論　文

機械学習を用いた複数の大語彙連続音声認識モデルの出力の混合

宇津呂武仁[†]　小玉　康広[††]　渡邉　友裕[†††]　西崎　博光[††††]
中川　聖一[†††]

Combining Outputs of Multiple LVCSR Models by Machine Learning

Takehito UTSURO[†], Yasuhiro KODAMA[††], Tomohiro WATANABE[†††], Hiromitsu NISHIZAKI[††††], and Seiichi NAKAGAWA[†††]

あらまし　本論文では，様々な認識特性をもった複数の大語彙連続音声認識モデルが利用できるような状況において，信頼性の高い認識結果を柔軟に組み合わせる混合規則を機械学習の手法により学習し，この規則を用いて，複数の大語彙連続音声認識モデルの出力を混合する方式を提案する．新聞読上げ音声及びニュース音声を評価音声データとして，デコーダ，音響モデルの異なる 26 種類の大語彙日本語連続音声認識モデルの出力を混合する評価実験を行ったところ，機械学習を用いた混合手法により，認識率最大の単独モデル，及び，ROVER 法のような（重み付き）多数決を用いた混合の単語認識率を上回る性能が達成できた．また，（重み付き）多数決に基づく混合手法の場合，認識率の低いモデルが多数派を占めると，混合結果の性能が認識率の低い多数派のモデルに強く影響されるという欠点があったが，機械学習（特に SVM — Support Vector Machines）を用いた混合手法では，認識率の高いモデルが多数派であるか少数派であるかにかかわらず，混合結果の単語認識率を安定して高く維持することができた．

キーワード　大語彙連続音声認識，機械学習，複数モデル混合，SVM，信頼度尺度

1. まえがき

近年，音声認識結果の正解部分と誤り部分を分離することを目的として信頼度（Confidence Measure）の研究が行われている（例えば，連続音声認識では [5], [17] など）．これまで提案されてきた信頼度尺度の多くは，いずれも，単一の認識エンジン・認識モデルが出力する認識結果を用いて，その正解部分と誤り部分を分離するというものであった．一方，連続音声認識の認識率そのものの向上を目的とする研究においては，複数の認識システムの出力を統合する方式（ROVER 法 [2]）も提案され，一定の効果が報告されている（例えば，文献 [11] など）．我々は，ROVER 法のような（重み付き）多数決法が認識率の改善に効果的であることを考慮して，音声認識結果の正解部分と誤り部分を分離するための信頼度尺度として，複数の大語彙連続音声認識モデルの出力の共通部分を用いる方法を提案し，次のようにその有効性を示した [14]．評価実験の結果では，デコーダ及び音響モデルが異なる二つのモデルについて，出力の共通部分の信頼度を評価したところ，最も高い性能が達成された．新聞読上げ音声では，認識結果の約 87%の単語について，99%近い精度で正解単語を推定することができ，ニュース音声でも，認識結果の約 64%の単語について，95%近い精度で正解単語を推定することができた．また，同一のデコーダを用いた場合の正解単語推定精度は，デコーダが異なる場合を約 1%程度下回るものの，ほぼそれに匹敵する性能を達成した．更に，デコーダ，音響モデルの異なる 26 種類の大語彙日本語連続音声認識モデルについて，品詞・音節数といった単語のもつ各種特徴と，二つのモデル組の出力の間の共通部分が認識正解となる

† 京都大学大学院情報学研究科知能情報学専攻，京都市
Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto-shi, 606–8501 Japan
†† ソニー株式会社，東京都
Sony Corporation, 6–7–35 Kitashinagawa, Shinagawa-ku, Tokyo, 141–0001 Japan
††† 豊橋技術科学大学工学部情報工学系，豊橋市
Department of Information and Computer Sciences, Toyohashi University of Technology, Tenpaku-cho, Toyohashi-shi, 441–8580 Japan
†††† 山梨大学大学院医学工学総合研究部，甲府市
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 4–3–11 Takeda, Kofu-shi, 400–8511 Japan

割合の相関を評価した．例えば，名詞の場合は，モデル A の出力とモデル B の出力の共通部分が認識正解となる割合が最も高く，逆に，動詞の場合は，モデル C の出力とモデル D の出力の共通部分が認識正解となる割合が最も高い，というように，単語の品詞によって，高い信頼性を示すモデルの組合せが異なり，また，単語の音節数によっても，高い信頼性を示すモデルの組合せが異なることを示した [12], [13].

以上の結果から，様々な認識特性をもった複数の大語彙連続音声認識モデルが利用できるような状況において，品詞・音節数などの単語の特徴に応じて信頼性の高い認識結果を柔軟に組み合わせることができれば，音声認識率そのものを改善できる可能性があることが分かった．ここで，複数の大語彙連続音声認識モデルの出力を混合する方式としては，多数決に基づく手法 [2], [11] が既に提案されているが，この手法では認識率の低いモデルが多数派を占めるような場合に，混合結果の性能が，認識率の低い多数派のモデルに強く影響されるという欠点がある．そこで，本論文では，機械学習の手法を用いて，品詞・音節数などの単語の特徴に応じて信頼性の高い単語を選択する規則を学習し，この規則を用いて，複数の大語彙連続音声認識モデルの出力を混合する方式を提案する．機械学習の方式としては，学習・適用が比較的容易である決定リスト [18]，及び，一般により性能が高いとされている SVM（Support Vector Machines）[15] を適用する．

新聞読上げ音声及びニュース音声を評価音声データとして，デコーダ，音響モデルの異なる 26 種類の大語彙日本語連続音声認識モデルの出力を混合する評価実験を行ったところ，機械学習を用いた混合手法により，認識率最大の単独モデル，及び，（重み付き）多数決を用いた混合の単語認識率を上回る性能が達成できた．また，SVM と決定リスト学習の比較では，SVM の方が高い性能を示した．更に，（重み付き）多数決による混合においては，混合に参加するモデル中で，認識率の高いモデルが多数派の場合は，多数派のモデルの影響により混合結果の単語認識率が高く維持されるが，認識率の低いモデルが占める割合が増えると，それらの影響により混合結果の性能が低下した．一方，機械学習（特に SVM）を用いた混合では，混合に参加するモデル中で，認識率の高いモデルが多数派であるか少数派であるかにかかわらず，混合結果の単語認識率は安定して高く維持できており，学習性能の高さを実証することができた．

2. 機械学習による複数の大語彙連続音声認識モデルの出力の混合

2.1 混合手順の概要

複数の大語彙連続音声認識モデルの出力の混合のタスクは，一般に，

- DP マッチングにより，複数の大語彙連続音声認識モデルによる認識結果の単語列の対応付けを行う，
- 単語列の対応付けの結果において，競合する単語のうち，どの単語が認識結果として適切かを判定する，

という二つの過程からなる．ここで，本論文では，このタスクに対して，機械学習の手法を適用するというアプローチをとる．具体的には，訓練用音声データを利用して，複数の大語彙連続音声認識モデルの出力の混合を行う規則を機械学習の手法により学習し，新規の音声データの認識においては，この規則を適用することにより，複数モデルの出力の混合を行い，認識結果を出力する．本節では，この各過程における処理の概要について述べる．

まず，音声データを，話者が重複しないように，訓練データセットと評価データセットに分割する．次に，複数の認識モデルを用いて訓練データセットの認識を行い，各々のモデルが第 1 位の認識結果として出力した結果を，各モデルによる認識結果の単語列とする．そして，訓練データセットの単語認識率が最大のモデルによる認識結果を基準とし，この基準と他のモデルによる認識結果の DP マッチングを行い，競合する単語の対応付けを行う．ここで，基準となる認識結果における各単語に対して，置換に相当する単語及び直前の単語との間の位置への挿入に相当する単語を競合する単語とみなし，競合単語集合を構成する．ただし，文末位置への挿入に相当する単語については，それらの挿入単語同士が競合しているとみなし，基準となる認識結果に含まれる単語を加えずに競合単語集合を構成する．

