単語を認識単位とした日本語の大語彙連続音声認識

西村雅史 伊東伸泰 山崎一孝 荻野紫穂
日本アイ・ビー・エム（株）東京基礎研究所
e-mail: nisimura@trl.ibm.co.jp

我々は先に，日本人が感覚的に捉えている単語単位を，既存の形態素解析プログラムの出力である形態素単位との統計的対応関係から自動推定する方法を提案し，それを認識および発声の単位とする離散単語発声の日本語ディクテーションシステムを構築した．今回，この人間の考える単語単位を連続音声認識の認識単位としても利用することを試み，特に，他の大語彙連続音声認識システムで用いられる事の多い形態素単位と比較し，その有効性について調査した．また，認識単位の定義が一意に決まらない現状を踏まえて，日本語の連続音声認識システムの評価方法を提案するとともに，不特定話者の大語彙音声認識実験結果について報告する．男女各 10 名に対する認識実験の結果，文字誤り率 3%，単語誤り率 4.3%が得られた．さらに，句読点の自動挿入方法や，未知語モデルを使った単語N-gramによる単語単位の自動分割方法などについても述べる．

Word-based approach to large-vocabulary continuous speech recognition for Japanese

Masafumi NISHIMURA, Nobuyasu ITOH, Kazutaka YAMASAKI, Shiho OGINO
IBM Research, Tokyo Research Laboratory, IBM Japan, Ltd.

In this paper, we discuss a word-based Japanese continuous dictation system. We have previously proposed a statistical method for segmenting a text into words on the basis of human intuition, and developed an isolated-word-based Japanese dictation system. By comparing this word unit used for the isolated word recognition with grammatical units, we show that this unit is also very useful for continuous speech recognition. Evaluation of the performance of this continuous dictation system showed that the character error rate was 3%, and that the word error rate was 4.3%. We also present a method for inserting punctuation marks in spoken texts automatically, and a method for segmenting Japanese text into words by using an N-gram model, focusing on how to handle unknown words.

1. はじめに

近年欧米では，統計的言語モデルを用いたディクテーションシステムが実用化され，現在では医療所見の入力といった特定分野から，徐々にではあるが個人ユーザーが日常的な文章を入力する手段へと市場が広まって来ている．そして研究の対象も，次第に読み上げ文からニュース音声の書き起こしなどのより自然な発話へと移行しつつある[1].

一方，日本語については，単音節発声をベースとした日本語音声ワープロが検討されて以来，長い間，音声による日本語文の入力システムは実用化されなかった．その直接の原因としては，欧米に比べて音声および言語データベースの整備が遅れたことが大きいが，技術的には，日本語の単位があいまいで，言語的な単位への自動分割が困難であったこと，基本的にN-gramなどの単純な統計モデルでは日本語を正確に表現することは難しいと信じられていたこと，また，欧米語で行われていたような離散単語発声が困難または，不適切であると考えられていたこと，などが理由として挙げられる．

我々は先に，形態素解析プログラムの出力である形態素と，日本人が感覚的に捉えている"単語単位"との対応関係を統計的なモデルで表現し，それによって日本語の単語単位を自動推定する方法を提案した[2]. そして，この単位を使えば日本語でも離散単語発声によるディクテーションシステムが，欧米語と同様にN-gram言語モデルと音素HMMによって構成される認識システムとして実現できることを示した．

ここでは，この単語単位を連続音声認識の認識単位としても使用することを試みる．このような単語単位は離散発声可能な発声の最小単位であることから，自由なポーズの挿入が可能となる利点がある．また，既存の形態素単位などと比較した場合，その単語カバレージは十分に高く，一方で単位あたりのモーラ長が長く，音響的には識別しやすい単位となっていることを示す．また，認識単位が異なる認識システムの性能を評価する方法についても検討する．さらに，句読点の自動挿入方法や，単語N-gramによって単語単位分割を行う方法についても触れる．

