自然言語処理 131－3
音声言語情報処理 26－3
（1999．5．28）

# 音声入力によるニュース音声検索システム

西崎 博光　中川 聖一

豊橋技術科学大学 情報工学系
〒441-8580 豊橋市天伯町字雲雀ヶ丘1-1
E-mail:{nisizaki,nakagawa}@slp.tutics.tut.ac.jp

近年、多くのニュース番組が放映されているが、過去のニュース番組から興味のある記事を見つけたいという欲求がある。

数多くのニュース番組の中から、必要なニュースを見つける場合、各ニュースに対してインデックスが付けられている場合はそれを使って検索することができる。しかし、ニュース音声データからの検索に対する需要もあり、この場合は放送された全てのニュースを予め文字化し、データベースとして蓄積しておく必要がある。この作業を人手で行なうのは不可能に近く、大語彙音声認識システムを用い、自動的に書き起こすこととなる。

本研究では、1) 自動的に書き起こしたデータベース (誤認識単語を含む) での検索性能を、テキスト入力のキーワードを用いて実験的に検討した。実験の結果、単語認識率が低いにもかかわらず、高い再現率を得ることができた。次に，2) 検索対象となる単語を音声で入力した際の問題点を挙げ、それに対する対処法を提案する。実際にキーワードを音声で入力し、提案した方法を使って実験を行ない、その有効性を示す。

# A Broadcast News Information Retrieval System via Voice

Hiromitsu NISHIZAKI and Seiichi NAKAGAWA

Department of Information and Computer Sciences
Toyohashi University of Technology, Tenpaku, Toyohashi, 441-8580, Japan
E-mail:{nisizaki,nakagawa}@slp.tutics.tut.ac.jp

To retrieve interesting broadcast news documents out of an enormous number of TV news programs, if no indexing is done on the news and word-based retrieval is required, it is inevitably necessary to transcribe all the broadcast news documents automatically and store them as a database. And this task can be done only by using a Large Vocabulary Continuous Speech Recognition(LVCSR) system.

In this paper, 1) the retrieval performance was experimentally compared between the system using automatically transcribed database(A) and the one using manually transcribed database(B). This experiment was done using text as the input to the system. As a result, high recall was obtained although the word recognition rate was low. Next, 2) to solve the inevitable problems which arise when the input to the system is realized as speech, i.e. misrecognition, a novel method was developed. In experiments, we retrieved news documents through inputed voice keywords to the system using by the method described above and represent its effectiveness.

# 1 はじめに

近年、多くのニュース番組が放映されているが、過去のニュース番組から興味のある記事を見つけたいという欲求がある [1][2]。ニュース音声の検索に関する研究は数多く行われており、さまざまな検索手法が提案されている。たとえば、Kenny らは単語単位のマッチングではなく、語彙サイズの増大という問題に着目し音素単位でのマッチングによる検索を行っている [3]。

数多くのニュース番組の中から、必要なニュースを見つける場合、各ニュースに対してインデックスが付けられている場合はそれを使って検索することができる。しかし、ニュース音声データからの検索に対する需要もあり、この場合は放送された全てのニュースを予め文字化し、データベースとして蓄積しておく必要がある。この作業を人手で行なうのは不可能に近く、大語彙音声認識システムを用い、自動的に書き起こすこととなる。

そこで本研究では、自動的に書き起こしたデータベースでの検索性能を調べるため、まず、実際のニュース音声に対して、音声認識システムにより書き起こし、検索用データベースを作成した。このデータベースと正確に書き起こしたデータベースに対して、キーワード群を使って検索された記事の再現率を求め、比較した。実験の結果、単語認識率が低いにもかかわらず高い再現率が得られた。

