

組込みソフトウェア：Inspirium

Embedded Software: Inspirium

● 徳田一郎 ● 森出茂樹

あらまし

Inspiriumは、富士通が開発した組込みソフトウェアの総称で、組込み機器の付加価値や品質を高めることを目的に、画像LSIに組み込まれる単体のミドルウェアからサーバ連携のミドルウェアまでを幅広く提供する。本稿では、ヒューマンセントリックなインタフェースとして、画像LSIでの画面表示とともにスマートフォンやタブレットPCなどで利用されているタッチパネルからの手書き文字入力を容易に実現させる、手書き文字認識ライブラリと、仮名漢字混じりの文章から正しい日本語で文章を音声で読み上げる音声合成ライブラリを紹介する。手書き文字認識ライブラリは、人が紙にペンで字を書くときの特徴を捉え、崩し文字や走り書きも高い認識率で入力できる。音声合成ライブラリは、仮名漢字混じりの文章から、聞きやすく正しい日本語で読み上げることができる。あらかじめ登録された単語辞書から単語の読みとアクセントを取り出し、韻律処理によってイントネーションとリズムを整え、聞きやすい日本語の音声を合成する。

本稿では、この2種類のライブラリの仕様・構成・応用例について述べる。

Abstract

Inspirium is a generic name for embedded software developed by Fujitsu. It is intended to increase the added value and quality of embedded devices and comes in a wide range of products including standalone middleware embedded in imaging chips (large scale integration or LSIs) and server linkage middleware. In addition to describing a screen displayed with imaging LSIs, this paper presents human-centric interfaces including a handwriting recognition library capable of easily realizing handwriting input from the touch panels used in smartphones and tablet PCs, and a speech synthesis library that provides an audio read-out in correct Japanese of text containing a mixture of kana (the syllabic Japanese scripts) and kanji (Chinese characters). The handwriting recognition library captures the characteristics of a person's handwriting on paper with a pen and is good at recognizing simplified characters and scribble. The speech synthesis library can read out text containing a mixture of kana and kanji in correct and listener-friendly Japanese. A predefined dictionary of words is used for adjusting the readings and stresses of words in terms of intonation and rhythm by prosodic processing to synthesize listener-friendly Japanese sounds. This paper describes the specifications and configuration of these two libraries and gives application examples.

まえがき

近年、組込み機器のソフトウェアの開発が非常に複雑化し、長期化、高コスト化している。また、製品の付加価値を高めるには、OS、ミドルウェア、アプリケーションとソフトウェアの比重が大きく、製品開発の足かせとなっている。そこで、富士通は自社の製品開発で培われたソフトウェアの技術を基に、お客様の組込み機器の開発を総合的に支援するミドルウェアとして組込みソフトウェアInspirium⁽¹⁾を開発した。Inspiriumを組込むことにより、短期間で高付加価値の製品を低コスト、高品質で開発することが可能となる。

本稿では、Inspiriumの製品ラインナップの中から、手書き文字認識ライブラリと音声合成ライブラリの2種類について述べる。

Inspirium手書き文字認識ライブラリ

情報通信の世界では、情報入力を中心としてキーボードが長く使われている。キーボードは、欧米のアルファベット入力に優れている反面、日本語のような複雑かつ多種類の文字を持つ言語には、仮名漢字変換などの文字の読みや音を入力し複数の候補から選ぶといった複雑な処理が必要となる。更に、情報を素早く入力するためには、キーボードの文字位置（配列）を覚える必要もあり、ある程度の訓練が必要となる。また、日本語の場合、漢字の読み方が分からないと入力すらできない状況にある。

そこで、富士通は普段、我々がペンで紙に文字を書く感覚で仮名漢字混じりの日本語を入力する手書きの文字認識をミドルウェア化した。このミドルウェアは日本語の文字をキーボードではなく、タッチパネルに指で書く感覚で入力できるミドルウェアである。複雑な日本語の入力を直接文字で入力することが可能で、読み方の分からない漢字も形をなぞることで簡単に入力できる。このことにより、漢字を苦手とする外国の方や小さな子供でも、難しい漢字を入力することが可能となる。

