

ユーザーと家電機器をつなぐインタフェースロボット

Interface Robot to Connect Users to Home Appliances

山本 大介 土井 美和子

■ YAMAMOTO Daisuke

■ DOI Miwako

情報家電の普及に伴い、ユーザーが音声で簡単に操作できるインタフェースが求められている。しかし、従来の音声認識技術では誤認識が避けられず、高精度な音声認識のためには、ユーザーはあらかじめ決められた音声指示語を発話する必要がある。

東芝は、この問題を解決するため、まちがいを許容されながらしだいにことばを覚える幼児を参考に、親和性を利用した音声指示語の習得技術の開発に取り組んでいる。そして、この技術をインタフェースロボットに搭載し評価したところ、誤操作が許され音声での操作性が向上すること、とりわけ高齢者にとって効果的であることを確認した。

With the dissemination of information appliances in recent years, demand has arisen for an interface that allows users to easily operate such equipment by voice instructions. However, erroneous recognition still occurs in conventional speech recognition technology, and to achieve accurate speech recognition it is necessary for users to register selected terms in advance.

To overcome this problem, Toshiba has developed a new voice instruction learning method with the innovative feature of incorporating familiar infant behaviors. This method was installed in an interface robot and evaluated. The results of experiments demonstrated that users accept cases of incorrect recognition and that operations by voice instructions from users are useful, especially when the user is elderly.

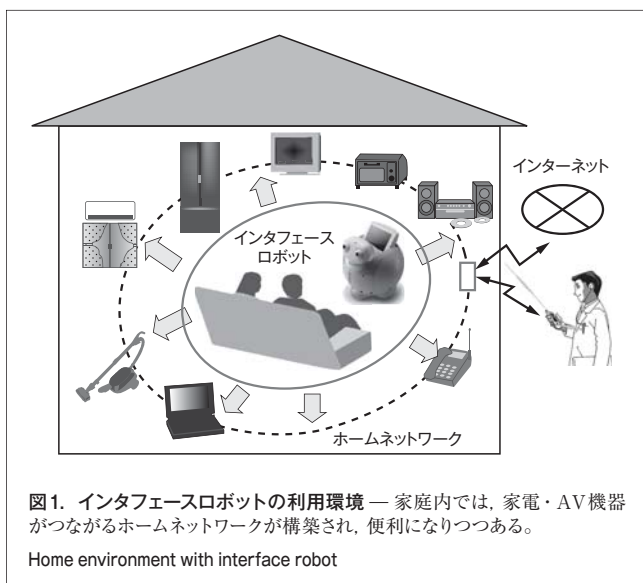
1 まえがき

近年、家庭内において、家電機器やAV機器を使ったホームネットワークの構築が進んでいる(図1)。しかし、これら情報家電のユーザーインタフェースは、必ずしも使いやすいものとは言えない。それは各家電機器、例えば、冷蔵庫が収納食品の在庫を表示したり、テレビ(TV)が画面を見ながら番組を選べる電子番組表(EPG: Electronic Program Guide)を提供したりと、多くの便利な機能を搭載するようになってきた

一方で、その多様さゆえに、ユーザーが欲する機能を実行するためには複雑な入力を強いられることが多いからである。

そこで、家電・AV機器とユーザーとの仲立ちとなり、簡単に操作できるインタフェースとして、音声で応答し、親しみやすい身体性を持つロボットを用いる方法が注目されている^{(1),(2)}。しかし、音声を用いたインタフェースには、音声の誤認識や、あらかじめ決められた音声指示語でしか操作できないという問題がある。