2. 単語の定義

日本語のように単語の概念が明確でない言語においてディクテーションシステムを実現しようとする場合，どのような単位を認識単位とするかが問題となる．連続発声に対処できるアルゴリズムを適用する限り，原理的には1音素以上のどのような単位を使用することも可能ではあるが，音響モデルとしても，言語モデルとしても，ある程度まとまった長さの単位が効率よく定義されていた方が良い．その意味で，日本語の解析的な最小単位である，形態素が用いられることが多い．

一方，我々は，大語彙音声認識の処理を軽減する手段として離散単語発声に着目し，日本語においても，離散発声入力が可能との条件を満たすような単位を単語として定義し，これを認識単位としてきた．この単位は日本語の最小発声単位をほぼ包含しているから，連続発声に適用した場合には，発話者が，文中に自由にポーズを置くことができるという利点がある．当然，これまでどおり離散単語発声による入力も可能である．また，離散発声は，認識結果の修正作業の際など，単語分割位置に関する誤りを回避したい場合には特に有効である．

3. 連続音声認識における認識単位の比較

離散単語発声のために定義した単語単位が，連続音声認識の認識単位として，どの程度有効であるかについて，他の単位との比較を行った．具体的には，形態素解析プログラム[4]の出力(以降，このプログラム名からこの単位を仮にJMA単位と呼ぶことにする[1])と，この単位の複合語部分を短単位に分割した，より本来の形態素に近い単位(以降，これを形態素と呼ぶことにする)と比較する．

3-1. カバレージの比較

日経新聞3ヶ月分のデータを，JMA単位，形態素単位，単語単位にそれぞれ自動分割し，各単位を頻度順にならべた辞書を用意した．また，このデータとは異

[1] この単位を単に形態素と呼ぶことも多い．ただ，この解析的単位には複合語などを1つの単位として多数含んでおり，文法学者が呼ぶところの形態素とはかなり異なっている．

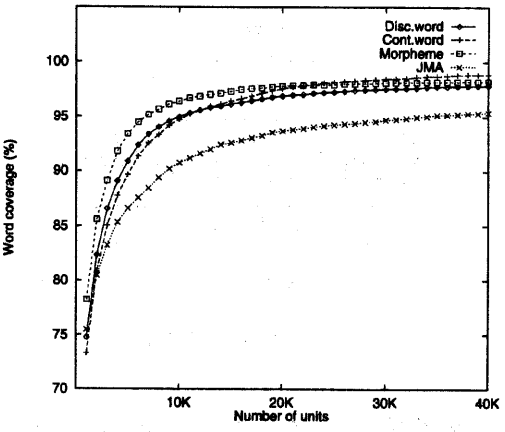

図1 認識単位毎の語彙サイズとカバレージの関係

なる600文のテストデータを用意し，それぞれ，(1)人間が単語単位に分割したもの，(1')単語のTri-gram言語モデルを用い，得られる尤度が最大になるように単語単位に自動分割したもの[2]，(2)形態素解析プログラムを用いて解析後，さらに人間が全数をチェックし，JMA単位に分割したもの，(3)複合語を分割し，再度人間が全数チェックして，形態素単位としたものの4種類のデータを作成し，それぞれの単位に対応する辞書を用いて，語彙サイズとテスト文のカバレージの関係を調べた．単語単位に関し，(1)は単語単位の離散発声入力を，そして(1')は同じ単語単位を使っているが，連続音声認識のアルゴリズムを適用した場合に相当すると考えられる．結果を図1に示す．

我々の提案した単語単位は，単語分割の揺らぎを反映した単位となっているが，離散単語発声(1)の場合にも，揺らぎを表現するために必要とされる単位の数はそれほど多くはなく，不利と思われた単語カバレージに関しても，形態素単位に近い効率を持ち，JMA単位をそのまま使うよりははるかに効率がよい単位であったことがわかる．

また，連続音声認識を想定した(1')の場合には言語モデルを参照して尤度が最大となるような単語列が選ばれることから，カバレージはさらに改善されることとなり，4万語では約99%に達する．この結果を見る限り，少なくとも2万語以上の語彙を用意する場合には，いずれの単位よりも単語単位のカバレージが高くなっている．

