

N-gram モデルのエントロピーに基づく パラメータ削減に関する検討

踊 堂 憲 道[†] 伊 藤 克 亘^{††}
鹿 野 清 宏[†] 中 村 哲[†]

人間の発話を文字に変換する大語彙連続音声認識(ディクテーション)技術は、キーボード入力の省力化や、様々な環境下における人間とコンピュータ間の音声インタフェースの実現のために必要不可欠な技術であり、活発に研究が行われている。認識システムには、人間の言語知識の役割を果たす言語モデルが組み込まれており、一般的には統計的言語モデルである N-gram が用いられている。しかし、数千語~数万語を対象とする場合、N-gram モデルのパラメータが指数関数的に増大し、システム構築に際して、大きな障害が生じることになる。本稿では、これまでに提案された種々の N-gram モデルのパラメータ削減手法の比較を行う。また、我々が提案する削減手法を (N-1)-gram に適用するための予備実験を行ったので、その結果について報告する。

A Study on Entropy-based Compression Algorithms for N-gram Parameters

NORIMICHI YODO,[†] KATSUNOBU ITO,^{††} KIYOHICO SHIKANO[†]
and SATOSHI NAKAMURA[†]

Large vocabulary continuous speech recognition (LVCSR), which is simply called as dictation, is an essential technology for the realization of voice typing and interface between human being and a computer in various conditions. An LVCSR system reduces search space using language models, where statistical N-gram models are generally used. However, they need a huge number of parameters that grow exponentially with N and the vocabulary size. Especially in the task with large vocabulary (from a few thousand of words to several ten thousands of words), their huge memory requirement results in the system implementation difficulty. In this paper we compare algorithms for reducing the number of parameters of an N-gram model. Preliminary experiments on the augmentation of our compression algorithm to deal with (N-1)-gram are carried out.

1. はじめに

連続音声認識技術は年々進歩をとげ、最近では人間の自然な発話をコンピュータに認識(文字化)させる大語彙連続音声認識(ディクテーション)の研究がさかんに行われている。たとえば、情報処理振興事業協会(IPA: Information Technology Promotion Agency)の「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^{1),2)}は複数の大学や国立研究所が連携して研究開発し、技術を一般に公開することで、我が国

の音声認識研究のレベル向上に貢献している。また、PC上で動くソフトウェアがIBM³⁾やNEC⁴⁾などから市販されており、音声認識技術の認知度もずいぶんと上がってきている。この技術の進歩と認知度の向上は、音声ワープロやニュース原稿書き起こし、あるいは障害者のコミュニケーション手段や自然発話によるカーナビゲーションシステムなどの実現への追い風となるだろう。さて、音声認識は一種の探索問題であり、大語彙連続音声認識などの複雑なタスクでは言語モデルを用いて探索空間を縮小する。音声認識における言語モデルといえば、かつては文法に基づくものもっぱらであったが、最近では、

1. 半自動的に構築できる(ツールも提供されている⁵⁾),
2. モデル作成に必要なテキストデータベースの整

[†] 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Nara Institute
of Science and Technology

^{††} 電子技術総合研究所
Electrotechnical Laboratory

備が行われた⁶⁾,

3. 統計的な音響モデルとの相性が良い, などの理由から, 統計的言語モデルである N-gram モデルが主流である^{7)~9)}. この N-gram モデルを用いたシステムの実用性は, 実行時間や必要なメモリ容量に大きく依存する. タスクが複雑になるほど, N-gram の学習に用いるデータベースが大きくなるほど, これらの問題は重要になる. この理由は, N-gram モデルは N が大きくなると, 必要なパラメータが指数関数的に増えるからである. 解決策として, これまでに, cutoff 手法や情報量尺度に基づいた N-gram モデルのパラメータ削減手法がいくつか提案されている. これら従来の手法については次章で紹介する.

