情報処理学会論文誌 Vol. 52 No. 12 3398–3407 (Dec. 2011)

音声対話システムにおける簡略表現認識のための自動語彙拡張

森 信介[†1] 駒谷 和範[†2] 勝丸 真樹[†3]
尾形 哲也[†3] 奥乃 博[†3]

音声対話システムにおいて，ユーザはしばしば名称の一部を省略して「簡略表現」として発話する．その結果，音声認識誤りを招く．我々は，簡略表現を元の表現の単語列の一部の単語を省略した表現として定義し，簡略表現を確率とともに自動生成して音声認識辞書に自動追加する．簡略表現の取得には，日本語では複合語を分割する必要があるが，形態素解析器のみの分割では固有名詞は必ずしも正確に分割できない．さらに，多くの簡略表現を辞書に追加すると，語彙サイズの増加により音声認識精度が劣化する．我々は，これらの問題の解決方法として，単語分割や発音推定のシステムの自動分野適応と元の表現との平仮名編集距離で表した音韻的類似度に基づく簡略表現候補の取捨選択を提案する．提案手法によって生成した簡略表現候補を推定確率とともに語彙に自動追加した結果，既存辞書内の語のみを含む発話に対する文字正解精度と簡略表現を含む発話の文字正解精度の両方が向上した．この結果から，提案手法により人手による簡略表現の追加を上回る音声認識精度が実現できることを示した．

Automatic Vocabulary Expansion for Abbreviation Recognition in Spoken Dialogue Systems

Shinsuke Mori,[†1] Kazunori Komatani,[†2]
Masaki Katsumaru,[†3] Tetsuya Ogata[†3]
and Hiroshi G. Okuno[†3]

Users of spoken dialogue systems often abbreviate long expressions. This causes errors in automatic speech recognition (ASR). To cope with this problem, we propose a method for generating abbreviation candidates with appropriate probabilities and adding to the ASR dictionary. Two issues arise during this vocabulary expansion. The first one is a low accuracy in word segmentation and pronunciation estimation for the expressions containing proper nouns. The second is an ASR degradation caused by inappropriate abbreviation candidates added to the vocabulary. As an solution, we propose an automatic adaptation of a word segmenter and a pronunciation estimator and a filter for the candidates according to the phonetic distance to the original expressions. The experimental results showed that our method improved the ASR accuracies for both the utterances containing abbreviated words and those containing words in the original expressions. This indicates that our method is capable of realizing a better ASR accuracy than a manual dictionary expansion.

1. はじめに

音声対話システムが持つ音声認識辞書内の語彙の表現と，ユーザが実際に発話する表現とはしばしば一致しない．特に，ユーザは名前の長い対象に言及する際にその一部を省略して発話する傾向がある[1)]．これはシステムに関する知識のない初心者によくあてはまる．従来，このような発話への対処として，システム運用中に得られたユーザの発話に基づき，開発者が人手で語彙を追加してきた．このような人手によるメンテナンスは時間やコストがかかる．そのうえ，運用中に認識されなかった発話に基づき語彙を追加するため，システム運用の初期段階ではそれらの表現を含むユーザの発話をシステムは認識できない．

本論文では，初期システムの辞書が与えられた際に，ユーザが発話しうる簡略表現をそれらの辞書に自動追加することを提案する．システム初期の段階の音声認識辞書の簡略化の対象となる項目を「元の表現」と呼び，元の表現の一部を省略した語を「簡略表現」と呼ぶ．本論文では，音声認識辞書内の元の表現の一部を省略することで簡略表現候補を自動生成し，適切な生成確率を推定したうえで，辞書に追加することを提案する．このような自動追加をシステムの運用開始以前に行っておけば，運用中の人手での追加が不要となるだけでなく，システム運用の初期から簡略表現を認識可能となる．

初期の音声認識辞書に含まれる元の表現から簡略表現を生成し，音声認識辞書に加えるうえで，主に以下の 3 つの課題がある．

(1) 簡略表現生成のための固有名詞の分割
元の表現は複合語であるが，日本語のような膠着言語では分かち書きしないため，省

†1 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University
†2 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University
†3 京都大学情報学研究科
Graduate School of Infomatics, Kyoto University

略する箇所を選択するために単語への分割が必要となる．単語への分割は一般的に形態素解析器や自動単語分割器によって行われる．しかし，音声対話システムにおける地名や商品名などのタスク遂行に必要な内容語はドメインに依存した固有名詞のため，一般的な形態素解析器や自動単語分割器では分割を誤る箇所が多くなる．

(2)　発音の推定

単語分割の結果得られた単語列の一部の単語を省略することで省略表現候補の単語列が得られるが，この発音を適切に推定する必要がある．これは，省略表現候補の単語列の個々の単語の発音を文脈に応じて推定するか，元の表現に付与されている発音とその単語分割結果とのアラインメントをとることで実現できる．いずれの方法でも，単語分割の場合と同様に，既存の発音推定システムの精度が高くないことが問題となる．

(3)　簡略表現追加に起因する認識率劣化の抑制

生成した簡略表現を単純に音声認識辞書に追加すると，語彙増加にともない音声認識において混同される候補が増加し，音声認識率が劣化する．これは，既存辞書の単語と音韻的に類似した簡略表現を追加した場合に特に問題となる．

我々は，これらの課題に対し，元の表現とその発音から単語分割と発音推定のシステムを適応し，それを用いて元の表現を単語に分割したうえで部分単語系列を生成し，適応された発音推定システムを用いてその発音を推定することを提案する．さらに，元の表現との発音の距離に基づき，生成した簡略表現候補の一部を棄却することを提案する．これらの手法により，語彙増加にともなう認識率の劣化を抑制しながら，ユーザが発する簡略表現を適切な生成確率とともに語彙に自動追加することが可能となる．

