

音声によるロボットへの作業指示とビジュアルフィードバック

Task-Level Voice Commands for Robots and Visual Feedback

尾崎 文夫
OZAKI Fumio

大明 準治
OAKI Junji

辰野 恭市
TATSUNO Kyoichi

ロボットが家庭や公共の場で人と共存して作業をするとき、「お茶を入れて」などの音声による作業指示を理解し、周囲環境を認識して作業を実行する必要がある。ここでは、作業を指示しただけでロボットが指示された作業を実行するのに必要な、タスクレベルのロボット言語と外界センサフィードバックについて述べる。一例として、実際のロボットが「赤いボールを取って」という音声指令を受け、赤いボールを画像認識により探し出した後、ボールの位置をリアルタイムで計測しながらビジュアルフィードバックを行ってボールに接近し、床の上のボールを拾う実験例を示す。

Robots need to be able to understand orders from humans when they work in homes and public places. Such orders are high-level commands such as "I'd like a cup of tea" or "bring me the newspaper," in which the detailed motion of the robot is not described. Robots must therefore analyze the human high-level command and break it down into primitive motion commands such as "move the arm tip to point A." It is also necessary for the robot to recognize and/or measure the concerned object of the task using sensors so as to accomplish the high-level command.

This paper describes an experiment, in which a robot picks up a ball on the floor upon hearing the human order "pick up the ball," to verify the validity of the high-level commands.

1 まえがき

ロボットは工場の中で広く活躍しているが、家庭や公共の場のような所での活躍はあまり聞かれない。工場では活躍できて、一般の場で活躍できない理由は何であろうか。

ここでは、まず、その原因と問題点を考察する。具体的には、タスクレベルのロボット言語の必要性和、ロボットが周囲環境を認識するためのセンサとフィードバックシステムに要求される項目を考え、問題解決のための答えとして開発したロボットシステムについて述べる。

2 ロボットが家庭や公共の場で活躍できない理由

ロボットが家庭や公共の場で活躍できていない理由には、次の二つが考えられる。

第1は、ロボットが作業をする対象物の位置を事前知っているかどうかである。工場では活躍するロボットはあらかじめ部品とロボットの位置関係(部品がどこにおいてあるか)を知っており、その位置関係を基にロボットの動作を事前に決めて、単にその動作を繰り返すことにより、例えば溶接作業を実現する。しかしながら、家庭や公共の場で人の話し相手をしたり、荷物を運んだりするロボットは相手や荷物の位置を事前には知ることはできず、その場で相手や荷物の位置を計測して知る必要がある。よって、一般の場で働くロボットには相手を認識し、その位置を知る機能が必須と

なるが、現状では多種多様な物の認識が難しく実用に至っていない。

第2は、ロボット動作の指令方法の違いである。工場では活躍するロボットはその動作を逐一動作レベルのロボット言語と呼ばれるロボット用のプログラム言語で指示する。動作レベルのロボット言語は、「ロボットアームの手先の位置をどこそこへ動かささい」というような命令であり、これを組み合わせることでロボットに複雑な動きをさせている。これに対して、一般の場で働くロボットの場合には、人間に対するのと同じように、行わせたい作業を指示するだけでロボットが自動的に作業を実現することが望まれる。例えば、高齢者が「お茶を入れて」、「新聞を持って来て」とロボットに音声で指示すると、ロボットがお茶を入れたり、新聞を持って来てくれることが望まれる(こちらは動作レベルに対してタスクレベルのロボット言語と呼ばれている)。これを実現するためには、指示を理解しロボットの動作に分解してロボットを制御する技術、お茶や新聞をロボットの見た映像の中から認識する技術などが必要となってくる。これらの技術もいまだ研究レベルである。

このように、現状のロボットは、①タスクレベルのロボット言語が構築できていない、②かりにタスクを与えられても、周囲の多種多様な物や相手を認識し、自律的に動作するためのセンサフィードバックが基礎研究の段階である、ために一般の場での活躍が少ないものになっていると考える。

