

報告番号 乙第 2033 号

会話音声自動認識システムに
関 する 研 究

鹿 野 清 宏

会話音声自動認識システムに
関 する 研 究

昭和56年2月

名古屋大学図書	
和	847042

鹿 野 清 宏

目 次

第1章 序 論	1
1.1 会話音声自動認識の研究	1
1.2 会話音声自動認識の研究動向	3
1.3 会話音声自動認識研究の問題点および本論文の構成	12
第2章 会話音声認識システム (Voice Q-A System) の基本構成	17
2.1 は し が き	17
2.2 システムの基本構成	17
2.3 認 識 対 象	18
2.3.1 質問回答の例	19
2.3.2 対象の複雑さの評価	21
2.4 あ と が き	24
第3章 Voice Q-A System I の構成と評価	25
3.1 は し が き	25
3.2 Voice Q-A System I の音響処理	25
3.3 言語処理の基本構成	27
3.4 単語認識部	28
3.4.1 音韻変形規則	28
3.4.2 単語辞書	29
3.4.3 単語認識アルゴリズム	30
3.4.4 動作例	33
3.5 構文解析部	37
3.5.1 プラグマティクス	37
3.5.2 構文解析	38
3.5.3 動作例	41
3.6 推 論 部	41

3.7	会話モデル	42
3.8	オンライン質問回答システムの構成	46
3.8.1	音響処理の高速化	46
3.8.2	言語処理の高速化	47
3.8.3	音声応答部	47
3.8.4	処理時間	49
3.9	認識実験による性能評価	50
3.9.1	音声資料	50
3.9.2	認識実験の結果	50
3.9.3	発声の制約による効果	53
3.10	質問回答実験による性能評価	56
3.10.1	音声資料	56
3.10.2	質問回答実験の結果	56
3.11	Voice Q-A System I の問題点	58
3.12	あとがき	59
第4章	音響処理の評価方法	61
4.1	はしがき	61
4.2	音韻ラティスの音韻認識率の推定モデル	63
4.2.1	考え方と問題点	63
4.2.2	モデル	68
4.3	音韻ラティスの情報量	69
4.3.1	音韻系列の情報量と認識率の関係	69
4.3.2	音韻ラティスの情報量	73
4.4	音韻認識率の自動推定アルゴリズム	76
4.4.1	マクロな統計量の自動抽出	76
4.4.2	ラティスの相互情報量の推定	82
4.4.3	音韻認識率の推定	82
4.4.4	音韻認識率の自動推定アルゴリズムのまとめ	83
4.5	アルゴリズムの評価	84

4.6	自動推定アルゴリズムの適用例	87
4.6.1	音響処理部の音韻認識率の推定	87
4.6.2	VCV音節認識におけるいき値の最適決定への適用	89
4.6.3	VCV音節認識における音韻の候補の数の制限の検討への適用	90
4.7	あとがき	90
第5章	LPCスペクトル・マッチング尺度の検討	93
5.1	はしがき	93
5.2	LPCスペクトル・マッチング尺度(LPC Distance Measures)	95
5.2.1	最尤スペクトル距離(Likelihood Ratio Measure)	95
5.2.2	正規化残差(Normalized Residual)	98
5.2.3	COSH尺度(COSH Measure)	99
5.2.4	LPCケプストラム距離(LPC Cepstrum Distance)	99
5.2.5	各尺度のマッチングの条件	101
5.2.6	マッチングの様子の観察	102
5.3	LPCスペクトル・マッチング尺度の評価	106
5.3.1	母音識別実験による評価	106
5.3.2	標準パターンについての検討	110
5.3.3	まとめ	112
5.4	Peak Weighted Measureの導出	112
5.4.1	LPCケプストラム距離からの導出	113
5.4.2	COSH尺度からの導出	115
5.4.3	最尤スペクトル距離からの導出	117
5.4.4	LPC相関距離	118
5.5	Peak Weighted Measureの評価	121
5.5.1	ピーク重みづけの解析	121
5.5.2	母音識別実験による評価	123
5.6	Peak Weighted Measureの母音認識における最適化	127
5.6.1	周波数軸における重みづけの効果	128
5.6.2	スペクトル平滑化による効果	131

5.6.3	ま と め	135
5.7	LPC スペクトル・マッチング尺度による大語い単語音声認識	136
5.7.1	LPC スペクトル・マッチング尺度	136
5.7.2	時間軸正規化マッチング	137
5.7.3	音 声 資 料	138
5.7.4	認識実験による尺度の評価	139
5.7.5	分析フレーム周期の影響	141
5.7.6	大語い単語音声認識実験	142
5.8	あ と が き	147
第6章	Voice Q-A System II の構成と評価	148
6.1	は し が き	148
6.2	言語処理の改良	149
6.3	COSH 尺度を用いた音響処理 (Voice Q-A System II-1)	149
6.3.1	前 処 理	150
6.3.2	標準パターン	150
6.3.3	基本特徴量の計算	151
6.3.4	基本構成	153
6.3.5	セグメンテーション	155
6.3.6	母音の識別	164
6.3.7	子音の識別	164
6.3.8	Voice Q-A System I の音響処理との比較	166
6.4	Voice Q-A System II-1 の性能評価	166
6.4.1	性能評価	167
6.4.2	標準パターンの効果	170
6.4.3	誤り・拒絶の原因	173
6.5	Peak Weighted Measure を用いた音響処理 (Voice Q-A System II-2)	174
6.5.1	Voice Q-A System II-2 の作成	174
6.5.2	動作例	176

6.6	Voice Q-A System II-2の性能評価	176
6.6.1	性能評価	176
6.6.2	標準パターンの効果	183
6.7	あとがき	186
第7章	結 言	188
謝	辞	192
文	献	193
付	録	201
1.	Voice Q-A System Iの言語情報	203
2.	Voice Q-A System Iの構文解析部の動作例	211
3.	Voice Q-A Systemの発声リスト	216
4.	FORTTRANによるリスト処理システムDLOP	220
5.	641都市名の発声リスト	229
6.	大語い単語音声認識実験における誤り	231
7.	Voice Q-A System IIの音韻変形規則	238
8.	Voice Q-A System II-1における発声者別の認識実験結果	240
9.	音声データのLPCスペクトル包絡の系列およびセグメンテーション に用いる基本特徴量の変化の図示	245
10.	Voice Q-A System IIの構文解析部の動作例	273
11.	Voice Q-A System II-2における発声者別の認識実験結果	286

第1章 序 論

1.1 会話音声自動認識の研究

音声は人間にとって意志の疎通に欠くことのできない手段である。人間は、自分の考えていることを音声で他人に伝えることが非常に多い。とくに、日常生活では、ほとんどすべての場合、自分の意志を音声で他人に伝えている。

人間は、頭の中で考えた事柄や感じた事を、単語の列に変換し、さらに、発声器官を駆動して声に出して、他人に伝える。聴き手は、無意識のうちに、会話状況、構文、単語などの言語情報を用いて、発声者の意図をは握している。このような音声は、人間の情報活動の中で、不可欠な媒体であり、最も重要なものである。しかしながら、人間の発声や聴きとり（認識）の過程は、ほとんど明らかにされていない。器官としての口や耳の構造や機能については、かなり明らかにされているが、頭の中での処理過程については、ほとんど何もわかっていない。

ここで、人間の発声と聴きとりの過程を単純化して、図1-1のように表わしてみよう。会話の対象や話しの流れから話したい意図が決まり、その意図から発声内容を決め、構文単語情報を用いて単語列が決まる。単語列には、意味内容や単語情報による韻律情報（ピッチ、ポーズなど）が付加され、音素列に変換される。この過程では、感情や周囲の環境などによっても韻律情報が付加される。さらに、方言の影響も付け加えられる。音素列は神経パルスに変換され、発声器官を動作させ、音声が出力される。このようにして発声された音声は、調音結合、発声器官の差、発声差などの様々な変形を受けている。音声波は空中を伝搬し、相手の耳に入る。この際、周囲雑音が音声波に加わる。聞き手は、発声者特有の歪や周囲雑音を含んだ音声から、言語情報を利用して、音声の冗長性によって歪や雑音を除去し、話し手の意図を聴きとっていると考えられる。聴きとりの過程は、発声の過程以上に複雑で、かつ、不明確である。

音声に含まれる情報としては、上で述べた(1)音韻情報（話し手の意図）の他にも、(2)個人性情報（話し手はだれか）、(3)情緒情報（話し手の感情）が考えられる。広い意味の音声認識には、(2)、(3)の情報の認識も含まれる。狭い意味での音声認識は、(1)の音韻情報だけを含む。本論文は、(1)の音韻情報に関するものであり、とくに、電子計算機と人間との

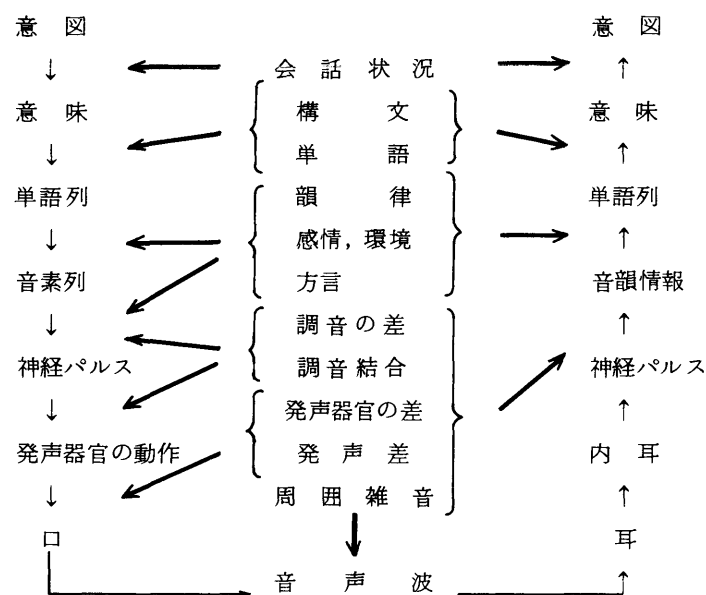


図 1 - 1 音声の発声・認識過程の概念図

コミュニケーションの手段としての音声認識の研究について述べたものである。

電子計算機による音声認識の困難さの要因として、次の事柄があげられる。

- (1) 認識対象の設定
- (2) 音韻の境界が不明瞭
- (3) 単語間の境界が不明瞭
- (4) 調音結合
- (5) 個人差
- (6) 発声ごとのばらつき
- (7) 経時変化

音声認識で、現時点までに実用化の段階に到達しているのは、発声者ごとに調整した単語数 100 以下の単語音声認識装置である。この装置では、単語数を 100 個以下に制限し、かつ、単語ごとに区切って発声させる制限を加えることによって、(1)、(3)の問題をさけ、単語全体でマッチングを行うことによって(2)、(4)の問題をさけている。さらに、発声者ごとの標準パターンを登録することによって(5)、(7)の問題をさけている。(6)の発声ごとのばらつきの正規化は、動的計画法で代表される時間軸の伸縮アルゴリズムである程度まで解決している。しかしながら、会話音声(連続単語音声)を認識しようとする場合には、単語間の境界が不明確になるために、(2)、(3)のセグメンテーションの問題に正面から取り組ま

ざるを得なくなる。また、単語を連続して発声するため、(1)の制限もゆるくならざるを得ない。よって、単語認識で用いられた手法は、ほとんど適用できなくなり、会話音声認識のための独自の手法を開発せねばならない。とくに、構文、単語辞書、韻律情報などの言語情報を最大限に利用して、(2)、(3)のセグメンテーションの問題の困難さを軽減して、認識システムを作成することが重要と考えられる。本論文は、このような観点からの会話音声認識システムの研究に関するものである。

音声認識の最終的な目標は、自由発声された会話音声を、話し手にかかわらず、認識できることである。しかし、最終的な目標が達成できなくても、一定の制約を加えることによって、認識システムが実現でき、音声の利点を生かすことのできる様々な用途で利用される。現在でも、単語ごとに区切って特定の発声者が発声するという制約を加えた単語認識装置でさえ、荷物の自動仕分けなどの仕事で利用されている。単語ごとに区切るという制約がとり除かれ、会話音声の認識ができるようになれば、機械と音声で比較的自由に対話できるようになり、対象の制限があっても、音声認識装置の適用分野は飛躍的に広がると考えられる。その分野も、予約サービス、銀行サービス、情報検索、ワードプロセッサと広がってゆくと予想される。このように、音声認識は、その研究の一部は実用化に移行し、さらに、不特定話者向きの認識の研究、会話音声認識の研究などが精力的に開始されており、認識技術も急速に蓄えられつつある状況にある。

1.2 会話音声自動認識の研究動向

この節では、文献(1)~(10)も参考にして、音声認識の歴史および研究動向について、会話音声認識に重点をおいて、概説する。

現在、音声認識は、音声情報処理の最も重要な分野の1つになっている。現在の音声情報処理の源の1つとして、1940年代のDudleyのボコーダ(vocoder)⁽¹¹⁾やPotter等のソナグラフ(visible speech)⁽¹²⁾をあげることができる。ボコーダの流れは、音声の分析の強力な手法である線型予測(LPC)分析⁽¹³⁾につながっていると考えられるし、ソナグラフは、その後改良され、現在でも、なお、音声の観察の重要な道具として用いられている。⁽¹⁴⁾ 最初の本格的な音声認識の研究は、1952年のDavis等による数字音声認識装置⁽¹⁵⁾に関する研究である。そこで用いられた手法は、第1ホルマントと第2ホルマ

ントの軌跡によって数字音声を認識しようとする試みであった。このように単語全体を一つのパターンとみなす考え方は、後の単語音声認識装置に受けつがれている。また、分析手法としても、帯域通過フィルタによる特徴抽出の先駆となっている。1960年代に入ると、Fant⁽¹⁶⁾、Steve⁽¹⁷⁾、Flanagan⁽¹⁸⁾などにより、音声情報処理の基礎が固められた。とくに、「Analysis by Synthesis」の考え方は、この後も広く受けつがれているし、会話音声認識における「Hypothesis and Test」の考えにも影響を与えている。この頃、音声タイプライター（phonetic typewriter）を目ざした研究がOlson⁽¹⁹⁾や坂井、堂下⁽²⁰⁾によって行われた。これらの結果は、音声現象の複雑さを研究者に再認識させ、音声認識の研究の困難さを思い知らせた。とくに、これらの研究結果は、言語情報を用いない会話音声認識へのアプローチは不可能に近いことを示した。その後、Reddy⁽²¹⁾やTappert⁽²²⁾、Alter⁽²³⁾などによって言語情報を利用した音声認識の研究が行われた。これらの研究は、現在の会話音声認識システムや音声理解システムの基礎となっている。この頃、アメリカでは、Pierceによる音声認識の研究への痛裂な批判⁽²⁴⁾が行われた。この批判の主旨は、「一般的な目的の連続音声認識は技術的に不可能に近く、また、特殊な目的の単語音声認識などは実用的でなく、かつ経済的でない」ということであった。批判の前半は現在の技術レベルをもってしても正しいと思われる。現在でも、対象を限定しない連続音声認識は、技術的に不可能に近いと考えられる。よって、対象を限定した音声理解システムの研究が行われている。しかし、批判の後半は、まったく的はずれである。現在では、単語音声認識装置は実用化され、様々な分野で使われ始めている。また、経済性の面でも、LSI技術の発展によって、充分廉価になり、実用化が可能になりつつある。Pierceの批判と前後して、日本で音声情報処理に関する2つの大きな進展があった。まず、板倉、斉藤によって、最尤スペクトル分析⁽¹³⁾、LPC分析⁽²⁵⁾等の音声分析の画期的な分析法の発見が行われた。このLPC分析の手法は、音声情報処理の研究を飛躍的に発展させた。第2に、音声認識での発声のばらつきを正規化する時間軸の伸縮法として、動的計画法（Dynamic Programming）の利用が行われ始めたことである。⁽²⁶⁾⁽²⁷⁾この手法は、発声者を限定した場合の単語音声認識の研究を実用化に近づけた。アメリカでは、自然言語処理の分野で、大きな進歩が見られた。WoodsはATN文法（Augmented Transition Network Grammar）を用いて自然言語による質問回答システムを作成した。⁽²⁸⁾また、Winogradは手続き（Procedure）という概念を押し進め、高度な自然言語による質問回答システムを作成した。⁽²⁹⁾これらの自然言語処理の研究は、対象を限定すれば、自然言

語による質問回答システムの作成が可能であることを実証した。自然言語処理の研究の発展とLPC分析の研究の伸展は、1971年のアメリカでのARPA(Advanced Research Projects Agency)の援助による音声理解システム(Speech Understanding Systems)のプロジェクトの開始につながった。⁽⁸⁾

アメリカでのARPAのプロジェクトは、人工知能や自然言語グループの先導によって開始された。⁽⁸⁾ このプロジェクトでは、従来からのSpeech Recognition Systemsという言葉にかわって、Speech Understanding Systemsという言葉を用いた。これは、音声認識で重要なことは、一語一語あるいは一音一音正しく認識することではなく、意味内容を把握することであるという考えに基づいている。このアプローチでは、言語情報の利用、つまり、構文情報、意味情報、プラグマティクス、単語辞書情報、韻律情報などの利用に重点が置かれた。ARPAのプロジェクトの開始から2～3年後に、日本でも同様のアプローチが、武蔵野通研⁽³⁰⁾、⁽³²⁾ 京都大学⁽³³⁾、京都工芸繊維大学⁽³⁴⁾、山梨大学⁽³⁵⁾ で開始された。その他、アメリカのIBMにおいても、会話音声認識の研究が地道に続けられている⁽³⁶⁾。ヨーロッパにおいても、同様な研究が1976年ごろから、フランス⁽³⁷⁾、イタリア⁽³⁸⁾などで開始されている。また、アメリカのベル研でも、1979年から開始されている⁽³⁹⁾。

これらの研究機関での会話音声認識システムの研究の概要を表1-1にまとめて示しておく。以下、これらの研究機関での会話音声認識システムの研究動向について簡単に述べる。

ARPAのプロジェクトによる音声理解システムの目標を表1-2に示す。このプロジェクトは、成果を1976年度末に表1-2の仕様を満足するデモンストレーション・システムを作成することによって示すという厳しい目標をもって開始された。ARPAのプロジェクトは、最初、アメリカの主な音声の研究機関を含んで開始された。途中で、プロジェクトは、CMU(Carnegie Mellon University)、BBN(Bolt Beranek and Newman Inc.)、SRI(Stanford Research Institute)とSDC(System Development Corporation)の3グループにしばられた。1976年の末に目標に近い性能を示したのは、CMUのHarpyシステムのみであった。このシステムは、比較的単純なnetwork modelを利用したtop-down的構成のシステムである。また、Harpyシステムを主として担当したのはLowerre⁽⁴²⁾1人であった。高度な言語情報を利用するというARPAの最初の目標は、十分に目的を達成しないでプロジェクトは終了した。この主たる原因としては、音響処理レベルの研究が余りにも弱かったことがあげられる。⁽⁹⁾⁽¹⁰⁾ ARPA

表 1-1 諸研究機関での会話音声認識の研究

研究機関	CMU ⁽⁴⁰⁾⁽⁴⁴⁾	CMU ⁽⁴¹⁾⁽⁴⁴⁾	CMU ⁽⁴²⁾⁽⁴⁴⁾	CMU ⁽⁴³⁾⁽⁴⁴⁾	BBN ⁽⁴⁵⁾	BBN ⁽⁴⁶⁾	
システム名	Hearsay I	Dragon	Harpy	Hearsay II	SPEECHLIS	HWIM	
対象語数	チェス 24~194	チェス 24~194	AIに関する情報検索 1011	AIに関する情報検索 1011	月の岩石に関する情報検索 250	旅行の予算管理 1097 (409)	
発声環境	端末室	端末室	端末室	端末室	無音室	端末室	
平均分岐数	—	—	33.4	33.4	—	196 (67)	
文章例	Pawn to queen four.	Castle on king side.	Does semantic nets get mentioned anywhere ?	Are any papers about semantic networks ?	Have any people done chemical analysis on this rock ?	When did Bill go to Mexico ?	
システム構成	音響処理	Lab. → Seg.	frame Lab.	Seg. → Lab.	Seg. → Lab.	Seg. → Lab.	
	言語処理	top-down(hypothesis and test)	network model	network model	blackboard model	(あいまいさあり) bottom-up	(Dip検出を主体) bottom-up
	音響処理結果	音韻系列	フレームごとの音韻のスコア	音韻系列とスコア	音韻とスコア	セグメント・ラティス	セグメント・ラティス
音響処理	音響分析	5・BPF, ZC	5・BPF, ZC	BPF, LPC, Itakura 尺度	BPF, LPC, Itakura 尺度	SLPC, LPC, 2次のLPC他	SLPC, LPC, ホルマント, ZC, ピッチ, Itakura 尺度
	話者適応化	—	—	音韻標準パターンの学習	—	—	声道長の補正
言語処理	単語認識	記号のマッチング	—	—	—	—	—
	構文解析	top-down (BN記法)	network model (全数探索)	beam search	bottom-up+top-down island driven (BN記法)	bottom-up(word lattice), island driven (ATNパーザー)	bottom-up+top-down (スペクトル合成による検証) ATNパーザー
性能	会話モデル	—	—	—	—	—	—
	発声者数	4名	1名	5名 (未知)	1名 (未知)	(HWIMに移行)	3名 (3名)
	文章総数	102文章	102文章	184文章 (5文章)	22文章 (5文章)		196文章 (67文章)
	意味理解率	31%	49%	95% (80%)	91% (60%)		44% (52%)
処理時間	実時間の9~44倍	実時間の48~174倍	27.9MIPSS	85.0MIPSS		実時間の1,350倍	
稼動年月	1973	1975	1976	1976	1975	1976	
備考	SUSの源。階層モデル。	Harpyの第1次システム。処理速度が遅い。	ARPAのSUSでの成功システム。Dragonの高速化。LPC利用。	blackboard model island driven parser	bottom-up system segment lattice word lattice ATNパーザー	比較的精密な音響処理。SPEECHLISの改良。合成による単語の認識。	

Lab... phoneme labeling, Seg... phoneme segmentation, BPF... bandpass filter, ZC... zero crossing
MIPSS... Million Instructions Per Second Speech

研究機関	MIT ⁽⁴⁷⁾	SRI ⁽⁴⁸⁾	SDC ⁽⁴⁹⁾⁽⁵⁰⁾	IBM ⁽³⁶⁾	IBM ⁽⁵¹⁾	BTL ⁽³⁹⁾	
システム名	CASPERS	—	VCDMS	—	CSAP200	—	
対象語数	—	水道栓の故障修理 54	戦艦の情報検索 約1,000	New Raleigh 言語 250	Laser Patent task 1,000	列車の座席予約 (ECL) 112	
発声環境	端末室	無音室	—	無音室	無音室	計算機室	
平均分岐数	9.15	—	105	7.32	(Perplexity 24.1)	2.00	
文章例	Set the frequency range to 100Hz through 1000Hz.	Put one washer in the faucet.	How fast is the Theodore Roosevelt?	Some workers make the radio in use.	The maser action that result has been explained above.	1日の新大阪から博多までの、6時22分の、ひかり19号の、グリーン券を、9枚、お願いします。	
システム構成	音響処理 言語処理 音響処理結果	Seg. → Lab. (ディップ重視) top-down 音韻系列	word function best-first 音響パラメータのデータベース	Seg. → Lab. (スペクトルの定常性) best-first A マトリックス	Seg. → Lab. (定常部) network model 音韻系列とスコア	Lab. → Seg. (200種類の音韻標準パターン) 音韻系列とスコア	ECLのVoice Q-A System II network model 音韻ラティスとスコア
音響処理	音響分析 話者適応化	LPC分析, ホルマント, ピッチ —	BPF, LPC分析 —	ZC, ホルマント, ピッチ, LPC分析 —	FFT —	20kHz サンプリング, FFT, 100Hzごとの band, —	ECLのVoice Q-A System II
言語処理	単語認識 構文解析 会話モデル	グラフ表現 best-first (BN記法) —	word function best-first parser (BN記法) —	グラフで記述 SRI を利用 —	PFSA, 動的計画法 FSA, best-first —	PFSA, 動的計画法 FSA, best-first —	network model 動的計画法 —
性能	発声者数 文章総数 意味理解率 処理時間	6名 275文章 49% 実時間の15~20倍	3名 71文章 62% —	— — 24% —	1名 363文章 81% 実時間の300倍	1名 20文章 (単語認識率91.3%) 実時間の300倍	4名 80文章 (496文節) 79.4% 実時間の1,000倍
稼動年月	1974	1974	1976	1976	1979	1980	
備考	よくまとまったシステム。 LPC分析。 全単語を一つのグラフで表現。 階層モデルとネットワークモデルの中間。	best-first parser word function	SRIと協同研究。	network model。 FFT。 精密な音響処理。		ECLと共同研究。 ECLのnetwork mode I版。 処理速度に難。	

研究機関	Instituto di (38) Electrotecnica(イタリア)	CNET(仏) (37)
システム名	—	KEAL
対象 語い数 発声環境 平均分岐数	交通機関の予約及び案内 約200 — —	電話番号案内 — — —
文章例	—	Je voudrais le numero de telephone de MARCEL DUPONT
システム構成	音響処理 言語処理 音響処理結果	Seg. → Lab. (CV単位) top-down 音韻系列
音響処理	音響分析 話者適応化	FFT, LPC分析, ホルマ ント, ピッチ —
言語処理	単語認識 構文解析 会話モデル	14BPF — word spotting(DP) top-down(BN記法) あり
性能	発声者数 文章総数 意味理解率 処理時間	3名 — 77% —
稼動年月	1975	1977
備考	システム全体としては 未動作?	精力的なシステム作り。 会話モデルあり。 応答部あり。

研究機関	京都大学 ⁽³³⁾	京都工芸繊維大学 ⁽³⁴⁾	山梨大学 ⁽³⁵⁾	ECL ⁽⁵²⁾⁽⁵³⁾	ECL ^{(30)~(32)}	ECL ⁽⁵⁴⁾
システム名	LITHAN	SPOKEN-BASIC	-	Voice Q-A System O	Voice Q-A System I	Voice Q-A System II
対象語い数	101	50	99	65	112	同左
発声環境	無音室	無音室	無音室	無音室	計算機室	
平均分岐数	5.0	15.0	-	-	20.0	
文章例	計算機言語の磁気ドラム装置は、どれがいているか。	100 DIM Y(10) CR	英語の得点の分散を求める。	東京から、新大阪まで、8時30分発の、ひかり59号を、お願いします。	1日の、新大阪から、博多までの、6時22分の、ひかり19号の、グリーンを、9枚、予約します。	同左
システム構成	音響処理	Lab. → Seg.	Lab. → Seg.	Seg. → Lab.	Seg. → Lab. (←)	Seg. → Lab. (帰納的構成)
	言語処理	top-down	top-down	top-down	top-down	top-down
音響処理結果	音韻系列	音韻系列	簡略化音素列	音韻ラティス	音韻ラティス、継続時間	同左
音響処理	音響分析	20 BPF	20 BPF	ZC, LPC分析, kパラメータ ホルマント	相関分析(正規化残差) VCV音節単位, DP	LPC分析(正規化残差, Itakura 尺度)
	話者適応化	母音と有声子音の自動修正	-	-	-	母音の手動変更
言語処理	単語認識	記号レベルDPマッチング	tree search	DPマッチング	tree search	高速 tree search
	構文解析	best-first(BN記法)	best-first(BN記法)	FSAで記述	リスト表現(ATN等価)	リスト表現
	会話モデル	-	-	-	あり	あり
性能	発声者数	10名	4名	4名	1名	8名
	文章総数	200文章	242文章	36文章	42文章(215文節)	320文章(1,984文節)
	意味理解率	64%	85.5%	77.8%	87%	86%
処理時間	-	-	-	-	実時間の5倍	実時間の10倍
稼動年月	1975	1975	1977	1975	1976	1980
備考	DPマッチングを用いている。 まとまったシステムである。			VCV単位 相関分析 音韻ラティス 構文のリスト表現 会話モデル	オンライン質問回答システム。 各種高速アルゴリズム。 LPC分析。	帰納的構成の音響処理。 COSH尺度, Peak weighted 尺度による精密化。

表 1-2 ARPA の音声理解システムの目標⁽⁸⁾

音声の種類	連続音声
発声者の数	多数
発声者の種類	協力的標準のアメリカ英語
発声環境	静かな部屋
変換器	高品質マイクロホン
発声者への適応	若干の調整
発声者の訓練	自然な適応
語い数	若干選択された 1,000 単語
言語の種類	人工的文法
タスクの種類	データマネジメント, 計算機状態の問い合わせなど
発声者の心理モデル	単純な心理モデル
対話の複雑さ	自然なやりとり
許容できる誤り	10%以下の意味の誤り
処理時間	実時間の数倍
稼動年月	1976 年

のシステムでは、意識的に異なったアプローチがとられた。CMUのHearsay II⁽⁴³⁾では、並列処理的なアプローチがとられ、認識システム構成での新しい手法、blackboard model を提案している。Hearsay II の性能は、Harpy に近いものであり、手法的には大きな成果を収めたといえる。BBNではbottom-up的なアプローチがとられた。⁽⁴⁶⁾ ARPAのプロジェクトの中では、BBNの音響処理は、もっとも精密なものであるが、bottom-up 的なシステム構成法による限界のために、システムとしては十分に動作しなかった。SRIとSDCのグループ⁽⁴⁹⁾⁽⁵⁰⁾では、SRIが言語処理を、SDCが音響処理を受けもって研究を行った。SRIの言語処理は、top-down 的構成で、best-first 法を基本として作成された。SDCでは比較的精密な音響処理が開発された。しかしながら、SDCの音響処理とSRIの言語処理とのインタフェースの不一致のためか、その性能は不十分であった。

日本では、我々電々公社武蔵野通信研究所をはじめ、京都大学、京都工芸繊維大学、山梨大学で、会話音声認識の研究が行われた。認識対象は、ARPAのプロジェクトよりも若干狭いけれども、より実際的なシステムの研究が行われ、各システムとも独自の成果をあげた。京都工芸繊維大学のSPOKEN-BASIC⁽³⁴⁾では、BASICのプログラムを対象に

とりあげて研究が行われた。システムは top-down 的構成からなっており、様々な制御法の比較、知識源の認識への貢献度、各処理部分の処理時間などに関する検討が行われた。現在では、韻律情報の導入が検討されている。⁽³⁵⁾ 京都大学の LITHAN⁽³³⁾では、単語認識レベルで動的計画法(DP)を用いた特徴のあるシステム構成がなされた。対象として、計算機ネットワークの問い合わせ、カレンダー、算術式がとりあげられ、対象の種類と認識率の関係が検討された。現在では、デパートの情報サービスを対象にとりあげ、新しいシステムの検討が開始されている。⁽⁵⁶⁾ 山梨大学のシステムでは、対象として、FORTRANのプログラムや統計に関する事象⁽³⁵⁾ をとりあげて研究が行われた。音響処理の特徴パラメータとして、零交差波、PARCOR係数、対数断面積比など様々な特徴の利用が試みられた。さらに、一般的な日本語を取り扱うことを目標としてシステムの改善が続けられている。

電々公社武蔵野電気通信研究所では、対象として新幹線の座席予約をとりあげ、1974年に研究を開始した。研究対象の選定にあたっては、候補として、音声ダイヤル、計算サービス、座席予約サービス、番号案内サービス、情報検索、ロボットの世界、音声タイプライタなどをとりあげ、次の6つの点から検討を行った。

- (1) 会話の話題が限定されている。
- (2) 入力の語いは、数10～数100程度である。
- (3) 入力の文章は比較的簡単な構文である。
- (4) 入力の文章に意味上の冗長性がある。
- (5) 会話のやりとりがある程度継続する。
- (6) 発声者ごとに音声の学習が可能である。

この検討結果を表1-3に示す。この結果から、我々は、新幹線の座席予約をとりあげた。⁽⁵³⁾ 1975年に予備的なシステム(Voice Q-A System O)を作成した。⁽⁵²⁾⁽⁵³⁾ その結果をもとに、1976年に実時間の5倍の処理時間で動作するオンラインのVoice Q-A System Iを作成した。^{(30)~(32)} このシステムでは、音響処理で特殊なハードウェアを用いて処理速度をあげ、また、言語処理でも単語認識アルゴリズムの高速化がはかられた。このシステムは、会話モデルをもち、質問回答形式で動作する本格的な会話音声認識システムであった。その後、音響処理に重点をおいて研究を進め、1980年に性能が飛躍的に向上したVoice Q-A System IIを作成した。本論文は、これらの武蔵野通研での会話音声認識システムについて、主として、著者が行ってきたVoice Q-A System OとSystem Iの言語処理とVoice Q-A System IIの音響処理の研究を中心に述べる。

ARPAのプロジェクト以外のアメリカの研究機関として、IBMとBell研究所がある。IBMでは、TappertとDixonによって1960年代後半に研究が開始された。⁽²²⁾ その後も、Jelinekのもとで、統計的手法と情報理論的手法を基本とした、network modelを利用した独自の研究が精力的に続けられている。⁽³⁶⁾⁽⁵¹⁾ とくに、非常に精密な音響処理の研究が行われている。しかしながら、実験に用いている発声者

表1-3 研究対象の適合性の比較

対 象 \ 検 討 事 項	話 題 の 限 定	入 力 の 語 意	入 力 の 構 文	入 力 の 冗 長 性	会 話 の 継 続 性	個 人 性 の 学 習
音声ダイヤル	◎	◎	◎	×	×	×
計算サービス	◎	◎	◎	×	×	△
座席予約サービス	○	○	○	○	◎	△
番号案内サービス	○	△	○	○	○	△
情報検索	○	△	○	○	◎	△
ロボットの世界	△	△	△	○	◎	△
音声タイプライター	×	×	×	×	×	○

が1名であること、言語情報の利用が弱いこと、処理時間がかかることなど問題が多い。Bell研では、Flanaganのもとで板倉の単語認識の研究⁽⁵⁸⁾をもとに、音声認識の研究が再開された。最初は、単語ごとに区切って発声した音声での航空券の座席予約システムがつくられた。⁽⁶⁾ その後、Levinsonによって連続音声認識の研究が行われている。⁽³⁹⁾

ヨーロッパでは、フランスのCNET (Centre National D'etudes des Telecommunications) の音声理解系KEALの研究が目につく。⁽³⁷⁾ ヨーロッパの会話音声認識の研究は盛んになるきざしがあり、その中心的な研究機関としてその動向が注目される。

会話音声認識システムの実現の可能性は、各研究機関のシステムで示されており、今後の認識技術の蓄積により、実用的なシステムの実現も遠くないように思われる。とくに、LSI技術などの部品材料の進歩、計算機における並列処理技術の進展、デジタル信号処理技術の普及により、システムの実用化はより加速されるものと思われる。また、電話網とデジタル網との有機的な結合、計算機ネットワークの普及、パターン認識装置の普及、マン・マシン・コミュニケーションの普及などにより、認識装置が社会に受け入れられる下地もでてくるものと考えられる。

1.3 会話音声自動認識研究の問題点および本論文の構成

本論文は、著者が数年間にわたって行ってきた会話音声の認識システムの研究に関する

ものである。

会話音声認識システムを構成するには、少なくとも次の項目について検討せねばならない。

- (1) 音響分析。
- (2) 単語認識アルゴリズム。
- (3) 言語情報の表現法および利用方法。
- (4) 認識システムの構成手法。
- (5) 認識システムの性能評価方法。

(1)の項目は、音声認識の基本的な部分であり、会話音声認識だけでなく、単語音声認識にも、直接にその技術を適用できる。本論文では、音響分析の手法として、分析精度が高いLPC分析をとりあげる。とくに、スペクトルマッチングの尺度として、COSH尺度、およびWLR尺度(Weighted Likelihood Ratio)をとりあげる。スペクトル・マッチング尺度は、母音などの標準パターンとのマッチングにおいて重要であるし、また、音韻単位のセグメンテーションの際のスペクトル変化を計算する尺度としても重要である。COSH尺度が、従来のLPCスペクトルマッチング尺度の中で、もっともすぐれていることを母音識別実験で実証し、認識システムにとり入れる。さらに、スペクトル包絡のピークに重みをおける新しい尺度として、WLRを提案し、従来の尺度よりすぐれていることを実証し、認識システムにとり入れる。WLRは、ピークに重みをおくだけでなく、周波数軸の重みづけもできるすぐれた尺度である。音韻単位のセグメンテーションの手法として、上記の尺度によるスペクトル変化の他に、パワーディップの新しい検出法を導入し、セグメンテーションの精密化をはかる。

(2)の単語認識に関しては、音韻の性質や調音結合の性質が考慮でき、かつ、高速で動作するアルゴリズムを開発することが必須となる。これを、stackを用いたtree searchのアルゴリズムで実現する。この単語認識アルゴリズムでは、tree searchの途中のプロセスを記憶することによって無駄なプロセスをできるだけ減らすことと、連想的な検索方法をとることによって高速化を達成する。また、単語辞書は、通常の音素表記で表わしており、容易に認識単語の変更が行えるようにしている。

(3)の言語情報の利用については、top-down的手法による利用、つまりhypothesis and test流の利用法を主体とする。構文情報のリスト表現を用いた新しい記述法を提

案する。このリスト表現は、構文の直接的な表現方法であるので、人にとっても理解が容易である。よって、構文表現の変更も容易である。意味情報は、ある程度まで構文表現の中に含めることにする。プラグマティクスに関しては、音声による質問回答システムの特徴を考慮し、会話モデルと密接な関連をもたせて作成する。とくに、認識誤りが生じた場合の処理方法、認識結果があいまいな場合の処理方法など、音声認識固有の問題に重点をおく。

(4)のシステム構成法に関しては、音響処理と言語処理とを明確に分離することにより、音響処理と言語処理の普遍性を高め、対象への適応性を高める。音響処理の結果のデータ形式として、音韻系列を一般化した形式の音韻ラティスという新しい形式を用いる。また、会話音声認識を質問回答システムの中で考える場合に、会話モデルは不可欠である。新幹線の座席予約を対象として、発声者とシステムとの対話が自然に行われるように会話モデルを作成する。

(5)の性能評価に関しては、音響処理、言語処理の各部を個々に評価することを考える。音響処理の評価は、従来通り、音韻認識率によって行うことを考える。そのため、音韻ラティスの情報量を用いた評価方法を提案する。さらに、その方法に基づいた自動評価アルゴリズムを確立する。言語処理は、入力音韻ラティスの音韻認識率と、処理結果の文節認識率（予約項目の認識率）とで評価を行う。また、質問回答システムの観点からの評価も試みる。

本論文では、上で述べた問題点に対する手法を中心に著者が数年間、日本電信電話公社武蔵野電気通信研究所で行ってきた会話音声の自動認識システムに関する研究について述べる。

第1章では、会話音声自動認識の研究動向、問題点などについて述べる。

第2章では、本論文でとりあげる会話音声認識システム（Voice Q-A System）の概要について述べる。まず、Voice Q-A Systemの基本構成について述べ、次に、対象としてとりあげる新幹線の座席予約の概要および対象の複雑さの評価について述べる。

第3章では、Voice Q-A System Iの言語処理の構成およびシステムの性能について述べる。そこでは、まず、言語処理における単語認識アルゴリズム、構文解析アルゴリズムについて述べ、さらに、認識実験による性能の評価を行う。次に、質問回答システムとしての立場から、会話モデルについて述べ、さらに、質問回答実験の結果について述べる。この章の最後で、Voice Q-A System Iでの問題点および、改良の方針について述べる。

第4章では、音響処理の評価方法について、情報量の立場から述べる。音響処理の結果は、一般化された音韻系列である音韻ラティスで表わされることが多い。相互情報量を利用した音韻ラティスの評価モデルを提案し、自動評価アルゴリズムを作成する。さらに、このアルゴリズムの評価も行う。

第5章は、Voice Q-A System I の問題点の改良およびVoice Q-A System II の音響処理の基本として、スペクトルのマッチング尺度について、理論的かつ実験的検討を行う。母音識別実験を行い、LPC分析を介して得られる尺度を評価する。まず、従来のLPCスペクトルマッチング尺度の中で、COSH尺度がすぐれていることを実証する。さらに、スペクトル包絡のピークに重みをおくことができるLPC Peak Weighed Measureを提案し、WLR(Weighted Likelihood Ratio)というすぐれた尺度を導出する。このWLRの有効性を母音識別実験および大語い単語音声認識実験で実証する。

第6章では、第3章から第5章までの検討結果をまとめて、最終的な会話音声認識システムVoice Q-A System IIを作成し、システムの性能評価を行う。Voice Q-A System IIでは、言語処理はVoice Q-A System I とほぼ同じアルゴリズムを利用している。音響処理では、第3章で摘出された問題点を解決することを目的とする。その中で、5章のスペクトル・マッチング尺度についての検討が生かされている。また、性能評価では、第4章で述べる音韻ラティスの自動評価システムを利用している。

第7章では、本論文で得られた結果についてのまとめを行う。

次に、本論文の構成と主な検討内容をまとめて示しておく。

(A) Voice Q-A System の概要(第2章)

(B) Voice Q-A System I (第3章)

(a) 言語処理

(i) 単語認識アルゴリズム

(ii) 言語情報の利用

○ 構文解析

○ 推論

○ 会話モデル

(b) システムの評価

(i) 認識システムの評価

- 音韻ラティスの評価法 (第4章)
- (ii) 質問回答システムの評価
- (c) Voice Q-A System I の問題点
 - (i) LPC スペクトル・マッチング尺度の検討 (第5章)
 - 従来のLPC スペクトル・マッチング尺度の評価
 - ピークに重みをおいた尺度 (WLR) の提案
- (C) Voice Q-A System II (第6章)
 - (a) 音響処理
 - (i) COSH 尺度を用いた音響処理 (Voice Q-A System II-1)
 - 音韻単位のセグメンテーション
 - 母音認識
 - 性能評価
 - (ii) WLR 尺度を用いた音響処理 (Voice Q-A System II-2)
 - 性能評価
- (D) 結言 (第7章)

第2章 会話音声認識システム(Voice Q-A System)の基本構成

2.1 はし が き

本章では、会話音声認識システム(Voice Q-A System)の基本構成、および、認識対象としてとりあげた新幹線の座席予約の概要について述べる。認識対象の節においては、情報量の立場から言語情報による冗長性について検討を加える。

2.2 システムの基本構成

会話音声認識システムを、次の方針に基づいて構成することにした。

- (1) 音声による質問回答システムを目標とする。そのため、質問回答のやりとりはすべて音声で行う。よって、音声認識部の他にも、音声応答部ももったシステムを作成する。
- (2) 音声認識部は、その役割が大幅に異なる音響処理と言語処理とに分離して構成する。
- (3) 音響処理と言語処理、および、言語処理と音声応答との間の情報の受け渡しを簡潔な形式にし、各部分の検討を並列的に進められるようにする。

以上のような方針に基づくシステムの構成を図2-1に示す。音響処理では、音声波形を入力として、音韻認識を行い、音韻ラティスの形で、音韻認識の結果と音韻のセグメンテーションの結果を、セグメントの継続時間の情報とともに言語処理に送る。音韻ラティスの例を図2-2に示す。この図からわかるように、音韻ラティスは、音韻のセグメンテーションのあいまいさと、音韻の識別のあいまいさを表現できる音韻系列の拡張表現である。言語処理では、この音韻ラティスを受けとり、単語を認識し、構文を解析して意味内

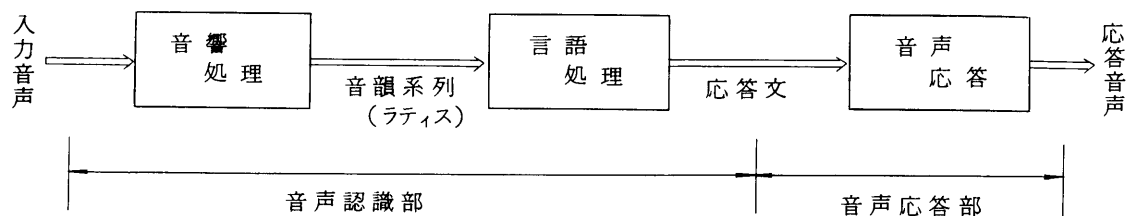


図2-1 Voice Q-A Systemの基本構成

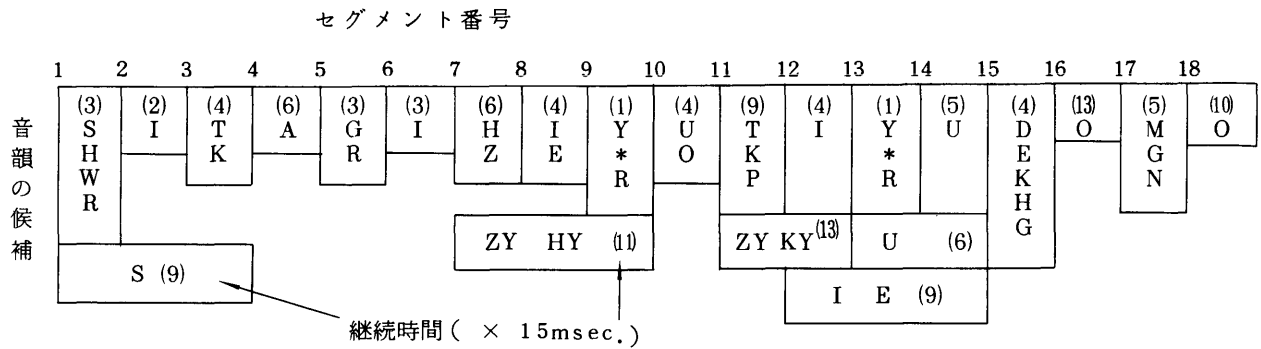


図 2-2 音韻ラティスの例(「ひかり 19 号の」)

容を把握し，さらに時刻表を利用して予約項目間の推論を行い，会話モデルに基づいて発声者への応答文を作成する。音声応答では，言語処理からの応答文に従って音声を合成し，発声者に音声で応答する。

2.3 認識対象

会話音声認識システムの対象として，現在の技術レベルや質問回答形式で行うことを考慮すると，入力の語い数が数 100 以下であること，文章には意味上の冗長性があること，会話のやりとりがある程度継続することが望ましいと考えられる。実用的ないくつかの候補をあげて，これらの事柄について比較検討した結果，座席予約サービスが適当であると判断した⁽⁵⁹⁾

具体的な対象として新幹線の座席予約サービスを取りあげた。システムは，予約項目ごとに区切って発声した入力音声を認識して，予約内容を把握し，発声者に音声で応答する。この座席予約の対象の概要を表 2-1 に示す。入力音声に含まれる単語は，表 2-2 に示す 112 個である。文節は，平均 3.5 個の単語からなり，短いものでは「今日 (KYO)」のように 1 単語からなり，長いものでは，「21 時 42 分発の (NI・ZYU・ITI・ZI・(PAUSE)・YON・ZYU・NI・HUN・HATSU・NO)」のように 11 個の単語からなっている。

以下，Voice Q-A System でのシステムと発声者の，質問回答の実例を示し，その後，対象の複雑さの評価を行う。

表 2 - 1 認 識 対 象

新幹線の座席予約	
(駅 28, 列車 181 本)	
単 語 数	112
平均分岐数	20
予 約 項 目	日付, 発駅, 着駅, 発時刻, 列車名, 等, 枚数
発声の節約	文節ごとに 0.5 秒以上のポーズ
(発声例)	明日の, ひかり 191 号で, 東京から, 博多駅までの, 指定券を, 4 枚, 予約します。

表 2 - 2 認識対象の単語

日付 16	TSUITACHI FUTSUKA MIKKA YOKKA ITSUKA MUIKA NANOKA YOKA KOKONOKA TOKA HATSUKA KYO ASU ASHITA ASATTE HONJITSU
数字 21	1 2 3 4 5 6 7 8 9 0 10 100 SHICHI KU YO I RO HA ZYU ZI RE
駅名 28	TOKYO SHINYOKOHAMA ODAWARA ATAMI MISHIMA SHIZUOKA HAMAMATSU TOYOHASHI NAGOYA GIFUHASHIMA MAIBARA KYOTO SHINOSAKA SHINKOBE NISHIAKASHI HIMEJI AIOI OKAYAMA SHINKURASHIKI FUKUYAMA MIHARA HIROSHIMA SHINIWAKUNI TOKUYAMA OGORI SHINSHIMONOSEKI KOKURA HAKATA
その他の 名詞 15	EKI NICHI JI FUN PUN PPUN HIKARI KODAMA GO GREEN FUTSU SHITEI KEN SEKI MAI
助詞 15	KARA HATSU YORI MADE YUKI IKI E NO O DE NOO NODEWA GA NOWA
動詞 12	YOYAKU ONEGAI ITA SHI MASU ARI MASEN KA DESU DES MAS MOSHIKOMI
その他 5	HAI IIE CHIGAI SO (PAUSE)

計 112 個 (音便変化した単語も含む)

2.3.1 質問回答の例

Voice Q-A System でのシステムと発声者との質問回答のやりとりの例を示す。最初の例は、典型的な質問回答の例である。第 2 の例は、認識誤りが生じたときの質問回答のやりとりの例である。認識誤りが生じたときにも自然な質問回答を通して誤りが訂正できるように会話モデルを設計することが重要である。

(例 1)

計算機 ; こちらは, 新幹線の, 座席予約センター, です。ご希望を, おっしゃって

ください。

発声者；姫路駅より，小郡駅までの，グリーン券で，5日の，こだま217号を，予約します。

計算機；3日ですか？

発声者；いいえ

計算機；何枚ですか？

発声者；5枚

計算機；あなたの，予約は，5日の，こだま217号で，姫路駅発，8時24分の，小郡駅までの，グリーン券を，5枚，ですね？

発声者；はい，そうです。

計算機；ご希望の，指定券は，とれました。予約番号は，4番です。最寄りの，緑の窓口で，お受けとり下さい。

この例では，日付が，3日と5日のどちらかであると認識され，第1の候補が3日であったので，まず，3日がチェックされた。しかし，回答が「いいえ」であったので，日付は5日に決定された。その後，まだ認識されていない枚数の問合せが行われた。発時刻は，発駅と列車名による時刻表を用いた推論によって推定された。最後に，全予約項目の確認が行われ，質問回答が終了した。

(例2)

計算機；(予約開始の案内)

発声者；新横浜から，浜松への，28日の，こだま215号の，グリーンを，1枚。

計算機；普通席，グリーン席の，どちらですか？

発声者；グリーン席

計算機；ご希望の予約は，こだま215号，新横浜駅発，9時2分の，浜松駅まで，と了解しましたが，該当する列車が，ございません，当方の，聞き違いと，思いますので，訂正してください。

発声者；8時34分発。

計算機；あなたの予約は，28日の，こだま215号で，新横浜駅発，8時34分の，浜松駅までの，グリーン券を，1枚，ですね？

発声者；そうです。

計算機；(予約終了の案内)

この例では、最初に入力した「グリーンを」を「9時2分を」に誤認識した。そのため、普通席かグリーン席かの問合せが行われた。そして、全部の予約項目が決まったところで、誤認識による予約項目間の矛盾の訂正が求められた。

2.3.2 対象の複雑さの評価⁽⁶⁰⁾

認識対象の複雑さの程度は、構文が表わし得る単語列の種類数と、言語情報による冗長度によって表わすことができる。構文が表わし得る単語列の種類数は、言語処理における処理量に関連しており、言語情報による冗長度は、音韻系列に含まれる誤りの訂正の可能性を示している。

本システムでの構文が表わし得る単語列の種類数は、 5.8×10^4 通りとなる。予約項目ごとの単語列の種類数を表2-3に示す。本システムでは、予約項目ごとに区切って発声しており、かつ、予約項目の発声順序は任意である。よって、入力の一文節当りの可能な単語列の種類数は、予約項目ごとの単語列の種類数の和となり、約 5.8×10^4 通りとなる。

次に、音韻空間における言語情報による冗長度について述べる。本システムでの各言語情報に対する冗長度を求めると表2-4に示すようになる。言語情報全体による冗長度は11.1になる。また、言語情報のうちでは、構文情報による冗長度が大きいことがわかる。以下、冗長度の計算方法を述べる。

表2-3 予約項目ごとの単語列の種類数と意味内容の種類数

	日付	発駅	着駅	発時刻	列車名	等	枚数	動詞
単語列の種類数	333	1,512	2,016	50,616	3,402	81	81	16
意味内容の数	31	28	28	1,020	189	2	9	1

日本語は、子音・母音の音節(CV音節)からなっていると考えられる。母音(V)として5母音/a, i, u, e, o/と撥音/N/の6種類を考える。子音(C)としては、この対象で現われる19種類の子音/p, t, k, s, h, b, d, g, z, r, m, n, w, y, ky, hy,

表2-4 言語情報による冗長度

		冗長度	情報量(bits)
言語情報	音韻空間	—	52.5
	単語辞書	1.42	36.9
	構文情報	3.10	11.9
	意味内容	1.57	7.6
	時刻表	1.60	4.8
全体		11.1	—

zy, pp, kk/と連続したVの間に考える仮想的な子音/*/の20種類を考える。よって、CV音節の種類数は101通り(/N/は×Nのみ)となる。

n個のCV音節からなる音韻系列が生起する確率は、

$$p_{(n)} \cdot (1/101)^n$$

である。ただし、 $p_{(n)}$ は長さnのCV音節の音韻系列の出現確率である。ここでは、 $p_{(n)}$ の値を、構文表現が表わしうる単語列が等確率で派生するとして計算した。予約項目の出現確率は等しいとした。以上より、音韻空間の情報量 I_s は、

$$\begin{aligned} I_s &= - \sum_{n=1}^N p_{(n)} \cdot \log(p_{(n)}/101^n) \\ &= 52.5 \text{ bits} \end{aligned}$$

となる。ただし、Nは音響系列の最長のCV音節の長さを表わす。

次に、単語辞書のみを用いた場合の空間の情報量(I_w)を求める。CV音節の長さがnのときに単語辞書の組合せで発生できる単語系列の数を $C_{(n)}$ で表わすと、単語辞書のみによる情報量 I_w は、

$$\begin{aligned} I_w &= - \sum_{n=1}^N p_{(n)} \cdot \log(p_{(n)}/C_{(n)}) \\ &= 36.9 \text{ bits} \end{aligned}$$

となる。よって、単語辞書の音韻空間に対する冗長度を求めると、

$$I_s / I_w = 1.42$$

となる。

予約項目の発声確率を等しいとして、構文を用いた場合の情報量(I_{syn})を求める。i番目の予約項目の構文が表わす単語列の種類数を $s_{(i)}$ とすると、

$$\begin{aligned} I_{syn} &= - \sum_{i=1}^M \frac{1}{M} \log\left(\frac{1}{M}\right) / s_{(i)} \\ &= 11.9 \text{ bits} \end{aligned}$$

となる。Mは文節の種類数で、7個の予約項目と動詞の文節からなり、 $M=8$ である。

よって、構文情報による冗長度は、

$$I_w / I_{syn} = 3.10$$

となる。

次に、意味レベルでの冗長度を考える。異なった単語列でも同じ意味内容を表わすことがある。各予約項目の意味内容の数を $d(i)$ とすると、意味内容による情報量 I_c は、

$$I_c = -\sum_{i=1}^M \frac{1}{M} \log \left(\left(\frac{1}{M} \right) / d(i) \right)$$

$$= 7.6 \text{ bits}$$

となる。よって、意味内容による冗長度は、

$$I_{\text{syn}} / I_c = 1.57$$

となる。

次に、時刻表を用いた場合の情報量 I_t を求める。時刻表によらない7項目からなる予約の種類は、 8.0×10^{10} 通りである。そのうち、時刻表に矛盾しない予約の種類は、 6.6×10^6 通りになる。よって、時刻表による冗長度は、

$$I_c / I_t = \log 8.0 \times 10^{10} / \log 6.6 \times 10^6$$

$$= 1.60$$

となる。よって、時刻表を用いた場合の情報量 I_t は、

$$I_t = I_c / 1.6$$

$$= 4.8 \text{ bits}$$

となる。

以上のようにして、冗長度が計算できる。この結果は、表2-4にまとめて示されている。この対象全体の冗長度は、

$$I_s / I_t = 11.1$$

となる。この対象では、構文表現による冗長度が最も大きいことがわかる。

冗長度の大きな対象は、言語情報による音韻系列の誤りの訂正の可能性が大きいといえる。この意味で、冗長度は、対象の困難さを表わすよい尺度である。しかしながら、各種の対象について、上記の冗長度の計算を行うことはむづかしい。とくに、意味情報、意味内容による冗長度の計算には困難が伴う。ここで述べた冗長度に匹敵する尺度で、様々な対象に対して計算が容易な尺度が望まれる。

アメリカのARPAの研究グループでは、対象の複雑さの尺度として、平均分岐数 (average branching factor) を用いている。⁽⁶⁰⁾ 平均分岐数は、文章中で、ある単語の後に続く平均の単語数として定義され、入力音声についての分岐数を計算し、その平均で表わされる。したがって、構文が複雑な場合でも比較的容易に計算できる。しかしながら、単語の長さなどを考慮していないため、冗長度とは直接の関係はない。本システムの平均分岐数は約20であり、ARPAの代表的なシステムであるCMUのHarpyやHearsay-II(表1-1を参照)の30よりも若干小さく、また、京

都大学の LITHAN の 5.0 , 京都工芸繊維大の SPOKEN-BASIC の 15 よりも大きい。

2.4 あとがき

会話音声認識システム (Voice Q-A System) の構成の方針および基本構成について述べた。さらに, 認識対象としてとりあげた新幹線の座席予約について説明した。また, 対象の複雑さの評価尺度として冗長度を提案した。

第3章 Voice Q-A System I の構成と評価

3.1 は し が き

この章では、会話音声認識システム Voice Q-A System I の言語処理に重点をおいて、その構成および性能評価の結果について述べる。

音響処理について簡単に説明した後、言語処理について詳しく説明する。言語処理の top-down 的かつ帰納的な構成法について述べ、順次、単語認識部、構文解析部、推論部、会話モデルについて、言語情報の表現方法とアルゴリズムに重点をおいて説明する。単語認識部では、音韻変形規則を利用した tree search による効率のよい単語認識アルゴリズムを提案する。このアルゴリズムは、様々な音韻の性質を考慮することができるうえに、非常に高速に動作する。構文解析部では、リスト表現を用いた構文の表現法と構文解析のアルゴリズムについて述べる。推論部では、時刻表を利用した推論の手法について、認識結果があいまいである場合や誤りが含まれる場合の手法を中心に述べる。会話モデルは、質問回答を通じて、できるだけ自然に、認識のあいまいさが解消でき、認識誤りが訂正できるように作成される。また、質問回答形式で動作させることを目的として、処理速度を高速化したオンライン Voice Q-A System I についても説明する。

Voice Q-A System I の性能評価を、認識性能の観点と、質問回答の性能の観点から行う。これらの性能評価の結果について詳しく述べる。この章の最後で、Voice Q-A System I の問題点を明らかにして、今後の改良すべき点をあげ、Voice Q-A System II の作成への指針を与える。

3.2 Voice Q-A System I の音響処理⁽³²⁾

この節では、Voice Q-A System I で用いた音響処理について、文献(32),(62)に従って概要を説明する。

音響処理では、図3-1にみられように、入力音声の特徴抽出と音韻認識を行い、音韻ラティスを言語処理に送る。以下、処理過程について簡単に説明する。

入力音声を3.2 kHz 低域通過フィルタに通し，標本化周波数8 kHzで12ビットに量子化し，そのデジタル音声を15 msecごとに分析して，音声パワーの極小点と10次までの自己相関係数の時系列を求める。

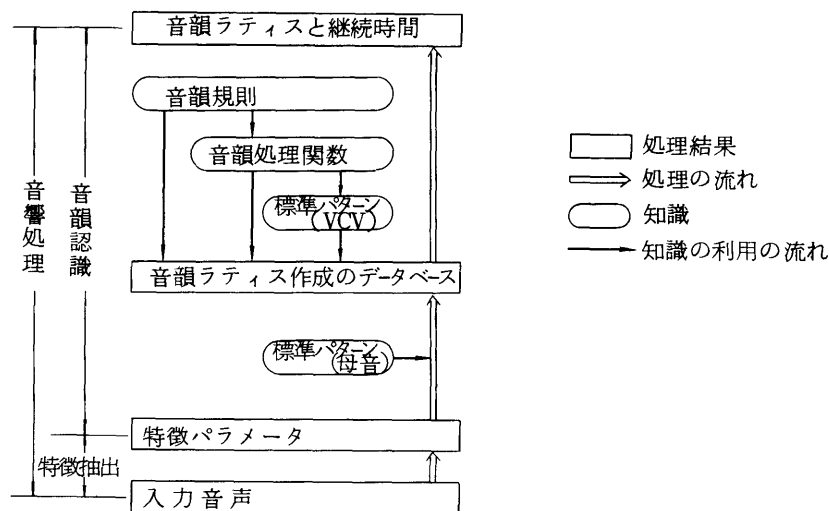


図 3 - 1 音響処理の構成

音韻認識では，まず，音韻ラティス作成のためのデータベースを作る。最初のデータベースの内容は次の通りである。

- (1) 音声パワーの極小点，
- (2) 自己相関係数の時系列，
- (3) 母音系列（母音の標準パターンを用いて，正規化残差で母音認識を行い，スコアが第1位と第2位の母音を求める），
- (4) 音韻ラティスの初期値（上の(1)と(3)を利用して，入力音声を母音区間，過渡区間，休止区間に分ける）。