次に，それらの競合単語集合の各単語が認識正解であるか，それとも，認識誤りであるかを，機械学習の際に判別対象とするクラス c とし，このクラスを決定するための規則を学習する．機械学習における各単語の素性としては，各単語を出力したモデルの情報・単語の品詞・音節数・音響スコア・言語スコアを用い，これらの有効性を評価する．

評価データセットに対して，機械学習により得られ

た正誤判別規則を適用し，複数モデルによる認識結果の混合を行う場合についても，上述の手順に従って，まず，競合単語集合を構成する．そして，競合単語集合中の各単語について，正誤判別規則によりその正誤を判別する．ただし，各正誤判別規則には信頼度が付与されており，更に，規則適用時には，信頼度の下限値が設定されている．そして，競合単語集合中で，この下限値以上の信頼度をもつ正誤判別規則により認識正解と判定された単語のうちで，最も大きい信頼度をもつ規則が適用された単語を認識結果として出力する．信頼度の下限値を超える規則が適用できない場合は，その競合単語集合中には高信頼度の認識結果は存在しないとして，単語を出力しない．

2.2 機械学習法の適用

前節の手順を踏まえて，本節では，SVM [15] 及び決定リスト学習 [18] という二通りの機械学習法を適用する場合の混合手法の詳細について述べる．どちらの機械学習法も，これまで，統計的自然言語処理の分野においてよく用いられており，構文解析におけるあいまい性の解消や語義のあいまい性の解消等の分類問題における特性などが知られている．それらの分野においては，一般には，SVM は最も高性能な学習法の一つとして位置付けられており，一方，決定リスト学習は，SVM などのような高い性能は期待できないが，実装が容易で簡便な学習法の一つと考えられている（両者の定量的比較については，例えば，語義あいまい性解消については，文献 [8] を参照．その他，日本語固有表現抽出においても，いくつかの文献を参照することにより，ほぼ同一条件での定量的比較が可能）．本論文では，複数の大語彙連続音声認識モデルの出力を混合するタスクに対して，このような対照的な機械学習法を適用することにより，このタスクにおける機械学習の枠組の有効性を検証する．

2.2.1 SVM

SVM [15] は，個々の素性を次元とする多次元空間中の点として記述された各事例に対して，全事例を二クラスに分類する境界面を学習することによりクラス判別規則を学習するという，機械学習の一つの手法である．SVM は，境界面をはさんで最も近い位置にある二つの事例の間の距離（マージン）を最大化するという基準により境界面を学習しており，また，一般に，カーネル関数を用いることにより非線形な境界面を学習することも可能である．

SVM を用いた複数モデルの出力の混合タスクにおいては，前節で述べたとおり，各単語の正誤を判別対象のクラス c とし，素性としては以下の 5 種類を用いた(注1)．i) その単語を出力したモデルの情報（**3.1** で述べた 26 種類のモデルのうちのどのモデルにより出力されたかの情報），ii) その単語の品詞（日本語形態素解析システム「茶筌(せん)」(注2)の最も粗い 9 品詞），iii) その単語の音節数，iv) その単語を出力したモデルの情報とその単語の音響スコアをフレーム数で割ったスコア [10] を結合したもの，v) その単語を出力したモデルの情報とその単語の言語スコア [10] を結合したもの．SVM 学習・適用のツールとしては，Tiny-SVM(注3)を用いた．カーネル関数としては，多項式カーネルの一次及び二次を評価したが，一次の方が性能がよかったので，**4.** では，一次の多項式カーネルを用いた場合の評価結果を示す．また，学習時における判別誤りの許容度を表す-c オプションについては，1/((訓練事例ベクトルの大きさ)^2 の平均) を用いた．更に，正誤判別規則の信頼度としては，評価事例と境界面との間の距離を用いた．

2.2.2 決定リスト学習

決定リスト [18] は，ある素性のもとでクラスを決定するという規則を優先度の高い順にリスト形式で並べたもので，適用時には優先度の高い規則から順に適用を試みていく．本論文では，各規則の優先度として，素性 f の条件のもとでの，クラス c の条件付き確率 $P(c \mid f)$ を用い，この条件付き確率順に決定リストを構成する．

決定リストを用いた複数モデルの出力の混合タスクにおいては，前節で述べたとおり，各単語の正誤を判別対象のクラス c とし，素性としては，以下の二通りの設定を評価した(注4)．

i) 各単語を出力した二つのモデルの組を用いる．
ii) 各単語を出力した二つのモデルの組，その単語

(注1)：音響スコア・言語スコアについては，音響スコアをフレーム数で割らない場合，文全体の音響スコア・言語スコアを素性として用いた場合や素性として追加した場合など，様々な設定を評価したが，ここで述べる設定の性能を超えるものはなかった．
(注2)：http://chasen.aist-nara.ac.jp/
(注3)：http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM/
(注4)：決定リスト学習における素性の設定においては，「複数モデルの出力の混合において，ある単語が出力として選択されるためには，少なくとも二つ以上のモデルによって出力される必要がある」という制約を課している．前節で説明した SVM の素性の場合には，このような制約を課していないが，このような制約を課した素性の設定のもとで，SVM により複数モデルの出力の混合を行う評価実験も行った．この評価実験の結果においても，**4.** の評価結果と同様に，SVM によるモデル混合の性能が決定リスト学習によるモデル混合の性能を上回った．

の品詞（「茶筌」の最も粗い 9 品詞），その単語の音節数を用いて，モデル組，モデル組・品詞の結合，モデル組・音節数の結合，モデル組・品詞・音節数の結合の 4 種類の素性を作成し，この全素性を用いる．

決定リストの適用の際には，決定リストの各規則に対して，素性 f の条件のもとでのクラス c の頻度 $freq(f,c)$ の下限，及び，条件付き確率 $P(c \mid f)$ の下限を設け，競合単語集合中で，これらの下限を満たす単語が存在すれば，その中で最も優先度（条件付き確率 $P(c \mid f)$）の高い単語を選択する．競合単語集合中で，頻度及び条件付き確率の下限を満たす単語が存在しない場合には，その競合単語集合中には高信頼度の認識結果は存在しないとして，単語を出力しない．

2.3 （重み付き）多数決による混合

本論文では，機械学習による複数モデルの出力の混合に対するベースラインとして，ROVER 法 [2] のような（重み付き）多数決法についても評価を行い，機械学習による混合との性能比較を行う．（重み付き）多数決法においては，訓練データを用いた学習は不用であるので，評価データに対する複数モデルの出力の混合のみを行う．この場合も，**2.1** で述べた手順により競合単語集合を構成し，各競合単語集合における（重み付き）多数決法により，認識結果となる単語を決定し出力する．ここで，重み付き多数決においては，各単語の重みとして，その単語が含まれる文に各モデルが付与するスコアを求め，この文スコアからその文全体の単語認識率の推定値を算出し，この値を，その文中に含まれる各単語の重みとした(注5)．多数決においては，競合単語集合中の各単語について，この重みの和を求め，重みの和が最大となる単語を認識結果として出力する．一方，重みなし多数決においては，各単語の重みをすべて 1 として多数決を行う．ただし，競合単語集合中において，重みの和が最大となる単語が複数個得られる場合は，その競合単語集合からは認識結果の単語を出力しない．

3. 実験条件

本章では，本論文の実験で用いる大語彙日本語連続音声認識モデル，及び，音声データの概要について述べる．本章で述べる内容の詳細については，文献 [14] を参照されたい．

3.1 大語彙日本語連続音声認識モデル

大語彙連続音声認識モデルとしては，SPOJUS [1]（音響モデル [9] は，12 kHz/16 kHz サンプリング，フレーム周期 8/10 ms，特徴ベクトルはセグメント単位/フレーム単位の MFCC の 2 種類，音節モデル，無音モデル有・無二通り，全/対角共分散行列，継続時間制御/自己遷移ループ，等合計 18 種類）及び Julius [4]（音響モデルは，16 KHz サンプリング，フレーム周期 10 ms，特徴ベクトルはフレーム単位の MFCC，トライフォン/モノフォン/PTM/音節モデル，無音モデル有・無二通り，の合計 8 種類）を使用した．言語モデルは，毎日新聞（45 カ月分）または NHK 汎用ニュース原稿（5 年分）から作成した tri-gram モデル（語彙サイズ 2 万）を用いた．言語モデルに含まれる各語は，単語・品詞・読みの 3 項組で表現されており，認識結果についても，単語・品詞・読みの 3 項組の列として出力される．

