

有限状態遷移の空間的拡張に基づく歩行者向け道案内のための 音声対話シナリオ編集システム

林 晃大[†] 山本 大介[†] 高橋 直久[†]

[†] 名古屋工業大学大学院工学研究科情報工学専攻 〒466-8555 愛知県名古屋市昭和区御器所町

E-mail: [†]hayashi@moss.elcom.nitech.ac.jp, ^{††}{yamamoto.daisuke, naohisa}@nitech.ac.jp

あらまし 近年、スマートフォンで利用可能な音声道案内システムが普及している。既存の音声道案内システムに関する研究では、システムから歩行者への一方向の音声案内を前提としている。そこで、歩行者からも話しかけることができる双方向の音声道案内が実現できれば、歩行者がシステムに話しかけることで道案内文の再確認、現在地の確認、進行方向の確認しながら目的地に到達可能な道案内や観光道案内などの実現が可能になると考えた。一方で、音声対話システムの分野では、双方向の音声対話システムに関する研究が広く行われおり、その中には、FST(有限状態遷移)形式で記述された音声対話シナリオをユーザ生成する研究も含まれる。本研究の目的は、音声対話の文脈や位置情報を考慮した、双方向の音声道案内シナリオの作成と利用を実現する手法を提案することである。具体的には、既存のFSTに基づく音声対話シナリオのユーザ生成手法を拡張し、新たに、対話的な文脈と空間的な文脈の両方を扱うSpatial-FST手法、空間的なイベントを処理可能なSpatial-Event手法、Spatial-FSTのテストを容易に実現可能にするSpatial-Test手法を提案する。本論文では、これらの提案手法に基づくプロトタイプシステムをAndroid Tablet上で開発を行い、提案手法の有効性や問題点について議論した。

キーワード 音声道案内, 有限状態遷移, 位置情報, ユーザ生成

1 はじめに

近年、Google Maps [1] など、モバイル端末上で利用可能な音声道案内システムが普及している。既存の音声道案内システムでは曲がり角等において「右折してください」など、システムから歩行者への一方向の案内しか行われていなかった。歩行者は、位置または進行方向を確認するために、システムに話しかけることができないため、進行方向が本当に正しいのか不安を感じてしまう問題があった。双方向の音声道案内が実現できれば、歩行者がシステムに話しかけることで道案内文の再確認、現在地の確認、進行方向の確認しながら目的地に到達可能な道案内や観光道案内などの実現が可能になると考えた。

一方で、音声対話システムの分野では、双方向の音声対話システムに関する研究が広く行われている。たとえば、MMDA-agent [2] は、FST(有限状態遷移)形式の対話シナリオを編集することによって、双方向の音声対話システムを構築することができる。EFDE [3] は、FSTをグラフとして表示・編集・利用することが可能なAndroidタブレット向けの統合開発環境である。

本研究の目的は、音声対話の文脈や位置情報を考慮した、双方向の音声道案内シナリオの作成と利用を実現する手法を提案することである。これにより、地図上に簡単な経路と案内情報を記述するだけで、双方向の音声道案内を可能にする対話シナリオの作成と利用が可能になる。具体的には、従来手法であるEFDEを、中心座標、半径からなる円と両端点の座標、幅からなる長方形で定義された領域を用いて、歩行者がどの領域にい

るか、入退したか、近づいているかまたは遠ざかっているかといった空間情報に基づく対話を扱うように拡張することを考えたが、以下に述べるような課題があった。

問題点 1 従来手法は(EFDE)は、対話の文脈に基づく状態遷移図のみ対応していた。提案手法は、対話の文脈だけでなく空間情報に基づいて、状態遷移図を作成する必要がある。

問題点 2 従来手法(EFDE)は、音声認識に基づくイベントのみに対応しており、空間情報に基づくイベント処理に対応していなかった。

問題点 3 提案手法は歩行者の空間情報によって対話内容が変化するため、実際に移動してデバックする必要があるが面倒であった。

上記の問題点に対して、FST形式に基づく音声対話システムに、空間的要素を組み合わせた、以下のアプローチで解決する。
アプローチ 1 状態遷移図に、空間情報を持たせた、Spatial-FSTを提案する。

アプローチ 2 歩行者の空間情報に応じたイベントとして、Spatial-Eventを提案する。

アプローチ 3 Spatial-FSTで作成された双方向の音声道案内シナリオをAndroidタブレット上でテスト可能なSpatial-Testを提案し、作成、テスト、実行の3つが一つのシステムで行える編集インタフェースを実装した。

上記のアプローチにより、対話の文脈だけでなく空間情報に基づいて、状態遷移図を作成できる。また、Spatial-FSTの構成要素であるSpatial-stateの領域ごとに状態遷移を記述することで、領域ごとにユーザとMMDA-agentの対話を記述が可能となり、Spatial-Eventを利用してSpatial-stateに対するイ

ベント・コマンドを記述することで、従来の音声認識イベントに加え、位置情報と Spatial-state を利用した空間的な情報をもとに状態遷移が可能となる。また、Spatial-Test により、Spatial-FST を用いて作成された双方向の音声道案内シナリオを Android タブレット上でテスト可能となる。