次に，このデータ・ベースをもとに，音韻レベルの音響的知識を利用して，最終的な音韻ラティスを作成する。利用する知識の内容は次の通りである。

- (1) 音韻規則

音韻ラティスの初期値における誤りを訂正するための規則で，音韻処理関数を起動させ，音韻ラティスにセグメントの追加を行う。

- (2) VCV 音節標準パターン

VCV音節の識別を，端点フリーのDPマッチング法によって行う。

これらの知識を用いた音韻認識の手順を次に示す。

- (1) 音韻規則を順に音韻ラティス上に適用する。音韻ラティス上で規則と合致する部分が見つかり、対応する音韻処理関数が呼ばれる。
- (2) 呼ばれた音韻処理関数は、音韻規則によって指示された VCV セグメントのチェックを行い、その結果に応じて、VCV 音節標準パターンとのマッチングで音韻を認識し、音韻ラティスに認識結果を付け加える。
- (3) 以上の手順をくり返すことにより、音韻認識を実行する。

Voice Q-A System I の音響処理は、以上のような手順で音韻ラティスを作成し、言語処理に音韻情報を渡す。詳しくは、文献(32),(62)を参照されたい。

3.3 言語処理の基本構成

言語処理は、単語認識部、構文解析部、推論部からなる認識部と、会話モデルからなっている。この構成を図3-2に示す。

認識部は、図3-2からわかるように、推論部を除けば、top-down 的な方向でのマッチングを用いて構成されており、単語認識部は構文解析部の一部になっている。処理の流れ(図3-2の太い矢印)は下位の概念(音韻ラティス)から上位の概念(構文表現や深層構造)に向かっているが、認識の手順は上位の概念を仮定し、下位の情報により、その仮定を検証するという、上位から下位へのマッチングの流れ(図3-2の細い矢印)に従っている。

マッチングは、left-to-right の順序で行われる。left-to-right のマッチングを採用する理由は、入力音声を文節ごとに区切って発声する場合、強く明確に発声されるキーワードが文節の先頭にくることが多いことと、処理の制御が簡単になることによる。

推論部では時刻表を用いて予約項目間の推論を行う。この部分は bottom-up 的な構成をとっている。これは、認識においてある程度の誤りが生ずるのはさけられなく、認識結果と時刻表を利用して構文解析での単語を予測す

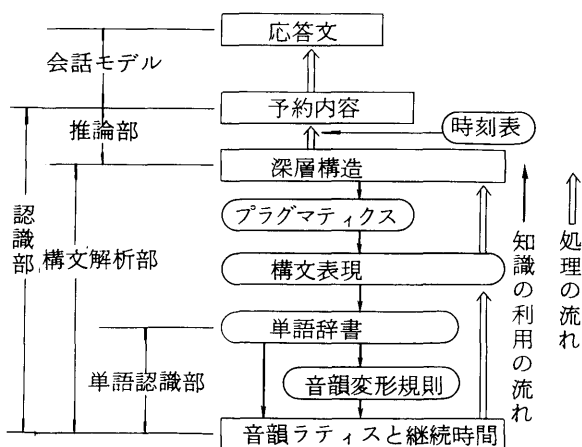


図3-2 言語処理の構成

るような top-down 的な構成にした場合、誤った認識結果が生じたときに、その誤りと時刻表による予測がシステムの動作に悪い影響を与えることがあると考えられるためである。

会話モデルは、発声者とシステムとが自然な質問回答をくりかえして、最終的には、システムが正しい予約内容を把握するように作成される必要がある。そのために、7個の会話状態を設け、各状態間の状態推移で会話モデルを表現した。

言語処理は、ミニコン PFU-400 上に、著者が作成した FORTRAN によるリスト処理システム DLOP(付録4)を用いて作成されている。

3.4 単語認識部

単語認識部では、構文解析部から提示された単語辞書と音韻ラティスとのマッチングが行われる。連続音声の中の単語を認識するためには、単語のセグメンテーションと単語の識別という2つの問題を解決する必要がある。この2つの問題のからみ合い、音韻ラティスのあいまいさ、種々の音韻変形規則の適用などが、連続音声の中の単語のマッチングのプロセスを非常に複雑にしている。よって、これらの複雑さを秩序づけて整理し、強力でかつ、能率のよい単語認識のアルゴリズムを得ることが必要となる。

音韻変形規則を用いた音声認識の手法は、単語音声認識に見られる^{(65)~(67)}。しかし、規則の種類も不十分であり、マッチングの能率もよくない。また、音韻変形規則を用いた連続音声の中の単語の認識は、本格的には試みられていない。本システムでは、音声学の知識や音響処理の性能を分析して、音韻の性質を、14種類の型からなる477個の書き換え規則で表わした。この多種多様な音韻変形規則を利用して、単語辞書と音韻ラティスとのマッチングを、tree searchのdepth-first methodによるアルゴリズムで実行する。単にdepth-first methodでマッチングを行うと、ステップ数が膨大になるので、むだな処理をできるだけ減らすようにアルゴリズムを改良して、単語認識の実行速度を飛躍的に向上させる。

3.4.1 音韻変形規則

音韻の性質を音韻変形規則として、書き換え規則の形で表わす。各々の書き換え規則は、規則の適用条件として、音韻ラティスの音韻の継続時間も利用する。さらに各々の規

則にはペナルティがつけられている。音韻変形規則は、表3-1に示すように、母音に書換える規則(6種の型)、子音に書換える規則(6種の型)、音韻の読みとばしの規則、および母音の無声化の訂正規則からなる。各々の書換え規則は、単語辞書の音韻をキーとして連想的に検索されるようにリスト形式で蓄えられており、単語認識アルゴリズムにおいて、能率よく検索されうる tree 形式のデータ構造になっている。Voice Q-A System I で用いた音韻変形規則のリスト表現を付録1に示しておく。

表3-1 音韻変形規則

	変形規則の内容	変形規則の型	継続時間
母音に関する規則	短母音から長母音への変形	$V_1 \rightarrow V_d$	$t(V_1) \geq D$
	調音結合による母音の誤りの訂正	$C_d V_1 C_{1d} \rightarrow C_d V_d C_{1d}$	$t(V_1) \leq D$
	母音のそう入 (1)	$V_1 \rightarrow V_d V_1$	$t(V_1) \geq D$
	母音のそう入 (2)	$V_{1d} \rightarrow V_{1d} V_d$	$t(V_{1d}) \geq D$
	母音と音節との間の誤りの訂正	$C_1 V_1 \rightarrow V_d$	—
	同じ母音ではさまれた子音のそう入	$V_{1d} \rightarrow V_{1d} C_d V_{1d}$	$t(V_{1d}) \leq D$
子音に関する規則	子音の類似性(1)*	$C_1 V_d \rightarrow C_d V_d$	$t(C_1) \geq D$
	子音の類似性(2)*		$t(C_1) \leq D$
	子音のそう入	$V_d V_{1d} \rightarrow V_d C_d V_{1d}$	—
	調音結合による誤りの訂正	$C_1 V_1 C'_1 \rightarrow C_d V_d C'_d$	—
	文頭の子音のそう入	$* \rightarrow * C_d$	—
	拗音のセグメント化誤りの訂正	$C_1 V_1 C'_1 \rightarrow C_d$	$t(V_1) \leq D$
その他	音韻の読みとばし	$P_d P_1 P_{1d} \rightarrow P_d P_{1d}$	$t(P_1) \leq D$
	母音の無声化の訂正	$C_1 \rightarrow C_d V_d C'_d$	$t(C_1) \geq D$

* 子音間の類似性の規則を、継続時間情報の利用方法の違いにより、2種類の規則にわけた

- (1) 添字 l, d, ld は、 l : 音韻ラティスの音韻、 d : 単語辞書の音韻、 ld : ラティスと辞書の両方に存在する音韻、を表わす。
- (2) $t(x)$ は音韻 x の継続時間を表わす。

3.4.2 単語辞書

単語辞書は、通常の音素表記で、リストで表わされている。音素として、 $/A, I, U, E, O/$ の5母音、 $/II, UU, EE, OO/$ の長母音、拗音 $/NN/$ 、子音 $/P,$

T, K, S, H, B, D, G, Z, R, M, N, W, Y, KY, HY, ZY, PP, KK/を用いている。単語辞書のリストには、単語名と同じ名前がつけられており、構文表現から直接に検索される。また、単語辞書には、音便変化も含まれている。単語辞書のリスト表現も付録1に示しておく。

3.4.3 単語認識アルゴリズム

単語辞書と音韻ラティスのマッチングを、tree searchによるアルゴリズムで実行する。tree searchによるマッチングは、動的計画法によるマッチング⁽³³⁾⁽³⁴⁾⁽⁶⁷⁾よりも、音韻の様々な性質を利用することができる。しかしながら、音韻の様々な性質を利用するために、計算量(ステップ数)が増大しやすい。よってtree searchによるアルゴリズムにおいては、ステップ数の削減が重要な問題となる。

tree searchによるアルゴリズムとして、記憶容量が少なくて済み、かつ、searchの制御構造が簡潔であるdepth-first methodを採用した。depth-first methodによる単語認識のアルゴリズムのフローチャートを図3-3に示す。

単語認識部では、構文解析で予測された単語の単語辞書と、その単語の認識を試みるラティス上のスタートセグメントのリストを受けとり、単語辞書と音韻ラティスとのマッチングをleft-to-rightの順序でおこなう。マッチングは、すべてのスタートセグメントから試みられ、マッチングが可能なラティス上のエンドアドレスをすべて求める(図3-4)。マッチングで音韻変形規則を適用すると、各々の規則のペナルティが加えられ、このペナルティの総和でマッチングの度合を表わす。単語認識の結果は、リ

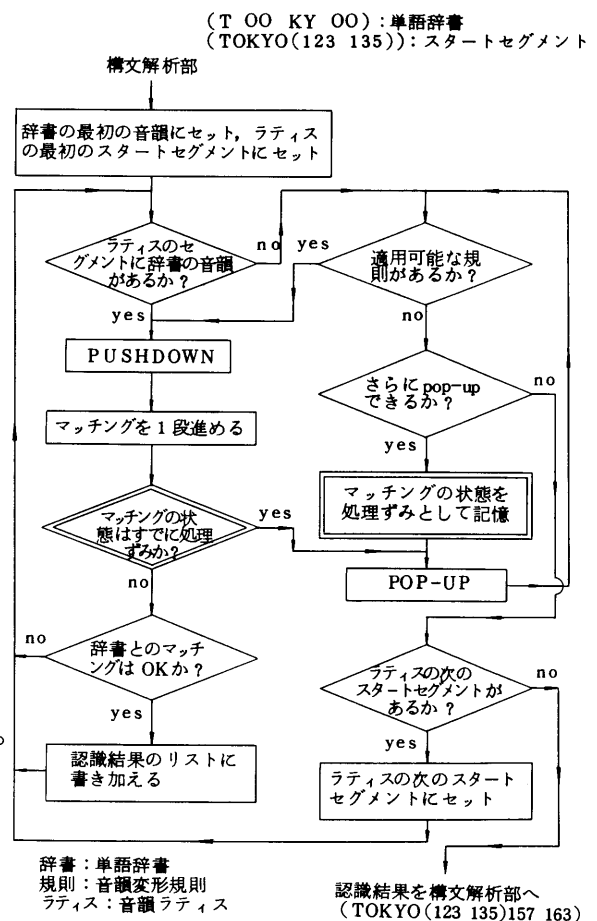


図3-3 単語認識アルゴリズムのフローチャート

ストで表わされ、ペナルティの総和とともに構文解析部に返される。depth-first method でプッシュダウン・スタックにプッシュダウンされるのは、表3-2の9項目である。

言語処理での処理時間の大部分は、単語認識部で費やされており⁽⁶⁸⁾、処理を高速化するには、単語認識アルゴリズムのステップ数を減らすことが必要である。そのために、次に示す4つの手法を併用した。

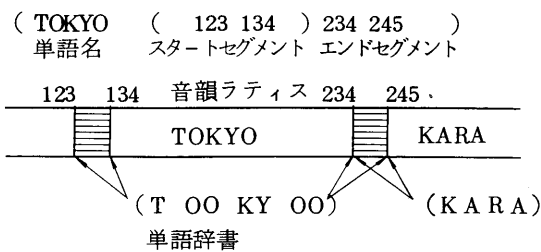


図3-4 単語認識の結果のリストと単語のセグメンテーションのあいまいさ

表3-2 単語認識アルゴリズムでpush-down stackにたくわえる項目

1	音韻ラティスの状況	音韻ラティスのアドレス
2		音韻ラティスのセグメントの番号(α)
3		音韻ラティスの音韻の候補のアドレス
4		セグメントの継続時間
5	単語辞書の状況	単語辞書の音韻のアドレス(β)
6	音韻変形規則の状況	音韻変形規則の種類
7		音韻変形規則のアドレス
8	マッチングの状況	ペナルティの総和(r)
9		直前に認識した音韻

- (I) 音韻変形規則のペナルティの総和に制限をもうけることによって tree search の枝切り (tree pruning) をおこなう。
- (II) tree search のプロセスを記憶し、それを参照することによって、同じプロセスをくり返さないようにする。
- (III) 単語の境界をあいまいなまま、まとめて処理し、単語認識の回数を少なくする。
- (IV) 単語認識の結果を記憶し、のちのプロセスで利用する。

手法(I)のペナルティの総和に対する制限は、depth-first method でのバックトラックが必ず生じるようにするためにも不可欠である。ペナルティの制限(P)には、単語辞書の音素表記の長さ按比例した次の値をもちいる。

$$P = P_0 + N + 1$$

P_0 は、文章の最初では、音韻認識の誤りが多いため、4の値をとり、その他では2の値をとる。 N は辞書の音素表記の長さである。入力音声が入力音節の構文で拒絶された場合には、このペナルティの総和に対する制限(P)の値を2だけ大きくして、再度、認識できなかった単語の認識を試みるようにしている⁽⁶⁹⁾。

手法(Ⅲ)は、音声認識でよく利用されている動的計画法において、途中の段までの最適値のみを求めて記憶し、後段でその最適値を利用して全体の最適値を求めるという考えの拡張とみなすことができる。この手法は、図3-3のフローチャートの2重の線で囲んだ部分の処理に相当する。適用する音韻変形規則がなくなってPop-upが生ずるとき、Pop-upする前の状態が処理済みとして、次の3つの組の形式で記憶される。

- (i) 音韻ラティスのセグメント番号(α),
- (ii) 単語辞書の音韻のアドレス(β),
- (iii) ペナルティの総和(r),

この3つ組(α , β , r)は音韻ラティスのセグメント番号をキーにして、連想的に検索できるように記憶される。このように蓄えられた処理済みの状態は、図3-3のフローチャートに示すように、マッチングを1段進めるごとに参照される。すなわち、マッチングの状態(α^0 , β^0 , r^0)を、処理済みとして記憶された状態(α , β , r)と比較して、もし、 $\alpha^0 = \alpha$, $\beta^0 = \beta$, $r^0 \geq r$ を満たす(α , β , r)が存在すれば、(α^0 , β^0 , r^0)は処理済みと判定され、Pop-upを生じさせる。こうすることにより、処理速度が約4倍向上した⁽⁷⁰⁾。

次に手法(Ⅳ)について説明する。入力音声は単語列からなっており、単語認識では、単語の境界も求めなければならない。そのとき、単語の境界は、通常、一意にはきまらない。ここで複数の単語境界の候補について、別々に、次に続く単語とのマッチングをおこなうと、単語認識の回数が増大する。これをさけるために、すべての単語境界の候補から単語辞書とのマッチングをbreadth-first的に同時におこなうようにした。これにより、処理速度が約1.2倍向上した⁽⁷⁰⁾。

手法(Ⅴ)では、単語認識の結果を、失敗した場合と成功した場合とにわけて記憶し、この結果を単語認識の直前に参照し、すでに認識が試みられていれば、その結果を構文解析部に返す。これによる処理速度の向上は、約1.1倍であった⁽⁶⁸⁾。

さらに、本システムでは、マッチングのプロセスや音韻変化規則の記憶形式として、

リスト形式をとり， tree 構造による連想的な検索ができるようにデータが蓄えられている。このデータ構造の改良によって，処理速度が約 6 倍向上した⁽⁶⁸⁾。

以上のステップ数削減のための手法(II)~(IV)とデータ形式の改良によって，処理速度が約 30 倍に向上した。現在，言語処理は実時間の約 2 倍の処理時間で動作している。 tree-search による本手法は，音韻の様々な性質を利用できるうえに，従来の動的計画法を利用した手法に近い処理速度で動作していると推定される。

連続音声中の単語認識の手法として tree search の手法を用いている。 tree search 以外の手法としては，動的計画法 (DP) による手法が存在している。一般に， DP は string 対 string のマッチングに関しては能率のよいアルゴリズムを与えてくれる。しかし， tree 対 string, tree 対 tree のマッチングには向いていないように思われる。我々のシステムでの単語認識では， tree (音韻ラティス) 対 string (単語辞書) のマッチングとなっているため， tree search 法によるアルゴリズムが適していると思われる。 tree search の手法としては， depth-first 法の他に， breadth-first, best-first, beam search などが存在している。この節で述べたように， depth-first 法は， tree search の手順をスタック・アルゴリズムで簡潔に記述でき，かつ，記憶容量も少なくすむという長所をもっている。しかしながら， tree search の過程で派生する branch の数は best-first 法より多くなる。 breadth-first 法は，並列的に tree を探索するために，非常に大きな記憶を必要とする。よって，そのまま用いられることは少なく，探索の範囲を発見的に有限個に限定した beam search 法に改良されて利用されることが多い。 best-first 法は，派生する branch の数が最小という意味で最適であるが， branch を派生するたびに，スコアが best である leaf を探索するという「並びかえのアルゴリズム」を必要とするため，1 個の branch を派生するための処理時間は大きくなる傾向にある。 beam search も本質的には「並びかえのアルゴリズム」を必要とするため同じような問題点をもつ。この節では，単語認識アルゴリズムの depth-first 法からの定式化について述べた。今後， depth-first 法， best-first 法， beam search 法の融合が必要になると考えられる。

3.4.4 動作例

実例をあげて，単語認識のプロセスを示そう。「熱海ゆきの」と発声した図 3-5

の音韻ラティスの先頭から，単語辞書（ A T A M I ）とのマッチングのプロセスについて説明する。マッチングは，図 3 - 3 のフローチャートに従って行われる。入力されるスタートセグメントは，

(ATAMI (1))

の形式で表わされる。ここで (1) は，音韻ラティスの先頭のセグメントの番号を示す。以下，マッチングのプロセスを示す。

(単語認識のプロセスの例)

β α r

push-down (# 1 0)

push-down (A 2 0)

push-down (T 3 0)

push-down (A 4 0)

push-down (M 5 0)

push-down (I 6 3) U → I (3)

OK (A T A M I (1) 7)

push-down (I 7 5) N → * (2)

OK (A T A M I (1) 7 8)

記憶 (N I L 8 5) ⇒ M 1

popup (I 7 5)

M 1 参照 R → * (2), M → * (3), B → * (3)

記憶 (I 7 3) ⇒ M 2

popup (I 6 3)

M 2 参照 NN → I (4)

記憶 (I 6 0) ⇒ M 3

popup (M 5 0)

push-down (M 5 ~ 7 1) N → M (1)

push-down (I 8 1)

OK (A T A M I (1) 7 8 9)

記憶 (N I L 9 1) ⇒ M 4

popup (I 8 1)

```

push-down ( I 5 ~ 7 3 ) I → I · I (2)
    記憶 ( NIL 8 3 ) ⇒ M 1
popup ( I 5 ~ 7 3 )
    記憶 ( I 7 1 ) ⇒ M 2
popup ( I 8 1 )
    記憶 ( I 8 1 ) ⇒ M 5
popup ( M 5 ~ 7 1 )
    M 2 参照 G → M(2), Z → M(3), MUH → M(3)
    M 3 参照 N → M(1), Z → M(3), R → M(3)
    記憶 ( M 5 0 ) ⇒ M 6
popup ( A 4 0 )
    記憶 ( A 4 0 ) ⇒ M 7
popup ( T 3 0 )
    M 7 参照 P → T(1), S → T(3)
    記憶 ( T 3 0 ) ⇒ M 8
popup ( A 2 0 )
push-down ( A 1 2 )
    記憶 ( A 1 2 ) ⇒ M 9
popup ( A 2 0 )
    記憶 ( A 2 0 ) ⇒ M 10
popup ( # 1 0 )
    M 10 参照 W → *(2), H → *(3), M → *(3)
push-down ( # 0 2 ) ⇒ M 11
popup ( # 1 0 )
    記憶 ( # 1 0 ) ⇒ M 12
popup ..... end

```

結果 (ATAMI (1) 7 8 9) ペナルティ(1)

ただし、上述のプロセスでは、便宜上、 β は単語辞書の音韻、 α は音韻ラティスのセグメントの番号、 r はペナルティの総和を表わしている。#は文頭を表わす記号である。「記憶(NIL 8 5)⇒M 1」は、処理済みの状態(NIL 8 5)を記憶し、

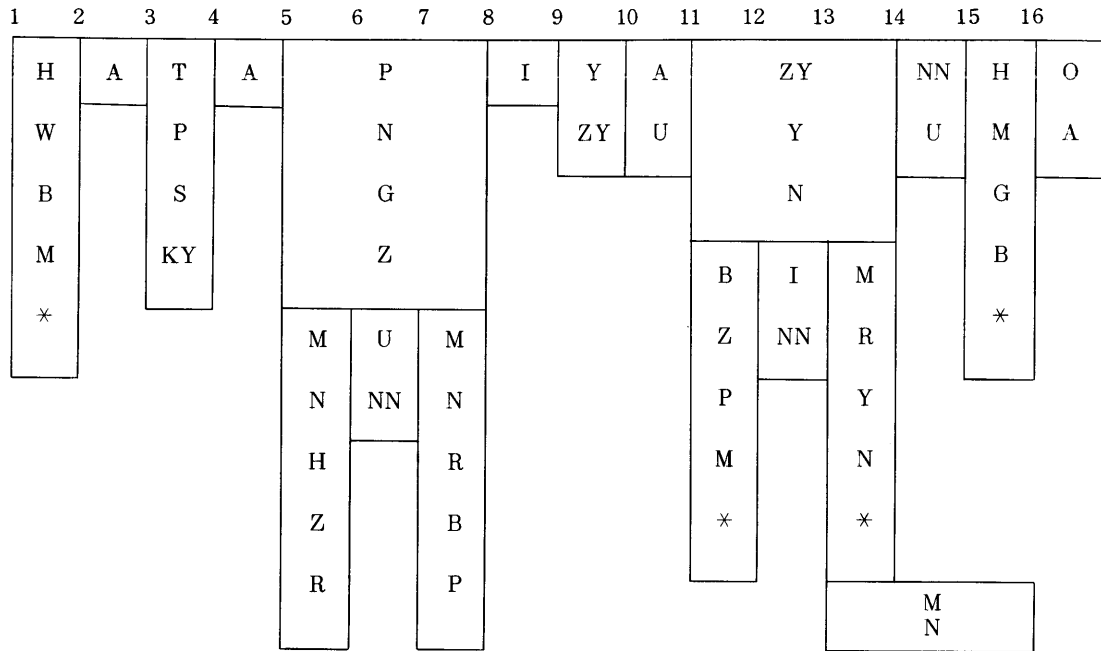


図 3-5 音韻ラティスの例(「熱海ゆきの」)

その記憶を M1 と名付けることを意味する。「M1 参照 R→*(2)」は、音韻変形規則 R→* がペナルティ 2 で適用できるが、適用した後の状態は、すでに処理済みの状態として、記憶されている M1 に含まれるので、その規則を適用する必要はないと判定されたことを示す。「OK(ATAMI (1) 7)」は、単語辞書と音韻ラティスのマッチングが成功し、認識結果を示すリスト(ATAMI (1)) にエンド・アドレス 7 を書き込んで(ATAMI (1) 7) をしたことを示している。

この例では、12 回の push-down で単語認識が終了している。この例を tree 状の状態推移図で表わすと、図 3-6 のようになる。図 3-6 でまるの中は音韻ラティスのセグメント番号を、実線の矢印は、push-down を、破線は処理済みの記憶の参照を示している。

次に、処理済みの記憶を利用しない場合との step 数の比較をこの例で行う。処理済みの記憶を利用しない場合には、blind search となり、push-down の回数は、

$$1 + 4 (1 + ((1 + 1) + 3 (1 + (1 + 4 (1 + 2 (1 + 4)) + 4 (1 + 2)))))) = 709 \text{ 回}$$

となる。この例からもわかるように、処理済みの状態を記憶し、参照にすることにより、step 数を大幅に削減することができる。

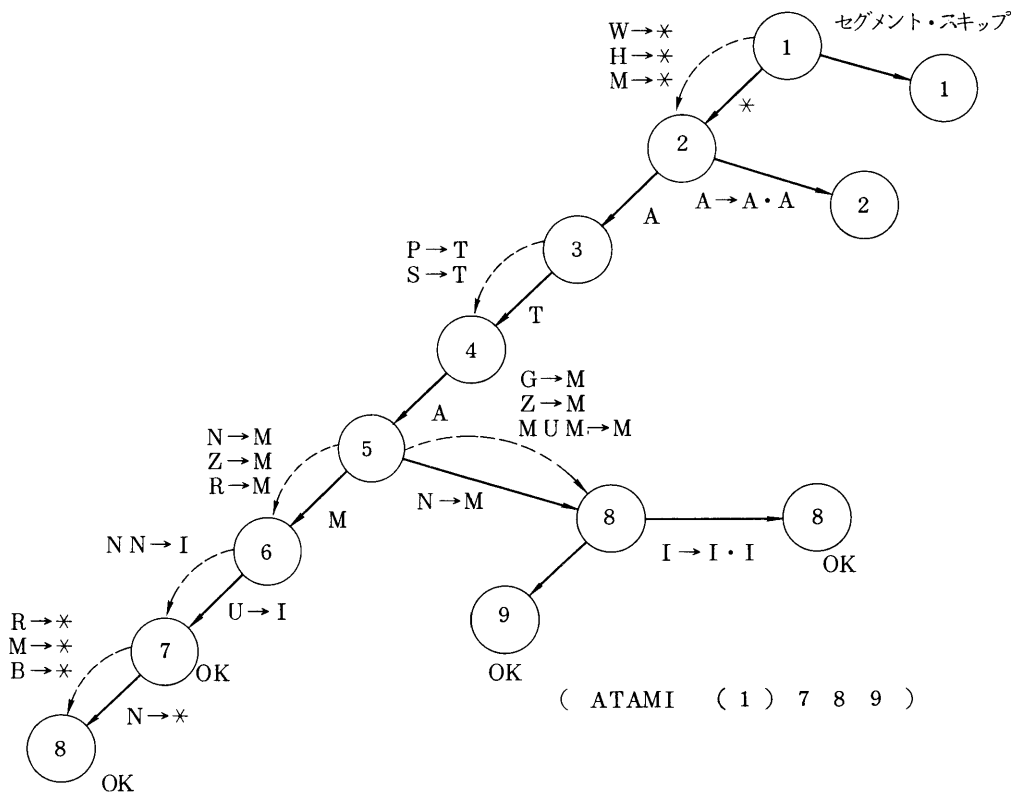


図 3 - 6 単語認識のプロセスの例

3.5 構文解析部

会話音声認識システムの多くは、構文情報をBNF (Backus-Naur Form) で表現する⁽³³⁾か、あるいは、transition network grammar⁽²⁸⁾で表現している⁽³⁴⁾⁽³⁵⁾⁽⁴⁵⁾。

BNFは、文脈自由形文法と等価であり、そのままでは、文章の構文を表わすのに不十分である。transition network grammarは、文章の構文の表現に適し、かつ、その構文表現に従って単語を予測するのに適した方法であると思われる。本節では、このtransition network grammar とほぼ等価な能力を持つ、リスト表現を用いた簡潔な構文の表現方法を提案する。

3.5.1 プラグマティクス

プラグマティクスでは、会話モデルの会話状態、予約項目のは握状態などを参考に、構文解析をする文節の順序を指定する。プラグマティクスは、リスト表現で手

続きとして、図3-7のように表わされている。このリスト表現では、次の演算子とセマンティックルーチンを用いている。

```
( SEM (YES-NO)(# YES-NO) SEM (VERB) SEM (LATTICE 4 20)(# VERB)
SEM (SYNTAX 1) SEM (LATTICE 4 24)(# DATE) SEM (SYNTAX 2) SEM
(LATTICE 6 26)(# STARTING-STATION) SEM (SYNTAX 3) SEM (LATTICE
8 26)(# ARRIVING-STATION) SEM (SYNTAX 4) SEM (LATTICE 12 64)(#
STARTING-TIME) SEM (SYNTAX 5) SEM (LATTICE 10 64)(# NAME-OF-TRA
IN) SEM (SYNTAX 6) SEM (LATTICE 2 18)(# SEAT-CLASS) SEM (SYNTAX
7) SEM (LATTICE 4 14)(# NUMBER-OF-TICKETS))
```

図3-7 プラグマティクスのリスト表現

SEM (YES-NO) 会話状態が「はい、いいえ」で予約内容を確認する状態であるか。

SEM (VERB) 入力音声が文章の最後の文節であるか。

SEM (LATTICE a b) 入力の音韻ラティスの長さがa以上でb以下であるか。

SEM (SYNTAX i) i番目の予約項目が未定か。ただし、会話状態が「矛盾の訂正の要求」のときには、常に未定とみなされる。

以上のセマンティックルーチンでは、条件が成立すれば、次のリスト表現に、さもなければ、(# -)の次のリスト表現に処理をとばす。

(# x) x という名前のリストで表わされる構文表現を構文解析せよ。

プラグマティクスの解析は、depth-first methodで、left-to-rightの順序で行なわれる。リスト表現で表わされているため、リストの変更やセマンティックルーチンの変更により、構文解析する文節の順序を変更したり、発声する文節の順序を固定したりすることが、容易におこなえる。プラグマティクスの変更による文節の発声の順序を固定した場合については、3.9.3節で述べる。

3.5.2 構文解析

構文は、7項目の予約項目、「はい、いいえ」、動詞の各々の文節ごとに、リスト表現で表わされる。各文節のリストには名前がつけられ、プラグマティクスのリスト表現(# -)から検索される。

構文表現の一部を図3-8に示す。構文表現は、次に示す演算子とセマンティックル

ーチンによって表わされる。

(AND $X_1 \dots X_n$) X_1, \dots, X_n のリスト表現が、この順序で存在する。通常、ANDを省略して表わす。

(OR $X_1 \dots X_n$) X_1, \dots, X_n のリスト表現のうち、1つの表現が存在する。

(OPT X) リスト表現Xは存在しなくてもよい。

SEM (x $a_1 \dots a_n$) セマンティックルーチン x を動作させる。 $a_1 \dots a_n$ はルーチン x のパラメータである。

(* x) x という名のリストにとぶ。

上記以外の演算子は、単語辞書と見なされて、単語認識部の単語認識プログラムを動作させ、その結果を受けとる。

```
DATE: ((OR (* TSUITACHI) (* FUTSUKA) (* MIKKA) (* YOKKA) (* ITSUKA) (* MUIKA)
(* NANOKA) (* YOKA) (* KOKONOKA) (* TOKA) (* HATSUKA) (* KYOASU)
((OPT (* 2)) (* 10) (OR ((OR (* 1) (* 2) (* 3) (* 5) (* 6) (* SHICHI) (* 7) (* 8)
(* KU) (( * NICHU) (* YOKKA))) (( * 3) (* 10) (OPT (* 1)) (* NICHU))) (* JOSHI))
STARTING-STATION: (( * EKIMEI) (OPT (* EKI)) (OR (* KARA) (* HATSU) (* YORI))
(* JOSHI))
ARRIVING-STATION: (( * EKIMEI) (OPT (* EKI)) (OR (* MADE) (* YUKI) (* IKI)
(SEM (LATTICE 35) (* E))) (* JOSHI))
STARTING-TIME: (( * SUJI6-22) (* JI) (OPT (* PAUSE)) (* SUJI0-59 FUN) (OPT
(* HATSU)) (* JOSHI))
NAME-OF-TRAIN: ((OR (( * HIKARI) (OPT (* PAUSE)) (* SUJI1-199)) (( * KODAMA)
(OPT (* PAUSE)) (* SUJI200-299))) (* GO) (* JOSHI))
SEAT-CLASS: ((OR (* SHITEI) (* FUTSU) (* GREEN)) (OPT (OR (* KEN) (* SEKI)))
(* JOSHI))
NUMBER-OF-TICKETS: (( * SUJI1-9) (* MAI) (* JOSHI))
YES-NO: (OR (* HAI) (* IIE) (( * SO) (* DESS)) (( * CHIGAI) (* MASS)))
VERB: (OR ((OR ((OR (* YOYAKU) (* ONEGAI)) (OPT (* ITA)) (* SHI)) (( * MOSHIKO
MI) (OPT (( * ITA) (* SHI)))))) (* MASS)) (( * ARI) (OR (* MASU) (* MASEN))
(* KA)) (* DESS))
```

図3-8 構文表現 (リストの名前のうち、JOSHI, EKIMEI, SUJI6-22, SUJI0-59 FUN, SUJI2-5, SUJI1-199, SUJI200-299, SUJI1-9, KYOASU, MASS, DESS, NICHUは構文リスト表現を、その他は単語辞書の音素表記したリスト表現を表わす)

構文解析は、リスト表現に従って、depth-first methodでleft-to-rightの順序でおこなわれ、各演算子は、スタックを用いた帰納的なプログラムで解釈される。

スタックには、表 3-3 の 5 項目がプッシュダウンされる。以下、各々の演算子の動作について簡単に説明する。

- AND, OR, OPT ……各々の演算子を示す指標を含む表 3-3 の 5 項目をプッシュダウンする。
- * …… (* X) として用いられ、構文表現のアドレスを、X という名のリストのアドレスにおきかえる。
- 単語辞書 ……検索された単語辞書と音韻ラティスのマッチングのプログラムを動作させる。単語認識が失敗のときは、back-track を生じさせる。単語認識が成功したときには、AND によって示される次の構文表現のアドレスを見つけ出し、構文表現のアドレスを変更する。さらに音韻ラティスのスタートセグメントのアドレスも進め、直前に認識された音韻もセットする。また、単語認識の結果をリストにして蓄え、そのリストのアドレスをセットする。

表 3-3 構文解析アルゴリズムで push-down stack にたくわえる項目

1	演算子の種類を示す指標
2	構文表現のアドレス
3	音韻ラティスのスタートアドレスのリスト
4	直前の単語で認識された語尾の音韻
5	直前の単語認識の結果のリスト

次に back-track のときの処理に

ついて説明する。back-track は、OR か OPT の指標が見つかるまで、スタックからポップアップしながら実行される。これらの指標が見い出されないときには、構文解析は終了となり、プラグマティクスに戻る。OR の指標

が見い出されたときには、構文表現のアドレスを、OR によって示される次の構文表現のアドレスに変更して構文解析をつづける。OR によって示される構文表現が存在しなときには、さらに back-track を生じる。OPT の指標が見い出されたときには、OPT のカッコでくくられている構文表現が存在していないとして、AND で示される次の構文表現の解析をおこなう。

入力音声全体が矛盾なく構文解析されると、そのキーワードが抽出され、予約内容がシステムには握される。同時に、単語列を構成する単語のマッチングの度合の総和もシステムに記憶される。また、1 つの予約内容が認識されても、同じ文節内の構文の他の解析も試みられる。

これら 4 個の演算子を用いて文節の構文を自然に表現することができる。また、リスト表現で手順として構文を記述しているため、リスト表現の変更により、構文の変

更も容易におこなえる。例えば、発駅の項目の発声では、語尾に「から」を常につけ、「東京から」と発声するというような発声の制約を加えることも容易にでき、予約項目の発声方法を制限したシステムに、簡単に変更できる。このような制約を加えた場合についても、3.9.3節で述べる。

リスト表現で表わされた構文表現を付録の1に示しておく。

3.5.3 動作例

「1日の、新大阪から、博多までの、6時2分発の、ひかり19号の、指定券を、9枚、お願いします」と発声した音声を入力した場合の構文解析部の動作の例を付録2に示す。

3.6 推論部

時刻表を用いて、発駅、発時刻、列車名の予約項目間で推論をおこなう。その推論の種類を表3

表3-4 推論の種類

	推論の型および予約項目の状態	他の予約項目の状態
1	発 駅(definite) → 発時刻(ambiguous)	列車名(not definite)
2	発 駅(definite) → 列車名(ambiguous)	発時刻(unknown)
3	発時刻(definite) → 発 駅(ambiguous)	列車名(unknown)
4	発時刻(definite) → 列車名(ambiguous)	発 駅(not definite)
5	列車名(definite) → 発時刻(ambiguous)	発 駅(not definite)
6	発 駅(definite), 発時刻(definite) → 列車名(not definite)	/
7	発 駅(definite), 列車名(definite) → 発時刻(not definite)	
8	発時刻(definite), 列車名(definite) → 発 駅(not definite)	

definite : 一意に予約内容が定まっている。

ambiguous : 予約内容があいまいである。

unknown : 予約内容が未定である。

not definite : 予約内容が未定あるいはあいまいである。

－ 4 に示す。推論では，予約内容のあいまいさを解消し，未定の予約項目をすでに認識した予約項目から推定する。推論の例を図 3－9 に示す。推論部は会話モデルとも密接に関連して動作する。推論を行う際に，用いている時刻表を付録 1 に示しておく。

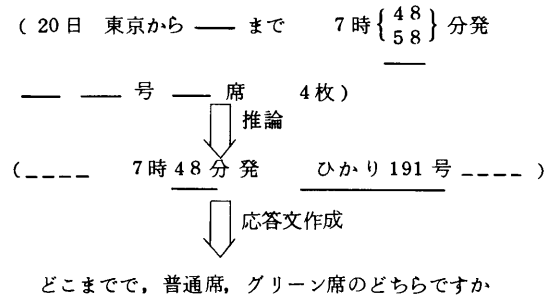


図 3－9 推論の例（推論の型，1，6）と応答文の例

3.7 会話モデル

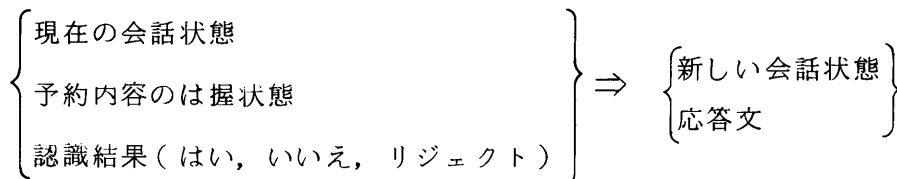
座席予約の質問回答をおこなうための会話モデルについて説明する。

会話モデルは次に示す 7 つの会話状態からなっている。

- (1) 予約開始の案内，
- (2) 未定の予約項目の問合せ，
- (3) あいまいな予約項目の確認，
- (4) 全予約項目の確認，
- (5) 予約項目間の矛盾の訂正の要求，
- (6) その他の訂正要求，
- (7) 予約完了の案内，

これらの会話状態の説明および各状態で発せられる応答文の例を表 3－5 に示す。

会話状態の状態推移図を図 3－10 に示す。会話状態間の推移は，



の形の推移規則に従っておこなわれる。すなわち，

- ・現在の会話状態：現在，7 つの状態のどこにいるか。
- ・予約内容のは握状態：まだわからない予約項目はどれか，認識結果が一意に決まらないであいまいな予約項目はどれか，全部の予約項目が決まったか，予約項目間に矛盾はないか，
- ・認識結果：システムからの確認の応答に対して入力の結果が「はい」であるか，

表 3 - 5 会話状態の説明と各状態における応答文の例

会話状態	説明	応答文の例
予約開始の案内	予約の開始を発声者に告げる。	こちらは新幹線の座席予約センターです。ご希望をおっしゃってください。
未定の予約項目の問合せ	まだ認識していない予約項目を問合わせる。	何日でどこからですか？ どこからですか？
あいまいな予約項目の確認	あいまいな予約内容を確認する。	東京駅からですか？ 9時40分発ですか？
全予約項目の確認	すべての予約項目が矛盾なく一意に認識されたとき、予約内容を確認する。	あなたの予約は、3日のひかり191号で、東京駅7時48分発、名古屋駅までの普通券を3枚ですね。
予約項目間の矛盾の訂正の要求	予約項目間に矛盾があるために、発声者に訂正を要求する。	ご希望の予約は、ひかり141号、東京駅、9時30分発、博多までと了解しましたが、該当する列車がございません。当方の聞き違いと思いますので、訂正をお願いいたします。
その他の訂正の要求	発声者に訂正を要求する。	訂正をお願いいたします。
予約完了の案内	予約が完了したことを、発声者に告げる。	ご希望の指定券はとれました。予約番号は4番です。最寄りの緑の窓口でお受取り下さい。

「いいえ」であるか、予約内容を訂正した入力が認識できたか、リジェクトされたか、に従って、

- ・新しい会話状態：7つの状態のいずれかに推移して、
- ・応答文：新しい会話状態に対応する応答文のいずれかで、予約内容の把握状態によって決まる、

を出力する。

質問回答のやりとりが、できるだけ自然な形で行えるように、会話状態間の推移規則は次の原則に従って作成した。

(1) 予約の内容に、あいまいな予約項目があ

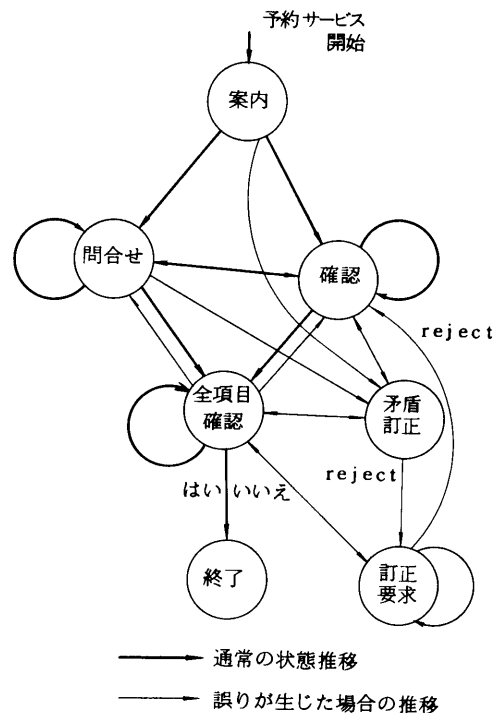


図 3 - 10 会話状態推移図

るときには、まず、その予約項目の確認を行って、あいまいさをなくすようにする。そのさい、認識結果の第1位の候補から順に確認し、最後の候補まで来たときには、確認しないで、それに決める。なお、認識結果の候補の数にはとくに制限を設けていなくて、認識のさいに音韻変形規則を適用したことにとまなうペナルティの総和がいき値を越えないものすべてを認識結果とし、そのペナルティの総和の少ないものから順に、第1位、第2位、……と順序づける。

(2) 次に、まだわからない予約項目の問合せを行う。

(3) 予約内容に、あいまいな予約項目やまだわからない予約項目が2つ以上あるときには、表2-1の予約項目順に、確認や問合せを行う。

(4) 予約項目間の矛盾のチェックは、全部の予約項目が決まってから行い、そこで矛盾が見つければ訂正を求める。

基本的な原則は以上であるが、質問回答のやりとりができるだけ少ない回数で済むように、次の考慮も払った。

(5) 発時刻と列車名の予約項目にあいまいさがあるときには、すぐには確認を行わないで、次のようにする。発駅が決まると時刻表を用いて発時刻と列車名の候補をかなり強力にしばることができるので、まず、発駅の予約項目を認識し、それに基づく推論によって、時刻表に合致しない発時刻と列車名を消して、まだあいまいさが残っていれば確認を行う。発時刻と列車名は、他の予約項目に比べて、認識結果が一意に決まらないことがとくに多いので、このような推論によって、確認の回数をかなり減らす効果がある。

(6) 通常は、推論を行いながら、予約内容を認識する。しかしながら、認識誤りが生じたときには、それに基づく誤った推論が行われて、認識誤りの訂正ができないことがありうる。このような事態をさけるために、推論によってある予約項目の認識結果が全部消された場合は、誤った推論が行われたかも知れないと判断して、そのあと全予約項目の確認を行うまで、推論の機能を一時止める。

時刻表を用いた推論は、発駅、発時刻、列車名の予約項目の間で行っており、その内容は表3-4に示した。推論が実行されて、予約内容が変わったときには、再び推論できるかどうか調べて、これを予約内容が変わらなくなるまで繰り返す。

次に、上で述べた会話モデルによって質問回答を行った会話例を示す。

(例1)

計算機：(予約開始の案内)

発声者：姫路駅より，小郡駅までの，グリーン券で，5日の，こだま217号を，予約します。

計算機：3日ですか？

発声者：いいえ

計算機：何枚ですか？

発声者：5枚，です。

計算機：あなたの予約は，5日のこだま217号で，姫路駅8時24分発，小郡駅までのグリーン券を，5枚ですね？

発声者：はい。

計算機：（予約終了の案内）

この例では，日付は，3日（第1位）と5日（第2位）に認識されたが，3日でないことを確認して，5日に決まった。そのあと，まだ入力していない枚数の問合せが行われた。また，発時刻は推論によって決まった。

（例2）

計算機：（予約開始の案内）

発声者：新横浜から，浜松への，28日の，こだま215号の，グリーンを，1枚。

計算機：普通席，グリーン席のどちらですか？

発声者：グリーン席

計算機：御希望の予約は，こだま215号，新横浜駅9時2分発，浜松駅までと了解しましたが該当する列車がございません。当方の聞き違いとしますので訂正をお願い致します。

発声者：8時34分発。

計算機：あなたの予約は，28日のこだま215号で，新横浜駅8時34分発，浜松駅までのグリーン券を，1枚ですね。

発声者：そうです。

計算機：（予約終了の案内）

この例では，最初に入力した「グリーン券を」を「9時2分を」に誤認識した。そのため，普通席かグリーン席かの問合せが行われ全部の予約項目が決まったところで，予約項目の矛盾による訂正が求められた。これに答えて，発声者が発時刻「8時34分発」を告げ，これの認識結果に基づいて推論が行われて，正しい予約項目がは握され，予約が完了した。

3.8 オンライン質問回答システムの構成

音声理解系の考え方の特徴の一つとして、発声者との質問回答形式で動作することがあげられる。Voice Q-A System Iでは、質問回答実験を行うために、実時間の約5倍で動作するオンラインシステムを作成した。

言語処理はミニコンPFU-400上に、音響処理、音声応答はミニコンNeac3200/70上に作成されている。2台のミニコンはチャンネル結合されており、音響処理からは音韻ラティスが、言語処理からは応答文が転送される。使用する計算機システムの構成を図3-11に示す。

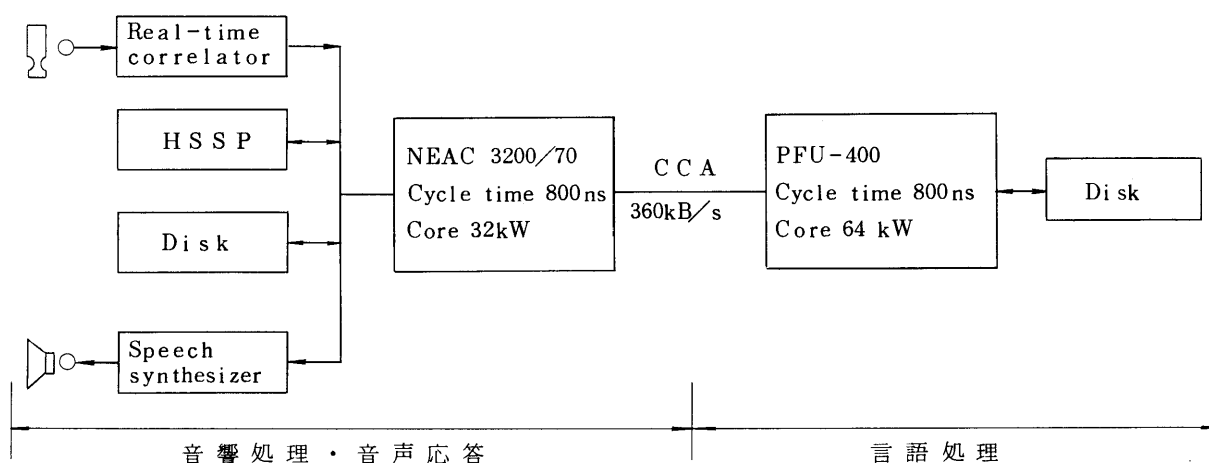


図3-11 Voice Q-A System Iの計算機システム

3.8.1 音響処理の高速化

音響処理部では、入力音声の特徴抽出と音韻認識を行い、音韻ラティスの形で、音韻の認識結果と継続時間の情報を言語処理部に送る。

特徴抽出では、分析窓として32ミリ秒の長さのハミング窓を用い、分析のフレーム周期を15ミリ秒にとって、音声波形の相関分析を実時間相関器(realtime correlator)で行い、音声パワーと自己相関係数の時系列を求める。

音韻認識では、まず、入力音声を母音区間、子音区間、過渡区間にセグメント化し、母音区間の認識を行う。子音区間の認識を、前後の母音区間と組み合わせて、VCV音節単位で行い、最終的な音韻ラティスを作る。VCV音節の認識には端点フリーDPマッチング法を用いて、母音の認識とともに、高速音声処理装置HSSP(high speed speech processor)で行う。

このように、特別に作成されたハードウェアである実時間相関器と高速音声処理装置 HSSP を用いることにより音響処理を実時間の 3 倍余りの処理時間で動作するようにしている。

3.8.2 言語処理の高速化

言語処理は、FORTRAN によるリスト処理システム DLOP⁽⁶⁸⁾⁽⁷³⁾ を用いて、PFU-400 上に作成されている。言語処理で最も時間を要する単語認識アルゴリズムは、tree search アルゴリズムの高速化の他にも、リスト処理の特質を生かした tree 構造による連想的な検索によって、さらに高速化がはかられている。また、リスト処理システム DLOP の基本ルーチンをアセンブラで記述することによる、DLOP 自体の高速化も行われている。このようなソフトウェアの高速化により、言語処理は実時間の約 2 倍の処理時間で動作している。

リスト処理システム DLOP の概要を付録 4 に示す。

3.8.3 音声応答部

各会話状態に対応した応答文を、音声合成して出力する。この音声出力を、単語または文節を単位とする録音編集方式で作る。

応答文には、表 3-6 に示されるように、9 種類の固定文と 14 種類の可変文がある。表 3-6 で可変文に下線を引いたところが可変部分であり、そこにそう入する単語または文節は、合計 460 種類である。その内訳は、次の通りである。

日 付 (31)	: 1 日 ~ 31 日
駅 名 (28)	: 東京駅 ~ 博多駅
時 間 (78)	: 6 時 ~ 23 時 0 分発 ~ 59 分発
列 車 名 (2)	: こだま, ひかり
列車番号 (294)	: 1 号 ~ 195 号 (ひかり) 200 号 ~ 298 号 (こだま)
券 (2)	: 普通券, グリーン券
枚 数 (9)	: 1 枚 ~ 9 枚
予約番号 (10)	: 4 番, 13 番, ……………, 811 番

その他 : 何日で, どこから, どこまで, 何時何分発で, どの列車で, 普通席グリーン席のどちらで

録音編集のための音声データは, 1名の女性が発声した。すべての応答文に必要な文章, 文節, 単語を, 3.2 kHzの低域通過フィルタに通して, 6.4 kHzで標本化し, 8ビットに量子化してムービングヘッド型のディスクに蓄えた。ディスクの使用領域は2.3 M語(1語16ビット)である。

表3-6 会話状態とその状態に対応する応答文

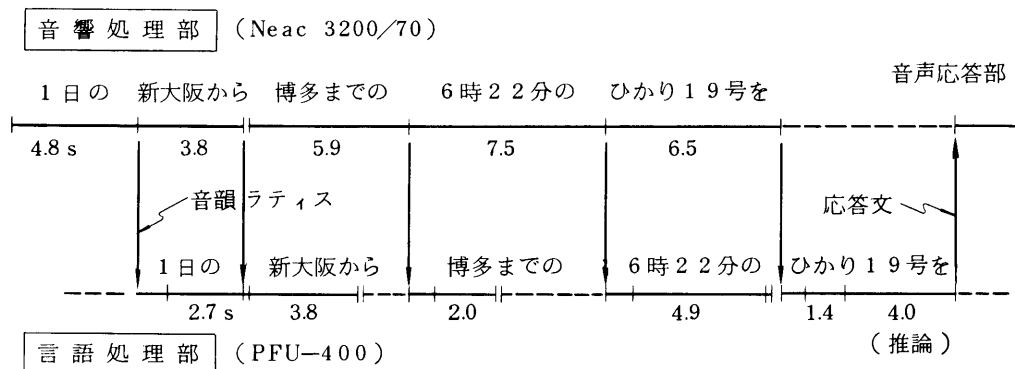
会話状態	固定文, 可変文	応答文の例
予約開始の案内	固定文 (1)	こちらは新幹線の座席予約センターです。ご希望をおっしゃって下さい。
未定の予約項目の問合せ	固定文 (7)	何日ですか? どこからですか? どこまでですか? 何時何分発ですか? どの列車ですか? 普通席・グリーン席のどちらですか? 何枚ですか?
	可変文 (6)	何日でどこからですか? どこからどこまでですか? どこまでで何時何分発ですか? 何時何分発でどの列車ですか? どの列車で普通席・グリーン席のどちらですか? 普通席・グリーン席のどちらで何枚ですか?
あいまいな予約項目の確認	可変文 (5)	<u>1日</u> ですか? <u>東京駅</u> からですか? <u>博多駅</u> までですか? <u>9時40分</u> 発ですか? <u>9時台</u> ですか?
全予約項目の確認	可変文 (1)	あなたの予約は <u>3日のひかり191号</u> で <u>東京駅7時48分</u> 発, <u>名古屋駅</u> までの <u>普通券</u> を <u>3枚</u> ですね?
予約項目間の矛盾の訂正の要求	可変文 (1)	ご希望の予約は, <u>ひかり41号</u> , <u>東京駅9時30分</u> 発, <u>博多駅</u> までと予解しましたが, <u>該当する列車がございません</u> 。 当方の聞き違いとしますので訂正をお願い致します。
その他の訂正の要求	固定文 (1)	訂正をお願い致します。
予約完了の案内	可変文 (1)	ご希望の指定券はとれました。予約番号は <u>4番</u> です。 最寄りの緑の窓口でお受け取り下さい。

(下線部は可変部分を表わす)

3.8.4 処理時間

処理のタイムチャートの一例を、図3-12に示す。入力は、「1日の、新大阪から、博多までの、6時22分の、ひかり19号を」という文章である。音響処理部では、入力音声の特徴抽出を実時間で処理したあと、最初の文節の音韻認識を行う。その音韻ラティスが得られると直ちに言語処理部に転送し、言語処理部で単語認識と構文解析を行っている間に、音響処理部では次の文節の音韻認識を開始する。このあとは、両方の処理部の処理が終了したことを確認してから、再び音韻ラティスを転送して、同様の処理を最後の文節まで繰り返す。そのあと、言語処理部で推論を行い、最終的な予約内容をは握する。

こうして、音響処理部と言語処理部が2台のミニコン上で並列に動作する。音韻認識に要する処理時間は、図3-12にみられるように、入力された各文節の長さにはほぼ比例する。それに対して、単語認識と構文解析に要する処理時間は、入力の長さよりも、むしろ、入力される文節の順序や構文の複雑さの程度に強く依存する。したがって、各文節の音韻認識に要する処理時間と単語認識・構文解析に要する処理時間の割合は、必ずしも一定してないのが普通である。



処理速度

システム全体	実時間の約5倍
音響処理	実時間の3.4倍
言語処理	実時間の2.0倍

図3-12 Voice Q-A System I のタイムチャートの例

平均の処理時間は、音響処理部が実時間の3.4倍、言語処理部が実時間の2.0倍、それら全体では実時間の5.0倍である。この結果は、連続音声の認識としては非常に高速であるといえる。なお、音声応答部では、一つの応答文を作るのに2~10秒か

かる。

3.9 認識実験による性能評価

Voice Q-A System Iの認識性能を評価するために、男性8名が発声した座席予約の文章音声データを用いて認識実験を行った。

3.9.1 音声資料

騒音レベル69dB(A)の計算機室で、文節ごとに区切って発声した座席予約の文章を対象として、認識実験および質問回答の実験をおこなった。

発声者は男性8名で、各発声者は20文章を2回発声した。1つの文章は平均6.2個の文節からなり、1つの文節は平均3.4個の単語からなる。

発声した文章を付録3に示しておく。ここでの認識実験では、グループ1の20文章を2回発声したものをを用いた。

音響処理は、文献(32)のシステムによる音響処理では、母音の標準パターンは各自ごとに作成し、子音の標準パターン(VCV音節)は、発声者RNのものに固定した。認識実験に用いる男性8名の発声した音韻ラティスに、正しい音韻が含まれている割合は、平均で78.6%(母音85.9%,子音71.4%)であった。音韻ラティスには、多くのセグメンテーションのあいまいさと音韻認識のあいまいさが含まれているので、音韻ラティスのもっている情報量で評価されるのが望ましい。しかし、情報量は直感的にわかりにくいので、同じ情報量をもつ通常の音韻系列の音韻認識率に変換して評価した。音韻ラティスの情報量から推定した音韻認識率は、平均で、58.5%(母音77.8%,子音39.1%)であった。誤りは、セグメンテーション誤りが26.0%,認識誤りが15.5%である。

音響処理の評価方法である音韻ラティスからの音韻認識率の推定方法については第4章で詳しく説明する。

3.9.2 認識実験の結果

音韻ラティスを入力として、言語処理の性能を調べた。発声者ごとの音韻認識率と

文節認識率を表3-7に示す。文節認識率を、文節中のキーワードの認識の正誤から求める。文節認識率には、複数個の認識結果が得られ、その中に正しい結果が含まれているものも含まれている。すべての音韻の標準パターンを学習した発声者RNでは、95.7%の文節認識率が得られた。男性8名の平均の文節認識率は、86.0%であった。また、予約項目別の文節認識率を表3-8に示す。構文の複雑な発時刻、列車名の認

表3-7 音声者ごとの音韻認識率と文節認識率 (%)

発声者		RN	HN	SF	KI	KS	SS	SA	MK	平均	
文節認識率	正 (correct)	95.7	82.2	89.1	90.0	85.2	88.3	74.3	83.0	86.0	
	拒絶(reject)	3.0	13.0	8.7	7.4	11.7	9.6	17.4	10.0	10.1	
	誤り(error)	1.3	4.8	2.2	2.6	3.1	2.1	8.3	7.0	3.9	
音韻認識率	ラティスに正しい音韻が含まれている割合		85.1	74.5	78.9	78.8	76.4	79.3	77.5	78.8	78.6
	情報量から推定	平均	62.7	55.2	60.6	59.1	58.2	59.3	55.6	57.1	58.5
		母音	80.7	76.0	80.0	78.6	76.4	80.2	74.8	75.9	77.9
		子音	44.7	34.4	41.2	39.6	39.9	38.4	35.0	38.3	39.0

母音の標準パターンは発声者ごとに、子音の標準パターンは発声者RNから作成された。

表3-8 予約項目別の認識結果 (%)

予約項目		日付	発駅	着駅	発時刻	列車名	等	枚数	全体
correct	total	87.1	87.5	83.9	80.2	80.3	89.7	90.0	86.0
	unique	37.9	70.0	68.1	56.8	57.7	75.0	63.3	62.5
	not unique (*)	49.2 (73.0)	17.5 (87.5)	15.8 (95.8)	23.4 (46.7)	22.6 (87.2)	14.7 (66.0)	26.7 (89.1)	23.5 (78.1)
reject		5.9	10.9	11.8	13.5	14.9	8.1	7.9	10.1
error		7.0	1.6	4.3	6.3	4.8	2.2	2.1	3.9

