社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－ＳＬＰ－49 (48)
2003／12／19

2パス探索アルゴリズムにおける高速な単語事後確率に基づく信頼度算出法

李 晃伸† 河原 達也†† 鹿野 清宏†

† 奈良先端科学技術大学院大学 情報科学研究科 〒630–0192 奈良県生駒市高山町 8916–5
†† 京都大学 学術情報メディアセンター 〒606–8501 京都市左京区吉田二本松町
E-mail: †{ri,shikano}@is.aist-nara.ac.jp, ††kawahara@ar.media.kyoto-u.ac.jp

あらまし 音声認識システムにおいて，認識結果に対して事後確率などを用いて信頼度を付与することで，発話検証や対話管理などの音声アプリケーションにおいて認識誤りを考慮したより高度な処理を行うことができる．この単語の事後確率を用いた信頼度算出では，通常，認識処理（デコーディング）の結果得られた仮説群のゆう度をもとに計算されるが，十分な精度の確信度を得るためにはN-best候補で数百以上の大量の文仮説を求める必要があり，多くの計算量を必要とする．本研究では，2パストリートレリス探索に基づくデコーディングにおいて，探索中に得られる部分文仮説の尤度から単語の信頼度を簡易かつ高速に算出するアルゴリズムを提案する．後段パスのスタックデコーディングにおける単語仮説展開時に，その次単語仮説の集合およびそれぞれから展開される新たな仮説のヒューリスティックを含む尤度から，その展開単語の事後確率を計算する．通常のデコーディング処理に対して極めて少ない計算量で信頼度を計算できる．認識エンジンJuliusにおいて，N-best候補から事後確率を算出する従来手法との比較を行った結果，提案手法は大量のN-best候補を求める必要がないことから認識処理全体を非常に高速に行え，また信頼度の精度も，簡易な計算法ながらN-best候補を用いる手法と同等以上の信頼度を算出できることが示された．
キーワード 単語信頼度，探索アルゴリズム，単語事後確率，対話音声認識，認識エンジンJulius

Real-Time Confidence Scoring Based on Word Posterior Probability on two-pass search algorithm

Akinobu LEE†, Tatsuya KAWAHARA††, and Kiyohiro SHIKANO†

† Graduate School of Information Science, Nara Institute of Science and Technology Takayama-cho 8916–5, Ikoma, Nara, 630–0192 Japan
†† School of Informatics, Kyoto University Nihonmatsu-cho, Yoshida, Sakyo-ku, Kyoto, 606–8501 Japan
E-mail: †{ri,shikano}@is.aist-nara.ac.jp, ††kawahara@ar.media.kyoto-u.ac.jp

Abstract Confidence scoring based on word posterior probability is usually performed as a post process of speech recognition decoding, and also needs a large number of word hypotheses to get enough confidence quality. We propose a simple way of computing the word confidence using estimated posterior probability while decoding. At the word expansion of stack decoding search, the local sentence likelihoods that contains heuristic scores of unreached segment are directly used to compute the posterior probabilities. Experimental result showed that, although the likelihoods are not optimal, it can provide slightly better confidence measures compared with N-best lists, while the computation is much faster because no N-best decoding is required.
Key words Confidence measure, search algorithm, word posterior probability, spoken language processing, recognition engine Julius

1. はじめに

大語彙連続音声認識のための探索アルゴリズムや大規模データベースを利用した高精度な不特定話者音響モデルの構築といった，近年の音声認識技術の発達により，音声認識を用いたさまざまな音声インタフェースや音声アプリケーションが試みられている．しかし，現在の音声認識システムは，実環境や広範囲のタスク，あるいは日常会話のようなより自由な発声を対象とするとき，認識率は著しく低下する．国内でも大規模な話し言葉データベースに基づく音声認識 [1] [2] や共有データベースを基盤とした雑音に対する頑健性の向上の取り組み [3] などが試みられているが，実環境の状況において精度の高い音声認識が行えるといった水準には至っていない．

音声認識の結果を利用するアプリケーションを考えるとき，この認識誤りは重要な問題となる．一般に音声インタフェースを持つシステムにおいては，音声認識の結果からユーザーの意図や意味の抽出を行うが，このとき前段の音声認識の誤りはアプリケーションの処理に深刻な影響を与える．たとえば音声対話において，認識誤りのためにユーザの意図解釈に矛盾が発声した場合に問い合わせやユーザへの再発話の要求を行うことがあるが，この場合も認識誤りがさらなる聞き返しや誤解などの混乱を引き起こす場合がある．

