F-02

ロボットの視覚と固有受容性感覚からの動作指令の意味獲得

Acquiring the Meaning of Action Instructions from Vision and Proprioception of a Robot

松岡 啓† 深田 智† 尾関 基行† 岡 夏樹†
Hiromu Matsuoka Chie Fukada Motoyuki Ozeki Natsuki Oka

概要:
ロボットが人のように言葉という記号を運動や知覚に接地させることは困難である．本研究では，ロボットに対し「右手を上げて」や「リンゴを前に出して」と話しかけながら，ロボットの手を取り直接動作教示する．ロボットは，視覚と固有受容性感覚に基づき，言葉と動作対象（自己身体やそれ以外の物体）の対応や，言葉と動作の対応を学習する．動作対象には，“右手”といったロボット自身の身体部位，及び“リンゴ”といったロボットの操作対象の 2 種類の物体を用いた．これらの動作対象の物体をロボットの視覚で捉えることにより，動作対象の物体と「右手を」や「リンゴを」といった動作対象を表す語句の対応を学習した．また，動作を物体の座標変化であると仮定し，視覚及び固有受容性感覚から入力された動作対象の動作を取得した．語句ごとに物体の座標変化をクラスタリングし，物体の座標変化のクラスタと「上げて」や「前に出して」といった動作を表す語句の対応を学習した．

1. 背景

ロボットが人のような体を持つことにより，ロボットは人間同士が行うコミュニケーションを習得することが可能となっている．ロボットが人と同様に言語を獲得することが可能になれば，意図の共有が容易になり，コミュニケーションを行う性能の向上が図れると考えられる．また，人はコンピュータの扱い方について習熟せずとも，ロボットの有する機能を利用することが可能となる．

ロボットが柔軟に言語獲得を行うためには，言語(記号)と運動(実世界)間が接地され，文の意味に沿う動作や内部処理をロボットが行うことができるようになる必要がある．

Harnad が提唱した Symbol Grounding（記号接地）問題とは，抽象的な記号を用いる高次認知において，それらの記号と現実の運動や知覚といった低次の要素との間に如何に柔軟なインタフェースがとれるか，という問題である[1]．つまり，ロボットに言語と，ロボット自身の運動や知覚との対応関係を持たせるためには，記号接地を柔軟にしなければならない，という問題がある．

また，言葉には Compositionality（組合せ合成可能性）という特性がある．Compositionality の原理とは「複合した構成要素の意味は，それを構成する各要素の意味と構成の規則によって決定する」ということである[2]．つまり，「文章の意味は，それを構成する単語の意味と単語の組み合わせにより生成される」ということである．また，ロボットの高次認知の記号処理において，言葉と運動の関係を扱うためには，Compositionality と呼ばれる特性を，ロボットの内部で表現する必要がある．

Tani ら[3]は，高次の記号の階層と低次のアナログな運動・知覚の階層の間に RNNPB（Recurrent Neural Network with Parametric Bias）を設けることによる接地を試みた．RNNPB は現在の状態を入力とし次の状態を出力とすることにより予測を行う，Jordan 型の RNN（Recurrent Neural Network）を基にした予測器である．これにより単文と動作間の多対多の写像を表現した．また未学習である単文に対応する運動を生成する汎用性を持たせることを実現した．Cappa らの研究[4]によれば，人は名詞と動詞では言語処理が行われる脳部分が異なる．このため本研究では，日本語の語句を，物体を表す言葉と動作を表す言葉という，compositional な要素ごとに分かれて学習することを目指した．

Roy[6]は，英語の音声と画像から名詞を獲得し，単語列を生成させた．Tani ら[3]は，英語を対象とし，名詞・動詞の意味と語順を獲得するシステムを構築した．実験では，ロボットのホームポジションからの移動の軌跡及びアームの挙動と，対象を示す語及び動作を示す語を組み合わせた「point red」や「push center」といった文の対応を学習させた．これに対し本研究では，ALDEBARAN Robotics 社のヒューマノイドロボット（ Nao T14 （以下 Nao ））を用い，直接教示を経てロボットに動作指令である語句の意味を獲得させることとした．ロボットに対し「右手を上げて」や「リンゴを前に出して」と話しかけながら，ロボットの手を取り直接動作教示を行った．

