

# 距離つき音節 n-gram 索引による 音声検索語検出の距離尺度の厳密化

岩見 圭祐<sup>1</sup> 坂本 渚<sup>1,a)</sup> 中川 聖一<sup>1,b)</sup>

受付日 2012年5月31日, 採録日 2012年11月2日

**概要:** ニュースや新聞記事のようにテキスト情報を含むものであれば既存のテキスト検索エンジンを用いることで、欲しい情報を高速に検索することができる。しかし、現在のところ音声ドキュメントに対しての有効な検索手法は確立されていない。その理由としてあげられるのが、未知語や認識誤りといった音声ドキュメント特有の問題である。我々は今までにサブワード単位の認識結果を用いた高速な音声検索語検出 (STD) 手法を提案してきた。音節ラティスから置換誤り、挿入誤りを考慮した距離つき n-gram 索引を構築し、脱落誤りを考慮したクエリで検索することで未知語と認識誤りに頑健な検索が可能となった。本論文では、挿入誤り距離と脱落誤り距離、およびダミー音節とクエリ中の音節との置換誤り距離の厳密化手法を提案し、その評価結果について述べる。

キーワード: 音声検索語検出, 音節認識, n-gram, 未知語, 既知語

## Strict Distance Measure for a Spoken Term Detection Method Based on a Syllable n-gram Index with Distance Metric

KEISUKE IWAMI<sup>1</sup> NAGISA SAKAMOTO<sup>1,a)</sup> SEIICHI NAKAGAWA<sup>1,b)</sup>

Received: May 31, 2012, Accepted: November 2, 2012

**Abstract:** We can find the information with an existing textual search engine if the target data consist of text information such as news and newspaper, but efficient spoken term detection (STD) method has not been established, because there are specific problems such as some recognition errors and out-of-vocabulary (OOV) terms for spoken documents. Therefore, sub-word unit based recognition and retrieval methods have been proposed. In our previous work, we proposed a very fast Japanese STD system that is robust for considering OOV words and mis-recognition of sub-units. We used individual syllables as sub-word unit in continuous speech recognition and an n-gram index of syllables with distance metric in a recognized syllable-based lattice. In this paper, we describe the improvement method, especially, the exact definition of insertion/deletion error distance and the distance between the dummy syllable and an arbitrary syllable in a query, and show an experimental result.

**Keywords:** spoken term retrieval, syllable recognition, n-gram, Out-of-Vocabulary, In-Vocabulary

### 1. はじめに

従来の音声ドキュメント検索語検出 (Spoken Term Detection: STD) 手法として、最も簡単な方法は、大語彙連

続音声認識の書き起こしの結果に対して単語単位のテキスト検索を行う方法で、既知語に対しては高精度な検索が可能である。しかし、この方法では、未知語や音声認識誤りの問題に対処することができない。未知語とは、大語彙連続音声認識の辞書にない単語のことである。辞書にない単語は、認識結果に現れることがないため、単語単位のテキスト検索では、未知語を検出することは不可能である。また、置換、挿入、脱落の認識誤りの問題もある。置換、脱

<sup>1</sup> 豊橋技術科学大学  
Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan

a) sakamoto@slp.cs.tut.ac.jp

b) nakagawa@slp.cs.tut.ac.jp

落の認識誤りによって、辞書に登録されている単語であっても認識結果に現れない場合があります、その場合はテキスト検索を使用しても検索することができなくなってしまう。そのため、大語彙連続音声認識システムの性能によってテキスト検索の性能も決まってしまう。未知語に対しては、サブワード列として音声認識した結果を使用するのが一般的である。サブワード単位の認識においても、単語単位の認識と同様に3種類の誤りが生じる。そのため、置換、挿入、脱落誤りを考慮した音節列どうしの近似マッチングに基づく検索方法が試みられている [1]。ドイツ語に対しては、5,000 個の音節を用いた音節どうしの重み付きレーベンシュタイン距離に基づく検索方法が提案されている (およそ、半分の単語が1音節語) [2]。中国語は音節数が416個で少ないため、検索の基本単位としてよく用いられる [3]。日本語の音節数は100余種類と比較的少なく扱いやすい。音節列として認識することによって、認識の際に単語辞書を使用しないので、文法の制約を無視でき、未知語の発音をそのまま認識できる可能性がある [4]。転置インデックスに基づく音素や音節の n-gram を用いた検索方法が種々提案されているが [5]、基本的には bag of words の使い方、n-gram の出現順序関係や音素認識誤りは考慮されていない [5], [6], [7]。そのため、文献 [8] は、高速検索の前処理として使用し、文献 [9] は高速性を犠牲にして、n-gram 間の距離の厳密化を行っている。

Katsurada らは音節列どうしの DP マッチングによる検索を高速化するために音節認識結果をサフィックスアレイとしてテーブル化しておき、検索時に置換、挿入、脱落誤りを許しながらテーブルを探索する方法を提案しているが [10]、音節ラティスへの適用は困難である。我々は、音節単位で認識した音節ラティスをサブワード列として用意しておき、音節の置換・挿入・脱落誤りを考慮した音節ラティスの n-gram とその出現順序を用いる検索手法を提案してきた [11]。

本論文では、我々が提案してきた音声ドキュメント内の未知語と大語彙連続音声認識による認識誤りの単語を高速で検索する手法 [12], [13], [14] の改善法を提案する。我々の手法は未知語を検索可能にするために音節単位での認識結果を用いている。認識した結果の音節ラティスからトライグラムを作成し、そのトライグラムを用いて転置インデックスを構築している。インデックスは辞書順にソートしておくことで高速な検索が可能となる [5], [6]。認識誤りへの対策としては、複数候補やダミー音節を用いて置換誤り、挿入誤り、脱落誤りに対処する手法を提案してきた [14], [15]。また、高速に検索を行うために認識誤り対策をどの程度行ったのかという情報を編集距離として導入し、標準的な編集距離に基づく音節列どうしの DP マッチングによる検索と同等以上の検索性能を実現した。しかし、距離の定義に改善の余地があった。本論文では挿入誤りと脱

落誤りに関する距離を編集距離からコンテキストを考慮した音節間の音響的類似度性を反映した距離に変更する方法を提案する。また、ダミー音節とクエリ中の音節との置換誤り距離を従来は固定値を用いていたのを、ダミー音節に対応する第1候補音節とクエリ中の音節との音節間距離に変更する方法を提案する。これにより、DP マッチングによる音節列どうしのマッチングと同様な距離が定義でき、検索性能を向上させることができる。しかも検索性能をあまり落とすことなくインデックスサイズと検索時間を小さくできることを示す。最後に提案法の有効性を評価実験で示す。

## 2. 認識誤り・未知語に頑健な高速検索手法

本章では、まず本論文のもととなる我々が提案してきた距離つき n-gram インデックスに基づく認識誤り・未知語に頑健な高速検索法 (従来法) について述べる。

### 2.1 n-gram に基づく未知語検索法 [11], [13]

本手法 (従来法と本論文の提案法との共通手法を本手法と呼ぶ) では、未知語に頑健な検索手法として、音節ラティス (コンフュージョンネットワーク) を使用して検索の際に認識誤りを考慮して検索を行う。まず、検索対象の音声ドキュメントに対して大語彙連続音声認識 (既知語検素用) と連続音節認識を行い、インデックス化する。