これらの手法を実装したプロトタイプシステムを実装することで、双方向の音声道案内シナリオの作成・利用・テストを可能にする統合開発環境の実現に貢献できるであろう。

2 関連研究

2.1 MMDAgent

音声対話システムの構築のためのツールキットとしては、MMDAgent などが提案されている。MMDAgent は、音声対話システムを実現するために必要な機能、たとえば、音声合成、音声認識、3D モデル描画、対話制御を統合したシステムである。PC だけでなくスマートフォン [4] での動作が可能である。音声認識では Julius [5] を、音声合成では Open JTalk [6] を採用している。対話シナリオの記述形式に FST(有限状態) 形式を採用している。



図 1 MMDAgent

2.2 FST

FST 形式による対話の記述例を図 2 に示す。FST 形式は、各機能部 (音声認識部、音声合成部など) から発生するイベントを入力、各機能部への命令コマンドを出力とした、状態遷移機械として記述可能である。Fig.1. の例では、音声認識機能部が「こんにちは」と認識すると、状態が 1 から 10 に変化する。続けて、モーション開始コマンド (MOTION_ADD)、音声合成開始コマンド (SYNTH_START) を出力し、状態 12 まで遷移する。音声合成終了イベント (SYNTH_EVENT_STOP) が発生すると状態 1 に遷移する。

```

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|こんにちは
12 1 SYNTH_EVENT_STOP|mei <eps>

```

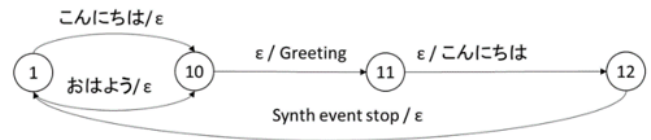


図 2 FST スクリプトの例 (上) とその状態遷移図 (下)

2.3 EFDE

FST 形式の対話シナリオの作成を支援するシステムも提案されている。EFDE は、タブレット端末を用いて状態遷移図を描くことにより、音声対話スクリプトを作成できるシステムである。このシステムでは見かけの状態数の削減のために複合状態を導入した状態遷移モデルを採用している。複合状態は UML のステートマシン図やハレルのステートチャート [7] に採用されており、内部状態を持つ状態のことである。複合状態を採用することで、内部状態を隠蔽することが可能になり、画面上に一度に表示すべき状態の数を減らすことが可能になる。また、テンプレートと入力フォームを用いて内部状態を編集できる仕組み [3] を採用することで、内部状態の編集を容易にした。

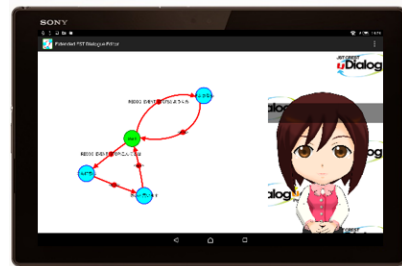


図 3 EFDE

2.4 複合状態を導入した状態遷移モデル

複合状態を導入した状態遷移モデル形式による対話の記述例を図 2 に示す。複合状態を導入した状態遷移モデル形式は、FST 形式と同じく各機能部 (音声認識部、音声合成部など) から発生するイベントを入力、各機能部への命令コマンドを出力とした、状態遷移機械として記述可能である。加えて、図のように複合状態を利用して一連の状態遷移を一つの状態にまとめることで見かけの状態数を削減することができる。

また、複合状態の出口を表す exit point を利用することで複合状態からの遷移を複合状態の内部状態すべてから出る遷移 (A から 2 の遷移) と exit point に対応する内部状態からの遷移 (A.Exit から 1 の遷移) の 2 種類を表現することができる。

図 4 の例では、数字を状態、英字を複合状態とする。また、3 行目は複合状態 A の開始状態と Exit point に対応する状態を表している。Event1 を入力すると Command1 が出力され、状態が状態 0 から複合状態 A に遷移し、複合状態 A の開始状態である A.0 に遷移する。続けて、Event2, Event3 を入力す

ると Command2, Command3 が出力され、複合状態 A の内部状態である A.1, A.2 に遷移する。次に、A.2 は exit point に対応しているため、Event5 が入力されると Command5 を出力して状態 1 に遷移する。また、現在の状態が複合状態 A の内部状態 A.1 である場合、Event6 が入力されると Command6 を出力して状態 2 に遷移するが、Event5 が入力されても exit point に対応していないため、状態 1 に遷移しない。

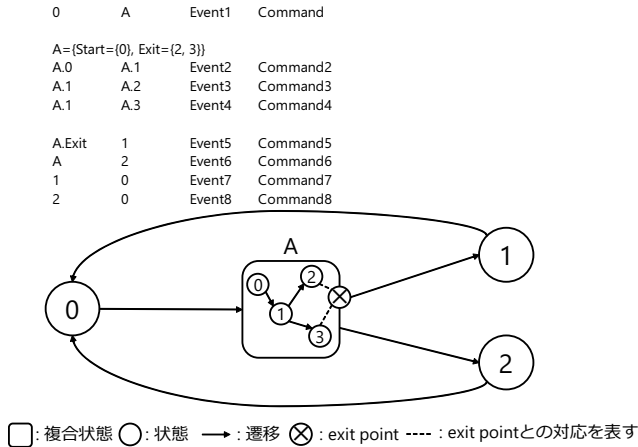


図4 複合状態を導入した状態遷移モデル（上）とその状態遷移図（下）

2.5 その他の関連研究

MMDAE [8] は、Web ブラウザを用いて編集可能な音声対話シナリオのエディタである。また、西村ら [9] は、Web ブラウザを用いて一問一答形式の音声対話をユーザ生成形式で生成する仕組みを提案した。また、FST を音声言語理解のための中間言語として用いる手法 [10] [11] も提案されている。音声対話シナリオの作成する方式として、XML を用いた Voice XML [12] も提案されている。

