

音声対話システムにおける 音韻的類似表現の混同を防ぐための確認の自動生成

浜辺 良二 駒谷 和範 尾形 哲也 奥乃 博

京都大学大学院 情報学研究科 知能情報学専攻

hamabe@kuis.kyoto-u.ac.jp

音声対話システムにおいて、聞き取りやすさを考慮したシステム応答側の発話内容や言語表現についてはほとんど研究されていない。本研究では、システムの語彙内の音韻的に類似した単語に対して、それらを聞き分けやすくする新たな単語を付加した確認を自動的に生成する。まず、複数の知識源を利用して、付加する単語の候補を複数取得する。さらに確認の適切性を測る尺度を定義することにより、候補から最適な確認表現を自動的に選択する。京都市バス運行情報システムおよびホテル検索システムの語彙に対して、本手法により確認を生成し評価を行った。被験者5名に対する聴取実験により、生成した確認によって音韻的に類似した単語の聞き分けが改善されることを確認し、本手法の有効性を示した。

Automatic Generation of Confirmation to Prevent Confusion between Phonologically Similar Expressions for Spoken Dialogue Systems

RYOJI HAMABE, KAZUNORI KOMATANI, TETSUYA OGATA and HIROSHI G. OKUNO

Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

In spoken dialogue systems, language expressions of systems' responses in terms of intelligibility have not been investigated. We design and develop an automatic response generation system that gives additional words if there exist phonologically similar words in the system's vocabulary. The system exploits multiple knowledge sources to get a set of candidates for additional expressions. We defined the appropriateness of candidates so as to select the best candidate. The effectiveness of the automatic response generation system was evaluated by applying it to the vocabulary of two existing systems, Kyoto City Bus Information System and Hotel Search System. The experimental result by 5 subjects showed that our method could improve intelligibility between phonologically similar words.

1. はじめに

近年の音声認識・音声合成技術の向上により、音声による対話システムの研究が活発化している¹⁾²⁾³⁾。しかし音声を用いたコミュニケーションでは、周囲のノイズにより、音韻的に類似した単語間の聞き分けが困難となる場合がある。そのため、システムは音声認識誤りを防ぐとともに、ユーザにとって聞き間違いにくい応答を行う必要がある。本研究では確認発話を対象とし、システムの確認表現に含まれる内容語をより聞き分けやすくする単語を付加することで、ユーザの聞き間違いの削減を図る。

本手法では、既存の知識源を用いることでタスクメインに依存せず自動的に、付加する単語を取得し確認表現を生成する。しかし単一の知識源のみから得ら

れた単語が、必ずしも聞き取りやすいものであるとは限らない。そこで複数の知識源を用いて、付加する単語の候補をできるだけ多く取得する。さらにこの際、候補の中から最適なものを選択する方法が必要となる。本研究では付加する単語の聞き分けやすさを定義することにより最適な候補を選択する。

2. ユーザの聞き間違いを考慮した確認応答

音声対話システムにおいて、確認応答から得られる情報はシステムの動作生成に重要な役割を果たしている。しかし、ユーザがシステムの応答を聞き間違える可能性については、これまで考慮されていなかった。本研究では、ユーザの聞き取りやすさに注目した確認応答を生成する。

2.1 システム応答の聞き取りやすさ

音声合成の品質が向上すれば、ユーザの聞き間違いの減少が期待できる。しかし、人間は紛らわしい単語を、強調や韻律の調整を行うことで伝達できるが、あらゆる固有名詞に対して自動的にこれらを柔軟にコントロールすることは困難である。よって音声合成の品質向上とともに、システム発話の言語表現をユーザが聞き取りやすいものにする必要もある。本研究では後者を取り扱う。

人間は他の単語と聞き間違いやすい語を伝達する際に、和文通話表を利用することがある。和文通話表とは「朝日のあ」「いろはのい」のように、日本語の1音を表現するために、その音で始まる簡単な単語を集めたものである。例えば「金閣寺」と「銀閣寺」を区別する場合には「切手の『き』で始まるか?」「ギリシャの『ぎ』で始まるか?」という質問をすることになる。しかしこの方法では同音異義語は区別できない。さらにこの例の場合「ゴールドか?」「シルバーか?」のように英語を用いて確認の方が直感的に分かりやすいと考えられる。

