最良照合STDによる音声ドキュメント索引付けの評価および分析*

☆堂元健太郎, 宇津呂武仁 (筑波大), 澤田直輝, 西崎博光 (山梨大院)

1 はじめに

一般に，音声中の検索語検出 (Spoken Term Detection, STD) においては，大語彙音声認識システムを用いて音声認識を行うため，音声認識誤りや未知語の対策が課題である．これらの問題に頑健な STD 手法として，10 種類の音声認識システムの認識結果から音素遷移ネットワーク (Phoneme Transition Network, PTN) 型のインデックスを構築し，これと音素列に変換した検索語の照合を行う方式 [1] が提案されている．しかし，音素照合型 STD においては，検索語と異なるキーワードの発話であっても音素列が類似していれば検出してしまうという，過照合による誤検出が重要な問題である．

そこで文献 [2] では，当該分野の音声中に出現する可能性のあるキーワード集合をあらかじめ用意しておき，これら全てをクエリとして音素照合型 STD(従来法である PTN 型インデックスを用いた STD [1]) を適用した後，照合音声区間が競合するキーワード集合に対して，照合コストを用いた順位付けを行い，照合コスト最小のキーワードのみを STD 結果として出力する「最良照合 STD によるキーワード集合の索引付け」方式を提案した．この方式においては，Fig. 1 左の「バブルソート」という音声区間の場合，競合するキーワード集合のうち最小コストで照合する「バブルソート」が優先され，Fig. 1 右の「二分探索」という音声区間の場合も，競合するキーワード集合のうち最小コストで照合する「二分探索」が優先される．

ここで，この「最良照合 STD によるキーワード集合の索引付け」方式においては，あらかじめ用意しておくクエリ・キーワード集合をどのように作成するか，という点について有効な方式を確立することが最も重要である．また，音声ドキュメントに対して検索を行う際に用いるクエリ・キーワード集合と，索引付け対象として用いるキーワード集合は，必ずしも一致する必要はないため，与えられたクエリ・キーワード集合に対して，検索性能を最適化するための索引付け対象キーワード集合をどのように用意するか，という点も同様に重要である．

以上をふまえて，本稿では，与えられたクエリ・キーワード集合に対して，各クエリ・キーワードを構成する構成形態素を索引付け対象キーワードとして用いる (Fig. 2 (a)) ことにより，クエリ・キーワード集合の検索性能が改善することを示す．また，各クエリ・キーワードごとに，「音素遷移ネットワークを用いた STD」と比較して検索性能の改善の有無を判定し，「キーワード集合をクエリとする最良照合 STD」方式によって検索性能が改善するクエリ・キーワードについては同手法を適用し，それ以外のクエリ・キーワードについては，「音素遷移ネットワークを用いた STD」を適用するという選択的併用手法 (Fig. 2 (b)) により，あらかじめ用意したクエリ・キーワード集合全体での検索性能が最も改善することを示す．

