

ニュース音声認識における言語モデルの検討

高木 一幸 小黒 玲 橋本 顕示 尾関 和彦

電気通信大学 情報工学科
〒182 東京都調布市調布ヶ丘 1-5-1
e-mail:takagi@cs.uec.ac.jp

あらまし ニュース音声は読み上げ音声に比べ音声認識が困難な題材であるが、高齢者・聴覚障害者のための字幕付与サービスなどへの応用が期待されている。今年になって、NHK ニュース放送音声のデータベースが構築され、これを対象とした音声認識タスクが研究されている。大語彙音声認識には精度の良い言語モデルが不可欠であるが、本研究では、(1) 附属語連鎖・複合名詞を連語として登録し単語として扱う連語モデル、(2) 名詞語彙をタスクに合わせて入れ換える名詞フラットモデル、(3) 品詞に基づいて語彙を分類した品詞クラスモデル、という3種類の N-gram 言語モデルを作成した。その結果、(1) についてはパープレキシティの改善が見られた。これらの言語モデルをニュース音声認識に適用した結果について報告する。

キーワード ニュース音声、大語彙連続音声認識、単語関連率、複合名詞、品詞クラス

Performance evaluation of language models for broadcast news speech recognition

Kazuyuki TAKAGI Rei OGURO Kenji HASHIMOTO Kazuhiko OZEKI

The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo, 182, Japan
e-mail:takagi@cs.uec.ac.jp

Abstract The focus of LVCSR research has shifted recently to more real world data, such as TV broadcast news speech, recognition of which aims at variety of applications such as the subtitling of TV program. Language modeling has a significant effect on recognition performance of LVCSR system, as well as the acoustic modeling. In this report, we evaluated the performance of N-gram language models for NHK TV news speech recognition task. Three types of models were created, i.e., (1) word phrase models of function words and compound nouns, (2) language model with uniform noun unigram, (3) POS-based class language model, among of which word phrase model achieved improvement on language perplexity.

key words broadcast news speech, LVCSR, word association ratio, compound noun, POS class

1 はじめに

近年の音声言語処理および計算処理技術の進歩により、従来に比べてより自然な発話や大語彙連続音声認識などに対する研究が盛んに行なわれるようになってきた。英語の大語彙連続音声認識に関しては、ARPAにおいてWall Street Journal コーパス [1] を用いた研究が各研究機関で進められ、着実に成果をあげている。日本でも、新聞記事のテキスト読み上げ音声のデータベースが構築され、これを対象とした大語彙音声認識の研究が行なわれている [2, 3]。

一方で、ニュース音声のように、読み上げ音声とは異なって、自然な発声により近い音声を対象とした大語彙音声認識についての研究は、従来ほとんど行なわれていなかったが、1995年APRAではニュース音声を対象としたプロジェクト (HUB4) [4] が発足し、ラジオおよびテレビ放送されたニュース音声認識の研究が始まっている。ニュース音声を読み上げ音声に比べ音声認識が困難な題材であるが、高齢者・聴覚障害者のための字幕付与サービス、放送素材へのインデックス付与などの広い応用が期待されている。最近になって、スタジオで収録したニュース音声のデータベースが整備されつつあり [5]、これを用いて字幕の自動生成のためのNHKテレビニュース音声の認識タスクが研究されている [6, 7]。

ニュース記事に使われる語彙は数万から十数万語にのぼり、分野も政治、経済、科学、文芸、スポーツなど幅広い。また、一文の長さも長いものが多い。このため、ニュース音声の認識においては、精度良い言語モデルの利用が不可欠である。

本研究では、NHKニュース番組の原稿テキストから、3種類のN-gram言語モデルを作成し、ニュース音声を対象とする大語彙音声認識に適用した。まず、高頻度の附属語連鎖、関連率の高い複合名詞などを加え、単語より長い認識単位を語彙に加えることによる、言語モデルの性能に与える影響を検討した。次に、学習済みの言語モデルにおいて、名詞語彙項目を記事に合わせて入れ換えることを目的とした言語モデルを試作した。さらに、名詞以外についても品詞・活用型で単語を分類したクラス言語モデルを作成して、ニュース音声認識に適用した結果を報告する。