我々は先ほど触れた IPA のプロジェクトにおいて, 言語モデルの圧縮に関する研究を行っており, 新しいパラメータ削減手法を提案し^{10)~12)}, その効果を確認した^{1),2)}. その手法は, 最尤推定に基づいており, 1つのパラメータを削除する際の言語モデルの性能劣化を定量的に評価し, 削除する優先度を決定する. そして, 優先度の高いパラメータから削除し, back-off 係数を更新していくことで, 任意のパラメータ数での言語モデルの作成を可能にした.

しかし, 語彙を大きくする場合や, N-gram モデルの拡張を考慮すると, さらなる圧縮の可能性を検討する必要がある. 本稿では, 2万語を用いた連続音声認識実験におけるパラメータ削減手法の比較結果, および提案手法の拡張として, N-gram と (N-1)-gram パラメータを同時に削減する効果について報告する.

2. 従来の N-gram モデルのパラメータ削減手法

この章では, これまでに提案されてきた, N-gram モデルのパラメータ削減手法について考察する. なお, 未知の単語系列は, back-off smoothing を用いて推定する back-off N-gram を対象とする.

2.1 cutoff 手法

パラメータ削減手法として最もよく使われる cutoff 手法は, 学習データに出現する回数がある値よりも小さいものをカウントしない方法である. 単純な手法であるが, パラメータ削減の効果は大きい. しかし, 以下の2点が考慮されていない.

1. 元の確率値と, back-off smoothing によって推定される値が大きく異なる場合, すなわち「N-gram は (N-1)-gram と似ている」という経験則が破綻する場合
2. 出現回数が同じでも確率値, すなわちコンテキ

ストに対する単語の出現確率が異なる場合 ($p_1 = 1/100, p_2 = 1/2$ が同時に削除される)

2.2 back-off 確率との比

元の N-gram 確率値と back-off により推定される確率値との比に着目する手法が提案されている¹³⁾. ここでは,

$$K * (\log(\text{元の確率}) - \log(\text{back-off 確率})) \quad (1)$$

を尺度としている. ここで K はディスカウントされた N-gram の出現回数である. この手法は, 頻度を考慮していない点が問題である. すなわち, 2つのパラメータの値の比が同じであれば区別されないことになる.

2.3 エントロピー尺度 (一括削減)

情報量尺度に基づく削減手法は多数提案されている. 基本的には, N-gram の確率分布と (N-1)-gram の確率分布の相関や距離を種々の情報量を用いて評価する考え方である. たとえば, “N-gram の確率分布と (N-1)-gram の確率分布の距離”として, 相対エントロピーが一般的に用いられている^{14)~16)}. これらの手法では, N-gram パラメータを (N-1)-gram パラメータで総替えするため, パラメータごとに見れば, 必ずしも性能劣化が少ないものから削除されるというわけではないという問題点がある.

3. エントロピーに基づく逐次削減手法

我々が提案したこの手法は, パラメータ削除による言語モデルの性能低下を, エントロピー尺度に基づいて定量的に, パラメータごとに評価するものであり^{10)~12),17)}, back-off 係数を更新していく点, 任意のパラメータ数を実現できるという点に特徴がある. Stolcke は, ニュース音声のタスクにおいて, この手法を 4-gram に適用した結果を報告している¹⁷⁾. ここでは, N-gram パラメータと (N-1)-gram パラメータが独立であるという仮定の下に, パラメータ削減が行われている. 1つのしきい値でパラメータ数が決まるため, N-gram と (N-1)-gram パラメータ数のバランスについては考慮されていないという点に問題が残る. 我々は, 上でいうところのバランスを考慮に入れた実験を行うが, ここではまず, アルゴリズムの説明を行う.

さて, 簡単化のために trigram を例にとって考えることにする (図 1). プロセスは, 次のようになる.

1. back-off N-gram モデルを作成する (ディスカウント手法は任意).
2. あるコンテキスト, すなわち N-1 単語列を固定して考えたとき, 次の2つの確率分布を得る.

