ロボットによる大規模言語学習に向けて -実世界知識の利活用とクラウドロボティクス基盤の構築-

杉　浦　孔　明*

＊ 国立研究開発法人情報通信研究機構，京都府相楽郡精華町光台 3–5
＊Published monthly by The Society of Instrument and Control Engineers
Hongo 1–35–28–303 Bunkyo-ku Tokyo Japan
＊E-mail: komei.sugiura@nict.go.jp

キーワード：模倣学習 (imitation learning)，音声対話 (spoken dialogues)，クラウドロボティクス (cloud robotics)
JL 0006/14/5306–0001

1．はじめに

半世紀以上も前から，人との音声コミュニケーションが可能なロボットは，音声翻訳機と並ぶ夢のシステムであった [1]．今日，音声言語処理技術は大きく進歩し，スマートフォンや PC などのコンシューマデバイスにおいて，音声エージェントが広く利用可能になっている．また，接客などを目的とした人型のロボットが市販されるなど，コミュニケーションを行うロボットに対して活発に研究開発が行われている．

ただし，あらかじめ決められたシナリオでは流暢に会話するように見えても，シナリオ外の内容については支離滅裂な会話を行うロボットは多い．一方，ハードウェア上の制約を持つロボットの中には，指示された物を運ぶなどの物理的な利便性を提供できないものもある．これらの事実が示すことは，コミュニケーションと物理的な利便性の双方において，実応用レベルのロボットを構築することは実際には簡単ではない，ということである．移動・物体認識・把持などの基本機能から，多様な話者に対する音声認識や指示の意味理解にいたるまで，要素技術の精緻化と高度な機能統合が必要とされる．

このように，時々刻々変化する実世界における感覚運動系にグラウンドした言語処理は，多くの関連課題を有する挑戦的な分野である．例として，日常環境で「新聞片付けておいて」という音声指示をロボットが実行するタスクを考える．環境中には「新聞（に分類され得るオブジェクト）」が複数存在する可能性があるうえ，「どれを」「どこに」「どうやって」片付けるか，など様々なレベルで曖昧性が存在するため，望ましい動作を実行することは簡単ではない．また，指示者にとって当たり前であるタスクの開始条件・終了条件や，行動途中で変化した状況に対しロボットがとるべき行動も，指示文の言語情報のみからは自明でないことが多い．

片付け指示の理解に限らず一般的な音声言語理解において，曖昧性解消手段は複数あり得る．例えば，単純な規則集合の適用，実世界情報に基づく推定（マルチモーダル発話理解），対話による曖昧性解消，などがある [2]．手段の評価尺度としては，タスク達成率，費用対効果，安全性，ユーザビリティなどが代表的な尺度であろう．ここで，本稿において「曖昧性解消」とは，タスク遂行に関する曖昧性を完全に解消することではなく，ある尺度で定義された曖昧性を減少させることを指すものとする．

近年，音声認識・自然言語処理・パターン認識などの分野で深層学習と大規模データを組み合わせたデータ指向アプローチが成功し，Visual QA などの実世界情報と言語を扱う課題を扱うベンチマークテストが出現している [3]．本分野は挑戦する価値がある，との認識が広まりつつあることの傍証であろう．また，本特集で紹介されるように，ロボティクスにおいても記号コミュニケーションに関する試みが実応用と結びつく事例も報告されるようになった．多くのセンサ・アクチュエータを扱うロボティクスは，大規模データの生成や利活用と親和性が高く，今後，ロボティクスにおいてもデータ指向アプローチがイノベーションを起こす可能性は十分ある．

本稿では，実世界知識を扱う音声対話に関して我々が取り組んできた事例（要素技術の関係を**図 1** に示す）を紹介するとともに，分野の動向と今後の展望について論じる．本稿の構成は以下の通りである．まず，2 で実世界知識を扱う音声対話の関連研究を紹介する．3 章ではこれまで我々が取り組んできた動作の言語化技術を紹介し，4 章では動作の言語化技術を用いたロボット対話技術を紹介する．5 章では，大規模データの利活用事例として，音声対話向けクラウドロボティクス基盤 rospeex につい