しかしながら、これら二つを解決するために必要な技術

は、今後ロボットがこのような場で働くために必須の技術である。ここで紹介するロボットは、ボールという限定した対象物についてではあるが、作業指示を認識し、画像による位置計測を用いてボールの位置にロボットを誘導し、ボールを拾い上げるといった作業を行った数少ない例である。

3 タスクレベルのロボット言語と外界センサフィードバック

現状のロボットは、そのほとんどが動作レベルのロボット言語により動作している。作業を実現するためには、この動作レベルのロボット言語を数十から数百組み合わせなければならない。工場などで行う作業が事前に決まっており、動作レベル言語を用いてプログラムを事前に作ることができる場合には使えるが、一般の場では作業がその場その場で決まり、また、利用者が複雑なプログラムを書くのも難しい。そのため、利用者が作業を指示して、あとはロボットが自動的に作業を実行できるような動作レベル言語に作業を分解して、それをロボットが実行することが望ましい。一般の場でロボットが活躍するためには、このようなタスクレベルのロボット言語が必要である。

次に外界センサフィードバックの必要性について述べる。一般の場での作業では、工場での作業とは違い刻々ロボットの周りの環境が変化している。そのため、タスクレベル言語では、変化する環境内での作業を実現するために周囲環境を計測する外界センサが必須となる。ロボットを環境に適応させつつ作業を行わせるためには、外界センサの情報をフィードバックしてロボットを動かすことが必要なのである。例えば、お茶を入れる場合には、まずお茶を入れる湯飲みを探し出して、これを持ち、お茶を入れなければならない。そのためには、外界センサにより湯飲みの映像などを抽出し湯飲みを認識する技術、湯飲みの位置を計測する技術、センサ情報を用いながら湯飲みの位置へロボットを動かす技術が必要となる。このような外界センサ情報のフィードバックがあって初めて、タスクレベルのロボット言語を実現できるのである。

4 ロボットへの作業指示とその実行

ここでは、実験に用いたロボットシステムについて説明し、タスクレベル言語を理解して実行する一例として、「赤いボールを取れ」と音声で指示されると、赤いボールを拾う場合について述べる。この作業指示がどのように分解されて、移動機構、複数のアームの動作命令プログラムになっていくのか、作業計画の一例を紹介する。また、このとき重要な役割を演じるビジュアルフィードバック、すなわち視覚により作業対象物とロボットの相対位置を計測しながら作業対象物と

ロボットの位置合せをする方法の一例を紹介する。すなわち、音声による指示で具体的な作業を、ビジュアルフィードバックを用いて実行した。

このロボットは、ボールを拾う、ボールを打ち合うなどのデモを行った⁽¹⁾。ボールを打つ動作ではボールをすばやく追いかけるためにボールの動きを予測して打点を計算している。そのため一度、視覚情報を物理法則(重力や空気抵抗)が考えられる座標系での情報に変換して、ボールの位置を計算し、ボールを打った。これを実現するためにはボールとの相対位置でなく、設定した座標系に対するロボットの位置関係を計測するか、あるいは座標系に対してロボットが固定しているというような絶対位置関係が必要になる。一方、ボールを拾う作業では、ボールの位置とロボットの位置を視覚情報の中で合わせるようにした。よって、ロボットとボールの相対位置関係だけが問題になる。家庭や公共の場で活躍するロボットでは絶対位置関係を保つことが難しいので、相対位置関係に頼った作業戦略をとる必要がある。ここでは、そのような例としてボール拾いの作業を説明する。

4.1 ロボットの概要

ロボット⁽¹⁾がボールを拾うようすを図1に示す。ロボットは、Gアームと呼ばれる大きな振り子状の移動機構に支えられ、頭部に視覚用のCCD(電荷結合素子)カメラを備え、Aアームと呼んでいる、人の腕とほぼ同じ大きさの左右のAアームを持つ。視覚(二つのCCDカメラによるステレオ視)によりボールの位置を測定しながら、Gアームでロボットをボール近傍まで近づけ、Aアームで地上のコートにある赤、青、緑のボールから指定された色のボールを拾う。

