アナウンサー発話の自動抽出とディクテーションによるニュース記事分類

西田昌史[†] 緒方 淳[†] 有木康雄[†]

ビデオ・オン・デマンドを目指したニュースデータベースを構築するには，ニュース記事を分類しておく必要がある．本論文では，ニュース音声に対してディクテーションを行い，キーワードを抽出することにより，自動的に記事分類を行う手法を提案している．ニュース記事を分類するには，記事中のキーワードと分類分野との関係をもとに分類を行う．このため χ^2 法によりキーワードを自動選択し，キーワードと分類分野との関連度を求め，記事分類を行っている．記事を分類するうえでは，アナウンサーの発話区間のみをディクテーションすれば十分であり，処理時間の短縮につながる．しかし，人手でニュース音声中からアナウンサーの発話区間を切り出すのは現実的ではない．そこで，本論文では，話者照合に基づきアナウンサーの発話区間のみを自動的に抽出する方法を提案している．NHK 5分間のニュース48記事に対して，アナウンサーの発話区間を自動抽出し，この区間に対してディクテーションして，記事を自動的に分類する実験を行った．その結果，分類精度を下げることなく処理時間を短縮できることを確認した．

News Article Classification by Speech Dictation for Automatically Extracted Announcer Utterance

Masafumi Nishida,[†] Jun Ogata[†] and Yasuo Ariki[†]

In order to construct a news database with a function of video on demand (VOD), it is required to classify news articles into topics. In this paper, we propose a system which can automatically dictate news speech, extract keywords and classify news articles into topics based on the extracted keywords. We employed χ^2 method to select keywords and to compute the association degree between keywords and topics. We also propose to dictate only the announcer utterance for classifying the news articles because it contributes to save the dictation time. In order to segment the announcer speech section from other speakers, we propose a speaker verification method based on subspace method. For 48 NHK news articles, we carried out the extraction of announcer utterance, speech dictation and article classification. As a result, we reduced the dictation time by restricting the dictation to the announcer utterance without losing the classification accuracy.

1. はじめに

近年，放送の多チャンネル化により，多くのニュース番組が放映されるようになった．これを受けて視聴者には，知りたいニュースだけを見たいという要求が生じている．この要求に対応するには，ニュース記事を分類してデータベースを構築しておく必要がある．このとき，人手でニュース記事を分類することは不可能であり，機械によるニュース記事の自動分類が望まれる．この点から，ニュース音声に対する話題分類，話題同定の研究が行われている[1)~4)]．

本研究では，ニュース音声に対してディクテーションを行い，キーワードを抽出することにより，自動的に記事分類を行うことを目的としている．ニュース音声は，レポータあるいはインタビュアーの発話部分に比較的雑音が重畳している場合が多いので，レポータあるいはインタビュアーの発話を正しくディクテーションすることは困難である．したがって，比較的雑音が少ないアナウンサーの発話区間のみをディクテーションすることにより，記事の分類に必要なキーワードを精度良く抽出できると考えられる．また，アナウンサーの発話のみをディクテーションすることは，処理の短縮にもつながる．この点から，本論文ではアナウンサーの発話区間のみを自動的に抽出してディクテーションし，ニュース記事を自動分類する方法を提

† 龍谷大学理工学部
Faculty of Science and Technology, Ryukoku University

案する．

本研究では，ニュース音声中からアナウンサーの発話区間を自動的に抽出する方法として，セキュリティ応用を目的に研究されている話者照合[5)~7)]の技術を用いている．現在，話者照合は，技術的にすでに高い照合精度が得られており，セキュリティ応用上，時期差に対するロバストネスが研究課題となっている[8),9)]．したがって，この話者照合技術は，時期差を問題としない限り，音声メディア処理として応用可能な技術である．たとえば，座談会などで特定の人の発言を拾い出して聞くといった目的にも利用可能である[10)]．この点に関しては，討論会において話者ごとに発話区間を切り出すインデキシングの研究が報告されている[11)]．この方法は，あらかじめ討論会の参加者について話者モデルを学習しておき，話者モデルを用いた話者認識により，インデキシングを行う方法である．しかし，話者モデルをあらかじめ学習しておくことは，自動化の点から望ましくない．特に，ニュース番組では，アナウンサーが日によって代わることもあり，またあらかじめアナウンサーの話者モデルを学習しておくと，時期差に対処しなくてはならなくなるからである．そこで，本研究では，特定の話者をあらかじめ学習することなく，入力音声中から各話者の話者モデルを自動学習し，話者照合に基づいて自動的に話者区間を切り出す方法を提案する．話者照合の方法としては，リアルタイム処理を目的として，部分空間法を用いている．提案手法の有効性を示すために，NHK 5分間のニュース45日分に対して，アナウンサーの発話区間の切出し実験を行った．

