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込み

岩切宗利† 松井甲子雄†

この論文では，デジタル音声の符号化方式であるITU-T勧告G.729 CS-ACELPによる音声符号の一部を利用してテキスト情報を密かに埋め込み伝送する一方法を提案する．その原理は，音声の符号化の際に用いられるマルチパルス音源コードブックの合成法に着目して，テキストから生成したランダムビット系列でパルス位置を制御しながら密かにテキスト情報を埋め込み伝送するものである．この方法によるとテキスト情報が埋め込まれた伝送符号をそのまま音声として再生しても聴感的にまったく違和感を与えないで済ますことができる．この論文では，200ビット/秒以下のテキスト情報を第三者に知られることなく埋め込めることを実験的に示す．この方法は音声ソフト等の著作権保護のための電子透かしの埋込みや通話者を認証するための特殊信号の秘密伝送にも応用できる．

Embedding a Text into Conjugate Structure Algebraic Code Excited Linear Prediction Audio Codes

MUNETOSHI IWAKIRI† and KINEO MATSUI†

In this paper we propose how to embed a text secretly into the conjugate structure algebraic code excited linear prediction audio codes. When 5 ms-speech sample is coded, four pulse positions of the sample are selected from the default parameters of the multipulse codebook in the G.729 specification. We try to control this selection according to a bit of the text. This scheme brings us a secret structure to embed our text data. The amount of the text data is estimated up to two hundred bits per second in 8 kbit/s CS-ACELP system.

1. はじめに

通話相手に密かに情報を伝達するための手段として，暗号化技術が広く用いられている．これは非常に有効な手段であるが，重要な情報の存在そのものを隠すことはできず，不正に情報を得ようと目論む第三者による解読の対象となりやすい．そこで，デジタル通信網を介して伝送される音声符号に秘匿情報を密かに埋め込むことで，秘密情報の存在自体を秘匿することを考える．

このような観点から，アナログ通信において，すでに2，3の研究がなされている．Steelら[1]は，音声スクランブラの特徴を巧みに利用して文字情報を埋め込み，Wongら[2]は，音声の位相を切り替えることによりビット情報を埋め込む方法を提案している．ほかにもデジタル化の際に生じる量子化雑音に見せかけて文書データを埋め込む方法が，松井ら[3]により提案されている．また，これに類する技術として電子透かしと呼ばれるマルチメディアの著作権を明確にするための技術が近年脚光を浴びている[4),5]．これは，画像や音声データに特殊な情報（著作権情報）を密かに埋め込むという点では同じ原理によるものと見なせる．

しかし，これらの手法は無圧縮状態のデジタル音声データを対象としたものがほとんどであり，音声通信において施される大幅なデータ圧縮によって，埋め込まれた情報のほとんどが消失すると考えられる．これに対して岩切ら[6]は，国際標準規格における高能率な音声符号化方式G.726への埋込みを試みている．この方法は，符号圧縮の過程において情報を埋め込むものであり，波形符号化方式を用いたデジタル通信系を対象とする場合に有効である．

そのほかの代表的な音声符号化方式に複数のデジタル波形値をフレームとしてまとめ，そのフレームごとに音声符号を生成する符号励振線形予測（CELP: Code Excited Linear Prediction Audio Codes）がある．ITU-T勧告G.729 8 kbit/s CS-ACELP（Conjugate Structure Algebraic CELP）[7]は，このCELPを原

† 防衛大学校情報工学科
Department of Computer Science, National Defense Academy

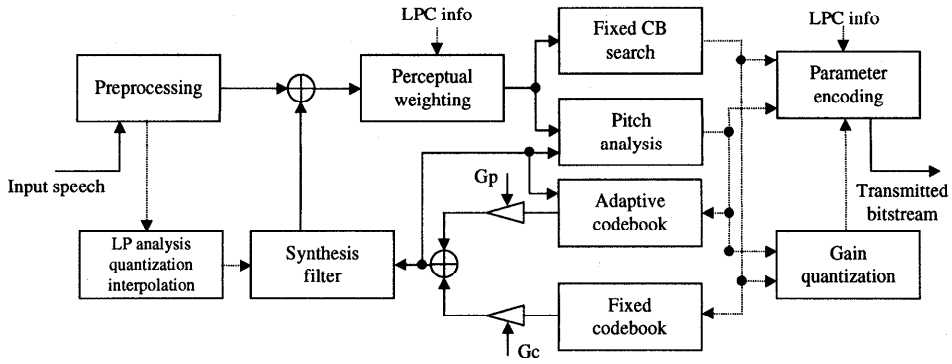


図1 G.729 8bit/s CS-ACELP 符号化ブロック図[7)]
Fig. 1 G.729 8bit/s CS-ACELP encoder block diagram [7)].