2. 一般公開したシステムにおける追加語彙の調査

京都市バス運行情報案内システム（以下，バス案内システムと呼ぶ）[2] は，電話を通じてバスの運行情報をリアルタイムで知ることができる音声対話システムである．出発地や目的地をバス停名や最寄り施設名（以下，施設名と呼ぶ）で指定する．音声認識用の言語モデルでは，施設名は 1 つのクラスとして表現されている．実際の施設名は，このクラスから等確率で生成されるとしている．この設計により，施設名の追加や削除に容易に対応することが可能となる．

このシステムの 2002 年 5 月から 2007 年 2 月までの 5 年間の運用中に実際に追加された語彙を調査することで，システムとユーザとの語彙の不一致を予備的に調べた．利用コール数は 15,290 コールであった．その期間中，開発者は，認識誤りとなった発話のログに基づいて音声認識辞書を更新していた[⋆1]．その結果，音声認識辞書に追加された施設名は 309 語であった．それらの分類と具体例を**表 1** に示す．表 1 において，既存辞書とは，バス案内システム運用開始時の音声認識辞書である．(a)，(b)，(c)，(d) は，既存の表現を意図した別の表現である．(e) はまったくの別名や，新しい施設名である．

表 1　京都市バス運行情報案内システムの運用を通じて追加された語彙
Table 1　Entries added to the Kyoto bus information system through operation.

運用中に追加された語彙の種類	追加数　（割合）
(a) 既存の語の簡略表現 例：蒔絵町（既存辞書の語：吉祥院蒔絵町）	242（78.3%）
(b) 既存の語内の単語を入れ替えた表現 例：烏丸四条（既存辞書の語：四条烏丸）	24　（7.8%）
(c) 既存の語に単語を加えた表現 例：阪急桂駅西口（既存辞書の語：桂駅西口）	5　（1.6%）
(d) その他の既存の語の別表現 例：三宝院（既存辞書の語：三宝寺）	12　（3.9%）
(e) 新停留所名・新施設名の追加 例：丹波口駅（新施設名）	26　（8.4%）

表 1 より (e) の新たな施設名の追加は 8.4%であり，(a)，(b)，(c)，(d) の別表現の追加と比較して少ない．これよりバス停に関連した施設名はシステム設計の段階で十分に用意されていたといえる．一方で，既存の表現を意図した別表現の追加は全体の 91.6%を占める．これは実際のユーザの多様な表現を，システム設計段階で開発者が想定できていなかったことを意味する．なかでも，元の施設名（以下，元の表現と呼ぶ）の一部を省略することで得られる簡略表現の追加が 78.3%を占めていることから，実際のユーザは元の表現の一部を省略して発話することが多いといえる．京都市バス運行情報案内システムを利用した初心者ユーザ（システム使用回数が 1 回のユーザ）の発話を実際に計数したところ，全 1,494 発話中 150 発話に，初期の音声認識辞書にはない簡略表現が含まれていた．以上の調査結果から，あらかじめ簡略表現を自動的に生成し，適切な確率を付与して語彙に追加しておくことで，運用開始時からより高い利便性を有する音声対話システムが実現可能であると考えられる．

⋆1 開発者は，ユーザの発話した語彙を選択的に音声認識辞書に加えている．というのも，バス案内システムは，混合主導型の対話システムであり，連続音声認識を行うため，孤立単語認識の場合ほど言語制約は強くない．そのため，湧き出し誤りの原因となりそうな短い語などは追加していない．

3. 関 連 研 究

自然言語処理の研究として，一般的な話者が共有する簡略表現を，コーパスや Web 文書から取得する研究[3),4)] が行われている．音声対話システムで用いる音声認識辞書内の固有名詞は非常にドメイン依存性が高く，Web 上での出現頻度は概して低い．音声対話システムにおける音声認識語彙の簡略表現を，Web などの外部の知識から取得するのは困難である．実際に，表 1 中の (a) 簡略表現の Web における出現頻度は，概して非常に低かった．たとえば，バス停名「新林公団住宅前」の簡略表現である「新林住宅前」や「新林公団前」の文字列としての Web 上でのヒット件数はそれぞれ 0 件と 5 件であった[⋆1]．これは意味を持たない語の断片のヒット件数よりずっと少ない．また，上述の先行研究では，簡略表現の発音推定に関しては何も述べていないが，音声対話システムへの簡略表現の追加には，発音の推定が不可欠である．しかしながら，ドメイン依存性により，一般的な学習データから構築された発音推定システム[5)] では，高い精度で発音を推定することが困難である．

多くの簡略表現は，元の表現を単語列に分解して，その部分単語系列として得られると考えられる[6)]．この研究では，単語境界を明示する英語を対象としているので，簡略表現生成時に複合語の分割の問題は生じない．日本語を対象とした研究として，榎ら[7)] は，既存の形態素解析器で単語分割を行っている．しかしながら，解析対象がドメイン特有の固有名詞を多く含むので，自動単語分割システムの精度向上や単語分割誤りへの対策が必要であると考えられる．

さらに，簡略表現を含めた音声認識精度の向上のためには，簡略表現の追加時には，語彙の拡大に起因する認識精度の低下を抑制する必要がある．Jan ら[6)] は生成簡略表現を辞書に単純に加えており，語彙拡大による悪影響を考慮していない．榎らは，簡略表現の生成規則や Web 上での頻度によって，簡略表現の絞り込みを行い，音声認識率の低下を防いでいる[7)]．これらの先行研究に対し，勝丸らは，形態素解析の結果を自動修正し，発音距離に応じて減少する関数を用いて生成確率を調整することにより人手による簡略表現追加に近い精度を報告している[8)]．この方法では，たとえば，形態素解析器による分割誤りに対処する方法として，各文字間に対して前向きと後ろ向きの文字 n-gram 確率を計算し閾値と比較するなど，アドホックな点が散見される．これに対して，本論文では，単語境界確率推定器を学習コーパスから学習し，単語境界確率[11)] を推定するなど，よりシステマティクな手法を提案する．

⋆1 2011 年 4 月に http://www.google.co.jp/で検索.

