2005－CVIM－150（1）
2005／9／5

ユビキタス環境における顔認識・音声認識を組み合わせたロボット対話インタフェースの試作

佐竹 純二† 近間 正樹† 坂上 文彦‡ 尺長 健‡ 上田 博唯†
†情報通信研究機構 けいはんな情報通信融合研究センター
‡岡山大学大学院自然科学研究科

概要 「ゆかりプロジェクト」では、ユビキタスホームと名付けた各種センサや家電品がネットワークで統合された環境において､コンテクストアウェア型サービスを実現することを目指している。そして、この新しい形のサービスをユーザに適切に提供していくための対話型インタフェースロボットの開発を進めてきた。本稿では、顔認識と音声認識を組み合わせたロボット対話インタフェースを試作し、ユビキタスホームでの生活実証実験を行った結果について報告する。

Prototyping of Dialog Interface Robot combining Face Recognition and Speech Recognition in Ubiquitous Environment

Junji SATAKE† Masaki CHIKAMA† Fumihiko SAKAUE‡
Takeshi SHAKUNAGA‡ Hirotada UEDA†
† National Institute of Information and Communications Technology
‡ Okayama University

Abstract In the UKARI project we are aiming at realizing new context-aware services in the environment, named Ubiquitous-Home, in which various sensors and electric appliances are combined with over the network. We had been developing a dialog interface robot to provide the context-aware services for the user suitably. This paper reports a prototyping of a dialog interactive robot that combines the face recognition and speech recognition.

1. はじめに

近年、居住空間に埋め込まれたセンサからの情報によって人間の状態や行動を解析し、その状態や行動に応じたサービスを提供するというシステムの研究が進められている[1-3]。筆者らが属している「ゆかり（Universal Knowledgeable Architecture for Real-Life appliance）プロジェクト」でも、ネットワークで結合された家電製品や情報機器、各種センサが協調動作することによって、どのような新しいサービスが実現できるようになるのかという観点で研究を進めている[4-7]。

ここで、ユーザ毎に異なる好みへの対応や、システム設計者は良かれと思って実装した機能が、その時のユーザの状況によっては迷惑と感じられる場合なども存在し、このような問題に柔軟に対応できる普遍的なユーザインタフェースとして、コンテクストアウェア処理能力を高度化した対話システムが必要だと考えられる。この課題への取り組み方の一つとして、ゆかりプロジェクトでは、ユーザの目に見えない存在であるアンコンシャス型ロボットシステムと、実体として存在するビジブル型ロボットとの分散協調の実現性および実用性を追及することをプロジェクトの大きな目標の一つとしている。

本稿では、ビジブル型の対話型インタフェースロボットを試作し、顔認識と音声認識を組み合わせた生活支援サービスについて、一般人に対して実際に生活実験を行った結果を報告する。

