社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2005－ＩＣＳ－139（７）
2005／3／14

音素片のカーネル主成分分析を用いたトピックセグメンテーション

佐土原　健†　　李　　時旭†　　児島　宏明†

† 産業技術総合研究所
〒 305-8568 茨城県つくば市梅園 1-1-1 つくば中央第二
E-mail: †sadohara@computer.org

あらまし　本論文では，語彙制約を用いることなしに，入力音声を意味的に等質な部分に分割する手法を提案する．この手法は，大語彙連続音声認識システム等によって，キーワードを抽出することなく，音声を，音素よりも粒度の細かい音素片の列として認識した上で，直接トピックセグメンテーションを行う．これにより，一定長以下の任意の音素片列に基づいた，語彙と文法に制約されないトピックセグメンテーションが可能になる．また，カーネル主成分分析を用いて，一つのトピックにおいて共起する音素片列を，まとめて一つの基底とすることによって，各分析区間を表現することも本手法の特徴である．これにより，ベクトルの余弦が，トピックに関する類似性を反映することになり，この余弦を類似性の指標として用いる階層的クラスタリング法により，トピック単位のクラスタリングを行う．また，このような手法の有用性を，ニュース音声のトピックセグメンテーションの実験により示す．

キーワード　主成分分析，カーネル法，トピックセグメンテーション，クラスタリング，音声認識

Topic Segmentation Using Kernel Pricipal Component Analysis for Sub-Phonetic Segments

Ken SADOHARA†, Shi-wook LEE†, and Hiroaki KOJIMA†

† National Institute of Advanced Industrial Science and Technology (AIST)
AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba-shi, Ibaraki, Japan
E-mail: †sadohara@computer.org

Abstract　This paper describes an open-vocabulary method for segmenting spoken documents into topically homogeneous blocks. Without transcribing the spoken documents into texts, the method builds the topical clusters directly from recognized sub-phonetic segments, and thus it is not constrained in term of vocabulary or grammar. Each analysis interval constituting the clusters is represented as a vector in a high dimensional space spanned by all sub-phonetic segments with given length. Then a kernel principal component analysis reduces the dimensionality by grouping co-occurred sub-phonetic segments in each topic. This yields that cosine similarity between vectors is related with topical similarity, and the hierarchical clustering method using the similarity measure is expected to form topically homogeneous clusters. In fact, effectiveness of the method is shown in an experiment on topic segmentation of broadcast news.

Key words　principal component analysis, kernel methods, topic segmentation, clustering, speech recognition

1. はじめに

今日，大容量記憶装置技術の進歩により，企業や社会はもちろん一個人に至るまで，テキスト，音声そして画像等の情報が大量に蓄積可能になった．しかし，その一方で，蓄積された情報の資源化技術は未だ発展途上であり，適切な構造化や索引化により，概要を素早く把握したり，欲しい情報に素早く到達するための技術が切実に求められている．特に，音声や画像の資源化技術は，テキストの資源化技術に比べて，その意味構造把握の困難さ故に，遅れをとっている．

本論文は，音声を手掛りとして，マルチメディア情報を，意味的に等質な部分に分割する技術について考察する．

このような技術については，これまでにも，例えば，ニュース音声等を対象として，多くの研究がなされてきた [1]〜[3]．これらの研究においては，大語彙連続音声認識システム等を用いて，キーワードを抽出し，分析区間を，これらキーワードを用いて

表現した後, 教師ありトピックセグメンテーションにおいては, トピックモデルとの類似性を, 教師なしトピックセグメンテーションにおいては, 隣接する分析区間との類似性を比較検討することで, 分析区間をトピック単位にまとめ上げる.

しかし, 現在広く用いられている音声認識システムでは, トップダウンに与えられる単語辞書や言語モデルに基づいた, 音韻列に対する音響的尤度によって, 認識処理が行なわれるため, 言語依存性が高く, 文法や語彙などの制約が不可欠である. したがって, トピックセグメンテーションにおいても, 辞書にないキーワードを用いてセグメンテーションを行うことができない. また, 認識段階での認識誤りが, 引き続くトピックセグメンテーションに悪影響を与えてしまう.

