TR-I-0224

音声認識結果の信頼性評価

Evaluation of Speech Recognition Candidates by Statistical Methods

小林尚弘　　竹沢寿幸

Naohiro Kobayashi　Toshiyuki Takezawa

1991.9.13

概要

音声認識システムから得られた11名40文節に対する認識スコアつき音声認識候補に対し、その認識スコアの分散統計特性を用いた二階差分法と分散法で実際に認識候補の絞り込みを行なった結果を述べる。また、別の観点から話者毎の認識率の違いにどのような要因があるかを分析する。さらに、話者適応や不特定話者音声認識における認識率向上のために本手法が利用できる範囲と課題を論ずる。

もくじ

1　まえがき

音声認識システムから得られる音声認識候補とその認識スコアから認識候補を絞り込む方法として二階差分法と分散法が提案されている [1] が、これまで、この手法では特定話者1名のデータを基にした結果しか出されていなかった。ここでは、話者適応を行なった男性9名、女性2名の合計11名のデータに二階差分法、分散法を適用し、認識候補の絞り込みを行なった結果を述べる。さらに、話者適応や不特定話者音声認識における認識率向上のために、本手法が利用できる範囲と課題を論ずる。

まず、第2節で、候補の絞り込みに関して説明する。第3節では、二階差分法の原理と実験結果を述べる。第4節では、分散法の原理と実験結果を述べる。第5節では分散法を用いた認識候補の正規化確率計算について述べる。第6節では、別の観点から発声速度と認識率の関係を調べたのでその結果を述べる。最後に全体をまとめ、今後の課題をあげる。

2　音声認識システムについて

2.1　音声認識システムにおける候補の絞り込み

話者適応機能を持った音声翻訳システム *SL-TRANS 2* [2] は、マイクから入力された音声を認識し、機械翻訳、音声合成を行なうシステムである。このシステムからは音声認識候補とそれに対応する認識スコアとが、文節毎に5個ずつ出力 される。この5個の認識候補をそのまま機械翻訳の処理系に渡すと、文候補が乗算的に増えるためにその計算量が大きくなり、処理系に多大な負荷をかけることになる。

この理由から、音声認識システムから出力される5個の候補を少数に絞り込む必要がある。この絞り込みを実現するために考えられた一つの手法が、二階差分法と分散法である。これらの方法は音声認識システムから出力される音声認識候補の認識スコアの統計的な特性を用いて絞り込みを実現している。

音声認識候補の絞り込みに関する処理の流れは、まず、二階差分法により絞り込みが行なえる場合は二階差分法である程度の絞り込みを行ない、さらに分散法で精度の高い絞り込みを行なう。また、二階差分法で候補の絞り込みが行なえなかった場合も分散法により候補の絞り込みを行なう。このようにして候補を少数に絞り込む目標は達成される 。

次節以下で、この二階差分法と分散法の原理と、11名分のデータを処理した結果を詳しく説明していくが、まず、今回使用したデータについて説明する 。

2.2 使用したデータ

今回使用したデータは男性9名、女性2名の合計11名の実際の発声による認識結果である。

11名の発声者はあらかじめ25単語により話者適応をし、国際会議に関する問い合わせの電話での会話を19文(40文節)マイクから文節毎にポーズをおいて1文単位で発声する(会話の内容は表1参照)。

音声認識システムは、この入力に対し文節毎に独立に処理し、認識候補とそれに対する認識スコアを上位から5個出力する。今回はその全データを用いて実験を行なった。

表 1: 国際会議に関する問い合わせの会話 (19 文)

文番号	会話文
1	もしもし
2	そちらは 会議事務局ですか
3	はい
4	そうです
5	会議に 申し込みたいのですが
6	登録用紙は 既に お持ちでしょうか
7	いいえ
8	まだです
9	分かりました
10	それでは 登録用紙を お送り致します
11	ご住所と お名前を お願いします
12	住所は 大阪市 北区 茶屋町 二十三です
13	名前は 鈴木真弓です
14	分かりました
15	登録用紙は 至急 送らせていただきます
16	分からない 点が ございましたら いつでも お聞き下さい
17	有難うございます
18	それでは 失礼します
19	どうも 失礼致します

3 二階差分法

3.1 原理

ある入力音声に対し、認識システムから認識スコアとして $\{D_i\}$ $(1 \leq i \leq 5)$ が得られたとする。このデータ列に対して二階差分を行なうと、次のようになる。

$$
\begin{aligned}
D''_1 &= D_2 - D_1 \\
D''_i &= D_{i-1} - 2D_i + D_{i+1} \quad (2 \leq i \leq 4)
\end{aligned}
$$

これがデータ D_i の二階差分である。

二階差分は、データが D_{i-1} から D_{i+1} へ変化する時の変化率を表している。

音声認識システムから出力される認識候補に対するスコア D_i は、その値が小さければ小さいほど候補の確からしさが大きくなっている。二階差分法は、スコア各々に対して二階差分を求め、その最大値を計算することによって候補の絞り込みを行なう手法である。 具体的には最大の二階差分値を求め、その値を D''_j とするならば j 番目より後の候補を切り捨てることにより絞り込みを行なう。

例えば、3番目に最大の二階差分値があった場合、4、5番目の候補を切り捨てるといった具合に絞り込みを行なう。ただし二階差分法では、後に説明するが、最大二階差分値が存在したとしても絞り込みが行なえない場合もある。

3.2 適用例

ここでは、実際の入力音声から得られる認識候補と認識スコアを用いて二階差分法の適用例を見ていくことにする。

入力音声「お持ちでしょうか」に対する認識結果が次のようになったとする。

(1)	お持ちでしょうか	(782634.827600)
(2)	思いましょうか	(1223711.659200)
(3)	用いましょうか	(1227388.122300)
(4)	思いますか	(2187719.190700)
(5)	用いますか	(2192537.229900)

この例の場合、1番目が正解である。カッコ内の数字はそれぞれの認識スコアである。この認識スコアに対して二階差分を計算する。

(1) 1223711.659200 - 782634.827600 = 441076.831600
(2) 782634.827600 - 2×1223711.659200 + 1227388.122300 = -437400.368500
(3) 1223711.659200 - 2×1227388.122300 + 2187719.190700 = 956654.605300
(4) 1227388.122300 - 2×2187719.190700 + 2192537.229900 = -955513.029200

この場合、二階差分値の最大値は3番目であるから、4、5番目の候補は切り捨てて良い。実際このようにして絞り込みが行なわれる。正解は1番目に存在するのでこの場合の絞り込みは成功したと言える。

ところが、二階差分法による絞り込みは先程述べたように、すべての場合に適用できるわけではない。二階差分法が適用されない例を次に示す。

この例は、入力音声「登録用紙は」に対する認識結果である。

(1) 登録用紙も (330930.786200)
(2) 登録用紙の (333588.717100)
(3) 登録用紙を (381039.177900)
(4) 登録用紙にも (407645.866800)
(5) 登録用紙は (416921.393500)

このスコアに対し、先程と同様に二階差分を計算してやると次の様になる。

(1) 333588.717100 - 330930.786200 = 2657.930900
(2) 2657.930900 - 2×333588.717100 + 381039.177900 = 44792.529900
(3) 333588.717100 - 2×381039.177900 + 407645.866800 = -20843.771900
(4) 381039.177900 - 2×407645.866800 + 416921.393500 = -17331.162200

この場合、二階差分値の最大値は2番目であるから3、4、5番目の候補を切り捨てて良いことになる。しかし、正解は5番目の候補にあるため絞り込みに失敗したことになる。

このように、二階差分法では絞り込みが行なえる場合と、行なえない場合とがある。

3.3 結果

二階差分法では、認識候補の絞り込みが行なえない場合があることは先に述べた。そこで絞り込みが行なえた場合の信頼度を1、行なえなかった場合の信頼度を0とする。最大の二階差分値を5万、10万単位で区切り、その間の平均の信頼度を求めた結果を示したのが図1、図2である。

このグラフは11名分の音声入力から得られるすべてのデータをもとにして作成したものである。最大の二階差分値は100万までのデータを採用し、その平均信頼度を縦軸にとりグラフにした。この図から、最大の二階差分値が大きければ比較的信頼度が高いことが分かる。ただ今回使用したデータは、正解が認識候補の1番目にあることが多く、最大の二階差分値も値の小さいものが多いため(表2参照)、グラフに少しばらつきが出ている。この傾向は、11名分個別にグラフにするとサンプル数の減少により顕著に現れる。参考までに11名個別のグラフを付録に載せておいたので参照してもらいたい。

ここで重要になってくるのは、二階差分法を用いるか否かを決めるしきい値を決めることである。このグラフからは分かりにくいが、接線の傾きがゼロになる点を求めれば良い。そしてその点をPとすれば、Pより大きい最大の二階差分値を持つ認識候補には二階差分法を用い、Pより小さい最大二階差分値をもつ場合には二階差分法は適用しないと決めて候補の絞り込みを行なう。