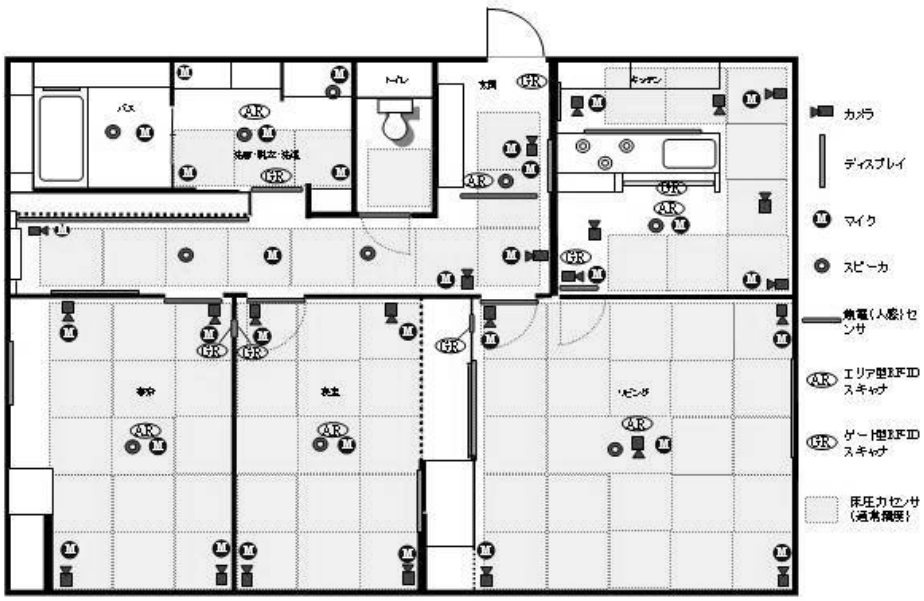


図1: ユビキタスホームの見取り図

2. ユビキタスホームと対話型インタフェースロボット

2.1 ユビキタスホーム

ゆかりプロジェクトでは、情報通信研究機構けいはんな情報通信融合研究センターのビル内に「ユビキタスホーム」と名付けた、マンションを模した居住空間を建設した。この施設は、図1の見取り図に示すようにリビングルーム、書斎、寝室、ダイニング・キッチン、浴室、トイレなどを完備し、ここで実際に一世帯の家族が生活することができる。また、この居住エリアの隣にはNOCと名付けた部屋があり、ここに各種処理を実行するためのプロセッサやデータベースサーバ、映像サーバなどを設置している。

この実証実験用住居には、ユビキタス関連技術を試すために、居住空間のあらゆる所に、ネットワークで結合されたセンサと家電製品を始めとする各種機器を設置している（図1参照）。図2はユビキタスホームの実際のリビングルームの様子である。天井に見えている黒い半球形のものがカメラであり、部屋の中央と四隅に設置されている。このように各部屋と玄関、廊下に天井カメラやマイク、RFIDタグ・スキャナ、床圧力センサ、人感（焦電）センサなどが取り付けられている。また、様々な場所にディスプレイや天井スピーカなども設置されており、ネットワーク経由で制御することができる。

以上のように、様々なセンサやアプライアンスをネットワークで結合し、それらを統合的に

図2：リビングルーム

管理することで、新しいサービスを実現することができる。このような環境においては、各種センサ情報に基づいて自律的にネットワーク上のアプライアンスを制御する、いわゆるアンコンシャス型ロボットの枠組みを基本とするのが一般的であり、ゆかりプロジェクトでもそうしている。

2.2 母親・子供メタファ

ユビキタスホームのような近未来型の住宅においては、機器や各種センサがネットワークで結合され、それらを統合的に管理することで、新しいサービスを実現することができる。ユーザが特に意識的に操作する必要がないアンコンシャス型のサービス提供は、エアコンの温湿度自動調整機能などでもお馴染みであり、かなりの範囲のサービスで有効で便利な機能であると

図3： 母親 - 子供メタファ

言える。しかし、今後ますます高機能化、多機能化が進むと、サービスの動作原理が複雑化してくる。そこで、ユーザのあいまいな要求を正確に受け取ってシステムを動作させ、同時にユーザに現在のシステムの状況をより的確に理解してもらうためには、対話インタフェースとして実体のあるビジブル型ロボットが有効であると考えられる。アンコンシャス型ロボットとビジブル型ロボットを協調的に動作させることで、ユーザに便利で快適なサービスを実現する。

複雑で高度なシステムにおいては、ユーザにシステムを適切に理解してもらうためのメンタルモデルが重要となる。ゆかりプロジェクトで提案している「母親・子供メタファ」[8]の概念図を図3に示す。家全体（アンコンシャス型ロボット）を母親ととらえ、いつも家の中にいて家族を見守り、必要なときにはどこからともなく現れてさりげなく家族を支援してくれる存在と位置付ける。一方、ユーザとの対話を受け持つビジブル型ロボットを子供とする。子供はユーザとの対話の内容に応じて、母親に対し適切なサービスの実行を依頼することができる。この子供–母親連携システムは、従来にない柔軟で高度な、そしてユーザに対する気配りの行き届いたサービスを実行することができると考えられる。