理とする符号化法の1つで，大幅に符号量を削減しているにもかかわらず高音質な音声を再生できる．

そこで，これまで検討されていないG.729による音声符号にテキスト情報等を密かに埋め込み伝送する方法について考える．その基本的なアイディアは，デジタル音声データを符号化する際に用いられるマルチパルス音源の構造に着目し，その合成過程においてビット系列化されたテキストデータを埋め込むものである．その際，埋込みを施す音声符号を不特定に選択し，埋込みの規則を変化させることで，埋込みの存在を隠すことができる．

まず，2章ではG.729を要約する．3章にマルチパルス音源を利用してテキストビット系列を音声符号に埋め込む方法を述べ，4章に埋込みの存在を隠す一方法を提案する．5章では，実験システムを構成して行った実験結果を示す．

2. G.729の概要

ITU-T 勧告 G.729 は，1フレームの長さを 10 ms（80サンプル）として，各フレームごとに80ビットの音声符号を生成することで，8 kbit/s のビットレートを実現している．この基本的な符号化処理ブロックを**図1**に示す．

G.729は，符号化に用いる固定コードブックに特徴がある．これは，40サンプルのサブフレームごとに**表1**に示す候補の中から4つのパルス位置 $m_0 \sim m_3$ と極性情報 $s_0 \sim s_3$ を決定するマルチパルス音源型のコードブックである．この探索処理は，次のようになる[7)]．

まず，10次の線形予測フィルタを量子化線形予測係数 $\hat{a}_i$, $i = 1, \cdots, 10$ を用いて，次のように定義する．

$$\frac{1}{\hat{A}(z)} = \frac{1}{1 + \sum_{i=1}^{10} \hat{a}_i z^{-i}} \tag{1}$$

表1 マルチパルスコードブック
Table 1 Multipulse codebook.

Pulse	Sign	Positions	
i_0	s_0: ±1	m_0:	0,5,10,15,20,25,30,35
i_1	s_1: ±1	m_1:	1,6,11,16,21,26,31,36
i_2	s_2: ±1	m_2:	2,7,12,17,22,27,32,37
i_3	s_3: ±1	m_3:	3,8,13,18,23,28,33,38
			4,9,14,19,24,29,34,39

また，聴感的重み付けフィルタ $W(z)$ を線形予測係数 a_i, $i = 1, \cdots, 10$ を用いて，次式のように定義する．ただし，γ_1, γ_2 は，$W(z)$ の特性を決定する重み付け係数である．

$$W(z) = \frac{1 + \sum_{i=1}^{10} \gamma_1 a_i z^{-i}}{1 + \sum_{i=1}^{10} \gamma_2 a_i z^{-i}} \tag{2}$$

ここで，これらの合成フィルタ $W(z)/\hat{A}(z)$ のインパルス応答を $h(n)$, $n = 0, \cdots, 39$ とする．また，表1から導かれる40次のコードベクトル $c(n)$ を次のように表現する．

$$\begin{aligned} c(n) &= s_0\delta(n - m_0) + s_1\delta(n - m_1) \\ &\quad + s_2\delta(n - m_2) + s_3\delta(n - m_3) \\ n &= 0, \cdots, 39 \end{aligned} \tag{3}$$

ただし，

$$\delta(l) = \begin{cases} 1 & l = 0 \\ 0 & l \neq 0 \end{cases} \tag{4}$$

まず，次のフィルタ $P(z)$ を用いて $c(n)$ を処理する．

$$P(z) = 1/(1 - \beta z^{-T}) \tag{5}$$

ここで，T はピッチ遅延である．また，β は前フレームにおける適応コードブックのゲイン $g_p^{(m-1)}$ の量子化値 $\hat{g}_p^{(m-1)}$ により定まる係数である．

$$\beta = \hat{g}_p^{(m-1)}, \quad 0.2 \leq \beta \leq 0.8 \tag{6}$$

ただし，遅延が40に満たない場合は，$h(n)$ を次のように修正する．

$$h(n)=\begin{cases} h(n) & n=0,\cdots,T-1 \\ h(n)+\beta h(n-T) & n=T,\cdots,39 \end{cases} \tag{7}$$

次に，ターゲット信号 $x'(n)$ を次式により求める．

$$x'(n)=x(n)-g_p y(n),\quad n=0,\cdots,39 \tag{8}$$

$x(n)$ は，重み付け音声 $sw(n)$ から $W(z)/\hat{A}(z)$ の0入力応答を減じたものである．また，$y(n)$ は，適応コードブック $v(n)$ と $h(n)$ の畳み込み積分値である．

