遺伝的アルゴリズムに基づく音声合成のための スペクトルパタン圧縮法

非会員 濱 上 知 樹 (千葉大学, セコムIS研究所)

正 員 小 坏 成 一 (千 葉 大 学)

正 員 平 田 廣 則 (千 葉 大 学)

Spectrum Pattern Compression Method for Speech Synthesis based on a Genetic Algorithm

Hamagami Tomoki, Non-member (CHIBA University., SECOM IS Laboratory, SECOM Co.,LTD.),
Koakutsu Seiichi, Member, Hirata Hironori, Member (CHIBA University.)

This paper proposes a new applied method to compress the VCV (Vowel-Consonant-Vowel) speech spectrum pattern for synthesis by rule using a Genetic Algorithm (GA). In this method synthesized speech is generated by using parameters from a subset of all of the analyzed VCV frame data. The subset which can generate high-quality synthesized speech is selected automatically by using an optimum combination technique based on GA. We regarded “individuals” in GA as the random selected subsets and defined “fitness” as the inverse of the perceptually based weighting error which is extracted from the difference between natural speech and synthesized speech. Furthermore by using a “generation-dependent mutation width”, the convergence speed can be greatly increased as compared to a ordinary mutation technique. The results from a listening experiment indicate that the synthetic speech generated from the compressed spectrum pattern is of sufficient quality to be of practical use.

キーワード：遺伝的アルゴリズム, 規則音声合成, 音声圧縮

1. ま え が き

音声合成技術には、音声符号化技術、分析合成方式、規則合成方式など、多岐にわたる方法が存在し、様々な音声パラメータの表現が提案されている。このうち、線形予測 (linear predictive coding, LPC) 係数は、少ない情報量で、効率的に音声のスペクトル包絡を表すことのできるパラメータとして、古くから研究され、広く利用されている⁽¹⁾⁽²⁾。また、LPC と等価な表現である偏自己相関係数 (partial auto-correlation coefficient, PARCOR) および線スペクトル対 (line spectrum pair, LSP) は、フィルタの安定性や、補間特性の面で優れた表現であることが知られている⁽³⁾⁽⁴⁾。

LPC を用いて音声进行分析する場合、一般に 15ms から 40ms の窓 (フレーム) 内の波形について 1 回の LPC 分析を行い、LPC 係数、PARCOR, LSP などの、スペクトル包絡に関するパラメータ (スペクトルパラメータ) およびピッチ、パワーなどの音源情報のパラメータ (音源パラメータ) を抽出する。さらにフレームを時間方向に 5ms から 20ms 程度ずらしながら連続分析を行い、時系列のパラ

メータ系列を得る。これらのスペクトルパラメータは、時間方向の冗長な成分を多く含むことが知られている。一般に母音や撥音、鼻音などの定常部では、数十 ms にわたってほぼ一定のスペクトルパタンが続き、隣接するフレーム間のスペクトルパラメータは連続的となる。一方、破裂音や摩擦音など、非定常な特徴を持つ子音では、隣接するフレーム間のスペクトルパラメータは、ほぼランダムな動きをする。これらスペクトルパラメータの時間方向の冗長性を利用したパラメータ圧縮の方法として、マトリクス量子化が提案されている⁽⁵⁾。マトリクス量子化では、スペクトルパラメータの空間領域における多次元ベクトルを、さらに時間方向に拡張し、時空間領域でのパラメータの偏りを利用することで、効率のよいパラメータ圧縮を実現している。しかし、符号化、復号化の際に必要なマトリクスパタンを、送受信双方で用意する必要があり、本質的なデータ圧縮にはなっていない。また、標準パタンの作成方法もヒューリスティックな方法が多く使われており、得られる合成音の品質は、マトリクスパタンの作成方法に依存する部分が多い。

本質的なパラメータの圧縮を行うためには、連続的に分

析された時系列音声パラメータ系列（フレーム系列）から冗長なフレームを取り除く、最適サンプリングの操作が必要である。すなわち、本問題は、全フレームの集合から、最適な部分集合を抽出する、組合せ最適化問題に帰着される。合成音の品質に大きく寄与するフレームだけを、効率的に選択するアルゴリズムを用いることで、例えば規則合成における音声単位データベースの圧縮などに応用できる。

本論文においては、以上のような背景に基づき、ヒューリスティックな方法とランダムな方法を合わせた最適化手法である、遺伝的アルゴリズム（GA）⁽⁶⁾に着目し、これを規則音声合成の基本単位である VCV データの圧縮に応用する方法を提案する。VCV データを対象とするのは、(1) 規則音声合成の基本単位であり、応用価値が高い、(2) 単語発声、文章発声の圧縮を行う際に応用が可能、などの理由による。