未知語を検索可能にするため、サブワード列として音節ラティスの上位 m ベストを使用する。音節ラティスのデータを保持させておくデータ構造として、主にテキスト検索で用いられる n-gram 転置インデックスを用いる。ここで用いる n-gram インデックスでは、音声ドキュメント内での出現位置情報と出現する n-gram の情報を保持させておく。n-gram インデックスの作成方法の概要を図 1 に示す。

この n-gram を辞書順に並べておけば、2分探索で高速に検索できる。同じ n-gram が複数個連続してインデックス表に格納されることがあるが、これに対しては種々の改良法がある。本手法では同じ n-gram を1つの n-gram に

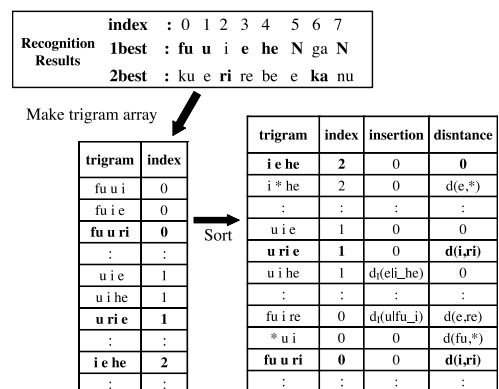


図 1 トライグラムアレイ作成手順

Fig. 1 Tri-gram array indexing procedure.

まとめ、これに対して複数のエントリ（位置情報など）を持つ索引構造としている。また、本論文では  $n=3$  としてトライグラムという固定長に限定しているため、トライグラムの種類とインデックスの位置が1対1に対応しているため、この関係を用いれば2分探索よりも高速に検索できる。4音節長以上の検索語に対しては、3音節の複数の組に分割し、それぞれで検索し、検索候補結果を連続性を考慮して、構成できるかどうかで、検索語の検索候補を出力する。たとえば、図1でクエリが“fu u ri e he N ka N”の場合、クエリはトライグラムに分割され、“fu u ri”、“e he N”、“N ka N”となる。ここで、最初のクエリトライグラム“fu u ri”が位置(index)0で検出された場合、次のクエリトライグラム“e he N”は位置が3のものを接合する。同様に、“e he N”は位置が3なので、“N ka N”の位置が5のものと接合する(2番目のトライグラムと3番目のトライグラムで“N”が重なっているため)。このようにそれぞれの検索結果の位置情報をマージアルゴリズムで比較し、接続を確認する。2つのトライグラムに分割して検索する場合、それぞれの検索結果の位置情報をマージアルゴリズムで比較し、接続を確認する。

## 2.2 認識誤りに関する対策 [11], [13]

### 2.2.1 置換誤り対策

置換誤り対策としては、文献[11], [13]に示されているように、 $m=5$ として音節ラティスの上位5ベストを用いる。本手法では索引を構築する際に音節ラティスの上位5ベストを組み合わせ、トライグラムを作成する(図1参照)。つまり、1つのインデックス位置に対して $5 \times 5 \times 5 = 125$ 個のトライグラム情報を持たせる。たとえば、「フーリエ変換」の1ベストの認識結果が「フエキエヘンカン」になっていたとしても、5ベスト中に正しい音節が含まれていれば検索することができる。第1候補以外のラティス中の音節は、置換誤り対策用の音節候補であり、置換誤りの距離は1ベストからの音節間距離(音節HMMの各状態のGMM間のバチャタリヤ(Bhattacharyya)距離)を使用した、1ベストのみから生成されたトライグラムを基準とし、置換誤り対策によって生成されたトライグラムの各音節との距離の合計を置換誤りの距離とする(図1参照)。もちろん、これ以外の距離の定義も種々考えられる。たとえば、クエリの音節がラティス中に存在すれば、完全マッチとして、距離を0にするという定義もありうる。

### 2.2.2 ダミー音節を含めた置換誤り対策 [14], [15]

置換誤り対策を行うことで、5ベストに含まれる誤りは対処することができる。しかし、それでも検出できないクエリが多く存在する。そこで、どんな音節にでもマッチするような音節をダミーとしてラティスに登録しておき、5ベストに含まれない場合でも検索できるようにした。たとえば、“ABC”といったトライグラムの場合、これに加え、

“AB\*”や、“A\*C”、“\*BC”などを加える。ここで“\*”はどんな音節にでもマッチするようなダミーとして扱う。ただし、3音節の内最大1音節のみ、もしくは2音節のダミー音節を許す。これにより、5ベストのトライグラムの組合せは125通りから200通り、もしくは215通りに増加する。

図1では、“i \* he”や“\* u i”などがダミーを含んだ例である。このときの距離 $d(e, *)$ と $d(fu, *)$ は等しく、比較的大きな固定値とする(従来法)。このダミーを含むトライグラムは、スキップバイグラム(1 distant bigram)に類似しているが、スキップした位置情報も使用する点が異なっている。これに対応して、クエリにもダミー音節を導入して検索する。そのため、クエリの検索回数は増加する。

### 2.2.3 挿入誤り対策

挿入誤りに関しては、索引構築時に1音節飛ばしたトライグラムを作成することで対処する(図1参照。3連続音節に対して1カ所)。したがって、1つのインデックス位置に対し、4つのトライグラムが索引に追加される。たとえば、音節列ABCDに対し、ABC, BCD, ACD, ABDを追加する。BCDを含め、これらのインデックス位置は同じである。挿入誤りを考慮したインデックスには固定値の距離を付随させておく(従来法)。

### 2.2.4 脱落誤り対策

脱落誤りに関しては上記2つの対策とは違い、検索時にクエリを数音節脱落させたものを含めて検索することで対処する。したがって、1つのクエリに対して複数回検索を行う。本実験では4音節以上のクエリに対しては1つの脱落、7音節以上のクエリに対しては2つの脱落を許すことをベースとする。ただし、脱落は連続する3音節以内に1カ所に制限する。脱落誤りを考慮したクエリのトライグラムには、固定値の距離を付随させておく(従来法)。

## 2.3 距離付き n-gram インデックスによる検索 [13]

前述した認識誤り対策を行うことで認識誤りに頑健な検索を行うことができる。しかし、同時に湧き出し誤りが増加するという問題がある。そこで、トライグラムインデックによる候補検出後にDPマッチングによる後処理による2段階検索法を以前提案した[11]。しかし、DPマッチングによる検出候補の検証では、検出候補すべてに対してDPマッチングを行わなければならないため、検索対象音声長が長くなったり、検出候補が増えたりすると処理時間が長くなるという問題があった。そこで、認識誤り対策を行う際にどれだけの誤りを許容したかという情報(距離)を用いて、検索時に検出候補の削減を行う方法を提案した[13]。距離を用いることでDPマッチングのように複雑な計算を行わずに、閾値との比較のみで検出候補の絞込みが高速に行える。

