

スマートメイちゃん：スマートフォンの機能を有する モバイル音声対話エージェント

山本大介^{†1†2} 西村良太^{†1†2} 打矢隆弘^{†1†2} 内匠逸^{†1†2}

本稿では、音声対話システム構築ツールキット MMDAgent を拡張し、GPS や NFC・GUI 等の Android スマートフォンの機能を効果的に活用するための仕組みを提案する。具体的には、全ての Android OS の機能を利用するためには Java 言語で開発する必要があるが、MMDAgent は C++ 言語での開発が想定されており、異なる言語間での連携は容易ではないという問題があった。そこで、Java と C++ の連携を実現するブリッジモジュールを開発し、容易に MMDAgent の各モジュールと Android の機能とを連携可能な Java プラグイン開発プラットフォームを提案した。また、本プラットフォームに基づく様々な Java プラグインの開発も行った。

Smart Mei-chan: A Mobile Voice Interaction Agent using Smartphone Functions

DAISUKE YAMAMOTO^{†1†2} RYOTA NISHIMURA^{†1†2}
TAKAHIRO UCHIYA^{†1†2} ICHI TAKUMI^{†1†2}

1. はじめに

近年、Siri[1]やしゃべってコンシェル[2]のように、携帯端末における音声対話システムが普及している。これらの音声対話システムでは、音声を端末側で録音してからサーバに転送し、サーバ側で音声認識した結果を端末側に転送する必要があるため、ネットワーク通信に起因する対話の遅延が発生する問題がある。そこで、我々は既存の音声対話システム構築ツールキット MMDAgent[3]を Android に移植することにより、図 1 に示す、端末内で音声合成・音声認識処理が可能なスマートメイちゃん[4]を開発してきた。これにより、ネットワーク通信に起因する遅延がなく応答時間が短いモバイル音声対話エージェントを実現した。また、感情音声合成技術により感情を込めた音声合成を行うと同時に、3D キャラクタの振る舞いも変更することも可能である。

その一方で、PC 版の MMDAgent をスマートフォンに移植しただけでは、モバイル環境での使いやすさや特徴を必ずしも生かすことはできない。たとえば、位置情報を利用した観光案内などを実現するためには、GPS や NFC(Near field communication)タグ、Web ブラウザや GUI(Graphical User Interface)などスマートフォンで標準的に利用されている機能と高度に統合できると良い。

そこで、本研究では、これらのスマートフォンの機能をモバイル音声対話システムに効果的に組み込んだ、モバイル音声対話プラットフォームを開発する。これにより、より使いやすく満足度の高いモバイル音声対話システムを実



図 1 スマートメイちゃん

現するだけでなく、よりユーザにとって魅力的な音声対話システムとは何かという問いの答えを追求していく。その際、我々が考慮した点は以下のとおりである。

1. MMDAgent の基本的な機能は全て C++ 言語で記述されているが、Android OS の機能の多くは Java からのみ利用可能である。そのため、Java から MMDAgent の機能と連携が可能であること。
2. MMDAgent は FST(Finite State transducer)スクリプトと内部メッセージによって対話制御されているため、これらの枠組みに基づいてスマートフォンの機

^{†1} 名古屋工業大学
^{†2} CREST, JST

能を利用できること。

2. MMDAgent

MMDAgent は、名古屋工業大学国際音声技術研究所で開発された音声対話システム構築ツールキットであり、音声認識、音声合成、3D モデルの描画や物理演算などのモジュールを統合したシステムである。音声認識エンジンとして Julius [5]を、音声合成エンジンとして Open JTalk [6]を、3D モデルとして MikuMikuDance 形式 [7]を、物理演算エンジンとして Bullet Physics [8]を採用している。

MMDAgent は実時間かつ低遅延で音声対話を行うことができ、かつ、3D キャラクタの描画などが可能など、本研究の目的を満たすため、MMDAgent を採用した。しかしながら、MMDAgent は PC 向けのソフトウェアであり、Windows や Linux, Mac OS X での利用はできるが、スマートフォンでは利用できない。そこで、我々は MMDAgent をスマートフォン(Android OS)に移植してきた。

