



「ロボカップ@ホームリーグ」

第8回

■ 杉浦孔明 (情報通信研究機構)

 応
般

ロボカップ@ホームとは

レフェリーがドアを開けると、1台のロボットが部屋の中に入ってくる。ロボットはレフェリーに尋ねる。

ロボット：「何を取ってきましょうか？」

レフェリー：「コーンフレークを取ってきて」

ロボット：「分かりました」

ロボットは移動を始める。部屋の外では、チームメンバが不安そうに見守っている。周りには、カメラを構える観客の姿もある。ロボカップ会場に現れたモデルルームに興味を持って足を止めたようだ。

ロボカップ@ホームリーグ(アットホームと読む。以下ロボカップ@ホームと略記)は、日常生活で人間を支援する自律ロボットの競技である。ロボカップ@ホームは、将来の日常生活でのロボットの応用を目指し、人間を支援するロボット技術を開発することを目的としている。

ロボカップ@ホームでは、家庭・オフィス・スーパーマーケットなどにおけるロボットの応用を想定したタスクが設定されている。主な競技フィールドは2LDK程度のモデルルームで、家具や食器等が配置された、実際の生活環境に近い複雑な環境が用意される。図-1に2009年グラーツ大会で用いられた競技フィールドを示す。この環境内で、ロボットが、日用品の探索、棚からユーザに言われたものを取ってくる、人を追従するなど、日常生活に役立つ機能を制限時間内にどれだけ達成できるかを競う。

ロボカップ@ホームは2006年にエキシビジョンとして始まり、2007年に正式競技となった比較的新しいリーグである。2011年には、11カ国の大学・研究機関から19チームの参加があった(表-1参照)。各チームは、6～10人程度のメンバを有している。ジャパンオープンでは参加チームが少ない年が続いていたものの、2011年には6チームの参加があった。

日常生活環境における自律ロボットの競技会は、移動マニピュレーションを主眼とした Mobile Manipulation Challenge や、物体探索を主眼とした Semantic Robot Vision Challenge などでも有名である



図-1 ロボカップ@ホーム 2009 グラーツ大会の様子

が、現状では参加チームが少ない。したがって、ロボカップ@ホームは世界最大規模の生活支援ロボットの競技会であるといえる。

タスクとスコア

タスク概要

2011年のタスクの概要を表-2に示す。タスクは計11種類あり、各タスクには、制限時間と配点(1000点または2000点満点)が設定されている。たとえば、Who Is Who タスクでは、音声認識に成功すると25ポイント、人検出に成功すると50ポイントが与えられる。

公式ルールの最新版は公式サイト¹⁾からダウンロードできる。2009年グラーツ大会のタスクに関

年	事前登録チーム (書類選考前)	参加チーム
2006	20	11
2007	16	11
2008	18	14
2009	25	18
2010	31	24
2011	30	19

表-1 世界大会参加チーム数の推移

しては、文献2)が詳しい。これまでの世界大会における得点傾向が文献3)にまとめられているほか、各チームの獲得スコア情報が大会Webサイト上で公開されている。

1次予選の総合得点により上位半数のチームが2次予選に進み、2次予選の上位5チームが決勝に進む。決勝では、各チームが自由デモを行う。デモに対する審査員の評価点と、それまでの得点を合計し

ステージ	タスク名	概要
1	Robot Inspection and Poster Session	競技形式で移動機能・停止スイッチ等の動作を確認する
	Follow Me	未知ユーザに対して1分以内にキャリブレーションを行い、追従する
	Go Get It!	フィールド内にある4つのオブジェクトを探索する
	Who Is Who	フィールド内にいる5人を見分けて、顔と名前を覚える
	General Purpose Service Robot 1	ランダムに生成された複文の指示(例「リビングへ移動して、緑茶を取って、サイドテーブルに届けて」)を実行する
	Open Challenge	ロボットの性能に関する、研究面で優れている点のデモ
2	Enhanced Who Is Who	フィールド内にいる5人を見分けて、飲み物を届ける
	General Purpose Service Robot 2	曖昧な指示を明確化(例「何の飲み物を持ってきますか」)し、指示を実行する
	Shopping Mall	未知環境で棚から3つのオブジェクトを持ってくる
	Demo Challenge	毎年変更されるテーマに沿ったデモ
決勝	Final	自由に選んだテーマに沿ったデモ

表-2 ロボカップ@ホーム 2011 イスタンブール大会のタスク



て最終順位が決定する。

これからロボカップ@ホームに参加しようとする場合、イメージをつかむ手段としては、過去の大会のビデオが効果的である。“RoboCup@Home”をキーワードとして、YouTube等を検索されたい。また、技術情報を共有するために、ロボカップ@ホームキャンプと題した講習会が開催されている。2010年より開催されているロボカップ@ホームキャンプには、50人程度の参加がある。

