単語 N-gram 言語モデルを用いた音声認識システムにおける未知語・冗長語の処理

甲斐充彦[†] 廣瀬良文[††,☆] 中川聖一[††]

対話音声認識システムや大語彙のディクテーションシステムにおいては，システムの辞書に登録されていない未知語や，間投詞・言い直し・言い淀みなどのユーザの要求に関係のない冗長語の扱いが重要である．このような問題に対処するために，本研究では単語 N-gram 言語モデルを用いた連続音声認識アルゴリズムにおいて，未知語処理を導入してその効果を調べた．未知語処理法として，サブワード単位の音響モデルを用いたサブワード系列デコーダを併用し，これによって未知語候補の生成と検証を行う方法を用いる．この方法は，以前に文脈自由文法を用いたシステムにおいて有効性を確かめている．本論文では，この方法に基づいて，単語 N-gram ベースの認識アルゴリズムに未知語処理を効率的に導入する方法を提案している．音声対話システムのタスクにおいて，未知語や冗長語を含む発話を用いて評価実験を行った結果，意味的な誤りが最大で 48%減少した．また，文脈自由文法に基づく同様なシステムと比較した結果，意味理解精度の向上に効果があることが分かった．さらに，大語彙連続音声認識タスクにおける効果を確かめるため，新聞記事の読み上げ音声を用いた評価実験を行った結果，単語単位での認識精度の改善は小さいが，文レベルでの高い未知語検出性能が示された．

Dealing with Out-of-vocabulary Words and Filled Pauses in Word N-gram Based Speech Recognition System

Atsuhiko Kai,[†] Yoshifumi Hirose[††,☆] and Seiichi Nakagawa[††]

For practical use of spoken dialog systems and dictation systems, it is important to cope with out-of-vocabulary words and filled pauses including the phenomena such as interjection, restart and hesitation. To address these problems, this study tries to use an unknown-word processing (UWP) method for a word N-gram language model based continuous speech recognition system. We investigate an UWP method which employs a subword sequence decoder with subword acoustic models to produce unknown-word hypotheses. This method has been shown to be effective on a small vocabulary task tested with a context-free grammar-based recognition system. This paper proposes an efficient method for incorporating the UWP into a word N-gram language model-based recognition system. We performed a series of experiments to show the effectiveness of the method for spoken dialog tasks and a dictation task. The experimental results show that a semantic accuracy was improved by 48% using the UWP method. Also, in compared with the result of a system using context-free grammar, the word N-gram based system could further improve the semantic accuracy for spontaneous speech. Furthermore, we performed a recognition experiment for a large-vocabulary dictation task. As a result, although only a slight improvement was observed in terms of the word accuracy, the high performance for detecting the existence of unknown-word in an utterance could be achieved.

1. はじめに

近年，大語彙を扱う音声認識システムの研究において，統計的な言語モデルである N-gram と，それを用いた効率的な探索アルゴリズムが提案され，種々のシステムが構築されている．しかし，これらの従来のシステムでは，システムに登録されていない単語（未知語や冗長語）を含む発話の扱いについては，十分に検討されていない．特に，N-gram をベースとした大語

† 豊橋技術科学大学情報処理センター
Computer Center, Toyohashi University of Technology
†† 豊橋技術科学大学情報工学系
Department of Information and Computer Sciences, Toyohashi University of Technology
☆ 現在，奈良先端科学技術大学院大学情報科学研究科
Presently with Graduate School of Information Science, Nara Institute of Science and Technology

彙音声認識システムにおいては，発話に含まれる未知語の割合以上に単語誤り率が増加することが報告されており[1]，その影響は無視できない．また，自然な発話を扱う対話音声認識システムとしての適用を考えると，未知語だけでなく，間投詞・言い直し・言い淀みなどのユーザの要求に関係のない冗長語の扱いが重要である．実際，日本音響学会連続音声データベースの書き起こしテキストの一部を使用した調査結果では間投詞が1文あたり1.126個，言い直しが1文あたり0.145個の割合で含まれていることが分かっている[2]．たとえ，単語認識精度が同じであっても，未知語や冗長語の部分が他の単語に誤認識されれば，未知語・冗長語部分として検出されるよりも理解に悪影響を及ぼす．

このような問題に対して，小・中語彙を対象としたシステムにおいては，未知語や冗長な単語を含む発話を扱うための方法としては，garbageモデルを使用する方法[3]や，音韻連鎖モデルによる未知語処理の方法などが報告されている[4)~6)]．我々も，文脈自由文法（CFG）を言語モデルとした音声認識システムにおいてサブワード単位の音響モデルに基づく未知語処理および，冗長語処理の効果を確かめた[7]．しかし，大語彙音声認識システム，特に，N-gram言語モデルをベースとしたシステムにおいて，未知語処理の検討は十分に行われていない．

本論文では，N-gram言語モデルをベースとした連続音声認識システムにおける未知語処理の適用について述べる．未知語処理の原理は，文献7)と同様に，未知語や冗長語を任意の音節系列としてモデル化する方法を採用した．未知語検出は，最適なサブワード系列の認識結果に基づく未知語仮説のスコアを用いた単純な仮説検定の原理に基づいている．文献7)で提案した未知語処理法は，一般的なone pass探索アルゴリズムに組み込んだ場合，未知語処理に関する計算量の大部分は語彙数に依存せず，サブワードのカテゴリ数nに対して$O(n)$相当のサブワード系列の認識処理の計算量の増加でほぼ実現できる．また，garbageを用いた方法と比べると，本手法は，音響モデルの高精度化にともなって，より確実に未知語の検出精度が改善されることが期待できる[8]．

本研究では，未知語・冗長語を含む発話に対して，N-gramをベースとした連続音声認識システムにおける未知語処理の効果を評価するため，2種類のタスクで実験的な評価を行った．1つは，音声対話システムを想定したタスクに関するユーザ発話を対象としたもので，文脈自由文法をベースとした連続音声認識システムに未知語処理を適用した場合との比較を行った．評価用音声データは，間投詞・冗長語を意図的に挿入して読み上げられたもののほか，実音声対話システムで収録された発話データでも評価を行った．さらに，大語彙音声認識のタスクの評価用として，新聞記事読み上げ音声データベースを用いた．本論文では，これらの評価実験結果を通して，音声対話システムにおけるN-gram言語モデルベースの連続音声認識システムの有効性と，大語彙音声認識における未知語処理の効果を示す．

2. 単語N-gramを用いた大語彙連続音声認識

N-gram言語モデルを用いた連続音声認識のアルゴリズムは，先行する$N-1$単語を文法的な状態と仮定した場合の有限状態文法に基づくone pass探索アルゴリズムとして実現できる．one pass探索アルゴリズムに基づく処理では，各フレームごとに各単語境界と仮定し，言語モデルによる確率の対数値をマッチング終了後の音響累積尤度に加えることを繰り返すことによって次の式を満たす最尤の単語列候補を計算する[9),10)]．

$$P(w^*|y_1^T) = \underset{\{w_1^N\}\{t_1^N\}}{\text{argmax}} \left\{ \sum_{n=1}^{N} \log(P_{\mathrm{A}}(y_{t_{n-1}+1}^{t_n}|w_n)) + weight \cdot \sum_{n=1}^{N} \{\log(P_{\mathrm{L}}(w_n|w_{n-1})) + \Delta \cdot \gamma(w_n)\} \right\}$$