次に，$x'(n)$ を用いて，次式により $d(n)$ を求める．

$$d(n)=\sum_{i=n}^{39} x'(i)h(i-n),\quad n=0,\cdots,39 \tag{9}$$

また，次式による値を C とする．ここで，m_i は各パルス位置である．

$$C=\sum_{i=0}^{3}|d(m_i)| \tag{10}$$

さらに，次式による値を E とする．

$$E=2\sum_{i=0}^{3}\sum_{j=i}^{3}\phi'(m_i,m_j) \tag{11}$$

ここで，$\phi'(i,j)$ は，次式により求まる．

$$\phi'(i,j)=sign[d(i)]sign[d(j)]\phi(i,j) \tag{12}$$

$$\phi'(i,i)=0.5\phi(i,i) \tag{13}$$

$$i=0,\cdots,39,\quad j=i+1,\cdots,39$$

また，$\phi(i,j)$ は，次式により求まる．

$$\phi(i,j)=\sum_{n=j}^{39}h(n-i)h(n-j) \tag{14}$$

$$i=0,\cdots,39,\quad j=i,\cdots,39$$

コードブックの探索は，C^2/E の値を最大にするように，Abs（合成分析：Analysis by Synthesis）の手法を用いて行われる．よって，その処理量は，一般に膨大となるので，次のように閾値を導入して探索処理量を削減している．

まず，初めの3パルスから得られる $\sum_{i=0}^{2}|d(m_i)|$ の最大値 $\max_3$，平均値 av_3 および係数 $K_3=0.4$ を用いて閾値 thr_3 を求める．

$$thr_3=av_3+K_3(\max_3-av_3) \tag{15}$$

ここで，4番目のパルス m_3 の探索は，thr_3 を超える m_0，m_1，m_2 との組合せについてのみ実施する．また，各フレームごとの最大探索処理量 $TI_{\max}$ を用いて，処理遅延の増大を抑制している．

一方，受信側は，送られてきた符号から各パラメータを抽出して出力音声を合成するため，その処理量は符号化に比べ少なく高速である．

3. 埋込み原理

CS-ACELP符号化方式において，音声符号中にテキスト情報を埋め込む原理を示す．表1において4番目のパルス情報 i_3 のパルス位置 m_3 は，他の $i_0\sim i_2$ のパルス位置の候補 $m_0\sim m_2$ と異なり隣接した候補を持つことが分かる．ここで，**図2** に示すように選択された最適パルス位置 m_3 をそれに隣接する候補 m_3' に置き換えて音声符号としても再生音声に与える影響は少ないと考えられる．これを利用して，音声符号のマルチパルス音源情報部分にテキスト情報等の特殊信号系列を埋め込むことを考える．

まず，m_3 の候補にラベル付けを行う鍵 k_p を導入する．たとえば，**図3** に示すように k_p="00001111" としたならば，k_p の最上位ビットが"0"であるので，m_3 の候補 {3} に"0"を割り当て，それに隣接する候

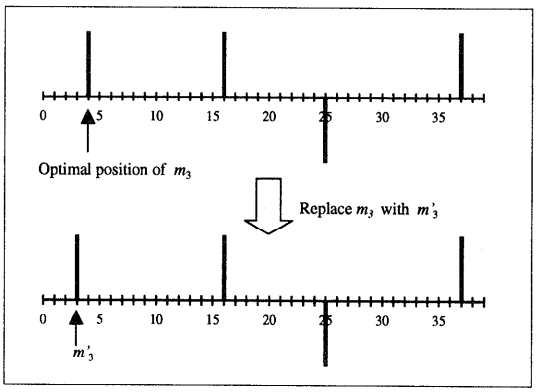


図2 m_3 の置換
Fig. 2 Replacement of position m_3.

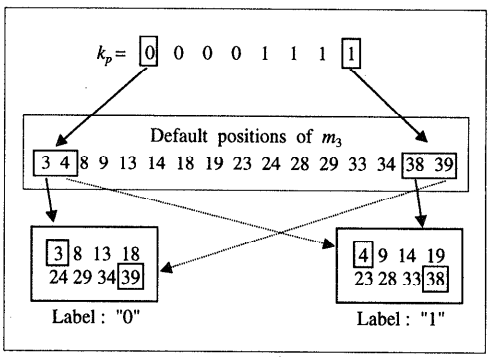


図3 k_p による m_3 の分類
Fig. 3 Grouping of m_3 positions by k_p.