2 テキストデータベース

NHKニュース原稿データベースの1992年8月から1996年6月までのうち38カ月分の原稿をテキストデータベースとして使用した。このうち、言語モデル作成には96年5月までの37カ月分、評価用に96年

6月1カ月分の原稿を用いることとした(表1)。放送原稿テキストデータは記事ごとに分けられており、各記事は日付、原稿名、および複数の本文からなるが、本研究では本文のみを対象とした。

表1: ニュース原稿テキスト

	原稿数	文数	文字数
学習用 (92/8-96/5)	91,115	477,926	97.1M
評価用 (96/6)	4,337	22,712	3.87M
計	95,452	500,638	101M

ニュース原稿の本文には音声化されない記号などの要素が含まれる。これらは以下のような前処理で取り除いておく。

☆★●○◎◇◆□■▲▼▽: ※@ \$ … * など項目の見出しや注意を引くため等の補助記号なので単純に削除する。

() 《 》 [] など主に文の補足説明、単語の読みがなに使われる例が多く、発声されないので削除する¹。

「」中身が引用文章である場合は括弧を外す。

3 言語モデル

3.1 ベースライン

ベースラインとなる言語モデルは、表1の学習用テキストから作成した。まず、JUMAN [8] を用いて形態素解析を行ない、出現頻度上位2万語の形態素を選び語彙とした。本研究では語彙の単位を形態素とし、これを以下「単語」と呼ぶことにする。言語モデルの作成には、CMU/Cambridge SLM Toolkit Ver.1 [9] を用いた。バックオフ・スムージングにはGood-Turingの推定を用い、cut-offの閾値はbigramに対して1とした。ベースライン言語モデルの諸元を表2に示す。

表2: ベースライン言語モデルの規模

学習テキスト	総形態素数: 42.9M 異なり形態素数: 115.8k
言語モデル	語彙: 20k、カバー率: 97.75% unigram: 20k、bigram: 545k

¹ただし固有名詞などの特殊読みが失われるデメリットがある。

3.2 連語の抽出

N-gramなどの統計的言語モデルを用いて音声認識を行なう場合、認識単位をどう設定するかが重要な問題となる。単位が短いと認識誤りが生じやすく、特に附属語ではその影響が大きい。一方、附属語は特定の連鎖が頻繁に起こることが多いと考えられるため、頻出するパターンについては、一単語として登録することが考えられる[10]。

また、語彙の大部分を占める名詞には、「総理-大臣」、「記者-会見」など共起頻度が高い連鎖(複合名詞)が含まれ、これらの名詞連鎖を新たな認識単位として加えることにより、言語モデルの精度向上が期待できる。

本節では、高頻度の附属語連鎖と複合名詞を抽出し、言語モデルの新たな語彙として追加した場合の効果を検討する。

3.2.1 附属語連鎖

学習用テキストコーパスから附属語の2連鎖と3連鎖を抽出し、そのうち出現頻度が100以上の連鎖をそれぞれ783個、1005個選び出した。全異なり連鎖パターンに対するカバー率は、それぞれ95.6%、82.5%であり、これらで大部分の附属語連鎖をカバーしている。なお、本稿では、形式動詞、サ変動詞、判定詞、助動詞、形式名詞、指示詞、助詞、および接尾辞の8品詞の単語を附属語としている。表3は、抽出された上位の附属語連鎖の一部である。

表3: 抽出された高頻度附属語連鎖(上位4)

連鎖数	附属語連鎖(頻度)
2	い・ます(141342)、で・は(106612)、 し・ました(61196)、こと・に(50023)
3	して・い・ます(39887)、 より・ます・と(27035)、 に・より・ます(27027)、 と・いう・こと(26022)