ここで，$P_{\mathrm{A}}(y_{t_{n-1}+1}y_{t_{n-1}+2}\cdots y_{t_n}|w_n)$は観測パターン系列$y_{t_{n-1}+1}y_{t_{n-1}+2}\cdots y_{t_n}$に対する単語$w_n$の音響モデルの尤度，$P_{\mathrm{L}}(w_n|w_{n-1})$は単語$w_{n-1}$の次に単語$w_n$が接続する言語確率，$\Delta$は文長に対するパラメータ，$\gamma(w_n)$は単語$w_n$の音節数である．すなわち，(HMMの音響尤度＋バイグラム言語モデルの尤度＋Viterbiベストスコア＋ビームサーチ)を基本的な認識の枠組みとして各時刻・各状態における累積尤度を計算する．

我々のシステムでは，tree状辞書表現による探索空間の効率的な構成や，音響的な先読みによる仮説の枝刈りの導入[10]，ユニバーサルなtree状辞書表現の探索空間を用いた近似的な探索処理法（1-best探索）[11]の適用などにより，処理の高速化を図っている．

3. N-gramに基づくシステムにおける未知語・冗長語処理

初めに，本論文で用いる未知語および冗長語という

例1 河口湖の近くに，えー，どんな宿泊施設がありますか．
例2 旅館に夕…，食事つきますか．
例3 山中湖ホテル，そのー，泊ま…，宿泊したいんですが．

図1 冗長語を含む発話の例
Fig. 1 Examples of utterances including filled pauses.

語の意味をより明確に定義する．冗長語は，1文程度の発話単位での意味理解においては無意味・冗長と考えられる単語や単語断片のことで，具体的には，間投詞の単語のほか，言い直しや言い淀みなどにともなって発話された語の断片部分のことである．一方，未知語は，連続音声認識システムの語彙に含まれない単語で，冗長語を除くすべての未登録単語のことを表す．本論文で扱う冗長語は，たとえば**図1**に示す発話の例では，「えー」，「夕…」，「そのー」，「泊ま…」，という単語に相当する．

3.1 未知語処理の原理

1章で述べたように，未知語処理の実現方法として，サブワード系列の認識結果を利用する方法をすでに提案している[7]．未知語に限らず，間投詞や言い直しの扱いにも適用され[7]，文法外の発話や未知語の音声入力に対する棄却の方法としての有効性も示されている[8),13)]．ここでは，簡単にその未知語処理の原理を述べる．

サブワード単位の音響モデルをベースとした連続音声認識システムを仮定したとき，任意の語彙単語はサブワード系列として表現できる．そこで，ある語彙単語 w の信頼性 $\hat{P}(\boldsymbol{W}|\boldsymbol{A})$ は，次式によって近似的に推定できる[14),15)]．

$$\hat{P}(\boldsymbol{W}|\boldsymbol{A}) = \frac{P(\boldsymbol{A}|\boldsymbol{W})P(\boldsymbol{W})}{\sum_{\boldsymbol{p}} P(\boldsymbol{A}|\boldsymbol{p})P(\boldsymbol{p})} \quad (1)$$

$$\approx \frac{P(\boldsymbol{A}|\boldsymbol{W})P(\boldsymbol{W})}{\max_{\boldsymbol{p}} P(\boldsymbol{A}|\boldsymbol{p})P(\boldsymbol{p})} \quad (2)$$

$$\approx \frac{P(\boldsymbol{A}|\boldsymbol{W})}{\max_{\boldsymbol{p}} P(\boldsymbol{A}|\boldsymbol{p})} \quad (3)$$

ただし，$\boldsymbol{p}$ は特定の言語で許される音素や音節の系列である．本研究で N-gram 言語モデルを用いる場合は，式(2)に対応する．式(3)は，単語や音素レベルでの確率が1または0の確定的な言語モデルを仮定した場合（たとえば CFG）である．この式で分母はサブワードモデルによる最適音素/音節系列の尤度と考えることができ，$\hat{P}(\boldsymbol{W}|\boldsymbol{A})$ は尤度比の形となる．ここで用いる CFG に対する未知語処理法は，原理的に式(3)の各尤度の利用に基づいたもので，具体的には $P(\boldsymbol{A}|\boldsymbol{p})$ に相当するモデルとして音節単位モデル（HMM）を用いたものである．

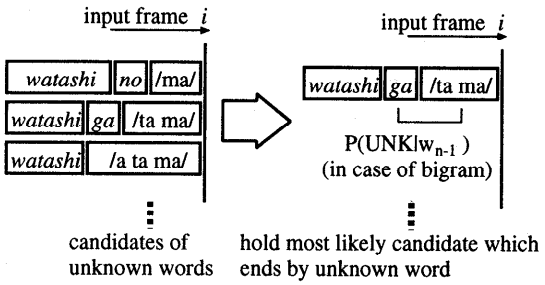


図2 未知語処理の概念図
Fig. 2 Concept of unknown-word processing.

3.2 未知語処理の適用

前述の未知語処理法を，N-gram 言語モデルを用いた連続音声認識システムに適用する．未知語の仮説とスコアを生成するため，one pass Viterbi 探索に基づく音節デコーダを用いる．前述の N-gram 言語モデルを用いた one pass 探索と，音節デコーダの処理は，ともに時間同期的なアルゴリズムで実現される．音節デコーダは，未知語仮説の集合，すなわち各処理フレームで終端固定で始端位置が異なる最適音節系列の部分系列を生成する．それらの未知語の仮説を，すでに計算されている未知語の始端位置までの部分仮説と接続し，未知語で終端する最適な仮説を求める．

以上の処理の概念図を**図2**に示す．i フレーム目の処理では，サブワードの系列長（ここでは音節数とする）が $1 \sim L$（L：未知語候補の音節表記として許容される最大長）の未知語候補の各々について，あらゆる可能な先行単語の累積尤度に未知語候補の尤度および，未知語との bigram 確率を掛け合わせて評価し，それらのスコアの最大値をフレーム i での未知語候補 x の累積スコアとする．

一般に，任意のサブワード系列による未知語仮説を許した場合，未知語の過剰検出が生じる．その改善策として，未知語仮説に関してある種の先見的な言語知識の制約を用いることが考えられる．ここでは，上述の未知語処理法に適用する場合に容易でかつ効果が高い方法として，1つの未知語を構成する音節数を制限（後述の評価実験では長さを2～10音節に制限）する方法を用いる[7]．なお，未知語仮説を最適音節系列としているため，その尤度は他のどの単語よりも高くなる．そこで，未知語の各候補については，音節数に比例したペナルティを与える．

ところで，図2に示すように，上述のアルゴリズムでは，言語モデルにおいて未知語が1つの単語（例では“UNK”）として扱われているものと仮定する．しか

し，未知語カテゴリ UNK の N-gram 確率 $p(UNK|w)$ は，統計量を求めるときに用いた学習コーパス中の未知語の割合に依存するため，未知語の割合や種類が多い場合には，個々の登録単語の N-gram 確率よりも高めに推定されるという問題がある．そこで，後述の評価実験結果で示されるように，実際に音声認識システムにおいて未知語の N-gram 確率を適用する際には，未知語の種類数を考慮するようなペナルティスコアを与えた方がよい[12]．