前述した置換誤り距離(ダミー音節距離を含む) $d_S$ 、挿入誤り距離 $d_I$ 、脱落誤り距離 $d_D$ の重みつき総和をクエリ

の長さ（クエリを分割したトライグラム数）で以下のように正規化し、閾値と比べて検出している。ここで、 $\alpha$ は置換距離、 $\beta$ は挿入距離、 $\gamma$ は脱落距離それぞれに対するペナルティの重みである。

$$\frac{\alpha \times \sum d_S + \beta \times \sum d_I + \gamma \times \sum d_D}{\text{number of trigram}} \quad (1)$$

同じ閾値では一般的に長いクエリは短いクエリに比べ検出しにくい。DP マッチングと異なり、索引に含まれないような誤りは閾値をどれだけ緩めても検出漏れとなってしまう。特に本研究の手法では誤り対策でカバーできなかった認識誤りは、検出漏れとなる。この問題を解決するために閾値の可変化手法を用いている。これは、クエリの長さ（音節数）に応じて閾値を緩くすることで、長いクエリに対する検出基準を弱める手法である。いい換えれば、短いクエリに比べ、長いクエリは閾値を緩めても信頼度良く検出できるからである。たとえば、音節数ごとに10%ずつ閾値を緩めていく場合、基準となる4音節のクエリの閾値を1.0とすると、5音節のクエリは1.1、6音節のクエリで1.2と閾値を可変にしていける。このようにすることで、閾値が比較的小さい箇所（適合率（Precision）が高く、再現率（Recall）が低い箇所）で長いクエリの検出が増加し、再現率（Recall）を高くすることができる。

### 3. 距離尺度の厳密化

本章では、我々が提案してきた従来法に対して、置換誤り距離、挿入誤り距離、脱落誤り距離の定義を厳密化し、検索精度の向上を図る方法を述べる。

#### 3.1 置換誤り距離の厳密化—ダミー音節とクエリ中の音節との距離の厳密化

従来法では、ダミー音節を使用したトライグラムは、ラティス中に存在しなかったというペナルティとして、固定値の距離を与えていた。たとえば、クエリ音節列がABCで、音節Bがラティスに認識候補として存在しておらず、その第1候補がDであった場合、つまり認識結果がADCであった場合を考えよう。このとき、A\*Cがインデックスとして登録されているため、このダミー音節とクエリ中の音節Bとの距離は固定値とはせず、 $d_S(B, *) = d_S(B, D)$ と定義する。これは、DP マッチングにおけるクエリ中の音節と認識結果の第1候補音節との距離の定義と同じである。すなわち、ダミー音節の導入と上述の距離の定義により通常のDP マッチングと同様な距離定義でラティスに存在しない音節を含むトライグラムが検索できることになる。ダミー音節に対応する第1候補の音節名はインデックスの位置情報から簡単に見い出せる。形式的に書けば、クエリの音節が $C_0V_0$  ( $C_0$ は子音、 $V_0$ は母音)、認識結果の第1候補が $C_1V_1$ であるとき、 $C_0V_0$ がラティス中に存在しないなら、

1best	ba
2best	da
3best	pa
4best	ka
5best	ta
dummy	*

図2 認識結果候補（音節ラティスの例）

Fig. 2 An example of recognition result (syllable lattice).

$$d_S(C_0V_0, *) = d_S(C_0V_0, C_1V_1) \times \lambda + \delta_S \quad (\lambda \geq 1.0) \quad (2)$$

ラティス中に存在するなら、

$$d_S(C_0V_0, *) = d_S(C_0V_0, C_1V_1) \times \eta \quad (\eta \leq 1.0) \quad (3)$$

と定義する。ペナルティを課するのは、第1～第5候補に入っている音節と比較して、ダミー音節は実際に発声された音節（第1候補に類似な音節と考えられる）との距離が大きいと仮定できるからである。なお、ダミー音節を利用する場合のペナルティ $\lambda$ と $\delta_S$ の付け方は種々考えられ、実験により評価する。

認識結果が図2の場合を考えよう。もし、クエリの音節が/ga/であれば、これは認識結果の候補（ラティス）に存在しないので、ダミー音節を採用し、置換距離は式(2)でペナルティを課さず $\lambda = 1$ 、 $\delta_S = 0$ なら、 $d_S(ga, *) = d_S(ga, ba)$ となる。もし、クエリの音節が/da/の場合のように、ラティスに存在する場合は、置換距離は、 $d_S(da, ba)$ となる(2.2.1項参照)。結局、クエリ中の音節が認識結果の候補に存在していても存在していなくても、同じ距離尺度となる（ペナルティ項を無視した場合）。この矛盾を防ぐために従来法はダミー音節を使用する場合は、大きめの固定値を置換距離としていた。一方、提案法では $\lambda$ と $\delta_S$ でペナルティを課した。これにより、クエリの音節が/so/のような場合は、ダミー音節との距離は固定値よりも大きな距離になり、過剰検出を防ぐことができる。

また、ダミー音節を採用したトライグラムは、正解音節に対応する認識候補はなくても検索できるようになるので、認識候補は不要になる。若干異なってくる点は、第2候補～第5候補に入っていた音節でもペナルティが課されることである。このように、ダミー音節を採用する場合は、対応する第1～第5候補を無視（第1候補はダミーの位置から同定できる）すれば、インデックスサイズを大幅に削減できる。たとえば、トライグラムの3音節のうち、ダミー音節を1個採用する場合は、インデックスとして登録する必要なトライグラム数は、25種類となる。したがって、各位置で登録する必要なトライグラム数は125通りから $25 \times 3 = 75$ 通りに減少する。さらに、ダミー音節を2個まで許すとすると、各位置で登録する必要なトライグラム数は、 $5 \times 3 = 15$ 通りとなり、大幅なインデックスの削減が達成できる。これは、重複する検索結果の削減に

もなり、検索時間の短縮にもつながる。なお、この場合、認識結果のトライグラム 3 音節のうち、少なくとも 1 音節は正しく認識されてラティス中に存在しなくてはならない。このため、音節認識結果は、1 ベストの結果だけでなく  $n$  ベストの結果まで出力するラティス表現が望ましい。たとえば、第 1 候補の音節正解率が 84%、第 5 候補以内に含まれる音節正解率が 91% の場合、ダミー音節をトライグラム中の 2 音節を使用する場合のカバー率（検索に対応できる割合）は、前者の第 1 候補だけを使用する場合は 99.6% ( $= 1.0 - (1.0 - 0.84)^3$ )、後者の第 5 候補までを使用する場合は 99.9% ( $= 1.0 - (1.0 - 0.91)^3$ ) となる。一方、トライグラム中でダミー音節を 1 音節に制限した場合のカバー率は、前者で 93.1%、後方で 97.7% となる。