キーワードを音声で入力することを考えた場合、必ずしも正しく認識されるとは限らない。また、機械には認識結果が正しいキーワードかどうかわからないので、誤りもありうる認識結果を使って検索を行なわざるを得ない。また、キーワードに同音意義語が存在する場合は、同じ読みの単語すべてをキーワードとして扱う必要がある。こういった場合、実際にユーザーが意図しない記事を大量に含む検索結果が得られたり、逆に全く結果が出力されないことになるので、これらの記事をうまく絞り込んでいく必要がある。そこで検索処理に先立って、単語間の関連度を用い、キーワード候補の語数を絞る手法を提案する。単語間関連度は正確に書き起こしたデータベースより学習し、キーワード候補をグルーピングする。その結果、キーワード候補は幾つかのグループに区分される。そして、単語数の最も多いグループ中の単語を用いて検索処理を行なう。検索用のキーワード候補が実際の入力数よりも増大するというのは、音声でキーワードを入力したときのみに起こる現象であり、検索前に必要なキーワードを選択するというこの手法は他に類を見ない。

本稿では実際にキーワードを音声で入力し、前述の手法で検索実験行ない、その有効性を示す。

# 2 ニュース検索システム

## 2.1 概要

今回作成した、ニュース検索システムの概略図を図 1に示す。

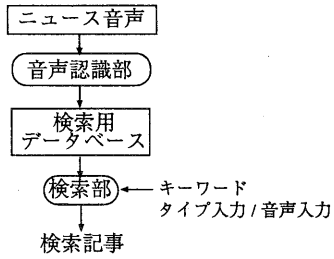

図 1: システムの概略

まず、ニュース音声を音声認識システムに通し、自動的に検索用データベースを作成する。これを基に、入力キーワード (タイピング入力、音声入力) に応じて記事を検索部で検索する。

検索部では全文検索 [4] を行なっているが、インデックス法 [5] を用いることで、高速な検索を可能にしている。検索キーワードは、テキスト入力でキーワードをいくつか入力する。すべてのキーワードが完全に一致した記事のみを出力する。ただし、これでは制約がきつ過ぎ必要な記事が検索されないので、入力キーワード数が多い場合は、全部が一致しなくてもその大部分が一致している記事を出力する。もし、入力キーワードが未知語だった (音声認識で使用した語彙辞書に入っていない) 場合は、音節列 (かな文字列) 単位の DP マッチングを行なうようにしている。

## 2.2 キーワードのグルーピング

キーワードの入力がテキストでなく、音声での入力も考えられる。音声によるキーワード入力では、キーワードが認識された時、

1. キーワードが正しく認識された
2. キーワードが違う単語として認識された (異なる音節列)
3. 正しい音節列ではあるが，異なる語 (同音意義語) として認識された

という場合が考えられるが、機械には認識結果が正しいキーワードかどうかわからないので、どの場合も得られた認識結果を使って検索処理を開始せざるを得ない。同音意義語が存在する場合は、全ての同音異義語を使って検索する必要があり、同音意義語がない場合でも、認識尤度の高い認識結果候補単語を複数個使って検索する必要性も考えられる。いずれにしても、発声単語数よりも多い単語セット (キーワード候補) を使って検索処理が行なわれるため、必要以上の記事が検索されたり、また逆に全く記事が検索されない恐れがある。こういった不具合を解決する方法として、キーワード間の関連度を用いたキーワードの絞り込み手法を提案する。関連度とは、ある 2 つのキーワードがどれくらい関係しているかを表す尺度で、以下の値を用いる。

- 共起頻度の利用

  2 つの単語間の関連度を求める際に、ある記事において、ある単語とどの単語が同時に同じ記事に出現しやすいかという情報を用いる。

2つの単語をそれぞれ、$W_1$,$W_2$としたとき、これらの $W_2$の $W_1$に対する関連度 $R(W_1,W_2)$ を以下のように計算する。

$$R(W_1,W_2) = \frac{1}{2}\{\frac{f(W_1,W_2)}{f(W_1)} + \frac{f(W_1,W_2)}{f(W_2)}\}$$