(*)はnot-uniqueのうちで第1位の候補が正しい割合

識率が若干低いことがわかる。誤りは平均で3.9%であり、その大半は、予約項目間にまたがる誤り(例, グリーン券を → 9時2分を)である。認識率86.0%の内訳は、正しい結果が一意に得られた場合が62.5%, 複数個の認識結果が得られて、正しい結

果が第1位にある場合が18.4%，第2位以下である場合が51%であった。認識結果が複数通りの場合は，会話モデルからの確認の質問に，発声者が「はい」または「いいえ」で答えることによって，正しい結果がシステムには握される。

単語の認識の様子を，8名の平均に近い文節認識率の発声者KSについて調べた。この結果を表3-9に示す。文章中の全単語のうち正しく認識された単語の割合（単語認識率に相当）は87.2%であった。正しい単語の約2.7倍の単語が認識され，そのうち，約1.8倍の単語が誤って認識された単語である。単語認識では言語処理での処理時間の大半が費やされるので，誤って認識された単語の割合を少なくすることが，処理速度の向上には，必要であると考えられる。

表3-9 単語の認識結果

文章数	20コ
文節数	124コ
単語数	423コ
正しく認識された単語	369コ (87.2%)
誤って認識された単語	767コ

(発声者KS)

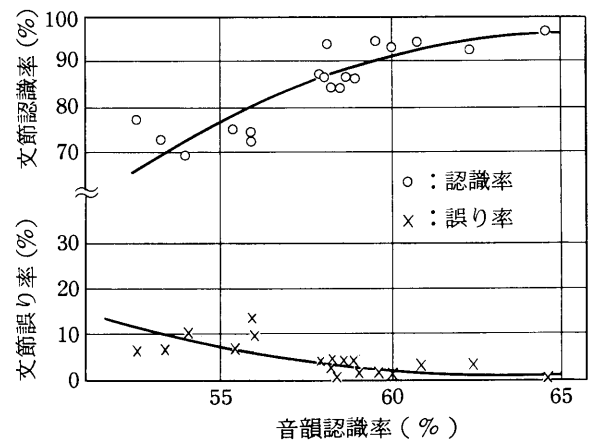


図3-13 音韻認識率と文節認識率の関係

音韻認識率と文節認識率との関係を調べるために，各発声者の20文章ごとの音韻認識率と文節認識率をプロットしたのが，図3-13のグラフである。システムが非常に良好に動作する目やすである95%の文節認識率を達成するには，63%以上の音韻認識率を得る必要があることが推定できる。

言語処理での誤り訂正能力を調べるために，音響処理での音韻認識の誤りと，言語処理で訂正できなかった誤りとの対比を，表3-10に示す。言語処理では，音響処理の誤りの96.8%を訂正している。音響処理での大半の誤りは，「余分なセグメントの付加」(43.1%)と，「子音間の認識誤り」(22.4%)であるが，この2種類の誤りは，言語処理で99.5%まで訂正され，システム全体の拒絶や誤りの原因の9.4%にすぎない。システム全体でみた場合の誤りや拒絶の原因の大半は，次の4項目である。

- (1) 文頭の母音の誤り(21.4%)

- (2) 拗音 (/ zy, ky, hy /), 半母音 (/ y /) のセグメンテーションの誤り (19.5 %)
- (3) 有声子音や音韻間の渡りの区間を母音とする誤り (10.7 %)
- (4) 撥音 / N / のセグメンテーションと認識の誤り (10.7 %)

認識システムの性能を向上させるためには、音響処理におけるこれらの誤りを少なくするとともに、言語処理もさらに精密化する必要がある。

表3-10 音響処理での音韻認識の誤りと言語処理で訂正できなかった誤り

音響処理部での誤り		言語処理部での誤り			誤り訂正率 $\frac{(A-B)}{A} \times 100(\%)$
割合(%)	個数 A	誤りの原因	個数 B	割合(%)	
2.3	104	文頭の母音の誤り	32	21.4	69.2
2.5	117	拗音のセグメント化誤り	29	19.5	75.2
1.6	71	子音過度区間→母音	16	10.7	77.5
2.5	116	撥音 / N / の認識誤り	16	10.7	86.2
8.8	406	無声子音の後の母音	8	5.4	98.0
1.8	84	連続母音	6	4.0	92.9
7.2	334	その他の母音の調音結合	12	8.1	96.4
0.9	41	子音のそう入	6	4.0	85.4
22.4	1038	子音間の認識誤り	5	3.4	99.5
0.4	20	母音→子音	3	2.0	85.0
0.2	10	ポーズ→無声子音	1	0.7	90.0
0.0	1	2重の無声化	1	0.7	0.0
43.1	1996	余分なセグメントの付加	9	6.0	99.5
6.3	290	子音の脱落	0	0.0	100.0
—	—	プラグマティクス	5	3.4	—
	4628	合計	149		96.8

(発声者8人、各自124文節)

3.9.3 発声の制約による効果⁽⁷¹⁾⁽⁷²⁾

Voice Q-A Systemの言語情報は、リスト表現で表わされており、リスト表現の変更によって、言語情報の変更が容易におこなえる。この節では、言語情報のリスト表現を変更することによって、2種類のシステムを作成し、発声制約した場合の認識実験の結果について述べる。

Voice Q-A Systemでは、予約項目の発声の順序は自由であり、予約項目の構文

も比較的自由である。この節では、このシステムを System A と呼ぶことにする。この System A のプラグマティクスのセマンティック・ルーチン SEM (SYNTAX i) の解釈を次のように変更して、予約項目の発声の順序を固定したのが System B である。System B では、SEM (SYNTAX i) を、i 番目入力の文節音声が入力された場合のみ、i 番目の予約項目を構文解析すると解釈する。さらに、構文の制約を設けたのが System C である。

System A, B, C における発声の制約と、各々のシステムに適合した発声例を表 3-11 に示す。この表からわかるように、System B では、予約項目の発声順序を、日付、発駅、着駅、発時刻、列車名、等、枚数、動詞と固定している。System C では、予約項目の発声順序を System B のように固定し、さらにそのうえ、表 3-11 の発声例において、下線で示された助詞を常に発声するようにした。

表 3-11 各認識システムの発声の制約とそれに適合した発声例

システム名(発声データ名)	発声の制約	発声例
System A (data a)	文節の発声順序自由	ひかり 151 で、名古屋から、相生までの、グリーン席を、1 枚おねがいします。
System B (data b)	文節の発声順序固定	10 日で、名古屋から、相生までの、8 時 27 分発、ひかり 151 号で、グリーン席を、1 枚、ありますか。
System C (data c)	文節の発声順序固定 構文制限	10 日 <u>の</u> 、名古屋から、相生まで、8 時 27 分 <u>の</u> 、ひかり 151 号 <u>の</u> 、グリーン <u>を</u> 、1 枚。

各システムの発声の制約に適合した座席予約の文章 (data a, data b, data c) を男性 4 名が各々 20 文章ずつ発声した。data a の発声内容を、付録 3 のグループ 1 に、data b はグループ 3 に、data c はグループ 4 に示しておく。これらの音声データを用いて認識実験をおこなった。

各音声データに付する各システムの文節認識率と、各音声データの音響処理の性能 (音韻認識率) とを表 3-12 に示した。System A は data a ~ c を認識することができる。

System B は data b, c を、System C は data c のみを認識することができる。data c は、比較的認識の容易な助詞をつけて発声するように、構文を制約している

ため、若干高い音韻認識率をもつ音声データとなり、いずれのシステムでも93~94%程度の文節認識率を達成している。data bはdata aとほぼ同じ音韻認識率であるが、data aよりも2.0%~3.4%高い文節認識率を示している。System Aでも

表3-12 各認識システムの発声の制約とそれに適合した発声例

	音響処理の性能		システム名	SystemA	SystemB	SystemC
data c (560文節)	音韻認識率	60.9%	文節認識率*	92.9%	93.9%	93.0%
	認識誤り率	15.2%	拒絶率	5.0%	4.8%	6.1%
	セグメント化誤り率	23.9%	誤り率	2.1%	1.3%	0.9%
data b (640文節)	音韻認識率	57.9%	文節認識率*	87.9%	89.3%	
	認識誤り率	14.8%	拒絶率	8.0%	7.9%	
	セグメント化誤り率	27.3%	誤り率	4.1%	2.8%	
data a (496文節)	音韻認識率	57.8%	文節認識率*	85.9%		
	認識誤り率	15.0%	拒絶率	9.3%		
	セグメント化誤り率	27.2%	誤り率	4.8%		

* 認識率のうち70~76%は一意に認識されたもので、残りは複数個の認識結果の中に正しいものが含まれている割合である。そのうち、約75%は第1位に正しい結果がある。

data bの文節認識率が高いのは、System Aでもdata bのような予約項目の出現順序をプラグマティクスで予想しているためと、予約項目の冗長性を利用した時刻表による推論のためである。data b, data cに関して、システムの制約の効果をみると、制約の強いシステムほど誤り率が小さくなる傾向にある。

しかし、認識率は余り変らない。次に、各音声データに対する文節当りの言語処理の処理速度を表3-13に示した。制約の強いシステムほど処理速度が速いことがわかる。さらに、data b, data cは、data aよりもかなり処理速度が速いこともわかる。

表3-13 各システムの文節当りの処理速度

	SystemA	SystemB	SystemC
data c	3.4sec	2.9sec	2.8sec
data b	3.5sec	2.9sec	
data a	4.5sec	(発声者1名, 20文章)	

data a~cによる認識率の向上ほど、System A~Cによる向上は大きくない。よって、システムを良好に動作させるには、発声者がdata cのように発声するようにしむけ、システムは自由な発声を受けつけるSystem Aのように作成するのがよいと考

えられる。

System A～System Cの対象の複雑さの指標として、文節あたりの、average branching factor⁽⁹⁾(表3-11の発声例)、構文と単語が表わしうる可能な単語列の数、意味内容の種類数をとりあげる。System A～System Cに対する各指標の値を表3-14に示す。各システムの文節認識率を考慮すると、branching factorよりも構文と単語辞書による冗長度の方が適切であると思われる。

表3-14 各システムの複雑さ

	branching factor	冗長度	単語列の数	意味内容の数
System A	26	4.4	18151	1318
System B	9	5.6	16-12312	1-1020
System C	8	8.2	3-1368	2-1020

3.10 質問回答実験による性能評価

質問回答システムとしてのVoice Q-A System Iの性能を評価するために、認識実験と同じ8名の発声者で質問回答のオンライン実験を行った。

3.10.1 音声資料

質問回答システムの性能の評価をおこなうために、7項目全部の予約が正しく認識できるまで、質問回答を繰返す実験をおこなった。発声者は、認識実験と同じ8名で、各自40種類の予約を1回ずつ、延べ320回の予約についておこなった。質問回答の途中で入力する予約項目の言いまわしは、各発声者の任意であるが、最初に入力する文章だけは、あらかじめ定めておいた。発声条件は、認識実験と同様である。最初に入力した文章は付録3のグループ1、グループ2の音声である。

3.10.2 質問回答実験の結果

各予約の最初に入力した文章の文節認識率を、発声者別に表3-15の左から1列目の欄に示す。平均の文節認識率は、84.6%であった。

質問回答の実験では入力プロジェクトされて、同じ内容の質問回答が5回連続した場合には、その予約を打ち切ることにした。その結果、途中で打ち切った予約は、320回のうち3回で、予約完了率99.1%であった。発声者別の結果を、表3-15の左から2列目の欄に示す。質問回答の打ち切りの3例の原因は、「6枚」の拒絶（発声者SSで2例）と「浜松→豊橋」の誤り（発声者SF）であった。会話モデルの不備によって、質問回答がループに落ち込むことはなかった。

表3-15 質問回答の回数と文節認識率

予約を終了するために必要とする平均の質問回答の回数									
	文節 認識率	予約を達成 できた 回数	各予約項目の最初の発声 (a_1)	確認 (b)		拒絶によるい いなおし(a_2)	誤りによる訂 正 (a_3)	全 体 ($b+a_2+a_3$)	
				はい(b_1)	いいえ(b_2)				
発 声 者	RN	95.3	40	1.48	1.78	0.26	0.18	0.00	2.22
	HN	87.2	40	1.48	1.35	0.48	0.34	0.20	2.37
	SF	87.2	39	1.48	1.51	0.18	0.52	0.49	2.70
	KI	88.5	40	1.48	1.55	0.50	0.37	0.43	2.85
	KS	85.0	40	1.48	1.75	0.50	0.57	0.43	3.25
	SS	78.4	38	1.48	1.74	0.63	0.63	0.47	3.47
	SA	78.0	40	1.48	1.80	0.53	1.46	0.33	4.12
	MK	76.7	40	1.48	2.23	0.90	0.79	0.83	4.75
平均	84.6	317 (99.1%)	1.48	1.71	0.50	0.61	0.39	3.21	
				2.21					

この結果から、76%~96%程度の文節認識率でも、適切な会話モデルを用いることによって、質問回答を通じて、非常に高い予約完了率を達成できることがわかった。音声理解系の主要な考えの1つである質問回答の重要さが確認できた。

質問回答の内容を分析するために、途中で打ち切った3回を除く317回の予約について、質問回答の回数を調べた。質問回答の種類は、(a)予約内容に関する発声、(b)あいまいな予約項目の確認の応答文（「……………ですか？」）に対する「はい(b_1)、いいえ(b_2)」の発声、に分けられる。(a)は、さらに、各予約項目の最初の発声(a_1)、発声内容がリジェクトされたための発声のしなおし(a_2)、誤認識を訂正するための発声(a_3)、にわけられる。全体での平均の質問回答数は4.69回である。これらの回数を

表3-15に示す。4.69回のうち、1.48回は予約が完了するために最少限必要な質問回答数(a_1)である。これが1よりも大きいのは、最初に、必ずしも、全予約項目を発声しないからである。よって、文節の認識が完全でないことによって生じる質問回答の回数は、321回である。予約1回あたり、発声のしなおし(a_2)は約0.61回、誤りの訂正(a_3)は0.39回であり、確認に関しては、「はい」(b_1)が1.71回、「いいえ」(b_2)が0.50回であった。「はい」の回数が多いが、これは発声者にあまり負担がかからない。それに対して発声者に負担がかかる誤り訂正の回数は、0.39回と比較的少ない。したがって、本システムは、発声者にとって、比較的使いやすいシステムであると考えられる。

次に、予約項目別の発声回数(「はい」, 「いいえ」は除く)を表3-16に示す。発時刻と列車名は、構文が複雑で、他の項目に比べて認識がむずかしいにもかかわらず、発声の回数が少なくなっており、時刻表を用いた推論の効果が現われている。

表3-16 予約項目別の質問回答の回数

予 約 項 目	回 数
日 付	1.12
発 駅	1.19
着 駅	1.26
発 時 刻	1.08
列 車 名	0.99
等	1.16
枚 数	1.16
平 均	1.14

3.11 Voice Q-A System Iの問題点

この節では、Voice Q-A System Iでの問題点をあげ、Voice Q-A System IIを作成する際の方針について述べる。

言語処理での問題点としては、システムの構成法、認識対象の拡大、言語情報の記述法などあげられる。しかし、現在の認識対象でさらに性能を向上させ、良好に動作するシステムを作成するには、音響処理を精密化することが不可欠である。音韻認識率と文節認識率との関係のグラフ(図3-13)からみると、システムが良好に動作する目安である95%の文節認識率を達成するには、63%以上の音韻認識率を得ればよい。よって、音響処理の性能を平均で5%余り向上させればよいことになる。よって音響処理の精密化を計り、Voice Q-A System IIを作成することにした。

Voice Q-A System Iでの音響処理と言語処理での認識誤りを分析することにより、

言語情報で訂正できる誤りと、訂正が困難である誤りとが明確になった(表3-10)。音響処理では、とくに、次の4点について検討を要する。

- 文頭の母音のセグメンテーション誤り。
- 拗音、半母音 / y / のセグメンテーション誤り。
- 母音区間の検出誤り。
- 撥音 / N / のセグメンテーションと認識誤り。

これらの誤りは、すべて音韻単位へのセグメンテーションが不完全であることから生じている。また、精密なセグメンテーションを行うには、スペクトルマッチングによる母音区間の検出も不可欠である。よって、次の2点について Voice Q-A System II の音響処理で検討することにした。

- (1) 音韻単位のセグメンテーションの精密化。
- (2) スペクトル・マッチングによる母音認識の精密化。

また、Voice Q-A System I の音響処理での子音認識は、VCV音節単位の標準パターンとのマッチングによって行われている。この方法では、VCV音節を発声者ごとに学習することを基本としている。しかし、VCV音節の種類は多いので学習は容易ではない。また、母音の標準パターンも連続音声中の母音の平均を用いている。この学習も必ずしも容易ではない。よって子音認識、母音標準パターンの学習に関して、次の方針で Voice Q-A System II の音響処理の検討を行う。

- (3) 発声者によらない子音認識方法。
- (4) 母音標準パターンの登録の簡単化。

(1)~(4)の方針に基づき、Voice Q-A System II を作成する。第5章でLPC分析に基づくスペクトル・マッチング尺度について、母音認識の見地から理論的かつ実験的検討を加える。第6章ではセグメンテーションの問題に重点をおいて、Voice Q-A System II の構成および性能について述べる。

3.12 あとがき

会話音声による質問回答システムの対象として、座席予約サービスを取りあげ、その言語処理の構成および性能について述べた。言語処理の特徴を以下にまとめる。

- (1) 言語処理は、top-down 的なマッチングを基本とし、帰納的なシステム構成からなっている。
- (2) 知識をリスト表現で表わし、リストの変更によって、動作の変更や認識対象の変更が可能である。
- (3) tree search を基本とした連続音声中の単語認識アルゴリズムを開発した。このアルゴリズムは、様々な音韻の性質を考慮することができ、かつ、高速に動作する。
- (4) リスト表現を用いた構文の表現方法と構文解析の方法を開発した。
- (5) 言語処理は、実時間の約 2 倍の処理時間で動作する。音韻認識率 58.5% の音韻ラティスを入力として、86.0% の文節認識率を得た。すべての音韻の標準パターンを学習した発声者 RN に対しては、95.7% の文節認識率が得られた。
- (6) 質問回答のための会話モデルを作成し、質問回答形式による予約達成率 99.1% を得た。このときの、最初の入力を除いた平均の質問回答の回数は、3.21 回で比較的少ない回数であり、自然な質問回答のやりとりで動作するシステムが実現できた。

本言語処理は、音響処理、音声応答を接続して、音声理解系を構成する。本システムは入出力とも会話音声で、オンラインで、しかも、質問回答形式で実際に動作する。現在までのところ、このようなシステムは、他には見あたらない。これは、会話の対象の選定や目標の設定が適切であったことによるところも大きいと考えている。

第4章 音響処理の評価方法

4.1 はし が き

パターンを認識する過程では、パターンを観測し、その特徴を抽出して情報を圧縮することが必要となる。この情報の圧縮は、認識のための処理をしやすく、かつ、処理に要する計算量を減らすために不可欠である。情報圧縮して得られた特徴は、通常、パターンの基本的構成要素（プリミティブ）からなり、プリミティブの切り出し（セグメンテーション）のあいまいさと、プリミティブの認識のあいまいさを含んでいる。音声のような時系列パターンにおいても、この2種類のあいまいさが含まれている。

会話音声認識システムは、通常、音韻を認識する音響処理部と意味内容を把握する言語処理部からなる。音響処理部での音韻認識の結果は、音韻系列で表わされ、言語処理部に送られる。多くの研究機関の会話音声認識システムでは、音響処理の結果を単なる音韻系列で表わすのではなく、音韻認識のあいまいさと音韻区間の切り出しのあいまいさを表わしうる音韻系列の形式を用いている。⁽³¹⁾⁽⁴⁶⁾⁽⁷⁴⁾ 文献(46)では、Segment Latticeという語を、文献(74)では、A-Matrixという語を用いている。我々は、このようなあいまいさを表わす音韻系列を音韻ラティスと呼んでいる。音韻ラティスの例を図4-1に示す。

系列の評価は、プリミティブの認識率によって行われることが多い。しかしながら、ラティスのようにあいまいさをもつ系列を評価する際には、認識率を求めることが困難になる。認識率より普遍的な評価尺度として、情報量⁽⁸³⁾を用いることが考えられるけれども、情報量は、直観的にわかりにくい評価尺度であり、通常のない系列の結果とも比較しにくい。よって、ラティスと同じ情報量をもつ通常の系列の認識率を求め、この認識率でラティスを評価することを考えることにする。

ラティスで表わされる系列の情報量を求めるには、まず、ラティスの構成要素（プリミティブの組み合わせからなる）についての統計量を得ることが必要となる。しかし、統計的に有効な処理結果を得るには、ラティスの構成要素のパラエティが非常に大きいので、膨大な処理結果が必要とされ、通常は不可能になる。このことは20個のプリミティブからなる入力で、5個のプリミティブのあいまいさをもつ場合の出力のパラエティが $15504 (= {}_{20}C_5)$ となることから容易に想像できる。この問題をさけるために、出力結果から、

ある程度のマクロな統計量を取り、その統計量内での出力のばらつき具合（エントロピー）を仮定することにより、ラティスのもつ情報量を推定する方法をとることにする（図4-2）。

このマクロな統計量も、人が入力と出力結果を対応させて求めるには、多くの労力がかかる。よって、計算機に、正しい系列と音韻ラティスとの対応を自動的にとらせて、統計量を求めることが望まれる。

本節は、以下の6つの節からなる。4.2節では、音韻ラティスからの音韻認識率の推定の考え方と、推定のためのモデルについて説明する。4.3節では、音韻ラティスの情報量の推定について詳しく述べる。とくに、情報量を推定するときの問題となる統計量内でのエントロピーのあいまいさについて詳しく検討する。4.4節では、正しい系列と音韻ラティスとの対応を取り、統計量を求めるアルゴリズムについて述べる。アルゴリズムは、tree searchのbest-first法⁽⁷⁵⁾で、相互情報量を評価関数として、実行される。さらに、

1	2	3	4	5	6	7	8	9	10
P	I	T	A	*	E	P	I	U	O
Z		P			NN	T		NN	
S		K				S		D	
H		S				H		Z	
K		KY				K		N	
				P				M	
				T				B	
				S					
				ZY					

図4-1 音韻ラティスの例（「TUITATINO」と発声した音声）

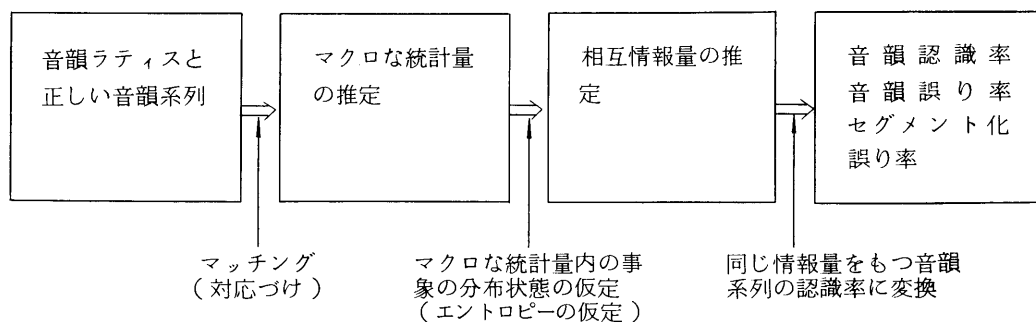


図4-2 音韻ラティスの評価の考え方

その対応づけのアルゴリズムから得られたマクロな統計量から、音韻認識率を推定する方法も定式化される。4.5節では、4.4節の自動推定アルゴリズムの評価として、実際に人間が対応づけをしてとったマクロな統計量との比較がおこなわれる。4.6節では、8人の発声者の音声を資料として、音響処理部の出力の音韻ラティスを評価した例を示す。4.7節では、全体のまとめと、今後の問題点について述べる。

4.2 音韻ラティスの音韻認識率の推定モデル

4.2.1 考え方と問題点

音韻ラティスの評価は、発声した音韻系列が、音韻ラティスからどの程度推定できるかによっておこなわれるのが望ましい。この意味において、音韻系列を入力として、出力を音韻ラティスと考えた場合の相互情報量⁽⁷⁶⁾を、ラティスの評価尺度として用いることがよいと考えられる。

モデルとして、図4-3のような通信路を考える。Xは入力空間で、発声した音声の中の音韻に対応し、正しい音韻系列が入力されると考える。Yは出力空間で、音韻ラティスの構成要素からなる。通信路は音響処理部⁽³²⁾に対応する。入力空間Xの入力事象をxで、出力空間Yの出力事象をyで表わす。相互情報量は、次式で定義される⁽⁷⁶⁾。

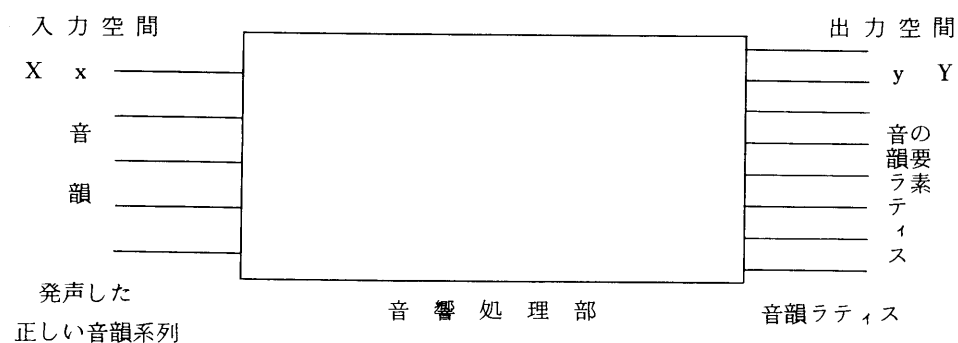


図4-3 音韻ラティスを出力する通信路

$$I(X; Y) = H(X) - H(X|Y) \quad (4-1)$$

$$= H(Y) - H(Y|X) \quad (4-2)$$

ここで、

$$H(Y|X) = - \sum_{x \in X} \sum_{y \in Y} P\{x, y\} \log P\{y|x\} \quad (4-4)$$

である。上式からわかるように、相互情報量 $I(X;Y)$ は、入力と出力の対応づけから得られる推移確率と、入力事象の出現確率がわかれば、計算することができる。

しかし、相互情報量は、評価尺度として用いる場合に、次の3点で問題がある。

(問題点Ⅰ) 直観に訴えにくいこと。

通常、音韻系列の評価には認識率が用いられている。これは、認識率が、直観に訴えやすく、具体的なイメージを人に与えやすいためと考えられる。一方、相互情報量は、その値のとり得る範囲が、入力事象の数や出現確率によって異なり、直観に訴えにくい。よって、単にラティスの相互情報量を示すだけでなく、それと同じ情報量をもつ音韻系列の認識率も示す必要があると考えられる。そのためには、相互情報量と認識率とを関係づけることが必要である。

(問題点Ⅱ) 相互情報量を厳密に求めることが困難であること。

ラティスの相互情報量を求めるには、入力事象と出力事象との間の推移行列を得るために統計量を求めねばならない。ラティスでは、構成要素の種類が多くなり、個々の統計量を求めることが、實際上、不可能になる。実用上は、入出力事象のマクロな統計量を取り、その統計量内の入出力事象のばらつき具合(エントロピー)の値を仮定することによって、相互情報量を推定するという手法をとらざるを得ない。このエントロピーの値の仮定をいかにするかが問題となる。

(問題点Ⅲ) 出力から入力を推定する際に、入出力事象の統計量を完全に利用することが困難であること。

言語情報を利用して、音韻ラティスから入力の音韻系列を推定するアルゴリズムで、入出力事象に関する統計量をすべて利用しているとは限らない。また、余りにも細かい統計量まで利用するアルゴリズムは、實際上、計算量の面で問題が生ずる。このようなことから、言語処理部⁽³¹⁾で利用可能な統計量だけから求まる評価尺度が望ましい。

以下、問題点Ⅰ～問題点Ⅲについて検討を加える。

通常のあいまいさのない音韻系列を出力し、かつ、セグメンテーションの誤りがない系を考えよう。簡単のため、入力の音韻の出現確率は等しく、かつ、次のような推移行列をもつ一様な通信系(音響処理部)を考える。

$$\begin{pmatrix} p & r_1 & r_2 & \cdots & r_{k-1} \\ r_{k-1} & p & r_1 & \cdots & r_{k-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & r_2 & r_3 & \cdots & p \end{pmatrix}$$

k ; 音韻数 (カテゴリ数)

p ; 認識率

$$r = \sum_{i=1}^{k-1} r_i ; \text{誤り率}$$

このような系の相互情報量は、次のようになる。

$$I(X; Y) = \log k + p \log p + \sum_{i=1}^{k-1} r_i \log r_i \quad (4-5)$$

この式から、相互情報量 $I(X; Y)$ は、認識率 p と誤り率 r のみによって定めることはできない。⁽⁷⁷⁾⁽⁷⁸⁾ 相互情報量を求めるには、誤り r 内の分布 $(r_1, r_2, \dots, r_{k-1})$ の状態も必要となる。(4-5)式は、次のように変形できる。

$$I(X; Y) = \log k + p \log p + r \log r - r \left\{ - \sum_{i=1}^{k-1} \left(\frac{r_i}{r} \log \frac{r_i}{r} \right) \right\} \quad (4-6)$$

$$= \log k + p \log p + r \log r - r H \left(\frac{r_1}{r}, \dots, \frac{r_{k-1}}{r} \right) \quad (4-7)$$

ここで、 $H \left(\frac{r_1}{r}, \dots, \frac{r_{k-1}}{r} \right)$ は誤りの分布のエントロピーであり、これを簡単に

H_r で表わす。よって、相互情報量と認識率は、 H_r の値を仮定すれば、一意の関係で結ばれる。ここで、 H_r の最大値は、

$$\frac{r_1}{r} = \frac{r_2}{r} = \dots = \frac{r_{k-1}}{r} = \frac{1}{k-1} \quad \text{のときで、}$$

$$\max H_r = \log(k-1) \quad (4-8)$$

であり、最小値は、

$$\frac{r_i}{r} = 1 \quad \text{and} \quad \frac{r_j}{r} = 0 \quad (j=1, \dots, k-1, j \neq i) \quad \text{のときで、}$$

$$\min H_r = 0 \quad (4-9)$$

である。よって、相互情報量の最大値、最小値は、

$$\max I(X;Y) = \log k + p \log p + r \log r \quad (4-10)$$

$$\min I(X;Y) = \log k + p \log p + r \log r - r \log(k-1) \quad (4-11)$$

となる。この相互情報量と認識率の関係を図4-4に示す。この例からわかるように、「相互情報量と認識率とは一意の関係では結ばれていない。⁽⁷⁷⁾⁽⁷⁸⁾一意の関係をつけるには、誤り内でのエントロピーの値の仮定が必要である」。

問題点Ⅱについても同じ通信系で説明する。

マクロ統計量を p と r とすると、(4-7)式で相互情報量を求めるには、統計量 r 内の事象のばらつき具合を表わすエントロピー H_r の値の仮定が必要となる。よって、次の事柄がいえ。

「音韻ラティスの相互情報量を推定するには、マクロな統計量内の事象のばらつき具合を表わすエントロピーの値の仮定が必要である」

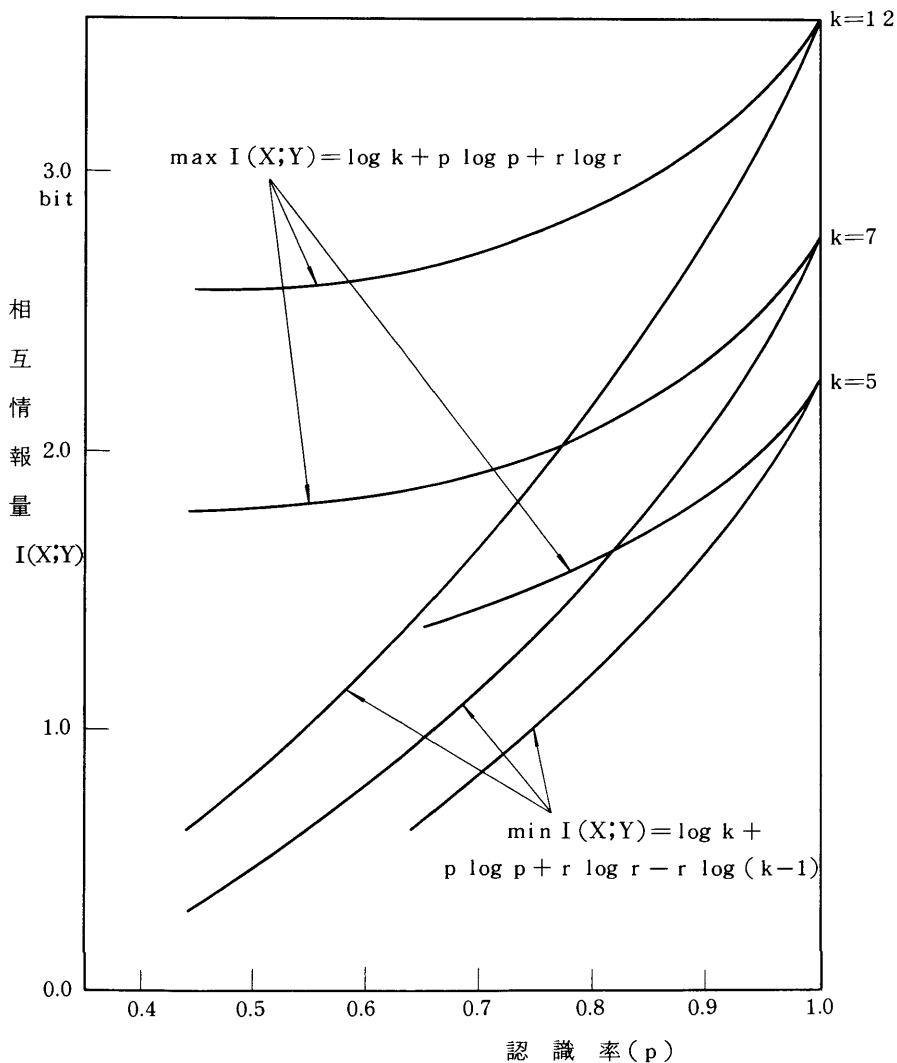


図4-4 相互情報量と認識率の関係(正解-誤り系)

問題点Ⅰ，問題点Ⅱの問題は，エントロピーの値を仮定すれば，解決できそうである。しかし，マクロな統計量内での分布のかたよりを仮定して相互情報量を求めても，実際に入力の音韻系列を推定する際に，マクロな統計量内でのかたよりまでも利用して入力の音韻系列を推定できるか否かは疑問である。この問題Ⅲについて，例をあげて説明を加える。

例として，図4-5のような系を考える。この系は， n 個の通信路を縦続接続したものである。 i 段目の出力現象 Y_i は第 $(n-i+1)$ 位までの音韻の候補であり，音韻数を k とすると，事象の数は ${}_k C_{n-i+1} \cdot (n-i+1)!$ 個である。 i 段目では， $(i-1)$ 段目の出力を入力とし，第 $(n-i+1)$ 位までの候補を出力する通信路である(図4-5)。最後の n 段目では，第1位の音韻だけが出力される。この系で，入力 X と i 段目の出力事象 Y_i との相互情報量 $I(X; Y_i)$ について検討する。縦続接続された通信路では，次のことが知られている。

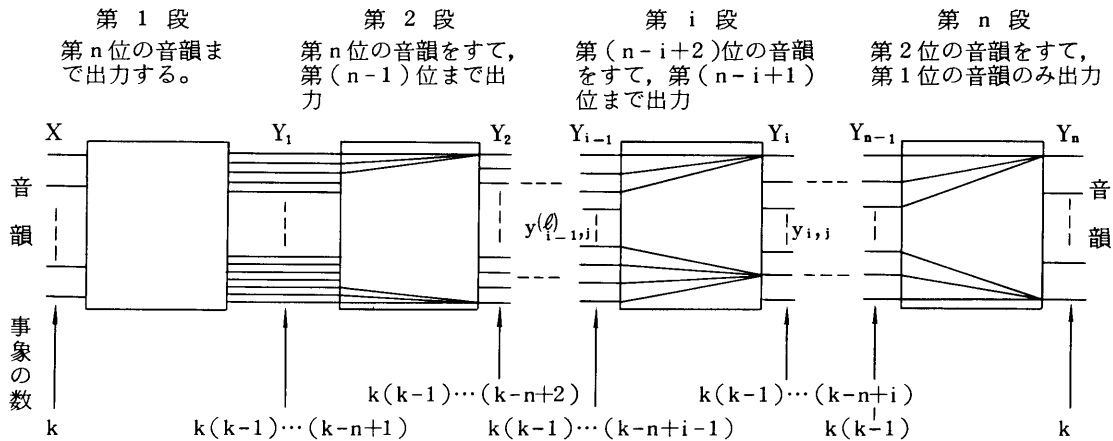


図4-5 縦続接続された認識系(音韻の尤度順を出力する系)

「相互情報量は，通信路を経由することによって，決して増大せず，一般に減少する」⁽⁷⁹⁾

よって，この系でも，

$$I(X; Y_1) \geq I(X; Y_2) \geq \dots \geq I(X; Y_{n-1}) \geq I(X; Y_n)$$

なる関係が成立する。(4-12)式で等号が成立するときは，(4-1)式より，各々，

$$H(X | Y_{i-1}) = H(X | Y_i) \quad (i=2, \dots, n) \quad (4-13)$$

が成立するときである。図4-5に示されるように， i 段目では， Y_{i-1} のうち

$(n-i+1)$ 番目まで同じ音韻列 $(y_{i-1, j}^{(l)}, l=1, 2, \dots, k-n+i-2)$ をまとめて，

$y_{i,j}$ として出力する働きをしている。よって、等号が成立する条件を次のように表わすことができる。

$$P \{ x | y_{i-1,j}^{(i)} \} = P \{ x | y_{i,j} \} \quad \begin{array}{l} l = 1, 2, \dots, k-n+i-2 \\ j = 1, 2, \dots, k(k-1) \dots (k-n+i) \end{array} \quad (4-14)$$

この(4-13)式,(4-14)式は、次のようにいうことができる。

「 Y_i からみたあいまい度 $H(X | Y_i)$ と Y_{i+1} からみたあいまい度 $H(X | Y_{i+1})$ が等しいことだけを要求しており、 $(i+1)$ 段目の通信路でロスが生じないことと雑音がないことを必要とはしていない」

実際には、上の条件はかなり厳しく、通常は、等号は成立しない。よって、(4-12)式は、相互情報量という評価尺度でみる限り、音響処理部では、出力事象の多い状態 (Y_1) で言語処理部に出力することがよいということを意味する。しかし、現実には、このような余りにも冗長な情報を利用することはできない。よって、

「すべての段で、(4-13)(4-14)式に近くなるように、統計量内でのエントロピーの値を仮定することが望まれる」。

以上の結果をまとめると次の2つになる。

- ① 統計量内でのエントロピーの値を仮定することによって、相互情報量を求めることができるし、かつ、認識率と相互情報量との関係を一意にすることができる。
- ② 単に出力事象の数を増加させて、相互情報量が増えるような評価方法はとるべきでない。

4.2.2 モデル

4.2.1 節の検討にもとづき、音韻ラティスの評価モデルを示す。評価モデルは、図4-6に示す3段階からなる。以下、各段階について簡単に説明する。

(第1段階) マクロな統計量の抽出

入力の正しい音韻系列と音響処理部の出力の音韻ラティスとをマッチングする。マッチングを抽出されたマクロな統計量から推定した相互情報量が最大になるようにおこなう。マッチングでの両者の対応関係からマクロな統計量をとる。セグメンテーションの誤りに対しては、それに対応する1つの統計量として定義する。マクロな統計量を多くの音韻系列と音韻ラティスに対して求め、それを集計し、合計のマクロな統計

量を第2段階への入力とする。詳しいアルゴリズムについては、4.4.1節で述べる。

(第2段階) 相互情報量の計算

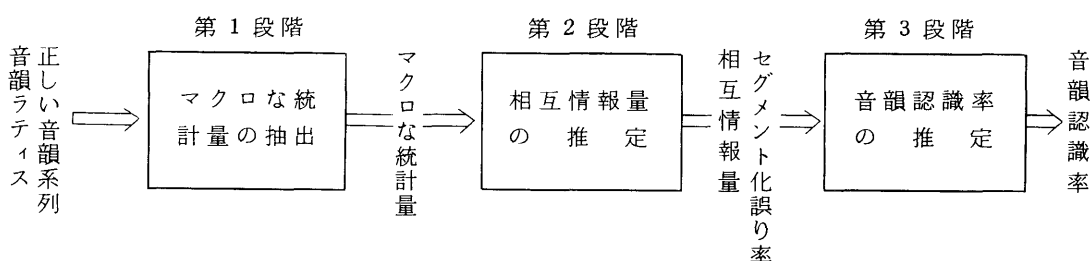


図4-6 音韻ラティスの評価モデルの概念図

第1段階からのマクロな統計量から相互情報量を推定する。そのためには、マクロな統計量内でのエントロピーの値を仮定しなければならない。詳しいアルゴリズムについては、4.3.2節および4.4.2節で述べる。

(第3段階) 認識率の推定

第2段階で推定された相互情報量と同じ相互情報量をもつ音韻系列の認識率をもとめる。詳しくは、4.3.1節および4.4.3節で述べる。

4.3 音韻ラティスの情報量

4.3.1 音韻系列の情報量と認識率の関係

この節では、音韻系列の認識率、誤り率と相互情報量との関係を導出する。この関係は、音韻ラティスの評価モデルの第3段階で利用され、ラティスの相互情報量から認識率を推定するのに用いられる。

正しい入力の音韻系列を、

$$\mathbf{x} = x_1 x_2 \cdots x_n$$

とする。このときの出力の音韻系列を、

$$\mathbf{y} = y_1 y_2 \cdots y_m$$

とする。通信系は入力 \mathbf{x} の要素に同期がとれているとみなし、系列 \mathbf{x} の要素 x_i が入力されたときの出力を観測するとする。このときの出力の type を次の4種類に分類する。

[type I] $x_i \rightarrow y_j$ ($x_i = y_j$) 正解

正しくセグメンテーションされ、かつ、正しく認識された場合。

[type II] $x_i \rightarrow y_j$ ($x_i \neq y_j$) 認識誤り

正しくセグメンテーションされているが、認識が誤っている場合。

[type III] $x_i \rightarrow y_j y_{j+1} \dots$ セグメンテーション・ミス(I)

細かくセグメンテーションされすぎている場合。

[type IV] $x_i \rightarrow \phi$ セグメンテーション・ミス(II)

x_i が検出されずに見落されている場合。

各々の音韻について、type I ~ type IV までの統計量を取り、通信路の推移行列を求める。

$$\begin{array}{c}
 X \\
 \\
 x_1 ; p(x_1) \\
 x_2 ; p(x_2) \\
 \vdots \\
 x_k ; p(x_k)
 \end{array}
 \begin{array}{c}
 Y ; y_1 \quad y_2 \quad \dots \quad y_k \\
 \\
 p(y_1) \quad p(y_2) \quad \dots \quad p(y_k) \\
 \left(\begin{array}{cccc}
 p_1 & r_{12} & \dots & r_{1k} \\
 r_{21} & p_2 & & r_{2k} \\
 \vdots & \vdots & \ddots & \vdots \\
 r_{k1} & r_{k2} & \dots & p_k
 \end{array} \right)
 \begin{array}{cc}
 * & ** \\
 u & v \\
 u_1 & v_1 \\
 u_2 & v_2 \\
 \vdots & \vdots \\
 u_k & v_k
 \end{array}
 \end{array}$$

- * セグメンテーション・ミス(I)
- ** セグメンテーション・ミス(II)

このときの相互情報量 $I(X; Y)$ を計算する。

$$H(Y) = - \sum_{y \in Y} p(y) \log p(y) \tag{4-15}$$

$$= - \sum_{i=1}^k p(y_i) \log p(y_i) - u \log u - v \log v \tag{4-16}$$

$$\begin{aligned}
 &= - (1 - u - v) \log (1 - u - v) + (1 - u - v) H_Y \\
 &\quad - u \log u - v \log v
 \end{aligned} \tag{4-17}$$

ただし、

$$u = \sum_{j=1}^k p(x_j) u_j \tag{4-18}$$

$$v = \sum_{j=1}^k p(x_j) v_j \tag{4-19}$$

$$H_Y = - \sum_{i=1}^k \frac{p(y_i)}{1-u-v} \log \frac{p(y_i)}{1-u-v} \quad (4-20)$$

である。

$$H(Y|X) = - \sum_{x \in X} \sum_{y \in Y} p(x) p(y|x) \log p(y|x) \quad (4-21)$$

$$= - \sum_{i=1}^k p(x_i) \left\{ p_i \log p_i + \sum_{\substack{j=1 \\ j \neq i}}^k r_{ij} \log r_{ij} + u_i \log u_i + v_i \log v_i \right\} \quad (4-22)$$

$$= - \sum_{i=1}^k p(x_i) p_i \log p_i - \sum_{i=1}^k p(x_i) r_i \log r_i + \sum_{i=1}^k p(x_i) \cdot r_i \cdot H_{r_i} \\ - \sum_{i=1}^k p(x_i) u_i \log u_i - \sum_{i=1}^k p(x_i) v_i \log v_i \quad (4-23)$$

$$= - p \log p + p H_p - r \log r + r H_r + r \overline{H_{r_i}} - u \log u + u H_u \\ - v \log v + v H_v - H_X \quad (4-24)$$

ただし,

$$p = \sum_{i=1}^k p(x_i) p_i \quad (4-25)$$

$$r = \sum_{i=1}^k p(x_i) r_i \quad (4-26)$$

$$r_i = \sum_{\substack{j=1 \\ j \neq i}}^k r_{ij} \quad (4-27)$$

$$H_{r_i} = - \sum_{\substack{j=1 \\ j \neq i}}^k \frac{r_{ij}}{r_i} \log \frac{r_{ij}}{r_i} \quad (4-28)$$

$$H_p = - \sum_{i=1}^k \frac{p(x_i) \cdot p_i}{p} \log \frac{p(x_i) \cdot p_i}{p} \quad (4-29)$$

$$H_r = - \sum_{i=1}^k \frac{p(x_i) \cdot r_i}{r} \log \frac{p(x_i) \cdot r_i}{r} \quad (4-30)$$

$$H_u = - \sum_{i=1}^k \frac{p(x_i) \cdot u_i}{u} \log \frac{p(x_i) \cdot u_i}{u} \quad (4-31)$$

$$H_v = - \sum_{i=1}^k \frac{p(x_i) \cdot v_i}{v} \log \frac{p(x_i) \cdot v_i}{v} \quad (4-32)$$

$$\overline{H_{r_i}} = \sum_{i=1}^k \frac{p(x_i) \cdot r_i}{r} H_{r_i} \quad (4-33)$$

$$H_X = - \sum_{i=1}^k p(x_i) \log p(x_i) \quad (4-34)$$

よって、(4-2)式より、

$$I(X; Y) = -(1-u-v) \log(1-u-v) + (1-u-v) H_Y + H_X + p \log p + r \log r - pH_p - r(H_r + \overline{H_{ri}}) - uH_u - vH_v \quad (4-35)$$

ここで、

H_Y ……出力Y(カテゴリ1~k)の確率によるエントロピー

H_X ……入力X(カテゴリ1~k)の確率によるエントロピー

H_p ……カテゴリ間の認識率のエントロピー

H_r ……カテゴリ間の誤り率のエントロピー

$\overline{H_{ri}}$ ……カテゴリ内の誤りのエントロピーの平均

H_u ……カテゴリ間の type III のエントロピー

H_v ……カテゴリ間の type IV のエントロピー

ということができる。

(4-35)式に示されるように、相互情報量は、認識率 p 、認識誤り率 r 、セグメンテーション誤り率 u および v と、エントロピーの値 H_Y 、 H_X 、 H_p 、 H_r 、 $\overline{H_{ri}}$ 、 H_u 、 H_v から求めることができる。入力信号を音韻に限れば、これらのエントロピーの値は、ほぼ、一定の値をとるものと考えられる。

よって相互情報量とセグメンテーション誤り u と v がわかれば、

$$p + r + u + v = 1 \quad (4-36)$$

の関係があるので、(4-35)式から、エントロピーの値を仮定することによって、認識率 p と誤り率 r が求められる。

各各のエントロピーの値は、その中の要素(事象)の数を k とすると、

$$0 \leq H \leq \log k \quad (4-37)$$

の間の値をとる。便宜上、エントロピーの値を、

$$H = a_H \log k \quad 0 \leq a_H \leq 1 \quad (4-38)$$

で表わすことにする。 a_H は、要素間のばらつき具合を表わす指標と解釈できる。

a_H が 0 のときは、要素の 1 つだけが生起することを意味し、 a_H が 1 のときは、全要素が等確率 ($1/k$) で生起することを意味している。

4.2.1 節の問題点 III で述べたように、言語処理部で利用しない統計量のかたよりにまで利用して相互情報量を求めることには問題がある。よって、マクロな統計量 p 、 r 、 u 、 v 内での統計的なかたよりはないとし、かつ、入力音韻生起確率も等しいとす

る。すなわち、すべてのエントロピーの指標 a_H を 1 と仮定すると、(4-35)式は、次のようになる。

$$\begin{aligned}
 I(X; Y)_{a_H=1} &= -(1-u-v) \log(1-u-v) + (1-u-v) \log k \\
 &+ \log k + p \log p + r \log r - p \log k - r(\log k + \log(k-1)) - u \log k - v \log k \\
 &= (1-u-v) \log k - (1-u-v) \log(1-u-v) + p \log p \\
 &+ r \log r - r \log(k-1) \qquad (4-39)
 \end{aligned}$$

我々は、この(4-39)式をもって、相互情報量と認識率の変換の式とした。

この(4-39)式は、4.2節の問題点 I～問題点 III を満す最善の解決方法である保証はないが、現在のところ、エントロピーの指標 a_H を 1 とすることが、もっとも妥当と思われる。これの厳密な検討は今後の課題である。

4.3.2 音韻ラティスの情報量

この節では、音韻ラティスからとり出されたマクロな統計量からの相互情報量の推定方法について述べる。この推定方法は、音韻ラティスの評価モデルの第2段階で利用される。また、この節で導出された相互情報量の推定の式は、評価モデルの第1段階でのマッチングの尺度としても利用される。

入力の音韻系列を \mathbf{x} 、出力の音韻ラティスを \mathbf{y} とする。出力 \mathbf{y} は、認識のあいまいさを表わす要素からなり、また、セグメンテーションのあいまいさも含んでいる。このラティスにおいて、認識のあいまいさが 1 個あるときには、 x_i という入力に対して、

$$(y_j^1 \quad y_j^2 \quad \dots \quad y_j^l)$$

という 1 個の音韻からなるラティスの要素が出力される。セグメンテーションのあいまいさがあるときには、各各のセグメント出力が等確率で生ずるものとする。通信系は入力に同期がとれているとし、音韻系列 \mathbf{x} の要素 x_i が入力されたときの出力を観測する。このとき観測される出力の type を次のように分類する。

[セグメンテーションが正しい場合]

音韻認識のあいまいさ	正	誤
(出力された音韻数)	(入力の音韻が含まれている)	(入力の音韻が含まれていない)
1	p_1	r_1
2	p_2	r_2
⋮	⋮	⋮
1	p_1	r_1

[セグメンテーションの誤りがある場合]

$$\begin{array}{l} x_i \rightarrow y_j \quad y_{j+1} \cdots \cdots \quad u \\ x_i \rightarrow \phi \quad \quad \quad \quad \quad \quad v \end{array}$$

これらのマクロな統計量 $p_1, \dots, p_l, r_1, \dots, r_l, u, v$ は、音韻ラティスの評価モデルの第1段階でのマッチングの結果から得られる。以下、これらのマクロな統計量と相互情報量の関係を導出する。

入力の音韻系列と出力の音韻ラティスとの推移行列は、次のように表わされる。

$$\left(\begin{array}{cccccccc} \underbrace{p_{11,k-1}C_0}_{p_{111}} & \underbrace{r_{11,k-1}C_1}_{r_{111} \cdots \cdots} & \underbrace{p_{21,k-1}C_1}_{p_{211} \cdots \cdots} & \underbrace{r_{21,k-1}C_2}_{r_{211} \cdots \cdots} & \cdots & \underbrace{p_{11,k-1}C_{l-1}}_{p_{111} \cdots \cdots} & \underbrace{r_{11,k-1}C_l}_{r_{111} \cdots \cdots} & u_1 \quad v_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \underbrace{p_{1k1}} & \underbrace{r_{1k1} \cdots \cdots}_{r_{1k1} \cdots \cdots} & \underbrace{p_{2k1} \cdots \cdots}_{p_{2k1} \cdots \cdots} & \underbrace{r_{2k1} \cdots \cdots}_{r_{2k1} \cdots \cdots} & \cdots & \underbrace{p_{1k1} \cdots \cdots}_{p_{1k1} \cdots \cdots} & \underbrace{r_{1k1} \cdots \cdots}_{r_{1k1} \cdots \cdots} & u_k \quad v_k \end{array} \right)$$

すると、

$$\begin{aligned} H(Y) = & \sum_{i=1}^l \{ -(p_i + r_i) \log(p_i + r_i) + (p_i + r_i) H_{Y_i} \} \\ & - u \log u - v \log v \end{aligned} \quad (4-40)$$

ただし、

$$u = \sum_{j=1}^k p(x_j) u_j \quad (4-18)$$

$$v = \sum_{j=1}^k p(x_j) \cdot v_j \quad (4-19)$$

$$H_{Y_i} = - \sum_{j=1}^{kC_i} \frac{p(y_{ij})}{p_i + r_i} \log \frac{p(y_{ij})}{p_i + r_i} \quad (4-41)$$

y_{ij} ; 認識のあいまいさ i で j 番目の出力事象

$$\begin{aligned} H(Y|X) = & \sum_{i=1}^l (- p_i \log p_i - r_i \log r_i + p_i H_{p_i} + r_i H_{r_i} \\ & + p_i \overline{H_{p_{ij}}} + r_i \overline{H_{r_{ij}}}) - u \log u + u H_u - v \log v + v H_v - H_X \end{aligned} \quad (4-42)$$

ただし、

$$H_{p_i} = - \sum_{j=1}^k \frac{p(x_j) \cdot p_{ij}}{p_i} \log \frac{p(x_j) \cdot p_{ij}}{p_i} \quad (4-43)$$

$$H_{r_i} = - \sum_{j=1}^k \frac{p(x_j) \cdot r_{ij}}{r_i} \log \frac{p(x_j) \cdot r_{ij}}{r_i} \quad (4-44)$$

$$H_{p_{ij}} = - \sum_{m=1}^{k-1} \frac{p_{ijm}}{p_{ij}} \log \frac{p_{ijm}}{p_{ij}} \quad (4-45)$$

$$H_{r_{ij}} = - \sum_{m=1}^{k-1} \frac{r_{ijm}}{r_{ij}} \log \frac{r_{ijm}}{r_{ij}} \quad (4-46)$$

$$\overline{H}_{p_{ij}} = \sum_{j=1}^k \frac{p(x_j) \cdot p_{ij}}{p_i} H_{p_{ij}} \quad (4-47)$$

$$\overline{H}_{r_{ij}} = \sum_{j=1}^k \frac{p(x_j) \cdot r_{ij}}{r_i} H_{r_{ij}} \quad (4-48)$$

$$H_X = - \sum_{i=1}^k p(x_i) \log p(x_i) \quad (4-34)$$

$$H_u = - \sum_{i=1}^k \frac{p(x_i) \cdot u_i}{u} \log \frac{p(x_i) \cdot u_i}{u} \quad (4-31)$$

$$H_v = - \sum_{i=1}^k \frac{p(x_i) \cdot v_i}{v} \cdot \log \frac{p(x_i) \cdot v_i}{v} \quad (4-32)$$

$$p_i = \sum_{j=1}^k p(x_j) \cdot p_{ij} \quad (4-49)$$

$$r_i = \sum_{j=1}^k p(x_j) \cdot r_{ij} \quad (4-50)$$

$$p_{ij} = \sum_{m=1}^{k-1} p_{ijm} \quad (4-51)$$

$$r_{ij} = \sum_{m=1}^{k-1} r_{ijm} \quad (4-52)$$

よって、 $H(Y)$, $H(Y|X)$ を (4-2) 式に代入して、相互情報量 $I(X; Y)$ は次のようになる。

$$\begin{aligned} I(X; Y) &= \sum_{i=1}^k \left(- (p_i + r_i) \log (p_i + r_i) + (p_i + r_i) H_{Y_i} \right. \\ &\quad \left. + p_i \log p_i + r_i \log r_i - p_i (H_{p_i} + \overline{H}_{p_{ij}}) - r_i (H_{r_i} + \overline{H}_{r_{ij}}) \right) \\ &\quad + H_X - u H_u - v H_v \end{aligned} \quad (4-53)$$

よって、エントロピーの値 H_{Y_i} , H_{p_i} , $\overline{H_{p_{ij}}}$, H_{r_i} , $\overline{H_{r_{ij}}}$, H_X , H_u , H_v の値を定めれば、相互情報量 $I(X; Y)$ はマクロな統計量 p_1, \dots, p_l , r_1, \dots, r_l , u , v から求められる。ここでも、4.2節の問題点 I ~ III を考慮して、エントロピーの指標 a_H をすべて 1 とする。すると、

$$\begin{aligned}
 I(X; Y)_{a_H=1} &= \sum_{i=1}^l \left(-(p_i + r_i) \log(p_i + r_i) + (p_i + r_i) \log_k C_i \right. \\
 &\quad \left. + p_i \log p_i + r_i \log r_i - p_i (\log k + \log_{k-1} C_{i-1}) - r_i (\log k + \log_{k-1} C_i) \right) \\
 &\quad + \log k - u \log k - v \log k \\
 &= (1 - u - v) \log k - \sum_{i=1}^l \left\{ (p_i + r_i) \log(p_i + r_i) - p_i \log p_i \right. \\
 &\quad \left. - r_i \log r_i + p_i \log i + r_i \log(k - i) \right\} \tag{4-54}
 \end{aligned}$$

が得られる。(4-54)式は、入力音韻の生起確率が等しく、かつ、マクロな統計量が与えられたときの相互情報量の最小値を表わしている。

我々は、(4-54)式で、音韻ラティスの相互情報量を推定することにした。

4.4 音韻認識率の自動推定アルゴリズム

4.4.1 マクロな統計量の自動抽出

正しい音韻系列と音韻ラティスとのマッチングをおこない、その対応関係からマクロな統計量を抽出する。マッチングは、抽出されたマクロな統計量から推定した相互情報量が最大になるようにおこなわれる。マッチングの際に利用できる変形規則の型を表4-1に示す。

母音と子音との出現頻度は著しく異なり、かつ、音響処理部⁽³²⁾での認識方法も異なっている。よって、マクロな統計量は、母音と子音とにわけてとることにした。さらに、音響処理部ではVCV音節(Vowel-Consonant-Vowel syllable)を基本とした認識方法をとっているため、VVの音韻列も、その間に仮想的な子音*を挿入して、V*Vとしている。このため、母音と子音の出現頻度は等しくなっている。

マッチングの際に、マッチングの程度を表わす評価尺度として、(4-55)式で表わされる相互情報量を用いる。(4-55)式は、母音、子音ごとに相互情報量を、

表 4-1 マッチングで利用できる変形規則

		母音 (xが母音)		子音 (xが子音)		
		対応づけの型	規則の名称 (マクロな統計量)	対応づけの型	規則の名称 (マクロな統計量)	
規則 I	セグメンテーション 正しい音韻が含まれている。 $x \rightarrow y$ $x \in y$	V → V	P _{v1}		P _{c1}	
規則 II	セグメンテーション誤りなし 正しい音韻が含まれていない。 $x \rightarrow y$ $x \notin y$		E _{v1}		E _{c1}	
規則 III	セグメンテーションミス(I)	V → VCV	V _{v1}	V _v	C → CVC	V _c
		V → VV	V _{v2}			
規則 IV	セグメンテーション誤り セグメンテーションミス(II)	V → φ	U _v		C → φ	U _c

ここで、x は正しい音韻系列中の音韻を、y は音韻ラティス中の要素 (y_1, y_2, \dots, y_l) を表わす。V は母音のセグメントを、C は子音のセグメントを表わす。例えば、V → VCV は音韻系列中のセグメント V を音韻ラティス中の 3 個のセグメント VCV (母音、子音、母音の連なりからなる) に対応づけることを意味する。φ は、セグメントが存在しないことを表わす。添字の v は母音を、c は子音を意味し、l はラティス中の要素 $y = (y_1, y_2, \dots, y_l)$ の中の音韻の個数を表わす。本システムでは、子音の l は 1 ~ 5 個、母音の l は 1 個あるいは 2 個である。