言語学習において，環境との相互作用のある身体は重要である．人は，身体をもって環境に働きかけ，身体を通して環境を知覚することで学習する[5]．ロボットも，身体性に基づいた，言語の獲得のための認知処理手法が必要である．Shinozawa ら[7]は、同じ外見のロボットとソフトウェアエージェントを比較して，ロボットのほうが情報伝達に際する影響が大きいことを明らかにした．本研究では，ロボットに対し「右手を上げて」や「リンゴを前に出して」と話しかけながら、ロボットを直接操作するインタラクションを行う．

環境との相互作用の一つとして，人は筋紡錘，腱，関節などの身体内部の受容器がもたらす固有受容性感覚により，自身の身体位置関係を把握することが出来る．また，対象が視界内にある場合は視覚によっても，その位置を把握することが出来る．本研究では，ロボットに，視覚と固有受容性感覚に基づいた，言葉と動作対象（自己身体やそれ以外の物体）の対応や，言葉と動作の対応を学習させる．ロボットは，関節角度や関節間距離によって自身のパーツ位置を把握することが出来る．リンゴやミカンが置かれている時は視覚によりその位置を捉えることが出来る．また，対象を持ち上げている時には視覚のみならず，自身の固有受容性感覚をも併せて位置を把握している可能性がある．本研究では学習する動作対象が，“右手”“左手”といったロボット自身の身体部位と，“リンゴ”“ミカン”といったロボットが操作する対象の 2 種類ある，という特徴がある．

†京都工芸繊維大学 大学院工芸科学研究科, Graduate School of Science and Technology, Kyoto Institute of Technology

2．研究の目的

ロボットに，視覚，つまりカメラからの入力と，固有受容性感覚，つまり関節間距離及び関節角度センサからの入力に基づいて，言葉と動作対象（自己身体やそれ以外の物体）の対応や，言葉と動作の対応を学習させる．

例として，ロボットが「右手を上げて」という動作の依頼を受けた状況について考える．

まず，機能語の「～（し）て」という部分から自分が動作主であり，自身が動作指令を受けたことを既に認識できるものとした．さらに，内容語同士の関係を表す「を」の助詞の働きに関する知識を Nao は既に持っているものとした．

本研究では，動作対象を表す語句と動作を表す語句の組み合わせ「右手を上げて」において，動作対象を表す「右手（を）」と動作を表す「上げ（て）」を，それぞれ「左手（を）」や「下げ（て）」と組み換え，あるいは入れ替えてもよいという知識も予め Nao は持っているものとした．

その上で，内容語「右手」と「上げ（る）」の獲得について，「右手」という語句が物体を表す語句であり，さらに自身の身体部位であることを学習させる．また，ひと通りではない物体の座標変化と語句「上げ（る）」の対応を学習させる．

本研究では，入力される言葉については，「右手を上げて」から機能語である「を」や「～（し）て」と，内容語「上げ（る）」及び「上げ（る）」を分け，内容語を学習するものとし，内容語と機能語の区別に関して，ロボットは既知とする．

動作対象にあたる物体には複数の属性があり，視覚つまりカメラを用いれば，画像特徴量としての RGB 値や大きさ，位置といった属性の情報が得られる．本研究では，これら属性の中から動作対象の中心点の位置情報を用いる．これを元に動作対象に当たる物体に対して「右手を」や「リンゴを」といった語句を対応させ，Nao に学習させる．また，動作については，動作対象である物体の中心点の位置変化を動作であると仮定して学習する．動作対象については，動作の開始時と終了時に視界に捉える物体を動作対象であると仮定して学習した．

つまり，「語句」を発話として，「物体」を動作対象として，「物体の位置座標の変化」を動作として仮定する．そして「語句」と「物体」対応，および，「語句と」と「物体の位置座標の変化」の対応を学習する．学習タスクでは，人が「右手を上げて」といった発話をしながら Nao の身体パーツを直接動かしポーズを変更する．この時与えられた「語句」，Nao の視覚で捉えた「物体」，及び「物体の位置座標の変化」が，学習時にロボットに与えられる入力となる．