音声認識結果に対して認識システム上でなんらかの信頼度尺度を付与することで，これらの誤認識の問題を緩和することができると考えられる．音声認識システムが出力する仮説のそれぞれについて，認識システムがどれだけの確信をもってその結果を出力したかの尺度を数値として付与することで，後段の音声認識処理でその値を考慮した処理が行える．確信度を用いた応用システムとしては，たとえば，音声対話システムにおけるタスク外発話の検証 [4]，確信度の高い部分を教師信号とする音響モデルの教師無し適応 [5]，あるいは音声対話システムにおける利用 [6] などの研究例が挙げられる．このように，認識エンジンが信頼度を出力することで，より高度な意図解釈や頑健なインタフェースが構築できる．なお本研究では，単語単位の確信度の付与について扱う．

この信頼度の算出方法の一つとして，単語の事後確率に基づく信頼度計算法がある．事後確率は，認識の結果得られた単語候補群の中における競合する単語仮説同士の相対的な尤度の比率を表すことができ，信頼度として有効に働くことが知られている [8]．

この単語事後確率を用いた信頼度算出には，通常のデコーディングのみを行う場合に比べて多くの計算コストを要する．十分な精度の確信度を得るためには，認識処理において一般には数百以上の大量の単語仮説を得る必要があることが知られている [8]．たとえ後段のアプリケーションが最尤の認識結果のみを必要とする場合でも，認識エンジン（デコーダ）はその確信度計算のためにより多くの仮説候補を求める必要がある．また，音声認識システム全体の応答時間を考えたとき，この確信度計算は認識処理が終了した後の後処理として行われるため，その確信度計算にかかる処理時間はそのまま応答の遅延となり，システム全体の応答速度に遅延をもたらす．

本研究では，2パスのトリートレリス探索において，この事後確率に基づく認識単語の確信度を非常に少ない計算コストで音声認識処理中に計算するアルゴリズムを提案する．解探索中に単語の確信度を求める手法については関連研究 [9] [10] があるが，[9] は事後確率に基づく確信度計算ではなく，[10] は前段パスで求めた単語グラフ上での再探索における再スコアリングである．本研究で提案する手法は，スタックデコーディング中の仮説展開時に，探索中の部分文仮説のヒューリスティックを含むゆう度から，近似的にその展開された単語の事後確率を計算しながら探索を行う．デコーダーが N-best 候補や単語グラフのように大量の上位仮説を求めなくて良いため，非常に高速に計算できる．

以下，2章で単語事後確率を用いた確信度計算の概要および計算量に関する考察を述べ，3章で提案手法であるデコーディング中の確信度計算についてアルゴリズムを示す．4章では認識実験において提案手法を N-best 候補を用いる従来手法と比較した結果を示し，5章で本論文をまとめる．

2. 事後確率を用いた単語の信頼度計算

単語の事後確率を用いた信頼度計算について述べる．認識処理の結果得られた単語グラフ，あるいは N-best 候補のリストにおいて，その中のある単語仮説 w が入力のフレーム τ から t に存在するとき，その単語仮説 $[w;\tau,t]$ の入力音声系列 X に対する事後確率 $p([w;\tau,t]|X)$ は，その仮説をパス上に含む全ての文仮説の出現確率の和より求められる．すなわち，

$$p([w;\tau,t]|X) = \sum_{W \in W_{[w;\tau,t]}} \frac{p(X|W)p(W)}{p(X)} = \sum_{W \in W_{[w;\tau,t]}} \frac{e^{g(W)}}{p(X)} \quad (1)$$

ただし $W_{[w;\tau,t]}$ は単語仮説 $[w;\tau,t]$ をパス上に含む全文仮説の集合であり，$g(W)$ はデコーダより得られる文仮説 W の言語モデル上および音響モデル上の出現確率の対数尤度である．$p(X)$ はその N-best 候補リストもしくは単語グラフにおける全ての文仮説の出現確率の和として計算できる．この事後確率から，単語仮説 $[w;\tau,t]$ の信頼度 $C([w;\tau,t])$ は以下のように定義される．

$$C([w;\tau,t]) = \sum_{W \in W_{[w;\tau,t]}} \frac{e^{\alpha \cdot g(W)}}{p(X)} \quad (2)$$