表2 c_p がとりうる値
Table 2 Expression of c_p.

c_p
6, 7, 11, 12, 16, 17, 21, 22, 26, 27,
31, 32, 36, 37, 41, 42, 46, 47, 51, 52,
56, 57, 61, 62, 66, 67, 71, 72, 76, 77,
81, 82, 86, 87, 91, 92, 96, 97,101,102,
106,107,111,112,116,117,121,122,126,
127,131,132,136,137,141,142,146,147

表3 実験音声
Table 3 Sound for experiment.

Data	Language	Type	Samples	Sec
Jm	Japanese	male	40,000	5
Jw	Japanese	female	40,000	5
Em	English	male	40,000	5
Ew	English	female	40,000	5
Ews	English	female	480,000	60

補 {4} に "1" を割り当てる．一方，k_p の最下位ビットは "1" であるので，m_3 の候補 {38} に "1" を割り当て，それに隣接する候補 {39} に "0" を割り当てる．この要領で，m_3 の全候補に "0" と "1" のラベルを付ける．

ここで，音声符号にテキストビット "0" を埋め込む場合は，k_p により "0" のラベルを付けられた候補の中から m_3 を選定する．一方，テキストビット "1" を埋め込む場合は "1" のラベルを付けられた候補の中から m_3 を選定する．これを繰り返すことで2値化したテキスト情報を埋め込むことができる．

また，k_p を知る受信者は，音声符号中に含まれる m_3 のラベルを調べることで，テキスト情報を容易に抽出できる．

4. 秘匿性の向上

3章で述べた方法により，全サブフレームに埋込みを施すと1秒あたり200ビットのテキスト情報を埋め込むことが可能である．しかし，同じ鍵 k_p を用いて全符号にテキストビットを埋め込むと，不正な第三者が k_p を解析する可能性が高くなる．そこで，次の方法により秘匿性の向上を試みた．

まず，$m_0 \sim m_3$ の合計値を c_p とする．

$$c_p = m_0 + m_1 + m_2 + m_3 \qquad (16)$$

この値は**表2**に示した58通りのいずれかになる．

また，音声符号に含まれる $m_0 \sim m_3$ の各値がランダムならば，表1から c_p の出現頻度は正規分布に近い特性を示すと考えられる．たとえば，英語の女声 Ews（**表3**参照）から得られた音声符号に含まれる $m_0 \sim m_3$ を用いて，c_p の出現頻度を調べると**図4**が

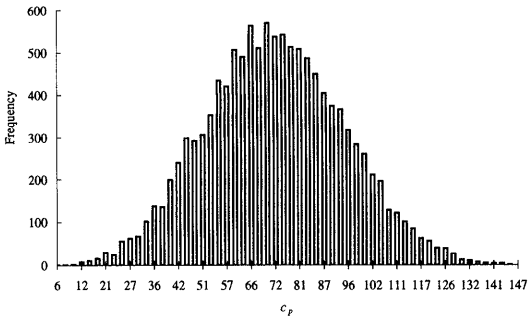


図4 c_p の出現頻度
Fig. 4 Frequency of c_p appearance.

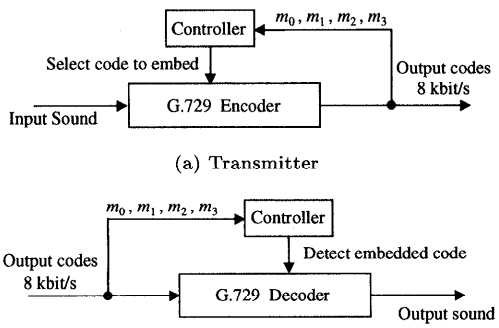


図5 フィードバック制御
Fig. 5 Feedback control.

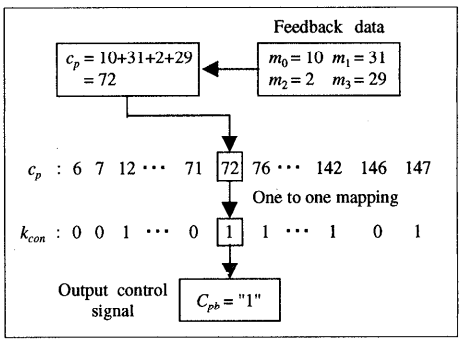


図6 c_{pb} 抽出処理
Fig. 6 Getting a signal c_{pb}.