音声道案内に関する研究も提案されている。柳ら [13] は、音声道案内のためのランドマーク情報をソーシャルアノテーションとして収集し、それらを利用した音声道案内システムを提案した。八木ら [14] は、仮想地図のような概念から音声案内文を生成する概念音声合成 (Concept-to-speech) を提案した。IFTTH [15] の Location trigger は、指定した位置の範囲内に入った時と出た時にイベントを生成するなど、位置情報イベントを扱うことが可能である。

3 提案手法の概要

本節では、1 章で述べた 3 つの問題点を解決するために、以下の手法を提案する。

手法 1 手法 1 は、領域の形状を紐づけた状態 (Spatial-state) を導入した状態遷移図である。これにより、対話の文脈だけでなく空間情報に基づいて、状態遷移図を作成でき、Spatial-Event に必要な領域情報を登録することができる。さらにプロトタイプシステムを用いて地図画面上で作成することで、Spatial-state の領域情報を自動取得することで作成者の負担を軽減した。ま

た、複合状態を採用することで Spatial-state の領域ごとに状態遷移を記述できるようになったため、領域ごとに歩行者と MMDAgent の対話を記述できる。

手法 2 歩行者の空間情報に応じたイベントとして、Spatial-Event を提案する。これにより、Spatial-Event を利用して Spatial-state に対するイベント・コマンドを利用することで、従来の音声認識イベントに加え、位置情報と Spatial-state を利用した空間的な情報をもとにした状態遷移が扱うことができる。

手法 3 Spatial-FST を用いて作成された双方向の音声道案内シナリオを Android タブレット上でテスト可能な Spatial-Test を提案する。これにより、歩行者の疑似的な経路 (緯度経度座標) を地図画面上を一本指でなぞることで入力することができ、実際に歩行することなく Spatial-FST で作成された音声道案内シナリオをテストすることができる。

4 実現方法

本節では、3 章で述べた手法とその実現方法について説明する。

4.1 Spatial-FST

双方向の音声道案内シナリオを実現するために、対話の文脈だけでなく、地理的な文脈に基づいた FST を作成できる必要がある。しかし、1 章の問題点 1 で述べたように、位置情報と状態の対応付けに課題があった。そこで、位置情報と状態の対応付けについては領域を持つ状態である Spatial-State を新たに定義し、その状態を画面上で地図と関連付けられるようにした。また、複合状態を採用することで、みかけの状態数を削減した。

通常の状態、複合状態、Spatial-state の違いを表 1, Spatial-state のデータ構造を表 2 に示す。

表 1 各状態の違い

	通常の状態	複合状態	Spatial-state
MMDAgent で実行可能	○	×	×
内部状態を持つ	×	○	○
領域を持つ	×	×	○

表 2 Spatial-state のデータ構造

Name	Type	Description
ID	int	Spatial-state の ID
Area	ArrayList<Double>	領域の緯度経度座標と半径
child_list	ArrayList<State>	内部の状態のリスト
transition_list	ArrayList<Transition>	内部の遷移のリスト
start	State	内部の開始状態を表す要素
end_states	ArrayList<State>	内部の終了状態のリスト

図 5 の例は、Spatial-FST と次節で説明する Spatial-Event を導入した状態遷移モデルである。数字を状態、英字を Spatial-state とする。また、3 行目は Spatial-state A の開始状態と領域を表している。次節で説明する Spatial-Event (SPATIAL_EVENT_IN|A) が入力される

と、状態 0 から Spatial-state A に遷移し、Spatial-state A の開始状態である A.0 に遷移する。続けて、音声合成開始コマンド (SYNTH_START) を出力し、Spatial-state A の内部状態 A.1 まで遷移する。音声合成終了イベント (SYNTH_EVENT_STOP) が発生すると Spatial-state A の内部状態 A.2 に遷移する。さらに、現在の状態が Spatial-state A である場合に、次節で説明する Spatial-Event (SPATIAL_EVENT_OUT|A) が入力されると、音声合成開始コマンド (SYNTH_START) を出力し、Spatial-state A から状態 1 に遷移する。音声合成終了イベント (SYNTH_EVENT_STOP) が発生すると状態 1 から状態 0 に遷移する。

0	A	SPATIAL_EVENT_IN A	<eps>
A={Start={0}, Area={lat, lon, R}}			
A.0	A.1	<eps>	SYNTH_START mei normal 右折してください
A.1	A.2	SYNTH_EVENT_STOP mei	<eps>
A	1	SPATIAL_EVENT_OUT A	SYNTH_START mei normal 戻ってください
1	0	SYNTH_EVENT_STOP mei	<eps>

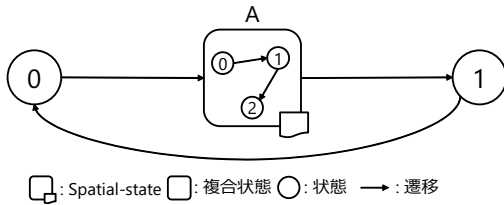


図 5 Spatial-FST と Spatial-Event を導入した状態遷移モデル (上) とその状態遷移図 (下)

