未知・複数信号源クラスタリング問題
— 未知話者クラスタリングへの応用 —

Unknown-Multiple Signal Source Clustering Problem
—An Application to Unknown Speaker Clustering—

渡辺 秀行[1]　　村上 仁一 †　　杉山 雅英 †
H.WATANABE　J.MURAKAMI　M.SUGIYAMA

†ATR 自動翻訳電話研究所
〒 619-02 京都府相楽郡精華町光台 2-2

†ATR Interpreting Telephony Research Labs.
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02

あらまし　入力信号系列を複数個の信号源に分割・同定する問題は、音声情報処理のみならず応用範囲が極めて広い。ここではこの問題の定式化、解法を述べる。解法として、Universal VQ 符号帳とその出現確率による出現確率クラスタリング法および ergodic HMM による方法を提案する。前者は各区間 (segmentation) を既知とするクラスタリングに基づく識別手法、後者はカテゴリとその遷移を ergodic HMM でモデル化し segmentation とカテゴリ識別との同時推定を行なうものである。一つの応用例として複数話者発話音声の分割同定を検討する。本報告では前者の手法に限定して実験により有効性を示し、今後の課題について述べる。

Abstract This report describes an unknown-multiple signal source clustering problem of input signal sequence, and its application to unknown speaker clustering. This type of problems has wide applications in various fields. This report formulates the problem and proposes the solution based on an Universal VQ codebook and its occurrence probability (Occurrence Probability Clustering Method), and ergodic HMM. The former requires the segmentation boundaries, the latter provides the segmentation and clustering simultaneously. This report also describes the former's application to unknown-multiple speaker clustering and evaluates the performance.

1　まえがき

未知・複数信号源クラスタリング問題 (N 信号源分割同定問題) は、観測された信号系列から異なる N 個の信号源より生成される区間 (ここではブロックと呼ぶことにする) を検出し (segmentation)、各ブロックが各々いずれの信号源 (ここではカテゴリと呼ぶことにする) から生成されたかを分類・識別 (discrimination) する問題である。この問題は、音声情報処理分野に限らず基本的かつ重要な問題である。例えば、複数話者発話の識別問題、音声・非音声区間の識別問題、複数言語識別問題、音響単位の自動決定問題 [12] (音響言語モデル) 等、この形式に定式化できる。

未知・複数信号源クラスタリング問題は大別して以下の 4 つの部分問題から構成される。

1. カテゴリを特徴付ける特徴量の問題
2. カテゴリ遷移の segmentation 問題
3. 各ブロックを N 個のカテゴリに識別する問題
4. カテゴリ数 N を推定する問題

これらの 4 つの部分問題を同時に解くことは容易ではないので、条件をある程度既知であると仮定して解法を検討する。

[1] 現在、北海道大学 電子科学研究所 博士課程 在学中

カテゴリ数 N 及び segmentation が既知の場合、Universal VQ 符号帳とその出現確率による解法 (出現確率クラスタリング法と略す) を提案する [1]。この方法ではブロック毎に算出される複数個の出現確率を N 個のカテゴリにクラスタリングすることにより識別が行なわれ、segmentation が未知の場合にも応用が可能である。segmentation が未知の場合、ergodic HMM による解法を提案する。この手法は、カテゴリ遷移を ergodic HMM でモデル化するものである。ここでは、Baum-Welch アルゴリズムで HMM パラメータを推定し、Viterbi アルゴリズムで最適状態遷移系列を推定することにより、segmentation とカテゴリ遷移との同時推定が行なわれる。ergodic HMM は話者認識や言語識別においてその有効性が示されている [2, 3, 4]。 本報告では前者の手法について未知話者クラスタリングに適用した実験結果を述べることとし、 ergodic HMM を用いる方法については次の報告 [5] で述べることとする。さらに、カテゴリ数 N の推定問題を検討する。

文献 [6, 7] において未知話者クラスタリングに関する類似の検討がなされている。しかしながら、文献で扱われた問題は少数の信号源 (pilot, controllers, noise) を扱っており、特定の信号源 (pilot) の同定 (追跡) に主眼がおかれている。

本報告における話者識別問題と通常の話者識別問題との違いは、あらかじめの学習 (事前学習) の有無である。本問題は与えられたひとつながりの音声に含まれる複数の信号源をその音声のみを用いて、事前学習することなく分割・分類することである。一方、通常の話者識別問題は事前に多量の音声を用いて話者モデルを作成し、それとは異なる未知の音声に対して識別を行なう。通常の話者識別で考慮しなければならないとされる話者特徴の時間変化などを本問題では考慮する必要がない。しかしながら、未知話者の分類のためには話者識別の問題・手法と密接に関わりあっている。

2 問題の定式化

図 1 に示すように、n 次元ベクトルで与えられる信号の系列を $\boldsymbol{X} = (\boldsymbol{x}_t)(t = 1, 2, ..., T)$ とする。この系列は、K 個のブロック $\boldsymbol{X}_k$ $(k = 1, 2, ..., K)$ からなり、各ブロックが $N(\leq K)$ 個のカテゴリ (状態) $\boldsymbol{C}_j (j = 1, 2, ..., N)$ のいずれかから生じた系列であるとする。ここで、ブロック $\boldsymbol{X}_k$ の構成要素を $(\boldsymbol{x}_{t_{k-1}+1} ... \boldsymbol{x}_{t_k})$ (ただし、$t_0 = 0, t_K = T$) 、継続長を M_k $(= t_k - t_{k-1})$ とする。