ここでは、東芝におけるインタフェースロボットの開発状況と、音声インタフェースの課題を解決する方法について述べる。

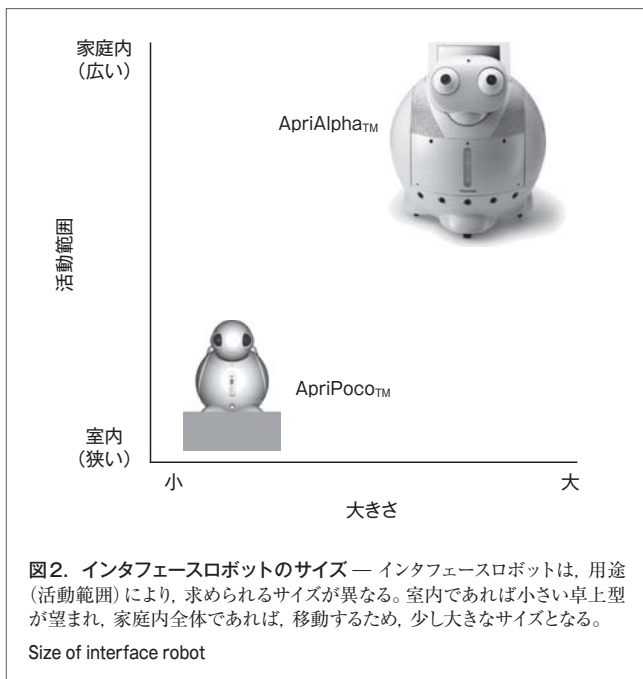


2 東芝におけるインタフェースロボットの開発状況

2.1 ロボットの大きさや用途

当社は、今までに、2種類のインタフェースロボットを発表している。2003年発表のApriAlpha™(アプリアルファ)は、留守番機能を特長の一つとし、床を移動することで家全体を活動範囲としている。一方、2008年に発表したApriPoco™(アプリアポコ)は、室内での使用に特化した、卓上型の小型ロボットである。

ロボットをユーザーインタフェースとして用いる場合の利点は二つある。一つは“親しみやすさ”である。詳細は4章で述べるが、擬人化された形状は、話しかけやすいという以上の利点がある。もう一つは“動く”ことである。ロボットは動くことで活動範囲を広げ、搭載されたカメラ、マイク、各種センサーで、



ユーザーや家庭内の状況を認識することができる。しかし、活動範囲を広げることは、ロボットのサイズを大きくすることにつながる。そして、サイズが大きくなると、一般的に“親しみやすさ”を損ないやすい。インタフェースロボットでは、その用途（活動範囲）に応じた適切なサイズを検討する必要がある（図2）。以下に、ApriAlpha™とApriPoco™の特長について述べる。

2.2 ApriAlpha™

ApriAlpha™⁽³⁾は、直径が350 mm、高さが380 mm、質量が約9.5 kgである。2眼のカメラ雲台である“顔”と、対象物へ移動する“脚”であり、体の向きを変える“腰”でもある2輪独立駆動方式の移動台車を備えている。顔の特徴として、一般のロボットは、首のパン・チルト動作で左右・上下の視野を確保するのにに対し、このロボットは、首のパン軸に加え眼球にもパン・チルト軸があり、大きな目が動くことが特徴となっている。また、機器とのインタフェース機能として、無線LANと赤外線リモコンの送受信器を備えており、ネットワーク経由でネット家電を操作することも、赤外線リモコンによりレガシー家電^(注1)を操作することもできる。また、背部には収納できるタッチパネル付きディスプレイを搭載し、画像を映したり、タッチパネルにより機器を操作することができる。

2.3 ApriPoco™

ApriPoco™（図3）は、卓上に置いてもじゃまにならないよう、210（縦）×210（横）×270（高さ）mmのサイズに抑えている。また、頭部に2軸、腕部には左腕、右腕それぞれに前後・上下の2軸ずつで計4軸、腰部に1軸の自由度を持たせ、様々

(注1) ネットワークへの接続機能を持たない従来の家電機器。

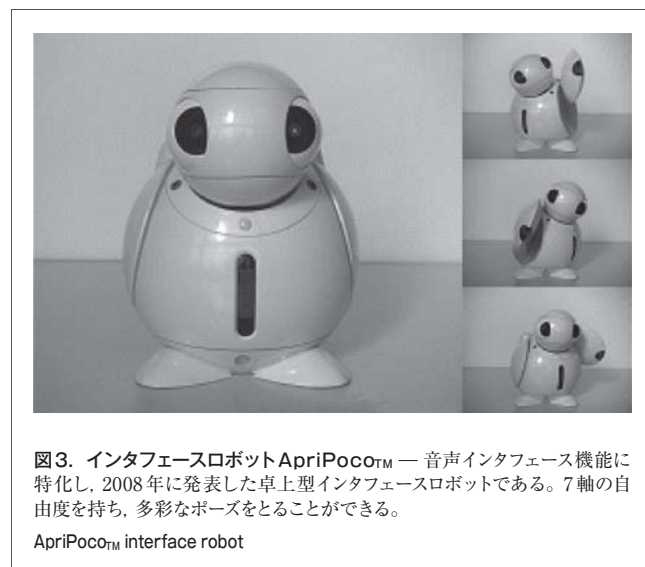


図3. インタフェースロボットApriPoco™ — 音声インタフェース機能に特化し、2008年に発表した卓上型インタフェースロボットである。7軸の自由度を持ち、多彩なポーズをとることができる。

