大語彙連続音声認識のための複数言語モデルの並列同時単語列探索法

中島邦男[†,☆] 阿部芳春[†]
伍井啓恭[†] 丸田裕三[†,☆☆]

複数の要素言語モデルを並置した並列言語モデルにおける効率的な単語列の探索方法を提案する．本方法は並列言語モデルの各要素言語モデルによる最尤単語列の中で尤度最大の単語列を同時並列的に求めるものである．この並列同時探索は，1つのスタック上で，要素言語モデルへのポインタと部分単語列とスコアとを1組とする仮説を最良優先探索に基づき展開することで実現される．本手法を筆者らが大語彙連続音声認識の手法として提案した「認識誤り傾向の確率モデルを用いた2段階探索法」の2段目に適用する．1段目は言語モデルと独立して入力音声に対して基本単位列の認識を行い，2段目は音響モデルと独立して基本単位列に対して単語列探索を行う．大語彙の連続音声認識実験を行い，各要素言語モデルの最尤単語列を求めてから尤度最大の単語列を選択する逐次探索法と性能を比較した．57カ月分の新聞記事を自動分類して複数言語モデルを作成し，並列言語モデルを構成した．実験の結果，同時探索法は，9並列の言語モデルの場合で，仮説展開回数を逐次探索法の1/5未満に削減するとともに，単一言語モデルに対する誤り削減率で約8.4%を得た．これは逐次探索法の約8.0%と同等以上であり，提案手法の有効性が確認された．

Simultaneous Word Sequence Search for Parallel Language Models in Large Vocabulary Continuous Speech Recognition

Kunio Nakajima,[†,☆] Yoshiharu Abe,[†] Hiroyasu Itsui[†]
and Yuzo Maruta[†,☆☆]

An efficient method to search parallel language models for a word sequence is proposed. The method finds simultaneously the most probable word sequence in word sequences of component language models. The simultaneous search is realized by a best-first search algorithm, in which a pointer to the component language model and the partial word sequence are paired as a hypothesis. In experiments for large vocabulary continuous speech recognition, the performance was compared with a serial search method, which selected the most probable word sequence after sequentially searching the component language models for the most probable word sequences. As a result, the simultaneous search method, in a case of parallel nine language models, reduced the count of generated hypotheses below one fifth of that of the serial search method, and obtained the error reduction rate of 8.4%, which was comparable to that of the serial search method of 8.0%. Thus the effectiveness of the proposed method was confirmed.

1. はじめに

音声認識の適用分野が広がり，より広い話題をカバーする言語モデルが必要となっている．言語モデルとして，単一の単語n-gramモデルを用いる方法では，探索空間が莫大になるとともに，音響処理と組み合わせたとき，出現確率の低いテキストの認識性能が低下する危険性が高く，カバー率と認識性能を両立させるのは難しい．

これに対し，文単位の制約を導入する方法[1),2)]が試みられている．また，話題の異なるコーパスから言語モデルを事前に作成し，少量の適応データにより言語モデルを話題適応する方法[3)~7)]が試みられている．

また，話題ごとに言語モデルを作成して，話題の制約をかけた単語列探索を行った後，各言語モデルに基づく正規化尤度が最大の単語列を認識結果とする方法が提案されている[8)]．この並列言語モデルの方法によれば，探索空間の削減，カバー率の改善，音響処理と

† 三菱電機株式会社情報技術総合研究所
Information Technology R&D Center, Mitsubishi Electric Corporation
☆ 現在，横浜国立大学産学連携推進本部
Presently with Office of Industry and Community Liaison, Yokohama National University
☆☆ 現在，三菱電機株式会社自動車機器開発センター
Presently with Automotive Electronics Development Center, Mitsubishi Electric Corporation

組み合わせたときの言語的制約性能の改善などの効果が期待される．

後者の並列言語モデルでは，尤度最大の単語列を求める計算を要素言語モデルごとに逐次行うと，計算コストが要素言語モデルの数とともに増大するという問題がある．この計算コストを削減する方法として，ユーザの発話から話題を同定し，その話題に依存した言語モデルを用いて再度認識することによって，精度を同等に維持して計算コストを削減する方法[9]が提案されている．しかし，この方法では，話題同定の処理が新たに必要になる．

本論文では，後者の並列言語モデルの枠組みで，尤度が最大の単語列の探索を効率的に行う並行単語列探索法について述べる．

提案する同時探索は，1 つのスタック上で，言語モデルへのポインタと部分単語列とからなる仮説の展開を行うことで実現される．

筆者らは，大語彙連続音声認識の手法として，認識誤りの確率モデルを用いた 2 段階探索法[10]を提案した．本論文では，同時探索をこの 2 段階探索法に適用する．2 段階探索法の 1 段目の基本単位の認識は言語モデルと独立している．また，2 段目の単語列探索は音響モデルと独立している．このため，単語列探索は 2 段目の言語モデルだけを拡張することで実現可能である．

2. 並列同時単語列探索

本章では，複数の言語モデルを並べた並列言語モデルにおいて，各言語モデルの中でスコア最大の単語列を認識結果とする場合の探索演算の削減について述べる．

2.1 基本単位列の認識結果に基づく単語列探索

本論文の対象とする方式は，2 段階の探索方式[10]である（**図 1**）．2 段階探索方式では，第 1 段階で基本単位列の認識を行う．第 2 段階では第 1 段階で得られた基本単位列に対して認識誤りの傾向を表す差分モデルと単語 n グラム言語モデルを用いて，単語列を探索する．

2 段階探索の演算量は，入力音声に対して基本単位列を求める 1 段目の演算と，基本単位列に対して，言語モデルによる単語列を探索する 2 段目の演算の和である．言語モデル数が多くなるに従い，2 段目の演算量が増大する問題がある．

2.2 単一言語モデルの場合の単語列探索

1 段目で得られた基本単位列を X とする．2 段目では，基本単位列 X に対して，差分モデルと言語モデルによる確率を最大とする単語列 $\hat{W}$ を探索する．

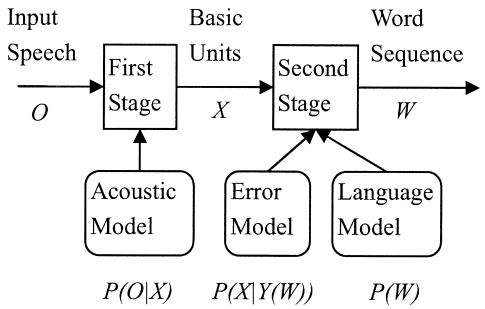


図 1 2 段階探索法
Fig. 1 A block diagram of the two-stage search method.

