オープンキャンパスにおけるモバイル端末向け音声対話システムの活用

打矢隆弘，山本大介，柴川元宏，吉田真基，西村良太，内匠逸，松尾啓志

名古屋工業大学大学院工学研究科

t-uchiya@nitech.ac.jp

概要：本研究ではオープンキャンパスにおいて高校生に学科案内やキャンパス案内を行う音声対話システムを開発した．このシステムはスマートフォンと VoIP を用いて利用場所を問わずに音声対話が利用可能であり，オープンキャンパス参加中の建物案内や大学周辺の地理案内に最適である．また，利用者は大学に関する情報を音声で気軽に入手することができる．本稿では，今年度のオープンキャンパスでの利用実績とユーザビリティの評価実験結果について報告する．

1　はじめに

近年，大型ディスプレイを用いて映像や文字を表示するデジタルサイネージシステムが急速に普及している．デジタルサイネージは店舗や公共施設，学校内などに設置され，商品やサービスの広告だけでなく，列車の運行状況や学生の呼出情報など，社会的にも重要な情報も掲示している．デジタルサイネージではユーザからの入力装置を持たず，デジタルサイネージからユーザへの一方向な情報提供を行うものが多いが，双方向の対話が可能なデジタルサイネージも開発されつつある[1]．

名古屋工業大学(以下，本学) 正門前に 2011 年 4 月から設置されている双方向音声案内デジタルサイネージ「メイちゃん」[2, 3]では，ユーザは音声入力により学内案内やイベント案内などの音声対話サービスを利用することができる．

しかし，現在の音声対話サービスはデジタルサイネージを使用した設置型サービスのため，利用場所の制限やデジタルサイネージ設置/運用/保守に掛かる費用などの問題がある．そこで本研究では，スマートフォンと VoIP(Voice over IP: 音声通信を IP 上で行う技術)を用いて利用場所を問わずに音声対話を行うことが可能な，モバイル端末向け音声対話システム「モバイルメイちゃん」を開発した．さらに，提案システムを活用して情報案内を行う具体的な事例として，オープンキャンパスにおいて高校生に学科案内やキャンパス案内を行う音声案内サービスを公開した．本稿では，本学の双方向デジタルサイネージで利用されている音声対話技術を説明し，この技術を応用した提案システムの設計･実装について述べる．さらに，オープンキャンパスにおける実証実験とユーザビリティの評価実験により，提案システムの有効性を実証する．

図 1　双方向音声案内デジタルサイネージ「メイちゃん」

2　双方向音声案内デジタルサイネージ「メイちゃん」

2.1 ハードウェア

本学には2011年4月から双方向音声案内デジタルサイネージが設置されている(図1)．70インチの

液晶ディスプレイを2台垂直に並べ，その前方にマイクブースが設置されている．また，マイクブースの下部には人感センサが設置されており，ユーザがマイクの前に来たかどうかを感知することが出来る．さらに，デジタルサイネージの上部にカメラを備え，顔認識技術等を用いて複数のユーザの状況を認識することが可能である．ディスプレイには等身大の仮想キャラクタ「メイちゃん」が表示される．本ハードウェアは，本学の正門前の屋外に設置され，多くの学生や教職員，来校者が利用可能である．

2.2 ソフトウェア

音声対話エージェントシステムとして名古屋工業大学国際音声技術研究所が開発した「音声インタラクションシステム構築ツールキット MMDAgent」[4](図2)が採用されている．MMDAgentとは，音声対話のための高度な機能を備えた基盤ソフトウェアであり，音声認識，音声合成，3Dモデルの描画や制御，対話管理を統合したシステムである．音声合成は「Open JTalk」，音声認識には「Julius」というモジュールを利用している．