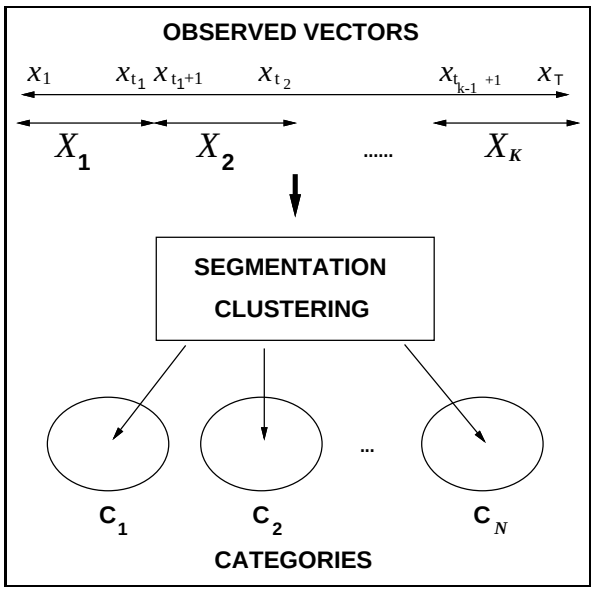

図 1: N 信号源分割同定問題

N 信号源分割同定問題は、与えられたベクトル信号系列に対し、信号源の特性の違いに着目してブロックの切れ目の位置 $t_k (k = 1, 2, ..., K-1)$ を探し (segmentation) この K 個のブロックを $N(\leq K)$ 個のカテゴリ (状態) に識別しカテゴリ数 N を推定することである。

最も簡単な場合として、カテゴリ数 N 及び segmentation が既知の場合を考える。この場合、N 信号源分割同定は、各ブロック $\boldsymbol{X}_k$ $(k = 1, 2, ..., K)$ を N 個のカテゴリ $\boldsymbol{C}_j$ $(j = 1, 2, ..., N)$ にクラスタリングする問題になる。

2.1 出現確率によるカテゴリ特徴付け

クラスタリングに際し各ブロックのカテゴリ特徴を表現する量の算出が重要となる。話者識別への応用のために、本報告ではカテゴリ特徴量として Universal VQ 符号帳の符号の出現確率を用いる。他の方法として混合ガウス分布も考えられる。これらは、既にテキスト独立型の話者認識や言語識別に対して有力であることが報告されている [8, 9]。

Universal 符号帳 $\boldsymbol{U} = \{\boldsymbol{u}_l\}(l = 1, 2, ..., L)$ (L は符号帳サイズ) は、$\boldsymbol{X}$ の全てのベクトルを用いてベクトル量子化することにより作成される。入力信号系列の各ブロック $\boldsymbol{X}_k$ $(k = 1, 2, ..., K)$ はそれぞれ $\boldsymbol{U}$ により量子化され、その量子化符号の出現頻度 (出現確率) はサンプルベクトル数がある程度多ければ各カテゴリ固有の特徴と

なり入力系列の各ブロックのカテゴリ性を表現する。出現確率 $\boldsymbol{p}_k$ は次式で与えられる。

$$\boldsymbol{p}_k = (p_{k1}p_{k2}...p_{kL}), \quad p_{kl} = \frac{K_{kl}}{M_k}, \tag{1}$$

ここで、K_{kl} は k 番目のブロックに対する符号 l の出現頻度、M_k は k 番目のブロックの継続長であり、$\boldsymbol{p}_k$ は L 次元ベクトルとなる。

2.2 出現確率クラスタリング法

L 次元ベクトル $\boldsymbol{p}_k(k = 1, 2, ..., K)$ を N 個のカテゴリにクラスタリングするために LBG 法 [10] を用いる。通常の LBG 法では、分割の個数が 2 の巾乗となり、任意カテゴリ数のクラスタリングを考える本問題には適していない。また 2 分割を行なうため比較的歪みの小さい集合 (partition) までも分割してしまう危険性を持っている。そこで本報告では歪み最大の集合のみを分割する変形 LBG 法を用いる [11]。これにより、2 の巾乗以外のカテゴリ数の場合でもクラスタリングが可能となり、かつ歪みの小さい集合が分割されることがなくなる。LBG 法と変形 LBG 法の比較実験を 4.1 で述べる。

クラスタリングにおける出現確率間の距離尺度としてここでは次式で表される Euclid 距離及び二つのタイプの Kullback-Leibler (KL) 情報量を用いる。

$$d_{\text{Euclid}}(\boldsymbol{p}, \boldsymbol{z}) = \sqrt{\sum_{l=1}^{L}(p_l - z_l)^2}, \tag{2}$$

$$d_{\text{KL Type1}}(\boldsymbol{p}, \boldsymbol{z}) = \sum_{l=1}^{L} z_l \log \frac{z_l}{p_l}, \tag{3}$$

$$d_{\text{KL Type2}}(\boldsymbol{p}, \boldsymbol{z}) = d_{\text{KL Type1}}(\boldsymbol{z}, \boldsymbol{p}). \tag{4}$$

ただし $\boldsymbol{z} = (z_l)$ である。Euclid 距離, KL Type 2 尺度に対する centroid は相加平均であり、KL Type 1 に対しては以下のように計算される。

- Euclid 距離, KL Type 2:

$$z_l^c = \frac{1}{\#(\boldsymbol{C})} \sum_{\boldsymbol{p} \in \boldsymbol{C}} p_l. \tag{5}$$

- KL Type 1:

$$z_l^c = \frac{q_l}{\sum_{l=1}^{L} q_l}, \quad q_l = \left[\prod_{\boldsymbol{p} \in \boldsymbol{C}} p_l \right]^{1/\#(\boldsymbol{C})}. \tag{6}$$

ここで、$\#(\boldsymbol{C})$ は集合 $\boldsymbol{C}$ の要素の数である。

2.3 カテゴリ数 (N) の推定

出現確率クラスタリング法では、クラスタリングの分割全体の平均歪み D が計算される。

$$D = \frac{1}{K} \sum_{k=1}^{K} \min_l d(\boldsymbol{p}_k, \boldsymbol{z}_l) \tag{7}$$