以下、第2章では、本論文で用いる GA の概略について述べる。第3章では、音声进行分析して得られるパラメータ系列に対し、GA を応用する方法について述べる。第4章では、人工的に作られた評価用の合成音に GA を応用し、GA の設定パラメータと動作の関係について調べた結果を述べる。第5章では、分析合成と規則合成を用いた音声パラメータの圧縮実験を行い、品質を評価した結果を述べる。

2. 遺伝的アルゴリズム

〈2・1〉単純GA GA は、自然界における進化をモデル化した、大域的な同時最適解探索が可能なアルゴリズムである。特に、多峰性の誤差関数をもつ最適化問題に適している。GA では、世代（Generation）を形成する個体（Individual）の集合（Population）において、環境への適応度（Fitness）の高い個体が次世代に受け継がれるように選択（Selection）される。次に交叉（Crossover）や突然変異（Mutation）を繰り返しながら、集団全体が環境に適応するようになる。

各個体は、染色体によって、その表現形質が決定される。また、各個体の形質によって、その個体の環境に対する適応度が決定される。選択、交叉は、染色体への操作によって行われ、現世代の形質の一部が次世代に受け継がれたり、現世代に存在しない新たな形質を加える操作が行われる。

このように、集団を形成する各個体の染色体に対し、適応度計算、選択、交叉、突然変異を行う方法は、単純GA⁽⁷⁾と呼ばれている。本論文においては、単純GAを用いることにする。以下、本論文で用いる GA の詳細について述べる。

〈2・2〉個体および染色体の符号化 N 個の要素 $\omega_n (1 \leq n \leq N)$ よりなる、以下のような集合 $S_N^{(0)}$ を考える。

$$S_N^{(0)} = \omega_1, \omega_2, \dots, \omega_N \quad (1)$$

$S_N^{(0)}$ から、 $M (1 \leq M \leq N)$ 個の要素をランダムに選択して作られる部分集合を $S_{M/N}$ と表す。このとき、全集合を $C = M/N$ の割合で圧縮したとみなし、 C を圧縮率と呼

図1 組合せパタンの染色体、個体への変換

Fig. 1. Conversion of combination to chromosome, and individual.

ぶことにする。

$$S_{M/N} \subset S_N^{(0)} \quad (0 < C = M/N \leq 1) \quad (2)$$

ここで、任意の部分集合を GA における個体とみなすことにする。また、全集合 $S_N^{(0)}$ から作られた j 番目の個体を $S_{M/N}^{(j)} (1 \leq j \leq J)$ と表し、 $S_{M/N}^{(j)}$ の染色体 $X_{M/N}^{(j)}$ を、以下のようなリストで表すことにする。ここで J は、全個体数である。

$$X_{M/N}^{(j)} = x_1^{(j)}, x_2^{(j)}, \dots, x_M^{(j)} \quad (3)$$

ここで、 $x_m^{(j)}$ は、全集合 $S_N^{(0)}$ から選択された要素 ω_n の要素番号 $n (1 \leq n \leq N)$ であり、 $x_m^{(j)} < x_{m+1}^{(j)}$ であるとする。すなわち。この染色体は、対立遺伝子が $1, 2, \dots, N$ であり、重複を許さない制限がある。このような簡単な染色体を用いることにより、後述の交叉や突然変異の操作と、物理的な要素選択操作との間に、直接的な対応関係が得られる。

染色体 $X_{M/N}^{(j)}$ を用いると、この染色体に対応する個体 $S_{M/N}^{(j)}$ は、以下のようにして復号できる。

$$S_{M/N}^{(j)} = \omega_{x_1^{(j)}}, \omega_{x_2^{(j)}}, \dots, \omega_{x_M^{(j)}} \quad (4)$$

〈2・3〉適応度 j 番目の個体 $S_{M/N}^{(j)}$ の適応度として、全集合 $S_N^{(0)}$ との類似度を評価する。評価にあたり、個体が占める空間を全集合が占める空間に拡張する。拡張された個体を $S_N^{(j)}$ と表し、以下の式で定義する。

$$S_N^{(j)} = \hat{\omega}_1, \hat{\omega}_2, \dots, \hat{\omega}_N \quad (5)$$

$$\hat{\omega}_k = \omega_{x_m^{(j)}} \quad (x_m^{(j)} \leq k < x_{m+1}^{(j)}) \quad (6)$$