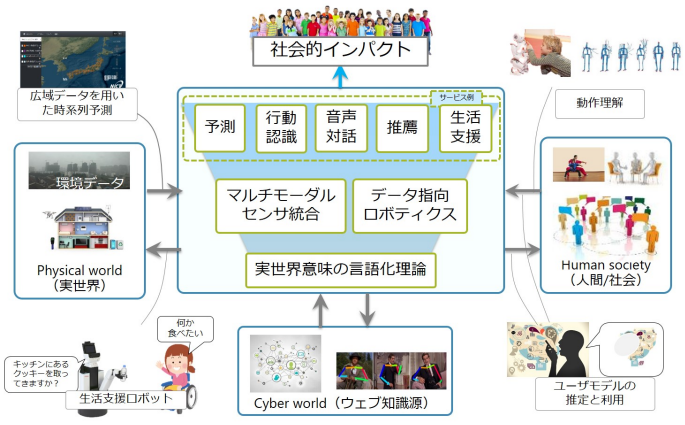


図 1　本稿で紹介する要素技術および応用の関係

て述べる．6 章で今後の展望について述べたのち，7 章で結論を述べる．

2. 実世界知識を扱う音声対話

実世界知識を扱う音声対話に関する分野は，実世界知識から記号・言語への一方向あるいは双方向の変換を扱う分野との関連が深い．1970 年代に Winograd により提案された SHRDLU は，本分野の嚆矢となる対話システムであった．SHRDLU が扱ったタスクはシミュレーション上の積み木の操作に限定されるものの，深い概念理解を伴う対話を行った点で意義深い．音声およびテキスト対話システム分野のその後の動向については，[1] が詳しい．

音声対話を行うロボットの先駆的事例としては，Jijo-2 が挙げられる [5]．また，稲邑らは，移動ロボットの障害物回避において，ロボットが段階的に行動決定モデルを獲得する機構を提案した [6]．センサ値と行動の関係をベイジアンネットを用いてモデル化し，推論結果の確信度を用いて応答決定を行う．

2000 年代以降，ロボティクスの分野では，動作-記号の相互変換に関する試みが近年広く行われてきた [2]．高野らは，運動の分節化を通じてヒューマノイドロボットが獲得した原始シンボルを用いて，運動認識・生成を行っている [7]．Ogata らは，動作系列と記号列の間の多対多対応問題を扱うリカレントニューラルネットに基づく手法を提案している [8]．Kollar らは，ロボットに与える移動指示に関して，ランドマークオブジェクトや動作にグラウンドした言語表現を学習する手法を提案した [9]．学習されたモデルを用いることにより，指示が示す最も確からしい経路を推論する．

本分野に関連するプロジェクトとしては，DARPA BOLT [10]，Robo Earth [11]，RoboBrain [12]，長井らによる CREST プロジェクト [13]，などがある．2011-15 年に推進された DARPA BOLT プロジェクトでは，翻訳と並び，グラウンドした言語学習が柱のひとつであった [10]．Robo Earth プロジェクトで提案されたシステムでは，タスクのサブタスクへの分割や物体情報のオントロジーを保持しており，複数のロボットが知識共有可能である [11]．また，RoboBrain プロジェクトでは，オンラインで入手可能なデータを可能な限り収集し，利用可能にすることを目指している [12]．[13] では，人間と機械が意味理解を伴ったコミュニケーションに基づき，日常的なタスクを協調して実行するための基盤技術の確立を目指している．

図 2　ロボカップ@ホーム環境においてタスクを行うボン大学のロボット [4] と音声指示の例

本分野と関連が深いベンチマークテストとしては，ロボカップ@ホーム [14, 15] がある．ロボカップ@ホームは世界最大の生活支援ロボットのコンペティションであり，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，などの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を統合したタスクが設定されている．**図 2** に，ロボカップ@ホームタスクにおいてタスクを行うボン大学のロボット [4] と音声指示の例を示す．

一方，我々は音声言語処理の基礎技術から実世界知識を扱う音声対話およびクラウドサービスにいたるまで包括的な研究を行っている．これまで，後述する事例のほか，京都観光案内対話システムやニュース動画への字幕付与システムを開発してきた [16]．さらに社会展開として，音声対話向けクラウドロボティクス基盤 rospeex の公開 (注1) や，大規模コーパスの公開 (注2)，多言語音声翻訳・音声対話システムの開発を容易にするツールキットの公開を行ってきた (注3)．