3.3 冗長語に対する未知語処理の適用

間投詞は音声認識システムにおいては一般に冗長語（不要語）と見なされ，意味理解に用いられることはほとんどない．少数の間投詞の出現頻度が多く，上位5個で全体の90%をカバーする[2]．しかし，これらを単語辞書に登録しても，実際の対話で観測される間投詞の単語の種類はかなり多いうえに発音が曖昧な場合が多い．また，言い直しや，言い淀みなどに対しても同様に対処する必要がある．以前の我々の研究において，冗長語に対処するために未知語処理を適用した評価実験において，頻度の高い10個の間投詞をシステムの辞書に登録した場合と同等の結果を得ている．そこで，本研究では，N-gram 言語モデルベースの音声認識システムへの適用の効果を検討する．

冗長語に関して前述の未知語処理を適用する場合，基本的には，未知語の扱いと同様に実現できる☆．しかし，冗長語に対してあらかじめ N-gram 言語モデルの統計量を求めるためには，自然な発話による音声対話の書き起こしテキストを大量に必要とし，一般に困難である．また，冗長語の出現位置に関する知見によれば[2]，文節間において比較的自由に出現しうることから，冗長語に関して連接制約（N-gram）を適用したとしても，制約が有効に働くとは考えられない．そこで，N-gram の学習時には冗長語をあらかじめ除外して冗長語の存在を無視するとともに，認識時においても冗長語の前後の単語を含むコンテキストに対する言語的なスコアの計算においては，冗長語の存在を無視する．すなわち，$P(w_n|UNK, w_{n-1}, w_{n-2}) = P(w_n|w_{n-1}, w_{n-2})$．

そこで，これと等価な処理は，概念的に**図3**に示す方法で実現できる．図3において，フレーム i で終わる冗長語の仮説は複数あり，それぞれに関して冗長語候補の境界で先行単語と接続し，接続して得られた尤度を先行単語の尤度として使用する．この例では，フレーム i で終わる冗長語を含む候補は，“が”と“の”に後続する場合である．アルゴリズムとしては，冗長語候補の先行する単語 n の累積尤度と冗長語候補の尤度とペナルティを掛け合わせたものがフレーム i での単語 n の尤度よりも大きければ，単語 n に冗長語が付属しているものとして処理する．なお，冗長語の処理の場合にも，未知語の場合と同様に，過剰検出を防ぐために冗長語の仮説に対してペナルティスコアを与える．

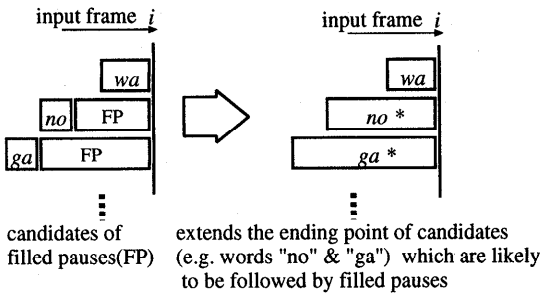


図3 冗長語処理の概念図

Fig. 3 Concept of unknown-word processing for filled pauses.

4. 評価実験

4.1 評価基準

4.1.1 言語モデルの評価基準

言語モデルに対する複雑性の尺度としては，パープレキシティを用いる[17]．本研究で言語モデルの構築に使用した CMU SLM Toolkit[19]では語彙に含まれないものはすべて1つの未知語のカテゴリにまとめられ，語彙に含まれる形態素と等価に未知語のカテゴリは扱われる．そのため語彙サイズのセットが小さいほど（カバー率が小さいほど），パープレキシティは小さくなるということになり好ましくない．そこで評価テキスト中に出現した未知語の種類 m と，未知語の出現回数 n_u を用いてパープレキシティを補正する[20]．補正パープレキシティは

$$APP = (P(w_1 \ldots w_n) m^{-n_u})^{-\frac{1}{n}}$$

で与えられる．これは，複数の未知語はそれぞれ等確率に生じると仮定して，補正したものである．

4.1.2 音声認識の評価基準

音声認識の性能評価として，単語正解率および，単語正解精度を用いた．以下に定義を示す．

単語正解率 = 100 − 置換率 − 脱落率

☆ 厳密には，言い直しや言い淀みのような現象の前後にはポーズが現れることが多いため，さらにポーズ情報を扱うことによってそれらを同定する手掛かりとなる．したがって，ポーズの情報を扱うことでさらに検出精度の改善が期待できるが，本研究ではあらかじめ処理対象とする発話からポーズを削除する方法をとり，冗長語処理においてはポーズを扱わない．

単語正解精度 = 100 − 置換率 − 挿入率 − 脱落率
ただし，正解文および認識結果において，未知語は 1 つのカテゴリ名（“UNK”）に置き換えて評価した．

未知語・冗長語の検出の性能評価として，再現率（recall）および，検出率（precision）を用いた．以下に定義を示す．

$$\text{再現率} = \frac{\text{正しく検出した数}}{\text{未知語（冗長語）の総数}}$$

$$\text{検出率} = \frac{\text{正しく検出した数}}{\text{未知語・冗長語として検出した数}}$$

認識文の意味理解の評価として，認識文を A：正解，B：助詞誤りのみ，C：（認識誤りがあるが）意味が分かる，D：意味が分からない，の 4 種類に人間により分類し，それぞれ文正解率 (A)，助詞誤りを無視した文正解率 (A + B)，意味理解率 (A + B + C) を求めた．実際の音声理解システムの意味理解率は，文正解率と意味理解率の中間ぐらいである[29]．

以下の節において，単語正解率，文正解率などについての改善の効果を検証するために示した有意差検定の結果は，2 項分布の正規分布近似により平均の差を検定する方法[21]を用いて求めた．

4.2 音声対話タスクによる評価

4.2.1 実験条件

表 1 に音声対話タスクによる評価実験に用いられた言語モデルと音響モデルの仕様を示す．タスクは “富士山観光案内に対する問合せ” である．本実験では，提案する未知語処理法との比較のために，音響モデルとして音節カテゴリの HMM に加えて未知語のモデルとして Garbage HMM を作成した．Garbage HMM は，キーワードスポッティングのタスクで主に用いられ，一般に非キーワード音声を集めて学習する[3]．本研究は，文発話に相当する単位での連続単語音声認識であり，あらゆる冗長語に関して学習用の音声データを収集することが実際的でない．そこで，Garbage HMM は音節カテゴリの HMM と同じ状態数・構造とし，音節カテゴリの HMM の学習に用いた音声データをカテゴリを区別せずに使用し，単一のモデルで学習することによって作成した．Garbage HMM を用いて未知語処理を行う場合は，音節系列デコーダによる未知語候補を出力する代わりに，Garbage HMM のループだけを許したデコーダによる出力を未知語候補とした．

表 1 音響モデルと言語モデルの仕様
Table 1 Specification of acoustic and language models.

