文法カテゴリ対制約を用いたA*探索に基づく大語彙連続音声認識パーザ

李 晃伸† 河原達也† 堂下修司†

文法に基づく連続音声認識（パージング）において，大語彙の条件下で効率の良いA*探索を実現するための手法を提案する．大語彙においては探索時に仮説の単語ネットワークが爆発するため，広く用いられている1パスのビーム探索では，ビーム幅を大きくとる必要があり効率が悪い．また文法による次単語予測のみでは候補の絞り込みが不十分である．これに対して，(1) 元の文法から抽出したコンパクトな単語対制約によって仮説ネットワークの大きさを抑え，かつ (2) 文法カテゴリごとに単語辞書を木構造化することで，効率的に強力なヒューリスティック計算を行う．さらに，(3) この第1パスの結果をインデックス化しその音響的照合結果から展開単語を絞り込むことで，大語彙で効率の良いA*探索を実現する．この手法を実装した汎用連続音声認識パーザ Julian を，5000語クラスの文法タスクでの認識実験において，標準的な1パスビーム探索のデコーダと比較した．その結果，本手法は大語彙ではるかに少ない計算量で探索が行え，また構文の複雑さによらずどのような文法でも安定して動作した．最終的に，実時間の2.2倍程度の処理時間で91.4%の単語認識精度を達成した．

Large Vocabulary Continuous Speech Recognition Parser Based on A* Search Using Grammar Category-pair Constraint

AKINOBU LEE,† TATSUYA KAWAHARA† and SHUJI DOSHITA†

We address an efficient A* search algorithm for grammar-based large vocabulary continuous speech recognition. While grammars can introduce long-distance constraint into search, the expanded word hypothesis network grows huge under large vocabulary. So convensional one-pass beam search needs extremely wide beam width to get optimum results. We propose an efficient two-pass search algorithm by (1) using word-pair constraint as heuristics and (2) tree-organizing the word lexicon for each grammar category, to represent the whole network in a compact loop structure. Futhermore, (3) the survived words on the first pass are indexed to eliminate candidates to be accessed on the second pass. We depeloped a portable FSA-based CSR parser named Julian and compared the performance with a typical one-pass beam decoder on 5,000-word task. Experimental results show that the proposed method achieves high accuracy with far less computation, and works stably with even more complex grammars. Finally, our parser archieved a word accuracy of 91.2% with process time of 2.5 times the real time.

1. はじめに

数千語から数万語の大語彙を対象とする連続音声認識は困難な問題の1つであり，特に効率良く最適解を見つける探索アルゴリズムが要求される．

大語彙の連続音声認識で近年さかんに用いられているのは，単語N-gramに代表される統計的言語モデルに基づくアプローチ[1)~3)]であるが，統計的に言語モデルを推定するには大量の整ったテキストコーパスが必要であるため，新聞記事などの大規模コーパスから汎用的なディクテーションシステムを作成することが多い．一方，情報検索や予約システムなどの音声インタフェースにおいては，タスクやドメインを反映した言語モデルが必要である．

しかし，タスクごとに大量のデータを収集しラベリングを行うのは大変な労力を要するので，統計的なモデルの構築は容易でない．また地名や商品名のように統計的なモデルにあまり意味がない場合もある．このため人手で文法や語彙を指定するほうがタスクに特化したモデルの構築が容易であり，語彙の入れ替えなどの変更が簡単に行える利点がある．

† 京都大学大学院情報学研究科知能情報学専攻
Graduate School of Informatics, Kyoto University

本論文では，大語彙の条件下における文法☆ベースの音声認識（パージング）アルゴリズムを研究の対象とする．パージングについては，これまで主に数百から千語程度の小中語彙での検証しか行われていない[4)~6)]．

大語彙のタスクでは，単語辞書が語彙に比例して巨大化する．また文法に従って仮説を展開すると単語パープレキシティの増大にともなって仮説ネットワークが爆発する．このため単純な1パスのビーム探索ではビーム幅を非常に大きくとる必要があり，効率が悪い．

これに対して，単語対制約をヒューリスティックとする A*探索[4)] を大語彙へ適用することを考える．中語彙のタスクでは解の最適性を重視してヒューリスティック計算を行っていたが，大語彙においてはこのヒューリスティック計算を候補の予備選択と位置づけ，全体として効率の良い探索アルゴリズムの実現を目指す．処理は2パスで構成され，第1パスでは，文法カテゴリ対制約（単語対制約と等価）を用いて探索空間である仮説ネットワークをコンパクトに抑える．その認識結果に基づいて第2パスで再探索を行うことで，効率良く高精度な解が得られる．さらに効率化のために，第1パスのビーム探索化，文法カテゴリ単位での木構造化，および第2パスでの音響的照合に基づく展開単語の絞り込みを導入する．

以下，まず大語彙における問題点，および従来手法である1パスビーム探索と単語対制約をヒューリスティックとする A*探索について述べる．そして A*探索を大語彙で適用するための手法について述べ，具体的アルゴリズムを示す．これを実装した大語彙連続音声認識パーザ Julian の仕様を述べた後，1パスビーム探索と 5000 語のタスクにおける認識実験による比較評価の結果を報告する．

