

Ubiquitous Talker: 拡張現実感における自然言語処理

長尾 確 暦本 純一

(株) ソニーコンピュータサイエンス研究所

1 はじめに

バーチャルリアリティ(仮想現実感)のようにコンピュータの生成する情報世界(仮想世界)に人間が入り込むのではなく、実世界と情報世界を融合し、実世界を情動的に拡張しようという考えは、拡張現実感あるいはオーグメンテッドリアリティ(augmented reality)と呼ばれる。

カーナビゲーションシステムは、GPS(Global Positioning System)などを使って、実世界における現在位置と地図(情報世界)上の現在位置の間に常に関連を持たせ、情報世界の内容の内、現在位置とユーザーの目的に関連のある情報を提示することができる。これは、実世界を情動的に拡張しているという意味で拡張現実感と言える。

実世界に電子のメディアを通して情報世界を融合するという拡張現実感においては、実世界の場面や状況の認識が大きな鍵となる。一方、自然言語処理において、ひとたび状況や文脈が特定されると、たとえ構文的に不完全であっても、言語表現の解釈を決定することはかなり容易になる。そこで、拡張現実感に、その状況認識を利用した効率的で頑健な自然言語処理を導入することが考えられる。

本稿では、拡張現実感と音声対話処理を統合したシステムについて述べる。

2 状況認識に基づく携帯型対話システム Ubiquitous Talker

Sony CSL では、日常的に存在する何らかの対象を認識すると、その対象に関する情報を対話的に提供するシステム Ubiquitous Talker を試作している [1]。このシステムは、状況認識を行ない関連情報を背景映像に重ねて液晶ディスプレイに表示するシステム(この部分は独立に動作させることができ、NaviCam と呼ばれている [4])と、認識された状況に従って適切な音韻モデルや知識ベースが選択される音声対話システムから構成される。

これは、持ち歩き可能なシステム(ただし、現在のところワークステーションとコードで繋がっている。将来は無線にすることを計画している)で、CCD カメラを装着した小型の液晶ディスプレイに写ったある対象に関する情報をディスプレイと合成音声を使ってユーザーに提供する。ユーザーは、音声言語を使って、システムに質問をしたり、システムの質問に答えることができる。

このようなシステムは、人間が特に意識しないで状況やドメインを特定でき、人間に過度な負担を与えないで自然言語処理を有効に利用する、重要な応用システムの一つと考えられる。

2.1 状況内インタラクション

ある特定の状況におかれていることがお互いの暗黙の了解になっている場合、コミュニケーションの効率はかなり向上する。それは、状況を共有しているために、省略や照応を使って言語表現を簡略化できるからである。ある場合は、言語表現自体が不適格であっても十分に意味のあるメッセージを送ることもできる。また参照表現などの状況を共有していないと使えないような言語表現も会話の効率を上げる原因になる。

また、人間は物理的な行為によってある状況から別の状況へ移動することができるが、この場合、ある状況に近づくことはその状況に関連した情報に近づくことを意味する。これは、情報空間を探索することに比べて、はるかに直感的で自然な情報検索であると言える。これはある意味で、状況に検索キーを持たせることになる。たとえば、本を見たいときは本棚に行くことが自然であるため、本棚のある状況そのものが本を検索するときのキーの一部になっていると考えることができる。

2.2 Ubiquitous Talker の構成

Ubiquitous Talker は、実世界の状況認識に基づく対話システムである。このシステムを使うと、あたかも注目している対象そのものと会話している雰囲気を作り出す。つまり、日常的に存在しているさまざまな対象を対話相手とすることが可能なのである。

図 1はこのシステムの構成を示している。

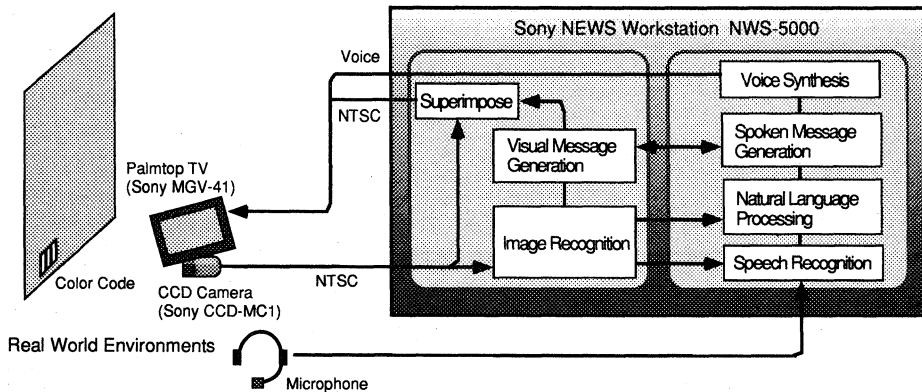


図 1: Ubiquitous Talker のシステム構成

また、図 2はこのシステムを使って、ある対象とインタラクションしようとしている様子と、対象が認識され、テキスト情報が背景画像にスーパーインポーズされた状態を示している。