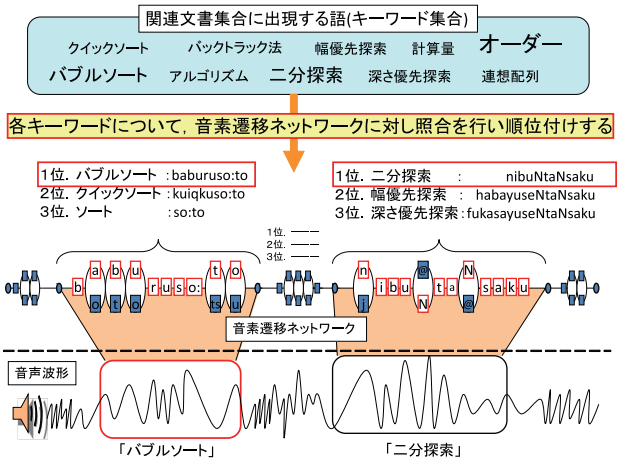


Fig. 1 最良照合 STD によるキーワード集合の索引付け

2 音素遷移ネットワーク型STD

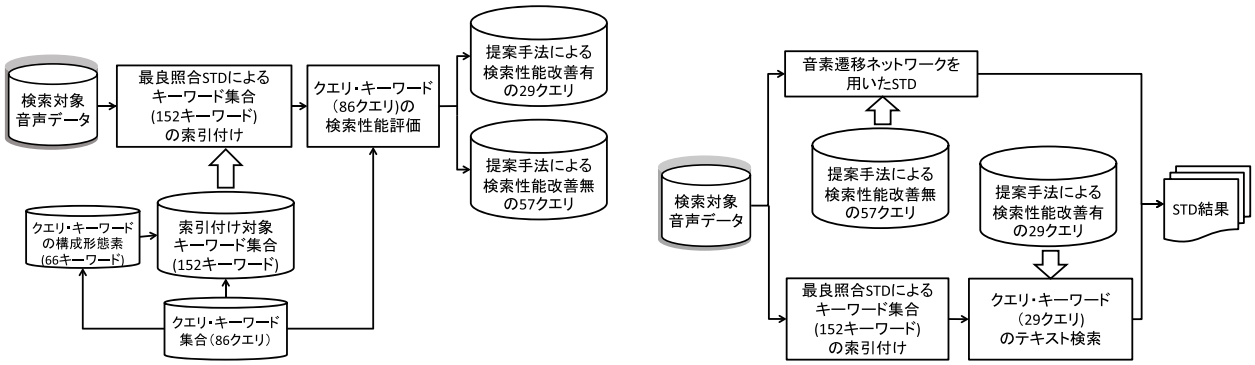
本稿では，PTN 型 STD として，文献 [3] における「Vot+Acw1」を用いた [2]．この方式においては，デコーダとして Julius rev. 4.1.3 を用い，2 種類の音響モデル (AM)，および，5 種類の言語モデル (LM) を用意して，AM と LM の組み合わせによって 10 種類の音声認識モデルを構築した．本稿では，模擬講義 (雑音の多い低品質な音声) [4] を評価対象として STD を適用する．この模擬講義を対象とした単語認識率は 26%，単語正解精度は 9%程度である [4]．

3 最良照合STDによるキーワード集合の索引付け

3.1 クエリ・キーワード集合

本稿の評価実験においては，模擬講義に対してクエリ・キーワード集合を人手で作成した．その際には，

*Evaluation and Analysis of STD by Selecting the Best Match from Candidate Query Keywords, by DOMOTO, Kentaro, UTSURO, Takehito (University of Tsukuba), SAWADA, Naoki, NISHIZAKI, Hiromitsu (University of Yamanashi)



(a) クエリ・キーワードの構成形態素の索引付け，および，性能改善の有無に基づくクエリ・キーワード集合の分割

(b) 「クエリ・キーワードの最良照合 STD」と「音素遷移ネットワークを用いた STD」の併用

Fig. 2 「クエリ・キーワードの最良照合 STD」と「音素遷移ネットワークを用いた STD」の併用による STD

講義の書き起こし文書に対して，専門用語抽出ツール TermExtract[1] を適用し，出力された専門用語候補に対して，検索語として利用するクエリ・キーワードを手作業によって選定した．ただし，4.2 節においてクエリ・キーワードの構成形態素を索引付けする方式の有効性を評価することを想定して，本節で作成するクエリ・キーワード集合中には，他のクエリ・キーワードの構成形態素となるキーワードを除外し，各々，最長となるクエリ・キーワードからのみクエリ・キーワード集合を構成した．

3.2 STD 結果の競合集合の作成

次に，キーワード集合のすべてのキーワードをクエリとして PTN 型 STD を行い，STD 結果を併合する．この結果，音声中の各区間ごとに複数の STD 照合結果が重複して得られる．このうち，検出フレーム時間が重複している照合結果を推移的に収集することにより，STD 結果の競合集合 C を作成する．

3.3 最長フレーム照合結果優先方式

提案手法においては,「最長フレーム法」(最長フレーム照合結果優先方式) および「最長フレーム法+リランキング」(最長フレーム照合結果優先方式+ 競合集合内の最小コストを用いたリランキング) の二種類の手法を適用するが，本節では，特に,「最長フレーム法」について述べる．