ただし α はスムージング係数 $(0 < \alpha \leq 1)$ である．一般に音声認識においては，対数尤度の値が非常に大きいダイナミックレンジを持つため，少量の上位単語仮説の値によって事後確率の値が支配される傾向にある [8]．係数 α はこの尤度のダイナミックレンジを補正するために用いられる．

単語グラフにおいて信頼度を付与した例を図1に，N-best 候補リスト上での信頼度計算の様子を図2に表す．N-best 候補として仮説空間が与えられる場合，各 N-best 候補文の事後確率を求めたのち，同じ位置にある単語仮説ごとにそれらを足し合わせることで信頼度が求められる．単語グラフにおいては，forward-backward アルゴリズムによってその単語を通る全て

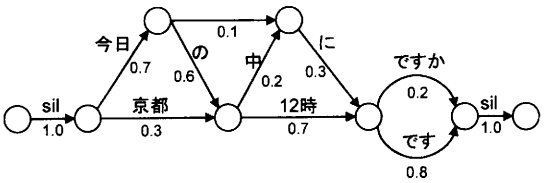

図 1 単語グラフにおける確信度付与の例
Fig. 1 An example of confidence scores on a word graph.

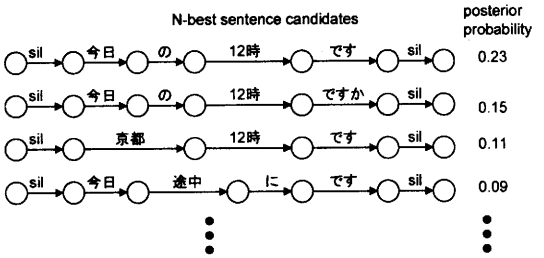

図 2 N-best 候補リストを用いた確信度計算の例
Fig. 2 Confidence scoring on N-best candidate list.

のパスの尤度の合計および全体の尤度の合計値が求められる．なお事後確率の性質より，任意の時刻 t においてその時刻上に存在する単語仮説の事後確率の合計は必ず 1 となる．

この単語グラフや N-best リストから事後確率を求める従来手法を，計算量およびシステムの応答時間への影響の観点から論じる．まず，この確信度の性能は仮説空間の広さ（単語グラフの深度，あるいは N-best 候補リストの候補数）に大きく依存するため，十分な性能を得るためには，認識処理の段階で大量の仮説候補を残す必要がある．仮説数が少ない場合は，上位の似かよった候補しか残らないため，確信度はほとんどの単語で 1 になってしまう．先行研究によると [8]，十分な性能を得るには N=100 程度が必要であるが，このような大量の文候補を得るためには，音声認識処理に多くの計算量が必要となる．

また，確信度計算の処理量の増大は応答時間の遅延につながる．ある単語仮説 $[w;\tau,t]$ の事後確率を求めるためには，その単語を通るすべてのパス $W_{[w;\tau,t]}$ の出現確率の合計を求める必要がある．また，$p(X)$ を求めるためには，同様に登場しうるすべてのパスを通る仮説の出現確率の合計を求める必要がある．確信度が認識処理が終了した後に計算されると，この処理時間はそのまま応答時間の遅延として現れる．この計算は，単語グラフにおいては forward-backward アルゴリズムによりある程度効率よく計算することができるが，このような応答の遅延は，特に音声対話システムなどの素早い応答が重要なアプリケーションでは問題となる．

3. 2 パストリートレリス探索における探索過程での信頼度計算法

デコーダーの解探索過程において，その探索時のスコアから簡便に信頼度の計算を行う手法を提案する．基本的なアプローチとしては，次単語を展開する際に，その時点で展開仮説が持つ推定尤度から，展開単語の事後確率を近似的に求める．

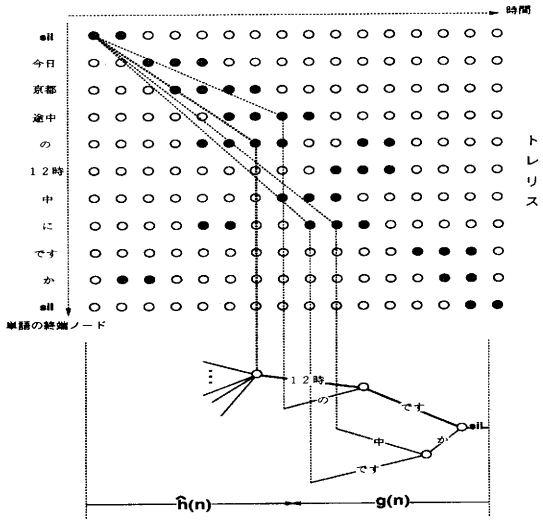

図 3 単語トレリスを用いたトリートレリス探索
Fig. 3 a tree-trellis search based on word trellis.

