

対面朗読者と視覚障害者の対話の分析

西本 卓也[†] 小川佳奈子^{††} 渡辺 隆行^{††}

[†] 東京大学大学院情報理工学系研究科

^{††} 東京女子大学現代文化学部

E-mail: nishi@hil.t.u-tokyo.ac.jp, nabe@lab.twcu.ac.jp

あらまし 視覚障害者の情報機器利用を支援する音声対話システムを設計するために、対面朗読者と視覚障害者による対話の収録と分析を行った。ケータリングのメニューから昼食を選ぶタスクについて検討した結果、階層化されたメニューから選択を行うタスクはウェブ上のショッピングカートと同様の構造でモデル化できること、音訳に固有の技術が有用であること、対面朗読に固有な要素として大局的理解や双方向性などが有効に利用されていること、などが示唆された。

キーワード 視覚障害, 音声対話, 対話分析, ウェブ

Analysis of human-to-human dialogs to assist visually-impaired people

Takuya NISHIMOTO[†], Kanako OGAWA^{††}, and Takayuki WATANABE^{††}

[†] The University of Tokyo

^{††} Tokyo Woman's Christian University

E-mail: nishi@hil.t.u-tokyo.ac.jp, nabe@lab.twcu.ac.jp

Abstract An experiment of performing human-to-human dialogs to assist visually-impaired people was conducted for acquiring knowledge required for developing dialog systems. A visually-impaired person was asked to choose the meal from the menu of a catering service. The assistant of the visually-impaired person had the experience of transliteration from the printed books to the audio recordings. As the result of our analysis, the conversation held between two persons can be modeled with the same structure as the shopping cart systems on the popular web sites. The technique of transliteration was effectively used in the dialogs. The overall views and the interactivity were also important factors for reduction of both mental workload and required time.

Key words Visually impaired people, human-to-human dialogs, dialog analysis, Web

1. はじめに

視覚障害者の日常生活の中で、情報の収集に関する制約は、最も大きな不自由のひとつである。現在コンピュータを利用している視覚障害者は、スクリーンリーダ（画面を音声や点字でユーザに伝えるソフトウェア）によって文書を作成したりウェブやメールを利用したりしている。また、コンピュータや組み込み機器における音声認識技術の利用事例も増えている。しかし、現在我々が利用可能な技術の多くは、人間同士で対話をするようにコンピュータを扱いたいという要求とは程遠い。音声認識や音声合成など諸要素技術の応用としてもさまざまな音声対話システムが研究開発されている。しかし、視覚障害者のニーズを的確に把握して音声対話システムが開発された事例は

少ない。

本研究の目的は、視覚障害者を支援する際に行われる人間同士の音声対話を分析し、音声対話の利点や、対話において人間が発揮している知的能力などを明らかにし、可能な限りそれらを音声対話システムの設計に反映することである。視覚障害者の日常生活における具体的なニーズに着目し、たとえ特定のタスクに話題を限定したとしても、そのタスクをあたかも人間同士のように自然な会話を通じて実行できる、そういったシステムの実現を目指している。

我々は、視覚障害者を支援する対話を上手に行えるのは対面朗読者である、と考えている。本報告の実験に協力した対面朗読者（女性1名）は、長期間にわたって音訳ボランティアの講習を受けており、録音図書等の作成に必要な知識や経験を持つ

ている。さらに、図書館等において、視覚障害者と対面で、依頼された書籍の朗読を行う、といったボランティアの経験を積んでいる。このような対面朗読者と視覚障害者の対話を分析することで、視覚障害者のニーズに近いシステムが実現できると期待される。

予備的な検討として、目を閉じた晴眼者が対面朗読者に書籍などの朗読を依頼する、という実験を行った。その結果、対面朗読者は 1) 目的に向かって誘導を行う、2) 双方向性を生かす、3) 音訳の技術を生かす、4) 聞き手の記憶負荷を考慮する、などの配慮を行っている、という知見を得た。

本報告では、この予備的な検討の結果を踏まえて、対面朗読者と視覚障害者が行う音声対話に関するいくつかの仮説を提示する。次に、メニュー選択タスクに関して行った対話実験について述べる。さらに、実験で行われた対話を分析し、提示した仮説の妥当性について論じる。最後に、まとめと今後の課題について述べる。

