

高齢者を対象とした音声入力インタフェース評価および対話実験

實 廣 貴 敏[†] 松 田 繁 樹[†] 葦 莉 豊[†]
中 村 哲[†] 矢 入 (江 口) 郁 子^{††} 猪 木 誠 二^{††}

すべての歩行者、特に高齢者や障害者に対し、バリアおよびバリアフリー情報を提供する歩行者支援地理情報システム (GIS) における音声入力インタフェースを提案する。雑音に強い音声認識方式として、マルチコンディション学習と SN 比別音響モデルを用いたパラレルデコーディングを用いた。また、主なユーザである高齢者対策として、高齢者専用音響モデルを用いることで、若年層と同等な認識精度を得ることができた。さらに、この音声入力インタフェースを高齢者がどの程度使えるか、評価実験を行った。マウスとの比較で条件入力の手間時間を 60%縮めることができた。さらにより自然なインタフェースを目指して、Wizard of Oz 法により、同じタスクを対話で行う実験を高齢者に対し行った。複数の対話パターンで対話時間を調べた。対話時間は各被験者の PC 使用経験度合に依存することがわかった。

Evaluation of Speech Interface and Dialog Experiments for Elderly People

TAKATOSHI JITSUHIRO,[†] SHIGEKI MATSUDA,[†] YUTAKA ASHIKARI,[†]
SATOSHI NAKAMURA,[†] IKUKO EGUCHI YAIRI^{††} and SEIJI IGI^{††}

We propose a speech input interface for the Mobility Support Geographic Information System (GIS). It can provide optimal paths to destinations with information on barrier and barrier-free objects related to sidewalks for all pedestrians, especially elderly people and disabled people. For robust speech recognition, our method includes multi-condition training and parallel decoding with SNR-dependent acoustic models. Using acoustic models for elderly people, our method obtained almost the same performance for elderly people as that for younger people. We also evaluated our speech input interface for elderly people. Results indicated that speech input is much easier than mouse input for them. Furthermore, we investigated elderly peoples' behavior toward a dialog system by using Wizard of Oz. In this experiment, we applied several dialog styles. Experiment results show that their dialog durations are dependent on degrees of their own experience with computers.

1. はじめに

近年、国をはじめ、各地方自治体により、歩行空間のバリアフリー化が精力的に進められている。しかし、完全にすべてバリアフリー化することは困難である。そこで、バリア情報およびバリアフリー情報を歩行者に提供することで、移動支援を行うことが考えられる。こういった観点から、情報通信研究機構 (NICT) では、移動支援を行う Robotic Communication Terminals (RCT) の実現を目指している¹⁾。RCT は、道路や駅などに設置される「環境端末」と、ユーザと共に移動する「ユーザ携帯型移動端末」および「ユーザ搭乗型移動端末」の 3 タイプに主に類別される。この内、一つの具体例として、歩行者用経路探索とバリアおよびバリアフリー情報を提供する「歩行者支援地理情報システム (Geographic Information

System: GIS)」の実用化を目指している²⁾³⁾。これは車いす使用などの身体的条件を入力することで障害者・高齢者を含む全ての歩行者が利用可能なものである。

歩行者支援 GIS にはインターネット上の公開サーバ型、計算機上のアプリケーション型がある。クライアントは一般的なパーソナルコンピュータ上で動作し、ユーザは自身の端末上で必要な条件を入力すると、バリアおよびバリアフリー情報を考慮した最適な歩行経路を得ることができる。これまでの検討では一般的な GUI を使い、入力事項をすべてマウス、キーボードで選択するというインタフェースのみを考慮してきた。しかし、このようなアプリケーションでは、高齢者のように PC に不慣れたユーザには必ずしも使いやすいものになっていない。

そこで、本報告では、音声入力インタフェースを検討する。音声認識ソフトウェアを用い、GIS での主要な機能である経路探索するための条件、身体状況、出発地、目的地などの設定に音声入力を用いる。また、本歩行者支援 GIS はその用途上、屋外での利用も想定されるため、環境騒音にも頑健に音声認識できる必要がある。そこで、雑音耐性の高い音声認識方式として、SN 比別に作成した

