NEW PRODUCTS 3

V850音声認識ミドルウエア ULTALKER-C

横　溝　隆　司★

はじめに

人が発声した音声を機械が認識する音声認識技術は，20～30年前からすでに研究が開始されています。最近では認識できる単語数が数千から数万へと増え，あらかじめ音声を登録しなくても認識することができるようになってきました。またシステム規模に関しても，演算量低減化とデバイス処理能力の飛躍的向上により，小型化が可能となってきました。このような背景のもと，近年，カーナビゲーションシステムや携帯電話などを中心に，音声認識技術の応用が進みつつあります。

これらの分野における音声認識技術の展開を受け，これまでコストなどが原因で導入が妨げられていた家電や玩具などの分野からローコスト音声認識の需要が増えてきました。

本稿では，これらの分野へ適した当社オリジナルのRISCマイコンであるV850ファミリMCU向けに開発を行った音声認識ミドルウエア「ULTALKER-C」の紹介と今後の展開について解説します。

当社の音声認識技術

当社では，古くから音声認識技術の開発を行っています。そのなかでも，ハイエンド組み込み分野向けに開発したV830，VRファミリ上で動作する音声認識ミドルウエア製品であるULTALKER-Vに搭載されている技術について説明します。

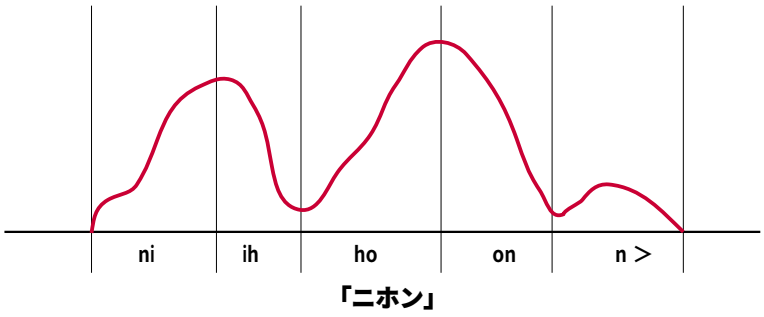


図1　半音節

おもな特長は以下の通りです。

- 超大語彙認識
  アルゴリズム上の認識語彙数に制限を設けていないので，市町村名まで含めると10万単語以上にもなる全国地名などを1回の発声で認識することができます。
- 高速応答
  従来の音声認識では，認識語彙数が増えると認識結果が出力されるまでの応答時間が長くなっていましたが，本製品では，認識語彙数に関わらず即時に認識結果が得られます。
- 不特定話者単語認識
  事前の発声登録を必要とせず，単語単位で発声された不特定な話者の音声を認識することができます。
- 簡単な認識対象語彙の設定
  単語の読みを記述するだけで認識語彙の設定ができます。よって，認識語彙の追加，変更を簡単に行うことができます。
- 高い耐ノイズ性能
  走行中の自動車内などの騒音環境でも使用できます。

この音声認識技術は，当社独自の半音節音声認識方式をベースにしています。半音節とは音節をその母音中心で2つに分割したものです（**図1参照**）。単語をこの半音節の組み合わせで表現するので，単語の読みだけで認識語彙を設定できます。また，演算量の大幅な低減のために，認識辞書をツリー状に構成し効率的にパタンマッチングを実現できる当社独自のダイナミック・ツリー・サーチ法を採用しています

★ 半導体ソリューション技術本部 応用ソフトウエア技術部

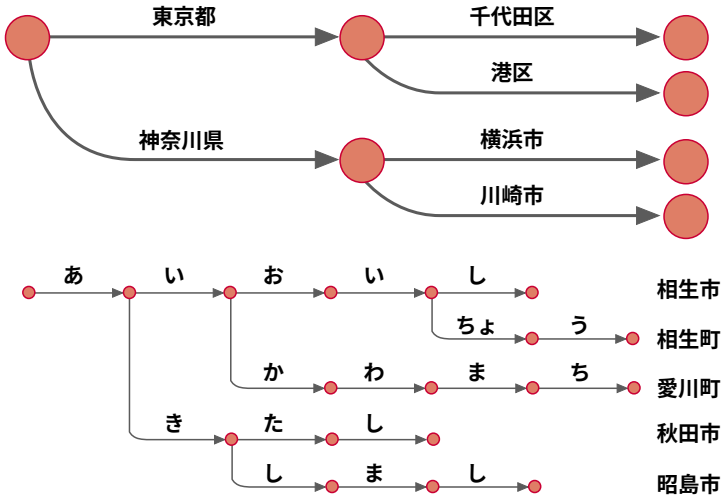


図2　ダイナミック・ツリー・サーチ

（**図2参照**）。これにより高速応答を実現しています。

本音声認識技術の構成を図示したものが**図3**です。多数話者の音声を用いた学習により，あらかじめ半音節単位のモデルを作成します。多数話者の音声を用いることにより，さまざまな音声の特徴が学習でき，不特定話者の発声にも対応できるようになります。学習のために発声する単語は，認識対象単語と同じである必要はありません。半音節単位に作成されたモデルと，単語の読みから作成された認識辞書を用いて認識を行います。