3. 動作の言語化と予測

動作にグラウンドした言語理解および言語生成には，動作のモデル化が必要不可欠である．動作の分類については種々の分類が議論されており，姿勢や物体位置時系列レベルでのモデル化，操作対象物への効果を含む行為レベルのモデル化，一連の行為の目的を含むレベルでのモデル化，などがある [17]．以下では，相対座標系を必要とする動作と必要としない動作を同時に学習可能な枠組みについて，我々が取り組んだ事例を紹介する．

「X を Y にのせる」や「Z を回す」など参照点に依存した動作の模倣では，世界座標系での動作軌道の模倣に意味はなく，適切な座標系を推定し軌道を汎化しなければならない．このために，参照点に依存した隠れマルコフモデル（Reference-Point-Dependent HMM，以下 RPD-HMM と略記）を開発した [18]．RPD-HMM は，物体位置の時系列を入力として，動作をモデル化するための最適な座標系を EM アルゴリズムにより推定し，軌道のモデルを学習している．

実験では，被験者 3 名に日用品が複数存在する環境で

(注1) http://rospeex.org/
(注2) https://alaginrc.nict.go.jp/resources/nict-resource/slc-info/slc-outline.html
(注3) http://www2.nict.go.jp/astrec-ast/mcml-sdk/index.html

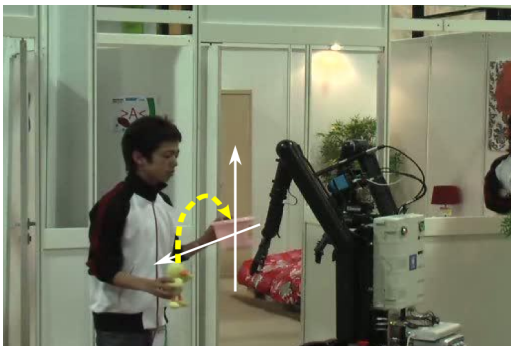


図 3　ロボカップ@ホーム環境における動作「捨てる」の学習

物体操作に関する動作を行わせ，学習およびテストデータを収集した．評価尺度としては動作の認識精度を用いた．ここで，本タスクにおける認識精度とは，動作の種類および関連するオブジェクトをすべて正解した場合のみ，動作認識結果が正解であるものとする．テストデータにおけるチャンスレベル（ランダムに動詞とオブジェクトを選択した場合の認識精度）は 3.03%であったが，RPD-HMM による認識結果は平均 90%であった．

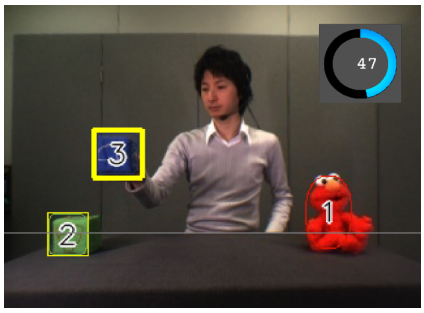
動作の生成時には，学習時とまったく同じように物体が配置されている訳ではないため，たとえ同じ「載せる」動作であっても，学習時の軌道をそのまま用いることは無意味である．学習した RPD-HMM は固有座標系において汎化されたものであるので，固有座標系 C から世界座標系 W へ変換する．RPD-HMM の位置・速度・加速度の平均ベクトルおよび共分散行列は，同時変換行列により C から W 上のモデルに変換される．HMM から滑らかな軌道を生成するために，音声合成の分野で用いられる HMM 軌道生成 [19] を用いる．実験の結果，7 回程度の教示で生成軌道の誤差が収束することが確認できた．

我々は，RPD-HMM の日常タスクへの応用も進めている．ロボカップ@ホーム環境における家事動作の学習への適用例を**図 3** に示す．RPD-HMM の拡張としては，RRT(Rapidly-exploring Random Trees) とのハイブリッドアプローチや，逐次軌道生成などをこれまで提案している [20]．さらに，時系列に特化した深層学習手法である Dynamically Pre-Trained Deep Recurrent Neural Network (DPT-DRNN) を提案し，関節角時系列を高い精度で予測可能であることを示した [21]．CATS ベンチマークにおいてベースライン手法と比較して DPT-DRNN は高い予測精度を得ただけでなく，実際の PM2.5 観測データ（50 都市，2 年分）から予測モデルを構築し，気象モデルベースの手法を上回る予測精度を達成している [8]．