本手法では、複数の知識源から確認表現の候補を取得し、その中から最適なものを選択する。これは前述の例のように、単一の知識源からでは最適な確認表現が得られるとは限らないからである。

2.2 タスクドメインに依存しない確認応答の生成

本研究では前節で述べた確認表現の生成を、対象とするドメインに依存しない方法により行う。このため本手法は、様々なドメインに対して自動的に適用できる。これに対して人手で確認表現を記述する場合には、一つ一つの単語に対して記述しなければならないうえ、単独のドメインにしか適用できない。

本手法内の処理は次のように分けることができ、全てドメインに依存しない。

- 音韻的に紛らわしい語の組の抽出
- 確認表現の候補の取得
- 最適な候補の選択

候補を取得するために用いる知識源もドメインごとに設計する必要はなく、既存のものを用いる。具体的には、一般に用いられている単語辞書・国語辞書・漢字辞書・英和辞書・和英辞書・和文通話表を組み合わせ用いる。

3. 音韻的類似表現に対する確認表現の生成

システムの語彙内に含まれる音韻的に類似した単語(これを音韻的類似表現と定義する)間の聞き分けを容易にする新たな単語(これを付加表現と定義する)を取得し、確認表現を生成する。以下では京都市バス運行

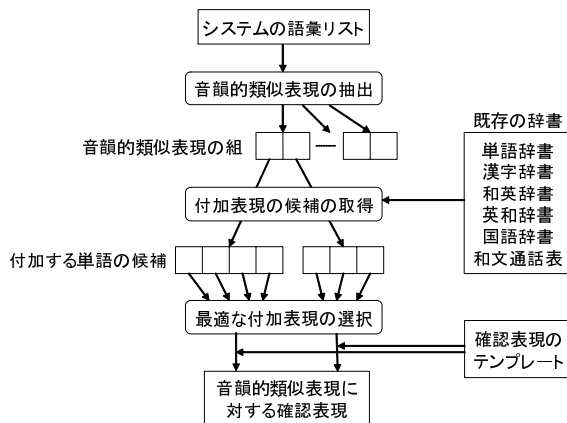


図1 確認表現の生成の流れ

S1:	乗車駅, 降車駅, 系統番号をおっしゃってください
U1:	金閣寺道から
S2:	金閣寺道か銀閣寺道か分からないのですが、 先頭に英語でシルバーの銀の字が付きますか?
U2:	いいえ
S3:	では、先頭に英語でゴールドの金の字が付きますか?
U3:	はい

図2 本手法を適用した対話の例

情報案内システム⁴⁾における例を用いて説明を行う。

付加表現の取得は次の手順で行う。その流れを図1に示す。

- (1) 単語間の音韻的距離を定義し音韻的類似表現の組をシステムの語彙リストから抽出する。
- (2) 音韻的類似表現の組に対し、それらの相違部分を含む単語を複数の知識源からできるだけ多く取得し、付加表現の候補とする。
- (3) 候補の中から最適なものを選択し、応答文のテンプレートに適用して確認表現を生成する。

本手法を適用することにより得られる対話の例を図2に示す。U1でユーザは「金閣寺道」という単語を発話しているが、バスシステムの語彙にはこれと音韻的に類似した「銀閣寺道」という単語が含まれている。従来では「銀閣寺道ですか?」のように確認が行われるが、ここではシステム発話が「銀閣寺道ですか?」「金閣寺道ですか?」のいずれであるかユーザは聞き取りにくい。そこで本手法を適用することでS2のように「ゴールド」「シルバー」という付加表現を加えた確認を生成する。

以下の節で付加表現を取得する手順の詳細を述べる。

3.1 システム語彙からの音韻的類似表現の抽出

まず、単語同士が音韻的にどれくらい類似しているかを測る指標として単語間の音韻的距離を定義する。単語同士が音韻的に類似している、すなわち聞き分けが難しいのは、次のような場合である。

表 1 京都市バス運行情報案内システムの語彙から得られた音韻的類似表現の例

類似表現 1	類似表現 2	p.d.
京都府立盲学校 (キョウトフリツモウガッコウ)	京都府立聾学校 (キョウトフリツロウガッコウ)	0.34
地下鉄・四条駅 (チカテツシヨウエキ)	地下鉄・二条駅 (チカテツニヨウエキ)	0.36
洛西高校前 (ラクサイコウコウマエ)	洛水高校前 (ラクスイコウコウマエ)	0.36
金閣寺道 (キンカクジミチ)	銀閣寺道 (ギンカクジミチ)	0.37

