日本語テキスト音声合成を利用した歌唱合成システム

5G－4

山本篤志　松本達郎　片江伸之　木村晋太
（株）富士通研究所

1　はじめに

当社では，音声合成システムの研究開発を行なっており，これまでに，通常の文章を自然に読み上げるシステムを開発した[1]。

今回は音楽，ゲームなどのアミューズメント分野への応用として，この音声合成システムをベースとした歌唱合成システムを開発した。このシステムの特徴は，以下のとおりである。

- 自然音声を分析して得られた歌唱モデル（遷移モデル，ビブラートモデル，ゆらぎモデル）が組み込まれており，自然な歌声の合成が可能である。
- 合唱の合成が可能である。

2　システム構成

本システムの構成を図1に示す。

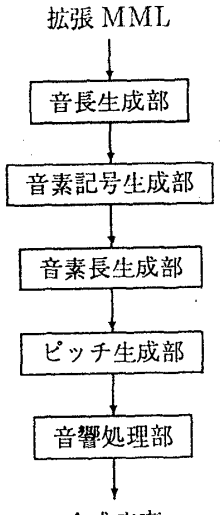

図1　システム構成

2.1 拡張 MML

計算機への入力に用いる楽譜と歌詞の情報の記述方式として従来パソコン上での自動演奏に使われていたMML(Music Macro Language)を拡張して使用した。これはMMLに各音符の記号に対応した歌詞を仮名で併記するものである(図3.b)。

2.2 音長生成

楽譜情報で表される音長は相対的な時間長であるから，音長の絶対時間長を求める。相対的な音長をNT，絶対時間長をANT，テンポをTeとすると次式のようになる。

$$ANT = 1000 \times \frac{60 \times 4}{NT} \times \frac{1}{Te} \quad [msec]$$

タイでつながれた音符をそれらの和の音長を持つ一つの音符に置き換える。

2.3 音素記号生成

拡張MMLの歌詞情報より音素記号を生成する。

2.4 音素長生成

1. 音長ANTを音符に割り当てられている歌詞の音節数で割って音節長とする。
2. その音節が1モーラであれば音節長をモーラ長とする。撥音節・促音節であればルールより撥音長・促音長を求め，それらを引いた値をモーラ長とする。
3. 単独母音，撥音，促音，無音であれば，モーラ長をそれらの音素長とする。ＣＶ音節であればルールより子音長を求め，音節長－子音長を母音長とする。
4. 一つの音符の歌詞として促音のみが割り当てられている時（図2），促音の前の音符に割り当てられている歌詞の母音部分の音長を伸ばし，その分促音の音長を縮める（ただし，音程の時間長は変更しない）。

2.5 ピッチ周波数生成

拡張MMLの音程記号から声の高さに相当する物理量であるピッチ周波数パタンを生成する（図3）。この時，単純な音程パタン（図3.c）で合成すると，金属音的で不自然な音質となる。そこで，自然音声の分析結果

Singing Synthesis System using Japanese Text-to-Speech Conversion
Atsushi YAMAMOTO, Tatsuro MATSUMOTO, Nobuyuki KATAE and Shinta KIMURA
Fujitsu Laboratories Ltd.

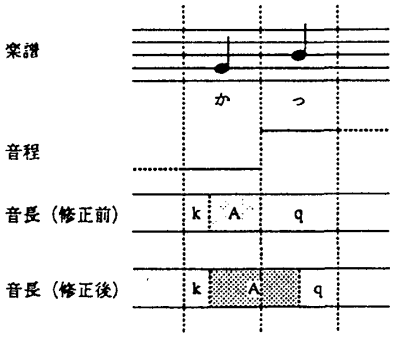

図 2 促音の音長

を基に以下のような三つのモデルを作成した。

遷移モデル　人が歌う声の高さは急激に変化せず，滑らかに連続的に変化する。自然音声を分析したピッチ周波数パタンではこの音程の遷移部分がS字型であったので，本システムでは次式を用いて近似した。