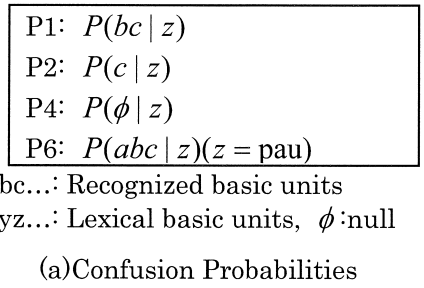
$$\hat{W} = \arg\max_{W} P(X|Y(W))P(W) \qquad (1)$$

ここで，$P(W)$ は単語列 W の確率，$Y(W)$ は単語列 W に対応する基本単位列で，$P(X|Y(W))$ は基本単位列 $Y(W)$ が基本単位列 X となる確率である．確率 $P(W)$ は単語 n-gram などの言語モデルで，また，確率 $P(X|Y(W))$ は差分モデルで計算される．

詳細には，基本単位列 X および $Y(W)$ を，それぞれ，$X = X_1X_2\cdots X_i\cdots X_I$ および $Y = Y_1Y_2\cdots Y_i\cdots Y_I$ と分解する．ここで，X_i および Y_i は部分音節列．確率 $P(X|Y(W))$ は，部分音節列間の対応確率 $P(X_i|Y_i)$（すなわち混同確率）の積として，$P(X|Y(W)) = \max\prod_i P(X_i|Y_i)$ と計算する．ここで，max は部分音節列への分解の仕方についての最大化を意味する．最大化は，DP マッチングにより行う．本論文では**図 2** に示すような非対称型の差分モデルを用いる．

式 (1) の単語列探索は，準 A*ヒューリスティクに基づく最良優先探索[11]で行う．最良優先探索は，スタックを用いて実現する．スタックに置かれた（プッシュされた）部分単語列仮説の中でスコアが最大の仮説を取り出す（ポップする）．ポップした仮説を言語モデルの下で展開する．展開した仮説をスタックに戻す．この操作を部分単語列が基本単位列の終端に達するまで繰り返す．基本単位列の終端に達した部分単語列仮説を解の候補とする．

A*探索では最初に得られる解が最適解となるが，準 A*探索では，解を展開し続けると，よりスコアの高い解が順次得られる．すなわち，漸近的に最適解が得られるという保証しかない．このため，展開の過程で得られた解をプールに保存しておき，保存した解の数が一定数に達したところで探索を終了し，プールした解の中からスコアが最大の解を選択するという手続き



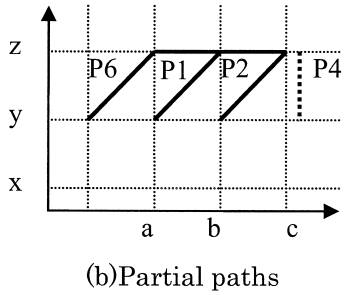


図 2 非対称型差分モデル
Fig. 2 Asymmetric error model.

をとる．以後，プールに保存する解の数を “探索解の数 n” とする．

2.3 並列言語モデルの逐次探索

逐次探索は最初にすべての言語モデル c 各々について次式の最大確率を与える単語列 $\hat{W}_c$ を探索する．

$$P(\hat{W}_c) = \max_{W_c} P(X|Y(W_c))P(W_c|c)P(c) \quad (2)$$

次に，上記の最大確率のさらに最大値を与える単語列 $\hat{W}$ を次式で求める．

$$\hat{W} = \arg\max_{\hat{W}_c} P(\hat{W}_c) \quad (3)$$

2.4 並列言語モデルの同時探索

同時探索は単語列 W_c と言語モデルの番号 c の組を同時に探索して，確率最大の解を求める．

$$\hat{W} = \arg\max_{(W_c,c)} P(X|Y(W_c))P(W_c|c)P(c) \quad (4)$$

この同時探索は，単一言語モデルの場合のスタックアルゴリズムを拡張することによって行う．**図 3** に処理フローを示す．スコア S と部分単語列 W_1^k，言語モデル番号 c の組を仮説とする．スタックから取り出した（ポップした）仮説の言語モデル番号が指す言語モデルに基づいて，部分単語列 W_1^k の展開を行う．展開の結果 W_1^{k+1} が生成される．展開の結果得られる単語列 W_1^{k+1} を新しいスコアとともにスタックに戻す（プッシュする）．

A*探索の性能はヒューリスティックに大きく依存

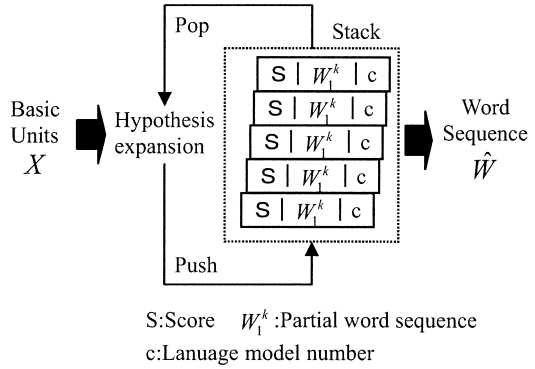


図 3 同時探索
Fig. 3 A flow diagram of the simultaneous search.

する．本論文では，文献 11) の手法を適用して，単語列を探索する．いま，入力の基本単位列を $X_1^I = \{x_1, x_2, \ldots, x_I\}$（$x_i$ は基本単位）とする．言語モデル c の仮説 i が単語列 W_1^k で時刻 t までの基本単位列 $X_1^t (0 \leq t \leq I)$ を覆うとき，仮説 i のスコア（対数確率）は

$$g_i(t) = \log P(X_1^t|Y(W_1^k))P(W_1^k|c)P(c) \quad (5)$$

$$\text{初期値 } g_i(0) = P(c) \quad (6)$$

仮説 i のヒューリスティック $h_i(t)$ は

$$h_i(t) = g_i(t) - Env(t) \quad (7)$$

$Env(t)$ は，評価時点で時刻 t に到達した全仮説 $\{j\}$ の最大スコア $Env(t) = \max_j g_j(t)$ で，初期値 $Env(t) = -\infty (0 \leq t \leq I)$ である．ヒューリスティックが同一の仮説がある場合，幅方向に展開する．