探索アルゴリズムとして，本研究では 2 パスのトリートレリス探索を想定する．以下，探索アルゴリズムの概要を述べた後，提案手法の解説および具体的アルゴリズムを示し，その得失について述べる．

3.1 探索アルゴリズムの概要

トリートレリス探索は，木構造化辞書とスタックデコーディングを用いた典型的な 2 パス探索法の一つである [11]．ここでは大語彙連続音声認識のために拡張したトリートレリス探索手法を用いる [12]．前段の第 1 パスでは木構造化辞書を用いたビーム探索を行い，各入力フレームにおいてビーム幅内に単語終端が残った単語について，その始終端時刻と入力先頭からの累積尤度を保存する．これを単語トレリスと呼ぶ．後段の第 2 パスはより詳細なモデルを用い，第 1 パスとは逆向きにスタックデコーディングを行うが，その際に第 1 パスで保存した単語トレリスを展開単語の絞り込み，および未探索部分の推定スコア（ヒューリスティック）として用いる．この様子を図 3 に示す．図中の上部が第 1 パスの結果の単語トレリスであり，各時間において終端ノードの残った単語が黒丸で表されている．第 2 パスでは図中下部のように逆向き探索を行うが，その際に部分文仮説の最終単語に対応するトレリス上の単語を参照し，その尤度を未探索部のスコアとして用いて探索を行う．また仮説展開時には，次単語集合として，該当時間において単語トレリス上に存在する単語のみを展開する．

展開仮説のスコア計算時には，最終単語の第 1 パスの単語トレリスと第 2 パスのトレリスを接続したときの仮説スコアが最大となる接続時刻 t を選択することで，単語履歴に依存する単語境界時間を再推定する．具体的には，第 2 パスにおいて，ある部分文仮説 $w_1^{n-1} = w_1, w_2, ..., w_{n-1}$ に対して単語 w_n を新たに接続する際，その新たな仮説のスコア $f(w_1^n)$ は以下のようにして求められる．

$$f(w_1^{n-1}, [w_n;\tau,t]) = g(w_1^{n-1}, t) + \hat{h}(w_n, t) \quad (3)$$

$$f(w_1^n) = \max_{0 \leq t < T} f(w_1^{n-1}, [w_n; t]) \quad (4)$$

ただし $g(w_1^{n-1}, t)$ は時刻 t における第2パスの部分文仮説 w_1^{n-1} の先頭部分の（前向き）尤度，$\hat{h}(w,t)$ は時刻 t における接続する単語 w の単語トレリス上での（後ろ向き）尤度である．

3.2 単語展開時の尤度からの単語事後確率算出の提案

このトリートレリス探索の第2パスにおいて，探索中の部分仮説の尤度である $f(w_1^n)$ から単語 w_n の事後確率 $p([w_n;\tau,t]|X)$ を求めることを考える．まず，尤度の和の計算においては最尤の確率が最終的な和の値を支配する場合が多いことから，単語 w_n を通るすべての仮説パスの出現確率の合計を求めるかわりに，単語 w_n を通る最尤パスの出現確率を用いることを考える．さらに，この最尤パスの出現確率として $f(w_1^n)$ を用いることを考える．$f(w_1^n)$ は探索途中の部分文仮説の評価スコアであるが，式 (3) に示すように探索済みの部分の尤度と未探索部分のヒューリスティックな尤度の合計となっており，これをその時点での単語 w_n を通るパスの最尤スコアとみなす．これは確率 $p(w_n, X)$ を，一部にヒューリスティックを含む最尤のパスの尤度を用いて近似することを意味する．

また，$p(X)$ の推定については，その展開単語 w_n と同じフレーム上に展開されるすべてのトレリス上の単語を展開候補として，その仮説 $[w;\tau,t]$ について $f(w_1^{n-1}, [w;\tau,t])$ を計算して足し合わせることで，それに近い値を求めることができる．