本論文では, このような問題点に対処するために, 入力音声を, 通常の音素よりも粒度が細かく, 言語依存性の低い音素片 [7] の列として認識し, この音素片の列から直接トピックセグメンテーションを行う手法を提案する. 音素片は, 通常の音素よりも粒度が細かいので, 誤認識によって必要な情報がそっくり抜け落ちてしまう危険性が少ない. また, 一定長以下の任意の音素片列をキーワードとすることにより, 固有名詞や, 省略語等の辞書に登録されていない未知語や, 一定長以下の任意のフレーズに基づくトピックセグメンテーションが可能になる.

このようなトピックセグメンテーションを実現するために, 本研究では, 分析区間に含まれる, 一定長以下の任意の音素片列を基底とするベクトルとして分析区間を表現する. さらに, 主成分分析を用いて, 主成分を基底とするより低次元のベクトルに変換する. このような手法は, テキストの索引化の分野においては, Latent Semantic Indexing (LSI) [4] と呼ばれる. このような基底の変換を行うのは, 次元の縮小により計算を容易にするという理由の他に, 各トピックにおける音素片列の共起構造に基づく, トピックに関する類似性が反映された表現に変換するという狙いもある. 例えば, 同一トピックに分類可能な分析区間であっても, 同義語の存在により, 同じ単語を共有しているとは限らないので, ベクトルの余弦等を計算するだけでは, トピックに関する類似性を見い出すことができない可能性がある. さらに, 音声を認識して得られる音素片列を対象とする場合, この問題はより深刻になる. 何故ならば, 音声の場合, 同一の単語が複数回発声されたとしても, 全く同一の記号列に認識される可能性は極めて低く, 上述した同義語と同質の問題が, より顕著に現われると考えられるからである. そこで, 主成分分析を用いて, あるトピックにおいて共起する音素片列をまとめて一つの主成分とするような基底の変換を行うことで, トピックに関する類似性を反映した表現を得ることが期待できる.

ただし, 音素片列を基底するベクトルは, 非常に高次元であるので, これを直接主成分分析することは, 計算量的に困難である. そこで, 本研究では, カーネル主成分分析 [5] を用いる. M 個の d 変量ベクトルの主成分分析は, $d \times d$ 共分散行列の対角化を行う必要があるが, カーネル主成分分析を用いると, 各ベクトルの内積を計算した $M \times M$ の行列 $\boldsymbol{K}_{ij} = \langle \boldsymbol{x}_i \cdot \boldsymbol{x}_j \rangle$ の対角化により, 高々 M 個の主成分計算することができ, $M \ll d$ である場合, 計算が容易になる. また, 内積の計算も, ストリングカーネル [6] を用いると, d に依存せずに, 音素片列の長さに対して線形な計算量で計算可能である.

このように, 各々の分析区間を, 主成分分析により抽出された主成分を基底するベクトルとして表現した上で, ある種の補正を加えた後, 階層的クラスタリング法により, 入力音声のトピックセグメンテーションを行う. 階層的クラスタリング法を用いることにより, 得られたクラスタは階層化されており, 必要に応じて大きな構造から細かな構造までを段階的に見ることを可能にするので, ユーザーが入力音声の概要を素早く把握することに寄与すると考えられる.

本論文では, このような語彙制約のない音声のトピックセグメンテーション手法を提案すると同時に, ニュース音声を対象として, セグメント境界の精度に関する簡単な評価実験を行った. その実験結果は, 厳密なセグメント境界を得ることは難しいが, どの部分で, どのような内容が話されているかを把握可能な程度の, おおまかな音声の分節化の可能性を示唆している.

2. トピックセグメンテーション

音声のトピックセグメンテーションは通常, 以下のような手順で行われる [2], [3].

(1) トピックとキーワードの関連度を表現するトピックモデルを構築する.

(2) 音声認識により, キーワードを抽出する.

(3) 任意の分析区間をキーワードを用いて表現し, トピックモデルに照らして, 各トピックとの類似度を計算する.

