非公開

TR-I-0127

音声合成素片の接続点でのスペクトル歪みと音韻環境について

Unit extraction context and inter-unit discontinuities

†橋本賢治　武田一哉　安部勝雄

†Kenji HASHIMOTO, Kazuya TAKEDA and Katsuo ABE

1990.1

内容更概

種々の音韻複合単位を用いる規則合成システムにおいて、単位間接続歪みを反映した単位選択基準を確立するために行った、統計的な分析の結果を報告する。報告内容は、(1) 種々の接続音韻環境下における接続歪みの分析、(2) 最適接続点検索アルゴリズムの改良、の2点である。

ATR 自動翻訳電話研究所
ATR Interpreting Telephony Research Laboratories

†早稲田大学理工学部
†Waseda Univ.

目次

1. はじめに 1
2. 使用データと実験条件 1
3. 各分類での距離 2
3.1 最適距離及び境界距離 2
3.2 最適点の再検討 2
4. 再分類 3
4.1 分類の細分化 3
4.2 境界距離との相関 4
5.まとめ 4

1. はじめに

種々の音韻複合単位を用いる規則合成方式[1]の開発には、単位選択基準の確立が重要である[2]。特に、合成単位間の不連続性が合成音の品質劣化要因であることから、単位間の連続性を選択基準に反映させることにより、より高品質な合成音を得ることが、期待される。

そこで本稿では、音韻環境から合成単位間の連続性を予測する尺度を確立するため、音韻環境と単位間接続歪みとの関係を定量的に分析する。

実際には、接続歪を表すパラメータとして、接続フレーム間でのFFT改良ケプストラム(30次元)のユークリッド距離を用いて、種々の音韻環境下における接続歪みの統計を取った。

2. 使用データと実験条件

(1)使用データ

統計を取るために用意したデータセットは、

模擬電話会話

- ○ k01~k08 (除:k06) 7文

音韻バランス文セット

- ○ a01~a50 (除:a30,a39,a50) 47文
- ○ b01~b50 (除:b37,b47,b42) 47文
- ○ c01~c50 (除:c14, c41) 48文
- ○ d01~d50 (除:d01,d20,d22,d29,d31,d36) 44文

の合計193文章である。除外対象は、現システムにとって長すぎる文章や外来音韻である「ディ、ティ、チェ」等を含む文章である。

(2)実験条件

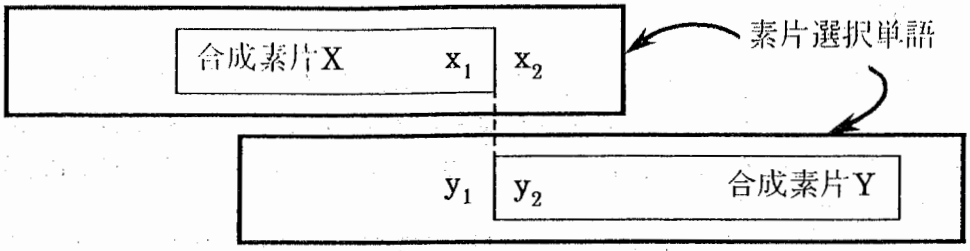
上記の各文を入力として、文献[1]に示す方法により最適な単位系列を得たのち、全ての単位間の接続点におけるケプストラム距離を算出した。ここで、単位間の接続点として、(1)文献[3]に述べる方法により得られた接続点、(2)音韻ラベルによりえられた音韻境界、の2点に関して統計を取った。

文献[3]に述べられた方法では、選択された合成単位間の最適接続点を、以下の手順で音韻環境に応じて探索している[3]。

合成素片をX,Yとしたときに

x_1, x_2 : 合成素片Xの末尾、後続音韻

y_1, y_2 : 合成素片Yの先行、先頭音韻



とすると、接続音韻環境を4分類し、それぞれ以下に示す探索区間において接続歪みが最小となる接続点を探索する。

	音韻環境	探索区間	
○	$x_1=y_1, x_2=y_2$	Same All	(有声音全体での接続)
○	$x_1=y_1, x_2\neq y_2$	Same Front	(母音での接続)
○	$x_1\neq y_1, x_2=y_2$	Same Back	(有声子音での接続)
○	$x_1\neq y_1, x_2\neq y_2$ & $x_1=y_2$	Different	(同一母音連続での接続)

しかし、音韻環境の分類は4種類では大まかすぎるので、本稿では後続合成素片Yの先頭音韻y_2により表1のように11種類に再分類した。ただし、この分類はあくまでプログラムから暫定的に使ったものである。結果として分類[3,5,6]のデータ数が少ないが、元々出現頻度の比較的少ない音韻であり接続点として選択される割合も低かったためと思われる。

3. 各分類での距離

3.1 最適距離及び境界距離

各分類による最適距離(AdjustDistance)及び境界距離(SegmentDistance)の平均値を表2に、また分散とあわせて図1に示す。さらに、最適距離のみの平均と分散を図2に示す。

全体的にどの分類においても、境界値より最適値の方が改善されており、平均値で約4.0~10.0ほど小さくなっている。

当初、分類ごとにみるとはっきりした傾向が見えてくるかと思われたが、母音、半母音、鼻音が平均で2.0をやや越えているだけで、他は1.0~2.0の間で分布しており大きな差はなかった。

3.2 最適点の再検討

個別にデータをみていくと大抵は境界値が最適値より大きかったが、時々、特に無声破裂音や促音等で境界値が最適値より小さくなる場合があった。この原因を知るためプログラムを再検討してみたところ、SameFrontとSameBackの場合、境界より1フレームずらした点より音韻の中心フレームまでを探索区間にしていた。同一音韻がある場合、常にその音韻で接続するという考え方ならばこのままで良いと思われる。しかし、例えばy_2が無声破裂音等でx_2がEnd、あるいはx_2が同じカテゴリである場合の接続では、閉鎖区間と無音部や閉鎖区間同志で接続する方がよりスペクトル歪が低減できるものと思われる。

そこで、このことを確認するために、SameFrontとSameBackにおいて境界及び境界前後5フレームも探索区間にして最適距離(変更距離[ChangedDistance])を求めなおした。2、3フレームしかない半母音等には長すぎるかもしれないが、プログラムを全面改定する時間も無かったので、暫定的に固定フレームとした。

各分類による変更距離、最適距離との差及び変更割合を表3に示す。変更されたもののみの最適距離、変更距離、その差、境界距離を表4に示す。また、変更距離の平均と分散を図3に示す。

結果として、全体的には無声破裂音、Unknown(促音、拗音)、半母音において値が平均で約0.5~1.0ほど小さくなっている。個別に検討してみると、例えば無声破裂音の分類で

