

## パソコン向け音声認識合成プラットフォームの 構築とアプリケーションの試作

7N-9

近藤 玲史 岩田 和彦 磯 健一 畑崎 香一郎 三留 幸夫  
渡辺 隆夫 (NEC) 水野 正典 (NEC 情報システムズ)

### 1. はじめに

パソコン上のユーザアプリケーションにおいて、音声認識および音声合成の機能を利用すると、よりユーザフレンドリなヒューマンインタフェースを構築することができる。我々はこれまで、これらの機能を手軽に利用できるよう、特定のアプリケーションに依存しない形で動作する音声認識合成プラットフォームの実現を行ってきた [1][2]。

現在、パソコン用 32bit マルチタスク OS が普及してきている。そこで、この機能を生かした新しい音声認識合成プラットフォームを開発し、その上で動作するアプリケーションの試作を行ったので報告する。

### 2. 音声認識合成プラットフォームとエンジン

#### 2.1 プラットフォーム

音声認識合成の機能を提供するプラットフォームを用いることで、アプリケーションに同機能を付加することが容易に可能となる。このプラットフォームとして、音声認識・合成エンジンおよび API を提供する (図 1)。

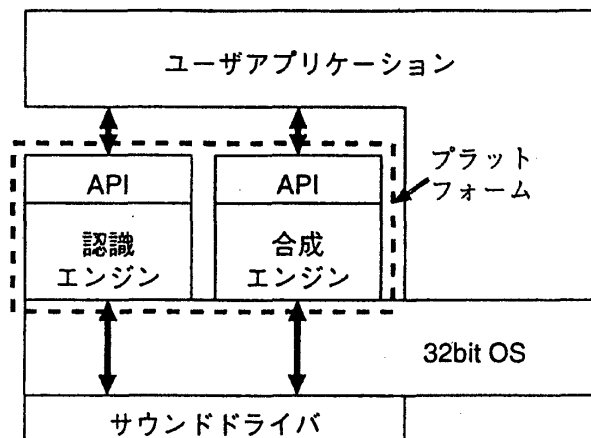


図 1. 音声認識合成プラットフォーム

本プラットフォームにおける API は、OS の定義する標準的な呼び出し方法に従っており、プロ

Speech Recognition and Synthesis Platform and Applications for Personal Computers.

KONDO Reishi, IWATA Kazuhiko, ISO Ken-ichi, HATAZAKI Kaichiro, MITOME Yukio, WATANABE Takao (NEC Corporation) MIZUNO Masanori (NEC Informatem Systems)

表 1. 音声認識合成エンジンの基本仕様

音声合成エンジン	
波形生成	波形編集方式
話者	男、女 (通常、老人、ロボット)
標本化周波数	11.025kHz, 8kHz
設定機能	テンポ、ピッチ、アクセント 話者、読みの区切り文字 数字・記号の読み方
ユーザ辞書数	最大 16
音声認識エンジン	
話者	不特定話者 (学習可能)
認識対象	連続単語
認識単語数	最大 200
辞書登録	テキストベース

グラミング上の特別な習熟の負担が少ない。

#### 2.2 音声認識合成エンジン

本エンジンの基本仕様を表 1 に示す。

音声合成エンジンは、発声速度に依存しない韻律的特徴を用いることで自然なイントネーションやリズムを実現している。また、音声生成においては、波形編集方式を用いて少ない演算量で高い明瞭性を得ている [1]。

音声認識エンジンは、独自の半音節モデルを用いた混合ガウス分布 HMM により、不特定話者の連続単語認識が可能となっている。また話者適応を行い、高い認識率を得ている [2]。

さらに、本エンジンでは以下のような機能を強化している。

- OS の標準的な呼び出し手段に従った API
- マルチタスク・マルチスレッド対応

本エンジンは、複数のアプリケーションから同時にロードされてマルチタスクで動作することができる。例えば、あるアプリケーションがパソコンに繋いだスピーカとマイクロフォンを使って音声対話を行っているのと同時に、システムのエージェントが電話回線に応答メッセージを送出するなどの使い方が可能である。