また、MMDAgent は、音声認識などのイベントをトリガーとした有限状態遷移(FST: Finite state transducer)に基づいて対話の実現した音声対話システムである。音声認識や音声合成などといった処理をマルチスレッドで実現することにより、特定の処理で状態遷移をブロッキングしない、リアルタイム性が高いシステムを実現している。例えば、天気予報の対話の例の場合、初期状態から、「名古屋」・「天気」などの単語を持つ音声を認識すると、名古屋の天気に関する説明文を音声合成しつつ、案内のモーションや天気予報のパネルを同時に表示することが可能である。各モジュール間の通信は、内部メッセージ（文字列）をやり取りすることによって行われる。

FST スクリプトの具体例を図 2 に示す。FST は、状態番号、遷移先状態番号、受入条件、コマンドの 4 つ組（スペース区切り）のリストから成る。図 2 の例では、現状態番号が 1 のとき、“RECOG_EVENT_STOP|こんにちは” という内部イベントメッセージを受信したとき、ε 遷移により状態番号 10 に遷移し、“MOTION_ADD|mei|greet|greet.vmd” という内部コマンドメッセージを出力して、状態番号 11 に遷移する。このように、状態遷移と内部メッセージのやり取りを繰り返すことによって、音声対話の制御を可能にしている。なお、<eps>は ε 遷移を表し、無条件で遷移することを表す。

3. 提案手法

3.1 Android OS との連携

PC 版の MMDAgent は C++ でプログラミングされているが、Android OS は基本的に Java でプログラミングされることを想定したシステムである。しかしながら、Android OS では Android NDK(Native Developer Kit) [9]を用いることによって、C++ でプログラムを記述することが可能である。

```
1 10 RECOG_EVENT_STOP|こんにちは <eps>
1 10 RECOG_EVENT_STOP|おはよう <eps>
10 11 <eps> MOTION_ADD|mei|greet|greet.vmd
11 12 <eps> SYNTH_START|mei|normal|おはよう。
12 13 SYNTH_EVENT_STOP|mei MOTION_ADD|mei|happy.vmd
13 14 <eps> SYNTH_START|mei|happy|素敵な朝ですね。
14 1 SYNTH_EVENT_STOP|mei <eps>
```

図 2 FST スクリプトの例（一部表記を簡約化）。

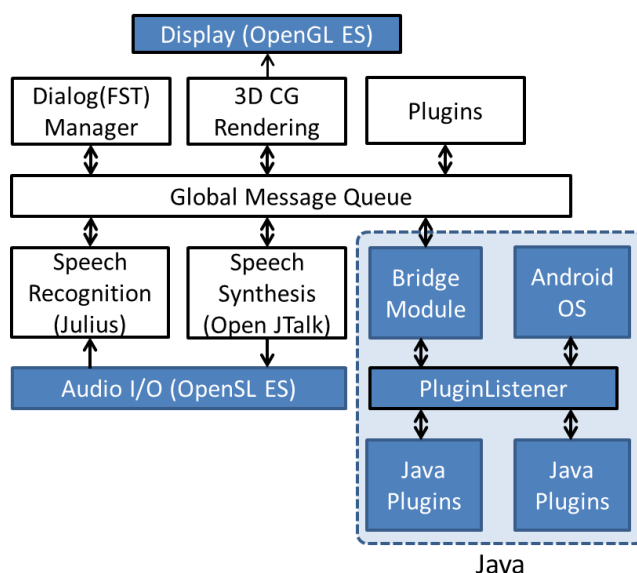


図 3 Android 版 MMDAgent のシステム構成図。Android OS や Java プラグインと連携するため Bridge Module と PluginListener を、OpenGL ES を用いた Audio I/O 周りを新たに実装した。

また、Android OS は Linux ベースの OS であるため、Linux 向けのソフトウェアが移植しやすいという特徴がある。そこで、Android NDK を活用することによって、Linux 版の MMDAgent を Android OS へ移植してきた。ただし、Android OS の OS 機能と連携する一部の機能（メール、カレンダー、ライブ壁紙、アプリ起動など）は Java からでしか利用できないため、図 3 に示すように、内部メッセージを、ブリッジモジュールを介してやり取りする方式で、Java で記述したプログラムや Android OS と C++ で記述したプログラムの連携を可能にした。

具体的な連携処理の手順を以下に述べる。MMDAgent は内部メッセージを Global Message Queue を通じてモジュール間通信を行うシステムである。そこで、Global Message Queue に流れる内部メッセージを、ブリッジモジュールを介して、Java 側の後述する PluginListener に転送する。ブリッジモジュールは JNI(Java Native Interface)機能を用いて、Java と C++ のプログラム間のデータ連携を可能にしている。反対に、Java 側から出力される内部メッセージは、同様に、