準 A*ヒューリスティックを用いる同時探索で，式 (4) のように $P(c)$ を加えると，探索性能への影響が考えられる．すなわち，最初に時刻 t に到達する仮説 k のヒューリスティック $h_k(t)$ は $Env(t) = g_k(t)$ と式 (7) よりつねに 0（最大値）である．このため，仮説展開の初期では，$P(c)$ が最も大きい仮説が優先的に展開される．しかし，仮説展開の進行につれ，展開される仮説のヒューリスティクは 0 から漸減していくため，$P(c)$ が相対的に小さい仮説でも，いずれは展開される．したがって，仮に，$P(c)$ が相対的に小さくても，並列言語モデル中の最適解が，仮説展開を繰り返すうちにいつか尤度最大の解として得られることは保証される．このように，準 A*探索では，$P(c)$ を加えると仮説展開の初期で探索に影響があるが，探索解の数 “n” を十分大きくすることにより，最適解が棄却される可能性を低下させることは可能であると考えられる．なお，本論文の実験では，$P(c)$ を加えることによる探索性能への影響を排除するため，$P(c)$ は

均等の値とする．

同時探索はスタック上の最尤仮説を優先的に展開する．ここで，最尤仮説はすべての要素言語モデルの仮説の競合の結果であるので，スコアの低い仮説の生成を避けることができ，逐次探索より少ない仮説の展開回数で最適解が得られると期待される．

3. 並列言語モデルのためのテキスト分類

評価用の並列言語モデルを作成するため，事前に分類されていないテキストを C 個のクラスタに自動分類する．まず，適当な方法で，テキストを C 個のクラスタに分類する．テキスト $s \in S$ とクラスタ $c \in C$ の類似度 $L(s,c)$ を測り，類似度が最大となるクラスタにテキストを分類する．テキストの単語列を $w_1, w_2, \cdots$ とするとき，次式の対数尤度 $\log P(w_i|c)$ の加重平均を類似度 $L(s,c)$ とする．

$$L(s,c) = \sum_i g(w_i) \log P(w_i|c) / \sum_i g(w_i) \quad (8)$$

ここで，$g(w_i)$ は単語 w_i の重要度である．

重要度として，ridf（residual inverse document frequency）を用いる．ridf は観測された df（document frequency）と Poisson 分布から予測される df の対数値の差である．Poisson 分布は非内容語の分布だけを良く近似するため，Poisson 分布からの偏りに対応する ridf は，内容語である度合いの良い推定子となる[12]．本論文では，助詞などの非内容語の加重を下げるとともに，出現文書数の少ない名詞などの内容語の加重を上げるため，ridf を用いる．

ridf は

$$\begin{aligned} \mathrm{ridf}_i &= \mathrm{idf}_i - \log(1 - p(0;\lambda_i)) \\ &= \log \frac{N}{\mathrm{df}_i} - \log\left(1 - e^{-\frac{\mathrm{cf}_i}{N}}\right) \end{aligned} \quad (9)$$

$p(k,\lambda)$：平均（=分散）λ のポアソン分布

cf_i：単語 w_i の総頻度

で与えられる．ここで，idf（inverse document frequency）は単語 w_i を含む文書（＝テキスト）数から

$$\mathrm{idf}_i = \log \frac{N}{\mathrm{df}_i} \quad (10)$$

N：文書総数　　df_i：単語 w_i を含む文書数

で計算される．

4. 評価実験

逐次探索と同時探索を仮説の展開回数および認識性能への影響により比較する．

4.1 コーパス

「毎日新聞 CD-ROM1991-1995」[13] のうち 4 年 9 カ月分の約 43 万記事〔約 1 億単語（＝形態素）〕を用いた．**表 1** にコーパスの概要を示す．異なり単語数（4 年 9 カ月分）は約 44 万 2 千語であった．以下の実験では，各記事を 1 つのテキストとして扱う．

表 1 テキストデータ（新聞記事）
Table 1 Text corpus (newspaper articles).

Purpose	Articles	Sentences	Words
Training	428.6×10^3	5029×10^3	102.5×10^6

表 2 言語モデル
Table 2 Language models.

Clustering	K-means algorithm [14]
Component LM	65K-word, 3-gram LM
Full corpus LM	Used jointly

4.2 テキストの分類

K-means アルゴリズム[14] を用いた．学習用のテキストをランダムに分配し，C 個の初期クラスタを作成した．その後，学習用のテキストを式 (8) に基づいて類似度が最大になるクラスに再分類する．その際，平滑化のため任意のクラス c の単語頻度 $N(c,w)$ に定数 ε（=1）を加えた．十分収束するよう 120 回反復した．

分類されたテキストから，要素言語モデルを作成した．これとは別に，コーパス全体から作成された共通の言語モデル（全体モデル）を 1 要素として加えた．このため言語モデルの総数はつねに $C+1$ となる．全体モデルおよびすべての要素言語モデルは，すべて trigram 言語モデルで，語彙数が 65,000 となるように頻度の低い単語を足切りした．なお，同時探索を行う場合でも，言語モデル間で語彙を共通化する処理はしていない．この結果として，並列言語モデル全体での総語彙数は 65,000 語を超えることが可能であり，実際，実験では並列言語モデル全体での総語彙数は 65,000（言語モデル数 1）から 202,847（言語モデル数 10+1）の間で変化した．**表 2** に言語モデルの概要を示す．

4.3 音声データ

評価音声は，「日本音響学会―新聞記事読上げ音声コーパス―」（JNAS corpus）の話者 f001〜f010 および m001〜m010 の新聞記事読み上げ音声から，掲載面が，スポーツと経済と社会である発話を選択した．また，掲載面と独立するセットとして，各話者の最初の 10 文からなる低 perplexity で普通の長さのセット J1 と，各話者の最後の 10 文からなる高 perplexity で長文のセット J2 を用意した．各セットは言語的にも音響的にもオープンなセットである．概要を**表 3** に

表 3 評価用音声データ（JNAS コーパス）
Table 3 Test speech data (Japanese News Article Speech corpus).