$f(W_i)$:$W_i$が出現した記事数 $(i=1,2)$
$f(W_1,W_2)$:$W_1$,$W_2$が共に出現した記事数

- 相互情報量の利用

相互情報量は、単語の共起や関連を客観的に表す尺度として用いられる。2つの単語 $W_1$,$W_2$ の相互情報量 $I(W_1;W_2)$ は、$W_1$ と $W_2$ を同じ記事で同時に観測する確率 $P(W_1,W_2)$ を、$W_1$ と $W_2$ を独立に観測する確率 $P(W_1)$,$P(W_2)$ と比較する。

$$I(W_1;W_2) = \log \frac{P(W_1,W_2)}{P(W_1)P(W_2)}$$

上記の式を変換して、

$$I(W_1;W_2) = \log \frac{\frac{f(W_1,W_2)}{N}}{\frac{f(W_1)}{N}\frac{f(W_2)}{N}}$$

$N$: 総記事数

2つの単語で、関連度が強いものはIの値が大きくなり、関連度がないものほど0に近づく。

関連度の学習は、正確に書き起こしたデータベースから学習した。ニュース記事から学習した前述の指標を使って、図 2に示すように関連度の高いキーワード候補どうしをグルーピングする。関連度には閾値を設けており、この閾値を越える関連度をもつグループどうしを関連づけるわけである。この例は、6個のキーワードの候補がありうる場合を示している。矢印で結んであるキーワードどうしが関連度の高いキーワードで、1グループを形成している。ここでは3つのグループが作られているが、最もキーワードの数が多い $G_1$のグループを使って検索を行なう。

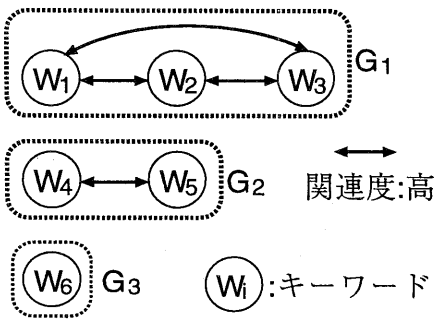


図 2: キーワード候補のグルーピング

図 3に実際の例を示す。これは、"公開"、"官庁"、"公務員"の3つのキーワードを音声入力したときの例である。"公開(こうかい)"の同音意義語として"更改"と"更改"、"官庁(かんちょう)"の同音意義語として"艦長"がある。この例では、"公開"—"官庁"間、"官庁"—"公務員"間、"公開"—"公務員"間の関連度が高くなっている(ある閾値を越えている)ので、これら3つのキーワード候補を1つのグループとしこれを検索キーワード群として用いる。

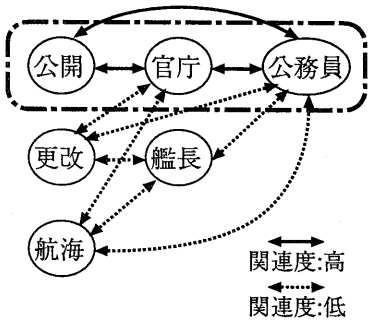


図 3: グルーピングの実例

## 2.3 検索方法

検索用のキーワード入力の違いにより次の4通りの方法で実験を行った。

1. テキスト入力
2. 音声入力 (1-best)
3. 音声入力 (3-best, グルーピングあり)
4. 音声入力 (3-best, グルーピングなし)

1-best とはキーワードの認識結果の 1-best 仮説のみを用い、3-best は 3-best 仮説までを使用するという意味である。音声入力の場合、すべてのキーワード候補がマッチングする記事のみを検索してくるのは制約がきつすぎるため、大部分がマッチングする記事のみを検索することにする。そこで、入力キーワード数に対し、どれだけマッチングすれば良いとするかの閾値を決めないといけないが、これは記事の検索率の期待値が 95%以上になるように設定した (但し書き起こしが 100%正しいと仮定した場合)。つまり、

$$\sum_{i=N}^{M} {}_M\mathrm{C}_i p^i (1-p)^{M-i} \geq 0.95$$

$M$ : 入力キーワード数
$N$ : 入力キーワードのうち記事中に実在する数

となるような N を求めた (表 1)。$p$ は入力キーワードが正しく音声認識される確率である。1-best の場合は $p=0.81$、3-best の場合は $p=0.85$ で計算した (表 5の結果より決定)。