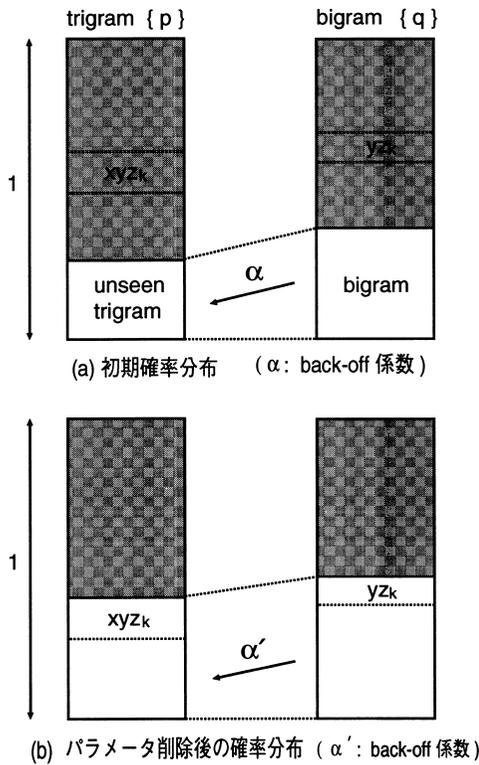


図1 back-off 係数の更新
Fig.1 Update of back-off coefficient.

$\{p\}$ 元のモデルの条件付き確率分布
 $\{p'\}$ 1つの N-gram パラメータを削除し、back-off により推定する場合に得られる確率分布

- 以上の2つの分布と、新しい back-off 係数 (α') を用いて、エントロピーの増加量を求める。その値は2つの分布間の相対エントロピーにコンテキストの頻度をかけることによって求められる。
- エントロピーの増加量が小さいものから、あるいは与えられたしきい値以下のものを削除し、必要であれば back-off 係数を更新する(終了)。
この過程における、back-off 係数の更新方法および、エントロピーの増加量の算出法について整理する。

3.1 back-off 係数の更新法

trigram モデルを考える場合、コンテキストである2単語列によってモデルの空間、すなわちすべての trigram パラメータを分類することができる。ここでは、3単語列 xyz_1, xyz_2, \dots は、「部分空間 Q_{xy} に含まれる」という表現を用いて、「各部分空間は他の部分空間に対して独立である」と仮定する。今、部分空間 Q_{xy} だけを考慮して、削除前の trigram パラメータ

xyz_i ($i = 1, 2, \dots$) の確率分布を $\{p\}$ 、対応する bigram パラメータ yz_i の確率分布を $\{q\}$ とする。また、未知の trigram および bigram に与えられる確率値の総和をそれぞれ p_{unks}, q_{unks} とする。このとき、back-off 係数 α は、 c_+ を $C(xyz_i) > 0$ である i の集合とすると、

$$\alpha = \frac{p_{unks}}{q_{unks}} = \frac{1 - \sum_{i \in c_+} p_i}{1 - \sum_{i \in c_+} q_i} \quad (2)$$

となる(図1(a))。

今、trigram パラメータ xyz_k を削除する場合、 $p_k = P(z_k|xy)$ を bigram から back-off smoothing により推定することになる。このとき、新しい back-off 係数を以下の値に更新する(図1(b))。

$$\alpha' = \frac{p'_{unks}}{q'_{unks}} = \frac{p_{unks} + p_k}{q_{unks} + q_k} \quad (3)$$

最終的に、初期状態の確率分布 $\{p\}$ と(1つの)パラメータ削除後の分布 $\{p'\}$

$$\{p'\} = \begin{cases} p_i & i \neq k, i \in c_+ \\ \alpha' \cdot q_k & k \in c_+ \\ \alpha' \cdot q_i & i \notin c_+ \end{cases} \quad (4)$$