2. 音声対話による視覚障害者支援

2.1 ショッピングカートモデルの有効性

情報機器やウェブコンテンツにおいては、多数の要素が階層構造に従って配置されており、メニューを何回もたどりながら最終的に何かを選択することが多い。また、日常生活においても、レストランのメニューや家電製品の操作マニュアル、辞書などを参照する場合には、同様の作業がよく行われる。

このようなメニュー選択タスクにおいて、メニュー階層の全体像が把握できない場合は、自分が求めるアイテムがメニューのどの位置にあるかを探ることが難しい。また、自分が求めているアイテムが自明ではない場合は、複数の候補に絞り込んで、それぞれの候補の詳細情報を比較しながら、最終的に適切なアイテムをひとつ選ぶ、といった作業が必要になる。

ここで我々は、次の仮説を提示する。

仮説 1: オンラインショッピングのウェブサイトで用いられるショッピングカートのシステムと、メニュー選択タスクの音声対話は、共通する構造をもつ。

ここで我々が「ショッピングカート」と呼ぶシステムは、以下の要素から構成されるウェブサイトである。

- リスト（アイテムまたはカテゴリ）
- 詳細説明（アイテムまたはカテゴリ）
- アイテムをカートに入れる操作、カートから出す操作
- カート内容の確認、比較、決定などの操作

このような仕組みを反映した音声対話は具体的には以下のような流れで行われる。

- (1) 対象となるアイテム群の全体把握
- (2) リストの読み上げ（気になったアイテムをカートに入れる。複数の候補アイテムがカートに入る）
- (3) カート内のアイテムの詳細説明を得る（詳細情報を聞いてアイテムの取捨選択を行う）
- (4) カート内に残ったアイテムを確定する

なお、実際のオンラインショッピングのウェブサイトにおいては、メニュー階層をひとつずつ辿らずに、キーワードで検索

を行い検索結果を得る場合もあるが、検索結果が複数の候補となる場合は、検索後の操作はやはりリストからの選択となる。つまり、このような検索機能は、メニュー階層の移動における一種のショートカット操作と見なすことができる。

2.2 音訳技術の有効性

音訳とは、視覚障害者のために墨字の書籍を朗読し、カセットテープや DAISY 規格の CD-ROM などの音声録音媒体を作成する作業である。主にボランティアによって行われており、マニュアル [2] [3] [4] も出版されており、以下のような技術が必要とされる。

- 音声表現技術：文意に沿って、情報を正しく伝える読み方
- 調査技術：漢字など、読めない文字の調べ方
- 配慮技術：視覚に頼った情報（同音異義語、写真、その他）の読み方
- 録音技術：雑音のないきれいな録音の仕方
- 校正技術：原本のとおり音声化されているかチェックの仕方

音訳とは、文学作品の朗読などとは異なり、書かれていることをできるだけ忠実に音声化することが求められる作業である。そのために、記号、同音異義語、ルビ、注など、印刷物に含まれるさまざまな情報を言語化する方法や、内容を正しく伝えるための抑揚、ポーズ、アクセントなど非言語情報における考慮など、さまざまな規則がある。

例えば、聞いていて抵抗がない、自然なピッチでの読み上げの例を以下に示す。

- (H 私たち) の学校は (H 山の上) に (L あります)。
- (H 今日は)(H 運動会) の予定でしたが、(H 朝から)(H 雨が) 降っていたので中止に (L なりました)。

また、意味の曖昧性を回避するための適切なポーズの例を以下に示す。ただし (P) はポーズがあるべき箇所、(P?) はポーズの有無によって文の意味が変わることを示す。

- かわいい (P) 女性のための喫煙具が人気の的です。
- 有名な (P?) 大学の (P?) 先生がいる。
- 太めの (P) 時計のバンドを買った。
- 私の背の高さ位の (P) リンゴがたくさんなっている木。

