1O4-OS-22a-5

キズ検査ロボットによる音声インタラクションを通した語彙の拡張

Method to increase vocabulary through speech interaction between human and inspection robot

藤本智也[*1] Tomoya Fujimoto　渡邉祐太[*1] Yuta Watanabe　呉比[*1] Wu Bi　田口亮[*1] Ryo Taguchi　服部公央亮[*2] Koosuke Hattori

保黒政大[*2] Masahiro Hoguro　梅崎太造[*1] Taizo Umezaki

[*1] 名古屋工業大学 Nagoya Institute of Technology　[*2] 中部大学 Chubu University

Nowadays, research on the service robots that work cooperatively with humans using speech interaction has attracted great attention. Conventional robots recognizes the predefined representations. If a user speaks undefined representations, the robot fails to recognize the utterance or rejects it. However, in the interaction among humans, they convey own intention by various linguistic representations. To improve the usability of the human-robot interaction, it has been desired that the robot can learn novel representations through the interaction. In this paper, we propose a method for a robot to add novel representations into the predefined vocabulary through human-robot interaction. In the experiment, the proposed method was embedded into a visual inspection robot as an example of service robots. The robot moves by recognizing user's voice commands and detects scratches on a surface of a car body. Moreover, it can change a parameter for image processing, and show the inspection results onto a display. The experimental results showed that novel representations for each voice command can be acquired by using the proposed method.

1. はじめに

近年，製造技術や音声・画像認識技術の発達に伴い,福祉や介護，警備やアミューズメント，教育など幅広い分野でのロボットの活躍が期待されている．これらのロボットは工業用ロボットと異なり，人の生活環境での協調動作が求められている.特に家庭やオフィスで活動するロボットは，初めてロボットに触れるユーザでも直感的にインタラクションできることが求められるため，人間が普段使用している言語を認識し応対する音声対話機能が必要となる．ロボットが人と対話するためには，言葉と実世界の事物・事象の対応関係をロボットが理解できなければならない．家庭やオフィスなどでは，未知の物や場所等に対応する必要があるため，それらを表す単語知識，すなわち語彙をユーザとのインタラクションを通して学習できることが望まれる．

ロボットによる語彙学習に関する先行研究では，ロボットに物や動作を見せながら対応する単語を発話することで，「箱」や「茶色」，「ペットボトル」などの物を表す単語や，「乗せて」や「近づけて」といった動作，「テレビの前」や「渡り廊下」など場所を表す単語を学習させた[岩橋 03，中村 08，田口 10，谷口 14]

これらの初期知識がゼロからの学習を対象とした研究に対して渡邉らはトップダウンで与えられた対話知識を元にして新たな知識を学習していくモデルを提案した[渡邉 15]．この手法では，命令発話の音節列とそれに対応したロボットの行動（意味 ID）を基に新たな言い回しを学習していく．まず，人が発話した内容が与えられた言語知識に従う発話（既知発話）か，未知発話かのどちらであるか判別する．そして未知発話と判別された場合には，「違う言い回しで言ってください」とロボットが発話し，既知発話で再度命令を与えるように促す．そして次に与えられた発話が既知発話と判別された場合に，先ほど未知とされた音節列に対して，正しい意味を対応づける．これにより，命令と教示というようにプロトコルを分けること無く，命令発話のみから新たな言い回しを学習していくことができる．しかし，先行研究では移動のみが可能なロボットを用いていたため，想定される発話の種類が限られており，十分な実験ができなかった．

そこで，本研究では，具体的なサービスロボットの一例として，キズ検査ロボット[呉 15]に提案手法を適用する．ロボットは，検査のための移動，位置決め，検査のパラメータ設定，検査結果の表示などが可能であり，これを作業員は音声またはタッチパネルで指示する．キズ検査ロボットと作業員の対話という具体的なシチュレーションを設定することで，自然で多様なインタラクションデータを収集することが可能となる．また，実用性の観点から，ロボットに与えるべき機能や事前知識を決定できるという利点もある．本稿では，音声入力機能および語彙拡張機能を追加したキズ検査ロボットについて紹介するとともに，タッチパネルで直接的に指示可能な命令について，語彙拡張の実験をした結果について述べる．