ただし、 $x_1^{(j)} = 1, x_{M+1}^{(j)} = N + 1$ とする。

各個体毎の適応度 $fitness^{(j)}$ を、全集合 $S_N^{(0)}$ と拡張された部分集合 $S_N^{(j)}$ との比較を行う関数 $f()$ を用いて、以下の式のように表す。

$$fitness^{(j)} = f(S_N^{(j)}, S_N^{(0)}) \quad 1 \leq j \leq J \quad \dots (7)$$

関数 $f(S_N^{(j)}, S_N^{(0)})$ の具体的な形式は第3章で述べる。

〈2・4〉 選 択 GA における選択操作には、エリート保存方式、ルーレット方式、ランキング方式などが知られている。ここでは、エリート保存方式とルーレット方式の変形を併用した「足切り確率」による選択操作を行う。

適応度の高い順に並びかえられた j 番染色体 $X_{M/N}^{(j)}$ において、足切り確率 P_s と、全個体数 J を用いて、 $P_s J$ 個の不良個体を、優秀な個体に置き換える操作を行う。ここで、もっとも優秀な個体染色体 $X_{M/N}^{(1)}$ は置き換えの対象とせず、2 番目に優秀な個体染色体 $X_{M/N}^{(2)}$ から $X_{M/N}^{(P_s J+1)}$ までを、 $X_{M/N}^{((1-P_s)J+1)}$ から $X_{M/N}^{(J)}$ の範囲に複写する。

$$X_{M/N}^{((1-P_s)J+j+1)} = X_{M/N}^{(j+2)} \quad (j = 0, 1, \dots, P_s J - 1) \quad (8)$$

〈2・5〉 交 叉 2.2 節の定義から、 j 番目の個体染色体 $X_{M/N}^{(j)}$ と、 k 番目の個体染色体 $X_{M/N}^{(k)}$ は、以下のように表すことができる。

$$X_{M/N}^{(j)} = x_1^{(j)}, x_2^{(j)}, \dots, x_M^{(j)} \quad \dots (9)$$

$$X_{M/N}^{(k)} = x_1^{(k)}, x_2^{(k)}, \dots, x_M^{(k)} \quad \dots (10)$$

このとき、 $X_{M/N}^{(j)}$ 、 $X_{M/N}^{(k)}$ を交叉させ、双方の形質を遺伝させた次世代の個体を、以下に示す方法で求める。まず、交叉直後の次世代個体染色体 $X_{M/N}^{(j)}$ 、 $X_{M/N}^{(k)}$ を、ランダムな値 m ($1 < m \leq M$) における一点交叉によって求める。

$$\begin{aligned} X_{M/N}^{(j)} &= x_1^{(j)}, x_2^{(j)}, \dots, x_M^{(j)} \\ &= x_1^{(j)}, \dots, x_{(m-1)}^{(j)}, x_{(m)}^{(k)}, \dots, x_M^{(k)} \quad \dots (11) \end{aligned}$$

$$\begin{aligned} X_{M/N}^{(k)} &= x_1^{(k)}, x_2^{(k)}, \dots, x_M^{(k)} \\ &= x_1^{(k)}, \dots, x_{(m-1)}^{(k)}, x_{(m)}^{(j)}, \dots, x_M^{(j)} \quad \dots (12) \end{aligned}$$

次に、制約条件より、任意の m に関して $x_m^{(j)} \leq x_{m+1}^{(j)}$ 、 $x_m^{(k)} \leq x_{m+1}^{(k)}$ となるように、リストの並びかえを行う。さらに、任意の m, m' ($m \neq m'$) に対して $x_m^{(j)} \neq x_{m'}^{(j)}$ 、 $x_m^{(k)} \neq x_{m'}^{(k)}$ であるという制約条件を満たすために、染色体を検査する。遺伝子の重複が発見された場合、重複している要素分について、新たにランダムな候補 m ($1 \leq m \leq M$) を発生させ、制約条件を満たすようにリストを修復する。以上の遺伝子操作を行った結果、新たな世代の染色体が得られる。最後に、新たに得られた染色体と、交叉を行った現世代染色体を交換する。

以上の遺伝子交叉操作は、1 から $m-1$ 番目までの組合せボタンと、 m から M 番目の組合せボタンを入れ換える操作に相当する。よって、本方法は各個体の形質を継承するしくみを、交叉操作中に直接反映した方法となっている。

図2 選択方法

Fig. 2. Selection operation on the chromosomes.