また、図表や写真などの視覚的な情報を言語化するためには、音訳者の創意工夫が必要となることもある。

ここで我々は、次の仮説を提示する。

仮説 2: 音訳の規則に従った対話を行うことで、視覚障害者に対して、情報が正確に伝わり、理解が容易になる。

2.3 対面朗読技術の有効性

対面朗読とは、視覚障害者に対して対面で本などを読むサービスである。専門書や新聞など録音図書が存在しない書籍の場合は対面朗読が必須である。朗読者にその場で、どこをどのように読んで欲しいか、といった指示を伝えることができるのが利点である。図書館などでは、利用者の要求に応じて、特定の分野の知識を持っている朗読者が選ばれる。また、我々の実験に協力した朗読者は、利用者の自宅を訪問して、家電製品のマニュアルを読みながら操作を支援したり、来客者のためにケータリングのメニューを読んで注文を代行したり、といった作業

の経験を持っている。

音訳のためのマニュアルの多くは録音図書の作成方法について説明しているが、対面朗読についての記述はわずかである。しかし、対面朗読の利用者は、何らかの基準で行為を選択しており、熟練した対面朗読者はそのような利用者の行為を支援することに長けていると考えられる。例えばメニュー選択タスクにおいては、利用者の行為選択基準は以下のようなものだと考えられる。

- 全体を把握して納得できる選択を行いたい
- できるだけ少ない時間や労力で作業を行いたい

これらの行為選択基準を満たすために、以下のような対面朗読の利点が特に有用だと考えられる。

仮説 3: 対面朗読の利点を生かしたメニュー選択対話においては、視覚障害者はよりよい選択を行うことができ、しかもそれが効率的に実現できる。

- 仮説 3a: 大局的理解の有用性: 全体を把握する、比較を行う、視点を変えながら動的に構造化を行う
- 仮説 3b: 双方向性の有用性: 確認を行う、共同作業で何かを発見する

大局的理解と双方向性は協調的に作用し、よりよい選択を効率的に実現することに貢献する、と考えられる。

3. 予備実験

まず、予備実験として、実際の対面朗読の様子を観察した。予備実験では晴眼者 1 人が目を閉じて視覚障害者の役を演じた。対面朗読者は約 10 年の朗読ボランティア経験を持つ S さん（女性）である。課題として以下を行った。

- 課題 1: 書籍を朗読する。
- 課題 2: 電話の子機の使い方を電話のマニュアルを使って説明する。
- 課題 3: ケータリングサービスのメニューの中から食べたいものを選ぶ。

対話の様子をビデオカメラで撮影し、分析を行った結果、対面朗読者は 1) 目的に向かって誘導を行う、2) 双方向性を生かす、3) 音訳の技術を生かす、4) 聞き手の記憶負荷を考慮する、などの配慮を行っていることが明らかになった。

具体的な例を以下に挙げる。

目的とする情報に早く誘導する工夫（ナビゲーション、ブラウジング、読み飛ばし、絞り込み）

対面朗読者: 「おおざっぱには…御膳ものでお弁当ですね。そのお弁当ものも和風、中華…中華はないですね、あるのはカレーの世界ですね。それにカツとかエビフライとかがトッピングで乗ったもの。それと後は、お重、いわゆる鰻重とかヒレカツ重とか、親子重だったりですね。」

利用者: 非言語「お膳ものでお弁当、和風、カレーの世界…」に合わせたうなずき。

全部を読んでもらおうとすると時間がかかる! 「おおざっぱには」という言い方で大まかなメニュー（和、洋、中など）を読み上げてブラウズしている。

対面朗読者: 「N さんが焼き肉系がいいかなとか、あっさり

が良いかなとか、中華彩り…があるのいいかな、によるのではないですか?」

相手の希望を聞いて絞り込もうとしている。

双方向性（音声対話、うなずきなどの頭の動き、聞き返し）
対面朗読者: 「今申し上げたのでイメージできますか? 決められますか?」

「イメージできましたか?」という言い方で、伝わったかどうか確認している。

対面朗読者: 「これあのちなみにね、特製カレーうどんプラス舞茸ご飯なんてなってますね」

利用者が迷っていると対面朗読者から提案することもある。

音声化の工夫（音訳マニュアル）

対面朗読者: 「カレーのチラシと…御膳…混ぜご飯みたいな…2000 円で二重、お重、あの二重三重ってありますでしょ、何重、そのお重になっていて」

利用者: 非言語「二重三重ってありますでしょ?」に対してうなずき。

「二重」の説明。聞き間違いやすい言葉を補う。写真を説明している。

対面朗読者: 「一番安いのは若鶏の唐揚げで 700 円ですね」

メニューの値段をはっきり発音している。

記憶の負荷

利用者: 「最初に言ってもらったカシミアルでしたっけ?」

以前に聞いた情報をはっきりと覚えておらず、記憶が正しいか確認している。

利用者: 「和風のお弁当と、カレーとお重とスパゲッティと…」

選択肢が多く、全部覚えるのが困難。

なお、予備実験の結果、ケータリングサービスのメニュー選択タスクが、被験者のモチベーションが高まること、課題のもつ構造に汎用性が期待できること、などの理由から本実験のタスクとして適切であるという知見を得た。

