東芝の音声認識・合成ソフトウエアの紹介

松浦博　正井康之　原義幸　新田恒雄　赤嶺政巳*　瀬戸重宣*
太田治徳**　鈴木孝子**　小林賢一郎++

（株）東芝マルチメディア研　　＊（株）東芝関西研究所
＊＊（株）東芝青梅工場　＋＋東芝ＡＶＥ（株）

１．　はじめに

東芝は95年11月、Windows(R) 95(注1)の日本での発売に合わせてWindows(R) 95パソコン用のテキスト音声変換（ＴＴＳ）による合成ソフトウエア「東芝音声システムVer1.0(1)」をプリインストールの形で発売した。96年 6月には音質を改良した「東芝音声システムVer1.5」、96年11月には音声認識ソフトを加えた「東芝音声システムVer2.0」を発売した。さらに97年６月、合成音質・認識機能を大幅に改善した「東芝音声システムVer2.5」を発売したので、ＴＴＳと音声認識の主な特長を中心に紹介する。

２．テキスト音声変換

「東芝音声システムVer2.5」には、いくつかのアプリケーションが標準搭載されている。そのうち「おしゃべりテキスト」など「おしゃべり」から始まるアプリケーションには、ＴＴＳ技術が使用されている。本ＴＴＳは図１に示すような構成になっており、主な特長として以下の４項目があげられる。

ａ．ＬＰＣ分析残差駆動合成器(2)

従来は、音と音の接続特性（補間特性）が比較的良好なケプストラム方式を使用していた。東芝音声システムVer2.5ではピッチ単位で分解した音声波形（ピッチ波形）を、ＬＰＣ係数と残差波形の対で表現したものを音声素片としている。そして、音声素片からピッチ波形を再生した後、それらを重畳することによって音声を合成する。これにより、原音の情報を欠落させることなく、肉声に近い高品質な音声の合成が可能になった。また音声素片の容量は１話者当たりで150kB と少なく、現在主流となっている波形合成方式に比べて10分の１以下で済む。

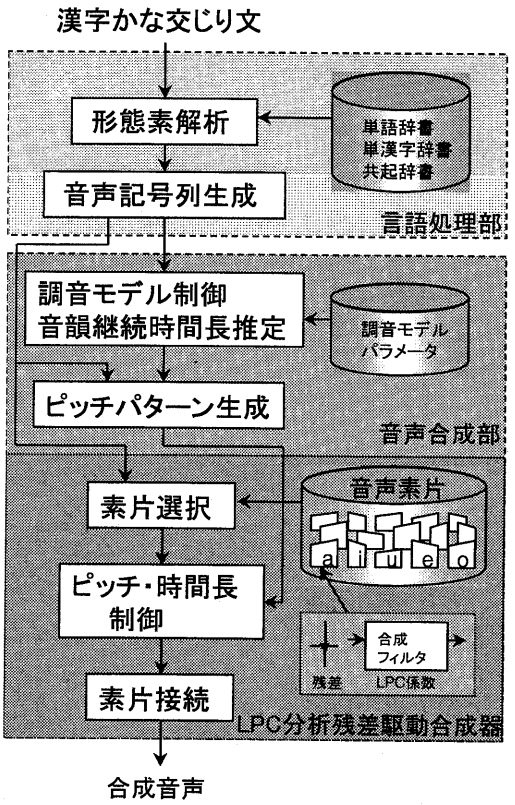

図１　テキスト音声変換の構成

また、音声素片はこれまで人手によって作成していたが、もとになる音声データが膨大になると選択肢が増え、人手では効果的に評価できないため、必ずしも品質の良い合成音とはならなかった。この問題を解決するために「閉ルー

プ学習法」した。本方法は音声データを一度分析・合成した後、合成音声と自然音声との歪みを計算し、様々な音韻環境・ピッチパターンに対して歪みが最小となるように素片を選択する。

b．調音モデルに基づく音韻長制御[3]

合成音をより自然にするためには、人間の音声生成過程における調音器官の動きの制約を考慮して、個々の音の長さを適切に制御する必要がある。そこで舌や顎などの調音器官の動きをパラメータ化し、そのパラメータを音声データベースを使って学習する。このパラメータの制限のもとで音韻継続時間長を制御することによって、合成音は人間の発声により近いものとなった。

c．声質制御

男女の音声素片をもとに一部の合成パラメータを変更することによって、お嬢ちゃん、お坊っちゃん、おじいさん、おばあさん、ロボットあるいはアニメのキャラクタのように、ユーザの好みの音声を合成することができる。合成パラメータの変更は「おしゃべりキャラクタ」というユーティリティーでユーザが簡単に行える。

d．地名の同字異音語の読み分け

日本語には、漢字が同じでありながら違う読み方をする場合がある。例えば「公園を通って学校へ通った。」という文では「通って」は「とおって」と読み、「通った」は「かよった」と読む。東芝音声システムVer2.0では、同字異音語の読み分け機能によって、これを正しく読み分けることができる。東芝音声システムVer2.5では、これに加えて地名の同字異音語についても対応した。例えば「彼女は神奈川の山北町から新潟の山北町へ嫁いだ」という文では、前者の「山北町」は「やまきたまち」、後者の「山北町」は「さんぽくまち」と正しく読み分けられる。

3．音声認識[4]

「東芝音声システム」中で「おきらく」が付くアプリケーションは認識機能を使っている。「おきらくコマンド」は、音声によるWindows操作を提供する。

「おきらくミミ」は、アニメーションで作成されたうさぎのキャラクタ「ミミ」と対話するアプリケーションである。例えば、「こんにちは」と話かけると「こんにちはミミです。よろしくね」のように「ミミ」のキャラクタに合った声質の合成音で応答し、おじぎの動作をする。動作は5種類の中から選ぶことになっているが、応答音はＴＴＳ機能を使っているので、ユーザが自由に設定できる。例えば「こんにちはゆうちゃん」ようにユーザの名前を呼ばせることもできる。

音声認識には不特定話者大語彙認識に実績のある統計的認識手法を基に、少ない演算量・メモリでも性能を維持できる次元圧縮手法を開発している。認識語彙はキーボードから入力すれば良いので、ユーザは話かける言葉を自由に設定できる。さらに設定した認識単語「おはよう」の前後に不要語がついた「えーおはようございます」のような発声にも対応している。本アプリケーションによって、一般ユーザには馴染みの薄い音声認識やＴＴＳに親しんでもらうとともに、コンピュータとの対話を体感してもらうことができた。

東芝音声システムはサウンドチップを備えたWindows(R) 95搭載のデスクトップタイプ、ノートブックタイプのパソコンにプリインストールされている。特にノートブックタイプは携帯機器として使用されることが多いので、外部マイクロホンだけでなく内蔵マイクロホンに対しても十分な認識性能を達成するように配慮した。また一部機種では学習に用いたマイクロホンとは特性が若干異なるため、入力データに対してマイクロホン特性を補正するような処理を行って認識性能を改善している。

文献 (1)桃崎ほか，音学講論2-4-10(96.9) (2)赤嶺ほか，97-SLP17-16(97.7) (3)志賀ほか，音学講論1-7-7(97.3)(4)正井ほか，信学総全大A-15-22(97.3)(注1)Windows(R) はMicrosoft社の登録商標