた

表 1 AMR 符号化における各パラメータへのビット割り当て量。

Mode (kbps)	Parameter	subframes			
		1st	2nd	3rd	4th
12.2	2 LSP sets	38			
	Pitch delay	9	6	9	6
	Pitch gain	4	4	4	4
	Algebraic code	35	35	35	35
	Codebook gain	5	5	5	5
10.2	LSP set	26			
	Pitch delay	8	5	8	5
	Algebraic code	31	31	31	31
	Gains	7	7	7	7
7.95	LSP set	27			
	Pitch delay	8	6	8	6
	Pitch gain	4	4	4	4
	Algebraic code	17	17	17	17
	Codebook gain	5	5	5	5

後、その強度 (ピッチゲイン: Pitch Gain) の値が求められる。固定コードブック検索部では、音声信号から周期成分を差し引いた波形を、コードブックで定められる単発パルス位置 (代数コードブック: Algebraic Code) とその極性を複数組み合わせたものにコードブックゲイン (Codebook Gain) を掛けあわせて表現し、残差が最小になるように代数コードブックと極性、コードブックゲインを選択する。

5.9 kbps 以上のモードでは、2 番めと 4 番めのサブフレームのピッチディレイは、1 番めと 3 番めのサブフレームのピッチディレイの整数値からの偏差として量子化される。12.2 kbps と 7.95 kbps モードでは、ピッチゲインとコードブックゲインは対数直線量子化が行われるが、それ以外のモードでは過去のフレームでのデータを元にした移動平均予測値について双方のパラメータ値を統合してベクトル量子化する。

代表的な 3 つのモードでの各パラメータへのビット割り当て量を、表 1 に示した。

### 2.2 従来の情報ハイディング手法

#### 2.2.1 固定コードブックへの埋め込み

岩切と松井 [3] は、CS-ACELP (ITU-T 勧告 G.729) 方式の固定代数コードブックにおける隣接したパルス位置どうしに 0 と 1 を割り当てて、どちらを利用するかによって情報を埋め込む手法を提案している。AMR の 7.95 および 7.4 kbps モードでは CS-ACELP と同じパルス位置候補をもつコードブックを採用しているので、彼らの方法はそのまま適用でき、200 bps の埋め込み量を実現する。しかし他のビットレートモードでは、隣接するパルス位置候補を持たないので、そのままでは適用できない。

Geiser and Vary [5] は AMR の 12.2 kbps モードにおいて、探索する固定代数コードブック候補を  $2^N$  個に分割し、どの範囲を探索するかによってサブフレームあたり  $N$  ビットの情報を埋め込む手法を提案している。この方法は、 $200N$  bps の埋め込み容量を持ち、最適コードブック探索アルゴリズムの改善

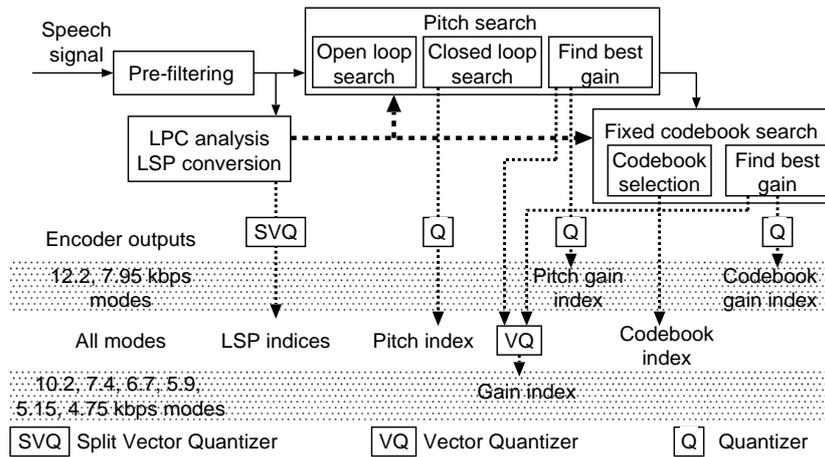


図 2 AMR 音声符号化方式の簡略化した演算過程。

も行った上で、400 bps での埋め込みを行っても、ほぼ埋め込み前と同じ客観的音声品質 (PESQ による MOS-LQO) が得られたと報告している。しかし、12.2 kbps モードでは固定代数コードブックの割り当てビット量はサブフレームあたり 35 ビットと、かなりの余裕と冗長性を持つが、7.95 kbps モード以下では 17 ビット以下であり、この条件で固定コードブックに情報を埋め込めば品質劣化は避けられないと予想される。