[†] ATR 音声言語コミュニケーション研究所
ATR Spoken Language Communication Research Laboratories

^{††} 独立行政法人 情報通信研究機構
National Institute of Information and Communications Technology

音響モデルをパラレルに用いて認識する方法を提案する。また、このアプリケーションではより高齢層への対応が必要となる。高齢者の発話は若年層の発話と音響特性が異なる。通常、音声認識用の音響モデルはせいぜい60歳代までの成人を対象に作られる場合が多いが、このような若年層用音響モデルでは、高齢者音声において認識精度が低下する。そこで、文献⁷⁾と同様、60歳代から80歳代の高齢者音声データベースを用いることで高齢者の発声に対応した音響モデルを作成する。さらにこの高齢者専用音響モデルを使い、若年層用音響モデルと同様の耐雑音性音声認識システムを構築する。

また、音声認識を用いた応用としては、ディクテーションソフトウェアや、音声 Web ブラウザーなどのように、パーソナルコンピュータ上で従来、キーボードやマウスで行ってきた入力を音声入力で置き換える、あるいは複数の入力デバイスを協調的または選択的に使うものがある。このような音声入力を用いたインタフェースの研究およびその評価には、作図ソフトウェアに19種類のコマンドを音声入力できる機能を付与したものが⁵⁾。この文献では操作時間を比較し、音声入力を用いる事でマウスだけによる作業よりも平均約21%短縮できたとしている。その他には、音声、マウス、キーボードを用いた入力インタフェースを備えたコンピュータ上の電話番号案内システムで評価を行ったものもある⁶⁾。また、高齢者での評価としては、情報検索システムにおいて、音声入力とキーボードを比較した研究がある⁸⁾。ただし、入力時間に差は得られなかったと報告している。対象が米国人であるので、タイプライターに慣れていない可能性もある。本研究では日本人高齢者を対象に音声入力インタフェースの検証を行う。その多くはキーボードやマウスに不慣れな場合が多い。子供等に教えてもらい、若干使っては見たものの、それ以降は自分自身では使った事がない人や、全く触った事すらない人が多い。また、現状の音声入力では、何を言っても機械が聞き分けてくれるわけではなく、決まった音声コマンドや、音声ボタンを押すなどの音声入力のための決まりを習得する必要がある。本報告では、実際の高齢者を被験者として、音声入力をどの程度習得できるのか、他の入力手法に比べ、どの程度有効なのかを検証する。

さらに、将来的に、より自然な音声インタフェースとして、対話システムを構築することが考えられる。Oxford Brookes 大学の音声プロジェクトでは、音声によるインタラクションを高齢者に対し、電話応答システムをもとに研究を行っている⁹⁾¹⁰⁾。高齢者向けに特に必要な要件を提案している。本報告では、まず、Wizard of Oz 法による GIS タスク対話実験において、高齢者がシステムの様々な対話スタイルに対し、どのような振舞いをするかについて調査した。

以下、2節で歩行者支援 GIS の概要を、3節で導入した音声入力インタフェースについて説明し、耐雑音音声認識システムとその音声認識評価実験を述べる。さらに、4節において、高齢者を対象とした音声入力インタフェースの評価実験について述べる。次に、5節において、Wizard of Oz 法による対話実験とそこで得られた知見について

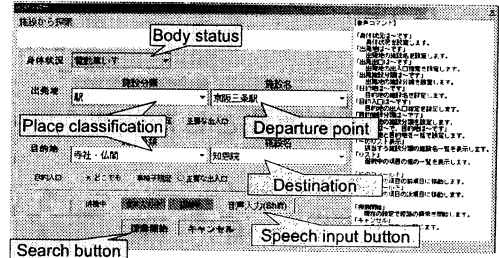


図1 「施設から探索」ダイアログボックス
Fig.1 Dialog box for a search by place names.

