Embedded Software: Inspirium

● Ichirou Tokuda ● Shigeki Moride

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 image processing 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 touchscreens 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.

1. Introduction

Recently, embedded device software has become extremely complicated, leading to longer development periods and higher costs. In addition, in order to increase the added value of embedded devices, software—such as OSes, middleware and applications—is getting important. To address this situation, Fujitsu has developed embedded software called Inspirium[1)] as middleware that comprehensively helps custom engineers develop embedded devices based on software technology nurtured with its own product development. By integrating Inspirium, high-value added and high-quality products (embedded devices) can be developed at low cost in short periods of time.

Of the Inspirium product line, this paper presents the handwriting recognition library and speech synthesis library.

2. Inspirium handwriting recognition library

In the world of information communication, keyboards have long been used as the main means of information input. While keyboards are useful for inputting Western alphabets, with languages including complex and diverse characters such as Japanese, complicated processing is required such as input of readings and sounds of characters and selection from multiple candidates as in kana (the syllabic Japanese scripts)-kanji (Chinese characters) conversion. In addition, the keyboard key locations (layout) must be learned for quickly inputting information, and this requires a certain level of training. With Japanese, information cannot be input in the first place without knowing the readings of kanji.

To deal with this situation, Fujitsu has built middleware capable of recognizing handwritten characters, and it accepts input of Japanese containing a mixture of kana and kanji in the same way as we usually write on paper with a pen. This middleware allows Japanese to be input by "writing" on a touchscreens with a finger rather than by using a keyboard. Complicated Japanese characters can be input as characters as they are, and

kanji with unknown readings can be easily input by tracing the shapes. This allows foreigners and small children who are not good at kanji to input complicated kanji.

Figure 1 shows an example of adopting handwriting recognition.

The following presents the features, configuration, specifications and application examples of the Inspirium handwriting recognition library.

2.1 Features

The Inspirium handwriting recognition library provides a handwriting input interface as a user interface of information communication devices. In addition to Japanese, it supports Chinese (GKB character set: simplified/traditional) and English (block letters). It features:

- A high recognition rate regardless of variation of stroke orders or counts
- Correct recognition of scribble and abbreviated forms (**Figure 2**)
- Real-time correction of recognition results based on the preceding and following characters
- Installability in various information communication devices including PCs, PDAs and mobile phones
- Support for various OSes (including Android, Linux and Windows)

Figure 3 shows how the Inspirium handwriting recognition library recognizes a character. It adopts hybrid recognition, which combines recognition by the shape of the character input (not affected by the stroke order) and recognition based on pen stroke order (not affected by simplified character shapes). This hybrid system has achieved a high recognition rate. It is also equipped with a context-based correction feature that corrects recognition results in real time based on the preceding and following characters.

2.2 Configuration

The Inspirium handwriting recognition library is fundamentally composed of the following three pieces

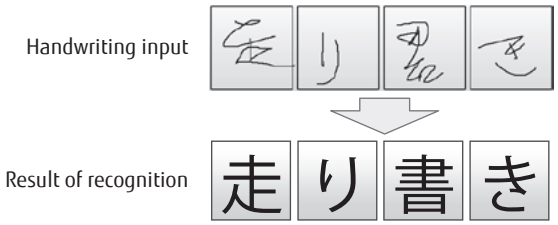

Figure 2
Example of scribble.