(4-54) 式から求めたものである。

$$I(X; Y) = \frac{1}{2} I_v(X; Y) + \frac{1}{2} I_c(X; Y) \quad (4-55)$$

ここで、

$$I_v(X; Y) = (1 - U'_v - V'_v) \log K_v - \sum_{i=1}^2 \{ (P'_{vi} + E'_{vi}) \cdot \log (P'_{vi} + E'_{vi}) - P'_{vi} \log P'_{vi} - E'_{vi} \log E'_{vi} + P'_{vi} \log i + E'_{vi} \cdot \log (K_v - i) \} \quad (4-56)$$

$$I_c(X; Y) = (1 - U'_c - V'_c) \log K_c - \sum_{i=1}^5 \{ (P'_{ci} + E'_{ci}) \cdot \log (P'_{ci} + E'_{ci}) - P'_{ci} \log P'_{ci} - E'_{ci} \log E'_{ci} + P'_{ci} \log i + E'_{ci} \cdot \log (K_c - i) \} \quad (4-57)$$

また、 $U'_v, V'_v, P'_{vi}, E'_{vi}$ は母音内での U_v, V_v, P_{vi}, E_{vi} に対応するマクロナ統計量の確率を表わし、 $U'_c, V'_c, P'_{ci}, E'_{ci}$ は子音内での U_c, V_c, P_{ci}, E_{ci} に対応するマクロナ統計量の確率を表わす。 K_v は母音の数（本システムでは、VCVの考えから撥音の /N/ も母音と考えるので $K_v=6$ ）を、 K_c は子音の数（ $K_c=20$ ）を表わす。

また、マッチングで用いる変形規則は、おおよそ、マクロナ統計量に対応している。以上より、マクロナ統計量を求める問題は、次のようにいうことができる。

〔問題〕 マクロナ統計量を求める問題

表 4-1 の変形規則を利用して、音韻系列と音韻ラティスを、(4-55)式で定義される相互情報量が最大になるように、マッチングする手法を求める。

この問題を、tree search の best-first method⁽⁷⁵⁾ を用いたマッチングで解くアルゴリズムを考える。入力の正しい音韻系列（以下、系列と略す）を、図 4-7 のようにリスト表現で表わす。

(T	U	*	I	T	A	T	I	N	O)
	1	2	3	4	5	6	7	8	9	10	
				↑							
				系列の位置 X_i							

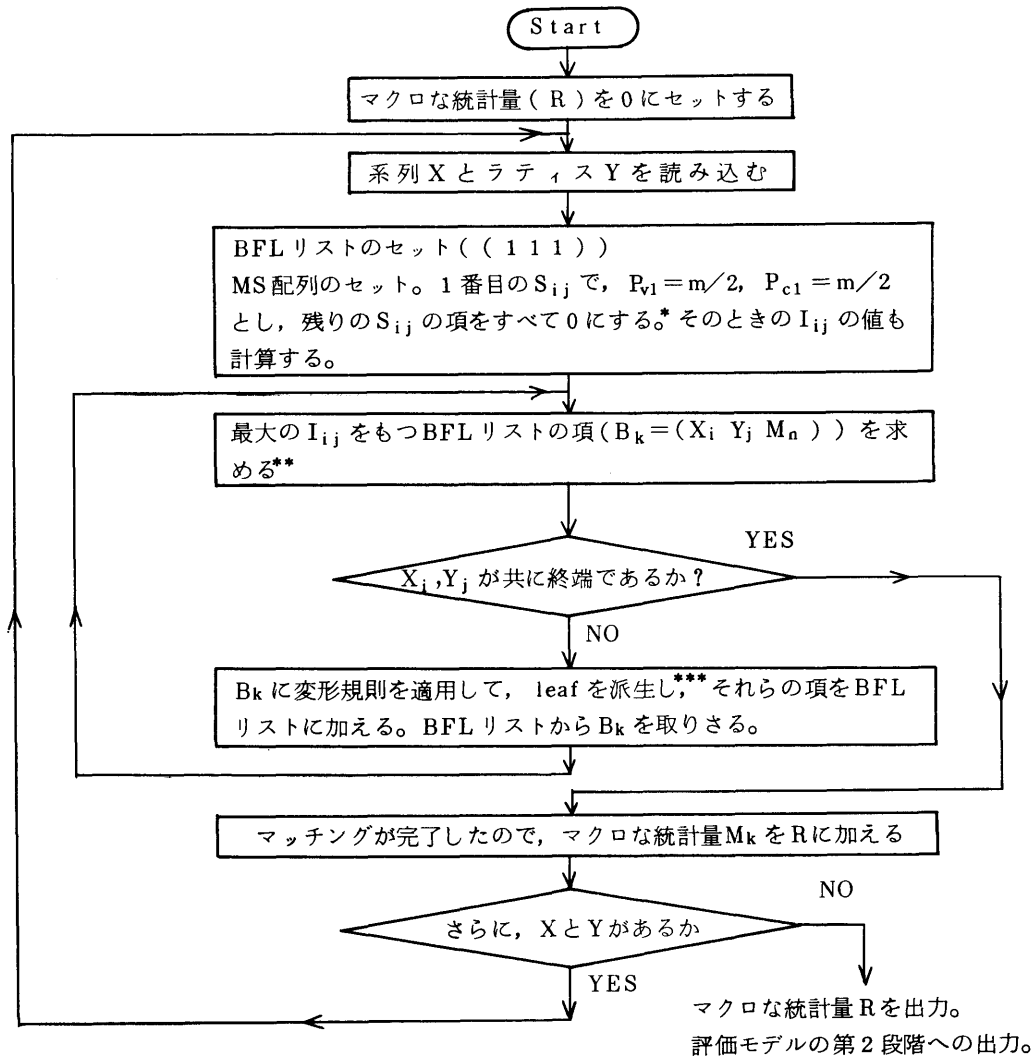
図 4-7 音韻系列のリスト表現

音韻ラティス（以下ラティスと略す）も図 4-8 のようにリスト表現で表わされている。

ラティスの位置 Y_i	セグメント・ポインター	
↓	↓	
((1 ((<u>P Z S H K</u>) 1 2)))		
(2 ((I) 5 3))		← 音韻認識のあいまいさ
(3 ((T P K S K Y) 4 4))		
(4 ((A) 3 5))		
(5 ((*) 1 6) ((P T) 5 8) ((S Z Y) 12 9))		
(6 ((E N N) 1 7))		↑ セグメンテーションのあいまいさ
(7 ((P T S H K) 5 8))		
(8 ((I) 3 9))		
(9 ((U N N) 6 10) ((D Z N M B) 6 10))		
(10 ((O) 6 11)))		

図 4-8 音韻ラティスのリスト表現

このマッチングのアルゴリズムを図4-11のフローチャートに示す。



- * マクロな統計量 R, S_{ij} は、実際には、確率ではなく、それが生じた頻度でたくわえる。m は系列 X の長さ。
- ** BFL リストの項がなくなったときは、マッチングが不可能となり、拒絶される。
- *** この項については、本文中で詳しく説明する。

図 4-11 系列とラティスのマッチングとマクロな統計量の抽出のフローチャート

図 4-11 のフローチャートで示されるマッチングのアルゴリズムの中で、leaf の派生する部分について説明を加える。最大の相互情報量をもつ leaf が BFL リストの内から選ばれ、その項が、

$$B_k = (X_i \ Y_j \ M_n)$$

であるとする。X_i が系列 X の終端であり、かつ、Y_j がラティス Y の終端であるかがチェックされる。もし、X_i, Y_j がともに終端であるならば、相互情報量が最大とな

表 4-2 変形規則を適用したときのマクロな統計量（出現数）の変化

		母 音 (X _i が母音)		子 音 (X _i が子音)		
規則 I	V → V	P _{v1}	P _{v1} - 1	C → C	P _{c1}	P _{c1} - 1
	x ∈ y	P _{v1}	P _{v1} + $\frac{1}{a}$		P _{c1}	P _{c1} + $\frac{1}{a}$
		V _v	V _v + $\frac{a-1}{a}$		V _c	V _c + $\frac{a-1}{a}$
規則 II	V → V	P _{v1}	P _{v1} - 1	C → C	P _{c1}	P _{c1} - 1
	x ∈ y	E _{v1}	E _{v1} + $\frac{1}{a}$		E _{c1}	E _{c1} + $\frac{1}{a}$
		V _v	V _v + $\frac{a-1}{a}$		V _c	V _c + $\frac{a-1}{a}$
規則 III	V → VCV	P _{v1}	P _{v1} - 1	C → CVC	P _{c1}	P _{c1} - 1
	V → VV	V _v	V _v + 1		V _c	V _c + 1
規則 IV	V → φ	P _{v1}	P _{v1} - 1	C → φ	P _{c1}	P _{c1} - 1
		U _v	U _v + 1		U _c	U _c + 1

a : ラティスの位置 y_j から始まるセグメントの数。

l : 変形規則を適用するセグメント内の音韻の数。

るマッチングが見つかったことを意味し、そのときの M_n 番目の MS 配列が、マクロな統計量 R に加えられる。さもなければ、B_k に変形規則（表 4-1）を適用し、可能な leaf をすべて派生し、BFL リストにつけ加える。各々の変形規則を適用したとき、M_n 番目の MS 配列の値から、マクロな統計量で変化させるべきものを表 4-2 に示す。規則 I および規則 II は、Y_j からはじまるすべてのセグメントに適用される。規則 III は、Y_j からはじまる一番短いセグメント列 VCV or CVC or VV にのみに適用される。規則 IV は、Y_j に無関係に、常に適用される。各々の変形規則から派生された leaf は、そのマクロな統計量から相互情報量をもとめ、それも MS 配列に蓄える。可能な leaf をすべて派生し、BFL リストに加える。その後、B_k は取りさる。さらに、BFL リストの項で、系列の位置 (X_i) とラティスの位置 (Y_j) が同じ項はまとめて、相互情報量をもっとも大きい項だけ残し、残りは取りさる。これによって leaf の数を減らし、アルゴリズムを高速化し、記憶容量を少なくしている。このようにし、また、BFL リストから最大の相互情報量をもつ項を選び、マッチングを続ける。

このアルゴリズムを図 4-7 の系列と図 4-8 のラティスのマッチングに適用すると、最終的に、次に示すような統計量が得られる。

$$P_{v1} = 4, U_v = 1$$

$$P_{c2} = 0.33, P_{c5} = 1.5, E_{c5} = 1, V_c = 1.17, U_c = 1$$

その他の統計量は 0

4.4.2 ラティスの相互情報量の推定

4.4.1 節のアルゴリズムで得られたマクロな統計量から, (4-56)式, (4-57)式, (4-55)式を用いて相互情報量 $I_v(X;Y)$, $I_c(X;Y)$, $I(X;Y)$ を求める。そして, 母音の相互情報量 $I_v(X;Y)$ と母音のセグメント誤り率

$$q_v = U'_v + V'_v$$

と, 子音の相互情報量 $I_c(X;Y)$ と子音のセグメント誤り率

$$q_c = U'_c + V'_c$$

の値を, 評価モデルの第 3 段階に送る。

4.4.3 音韻認識率の推定

第 2 段階から送られた $I_v(X;Y)$, $I_c(X;Y)$, q_v , q_c の値から, 音韻認識率 p_v, p_c 認識誤り率 r_v, r_c を求める。音韻系列の認識率と相互情報量の関係式 (4-39)式から,

$$I_v(X;Y) = (1 - q_v) \log K_v - (1 - q_v) \log (1 - q_v) + p_v \log p_v + r_v \log r_v - r_v \log (K_v - 1) \quad (4-58)$$

$$I_c(X;Y) = (1 - q_c) \log K_c - (1 - q_c) \log (1 - q_c) + p_c \log p_c + r_c \log r_c - r_c \log (K_c - 1) \quad (4-59)$$

の関係がある。また,

$$p_v + r_v + q_v = 1 \quad (4-60)$$

$$p_c + r_c + q_c = 1 \quad (4-61)$$

の関係がある。よって, (4-58)式と(4-60)式から, p_v, r_v の値を, (4-59)式と(4-61)式から, p_c, r_c の値を求めることができる。

(4-58)式, (4-59)式は, 必ずしも p_v, p_c の一価関数ではない。よって, それらの式の性質について検討を加える。(4-58)式で表わされる相互情報量の最大値は,

$$p_v = 1 - q_v, r_v = 0$$

のときで,

$$I_v(X;Y)_{\max} = (1 - q_v) \log K_v \quad (4-62)$$

である。つぎに、最小値をラグランジュの未定数法を用いて求める。(4-58)式と(4-60)式から、

$$F = I_v(X; Y) + \lambda (p_v + r_v + q_v) \quad (4-63)$$

$$\frac{\partial F}{\partial p_v} = \log p_v + 1 + \lambda = 0 \quad (4-64)$$

$$\frac{\partial F}{\partial r_v} = \log r_v + 1 - \log(K_v - 1) + \lambda = 0 \quad (4-65)$$

(4-64)式から(4-65)を引くと、

$$\log p_v - \log r_v + \log(K_v - 1) = 0 \quad (4-66)$$

$$r_v = (K_v - 1) p_v \quad (4-67)$$

(4-60)式に代入すると、

$$p_v = \frac{1 - q_v}{K_v} \quad (4-68)$$

$$r_v = \frac{(K_v - 1)(1 - q_v)}{K_v} \quad (4-69)$$

となる。 p_v , r_v が上式の値をとるとき、相互情報量は、最小

$$I_v(X; Y)_{\min} = 0 \quad (4-70)$$

となる。さらに、(4-58)式は、 p_v が、

$$\frac{1 - q_v}{K_v} \leq p_v \leq 1 - q_v \quad (4-71)$$

の間では、単調増加であり、

$$0 \leq I_v(X; Y) \leq (1 - q_v) \log K_v \quad (4-72)$$

の値をとることが容易にいえる。以上より、 $I_v(X; Y)$ と q_v の値が与えられたときの p_v の値は、容易に求められる。

以上の議論は、 $I_c(X; Y)$ と p_c についても同様にすることができる。

4.4.4 音韻認識率の自動推定アルゴリズムのまとめ

音韻ラティスの音韻認識率の自動推定アルゴリズムの各段階を、4.4.1節、4.4.2節、4.4.3節で述べた。この節では、これらを整理して示す。アルゴリズムは、3段階からなり、各々の機能を表4-3に示す。

表 4-3 アルゴリズムのまとめ

第一段階	入力	多量の音韻系列と音韻ラティスの組	
	機能	マクロな統計量の自動抽出	相互情報量を評価尺度とした best-first method によるマッチング
第二段階	出力	マクロな統計量, (母音: $\{P_{v1}, E_{v1}, i=1, 2\}, V_v, U_v,$ 子音: $\{P_{c1}, E_{c1}, i=1\sim 5\}, V_c, U_c$) (表4-1)	
	機能	ラティスの相互情報量の推定	(4-56)式, (4-57)式を利用。
	出力	相互情報量とセグメント化誤り率(母音: $I_v(X;Y), q_v,$ 子音: $I_c(X;Y), q_c$)	
第三段階	機能	音韻認識率の推定	(4-58)式, (4-59)式を利用。
	出力	音韻認識率(母音: $p_v, r_v, q_v,$ 子音: p_c, r_c, q_c)	

4.5 アルゴリズムの評価

音韻認識率を自動推定するアルゴリズムを評価するために、そのアルゴリズムでの推定結果と、人間が視察でとったマクロな統計量より求めた音韻認識率とを比較する。

比較のためのデータは、男性1名が文節ごとに区切って発声した列車の座席予約の20文章で、文節数は124文節である。そのデータに含まれる音韻は、1540個で、母音、子音、各各770個である。その音韻別の出現頻度を表4-4に示す。

自動推定のアルゴリズムによって求めたマクロな統計量を表4-5に、それから求めた相互情報量とそれに対応する音韻認識率を表4-6に示す。

次に、人が正しい音韻系列と出力の音韻ラティスとの対応関係をとることによって求めたマクロな統計量を表4-7に示す。そのマクロな統計量から推定した相互情報量と、音韻認識率を表4-8に示す。

以上のデータを基に、自動推定アルゴリズムと視察との比較検討をおこなう。全体の音韻認識率をみると、60.5%と59.9%で、約0.6%自動推定アルゴリズムの方が高い認識率を推定する。この0.6%は、実用上、ほぼ無視できる値である。さらに、表4-5と

表 4-4 音韻の出現頻度 (124 文節, 音韻数 1,540 個)

母 音	音韻名	A	I	O	U	E	NN														
	出現数	213	180	150	109	68	50														
	出現頻度	0.28	0.23	0.20	0.14	0.09	0.06														
子 音	音韻名	*	K	N	T	M	R	S	H	G	D	ZY	Y	Z	KY	HY	W	PP	KK	B	P
	出現数	126	102	98	61	59	53	52	47	35	33	26	22	17	12	9	8	5	3	2	0
	出現頻度	0.16	0.13	0.13	0.08	0.08	0.07	0.07	0.06	0.04	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01	0.01		

本システムは、VCV音節を単位としているため撥音の /N/ (表では NN で表わされている) も母音と見なされる。また、* は連続母音の間の仮想的な子音を表わす。

表 4-5 自動推定アルゴリズムで求めたマクロな統計量の出現数

母 音				子 音			
P_{v1}	5 2 3	E_{v1}	6	P_{c1}	7 4	E_{c1}	6
P_{v2}	8 3	E_{v2}	5	P_{c2}	1 0 8	E_{c2}	1 1
				P_{c3}	9 3	E_{c3}	1 5
				P_{c4}	6 7	E_{c4}	1 3
				P_{c5}	9 0	E_{c5}	5 4
$V_v + U_v$	1 5 3			$V_c + U_c$	2 3 9		

(記号は, 表 4-1 に準ずる)

表 4-6 自動推定アルゴリズムから求めた音韻認識率

		全 体	母 音	子 音
相互情報量 (bit)		1.5 4	1.8 3	1.2 6
認 識 率 (%)		6 0.5	7 7.3	4 3.7
誤 り 率	認 識 (%)	1 4.0	2.8	2 5.2
	セグメント化 (%)	2 5.5	1 9.9	3 1.1

表 4-7 視察によって求めたマクロな統計量の出現数

母 音				子 音			
P _{v1}	5 2 5	E _{v1}	8	P _{c1}	6 6	E _{c1}	9
P _{v2}	8 9	E _{v2}	3	P _{c2}	1 0 1	E _{c2}	1 4
				P _{c3}	9 3	E _{c3}	1 4
				P _{c4}	6 3	E _{c4}	1 4
				P _{c5}	9 2	E _{c5}	5 0
V _v + U _v	1 4 5			V _c + U _c	2 5 4		

表 4-8 視察から求めた音韻認識率

		全 体	母 音	子 音
相 互 情 報 量 (bit)		1.5 1	1.8 5	1.1 7
認 識 率 (%)		5 9.9	7 8.3	4 1.5
誤 り 率	認 識 (%)	1 4.2	2.9	2 5.5
	セグメント化 (%)	2 5.9	1 8.8	3 3.0

表 4-7 との比較から、自動推定アルゴリズムで妥当なマクロな統計量を推定できることがわかる。0.6% 高めに認識率を出す原因として、自動推定アルゴリズムでは、正しい系列と音韻ラティスとの対応づけを、(4-55)式~(4-57)式で定義される相互情報量が最大になるようにしているため、視察よりも、功妙な対応づけになってしまう場合が若干、存在しているためと考えられる。母音、子音別の認識率を比較すると、自動推定アルゴリズムでは、視察より、母音認識率では1.0%低く、子音認識率では、逆に、2.2%高いことがわかる。これは、子音と母音との情報量の差によるものと考えられる。つまり、子音の種類は20個で、その情報量は、4.32ビット($\log_2 20$)であり、母音の種類は6個で、その情報量は2.58ビット($\log_2 6$)であり、1個の子音は、1個の母音の約1.8倍の情報量を持っていることがわかる。よって、相互情報量に対応づけの評価尺度として用いる場合、大きな情報量をもつ子音の方を重視した対応づけがおこなわれがちになることが予想される。このことは、マクロな統計量を示す表 4-5、表 4-7 を比べ、自動推定ア

ルゴリズムの方が、 P_{c1} 、 P_{c2} の出現数が多くなっていることから推測される*

視察によって124文節の統計量を求めるには、セグメンテーションのあいまいさなどのために、約16時間・人かかる。自動推定アルゴリズムでは、ミニコンで70分余りである。さらに、自動推定アルゴリズムは、視察とはほぼ同じ認識率の値を推定することができるため、音響処理部の性能を測る客観的な評価系として有効であり、すでに、実際に音響処理部の評価に用いている。また、表4-6と表4-8の母音と子音の相互情報量の比較から、現在の音響処理部では、母音の方が多くの情報量を持っていることもわかる。

4.6 自動推定アルゴリズムの適用例

この節では、自動推定アルゴリズムの応用例をあげる。まず、8人の男性が発声した音声の音響処理部での音韻認識の評価に適用した場合を示す。次に、音響処理部でのVCV音節を単位とした認識の際に、子音認識のあいまいさの程度を決めるいき値の最適設定に適用した結果について述べる。最後に、あいまいな音韻の個数の制限による認識率の変化を調べるために、このアルゴリズムを適用した結果について述べる。

4.6.1 音響処理部の音韻認識率の推定

男性8名が列車の座席予約の20文章(124文節、出現音韻は表4-4と同じ)を2回ずつ発声したものを対象として、音響処理部⁽³²⁾の各々の出力の音韻ラティスの音韻認識率を推定した。この結果を表4-9に示す。

従来、音響処理部では、音韻ラティスの評価を、ラティス中に含まれる正しい音韻の割合でおこなっていた。この評価方法は、人手がかかるうえに、次のような大きな欠点をもっている。ラティス中に含まれる音韻の割合で評価するとき、ラティスの音韻の候補の数を増し、かつ、セグメントの候補を増すことによって、ラティス中に含まれる正しい音韻の割合は容易に大きくなる。

ここで提案した音韻認識率の自動推定アルゴリズムは、人手がかからないだけでな

* 対応づけの尺度として、相互情報量のかわりに認識率を用いることが考えられる。計算量は多くなるが、対応づけの精度が向上することが確かめられている。

表 4-9 音響処理部の音韻認識率

発声者		全体 (%)			母音 (%)			子音 (%)		
		正	誤		正	誤		正	誤	
			認識	セグメンテーション		認識	セグメンテーション		認識	セグメンテーション
MK	1	58.2	17.2	24.6	77.3	4.4	18.3	39.2	29.9	30.9
	2	55.9	15.3	28.8	74.4	3.7	22.0	37.4	27.0	35.7
KI	1	59.4	15.6	25.0	78.8	3.1	18.2	40.0	28.2	31.8
	2	58.7	16.8	24.5	78.3	2.9	18.9	39.2	30.6	30.2
SA	1	55.3	16.3	28.4	74.8	3.8	21.4	35.8	28.8	35.4
	2	55.9	16.0	28.1	74.7	3.3	22.0	37.2	28.6	34.2
SS	1	58.6	14.5	26.9	79.0	2.3	18.8	38.3	26.8	35.0
	2	60.0	15.4	24.7	81.4	2.3	16.3	38.5	28.5	33.0
RN	1	64.6	12.8	22.6	83.4	1.9	14.6	45.8	23.7	30.6
	2	60.8	14.1	25.1	78.0	2.8	19.2	43.6	25.3	31.1
KS	1	58.1	15.4	26.6	76.3	3.1	20.6	39.9	27.6	32.5
	2	58.2	16.2	25.6	76.5	2.7	20.8	39.9	29.6	30.5
SF	1	59.0	16.3	24.8	78.7	2.5	18.8	39.2	20.1	30.7
	2	62.2	15.2	22.6	81.2	2.4	16.4	43.2	28.1	28.7
HN	1	52.5	15.8	31.7	73.1	2.7	24.2	31.9	28.8	39.3
	2	57.9	15.6	26.6	78.8	2.7	18.4	36.9	28.4	34.7
平均		58.5	15.5	26.0	77.8	2.9	19.3	39.1	28.1	32.8

く、相互情報量という尺度に基づいているため、上のような欠点はなく、かつ、認識率という人間にわかりやすい尺度を出力する。

表 4-9 に示すような多量の音韻ラティスの評価も、自動推定アルゴリズムによって可能となり、音韻認識率、母音認識率、子音認識率を容易に得ることができる。さらに、母音、子音別の認識誤り率、セグメンテーション誤り率もわかり、音響処理部の性能評価のための重要な情報を与える。

さらに、このアルゴリズムで推定された音韻認識率と、言語処理部での文節認識率

との関係をグラフにすることから、音響処理部の音韻認識率の目標値の設定にも利用できる。

4.6.2 VCV 音節認識におけるいき値の最適決定への適用

音響処理部では、子音の認識を VCV 音節を単位としたマッチングでおこなっている。このとき、ラティスのセグメントに音韻の候補として、マッチングの程度がある程度以上の音韻だけが入れられる。現在では、このマッチングの尤度値のいき値として、もっともよくマッチングできた VCV 音節の尤度から、あるいき値 (T) 以内の VCV 音節中の子音を、ラティス中の

表 4-10 いき値と音韻認識率 (%)

T	音韻認識率	母音認識率	子音認識率
0.05	56.7	72.0	41.7
0.10	59.3	75.1	43.5
0.15	60.1	77.1	44.0
<u>0.20</u>	<u>60.6</u>	77.2	44.0
0.25	60.2	78.1	42.3
0.30	59.2	78.4	39.9
0.35	58.2	78.2	38.3
0.40	58.1	78.2	37.9

セグメントに加えている。このいき値の最適値 (T_{OPT}) を自動推定アルゴリズムによる推定結果から求める。

音響処理部で、いき値 T を 0.05 から 0.40 まで 0.05 ずつ増加し、各々の T に対する音韻ラティス (124 文節) を得、自動推定アルゴリズムで音韻認識率を推定した。各々のいき値 T に対する音韻認識率、母音認識率、子音認識率を表 4-10 に示す。

表 4-10 から、いき値 T の最適値 (T_{OPT}) は、

$$T_{OPT} = 0.20$$

であることがわかる。音響処理部では、このようにして得られたいき値 T_{OPT} を用いて、音韻ラティスを作成している。

いき値 T は、VCV 音節の認識にのみ、つまり、子音の認識率のみに影響を与えるはずであるが、実際には、母音認識率もかなり変動している。これは 4.5 節でも述べたように、母音のもつ情報量と子音のもつ情報量の違いから、相互情報量を評価尺度にしているために、T が小さくなるにつれて、ますます子音に重点がおかれすぎるようになるためである。T が 0.05 では、母韻認識率が 6% 余りも低く見つもられてお

り、子音がかなり高く見つもられている恐れがある。

4.6.3 VCV 音節認識における音韻の候補の数の制限の検討への適用

音響処理部では、セグメント内の音韻の候補数の制限をおこなっている。通常は、母音は2個まで、子音は5個まで許している。したがって、4.6.2節のいき値 (T_{OPT}) 内の子音でも、5位以内でないと候補に加えられないことになる。

表 4-11 候補数の制限と音韻認識率 (%)

L	音韻認識率	母音認識率	子音認識率
1	53.2	69.1	37.3
2	59.4	76.7	42.0
3	60.1	77.1	43.2
4	60.3	77.1	43.5
5	60.5	77.3	43.7

ここでは、124文節から得られる音韻ラティスを対象として、候補数 (L) の影響を調べるために、音韻認識率の自動推定アルゴリズムを用いた。この結果を表 4-11 に示す。

候補の数の制限 (L) が1のときは、認識率が7%以上も低くなり、複数の音韻の候補を許すことが非常に有効であるのがわかる。また、L=2のときの母音認識率は、L=1のときよりも7%余り向上し、母音認識があいまいなときは、2個の母韻の候補を許すことの妥当性を示している。さらに、3個以上の母音の候補を許すことについても検討する必要があると考えられる。L=3, 4, 5では、徐々に認識率が向上しており、最大となる値は存在していない。これは、 T_{OPT} 内の子音を無理にLで打ち切ることによる情報の損失があることを示している。理想的には、このようなLによる制限がない方がよいことを暗示している。しかしながら、L=3, 4, 5での認識率の増加は、0.2%と小さいので、実用上は、現状のL=5でもよいと考えられる。

4.7 あとがき

音韻認識のあいまいさと、セグメンテーションのあいまいさを含む時系列である音韻ラティスの評価方法を提案し、その有効性を確かめた。

音韻ラティスを評価するために、ラティスと入力系列との間の相互情報量を求め、それ

と等価な情報量をもつ音韻系列の音韻認識率をもって、ラティスの音韻認識率とした。このようにして、直観的にわかりやすい評価尺度として、音韻認識率、認識誤り率、セグメンテーション誤り率を得ることができた。ラティスから相互情報量を求める際に、ラティスの出力事象の種類が非常に多く、有意な統計量をとることが不可能であるので、ラティスからマクロな統計量を求め、それより相互情報量を推定する手法をとった。この推定の際、マクロな統計量内での分布のかたより（エントロピーの値）が問題となる。このエントロピーの値の問題について、我々は、ラティスから入力音韻系列を推定するのに実際に利用可能な範囲の分布のかたより（エントロピーの値）だけを考慮すべきであり、単に出力事象を増すことによって相互情報量が増加するような推定方法をとるべきでないなどの点を考え、マクロな統計量から求まる最小の相互情報量が適当であるとした。このときの、マクロな統計量内での事象の生起確率はすべて等確率で、エントロピーの値は最大値をとっていることにあたる。さらに、マクロな統計量を求めるにも、人手がたいそうかかるため、正しい音韻系列と音韻ラティスとの対応づけを自動的におこなない、その対応づけからマクロな統計量を得るアルゴリズムを提案した。この対応づけは、正しい音韻系列と音韻ラティスとの相互情報量が最大になるようにおこなわれる。この対応づけのアルゴリズムを、tree search で実行し、相互情報量を尺度とした best-first method でおこなった。このようにして、音韻ラティスの音韻認識率を自動推定するアルゴリズムを確立した。

この音韻認識率自動推定アルゴリズムによって求めた音韻認識率と、視察で求めたマクロな統計量から推定した音韻認識率とは、ほとんど同じ認識率の値をとることも確かめられ、このアルゴリズムの妥当性も確かめられた。

現在、この自動推定アルゴリズムは、音響処理部のパラメータのいき値の決定や、音響処理部の評価に用いられ、その効力を発揮している。

この自動推定アルゴリズムは、ミニコン U-400（サイクルタイム 0.8 μ sec, コア 16 bits \times 64k words）上に FORTRAN でつくられている。リスト処理の部分のプログラムは、FORTRAN によるリスト処理言語 DLOP で書かれている。

本稿で述べた音韻ラティスの評価手法は、音韻ラティスの評価だけでなく、パターン認識における認識とセグメンテーションのあいまいさを含むさまざまな処理結果の評価にも適用できる。また、相互情報量という尺度は、音声認識でのマッチングの尺度としても利用できるし、2つの音韻系列間の類似性の尺度としても利用できる。とくに、この尺度は、

長さの異なる音韻系列間の類似性の尺度にもなっており、音声認識系の解析のための有力な手段にもなり得ると考えられる。

本アルゴリズムの今後の課題として、次の事柄があげられる。まず、マクロな統計量から相互情報量を推定する際のエントロピーの値に対する理論的検討をすすめ、その意味を明らかにする必要がある。次に、対応づけの評価尺度として相互情報量を用いると、多少、子音に重点を置いた対応づけになるようであり、母音にも子音にも適切に重みをおいた対応づけにするために、評価尺度の改良の検討が必要となるかもしれない。また、相互情報量という評価尺度は、計算量が多く、プログラムの速度が遅くなる（124文節で約70分）主要原因となっているので、実行速度をあげるには、計算量が少ない他の評価尺度も検討する必要がでてくるであろう。

第5章 LPCスペクトル・マッチング尺度の検討

5.1 はしがき

会話音声認識の研究は、ここ数年間精力的に行なわれてきた。アメリカではARPAのプロジェクトも終了し、今後重点をおいて研究を続けるべき問題点について検討されている。⁽⁹⁾⁽¹⁰⁾ 我国でも、会話音声認識の研究が行なわれ、かなりの成果をあげた。言語情報の利用に関しては、具体的なアルゴリズムが提案され、その有効性が明確になった。しかしながら、音響処理については不十分で、基礎的な研究の継続が必要である。我々も列車の座席予約を対象として、オンラインで動作する会話音声による質問回答システムを作成した。^(第2章~第3章) このシステムの対象を拡大し、性能を向上させるには、音響処理の精密化が必要であると考えられる。我々のシステムの音響処理⁽³²⁾は、LPC(Linear Predictive Coding)分析⁽¹³⁾を介して得られたスペクトル包絡のマッチングを基本にして構成されている。本章では、音響処理の精密化の一環として検討したスペクトル・マッチングの尺度について述べる。様々なスペクトル・マッチングの尺度が、音韻認識⁽⁸⁴⁾、単語認識⁽²⁷⁾⁽⁵⁸⁾、会話音声認識⁽³²⁾⁽⁴²⁾⁽⁴⁶⁾で利用されている。

音声中の音韻を認識する一方法として、標準パターンとのマッチングによって行なうことが考えられる。その際、母音区間を切り出して、その区間の母音を識別するという単純な過程によるのではなく、この逆の過程、すなわち、マッチングによって母音を識別して母音区間を切り出すという過程も必要と考えられる。このような両方向の過程で母音を認識するには、入力区間のフレームが特定の母音であるか否かを判定できる絶対的な尺度となりうるマッチングの手法が望まれる。

従来、音声認識では、スペクトル・マッチングの尺度として、対数スペクトル距離⁽⁸⁵⁾、正規化対数予測残差⁽⁵⁸⁾⁽⁴²⁾⁽⁴⁶⁾、対数予測⁽²⁷⁾⁽³²⁾などが用いられてきた。さらに、LPC分析をしたスペクトル包絡のスペクトル・マッチング尺度として、COSH尺度⁽⁸⁶⁾、LPCケプストラム距離⁽⁸⁷⁾も提案されている。本章では、まず、会話音声での母音認識のための絶対的な基準をもつ尺度を見出すことを目的として、LPC分析をしたスペクトル包絡上での従来からのLPCスペクトル・マッチングの尺度の評価を行なう。評価の対象として、最尤スペクトル距離、予測残差、COSH尺度、LPCケプストラム距離を取りあげる。パ

ワーの値を様々な条件に設定し、各尺度の評価を、母音の識別実験で行う。これらの LPC スペクトル・マッチング尺度の評価を 5.2 節と 5.3 節で行う。

従来から、ピーク付近での重みづけや周波数軸での重みづけが、スペクトル・マッチングによる音声認識に有効であるとされてきた。周波数軸に関しては、線形な周波数軸より対数軸の方が妥当と考えられる。⁽⁴⁾ また、ピークに関しては、聴覚的に、ピークでの聴覚の感度は、谷よりも強いことが確かめられている。⁽⁸⁸⁾ 本章の 5.4 節以降で、LPC スペクトル包絡のピークに重みをおく、新しい LPC スペクトル・マッチング尺度を提案し、評価を行う。

音声認識におけるスペクトル・マッチングの手法として、代表的な 2 種類の方法がある。1 つには、周波数領域をいくつかの帯域に分割し、各帯域での周波数強度を用いる band-pass filter (BPF) による方法である。もう 1 つは、LPC 分析によって LPC スペクトル包絡を求め、スペクトル包絡間の差を計算する方法である。BPF による方法は、周波数軸の重みづけやピークでの重みづけが可能であるという利点を持つが、ピッチの影響のため、分割帯域数が 20 ~ 30 程度しかとれず、周波数分解能が悪いという欠点を持っている。この 5.2 ~ 5.3 節で検討する従来からの LPC 分析によるスペクトル包絡間の差を求める方法はピッチを除いたスペクトル包絡間の差を求めることができる点で、BPF による方法よりも周波数分解能はよいけれども、ピークの重みづけや周波数軸の重みづけができないという欠点を持っている。以上の観点から、LPC スペクトル包絡の差を計算でき、かつ、ピークの重みづけと周波数軸の重みづけができる尺度が望まれてきた。5.4 節では、このような条件も満す LPC Peak Weighted Spectral Matching Measure (以下、Peak Weighted Measure と略す) を導出する。

さらに、Peak Weighted Measure の評価を、母音識別実験を通して、5.5 ~ 5.6 節で行い、Peak Weighted Measure が有効であることを実証する。

また、5.7 節では、大語いの単語音声 (641 都市名) を対象にとりあげ、LPC スペクトル・マッチング尺度の評価を単語音声認識の立場から行う。

この節での検討をもとに、第 6 章で、新しい会話音声認識システム Voice Q-A System II を作成する。

5.2 LPC スペクトル・マッチング尺度 (LPC Distance Measure)

LPC 分析を介して得られた標準パターンのスペクトル包絡を $f(\theta)$ とし、入力音声のピリオドグラム⁽¹⁾のスペクトル包絡を $g(\theta)$ とする。 $f(\theta)$ と $g(\theta)$ に関する LPC 分析でのパラメータの記法を表 5-1 に示す。ただし、 θ は角周波数である。 $f(\theta)$ 、 $g(\theta)$ の LPC スペクトル包絡は次式で表わされる。

$$f(\theta) = \frac{1}{2\pi} \frac{u \cdot R_f}{|1 + a_1 e^{j\theta} + \dots + a_p e^{jp\theta}|^2}$$

$$= \frac{1}{2\pi} \frac{u \cdot R_f}{\sum_{k=-p}^p A_k \cdot e^{jk\theta}} \quad (5-1)$$

$$g(\theta) = \frac{1}{2\pi} \frac{v \cdot R_g}{|1 + b_1 e^{j\theta} + \dots + b_p e^{jp\theta}|^2}$$

$$= \frac{1}{2\pi} \frac{v \cdot R_g}{\sum_{k=-p}^p B_k \cdot e^{jk\theta}} \quad (5-2)$$

表 5-1 標準パターンと入力音声の LPC 分析で用いるパラメータの記法

	パワー	自己相関係数	予測係数	最ゆうスペクトル パラメータ	正規化残差	ケプストラム 係数
標準パターン $f(\theta)$	u	r_j	a_i	A_j	R_f	$C_n^{(f)}$
入力音声 $g(\theta)$	v	s_j	b_i	B_j	R_g	$C_n^{(g)}$

$i = 1 \dots p$, $j = -p \dots p$, p : AR モデルの次数, $n = -q \dots q$, q : ケプストラム分析の次数

$$A_j = \sum_{i=|j|}^p a_i a_{i-|j|}, B_j = \sum_{i=|j|}^p b_i b_{i-|j|} \quad (\text{但し, } a_0 = 1, b_0 = 1)$$

5.2.1 最尤スペクトル距離 (Likelihood Ratio Measure)⁽¹³⁾

最尤スペクトル距離 (LR) は、LPC 分析でスペクトル包絡を推定する際に、評価関数として用いられた尺度で、(5-3) 式で表わされる。⁽¹³⁾

$$LR = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \log \frac{f(\theta)}{g(\theta)} + \frac{g(\theta)}{f(\theta)} - 1 \right\} d\theta \quad (5-3)$$

ここで,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\theta) d\theta = \log \frac{uR_f}{2\pi} \quad (5-4)$$

の関係式を用いる。

これは,

$$1 + a_1 e^{j\theta} + \dots + a_p e^{jp\theta}$$

が, 単位円の内にすべての0点をもっていることから導びかれる。⁽¹³⁾ 同様にして,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log g(\theta) d\theta = \log \frac{vR_g}{2\pi} \quad (5-5)$$

の関係がある。さらに,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{g(\theta)}{f(\theta)} d\theta = \frac{v}{uR_f} \sum_{k=-p}^p A_k s_k \quad (5-6)$$

の関係式も得られる。

この式は以下のようにして導出される。

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{g(\theta)}{f(\theta)} d\theta &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{v}{uR_f} \sum_{\tau=-\infty}^{\infty} s_{\tau} e^{j\tau\theta} \cdot \sum_{k=-p}^p A_k \cdot e^{jk\theta} d\theta \\ &= \frac{v}{uR_f} \sum_{k=-p}^p A_k s_k \end{aligned}$$

同様にして

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\theta)}{g(\theta)} d\theta = \frac{u}{vR_g} \sum_{k=-p}^p B_k r_k \quad (5-7)$$

の関係も得られる。

(5-4)~(5-7)式の関係を(5-3)式に代入すると, 最尤スペクトル距離LRは,

(5-8)式で簡単に計算される。

$$LR = \log \frac{uR_f}{vR_g} + \frac{v}{uR_f} \sum_{k=-p}^p A_k s_k - 1 \quad (5-8)$$

対数スペクトルの差 $V(\theta)$ に対するスペクトルマッチングの様子を調べるために,

$$V(\theta) = \log f(\theta) - \log g(\theta) \quad (5-9)$$

とおくと、(5-3)式は、

$$LR = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ e^{-V(\theta)} + V(\theta) - 1 \} d\theta \quad (5-10)$$

と表わされる。⁽¹³⁾ (5-10)式の被積分項を $V(\theta) = 0$ のまわりでテーラ展開すると、

$$\begin{aligned} e^{-V(\theta)} + V(\theta) - 1 &= \sum_{i=2}^{\infty} \frac{1}{i!} (-V(\theta))^i \\ &= \frac{1}{2} V(\theta)^2 - \frac{1}{6} V(\theta)^3 + \frac{1}{24} V(\theta)^4 - \dots \end{aligned} \quad (5-11)$$

となり、 $V(\theta)$ が小さいときには、自乗の重みに近いことがわかる。 $V(\theta) \gg 0$ のときには、 $|V(\theta)|$ に比例する重みを、 $V(\theta) \ll 0$ のときには、指数関数($e^{|V(\theta)|}$)の重みをもつことがわかる。最尤スペクトル距離LRの $V(\theta)$ に対する重みを図5-1に示す。

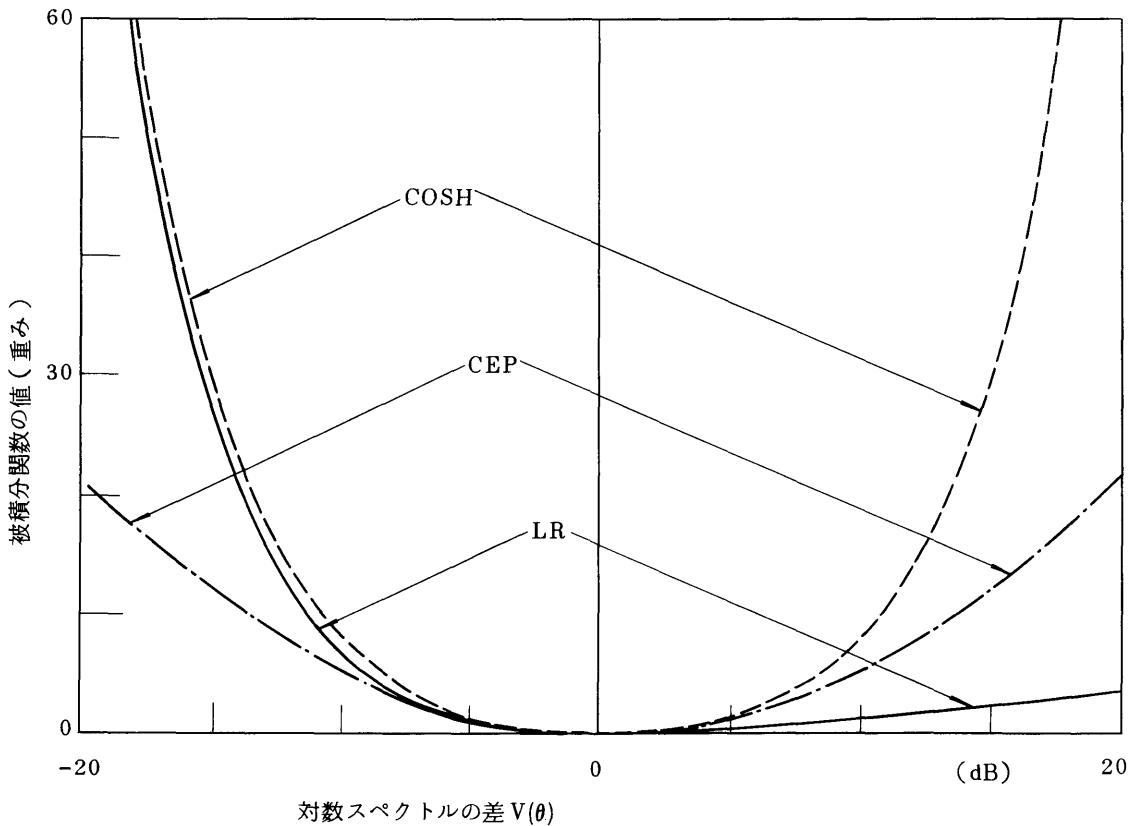


図5-1 尺度COSH, CEP, LRにおける対数スペクトル差 $V(\theta)$ に対する重み

5.2.2 正規化残差 (Normalized Residual)⁽²⁷⁾

正規化残差 (NR) は、入力音声を標準パターンのフィルターに通したときの出力を入力パワーで正規化したものと見なせる。正規化残差 NR は次式で表わされる。

$$NR = \sum_{k=-p}^p A_k s_k \quad (5-12)$$

(5-12) 式は、最尤スペクトル距離 (5-8) 式が最小の値をとるように u と v の比を設定して、かつ、入力音声だけに関する項を省略した式に等しくなる (表 5-2 の LR 参照)。また対数尤度を表わす式⁽¹³⁾

$$L = -\frac{N}{2} \left(\log 2\pi u R_f + \frac{v}{u R_f} \sum_{k=-p}^p A_k s_k \right)$$

に最尤法を適用して u に関して最大をとることからも導かれる。

$$L = \max_u \left(-\frac{N}{2} \left(\log 2\pi u R_f + \frac{v}{u R_f} \sum_{k=-p}^p A_k s_k \right) \right)$$

L を $(u R_f)$ で偏微分して 0 とおく。

$$\frac{\partial L}{\partial (u R_f)} = -\frac{N}{2} \left(\frac{1}{u R_f} - \frac{v}{(u R_f)^2} \sum_{k=-p}^p A_k s_k \right) = 0$$

$$u R_f = v \sum_{k=-p}^p A_k s_k$$

よって、

$$\begin{aligned} L &= -\frac{N}{2} \left(\log 2\pi + \log v \cdot \sum_{k=-p}^p A_k s_k + 1 \right) \\ &= -\frac{N}{2} \left(\log 2\pi + \log v + \log \sum_{k=-p}^p A_k s_k + 1 \right) \end{aligned} \quad (5-13)$$

よって、入力の音声がどの標準パターンに合致するかは、正規化残差のみで判定できることがわかる。ただし、上式は最尤法を用いて導出されているため、 L の値が尤度の絶対値を表わしているか否かについては疑問が残される。

この正規化残差 NR による尺度では、入力音声の LPC 分析は不必要で、相関係数 (s_τ) のみを計算すればよい。これに基づいて、通研での単語音声認識は行なわれてきた。⁽²⁷⁾

5.2.3 COSH尺度 (COSH Measure)⁽⁸⁶⁾

COSH尺度 (COSH) は、最尤スペクトル距離 (LR) のマッチングの重みの非対称性を取り除くために考え出され、最尤スペクトル距離 (5-3) 式と、 $f(\theta)$ と $g(\theta)$ とを入れ換えた式とを加えた式で表わされる。

$$\begin{aligned} \text{COSH} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \log \frac{f(\theta)}{g(\theta)} + \frac{g(\theta)}{f(\theta)} - 1 \right\} d\theta \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \log \frac{g(\theta)}{f(\theta)} + \frac{f(\theta)}{g(\theta)} - 1 \right\} d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} 2 \cdot \left\{ \cosh (\log f(\theta) - \log g(\theta)) - 1 \right\} d\theta \quad (5-14) \end{aligned}$$

COSH尺度は、(5-14) 式で示されるように、対数スペクトルの差の \cosh 関数で表わされる。COSH尺度は、(5-8) 式を用いて簡単に計算される。

$$\text{COSH} = \frac{v}{u R_f} \sum_{k=-p}^p A_k s_k + \frac{u}{v R_g} \sum_{k=-p}^p B_k r_k - 2 \quad (5-15)$$

次に、対数スペクトルの差 $V(\theta)$ に対する重みを調べるために、(5-14) 式を $V(\theta)$ で表わす。

$$\text{COSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (e^{V(\theta)} + e^{-V(\theta)} - 2) d\theta \quad (5-16)$$

(5-16) 式の被積分項を $V(\theta) = 0$ のまわりでテーラ展開すると、

$$\begin{aligned} e^{V(\theta)} + e^{-V(\theta)} - 2 &= \sum_{i=1}^{\infty} \frac{2}{(2i)!} V_{(\theta)}^{2i} \\ &= V_{(\theta)}^2 + \frac{1}{12} V(\theta)^4 + \frac{1}{360} V(\theta)^6 + \dots \quad (5-17) \end{aligned}$$

となり、 $|V(\theta)|$ が小さいときには、自乗にきわめて近い重みをもっていることがわかる。

$|V(\theta)| \geq 0$ のときには指数関数 ($e^{|V(\theta)|}$) の重みをもつ。この重みの様子を 図 5-1 に示す。

5.2.4 LPCケプストラム距離 (LPC Cepstrum Distance)⁽⁸⁶⁾

LPC分析から得られたスペクトル包絡のLPCケプストラム係数は、予測係数(a_i :

$f(\theta)$, $b_i : g(\theta)$) から次式によって求められる。⁽⁸⁷⁾

$$C_0^{(f)} = \log \left(\frac{u \cdot R_f}{2\pi} \right), \quad C_0^{(g)} = \log \left(\frac{v \cdot R_g}{2\pi} \right) \quad (5-18)$$

$$\left. \begin{aligned} C_n^{(f)} &= -a_n - \sum_{k=1}^{n-1} \frac{n-k}{n} a_k \cdot C_{n-k}^{(f)} \\ C_n^{(g)} &= -b_n - \sum_{k=1}^{n-1} \frac{n-k}{n} b_k \cdot C_{n-k}^{(g)} \\ C_n^{(f)} &= C_n^{(f)}, \quad C_{-n}^{(g)} = C_n^{(g)} \end{aligned} \right\} \quad (5-19)$$

ただし, $k > p$ に対しては, $a_k = 0$, $b_k = 0$ 。ケプストラム距離 (CEP) は次式で定義される。

$$CEP = \sum_{i=-\infty}^{\infty} (C_i^{(f)} - C_i^{(g)})^2 \quad (5-20)$$

$$= \left(\log \frac{uR_f}{vR_g} \right)^2 + 2 \cdot \sum_{i=1}^{\infty} (C_i^{(f)} - C_i^{(g)})^2 \quad (5-21)$$

ところで, ケプストラム係数の推定値は

$$C_n^{(f)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\theta) e^{-jn\theta} d\theta$$

$$C_n^{(g)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log g(\theta) e^{-jn\theta} d\theta$$

と定義されるので,

$$C_n^{(f)} - C_n^{(g)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ \log f(\theta) - \log g(\theta) \} e^{-jn\theta} d\theta$$

となる。Parseval の定理を用いると (5-20) 式は,

$$\begin{aligned} CEP &= \sum_{i=-\infty}^{\infty} (C_i^{(f)} - C_i^{(g)})^2 \\ &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \{ \log f(\theta) - \log g(\theta) \}^2 d\theta \\ &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} V(\theta)^2 d\theta \end{aligned} \quad (5-22)$$

となる。CEP の対数スペクトル差 $V(\theta)$ に対する重みも図 5-1 に示す。

5.2.5 各尺度のマッチングの条件

入力音声のパワー(v)と標準パターンのパワー(u)との比の設定によって、各尺度(LR, COSH, CEP)は異なる値をとる。ここでは、各尺度の次の条件でのマッチングを考える。

〔条件Ⅰ〕 パワーの値をそのまま用いる。

〔条件Ⅱ〕 パワーを正規化する。

$$u = v \quad (5-23)$$

〔条件Ⅲ〕 残差パワーを正規化する。

$$u \cdot R_f = v \cdot R_g \quad (5-24)$$

〔条件Ⅳ〕 尺度の値が最小になるように、u, vの値を設定する。

最尤スペクトル距離(LR)を最小にするuとvのパワー比は、(5-8)式をuあるいはvで偏微分して、0とおくことによって求まる。(27)(58)

$$\frac{\partial (LR)}{\partial (uR_f)} = \frac{1}{uR_f} - \frac{v}{(uR_f)^2} \sum_{k=-p}^p A_k S_k = 0$$

$$uR_f = v \cdot \sum_{k=-p}^p A_k S_k \quad (5-25)$$

すると、

$$LR = \log \sum_{k=-p}^p A_k S_k - \log \sum_{k=-p}^p B_k S_k$$

$$= \log \sum_{k=-p}^p A_k S_k - \log R_g$$

となる。この尺度は、板倉によって導出され、アメリカで音声認識のスペクトル・マッチングの尺度として、広く用いられている。(58)(42)(46) 同様に、COSH尺度を最小にするパワー比も(5-15)式をuあるいはvで偏微分して0とおくことによって求められる。

$$\frac{\partial (COSH)}{\partial (uR_f)} = -\frac{v}{(uR_f)^2} \sum_{k=-p}^p A_k S_k + \frac{1}{vR_g} \cdot \frac{1}{R_f} \sum_{k=-p}^p B_k r_k = 0$$

$$\therefore u \cdot \sqrt{R_f \cdot \sum_{k=-p}^p B_k r_k} = v \cdot \sqrt{R_g \cdot \sum_{k=-p}^p A_k S_k} \quad (5-26)$$

このときのCOSH尺度COSHは

$$\text{COSH} = 2 \sqrt{\frac{\sum A_k s_k \cdot \sum B_k r_k}{R_f \cdot R_g}} - 2$$

で表わされる。

さらにケプストラム距離 CEP を最小にする条件は、(5-21)式より、残差パワーを等しくしたときであることがわかる。なぜなら、 $C_i^{(f)}$ 、 $C_i^{(g)}$ ($i \geq 1$) は、(5-19)式から導出され、パワーの値は、(5-21)式の第1項のみに影響するからである。

条件 I ~ IV での各尺度の計算式を表 5-2 に示す。今後、各尺度を表 5-2 の略号 (LR_i 、 COSH_i 、 CEP_i ; 添字 i は条件 i を表わす) を用いて表わすことにする。

表 5-2 LPC 距離尺度

		条件	パワー設定値	LPC 距離尺度
最ゆうスペクトル距離 (LR)	LR_1	1	$\log(uR_f/vR_g) + (v/uR_f)\sum A_k s_k - 1$
	LR_2	2	$u = v$	$\log(R_f/R_g) + (1/R_f)\sum A_k s_k - 1$
	LR_3	3	$uR_f = vR_g$	$(1/R_g)\sum A_k s_k - 1$ (similar to LR_4)
	LR_4	4	$uR_f = v\sum A_k s_k$	$\log(\sum A_k s_k) - \log(R_g)$
予測残差 (NR)	NR_4	4	$(uR_f = v\sum A_k s_k)$	$\sum A_k s_k$
COSH 尺度 (COSH)	COSH_1	1	$(v/uR_f)\sum A_k s_k + (u/vR_g)\sum B_k r_k - 2$
	COSH_2	2	$u = v$	$(1/R_f)\sum A_k s_k + (1/R_g)\sum B_k r_k - 2$
	COSH_3	3	$uR_f = vR_g$	$(1/R_g)\sum A_k s_k + (1/R_f)\sum B_k r_k - 2$
	COSH_4	4	$\frac{u^2 R_f \sum B_k r_k}{= v^2 R_g \sum A_k s_k}$	$(2/\sqrt{R_f R_g})\sqrt{\sum A_k s_k \sum B_k r_k} - 2$
LPCケプストラム 距離 (CEP)	CEP_1	1	$(\log(uR_f/vR_g))^2 + 2 \sum_{i=1}^{16} (C_i^{(f)} - C_i^{(g)})^2$
	CEP_2	2	$u = v$	$(\log(R_f/R_g))^2 + 2 \cdot \sum_{i=1}^{16} (C_i^{(f)} - C_i^{(g)})^2$
	$\text{CEP}_3, \text{CEP}_4$	3, 4	$uR_f = vR_g$	$2 \cdot \sum_{i=1}^{16} (C_i^{(f)} - C_i^{(g)})^2$

5.2.6 マッチングの様子を観察

各条件によって設定されたパワー比におけるスペクトルマッチングの様子を観察する。マッチングの様子と各尺度 (LR, COSH, CEP) の対数スペクトルの差の様子を図示し、検討する。図示するのは、下記の値にパワー比を設定した場合である。

(a) パワーの値 (u , v) をそのまま利用。

(b) パワーを正規化 (条件 II) 。

$$u = v$$

(c) 残差パワーを正規化 (条件 III) 。

$$uR_f \approx vR_g$$

(d) 最尤スペクトル距離の値が最小 (条件 IV の LR)

$$uR_f = v \sum_{k=-p}^p A_k S_k$$

(e) COSH 尺度の値が最小 (条件 IV の COSH)

$$u^2 \cdot R_f \cdot \sum_{k=-p}^p B_k r_k = v^2 \cdot R_g \cdot \sum_{k=-p}^p A_k S_k$$

正規化残差 (NR) は、パワー比 (d) の値に設定した場合のマッチングの状態を表わす相対的な尺度と考えられる。

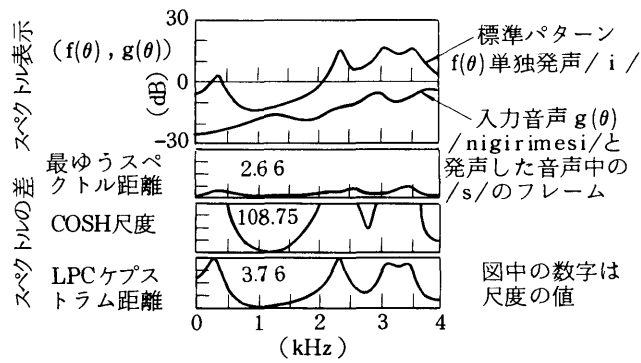
図 5-2 に、単独に発声した母音 / i / から作成した標準パターン / I / と / n i g - i r i m e s i / と発声した / S / の部分とのスペクトル・マッチングの様子を示す。パワーも考慮した場合 (a) では / I / と / S / のパワーが非常に異なるため、いずれの尺度も大きな値をとっている。次にパワーを正規化した場合 (b) では、 / S / のスペクトルがもち上げられぎみでマッチングされている。COSH と CEP の値は (a) のときより小さくなっているが、尺度 LR の値は大きくなっている。視覚的には、 (a) より (b) の方がよくマッチングされているように思われる。よって LR の値は、視覚で見たマッチングの状態と一致しない。尺度 LR が最小となる場合 (d) でも、 / I / のスペクトルが下がりみになっている。これは、マッチングの重みの非対称性による。この場合も、 (c) や (e) のマッチングの状態よりも視覚的にはよくマッチングされていない。この例では、パワー比を (c) , (e) に設定した場合が、よくマッチングされていると思われる。また、尺度 COSH, CEP の値がよくマッチングの状態を表わしている。

さらに、図 5-3 に、単独に発声した / a / から作成した標準パターン / A / と / i e a o u N / と発声した音声の / a / の部分から作成した / A / とのマッチングの様子を示す。この例でも、 (c) , (e) のマッチングの状態がよいように思われ、COSH, CEP が妥当な値を示していることがわかる。

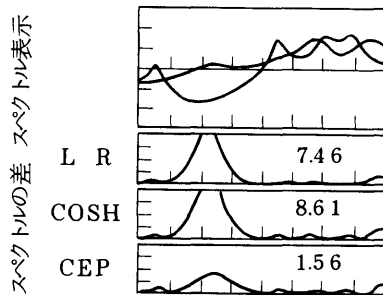
以上より、パワー情報を無視して、スペクトル・マッチングを行なう場合、パワー

比を(c)あるいは(d)に設定し, COSH尺度 $COSH_3$, $COSH_4$, あるいは, ケプストラム距離 CEP_3 でマッチングの度合を表わすのがよいと考えられる。

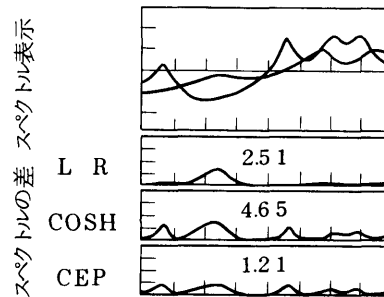
図5-2, 図5-3の場合の, (a)~(e)のパワー比の設定条件での各尺度の値を表5-3にまとめておく。表5-3で, 下線で示された値は, 最小の値を示す条件(条件IV)での各尺度の値である。パワー比の設定条件(d)では, LRの値は最小であるが, COSH, CEPの値はかなり大きくなっていることがわかる。また, パワー比の設定条件(c)と(e)では, 各尺度ともかなり近い値をとっており, きわめて類似したマッチングの状態になっていることが推測される。



