Vol.2011-SLP-89 No.5
2011/12/19

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

岩見圭祐[†1]　山本一公[†1]　中川聖一[†1]

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．その理由として挙げられるのが，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ユニットの認識結果を用いた高速な STD 手法を提案してきた．音節ラティスから置換，挿入を考慮した 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る．本稿では特にダミー音節を用いた検索性能の改善と索引サイズの削減に焦点を当てる．

Using Augmented Syllable lattice and Reduction of the n-gram Index for Spoken Term Detection

Keisuke Iwami,[†1] Kazumasa Yamamoto[†1]
and Seiichi Nakagawa[†1]

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 is not currently the established, because spoken document has specific problems such as some recognition errors and out-of-vocabulary(OOV) terms. 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 sequence of syllables in a recognized syllable-based lattice. Specially, in this paper, we introduced a dummy syllable symbol for attacking the substitution errors and the index reduction.

1. はじめに

本稿では，音声ドキュメント内の未知語と認識誤りの単語を高速で検索する手法[5)6)] 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．認識した結果から音節ラティスのトライグラムを作成し，そのトライグラムを用いて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．[2)1)] 認識誤りへの対策としては，今までにも複数候補を用いて置換誤り，挿入誤り，脱落誤りに対して対処する手法を提案している[6)]．また，高速に検索を行うために認識誤り対策をどの程度おこなったのかという情報を距離として導入している．本稿では，置換誤り対策で対処できないような誤りをダミーの音節を索引に取り入れることで解決した．索引サイズの問題に対しては索引の構造をよりコンパクトにし，さらに索引の構築時の誤り対策に制限を設けることで索引のサイズを 80% 程度削減した．また，既知語に対しては大語彙認識システム (LVCSR) の結果と併用することで検索精度を向上させた．

2. 音声ドキュメント検索手法

2.1 大語彙連続音声認識の利用

従来の音声ドキュメント検索手法として，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ある．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，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．

2.2 サブワード単位認識の利用

未知語に対しては，サブワード列として音節単位で認識した結果を使用する．音節単位の認識のおいても，単語単位の認識と同様に 3 種類の誤りが生じる．ドイツ語に対しては，

†1 豊橋技術科学大学
Toyohashi University of Technology

© 2011 Information Processing Society of Japan

5000 個の音節を用いた音節同士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)[9]．なお，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[4]．また，置換，挿入，脱落誤りを考慮した音節列同士のマッチングに基づく検索方法も試みられている[10]．音素の n-gram を用いた検索方法も種々提案されてはいるが[1]，基本的には bag of words の使い方で，音素認識誤りは考慮されていない.[3][11]

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.[8] そこで，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，音節ラティスの n-gram を用いる．手島らは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るが,[7] 音節ラティスへの適用は困難である．

3. 認識誤り・未知語に頑健な高速検索手法

3.1 n-gram に基づく未知語検索法[6]

本手法では，未知語に頑健な検索手法として，音節ラティスを使用して検索の際に認識誤りを考慮して検索を行う．本手法の概略を図 1 に示す．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．そして，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 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 2 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，本稿では 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を 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，検索後の検索候補を出力する．例えば，図 2 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，連接を確認する．

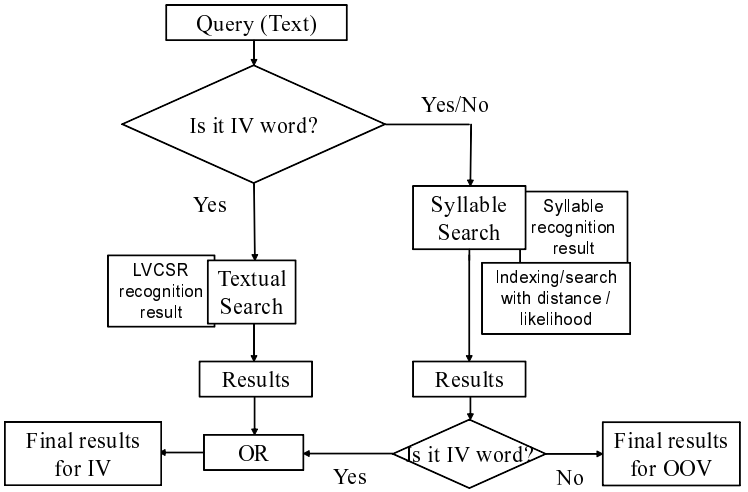

図 1 提案手法のフローチャート

3.2 認識誤りに関する対策[6]