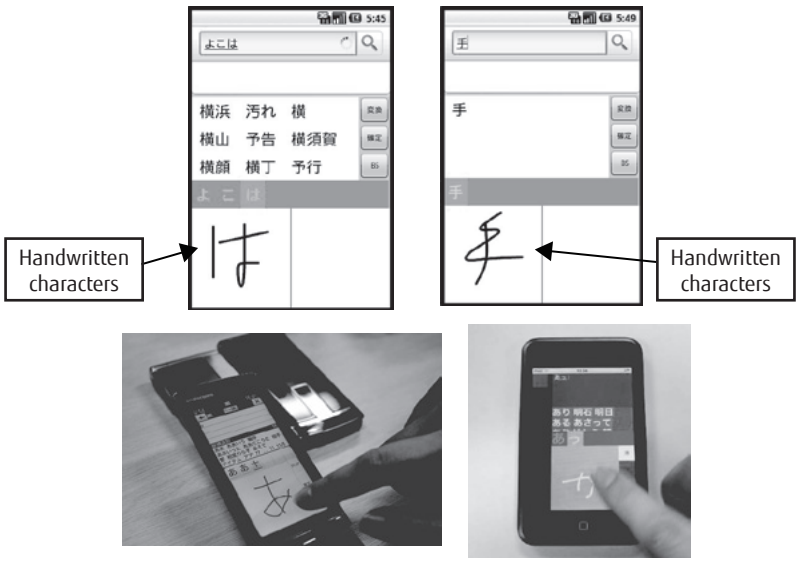

Figure 1
Example of adoption of handwriting recognition.

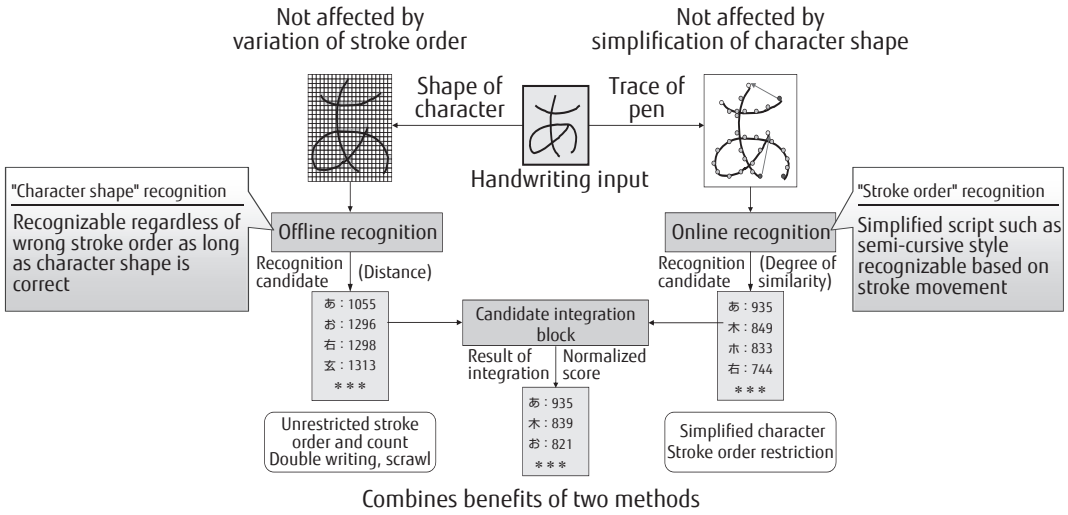

Figure 3
Process of handwriting recognition.

of software:

- Handwriting recognition engine
- Recognition library
- Handwriting input user interface block

This simple configuration makes it functional as a handwriting recognition engine that maximizes the characteristics of the various devices of customers. The screen display designs and touchscreens types can be freely selected and handwriting recognition can be integrated easily.

Figure 4 shows an example of implementation in Android. The handwriting recognition engine uses the coordinate data sent from an input device such as a touchscreens to recognize characters. Although this paper mainly uses a touchscreens for explanation purposes, the character recognition engine functions as long as the coordinate data of the character to input can be input in this library.

Another possible example is reading hand movements from a game machine console as coordinate data, and this data can then be converted into characters by using the character recognition engine to input character information to the game.