2. 記述文法に基づく大語彙連続音声認識

記述文法に基づく連続音声認識の大語彙における問題点，および探索法として1パスビーム探索と基本的な A*探索について述べる．

一般に連続音声認識は，与えられた言語的・音響的制約のもとで，最も確率の高い単語列からなる文候補を見つけ出す探索問題として定式化される．言語制約として文法を用いる場合の探索空間は，その文法および単語辞書をオートマトンとして展開した文仮説のネットワークとなる．有限状態文法を用いる場合は，すべての規則をオートマトンとしてあらかじめ展開しておくことが可能である．また文脈自由文法の場合は，探索時にパーザにより次単語を逐次的に予測しつつ，必要に応じて展開する．

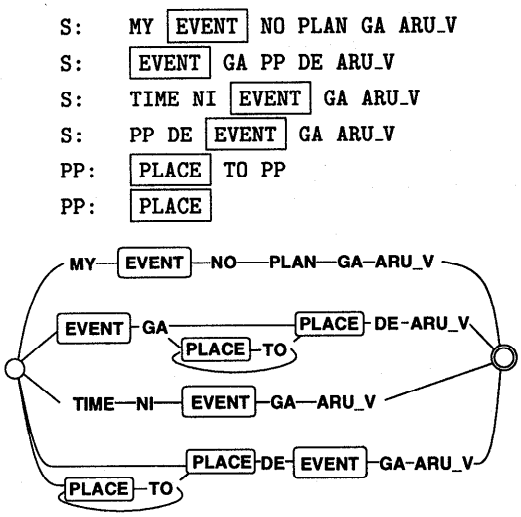


図1 文法から展開される単語仮説ネットワークの例
Fig. 1 Example of word hypothesis network expanded from a grammar.

2.1 大語彙の問題

大語彙においては，一般にパープレキシティが増大し，探索時に多くの仮説を扱う必要が生じる．特に文法ベースの連続音声認識においては，以下の要因から単語仮説ネットワークが爆発し，探索が困難となる．

文法カテゴリの多重性 単語 N-gram モデルが1，2単語の履歴しか考慮しない単純なモデルであるのに対して，文法では文全体に対する履歴を考慮した制約が与えられる．そのため大語彙においては仮説ネットワークの大きさが爆発的に増大しやすい．図1に文法と展開後の仮説ネットワークの例を示す．なお文法は単語のカテゴリのみを記述してあり，実際には各パス上でカテゴリに含まれる全単語が展開される．

このように文法には，コンテキストが異なるものの局所的には同じ構文規則が頻繁に含まれるため，同じ単語が類似した位置に何度も展開される（図1の EVENT や PLACE など）．特に大語彙のタスクでは，地名や商品名といった特定のカテゴリに大量の単語が登録される傾向が強いが，このカテゴリが多重に展開されると仮説ネットワークの増大が著しい．

文法においてカテゴリが重複せずに，展開後のネットワークが小さくなるようなコンパクトな文法を意識して記述することも可能であるが，文法を作成するア

☆ 以下，本論文では特に記述がない場合，文法とは，統計的言語モデルに対する文脈自由文法や有限状態文法などのルールベースの記述文法をさす．

プリケーション開発者が，つねに文法の専門家であるとは限らない．また展開後のネットワーク（オートマトン）の最小化問題はNP困難であり，コンパイラなどによる自動最小化も限界がある．このように文法ノードの多重性は避けがたい問題である．

単語辞書の巨大化 認識用の単語辞書は語彙サイズに比例して巨大化する．このためこれを文法に従って展開した単語仮説ネットワーク全体の大きさも，これに比例して増大する．

これに対して，同じプレフィックスを持つ単語どうしで状態を共有することで，単語辞書の大きさを抑えることができる[7]．この木構造化の効果は語彙が大きくなるに従って大きくなり，単語N-gramベースの大語彙連続音声認識では不可欠である．ただし文法を用いたパージングでは，単純に全体で1つの木を構成すると，異なる構文制約を持つ単語どうしで単語の先頭（=木の根ノード）を共有するため，静的なネットワークで文法制約を表現できないという問題が生じる．探索時に動的に文法制約を参照しながら仮説展開を行うのは複雑な制御機構を必要とし，特に大語彙では処理のオーバヘッドが大きい．

不十分な候補の絞り込み 文法ベースの探索では，単語N-gramと異なり，文法制約から展開単語を決定的に絞り込むことができる．しかし語彙サイズが大きくなるにつれて，特に地名などのように特定の文法カテゴリの語彙が大きい場合は，この絞り込みが十分に機能しない．展開のたびに多くの単語を照合する必要から，計算コストは文法の単語パープレキシティに比例して増大する．

これらの問題は，特に語彙サイズの増大にともない顕著になり，複合的に作用して探索を困難にする．

2.2 1パスビーム探索

一般に広く用いられている1パスのビーム探索法[5),8)]においては，探索時点までのスコアに基づいて枝刈りをしながら探索を進める．そのため局所的なノイズや類似性の影響を受けやすく，最適解が探索途上で失われないようにするためにはビーム幅を大きくする必要がある．