図3 交叉方法

Fig. 3. Crossover operation on the chromosomes.

図4 $GA(N, C, J, P_s, P_m)$ のフロー

Fig. 4. Processing flow of simple Genetic Algorithm.

〈2・6〉 突然変異 j 番個体染色体 $X_{M/N}^{(j)}$ に対し、突然変異率 P_m の確率で、突然変異が発生する。変異直後の個体染色体を、 $X_{M/N}^{(j)}$ と表し、以下のようなリスト同士の加算で実現する。

$$D_M = d_1, d_2, \dots, d_M \dots\dots\dots (13)$$

$$\begin{aligned} X_{M/N}^{(j)} &= X_{M/N}^{(j)} + D_M \\ &= x_1^{(j)} + d_1, x_2^{(j)} + d_2, \dots, x_M^{(j)} + d_M \dots\dots\dots (14) \end{aligned}$$

ここで、突然変異リスト D_M は、 M 個の整数要素 $d_1, d_2, \dots, d_M$ をもつリストである。突然変異を起こした後の染色体 $X_{M/N}^{(j)}$ は、前述の制約条件を満たすために検査され、必要ならば再度変異を生じさせることで、制約条件を満たすように修復される。 D_M の具体的な値については、3 章で述べる。

〈2・7〉 遺伝的アルゴリズムのパラメータ 以上で述べた、符号化、適応度計算、選択、交叉、突然変異、復

号化を繰り返す GA のプロセスを、以下の変数を用い、 $GA(N, C, J, P_s, P_m)$ と表すことにする。例えば、集合の要素数 $N = 48$ 個、圧縮率 $C = 0.5$ ($M = 24$)、個体数 $J = 100$ 、足しり確率 $P_s = 0.25$ 、突然変異率 $P_m = 0.05$ の場合を、 $GA(48, 0.5, 100, 0.25, 0.05)$ と表す。

$GA(N, C, J, P_s, P_m)$ のアルゴリズムを図 4 に示す。

3. 最適フレーム組合せ問題への GA の適用

〈3・1〉 音声分析値への GA の適用 音声 $s_n(t)$ ($0 \leq t < T$) に対し、 P 次の LPC フレーム分析を行い、各フレーム毎に P 次の LSP ベクトルを求める。このとき、フレーム総数 N は、フレーム周期 n_s に対し、 $N = T/n_s$ で表すことができる。 N フレームによって作られる LSP ベクトルの集合を $S_N^{(0)}$ と表す。また、 n 番フレームの LSP ベクトルを、 ω_n と表す。以上の表現により、最適フレーム組合せ問題を、2.2 節以降で述べた GA を用いて解くことができる。

〈3・2〉 適応度の計算 本問題は、全フレーム（全集合）によって作られる合成音 ($s_0(t)$) と、ある個体（部分集合）によって作られる合成音 ($s^{(j)}(t)$) との聴覚上の差異を最小とする部分集合 $S_{M/N}^{(j)}$ を見つけることが目的である。そこで、 $s_0(t)$ 、 $s^{(j)}(t)$ 2 つの合成音間の誤差波形に、さらに聴覚的重み付けを施した波形の平均自乗値を用いることで、適応度を表すことにする。

$$e^{(j)}(t) = (s_0(t) - s^{(j)}(t))W(t, Z) \dots\dots\dots (15)$$

$$s_0(t) = \epsilon(t) + \sum_{p=1}^P \alpha_p^{(t/n_s)} s_0(t) Z^{-p} \dots\dots\dots (16)$$

$$s^{(j)}(t) = \epsilon(t) + \sum_{p=1}^P \beta_p^{(t/n_s)} s^{(j)}(t) Z^{-p} \dots\dots\dots (17)$$

$$W(t, Z) = \frac{1 - \sum_{p=1}^P \alpha_p^{(t/n_s)} Z^{-p}}{1 - \sum_{p=1}^P \gamma^p \alpha_p^{(t/N_s)} Z^{-p}} \dots\dots\dots (18)$$

ここで、 $\epsilon(t)$ は、LPC 分析の際に得られる予測誤差波形である。全フレーム $S_N^{(0)}$ をフィルタ係数とし、 $\epsilon(t)$ を音

図5 聴覚的重み付けを施した誤差評価

Fig. 5. Error evaluation using perceptually based weighting filter.