V850音声認識［ULTALKER-C］

先に説明したULTALKER-Vに対して，V850/V850Eファミリ上で動作させたいとのマーケットからの声に応え，V850ファミリMCU向けにULTALKER-Cを開発しました。今回，ULTALKER-Cの開発に先立ち想定したマーケットは，おもに「モバイル系」「家電系」「玩具系」です。

「モバイル系」は，携帯電話をはじめとする携帯端末の類で，実際にいくつかの製品にはすでに音声認識が搭載されはじめています。小型化を行うと同時に使いやすさも確保しなければならない分野に対し，音声認識による操作性の向上が期待されます。

「家電系」は，おもにリモコンなどへの導入を想定しています。エアコンの温度調整など人間の感覚に依存する操作や，高機能化し操作が複雑になってきているAV機器などの操作に，音声認識を用いた自然言語操作を用いることで容易な機器操作が実現できると考えています。

「玩具系」は，TVゲームですでに存在している音声認識を用いたキャラクタとの対話をはじめ，音声認識によるロボットなどの玩具の操作などが考えられます。

これらのマーケットに共通した音声認識への要求は「コスト低減」であり，ULTALKER-Cの開発にあたってはアプリ

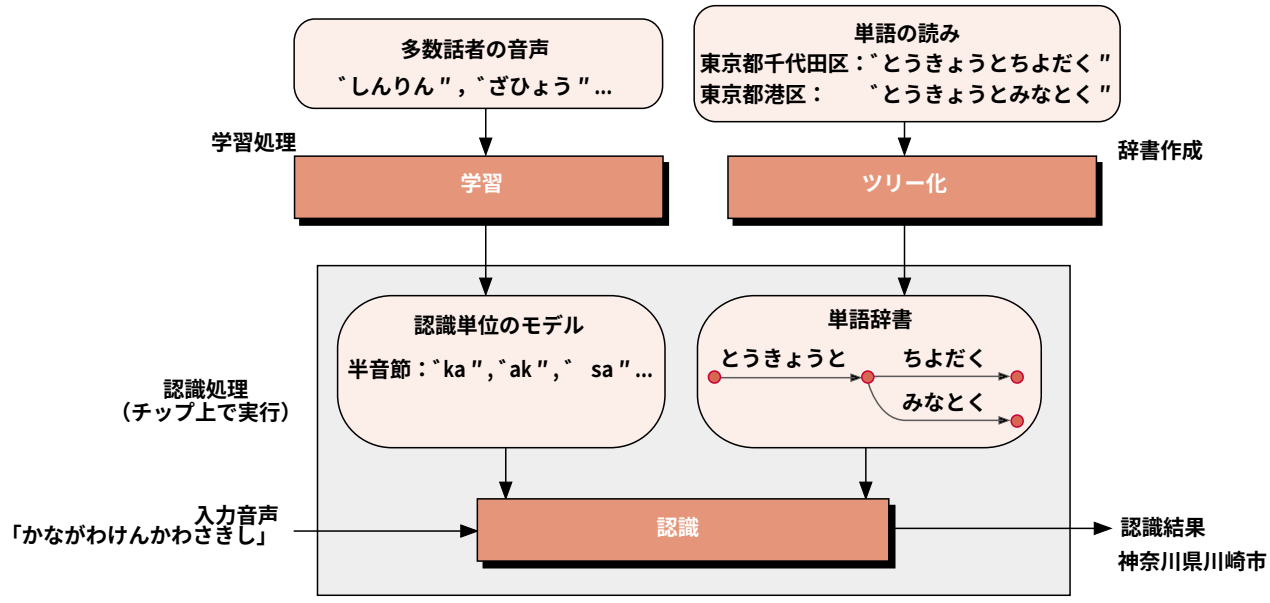


図3　音声認識の構成

ケーションシステムのコスト削減に寄与することを最優先として仕様化を行いました。以下に, ULTALKER-Cのおもな特長を示します。

1. 不特定話者音声認識

ULTALKER-CもULTALKER-Vと同様, 事前の発声登録を必要とせず, 単語単位で発声された不特定な話者の音声を認識できます。他社の廉価版の音声認識では, 特定話者方式を採用しているため, 事前に音声認識メーカに対し話者データを提出し, 学習データを作成しなければならないなどのコストが発生するものもありますが, ULTALKER-Cではこのようなコストを削減することができます。