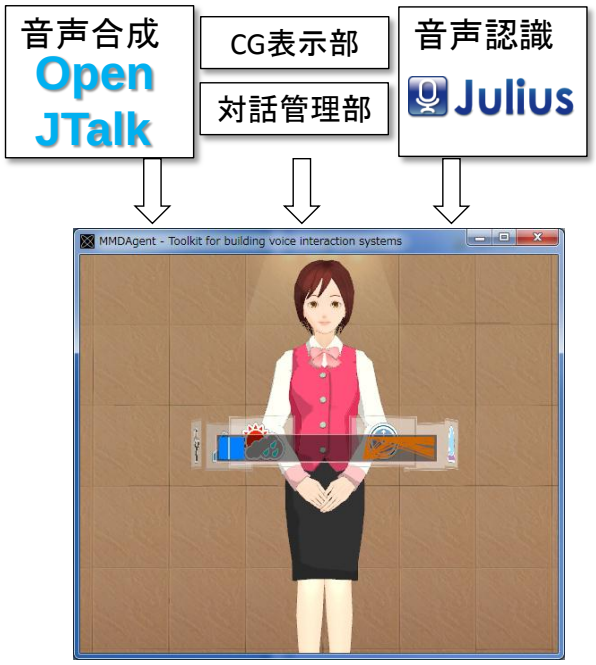

図 2　音声対話エージェントシステム MMDAgent

・Open JTalk

「Open JTalk」[5]は本学の国際音声技術研究所が開発した統計モデルに基づいた音声合成基盤ソフトウェアであり，感情音声を表現することができる．これによりユーザと対話する「メイちゃん」の嬉しい/悲しいといった感情をユーザは音声で感じることが可能である．

Open JTalkでは統計的パラメトリック手法である隠れマルコフモデル（HMM）に基づく音声合成手法を採用し，その音響モデル構築には HMM 音声合成のための研究基盤ソフトウェアである HMM-based speech synthesis system (HTS)[6]を用いている．

HTSの特徴は以下のとおりである．

- 隠れマルコフモデルに基づく新しい音声合成
- 少量のデータで音声を合成可能
- 多様な話者性・話者スタイル・感情表現の実現
- 多言語での動作実績

・Julius

「Julius」[7]も本学国際音声技術研究所で開発された汎用大語彙連続音声認識エンジンである．高性能な音声認識を実現し，現在学術・応用の両面で広く普及している．

Juliusの特徴は以下のとおりである．

- 商用エンジンと同様の能力
- フリーでは世界一
- 高い拡張性・オープンなインタフェース（組込用プロセッサでも動作）
- 国内ではほとんどの音声研究開発機関で採用
- 多言語での動作実績

・CG表示部

「CG表示部」は3Dモデル「メイちゃん」の描画や制御を実行する．合成音声と同期して口を動かすリップシンク機能や，発話内容に即して顔の表情や体全体の動作を変化させる機能を有する．

・対話管理部

「対話管理部」には有限状態トランスデューサ(Finite State Transducer :FST)を採用している．ユーザと「メイちゃん」との音声対話は，メッセ

ージを入出力する状態遷移の形で，専用の FST ファイルに記述される．具体的には，特定のキーワードの認識やセンサによるユーザの検出をトリガーとして，メイちゃんがどの状態に遷移し，どのようなアクション(発話・モーション変化等)を実行するかが記述されている．

3 モバイル端末向け音声対話システム「モバイルメイちゃん」の開発

3.1 検討項目

本研究では，スマートフォンを用いて利用場所を問わずに音声対話を行うことが可能な，モバイル端末向け音声対話システム「モバイルメイちゃん」を提案する．システム開発に先だち，「基盤音声対話システム」「システム構成」「音声コーデック」「利用メディア」について検討を行った．

・「基盤音声対話システム」

ユーザとシステムが音声で円滑にコミュニケーションを行うためには，高性能な音声認識と音声合成のメカニズムが必要となる．また，システムが擬人化されており，ユーザが抵抗なくシステムと対話できることが望ましい．本研究では，本学のデジタルサイネージで利用実績のある MMDAgent を基盤音声対話システムとして採用した．

・システム構成

システムの構成として，「スタンドアロン型」と「ビデオ通話型」の 2 種類を検討した．

「スタンドアロン型」はスマートフォン単体で MMDAgent を動作させ音声対話を実現するタイプである．単体で動作し対話の遅延が無いことから対話システムとして理想的である．一方，スマートフォンで音声認識や音声合成，3D モデルの表示を行うため，端末には高い処理能力が要求される．

「ビデオ通話型」は，ネットワークで接続されたサーバ上で MMDAgent を動作させ，ユーザがスマートフォンから MMDAgent に対してビデオ通話を行い音声対話を実行するタイプである．スマートフォン上で動作するビデオ通話のアプリには Skype などがあり，処理能力の低い端末でも安定動作する．一方，無線ネットワークを介してスマートフォンとサーバが通信を行うため，広帯域の無線ネットワーク環境が必要とされる．