[2] 未知語の処理方法を含めN-gramモデルによる単語の自動分割方法については5-2と5-3.で述べる．

表1 認識単位と単位あたりの平均モーラ長

	認識単位		
	単語	形態素	JMA
日経新聞	2.5	2.2	2.4
産経新聞	2.3	1.9	1.9
電子会議室	2.3	1.8	1.8
ビジネストーク	2.3	1.9	1.9
小説	2.2	1.7	1.7
平均	2.3	1.9	1.9

3-2. 平均モーラ長の比較

一方，種々のタスクに対し，上記3種類の単位(JMA，形態素，単語)において，テスト文中の単位あたりの読みのモーラ長を比較したのが表1である．日経新聞ではいずれの単位のモーラ長も差は小さいが，他のタスクでは平均20%以上単語単位のモーラ長が長い．少なくとも我々の認識システムを含む多くの連続音声認識システムでは認識単位間に渡る右側音素環境の予測は十分でないので，認識単位内のモーラ長が長い分，音響的な識別にとっては有利な単位であると考えられる．

なお，各単位の出現頻度を無視し，辞書中の読みの平均モーラ長をカウントすると，単語が3.9，形態素が3.4，JMA単位が4.6となり，複合語をたくさん含む分，JMA単位が突出してモーラ長が長い．しかしそれらは主に低頻度語なので，実際のテスト文に対しては形態素とほとんど差がない．

3-3. 言語モデルの能力の比較

次に，約200万文のテキストデータを使って3種類の単位それぞれに対してTri-gram言語モデルを構築し，その性能を評価した．なお，テスト文は主にビジネストークからとった679文であり，3-1.における(1),(2),(3)の基準で分割したものを用いた．

未知語の有無がパープレキシティの値に大きな影響を与えるためいずれの場合もパープレキシティの算出時には未知語の予測を計算から除外している．また，単語単位についてはすでに用意してあった約4万語の辞書を用いたが，形態素ならびにJMA単位に対して

表2 認識単位による言語モデルの能力の比較

評価尺度	認識単位		
	単語	形態素	JMA
文あたりの平均単位数	24.2	30.2	29.5
文あたりのEntropy	169.2	167.3	165.3
単位あたりのPerplexity	127.9	46.6	49.0
形態素あたりのPerplexity	48.6	46.6	44.5

はテストデータに対して単語単位と同等のカバレージ（97.5%）が得られるように高頻度のものから順次単位を追加し，単位ごとに新たに辞書を作成した上でそれぞれの言語モデルを推定した．

これらの単位はそれぞれ長さが異なり，文を構成する単位数が異なるため，何らかの基準で正規化を行わないと比較ができない．ここでは各単位に対し，一文あたりのエントロピー(H_{snt})を推定した後，次式で形態素あたりのパープレキシティ(PP_{mor})に正規化して比較を行った．

$$H_{mor} = H_{snt} / N$$
$$PP_{mor} = 2^{H_{mor}}$$

ここで，H_{snt}は文あたりのエントロピー，Nは文あたりの形態素の単位数である．単位あたりのパープレキシティおよび，形態素あたりに換算したパープレキシティを表2に示す．

この結果から分かるように，各単位を形態素の単位あたりに換算して求めたパープレキシティに本質的な差はないことから，言語モデルの能力に関しては，いずれの単位を用いてもそれほど大きな差はないと思われる．

本来ならば，単位長の長い単語単位が，同じN-gramによってより広い範囲を表現できる分優位であるはずであるが，一般的な単語の切り出し揺らぎを表現するため語彙サイズが大きくなっていた分曖昧さが増し，結局いずれの単位も能力的に差がない程度に落ち着いているのだと考えている．

いずれにせよ，このように人間の振る舞いに基づいて推定した単語単位が，形態素などの単位と比較して，言語モデルの能力という点ではあまり差がなかったものの，単語カバレージに優れ，また，音響的にも識別しやすい単位となっていることが分かった．