また探索の幅優先的性質から，処理量は文法のパープレキシティに比例する．特に前述のような文法の多重性によってあちこちの文法ノードに同じ単語が出現すると，評価値の高い同じ単語がビームの上位を占めるため，これもビーム幅を大きくさせる要因となる．

2.3 A* 探索

これに対してA*探索は，best-first探索の一種であり，評価値の最も高い仮説を展開することで探索を進める．仮説の評価値には，未探索部分のスコアのヒューリスティックな推定値を加える．すなわち，仮説 n について，その評価値 $\hat{f}(n)$ を次のように定義する．

$$\hat{f}(n) = g(n) + \hat{h}(n) \tag{1}$$

ただし，$g(n)$ はすでに展開された区間のスコア（対数尤度），$\hat{h}(n)$ は未展開部分のヒューリスティックな推定スコアである．このとき，最適解が必ず得られるようにするためには，$\hat{h}(n)$ を実際のスコア $h(n)$ より厳しくしない，つまり

$$|\hat{h}(n)| \leq |h(n)| \tag{2}$$

という条件（A*適格性）が必要である．また，できる限り無駄な仮説を展開することなく最適解を早く見つけるためには，この推定スコア $\hat{h}(n)$ ができるだけ実際の値に近いことが望ましい．

トリートレリス探索[9)]においては，$h(n)$ を求めるために，探索の前処理として探索とは逆方向に認識処理を行い，そのトレリス上の評価値を保存しておく．そして探索時にこの評価値をヒューリスティックとして用いる．つまり認識処理は，ヒューリスティックの計算とパージングの2パスから構成される．

我々はこれまでに，このヒューリスティックの制約として単語対文法を用いることを提案した[4)]．これは元の文法から単語間の接続に関する情報のみを抽出したものである．元の文法のサブセット（言語的にはスーパセット）であり，探索に用いる制約よりも弱いためA*適格性を満たす．また単純な単語連接や音素連接の制約よりも強力であり，良いヒューリスティックとなる．

このA*探索は，これまでに小・中語彙のタスクにおいて有効性が示されている[4),10)]．しかしA*探索において適格性を厳密に満たした探索を行うには，パージング中に出現しうる全単語について，あらかじめヒューリスティック評価値を計算しておく必要がある．そのため第1パスでは基本的に全探索となり，大語彙ではヒューリスティックの計算量からこのままでは適用は不可能である．

3. 大語彙におけるA*探索の実現

A*探索を大語彙で実現するためには，ヒューリスティックの計算量の増大が大きな問題となる．これは1パスのビーム探索における計算量増大の問題と基本的に同じであるが，異なるのは，A*探索ではマルチパスで処理を行うことにより，第1パスで近似を導入しても，第2パスでそのエラーの回復が行えることで

ある．すなわち，ヒューリスティック計算のプロセスを単語候補の予備選択と見なして，コンパクトな制約によって候補の絞り込みを行い，その結果に基づいて第2パスで再探索することによって最終的に高精度で効率も良い探索を実現する．ヒューリスティックが厳密なA*適格性を満たさなくなるが，十分な候補を後段へ残せば実際の認識において大きな問題にはならないと考えられる．

具体的には，ヒューリスティック計算におけるビーム幅の設定のほかに，2.1節で述べた大語彙での各問題点に対して以下のアプローチをとる．

3.1 単語対制約による束ね効果

単語対制約をヒューリスティックのための制約として用いることで，ネットワークの増大を大幅に抑えられる．1パスビーム探索では，巨大な探索ネットワークを1回で走査するため探索効率が悪い．一方，単語対制約による文法ネットワークは単純なループで表現されるため，文法上で異なるノードの同一単語がネットワーク上に複数現れることがない．またどのような文法からも安定して語彙サイズに比例したノードからなるコンパクトなネットワークが得られる．またこの単語対制約による認識処理結果をヒューリスティックとして再び第2パスで文法による探索を行うことで，最終的には同じ制約による結果が得られる．

単語対制約によって同じ仮説単語を1つにまとめることは，仮説の束ね効果の一種といえる．束ね処理に関しては，1パスのフレーム同期ビーム探索において，伊藤ら[11]が音韻レベルで，渡辺ら[12]が単語レベルで提案した．どちらも近傍に出現する同一の音韻もしくは単語の照合は1つに対してのみ行い，他の候補は近似的にその結果を再利用する．しかしこれらは束ね処理を照合と別に行う機構が必要であるのに比べて，単語対制約はネットワークそのものをコンパクトにするため特別な機構は必要なく，探索の制御が単純である．また認識結果が束ねによる誤差を含むのに対して，A*探索では第2パスの再探索でより正確な結果を得ることができる利点がある．

3.2 単語辞書のカテゴリ単位の木構造化

単語辞書に関しては，制約の単位を単語ではなく文法のカテゴリと見なし，各々のカテゴリ内で個別に木構造化を行うことで，単語対制約全体を静的なネットワークで表現できる[13]．カテゴリごとに木構造化した文法カテゴリ対ネットワークの例を**図2**に示す．