述べる。最後に、6節で全体のまとめを行う。

2. 歩行者支援地理情報システムの概要

歩行者支援 GIS は、電子地図を管理する電子地図サーバ、歩行空間のアクセシビリティに関するデータを蓄積したアクセシビリティに関するデータを蓄積したアクセシビリティデータベースサーバ、インターネット公開サーバから構成される。歩行空間のアクセシビリティとは、「歩行者がそこを通行する場合の肉体的・精神的負担の少なさ、および快適さ」を一般的に指す言葉である。またバリアはアクセシビリティを阻害する事物を、バリアフリーはアクセシビリティを高める事物を一般に意味する。ユーザは PC または携帯電話や PDA などのモバイル端末を用いて GIS にアクセスし、最適経路や地域のバリア・バリアフリー情報の検索、ナビゲーションなどのサービスを利用する。

具体的に実現されている例として、東京都小金井市、あるいは京都市東山区の一部に対し、観光スポットのバリアおよびバリアフリー情報を含んだ「小金井バリア・バリアフリーマップ」「京都東山バリア・バリアフリーマップ」があり、インターネットで現在公開されている¹¹⁾。出発地および目的地だけでなく、種々の身体状況(車椅子使用、杖使用、人ごみさけるなど)を設定することで、その状況に応じた最適な歩行経路を地図上に示すことができる。なお、2005年9月現在、公開されているバージョンは新しいもので、本報告でベースにしているものは、一つ前のバージョンである。

3. 音声入力インタフェースの導入

本 GIS をより自然に扱うことができるように、音声入力インタフェースを導入する。具体的には、施設名を指定して探索するダイアログボックスと経路を再探索するダイアログボックスに含まれる項目を音声認識により入力、または制御できるシステムを構築する。

3.1 音声入力インタフェース設計

図1に経路探索条件を設定する音声化された「施設から探索」ダイアログボックスを示す。他には、最適経路を表示した後、部分的に条件を変えて再度探索できる再探索ダイアログボックスもある。ダイアログボックス内の各フィールドは、従来システムと同じものでマウス、キーボードにより入力できる。これらを内部的に音声コマンドに応じて入力できるように変更している。また、音声

入力ボタン、音声コマンドの説明、上部の空白部分に音声入力に対するエラー表示窓、の3つが外見上、音声入力対応用に追加されている。

音声コマンドは、基本的に「コマンド」+「キーワード」という定型句で、「身体状況は～」、「出発地は～」といったパターンである。さらに、間投詞を含んだ発話を認識可能にするために、各コマンドの前に「あー」「えーと」、後ろに「です」「お願いします」などを付与したのもも受理できるようにする。さらに、「出発地は○○、目的地は△△」といった文型で、一度に複数の情報を入力できる。これにより、音声入力に慣れてくれば、より速い入力が可能になる。

3.2 雑音に頑健な高齢者用音声認識システム

歩行者支援 GIS は障害者や高齢者を含む全ての歩行者が利用する事を前提とし、さらに、室内ばかりでなく、野外などの周囲が騒がしい場所でも利用されることが予想されるため、雑音に対する頑健性が求められる。そこで、高齢者用専用の音響モデルを作成し、さらに、雑音に対する頑健性を高めるため、複数の雑音を用いたマルチコンディション学習及び、SN 比別パラレルデコーディング法を使用する。

文献⁷⁾では、大規模な高齢者音声データベースを用いて音響モデルを作成し、新聞読上げタスクにおいて若年層データベースから作成した音響モデルより3~5%の単語認識率の改善を得ている。本報告では、同様な高齢者音声のデータベースを用いて作成した音響モデルをベースに用いる。

雑音の種類に対する頑健性を高めるため、マルチコンディション学習により作成された音響モデルを用いる。更に、SN 比に対する頑健性を高めるため、SN 比に依存した音響モデルを複数用い、同時に認識処理を行う。この手法を用いることにより、発話された音声の SN 比に依存しない音声認識が実現できる。実際の認識システムでは、性別依存モデルをベースに用い、それぞれ雑音の SN 比に依存した音響モデルを用意する。最終的に複数の仮説から最大音響ゆう度を持つ仮説を最終的な認識結果として用いる。

3.3 音声認識評価実験

3.3.1 実験条件

雑音環境における高齢者音声の認識性能を評価するため、性別及び、雑音の SN 比、雑音の種類を未知とした不特定話者音声認識実験を行った。認識性能の評価基準として、間投詞の認識精度を考慮しない正解率を用いた。ここではコマンド正解率と呼ぶことにする。評価用データとして、若年層として20歳代から40歳代の男女各5名、高齢層として60歳代から70歳代の男女各10名、それぞれ21文発声したものをを用いた。