手書き文字認識の導入例を図-1に示す。

以下、Inspirium手書き文字認識ライブラリの特徴・構成・仕様・応用例を紹介する。

● 特徴

Inspirium手書き文字認識ライブラリは、情報通信機器のユーザインタフェースとして、手書き文字による入力インタフェースを提供する。日本語に加え、中国語（GKB文字セット：簡体字、繁体字）・英語（ブロック体）にも対応している。以下に特徴を示す。

- ・筆順や画数が変動しても高い認識率を提供
- ・走り書き略字も正しく認識（図-2）
- ・前後の文字を見てリアルタイムに認識結果を補正
- ・PC、PDA、携帯電話など各種の情報通信機器への搭載が可能
- ・各種OSに対応（Android、Linux、Windows、…）

Inspirium手書き文字認識ライブラリが文字を認識する方法を図-3に示す。入力された文字の形での認識（筆順に左右されない）とペンの筆順で



図-1 手書き文字認識の導入例



崩し文字・筆順変動をリカバー
精度が必要な指入力に対応

図-2 走り書きの例

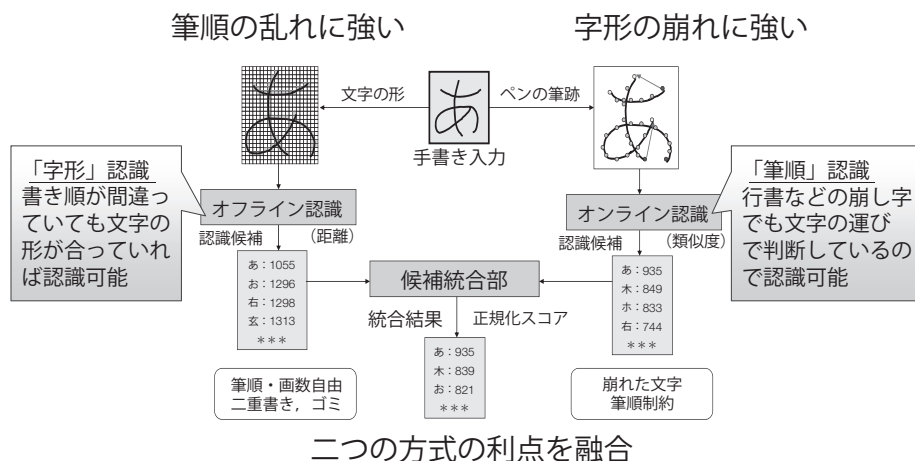


図-3 文字認識方法

認識（字形の崩れに強い）する、ハイブリッド認識を採用している。このハイブリッド方式により、高い認識率を誇る。また、前後の文字を見てリアルタイムに認識結果を補正する文脈補正機能を備えている。

● 構成

Inspirium手書き文字認識ライブラリは、大きく三つのソフトウェアから構成される。

- ・手書き文字認識エンジン
- ・認識辞書
- ・手書き入力ユーザインタフェース部

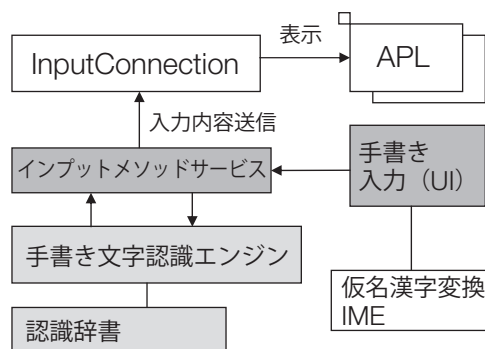
このようなシンプルな構成により、お客様の各種機器の特徴を生かせる、手書き文字認識エンジンとして機能する。画面表示のデザインやタッチパネルの種類などが自由に選択可能で、手書き文字認識を容易に組み込むことが可能となる。