### 3.2 挿入誤り距離の厳密化 [16]

従来法では挿入誤り対策や脱落誤り対策を行ったトライグラムに対し、ペナルティとして固定値を付随させていた。しかし、固定値では距離の定義が正確にできておらず、検索を厳密に行うことが不可能である。そこで、提案法として挿入距離についての厳密化を行う。たとえば音節列の認識結果が、 $C_1V_1 C_2V_2 C_3V_3$  ( $C$  は子音、 $V$  は母音) であるとき、 $C_2V_2$  を挿入誤りと仮定したときの挿入誤りの距離を以下のように定義する。

$$d_I(Z|X\_Y) \tag{4}$$

$$= \min \{d(C_2V_2, C_1V_1), d(C_2V_2, V_1), d(C_2V_2, C_3V_3)\}$$

$$+ \delta_I$$

ここで、 $Z = C_2V_2$ 、 $X = C_1V_1$ 、 $Y = C_3V_3$ 、挿入音節と仮定した  $Z$  の左コンテキストが音節  $X$ 、右コンテキストが音節  $Y$  であることを示している。右辺の第 2 項は左コンテキスト音節の母音部が分かれて  $Z$  として挿入誤りになった場合に相当する。すなわち、音節の挿入誤りは、 $Z$  の左右のどちらかの音節が誤って分かれて認識されたものと考え、特に、左の音節の母音部が分かれて誤認識されたと解釈する。たとえば、 $/sa/$  が  $/sa ga/$  となり  $/ga/$  が挿入されたと仮定する場合、直前の音節が  $/sa/$  であった場合、 $/sa/$  が分かれて  $/sa ga/$  となったというよりも母音部の  $/a/$  が分かれて  $/a ga/$  となったと解釈するのが妥当と考えた距離の定義になっている。ただし、 $/sa/$  が分かれる場合と  $/a/$  の部分に分かれる場合の 2 通りの場合があることを想定した。 $\delta_I$  は挿入のペナルティであり、今回は  $\delta_I = 0$  としている。なお、従来法は  $d_I(Z|X\_Y) = 1$  としていたことに相当する。

### 3.3 脱落誤り距離の厳密化 [16]

クエリの音節列  $XZY$  の  $Z$  を音声認識で脱落誤りと仮定した場合の脱落距離も挿入誤り距離と同様に定義する。

$$d_D(Z|X\_Y) \tag{5}$$

$$= \min \{d(C_2V_2, C_1V_1), d(C_2V_2, V_1), d(C_2V_2, C_3V_3)\}$$

$$+ \delta_D$$

$\delta_D$  は脱落のペナルティであり、今回は  $\delta_D = 0$  としている。すなわち、音節の脱落誤りは、この左右のいずれかの音節と音響的に類似しているため誤って 1 音節に誤認識されたと解釈する。特に左の音節の母音部に併合されて誤認識されたと解釈する。たとえば、 $/sa ra da/$  が、 $/sa da/$  と誤認識された場合は、特に  $/sa/$  の母音部と  $/ra/$  が併合して  $/sa/$  となったと解釈するのが妥当と考えた距離の定義になっている。ただし、 $/ra/$  が  $/sa/$  と併合した場合と  $/sa/$  の母音部の  $/a/$  と併合した場合の 2 通りがあることを想定した。なお、従来法では  $d_D(Z|X\_Y) = 1$  としていたことに相当する。

### 3.4 提案法と DP マッチング法の類似点と相違点

提案法と DP マッチングとの類似点として、本論文で提案した距離の厳密化法によって、両方とも置換、脱落、挿入距離（コスト）を任意に設定することができる点があげられる。ただし、本提案法では、クエリ中の音節が認識結果のラティス中に存在しない場合は、クエリとのマッチングができなくなるのでダミー音節を導入しているが、トライグラムの 3 音節ともダミー音節というインデックスは認めていない（音節トライグラムの意味をなさなくなるため）。このため、認識結果のトライグラム 3 音節のうち、少なくとも 1 音節は正しく認識されてラティス中に存在しないとならないことが DP マッチング法とは異なる。一方、文献 [8] のサフィックスアレイによる DP マッチングの高速化法は、認識結果がラティス（コンフュージョンネットワーク）表現になっている場合は、適用が難しい。

マッチングの際の相違点として、DP は非線形的なマッチングを許しており、通常は  $n$  音節に対して  $n$  個の挿入と  $n/2$  個の脱落を許すため（傾き  $1/2 \sim 2$  の DP パス）、挿入と脱落の制限が弱い。そのため再現率は高くなるが適合率は低くなる。一方、提案法は挿入、脱落は 3 音節あたり 1 音節に制限しているので、再現率は減少するが、適合率が向上すると考えられる。

### 3.5 トライグラムの DP マッチングによる索引化との相違点

提案法は、DP マッチング法と比べ、ほぼ同等以上の検索性能を示すが、インデックス生成の制約が強く、必要なインデックスの漏れが生じ、再現率を高くできないという欠点がある。そこで認識結果のラティス表現に対して、DP マッチングを高速に実現する方法として、クエリに現れる全音節のトライグラムと認識結果の音節列の DP マッチングを行っておき、そのマッチング位置と DP 距離を索引

化する手法が考えられる [9], [17]. すべてのトライグラムを保存するのは必要メモリが膨大になるため難しいが, 距離が閾値よりも小さいものだけを登録することで実現可能なサイズへと調整することができる [17].

認識システムで使用した音節トライグラム言語モデルの登録数は約 21 万であり, 提案法により 44 時間の索引に登録されているトライグラムの種類は約 82 万となっている. そのため, 可能なトライグラム音節数  $116^3 = 1,560,896$  通りのうち, 100 万通り程度を索引付けしておくことは非現実的でない. また, 距離はあらかじめ計算することができるため, 検索の絞り込みの際にも高速に行うことができるという利点がある. ただし, 事前に行う DP マッチングの計算量が膨大となるという問題が生じる. さらに, DP による検索区間が限定され, 隣接トライグラムとのマージングが精密にできないため, 検出された候補区間に対して, クエリと DP マッチングをやり直す必要がある [17]. これは, 高速化という点で大きな欠点となる.

#### 4. 評価実験

##### 4.1 実験データ

実験データには CSJ (日本語話し言葉コーパス) [18] のコアデータ 44 時間分を用い, 本研究室で開発された SPOJUS++ [19] による音節認識結果を対象とし, 検索, 評価を行った. 大語彙連続音声認識 (LVCSR) の辞書 (約 28,000 語) には, コア講演以外の CSJ2702 講演を学習データとし, カットオフを 4 とした. したがって, 出現回数が 4 回以上あるものは未知語にはならない. 今回検索語セットは NTCIR9 から提供されているコア講演用のドライランテストセット [20] を使用した. すなわち, 既知語 50 個 (総出現数は 766 語), 未知語 50 個 (総出現数は 239 個) を使用した. 連続音節認識には音節の 4 グラムの言語モデルを用いた. 連続音節認識による音節認識率を表 1 に示す. 連続音節認識結果の第 5 候補までを考慮すると音節認識結果の正解率は 91% とかなり高い. なお, ベースラインである従来法は, 断らない限り, 挿入誤り距離 ( $d_I$ ) と脱落誤り距離 ( $d_D$ ) は固定値 1, 式 (1) で  $\alpha = 2$ ,  $\beta = 3$ ,  $\gamma = 4$  とした結果を示す. 評価尺度としては以下のものを用いた [20].