若年層用の音響モデル学習データとして、20歳代から50歳代の話者による ATR 旅行対話データベースを使用した。407名が発声した対話及び、音素バランス503文(計26,380発声)である。高齢者用の音声データベースとして、ATR で収集した「日本語高齢者読み上げ音声」、および財団法人イメージ情報科学研究所らにより作成された「高齢者音声データベース」(S-JNAS)を用いた。「日

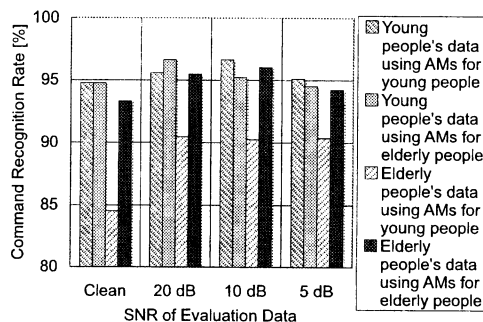


図2 性別依存雑音重畳音声用音響モデルのパラレルデコーディングによる若年層および高齢者に対するコマンド正解率

Fig. 2 Command recognition rates both for young and elderly people by using SNR-dependent acoustic models and parallel decoding.

本語高齢者読み上げ音声」は ATR 音素バランス文と旅行会話基本本文を含み、そこから男女それぞれ27名、26名を用いた。S-JNASからは新聞読上げ音声をも男性178名、女性176名分用いた。合計で41,101発声であった。

音響モデルは最尤逐次状態分割法 (Maximum Likelihood Successive State Splitting: ML-SSS)¹³⁾により生成された1400状態10混合のHMMで、無音モデルは3状態20混合のHMMを用いた。雑音重畳音声データは、4種類(空港ロビー、車道脇、街頭、電車内)の雑音を用いて、SN比を5dB, 10dB, 20dBの条件でクリーン音声に重畳した。音声波形のサンプリング周波数は16kHzである。フレーム長20ms, フレーム周期10msで、ハミング窓を掛けて音響分析を行った。音響特徴パラメータは、12次元MFCC, Δパワー, 12次元ΔMFCCの計25次元を使用した。認識用の文法は状態遷移ネットワークで表現され、キーワードとしては約100単語であるが、間投詞や重複も入れて文法には約300単語含まれている。

3.3.2 認識性能評価

図2に性別依存雑音重畳音声用音響モデルのパラレルデコーディングによる若年層および高齢者に対する認識結果を示す。若年層に対しては、どちらの年齢層のモデルでも同様な精度であった。高齢者に対しては、若年層モデルでは精度が低いが、高齢者モデルでは高い精度が得られた。また、パラレルデコーディングにより、どのSN比でも高い精度が得られた。

4. インタフェース評価実験

現状の音声入力インタフェースでは、使い方や音声コマンド自体を覚える必要がある。特に、本GISの主な対象ユーザは高齢者であり、使用方法を習得可能かどうかは問題となる。そこで、実際に高齢者に使ってもらい、既存インタフェースとの比較を行う。

4.1 マウス入力と音声入力の比較

マウス入力では、まず施設分類を選択し、その分類内に含まれる施設名を選択する。施設名から施設分類を推定する段階で、どの分類に属するか判断に迷う場合が結構多い。また施設名を選ぶ際、数が多いと目的のものを

選ぶ時、一般に時間がかかる。特に、高齢者においては、視力の低下により、目で確認するのも困難が伴うと考えられる。

これに対し、音声入力の場合は、施設名を発声することで直接的に入力できる。どの施設分類に分類されているかを気にする必要がない。ただし、音声入力方法と音声コマンドは覚える必要がある。「音声入力」ボタンをマウスで押し続けている間か、あるいは、shift キーを押し続けている間、音声の取り込みが行われる。また、音声コマンドは忘れてもすぐ思い出せるよう、ダイアログボックス自体に簡単に明記してある。

