

人とのインタラクション機能を持つパーソナルロボット PaPeRo の紹介

大中慎一 安藤友人 岩沢 透

NEC マルチメディア研究所

パーソナルロボット PaPeRo は、ボタンを押して操作することで指示した情報や機能を引き出すだけのものではなく、人間と共に暮らし、家族の好みを覚えたりしながら、IT の恩恵を誰もが無意識に得られるインターフェースを目指して開発した「キーボードの無い将来の家庭用コンピュータ」の試作機であり、個性や表情を持った小型ロボットである。本稿では、PaPeRo の全体構成と、人とのインタラクションの中の特に音声インタラクション部分について紹介する。

The introduction of the personal robot PaPeRo

Shin-ichi Ohnaka, Tomohito Ando, Toru Iwasawa

NEC Multimedia Research Laboratories

A new personal robot named 'PaPeRo' is unlike anything seen so far in robot development, with its natural expressions and ability to remember its owners' interests and preferences. Following in line with NEC' s goal of finding solutions for the "i-society" , PaPeRo was designed to bring about a more natural interface with which people can easily and unconsciously benefit from the Internet. In this paper, we show the architecture and the interaction with humans, especially the verbal interaction, of the PaPeRo.

1. はじめに

近年、コンピュータや通信機器における技術革新のスピードは、人々の理解のスピードを越えて発展しており、新しい便利な機器が次々と登場するものの、同時に操作自体が難しい機器が多いのが現状であり、子供から高齢者まで、誰もが新しいテクノロジーを簡単に使えるようにする技術が求められている。また、半導体技術やメカトロニクス技術の向上によって、家庭用ロボットが現実のものになってきたことや、従来の機能中心のものではなく、一緒に暮らすことができるパーソナルロボットへの興味が高まってきている [1]。

我々は、画像認識・音声認識技術の家庭用機器への応用を研究するため、パーソナルロボットの試作 1 号機 R100 を開発した [2]。そこで提案した、人のパートナーとなり、いろいろな「やりとり」ができるロボットは、未来の様々な可能性を示すものとして注目された。しかし R100 では、1) 外部パソコンや無線ターミナルが必要で実験室から持ち出すことができなかった、2) 信頼性、安全性や使い勝手

が不十分、などの欠点があった。そこで今回、これらの欠点を改良した試作2号機 PaPeRo (Partner-type Personal Robot) を開発した。

PaPeRo には、(1) 両目に埋め込まれたカメラを用いた画像認識技術、(2) 内蔵された1つのマイクを用いた音声認識技術と3つのマイクによる音源方向検出技術、(3) モータとモジュール化された制御回路によるメカトロニクス技術、(4) インターネット通信機能やマルチメディア処理機能、(5) 動作や行動のプログラミングを容易にするソフトウェア技術、(6) スタンドアロンでの動作を可能にする高集積化技術、などが統合されている。

PaPeRo は、ボタンを押して操作することで指示した情報や機能を引き出すだけのものではなく、一緒に暮らし、家族の好みを覚えさせたりしながら、ITの恩恵を誰もが無意識に得られるインターフェースを目指して開発した「キーボードの無い将来の家庭用コンピュータ」の試作機であり、個性や表情を持った小型ロボットである。

2. PaPeRo の構成

PaPeRo の外観を図1に、構成を図2に示す。

両目にはCCDカメラが内蔵されており、頭部前方には音声認識用のマイクがある。また頭部のマイク後方には、タッチセンサ(スイッチ)が内蔵されており、叩かれたことが分かる。顔面には目と口と頬にLEDが埋め込まれており、表情を出すことができる。また、耳の部分にもLEDがあり、耳が点灯しているときに音声認識可能であることを人に示す。首の部分には撫でセンサーがあり、頭を撫でられたことが分かる。