源とする合成音 $s_0(t)$ は、自然音 $s_n(t)$ と同じ波形となる。 $\alpha_p^{(t/n_s)}$ は、サンプル点 t の近傍に位置する t/n_s 番フレームの LSP パラメータベクトル ω_{t/n_s} を変換して得られる LPC 係数である。一方、 $\beta_p^{(t/n_s)}$ は、拡張された個体 $S_N^{(j)}$ における、サンプル点 t の近傍に位置する t/n_s 番フレームの LSP パラメータベクトルを変換して得られる LPC 係数である。 γ は重み係数で、 $0 < \gamma < 1$ の値をとる定数である。聴覚重み付けフィルタ $W(Z)$ は、フォルマント近傍に誤差を集中させる役割があり、これによって、聴覚マスキング効果を利用した誤差の評価が可能となることが知られている⁽⁹⁾。

$e^{(j)}(t)$ の自乗和平均値の逆数をもって、適応度関数とする。

$$error^{(j)} = \sqrt{\frac{\sum_{t=0}^{T-1} (e^{(j)}(t))^2}{T}} \dots\dots\dots (19)$$

$$fitness^{(j)} = (error^{(j)})^{-1} \dots\dots\dots (20)$$

〈3・3〉 選 択 前述の適応度関数を用いて、各個体毎に適応度が与えられる。適応度の高い順に並びかえられた集団に対し、2.4 節に基づく選択操作が行われる。

〈3・4〉 交 叉 選択によって新たに残される個体染色体は、2.4 節で述べたように、最も適応度の高い個体染色体 $X_{M/N}^{(1)}$ が 1 つ、2 番目に適応度の高い個体染色体 $X_{M/N}^{(2)}$ から $P_s J + 1$ 番目までが 2 つずつ、 $P_s J + 2$ 番目から、 $(1 - P_s) J$ 番目までが 1 つという構成になる。これらの個体染色体のうち、 $X_{M/N}^{(1)}$ を除く $J - 1$ 個の染色体をランダムに並び替え、先頭から 2 つずつの組を順に作り、すべての組に対して 2.5 節に基づく交叉を行う。 $J - 1$ が奇数の場合、最後に残された 1 つの個体染色体は、そのまま次世代に残す。このように、 $X_{M/N}^{(1)}$ を除くすべての個体を用いて交叉を行い、多様な子個体を生成することで、次世代で $X_{M/N}^{(1)}$ より適応度の高い個体が発生する確率をできるだけ大きくする。

〈3・5〉 突然変異 各交叉操作時に、突然変異率 P_m に基づき、新たに作成する個体に突然変異を生じさせる。本方式における突然変異は、現在選択されているフレーム番号 $x_m^{(j)}$ を、別の重複しないフレーム番号 $x_m^{(j)} + d_m$ に変更する手続きである。 d_m が小さい場合は、 $x_m^{(j)}$ 番フレーム近傍のフレームを新たに選択する局所的な変異となる。逆に、 d_m が大きい場合は、 $x_m^{(j)}$ 番フレームとは大きく離れた場所から新たなフレームを選択する大域的な変異となる。

ここでは、簡単のため、突然変異を生じさせるリスト D_M の要素 d_m を以下のように設定する。

$$d_m = \begin{cases} d_x & (m = m_r) \\ 0 & (m \neq m_r) \end{cases} \dots\dots\dots (21)$$

$m_r (1 < m_r \leq M)$ は、突然変異が発生した時点で、ランダムに決定される整数である。また、 d_x は、 $-L \leq d_x \leq L$ の範囲でランダムに決定される整数と定義する。 L を突然変異の幅と呼ぶことにする。

4. GA による音声フレームの最適組合せ実験

〈4・1〉 評価用合成音による実験 GA による音声フレームの最適組合せの動作を検証するために、合成音として作られた音声フレーム系列に対し、GA を適用する実験を行う。女声の 5 母音 /a/, /o/, /i/, /u/, /e/ を表す代表的なフレームを、それぞれ 10 フレーム (62.5ms) ずつ連続させ、合計 50 フレーム (312.5ms) のスペクトルデータ /aoiue/ を用意する。これを基本周波数 350Hz のパルス音源で駆動した合成音は、各母音間遷移部が存在しない合成音となる。この合成音に対する最適フレームの組合せは、各母音の先頭フレーム部分 (5 フレーム分) を含む部分集合である。このとき、誤差は 0 となる。