Set	Amounts	Tag/Comment
J05	178sent.	スポーツ
J09	146sent.	経済
J12	270sent.	社会
J1	200sent.	Low perplexity and normal length
J2	200sent.	High perplexity and long length

示す．

4.4 認識性能の評価

評価用音声データを大語彙連続音声認識システム[10]を用いて認識した．音響モデルは，状態共有化 3 状態音素トライホンモデル（音素数 25，総状態数 2,000，16 混合対角正規分布）である．特徴ベクトルは，60 次元セグメント統計量である．セグメント統計量は，サンプリング周波数 16 kHz の原データを 11 kHz にダウンサンプリングした後，10 ms のフレーム周期で，0 次から 12 次までのメルケプストラム係数 13 次元に変換後，連続する 9 フレームを切り出して，固定長のセグメントを求め，主成分分析により 60 次元に圧縮して求めた．音響モデルの学習は，評価音声とは異なる男女各 99 名が発声した新聞記事と ATR 音素バランス文（合計約 3 万文）を用いた．

図 1 のシステム構成において，第 1 段階で，評価用音声データに対して，音響モデルを用いて，音素間に日本語音節内の音素並びの制約を課した連続音素認識を行い，音素列を求めた後，基本単位としての音節の列に変換した．第 2 段階で，第 1 段階で得た音節列に対して，並列言語モデルを用い逐次探索または同時探索を行い，最適な単語列を求め，単語列中の各単語の表記を連結して認識結果の漢字仮名交じりのテキストを得た．

認識性能を評価するため，認識結果の漢字仮名交じりのテキストにおける単語単位の誤り率（Word Error Rate; WER）を求めた．WER は正解の単語数に対するすべての種類の誤り（置換，挿入，脱落）の合計数の割合として定義する．単語誤り率は単語辞書の単位のとり方に依存するため，日本語形態素解析システム「茶筅」[15]による形態素解析結果の単位を用いて評価した．

また，単語列探索における探索回数の評価のため，仮説の生成回数（1 文あたりの平均）を求めた．仮説展開では，最初に得られる最大 n 個の解のうち尤度最大の解を認識結果とし，以後の探索を打ち切った．

4.5 並列言語モデルの効果

並列言語モデルの効果を調べるため，単一の言語モデルと並列言語モデルの比較を行った．本節では**表 4** の学習テキストを用いた．単一言語モデルは，掲載紙面がスポーツ，経済，社会であるテキストをあわせたデータから作成した．並列言語モデルは，これらのトピックごとに言語モデルを作成した（$C = 3$）．並列言語モデルでは，一要素として，単一言語モデルを加える場合（全体モデルあり）と，加えない場合（全体モデルなし）を比較した．認識結果は逐次探索で得た．

表 4 4.5 節における学習テキストのデータ量
Table 4 Amounts of training texts in Section 4.5.

掲載紙面	スポーツ	経済	社会
記事数	36461	36213	121925
文数	583818	314214	1038126
単語数	7760545	7345583	23768461

表 5 単一言語モデルと並列言語モデルの比較
Table 5 Comparison of the single language model and the parallel language models.

WER(%)

評価セット	単一言語モデル	並列言語モデル ($C = 3$) 全体モデルあり	並列言語モデル ($C = 3$) 全体モデルなし
J05	21.54	19.81	21.13
J09	17.96	16.10	15.87
J12	21.29	20.83	20.80
平均	20.26	18.91	19.27

評価セットをスポーツ，経済，社会の音声データとして，単語単位の誤り率を**表 5** に示す．単一言語モデルにくらべて，全体モデルありの並列言語モデルを用いることにより，平均約 1.35%誤り率が減少している．また，全体モデルを用いることで，誤り率が平均約 0.36%減少していることが分かる．

4.6 探索解の総数と言語モデル数の比を一定とした場合の比較

逐次探索で，1 言語モデルあたりの探索解の数を 20 とした．また，同時探索では，探索解の総数を $n = 20 \times (C + 1)$ とし，逐次探索と同時探索の両者で，探索解の総数が同じになるようにした．

仮説生成回数（プッシュの回数）の比較を**図 4** に示す．**図 5** に逐次探索の仮説生成回数に対する同時探索の仮説生成回数の比を示す．これらの図から，逐次探索では，要素言語モデル数にほぼ比例して，仮説生成回数が増加しているのに対して，同時探索では，要素言語モデル数の増加による仮説生成回数の増加が抑制されていることが分かる．同時探索では，言語モデル数が 1 から 1+1 になるときに約 1.6 倍に増加するが，それ以後はあまり増加していない．平均すると，言語モデル数が 10+1 のときの仮説生成回数は，言語

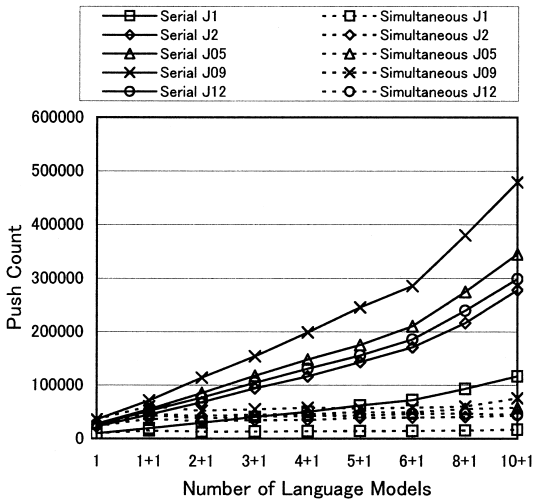


図 4 仮説生成回数の比較（$n = 20 \times (C + 1)$）
Fig. 4 Comparison of the counts of generated hypotheses ($n = 20 \times (C + 1)$).

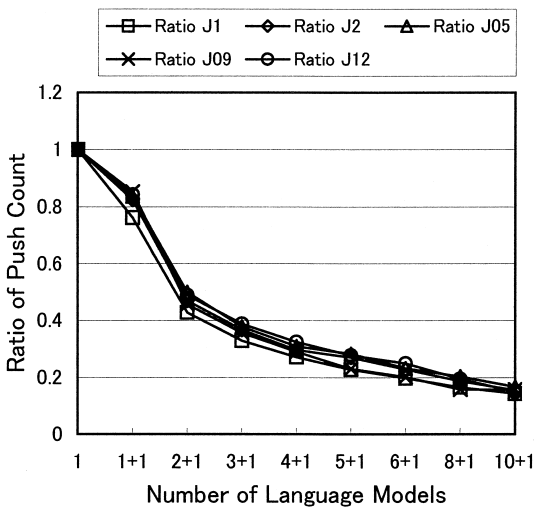


図 5 仮説生成回数の比（$n = 20 \times (C + 1)$）
Fig. 5 Ratio of the counts of generated hypotheses ($n = 20 \times (C + 1)$).