4. 本実験

4.1 本実験の手順

本実験における対面朗読者は予備実験と同じ S さん、視覚障害者は大学院生の M さん（男性）である。M さんは 3 歳のときに視力を失い、現在の視力は 0 である。

ケータリングサービス（弁当屋）のメニューの中から食べたいものを選ぶという課題を用意した。被験者のモチベーションを高めるために、昼食の時間帯から実験を始め、実際に選んだメニューを実験の後に注文するという設定で行った。

対面朗読者 S さんと視覚障害者 M さんにはテーブルをはさんで向かい合って座ってもらい、すべての課題について対話の様子をビデオに録画して観察した。録画した対話をすべて書き起こして分析に用いた。また、以下のように、発話内容の意図を踏まえて簡潔に書き直した意図情報を作成した。

- 発話: スタンダード 1, スタンダード 2, なにこれ、見えますがどこが違うの? という感じでほとんど違わないんじゃないの? みたいなんですが、ご希望に沿いながら読み上げていきますが

● 意図：スタンダード 1 と 2 の違いはほとんどないようです。

さらに、個々の発話に意図タグを付与した（文献 [5] [6] を参考にした）。意図タグは談話行為、動作、対象、詳細情報の 4 階層で構成されている。本研究では談話行為、動作の第 2 レイヤーまでを利用した。具体的には以下のカテゴリから構成される。

● 談話行為（話者の発話内行為）：依頼、表明、陳述

依頼：ある動作を依頼する、何か相手に頼む。

示唆：ある行為を示唆する、行ってほしいことをそれとなく示す

提案：ある行為を相手に提案する。

陳述：行為を述べる、ある動作を相手に働きかける発話でないもの

● 動作（話者の行為）：提示、検索、再検索、選択

確認：相手にある対象を確認する。

提示：ある情報（対象）を相手に提示する。

検索：対象物を検索する。

選択：いくつかの対象物から、ある対象物を選択する。

● 対象：ジャンル、メニュー、カテゴリ、メニュー情報、検索結果、意志内容、真偽情報、提示内容

● 詳細情報：名前、値段、中身、種類

4.2 本実験の結果

本実験の対話の所要時間は約 40 分であり、朗読者と利用者の発話数の合計は約 700 となった。

行われた対話において、被験者はまずメニューの全体を対面朗読者に読み上げさせ、興味があったものをいくつか記憶しておき、後ほど詳細情報を確認しながら取捨選択を行っていく、という流れが観察できた。

本実験では、視覚障害者 M には、ほとんどうなずきが見られず、予備実験において晴眼者 N が頻繁にうなずいていたのとは対照的であった。

相槌の回数は、約 40 分間の会話で 183 回であった。予備実験における晴眼者と比較して、時間あたりの相槌の頻度はほぼ同じであった。相槌を打たない場合の多くは、朗読者の発話に対してすぐに発話しているときであった。観察の結果から、相槌は話を聞いていることを示し、また、対話が順調に進んでいることを示していたと考えられる。