$$\frac{1-cos\frac{\pi}{T}t}{2} \qquad (\text{ただし，} T:\text{遷移時間})$$

ビブラートモデル　ビブラートは一定の音程が続く箇所でも，歌う人が意識して音程を揺らがせる現象であり，文献[2]によると周波数約6Hzの周波数変調である(図3.e)。本システムでは次式の変調をログ軸上で加えることにより，この音程の揺れをモデル化した。ただし，Aは振幅，A_{max}は振幅の最大値，T_Aは振幅が最大値になるまでの時間，Cは変調の周期である。

$$Asin\frac{2\pi}{C}t$$

ただし，

$$A=\begin{cases} A_{max} & t>T_A \\ \dfrac{t}{T_A}A_{max} & 0\le t\le T_A \end{cases}$$

ゆらぎモデル　ゆらぎは一定の音程が続く箇所でも，歌う人の意識とは無関係に音程が揺らぐ現象である。本システムでは一定の音程が続く箇所でピッチ周波数に対し，次式に示す正弦波の和によるゆらぎ[3]をログ軸上で加える。ただし，A_iは振幅，C_iは周期である。

$$\sum_{i=1}^{n} A_i sin\frac{2\pi}{C_i}t$$

3 合唱生成

以下の項目を考慮しながら各パートの歌声を合成した後で，それらの定位を調整し重ね合わせることにより合唱を合成する。

1. 波形データの伸縮により，声質を変換する。
2. 音程を微妙にずらす。

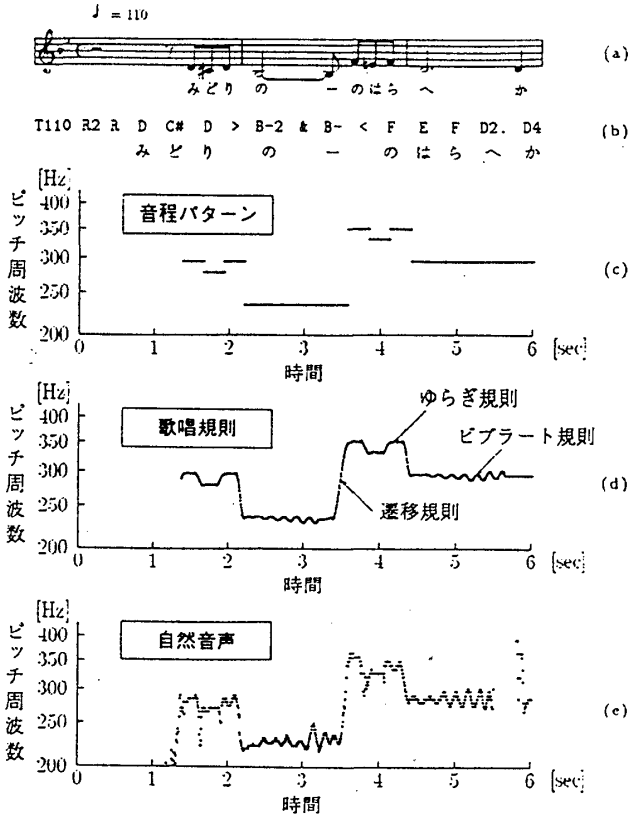

図 3 楽譜と音程の変化

3. タイミングを微妙にずらす。

4 今後の課題

- 声の大きさの制御：波形を相似形に保ったまま振幅を変化させても，音量を大小させたような声となり，人が強弱を付けたような声にならない。
- 高い声での品質の向上：高い声の合成時には合成方式の影響により，波形データの中心付近しか使われないため，情報が落ちている。

上記の問題に対し，声の大小及び高さに応じた波形データを持つことにより解決を図る。

文献

[1] 小林俊一他:「高品質日本語テキスト音声合成システムの開発」,情報処理学会第49回全国大会

[2] 難波精一郎他:「音の科学」,発行 株式会社朝倉書店, 1989年1月20日初版第1刷

[3] Dennis. H. Klatt and Laura C. Klatt "Analysis, synthesis, and perception of voice quality variations among female and male talkers", J. Acoust. Soc. Am.87(2),February 1990