4.2 評価方法

評価実験として、一人当たり下記の3パターンに対し、10個の課題をこなしてもらった。

- (1) マウス入力のみ使用
- (2) 音声入力のみ使用
- (3) 両方をできるだけ使用

高齢者10名を2グループに分け、[Aグループ] (1) マウス、(2) 音声、[Bグループ] (2) 音声、(1) マウス、の順に取り組んでもらい、どちらのグループも最後に(3) 両方をこなしてもらった。また個々の課題は16の課題をランダムに並べたものを3パターン作成して用いた。当初、被験者には1パターン16課題行う予定であったが、予定した拘束時間では10課題が限度であった被験者が数名存在し、集計は最初の10課題とした。各課題は、それぞれ身体条件、出発地施設、目的地施設の3つを明記した紙を用意し、各課題の開始時に被験者に提示した。また、施設によっては分類が難しい場合もあり、分類のヒントを与えたものもある。これらをしてできるだけ速く効率的に入力するよう指示した。教示の仕方は、はじめにスタッフが入力例を示した後、1つ目の課題は、1条件の入力ごと入力方法を指示しながら、被験者自身に入力してもらった。2回目から4回目までは、被験者が困っているときのみ、簡単なアドバイスを与えて補助した。この回数は被験者に依存する。5回目以降は被験者自身のみで課題をこなしてもらった。1つの課題をこなす時間を

- (a) 条件設定開始から探索開始まで
- (b) 再探索のための条件設定開始から探索開始までの合計とした。

音声認識には音声コマンドを覚える必要がある。このタスクでは、音声認識の文法上、2つ以上のスロットを同時入力することも可能である。しかし、今回の被験者では音声認識自体だけでなく、PCも初体験という人も多く、下記のような簡単に使える1対1の音声コマンドのみを教えて使ってもらった。

[この評価で教示した主な音声コマンド]

- 「身体状況は～です」
- 「出発地は～です」
- 「目的地は～です」
- 「探索開始」

4.3 音声認識における実験条件

音声認識における詳細な実験条件を述べる。音響モデルには各話者の話者適応モデルと若年層の性別依存音響

表1 オフラインでのコマンド正解率
Table 1 Offline command correct rates.

	Command correct rate [%]
Baseline model	95.97
Adapted model	97.58

モデルを併用した。これにより、高齢者用不特定話者モデル以上の性能を得ることができる。この実験を行った時点では高齢者用の不特定話者音響モデルを使用することができなかったこともあるが、実験をなるべくうまく運ぶために最も条件の良いモデルを用いた。適応データとして各話者が発声している音素バランス503文100発声を用い、適応手法としてMAP-VFS法¹⁴⁾を用いた。これらは防音室においてヘッドセットマイクロホン(Sennheiser HMD-410)により収録した。また、インタフェース評価実験では、静かな部屋で同種のヘッドセットマイクロホンを用いて行った。

4.4 オフラインでの音声認識性能

まず、インタフェース評価の前に、今回の被験者に対して、歩行者支援GISに用いる音声認識の評価を行った。評価データとして、歩行者支援GISタスクの各話者21文の発声を用いた。適応データと同一条件で防音室にて収録した音声であり、人手により1発話ずつ切り出されている。表1に被験者10名の平均コマンド正解率を示す。話者適応モデルを用いることで、精度を上げることができる。

4.5 インタフェース評価実験結果

表2に試行回数2~10回までに対して、課題達成までにかかった平均所要時間を、図3に各課題に対する課題達成までにかかった時間をそれぞれの入力方法別に示す。それぞれのデータは10名分の平均値である。またマークなしの細線で示したのは2~10回までの平均値である(表2と同じ値)。1回目はスタッフの教示する時間が大きく入るため、この平均値の計算では含めていない。

マウス入力のみの方は音声入力の方に比べ、約2.4倍時間がかかっており、明らかに音声入力のみの方が効率が良い。約60%所要時間を縮めた事になる。音声入力の場合は最低で「身体状況」「出発地」「目的地」「探索開始」の4つを入力すればよい。これに対して、マウスでは一度、項目を選んでリスト表示してから選択が必要がある。つまり一つ項目を入力するには2回クリックする必要がある。さらに「施設分類」も選択する必要がある。そのため、普通には11回クリックが必要である。