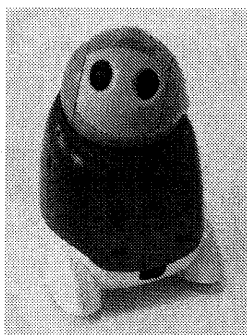


図1：PaPeRo の外観

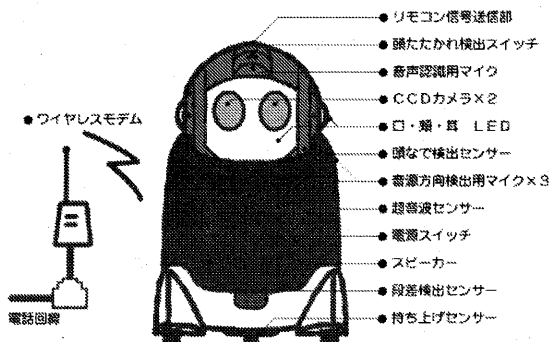


図2：PaPeRo の全体構成

胴体の上部には正面に1つ、サイドに2つ、計3つのマイクが埋め込まれており、それらのマイクへ到達した音の位相差から、360度の音源方向推定を行なっている。超音波センサを胴体正面に3個、後方に2個有し、画像認識による障害物検出と併用して障害物の検出を行なう。胴体正面にはスピーカもある。また胴体下部には持ち上げセンサと段差センサを有し、持ち上げられたことや、前方の段差を検出することができる。

図3に、PaPeRo の内部構成を示す。PaPeRo はノートPCをベースに構成されており、モータ制御部、センサ制御部、LED制御部、音源方向検出部は USB インタフェースにより接続されている。また、2

つのCCDカメラは IEEE1394インターフェースで接続されている。

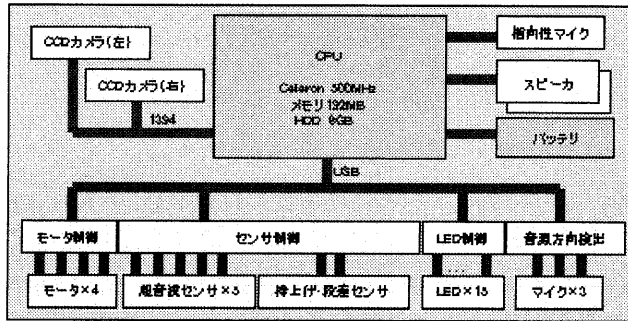


図3：PaPeRo の内部構成

3. PaPeRo の機能

次に、上記の構成のロボットで今回実現している機能を以下に記す。

PaPeRo は対話モードにおいて同時に約 650 種類の言葉を認識し、約 3000 種類の言葉を話すことができ、さらに人の顔を識別できるので、対話しながらキーボード無しで、(1) メールやメッセージを伝えたり、必要な情報を自発的に伝えてくれたりするなど、容易に操作できるインターネット機能、(2) 画像録画や音声録音を用いて家族間で伝言をやりとりすること、(3) ダンス、目覚まし、テレビリモコン機能など多様なやりとりを行なうことができる。

図4に PaPeRo の動作の流れと使用している技術を模式的に示した。PaPeRo には「散歩モード」と「対話モード」がある。散歩モードのときには、画像認識と超音波センサを利用した障害物検出を利用して家具や物を避けながら部屋の中を移動する。耳のLEDが点灯しているときに呼びかけると、音源方向検出により音源の方向を検出し、声のした方向に振り返り、画像認識によって人を探す。人が見つかったら距離を測り近づき、誰だか分かると名前を呼び、対話モードへ入る。

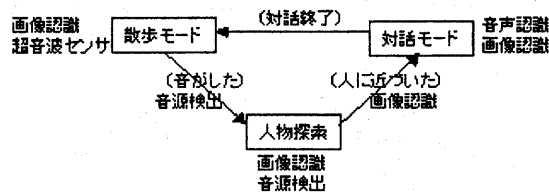


図4：PaPeRo の動作の流れ

対話モードでは、画像認識により、対話している相手の方を常に見つけるように、首や体を回転する。人が予め決められた言葉を話すと、ダンス、なぞなぞ、おみくじ、だるまさんがころんだ、TVリモコン、伝言、メール受信、時刻、タイマー、目覚ましなどの機能を使うことができる。

TVリモコン機能は、ロボットにTVのオンオフやチャンネルを音声で指示すると、赤外線によりTVをコントロールするものである。伝言機能では、伝言相手の家族を指定すると、内蔵マイクで取得した音声が入力され、次にロボットが伝言相手に出会ったときに伝言を再生する。メール受信では、家族宛てに