を得ることになる。

3.2 エントロピー変化量

一般に「言語モデルが表現する言語の複雑さ」を表す尺度として、エントロピーが用いられる。3単語列 xyz の生起確率を $p(xyz)$ で表すと、エントロピーは、

$$\mathcal{H} = - \sum_{xyz} p(xyz) \log p(xyz) \quad (5)$$

で表され、最尤推定式と等価である。部分空間 Q_{xy} の生起確率を $P(Q_{xy})$ とすると、式(5)は、trigram 確率(条件付き確率) $p_i = P(z_i|xy)$ を用いて、

$$= - \sum_{xy} P(Q_{xy}) \log P(Q_{xy}) + \sum_{xy} P(Q_{xy}) \left(- \sum_i p_i \log p_i \right)$$

と書き直すことができる。

さて、全空間でのエントロピーの変化量は、部分空間 Q_{xy} における変化量と等しいと仮定しているので、 $C(\cdot)$ を事象の生起回数として、

“unks”は unknown words を意味する。

$$\begin{aligned} \Delta \mathcal{H} &= \mathcal{H}' - \mathcal{H} \\ &= P(Q_{xy}) \sum_i p_i \log \frac{p_i}{p'_i} \\ &= \frac{C(xy)}{C(all)} \times \mathcal{D}(p||p') \end{aligned} \quad (6)$$

となる．ここで， $C(all)$ は，bigram の総出現回数であり， \mathcal{D} は，相対エントロピーである．

同様に，パープレキシティーの変化量は，空間 Q_{xy} だけで考えると，

$$\begin{aligned} \log(\Delta \mathcal{P}\mathcal{P}) &\propto \sum_i C(xy z_i) \log \frac{p_i}{p'_i} \\ &= C(xy) \times \mathcal{D}(p||p') \end{aligned} \quad (7)$$

という式で求めることができる．

4. パラメータ削減手法の評価 (パープレキシティー)

前章で取り上げた種々の方法の比較を行う．ここでは，CD-ROM 版毎日新聞記事から言語モデル (back-off trigram) を作成し，テストセットパープレキシティーを評価尺度に用いた．

学習には 45 カ月分 (91 年 1 月～94 年 9 月) を，テストセットには 3 カ月分 (94 年 10～12 月) のデータを用いた．なお，言語モデルの学習には不要と思われるような記事や文はあらかじめ除いてある¹⁸⁾．形態素解析には，奈良先端科学技術大学院大学で開発された茶釜システム¹⁹⁾ をベースに，IPA が改良を加えた解析システムを用いた．さらに，数詞や助動詞に対する読みや，連濁に対応した読み付とツールを用いて処理を行った²⁰⁾ (2 つの処理をまとめて形態素解析と呼ぶことにする)．

学習データの形態素解析結果を表 1 に示す．本実験では，形態素，その原形，品詞番号を考慮して単語を識別している．すなわち，表 2 の下線部の形態素はすべて区別される．以下では，形態素を単に単語と呼ぶことにする．

4.1 trigram パラメータの削減

このデータの頻度上位 20,000 単語を用いて cutoff 手法，相対エントロピーによる一括削減，逐次削減モデルを作成し，パラメータの数 (bigram + trigram) とテストセットパープレキシティーとの関連を調べた．図 2 にその結果を示す．図中の () 内の数字は trigram パラメータの削減率を示している．グラフ中の最も右のポイントに対応するモデルは，bigram に対して 1，trigram に対して 1 の cutoff を行っている．なお，言語モデルの作成には，CMU-Cam-Toolkit⁵⁾ を用い，back-off smoothing の際のディスカウンティングは

表 1 学習データ解析結果 (45 カ月)
Table 1 Training data (45 months).

総文数	約 237 万文
総単語数	約 6,623 万語
異なり単語数	約 19 万語
一文あたりの単語数	28.16 語
5 K, 単語被覆率	88.23%
20 K, 単語被覆率	96.47%
90%被覆単語数	6,315
97%被覆単語数	22,959

表 2 単語の定義

Table 2 The definition of a word.