2.3 対話型インタフェースロボット

試作した対話型インタフェースロボットを図4に示す。自然言語による音声対話に関しては、残念ながら現在の音声認識の性能は十分とは言えないため、子供メタファを3歳児程度と位置付ける。ロボットの大きさは、日常生活の邪魔にならないサイズにするため、台所にあるサラダ油徳用瓶程度とした（体長約25cm）。また、家の中では移動速度の遅いロボットは邪魔になる（逆に高速移動は危険を伴う）ことと、物を持って運ぶような仕事ではなく、情報処理を主体とするため、移動はしない代わりに部屋のあちこちに置き、ユーザには同一のロボットがワープしてきて、対話を継続しているように見せる

図4： 対話型インタフェースロボット

図5：配置例（リビングルーム）

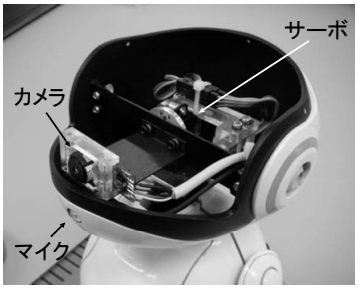


図6：ロボット頭部

図7：動作例（指差し）

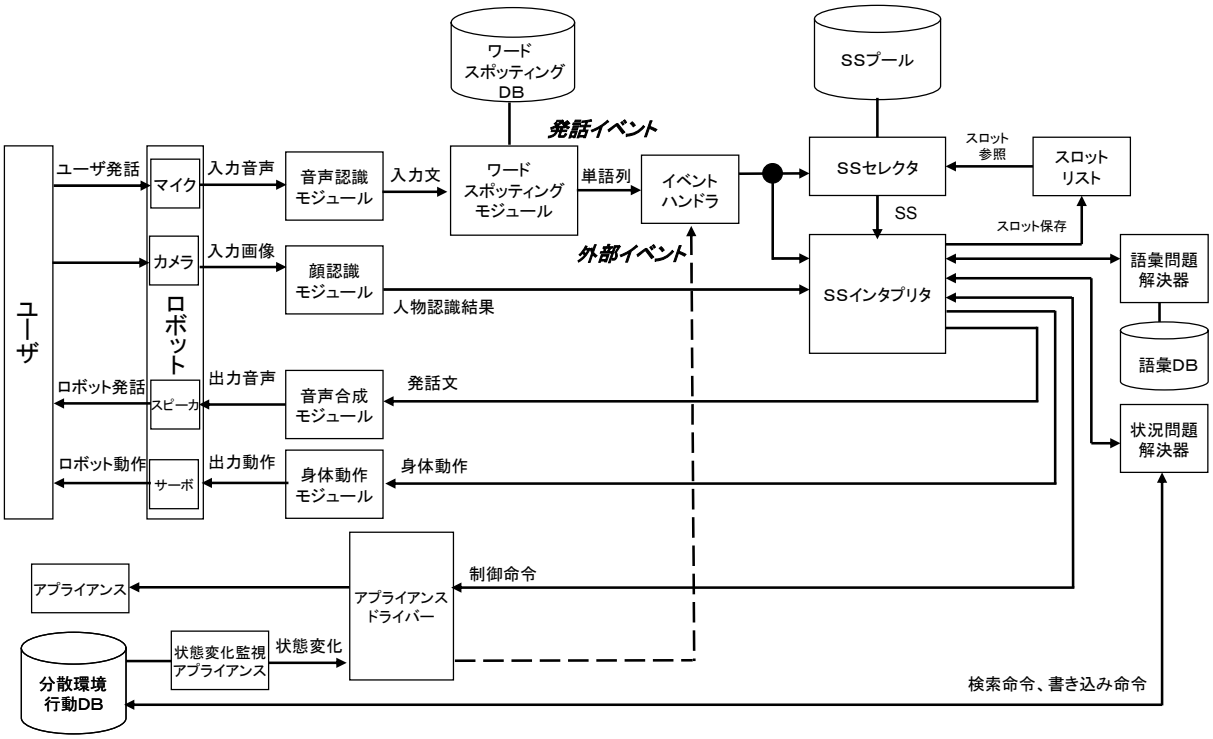


図8：対話システム構成図

ことにする。リビングルームで想定される配置の例を図5に示す（キッチンや玄関の配置例については後述の図10を参照）。

ロボットにはユーザとの対話を行うためのデバイスとして、頭部にUSBカメラ（Kanebo KBCR-M01VU-RUB03を横向きに設置）と単一指向性マイク（Sennheiser ME105）を取り付け、ロボット足元の台座内にスピーカを取り付けた（図6）。また、子供らしい仕草を可能とするため、首に前後回転・左右回転・傾きの3自由度、左右の手にそれぞれ上下1自由度、胴に左右回転1自由度、合計6自由度の動作を行うためのサーボを取り付けた。動作の一例を図7に示す。なお、ロボット足元の台座内には、スピーカの他、電源やサーボの制御ボードが入っている。