3. 学習データ取得実験

ユーザーが Nao に対し「右手を上げて」や「りんごを前に出して」と入力しながら，Nao の手を取り直接動作教示を行う．Nao は，視覚としてカメラから取得される動作対象の中心点の座標，及び固有受容性感覚として自身の関節間距離・関節角度を元に算出される動作対象の中心点の座標を元に，言葉と動作対象（自己身体やそれ以外の物体）の対応や，言葉と動作の対応を学習する．

ただし，現段階では直接教示時に，どの動作対象の中心点が視野内にあるのかを判別する際にのみカメラからの入力を使用しており，カメラによる動作対象の中心点の座標取得は未実装である．本来視覚によって取得される，物体“リンゴ”や“ミカン”の初期位置の中心点の座標は，カメラを用いず初期位置を固定として予め入力している．また，現在は物体“リンゴ”や“ミカン”を移動させた場合の座標は関節角度を元に算出しており，将来的には外部カメラを用いることを予定している．

3.1 実験環境

実験環境を図 1 に示す。本実験で用いた Nao は高さ 31cm の上半身モデルであり，これを高さ 5cm の台の上に載せている．Nao への発話は，将来は音声認識により検出する予定であるが，現時点では図 2 に示す画面上のボタンをクリックすることで代用している．

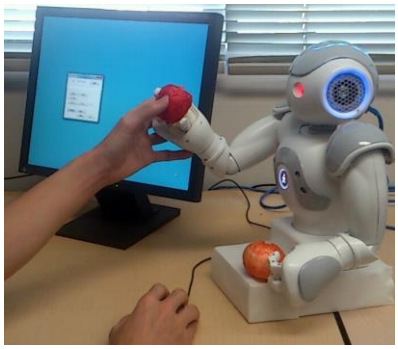


図 1 実験環境

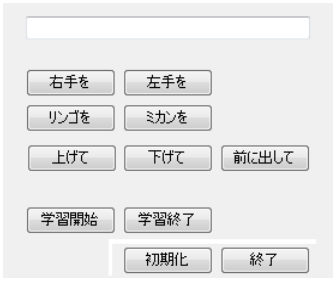


図 2 語句入力画面

3.2 タスク

言語入力はボタンをクリックすることで代用する．「右手を上げて」という語句に対する動作を学習させる際には，「右手を」「上げて」と入力をしながらロボットの身体パーツを直接動かし，Nao のポーズを変更する．この時，教示者が Nao の頭部を動かすことにより，学習する動作対象となる物体を追従するようにした．これにより，Nao が注意を向ける Nao 自身が物体を決めるのではなく，教示者が決めるようにした．これにより動作前と動作後の動作対象，及びその周辺を頭部カメラで捉えるものとする．以上を 1 回の試行とし，10 回試行する．

同様にして他の語句の組合せに対しても教示を行う．物体“リンゴ”，及び“ミカン”については，試行開始時，リンゴとミカンの両方をロボットの前に置いた状態で始めるが，リンゴとミカンのどちらを左に置くかを 5 回ごとに入れ替え，「上げて」「下げて」「前に出して」の動作に際して Nao の左右どちらの腕を使うかも併せて入れ替える．この際 Nao の手のひらにより，対象を把持させたまま動作を教示する．

3.3 動作対象の中心座標の設定

動作対象の座標設定は，右手系の xyz 座標に従う．リンゴ，及びミカンの中心座標については，初期位置は固定であり，Nao に対する相対位置を予め入力している．

「リンゴを」「ミカンを」，「上げて」「下げて」「前に出して」の語句を組み合わせた動作の際には，リンゴやミカンを把持している腕の関節角度，関節間距離，及び手のひらと掌握対象の距離を元に位置を算出している．将来的に，カメラから取得した画像，及び深度情報を元に Nao に対する相対座標を算出することを予定している．