3.2.1 置換誤り対策

置換誤り対策としては，文献[6] 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 ベストを組みあわせ，トライグラムを作成する (図 2 参照)．つまり，1 つのインデックスに対して $3 \times 5 \times 3 = 125$ 個のトライグラム情報を持たせる．例えば,「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

© 2011 Information Processing Society of Japan

Vol.2011-SLP-89 No.5
2011/12/19

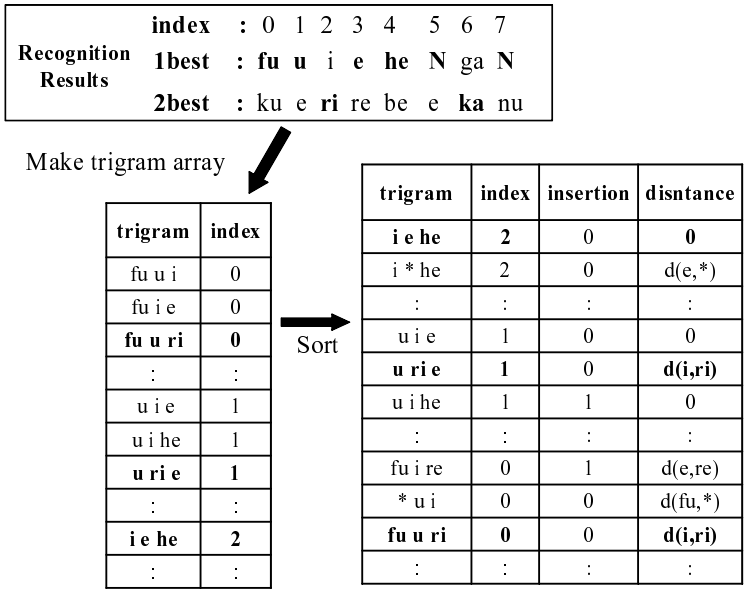

図 2 トライグラムアレイ作成手順

ていれば検索することができる．5 ベスト中に含まれない場合でも，次節に示すダミー音節や挿入誤り対策と脱落誤り対策の併用で対処できる．

3.2.2 ダミー音節を含めた置換誤り対策

置換誤り対策をおこなうことで，5 ベストに含まれる誤りは対処する事ができる．また，5 ベストに含まれない場合でも挿入誤りと脱落誤り対策の併用によって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索引に登録しておき，5best 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み合わせは 125 通りから 200 通りに増加する．図 2 では，”i * he”や”* u i”などがダミーを含んだ例である．この時の距離 $d(e,*)$ と $d(fu,*)$ は等しく，比較的大きな固定値とする．

3.2.3 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 2 参照．3 連続音節に対して 1 箇所) ．したがって，1 つの位置に対し，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．

3.2.4 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おこなう．本実験でははは 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．但し，脱落は連続する 3 音節以内に一ヶ所に制限する．

3.3 距離付き n-gram インデックスによる検索[6)]

前述した認識誤り対策をおこなうことで認識誤りに頑健な検索をおこなうことができる．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スの候補検出後に DP マッチングによる後処理による検索法を，以前提案した.[6)] しかし，DP マッチングによる検出候補の削減では，検出候補すべてに対して DP マッチングをおこなわなければいけないため，検索対象音声が長くなったり，検出候補が増えると処理時間が長くなるという問題があった．そこで，認識誤り対策をおこな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おこなう方法を提案した.[6)] 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おこなえる．

置換誤りの距離は 1 ベストからの音節間距離 (Bhattacharyya 距離)[6)] 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 2 参照)．つまり，1 ベストの音節の距離は 0 となる．また，ダミー音節の距離には大きな固定値を与えている．次に，挿入誤りの距離は挿入誤り対策をおこなったかの有無を 0，1 の 2 値で表現する (図 2 参照)．最後に脱落誤りの距離に関しては，検索時に脱落させた音節数を脱落誤りの距離とする．本実験では最大 2 音節の脱落を許すため (3 連続音節につき 1 箇所)，0，1，2 の 3 値となる．