3. 顔認識・音声認識を用いた生活支援サービス

3.1 対話システムの概要

試作した対話システムの構成図を図8に示す。音声認識モジュールではマイクから入力された音声を認識し、ユーザ発話をテキストに変換する。顔認識モジュールではカメラで撮影した画像を用いて、ユーザが誰であるかを認識する。合成音声モジュールはスピーカを通してロボットの発話を行い、身体動作モジュールはロボットのサーボを制御し、ロボットに動きを与える。また、アプライアンスドライバーはアプライアンス（機器）の制御を行う。対話システムは、これらのモジュールを用いて生活支援サービスを実現する。具体的な流れについては**3.2**で述べる。

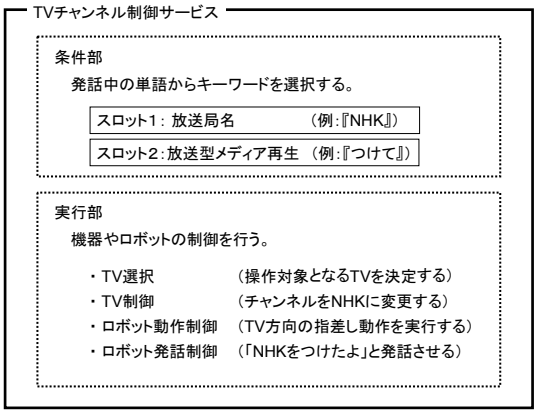


図9：サービスシナリオ（SS）の例

サービスのトリガとなるイベントには、発話イベントと外部イベントの2種類が存在する。発話イベントはユーザがロボットに向かって命令する発話、外部イベントは各種センサにより得られた家庭内の状態の変化である。本稿では、特に発話イベントをトリガとする生活支援サービスを取り扱う。

今回のシステムの音声認識には連続音声認識

ソフトウェアJulius[9]、顔認識にはDecomposed Eigenface法[10,11]を使用した。また、音声合成にはWizardVoice[12]を使用し、実際の子供の声をモデルにした音声データベースを用いた。

なお、音声認識にユーザがロボットの頭を撫でる音を学習させ、それを認識できるようにした。ロボットは頭を撫でられた時、手と頭を上下に振って喜び、直前に提供したサービスにユーザが満足したとみなす。

3.2 生活支援サービスの流れ

それぞれの生活支援サービスは、サービスシナリオ（SS）という形でSSプールに保持されている。そして、イベントにより発火条件が満たされると、SSはSSセレクタにより選択され、SSインタプリタにより実行される。

TVチャンネル制御サービスの例を用いて、サービス実行の流れを説明する（図8,9）。「NHKをつけて。」というユーザ発話が入力されると、その中から『NHK』と『つけて』というキーワードがワードスポッティングモジュールにより抽出される。SSセレクタは各SSの条件部をチェックし、全てのスロットが対応するキーワードで埋まったSSがあればそれをSSインタプリタに送る。この場合、スロット1に『NHK』、スロット2に『つけて』が埋まったTVチャンネル制御サービスがSSインタプリタに送られる。SSインタプリタはアプライアンスドライバー（TVの決定・チャンネルの制御）や身体動作モジュール（TV方向の指差し動作を実行）、音声合成モジュール（「NHKをつけたよ」と発話）を呼び出しながら、受け取ったSSの実行部の内容を実行する。