4. 複合語からの簡略表現の自動生成法

本研究で目的としているバス案内システムの音声対話においては，施設名が発音とともに与えられている[⋆2]．利用者はしばしば施設名の簡略表現を発話する．施設名の元の表現は複合語であり，ほとんどの簡略表現は，元の表現の一部を省略することによって生成される．その過程で以下の問題が生じる．

(1) 元の表現からの簡略表現の生成方法とその場合の確率の推定
　簡略表現は，元の表現の部分文字列である．音声認識精度の向上のためには，ユーザが発話しそうな簡略表現を網羅的に生成し，その度合いに応じて適切な確率を付与する必要がある．

(2) 分野特有の複合語に対する単語分割と発音推定
　簡略表現は，元の表現の単語とその発音を保持している傾向があると考えられる．このため，元の表現の単語の認定と発音推定が重要である．しかしながら，一般に，元の表現は分野特有の複合語であり，既存の自動単語分割器や発音推定器の精度が低い．

(3) 既存の語彙に近い発音を付与されているなどの不適切な簡略表現による認識精度の低下
　簡略表現は，網羅性を重視して多めに生成する必要がある．一方でユーザは，別の元の表現の発音に近くなる簡略表現を避けられると考えられる．したがって，このような簡略表現を音声認識辞書に追加することは，認識精度の低下を招くと考えられる．

本章では，これらの問題を解決し，簡略表現を含む音声対話における音声認識精度の向上を実現する方法を提案する．以下では，提案手法の各処理について詳述する（**図 1** 参照）．

4.1 複合語の部分単語系列の生成

多くの単語分割基準では，施設名は複合語である．ある複合語から簡略表現の候補となる文字列を生成するために，自動単語分割器を用いて複合語を単語列に自動的に分解し，簡

⋆2 実際には，多少の作業を要したと推測される．具体的には，バス停名（http://www.city.kyoto.jp/kotsu/busdia/hyperdia/mnukana.htm，2011 年 4 月確認）には仮名表記があるが，発音とは少し異なる．また，施設名（http://www.city.kyoto.jp/kotsu/busdia/hyperdia/mnufac.htm，2011 年 4 月確認）には仮名表記も発音もない．

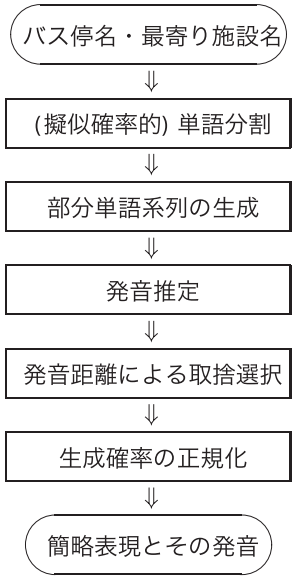

図 1　簡略表現自動生成の流れ
Fig. 1　Flow of the abbreviation candidate generation.

略表現をその部分単語列として生成する．このとき，ある単語は文脈によらず確率 $p_{d,0}$ で消去されるとする．この結果，ある複合語 $w_1w_2\cdots w_k$ の簡略表現は，採否を表すビット列 $b_1b_2\cdots b_k$ と等価になる．対応するビットが 1 の場合を採用とすると，すべての単語を採用する $b_1b_2\cdots b_k = 11\cdots 1$ の場合に元の複合語と同じ文字列になり，すべての単語を棄却する $b_1b_2\cdots b_k = 00\cdots 0$ の場合に空文字列になる．このように，全ビットが 1 あるいは 0 の場合は簡略表現とならないので，これを除外した場合の簡略表現の生成確率は，これらの確率の補正も加味して，以下のようになる．

$$P(b_1b_2\cdots b_k) = \frac{1}{1-(p_{d,0}^k + p_{d,1}^k)}\prod_{i=1}^{k} p_{d,b_i} = \frac{1}{1-(p_{d,0}^k + p_{d,1}^k)} p_{d,0}^{k-l} p_{d,1}^{l}$$

ここで，$p_{d,1} = 1 - p_{d,0}$ であり $l = \sum_1^k b_i$（簡略表現の単語数）である．たとえば，表 1 の (a) の例の単語分割結果 (w_1, w_2, w_3, w_4) = (吉祥, 院, 蒔絵, 町) の簡略表現として「蒔絵町」（$k = 4$, $l = 2$, $b_1b_2\cdots b_k = 0011$）が生成される確率は，

$$P(0011) = \frac{1}{1-(p_{d,0}^4 + p_{d,1}^4)} p_{d,0}p_{d,0}p_{d,1}p_{d,1} = \frac{1}{1-(p_{d,0}^4 + p_{d,1}^4)} p_{d,0}^2 p_{d,1}^2$$

となる．なお，ごく稀に異なるビット列から同一の文字列が生成されることがある．これは，同じ文字を複数含む複合語のある種の単語分割結果に対して起こる．その場合には各ビット列による生成確率の合計を文字列の生成確率とする．

得られた部分単語系列の各単語に対して文脈を考慮した発音推定を行い，各単語の発音を順に連接することで簡略表現の発音とする．「蒔絵 町」の例では，発音推定の結果「蒔絵/まきえ 町/ちょー」が得られるので，発音を「まきえ ちょー」とする．

4.2　自動単語分割と発音推定の適応

施設名は，地名などの固有名詞を多く含む．このため，既存の形態素解析器[5),9)]では，正しく単語分割や発音推定がなされない．この問題を軽減するために，元の表現とその発音に対して，以下のような処理を行う．

(1)　元の表現と発音の組を集合 X に格納する．
　　例）X = {JR 京都駅/じぇーあーるきょーとえき, 烏丸御池/からすまおいけ, 烏丸下立売/からすましもだちうり, ···}