また，これらの距離をクエリの長さ (トライグラム数) で正規化し，閾値と比べて検出するが，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．この問題を解決するために閾値の可変化手法を提案する．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．言い換えれば，短い

© 2011 Information Processing Society of Japan

Vol.2011-SLP-89 No.5
2011/12/19

表 1 組み合わせとカバー率

	組み合わせ	カバー率
従来法：5best	125	0.753
提案法：5best+dummy	200	0.977
提案法：4best+dummy	112	0.972
削減法① ：reduced1	108	0.967
削減法② ：reduced2	88	0.975
削減法③ ：reduced3	64	0.961

クエリに比べ，長いクエリは閾値を緩めても信頼度よく検出できるからである．たとえば，音節数ごとに 10% づ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く．このようにすることで，閾値が比較的小さい箇所 (Precision が高く，Recall が低い個所) で長いクエリの検出が増加し，Recall を増加することができる．

3.4 置換誤り対策に対する制限

索引を構築する際に置換誤り対策で 5best まで考慮すると，1 つの位置でトライグラムの組み合わせは 125 通りにもなる．そうなると単純計算で索引のサイズも 125 倍となり，扱いが難しくなる (挿入誤りを許すと 600 通り)．また，音節のうち 1 音節ダミーを許すと組み合わせは 200 通りになる (挿入を許すと 800 通り)．そこで，置換誤り対策に制約を追加し，索引の削減をおこなった．今回は以下の制約条件をそれぞれ加えてインデックスを構築した．いずれもトライグラムを構成する音節のうちダミー音節を使うのは 1 音節だけである．

- **削減法①**トライグラムを構成する 3 音節の内，2 音節は 1〜3 ベストの結果を使う．残りの 1 音節は 1〜5 ベスト (+dummy) の結果を用いる $\Longrightarrow$ 組み合わせ数は 108 通り
- **削減法②**トライグラムを構成する 3 音節の内，1 音節は 1 ベストの結果を使う．そのほかの音節は 1〜5 ベスト (+dummy) の結果を用いる $\Longrightarrow$ 組み合わせ数は 88 通り
- **削減法③**トライグラムを構成する 3 音節の内，1 音節は 1 ベストの結果を使い，もう 1 音節は 1〜3 ベストの結果を用いる．そのほかの音節は 1〜5 ベスト (+dummy) の結果を用いる $\Longrightarrow$ 組み合わせ数は 64 通り

種々のインデックス削減法によるトライグラムの組み合わせ数とこれらのトライグラムによる正解カバー率を表 1 に示す (第一候補 0.836，第 3 候補まで 0.891，第 4 候補まで 0.90，第 5 候補まで 0.910 とした．表 3 参照)．

3.5 本手法と DP マッチング法の類似点と相違点

本手法と DP マッチングとの類似点として，両方とも脱落，挿入コストを任意に設定することができる点が挙げられる．相違点として DP 手法は入力されたクエリの音節と第 n 候補間の距離で評価するのに対し，本手法は置換コストに関して第 5 候補まで考慮し，第一候補と第 n 候補間の距離で評価する点である．したがって，本手法では正確にクエリとの距離を計算することができない．我々の方法でも，第 n 候補の認識尤度 (の逆符号) を使用すれば DP マッチング距離相当のものになりうる．一方で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 1/2〜2 の DP パス)，挿入と脱落の制限が弱いが，本手法は上記で述べたとおり挿入，脱落は 3 音節あたり 1 音節に制限しているので，再現率 (Recall) は減少するが，適合率 (Precision) が向上すると考えられる．

3.6 大語彙認識結果を併用した既知語検索

未知語に対しては音節認識結果を用いて索引を構築し，検索をおこなう．加えて，既知語では大語彙認識結果を併用することでさらに検索性能が向上する．クエリには以下のような 3 タイプに分類することができる．

