合成処理時間が設定可能な文音声合成ソフト

6B−3

原　義幸　新田　恒雄　小林　賢一郎†
(株)東芝　情報処理・機器技術研究所　†東芝ＡＶＥ(株)

1．はじめに

近年、電子メール文の読み上げのように、漢字かな混じり文を音声に変換する「文音声合成(Text-to-Speech;以下ＴＴＳと略す)」技術利用の要求が高まってきている。このような背景のもと著者らは、先にケプストラム方式を用いた文音声合成ボードの試作について報告した[1][2]。

一方、現今のワークステーション(ＷＳ)は、処理能力が向上し(数十〜数百ＭＩＰＳ)、同時に、オーディオデバイス(ＣＯＤＥＣ，スピーカ)を標準で搭載する機種が増えつつある。このようなＷＳを用いると、専用ハードウェアなしにソフトウェアのみでＴＴＳを実行できる。しかし、サーバ／クライアント、あるいはマルチタスク処理環境のもとでは、ＴＴＳの実時間処理が困難となる場合を生ずる。こうした問題に対処するため、処理時間の設定が可能なＴＴＳソフトをＷＳ(AS4075)上に構築したので、概要を述べる。

2．処理の概要

本ソフトは図1に示すように大きく分けて言語処理部と音声合成部から成る。

2．1　言語処理部

入力された漢字かな混じり文は、最初、単語辞書を照合しながら総当たり法によって単語単位に分割され、読み・品詞・アクセント型などの情報に置き換えられる(形態素解析)。次に、読みの変形処理が行われ、続いてアクセント句結合規則により単語をアクセント句に結合してアクセント型を求めるとともに、ポーズ挿入規則に従ってポーズ記号が挿入される(音声記号列生成)。得られた音声記号列(韻律情報と音韻系列)は、音声合成部に送られる。

2．2　音声合成部

ここでは、最初、音韻系列に対応する音声素片(ケプストラムパラメータ)を取り出すとともに、音節時間配置パラメータをもとに音韻長を決定する。次に、補間結合規則に基づいて各素片間を補間することにより、音韻パラメータ列を生成する。一方、韻律パラメータ生成では、アクセント型やポーズ長などの韻律情報と音韻長に基づきピッチパターンが与えられる。

合成器における音源生成は、有声としてインパルス系列を、無声としてＭ−系列を用いた。また、合成フィルタは、ケプストラムパラメータを直接その係数とするＬＭＡ(Log Magnitude Approximation)フィルタ[3]を用いている。

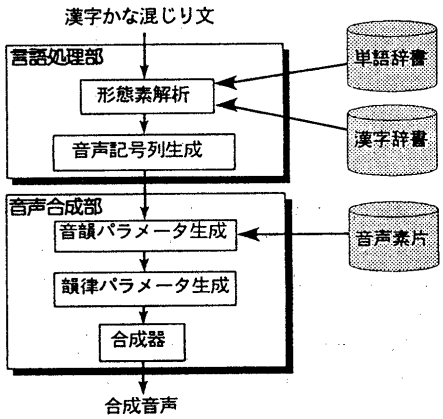

図1　ＴＴＳ処理の流れ

3．実時間処理のための手法

一般にＷＳはマルチタスク処理が可能なＯＳ下で動作している。したがって、ＴＴＳ以外のタスクが増えると、音声生成が実時間で行えなくなり、途中で途切れや異音を生じることになる。同様の問題は、ＴＴＳサーバとしての利用の際にも起こる。こうした問題に対処するため、音声生成に要する時間を可変にする方法と、実時間処理できない場合でも途切れによる聞きづらさを低減する方法の２つを採用した。

3．1　合成次数可変フィルタ

文を音声に変換する処理の中で、最も時間を要するのは合成フィルタ(ＬＭＡフィルタ)である。このフィルタは、サンプリング周期内に

TTS (Text-to-Speech Synthesis) Software with Controllable Processing Time.

Yoshiyuki Hara, Tsuneo Nitta, and Ken-ichiro Kobayashi †
Information Systems Engineering Lab., Toshiba Corp.　† Toshiba AVE Co., Ltd.