ニュース記事を分類するには，記事中のキーワードを抽出し，キーワードと分類分野との関係を用いるのが一般的である．ここで問題となるのは，ニュース記事の分類性能が，ニュース音声記事からキーワードを抽出する精度だけではなく，キーワードと分類分野との関係をどのように設計するかに依存している点である．我々はこれまで，人手で与えたキーワードをもとにニュース記事を分類する研究を行ってきた[12)]．今回，χ^2 法を用いてキーワードの自動選択を行い，選択されたキーワードと分類分野との関連度を計算した．記事の分類は，単語 bigram と不特定話者 HMM により，ニュース音声をディクテーションすることから始まる．ディクテーションの結果得られたキーワードをもとに，分類分野ごとにその関連度を積算する．最後に，積算された関連度が最大となる分野にニュース音声記事を分類する．

2. アナウンサーの発話区間の抽出

2.1 話者照合

話者照合[13)]とは，入力音声と同時に本人IDを入力し，入力音声がそのIDに対応する人の発話であるかどうかを判定するものである．**図1**に示すように，入力音声と本人の話者モデルとの距離が，閾値よりも小さければ本人の発話であるとして受理し，そうでなければ他人の発話であるとして棄却するものである．

2.2 主成分分析法による話者照合

本研究では，リアルタイム処理を目的として，従来法に比べて計算量の少ない部分空間法（主成分分析法）[14)]を話者照合のベースにしている．主成分分析は，多変量解析における次元削減のための手法であり，分散の小さな一次結合成分を取り除き，大きな分散を持つ成分のみでパターンを表現する手法である．**図2**に主成分分析法の概念図を示す．

主成分分析法による話者照合では，観測空間で観測される各話者の学習用特徴ベクトル x_k $(1 \leq k \leq N)$ から平均ベクトル $\mu^{(i)}$ を求め，式(1)により分散共分散行列 $R^{(i)}$ を求める．ここで，(i) は話者の識別番号を表す．

$$R^{(i)} = \frac{1}{N}\sum_{k=1}^{N}(x_k - \mu^{(i)})(x_k - \mu^{(i)})^T \qquad (1)$$

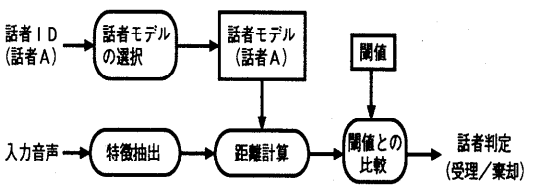


図1 話者照合の概念図
Fig. 1 Block diagram of speaker verification.

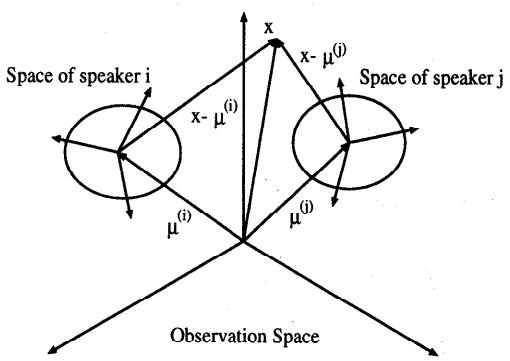


図2 主成分分析法の概念図
Fig. 2 Concept of principal component analysis.

各話者の分散共分散行列 $R^{(i)}$ を固有値分解して，固有ベクトル $v_j^{(i)}$，$j=1,2,\cdots,p^{(i)}$ を求める．この固有ベクトルで各話者の部分空間を張る．ここで，$p^{(i)}$ は部分空間の次元数を表す．この部分空間は，図1中の話者モデルに対応しており，本論文では話者部分空間と呼ぶ．

照合は，本人 A であると申告された話者の部分空間と，入力特徴ベクトルの集合 $\{x_k\}$ との平均距離 $y^{(A)}$ を式(2)により求める．ここで，N は入力特徴ベクトル x_k の総数である．

$$y^{(A)} = \frac{1}{N}\sum_{k=1}^{N}\left\|x_k-\left\{\sum_j v_j^{(A)}(v_j^{(A)T}(x_k-\mu^{(A)}))+\mu^{(A)}\right\}\right\|^2 = \frac{1}{N}\sum_{k=1}^{N}\left\|\left(I-\sum_j v_j^{(A)}v_j^{(A)T}\right)(x_k-\mu^{(A)})\right\|^2 \quad (2)$$