以上より，単語展開における次単語 w_n について，近似的な事後確率 $\hat{p}(w_n|X)$ を以下の式で計算する．

$$W_c = [w;\tau,t] : \tau \leqq t_n \leqq t \quad (5)$$

$$\hat{p}(w_n|X) = \frac{e^{f(w_1^n)}}{\sum_{W_c} e^{f(w_1^{n-1}, [w;\tau,t])}} \quad (6)$$

ただし t_n は式 (4) において最大を与える時刻 t である．

この近似により，探索過程において単語展開時に単語事後確率を用いた信頼度計算を行う．

3.3 具体的アルゴリズム

提案した信頼度計算法を組み込んだ第2パスの探索アルゴリズムは以下のとおりである．(e) の部分が信頼度計算のために追加された部分であり，他の部分は従来のアルゴリズムと同一の処理である．

（1） 初期仮説を仮説スタックに入れる．

（2） 以下の (a) から (f) のステップを繰り返す．

（a） 仮説の取り出し：

仮説スタックからスコアが最も高い仮説 w_1^{n-1} を取り出す．

（b） 終了判定：

取り出した仮説の探索が入力終端に達していたら，認識結果としてそれを出力し，(a) に戻る．

（c） 次単語集合の抽出:

取り出した仮説の終端フレームを推定し，その推定フレームの周辺に存在するトレリス上の単語の集合 W_c を次単語集合として抽出する（式 (5)）．

（d） 次仮説の生成:

次単語集合 W_c に含まれる各単語 w_n について，それを接続した新たな部分文仮説を生成してスコア $f(w_1^n)$ を計算する．仮説のスコアは，第2パスのトレリスと第1パスのトレリスを接続して求める（式 (3)）が，その前後に存在する同じ単語のトレリス単語の中からスコアが最も高くなるものを選んで，それを新たな仮説のスコアとする（式 (4)）．

（e） 信頼度計算：

新たに生成した仮説の各スコアから，式 (6)）に従って単語事後確率に基づく信頼度を計算し，各仮説に保持させる．

（f） 仮説の格納：

生成した仮説を仮説スタックに入れる．(a) に戻る．

3.4 提案手法の得失について

本提案手法の長所は，未探索部分の推定値を含み，かつベストなコンテキストを考慮した尤度を用いることで，認識の後処理ではなく認識の探索処理の途中で高速に事後確率を計算できることにある．信頼度の算出のために行う実質の計算は前節の (e) のみであり，その算出に必要な値のほとんどは探索の過程で常に求められる値のため，信頼度の計算のための計算コストは非常に少なくてすむ．また，探索途中に展開する単語間のスコアから直接計算するため，N-best リストや単語グラフのように大量の文仮説を求める必要がなく，また認識終了と同時にCM 値を得ることが可能である．これらの特徴により，認識処理に対するオーバヘッドが非常に小さい高速な確信度計算を行うことができる．

また，N-best リストや単語グラフを用いる従来手法では，認識結果の上位候補を表現するため，上位の仮説のみでは似通ったスコアのものが多く事後確率をうまく計算できない．有効な確信度を得るためには大量の仮説を生成する必要がある．，このため認識処理で大量の仮説を出力する必要があり，CM 計算のコストのみならず，この認識のための計算コストも増大する．これに対して，本手法では探索の過程で単語終端に達したすべての単語仮説を対象とするので，単語グラフや N-best リストでは残らない下位の仮説まで考慮した評価となる．このため，よりよい信頼度の値が得られる可能性がある．また，確信度計算のために大量の候補を求める必要がなく，認識処理は通常の1位のみを求める計算で済むため，トータルの計算量が大幅に削減される．

ただし，用いる仮説スコアは第1パスの結果から得られる未探索部分のヒューリスティックな値を含んでいるため，正確な最尤スコアとの間に誤差が生じる．算出に用いる尤度は未探索部の推定値を含む値であり，また尤度の合計 $p(X)$ も真の値ではない．このように，用いる確率が入力全体を考慮した真の確率ではなく，その探索時点での情報をもとにした近似となっているため，誤差が生じる可能性がある．