原稿テキスト中に現れたこれらの連鎖を一単語に置き換えた。これをもとにして、出現頻度上位の単語から、ベースライン言語モデルの語彙カバー率97.75%に合わせて、21k語を語彙サイズとして採用し、SLMを用いてbigram言語モデルを作成した。表5に学習用テキストと評価用テキストに対するパープレキシティを示す。高頻度附属語連鎖を加えたことにより、パープレキシティは学習セットに対しては17%低下するが、評価用セットについては4.5%増加している。

3.2.2 複合名詞

附属語に比べ名詞は種類が非常に多く、ベースライン言語モデルの語彙2万語のうち61%が名詞である。また、学習用テキストに出現した名詞単語2連鎖の異なり総数は412kで、単語2連鎖の異なり総数1,717kの42%を占める。名詞が言語モデルに大きな影響を与えていることは確実で、名詞に関するモデルの改善によって認識精度が向上することが期待できる。

高頻度名詞連鎖を抽出して、附属語と同様に語彙に登録することも考えられる。しかし、附属語に比べ名詞連鎖は種類が非常に多く、登録語数を増加させてもカバー率は上がらない。出現頻度が5回以下の地名・人名・団体名などの固有名詞が、名詞連鎖の異なり総数の85%以上を占めており、ニュース記事の性質を反映している。

「総理-大臣」「記者-会見」「警察-本部」などの名詞連鎖は出現頻度が高く、記事の年度によらず共通して現れているため、一単語として登録しておけば、他年度の言語モデルにも利用できると考えられる。一方、これらの名詞は他の多くの名詞とも結び付いて複合名詞を作るため、単に高頻度名詞連鎖から語彙に追加するのは効率が悪い。事実、予備調査によると、学習テキストに出現した頻度100以上の複合名詞4051種類の、全複合名詞に対するカバー率は47.4%に留まる。

したがって、名詞連鎖に関しては、頻度がある程度高く、ほとんど確実に同じパターンで出現する連鎖を選ぶのが効果的であると考えた。そこで本稿では、次節の単語関連率を用いて、新たな認識単位として追加する複合名詞を抽出することにする。

3.2.3 単語関連率

単語関連性の尺度として、Churchらは相互情報量に基づく関連率を提案し、doctor - nurses、United - Statesのような関連性の高い単語対をコーパスから抽出できることを示した[11]。単語xとyの単語関連率(word association ratio)は、xとyを同時に観測する確率を、独立に観測する確率と比較する。

$$I(x; y) = \log_2 \frac{P(x, y)}{P(x)P(y)} \quad (1)$$

$P(x, y)$ の値は、幅wの窓において単語yが単語xに後続する頻度 $f_w(x, y)$ から求める。したがって、 $I(x; y) \neq I(y; x)$ であって、 $P(x, y) = P(y, x)$ を仮定している相互情報量 $I(x, y) = I(y, x)$ とは異なり、単語x、yに関して非対称である。本研究では $w = 2$ とし、ベースライン言語モデルのbigram値をもとに単語関連率を計算した。

表 4: 関連率の高い名詞連鎖 ($I(x; y) \geq 14$)

$I(x; y)$	x	y	$f(x, y)$
16.5	内回り	外回り	25
16.5	マリアナ	海溝	26
16.5	遠心	分離機	31
16.4	COD	化学的酸素要求量	37
15.0	預け	払い	94
...
14.3	円売り	ドル買い	144
14.1	罪状	認否	118
14.0	次席	検事	50
14.0	実況	中継	33

ただし、 $I(x; y)$ の値は共起頻度 $f(x, y)$ が小さいと不安定である。例えば、めったに出現しないが出現すると必ず共起する単語対の $I(x; y)$ は高く、このような単語対は連鎖として登録すべきではない。

表 4 には、ニュース原稿コーパス中に現れた複合名詞のうち、単語関連率が 14 以上のもの 920 個の一部を示す。附属語連鎖と同様に、原稿テキスト中に現れたこれらの複合名詞を一単語に置き換え、これをもとに語彙サイズ 21k 語の bigram 言語モデルを作成した。表 5 に示されているように、パープレキシティはベースラインに比べて、学習セットでは 20%、テストセットで 2.5% 改善されている。