本研究では端末の種類によらず音声対話システムを利用できることを最優先事項とし，「ビデオ通話型」のシステム構成を採用した．

・音声コーデック

「ビデオ通話型」のシステムでは，スマートフォンとサーバの間で音声データのやり取りが行われる．人間の発声はスマートフォン上で音声コーデックにより音声符号化され，圧縮データとしてサーバに送信される．サーバは圧縮データを伸張することで元の音声データを獲得する．音声コーデックにはサンプリング周波数・ビットレート等の異なる様々なコーデックが存在し，音声品質もコーデックによって異なる．

本研究では，SILK，G.711，Speex といった複数の音声コーデックを用いて音声対話のテストを行い，音声品質の優れていた SILK を音声対話コーデックとして採用した．SILK はビデオ通話アプリ Skype で利用されており，使用可能なネットワーク帯域に応じて音声品質を自動調節する仕組みが備わっている．

・利用メディア

音声対話サービスの基本となるメディアは音声メディアであるが，音声で表現しにくい情報は映像メディアあるいは文字メディアで伝えることができることが望ましい．例えば，大学のキャンパス案内において建物の位置をユーザに案内する場合，音声情報に加えて映像で地図を表示することで，ユーザは目的の建物を発見しやすくなる．また，対話中に Web のコンテンツを引用して案内するような場合には，コンテンツの URL を文字データとしてユーザに届けることで，ユーザはコンテンツの参照が容易になる．

そこで本研究では，「映像」「音声」「文字」の 3 つのメディアを用いて音声対話が行える環境を整備する．

サーバからスマートフォンに送信する「映像」については，VP8，H.264/SVC，H.263，H.263p 等のビデオコーデックの中から，映像の滑らかさに優れていた VP8 を採用する．このビデオコーデックは Skype で利用されている．

サーバとスマートフォンの間で送受信する「音声」については，前述したとおり，SILK という音声コーデックを利用する．

サーバからスマートフォンに送信する「文字」については，ビデオ通話アプリに備わっているテキストメッセージ送受信機能を利用する．

以上の検討から，本研究ではビデオ通話アプリとして「Skype」を利用する．

3.2 設計・実装

・概要

モバイル端末向け音声対話システム「モバイルメイちゃん」(図 3)は，スマートフォン上のビデオ通話アプリを用いて，ユーザに対して音声対話に基づく各種案内を行う．「モバイルメイちゃん」はデジタルサイネージ版「メイちゃん」同様に，大学案内・建物案内&道案内・時刻案内・天気予報・占いなどを行うことができる．また，鶴舞駅や鶴間公園・飲食店など大学周辺のスポット案内も可能である．ユーザは 3D キャラクタ「メイちゃん」と音声だけでなく視覚的にもコミュニケーションでき，まるで生きているかのような会話を楽しむことが可能である．

実際の利用イメージを図 4 に示す．

図 3　モバイル端末向け音声対話システム「モバイルメイちゃん」

図 4　「モバイルメイちゃん」の利用イメージ

・特徴

「モバイルメイちゃん」の特徴は以下の通りである．

－「独自の音声技術」を採用

MMDAgent を採用することで、画面内の 3D キャラクタ「メイちゃん」がまるで実在する人間のように感情のこもった対話を行う．

－スマートフォン対応

スマートフォンからビデオ通話機能により 3D キャラクタ「メイちゃん」といつでもどこでも音声対話が可能である．

－「映像と音声と文字」の活用

音声だけでなく，パネル画像やジェスチャで視覚的に案内を行う．さらにテキストメッセージも活用する．

・システム構成

図 5 にモバイルメイちゃんのシステム構成を示す．音声対話を実現する MMDAgent「メイちゃん」は大学内のメイちゃんサーバ上で動作している．ユーザはスマートフォンから、サーバ上のメイちゃんに対してビデオ通話を行うことで、メイちゃんとの対話を楽しむことができる．