- 単語間で異なる音素が少ない場合
- 単語間で共通する音素が多い場合
- 単語のモーラ数が近い場合

このことから，単語 w, w' 間の音韻的距離 (phonetic distance) $p.d.(w, w')$ を次式のように定義する．

$$p.d.(w, w') = \frac{e.d.(w, w')}{\log(\min(|w|, |w'|) + 1)} \quad (1)$$

$|w|, |w'|$ はそれぞれの単語の音素数を表す． $e.d.(w, w')$ は単語 w, w' を音素列にしたときの編集距離 (edit distance) であり，DP マッチング⁵⁾ により計算する．ただし，編集の際にコストを導入し，モーラ数が変化する場合 2，変化しない場合 1 とする．音素を母音・撥音・促音のクラスと，子音・半母音のクラスに分類すると，モーラ数が変化するのは前者のクラスに属する音素の挿入 (または削除)，および異クラスの音素の置換が行われたときである．

表 1 に，京都市バス運行情報案内システムの語彙の中で音韻的距離の小さかったものの例を示す．

3.2 付加表現の候補の取得

本節では，前節で得られた音韻的類似表現間の確認に用いる付加表現の候補を取得する方法を説明する．本研究では音韻的類似表現間の文字の違いおよび音の違いに着目する．複数の知識源を用いてこれらの違いを含む単語を得て，テンプレートに当てはめることで確認表現を生成する．

まず，音韻的類似表現間の非共通文字 (漢字) および非共通音を抜き出す．以下の説明では，これらをそれぞれ $\langle \text{char} \rangle$, $\langle \text{phone} \rangle$ で表す．また，これらの音韻的類似表現内での位置をそれぞれ $\langle \text{charLoc} \rangle$, $\langle \text{phoneLoc} \rangle$ で表す．位置は，音韻的類似表現間の先頭から非共通文字 (非共通音) の直前までの共通文字列 (共通音) を用いて表す．例えば，表 1 に含まれる「洛西高校前」と「洛水高校前」の例では，

$$\left\{ \begin{array}{l} \langle \text{char} \rangle = \text{「西 (サイ) または「水 (スイ)」} \\ \langle \text{phone} \rangle = \text{「サ または「ス」} \\ \langle \text{charLoc} \rangle = \text{「洛 (ラク) の後」} \\ \langle \text{phoneLoc} \rangle = \text{「ラクの後」} \end{array} \right.$$

となる．ただし，音韻的類似表現間の非共通文字 (非共通音) が先頭 (または最後) にある場合は， $\langle \text{charLoc} \rangle$

表 2 「洛西高校前」と「洛水高校前」に対する確認表現

付加表現を取得する手法	「洛西高校前」に対する確認表現 「洛水高校前」に対する確認表現
熟語を用いる方法 (一部のみ)	洛の後に西の西という字が付きますか？ 洛の後に西暦の西という字が付きますか？ 洛の後に西陣織の西という字が付きますか？ … 洛の後に水曜日の水という字が付きますか？ 洛の後に水族館の水という字が付きますか？ 洛の後に雨水の水という字が付きますか？ …
漢字の訓読みを用いる方法	洛の後に西 (にし) という字が付きますか？ 洛の後に水 (みず) という字が付きますか？
漢字の英訳を用いる方法	洛の後に英語でウエストの西の字が付きますか？ 洛の後に英語でウォーターの水の字が付きますか？
音の違いを強調する方法	ラクの後の音は「さくら」のサですか？ ラクの後の音は「すずめ」のスですか？

= $\langle \text{phoneLoc} \rangle$ = 「先頭」 (または「最後」) とする．

非共通文字または非共通音を含む単語を以下に述べる 4 つの方法で取得し，テンプレートによる確認表現を生成する．この例に対して，それぞれの方法で得られた付加表現を用いた確認表現を表 2 に示す．

3.2.1 漢字を含む熟語を用いる方法

非共通文字を含む熟語を付加表現として用いる．熟語の候補は Juman の JC1⁶⁾ という単語辞書から得る．この辞書の語彙サイズは 35,075 語で，図 3 のように各単語に対して「品詞・活用型・表記・読み」が与えられている．また後述するように各単語に難易度が与えられているため，本研究では JC1 を熟語の知識源として用いた．確認のテンプレートは「 $\langle \text{charLoc} \rangle$ に，(熟語)の $\langle \text{char} \rangle$ の字が付きますか？」とする．