- 既知語のみから構成される既知語クエリ
- 未知語のみから構成される未知語クエリ
- 未知語，既知語の両方を含むような複合語クエリ (例：”名犬ラッシー”では”名犬”は既知語で，”ラッシー”は未知語となる)

ここでは既知語のみから構成される既知語クエリと，未知語，既知語両方を含む複合クエリに対して，2 つの索引を用いて検索をおこなう．はじめに，大語彙認識結果のコンフュージョンネットワークから成る，単語単位の転置インデックスを用いて検索をおこなう．大語彙認識結果を用いることで正しく認識された既知語を検出することができる．2 つ目は, 未知語と同様の音節単位の認識結果から構築した n-gram インデックスからの検索である．これは大語彙認識結果で誤ったものを補うために用いられる．この 2 つの結果を OR 演算を用いて結合する．n-gram インデックスからの検索は false alarm は増加するが，認識誤りに対してより頑健になる．

3.7 検 索 性 能

3.7.1 メモリ使用量

我々が提案した n-gram インデックスの構造を説明する．ここでは $n = 3$ として説明する．１つのトライグラムは”音節の 3 つ組”，”そのトライグラムが持つエントリ数”から成る．エントリは”出現位置”，”挿入距離”，”置換距離”を持つ．索引は 1 つのトライグラムに対して複数のエントリが存在する．n-gram インデックスを構築する際に必要とされるメ

© 2011 Information Processing Society of Japan

表 2 索引に必要なメモリ量

(a) Memory size of S_1

手法	3 つ組	エントリ数	合計
従来手法	4bytes	4bytes	8bytes
コンパクト	3bytes	4bytes	7bytes

(b) Memory size of S_2

手法	出現位置	挿入距離	置換距離	合計
従来手法	4bytes	4bytes	4bytes	12bytes
コンパクト	4bytes	1bit	7bits	5bytes

モリ使用量は式 (1) で表される．

$$M = M_1 \times S_1 + M_2 \times S_2 \quad (1)$$

S_1 = memory size of {n-gram type + number of entries}

S_2 = memory size of {position
+insertion distance
+substitution distance}

ここで M_1 は n-gram の種類数を表しており，M_2 は n-gram の総エントリ数を表している ($M_2 \gg M_1$)．S_1 は一つの n-gram に対して必要とされるメモリを表しており，S_2 はエントリごとに必要とされるメモリの量を表している．それぞれの必要なメモリ量は表 2 のようになる．n-gram のエントリ数は誤り対策をおこなうことで増加する．たとえば，トライグラムの場合に置換誤りで 5best まで考慮すると通常のインデックスの 125 倍にもなる．さらに，挿入誤りを考慮するとその 4 倍になる．

従来の手法[5)6)] では置換，挿入の距離を表すのにそれぞれ 4byte のメモリを確保していたため (表 2：従来手法)，必要とするインデックスサイズが大きくなっていた．実際には，挿入距離は 0，1 の 2 値のため 1bit で表すことができる．置換の距離はバタチャリヤ距離を採用しているが，7bits(128 通り) で量子化することで性能の低下なしに圧縮することができる (表 2：コンパクト)．挿入距離もバタチャリヤ距離等を採用する場合は，挿入距離と置換距離をそれぞれ 4bits で表現する．このようにすることでインデックスのサイズを半分以下にすることができた．実際に従来手法で 1 時間の音声ドキュメントを第 5 候補まで用いて構築したインデックスのサイズは約 $40MB$ 程度であるが，索引をコンパクトに表現すると約 $17MB$ まで圧縮することができる．さらに，3.4 節で述べた索引構築の際に置換誤り対策に制限を設けることで性能の低下なしに半分にまで削減することができる．

表 3 音節認識率 (%)

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
単語認識結果	5.4	4.6	22.7	71.9	67.3