4. 実験的評価

4.1 実験条件

認識実験によって提案手法の評価を行った．提案法および N-best 方式の両手法をトリートレリス探索を行う認識エンジン Julius [14] rev. 3.4 (注1) に実装し，同一の条件のもとで性能の比較を行った．なお N-best 候補については $N = 100$ とする．

評価タスクは，生駒市の市民ホールに常設された音声情報案

(注1)：http://julius.sourceforge.jp/にて入手可能．バージョン 3.4 以降では，本論文で提案する信頼度算出法が標準で実装されている．

内エージェントに対する質問タスクとする [15]．このタスクは，市民会館のホールに設置されたソフトウェアエージェントに対する問いかけで，館内や周囲の道案内，呼びかけ応答などを対象とする．エージェントに対して自由に発話されたユーザ発話のうち，成人の比較的クリアな発声を 500 文抽出してテストセットとした．平均発話長は 2.1 秒である．言語モデルは第 1 パスに単語 2-gram，第 2 パスに後ろ向きの単語 3-gram を使用する．このホール名や市の名前をキーワードとして Web 検索エンジンを用いて収集したテキストから学習したモデル，およびこのタスクでの書き起こしテキストから作成したモデルを融合したものを用いている．語彙数は 41248 語で，この言語モデルによるテストセットパープレキシティは 11.4 である．音響モデルはそのシステムの設置された雑音環境に比較的近い音響モデルとして，展示会雑音をデータベースの音声に重畳して学習した不特定話者の phonetic tied-mixture (PTM) モデルを用いる．なお，信頼度計算におけるスケーリング係数 α は，テストセットに対する最適な値（N-best リスト・提案手法とも 0.04）を採用した．

4.2 評価尺度

信頼度の精度を評価するために，しきい値 θ を定め，認識結果に対して $C(w) \geq \theta$ であれば受理，$C(w) < \theta$ であれば棄却するとして認識結果のタグ付けを行い，そのタグと認識結果の正解・不正解を比較する．

評価には複数の尺度を用いる．まず音声認識結果に対するラベル付けの精度の観点から，Confidence Error Rate (CER) および Detection-error-tradeoff (DET) カーブを算出する．CER は誤りタグ数（受理した不正解単語数と棄却した正解単語数）の認識単語数に対する割合である．baseline CER はすべての単語を受理したときの値であり，置換誤りと挿入誤りの和を認識単語数で割ったものとなる．また，DET カーブは誤受理率（不正解単語のうち，誤って受理した単語の割合）と誤棄却率（正解単語のうち誤って棄却した単語の割合）をさまざまな θ についてプロットしたものである．

また別の尺度として，確信度がしきい値以上の単語のみを認識結果として受理したときの，正解文に対する適合率と再現率を表す尺度として，受理誤り率 (failure of acceptance: FA) とスロット誤り (Slot Error: SErr) を以下のように定義する．

$$\mathrm{FA} = 1 - \frac{\text{受理した正解単語数}}{\text{受理された単語の総数}} \tag{7}$$

$$\mathrm{SErr} = 1 - \frac{\text{受理した正解単語数}}{\text{総正解単語数}} \tag{8}$$

FA はあるしきい値における受理した単語の置換誤りと挿入誤りの和を受理した単語数で割ったものであり，すべての単語を受理するとき，FA は baseline CER と等しくなる．SErr は受理した単語の置換誤りと削除誤りの和を表す．

また，一般に音声認識の後処理においては，誤り単語が受理される誤りの方が，正解単語が棄却される誤りに比べてシステム全体に対する損失が大きい．たとえば音声対話において誤った単語をもとに誤った意図解釈がなされると，システムとユーザの間に誤解が生じ，その修正にさらに多くの発話が必要となるからである．これを考慮して，FA に重み λ を与えた値 $(\mathrm{FA} \cdot \lambda + \mathrm{SErr}) \times 2/(\lambda + 1)$ についても求める．

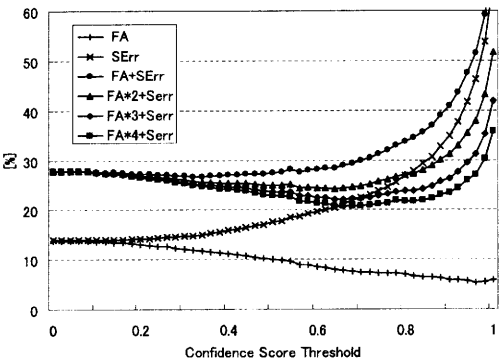

図 4 N-best 候補を用いた手法における FA および SErr (N=100)
Fig. 4 FA and SErr by N-best method (N=100).