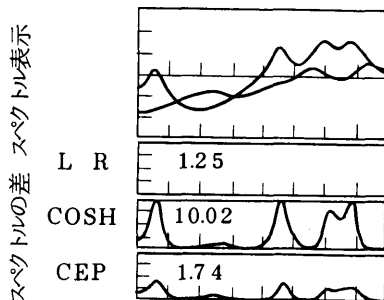
(a) (条件1) パワー利用



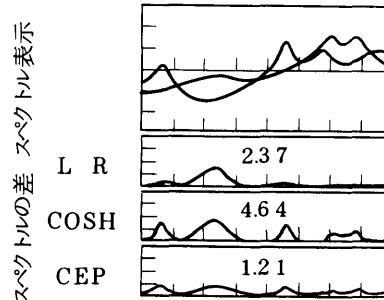
(b) (条件2) パワー正規化, ($u=v$)



(c) (条件3) 残差パワー正規化 ($uR_f = vR_g$)



(d) (条件4-1) 最ゆうスペクトル距離 最小



(e) (条件4-2) COSH尺度 最小

図5-2 標準パターン/i/と入力音声/s/とのスペクトルマッチングの状態

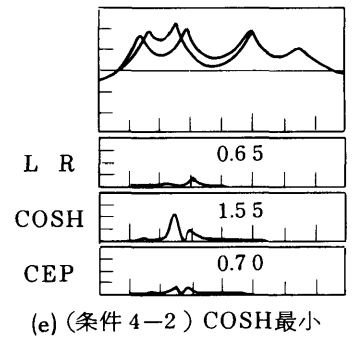
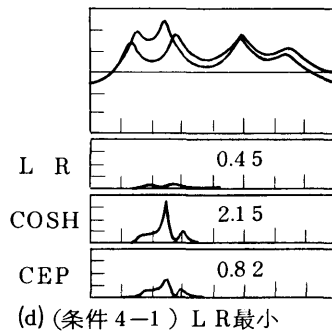
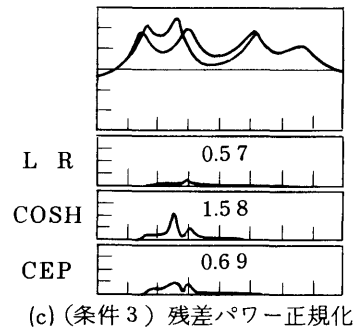
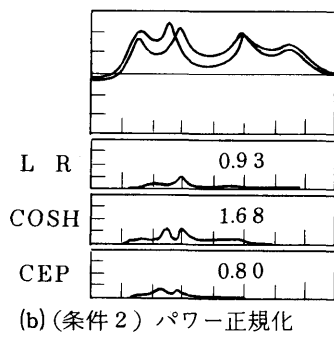
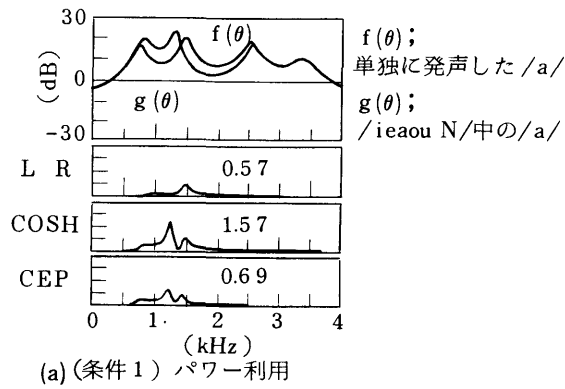


図5-3 単独に発声した/a/と連続母音中の/a/とのスペクトルマッチングの状態

表5-3 さまざまな条件でのLPC距離尺度の計算値の例

標準パターン		/i/			/a/		
入力音声		/s/ in /nigirimesi/			/a/ in /ieaou N		
パワーの設定値		LPC距離尺度					
		L R	COSH	CEP	L R	COSH	CEP
(a) (1)	...	2.66	1088	3.76	0.57	1.57	0.69
(b) (2)	$u = v$	7.46	861	1.56	0.93	1.68	0.80
(c) (3)	$uR_f = vR_g$	2.51	465	<u>1.21</u>	0.57	1.58	<u>0.69</u>
(d) (4-1)	$uR_f = v \sum A_k S_k$	<u>1.25</u>	1002	1.74	<u>0.45</u>	2.15	0.82
(e) (4-2)	$u^2 R_f \sum B_k \Gamma_k = v^2 R_g \sum A_k S_k$	2.37	<u>464</u>	1.21	0.65	<u>1.55</u>	0.70

5.3 LPCスペクトル・マッチング尺度の評価

5.3.1 母音識別実験による評価

条件 I ~ IV の条件での各尺度の評価を，母音識別率で行なう。標準パターンとして，母音 / a , i , u , e , o / と撥音 / N / を用いる。各尺度の評価は，各尺度の値が，音韻の絶対的な識別（音韻 / X / であるか否かの識別）に役立つ程度によって行なう。

音声データを 4 kHz の低域通過フィルタに通し，8 kHz サンプルングで AD 変換し，AD データの差分をとってプレエンファシスを行なう。LPC 分析を 32 msec のハミング・ウィンドをかけ，8 msec ステップで行なう（図 5-4）。なお，入力音声は，あらかじめデジタル・ソナグラムによる視察で，各フレームごとに音韻名がつけられている。

発声者は男性 4 名（KS, KI, SF, MK）で，各自，表 5-4 の音声データ（データ 1 ~ 3）を発声した。

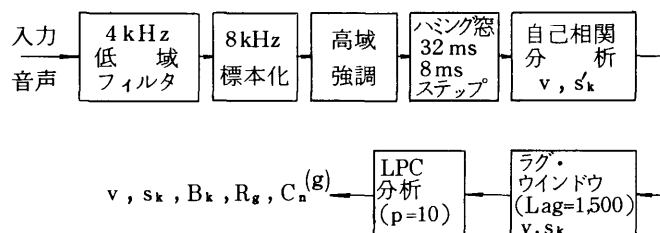


図 5-4 LPC 分析のブロック図

この音声データを用いて母音の識別実験を行なった。

標準パターンは，5 母音と鼻音に対して，「ア，イ，ウ，エ，オ，ン」と区切って発声した単独母音から発声者ごとに作成した。標準パターンの例を図 5-5 に示す。

識別実験の手順は，まず，標準パターンの音韻名（ / X / とする ）のラベルがつけられている分析フレームについて，尺度の値のヒストグラムを求める。次に，その音韻 / X / と / X / に隣接する過渡区間を除いたフレームのヒストグラムを求める。この 2 つのヒストグラムの度数が等しくなるように正規化する。この手順で得られたヒストグラム上で，誤り率が最小になるようにいき値 (z_T) を設定し，この誤り率で尺度の評価を行なう（図 5-6 参照）。

表 5-2 に示した各尺度で母音の識別実験を行なった。その男性 4 名の平均の結果

を表5-5に示す。ケプストラム距離では、ケプストラム係数を16次まで利用して計算した。5母音と鼻音に対する平均で見ると、COSH尺度のCOSH₃（残差パワ正規化）とCOSH₄（COSHの値の最小化）の誤り率が小さいことがわかる。また、表5-2の尺度の計算式からわかるようにCOSH₃は $\sum A_k s_k / R_f$ と $\sum B_k r_k / R_g$ の算術平均であり、COSH₄は幾何平均である。よって、計算はCOSH₃の方がわずかに簡

表5-4 音声データ

データ1	すべての音韻の出現を考慮した発声データ。58単語、10文節。 (約3分)
データ2	主として、無声化母音、長母音、連続母音、半母音、抑音を含む発声データ。81単語、24文節。 (約5分)
データ3	新幹線の座席予約文(付録3のグループ3) (約7分)

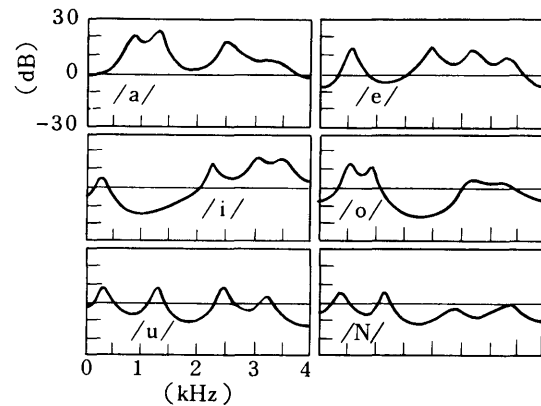


図5-5 発声者KIの標準パターン

単である。さらに、この2項の値が近いときにはCOSH₃とCOSH₄はきわめて近い値をとることもわかる。図5-2、図5-3、表5-3からも、COSH₃とCOSH₄のマッチングの状態がきわめて近いことがわかる。

音韻別の発声者ごとの各尺度の誤り率を表5-6に示す。また、表5-7にケプストラムの係数を8次、16次、32次まで計算したときの誤り率(発声者KS)を示す。

表5-5からわかるように、音韻/A/, /U/, /N/では、必ずしもCOSH₃、COSH₄がよいとはいえない。これらの音韻では、パワー情報を用いた尺度(COSH₁、LR₁)の誤り率が小さいことがわかる。これら/A/, /U/, /N/の音韻でも、パワー情報を適切に用いれば、COSH₃の誤り率も、COSH₁あるいはLR₁の誤り率よりも小さくできる可能性がある。これの簡単な検討として、図5-7に示す手順でパワー情報の利用を試みた。まず、パワー情報を用いたCOSH尺度COSH₁でスペクトルの差を計算し、適当ないき値(y_T)以上の値をとる入力フレームは、その音韻/X/でないとする。y_T以下の値のフレームは、さらにD₃でスペクトルの差を計算する。この手順の有効性を発声者KIの音声データで調べた。この結果を表5-8に示す。/N/以外の音韻で、もっとも誤り率が小さい尺度よりも小さい誤り率が得ら

れた。鼻音 /N/ については、 LR_1 よりも小さい誤り率を得ることができなかった。この原因として、/N/ のスペクトルの不安定さと、/N/ の標準パターンのパワー値が、単独に発声した /N/ から作成したため大きすぎたことが考えられる。

以上より、COSH 尺度 $COSH_3$ は、パワー情報を適切に用いると、すべての母音に対して、絶対的なスペクトルの差を表わす妥当な尺度になり得ると考えられる。

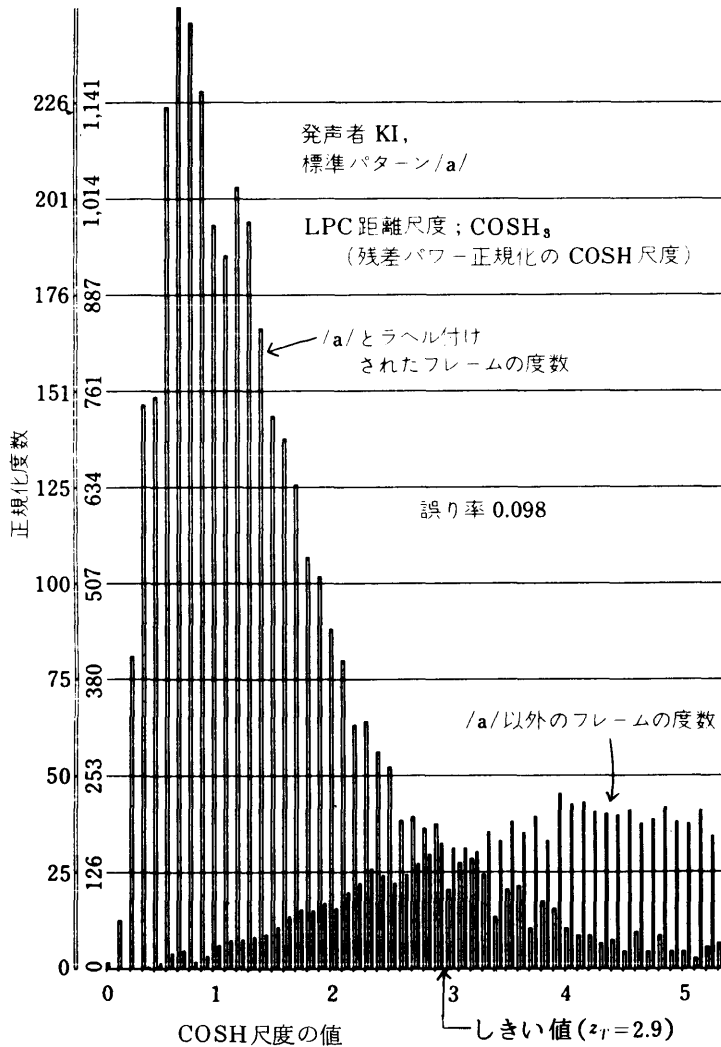


図 5-6 COSH 尺度によるヒストグラムの例

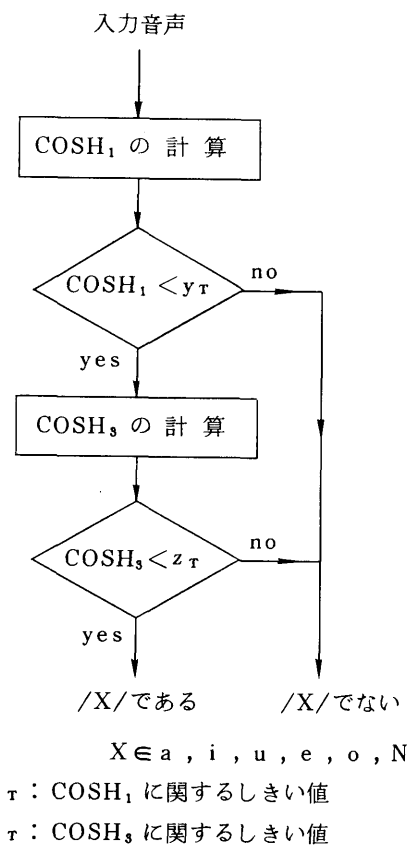


図 5-7 パワー情報を利用した $COSH_3$ による識別手順

表 5-6 LPC距離尺度による識別実験の結果（誤り率）

パワーの設定条件		(I) ...			(II) $u = v$			(III) $uR_f = vR_g$				(IV) 尺度の値最小	
尺 度		LR	COSH	CEP	LR	COSH	CEP	NR	LR	COSH	CEP	LR	COSH
音	/a/	0.244	<u>0.142</u>	0.152	0.202	0.178	0.187	0.214	0.213	0.188	0.200	0.213	0.190
	/i/	0.119	0.153	0.142	0.106	0.101	0.105	0.108	<u>0.090</u>	<u>0.099</u>	<u>0.095</u>	<u>0.090</u>	<u>0.095</u>
	/u/	0.344	<u>0.237</u>	0.242	0.314	0.265	0.261	0.355	0.309	0.258	0.269	0.309	0.262
韻	/e/	0.187	0.129	0.141	0.108	<u>0.084</u>	0.088	0.127	0.097	0.083	0.087	0.097	<u>0.083</u>
	/o/	0.221	0.155	0.166	0.126	0.090	<u>0.084</u>	0.133	0.116	0.085	0.087	0.116	<u>0.086</u>
	/N/	<u>0.208</u>	0.324	0.355	0.307	0.285	0.289	0.345	0.282	0.280	0.282	0.282	0.280
平均値		0.221	0.190	0.197	0.194	0.167	0.169	0.214	0.184	0.166	0.170	0.184	0.166

下線は比較的好いとみなされる誤り率

表 5-7 ケプストラム距離の次数と誤り率

phoneme	order	data 1	data 2	data 3	average
/a/	1 ~ 8	0.311	0.366	0.398	0.377
	1 ~ 16	0.308	0.362	0.386	0.368
	1 ~ 32	0.307	0.361	0.384	0.367
/i/	1 ~ 8	0.062	0.096	0.122	0.104
	1 ~ 16	0.061	0.092	0.120	0.101
	1 ~ 32	0.060	0.092	0.120	0.101
/u/	1 ~ 8	0.264	0.292	0.343	0.307
	1 ~ 16	0.252	0.279	0.330	0.294
	1 ~ 32	0.260	0.284	0.333	0.299
/e/	1 ~ 8	0.065	0.086	0.105	0.090
	1 ~ 16	0.060	0.074	0.102	0.084
	1 ~ 32	0.060	0.071	0.102	0.083
/o/	1 ~ 8	0.141	0.102	0.142	0.133
	1 ~ 16	0.137	0.096	0.135	0.127
	1 ~ 32	0.137	0.097	0.136	0.128
/N/	1 ~ 8	0.265	0.309	0.395	0.359
	1 ~ 16	0.264	0.308	0.389	0.355
	1 ~ 32	0.266	0.306	0.388	0.354

(発声者 KS)

表 5-8 パワー情報と COSH₃を併用したときの誤り率の改善

	COSH ₃ の誤り率 (y _T)	最良の尺度による 誤り率(最良の尺度)
/ a /	0.065 (10)	0.067 (COSH ₁)
/ i /	0.060 (50)	0.065 (LR ₃)
/ u /	0.153 (10)	0.185 (COSH ₄)
/ e /	0.051 (50)	0.070 (COSH ₃)
/ o /	0.050 (50)	0.051 (COSH ₃)
/ N /	0.282 (30)	0.211 (LR ₁)

(発声者 KI)

5.3.2 標準パターンについての検討

前節の母音の識別実験では、単独発声した母音から作成した標準パターンを用いた。この節では、標準パターンをかえたときの COSH 尺度 COSH₃ の誤り率について検討する。標準パターンの作成方法として、次の 4 通りを考える。

- (1) 単独に発声した母音から作成(/ X /)。
- (2) Lag Window⁽⁸⁹⁾ のパラメータ(L)を 200 あるいは 50 にして標準パターン / X / を平滑化。
- (3) 連続音声 / ieaou N / から標準パターンを作成(/ ieaou N /)。
- (4) 多量の音声データから相関係数上での平均の標準パターンを作成(average)。

これら(1)~(4)の方法で作成した標準パターンを用いて、男性 4 名のデータ 1 とデータ 2 の音声データ(表 5-4)を用いて、COSH 尺度の COSH₃ による識別実験を行った。方法(2)で作成した標準パターンを用いると、/ A / ~ / N / のすべての音韻で誤り率が大きくなり、単純な平滑化は効果のないことがわかった。この実験における、発声者 KI のデータ 1, データ 2 に対する誤り率を表 5-9 に示す。

表 5-9 ラグ窓による平滑化による誤り率の変化

lag =		スペクトルの平滑化		
		1500	200	50
母音	データ	誤り率		
/a/	1	0.055	0.058	0.112
	2	0.088	0.079	0.113
/i/	1	0.058	0.058	0.064
	2	0.059	0.063	0.078
/u/	1	0.160	0.170	0.199
	2	0.176	0.185	0.207
/e/	1	0.059	0.079	0.202
	2	0.058	0.092	0.235
/o/	1	0.066	0.067	0.075
	2	0.048	0.047	0.053
/N/	1	0.276	0.293	0.301
	2	0.268	0.277	0.270

(尺度 COSH₃, 発声者 KI)

次に、方法(3)、(4)によって作成した標準パターンを用いて識別実験を行なった。その結果を図 5-8 に示す。その結果、母音 /I/, /E/, /O/ については、単独母音から作成した標準パターンでも、良好な識別結果が得られていることがわかる。母音 /A/ については、4 名中 3 名については、単独母音から作成した標準パターンで十分な識別結果が得られたが、発声者 KS のみ、平均した /A/ を用いることによる識別率の向上が著しかった。母音 /U/ と鼻音 /N/ については、平均の標準パターンを作成したり、連続音声 /i e a o u N/ から作成する

ことが有効であることがわかる。しかしながら、/U/, /N/ については、誤り率がかなり大きく、識別方法についてさらに検討する必要がある。

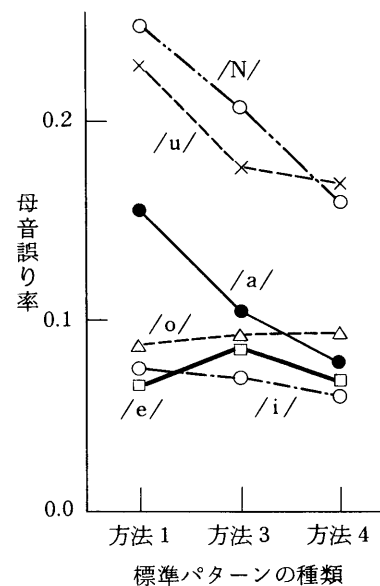


図 5-8 標準パターンの種類に対する COSH₃ の誤り率

5.3.3 まとめ

LPC分析を介して得られるスペクトル包絡のスペクトルの差を表わす距離尺度について検討した。尺度として、最尤スペクトル距離，正規化残差，COSH尺度，ケプストラム距離をとりあげた。マッチングの様子を観察からは，残差パワーを正規化したCOSH尺度 $COSH_3$ とケプストラム距離 CEP_3 と，さらに尺度の値が最小になるようにパワー比を設定したCOSH尺度 $COSH_4$ がすぐれているように思われる。

さらに，母音の識別実験によって各尺度の性能を調べた。その結果として，COSH尺度 $COSH_3$ ， $COSH_4$ が比較的すぐれていることがわかった。計算の簡単さも考慮すると $COSH_3$ がもっともよいと考えられる。

パワー情報をCOSH尺度 $COSH_1$ の形で，前処理的に用いると， $COSH_3$ の誤り率をさらに小さくできることが確認できた。

標準パターンの作成方法についても検討し，母音 / I /， / E /， / O / については，単独に発声した母音から作成した標準パターンで十分であることが推測できた。

これらの尺度を利用して，音声認識システムの音響処理の性能を向上させる可能性が確かめられた。

5.4 Peak Weighted Measureの導出⁽⁹⁰⁾⁽⁹¹⁾⁽⁹²⁾

音声に関する聴覚の特性として，スペクトルの谷形部分に比べて，山形部分 (peak) により敏感であるといわれている。⁽⁴⁾⁽⁸⁸⁾ 従って音声認識においても，スペクトルのピークに重みをおくスペクトル・マッチング尺度は，重要であると考えられる。

本節では，既存のLPCスペクトル・マッチング尺度からスペクトル包絡のピークに重みをおく新しい尺度を導出する。

既存のスペクトル・マッチング尺度を次にあげる。ここで， f を標準パターンのLPCスペクトル包絡， g を入力音声のLPCスペクトル包絡とする。

(a) LPCケプストラム距離

$$CEP = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} (\log f - \log g)^2 d\theta \quad (5-27)$$

(b) COSH尺度

$$\text{COSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{g}{f} + \frac{f}{g} - 2 \right) \cdot d\theta \quad (5-28)$$

(c) 最尤スペクトル距離

$$\text{LR} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\log \frac{f}{g} + \frac{g}{f} - 1 \right) \cdot d\theta \quad (5-29)$$

(d) LPC 相関距離

$$\text{COR} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f - g)^2 \cdot d\theta \quad (5-30)$$

上式のうち、すでに、LPC 相関距離は、対数スペクトルを基準とすれば、ピークに重みをおいた尺度であると言えることができる。

Peak Weighted 尺度は、上述の LPC スペクトル・マッチング尺度の被積分関数に、ピークに対応する重みをかけて積分することにより導出できる。

LPC スペクトル包絡 f 、 g がスペクトルの傾きがないように平坦化されているとき、パワーで正規化された関数 f/u 、 g/v の値の大小は、 f 、 g のピークでは大きく、谷では小さくなる傾向にある。この性質を用いて、重み関数として、次のものを取りあげる。

$$W_1(\theta) = \frac{1}{2} (f(\theta)/u + g(\theta)/v) \quad (5-31)$$

$$W_2(\theta) = (f(\theta) + g(\theta)) / (u + v) \quad (5-32)$$

$$W_3(\theta) = f(\theta) / u \quad (5-33)$$

$$W_4(\theta) = g(\theta) / v \quad (5-34)$$

以下、LPC スペクトル・マッチング尺度からの Peak Weighted Measure の導出を行なう。これらの導出では、LPC ケプストラム係数 ($C_n^{(f)}$ 、 $C_n^{(g)}$) や、LPC 相関係数 (r_n 、 s_n) が、比較的次元で 0 に収束することを利用する。⁽⁹¹⁾

5.4.1 LPC ケプストラム距離からの導出

(i) Symmetric Normalized weighted CEP strum (SNCEP)

$$\text{SNCEP} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\log f(\theta) - \log g(\theta))^2 \cdot \frac{1}{2} \left(\frac{f(\theta)}{u} + \frac{g(\theta)}{v} \right) d\theta \quad (5-35)$$

$$= h_0 + \sum_{n=1}^{\infty} (r_n + s_n) h_n \quad (5-36)$$

なぜなら

$$d_n = C_n^{(f)} - C_n^{(g)}$$

とすると、被積分項のケプストラムの項は、

$$\begin{aligned} & (\log f - \log g)^2 \\ &= \left(\sum_{n=-\infty}^{\infty} d_n e^{jn\theta} \right)^2 = \sum_{n=-\infty}^{\infty} \left(\sum_{s-t=n} d_s d_t \right) e^{jn\theta} \\ &= \sum_{n=-\infty}^{\infty} h_n e^{jn\theta} \end{aligned} \quad (5-37)$$

となる。ただし、 h_n は

$$h_n = \sum_{s-t=n} d_s d_t \quad (5-38)$$

である。また、重み関数 $W_1(\theta)$ は、

$$\begin{aligned} W_1(\theta) &= \frac{1}{2} \left(\frac{f}{u} + \frac{g}{v} \right) \\ &= \frac{1}{2} \sum_{n=-\infty}^{\infty} (r_n + s_n) e^{jn\theta} \end{aligned} \quad (5-39)$$

となる。よって、(5-38)式、(5-39)式から、(5-36)式が導びかれる。以上の式の導出では、次式で計算されるLPC相関係数(r_n, s_n)を用いる。

$$\left. \begin{aligned} r_n &= - \sum_{i=1}^p a_i r_{n-i} \\ s_n &= - \sum_{i=1}^p b_i s_{n-i} \end{aligned} \right\} \quad (5-40)$$

さらに、尺度の値が最小になるようにパワー比を設定する。(5-36)式をパワーを含む d_0 で偏微分して0とおく。

$$2d_0 + 2 \sum_{n=1}^{\infty} d_n (r_n + s_n) = 0 \quad (5-41)$$

よって、

$$d_0 = - \sum_{n=1}^{\infty} d_n (r_n + s_n) \quad (5-42)$$

のとき, SNCEPは最小値をとる。

5.4.2 COSH尺度からの導出

(i) Symmetric Normalized weighted COSH (SNCOSH)

$$\text{SNCOSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{g}{f} + \frac{f}{g} - 2 \right) \cdot \frac{1}{2} \left(\frac{f}{u} + \frac{g}{v} \right) d\theta \quad (5-43)$$

$$= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{1}{u} \frac{f^2}{g} + \frac{1}{v} \frac{g^2}{f} \right) d\theta + \frac{1}{2} \left(\frac{v}{u} + \frac{u}{v} \right) - 2$$

$$= \frac{1}{2} \left(\frac{u}{vR_g} \sum_{n=-p}^p B_n \tilde{r}_n + \frac{v}{uR_f} \sum_{n=-p}^p A_n \tilde{s}_n + \frac{1}{2} \left(\frac{v}{u} + \frac{u}{v} \right) - 2 \right) \quad (5-44)$$

ここで, \tilde{r}_n , \tilde{s}_n は次の様に定義される。

$$\tilde{r}_n = \frac{1}{u^2} \int_{-\pi}^{\pi} f^2 e^{-jn\theta} \frac{d\theta}{2\pi}$$

$$= \sum_{s-t=n} r_s r_t = \sum_{t=-\infty}^{\infty} r_t r_{t+n} \quad (5-45)$$

同様に,

$$\tilde{s}_n = \sum_{t=-\infty}^{\infty} s_t s_{t+n} \quad (5-46)$$

\tilde{r}_n , \tilde{s}_n は, LPC相関係数 (r_n , s_n) が, 有限項で0に収束するため, 近似的に計算することが可能である。

(5-44)式で, $u/v = x$ とおいて, x で偏微分し, 0とおくと,

$$x = \sqrt{\frac{\frac{1}{R_f} \sum_{n=-p}^p A_n \tilde{s}_n + 1}{\frac{1}{R_g} \sum_{n=-p}^p B_n \tilde{r}_n + 1}} \quad (5-47)$$

のとき, SNCOSHの値は最小となる。

(ii) Symmetric Weighted COSH (SWCOSH)

$$\text{SWCOSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{g}{f} + \frac{f}{g} - 2 \right) \left(\frac{f}{u} + \frac{g}{v} \right) d\theta \quad (5-48)$$

$$= \frac{1}{u+v} \left(\frac{u^2}{vR_g} \sum_{n=-p}^p B_n \tilde{r}_n + \frac{v^2}{uR_f} \sum_{n=-p}^p A_n \tilde{s}_n \right) - 1 \quad (5-49)$$

(iii) Unsymmetrical Weighted COSH (UWCOSH)

$$\text{UWCOSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{g}{f} + \frac{f}{g} - 2 \right) \frac{f}{u} d\theta \quad (5-50)$$

$$= \frac{u}{vR_g} \sum_{n=-p}^p B_n \tilde{r}_n + \frac{v}{u} - 2 \quad (5-51)$$

f と g とに関して、非対称な関数なので、スペクトル値が最小になるようにパワー比を設定して用いるには問題がある。ここでは、残差パワーを正規化した。

$$u/v = R_g / R_f$$

の条件での式を示しておく。

$$\text{UWCOSH} = \frac{R_f}{R_g} + \frac{1}{R_f} \sum_{n=-p}^p B_n \tilde{r}_n - 1 \quad (5-52)$$

(iv) Symmetric Multiplicative weighted COSH (SMCOSH)

重み関数 $W_5(\theta)$ として、

$$W_5(\theta) = \frac{f(\theta) \cdot g(\theta)}{\frac{1}{2\pi} \int_{-\pi}^{\pi} f \cdot g d\theta} \quad (5-53)$$

$$= \frac{f \cdot g}{u \cdot v \sum_{n=-\infty}^{\infty} r_n s_n} \quad (5-54)$$

を考える。

$$\text{SMCOSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{g}{f} + \frac{f}{g} - 2 \right) \cdot \frac{f \cdot g}{u \cdot v \sum_{n=-\infty}^{\infty} r_n s_n} d\theta \quad (5-55)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{(f-g)^2}{u \cdot v \sum_{n=-\infty}^{\infty} r_n s_n} d\theta \quad (5-56)$$

$$= \frac{\sum_{n=-\infty}^{\infty} \left(\frac{u}{v} r_n - s_n \right)^2}{\frac{u}{v} \sum_{n=-\infty}^{\infty} r_n s_n} \quad (5-57)$$

スペクトルの差が最小になるように，パワー比を定めると，パワー比は，

$$\frac{u}{v} = \sqrt{\frac{\sum_{n=-\infty}^{\infty} s_n^2}{\sum_{n=-\infty}^{\infty} r_n^2}} \quad (5-58)$$

となる。このときの尺度値は，

$$\text{SMCOSH} = 2 \left(\frac{\sqrt{\sum_{n=-\infty}^{\infty} r_n^2 \sum_{n=-\infty}^{\infty} s_n^2}}{\sum_{n=-\infty}^{\infty} r_n s_n} - 1 \right) \quad (5-59)$$

となる。(5-59)式は，相関係数だけで尺度値が計算できる。もし，尺度値の近似がLPC分析のp次以下でも充分であるならば，LPC分析を必要としなくなり，計算量の少ない有効な尺度となり得る。

5.4.3 最尤スペクトル距離からの導出

最尤スペクトル距離から COSH 尺度を導出する過程 (5.2.3 節) と同様にして，Peak weighted measure を導出する。

(i) Weighted Likelihood Ratio (WLR)

$$\text{WLR} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \left(\log \frac{f}{g} + \frac{g}{f} - 1 \right) \cdot \frac{f}{u} + \left(\log \frac{g}{f} + \frac{f}{g} - 1 \right) \cdot \frac{g}{v} \right\} d\theta \quad (5-60)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f}{u} - \frac{g}{v} \right) (\log f - \log g) d\theta + \frac{v}{u} + \frac{u}{v} - 2 \quad (5-61)$$

$$= \sum_{n=-\infty}^{\infty} (r_n - s_n) (C_n^{(f)} - C_n^{(g)}) + \frac{v}{u} + \frac{u}{v} - 2 \quad (5-62)$$

尺度の値は， $u = v$ のとき最小となり，

$$\text{WLR} = 2 \cdot \sum_{n=1}^{\infty} (r_n - s_n) (C_n^{(f)} - C_n^{(g)}) \quad (5-63)$$

となる。WLRは，Peak weighted measure のうち，計算量も少なく有効な尺度である。

(ii) Symmetric Weighted EXP onential likelihood ratio (SWEXP)

$$\text{SWEXP} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \left(\log \frac{f}{g} + \frac{g}{f} - 1 \right) \frac{g}{v} + \left(\log \frac{g}{f} + \frac{f}{g} - 1 \right) \frac{f}{u} \right\} d\theta \quad (5-64)$$

$$\begin{aligned}
&= 2 \cdot \sum_{n=1}^{\infty} (r_n - s_n) (C_n^{(g)} - C_n^{(f)}) \\
&+ \frac{u}{v \cdot R_g} \sum_{n=-p}^p B_n \tilde{r}_n + \frac{v}{u \cdot R_f} \sum_{n=-p}^p A_n \tilde{s}_n - 2 \quad (5-65)
\end{aligned}$$

5.4.4 LPC 相関距離

LPC 相関距離 (COR) も Peak weighted measure の一種と考えることができる。パワーを 1 に正規化した LPC 相関距離は次式で表わされる。

$$\text{COR} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f/u - g/v) d\theta \quad (5-66)$$

$$= 2 \cdot \sum_{n=1}^{\infty} (r_n - s_n)^2 \quad (5-67)$$

以上の導出した尺度を，従来からの LPC スペクトルマッチング尺度とともに表 5-10 にまとめておく。

表5-10 LPCスペクトル・マッチング尺度

	Integrand	Power use	Power normalize	Residual normalize	Minimal condition	Remark
CEP	$(\log f - \log g)^2$	$(C_o^{(f)} - C_o^{(g)})^2$ $+ 2 \sum_{n=1}^{\infty} (C_n^{(f)} - C_n^{(g)})^2$	$(\log R_f - \log R_g)^2$ $+ 2 \sum_{n=1}^{\infty} (C_n^{(f)} - C_n^{(g)})^2$	$2 \sum_{n=1}^{\infty} (C_n^{(f)} - C_n^{(g)})^2$	$2 \sum_{n=1}^{\infty} (C_n^{(f)} - C_n^{(g)})^2$	
SNCEP	$(\log f - \log g)^2$ $\times \frac{1}{2} \left(\frac{f}{u} + \frac{g}{v} \right)$	$e_o + \sum_{n=1}^{\infty} e_o(r_n + s_n)$	$(\log \frac{R_f}{R_g})^2 + \sum_{n=1}^{\infty} d_n^2 + \sum_{n=1}^{\infty} (r_n + s_n)(2d_n \log \frac{R_f}{R_g} + h_n)$	$\sum_{n=1}^{\infty} d_n^2$ $+ \sum_{n=1}^{\infty} (r_n - s_n) h_n$	$d_o = - \sum_{n=1}^{\infty} d_n (r_n - s_n)$ とおくと, $d_o^2 + 2 \sum_{n=1}^{\infty} d_n^2 + \sum_{n=1}^{\infty} (2d_o d_n + q_n)(r_n + s_n)$	$d_k = C_k^{(f)} - C_k^{(g)}, e_n = \sum_{k=-\infty}^{\infty} d_k d_{k+n}$ $q_n = 2 \sum_{k=1}^{\infty} d_k d_{k+n} + \sum_{k=1}^{n-1} d_k d_{n-k}$
COSH	$\frac{g}{f} + \frac{f}{g} - 2$	$\frac{v}{u R_f} \sum_{n=-p}^p A_n s_n$ $+ \frac{u}{v R_g} \sum_{n=-p}^p B_n r_n - 2$	$\frac{1}{R_f} \sum_{n=-p}^p A_n s_n$ $+ \frac{1}{R_g} \sum_{n=-p}^p B_n r_n - 2$	$\frac{1}{R_g} \sum_{n=-p}^p A_n s_n$ $+ \frac{1}{R_f} \sum_{n=-p}^p B_n r_n - 2$	$2 \left(\sqrt{\frac{\sum A_n s_n \cdot \sum B_n r_n}{R_f \cdot R_g}} - 1 \right)$	
SNCOSH	$\left(\frac{g}{f} + \frac{f}{g} - 2 \right) \frac{1}{2}$ $\times \left(\frac{f}{u} + \frac{g}{v} \right)$	$\frac{1}{2} \left\{ \frac{u}{v R_g} \sum_{n=-p}^p B_n \tilde{r}_n + \frac{v}{u R_f} \sum_{n=-p}^p A_n \tilde{s}_n + \frac{v}{u} + \frac{u}{v} \right\} - 2$	$\frac{1}{2} \left\{ \frac{1}{R_g} \sum B_n \tilde{r}_n + \frac{1}{R_f} \sum A_n \tilde{s}_n \right\} - 1$	$\frac{1}{2} \left\{ \frac{1}{R_f} \sum B_n \tilde{r}_n + \frac{1}{R_g} \sum A_n \tilde{s}_n + \frac{R_f}{R_g} + \frac{R_g}{R_f} \right\} - 2$	$\left(\frac{1}{R_g} \sum B_n \tilde{r}_n + 1 \right)^{\frac{1}{2}}$ $\times \left(\frac{1}{R_f} \sum A_n \tilde{s}_n + 1 \right)^{\frac{1}{2}} - 2$	\tilde{r}, \tilde{s} ; LPC 相関係数の自己相関
SWCOSH	$\left(\frac{g}{f} + \frac{f}{g} - 2 \right) \frac{f+g}{u+v}$	$\frac{1}{u+v} \left(\frac{u^2}{v R_g} \sum_{n=-p}^p B_n \tilde{r}_n + \frac{v^2}{u R_f} \sum_{n=-p}^p A_n \tilde{s}_n \right) - 1$	$\frac{1}{2} \left(\frac{1}{R_g} \sum B_n \tilde{r}_n + \frac{1}{R_f} \sum A_n \tilde{s}_n \right) - 1$	$\frac{1}{R_f + R_g} \left(\frac{R_g}{R_f} \sum B_n \tilde{r}_n + \frac{R_f}{R_g} \sum A_n \tilde{s}_n \right) - 1$	未定	
UWCOSH	$\left(\frac{g}{f} + \frac{f}{g} - 2 \right) \cdot \frac{f}{u}$	$\frac{u}{v R_g} \sum_{n=-p}^p B_n \tilde{r}_n + \frac{v}{u} - 2$	$\frac{1}{R_g} \sum B_n \tilde{r}_n - 1$	$\frac{1}{R_f} \sum B_n \tilde{r}_n + \frac{R_f}{R_g} - 2$	$2 \left(\sqrt{\frac{1}{R_g} \sum B_n \tilde{r}_n} - 1 \right)$	

表 5-10 (続き)

	Integrand	Power use	Power normalize	Residual normalize	Minimal condition	Remark
SMCOSH	$\left(\frac{g}{f} + \frac{f}{g} - 2\right) \frac{fg}{r_{fg}}$ $= \frac{1}{r_{fg}} (f - g)^2$	$\frac{\sum_{n=-\infty}^{\infty} \left(\frac{u}{v} r_n - s_n\right)^2}{\frac{u}{v} \sum_{n=-\infty}^{\infty} r_n s_n}$	$\frac{\Sigma (r_n - s_n)^2}{\Sigma r_n s_n}$	$\frac{\Sigma \left(\frac{R_g}{R_f} r_n - s_n\right)^2}{\frac{R_g}{R_f} \Sigma r_n s_n}$	$2 \left(\frac{\sqrt{\Sigma r_n^2 \Sigma s_n^2}}{\Sigma r_n s_n} - 1 \right)$	$r_{fg} = uv \sum_{n=-\infty}^{\infty} r_n s_n$ $= \int_{-\pi}^{\pi} fg \cdot \frac{d\lambda}{2\pi}$
L R	$\log \frac{f}{g} + \frac{g}{f} - 1$	$\log \frac{u}{v} \frac{R_f}{R_g} + \frac{v}{u R_f}$ $\times \sum_{n=-p}^p A_n s_n - 1$	$\log \frac{R_t}{R_g} + \frac{1}{R_f} \Sigma A_n s_n$ $- 1$	$\frac{1}{R_g} \Sigma A_n s_n - 1$	$\log \left(\frac{1}{R_g} \Sigma A_n s_n \right)$	
WLR	$\left(\log \frac{f}{g} + \frac{g}{f} - 1\right) \frac{f}{u} +$ $\left(\log \frac{g}{f} + \frac{f}{g} - 1\right) \frac{g}{v}$	$\sum_{n=-\infty}^{\infty} (r_n - s_n)(C_n^{(f)} - C_n^{(g)}) -$ $C_n^{(g)} + \frac{u}{v} + \frac{v}{u} - 2$	$\Sigma (r_n - s_n)(C_n^{(f)} - C_n^{(g)})$	$\Sigma (r_n - s_n)(C_n^{(f)} - C_n^{(g)})$ $+ \frac{R_g}{R_f} + \frac{R_f}{R_g} - 2$	$\Sigma (r_n - s_n)(C_n^{(f)} - C_n^{(g)})$	
SWEXP	$\left(\log \frac{f}{g} + \frac{g}{f} - 1\right) \frac{g}{v}$ $+ \left(\log \frac{g}{f} + \frac{f}{g} - 1\right) \frac{f}{u}$	$\sum_{n=-\infty}^{\infty} (r_n - s_n)(C_n^{(f)} - C_n^{(g)})$ $+ \frac{u}{v R_g} \sum_{n=-p}^p B_n \tilde{r}_n + \frac{v}{u R_f}$ $\times \sum_{n=-p}^p A_n \tilde{s}_n - 2$	$\Sigma (r_n - s_n)(C_n^{(g)} - C_n^{(f)})$ $+ \frac{1}{R_g} \Sigma B_n \tilde{r}_n + \frac{1}{R_f} \Sigma A_n \tilde{s}_n$ $- 2$	$\Sigma (r_n - s_n)(C_n^{(f)} - C_n^{(g)})$ $+ \frac{1}{R_f} \Sigma B_n \tilde{r}_n + \frac{1}{R_g} \Sigma A_n \tilde{s}_n$ $- 2$	$\Sigma (r_n - s_n)(C_n^{(g)} - C_n^{(f)})$ $+ 2 \sqrt{\frac{\Sigma B_n \tilde{r}_n \Sigma A_n \tilde{s}_n}{R_f \cdot R_g}}$ $- 2$	
COR	$\left(\frac{f}{u} - \frac{g}{v}\right)^2$	$-$	$-$	$-$	$\sum_{n=-\infty}^{\infty} (r_n - s_n)^2$	

5.5 Peak Weighted Measureの評価⁽⁹⁰⁾⁽⁹¹⁾⁽⁹²⁾

5.4節で導出したPeak weighted measureのスペクトル・マッチングの対数スペクトルの差に対する重み(被積分項の値)の解析を5.5.1節で行う。さらに、各尺度の評価を母音識別実験によりPeak weighted measureの性能評価を行う。

5.5.1 ピーク重みづけの解析

Peak weighted measureのピーク重みづけの様子を解析するために、代表的な次の尺度をとりあげる。

$$\text{SNCEP} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\log f - \log g)^2 \cdot \frac{1}{2} \left(\frac{f}{u} + \frac{g}{v} \right) d\theta \quad (5-35)$$

$$\text{SNCOSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{g}{f} + \frac{f}{g} - 2 \right) \cdot \frac{1}{2} \left(\frac{f}{u} + \frac{g}{v} \right) d\theta \quad (5-43)$$

$$\text{WLR} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \left(\log \frac{f}{g} + \frac{g}{f} - 1 \right) \cdot \frac{f}{u} + \left(\log \frac{g}{f} + \frac{f}{g} - 1 \right) \frac{g}{v} \right\} d\theta \quad (5-60)$$

$$\text{SMCOSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{(f-g)^2}{u \cdot v \cdot \sum_{n=-\infty}^{\infty} r_n s_n} d\theta \quad (5-56)$$

スペクトル f , g は平坦化されているとし、 f および g のピークの度合を表わすパラメータとして、

$$r = \frac{1}{2} \left(\frac{f(\theta)}{u} + \frac{g(\theta)}{v} \right)$$

をとりあげる。 r の大きい時には、 $f(\theta)$ あるいは $g(\theta)$ が大きな値をとっていることを意味し $f(\theta)$ あるいは $g(\theta)$ がスペクトル包絡のピーク付近の値をとっていることになる。また、 r の値が小さいときには、 $f(\theta)$ および、 $g(\theta)$ がともに小さい値をとっていることを意味し、 $f(\theta)$ および $g(\theta)$ がスペクトル包絡の谷付近の値をとっていることになる。

以下、対数スペクトルの差

$$V(\theta) = \log f - \log g \quad (5-69)$$

に対する重みの解析を行う。各尺度の被積分項は、 $V(\theta)$ と r とによって、次のように

表わされる。ただし，WLRとSMCOSHではパワーの設定条件を $u = v$ とする。

$$\begin{aligned} \text{SNCEP} &: (\log f - \log g)^2 \cdot \frac{1}{2} \left(\frac{f}{u} + \frac{g}{v} \right) \\ &= r \cdot V(\theta)^2 \end{aligned} \quad (5-70)$$

$$\begin{aligned} \text{SNCOSH} &: \left(\frac{g}{f} + \frac{f}{g} - 2 \right) \cdot \frac{1}{2} \left(\frac{f}{u} + \frac{g}{v} \right) \\ &= r \cdot (e^{V(\theta)} + e^{-V(\theta)} - 2) \end{aligned} \quad (5-71)$$

$$\begin{aligned} \text{WLR} &: \left(\log \frac{f}{g} + \frac{g}{f} - 1 \right) \frac{f}{u} + \left(\log \frac{g}{f} + \frac{f}{g} - 1 \right) \frac{g}{v} \\ &= \left(\frac{f}{u} - \frac{g}{v} \right) \cdot V(\theta) \\ &= 2r \left(\frac{e^{V(\theta)} - 1}{e^{V(\theta)} + 1} \right) V(\theta) \\ &= 2r V(\theta) \cdot \tanh \frac{V(\theta)}{2} \end{aligned} \quad (5-72)$$

$$\begin{aligned} \text{SMCOSH} &: \frac{(f - g)^2}{u \cdot v} \\ &= \left(\frac{g}{f} + \frac{f}{g} - 2 \right) \cdot \frac{f \cdot g}{u \cdot v} \\ &= 4 \cdot r^2 \cdot \tanh^2 \left(\frac{V(\theta)}{2} \right) \end{aligned} \quad (5-73)$$

SNCEP, SNCOSH, WLRは, r の値に比例した重みを持つことがわかる。SMCOSHは, r^2 に比例する重みをもつ。

$f(\theta)$ と $g(\theta)$ の値が近い場合の様子を, $V(\theta)=0$ のまわりでテーラ展開して調べる。

$$\text{SNCEP} : r \cdot V(\theta)^2 \quad (5-70)$$

$$\text{SNCOSH} : r \cdot \left(V(\theta)^2 + \frac{1}{12} V(\theta)^4 + \dots \right) \quad (5-74)$$

$$\text{WLR} : r \left(V(\theta)^2 - \frac{1}{12} V(\theta)^4 + \dots \right) \quad (5-75)$$

$$\text{SMCOSH} : r^2 \left(V(\theta)^2 - \frac{V(\theta)^4}{8} + \dots \right) \quad (5-76)$$

このように、 $V(\theta) = 0$ の付近では、いずれの尺度も $r V(\theta)^2$ の重みをもっていることがわかる。

次に、 $|V(\theta)|$ が、十分大きいところでの様子を解析する。

$$\text{SNCEP} : r \cdot V(\theta)^2 \quad (5-77)$$

$$\text{SNCOSH} : r \cdot e^{|V(\theta)|} \quad (5-78)$$

$$\text{WLR} : 2r \cdot |V(\theta)| \quad (5-79)$$

$$\text{SMCOSH} : 4r^2 \quad (5-80)$$

$|V(\theta)|$ が十分大きいところでは、各尺度の重みは特徴的である。すなわち、SNCOSHは $O(e^{|V(\theta)|})$ のオーダーで、SNCEPは $O(V(\theta)^2)$ 、WLRは $O(|V(\theta)|)$ で無限大に近づく。また、SMCOSHは、 $4r^2$ という定数で押えられる。

従来からのLPCスペクトル・マッチング尺度では、 $|V(\theta)|$ が十分大きいところで、 $O(|V(\theta)|)$ という対数スペクトル差 $|V(\theta)|$ に対して一次の重み $|V(\theta)|$ をもつ尺度は存在していない。この意味からも、WLRは興味深い尺度である。

r をパラメータとして、 $V(\theta)$ に対する各尺度の重みの様子を図5-9から図5-11に示す。図5-9は、 $r = 0.25$ のときで、 f 、 g が谷部に存在するときである。図5-10は、 $r = 1$ で f 、 g が平均部に、図5-11は $r = 4$ で f あるいは g が山部に存在しているときの重みの様子を表わす。さらに、図5-12に、 $r = 0.25, 1.0, 4.0$ のときの重みをまとめて示す。この図から、ピーク重みづけの様子がわかる。

5.5.2 母音識別実験による評価

母音識別実験によってPeak weighted measureの評価を行なう。

実験に用いる音声データは、男性4名が発声した新幹線の座席予約文(付録3の発声リストのグループ1)であり、あらかじめ視察によりフレームごとに音韻名がつけられている。音声データを、4kHzのHamming窓にかけ、8msecのステップで相関分析を行う。さらに、1次系 $(1 - z^{-1})$ で高域強調をかけ、発声ごとに、1次系 $(1 - \alpha z^{-1})$ でスペクトルの平坦化を行なう。ここで α は単独に発声した5母音の第1次の相関係数の平均値である。その後、10次のLPC分析を行う。母音標準パターンは入力音声の中の各母音のフレームを相関係数上で平均した平均化母音である。

次に母音識別実験の手順について説明する。まず、母音標準パターン $/x/$ と音韻名が $/x/$ である入力フレームとの尺度値のヒストグラムを求める。次に、音韻名が

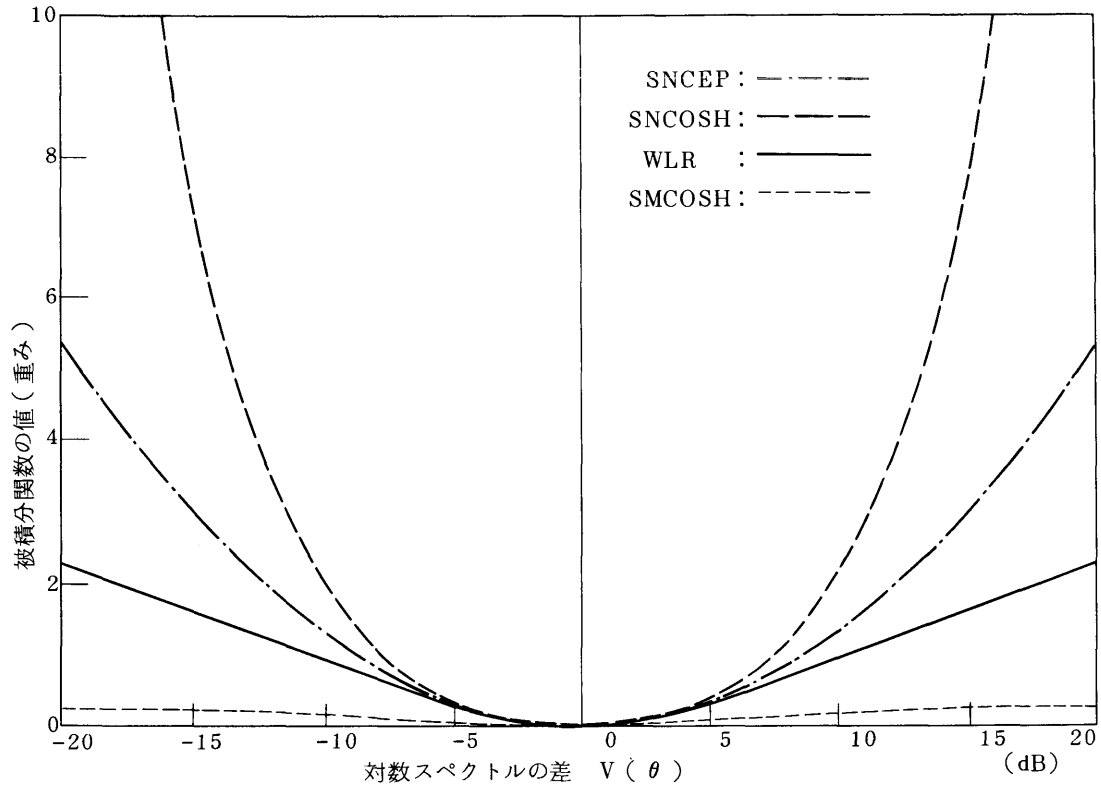


図 5-9 対数スペクトル差 $V(\theta)$ に対する Peak weighted measure の重み。 f, g とも谷の場合 ($r = 0.25$)

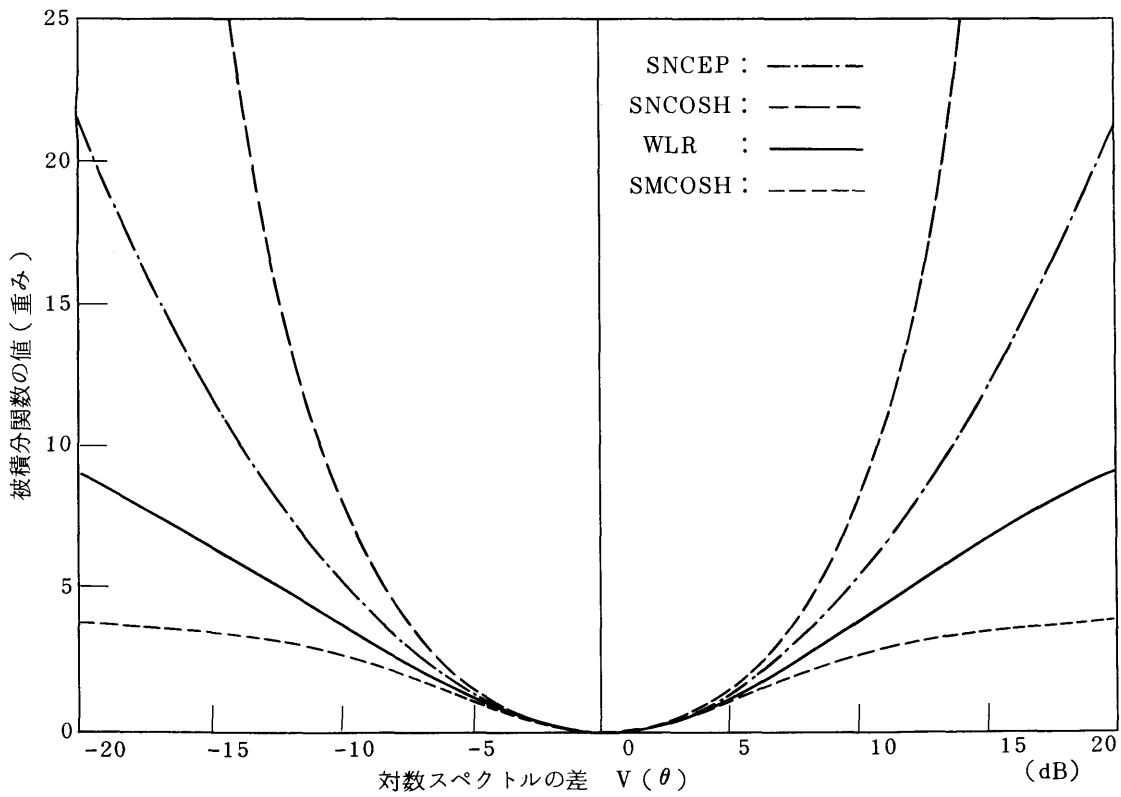


図 5-10 対数スペクトル差 $V(\theta)$ に対する Peak weighted measure の重み。 ($r = 1.0$)

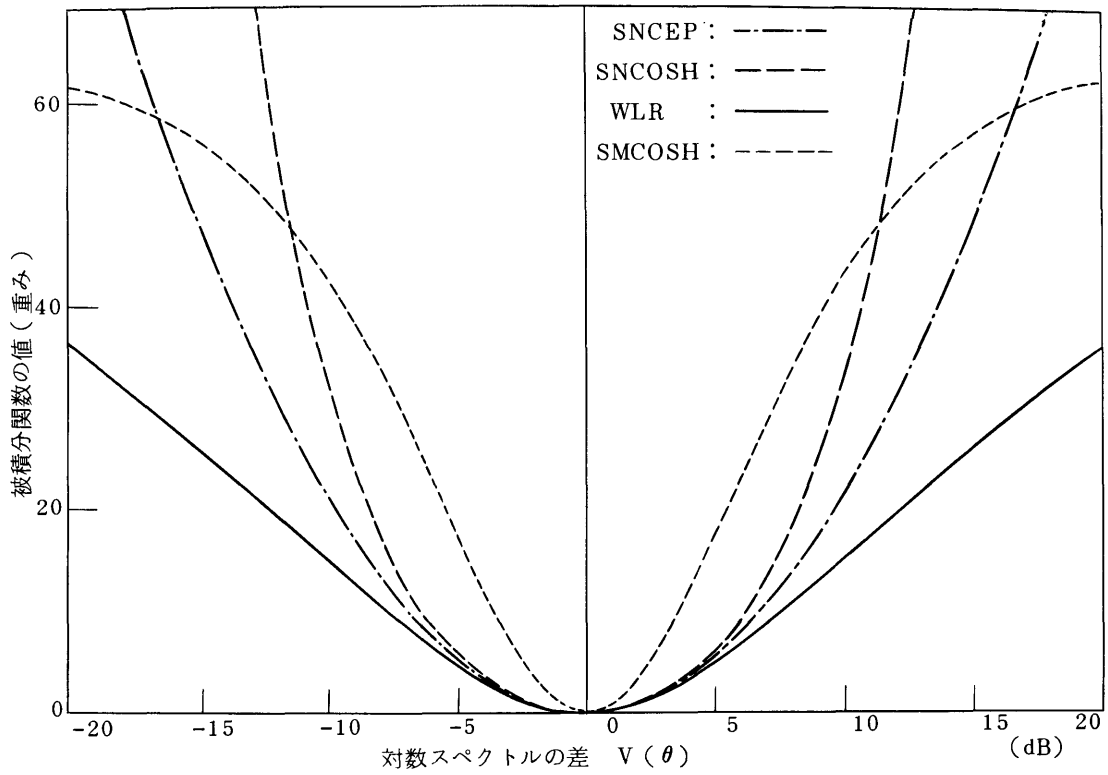
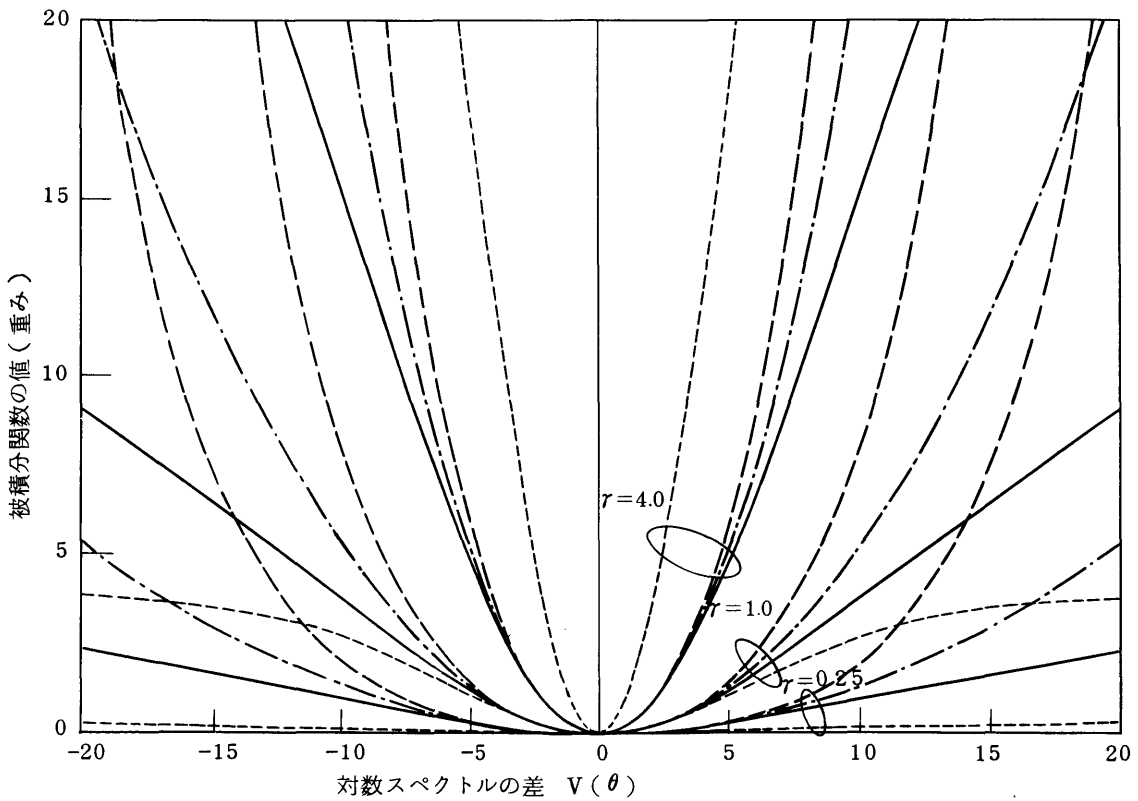