3.8 検索時間

我々の提案する手法は 2 分探索に基づいており，1 つのトライグラムの検索にかかる計算量は $O(\log_2 M_1)$ となる．実際には長いクエリは分割されるため，クエリの分割数を k とすると，計算量は $O(k \log_2 M_1)$ となる．また，脱落誤りを考慮した場合，さらに検索回数は増える．たとえば，6 音節のクエリを検索する場合，クエリの各位置を脱落させて検索するため，6 回のトライグラムの検索が必要となる．次に，検出したトライグラムの連接しているかのチェックをおこなうため，検出数に比例し処理が必要となる．これはトライグラムの検索よりも時間がかかる．したがって，全体的な処理時間はクエリの長さと検出数に依存する．一方で，LVCSR の結果を用いた単語の転置インデックスの検索は n-gram インデックスからの計算よりも高速おこなえる．単語の転置インデックスでは認識誤りなどを考慮しないため，インデックスのサイズも n-gram インデックスと比べると比較的小さい．

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) のコアデータ 44 時間分を用い，本研究室で開発された SPOJUS++[13)] による音節認識結果を対象とし，検索，評価をおこなった．大語彙連続音声認識 (LVCSR) の辞書（約 28000 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秋葉らの報告[12)] にあるものを使用した．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と LVCSR の結果による単語認識率と音節列に変換後の音節認識率を表 3 に示す (単語正解率は 72%)．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．

4.2 既知語検索結果

既知語検索をおこなった結果を表 4 に示す．”LVCSR”は大語彙認識結果 (confusion network) を用いた単語の転置インデックスからの検索結果を，”n-gram”は従来法で音節認識

© 2011 Information Processing Society of Japan

表 4 既知語検索結果 (LVCSR+n-gram)

	LVCSR	n-gram(5-best)	LVCSR+n-gram	DTW	LVCSR+DTW
Recall	0.51	0.44	0.68	0.43	0.69
Precision	0.95	0.86	0.88	0.94	0.93
F 値	0.67	0.59	0.77	0.59	0.79

結果を用いた n-gram インデックスからの検索結果を表している．”LVCSR+n-gram”はこれらを組み合わせた結果である．”DP”は 5 ベストまで考慮するバタチャリヤ距離を用いた音節列に対する音節 DP マッチングである[5]．”LVCSR+DP”は大語彙認識結果と DP マッチングを組み合わせた結果である．図 3 はこれらの結果を比較した図である．”dummy”は 5 ベストまで用いた索引にダミー音節を加え検索する提案手法である．”dummy+LVCSR”は提案法での検索に LVCSR を併用した結果である．従来法と比べ，F 値の最大値はほぼ同じであるが，提案手法では Recall の最大値が増加した．

再現率と精度のバランスを考えたとき，DP マッチングと提案手法に差がないことがわかる．加えて，大語彙認識結果を併用することで提案手法，DP マッチング共に性能が大幅に改善している．この時でも両者に大きな性能の差はない．

4.3 未知語検索結果

連続音節認識の結果の 5 ベストまでを用い，それぞれの対策をおこなった際の検索結果を表 5 に示す．(1)，(2)，(3) はそれぞれ置換誤り対策，挿入誤りを対策，脱落誤り対策をおこなった結果となっている．ベースラインの DP マッチングは既知語検索のときに用いたものと同じ音節単位の DP マッチングである．この結果から誤り対策で最も効果があるのは置換誤り対策であることがわかる．全ての対策をおこなったとき，ベースラインの DP よりも F 値がよくなっている．

次に，ダミー音節を加えた際の結果を図 4 に示す．”DP”は DP マッチングを表し，”n-gram(5best)”は従来手法を示している．”5best+dummy”は 5 ベストまで用いた索引にダミー音節を加え検索する提案手法である．同様に”4best+dummy”は 4 ベストまで用いた索引にダミー音節を加えた提案手法である．ダミー音節を用いた手法では Precision は多少低下するが，Recall が増加する．5best に含まれず，検出できなかったものが検出できるようになるため，Recall の最大値も約 10% 程度増加した．F 値の最大値は 0.51 から 0.52 に向上した．

また，これらに索引の削減をおこなった結果を図 5 に示す．”dummy”は図 4 の”5best+dummy”に相当する．”dummy+reduced1”は 3 音節の内 2 つの音節は 1〜3 ベストを用いる削減法