- 適合率 =  $\frac{\text{正しく検出された数}}{\text{検出された検索語数}}$
- 再現率 =  $\frac{\text{正しく検出された数}}{\text{正しく検出されるべき検索語数}}$
- F 値 =  $\frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$
- MAP = 各クエリの検出数ごとの適合率の平均の全クエリにわたる平均

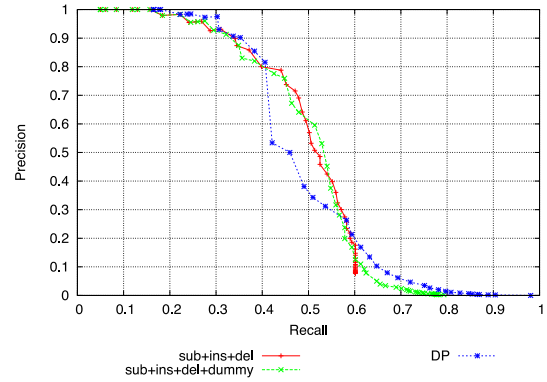
##### 4.2 ダミーを含めた索引を用いた検索—従来法 [15]

置換誤り対策を行うことで 5 ベストに含まれる置換誤り

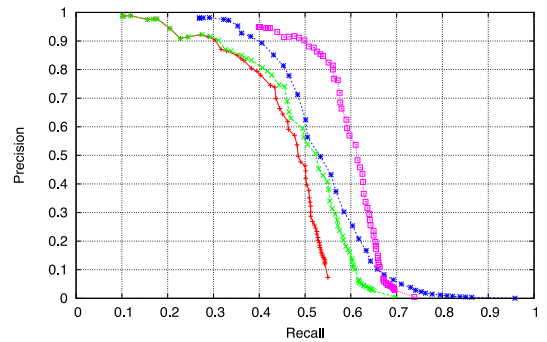
表 1 SPOJUS++ による音節認識率 (%)

Table 1 Syllable recognition rates by SPOJUS++.

output	Del	Ins	Subs	Corr	Acc
音節 (1best)	3.9	3.6	12.5	83.6	80.0
音節 (3best)	3.9	2.2	6.9	89.1	86.9
音節 (5best)	4.1	1.9	4.9	91.0	89.1



(a) 未知語



(b) 既知語

図 3 ダミーを含めた索引からの検索結果—従来法

Fig. 3 Retrieval result based on indices after introducing dummy syllables (previous method).

にはすべて対処することができる. また, 挿入誤り対策と脱落誤り対策を組み合わせる (distant 3gram に相当) ことで 5 ベストに含まれない置換誤りにも対処することができる. しかし, それでも検出することができない検索語が多数存在する. この問題を解決するために 2.2.2 項で述べたダミーの音節を含めた索引を構築し, 検索実験を行った. 本評価実験ではトライグラムの 3 音節中, 1 音節のみダミー音節を許した. ダミー音節を使用したときの置換距離は固定値 4.5 を用いた (すなわち, 式 (2) で,  $\lambda = 0$ ,  $\delta_s = 4.5$ ). 未知語の結果を図 3 の (a) と表 2 の (a), 既知語の結果を図 3 の (b) と表 2 の (b) に示す. 表の “trigram” は, 置換・挿入・脱落誤りを考慮したトライグラムインデックスを用いる方法を示している. また, “+dummy” はこの手法にダミー音節を導入したインデックスを用いる方法を示している. この結果からダミーを入れることで F 値はほぼ同じ程度となるが, MAP は大きく向上している. また,

表 3 挿入と脱落距離の厳密化を行った未知語検索結果—提案法

Table 3 Retrieval results for out-of-vocabulary words after introducing strict distance measure of insertion and deletion errors: proposed method.

尺度	手法	音節認識 (5 ベスト+ダミー)						(a'')+ (b)
		挿入距離の厳密化			脱落距離の厳密化			
		(a) 母音のみ考慮	(a') 母音も考慮	(a'') 考慮なし	(b) 母音のみ考慮	(b') 母音も考慮	(b'') 考慮なし	
再現率		0.444	0.444	0.453	0.487	0.491	0.517	0.479
適合率		0.812	0.806	0.791	0.735	0.723	0.672	0.812
F 値		0.575	0.573	0.576	0.586	0.585	0.585	0.602
MAP		0.575	0.581	0.583	0.622	0.624	0.618	0.609

表 4 挿入と脱落距離の厳密化を行った既知語検索結果—提案法

Table 4 Retrieval results for in-vocabulary words after introducing strict distance measure of insertion and deletion errors: proposed method.

尺度	手法	音節認識 (5 ベスト+ダミー)						(a'')+ (b)
		挿入距離の厳密化			脱落距離の厳密化			
		(a) 母音のみ考慮	(a') 母音も考慮	(a'') 考慮なし	(b) 母音のみ考慮	(b') 母音も考慮	(b'') 考慮なし	
再現率		0.445	0.447	0.442	0.408	0.416	0.449	0.471
適合率		0.727	0.721	0.751	0.837	0.798	0.670	0.736
F 値		0.552	0.552	0.557	0.549	0.547	0.537	0.574
MAP		0.513	0.509	0.514	0.494	0.501	0.490	0.520

表 2 ダミーを含めた検索結果—従来法

Table 2 Retrieval results after introducing dummy syllables (previous method).

(a) 未知語			
尺度	手法	音節認識 (5 ベスト+ダミー)	
		trigram	+dummy
再現率		0.452	0.448
適合率		0.701	0.759
F 値		0.550	0.563
MAP		0.486	0.561
メモリサイズ		750 MB	1,309 MB

(b) 既知語				
尺度	手法	音節認識 (5 ベスト+ダミー)		
		trigram	+dummy	+LVCSR
再現率		0.443	0.439	0.557
適合率		0.757	0.718	0.803
F 値		0.551	0.545	0.658
MAP		0.414	0.512	0.611

検出できる検索語が増加し、再現率の最大値が増加していることが分かる (図 3 参照)。これは検索語の検出が増えると同時に誤検出も増えたためだと考えられる。メモリサイズに関してはダミーを導入することで増加し、約 1.3GB 必要となる。しかし、メモリの削減手法を用いることで性能の低下なしに 680 MB まで削減することができる [15]。

### 4.3 挿入誤り距離と脱落誤り距離の厳密化

挿入誤りと脱落誤りの距離を 0 もしくは 1 の 2 値の従来法からバタチャリヤ距離を用いて距離の厳密化を行った。その結果を表 3 と表 4, 図 4 の (a) と (b) に示す。このとき、式 (1) で  $\alpha = 2, \beta = 3, \gamma = 3$  としている。表 3 と