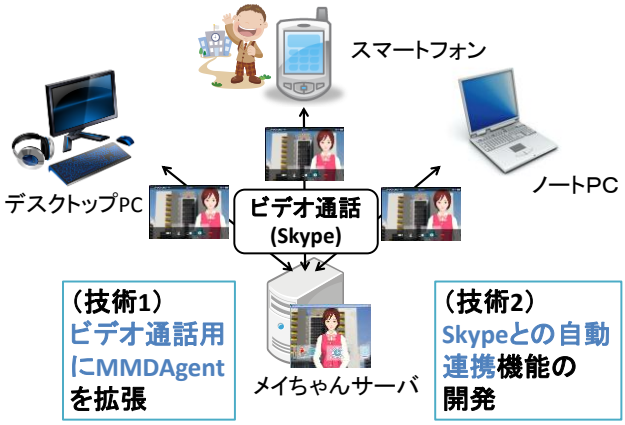

図 5　「モバイルメイちゃん」のシステム構成

今回のシステムではビデオ通話用に MMDAgent を拡張し，さらに Skype との自動連携機能を新規開発した．これら 2 点の技術について詳しく説明する．

・(技術 1) ビデオ通話用 MMDAgent

我々は MMDAgent をモバイル向けに拡張した，「ビデオ通話用 MMDAgent」(図 6)を開発した．この拡張版では，従来の MMDAgent の特徴である「音声認識・感情音声合成」，「CG キャラクタやオブジェクトの表示・制御」等を引き継ぎながら，さらに画面をモバイル向けにカスタマイズした．

また，屋外環境における雑音や無線通信にお

けるパケットロスを想定し，音声認識エンジンを最適化するなどの工夫を施した．

図 6　ビデオ通話用 MMDAgent

・(技術 2) Skype との自動連携機能

MMDAgent と Skype の自動連携機能の概要について説明する．本機能では Skype によるビデオ通話と音声対話システム MMDAgent の自動連携を実現している(図 7)．

[音声メディア] メイちゃんサーバ上の Skype デスクトップ版(以降，Skype-D)と MMDAgent の音声メディアのやり取りには，サーバサウンドカードの再生リダイレクト(ステレオミックス機能)を利用する．具体的には，Skype-D の音声出力(ユーザの音声)を再生リダイレクトにより MMDAgent の音声入力として処理する．また，MMDAgent の音声出力(メイちゃんの音声)を再生リダイレクトにより Skype-D の音声出力として処理する．

[映像メディア] MMDAgent の表示部分を映像として Skype-D でストリーム配信するために，映像キャプチャソフト「ニコ生デスクトップキャプチャ(NDC)」を利用する．Skype-D では，ビデオ設定画面で Web カメラの映像ソースを NDC に設定する．

[文字メディア] Skype-D のテキストメッセージを操作する C#プログラムを新規開発し，MMDAgent に組み込んだ．また，同プログラムと MMDAgent との接続には，MMDAgent ソケット通信プラグインを利用した．これにより，MMDAgent からユーザに対して，文字メディアによる各種案内が可能となった．

以上の工夫により，「モバイルメイちゃん」では音声，映像，文字での対話が新たに行えるようになった．また，会話の状況に即して最適なメディアでコミュニケーションを行える．さらに，必要に応じてメディアの複合利用も可能である．よって，音声だけの各種案内よりも，より表現力の高い案内サービスを提供することが可能となった．

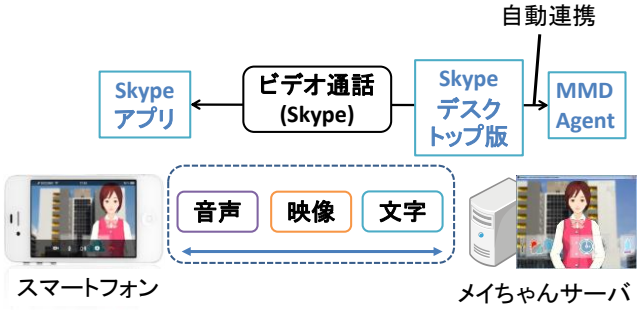

図 7　Skype との自動連携機能

・(その他の工夫)

[自動応答]

ユーザがビデオ通話を開始した際に，Skype-D が自動的に応答を開始するよう設定を行った．これにより，ユーザのスマートフォンの画面には，メイちゃんの映像が即座に表示される．

[アイドル状態の検出]

ビデオ通話開始後，ユーザの発話が一定期間ない状態(アイドル状態)を検出する仕組みを導入した．これにより，ユーザが沈黙状態の場合に，メイちゃん側からユーザに対して呼びかけを行うことができるようになった．

[メイちゃんの状態表示]

Skype を用いたビデオ通話では，同時通話が 1 名に限定されるため，複数人が同時に音声対話サービスを受けることができない．本研究では，他人がこのサービスが利用中かどうかを判断する仕組みを導入した．メイちゃんが既に通話中の場合，C#プログラムが Skype-D のステータスを「取り込み中」にすることで，サービスが他者によって利用中であることをユーザが把握できる．また，通話終了時には Skype-D のステータスを「オンラ

イン」に変更することで，サービスが利用可能であることを把握できる．

4　実験と評価

4.1 オープンキャンパスにおける実証実験

(実験内容)