制約の適用方法に関しては，語彙全体で単一の木を構築し，探索時に単語対近似[14]によって必要な部分を多重化しながら動的に仮説展開を行う方法も考えられる．しかし静的なネットワークに展開しておくことで，探索時に文法制約を参照しながら展開を行うための処理のオーバヘッドを避けられる．

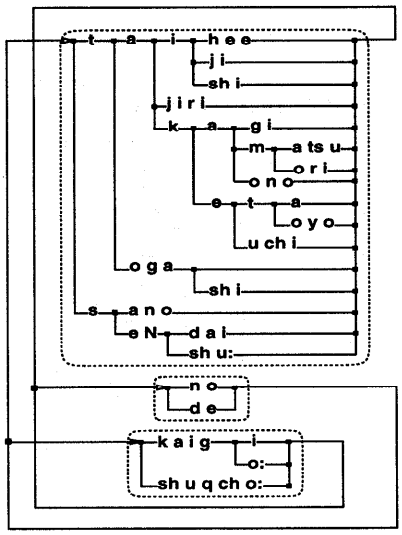


図2 木構造化された文法カテゴリ対ネットワーク
Fig. 2 Tree-organized category-pair network.

カテゴリ単位の木構造化による探索空間の削減は，特に地名などのように1カテゴリに多数の単語が登録される場合に効果が大きい．逆にカテゴリが細分化され1カテゴリあたりの単語が少ない場合は，木構造化の効果が小さくなるが，それだけ文法による絞り込みの効果が大きくなるので，全体として処理量に大きな差はない．

3.3 音響的照合による絞り込みの導入

大語彙においては文法による単語予測のみでは十分に候補を絞れない．これに対して，第1パスのヒューリスティック計算の結果のトレリスに，フレームごとにビーム内に残った単語のインデックスを付加する[3]．第2パスの探索における仮説の展開の際には，文法制約に加えてこの予備選択された単語のインデックスを参照することで，高速に照合を行える．

地名などのように，そのカテゴリだけで語彙のほとんどを占める場合に，この音響的照合からの絞り込みは特に効果的であると考えられる．

3.4 パージングアルゴリズム

このA*探索に基づく大語彙連続音声認識は，**図3**に示すように2パスで構成される．第1パスは探索の前処理であり，文法カテゴリ対制約に従って全入力に対してフレーム同期にleft-to-rightに認識処理を行い，途中の各フレームごとに，ビーム内に残った単語の終端の評価値およびそのインデックスをトレリスの

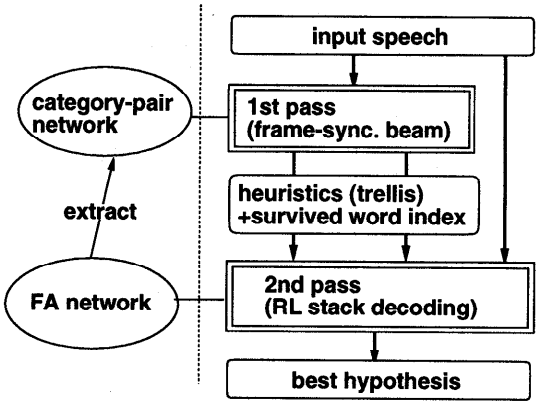


図 3 探索アルゴリズムの構成
Fig. 3 Search algorithm overview.

形で保存しておく．第2パスでは文法に従ってスタックデコーディング探索を行うが，その際探索を逆の right-to-left 方向に進め，仮説の未探索部分のヒューリスティックとして，対応する第1パスのトレリスを接続する．

以下に探索（第2パス）の具体的アルゴリズムを示す．ここで部分文仮説を $W_n = w_n, w_{n-1}, ..., w_1$（$w_1$ が文末）と表し，入力長を T とする．時刻 T から t までの文仮説 W_n の後向きの尤度を $g(W_n, t)$，時刻 t に単語 w で終わる前向きヒューリスティックを $h(w, t)$，文仮説 W_n の評価値を $f(W_n)$ とそれぞれ表す．また第1パスで得られた，フレーム t においてビーム内に残った単語の集合を $index(t)$ とする．

(1) 文末に出現しうる各単語 w_1 について，1単語からなる部分文仮説 W_1 を生成し，以下の評価値を計算して仮説スタックに入れる．

$$f(W_1) = h(w_1, T)$$

(2) 仮説スタックから評価値の最も高い部分文仮説 W_n を取り出す．W_n が受理状態にありかつ入力始端に達していれば，解として出力し探索を終了．

(3) 仮説 W_n の最終単語 w_n に対する HMM を構成し，その（後向き）トレリスを展開する．

$$g(W_n, t) = \begin{cases} \max_{t'} \{\beta(w_n, t, t') + g(W_{n-1}, t')\} & \text{if } n > 1 \\ \beta(w_n, t, T) & \text{if } n = 1 \end{cases}$$