このグラフには多少ばらつきがあるが、しきい値をだいたい40万程度に決めてやれば、最も精度の良い絞り込みが行なわれる。

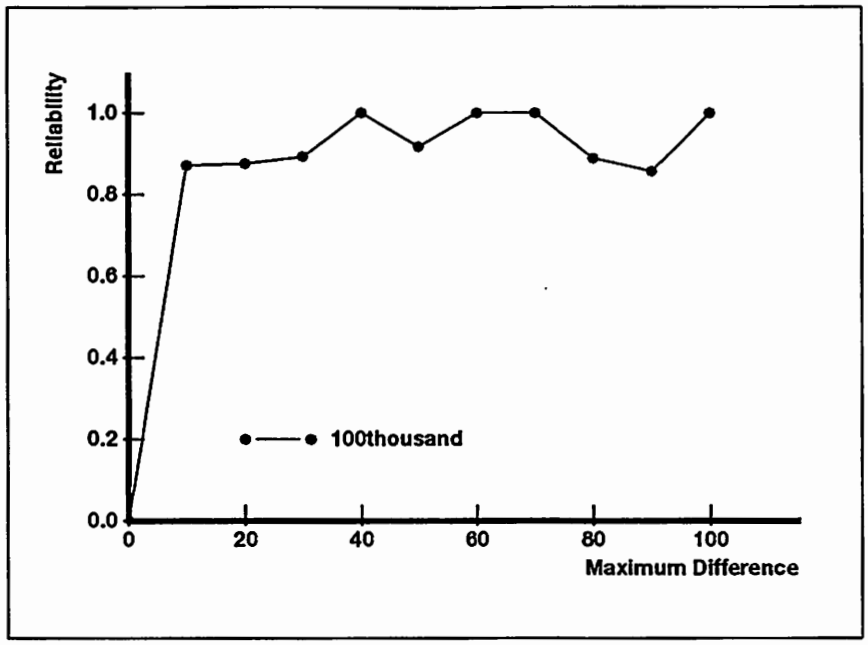

図 1: 二階差分法の信頼度 (10 万きざみ)

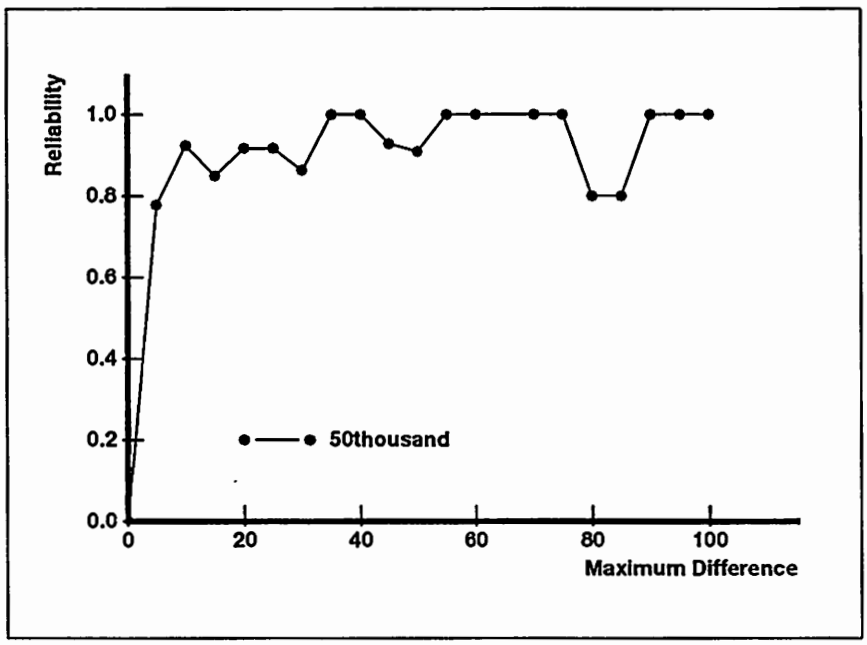

図 2: 二階差分法の信頼度 (5 万きざみ)

表 2: 最大二階差分値による度数分布表

最大二階差分値 ($\times 10^4$)	度数	正解数	平均信頼度
1 ～ 5	36	28	0.777778
6 ～ 10	65	60	0.923078
11 ～ 15	73	62	0.849315
16 ～ 20	48	44	0.916667
21 ～ 25	36	33	0.916667
26 ～ 30	29	25	0.862069
31 ～ 35	15	15	1.000000
36 ～ 40	16	16	1.000000
41 ～ 45	14	13	0.928571
46 ～ 50	22	20	0.909091
51 ～ 55	14	14	1.000000
56 ～ 60	16	16	1.000000
61 ～ 65	6	6	1.000000
66 ～ 70	7	7	1.000000
71 ～ 75	4	4	1.000000
76 ～ 80	5	4	0.800000
81 ～ 85	5	4	0.800000
86 ～ 90	2	2	1.000000
91 ～ 95	4	4	1.000000
96 ～ 100	1	1	1.000000

3.4　二階差分法の信頼性の向上について

二階差分法の信頼性を上げるためには、先ほどの結果ではまだ不十分である。そこで、ここでは、文節ごとの二階差分法の信頼度を見ていくことにする。

文節ごとの信頼度を表したのが、表3である。

この表から分かるように、二階差分法で認識されにくい文節は極めて偏っている。特に「ご」で始まる「ございましたら」(0.272727)、「ご住所」(0.545455)が認識されにくい。

二階差分法を用いる前に、このような認識されにくい文節を予め二階差分法のルーチンから除外してやることができれば、二階差分法の信頼度をさらに上げることができる。しかし、このような文節を除外することは極めて難しい。入力音声「ございましたら」に対して、音声認識システムは「ご」を認識できない。例として「ございましたら」に対する認識結果の一例を示してみると次のようになる。

(1)　お名前から　　(2056206.202300)
(2)　お願いしたら　(2134773.945200)
(3)　お願いするが　(2483957.827100)
(4)　入れませんが　(2513440.662000)
(5)　入れますか　　(2571402.464900)

「ご」は、5個の認識候補のどれについても認識されていない。よって、認識候補列から認識されにくい文節を除くことは難しい。

しかし、もし、なんらかの方法(例えば認識されにくい語の音声波形を予めシステムに入力しておくなど)でその文節を除外することができれば、図1、図2よりも精度の高い信頼度が得られる。認識されにくい文節(この場合は「ご」で始まる文節)を除外したときのグラフを図3に示す。

図3より、若干、信頼度が上がっているのが分かる。

なお、この結果は、アナウンサー1名の音声から作られた[g]の音韻モデル自体が、アナウンサー以外の[g]の発声に適していないことを示唆しているのかもしれない。[g]の音韻は鼻音化して発声されることもあるので、[g]の音韻モデルをいずれにしても見直すことが必要である。

表 3: 文節別信頼度

番号	文節	信頼度	番号	文節	信頼度
1	もしもし	1.000000	21	大阪市	1.000000
2	そちらは	1.000000	22	北区	0.818182
3	会議事務局ですか	1.000000	23	茶屋町	1.000000
4	はい	1.000000	24	二十三です	1.000000
5	そうです	1.000000	25	名前は	1.000000
6	会議に	1.000000	26	鈴木真弓です	1.000000
7	申し込みたいのですが	0.909091	27	分かりました	0.909091
8	登録用紙は	0.909091	28	登録用紙は	1.000000
9	既に	1.000000	29	至急	1.000000
10	お持ちでしょうか	0.909091	30	送らせていただきます	0.909091
11	いいえ	1.000000	31	分からない	0.909091
12	まだです	0.727273	32	点が	1.000000
13	分かりました	1.000000	33	ございましたら	0.272727
14	それでは	0.909091	34	いつでも	0.545455
15	登録用紙を	0.909091	35	お聞き下さい	1.000000
16	お送り致します	1.000000	36	有難うございます	1.000000
17	ご住所と	0.545455	37	それでは	0.909091
18	お名前を	0.818182	38	失礼します	0.636364
19	お願いします	0.909091	39	どうも	1.000000
20	住所は	0.909091	40	失礼致します	0.545455