次に，入力特徴ベクトルの集合 $\{x_k\}$ と各話者の部分空間との平均距離 $y^{(i)}$ を求め，距離総和 $\sum_i y^{(i)}$ を計算する．最後に式(3)に示すように，$y^{(A)}$ を距離総和で正規化し，閾値 r より小さければ，本人の音声であると判定する．ここで，$y^{(A)}$ を距離総和 $\sum_i y^{(i)}$ で正規化する理由は，時期差などによる発話の特徴変動の影響を正規化して，閾値処理を頑健にするためである[5]．

$$\frac{y^{(A)}}{\sum_i y^{(i)}} < r \quad (3)$$

2.3 アナウンサーの話者区間抽出法

ここでは，異なる話者が発話したそれぞれの区間のことを話者区間と呼ぶ．同一話者がポーズをおいてしゃべっても，同一話者の区間と見なす必要がある．

部分空間法による話者照合に基づいて話者区間を切り出す方法[10]を**図3**に示し，処理内容を以下に述べる．ここで，話者を判定する閾値は，話者モデルごとに設定している．

(1) 入力音声の1秒ごとに，1秒間の平均パワーを求め，閾値処理により音声か無音かを判定する．本研究では，閾値の設定方法は，評価データとは異なるNHK 5分間のニュース5日分を用いて，無音区間と音声区間の平均パワーを求め，音声区間の抽出率が100%になるように，閾値を設定した．これにより抽出された無音から無音までの区間を発話区間と呼び，この発話区間に対して，以下の処理を行う．

(2) 最初に抽出した発話区間で話者モデル（話者部分空間）を学習するとともに，話者照合のための閾値を式(4)により設定する．この発話区間が最初の話者区間となる．

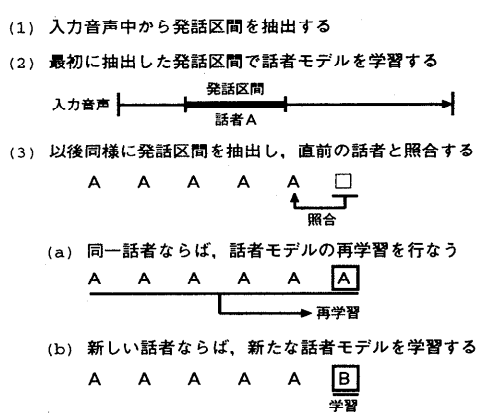


図3 話者区間の切り出し法
Fig. 3 Segmentation of speaker section.

(3) 次の発話区間に対して，直前の発話区間の話者（以後，直前の話者と呼ぶ）と照合する．ここで，入力特徴ベクトルと部分空間との平均距離に対して，式(3)に示す正規化は行っていない．なぜならば，本研究では，あらかじめ話者の数や話者モデルが既知ではなく，処理の進行につれて話者モデルの数が増えるためである．さらに，直前の話者のモデルとつねに照合を行うため，時期差による特徴変動の影響を考慮する必要がないからである．

(a) 直前の話者と判断されれば，直前までの話者区間と今抽出した発話区間を合わせて，その話者の新しい話者区間とする．この話者区間に対して話者モデルの再学習，閾値の再設定を行う．

(b) 直前の話者でないと判断されれば，直前の話者以外と照合し，閾値を下回った話者の中で最も距離が小さい話者に判定する．この場合，(a)と同様に，話者モデルの再学習，閾値の再設定を行う．一方，直前の話者以外と照合した結果，閾値を下回る話者がいなければ，新しい話者と判断し，今抽出した発話区間をその話者の話者区間とする．この話者区間に対して話者モデルの学習，閾値の設定を行う．

このように最初に抽出した発話区間で話者のモデルを学習し，以後抽出した発話区間に対して，直前の話者モデルとの照合を繰り返すことで，事前に話者モデルを作成しておくことなく，話者の数と話者区間を抽

出することが可能となる．

本研究では，ニュース音声中のアナウンサーの話者区間を自動的に抽出することを目的としている．したがって，ニュース1日分ごとに図3に示した方法により，話者区間を切り出すが，どの話者区間がアナウンサーの話者区間であるかを判定する基準を持っていない．そこで，話者区間を話者ごとに抽出した後，最も長時間発話している話者をアナウンサーと判定している．