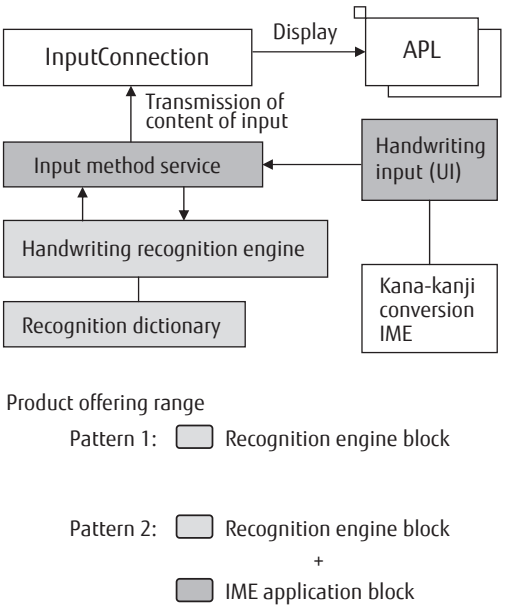

Figure 4
Example of implementation in Android.

2.3 Specifications

The library can be used with OSes and CPUs that are often installed in embedded devices. The required memory size varies depending on the type of characters to be recognized. The following details the specifications.

1) OS: Android, Linux, Windows and others
2) CPU: ARM, FR, SH, Pentium, etc.
3) Required memory size

- Japanese-capable standard version: approximately 2 MB
  341 KB for recognition block, 1.2 MB for dictionary (JIS1 + JIS2)

476 KB for context processing block, rest for work area

- Japanese-capable memory-saving version: approximately 1.4 MB
  341 KB for recognition block, 0.6 MB for dictionary (JIS1)
  476 KB for context processing block, rest for work area
- Chinese-capable version: approximately 9.3 MB
  550 KB for recognition block
  8.3 MB for dictionary
  (GBK full support, GBK 21 866 characters)
  525 KB for context processing block, rest for work area
- Recognition speed
  Pentium III, 400 MHz: 21.3 ms
  StrongARM, 206 MHz: 79.8 ms

2.4 Application examples

As a user interface of a device that integrates a touchscreens, character input by handwriting allows users who are not familiar with keyboard input to easily input characters. Possible examples of application include devices requiring direct provision of characters such as input of destinations and addresses of car navigation systems, text input for e-mailing and other purposes with mobile devices, and devices requiring character input of ATM and kiosk terminals (**Figure 5**).

3. Inspirium speech synthesis library

In the world of information communication, use of speech for communicating information has been in practical application since over 100 years as telephone and radio. Telephones have now evolved into smartphones capable of sending and receiving various types of information such as text, pictures and photographs, not to mention speech. In the world of information processing (computer), however, information is often communicated by text and pictures and the method of communicating information by speech is not used frequently. As compared with text information, speech information is very difficult to handle for reasons including the fact that it requires large amounts of data. But speech information is also characterized by its ability to communicate by hearing in environments where text or pictures are not visible. Accordingly, Fujitsu has developed speech synthesis middleware that can be integrated into relatively small embedded devices.

3.1 Features

The speech synthesis library provides an output interface capable of offering an audio read-out of text containing a mixture of kana and kanji and allows use of not only male and female voices but also high and low sound quality according to the memory environment available.

It features:

- Read-out of Japanese text containing a mixture of kana and kanji
- Natural read-out with speech similar to human voice
- Read-out of successive digits as one number
- Language dictionary to support names of people and places (proper nouns) and words with special readings
- Availability of one type of male voice and two

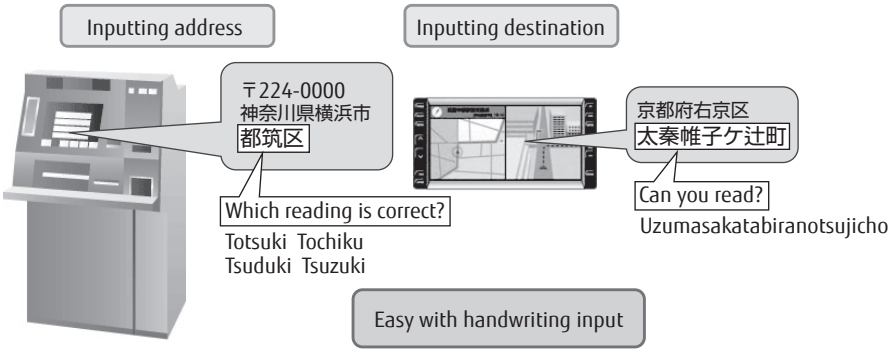

Figure 5
Application examples.