タスクの標準化

ルールは、テクニカルコミッティ (TC) による議論に基づき、毎年更新される。毎年1～3月にルールブックの暫定版が公開され、チームからのフィードバックや地区大会での試用を経て、5～6月に最終版が公開される。TCメンバは前年の大会においてチームリーダーからの投票により選出される。アジア、ヨーロッパ、アメリカからの人数バランスを考慮して、3～5人が選出される。筆者は2008年からTCを務めている。

タスクの設計に際しては、生活支援ロボットのベンチマークテストとして標準化することを重視している。つまり、生活支援ロボットの研究開発のための、統一的な実験設定や評価尺度となることを目指している。生活支援ロボットの研究開発においては、独自の環境や評価尺度が用いられることが多く、手法同士を比較することが難しい。タスクを標準化することで、比較評価のコストを低減すれば、コミュニティ全体の研究開発に貢献できるであろう。

タスクの難易度は、前年の結果を受け修正される。たとえば、2008年蘇州大会では既知ユーザの追跡に多くのチームが成功したため、2009年グラーツ大会以降は、未知ユーザの追跡に修正された。個々のスコアに関しては、音声認識と物体認識のバランスをいかにするべきか、など議論すべき点が多い。複数の案がある場合は、TC内で投票が行われ多数決により決定する。

実際のTCの議論は、メールや掲示板を用いて行われる。2010年からは、予備議論はWeb上で公開

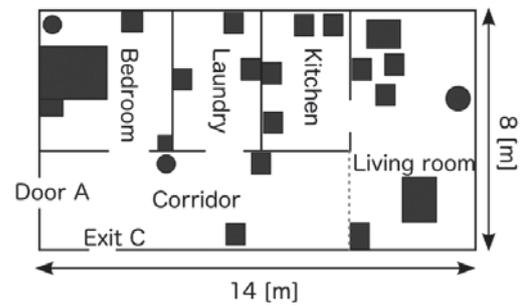


図-2 2011年イスタンブール大会におけるフィールド

されるようになった。ロボカップ@ホームの参加チームに所属していれば、投票権はないものの予備議論自体には参加することもできる。

これまでの経験から感じたことは、ルールに関する議論に加わらないと「もったいない」ということだ。他のリーグでも似た状況であると聞かすが、議論に参加する日本人は少ない。仮に、自分が疑問を感じるルール変更がなされようとしている場合は、具体的な数字・根拠を出して反駁する方法がよいようだ。たとえば、「現状最高性能の雑音抑圧技術を用いても、ロボカップ@ホームの環境では信号対雑音比 (SN比) は0程度であり、話者-マイク間の距離を50cmから100cmに伸ばすことでほとんどの音声認識率は失敗するだろう。そのような改訂は非現実的である」など、具体的な反例が効果的である。掲示板における議論では、焦らず文章と具定例を推敲すればよい。実際に、筆者は1つのコメントに数時間かけて推敲することもある。

タスク環境

図-2に2011年イスタンブール大会のフィールドを示す。図において、濃色の領域は家具を示す。フィールドは、複数の部屋からなる。2011年イスタンブール大会のフィールドは、リビングルーム、キッチン、ランドリー、ベッドルーム、廊下の5部屋から構成されている。部屋同士が壁で明確に仕切られていない場合もあり、その場合は境界線が定義される。

タスクで用いられる家具・オブジェクト・名称は準備日に公開される。以下に、準備日に公表される



図-3 2011年イスタンブール大会で用いられたオブジェクト

項目を示す。

○：標準オブジェクト集合

ペットボトルや菓子など、物体認識や把持のために用いられるオブジェクトの候補集合である。ただし、後述するように、○に含まれているがタスクで用いられないオブジェクトは存在する。図-3に2011年イスタンブール大会で用いられたオブジェクトを示す。オブジェクトが公表されたのち、同じものを食品店で購入し、手元に置いて把持のテストを行うチームも見られる。

○_s：チーム固有オブジェクト集合

チームが○から10個選択する。タスク内で把持が求められる場合、オブジェクトは○_sから選択される。ただし、○_sからどのオブジェクトが選択されるかは、レフェリーに委ねられている。これは、特定のオブジェクトに特化した手法が高得点を上げることがを防ぐためである。

○と○_sの違いは、タスク中に把持する必要があるかどうかである。把持するオブジェクトは2009年グラーツ大会までは○から選択されていたが、ハンドが小さいなどの理由で、明らかに把持できないようなオブジェクトが選択される場合があった。そのため、2010年シンガポール大会からはチームが○_sを選択できるようになった。