(4) 分析区間をずらしながら, 3 の計算を行い, 各トピックの時間関数を求め, この関数の値が大きな部分を該 当するトピックの区間として切り出す.

このような手法は, あらかじめ, トピックを付与されたデータを必要とするという意味で, 教師ありトピックセグメンテーションと呼ばれる.

一方, トピックとの類似度の計算 (3) を行う代りに, 各分析区間を, キーワードの重要度を成分とするベクトルとして表現し, 隣接する分析区間において, このベクトルが類似していればトピックが継続していると判断し, 類似していなければトピック境界と判断する手法も提案されている [1]. このような手法は, あらかじめ, トピックを付与されたデータを必要としないという意味で, 教師なしトピックセグメンテーションと呼ばれ, あらかじめ与えられたトピック以外のトピックを切り出せることが利点である.

本論文で提案するトピックセグメンテーションは, 教師なしトピックセグメンテーションであるが, 音声認識によるキーワード抽出 (2) を必要としないという点で, 上述した両手法とも異る. 本手法は, 次節で延べる音素片の列, すなわち通常の音素よりも粒度の細かい音韻列から, 直接トピックセグメンテーションを行う. これにより, 未知語を含む任意のフレーズをキーワードとして, トピックセグメンテーションを行うことが可能になる.

3. 音 素 片

現在広く用いられている音声認識システムでは, トップダウ

ンに与えられる単語辞書や言語モデルに基づいた，音韻列に対する音響的尤度によって，認識処理が行なわれる．現在の音声認識システムの高精度化は，このような言語的知識の積極的な利用によるところが大きいが，その反面，言語依存性が高くなり，文法や語彙などの制約が不可欠になってしまうという問題がある．上述したトピックセグメンテーションにおいても，辞書にないキーワードを用いてセグメンテーションを行うことができない．また，認識段階での認識誤りが，引き続くトピックセグメンテーションに悪影響を与えてしまう．

本論文では，このような問題点に対処するために，入力音声を，通常の音素よりも粒度が細かく，言語依存性の低い音素片 [7] の列として認識し，この音素片の列から直接トピックセグメンテーションを行う手法を提案する．音素片は，既に，音声文献の検索に適用され，その有効性が示されている [8]．図 1 は，実際のニュース音声の中の「ニュースです」という音声を音素片の列として表現したものであり，連続する 3 つの音素片が音素に対応している．認識誤りや，ノイズも見受けられるが，通常の音

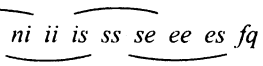

図 1 音 素 片

素よりも粒度が細かいので，必要な情報がそっくり抜け落ちてしまう危険性が少ない．また，長さ p 以下の任意の音素片列をキーワードとすることにより，固有名詞や，省略語等の辞書に登録されていない未知語や，長さ p までの任意のフレーズに基づくトピックセグメンテーションが可能になる．

本論文では，任意の音素片列が，ある分析区間 S_j においてどの程度特徴的であるかを示す指標を成分とするベクトルとして，各分析区間を表現し，分析区間ベクトルと呼ぶことにする．具体的には，S_j に現われる，長さが d $(0 < d \leqq p)$ の音素片列 σ に対して，

$$d \log N(S_j, \sigma) - \log N(S, \sigma)$$

という指標を用いている．ここで，$N(S_j, \sigma)$ は，区間 S_j における音素片列 σ の頻度であり，S は，全区間を表わす．

しかし，このベクトルの余弦は，トピックに関する類似性を必ずしも反映していない．例えば，同じトピックに分類可能な分析区間であっても，同義語の存在により，同じ単語を共有しているとは限らず，その場合，これら分析区間は類似性を見出すことができない可能性がある．さらに，音素片列に含まれる認識誤りやノイズは，この問題を一層深刻なものとする．何故ならば，音声の場合，同じ単語が複数回発声されたとしても，それらが同一の記号列として認識される可能性は非常に低く，上述した同義語の問題がより顕著に現われる考えられるからである．

こうした問題に対処するために，テキストの索引化の分野で利用されている Latent Semantic Indexing (LSI) 法 [4] に習い，主成分分析を用いて，音素片列の共起構造を考慮した新たな基底を抽出し，各分析区間を，トピックに関する類似性を反映した低次元のベクトルで表現する．