types of female voices
- Resource-saving ROM size suitable for embedded devices
- Installability in various information communication devices including PCs, PDAs and mobile phones
- Support for various OSes (including Android, Linux and Windows)

3.2 Issue with Japanese reading 1

In the Japanese language, one character often has more than one reading.

For example,

橋: Pronounced "kyo" in the Chinese-derived reading, and "hashi" in the Japanese reading

日本橋: Nihombashi or Nippombashi

Even with one kanji –"橋"– there are two readings: "kyo" in the Chinese-derived reading and "hashi" in the Japanese reading.

When it is incorporated in a geographical name –"日本橋"– the reading is changed from "hashi" to "bashi." In addition, the same set of kanji may be read "Nihombashi" or "Nippombashi" depending on the area. In Japanese, one character has more than one reading and the correct meaning may not be communicated with a wrong reading.

3.3 Issue with Japanese reading 2

Without correct stressing and intonation, certain text may be interpreted differently from the original text when it is read out.

For example,

The phrase "*Hashi wo wataru*" may use the characters:

1) 橋を渡る ("Cross a bridge"): the most common interpretation
2) 端を渡る ("Cross on the edge"): may be used when calling attention
3) 箸を渡る ("Cross chopsticks"): interpreted wrongly

If the text can be correctly recognized in writing, it may sound totally differently in the form of speech.

To describe another point, let us suppose that the processing speed of the K computer[note)] is read out:

[Mokuhyochi (target)] 10^{16} (10 000 000 000 000 000)

In this context, the number should be read "*ikkei*," which means one trillion, and reading out each digit as in "*ichi* (one), zero, zero,..." does not make sense.

3.4 Natural-sounding and correct speech synthesis

The Inspirium speech synthesis library uses the following schemes to take these characteristics of the Japanese language into account to achieve natural-sounding and correct speech synthesis of arbitrary text containing a mixture of kana and kanji.

1) Scheme to deal with text containing words and/or phrases with more than one reading

The library supports difficult-to-read names of places and people by adding Japanese readings to a dictionary and contains 170 000 words as the basic word dictionary. New words are added to dictionaries.

2) Scheme to realize mild synthesized speech

The conventional speech synthesis produced synthesized speech in a flat tone that sounded like a robot speaking and was very unfriendly to listeners. To address this problem, processing as shown in **Figure 6** is carried out to generate mild synthesized speech that does not give an impression of being artificial. For the text that has been input, information including reading, stress (pitch) and tempo is first retrieved from the basic word dictionary as the language processing. Next, as the prosodic generation, intonation and rhythm patterns are determined. Then, as the speech waveform processing, waveform data are retrieved from the speech waveform database according to the intonation pattern, and the waveform data are concatenated. The waveforms concatenated are subjected to digital-to-analog conversion to generate speech data. To realize mild speech synthesis, this prosodic processing has great importance.

3.5 Configuration

The Inspirium speech synthesis library is fundamentally composed of three pieces of software and two dictionaries.

1) Language processing block
   Retrieves the reading and stress from the word dictionary.
2) Prosodic processing block
   Determines the intonation and rhythm patterns.
3) Waveform synthesis block
   Retrieves waveform data from the speech

note) English name given by RIKEN in July 2010 for the supercomputer.

