

パソコン向け音声合成ソフトウェア

5V-6

高橋一裕 岩田和彦 三留幸夫 永野敬子 (NEC 情報メディア研究所)
森谷祐一 (NEC 技術情報システム開発)

1 はじめに

近年、パーソナルコンピュータ(以下PC)の性能が急速に向上し、さらにオーディオ信号の入出力装置を備えたものも登場してきた。また、PC上での音声合成への期待も高まってきている。しかし、従来の音声合成システムは信号処理用LSIを搭載した専用ハードウェアを必要とするもの[1]がほとんどであったため、高価なものになっていた。音声合成システムの普及には、性能自体の向上に加えて、低価格化も必要である。

そこで、音声信号生成に演算量の少ない波形編集方式を採用することによって、全ての処理をソフトウェアで行なうテキスト音声合成システムを実現した。専用ハードウェアを必要としないため、従来のものに比べて低コストで利用が可能であり、他のアプリケーションとの結合も容易である。

2 システム概要

システムは、図1に示すように、大きく「テキスト解析部」「韻律制御部」「音声生成部」に分かれている。32bit, 20MHz程度以上の性能を持つCPUとD/Aコンバータ(サンプリング周波数11KHz)とを持つPC上で動作し、システム全体で750Kbyteのメモリ、テキスト解析用辞書1Mbyte、音声データ2Mbyteを必要とする。

また、

- ユーザ辞書のメンテナンスや複数のユーザ辞書間の優先順位の設定
- 読み方(発話テンポ、声の高さ、記号を読む/読まない、など)の設定

などの機能を提供しており、これによって読み上げ方をユーザの好みに合わせることが可能である。

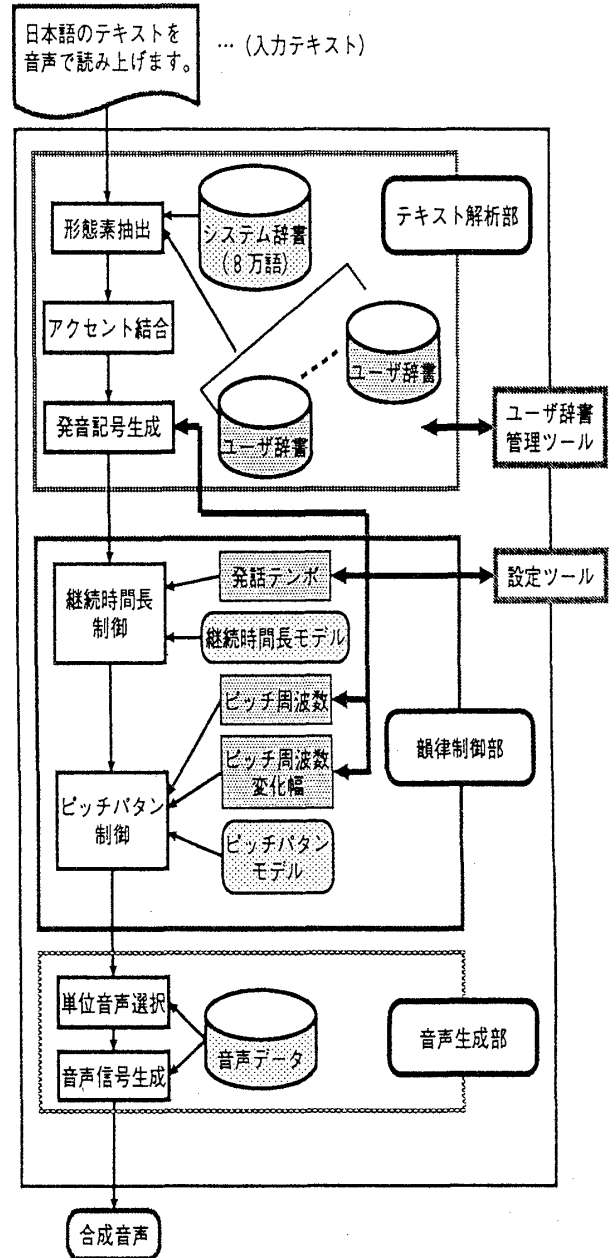


図1. 日本語テキスト音声合成システム

Text-to-Speech Software for Personal Computers
Kazuhiro TAKAHASHI†, Kazuhiko IWATA,†
Yukio MITOME†, Keiko NAGANO†
Yuichi MORIYA†
†Information Technology Research Laboratories, NEC
‡NEC Scientific Information System Development