SPOJUS 及び Julius のどちらのデコーダも，2 パスで認識結果の探索を行う．1 パス目では，bi-gram 言語モデルと木構造辞書による 1-best 近似探索を行う．SPOJUS は，1 パス目の中間結果を N-best（200-best）候補として出力し，2 パス目で，前向き tri-gram 言語モデルを用いて N-best 候補のリスコアリングを行い，認識結果を出力する．一方，Julius は，1 パス目の中間結果を単語トレリスで出力し，2 パス目で，逆向き tri-gram 言語モデル及び音響モデルを用いて，単語トレリス空間の後向き探索を行い，認識結果を出力する．両者は，中間結果を N-best 候補で出力するか，単語トレリスで出力するかという点で異なっており，Julius の方がより多くの解候補を探索しているといえる．なお，SPOJUS のデコーダについては，改良版 [6] が報告されているが，本実験では文献 [1] のものを用いた．

3.2 音声データ

実験用音声データとしては，音声認識が比較的容易な新聞読上げ音声，及び，相対的に音声認識が容易でないニュース音声の 2 種類を用いる．

（1） IPA「日本語ディクテーション基本ソフトウェアの開発」プロジェクト [4] において，新聞記事読上げ音声データベース（JNAS）[3] から選定した 100 文

(注5)：具体的には，機械学習による混合において訓練用データとして用いた音声データを用いて，文ごとの単語認識率を求めておき，文スコアと文単位の単語認識率の間の線形変換式を推定する．そして，この線形変換式を用いて，評価用音声データにおける各文の文スコアから，その文の単語認識率の推定値を算出する．なお，この方式による重み付き多数決とは別に，単純に，訓練用データに対する各モデルの単語認識率を求めておき，各モデルごとに単語の重みを固定した重み付き多数決法の評価も行ったが，文ごとに単語の重みを変動させる方式の方が高性能であった．

（男性話者 10 人，1,565 語）．

（2） NHK のニュース「ニュース 7」と「おはよう日本」（1996 年 6 月 1 日）の 175 文（男性話者 10 人 — アナウンサー 2 人，レポーター 8 人，6,813 単語）．いずれの音声データも，音響モデルの学習には用いていない．新聞読上げ音声の認識精度は，SPOJUS で単語正解率 90.2 ～ 78.1%，単語正解精度 85.3 ～ 51.0%，Julius で単語正解率 93.0 ～ 72.7%，単語正解精度 90.4 ～ 69.4%，ニュース音声の認識精度は，SPOJUS で単語正解率 70.7 ～ 55.4%，単語正解精度 62.8 ～ 36.2%，Julius で単語正解率 71.7 ～ 49.0%，単語正解精度 68.8 ～ 39.7%であった(注6)．

4. 実験及び評価

3.1 で述べた全 26 種類の大語彙日本語連続音声認識モデルを用いて，複数モデルによる認識結果を混合する実験を行った．認識結果混合の対象となるモデルを選択する方法としては，二通りの方法を評価した．一つ目の方法は，（訓練用データにおける）単語正解率の高い順に $n(3 \leq n \leq 26)$ 個のモデルを選択する方法で，この方法を「単語正解率順」のモデル選択法と呼ぶ．二つ目の方法は，複数のモデルによる認識結果を足し合わせた結果，正解単語がなるべく多く含まれるような順にモデルを選択する方法である．具体的には，複数のモデルによる認識結果を足し合わせた上で，（訓練用データの）正解単語中のどれだけの単語が復元できるかの割合である「和の再現率」を測定し，この「和の再現率」がなるべく大きくなる順にモデルを選択する．この方法を「和の再現率順」のモデル選択法と呼ぶ．ただし，この方法では，1 番目のモデルとしては，単語認識率最大のモデルを選ぶ．

評価実験全体を通して，機械学習手法として SVM を用いて認識結果の混合を行う方法の性能が最も高く，また，SVM による混合の場合，混合の対象となるモデルが多ければ多いほど，混合結果における認識率が高くなるという傾向であった．特に，モデル選択法としては，「和の再現率順」のモデル選択法の方が，混合対象となるモデルが少ない時点での認識率が相対的に高く，優れた特性を示した．そこで，以下の評価結果においては，**4.3** において，「単語正解率順」及び「和の再現率順」という二つのモデル選択法を比較する以外は，すべて，「和の再現率順」でモデル選択を行った場合の評価結果を用いている．

以下では，まず，**4.1** において，単語認識率の評価尺度について述べる．**4.2** では，複数モデルの出力を混合する手法として，SVM によって学習した混合規則を用いる場合，決定リスト学習によって学習した混合規則を用いる場合，及び，（重み付き）多数決による場合の性能の比較を行う．**4.3** では，SVM を用いた混合について，「単語正解率順」に混合するモデルを選択する手法と，「和の再現率順」に混合するモデルを選択する手法の性能比較を行う．**4.4** では，SVM を用いた混合について，学習に用いる素性の性能の比較を行う．最後に，**4.5** では，複数の大語彙連続音声認識モデルの出力の混合において，異なるデコーダを用いた複数モデルの出力を混合する場合と，同一デコーダを用いた複数モデルの出力を混合する場合の性能の比較を行う．

4.1 評価尺度

本論文では，単語認識率の評価において，**3.2** において述べた単語正解率・単語正解精度の他に，以下の再現率・適合率・F 値 $_{\beta=1}$ も用いる(注7)．

$$\text{再現率} = \frac{\text{正解単語数}}{\text{正解文の単語数}}$$

$$\text{適合率} = \frac{\text{正解単語数}}{\text{認識結果の単語数}}$$

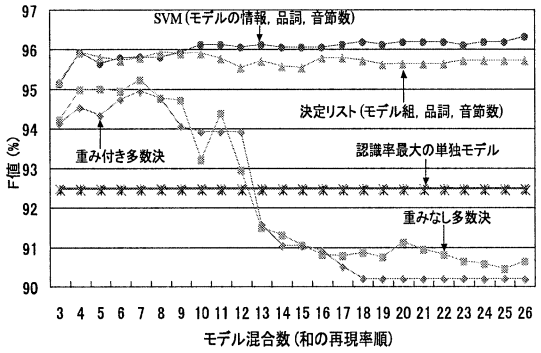
$$F\text{ 値}_{\beta=1} = \frac{2}{\frac{1}{\text{再現率}} + \frac{1}{\text{適合率}}}$$

4.2 混合手法の性能比較

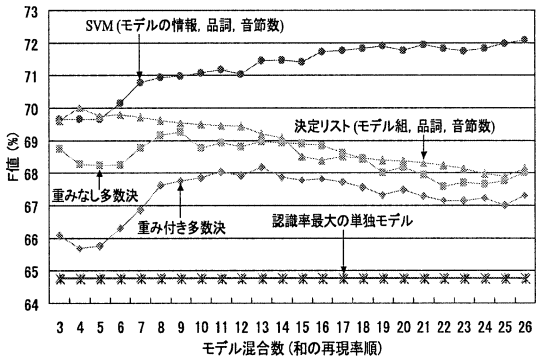
まず，複数モデルの出力を混合する手法として，SVM によって学習した混合規則を用いる場合，決定リスト学習によって学習した混合規則を用いる場合，及び，（重み付き）多数決による場合の単語認識の性能を比較した．図 1 ～ 図 3 に，混合対象となるモデルの数 n を変化させた場合の，各混合手法の F 値，単語正解率，単語正解精度の変化を，それぞれ示す．ただし，SVM においては，評価事例と境界面との間の距離の下限の値に応じて，単語認識における再現率及び適合率のトレードオフが生じ，また，決定リスト学習においても，素性 f の条件のもとでのクラス c の