### 2.2.2 適応コードブックへの埋め込み

岩切は CELP 系符号化方式である ITU-T 勧告 G.723.1 における、ピッチディレイデータの LSB にデータを埋め込む方法を提案している [4]。これは、クローズドループでのピッチディレイ探索範囲を、埋め込むデータビット値に応じて偶数範囲か奇数範囲かに限定する方法であり、LSB を 1 ビット分置換することにも相当する。符号化アルゴリズムの中に埋め込みアルゴリズムを採り入れる (図 1 中 (a) Integrated implementation 参照) ことによって、標準の符号化アルゴリズムに対する演算処理の低減と、ピッチディレイ算出後のパラメータ値 (ピッチゲイン, 固定コードブック, コードブックゲイン) の探索時に、ピッチディレイへのデータ埋め込みに伴って生じた誤差を低減する方向に最適化が行われるので、音質劣化は少ないとしている。AMR 方式に適用する場合、LSB を 1 ビット幅とすると、どのビットレートモードでも 200 bps の埋め込み量を実現する。ここではこの方法を、ピッチ LSB (PLSB) 法とよぶことにする。

大田らは、CELP 系符号化方式において、ピッチゲインの値に閾値を設け、閾値以下のサブフレームではピッチディレイデータの下位ビットを、埋め込み情報ビットと置き換える方法を提案している [6]。ピッチゲイン値が小さいサブフレームは、音声波形に周期成分が少ない (無声音部, あるいは背景雑音が優勢な部分) あるいは振幅自体が小さいといえ、このときにピッチディレイの値が任意であっても、音質劣化への影響は少ない。ここではこの方法を、ゲイン閾値ピッチ LSB (GTPLSB) 法とよぶことにする。

AMR 方式の 12.2 kbps と 7.95 kbps モード以外では、ピッチゲインの値と過去のサブフレームで得られたコードブックゲイン値の移動平均予測で得られるコードブックゲインの値は、

結合されたのちにベクトル量子化されるため、GTPLSB 法では、符号化処理内部に埋め込み処理を組み込むことは実質不可能である。よって、7.95 kbps と 12.2 kbps のモードにおいて、符号化器が出力するビットストリームに対して埋め込み処理を行った (図 1 中 (b) Separated implementation 参照)。

### 2.3 評価方法

前節までに見てきたように、固定コードブックパラメータへの情報埋め込みは、高いビットレートの場合は良いが、AMR 方式のように低ビットレートもサポートする場合には、原理的に無理であったり、音質劣化が大きくなる可能性が高い。よって AMR 方式においては、適応コードブック (ピッチ) パラメータへの埋め込みが有望だと考えられる。

従来のピッチパラメータへのデータ埋め込み済音声品質の客観的評価は、短い時間間隔毎に原音と符号化音声の波形の差分をノイズ成分と考える Segmental SNR 値で評価されてきた [4], [6]。Segmental SNR 値は、大きいほど符号化音声波形が原音波形に近いことを意味するが、CELP 系音声符号化方式では、知覚的な劣化を最小化する処理が行われるので、波形上の相違は大きく、SNR 10dB 前後になるのが普通である。つまり標準符号化音声と埋め込み済符号化音声の SNR 値が同程度であったとしても、それらの波形上の違いは大きい場合もありえるため、知覚的な劣化が同程度であるとは言えない。