得られる．この図から各 $m_0 \sim m_3$ は，ほぼランダムに選択されていることが分かる．そこで，**図5**に示すフィードバック処理構造によりテキストビットを分散配置することを試みた．

ここで，鍵 k_{con} を導入する．これは c_p がとりうる58通りの全値に対し "0" と "1" を割り当てるもので，58ビットの2進数である．まず，**図6**に示すようにフィードバックした出力音声符号から c_p を求める．次に，その c_p に対応する k_{con} のビット値 c_{pb} を

抽出する．この c_{pb} が "1" のときは，音声符号への埋込みを実施する．一方，"0" のときは埋込みを実施しない．これを繰り返すことで，テキストビットを音声符号全体に分散配置できる．この方法により，k_{con} を知らない第三者がテキスト情報を含む音声符号を特定するのは難しくなる．

しかし，特殊な k_p（"00000000" や "11111111"）を用いて長期間にわたって埋込みを施すと，埋め込んだテキストデータの統計的特性が音声符号に反映されると考えられる．よって，同じ k_p の長期使用は，埋込みの存在を隠すうえで望ましくない．そこで，鍵 k_p を，短間隔で $\bar{k}_p$ として置き換えながら埋込みを施せば，c_p の出現頻度の偏りを拡散できるが，c_p は 1/2 の確率で偶数になることが表 2 から明らかである．したがって，c_p が偶数になった場合に式 (17) を用いて k_p を $\bar{k}_p$ に置き換えることにする．

$$k_p \leftarrow \bar{k}_p \tag{17}$$

この方法を用いて，不特定な短期間に k_p を換えれば，統計的偏りを拡散できることになる．

5. アルゴリズム

各パルス位置の探索アルゴリズムについて詳細を示す．ここでは，次のように表記する．

thr_3：探索処理閾値，$S_{\max}$：C/E の最大値，$TI_{\max}$：最大探索処理量，$time$：探索処理量，L0, ⋯, L4：ループ処理，$i_0, \cdots, i_4$：パルス位置候補，$m_0, \cdots, m_3$：最適パルス位置，$CE(x_0, x_1, x_2, x_3)$：パルス位置 x_0, x_1, x_2, x_3 を用いて C/E を求める関数，$tbit$：埋め込むビット値，$mode$：埋込み実施フラグ（埋め込む場合：1，埋め込まない場合：0，初期値：0），$tap(x, y)$：x の上位から y ビット目のビット値を抽出する関数，$check(k_{con}, c_p)$：k_{con} の c_p に対応するビット値 c_{pb} を抽出する関数，$get(T)$：埋め込むデータファイル T から 1 ビットずつ抽出する関数，$put(tbit, T)$：抽出ビット値 $tbit$ をデータファイル T に出力する関数．

5.1 符号化手順

(Step 1) thr_3 を計算する．
(Step 2) $S_{\max} \leftarrow 0$
(Step 3) $time \leftarrow 0$
(Step 4) begin L0: $i_0 = 0, 5, \cdots, 35$
(Step 5) begin L1: $i_1 = 1, 6, \cdots, 36$
(Step 6) begin L2: $i_2 = 2, 7, \cdots, 37$
(Step 7) もし，$\sum_{k=0}^{2} d(i_k) > thr_3$ ならば，Step8～15 を実行する．
(Step 8) begin L3: $i_3 = 3, 8, \cdots, 38$
(Step 9) もし，$mode = 0$ または，$mode = 1$ かつ $tap(k_p, (i_3 - 3)/5) = tbit$ ならば，次の処理を実施する．
  (1) $S \leftarrow CE(i_0, i_1, i_2, i_3)$
  (2) もし，$S > S_{\max}$ ならば，次の処理を行う．
    (a) $S_{\max} \leftarrow S$
    (b) $m_0 \leftarrow i_0$; $m_1 \leftarrow i_1$; $m_2 \leftarrow i_2$ $m_3 \leftarrow i_3$
(Step10) end L3
(Step11) begin L4: $i_4 = 4, 9, \cdots, 39$
(Step12) もし，$mode = 0$ または，$mode = 1$ かつ $tap(k_p, (i_4 - 4)/5) \neq tbit$ ならば，次の処理を実施する．
  (1) $S \leftarrow CE(i_0, i_1, i_2, i_4)$
  (2) もし，$S > S_{\max}$ ならば，次の処理を行う．
    (a) $S_{\max} \leftarrow S$
    (b) $m_0 \leftarrow i_0$; $m_1 \leftarrow i_1$; $m_2 \leftarrow i_2$ $m_3 \leftarrow i_4$
(Step13) end L4
(Step14) $time \leftarrow time + 1$
(Step15) もし，$time > TI_{\max}$ ならば，Step 19 へ飛ぶ．
(Step16) end L2
(Step17) end L1
(Step18) end L0
(Step19) $c_p \leftarrow m_0 + m_1 + m_2 + m_3$
(Step20) $mode \leftarrow check(k_{con}, c_p)$
(Step21) もし，$mode = 1$ ならば，$tbit \leftarrow get(T)$
(Step22) もし，c_p が偶数ならば，$k_p \leftarrow \bar{k}_p$
(Step23) m_0, m_1, m_2, m_3 を出力する．