表 5: 連鎖モデルのパープレキシティ

言語モデル	学習セット	テストセット
Baseline	90.85	91.89
附属語連鎖	75.40	96.08
複合名詞	72.75	89.61

3.3 名詞フラットモデル

言語モデルを作成する際に考慮しなければならない要素の一つとして、学習セットとテストセットに含まれる語彙がある。限られた学習データから言語モデルを推定するため、学習データでは出現しなかった、あるいは頻度が閾値より低かった単語でも、テストデータに現れることがある。このような未知語を認識するため、学習済みの言語モデルに語彙を増やすことを考えた。

未知語は名詞であることが多いことが予測されたので、言語モデルを作成する際に名詞であれば、見出し語を「普通名詞、副詞の名詞、時相名詞、形式名詞、人名地名、数詞」の 6 つにセグメント化してバイグラ

ムを生成した。それぞれのセグメントでは、等確率で単語が出現することを仮定したユニグラムを作成し、言語バイグラムと名詞ユニグラムを融合した新たな名詞フラット言語モデルを作成した。

名詞のセグメント内のユニグラムは学習モデルによらないものなので、認識時にその項目をタスクに合わせて適宜変更することができる。今回の実験では、ベースラインのモデルから、一旦名詞となる単語を削除し、テストデータ中に現れる名詞だけを加えた名詞フラットモデルを作成し、その他の品詞については変更をしなかった。名詞は 1208 あった。

3.4 品詞クラスモデル

3.4.1 クラス言語モデル

単語 N-gram モデルは、単語数を増加させるとモデルのパラメータ数が膨大となり、ニュース音声のような大語彙ではその影響は顕著である。さらに、3.2.2 節で述べたように、ニュース記事の性質として次々と新しい話題が現れ、語彙(特に固有名詞など)は流動的であるため、それに合わせて言語モデルを再学習させることが必要である。

そこで、本稿では品詞によるクラス言語モデルについて検討した。すなわち、品詞別の辞書を用意し、それぞれ品詞 C_1, C_2 に属する単語 w_1, w_2 の bigram を、式 (2) で近似する。 $p(C_2 | C_1)$ は記事内容に依存しないと考えられるので、語彙更新の場合は $p(w_2 | C_2)$ の再計算により言語モデルの再学習ができる。

$$p(w_2 | w_1) \simeq p(C_2 | C_1) \times p(w_2 | C_2) \quad (2)$$

3.4.2 言語モデルの作成

品詞クラスによるクラス言語モデルは、品詞クラス間 bigram と品詞内単語 unigram による階層構造で実現する。品詞クラス間 bigram は、品詞列に変換した学習テキストから作成した。品詞内単語 unigram は、品詞毎に学習テキストに出現した単語の頻度をもとに計算する。語彙カバー率がベースラインモデルと同等の 97.75% になるように、頻度 32 未満の単語を切り捨てた。ただし、イ形容詞接頭辞、動詞接頭辞の 2 品詞の単語は元々出現頻度が小さい。前者は単語数 3 ですべて頻度が 1、後者は単語数 4 で最大でも頻度 23 の品詞クラスである。従ってこの 2 者については切り捨てずに残した。その結果、語彙サイズは 19,655 語となった。ちなみに最大の品詞クラスは普通名詞で語彙サイズは 7207、最小は引用助詞で語彙サイズ 1(「と」)であった。品詞クラスは JUMAN Version3.0 における 63 品詞(表 6)とする。