ただし，各分析区間ベクトルは非常に高次元 (注1) であるので，主成分分析を直接扱うことは，計算効率とメモリ効率の観点から問題が大きい．そこで，次節で述べるカーネル主成分分析を用いて，陽に分析区間のベクトルを計算することなく，各分析区間の内積のみを計算することで，任意の音素片列が張る高次元空間で主成分分析を行う．

4. カーネル主成分分析

カーネル主成分分析は，主成分分析の非線形拡張の一種として提案されたもので [5]，任意の写像 ϕ により，データ $\boldsymbol{z}$ を写像した空間で，陽に $\phi(\boldsymbol{z})$ を計算することなく主成分分析を行う．

我々の問題の場合，分析区間 S を分析区間ベクトルに変換する写像を ϕ と考え，分析区間ベクトルの空間で主成分分析を行う．すなわち，M 個の分析区間ベクトル $\boldsymbol{x}_i (i = 1, \ldots, M)$ が与えられるとき，共分散行列

$$C = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{x}_i \boldsymbol{x}_i^{\top}$$

に対して，

$$\lambda \boldsymbol{V} = C\boldsymbol{V}, \lambda \geqq 0, \boldsymbol{V} \neq \boldsymbol{0} \tag{1}$$

を解く．ただし，このとき，分析区間ベクトルはセンタリングされていると仮定する．すなわち，

$$\sum_{i=1}^{M} \boldsymbol{x}_i = \boldsymbol{0}. \tag{2}$$

$\boldsymbol{x}$ の次元を d とすれば，C は $d \times d$ 行列であるので，$M \ll d$ の場合，これを直接解くことは計算量的に困難である．

これに対して，

$$\boldsymbol{V} = \sum_{i=1}^{M} \alpha_i \boldsymbol{x}_i$$

と書けて，さらに，

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)^{\top}$$
$$\boldsymbol{K}_{ij} = \langle \boldsymbol{x}_i \cdot \boldsymbol{x}_j \rangle$$

とおくと，

$$M\lambda\boldsymbol{\alpha} = \boldsymbol{K}\boldsymbol{\alpha}$$

の解は，(1) の解でもあることが知られている．

$\boldsymbol{K}$ は，実対称行列かつ半正定値であり，$\boldsymbol{K}$ の ℓ 個の非零の固有値 $\lambda^1 \geqq \cdots \geqq \lambda^\ell > 0$ に対応する固有ベクトル $\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^\ell$ を用いると，k $(1 \leqq k \leqq \ell)$ 番目の主成分は，

$$\langle \boldsymbol{V}^k \cdot \boldsymbol{x} \rangle = \sum_{i=1}^{M} \alpha_i^k \langle \boldsymbol{x}_i \cdot \boldsymbol{x} \rangle$$

と表現できる．ここで，$\boldsymbol{V}^k$ は正規化されている必要があるが，

(注1)：現在用いている音素片の数は 411 個 で，分析区間ベクトルの次元は $O(411^p)$.

これは, $\boldsymbol{\alpha}^k$ を $\lambda^k \langle \boldsymbol{\alpha}^k \cdot \boldsymbol{\alpha}^k \rangle = 1$ を満すようにすることで実現できる. 従って, M 個のデータは, ℓ 個の主成分を用いて,

$$\left[\boldsymbol{x}'_1 \cdots \boldsymbol{x}'_M\right]^\top = K \left[\boldsymbol{\alpha}^1 \cdots \boldsymbol{\alpha}^p\right]$$

と書ける.

さらに, ここまでの議論は, (2) を仮定しているが, この仮定も,

$$\tilde{\boldsymbol{K}} = \boldsymbol{K} - \boldsymbol{U}\boldsymbol{K} - \boldsymbol{K}\boldsymbol{U} + \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}, \quad \boldsymbol{U}_{ij} = \frac{1}{M}$$

のように, 行列 $\boldsymbol{K}$ を補正することで実現できることが知られている.