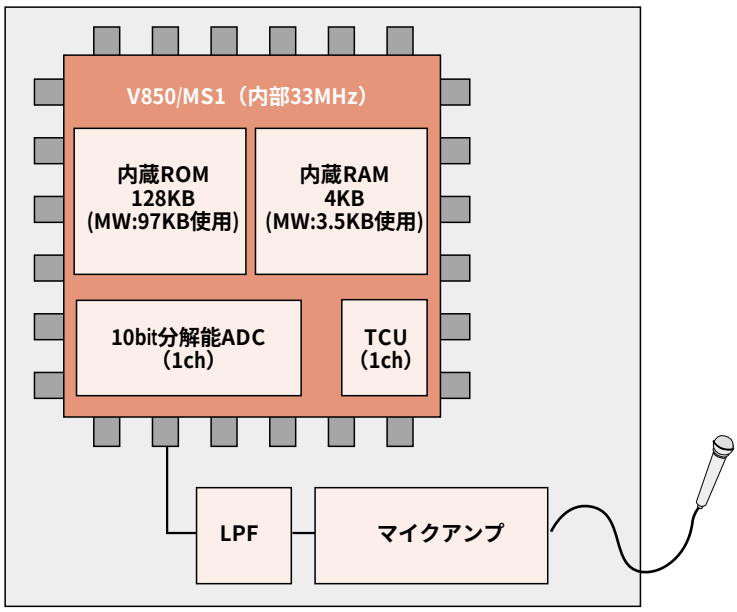
2. 外部ROM/RAM不要

V850ファミリMCUに搭載されている内蔵ROM/RAMだけで動作することを前提に設計を行っており, 外部に別途ROM/RAMを追加する必要はありません。(一部少メモリ品種上で動作させる場合を除く。) ULTALKER-Cを動作させるのに必要なROM/RAMを**表1**に示します。

3. 内蔵A/Dコンバータを使用可能

A/Dコンバータが内蔵されているMCUを使用する場合, 内蔵されている10ビット精度のA/Dコンバータを入力音声の量子化に使用できます。ですから音声用のCODECデバイスを別途システムに追加する必要がありません。

ULTALKER-Cでは, 以上のような特長を有しています。**表2**にULTALKER-Cで実装されている機能(関数)を示します。



設定条件

認識単語数：15単語
プログラム・テーブル配置：内蔵ROM
辞書配置　　：内蔵ROM
ワーク・スタック配置　：内蔵RAM

LPF：カットオフ周波数4KHz（8KHzサンプリング時）
マイクアンプ：ゲイン 約10倍

図4　システムの構成図

このULTALKER-Cを用いて音声認識システムを構築する場合のシステム構成例を**図4**に示します。この構成では, マイクから入力された音声をアナログマイクアンプで増幅し, LPFを通した音声信号をV850E/MS1内蔵のA/Dコンバータへ入力しています。ここで使用しているアナログ回路は, **図5**に示すような非常に簡単な回路です。この例ではA/Dコンバータにて量

容量	ROM	プログラム	約28KB
		テーブル	約63KB
		辞書（認識単語数＝15）	約0.3KB※1
	RAM	固定ワークサイズ	約2.0KB
		変動ワークサイズ	66×N バイト※2
		スタック	約0.4KB
CPU占有率※3		5単語認識　約43%	
		10単語認識　約53%	
		15単語認識　約63%	

※1　標準辞書サイズは1単位あたりの平均文字数を5文字とした場合の値です。
※2　変動ワークサイズは認識単語数 N に比例する。
※3　V850E/MS1（43MIPS）によるマッピングは内蔵ROM/RAMのみ使用。

表1　必要なROM/RAM容量とCPU占有率

関数名	分類	機能
vrg_Recog vrg_Stop vrg_InputWave	認識処理	1発声分の認識処理を指定された認識辞書を用いて行う vrg_Recogの処理を中断する A/Dコンバータにより量子化された音声データをvrg_Recogに引き渡す
vrg_MakeDic	認識辞書管理	指定された単語リストから音声認識辞書を作成する