(2)　単語境界確率推定器 `WS` と発音確率推定器 `PE`[5)] を単語分割済みかつ発音付与済みの一般分野のコーパス C_g と辞書 D_g から学習する．

(3)　集合 X の要素に対して以下の処理を行う．集合 X が空であれば終了する．
　(a)　元の表現を `WS` で単語分割する．結果を $w_1w_2\cdots w_k$ とする．
　　例）JR 京都 駅, 烏丸 御池, 烏丸 下立売
　(b)　各単語の発音を `PE` で列挙する．単語 w_i の発音集合を $\boldsymbol{\mathcal{H}}_i \in \mathcal{H}^+$ とする．ここで $\mathcal{H}^+$ は，発音のアルファベット $\mathcal{H}$ の正閉包である．
　　例）JR/じぇーあーる 京都/きょーと 駅/えき,
　　烏丸/からすまる 御池/{おいけ, みいけ, おち, みち},
　　烏丸/からすまる 下立売/{したたう, しもたう, ···}
　(c)　元の表現に付与された発音 $\boldsymbol{h} \in \mathcal{H}^+$ が満たす条件に応じて以下の処理を行う．
　　if　$\boldsymbol{h} \in \boldsymbol{\mathcal{H}}_1\boldsymbol{\mathcal{H}}_2\cdots\boldsymbol{\mathcal{H}}_k$（元の表現の発音が各単語の発音列に分解可能）
　　条件を満たす単語と発音の組の列 $\langle w_1, h_1\rangle, \langle w_2, h_2\rangle, \cdots, \langle w_k, h_k\rangle$ を適応分野のコーパス C_t（初期値は空）に追加し，元の表現を集合 X から消去する．ここで $h_1 \in \boldsymbol{\mathcal{H}}_1, h_2 \in \boldsymbol{\mathcal{H}}_2, \cdots, h_k \in \boldsymbol{\mathcal{H}}_k$ であり，複数の分解が可能な場合は前の単語から最長一致で探索を行った結果最初に得られる分解

とする．

例）C_t = {JR/じぇーあーる 京都/きょーと 駅/えき}

else if $\boldsymbol{h} \in \boldsymbol{\mathcal{H}}_1\boldsymbol{\mathcal{H}}_2\cdots\boldsymbol{\mathcal{H}}_{i-1}\mathcal{H}^+\boldsymbol{\mathcal{H}}_{i+1}\boldsymbol{\mathcal{H}}_{i+2},\cdots,\boldsymbol{\mathcal{H}}_k$, $1 \leq \exists i \leq k$（ある 1 つの単語の発音を任意とすることで元の表現の発音が各単語の発音列に分解可能）

条件を満たす単語と発音の組の列 $\langle w_1,h_1\rangle,\langle w_2,h_2\rangle,\cdots,\langle w_k,h_k\rangle$ を適応分野のコーパス C_t に追加し，元の表現を未処理の集合 X から消去する．ここで $h_1 \in \boldsymbol{\mathcal{H}}_1, h_2 \in \boldsymbol{\mathcal{H}}_2,\cdots,h_{i-1} \in \boldsymbol{\mathcal{H}}_{i-1}, h_i \in \mathcal{H}^+, h_{i+1} \in \boldsymbol{\mathcal{H}}_{i+1}, h_{i+2} \in \boldsymbol{\mathcal{H}}_{i+2},\cdots,h_k \in \boldsymbol{\mathcal{H}}_k$ であり，複数の分解が可能な場合は，小さい i から上記と同じ処理を行う．

例）C_t = {JR/じぇーあーる 京都/きょーと 駅/えき,
烏丸/からすま 御池/おいけ}

この処理で，未知語の発音がある程度の確度で獲得される．上の例では，「烏丸」に対する発音が「からすま」であることが分かる．

else　何もしない（元の表現と発音の組を未処理の集合 X に残す）．

例）X = {烏丸下立売/からすましもだちうり, ⋯}

(d)　処理 (c) の結果，未処理の集合 X が不変の場合，すべてを C_t に移す．

(4)　WS と PE を一般分野のコーパス C_g と辞書 D_g および適応分野のコーパス C_t から学習する．例では，「烏丸/からすま 御池/おいけ」が学習コーパスに加わることにより，次の繰返しでは，「烏丸下立売/からすましもだちうり」$\in X$ が正しく分割され発音が付与されることが期待される．

(5)　処理 (3) に戻る．

この処理により，元の表現を自動単語分割した場合の各単語に対する発音を高い精度で推定可能な発音推定システムが構築される．

上述の手続きで単語分割の精度も多少向上することが期待されるが，与えられた言語資源（元の表現と発音の組）は，単語境界推定のための情報はほとんど含まないので，精度向上の程度は非常に限定的であろう．しかしながら，自動単語分割の誤りは，後述する単語の取捨選択において，その判断が単語境界をまたいで異なる場合にのみ悪影響を及ぼす．すなわち，誤って単語境界と推定された箇所があっても，その前後の単語の両方を簡略表現においても採用（棄却）することになれば悪影響はない．したがって，自動単語分割の精度は，発音推定ほど重要ではない．また，自動単語分割の誤りは，後述する後段の擬似確率的単語分割によってある程度緩和できる．

4.3　擬似確率的単語分割

施設名に対する高い自動単語分割の精度は期待できない．本論文では，分野適応において自動単語分割の誤りの影響を軽減する擬似確率的単語分割[10] を用いることを提案する．ある文字列 $x_1x_2\cdots x_n$ に対する擬似確率的単語分割の結果は以下のようにして得られる．

(1)　文字 x_ix_{i+1} の間に単語境界がある確率 P_i を推定する[11]．

(2)　各文字間に対して 0 以上 1 未満の乱数 r を発生させ，結果を単語境界確率 P_i と比較し，その大小関係に応じて単語境界がない（$r < P_i$ の場合）またはある（$P_i \leq r$ の場合）とし，文字列に対する単語列を得る．