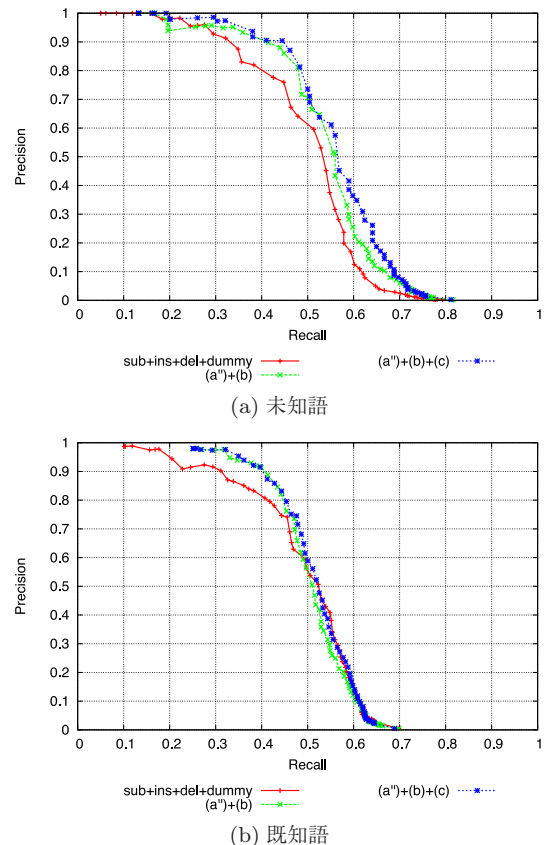


図 4 距離を厳密化した検索結果—提案法

(a'') : 挿入誤り距離の厳密化 (母音を考慮しない), (b) : 脱落誤り距離の厳密化 (母音のみを考慮), (c) : 置換誤りの厳密化 ( $\lambda = 1.5/\eta = 1.0/\delta_S = 2.0$ )

Fig. 4 Retrieval result based on indices after introducing strict distance measure (proposed method).

表 5 種々の置換距離の厳密化による未知語検索結果—提案法

Table 5 Retrieval results for out-of-vocabulary words after introducing strict distance measure of substitution errors: proposed method.

(a)  $\delta_S = 0.0$

手法 \ 尺度	音節認識 (5 ベスト+ダミー) 左: $\lambda$ /右: $\eta$						
	1.0/1.0	1.5/1.0	2.0/1.0	2.5/1.0	1.0/0.5	1.5/0.5	2.0/0.5
再現率	0.491	0.449	0.483	0.462	0.385	0.496	0.543
適合率	0.550	0.820	0.753	0.812	0.826	0.699	0.641
F 値	0.519	0.580	0.589	0.589	0.525	0.580	0.588
MAP	0.611	0.626	0.642	0.646	0.608	0.637	0.647

(b)  $\delta_S = 2.0$

手法 \ 尺度	音節認識 (5 ベスト+ダミー) 左: $\lambda$ /右: $\eta$						
	1.0/1.0	1.5/1.0	2.0/1.0	2.5/1.0	1.0/0.5	1.5/0.5	2.0/0.5
再現率	0.483	0.483	0.483	0.483	0.463	0.483	0.530
適合率	0.813	0.813	0.813	0.813	0.790	0.796	0.697
F 値	0.606	0.606	0.606	0.606	0.600	0.601	0.602
MAP	0.646	0.658	0.656	0.646	0.662	0.656	0.654

表 6 種々の置換距離の厳密化による既知語検索結果—提案法

Table 6 Retrieval results for in-vocabulary words after introducing strict distance measure of substitution errors: proposed method.

(a)  $\delta_S = 0.0$

手法 \ 尺度	音節認識 (5 ベスト+ダミー) 左: $\lambda$ /右: $\eta$						
	1.0/1.0	1.5/1.0	2.0/1.0	2.5/1.0	1.0/0.5	1.5/0.5	2.0/0.5
再現率	0.453	0.456	0.482	0.450	0.420	0.492	0.484
適合率	0.786	0.767	0.701	0.783	0.835	0.664	0.678
F 値	0.574	0.572	0.572	0.571	0.559	0.565	0.565
MAP	0.529	0.534	0.534	0.536	0.531	0.537	0.539

(b)  $\delta_S = 2.0$

手法 \ 尺度	音節認識 (5 ベスト+ダミー) 左: $\lambda$ /右: $\eta$						
	1.0/1.0	1.5/1.0	2.0/1.0	2.5/1.0	1.0/0.5	1.5/0.5	2.0/0.5
再現率	0.476	0.476	0.476	0.472	0.477	0.476	0.472
適合率	0.744	0.745	0.745	0.750	0.724	0.725	0.727
F 値	0.580	0.581	0.581	0.580	0.575	0.575	0.573
MAP	0.533	0.534	0.536	0.534	0.540	0.537	0.536

表 4 の“(a)”,“(b)”はそれぞれ挿入,脱落の距離を式(4),式(5)のバタチャリヤ距離で定義し( $\delta_I = \delta_D = 0$ ),音節の左コンテキストの母音に関してのみ考慮したものである。“(a)+(b)”は挿入と脱落を同時に仮定したものである。“(a’)”,“(b’)”は挿入,脱落誤りの母音の分割も考慮した場合である。“(a’’)”,“(b’’)”は母音の分割を考慮しなかった場合である。挿入距離の厳密化は未知語と既知語に対して改善の効果があつた。組合せでF値が最も高かつたのは挿入,脱落の両方の距離を厳密化し,脱落に関しては左コンテキストの母音の分割だけを考慮した場合であつた。未知語と既知語に対して,F値とMAPともに改善できた。

#### 4.4 ダミー音節とクエリ中の音節との距離の厳密化

挿入誤り距離と脱落誤り距離の厳密化は,表3の右端欄の(a’)+(b)の条件に固定し,クエリ中の音節とダミー

音節(およびラティス中の音節)との置換誤り距離の定義を種々変えて行った未知語と既知語の検索結果を表5と表6および図4に示す。表中の $\lambda$ は,クエリ中の音節とダミー音節の厳密化距離(式(2)参照)を $\lambda$ 倍して用いることを示す。また, $\eta$ は,クエリ中の音節とラティス中の音節との距離を $\eta$ 倍して用いることを示す(式(3)参照)。つまり表5(a)と表6(a)は式(2)の $\delta_S$ を0.0に,表5(b)と表6(b)は式(2)の $\delta_S$ を2.0にしたときの結果である。

まず表5(b)と表6(b)から,クエリ音節とダミー音節との距離を厳密化( $\lambda/\eta = 1.0/1.0$ )することにより,検索性能が向上していることが分かる(表3と表4の最右欄との比較)。さらに,ラティス中に含まれていない音節をダミー音節として扱う場合は,距離を大きめの固定値に設定していた従来法と同様に,ペナルティを乗じることになる1.5倍とか2倍にすることにより,MAPの性能が少し向上



表 7 脱落数による未知語検索結果 ( $\theta_1$  以上に対して脱落 1 つ,  $\theta_2$  以上に対して脱落 2 つ)

Table 7 Retrieval results for out-of-vocabulary words in accordance with the permissible number of deletion errors.