ただし，$\beta(w_n, t, t')$ は単語 w の t' から t までの後向き尤度である．

(4) 文法制約上，仮説 W_n に接続しうる単語のうち，第1パスのインデックスに残っていたもののみについて，W_n に接続して新たな仮説 W_{n+1} を生成する．新たな仮説の評価値は次のように求める．

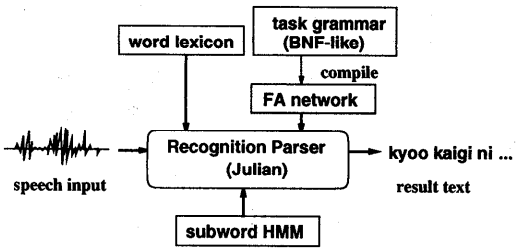


図 4 連続音声認識パーザ Julian を用いた認識システム
Fig. 4 Recognition system with CSR parser Julian.

$$f(W_{n+1}) = \max_{t:w_{n+1} \in index(t)} \{h(w_{n+1}, t) + g(W_n, t)\}$$

(5) 生成したすべての W_{n+1} を仮説スタックに入れる．(2) へ戻る．

4. 大語彙連続音声認識パーザ Julian

前章の探索アルゴリズムを，大語彙に適用可能な汎用の連続音声認識パーザとして実装した．京都大学における名称は Julian である．これに文法および音響モデルを与えることで，任意のタスクドメインに対する認識システムが構成できる．Julian を用いた連続音声認識システムの構成例を図4に示す．

文法と単語辞書は別々に用意する．文法は，単語のカテゴリを終端記号として BNF 形式で記述する．単語辞書は，各文法カテゴリごとに，それに属する単語の名前およびサブワード（音素）列を列挙する．文法は専用コンパイラによって有限状態オートマトン（FA）に変換される．BNF 形式では文脈自由文法のクラスまで記述可能であるため，コンパイル時に正規文法のクラスに収まるかどうかがチェックされる．再帰性は右再帰性のみ扱える．また文法カテゴリ対制約は，パーザの実行時に自動抽出される．

音響モデルは，HTK[15] のフォーマットに対応している．使用可能な HMM の型は連続混合型で，モデル数・全状態数・混合数について任意のモデルを扱える．また triphone などの音素環境依存モデルを扱うこともできる．このとき環境依存性は，扱いの簡便さから第1パスでは単語内のみ扱い，第2パスで単語間まで考慮する．

このようなインタフェースにより，Julian は高いポータビリティを実現しており，様々なアプリケーションに容易に適用することができる．

5. 実験的評価

同一の音響モデルおよび等価な文法を用いて，等しい言語的・音響的制約のもとで認識実験を行い，探索

- 明日の2時から3時まで、二研で音声研究会を開きたい。
- 金曜日の予算会議の場所を小会議室に変更したい。
- 今日のセミナーは何時からですか？
- では、16時からにします。

図5 評価に用いた文例
Fig. 5 Example of testset sentencesl.

表1 タスク文法の諸元
Table 1 Specification of task grammars.

grammar	lexicon size	word perp.	FA nodes Julian	FA nodes HTK
PG1	806	91.6	257	280
PG2	4439	257.1	257	280
SG1	833	28.7	5061	2849
SG2	5023	76.0	5061	2849

FA 状態数：有限状態文法に変換後のノード数

アルゴリズムの精度と効率を調べる．

比較対象として，フレーム同期の1パスビーム探索を行う標準的なソフトウェアであるHTKの連続音声認識モジュールHViteを用いた．

5.1 実験条件

タスクドメインは個人スケジュール管理であり，スケジュールの登録・削除・変更・問合せに関する発話を認識対象とする．テストセットとして男性8名によって発声された50種のサンプル文の計400サンプルを用いた．文の平均発話長は3.2秒であり，1文あたり平均6.2単語からなる．文例を**図5**に示す．

タスク文法として，任意フレーズの繰返しを許すフレーズ単位の文法PGと，フレーズの文中での位置を固定的に記述した文単位の文法SGの2種類を記述した．またそれぞれについて語彙サイズの異なる2種類の辞書を用意した．各文法の諸元を**表1**に示す．

PG1/SG1はタスクドメインで典型的に出現しうるフレーズを人間が1つずつ登録した文法であり，PG2/SG2はそれに辞書から抽出した名詞（特に地名・時間を表す名詞）や動詞を追加して語彙を増やしたものである．SG1，SG2は構文制約を発話のパターンごとに文単位で文法を記述した複雑な文法であり，パープレキシティは小さいが文法ノード数はPG2の約10倍となっている．

音響モデルはmonophone HMMおよびtriphone HMMを用いる．日本音響学会研究用連続音声データベースの男性話者30人分と日本音響学会新聞記事読み上げ音声コーパスのうち男性話者100人分の発声データで学習した不特定話者HMMである[16]．monophone HMMの総状態数は147，triphone HMMの総状態数は2110であり，それぞれ1状態あたり16混合分布を持つ．