連絡先： 藤本智也，名古屋工業大学大学院工学研究科，fujimoto@ume.mta.nitech.ac.jp

2. キズ検査ロボット

図 1 キズ検査ロボット

本研究で用いるロボットは，車両側面のキズを検査することを目的に開発されたロボットであり，検査のための LED 直線照明とカメラ，動作制御用のタッチパネル，PC2台を搭載している（図1）．検査の際には，対象に照明光を照射し，その反射光をカメラで撮影する．キズが生じている部分は周辺と法線方向が異なるため，直接反射が得られず黒く浮かび上がる．そこで，画像の局所領域で分散を求めることでキズを強調し検出する[村瀬 13]．

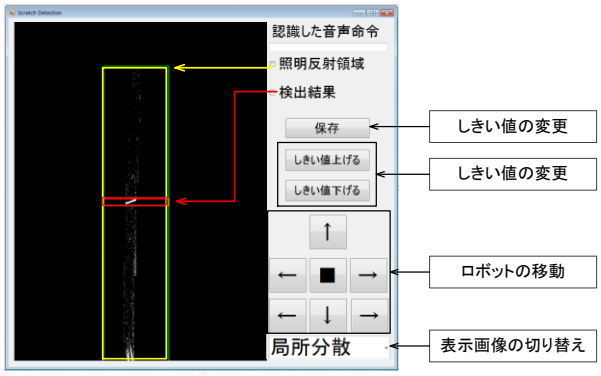


図 2 タッチパネルの操作画面

本ロボットの運用場面では，検査対象までの移動や，検査開始姿勢の調整，カメラパラメータの調整を，検査員によるマニュアル操作で行う必要がある．操作で用いるタッチパネルの画面を図 2 に示す．画面上には撮影画像や画像処理結果の画像をリアルタイムに表示することが可能であり，検査員はこの画像を確認しながら，ロボットの姿勢や各種パラメータを調整する．

本研究では，タッチパネル操作だけでなく，音声でも入力を可能とするために，ロボットに音声理解機能を追加する．音声認識にはオープンソースの音声認識エンジンである Julius[1] を用いた．Julius は，単語 n-gram 言語モデルを用いた大語彙連続音声認識が可能であるが，本研究ではキズ検査という小規模なタスクであり，ロボットへの命令を高精度に認識することが求められるため，記述文法による音声認識を採用する．記述文法音声認識では，ロボットが受理可能な発話を，単語辞書と構文制約（これを文法と呼ぶ）で定義する．本稿では，各単語に意味 ID を付与し，意味 ID の列を文法として構文制約に登録する．そして各意味 ID 列に対し，ロボットの応答を定義する．例えば，「元画像見せて」という発話に対し，「画面に表示する画像を元画像に切り替える」という処理を応答として定義する場合には，まず，「元画像」と「見せて」の読み「げんがぞう」，「みせて」と意味 ID「IMG」，「SHOW」の対応を単語辞書に登録する．次に，文法として“IMG SHOW”を構文制約に登録し，その意味 ID 列に対応する応答処理を「画面に表示する画像を元画像に切り替える」と定義する．

3. 語彙拡張手法

3.1 提案手法の概要

従来の音声対話システムでは，予め決められた単語や文法通りの発話でなければ認識・応答することができない．しかし，一般的な人間同士の対話では，意味は同じだが異なる言い表し方（本稿ではこれを「言い回し」と呼ぶ）が多く用いられ，それら全てを事前にシステムに与えることは困難である．そこで，本稿では登録されていない未知の音韻系列からなる単語と，その意味 ID を人とのインタラクションを通してロボットに学習させる．

提案手法は以下に示すような流れで行われる．

1. ユーザの発話を音声認識
2. 認識した音声が既知の発話か未知の発話かを判定
3. 既知発話の音声に認識した意味 ID 列を割り当て
4. 未知の発話の場合は「タブレットで命令してください」と発話し，ボタン操作から得られた意味 ID 列を割り当てる．
5. 教師なし形態素解析と Vitabi アルゴリズムによる語彙獲得