分割の個数が真のカテゴリ数と等しい場合における D の値が、カテゴリー数 (話者数) やカテゴリーの種類 (性別) などによらず変化しないならば、D の絶対的な値を基に N を推定できることになる。これについては 4.3 で考察する。

ergodic HMM の場合、Baum-Welch のパラメータ推定アルゴリズムにおいて、シンボル系列に対するモデルの尤度が計算される。出現確率による解法と同様、尤度の絶対的な値を基に N を推定できると考えられる。また、N の推定値を客観的に与える方法として、AIC 基準や MDL 基準等の情報量基準の適用も考えられる [13, 14]。

3 実験の準備

3.1 実験条件

音声分析条件を表 1 に示す。ここでは簡単化のため、特徴量として LPC ケプストラム (1 次 ~ 16 次) を用いた。従って、得られる音声特徴量 (x_t) は LPC ケプストラムベクトルの系列である。

実験には単語音声 (ATR 5240 単語データ) を並べて疑似的な複数話者による対話音声を作成した。話者の一発話 (1 ブロック) は複数単語 (W) からなり、これを各話者毎に 10 ブロック作成し、総数 $10W$ のブロックを連接して 1 つの音声資料とした。さらにこれを 10 セット作成し 10 回の実験を行なって認識単位毎の分類率を算出しその平均で評価を行なうこととした。話者内では 10 セットに含まれる単語は全て異なっている。各単語には前後に無音区間が含まれている。

表 1: 音声分析条件

音声特徴量	LPC ケプストラム係数
標本化周波数	12 kHz
LPC 分析	14 次
打ち切り次数	16 次
フレーム長	21.3 ms (256 点)
フレーム周期	10.7 ms (128 点)
高域強調	$(1 - 0.97z^{-1})$

3.2 識別評価方法

信号系列を複数のカテゴリに分類した場合、各カテゴリーがどの話者に対応しているのかは同定できない。すなわち、求まる系列はカテゴリーの番号であり、番号と話者の対応には $N!$ 個の組み合せが存在する。そのため以下の式で識別率を算出した。

$$R = \frac{1}{K} \max_{\sigma} \sum_{k=1}^{K} d(\tau(\boldsymbol{X}_k), \sigma(S_k)) \qquad (8)$$

ここで、τ, σ, S_k, d はクラスター番号を与える写像、$(1, 2, \ldots, N)$ の任意の置換、各ブロックの正解カテゴリー番号、値が一致した時のみ 1、それ以外は 0 である関数である。手順を表 2 に示す。各カテゴリー番号を σ で置換し、$N!$ 通りの置換について正解率を算出しその中の最大値を識別率とする。この例の場合、$3! = 6$ 通りの組み合せがあり識別率が $R = 70\%$ となる。識別率 R_{AVE} は各試行の識別率の 10 試行分を平均して算出される。R の算出は出現確率による方法では上式のようにブロック単位で、ergodic HMM の方法ではフレーム単位で行なわれる。 ergodic HMM に対して R の算出には多大の計算量を要するので分岐限定法などの組み合せ最適化法による高速化が必要である。

3.3 Universal 符号帳の作成手順

Universal 符号帳は音声資料 $\boldsymbol{X}$ を用いて作成するべきであるが、このようにすると音声資料毎にベクトル量子化が必要となるので、ここでは簡単化して $\boldsymbol{X}$ とは独立な評価に用いない単語音声データから作成することにする。1 人あたり 100 単語を用いることにし、話者が 4 人以下の場合は 1 段階で、話者 8 人の場合はフレーム数が多く (約 80000) 計算量が膨大となるため 2 段階で作成する。まず話者毎に、100 単語 (約 40000 フレーム) から 512 個の符号を求め、次に 8 名分の符号 (計 4096 個) から、所定個の符号を求める。実験では符号サイズを $L = 256$ とし、距離尺度として Euclid 距離を用いる。

4 未知・複数話者クラスタリング実験

4.1 LBG 法と変形 LBG 法との比較

LBG 法と変形 LBG 法の識別性能の比較を図 2 に示す。ここでは話者 4 名 (男性 2 名、女性 2 名) である。図の横軸は各ブロック中の単語数、縦軸は識別率 (R_{AVE}) を表す。図からいずれの尺度を用いても変形 LBG 法は LBG 法よりも高い識別率を与えることが判る。以降の実験では変形 LBG 法を用いることにする。

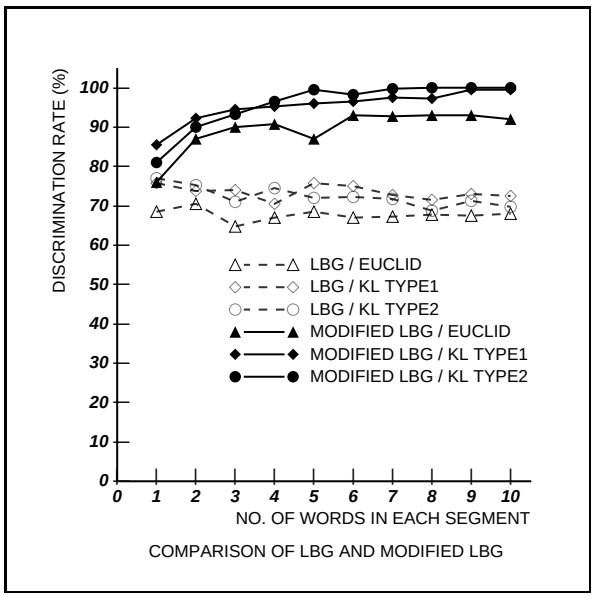

図 2: LBG 法と変形 LBG 法の比較 (男 2 名・女 2 名)

