

マルチモーダルインタフェースによる知能化技術に関する調査研究

後藤 和弘・佐藤 辰雄・吉岡 孝*
情報産業部・*大分県立芸術文化短期大学

A Study on Intelligent Environment with Multimodal Interface

Kazuhiro GOTO・Tatsuo SATO・Takashi YOSHIOKA*
Information Technology Division・*Oita Prefectural College of Arts and Culture

1. はじめに

インターネットの普及によってさまざまな電子的な情報がネット上に溢れているが、その「情報」を人が活用する場合、従来はコンピュータを操作して「実体を持たない情報」のまま扱うことが一般的である。しかし、コンピュータを利用するにはまずマウスやキーボードなど基本的な操作方法を覚える必要があるため、初心者には馴染みにくく、デジタルデバイドを生む要因のひとつにもなっている。近年、家庭内やオフィス内にセンサ類を埋め込んだセンサネットワークや情報家電、人を支援するロボットが開発されてきたことで、今後は「実体のあるモノ」との相互作用を通じた情報の入出力技術が重要になると考えられる。ロボットを操作するためのユーザインタフェースについてはジェスチャや音声の利用が多く研究されており^{1,2)}、コミュニケーションが可能なロボットのニーズが高まっている³⁾。

我々はこれまでにシンクライアントを用いて、テレビ等の家電品や移動ロボットを音声で操作可能な遠隔操作システムを開発しており、PDAをクライアントにすることで小型軽量の端末による遠隔操作を実現している⁴⁾。しかし、PDAを常に携帯しておく必要があり、利用者には提示できる情報はテキスト、音（音声、音楽）、画像、動画などに限定されるという制約があった。

そこで本研究では、自ら移動できる市販のペット型ロボットをユーザインタフェースとすることにより、利用者が常時携帯することなく、音声での動作指令やペット型ロボットの行動などによるマルチモーダルな移動型ユーザインタフェースの実現を目標とする。

2. 移動型インタフェースによる遠隔操作

図1にシステム構成の概要を示す。ペット型ロボット（Sony, ERS-210）、大語彙連続音声認識エンジン Julius⁵⁾、TVなどを操作するためのコントローラをネットワークに接続する。ペット型ロボットは専用のCPUとOSからなる一種の組み込みシステムであるため、一般的なパーソナルコンピュータと比べてCPUの処理能力が低く、メモリ容量が少ない。今回利用したペット型ロボットはもとも音声認識機能を有しているものの、限られた容量で処理できる語彙力は十分とはいえない。そこで、ペッ

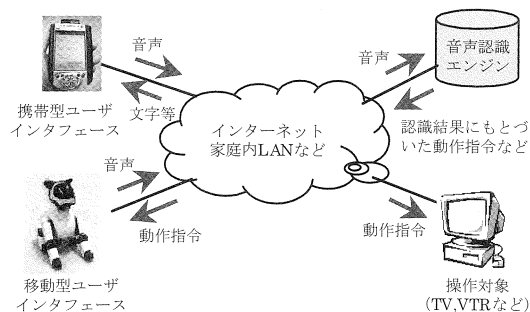


図1 システム構成の概要

ト型ロボットのマイクから入力した音声ネットワーク経由で音声認識サーバへ送信するユーザインタフェース・アプリケーションを開発した。音声認識エンジンは、ペット型ロボットから受信した音声データについて認識処理を行ない、登録してある動作指令と認識結果が一致すれば操作対象へ動作指令を送信する。これまでに開発した遠隔操作システムは柔軟性や拡張性に優れているため、従来のシステムにペット型ロボットを容易に組み込むことができる。

ペット型ロボットに組込むアプリケーションは、無償配布されている OPEN-R SDK⁶⁾を利用して開発した。OPEN-R はエンターテインメントロボット用のアーキテクチャーで、ハードウェアのモジュール化、ソフトウェアのモジュール化、ネットワーク対応などを特徴とする標準インタフェースである。関節部の制御、各種センサ情報やカメラ画像の取得、無線 LAN による通信など、基本的な動作を制御する API がソフトウェア仕様で規定され、OPEN-R SDK によって Windows 等のパソコン上で動作プログラムを開発できる。マイクから音声を入力するときに、入力されたことを検知する方法には幾つかの手法が考えられる。常にマイク入力を監視してそのレベルがあるしきい値を越えた時点トリガとする場合、入力された音がユーザの発声なのか雑音なのかを識別する必要があるうえ、ユーザが遠くにいたり音声小さいときには正しく音声認識できない可能性がある。また組み込みシステムであるため、あるプログラムが常に実行されていると CPU 資源やメモリ資源の利用の面から望ましいとは言えない。そこで本研究では本体のタッチセンサ