5.2 評価基準

評価値として，認識率とビーム幅を用いる．ここでビーム幅とは，1フレームごとにViterbi計算を行うHMMの（平均）状態数と定義する．

HTKについては，音素モデル数によるビーム設定値をHMM状態数に換算した値を用いる．実際には使用した音素モデルはすべて3状態からなるので，上記設定値を3倍した値がここでのビーム幅となる．

2パス探索のJulianについては，そのビーム幅$\hat{b}$を次式のように定める．

$$\hat{b} = b_1 + \hat{b_2} \tag{3}$$

ここでb_1は第1パスのヒューリスティック計算でのビーム幅であり，$\hat{b_2}$は第2パスの計算量から換算した仮想的なビーム幅である．$\hat{b_2}$は，パージングの際のすべての仮説に含まれるHMM状態数から以下の式によって正規化した値を求め，その全サンプルでの平均値を求める．ここで*npop*はある解が得られるまでに仮説が展開された回数，*nword*はその得られた文仮説の単語数，そして*avg_state*は1単語あたりの平均状態数である．

$$\hat{b_2} = \frac{npop}{nword} \times avg\_state \tag{4}$$

5.3 1パスビーム探索との比較

PG1およびPG2における1位候補の単語認識精度を，ビーム幅$\hat{b}$ごとに**図6**に示す．なお，音響モデルはmonophoneモデルを用いている．

Julianでは，HTKに比べてはるかに小さいビーム幅で同等の認識精度が得られた．1パスビーム探索のHTKでは，枝刈りの精度が劣るのに加えて，パープレキシティの増大にともなって単語ネットワークが組合せ的に巨大化するため，十分な精度を得るにはビーム幅をかなり大きくする必要があった．一方，A*探索のJulianでは，第1パスでよりコンパクトなカテゴリ対制約を用いることでビーム幅を抑えることができた．また強力なヒューリスティックによる深い先読みが行えるため，第2パスではほとんどのサンプルでbest-firstに解が求められた．また第2パスのビーム幅$\hat{b_2}$は第1パスの幅によらず24から26と，第1パスに比べてはるかに少なく抑えられた．

語彙数に関しては，PG1，PG2とも語彙サイズにほぼ比例した大きさのビーム幅が必要であった．このとこから，この差はさらに大語彙ではより顕著であろうと考えられる．

次に文法SG1，SG2で評価したところ，HTKはネットワークの爆発から動作が困難となり，安定した

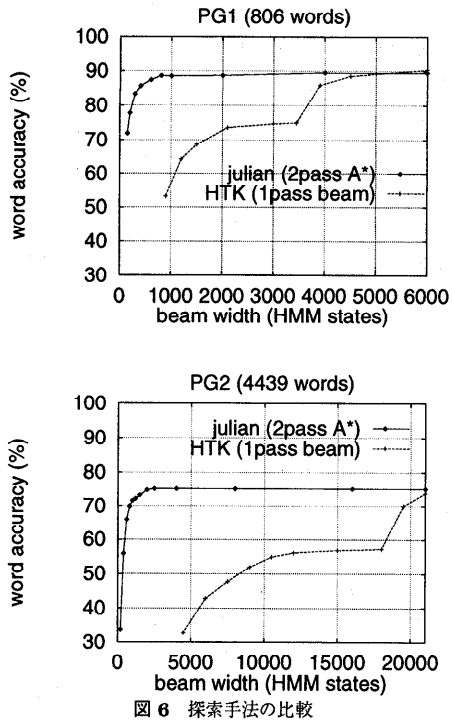


図 6 探索手法の比較
Fig. 6 Comparison of search algorithms.

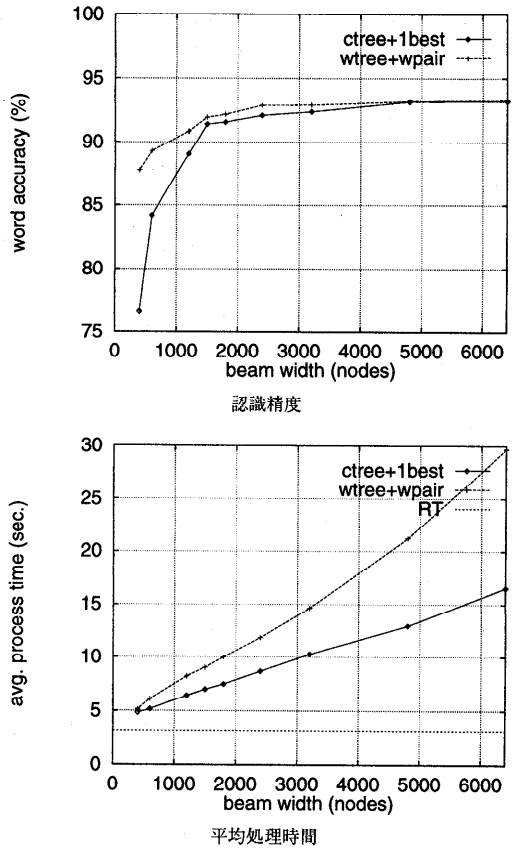
表 2 各文法における認識精度と平均処理時間
Table 2 Recognition accuracy and average process time for each grammar.