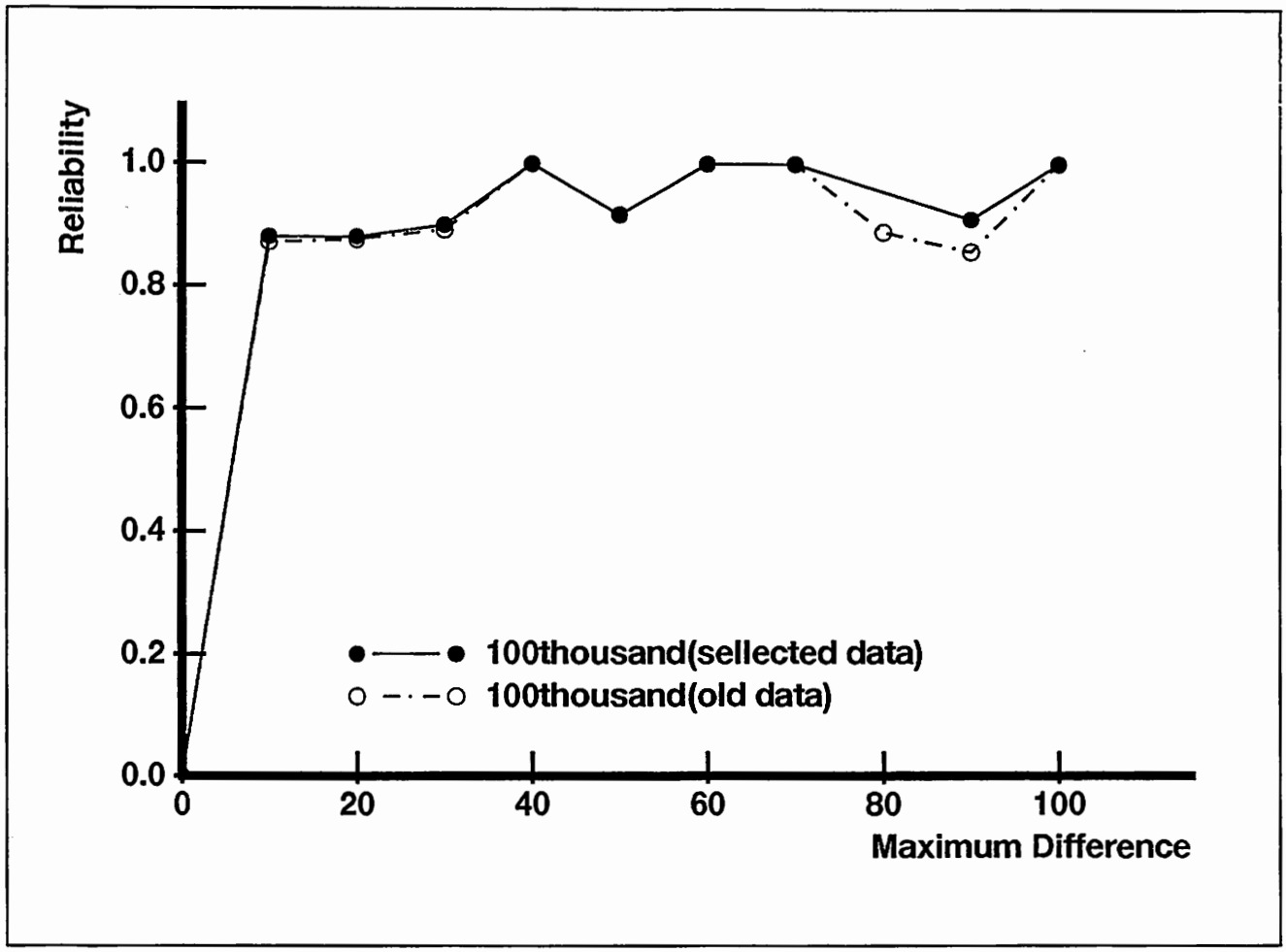

図 3: 二階差分法の信頼度 (改良後)

4 分散法

4.1 原理

この手法は、音声認識候補の認識スコアの分散を計算し、スコアの散らばり具合から候補を絞り込む方法である。

音声認識システムから得られる認識候補列のうち、入力されたものと同じ文節の候補の番号を正解ランクと定義する。

例えば、入力音声「分からない」に対する認識結果が次にようになったとする。

(1) 送らない (984717.430100)
(2) 分からない (1009240.651700)
(3) 宝に (1091326.914200)
(4) かからない (1200081.605100)
(5) ならない (1204289.039000)

このときの正解は2番目にあるので、正解ランクは2である。このように正解ランクを決めていく。正解がない場合には、正解ランクを0と決める。

次に、認識候補5個に対する認識スコアから分散を計算する。分散は、認識スコアを D_i、認識スコアの平均値を m としたとき、次のようにして計算される。

$$v = \frac{1}{5}\sum_{i=1}^{5}(D_i - m)^2 \quad (1)$$

$$m = \frac{1}{5}\sum_{i=1}^{5}D_i \quad (2)$$

分散法では、データの散らばりと正解ランクの関係を用いて、候補の絞り込みを行なっている。具体的には、分散が大きくなると正解ランクが低くなる傾向を用いて、絞り込みを行なう。

4.2 結果

まず始めに、入力音声に対して得られる各候補に対する認識スコアの平均値と分散を計算する。各分散に対する正解ランクの関係を調べ、ヒストグラムを作る。実際にデータから得られた分散と正解ランクから作成したヒストグラムを図4に示す。

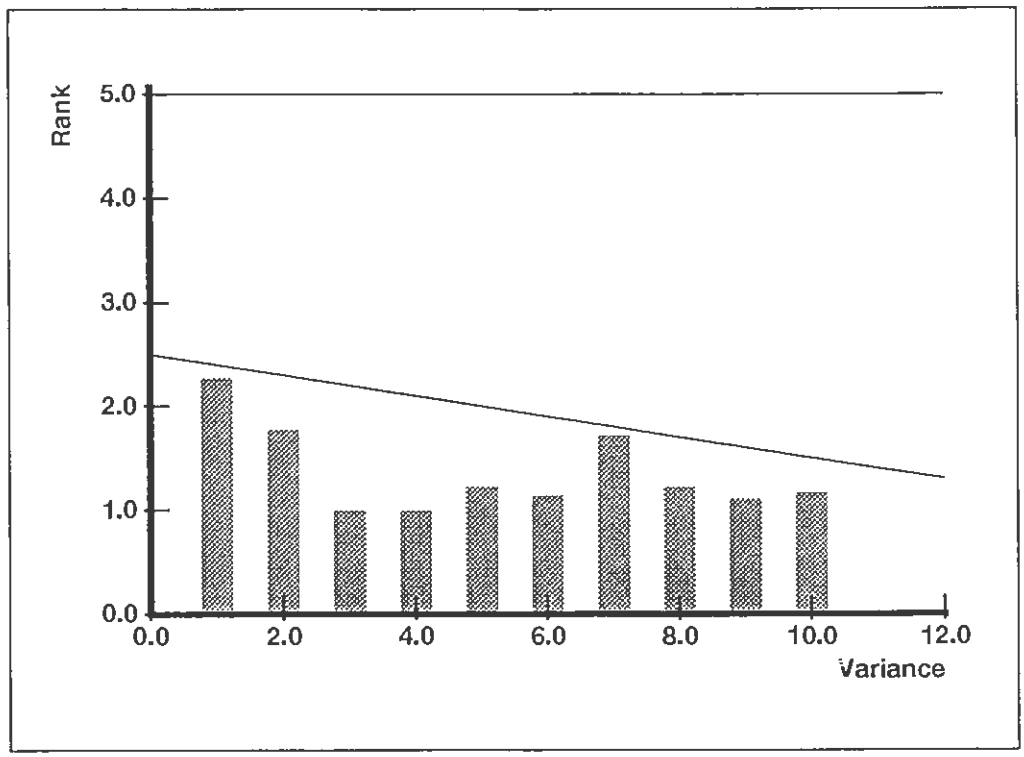

図 4: 分散と正解ランク

このグラフは縦軸に正解ランク、横軸に分散 ($\times 10^9$) をとったものである。グラフから、データの散らばりが大きくなれば正解ランクが低くなることが分かる。音声認識システムからは、5個の候補が一様に出力される。これはグラフで言えば、正解ランク＝5の直線よりも下の部分すべてを出力していることに相当する。

分散法では、グラフで言う斜めの直線より下の部分の候補のみを出力することにより、候補の絞り込みを行なっている。この近似グラフは単調減少になる。このデータの場合、認識候補が半分以下に絞り込まれている。

但し、先ほども述べたが、このデータは正解ランクがほとんどの場合1であるため、近似があまりうまくいっているとは言えない。これも個人別のグラフを見ると良く分かる。参考

までにそのグラフは付録に載せておく。

しかしながら、図4からは、実際的に絞り込みを行なうのは難しい。

そこで、次に、認識候補のランクと正解確率について見てみると正解ランク1の場合が、圧倒的に正解率が高い。正解ランク4、正解ランク5、に至っては正解確率はほとんど0と言ってもいい程である。それは図5を見れば明らかである。