以上の議論から, $\boldsymbol{K}$ が計算できれば, 主成分分析を通して次元の縮小を行い, データを ℓ ($\ell \leq M$) 個の主成分を基底とするベクトルに変換することが可能になる. 関数 $\boldsymbol{K}_{ij} = \langle \phi(S_i) \cdot \phi(S_j) \rangle$ は, カーネル関数と呼ばれる. 特に S が文字列の場合には, ストリングカーネルと呼ばれ, 非連続な部分列を許すか否かや, 文字のソフトマッチングを許すか否か等の, 様々な種類のストリングカーネルが提案されている [6]. これらの中で, 本論文では, 連続する長さ p までの部分文字列が張る空間の内積を, トライを用いて $O(p(|S_i| + |S_j|))$ で計算するカーネル関数を用いた. ただし, 分析区間ベクトルにおいては, 音素片列のその区間における頻度に加えて, 全区間の頻度 $N(S, \sigma)$ を計算する必要があるため, 事前に計算した全区間の頻度を参照する計算が余分に必要となる. 表 1 に, $\boldsymbol{K}_{ij}$ の計算手続きの概略を示す.

入力: S_i, S_j, p
$L_i(\epsilon) = \{(s,0) | s$ は, S_i に含まれる長さ p の部分文字列 $\}$
$L_j(\epsilon) = \{(s,0) | s$ は, S_j に含まれる長さ p の部分文字列 $\}$
$\boldsymbol{K}_{ij} = 0$
kernel(ϵ, 0)

procedure kernel(v, d)
  If $d \leq p$ かつ $L_i(v) \neq \emptyset$ かつ $L_j(v) \neq \emptyset$
    If $d > 0$
      $f_i = d \log |L_i(v)| - \log N(S, v)$
      $f_j = d \log |L_j(v)| - \log N(S, v)$
      $\boldsymbol{K}_{ij} = \boldsymbol{K}_{ij} + f_i f_j$
    If $d < p$
      for each $(u,k) \in L_i(v)$, $(u, k+1)$ を $L_i(vu_{k+1})$ に追加
        ただし, u_{k+1} は u の $k+1$ 文字目を表わす.
      for each $(u,k) \in L_j(v)$, $(u, k+1)$ を $L_j(vu_{k+1})$ に追加
      for each $a \in \Sigma$ kernel(va, $d+1$)

表 1 カーネル関数の計算アルゴリズム

5. 階層的クラスタリング

提案手法おいては, 入力音声は, まず, M 個のセグメントに分割される. そして, 図 2 のように, 連続する ℓ 個のセグメントからなる $M - \ell + 1$ 個の分析区間を前節で述べたカーネル主成分分析にかけ, 主成分で記述された分析区間ベクトルを得る. 以降, 本論文では, $\ell = 3$ とする. 分析区間ベクトルは隣接するベクトルと重複しているので, 重複部分を以下のような

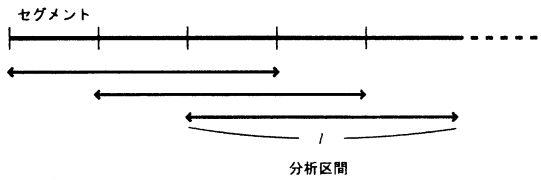

図 2 分析区間

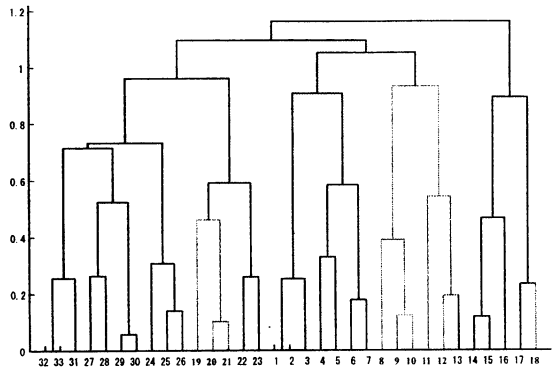

図 3 樹形図

補正で取り除き, 各セグメントに対応するセグメントベクトル $\boldsymbol{s}_i$ $(i = 1, \ldots, M)$ を得る.