意図タグの集計結果は以下の通りである。

● 利用者の談話行為は 69% が「依頼」と「示唆」。
● 朗読者の発話は 96% が「陳述」。利用者の質問や要望に答える場面が多い。

詳細な分析結果は次章で述べる。

5. 考 察

5.1 仮説 1) の検証

まず仮説 1) について検証する。本実験で行われた対話は基本的にショッピングカートモデルに従っていた。特に興味深い現象の一部を以下に挙げる。

● 候補リストへの追加をするかどうか、判断を促す対話が

見られた。

例：「どれを捨ててどれを残しますか？」

● カートの中身に相当する候補リストは、利用者および朗読者の記憶にのみ存在した。現在の候補リストの内容に関するお互いの記憶を確認する対話がときどき行われた。

● 利用者が最後まで迷っていたことから、候補に追加する、候補から除外する、といった操作が頻繁に行われた。

● それぞれのアイテム（弁当の名前）だけでなく、「カレー 3 種類」のように、カテゴリを単位とした候補リストへの追加・削除が起きた。

● 比較や差分などの説明が有効。

例：2500 円と 3000 円の「違いはフルーツのみです」

● 朗読者が絞り込み戦略の提案を行った。

例：「今日はこれが食べたい… てなやり方もありますでしょ？」

5.2 仮説 2) の検証

次に仮説 2) について検証する。本実験で行われた対話においては、音訳技術と同一の配慮が多く見られた。特に興味深い現象の一部を以下に挙げる。

● 詳細説明において写真の内容を言語化していた。

例：「お弁当の折り詰めの中の配置は…」

● リストの読み上げにおいて、アイテム総数の読み上げ、アイテムの通し番号の読み上げ、リスト読み上げの終了の宣言、などが行われた。

例：「(スタンダード 1 の種類は) 10 かそのくらい」

「1, スペシャルミックス」

「2, チキンカツ」

…

「はい、これだけです。いま名称を申し上げました」

5.3 仮説 3a) の検証

仮説 3a) は大局的理解の有用性に関する仮説であるが、これを裏付ける現象がいくつか見られた。本節ではいくつかのパターンに分類して具体例を示す。

● 全体把握：おおざっぱな系統や印象、違いなどを説明していた。

例：「全体的な系統は和食系です」

● 全体把握：「何があるか」に加えて「何がないか」を概説した。

例：「全体に和食です」

「麺類はなさそうですよ」

● 例外確認：利用者は、全体把握の情報を得たときに、例外の有無を確認した。

例：「冷やし中華とかはありますか？」

「うなぎは？」

● ヒントの提供：利用者が迷っているときに、朗読者は全体把握に基づくヒントを与えた。

例：「オードブルはあまりリーズナブルでないようです」

● メタ情報：利用者はカタログの視覚的構造やアイテムの命名規則などのメタ情報について質問を行った。

例：「1 つのメニューに 1 つの写真がついている？」

「名前から内容がわからないネーミングですね？」

- 対比による特徴の説明が行われた。

例：天むすは「白いご飯にごまかけじゃなくて」

- 2つのアイテムの差分による説明が行われた。

例：2500円と3000円の違いは「フルーツのみ」

- 省略が利用された。

例：「御膳、省略しますよ。洋風ハンバーグ（御膳）…」

- 視点の変更：ユーザの興味に応じてアイテムの呼び方を変えていた。

例：「AB 若菜，桐」

「（価格で）1500, 2000, 2500, 3000 …」

これらの行為を行うためには、利用者の要求に関する推論や料理に関する常識などが朗読者において必要となる。

5.4 仮説 3b) の検証

仮説 3b) は双方向性に関する仮説であるが、これを裏付ける現象もいくつか見られた。本節ではいくつかのパターンに分類して具体例を示す。

- 要求の確認が行われた。

例：（朗読者）「希望は和風，洋風，中華，天むす，だいたいこんな感じ？」

- 理解が可能であるかどうかの確認が行われた。

例：（朗読者）「かんぴょう分かりますか？」

例：（予備実験より）「いま申し上げたのでイメージできますか？」

- 利用者が受け取った情報を正しく記憶しているかどうかの確認が行われた。

例：（約 10 分前に言及された「五目ごはんデラックス」について）（利用者）「五目なんとなかってのが確か。五目ごはん …」

例：（利用者）「天むすが3つで、あとは …」

（朗読者）「そうですね」

例：（利用者）「天むすってのが何点があったと思うんですけど」

（朗読者）「いろいろなパーツのところがありましたね」

（利用者）「スタンダードの1と2の両方があったと思うんですが」