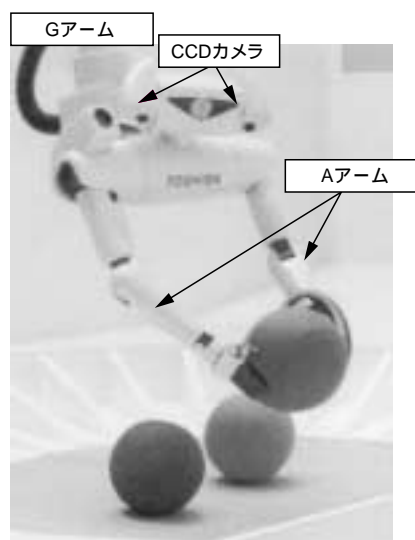


図1 . ロボットがボールを拾うようす 視覚によりボールの位置を測定しながら、移動機構(Gアーム)でボール近傍まで近づき、人腕型Aアームでボールをつかむ。
Robot picking up ball

ロボットコントローラの機能構成を図2に示す。全体管理、動作管理、音声認識、左右のAアーム制御系、Gアーム制御系、雲台制御系、ボール位置計測系のサブシステムから成る。各サブシステムへの動作レベルのロボット言語は、例えば手先の移動のためのMove命令は、左右AアームにはLarm.Move(…)やRarm.Move(…),GアームにはGarm.Move(…)というような形になる。

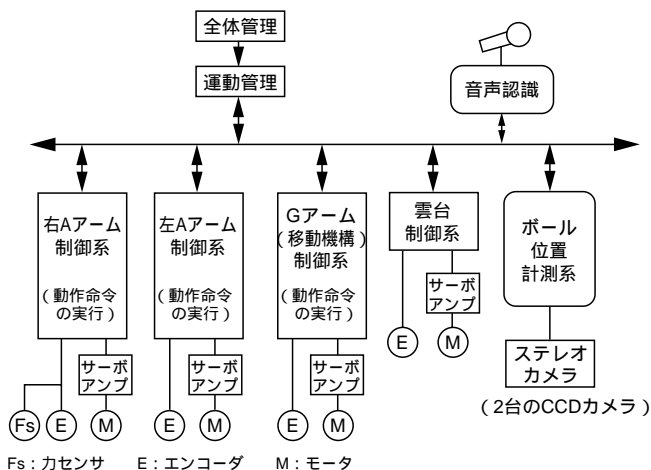


図2. ロボットコントローラの機能構成 全体管理,各サブシステムに動作指示を送る運動管理と,左右の人腕アーム制御系,移動用アーム制御系,ボール位置計測系などのサブシステムから成る。

Functional structure of robot controller

4.2 ボールを拾うときのシステムの動作

作業レベルの言語の一例として、「ボールを拾う」という作業をするときのロボットシステムの動作を以下に説明する。「ボールを拾う」という作業は次のような、より小さい基本作業に分割される。

- (1) 作業指示を入力する(音声認識 + 作業理解)。
- (2) ボールを探す。
- (3) ボールを拾う位置へ移動する。
- (4) ボールをつかむ。

この分解された基本作業を実行するように、全体管理から音声認識、運動管理に指示が出される。

運動管理は、サブシステム(音声認識,各アーム制御系,雲台制御系,ボール位置計測系)に、動作レベルロボット言語の起動命令を出して、上記(1)~(4)の基本作業を実行する。

次に、各基本作業でのサブシステム動作について述べる。

4.2.1 作業指示入力 音声認識(サブシステム)

音声入力待ち状態にし、「赤いボールを取れ」と人が音声を入力(作業指示(タスクレベルの言語での指示))すると、音声認識する。ここで用いたHidden Markov Model(HMM)による音韻認識に基づく音声認識の方法を簡単に紹介しておく。