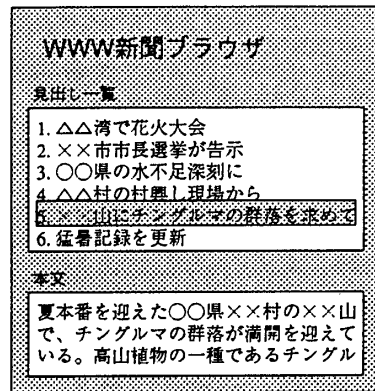
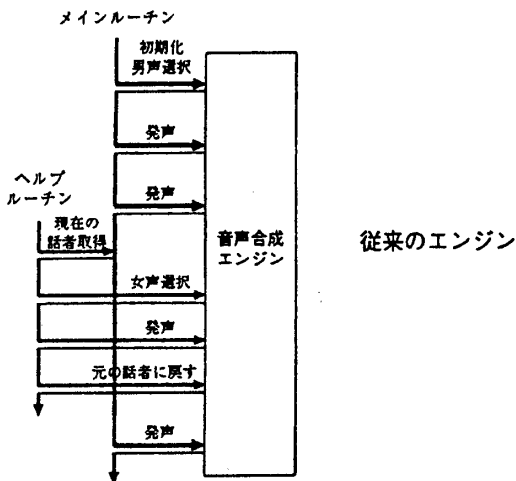


図 3. 試作したアプリケーション例 (概念図)

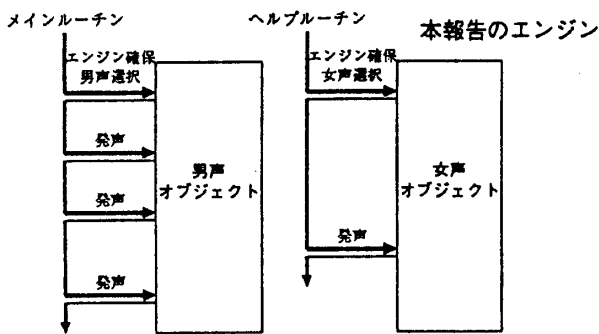


図 2. 複数オブジェクトの利用例

また、例えばあるアプリケーションにおいて、通常は男声の音声合成音を出力して女声のヘルプメッセージが任意の時点で割り込む場合、従来のエンジンではヘルプの割り込み開始時点での話者情報を保存し、話者を変更してヘルプを女声で発声し、その後また話者を元に戻さねばならず、複雑であった(図2)。本エンジンにおいては、アプリケーションの初期化時に男声および女声の音声合成エンジンのインスタンスを、男声オブジェクト・女声オブジェクトとして生成する。その後、男声で発声したい場面では男声オブジェクトに、女声で発声したい場面では女声オブジェクトにテキストを送ることで、直前の発声に影響されずに指定した声質(話者、ピッチなど)で発声を行うことができる。

### 3. アプリケーションの試作

音声認識合成プラットフォームを用いて容易にユーザアプリケーションにこれらの機能を組み込むことができる。以下のような簡単なアプリケーションを試作した。

新幹線検索 新幹線の発着駅を連続音声認識し、時刻と料金を検索して音声合成で応答する。

WWW 新聞ブラウザ WWW の新聞情報にアクセスして見出しを取り出し、音声認識で任意

の見出しを選択。選択された本文を、音声合成で読み上げる(図3)。

テキスト読み上げ テキストファイルの内容や、ユーザがキーボードから入力したテキストを音声合成で読み上げる。

これらの試作により、音声認識・音声合成とも数十行のコードを付加することで各機能を使えることが判った。更に、それらのコードの多くの部分は、エンジンの動作終了などの通知を受け取るための部分であってほぼ定型的に書くことができる。これにより、本プラットフォームを用いることで容易に音声認識・合成機能をアプリケーションに組み込むことができることを示した。

### 4. まとめ

パソコン上の 32bit マルチタスク OS 環境において、その機能を生かして動作する音声認識合成プラットフォームを開発した。マルチタスクやマルチスレッドに対応し、複数アプリケーションにおける同時使用や、一つのアプリケーション内で複数のインスタンスを確保して動作することが可能である。

また、このプラットフォームの上で動作するアプリケーションを試作した。音声合成・音声認識とも、数十行のコードを記述することで機能を取り入れることができ、本プラットフォームの有効性を示した。

今後は、プラットフォームの機能強化、および性能向上を図る予定である。

### 文献

- [1] 高橋ほか「パソコン向け音声合成ソフトウェア」情報処理学会第47回全国大会, 2-377 (1993)
- [2] 篠田ほか「パソコンソフト連続音声認識」情報処理学会第50回全国大会, 2-465 (1995)