grammar	word acc. / time (sec.) Julian	HTK
PG1	88.6 / 3.2 (1000)	88.6 / 10.5 (4500)
PG2	74.7 / 6.2 (2000)	73.7 / 44.2 (21000)
SG1	97.1 / 3.4 (600)	— (—)
SG2	91.4 / 7.0 (1500)	— (—)

CPU：UltraSPARC 300 MHz
実時間：3.2 sec.
括弧内はビーム幅

結果が得られなかった．一方 Julian は安定して動作し，SG1 で 97.1%，SG2 で 91.4%の単語認識精度を示した．各文法における認識精度と平均実行時間の一覧を**表 2** に示す．一般に BNF などの文法記述からコンパクトな文法ネットワークを生成するのは容易ではないが，カテゴリ対文法は単純なため元文法の複雑さによらず安定して抽出できる．このため Julian はどのような文法に対しても安定して動作すると考えられる．

認識の実行時間に関しては，Julian は HTK に比べて同じ認識率を得るのに必要な平均処理時間がはるかに少なかった．SG1 においてほぼ実時間，SG2 においても実時間の 2.2 倍程度で解が得られた（表 2）．



（注）文法：SG2

図 7 文法カテゴリ対制約の静的展開と動的展開の比較
Fig. 7 Comparison of static and dynamic expansion of category-pair network.

5.4 文法カテゴリ対制約の展開方法の比較

第 1 パスの文法カテゴリ対制約の実装手法として，文法カテゴリ対ネットワークによる静的な展開方法と，語彙全体で単一の木を用いて単語対文法を直接駆動する動的な展開方法とを比較する．文法 SG2 での認識実験の結果を**図 7** に示す．ctree+1best が前者の静的展開，wtree+wpair が後者の動的展開を表す．十分なビーム幅において認識精度の差はほぼ 1%以内にとどまり，両実装法の間に精度上の大きな差は認められなかった．また木構造化の効果については，状態数が前者（カテゴリ単位）で 70.0%，後者（語彙全体）で 51.3% に削減された．小さいビーム幅で動的な展開の方が認識率の減衰が小さかったのは，ノード数が少ないため相対的なビーム幅が広くなるためであると考えられる．

一方，平均処理時間に関しては，探索時に展開を行

表 3 音素環境依存モデルを用いた認識性能
Table 3 Performance with context-dependent model.

	monophone	triphone
word acc. (%)	91.4	93.2
avg. time (sec.)	7.0	20.0
memory used (MB)	12	29

(注) 文法：SG2

わない文法カテゴリ対ネットワークの方が探索全体の平均処理時間を3分の2に抑えられた．ネットワークを動的に展開する方が同じビーム幅でも明らかに処理コストが大きいので，同じ処理時間で実現される認識率を比較すると静的なネットワークの方が優れているといえる[17),18)]．

5.5 音素環境依存モデルの効果

音素環境依存を考慮した triphone モデルを導入し，monophone モデルと認識精度および処理コストを比較した．結果を**表 3** に示す．文法 SG2 において認識誤りの 21% が改善され，認識精度は 1.8% 向上した．

一方，大規模な triphone モデルは大量のメモリを必要とする．また状態数が多いため出力確率計算のキャッシュが効きにくいことから出力確率計算のコストが高く，約3倍の処理時間を要した．

6. おわりに

有限状態文法に基づく連続音声認識において，A*探索を大語彙の条件下で実現し，評価を行った．5000語レベルの大語彙タスクによる認識実験の結果，文法カテゴリ対制約をヒューリスティックとする A*探索は

(1) 元文法から導出したコンパクトな文法カテゴリ対制約を用いた効率の良い探索
(2) その処理結果をヒューリスティック（先読み情報）とした高精度で best-first な再探索

の2パスの組合せによって，大語彙においても，単純な1パスのビーム探索に比べて小さいビーム幅で高精度な探索を行えることが示された．

語彙や複雑さの異なるいくつかの文法で比較した結果，語彙が大きくなるほど両者の性能の差が顕著に現れることが確認された．また文法カテゴリ対制約によるヒューリスティック計算は，タスク文法の規模や複雑さに対して頑健であり，どのような文法でも安定して動作することが示された．

実装した大語彙連続音声認識パーザ Julian は，5000語（パープレキシティは 76.0）のタスクにおいて，実時間の 2.2 倍程度の平均処理時間で 91.4%の単語認識精度を示した．また 1000 語のタスクではほぼ実時間で 97.1%を達成した．このプログラムは一般に公開される予定である．

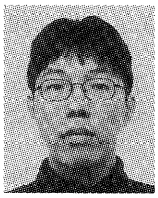
謝辞 音響モデルは IPA の「日本語ディクテーション基本ソフトウェア 97 年度版」のものを使用した．