GA のパラメータのうち、 $N = 50$, $C = 0.3$ ($M = 14$), $P_s = 0.25$ および $L = 5$ は固定とする。個体数 J は 200, 100, 50, 30, の 4 種類とする。突然変異率 P_m は、0.0, 0.05, 0.10 の 3 種類とする。これらの組合せ (12 種類) に基づく GA を、初期個体を変えてそれぞれ 10 回ずつ実行し、合計 120 回の実験を行った。

実験に先立ち、アルゴリズムの収束条件を定める予備実験を行った。上記パラメータのうち、 $P_m = 0.0$, $J = 200$ の条件にて 10 回の試行を行い、各世代の集団を観測した。その結果、最も適応度の高い個体の染色体が、20 世代以上に渡って変化しなかった時に、集団中のすべての個体が同じ染色体の個体で占められた状態となっていることがわかった。これより、本アルゴリズムの収束条件を、20 世代にわたって、同じ適応度の個体が、最も適応度が高いと判定された場合とした。

図 6 に、評価用合成音の波形、スペクトログラムおよび GA(50, 0.3, 100, 0.25, 0.0) で選択されたフレーム位置の世代変化を示す。図 7 に、誤差の世代変化を示す。表 1 に、収束結果中に、誤差が 0 となる最適フレームの組合せを含む確率の平均値、および平均収束世代数を、各 J および P_m 毎に示す。

以上の実験から、以下のことがいえる。

- (1) $J > 100$ の場合、 $P_m = 0$ としても、初期個体からの選択、交叉のみで最適解を導き出すことができる。
- (2) $J < 50$ の場合、 $P_m = 0$ とすると、少ない世代数で収束してしまい、最適解を得る確率が低くなる。
- (3) 個体数の大小に拘らず、 P_m が大きくなると、収束に必要な世代数が大きくなる。

〈4・2〉 突然変異の幅 L の調整 前述の実験結果から、最適な解を確実に見つけるためには、十分な個体数を用意することが重要であると結論づける。一方、計算の効率を考えた場合、少ない個体数で、最適解に速く収束する方法が望ましい。一般に、個体数を少なくした場合は、大域探査が行えずに、局所解に陥り易い傾向がある。これに対し、突然変異を有効に用いることによって、個体のバリエーションを増やし、大域的な探査が可能になる。一方、本問題の

表 1 最適なフレームが選択された割合/平均収束世代数の変化 ($L = 5$)
 Table 1. Selection rate of optimum frame and averaged convergence generation number.

P_m ($L = 5$)	J			
	200	100	50	30
0.00	1.00/27.2	1.00/28.3	0.94/23.7	0.80/20.2
0.05	1.00/42.1	0.98/40.2	0.94/65.5	0.94/61.2
0.10	1.00/54.5	1.00/62.1	1.00/82.1	1.00/95.8

特徴として、最適解は準最適解の近傍にある可能性が高いことがわかっている。すなわち、ある程度の収束が行われた後は、局所的な探索を積極的に行うことにより、高い適応度の個体を効率よく見つけることができると考えられる。以上の考察から、突然変異の幅 L を、以下に示す式により、世代数 g に応じて順次小さくする方法を採り入れることにする。

$$L(g) = L_0 \beta^g \dots\dots\dots (22)$$

ここで、 β は、 $0 < \beta < 1$ の実数である。これより、世代数が小さいときには、選択フレームを大きく移動させ、個体のバリエーションを増やす大域的探索が行える。また、世代数が増すにつれ、選択フレームを小さく変化させる局所的な探索が行える。

表 2 に、4.1 節の評価用合成音に対し、 $L(g)$ を用いた場合の、最適なフレームが選択された割合/平均収束世代数を示す。試行回数は 10 回である。突然変異のパラメータは、試行錯誤の結果、 $L_0 = 30$ 、 $\beta = 0.8$ 、 $P_m = 0.3$ としている。以上の結果から、本問題においては、突然変異の幅を

図 6 $GA(50, 0.3, 100, 0.25, 0.0)$ のフレーム選択結果

Fig. 6. The result of selecting frames from synthetic vowel spectrum pattern.

図 7 $GA(50, 0.3, 100, 0.25, 0.0)$ のエラー変化
 Fig. 7. Relation between error (=1/fitness) and generation number.

表 2 最適なフレームが選択された割合/平均収束世代数 (突然変異の幅が可変の場合)
 Table 2. Selection rate of optimum frame and averaged convergence generation number.