4.　ロボット対話

4.1　LCore におけるマルチモーダル発話理解

一般的にロボットの対話処理機構では，実世界情報と音声認識は別々に処理されていることが多い [22]．一方，



【状況】オブジェクト 2 が直前に操作された
U: ハコ エルモ ちかづけて．
R: ミドリハコをちかづけて？
U: いいえ．
R: アオイハコをちかづけて？
U: はい．
R: (動作実行：オブジェクト 3 をオブジェクト 1 に近づける)

図 4　グラウンドした語彙による確認発話の生成

画像，動作，アフォーダンス，履歴などさまざまな実世界情報の学習と統合は知能ロボティクスの到達点の一つであろう．

このような背景から我々はコミュニケーション学習基盤 LCore [23–25] を開発してきた．LCore はマルチモーダル入力から学習されたモデルを用いて，音声言語理解を行う．具体的には，発話 s と状況 O（オブジェクトの位置や視覚特徴など）を入力として，以下の発話理解スコア Ψ を最大化する行動 a_k を出力する．

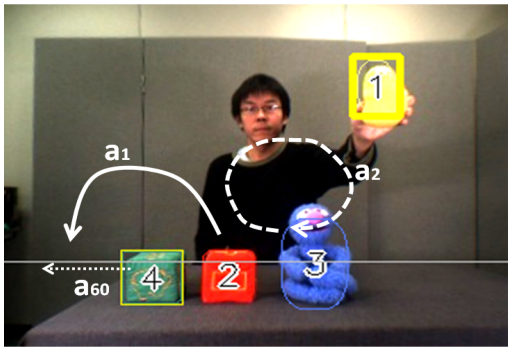
$$\Psi(s, a_k, O, \boldsymbol{q}^{\langle i_t \rangle}) = \max_z (\gamma_1 B_S + \gamma_2 B_I + \gamma_3 B_M + \gamma_4 B_R + \gamma_5 B_H)$$

ここに，z は各単語の概念構造への分割であり，$(s, a_k, O, \boldsymbol{q}^{\langle i_t \rangle})$ はそれぞれ，発話，行動，状況，トラジェクタに関するコンテキストを表す．また，$\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_5)$ は各項に対する重みであり，MCE 学習 [26] を用いて学習される．B_S, B_I, B_M, B_R, B_H はそれぞれ，音声，視覚情報，予測軌道，動作-オブジェクト関係，行動コンテキストの尤度スコアを表す．B_M の学習については，前章で述べた手法が用いられる．その他の詳細については [23, 24] を参照されたい．

4.2　期待効用最大化に基づく確認発話の生成

1 章で述べたように，実世界知識を扱う音声対話における曖昧性解消は簡単な課題ではない．例えば，ユーザが「コップ（テーブルに）置いて」などの曖昧な発話を行ったとする．曖昧性を解消するためには，「赤いコップですか，青いコップですか」のような確認発話を毎回行なってもよいが，曖昧なときのみ確認発話を行う方が望ましい．

そこで我々は，発話理解確率（発話を正しく解釈できる確率）を推定し，効用最大化により応答を生成する手法を提案した [27]．発話が曖昧であるとき，発話理解結果



Target action	Robot utterance	Margin	ICM	ELL
a_1	*"Jump-over red box"*	15.0	0.913	3.41
a_1	*"Jump-over box red box green"*	17.9	0.945	3.43
a_1	*"Jump-over"*	0.077	0.473	4.33
a_1	:	:	:	:
a_2	"Rotate Glover"	574	1.000	3.52
a_2	:	:	:	:
:	:	:	:	:
a_{60}	*"Move-away box green box red"*	26.8	0.987	3.48

図 5　能動学習に基づく発話の生成例

の第 1 候補と第 2 候補のスコアの差（マージン）が小さいことから，これを曖昧性の尺度として用いている．提案手法では，マージン d を入力として，ベイズロジスティック回帰により発話理解確率の推定を行う．動作応答 b_1 と確認発話応答 b_2 は，対応する期待効用 $\mathbb{E}[R_i](i = 1, 2)$ の最大化により選択される．

$$\mathbb{E}[R_i] = r_{i1} f(d; \boldsymbol{w}) + r_{i2}(1 - f(d; \boldsymbol{w})) \quad (1)$$

$$d = \Psi(s, \hat{a}, O, \boldsymbol{q}) - \max_{a_j \neq \hat{a}} \Psi(s, a_j, O, \boldsymbol{q}) \quad (2)$$