参考文献

1) 西村雅史，伊東伸泰：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，Vol.J81-D-II, No.1, pp.10–17 (1998).
2) 松岡達雄，大附克年，森 岳至，古井貞熙，白井克彦：新聞記事データベースを用いた大語い連続音声認識，電子情報通信学会論文誌，Vol.J79-D-II, No.12, pp.2125–2131 (1996).
3) 李 晃伸，河原達也，堂下修司：単語トレリスインデックスを用いた大語彙連続音声認識エンジン JULIUS，信学技報，SP98-3 (1998).
4) 河原達也，松本真治，堂下修司：単語対制約をヒューリスティックスとする A* 探索に基づく会話音声認識，電子情報通信学会論文誌，Vol.J77-D-II, No.1, pp.1–8 (1994).
5) 北 研二，川端 豪，斎藤博昭：HMM 音韻認識と拡張 LR 構文解析法を用いた連続音声認識，情報処理学会論文誌，Vol.31, No.3, pp.472–480 (1990).
6) 伊田政樹，中川聖一：音声認識におけるビームサーチ法と A*探索法の比較，電子情報通信学会技術研究報告，SP96-12 (1996).
7) Klovstad, J.W. and Mondshein, L.F.: The CASPERS Linguistic Analysis System, *IEEE Sympo. Speech Recognition*, pp.234–240 (1974).
8) 中川聖一：文脈自由文法のフレーム同期型構文解析法による連続音声認識，電子情報通信学会論文誌，Vol.J70-D, No.5, pp.907–916 (1987).
9) Soong, F.K. and Huang, E.-F.: A tree-trellis based fast search for finding the N best sentence hypotheses in continuous speech recognition, *Proc. IEEE-ICASSP*, pp.705–708 (1991).
10) Paul, D.B.: An efficient A* stack decoder algorithm for continuous speech recognition with a stochastic language model, *Proc. IEEE-ICASSP*, Vol.1, pp.25–28 (1992).
11) 伊藤克亘，速水 悟，田中穂積：音素文脈依存モデルと高速な探索手法を用いた連続音声認識，電子情報通信学会論文誌，Vol.J75-DII, No.6, pp.1023–1030 (1992).
12) 渡辺隆夫，吉田和永，畑崎香一郎：バンドルサーチ法を用いた連続音声認識の高速化，電子情報通信学会論文誌，Vol.J75-DII, No.11, pp.1761–1769 (1992).
13) 李 晃伸，河原達也，堂下修司：A*探索に基づく大語彙連続音声認識，情報処理学会研究報告，96-SLP-11-4 (1996).
14) Schwartz, R., et al.: A Comparison of Several Approximate Algorithms for Finding Multiple

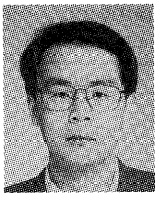
(N-best) Sentence Hypotheses, *Proc. ICASSP*, Vol.1, pp.701–704 (1991).
15) Young, S., Jansen, J. and Woodland, J.D.P.: *The HTK Book* (1995).
16) 武田一哉，峯松信明，伊藤彰則，伊藤克亘，宇津呂武仁，河原達也，小林哲則，清水 徹，田本真詞，荒井和博，山本幹雄，竹沢寿幸，松岡達雄，鹿野清宏：大語彙日本語連続音声認識研究基盤の整備—汎用音素モデルの作成，情報処理学会研究報告，97-SLP-18-3 (1997).
17) Lee, A., Kawahara, T. and Doshita, S.: An Efficient Two-pass Search Algorithm using Word Trellis Index, *Proc. ICSLP*, pp.1831–1834 (1998).
18) Nguyen, L. and Schwartz, R.: The BBN Single-Phonetic-Tree Fast-Match Algorithm, *Proc. ICSLP*, pp.1827–1830 (1998).

(平成 10 年 10 月 5 日受付)
(平成 11 年 2 月 8 日採録)



李 晃伸

1996 年京都大学工学部情報工学科卒業．1998 年同大学大学院工学研究科修士課程修了．現在，同大学大学院情報学研究科博士後期課程在学中．音声認識の研究に従事．電子情報通信学会，日本音響学会各会員．



河原 達也（正会員）

1987 年京都大学工学部情報工学科卒業．1989 年同大学大学院修士課程修了．1990 年同博士後期課程退学．同年同大学工学部助手．1995 年同助教授．現在，同大学情報学研究科助教授．1995 年から 1 年間米国ベル研究所客員研究員．音声認識・理解の研究に従事．京都大学博士（工学）．1997 年度日本音響学会 粟屋賞受賞．電子情報通信学会，日本音響学会，人工知能学会，IEEE 各会員．

堂下 修司（正会員）

1958 年京都大学工学部電子工学科卒業．1960 年同大学大学院修士課程修了．1963 年同博士課程単位取得退学．同年同大学工学部助手．1965 年同助教授．1968 年東京工業大学助教授．1973 年京都大学工学部教授．1996 年大型計算機センター長（併任）．現在，同大学情報学研究科教授．その間，音声の分析と認識，オートマトンの学習的構成，自然言語処理，人工知能等知的情報処理の研究に従事．京都大学工学博士．1959 年通信学会稲田賞受賞．1988 年人工知能学会論文賞受賞．1990 年情報処理学会創立 30 周年記念論文賞受賞．人工知能学会（元）会長．電子情報通信学会，日本音響学会等会員．