ApriPoco™ interface robot

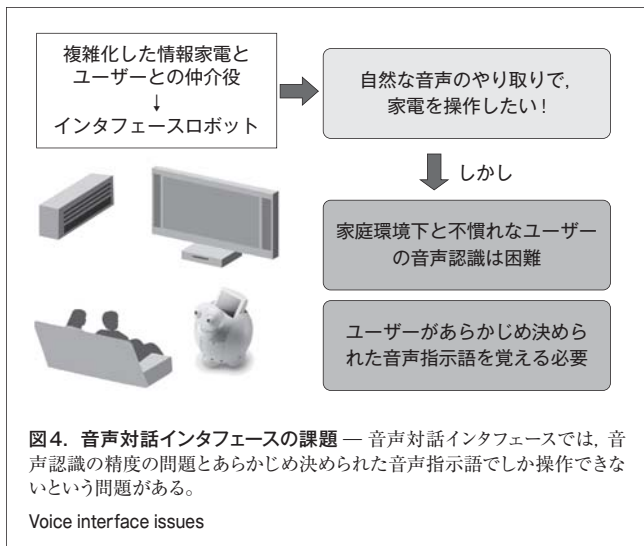
なしぐさや物を指し示す動作ができるようにしている。頭部は、肯定を示す“うなずき”，否定を示す“首振り”，及び疑問を示す“かしげ”を表現するため、チルト軸を斜めに配置した。ユーザーがロボットの顔を見ながら発話するとマイクがユーザーの口の正面になるように配置し、入力音声の音量レベルを示すインジケータを付けた。また、ユーザーからの“たたく”，“なでる”といった自然な入力を受けるために、前後・左右・上下の入力ができるタッチスイッチを頭部に搭載するとともに、ユーザーとの距離を測定するための距離センサや、人感センサ、照度センサ、温度・湿度センサも搭載した。また、赤外線リモコン送信部を右腕の先端に取り付け、あたかもロボット自身がリモコンで操作している感じを表現している。

3 音声対話インタフェースの課題

音声によるインタフェースは、複雑な入力操作なしで、ひと言で操作できるという利点がある。しかし、これには二つの問題がある。一つは音声の誤認識である。家庭環境の多様な騒音に加え、ロボットの音声認識ではマイクと発話者の位置関係が定まらず、音響特性は変動する。更に、子供から老人まで幅広くかつ不慣れたユーザーに対応する必要があり、誤認識は避けられない。もう一つは音声指示語の問題である。あらかじめ決められた指示語が発話されれば、高精度の音声認識が期待できる。しかし、ユーザーがそれを覚えるのは難しく、ユーザーは自分の使い慣れたことばで指示できることを望んでいる（図4）。

この二つの問題を解決するために、インタフェースロボットには“幼児の特質”を持たせている。

まず、音声の誤認識に対しては、誤認識しても許容されやすいよう、幼児らしい親しみやすい外見を取り入れた。すなわ



ち、幼児らしさを引き立てるため、ローレンツの幼児図式^(注2)にかなった形状を持たせており、これらは成人や成体に“かわいらしい”とか“愛らしい”という感情や気持ちを起こさせると言われている。

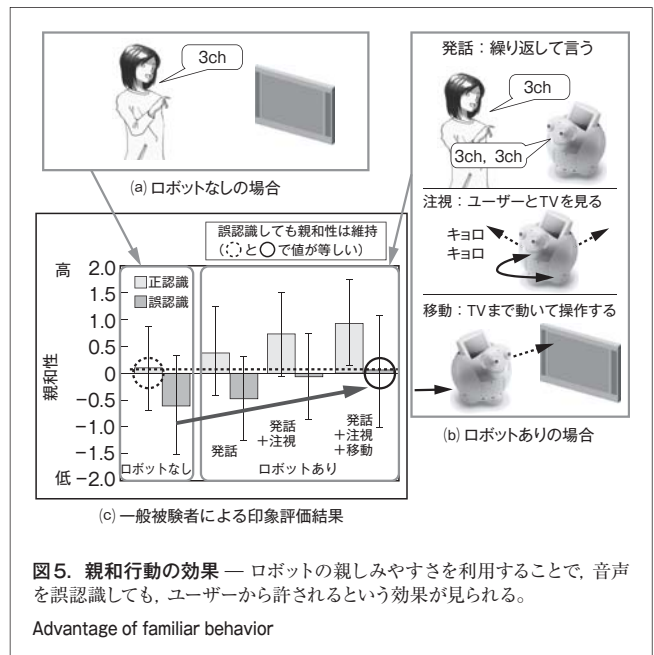
また、音声指示語に対しては、幼児がことばを覚えるときのように、ロボットがユーザーの音声指示語を聞いてそのことばを習得する手法を開発することで、ユーザーが決まった音声指示語をあらかじめ覚えなくても使えるようにした。以下、その手法と評価結果について述べる。