また、音声入力にはshiftキーを使うのでマウス移動の必要がなく、時間的に有利である。一方、音声入力では、誤認識や、音声入力ボタンから手を離すのが早すぎて途中で音声途切れる、あるいは音声認識結果がすぐに反映されないバグ、など負の要因が存在したが、認識率はある程度高いので、数回言い直すことで正しく入力できた。

両方使う場合は、人によってどの項目を入力するときを使うか異なるが、マウスで操作する場合、移動させる時間が必要なため、音声入力のみより若干時間が増えている。ボタン1つクリックする「探索開始」や、選択項目数の少ない「身体状況」(7種類)をマウス入力で行う

表 2 高齢者 10 人による 2 回から 10 回までの試行における各課題を達成するまでの平均所要時間

Table 2 Average time to complete each scene from second trial to tenth trial for ten elderly subjects.

	平均所要時間 [s]		
	All	PC users (S1)	PC non-users (S2)
Mouse	93.2	75.2	120.3
Speech	38.6	37.3	40.5
Both	41.6	36.2	49.8

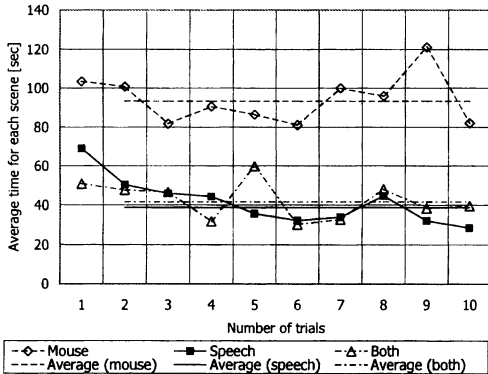


図 3 試行回数に対する課題達成までにかかった時間
Fig. 3 Average time to complete each trial.

人が多かった。

4.6 PC 経験有無による所要時間比較

被験者 10 人のうち 6 人がパーソナルコンピュータ経験者で、そのうち日常的に使用しているのは 3 名であった。これら以外の 4 人は自分の子供等に教えられて触ったことはあるが、それ以降全く触っていない人、本当に全く触ったことがない人であった。

そこで、PC 経験のある被験者と経験のない被験者に分けて所要時間を集計した。試行回数 2~10 回までの平均所要時間を表 2 右側の列 2 つに、試行回数に対する所要時間を図 4 に示す。経験者と未経験者では、マウス操作に対する慣れが全く異なり、所要時間に大きな差があった。マウス入力のみの場合、未経験者は経験者と比べて約 1.6 倍かかった。

また、音声入力においては、PC 経験者と未経験者間で差が小さい。平均での比較では、経験者はマウス入力の場合の約 1/2 倍であったのに対し、未経験者では約 1/3 倍になっている。音声入力は今回の被験者全員にとって、初体験であり、shift キーを押して入力するため、PC に対する若干の慣れは必要である。しかし、これ自体は簡単な操作なため、誰にとっても簡単に習得できたと考えられる。

さらに、両方使う場合には差が開いた。表 2 から、音声入力のみの場合より、経験者は所要時間を短くできているが、未経験者はむしろ時間がかかった。これは、未経験者はマウスの使いみちだけでなく、2 種類の入力方法があると、どう使い分けるかとまどいがかちであったのに対し、たいいてい経験者は効率よく使うことができていたことによると考えられる。

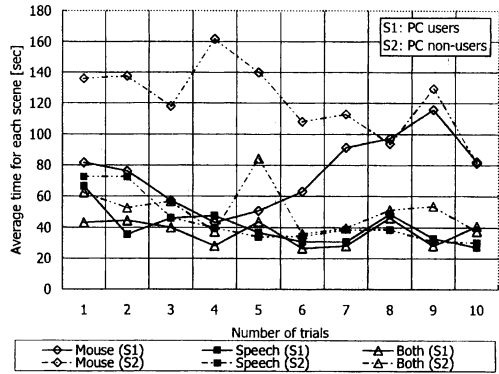


図 4 試行回数に対する課題達成までにかかった時間 (PC 経験別)
Fig. 4 Average time to complete each trial by each group divided by PC experience.