◇ kaette(kaette)とsekitaN(kita)の接続
(Segment=6.92) Adjust=3.27 Changed=0.47 (差)2.80

◇ gakkari(kkari)とmitomeru(to)の接続
(Segment=1.40) Adjust=2.82 Changed=0.39 (差)2.43

の様に、特にEndの無音部との接続で約1.0~3.0ほど値が小さくなった。また、傾向としては、変更距離をもとに分類[1,3,7]、[4,5,6,10]、[2,8,9,11]の順で平均値が大きくなっている。

4. 再分類

4.1 分類の細分化

後続合成素片Yの先頭音韻y_2による分類だけでは、各々の差はそれほど明確ではなかった。そこで、データ数の少ない分類[3,5,6]を除いた8種類の分類をさらに各々以下のように分類した。

(1) 分類[1,2,7] : SameFrontのときのみ計算
先行合成素片Xの後続音韻x_2により、以下の3つに分類する。
- Same Category (x_2とy_2が同じカテゴリ)
- End (x_2がEnd、すなわち素片Xが選択単語の語尾までのとき)
- else (上記以外)

(2) 分類[4,9] : SameAll、SameFrontのとき計算
以下の4つに分類する。
- Same (SameAllのとき)
- Same Category
- End
- else

(3) 分類[8.10] : SameBack、SameAll、SameFrontのとき計算
以下の6つに分類する。
- Same
- Top (y_1がTop、すなわち素片Yが選択単語の語頭からのとき)
- S.F else (SameBackでTop以外)
- Same Category

- End
- S.B else (SameFrontで上記以外)

(4) 分類[11] : 4種類すべてで計算
以下の6つに分類する。
- Different (Differentのとき)
- Same
- S.F else (SameBackでTop以外)
- Same Category
- End
- else

(1)~(4)の分類のデータ数を表5に、分類による境界距離、最適距離、変更距離の各平均値を表6~9に示す。また、(1)~(3)の最適距離、変更距離を図4~6に、境界距離を図7に各々の分類で示す。ただし、(1)で分類[7]の促音、拗音の場合には同じカテゴリでの接続がほとんどなかったので、無声破裂音の[p, t, k]を代用した。分類(4)は、他と違ってDifferentがデータの半分以上を占めているため、図示からは除外した。また、分類[10]の撥音には(Same Category) (Top)がないので空欄とした。

(1)の分類では、無声破裂音、Unknown(促音、拗音)が変更距離においてEndやSame Categoryで大きく改善されているのに対し、無声摩擦音はSame Categoryでは良いものの、逆にEndでは悪くなっている。これは全体を通して言えることだが、Endと有声の渡りとの接続は良くなく、無声摩擦音もこの中に含まれるものと考えられる。

(2)~(4)の分類では、SameやSame Categoryで良く、EndやTopは悪い。個別には撥音が、他の有声音である母音や鼻音に比べてどの分類においても安定した値を出している。

個別の具体的な例を表10に示す。

4.2 境界距離との相関性

図7より、境界距離の傾向と変更距離の傾向とが各分類で類似していた。そで、境界距離と変更距離の相関性を見るために、例として無声破裂音と半母音についてその相関を図8、9に示す。

両者とも相関性は見られるが、分布も広く例外も幾つか存在する。境界距離が大きくても変更距離が小さくなるのは、たいてい母音中心点まで大きくシフトして定常部で接続したときである。一方、境界距離が小さい割に変更距離が小さくならないのは、探索区間が狭かったり、違うカテゴリへの渡りでの接続で見られる。素片選択のひとつの要素としてこの境界距離を閾値として用いることは、可能であると思われる。

5. まとめ

今回の結果はあくまでケプストラム距離のみをデータとしており、受聴試験は都合により行えなかったが、裏付けするために必要不可欠である。

境界前後5フレームを考慮するだけで、無声破裂音や促音で大きく改善された。他のものも多少改善されているが、5フレーム固定でなく音韻の中心フレー

ムまでを探索区間とする方が、更に改善されるものと思われる。

無声破裂音や促音では同じカテゴリや先行素片の切り出しが語尾の場合、前部の同一音韻で接続するよりも、無音部と安定した閉鎖区間あるいは閉鎖区間同志で接続する方が小さい値になった。有声音や無声摩擦音では、逆に語尾から切り出した素片との接続は良くなっかた。すなわち、有声間の渡りと、有声無声間の渡りでの接続が良くないと言える。以上の結果から、合成素片境界の音韻環境だけでなく、素片の切り出しの音韻環境との関係についても考慮して素片の選択を行うことで、より接続歪を抑えることになる。

接続歪(ケプストラム距離)と実際の聞こえの関係を明らかにするため、原音声の音韻境界における歪との比較や、受聴試験が必要である。境界距離との相関性より素片選択において境界距離を閾値にすることも可能であると思われ、より多くのデータを用いて受聴試験と併せてこれらの値を決定していくことが望まれる。また、分類自体も大まかであるので、どの様なときにどの程度接続歪が抑えられるかを更に細かく分類していくことも望まれる。

謝辞

実習の機会を与えて下さった榑松社長、貴重な助言をいただいた鹿野室長、小森研究員および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] 匂坂「種々の音韻連接単位を用いた日本語音声合成」
音講論集,1987.10

[2] 武田, 安部, 匂坂「入力音韻系列に応じた音声合成素片選択法の改良」
音講論集,1989.10

[3] 安部, 匂坂, 武田「音韻環境に応じた音声合成素片の接続方法の検討」
音声研資,SP89-66, 1989.11

図1 各分類によるDISTANCEの平均と分散
(Adjust & Segment)