本稿では、AMR 音声符号化におけるピッチディレイパラメータのみを用いて、データ埋め込みを行う新しい方法を提案する。また、人間のピッチ知覚特性を利用して、ピッチディレイ値への LSB 埋め込み方式を拡張する。これらを組み合わせた新しい情報埋め込み方式と、従来のピッチディレイパラメータへの埋め込み方式を、情報埋め込み量の点と、知覚的な音質劣化度を客観的に予測する ITU-T 勧告 P.862.1 (PESQ) を用いて比較する。

## 3. 新しい埋め込み方式

### 3.1 ピッチ変化点埋め込み法

ピッチディレイパラメータ値は、音声波形の基本周期をあらわす。音声の有声部 (声帯振動が音源となる部分) では、音声

波形は周期をもち、声帯振動の生理的な制約により、その周期は時間的にゆるかかつ連続的に変化する。よって、急激にピッチディレイパラメータ値が変化する部分は、有声部ではなく、ピッチディレイパラメータ値が音声復号化時に寄与する割合は低いと言え、ここにデータを埋め込んでも品質劣化は聞き取られ難い。この新しい埋め込み方法を、ピッチ変化点埋め込み法 (Pitch Change Point embedding: PCP 法) とよぶ。

具体的な方法は、連続した奇数サブフレーム (10 ms 間隔) のピッチディレイ値の変化を調べ、周波数上で  $\pm 10\%$  以上の変化がある場合は連続でないとみなし、偶数サブフレームのピッチディレイデータを埋め込み情報ビット値で置換する。しかし、ピッチ検出誤りにより、オクターブ下あるいはオクターブ上のピッチが連続して見られる場合もある。よって、オクターブ上下の周波数の  $\pm 10\%$  以内であれば、ピッチが連続であるとみなし埋め込みは行わない。

### 3.2 適応ピッチ LSB 埋め込み法

PLSB 法では、ピッチディレイ値によらず LSB 置換を行うが、ピッチディレイの値が大きいほど、この LSB 置換の幅を広げる新しい手法を、適応ピッチ LSB 埋め込み法 (APLSB 法) とした。人間の音色や高さの感覚は、対数周波数軸にほぼ対応しているため、ピッチディレイ値へのデータ埋め込みに伴って生じる誤差の音質への影響は、長いピッチディレイの方が、短いピッチディレイよりも少ないことを利用した手法といえる。

AMR 方式では、この人間の聴覚特性に合ったピッチディレイの量子化を行うために、奇数サブフレームでは、12.2 kbps モードの場合ピッチディレイの値が 462 以下 (84.7 Hz 以上) では  $1/6$  サンプルの精度で基本周期を量子化し、それ以外のモードの場合、196 以下 (94.1 Hz 以上) では  $1/3$  サンプルの精度で量子化している。ここでは、12.2 kbps モードの場合、ピッチディレイ値が 256 以上の時に  $w$  ビット幅で、それ未満では  $(w-1)$  ビット幅の LSB 埋め込みを行うことにした。12.2 kbps モード以外の場合、128 を LSB 幅の境界とした。

さらに AMR 符号化では、偶数サブフレームではピッチディレイ値によらず、12.2 kbps モードで  $1/6$  サンプル、それ以外のモードで  $1/3$  サンプルの精度で量子化を行うため、より冗長度が高い。よって、奇数サブフレームのピッチディレイ値が 12.2 kbps モードの場合 384 以上のときに偶数サブフレームでは  $(w+1)$  ビット幅で、それ以外のモードのときには 192 以上のときに  $(w+1)$  ビット幅で埋め込みを行う。

なお、5.15 kbps と 4.75 kbps のモードでは 2-4 番めのサブフレームのピッチディレイ値は、ひとつ前のサブフレームのピッチディレイの整数値からの偏差として表現されるため、ここで提案する PCP 法、APLSB 法いずれもそのままでは適用できない。よって次節以降ではこれらのモードは扱わないが、埋め込み処理の若干の修正によりこれらのモードに適用することは可能である。

さらに、奇数サブフレームに APLSB 法を適用し、埋め込み後のピッチディレイ値を元に偶数サブフレームでは PCP 法を適用した上で、PCP 法が適応されなかった偶数サブフレームには APLSB 法を適用する手法を、MIX 法とした。