P_m ($L = L(g)$)	J			
	200	100	50	30
0.30	1.00/28.5	1.00/32.4	1.00/34.7	1.00/33.3

世代数に応じて可変にすることにより、少ない個体数でも、最適な解が比較的速く求められることがわかった。

5. 合成音声による実験

〈5・1〉 分析合成音による評価実験 本方法を用いて圧縮を行った VCV データと、圧縮を行っていない VCV データを用いた合成音をそれぞれ作成し、5 段階オピニオン評価による品質の比較試験を行った。以下に示す手順に従い、分析合成音をベースとした合成音を作成した。

- (1) 10 種の単語発声データ (16kHz, 16bit) を用意し、フレーム長 25.0ms、フレーム周期 6.25ms、ハミング窓による 16 次 LPC 分析を行い、LSP パラメータ系列、ピッチ、パワーなどの連続パラメータを得る。
- (2) それぞれの発声データに対し、目視にて VCV 単位のラベリングを施し、VCV 単位毎にフレームを分割する。
- (3) 各 VCV フレーム系列に対し、 $GA(N, C, 30, 0.25, 0.3)$ および $L(g)$ ($L_0 = 30$, $\beta = 0.8$) を用いた GA により、最適フレーム選択を行い、LSP パタンの圧縮をはかる。ここで N は、当該 VCV のフレーム数である。
- (4) 圧縮を施したパラメータを用いて、合成音を作成する。

圧縮率 C は、1.0, 0.8, 0.5, 0.3 の 4 種類とする。それぞれの合成音を $s_{1.0}$, $s_{0.8}$, $s_{0.5}$, $s_{0.3}$ と表す。さらに比較用データとして、 $C = 0.8, 0.5, 0.3$ の場合の初期個体を用いて、ランダムにフレームを選択した合成音を用意する。これを、 $r_{0.8}$, $r_{0.5}$, $r_{0.3}$ と表す。音源には残差波形を用い

る。これらの7サンプルをランダムに提示し、5段階評価による絶対評価を行う。被験者は3名であり、1人に対し3回の評価を行う。図8に評価結果を示す。本実験結果より、LSPの圧縮率を1/3程度に圧縮しても、普通程度の品質が保たれることがわかった。また、フレームをランダムに選択して作られたいずれの合成音に対しても、本アルゴリズムを用いた合成音は、より高い品質が得られることがわかった。

〈5・2〉 規則合成音による評価実験 自然発声されたVCV音節に対して、本アルゴリズムによるフレーム選択を行い、規則合成で用いるVCVデータベースを以下に述べる方法で作成する。発話者は女性ナレータである。VCVを無意味単語 /arVCVra/中に埋め込み、5.0mora/sの発話速度で発声を収録する。収録された音声から、目視にてVCVの両端の母音部エネルギー重心位置でVCVデータを切り出す。総VCV数は920個である。これらのVCVデータ(16kHz, 16bit)それぞれに対し、フレーム長31.25ms、フレーム周期6.25ms、ハミング窓による16次のLPC分析を行い、LSP係数ベクトル時系列データ($S_N^{(0)} = \omega_1, \omega_2, \dots, \omega_{N_{VCV}}$)を得る。 N_{VCV} は、当該VCV音節の分析フレーム数である。平均フレーム数は42フレーム(262.5ms)/VCVであった。

これら920個のVCV音節それぞれに対し、 $GA(N_{VCV}, 0.3, 100, 0.25, 0.3)$ のアルゴリズムを用いてLSPパターンの圧縮を行う。ただし、 $L = L(g)$ ($L_0 = 30$, $\beta = 0.8$)を用いる。

図9に、/amyo/ ($N_{VCV} = 47$)について、本アルゴリズムを適用し、合成音を作成した例を示す。また、図10にこのときのエラーの世代変化を示す。これらの結果より、本方法を用いて選択されたフレームは、スペクトル変化の緩やかな箇所では、まばらに選択され、スペクトル変化の大きい箇所では、細かく選択されていることが確認された。

圧縮されたVCVデータベースを用いて作られた規則合成音($S_{0.3}$)と、圧縮を行っていないデータベースによって作られた規則合成音($S_{1.0}$)をそれぞれ10単語について作成し、対比較による評価を行った。2人の被験者に対し、2回の試験を行い、どちらがより自然であるかを判定する。

図8 分析合成音による5段階オピニオン試験

Fig. 8. The evaluation result of the nature for analysis-synthesis speech by opinion test.