Androidでの実装例を図-4に示す。手書き文字認識のエンジンはタッチパネルなどの入力装置から送られてくる座標データを基に文字認識を行っている。本稿ではタッチパネルを主に説明しているが、入力したい文字の座標データを本ライブラリに入力できれば、文字認識エンジンとして機能する。

例えば、ゲーム機のリモコンから手の動きを読み取り、座標データとして求める。座標データから、文字認識エンジンで文字に変換しゲームに文字情報を入力するといったことも考えられる。

● 仕様

組み込み機器で多く搭載されている、OS/CPUでの利用を可能としている。また、認識させる文字



製品提供範囲

パターン1： 認識エンジン部

パターン2： 認識エンジン部
+ IMEアプリ部

図-4 Androidの実装例

種により必要とするメモリ容量が異なる。下記に詳細を示す。

- (1) OS：Android, Linux, Windows, そのほかのOSに対応可能
- (2) CPU：ARM, FR, SH, Pentiumなど
- (3) メモリ容量

- ・日本語対応版 標準版 約2 Mバイト
認識部341 Kバイト, 辞書1.2 Mバイト (JIS1+2)
文脈処理部476 Kバイト, そのほかワーク域
- ・日本語対応版 省メモリ版 約1.4 Mバイト
認識部341 Kバイト, 辞書0.6 Mバイト (JIS1)

- 文脈処理部476 Kバイト，そのほかワーク域
- ・中国語対応版 約9.3 Mバイト
- 認識部550 Kバイト
- 辞書8.3 Mバイト
- (GBKフルサポート，GBK21 866文字)
- 文脈処理部525 Kバイト，そのほかワーク域
- ・認識速度
- Pentium III，400 MHz：21.3 ms
- StrongARM 206 MHz：79.8 ms

● 応用例

タッチパネルを搭載した装置のユーザインタフェースとして，日頃からキーボードに慣れ親しんでいない方でも，手書きの文字入力であれば容易に文字入力が可能となる。応用例として，カーナビゲーションの目的地や住所の入力，モバイル機器におけるメールなどの文章入力，また，ATM端末，キオスク端末の文字入力を必要とする機器など，文字を直接記述する必要がある機器での応用が考えられる (図-5)。

Inspirium音声合成ライブラリ

情報通信の世界では，音声で情報を伝えることは100年以上の昔から電話機やラジオとして多く利用されてきた。現在の電話機は，音声はもちろん，文字，絵，写真など，多種の情報が送受信できるスマートフォンへと進化している。一方で，情報処理 (コンピュータ) の世界では，文字や絵で情報を伝えることが多く，音声で情報を伝える方法は，あまり利用されていない状況にある。音声情報は文字情報に比較し，大容量のデータを必要とするなど非常に扱いにくい情報となっている。しかし，音声による情報は，文字や絵が見えない環

境下でも，聴覚によって伝達を可能とする特徴もある。そこで富士通は，比較的，小規模な組込み機器でも実装可能な，音声合成ミドルウェアを開発した。

● 特徴

仮名漢字混じりの文章から，音声で文章を読み上げる出力インタフェースを提供するライブラリで，男性の声と女性の声，また利用できるメモリ環境により高音質版と低音質版に対応している。

以下に特徴を示す。

- ・日本語の仮名漢字混じりの文章を読上げ
- ・人間の肉声に近い音声での自然な読上げ
- ・連続する数字は位取りをして読上げ
- ・人の名前，地方名 (固有名詞)，特別な読み方をする単語を，言語辞書で対応
- ・男性の声1種，女性の声2種
- ・組込み機器に向けた小資源なROMサイズ
- ・PC，PDA，携帯電話など各種の情報通信機器への搭載が可能
- ・各種のOSに対応 (Android，Linux，Windows など)