5. 高齢者対話実験

将来的には、よりユーザフレンドリーな音声インタフェースとして対話システムが考えられる。今回は、まず、高齢者が対話システムにおいてどのような振舞いをするか調査した。

5.1 対話実験パターン

本 GIS の条件入力を対象とした Wizard of Oz 法による対話実験を行った。対話パターンとしてユーザ主導/システム主導、あるいはアプリ表示あり/なしで調べた。高齢者 20 名の各被験者に対し、(1) アプリ表示なしで、a. ユーザ主導、または b. システム主導、(2) アプリ表示ありで、a. ユーザ主導、または b. システム主導、(3) アプリ表示なしで a. ユーザ主導、の順に 3 シーンずつ対話を行った。1 シーンごとに身体状況、出発地、目的地を含んだユーザの状況を示す文章を提示し、それを一通り理解してもらってから各対話実験を行った。

5.2 対話実験結果

図 5 に各シーンに対する平均対話時間を示す。各対話パターンの平均から、システム主導対話のとき対話時間が減少していることがわかる。さらに最後のパターン (3) アプリ表示なし、a. ユーザ主導、では、初めのパターンと同じではあるが、対話時間が減少していることがわかる。学習効果によるものと考えられる。

より詳細に分析するため、被験者を PC 使用頻度に応じてグループ分けし、集計した。PC を日常的に使う “PC user” 4 人、年に数回使う程度の “PC beginner” 9 人、使用経験のない “No PC experience” 7 人に分かれた。図 6 に PC 経験度合別の平均対話時間を示す。“PC user” ではアプリ表示ありのとき対話時間が大きく減少した。GUI に慣れており、対話の目的に気がつきやすいと考えられる。“PC beginner” ではアプリ表示あり/なしでは差はないが、ユーザ主導/システム主導では差が大きく、システムに合わせた対応をしたことがわかる。“No PC experience” では、全体的な対話時間の減少はあるが、対話パターンに依存した変動は少ない。システム主導で一項目ずつ聞く場合でもすべての情報を話す人が多かったことが原因であった。実験後のアンケートによると、こ

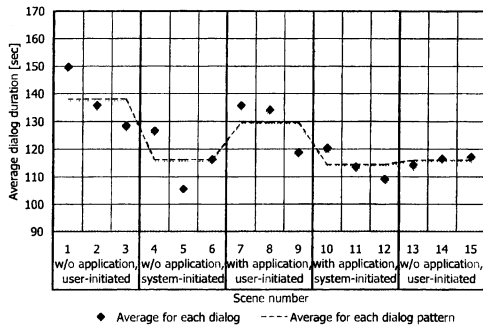


図5 各シーンに対する平均対話時間
Fig. 5 Average dialog durations for each scene.

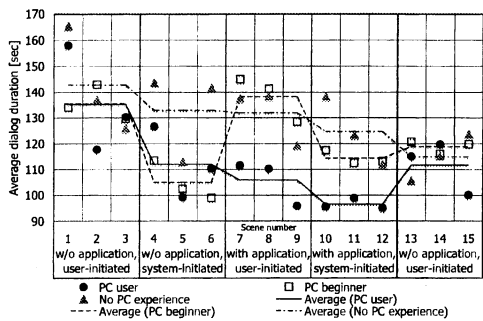


図6 PC経験度合による各シーンに対する平均対話時間
Fig. 6 PC experience-dependent average dialog durations for each scene.

のグループは他と異なり、ユーザ主導の方を好むことがわかった。システム主導は制約が強すぎ、それに合わせるのは煩わしく、自由に話せるユーザ主導の方が良いと答える人が多かった。

6. まとめ

すべての歩行者に対し、バリアおよびバリアフリー情報を提供する歩行者支援地理情報システム (GIS) において、音声入力インタフェースを導入した。既存のマウスやキーボード入力インタフェースを残しつつ、音声入力を効果的に利用できるようなインタフェース、音声コマンドを実現した。本歩行者支援 GIS は野外での使用が多く考えられることから、複数の雑音を予め用いたマルチコンディション学習と SN 比別音響モデルを用いたパラレルデコーディングを導入し、雑音耐性のある音声認識システムを構築した。また、主なユーザとして高齢者が考えられるため、高齢者用の音響モデルを作成して用いる事で、若年層と同様に、様々な SN 比における高齢者の発声に対しても十分な認識精度を得る事ができた。