(注6)：正解文の単語数を N，認識結果における正解単語数を C，置換誤り単語数を S，挿入誤り単語数を I，脱落誤り単語数を D とすると，単語正解率は $C/N = (N - S - D)/N$，単語正解精度は $(N - S - D - I)/N$ として定義される．

(注7)：本論文では，ここで導入した適合率を信頼度とする評価尺度 [14] との整合性をとるために，再現率・適合率・F 値 $_{\beta=1}$ を用いて，混合結果の単語認識率を評価する．なお，ここでの再現率は単語正解率と同じ評価式となっている．



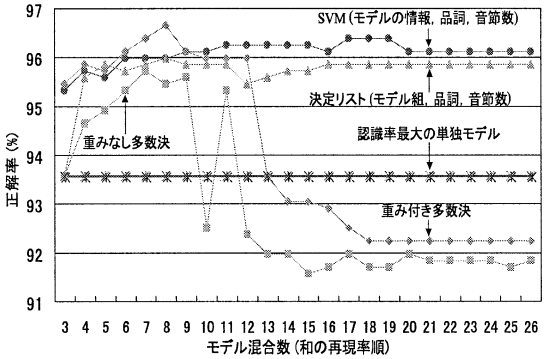
(a) 新聞読上げ音声



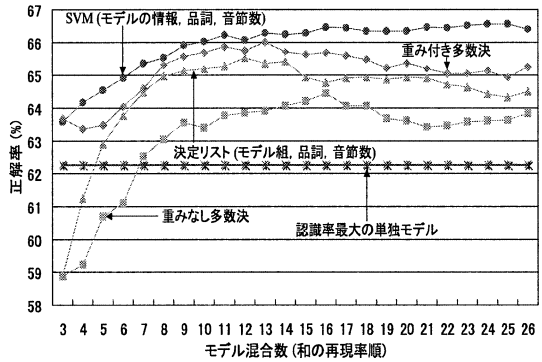
(b) ニュース音声

図 1　和の再現率順に選択した n ($3 \leq n \leq 26$) モデルの出力の混合における混合手法の性能比較（F 値）

Fig. 1　Comparing F-measures of methods for combining outputs of n ($3 \leq n \leq 26$) models. (selected so as to maximizing recall of union)



(a) 新聞読上げ音声



(b) ニュース音声

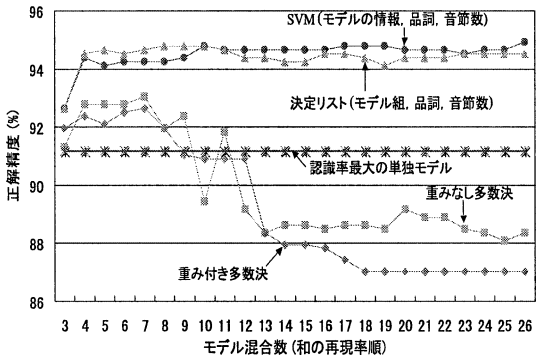
図 2　和の再現率順に選択した n ($3 \leq n \leq 26$) モデルの出力の混合における混合手法の性能比較（単語正解率）

Fig. 2　Comparing word correct rates of methods for combining outputs of n ($3 \leq n \leq 26$) models. (selected so as to maximizing recall of union)

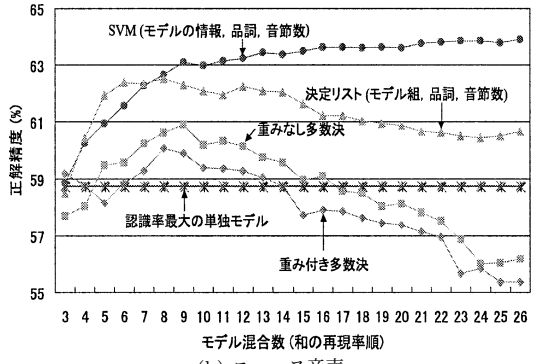
頻度 $freq(f, c)$ の下限，及び，条件付き確率 $P(c \mid f)$ の下限の値に応じて，単語認識における再現率及び適合率のトレードオフが生じる．そこで，これらの下限値については，図 1 の場合は，F 値が最大となる値を，図 2 の場合は，単語正解率が最大となる値を，図 3 の場合は，単語正解精度が最大となる値を，それぞれ用いる(注8)．同様に，図 1 ～ 図 3 の「認識率最大の単独モデル」としては，図 1 では F 値が最大となる単独モデルの値を，図 2 では単語正解率が最大となる単独モデルの値を，図 3 では単語正解精度が最大となる単独モデルの値を，それぞれ用いる．また，各手法間で，素性等の利用する情報を合わせるために，SVM の素性として **2.2.1** の iv) の音響スコア，及び，v) の言語スコアを用いない場合の性能を，また，決定リスト学習の素性としては，**2.2.2** の ii) の設定を用いた場合の性能を，それぞれ示す．

この結果から分かるように，新聞読上げ音声・ニュース音声とも，全 26 モデルの出力を混合した場合の性能を，F 値・単語正解率・単語正解精度の観点から総合的に比較すると，性能の高い順に，SVM，決定リスト学習，重みなし多数決，重み付き多数決，となった．また，機械学習を用いた混合においては，F 値・単語正解率・単語正解精度のいずれにおいても，単語認識

(注8)：今回の実験では，最適な下限値が選択できた場合にどの程度の性能が達成できるのかという観点で，評価データにおける性能が最大となるように下限値を設定した．ただし，実際には，より性能の高い機械学習法である SVM では，F 値，単語正解率，単語正解精度のそれぞれについて，音声データの種類・モデル混合数に依存せず，ほぼ一定の距離下限値において，最大の性能を達成していた．具体的には，F 値を最大とする距離下限値は，距離 0（すなわち境界面）よりわずかに認識誤り側となり，実際には，境界面をそのまま用いて認識結果の単語の正誤判定を行った場合とほぼ同等の性能であった．また，単語正解精度に関しては，F 値での下限値とほぼ同じかやや認識誤り側の距離下限値となり，境界面によって認識誤りと判定された単語のうちの一部についても，認識結果として出力する場合が最大となった．更に，単語正解率に関しては，距離下限値をもう少し認識誤り側に設定し，認識結果として出力する単語の割合を増やした場合に最大となった．



(a) 新聞読上げ音声



(b) ニュース音声

図 3 和の再現率順に選択した n ($3 \leq n \leq 26$) モデルの出力の混合における混合手法の性能比較（単語正解精度）

Fig. 3 Comparing word accuracy rates of methods for combining outputs of n ($3 \leq n \leq 26$) models. (selected so as to maximizing recall of union)

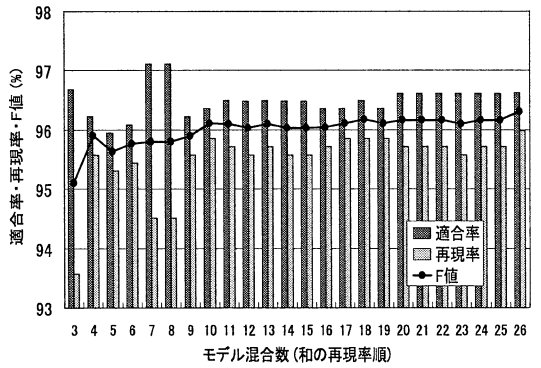
率最大の単独モデルの性能を上回っており，モデル混合の効果が確認できた．ただし，SVM と決定リスト学習の間には一定の性能差が認められており，統計的自然言語処理における両者の性能差と合致する傾向であるといえる(ニュース音声の場合は，モデル数が多くなると決定リスト学習の性能が低くなる傾向があるが，素性として **2.2.2** i) モデル組のみを用いた場合は，この傾向はそれほど顕著ではないという結果を得ている．この結果から，決定リスト学習は，SVM と比べて，素性の設定の影響を受けやすいという不安定性をもつことが分かる．例えば，今回の実験では，モデル数が多くなり，認識率の高いモデルと低いモデルが混在すると，特殊性の高い素性がかえって悪影響を及ぼすという結果になっていた)．