表 1: 閾値の設定

| M | N | |
|---|---|---|
| | 1-best | 3-best |
| 3 | 2 | 2 |
| 4 | 2 | 2 |
| 5 | 3 | 3 |
| 6 | 3 | 4 |
| 7 | 4 | 4 |
| 8 | 5 | 5 |

# 3 検索実験

## 3.1 データベース

実験対象の音声データは、NHK ニュース (1996 年 6 月 1 日～7 月 14 日) で、記事の数は 976 記事 (このうちすべて clean な音声は 535 記事)、文数で 7099 文 (このうち clean な音声 4439 文, noisy な音声 2660 文) である。ニュース音声の書き起こしに使用した音声認識システムの条件を表 2に示す。言語モデルは第 1 パスでは語彙サイズ 20000 の単語 bigram、第 2 パスでは単語 trigram を使用している。この音声認識システム [6] を使用した場合、ニュース音声 (バックグラウンドミュージック、紙をめくる音などの背景雑音などが混入されている) の単語カバー率は 96.7%、単語正解率は 54.3%、単語正解精度は 38.0%と非常に低くなった。また clean と noisy 別だと、clean な音声の単語正解率は 57.7%、正解精度で 46.3%であり、noisy な音声の単語正解率は 47.3%、正解精度で 21.3%であった。

表 2: 認識実験の実験条件

| | |
|---|---|
| 音響モデル | |
| 5 状態 4 出力分布 (4 混合ガウス分布, 全共分散行列) | |
| 離散継続時間分布付き連続出力分布型 HMM | |
| 音節カテゴリ数 | 113 音節 |
| サンプリング周波数 | 12kHz |
| 窓関数 | 21.33ms ハミング窓 |
| フレーム周期 | 8ms |
| 分析 | 14 次元 LPC 分析 |
| 学習データ | |
| ASJ ATR503 文 A～J セットの 6 名の男性話者と 216 単語の音声データから初期モデルを作成 | |
| ASJ ATR503 文 A～J セットの 30 名の男性話者と JNAS 新聞記事文 125 名の男性話者を MAP 推定で追加学習 (総発話数 17221 文) | |
| 特徴パラメータ | |
| LPC メルケプストラム (10 次元 × 4 フレームの特徴量を KL 展開で 20 次元に圧縮) | |
| ＋ Δ ケプストラム (10 次元) | |
| ＋ ΔΔ ケプストラム (10 次元) | |
| ＋ Δ パワー ＋ ΔΔ パワー | |

キーワードを選択するため、検索対象記事ごとに 3 人の被験者にキーワードを 3～5 個選んでもらい、3 人とも共通に選んだ単語の集合をその対象記事のキーワード群とした。

正確に書き起こしたデータベース (以後「データーベース (A)」と記す) と、自動的に書き起こしたデータベース (以後「データーベース (B)」と記す) に対して、キーワード群を使って検索し、下記の再現率、適合率、検索率を求める。

**再現率 (recall)：** あるキーワード群のテキスト入力でデータベース (A) に対して検索された記事のうち、同じキーワード群でデータベース (A) または (B) に対して検索した場合に、どれだけ検索されたかを表す割合。

**適合率 (precision)：** 検索されたすべての記事のうち、正解の記事数の割合で、余計な記事の湧きだしが多いほど小さくなる。

**検索率：** 対象記事がどれだけ正しく検索されたかを表す割合 (検索対象記事だけを正解とみなした再現率)。

## 3.2 キーワードのテキスト入力による実験結果

実験結果を表 3に示す。1 キーワード群当たり検索された記事数はデータベース (A) で平均 7.4 記事、データベース (B) で平均 11.3 記事、再現率は 76.2%であり (clean データでの再現率は 80.9%、noisy なデータでの再現率は 70.5%)、対象記事 30 記事中で (B) のデータベースで 27 記事 (90.0%) が正しく検索された。再現率と検索率は本来同等の精度になりうるはずだが、差が生じたのは今回の検索対象記事の書き起こし文の正解率が良かったためと考えられる。また、適合率は 50.1%となり、検索された記事のうち約半分が、余計な記事の湧きだしということになる。