[1]使用バージョン：dictation-kit-v4.0-win, http://julius.sourceforge.jp/index.php

3.2 発話の音声認識

未知の発話を単語辞書に登録するために発話の音節列を得る必要がある．そこで，音声認識の際には，前章で述べた命令発話を理解するための文法認識だけでなく，日本語音節が登録された単語辞書と，任意の音節列を受理可能な構文制約を用いた音節認識を同時に実行する．

3.3 発話の判別

ユーザからの発話が入力された際にそれが事前に定義されている発話（既知発話）であるか，それとも未知の発話（未知発話）であるかを判別する必要がある．そこで，文法認識結果の尤度と音節認識結果の尤度の比を用いて，未知発話の判定を行う．既知発話の場合，両方の尤度は近い値となり，未知発話であれば音節認識結果の尤度に比べて，文法認識結果の尤度は大きく低下する．本稿では，尤度比に閾値を設けることで，未知発話の判定を行う．

3.4 意味 ID の付与

未知の発話と判別された際にはロボットは命令を理解できないため，タッチパネル押して命令するよう求める．ユーザは表示画像を元画像に切り替えるボタン操作で命令を与える．ロボットは前に発話された命令が，今操作されたボタンと同じ意味 ID 列を示すとみなし，未知の発話に対し意味 ID 列を付与する．

3.5 語彙獲得

複数回の発話データを蓄積，オフラインで学習を行う．形態素解析によって分けられた単語を，単語同士結合することで意味 ID に対応した一単語とする．

命令を複数回行った後，ロボットは聞き取った発話を辞書なし形態素解析により分節し，単語を学習する．語彙の拡張は発話が一定数集められたのちに提案手法を用いて実行される．また，形態素解析には，G. Neubig らの教師なし形態素解析の手法を用いた．[Neubig 12]．そして，形態素を結合して単語を生成するために Vitabi アルゴリズムを用いる．

たとえば，「原画像表示」という発話が“えんがぞうじょおおじ”と認識された場合，形態素解析により“え/んがぞ/ひょ/お/お/じ”と分けられる．そして，それぞれの形態素について，文法に登録されている意味 ID 列のどれに近いかを，Vitabi を用いて計算する．次に Vitabi アルゴリズムにより，計算された確率にを用いて形態素が結合されて”えんがぞ/ひょおおじ”となる．このとき“えんがぞ”に意味 ID“IMG”，“ひょおおじ”に“SHOW”が割り当てられる．

4. 実験

4.1 実験条件

Julius の単語辞書は，既存の大語彙が登録された辞書を用いず，表 1 の単語とそれに対応する意味 ID を登録する．また，文法として表 2 に示す日本語音節，意味 ID 列を登録する．そして 1-best 認識によって結果を得る．マイクには，オーディオテクニカ社の AE6100 を使用した．また，USB 接続オーディオインターフェースとして ROLAND 社の QUAD-CAPTURE UA-55 を用いてマイクと PC の接続をした．本実験では話者1名が実際に発話した音声データを与え，提案手法による語彙の拡張を行う．そして，語彙の拡張を行う場合の単語辞書と行わない場合の単語辞書により，同一の評価データを認識する．二つの認識結果を比較することで提案手法を評価する．この時，命令それぞれに対して 2 種類の未知語を含む文章をそれぞれ 10 回ずつと既

知の発話を 10 回ずつ，計 390 発話をロボットに与えた． また，形態素解析には，latticelm[2] を使用した．latticelm のパラメータは初期設定のままとした．

前章でも触れたように，人の発話が既知発話であるか，それとも未知発話であるかを，尤度比に閾値を定めて判別する必要がある．その閾値を決定するために，既知発話の尤度比と未知発話の尤度比の間に閾値を設定した時のグラフが図 3 に示した．これを見ると，未知発話と既知発話の境界は尤度比が 0.983 である．そこで今回の実験における尤度比の閾値は 0.983 と定めた．この値を用いて以降の実験を実施する．