（重み付き）多数決法では，$n = 26$ よりも少ないモデル数の段階で F 値が最大値に達し，その後，モデル数が増えるに従って，混合結果の単語認識率（F 値）が低下する傾向にある．一方，機械学習（特に SVM）を用いた混合の場合，$n = 26$ よりも少ないモデル数の段階で F 値が最大値に達し，その後，モデル数が増えても，単語認識率の大幅な低下は起こらない．ここで，「和の再現率順」のモデル選択法で選択された上位 10 数モデルの多くは，「単語正解率順」のモデル選択法でもやはり上位 10 数モデルに含まれていた（ただし，それら 10 数モデルの順序関係は，「和の再現率順」の場合とは異なっていた)．つまり，「和の再現率順」のモデル選択法においても，混合に参加するモデル数が少ない段階では，比較的認識率の高いモデルが多数派であるが，混合に参加するモデル数が増えると，相対的に認識率の低いモデルが多くなるという傾向にあった．このことから，（重み付き）多数決による混合においては，認識率の高いモデルと低いモデルが混在すると，混合結果の性能が認識率の低いモデルに影響されることが分かった．一方，機械学習（特に SVM）を用いた混合では，性能の高いモデルと低いモデルが混在しても，学習の段階でこれに対処することが実現でき，混合結果の性能が認識率の低いモデルに影響されることはなかった(注9)．

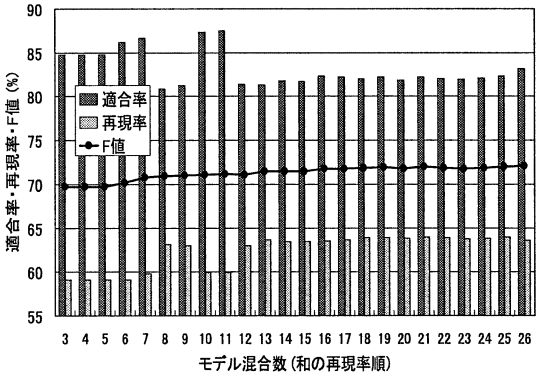
また，機械学習を用いた混合と（重み付き）多数決による混合の性能差は，特に，単語正解精度において大きい．機械学習を用いた混合においては，それぞれの学習手法におけるしきい値を操作することで，高信頼度な認識結果のみを出力することが可能であり，この機能により挿入誤りを抑制し，単語正解精度を改善している．図 4 には，図 1 における SVM を用いた混合の F 値の変化のプロットについて，適合率・再現率の変化を併せてプロットした結果を示しているが，この結果から分かるように，$n = 3 \sim 26$ のどのモデル数の混合においても，一貫して高い適合率（新聞読上げ音声で 96%以上，ニュース音声で 80%以上）を維持しており，高信頼度な認識結果のみを出力していることが分かる．

次に，図 5 では，単独モデルによる認識結果及び複数モデルの出力の混合結果について，単語正解精度を重視し，単語正解精度が最大となる認識結果を比較する．この結果から単語誤り改善率を算出すると，新聞読上げ音声においては，認識率最大の単独モデルに対

(注9)：「単語正解率順」のモデル選択法の場合でも，（重み付き）多数決による混合と機械学習（特に SVM）による混合の間では．同様の比較結果が得られた．



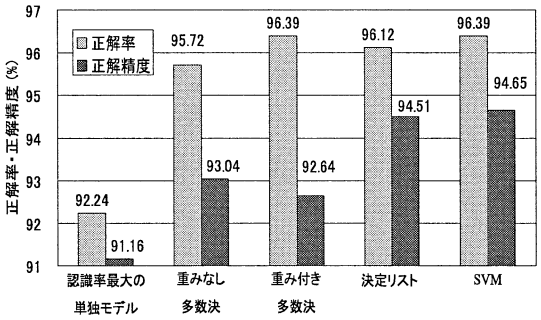
(a) 新聞読上げ音声



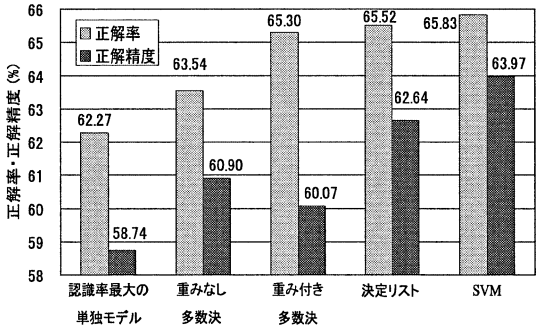
(b) ニュース音声

図 4　和の再現率順に選択した n ($3 \leq n \leq 26$) モデルの出力の混合における適合率・再現率・F 値の変化（SVM による混合）

Fig. 4　Performance of combining outputs of n ($3 \leq n \leq 26$) models. (selected so as to maximizing recall of union, combination by SVM): Changes of precision/recall/F-measures.



(a) 新聞読上げ音声



(b) ニュース音声

図 5　SVM による混合・決定リスト学習による混合・（重み付き）多数決による混合・単独モデルの出力の正解率・正解精度の比較

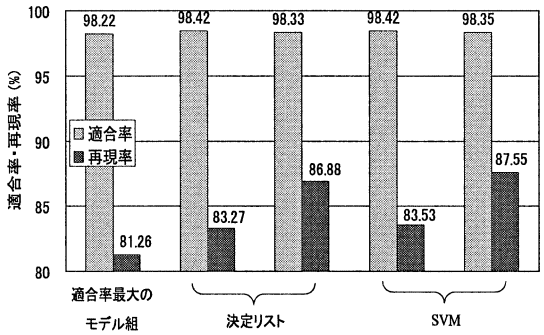
Fig. 5　Comparing word correct/accuracy rates among combination by SVM / by decision list learning / (weighted) majority votes and individual models.

して，SVM による混合により，単語正解率において 53%，単語正解精度において 39%の誤り改善率を達成した．また，（重み付き）多数決との比較では，単語正解率の改善はわずかであるが，単語正解精度において 23%の誤り改善率であった．同様に，ニュース音声では，SVM による混合により，認識率最大の単独モデルに対して，単語正解率において 9%，単語正解精度において 13%の誤り改善率であり，（重み付き）多数決との比較では，単語正解率の改善はわずかであるが，単語正解精度において 8%の誤り改善率であった．

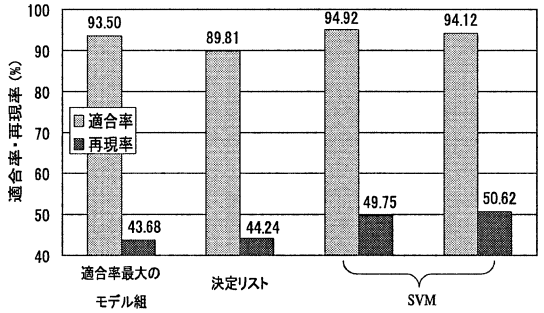
最後に，二つのモデルの出力の共通部分について，適合率が最大となるモデル組を求め [14]，そのモデル組の出力の共通部分の再現率・適合率を，機械学習による複数モデルの混合における再現率・適合率と比較した結果を図 6 に示す．ただし，機械学習を用いた混合においては，混合対象となるモデルの数 n を変化させ，最も性能が良かった結果について，それぞれの学習手法におけるしきい値を操作して，適合率と再現率の両方が比較できるような結果を何通りか示す．この結果から分かるように，SVM を用いた混合においては，適合率最大のモデル組に対して，その適合率をやや上回り，更に再現率については十分な改善が達成できた．したがって，二つのモデルの出力の共通部分が高い信頼度をもつという特性を十分に保持したまま，複数モデルによる認識結果を相補的に利用して，信頼性の高い認識結果を組み合わせることができていることが分かる(注10)．

以上の結果から，機械学習（特に SVM）を用いた混合においては，学習手法におけるしきい値を最適に調整することにより，適合率（認識結果の信頼度に相

(注10)：（重み付き）多数決による混合の適合率は，新聞読上げ音声の場合で，たかだか 95%，ニュース音声の場合で，たかだか 76%であり，適合率・再現率の両方において，機械学習による混合の性能を下回る．