言語モデル	単語 bigram
音響モデル	5 状態 4 出力分布・連続出力分布型 HMM
・音節カテゴリ数	113 音節
・学習データ	ATR 研究用日本語音声データベース （男性 6 名，計 503 文 + 音韻バランス 216 単語（男性 10 名）） 日本音響学会研究用連続音声データベース （男性 30 名，計 4,500 文）
・話者適応化文数	50 文/話者
サンプリング周波数	12 kHz
窓関数	21.33 ms ハミング窓
フレーム周期	8 ms
分析	14 次元 LPC 分析
特徴パラメータ	10 次 LPC メルケプストラム +10 次回帰係数

以下では，初めに N-gram 言語モデルを用いた場合の未知語・冗長語処理に関する評価実験について述べる．次に，自然な発話に対する bigram 言語モデルベースのシステムの性能と未知語・冗長語処理の導入の効果を確かめるため，音声対話システムの利用を通して収集された対話音声を用いた場合について述べる．後者の実験では，文脈自由文法（CFG）に基づくシステムとの比較実験結果を示す．

4.2.2 未知語・冗長語処理の評価

単語 bigram の構築のために，53 人の発話者にあらかじめ使える単語（名詞と動詞）を教えた条件でタスクに関する文を発声してもらい，計 1,020 文の書き起こしのテキストデータを作成した．このうち，106 文を言語モデルの評価用として除外し，残りの 914 文（異なり単語数 357）を言語モデルの学習用データとして用いた．評価用音声データとしては，これらの書き起こしテキストとは独立に作成した未知語や冗長語を含む 115 文のセットを 2 人の話者が読み上げたもの，のべ 230 発話を使用した．冗長語を含む評価用文の例は，図 1 に示したようなものである．115 文から間投詞，言い直しにともなう語の断片を除いたときの全単語数は 734 で，1 文あたりの平均長は 6.4 単語である．115 文の内訳は，間投詞を含んだ文が 16 文，言い直しにともなう語の断片を含んだ文が 14 文，両方を含んだ文が 3 文で，計 33 文の冗長語を含む文がある．未知語を含む文は 8 文で，間投詞および言い直しにともなう語の断片を含んだ文がそれぞれ 1 文ずつ含まれている．発話内容はテキストに示されているが，なるべく自然に話すように発声してもらった．

表 2 に作成した bigram 言語モデルの学習・評価用データ，パープレキシティなどを示す[16]．この表の評価用データは音声認識評価用と同じ 115 文で，カバー率，パープレキシティの計算では冗長語を除いている．辞書としては，学習用データに出現する 357 単語を語彙として登録している．すなわち，bigram 言語モデルの学習用データには未知語が含まれないため，未知語との bigram 確率は unigram 確率の flooring 用の

表2 模擬対話文で作成したbigramの言語モデル
Table 2 bigram language model constructed using simulated dialog text.

語彙数			357
学習用データ	文数		914
	パープレキシティ		7.3
評価用データ	未知語数［総数］		17［19］
	未知語率		2.2%
	冗長語率		4.4%
	カバー率	bigram	81.5%
		unigram	97.8%
	パープレキシティ		19.9
	補正パープレキシティ		25.5

値を用いたback-offスムージングに基づく値が適用されている．

未知語および冗長語に対する音響尤度の1音節あたりのペナルティを変化させ，未知語・冗長語を含む評価用文の認識実験を行った．ペナルティの値は，音節のHMMの累積対数尤度に対して加算される値である．用いるペナルティ値は，言語制約のない連続音節認識による尤度最大の最適解の累積対数尤度と，正解発話の音節系列に従って連結したHMMの累積対数尤度との音節あたりの平均的な差の値にほぼ相当し，あらかじめほぼ最適な値を推定できる．我々が以前に行ったCFGベースの音声認識システムにおける未知語処理の評価実験では，話者6人それぞれの最適なペナルティ値を用いた場合と，話者共通のペナルティ値を用いた場合とで未知語を含む文の誤認識率を比較した結果では，話者ごとの最適なペナルティ値での誤認識率が話者間で15〜65%の開きがあったのに対し，話者ごとかまたは共通のペナルティ値を用いるかの違いによる誤認識率の変動はおよそ5%以内に収まっていた．本実験で未知語と冗長語でペナルティ値が異なるのは，冗長語では言語確率を適用しないのに対して未知語ではflooring用の値を適用していること，bigram言語モデルでは未知語を1単語として扱うために未知語の種類数に相当する言語確率の補正が必要であることに起因する．

認識された文の結果例を**図4**に，認識実験の評価結果を**表3**に示す．表においてbaselineは未知語・冗長語処理を行わなかった場合，Garbage HMMは提案する未知語処理法との比較のため音節系列デコーダの代わりにGarbage HMMを用いた場合の結果である．Garbage HMMの結果は，文正解率および正解精度が最大になるペナルティ値のときの結果のみを示した．単語認識率を計算するときには，正解文としては未知語・冗長語に対しては1つのシンボルとして扱っている．未知語・冗長語処理を行わなかった場合に比べて，

発話例1： {ヘー}，サイクリングできるんですね，{河‥}，山中湖で．
認識結果1：{{ ヘ エ エ }} サイクリング も できる んです ね {{ カ ワ }} 山中湖 で
発話例2： 富士博物館の<入館料>はいくらですか．
認識結果2：富士博物館 の << ニュ ウ カ ン リョ オ >> は いくら です か
発話例3： <富士>周辺に遊園地はありますか．
認識結果3：{{ フ ジ }} 周辺 に 遊園地 は あります か
発話例4： 旅館に{タ‥}，食事つきますか．
認識結果4：旅館 に 夕食 {{ ジ ヨ }} は つき ます か

ただし，“{ }”，“{{ }}” の部分はそれぞれ冗長語の発話部分と冗長語として検出された部分の音節列．

“<>”，“<< >>” の部分はそれぞれ未知語の発話部分と未知語として検出された部分の音節列．

図4 未知語・冗長語処理による認識結果例
Fig. 4 Examples of recognized sentences using unknown-word processing.

表3 未知語・冗長語を含む文の単語認識率［%］
Table 3 Recognition performance for utterances including OOV words and filled pauses.

ペナルティ	話者	正解率	正解精度	置換	挿入	脱落
baseline	TI	89.3	84.8	9.0	4.5	1.7
	MH	86.0	79.3	11.1	6.7	2.9
	平均	87.6	82.0	10.1	5.6	2.3
未知語 −35 冗長語 −30	TI	92.8	89.3	5.7	3.5	1.6
	MH	88.7	84.8	7.8	3.9	3.5
	平均	90.7	87.0	6.7	3.7	2.5
未知語 −40 冗長語 −35	TI	92.8	89.3	5.8	3.5	1.5
	MH	88.4	83.8	8.4	4.6	3.2
	平均	90.6	86.5	7.1	4.1	2.3
未知語 +25 冗長語 +30 (Garbage HMM)	TI	89.9	85.7	8.7	4.2	1.4
	MH	87.3	81.0	10.1	6.2	2.6
	平均	88.6	83.4	9.4	5.2	2.0

単語正解率および単語正解精度がそれぞれ向上している．両者とも，4.1.2項で述べた有意差検定法を用いて改善の割合に対して検定を行った結果は，危険率が1%以下で有意であった☆．挿入誤りが減少しているのは未知語・冗長語処理を行わなかった場合には，未知語・冗長語が複数の助詞として認識されるために，挿入誤りが多いからである．一方，Garbage HMMを用いた結果は，未知語・冗長語処理を行わなかった場合に比べて顕著な改善の効果はなかった（正解率，正解精度の改善割合に対して，それぞれ危険率が40.3%および31.6%で有意）．1発話中の単語やキーワードの出現数の制限を課しやすいキーワードスポッティングのタスクに比べて言語的な制約が緩いため，特に効果の違いが生じたと考えられる．

☆ これ以降で述べる危険率の値はすべて同じ検定法による結果である．

表 4 は発話文中の未知語・冗長語部分に対する未知語・冗長語処理の再現率・検出率を示す．ただし，未知語・冗長語として検出した音節列と実際の発話との違いは無視している．この表から，未知語に対するペナルティが −35，冗長語に対するペナルティが −30 のときが最も良い結果を得ている．未知語の検出数が少ないため，未知語を未知語として検出することが困難である．しかし，冗長語として未知語を検出できているので，検出率としては良い結果を得ている．学習データにおいて，未知語の bigram を学習することで未知語の検出数が多くなると考えられる．