$$\boldsymbol{s}_i = \begin{cases} \boldsymbol{x}_1 & i = 1 \text{ のとき} \\ \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2}{2} & i = 2 \text{ のとき} \\ \frac{\boldsymbol{x}_{i-2} + 2\boldsymbol{x}_{i-1} + \boldsymbol{x}_i}{4} & 2 < i < M-1 \text{ のとき} \\ \frac{\boldsymbol{x}_{M-3} + \boldsymbol{x}_{M-2}}{2} & i = M-1 \text{ のとき} \\ \boldsymbol{x}_{M-2} & i = M \text{ のとき} \end{cases}$$

このようにして得られたセグメントベクトル $\boldsymbol{s}_i$ を, 階層的クラスタリング法により, 距離が近いものからボトムアップにクラスタにまとめ上げる. 本論文では, その際, 以下で定義される距離を用いた.

$$d_0(\boldsymbol{s}_i, \boldsymbol{s}_j) = 1 - \frac{\langle \boldsymbol{s}_i \cdot \boldsymbol{s}_j \rangle}{\sqrt{\langle \boldsymbol{s}_i \cdot \boldsymbol{s}_i \rangle} \sqrt{\langle \boldsymbol{s}_j \cdot \boldsymbol{s}_j \rangle}}$$

$$d(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{m=1}^{|C_i|} \sum_{n=1}^{|C_j|} d_0(\boldsymbol{s}_m^i, \boldsymbol{s}_n^j)$$

ただし, C_i, C_j はセグメントベクトルの集合とし, $\boldsymbol{s}_m^i$ は, 集合 C_i の m 番目の要素を表わす.

図 3 は, 5 分のニュース音声に階層的クラスタリング法を適用して得られた樹形図である. 横軸の数字 k は, 総数 31 個のセグメントの内, k 番目のセグメントを表わしている. また, 縦軸は, 上で定義したクラスタ間の距離を表わしている. なお, この例では, 大きく 5 つのトピックを含んでおり, 図では, 5 つのトピックが変わるごとに書体を交互に変えている.

階層的クラスタリングを用いたことにより, 得られたクラスタは階層化されており, 必要に応じて大きな構造から細かな構造までを段階的に見ることを可能にするので, ユーザーが入力音声の概要を素早く把握することに寄与すると考えられる.

6. 実　　験

本研究では，NHK の 15 分のニュース番組 6 回分を対象にして実験を行った．各々の番組は，人手により，基本的に 1 文ごとに 1 つのセグメントに分割したが，音素片の認識エンジンの実装上の都合により，20 秒を超える文章は，適当な無音区間で，20 秒未満のセグメントに分割した．なお，1 つのセグメントを 1 つの文章にしなければいけない本質的な理由は存在しないが，今回の実験では，以下で述べるようにトピックの境界を，どの程度正確に予測できるかを調べることを目的としているので，セグメントの中に境界が現われないように，このようなセグメントの分割を行った．表 2 は，これら 6 つの番組の詳細を示している．なお，トピックの境界の判定については，ニュース映像のキャプションが変化した場所を，トピックの境界と考えた．

番組	セグメント数	トピックの数	トピックの平均長
1	75	11	6.8
2	71	10	7.1
3	71	11	6.5
4	73	12	6.1
5	73	11	6.6
6	83	13	6.4
平均	74.3	11.3 (セグメント数)	6.6

表 2　番組の詳細

これらの番組に対して，前節までに述べたカーネル主成分分析と階層的クラスタリングを適用した．その結果，得られた樹形図に対して，あるクラスタ間の距離 d が与えられると，d よりも距離の大きなクラスタを分離することができるが，このときに生じるセグメント境界の再現率と適合率を評価した．図 4 は，d を変化させたときの，再現率と適合率の変化の様子を示している．図中，誤差 0 とラベル付けされたプロットが，通常の ROC カーブであるのに対して，誤差 1 とラベル付けされたプロットは，前後に 1 セグメントずれた境界であっても正解とした場合の ROC カーブである．また，図には，誤差 0 の場合に，ランダムに選んだ境界が正しい境界である確率 0.141 と，誤差 1 の場合の確率 0.196 が示されている．誤差 0 のプロットを見ると，提案手法で，厳密な境界を求めることは難しいことが分かる．しかし，誤差 1 のプロットを見ると，どの部分で，どのような内容が話されているかを把握可能な程度の，おおまかな構造に分割できる可能性はあるように思われる．