4.2 Spatial-Event

従来手法 (EFDE) は、1 章の問題点 2 音声認識に基づくイベントのみに対応しており、空間情報に基づくイベント処理に対応していなかった。そこで、歩行者の位置情報 (GPS) と Spatial-state の領域をもとに歩行者の位置と移動に応じてイベントを発行する Spatial-Event プラグインを提案することで、状態遷移の枠組みで位置情報を扱うことを可能にした。

具体的には、歩行者の領域への入退と移動に関するイベント、Spot までの距離と方向に関するコマンドを定義した。また、さらに、Spatial-Event を生成する機能を MMDAgent のプラグインとして開発した。これにより、提案システムで作成した音声道案内シナリオ (FST 形式) を他端末でも実行可能となる。具体的には、以下に挙げる、コマンドとイベントをサポートする。それぞれのコマンドとイベントの詳細と、そのコマンドとイベントの検出方法について以下に述べる。

4.2.1 Spatial-state の登録に関するコマンド

Spatial-state の領域と名前を MMDAgent に登録するコマンドである。提案システムでは MMDAgent を用いて道案内を行う前に、GPS_SET を利用して各 Spatial-state の情報を登録する。ただし、ID は Spatial-state の識別子、latitude は Spatial-state の緯度、longitude は Spatial-state の経度、radius は Spatial-state の半径とする。

- SPATIAL_SET|ID|latitude|longitude|radius

4.2.2 ユーザの領域の入退に関するイベント

ユーザが Spatial-state に入退したときに、イベントを発行する。Spatial-state の領域に入ったときは SPATIAL_EVENT_IN イベントが、Spatial-state から出た時は SPATIAL_EVENT_OUT イベントが発行される。それぞれのイベントの定義は以下のとおりである。ただし、ID は Spatial-state の識別子とする名前とする。

- SPATIAL_EVENT_IN|ID
- SPATIAL_EVENT_OUT|ID

ある Spatial-state の識別子 ID に対して、イベント発行のアルゴリズムを以下に示す。ただし、領域を R とする。

Algorithm 1 ユーザの領域の入退に関するイベント

```

while x 秒おき do
  if  $C \in R \cap C' \notin R$  then
    SPATIAL_EVENT_IN|ID
  end if
  if  $C' \in R \cap C \notin R$  then
    SPATIAL_EVENT_OUT|ID
  end if
end while

```

4.2.3 ユーザの移動に関するイベント

ユーザが Spatial-state の領域から出て次の Spatial-state に移動中に、Spatial-state に遠ざかる、または近づいたときにイベントを発行する。Spatial-state から L メートル離れるたびに SPATIAL_EVENT_AWAY イベントが、Spatial-state から L メートル近づくたびに SPATIAL_EVENT_APPROACH イベントが発行される。それぞれのイベントの定義は以下のとおりである。ただし、ID は Spatial-state の識別子とする。

- SPATIAL_EVENT_AWAY|ID
- SPATIAL_EVENT_APPROACH|ID

ある Spatial-state の識別子 ID に対して、イベント発行の手順を以下に示す。ただし、領域の中心座標を R_C 、距離の閾値を L_{th} 、x 秒で近づいた距離を L_X とする。

Algorithm 2 ユーザの移動に関するイベント

```

while x 秒おき do
   $L_1 = dist(C', R_C)$ ;
   $L_2 = dist(C, R_C)$ ;
   $L_X = L_2 - L_1$ ;
   $sum += L_X$ ;
  if  $sum > L_{th}$  then
    SPATIAL_EVENT_APPROACH|ID
     $sum = 0$ 
  end if
  if  $sum < -L_{th}$  then
    SPATIAL_EVENT_AWAY|ID
     $sum = 0$ 
  end if
end while

```

4.3 Spatial-FST から FST への変換

Spatial-FST の Spatial-state は、緯度経度座標と半径からなる領域を持つ。しかし、複合状態は領域の情報を扱うことができない。また、Spatial-state はそのまま MMDAgent で実行することができない。そこで、Spatial-Event プラグインで定義された領域を登録する SPATIAL.SET コマンドを利用して、図 6 のように、あらかじめ領域情報 (Spatial-state A の場合、Area={lat_A, lon_A, R_A}) を MMDAgent に登録し、Spatial-state の領域情報を削除することで複合状態への変換を可能にした。また、変換された複合状態は若林の手法 [3] により FST に変換する。これにより、Spatial-FST を FST に変換し MMDAgent で実行可能にした。変換方法は図 6 の例を用いて流れを説明する。

(1) SPATIAL.SET コマンドで Spatial-state の領域情報を登録し、Spatial-state から領域情報を削除し複合状態に変換する。

(2) 若林に手法により、複合状態間の遷移に対して、遷移前の複合状態の全内部状態から遷移後の複合状態の開始状態への遷移に変換することで複合状態を展開し MMDAgent で実行可能な FST に変換する。