確率的単語分割に対する近似誤差を低減するために，m 回の試行を行い m 個の単語列を生成する．このときの m を倍率と呼ぶ．各単語列の生成確率は，単語列の頻度を $f(\boldsymbol{w})$ として，$f(\boldsymbol{w})/m$ とする．

たとえば，バス停名「東北園町」を倍率 16 で疑似確率的単語分割した結果，f(東北 園町) = 13，f(東北 園 町) = 1，f(東 北園 町) = 2 となったとすると，正しい単語分割結果の生成確率は P(東 北園 町) = 2/16 となる．このように，決定的な単語分割の結果が「東北 園町」であったとしても，一定の確率が正しい単語分割結果に割り当てられ，自動単語分割の誤りの影響が軽減されることが期待される．

4.4　生成簡略表現の追加に起因する音声認識率低下の抑制

音声認識の発音辞書への簡略表現候補の追加のために，決定的あるいは疑似確率的単語分割の結果から生成された部分単語列に対して発音推定を行う．発音推定のための分類器は，一般分野のコーパス C_g と辞書 D_g から，あるいはそれらと適応分野のコーパス C_t から学習する．簡略表現候補の発音は，それを構成する各単語の発音の連接として得られる．

このようにして得られる簡略表現候補の発音が，既存の施設名の発音に非常に近くなり，誤認識の原因となることが考えられる．この問題を解決するために，発音の距離による簡略表現候補の取捨選択を行う．これは以下のように行われる．

まず，ある表現 x の発音の平仮名表記を $h(x)$ とし，2 つの表現 x_i と x_j の発音の距離を平仮名表記の編集距離[12] と定義する[⋆1]．これを平仮名編集距離と呼び，$d(h(x_i),h(x_j))$ と表記する．次に，ある簡略表現候補 t について，その生成元であった施設名 s 以外に，平仮

⋆1 同一母音の平仮名間の削除や挿入のペナルティを下げることや，国際音声記号音韻を発音の表記に用いることで多少の改善が実現できる可能性がある．

名編集距離が閾値 D_{min} 以下になる元の表現 s' がある場合には，この簡略表現候補を削除する．すなわち以下の条件を満たす場合には，簡略表現候補 t を削除する．

$d(h(s'), h(t)) \leq D_{min}$　ここで $\exists s' \in \mathcal{S} - \{s\} \wedge t \in \mathcal{T}_s$

ただし，$\mathcal{T}_s$ は元の表現 s から生成された簡略表現の集合を表す．たとえば，$s =$ “知恩院” の簡略表現候補として $t =$ “知恩” が生成され，その発音が $h(t) =$ “ちおん” と推定されるが，別の元の表現 $s' =$ “祇園” の発音が $h(s') =$ “ぎおん” であり，それらの平仮名編集距離が $d(h(s'), h(t)) = 1$ と小さいため候補 $t =$ “知恩” は削除される．最後に，削除された候補の確率を削除されなかった候補に比例配分（正規化）する．

4.5　各語彙項目の生成確率

2 章で述べたように，バス案内システムの音声認識の言語モデルでは，施設名は，1 つのクラス c として表現されている．実際の施設名は，このクラスから等確率で生成されるとしている．すなわち，施設名の集合を $\mathcal{S}$ とすると，施設名クラスから施設名 s が生成される確率は $P(s|c) = 1/|\mathcal{S}|$ としている．

簡略表現も元の表現と同じ文脈で認識されるように，簡略表現クラス c_a と元の表現クラス c_f を導入し，それぞれが施設名クラスから確率 $P(c_a|c)$ および $P(c_f|c)$ で生成されるとする．ここで，$P(c_f|c) + P(c_a|c) = 1$ である．元のバス案内システムと同様に，元の表現 s は，以下のように元の表現クラスから等確率で生成されるとする．

$$P(s|c) = P(s|c_f, c)P(c_f|c) = \frac{1}{|\mathcal{S}|}P(c_f|c)$$

簡略表現候補 t は，本節で説明した確率を加味して，以下の確率で生成される．

$$P(t|c) = P(t|c_a, c)P(c_a|c) = \frac{1}{|\mathcal{S}|}P(w_1 w_2 \cdots w_k|\boldsymbol{x})P(b_1 b_2 \cdots b_k)P(c_a|c)$$

$$P(w_1 w_2 \cdots w_k|\boldsymbol{x}) = \begin{cases} 1 & \text{決定的単語分割の場合} \\ \frac{f(\boldsymbol{w})}{m} & \text{疑似確率的単語分割の場合} \end{cases}$$

ここで $w_1 w_2 \cdots w_k$ は元の表現の文字列 $\boldsymbol{x}$ の単語分割結果である．

発音の平仮名編集距離による取捨選択を行う場合は，棄却された簡略表現の確率を比例配分（正規化）する．

5. 評　　価

実際の音声対話システムの音声認識辞書から簡略表現を生成し，これを辞書に加えて音声認識の実験を行った．本章では，この結果を提示し，提案手法を評価する．

5.1　評価対象発話データ

評価には，京都市バス運行情報案内システムにおいて収集した実際のユーザの発話データを用いた．本研究では，システムの持つ語彙を知らないユーザに対する性能をみるため，システムの使用回数が 1 回だけの初心者ユーザの発話を集めた．ここで無音やタスクの進行に関係のない発話は除いた．その結果，183 名の 1,494 発話を得た．1,494 発話のうち 150 発話に，既存辞書の語を簡略化した表現が 69 種類 161 個含まれていた．また，既存の音声認識辞書で認識できる語彙のみを含む発話は 1,142 発話あった．残りの 202 発話は，簡略表現を含む発話でも既存辞書で認識できる語彙を含む発話でもなかった．たとえば，「乗り換えはだめなんですね」といった発話である．これらの発話は，以下のいずれの実験条件でも語彙外となるため，正しく認識することはできない．