4 親しみやすさで許容されるインタフェース

音声の誤認識に対する親しみやすさの効果を調べるため、幼児の親和行動の中で最初に現れるコミュニケーション動作の注視・発話動作と、運動系の発達に伴って現れる移動動作を実装した。以下に、幼児の発達過程と対応する動作について述べる。

注視動作として、幼児が人の顔を好むことも知られており、ロボットにも人の顔を検出するとその方向を見る動作を持たせた。発話動作として、一語文^(注3)発話期の幼児を想定し、ユーザーの発話を単語レベルで認識して、その認識した単語を2回繰り返して発話させるようにした。移動動作としては、はいはいや伝い歩きを経て、興味のある方へ行く探索動作が現れる。ロボットには、車輪による移動ではあるが、操作対象までの移動機能を搭載した(図5(b))。

- (注2) 人間の幼児や動物の幼体を持っている、丸い体型、身体に比較して大きな頭、顔の中央よりやや下に位置する大きな目、丸々とした手足やおなどの形態的特徴。
- (注3) 「ワンワン」、「ブーブー」、「マンマ」など、1歳時期を過ぎたころに発話される一語で話すことば。
- (注4) 複数の形容詞対を用いたアンケートにより印象を評価する主観評価手法。



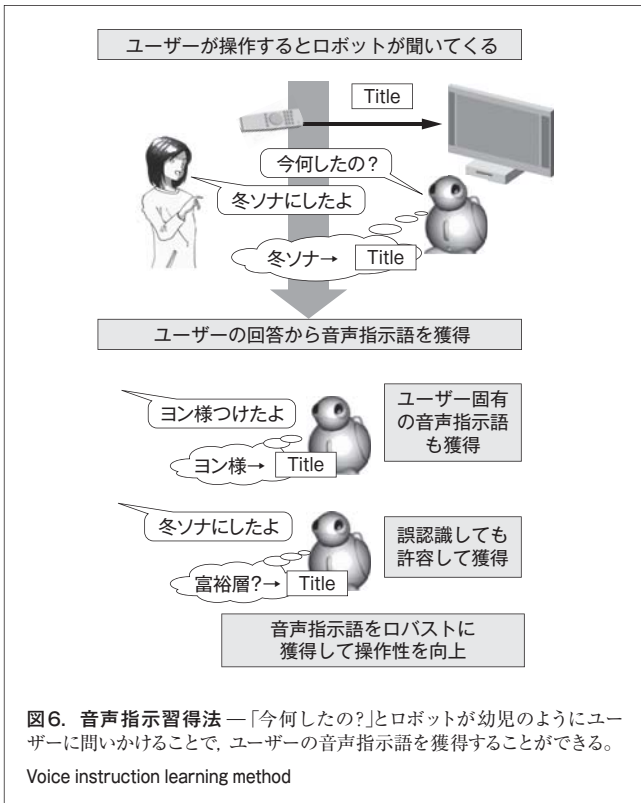
前述の音声の誤認識に対する許容への有効性を調べるため、高齢者や若者28名から成る被験者を対象に、SD (Semantic Differential) 法^(注4)による印象評価実験を行った。実験内容は、「2チャンネル (ch) にして」など、ユーザーの音声指示によりロボットがTVのchを切り替えるタスクを設定した。親和行動としてユーザーの指示を繰り返す発話動作、ユーザーとTVを交互に見る注視動作、及びchを切り替える際にTVの近くへ移動する移動動作を順に加えていき、評価基準となる、ロボットなしでTVに直接音声指示するパターン(図5(a))と合わせ、計4パターンの比較を行った。実験では、誤認識の影響を評価するため50%の誤認識を加えた。