表 3: データベース (B) に対する実験結果 (テキスト入力)

| | | |
|---|---|---|
| 検索対象記事 | : | 30 |
| 再現率 | : | 76.2% |
| 適合率 | : | 50.1% |
| 検索率 | : | 90.0% |

表 3の実験結果を見ると、単語正解率 54.3%、単語正解精度 38.0%とかなり低い値になっているにもかかわず再現率が比較的高くなっている。これは、評価実験で入力したキーワード (異なり数で 104 単語、総数で 8185 単語、但し複合語が多い) の認識率 (93.0%, clean で 93.2%、noisy で 90.8%) が全体の認識率よりも高くなっているためである。音声認識を使って書き起こしたデータベースを用いると、2 割程度性能が低下してしまったが、全体の音声認識率は検索性能にそんなに影響しないということが言える。これは、文献 [1] の結果と符合する。

### 3.3 キーワード音声入力による検索実験

#### (a) キーワードの音声認識

2名の話者にキーワードもしくはキーワード列を発話してもらい、認識実験を行なった。キーワード列とは、キーワードの連続のことであり、複合名詞などに該当する。発話してもらったキーワード(列)は、3.2節のテキスト入力の検索実験で用いたキーワード(列)と同じものである。

キーワード(列)の認識には、ニュース音声の時と同じ大語彙連続音声認識システム(語彙サイズは20000単語)を使用した。このため、キーワードに助詞が挿入された結果が多く生じた。認識結果を表4と表5に示す。キーワード数は104個(このうち同音異義語があるものは21個)、キーワード列で数えると54個である。表4は、キーワードごとの認識率を求めた表で、表5は54個のキーワード列の内、どれだけが正しく認識できたかを示した表である。

表4、表5を見てもわかるが、比較的高い認識率が得られた。これは、キーワード列が複数の形態素から構成されており、認識時にbigramの言語確率がうまく機能しているためだと考えられる。これに対して、単一形態素のキーワードの認識は悪かった(69.6%)。

認識結果を調べて判明したこととして、キーワードの認識結果には余計な文字(とくに助詞)が挿入されてしまっているのがほとんどである(表4の挿入率の高さからも言える)、正解のキーワード(列)はほぼ3-bestまでに入っている、ということが挙げられる(本実験では3-bestで正解率は飽和している)。

キーワードの認識時、キーワードの脱落により検索結果が受ける影響に比べ、余計な単語(特に助詞)の挿入により受ける影響は少ないと思われる(余計な助詞、動詞などキーワードになり得ない単語(ストップワード)が挿入された場合は、キーワード候補を図1の検索部に入力する際に取り除くようにしているため)。つまり、表4の正解率と表5でいう準正解にあたる結果が、実質的なキーワード(列)の認識率に相当する。

表4: キーワードの音声認識率[%]

| | 置換 | 挿入 | 脱落 | 正解 | 正解精度 |
|---|---|---|---|---|---|
| 話者1 | 11.5 | 30.8 | 0.0 | 88.5 | 57.7 |
| 話者2 | 12.5 | 23.1 | 0.0 | 87.5 | 64.4 |

表5: キーワード列での認識結果
(注:括弧内数値は割合)

正解1:認識結果の1-bestのみの正解数
正解3:認識結果の3-bestまでの正解数
準正解:挿入を許す場合の正解数
(正解1, 正解3は、置換・挿入は不正解とする)

| | 正解1 | 正解3 |
|---|---|---|
| 話者1 | 29(53.7%) | 32(59.3%) |
| 話者2 | 26(48.1%) | 29(53.7%) |

| | 準正解1 | 準正解3 |
|---|---|---|
| 話者1 | 45(83.3%) | 48(88.9%) |
| 話者2 | 42(77.8%) | 44(81.5%) |

表6: データベース(A)に対する検索結果(音声入力)

