

発声器官を考慮した仮想生物の音声生成

函館工業高等専門学校 ○仲谷彰浩 坂東 樹 竹原直美 石若裕子

要旨

本研究では、仮想生物の声を、モンタージュ・ヴォイス技法を基に声道の共鳴データを遺伝的アルゴリズムによって求め、生成する手法を提案する。本研究では、まず提案手法が期待される効果を発揮するかを検証し、実在する生物の声を対象にした実験を行い、その結果を仮想生物に反映させることで、その声を生成する。

1. はじめに

通常聞くことのできない声を合成する技術に、モンタージュ・ヴォイス[1]という技術がある。これは、写真・銅像などから性別・身長・年齢・顔形を特定し、その人の発声器官をコンピュータ上で電子回路に置き換え、シミュレーションして声を合成する技術である。

本研究では、モンタージュ・ヴォイス技法を基に生物の声道・外形による複雑な共鳴周波数を、遺伝的アルゴリズムを用いて電子等価回路式に置き換えることで、複雑な数値解析を行わずに仮想生物の声を合成する手法を提案する。

2. 発声のメカニズム

人間の発声は次のように知られている。まず肺からの空気によって声帯を振動させることで三角間隙鋸歯状波と呼ばれる声帯の基本振動音が生成される。このときはまだ“言葉”として認識できるものではなく、ブザーのような音である。この基本振動が口腔・鼻腔・喉等からなる声道を通過する際に、声道内の容積に共鳴したものが口や鼻から外部へ放射されることで“声”となるのである[2]。

3. 提案手法

発声器官を簡略化し、声帯の基本振動音である鋸波と声道の容積による共鳴周波数から構成されると仮定する。ここで発声器官を基本振動（鋸波）と電子回路（等価の共鳴周波数が得られるもの）に置き換え、それらを構成するパラメータを遺伝子として遺伝的アルゴリズム（GA）を適用する。それにより発声器官の最適な情報を得る。

3.1 共鳴と電子等価回路

声道の容積による共鳴部分を電子等価回路に置き換えて考えると、共鳴部は基本振動音を入力として与えた場合に生物の声が出力されるような伝達関数として捉えることが可能である。

そこで図2のような回路を考える。これはLC回路の繰り返しであるので、その一つをとりだして（図3）伝達関数を求めると

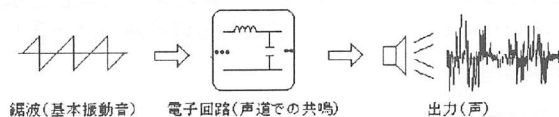


図1 発声の仕組み

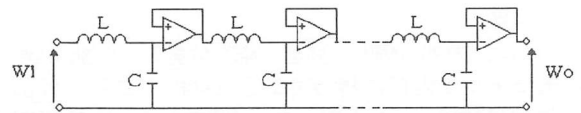


図2 電子回路（共鳴部）の構造

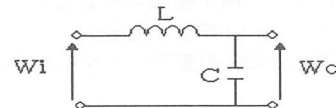


図3 LC回路

$$W_o = \frac{1}{\sqrt{LC}} \sin\left(\frac{1}{\sqrt{LC}}t\right)W_i \dots (1)$$

となる。これが複数になると

$$W_o = \frac{1}{\sqrt{LC_1}} \frac{1}{\sqrt{LC_2}} \dots \sin\left(\frac{1}{\sqrt{LC_1}}t\right) \sin\left(\frac{1}{\sqrt{LC_2}}t\right) \dots W_i \dots (2)$$

と考えることができる。

3.2 遺伝子

GAによる学習で最適な鋸波と電子回路のデータを求める。遺伝子は鋸波の振幅 A と周期 T、電子回路を構成する素子 L と C 複数で構成され、次のような配列データとなっている。但し、L と C については、3.1 から分かるように LC 回路 1 つについて考えると各素子は別個ではなく、 $L \times C = LC$ で表現できることが分かる。よって $LC_1 \cdot LC_2 \dots$ のようにデータを表現する（図4）。

ここでは、エリート保存戦略を用いる。このエリート保存戦略の弱点である局所解への陥りやすさを低く抑えるため、初期集団から派生する遺伝子（突然変異体を含む）の情報に全く左右されていない遺伝子（以下外来体）を混合させる。また、交叉時に一定の確率で bit 列の交換ではなく、ランダムに新たな bit 列を組み込む（交叉変異）。これらにより遺伝子の局所解に近づく偏りを抑止させる。

A	T	LC1	LC2	LC3	LCn
---	---	-----	-----	-----	-------	-----

鋸波振幅 A : 10bit 0.00097~1.0
 鋸波周期 T : 11bit 0.00049~1.0
 素子 LC : 9bit 0.0512~0.0001