2011 年 8 月 1 日に開催されたオープンキャンパスの前後 2 週間の期間において，オープンキャンバス参加者に学科案内サービス，及び，キャンパス案内サービスを公開した．なお，本サービスでは一ユーザのサービス占有を防ぐため，一度のサービス利用時間を 3 分に限定した．メイちゃんサーバは 6 台準備し，最大 6 名まで同時接続が可能な環境を構築した．本実証実験では，主に提案システムの動作検証を行った．また，実験終了後，システム利用状況の調査を行った．

図 8　学科リストの表示

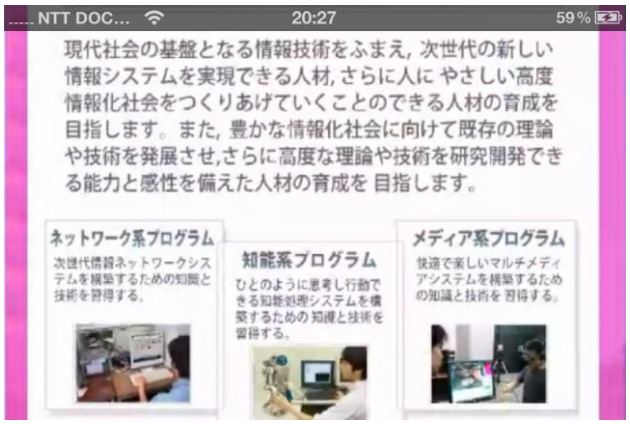

図 9　情報工学科の学科案内

(実験結果 1：学科案内の動作検証)

学科案内では，学科リストの表示→各学科のパネル表示の順で案内を行う．初めにユーザが「名工大にはどんな学科があるの？」と質問を行った．この際，システムはユーザが学科の一覧を必要としていることを認識し，画面に学科リストを表示した(図 8)．次に，ユーザが「情報工学科について教えて！」とシステムに要求を行った．この際，システムは情報工学科の概要を示すパネルを画面に表示し，併せて，音声による学科案内を行った(図 9)．さらに，より詳しい学科の内容をユーザに伝えるため，文字メディアにより学科案内の Web ページをユーザに通知した(図 10)．

図 10　文字メディアによる案内

以上の動作におけるシステムの対話レスポンスは概ね 1 秒以内に収まっており，実システムとして十分な性能を有していることが確認できた．

(実験結果 2：キャンパス案内の動作検証)

キャンパス案内では，ユーザの質問に対し，システムが地図を用いて建物や周辺の飲食店の案内を行う．ユーザが，「学食はどこですか？」とシステムに質問した場合，システムは学食の場所である大学会館を地図で表示する．併せて，音声により正門から大学会館までの道案内を行う(図 11)．

ユーザが，「グルメマップを見せて！」とシステ

ムに要求した場合，システムは大学周辺の飲食店を地図で表示する(図 12).

図 11　大学の食堂案内

図 12　大学周辺の飲食店案内

本システムでは，提示する地図の大きさをスマートフォン向けに最適化しており，地図内の文字が問題なく認識できることが確認できた.

[実験結果 3：システム利用状況]

本実験におけるシステム利用者数，システム利用回数, 平均利用時間, 利用時間帯を以下に示す.

システム利用者数：27
システム利用回数：86
平均利用時間：97.27 秒
利用時間帯：
0 時～3 時：1　　3 時～6 時：0　　6 時～9 時：7
9 時～12 時：12　12 時～15 時：13　15 時～18 時：11　18 時～21 時：16　21 時～24 時：26

システムの利用者数に対して，システムの利用回数が約 3 倍程度になっている．これは，提案システムに魅力を感じたユーザが，複数回にわたってサービスを利用したことが原因であると考えられる．また，平均利用時間も 90 秒を超え，多数のユーザが一度の会話でメイちゃんと複数回のインタクションを行っていることが明らかになった．これは提案システムにおける 3D キャラクタ付きの情報案内が，インタクションの楽しさをユーザに提供しているためと考えられる.
システムの利用時間帯については，午後の時間帯の利用が多いものの，ほぼすべての時間帯で利用されていることが確認できた.

4.2 ユーザビリティの評価

被験者（20 代男子大学生）16 人に対して，ユーザビリティの評価を実施した．被験者は，Apple 社が提供する音声対話アシスタント Siri, モバイルメイちゃんの順にそれぞれ対話を実施した後に，以下の項目に対してアンケートによる 5 段階評価と自由コメントを記述した.