チームが○_sを選択する際には、大きさに注意する必要がある。つまり、大きいオブジェクトの方が画像認識に有利であるものの、把持する際には大き

すぎるオブジェクトは不利である。また、ハンドの形状により把持が難しいオブジェクトもある（本など）。

F：家具集合

家具は、1部屋に数個置かれており、実際の家庭を再現している。Fには、ダイニングテーブル、ソファ、椅子などが含まれる。椅子には人が座っている場合があり、机にはオブジェクトが載っている場合がある。一部のタスクでは、未知環境を模擬するために、家具の位置がタスク直前に変更される。そのため、オンラインSLAM (Simultaneous Localization and Mapping) 機能を有することが望ましい。

N_o：オブジェクト名集合

オブジェクト名は、1)「コーラを取ってきて」など、ユーザがロボットに指示を与えるため、2)「コーラを見つけました」など、ロボットが認識結果をレフェリーに伝えるために用いられる。○とN_oは1体1対応する。N_oは準備日に公表されるため、名称を音声認識辞書に人手で登録することが可能である。

N_f：家具名集合

N_fの目的は、N_oと同様である。FとN_fは1体1対応する。

N_h：人名集合

N_hは、Who Is Whoタスクなど、顔と名前を記憶するタスクに用いられる。理想的には、どのような名前でも音声認識できることが望ましいが、あらかじめ音声認識辞書にすべての人名を登録しておくことは現実的でない。そのため、準備日にN_hが公表される。人名が必要な場合、このリストからランダムに選択される。2011年イスタンブール大会では、N_hのリストは、マイケル、ソフィアなどの一般的な英語名20人分が用いられた。

2008年蘇州大会まで、Who Is Whoタスクに観客からランダムに選ばれた人が参加する機会があった。しかし、その観客の名前が英語名でない場合、音声認識辞書に名前が登録されている可能性は低いため、タスクがまったく進まなくなってしまう。こ



図-4 Go Get It! タスクにおいてオブジェクトを探索するロボット

のような事態を避けるため、 N_h が用意されるようになった。

タスク例：Go Get It!

本節では、Go Get It! タスクを例として、スコアを説明する。Go Get It! タスクは、フィールドにランダムに置かれたオブジェクトを見つけて、持ってくるタスクである (図-4)。本タスクの主眼は、物体探索、物体認識、移動、把持などの機能である。本タスクは、ロボティクスで近年注目されている、移動マニピュレーションを拡張した課題と捉えることができる。移動マニピュレーションとの主な違いは、オブジェクト名の音声認識を含むことである。

フィールドの準備は以下のように行われる。まず、レフェリーは、 O_s から 4 個のオブジェクトを選択し、フィールド内の床と棚に 2 個ずつ配置する。ダミーのオブジェクトとして、 $O \setminus O_s$ (O から O_s を除いた集合) から 4 個のオブジェクトが同様に配置される。ロボットには、 O_s のうち 1 個を見つけ、把持してユーザに手渡すことが求められる。

Go Get It! タスクでは、2 チームが同時にスタートする。ロボットはスタートの合図があるまでフィールド外で待機する。ロボットの待機位置は別々のドアの前であり、チームごとにあらかじめ指定される。レフェリーによりドアが同時に開かれることをスタートの合図として、タスクが開始される。不正 (オブジェクトの位置を手で入力する、など) がな

いように、待機位置に移動後は、チームメンバがロボットに触れることはできない。そのため、ロボットはドアが開いたことを検知し、フィールドに入場しなければならない。

ロボットはフィールドに入場後、オブジェクトの置かれた部屋へ行くよう、レフェリーから音声指示 (「ダイニングルームに行って」など) を与えられる。正しい部屋に到着できれば 200 点が与えられる。

ロボットは、正しい部屋に移動したのち、 O_s を探索する。正しいオブジェクトを発見できれば 200 点が与えられる。そのオブジェクトを把持できれば、さらに 200 点が加算される。このとき、オブジェクトを 5cm 以上持ち上げて、10 秒間保持できれば把持の成功とみなされる。

次に、ロボットは、そのオブジェクトをレフェリー (指示を与えた人物) に渡すために入り口に移動する。オブジェクトをレフェリーに渡すことができれば、300 点が与えられる。最後に、フィールドから退出できれば、100 点が与えられる。タスクの制限時間は 5 分である。

タスク開始後にチームメンバがキーボード、タッチパネル等を使ってロボットを直接操作することはできない。また、特別な場合を除き、無線による通信も許可されていない。ロボットへの指示は、音声やジェスチャなどロボットに直接触れないようなインタラクションのみが許可されている。これは、チームメンバが直接操作するような不正を防ぐためである。チームがオブジェクト位置をあらかじめ登録できないように、オブジェクトはスタート直前にレフェリーが配置する。