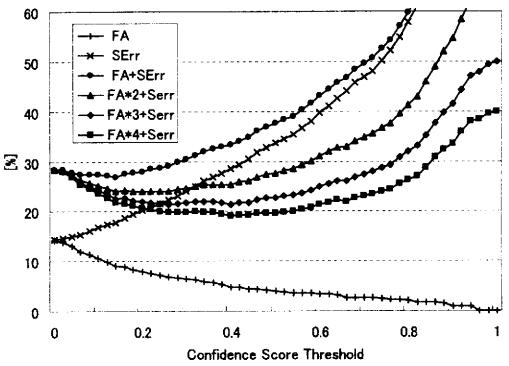

図 5 提案手法における FA および SErr
Fig. 5 FA and SErr by proposed method.

各評価尺度の位置づけは，CER および DET カーブは信頼度を発話検証のための尺度として用いる場合の評価尺度であり，FA および SErr は音声対話システムを想定した発話解釈を考慮した尺度である．

4.3 実験結果

さまざまなしきい値 θ における FA，SErr および FA+λSErr の値について，従来法である N-best リスト法での結果を図 4 に，提案手法での結果を図 5 に示す．提案手法は近似手法ながら，従来の N-best と同等の信頼度を算出できることが示された．特に，FA を比較的低く抑えられる傾向が見られた．これは，N-best 方式が上位 N 個の候補のみを計算に用いるため確率の低い仮説が計算から除外されるのに対して，提案手法は展開対象となった全単語から計算するため，特に競合候補が多い単語についてより厳しい信頼度を算出するためと考えられる．しかし，一方で SErr はかなり大きくなる傾向にあった．これは提案手法では尤度が第 1 パスに基づくヒューリスティックなスコアを含む，探索中のスコアで算出するためであるとみられる．特に，Julius では第 1 パスで単語 2-gram，第 2 パスで単語 3-gram を用いているが，この言語スコアの差が大きくなる場所では，最終的に 3-gram によってスコアが良くなるはずの

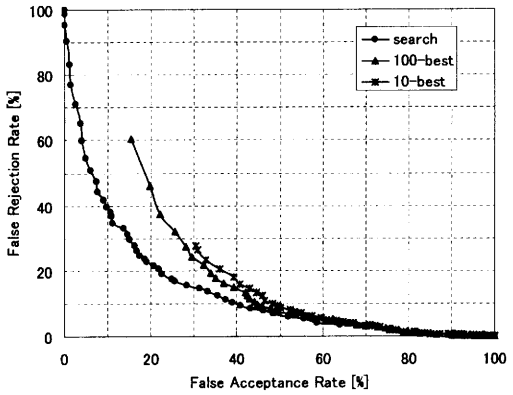

図 6 Detection-error tradeoff カーブ
Fig. 6 Detection-error tradeoff curves.

表 1 最小誤り率 [%] および平均処理時間
Table 1 Minimal error rate [%] and average process time

	FA +SErr	FA*4 +SErr	CER	avg. time (sec)
10-best	27.1	21.9	12.3	2.0
100-best	26.7	20.8	12.3	2.4
search	26.9	19.1	11.8	2.1

CPU: Intel Xeon 2.4GHz, baseline CER = 14.1

単語が，展開時の第 1 パスの 2-gram の影響で低い確信度が与えられてしまう例が見られた．

同じ実験結果から算出した DET カーブを図 6 に示す．N-best 候補を用いる方法が，似た仮説を多く生成するため多くの単語について 1.0 の確信度を与えてしまうため誤受理率の下限を下げられないのに対して，提案手法（図中 “search”）はより低い誤受理率と誤棄却率を示した．

最後に，表 1 に，誤り率が最小となる θ における各種の誤り率と一文あたりの平均処理時間を示す．表中の “search” が提案手法にあたる．また N-best 方式において $N = 10$ の場合の値も示す．$\lambda = 4$ の時の FA*4+SErr の最小値が，N-best 手法 (N=100) の 21.0 に対して提案手法では 18.9 となり，提案手法による計算方法が確信度として有効に機能することが示された．また，受理／棄却と正解／不正解の適合率を示す確信度誤り率 (confidence error rate: CER) についても，提案手法が N-best 手法を若干上回った．処理量については，信頼度計算そのものにかかる処理量は認識の処理量に比べて微少であるが，N-best 方式では大量の N-best 候補を求める必要があるため認識処理量が多くかかった．これに対して，提案手法では 1 位候補のみですみ，また確信度計算にかかる計算量も極めて小さく，通常のデコーディングのみの場合とほとんど変わらない速度で確信度を計算することができた．