同一話者か新しい話者かを判定する話者照合の閾値 θ は，学習に用いた音声区間中の特徴ベクトル集合と学習時に作成された話者部分空間との平均距離 μ，ならびに標準偏差 σ を求めて次式のように設定している．

$$\theta = \mu + \frac{\sigma}{3} \tag{4}$$

この閾値は，NHK 5分間のニュース5日分を設定データとして用い，値を変えて実験を行い，切出し精度が最も高かったときの値として決定している．

2.4 アナウンサーの話者区間抽出実験

2.4.1 実験条件

NHKの5分間のニュース45日分を用いて，アナウンサーの話者区間を切り出す実験を行った．音声データの作成方法は，NHK 5分間のニュースを8mmテープに録画し，8mmテープからSGI社のOctaneにより，チャンネルモノ，サンプリング周波数44.1kHz，量子化ビット数16ビットで音声を録音した後，サンプリング周波数12kHzにダウンサンプリングを行った．音声区間検出のための閾値および部分空間の次元数は，式(4)で求められる話者の閾値の設定データと同じで，評価データとは異なるNHK 5分間のニュース5日分を用いて設定した．その結果，部分空間の次元数は，最も切出し率が高かった7次元とした．また，実験に用いた音声データの平均SNRは，20.5dBであった．実験条件を**表1**に示す．実験の評価は，アナウンサーの話者区間の切出し率と適合率で評価した．これらは次式で定義される．また，切り出された発話区間において，アナウンサーの発話とその他の発話が混ざった区間は，アナウンサーの発話として評価している．

$$\text{切出し率} = \frac{\left\{\begin{array}{l}\text{アナウンサーと正しく照合した}\\ \text{発話区間数}\end{array}\right\}}{\left\{\begin{array}{l}\text{全ニュース中のアナウンサーの}\\ \text{発話区間数}\end{array}\right\}} \tag{5}$$

$$\text{適合率} = \frac{\left\{\begin{array}{l}\text{アナウンサーと正しく照合した}\\ \text{発話区間数}\end{array}\right\}}{\left\{\begin{array}{l}\text{アナウンサーとして切り出された}\\ \text{発話区間数}\end{array}\right\}} \tag{6}$$

表1 実験条件
Table 1 Experimental condition.

データ	NHK 5分間のニュース45日分
サンプリング周波数	12kHz
フレーム長	20ms
フレーム周期	5ms
窓タイプ	ハミング窓
特徴パラメータ	LPCケプストラム（16次）
部分空間の次元数	7
閾値 θ	$\theta = \mu + \frac{\sigma}{3}$

表2 アナウンサーの話者区間切出し結果（%）
Table 2 Segmentation result of announcer utterance (%).

切り出し率	92.1
適合率	91.1

実験に用いたNHK 5分間のニュース45日分すべてに，レポータあるいはインタビュアーが含まれている．また，5分間のニュース45日分に含まれる総記事数は，136記事であり，その中でレポータあるいはインタビュアーを含む記事は48記事である．ニュース音声中のアナウンサーの数は1名であった．

2.4.2 実験結果と考察

アナウンサーの話者区間切出し実験を行った結果を**表2**に示す．

アナウンサーの話者区間の切出し率，適合率はそれぞれ92.1%，91.1%と高い値を得ることができた．これは，アナウンサーの話者モデルを1日のニュースごとに自動作成していることから，時期差の問題を回避できたためと考えられる．本手法は，最初の話者区間の継続時間がアナウンサーの発話区間の抽出精度に大きく依存する．したがって，最初の話者区間（45日分）の継続時間が問題となるので，この分布を図**4**に示す．図4より，全体の75%は，最初の話者区間の継続時間が5秒以下でかなり短い区間が多く含まれていることが分かる．また，抽出された音声区間の継続時間の分布を図**5**に示した．音声区間の平均継続時間は，11秒であった．さらに，最初の話者区間の継続時間に対する本手法の性能の依存性を調べるために，最初の話者区間の継続時間を擬似的に80，60，40，20%にそれぞれ削減した場合のアナウンサーの発話区間抽出率を図**6**に示す．削減していない場合は抽出率が92.1%であるのに対して，最初の話者区間の継続時間を80%に削減すると抽出率が84.9%となった．

切出し誤りが生じる原因としては次のような場合があった．各話者の部分空間の学習に用いる音声データが短かすぎると，一般に閾値が低く設定されるため，以後その話者の音声は棄却されやすくなる．また，話

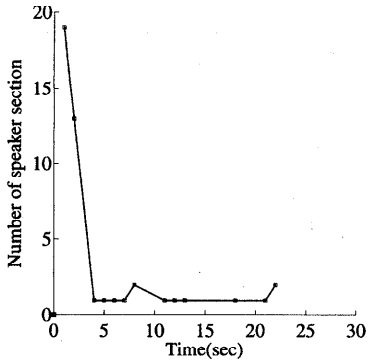


図4 最初の話者区間の継続時間分布
Fig. 4 Duration distribution of the first speaker section.