ハードウェア

eR@sers が用いたハードウェア

ロボカップ@ホームでは、各チーム 2 台までのロボットをエントリーすることができる。チーム eR@sers は、電気通信大学が中心になって開発したロボット DiGORO (図-5) と、玉川大学が中心になって開発した eR@ser の 2 台で出場した。両者は、

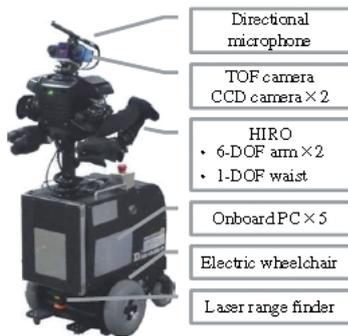


図-5 2010年優勝のチーム eR@sers のロボット DiGORO

NICT が開発を担当した共通の音声モジュールを用いている。

図-5 に示すロボットのハードウェア構成は以下の通りである。

- 指向性マイク (三研マイクロホン製 CS-3e)
- スピーカ (Yamaha 製 NX-U10)
- Imaging Source 製 CCD カメラ
- 赤外線 TOF 方式 3 次元距離測定カメラ (MESA Imaging 製 SR-4000)
- 上体ヒューマノイド (川田工業製 HIRO)
- Okatech 製電動車椅子用台車
- レーザレンジファインダ (北陽電機製 UTM-30LX)

以下では、地図作成、画像処理、音声処理の面から、DiGORO のハードウェアについて概説する。なお、DiGORO の詳細については、文献 4) を参照されたい。

DiGORO は、レーザレンジファインダ (LRF) を用いて地図作成を行っている。LRF には北陽電機製の UTM-30LX を採用した。移動機能のため、Okatech 製台車を用いている。カメラやマイクは、川田工業製上体ヒューマノイド HIRO に搭載されている。HIRO は、C++ や Python などの言語に対応しており開発効率が高い。

画像処理用には、Imaging Source 製 CCD カメラ 2 台を用いている。さらに、MESA Imaging 製赤外線 TOF 方式 3 次元距離測定カメラ SR-4000 から得られた情報を統合して、RGB-D データを得ている。SR-4000 は、前面の LED より強度変調して発

光された赤外線が対象物体で反射して戻ってくるまでの時間を計測することで、物体までの距離を取得する。計測範囲は 0.3 ~ 5m 程度で、精度は 10mm 程度である。

音声認識の精度を高めるためには、マイクの選定に注意が必要である。我々は、指向性マイクを用いることでユーザ方向以外の音声を除去する方法を採用した。さらに、後述するようにソフトウェア的にも雑音抑圧を行っている。マイクとしては、三研マイクロホン製 CS-3e を用いている。CS-3e は良好な指向性を持ち、我々が用いた音響モデルとの相性もよいという特徴を持つ。

音声波形をデジタル化して音声認識器に入力するための AD コンバータとしては、M-AUDIO 製 Mobile Pre USB を用いている。Mobile Pre USB は USB バスパワーのため、追加で電源を供給しなくてよいという利点がある。標準的な音声認識器は 16kHz でサンプリングした音声を前提としているので、他の AD コンバータを用いる場合は確認が必要である。

高騒音環境下でもロボットの発話が聞き取れるように、高出力なスピーカを用いることは重要である。たとえば、ロボットが正しい物体を発見したとしても、「見つけました」という発話がレフェリーに伝わらなければ点数が与えられないこともあり得る。我々は、Yamaha 製 NX-U10 を採用した。NX-U10 は、USB バスパワーで動作するスピーカとして十分な出力を有している。

他チームのハードウェア

本節では、ロボカップ@ホームに出場しているチーム、特に上位チームの用いている代表的なデバイスについて概説する。表-3 に上位チームが用いている代表的なデバイスを示す。

移動手段としては、台車移動が一般的である。2011 年イスタンブール大会では、19 チーム中 18 チームが台車移動のロボットを用いていた。代表的な台車としては、Adept Technology (MobileRobots) 製 Pioneer シリーズ、Fraunhofer IAIS 製 Volksbot、



	デバイス名
台車	Adept Technology製 Pioneer3-AT, Fraunhofer IAI製 Volksbot, Segway製 RMP-100
LRF	SICK製 S300, 北陽電機製 URG-04LX
アーム	Neuronic製 katana 6M180, Exact Dynamics製 iARM
カメラ	PointGrey製 Flea2, MESA Imaging製 SR-4000, Microsoft製 Kinect
マイク	三研マイクロホン製 CS-3e

表-3 代表的なデバイス

Segway 製 RMP-100 などがある。RMP-100 は 2 輪のみで走行可能であるが、スイッチが押された場合に停止させるというルールがあるため、キャストを取り付けるなど転倒予防策が必要である。