評価方法としては 2-fold cross validation を用いる．その際には，全発話数のうちユーザが意図した発話の意味と実行された処理が一致した割合を認識成功率として表す．

表 1　命令認識用単語辞書

	単語	読み	意味 ID
名詞	元画像	げんがぞう	IMG
	二値画像	にちがぞう	BINARY
	局所分散	きょくしょぶんさん	LOCAL
	検出結果	けんしゅつけっか	RESULT
	反射領域	はんしゃりょういき	RANGE
	閾値	しきいち	TH
動詞	見せて	みせて	SHOW
	保存	ほぞん	SAVE
	消して	けして	OFF
	上げて	あげて	UP
	下げて	さげて	DOWN
	止まって	とまって	STOP

表 2 初期知識の文法と意味 ID 列

初期文法	意味 ID 列
元画像　見せて	IMG SHOW
元画像　保存	IMG SAVE
二値画像　見せて	BINARY SHOW
二値画像　保存	BINARY SAVE
局所分散　見せて	LOCAL SHOW
局所分散　保存	LOCAL SAVE
検出結果　見せて	RESULT SHOW
検出結果　消して	RESULT OFF
反射領域　見せて	RANGE SHOW
反射領域　消して	RANGE OFF
閾値　上げて	TH UP
閾値　下げて	TH DOWN
止まって	STOP

表 3　追加する単語

意味 ID	追加する単語	処理
SHOW	ひょうじ	表示する
	きりかえて	
SAVE	さつえい	保存する
	とって	
OFF	さくじょ	消す
	なくして	
UP	あっぷ	上げる
	たかく	
DOWN	だうん	下げる
	ひくく	
STOP	すとっぷ	停止する
	とまれ	

[2]使用バージョン：latticelm 0.4,
http://www.phontron.com/latticelm/index-ja.html

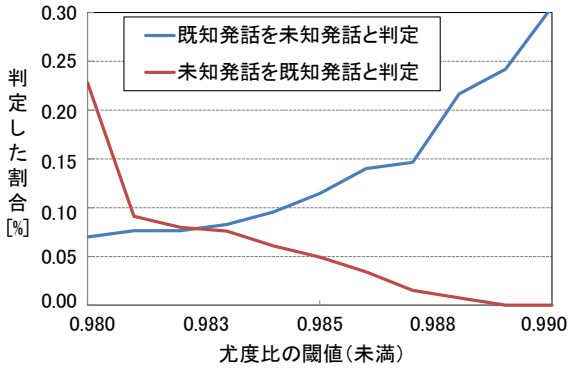


図 3 閾値による誤判定率の推移

4.2 実験結果と考察

まず，表 4 に対話を通じて拡張された単語数を示す．表を確認すると，単語により新たに登録された単語数が大きく異なることが分かる．この中で，「SHOW」は他の意味 ID と比べて発話される回数が多いため，登録語数も多いと考えられる．また，「IMG」と「BINARY」の二つを比べると「BINARY」の登録単語数が多い．これは「IMG」について「BINARY」と比較して，既知発話と認識される割合が高いことが理由として考えられる．次に拡張前と拡張後における認識成功率の変化を図 4 に示す．さらに，拡張前において未知発話の判定をした場合としない場合，学習後，の 3 種類に分けてそれぞれの割合を図 5 に示す．さらに表 7 に学習することで得られた単語辞書の一部を示す．

図 5(b)から確認できるように，単語辞書を拡張した結果，未知発話の認識成功率が上昇している．このことから提案手法の優位性が確認できる．また，図 5(a)より，既知発話についても提案手法の認識成功率が高いことが確認できる．これについて考察すると語彙の拡張をすることでユーザの発話に適応し尤度が上昇し，既知発話と判別されて正解数が増加したと考えられる．