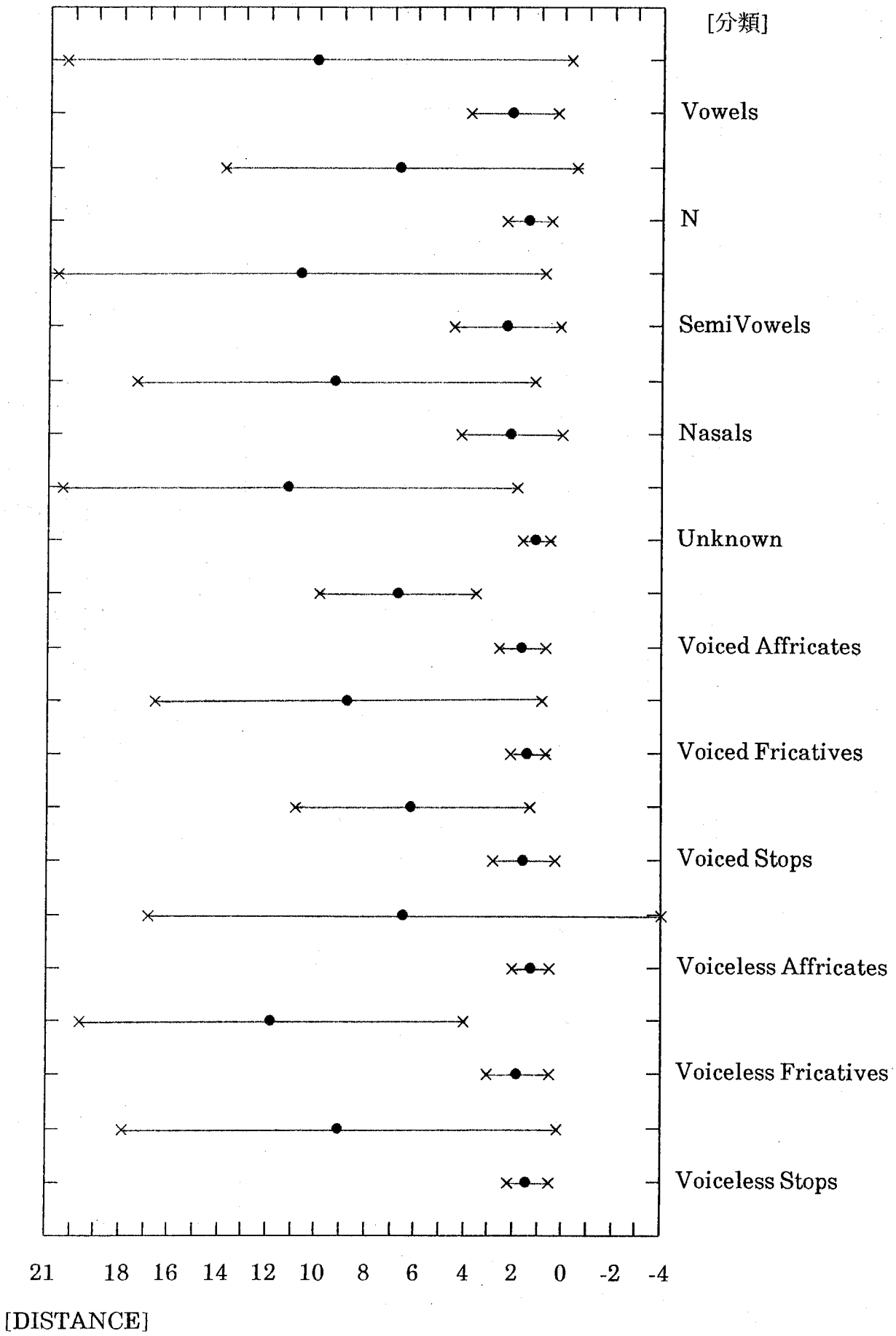


図2 DISTANCEの平均と分散
(Adjust)

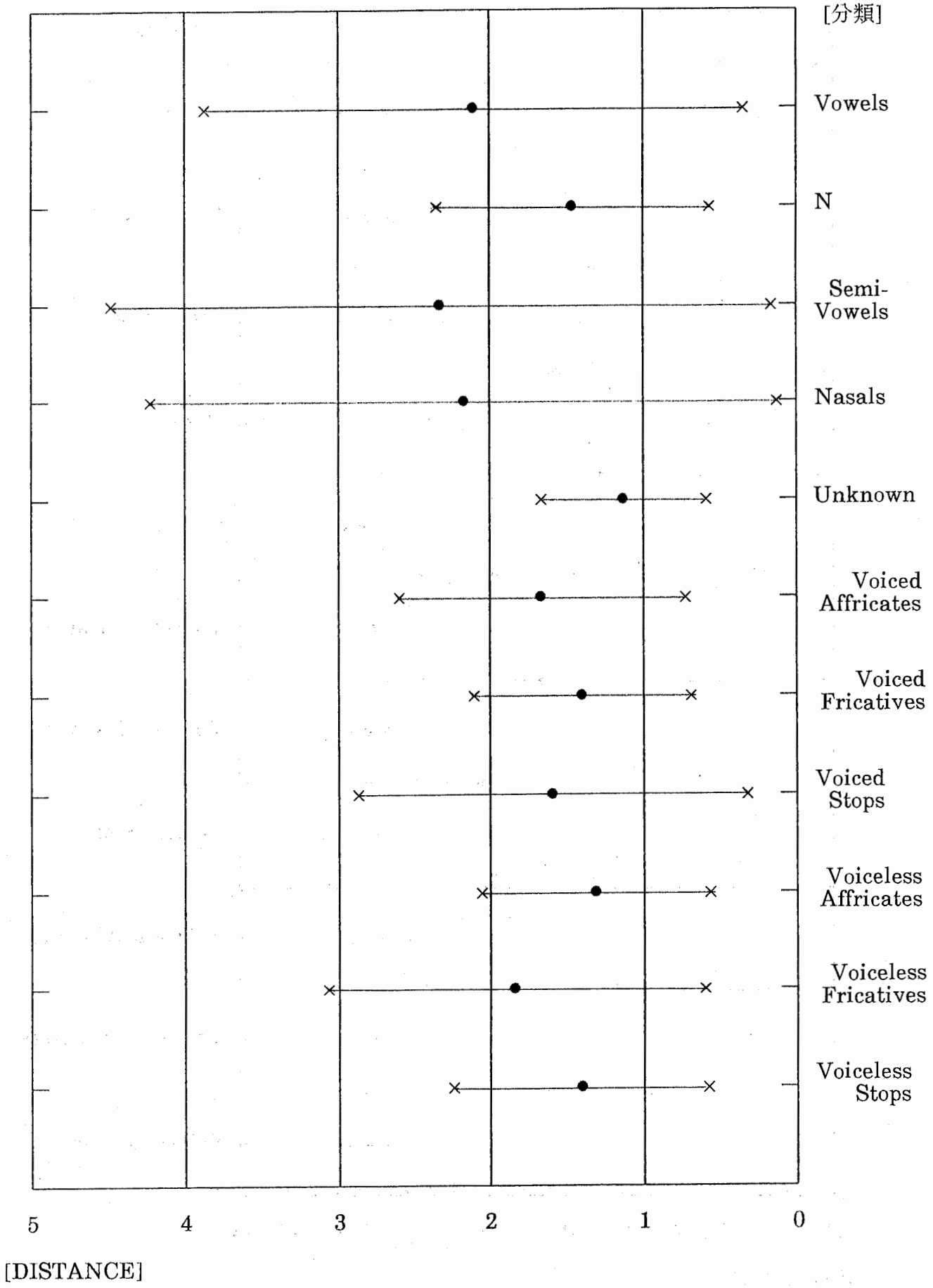


図3 DISTANCEの平均と分散
(Changed)