LRF の数については、上位チームとそれ以外のチームの間に大きな違いがあり、上位チームは複数の LRF を持つことが多い。2011 年イスタンブール大会優勝のチーム Nimbrow のロボットには、地図作成用に 1 台 (SICK 製 S300)、前後の障害物回避用に 2 台 (北陽電機製 URG-04LX)、3 次元環境認識用に 1 台 (北陽電機製 UTM-30LX) の LRF が搭載されていた。

アームやグリップを持たないロボットの出場は可能であるが、スコア上有利ではない。上位 5 チームのうちアームやグリップを持たないロボットは、2008 年 3 台、2009 年 1 台、2010 年 0 台であった。2011 年優勝の Nimbrow は、Dynamixel 製モータによる自作アームを用いている。一方、既製アームとしては、Neuronic 製の katana 6M180 を使用するチームが多い。他には Exact Dynamics 製 iARM などを用いられている。上位チームのロボットの可搬重量は、500g から 1000g 程度である。

Aldebaran Robotics 製の NAO を用いたチームもあったが、物体の把持に苦労していたようである。床に近い物体しか把持することができず、把持する場合も両腕で届く範囲がきわめて短い。

音声入力デバイスとしては、指向性マイクを用いるチームが多い。スピーカに関しては、研究室ではよく聞こえる音量であっても、高騒音環境ではまったく聞き取れないこともあり得るので注意が必要である。実際に、ほとんどのチームは低出力なスピー

カを用いており、改善が必要だと考えられる。なお、ルール上は会場スピーカと接続するためのワイヤレストランスミッタが用意されることになっている。しかし、2009 年グラーツ大会を除き、正しい仕様の機器が用意されたことはないで、高出力のスピーカを用意の方が無難である。この点に関しては、運営側に改善を求めていると考えている。

サイズに対する制約

ロボットのサイズは、幅 800mm 高さ 2000mm のドアを通り抜けられるサイズでなければならない。自律移動によりドアを通り抜ける機能は、ほとんどのタスクで必要になる重要な機能であるものの、これに失敗するロボットを大会でよく見かける。設計の際には、以下の点を考慮することが重要である。

- 必ずしも正面からドアに接近できるとは限らない。
- ドアを開けた場合、有効幅は 750mm 程度である。幅 650mm のロボットであれば、片側の余裕は 50mm 以下である。
- ドア付近に段差があることが多い。ルール上は 25mm までの段差は許容される。

ロボカップ@ホームでは、重量が大きすぎるロボットの出場は許可されていない。これは、家庭やオフィスでの応用を目指しているためである。最大重量は、大会が開催される地域における、住宅またはオフィスの建築に関する法律に従うことになっている。たとえば日本の場合、建築基準法において住宅の居室の積載荷重の制限 (1800N/m^2) がある。そのため、180kg を超えるロボットは出場できない。2011 年イスタンブール大会優勝のチーム Nimbrow のロボット Cosero (図-6) は、全体で 32kg と非常に軽量である。

アームの設置に関しては、オブジェクトが置かれる位置を考慮することが重要である。オブジェクトは、床、棚、机 (70cm 程度)、調理台 (85cm 程度) などに置かれることが多い。これらの高さに置かれたオブジェクトを把持できるように、アームを配置することが望ましい。ただし、アームを設置すること



図-6 ドアを通り抜ける
ロボット Cosero



図-7 Who Is Who タスクにおける人検出

と台車のサイズを小さくすることを、同時に満たす設計は簡単ではない。

ヒューマンロボットインタラクション

本章では、ヒューマンロボットインタラクションの観点から、ロボカップ@ホームで必要とされる機能について述べる。

人検出・顔認識

ロボカップ@ホームで高得点を取るためには、高精度の人検出・トラッキング・顔認識機能を持つことが望ましい。たとえば、Who Is Who タスクや Enhanced Who Is Who タスクでは、制限時間内に顔を記憶し、フィールド内にいる人物を見分ける必要がある(図-7 参照)。

ロボカップ@ホーム特有の難しさとして、フィールド外にいる観客の誤検出が挙げられる。誤検出を防ぐためには、顔までの距離を計測してフィールド内外を判定することが重要である。一方、ロボカップ@ホームでは認識対象の人物は高々5人であり、それほど多いわけではない。これは一部屋に入る人数の制限によるものである。

Follow Me や Shopping Mall タスクでは、動的環境で人を追跡する機能が求められる。これらのタスクでは、人検出のために LRF が用いられることが多い。上位チームは、LRF を複数設置し精度を高

めていることが多い。たとえば、足位置の LRF を候補検出のために用い、腰位置の LRF で信頼度の高い候補を選別している。腰位置の LRF の代わりにカメラや Kinect を用いるチームもある。