まず，キーワードを w，その STD 検出開始フレームを t，STD 検出終了フレームを t'，STD 照合コストを $cost$ とすると，n 個の四つ組 $\langle w, t, t', cost\rangle$ から成る競合集合 C は，次式のように書ける．

$$C = \Big\{\langle w_1, t_1, t'_1, cost_1\rangle,\ \ldots,\ \langle w_n, t_n, t'_n, cost_n\rangle\Big\}$$

このとき，従来手法による STD においては，Fig. 3 に示すように，競合集合 C のすべての照合結果を出力する．

一方,「最長フレーム法」では，競合する n 個の照合結果のうち，最小コストとなる照合結果

$$\langle\ w_{min},\ t_a,\ t'_a,\ cost_{min}\ \rangle$$

をまず選定する．次に競合集合 C 内の照合結果について，最小コスト照合結果からコスト幅 Δ 以内にある照合結果を索引付けの候補集合 $C(\Delta)$ とする．

$$C\,(\Delta) \quad = \quad \Big\{\ \langle\ w, cost\ \rangle \in C \ \Big|\ cost \le (cost_{min} + \Delta)\ \Big\}$$

最後に，$C\,(\Delta)$ の要素のうち，検出フレーム長 $t'_b - t_b$ が最大となる照合結果

$$\langle\ w_{lg},\ t_b,\ t'_b,\ cost_{lg}\ \rangle$$

を選定する．これが当該音声区間の STD 結果となる．そして，この STD 結果と検出フレーム時間が重複している照合結果を削除する．

競合集合が空になるまで以上の処理を繰り返す．

3.4 競合集合内の最小コストを用いたリランキング

一般に，音素数の多いキーワードは，STD における照合コストが大きくなる傾向がある．そこで，提案手法「最長フレーム法」における照合コストを，競合集合内のより小さいコストに置き換える．ただし，その際には，検出フレーム長 $t'_b - t_b$ が最大となる照合結果

$$\langle\ w_{lg},\ t_b,\ t'_b,\ cost_{lg}\ \rangle$$

との間で検出フレーム時間が重複している照合結果のうち，最小コストとなる照合結果

$$\langle\ \overline{w}_{min},\ t_c,\ t'_c,\ \overline{cost}_{min}\ \rangle$$

[1] http://gensen.dl.itc.u-tokyo.ac.jp/

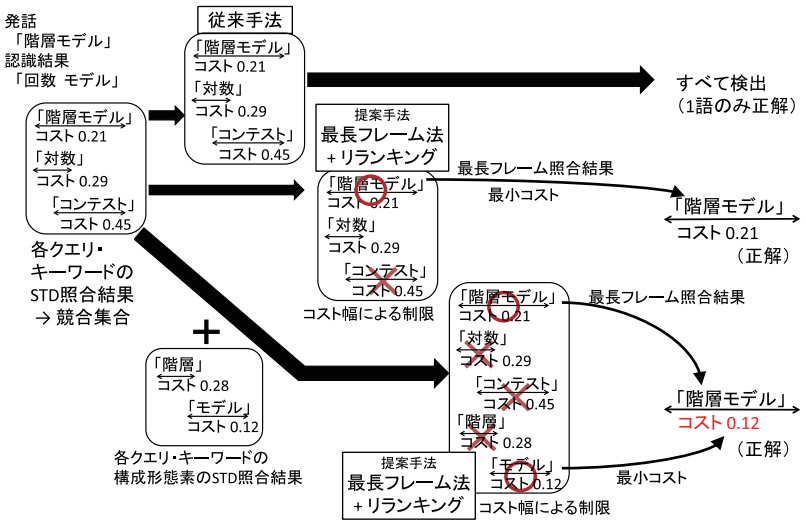


Fig. 3　従来手法，および,「最長フレーム法+リランキング」(クエリ・キーワードの構成形態素の索引付け有/無) による STD 結果 (最小コストを用いたリランキングにより「正解」の索引付けの照合コストを減らすことに成功する例)

