

連続音声認識コンソーシアムの活動報告 及び最終版ソフトウェアの概要

河原 達也[†] 武田 一哉^{††} 伊藤 克亘^{††} 李 晃伸^{†††} 鹿野 清宏^{†††}
山田 篤^{††††}

[†] 京都大学 学術情報メディアセンター

^{††} 名古屋大学 情報科学研究科

^{†††} 奈良先端科学技術大学院大学 情報科学研究科

^{††††} (財) 京都高度技術研究所

E-mail: †kawahara@i.kyoto-u.ac.jp

あらかし 連続音声認識コンソーシアム (CSRC) は、IPA プロジェクトで開発された「日本語ディクテーション基本ソフトウェア」の維持・発展をめざして、情報処理学会 音声言語情報処理研究会のもとで 2000 年度から 2002 年度まで (2003 年 9 月まで) 活動を行ってきた。本稿では、この活動の報告を行うとともに、このたび編集した最終版ソフトウェアの概要を述べる。本プロジェクトでは、大語彙連続音声認識エンジン Julius の機能拡張と Windows SAPI 対応を行うとともに、非常に大規模なデータベースを用いた高精度な音響モデル・言語モデルの構築を行った。また音響モデルについては、多様な話者層 (高齢者・小児) や入力環境 (電話・車内環境など) に対応したモデルを整備した。キーワード 連続音声認識, コンソーシアム, ソフトウェア

Overview of Activities and Software of Continuous Speech Recognition Consortium

Tatsuya KAWAHARA[†], Kazuya TAKEDA^{††}, Katunobu ITOU^{††}, Akinobu LEE^{†††},
Kiyohiro SHIKANO^{††††}, and Atsushi YAMADA^{††††}

[†] Kyoto University, School of Informatics

^{††} Nagoya University, School of Information Science

^{†††} Nara Institute of Science and Technology, School of Information Science

^{††††} ASTEM, Kyoto

E-mail: †kawahara@i.kyoto-u.ac.jp

Abstract Continuous Speech Recognition Consortium (CSRC) was founded under IPSJ SIG-SLP for further enhancement of Japanese Dictation Toolkit that had been developed by the IPA project. An overview of its activities and final version of the developed software is given in this report. The LVCSR (large vocabulary continuous speech recognition) engine Julius has been improved both in functionality and stability, and ported to Windows in compliance with SAPI (Speech API). A set of acoustic and language models are trained using very large-scale databases. We also set up a variety of acoustic models to cover wider user generations and speech-input environments.

本ソフトウェアの申込み先 <http://www.lang.astem.or.jp/CSRC/>
[mailto: csrc@astem.or.jp](mailto:csrc@astem.or.jp)

1. 活動の経緯

日本の情報処理技術において、現在、日本語音声認識技術が注目され、実用化も視野に入れた研究・開発が活発に行われている。しかしながら、基本性能・頑健性、そしてユーザインタフェースにおいて、一層の改善を必要とするのが実情である。個別要素技術の研究とシステムの開発をバランスよく推進するためには、データベースだけでなくモデルやプログラムを含めたプラットフォームを整備することが必要である。また、これらがソースコードを含めてオープンになっていることも重要である。

そこで我々は平成9年度から3年間にわたって、情報処理振興事業協会 (IPA) の「独創的先進の情報技術に係わる研究開発」の受託事業として、「日本語ディクテーション基本ソフトウェア」[1][2][3]の開発を進めてきた。この成果は、標準的な日本語音響モデル、言語モデル、大語彙連続音声認識エンジン Julius、及び種々のツールから構成され、フリーソフトウェアとして公開し、多数の研究機関でベースライン・リファレンスとして利用されている。^(注1)

平成12年10月には、本ソフトウェアの一層の拡充・発展とともに、音声認識を用いたアプリケーション開発の促進を目指して、連続音声認識コンソーシアム (CSRC) が情報処理学会 SLP 研究会のもとで発足し、約70の企業・大学の参加を得て、平成15年9月まで3年間にわたり活動を行った。