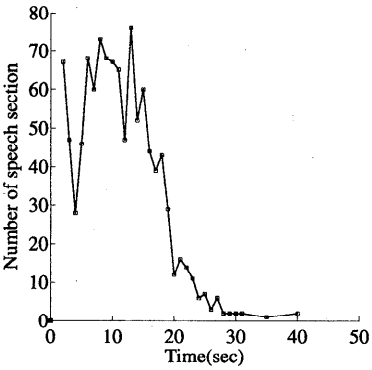


図5 音声区間の継続時間分布
Fig. 5 Duration distribution of speech section.

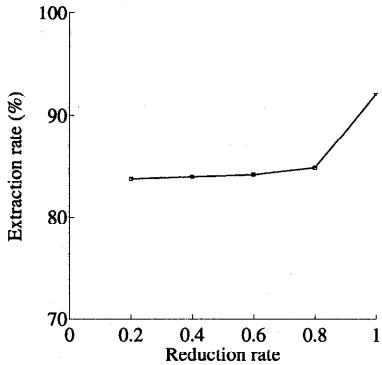


図6 最初の話者区間の継続時間削減率と抽出率
Fig. 6 Relation between extraction rate and reduction rate for duration time of the first speaker section.

者部分空間の作成後，話者照合のための発話区間が短い場合にも，同一話者の発話を棄却したり，異なる話者の発話を同一話者として受理しやすくなったりする傾向がある．さらに，発話区間に雑音が重畳すると，話者照合を誤る場合があった．

表3 音響分析と HMM
Table 3 Acoustic analysis and HMM.

音響分析	サンプリング周波数	12 kHz
	特徴パラメータ	MFCC（39 次元）
	フレーム長	20 ms
	フレーム周期	5 ms
	窓タイプ	ハミング窓
HMM	状態数	5 状態 3 ループ
	タイプ	Triphone HMM
	混合数	8
	学習方法	連結学習
	発話数	21,782 文
	発話者数	男性 137 人

3. ディクテーション

3.1 実験条件

用いた言語モデルは，語彙数 20,000 の back-off bigram で，毎日新聞 CD-ROM 版の 45 カ月分（91 年 1 月～94 年 9 月）の記事から学習した．Back-off smoothing には witten-bell の推定を用い，bigram に対する cut-off は 1 とした．

音響モデルは，男性不特定話者 HMM で，単語間の音素文脈依存も考慮した cross-word triphone モデルである．学習には，日本音響学会新聞記事読み上げ音声コーパス（JNAS）のうち，男性話者 137 人分の 21,782 発話を用いた．音響特徴量には 39 次元の MFCC 特徴パラメータ（12 次元のメルケプストラム係数とパワー，およびそれぞれの Δ，ΔΔ 係数）を用いた．実験条件を**表 3** に示す．

評価用音声データは，NHK 5 分間のニュース 45 日分からレポータあるいはインタビュアーを含む 48 記事を用いた．この 48 記事に対して評価したアナウンサーの話者区間切出し率は 92.6%，適合率は 82.9%であった．こうして切り出されたアナウンサーの発話に対してディクテーションを行う．

3.2 連続音声認識実験

ディクテーションの評価用データは，話者区間抽出実験により切り出されたアナウンサーの話者区間（Anchor），アナウンサー以外のインタビュアーやレポータ等の話者区間（Other），Anchor と Other 区間を両方ともに含んだ全区間（All）の 3 セットである．この評価用データ 3 セットのテストセットパープレキシティを**表 4** に示す．

表 4 より，アナウンサーの話者区間におけるテストセットパープレキシティは 127.4 であり，アナウンサー以外の話者区間におけるテストセットパープレキシティ 234.5 に比べてかなり低いことが分かる．これは，アナウンサーの発話がニュース原稿を正確に読み

表4 評価用ニュース音声データ
Table 4 News speech data for evaluation.

	発話数	20K未知語率	perp
Anchor	250発話	10.8%	127.4
Other	113発話	24.3%	234.5
All	363発話	17.6%	178.6

perp: test-set perplexity

表5 ニュース音声認識実験の結果 (%)
Table 5 News speech recognition result (%).

	単語誤り率	単語正解率	単語正解精度
Anchor	33.6	75.1	66.4
Other	77.2	39.4	22.8
All	45.6	65.8	54.4

上げているのに対して，レポータやインタビュアーの発話は，原稿の正確な読み上げとは異なっていることを示している．

前節で述べた音響モデルと言語モデルを用いて，ニュース音声に対してbeam-searchを用いたビタビデコーディングを行った．デコーダーにはHTK (HMM Toolkit)[15]を用いた．実験結果を表5に示す．