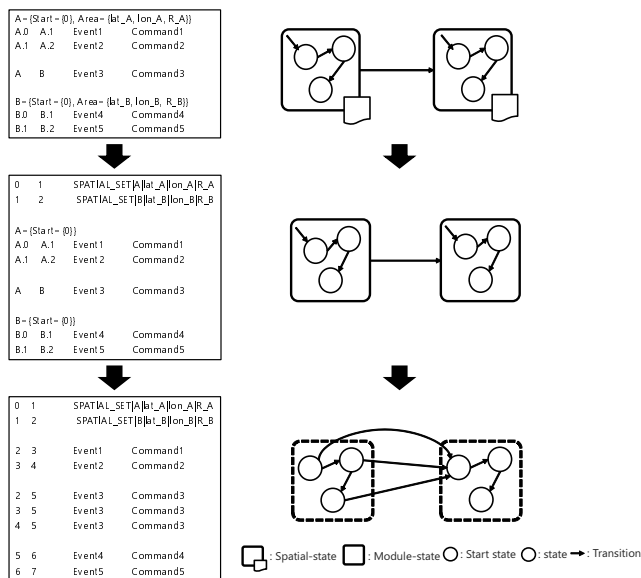


図 6 Spatial-FST から FST への変換

4.4 Spatial-Test

Spatial-Test では、作成者が Test layer を通して地図画面上を一本指でなぞることで歩行者の経路をリアルタイムで入力することが可能になる。また、作成者が音声入力を行うこともできる。これにより、作成者は実際に歩行することなく Spatial-FST に基づいた音声道案内シナリオのテストを行うことができる。

4.4.1 Spatial-Test の構成

Spatial-Test は、一本指でなぞることにより歩行者の経路を入力することができる Test layer と、Test layer から入力された画面座標を地図が表示された Map layer の緯度経度座標に変

換し、MMDAgent に送信する座標変換・送信機能によって構成される。

作成者が入力する経路の取得のために提案システムでは、Test layer, Map layer, 音声道案内コンテンツを作成するための Edit layer に分割することで解決を図った。これにより、テストを行う際には Test layer を図 7 のように最上層に移動させることで Edit layer への誤ったタッチ反応を防ぎ、Test layer に一本指でなぞることで歩行者の経路をリアルタイムで入力することができる。また、音声道案内シナリオを編集する際には、Edit layer を最上層とし、Test layer と入れ替えることで編集を可能にしている。

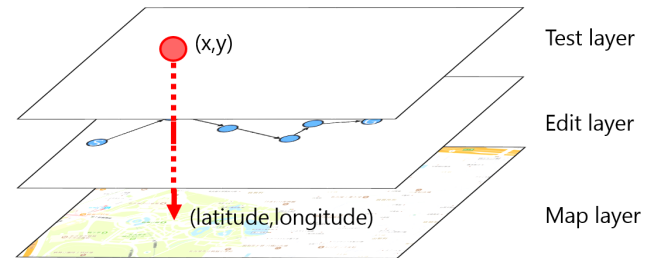


図 7 レイヤ構造

そして、座標変換・送信機能によって Test layer に入力された画面座標 (x,y) を Map layer の緯度経度座標 (latitude, longitude) に変換し MMDAgent に送ることでテストを行っている。

5 プロトタイプシステム

プロトタイプシステムのイメージを図 8 に示す。プロトタイプシステムでは、曲がり角を表す Spot (丸) と 2 つの Spot 間の道を表す Move (矢印) を地図上に交互につなぎ合わせることで、音声道案内シナリオの経路を決定することができる。編集インタフェースで作成できる音声道案内シナリオの経路は一本道で分岐がないものとする。次に各要素において、入力フォームに従って道案内に必要な情報である案内情報を入力する。これにより Spot では曲がり角で有効な対話、Move では 2 つの Spot 間の道、つまり次の Spot に移動しているときに有効な対話が生成され、音声道案内シナリオを作成することができる。

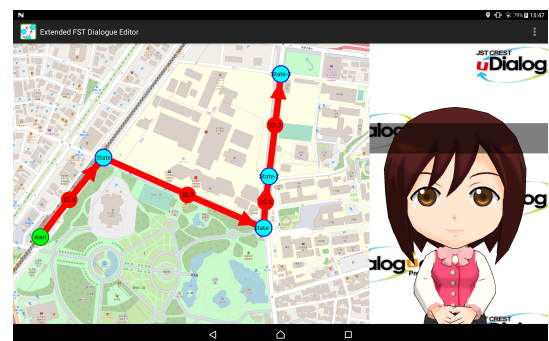


図 8 プロトタイプシステム

5.1 プロトタイプシステムの構成

プロトタイプシステムは、音声道案内シナリオの編集インタフェース、空間イベント生成機能 (Spatial-Event plugin)、FST 変換機能、音声道案内シナリオのテストを行う Spatial-Test 機能から構成される。編集インタフェースで作成された対話シナリオ (Spatial-FST 形式) は、音声道案内シナリオから Spatial-Event plugin で定義されたコマンド (SPATIAL.SET) を用いて FST 変換機能により、MMDAgent で利用可能な FST 形式に変換する。MMDAgent では、変換された FST に基づいて動作する。実際の歩行者の緯度経度座標から Spatial-Event plugin でイベント・コマンドが生成され、FST が遷移することで音声道案内シナリオが実行される。また、作成者は Spatial-Test 機能によって、地図画面上でなぞった歩行者の擬似的な経路が緯度経度座標を変換し、MMDAgent に送信することで、実際に歩行することなく音声道案内シナリオのテストを行うことができる。