2001年1月に設立総会を開催して以来、ASTEMに事務局において、毎年ほぼ以下のような事業を行ってきた。

- [1月] 総会... 当年度の事業・予算・開発計画の策定
- [7月] セミナー... ベータ版の配布と説明・デモ
- [8月] 講習会... 音声認識の講義とソフトウェアを用いた実習
- [9月] ソフトウェア正式版配布及び事業・決算報告

いずれのイベントも盛況であり、特に8月に4日間にわたって山形大学で行ってきた講習会には、定員いっぱいの70名ほどの参加者があった。

会費等による活動資金は、上記のイベント開催や CD-ROM 作成以外に、大規模な商用データベースを購入して高精度な音響モデルを構築したり、学生アルバイトを継続的に雇用してプログラムを開発・維持するのに使われた。また、名古屋大学の COE(CIAIR) プロジェクトと連携してさまざまな実環境向きのモデルを構築したり、オープンソースの音声認識エンジン Julius に Speech API (SAPI) を実装するなど、かなり大規模なソフトウェア開発を行った。これらのソフトウェアは原則としてすべてオープンソースで、(無保証ではあるが) 商用を含めた利用が可能である。

このように当初掲げた以下の目的はおおむね達成できたものと考えられる。

- (1) 基礎研究の共有基盤の維持
- (2) 高精度化と実環境への対応

(3) 産業界や他分野のための利便性の改善

この3年間にわたり、毎年ソフトウェアを編集し、CD-ROM として頒布してきた [4][5][6] が、このたびそれらを総集し、一部を更新した最終版を作成したので、その概要を以下に紹介する。

2. 音響モデル

IPA「日本語ディクテーションソフトウェア」では、日本音響学会の新聞記事読み上げ音声コーパス [7] で学習した音響モデルを提供していた。コンソーシアムでは、ATR の多数話者音声データベース [8] を利用することにより、より高精度なモデルを構築した。

また音響モデルは、話者層や入力環境が大きく変わると大幅な性能低下を引き起こすので、それらに応じて適切なものを用いる必要がある [9]。そこで、高齢者や子供などの話者層、電話や自動車内などの入力環境のためのモデルを構築した。

いずれも、各音素3状態の対角共分散の混合連続分布 HMM に基づいており、HTK フォーマットである。また、音素体系・表記、及び音響分析や特徴量も、特記していない限り、IPA モデル [2] と同一である。ただしコンソーシアムで作成したのは、原則としてすべて性別非依存 (GID) モデルである。

さらに、実環境において高い性能を得るためには、適応を行うことが有効であるので、できる限り MLLR 適応が可能なモデルを用意した。

2.1 高精度成人モデル (CSRC モデル)

成人音響モデルの学習には、日本音響学会の音素バランス文からなる研究用連続音声データベース (ASJ-PB) 及び新聞記事読み上げ音声コーパス (ASJ-JNAS) に加えて、ATR の多数話者音声データベース音素バランス文セット (ATR/BLA) を用いた。

学習データの概要を表1に示す。このデータ量は、IPA モデルに比べて話者数で15倍、学習量で3.5倍の規模を持つ。ただし ATR/BLA コーパスでは女性のデータが男性に比べて2倍近く多い。

作成した音響モデルの一覧を表2に示す。種々の混合分布数からなる音素環境独立 (monophone) モデル、状態共有 triphone モデル、PTM triphone モデル [10] を作成した。最も高精度なものでは、30万以上のガウス分布を持ち、IPA モデルの6倍以上の規模である。PTM モデルでは、混合分布のコードブックは monophone の状態 (129個) 毎に用意し、triphone の共有状態 (3000個) 毎に異なる重みを推定する。

音響モデルの評価を、読み上げ音声と対話音声を用いて行った。

読み上げ音声に対する認識精度を表3に示す。評価データは新聞記事読み上げコーパス (JNAS) から選択された IPA-98-TestSet である。同一の複雑さのモデルで比べると、学習データ量や話者数の増強による認識精度の改善は見られない。しかし、IPA モデルがこのパラメータ数ではほぼ飽和していたのに対して、大規模な ATR/BLA コーパスを用いた CSRC モデルではパラメータ数の増加につれて認識精度も着実に上昇してい