## 4. 評価実験

### 4.1 客観的符号化音声品質劣化評価法: PESQ

PESQ (Perceptual evaluation of speech quality) は、ITU-T 勧告 P.862 で示され、コンピュータプログラムとして ITU-T より提供されている、符号化音声に対する客観的音質劣化評価アルゴリズムである。PESQ の入力は、劣化のない原音声信号と、符号化を経た劣化音声信号である。最初に双方の時間ずれと平均的な強度ずれを合わせたあと、FFT による周波数分析を短時間フレーム毎に行い、知覚的な周波数 (Bark) 軸と知覚的な強度尺度 (Sone) に基づいた差分を計算して、音声に対する妨害成分を算出する。この妨害成分は平均の変動の値と平均の非対称な変動の値として計算され、これらの線型結合で PESQ の出力である PESQ score が得られる。PESQ score は、被験者を用いた絶対カテゴリ評価 (ACR: Absolute Category Rating) 実験で得られる音声品質の MOS (Mean Opinion Score) 値と対応づけられることが意図されている。

その後の ITU-T 勧告 P.862.1 では、より多くの音声サンプルと主観評価実験の結果を元に、主観的な MOS 値 (MOS-LQS: Listening Quality Subjective) と、より相関が高くなるよう、PESQ score を変換する関数が導入された。この変換後の MOS-LQO (Listening Quality Objective) 値は 1.02-4.56 の値をとり、1-5 はそれぞれ主観的な MOS での「非常に悪い」、「悪い」、「まあよい」、「よい」、「非常によい」に対応する。

### 4.2 シミュレーション実験

日本音響学会研究用連続音声データベース Vol.1 より、22 名の話者 (男 10 名、女 12 名) による 1200 の音素バランス文を 2 つづつ繋げて作成した 600 文の 6 秒から 12 秒程度の音声信号を用いた。サンプリング周波数は 8kHz にあらかじめ変換し、量子化ビット数は 16bit とした。AMR 符号化時の入力音声の平均レベルは、主観的符号化音声品質試験時の標準 (ITU-T 勧告 P.830) で定められた 3 条件、 $-26$  dBov (0 dBov はデジタル音声振幅クリッピング限界のレベル) と、 $-14$  dBov、 $-38$  dBov で行った。

埋め込みと AMR 符号化には、3GPP TS 26.073 [8] に付属している ANSI-C コードにデータ埋め込み用コードを付加してコンパイルしたものを用いた。PLSB 法、PCP 法、APLSB 法、MIX 法は、5.9、7.95、12.2 kbps の 3 つのモードについて、PGTSLB 法は、ピッチゲインパラメータが独立して得られる 7.95、12.2 kbps モードについて、データ埋め込み量の測定と、PESQ による客観的音質評価を行った。

各埋め込み方式で用いたパラメータ値を図 2 に示した。本実験での AMR 符号化においては、音声区間検出 (VAD) と断続伝送 (DTX) オプション [9] を有効にしたため、音声の休止区間となるフレームにはデータが埋め込まれない。また、音声信号の周波数 (APLSB 法) やその時間変動 (GTPLSB 法、PCP 法、MIX 法) にデータ埋め込み量は依存する。よって、埋め込み法の評価は、データ埋め込み量を横軸に、埋め込みを行った符号化音声の MOS-LQO 値から、埋め込みを行わない標準 AMR 符号化音声の MOS-LQO 値を差し引いた値を縦軸に示

表 2 埋め込みパラメータ

Method	Parameters			Embedding
	LSB	Subframes	Threshold	bit rate [bps]
PLSB	1	1,2,3,4	—	max. 200
	1	1,2	—	
	2	3,4	—	max. 300
	2	1,2,3,4	—	max. 400
GTPLSB	3	1,2,3,4	4	90 — 220
	4	1,2,3,4	5	150 — 340
APLSB	2	1,2,3,4	—	180 — 360
PCP	6	2,4	—	70 — 220
MIX(APLSB)	2	1,2,3,4	—	
(PCP)	6	2,4	—	240 — 480

