

|              |                                                                                 |
|--------------|---------------------------------------------------------------------------------|
| Title        | 基本周波数とスペクトル包絡を制御した歌声合成に関する研究                                                    |
| Author(s)    | 清水, 一郎                                                                          |
| Citation     |                                                                                 |
| Issue Date   | 2003-03                                                                         |
| Type         | Thesis or Dissertation                                                          |
| Text version | author                                                                          |
| URL          | <a href="http://hdl.handle.net/10119/1663">http://hdl.handle.net/10119/1663</a> |
| Rights       |                                                                                 |
| Description  | Supervisor: 赤木 正人, 情報科学研究科, 修士                                                  |

# 基本周波数とスペクトル包絡を制御した 歌声合成に関する研究

清水 一郎 (110057)

北陸先端科学技術大学院大学 情報科学研究科

2003 年 2 月 14 日

キーワード: スペクトル包絡, 基本周波数, 歌声合成, 話声, スペクトル制御.

本論文は話声とメロディから歌声への高品質な歌声合成を実現することを目的とする。ソースフィルタモデルの STRAIGHT を用いることにより、基本周波数、スペクトルを制御しながら、話声から歌声を合成する手法を提案する。

人が歌を歌うことは感情や想いを伝える重要な感情表現の一つであり、歌がどのようなものを研究することは大事である。人間は歌を歌うとき自分の音色を保ちつつ、メロディにあわせ音程を激しく変化させている。つまり、人間は歌を歌うとき基本周波数を大きく変化させながら音色をたもっている。

歌声の音色に関係しているといわれるスペクトルについて述べる。話声の母音において、第 1 フォルマント、第 2 フォルマントにより音韻性が決まり、第 3 フォルマント以上のフォルマントはあまり変化せず、個人的特徴に関連していると言われている。一方、歌声において、必ずしも 2 つのフォルマントが母音らしさを決定しているのではないことがいわれている。たとえば、男性のオペラ歌手が歌った [i:] という母音の第 1 と第 2 フォルマントの周波数は話声における母音図では [y:] に相当する。また、女性のソプラノ歌手の高い声は 1000Hz ないし 1400Hz もの高さの基本周波数で歌うので、話声の第 1 フォルマントより高いという現象が起る。そこで、第 1 フォルマントを基本周波数の近傍へ動かす操作を行う。声種による音色の違いは各声種の歌声について母音のフォルマントの違いに着目して研究されており、一般的に高い声種の場合フォルマント周波数が高いほうに偏移することが述べられている。音色を保つためには基本周波数の変化にともないスペクトルが変化しているのではないかと考えられる。そこで、基本周波数の変化に対応してスペクトルを制御することを考える。

話声から高品質な歌声を合成するために、話声と歌声を分析し比較する。話声と歌声の F0 に対する フォルマント周波数を分析した。話声において、基本周波数が上昇すると第 1 フォルマント、第 3 フォルマントは上昇し、第 2 フォルマントが下降する傾向がみられた。話声において、基本周波数が上昇すると第 1 フォルマント、第 2 フォルマント、第 3

フォルマントが上昇する傾向がみられた。そして、話声のスペクトル  $P$  と話声の  $F_0$  に対する歌声の  $F_0$   $\bar{F}_0$  から歌声のスペクトル  $I$  を算出する式  $I = 0.13\bar{F}_0P + 0.84P$  を最小 2 乗法により導出した。また、歌声のヴィブラート部の  $F_0$  と  $F_1$  の微細変動に時間のラグなく相関があることがわかった。この分析により得られた結果からスペクトルの周波数方向の制御を行う。スペクトルの周波数方向の制御は  $I = 0.13\bar{F}_0P + 0.84P$  の式を用い、1[ms] ごとに話声のスペクトルを写像することにより歌声のスペクトルを作成する。そして、ヴィブラート部の  $F_0$  に対応した揺れに同期して  $F_1$  の微細変動として  $F_1$  に付け加える。

次に、歌声での音節の長さはメロディにより決定されるので、歌声の子音の時間の長さがわかれば母音の時間の長さが決定されと考え、話声の子音の時間の長さに対する歌声の子音の時間の長さを分析する。分析により歌声の子音の時間の長さは、子音の調音方式ごとに話声の摩擦音のとき 1.28 倍、破裂音のとき 1 倍、半母音のとき 2.37 倍、鼻音のとき 1.43 倍、/y/ のとき 1.22 倍にすればよいことがわかった。これらの分析結果を用い、スペクトルの時間方向の制御方法を考案した。時間方向における話声のスペクトルから歌声のスペクトルへの変形のために、子音定常部、調音結合部 (40ms)、母音定常部にセグメンテーションする。歌声のスペクトルの制御は調音結合部はそのまま話声のものをを用い、話声の母音、子音の定常部のスペクトルを伸長する。歌声の子音の長さは話声の子音の時間の長さで調音方式により分析結果から決定される。歌声の母音の長さはメロディから子音の長さを引き算すればよい。母音、子音の定常部の時間を伸長する手法として 3 つを提案する。1 つ目は話声の定常部の 1 点を取ってきて、それを伸ばしたい時間の長さだけ並べることにより歌声の定常部を作成する。2 つ目は FFT8192 点で分析した話声の定常部を伸ばしたい時間の長さにするために各周波数方向のチャンネルごとに時間方向へ線形補間を行う。3 つ目は FFT8192 点で分析した話声の定常部を伸ばしたい時間の長さにするために各周波数方向のチャンネルごとに時間方向へスプライン補間を行う。3 つの方法でスペクトルの制御を行って作成された歌声合成音の受聴により、スプライン補間を行ったものが最も高品質だと思われたので、スペクトルに時間方向への制御はスプライン補間を採用することとした。

話声とメロディから、歌声の  $F_0$ 、スペクトル包絡の時間方向、周波数方向の制御を行い歌声合成音を作成した。歌声合成音が高品質であるか、スペクトル包絡制御が高品質な歌声を合成するために役立っているかを調べるために、聴取実験を行った。これらの研究の手法で作成された歌声合成音を シェツフェの対比較法を用い評価した。

聴取実験の結果から、スペクトルの時間方向、周波数方向の制御の有効性が示された。また、話声メロディから歌声を合成する手法の有効性が示せた。