ブリッジモジュールを介して、Global Message Queue に転送される。これにより、内部メッセージとブリッジモジュールを介して、C++言語で開発されたモジュールと、Java で開発されたモジュールの連携を容易にしている。

3.2 プラグインによる機能拡張

MMDAgent はプラグインを作成することによって機能拡張が可能である。具体的には、図 4 に示すように、extProcCommand や extProcEvent 関数を実装することにより、内部メッセージ (char *type, char* args に格納されている) のやり取りを行うことができる。開発されたプラグインは DLL(Dynamic Link Library)として MMDAgent に容易に追加することが出来る。しかしながら、既存の MMDAgent のプラグインは、C++言語で開発する必要がある。そのままでは、Android OS の全ての API へアクセスすることができない。また、人によっては C++よりも Java で機能拡張をしたい人がいるかもしれない。そこで、Java でもプラグインを作成できる環境を構築する必要があると考えた。

そこで、図 5 に示す、PluginListener インタフェースを設計し、PluginListener インタフェースを継承することでプラグインを追加できるようにした。具体的には、onCommandMessage 関数や onEventMessage 関数を実装することによって、ブリッジモジュールを介して間接的に Global Message Queue と内部メッセージの連携が可能になる。また、PC 版 MMDAgent にない機能として、スマートフォンのライフサイクルに関するイベント (ポーズ、レジュームなど) をトリガーとした処理を onResume 関数、 onPause 関数などに記述することによって、きめ細やかな制御を実現可能にしている。

4. 実装

3章で述べた機能は、容易に Java でプラグインを実装可能なプラットフォームとして機能する。そこで、これらのプラットフォームを利用することによって、Android の機能を利用したプラグインをいくつか開発した。Android で利用可能な機能としては、例えば、センサー系の機能 (GPS, 加速度センサー, 地磁気センサー, NFC など), 通信系の機能 (ネットワークなど), GUI 系の機能(GUI, Web ブラウザモジュールなど), OS サービス (カレンダー, メールなど) が存在している。これらの機能は前述の PluginListener を実装することによって、プラグインとして開発している。内部メッセージを介して MMDAgent 本体の FST スクリプトや他のモジュールと連携することが可能である。

4.1 センサー系機能

センサー系機能は、主に、短い一定間隔でセンサーの値を検出するイベントメッセージを発行すれば良いと考えた。具体的には、GPS の場合、GPS の値が更新される毎に “GPS_EVENT_DETECT|(緯度)|(経度)|(精度)” の値を内部

```
EXPORT void extAppStart(MMDAgent *mmdagent){}
EXPORT void extProcCommand(MMDAgent *mmdagent,
    const char *type, const char *args) {
}
EXPORT void extProcEvent(MMDAgent *mmdagent,
    const char *type, const char *args) {
}
```

図 4 C++による拡張 (ext***関数を実装、DLL化)

```
public class Sample implements PluginListener {
    public void onCommandMessage(String message){
    }
    public void onEventMessage(String message){
    }
    ...
}
```

図 5 Javaによる拡張 (PluginListener を実装)

■ イベントメッセージの例
NFC_EVENT_DETECT|(ESID)
 (ESID)のタグを感知したとき
CALENDAR_EVENT_GET|(alias)|(num)|(schedule)
 (num)件の 予定を受信したとき

■ コマンドメッセージの例
CALENDAR_GET|(calendar alias)|(date)
 (date)の予定を獲得する
BROWSER_ADD|(URL)
 (URL)の HP を表示する

図 6 Android 固有のメッセージの例

```
1 10 NFC_EVENT_DETECT|04716c6a472980 <eps>
10 11 <eps> SYNTH_START|mei|normal|音声認識・・・
11 1 SYNTH_EVENT_STOP|mei <eps>
```

図 7 NFC タグのタッチイベントに基づいた処理の例

メッセージとして出力する。

また NFC のタッチイベントのように特定のイベントをトリガーとして実行される仕組みの開発も行った。具体的には、NFC タグを端末が検知すると “NFC_EVENT_DETECT|(ESID)” という形式の内部イベントメッセージを出力し、他のモジュールに通知することが可能になる。これにより、たとえば、ユーザがスマートフォンをタグにタッチすると、そのタグの ID に応じた音声やモーションの再生を行うことが可能である。NFC のタ

