東芝における最近の音声合成・認識の応用

松浦博　正井康之　田中信一　原義幸　桃崎浩平　新田恒雄　*太田治徳　**小林賢一郎
（株）東芝　マルチメディア技術研究所　＊青梅工場　＊＊東芝ＡＶＥ（株）

１．東芝音声システム(1)

東芝音声システムは日本語 Windows(注1) 95で動作し、ＴＴＳを用いたユーザインタフェースを提供するソフトウェア環境である。本システムの主な特長を以下にあげる。

・直交化残差方式とケプストラム合成方式によって音質を向上させている。

・約12万語の単語辞書、２千語登録可能なユーザ辞書により正確なアクセントを付ける。

・知識処理を用いた解析によって、「質の良い宝石を質に入れる」の「質（しつ）」と「質（しち）」のような同字異音語を読み分ける。

・豊富なアプリケーションを備えている。

本システムは図１に示すように主にユーザインタフェース部とＴＴＳエンジンからなる。

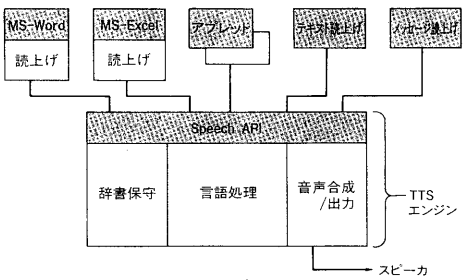

図１　東芝音声システムの構成

ユーザインタフェース部はユーザがＴＴＳを利用するためのツール群である。現状ではＴＴＳを標準にサポートしている市販のアプリケーションが少ないため、独自に作成した。ＴＴＳエンジンはＴＴＳ処理を行うモジュールであり、それと共に上位ソフトウェアに対して、ＴＴＳ機能を使用するためのＡＰＩ（Application Programming Interface)を提供する。本ＡＰＩは Microsoft(注2) 社が提案するSpeechＡＰＩのサブセットになっている。

１．１　ＴＴＳエンジン

ＴＴＳエンジンは、実際のＴＴＳ処理を行う文音声合成エンジン部と、ＡＰＩを提供するSpeechＡＰＩ部の２層からなる。文音声合成エンジン部は、言語処理、音声合成／出力および辞書保守の各処理を実行するモジュールで、32ビットDLL(Dynamic Link Library)になっている。

SpeechＡＰＩ部は、文音声合成エンジン部の機能をユーザに提供するためのインタフェースモジュールである。ユーザはこのＡＰＩを介してだけ文音声合成エンジンの機能を使用することができる。テキストあるいは音声記号列から音声への変換、音声出力の停止／再開、音声属性の各種設定、ユーザ辞書登録などの機能を提供する。ＡＰＩはＯＬＥ２（Object Linking and Embedding）に基づいて定義されており、ＡＰＩ部の上層に実装されたラッパーによって、ＴＴＳエンジンをオブジェクトとして扱うことができる。

１．２　ユーザインタフェース部

ＴＴＳを利用するための支援ツール群として以下のものを提供する。

(1) テキスト読上げ

アプリケーションプログラムであり、SpeechＡＰＩを使用して、テキストファイルの読上げ、テキストファイルの作成、クリップボードの読上げ、ＯＬＥサーバ機能の動作を行う。ＯＬＥサーバ機能とは読上げ文書の埋込みおよびリンクのサポートである。さらに、読み上げ音声の種類の設定、読み上げモードの各種設定、ユーザ辞書登録／削除／更新／検索を行う。

(2) アプレット

入力したキーおよび計算結果を読上げる音声電卓アプレット、時刻を読み上げる音声時

刻アプレット、キーボードから入力されたコードを読み上げるキーボード読上げアプレットがある。

(3) ＭＳ－Ｅｘｃｅｌ読上げ

Microsoft Excel のデータを読み上げるためのマクロ機能を提供する。

(4) ＭＳ－Ｗｏｒｄ読上げ

Microsoft Wordの文書を読み上げるためのマクロ機能を提供する。

(5) メッセージ読上げ

システムの状態を常時監視して、システムからのエラーメッセージを読み上げたり、システムの開始／終了時などに特定のメッセージを読み上げるプログラムである。

2．音声認識を利用した地図検索システム

音声認識を利用するメリットとして、多項目からの検索が迅速かつ容易に行えることがあげられる。我々は警察や消防の通信指令室などで使用される、目的地の周辺地図を音声認識を用いて迅速に検索する地図検索システムを開発した。今回、作成したシステムには地名（字名まで）３万４２３０件、目標物名７万４９３７件、公衆電話ボックス名２３４７件の計１１万１５１４件が登録されている。このように大規模なシステムを構築できるようになった理由の一つは、単語の読みを入力するだけで、任意の単語を認識できるようになったためである。

2．1　本システムの操作例

地名「福島市森合字森下」を検索する場合を例にとって、本システムの操作方法を説明する。まず「フクシマ」を発声すると、「森合」などの福島市の大字名が表示される。次に「モリアイ」と発声すると、「森下」などの森合に属する小字名が表示される。さらに、「モリシタ」と発声すると、目的地が特定され相当する地図が表示される。

目標物「東芝福島工場」を検索するに場合について説明する。まず市町村を確定し、「カイシャ」などの分類名を発声し決定する。次に目標物名の「トウシバフクシマコウジョウ」を発声する。会社名の場合、３千単語を越えるような大語彙となるが、本システムではこれを一括で認識する。認識単語数を大きく増加させることによって、地図検索など多くの項目を入力する必要があるシステムへの音声認識の応用が可能になった。さらに、「市町村－大字－（小字）」、「市町村－分類－目標物」のように認識対象を階層化することよって、総検索単語数を十万件以上にすることが可能となった。ここで、括弧で囲われた項目は存在しない場合があることを示す。また、公衆電話ボックスは分類の一つとして扱っている。

2．2　システム構成

本システムは図２に示すように音声認識装置とワークステーションから構成されている。

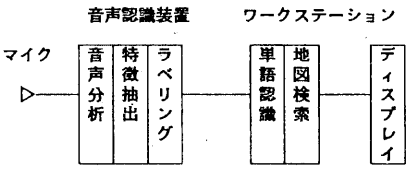

図２　音声認識地図検索システムの構成

音声認識装置は音声分析、特徴抽出、ラベリングを行う。一方、ワークステーションは単語単位の認識を行い５位までの認識結果を求めディスプレイに表示させると共に、目的地が特定された場合、１位の結果に対する地図を表示する地図検索ソフトを備えている。ここで、２位から５位の認識結果をマウスの操作によって指定すれば、対応した目的地の地図が改めて表示される。

文献(1) 太田治徳、高橋勉、原義幸：「ﾊﾟｿｺﾝにおける文音声合成を利用したﾋｭｰﾏﾝｲﾝﾀﾌｪｰｽ」東芝ﾚﾋﾞｭｰ, pp14-17, Vol. 51, No. 1(1996)

（注１）Windows はMicrosoft 社の商標。

（注２）Microsoft はMicrosoft 社の商標。