(注1): 「日本語ディクテーションソフトウェア」最終版は、文献 [1] の付録 CD-ROM として取られている。

表1 CSRC 成人モデルの学習データの概要

		話者数	発話数	時間
ATR/BLA	男性	1379	42057	59h
	女性	2390	70483	103h
	計	3769	112540	162h
ASJ/JNAS +	男性	179	28127	47h
	女性	182	28681	51h
	計	361	56808	98h
全体	男性	1558	70184	106h
	女性	2572	99104	154h
	合計	4130	169348	260h

表2 CSRC 成人モデルの一覧

	状態数	混合分布数
monophone	129	16, 32, 64, 128, 256
triphone 2000	2000	16, 32
triphone 5000	5000	16, 32, 64
PTM triphone	3000/129	64, 128, 256

表3 JNAS 読み上げ音声に対する単語認識精度 (%)

	PTM	PTM	PTM	triphone
状態数	3000	3000	3000	2000
コードブック	129	129	129	-
混合数	64	128	256	16
IPA モデル	92.7	NA	NA	93.6
CSRC モデル	92.1	93.0	93.6	93.3

表4 ATR 対話音声に対する単語認識精度 (%)

	PTM	PTM	PTM	triphone
状態数	3000	3000	3000	2000
コードブック	129	129	129	-
混合数	64	128	256	16
IPA モデル	82.1	NA	NA	80.5
CSRC モデル	83.1	84.5	84.8	82.4

る。IPA モデルが主に新聞記事読み上げ文から学習されているのに対して、ATR/BLA が音素バランス文であることを考慮すると、CSRC モデルの方が汎用性が高いと考えられる。

対話音声に対する認識精度を表4に示す。ATR 自然発話音声データベース(旅行会話タスク)から対面対話音声(ATR/SDB)212文、通訳対話音声(ATR/SLDB)108文を用いた。通訳対話は対面対話と読み上げの中間の性質を持つと考えられる。ここでは、男性のみのサンプルを用いている。言語モデルは単語 3-gram である[11]。CSRC モデルが全般に高い認識精度を得ており、効果が確認された。

特に 128 混合や 256 混合の PTM モデルは、高い認識精度と実時間に近い処理効率の両立を実現するものである。

2.2 高齢者音声モデル

高齢者に対する音声インタフェースの開発は今後重要になると考えられるが、高齢者の発声の音響的特徴は若年層と異なるため、成人のモデルでは十分に対応できない。そこで、NEDO の委託事業「シニア支援システムの開発」プロジェクトで構築

されたシニア音声認識用大規模データベース[12]を用いて、高齢者向けの音響モデルを作成した。本データベースは、60~90歳の301名の被験者が各音素バランス100文、新聞記事文100文を読み上げたものである。

性別非依存(GID)と性別依存(GD)の両方の monophone と PTM triphone (2000 状態 64 混合)がある。

2.3 小児音声モデル

小児は声道長が短いため、成人用の音響モデルで音声認識するのは困難である。そこで、名古屋大学の CIAIR プロジェクト[13]で構築が進められている子供の声データベースを用いて、小児用の音響モデルを作成した。小学生400名が各100単語程度を発話したデータを使用している。ただし、十分な学習サンプルを確保できない音素コンテキストがいくつか存在するため、成人女性モデルから MAP 適応学習を行うことにより作成した。作成したモデルは、monophone(16 混合)と PTM triphone(64 混合)である。

2.4 電話用音響モデル

IVR などの電話を介した音声認識アプリケーションの需要が大きいが、帯域制限や回線上のノイズがあるために、電話音声用の音響モデルが必要である。そこで、京都大学で収集された電話音声データベースを使用して、音響モデルを作成した。このデータベースでは、517名が音素バランス文各50文を発声しており、固定電話と携帯電話の割合は約半数である。収集に用いられた電話インタフェースボードは、Dialogic 社の D/41E-PCI である。作成したモデルは、triphone (2000 状態 16 混合)と PTM triphone (32 混合)である。