イベントに基づく FST スクリプトの制御例を図 6 に示す。この例では、特定の ESID を持つ NFC タグを検知すると、状態番号 10 に遷移し、そのタグに応じた音声を再生することができる。

4.2 通信系の機能

通信系の機能としてファイルダウンロード機能を実装した。“DOWNLOAD|(alias)|(local file name)|(URL)”の形式の内部メッセージを受信すると、URL にあるファイルをダウンロードして(local file name)の位置にそのファイルを保存することが可能になる。これにより、動的に、インターネット上の情報を獲得することが可能になる。なお、ファイルのダウンロードを完了時や失敗時にも “DOWNLOAD_EVENT_STOP” イベントや “DOWNLOAD_EVENT_FAIL” イベントが発生し、これによりダウンロード失敗時の例外処理などを FST に記述することが可能になる。

4.3 GUI 系の機能

GUI 系の機能として、ブラウザ機能と標準的な GUI を用いた機能を実装した。

ブラウザ機能では、“BROWSER_ADD|(URL)”という内部コマンドメッセージを受信すると、画面内に、(URL)の Web ページを表示させることが可能である。

また、Android OS 標準の文字入力フィールドや選択フィールドなどを画面内に表示させることによって対話的な操作も可能である。実際の画面例を図 7 に示す。これは、Android の標準 GUI を画面の上に重ねて表示することにより、文字入力や項目の選択を可能にした。Preview ボタンを押すと、入力した内容の音声対話コンテンツを再生することが可能になる。

4.4 OS サービス系の機能

OS サービス系の機能として、カレンダー連携機能を実装した。これは、メイちゃんに“今日の予定は?”と聞くと、Android OS 標準のカレンダーシステムの API を参照して、今日の予定の一覧を音声で教えてくれるシステムである。

具体的には、“CALENDAR_GET|(alias)|(date)”という内部コマンドメッセージを受信すると、その日のスケジュールの件数や、コメントを、“CALENDAR_EVENT_GET” イベントメッセージとして獲得することが可能になる。これらの内部メッセージを処理する FST スクリプトを記載することによって、音声秘書アプリの作成も可能になる。

5. 関連研究

MMDAgent を利用した音声対話システムとして、全天候型双方向音声案内デジタルサイネージ[10]とモバイルメイちゃんが挙げられる。

全天候型双方向音声案内デジタルサイネージは、名古屋工業大学の正門前に設置されたデジタルサイネージ[11]上に実装され、等身大 3D キャラクタ「メイちゃん」が表示



図 8 GUI を利用したプラグインの例

されている。学生や教職員が Web ブラウザを用いてイベントパネルを投稿することが可能[12]であり、それらに基づいて動的にスクリプトを生成することが可能である。

モバイルメイちゃん[13]は、スマートフォン等のビデオ通話機能を用いて、3D エージェントとの音声対話を実現したシステムである。サーバ上に、Skype API と MMDAgent を連携したシステムが構築されており、ユーザがビデオ通話を開始することによって、音声対話を実現している。

3D キャラクタを表示した初期の音声対話システムとしては、Talkman [14]がある。Talkman は、人間と自然なインタラクションを行うエージェントを目指して構築されており、人間的な顔と表情を持ち、音声認識技術による音声対話を実現している。キャラクタを用いた音声対話システムとしては他に、たけまるくんやキタちゃん[15]もある。音声対話システムの実用化例としては、Let's GO バスシステム[16]、京都市バス運行情報案内システム[17]などがあり、携帯電話などから音声通話でバス運行情報を問い合わせることが可能である。

擬人化エージェントツールキットとしては、本研究で利用した MMDAgent の他に、Galatea ツールキット[18]が存在する。

6. おわりに

本稿では、MMDAgent を拡張し、GPS や NFC、GUI などといったスマートフォンならではの機能を効果的に活用するための仕組みを提案した。具体的には、異なる言語間の連携が困難であるという問題に対して、ブリッジモジュールと PluginListener を新たに開発することによって、容易に C++言語で開発された MMDAgent の各モジュールと連携可能な Java プラグイン開発プラットフォームを提案した。ま

た, 実際に, このプラットフォームに基づく, センサー系のプラグイン, 通信系のプラグイン, GUI系のプラグイン, OS サービス系のプラグインなどを開発した. 今後の課題は, Java プラグイン開発プラットフォームを利用し, より柔軟により面白くネットワークと連携する新しいサービスの開発などが挙げられる.