顔検出には、Haar-like 特徴量⁵⁾ が用いられることが多い。ただし、OpenCV で提供される関数を単純に使用するのみでは、高い精度を得ることは難しい。顔認識ソフトウェアとしては、他に VeriLook SDK などがある。我々のチームは OMRON OKAOVision を用い十分な精度を得ている。

音声対話

■ 音声対話機能の概要

図-8 に、ロボットの音声対話機能の一般的な構成を示す。まず、マイクから入力された音声は、音声認識部において単語列に変換される。次に、発話理解部において、より抽象度の高い表現にマッピングされる。たとえば、「ペットボトル持ってきて」や「ペットボトルを持ってきてください」などの発話は、同じ機能にマッピングすることが一般的である。さらに、応答生成部において、ロボットの応答が生成される。

音声認識部の前段処理として、雑音抑圧や音声区間検出 (Voice Activity Detection; VAD) がある。雑音抑圧とは、音声信号中の目的音声以外の信号 (たとえば音楽) を低減する機能である。また、VAD は音声信号から目的音声部分を切り出す機能である。

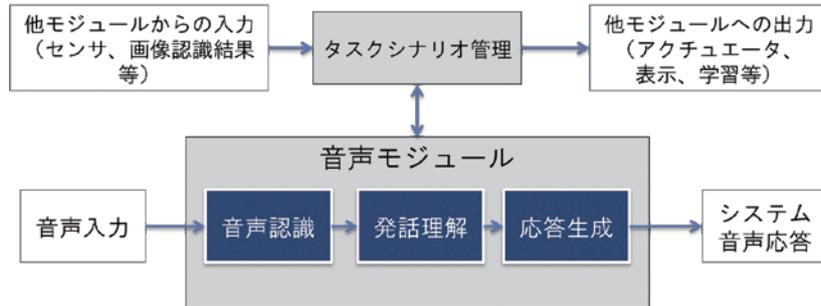


図-8 一般的な音声対話機能の構成

応答生成部では、ロボットの発話が生成される。ほとんどのチームはテキスト音声合成を用いている。物体名・人名リストは準備日に公開されるものの、辞書に登録されていない単語が用いられる場合がある。その場合は人手による登録作業が必要である。我々のチームは、2008年の蘇州大会において、未登録語を学習するデモンストレーションを行った。ロボットの声で未登録語を発声するために、Eigenvoice GMMに基づく声質変換技術を用いている⁴⁾。

■ 高騒音環境における音声認識

ロボカップ@ホーム環境は、音声認識技術にとって過酷な環境である。フィールドの等価騒音レベル (Leq) は約 75dB であり、最大騒音レベルは約 85dB (会場アナウンスなど) である。さらに、マイクと話者の距離が離れると、音声認識の精度が劇的に悪化することにも注意が必要である。たとえば、ヘッドセットを用いた場合に 90% 以上の認識精度が得られたとしても、ロボットに搭載したマイク (オンボードマイク) を用いた場合には 50% 以下になることも珍しくない。2011年イスタンブール大会で収録した騒音のデータベースが公開されているので、性能試験に利用されたい^{☆1}。

ロボカップ@ホームに特有の騒音としては、サッカーリーグの観客による歓声、音楽、アナウンスなどが挙げられる。特に、アナウンス音声とユーザの声の分離は、どちらも人の声であるため難しい。さらに、会場の高出力スピーカから再生されたロボット自身の発話を、ユーザ発話として誤検出するケ

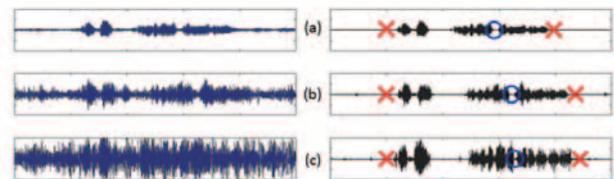


図-9 雑音抑圧前(左列)と雑音抑圧後(右列)の音声波形。(a) 60dBA, (b) 70dBA, (c) 80dBA を示す。

スも多く見受けられる。図-9に、ロボカップ@ホーム環境と同等の騒音環境における音声波形を示す。左列は雑音抑圧前、右列は我々が用いた手法⁶⁾による雑音抑圧後の波形を表す。騒音が大きい場合、発話の開始・終了を見つけることが非常に難しいことが分かる。

ヒューマノイドではモータ・ファンから生じる騒音が問題であるが、ロボカップ@ホーム環境では状況が異なる。ロボカップ@ホームでは、周囲の騒音への対策が最も重要である。これは、前述のようにファン音に比べて非常に大きい騒音の存在と、台車移動のロボットが多い事情による。また、残響に対する対処を行っているチームは現状では少ないと推定される。