用例	形態素, 原形, 品詞
私は <u>知らない</u> ．	(知ら, 知る, 連用形ナイ接続)
そうとは <u>知ら</u> ず	(知ら, 知る, 連用形ズ接続)
100 <u>円</u>	(円, 円, 助数詞)
<u>円</u> が回復	(円, 円, 名詞)

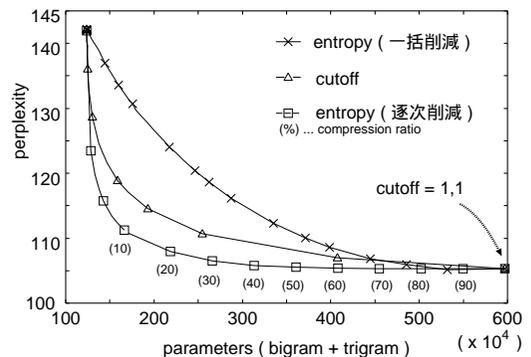


図 2 パラメータ削減手法の比較

Fig. 2 Comparison of various compression algorithm.

Witten-Bell 法で行っている²¹⁾．同じパラメータ数では，逐次的削減手法によるモデルが最も良い性能を示し，次に cutoff によるモデル，一括削減によるモデルとなっている．とりわけ逐次削減手法では，trigram パラメータを 1/2 に削減しても精度を維持している．

4.2 逐次削減手法の bigram パラメータへの適用
前節の結果から，削減手法としては，相対エントロピーに基づく逐次削減が最も効果的であることが分かった．さらなる言語モデル圧縮を実現するために，(N-1)-gram (ここでは bigram) にもこの削減手法を適用した．具体的には，trigram パラメータを式 (7) に基づいて削減し，次に全 bigram パラメータの中で，bigram における式 (7) の値が小さいものから削減した．テストセットパープレキシティーを求めた結果を図 3 に示す．図の太線は，trigram パラメータだけを削減したときのパープレキシティーの変化を示している．

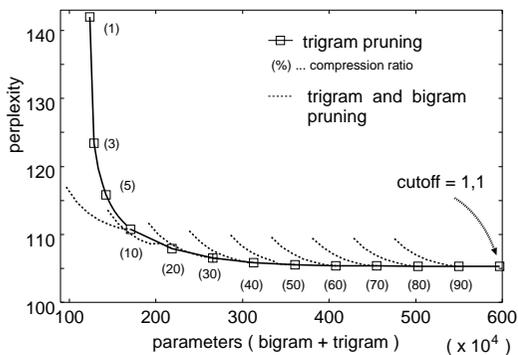


図3 逐次削減手法による言語モデル性能

Fig. 3 The performance of language models compressed by proposed method.

そして、太線から分岐する各線は trigram パラメータ数を固定して、bigram パラメータを減らして(40%まで)いったときの変化を示している。

trigram パラメータの削減率が小さいとき(多くが残っている)は、bigram パラメータの削減が言語モデルの性能を大きく低下させることが分かる。しかし、trigram パラメータの削減率が大きくなってくると、ある領域では、bigram パラメータを削減した方が高い性能が得られることが分かる。これは、1 単語をアルファベット 1 字で表すと、trigram ABC を削除して bigram BC から推定するよりは、bigram XY を unigram Y から推定する方が、評価データのパープレキシティーの劣化が少ない、そういう trigram, bigram があることを示している。

5. 大語彙連続音声認識による評価

2 万語を対象とする大語彙連続音声認識実験を行い、パラメータの削減が認識精度にあたる影響を調べる。実験には、情報処理振興事業協会 (IPA) の「日本語ディクテーション基本ソフトウェア, 1998 年度版 (CD-ROM)」に納められている音響モデル、言語モデル、認識エンジン JULIUS、および「JNAS データベース (毎日新聞記事読み上げ文)」を用いた。

音響モデルは表 3 に示された triphone を用い、音声データは JNAS データベースから男女各 23 名の発話による 804 文章を選んだ。総単語数は 6,331 単語で、そのうち 114 単語 (約 1.8%) は未知語である。このデータは音響モデル、言語モデル双方に対してオープンなものである。

京都大学で開発されている認識エンジン JULIUS²²⁾ は、探索を 2 パスで行っている。第 1 パスでは、left-to-right の bigram モデルを、第 2 パスでは、right-to-left

表 3 音響モデル

Table 3 Acoustic model.