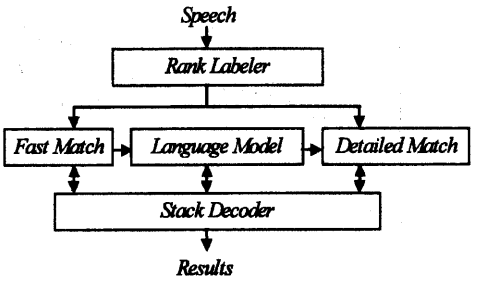

図2 認識システムの構成

4. 連続音声のディクテーションシステム

4-1. 認識システムの構成[8,2]

本認識システムは図2に示すように，ランクラベラー，ファーストマッチとディーテイルマッチを併用したスタックデコーダーおよびTri-gram言語モデルなどから構成される．なお，音響モデルは混合正規分布で表現された音素環境依存型のHMMである．

認識対象語彙は主に出現頻度に基づいて選択した39,295語（40K語彙）とした．

4-2. 言語モデル[2]

約200万文のテキストデータに対して，形態素解析処理を行った後，単語分割モデルを用いて単語単位に自動分割し，これを言語モデルの学習データとした．なお，学習データには句読点や括弧類など読み上げが可能と思われる記号はすべてそのまま含んでいるが，極端に記号が多い文章など，読み上げに不適切な文章はあらかじめ除いてある．この学習データを用いてTri-gram言語モデルの推定を行った．

4-3. 音響モデル[2,3]

20代～60代の男性110名，女性105名がそれぞれ120文（Set-25K: 計25,800文，355,454単語），20代～50代の男性312名，女性346名がそれぞれ148文(Set-100K: 計97,384文，1,338,307単語）を連続発声したデータを用いて男女共通の不特定話者用音響モデルを作成した．なお，各発声文には新聞記事，手紙，小説など多様な文書を用意し，各人がそれぞれ異なる文章を読み上げ

ている．採取は比較的静かな室内で，接話型のマイク(SHURE-SM10)を用いて行った．

4-4. 句読点の自動挿入

連続音声によるディクテーションの場合，発声をより自然なものにするため，句点(。)および読点(、)を自動的に挿入したいという要望がある．また，今後放送音声の書き起こしや会議の議事録採取などへの応用を考える場合，句読点があると認識結果の可読性を高めることが出来る．

ここでは息継ぎ位置と句読点位置がある程度対応することを利用して，句読点を自動挿入する．具体的には句点および読点に対して「マル」，「テン」といった読みをあらわすモデルに加え，無音のモデルを割り当てておく．この結果，ポーズが挿入された部分は音響的には無音として認識されるが，言語的には句読点または"透過単語"（言語モデルにとっては何もないのと同じ扱いの単語）として処理され，言語的に可能性の高い方の系列が認識結果として出力されることになる．

5. 評価方法

日本語を対象とする場合，文字単位のような明確な単位を用いる場合は別として，１つの文を認識単位列に分割するやり方には曖昧さが残る．特に，ここで定義したような単語単位を用いた場合，１つの入力文にあてはまる単語列候補は多数存在し，入力単語数ですら一意には決められない．このため，ここではシステムの性能評価（誤認識率，単語カバレージ，パープレキシティ）を次のようにして行う．

5-1. 誤認識率

ディクテーションシステムを考えた場合には，入力後の修正作業も含めてその性能を評価できたほうがよい．その意味では認識結果の単語分割の正確さも重要な要因となりうるが，認識単位の異なるシステムを比較するには，単純に認識結果の表記上の誤り，つまり文字誤り率（*CER*）と文字正解精度（*Accuracy*）で認識性能を評価するのが望ましい．

$$CER = 100 \cdot \frac{Ins. + Del. + Sub.}{N}$$

$$Accuracy = 100 - CER$$

なお，*Ins.*，*Del.*，*Sub.*，はそれぞれ文字の挿入，脱落，置換誤りを，*N*は入力文字数をあらわす．

5-2. 単語カバレージとパープレキシティ

音響上の性能も含めた場合には，上記文字誤り率がよい指標となるが，言語モデルの観点からは，単語カバレージやパープレキシティも調査しておきたい．

ここでは，認識時と同じTri-gram言語モデルを使用し，言語モデルの観点から認識対象文を生成する確率のもっとも高いパス（最適パス）を探索し，そのパス上の単語列によって単語カバレージおよび単語あたりのパープレキシティを評価する．ただし，最大化したのは文としての尤度であり，単語あたりのパープレキシティを最小化したのではない．単語誤り率を推定する必要がある場合には，最適パス上の単語列を仮に正解単語列としている．