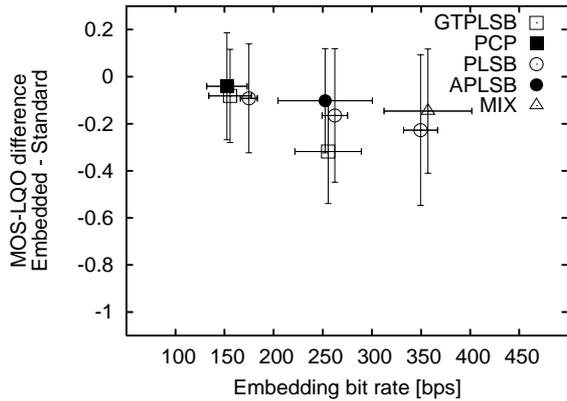


図 3 12.2 kbps モードでの埋め込み量と音声品質劣化度合. 入力レベルは -26 dBov.

した. MOS-LQO 値の差分は, 負の値がデータ埋め込みに伴う劣化の度合を示すことになる.

#### 4.3 実験結果

図 3 に, 12.2 kbps モードでの平均の結果を, 図 4 に, 7.95 kbps モードでの結果を, 図 5 に, 5.9 kbps モードでの結果を示した. 入力レベルはすべて -26 dBov の場合である. それぞれの点は, 埋め込み方式を示し, 縦の誤差棒は MOS-LQO 値の  $\pm 1$  標準偏差を, 横の誤差棒は埋め込みデータ量のビットレート値の  $\pm 1$  標準偏差を示している.

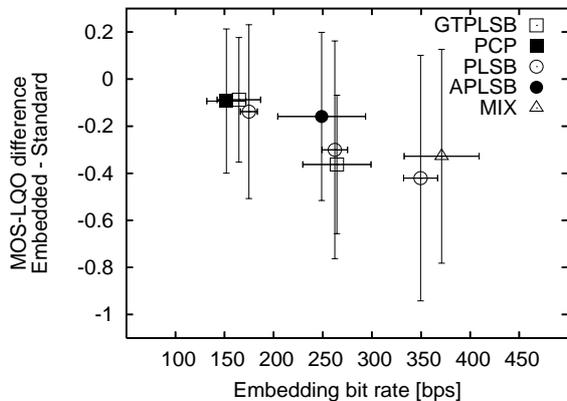


図 4 7.95 kbps モードでの埋め込み量と音声品質劣化度合. 入力レベルは -26 dBov.

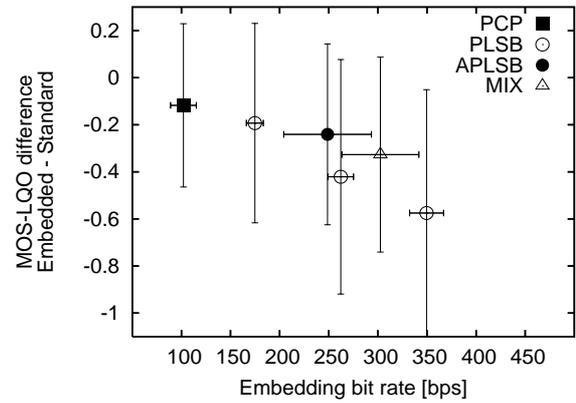


図 5 5.9 kbps モードでの埋め込み量と音声品質劣化度合. 入力レベルは -26 dBov.

平均埋め込みビットレートが 150 bps 程度では, PCP 法, GTPLSB 法, PLSB 法の三者は, 音質においてほぼ同等である. PCP 法は劣化が最も少ないが, 偶数サブフレームにのみ埋め込みを行うため, そのビット長が 4 bit となる 5.9 kbps モードでは埋め込み量が 100 bps 程度となっている. 平均埋め込みビットレートが 250 bps 程度では, すべてのモードを通じて APLSB 法が最も音質劣化が少ない. 350 bps 付近では, MIX 法の音質劣化が少ないが, 5.9 kbps モードでは PCP 法による埋め込み量が少なくなるため, 300 bps 程度に埋め込み量が下がる.