サンプリング周波数	16 [kHz]
プリエンファシス	0.97
分析窓	Hamming Window
分析窓長, 窓間隔	25 [ms], 10 [ms]
特徴パラメータ	MFCC (12) + ΔMFCC (12) + Δ Pow (計 25 次)
混合数	16
状態数	2000

表 4 初期モデルの選択

Table 4 The selection of an initial model.

	cutoff なし	cutoff 1
bigram	2,329,197	1,238,929
trigram	14,505,266	4,731,760
Accuracy (%)	84.52	84.13

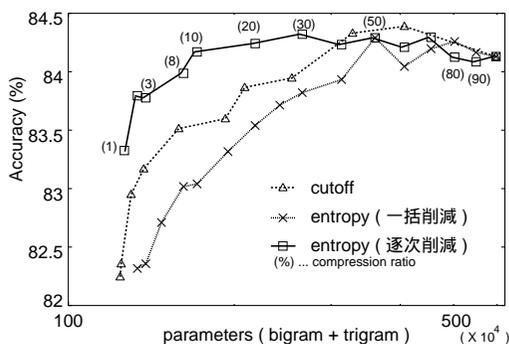


図 4 パラメータ削減手法の比較

Fig. 4 Comparison of three algorithms.

(逆向き)の trigram モデルを用いる仕様となっている。JULIUS の実行に必要なメモリの半分以上が trigram モデルによって占められている。ここでは、第 1 パスの bigram を固定し、第 2 パスで用いる trigram を種々変えて、認識精度の変化を調べた。cutoff をまったく行わないモデルを用いた場合と、bigram と trigram に対して cutoff (それぞれ 1) を行ったモデルを用いた場合で、認識率に大きな差はみられなかったため後者を初期モデルとして使うことにする。表 4 に両モデルのパラメータ数、認識率を示しておいた。

5.1 cutoff と逐次削減手法の比較

cutoff と逐次削減手法を用いて、trigram パラメータの削減を行ったところ、図 4 のような結果が得られた。逐次削減法は、trigram パラメータ数を約 1/3 に減らしても精度を維持しており、cutoff よりも精度の劣化が少ないことが分かる。特に、大幅な削減を行った場合(グラフ左側部分)にはその差が大きくなり、パープレキシティーで見られた傾向が反映されている。グラフでは、パラメータ数を横軸にとっているが、実

表5 言語モデルのサイズと性能
Table 5 The size of language model and their performance.

圧縮方法	PP	file size	Acc (%)
提案手法 (10%)	110.75	83.6 MB	84.17
cutoff	110.68	124.2 MB	83.95

PP... perplexity Acc... Accuracy

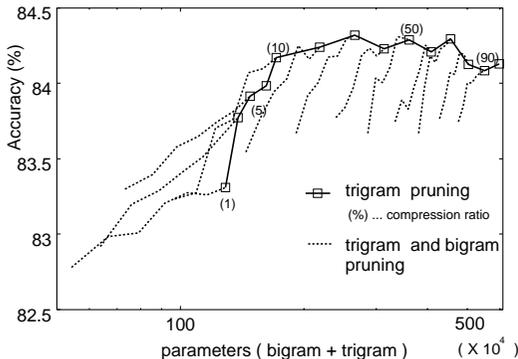


図5 逐次削減手法によるモデルと単語正解率

Fig. 5 The performance of models compressed by proposed method (Accuracy).