3.3 生活支援サービスの例

ユビキタスホームにおいて現在検討中の生活支援サービスの内、顔認識・音声認識を組み合わせたロボット対話により実現されるものを紹介する（図10参照）。

・TV番組推薦　ユーザがロボットに「TVをつけて。」と言うと、顔認識によりユーザを認識し、放映中のTV番組の中からそのユーザの好みに合った番組をつける。ユーザのTV番組嗜好情報は普段見ている番組から学習する。また、番組を推薦した時、ユーザが「ありがとう。」と言うと、その推薦が正しかったと判断し、嗜好情報にフィードバックを行う。

・料理レシピ提示　顔認識によりユーザを認識し、ユーザが発話したキーワード（食材や料理名など）をもとにした料理に関する対話（連想しりとり[13]）や検索・推薦を行う。そして、ユーザが選択した料理のレシピをTVに表示したり、図10(a)に示すようにキッチン内のハッチの半透明スクリーンに投射して表示したりする。将来的には、冷蔵庫に残っている食材の情報なども推薦に利用する。

(a) 料理レシピ提示

(b) 忘れ物チェック

図10：サービス例

・忘れ物チェック　予め各持ち物にRFIDタグを取り付けておき、玄関の下駄箱内に取り付けたリーダで何を持っているかを読み取ることで、忘れ物のチェックを行う。床圧力センサによりユーザが外出しようとしているのを検知し、ロボットが声をかけ、忘れ物があれば発話により知らせる。顔認識によりユーザを認識し、そのユーザが持っているべきもののリストをデータベース（その日の行動予定なども格納されている）から参照し、チェックする。

4. 実験

4.1 実験環境

・被験者 … A（30代男性）、B（30代女性）、C（3歳児）の3人家族。

・実験期間 … 食事や睡眠を含め、12日間ユビキタスホームで生活してもらった。

・対話型ロボット … リビング、キッチン、玄

関、寝室、書斎の５ヵ所にそれぞれ設置した。

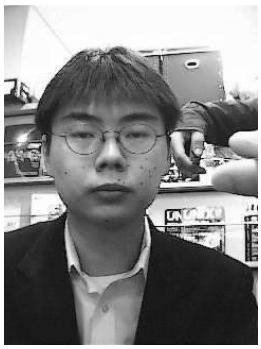
・ロボットの基本姿勢 … サーボの駆動音が音声認識に悪影響を与えるため、人物顔の追従はさせず、ロボットの基本姿勢は図5のような向きで固定とした。また、被験者にはロボットの正面から話しかけてもらうように指示した。リビング以外についても同様に、話しかける被験者の顔があると思われる方向を向くようにロボットの基本姿勢を調整した。

・生活支援サービス … ロボット対話によるTV制御（電源、チャンネル）、TV番組推薦、料理レシピ提示、目覚ましを対象とした。

・センサ類 … ユビキタスホームに設置された天井カメラ、マイク、床圧力センサ、扉や戸棚・引き出しの開閉センサ、睡眠センサ、RFIDタグリーダ等を用いてデータ収集を行った。ただし、風呂場やトイレにはカメラは無く、寝室もカメラにカバーを被せて撮影できないようにした。

・アンケート等 … サービスやロボットに関するアンケートを毎日記入してもらうものと、期間全体に関して最終日に記入してもらうものの2種類行った。また、後日、被験者に対してインタビューを行った。各サービスに関する細かい点については、必要に応じてメールで被験者に問い合わせた。

図11：実環境下での使用状況



(a) 距離60cm

(b) 距離20cm

図12：ロボットとの距離による撮影画像の違い

4.2 実験結果

センサ情報や各サービスの実行結果、アンケートやインタビューから、以下のような知見が得られた。

1) 天井カメラに比べてロボットのカメラは抵抗が少ない … 「最初は天井カメラで撮影されていることに抵抗を感じた」という意見が得られたが、ロボットのカメラについては特に指摘されなかった。天井カメラでは監視されている感があるのに対し、ロボットの視点で見られることには抵抗が少ないものと考えられる。また、天井カメラについても、「3日目には撮られていることに抵抗が無くなった」という意見が得られている。

2) 音声認識は3日程で慣れる … 実験開始直後は何度繰り返しても正しく音声認識できなかったが、3日後には多くても2,3回の言い直しでほぼ認識できるようになっていた。実験で使用した音声認識には学習機能が無いため、どのように発話すればうまく認識してくれるのかを被験者が学習した。ただし、被験者C（3歳児）の声は最後までほとんど音声を認識することができなかった。