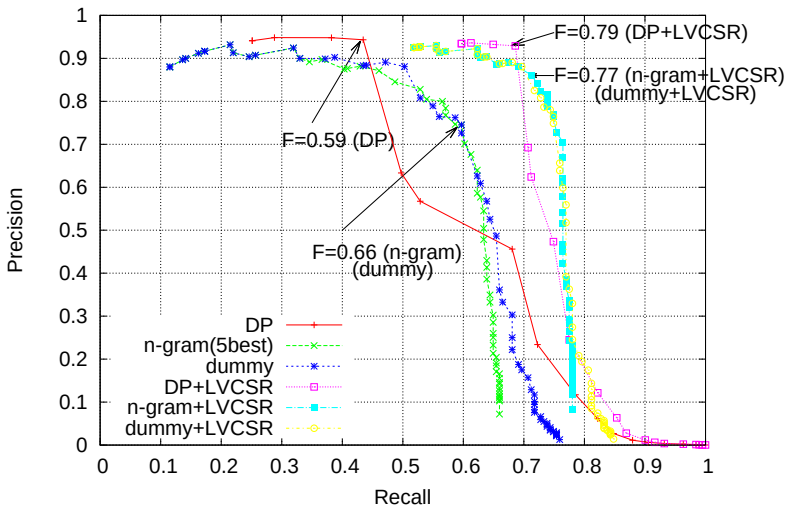

図 3 既知語検索結果

①である．”dummy+reduced2”は 3 音節の内 1 つの音節は必ず 1 ベストを用いる削減法②である．”dummy+reduced3”は 3 音節の内 1 つの音節は必ず 1 ベストを用い，もう 1 音節は 3best までの結果を用いる削減法③である．索引の削減をおこなった手法では F 値の最大値の低下はほとんど見られない．Recall は従来法の n-gram より高い値になっている．索引のサイズでは，”reduced2”は 1.3GB から 680MB まで削減でき，”reduced3”は 410MB まで削減できた．”dummy+reduced1”は特殊で，索引のサイズは 1.3GB から 1.1GB となっており削減量は少ないが，F 値の改善が得られた．これは索引構築の際の制限が”reduced2”や”reduced3”よりもうまく働き，誤検出を減らすことができたためだと考えられる．

また，3.3 節で述べた閾値を可変にした手法の結果を図 6 に示す．今回は閾値を 5%，10%，15% とクエリの音節長に応じて緩めていく 3 パターンでおこなった．Precision が高く，Recall が低い箇所での改善が顕著であり，最も性能が改善したのは閾値を 15% づつ緩めていく方法であり，F 値の最大値は 0.55 となった．実際には 20% などのパターンでもおこなったが，これ以上の改善は得られなかった．今回は単純な方法でおこなったが，このほかにも閾値の可変方法は数多くあり，今後はこれについても調べていく予定である．