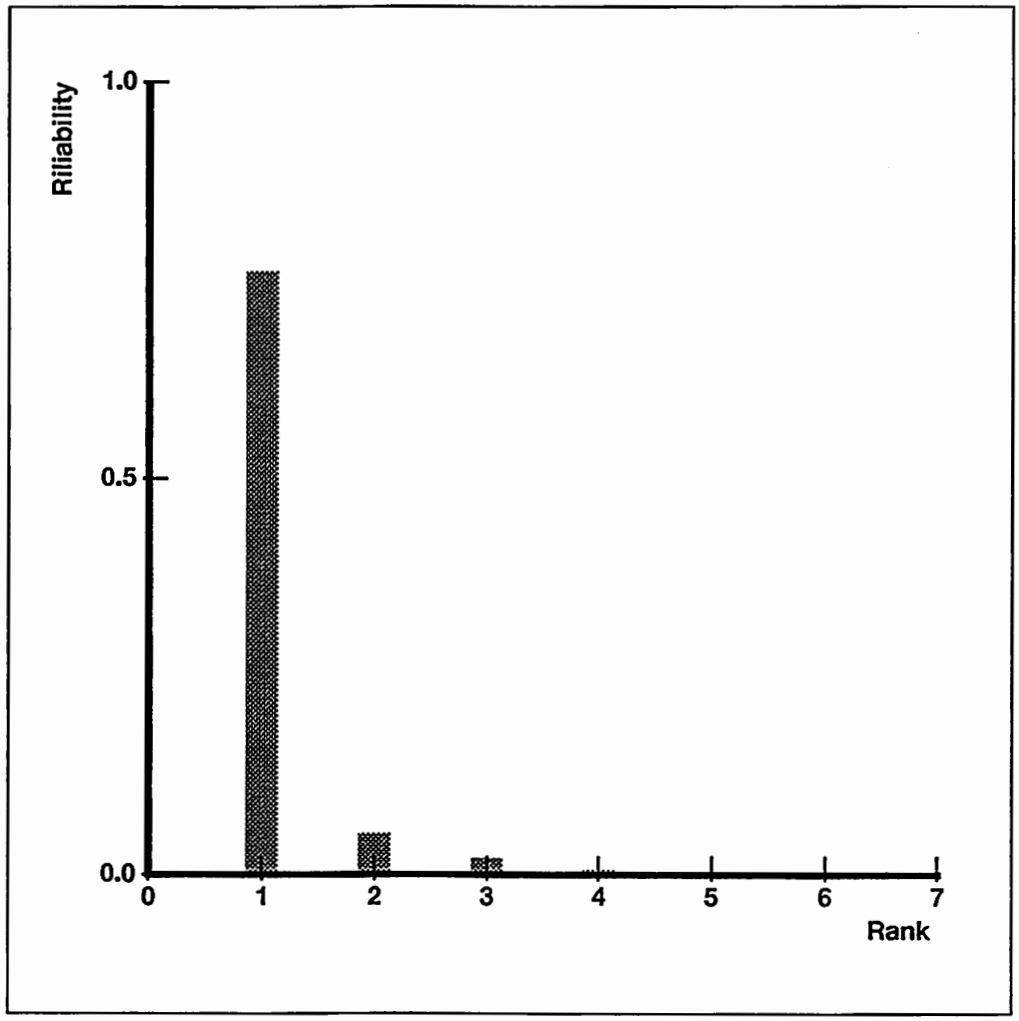

図 5: 正解確率とランク

さらに、図5の各ランクによる正解確率を分散の値によって1次近似すると次のような図6が書ける。グラフは、縦軸に平均信頼度、横軸に分散 ($\times 10^7$) をとった。

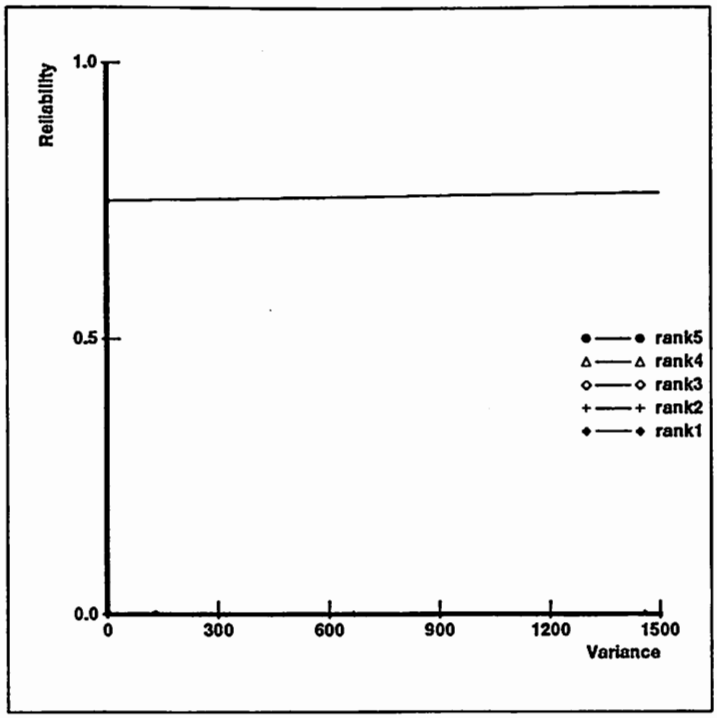

図 6: 正解確率とランクの関係の一次近似

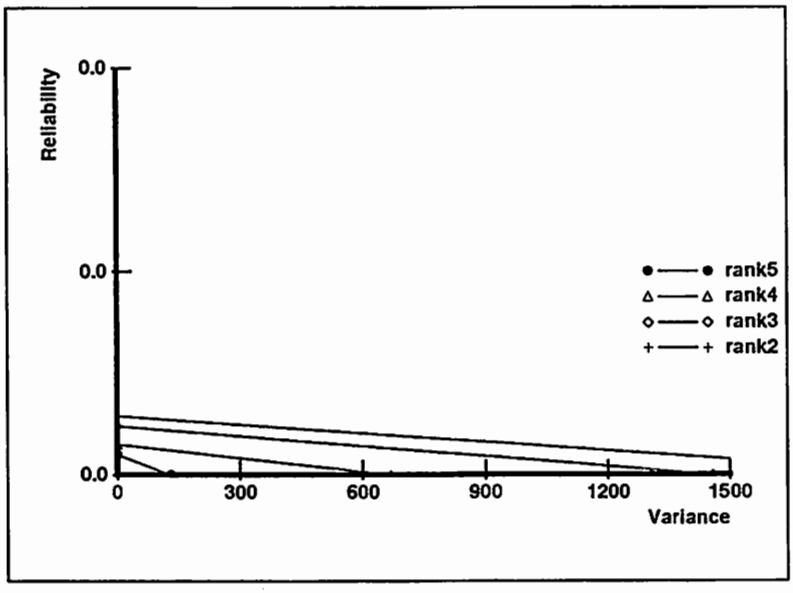

図 7: 正解確率とランクの関係の一次近似 (拡大)

図6からはランク2からランク5までは単調減少のグラフになるのだが、値が小さいために確認できない。これはこのサンプルの場合、ランク1の正解率が非常に高いためである。実際には、ランク2は分散が20000($\times 10^7$)で正解確率0、ランク3は分散が1458($\times 10^7$)で正解確率0、ランク4は分散が668($\times 10^7$)で0、ランク5は分散が130($\times 10^7$)で正解確率0になる。先ほどのグラフを、拡大したものを図7として載せておく。

図7の正解確率は0.01を、上限にとった。これを見れば、ランク1がいかに多いかが分かる。

4.3 絞り込みの実施例

分散法では、音声認識システムから得られたスコアから、分散を計算しその分散値を図6に当てはめて、正解確率が0になっているランクを取り除くことによって絞り込みを行なう。

入力音声「お持ちでしょうか」に対する認識結果が次のようになったとする。

(1) お持ちでしょうか (743054.063700)
(2) 用いましょうか (1044143.067800)
(3) まずでしょうか (1141898.289200)
(4) 思いましょうか (1382484.833700)
(5) お持ちしましょうか (1497771.713700)

分散 = 70233230882.129730

この分散値を図6に当てはめると、ランク3からランク5は正解確率が0なので、ランク3～ランク5、つまり第三候補から第五候補を候補から取り除く。正解は1番目にあるのでこの場合の絞り込みは成功したといえる。このようにして分散法での絞り込みを行なうことができる。

5 正規化確率の計算について

5.1 音声認識候補の尤度について

まず、音声認識候補に対する尤度について説明する。音声認識システムでは入力音声に対して音韻の照合スコアを計算し 、あらかじめ与えられた語彙と文法により許される音韻連鎖の枝を伸ばしながら処理を進めている。そして、音声の存在する全区間の照合が終った時に、照合スコアの良い順に順序をつけて認識候補として出力する。

例えば、音声入力「いいえ」[iie] に対する出力結果が次のようになったとする。

(1) いいえ [iie]
(2) 意味で [imide]
(3) 入れて [irete]
(4) 意味に [imini]
(5) 入れない [irenai]

音声認識システムは、入力に対する一音韻ごとの確率を出し、それらの確率を音韻数だけ掛け合わせることにより、一つの候補の音響的出力確率値を計算している。例えば、第一候補 [iie] では、最初の音韻が [i] である確率 a_1 を出し、次の音韻が [i] である確率 a_2 を出し、その次の音韻が [e] である確率 a_3 を出し、その後その三つの確率を掛け合わせて [iie] である確率 $p(=a_1 \times a_2 \times a_3)$ を出している。実際には計算コストを下げるために確率値の log をとり、乗算演算を加算演算としている。このようにして得られた音響的出力確率値が認識システムから出力される 。

一般に、この値は対象とする入力音声に依存するため、異なる入力に対する認識候補の確からしさを比較する場合には、この値を単純に用いることはできない。

音声認識処理により入力音声の音響的な出力候補として、

$$d_1 \cdots , d_n$$

が得られたとする。各々の出力に対する音響的出力確率を P_d で表すと、

$$\sum_{i=1}^{n} P_d(d_i) = 1$$

となる。この音響的な出力候補 $d_1,\cdots,d_n$ のうちで、言語的に意味をなす語に変換できるものを

$$d_1,\ \cdots\ ,\ d_m \quad (m \leq\ n)$$

とし、それらを単語に変換したものを $w_1,\cdots,w_m$ とする。

音声認識システムから出力されるのは $w_1,\cdots,w_m$ であり、その出力確率として

$$P_d(d_1),\cdots,P_d(d_m)$$

(音響的出力確率値) が出力されている。しかし、この確率値は正確には純粋な尤度としては扱えない。なぜなら、言語的に意味をなさない語 $P_d(d_{m+1}),\cdots,P_d(d_n)$ を含めた確率値であるからである。つまり下式

$$\sum_{i=1}^{m} P_d(d_i)$$

が 1 にならない。ここで欲しいのは $w_1,\cdots,w_m$ の正確な尤度 P_w(言語的出力確率値) である。この尤度が求まれば、この値が極端に低い候補を切り捨てることにより候補の絞り込みが可能となる。

$$\sum_{i=1}^{m} P_w(w_i) = 1$$

この値はすでに述べてきた分散法により求めることができる。このことを次節以下で説明する 。

5.2 尤度の計算

分散法で言語的出力確率値を求めるためには図 6を用いて絞り込みを行なったのと、同じ要領で行なう。

まず、入力音声に対する認識スコアの分散を計算し、その値をグラフに当てはめる。その分散値で正解確率が 0 になっている候補に関しては取り除き、その候補の音響的出力確率を 0 にする。残っている認識候補についても、グラフから正解確率を求める。音響的出力確率 $P_d(d_i)$ を p_1, p_2, p_3, p_4, p_5 とし、言語的出力確率 $P_w(w_i)$ を q_1, q_2, q_3, q_4, q_5 とすれば、尤度である言語的出力確率は次のようにして求まる。

$$q_i = \frac{p_i}{\sum_{j=1}^{5} p_j}$$
$$(1 \leq i \leq 5)$$

5.3 尤度の計算例

次に、実際に、例を用いて説明する。入力音声「既に」に関する認識結果が次のようになったとする。

(1) 既に (371817.604600)
(2) 千円に (552406.191600)
(3) 千に (606181.220700)
(4) 三千に (704065.577000)
(5) 三人に (872411.120800)