動作対象の中心座標の変化を用いた動作の表現にはリストを用いる．リストの要素は先頭から x，y，z 方向の変化量 Δx，Δy，Δz である．

物体を表す言葉は，いずれかの動作対象（すなわちリスト）に対応する，という前提知識をロボットは保有するものとする．すると，例えば，「右手を上げて」という発話と，直接教示による動作がロボットに与えられたとき，「右手を」あるいは「上げて」という言葉が動作対象のどれかと対応する可能性があることをロボットは推測できる．また， 動作を表す言葉として，リストの 3 つの要素 Δx，Δy，Δz のパターンに，「右手を」あるいは「上げて」といった名前がつくかもしれない，ということを既知の知識として与える．

これは，物体を表す概念が何らかの物理的な対象を指し，動作を表す概念が何らかの数量変化あるいはその組みを指していることが，ロボットにとって既知であるということである．また，内容語が物体を表す言葉や動作を表す言葉であることが既知であり，この状況下で内容語とその指示対象である物体や動作の対応付けを行う，という学習問題となっている．

4. 学習

（1） 動作対象

1 回の試行において，試行開始時及び終了時共に Nao のカメラの視界に収まった物体“右手”“左手”“リンゴ”“ミカン”を，語句「右手を」「左手を」「リンゴを」「ミカンを」「上げて」「下げて」「前に出して」に対応させる動作対象とする．

（2） 動作

視覚及び固有受容性感覚に対して入力があった物体の中心点の変化に対し，語句「右手を」「左手を」「リンゴを」「ミカンを」「上げて」「下げて」「前に出して」ごとにクラスタリングを行う．物体の座標変化 Δx，Δy，Δz のリスト間に，同じクラスタとなる共通の動作パターンがあった場合，その中心点の動作パターンを，動作時に入力された語句に対応させる動作とする．

物体の中心点の変化のクラスタリングには，Pelleg ら[8]の提案した x-means 法を用いた．x-means 法は，非階層的クラスタリングでありクラスタ形状を超球状で表現する k-means 法を再帰的に実行していくというものである．あらかじめクラスタ数を決めておかなければならない k-means 法とは異なり，最適なクラスタ数を推測することが出来る．クラスタリングには，データマイニングツールの Weka 3.6.9[9]を用い，最小クラスタ数 2，最大クラスタ数 5 でクラスタリングを行った．

5. 学習結果と考察

入力された語句に対応する物体の座標変化を図 3 に示す．動作対象を示す語句「右手を」「左手を」「リンゴを」「ミカンを」，及び動作を示す語句「上げて」「下げて」「前に出して」の組合せ 12 通りの中から代表的な 4 つの組み合わせを選び（a）（b）（c）（d）に図示した．

固有受容性感覚で取得しているものは，常に物体“右手”と“左手”の中心点の座標である．

（b）のように語句「右手を」または「左手を」，及び語句「上げて」または「前に出して」が入力されているときは，それぞれ物体“右手”または“左手”のみが視覚から入力された．

図 3（a）（d）のように語句「下げて」が入力されている時，動作対象だけでなく，動かされることなく置かれたままの物体“リンゴ”や“ミカン”も常に視野内に収まり，中心座標が視覚から入力された．

図 3（c）（d）のように語句「リンゴを」または「ミカンを」が入力されている場合，物体“リンゴ”や“ミカン”は 10 回の試行のうち，5 回で把持する左右の手を持ち変える．このため，固有受容性感覚に入力される物体“右手”及び“左手”の座標は，5 回はほぼ変わらず，5 回は大きく変化する結果となっている．

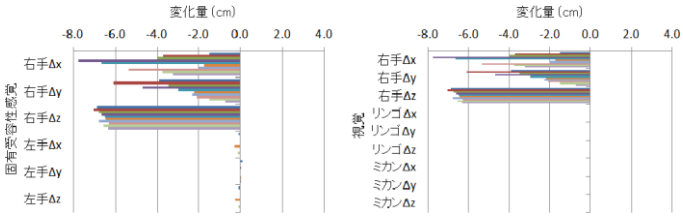
動作対象の学習の結果は次の通りであった．語句「右手を」「左手を」「リンゴを」及び「ミカンを」については，それぞれ物体“右手”“左手”“リンゴ”及び“ミカン”が対応した．また，語句「上げて」「前に出して」については物体と対応することはなかった．しかし「下げて」については，物体“リンゴ”“ミカン”が対応してしまった．原因は，動作していない物体と動作している物体を，同じように視覚に入力された動作対象として扱ったためだと考えられる．