図 5-11 対数スペクトルの差 $V(\theta)$ に対する Peak weighted measure の重み。
 f あるいは g がピーク付近の場合 ($r = 4.0$)



SNCEP: - · - · - · - SNCOSH: - - - - - WLR: ——— SMCOSH: - - - - -

図 5-12 対数スペクトル差 $V(\theta)$ に対する Peak weighted measure の重み。
 $(r = 0.25, 1.0, 4.0)$

／x／でない入力フレームと標準パターン／x／との尺度値のヒストグラムを求める。ただし、両ヒストグラムでは、母音／x／に隣接する過渡区間を除いている。この2つのヒストグラムの各々の度数が等しくなるように正規化した後、このヒストグラム上で誤り率が最小になる様に、いき値（両ヒストグラムの交点）を設定し、このときの誤り率で尺度の評価を行う。ヒストグラムの例を図5-13に示す。

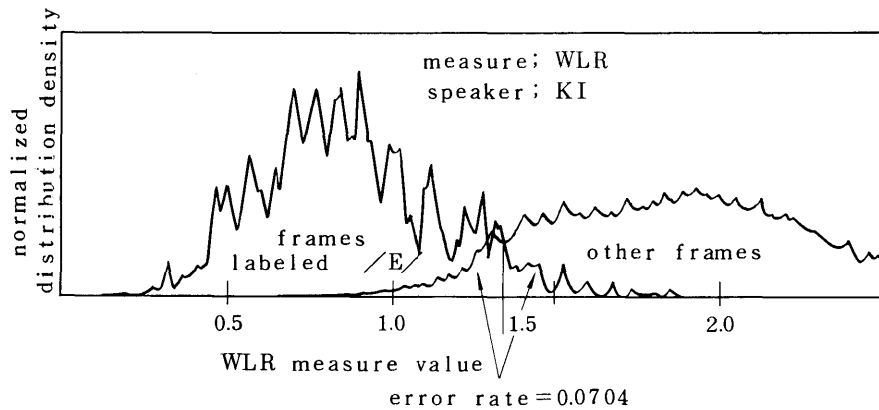


図5-13 母音識別実験でのヒストグラムの例

各尺度値を計算する際、LPC ケプストラム係数 ($C_n^{(f)}$, $C_n^{(g)}$), LPC 相関係数 (r_n , s_n) を有限次数で打ち切る必要が出てくる。LPC スペクトル包絡と有限次数で打ち切った場合のスペクトル表示の比較から、LPC ケプストラム係数は16次まで、相関係数は32次まで、両係数が表われるWLRの場合には24次で打ち切ることとする。

各尺度の5母音の平均の誤り率を表5-11に示す。いずれのPeak weighted measureも既存のLPCスペクトルマッチング

表5-11 母音識別実験によるPeak weighted measureの評価

尺度	誤り率	最良条件
CEP	0.117	Min (= Resid: 正規化)
SNCEP	0.096	Min.
COSH	0.115	Resid. 正規化
SNCOSH	0.106	Min.
SWCOSH	0.107	Power 正規化 (Min 条件が未定)
UWCOSH	0.115	Resid 正規化
SMCOSH	0.099	Min.
LR	0.120	Min.
WLR	0.093	Min. (= Power 正規化)
SWXP	0.108	Min.
COR	0.163	—

尺度 (CEP, COSH, LR) より誤り率が小さく、母音認識には適していることがわか

る。とくに、WLRは、誤り率が最も小さく、かつ、計算量も少ないため、すぐれた尺度であるといえる。母音別の誤り率を図5-14に示す。母音/A/, /U/, /O/に対して、ピークの重みづけが非常に有効であるといえる。また、話者別にみると、図5-15に示されるように、既存のLPC距離尺度では、誤り率の各話者間でのばらつきが非常に大きいですが、Peak weighted measureでは、ばらつきが小さくなっている。以下5.6節では、Peak weighted measureのうち最も有望と推測されるWLRをとりあげ、周波数軸の重み付け、および、スペクトルの平滑化の効果について母音認識の立場から検討を加える。

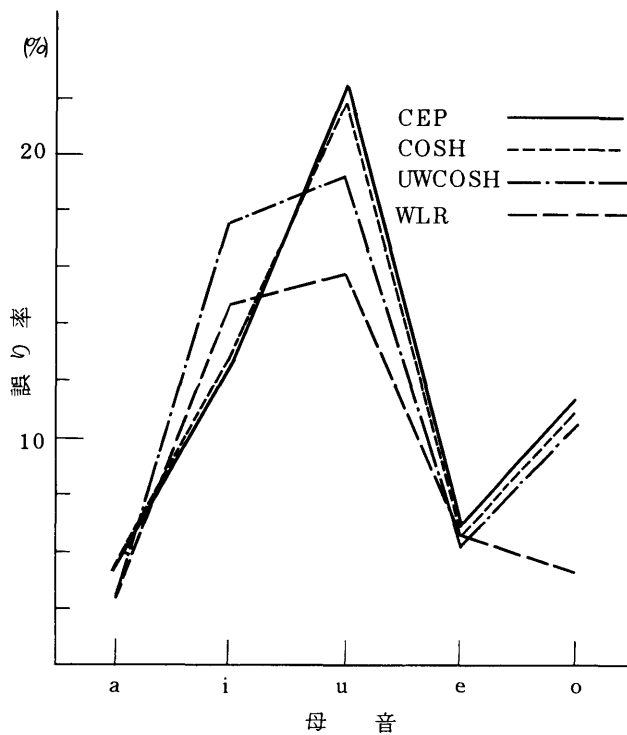


図5-14 母音識別実験における母音別の誤り率

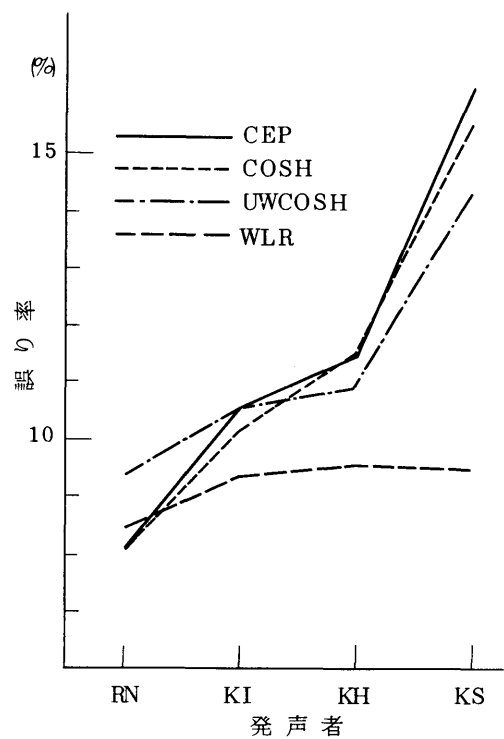


図5-15 母音識別実験における発声者別の誤り率

5.6 Peak Weighted Measureの母音認識における最適化

Peak weighted measureは、ピークに着目したマッチングが可能である上に、周波数軸上での重みづけも可能である。また、スペクトルの平滑化によるピークの重みづけの最適化の可能性もある。この2項目の検討の対象として、WLR(Weighted Likeli-

hood Ratio) 尺度をとりあげる。WLRは、前節の母音識別実験で最もよい母音 識別率を示した尺度である。また、計算量も比較的少ない尺度である。

5.6.1 周波数軸における重みづけの効果 ⁽⁹³⁾

音声認識におけるスペクトル・マッチングの手法として、フィルターバンクによる手法とLPCスペクトル包絡の差による手法とがある。前者は、周波数軸上での重みづけができるという利点がある反面、周波数分解能が悪いという欠点をもっている。WLRをはじめとするPeak weighted measureは、ピークに着目したマッチングが可能である上に、周波数軸上での重みづけも可能である。本節では、Peak weighted measure のうち、計算量の比較的少ないUWCOSH (Unsymmetrical Weighted COSH)とWLR(Weighted Likelihood Ratio)をとりあげ、母音識別実験における周波数重みづけの有効性を示す。

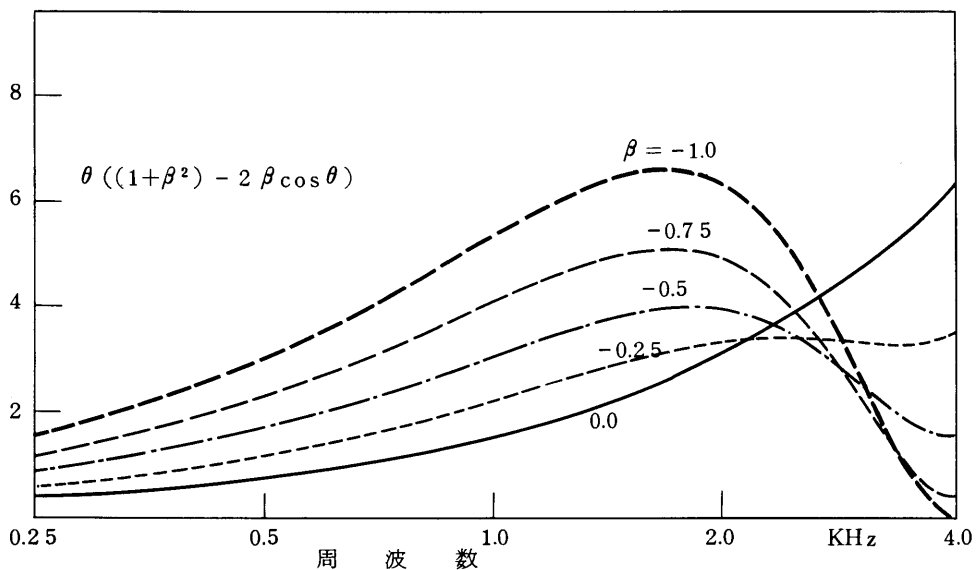


図5-16 1次系 $(1 - \beta z^{-1})$ による周波数軸の重みづけ

Peak weighted measure との比較のため、従来からの尺度として、COSH尺度 (COSH) とLPCケプストラム距離 (CEP) もとりあげる。

$$WLR = \sum_{n=1}^{24} (r_n - s_n)(C_n^{(f)} - C_n^{(g)}) \quad (5-81)$$

(尺度値最小の条件, $u = v$)

$$UWCOSH = \frac{1}{R_f} \sum_{k=-10}^{10} B_k \tilde{r}_k + \frac{R_f}{R_g} - 2 \quad (5-82)$$

(残差パワー正規化の条件, $uR_f = vR_g$)

$$COSH = \sqrt{\frac{\sum_{k=-10}^{10} A_k S_k \cdot \sum_{k=-10}^{10} B_k r_k}{R_f \cdot R_g}} - 1 \quad (5-83)$$

(尺度値最小の条件)

$$CEP = \sum_{i=1}^{16} (C_i^{(f)} - C_i^{(g)})^2 \quad (5-84)$$

(尺度値最小の条件, $uR_f = vR_g$)

(5-50)式からわかるように, UWCOSHはCOSH尺度にピークの重みづけ関数として, 標準パターンのスペクトル強度(f/u)をかけたものである。また, WLRは(5-60)式からわかるように, 最尤スペクトル距離に重み(f/u)と(g/v)をかけることによって導びかれている。よって, スペクトルが平坦化されている場合には, ピークをスペクトル強度(f/u , g/v)が大きい部分として, 谷を小さい部分として表わしうる。逆にスペクトル f , g に負の傾きがある場合には, f/u , g/v は低域ほど大きな値をとる傾向になるため, 低域を重視したマッチングとなる。

従来から, 母音の識別には, 第1ホルマントと第2ホルマントが有効であるとされている。⁽⁴⁾ 第1ホルマントと第2ホルマントは, 比較的low域に存在している。この意味からもlow域を重視したマッチングは, 母音識別において, 有効であると考えられる。平坦化されたスペクトルに一定の負の傾きをつけるのに1次系,

$$1 - \beta z^{-1} \quad (5-85)$$

を用いる。ただし, β は負の値である。1次系での対数周波数軸に対する重みを図5-16に示す。

母音の識別実験 5.5.2節と同じく男性4名が発声した会話音声データ(列車座席予約の496文節)を用いて, 上記の4種類の尺度によるスペクトル・マッチングによって行なう。母音標準パターンとして, 連続音声での平均を発声者ごとに用いる。誤り率を求める手法は, 5.5.2節での実験とまったく同じである。5.5.2節と同様の手順で平坦化したスペクトルに, (5-85)式の1次系で負の傾きをつけて母音識別実験を行った。

1次系($1 - \beta z^{-1}$)の β の値に対する各尺度の誤り率を図5-17に示す。COSH,

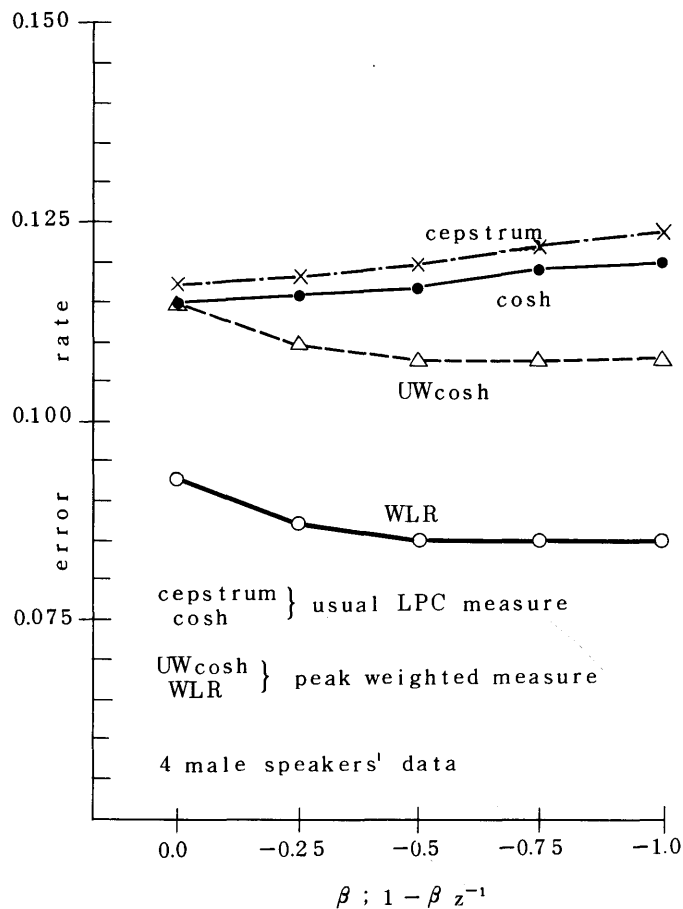


図5-17 母音識別実験における周波数重みづけの効果

CEPはスペクトルの傾きに対して不変であるべき尺度であるが、傾きによるLPC分析の精度の劣化により、誤り率は微増する。WLRとUWCOSHでは、低域の重みを増すことにより誤り率は減少し、周波数軸での重みづけの効果が示されている。次に、WLRの β に対する母音別の誤り率を図5-18に示す。すべての母音に対して、周波数重みづけの効果がみられる。さらに、 β に対する発声者ごとの誤り率を図5-19に示す。いずれの発声者に対しても周波数重みづけの効果がみられる。最後に、発声者別の各尺度の誤り率を図5-20に示す。COSHとCEPは $\beta = 0$ のときの誤り率、WLRとUWCOSHは $\beta = -0.5$ のときの誤り率である。いずれの発声者に対しても、WLRは最小の誤り率を示しており、かつ、発声者間のばらつきも小さい。この点でもWLRは有効な尺度である。

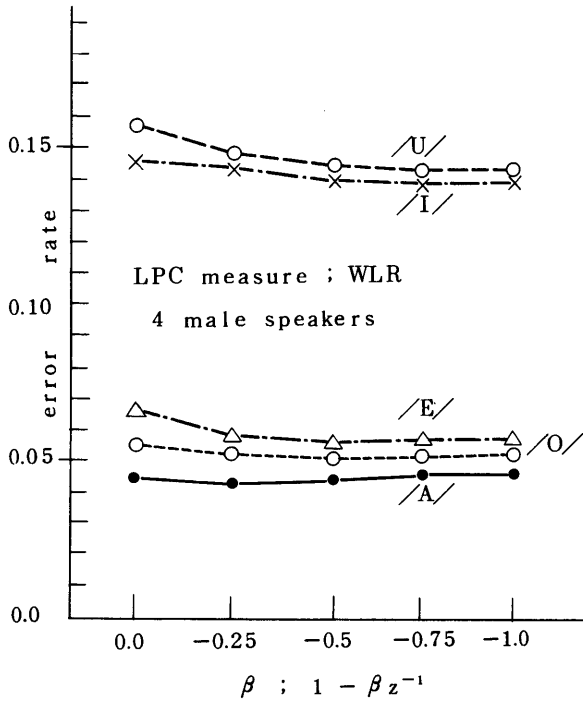


図 5-18 母音識別実験における周波数重みづけの母音別の効果

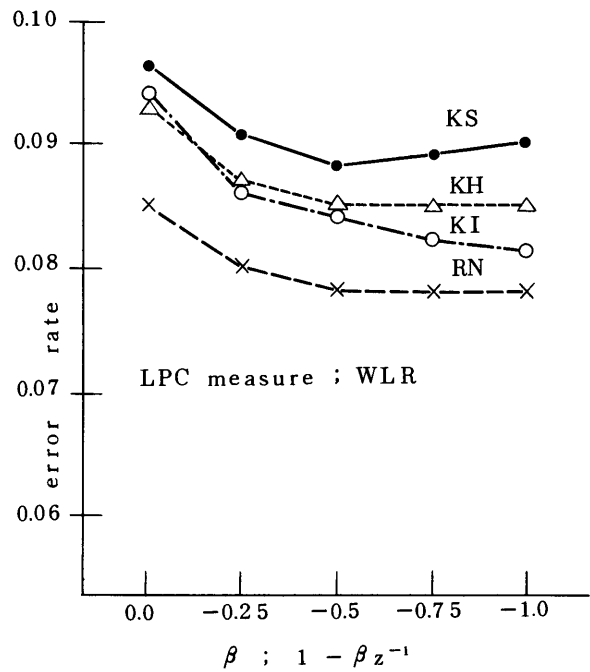


図 5-19 母音識別実験における周波数重みづけの発声者別の効果

5.6.2 スペクトル平滑化による効果⁽⁹⁵⁾

Peak weighted measure は、ピーク付近での重みづけが強すぎる恐れがある。スペクトルを平滑化することによって、ピーク重みづけを弱くすることも可能である。しかしながら、スペクトルの平滑化によって、スペクトル情報が失われ、母音識別率が低下する恐れもある。よって、ピークの重みづけの最適化と、スペクトル情報の損失の2つの要因により、最適な、スペクトルの平滑化の値が存在する可能性がある。

スペクトルの平滑化の手法として

は、LPC ケプストラム係数やLPC 相関係数を有限次数で打ち切ることによる手法⁽⁴⁴⁾

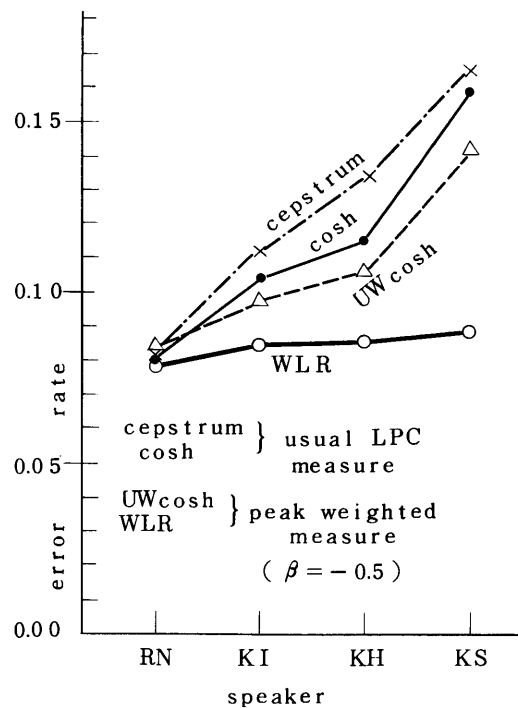


図 5-20 周波数軸の重みづけをした母音識別実験での各尺度の比較

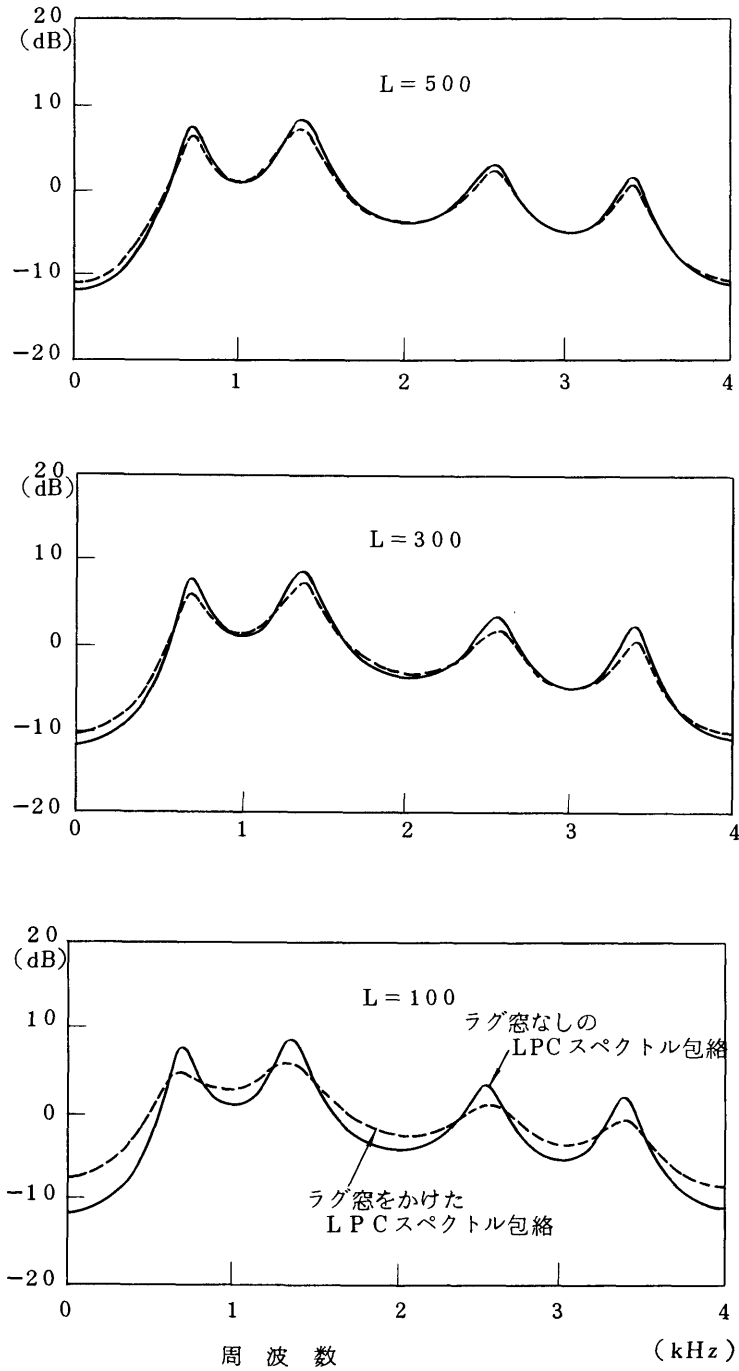


図 5 - 21 ラグ窓によるスペクトルの平滑化

と、ラグ窓⁽⁸⁹⁾などによる手法とが考えられる。本節では、ラグ窓によるスペクトルの平滑化の効果について検討する。

この節では、5.6.1節でとりあげた COSH, CEP, WLR, UWCOSHの他に、下に示す4種の Peak weighted measure もとりあげる。

$$\text{SNCEP} = d_0^2 + 2 \sum_{n=1}^{16} d_n^2 + \sum_{n=1}^{24} (2 d_0 d_n + q_n)(r_n + s_n) \quad (5-86)$$

ただし,

$$d_0 = \sum_{n=1}^{24} d_n (r_n + s_n)$$

$$d_k = C_k^{(f)} - C_k^{(g)}$$

$$q_n = 2 \cdot \sum_{k=1}^{16} d_k d_{k+n} + \sum_{k=1}^{n-1} d_k d_{n-k}$$

である。

$$\text{SNCOSH} = \left(\frac{1}{R_g} \sum_{n=-10}^{10} B_n \tilde{r}_n + 1 \right)^{1/2} \cdot \left(\frac{1}{R_f} \sum_{n=-10}^{10} A_n \tilde{s}_n + 1 \right)^{1/2} - 2 \quad (5-87)$$

$$\text{SMCOSH} = \frac{\sqrt{\sum_{n=-32}^{32} r_n^2 \sum_{n=-32}^{32} s_n^2}}{\sum_{n=-32}^{32} r_n s_n} - 1 \quad (5-88)$$

$$\text{SWEXP} = \frac{\sqrt{\sum_{n=-10}^{10} B_n \tilde{r}_n \sum_{n=-10}^{10} A_n \tilde{s}_n}}{R_f \cdot R_g} - \sum_{n=1}^{24} (r_n - s_n)(C_n^{(f)} - C_n^{(g)}) - 1 \quad (5-89)$$

母音識別実験には、前節と同じく男性4名が発声した座席予約文を用いる。1次系 $(1 - \beta z^{-1})$ による周波数重みづけの値として、前節の結果に基づき、 $\beta = -0.5$ を用いる。スペクトルの平滑化を行うために、相関係数に2項目係数窓によるラグ窓をかける。Lを2項目係数窓のラグ係数とすると、i次の相関係数にかける係数 ω_i は次式で定義される。

$$\omega_i = \binom{2L}{L-i} / \binom{2L}{L} \quad (5-90)$$

このラグ係数(L)を、1600から50まで変化させる。標準パターン/a/のラグ窓による平滑化の様子を図5-21に示す。このときの各尺度の誤り率を図5-22に示す。従来からのLPCスペクトル・マッチング尺度のCEP, COSH, および非対称なPeak weighted measureであるUWCOSHでは平滑化とともに誤り率は増加している。その他のPeak weighted measureでは、ラグ係数400あたりまで、誤り率は、

ほとんど増加しない。WLRでは、従来からの尺度 CEP や COSH よりも、3% 余り誤り率が減少している。WLR の母音別の誤り率とラグ係数の関係を図 5-23 に示す。後舌母音 /a/ および /o/ での誤り率の改善が大きいことがわかる。

この実験では、母音標準パターンとして、連続音声での平均を用いているが、次の章でのべる会話音声認識系 Voice Q-A System II では、単独に発声した母音から標準パターンを作成している。⁽⁹⁶⁾ そのため、連続音声での平均化母音の標準パターンよりスペクトルのずれが大きくなる。よってこのような場合には、ずれに対して強いスペクトル・マッチング尺度が望まれる。また、ずれに対処する一方法としてスペクトルの平滑化が考えられる。この意味でも、ラグ係数 300 あたりまで、誤り率の増加しない Peak weighted measure は、母音認識に適した尺度であると推察される。

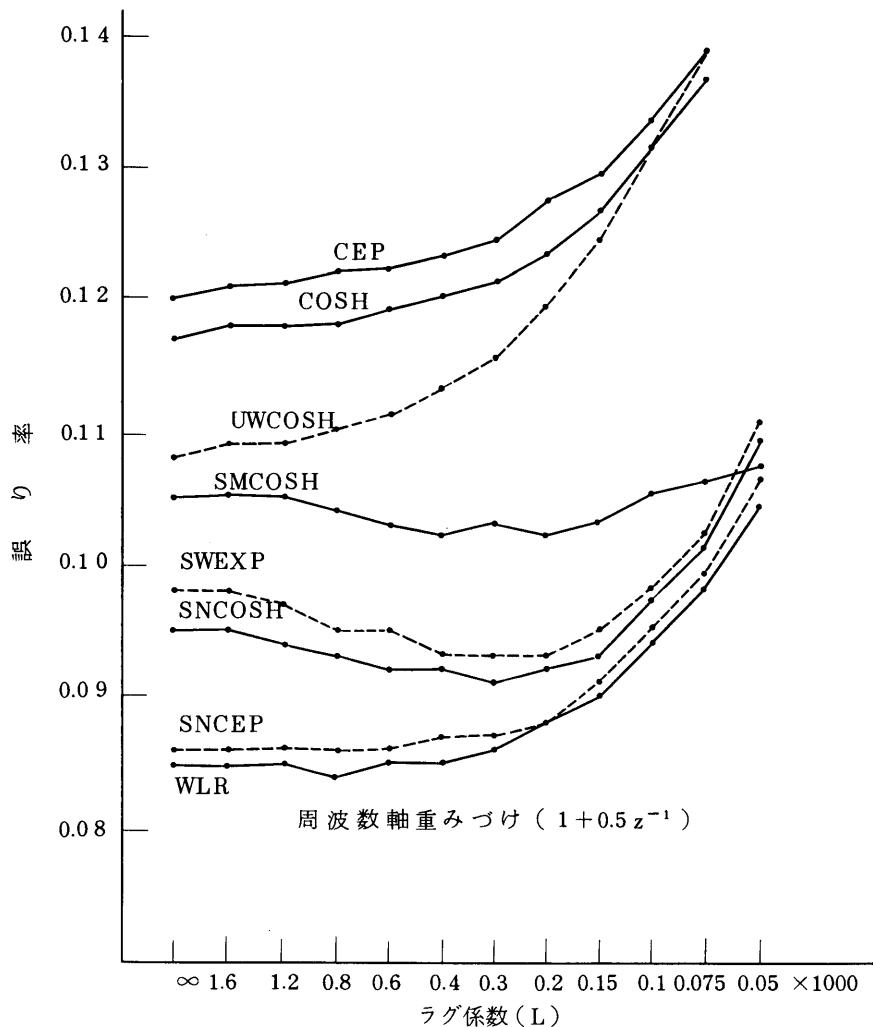


図 5-22 母音識別実験におけるラグ窓による平滑化の効果

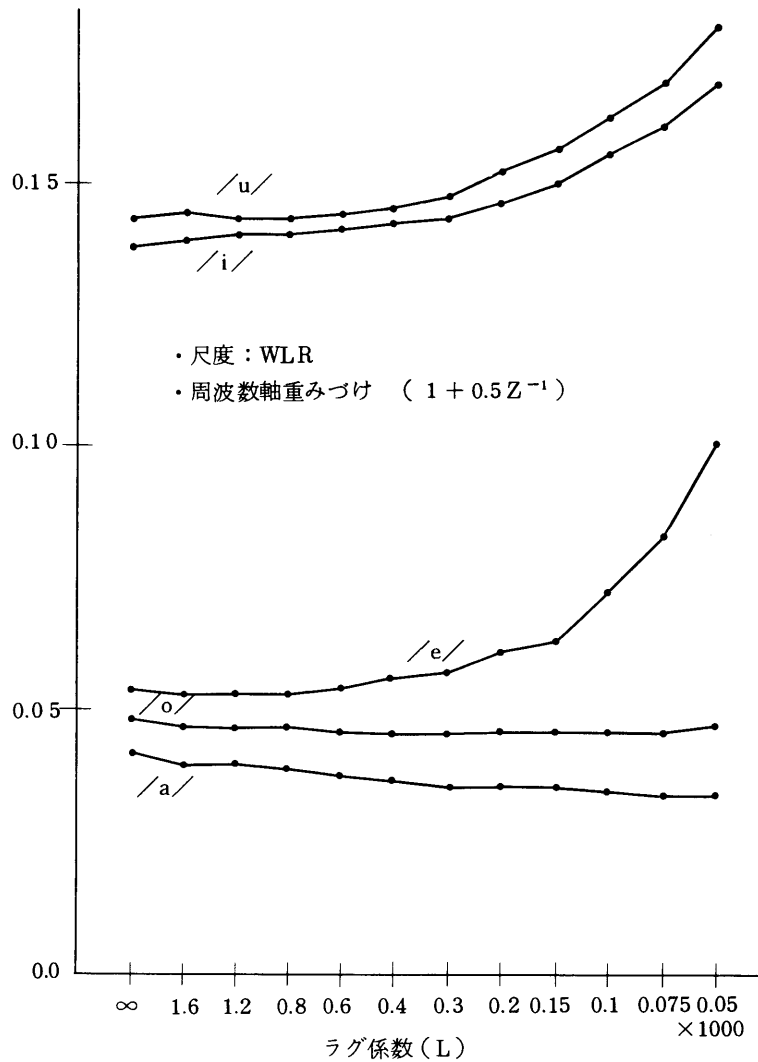


図 5-23 WLRによる母音識別実験での母音別の平滑化の効果

5.6.3 まとめ

LPC分析の周波数精度をもち、かつ、ピークの重みづけと周波数軸での重みづけが可能な Peak weighted measure を提案した。さらに、Peak weighted measure のピーク重みづけの様子を解析し、また、母音識別実験でその有効性を確かめた。とくに、WLR(Weighted Likelihood Ratio)は計算量も少なく、すぐれたスペクトル・マッチング尺度であることがわかった。さらに、周波数軸の重みづけの有効性とスペクトルの平滑化の効果を高め、WLRによる会話音声認識系 Voice Q - A System II - 2の母音認識の改良のためのパラメータの値を設定した。

5.7 LPC スペクトル・マッチング尺度による大語い単語音声認識⁽⁹⁷⁾

LPC分析にもとづくスペクトル・マッチング尺度の評価を641都市名の単語音声認識によって行う。LPCスペクトル・マッチング尺度として、従来からのLPCケプストラム距離(CEP), COSH尺度(COSH), 最尤スペクトル距離(LR)をとりあげる。また, 本論文で提案したPeak weighted measure から, 比較的計算量が少ない尺度として, WLR, SMCOSH, CORをとりあげる。男性4名の音声データによる認識実験で, WLRは96.7%の単語認識率を示し, 従来からの尺度CEPの95.3%よりも1.4%高い性能を示した。

5.7.1 LPC スペクトル・マッチング尺度

スペクトル・マッチング尺度として, 以下に示す尺度をとりあげる。

$$CEP = \sum_{i=1}^{16} (C_i^{(f)} - C_i^{(g)})^2 \quad (5-91)$$

$$COSH = \frac{1}{R_g} \sum_{k=-10}^{10} A_k s_k + \frac{1}{R_f} \sum_{k=-10}^{10} B_k r_k - 2 \quad (5-92)$$

$$LR = \log \sum_{k=-10}^{10} A_k s_k - \log R_g \quad (5-93)$$

$$WLR = \sum_{i=1}^{16} (r_i - s_i)(C_i^{(f)} - C_i^{(g)}) \quad (5-94)$$

$$SMCOSH = \frac{\sqrt{\sum_{i=-10}^{10} r_i^2 \cdot \sum_{i=-10}^{10} s_i^2}}{\sum_{i=-10}^{10} r_i s_i} - 1 \quad (5-95)$$

$$COR7 = \sum_{i=1}^7 (r_i - s_i)^2 \quad (5-96)$$

$$COR10 = \sum_{i=1}^{10} (r_i - s_i)^2 \quad (5-97)$$

各尺度の近似の次数として, 文献(94)での母音識別実験での打ち切りによる平滑化の手法の検討で, 誤り率が最小となる次数を用いた。CEP, COSH, LRおよびWLRはLPC分析を必要とする尺度である。SMCOSH, COR7, COR10は相関分析だけでLPC分析を必要としない尺度である。SMCOSHによる母音識別実験では, LPC

相関係数の最適次数は、16次とされている。しかし、LPC分析の次数10次以上のLPC相関係数を求めるには、LPC分析が必要となる。よって相関分析からだけで計算できるというSMCOSHの特質を生かすために、敢えて近似の次数は10次とした。相関距離に関して、文献(94)の母音識別実験での検討では、6次ないし7次が最適であるという結論に達したが、単語音声認識では、子音区間でのスペクトル・マッチングも行われる。よって、7次までの相関距離COR7と10次までの相関距離COR10とをとりあげた。

5.7.2 時間軸正規化マッチング

単語音声認識での時間軸の正規化マッチングの手法として、動的計画法(DP)を用いる。⁽²⁶⁾⁽²⁷⁾ 本節で用いるDP(Dynamic Programming)の説明を図5-24に示す。このDPは文献(99)の窓制限(adjustment window)および傾斜制限(slope constraint)に、端点フリーの手法⁽³²⁾をつけ加えた手法である。従来、端点フリーの手法を用いたDPでは、始端(start region)および終端(end region)での端点フリー領域でのコスト関数の取り扱いに問題があり、DPのパスを非対称にし、かつ、入力音声のみ端点フリーにすることにより、この問題をさけていた。⁽³²⁾⁽¹⁰⁰⁾ ここでのDPでは、上記のコスト関数の取扱いを厳密に行えるように、始端領域で、上方向ないし右方向のみの特殊なパスを許すことによってこの問題を解決した。以下、ここでのDPについて説明を加える。入力音声をB、単語標準パターンをAとする。AのiフレームとBのjフレームとのスペクトル距離を $d(i, j)$ で表わす。また $g(i, j)$ で (i, j) フレームまでのコスト関数を表わす。

(初期設定)

$$\left. \begin{aligned}
 g(1, 1) &= 2 \cdot d(1, 1) \\
 g(1, j) &= g(1, j-1) + d(1, j); 2 \leq j \leq r': \text{始端領域} \\
 g(i, 1) &= g(i-1, 1) + d(i, 1); 2 \leq i \leq r': \text{始端領域} \\
 g(1, j) &= \infty \quad r' < j \leq r \\
 g(i, 1) &= \infty \quad r' < i \leq r
 \end{aligned} \right\} (5-98)$$

ここで、 r' ; 始端領域の幅

r ; 窓制限の幅($2r+1$)

(繰返し計算)

$$g(i, j) = \min \begin{cases} g(i-1, j-2) + 2 \cdot d(i, j-1) + d(i, j) \\ g(i-1, j-1) + 2 \cdot d(i, j) \\ g(i-2, j-1) + 2 \cdot d(i-1, j) + d(i, j) \end{cases} \quad (5-99)$$

$$2 \leq i \leq iend, \quad 2 \leq j \leq jend$$

ここで、 $iend$ は標準パターンAのフレーム長、 $jend$ は入力音声Bのフレーム長である。

(マッチングの度合 $D(A:B)$ の決定)

$$D(A:B) = \min_{\{(i,j)\}} \{g(i, j) / (i+j)\}; (i, j) \in \text{終端領域} \quad (5-100)$$

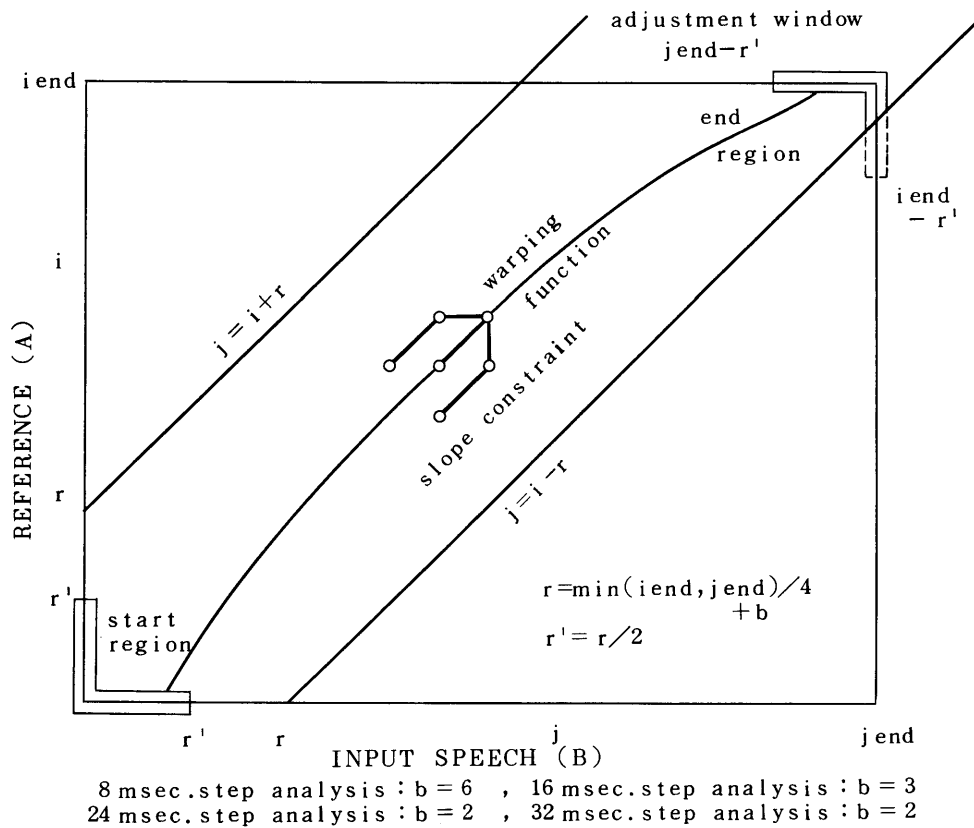


図 5-24 動的計画法による時間軸正規化マッチング

5.7.3 音声資料

100 単語余りを対象とした単語音声認識実験では、認識率が 98 ~ 99 % 余りとなり、認識率による尺度の評価は不可能となっていた。⁽⁹⁸⁾ 認識率で、尺度の評価を行うには、大語い単語を対象とする必要がある。ここでは、日本の全都市名(付録 5)を対

象にとりあげ各尺度の評価を行う。発声者は男性1名(KS)で、全都市名を2回発声した。発声場所は騒音レベル約70 dB(A)の計算機室でダイナミックマイクを用いた。2回の発声時期は約2週間離れている。最初に発声した音声データをAセット、2週間後発声した音声データをBセットとする。

音声を4 kHzの低域通過フィルターに通し8 kHzで標本化し、8 m sec,ステップで32m secのHamming窓をかけ、自己相関分析を行う。1次系($1-z^{-1}$)で高域強調ののち、10次のLPC分析を行う。単語音声区間の切り出しは、パワー情報を用いて自動的に行なう。単語間のマッチングの際に、端点フリー的な手法を用いるため語頭および語尾に若干の雑音区間がつくように単語音声区間を切り出している。

5.7.4 認識実験による尺度の評価

まず、Aセットの単語音声データを標準パターンとして、Bセットの単語音声データの認識実験を行い、次に、Bセットを標準パターンとしてAセットの認識実験を行う。まず、3フレームおきのデータ(24m sec・step)を用いて、尺度CEP, COSH, LR, WLR, SMCOSH, COR7, COR10による認識実験を行った。その結果を表5-12に示す。WLRは、従来からのLPCスペクトル・マッチング尺度COSH, CEP, LRよりも高い単語認識率を示している。この結果より、単語音声認識においてもPeak weighted measureが有効である可能性が示された。LPC分析を必要としない尺度, SMCOSH, COR10, COR7も比較的高い単語認識率を示している。とくに、SMCOSHは、従来からのLPCスペクトル・マッチング尺度に近い単語認識率を示している。SMCOSHは10次までの相関係数から直接に計算でき、簡易形の単語音声認識装置に適した尺度である。7次までの相関係数を用いたCOR7は、COR10とほぼ同じ単語認識率を示している。計算量が非常に少ないという点からすれば、COR7も有望である。

また、単語音声認識系の語い数によらない評価尺度として、ヒストグラムからの誤り率(error rate by threshold)がある。つまり、図5-25に示すようにマッチングスコア $D(A:B)$ の $A=B$ の場合のヒストグラムと $A \neq B$ の場合のヒストグラムとの度数を正規化し、誤り率最小になるようにいき値を定め、そのときの誤り率を用いる。この方法は、この章の5.3.2節, 5.5.2節の母音識別実験の場合と同様の評価手法である。この誤り率は、図5-25からわかるように、単語認識系の拒絶能力と

も密接な関係がある。この誤り率による評価結果を表 5-12 の最下段の誤り率 (error rate by threshold) の欄に示す。誤り率の評価でも WLR はすぐれていることがわかる。

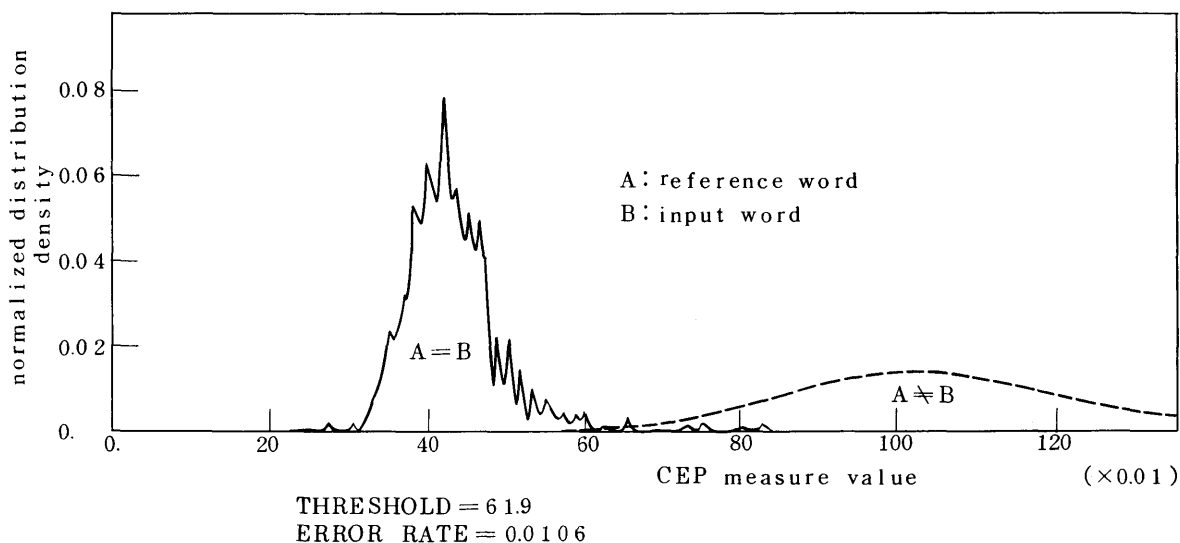


図 5-25 尺度 CEP によるヒストグラムといき値による誤り率

次に、分析のステップを 8 m sec にした場合の WLR, CEP, SMCOSH による認識実験の結果を表 5-13 に示す。WLR は 96.4% と従来からの LPC スペクトル・マッチング尺度の CEP よりも 1% 高い単語認識率を示しており、Peak weighted measure の有効性を示している。WLR では、正しい単語が第 2 位以内に 98.2% まで、第 3 位以内に 98.4% まで含まれる。付録 6 に、WLR, CEP, SMCOSH の誤りを参考のために示しておく。

分析ステップ 8 msec の場合の尺度 WLR, CEP, SMCOSH での誤りの種類を表 5-14 に示す。表 5-14 の誤りの例の欄に、WLR による全誤りを示しておく。誤りの大半は、音声区間の切り出し誤りに起因していることがわかる。音声区間の切り出し誤りの基準を次のように設定した。

- (1) 1 モーラ分の音声の脱落がみとめられる。
- (2) 語頭あるいは語尾に 160 msec 以上の雑音区間が付加されている。

表 5-12 大語い単語音声認識における LPC スペクトル・マッチング尺度の比較
(24 msec ステップ分析)

	WLR	CEP	COSH	L R	SMCOSH	COR10	COR7
単語認識率	94.9 $(\frac{1216}{1282})$	94.4 $(\frac{1210}{1282})$	94.6 $(\frac{1213}{1282})$	94.2 $(\frac{1207}{1282})$	93.1 $(\frac{1193}{1282})$	93.0 $(\frac{1192}{1282})$	92.4 $(\frac{1185}{1282})$
2位以内	97.8	97.7	97.5	97.3	97.0	97.1	96.3
3位以内	98.1	98.1	98.1	98.1	97.5	97.8	97.7
5位以内	98.3	98.4	98.4	98.4	98.1	98.0	98.1
10位以内	98.4	98.7	98.5	98.5	98.2	98.4	98.2
error rate by threshold	1.15	1.30	1.30	1.35	1.56	1.41	1.42

(%)

表 5-13 大語い単語音声認識における LPC
スペクトル・マッチング尺度の比較
(8 msec ステップ分析)

	WLR	CEP	SMCOSH
単語認識率	96.4 $(\frac{1236}{1282})$	95.4 $(\frac{1223}{1282})$	94.3 $(\frac{1209}{1282})$
2位以内	98.2	98.2	97.4
3位以内	98.4	98.6	98.0
5位以内	98.8	98.8	98.2
10位以内	99.2	99.1	98.8
error rate by threshold	0.99	1.06	1.42

(%)

切り出し誤り以外の主な誤りの種類は、文頭の子音の認識誤りである。WLRは、母音の識別に非常に有効であった。その性質は、この単語認識実験でも見受けられ、WLRは CEPでの第1モーラにおける母音の誤りをすべて救済している。表 5-14の第1モーラにおける母音の誤りの例には、CEPによる誤りを示しておいた。SMCOSHの誤りと

して、モーラ数が異なる単語に誤るなどの、その他の項に分類される誤りが多いことがわかる。これは、SMCOSHが、ピークの重みづけが強すぎる尺度であることと、相関係数を10次で打ち切っていることによると考えられる。

5.7.5 分析フレーム周期の影響

各尺度による単語認識における分析周期の認識率に対する影響をみるために、分析フレーム周期を 8m sec ~ 32m sec にして単語音声認識を試みた。その結果を図 5-26

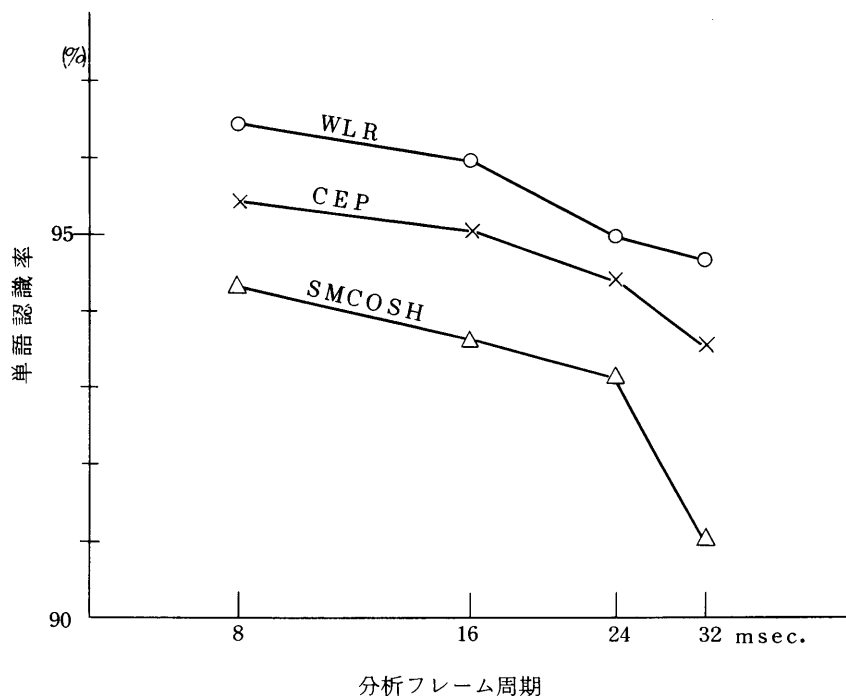


図 5-26 大語い単語音声認識実験における分析フレーム周期の影響

に示す。分析フレーム周期を長くするにつれて認識率が低下する。単語認識の計算量は分析フレームが n 倍になれば、ほぼ $1/n^2$ となる。このことを考えると、単語認識における分析フレーム周期としては、16m sec 付近がよいと推測される WLR は、分析フレーム周期を長くしても比較的高い認識率を保っている。

5.7.6 大語い単語音声認識実験

新たな 3 名の男性の発声者 (KA, NS, KI) の発声した音声データを用いて、641 都市名を対象とした単語音声認識実験を行った。発声場所、発声時期、音声区間の切り出し方法は、5.7.3 節の発声者 KS の場合と同様である。分析のステップは、5.7.5 節の検討をもとに、16m sec ステップにした。スペクトル・マッチング尺度として、WLR, CEP, SMCOSH をとりあげた。発声者 (KS, KA, NS, KI) ごとの単語認識率を表 5-15 に示しておく。いずれの発声者に対しても、Peak weighted measure の WLR による単語認識率が、従来からの尺度 CEP よりも高いことがわかる。また、LPC 分析を必要とせず、相関分析から直接に計算できる尺度 SMCOSH も比較的高い単語認識率を示しており、尺度 CEP より 0.5% 低いだけである。

表5-14 大語い単語音声認識における誤りの種類と誤りの例
(8m sec ステップ分析, 発声者: KS)

	WLR	CEP	SMCOSH	誤りの例 (尺度: WLR)
文頭の子音のみの誤り	11	13	14	oga → koga, koga → oga, kimizu → shimizu, akikawa → takikawa, saNjo → aNjo, oga → koga, hasuda → masuda, ko:fu → ho:fu, chino → hino, fuji → uji, tagawa → sūkagawa
2個以下の子音の誤り	6	8	13	ko:fu → o:bu, nakano → nagano, so:ja → o:da, ho:fu → o:bu, ishikawa → ichikawa, ichikawa → ishikawa
第1モーラの母音の誤り	0	6	3	(CEP) kuki → toki, kumamoto → kitamoto, kuki → toki, uozu → o:zu, nishinomiya → fujinomiya, tokushima → fukushima
音声区間の切り出し誤り	24	27	28	—
その他	5	5	15	suzu → tsu, mino → minoo, uwajima → wajima, i:zuka → iida, tsu → suzu
	46	59	73	

4名の発声者の平均の単語認識率で、WLRは96.7%となり、従来からの尺度CEPよりも1.4%高い性能を示している。単語認識誤りの約半数は、音声区間の切り出しに起因する誤りである(表5-16参照)。この切り出しに起因する誤りを、スペクトル・マッチング尺度のみで救済することは、ほとんど不可能である。このことを考慮すると、この1.4%の性能の向上は大きいと考えられる。つまり、音声区間の切り出しによる誤りを除いた単語認識率のCEPを基準にした改善率でみると、発声者KSで32%($= (95.9 - 95.0) / (97.8 - 95.0)$), CEP: 95.0% → WLR: 95.9%, 切り出し誤り2.2%), 発声者KAで41%($= (96.3 - 94.4) / (99.0 - 94.4)$), CEP: 94.4% → WLR: 96.3%, 切り出し誤り1.0%), 発声者NSで39%($= (97.7 - 97.0) / (98.8 - 97.0)$), CEP: 97.0% → WLR: 97.8%, 切り出し誤り1.2%), 発声者KIで49%($= (96.9 - 94.6) / (99.3 - 94.6)$), CEP: 94.6% → WLR: 96.9%, 切り出し誤り0.7%), となる。よって、CEPを基準とするWLRの平均の改善率は約40%となり、かなり大きな改善効果が得られていることがわかる。

次に、尺度WLRの場合の4名の発声者ごとの誤りの種類を表5-16に示しておく。分類項目は表5-14と同じである。誤りの半数弱(38%)は、音声区間の切り出し誤りに起因している。それ以外の誤りでは、文頭の子音のみの誤りが平均で27%と多

表5-15 大語い単語音声認識実験の結果
(16msec step分析, 発声者: KS, KA, NS, KI)

発声者	K S			K A			N S			K I			平均		
尺度	CEP	WLR	SM COSH	CEP	WLR	SM COSH	CEP	WLR	SM COSH	CEP	WLR	SM COSH	CEP	WLR	SM COSH
単語認識率	95.0	95.9	93.6	94.4	96.3	93.8	97.0	97.7	96.3	94.6	96.9	95.5	95.3	96.7	94.8
	$\frac{1218}{1282}$	$\frac{1230}{1282}$	$\frac{1200}{1282}$	$\frac{1210}{1282}$	$\frac{1234}{1282}$	$\frac{1202}{1282}$	$\frac{1244}{1282}$	$\frac{1252}{1282}$	$\frac{1235}{1282}$	$\frac{1213}{1282}$	$\frac{1242}{1282}$	$\frac{1224}{1282}$	$\frac{4885}{5128}$	$\frac{4958}{5128}$	$\frac{4861}{5128}$
2位以内	97.7	97.9	97.3	98.1	98.9	98.4	98.7	98.8	97.9	98.0	99.1	97.8	98.1	98.7	97.9
3位以内	98.4	98.3	97.9	98.8	99.1	98.9	99.1	99.1	98.4	98.8	99.3	98.4	98.8	99.0	98.4
5位以内	98.6	98.6	98.1	99.4	99.5	99.2	99.2	99.1	98.8	99.3	99.5	98.9	99.1	99.2	98.8
10位以内	98.7	98.7	98.5	99.5	99.5	99.4	99.2	99.2	98.9	99.5	99.5	99.3	99.2	99.2	99.0
error rate by threshold	1.13	1.11	1.35	0.69	0.66	0.86	0.57	0.66	1.11	0.63	0.53	1.12	0.76	0.74	1.11

表 5 - 16 大語い単語認識における誤りの種類
(16msec ステップ分析, 尺度WLR)

発 声 者	KS	KA	NS	KI	合 計 (%)
文頭の子音のみの誤り	11	21	6	8	46 (27%)
2個以下の子音の誤り	5	7	4	13	29 (17%)
第1モーラの母音の誤り	1	2	3	2	8 (5%)
音声区間の切り出し誤り	28	13	15	9	65 (38%)
そ の 他	7	5	2	8	22 (13%)
合 計	52	48	30	40	170

い。この文頭での子音の誤りが多い原因として、文頭での子音のスペクトルの特徴が明確でないことの他に、次の3つの事柄が考えられる。

- (1) 70dB(A)の騒音下では、音声区間の切り出しが不確実となり、数フレーム(40～70msec)余りの切り出しの誤りはされることができない。この切り出しの不確実さによって、子音の特徴が失なわれている可能性がある。
- (2) 端点フリー的なスペクトル・マッチングの手法により、文頭のスペクトルの系列としての子音情報を殺している可能性がある。
- (3) LPC分析で用いている32msecのHamming窓の窓長が長すぎる可能性がある。この文頭の誤りに対しては、文頭の子音部分のみに対するtop-down的かつ局所的な処理によって、ある程度まで救済できるものと考えられる。

その他、モーラ数が異なる単語への誤りは非常に少ない。このことは、時間軸の非線形伸縮に用いたDPの手法における窓制限、傾斜制限、端点フリー的処理の有効性を示しているものと考えられる。

発声者4名による認識誤りの全誤りを、誤りの種類別に表5-17に示しておく。ただし、音声区間の切り出し誤りに起因するものは除いてある。モーラ数が異なる単語への誤りは、誤り全体の18%である。子音1個のみの誤りは、誤り全体の54%をしめている。

表5-17 大語い単語音声認識における誤り(16mec ステップ分析, 発声者: KS, KA, NS, KI)