表 5 に未知語・冗長語処理を行った場合の意味理解率を示す．結果をみると，未知語に対するペナルティが −40，冗長語に対するペナルティが −35 のときが最も良く，未知語・冗長語処理を行わなかった場合と比較して，正解率 (A) が 51.1%から 66.0%に向上している（危険率 1%以下で有意）．また助詞誤りを許した正解率 (A + B) では 62.0%から 80.3%と大幅に改善されており（危険率 1%以下で有意），意味理解率 (A + B + C) も 80.8%から 89.5%に改善されている（危険率 1%以下で有意）．以上のことから，未知語・冗長語処理の追加により，未知語・冗長語を含んだ文の認識精度が改善できることが分かる．

表 4 未知語・冗長語の検出性能
Table 4 Detection performance of OOV words and filled pauses using unknown-word processing.

ペナルティ	話者	再現率（%）		検出率（%）
		未知語	冗長語	
未知語 −35	TI	71.4	63.2	81.8
冗長語 −30	MH	65.7	52.6	71.7
	平均	68.6	57.9	76.8
未知語 −40	TI	62.9	57.9	94.4
冗長語 −35	MH	51.4	47.4	73.0
	平均	57.2	52.7	83.7

表 5 未知語処理・冗長語処理を行った場合の文理解率
Table 5 Semantic accuracy using unknown-word processing.

A：正解，B：助詞誤りのみ，C：意味が分かる

		理解率［%］		
ペナルティ	話者	A	A+B	A+B+C
baseline	TI	53.9	64.3	82.6
	MH	48.2	59.6	78.9
	平均	51.1	62.0	80.8
未知語 −35	TI	70.4	85.2	93.9
冗長語 −30	MH	64.9	75.4	82.5
	平均	68.6	75.4	82.5
未知語 −40	TI	69.6	86.0	94.8
冗長語 −35	MH	62.3	74.6	84.1
	平均	66.0	80.3	89.5

4.2.3 自然発話音声を用いた比較評価実験

音声対話システムのプロトタイプとして富士山観光案内システムが開発され，以前に本システムを用いた音声対話の被験者実験が行われている[24)~26)]．ここでは，そのときに収録された対話音声を使用した評価実験について述べる．

評価実験では，文脈自由文法（CFG）を用いたシステムとの比較を行った．なお，文献 16) において，同様な評価用データを用いて，文脈自由文法と単語クラスペアおよび確率付の CFG（SCFG）と単語クラス bigram の比較実験が行われている．その結果，CFG よりも単語クラスペアの方がパープレキシティおよび認識結果においてともに良い結果を得ている．また，SCFG よりも単語クラス bigram の方が良い結果を得ている．

被験者実験は 96 年 4 月，96 年 12 月，97 年 9 月の 3 回行われていて，25 人の被験者から計 2500 発話の対話音声が得られている．これらの発話から，97 年 9 月に行われた実験のうちの 4 話者の発話，計 437 発話（2,592 単語）を評価用データとし，残りの 2,063 発話（12,763 単語）を単語 bigram の学習用データとして使用した．評価用データの平均文長は 5.9 単語/文であった．

CFG と条件を合わせるために語彙数 359 単語の単語 bigram を学習した．**表 6** に使用した言語モデルの詳細を示す．CFG は人手で綿密に作成されたものであるが，受理できた文に対するパープレキシティが比較的大きい．これは，CFG が自然な対話音声を考慮して一部の倒置表現や任意の助詞部分での助詞落ちを受理するようになっていること，確率を用いていないことが主な要因である．評価用データの平均文長が比較的短いことから，特にその影響が大きかったものといえる．なお，実際の発話には倒置表現はなかった．パープレキシティが大きい CFG に対して，受理率でも単語 bigram の方が有利であり，CFG で対話音声向けに制約の強い文法を作成することが困難であることが分かる．

富士山観光案内システムの被験者実験で収集した入

表 6 spontaneous speech を用いた評価実験用の言語モデル
Table 6 Language models for experiments using spontaneous speech.

言語モデル	CFG	bigram
語彙サイズ	359	
未知語率（%）	2.0	
冗長語率（%）	0.2	
文受理率（%）	81.5	92.0
パープレキシティ	130	9

表7 自然な対話音声に対する認識結果
Table 7 Experimental result of spontaneous speech.

A：正解，B：助詞誤りのみ

言語モデル & UWP	単語正解精度（%）	文正解率 (A)（%）	文正解率 (A+B)（%）
CFG, with UWP	73.3	39.3	76.9
bigram, no UWP	91.5	72.8	84.7
bigram, with UWP	91.8	69.1	88.6

力発話に対して認識実験を行った．**表7** に実験結果を示す．前項と違って bigram の学習データに未知語が含まれていることから，改めて未知語処理における未知語・冗長語に対するペナルティ値を実験的に求めた．評価データを各話者で約半分に分割して話者共通の最適なペナルティ値を求めた結果は一致しており，本実験の結果では，未知語に対するペナルティは −60，冗長語に対するペナルティは −40 を用いている．この実験結果においては，baseline の場合と比較して単語正解率および単語正解精度の差はほとんどない．その主な理由は，未知語の種類が少なく，評価用データに多く含まれる未知語の “知りたい” や “行きたい” が，辞書に登録されている登録単語である “したい” と比較的近い単語であったため，未知語として検出できていないためであった．文正解率については，未知語処理によって逆に低下しているが，その原因はほとんど登録語部分を未知語として認識した過剰検出による誤りのためであった．しかし，未知語・冗長語処理によって，助詞誤りを許した意味的な正解率の改善はその影響を上回っている．この表の結果において，bigram 言語モデルを使用した方が CFG の場合と比較して文正解率で 29.8%，助詞誤りを許した文正解率で 11.7%改善されている（両者とも危険率 1%以下で有意）．なお，未知語処理を用いた場合に，CFG および bigram の両者で受理できる評価用文だけで評価すると，文正解率は CFG ベースのシステムで 48.9%，bigram ベースのシステムで 69.4%となり，助詞誤りを許した正解率ではそれぞれ 89.0%と 88.8%という結果であった．すなわち，言語モデルで受理できる文において bigram の有効性が顕著であり，未知語を含む受理されない発話に対しても bigram ベースのシステムの方がより頑健であったといえる．また，この実験結果に基づいた対話システムの評価[18]においても，CFG ベースのシステムの認識結果による意味理解率が 57%であったのに対し，bigram ベースでは 68%に改善されている（危険率 1%以下で有意）．CFG のパープレキシティが 130 であるのに対して，bigram は 9 と非常に差があることも大きな理由であるが，対話システムの言語解析処理において助詞誤りがある程度許される場合には[23]，bigram 言語モデルによる対話音声の認識法が CFG よりも有効であることが実対話データにおいても示された．

表8 音響モデルと言語モデルの仕様（大語彙音声認識システム）
Table 8 Acoustic and language models for experiments using a large-vocabulary speech recognition system.