表 6: 品詞クラスの分類

イ形容詞アウオ段、イ形容詞イ段、イ形容詞イ段特殊 イ形容詞接頭辞、カ変動詞、カ変動詞来、サ変動詞 サ変名詞、サ変動詞、タル形容詞、ナノ形容詞、ナ形容詞 ナ形容詞接頭辞、ナ形容詞特殊、引用助詞、格助詞、感動詞 形式名詞、形容詞性述語接尾辞、形容詞性名詞接尾辞 固有名詞、子音動詞カ行、子音動詞カ行促音便形 子音動詞ガ行、子音動詞サ行、子音動詞タ行、子音動詞ナ行 子音動詞バ行、子音動詞マ行、子音動詞ラ行 子音動詞ワ行イ形、子音動詞ワ行、子音動詞ワ行文語音便形 時相名詞、終助詞、助動詞く型、助動詞そうだ型 助動詞だらう型、助動詞ぬ型、人名、数詞、接続詞、接続助詞 地名、動詞性接尾辞、動詞性接尾辞ます型、動詞接頭辞 判定詞、普通名詞、副詞、副詞形態指示詞、副詞の名詞 副助詞、母音動詞、無活用型、名詞形態指示詞 名詞性述語接尾辞、名詞性特殊接尾辞、名詞性名詞助数辞 名詞性名詞接尾辞、名詞接頭辞、連体詞、連体詞形態指示詞
--

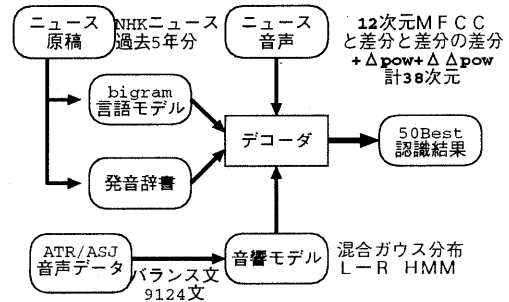


図 1: ニュース音声認識システム

4 音響分析・音響モデル

音響モデルの構築には ATR 音声データベース (B,C セット) と ASJ 音声データベース (Vol. 1-3) を用いた。評価は女性キャスター 1 名の発声で行なうことにし、女性不特定話者音響モデルを HTK(HMM Toolkit)[13] を用いて作成した。ATR データベースから 503 文×2 名、150 文×20 名、ASJ データベースから 150 文×34 名 (一部 153 文発声) 計 9124 文のデータで学習を行なった。

各モデルは 40 音素に無音とショートポーズを加えたもので、各状態単一ガウス分布を持つ 5 状態 3 ループの一方方向性 HMM をもとに、状態共有化した 4 混合ガウス分布型トライフォン HMM を作成した。最終的には 1536 のトライフォンを作成し、これに無音と省略可能な無音の 2 つを加えた 1538 モデルを得た。

音響パラメータは、12 次元の MFCC とその 1 次・2 次差分に対数パワーの 1 次差分と 2 次差分を加えた 38 次元を用いた。

5 認識実験

5.1 認識システム

図 1 に認識システムの概要を示す。認識には HTK を使い、ビームサーチによる N-best デコーディングを行なう。本稿の実験では、ビーム幅は 250 (最大尤度を与えるモデルとの尤度差) とし、認識結果は 50-best までの仮説単語列によって評価した。なお、デコーダのスコア計算における音響モデルに対する言語モデルの重みは、16~20 までの値を用いて実験を行なっている。

5.2 実験条件

5.2.1 評価

認識結果は正解ラベルとの DP マッチングにおいて、N を正解文の単語数、D を脱落数、S を置換数、I を挿入数としたとき、式 (3) で計算される正解精度 (Accuracy) と正解率 (Correct) で評価した。

$$Accuracy = \frac{N - D - S - I}{N} \times 100\%$$

$$Correct = \frac{N - D - S}{N} \times 100\% \quad (3)$$

なお、DP マッチング時には、「今日」と「きょう」を別の単語として評価した。また、キャスター発声と正解単語の読みが一致しない — 例えば「早急」は「さっきゅう」と発声されているが発音辞書には「そうきゅう」しか登録されていない — という問題点もあるが、今回の評価の際には考慮していない。

5.2.2 音声データの分割

システムの評価は、ニュース音声データベース [5] から、女性のアナウンサーによる音声を用いた (表 7)。

ニュース音声における 1 文は 50 単語を越えるような長いものも存在する。推定区間が長い程、計算に必要なメモリ量も計算量も多くなる。また、推定区間を短くすると、単語列候補を同数とった場合には実効探索空間が広くなると考えられる。