7. おわりに

本論文では，音声を手掛りとして，マルチメディアコンテンツを構造化し，コンテンツの概略を素早く把握することを可能にする技術の開発を目的として，入力音声を意味的に等質な部分に分割する手法を提案した．この手法は，大語彙連続音声認識システム等によるキーワード抽出を行うことなく，音声を，音素よりも粒度の細かい音素片の列として認識した上で，直接トピックセグメンテーションを行う．これにより，一定長以下の任意の音素片列に基づいた，語彙と文法に制約されないトピックセグメンテーションが可能になる．このようなセグメンテーションを実現するために，任意の音素片列が張る高次元空間のベクトルで表現される分析区間に対して，カーネル主成分分析を適用し，一つのトピックにおいて共起する音素片列を，まとめて一つの基底とすることによって，各分析区間の次元の縮小を行う．これにより，ベクトルの余弦は，トピックに関する類似性を反映し，この余弦を類似性の指標として用いる階層的クラスタリング法により，トピックセグメンテーションを行う．このような手法を用いて，ニュース音声のトピックセグメンテーションの実験を行い，トピック境界の精度を調べた．その結果，厳密な境界を得ることは難しいものの，ある程度の誤差を許容して，どの部分で，どのような内容が話されているかをおおまかに把握可能な程度の，トピックセグメンテーションの可能性を示すことができた．今後は，画像等を手がかりとする別の手法との融合を図りながら，さらなる精度向上を目指すと同時に，トピックの階層構造の正しさに関する評価等も行って行きたい．

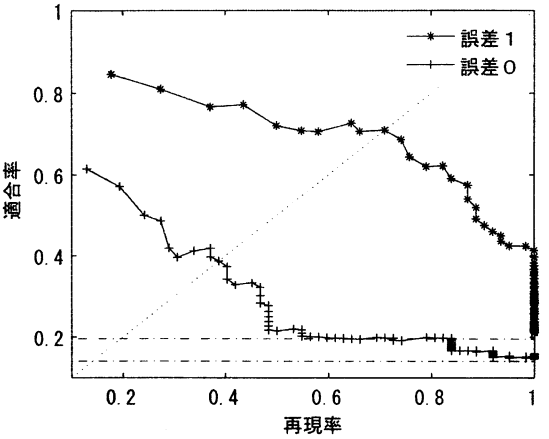

図 4　ROC カーブ

文　　献

[1] 鷹尾 誠一, 他.: "ニュース音声に対するトピックセグメンテーションと分類", 情報処理学会研究報告, 24, pp. 55–62 (1998).

[2] P. Mulbregt, et al.: "Text segmentation and topic tracking on broadcast news via a hidden Markov model approach", ICSLP, Vol. VI, pp. 2519–2522 (1998).

[3] K.Ohtsuki, et al.: "Topic extraction based on continuous speech recognition in broadcast news speech", IEICE Trans. Inf. and Syst., **E85-D**. 7, pp. 1138–1144 (2002).

[4] S. Deerwester, et al.: "Indexing by latent semantic analysis", Journal of the American Society of Information Science, **41**, 6, pp. 391–407 (1990).

[5] B. Schölkoph, A. Smola and K.R.Müller: "Nonlinear component analysis as a kernel eigenvalue problem", Neural Computation, **10**, 5, pp. 1299–1319 (1998).

[6] J.Shawe-Taylor and N.Cristianini: "Kernel methods for pattern analysis", Cambridge University Press (2004).

[7] K.Tanaka, et al.: "Speech data retrieval system constructed on a universal phonetic code domain", Proc. of ASRU2001, pp. 1–4 (2004).

[8] S. Lee, et al.: "Combining multiple subword representations for open-vocabulary spoken document retrieval", to appear in ICASSP (2005).