5.2　比 較 手 法

京都市バス運行情報案内システムの初期の音声認識辞書（語彙サイズ 1,658）から，内容語である施設名（1,471 個）に対して簡略表現を自動生成した．次に示すベースラインと手動追加と 4 つの自動簡略表現生成手法を実験的に比較した．

BL:　既存辞書（ベースライン；文献 2) 参照）
　各施設名の生成確率を等確率としている．すなわち $P(s|c) = 1/|\mathcal{S}| = 1/1471$ である．

Man:　発話された 69 種類の追加（手動追加）
　施設名のクラスの生成確率を $P(c_f|c) = 1471/(1471 + 69)$ とし，簡略表現クラスが生成される確率を $P(c_a|c) = 69/(1471 + 69)$ とした．これにより，施設名とその簡略表現が等確率（$1/(1471+69)$）で出現することになる．

M1:　BL+部分単語系列の生成（4.1 節参照）
　単語が省略される確率を $p_{d,0} = 1/4$ とした．これは，$1/2$ より小さく 0 より大きくするべきでかつそれ以上の情報がないので，この区間の一様分布の平均値とした．実際の省略表現を収集し，実例から推定することで多少のさらなる精度向上が可能であると考えられる．なお，$P(c_f|c)$ と $P(c_a|c)$ は Man と同じとした．

M2:　M1+単語分割と発音推定の適応（4.2 節参照）

M3:　M2+確率的単語分割（4.3 節参照）
　文献 10) に従って，倍率を $m = 16$ とした．

表 2　単語分割と発音推定の精度
Table 2　Accuracy of word segmentation and pronunciation estimation.

言語資源	単語分割（F 値）	発音推定（F 値）
BCCWJ, UniDic	83.58	90.05
BCCWJ, UniDic, 自動生成した適応分野のコーパス	83.85	99.55

M4:　M3+発音の平仮名編集距離による取捨選択（4.4 節参照）
発音の平仮名編集距離の閾値を $D_{min} = 1$ とした．この値は直感的に妥当であろう．すなわち，結果として，他の施設名とまったく同じ発音になるか平仮名表記で 1 文字のみ異なる発音の簡略表現が棄却される．

単語境界確率推定と発音確率推定には，KyTea（version 0.2.1）[5] を用いた．パラメータの学習に用いた言語資源は，現代日本語書き言葉均衡コーパス 2009 年度版のモニタ公開データ[13] のコアデータ（BCCWJ）と UniDic（version 1.3.12）[14] である．単語分割と発音推定の適応を行う場合には，BL に含まれる 1,471 個の施設名とその発音を 4.2 節の手法で処理した結果も学習コーパスとして利用する．

5.3　自動単語分割と発音推定の分野適応

まず，自動単語分割と発音推定の分野適応を単独で評価する．1,471 個の施設名を人手で単語分割した．なお，正しい発音は，初期システムに含まれており，改めて作業する必要はない．これをテストデータとして，4.2 節で説明した方法で適応を行う場合と行わない場合の精度を計算した．**表 2** がその結果である．この表から，単語分割の精度向上はわずかであるが，発音推定の精度向上は著しいことが分かる．これは，元の施設名から得られる単語境界に関する情報は，せいぜい両端に単語境界があるという程度であるが，発音は適切に付与されていることから，当然の結果である．

5.4　生成簡略表現の評価

それぞれの分割手法ごとの追加語数とユーザが実際に発話した 69 種類の簡略表現に対する再現率と適合率を**表 3** に示す．

M1 の結果から，元の表現を単語分割し，部分単語系列を生成することで必要な簡略表現の 6 割程度が生成できていることが分かる．次に，M2 の結果から，単語分割と発音推定を分野適応することで，さらに 4 つの必要な簡略表現を生成できていることが分かる．分野適応の効果は主に発音推定精度の向上であると考えられる（表 2 参照）．実際に，この 4 つを調査した結果，表記はいずれも適切であるが，M1 の発音は誤りであり，M2 の発音が人手による追加と一致した（**表 4** 参照）．M3 の結果から，発話された 69 種類に対する確率的単語分割による発音辞書のカバー率の向上はなかったといえる．M4 の結果から，発話された 2 種類を含む 867 の簡略表現候補が削除されたことが分かる．

表 3　簡略表現候補生成手法による再現数
Table 3　Recall numbers of candidate generation methods.

ID	簡略表現候補生成手法	追加簡略表現数	再現数
BL	既存辞書（ベースライン）	0	0
Man	発話された 69 種類の追加（手動）	69	69
M1	BL+部分単語系列生成	8,733	41
M2	M1+単語分割と発音推定の適応	8,606	45
M3	M2+確率的単語分割	17,479	45
M4	M3+発音の距離による取捨選択	16,612	43

表 4　手法 M1 と M2 の候補集合の差
Table 4　Difference in the candidate sets of M1 and M2.

文字列	M1 での推定発音	M2 での推定発音	Man の発音
上終町	あおちょー	かみはてちょー	かみはてちょー
本町	ほんちょー	もとちょー	もとちょー
花園	かえん	はなぞの	はなぞの
烏丸	からすまる	からすま	からすま

5.5　発話データの音声認識実験

各自動簡略表現生成手法を用いて，発話データの音声認識を行った．実験条件は以下のとおりである．言語モデルには，統計的言語モデルを用いた．このモデルは，京都市バス運行情報案内システムの音声認識用文法から内容語（バス停・施設名，系統番号）をクラス化した文法を作り，その文法から生成したすべての文パターンを用いて作成した．言語モデルの作成には CMU Toolkit [15] を用いた．単語 n-gram モデル学習後，クラスごとに内容語を割り当てた．系統番号のクラスから個々の系統番号が生成される確率は等確率とした．施設名のクラスから個々の表現が生成される確率は，各手法において説明したとおりである．音響モデルは，電話用 2,000 状態 16 混合トライフォンモデルであり，音声認識エンジンには Julius [16] を用いた．各手法による，簡略表現を含む 150 発話，既存辞書の語彙内の 1,142 発話，全体の 1,494 発話の内容語に対する文字正解精度を**表 5** に示す．なお，単語正解精度ではなく，文字正解精度を評価基準としている理由は，簡略表現の単語分割が明確ではなく，単語正解精度を計算することが困難であることである．

まず，BL と Man の比較から，人手によって発話された語彙を追加することは非常に効果

表 5　文字正解精度 [%]
Table 5　Character accuracy [%].