3.2.2 漢字の訓読みを用いる方法

単独で意味を持つ漢字の訓読みを付加表現として用いる．図 3.2.3 のような，漢字に対して音読み・訓読みが与えられているテキストデータを用いて，非共通文字の訓読みを調べる．その訓読みが単語辞書 (JC1) に含まれていれば，単独で意味を持つとし，候補に加える．このような訓読みは全部で 1,909 個あった．確認のテンプレートは「 $\langle \text{charLoc} \rangle$ に，(訓読み)という字が付きますか？」とする．

3.2.3 漢字の英訳を用いる方法

非共通文字が，単独で意味を持つ場合，その英訳を付加表現とする．単独で意味を持つ漢字は和英辞書の見出し語に含まれるものとし，和英辞書のテキストデータから漢字一文字から成る見出し語を取り出し，その訳語を得る．さらに，国語辞書のテキストデータ (カタカナ英語として一般的に用いられる英単語が見出し語に含まれている) から訳語のカナ発音を得て付加表現の候補に加える．これにより図 5 の例を含む 275 種類の漢字・英訳の組 (以下では英訳リストと呼ぶ) を得た．確認のテンプレートは「 $\langle \text{charLoc} \rangle$ に，英語で (英訳) の $\langle \text{char} \rangle$ という字が付きますか？」とする．

品詞	活用型	表記	読み
名詞 普通名詞	-	朝日	あさひ
形容動詞	ナ形容詞	静か	しずか
話す	子音動詞サ行	話す	はなす
形容詞	イ形容詞イ段	難しい	むずかしい

図3 JC1のデータ形式

漢字	音読み	訓読み	漢字	英訳
音	イン, オン	おと, ね	雨	レイン
声	セイ, ショウ	こえ, こわ	黄	イエロー
読	ドク, トク	よ・む	石	ストーン

図4 漢字の音読み・訓読みデータ

図5 英訳リスト

3.2.4 音の違いを強調する方法

和文通話表に含まれる単語を付加表現として、音韻的類似表現間の最も先頭にある非共通音を強調する。確認のテンプレートは、「<phoneLoc>の音は、(和文)の<phone>ですか?」とする。非共通音が濁音もしくは半濁音である場合は、<phone>は非共通音から濁点(半濁点)を取り除いたものとし、テンプレートは、「<phoneLoc>の音は、(和文)の<phone>に濁点(半濁点)ですか?」とする。

特殊な場合として「金閣寺道」と「銀閣寺道」のように、音韻的類似表現間の非共通音が濁点の有無のみの違いである場合は「<phoneLoc>の音は、濁点(半濁点)が付かないですか(付きますか)?」といった、濁点の有無を強調する確認を行う。

3.3 最適な付加表現の選択

得られた付加表現の候補から最適となるものを、付加表現の難易度および聞き分けやすさによって選択する。

3.3.1 付加表現の難易度

単語辞書(JC1)・英和辞書では、辞書に含まれる単語の難易度が5段階に分けられている。難易度の高い単語はユーザが未知である可能性が高く、その場合は何度聞き直しても伝わらないため、対話の進行が妨げられる。よって、単語辞書(JC1)、英訳リストから得られた付加表現の中で難易度が最小のもの以外を候補から削除する。単語辞書・英訳リスト内の各難易度の単語の例と諸元を表3、表4に示す。表4では各難易度に該当する、英語辞書全体に含まれる英単語の数と、和訳が漢字一文字からなる英単語の数を示している。難易度の高い英単語は、漢字一文字では表せないため、ここでは現れていない。なお和文通話表は、簡単な単語を用いることで一音を強調するものであるため、和文通話表から得た付加表現は全て候補に残した。

3.3.2 付加表現の聞き分けやすさ

付加表現の聞き分けやすさを測る尺度として、付加表現 w_i の一意性 u_i 、および相違性 d_i を定義する。

一意性とは、本手法で用いた知識源全体(単語辞書・英訳リスト・和文通話表)の中で、ある付加表現と音

表3 単語辞書(JC1)の難易度の区分

難易度	単語数	熟語の例	訓読みの例
A1(低)	2,337	反対	見る
A2	2,618	主要	裂ける
B	4,892	警告	朽ちる
C	8,989	歓待	憤る
F(高)	16,239	違憲	悪い