x-means 法を用いて物体の座標変化を語句ごとにクラスタリングした結果を表 1 に示す．

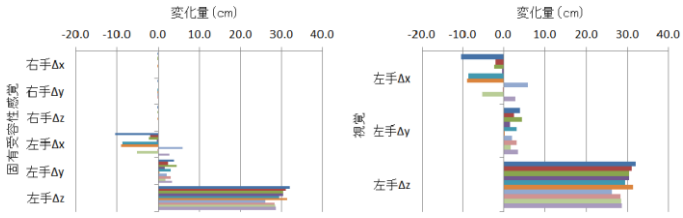
入力語句とクラスタ数の関係は「右手を」が 3，「左手を」が 3，「リンゴを」が 2，「ミカンを」が 2，「上げて」が 2，「下げて」が 2，「前に出して」が 2 となった．

動作に関する語句「上げて」「下げて」「前に出して」は，動いた物体の座標変化と，動かなかった物体の座標変化の 2 つにクラスタリングされた．また，語句「上げて」「下げて」「前に出して」3 つとも，動作を表す，動いた物体の座標変化のクラスタが，最もインスタンス数の多いクラスタを形成した．

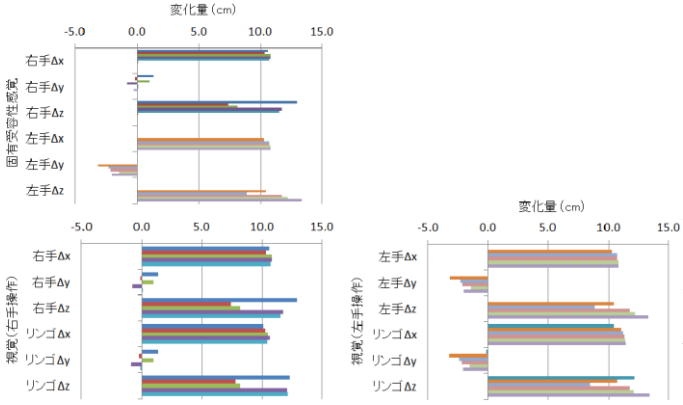
動作を表す語句は座標変化の値のパターンに共通点が現れるためクラスタ数が少なく，動作対象を表す語句は異なる変化を持つ動作を表す語句「上げて」「下げて」「前に出して」とそれぞれ対応するためクラスタ数は多くなる，と予想していた．しかし，結果は「リンゴを」や「ミカンを」もクラスタ数は動作を表す語句と同数となった．「リンゴを」や「ミカンを」共に Cluster0 は語句「下げて」の Cluster0 のように Δz の値減少，Cluster1 は語句「前に出して」「上げて」の Cluster0 の間の値となった．これは，手のひらで動作対象を把持した状態では語句「上げて」が入力された場合でも「右手を」や「左手を」の場合ほど手を上げなかったため，「前に出して」の動作と大きな差が出来ず同一のクラスタになったためである．