資料として用いた音声データは平均SNR 20.5 dBで比較的背景雑音等が多く，全体的に認識精度は低い値となった．特にOtherセットのインタビュアーなどの区間では，平均SNR 11.2 dBで特に背景雑音が多く，Anchor区間の平均SNR 29.8 dBに対して平均SNRが約63%低かった．また，テストセットパープレキシティも高い値を示し，アナウンサーに対するディクテーションに比べて単語正解精度はかなり低い結果となった．

4. ニュース記事分類

4.1 記事分類の概要

毎日新聞記事の形態素解析結果であるRWCテキストデータベース中の93年度の新聞記事10,000件を，92年度の朝日新聞記事データベース分類表索引[16]を用いて分類した．この分類表索引には，約1万2千のキーワードが載っており，キーワードごとに対応する分類分野が付与されている．この分類表索引には，大分類として総類，政治，経済，労働，文化，科学，社会，事件，スポーツ，国際の10分野が設定されている．ニュース記事分類の実験では，この10分野に対して記事の分類を行った．

10,000記事の正解分野の決定方法は次のとおりである．10,000記事に含まれるキーワードのうち，朝日新聞記事データベース分類表索引に含まれるキーワードを抽出する．このキーワードから，文献17)に記述されている方法でキーワードの各分野に対する寄与率を求める．次に，キーワードの音響尤度を1として分類確率を計算し，分類を行っている．この分類結果を用いて，キーワード選択，キーワードと分類分野の関連度計算を行った．10,000記事の正解分野を人手で行うことも可能であるが，その場合においても，10,000件の正解分野が少し異なるだけで，キーワード選択と関連度の計算法はそのまま用いることができる．

分類した10,000記事を学習データとして，この中から名詞だけを抜き出し，この名詞からキーワードを選択するとともに，キーワードと分野間の関連度を求めた．次に，選択したキーワードと関連度に基づいて，ニュース音声記事のディクテーション結果に対して分類実験を行った．

4.2 新聞記事からのキーワード選択法

新聞記事中の単語 w_i と，その記事が属する分野 t_j との間で関連度 γ_{ij} を求め，閾値処理を行って，関連度 γ_{ij} の大きいものをキーワードとして抽出する[18]．この関連度 γ_{ij} として，本研究では ${\chi_{ij}}^2$ 値を用いている．

χ^2 検定で用いられる χ^2 値は，分野における単語の偏りを示す指標として利用できる．χ^2 値の計算方法は，まず，各単語 w_i について分野 t_j における頻度 x_{ij} を求める．次に，単語 w_i の出現確率は全分野を通じて等しいという仮説を設定し，この仮説に基づいて単語 w_i の分野 t_j における予測頻度 m_{ij} を式(8)により計算する．この x_{ij} と m_{ij} をもとに，式(7)に従って，${\chi_{ij}}^2$ 値を求める．もし，この ${\chi_{ij}}^2$ 値が十分大きな値になれば特定の分野に偏って表れる単語ということになり，分野の識別に有効な単語と見なすことができる．別の見方をすれば，分野 t_j との共起が高い単語 w_i を，キーワードとして選んでいると解釈できる．

$$ {\chi_{ij}}^2 = \frac{(x_{ij} - m_{ij})^2}{m_{ij}} \tag{7} $$

$$ m_{ij} = \frac{\sum_{j=1}^{n} x_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}} \times \sum_{i=1}^{m} x_{ij} \tag{8} $$

ここで，m は異なり単語数，n は分野数，x_{ij} は単語 w_i の分野 t_j における頻度，m_{ij} は単語 w_i の分野 t_j における予測頻度を表している．

4.3 記事の分類方法

1つの記事が与えられると，記事中のキーワード w_i をすべて抽出する．このキーワード w_i と分野 t_j との

表6 キーワード「日米」の3つの分野に対する出現回数と関連度，分類寄与率の例
Table 6 Example of keyword frequency, association degree and contribution rate for 3 topics.

分野	出現回数	関連度	分類寄与率
政治	3	30	0.6
経済	2	20	0.4
社会	0	0	0.0

関連度 γ_{ij} をもとに，キーワード w_i が分野 t_j の分類に寄与する割合（分類寄与率）C_{ij} を式(9)で計算する．たとえば，キーワード「日米」が分野「政治」,「経済」,「社会」に対して**表6**のような関連度を持っている場合，分類寄与率 C_{ij} は表6の右欄のように表される．ここで，表6中の出現回数とは，分類すべき1つの記事にキーワードが現れた回数を表している．

$$C_{ij} = \frac{\gamma_{ij}}{\sum_j \gamma_{ij}} \tag{9}$$

この方法は，キーワード w_i の分野 t_j に対する関連度を，キーワード間で比較可能にする効果がある．

最後に，1つの記事中に含まれているキーワードの出現回数を N_i とすると，分類寄与率 C_{ij} を次式のように総和して，記事 x と分野 t_j との類似度を求める．