2.5 自動車内用音響モデル

カーナビなどの自動車内での音声認識インターフェースの需要が大きいが、入力環境が前述の大規模音声コーパスと大きく異なるので、専用の音響モデルが必要となる。

そこで、名古屋大学の CIAIR プロジェクト[13]で構築が進められている車内対話音声データベースを利用して、自動車内用の音響モデルを作成した。音素バランス文を中心に計700名による約22000文の音声データを使用している。アイトリング中と市街地走行中の両方の状況で、バイザー位置に設置したマイクロフォンから入力されている。SN比の平均は20dBである。250Hz以上に帯域制限した分析条件で、作成したモデルは triphone (2000 状態 32 混合)である。

2.6 話者・環境適応のサポート

このように、ある程度多様な音響モデルを用意したが、実際に使用する際には、適切なモデルを選択した上で、さらに個々のユーザや周囲の環境・入力チャンネルに適応することが望ましい。

利用話者・使用環境への事前適応の手法として、MLLR(Maximum Likelihood Linear Regression)法が近年最も広く使われており、HTK[14]のパッケージにも含まれている。

そこで、ほとんどの音響モデル(HTKフォーマット)にHTKのMLLR適応が適用できるように必要な回帰情報を埋め込んでいる。

2.7 ハンズフリーツールキット

接話型マイクを用いないようなアプリケーションにおいては、マイクロフォンアレー (2チャンネルの場合も含む) が有望な入力装置と考えられる。

そこで、マイクロフォンアレーのための信号処理ツールキットを用意した。遅延和アレーと適応型アレー (AMNOR) を実装している。また、TSP を用いたインパルス応答の測定プログラムも含まれている。

3. 言語モデル

IPA「日本語ディクテーションソフトウェア」では、毎日新聞記事データ (1991~1997年分) で学習した単語 N-gram モデルを提供していた。コンソーシアムでは、この新聞記事モデルを更新するとともに、より日常的な言葉を指向したモデルを作成した。

いずれも、形態素解析に Chasen を用いている。N-gram モデルのフォーマットは原則として Julius 用のバイナリ形式であるが、一部のみ ARPA 形式のものもある。

3.1 新聞記事モデルの更新

毎日新聞記事データ 1991年~2002年の12年分のテキスト (3.5億形態素) を用いて、言語モデルを構築した。このテキストから高頻度語を選定して、6万語彙 (60K) の単語辞書を作成した。そして、Julius 用に前向き 2-gram と後向き 3-gram を学習した。

3.2 Web 上テキストから学習したモデル

新聞記事データよりも、World Wide Web の方がより大規模なテキストを収集することができる。また、Web ページの方がより日常的な言葉や話し言葉が含まれている。

そこで、テキストサイズが約 27 億形態素のデータを収集し、言語モデルを構築した。語彙サイズは 6 万 (60K) である。

3.3 言語モデル作成ツール

- N-gram 学習ツールキット Palmkit [15]

CMU-Cambridge SLM Toolkit とコマンドレベルではば互換で、さらに、クラス N-gram をサポートし、また異なるタイプのモデルや、異なる長さの N-gram を組合わせて利用することもできる。^(注2)

- N-gram モデル融合ツール [16]

複数の N-gram モデルを直接融合する。相補的バックオフを適用することで、元コーパスを必要とせず、任意の融合重みや語彙サイズを設定することができる。

- Web からの N-gram モデル自動作成ツール [17]

指定されたタスクやドメインのキーワードを基に、Web ページを収集し、テキスト処理や文選択を行って、N-gram モデルを作成する。

- N-gram モデル圧縮ツール [18]

エントロピーに基づいて、N-gram エントリを削減する。

4. 認識エンジン Julius

大語彙連続音声認識エンジン Julius [19] [20] ^(注3) については、引き続き機能の拡張と安定性の強化を行うとともに、コンソーシアムではネットワーク文法を扱えるパーザ Julian を統合した。さらに、これらの利便性の向上を図るために、Windows 上への移植と SAPI (Speech API) の実装を行ってきた。

4.1 記述文法用認識エンジン (Julian)

IPA「日本語ディクテーション基本ソフトウェア」の Julius では、言語モデルとして単語 N-gram モデルしか扱えなかった。しかし、音声認識の比較的単純なアプリケーションでは記述文法を用いる場合が多い。そこで、ネットワーク文法のための認識エンジン Julian [21] を統合した。