際にはファイルサイズが重要である。表5に、パープレキシティーがほぼ等しい言語モデルの大きさと認識率をあげておいた。

5.2 bigram への適用 (逐次削減法)

次に, trigram と bigram の両パラメータを逐次削減手法により同時に削除したモデルを用いて認識実験を行った (図5)。太線は, trigram パラメータだけを削減した場合の認識率の変化を示している (前節と同じ)。太線から分岐している各線は, trigram パラメータを固定し bigram パラメータを削減したモデルを用いた場合の結果を示している。trigram パラメータの削減率が小さいときは bigram パラメータの削減は精度の低下を引き起こし, trigram パラメータの削減率が大きくなってくると, ある領域では bigram パラメータを削減した方が高い認識率を示すことが分かる。これはパープレキシティーでもみられた傾向であり, パープレキシティーと認識率の間に相関性があることが分かる。

6. おわりに

本稿では, 種々の N-gram モデルのパラメータ削減手法を trigram モデルに適用し, パープレキシティーを比較した。結果, 我々が提案する手法の効果を確認した。また, 削減手法の拡張として, trigram パラメータと bigram パラメータを同時に削減する予備実験を

行い, ある領域では, trigram パラメータを減らすよりも bigram を減らした方がよいという事実を確認した。さらに, 提案手法を用いて 20,000 単語を対象とする大語彙連続音声認識実験を行い, パープレキシティーと同様の効果を確認した。

以上の結果から, パープレキシティーを尺度として, 逐次削減法により N-gram と (N-1)-gram パラメータを同時に削減することで, N-gram のみを削除するよりも高い精度を得られることを確認した。

謝辞 本研究は情報処理振興事業協会 (IPA) の「日本語ディクテーション基本ソフトウェア」プロジェクトの支援を受けて行われた。関係者の皆様にこの場を借りてお礼申し上げます。

参考文献

- 1) 情報処理振興事業協会: 日本語ディクテーション基本ソフトウェアの開発, 調査研究報告書, 1998年度 (1999).
- 2) 河原達也, 李 晃伸, 小林哲則, 武田一哉, 峯松信明, 伊藤克亘, 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 (98年度版) の性能評価情報処理学会研究報告, SLP26-6, pp.39-46 (1999).
- 3) <http://www.ibm.co.jp/voiceland/>
- 4) <http://www.nec.co.jp/japanese/product/kiban/control/product/ninsiki/mt.html>
- 5) Clarkson, P. and Rosenfeld, R.: Statistical Language Modeling Using The CMU-Cambridge Toolkit, *ESCA Eurospeech 1997*, Vol.5, pp.2707-2710 (1997).
- 6) 武田一哉, 伊藤克亘, 松岡達雄, 竹沢寿幸, 鹿野清宏: 大語彙連続音声認識研究のためのテキストデータ整備, 情報処理学会研究報告, SLP11-9, pp.49-32 (1996).
- 7) 伊藤彰則, 代島直人, 丸山 敦, 加藤正治, 好田正紀: 大語彙言語データベースからの N-gram 構築とタスク適応の検討, 情報処理学会研究報告, SLP11-5, pp.25-30 (1996).
- 8) 大附克年, 森 岳至, 松岡達雄, 古井貞照, 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, pp.63-68 (1995).
- 9) 吉田航太郎, 松岡達雄, 大附克年, 古井貞照: 単語 trigram を用いた大語彙連続音声認識, 情報処理学会研究報告, SLP14-14, pp.99-104 (1996).
- 10) 踊 堂: 形態素単位の N-gram モデルの構築と圧縮に関する研究, 奈良先端科学技術大学院大学修士論文 (1998).
- 11) 踊堂憲道, 鹿野清宏, 中村 哲: 情報量に基づく trigram パラメータの逐次的削減手法, 情報処理学会研究報告, SLP22-17, pp.91-96 (1998).
- 12) Yodo, N., Shikano, K. and Nakamura, S.:

Compression Algorithm of Trigram Language Models Based on Maximum Likelihood Estimation, *Proc. ICSLP-98*, pp.716-719 (1998).