5-3. 未知語モデルとN-gramモデルによる単語の自動分割

上記のように，N-gramモデルを使って最適な単語系列を推定するためには未知語を何らかの方法で統計的にモデル化する必要がある．統計的な未知語モデルとしては文字N-gramモデルを用いる方法がすでに提案されている[5,6]．ただ，状態数が多くなり，その推定には非常に多くの訓練データが必要となることから，ここでは文字種（漢字，ひらがな，カタカナ）をクラスとみなした以下のモデルで表現することにした[7]．ここでcは未知語の文字，gは文字種のクラスをあらわす．

$$\begin{aligned}\Pr(c_1c_2\cdots c_n) &= \Pr(c_1c_2\cdots c_n \mid g_1g_2\cdots g_n)\Pr(g_1g_2\cdots g_n) \\ &\cong \prod_{i=1}^{n}\Pr(c_i \mid g_i)\Pr(g_1g_2\cdots g_n)\end{aligned}$$

なお，N-gramモデルによる単語の自動分割は，未知語の自動抽出や，既存の形態素解析プログラムに依存しない言語モデルの構築手段を提供するという意味でも重要である．

表3 40K語彙のカバレージとパープレキシティ

	Coverage(%)		Perplexity	
	Disc.	Cont.	Disc.	Cont.
日経新聞	98.0	99.5	92.2	73.9
産経新聞	95.5	97.6	166.7	170.7
電子会議室	95.2	97.3	246.8	225.6
毎日新聞	96.1	98.0	138.6	138.9
ビジネストーク	97.5	98.8	122.4	113.9
小説	94.6	96.5	196.4	189.3

6. 実験結果

6-1. カバレージとパープレキシティ

6種類のタスクに対し，40K語彙のカバレージおよびパープレキシティを，先の定義に基づいて調査した．結果を表3に示す．ここで，毎日新聞，ビジネストークおよび小説は言語モデルの学習時には用いておらず，学習領域外のデータである．

表中の*Cont.*は5-2.で述べた方法で自動推定された単語系列に対する結果を示しており，言語モデルを用いて連続音声認識を行う場合に相当すると考えられる．また，離散単語認識との比較のため，同じ文章を人間が単語単位に切り出した場合のカバレージとパープレキシティも示す（*Disc.*と表示）．

予想どおりこの定義では，連続音声認識においてカバレージが一様に改善されることが分かる．一方，単語パープレキシティは離散に比べ若干増大する場合も見られるが，さほど大きな変化ではなく，離散に対する場合とタスクごとのパープレキシティの順序関係は変化していない．

6-2. 不特定話者連続音声認識

訓練時とは異なる20代～50代の男女各10名（20代3名：話者番号1-3，30代4名：話者番号4-7，40代2名：話者番号8-9，50代1名：話者番号10）が，それぞれ30文ずつ読み上げた合計600文（入力総数=15,768文字：8,252単語，単語Perplexity=148.3）を用いて認識実験を行った．なお，この実験では全体の4/5の文章に対しては句読点も読み上げるように指示しており，それ以外の文章に対しても句読点の自動挿入は行っていない．内容としては，新聞，雑誌，電子会議室の発言をそれぞれ1/3ずつ含む．結果を，単語あたりのパープレキシティ，単語誤り率（*WER*）とともに表4に示す．なお，音響モデルの訓練には先に示した学習データ，Set-25K，Set-100Kの両方を用い，約60K個のプロトタイプ（正規分布）を推定した．