モデル数が 1 のときの約 1.9 倍，また，言語モデル数が 1+1 のときの約 1.2 倍であった．このように同時探索により，仮説生成回数を抑制することができ，探索空間が減少していることが分かる．

図 6 にスタックに積まれた最大の仮説数☆（スタックの深さ，平均値）を示す．逐次探索と同時探索で，

☆ 逐次探索は，各言語モデルごとに個別のスタックを用意して，探索するものであるが，これでは，スタックの総使用量が要素言語の数に比例して増大する．そこで，逐次探索の実験では，各言語モデルの探索で，1 本のスタックを使いまわすこととし，このときスタックに積まれた最大の仮説数の最大値を計測した．

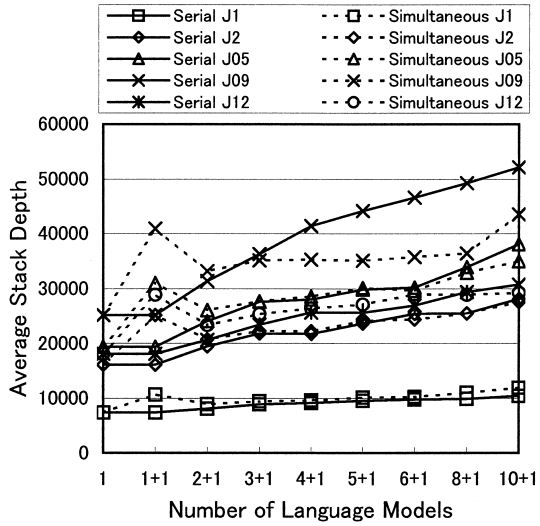


図 6 スタックの深さ（$n = 20 \times (C + 1)$）
Fig. 6 A maximum depth of the stack ($n = 20 \times (C + 1)$).

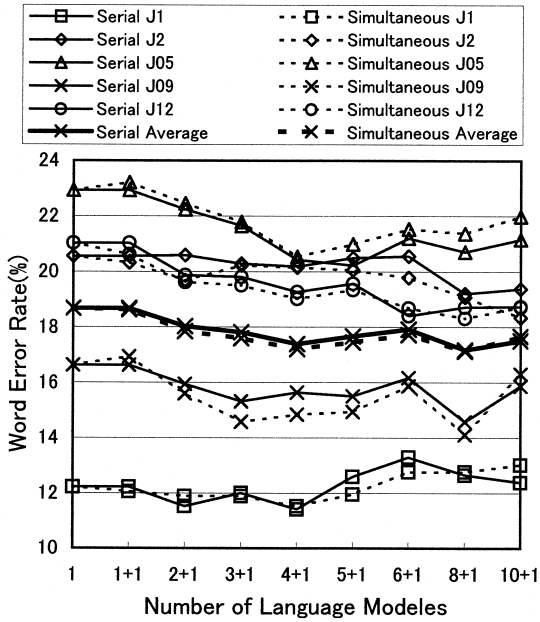


図 7 単語誤り率の比較（$n = 20 \times (C + 1)$）
Fig. 7 Comparison of the word error rates ($n = 20 \times (C + 1)$).

スタックの深さに，評価セットによる差が見られるものの，オーダ的な違いは見られなかった．

図 7 に単語誤り率の比較を示す．逐次探索から同時探索にすることで，評価セットにより，単語誤り率の傾向に，相違が見られる．**図 8** は逐次探索に対する同時探索の誤り率の変化を平均値と標準偏差で示す．言語モデル数が 10+1 のとき，単語誤り率が 0.16±0.75%増加している以外は，個々の評価セットによるばらつき

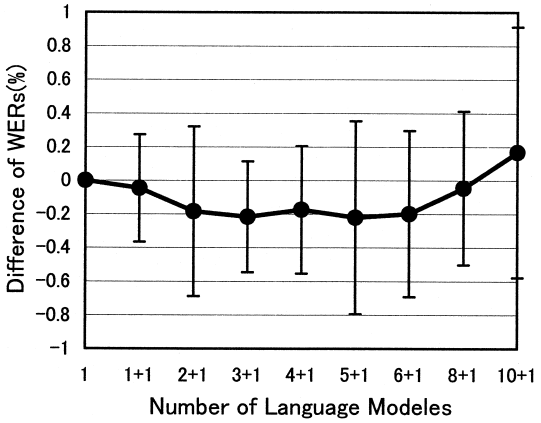


図 8 単語誤り率の差異（$n = 20 \times (C + 1)$）
Fig. 8 Difference of the word error rates ($n = 20 \times (C + 1)$).

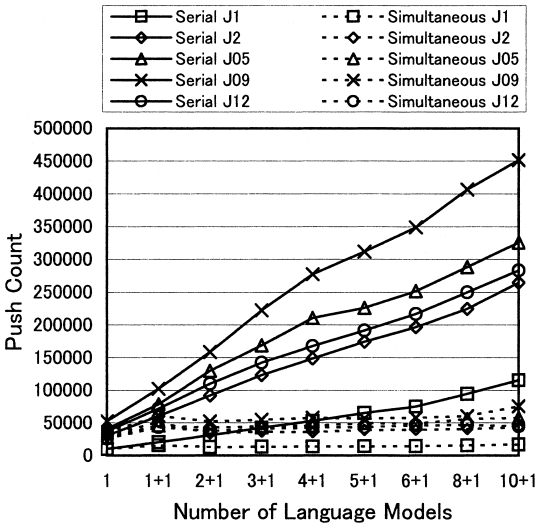


図 9 仮説生成回数の比較（$n \simeq 200$）
Fig. 9 Comparison of the counts of generated hypotheses ($n \simeq 200$).

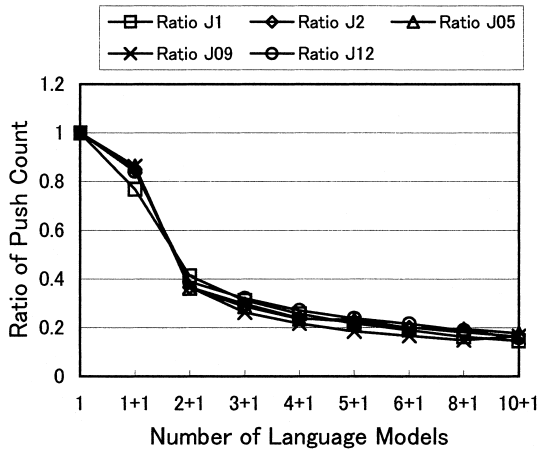


図 10 仮説生成回数の比（$n \simeq 200$）
Fig. 10 Ratio of the counts of generated hypotheses ($n \simeq 200$).