プロトタイプシステムの構成図を図 9 に示す。

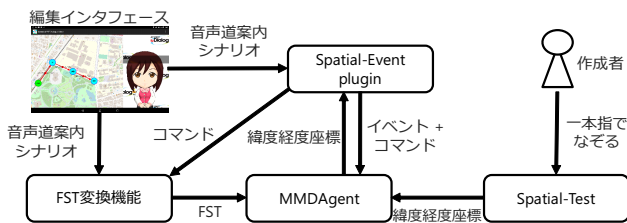


図 9 プロトタイプシステムの構成図

5.2 Spot, Move と Spatial-FST の関係

Spot は円状の領域を持つ Spatial-state、Move は長方形の領域を持つ Spatial-state とその Spatial-state の前後に接続されている遷移で構成される。前後の遷移は Spatial-Event の SPATIAL.EVENT.IN がイベントとなっている。前側の遷移のイベント SPATIAL.EVENT.IN の引数は Move の前に接続された Spot の Spatial-state の ID、後側の遷移のイベント SPATIAL.EVENT.IN の引数は Move の先に接続された Spot の Spatial-state の ID とする。このように Spatial-FST を簡略化し、Spot を曲がり角や交差点、Move をその間の道とすることで、作成者が音声道案内シナリオを作成しやすいようにした。

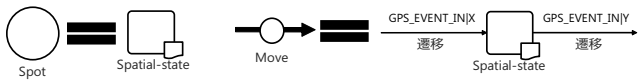


図 10 Spot, move と Spatial-FST の関係

5.3 Spatial-FST テンプレート

プロトタイプシステムでは、Spot と Move の構成要素である Spatial-state を作成するために Spatial-FST テンプレートを利用している。Spatial-FST テンプレートは Spatial-state 内部の状態遷移集合、変数名集合、領域情報で構成される。それに対応する入力フォームは (変数名, 値) の集合と領域情報で

構成されている。入力フォームから変数名に対応する値を入力することで、入力された値と Spatial-FST テンプレートから Spatial-state を作成する。

プロトタイプシステムの Spot の場合では、入力フォームでどのような Spot (ランドマークの有無) であるか、どのような道案内 (一方または双方向) を行うかを選択できるようになっており、計 4 種類の音声道案内を選択できる。また、Spot の領域情報 (中心座標と半径) と Spot の案内情報 (ランドマークや曲がる方向、次までの距離と時間) を入力することができる。今回は作成者が地図画面上に配置した経路の座標から Spot の領域情報と Spot の案内情報の一部を自動入力している。作成者は残りの案内情報 (ランドマーク) を入力することにより、Spot 内で有効な対話を設定する。図 11 の入力フォームから生成される対話は以下の通りである。

- (1) 周囲に関する対話
 - ・ ユーザ「周囲に何がある?」
 - ・ システム「コンビニがあります」
- (2) 曲がる場所と方向に関する対話
 - ・ ユーザ「どこを曲がるの?」
 - ・ システム「コンビニを右に進んでください」
- (3) 次までの時間と距離に関する対話
 - ・ ユーザ「次までどれくらい?」
 - ・ システム「約 400m で 5 分くらいかかります」

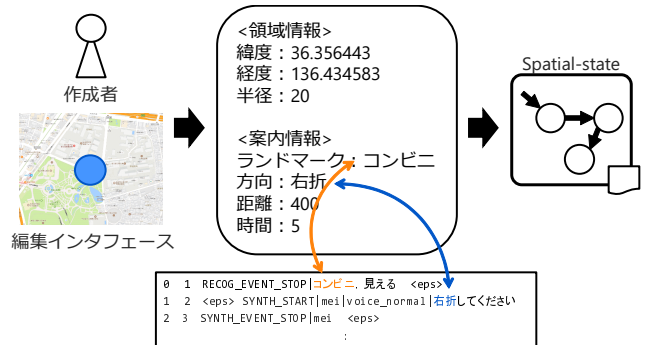


図 11 Spot の入力フォーム

このように作成者は Spot と Move を交互につなぎ合わせて経路を入力し、各要素に案内情報の一部を入力することで MMDAgent で実行可能な音声道案内シナリオを作成することができる。

6 評価実験

6.1 実験の目的

提案手法に基づいて実装したプロトタイプシステムで作成した双方向の音声道案内の有用性と利便性を検証することである。

6.2 実験方法

本学学生 5 名ずつにプロトタイプシステムで作成した一方と双方向の 2 種類の音声道案内を実際に利用してもらい、歩行時間、道を間違えた回数、歩行者が MMDAgent に話しかけた

回数を計測する。また、実験後に 2 種類の音声道案内について 5 段階のアンケートに回答してもらう。

音声道案内を行う経路は、開始地点を名古屋工業大学、目的地点を吹上駅とする図 12 のようなルートで行った。地図画面左下の緑色の Spot が開始地点、右上の Spot が目的地点である。今回は 11 つの Spot と 10 つの Move で作成した。各 Spot と Move に登録したランドマークの例としてはガソリンスタンド、寺、公園などが挙げられる。