の照合コストに置き換えることとし，この方式を「競合集合内の最小コストを用いたリランキング」方式とする (Fig. 3 中段の提案手法)．この手法による STD 結果は

$$\langle\ w_{lg},\ t_b,\ t'_b,\ \overline{cost}_{min}\ \rangle$$

と表記される．

4　「クエリ・キーワードの最良照合 STD」と「音素遷移ネットワークを用いた STD」の併用

4.1　クエリ・キーワードの最良照合 STD

「クエリ・キーワードの最良照合 STD」においては，前節の「最良照合 STD によるキーワード集合の索引付け」方式により索引付けされたテキスト・キーワードを対象として，テキスト検索を行うことにより STD を行う．具体的には，前節の「最長フレーム法+リランキング」による索引付けを用いたテキスト検索により STD を行う．

ここで，本稿で評価対象とした模擬講義においては，クエリ・キーワード集合のキーワード種類数は 86 キーワード，クエリ・キーワード出現箇所数は 323 箇所であった．また，提案手法におけるコスト幅 Δ は，$\Delta = 0.10$ とした．以上の設定における従来手法 (86 クエリ)，および，提案手法「最長フレーム法+リランキング」による索引付けを用いたテキスト検索 (Fig. 4 の「提案手法 (86 クエリ): 構成形態素の索引付無」．具体例は，Fig. 3 中段の提案手法．) の検索性能の比較を Fig. 4 に示す．この結果から分かるように，両者の間には検索性能の差はほとんどない．

4.2　クエリ・キーワードの構成形態素の索引付け

ここで，提案手法「最長フレーム法+リランキング」においては，照合結果の競合集合内において，検出フレーム長が短くてもよいので，照合スコアができるだけ低い語を含むことができれば，結果的に，クエリ・キーワードの適切なリランキングを実現できる可能性が高くなる (ただし，沸き出し誤りの索引付けのリランキングを引き起こす可能性も同時に高くなる)．このことを考慮して，クエリ・キーワード全 86 語の構成形態素全 66 語を索引付け対象キーワードとして追加する．この後，86 クエリ・キーワードを対象として検索性能を再測定した結果を Fig. 4 の「提案手法 (86 クエリ): 構成形態素の索引付有)」のプロットとして示す (具体例は，Fig. 3 下段の提案手法)．この結果においては，従来手法 (86 クエリ)，および,「提案手法 (86 クエリ): 構成形態素の索引付無」の検索性能を改善していることが分かる．

4.3　性能改善の有無に基づくクエリ・キーワード集合の分割

次に，各クエリ・キーワードごとに，従来手法「音素遷移ネットワークを用いた STD」と比較して検索性能が改善したか否かによって，従来手法，提案手法の選択的併用を行う．

まず，各クエリ・キーワードごとに，検索性能の改善の有無を判定するため，

> 照合コストの上限を 0.01~1.0 の間で 0.01 間隔で変化させた 100 個の点のうちの 3 点以上において，再現率および F 値とも従来手法より改善する

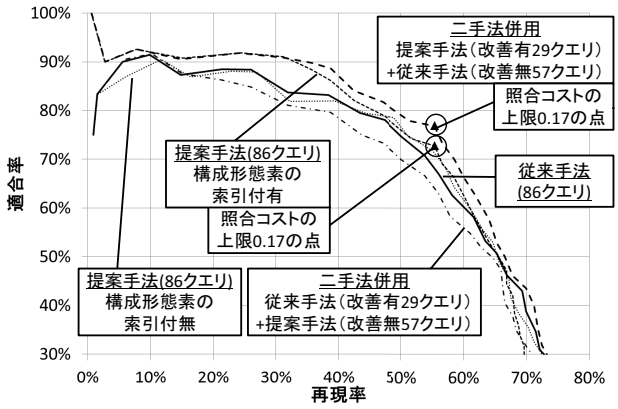


Fig. 4　評価結果

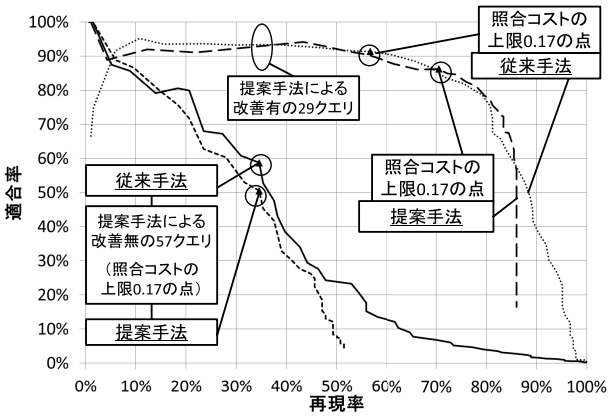


Fig. 5　「改善有 29 クエリ」または「改善無 57 クエリ」での検索性能 (索引付けキーワード: 全 86 キーワード+構成形態素 66 キーワード)