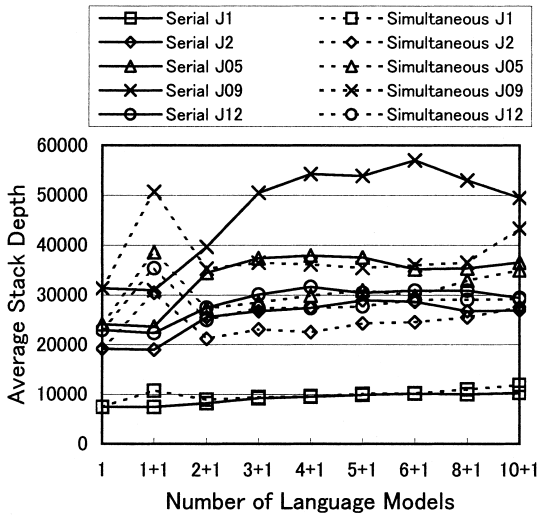


図 11 スタックの深さ（$n \simeq 200$）
Fig. 11 Maximum depth of the stack ($n \simeq 200$).

を除くと，平均的には，同時探索による誤り率の劣化は見られない．

4.7 探索解の総数を一定にした比較

逐次探索と同時探索で探索解の総数が言語モデル数によらずほぼ 200 になるように設定した．仮説生成回数を**図 9** に示す．**図 10** に逐次探索に対する同時探索の仮説生成回数の比を示す．**図 11** にスタックに積まれた最大の仮説数（スタックの深さ，平均値）を示す．また，単語誤り率を**図 12** に示す．また，逐次探索から同時探索にするときの単語誤り率の変化を平均と標準偏差で**図 13** に示す．仮説生成回数は，同時探索では，平均すると，言語モデル数 10+1 のとき，言語モデル数 1 のときの約 1.8 倍，言語モデル数 1+1 のときの約 1.1 倍であった．また，逐次探索から同時探索への単語誤り率の増加は，言語モデル数 8+1 のときで，-0.07 ± 0.52%であった．前節の $n = 20 \times (C + 1)$ としたときの結果と比較すると，平均の単語誤り率の差は，言語モデル数 8+1 のとき，逐次探索で，0.01%（$= 17.18 - 17.17$），同時探索で -0.02%（$= 17.11 - 17.13$）であり，性能的には同等であった．

4.8 同時探索の探索解総数を増大する効果

同時探索で探索解の総数を増大させ，$n = 2000$ の場合を試みる．仮説生成回数を $n = 200$ の結果と比較して**図 14** に示す．また，単語単位の誤り率を**図 15** に示す．$n = 200$ から $n = 2000$ としても仮説生成回

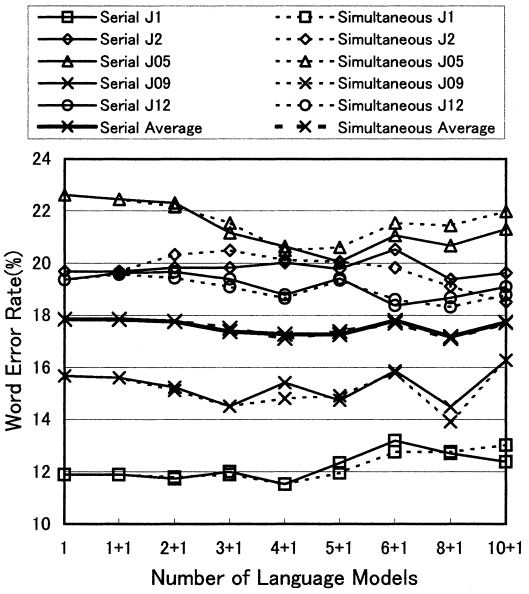


図 12 単語誤り率の比較（$n \simeq 200$）
Fig. 12 Comparison of the word error rates ($n \simeq 200$).

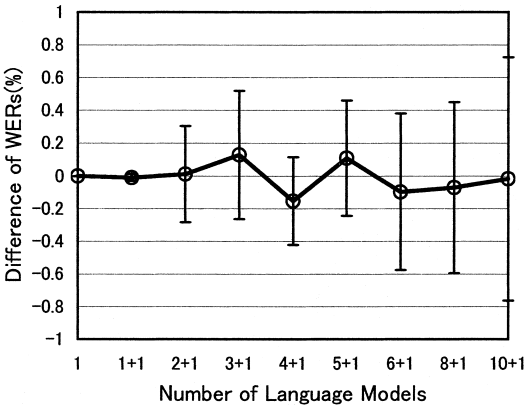


図 13 単語誤り率の差異（$n \simeq 200$）
Fig. 13 Difference of the word error rates ($n \simeq 200$).

数の増加はわずかである．また，単語誤り率はほとんど変化していない．したがって，$n = 200$ でも，最適な解が得られていると考えられる．

4.9 考　　察

表 6 にここまでの実験で得られた代表的な結果をまとめる．表から，並列同時探索により仮説展開数が削減される様子が分かる．たとえば，言語モデル数が 8+1 で $n \simeq 200$ のとき，逐次探索では言語モデル数 1 のときの 10.46 倍の仮説展開が必要である．一方，同時探索では 1.82 倍の仮説展開ですんでいる．このため，逐次探索に対する同時探索の仮説展開数は約 0.17 倍（=1.82/10.46）となっており，仮説展開数削減に関する提案手法の効果が確認された．また，このとき，

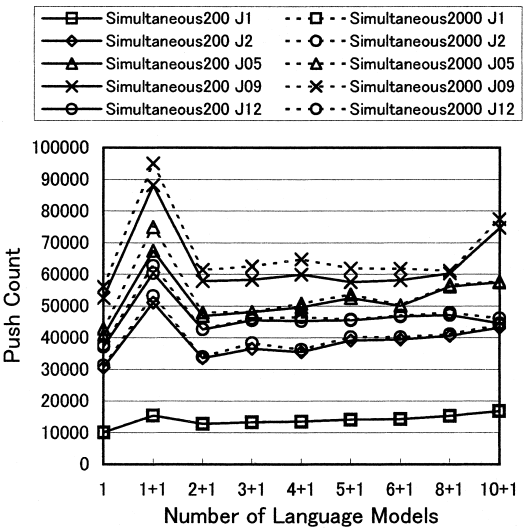


図 14 仮説生成回数（$n = 200$ と $n = 2000$）
Fig. 14 Counts of generated hypotheses ($n = 200$ and $n = 2000$).