5.2 復号手順

(Step 1) $mode \leftarrow 0$
(Step 2) 音声符号から m_0, m_1, m_2, m_3 を取り出す．
(Step 3) $c_p \leftarrow m_0 + m_1 + m_2 + m_3$
(Step 4) もし，$mode = 0$ ならば，Step8 へ飛ぶ．
(Step 5) もし，$m_3 - 3 \equiv 0 \pmod 5$ ならば，$tbit \leftarrow tap(k_p, (m_3 - 3)/5)$
(Step 6) もし，$m_3 - 4 \equiv 0 \pmod 5$ ならば，$tbit \leftarrow tap(k_p, (m_3 - 4)/5) \oplus 1$
(Step 7) $put(tbit, T)$
(Step 8) $mode \leftarrow check(k_{con}, c_p)$
(Step 9) もし，c_p が偶数ならば，$k_p \leftarrow \bar{k}_p$
(Step10) Step2～9 を繰り返す．

6. 実験結果

6.1 実験システムの概要

第三者に再生音質の異常から埋込みの存在を知られないためには，埋込みによって音質が大きく劣化しないことが重要である．そこで，G.729のアルゴリズムに従ったシミュレータを作成し実験を行った．また，表3に示した実験音声は，FMラジオならびに英会話テープから抽出した日本語と英語の男性と女性の発声音を8 kHz16 bitで量子化したものである．

また，埋込みには，次の鍵 k_p，k_{con} を用いた．これらの鍵の値は，0～Fの16進数で表現している．

$$k_p = 00_{(16)} \tag{18}$$

$$k_{con} = 3FFFFFFFFFFFFF_{(16)} \tag{19}$$

ここで，k_p を式(18)のような特殊な値にすると音声符号から得られる c_p の統計値に大きく影響が現れると考えられる．さらに，k_{con} を式(19)のように設定することで全符号に対して埋込み処理を行い，再生音質の劣化を最大にしている．すなわち本実験では，最悪のケースを扱っていることに注意する．通常このような事態は避けなければならない．また，埋め込むテキスト情報としてインターネット規格のRFC（Request For Comments）に含まれる英文テキストデータを用いた．これは，各文字を8ビットのアスキーコードで表現しており，統計的にはビット“0”を多く含む特徴がある．

6.2 音質の評価法

主観的評価法として，評価者の絶対判断によるオピニオン評価を用いた[8]．これは，複数の評価者に音質を5段階に絶対評価させ，得られた評価値から平均オピニオン値（MOS: Mean Opinion Score）を求めるものである．本実験では，オピニオン評価の基準を，非常に良い：5，良い：4，普通：3，悪い：2，非常に悪い：1とした．また，被験者の先入観による影響を避けるため，各音声ごとに埋込みのないものと埋込みのあるものの2種類を外見上区別できない状態で準備した．また，各音声を任意に参照できるシステムを準備し，自由に聞き比べることで評価させた．これにより，埋込みによる聴感的な音質の違いがあるならば，埋込みのある音声とない音声のMOSに大差が生じると考えられる．

6.3 実験結果と考察

本実験では，式(19)の鍵 k_{con} を用いたため，音声符号への埋込み量は1秒あたり200ビットになる．この埋込みを施した再生音声と埋込みのない再生音声の音質を，20代の健聴者8名により評価した場合のMOSを**表4**に示す．この結果から，埋込みのある場合とない場合のMOSは，それぞれ約3.7程度になっている．これは，テキストデータの埋込みによる聴感的な音質の違いがほとんどないため，被験者が埋込みのある音声を特定できなかったことを示している．よって，再生音質の違いから埋込みが施された音声符号を特定するのは難しいと考えられる．

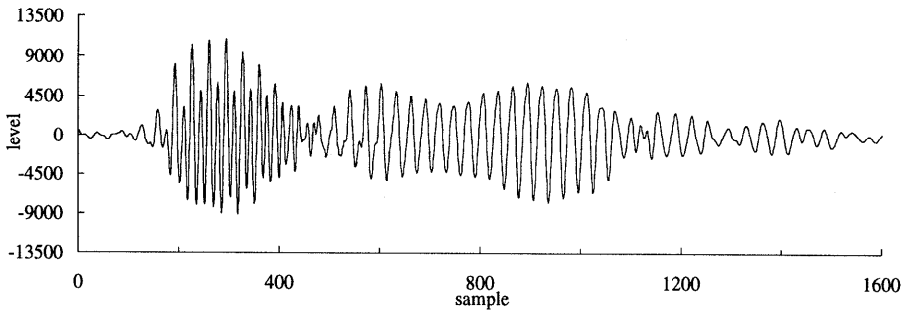
表4 再生音質（MOS）
Table 4 Mean opinion score.