分散 = 27424661304.450531

この値を図6に当てはめると、第三候補、第四候補と、第五候補は切り捨てられる。次に、第一候補と第二候補の、正解確率をグラフから求める。

その値が次のように決まったとする。

$$P_d(d_1) = 0.750000000$$
$$P_d(d_2) = 0.000010000$$
$$P_d(d_3) = 0.000000000$$
$$P_d(d_4) = 0.000000000$$
$$P_d(d_5) = 0.000000000$$

この値から正規化確率を求めると次の様になる。

$$P_w(w_1) = 0.999998667$$
$$P_w(w_2) = 0.000001333$$
$$P_w(w_3) = 0.000000000$$
$$P_w(w_4) = 0.000000000$$
$$P_w(w_5) = 0.000000000$$

6 発声速度と認識率について

ここでは、統計分散特性からは少し外れるが、発声速度と認識率について見ていきたい。

今回の実験で用いた音声認識システムでは、標準となるアナウンサーの音声をもとにして、入力音声の認識を行なっている。入力音声の速度は、1秒間に発声するモーラ数(モーラとはおおよそ日本語のカナに相当する)で表現する。標準としたアナウンサーの音声は、単語発声で 5.4mora/sec、 文節発声で 7.1mora/sec 、文発声で 9.1mora/sec となっている。今回の認識実験では、標準パターンの作成に単語発声のデータを用いて、文節発声の音声で評価実験を行なっている。

発声速度と認識率の関係をグラフにしたのが図8である。横軸に秒単位のモーラ数をとり、縦軸に認識率をとってグラフにしてみた。

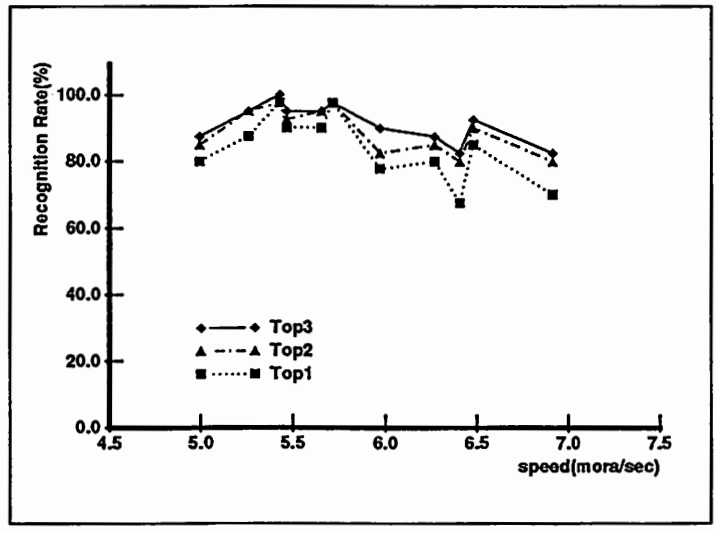

図 8: 発声速度と認識率の関係

Top1、 Top2、 Top3 はそれぞれ、第一位までの累積認識率、第二位までの累積認識率、第三位までの累積認識率を示している。

図8をみると、だいたい 5.4 〜 5.5 mora/sec で最高の認識率になっており、それより発声速度が速くても遅くても認識率は下がっている。したがって、今後は、発声速度の違いに対して頑健な音声認識技術の研究が望まれる。

また、 6.5mara/sec 位でも認識率の高い話者がいるので、発声速度以外の要因、例えば子音を明瞭に発音したり、口を大きくあけて発音したりなどがあるかもしれない。

7　むすび

本報告では、二階差分法、分散法、発声速度と認識率の関係を11名のデータ分析結果と共に見てきたわけであるが、今回使用したデータはサンプル数が十分でなく、さらにかなり偏ったデータと考えられるため、この結果をそのまま一般の場合に利用するのは危険かもしれない。だが、話者適応や不特定話者音声認識において認識率を上げるには、何を研究すべきか示唆してくれるものがある。

まず、今回の実験で明らかになったことを整理すると、個人差がかなり広範囲で見られたこと、データの種類によってかなりの偏りが見られたことなどがある。

個人差に対する問題に関しては、ここで扱った手法も含めた統計的な手法で正規化することは有効であると考えられる。しかし、個人差自体を無くすことが不可能である以上、正規化することにより生じる弊害を無くすことはできない。そこで、話者適応や不特定話者音声認識システムでは、これらの弊害をいかに少なく抑えることができるかが重要な問題になってくる。

その問題を解決するためには、まず、今回の実験に用いたデータのような偏りがない、十分幅広いデータを用いて正規化を行なうこと、さらに音声データ入力時に防げる個人差の要因、つまり発声方法や発声速度などをできる限り取り除くことなどが挙げられる。

さらに、特に、ばらつきの激しいデータに関してはシステム内にその情報を記憶させ、その種の入力があった場合には、統計的な手法以外の方法で処理を行なうようにすることも、必要になってくるであろう。

謝辞

本報告書は著者の1人が実習生として1991年8月19日から9月13日までATR自動翻訳電話研究所に滞在した際に行なった研究成果をまとめたものである。

プログラムの作成からレポートの書き方まで、何から何までご教授いただいた竹沢寿幸氏、実習の機会を与えてくださった樽松明社長並びに森元逞データ処理研究室長をはじめとするATR自動翻訳電話研究所の皆様に感謝致します。

参考文献

[1] 坂野, 森元: “ 統計による音声認識候補の絞り込みに関する考察 ”, *ATR Technical Report*, TR-I-068, (1990-9).

[2] 竹沢, 大倉, 森元, 嵯峨山, 樽松: “ 日英音声言語翻訳実験システム *SL-TRANS 2*”, 日本音響学会講演論文集, (1991-10).

A 付録 (二階差分法のグラフ)

ここでは、本文中で紹介できなかった二階差分のグラフをまとめておく 。二階差分のグラフは横軸に最大二階差分値 ($\times 10^4$)、縦軸に平均信頼度をとってある。