表4 不特定話者連続音声認識実験結果

話者	Perplexity	Error Rate(%)				
		Sub.	Ins.	Del.	CER	WER
男性1	168.5	2.89	0.52	0.26	3.68	5.60
男性2	125.6	1.34	0.36	0.48	2.19	3.44
男性3	126.7	2.53	0.0	0.80	3.33	3.64
男性4	124.1	1.95	0.12	0.24	2.32	2.91
男性5	218.9	0.87	0.12	0.50	1.50	1.23
男性6	168.5	1.97	0.52	0.13	2.63	4.14
男性7	125.6	1.7	0.60	0.24	2.56	4.58
男性8	126.7	3.46	0.40	0.13	4.00	5.58
男性9	124.1	1.95	0.12	0.36	2.44	3.40
男性10	218.9	1.37	0.25	0.37	2.00	2.21
男性平均	148.3	1.98	0.30	0.35	2.64	3.68
女性1	168.5	1.30	0	0.26	1.58	2.68
女性2	125.6	2.56	0.12	0.36	3.04	4.81
女性3	126.7	7.73	2.26	0.13	10.1	12.6
女性4	124.1	1.46	0.12	0.24	1.83	4.13
女性5	218.9	0.50	0.12	0.25	0.87	0.24
女性6	168.5	3.29	0.26	1.05	4.61	6.34
女性7	125.6	1.70	0.24	0.24	2.19	3.89
女性8	126.7	2.13	0.66	0.93	3.73	4.61
女性9	124.1	1.96	0.36	0.61	2.93	4.62
女性10	218.9	2.25	0.50	0.75	3.50	4.18
女性平均	148.3	2.45	0.45	0.48	3.39	4.81
全体平均	148.3	2.22	0.38	0.42	3.02	4.25

表5 パープレキシティの範囲とその範囲に収まるテスト文に対する文字誤り率

Perplexity	該当文数	CER(%)
40-80	88	2.20
81-120	140	2.02
121-160	108	2.97
161-200	72	3.57
201-240	72	3.64
241-	120	4.21

このタスクにおける一単語あたりの平均文字長は1.91であり，平均文字誤り率から推定される単語正解精度は$0.97^{1.91}$=0.943であるが，誤りの分布に偏りがあるため，それよりは高い単語正解精度が得られている．20名の話者の中では1名だけ，10%を超える文字誤り率を示した話者がいるが，それ以外の話者についてはおおむね良好な結果が得られており，性別，年齢あるいはパープレキシティと，誤り率の関連は見出せない．また，この表にはないが，発声速度も人によって大きな違いがあった（1.3～2.2単語/秒）が，その影響も見られない．なお，認識率が特に悪かった話者について

も，音声品質には問題がなく，ただ舌足らずな発声が特徴的な話者であった．

6-3. パープレキシティと文字誤り率の関係

表４について，単語パープレキシティと文字誤り率(*CER*)の関係を，パープレキシティが一定範囲内に収まる文ごとに集計した結果が表５である．表４を見る限りではパープレキシティと*CER*の相関は非常に低いように思われたが，話者の要因を取り除くと，この表のようにおおむね単調な相関関係が見て取れた．ただ，実際に言語モデルの能力が認識率に影響を及ぼすのは，パープレキシティとして求まるような平均分岐数ではなく，局所的に出現確率が低下する部分であり，今後そのような局所的な統計値との対応関係も調べる必要があると考えている．

6-4. 学習データ量とモデルの複雑さの影響

次に，音響モデルの学習データ量およびモデルの複雑さが認識精度に与える影響について調査した．結果を表６に示す．なお，テストデータは，表４と同じ物である．音素環境依存モデルの数はいずれの場合も約3,000に固定したため，プロトタイプ数30Kが，ほぼ正規分布の混合数10に，60Kが混合数20の場合に相当する．この結果を見る限りでは混合数を増加させるメリットは学習データ量によらずあまりない．一方，モデルの学習データ量については，215名から得られたSet-25K（計25,800文，355,454単語）では明らかに不足であり，658名から得たSet-100K（計97,384文，1,338,307単語）を追加することで，認識率が顕著に改善されたことが分かる．

ただ，表４の結果に見られるように，突出して誤り率の高い話者がまだ存在することから，学習データ量についてはさらなる増量が必要であると思われる．

6-5. 句読点の自動挿入実験

話者4名が句読点のある文章を句読点は含まずに読み上げた合計120文を認識させて，句読点の自動挿入実験を行った．なお，句読点で息継ぎをする等の指示は与えず，あくまでも自然に読み上げさせている．