印象評価の分析の結果、ユーザーとの親和性を示す“親和性因子”が得られた。親和性因子は、音声認識の正誤にかかわらず、発話、注視、移動という親和行動を加えるにつれて増加し、これらすべてを加えて音声を誤認識した場合と、ロボットなしで音声を正しく認識した場合がほぼ同じとなった(図5(c))。このことから、ロボットの親和行動が誤認識を許容させる効果があったと言える。

また、この結果を20代の若者と60代の高齢者に分けて集計したところ、高齢者は若者よりも親和行動による親和性の効果が高く、特に音声を誤認識した際にもあまり影響を受けないという結果が得られた。このことから、これらの親和行動が高齢者に対して特に有効であることがわかった。

5 音声指示を習得するインタフェース

一般的な音声操作では、機器の操作と対応する音声指示語があらかじめ決まっており、ユーザーは決まった音声指示語



を発することで機器を操作する。この場合、操作の数だけ音声指示語を覚えなければならず、多くの操作を音声で行う場合、決して使いやすいものとは言えない。

当社は、自然な音声のやり取りでユーザーのことは獲得する音声指示習得法を開発した。これは、幼児がわからないことを親に聞くように、ロボットがわからない操作をユーザーに聞き、教示してもらうことで、ユーザーのことはと家電機器操作の対応付けを獲得する(図6)。この手法の利点として、音声指示語をあらかじめ登録する必要がないだけでなく、通常では登録されえないユーザー固有の音声指示語で操作ができること、及び音声誤認識に対するロバスト性が高いことが挙げられる。また、この手法は、ネット家電でもレガシー家電でも、受信した信号をそのまま送信して操作できる機器であれば、ロボットが音声指示語習得の際に受信した信号を、そのまま送信することで操作ができるという利点がある。

この手法に関して、一般の被験者による評価実験を行った。その内容は、被験者が音声で家電機器を操作し、その際の音声認識率と音声操作率^(注5)を測定するとともに、SD法による印象評価も行った。その結果、音声指示習得法を用いたほうが、従来のようにあらかじめ決まった音声指示語で操作するよりも、音声操作率が向上することを確認した。また、高齢者と

(注5) 音声認識率は、正しく認識された回数/発話回数を、音声操作率は、正しく操作された回数/発話回数を表す。

若者の比較実験も行い、若者には「いちいちことばを教えるのはめんどろ」という否定的な意見がある一方で、高齢者からは「2歳の孫にことばを教えているよう」といった肯定的な意見を得ている。更に、SD法による評価でも、若者よりも高齢者にとって親和性が高いという結果を得ている⁽⁴⁾。

6 あとがき

家電機器とユーザーのインタフェースとして当社が開発した、音声操作を行うインタフェースロボットとその技術について述べた。

このインタフェースロボットは、音声誤認識の許容と音声指示の習得という、従来の音声操作を行うロボットの課題を解決する新技術を搭載しており、便利ではあるが複雑になってきている家電機器やAV機器の操作を簡便化し、とりわけ情報弱者と呼ばれる高齢者にとって有効であることが実験的に確認された。これは、このロボットが、ユーザーと複雑になった家電機器をつなぐインタフェースとして実用性があることを示しており、今後、実用化を目指して研究開発を進めていく。

文献

- (1) 藤田善弘. NECにおけるパーソナルロボットの開発. 日本ロボット学会誌. 20, 7, 2002, p.676-679.
- (2) 新庄 宏, ほか. マルチモーダル対話技術による知的ユーザーインタフェース. 日立評論. 87, 10, 2005, p.47-50.
- (3) 松日楽信人, ほか. 特集:ホームロボット技術. 東芝レビュー. 59, 9, 2004, p.1-55.
- (4) 山本大介, ほか. “ユーザーの音声指示を覚えるインタフェース-桌上インタフェースロボットへの適用”. 情報処理学会 第70回全国大会, 2D-3. 筑波大学, 2008-03. 情報処理学会. p.2-67-2-68.



山本 大介 YAMAMOTO Daisuke, D. Eng.
研究開発センター ヒューマンセントリックラボラトリー研究主務, 博士(工学)。ロボット技術の研究・開発に従事。日本ロボット学会, 日本機械学会, 情報処理学会会員。
Human Centric Lab.



土井美和子 DOI Miwako, Ph. D.
研究開発センター首席技監, 博士(工学)。ヒューマンインタフェースの研究・開発に従事。IEEE, 電子情報通信学会, 情報処理学会, ヒューマンインタフェース学会会員。
Corporate Research & Development Center