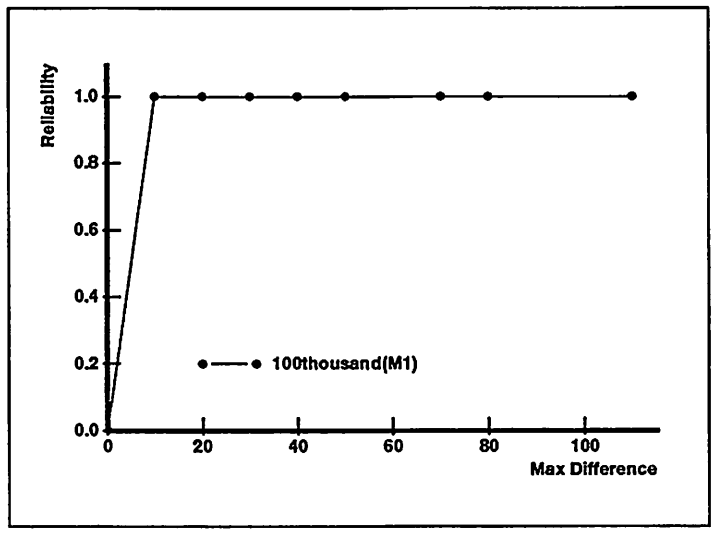

図 9: 二階差分法 男性話者 A

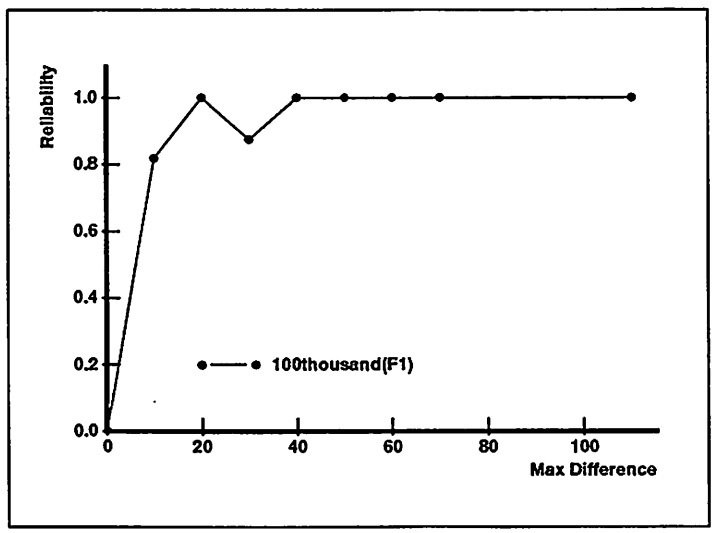

図 10: 二階差分法 女性話者 A

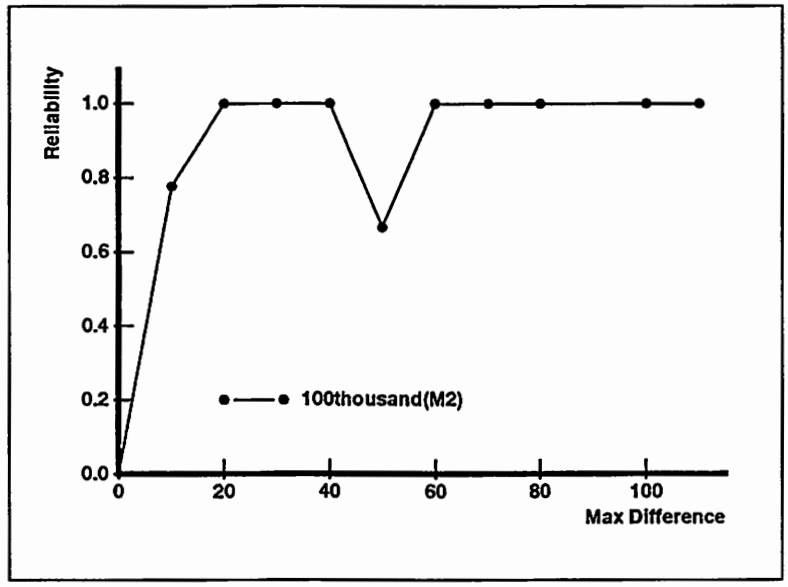

図 11: 二階差分法 男性話者 B

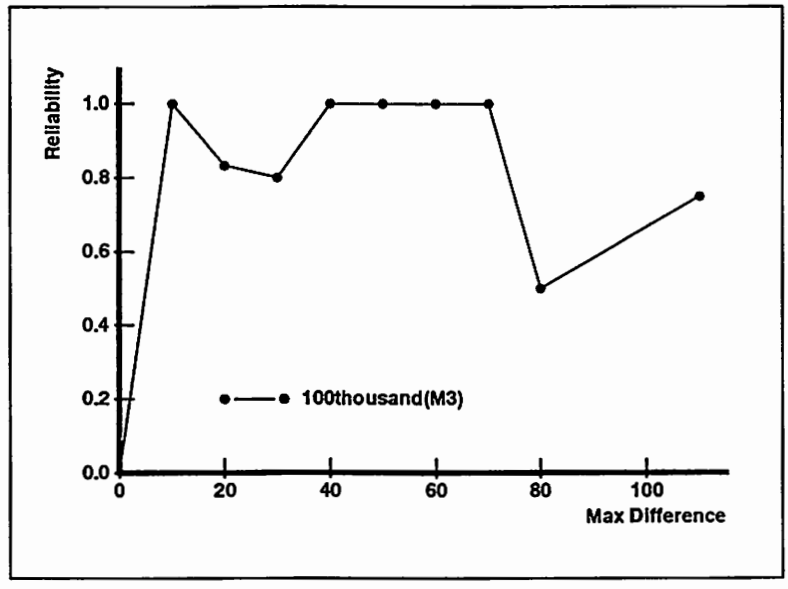

図 12: 二階差分法 男性話者 C

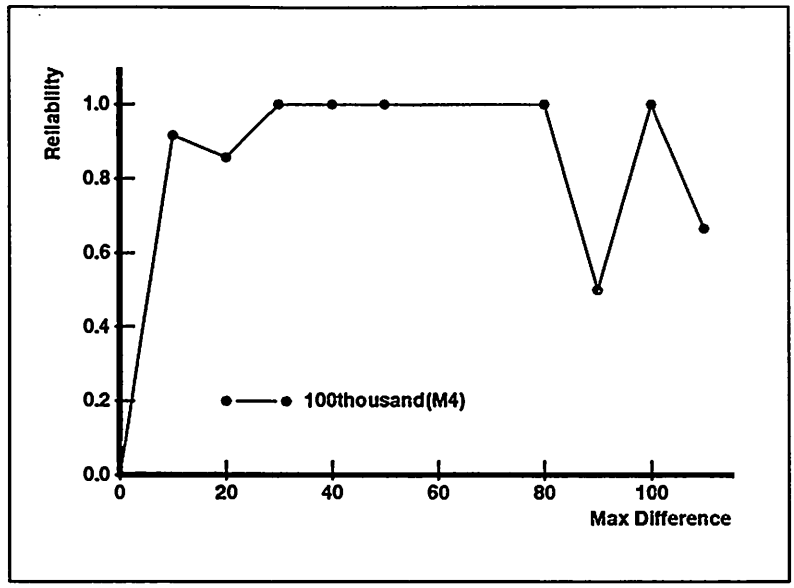

図 13: 二階差分法 男性話者 D

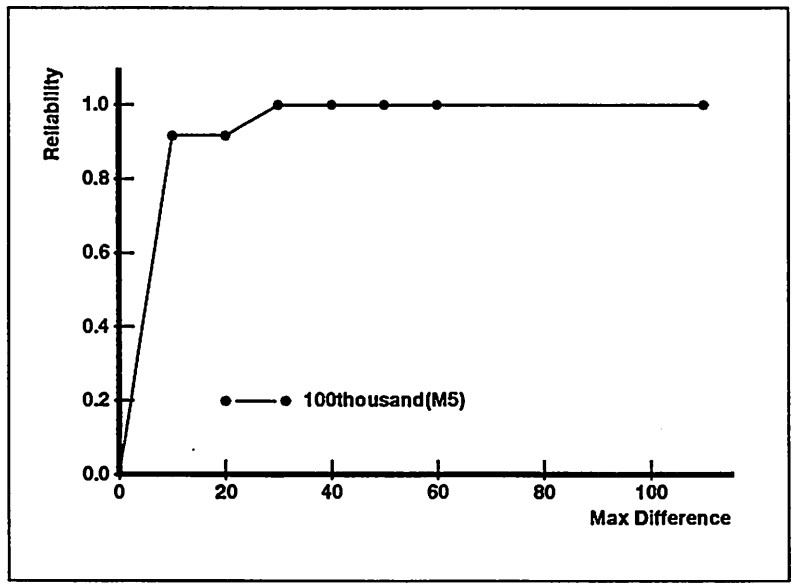