(a)1-bestの結果のみを用いた場合

| 話者 | 再現率[%] | 適合率[%] | 検索率[%] |
|---|---|---|---|
| 話者1 | 65.1 | 29.2 | 76.7 |
| 話者2 | 58.7 | 43.1 | 76.7 |

(b)3-bestまでの結果を用いた場合

| 関連度 | 話者 | 再現率[%] | 適合率[%] | 検索率[%] |
|---|---|---|---|---|
| 共起頻度 | 話者1 | 86.5 | 57.4 | 86.7 |
| | 話者2 | 81.2 | 47.3 | 86.7 |
| 相互情報量 | 話者1 | 89.7 | 68.0 | 90.0 |
| | 話者2 | 83.6 | 58.5 | 86.7 |

表7: データベース(B)に対する検索結果(音声入力)

(a)1-bestの結果のみを用いた場合

| 話者 | 再現率[%] | 適合率[%] | 検索率[%] |
|---|---|---|---|
| 話者1 | 48.4 | 28.6 | 73.3 |
| 話者2 | 55.6 | 18.0 | 73.3 |

(b)3-bestまでの結果を用いた場合

| 関連度 | 話者 | 再現率[%] | 適合率[%] | 検索率[%] |
|---|---|---|---|---|
| 共起頻度 | 話者1 | 70.4 | 32.2 | 83.3 |
| | 話者2 | 69.5 | 22.6 | 86.7 |
| 相互情報量 | 話者1 | 71.7 | 47.5 | 90.0 |
| | 話者2 | 68.6 | 37.5 | 86.7 |

#### (b) 検索実験

キーワードの音声認識結果を入力キーワードとして、検索実験を行なった。データベース(A)を使った場合と、データベース(B)を使った場合との検索実験で、検索結果にどれくらいの違いが現れるかを調べた。

キーワードの認識結果の1-bestのみを用いた場合と、3-bestまでを用いた場合とで検索実験を行なった。3-bestの場合、同じ読みの単語や認識により発生した余計な単語も一緒にキーワード候補として入力した。例えば、"フィルム 史上"を入力したとすると、"史上(しじょう)"と同じ読みの単語("市場"、"市上"、"試乗")を辞書から検索し、キーワード候補とする。このままでは、ユーザーが意図しない記事が検索されたり、また、欲しい記事が検索されないということになるので、2.2節で述べたキーワード候補間の関連度(共起情報、相互情報量)を用いてキーワード候補のグルーピングを行ない、そのグループ内の候補を使って検索を行なう。今回の検索実験では、複数のグループの中で一番多くの候補をもっているグループを使用した。これに対して、1-bestの場合は認識結果そのものを入力とし、同音意義語は考えない。また、グルーピング処理も行なっていない。

ある記事に対するキーワード群を用いて(A),(B)2種類のデータベースに対しての記事の検索結果を表3, 表6, 表7に示す。表3はデータベース(B)に

対してキーワードがテキスト入力の場合、表 6はデータベース (A) に対してキーワードが音声入力の場合、表 7はデータベース (B) に対してキーワードが音声入力の場合である。検索対象記事は全部で30 記事で、再現率、適合率、検索率は全体の平均である。テキスト入力 (表 3) と音声入力 (表 7(b)) とは性能にあまり差のないことがわかる。

表 6, 表 7で、キーワード (列) の認識率が高かったので、再現率、検索率とも比較的高い値になっている。しかし、適合率が低くなっているということから、余計な記事の湧きだしがかなり多くなっていることが分かる。また、表 3, 表 7のデータベース (B) に対しての検索結果に対しても同様なことが言えるが、両方とも音声の場合は、若干再現率が落ちている。これは、キーワードの認識が完全でないこと、グルーピングが必ずしもうまくいっていないことが考えられる。グルーピングは余計な単語を取り除くのには効果があるが、取り除きすぎ (単語間の関連度が低いとき) になる場合もある。また、正解でない候補どうしでグループを構成する場合もあった。