条件	簡略表現を含む発話（150 発話）	既存辞書の語を含む発話（1,142 発話）	全発話（1,494 発話）
BL　既存辞書（ベースライン）	6.2	65.7	48.4
KA　勝丸ら[8] の方法	45.2	66.7	54.0
Man 発話された 69 種類の追加（手動）	45.3	66.7	54.1
M1　BL+部分単語系列生成	39.0	66.2	53.7
M2　M1+単語分割と発音推定の適応	45.3	67.8	55.6
M3　M2+確率的単語分割	43.9	68.1	55.7
M4　M3+発音の距離による取捨選択	46.8	68.1	56.0

があることが分かる．単語分割誤りの緩和や発音の距離による確率の調整を提案している勝丸ら[8] の方法 KA は，アドホックではあるが Man に近い精度を実現している．提案する各処理をすべて行う M4 は KA や Man よりも高い精度となっていることが分かる．表 3 に示したように，M4 は発話された 69 種類のうち 43 種類しか再現できていないが，確率が適切に推定された結果である．すなわち，人手での簡略表現の追加は，内省により表記と発音を適切に選択することはできるが，クラスからの生成確率に関しては適切な値を設定するのは容易ではなく，これも含めて自動生成する提案手法が音声認識精度の向上に有効であるといえる．検定の結果，M4 と KA との精度の差は有意水準 5%で有意であった．KA は，様々な箇所で確率とは異なるアドホックな選択がなされているが，提案手法では，これらを適切に確率にしたことが精度向上の主たる理由であろう．

次に，本論文で提案する自動生成手法の各処理の効果についてである．部分単語系列の追加のみを行う M1 の結果は，BL よりかなり良いが，Man には及ばない．単語分割誤りや発音推定誤りがあっても，かなりの簡略表現は一般的な単語分割システムと発音推定システムの結果の部分単語系列として生成できることが分かる．これらの分野適応を行うことで得られた M2 の結果と表 2 から，発音推定の誤りの影響は大きく，提案する自動分野適応手法が有効であることが分かる．M3 の結果から，疑似確率的単語分割を行うことで，分割誤りの悪影響の緩和によりさらに少し精度が向上することが分かる．表 5 を詳しくみると，簡略表現を含む発話のみの精度が低下し，既存辞書の語を含む発話の精度が向上していることが分かる．このことと表 3 から，確率の推定がより適切になったことが精度向上の要因であるといえる．最後に，発音の平仮名編集距離による取捨選択の効果についてである．表 5 における M3 と M4 の精度は，既存辞書の語を含む発話においては同じであるが，簡略表現を含む発話に対しては，M4 の方が高くなっている．このことから，擬似確率的単語分割により大量に追加された簡略表現候補（表 3 によると 8,873 個）のうち不適切な発音を持つ候補を棄却することで，既存辞書の語を含む発話に対する認識精度を維持しつつ簡略表現を含む発話に対する認識精度を大幅に向上させ，結果として全体の精度がさらに向上していることが分かる．

以上のことから，提案手法の各処理はそれぞれ有効であり，すべてを用いることで発話された簡略表現を人手で追加する方法よりもより高い精度を実現できることが分かる．

6. おわりに

本研究では，音声対話システムにおける音声認識精度の向上を目的として，ユーザ簡略表現に焦点をあて，音声認識辞書の語から簡略表現を自動生成し認識辞書に加える手法について述べた．本研究では，ユーザが発する簡略表現の再現率を高めるために，単語分割結果に対する部分単語系列から簡略表現候補を生成することを提案した．また，分野特有の表現を単語に分割し発音推定をする際に，単語分割や発音推定の自動分野適応を行うとともに，擬似確率的単語分割を用いることを提案した．さらに，不適切な発音を持つ簡略表現の追加による音声認識率の低下を抑えるために，生成した簡略表現候補に対して，既存の音声認識辞書内の語との発音の平仮名編集距離を計算し，その値に応じて候補を棄却することを提案した．

実際の初心者ユーザの発話を用いた評価実験では，部分単語系列を簡略表現候補として追加することで，文字認識精度が 5.3 ポイント向上した．加えて，単語分割や発音推定の自動分野適応を行うことで，文字認識精度が 1.9 ポイント向上した．さらに，擬似確率的単語分割を用いることで，文字認識精度が 0.1 ポイント向上した．最後に，発音の平仮名編集距離による取捨選択により，文字認識精度が 0.3 ポイント向上した．以上のすべてを用いることで，初期のシステムや，人手によって発話された簡略表現を追加したシステムよりも高い認識精度が実現された．以上のことから，提案する簡略表現候補の自動追加手法によって，ユーザの固有名詞を簡略化した発話に対しても，音声認識システムが頑健動作することを示した．

参 考 文 献

1) Zweig, G., Nguyen, P., Ju, Y., Wang, Y., Yu, D. and Acero, A.: The Voice-Rate Dialog System for Consumer Ratings, *Proc. InterSpeech2007*, pp.2713–2716 (2007).