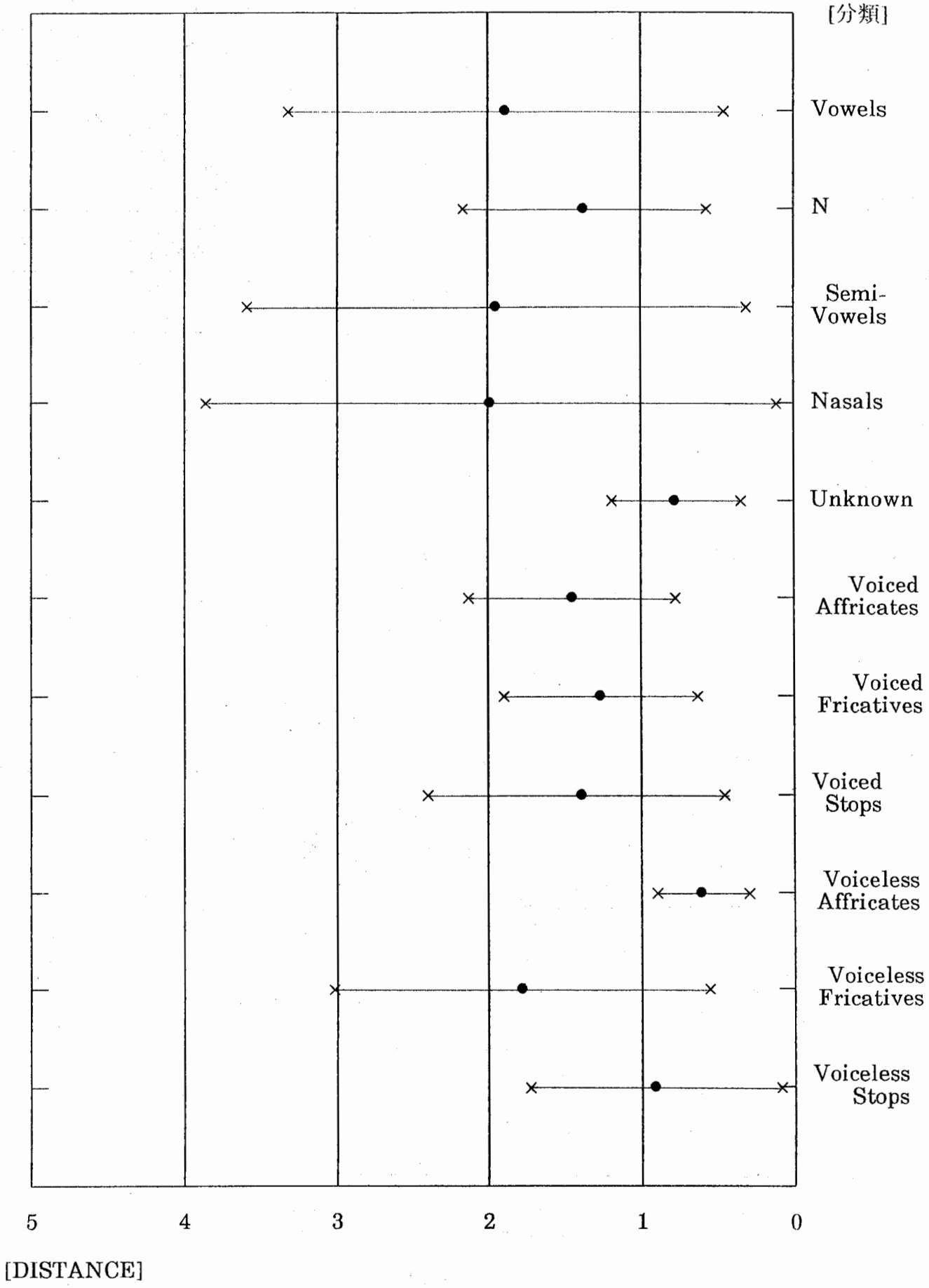


図4 SameFrontのみ [1,2,7]の分類

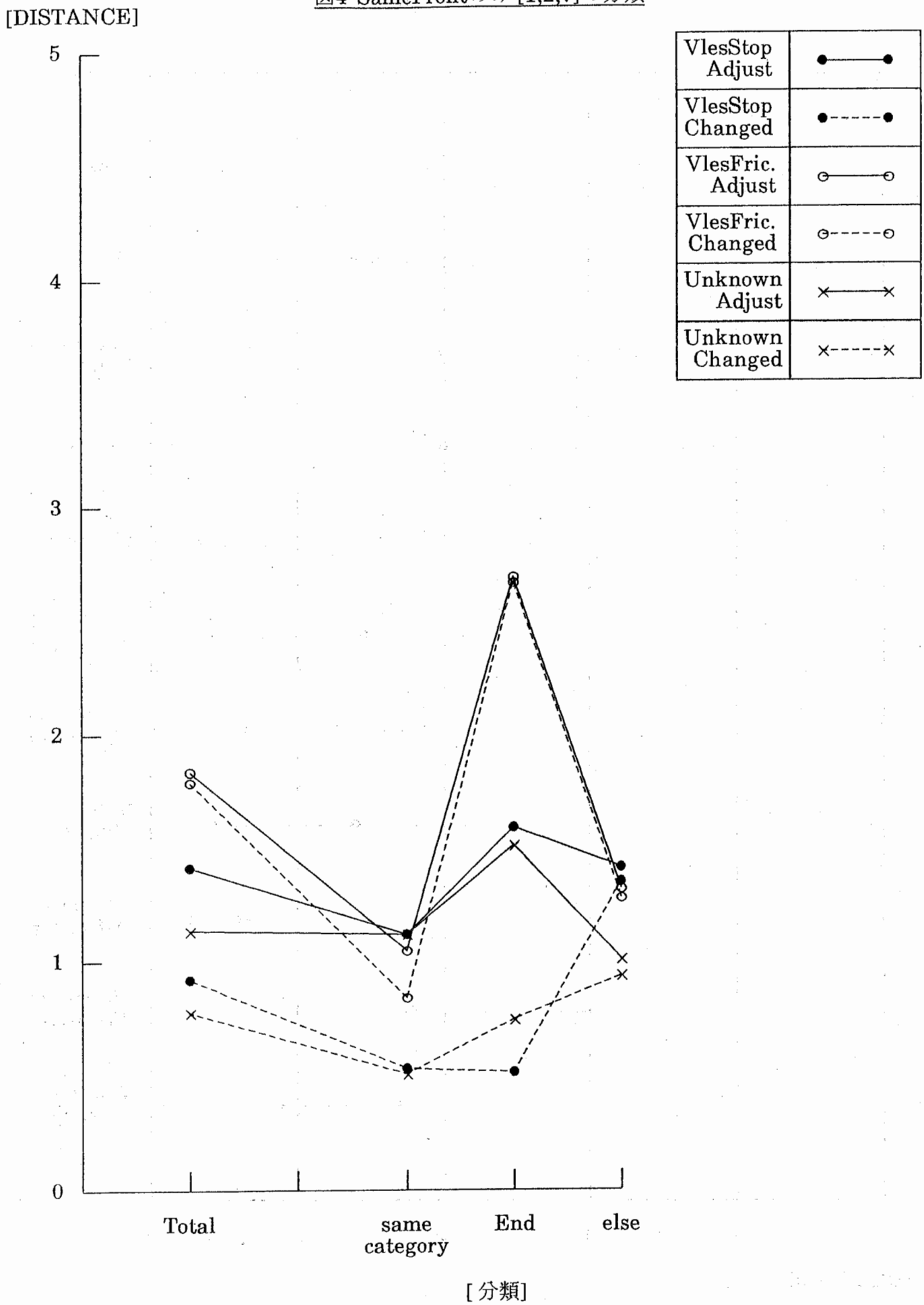