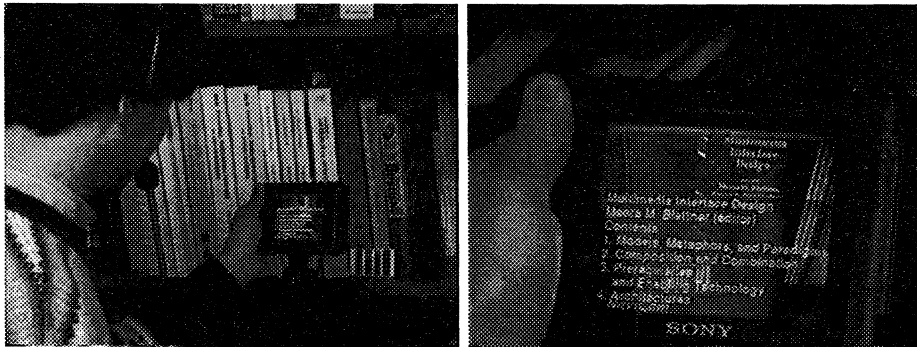


図 2: Ubiquitous Talker の使用形態 (左) とディスプレイ (右)

システムが実世界の対象を認識すると、音声合成により「これは〇〇です。私の提供できる情報は次の通りです」のようなメッセージがユーザーに与えられ、ディスプレイに簡単なメニューが表示される。このとき音声認識が起動される。音声認識のための辞書や文法は、認識された対象によって制限される。ユーザーは、声でメニューから項目を選択したり、その項目に関する質問をしったりすることができる。

状況認識 実世界状況を認識する手段として、液晶ディスプレイに装着した小型ビデオカメラを利用する。ビデオカメラからの映像はリアルタイムにワークステーションに取り込まれて処理される。実世界状況の認識を容易

にするために、実世界の対象にカラーコード (赤と青のストライプによって ID をエンコードしたもの) を添付する。このカラーコードによって複雑な認識技術を用いることなく、実世界の状況や対象を同定することができる。システムは、カメラから得られた映像の上に、認識したカラーコードやユーザーからの要求に従って生成されたメッセージ (テキストやグラフィックス) を重ねた画像を合成する。

生成された画像は液晶ディスプレイによってユーザーに提示される。画像認識を含む処理は、NTSC 信号とビットマップ画像の変換以外はすべてソフトウェアで行なっており、毎秒 10 フレーム程度の頻度で画像を更新している。

音声言語処理 状況認識に基づく音声言語処理は、音韻モデルや文法や知識ベースを状況に応じて切り替えることができるため、より頑健で効率的に行なうことができる。たとえば、カレンダーに向かっているときは、日付やスケジュールに関する質問に限定して語彙や知識を用意すればよい (ただし、スケジュールに関する質問は領域を限定できない場合があるので設計に注意が必要である)。

言語以外のモダリティを用いて文脈を限定して、より頑健な音声言語処理を行なうシステムに、SRI の Cohen らの開発したシステムがある [2]。これは、レンタカーの予約を液晶タブレットと音声を使って行なうシステムで、タブレットに表示されている予約フォーム (それをマルチモーダルフォームと呼んでいる) の各項目をペンで指しながら (手書き文字入力も許されている)、音声入力を行なう。ペンで指された項目に注意が当てられていることは明らかなので、その項目に関する語彙的に絞って音声認識を行なうことができる。

このシステムと Ubiquitous Talker との違いは、Ubiquitous Talker が携帯型であり、ユーザーが直面する実世界のさまざまな状況をシステムに認識させようとする点である。

2.3 Ubiquitous Talker の応用例 – もう一つの電子図書館 –

いわゆる電子図書館は本や雑誌などを電子化し、ネットワークで繋がれたコンピュータ上に仮想的な図書館を構築するという構想 [3] であるが、現実の図書館や本が存在する限りにおいて、現実と仮想の図書館の間の連携を考える必要がある。つまり現実の図書館におけるさまざまな状況において適切な情報が仮想図書館から呼び出され、人間に提示されなければならない。