届いたメールをロボットが取り込み、宛先の家族に出会ったときにメールの読み上げを行なう。

4. 人との音声インタラクション

4-1. 音声認識インタフェース

移動ロボットと人間の自然なインタラクションを実現するためには、利用者の身体的自由度の高い音声認識インタフェースが必要だと考えられる。すなわち、我々の目指す音声認識インタフェースの特徴は、

- 1) 不特定話者音声認識
- 2) ハンズフリーな音声認識インタフェース
- 3) 人間-マイク間距離に融通性のある音声認識インタフェース

である。

このような音声認識インタフェースを構築する上で問題になるのが雑音や周囲会話、未知発話といった不要音声の棄却であるが、信号処理レベルで雑音と発話を分離したり、音声認識レベルにおいて複数の発話を個別に認識するようなことは現在研究がなされている段階であり、PaPeRo のシステムに組み込んで利用するにはまだ時間がかかる。

今回の PaPeRo の音声認識では NEC 製の音声認識エンジン Ver4.1 (SmartVoice Ver3.0 付属の音声認識エンジン) を、音声認識方式としては離散単語認識を使用することとした。人間がマイクから 1.5m くらい離れてもはっきりと話せば認識できることを目標としたので、マイクレベルを通常 (接話の場合) よりも高く設定する必要があるが、このために環境雑音を拾いやすい。また、ディスプレイなどの表示装置を持っていないので、ユーザが認識語でない未知語を発話する可能性が高い。そこで、環境雑音と認識語以外の不要発話に対処することにした。

我々は棄却専用の音声認識辞書を音声認識語彙に適合させ構築するという手法により、音声認識率を劣化させることなく、環境雑音や不要発話の棄却率を従来より向上させることができた [3]。また、棄却用音声認識辞書は、環境雑音を棄却するための雑音棄却辞書と、認識語以外の未知語を棄却するための発話棄却辞書から成り、環境雑音と認識語以外の未知語を区別して認識することが可能となっている。

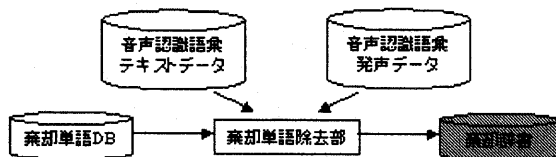


図5：棄却辞書構築システムの構成図

図5に棄却辞書構築システムの構成図を示した。雑音棄却辞書と発話棄却辞書は共に音声認識語彙とは無関係に構築された棄却単語DBから、音声認識語彙の認識に影響を及ぼす語彙を除去する方法により構築される。棄却単語の除去には音声認識語彙のテキストデータと音声認識語彙の発話音声収録した音声認識語彙発声データを利用している。詳細は文献 [3] を参照されたい。

4-2. 音声によるインタラクション

次に「対話モード」の概要を説明する。図6に人を見つけてから音声インタラクションへ入るまで、図7には音声インタラクションの主要部分を模式的に表した。

PaPeRo は人を見つくとステレオ視で距離を測りながら近づき、一定の距離まで近づくとあいさつをする。このとき人が識別できていれば、「〇〇ちゃん、おはよう」などという（“対話相手確認”）。ここでユーザが「ちがうよ」などと言う事によって“対話相手認証”へ行き、画像による顔識別をやり直すことが可能である。“記念日チェック”では予め登録された記念日や、後からユーザが登録した家族の誕生日などを教えてくれる。“履歴に基づく対話”では、これまでに対話をした履歴に基づいて「また会ったね」「久しぶりだね」などと言う。“今日は何の日”では今日が何の日であることを教えてくれる。そして家族から伝言があれば、“伝言チェック”でそれを教えてくれる。