また、音声入力インタフェースとしての評価を主な対象である高齢者を被験者として行った。音声入力により、マウス入力に対して所要時間を 60%縮める事ができ、効果がある事が分かった。被験者のうち半分は PC に不慣れであり、2つのグループ間で比較するとマウス入力に対する時間が大きく異なる事が分かった。しかし、音声

入力に関しては同様な所要時間で入力可能であり、どちらのグループにおいても効率的な入力方法であることが実証された。

対話システム構築のための実験として、Wizard of Oz 法により、高齢者と機械の対話実験を行った。数種類のパターンを行った結果、対話時間が PC 使用経験に依存することがわかった。

今後の課題として、高齢者、さらには発話に困難さを持つ障害者に対する音声認識の精度向上及び頑健性向上を図るとともに、音声コマンドに縛られないような、より柔軟性のある音声インタフェースの検討があげられる。

謝辞 本研究は独立行政法人 情報通信研究機構の研究委託「歩行者支援 GIS のための音声入力インタフェースの改良と評価」及び「大規模コーパス音声対話翻訳技術の研究開発」により実施したものである。

参考文献

- 1) 矢入 (江口) 郁子, 猪木 誠二: 高齢者・障害者の自立的移動を支援する Robotic Communication Terminals(3), 人工知能学会論文誌, Vol.18, No.1, pp.29-35 (2003).
- 2) 矢入 (江口) 郁子, 吉岡 裕, 小松 正典, 猪木 誠二: 歩行者支援 GIS のための歩行空間アクセシビリティ情報の蓄積と評価, ヒューマンインタフェース学会論文誌, Vol.5, No.4, pp.17-23 (2003).
- 3) 矢入 (江口) 郁子, 猪木 誠二: 高齢者・障害者を含む全てのユーザを対象とした歩行者支援 GIS プロジェクト, 信学技報, WIT2003-39, pp.17-22 (2004).
- 4) 實廣 貴敏, 松田 繁樹, 葦苅 豊, 中村 哲, 矢入 (江口) 郁子, 猪木 誠二: 歩行者支援 GIS における音声入力インタフェースの検討; 信学技報, NLC2004-1, WIT2004-28, pp.1-6 (2004).
- 5) Pausch, R., Leatherby, J. H.: An empirical study: adding voice input to a graphical editor, *Journal of the American Voice Input/Output Society*, 9:2, pp.55-66 (1991).
- 6) Yoshioka, O., Minami, Y., Shikano, K.: A multi-modal dialogue system for telephone directory assistance, *Proc. ICSLP '94*, Vol. 2, pp.887-890 (1994).
- 7) 馬場 朗, 芳澤 伸一, 山田 実一, 李 晃伸, 鹿野 清宏: 高齢者音響モデルによる大語彙連続音声認識, 信学論 (D-II), Vol. J85-D-II, No.3, pp.390-397 (2002).
- 8) Anderson, S., Liberman, N., Bernstein, E., Foster, S., Cate, E., Levin, B.: Recognition of elderly speech and voice-driven document retrieval, *Proc. of ICASSP*, Vol.1, pp.145-148 (1999).
- 9) <http://cms.brookes.ac.uk/computing/speech/>
- 10) Zajicek, M., Wales, R., Lee, A.: Speech interaction for older adults, Special Issue 'Design Principles to Support Older Adults' of Universal Access in the Information Society, (ed. M. Zajicek & S. Brewster) Springer, Vol. 3/2, pp. 122-130 (2004).
- 11) 歩行者支援 GIS <http://bfms.nict.go.jp/>
- 12) 松田 繁樹, 實廣 貴敏, コンスタンティン マルコフ, 中村 哲: 雑音や発話スタイルの変動に頑健な日本語大語彙連続音声認識; 情処学研報, 2004-SLP-50-6. (2004).
- 13) Ostendorf, M., Singer, H.: HMM Topology Design Using Maximum Likelihood Successive State Splitting, *Computer Speech and Language*, Vol.11, No.1, pp.17-42 (1997).
- 14) Tonomura, M., Kosaka, T., Matsunaga, S.: Speaker adaptation based on transfer-vector-field-smoothing using maximum a posteriori probability estimation, *Proc. of ICASSP '95*, Vol.1, pp.688-691 (1995).