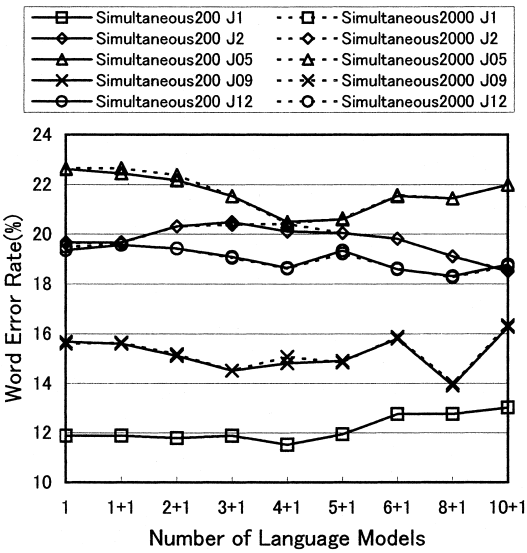


図 15 単語誤り率（$n = 200$ と $n = 2000$）
Fig. 15 Word error rates ($n = 200$ and $n = 2000$).

単語誤り率について，逐次探索では 8.0%の削減率であるのに対して，同時探索では 8.4%の削減率が得られており，同時探索でも，逐次探索の単語誤り率はほぼ維持されることが分かる．

表 7 に単一言語モデルおよび言語モデル数 8+1 の並列言語モデルに対する各探索法の実行時間（CPU: Pentium4-2.0 GHz，メモリ 2 G バイト，OS: Linux の計算機上で計測）を示す．同時探索による並列言語モデルの探索時間は，単一言語モデルの探索時間の平均約 1.5 倍である．一方，逐次探索では単一言語モデ

表 6 探索条件と平均探索性能の関係
Table 6 Relation between the search conditions and the average performances.

探索条件			仮説展開回数（比）	単語誤り率%（削減率%）
探索	解の総数 n	言語モデル数 $C+1$		
逐次	20	1	24178 (1.00)	18.67 (0.0)
逐次	200	1	34171 (1.41)	17.84 (4.4)
逐次	2000	1	35581 (1.47)	17.86 (4.3)
逐次	180	8+1	240748 (9.96)	17.17 (8.0)
逐次	198	8+1	252834 (10.46)	17.18 (8.0)
同時	180	8+1	43715 (1.81)	17.13 (8.2)
同時	200	8+1	43895 (1.82)	17.11 (8.4)
同時	2000	8+1	44419 (1.76)	17.12 (8.3)

表 7 探索時間の比較（$n \simeq 200, C = 8$）
Table 7 Comparison of search times ($n \simeq 200, C = 8$).

評価セット	探索時間（入力音声 1 秒あたりの処理時間，秒）		
	単一言語モデル	同時探索	逐次探索
JNS05	26.86	51.56	93.12
JNS09	29.06	33.24	119.60
JNS12	24.50	34.90	61.38
JNSJ1	2.70	6.79	22.22
JNSJ2	17.98	25.39	53.78
平均	20.22	30.37	70.02

ルの約 3.5 倍の時間を要している．この結果から，探索時間についても，提案手法の効果が確認された．

並列言語モデル化の効果は，対象分野の広がりの程度や要素言語モデルの構成（要素言語モデル数や各要素言語モデルがカバーする範囲など）によって変化すると考えられる．本実験では表 6 下から 2 番目の条件で単一言語モデルに対して 1.56%の誤り率減少（誤り削減率では 8.4%）の効果が確認できた．一方，実行時間は単一言語モデルに対して平均約 1.5 倍増加した．並列言語モデルによる認識性能向上という利点と実行時間増加という欠点を比較すると必ずしも並列言語モデルの有効性は明確ではない．しかし，要素言語モデルの構成法を改良することによって，認識性能向上という利点をより大きくできる可能性はある．今後，このような並列言語モデル自体の認識性能向上により，実行時間増加の欠点は問題にならない程度のものであると考えられる．なお，並列言語モデルの認識性能向上については本論文の対象の範囲外とする．

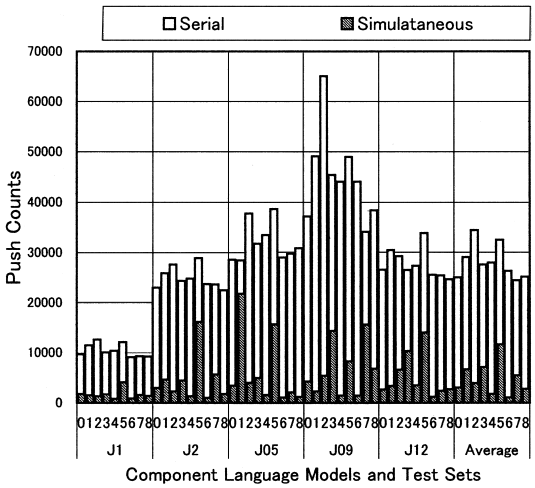


図 16 並列言語モデルの各要素の仮説生成回数（$n \simeq 200, C = 8$）
Fig. 16 Counts of generated hypotheses for each component language model ($n \simeq 200, C = 8$).