ここに，$f(d; \boldsymbol{w})$ は，$\boldsymbol{w} = (w_0, w_1)$ をパラメータとするロジスティックシグモイド関数である．また，r_{i1}, r_{i2} はそれぞれ，行動 $\hat{a}$ がそれぞれ正解，不正解のときの応答 b_i に対する効用である．

図 4 は，ユーザ (U) とロボット (R) の対話例を示したものである．図において，右上の数値は発話理解確立の推定値 $f(d)$ を表す．図 4 では $f(d)$ は約 47%であり，確認発話「アオイ ハコをちかづけて」が最適応答であった．定量評価の結果，動作や視覚などの情報を用いず，音声のみで発話理解を行った場合の行動失敗率は 83.4%であった．ベースライン手法（発話を行わず動作のみで応答する）における行動失敗率は 12.0%である一方，提案手法による行動失敗率は 2.6%であった．このことから，提案手法はベースライン手法に比べて行動失敗率を大幅に低減できたといえる．

4.3　能動学習に基づく対話戦略最適化

ここまで，ユーザが指示を行いロボットが確認発話や物体操作を行う問題について述べてきた．このようなインタラクションでは，学習フェーズと実行フェーズを明示的に分けるか否かに関わらず，正事例と負事例を含む学習サンプルを収集する必要がある．一方，本タスクでは負事例は動作失敗を意味するので，安全面の理由から動作失敗は少ないことが望ましい．

この課題に対し，ユーザがロボットに指示するのではなく，ロボットが指示発話を行ってユーザが動作することで学習サンプルを得るアプローチを提案した [27]．ここで，ユーザがロボットに指示することによって開始されるタスクをターゲットドメイン，逆にロボットがユーザに指示することで開始されるタスクをソースドメインと呼ぶことにする．厳密には，ソースドメインとターゲットドメインのサンプルは異なる分布から得られたものであるが，我々はソースドメインで得た知識がターゲットドメインの事前分布として有効であることを示した．

ソースドメインにおける研究課題のひとつは，無数に存在する指示発話の候補から，いかに学習に有効な発話を生成するかである．本研究では，発話の生成には，能動学習の一種である Expected Log Loss Reduction(ELLR) [28] を用いた．ELLR は，テストセットにおける誤分類の期待値を最も低くするような入力を選択する．ELLR ベースの発話生成では，マージンを入力として確信度関数のパラメータを効率的に学習するように発話が選択される．実験では，発話指示が曖昧なため行動失敗を続けるユーザに対しては，発話がより明確になることが確認された．

提案手法による発話生成の例を**図 5** に示す．図 5 左図の状況においては，182 個の発話候補が生成された．図 5 右表は，各発話のマージン，発話理解確率の推定値 (ICM)，対数損失の期待値 (ELL) を示したものである．この状況では，発話「赤い 箱 飛び越えさせて (Jump-over red box)」が最小の ELL を持つため，この発話が音声合成され，それを受けてユーザが動作を実行する．実験の結果，能動学習による対話戦略により，学習フェーズにおけるサンプル数および動作失敗数を低減できることがわかった．

5.　クラウドロボティクス基盤 rospeex による音声対話

スマートフォンへの音声インタフェース機能の導入は，音声言語処理機能を広く一般に認知させた．この背景には，機械学習手法の進展とその大規模データへの適用とともに，音声認識機能がクラウドサービス化され，非力な端末でも高品質なサービスを利用できるようになったことがある．今後，ネットワークに接続されるロボットや IoT デバイスが増加するに従い，それらの機器を対象としたクラウドサービスも増加すると考えられる．ただし，このような基盤構築は，音声認識・合成についての基礎技術からクラウド基盤の構築・運用やロボットへの