音声入力デバイスとして、ヘッドセットを用いることも許可されている。リーグ開始初期には上位チームであってもヘッドセットを用いるチームがあったものの、2010年には数チームを除きほとんどのチームがロボットに搭載したマイクを用いていた。これは、オンボードマイクによる音声認識を促進するために、ボーナス点が与えられるようになったためである。ボーナス点は、マイクと話者の距離が 50cm 以上の場合に与えられる。2008年には、

☆1 http://mastarpj.nict.go.jp/~ksugiura/software/software_jp.html

上位チームであっても苦労するチームがあったが、2010年にはほとんどの上位チームが問題なく距離の問題に対処していた。

我々のチームで採用した騒音への主な対策は、以下の通りである

1. 指向性マイクロホンの利用

マイクを話者に向け、話者の方向から到来する音声のみを入力とする。他のチームでは、マイクロホンアレーの利用も行われている。

2. 非定常な雑音の推定と抑圧

パーティクルフィルタとスイッチングダイナミカルシステムにより、雑音を逐次推定し、雑音抑圧を行う⁶⁾。スペクトルサブトラクションにより雑音抑圧を試みるチームもある。

3. ロボット名による呼びかけ

「イレイサー、ペットボトルを取ってきて」のように、音声命令にロボット名が含まれるようにする。つまり、ロボット名のキーワードスポットティングにより命令を受理するかどうかを決定する。

ロボカップ@ホームタスクのうち、話者がチームメンバであるタスクでは、さまざまな言い回しに対処する必要はない。たとえば、そのようなタスクでは、「私の名前は～です」と発話すると決めておけばよい。そのため、音声認識器の言語モデルでは、記述文法に基づくものが広く用いられている。一方、チームメンバ以外の話者との対話を必要とするタスクに対して、統計的言語モデルを用いた連続音声認識を用いるチームもあった。しかし、現状では話者に「～と指示してください」のように言い回しを指定するチームがほとんどである。

■ 対話履歴の管理

図-8のタスクシナリオ管理部は、行動・対話履歴や時間管理を行う。多くのチームでは、ロボットの音声対話はタスクシナリオ管理部により制御されている。たとえば、「顔を検出した後でロボットが挨拶する」などの状態管理が行われている。

我々のチームを含め、音声認識精度を向上させるため、対話履歴を利用するチームも多い。以下の例

を考える。

ユーザ： 「ペットボトルを持ってきて」

ロボット： 「ペットボトルを持ってきます。いいですか？」

ユーザ： 「はい/いいえ」

このとき、ロボットの発話後に「はい」「いいえ」の2単語のみを受け付けるような認識文法に切り替えることにより、認識精度を向上させることが可能である。

上記のような問題は、マルチモーダル音声対話システムの分野においても広く研究されてきた。ただし、ロボットの対話システムでは、入出力の多様性や処理の非直列性を扱う点が技術的に難しい。現状では、ロボットにおける音声対話を扱う研究は端緒についたばかりであるが、2010年にはAAAI Fall Symposiumにおいて“Dialog with Robots”をテーマとしたシンポジウムが開かれるなど、今後の発展が期待される。

■ 音声認識・合成ソフトウェア

上位チームが用いている音声認識ソフトウェアは、CMU Sphinx^{☆2}、HTK^{☆3}、CSLU Toolkit^{☆4}、Microsoft Speech APIなどが挙げられる。音声合成ソフトウェアとしては、Festival^{☆5}、Loquendo TTS、MAC OS X Speech Synthesis APIなどが用いられている。我々のチームでは、NICTで開発されているATRASRおよびXIMERAを、それぞれ音声認識および音声合成のために用いている。

ジェスチャ認識

Follow Meなどのタスクでは、ユーザからの命令は、音声またはジェスチャのいずれかで与えられることになっている。ジェスチャ命令と動きの対応は、各チームが定義できる。この理由は、ジェスチャの意味が文化依存であるためである。有名な例では、日本での手招きジェスチャが、「あっちへ行け」という意味と解釈される文化もある。チームが定義した

☆2 <http://cmusphinx.sourceforge.net/>

☆3 <http://htk.eng.cam.ac.uk/>

☆4 <http://www.cslu.ogi.edu/toolkit/>

☆5 <http://www.cstr.ed.ac.uk/projects/festival/>



図-10 ロボット頭部に設置した Kinect を用いたジェスチャの認識

ジェスチャを未知ユーザに伝えるために、タスク開始直前に準備時間が用意されている。

ジェスチャの認識では、色ベースの単純な手法を用いるチームが多い。まず、顔を検出したうえで同色の領域を手と推定し、手の軌道等を入力として認識を行う。認識器としては、サポートベクトルマシン (SVM)、naïve Bayes、隠れマルコフモデル (HMM) などの手法が用いられている。