同時探索では，評価セットにより，特定の言語モデルに偏って，仮説展開が行われているかを調べた．**図 16** に言語モデル数 8+1 の並列言語モデルに対する各探索法の仮説生成回数を評価セット，要素言語モデル（要素言語モデルの “0” は全体モデル）ごとに示す．逐次探索では，評価セット “J09” に対して，要素言語モデル “2” で仮説生成回数の増加が見られるが，平均すると，要素言語モデルに依存する際立った偏りはない．一方，同時探索では，たとえば，“J05” の “1” と “5”，“J09” の “3” と “7”，“J12” の “5” のように，特定の要素言語モデルで仮説生成回数の増大が見られる．また，平均すると要素言語モデル “5” で増大が見られる．このように，同時探索では，評価セットにより，特定の要素言語モデルに偏った仮説の展開が行われることが分かる．同時探索におけるこのような仮説生成回数の要素言語モデル間の偏りが逐次探索に対する探索効率の向上をもたらしていると考えられる．

5. むすび

話題ごとに用意された複数の言語モデルを備える大語彙連続音声認識システムにおいて，これらの複数の言語モデルを同時に考慮して単語列を探索する手法を提案した．

提案した同時探索法によると，全言語モデルを順番に用いて単語列を探索する逐次探索に比較して，仮説展開回数を抑制でき，逐次探索と同等の認識性能が得られる．

新聞記事に対する実験結果から，言語モデル数を 9

としたとき提案手法による仮説展開数は言語モデル数1のときの1.8倍である．一方，これは逐次探索の10倍に対して1/5未満にずぎなかった．また，このとき単語単位の誤り率削減率は8.4%であり，逐次探索とほぼ同じ誤り削減率を得，本提案手法の有効性が確認された．

なお，本論文では，同時探索を，音響モデルと言語モデルを独立に扱う2段階探索法の2段目の並列言語モデルの探索に適用した．この場合，各モデルの尤度は言語モデルの尤度のみからなる．しかし，同時探索は，複数のモデル中で尤度最大解を求める一般性のある探索手法であるため，各モデルの尤度が音響モデルと言語モデルの尤度の組合せである場合（たとえば，文献16)の探索方式）でも，ヒューリスティックを適切に与えることによって，本手法の適用が可能であると考えられる．これについては，今後の検討課題とする．

謝辞 本論文の提案手法の評価においては，「日本音響学会―新聞記事読上げ音声コーパス―」，毎日新聞社の「毎日新聞CD-ROM 1991–1995」，奈良先端科学技術大学院大学の松本研究室より公開されている「日本語形態素解析システム―茶筅―」を用いた．ここに謝意を表する．

参 考 文 献

1) Iyer, R., Ostendorf, M. and Rohlicek, J.R.: Language modeling with sentence-level mixtures, *Proc. ARPA Workshop on Human Language Technology*, pp.82–87 (1994).
2) Carter, D.: Improving language models by clustering training sentences, Technical report, SRI International (1994).
3) Iyer, R. and Ostendorf, M.: Modeling long distance dependence in language: topic mixtures vs. dynamic cache models, *Proc. ICSLP96*, pp.236–239 (1996).
4) Clarkson, P.R. and Robinson, A.J.: Language model adaptation using mixtures and an exponentially decaying cache, *Proc. ICASSP97*, pp.799–802 (1997).
5) Martin, S.C., Liermann, J. and Ney, H.: Adaptive topic-dependent language modelling using word-based varigrams, *Proc. Eurospeech97*, pp.1447–1450 (1997).
6) 清水 徹，大野晃生，樋口宜男：文のクラスタリングに基づく統計的言語モデル，音講論，pp.1–6–14 (1998-03).
7) 政瀧浩和：MAP推定に基づくN-gram言語モデルの自動分類されたコーパスへの適応，音講論，pp.1–6–19 (1998-03).
8) Itsui, H., Maruta, Y., Abe, Y. and Nakajima, K.: A study on topic-dependent language modeling, *Proc. WESTPRAC VII*, pp.137–140 (2000).
9) Lane, I.R.，河原達也，松井知子，中村 哲：話題同定に基づく言語モデル切替えによる対話音声認識，情報処理学会音声言語情報処理研究会資料，pp.SLP–44–25 (2002-12).
10) 阿部芳春，伍井啓恭，丸田裕三，中島邦男：認識誤り傾向の確率モデルを用いた2段階探索法による大語彙連続音声認識，信学論，Vol.J83-D-II, No.12, pp.2545–2553 (2000).
11) Paul, D.B.: An efficient A* stack decoder algorithm for continuous speech recognition with a stochastic language model, *Proc. ICASSP92*, pp.I-25–I-28 (1992).
12) Manning, C.D. and Schütze, H.: *Foundations of statistical natural language processing*, The MIT Press (1999).
13) 毎日新聞社：毎日新聞CD-ROM1991-95.
14) 阿部芳春，伍井啓恭，丸田裕三，中島邦男：混合言語モデル作成のためのコーパスのクラスタ分割の検討，音講論，pp.3–P–17 (2001-03).
15) 松本裕治，北内 啓，山下達雄，今 一修，今村友明：日本語形態素解析システム『茶筅』version1.0使用説明書，Technical report，奈良先端科学技術大学院大学 (1997).
16) 河原達也，李 晃伸，小林哲則，武田一哉，峯松信明，伊藤克亘，伊藤彰則，山本幹雄，山田 篤，宇津呂武仁，鹿野清宏：日本語ディクテーション基本ソフトウエア（97年度版），音響誌，Vol.55, No.3, pp.175–180 (1999).

(平成15年10月27日受付)
(平成16年10月 4 日採録)

中島 邦男（正会員）

1968年名古屋工業大学工学部電子工学科卒業．1970年同大学大学院修士課程修了．同年三菱電機（株）入社．以来，データ伝送，音声認識および音声合成・音声符号化の研究開発に従事．同社情報技術総合研究所音声言語技術部長，インタフェース部門長等を経て，2004年より横浜国立大学産学連携推進本部．1995年度日刊工業新聞社十大新製品賞，2003年度情報処理学会業績賞を受賞．電子情報通信学会，日本音響学会各会員．

阿部　芳春

1976 年東京工業大学工学部電子工学科卒業．1981 年同大学大学院博士課程修了．同年三菱電機（株）入社．以来，音声認識の研究に従事．現在，同社情報技術総合研究所音声処理技術部主席研究員．日本音響学会，電子情報通信学会各会員．工学博士．

伍井　啓恭（正会員）

1986 年東京電機大学工学部電気通信工学科卒業．同年三菱電機（株）入社．1996 年〜1998 年日本電子化辞書研究所に出向．1997 年〜1998 年東京工業大学大学院情報理工学研究科研究生．現在，三菱電機（株）情報技術総合研究所音声処理技術部主席研究員．音声認識処理，自然言語処理に関する研究に従事．日本音響学会会員．

丸田　裕三（正会員）

1984 年千葉大学理学部物理学科卒業．1986 年同大学大学院修士課程修了．同年三菱電機（株）入社．以来，文字認識および音声認識に関する研究に従事．現在，同社自動車機器開発センター開発第二部専任．