図11に結果を示す。本実験の結果、1/3に圧縮されたVCVデータベースによる合成音は、全フレームを用いた合成音より、やや品質が劣るものの、その差異は小さいことが明らかとなった。これより、本方法を用いてスペクトルパターンを圧縮したVCVデータベースからでも、大きく品質を損なわない程度の規則合成音を作ることが可能であることが明らかとなった。

図9 /amyo/ $GA(47, 0.3, 100, 0.25, 0.30)$, $(L = L(g))$ の結果

Fig. 9. The result of selecting frames and synthetic spectrum pattern from natural VCV speech.

図10 /amyo/ $GA(50, 0.3, 100, 0.25, 0.30)$, $L = L(g)$ ($L_0 = 30$, $\beta = 0.8$) のエラー変化

Fig. 10. Relation between error(=1/fitness) and generation number. (GA applied to natural VCV spectrum pattern)

C=0.3	C=1.0
42%	58%

図 11 規則合成音のプリファレンススコアによる
評価

Fig.11. The evaluation result of synthesized
speech (TTS) by preference score.

6. む す び

GA に基づく、音声合成のための VCV パタン圧縮方法について検討を行った。本方法は、分析されたフレーム系列を全集合として、全集合で作られる合成音に最も聴覚的に近い部分集合を選択する問題を GA で解くものである。評価用に作られた人工データを用いた実験から、高い精度で最適解を導きだせることがわかった。また、個体数が少ない場合でも、突然変異の幅を可変にすることで、比較的速く、最適解が得られることがわかった。本方法を自然音の VCV 音節に適用した結果、スペクトルパラメータを時間方向に 1/3 程度圧縮しても、自然性を大きく損なわない合成音を得られることがわかった。

謝辞

研究の機会を与えていただいた、セコム（株）常務取締役、橋本新一郎 IS 研究所長、並びに、日頃御指導頂く音声情報処理研究室、古村光夫室長に感謝いたします。

（平成 10 年 5 月 29 日受付，同 10 年 8 月 17 日再受付）

文 献

- (1) Atal, B.S. and Hanauer, S.L. : "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Amer., 50, 2(pt.2), pp.637-655 (1971).
- (2) 板倉 文忠, 斎藤 収三 : "最尤スペクトル推定法を用いた音声情報圧縮," 音響誌, 27, 9, pp.17-26 (1971).
- (3) 北脇 信彦, 板倉 文忠, 斎藤 収三 : "PARCOR 形音声分析合成系における最適符号構成," 信学論, J61-A, 2, pp.119-126 (1978).

- (4) 菅村 昇, 板倉 文忠 : "線スペクトル対 (LSP) 音声分析合成方式による音声情報圧縮," 信学論, J64-A, 8, pp.599-606 (1981).
- (5) 白木 善尚, 菅田 雅彰 : "音声の時空間パタン符号化の最適構成法," 音響学会音声研究会資, S85-45 (1985).
- (6) 北野 宏明 編: 遺伝的アルゴリズム (1993) 産業図書
- (7) 北野 宏明 編: 遺伝的アルゴリズム 2 (1995) 産業図書
- (8) B.S. Atal et al, "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates," Proc. ICASSP 82, pp 614-617 (1982).

濱 上 知 樹（非会員） 1966 年 3 月 7 日生。1988 年 千葉大学工学部電気工学科卒業。同年、セコム（株）入社。IS 研究所音声情報処理研究室所属，1998 年在職のまま千葉大学大学院自然科学研究科生産科学専攻後期博士課程に入学，現在に至る。音声の分析・合成の研究に従事。電子情報通信学会，日本音響学会会員。

小 坪 成 一（正員） 1964 年 7 月 7 日生。1992 年，千葉大学大学院自然科学研究科生産科学専攻博士課程修了。同年，千葉大学工学部助手，1997 年同助教授，同年同大学院自然科学研究科助教授となり現在に至る。1994-1995 年，カリフォルニア大学サンタクルズ校客員研究員。VLSI レイアウト設計，確率的最適化手法，ニューラルネットワーク，学習オートマトンの研究に従事。博士（工学）。IEEE, INNS, 電子情報通信学会，計測自動制御学会会員。

平 田 廣 則（正員） 1948 年 6 月 2 日生。1976 年，東京工業大学大学院電気工学専攻博士課程修了。同年，千葉大学工学部助手，1981 年同助教授，1994 年同教授，1997 年同大学院自然科学研究科教授となり現在に至る。大規模システムのモデリング，解析と設計，特に生態システム，VLSI レイアウト，並びに分散型コンピュータの基礎的理論に興味を持つ。工学博士。IEEE, ISEM, INNS, 電子情報通信学会，情報処理学会，計測自動制御学会，システム制御情報学会，人工知能学会会員。