表4 英訳リストの難易度の区分

難易度	英語辞書全体での単語数	和訳が漢字1文字の単語数	和訳の例	英単語の例
中学必修(低)	1,000	148	夜	ナイト
高校必修	3,000	110	巢	ネスト
高校学習	3,500	17	岬	ケープ
大学教養	4,000	0	-	-
その他(高)	約90,000	0	-	-

韻的に類似した単語がどれくらい含まれるかを測る指標である。付加表現と類似した単語が知識源中に少ないほうが、ユーザはその付加表現を他の単語と混同しにくく、聞き取りやすいといえる。

付加表現 w_i の一意性 u_i は次のように定義する。本手法で用いた知識源全体に含まれる単語の中で、 w_i と音韻的距離の小さいもの上位 N 個を $\tilde{w}_k (1 \leq k \leq N)$ としたとき、その平均値を u_i とする。

$$u_i = \frac{1}{N} \sum_{k=1}^N p.d.(w_i, \tilde{w}_k) \quad (2)$$

$p.d.(w_i, \tilde{w}_k)$ は3.1節で定義した単語 w_i, \tilde{w}_k 間の音韻的距離である。ここでは $N = 10$ とした。

次に相違性とは、確認対象となっている単語 w が、付加表現によりどれだけ変化したかを測る指標である。直前の対話に出現した単語と音韻的に類似した付加表現を用いた確認を行うことは、ユーザにとって聞き分けの難しい単語を繰り返すことになるので避けるべきである。つまり相違性は、その時点の対話の状況での聞き分けやすさを表す。

付加表現 w_i の相違性 d_i を次のように定義する。確認対象となっている単語 w と付加表現 w_i の、音素数の多い方の任意の部分音素列(長さは少ない方の音素数)と、音素数の少ない方の音素列との音韻的距離の最小値を d_i とする。 $w[k..k']$ は w の k 番目から k' 番目までの部分音素列を表す。

$$d_i = \begin{cases} \min_{k=0,1,\dots,|w|-|w_i|} \{p.d.(w[k+1..k+|w_i|], w_i)\} (|w| \geq |w_i|) \\ \min_{k=0,1,\dots,|w_i|-|w|} \{p.d.(w, w_i[k+1..k+|w|])\} (|w| < |w_i|) \end{cases} \quad (3)$$

難易度が最小の候補の中で、一意性と相違性の重み付き和で定義したスコアが最大となるものを最適な付加表現とする。付加表現 w_i に対するスコア $score(w_i)$ は次式で定義する。 E は難易度最小の単語の集合とする。

$$score(w_i) = W_1 \cdot u_i + W_2 \cdot d_i (w_i \in E) \quad (4)$$

表5 京都市バス運行情報案内システムの語彙に対して得られた付加表現の候補数とその割合

	付加表現の候補の平均数	付加表現を取得できた割合			
		熟語	訓読み	英訳	音の違い
候補全体	51.1 個	96%	67%	17%	100%
難易度最小	3.9 個	44%	46%	13%	100%

表6 「金閣寺道」に対する付加表現の主観評価の例

評価	「金閣寺道」に対する確認表現
A	先頭に金曜日の金の字が付きますか？
A	先頭に英語でゴールドで金の字が付きますか？
B	先頭に金(かね)の字が付きますか？
C	先頭に料金の金の字が付きますか？
C	先頭に金額の金の字が付きますか？

表7 京都市バス運行情報案内システムでの付加表現に対する主観評価の分布

	A	B	C
得られた候補全体に対する評価(個)	113	102	13
選択された付加表現に対する評価(個)	49	9	0

予備実験の結果, $W_1 = 4, W_2 = 1$ とした.

4. 評価実験

4.1 京都市バス運行情報案内システムへの実装

京都市バス運行情報案内システムは, ユーザの指定するバスが乗車する停留所のいくつ手前まで接近しているかを知らせるものである⁴⁾. 語彙サイズは1,574語で, 語彙には「京都駅前」「祇園」といったバスの停留所名や「206番」「南11」といったバスの系統番号が含まれる.