Ubiquitous Talker は、現実の状況や対象と、それに関連した情報世界を融合するのであるから、図書館という状況においては図書館を補強する情報世界を創り出すことになる。この情報世界は仮想図書館と呼べるほど整備されたものでなくてもとりあえず構わない。たとえば、本そのものがその情報に含まれていなくても構わない。本は実世界の図書館に存在するので、情報世界に存在しなくても人間がアクセスすることは可能だからである。この場合、図書館と情報世界を合わせたものが電子図書館と呼べるものになる。

実際の利用例は次のようになる。

まず、図書館の入口で Ubiquitous Talker を使うと (たとえば、入口の案内板にかざして見る)、「ここは〇〇図書館です。どんな分野の本をお探しですか?」というメッセージが聞こえる。そこで「コンピュータサイエンス」と言うと、「こちらにどうぞ」という声と共にディスプレイに地図と行き方が表示される。地図に従って行くと分野ごとに整理された書棚のうちのコンピュータサイエンスの棚にたどり着く。そこで、次に Ubiquitous Talker をその棚にかざして見ると、「ここはコンピュータサイエンス関連図書の本棚です。何をお探しですか?」のようなメッセージが発せられる。「言語に関するもの」と言うと、「プログラミング言語ですか、それとも自然言語ですか?」と聞いてくる。「自然言語とは?」と聞くと、「人間の話す言葉のことです」のように答えてくれる。「プログラミング言語の本が見たい」と言うと、「プログラミング言語に関する本は、上から 3 段目にあります」と教えてくれる。そこで、そのあたりにある本のどれかを Ubiquitous Talker に映して見ると、「この本のタイトルは〇〇です」というメッセージと共に画面にタイトルなどが表示される。「著者に関して教えて」と聞くと、その略歴や業績について教えてくれる。画面には顔写真などが表示される。

表 1 に対話例を示す。

このように状況に応じたインタラクションを行なうことで、適切に情報が検索され、ユーザーをサポートしガイドしていくことができる。

表 1: Ubiquitous Talker の対話例

状況 (対象)	話者	発話内容	画面の表示内容
図書館の入口 (案内板)	システム ユーザー システム	「どんな分野の本をお探しですか?」 「コンピュータサイエンス」 「こちらにどうぞ」	分野リスト フロアの地図と本棚への道順
コンピュータ サイエンス 関連の本棚	システム ユーザー システム ユーザー システム	「何をお探しですか?」 「言語に関するもの」 「プログラミング言語ですか、 それとも自然言語ですか?」 「自然言語とは?」 「自然言語とは人間が話す言葉のことで、 その理解や生成に関する研究領域です」 「プログラミング言語の本が見たい」 「プログラミング言語の本は、 上から3段目にあります」	分野の下位分類リスト 説明文 場所情報
プログラミング 言語の本	システム ユーザー システム ユーザー システム	「この本のタイトルは〇〇です」 「著者は?」 「この本の著者は〇〇です」 「目次を見せて」 「この本の目次はご覧の通りです」	タイトルなどの情報 著者情報 目次

3 おわりに

マルチモーダル対話システムの一つの実現例として、拡張現実感と音声対話処理の技術を統合した Ubiquitous Talker を紹介した。このシステムは、実世界の対象にある種の能動的な存在 (エージェント) とすることが可能である。

このシステムの重要な点は、実世界状況に依存した情報処理を行なうことである。会話しているときの場所などの状況は、音声認識や言語処理の探索空間を制限するだけでなく、不完全で曖昧な発話から意味を抽出する手助けにもなる。また、人間は情報とそれをアクセスした状況とを結び付けて記憶しているので、場所などの状況を情報検索のキーとするやり方は有効である。

今後、現実の図書館における問い合わせ会話を分析し、状況に応じたインタラクションのより詳細な設計を行なう予定である。

謝辞

システムの設計と実装に協力してくれた計量計画研究所の乾裕子さんと東大の宮田高志氏に感謝します。

参考文献

- [1] Katashi Nagao and Jun Rekimoto. Ubiquitous Talker: Spoken language interaction with real world objects. Technical Report SCSL-TR-95-003, Sony Computer Science Laboratory Inc., Tokyo, Japan, 1995.
- [2] Sharon L. Oviatt, Philip R. Cohen, and Michelle Wang. Reducing linguistic variability in speech and handwriting through selection of presentation format. In *Proceedings of the International Symposium on Spoken Dialogue (ISSD-93)*, pp. 227-230. Waseda University, Tokyo, Japan, 1993.
- [3] 長尾真. 電子図書館. 岩波書店, 1994.
- [4] 暦本純一. Augmented Interaction: 状況認識に基づく新しいインタラクションスタイルの提案. 竹内彰一 (編), *インタラクティブシステムとソフトウェア II*, pp. 9-17. 近代科学社, 1994.