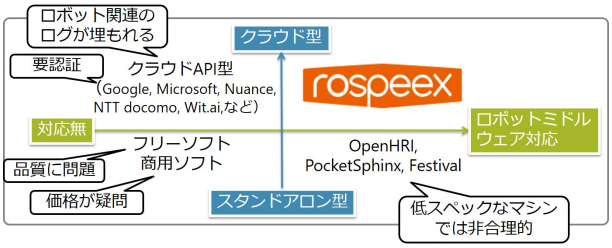


図 6　rospeex と他のシステムの位置付け

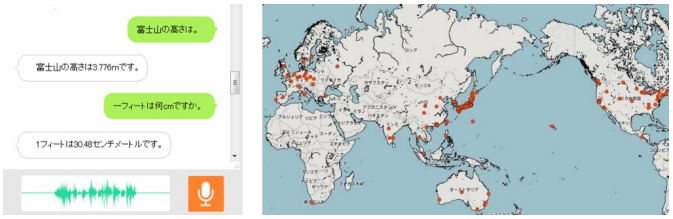


図 7　左：rospeex のユーザインタフェース．右：rospeex ユーザの IP 分布．

適用に至るまでの包括的な技術開発が必要であり，簡単な課題ではない．

上記の背景から，音声対話向けクラウドロボティクス基盤 rospeex を我々は構築し，2013 年 9 月から運用してきた [29, 30]．**図 6** に rospeex の位置付けを示す．現在，4 か国語（日英中韓）の音声認識・合成に対応しており，学術研究目的に限り無償・登録不要で公開している．これまでに，高齢者施設におけるコミュニケーションロボット，ヒューマノイド，対話エージェント，スマートホーム向け音声インタフェース，ホテルでの受付ロボット，などに利用されている (注4)．**図 7** に rospeex のユーザインタフェースおよびユーザ IP の分布を示す．

5.1　分割送信によるレスポンス時間の短縮

ロボット対話における自然性を向上させるため，クラウドサービスへの通信時間を短縮することは重要な課題である．そのため，音声認識において分割送信方式に対応している．分割送信とは，発話区間検出（Voice Activity Detection; VAD）終了後に一括で送信するのではなく，音声入力中に前もって音声を分割・送信する方法を指す．

分割送信の効果を検討するため，一括送信との比較実験を行った．一括送信でも分割送信でもサーバ上での音声の処理時間には大きく違いがあるものではない．違いは VAD 終了後に一括で送信するか，音声入力中に前もって送信するかの違いである．分割数を多くすれば処理時間を減少させられるものの，多くしすぎても性能は頭打ちになる．

コーパスとして ATR503 文を用い，声優 1 名による収録を行った．そのうち発話区間検出に問題がない 495 文

(注4) rospeex は 2016 年 9 月までに 4 万ユニークユーザに利用されている．なお，「rospeex」という名称は，「robot」「ROS」「speak」「speex（音声圧縮フォーマットのひとつ）」などから発想された造語である．

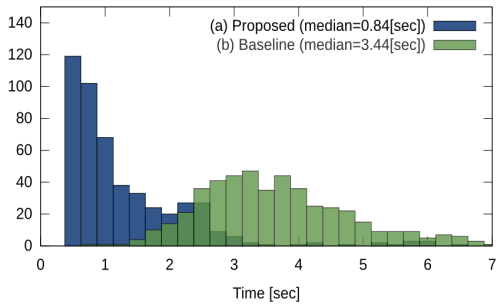


図 8　音声認識におけるレスポンス時間のヒストグラム

についてレスポンス時間を求めた．レスポンス時間とは，クラウドサービスからのレスポンスが得られた時刻から VAD 終了時刻を引いたものである．本実験では，サーバの負荷とのバランスから 3.52kB を単位として分割を行った．ATR503 文であれば，平均的に 50 ほどに分割されサーバに送信される．実験の結果，一括送信でのレスポンス時間の中央値は 3.44 秒であったが，分割送信では 0.84 秒に短縮された（**図 8** 参照）．

5.2　非モノローグ音声合成

日本語の音声合成としては非モノローグ HMM 音声合成 [31] に対応し，ロボットとの対話に特化して開発されたボイスフォントを利用可能である．ロボットのコミュニケーション機能の開発においては自然な音声合成が求められているが，一般的な音声合成器は人-ロボット対話に最適化されている訳ではない．そのため，ロボットの合成音声は自然さや親しみやすさに欠け，抑揚が適切でないため話者が質問されたことに気づかないなどの事態も起こり得る．我々はサービスロボットタスクとの対話タスクにおいて非モノローグ HMM 音声合成の評価を行い，理論的上限にせまる MOS 値を持つことを示している．性能評価の詳細については，[31] を参照されたい．