図4 遺伝子配列

3.3 評価

適応度はGAによる合成波形と教師波形を周波数解析 (FFT) し、二つの波形を各周波数成分について比較し、誤差の総和が小さいほど高くする (1が最大)。

$$fitness = \frac{1}{1 + \sum gosa}$$

4. 実験

4.1 手法の有効性を調査

図5(c)の音声波形について下の条件で実験を行い、この手法で発声器官の情報が正しく得られるか調査した。このとき用いたGAのパラメータを以下に示す。

表1 実験に用いたGAのパラメータ

個体数	100	個	世代数	5000	世代
LC素子	10	個	選択方式	トーナメント方式	
エリート保存	5	%	突然変異	1	%
交叉	2点交叉		交叉変異	2	%
外来体	55	%			

世代ごとの結果を見ると (図5)、波形は世代が進むにつれ教師波形に近づいている。つまり手法としては有効であるといえる。しかし今回実験で求めた5000世代目でも、まだ発声器官はこうであるという確証を持つ段階ではない。これは、波形 (遺伝子) の評価に利用したFFTの結果 (図6) でも同様の結果が得られている。

4.2 手法の再検討

今回提案したGAによる手法を、実験結果を基に検討する。発声器官の断定に確証を持っていないのは、適合度が、500世代を越してから変化しない状況があるからである。これは単純に実行世代数が少なからざるものだけでなく、適合度の評価が単一方法であることにも要因があると考えられる。

また、仮想生物の音声生成に反映するにあたって電子回路の構成も見直す必要がある。(1)、(2)を見ればわかるが、LC回路がいくら増えても、各周波数について位相が考慮されていない。今回この手法の確認をするのに用いた波形はそうでもないが、生物の声のように複雑な波形になると無視できない情報であると思われる。

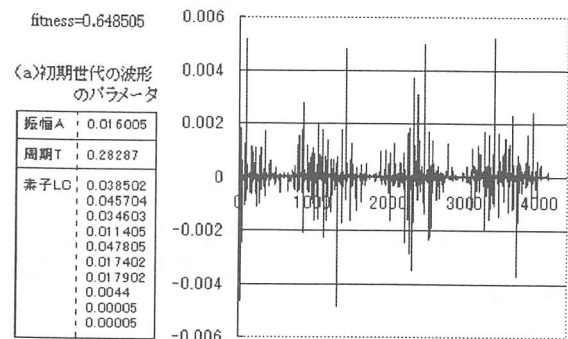
5. 終わりに

今回の実験では、発声器官の共振周波数をGAを利用して求める手法の有効性を検証した。結果、改善が必要な点が見られたが方法としては有効であると思われる。

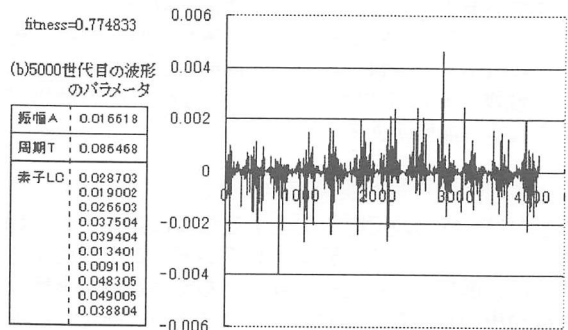
今後の課題は、得られたデータから改善点を見出し、このシステムを仮想生物に応用させ、その音声を合成することである。

参考文献

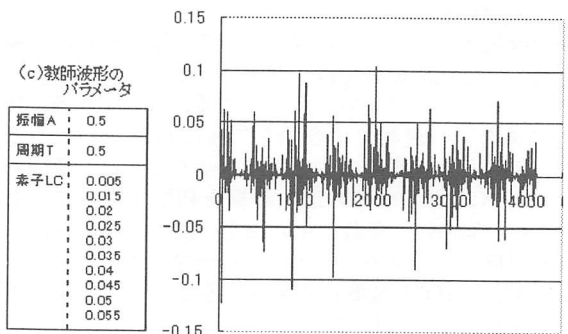
- [1] 日本音響研究所 : モンタージュ・ヴォイス
<http://www3.alpha-net.ne.jp/users/japaco/MV.htm>
 [2] 今井聖 : 音声信号処理 1996 森北出版



(a) 初期集団の波形



(b) 5000世代目の波形



(c) 期待される波形 (教師波形)

図5 世代ごとの波形と教師波形

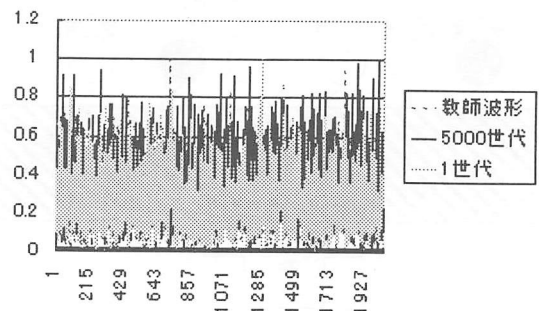


図6 世代ごとの合成波形と教師信号のFFTの結果