3) 顔認識と音声認識の適正距離が異なる … 顔認識は図5のようにロボットとユーザの距離が60～100cmの場合を想定していたのに対し、音声認識では実環境における騒音を考慮した場合、現在のシステムでの適正距離が図11のように20～30cmであることが分かった。また、音声認識の失敗が続くと、被験者はロボットにどんどん近づいて話しかける傾向があることが分かった。

4) 近距離では視野から顔が切れる …ロボットとユーザとの距離が60cmの場合には人物顔のサイズが約100pixel（垂直画角38度、画像サイズ240×320pixels）であるのに対し、距離が20cmの場合には人物顔が視野いっぱいになってしまう（図12）。このため、近距離では顔が少しずれただけで視野から切れてしまい、顔認識を行うことができない。

5) ロボットの環境差異を個性と感じた … 被験者Bのアンケートより「台所のロボットは割りと私の声に応えてくれて、仲の良い友人のような感じ。リビングの子はよく考え込んだりする真面目で頑張りすぎるタイプ。寝室の子はあまり言うことを聞かないけれど、よく手を振って

喜んでいる無邪気な子供。」という感想が得られた。音声認識のプログラムは同一であるが、被験者は環境による音声認識率や誤認識の違いをロボットの個性と感じていると考えられる。（寝室の動作は、ドア開閉音を頭を撫でられた音と誤認識した結果の喜び行動によるものであった。）

6) 同じロボットが複数存在する違和感 … 被験者Bから「キッチンでロボットと対話している時、同時にリビングで同じ声で被験者Aと対話するロボットを見て、さみしい気がした」という意見が得られた。同じ形状、同じ声、同じ名前で存在する複数のロボットをどのように位置付けるか（ユーザにどのようなメンタルモデルを形成させるか）を検討する必要がある。

7) リアクションの必要性 … 被験者が話しかけてもロボットが無反応だった時、ロボットに「怖さ」「不気味さ」を感じていた。また、反応が無いロボットに対し、電源が入っていないのか、音声認識に失敗したのか、正しく認識したが理解できない（反応できない）のか等が分かるようにして欲しいという意見が得られた。

以上の結果は、実生活の中で実験を行うことで明らかになった事項であると考えられる。

4.3 考察

実験による知見**3)**,**4)** に対応するため、ワイドコンバーションレンズ（KenkoDIGITAL MPL-WA）を取り付けた（図13）。これにより、垂直画角が58度となった。ただし、広角レンズを取り付けたことで映像に歪みが生じたため、歪み補正[14]を行っている。この時の撮影画像の例を図14に示す。実環境下で音声認識と顔認識を併用する場合に、より適した画角が得られていると考えられる。また、ユーザはロボットに話しかけるために近づく際、特に自分の口をマイクに寄せようとするため、図13(b)のように顔の上部が切れやすい傾向がある。これはロボットのマイクとカメラの距離が近く、かつカメラ位置がロボット頭部のやや低い位置にあるためである。そこで、図15のようにカメラを少し上向き（ロボットの視線方向より約6度上向き）にすれば、さらに改善されると考えられる。

知見**7)** に対応するため、LEDを用いて状態を表示することや、ユーザ発話に対して「○○って何？」「わからないよ」とリアクションを返すことを検討している。ただし、実環境下での使用を想定した場合、環境雑音やロボット発話、ロボットの動作音などに反応して、延々とループしてしまう危険性もある。