$$S(x,t_j) = \sum_i N_i \cdot C_{ij} \tag{10}$$

この類似度の大きな分野 t_j に記事を分類する．

5. ニュース音声記事の分類実験

5.1 実験条件

自動抽出されたアナウンサーの話者区間に対してディクテーションを行い，その結果をもとに記事分類を行った．また，比較として，インタビュアー，レポータ等も含んだ場合についても記事分類を行った．記事分類に用いた48記事における分野の分布を**表7**に示す．今回記事分類に用いた48記事には，10分野のうち科学と文化を除く8つの分野が含まれていた．実際のニュースにおいては，政治，経済といった分野が多く存在していることから，今回記事分類に用いた48記事の分野の分布は，実際のニュースの分野の分布を反映していると考えられる．

キーワードと分類分野との関連度の学習には，94年の毎日新聞記事データを用いている．記事分類を行う際には，新聞記事データ中の単語に対して分野数だけ χ^2 値を求め，分野ごとにある一定の閾値以上の χ^2 値を持つ単語をキーワードとして選択した．

表7 記事の分野の分布
Table 7 Distribution of topics.

分野	頻度	分野	頻度
政治	19	総類	3
経済	9	社会	2
国際	6	スポーツ	2
事件	6	労働	1

表8 ニュース音声記事の分類結果（%）
Table 8 Classification result of news speech articles (%).

	単語正解率	単語正解精度	記事分類率
Anchor	75.1	66.4	70.8
All	65.8	54.4	66.7

5.2 実験結果と考察

ニュース音声記事の分類結果を**表8**に示す．表には，χ^2 値が80以上の値を持つキーワード1,376単語を用いた記事分類率を示している．表中の分類率とは，記事の総数に対して正解分野に分類された記事数の割合である．正解分野は，NHK 5分間のニュース45日分から選んだ48記事に対して人手で付与した．

表8より，アナウンサーの話者区間のみで記事分類を行った結果，単語正解精度66.4%のとき記事分類率70.8%を得た．比較として行ったAllセットに対する記事分類率は66.7%であり，アナウンサーの話者区間のみで記事分類した方が，約4%ほど高い結果を得た．しかし，この差は1記事分の差であり，有意な差であるとはいえず，2つのデータセットに対する記事分類率はほぼ同じであると考えた方がよい．これには次のような理由がある．(1) 評価データである48記事に対するアナウンサーの話者区間切出し率が92.6%，適合率82.9%という結果から，レポータあるいはインタビュアーの区間を誤ってアナウンサーの区間と判定してしまった場合が多く，Anchorに対する記事分類率が低下した．(2) アナウンサーと比べてレポータあるいはインタビュアーの区間が少ないため，Allに対する記事分類率が低下しない．(3) レポータあるいはインタビュアーの区間に対するディクテーションの精度は低く，記事分類に貢献するキーワードが抽出できていない．

以上のことから，Anchorに対する記事分類率を向上させるためには，話者区間の切出し精度をさらに向上させる必要がある．また，キーワードの選択や関連度の計算には，新聞記事10,000件に対して正解分野を機械的に付与したのに対し，ニュース音声記事48件に対しては正解分野を人手で与えている．この新聞記事10,000件から抽出されたキーワードは，機械的に付与した正解分野に依存している．一方，ニュース

音声記事48件の分類結果の評価では，人手で付与した正解分野と比較している．このように，キーワードの選択と選択されたキーワードを用いた記事分類では，新聞記事とニュース記事，機械的に付与した正解分野と人手で付与した正解分野といった違いがある．したがって，新聞記事から選択したキーワードは，必ずしもニュース音声記事の分類にベストなキーワードではない．このため分類精度には上限ができ，2つのデータセットに対する分類精度に差が出なかったものと考えられる．今後，この点の改善が必要であると考えている．

6. むすび

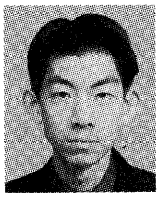
部分空間法（主成分分析法）をベースとする話者照合に基づき，NHK 5分間のニュース45日分に対して，アナウンサーの話者区間抽出実験を行った．その結果，アナウンサーの話者区間切出し率92.1%，適合率91.1%を得た．

また，自動抽出されたアナウンサーの話者区間に対して，ディクテーションを行い記事分類を行った．記事分類を行う際，χ^2 法によりキーワードを自動選択し，キーワードと分類分野との関連度を求め，この関連度により記事の分類を行った．その結果，インタビュアーなどを含めた場合に比べ，分類率を落とすことなく，処理時間を短縮することができる可能性があることが分かった．実験で用いたニュースデータには，アナウンサーの話者区間にも背景雑音が重なっているものもあった．そのようなデータについては，比較的明瞭に，また文法的にも正しいアナウンサーの話者区間であっても，低い認識結果となった．今後，そのような雑音下でもロバストなディクテーションができるよう考慮する必要がある．