図5 SameFrontとSame[4,9]の分類

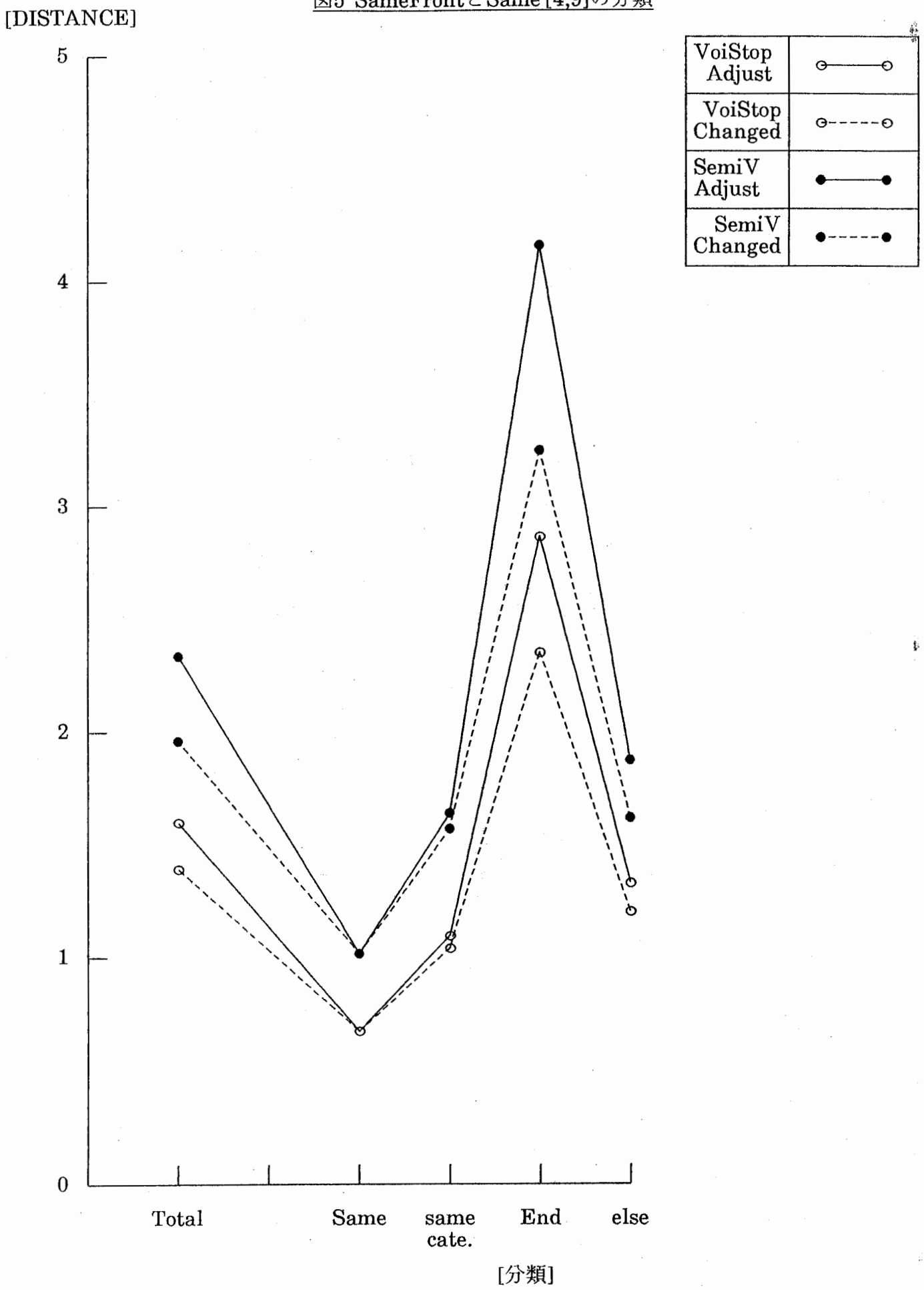


図6 Same,SameFront,SameBack [8,10]の分類