Julian では単語カテゴリという概念を導入しており、文法ファイルでは単語カテゴリ (非終端記号) を用いて BNF 記法で書き換え規則を記述し、語彙ファイルで各カテゴリに属する単語を記述する。BNF 記法では文脈自由文法を記述できるが、認識時には効率化のため決定性有限状態オートマトン (DFSA) を使用するため、文法はこれにコンパイルできるクラス (左再帰を許さない) に制限される。ただし、実際には大半のタスクに適用可能である。

コンパイルや文法チェックのためのツール、さらに SAPI の XML 形式への (半自動) 変換スクリプトもパッケージに含まれている。

4.2 クラス N-gram のサポート

統計的言語モデルにおいて学習データのスパースネスに対処するために、クラス N-gram がしばしば用いられる。特に人名や商品名などの固有名詞のモデル化において有効である。クラス N-gram のクラスを文法の単語カテゴリと対応づけることにより、単語 N-gram と記述文法の間接形として捉えることもできる。また、発音の変形のモデル化もクラス N-gram の枠組みで実装することができる。

そこで、Julius においてクラス N-gram のサポートを行った。言語モデルファイルは通常の単語 N-gram と同様であるが、クラス内の単語生起確率を単語辞書で指定する。Palmkit で生成されるクラス N-gram と若干フォーマットが異なるため、変換スクリプトを用意している。

4.3 複数の文法のサポート

文法を用いた認識においては、サブタスク毎に文法を用意しておき、文脈情報などから適当なものに絞り込んだり、切り替えたりできると、認識精度・処理効率の両面で効果的である。例えば、対話システムではプロンプト毎に、フォームフィリングでは項目毎に、ブラウザではページ毎に、受理する文法を切り替えることが考えられる。

そのため、同時に複数の文法で認識を行えるように認識エンジンを拡張し、また動作中にアクティブな文法を切り替えられるようにした。ただし、この実装は、Unix 版と Windows 版で異なる。Unix 版では、アクティブな文法はクライアントから

(注2) : <http://palmkit.sourceforge.net>

(注3) : <http://julius.sourceforge.jp>

専用の API で指定し、サーバによる認識は、単一のデコーディングでアクティブな文法ノードのみを展開する方式となっている。これに対して Windows 版では、文法の指定は SAPI に準拠しており、アクティブな文法毎に (複数の) デコーディングを行い、尤度の最も高いものを出力する。なお、(ディクテーション用に) N-gram も記述文法と同時にアクティブにできるが、読み込める N-gram は高々 1 個である。

4.4 マルチパス音響モデルのサポート

音響モデルについても、複数のモデルを読み込んで、入力に応じて動的に選択できると、多様な話者層や発声環境の変化に対応できると考えられる。例えば、成人モデルと小児音声モデルを並列して用いたり、種々の SN 比に応じた雑音適応モデルを用意することが考えられる。

このような複数の音響モデルを、HMM のトポロジー上でマルチパスとして (事前に) 結合したモデルを扱えるようにした。ただし、この機能は Unix 版のみの実装で、また処理速度上の理由から別のパッケージになっている。

4.5 雑音対策

雑音下での音声認識のために、スペクトルサブトラクションを実装した。また、リアルタイムでの認識のために CMN パラメータを保存できる機能も実現した。なおデフォルトでは、前の発話のケプストラム平均を用いている。これらの機能も Unix 版のみの実装である。

4.6 認識信頼度の算出

音声対話システムにおいてユーザへの確認を制御したり、音声メディアの自動書き起こしに基づいた検索や要約において取捨選択するための指標として、認識結果の単語毎に信頼度が付与されていることが望ましい。

そこで、Julius においても認識信頼度を算出する機能を追加した。これは、事後確率と同様に 0~1 の値をとるものである。一般的に認識信頼度は、単語グラフをいったん出力してから事後的に計算することが多い。しかし、Julius で採用しているスタックデコーディングサーチでは単語グラフを求めることが容易でないため、

(1) N-best 単語列候補から求める方法 [22]

(2) スタックデコーディング途上で計算する方法 [?]