言語モデル	bigram	trigram
語彙サイズ	5,000	
学習データ	毎日新聞記事テキストデータベース（1991 年 1 月～1994 年 10 月，約 8,600 万形態素，RWC の形態素解析結果を使用）	
未知語率（%）	10.6	
テストセットパープレキシティ	74.8	50.4
テストセット補正パープレキシティ	141	94.9
音響分析	表 1 と同じ	
音響モデル ・学習データ	セグメント入力型 HMM（5 状態 4 出力分布，4 混合ガウス分布）表 1 のデータに加えて，日本音響学会 JNAS 新聞記事読み上げコーパスに含まれる 125 名の男性話者の計 17,221 発話を使用	
特徴パラメータ	20 次元のセグメント特徴量，10 次のメルケプストラム係数の 1 次および 2 次の回帰係数，パワーの 1 次および 2 次の回帰係数	
話者適応化	なし（不特定話者音声認識）	

4.3 大語彙ディクテーションタスクによる評価実験

大語彙音声認識システムにおいて，前述の音声対話システムの場合と同様に未知語処理の評価実験を行った．言語モデルと音響モデルの仕様を **表8** に示す．N-gram 言語モデルの構築には CMU SLM Toolkit ver1[19] を使用し，bigram カウントのカットオフを 0 とした backoff bigram モデルを標準の設定で作成した．また，音響モデルとして，スペクトルの動的な特徴をより精密にモデル化したセグメント入力型 HMM を用いた[22]．4.2 節と同様に，比較のために音節カテゴリ HMM と同じ状態数・混合数・構造の Garbage HMM を作成した．言語モデル，音声認識システムの評価用としては，日本音響学会新聞記事読み上げ音声コーパス（JNAS）から，音響モデルの学習用とは異なる男性 3 人（NM006, NM014, NM017）の各々異なる約 100 文の発話，計 311 発話（5,572 単語）を用いた．

使用した N-gram ベースの大語彙音声認識システムは，第 1 パスで bigram 言語モデルを用いた上位 200 文の認識結果を出力し，第 2 パスで trigram 言語モ

表 9 大語彙タスクにおける未知語処理の効果
Table 9 Effect of unknown-word processing on a large-vocabulary dictation task.
$PP_{UNK} = 37924$：学習データ中の未知語の unigram によるパープレキシティ
$m_{UNK} = 292529$：学習データ中の未知語の種類数
GB：Garbage HMM を使用した未知語処理法

言語モデル UWP & ペナルティ	単語正解精度（%）		単語レベル		文レベル		文レベルでの棄却後の内訳		
	全発話データ	文レベルでの棄却後	再現率（%）	検出率（%）	再現率（%）	検出率（%）	未知語なし文数	未知語あり文数	棄却誤り文数
bigram, 未知語処理なし	65.0	(85.9)	–	–	–	–	–	–	–
bigram, $-\log 10$	58.1	87.5	60.3	70.2	95.0	84.7	31	12	41
bigram, $-\log 5000$	64.5	74.5	29.5	84.9	60.3	91.1	58	95	14
bigram, $-\log(PP_{UNK})$	65.9	71.3	20.0	88.7	43.5	92.9	64	135	8
bigram, $-\log(m_{UNK})$	65.8	68.9	12.2	91.1	28.5	95.8	69	171	3
bigram, GB, $-\log(m_{UNK})$	65.1	65.8	2.7	88.9	6.7	88.9	70	223	2
trigram, 未知語処理なし	66.7	(87.3)	–	–	–	–	–	–	–
trigram, $-\log 10$	62.6	86.8	67.6	68.6	95.8	82.4	23	10	49
trigram, $-\log 5000$	67.4	79.8	39.0	84.2	75.7	89.2	50	58	22
trigram, $-\log(PP_{UNK})$	67.8	75.2	29.5	88.8	61.5	94.2	63	92	9
trigram, $-\log(m_{UNK})$	68.1	72.4	19.7	89.9	43.1	93.6	65	136	7
trigram, GB, $-\log(m_{UNK})$	67.2	68.2	4.6	93.1	11.3	93.1	70	212	2

デルでリスコアリングを行っている[10]．また，未知語候補のペナルティは 1 音節あたり −50 とし，未知語の N-gram 確率に対して未知語の種類数に相当するペナルティとして，$\log 10$ から $\log 292$，529 まで数段階に変えて実験を行った．未知語候補の音節あたりのペナルティは，前節までと音響モデルが異なるため改めて決定しているが，予備実験において話者ごとの最適なペナルティの変動が少ないことから，1 話者の評価セットにおいてほぼ最適なペナルティを決定した．ここで，$\log 292$，529 という値は，N-gram 言語モデルの学習データ中に現れた未知語の種類数 $m = 292$，529 に基づいて，各未知語の出現確率を等確率と仮定した場合である．また，PP_{UNK} という値は，未知語として学習データ中に現れる単語の頻度情報 $P(w_i)$（ただし，$\sum_{w_i \in OOV} P(w_i) = 1$）を用いて，未知語となる全単語のエントロピー H から 2^H として計算されるパープレキシティに相当する値である[12]．

表 9 に，未知語処理を用いた N-gram ベースの大語彙音声認識システムの性能を示す．本実験では，語彙数が少ないため評価データの未知語率が大きく，1 発話中に複数の未知語が含まれる場合もあり[27]，比較的難しいタスクとなっている．また，認識の単位が比較的短い形態素であるため，未知語の検出位置の始・終端のずれが前後の単語の認識結果に影響を及ぼしやすい．結果として，表の “全発話データ” の結果に示されるように，評価セット全体の単語正解精度においてわずかであるが効果がみられた（bigram および trigram 言語モデルでペナルティとして $-\log(m_{UNK})$ を用いた場合に，それぞれ危険率 37.5%および 11.5%で有意）．また，Garbage HMM を用いる方法の結果（表の “GB” と記述された行）では提案手法に比べて改善の効果は小さかった（bigram および trigram 言語モデルを用いた場合に，それぞれ危険率が 91.2%および 57.5%で有意）．

さらに，発話の文レベルでの未知語の棄却後の性能として再評価を行った．表の “文レベルでの棄却後” および “文レベルでの棄却後の内訳” の項目にその結果を示している．文レベルでの未知語の棄却後の評価は，未知語を一部に含む発話に対して未知語を任意の部分で検出した場合にその発話全体を棄却し，それらを除いた発話で評価を行った結果である．“文レベル” および “発話ごとの棄却後の内訳” の結果は，文レベルでの未知語を含む文の棄却（正しい検出）の性能と結果の内訳を示している．“棄却誤り文数” は，未知語を含まない文で誤って棄却された文数である．一方，“単語レベル”（未知語）の再現率と検出率の評価では，テキストレベルでの正解発話との DP マッチングにおいて，入力の未知語部分に対して認識結果の未知語候補が対応づけられたものを正しい検出と見なしている．表の括弧内の値は，未知語を含む発話を除いた全発話に対する認識結果である．“未知語処理なし” の場合において，実際に未知語を含まない文（311 文中 72 文）では，bigram の言語モデルで約 86%の単語正解精度が得られている．一方，未知語を含んだ文（239 文）を含めた全体の単語正解精度は 65%で，未知語率が 10.6%であるので，単語正解精度の期待値（$0.894 \times 0.86 = 0.769$ (76.9%)）よりもかなり悪い．未知語を含まない文集合に対するパープレキシティは