Sound	Sound without text	Sound with text
Jm	3.29	3.43
Jw	4.43	4.14
Em	3.57	3.71
Ew	3.71	3.71
Average (MOS)	3.75	3.75

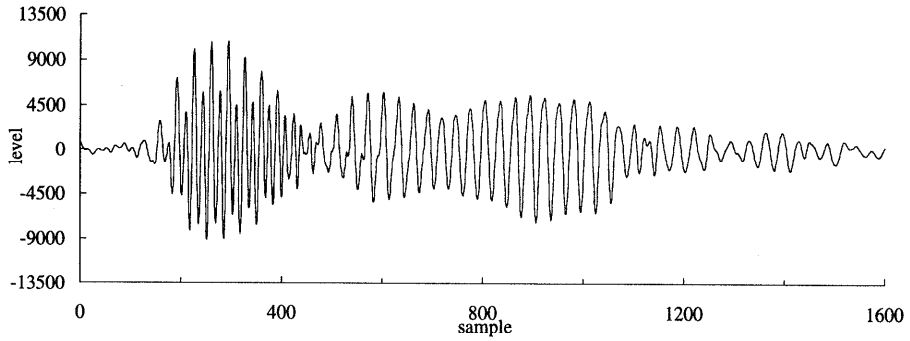
次に，再生波形の一部を切り出して，埋込み処理が波形の形状に与えた影響を観察してみる．**図7**に埋込みのない再生音声波形(a)と埋込みを施した再生音声波形(b)およびそれらの差分波形(c)を示した．これら波形は，Emにおける発音“think”に相当する部分で0.2 sの音声区間である．したがって，波形(b)には，40ビットのテキスト情報が埋め込まれていることになる．

ここで，図7の差分波形(c)から(a)と(b)の波形に違いが生じていることが分かる．これは，コードブックのパルス位置の変更により生じた位相の変化によるものと考えられる．しかし，人間の聴覚は，位相のずれを感じることが一般に苦手である．そのため，波形の形状に不自然な歪みを生じなければ，位相が若干変化しても音声として不自然には感じられない．よって，これらの再生音声を聞き比べても聴感的に大きな違いをほとんど感じないと考えられる．これは，先に示した表4からも推察できる．また，本方式を用いて通常公開する音声符号は，埋込みのあるもののみである．よって，不正な手段で音声符号を傍受されても埋込みのない波形と比較することはできないため，差分波形を得ることはできない．したがって，再生波形の形状から埋込みのある音声符号を特定するのは難しいと考えられる．

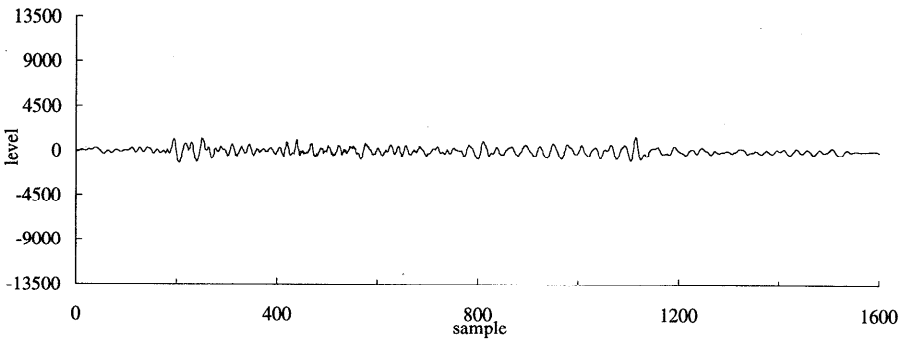
次に，式(18)のような k_p を使用して，大量に埋込みを施すと埋め込むデータの統計的なビット特性が音声符号に反映されると考えられる．そこで，図4と同様にEwsの全音声符号に埋込みを施した場合の c_p の出現頻度を調べると**図8**が得られた．これと図4を比較すると明らかに埋込みの影響を観察できる．したがって，ここで用いたような特殊な k_p を長期にわ



(a) Decoded waveform of "think" without text



(b) Decoded waveform of "think" with text



(c) Difference waveform

図7 音声波形の比較
Fig. 7 Comparison of sound waveform.