という条件を課すことにより，全 86 クエリ・キーワードを，改善有 29 クエリ，および，改善無 57 クエリに分類する．

この後，前節において，クエリ・キーワード全 86 語に加えて，構成形態素全 66 語を索引付け対象キーワードとして追加して索引付けを行った後，検索性能改善有 29 クエリ，および，検索性能改善無 57 クエリの各々について，従来手法，および，提案手法による索引付けを用いたテキスト検索の検索性能の比較を Fig. 5 に示す．この結果において，検索性能改善有 29 クエリの検索性能は検索性能改善無 57 クエリの検索性能を上回っている．また，検索性能改善無 57 クエリにおいては，両者の性能の差はほとんどない．一方，検索性能改善有 29 クエリにおいては，プロット上の差はほとんどないものの，Fig. 4 の「提案手法 (86 クエリ): 構成形態素の索引付有)」において F 値最大となる照合コスト上限に該当する点における再現率は，従来手法に比べて，提案手法による索引付けの方が 10%以上改善している．ここで再現率が改善することによって，次節において二手法を併用した結果，Fig. 4 のプロットにおける検索性能改善が可能となる．

4.4　二手法の選択的併用

最後に,「クエリ・キーワードの最良照合 STD」方式によって検索性能が改善する 29 クエリについては同手法を適用し，それ以外の 57 クエリについては，従来手法である「音素遷移ネットワークを用いた STD」を適用するという選択的併用手法 (Fig. 2 (b)) の検索性能を Fig. 4 の「二手法併用: 提案手法 (改善有 29 クエリ)+ 従来手法 (改善無 57 クエリ)」に示す．同様に，検索性能が改善する 29 クエリについては従来手法を適用し，それ以外の 57 クエリについては，提案手法を適用することにより，検索性能が低下する手法を併用する方式の検索性能を Fig. 4 の「二手法併用: 従来手法 (改善有 29 クエリ)+ 提案手法 (改善無 57 クエリ)」に示す．この結果から,「二手法併用: 提案手法 (改善有 29 クエリ)+ 従来手法 (改善無 57 クエリ)」が最も高くなっていることが分かる．

5　おわりに

本稿では,「最良照合 STD によるキーワード集合の索引付け」方式 [2] において，検索性能が改善するクエリ・キーワードについては同手法を適用し，それ以外のクエリ・キーワードについては，従来手法である「音素遷移ネットワークを用いた STD」を適用するという選択的併用手法により，あらかじめ用意したクエリ・キーワード集合全体での検索性能が最も改善することを示した．今後は，索引付けすべきキーワード集合を自動選定する方式に取り組む予定である．また，検索性能を改善する目的において，各クエリ・キーワードに対して，提案手法により検索性能が改善するか否かを自動判定することにより，クエリ・キーワード集合全体での検索性能を最適化する方式を開発する．

参考文献

[1] S. Natori, et al.: “Spoken term detection using phoneme transition network from multiple speech recognizers’ outputs”, Journal of Information Processing, **21**, 2, pp. 176–185 (2013).

[2] 堂元他：“キーワード集合をクエリとする最良照合 STD 方式”, 第 8 回音声ドキュメント処理ワークショップ SDPWS2014-09 (2014).

[3] 古屋他：“音声中の検索語検出における検出誤り抑制パラメータの検討”, 第 6 回音声ドキュメント処理ワークショップ SDPWS2012-11 (2012).

[4] 米倉他：“電子ノート作成支援システムで利用する音声からのキーワード検出技術”, 信学技報 NLC, **113**, 83, pp. 13–18 (2013).