発話例1 ： ベルリン 出身 の <指揮者> <ハインツ> <フリッケ> が ワシントン <歌劇> 場 の 音楽 監督 から 音楽 総 監督 に 昇進 した
認識結果1 ： ベルリン 出身 の 趣旨 が 相次い で だまし 取っ て 各 劇場 の 音楽 監督 から 音楽 総 監督 に 昇進 した
認識結果1'： <UNK> 高い <UNK> か 劇場 の 音楽 監督 から 音楽 総 監督 に 昇進 した
認識結果1"： ベルリン 出身 の <UNK> が ワシントン と 劇場 の 音楽 監督 から 音楽 総 監督 に 昇進 した
発話例2 ： 昨年 同 <大橋> の 危険 性 が 指摘 され ながら <見過ごさ> れ た 際 に は 副 市長 と して <在職> して い た
認識結果2 ： 昨年 同 王 橋 の 危険 性 が 指摘 され ながら 組長 され た 際 に は 副 市長 と して 退職 して い た
認識結果2'： 昨年 同 <UNK> の 危険 性 が 指摘 され ながら <UNK> され た 際 に は 副 市長 と して <UNK>
認識結果2"： 昨年 同 王 橋 の 危険 性 が 指摘 され ながら 組長 され た 際 に は 副 市長 と して 退職 して い た
発話例3 ： 元 大手 <製薬> 会社 の <研究員> だっ た <芦田> 信 社長 が 七五 年 に 設立 人間 の <尿> から 生産 し た <血栓> <溶解> 剤 の 製造 を 始め た
認識結果3 ： 元 を 徹底 や 子会社 の 研究 院 だっ たし た 新 社長 が 七五 年 に 設置 する 人間 の 夜 から 制裁 的 信用 内外 の 制度 を 始め た
認識結果3'： 元 <UNK> 健 <UNK> 新 社長 が 七五 年 に 設立 に <UNK> 再生 専用 <UNK> 始め た
認識結果3"： 元 を 徹底 や 子会社 の 研究 院 だっ たし た 新 社長 が 七五 年 に 設置 する 人間 の 夜 から 制裁 的 信用 内外 の 制度 を 始め た

(<>部分が未知語，<UNK> が未知語として検出された部分．「認識結果 n」は未知語処理なしの場合．「認識結果 n'」，「認識結果 n"」は，それぞれ未知語の言語確率に与えるペナルティが $-\log 10$ および $-\log(m_{UNK})$ の場合.)

図5 ディクテーションタスクでの未知語処理例
Fig. 5 Results of unknown-word processing in dictation task.

bigram が 90.3，trigram が 58.8 であるのに対し，未知語を含む文集合のパープレキシティ（補正パープレキシティ）は bigram が 102.3 (156.7)，trigram が 66.8 (106.3) と高く，より難しい文であることや，未知語を含んでいる文では平均 2.5 単語の未知語を含んでいることから，未知語の存在が既知語の認識に大きく影響を及ぼしたと考えられる．

結果として，文単位での正しい棄却割合（検出率）が高い条件となるペナルティ（$-\log(m_{UNK})$）の場合では単語正解精度に関して未知語処理による顕著な効果は見られないが，ペナルティとして小さい値（$-\log 10$）を用いた場合において，発話の文レベルごとでの未知語の棄却の性能（再現率と検出率）および棄却後の単語正解精度が全体的に高い結果が得られている．実際，認識結果を分析すると，**図5** に示す例のように，未知語の検出位置のずれが多く見られ，このことから，全発話データでの単語正解精度において未知語処理の効果が小さい結果となっている．

図5の発話例1，2のように，未知語の確率に対するペナルティによって，未知語の前後や文末の未知語を含まない単語系列全体を1つの未知語として誤って検出する例や，未知語が検出されない場合がある．特に，未知語に類似した登録語がある場合や，1文中の未知語が多い場合などには，未知語の検出や登録語との境界の検出が困難であった．したがって，これらの問題の改善のためには，未知語に関する言語的な制約の強化や音響モデルの精密化がさらに必要である．

この実験結果において，ペナルティが大きい場合（たとえば $-\log(m_{UNK})$ のとき）に全発話に対して未知語処理による認識精度の改善は1%程度であるが，誤りと棄却では理解率に大きな差を及ぼす．たとえば，未知語処理なしの場合に認識誤りとなっていた部分が，未知語処理によって未知語として検出される場合でも認識精度には影響しないが，対話システムにおける影響を考えると，発話中のある部分が認識誤りとなるより未知語として検出される方が意味理解の過程においてより望ましいといえる．したがって，ディクテーションを目的とする場合，より高い単語正解精度が得られるペナルティ（たとえば $-\log(m_{UNK})$）を用いるほうが良いが，音声対話システムへの適用を考えた場合では，文や単語レベルでの未知語の検出割合（再現率）が高い方が発話の理解率や対話の効率が向上する可能性がある．たとえば，trigram 言語モデルでペナルティとして $-\log 10 \sim -\log(PP_{UNK})$ を選択すれば文レベルでの検出割合（再現率）が50%以上，$-\log 10$ を選択すれば単語レベルでの検出割合も 50%以上となる．また，提案した未知語処理を用いる方法は，未知語の辞書情報を新たに追加登録することを前提とした場合においても[28]，発話の文レベルごとのリジェクションによって未知語の存在を検出することへの適用も考えられる．これらのことから，提案した未知語処理は大語彙音声認識システムにおいても有用と考えられる．

5. まとめ

本研究では，音声対話システムにおいてユーザの発話文として予想される未知語および間投詞や言い直し，言い淀みなどの冗長語を含んだ文に対処するため，

bigram 言語モデルを用いた音声認識システムにおいて未知語・冗長語処理の検討を行った．

音声対話タスクにおける未知語・冗長語を含む評価データに対して，未知語処理と冗長語処理を併用するシステムで認識実験を行った結果，単語正解精度が改善され，意味理解率も大きく向上した．また，未知語・冗長語を含む文において，文脈自由文法（CFG）による認識実験と比較した結果，助詞誤りを許す文正解率で CFG よりも良い結果を得た．さらに，対話システムにおいて収集した対話音声に対しての認識実験においても，CFG と比較して，文正解率および助詞誤りを許した文正解率で CFG よりも良い結果を得ることができた．

協力的に発声された発話文に対しては学習データ量が十分多い場合（数百単語のタスクなら 2,000 文以上）は，bigram 言語モデルの方が CFG による言語モデルよりも認識精度の良いことが示されていたが[16]，冗長語・未知語が現れる対話音声文おいても，本研究で構築した bigram 言語モデルによる未知語・冗長語処理を行う認識アルゴリズムが文脈自由文法に基づく方法よりも有効であることが分かった．

また，大語彙連続音声認識タスクにおいて，未知語処理の評価実験を行い，未知語の種類数を考慮した条件において単語レベルの認識精度では若干の改善がみられた．また，文レベルの評価を行った場合には，未知語の検出をより増加させ，単語正解精度が未知語処理なしの場合と同程度になる条件において，文レベルでの未知語の検出での再現率がそれぞれ 95.8%および 61.5%のときに 82.4%および 94.2%の検出率が得られた．より高い未知語の検出性能が得られることで，対話システムへの応用に対して有利になると考えられる．

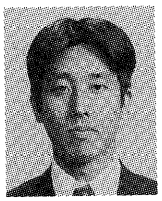
提案した方法は，未知語と発音が近い単語が語彙に含まれる場合や，1 文中の未知語が多い場合などにおいて，未知語の検出漏れや，単語境界の誤りが生じやすい問題がある．このような問題を改善するためには，未知語のカテゴリを設けるなど，未知語に関する言語的な制約の強化や音響モデルの精密化のためのさらなる改良が必要といえる．

