

## 東芝の音声認識製品化への取り組み

株式会社 東芝 研究開発センター  
松浦 博 金沢博史

### Toshiba's Approach to Speech Recognition Systems Producing

TOSHIBA Corp., Corporate Research and Development Center  
Hiroshi Matsuura, Hiroshi Kanazawa

#### 1. はじめに

東芝は以下の三つの基本的な考え方に基づき、音声処理技術の研究実用化(製品化)を行っている。

音声は対話手段である  
実環境で使える

人間が常に中心である

我々は、音声を単なる入出力手段として捉えるのではなく、機械(コンピュータ)との対話手段と考えている。すなわち、音声から文字および文字から音声への変換を行うだけでなく、人間の意図を正しく伝達することが重要であると考えている。また、“いつでも、どこでも、だれでも”をキーワードに実環境でも快適に使える音声処理技術の開発に注力している。さらに、ユーザのニーズを重視し、人間を中心に考えた自然なコンピュータとの対話の実現を目指している[1]。

#### 2. 最近の東芝の音声認識製品の紹介[2]

東芝の製品のうち「東芝音声システム」とマイコン用ミドルウェアを紹介する。

##### 2.1 東芝音声システム

「東芝音声システム」は、Windows® 95 および Windows® 98 に対応したパソコン用テキスト音声認識/テキスト音声合成ソフトウェアであり、当社のパソコンにプレインストール販売されている[3]。

1995年11月、Windows® 95 の日本での発売に合わせてテキスト音声合成(TTS:Text-To-Speech)ソフト「東芝音声システム Ver1.0」[3]をプレインストールの形で発売したのに始まり、1996年11月には音声認識機能を加えて「同 Ver2.0」[3]を、1997年6月には合成音質と認識機能を大幅に改善した「同 Ver2.5」を発売した。また、「東芝

音声システム Ver4.0」が、東芝ダイレクトPC(国内モデル販売)にてインターネット通信販売された。2000年6月時点では、ディクテーション機能を備えた「東芝音声システム Ver5.5」がプレインストールされている。

「東芝音声システム」には、孤立単語認識エンジン、キーワードスポッティングエンジン、連続単語認識エンジン、連続音声認識エンジンの4種類の不特定話者認識エンジンが含まれる。また、これらのエンジンを用いたいくつかのアプリケーションが提供されている。認識語彙は単語の読みを平仮名で入力することによって自由に設定可能であり、登録のための発声は一切不要である。

##### (1) 孤立単語認識エンジン

音声認識の基本機能である孤立単語音声認識は、認識が初めて搭載された Ver2.0 からサポートしている。当初は最大アクティブ(同時認識可能)単語数が500程度であったが、現在では50000に増やしている。アプリケーションの起動や操作を行う「おきらくコマンド」はマイク入力、「きこえるメール」は電話音声入力のアプリケーションである。

##### (2) キーワードスポッティングエンジン

キーワードスポッティングエンジンは、キーワードの前後に「えーと」などの不要な言葉が付加した音声も認識する。このエンジンを用いれば、エージェントなどとの対話を実現することができる。本エンジンは、ユーザの語りかけを認識してアニメーションのうさぎ「ミミ」が動きながらおしゃべりするアプリケーション「おきらくミミ」に使用されている。本アプリケーションは Ver2.5 から搭載されており、一般ユーザには馴染みの薄かった音声認識やテキスト

音声合成 (TTS) に親しんでもらうことを目的としている。近ごろ、はやっている対話型アプリケーションの先駆けと言える。

### (3) 連続単語認識 (CFG) エンジン

連続単語認識エンジンは CFG(Context Free Grammar) で表現した文を認識する。連続数字入力や経路入力 (〇〇から△△まで) などのように組合せが多い文を簡単な文法で表現し、簡単に認識対象として設定できる。本エンジンは、Microsoft® Agent 「スーパーミミ」への指示の伝達などに使用されている。

### (4) 連続音声認識 (ディクテーション) エンジン

連続音声認識エンジンは 10 万 5 千語からなる豊富な語彙を持ち、発声した言葉を仮名漢字混じりの文章に自動変換するディクテーション機能を提供する。誰でも簡単にディクテーション機能を体験できるようにするため、完全な不特定話者認識方式とし、使用に先立って面倒なトレーニング (話者適応) の必要は一切なく、ユーザのニーズにも合致している。また、社会システム機器などパーソナルユース以外への展開も可能となっている。

さらに、辞書登録機能を使用して人名や会社名など、ユーザやタスクの必要に応じた単語を簡単に追加登録することができる。辞書登録は平仮名で読みを登録するだけで良く、登録のための発声を必要としない。本エンジンは、ディクテーションおよび TTS のアプリケーション「おしゃべりテキスト」に使用されている。また、Microsoft® IME 2000 の「IME パッド 音声入力アプレット」を使用すれば、様々なアプリケーションに音声で日本語入力することも可能である。

## 2.2 東芝マイコン向け認識ミドルウェア

頑健な認識ミドルウェアを開発し東芝マイコン TX 39 用に製品化した。

### (1) TMW39-SR-A (汎用版ミドルウェア)

TMW39-SR-A は 50MIPS 程度以上の MPU を対象としている。不特定話者単語音声を認識するので話者の声の登録が不要であり、語

彙登録のみで良い。語彙登録は平仮名で 31 文字以内で入力すれば良い。なお、最大発声長は 4 秒で、同時に認識する最大単語数は 1000 語 (1 セット当たり) である。単語セットは最大 100 セットまで切替え可能である。また、ノイズキャンセル機能として環境適応辞書によって S/N 比 20dB 程度までの定常ノイズに対応することができる。入力モードとしては発声前にボタンなどを押す発声通知モードと、その必要のないハンズフリーモードがある。必要メモリは RAM 2.7MB、ROM 1.9MB である。なお、平均処理時間は評価ボード R3904@ 50MHz にて計測した結果で 0.7 sec 以内であった。特に、本ミドルウェアはカーナビゲーションをはじめ多様な実環境での使用に対応する。

### (2) TMW39-SR-B (コンパクト版ミドルウェア)

TMW39-SR-B は現在、製品化手続き中であり、40MIPS 程度の MPU で動作する。当社の省電力タイプのマイコン TX 19 用としても使用可能である。コンパクト版は汎用版に対して必要メモリを RAM 18kB、ROM 100kB 強に押さえていることが特長で、認識単語数は 100 語である。本ミドルウェアは携帯機器への使用などに適している。

## 3. おわりに

東芝の代表的な音声認識製品の紹介を通して、ユーザのニーズを重視した製品化への取り組みについて紹介した。

### 参考文献

- [1] 竹林洋一、「デジタルメディア時代のヒューマンインタフェースの研究」、信学論、vol.82, no.1, pp.22-27, 1999
- [2] 正井康之、桃崎浩平、松浦博、田中信一、鈴木孝子、赤嶺政巳、山中紀子、原義幸、「東芝パソコンの音声認識/合成ソフトウェア」、信学全大、2000.3
- [3] 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、情処学会研究報告 97-SLP-17-15, 1997.3

(注) Windows、Microsoft は米国 Microsoft Corporation の商標または登録商標です。