なお，音声認識または合成単体としても利用可能であり，その場合は JSON ファイルをインタフェースとする．ユーザが用いるプログラミング言語には依存しないため，C++や Python など各種のプログラミング言語を利用可能である．

6.　おわりに

国連の調査によると，2030 年には主要 7 カ国（G7）すべてで高齢化率が 20%を超えると予測されている [32]．この背景のもと，人と自然なコミュニケーションを行い自立した生活を支援するロボット技術に，社会からの期待が高まっている．一方，現状の技術レベルと期待の間に隔たりは大きい．1 章で述べたように，例えば，「新聞片付けておいて」という一見単純な指示であっても，現在の技術では「どれを」「どこに」「どうやって」などの情報を指定しない限り，望ましい挙動をロボットに行わせることは難しい．

本稿では，動作の言語化および予測と，それを利用したマルチモーダル対話，大規模データの利活用の試みについて紹介した．音声認識・自然言語処理・パターン認識などの分野で成功したデータ指向アプローチが，今後，ロボティクスにおいてもイノベーションを起こす可能性は十分あると考えられる．本研究に関する動画は `http://komeisugiura.jp` を参照されたい．

謝辞

本研究の一部は，JST CREST および JSPS 科研費 15K16074 の助成を受けて実施されたものである．

参 考 文 献

1） 河原達也，“音声対話システムの進化と淘汰–歴史と最近の技術動向–,” 人工知能学会誌，vol.28，no.1，pp.45–51，2013.
2） T. Taniguchi, T. Nagai, T. Nakamura, N. Iwahashi, T. Ogata, and H. Asoh, “Symbol emergence in robotics: a survey,” Advanced Robotics, vol.30, no.11-12, pp.706–728, 2016.
3） R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D.A. Shamma, et al., “Visual genome: Connecting language and vision using crowdsourced dense image annotations,” arXiv:1602.07332, 2016.
4） J. Stückler, D. Holz, and S. Behnke, “RoboCup@Home: Demonstrating Everyday Manipulation Skills in RoboCup@Home,” Robotics & Automation Magazine, IEEE, vol.19, no.2, pp.34–42, 2012.
5） 松井俊浩，麻生英樹，F. John，浅野　太，本村陽一，原　功，栗田多喜夫，速水　悟，山崎信行，“オフィス移動ロボット jijo-2 の音声対話システム,” 日本ロボット学会誌，vol.18，no.2，pp.142–149，2000.
6） 稲邑哲也，稲葉雅幸，井上博允，“ユーザとの対話に基づく段階的な行動決定モデルの獲得,” 日本ロボット学会誌，vol.19，no.8，pp.983–990，2001.
7） 高野　渉，中村仁彦，“統計的相関に基づく動作パターンのリアルタイム教師なし分節化と原始シンボルの自律的獲得,” 日本ロボット学会誌，vol.27，no.9，pp.1046–1057，2009.
8） T. Ogata, M. Murase, J. Tani, K. Komatani, and H.G. Okuno, “Two-way translation of compound sentences and arm motions by recurrent neural networks,” Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and System, pp.1858–1863, 2007.
9） T. Kollar, S. Tellex, D. Roy, and N. Roy, “Toward understanding natural language directions,” Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction, pp.259–266, 2010.
10） S. Mohan, A. Mininger, J. Kirk, and J.E. Laird, “Learning grounded language through situated interactive instruction.,” AAAI Fall Symposium: Robots Learning Interactively from Human Teachers, pp.30–37, 2012.
11） M. Tenorth, A.C. Perzylo, R. Lafrenz, and M. Beetz, “The RoboEarth Language: Representing and Exchanging Knowledge about Actions, Objects, and Environments,” Proc. ICRA, pp.1284–1289, 2012.
12） A. Saxena, A. Jain, O. Sener, A. Jami, D.K. Misra, and H.S. Koppula, “Robobrain: Large-scale knowledge engine for robots,” 2014.
13） 長井隆行，谷口忠大，尾形哲也，岩橋直人，杉浦孔明，稲邑哲也，岡田浩之，“記号創発ロボティクスによる人間機械コラボレーション基盤創成,” 第 19 回クラウドネットワークロボット研究会，pp.23–27，2015.
14） 岡田浩之，大森隆司，“ロボカップ@ホーム：人とロボットの共存を目指して,” 人工知能学会誌，vol.25，no.2，pp.229–236，2010.
15） L. Iocchi, D. Holz, J. Ruiz-delSolar, K. Sugiura, and T. van derZant, “RoboCup@Home: Analysis and results of evolving competitions for domestic and service robots,” Artificial Intelligence, vol.229, pp.258–281, 2015.