誤りの種類	誤りの個数	誤り
文頭の子音のみの誤り	46	oga→koga(3), koga→oga(3), kimi zu→shimi zu(3), akikawa→takikawa(2), saNjo:→aNjo:(2), tsushima→kushima, itami→kitami, hasuda→masuda, ko:fu→ho:fu(2), saga→kaga, kuji→fuji(3), kaga→saga(2), chiryu:→kiryu:, uji→fuji, ako:→wako:, saNda→haNda, ho:fu→ko:fu(2), taku→saku(3), takikawa→akikawa, kuji→uji, mito→ito:, saku→taku(2), fuji→uji, tagawa→sukagawago:tsu→o:tsu, kashima→hashima, kiryu:→chiryu:(2), oyama→toyama, o:no→to:no
2個以下の子音の誤り	29	ko:fu→o:bu(2), nakano→nagano, so:ja→o:da, ho:fu→o:bu, o:mura→o:mura(2), date→naze, nikko:→mito, ichikawa→ishikawa, o:me→ko:be, yamada→yamaga, kanoya→nagoya, o:ta→o:da, hamada→tamana, yamaga→yamada, oga→toda, kawagoe→kawanoe, hadano→katan(2), ko:be→o:me, nagato→nakano, kasuga→hasuda, hirara→hirata, ho:jo:→hoNdo, oga→toba, sagae→sabae, nagato→nagano, saito→saijo:
第1モーラの母音の誤り	8	kuki→toki, fukui→haku(2), fukushima→tokushima, fujinomiya→nishinomiya, fujinomiya→ichinomiya, izumi→tsukumi(2)
その他	22	suzu→tsu, mino→minoo(2), uwajima→wajima, i:zuka→iida, wajima→uwajima(2), tsu→suzu, kuki→kikuchi, zushi→kikuchi, shiki→kitsuki, ni:za→iida, kuki→usuki, fukushima→tsushima, hasuda→katsuta(2), shimoda→o:ta, nagai→imabari, fujimi→fuji, nabari→imabari, o:da→o:mura, kushikino→chikushino

(尺度: WLR)

5.8 あとがき

会話音声認識システム Voice Q - A System II における母音認識の基礎検討として、LPC スペクトル・マッチング尺度について検討した。

まず、従来からの LPC スペクトル・マッチング尺度として、LPC ケプストラム距離、COSH 尺度、最尤スペクトル距離、正規化残差をとり上げ、母音標準パターンとのスペクトル・マッチングによる母音識別実験での評価を行った。この実験で、スペクトルのパワー比を様々な条件に設定した。その結果、尺度値が最小になるようにパワー比を設定した条件での誤り率が小さく、マッチングの条件としてすぐれていることが確かめられた。これらの従来からの LPC スペクトル・マッチング尺度のうち、残差パワーを正規化した条件、あるいは、尺度値最小の条件での COSH 尺度がすぐれていることが確かめられた。第 6 章では、まずこの COSH 尺度を用いて、会話音声認識系 Voice Q - A System II の音響処理を作成する。

5.4 節から 5.6 節において、新しい LPC スペクトル・マッチング尺度、LPC Peak Weighted Spectral Matching Measure を提案し、さらに、尺度の性質の理論的解析および母音識別実験による評価を行った。この新しい尺度は、LPC スペクトル包絡のピーク付近に重みをおいたマッチングが可能であり、ホルマント・マッチングと LPC スペクトル・マッチングの両者の長所をあわせもつ尺度であることが確かめられた。また、Peak Weighted Measure においては、周波数軸の重みづけも可能である。とくに、Peak Weighed Measure のうち WLR (Weighted Likelihood Ratio) は、すぐれた母音識別能力を持っていることがわかった。第 6 章の後半では、この WLR を利用して、会話音声認識系 Voice Q - A System II の音響処理の改良を行う。

5.7 節では、WLR を 641 都市名を対象とするスペクトル・マッチングによる大語い単語音声認識に適用し、男性 4 名の平均で、96.7% 単語認識率を得、従来の LPC 尺度よりも大幅にすぐれていることを確かめた。

第6章 Voice Q-A System IIの構成と評価

6.1 はしがき

会話音声認識の研究を、新幹線の座席予約を対象にとりあげて、1974年に開始し⁽¹⁰¹⁾、1976年末に実時間の5倍の処理時間で動作するオンライン・システム(Voice Q-A System I)⁽³⁰⁾⁽³¹⁾⁽¹⁰²⁾を作成した。このシステムについては、言語処理を中心に本論文の2章、3章で詳しく述べた。このVoice Q-A System Iの評価を、男性8名の音声データで行い、86.0%の文節認識率を得た。このSystem Iの性能を向上させるための要因を調べた結果、音響処理の精密化とシステム全体の構成法の問題点がでてきた。とくに、音響処理での音韻単位へのセグメンテーションの改善が急務であるとの結論を得た。これと同様の結論は、ARPAの最終報告⁽⁹⁾⁽¹⁰⁾にも見受けられる。この結論に基づき、1978年4月から音響処理の精密化の研究を再開した。この精密化の効果をVoice Q-A System IIを作成する中で評価することにした。Voice Q-A System IIの対象も、System Iと同様、新幹線の座席予約である。

Voice Q-A System IIでは、次の項目に重点をおいて検討を行った。

- (1) 音韻単位へのセグメンテーションの精密化。
- (2) 母音認識の精密化。
- (3) 母音標準パターンの登録(学習)の容易化。
- (4) 発声者によらない子音認識。

本章では、以上の方針に基づいて作成したVoice Q-A System IIについて詳しく述べる。この章の前半では、第5章の前半の検討に基づき、スペクトル・マッピングの尺度としてCOSH尺度を取りあげ、Voice Q-A System II-1の音響処理の作成を行う。後半では、第5章の後半の検討に基づき、スペクトル・マッチング尺度をCOSH尺度からPeak Weighted MeasureのWLR(Weighted Likelihood Ratio)に変更し、Voice Q-A System II-1の改善を行い、Voice Q-A System II-2を作成する。

6.2 言語処理の改良

Voice Q-A System II の音響処理の評価を，言語処理と結合することによって行う。言語処理は，第3章で述べた Voice Q-A System I の言語処理とほとんど同じである。ただし，次の2点の改良を行った。

- (1) 母音に関する音韻変形規則(表3-1)に VCV→V型を追加した。
- (2) 入力音声に対して，予約項目間にわたってのスコアの比較を許した。

改良点(1)；Voice Q-A System I の音響処理での誤りの中に，長母音を2個の母音に切断する誤りが多く，かつ，この訂正規則が既存の音韻変形規則では記述が迷倒であった。よって， $V * V \rightarrow V$ にする変形規則を付け加えた。System II での音韻変形規則を付録7に示しておく。

改良点(2)；Voice Q-A System I の言語処理では，入力の音韻ラティスに対して，予約項目別に構文解析を行い，予約項目間にわたるスコアの比較を行っていなかった。このような depth-first 的な手法を，予約項目間にわたってスコアの比較ができるように breadth-first 的な手法に変更した。

6.3 COSH尺度を用いた音響処理 (Voice Q - A System II - 1)

新幹線の座席予約文に含まれる音韻は， $/a, i, u, e, o/$ 5母音と撥音 $/N/$ ，そして20種類の子音 $/k, s, t, n, h, m, y, r, w, g, z, d, b, p, zj, hj, kj, tt, pp, kk/$ である。音響処理では，「日本語には子音の連続はない」という性質を利用して，VCV音節 (Vowel - Consonant - Vowel concatenation units) を単位とする処理方法をとっている。この処理方法は，日本語のCV音節のリズムを利用した処理である。このリズムを保つために，母音の連続の間には，仮想子音 $/*/$ を挿入する。さらに，撥音 $/N/$ を母音的に取り扱う。子音に関しては，通常の子音はすべて含まれている。拗音は，3種類 $/zj, kj, hj/$ が含まれている。促音は $/tt, pp, kk/$ が含まれている。その他，ポーズも一種の音韻として扱っており $/**/$ という記号で表わしている。

以下，音韻単位のセグメンテーションの手法に重点をおいて，音響処理について説明す

る。

6.3.1 前処理

入力音声を 4 kHz の低域通過フィルタに通し，8 kHz で標本化したのち，32msec のハミング窓をかけ，8msec ステップで自己相関分析を行う。さらに，相関領域で $(1-z^{-1})$ のフィルタで高域強調を行い，さらに， $(1-\alpha z^{-1})$ のフィルタを用いてスペクトルの平坦化を行う。ここで， α は，高域強調をかけた単独に区切って発声した母音 /a, i, u, e, o/ の第 1 次の相関係数の平均値である。これらの処理の後，ラグ窓をかけ，10 次の LPC 分析を行う。

6.3.2 標準パターン⁽⁹⁶⁾

音響処理システムは，LPC 分析を介したスペクトル包絡のマッチングを基本としている。標準パターンとして，母音 /a, i, u, e, o/ の標準パターンと撥音 /N/ の標準パターンと雑音の標準パターンを利用する。これらの標準パターンは，LPC 分析された形で蓄えられる。

母音の標準パターンは，区切って発声した母音 /a, i, u, e, o/ のスペクトル包絡をもとに，各々の母音の第 1 ～ 第 3 ホルマントの周波数とバンド幅を変更することによって，自動的に作成される。この推定法によって，単独発声した母音から連続音声の中の平均の母音標準パターンを作成する。しかし，拗音に続く母音や半母音 /y/ の前後の母音は，強い調音結合の影響を受けるため，平均の母音標準パターンで認識できないことが多い。このような条件では，中立母音の標準パターンも併用して母音認識を行う。中立母音として，/ie/, /iu/, /uo/, /ea/, /eu/, /eo/ の 6 種のわたり音を用意している。これらの中立母音の標準パターンを，母音の標準パターンから，ホルマントの移動合成によって作成する。これらの中立母音は，各々，/e/, /u/, /o/, /a/, /u/, /o/ の標準パターンとみなされる。また，撥音 /N/ の標準パターンは，単独に発声した /N/ をそのまま用いている。

雑音の標準パターンは，音声区間の検出，無声区間の検出などのために用いられる。雑音の標準パターンは，次に示す手順のように，そのパワー情報とスペクトル情報が，入力音声ごとに更新される。

(i) 入力フレームと雑音標準パターンとの差(d)を COSH 尺度で計算する。入力フレー

ムがなければ、(V)へ。

(ii) $d \geq 2.0$ ならば、次のフレームを計算 (i)へ)。

(iii) 雑音の標準パターンのパワー(u)に、入力フレームのパワー(v)を次式のように加え、uの値を更新する。

$$u \rightarrow 0.975 u + 0.025 v \quad (6-1)$$

(iv) $d \geq 0.75$ ならば、(i)へ。さもなければ、雑音の標準パターンの相関係数(r_i)に、入力フレームの相関係数(s_i)を次式のように加える。

$$r_i \rightarrow 0.975 r_i + 0.025 s_i \quad (i = 1, \dots, 10) \quad (6-2)$$

次のフレームへ (i)へ)。

(V) 相関係数 r_i を LPC 分析し、新しい雑音の標準パターンを作成する。

上記の手順によって雑音標準パターンを更新するので、周囲雑音の変動に、ある程度まで追従することができる。

6.3.3 基本特徴量の計算

入力音声のセグメンテーション、母音認識、子音認識に用いる基本特徴量として、次の①～⑧を、入力音声のパワー値(v)、相関係数(s_i)、残差パワー(R_g)、最尤スペクトルパラメータ(B_i)から計算する。

① 雑音スペクトルとの差 (NOISE)

入力音声の各フレームのスペクトルと、雑音スペクトル(u, r_i, R_f, A_i)との差を次式の COSH 尺度で計算する。

$$\text{NOISE} = \frac{v}{u \cdot R_f} \sum_{k=-10}^{10} A_k \cdot s_k + \frac{u}{v \cdot R_g} \sum_{k=-10}^{10} B_k \cdot r_k - 2 \quad (6-3)$$

② パワー・レベル (POWER)

入力音声パワーの最大値 (u_{\max}) と最小値 (u_{\min}) を基準にして、かつ対数をとることによってパワーレベルを正規化する。

$$\text{POWER} = \frac{\log_{10} v - \log_{10} u_{\min}}{\log_{10} u_{\max} - \log_{10} u_{\min}} \quad (6-4)$$

ただし、 u_{\max} 、 u_{\min} は、入力音声ごとに更新される。

$$\log_{10} u_{\max} \rightarrow 0.8 \cdot \log_{10} u_{\max} + 0.2 \cdot \log_{10} v_{\max} \quad (6-5)$$

$$\log_{10} u_{\min} \rightarrow 0.8 \cdot \log_{10} u_{\min} + 0.2 \cdot \log_{10} v_{\min} \quad (6-6)$$

v_{\max} , v_{\min} は、入力音声での最大パワーと最小パワーである。

③ パワー・ディップ (POW - DIP)

入力フレームの前後7フレームの正規化対数パワー (POWER) を2次曲線で最小2乗近似したときの2次微係数 (ラプラシアン) でパワー・ディップを計算する。

$$\text{POW-DIP} = \sum_{i=-7}^7 \omega_i p_i \quad (6-7)$$

p_i は、 i フレームの正規化された POWER の値、 ω_i はラプラシアンの重みであり、次の値をとる。

$$\begin{aligned} & \{ \omega_0, \omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7 \} \\ & = \{ -280, -265, -220, -145, -40, 95, 260, 455 \} \quad (6-8) \end{aligned}$$

$$\omega_i = \omega_{-i}$$

④ 短時間スペクトル変化 (SPEC.(S))

16msec つまり、1フレーム離れたフレーム間の LPC スペクトル包絡の差を COSH 尺度で計算する。

⑤ 長時間スペクトル変化 (SPEC.(L))

48msec つまり、5フレーム離れたフレーム間の LPC スペクトル包絡の差を COSH 尺度で計算する。

⑥ 母音の候補 (VOWEL)

入力フレームと、母音の標準パターン / a, i, u, e, o / および撥音 / N / とのスペクトルの差を、残差パワーを正規化した条件での COSH 尺度で計算する。

$$\text{VOWEL} = \frac{1}{R_g} \sum_{k=-10}^{10} A_k s_k + \frac{1}{R_f} \sum_{k=-10}^{10} B_k r_k - 2 \quad (6-9)$$

ただし、 $\{ A_k, r_k, R_f \}$ は母音の標準パターン、 $\{ B_k, s_k, R_g \}$ は入力フレームの LPC パラメータである。

⑦ 中立母音の候補 (NEUT. VOW.)

入力フレームと、6種の中立母音との差を残差パワーを正規化した条件での COSH 尺度で計算する。

⑧ 高域と低域のスペクトル・エネルギーの比 (BANDPASS)

低域 (0 ~ 500 Hz) の対数スペクトル・エネルギーの最大値と、高域 (3 kHz ~ 4 kHz) の対数エネルギーの平均値の差を、入力フレームの最尤スペクトルパラメ

ータ (B_i) から DFT (cos変換) で計算する。

以上の基本特徴量①～⑧を用いて、音声区間の検出、音韻のセグメンテーション、母音認識、子音認識を行う。

6.3.4 基本構成

計算された基本特徴量をもとに、まず、音声区間の切り出しを行う。次に、子音区間の候補と母音区間の候補を抽出する。その後、図6-1に示す過程で音韻のセグメンテーションおよび認識を行い、その結果を音韻ラティス上に出力する。

音声区間の切り出しを、主として雑音スペクトルとの差①を用いて行う。さらに精密な文頭の位置を短時間スペクトル変化④を用いて決定する。つまり、 $NOISE \leq 1.0$ のフレームを見出し、その付近で、 $SPEC. (S) \geq 0.5$ のフレームが存在すれば、そのフレームを文頭とする。

子音区間の候補のラベリングを、雑音スペクトルの差①、パワー・レベル②、パワー・ディップ③、短時間スペクトル変化④を用いて行う。このラベリングによって、ほとんどの子音区間が検出される。子音区間のラベリングについては、6.3.5の(I)で詳しく述べる。

ラベリングによって子音区間の候補、母音区間の候補が抽出されたのち、図6-1に示す系で各区間の処理を行う。母音区間の候補は、VOWEL RECOGNITION PROGRAMで処理され、子音区間の候補は、CONSONANT RECOGNITION PROGRAMで処理される。図6-1で小文字の leaf はサブルーチンであり、同じ名前の leaf では同一のサブルーチンを用いている。子音区間の候補のラベリングで、必ずしも、子音区間と母音区間とが完全に分離されるわけではない。よって、図6-1の tree の小文字の node に示される6種類のセグメンテーションを主体とする check program を動作させる。母音区間の候補に対しては、

- (a) 連続母音の検出 (check of successive vowels) のプログラムを動作させる。子音区間の候補に対しては、
- (b) 拗音、半母音 / y / の検出 (check of contracted consonants) ,
- (c) 撥音 / N / の検出 (check of nasal / N /) ,
- (d) 無声化母音の検出 (check of devocalized vowels) ,
- (e) 文頭の子音のセグメンテーション (check of head consonants) ,

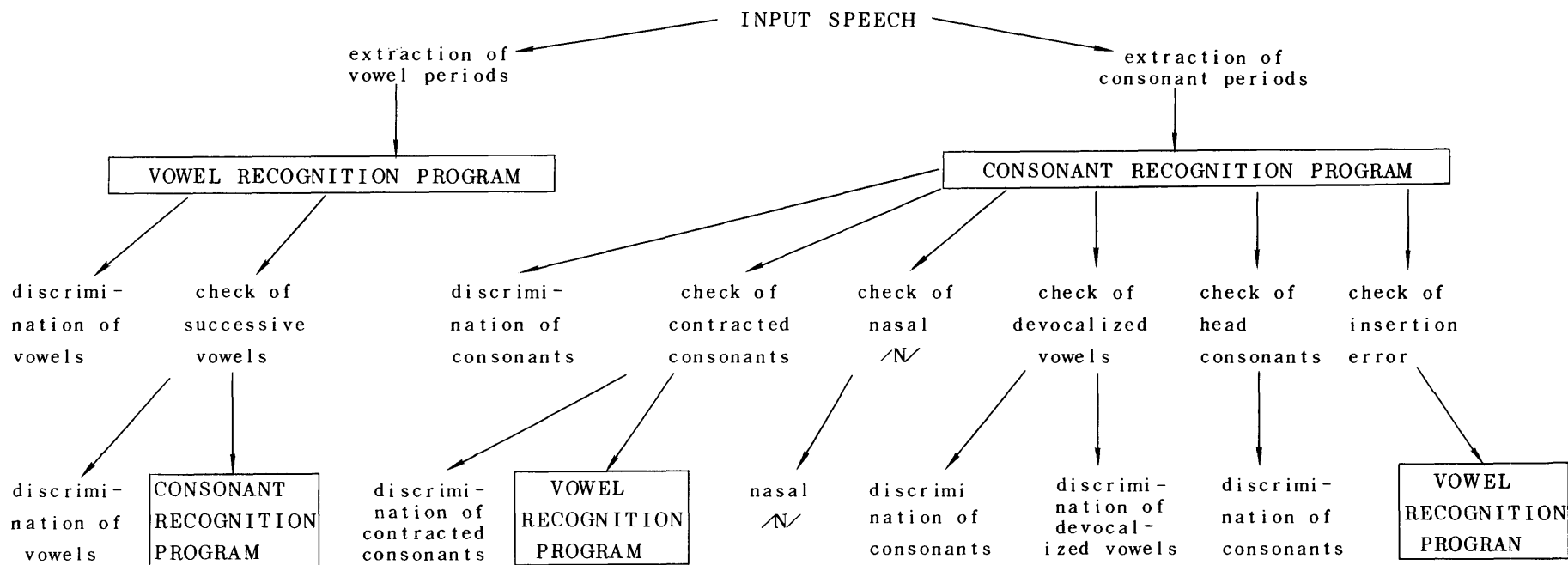


図6-1 音響処理の帰納的構成

(f) 子音区間の挿入誤りの訂正 (check of insertion error)

のプログラムを動作させる。これらチェックプログラムの中には、大文字の leaf で示してある VOWEL RECOGNITION PROGRAM や CONSONANT RECOGNITION PROGRAM を必要とするものがある。これらの REC. PROG. は、複雑な認識プロセスを必要としており、tree 上部の node にある同じ名前のプログラムを利用するような再帰的 (recursive) な構成が望ましい。しかし、このような再帰的な系を実現することは、再帰的呼び出しを禁止している通常の FORTRAN では不可能である。よって、再帰的な系とほぼ同等の系を擬似的に作成するために、最下段の leaf の REC. PROG. の部分に、tree 上部の node にある各々の REC. PROG. の簡略化したプログラムを付け加えた。これにより、深さ 2 あるいは 3 までの再帰的呼び出しのできる系と同等の系を作成した。よって、REC. VOW. → REC. CONSO. → REC. VOW. → REC. CONSO. → REC. VOW. → REC. CONSO. までの呼び出しと、REC. → CONSO. → REC. VOW. → REC. CONSO. → REC. VOW. → REC. CONSO. までの擬似的な再帰的呼び出しが可能となった。このような再帰的な構成は、システムのパフォーマンスの上でも有効であることを文献 (104) での認識実験で確かめた。

以下、図 6-1 の tree で小文字で表わされている各 node と leaf について説明を加える。

6.3.5 セグメンテーション

音韻単位のセグメンテーションを、子音区間候補のラベリングをもとに行う。子音区間のラベリングの手順は通常の子音区間の検出にはよく動作するが、母音と子音の中間的な性格をもつ音韻 (例えば、撥音 /N/, 半母音 /y/, 拗音) や、連続母音のセグメンテーションに対しては不十分である。これらの音韻に対しては、その音韻の性質を利用した個別の処理が不可欠となる。

(I) 子音区間の候補のラベリング

子音区間の候補を、雑音スペクトルの差①、パワー・レベル②、パワー・ディップ③、短時間スペクトル変化④を用いて検出する。このラベリングの手順を図 6-2 に示す。NOISE①と POWER②によって、無音区間や有声子音の buzz 部が検出される。さらに、POW-DIP③によって、鼻子音、有声子音が検出される。

また、SPEC. (S)^④によって、母音区間と子音区間の境界、 / r / などが検出される。この手順によって、ほとんどすべての無声子音、有声子音、半母音 / w / の区間と、一部の半母音 / y / と撥音 / N / の区間が検出される。

このラベリングの例を、図 6-3 ~ 図 6-5 に示す。図 6-3 ~ 図 6-5 は、各々、「6時22分の」、「姫路駅より」、「新倉敷から」と発声した音声である。これらの図の CONSO. LAB. の欄に、ラベリングの結果（スムージングの直前の結果）を示す。これらの図には、基本特徴量も示されている。この図で、C は雑音に近い (< 10) かパワーが小さい (< 0.25) フレームを表わす。D はパワー・ディップ (> 300) を、S は短時間スペクトル変化が大きい (> 2.4) フレームを表わしている。C, D, S での子音区間のラベリングを行い、さらに、ラベルのスムージングを次の手順で行う。

(ラベルのスムージング)

(i) 2 フレーム以下の空白を埋める。

DDDS □ □ DDD → DDDSSSSDDD

(ii) 長さ 6 以上の S のラベルはその 1 部を消去する。

DDSSSSSSSSDDD → DDDSS □ □ SSDDD

DDSSSSSSSS □ □ → DDD □ □ □ SSSS □ □

(iii) 長さ 6 以上の D だけからなる子音区間の候補を抽出し、SPEC. (S) が 1.5 以下で、かつ、BANDPASS が -0.75 以下ならば、この区間のラベルを消去する。

スムージングの後に、音声の文頭の処理と、音声の語尾の処理を行う。各々について

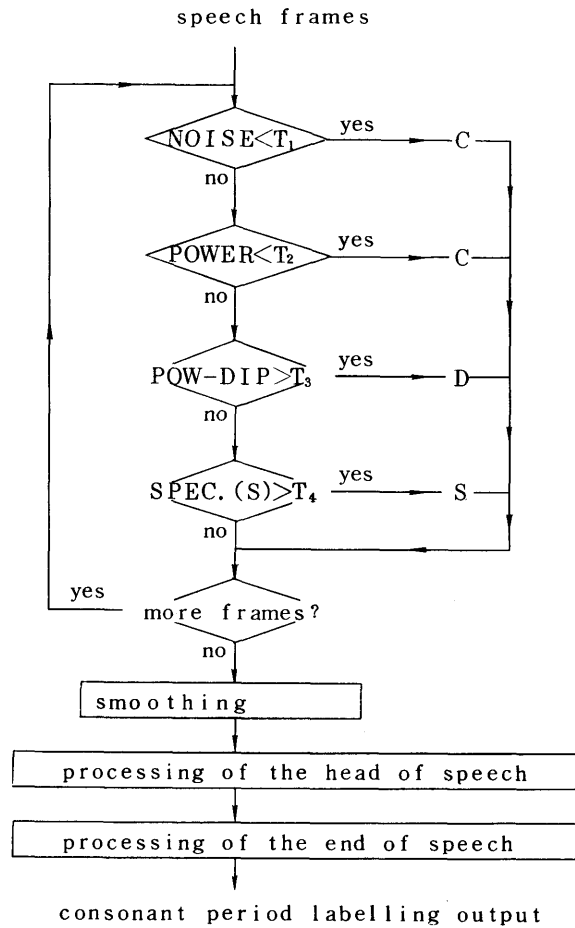


図 6-2 子音区間の候補のラベリングの手順

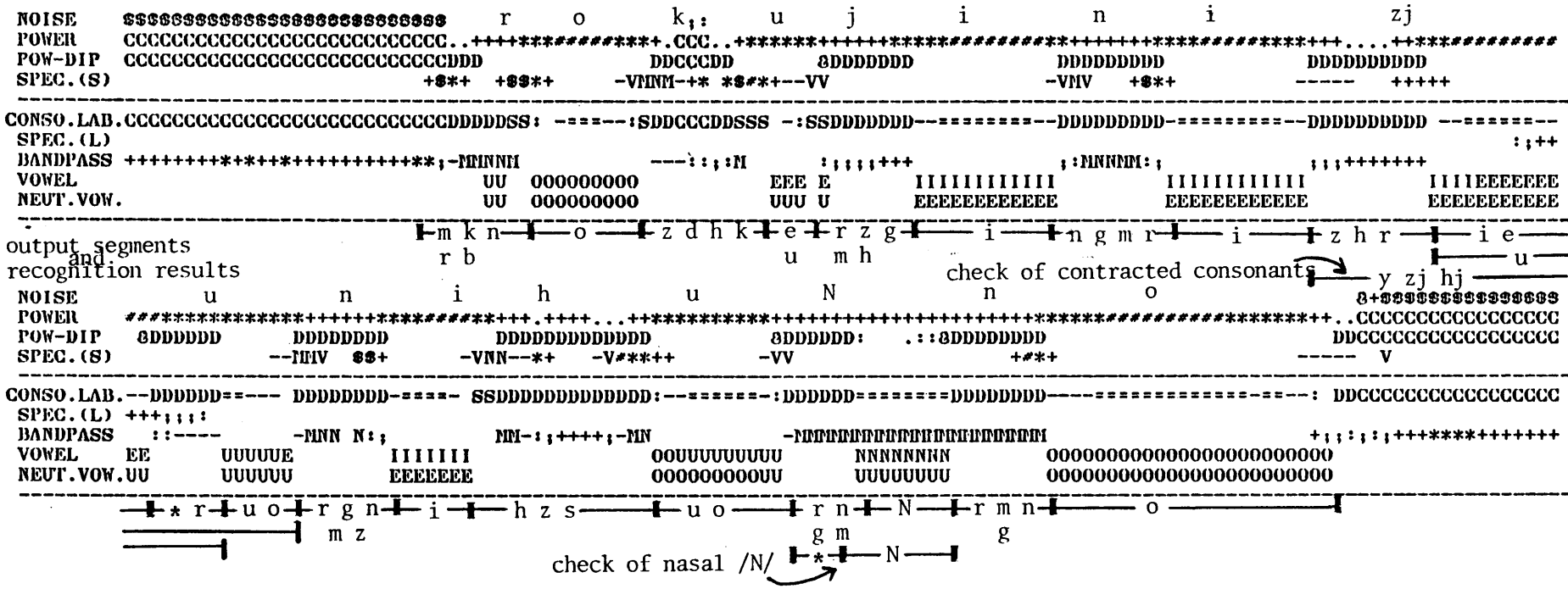


図 6-3 拗音のセグメンテーションと撥音 /N/ のセグメンテーションの例
 入力音声は「6時22分の」

簡単に説明する。

(文頭の処理)

- (i) 文頭で POWER の値が 0.5 以下で、かつ、母音の第 1 位の候補が /N/ であるときには voice bar であるとして、ラベル C をつける。

(語尾の処理)

- (i) 語尾の母音区間の候補で、その継続時間が短く、かつ、POWER の値が 0.5 以下の区間は、ラベル C をつける。

上記の子音区間の候補ラベリングの後、図 6-1 に示した過程に従って、次のセグメンテーションのプログラムを動かす。

- (a) 連続母音の検出 (check of successive vowels)
- (b) 拗音、半母音 /y/ の検出 (check of contracted consonants)
- (c) 撥音 /N/ の検出 (check of nasal /N/)
- (d) 無声化母音の検出 (check of devocalized vowels)
- (e) 文頭の子音のセグメンテーション (check of head consonants)
- (f) 子音区間の挿入誤りの訂正 (correct of insertion error)

これらのセグメンテーションのためのチェックプログラムは独立に動作し、チェックの結果、セグメントの存在の可能性があるならば、そのセグメントと音韻認識の結果を音韻ラティス上に出力する。

(II) 連続母音のセグメンテーション

15 フレーム (120 msec) 以上続く母音区間は、連続する 2 つ以上の母音が存在している可能性があるとして、長時間スペクトル変化⑤を用いて、その可能性がチェックされる。子音部の影響がないフレームでの長時間スペクトル変化のみを用いてチェックすべきであるので、母音区間の候補の先頭の 7 フレームと最後の 7 フレームのチェックは行わない。SPEC. (L) が 1.8 以上のフレームが存在すれば、2 個以上の母音が存在するとして、そのフレームで母音区間を分割する。SPEC. (L) の最大値が、3.0 以下のときには、バイパス・セグメントの存在も許す。この処理の例を、「himejiekiyorori」と発声した図 6-4 に示す。この例では、2 箇所 (ie と iyo) に連続母音区間が存在している。ie の区間の SPEC. (L) の値は余り大きくなく、よって、i * e のセグメントの他に、バイパス・セグメント (i e) の可能性も存在する。また、

iyoの区間は、SPEC.(L)の値が大きく(Sで表わされている)、一意にセグメンテーションされている。なお、母音に狭まれた子音は、CONSONANT RECOGNITION PROGRAMで認識される。

また、母音区間の候補が/o ya/, /a ya/などのような場合には、SPEC.(L)の値が2つの近接したピークをもつことがある。このチェックを、ピーク間のフレームの距離、および第1位の母音候補が/i/あるいは/e/であることを用いて行う。

(Ⅲ) 拗音、半母音/y/のセグメンテーション

拗音は子音と母音とが結合したものとみなすことができ、両者の性質を合せもつ。その特徴の一つは、子音部の直後のスペクトルの動き、すなわち、母音/i/付近から/e/付近を通して、ターゲットの母音に接近するスペクトルの動きである。また、このスペクトルの動きの間には、急激なスペクトル変化④のないことも特徴である。これらの特徴を用いて、下記の手順で拗音を切り出すことを試みる。

(i) 母音の候補⑥の第1位の母音系列(VOWEL 1)で、I…IE…EX…Xの系列、あるいは、E…EX…Xの系列を深がす。ここで、Xは、A、U、Oのいずれかである。系列の途中へのIあるいはNの2フレーム以下の挿入は無視する。ただし、系列の途中で、SPEC.(S)が0.3以上、あるいはPOWERが0.5以下のフレームがある場合には、上記の系列は存在しないとする。さらに、I…IE…Eの部分あるいはE…Eの部分にPOW-DIPが5以上か、あるいは、SPEC.(L)が2以上のフレームがあれば、そこまでの区間内に拗音区間が存在するとして、I…IE…Eあるいは、E…EE部分の長さを出力する。

(ii) 系列がI…IE…EあるいはE…Eのみで終わっているときには、/i/と/e/の先頭付近のCOSH尺度の値と、終りの部分の/i/と/e/のCOSH尺度の値をチェックする。その他のチェックの条件は(i)と同じである。

上記の手順で、I…IE…EあるいはE…E部分の長さが出力される。この長さをもとに拗音あるいは半母音/y/の部分の切り出す。

図6-3は、拗音/zj/のセグメンテーションの例を含む「rokujinizjunifuNno」と発声した音声の処理例である。この例で拗音/zj/の区間は、母音系列のI→E→Uの系列から決定されている。

半母音/y/についても同様のアルゴリズムで/y/の区間の検出と識別が行われ

る。

(IV) 撥音／N／のセグメンテーション

撥音／N／は、母音と子音の中間的な性質を有している。／N／の過半数は、パワー・ディップ③によって始端が検出される。パワー・ディップが生じない／N／は、母音的に取り扱われ、連続母音のセグメンテーションのアルゴリズムによって、標準パターンとのスペクトル・マッチングで識別される。ディップで始端が検出された／N／は、区間全体が、短時間スペクトル変化④と高域と低域の比⑧を用いて検出される。この手順を次に示す。

- (i) ディップの終端付近の SPEC. (S) の値が 1.0 以下で、かつ、BANDPASS の値が 0.75 以上、POWER の値が 0.8 以下である。
- (ii) 継続フレームが 15 以上のときは／N／+ (／n, m, g／) の識別も行う。
- (iii) この他に、／N／のチェックとして、NOISE の値、継続時間の情報も補助的に用いる。

撥音／N／の処理の例を、図 6-3 (「rokujinizjunihuNno」) と図 6-5 (「siNkurasikikara」) に示す。これらの図の BANDPASS の欄で、V あるいは N の印は、低域のスペクトル・エネルギーが大きいことを示している。

(V) 無声化母音のセグメンテーション

無声化しかけた母音／i, u／の検出を、雑音スペクトルとの差①を用いて行う。／i／と／u／の識別は、低域と高域のエネルギーの比⑧を用いて行う。NOISE が 10.0 以下のフレーム数が 20 フレーム以上の子音区間の候補は、無声化母音のチェックを行う。区間の中央部分に、そのまわりの区間より NOISE の値が大きい区間が存在するならば、無声化母音が存在する可能性があるとして、無声化母音の識別および、前後の無声子音の識別を行う。この処理の例を図 6-5 (「siNkurasikikara」) に示す。

(VI) 文頭の子音のセグメンテーション

文頭の有声子音／m, n, r, b, d, g／や無声子音／s, k, t, h／などでは、子音区間全体が完全にはラベリングされないことがある。よって、文頭では、パワー②、

短時間スペクトル変化④，低域と高域のエネルギー比⑧を用いて，子音区間の検出を精密に行う。なお，文中の無声子音 / s , k , t / についても摩擦部や破裂部がラベリングされないことがあるので，同様の処理を文中でも行う。以下詳しく述べる。

(1) 文頭の / g , m , h / のチェック

文頭の母頭区間の候補で，すべてのフレームが， $\text{SPEC. (S)} < 0.2$ and $\text{VOWEL 1} = /N/$ and $\text{POWER} < 0.6$ で，次の子音区間の候補のラベリングが D ,あるいは， $\text{SPEC. (S)} < 0.2$ であるならば，文頭の母音区間も子音区間の一部とみなして，子音認識を行う。

(2) 文頭の / r / のチェック

最初の子音区間の終端部分で，パワーが少しずつ減少しているか，あるいは， $\text{SPEC. (S)} < 0.2$ であり，次の子音区間の候補のラベリングがすべて S であるならば，文頭の母音区間を子音区間の一部と見なして子音認識を行う。

(3) / s , h / のチェック

母音区間が / s , h / の一部であるか否かをチェックする。母音区間のすべてのフレームで，次の条件を満たせば，子音区間の一部と見なす。

$$\left\{ \begin{array}{l} \text{POWER} < 0.85 \\ \text{NOISE} > 0.1 \\ \text{SPEC. (S)} < 0.4 \\ \text{BANDPASS} < -0.15 \end{array} \right.$$

このチェックは，文中の区間にも適用される。

(4) 文頭の有声子音のチェック

母音区間の候補の長さが 7 以下で，

$$\text{BANDPASS} > 0.0 \text{ and } \text{POWER} < 0.4$$

であるときには，子音区間の一部とみなす。

(5) / k , t / のチェック

無声子音の破裂部が母音区間の候補となることがある。このチェックを無音部 ($\text{noise} < 1.0$) の長さ と破裂部の長さの比，および， BANDPASS の値 (< -1.0) を用いて行う。

(VII) 子音区間の候補の挿入誤りの訂正

子音区間であるとしてラベリングされた区間が、本当に子音区間であるか否かのチェックを、継続時間、短時間スペクトル変化④などを用いてチェックする。もし、存在が不明確であるならば、両側の母音区間を併合して、母音認識のプログラムも動作させる。チェックの手順を以下に示す。

- (i) 子音区間の候補のフレーム中に、 $\text{NOISE} < 0.1$ のフレームがあれば、子音区間は存在するとする。
 - (ii) 子音区間の長さが7フレーム以上であれば、子音区間は存在するとする。
 - (iii) 子音区間の長さが、7フレーム以下であり、かつ、 SPEC. (S) が1.5以下であれば、両側の母音区間を併合も行い、母音認識プログラムを動作させる。
- この例を、図6-5の「siNkurasikikara」の語尾の部分に示す。

6.3.6 母音の識別 (discrimination of vowels)

母音区間内の母音の識別を、短時間スペクトル変化④と、母音標準パターンとのCOSH尺度⑥の値を用いて行う。まず、短時間スペクトル変化 SPEC. (S) のもっとも小さいフレームを見出し、そのフレームでのCOSH尺度の値で母音を認識する。

図6-3～図6-5における CONSO. LAB. の欄で、「=」の記号は、スペクトル変化の小さいフレームを表わしている。ただし、拗音の後や半母音/y/の前後である可能性のある母音区間では、中立母音の標準パターンとのCOSH尺度の値⑦も併用して母音の識別を行う。この手法で、従来の手法よりも高い母音認識率を達成できる⁽¹⁰⁷⁾。さらに、スペクトルの安定したフレームだけで標準パターンとのマッチングを行えばよく、大幅な計算量の削減も可能である。

6.3.7 子音の識別 (discrimination of consonants)

子音区間の子音の識別を、次の特徴量を用いて行う。

- (i) 子音区間の長さ
(子音区間のフレーム数)
- (ii) 雑音スペクトル①に近い区分数(1)
($\text{NOISE} < 0.2$ のフレーム数)
- (iii) 雑音スペクトル①に近い区分数(2)

- (NOISE < 1.2 のフレーム数)
- (iv) 雑音スペクトル①に近い区分数(3)
(NOISE < 1.2 かつ NOISE \geq 0.2) のフレーム数)
- (v) パワー②の最小値
(POWER の値が最小のフレームの POWER の値)
- (vi) 短時間スペクトル変化④の最大値
(SPEC. (S) の値が最大のフレームの SPEC. (S) の値)
- (vii) 短時間スペクトル変化の小さい区分数
(SPEC. (S) < 1.5 のフレーム数)
- (viii) バーストの強さ
(SPEC. (S), POWER, NOISE の値から, 擬似的に, バーストの強さを計算する)
- (ix) 低域と高域のエネルギー比⑧
(子音区間の中央のフレームでの低域と高域のエネルギー比⑧を計算する)
- また, 文頭の子音では, 次の特徴量を用いて行う。
- (i) 子音区間の長さ
- (ii) 雑音スペクトルに近くない区分数
(NOISE > 0.2 のフレーム数)
- (iii) バースト強さ
- (iv) 短時間スペクトル変化の最大値
- (v) 低域と高域のエネルギー比

なお, 拗音と半母音の識別には, (Ⅲ) で抽出される / j / 部の長さの情報も利用する。

これらの特徴量に関して, 子音ごとに, 図 6-6 に示す台形の関数で, 各々の特徴量によるスコアを近似する。

よって, 子音のスコアを, これらの特徴量によるスコアの合計で表わす。音韻のスコアは, -100 ~ 50 の値をとりうる。すべての音韻のスコアを比較して, 子音の候補を音韻ラテイスに出力する。

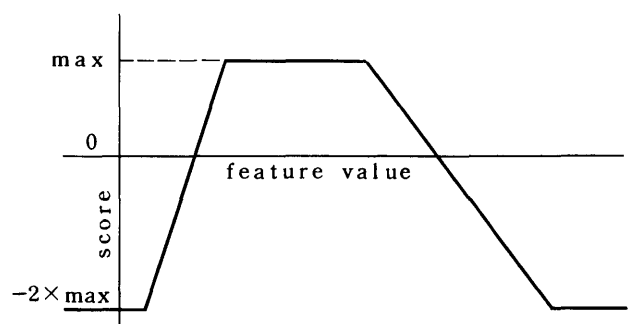


図 6-6 特徴量に対するスコア関数

6.3.8 Voice Q-A System I の音響処理との比較

日本語音声の処理の基本単位として、CV音節のリズムを用いるという System I の考え方を Voice Q-A System II-1 でも受けついでいる。しかし、多くの点で、新しい手法を用いている。表 6-1 に、System I と System II-1 の主な相違点をまとめておく。

表 6-1 Voice Q-A System I と Voice Q-A System II-1 との主な相異点

		Voice Q-A System I	Voice Q-A System II-1
前 処 理	帯域フィルタ	3.2 kHz low-pass	4.0 kHz low-pass
	入力レベルへの適応	なし	あり
	分析ステップ	15 msec (1/2 オーバラップ)	8 msec (3/4 オーバラップ)
	高域強調	なし	あり(差分)
理	スペクトル平坦化	なし	あり(母音標準パターンを基準)
	Lag-window	なし	あり
セグメンテーションで用いる情報		パワー, 母音系列	対数パワー, パワーディップ, スペクトル変化
標準パターン		母音(6種); 連続発声中の平均, 子音(538種); VCV音節, 特定発声者のを流用	母音(6種), 中立母音(6種), 周囲雑音。単独に区切って発声した母音から推定
母音認識		最尤法(正規化予測残差)で求めた母音系列から決定	スペクトルの安定点での COSH 尺度の値から決定
子音認識		VCV音節標準パターンとの端点フリー DP マッチング	継続時間, 雑音との距離, パワー, スペクトル変化, 低・高域のスペクトル・エネルギーから識別

6.4 Voice Q-A System II-1 の性能評価

騒音レベル約 70 dB(A) の計算機室で、ダイナミック・マイクロホンから音声を入力した。入力音声は、文節ごとに区切って発声した新幹線の座席予約の文章である。予約項目の発声順序は自由である。発声者は男性 9 名で、各発声者とも 20 文章(124 文節)を発声した。認識実験に用いた発声リストは付録 3 のグループ 1 のものである。発声者 9 名のうち 6 名は、Voice Q-A System I と共通の発声者である。System II-1 では、母音の標準パターンを、単独に区切って /a, i, u, e, o, N/ と発声した音声から、ホルマント

の周波数とバンド幅の変更によって作成した。⁽⁹⁶⁾

6.4.1 節では、認識実験の結果と、Voice Q-A System I との性能の比較について述べる。6.4.2 節では、母音の標準パターンについて考察する。ここでは、標準パターンとして、単独に発声したものをそのまま用いた場合と平均化母音を用いた場合とをとりあげ、各々の場合の性能を示し、推定母音の場合との比較を行う。6.4.3 節では、System II-1 での文節認識における誤りと拒絶の原因について述べる。

以上の性能評価を、文節認識率（予約項目の認識率）と音韻ラティスの音韻認識率とから行う。音韻ラティスの音韻認識率は、4 節で述べた手法により、音韻ラティスの相互情報量から推定した。本論文では、日本語を CV の単位でとらえているため、母音の中に撥音 /N/ を含め、さらに、連続母音の境界に仮想子音 /*/ を考える。よって、母音認識率は /N/ を含んでおり、子音認識率は /*/ を含んでいる。

6.4.1 性能評価

Voice Q-A System II-1 の認識実験の結果を表 6-2 に示す。また、表 6-2 の最下段に従来の Voice Q-A System I の性能も示しておく。System II-1 では、9 人の発声者の平均で 96.5% の文節認識率が得られた。よって、文節認識率で、System I の 86% に比べ、10.5% の向上が達成できた。また、System II-1 の音韻認識率の平均は、66.3% であり、System I の 58.5% に比べ、7.8% の向上が達成できた。System I と共通の 6 名の発声者 (RN, HN, KI, KS, SA, MK) の全員について、文節認識率、音韻認識率の向上がみられた。とくに、System I での文節認識率が 70%~90% の間の発声者 5 名については、10% 余りの文節認識率の向上が達成できた。これらの結果は、System II-1 の音響処理が有効であることを示している。

Voice Q-A System では、複数の認識結果の出力を許している。文節認識率 (total の欄) には、正しい出力を含んだ複数出力も含まれている。文節認識率のうち約 37% 余りは一意の出力 (unique の欄) である。残りの 60% 余りが、複数出力である。とくに似かよった音韻系列が異なる意味をもつ発時刻、列車名、日付の予約項目では複数出力が多い。複数出力 (not unique の欄) のうち、第 1 位の候補が正しい割合 (= 1 の欄) は、平均で 72.5% である。第 2 位 (≤ 2 の欄) までに 91.1% まで、第 3 位 (≤ 3 の欄) までに 96.7% まで正解が含まれる。実際には、これらの

表 6-2 Voice Q-A System II-1 の性能 (推定母音標準パターン)

(%)

		speaker	RN	HN	KI	KS	SA	MK	NA	KH	MS	average		
		native place	兵庫	千葉	東京	岐阜	兵庫	愛知	東京	茨城	栃木	—		
Voice Q-A System II-1	estimated vowel templates	phrase recognition rate	correct	total	100.0	98.3	100.0	98.3	92.2	94.8	94.8	95.7	94.8	96.5
				unique	44.3	32.7	46.1	37.2	42.5	35.8	38.5	32.7	25.7	37.3
				not + unique	= 1	73.4	80.3	80.6	62.0	75.4	74.3	74.6	68.9	63.0
		≤ 2	90.6	92.1	95.2	88.7	98.4	88.6	91.0	90.5	85.2	91.1		
		≤ 3	96.9	97.4	95.2	93.0	100.0	95.7	98.5	97.3	95.1	96.7		
		reject	0.0	0.9	0.0	1.7	1.7	3.5	1.7	1.7	1.7	1.7	1.4	
	error	0.0	0.9	0.0	0.0	6.1	1.7	3.5	2.6	3.5	2.1			
	phoneme recognition rate	average	correct	65.6	63.6	71.2	67.5	68.1	66.2	66.4	66.6	61.4	66.3	
			rec. error	22.1	23.9	18.2	20.2	19.6	23.4	22.2	22.0	24.0	21.7	
		seg. mis.	12.3	12.5	10.6	12.3	12.3	10.4	11.4	11.4	14.6	12.0		
		(no seg.)	(5.0)	(4.2)	(4.1)	(4.8)	(4.4)	(3.9)	(2.9)	(3.3)	(5.8)	(4.3)		
	vowel ⁺⁺	79.7	76.5	85.2	82.4	80.4	78.8	78.9	79.1	76.5	79.7			
consonant	51.5	50.7	57.1	52.5	55.8	53.6	53.9	54.1	46.3	52.8				
System I	ave- raged vowels	phrase rec. rate	95.7	82.2	90.0	85.2	74.3	83.0	—	—	—	86.0		
		phoneme rec. rate	62.7	55.2	59.1	58.2	55.6	57.1	—	—	—	58.5		

+ Not unique means plural output. Score in = 1 is the percentage of the correct 1st candidate in plural output. Score in ≤ 2 is the percentage the 1st or 2nd candidate is correct. Score in ≤ 3 is the percentage the best 3 candidates include correct one.

++ In our system, vowels include nasal /N/.

あいまいな認識結果は、発声者に確認することなく、質問回答の過程での時刻表を用いた推論によって一意に決定されることが多い。また、拒絶率(rejectの欄)は1.4%、誤り率(errorの欄)は2.1%であった。発声者ごとの詳しい認識結果を付録8に示しておく。

音韻認識におけるセグメント化誤り率(seg. mis.の欄)は12.0%であり、System Iのセグメント化誤り率26.0%の $\frac{1}{2}$ 以下になっている。これから、System II-1の音韻単位へのセグメンテーションのアルゴリズムが大幅に改善されていることがわかる。また、System II-1において、音韻ラティス中に正しいセグメントの存在していない割合(no seg.の欄)は、4.3%である。よって、セグメント化誤り率のうちの7.7%は、音韻ラティス中に含まれるバイパス・セグメントによっていることがわかる。また、子音認識率(consonantの欄)は52.8%であり、System Iより10%余り向上している。この10%の向上は、主として、セグメンテーションの精密化によるものと考えられる。母音認識率(vowelの欄)は79.7%で、System Iより約2%向上している。母音認識率の向上が比較的小さい原因として、次の事柄が考えられる。

- (i) 母音標準パターンを単独発声母音から推定していること。
- (ii) 音韻ラティスで母音の候補を3個(従来は2個)まで許していること。
- (iii) 撥音/N/の処理が不充分であること。

今後とも、これらの母音認識の問題点に留意して、撥音/N/の処理、LPC スペクトル・マッチングの尺度などについて検討する必要がある。

従来のシステム(Voice Q-A System I)と現システム(Voice Q-A System II-1)との性能の比較を図6-7に示す。図6-7で、横軸は音韻認識率、縦軸は文節認識率である。図中で、白丸○はSystem Iの発声者ごとの性能を表わす。右上りの曲線は、System Iの認識実験の結果から推定された音韻認識率と文節認識率との関係を示している。System IIの研究を開始する際に、このグラフより、System IIの最終的な性能の目標として、すべての発声者で音韻認識率63%を越えるという目標値を設定した。黒丸●はSystem II-1での発声者ごとの音韻認識率と文節認識率を示している。Voice Q-A System II-1では、音韻認識率で平均7.8%、文節認識率で平均10.5%向上している。しかしながら、すべての発声者で、当初の目標(音韻認識率63%、文節認識率95%)を達成するには、なお一層の努力が必要である。

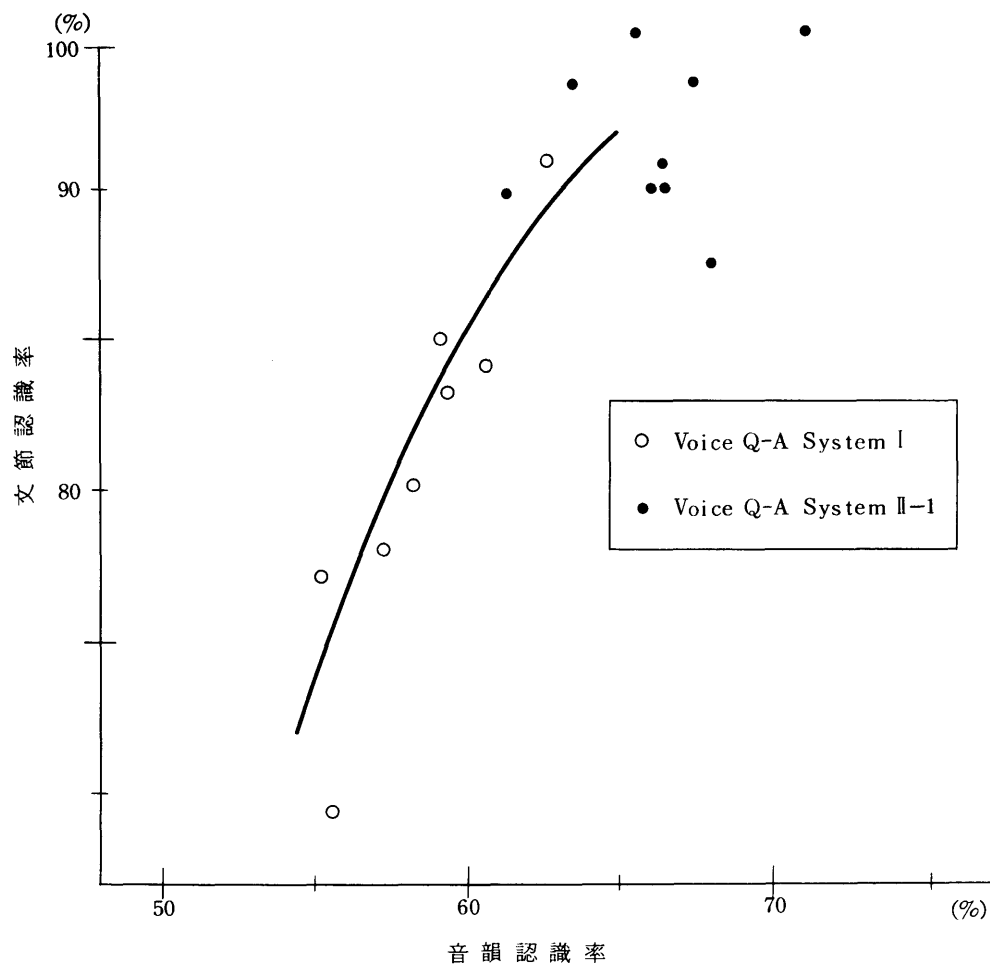


図 6-7 音韻認識率と文節認識率の関係

6.4.2 標準パターンの効果

Voice Q-A System Iでは、連続音声中の母音区間の相関係数を平均して、母音の標準パターンを作成していた。System IIでは、単独に区切って発声した母音から、ホルマントとバンド幅を変更することによって推定している。⁽⁹⁶⁾ この推定方法の効果を9人の発声者で確かめるために、母音の標準パターンを変更して、System II-1で認識実験を行った。母音標準パターンとして、単独に発声した母音をそのまま用いた場合 (isolate vowels) と、System Iと同様に連続音声中の平均を用いた場合 (averaged vowels) とをとりあげた。この結果を表6-3、表6-4に示す。

単独発声した母音をそのまま標準パターンとして用いた場合には、文節認識率92.1%、音韻認識率63.3%であり、推定母音の標準パターンを用いた場合より、各々、4.4%、3.0%低い。これから、推定母音を用いることの有効性がわかる。

表6-3 Voice Q-A System II-1の性能(単独母音標準パターン)

(%)

speaker		RN	HN	KI	KS	SA	MK	NA	KH	MS	average			
native place		兵庫	千葉	東京	岐阜	兵庫	愛知	東京	茨城	栃木	—			
Voice Q - A System II - 1	isolatedly uttered vowel templates	phrase recognition rate	correct	total	99.1	93.9	100.0	86.1	89.6	87.8	90.4	92.2	89.6	92.1
				unique	44.7	34.3	41.7	33.0	36.9	39.6	46.2	34.0	31.1	37.9
			not unique	≦ 1	71.4	81.7	74.6	72.7	67.7	73.8	73.2	62.9	60.6	71.0
				≦ 2	93.7	94.4	89.6	95.5	87.7	88.5	83.9	87.1	88.7	89.9
				≦ 3	98.4	97.2	98.5	98.5	100.0	100.0	92.9	90.0	94.4	96.7
			reject	0.9	0.9	0.0	5.2	4.3	3.5	4.3	2.6	4.3	2.9	
	error	0.0	5.2	0.0	8.7	6.1	8.7	5.2	5.2	6.1	5.0			
	phoneme recognition rate	average	correct	63.5	63.5	67.3	64.2	63.6	64.3	64.4	60.6	58.5	63.3	
			rec. error	23.2	22.9	21.3	21.9	23.7	24.6	23.3	28.3	26.0	23.9	
			seg. mis.	13.3	13.6	11.4	13.9	12.8	11.1	12.3	11.1	15.5	12.8	
			(no seg.)	(5.2)	(4.6)	(4.8)	(6.2)	(4.6)	(3.7)	(3.3)	(3.1)	(6.0)	(4.6)	
		vowel	76.6	76.7	79.3	76.2	74.4	74.2	74.9	69.7	74.2	75.1		
consonant		50.4	50.3	55.2	52.2	52.7	54.4	53.9	51.6	42.8	51.5			

表 6-4 Voice Q - A System II - 1 の性能 (平均化母音標準パターン)

(%)

		speaker	RN	HN	KI	KS	SA	MK	NA	KH	MS	average		
		native place	兵庫	千葉	東京	岐阜	兵庫	愛知	東京	茨城	栃木	-		
Voice Q - A System II-1	averaged vowel templates	phrase recognition rate	correct	total	97.4	98.3	100.0	99.1	92.2	98.3	94.8	93.9	93.9	96.5
				unique	43.8	35.4	51.3	44.7	46.2	43.4	46.8	43.5	38.0	43.7
			not unique	= 1	73.0	80.8	82.1	76.2	78.9	73.4	70.7	75.4	61.2	74.6
		≧ 2		92.1	95.9	94.6	95.2	96.5	92.2	93.1	91.8	91.0	93.6	
		≧ 3		100.0	98.6	98.2	96.8	100.0	98.4	100.0	96.7	97.0	98.4	
		reject	1.7	0.9	0.0	0.0	4.3	0.9	2.6	1.7	3.5	1.7		
	error	0.9	0.9	0.0	0.9	3.5	0.9	2.6	4.3	2.6	1.8			
	phoneme recognition rate	average	correct	65.1	65.1	72.8	68.9	67.5	68.7	69.2	69.2	62.0	67.6	
			rec. error	22.4	21.8	16.6	19.8	20.0	21.9	19.5	20.2	23.7	20.7	
			seg. mis.	12.5	13.0	10.6	11.3	12.5	99.5	11.2	10.6	14.4	11.7	
			(no seg.)	(5.0)	(5.0)	(5.2)	(4.6)	(4.6)	(2.5)	(3.1)	(2.9)	(6.8)	(4.4)	
		vowel	78.7	78.7	86.9	83.8	79.8	81.6	83.7	83.6	78.5	81.7		
consonant	51.6	51.6	58.6	54.0	55.3	55.7	54.8	54.9	45.5	53.6				

平均化母音の標準パターンを用いた場合には、文節認識率 96.5%，音韻認識率 67.6% であり、推定母音の場合と比較して、文節認識率では同じで、音韻認識率で 1.3% 高いだけである。この結果より、母音標準パターンの推定方法は有効であることがわかる。

6.4.3 誤り，拒絶の原因

Voice Q-A System II-1 の言語処理で訂正できなかった誤りや拒絶の原因を分析した。母音標準パターンとして、単独に区切って発声した母音から推定したものをを用いた。誤り拒絶の総数は 36 個（文節）であった。その主要原因をまとめて表 6-5 に示す。

おもな原因は次の 3 項目である。

- (1) 拗音 / zj, kj, hj / のセグメンテーション誤り (17%)。拗音 / zj, kj, hj / の / j / の部分の検出ができなくて、かつ、その誤りの訂正が言語処理でできなかったことに起因する。拗音の検出 (check of contracted consonants) のプログラムを改良する余地がある。
- (2) 子音のセグメンテーション誤り (17%)。セグメンテーション誤りは、調音結合などの影響を受けやすいと推測される子音 / s, n / などで生じている。
- (3) 母音の認識誤り (14%)。前後の音韻環境による母音のなまけによる認識誤りで、言語処理の音韻変形規則で誤り訂正できなかったことに起因する。個人の発声のくせや、方言の影響も見受けられる。

表 6-5 Voice Q-A System II-1 における誤り・拒絶の原因

拗音 / zj, kj, hj / のセグメント化誤り	6
子音のセグメント化誤り	6
母音の認識誤り	5
連続母音のセグメント化誤り	3
半母音 / y / の検出誤り	3
拗音 / N / に関する誤り	3
言語処理による誤り	3
語尾の切り出し誤り	2
ポーズの挿入による誤り	2
子音の認識誤り	2
発声ミス	1

合計 36
(発声者 9 名，各自 124 文節発声)

6.5 Peak Weighted Measureを用いた音響処理⁽⁹⁵⁾

(Voice Q - A System II - 2)

6.3節, 6.4節では, COSH尺度を用いた音響処理 Voice Q - A System II - 1について述べた。COSH尺度は, 従来からのLPCスペクトル・マッチング尺度の中ではすぐれた尺度である。5章の後半で検討したPeak Weighted MeasureのWLR(Weighted Likelihood Ratio)は, 人間の聴覚の特性にあったすぐれたスペクトル・マッチング尺度と考えられる。特にWLRは母音の識別能力が高い。本節では, 5.6節での母音認識におけるWLRのピークの重みづけおよび周波数軸の重みづけの検討に基づいて, Voice Q - A System II - 1の改良を行う。

6.5.1節でVoice Q - A System II - 2の改良点について説明する。6.5.2節では, 実際の動作例を示す。6.6節で, その性能について述べる。

6.5.1 Voice Q - A System II - 2の作成

音響処理における母音認識の精密化を目的として, LPCスペクトル・マッチング尺度をCOSH尺度からWLRに変更する。WLRはLPCスペクトル包絡のピークに重みをおいたスペクトル・マッチング尺度であり, かつ, 周波数軸上での重みづけも可能である。

周波数軸上での重みづけを, 1次系 $(1 - \beta z^{-1})$ で相関関数上で行う。5.6.1節での検討結果に基づいて, $\beta = -0.5$ とする。