謝辞 本研究は, 独立行政法人科学技術振興機構 CREST 「共生社会に向けた人間調和型情報技術の構築」による支援を受けた.

参考文献

- 1) Apple Inc., Siri, <http://www.apple.com/ios/siri/> (accessed October 26, 2012)
- 2) NTT ドコモ, シャベってコンシェル, http://www.nttdocomo.co.jp/service/information/shabette_concier/ (accessed October 26, 2012)
- 3) 李 晃伸, 大浦 圭一郎, 徳田 恵一, 魅力ある音声インタラクティブシステムを構築するためのオープンソースツールキット, 情報処理学会研究報告, Vol. 2011-SL-89, pp. 1-6 (2011)
- 4) 山本 大介, 大浦 圭一郎, 西村 良太, 打矢 隆弘, 内匠 逸, 李 晃伸, 徳田 恵一, "スマートフォン単体で動作する音声対話 3D エージェント「スマートメイちゃん」の開発", インタラクティブ ショウ 2013, IPSJ Symposium Series Vol. 2013, No. 1, pp. 675-680, 東京, (2013)
- 5) Lee, A. and Kawahara, T.: Recent Development of Open-Source Speech Recognition Engine Julius, APSIPA, pp. 131137 (2009)
- 6) 大浦 圭一郎, 酒向 慎司, 徳田 恵一: 日本語テキスト音声合成システム Open JTalk, 日本音響学会春季講義集, Vol. 1, No.2-7-6, pp. 343344 (2010)
- 7) MikuMikuDance, <http://www.geocities.jp/higuchuu4/index.htm>, (accessed October 26, 2012)
- 8) Bullet Physics, <http://bulletphysics.org>, (accessed October 26, 2012)
- 9) Google, Android NDK, <http://developer.android.com/tools/sdk/ndk/index.html> (accessed October 26, 2012)
- 10) 大浦 圭一郎, 山本 大介, 内匠 逸, 李 章伸, 徳田 恵一, キャンパスの公共空間におけるユーザ参加型双方向音声案内デジタルサイネージシステム, 特集「音声対話」, 人工知能学会誌, Vol.28, No.1 (2013) (in press)
- 11) 川出 陽一, 双方向音声案内デジタルサイネージ, 印刷雑誌, Vo.94, No.10, pp.25-29 (2011)
- 12) 山本 大介, 大浦 圭一郎, 李 晃伸, 打矢 隆弘, 内匠 逸, 徳田 恵一, 松尾啓志, 双方向音声デジタルサイネージのための学内イベント登録システム, 大学 ITC 推進協議会 2011 年度年次大会, CD-ROM, 2011 年 12 月.
- 13) 打矢 隆弘, 山本 大介, 柴川 元宏, 吉田 真基, 西村 良太, 内匠 逸, ビデオ通話型音声対話サービス「モバイルメイちゃん」の開発, JAWS2012 講演論文集 (2012)
- 14) Nagao, K., Takeuchi, A.: Speech dialogue with facial displays: Multimodal human-computer conversation, Proc. ACL-94, pp. 102-109 (1994)
- 15) 鹿野清宏, Tobias, C., 川波弘道, 西村竜一, 李 晃伸: 音声情報案内システム「たけまるくん」および「キタちゃん」の開発, 情報処理学会研究報告, 2006-SLP-63-7 (2007).
- 16) Raux, A., Bohus, D., Langner, B., Black, A.W. and Eskenazi, M.: Doing Research on a Deployed Spoken Dialogue System: One Year of Let's Go! Experience, Proc. Int'l Conf. Spoken Language Processing (INTERSPEECH) (2006)
- 17) 駒谷 和範, 上野 晋一, 河原 達也, 奥乃 博, ユーザモデルを導入したバス運行情報案内システムの実験的評価, 情報処理学会研究報告, Vol. 2003-SLP-75, pp. 59-64 (2003)
- 18) 川本 真一, 下平 博, 新田 恒雄, 西本 卓也, 中村 哲, 伊藤 克亙, 森島 繁生, 四倉 達夫, 甲斐 充彦, 李 晃伸, 山下 洋一, 小林 隆夫, 徳田 恵一, 広瀬 啓吉, 峯松 信明, 山田 篤, 伝康 晴, 宇津呂 武仁, 嵯峨山 茂樹, カスタマイズ性を考慮した擬人化音声対話エージェントツールキットの設計, 情報処理学会論文誌, Vol. 43, No. 7, pp.2249-2263 (2002)