2010年シンガポール大会におけるスコア取得率は、上位チームの平均が約60%であり、全チームの平均は約20%であった。ジェスチャ認識率が低いチームでは、未知ユーザに対する頑健性が低いことが多い。研究室でのテスト(性能試験)ではユーザが慣れてしまい問題が発見できないことが多いため、多様(年齢・性別・国籍など)なユーザ群を集めてテストを行うことが重要である。図-10に示すように、Kinect等のRGB-Dカメラを利用するチームも見られるようになり、今後の精度向上が期待される。

今後の展望

最後に、ロボカップ@ホームの展望についてまとめ、結言としたい。

本稿で述べたように、ロボカップ@ホームの主要課題は、移動マニピュレーションとヒューマンロボットインタラクションの2つである。

マニピュレーションに関しては、認識・把持機能

の高度化のため、物体数を徐々に増加させるルール改正が、TCで議論されている。また、一部のタスクにはドアを開ける課題が含まれているが、現時点でどのチームも成功していない。筆者としては、ボーナス点や関連する課題数を増やすなど、ドアや引き出しを開ける機能の開発を促進したいと考えている。

ヒューマンロボットインタラクションに関しては、言語処理と画像処理による実世界情報処理の重点化が進められている。たとえば、2010年より開始されたGeneral Purpose Service Robot (GPSR) タスクは、2011年には1stステージと2ndステージに分割され、事実上配点が加算された。GPSR I タスクでは、ロボットは、「キッチンテーブルまで行って、シーフードヌードルを取ってからフィールドから出て」などランダムに生成された長文の命令を実行する。タスクで高得点を取るためには、頑健な発話理解と多彩な機能を実行できることが必要であるが、2011年現在ほとんどのチームがスコアを取ることができていない。

日本発のRTミドルウェアや、Willow Garageが提唱するROS (Robot Operating System) などの登場により、ロボット開発に必要な開発コストは、急速に減少していると言えるだろう。ロボカップ@ホームでは、特にROSを統合プラットフォームとして採用するチームが増えている。ROSはオープンソースのロボットライブラリであり、シミュレータ、経路計画、画像処理、通信など、豊富なライブラリに特徴がある。

ロボカップ@ホームでは、個々の研究分野においても困難な問題に取り組むと同時に、一般環境で競技を行うなど、応用をイメージしやすい課題を設定している。冒頭で述べたように、ジャパンオープンの参加チームは少ない状況が続いていたが、ロボカップ@ホームキャンプなど情報共有の試みもあり、2011年には参加チームが6チームに増加した(図-11)。これまでの経験から、ロボカップ@ホームコミュニティは、新しい技術やシステムの導入スピードが速いと感じている。新規に参加される方に



図-11 ロボカップジャパンオープン 2011 の集合写真

は、情報交換や教育の場としても有効に活用していただけるのではないかと考えている。

参考文献

- 1) RoboCup@Home Rulebook,
<http://www.ai.rug.nl/robocupathome/>
- 2) 岡田浩之, 大森隆司: ロボカップ@ホーム一人とロボットの共存を目指して一, 人工知能学会誌, Vol.25, No.2 (2010).
- 3) Iocchi, L. and van der Zant, T.: Robocup@Home: Adaptive Benchmarking of Robot Bodies and Minds, Proceedings of the International Conference on Simulation, Modeling and Programming for Autonomous Robots, pp.171-182 (2010).
- 4) Nakamura, T., Sugiura, K., Nagai, T., Iwahashi, N., Toda, T., Okada, H. and Omori, T.: Learning Novel Objects for Extended Mobile Manipulation, Journal of Intelligent and Robotic Systems, pp.1-18 (2011).
- 5) Viola, P. and Jones, M.: Rapid Object Detection Using a Boosted Cascade of Simple Features, in Proceedings of CVPR (2001).

- 6) Fujimoto, M. and Nakamura, S.: Sequential Non-Stationary Noise Tracking Using Particle Filtering with Switching Dynamical System, in Proceedings of ICASSP, pp.769-772 (2006).

(2011年12月28日受付)

■杉浦孔明 komei.sugiura@nict.go.jp

2007年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。日本学術振興会特別研究員, ATR 音声言語コミュニケーション研究所研究員を経て, 2009年より情報通信研究機構専攻研究員。ロボット対話の研究に従事。

謝辞 岡田浩之教授, 長井隆行准教授, 大森隆司教授, 岩橋直人専攻研究員をはじめとするチーム eR@sers のメンバに感謝の意を表す。紙面の都合から全員の名前を挙げられないことをご容赦いただきたい。ジャパンオープンでの@ホームリーグ立ち上げに尽力いただいた九工大の大橋健准教授に感謝する。