4.2 話者数・単語数と識別率との関係

話者数と識別率との関係を、図 3 (男女同数) 、図 4 (男性のみ) に示す。話者数が増加するに従い識別率が下降する。男女の識別は比較的容易であるので同数の場合は男性のみの方がより識別が難しくなる。話者 2 人の発話の場合は男女の場合でも男性のみの場合でも 1 単語の発話でほぼ完全に識別できる。10 単語を用いる場合、尺度として KL Type2 を適用すれば男女同数の場合でほぼ 100%、男性のみの場合は 8 人で約 90%の識別率が得られる。特に話者 8 人の場合において、Euclid 距離では単語数が増加しても識別率があまり上昇しないのに対し、KL 情報量では識別率が上昇する。これは単語数の増加に従い各ブロックの話者を表す出現確率のばらつきが小さくなり、KL 情報量に基づいたクラスタリングが安定するためであると考えられる。逆に単語数が少ない場合は Euclid 距離が最も良い識別性能を与えている。

各ブロック (発話) 中の単語数と識別率との関係を図 5 (話者 4 人の場合) 、図 6 (話者 8 人の場合) に示す。単語数が増加するに従い識別率が上昇する。

話者 8 人の場合において、3 種の尺度による識別率の関係は、音声長が 5 単語より短い場合、Euclid > KL Type1 > KL Type2、音声長が 5 単語以上の場合、Euclid < KL Type1 < KL Type2 である。このことは、各ブロックの音声長により尺度を使い分ける必要があることを示して

表 2: 識別評価方法 (話者 3 人・10 系列の場合)

真の系列	*1*	*1*	*1*	*2*	*2*	*2*	*3*	*3*	*3*	*3*	
推定系列	1	1	0	0	2	0	0	2	2	2	
$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{pmatrix}$	0	0	0	1	1	1	2	2	2	2	4/10
	×	×	◯	×	×	×	×	◯	◯	◯	= 40 %
$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \end{pmatrix}$	0	0	0	2	2	2	1	1	1	1	2/10
	×	×	◯	×	◯	×	×	×	×	×	= 20 %
$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 2 \end{pmatrix}$	1	1	1	0	0	0	2	2	2	2	7/10
	◯	◯	×	◯	×	◯	×	◯	◯	◯	= 70 %
$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 0 \end{pmatrix}$	1	1	1	2	2	2	0	0	0	0	4/10
	◯	◯	×	×	◯	×	◯	×	×	×	= 40 %
$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \end{pmatrix}$	2	2	2	0	0	0	1	1	1	1	2/10
	×	×	×	◯	×	◯	×	×	×	×	= 20 %
$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \end{pmatrix}$	2	2	2	1	1	1	0	0	0	0	1/10
	×	×	×	×	×	×	◯	×	×	×	= 10 %
											⇓
											最大値 = 70%
											= 識別率

いる。

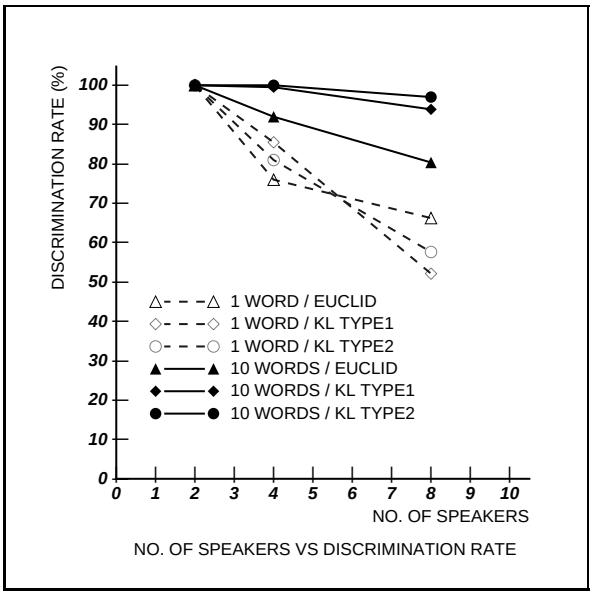

図 3: 出現確率クラスタリング法における話者数と識別率との関係 (男女同数)

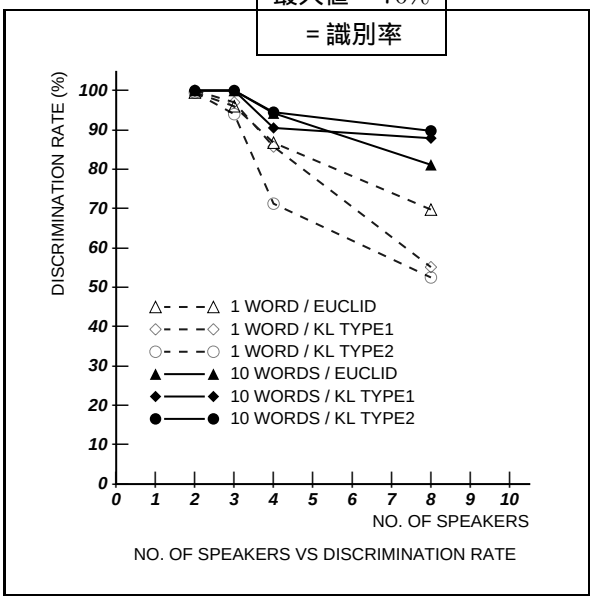

図 4: 出現確率クラスタリング法における話者数と識別率との関係 (男性のみ)

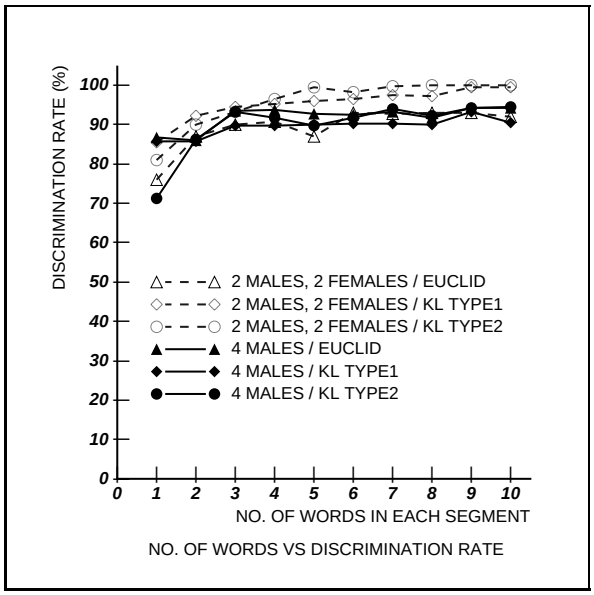

図 5: 出現確率クラスタリング法における単語数と識別率との関係 (話者 4 名)