実験の結果，句点に関しては120個の指定に対し，121個が検出され，ほぼ100%正解を得ることが出来たが，これは文の間には十分長いポーズを置くように指示したために，文の終わりである可能性が別途音響的に検出されていたことにも起因する．一方，読み上げ文に記載されていた読点は52個所であったが，134個の読点を自動検出した．なお，文中のポーズは232個所で検出されている．脱落は皆無であったものの，明らかに多くの読点が挿入されていることになる．ただし，文章としては特に不自然になるような挿入ではない．

表6 学習データ量とプロトタイプ数の影響

学習データ	*プロトタイプ数*	*CER(%)*
Set-25K	30K	5.0
	60K	4.9
Set-25K +Set-100K	30K	3.2
	60K	3.0

7. おわりに

人間の考える単語単位を認識単位とし，連続音声を認識対象とした不特定話者用の日本語ディクテーションシステムの実現可能性について検討した．

認識単位間の比較に関しては，単位によってカバレージや単位長さが異なり，また自動解析精度にも差が出るため，言語モデルや音響モデルまで含めた正確な単位間の性能比較は非常に難しい．このため，この認識単位が他の認識単位と比べて特に優れているということを客観的に示す証拠は必ずしも多くはない．しかしながら，ここで示した実験結果からは，言語モデルとしての性能は他の形態素単位などとは大差ないものの，その一方で，単語カバレージに優れ，また，一文を構成する単位数が少なく，単位あたりの平均モーラ長が長いなどの特徴が明らかになった．

実際，この認識単位を用いた不特定話者大語彙認識実験の結果は良好なもので，約4万語の語彙からなるタスクにおいて，Perplexity=148.3のテスト文に対し，文字正解精度は97%，単語正解精度も95.7%が得られた．

また，連続音声認識時の言語モデルの性能評価のために，未知語モデルとN-gramモデルによる単語の自動分割手法を導入した．この手法は既存の単語N-gramモデルを生かし，新たなコーパスを単語単位に自動分割する方法としても重要である．

さらに，ポーズ位置と句読点の挿入位置にある程度対応がつくことを利用して句読点の自動挿入を試み，おおむね良好な結果が得られることを示した．今後放送音声の書き起こし等に応用したいと考えている．

謝辞

データ使用を許可してくださった，産経新聞社，日本経済新聞社，毎日新聞社（CD-毎日新聞94）ならびに(株)ピープルワールドカンパニーに感謝します．

参考文献

[1] J.L.Gauvain, L.Lamel, "Large vocabulary continuous speech recognition: From large vocabulary systems towards real-world applications," 電子情報通信学会論文誌，D-II, Vol.J79-D-II, No.12, pp.2005-2021, 1996.12.

[2] 西村，伊東，"単語を認識単位とした日本語ディクテーションシステム," 電子情報通信学会論文誌, D-II, Vol. J81-D-II, No.1, pp.1-8, 1998.1.

[3] 西村，伊東，山崎，荻野，"単語を認識単位とした日本語大語彙連続音声認識," 日本音響学会平成9年度秋季研究発表会，3-1-5, pp95-96, 1997.9.

[4] 丸山，荻野，"正規文法に基づく日本語形態素解析"，情報処理学会論文誌，35-7，pp.1293-1299, 1994.

[5] 永田，"単語頻度の期待値に基づく未知語の自動収集," 情報処理学会自然言語処理研究会，116-3, pp. 13-20, 1996.

[6] 森，山路，"日本語の情報量の上限の推定," 情報処理学会論文誌, Vol. 38, No. 11, 1997.11.

[7] 伊東，西村，"N-gram を用いた日本語テキストの単語単位への分割," 情報処理学会自然言語処理研究会，122-9，pp.57-62，1997.11.

[8] L.R.Bahl et al., "Performance of the IBM large vocabulary continuous speech recognition system on the ARPA wall street journal task," Proc. ICASSP'95, pp.41-44, 1995.