これらの理由から、評価用 53 文を検聴により呼気段落で分割したものも用意した。なお、テスト 53 文のベースラインによる認識時間は、UltraSPARC 200MHz CPU のユーザー時間で、分割前が 9 時間 37 分に対して、分割後は 8 時間 54 分であった。正解率と正解精度は、分割前がそれぞれ 70.5%、68.2%、分割後は 77.0%、75.5% であり、認識性能への効果が確かめられたので、評価には分割音声データを用いることにする。

表 8: 言語モデルによる正解精度 (正解率)

言語モデル	テスト 5 文			テスト 53 文		
	s16	s18	s20	s16	s18	s20
Baseline	83.2 (85.3)	85.3 (87.4)	85.3 (86.7)	75.5 (77.2)	75.5 (77.0)	74.4 (76.0)
附属語連鎖	77.8 (79.1)	76.4 (79.1)	73.6 (76.4)	—	—	—
複合名詞	83.2 (85.3)	82.5 (85.3)	81.1 (84.6)	74.5 (76.3)	73.9 (76.0)	72.7 (74.8)
名詞フラット	73.6 (74.3)	75.0 (74.3)	—	72.8 (73.9)	72.6 (73.5)	—

s16, 18, 20 は言語尤度の音響尤度に対する重み条件。

表 7: テスト用ニュース音声

放送日	1996年6月4日
番組名	「ニュースセンター7」
話者	女性1人
テスト 5 文	総時間: 58 秒、単語数: 161 分割音声ファイル数: 16
テスト 53 文 (評価用)	総時間: 10 分 48 秒、単語数: 4920 分割音声ファイル数: 160

テスト 53 文はテスト 5 文を含む。

5.3 認識結果

5.3.1 連語モデル

附属語連鎖の登録による認識性能の評価をテスト 5 文で行なった結果、ベースラインに比べて単語正解精度は 2.2% から 6.4%、正解率は 4.2% から 6.9%、言語モデルの重みに比例して低下している。附属語連鎖が正しく認識されたことにより、挿入誤り数がベースラインよりも減少する効果はあったが、一方で置換エラーが増加した結果、認識率は向上しなかった。言語モデルのテストセットパープレキシティ(表 5)の増加が影響しているものと考えられる。

複合名詞による言語モデルを用いた場合も、テスト 5 文、テスト 53 文に関しては改善が見られず、現状では言語モデルの性能向上を音声認識に反映させることができていない。評価用データに登録した複合名詞が現れなかったことが原因と考えられる。

5.3.2 名詞フラットモデル

語彙サイズの大幅な減少にもかかわらず、認識率の向上が見られなかった。実験に用いられた名詞群は、テストデータに対して完全にクローズなものであり、

名詞に関しては理想的なモデルであると考えられる。このモデルで認識率が低下したのは、名詞をセグメント化し、そのセグメント内ではフラットな出現確率を用いたことによるものである。名詞をセグメント化する際に数詞や人名、地名などで細分化したが、効果が見られなかった。

このことより、単語バイグラムで重要な要素は、単語→品詞の遷移ではなく、単語→単語の遷移であることが推察できる。しかしながら、文中に出現する名詞は、何らかの規則を適応することで置き換えが可能であり、名詞を意味によって細分化することで効果が得られるかもしれない。

5.3.3 品詞クラスモデル

品詞クラスモデルの認識性能をテスト 53 文で行なった結果、言語重み s14 の場合単語精度は 24.9%、単語正解率 27.1% である。

ベースラインの bigram 値が高い単語遷移でも、品詞クラス間 bigram 値と品詞内単語 unigram 値のどちらかが小さくても積は小さくなってしまいます。その結果、認識時に遷移が本来の単語遷移の確率よりも低く評価されることが多かったのが正解率の低くなる主な原因の一つだと思われる。今回は活用型も含めた品詞分類による品詞クラスモデルの作成を試みたが、結果として活用語の活用型によるクラスは小さ過ぎるという問題があった。品詞のみで分類するなど、より大きなクラスを用いることが考えられる。逆に、名詞クラスなどに関しては、たとえば意味分類 [12] により細分化するなどの工夫をしてどの程度細分化できるかを検討して行きたい。