図 14: 二階差分法 男性話者 E

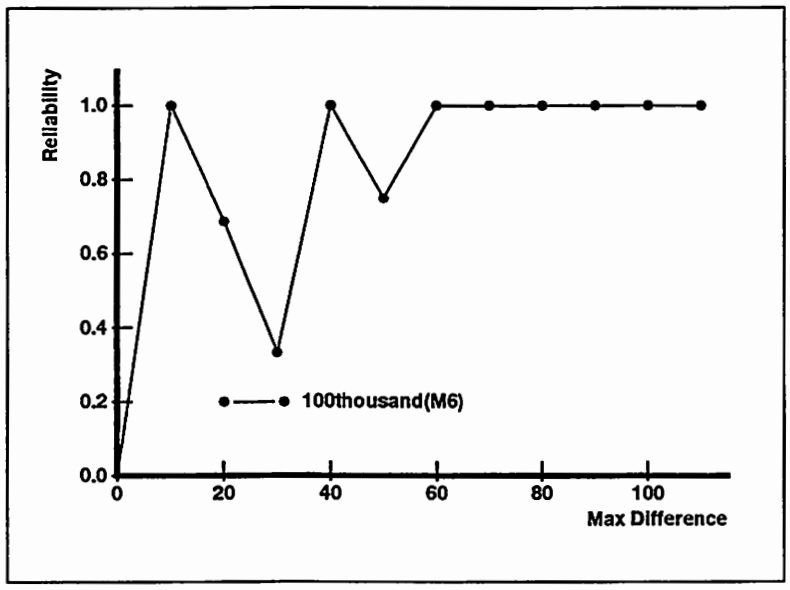

図 15: 二階差分法 男性話者 F

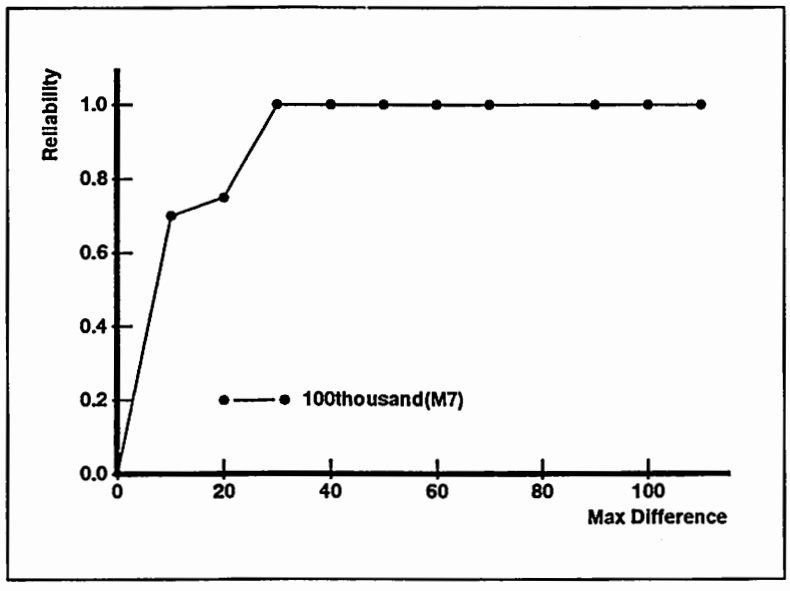

図 16: 二階差分法 男性話者 G

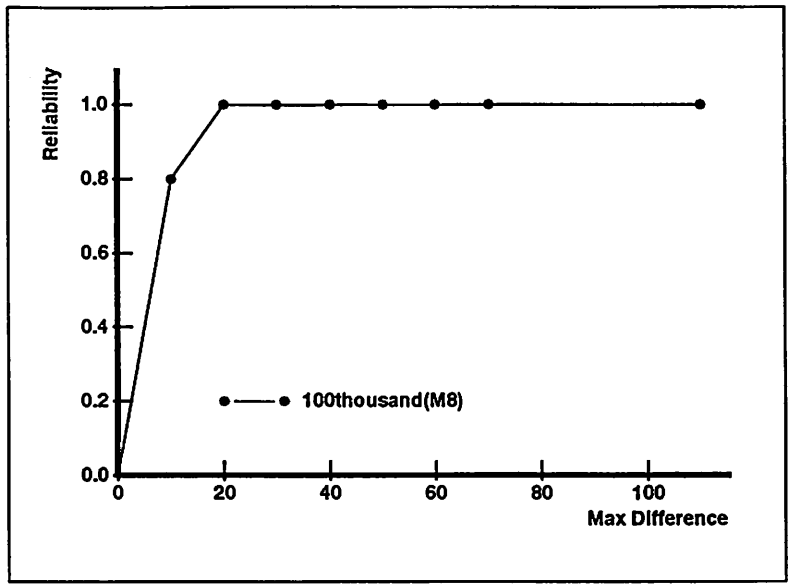

図 17: 二階差分法 男性話者 H

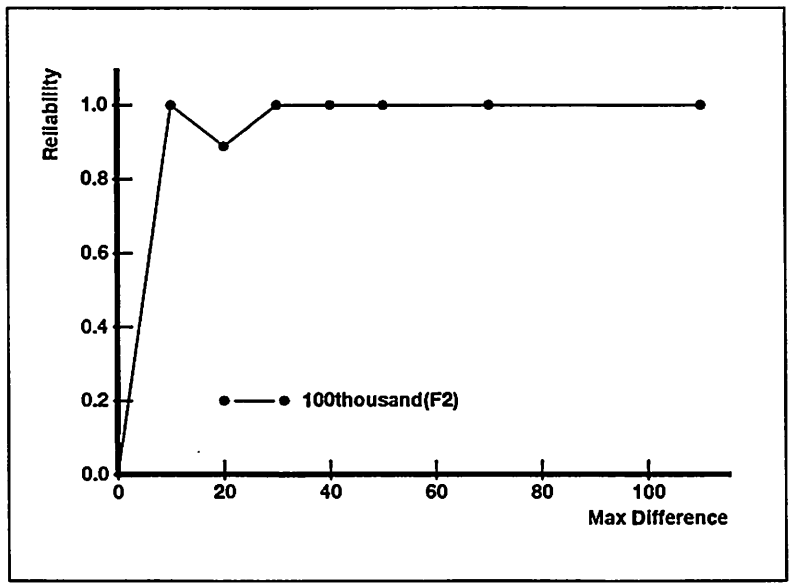

図 18: 二階差分法 女性話者 B

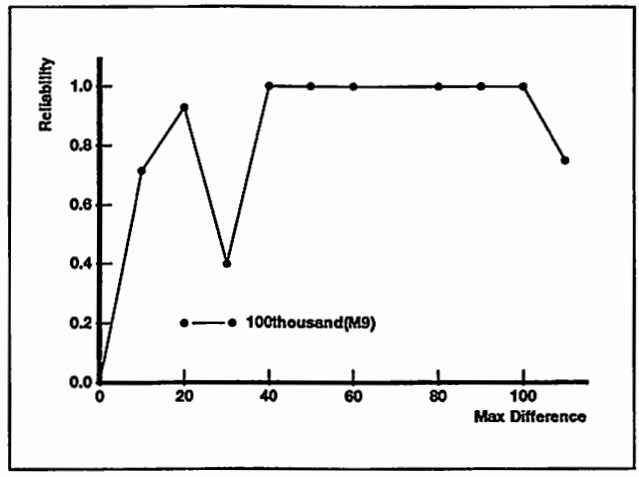

図 19: 二階差分法 男性話者 I

B　付録 (分散法のグラフ)