尺度	手法	音節認識 (5 ベスト+ダミー) 左: $\theta_1$ /右: $\theta_2$			
		4/4	4/7	5/7	6/7
再現率		0.483	0.483	0.483	0.474
適合率		0.807	0.813	0.813	0.874
F 値		0.604	0.606	0.606	0.615
MAP		0.650	0.658	0.658	0.644

している. ベースラインである DP 法と比べて, 未知語に対しては F 値と MAP で大幅に上回る性能が得られた. 一方, 既知語に対しては F 値は DP 法と同等の性能が得られたが, MAP に対しては DP 法よりもまだ悪い. 図 4(b) から分かるように既知語に対してはまだ再現率が向上していないためであり, 今後この原因を改善していく必要がある.

#### 4.5 脱落誤り数制約条件の比較

今までの検索評価実験では, 音声認識の挿入誤り数は, 3 音節に 1 カ所以内, 脱落誤り数は, クエリ語の音節長が 3 音節未満なら 0 個, 4 音節~6 音節なら 1 個以内, 7 音節以上なら 2 個以内と仮定して行った. そこで, この条件を変えた評価結果表 7 をに示す. 表中の  $\theta_1$  は, クエリ語の音節長が  $\theta_1$  以上なら 1 個の脱落誤りを許容することを示す. また,  $\theta_2$  は, クエリ語の音節長が  $\theta_2$  以上なら 2 個の脱落誤りを許容することを示す. 表から, 脱落誤り数を許容しすぎる ( $\theta_2 = 4$ ) と, 検索性能が劣化し, むしろ厳しくした条件  $\theta_1=6, \theta_2=7$  (5 音節以下のクエリ語は脱落誤りなし, 6 音節のクエリ語は脱落誤り数は 1 個以内, 7 音節以上のクエリ語の脱落誤り数は, 2 個以内) で F 値の性能が良くなり, 0.615 となった. 一方, MAP は 0.644 と悪くなった. 既知語に対しては条件を厳しくすると F 値, MAP とともに悪くなった. このことから  $\theta_1 = 4, \theta_2 = 7$  が適切であるといえる.

#### 4.6 検索時間

従来手法である n-gram インデックスからの検索とベースラインの DP マッチングを比較した. その結果を図 5 に示す. 44 時間の音声ドキュメントを検索する場合, DP マッチングの平均検索時間が 675 [ms] なのに対し, ダミー音節を用いない従来の n-gram インデックスを用いた場合は平均 1 [ms] となった. 検索対象音声の時間長が大きくなったとしても, 索引からトライグラムを検索する時間は対数スケールの増加で済む. しかし, クエリを分割して検索した場合, それぞれのトライグラムの接続を考慮するマージ処理が必要となる. この処理が線形で増加するため, 全体的な検索時間は線形で増加する. しかし, ダミー音節を使用する場合でも 1,000 時間の音声に対して 300 ms

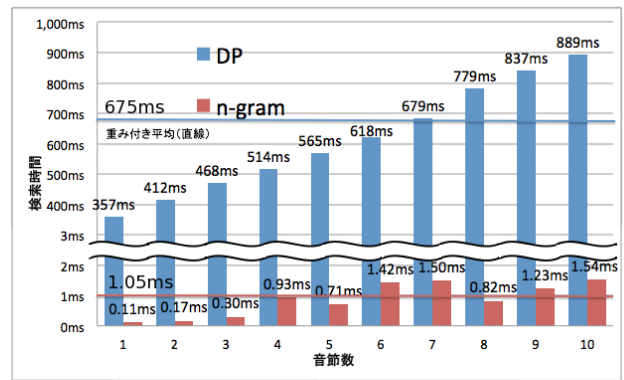


図 5 1 クエリあたりの検索時間の比較 (DP マッチングとダミー音節を使用しない従来手法) (音声ドキュメント量: 44 時間)

Fig. 5 Comparison of retrieval times per query (DTW method and previous method before introducing dummy syllables): amount of spoken documents = 44 hours.

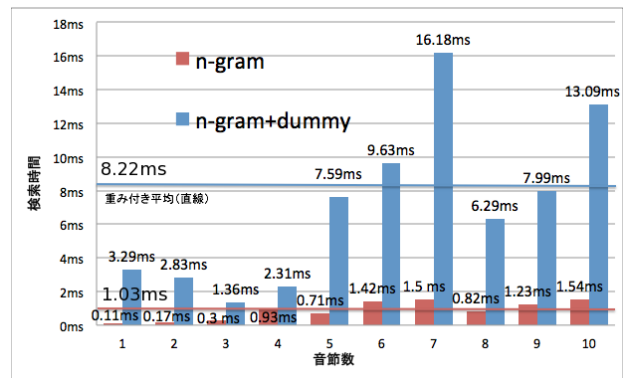


図 6 1 クエリあたりの検索時間の比較 (ダミーを導入した提案法) (音声ドキュメント量: 44 時間)

Fig. 6 Comparison of retrieval times per query (proposed method after introducing dummy syllables): amount of spoken documents = 44 hours.

程度で検索できると考えている. また, ダミー音節を含めた提案法と従来法を比較した結果を図 6 に示す. ダミー音節を含めた場合の検索はアルゴリズムは最適化されておらず, n-gram インデックスを用いた場合よりも 8 倍程遅い平均 8 [ms] であった.

なお, 3.1 節でも述べた無用なダミーつき n グラムインデックスの削減法により, 60%以上のインデックスの削減と 30%程度の検索時間の高速化ができる見通しである. しかしながら, 現実装方法でも, DP マッチングより約 80 倍高速である.

#### 5. おわりに

本論文では, 我々が従来提案してきた距離つき n-gram 索引による高速検索法に対し, 置換誤り距離, 挿入誤り距離, 脱落誤り距離の厳密化法を提案し, 検索実験により評価した. 挿入距離と脱落距離を厳密に定義し直したところ, 未知語検索に対し, F 値で従来法の 0.563 から 0.602

まで改善することができた。これに、ダミー音節とクエリ中の音節との置換距離を厳密化などの条件を変えることにより、さらに F 値は、0.606 まで向上した。ベースラインの DP マッチング法 (5 ベストの音節ラティス利用) の F 値 0.542 よりも相当性能が良い。44 時間の音声ドキュメントから 1 タームあたり約 8 ms で検索することができ、DP マッチング法と比べて、約 80 倍の高速化を実現した。

本検索法の検索性能に影響を与えるインデックス作成の制約や検索条件を列挙すると、以下のものがあり、今後検討する必要がある。適合率を上げたいのか、再現率を上げたいのかによって、検索条件などを適応させる必要がある。