京都市バス運行情報案内システムの語彙内で, 音韻的距離が0.6未満の単語の組(計29組)に対して付加表現の候補を取得する. 熟語を用いる手法・漢字の訓読みを用いる手法・漢字の英訳を用いる方法・音の違いを強調する方法のそれぞれで, 少なくとも1つの付加表現の候補が取得できた単語の割合と, これらの方法で得られた付加表現の候補の平均数を表5に示す. ここでは, 候補を難易度が最小のものだけに限定した場合の値も共に載せている. 同音異義語は含まれなかったため, 音の違いに着目する方法では, すべての単語に対して付加表現が取得できている. また, 難易度が最小のものだけに限った場合でも, 複数の知識源を用いていることで, 音の違いに着目した候補以外にも, 平均で2.9個の候補が得られている.

次に, 実際に分かりやすい付加表現が選択されることを確かめた. 得られた各候補に対して, 分かりやすさの評価を, A(分かりやすい), B(どちらともいえない), C(分かりづらい)の3段階で主観的に与えた「金閣寺道」と「銀閣寺道」という音韻的類似表現の組で前者に対する付加表現の候補に評価を与えた場合, 表6のようになった「金曜日」や「ゴールド」は, これと混

表8 ホテル検索システムの語彙に対して得られた付加表現の候補数とその割合

	付加表現の候補の平均数	付加表現を取得できた割合			
		熟語	訓読み	英訳	音の違い
候補全体	51.4 個	90%	52%	16%	96%
難易度最小	3.3 個	40%	48%	15%	96%

表9 ホテル検索システムでの付加表現に対する絶対評価の分布

	A	B	C
得られた候補全体に対する評価(個)	102	60	9
選択された付加表現に対する評価(個)	47	5	0

同しやすい音韻的に類似した単語は少ないと思われるため, 評価をAとした「金(かね)」は, 音素数が少ないためユーザが他の単語と混同する可能性があるため, 評価をBとした「料金」は「両親」と聞き間違える可能性が高く, また「金額」は「金閣寺道」の「金閣」の部分と非常に類似しているため分かりづらいと思われるため, 評価をCとした.

各候補に与えた評価の分布と, スコアが最大となり選択された付加表現に対する評価の分布を表7に示す. これより, 評価の良い付加表現が多く選ばれて, 逆に評価の悪いものは全く選ばれていないことが分かる.

4.2 他ドメインへの適用

本手法をホテル検索システム⁷⁾に適用し, 本手法がドメイン非依存であることを確かめる. ホテル検索システムは, ユーザが指定した条件を満たすホテルを検索し, その情報を提示する. 検索条件には関西地区のホテルの所在地・上限(下限)料金・付帯施設などがある. 語彙サイズは865語で, 語彙には「和歌山県」「宇治市」といった地名や「プール」「レストラン」といった付帯施設名が含まれる. ホテル検索システムの語彙内で音韻的距離が同じく0.6未満の単語は26組あり, その中には「海山町」「美山町(みやまちょう)」という同音異義語の組が含まれた.

これらの単語の組に対して, 少なくとも1つの付加表現の候補が取得できた単語の割合と, これらの方法で得られた付加表現の候補の平均数を表8に示す. 難易度が最小のものについて, 音の違いに着目した候補以外にも, 平均で2.3個の候補が得られている.

各候補に対する主観評価の分布を表9に示す. ホテル検索システムでも, 評価の良い候補が多く選ばれ, 評価の悪い候補は選ばれていない. 以上の結果から, 本手法はタスクドメインに依存せず用いることができるといえる.

4.3 生成された確認の聴取実験

本手法により生成された確認によって, 実際に音韻的類似表現間の聞き分けが容易になったかどうかを評価実験により検証する.

4.3.1 聴取実験の方法

音韻的類似表現の組 A,B に対して、まず、ユーザに A だけをテキストで提示する。その後、音声合成により、付加表現を用いない従来の確認(1)を聞かせて、ユーザに「はい」「いいえ」「わからない」のいずれかの回答を選択してもらおう。ユーザの回答が「わからない」であった場合にのみ、付加表現を用いた確認(2)を聞かせて、同様に回答を選択してもらおう。それぞれの確認は3回まで繰り返し聞くことができたとして、

- (1) 「A か B (B か A) か分からないのですが、X ですか？」
- (2) 「A か B (B か A) か分からないのですが、(X の付加表現を用いた確認)?」