2) Komatani, K., Ueno, S., Kawahara, T. and Okuno, H.G.: User Modeling in Spoken Dialogue Systems for Flexible Guidance Generation, *User Modeling and User-Adapted Interaction*, Vol.15, No.1, pp.169–183 (2005).
3) Park, Y. and Byrd, R.J.: Hybrid Text Mining for Finding Abbreviations and their Definitions, *Conference on Empirical Methods in Natural Language Processing*, pp.126–133 (2001).
4) 酒井浩之，増山　繁：略語とその原型語との対応関係のコーパスからの自動獲得手法の改良，自然言語処理，Vol.12, No.5, pp.207–231 (2005).
5) Neubig, G. and Mori, S.: Word-based Partial Annotation for Efficient Corpus Construction, *Proc. 7th International Conference on Language Resources and Evaluation* (2010).
6) Jan, E.E., Maison, B., Mangu, L. and Zweig, G.: Automatic Construction of Unique Signatures and Confusable Sets for Natural Language Directory Assistance Applications, *Proc. 8th European Conference on Speech Communication and Technology*, pp.1249–1252 (2003).
7) 榎　将功，皇甫美華，大田健紘，柳田益造：日本語における略語自動生成法の検討とその音声インタフェースへの応用，情報処理学会研究報告，SLP69, pp.313–318 (2007).
8) 勝丸真樹，駒谷和範，尾形哲也，奥乃　博：音声対話システムにおける簡略表現認識のための誤認識増加を抑制する自動語彙拡張，情報処理学会研究報告，SLP71, pp.71–76 (2008).
9) 工藤　拓，山本　薫，松本裕治：Conditional Random Fields を用いた日本語形態素解析，情報処理学会研究報告，NL161 (2004).
10) 森　信介，小田裕樹：擬似確率的単語分割コーパスによる言語モデルの改良，自然言語処理，Vol.16, No.5, pp.7–21 (2009).
11) 森　信介，宅間大介，倉田岳人：確率的単語分割コーパスからの単語 N-gram 確率の計算，情報処理学会論文誌，Vol.48, No.2, pp.892–899 (2007).
12) Cormen, T.H., Leiserson, C.E. and Rivest, R.L.: *Introduction to Algorithms*, The MIT Press (1990).
13) 前川喜久雄：代表性を有する大規模日本語書き言葉コーパスの構築，人工知能学会誌，Vol.24, No.5, pp.616–622 (2009).
14) 伝　康晴：多様な目的に適した形態素解析システム用電子化辞書，人工知能学会誌，Vol.24, No.5, pp.640–646 (2009).
15) Clarkson, P.R. and Rosenfeld, R.: Statistical Language Modeling Using the CMU-Cambridge Toolkit, *Proc. 5th European Conference on Speech Communication and Technology*, pp.2707–2710 (1997).
16) Kawahara, T., Lee, A., Takeda, K., Itou, K. and Shikano, K.: Recent Progress of Open-Source LVCSR Engine Julius and Japanese Model Repository, *Proc. 8th International Conference on Speech and Language Processing*, pp.3069–3072 (2004).

(平成 23 年 4 月 11 日受付)
(平成 23 年 9 月 12 日採録)

森　　信介（正会員）

1998 年京都大学大学院工学研究科電子通信工学専攻博士後期課程修了．同年日本アイ・ビー・エム（株）入社．2007 年より京都大学学術情報メディアセンター准教授．京都大学博士（工学）．1997 年情報処理学会山下記念研究賞受賞．2010 年情報処理学会論文賞受賞．2010 年第 58 回電気科学技術奨励賞受賞．言語処理学会会員．

駒谷　和範（正会員）

2002 年京都大学大学院情報学研究科知能情報学専攻博士後期課程修了．京都大学博士（情報学）．同年京都大学大学院情報学研究科助手．2007 年同助教．2010 年より名古屋大学大学院工学研究科准教授．同年より JST さきがけ「情報環境と人」領域研究員兼務．2008 年から 2009 年まで米国カーネギーメロン大学客員研究員．情報処理学会平成 16 年度山下記念研究賞．FIT2002 ヤングリサーチャー賞等を受賞．電子情報通信学会，言語処理学会，人工知能学会，ISCA 各会員．

勝丸　真樹

2008 年京都大学工学部情報学科卒業．2010 年同大学院情報学研究科知能情報学専攻修士課程修了．現在パナソニック株式会社勤務．在学中は音声対話システムの研究に従事．情報処理学会平成 20 年度音声言語処理研究会（SIG-SLP）学生奨励賞受賞．

尾形　哲也（正会員）

1993 年早稲田大学理工学部機械工学科卒業．日本学術振興会特別研究員，早稲田大学理工学部助手，理化学研究所脳科学総合研究センター研究員，京都大学大学院情報学研究科講師を経て，2005 年より同助教授（現，准教授）．博士（工学）．JST さきがけ研究「情報環境と人」領域研究員（5 年）．この間，早稲田大学ヒューマノイド研究所客員准教授，同大学理工学研究所客員准教授，理化学研究所脳科学総合研究センター客員研究員等を兼務．研究分野は人工神経回路モデルおよび人間とロボットのコミュニケーション発達を考えるインタラクション創発システム情報学．日本ロボット学会，日本機械学会，人工知能学会，計測自動制御学会，ヒューマンインタフェース学会，バイオメカニズム学会，IEEE 等会員．

奥乃　　博（正会員）

1972 年東京大学教養学部基礎科学科卒業．日本電信電話公社，NTT，JST，東京理科大学を経て，2001 年より京都大学大学院情報学研究科知能情報学専攻教授．博士（工学）．この間，スタンフォード大学客員研究員，東京大学工学部客員助教授．人工知能，音環境理解，ロボット聴覚，音楽情報処理の研究に従事．1990 年度人工知能学会論文賞，IEA/AIE-2001，2005，2010 最優秀論文賞，IEEE/RSJ IROS-2001，2006 Best Paper Nomination Finalist，IROS-2010 NTF Award for Entertainment Robots and Systems，第 2 回船井情報科学振興賞等受賞．本学会理事，人工知能学会，日本ロボット学会，日本ソフトウェア科学会，ACM，IEEE，AAAI，ASA 等会員．