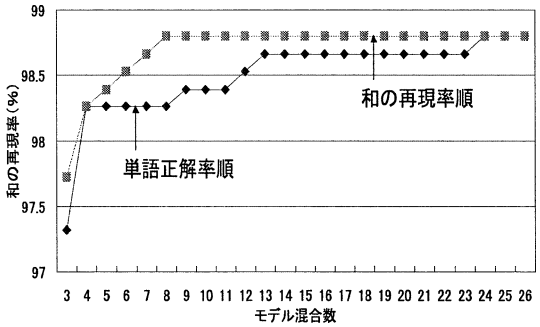
(a) 新聞読上げ音声



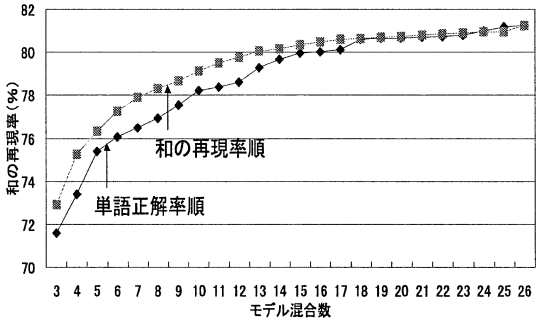
(b) ニュース音声

図 6 SVM による混合・決定リスト学習による混合と 2 モデルの出力の共通部分との適合率・再現率の比較

Fig. 6 Comparing precision/recall among combination by SVM / by decision list learning and agreement between outputs of two models.



(a) 新聞読上げ音声



(b) ニュース音声

図 7 n ($3 \leq n \leq 26$) モデルの出力の和の再現率の推移（和の再現率順と単語正解率順の比較）

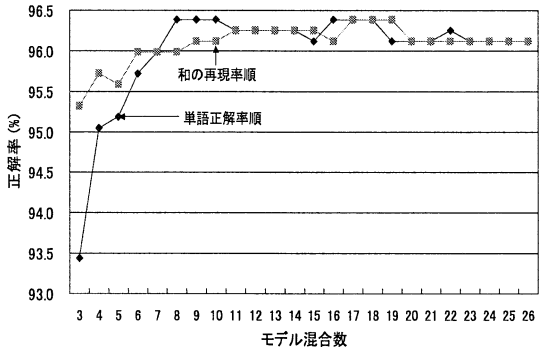
Fig. 7 Comparing recall of union of the outputs of n ($3 \leq n \leq 26$) models. (maximizing recall of union vs. descending order of word correct rates)

当）あるいは単語正解率（=再現率，認識結果の被覆率に相当）のいずれかを優先した混合結果を出力することが容易に実現可能であることが分かる．

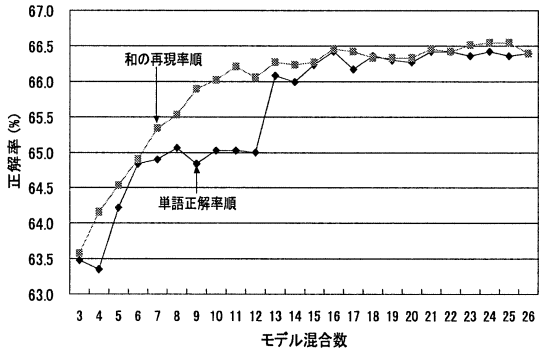
4.3 複数モデルの選択法の比較

SVM を用いた混合について，「単語正解率順」に混合するモデルを選択する手法と，「和の再現率順」に混合するモデルを選択する手法の性能比較を行った．まず，二つのモデル選択法について，混合するモデルを一つずつ増やした場合の和の再現率の推移を，図 7 に示す．この結果から，「単語正解率順」に混合するモデルを選択した場合は，和の再現率の立上りが遅く，比較的類似したモデルが選択されている可能性があるといえる．また，新聞読上げ音声では，「和の再現率順」のモデル選択により，モデル混合数が少ない段階で和の再現率が最大値に達するのに対して，ニュース音声では，ほぼ全モデルを混合するまでは，和の再現率は最大値には達しない．これは，ニュース音声の認識の方がタスクとして難しく，結果として，音声認識モデルの多様性も大きくなっているためであると考えられる．

次に，二つのモデル選択法について，単語正解率の推移をプロットした結果を，図 8 に示す（F 値及び単語正解精度についてもほぼ同様の推移となる）．「和の再現率順」のモデル選択法においては，モデル混合数の少ない立上りの段階において「単語正解率順」のモデル選択法の性能を上回っており，図 7 の和の再現率の推移における両者の差をそのまま反映している．したがって，SVM を用いた混合においては，モデル混合数が少なく，混合されるモデル間の多様性が大きい場合でも，信頼性の高い認識結果を相補的に選択して認識結果を出力することが実現できており，SVM の高い学習能力を実証する結果といえる．逆に，複数モデルによる音声認識結果の混合のタスクにおいて，SVM のような高い学習能力をもつ学習器を適用できる場合には，認識結果の和の再現率がなるべく大きくなるようなモデル選択法が効果的であるといえる．ただし，図 7 の和の再現率の最大値と，図 8 の単語認識率の



(a) 新聞読上げ音声



(b) ニュース音声

図 8　SVM による n ($3 \leq n \leq 26$) モデルの出力の混合における単語正解率（=再現率）の推移（和の再現率順と単語正解率順の比較）

Fig. 8　Comparing word correct rates (= recall) of combining outputs of n ($3 \leq n \leq 26$) models by SVM. (maximizing recall of union vs. descending order of word correct rates)

最大値を比べると，混合結果の単語認識率はまだ上限には達していないことが分かる．特に，認識の困難なニュース音声の場合は，両者の間に約 15%の大きな差がある．したがって，今後は，複数モデルによる認識結果の混合により有効な素性を調査するなどして，混合結果の認識率改善の可能性について検討する必要があるといえる．

また，表 1 には，和の再現率順で選択された上位 10 モデルを，新聞読上げ音声とニュース音声の間で比較した一覧を示す．表中では，太字で示す 7 モデルが，両方の音声データにおいて共通に上位 10 モデルに選ばれており，このことから，個々の音声データ固有の特性を超えて，汎用的に有効なデコーダ・音響モデルの組合せが存在するといえる．

4.4　機械学習の素性の性能比較

SVM について素性の性能の比較を行った．**2.2.1** の素性 i)（各単語を出力したモデルの情報）のみを用いた場合，素性 i) ～ iii)（モデルの情報，品詞，音節数）を用いた場合，及び，i) ～ v) の全素性（モデルの情報，品詞，音節数，音響スコア，言語スコア）を用いた場合について，混合対象となるモデルの数 n を変化させて，F 値の変化をプロットしたものを図 9 に示す．この結果では，品詞，音節数，音響スコア，言語スコアを用いることにより，混合結果の単語認識率は向上するものの，その効果はわずかである．したがって，SVM は素性の設定の違いの影響を受けにくく，利

表 1　和の再現率順で選択された上位 10 モデルの一覧
Table 1　First 10 models selected when maximizing recall of union.