音声符号データビットレートが小さくなるに従い, 同じデータ量を埋め込む際の音質劣化度合は増えていくことが分かった. これは, 音声符号データビットレートが小さいときは冗長度が小さいため, 同じ埋め込み量でも音質を劣化させやすいことを意味している.

音声入力レベルが -14 dBov の場合の 12.2 kbps モードの結果を図 6 に示した. -38 dBov の場合の 12.2 kbps モードの結果は, 図 7 に示した. 入力レベルが -14 dBov での結果は, -26 dBov と比べて 10%前後埋め込みビットレートが向上した. また, -38 dBov での結果は, 埋め込みビットレートが 10%前後下がった. これは音声レベルが大きいと VAD 機能により無音声区間と判定されるフレームが少なくなり, 逆に音声レベルが小さいと無音声区間と判定されるフレームが多くなりデータが埋め込まれなかったためである. このような傾向は, 他のモードでも同様に見られた.

音声品質の劣化度合は, -38 dBov の場合には埋め込みビットレートが低い分, 劣化がやや小さくなる結果が得られた. -14 dBov の場合には, 埋め込みビットレートは増えているが, 品質劣化もわずかに小さくなる結果が得られた. これは音声波形の振幅が大きくクリップするような箇所では, 埋め込み量が多くても知覚的な音質劣化が少なくなるためと考えられる.

## 5. 考察

### 5.1 埋め込みおよび検出演算負荷

今回テストした埋め込み手法のうち, GTPLSB 法以外は, サブフレーム毎のビット置き換え処理となるので, 演算負荷は非

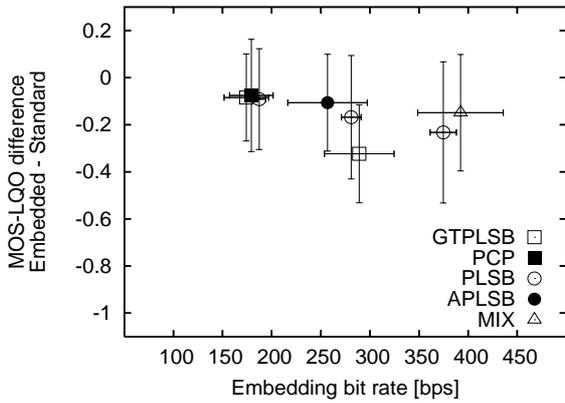


図6 12.2 kbps モードでの埋め込み量と音声品質劣化度合. 入力レベルは -14 dBov.

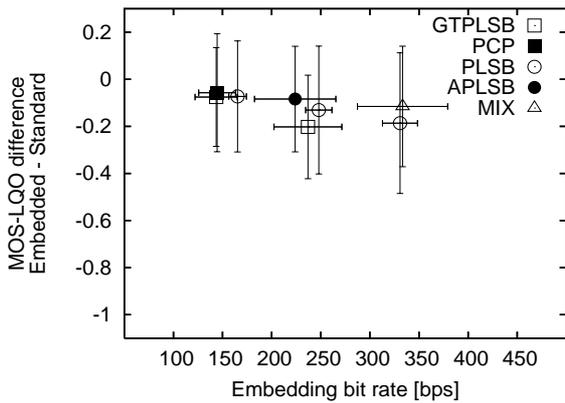


図7 12.2 kbps モードでの埋め込み量と音声品質劣化度合. 入力レベルは -38 dBov.