表2 実装されている関数

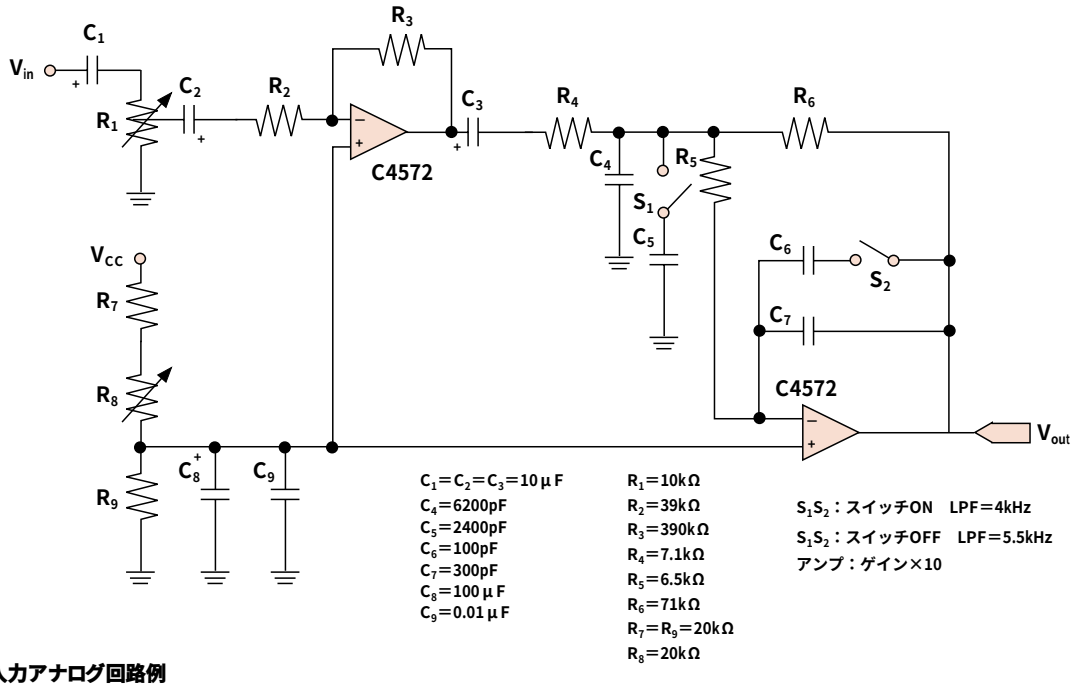


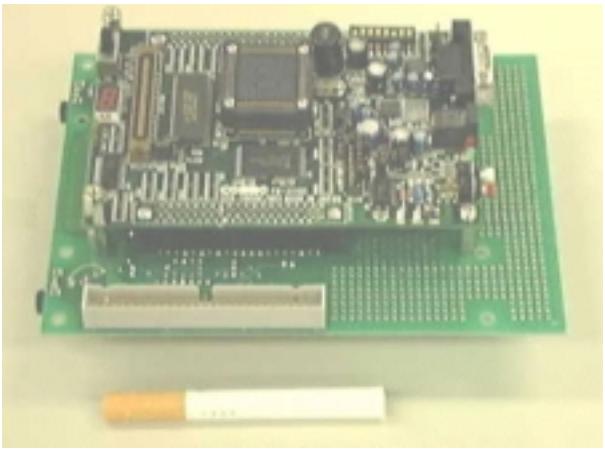
図5　音声入力アナログ回路例

子化された音声データの取得には，TCUにてサンプリング周波数に同期したタイマ割り込み（8KHzサンプリングの場合125μs毎）を使用することを前提としていますが，DMAを搭載したMCUでは，タイマ割り込みの代わりにDMA転送を用いて音声データの取得を行うこともできます。このようにULTALKER-Cでは，非常に簡単に音声認識システムを構築することができます。

ULTALKER-Cの評価環境

先に述べたように，ULTALKER-Cではシステム構築のコスト低減を念頭に設計・開発を行いました。このコスト削減はデバイスコストだけでなく，ユーザがアプリケーションシステムの開発を行う際のコストにも当然あてはまります。そこで，ユーザが音声認識の評価やアプリケーション検討に手軽に使用できる評価システムを用意しました（**写真1参照**）。