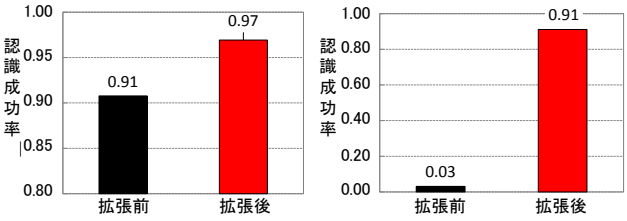
また表 5 に示すように，ユーザが発話した「とまって」は Julius により「おあって」と音節認識される．初期単語辞書による認識では「おあって」は「とまって」と正しく認識されたが，尤度比が低いため未知発話と判定され，語彙の拡張が行われる．そして拡張後に同じ発話を行うと「おあって」と認識され，「とまって」の時より尤度比が高くなり既知の発話と判定される．このことから提案手法は単語辞書を拡張することで未知発話の認識を可能とし，なおかつ，既知発話を用いた際にユーザの発話のクセを吸収した認識をしているということが考えられる．

また，図 5 の(a),(c)からわかるように未知発話の判定（リジェクト）を行うことで，元々認識間違いをしている割合よりもリジェクトされる割合が上昇している．しかし，(a)の認識誤りの割合が 2%存在するのに(c)では認識誤りは 0%となり，認識誤りは減少している．本システムは語彙の拡張を目的としているため，誤認識をしてしまうより，未知発話と認識した方が最終的な認識に良い影響を与えるため有効である．また，(b),(d)を見ると(d)では多くの発話をリジェクトしているがそれにより認識誤りの割合を減少させることができているため，同様にこちらも有効である． また図 5(a),(e)から既知語の認識において，拡張後の単語辞書では誤りが増加している．増加した誤り例を表 6 に示す．

この原因として，拡張した語彙により，発話の認識間違いが増加していることが考えられる．増加した誤りの例を参照すると，拡張前では正しい認識が行われていたが，単語辞書の拡張をすることにより，誤った単語認識が行われる例が見られた．このように短い単語に対して意味 ID が学習されている際にこのような間違いが見られる．これに対する処理が必要だと思われる．

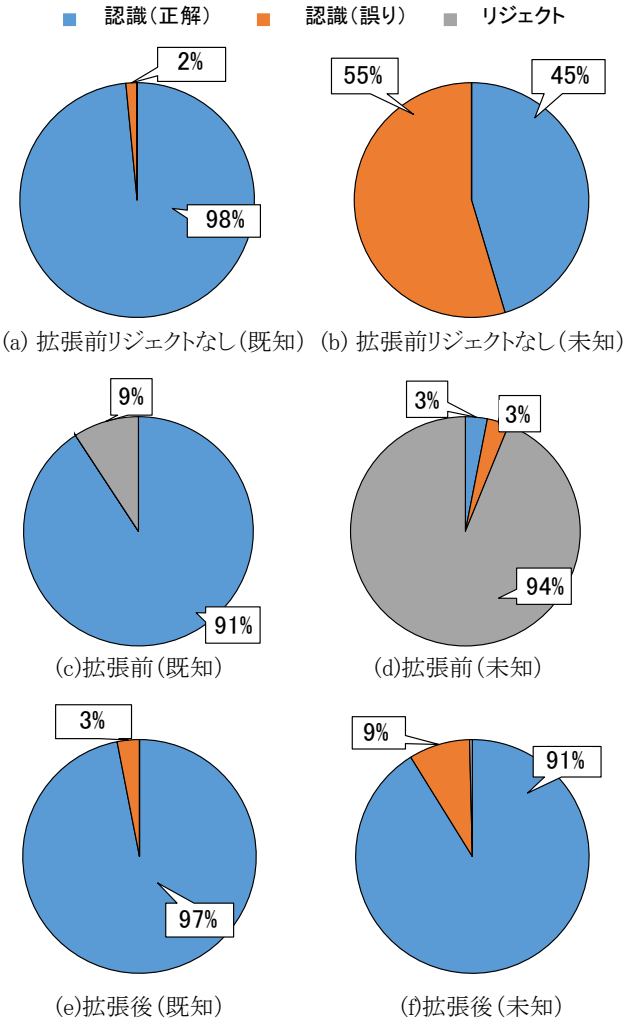
表 4 登録された単語数

単語	意味 ID	拡張された単語数(語)
元画像	IMG	7
二値画像	BINARY	17
局所分散	LOCAL	23
検出結果	RESULT	31
反射領域	RANGE	33
閾値	TH	35
見せて	SHOW	59
保存	SAVE	39
消して	OFF	19
上げて	UP	6
下げて	DOWN	6
止まって	STOP	13