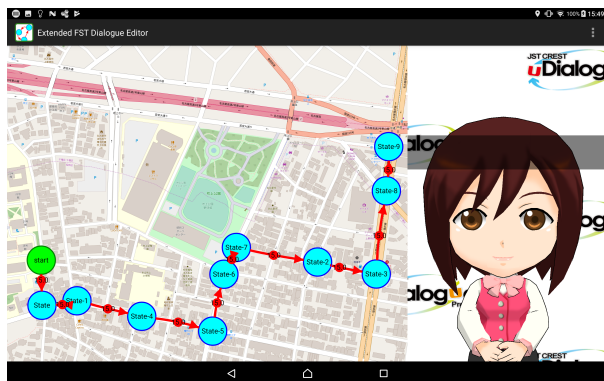


図 12 音声道案内を行う経路

6.3 アンケート内容

アンケートは 5 段階評価と自由コメント欄を用意した。アンケートの項目は以下の通りである。評価は 5 段階でしてもらい、評価基準は「1:あてはまらない、2:ややあてはまらない、3:どちらともいえない、4:ややあてはまる、5:あてはまる」とした。アンケート項目は以下のとおりである。

- (1) 音声道案内の満足度
- (2) 曲がり角で不安を感じた
- (3) 途中の道で不安を感じた
- (4) 道案内で歩行者から話しかけることは必要だと感じた
- (5) また使ってみたい
- (6) システムが話す、返答する内容は適切か
- (7) 案内がわかりやすい

6.4 結果と考察

アンケートの各項目について評価の平均をとったグラフを図 13、歩行時間、道を間違えた回数、歩行者が話しかけた回数の平均を表 6.3 に示す。項目 1 から項目 7 については数値が高いほど良い評価である。

すべての項目において双方向の音声道案内の方が良い評価を得ていることがわかる。特に項目 2、3 では双方向の音声道案内と一方向の音声道案内で評価の差が開いている。また、自由コメントでも、「案内が聞き取れなかったときに再確認できて良かった」、「ランドマークを使って自分の進むべき方向が確認できるのが良かった」、「自分が歩行している道が合っているかの確認ができるのが良かった」といった好感触なコメントが得られた。この結果から、プロトタイプシステムで作成した双方向の音声道案内シナリオは有用であるといえる。歩行時間を比較しても 1 分以上短縮できており、歩行者が話しかけることによ

て歩行者が道を確認でき、迷うことなく歩行できたためであると考えられる。話しかけた回数の平均は 4.8 回となっており、特に歩行者が話しかけていた内容としては、「次の目的地の目印は?」といった次の目印の確認や「(ランドマーク名) が見える」といったランドマークを使った進行方向の確認や「この道で大丈夫」といった道の確認が多かった。また、話しかける場所については次までの目的地が遠いときに特に話しかける傾向があった。これらの結果から、歩行者はシステムからの案内が一定時間ないときに不安を感じ、システムを確認することで不安を解消したいことがわかった。このことから、双方向の音声道案内が有効であることがわかる。

一方で項目 6、7 と道を間違えた回数については、あまり差が開かなかった。これは一方向の案内においてもランドマークを利用して曲がる方向を案内する (例、コンビニの角を右に進んでください) といった音声案内を行っているため、歩行者にとってわかりやすかったこと理由として考えられる。また、双方向の音声道案内を行っている際に、登録されていないランドマークについて話しかける場合があり、システムが何も返答できない場合があった。このような問題に対して、システムにとって未知のランドマークが話された場合に、システムが「(ランドマーク名) があるのですか」と返答し、話しかけられたランドマーク名と話しかけられた領域 (どの Spot、Move) だったのかを記憶し新たに音声道案内シナリオに追加することで、次の案内に活かすような機能が必要であると考えられる。

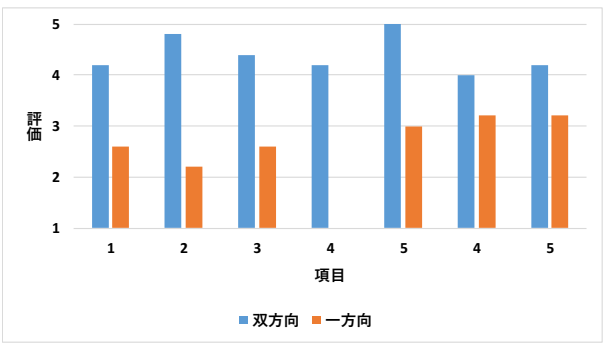


図 13 評価実験 (歩行) の結果

表 3 各項目の平均

	双方向	一方向
歩行時間	16 分 20 秒	17 分 29 秒
道を間違えた回数	0	0.6
歩行者が話しかけた回数	4.8	

7 考 察

従来手法では MMDAgent で道案内を行う際に空間的な情報が扱えなかったため、音声認識イベントのみで状態遷移することで、歩行者に対して道案内を行っていた。しかし、音声認識イベントのみの状態遷移では歩行者のしゃべる場所によっては誤った案内になってしまうため、歩行者が道を間違えてしま

うという問題が発生する可能性がある。

これに対して、提案手法では Spatial-FST と Spatial-Event により、Spatial-state を利用して複合状態に位置情報を持った領域を持たせることができるようになった。これにより、Spatial-state の領域ごとに状態遷移を記述できるようになったため、領域ごとにユーザと MMDAgent の対話を記述できるようになった。また、Spatial-Event を利用して Spatial-state に対するイベント・コマンドを利用することで、従来の音声認識イベントに加え、位置情報と Spatial-state を利用した空間的な情報をもとに状態遷移ができるようになった。