3 テキスト解析部

形態素抽出には、演算量の少ない左最長一致法を採用した。

辞書としては、8万語のシステム辞書を用意したほか、ユーザ辞書（最大15個）を参照することもできる。ユーザ辞書には複数アクセント句からなる語句を登録することが可能なので、長い語句（例：【老いも若きも】お' いも／わ' かきも）や文などにも正しい読みを与えることが出来る。

又、通常の読み上げでは句読点や括弧などを読み上げる必要はないが、テキストの校正を行なう場合などでは読み上げる方が便利である。そのため、ユーザが句読点や括弧などを読み上げるかどうかを指定することが出来るようにした。

4 韻律制御部

異なる発話テンポで読み上げたときの音声を持つ韻律的特徴の分析結果 [2] に基づいたモデルを採用している。これにより、テキスト情報からの自然なリズムやイントネーションの生成が可能である。

継続時間長制御は、母音中心間長の制御を基本としたモデル [3] に基づいて行なわれ、ユーザの指定する発話テンポに合わせた適切なリズムを生成する。ピッチパタン制御は、正規化ピッチパタンを用いたモデル [4] に基づいている。正規化ピッチパタンは、品詞の並びなどのいくつかの条件を考慮して決定する。

5 音声生成部

音声データとして、男性アナウンサ1名の発声による1600単語の中から抽出された単位音声（約1130音素）を用意した。これらの中から、合成する音素系列と抽出元の音素環境とを考慮して、実際に使用する単位音声を選択する。

音声生成方式は、波形編集方式 [5] を基本として有声音全体のピッチ制御を行なう方式を用いている。これにより、高い明瞭性と滑らかなピッチ変化を実現した。演算量低減のため、予め残差駆動によって生成した長めのピッチ周期程度の時間長を持った素片波形を蓄えておき、これを編集するようにした。単位音声の接続部では、前後の単位音声波形の重み付け加算による補間を行ない、接続部でのスペクトル不連続の影響を緩和させている。

これらの音声信号生成に必要な積和演算は、線

形予測フィルタ演算の1/10程度と少なく、PCのCPUでも実時間の音声生成が可能となった。

6 応用

本システムには、読み上げさせたい漢字仮名交じりの文字列をそのまま入力すればよい。また、PC用マルチタスク環境上のアプリケーションとして作成してある。そのため、他のアプリケーションソフトに、本システムに文字列を転送する機能を付加するだけで、音声を出力することができるようになる。アプリケーション特有の単語がある場合でも、その単語をユーザ辞書に登録し、そのユーザ辞書の優先順位を高く設定しておけば適切に読み上げることができる。

一例として、市販の表計算ソフトにデータ校正用読み上げ機能を付加してみた。このソフトにはユーザが命令を定義できるマクロ機能が付属しており、このマクロ機能を利用することによって、表計算ソフト自体を改造することなく音声出力が実現できた。

7 おわりに

任意の日本語テキストを読み上げるテキスト音声合成システムを、PCのソフトウェアで実現した。今後は、入力文の解析精度の向上、話者（出力音声）の多様化、単位音声の拡張による音質の向上などを図る予定である。

また、音声合成システムへの要望に応えるために必要となる機能・インタフェースなどについても引き続き検討していきたい。

参考文献

- [1] 岩田他：信学技報, SP90-56(1990)
- [2] 岩田他：春季音学講論集, 1-2-21(1992)
- [3] 岩田他：春季音学講論集, 3-8-3(1993)
- [4] K.Iwata, Y.Mitome: Proc. ICSLP-92, pp.1155-1158(1992)
- [5] 三留他：音声研資, S82-06(1982)