の 2 通りを実装した。(1) の方が簡便であるが、多くの候補に対して信頼度 1 になってしまう。(2) は近似計算であるが、すべての候補について実時間に計算できる。ただし、Windows 版では (1) のみを実装している。

4.7 Windows への移植と SAPI の実装

音声認識を利用したアプリケーションの開発やマルチモーダルインタフェースに適用するには、標準的な API を提供することが重要である。そのため、Windows への移植を行うとともに、マイクロソフト社が策定した Speech API (SAPI-5) ^(注 4) の Julius への実装を進めてきた [24]。

Windows XP では、SAPI が標準に含まれているので、Speech SDK をインストールしなくても Julius が動作する。コントロー

ルパネルの「音声認識」のプロパティから、エンジン自体の指定やマイクの設定、そして音響モデル・言語モデル・デコーディングオプションの指定を行うことができる。

SAPI では標準の文法が XML 形式であるので、Julian 用の外部形式・内部表現との相互変換を行う。また、MS-IME を用いて読み付与を行うこともできる。複数の文法 (1 つの N-gram を含む) を扱う場合は、各文法毎にインスタンスを生成し、並列・独立に認識処理が行われる。

現在、Speech SDK 5.1 付属のアプリケーションや MS-IME2002 の音声入力パッド、Office XP など動作確認を行っている。また、.NET Speech SDK 1.0β に含まれている SALT (Speech Application Language Tags) ^(注 5) に対しても、動作確認ができています。

ただし、この Windows SAPI 版 Julius では、Unix 版に比べて、オプションや機能が一部制限される。

またこれとは別に、名古屋大学で SAPI を介さない DLL 版 (Windows, MacOS X で動作) も作成されており、このパッケージも収めている。 ^(注 6)

5. おわりに

本ソフトウェアは、IPA「ディクテーション基本ソフトウェア」と同様に、各モジュールのフォーマットとインタフェースに一般性があり、またソースコードも公開されているので、汎用性と拡張性に富んでいる。今回さらに、Windows に移植し、SAPI 対応になったことにより、アプリケーション開発の利便性が向上した。また、種々の話者や環境に対する音響モデルの整備を行ったことで、やはり多様なアプリケーションへの適用の可能性が広がった。

謝辞：このような活動及びソフトウェアの開発は、コンソーシアムという活動基盤がなければできなかったものと考えられます。ここに、設立・運営に対して協力を頂きました SLP 研究会及び情報処理学会の関係各位、そして活動に対して支援を頂きました会員各位に深い感謝の意を表します。