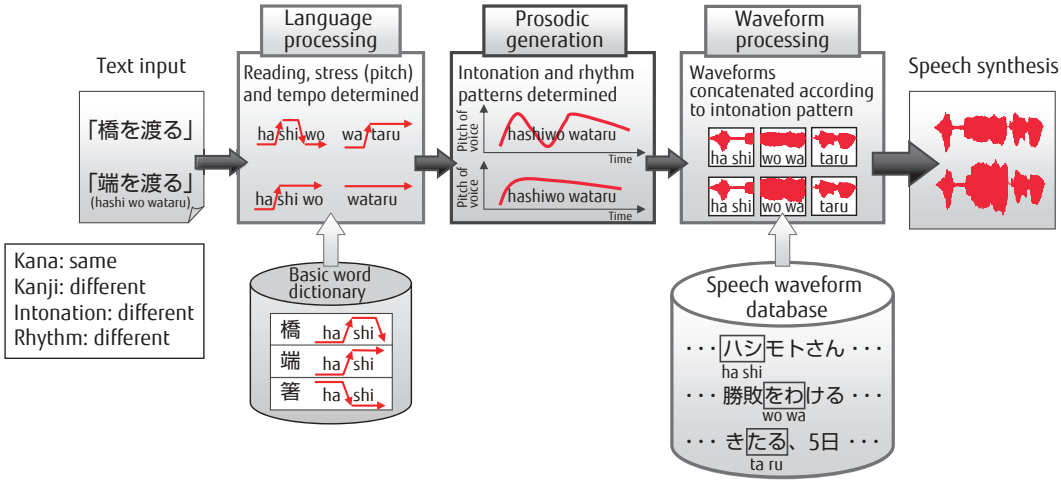

Figure 6
Speech synthesis processing.

waveform database.

4) Basic word dictionary
Contains readings and stresses.

5) Speech waveform database
Contains speech waveforms.

3.6 Specifications

The Inspirium speech synthesis library offers various settings that can be configured according to the specifications of the customer's embedded device. Such specifications may stipulate whether or not to read a set of successive digits as a number, and whether or not to read symbols for the language processing and the voice pitch, speed, intonation, volume and high emphasis for the acoustic processing. In addition, three sound quality settings are available including the resource-saving, SPMU and high sound quality versions according to the memory size. The type of voice is also selectable between male (one type) and female (two types) voices.

The system requirements include:

1) OS: Android, Linux, Windows and others
2) CPU: ARM, FR, SH, Pentium, etc.
3) Required memory size

- 1.0 MB for program
- 1.0 MB max. for work area
- 1.5 MB (170 000 words) for basic language dictionary
- 1.5 MB to 102 MB for waveform dictionary

Figure 7
Scenes of use.

3.7 Application examples

The library can be applied in many ways as a user interface by means of speech. Examples of practical application include reading out TV show schedules, reading out e-mail and news on devices such as mobile phones, and reading out audio instructions and confirmation in environments where screens are not visible while driving or operating machines. Furthermore, double confirmation can be made through instructions

by text on the screen and instructions by speech, and this can be applied to ensure safety.

Figure 7 shows some scenes of use.

4. Conclusion

This paper has presented the handwriting recognition library and speech synthesis library of Inspirium, Fujitsu's embedded software. We believe that it will be used in various devices as a human-centric interface of microcontrollers that are becoming increasingly sophisticated. In the future, we intend to further enhance the products by providing multi-language support and other features for the development of products that can be used worldwide.

Fujitsu's embedded software Inspirium includes various types of software in addition to the products presented in this paper, and they are offered as products that increase the added value of customers' products.

References

1) Fujitsu: General-Purpose Embedded Software: Inspirium. (in Japanese).
http://edevice.fujitsu.com/jp/products/embedded/

Ichirou Tokuda
Fujitsu Electronics Inc.
Mr. Tokuda is currently engaged in proposal of solutions using LSIs and customer support.

Shigeki Moride
Fujitsu Ltd.
Mr. Moride is currently engaged in planning and marketing of embedded middleware.