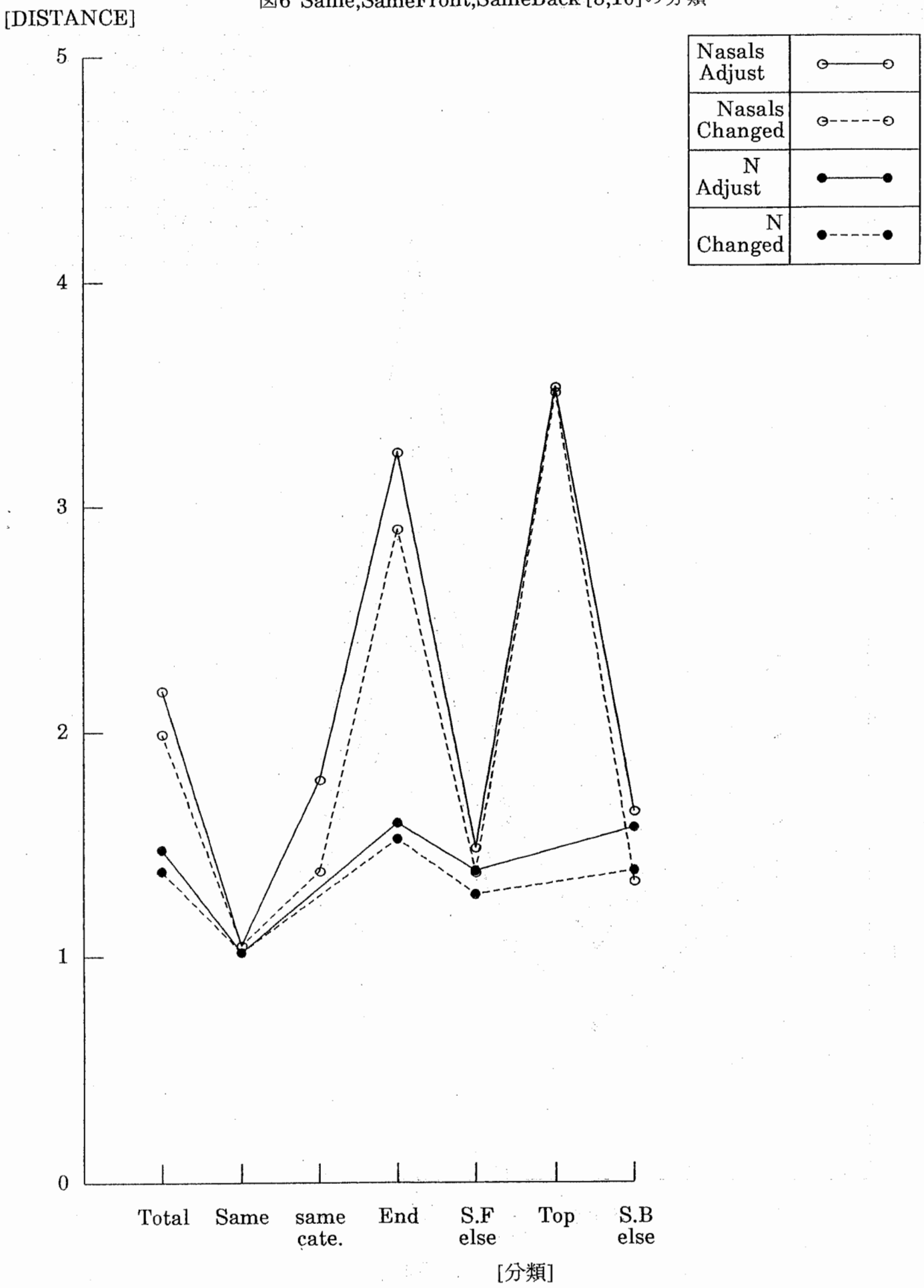


図7 SEGMENT DISTANCEの分類

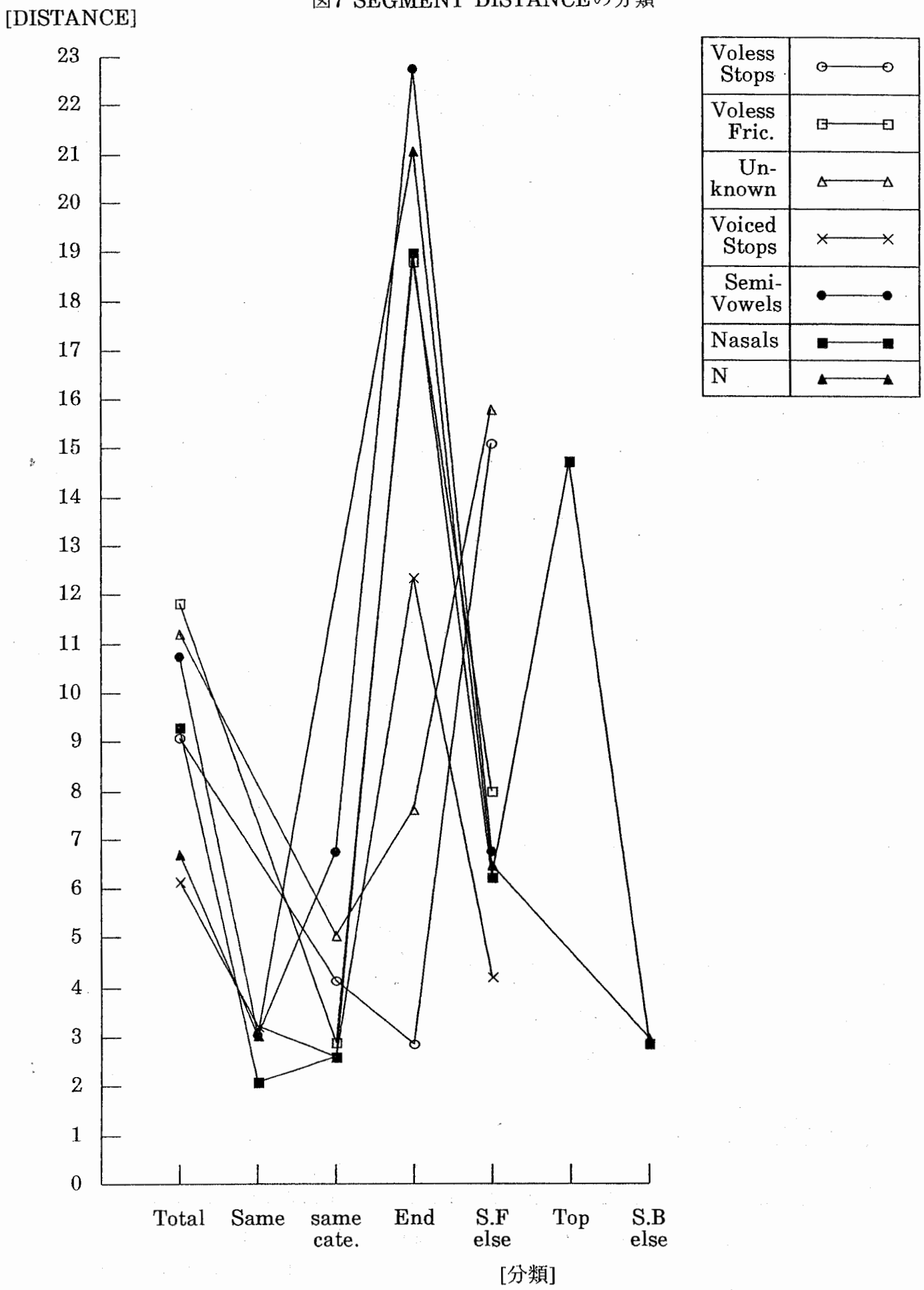