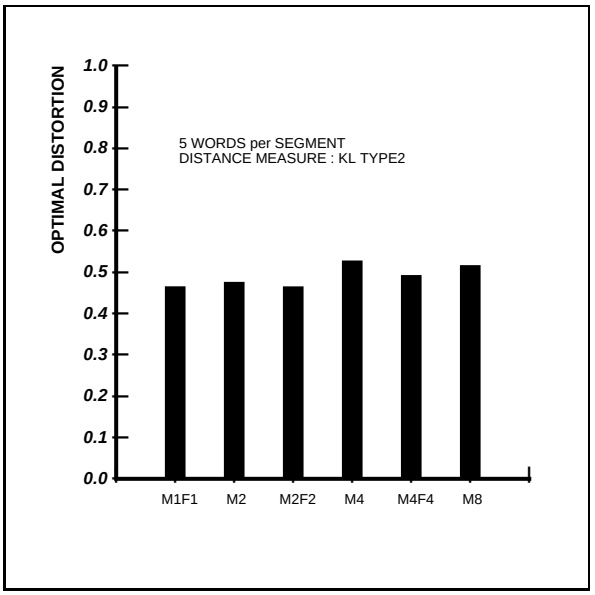

図 7: 最適量子化歪みと話者セットとの関係 M1F1, M2, M2F2, M4, M4F4, M8 (M:男, F:女, 5 単語, 尺度 : KL Type2)

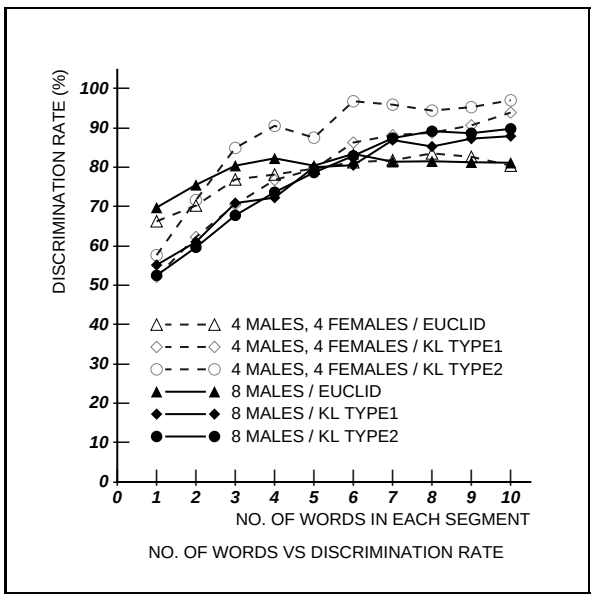

図 6: 出現確率クラスタリング法における単語数と識別率との関係 (話者 8 名)

4.3 量子化歪みによる話者数の推定

変形 LBG 法で算出される平均量子化歪み D を基準として、話者数 N を推定する問題を考える。 単語数を固定して話者数や性別を変化させた場合の最適量子化歪みの変化を図 7 に示す。この図では、 5 単語、KL Type2 を用いている。図より多少のばらつきはあるものの話者数や性別が異なっても最適量子化歪み値は大きく変化しないことがわかる。他の単語数及び尺度の場合も同様の傾向が得られている。 男性 2 名・女性 1 名、 5 単語、KL Type2 の場合における分割の個数と量子化歪みの関係を図 8 に示す。図では 3 回試行の結果を重ね書きしている。この例では識別率は 100% である。図からわかるように分割を増加させた場合、分割個数が真の話者数になるまでは量子化歪みの減少が急激であるがそれ以降の減少度が小さい。図 9 は、男性 4 名、5 単語、KL Type2 の場合であるが、量子化歪み変化の曲線がなだらかになっており話者数推定が図 8 に比べて困難である。これは話者が全て男性であるのでクラスタ形成が明確にできないためであると考えられる。この例では識別率は約 90% である。

次に話者を固定して単語数を変化させた場合の最適量子化歪みの変化を図 10 に示す。この図は、男性 2 名・女性 2 名、KL Type2 の場合である。図から単語数の増加に従い最適量子化歪みの値が減少することがわかる。これは単語数の増加により同一話者における符号出現確率

のばらつきが減少するためであると考えられる。 以上より各発話 (ブロック) の長さ (フレーム数) がある程度限定されれば量子化歪みの値に基づいて話者数を推定することができると考えられる。また図 10 から単語数がある程度多くなれば最適量子化歪みの減少度が小さくなるので長さが未知であっても量子化歪みの値から話者数を推定することができると考えられる。

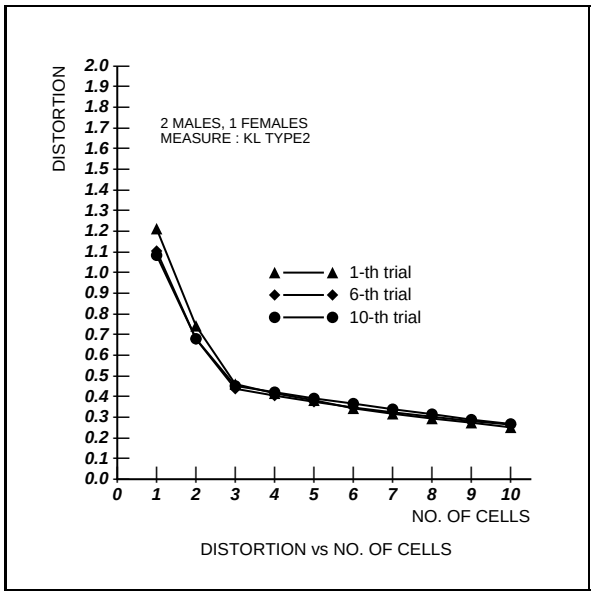

図 8: 量子化歪みと分割個数との関係 (男性 2 名・女性 1 名、 5 単語、尺度：KL Type2)