● 日本語の読み方の問題点1

日本語は，一つの文字に複数の読み方がある。例を挙げると

『橋』音読み：キョウ 訓読み：はし
 『日本橋』：にほんばし・にっぽんばし
 漢字1文字の『橋』でも，「キョウ/はし」と音読みと訓読みの2種類の読み方がある。

また，『日本橋』の読みは，「にほんばし」と「はし」が「ばし」と濁る。更に大阪では同じ『日本橋』を「にっぽんばし」と読む。日本語には，文字の読み方が複数あり，読み方が正しくないと意味す

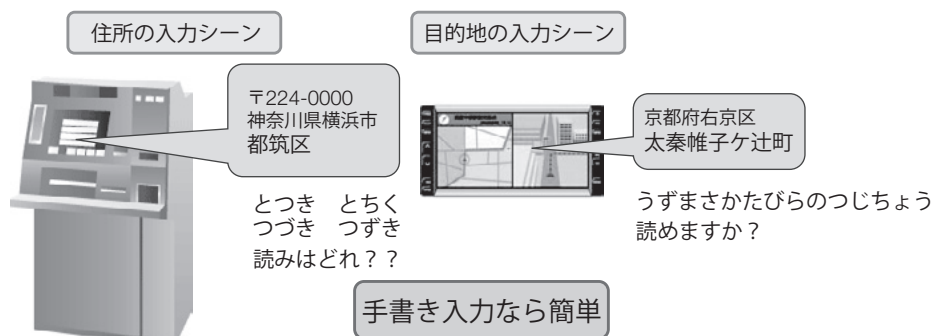


図-5 応用例

ら通じなくなる場合もある。

● 日本語の読み方の問題点2

文章により、アクセントとイントネーションが正しくないと意味の違った文章として聞こえる。

例を挙げると

- はしをわたる⇒『橋を渡る』：普通はこの文
- 『端を渡る』：注意喚起文では
- 『箸を渡る』：間違えると

文字では正しく認識できても音声では全く違った内容に聞こえる。

また、スーパーコンピュータ「京」^(注)の演算処理速度を音声で読み上げるとすると。

(目標値) 10¹⁶ (10000000000000000)

この文は『いっけい』と読んでほしいが、『いち、ぜろ、ぜろ、…』と読み上げても意味が通じない。

● 自然でかつ正確な音声合成の方法

Inspirium音声合成ライブラリでは、これらの日本語の特徴を捉え、任意の仮名漢字混じりの文章を自然でかつ正確な音声合成を行うため、下記の方法で対応する。

(1) 単語、熟語に複数の読み方がある文章の対応

日本語の読み方を辞書として持つことによって、難しい地方名や人名に対応し、基本単語辞書として17万語を収録している。更に、新しい言葉には辞書を追加することにより対応する。

(2) 柔らかな合成への対応

従来の音声合成では、ロボットが話す声のよう

な、平坦な音声で、非常に聞きづらい合成音であった。そこで、人工的な印象をなくした柔らかな合成音を作るために図-6に示す処理を行う。入力された文章からまず言語処理として、読み、アクセント（高低）、間の取り方などの情報を基本単語辞書から取り出す。次に韻律生成として、イントネーション、リズムのパターンを決定する。更に、音声波形処理として、イントネーションのパターンに沿って、音声波形データベースから、波形データを取り出し、波形を接続する。接続された波形をデジタルアナログ変換により音声データが生成される。柔らかな音声合成には、この韻律処理が重要なウエイトを占める。