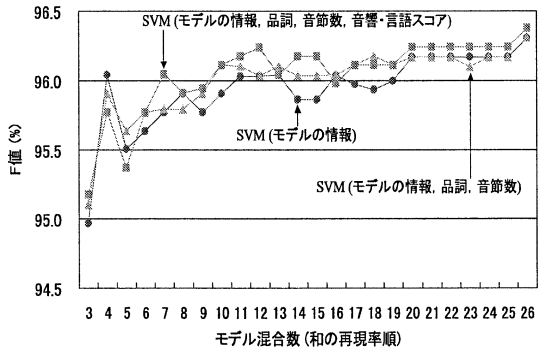
(a) 新聞読上げ音声

デコーダ	音響モデル	順位 新聞	順位 ニュース
Julius	無音あり/トライフォン	1 位	12 位
SPOJUS	**無音あり/継続時間制御/16 kHz/10 ms/全共分散/MFCC-frm**	**2 位**	**2 位**
Julius	無音なし/PTM	3 位	20 位
Julius	**無音あり/音節モデル**	**4 位**	**1 位**
SPOJUS	**無音なし/自己ループ/12 kHz/8 ms/全共分散/MFCC-seg**	**5 位**	**5 位**
Julius	**無音あり/PTM**	**6 位**	**8 位**
Julius	**無音なし/トライフォン**	**7 位**	**3 位**
SPOJUS	**無音あり/継続時間制御/12 kHz/8 ms/全共分散/MFCC-seg**	**8 位**	**6 位**
SPOJUS	**無音あり/継続時間制御/16 kHz/10 ms/全共分散/MFCC-seg**	**9 位**	**10 位**
SPOJUS	無音なし/継続時間制御/12 kHz/8 ms/全共分散/MFCC-seg	10 位	21 位

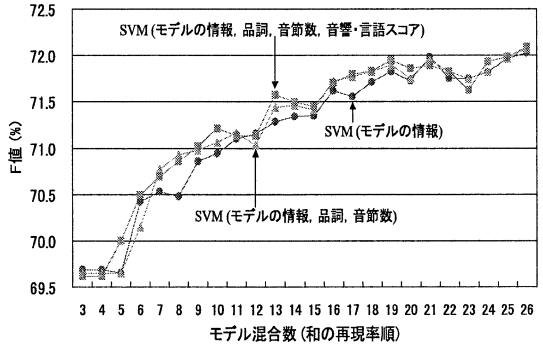
(b) ニュース音声

デコーダ	音響モデル	順位 新聞	順位 ニュース
Julius	**無音あり/音節モデル**	**4 位**	**1 位**
SPOJUS	**無音あり/継続時間制御/16 kHz/10 ms/全共分散/MFCC-frm**	**2 位**	**2 位**
Julius	**無音なし/トライフォン**	**7 位**	**3 位**
SPOJUS	無音なし/継続時間制御/16 kHz/8 ms/全共分散/MFCC-seg	17 位	4 位
SPOJUS	**無音なし/自己ループ/12 kHz/8 ms/全共分散/MFCC-seg**	**5 位**	**5 位**
SPOJUS	**無音あり/継続時間制御/12 kHz/8 ms/全共分散/MFCC-seg**	**8 位**	**6 位**
Julius	無音なし/音節モデル	11 位	7 位
Julius	**無音あり/PTM**	**6 位**	**8 位**
SPOJUS	無音なし/継続時間制御/16 kHz/10 ms/全共分散/MFCC-frm	22 位	9 位
SPOJUS	**無音あり/継続時間制御/16 kHz/10 ms/全共分散/MFCC-seg**	**9 位**	**10 位**

12 kHz/16 kHz：サンプリング周波数，8 ms/10 ms：フレーム周期
MFCC-frm/MFCC-seg：フレーム単位/セグメント単位 MFCC
太字：新聞・ニュースともに 10 位以内



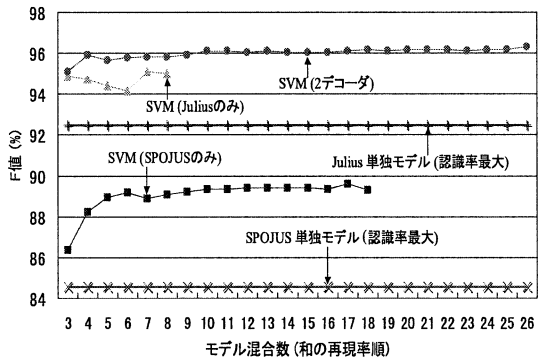
(a) 新聞読上げ音声



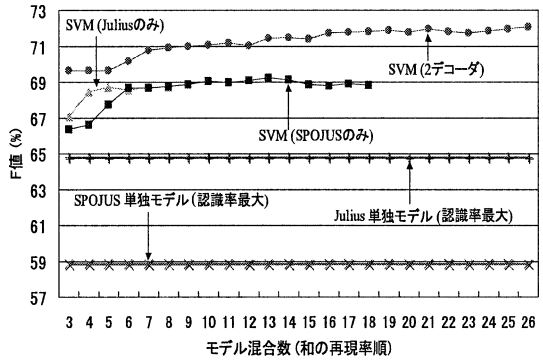
(b) ニュース音声

図 9 和の再現率順に選択した n ($3 \leq n \leq 26$) モデルの出力の混合における素性情報の性能比較（F 値, SVM による混合）

Fig. 9 Performance comparison for features of combining outputs of n ($3 \leq n \leq 26$) models. (selected so as to maximizing recall of union, in F-measures, combination by SVM)



(a) 新聞読上げ音声



(b) ニュース音声

図 10 和の再現率順に選択した n ($3 \leq n \leq 26$) モデルの出力の混合におけるデコーダの組合せの性能比較（SVM による混合）

Fig. 10 Performance comparison for decoder combination in combining outputs of n ($3 \leq n \leq 26$) models. (selected so as to maximizing recall of union, in F-measures, combination by SVM)

用可能な素性が限られている場合でも，効果的な学習を行えていることが分かる．

4.5 デコーダの組合せの性能比較

最後に，複数の大語彙連続音声認識モデルの出力の混合において，異なるデコーダを用いた複数モデルの出力を混合する場合と，同一デコーダを用いた複数モデルの出力を混合する場合の性能の比較を行う．機械学習手法として SVM を用いた場合について，26 モデルすべてに対して，和の再現率順に $n(3 \leq n \leq 26)$ 個選択して複数モデルの出力の混合を行った場合，デコーダが Julius の場合について，8 種類のモデルを和の再現率順に $n(3 \leq n \leq 8)$ 個選択して複数モデルの出力の混合を行った場合，及び，デコーダが SPOJUS の場合について，18 種類のモデルを和の再現率順に $n(3 \leq n \leq 18)$ 個選択して複数モデルの出力の混合を行った場合の混合結果の単語認識率（F 値）の変化を図 10 に示す．ただし，SVM の素性としては，**2.2.1** の i) ～ iii)（モデルの情報，品詞，音節数）を用いた．また，「単独モデル（認識率最大）」としては，図 1 の場合と同様に，それぞれのデコーダで F 値が最大となる単独モデルの値を用いる．

この結果から分かるように，異なるデコーダを用いた複数モデルの出力を混合する場合の性能の方が高くなっている．今回の実験では，デコーダが異なるモデルの方が，認識結果として異なる（正解）単語を出力する傾向があり，複数モデルの出力の混合においても，個々のモデルができるだけ異なる（正解）単語を出力した方（すなわち，和の再現率が高くなる方）が，混合結果の性能が高くなったものと考えられる．また，新聞読上げ音声では，デコーダが Julius の場合，モデル混合数が 3 において，混合結果の F 値がほぼ飽和

表 2　和の再現率順で選択された上位モデル（単一デコーダ，音響モデルの一覧）
Table 2　Models selected first when maximizing recall of union. (one decoder, list of acoustic models)

(a) 新聞読上げ音声

Julius（3 モデル）	無音あり/トライフォン，無音なし/PTM，**無音あり/音節モデル**
SPOJUS（6 モデル）	**無音あり/継続時間制御/16 kHz/10 ms/全共分散/MFCC-seg** 無音なし/自己ループ/12 kHz/8 ms/全共分散/MFCC-seg **無音あり/継続時間制御/12 kHz/8 ms/全共分散/MFCC-seg** 無音あり/自己ループ/16 kHz/8 ms/全共分散/MFCC-seg 無音なし/継続時間制御/16 kHz/10 ms/全共分散/MFCC-seg 無音なし/自己ループ/16 kHz/8 ms/全共分散/MFCC-seg

(b) ニュース音声

Julius（5 モデル）	**無音あり/音節モデル**，無音なし/トライフォン，無音なし/音節モデル，無音あり/PTM，無音あり/モノフォン
SPOJUS（6 モデル）	無音あり/継続時間制御/16 kHz/8 ms/全共分散/MFCC-seg **無音あり/継続時間制御/16 kHz/10 ms/全共分散/MFCC-seg** 無音なし/継続時間制御/12 kHz/8 ms/全共分散/MFCC-seg **無音あり/継続時間制御/12 kHz/8 ms/全共分散/MFCC-seg** 無音なし/継続時間制御/16 kHz/8 ms/全共分散/MFCC-seg 無音あり/自己ループ/16 kHz/10 ms/全共分散/MFCC-seg