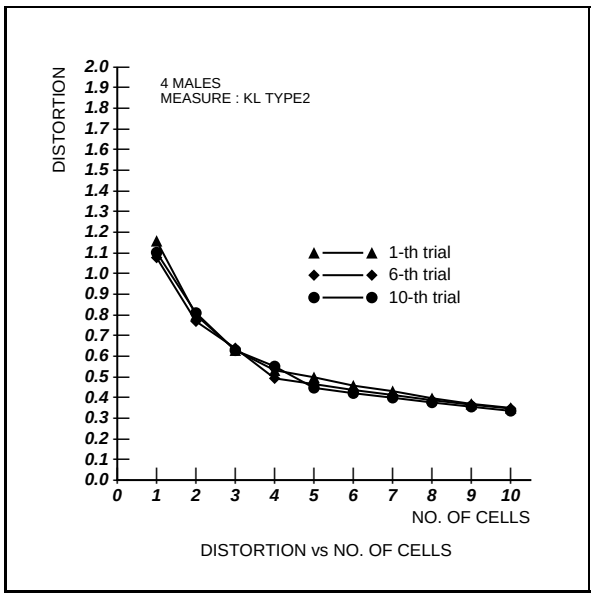

図 9: 量子化歪みと分割個数との関係 (男性 4 名、 5 単語、尺度：KL Type2)

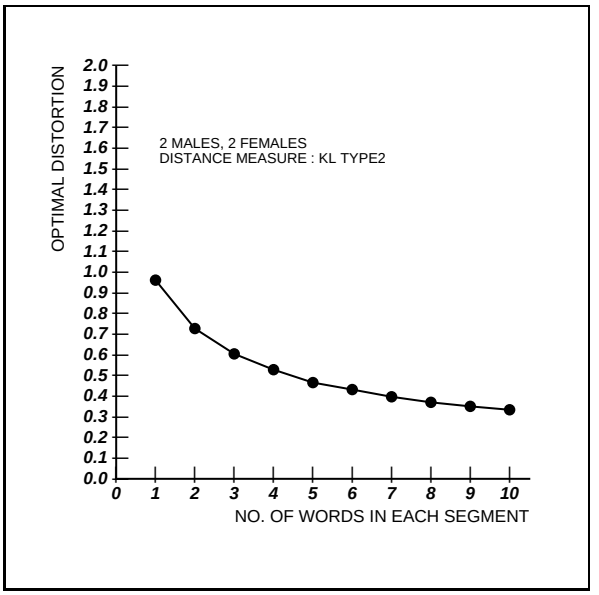

図 10: 最適量子化歪みと単語数との関係 (男性 2 名・女性 2 名、尺度：KL Type2)

4.4 無音区間を除去する効果

音声データベースは話者毎に独立に収録・作成されているため無音区間が収録環境などに影響されその周波数特性が話者毎に異なっている可能性がある。各単語の前後に付けられている無音区間を取り除いた時の識別率を図 11(男女同数)、図 12(男性のみ) に示す。話者により周波数特性が異なることに加えて無音区間が除かれ全体の継続長が短くなるため 1 単語を用いた時の識別率は大幅に減少している。図 13(話者 4 人)、図 14 (話者 8 人) に単語数と識別率との関係を示す。単語数の増加にほぼ比例して識別率が向上している。無音区間がある場合と異なって、男性のみの場合はどの条件に対しても識別率は KL Type 2 > Euclid > KL Type 1 の順となっているのに対して、男女混合の場合は KL Type 2 が最良であり、単語数によって Euclid と KL Type1 との識別率の逆転がおこっている。以下の実験では無音区間を取り除くことにする。

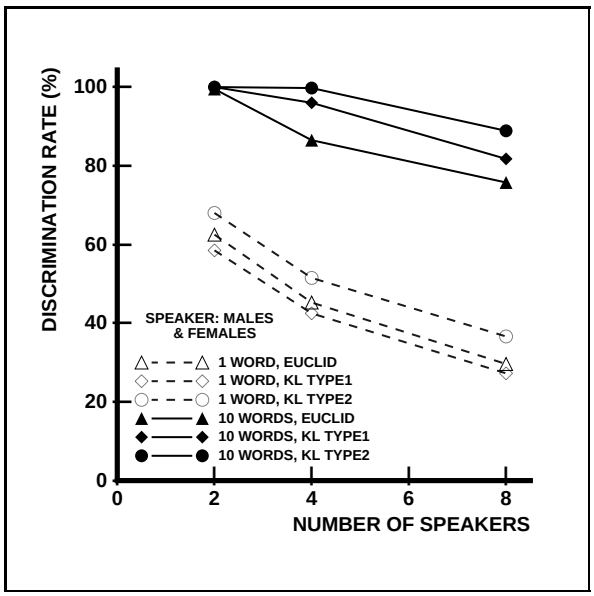

図 11: 無音区間を取り除いた時の出現確率クラスタリング法における話者数と識別率との関係 (男女同数)