● 構成

Inspirium音声合成ライブラリは大きく三つのソフトウェア、二つの辞書から構成される。

(1) 言語処理部

単語辞書から、読み、アクセントを取り出す。

(2) 韻律処理部

イントネーション、リズムのパターンを決定する。

(3) 波形合成部

音声波形データベースから、波形データを取り出す。

(4) 基本単語辞書

読み、アクセントを収録する。

(5) 音声波形データベース

音声波形を収録する。

● 仕様

Inspirium音声合成ライブラリには、お客様の組

(注) 理化学研究所が2010年7月に決定したスーパーコンピュータの愛称。

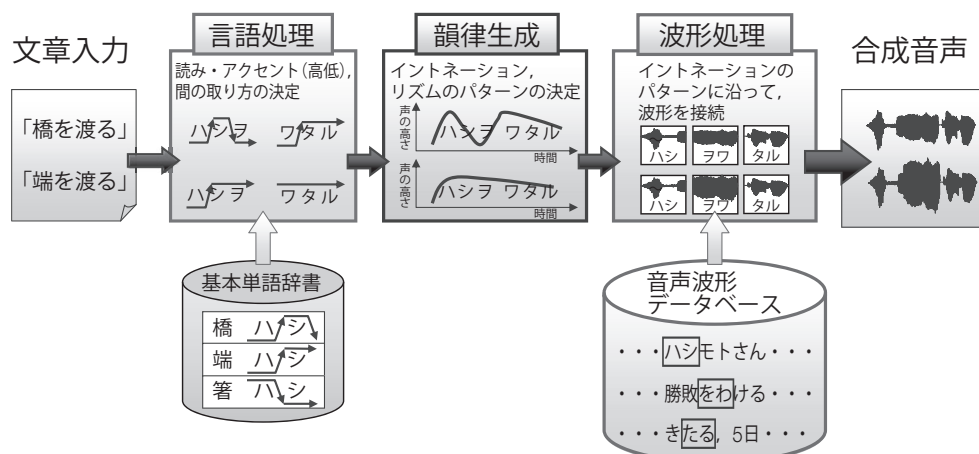


図-6 音声合成処理



図-7 利用シーン

込み機器の仕様により、言語処理として、数値の桁読みの有無、記号の読みの有無、音響処理として、声の高さ・スピード・イントネーション・ボリューム・高域強調の設定が可能となっている。メモリ容量の大きさにより、省資源版、SPMU版、高音質版の3種類の音質設定が選択可能となっている。また、男性の声もしくは女性の声（2種）の選択も可能としている。

下記に利用可能な資源を示す。

- (1) OS：Android, Linux, Windows, そのほかのOSに対応可能
- (2) CPU：ARM, FR, SH, Pentiumなど
- (3) メモリ容量
 - ・プログラム容量：1.0 Mバイト

- ・ワーク領域：最大1.0 Mバイト
- ・基本言語辞書：1.5 Mバイト（17万語）
- ・波形辞書：1.5 Mバイト～102 Mバイト

● 応用例

音声でのユーザインタフェースとして様々な応用が考えられる。実用例としてTV番組表の読上げ、携帯電話などのメール・ニュースの読上げ、また、車や機械の運転中に画面が見えない環境下での音声による指示や確認などが挙げられる。更に、画面の文字による指示と、音声による指示で2重の確認が可能で、安全の確保にも応用可能である。

利用シーンを図-7に示す。

む す び

本稿では、富士通の組込みソフトウェアInspiriumの手書き文字認識ライブラリと音声合成ライブラリについて述べた。高性能化するマイクロコントローラのヒューマンセントリックなインタフェースとして様々な機器で利用されていくものと考えている。今後は、多国語対応など、更に製品を充実させ、ワールドワイドに利用可能な製品開発に努めていく。

なお、富士通の組込みソフトウェアInspiriumは本稿で紹介した製品以外でも各種のソフトウェアがあり、お客様の製品の付加価値を高める製品として提供している。

参考文献

- (1) 富士通：汎用組込みソフトウェアInspirium（インスピリウム）。

<http://edevicet.fujitsu.com/jp/products/embedded/>

著者紹介



徳田一郎 (とくだ いちろう)

富士通エレクトロニクス（株）
ソリューションサポートセンターソリューション推進部 所属
現在、LSIを利用したソリューションの提案、お客様のサポート業務に従事。



森出茂樹 (もりで しげき)

ミドルウェア事業本部組込みソフトウェアテクノロジー事業部 所属
現在、組込みミドルウェアの企画・マーケティングに従事。