太字：新聞・ニュースともに上位のモデル

し，デコーダが SPOJUS の場合は，モデル混合数が 6 においてほぼ飽和している．ニュース音声では，デコーダが Julius の場合，モデル混合数が 5 において，混合結果の F 値がほぼ飽和し，デコーダが SPOJUS の場合は，モデル混合数が 6 においてほぼ飽和している．したがって，同一デコーダを用いる場合でも，これらの少数のモデルを用いるだけで，ほぼ最適に近い性能が得られるといえる．表 2 にこれらのモデルの内訳を和の再現率順に示す．

5. むすび

本論文では，機械学習を用いて複数の大語彙連続音声認識モデルの出力を混合する手法を提案した．機械学習手法として，SVM 及び決定リスト学習を適用し，複数モデルの出力を混合する規則を学習した．この混合規則を用いて，デコーダ，音響モデルの異なる 26 種類の大語彙日本語連続音声認識モデルの出力の混合を行ったところ，認識率最大の単独モデル，及び，（重み付き）多数決を用いた混合の単語認識率を上回る性能が達成できた．本論文では，新聞読上げ音声，及び，ニュース音声を評価音声データとして，提案手法の有効性を評価したが，その他，旅行会話読上げ音声の認識においても，提案手法により単語認識率が改善できることを示している [16]．また，音声入力によるウェブ検索タスクにおいても，提案手法により，音声による検索課題の単語認識率・キーワード認識率・検索精度が改善できている [7]．特に，この結果は，助詞・助動詞といった付属語だけでなく，名詞・動詞といった自立語の認識率が改善できていることを示しており，今後，音声認識結果に対して言語処理を行うといった様々な応用的局面において，提案手法を有効に活用することができると期待される．

謝辞　本研究に協力して頂いた豊橋技術科学大学工学部情報工学系中川研究室の関係者に深く感謝する．また，ニュース音声データベース，ニューステキストデータベースを提供して頂いた NHK 放送技術研究所の関係諸氏に感謝する．なお，本論文の研究の大部分は，筆者らが豊橋技術科学大学工学部情報工学系に在籍中に行ったものである．

文　献

[1] 赤松裕隆，花井建豪，甲斐充彦，峯松信明，中川聖一，“新聞・ニュース文をタスクとした大語彙連続音声認識システムの評価，” 情処学会第 57 回全大，pp.35–36, 1998.

[2] J.G. Fiscus, “A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (ROVER),” Proc. IEEE Workshop on Automatic Speech Recognition and Understanding, pp.347–354, 1997.

[3] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, “JNAS Japanese speech corpus for large vocabulary continuous speech recognition research,” J. Acoust. Soc. Jpn. (E), vol.20, no.3, pp.190–206, 1999.

[4] 河原達也，李　晃伸，小林哲則，武田一哉，峯松信明，嵯峨山茂樹，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ェア（99 年度版），” 音響誌（技術報告），vol.57, no.3,

pp.210–214, 2001.
[5] T. Kemp and T. Schaaf, “Estimating confidence using word lattices,” Proc. 5th Eurospeech, pp.827–830, 1997.
[6] 北岡教英，高橋伸寿，中川聖一，“N-best 線形辞書探索と 1-best 近似木構造辞書探索の併用による大語彙連続音声認識,” 信学技報，SP2003-26, 2003.
[7] 松下雅彦，西崎博光，宇津呂武仁，中川聖一，“音声入力による Web 検索のためのキーワード認識・抽出法の検討,” 情処学研報，2003-SLP-48, pp.21–28, 2003.
[8] 村田真樹，内山将夫，内元清貴，馬　青，井佐原均，“SENSEVAL2J 辞書タスクでの CRL の取り組み — 日本語単語の多義性解消における種々の機械学習手法と素性の比較,” 自然言語処理，vol.10, no.3, pp.116–133, 2003.
[9] 中川聖一，花井建豪，山本一公，峯松信明，“HMM に基づく音声認識のための音節モデルと triphone モデルの比較,” 信学論（D-II），vol.J83-D-II, no.6, pp.1412–1421, June 2000.
[10] 中川聖一，堀部千寿，“音響尤度と言語尤度を用いた音声認識結果の信頼度の算出,” 情処学研報，2001-SLP-36, pp.87–92, 2001.
[11] H. Schwenk and J.-L. Gauvain, “Combining multiple speech recognizers using voting and language model information,” Proc. 6th ICSLP, vol.II, pp.915–918, 2000.
[12] T. Utsuro, Y. Kodama, T. Watanabe, H. Nishizaki, and S. Nakagawa, “Confidence of agreement among multiple LVCSR models and model combination by SVM,” Proc. 28th ICASSP, vol.I, pp.16–19, 2003.
[13] 宇津呂武仁，原田哲志，渡邊友裕，西崎博光，中川聖一，“複数の大語彙連続音声認識モデルの出力の共通部分を用いた信頼度—信頼度を利用した複数モデルの出力の混合,” 信学技報，SP2002-22, 2002.
[14] 宇津呂武仁，西崎博光，小玉康広，中川聖一，“複数の大語彙連続音声認識モデルの出力の共通部分を用いた高信頼度部分の推定,” 信学論（D-II），vol.J86-D-II, no.7, pp.974–987, July 2003.
[15] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, 1995.
[16] 渡邊友裕，山本博史，小窪浩明，菊井玄一郎，西崎博光，小玉康広，宇津呂武仁，中川聖一，“機械学習を用いた複数の大語彙連続音声認識モデルの出力の混合—旅行会話音声における評価,” 2003 春季音講論集，vol.I, pp.209–210, 2003.
[17] F. Wessel, K. Macherey, and H. Ney, “A comparison of word graph and N-best list based confidence measures,” Proc. 6th Eurospeech, pp.315–318, 1999.
[18] D. Yarowsky, “Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French,” Proc. 32nd ACL, pp.88–95, 1994.

（平成 15 年 10 月 22 日受付，16 年 1 月 13 日再受付）

宇津呂武仁

1989 京大・工・電気工学第二卒．1994 同大大学院工学研究科博士課程電気工学第二専攻了．京都大学博士（工学）．同年，奈良先端科学技術大学院大学情報科学研究科助手．1999 ～ 2000 米国ジョンズ・ホプキンス大学計算機科学科客員研究員．2000 豊橋技術科学大学工学部情報工学系講師，2003 京都大学大学院情報学研究科知能情報学専攻講師，現在に至る．自然言語処理，音声言語情報処理の研究に従事．情報処理学会，人工知能学会，日本ソフトウェア科学会，言語処理学会，日本音響学会，ACL 各会員．

小玉　康広

2001 豊橋技科大・工・情報工学卒．2003 同大大学院工学研究科修士課程情報工学専攻了．現在，ソニー（株）に勤務．在学中は音声言語情報処理に関する研究に従事．

渡邊　友裕

2003 豊橋技科大・工・情報工学卒．現在，同大大学院工学研究科修士課程情報工学専攻在学中．音声言語情報処理に関する研究に従事．

西崎　博光　（正員）

1998 豊橋技科大・工・情報工学卒．2000 同大大学院工学研究科修士課程情報工学専攻了．2003 同大学院工学研究科博士後期課程電子・情報工学専攻了．博士（工学）．2003 山梨大学大学院医学工学総合研究部助手，現在に至る．音声言語情報処理に関する研究に従事．情報処理学会，日本音響学会各会員．

中川　聖一　（正員）

1976 京都大学大学院工学研究科博士課程了．同年京都大学工学部情報工学科助手．1980 豊橋技術科学大学工学部情報工学系講師．1990 同教授．1985 ～ 1986 カーネギーメロン大学客員研究員．音声言語情報処理，自然言語処理，人工知能の研究に従事．工学博士．1977 本会論文賞，1998 年度 IETE 最優秀論文賞，2001 本会論文賞受賞．著書「確率モデルによる音声認識」（本会編），「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），「パターン情報処理」（丸善）など．