(a) 外観　　(b) 内側

図13：ワイドコンバーションレンズの取付け

(a) 距離60cm　　(b) 距離20cm

図14：ワイドコンバーションレンズ取付け時の撮影画像例

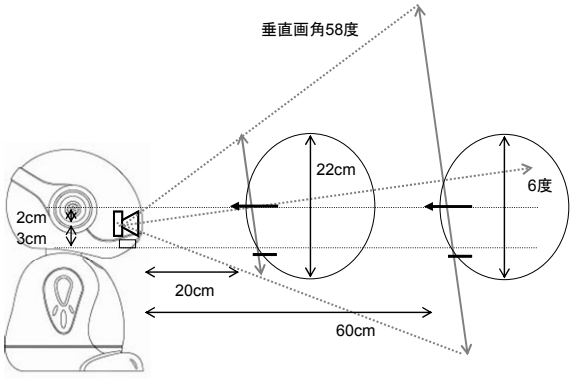


図15：ロボットとユーザの視線を合わせた場合の位置とカメラ画角の関係

知見**5)**,**6)** に関しては、ロボットに個体差をつけるべきなのかどうか、より深い検討を行う必要があり、今後の大きな課題の一つである。

5. まとめ

本稿では、ユビキタスホームにおけるインタフェースとなる対話型ロボットを試作し、顔認識と音声認識を組み合わせた対話システムを提案した。また、一般人に対して生活実験を行い、

実際に生活を行うことで得られた知見について報告した。

今後の課題としては、実環境における顔認識・音声認識の精度向上や、顔認識と音声認識を用いた新しいサービスの提案、また、複数のロボットをどのように取り扱うのか等を深く検討する必要がある。

参考文献

[1] Cory D. Kidd, Robert Orr, Gregory D. Abowd, Christopher G. Atkeson, Irfan A. Essa, Blair MacIntyre, Elizabeth Mynatt, Thad E. Starner, and Wendy Newstetter, "The Aware Home: A Living Laboratory for Ubiquitous Computing Research", In the Proceedings of the Second International Workshop on Cooperative Buildings — CoBulid'99. Position paper, October (1999)

[2] 佐藤知正, 森武俊, 西田佳史, 平成 13 年度未踏ソフトウェア創造事業 佐藤・森・西田プロジェクト研究計画, http://www.ics.t.u-tokyo.ac.jp/ipa/ipa2001/

[3] 独立行政法人 産業技術総合研究所デジタルヒューマン研究センター, http://www.dh.aist.go.jp/

[4] 美濃導彦, "ゆかりプロジェクトの目的と概要 —UKARI プロジェクト報告 No.1—", 情報処理学会第 66 回全国大会, pp.5-5～5-8(2004)

[5] 山崎達也, 沢田篤史, 多鹿陽介, 大倉計美, 中尾敏康, ヌリ シラジ マハダド, 佐野睦夫, 金田重朗, "ゆかりプロジェクトにおける分散協調基盤ミドルウェア —UKARI プロジェクト報告 No.2—", 情報処理学会第 66 回全国大会, pp.5-9～5-12(2004)

[6] 土井美和子, "分散環境行動 DB と場モデルに基づくユビキタスインタフェース設計—UKARI プロジェクト報告 No.3—",情報処理学会第 66 回全国大会, pp.5-13～5-16(2004)

[7] 上田博唯, "ユビキタス生活支援のためのロボットインタフェース —UKARI プロジェクト報告 No.4—",情報処理学会第 66 回全国大会, pp.5-17～5-20(2004)

[8] 上田博唯, 佐藤淳, 近間正樹, 木戸出正継, "アンコンシャス型ロボットとビジブル型ロボットの協調メカニズム —母親・子供メタファ—", 第 28 回ヒューマンインタフェース学会研究会, ヒューマンインタフェース学会研究報告集, vol.6, no.3, pp.57-64(2004)

[9] http://julius.sourceforge.jp/

[10] F. Sakaue, K. Shigenari and T. Shakunaga, "Robust Face Recognition by Combining Projection-Based Image Correction and Decomposed Eigenface," Proc. International Conference on Automatic Face and Gesture Recognition, pp.241-247, 2004.

[11] T. Shakunaga and K. Shigenari, "Decomposed Eigenface for Face Recognition under Various Lighting Conditions", Proc. IEEE conference on Computer Vision and Pattern Recognition, vol. 1, pp.864-871, 2001.

[12] http://www.ai-j.jp/

[13] 佐藤淳, 近間正樹, 上田博唯, 木戸出正継, "対話型ロボットにおける連想しりとり型対話戦略とその評価", 情報処理学会第 67 回全国大会, (2005)

[14] Z. Zhang, "A flexible new technique for camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22(11), pp.1330-1334, 2000