- 挿入誤りは 3 音節あたり 1 個に限定している (これにより適合率は高いが、高い再現率が得られない)。
- 脱落誤りはクエリの長さ依存し、6 音節以下は 1 個の脱落、7 音節以上は 2 個の脱落を許容している (これにより適合率は高いが、再現率を高くできない)。
- 音節ラティス中に現れない正解音節 (ダミー音節で対処) は 3 音節に 1 個または 1~2 個に限定している (本実験では 1 個のみ)。いい換えれば、3 音節のうち、少なくとも 1 音節はラティス中に正しく認識されていないと、正しく検出できない (本論文の実験では 3 音節のうち 2 音節以上がラティスに含まれていないと正しく検索できない)。
- 長い検索語は複数個のトライグラムに分割して検索するが、4 分割以上された場合のみ、1 つのトライグラムが検出されなくてもよいと限定している (まだ制約が強い可能性がある)。

その他の今後の課題としては、トライグラムインデックス中のダミー音節を 2 つまで許容する実装と評価および 3.1 節で述べたようにダミー音節を用いた際のトライグラムインデックス登録の削減による検索アルゴリズムの高速化と省メモリ化の手法の実装と評価が考えられる。これにより、検索性能を落とすことなく、検索速度の向上とインデックスサイズの削減が可能と考えている。また、クエリをトライグラム列に分割する際、重複する音節が生じる場合があり、この場合はマージング処理が複雑になっている。たとえば、クエリが ABCDE の場合、ABC と CDE の 2 つのトライグラムに分割すれば C が重複しているが、これを ABC と DE のようにトライグラムとバイグラムに分割するとマージング処理が単純化・厳密化でき、インデックスサイズは増加するが、検索性能の向上につながると考えられる [21]。検索精度の向上のために、コンテキスト情報を用いて WEB や言語モデルの情報から検索結果の検証を行うことなども今後の課題である。

## 参考文献

[1] Wechsler, M., Munteanu, E. and Schauble, P.: New techniques for open-vocabulary spoken document retrieval,

- SIGIR*, pp.20-27 (1998).
- [2] Larson, M. and Eickeler, S.: Using syllable-based indexing features and language models to improve German spoken document retrieval, *EuroSpeech*, pp.1217-1220 (2003).
- [3] Wang, H.: Experiments in syllable-based retrieval of broadcast news speech in Mandarin Chinese, *Speech Communication*, Vol.32, pp.49-60 (2000).
- [4] Ng, K.: Towards robust methods for speech document retrieval, *ICSLP*, pp.1088-1091 (1998).
- [5] Chen, B., Wang, H. and Lee, L.: Retrieval of broadcast news speech in Mandarin Chinese collected in Taiwan using syllable-level statistical characteristics, *ICASSP*, pp.2985-2988 (2000).
- [6] Allauzen, C., Mohri, M. and Saraclar, M.: General indexing of weighted automata-application to spoken utterance retrieval, *Workshop on Interdisciplinary Approaches to Speech Indexing and Retrieval*, pp.33-40 (2004).
- [7] Ng, C., Wilkinson, R. and Zobel, J.: Experiments in spoken document retrieval using phoneme n-grams, *Speech Communication*, Vol.32, pp.61-77 (2000).
- [8] 神田直之, 住吉貴志, 小窪浩明, 佐川浩彦, 大淵康成: 多段リスコアリングに基づく大規模音声中の任意検索語検出, 電子情報通信学会論文誌, Vol.J95-D, No.4, pp.969-981 (2012).
- [9] Chaudhari, U.V. and Picheny, M.: Matching criteria for vocabulary-independent search, *IEEE Trans. Audio, Speech and Language Process*, Vol.20, No.5, pp.1633-1642 (2012).
- [10] Katsurada, K. Teshima, S. and Nitta, T.: Fast keyword detection using suffix array, *Interspeech*, pp.2147-2150 (2009).
- [11] 中川聖一, 高橋将史: 未知語に頑健な音声ドキュメント検索手法の検討, 第 3 回音声ドキュメント処理ワークショップ論文集, pp.7-14 (2009).
- [12] Iwami, K., Fujii, Y., Yamamoto, K. and Nakagawa, S.: Out-of-vocabulary term detection by n-gram array with distance from continuous syllable recognition results, *SLT*, pp.200-205 (2010).
- [13] Iwami, K., Fujii, Y., Yamamoto, K. and Nakagawa, S.: Efficient out-of-vocabulary term detection by n-gram array indices with distance from a syllable lattice, *ICASSP 2011* (2011).
- [14] Nakagawa, S., Iwami, K., Fujii, Y. and Sekiguchi, Y.: A robust/fast spoken term detection method based on a syllable n-gram index with a distance metric, *Speech Communication* (to appear, 2013).
- [15] 岩見圭祐, 山本一公, 中川聖一: 音声ドキュメント検索のための音節ラティスの拡張と n-gram 索引の削減手法, 第 89 回音声言語情報処理研究会 (2011).
- [16] 岩見圭祐, 山本一公, 中川聖一: 複数音声認識システムを併用した音節 n-gram 索引による検索性能の改善, 第 6 回音声ドキュメント処理ワークショップ論文集, SDPWS2012-10 (2012).
- [17] 斉藤裕之, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李 時旭: 複数音節の事前検索結果を利用した音声中の検索語検出の高速化, 第 6 回音声ドキュメント処理ワークショップ論文集, SDPWS2012-6 (2012).
- [18] 前川喜久雄: 日本語話し言葉コーパス, 外観, 入手先 (<http://www.kokken.go.jp/katsudo/seika/corpus/releaseinfo/040/overview.html>).
- [19] Fujii, Y., Yamamoto, K. and Nakagawa, S.: Large vocabulary speech recognition system: SPOJUS++, *MUSP*, pp.110-118 (2011).

- [20] Akiba, T., Nishizaki, H., Aikawa, K., Kawahara, T. and Matsui, T.: Overview of the IR for spoken documents task in NTCIR-9 workshop, *Proc. 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access* (2011).
- [21] 坂本 渚, 山本一公, 中川聖一: 距離つき音節トライグラムを用いた音節入力による音声ドキュメントの検索語検出法の評価, 日本音響学会春季講演論文集 (2013).



岩見 圭祐

1987年生。2010年豊橋技術科学大学工学部卒業。2012年同大学大学院修士課程情報工学専攻修了。同年パナソニックアドバンステクノロジー株式会社入社。



坂本 渚

1990年生。2011年豊橋科学技術大学情報工学課程入学。音声検索語検出に関する研究に従事。



中川 聖一 (フェロー)

1976年京都大学大学院博士課程修了。同年京都大学情報工学科助手。1980年豊橋技術科学大学情報工学系講師。1990年同大学教授。1985~1986年カーネギメロン大学客員研究員。音声情報処理, 自然言語処理, 人工知能の研究に従事。工学博士。1977年電子通信学会論文賞, 1988年IETE最優秀論文賞, 2001年電子情報通信学会論文賞, 各受賞。電子情報通信学会フェロー。著書『確率モデルによる音声認識』(電子情報通信学会編), 『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), 『Spoken Language Systems』(編著, IOS Press)等。