(a)既知発話　(b)未知発話

図 4 学習の前後における認識成功率の比較



(a) 拡張前リジェクトなし(既知)　(b) 拡張前リジェクトなし(未知)

(c)拡張前(既知)　(d)拡張前(未知)

(e)拡張後(既知)　(f)拡張後(未知)

図 5　認識結果の内訳

表 5　既知の発話が未知の発話として登録される例

ユーザの発話	音節認識	登録単語	与えられる意味 ID
とまって	おあって	おあって	STOP

表 6　増加した誤りの例

ユーザの発話	認識結果	意味 ID
元画像見せて	け　がぞおびせて	RESULT SHOW
閾値下げて	すきいちぇさげて　う	TH UP

表 7 学習した単語(一部抜粋)

IMG	BINARY	LOCAL
えんがぞ	いち	おくしょぶさ
げ	いちぇが	きょくしょぶさ
けん	いちぇがぞ	ぞくしょぶさ
RESULT	RANGE	TH
ぎへしつけた	はしゃにょおいき	しきしぇさげ
けしゅつけっか	はっしゃでおういき	すきちっらう
きへしょつけか	はっしゃれのいき	しきちざめっ
SHOW	SAVE	OFF
おみせて	おとって	なくして
がぞうひょおち	さつえ	さくじょう
がぞおびせって	ぞふぉぞう	さくじょお
UP	DOWN	STOP
あぷ	おん	すとうぷ
う	か	すとっぷ
く	くく	すとっぷう

5. おわりに

本稿では，キズ検査という具体的な場面の下で事前知識を持つロボットがユーザとのインタラクションを通して語彙を拡張していく手法について述べた．提案手法は，様々な語彙を拡張できることを確認した．さらに語彙の拡張を行うことにより，認識できる言葉が増加することを確認し，さらに発話者のクセに適応した認識も確認できた．しかし，単語の拡張を行うことで，非常に短い単語に対して意味が割り当てられ，それにより認識間違いが増加することも確認された．解決のためには，登録する単語の判別を行う手法が必要である．今後は判別手法の研究や別のシチュレーションにおける実験等を行いたい．

参考文献

[岩橋　03] 岩橋直人: ロボットによる言語獲得　－言語処理の新しいパラダイムを目指して，人工知能学会誌，Vol.18，No.1，pp.49-58，2003.

[中村　08]中村友昭，長井隆行，岩橋直人，ロボットによる物体のマルチモーダルカテゴリゼーション，電子情報通信学会和文論文誌，Vol.J91-D，No.10，pp.2507-2518，2008.

[田口　10] 田口亮，岩橋直人，船越孝太郎，中野幹生，能勢隆，新田恒雄: 統計的モデル選択に基づいた連続音声からの語彙学習，人工知能学会論文誌，Vol. 25，No. 4，pp. 549-559，2010.

[谷口　14] 谷口彰，吉崎陽紀，稲邑哲也，谷口忠大: 自己位置と場所概念の同時推定に関する研究，システム制御情報学会論文誌

[渡邊　15] 渡邊祐太，田口亮，服部公央亮，保黒政大，梅崎太造: サービスロボットによるインタラクションを用いた語意の拡張，人工知能学会全国大会，2015.

[呉　15] 呉比，田口亮，服部公央亮，保黒政大，梅崎太造: 自動外観検査ロボットのための検査位置推定手法の検討，計測自動制御学会システム・情報部門　学術講演会，2015.

[村瀬　13] 村瀬智光，Yu Quiyue，田口亮，保黒政大，梅崎太造: カーシェアリングにおける自動車の自動外観検査システム,精密工学会誌 80(12)，1102-1108，2014.

[Neubig　12] Graham Neubig, Masato Mimura, and Tatsuya Kawahara: Bayesuan learning of a language model from continuous speech, IEICE TRANSACTIONS on Information and Systems, Vol. 95, No.2, pp. 614-625, 2012.