- 共同作業：リスト読み上げ中に利用者が、アイテムごとに相槌を行った。

- 共同作業：朗読者の誤りを相槌の代わりに指摘していた。

例：「5. 鳥照り焼き」

「はい」

「7. ビーフステーキ」

「あ、6ですね」

- 共同作業：二人で協力して比較・発見が行われた。

例：（スタンダード1と2の違いは価格帯であることを知るために）

「1と2の違いってわかりますか？価格帯ですか？」

「700円800円代が1，2が1200円クラス、みたいですわね」

- 割り込み：対面朗読者の発話を遮って利用者が自分の要望を伝えるという場面が見られた。

例：「1，スペシャルミックス。見た目はね …」

「名前だけ」

「名前だけ言いますね」

これらの双方向的なやりとりには、朗読者が利用者の要求を

理解している必要がある。また、選択作業の時間や労力を軽減するためにどのような配慮を行えばよいか、という朗読者の適切な判断が重要である。さらに、発話中の相槌や割り込みなど実時間性の高いコミュニケーションが、双方向性を生かすために重要であると考えられる。

6. まとめ

視覚障害者の支援技術としての音声対話のあり方を検討するために、対面朗読者と視覚障害者の対話を検討した。対面朗読者が行う対話がなぜ有用なのか、といった点に関する仮説を以下のように整理した。

- 仮説 1) ショッピングカートシステムの有用性：階層化メニューを辿るタスクにおいて、ショッピングカートシステムを音声化したような情報構造やナビゲーションモデルが有用である。

- 仮説 2) 音訳技術の有用性：録音図書を作成する際に使われる音訳技術は対面朗読においても有用である。

- 仮説 3a) 大局的理解の有用性：対面朗読に固有な技術として、リストの全体を見通して大局的理解を行い、それを言語化して伝達すること、あるいは効率的に比較を行うこと、などはメニュー選択タスクにおいて有用である。

- 仮説 3b) 双方向性の有用性：対面朗読に固有な技術として、双方向で実時間のコミュニケーションが特に有用である。特に、音声対話を効果的なものにする上で、確認を頻繁に行うことや、情報の一方的な提供ではない共同作業による比較や発見などが重要である。

さらに、実際に行われた対話における現象を、これらの仮説を裏付けているかという観点から分析した。

今後の予定は以下の通りである。

- 朗読者と利用者の個人性について検討するために、他の被験者による実験を行う。

- 本実験において録音された音声を詳細に分析し、特に非言語情報やポーズなどに着目し、音訳技術の有効性の検討を引き続き行う。

- 本実験において対話中に採取した頭部モーションセンサーの情報を分析し、朗読者および利用者の頭部運動を分析することで、音声入力以外の手段で人間の意図や態度を理解することが可能であるかを検討する。

- それぞれの仮説をなるべく個別に検証できる対話実験を検討する。例えば、ショッピングカートシステムの有用性に注目し、他のシステムとの比較を行う対話実験を行う。また、音訳技術のうち、リスト読み上げに関する配慮の有無を比較する実験、大局的理解のうち概略説明や比較に関する一部の行為を統制した比較実験を行う。

- 個々の仮説のうちシステム化しやすい部分を、音声ブラウザで利用可能なウェブアプリケーションとして実装する。また、音声認識を用いた対話システムとして実装する。これらのシステムを視覚障害者や晴眼者に利用させ、評価する。

謝 辞

本研究は、科研費特定研究「情報福祉の基礎」Kiki 班（視覚障害者の聴覚認知の解明と音声対話への利用）の成果である。安村通晃・慶應義塾大学教授，小田浩一・東京女子大学教授，Kiki 班メンバー各位，実験に協力していただいた被験者の方々に感謝する。

文 献

- [1] 渡辺隆行, 安村通晃, 小田浩一, 西本卓也: “視覚障害者の聴覚認知の解明と音声対話への利用に向けて,” 電子情報通信学会技術報告, WIT2004-74, pp.7-12, Mar 2005.
- [2] 全国視覚障害者情報提供施設協会: 初めての音訳, 1999.
- [3] 全国視覚障害者情報提供施設協会: 音訳マニュアル 音訳・調査編, 2001.
- [4] 全国視覚障害者情報提供施設協会: 音訳マニュアル 処理事例集, 2004.
- [5] 入江 友紀, 松原 茂樹, 河口 信夫, 山口 由紀子, 稲垣 康善: “意図タグつきコーパスを用いた発話意図推定手法,” 人工知能学会 SIG-SLUD, Jul 2003.
- [6] 入江 友紀: 意図タグ付けマニュアル Version2.0, 2004.