一方で、提案手法では以下のような課題がある。図 14 の左のようにある複合状態から別の状態に遷移し、再度複合状態に遷移する場合である。これは前回、複合状態内部のどの状態から遷移したかを記憶していないため、再度複合状態に遷移した場合に前回遷移した複合状態の内部状態に遷移することができない。この問題を解決するためには、図 14 の右の UML のステートマシン図で利用されている履歴状態を導入して、複合状態から別の複合状態に遷移した際に複合状態のどの内部状態から遷移したかを記憶し、再度複合状態に遷移した際に記憶した状態に遷移する必要がある。

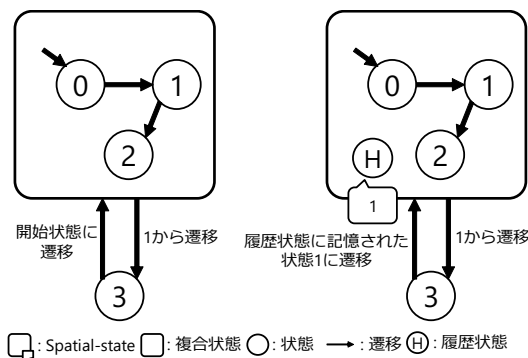


図 14 履歴状態がない場合 (左) と履歴状態がある場合 (右)

8 おわりに

本研究では、既存の FST に基づく音声対話シナリオのユーザ生成手法を拡張し、新たに、空間的なイベントを処理可能な Spatial-Event 手法、対話的な文脈と空間的な文脈の両方を扱える Spatial-FST 手法、地理的なデバッグを容易に実現可能にする Spatial-Test 手法を提案した。これにより、音声道案内シナリオでは状態数が膨大になり編集が難しいという問題や、案内経路と音声対話シナリオとの対応付けが難しいという問題を解決した。本論文では、これらの提案手法に基づくプロトタイプシステムを Android Tablet 上で開発を行い、提案手法の有効性や問題点について議論した。

今後の課題として、開始地点と最終目的地を入力した際に、地図上からランドマークを探索し、ランドマークの多い経路を選択で自動で経路を決定する機能やシステムにとって未知のランドマークが話しかけられた際の対処方法を考え、その機能を実装することが挙げられる。また、提案手法の考察で述べた課

題を解決する手法を考え、新たな状態遷移モデルを提案することが挙げられる。

謝 辞

本研究は JSPS 科研費 25700009, 科学技術振興機構 CREST, および、総務省 SCOPE の助成を受けたものです。

文 献

- [1] GoogleMaps, <https://www.google.co.jp/maps/> 2017-04-19.
- [2] A. Lee, K. Oura, and K. Tokuda, MMDAgent - A fully open-source toolkit for voice in-teraction systems, Proceedings of the ICASSP 2013, pp. 8382-8385, 2013.
- [3] Wakabayashi, K., Yamamoto, D., and Takahashi, N. A voice dialog editor based on finite state transducer using composite state for tablet devices, Computer and Information Science 2015, Studies in Computational Intelligence, Vol. 614, pp.125139, 2016.
- [4] D. Yamamoto, K. Oura, R. Nishimura, T. Uchiya, A. Lee, I. Takumi, and K. Tokuda, Voice Interaction System with 3D-CG Human Agent for Stand-alone Smartphones, Proceedings of the 2nd International Conference on Human Agent Interaction, pp.320-330, 2014.
- [5] A. Lee and T Kawahara, Recent development of open-source speech recognition engine julius, in Proceedings of the AP-SIPA 2009, pp. 131137, 2009.
- [6] Open JTalk, <http://open-jtalk.sp.nitech.ac.jp/> 2019.1.9.
- [7] D. Harel, Statechart: A visual formalism for complex systems, Science of Computer Prgramming, Vol.8, No.3, pp.231-27, 1987.
- [8] R. Nishimura, D. Yamamoto, T. Uchiya, and I. Takumi, Development of a Dialogue Scenario Editor on a Web Browser for a Spoken Dialogue System, Proceedings of the 2nd International Conference on Human Agent Interaction, pp. 129-132, 2014.
- [9] R. Nishimura, D. Yamamoto, T. Uchiya, and I. Takumi, Web-based envi-ronment for user generation of spoken dialog for virtual assistants, EURASIP Journal on Audio, Speech, and Music Processing, 2018:17, 2018.
- [10] C. Raymond, F. Bchet, R. De Mori, and G. Damnati, On the use of finite state trans-ducers for semantic interpretation, Speech Communication, Vol. 48, No.3-4, pp. 288304, 2006.
- [11] C. Hori, K. Ohtake, T. Misu, H. Kashioka, and S. Nakamura, Proceedings of the INTERSPEECH 2008, pp. 211214, 2008.
- [12] VoiceXML Forum, VoiceXML: The standard application language for voice dialogues.
- [13] T. Yanagi, D. Yamamoto, and N. Takahashi, Development of Mobile Voice Navigation System Using User-Based Mobile Maps Annotations, Proceedings of the IEEE/ACS ICIS 2015, pp. 373-378, Las vegas, 2015.
- [14] Y. Yagi, S. Takada, K. Hirose, and N. Minematsu, Realization of Concept-to-speech Conversion for Reply Speech Generation in a Spoken Dialogue System of Road Guid-ance and its Evaluation, IPSJ Journal, Vol.48, No.9, pp.3300-3308, 2007.
- [15] IFTTT, <https://ifttt.com/> 2018.