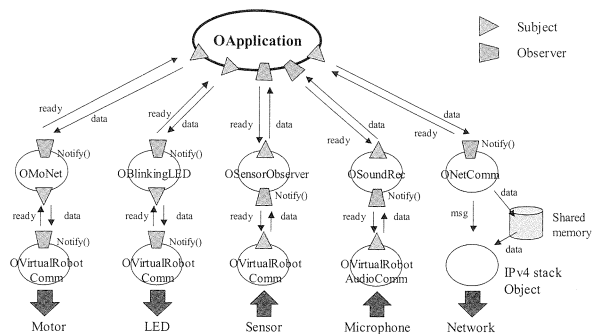


図2 機能オブジェクトの構成

を利用し、センサが押されたことをトリガとして音声入力を検知する。ユーザインタフェース・アプリケーションは図2に示すように複数のオブジェクトで構成し、音声やタッチセンサ、動作指令などの各データはオブジェクト間通信によって受け渡す。タッチセンサが押されると音声の入力を開始するとともに、ユーザへ開始を提示するために目の部分のLEDを点滅させる。音声を一時的に格納するバッファが一杯になると入力を終了し、LEDを消灯する。本体の仕様によりサンプリング16KHz、分解能16ビットで音声を入力し、2秒間に相当する64Kbyteの音声データを音声認識エンジンへ送信する。

音声認識サーバではDSR (Distributed Speech Recognition) 方式によって音声認識処理を行ない、認識結果をもとに操作対象を遠隔操作するための動作指令を送信できるよう、音声認識エンジンJuliusのソースを改造した。動作指令としては、テレビ等の電源のON-OFFなどのほか、ペット型ロボットの行動を指定することも可能である。これにより、ユーザとの対話内容と外部の他の情報（例：天気、イベント、メール確認など）を組み合わせ、次の行動を決定できるようになる。

3. システムの評価

開発したシステムの動作確認を行なった。ペット型ロボットのタッチセンサを押下して音声を入力し、音声認識の結果をもとに操作対象を遠隔操作できることを確認した。ユーザの発声から実際に対象物が動作するまでには、オブジェクト間での音声データの受け渡しや無線ネットワーク経由での音声データの転送等があるにもかかわらず、体感的には時間遅れはほとんど気にならない程度であった。

一方、ペット型ロボットがどのような仕組みでできているかを普段からほとんど意識しない女子短大生を被験者とした場合、座って待機状態であるペット型ロボットに対してほとんどの学生が頭に触れることからコミュニケーションを始める傾向があった⁷⁾。さらに顔の正面方向へ移動して接する傾向が強く、話し掛ける際は耳に近

付いて声を掛けようとする傾向が強いものの、ささやく程度の声であったために音声認識が困難な場面も見られた。また、「どのように接したらいいのかわからない」「どのぐらいの声を出したら伝わるのかわからない」「分からなかったら聞き返すそぶりをしてほしい」など、機械を意識した意見もあった。

4. まとめ

本研究ではペット型ロボットを移動型ユーザインタフェースとした、音声による遠隔操作システムを開発した。従来のPCのように文字や画像による情報の提示と比べ、行動による情報提示はコミュニケーションしやすいと考えられる。また、ネットワークを前提としたシステム構成であることから、ペット型ロボットへ直接話しかけなくても、PDAやPCなどの他の端末を用いて音声による動作指令を与えることも可能である。

さらに、音声指令に対応する行動プログラムを作成してネットワーク上のアプリケーションサーバに蓄積するとともに、WEBサービスと組み合わせることで、生活者を知的に支援する環境を構築できると考えられる。

参考文献

- 1) 松井俊浩, 麻生英樹, John Fry, 浅野太, 本村陽一, 原功, 栗田多喜夫, 速見悟, 山崎信行, "オフィス移動ロボットJijo-2の音声対話システム," 日本ロボット学会誌, Vol.18, No.2, pp. 300-307, 2000.
- 2) 水川真, 神名篤史, 松原安彦, 安藤吉伸, "物理エージェント (PAS) における, 音声操作系の基本検討," 日本機械学会ロボティクス・メカトロニクス講演会'01講演論文集, 2P2-K9, 2001.
- 3) 石黒浩, 日浦亮太, "コミュニケーション支援ロボットビジネス," 日本ロボット学会論文集, Vol.20, No.7, pp.672-675, 2002.
- 4) 後藤和弘, 佐藤辰雄, 大城英裕, 吉岡孝, 築根秀男, "シンクライアントを用いた音声コマンドによる移動ロボットシステム," ロボティクス・メカトロニクス講演会2002, 1P1-D10, 2002.
- 5) 大語彙連続音声認識エンジン Julius . <http://julius.sourceforge.jp>
- 6) OPEN-R SDK, <http://openr.aibo.com/>
- 7) 吉岡孝, 後藤和弘, 佐藤辰雄, 大城英裕, 本田久平, 築根秀男, "音声コマンドの転送を用いたペットロボットの遠隔制御," ロボティクス・メカトロニクス講演会2003, 2P1-1F-E6, 2003.