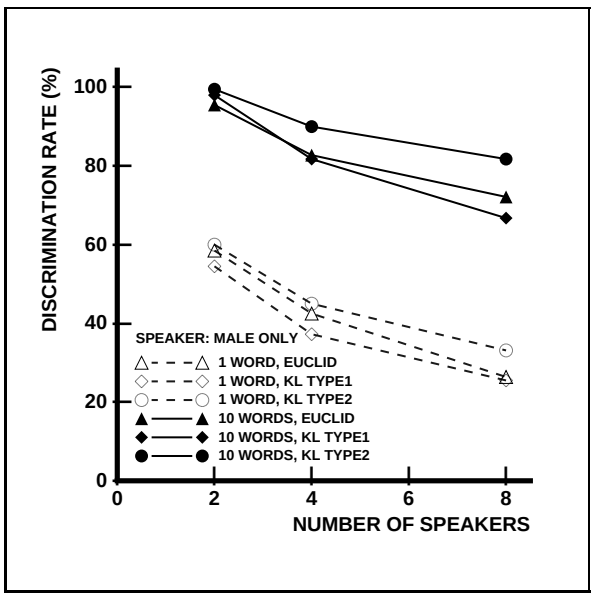

図 12: 無音区間を取り除いた時の出現確率クラスタリング法における話者数と識別率との関係 (男性のみ)

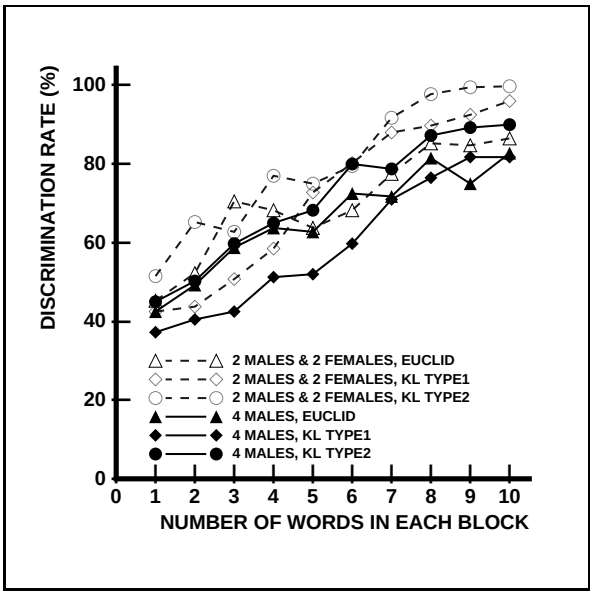

図 13: 無音区間を取り除いた時の出現確率クラスタリング法における単語数と識別率との関係 (話者 4 人)

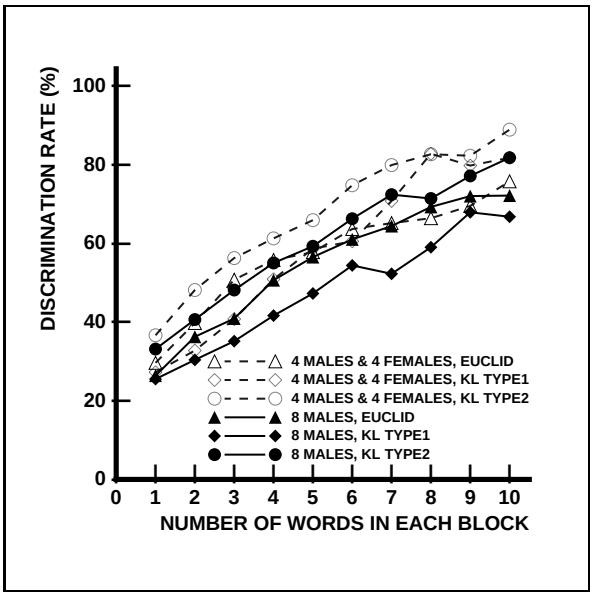

図 14: 無音区間を取り除いた時の出現確率クラスタリング法における単語数と識別率との関係 (話者 8 人)

4.5 Fuzzy VQ の効果

出現確率の算出に Fuzzy VQ を適用する。これは式 (9) に示すように入力ベクトル x_t に対する Fuzzy membership 関数 w_{tl} から算出される。式中で N_k は x_t の $\boldsymbol{U}$ の k 近傍の符号番号の集合である。fuzziness (F) と k 近傍数とを変化させた時の識別率を図 15 に示す。話者は男性 8 名で KL Type 2 を用いており、無音区間を除いている。Hard VQ の時、81.8% の識別率が $F = 1.1, k = 2$

の時、 85.3% に向上している。一方、近傍数が増加するにつれて逆に劣化している。この傾向は他の話者組み合せに対しても同様である。話者数が少ない場合のように問題が簡単なほどこの傾向が顕著であり、さらに Fuzzy VQ の効果が大きい。

$$p_l = \frac{K_l}{M}, \quad K_l = \sum_t w_{tl}, \tag{9}$$

$$w_{tl} = \begin{cases} 1/(\sum_{j=1}^{k} (\frac{d_l}{d_j})^{\frac{1}{F-1}}), & l \in N_k, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

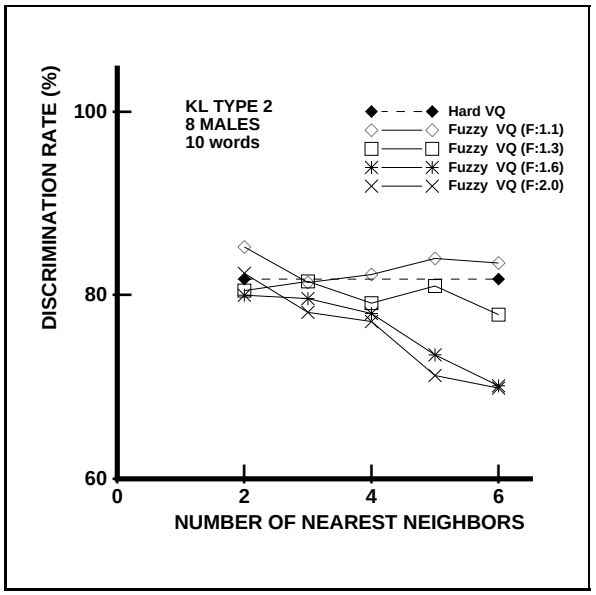

図 15: Fuzzy VQ と Hard VQ との比較 (話者 8 人)