たって使用すると埋め込むビット系列の統計的な特徴が反映されてくるので好ましくない．一般的には，テキストの存在を気付かれにくくすることが望ましいので，4章に述べた簡単な手法を用いて k_p のみを変動させ，統計的な偏りを拡散することを試みた．その結果，**図9** が得られた．この図から，k_p の変動処理により統計的な偏りが拡散されていることが分かる．よって，k_p の変動処理は，埋込みにより生じる統計的な偏りを解消する方法として有効であると考えられる．

7. むすび

この論文では，G.729 8 kbit/s CS-ACELP による音声符号にテキスト情報を密かに埋め込む手法を提案し，音声符号にテキスト情報等が埋め込まれた状態でも聴感的な違和感を与えない音声を再生できることを実験的に示した．本手法によれば，不正に情報を得ようとする第三者から秘密情報の存在自体を秘匿し，特定の相手にのみ伝送することが可能である．これは，将来の秘匿通信技術の1つとして有効な手段であると考える．

また，本手法はバックワード型の適応化処理構造を持つため符号誤りにより受ける影響は大きい．その対策として，埋め込むデータに誤り訂正機能を付加すれば，ある程度のランダム符号誤りには対処できる．し

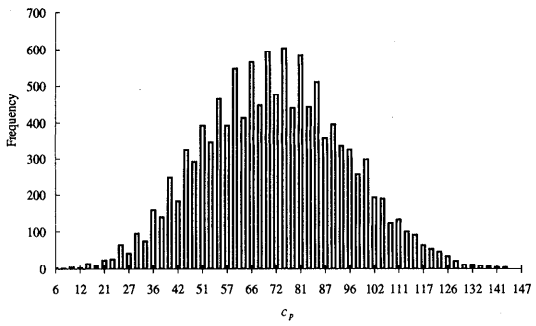


図8 c_p の出現頻度（k_p：固定）
Fig. 8 Frequency of c_p appearance (k_p: fixed).

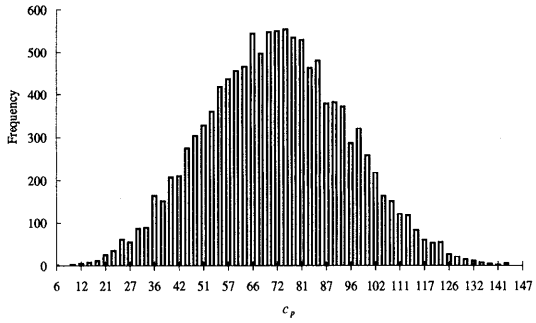


図9 c_p の出現頻度（k_p：変動）
Fig. 9 Frequency of c_p appearance (k_p: fluctuated).

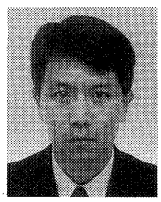
かし，実用化にあたっては，さらに効果的な方法について検討する必要がある．

謝辞 主観的音質評価に協力していただいた皆様に深く感謝します．

参 考 文 献

1) Steel, R. and Vitello, D.: Simultaneous transmission of speech and data using code-breaking techniques, *The Bell System Technical Journal*, Vol.60, No.9, pp.2081–2105 (1981).
2) Wong, W.C., Steel, R. and Xydeas, C.S.: Transmitting data on the phase of speech signals, *The Bell System Technical Journal*, Vol.61, No.10, pp.2947–2970 (1982).
3) 松井甲子雄，中村康弘，ナタウットサムパイブーン：音声通信への文字情報の埋め込み，第18回情報理論とその応用シンポジウム，pp.389–392 (1995).
4) 松井甲子雄：ディジタル透かし，画像電子学会誌，Vol.26, No.3, pp.266–274 (1997).
5) Boney, L., Tewfik, A.H. and Hamdy, K.N.: Digital watermarks for audio signals, *Proc. International Conference on Multimedia Computing and Systems*, pp.473–480 (1996).
6) 岩切宗利，松井甲子雄：適応差分PCM符号化における音声符号へのテキスト情報の埋込み，情報処理学会論文誌，Vol.38, No.10, pp.2053–2061 (1997).
7) Recommendation G.729, ITU (1996).
8) 小澤一範：ディジタル移動通信のための高能率音声符号化技術，トリケップス (1992).

(平成 9 年 11 月 4 日受付)
(平成 10 年 7 月 3 日採録)



岩切 宗利

昭和 45 年生．平成 5 年防衛大学校情報工学科卒業．平成 10 年同大理工学研究科情報数理専攻修了．

松井甲子雄（正会員）

昭和 14 年生．昭和 36 年防衛大学校電気工学科卒業．昭和 40 年九州大学大学院工学研究科電子専攻修了．昭和 56 年防衛大学校電気工学科教授．平成元年同大情報工学科教授．この間暗号学，情報セキュリティ，電子透かし，音声・画像データの符号化に関する研究に従事．著書「画像深層暗号」（森北出版）．工学博士．電子情報通信学会，画像電子学会，映像情報メディア学会各会員．