X は、A か B のいずれかで、音韻的類似表現の組ごとにランダムに定める。

本研究室の学生5人を被験者として、京都市バス運行情報案内システム、ホテル検索システム内の音韻的類似表現(計55組)に対して上記の実験を行った。この実験では、周囲に一定の雑音がある状況下でのシステムの使用を想定し、その状況を再現するため、背景雑音を音声合成と同時に被験者に聞かせた。背景雑音には TARGET ENTERTAINMENT 社の新・効果音大全集に収録されている「駅前広場のガヤ」の音を用い、SN比は0(音声合成と背景雑音の大きさは同等)とした。音声合成にはクリエートシステム開発株式会社の synth を用いた。

4.3.2 聴取実験の結果と考察

聴取実験の結果を表10、表11に示す。表10には、従来の確認でユーザ正しく回答できた数、誤った回答をした数、「わからない」と回答した数を示している。表11には、従来の確認で聞き取りが不可能であった単語に対して付加表現を用いた確認を行った場合の結果を示している。

表11によると従来の確認でユーザが聞き取れなかったもののうちの69%が、付加表現を用いた確認により正しく聞き取れるようになっている。例えば「波賀町(はがちょう)」という単語に対して「先頭に波(なみ)という字が付きますか？」という確認を行った場合、被験者5人中全員が正しく回答できた。この結果は本手法の有効性を表している。

一方、最適であるとして選択されたにもかかわらず、被験者の聞き取りが困難である付加表現があった。例えば「養源院(ようげんいん)」という単語に対して、「先頭に養(やしなう)という字が付きますか？」という確認を行った場合、正しく回答できた被験者は5

表10 従来の確認に対する回答の正解数

被験者	正解	誤り	不可
a	21	0	34
b	38	8	9
c	23	1	31
d	25	0	30
e	32	0	23
平均	27.8 (51%)	1.8 (3%)	25.4 (46%)

表11 付加表現を用いた確認に対する回答の正解数

被験者	総数	正解	誤り	不可
a	34	26	0	8
b	9	5	2	2
c	31	14	3	14
d	30	21	0	9
e	23	22	1	0
平均	25.4	17.6 (69%)	1.2 (5%)	6.6 (26%)

人中2人であった。これは「養う」という付加表現が、難易度は低いとされているものの比較的口語的でないため、思い浮かべにくかったためと考えられる。よって、より適した付加表現を選択するためには、意味的な難易度ではなく、口語表現としての頻度も考慮した難易度の定義が必要となる。

このように、付加表現を用いてもなお、ユーザが理解できなかった確認表現が26%あった。原因の1つとして、本手法による確認では、それまでの対話の話題と全く関係のない単語が突然現れるため、ユーザがその単語を予測できないことが挙げられる。このことから、付加表現の難易度に、タスクや文脈との関連性を加える必要があるといえる。

5. まとめ

本稿では、音声対話システムにおける音韻的類似表現間で、ユーザの聞き間違いを防ぐための確認表現を自動的に生成する手法について述べた。本手法により、適切な確認表現を生成できることを、評価実験により示した。さらに、生成された確認によって音韻的類似表現間の聞き分けがしやすくなることを聴取実験により確かめた。

謝辞

JC1の使用を許可して下さった、京都大学大学院情報科学研究科の佐藤理史先生に感謝します。

参考文献

- 1) 植田喜代志, 秋田祥史, 荒木雅弘, 西本卓也, 新美康永. VoiceXMLのマルチモーダル化の検討. 情報処理学会研究報告, 2001-SLP-38-7, 2001.
- 2) 伊藤敏彦, 小暮悟, 中川聖一. 協調的応答を備えた音声対話システムとその評価. 情報処理学会論文誌, 第39巻5号, pp. 1248-1257, 1999.
- 3) 伊東幸宏, 小西達裕, 伊藤敏彦, 桂川景子. ドライブプランニングシステムの自然言語インタフェース. 人工知能学会誌, 第17巻3号, pp. 285-290, 2002.
- 4) 音声ポケロケ. <http://www.lang.astem.or.jp/bus>.
- 5) G. Navarro. A guided tour to approximate string matching. In *ACM Computing Surveys*, Vol. 33(1).
- 6) 佐藤理史. 異表記同語認定のための辞書編纂. 情報処理学会研究報告, 2004-NL-161, pp. 97-104, 2004.
- 7) 駒谷和範, 河原達也. 音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理. 情報処理学会論文誌, 第43巻10号, pp. 3078-3086, 2002.

<http://www.target-ent.com/music.html>

<http://www.createsystem.co.jp/linux.html>