6 まとめ

本稿では、ニュース音声認識タスクに対して、連語モデル、名詞フラットモデル、品詞クラスモデルの 3

種類の N-gram 言語モデルを適用し、その効果を検討した。

連語モデルでは、より長い認識単位を用いることによって認識性能の改善を図るという目的で、高頻度付属語と関連率の高い複合名詞を新たな語彙として認識辞書に登録して言語モデルを作成した。名詞フラットモデルは、ベースラインの言語モデルから、一旦名詞を削除した後、テストデータ中に現れる名詞だけを登録し、名詞語彙をタスクに合わせて適宜変えることを試みた。品詞クラスモデルでは、名詞以外についても品詞・活用型で分類したクラス言語モデルを作成した。その結果、複合名詞による連語モデルを用いることにより言語モデルの性能は改善されたが、現状では音声認識性能には反映できていない。

今回、連語モデルについては付属語と名詞の2連鎖・3連鎖のみを取り扱ったが、今後はコーパス中に頻出するさらに長い単語連鎖についても適用を広げ、検討したい。名詞フラット、品詞クラスモデルにおけるクラス分類法については、さらに意味分類を考慮する、単語のクラスタリングを行なうなど、検討を重ねる必要がある。

謝辞

この研究は、NHK 放送技術研究所との共同研究「ニュース音声認識アルゴリズムの研究」の一部として行なわれた。ニュース音声データベース、各種分析ツールを提供されたNHK 放送技研の関係諸氏に深く感謝する。ニュース原稿テキストの形態素解析にはJUMAN、言語モデルの構築にはCMU/Cambridge SLM Toolkitを利用した。

参考文献

- [1] D. B. Paul, J. M. Baker, "The design for the Wall Street Journal-based CSR corpus," Proc. DARPA Speech and Natural Language Workshop, pp. 357-362 (1992.02)
- [2] 大附克年、森岳至、松岡達雄、古井貞熙、白井克彦, "新聞記事を用いた大語彙連続音声認識の検討," 信学技報, SP95-90 (1995.12)
- [3] 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彙連続音声認識," 信学論, Vol. J79-D-II, No. 12, pp. 2125-2131 (1996.12)

- [4] A. I. Rudnicky, "HUB-4: Business broadcast news," Proc. DARPA Speech Recognition Workshop, pp. 8-11 (1996.02)
- [5] 安藤彰男、宮坂栄一, "ニュース音声データベースの構築," 音講論, 2-Q-9 (1997.03)
- [6] 田口雄一、大附克年、松岡達雄、古井貞熙、白井克彦, "ニュース音声を対象とした大語彙連続音声認識," 音講論, 2-6-11 (1997.03)
- [7] 小林彰夫、今井亨、安藤彰男、宮坂栄一、赤松裕隆、中川聖一、小黒玲、尾関和彦、古井貞熙、鈴木順子、白井克彦, "ニュース音声認識システムの検討," 音講論, 3-1-9 (1997.09)
- [8] 日本語形態素解析システムJUMAN, Version 3.0, <http://pine.kuee.kyoto-u.ac.jp/nl-resource/juman.html>
- [9] P. Clarkson, R. Rosenfeld, "Statistical language modeling using the CMU-Cambridge Toolkit," Proc. EUROSPEECH-97, pp. 2707-2710 (1997-09)
- [10] 小林紀彦、中村裕一郎、肥田木康明、小林哲則, "統計的言語モデルにおける付属語の扱いに関する一考察," 音講論, 2-1-6 (1997.09)
- [11] K. W. Church, P. Hanks, "Word association norms, mutual information and lexicography," Computational Linguistics, Vol.16, No. 1, pp. 22-29 (1990.03)
- [12] 安藤彰男、今井亨、鈴木順子、白井克彦, "意味分類を利用した統計的言語モデルの検討" 音講論, 2-1-8 (1997.09)
- [13] HTK: Hidden Markov Model Toolkit, Version 2.01, <http://www.wntropic.com/htk/htk.html>