図8 Voiceless Stops 相関図

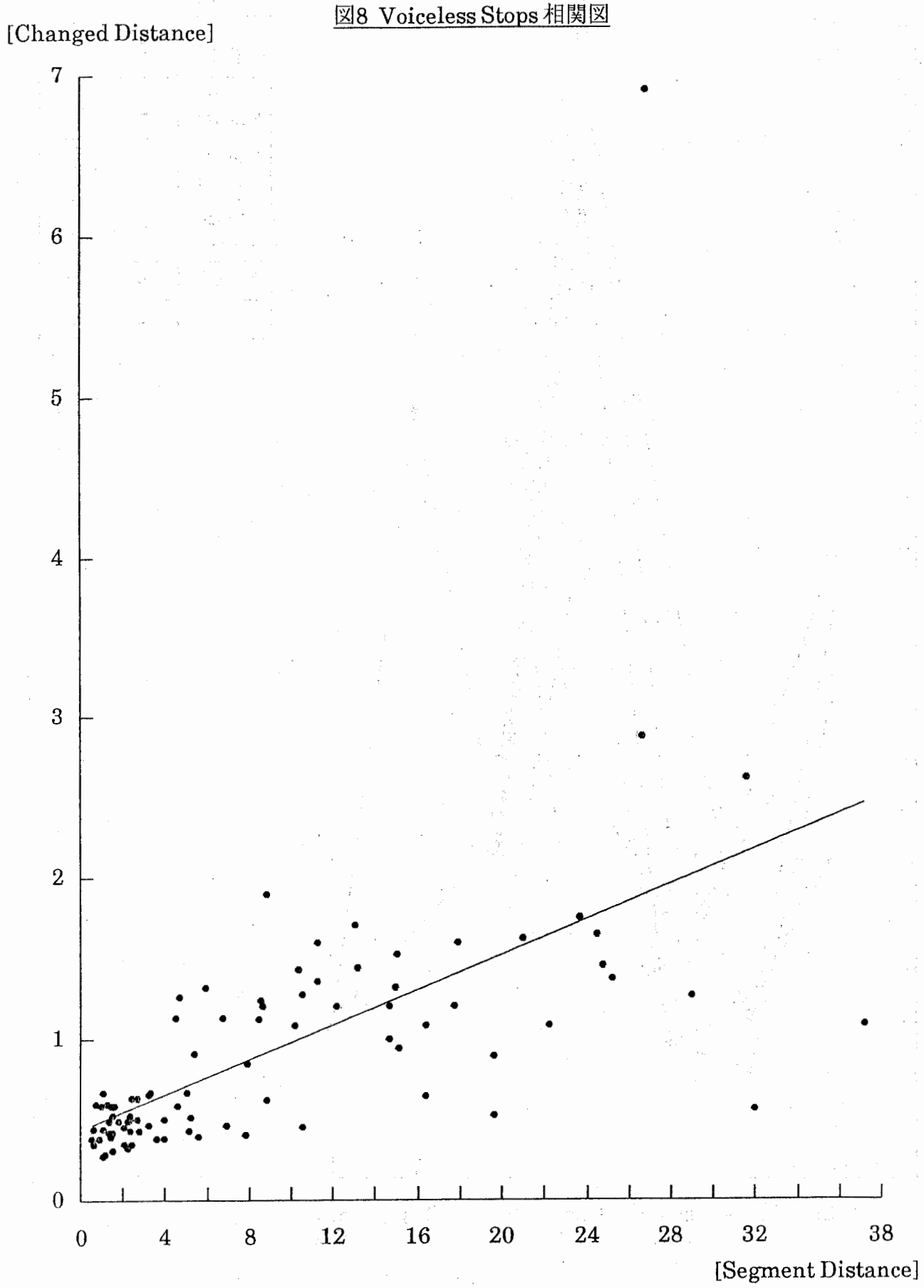
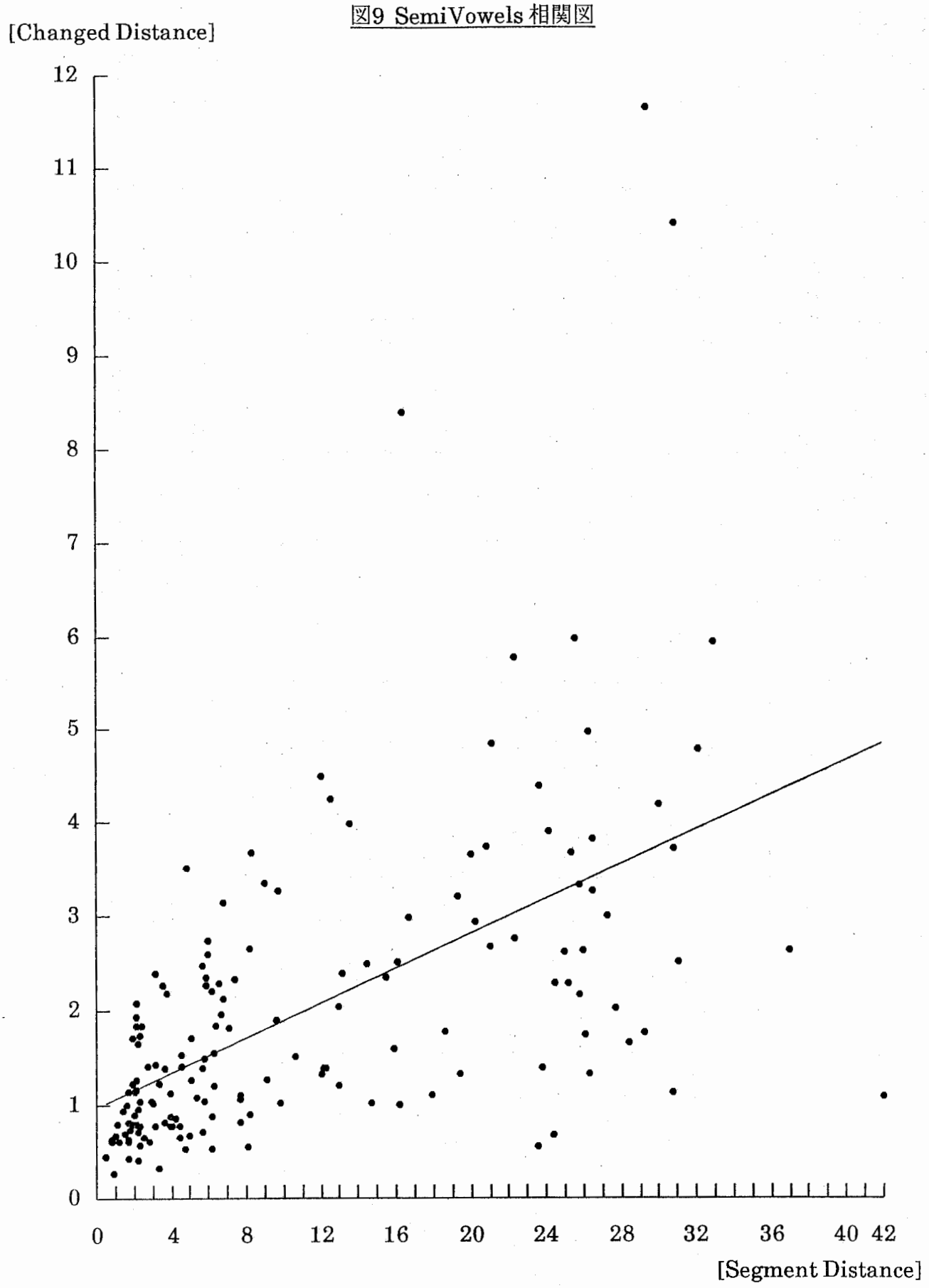