ここでは、本文中で紹介できなかった分散のグラフをまとめておく 。分散のグラフは横軸に分散 ($\times 10^9$)、縦軸に正解ランクをとってある。

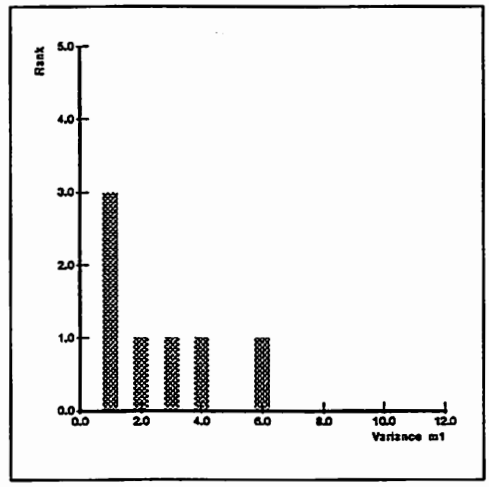

図 20: 分散法 男性話者 A

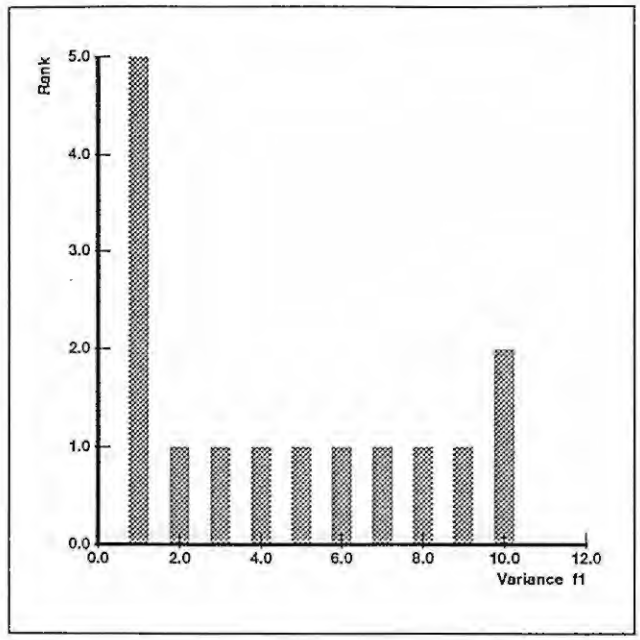

図 21: 分散法 女性話者 A

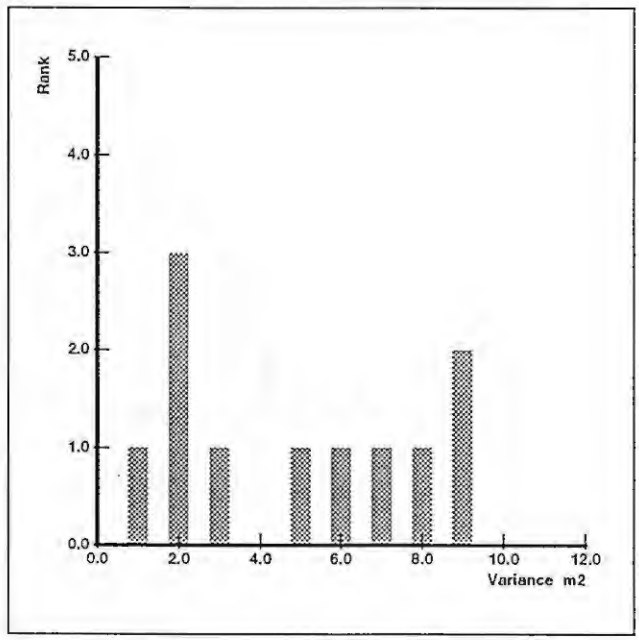

図 22: 分散法 男性話者 B

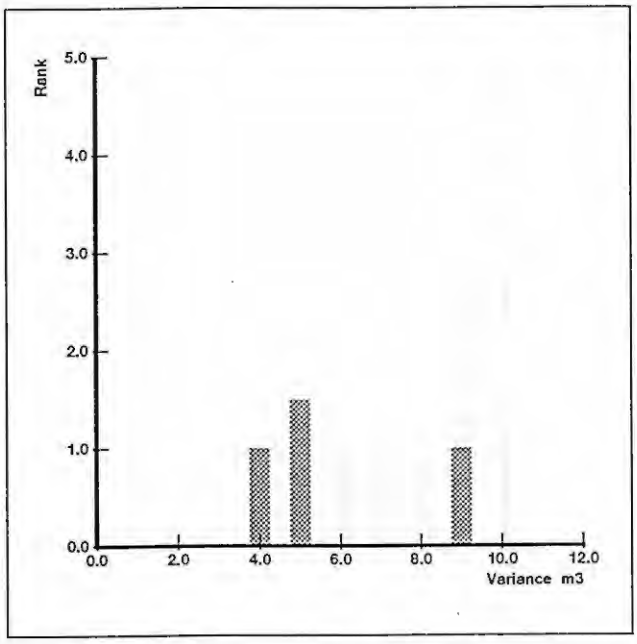

図 23: 分散法 男性話者 C

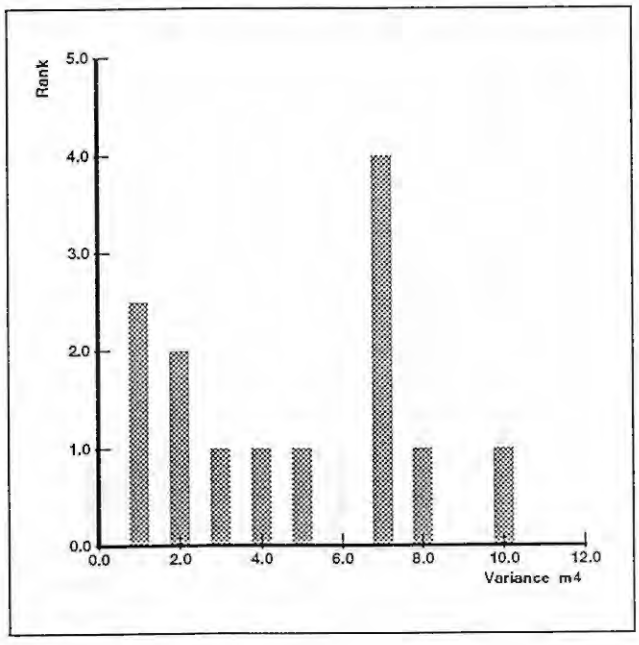

図 24: 分散法 男性話者 D

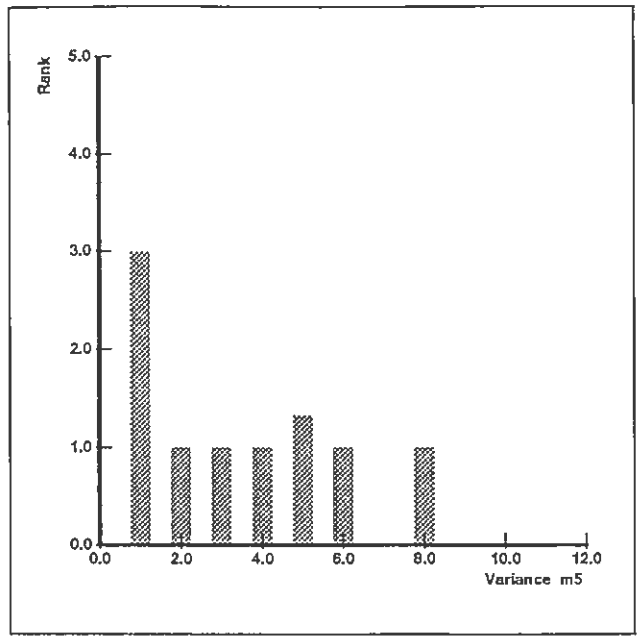

図 25: 分散法 男性話者 E

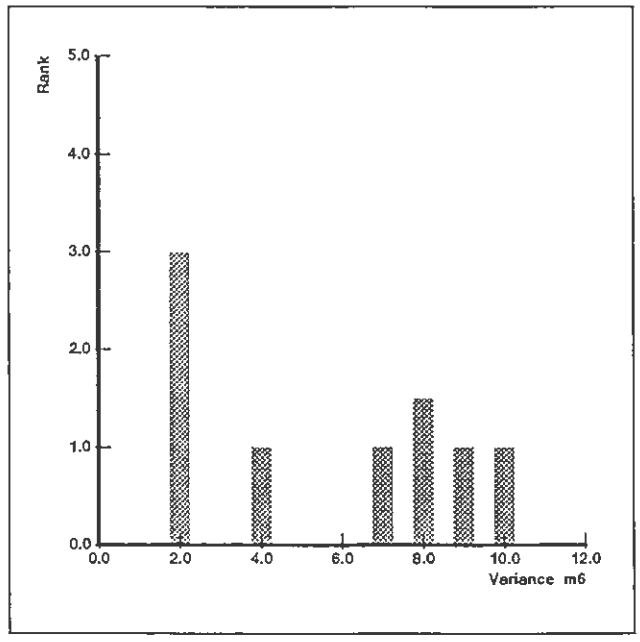

図 26: 分散法 男性話者 F

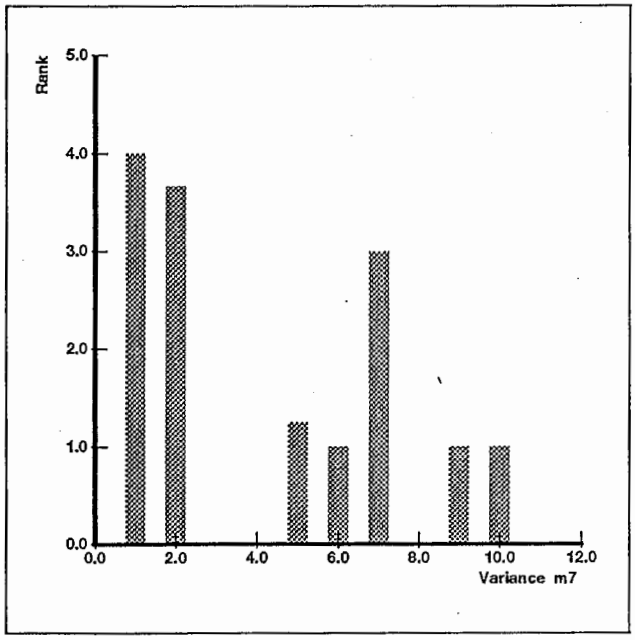

図 27: 分散法 男性話者 G

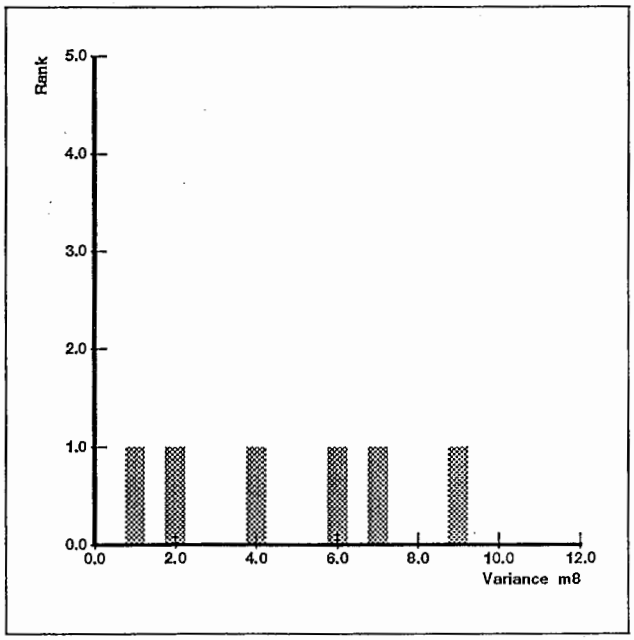

図 28: 分散法 男性話者 H

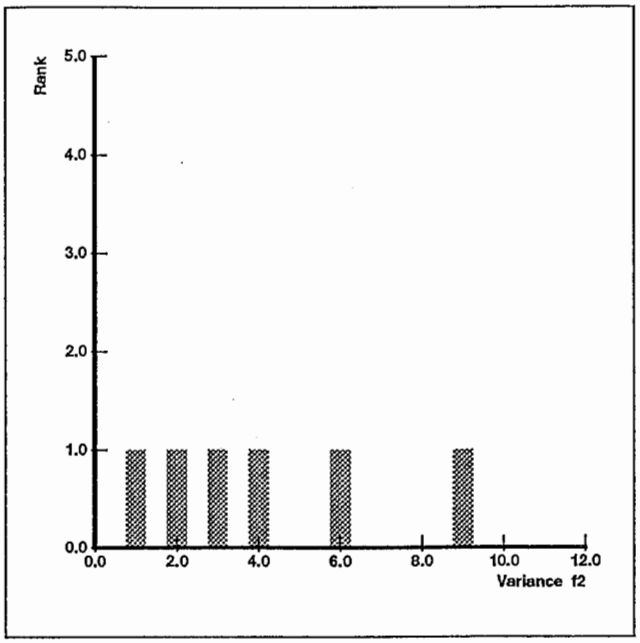

図 29: 分散法 女性話者 B

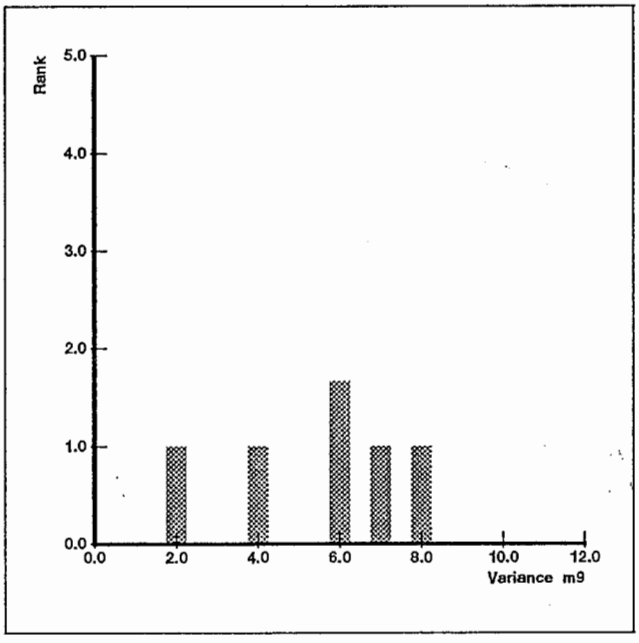

図 30: 分散法 男性話者 I