参 考 文 献

1) Gauvain, J-L. and Lamel, L.: Large vocabulary continuous speech recognition: From laboratory systems towards real-world applications, 信学論, Vol.J79-D-II, No.12, pp.2005–2021 (1996).
2) 中川聖一，小林 聡：自然な音声対話における間投詞・ポーズ・言い直しの出現パターンと音響的性質，日本音響学会誌，Vol.51, No.3, pp.202–210 (1995).
3) Wilpon, J.G., Rabiner, L.R., Lee, C-H. and Goldman, E.R.: Automatic recognition of keywords in unconstrained speech using hidden Markov models, *IEEE Trans. ASSP*, Vol.38, No.11, pp.1870–1878 (1990).
4) Asadi, A., Schwartz, R. and Makhoul, J.: Automatic detection of new words in a large vocabulary continuous speech recognition system, *Proc. ICASSP '90*, pp.125–128 (1990).
5) Kita, K., Ehara, T. and Morimoto, T.: Processing unknown words in continuous speech recognition, *IEICE Trans.*, Vol.E74, No.7, pp.1811–1816 (1991).
6) 伊藤克亘，速水 悟，田中穂積：連続音声認識における未知語処理の扱い，信学技報，SP91-96 (Dec. 1991).
7) 甲斐充彦，中川聖一：冗長語・言い直し等を含む発話のための未知語処理を用いた音声認識システムの比較評価，信学論，Vol.J80-D-II, No.10, pp.2615–2625 (1997).
8) Kai, A. and Nakagawa, S.: Relationship among recognition rate, rejection rate and false alarm rate in a spoken word recognition system, *IEICE Trans. Inf. & Syst.*, Vol.E-78-D, No.6, pp.698–704 (1995).
9) 周 旻，堤真理子，中川聖一：確率モデルにおける大語彙連続音声認識の評価，情報処理学会研究報告，96-SLP-11-6, pp.31–36 (May 1996).
10) 甲斐充彦，廣瀬良文，中川聖一：N-gram 言語モデルと効率的探索法を用いた大語彙連続音声認識システムの検討，信学技報，SP97-99, pp.31–38 (Jan. 1998).
11) Koga, S., Isotani, R., Tsukada, S. and Yoshida, K.: A Real-Time Speaker-Independent Continuous Speech Recognition System Based on Demi-Syllable Units, *Proc. ICSLP*, pp.1483–1486 (1992).
12) 中川聖一，赤松裕隆：未知語を含む文集合のパープレキシティの算出法—新補正パープレキシティ，日本音響学会平成 10 年度秋季研究発表会講演論文集，2-1-13, pp.63–64 (Sep. 1998).
13) 渡辺隆夫，塚田 聡：音節認識を用いたゆう度補正による未知発話のリジェクション，信学論，Vol.J75-D-II, No.12, pp.2002–2009 (1992).
14) Ariki, Y. and Kawamura, T.: Simultaneous spotting of phonemes and words in continuous speech, *Proc. ICSLP 94*, pp.2191–2194 (1994).
15) 新美康永，高橋一城，小林 豊：音声認識結果の認識誤り区間と未知語区間の推定，信学技報，SP95-31 (June 1995).
16) 中川聖一，大谷耕嗣：Bigram の使用による話し言葉用確率文脈自由文法の自動学習，情報処理学

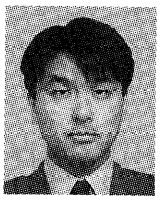
会論文誌，Vol.39, No.3, pp.575–584 (1998).
17) 中川聖一：確率モデルによる音声認識，電子情報通信学会 (1988).
18) 小暮 悟，伊藤敏彦，廣瀬良文，甲斐充彦，中川聖一：CFG/bigram を使用した対話音声認識における意味理解の比較検討，情報処理学会全国大会予稿集，4R-04 (Oct. 1998).
19) Rosenfeld, R.: The CMU statistical language modeling toolkit, and its use in the 1994 ARPA CSR Evaluation, *Proc. ARPA Spoken Language Systems Technology Workshop*, pp.47–50 (1995).
20) Ueberla, J.: Analysing a simple language model – some general conclusions for language models for speech recognition, *Computer Speech and Language*, Vol.8, No.2, pp.153–176 (1994).
21) 中川聖一，高木英行：パターン認識における有意差検定と音声認識システムの評価法，日本音響学会誌，Vol.50, No.10, pp.849–854 (1994).
22) Nakagawa, S. and Yamamoto, K.: Evaluation of segmental unit input HMM, *Proc. ICASSP '96*, pp.439–442 (1996).
23) 山本幹雄，伊藤敏彦，肥田野勝，中川聖一：人間の理解手法を用いたロバストな音声対話システム，情報処理学会論文誌，Vol.37, No.4, pp.471–482 (1996).
24) 傳田明弘，伊藤敏彦，中川聖一：マルチモーダルインターフェースを備えた観光案内対話システムの評価，人工知能学会全国大会，15-09, pp.431–434 (1996).
25) 傳田明弘，伊藤敏彦，小暮 悟，中川聖一：マルチモーダルインターフェースを備えた観光案内対話システムの評価実験，情報処理学会研究報告，SLP-15-8, pp.47–52 (1997.2).
26) 中川聖一，傳田明弘，伊藤敏彦：マルチモーダル観光案内対話システム，人工知能学会誌，Vol.13, No.2, pp.241–251 (1998).
27) 萬崎 弘，山本幹雄，板橋秀一：日本音響学会新聞記事読み上げコーパスからの評価用発話セットの作成，日本音響学会平成 10 年度秋季研究発表会講演論文集，1-R-13, pp.143–144 (Sep. 1998).
28) 西村雅史，伊東伸泰：ニュース音声書き起こしシステムに関する検討，日本音響学会平成 10 年度秋季研究発表会講演論文集，1-R-14, pp.145–146 (Sep. 1998).
29) 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，Vol.39, No.5, pp.1240–1249 (1998).

(平成 10 年 10 月 6 日受付)
(平成 11 年 2 月 8 日採録)



甲斐 充彦（正会員）

昭和 43 年生．平成 3 年豊橋技術科学大学工学部情報工学課程卒業．平成 5 年同大学院修士課程修了．平成 8 年同大学院博士後期課程修了．同年豊橋技術科学大学工学部助手．音声認識と言語処理に関する研究に従事．工学博士．日本音響学会，電子情報通信学会各会員．



廣瀬 良文

昭和 50 年生．平成 10 年豊橋技術科学大学情報工学課程卒業．現在奈良先端科学技術大学院大学情報科学研究科在学中．音声認識に関する研究に従事．日本音響学会会員．

中川 聖一（正会員）

昭和 23 年生．昭和 51 年京都大学大学院博士課程修了．同年京都大学工学部情報工学科助手．昭和 55 年豊橋技術科学大学情報工学系講師．平成 2 年教授．昭和 60～61 年カーネギーメロン大学客員研究員．音声情報処理，自然言語処理，人工知能の研究に従事．工学博士．昭和 52 年電子情報通信学会論文賞，1988 年度 IETE 最優秀論文賞授賞．著書「確率モデルによる音声認識」（電子情報通信処理学会編），「音声・聴覚と神経回路網モデル」（共著，オーム社），「情報理論の基礎と応用」（近代科学社）等．