© 2011 Information Processing Society of Japan

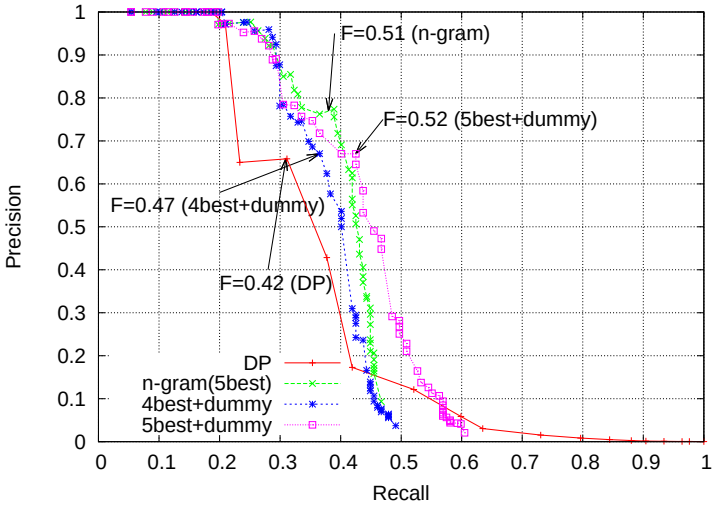

図 4 未知語検索結果

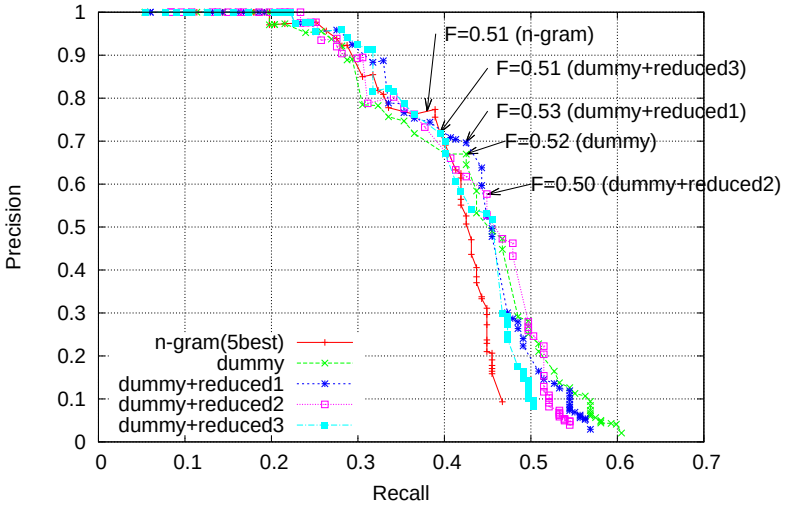

図 5 索引の削減をおこなった未知語検索結果

表 5 未知語の誤り対策別の検索結果 (連続音節認識)

OOV	n-gram index								DP
	対策無し	(1) 置換	(2) 挿入	(3) 脱落	(1)+(2)	(1)+(3)	(2)+(3)	(1)+(2)+(3)	
Recall	0.05	0.17	0.05	0.09	0.23	0.26	0.11	0.38	0.31
Precision	1.0	0.90	1.0	1.0	0.95	0.81	1.0	0.77	0.66
F 値	0.10	0.28	0.10	0.16	0.37	0.40	0.19	0.51	0.42

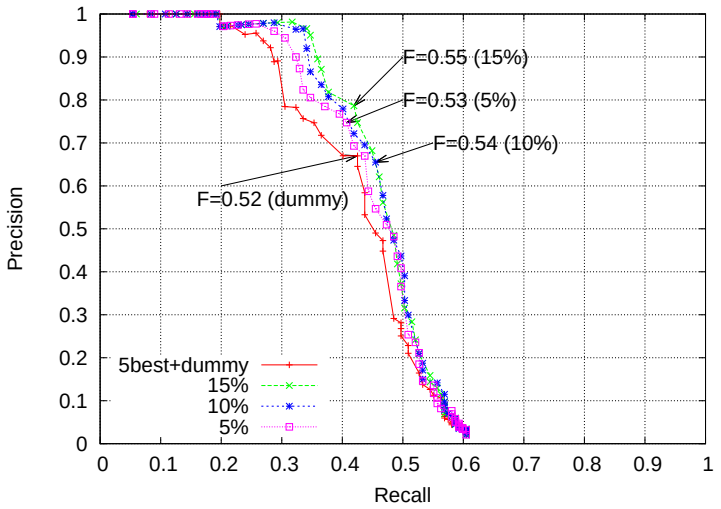

図 6 閾値を可変にした未知語検索結果

4.4 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 7 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 500[ms] なのに対し，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連接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，1000 時間の音声に対して 30ms 程度

© 2011 Information Processing Society of Japan

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手法と従来手法を比較した結果を図 8 に示す．提案法である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 7 倍程遅い平均 7[ms] であった．しかしながら，DP マッチングより約 70 倍高速である．

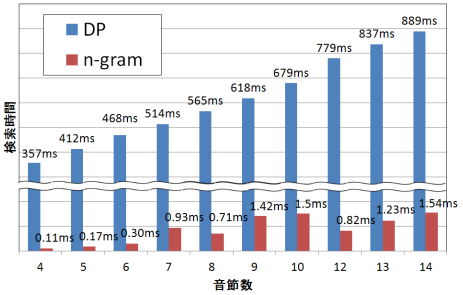

図 7 1 クエリあたりの検索時間の比較 (DP マッチングと従来手法)(音声ドキュメント量:44 時間)