計算機には、あらかじめ音韻モデル(声紋がA,K,A,Iなどの音韻のいずれに近いかを比較するためのモデル)、認識対象となる単語、及び「赤い、青い、ボール、取れ、ください、…」とどの単語とどの単語がつながるかという文法、すなわち「赤い→ボール、×赤い→取れ、…」というルールを登録しておく。認識したい文全体の声紋を登録するのではない。

一定時間(8ms)ごとに入力音声进行分析し、音声をスペクトルパターン、いわゆる“声紋”に変換する。音韻モデルと声紋を照合し、類似度により認識のための判定スコアを付ける。このスコアに基づいて、単語のスコア(例えば、「赤い/AKAI/」という単語の場合は、/A/と/K/と/A/と/I/のスコアの和)を順次求めていく。同時に、どの単語とどの単語が接続するかを、あらかじめ登録してある文法ルールに従って判断し、単語と単語のつながりによって構成される文の候補を作成する。候補に挙がった各文に対して、文を構成する単語の和と処理中の単語の部分が獲得しているスコアを加えていき、音声が入力されなくなったときに最大スコアとなった文の候補、例えば「赤いボールを取れ」と言った場合は、/A/K/A/I/-/B/O/R/U/-/W/O/-/T/O/R/E/という音韻列の文が最大スコアを獲得し、「赤いボールを取れ」と認識する。登録されているのは文ではなく単語であるので、少し異なった表現「赤いボールを取ってください」や「赤いボール」なども同様の意味のことを言ったと認識できる。

4.2.2 ボールを探す 運動管理は、ボールを探すための初期位置にロボットを移動する指令を出す。各アーム制御系に初期位置へ関節角座標で移動するロボット言語命令Larm.Move(L_0),Rarm.Move(R_0),Garm.Move(G_0)を出し、ロボット全体として初期位置・姿勢をとるようにする。

次にボールを探す動作を雲台制御系に指示する。カメラの視野がコート上を走査するように、雲台を振ってボールを探す。走査のための雲台の動作はロボット言語でプログラミングしておく。

一方、ボール位置計測系には、赤いボールを認識するように指示する。ボール位置計測系は色(赤)と赤い一塊の領域の形状・大きさでボールを認識し、ボールがあるかどうか判断する。

一度ボールがカメラの視野に入ったと認識すると、ボール位置計測系はボールの三次元位置を計測し続け、雲台制御系は、計測したボール位置を用いて、ボールの映像の中心が左カメラの画面の中心にくるように雲台を制御しながらボールをトラッキングする。以後、ボール位置計測系、雲台制御系はボールをトラッキングしながら三次元位置を計測し続ける。

4.2.3 ボールを拾う位置へ移動 運動管理は、ビジュアルフィードバックでボールを拾う位置まで移動するように、Gアーム制御系に指令する。

Gアームは、以下のようにビジュアルフィードバックで移動

する。図3に示すように、ボールを拾う位置はボールに設定した座標系 b で表示した P_{pick} であるとする。カメラ基準座標系から見ると、 X_r の位置にボールがきたときである。図3のように、カメラから見たボールの位置を X (現在値)、拾うための位置(目標値)を X_r とする。ボールの現在値 X は、ボール位置計測系で $1/60$ s に1回計測され、ボール打合い管理に格納されているものである。ロボットの姿勢は固定で、あらかじめ決めた拾うための姿勢を保つ。