数十回の積和演算処理を行っている。したがって、この演算回数を可変にすることによって、音声生成に要する時間を増減することができる。これを実現するには、ＬＭＡフィルタにおける基礎フィルタの修正ｐａｄｅ近似式の次数、つまりフィルタの内部構造を可変にする方法と、合成するときの音韻パラメータの次数を可変にする方法とがある[3]。今回は、簡単に実現できる後者の方法を採用した。

図２に、このフィルタのブロック図を示す。

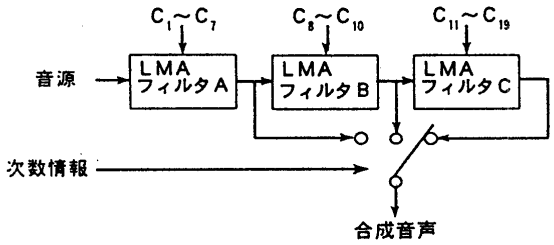

図２　合成次数可変フィルタ

図のように次数情報の入力により、合成次数を変えることが可能である。したがって、演算量を変えることができ、音声生成に要する時間を増減できる。

図３は３通りの合成次数Ｍで生成した音声のスペクトル包絡（「toshiba」）を示している。

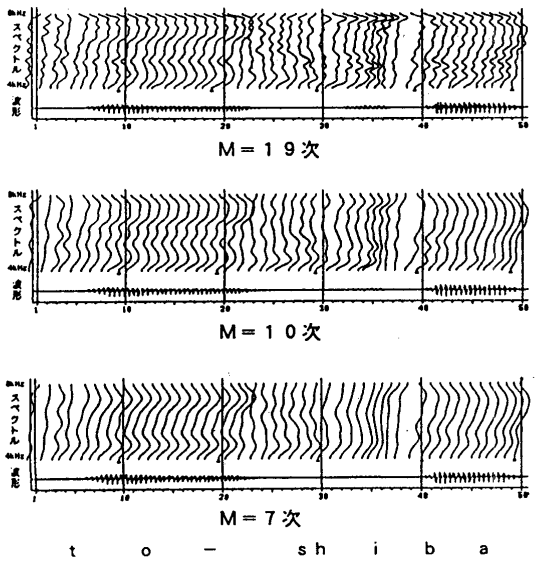

図３　「ｔｏｓｈｉｂａ」のスペクトル包絡

１９次、１０次、７次と、スペクトルの山谷がなだらかになっているのが分かる。音質は、１９次が歯切れの良いのに対して、次第になまった感じの音になるが、７次までは実用に耐えると考えている。

３．２　音声の途切れへの対応

通常、ＴＴＳを専用ハードで処理する際は１サンプルの音声データが生成される毎にＤＡコンバータに転送する方法が採られる。しかし、この方法では実時間処理が不可能な場合、音声の途切れも１サンプル毎になるため、非常に聞きづらくなる。そこで、無音区間を表す記号で挟まれたフレーズの音声データを生成した後、ＣＯＤＥＣに転送する方式を採用した。この方式では実時間に処理できない場合にも、無音の箇所に新たに無音が加わるだけのため、音声途中で途切れることによる聞きづらさが解消される。

４．処理時間の例

０．５秒間の音声を生成するために必要な時間をＡＳ４０７５上で測定したので、その結果を表１に示す。

表１　処理時間

合成次数	処理時間［s］	
	A	B
１９	０．５８	０．７
１０	０．３９	０．４８
７	０．３１	０．３８

A：合成フィルタのみ
B：全処理

この例では、次数が１９次の場合、実時間で処理されていないが、本来はいるべき無音区間が少し長くなるだけのため、文了解に支障はない。

５．まとめ

合成処理時間が設定可能な文音声合成ソフトについて報告した。本ソフトは、電子メール、文章校正支援、ドキュメント読み上げ、音声による通知などに幅広く応用できると思われる。今後は、音質向上とともに応用ソフトの開発を行いたい。

《参考文献》

(1) 原ほか：　“文音声合成ボードの試作”
信学全大, A-231 (1992.3)
(2) 小林ほか：“日本語解析処理を用いた音声読み上げ方式の開発”
情報全大, 7N-6 (1992.3)
(3) 今井：　“対数振幅近似（ＬＭＡ）フィルタ”
信学論(A), J62-A, 12(1980.12)