図9 SemiVowels 相関図



NO.	分類	音韻	データ数
1	Voiceless Stops	(p, t, k)	90
2	Voiceless Fricatives	(s, sh, h, f)	82
*3	Voiceless Affricates	(ts, ch)	13
4	Voiced Stops	(b, d)	108
*5	Voiced Fricatives	(z)	12
*6	Voiced Affricates	(j)	15
7	Unknown	(例) tta(促音), kyo(拗音)	71
8	Nasals	(m, n, g)	319
9	SemiVowels	(r, w, y)	168
10	N	(N (撥音))	134
11	Vowels	(a, i, u, e, o)	335
(合計)			1,347

表1.　接続方法の分類

NO.	Adjust	Segment
1	1.42	9.08
2	1.84	11.84
3	1.32	6.45
4	1.60	6.15
5	1.41	8.78
6	1.67	6.73
7	1.14	11.19
8	2.18	9.31
9	2.33	10.73
10	1.47	6.72
11	2.12	10.09
Total	1.92	9.35

表2.　Distanceの平均値

NO.	Changed	Adjust－Changed	change rate[%]
1	0.92	0.50	57.8
2	1.79	0.05	13.4
3	0.61	0.71	69.2
4	1.39	0.21	22.2
5	1.27	0.14	16.7
6	1.46	0.21	53.3
7	0.78	0.36	50.7
8	1.99	0.19	23.2
9	1.96	0.37	25.6
10	1.38	0.09	20.1
11	1.90	0.22	28.5
Total	1.68	0.24	28.7

表3.　ChangedDistanceの平均、差、割合

NO.	Adjust	Changed	Adjust－Changed	Segment
1	1.48	0.63	0.85	5.05
2	1.95	1.61	0.34	6.63
3	1.69	0.66	1.03	2.09
4	2.80	1.86	0.94	7.10
5	2.32	1.48	0.84	2.71
6	2.17	1.78	0.39	6.50
7	1.32	0.61	0.71	5.64
8	2.59	2.56	0.03	10.86
9	3.89	2.45	1.44	12.97
10	2.05	1.61	0.44	5.84
11	2.87	2.12	0.75	7.66
Total	2.39	1.75	0.64	7.66

表4.　変更されたもののみの各平均値

NO.	Total	Only chn.	Same	SameFront			SameBack		Diff.
				same cate.	End	else	Top	else	
1	90	52	---	17	29	44	---	---	---
2	82	11	---	6	32	44	---	---	---
4	108	24	25	11	31	41	---	---	---
7	71	36	---	19	15	37	---	---	---
8	319	74	50	25	83	79	42	40	---
9	168	43	44	18	52	54	---	---	---
10	134	27	38	---	17	54	---	38	---
11	335	101	5	11	10	25	---	71	213

表5.　各分類のデータ数

分類	Voiceless Stops			Voiceless Fricatives			Unknown		
	Adj.	Chn.	Seg.	Adj.	Chn.	Seg.	Adj.	Chn.	Seg.
Total	1.42	0.92	9.08	1.84	1.79	11.84	1.14	0.78	11.19
Only Change	1.48	0.63	5.05	1.95	1.61	6.63	1.32	0.61	5.64
same cate.	1.12	0.53	4.14	1.05	0.84	2.90	1.12	0.51	5.07
End	1.59	0.52	2.86	2.69	2.67	18.80	1.51	0.75	7.60
else	1.42	1.35	15.10	1.32	1.28	7.99	1.01	0.94	15.79

表6. (1)の分類による各平均値

分類	Voiced Stops			SemiVowels		
	Adjust	Changed	Segment	Adjust	Changed	Segment
Total	1.60	1.39	6.15	2.33	1.96	10.73
Only Change	2.80	1.86	7.10	3.89	2.45	12.97
Same	0.68	0.68	3.21	1.02	1.02	3.04
same cate.	1.10	1.04	2.59	1.64	1.57	6.74
End	2.87	2.35	12.35	4.16	3.25	22.74
else	1.33	1.20	4.21	1.87	1.62	6.75

表7. (2)の分類による各平均値

分類	Nasals			N		
	Adjust	Changed	Segment	Adjust	Changed	Segment
Total	2.18	1.99	9.31	1.47	1.38	6.72
Only Change	2.59	2.56	10.86	2.05	1.61	5.84
Same	1.05	1.05	2.09	1.02	1.02	3.04
same cate.	1.78	1.38	2.61	-----	-----	-----
End	3.24	2.90	18.97	1.59	1.52	21.05
S.F else	1.48	1.37	6.22	1.38	1.27	6.48
Top	3.53	3.50	14.71	-----	-----	-----
S.B else	1.64	1.33	2.86	1.57	1.38	2.98

表8.　(3)の分類による各平均値

分類	Vowels		
	Adjust	Changed	Segment
Total	2.12	1.90	10.09
Only Change	2.87	2.12	7.66
Same	1.40	1.40	3.08
Different	2.24	1.96	10.77
SameBack	2.02	1.83	5.24
same cate.	1.40	1.38	3.71
End	2.96	2.96	31.41
S.F else	1.61	1.50	13.74

表9.　(4)の分類による各平均値

分類 No.	具体例	
	DISTANCE Best & Worst 5	接続点
1	0.27~0.34	V(k)とV(t) or V(End)とV(t)
	1.75~6.91	連母音または有声音への渡り
2	0.35~0.47	母音定常点
	4.17~7.33	V(End)
4	0.26~0.42	Same
	3.09~5.08	V(End)
7	0.25~0.35	V(End) or V(t)
	1.46~2.33	V(r) or V(End)
8	0.28~0.35	Sameまたは母音定常点
	7.23~17.01	V(End) or (Top)C [2.50以上ほとんど]
9	0.27~0.44	一つ以外Same
	5.93~11.65	V(End)
10	0.34~0.49	Same or Nasalsへの渡り
	3.20~4.71	V(End) or V(k)
11	0.30~0.44	母音定常点
	7.98~11.25	V(End) or V(t)

表10.　各分類の具体例