4.6 ランダムテキストの効果

以上の実験でのテキスト (単語集合) は同一話者内では全て異なるが全話者共通となっている。各話者が同一テキスト発話であるのでテキスト固定型と言える。この場合、共通のテキストを用いるので話者の違いが強調されより識別がしやすくなると考えられる。そこでランダムテキストすなわち単語がランダムに並ぶ時の識別性能を評価する。ここでは乱数を用いて単語列を作成した。結果を図 16 に示す。ここでは前節の結果から Fuzzy VQ を適用している。入力単語が少ない場合はテキスト固定とランダムテキストとの性能の差は小さいが単語数が増加するにつれてランダムテキストの方がはるかに高い識別率を与えることがわかる。これは予想に反する結果であるが、テキスト固定の場合、使われる単語列が 5 0 音順に並んでおり入力中の音素が偏って分布し話者特徴量である出現確率が良好に推定されないためと考えられる。実際の対話音声では全ての音素が偏らずにランダムに含むことは強い仮定であるので、これほど高い性能は得られないと考えられる。

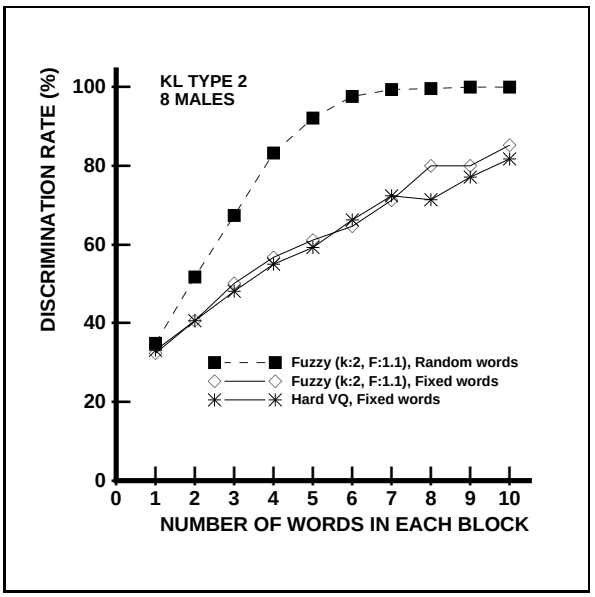

図 16: ランダムテキストの時の単語数と識別率との関係 (男性話者 8 人)

5 むすび

入力信号系列の複数信号源分割同定問題を定式化しその解法を検討した。実験では複数話者発話の分割同定に対して単語数・話者数と識別性能との関係を述べ、以下の結論を得た。

1. 出現確率クラスタリング法における距離尺度として Kullback-Leibler 情報量 (KL Type 2) を用いることにより良好な識別が行なえる。
2. 話者数 N はクラスタリングの量子化歪みを基準にしてある程度推定できる。
3. Fuzzy VQ による出現確率の計算は有効であり、fuzziness や k 近傍数を小さくする必要がある。
4. テキスト固定型とランダム型とでは後者の方の認識性能が高い。

今後の課題を以下に述べる。

- Universal 符号帳の大きさと識別率との関係
- 長時間窓音声分析の効果
- 出現確率を用いたカテゴリー自動セグメンテーション
- 分岐限定法などの組合せ最適化法の適用による評価法の高速化

謝辞

研究の機会を与えて頂いた北海道大学電子科学研究所 永井信夫教授、三木信弘助教授、ATR 自動翻訳電話研究所 樽松明社長、熱心に御討論していただくと同時に有益な御助言をいただいた音声情報処理研究室の皆様に感謝致します。

参考文献

[1] 杉山, 渡辺, 村上, 事前登録なしの話者識別問題, 電子通信学会部門別全国大会, (1992-09), 発表予定.

[2] 松井, 古井, エルゴード的 HMM による話者認識, 音響学会講演論文集, 3-6-14 (1991-10).

[3] M.A.Zissman, HMM Language Identification, Proc. of Speech Research Symposium XII (Jun.1992).

[4] 杉山, 吉田, スペクトル遷移の確率モデルによる言語識別の検討, 音響学会講演論文集, 1-P-1 (1992-03).

[5] 村上, 渡辺, 杉山, HMM を用いた未知・複数信号源クラスタリング問題の検討, 音声研究会資料 (1992-09), 発表予定.

[6] G.Yu, M.Siu, H.Gish, An Unsupervised, Sequential, Learning Algorithm for the Segmentation of Speech Waveform with Multiple Speakers, Proc. of ICASSP92, 86.10 (Apr.1992).

[7] H.Gish, G.Yu, Clustering of Speakers Engaged in Dialog, Proc. of Speech Research Symposium XII (Jun.1992).

[8] 石毛, 佐野, 間野, 白井, スペクトルの量子化分布による話者認識, 信学技報, SP86-69, pp.59-64 (1986).

[9] M.Sugiyama, Automatic Language Recognition Using Acoustic Features, Proc. ICASSP91, pp.813-816 (1991).

[10] X.D.Huang, Y.Ariki, M.A.Jack, Hidden Marcov Models for Speech Recognition, Edinburgh University Press, Edinburgh (1990).

[11] 鹿野, 入力音声のベクトル量子化による単語音声認識, 音声研資, S82-60, pp.473-480 (1982-12).

[12] 児島, 田中, 速水, 単語音声サンプルからの階層的な音韻概念の獲得, 音響学会講演論文集, 2-P-4 (1991-10).

[13] 森田, 算術符号から MDL 基準へ, 数理科学, No.290, pp.25-31 (1987).

[14] J.Rissanen, A Universal Prior for Integers and Estimation by Minimum Description Length, Ann. Stat., Vol.11, pp.416-431 (1983).