5. まとめ

単語の事後確率に基づく信頼度を簡易かつ高速に計算する手法を提案した．認識過程において仮説展開時の部分文スコアを直接用いることで，従来手法のように大量の文候補を求めることなしに，デコーダ上で認識と同時に信頼度の計算を行うことができる．評価実験の結果，探索過程で出現するすべての競合する部分文仮説から信頼度を算出することで，従来手法よりも若干精度の高い信頼度を付与することができることが示されたが，一方で探索途中のスコアを用いることで低い確信度が与えられてしまう例が見られた．今後の課題としては，探索中のスコアを事後補正する手法の検討や，他のタスクドメインにおける評価，具体的なアプリケーションにおける評価が挙げられる．

文献

[1] 南條浩輝, 加藤一臣, 李晃伸, 河原達也. 大規模な日本語話し言葉データベースを用いた講演音声認識. 電子情報通信学会論文誌, Vol. J86-DII, No. 4, pp. 450–459, 2003.

[2] 篠崎隆宏, 古井貞熙. 日本語話し言葉コーパスを用いた講演音声認識. 情報処理学会論文誌, Vol. 43, No. 7, pp. 2098–2107, 2002.

[3] 山本一公, 中村哲, 武田一哉, 黒岩眞吾, 北岡教英, 山田武志, 水町光徳, 西浦敬信, 藤本雅清. AURORA-2J/AURORA-3J データベースとその評価ベースライン. 情報処理学会研究報告, 2003-SLP-47-19, 2003.

[4] Ananth Sankar and Su-Lin Wu. Utterance verification based on statistics of phone-level confidence scores. In *Proc. ICASSP*, Vol. 1, pp. 584–587, 2003.

[5] 緒方淳, 有木康雄. 音素事後確率に基づく信頼度を用いた音響モデルの教師なし適応化. 情報処理学会研究報告, 2001-SLP-39-22, 2001.

[6] 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, Vol. 43, No. 10, pp. 3078–3086, October 2002.

[7] Thomas Kemp and Thomas Schaaf. Estimating confidence using word lattices. In *Proc. EUROSPEECH*, Vol. 2, pp. 827–830, September 1997.

[8] Frank Wessel, Ralf Schluter, Klaus Macherey, and Hermann Ney. Confidence measures for large vocabulary continuous speech recognition. *IEEE Trans. Speech & Audio Process.*, Vol. 9, No. 3, pp. 288–298, March 2001.

[9] C. Neti, S. Roukos, and E. Eide. Word-based confidence measures as a guide for stack search in speech recognition. In *Proc. ICASSP*, Vol. 2, pp. 883–886, 1997.

[10] G. Evermann and P. Woodland. Large vocabulary decoding and confidence estimation using word posterior probabilities. In *Proc. ICASSP*, Vol. III, pp. 1655–1658, 2000.

[11] F. K. Soong and E.-F. Huang. A tree-trellis based fast search for finding the N best sentence hypotheses in continuous speech recognition. In *Proc. ICASSP*, Vol. 1, pp. 705–708, 1991.

[12] 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段階的探索による大語彙連続音声認識. 電子情報通信学会論文誌, Vol. J82-D-II No.1, pp. 1–9, 1999.

[13] 李晃伸, 河原達也. 大語彙連続音声認識エンジン Julius における A*探索法の改善. 情報処理学会研究報告, 99-SLP-27-5, 1999.

[14] A. Lee, T. Kawahara, and K. Shikano. Julius — an open source real-time large vocabulary recognition engine. In *Proc. EUROSPEECH*, pp. 1691–1694, September 2001.

[15] R. Nisimura, Y. Nishihara, R. Tsurumi, A. Lee, H. Saruwatari, and K. Shikano. Takemaru-kun: Speech-oriented information system for real world research platform. In *Proc. First International Workshop on Language Understanding and Agents for Real World Interaction*, pp. 70–78, July 2003.