そこで、グルーピングが本当に有用であるかを確かめるため、3-best までのキーワードの認識仮説を用い、グルーピングを行わずに実験を行った。結果を、表 8に示す。実験結果を見ると、表 8(a)(b) ともグルーピングありの結果よりも再現率が10%程度高くなっている。しかし、湧き出した記事が非常に多かったため、適合率が非常に低くなった。この結果から、再現率の点においては、グルーピングを行った方が若干劣るけれども、適合率の点では大幅な改善が見られた。つまり、グルーピングによって関係のないキーワードがうまく除去されているということが分かる。本手法は不必要な記事をできるだけ検索しないようにするためには、たいへん有効な手法であると言える。

表 8: グルーピングなしでの検索結果

(a) データベース (A) に対する結果

| 話者 | 再現率 [%] | 適合率 | 検索率 [%] |
|---|---|---|---|
| 話者 1 | 95.4 | 16.3 | 76.7 |
| 話者 2 | 85.7 | 16.2 | 76.7 |

(b) データベース (B) に対する結果

| 話者 | 再現率 [%] | 適合率 | 検索率 [%] |
|---|---|---|---|
| 話者 1 | 84.8 | 6.0 | 66.7 |
| 話者 2 | 79.4 | 7.2 | 63.3 |

次に、1-best のみ使用した場合と、3-best までを使用した場合とでは、再現率、検索率ともに 3-best までの方が5〜15%程良くなっている。また適合率に関しては明らかに 3-best の方がよい。これは、キーワードの認識率 (表 5) より、3-best までを考慮した方が認識率が高くなること、グルーピングを行うことによる余計なキーワード候補の除去がうまく働いているためである。この結果から、1-best のみの結果を使用するよりも、3-best までの結果を使用する方が良いことがわかる。

最後に表 6, 表 7で、関連度の尺度を共起表現を用いた場合と、相互情報量を用いた場合での結果を載せているが、相互情報量を用いた方が、適合率が良く (つまり、記事の湧きだしが少なくなっている)、検索率も若干良い。再現率はほとんど変わらなかった。

## 4 むすび

今回、ニュース音声データベースから、ニュース記事の検索システムを作成し、音声認識による書き起こしのデータベースを用いても検索能力が高いことを示した。また、キーワードが音声入力の場合に考えられるキーワード候補の増大に対処する方法を提案した。実際に音声入力による実験では、提案したグルーピング手法を用いキーワード候補を絞り込むことで、高い検索率を得られることがわかった。しかし、多数の余計な記事が検索されている (適合率が低い) ので、グルーピングの改良 (関連度の尺度など) が必要である。

本稿では、検索を行なう前にキーワード候補を絞り込んだが、その検索結果をさらに絞り込む方法として、音響的類似性などを使った方法を試みたい。また、同義語 (例えば、首相⇔総理大臣) や固有名詞の取り扱い方も検討していく必要がある。

## 謝辞

この研究は、NHK 放送技術研究所のニュース音声データベース、ニューステキストデータベースを使わせていただいた。これらのデータベースを提供された NHK 放送技術研究所の関係諸氏に深く感謝する。

## 参考文献

[1] A.G. Hauptmann,H.D. Wactlar: Indexing and Search of Multimodal Information, Proc.ICASSP,pp.195–198(1997)

[2] Dave Abberley,Steve Renals,Gary Cook: Retrieval of Broadcast News Documents with the THISL System, Proc.ICASSP,pp.3781–3784(1998.5)

[3] Kenney NG: Towards Robust Methods for Spoken Document Retrieval, Proc.ICSLP, pp.939–942(1998.12)

[4] 長尾 真編:自然言語処理, 岩波書店 (1996)

[5] 福島, 赤峯:全文検索システム Retrieval Express の開発と評価, 言語処理学会, 第 3 回年次大会,pp.361–364(1997.3)

[6] 赤松, 花井, 甲斐, 峯松, 中川:新聞・ニュース文をタスクとした大語彙連続音声認識システムの評価, 情報処理学会, 第 57 回全国大会, pp.35–36(1998.10)