1. 音声対話の応答時間は短かったか？
2. 合成音声の品質は良かったか？
3. 音声認識の精度は良かったか？
4. エージェントに実在感を感じたか？
5. エージェントに魅力を感じたか？
6. 音声対話をして楽しかったか？
7. 音声対話を自然に感じたか ？
8. 映像の品質は良かったか？(提案システムのみ)
9. エージェントを表示させる必要性を感じたか？(提案システムのみ)

5 段階評価の平均結果を図 13 に示す．提案システムは「応答時間の短さ」「合成音声の質」「実在感」「魅力」「楽しさ」の項目で Siri の評価を上回る結果となった．「応答時間の短さ」「合成音声の質」は基盤として採用した MMDAgent の性能の高さを示しており, この 2 点がエージェントの「実在感」を高めることに成功しているといえる．また,「魅力」と「楽しさ」の項目は 3D キャラクタエージェントの存在が非常に大きい．項目 9 の結果も 4.2 と非常に高く，エージェントを画面に表示させる必要性は十分にあるといえる.

一方，「音声認識精度」「自然さ」については，Siri を下回る結果となった．これは提案システムがビデオ通話型のシステムであるため，スマートフォンとサーバ間のネットワークの影響を大きく受けることが原因である．ネットワーク混雑時のパケットロスと遅延の影響を抑え，音声認識精度の低下と映像品質の低下を防ぐことが今後の重要な課題である．

提案システムに対する自由コメントとしては，応答が早い，エージェント表示によって表情などの感情表現ができる，会話の内容が機械的ではなく人間味があって良い，映像があるとエージェントを身近に感じやすいというポジティブな意見があった．その一方で，Siri と比べて応答パターン数が少ない，画面が小さいのでパネルによる情報提示は効果的でない，音量バーが無いので入力されているかどうかがわからない，通信環境が悪い場合映像がカクカクしていて親近感がわかないというネガティブな意見もあった．

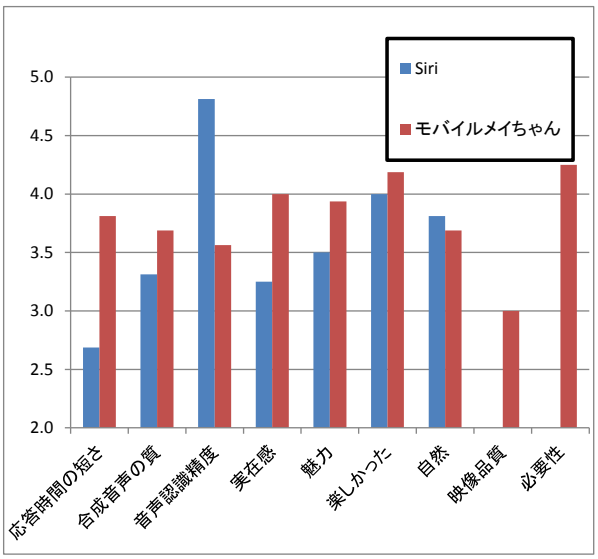

図 13　ユーザビリティに関するアンケート結果

5　おわりに

本研究では，スマートフォンと VoIP を用いて利用場所を問わずに音声対話を行うことが可能な，モバイル端末向け音声対話システム「モバイルメイちゃん」を提案した．またオープンキャンパスにおける実証実験とユーザビリティの評価実験により，提案システムの有効性を実証した．今後はマルチユーザ対応や，音声認識精度向上のための工夫を施していく．

謝辞

本研究の一部は，独立行政法人科学技術振興機構 CREST「コンテンツ生成の循環系を軸とした次世代音声技術基盤の確立」の助成を受けている．

参考文献

[1] 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

[2] 大浦 圭一郎，山本 大介，内匠 逸，徳田 恵一，李 晃伸，「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」，人工知能学会論文誌 28 巻 1 号，(to be appeared in 2013)

[3] 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ICT推進協議会2011年度年次大会講演論文集, 2011

[4] 李 晃伸，大浦 圭一郎，徳田 恵一，「魅力ある音声インタラクション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MMDAgent」，Technical Report of IEICE, Vol.2011-SLP-89, No.27, pp.1-6, 2011

[5] 統計モデルに基づく音声合成基盤ソフトウェアOpen JTalk, http://open-jtalk.sourceforge.net/

[6] HMM-based Speech Synthesis System, http://hts.sp.nitech.ac.jp/

[7] 汎用大語彙連続音声認識エンジンJulius http://julius.sourceforge.jp/