- 音声認識デモシステム
  ユーザが任意の認識単語セットを用意して，音声認識性能の評価を行うためのデモシステムです。電源ONで即，音声認識を行うことができ，パソコンから任意の認識単語辞書をダウンロードする事によりアプリケーションに特化した認識単語による評価も行うことができます。認識した結果は，RS232-C経由で出力すると同時に，簡易音声合成による音声出力も行います。
- 音声認識プロトタイプシステム
  ＲＯＭ化された音声認識エンジンをRS232-Cインタフェースを有するホストマシンからコマンドにより操作するプロトタイプシステムです。本システムを使用することにより，RS232-C経由で実機上で動作する音声認識を直接操作することができるので，アプリケーションシステムとしての簡易評価などがパソコンや既存



－V850E/MS1によるリアルタイム応答
－任意辞書使用可能
－簡易音声合成による応答

写真1　ULTALKER-C評価システム

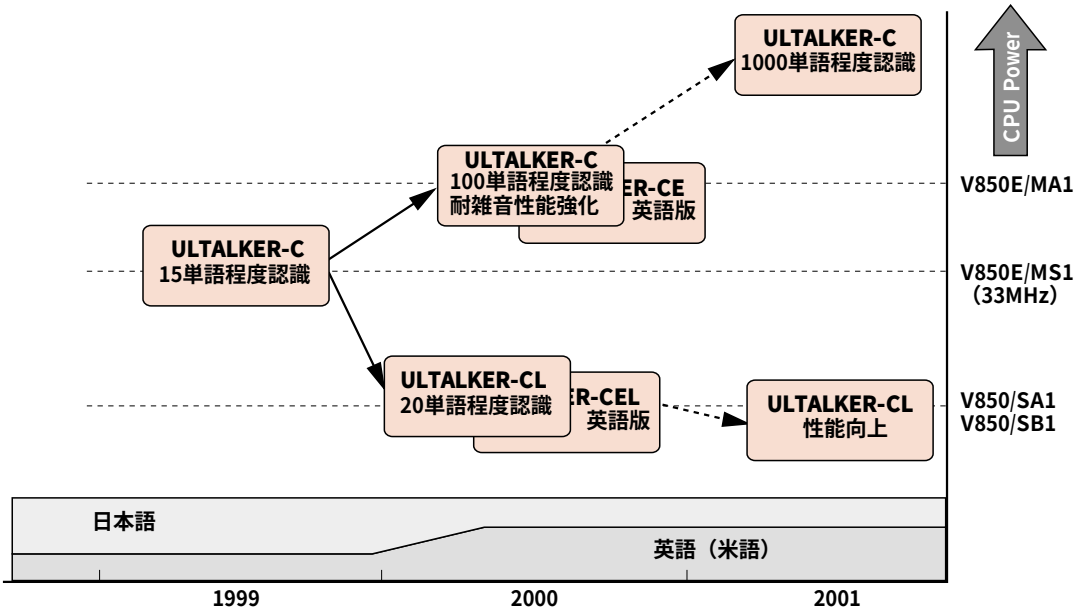


図7　ULTALKER-Cロードマップ

のシステムを用いて容易に行えるようになります。

これらの評価環境により，開発の初期投資によるユーザの負担を大幅に軽減できると考えています。なお，これらの環境は，サードパーティから入手できます。

今後の展開

これまで，今回開発を行ったULTALKER-Cについて説明してきましたが，今後の展開として，V850ファミリのデバイス展開に合わせ「ハイエンドラインナップ」と「ローエンドラインナップ」に合わせた製品展開を行っていく予定です（**図6参照**）。

- ハイエンドラインナップ
  V850E/M系MCUのコア性能をフルに活かし，認識単語数の向上をはじめとする機能追加を行っていく予定です。VR系向けの音声認識との性能差を埋めるものになると予想しています。
- ローエンドラインナップ
  V850/S系MCUに適応する，よりコンパクトな小語彙音声認識エンジンとして，リソースの削減，CPU占有率の削減などさらなるコストダウンを主眼に開発を行っていく予定です。

また，音声認識のみならず，音声ヒューマンインタフェースとして音声合成や音声圧縮などの音声応答を司るミドルウェアの充実も行っていきたいと考えています。

おわりに

今回，V850ファミリ向けの音声認識ミドルウエア「ULTALKER-C」について紹介してきました。カーナビゲーションなど特定の情報機器に音声認識技術が普及し，世の中に音声認識技術が認知され，ようやく音声認識技術が一般に普及する時期にきたといえます。今回開発したULTALKER-Cは，パーソナルユースとして使われる小規模なセットへの搭載に向けた音声認識エンジンであり，広く多くの人々に音声認識を使ってもらえるものと期待しています。

関連するHPの紹介

http://www.nec.co.jp/japanese/product/nuivoice/