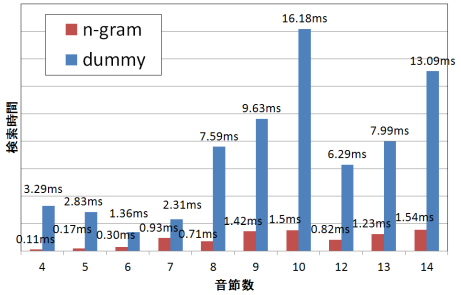

図 8 1 クエリあたりの検索時間の比較 (提案手法と従来手法)(音声ドキュメント量:44 時間)

5. おわりに

本稿では，ダミー音節を導入した n-gram インデックスによる高速検索法を提案し，5 ベストの音節正解率 91% の音声ドキュメントから既知語に対して 0.79, 未知語に対して 0.52 の F 値を得た．また，ダミー音節をインデックスに組み込むことで，置換誤りで対処できなかった箇所を検出することができ，Recall の最大値が増加した．インデックスが大規模になる問題に関しても，インデックス構造を見直すことで 50% 以上の削減ができた．NTCIR9 Formal run での F 値は 0.645(IV+OOV) であったが，今回提案したダミー音節と閾値可変化の導入により 0.669 と向上した．従来法と比べ，適合率を維持しながら再現率が向上したことから，MAP 尺度[12] では性能が大きく向上したと考えられる．今後の課題として，置換誤り距離や脱落誤り距離の検討とダミー音節を用いた手法，マージ処理の高速化が挙げられる．

参 考 文 献

1) B.Chen, H.Wang and L.Lee: Retrieval of broadcast news speech in Mandarin Chinese collected in Taiwan using syllable-level statistical characteristics, ICASSP, pp.2985–2988 (2000).
2) C.Allauzen, M.Mohri and M, S.: General indexation of weighted automata - application to spoken utterance retrieval, Workshop on interdisciplinary approaches to speech indexing and retrieval, pp.33–40 (2004).
3) C.Ng, R.Wilkinson and J.Zobel: Experiments in spoken document retrieval using phoneme n-grams, Vol.32, Speech Communication, pp.61 – 77 (2000).
4) H.Wang: Experiments in syllable-based retrieval of broadcast news speech in Mandarin Chinese, Vol.32, Speech Communication, pp.49 – 60 (2000).
5) K.Iwami, Y.Fujii, K.Yamamoto and S.Nakagawa: Out-of-vocabulary term detection by n-gram array with distance from continuous syllable recognition results, SLT, pp.200–205 (2010).
6) K.Iwami, Y.Fujii, K.Yamamoto and S.Nakagawa: EFFICIENT OUT-OF-VOCABULARY TERM DETECTION BY N-GRAM ARRAY INDICES WITH DISTANCE FROM A SYLLABLE LATTICE, ICASSP 2011 (to appear) (2011).
7) K.Katsurada, S.Teshima and Nitta, T.: Fast keyword detection using suffix array, Interspeech, pp.2147–2150 (2009).
8) K.Ng: Towards robust methods for speech document retrieval, ICSLP, pp. 1088–1091 (1998).
9) M.Larson and S.Eickeler: Using syllable-based indexing features and language models to improve German spoken document retrieval, EuroSpeech, pp.1217 – 1220 (2003).
10) M.Wechsler, E.Munteanu and P.Schauble: New techniques for open-vocabulary spoken document retrieval, SIGIR, pp.20 –27 (2008).
11) S.Dharanipragada and S.Roukos: A multistage algorithm for spotting new words in speech, Vol.10, IEEE Transactions on Speech and Audio Processing, pp.542 – 550 (2002).
12) T.Akiba, H.Nishizaki, K.Aikawa, T.Kawahara and T.Matsui.: Overview of the IR for Spoken Documents Task in NTCIR-9 Workshop, Proceedings of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access (2011).
13) Y.Fujii, K.Yamamoto and S.Nakagawa: Large Vocabulary Speech Recognition System: SPOJUS++, MUSP, pp.110 – 118 (2011).

© 2011 Information Processing Society of Japan