(注 5) : <http://www.saltforum.org>

(注 6) : <http://www.itakura.nuee.nagoya-u.ac.jp/people/banno/julius.html>

(注 4) : <http://www.microsoft.com/speech>

CSRC 実行委員リスト

代表：河原達也（京大）...2002 年 3 月より

鹿野清宏（奈良先端大）...2002 年 3 月まで

幹事：武田一哉（名大）

伊藤克亘（名大）

李 晃伸（奈良先端大）

山田 篤（ASTEM）

委員：伊藤彰則（東北大）

宇津呂武仁（京大）

峯松信明（東大）

山本幹雄（筑波大）

小林哲則（早大）

嵯峨山茂樹（東大）

岩野公司（東工大）

坂野秀樹（和歌山大）

北岡教英（豊橋技科大）

山田武志（筑波大）

西浦敬信（和歌山大）

西田昌史（千葉大）

三村正人（ASTEM）

文 献

- [1] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄. 音声認識システム. オーム社, 2001.
- [2] 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) の性能評価. 情報処理学会研究報告, 2000-SLP-31-2, 2000.
- [3] T.Kawahara, A.Lee, T.Kobayashi, K.Takeda, N.Minematsu, S.Sagayama, K.Itou, A.Ito, M.Yamamoto, A.Yamada, T.Utsuro, and K.Shikano. Free software toolkit for Japanese large vocabulary continuous speech recognition. In *Proc. ICSLP*, Vol. 4, pp. 476-479, 2000.
- [4] 河原達也, 住吉貴志, 李晃伸, 武田一哉, 三村正人, 伊藤彰則, 伊藤克亘, 鹿野清宏. 連続音声認識コンソーシアム 2000 年度版ソフトウェアの概要と評価. 情報処理学会研究報告, 2001-SLP-38-6, 2001.
- [5] 河原達也, 住吉貴志, 李晃伸, 坂野秀樹, 武田一哉, 三村正人, 山田武志, 西浦敬信, 伊藤克亘, 伊藤彰則, 鹿野清宏. 連続音声認識コンソーシアム 2001 年度版ソフトウェアの概要. 情報処理学会研究報告, 2002-SLP-43-3, 2002.
- [6] 河原達也, 住吉貴志, 李晃伸, 坂野秀樹, 武田一哉, 三村正人, 伊藤克亘, 伊藤彰則, 鹿野清宏. 連続音声認識コンソーシアム 2002 年度版ソフトウェアの概要. 情報処理学会研究報告, 2003-SLP-48-1, 2003.
- [7] 板橋秀一, 山本幹雄, 竹沢寿幸, 小林哲則. 日本音響学会新聞記事読み上げ音声コーパスの構築. 音響誌, 3-P-22, 秋季 1997.
- [8] 奥田浩三, 松井知子, 内藤正樹, 匂坂芳典, 中村哲. 大規模日本語音声データベースの構築と評価. 音響誌, Vol. 58, No. 9, pp. 569-578, 2002.
- [9] 河原達也. ここまできた音声認識技術. 情報処理, Vol. 41, No. 4, pp. 436-439, 2000.
- [10] 李晃伸, 河原達也, 武田一哉, 鹿野清宏. Phonetic Tied-Mixture モデルを用いた大語彙連続音声認識. 電子情報通信学会論文誌, Vol. J83-DII, No. 12, pp. 2517-2525, 2000.
- [11] 三村正人, 河原達也. ディクテーションと対話音声認識における音響モデルの差異. 日本音響学会研究発表会講演論文集, 2-8-4, 春季 2000.
- [12] 馬場朗, 芳澤伸一, 山田実一, 李晃伸, 鹿野清宏. 高齢者向け音響モデルによる大語彙連続音声認識の評価. 情報処理学会研究報告, 2001-SLP-35-3, 2001.
- [13] 武田一哉, 板倉文忠. 文部省 COE プログラム統合音響情報研究拠点 (CIAIR). 音響誌, Vol. 56, No. 11, pp. 748-751, 2000.
- [14] S.Young, J.Jansen, and J.Odell D.Ollason P.Woodland. *The HTK BOOK*, 1995.
- [15] 伊藤彰則, 好田正紀. 単語およびクラス n-gram 作成のためのツールキット. 情報処理学会研究報告, 2000-SLP-34-32, 2000.
- [16] 長友健太郎, 西村竜一, 小松久美子, 黒田由香, 李晃伸, 猿渡洋, 鹿野清宏. 相補的バックオフを用いた言語モデル融合ツールの構築. 情報処理学会研究報告, 2001-SLP-35-9, 2001.
- [17] 西村竜一, 長友健太郎, 小松久美子, 黒田由香, 李晃伸, 猿渡洋, 鹿野清宏. Web からの音声認識用言語モデル自動生成ツールの開発. 情報処理学会研究報告, 2001-SLP-35-8, 2001.
- [18] 踊堂憲道, 鹿野清宏, 中村哲. 情報量に基づく trigram パラメータの逐次の削減手法. 情報処理学会研究報告, 98-SLP-22-17, 1998.
- [19] 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-DII, No. 1, pp. 1-9, 1999.
- [20] A.Lee, T.Kawahara, and K.Shikano. Julius - an open source real-time large vocabulary recognition engine. In *Proc. EUROSPEECH*, pp. 1691-1694, 2001.
- [21] 李晃伸, 河原達也, 堂下修司. 文法カテゴリ制約を用いた A* 探索に基づく大語彙連続音声認識パーザ. 情報処理学会論文誌, Vol. 40, No. 4, pp. 1374-1382, 1999.
- [22] 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, 2002.
- [23] 李晃伸, 鹿野清宏, 河原達也. 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. 情報処理学会研究報告, 2003-SLP-49-48, 2003.
- [24] 住吉貴志, 李晃伸, 河原達也. 音声認識エンジン Julius/Julian の API 実装. 情報処理学会研究報告, 2001-SLP-37-16, 2001.