16） 柏岡秀紀，翠輝久，水上悦雄，杉浦孔明，岩橋直人，堀智織，“観光案内への音声対話システムの活用,” 情報処理学会デジタルプラクティス，vol.3，no.4，pp.254–261，2012.
17） 人工知能学会（編），人工知能学事典，共立出版，2005.
18） K. Sugiura, N. Iwahashi, H. Kashioka, and S. Nakamura, “Learning, Generation, and Recognition of Motions by Reference-Point-Dependent Probabilistic Models,” Advanced Robotics, vol.25, no.6-7, pp.825–848, 2011.
19） K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Kitamura, “Speech Parameter Generation Algorithms for HMM-Based Speech Synthesis,” Proceedings of ICASSP, pp.1315–1318, 2000.
20） 杉浦孔明，是津耕司，“Rpd-hmm に基づく逐次動作生成による物体操作の模倣学習,” 第 28 回人工知能学会全国大会資料，pp.1I5–OS–09b–5，2014.
21） 杉浦孔明，是津耕司，“Dynamic pre-training を導入した deep neural network による関節角時系列の予測,” 第 33 回日本ロボット学会学術講演会資料，pp.RSJ2015AC1B2–08，2015.
22） R. Hartanto, A Hybrid Deliberative Layer for Robotic Agents, Springer, 2011.
23） N. Iwahashi, “Robots that learn language: Developmental approach to human-machine conversations,” Human-Robot Interaction, ed. by N. Sanker, pp.95–118, I-Tech Education and Publishing, 2007.
24） N. Iwahashi, R. Taguchi, K. Sugiura, K. Funakoshi, and M. Nakano, “Robots that Learn to Converse: Developmental Approach to Situated Language Processing,” Proceedings of International Symposium on Speech and Language Processing, pp.532–537, 2009.
25） 杉浦孔明，“ロボット対話 -実世界情報を用いたコミュニケーションの学習-,” 人工知能学会誌，vol.27，no.6，pp.580–586，2012.
26） S. Katagiri, B.H. Juang, and C.H. Lee, “Pattern Recognition Using a Family of Design Algorithms based upon the Generalized Probabilistic Descent Method,” Proceedings of the IEEE, vol.86, no.11, pp.2345–2373, 1998.
27） K. Sugiura, N. Iwahashi, H. Kawai, and S. Nakamura, “Situated Spoken Dialogue with Robots Using Active Learning,” Advanced Robotics, vol.25, no.17, pp.2207–2232, 2011.
28） N. Roy and A. McCallum, “Toward optimal active learning through sampling estimation of error reduction,” Proceedings of 18th International Conference on Machine Learning, pp.441–448, 2001.
29） K. Sugiura, Y. Shiga, H. Kawai, T. Misu, and C. Hori, “A Cloud Robotics Approach towards Dialogue-Oriented Robot Speech,” Advanced Robotics, vol.29, no.7, pp.449–456, 2015.
30） K. Sugiura and K. Zettsu, “Rospeex: A cloud robotics platform for human-robot spoken dialogues,” Proc. IEEE/RSJ IROS, pp.6155–6160, 2015.
31） K. Sugiura, Y. Shiga, H. Kawai, T. Misu, and C. Hori, “Non-Monologue HMM-Based Speech Synthesis for Service Robots: A Cloud Robotics Approach,” Proc. ICRA, pp.2237–2242, 2014.
32） United Nations, “World population prospects: The 2012 revision,” 2013.

［著　者　紹　介］

杉浦孔明（すぎうらこうめい）君(正会員)

2002 年京都大学工学部電気電子工学科卒業．2004 年同情報学研究科修士課程修了．2007 年同博士後期課程修了．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，現在，情報通信研究機構先進的音声翻訳研究開発推進センター主任研究員．サービスロボット，機械学習，クラウドロボティクス，音声対話システム，情報検索，推薦システムなどの研究に従事．情報処理学会，人工知能学会，日本ロボット学会，IEEE などの会員．