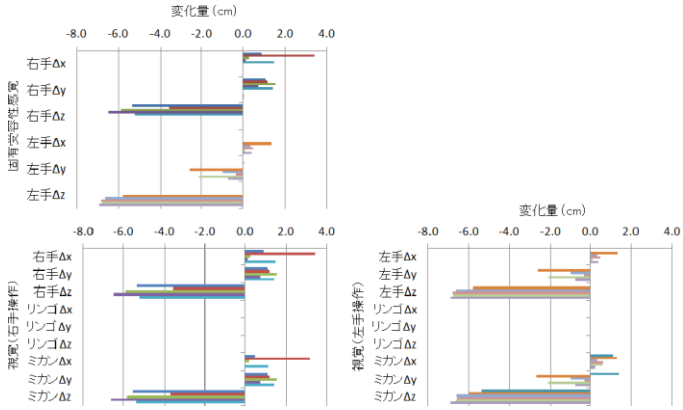
(a) 語句「右手を下げて」に対する物体の座標変化量



(b) 語句「左手を上げて」に対する物体の座標変化量



(c) 語句「リンゴを前に出して」に対する物体の座標変化量



(d) 語句「ミカンを下げて」に対する物体の座標変化量

図 3 各語句の入力に対する物体の座標変化量

表 1 物体の座標変化量をクラスタリングした結果

入力語句	Cluster 0				Cluster 1			
	Δx (cm)	Δy (cm)	Δz (cm)	Clustered (%)	Δx (cm)	Δy (cm)	Δz (cm)	Clustered (%)
右手を	2.7	0.7	1.7	64	−8.6	0.4	31.2	18
左手を	0.0	0.0	0.0	45	3.8	0.2	19.0	36
リンゴを	0.5	−0.1	−2.6	54	8.6	−1.2	14.6	46
ミカンを	0.4	0.0	−2.6	54	7.8	−1.3	15.8	46
上げて	1.1	−0.2	23.4	71	0.0	0.0	0.0	29
下げて	−0.8	−0.2	−6.2	53	0.0	0.0	0.0	47
前に出して	10.4	−0.7	9.9	71	−0.1	−0.1	−0.1	29

入力語句	Cluster 2			
	Δx (cm)	Δy (cm)	Δz (cm)	Clustered (%)
右手を	−4.0	−3.0	−6.6	18
左手を	−3.0	2.5	−6.5	18
リンゴを	—	—	—	—
ミカンを	—	—	—	—
上げて	—	—	—	—
下げて	—	—	—	—
前に出して	—	—	—	—

6. 結論と今後の課題

本稿では，ロボットが，視覚と固有受容性感覚に基づき，言葉と動作対象（自己身体やそれ以外の物体）の対応や，言葉と動作の対応を学習した．動作対象の物体をロボットの視覚で捉えることにより，動作対象の物体と「右手を」や「リンゴを」といった動作対象を表す語句の対応を学習した．また，動作を物体の座標変化であると仮定し，語句ごとに物体の座標変化をクラスタリングし，物体の座標変化のクラスタと「上げて」や「前に出して」といった動作を表す語句の対応を学習した．

今後は，カメラからの物体の中心点の座標取得を実装する．また，今回の実験では物体“リンゴ”“ミカン”を動作対象とするとき，常に Nao に把持させて学習させた．リンゴやミカンが Nao 自身の身体部位ではなく別の物体である特性が活かしきれていなかったため，今後，Nao にリンゴやミカンを操作させるだけでなく，人が Nao の前でリンゴやミカンを操作する，といった学習を行なう予定である．

参考文献

[1] Stevan Harnad.: The symbol grounding problem. *Physica D*, Vol. 42, pp. 335-346, 1990.
[2] Jun Tani, Ryunosuke Nishimoto, and Rainer W.Paine.: Achieving “Organic Compositionality” through self-organization: Reviews on Brain-Inspired Robotics Experiments. *Neural Networks*, Vol. 21, pp. 584-603, 2008.
[3] Yuuya Sugita and Jun Tani.: Learning Semantic Combinatoriality from the Interaction between Linguistic and Behavioral Processes. *Adaptive Behavior*, Vol. 13, No. 1, pp. 33-52, 2005.
[4] Stefano F. Cappa and Daniela Perani.: The neural correlates of noun and verb processing. *J.Neurolinguistics*, Vol. 16, pp. 183-189, 2003.
[5] 國吉康夫, ベルトゥーズリュク. 身体性に基づく相互作用の創発に向けて. 日本ロボット学会誌, Vol. 17, No. 1, pp.29-33, 1999.
[6] Deb Kumar Roy.: Learning visually grounded words and syntax for a scene description task. *Computer Speech and Language*, Vol. 16, No. 3, pp. 353-383, 2002.
[7] Kazuhiko Shinozawa, Futoshi Naya, Junji Yamato, and Kiyoshi Kogure.: Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies*, Vol. 62, No. 2, pp. 267-279, 2005.
[8] Dan Pelleg and Andrew W. Moore.: X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *In Proceedings of the 17th International Conf. on Machine Learning*, Morgan Kaufmann ,pp. 727-734, 2000.
[9] WEKA The University of Waikato, http://www.cs.waikato.ac.nz/ml/weka/, (最終閲覧日2013年7月24日)