参考文献

1) 横井謙太郎，河原達也，堂下修司：キーワードスポッティングに基づくニュース音声の話題同定，情報処理学会研究報告，95-SLP-6-3，pp.15–20 (1995).
2) 大附克年，松岡達雄，松永昭一，古井貞熙：ニュース音声を対象とした大語彙連続音声認識と話題抽出，信学技報，SP97-27，pp.67–74 (1997).
3) Ohtsuki, K., Matsuoka, T., Matsunaga, S. and Furui, S.: TOPIC EXTRACTION MULTIPLE TOPIC-WORDS IN BROADCAST-NEWS SPEECH, *ICASSP98*, pp.329–332 (1998).
4) 恒川俊克，山下洋一，溝口理一郎：キーワードスポッティングに基づくニュース音声の話題分類，情報処理学会研究報告，98-SLP-20-11，pp.61–68 (1998).
5) 松井知子，古井貞熙：音韻・話者独立モデルによる話者照合尤度の正規化，信学技報，SP94-22，pp.61–66 (1994).
6) 松井知子，古井貞熙：テキスト指定型話者認識，信学論，Vol.J79-DII，No.5，pp.647–656 (1996).
7) Markov, K.P. and Nakagawa, S.: EVALUATION OF FRAME-BASED LIKELIHOOD NORMALIZATION FOR SPEAKER VERIFICATION，日本音響学会平成9年度春季研究発表会，2-6-13，pp.69–70 (1997).
8) 松井知子，西谷 隆，古井貞熙：話者照合におけるモデルとしきい値の更新法，信学論，Vol.J81-DII，No.2，pp.268–276 (1998).
9) 松井知子，相川清明：時期差による発声変動を考慮した話者モデルの生成法，日本音響学会平成9年度秋季研究発表会，1-1-23，pp.45–46 (1997).
10) 西田昌史，有木康雄：自動学習による話者セグメンテーション，信学技報，SP97-57，pp.1–6 (1997).
11) 三村正人，河原達也，堂下修司：パネル討論音声の話者と話題に関する自動インデキシング，情報処理学会研究報告，96-SLP-11-3，pp.13–18 (1996).
12) 緒方 淳，森 晴，有木康雄：単語bigramを用いた日本語ニュースディクテーションによる記事分類，日本音響学会平成10年度春季研究発表会，2-Q-16，pp.151–152 (1998).
13) 松井知子，古井貞熙：VQひずみ，離散/連続HMMによるテキスト独立型話者認識法の比較検討，信学論，Vol.J77-A，No.4，pp.601–606 (1994).
14) エルッキ・オヤ（著），小川英光，佐藤 誠（訳）：パターン認識と部分空間法，産業図書 (1986).
15) Cambridge University Engineering Department Speech Group and Entropic Research Laboratory Inc.: HTK Hidden Markov Model Toolkit V2.0.
16) 朝日新聞社ニューメディア本部：朝日新聞記事データベース分類表索引 (1992).
17) 櫻井光康，有木康雄：キーワードスポッティングによるニュース音声の分類と索引付け，信学技報，SP96-66，pp.37–44 (1996).
18) 鷹尾誠一，緒方 淳，有木康雄：ニュース音声の記事分類におけるキーワード選択法の比較，情報処理学会研究報告，98-SLP-22-15，pp.75–82 (1998).

(平成10年10月1日受付)
(平成11年2月8日採録)

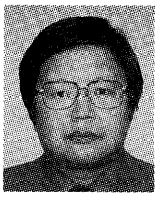


西田　昌史（学生会員）

昭和49年生．平成9年龍谷大学理工学部電子情報学科卒業．現在，同大大学院修士課程在学中．話者認識に関する研究に従事．日本音響学会会員．

緒方　淳

昭和51年生．平成10年龍谷大学理工学部電子情報学科卒業．現在，同大大学院修士課程在学中．音声認識に関する研究に従事．日本音響学会会員．



有木　康雄（正会員）

昭和25年生．昭和51年京都大学大学院修士課程修了．昭和54年同大学院博士課程修了．昭和55年京都大学工学部情報工学科助手．平成2年龍谷大学理工学部電子情報学科助教授，平成4年教授，現在に至る．工学博士．昭和62～平成2年エディンバラ大学客員研究員．画像処理，音声情報処理に従事．電子情報通信学会，日本音響学会，人工知能学会，画像電子学会，IEEE各会員．