まず、現在値 X から目標値 X_r までの直線軌道を生成し、この軌道上の微小ステップを X とする。すなわち、 $X = (X - X_r)/N$ (N :分割数)。 X は、ロボットを移動させるGアームの関節角(第1,2,3関節)をとすると、

$$X = J \cdot \theta \quad (1)$$

の関係がある。ここで J は、 θ と X との関係を示すヤコビアンで、設計値である。したがって (1) 式の逆、

$$\theta = J^{-1} X \quad (2)$$

を用いて、 2 ms ごとに θ を計算し、Gアーム制御系移動命令 $\text{Garm.Move}(\theta)$ を送る。 $1/60 \text{ s}$ ごとに直線軌道を修正しながら、移動命令 $\text{Garm.Move}(\theta)$ を送り続ける。このようにして、ロボットがボールを拾う位置 X_r すなわち P_{pick} にくるように制御する。用いたヤコビアン J は設計値であるが、実際のヤコビアン J と少し違っていても、漸近的に目標値 X_r になるに収束する。

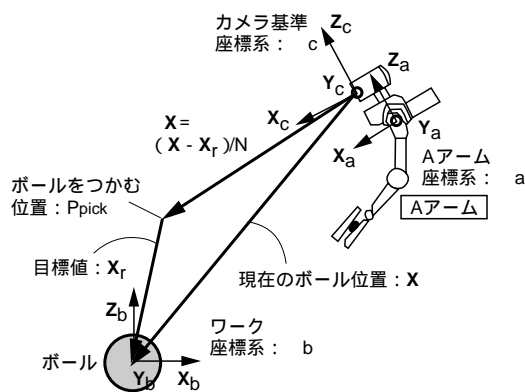


図3 . ビジュアルフィードバックの座標系 ボールの位置を CCD カメラで測定し、目標位置 P_{pick} に移動する。
Coordinates for visual feedback

4.2.4 ボールをつかむ 動作管理は、左右の A アーム制御系に、あらかじめ教示した挟む動作命令を順次送る。このときの動作は、A アーム座標系 a で教示されたものがあるが、ロボットの位置がボールを拾う位置にあるので、A

ーム先端は図4のようにボールに設定したワーク座標系で $\pm Y_{Re}$ まで移動してボールをつかむことになる。

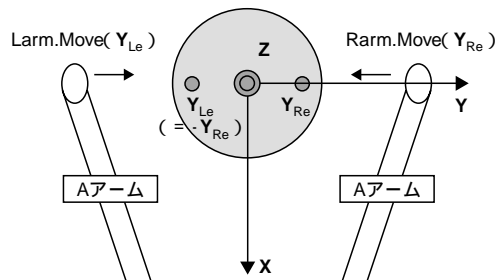


図4 . ワーク座標系で表したつかむ動作 ボールに設定したワーク座標系でつかむ作業を定義する。そうすると、ロボットに依存しない作業が定義できる。

Ball picking up motion expressed by work coordinates

5 おわりに

タスクレベルの音声言語指示によりボールを自動的に拾うロボットの例を示し、家庭や公共の場で働くロボットに必要な機能の一部を実際に示した。ロボットが作業をするためには、ロボットがするにせよ、人がするにせよ、最終的にはロボットの動作をプログラミングしなければならない。今後、ロボットがいろいろな作業をするためには、いろいろな作業対象物に対して、“把持”、“挿入”などの基本的な作業の作業プログラムを蓄積しなければならない。この作業プログラムの蓄積がロボット普及の最大のポイントでもある。

文献

- (1) 辰野恭市 “ビーチボールを打ち合うロボット”. 日本ロボット学会誌 .18 ,5 , 2000 , p.721 - 727 .



尾崎 文夫 OZAKI Fumio

研究開発センター 機械・システムラボラトリー主任研究員。
ロボットのソフトウェアシステムの研究・開発に従事。日本ロボット学会会員。
Mechanical Systems Lab.



大明 準治 OAKI Junji

研究開発センター 機械・システムラボラトリー研究主務。
ロボット制御の研究・開発に従事。日本ロボット学会、計測自動制御学会会員。
Mechanical Systems Lab.



辰野 恭市 TATSUNO Kyoichi, D.Eng.

研究開発センター 機械・システムラボラトリー研究主幹, 工博。ロボット及び光応用計測技術の研究・開発に従事。日本機械学会, 日本ロボット学会, 計測自動制御学会会員。
Mechanical Systems Lab.