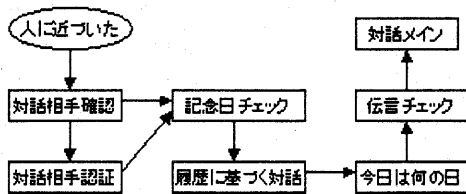


図6：人を見つけてから音声インタラクションへ入るまで

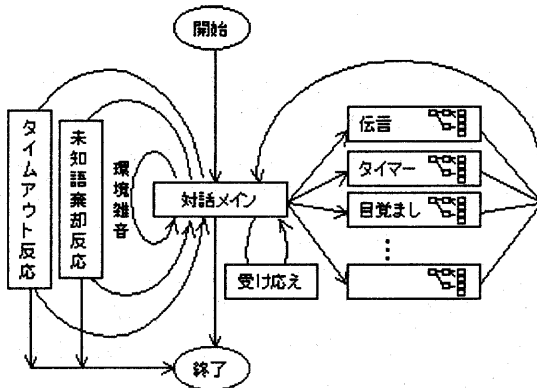


図7：音声インタラクションの流れ

PaPeRo はこの対話メインにおいて同時に約650の語彙を認識する。ユーザが予め決められたキーワードを言うと、伝言やタイマーや目覚ましなどへと遷移し、それが終了すると再び“音声入力待ち”へ戻るのが基本の流れである。“伝言”、“タイマー”、“目覚まし”などはそれ自体が独自の対話フローを持つが、ユーザに「おはよう」と言われて PaPeRo が「おはよう」と言い返すだけの一言の応対（“受け応え”）もある。

先に述べたように、PaPeRo の音声認識インターフェースは環境雑音と認識語以外の未知語を別々に認識するようになっているので、音声インタラクションにおいても、各々に対して別の反応をするようにしてある。すなわち環境雑音を認識した場合には何もせずにメインに戻る。一方、未知語として認識さ

れた場合には、ユーザが PaPeRo に向かって未知語を話したと判断して未知語棄却反応をする。未知語棄却反応では「ん?」「え?なあに?」などと言ったり、「ダンスして欲しいときは“ダンス”って言ってね」などというガイダンスを行なう。

人間の発話ではない雑音であるとか、人間の発話でも距離が離れている場合には、環境雑音として認識される。また、ユーザが認識しないキーワードを発話した場合や、正しいキーワードを発話したがはっきり発音しなかった場合は、未知語棄却反応が起こる。

タイムアウト反応は、一定時間音声の入力がなかった場合になされる処理であり、発話を促したり間をうめる発話を行なう。

発話棄却反応とタイムアウト反応については、各々について反応を行なうごとにカウントを行い、予め決められた回数に達すると対話を終了する。もちろん、キーワードで対話を終了することもできる。

さらに PaPeRo は感情や人物毎の好感度や親密度といった内部状態を持っており、接し方によって人毎に反応が変化する機能や、ロボットから相手に対して趣味や好みの質問をし、その答えを人毎に記憶して、相手によって話す内容や反応が変わるといった機能が組み込まれている。

内部状態は感情・好感度・親密度の値によって七つのカテゴリに分けられており、各々のカテゴリごとに反応が異なる。たとえば一例としてユーザが「おはよう」と言った場合を挙げると、ロボットはそのときの内部状態のカテゴリに応じて、「グッモ〜ニ〜ン!」「おーす」「おはよー」「おはよー」「おはよう(元気に・普通・暗く)」など様々な発話をする。

5. おわりに

PaPeRo は、R100 に比べて一回り小柄になり、信頼性・安定性・安全性や認識性能が向上した。さらに、全ての処理を本体内部で行なうことができるので、どこにでも手軽に持ち運ぶことが可能となった。将来的には、(A) ホームネットワーク、ホームセキュリティシステムとのインタフェースの実現、(B) 高齢者の見守り、遠隔医療システムや障害者のケアシステムとの連携、(C) 子供が楽しみながら勉強したり質問したりする教育ツール機能の提供、などの用途の広がり期待できる。

音声インタフェースには、棄却率をより向上させること、周囲で人間同士が話しているときにロボットが認識する語彙を発話すると認識してしまうこと、認識できる語彙を増やす必要があるなど、まだまだ改善の余地がある。ワードスポッティングの利用であるとか、人物検出とマイクアレイを組み合わせ使用することなども今後考えていきたい。

また、認識性能や動作・行動などの技術面と、人とロボットの関わり合いといったメンタルな面の2つの領域において、技術系大学や人文系大学と共同研究を進め、家庭用ロボットの可能性をさらに追求する計画である。

参考文献：

- [1]：田代泰典、「パーソナルロボットー人間との共存ー」、学術月報, Vol.53, No.9, p989, 2000.
- [2]：藤田善弘、「パーソナルロボット R100」、日本ロボット学会誌, Vol.18, No.2, p.40, 2000.
- [3]：岩沢透、「パーソナルロボット PaPeRo の音声認識インタフェース」、人工知能学会 AI チャレンジ研究会, Vol.13, p.17, 2001.