常に小さい。検出処理については全ての手法で IF THEN ルールに従い簡単に埋め込みデータを検出できる。

GTPLSB 法の埋め込み処理の場合は、7.95 kbps と 12.2 kbps モード以外では、符号化器のビットストリームに対して、復号化器と同じ処理を適用してピッチゲインの値を取り出し、閾値判定を行ってからビットストリーム中のピッチディレイの値を変更する必要がある。これは、既に記録されている音声符号化ファイルへの埋め込み処理には有効であるが、音声通話中のリアルタイムな埋め込み処理においては、音声にさらなる遅延が生じることと演算負荷の面で、不利となる。

### 5.2 実際の利用環境での性能および要求

実際の音声通信環境では、音声と背景雑音の混合信号が符号化される。非定期的な雑音が不可されると、非音声区間と判定されるフレームが少なくなり、いずれの埋め込み法においても埋め込み量が増大する。非公式なシミュレーション実験では 10~20% 程度の埋め込み量の増大が見られた。

いずれの埋め込み法も、埋め込みデータ量は音声信号の活性度に多少なりとも依存するため、適切なデータ区間同期法を導入する必要がある。また、実際の通信環境では伝送路上でのエラーによるフレームデータの欠如やビットエラーが起こりうる。これらに対応する適切なエラー訂正法の導入も必要である。

## 6. ま と め

CELP 系音声符号化方式のひとつであり、第三代携帯電話に採用されている AMR 音声符号化方式にデータを埋め込む手法について検討した。AMR 方式におけるピッチディレイパラメータにデータを埋め込む手法として、従来法と新たに考案した手法をとりあげ、埋め込み情報量と、情報埋め込み済み符号化音声を復号化した音声の客観的品質劣化度合について比較を行った。その結果、5.9 kbps 以上の音声符号化モードにおいて 100~150 bps 程度の埋め込みを行えば、主観的な品質尺度に対応する客観的尺度である MOS-LQO における劣化度合が平均 0.1 程度に抑えられることが分かった。また、250 bps 程度の情報埋め込みであれば、劣化度合は平均 0.2 程度以内であることが分かった。埋め込み情報量と品質劣化の双方を勘案すると、新規に考案した 2 つの手法が最適な埋め込み手法であることが分かった。それらは、ピッチディレイの値が大きいほど、埋め込み情報と置き換えるピッチディレイデータの LSB 幅を広くとる手法と、これと急激なピッチ変化点のピッチディレイデータを埋め込み情報と置き換える手法を併用する手法であった。符号化器への音声信号の入力レベルは  $\pm 12$  dB の範囲では、いずれの手法についても、埋め込み量が  $\pm 10\%$  前後増減し、音質劣化はわずかに小さくなることが分かった。

謝辞

本研究の一部には平成 20 年度東京情報大学共同研究 2、および科研費 (20560365) による補助を受けた。

### 文 献

- [1] M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): High quality speech at very low bit rates," in *Proc. ICASSP'85*, 937-940, (1985).
- [2] 岩切宗利, 松井甲子雄, "音声符号への電子透かしに関する一検討," 電子情報通信学会 1997 年基礎境界ソサエティ大会, SA-7-5, 250-251 (1997).
- [3] 岩切宗利, 松井甲子雄, "共役構造代数符号励振線形予測による音声符号へのテキスト情報の埋め込み," 情報処理学会論文誌, 39, No. 9, 2623-2630 (1998).
- [4] 岩切宗利, "ITU-T 勧告 G.723.1 による音声符号化方式を用いたステガノグラフィ," 2002 年暗号と情報セキュリティシンポジウム, 289-294 (2002).
- [5] B. Geiser and P. Vary, "Backwards Compatible Wideband Telephony in Mobile Networks: CELP Watermarking and Bandwidth Extension," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. IV, 533-536, (2007).
- [6] 大田恭士, 鈴木政直, 土永義照, 田中正清, 佐々木繁, "音声符号に対するデータ埋め込み/抽出方法および装置並びにシステム," 特許公開 2003-295879 (2003).
- [7] 3rd Generation Partnership Project, "Mandatory Speech Codec speech processing functions AMR speech codec; Transcoding functions," 26.090, (1999).
- [8] 3rd Generation Partnership Project, "ANSI-C code for the Adaptive Multi Rate speech codec," 26.073, (2001).
- [9] 3rd Generation Partnership Project, "Mandatory Speech Codec speech processing functions AMR Speech Codec; Source Controlled Rate operation," 26.093, (2000).