- 13) Seymore, K. and Rosenfeld, R.: Scalable Back-off Language Models, *Proc. ICSLP-96*, pp.232-235 (1996).
- 14) Bonafonte, A. and Marino, J.B.: Language Modeling Using X-grams, *Proc. ICSLP-96*, Vol.1, pp.394-397 (1996).
- 15) Kneser, R.: Statistical Language Modeling Using a Variable Context Length, *Proc. ICSLP-96*, Vol.1, pp.494-497 (1996).
- 16) Ron, D., Singer, Y. and Tishby, N.: Learning Probabilistic Automata with Variable Memory Length, *7th Annual ACM Conf. on Computational Learning Theory*, pp.35-46 (1994).
- 17) Stolcke, A.: Entropy-based pruning of back-off language models, *Proc. Broadcast News Transcription and Understanding Workshop*, pp.270-274 (1998).
- 18) 伊藤克亘, 松岡達雄, 竹沢寿幸, 武田一哉, 鹿野清宏: 大語彙連続音声認識のためのテキストデータ処理, 音学講論, pp.105-106 (Sep. 1996).
- 19) 松本裕治, 北内 啓, 山下達雄, 平野善隆: 日本語形態素解析システム「茶筌」version2.0 使用説明書, Information Science Technical Report NAIST-IS-TR99008, 奈良先端科学技術大学院大学 (1999).
- 20) 伊藤克亘, 山田 篤, 天白成一, 山本俊一郎, 踊堂憲道, 山本幹雄, 宇津呂武仁, 鹿野清宏: 日本語ディクテーションのための言語資源・ツールの整備, 情報処理学会研究報告, SLP26-5, pp.31-38 (1999).
- 21) Placeway, P., Schwartz, R., Fung, P. and Nguyen, L.: The Estimation of Powerful Language Models from Small and Large Corpora, *Proc. ICASSP-93*, Vol.II, pp.33-36 (1993).
- 22) 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大語彙連続音声認識エンジン JULIUS, 信学技報, SP98-3, pp.17-24 (1998).

(平成 11 年 11 月 11 日受付)

(平成 12 年 12 月 1 日採録)



踊堂 憲道

平成 8 年京都大学工学部情報工学科卒業。平成 10 年奈良先端科学技術大学院大学情報科学研究科情報処理学専攻博士前期課程修了, 同博士後期課程進学, 現在に至る。音や言葉が人の心身に与える様々な影響に関心を持っている。

音響学会会員。



伊藤 克亘 (正会員)

昭和 63 年東京工業大学工学部情報工学科卒業。平成 5 年同大学院博士課程修了。同年, 電子技術総合研究所入所, 現在に至る。平成 12 年, 奈良先端科学技術大学院大学客員助教授。音声認識/音声対話システムが本当に役に立つのかどうかを見極めたいと思っている。日本音響学会会員。



鹿野 清宏 (正会員)

昭和 45 年名古屋大学工学部電気工学科卒業。昭和 47 年同大学院修士課程修了。同年電電公社武蔵野電気通信研究所入所。昭和 59~61 年カーネギーメロン大学客員研究員。昭和 61~平成 2 年 ATR 自動翻訳電話研究所音声情報処理研究室長。平成 4 年 NTT ヒューマンインタフェース研究所主席研究員。平成 6 年より奈良先端科学技術大学院大学情報科学研究科教授。音情報処理学講座を担当。工学博士。主として音声・音情報処理の研究および研究指導に従事。昭和 50 年電子通信学会米沢賞。平成 3 年 IEEE SP 1990 Senior Award, 平成 6 年日本音響学会技術開発賞。平成 12 年情報処理学会山下記念研究賞, IEEE, ISCA, 音響学会, 電子情報通信学会各会員。



中村 哲 (正会員)

昭和 56 年京都工芸繊維大学工学部電子工学科卒業。昭和 56~平成 6 年シャープ(株)中央研究所および情報技術研究所に勤務。昭和 61~平成元年 ATR 自動翻訳電話研究所に出向。平成 6~12 年奈良先端科学技術大学院大学情報科学研究科助教授。平成 8 年 3 月~8 月 Rutgers University・CAIP Center 客員教授。平成 12 年 4 月より ATR 音声言語通信研究所第一研究室長, 音声情報処理, 主として音声認識の研究に従事。京都大学博士(工学)。平成 4 年日本音響学会粟屋学術奨励賞受賞。IEEE, 日本音響学会, 人工知能学会各会員。