ピークの重みづけも, 5.6.2節の結果に基づいて, 相関関数にラグ窓をかけ, スペクトルを平滑化することによって行う。5.6.2節の母音識別実験では, 母音標準パターンとして連続音声での平均を用いているが, Voice Q - A System IIでは, 単独に発声した母音から母音標準パターンを作成している。そのため, 連続音声での平均の母音標準パターンより, スペクトルのずれが大きくなる。よってVoice Q - A System IIでは, スペクトルのずれに対して強いスペクトル・マッチング尺度が望まれる。このスペクトルのずれに対処する一手法として, スペクトルの平滑化が考えられる。5.6.2節の図5-22からわかるように, COSH尺度ではラグ窓による平滑化とともに誤り率が増加するが, WLRはラグ係数300あたりまで, 一定の誤り率を保っている。よって, $\text{lag} = 300$ の値を用いて, 相関関数に重みづけすることにより,

スペクトルの平滑化を行なうことにする。

尺度を COSH 尺度から WLR に変更し、かつ、周波数軸上での重みづけにともない、6.3.3 節で述べた基本特徴量の計算の一部を変更する。以下、変更点について述べる。

- (i) 雑音スペクトルとの差 (NOISE) の計算を COSH 尺度から LPC ケプストラム距離に変更する。よって、雑音スペクトル ($u, r_i, R_f, C_n^{(f)}$) と入力音声 ($v, s_i, R_g, C_n^{(g)}$) の差を次式で計算する。

$$\text{NOISE} = (\log u \cdot R_f - \log v \cdot R_g)^2 + 2 \cdot \sum_{i=1}^{16} (C_i^{(f)} - C_i^{(g)})^2 \quad (6-10)$$

- (ii) 短時間スペクトル変化 (SPEC, (S)) の計算を COSH 尺度から LPC ケプストラム距離に変更する。

- (iii) 長時間スペクトル変化 (SPEC, (L)) の計算を COSH 尺度から WLR に変更する。

5 フレーム離れたフレーム間のスペクトルの差をパワーを正規化した条件での WLR で計算する。WLR は Peak weighted measure であるので、LPC スペクトル包絡でのピーク (ホルマント) の変化に、COSH 尺度よりも敏感であると考えられる。

- (iv) 母音標準パターンと入力フレームとのスペクトルの差と、COSH 尺度を WLR に変更する。

$$\text{VOWEL} = \sum_{i=1}^{24} (r_i - s_i)(C_i^{(f)} - C_i^{(g)}) \quad (6-11)$$

ここで母音の標準パターンは $\{r_i, C_i^{(f)}\}$ 、入力フレームは $\{s_i, C_i^{(g)}\}$ である。

- (v) 高域と低域のスペクトル・エネルギーの比 (BANDPASS) の計算の際、1 次系 ($1 - \beta z^{-1}$) による傾きの影響を補償する。

このように基本特徴量の計算を変更するのみで、その後のプロセスは Voice Q-A System II-1 とまったく同様である。以上の変更により、Voice Q-A System II-2 では、母音認識率の向上が期待できる。

付録 9 に、4 名の発声者 KI, KS, RN, KH が発声した座席予約文の音声データの LPC スペクトル包絡の系列と、音韻単位のセグメンテーションに用いる基本特徴量の変化を図示しておく。

6.5.2 動作例

発声者K Iが発声した音声「1日の、新大阪から、博多までの、6時2分発の、ひかり19号の、指定券を、9枚、お願いします」を入力をしたときの言語処理の動作の過程を図6-8に示す。この図6-8は、音響処理から送られた音韻ラティス、および、言語処理での単語認識部でマッチングを試み、成功した単語名とその位置のリストを出力したものである。また、各文節入力の最後に、そこまでに把握された予約項目を示すリストが出力されている。この図の単語辞書の出力をもとに、構文解析部の動作を追いかけることができる。とくに、構文が複雑である発時刻の入力(4番目の入力)に対する構文解析部の動作は複雑であり、付録1の構文表現をもとにこの過程をたどることは、構文解析部の理解につながると思われる。

付録10に、同じ発声内容に対する発声者KS, RN, KHの計算機出力結果を示しておく。

6.6 Voice Q-A System II-2の性能評価

男性9名が計算機室で予約項目ごとに区切って発声した新幹線の座席予約を対象として認識実験を行った。この音声データは、Voice Q-A System II-1の評価に用いたものと同じデータである。母音標準パターンもSystem II-1と同じものを用いた(6.4節)。この節では、6.6.1で認識実験の結果についてSystem II-1との比較を行う。6.6.2では、母音標準パターンの効果についても、System II-1との比較を行う。

6.6.1 性能評価

Voice Q-A System II-2の認識実験の結果を表6-6に示す。母音の標準パターンとして、単独に/a, i, u, e, o, N/と切って発声した母音から推定したものを用いている。また、Voice Q-A System II-1と比較するために、System II-1の性能を表6-6に右端の欄に示しておく。LPCスペクトル・マッチング尺度をCOSH尺度からWLRに変更したことにより、0.7%の母音認識率の向上が見られる。また、文節認識率でも若干の向上が見られる。発声者別の詳しい認識結果を付録11に示しておく。言語処理で訂正できなかった誤りや拒絶の原因は、ほとんどVoice

RECOGNITION PROCESS (SPEAKER KI)
 (SYSTEM VOICE Q-A SYSTEM II-2)

(TSUITACHINO SHINOSAKAKARA HAKATAMADE 6JI2FUNHATSUNO
 HIKARI19GONO SHITEKENO 9MAI ONEGAI SHIMASU)

(TSUITACHINO)

PHONEME LATTICE
 (1 ((S R H B T) 4 2))
 (2 ((I E) 7 3)))
 (3 ((K T S) 6 4) ((ZY KY HY) 7 5)))
 (4 ((A) 4 6)))
 (5 ((A) 3 6)))
 (6 ((K H D T S) 7 7)))
 (7 ((I) 4 8)))
 (8 ((N G R M) 2 9)))
 (9 ((O) 11 END)))

***** RECOGNITION PROCESS *****

(TSUITACHI (1) 8 9)
 (NO (8 9) END 9)
 (MIKKA (1) 6)
 (ITSUKA (1) 6)
 (MUIKA (1) 6)
 (2 (1) 3 2 1)
 (SHITEI (1) 3)
 (GA (3) 6)
 (1 (1) 3 2 4))

RECOGNITION RESULT
 (((1) NI CHI) (---KARA) (---MADE) (---JI---FUN HATSU) (--- ---GO) (---SEKI
) (---MAI))

(SHINOSAKAKARA)
 PHONEME LATTICE
 (1 ((S H R) 6 2) ((S) 10 4) ((ZY KY HY) 21 6)))
 (2 ((I) 2 3)))
 (3 ((R S H) 1 4) ((Y HY ZY) 12 6)))
 (4 ((I) 8 5)))
 (5 ((* R) 1 6)))
 (6 ((U) 4 7) ((U O) 4 8) ((U O NN) 6 9)))
 (7 ((Y * R) 2 10)))
 (8 ((Y * R) 2 11)))
 (9 ((*) 1 10)))
 (10 ((*) 1 11) ((O E) 8 12)))
 (11 ((O E) 7 12)))
 (12 ((T K S) 6 13)))
 (13 ((A) 2 14)))
 (14 ((K T D S) 5 15)))
 (15 ((A) 3 16)))
 (16 ((T K D) 5 17)))
 (17 ((A) 3 18) ((A) 6 20) ((A) 12 END)))
 (18 ((* R) 2 19)))
 (19 ((A) 6 END)))
 (20 ((*) 1 21)))
 (21 ((A) 5 END)))

***** RECOGNITION PROCESS *****

(YOKA (1) 14)
 (DE (14) 16)
 (GA (14) 16)
 (KYO (1) 8 11 9 10 12)
 (2 (1) 3 4 5 6))
 (10 (3) 4 5 6 7) 10 11 8 9 12))
 (2 (7 10 11 8 9 12) 10 11))
 (5 (7 10 11 8 9 12) 12 10 11 14))
 (6 (7 10 11 8 9 12) 13))
 (7 (7 10 11 8 9 12) 16))
 (KU (7 10 11 8 9 12) 13))
 (YOKKA (7 10 11 8 9 12) 14))
 (10 (1) 7 10 11 8 9 12))
 (SHIZUOKA (1) 14))
 (KARA (14) 18 19 20 21 END 17))
 (WA (18 19 20 21 END 17)) END 18 19 20 21 17))
 (KYOTO (1) 14))
 (KARA (14) 18 19 20 21 END 17))
 (SHINOSAKA (1) 16))
 (KARA (16) 18 19 20 21 17))

RECOGNITION RESULT
 (((1) NI CHI) (SHINOSAKA SHIZUOKA KYOTO)KARA) (---MADE) (---JI---FUN HATSU)
 (--- ---GO) (---SEKI) (---MAI))

図6-8 Voice Q-A System II の動作例

(HAKATAMADENO)

```

PHONEME LATTICE
( 1 ( ( B S R D M ) 4 2 ) )
( 2 ( ( A ) 2 3 ) )
( 3 ( ( K T D S ) 5 4 ) )
( 4 ( ( A ) 3 5 ) )
( 5 ( ( S K T H ) 5 6 ) )
( 6 ( ( A ) 3 7 ) )
( 7 ( ( M N Z G ) 4 8 ) )
( 8 ( ( A ) 5 9 ) )
( 9 ( ( G Z ) 3 10 ) )
( 10 ( ( E U O ) 4 11 ) )
( 11 ( ( M N G R Z ) 4 12 ) )
( 12 ( ( O ) 9 END ) )

```

*** RECOGNITION PROCESS ***

```

( ODAWARA ( 1 ) 9 )
( HAKATA ( 1 ) 7 8 )
( MADE ( 7 8 ) 11 12 )
( NO ( 11 12 ) END 12 )
( 7 ( 1 ) 5 )
( KU ( 1 ) 2 )
( 5 ( 1 ) 3 )

```

RECOGNITION RESULT

(((1) NI CHI) ((SHINOSAKA SHIZUOKA KYOTO) KARA) ((HAKATA) MADE) (-- J I -- FUN HATSU) (--- GO) (--- SEKI) (--- MAI))

(ROKUJ INI HUNHATSUNO)

```

PHONEME LATTICE
( 1 ( ( P R H * B ) 1 2 ) ( ( T N M R B ) 4 4 ) )
( 2 ( ( U ) 1 3 ) ( ( O ) 6 5 ) )
( 3 ( ( R M ) 1 4 ) )
( 4 ( ( O ) 3 5 ) )
( 5 ( ( T K D ) 6 6 ) )
( 6 ( ( U E ) 1 7 ) )
( 7 ( ( G Z N D M ) 8 8 ) )
( 8 ( ( I ) 4 9 ) )
( 9 ( ( M N R G Z ) 4 10 ) )
( 10 ( ( I ) 3 11 ) )
( 11 ( ( K S H ) 7 12 ) )
( 12 ( ( U O NN ) 11 13 ) )
( 13 ( ( Z H M ) 6 14 ) )
( 14 ( ( A ) 2 15 ) )
( 15 ( ( T S K P ) 9 16 ) )
( 16 ( ( NN ) 2 17 ) )
( 17 ( ( R M ) 1 18 ) )
( 18 ( ( O ) 12 END ) )

```

*** RECOGNITION PROCESS ***

```

( FUTSUKA ( 1 ) 7 )
( ITSUKA ( 1 ) 7 )
( MUIK ( 1 ) 7 )
( TOKA ( 1 ) 7 )
( KYO ( 1 ) 5 )
( 2 ( 1 ) 3 4 1 2 )
( 6 ( 1 ) 7 8 )
( JI ( 7 8 ) 9 10 11 )
( PAUSE ( 9 10 11 ) 12 )
( I ( 12 ) 12 )
( RO ( 12 ) 13 12 )
( 2 ( 9 10 11 ) 11 )
( I ( 9 10 11 ) 11 )
( FUN ( 11 ) 14 15 13 )
( HATSU ( 14 15 13 ) 17 18 16 )
( NO ( 17 18 16 ) END 18 )
( 9 ( 9 10 11 ) 13 )
( ZI ( 9 10 11 ) 11 )
( 8 ( 1 ) 7 8 )
( JI ( 7 8 ) 9 10 11 )
( KU ( 1 ) 3 4 5 2 2 )
( JI ( 3 4 5 2 ) 7 )
( 2 ( 7 ) 9 10 11 )
( I ( 7 ) 9 )
( FUN ( 9 10 11 ) 14 13 15 )
( HATSU ( 14 13 15 ) 17 18 16 )
( ZI ( 7 ) 9 10 11 )

```

RECOGNITION RESULT

(((1) NI CHI) ((SHINOSAKA SHIZUOKA KYOTO) KARA) ((HAKATA) MADE) ((6 8 9) J I (2) FUN HATSU) (--- GO) (--- SEKI) (--- MAI))

(HIKARIZYUKYUGONO)

```

PHONEME LATTICE
1 ( ( H S * P R ) 2 2 ) ( ( S ) 10 4 ) )
2 ( ( I ) 1 3 ) )
3 ( ( K T D ) 5 4 ) )
4 ( ( A ) 5 5 ) )
5 ( ( M R G Z N ) 3 6 ) )
6 ( ( I ) 4 7 ) )
7 ( ( D Z H ) 5 8 ) ( ( ZY KY ) 13 10 ) )
8 ( ( E ) 5 9 ) ( ( E U A ) 11 11 ) )
9 ( ( Y * ) 1 10 ) )
10 ( ( U E ) 3 11 ) )
11 ( ( K T ) 8 12 ) ( ( ZY KY ) 16 14 ) )
12 ( ( E ) 5 13 ) ( ( U ) 11 15 ) )
13 ( ( Y * R ) 2 14 ) )
14 ( ( U O ) 3 15 ) )
15 ( ( G Z M N H ) 4 16 ) )
16 ( ( O ) 13 17 ) )
17 ( ( G Z N M H ) 3 18 ) )
18 ( ( O ) 11 END ) )

```

**** RECOGNITION PROCESS ****

```

( FUTSUKA ( 1 ) 5 6 )
MUIKA ( 1 ) 5 )
2 ( 1 ) 3 )
3 ( 1 ) 6 7 5 )
10 ( 6 7 5 ) 11 )
1 ( 11 ) 13 14 )
HIKARI ( 1 ) 7 )
3 ( 7 ) 11 )
7 ( 7 ) 11 )
8 ( 7 ) 13 14 12 )
9 ( 7 ) 11 )
10 ( 11 ) 15 16 )
2 ( 15 16 ) 16 )
4 ( 15 16 ) 17 )
GO ( 17 ) END )
5 ( 15 16 ) 17 18 16 )
GO ( 17 18 16 ) END )
GO ( 15 16 ) 17 18 )
NO ( 17 18 ) END 18 )
100 ( 7 ) 15 16 13 14 12 )
2 ( 15 16 13 14 12 ) 16 15 )
10 ( 16 15 ) 17 )
2 ( 17 ) 18 )
4 ( 17 ) END )
5 ( 17 ) END 18 )
GO ( 17 ) END )
5 ( 15 16 13 14 12 ) 17 18 16 15 )
10 ( 17 18 16 15 ) 17 )
6 ( 15 16 13 14 12 ) 16 )
10 ( 15 16 13 14 12 ) 17 )
10 ( 7 ) 11 )
9 ( 11 ) 15 )
GO ( 15 ) 17 )
GO ( 16 15 ) 17 18 )
4 ( 15 16 13 14 12 ) 17 16 18 )
GO ( 17 16 18 ) END 17 )
GO ( 17 18 16 15 ) END 17 18 )
GO ( 15 16 13 14 12 ) 17 18 )
SHITEI ( 1 ) 3 )

```

RECOGNITION RESULT

```

(((1)NICHII)((SHINOSAKA SHIZUOKA KYOTO)KARA)((HAKATA)MADE)((6 8 9)JI(2)
FUN HATSU)((HIKARI)(100 105 90 95 104 19 102 110 150 94 120)GO)(---SEKI)(---
MAI))

```

(SHITEKENO)

PHONEME LATTICE

```

1 ( ( S H ) 4 2 ) ( ( S ) 12 4 ) )
2 ( ( I U ) 5 3 ) )
3 ( ( K T S ) 5 4 ) )
4 ( ( E ) 11 5 ) )
5 ( ( K T ) 6 6 ) ( ( ZY KY ) 10 7 ) )
6 ( ( E ) 6 8 ) )
7 ( ( E U ) 2 8 ) )
8 ( * ) 1 9 ) ( * ) 8 11 ) )
9 ( NN ) 7 10 )
10 ( * ) 1 11 )
11 ( O ) 13 END )

```

***** RECOGNITION PROCESS *****

```

2 ( 1 ) 3 )
SHITEI ( 1 ) 5 4 10 11 9 8 )
KEN ( 5 4 10 11 9 8 ) 10 11 8 9 )
NO ( 10 11 8 9 ) END 11 )
FUTSU ( 1 ) 3 )
KEN ( 3 ) 5 )
DE ( 5 ) 8 9 11 )
SEKI ( 3 ) 8 9 11 6 )
NO ( 8 9 11 6 ) END 11 )
1 ( 1 ) 4 )
8 ( 1 ) 4 )

```

RECOGNITION RESULT

((1)NICHI)((SHINOSAKA SHIZUOKA KYOTO)KARA)((HAKATA)MADE)((6 8 9)JI(2) FUN HATSU)((HIKARI)(100 105 90 95 104 19 102 110 150 94 120)GO)((FUTSU)SEKI)(--MAI))

(KYUMAI)

PHONEME LATTICE

```

1 ( ( S T K H R ) 5 2 ) ( ( Y KY HY ) 12 5 ) )
2 ( ( E ) 5 3 ) ( ( E I U ) ( 6 4 ) ( ( U ) 11 6 ) ) )
3 ( * R ) 1 4 )
4 ( * ) 1 5 ) ( ( U O ) 4 6 ) )
5 ( U O NN ) 4 6 ) )
6 ( M N Z H ) 6 7 ) )
7 ( A ) 7 8 ) )
8 ( Y * R G ) 3 9 ) )
9 ( E I U ) 3 END ) )

```

***** RECOGNITION PROCESS *****

```

FUTSUKA ( 1 ) 8 9 )
DE ( 8 9 ) END )
ITSUKA ( 1 ) 8 9 )
MUIKA ( 1 ) 8 9 )
YOKA ( 1 ) 8 9 )
KYO ( 1 ) 6 )
WA ( 6 ) 8 9 )
GA ( 6 ) 8 9 )
2 ( 1 ) 4 5 6 3 2 1 ) 6 )
10 ( 4 ) 5 6 3 2 1 ) 6 )
2 ( 6 ) 7 )
3 ( 6 ) 9 END )
10 ( 1 ) 6 )
SHIZUOKA ( 1 ) 8 9 )
TOYOHASHI ( 1 ) 8 9 )
SHITEI ( 1 ) 6 3 4 5 )
NO ( 6 3 4 5 ) 6 6 )
O ( 6 3 4 5 ) 6 6 )
WA ( 6 3 4 5 ) 8 9 )
GA ( 6 3 4 5 ) 8 9 )
NOWA ( 6 3 4 5 ) 8 9 )
FUTSU ( 1 ) 4 5 6 )
NO ( 4 5 6 ) 6 6 )
O ( 4 5 6 ) 6 6 )
WA ( 4 5 6 ) 8 9 )
GA ( 4 5 6 ) 8 9 )
NOWA ( 4 5 6 ) 8 9 )
GREEN ( 1 ) 9 )
WA ( 6 ) 8 9 )
GA ( 6 ) 8 5 )
1 ( 1 ) 4 6 3 2 )
MAI ( 4 5 6 3 2 ) END 8 9 )
MAI ( 4 5 6 3 2 1 ) END 8 9 )
4 ( 1 ) 7 6 )
MAI ( 7 ) 6 ) END 8 9 )
9 ( 1 ) 6 4 5 )
MAI ( 6 4 5 ) END 8 9 )

```

RECOGNITION RESULT

((1)NICHI)((SHINOSAKA SHIZUOKA KYOTO)KARA)((HAKATA)MADE)((6 8 9)JI(2) FUN HATSU)((HIKARI)(100 105 90 95 104 19 102 110 150 94 120)GO)((FUTSU)SEKI)(4 9 2 1)MAI))

図 6 - 8 (続 き)

(ONEGAISHIMASU)

PHONEME LATTICE

```
( 1 ( ( T * B R G ) 2 2 ) )
( 2 ( ( O ) 3 3 ) )
( 3 ( ( G R N M Z ) 3 4 ) ( ( Y HY ZY ) 11 6 ) )
( 4 ( ( E ) 5 5 ) )
( 5 ( ( * R ) 1 6 ) )
( 6 ( ( A O ) 12 8 ) )
( 7 ( ( A O ) 11 8 ) )
( 8 ( ( S HY ) 7 9 ) )
( 9 ( ( E I U ) 1 10 ) )
( 10 ( ( M R N G Z ) 3 11 ) )
( 11 ( ( A ) 7 END ) )
```

*** RECOGNITION PROCESS ***

```
( YOYAKU ( 1 ) 10 11 9 )
( SHI ( 10 11 9 ) 10 11 )
( MAS ( 10 11 ) END )
( ONEGAI ( 1 ) 6 8 7 )
( SHI ( 6 8 7 ) 10 11 9 )
( MAS ( 10 11 9 ) END )
( 10 ( 1 ) 3 4 )
( 2 ( 3 4 ) 5 6 7 )
( 3 ( 3 4 ) 8 )
( TOYOHASHI ( 1 ) 10 11 9 )
( E ( 10 11 9 ) 10 11 )
( WA ( 10 11 ) END )
( KOKURA ( 1 ) 8 6 7 )
```

RECOGNITION RESULT (WITHOUT INFERENCE)

(((1)NICHI)((SHINOSAKA SHIZUOKA KYOTO)KARA)((TOYOHASHI HAKATA)MADE)((6 8 9)JI
(2)FUN HATSU)((HIKARI)(100 105 90 95 104 19 102 110 150 94 120)GO)((FUTSU
)SEKI)((4 9 2 1)MAI))

PENALTY SCORES FOR RECOGNIZED ITEMS

(((2))((0 7 8))((4 4))((4 7 8)JI (4))((2))(2 2 3 3 4 4
5 6 6 8 9))((3))((1 2 2 3))

RESPONSE SENTENCE

TOYOHASHI MADE DESUKA

FINAL RESULT AFTER INFERENCE

(((1)NICHI)((SHINOSAKA)KARA)((TOYOHASHI HAKATA)MADE)((6)JI(2)FUN HATSU
(HIKARI)(19)GO)((FUTSU)SEKI)((4 9 2 1)MAI))

PENALTY SCORES

(((2))((0))((4 4))((4)JI(4))((100)(100))((3))((1 2 2
3))

表 6-6 Voice Q-A System II-2 の性能 (推定母音標準パターン)

(%)

		speaker		RN	HN	KI	KS	SA	MK	NA	KH	MS	average	System	
		native place		兵庫	千葉	東京	岐阜	兵庫	愛知	東京	茨城	栃木	—	II-1	
Voice Q-A System II-2	estimated vowel templates	phrase recognition rate	correct	total	98.3	95.7	99.1	99.1	93.9	96.5	98.3	93.0	98.3	96.9	96.5
				unique	46.9	45.5	54.4	43.9	41.7	40.5	47.8	40.2	33.6	43.8	37.3
				not unique	≦ 1	68.3	71.7	78.8	64.1	71.4	71.2	76.3	68.8	61.3	70.2
			≦ 2		93.3	95.0	90.4	84.4	92.1	89.4	89.8	92.2	80.0	89.6	91.1
			≦ 3		96.7	96.7	96.2	93.8	96.8	97.0	93.2	98.4	89.3	95.3	96.7
		reject	0.0	2.6	0.0	0.9	2.6	1.7	1.7	1.7	0.0	1.3	1.4		
	error	1.7	1.7	0.9	0.0	3.5	1.7	0.0	5.2	1.7	1.8	2.1			
	phoneme recognition rate	average	correct	65.8	65.0	71.2	65.6	68.4	64.8	69.3	67.0	63.6	66.8	66.3	
			rec. error	21.6	19.7	17.7	21.4	19.7	22.3	19.5	20.9	22.0	20.5	21.7	
			seg. mis.	12.6	15.3	11.1	13.0	11.9	12.9	11.2	12.1	14.5	12.7	12.0	
			(no seg.)	(4.6)	(6.4)	(4.3)	(5.4)	(4.1)	(4.6)	(3.7)	(4.4)	(5.4)	(4.8)	(4.3)	
		vowel	77.9	81.8	84.6	79.2	80.1	78.1	83.3	81.1	77.6	80.4	79.7		
consonant	53.7	48.2	57.7	52.0	56.8	51.4	55.2	53.0	49.5	53.1	52.8				

Q-A System II-1と同じである。Voice Q-A System Iでの認識実験の結果から推定した音韻認識率と文節認識率の関係のグラフと、Voice Q-A System II-2の性能を図6-9に示す。System II-1より、若干、個人による認識率のばらつきが小さくなっているようである。当初の目標である音韻認識率63%を発声者全員の音声データで達成することができた。また、文節認識率の目標である95%は、9人の発声者中7人まで達成することができた。

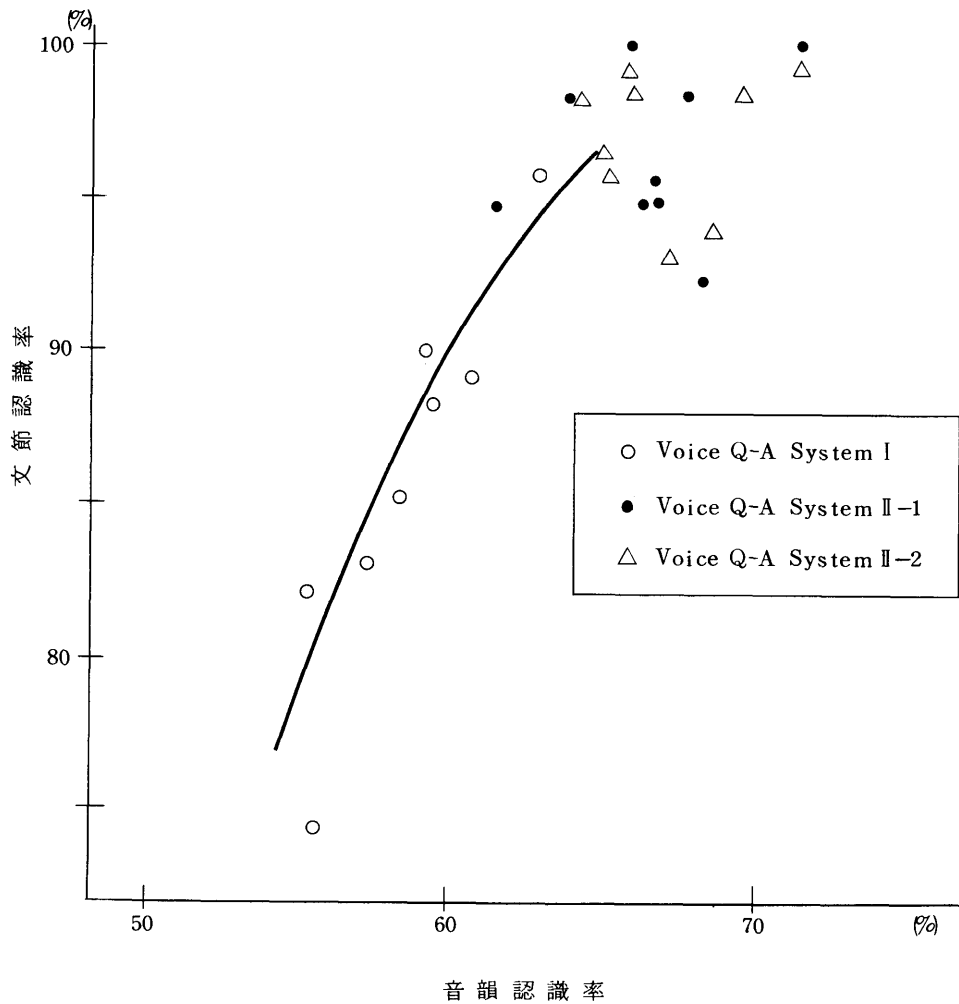


図6-9 音韻認識率と文節認識率の関係

6.6.2 標準パターンの効果

Voice Q-A System II-1と同様に、推定母音以外に、次の2種類の母音標準パターンを用いた場合の認識実験も行った。

(i) 単独に区切って発声した母音をそのまま標準パターンとした場合。(isolate

表6-7 Voice Q-A System II-2の性能(単独母音標準パターン)

(%)

		speaker		RN	HN	KI	KS	SA	MK	NA	KH	MS	average	System II-1	
		native place		兵庫	千葉	東京	岐阜	兵庫	愛知	東京	茨城	栃木	-		
Voice Q-A System II-2	isolatedly uttered vowel templates	phrase recognition rate	correct	total	99.1	91.9	98.3	98.2	88.7	89.6	98.2	93.9	93.1	94.8	92.1
				unique	50.0	35.2	53.1	43.4	40.2	42.7	47.8	37.0	31.8	42.4	37.9
				not unique	= 1	66.7	72.9	84.9	59.4	70.5	74.6	72.9	67.6	57.5	69.7
			≧ 2	96.5	97.1	88.7	87.5	86.9	91.5	94.9	94.1	90.4	92.0	89.9	
			≧ 3	96.5	98.6	98.1	93.8	90.2	94.9	100.0	98.5	98.6	96.6	96.7	
			reject	0.0	3.5	0.0	0.9	3.5	4.3	0.9	2.6	4.3	2.2	2.9	
	error	0.9	2.6	1.7	0.9	7.8	6.1	0.9	3.5	2.6	3.0	5.0			
	phoneme recognition rate	average	correct	63.9	62.3	69.3	65.5	64.2	60.9	68.1	65.0	61.3	64.5	63.3	
			rec. error	22.7	21.9	18.8	21.9	22.5	25.5	20.2	22.1	25.0	22.3	23.9	
			seg. mis.	13.4	15.8	11.9	12.6	13.3	13.7	11.7	12.9	13.7	13.2	12.8	
			(no seg.)	(5.2)	(6.0)	(5.2)	(5.0)	(4.4)	(5.2)	(4.2)	(5.0)	(5.4)	(5.1)	(4.6)	
		vowel	74.7	76.7	82.5	78.5	76.1	72.3	82.3	75.9	72.7	76.9	75.1		
consonant	53.1	47.9	56.2	52.6	52.3	49.5	53.8	54.1	49.9	52.2	51.5				

表6-8 Voice Q-A System II-2の性能(平均化母音標準パターン)

(%)

		speaker		RN	HN	KI	KS	SA	MK	NA	KH	MS	average	System II-1	
		native place		兵庫	千葉	東京	岐阜	兵庫	愛知	東京	茨城	栃木	—		
Voice Q-A System II-2	averaged vowel templates	phrase recognition rate	correct	total	99.1	96.5	99.1	96.5	92.2	96.5	98.3	95.7	95.6	96.6	96.5
				unique	51.8	41.4	53.5	51.4	51.9	43.2	48.7	48.2	43.6	47.1	43.7
				not unique	= 1	69.1	76.9	83.0	79.6	72.5	73.0	77.6	80.7	78.5	75.7
			= 2	92.7	96.9	94.3	94.4	96.1	90.5	94.8	94.7	97.7	93.6	93.6	
			= 3	96.4	98.5	100.0	98.1	98.0	96.8	98.3	100.0	95.9	98.0	98.4	
			reject	0.0	2.6	0.0	2.6	3.5	0.9	0.9	2.6	0.9	1.6	1.7	
		error	0.9	0.9	0.9	0.9	4.3	2.6	0.9	1.7	3.5	1.8	1.8		
		phoneme recognition rate	average	correct	68.0	65.8	73.4	68.5	68.1	66.8	70.2	71.1	64.9	68.5	67.6
				rec. error	19.5	19.0	15.6	18.9	19.6	20.7	18.7	17.5	11.5	19.0	20.7
				seg. mis.	12.5	15.2	11.0	12.6	12.3	12.6	11.1	11.4	13.6	12.5	11.7
				(no seg.)	(5.0)	(6.2)	(4.2)	(5.4)	(4.6)	(4.4)	(3.5)	(4.1)	(5.4)	(4.8)	(4.4)
			vowel	81.3	81.5	88.4	83.4	81.4	80.4	86.0	77.5	89.8	83.3	81.7	
		consonant	54.8	50.2	58.5	53.6	54.7	53.2	54.4	54.7	50.0	53.8	53.6		

vowels)。

(ii) 連続音声の中の平均を用いた場合 (averaged vowels)。

各々の標準パターンを用いた場合の認識実験の結果を表 6-7, 表 6-8 に示す。各々の表の右端の欄に Voice Q-A System II-1 の結果を示しておく。単独母音を用いた場合には, 母音認識率が System II-1 に比べて, 1.8% 向上し, 文節認識率も 2.7% 向上した。平均化母音を用いた場合には, 母音認識率で 1.6% 向上して 83.3% となった。いずれの標準パターンに対しても, COSH 尺度を WLR に変更することによる母音認識率の向上がみられる。特に, 単独に発声した母音を標準パターンとしたときの音韻認識率, 文節認識率の向上が大きい。このことは, WLR が LPC スペクトル包絡の多少のずれに対しても強い尺度であることを示唆している。

6.7 あとがき

列車の座席予約を対象とした会話音声認識システム Voice Q-A System II-1 の音響処理の構成および性能について述べた。本音響処理で, 2 章, 3 章で述べた Voice Q-A System I における問題点を大幅に改善した。その特徴は次の 5 点である。

- (1) 再帰的な構成からなる音響処理を作成した。
- (2) 音韻単位のセグメンテーションを精密化した。
- (3) COSH 尺度によって母音認識の精密化を行った。
- (4) 母音標準パターンの学習法の単純化を行った。
- (5) 発声者によらない子音認識方法を試みた。

本論文では, (1), (2) に重点をおいて説明した。

言語処理と結合して, 男性 9 名の音声データで認識実験を行った結果, 66.3% の音韻認識率, 96.5% の文節認識率を得た。Voice Q-A System I に比べ, 音韻認識率で 8%, 文節認識率で 10% の向上を達成した。また, 母音標準パターンの簡単な学習法の効果も, 認識実験により確かめた。さらに, 誤り拒絶の主原因を分析し, 今後の問題点を明らかにした。

さらに, Voice Q-A System II-1 の母音認識部のスペクトルマッチング尺度を COSH 尺度から Peak weighted measure の WLR に変更した Voice Q-A System

Ⅱ－２を作成し，性能の向上をはかった。男性 9 名の音声データで認識実験を行った。母音認識率で 0.8 % の向上がみられ，音韻認識率 66.8 %，文節認識率 96.9 % が達成できた。

第7章 結 言

本論文では、著者が数年間、日本電信電話公社武蔵野電気通信研究所で行ってきた会話音声の自動認識システムの研究について述べた。

第1章では、会話音声自動認識の研究動向について述べるとともに、会話音声自動認識の問題点をあげ、研究の基本方針及び本論文の構成について述べた。

第2章では、会話音声認識システム (Voice Q-A System) の基本構成について述べ、さらに、認識対象としてとりあげた新幹線の座席予約の概要について述べた。Voice Q-A Systemの基本構成として、

- (1) 音声による質問回答システムを目標とする。そのため音声認識部の他に音声応答部も作成する。
- (2) 音声認識部は、役割が大幅に異なる音響処理と言語処理とを分離して作成し、各々が並列に動作できるようにする。
- (3) 音響処理と言語処理の情報の受け渡しの形式として音韻ラティスの形式をとることにする。

の方針をとることにした。また、認識対象の困難さの評価尺度として冗長度の概念を用いることを提案し、各種の言語情報の冗長度の大きさを評価した。この新幹線の座席予約の対象では、構文情報による冗長度が最も大きいことがわかった。

第3章では、Voice Q-A Systemの言語処理の構成法及びVoice Q-A System Iの性能評価について述べた。言語処理の構成法及びアルゴリズムの特徴を以下にまとめておく。

- (1) 言語処理は単語認識部、構文解析部、推論部からなる認識部と会話モデルからなる。認識部は、top-down的な構成からなっており、left-to-rightのマッチングを基本としている。
- (2) 音韻ラティスと単語辞書とのマッチングのアルゴリズムとして、tree searchのdepth-first methodを基本とする手法を提案した。この手法では、様々な音韻の性質を音韻変形規則の形で利用できる。tree searchの途中過程の記憶や連想的な検索によって実行速度も飛躍的に向上している。
- (3) 構文情報の表現方法として、リスト表現による手法を提案した。この方法は、transition network grammar とほぼ等価な能力を持つ。この構文表現に基づく構文解析

のアルゴリズムを提案した。

- (4) 質問回答が自然に行われることを目的として、会話モデルを作成し、認識システムにおける会話モデルを作成する上の問題点を明らかにした。また、Voice Q-A System I の実時間の5倍の処理時間で動作するオンラインシステムを作成し、男性8名の音声データを用いてSystem Iの言語処理の評価を行った。その結果、音韻認識率58.5%の音韻ラティスを入力として、86.0%の文節認識率を得た。また、質問回答実験も行い、99.1%の予約達成率を得た。この章の最後で、Voice Q-A System Iの問題点を明らかにし、Voice Q-A System IIの作成方針を示した。

第4章では、音響処理の出力である、音韻ラティスの評価方法を提案し、その有効性を確かめた。この章で提案した方法により、音韻認識のあいまいさとセグメンテーションのあいまいさを含む時系列である、音韻ラティスの評価方法が確立された。この方法は、相互情報量に基づいて音韻ラティスの音韻認識率を推定する手法である。この方法では、音韻ラティスからマクロな統計量を抽出することが必要となる。このマクロな統計量を自動的に求めるアルゴリズムをtree searchのbest-first methodで作成し、音韻ラティスの評価を自動化した。この方法によって、音響処理からの多量の音韻ラティスの自動評価が、初めて可能になり、音響処理を客観的に評価する方法が確立された。この方法を、Voice Q-A System I及びVoice Q-A System IIの評価に用いた。

第5章では、音響処理の精密化の一環として、LPCスペクトル・マッチング尺度について検討した。5章の前半では、従来からのLPCスペクトル・マッチング尺度として、LPCケプストラム距離、COSH尺度、最尤スペクトル距離、正規化残差をとりあげ、母音標準パターンとのスペクトル・マッチングによる母音識別実験を行った。その結果、次の結論を得た。

- (1) スペクトルのパワー比を様々な条件に設定し、母音識別実験を行った。その結果、尺度値が最小になるようにパワー比を設定した条件でのマッチングが適当であることを確かめた。
- (2) 従来からLPCスペクトル・マッチング尺度のうち、残差パワーを正規化した条件、あるいは尺度値最小の条件でのCOSH尺度が母音の識別にはすぐれていることが確かめられた。

5章の後半では、新しいLPCスペクトル・マッチング尺度として、LPC Peak weighted measureを提案し、理論的解析及び母音識別実験による評価を行った。その結果、次の結論を得た。

- (1) LPC Peak weighted measureは、LPCスペクトル包絡のピーク付近に重みをおい

たマッチングが可能であり、ホルマント・マッチングとLPCスペクトル・マッチングの両者の長所をあわせもつ尺度であることが母音識別実験で確かめられた。

(2) LPC Peak weighted measureは、周波数軸の重みづけも可能であり、この有効性も、母音識別実験で確かめられた。

(3) LPC Peak weighted measureのうち、WLR(Weighted Likelihood Ratio)は、もっともすぐれた尺度であることが、母音識別実験により確かめられた。

5章の最後では、641都市名を対象として、LPCスペクトル・マッチング尺度による大語い単語音声認識実験を行った。その結果、WLRはもっとも高い認識率を示し、従来からのLPCスペクトル・マッチング尺度よりもすぐれていることが確かめられた。

第6章では、Voice Q-A System Iにおける問題点を大幅に改善することを目的として、Voice Q-A System IIを作成した。とくに、音響処理では、次の点の改善を行った。

- (1) 再帰的な構成からなる音響処理。
- (2) 音韻単位のセグメンテーションの精密化。
- (3) LPCスペクトル・マッチング尺度による母音認識の精密化。
- (4) 母音標準パターンの学習法の簡易化。
- (5) 発声者によらない子音認識法。

まず、LPCスペクトル・マッチング尺度として、COSH尺度をとり上げ、Voice Q-A System II-1を作成した。男性9名の音声データで認識実験を行い、66.3%の音韻認識率、96.5%の文節認識率を達成し、Voice Q-A System Iに比べ、音韻認識率で8%、文節認識率で10%の向上を達成した。次に、マッチング尺度として、Peak weighted measure WLRをとりあげ、Voice Q-A System II-2を作成し、性能の向上をはかった。System II-1に比べ母音認識率で0.8%の向上が達成でき、66.8%の音韻認識率、96.9%の文節認識率を達成した。これにより、すべての発声者で音韻認識率63%、文節認識率95%を達成するという当初の目標をほぼ達成した。

最後に、本論文の研究成果で、音声認識の研究にとって大きく貢献すると思われる項を列挙する。

- (1) 質問回答システムの観点から会話音声認識システムを作成し、高い認識性能を達成した。このことにより、会話音声認識システムの実現の可能性が示された。本論文で作成した、Voice Q-A System IIは、今後の会話音声認識システムの研究の基礎となりうる。
- (2) 言語処理での具体的アルゴリズムを提案し、その有効性を確かめた。とくに、音韻ラティス

を入力した場合の tree search による単語認識アルゴリズム, リスト表現による構文解析アルゴリズムなどを提案した。これらの手法は, 会話音声認識システムの言語処理での基本的な手法となりつつある。また音響処理と言語処理との間の情報の受け渡しとして, 音韻ラティスの形式を提案した。音韻ラティスの形式も様々な研究機関で用いられ始めており, その有効性が確かめられている。

(3) Voice Q-A System II の音響処理を音韻単位のセグメンテーションに重点をおいて再帰的な構成で作成した。この構成は, 今後の音響処理の基本となり得る。また, この音響処理は, LPC 分析に基づいて作成されている。諸研究機関でも LPC 分析に基づいた音声認識の研究を始めており, それらの研究に対して, 先導的な役割を果たすと考えられる。

(4) LPC スペクトル・マッチング尺度について, 母音識別実験の観点から検討した。その結果, 従来の尺度のうちでは, 残差パワーを正規化した条件での LPC ケプストラム距離と COSH 尺度がすぐれていることを明らかにした。さらに, ホルマント・マッチングとスペクトル・マッチングの長所を兼ね備えた Peak weighted measure を提案し, 理論的解析及び母音識別実験でその有効性を確かめた。また, Peak weighted measure を大語い単語認識に通用し, その有効性を確かめた。この Peak weighted measure の提案により, スペクトル・マッチングの新しい分野が開けた。

(5) 音韻ラティスの自動評価システムを作成し, 音響処理の客観的な評価方法を確立した。

この方法は, 会話音声認識システムの研究を進めるうえで, 今後とも役立つと考えられる。以上に述べたように, 本論文の研究成果は音声認識の基礎的な技術レベル向上に寄与するところが極めて大きい。この研究成果をもとに, 近年中に実用的な会話音声認識システムが作成できることを確信する。

謝

辞

本研究の開始以来、長期間にわたって御指導賜った日本電信電話公社武蔵野電気通信研究所研究専門調査役・好田正紀博士に心から感謝の意を表します。また、この研究期間に当所基礎研究部第四研究室長及び齊藤特別研究室長として御指導を賜った東京大学教授・齊藤収三博士に心から感謝の意を表します。また、研究の経過や論文のとりまとめにおいて種種の面から御教示を賜った名古屋大学・福村晃夫教授，御鞭撻頂いた名古屋大学・池谷和夫教授，名古屋大学・杉江昇教授に深く感謝いたします。

また、この研究の機会を与えられ、また御鞭撻下さった武蔵野電気通信研究所長・豊田博夫博士，前基礎研究部長・野田健一博士，基礎研究部長・畔柳功芳博士，基礎研究部統括役・橋本新一郎博士，前基礎研究部第四研究室長・小池恒彦氏，基礎研究部第四研究室長・板倉文忠博士に感謝の意を表します。さらに、日頃有益な討論をして下さった横須賀電気通信研究所研究専門調査員・中津良平氏，研究主任・箱田和雄氏，武蔵野電気通信研究所研究専門調査役・古井貞熙博士，研究専門調査員・嵯峨山茂樹氏，杉山雅英氏ならびに基礎研究部第四研究室の諸氏に厚くお礼申し上げます。

文 献

- (1) 新美康永；音声認識，
情報科学講座 E・19・3，共立出版，(1979)
- (2) 中田和男；音声，
日本音響学会編，音響工学講座 7，コロナ社，(1977)
- (3) 中田和男；パターン認識とその応用，
現代自動制御双書 11，コロナ社，(1978)
- (4) 電子通信学会編；聴覚と音声，
電子通信学会，コロナ社，(1980)
- (5) J.L.Flanagan；Speech Analysis, Synthesis and Perception, 2nd Edition,
Springer-Verlag, (1972)
- (6) J.L.Flanagan;Computers that Talk and Listen;Man-Machine Communication
by Voice, Proceeding of the IEEE, Vol.64, No 4, (1976-04)
- (7) D.R.Reddy;Speech Recognition by Machine:A Review, Proceeding of the
IEEE, Vol.64, No 4, (1976-04)
- (8) A.Newell, J.Barnet, J.W.Forgie, C.Green, D.Klatt, J.C.R.Licklider, J.
Munson, D.R.Reddy and W.A.Woods:Speech Understanding Systems-Final
report of a study group, North-Holland, (1973)
- (9) D.H.Klatt:Review of the ARPA Speech Understanding Project, Journal
of the Acoustic Society of America, 62, (1977-06)
- (10) W.A.Lea and J.E.Shoup:Gaps in the Technology of Speech Understanding,
draft of oral presentation at IEEE International Conference on
Acoust., Speech, and Signal Process.,(1978)
- (11) H. Dudley:The Carrier Nature of Speech, Bell Sys. Tech. Journal,
19, (1940)
- (12) R.K.Potter, G.A.Kopp and H.C.Green:Visible Speech, D.Van Nostrand
Co.New York (1947)
- (13) 板倉，斉藤：統計的手法による音声スペクトル密度とホルマント周波数の推定，信学会論

文誌, 53A (1971-01)

- (14) D.H.Klatt and K.N.Stevens: Strategies for Recognition of Spoken Sentences from Visual Examination of Spectrograms, BBN report, No 2154, 10, (1971)
- (15) K.H.Davis, R.Biddulph and S.Balashchek: Automatic Recognition of Spoken Digits, Journal of the Acoustic Society of America, 24, (1952-11)
- (16) G.Fant: Acoustic Theory of Speech Production, Mouton & Co's-Gravenhage (1960)
- (17) K.N.Stevens: Toward a Model for Speech Recognition, Journal of the Acoustic Society of America, 32 (1960)
- (18) J.L.Flanagan: A Difference Limen for Vowel Formant Frequency, Journal of the Acoustic Society of America, 27, (1955)
- (19) H.F.Olson, H.Belar: Phonetic Typewriter, Journal of the Acoustic Society of America, 28, (1956-11)
- (20) 坂井, 堂下; 会話音声識別装置, 信学誌, 46, (1963-11)
- (21) D.R.Reddy; Computer Recognition of Connected Speech, Journal of the Acoustic Society of America, 42, (1967-04)
- (22) C.C.Tappert, N.R.Dixon and A.S.Rabinowitz; Application of Sequential Decoding for Converting Phonetic to Graphic Representation in Automatic Recognition of Continuous Speech (ARCS), IEEE trans. Audio Electroacoust., AU-21 (1973-06)
- (23) R.Alter: Utilization of Contextual Constraints in Automatic Speech Recognition, IEEE trans. Audio Electroacoust., AU-16 (1968-03)
- (24) J.R.Pierce: Whither Speech Recognition, Journal of the Acoustic Society of America, 46, (1969-10)
- (25) 板倉, 齊藤; 偏自己相関係数による音声分析合成系, 音響学会講論集, 2-2-6 (1969-10)
- (26) 迫江, 千葉; 動的計画法を利用した音声の時間軸正規化に基づく連続単語認識, 音響学会誌, 27 (1971-09)
- (27) 好田, 橋本, 齊藤; 数字音声の機械認識系, 信学論(D), 55-D, (1972-03)

- (28) W.A.Woods: Transition Network Grammars for Natural Language Analysis, Commun. ACM, 13, (1970-10)
- (29) T.Winograd: Understanding Natural Language, Academic Press(1972)
- (30) 好田, 中津, 鹿野, 伊藤: 音声によるオンライン質問回答システム, 音響学会誌, 34-3, (1978-03)
- (31) 鹿野, 好田: 会話音声の機械認識における言語処理, 信学論(D), J61-D, 4, (1978-04)
- (32) 中津, 好田: 会話音声の機械認識における音響処理, 信学論(D), J61-D, 4, (1978-04)
- (33) S.Nakagawa: A Machine Understanding System for Spoken Japanese Sentences, 京都大学博士論文(1976-10)
- (34) 新美, 小林, 浅見, 三木: 「SPOKEN BASIC」の認識システム, 情報処理, 5, (1977-05)
- (35) 関口, 重永: 日本語文章の音声認識システム, 音響学会誌, 34-3(1978-03)
- (36) F.Jelinek: Continuous Speech Recognition by Statistical Methods, Proc. IEEE, 64-4(1976-04)
- (37) G.Mercier, P.Quinton and R.Vives: Man-Machine Dialogue with KEAL, Centre National D'etudes des Telecommunications, Vol.4(1977)
- (38) R.De Mori, S.Rivoira and A.Serra: A Speech Understanding System with Learning Capability, Proc. 4th Int. Joint Conf. on AI, P468(1975)
- (39) S.E.Levinson; Maximum Likelihood Parsing of Speech in the Presence of Segmentation Errors, Meeting of the Acoustic Society of America, 67 (S1), S14(A), (1980-04)
- (40) D.R.Reddy, L.D.Erman and R.B.Neely; A Model and a System for Machine Recognition of Speech, IEEE, Trans. Audio Electroacoust., Vol. AU-21(1973-06)
- (41) J.K.Baker; The DRAGON System-An Overview, IEEE, Trans. Acoust, Speech, and Signal Process., Vol ASSP-23, (1975-02)
- (42) B.T.Lowerre: The HARPY Speech Recognition System, Ph.D.thesis, Carnegie-Mellon Univ.(1976)

- (43) V.R.Lesser, R.D.Fennell, L.D.Erman and D.R.Reddy: Organization of the Hearsay II Speech Understanding System, IEEE Trans. Acoust., Speech, and Signal Process., Vol.ASSP-23(1975-02)
- (44) D.R.Reddy et al.: Speech Understanding System-Summary of results of the five-year research effort at Carnegie-Mellon Univ., CMU Tech.Rep. (1977)
- (45) W.A.Woods: Motivation and Overview of SPEECHLIS: An experimental prototype for speech understanding research, IEEE Trans. Acoust., Speech, and Signal Process., Vol.ASSP-23(1975-02)
- (46) W.A.Woods et al.: Speech Understanding Systems-Final technical Progress report, BBN Tech.Rep. 3438(1976)
- (47) J.W.Klovstad and L.F.Mondschein: The CASPERS Linguistic Analysis System, IEEE Trans. Acoust., Speech, and Signal Process., Vol.ASSP-23(1975-02)
- (48) D.E.Walker: The SRI Speech Understanding System, IEEE Symposium on Speech Recognition, P32-P37, (1974-04)
- (49) D.E.Walker, et al.: Speech Understanding Research-Final technical report, SRI Tech.Rep.(1976)
- (50) M.I.Bernstein et al.: Interactive Systems Research-Final report, SDC Tech.Rep.(1976)
- (51) L.R.Bahl, R.Bakis, P.S.Cohen, A.G.Cole, F.Jelinek, B.L.Lewis, and R.L.Mercer: Further Results on the Recognition of a Continuously Read Natural Corpus, IEEE Int.Conf.on Acoust., Speech, and Signal Process., P872(1980-04)
- (52) 鹿野, 好田, 会話音声認識の言語処理系の構成, 信学技報 PRL75-45(1975-10)
- (53) 好田, 中津, 鹿野, 会話音声の認識系の構成, 信学技報 PRL75-55(1975-11)
- (54) 鹿野, 箱田, COSH尺度を用いた会話音声認識システム Voice Q-A System II の音響処理, 信学論(D), 掲載予定(1981-04)
- (55) Y.Kobayashi, Y.Niimi: Word Boundary Detection by Pitch Contours in an Artificial Language, IEEE Int. Conf. on Acoust., Speech, and Signal

Process. P900(1980-04)

- (56) 石川, 浮田, 中川, 坂井; 音声による質問応答システムの構成—会場案内システムの場合—
情報処理学会全国大会, 2H-10, P773, (1980-05)
- (57) 関口, 来嘉, 重永; 日本語文章の音声認識システムの改善, 音響学会講論集, 1-7-21
(1980-05)
- (58) F.Itakura; Minimum Prediction Residual Principle Applied to Speech
Recognition, IEEE, Trans.Acoust., Speech, and Signal Process.,Vol. ASSP-
23(1975-02).
- (59) 齊藤; 最近の音声の研究について, 信学技報, EA75-26(1975-07)
- (60) 鹿野, 好田; 会話音声の機械認識における言語処理, 通研実報, 27, No 6, (1978-06)
- (61) R.G.Goodman; Analysis of Language for Man-Machine Voice Communica-
tion, Ph.D.thesis, Carnegie-Mellon Univ.(1976)
- (62) 中津, 好田; 会話音声の機械認識における音響処理, 通研実報, 27, No 6, (1978-06)
- (63) K.Shikano, M.Kohda; A Linguistic Processor in a Conversational
Speech Recognition System, ECL Review No6 (1978-11)
- (64) 板橋, 城戸; 辞書と音形規則を利用した単語音声の認識, 音響学会誌, 27, No 9, (1971
-09)
- (65) 樽松, 武田, 井上; 書き換え規則を用いて音声認識を行う場合の規則の一構成法, 信学論
(D), 55, No 2, (1972-02)
- (66) 三輪, 新津, 牧野, 城戸; 音声スペクトルの概略形とその動特性を利用した単語音声認識
システム, 音響学会誌, 34, No 3, (1978-03)
- (67) 関口, 大輪, 青木, 重永; フォートランプログラムの音声認識, 情報処理, 18, No 5,
(1977-05)
- (68) 鹿野, 好田; オンライン会話音声認識系における言語処理部の構成, 信学技報, PRL76
-40(1976-10)
- (69) 鹿野, 好田; オンライン会話音声認識系の言語処理部の改良, 信学会情報部門全大, No
223, (1977-10)
- (70) 鹿野, 好田; 会話音声認識の言語処理系における tree search の検討, 信学会全大, No
1212, (1976-03)
- (71) 鹿野, 好田; 会話音声認識における発声の制約とその効果, 信学会全大, No 1240, (1978

- (72) 鹿野清宏：Voice Q-A Systemの言語処理について，情報処理学会人工知能と対話技法研究会 9 - 2 (1979-01)
- (73) 吉田，福村：FORTRANに基づくリスト処理システム (DLOP-2)とそのガーベージ・コレクションについて，情報処理全大，P52(1972)
- (74) R.V.Weeks: Predictive Syllable Mapping in a Continuous Speech Understanding System, IEEE Symp. Speech Recognition, P154(1974)
- (75) B.Raphael: The Thinking Computer (Mind Inside Matter), A Series of Books in Psychology, W.H. Freeman and Company San Francisco, (1976), P75, Optimum Search for Minimum-Cost Path.
- (76) 福村晃夫：情報理論，コロナ社，(1970)
- (77) 阿部圭一：統計的手法によるパターン認識の研究，名古屋大学博士論文(1971)，P2～P29，第1章，パターン認識系の基礎的考察
- (78) 坂井利之：パターン認識の理論，共立出版(1967)，P24～P28，情報量と識別率の関係。
- (79) 三根久：情報理論入門，朝倉書店(1964)，P122～P124，通信路の縦続接続。
- (80) 鹿野清宏；音韻ラティスの評価システム，信学論(A)，J63-A，3，(1980-03)
- (81) 鹿野清宏；音韻ラティスの評価システム，信学技報，PRL78-42(1978-09)
- (82) 鹿野，好田；音韻ラティスの評価システム，音響学会講論集，3-1-17(1977-10)
- (83) 斉藤，松田；音韻識別における付加情報について，音響学会講論集，2-2-17，(1962-05)
- (84) H.Wakita: Residual Energy of Linear Prediction Applied to Vowels and Speaker Recognition, IEEE Trans, Acoust., Speech, and Signal Process., ASSP-24, (1976-06)
- (85) H.F.Silverman and N.R.Dixon; A Comparison of Several-Spectra Classification Methods, IEEE Trans. Acoust., Speech, and Signal Process., ASSP-24, (1976-10)
- (86) A.H.Gray and J.D.Markel; Distance Measures for Speech Processing, IEEE Trans. Acoust., Speech and Signal Process., ASSP-24, (1976-10)
- (87) B.S.Atal; Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification,

- IEEE Trans. Acoust., Speech, and Signal Process., ASSP-23, (1975-02)
- (88) 松田亮一；谷形特性を持つ伝送系の弁別限に対する入力信号の時間的変動の影響，信学誌 49, 10, P1865(1966)
- (89) 東倉，板倉；パーコール帯域圧縮方式における音声品質向上，信学論(A), 61-A, 3, (1978-03)
- (90) 杉山，鹿野；ピークに重みをおいたLPCスペクトラム・マッチング尺度の提案，音響学会講論集，1-7-5, (1980-05)
- (91) 杉山，鹿野；ピークに重みをおいたLPCスペクトル・マッチング尺度の提案，音響学会音声研資料，S80-13, (1980-05)
- (92) 杉山，鹿野；ピークに重みをおいたLPCスペクトラム・マッチング尺度，信学論(A), 掲載予定(1981-05)
- (93) 鹿野，杉山，箱田；ピークに重みをおいたLPCスペクトラム・マッチング尺度による周波数軸の重みづけ，音響学会講論集，1-7-6, (1980-05)
- (94) 杉山，鹿野；Weighted Likelihood Ratioのピーク重み付けの最適化，音響学会講論集，1-1-12, (1980-10)
- (95) 鹿野清宏；Peak weighted measureによる会話音声認識系Voice Q-A System IIの改良，音響学会講論集，1-1-22, (1980-10)
- (96) 箱田，鹿野，杉山；会話音声認識における母音標準パターンの検討，音響学会音声研資料，S79-52(1979-12)
- (97) 鹿野清宏；大語い単語音声認識におけるLPCスペクトル・マッチング尺度の評価，音響学会講論集，1-1-10(1980-10)
- (98) 中津良平；単語音声認識における距離尺度の検討，信学会全国大会，1292(1979-03)
- (99) H. Sakoe and S. Chiba; Dynamic Programming Algorithm Optimization for Spoken Word Recognition, IEEE Trans. Acoust., Speech, and Signal Processing, ASSP-26-1(1978-02)
- (100) L. R. Rabiner, S. E. Levinson, A. E. Rosenberg, and J. G. Wilpon; Speaker-Independent Recognition of Isolated Words Using Clustering Techniques, IEEE Trans. Acoust., Speech, and Signal Process., ASSP-27, No. 4 (1979-08)
- (101) M. Kohda, R. Nakatsu, K. Shikano; Speech Recognition in the Question-Answering System Operated by Conversational Speech, IEEE International Conference on Acoust., Speech, and Signal Process., (1976-04)

- (102) K.Shikano, M.Kohda; A Linguistic Processor for Train Seat Reservation in a Conversational Speech Recognition System, International Joint Conference on Pattern Recognition(1978-10)
- (103) K.Shikano; Acoustic Processing in the Conversational Speech Recognition System, IEEE International Conference on Acoust., Speech, and Signal Process.,(1981-03), to be appeared.
- (104) 鹿野清宏; 会話音声認識における音響処理システムの再帰的構成の検討, 音響学会講論集, 2-2-20(1979-10)
- (105) 鹿野清宏; 会話音声認識システム (Voice Q-A System II) について, 信学技報 PRL 79-61(AL79-77), (1979-12)
- (106) 鹿野, 箱田; 会話音声認識システム (Voice Q-A System II) の音響処理, 信学技報 EA 79-89, (1980-01)
- (107) 鹿野, 箱田; LPCスペクトル・マッチングに基づく会話音声中の母音認識, 信学会部門別全国大会, 51(1979-10)
- (108) 吉田, 福村; FORTRANに基づくデータ・リスト向きリスト処理システム (DLOP-3), 情報処理全国大会, 142(1976)

付 録

1. Voice Q-A System I の言語情報
2. Voice Q-A System I の構文解析部の動作例
3. Voice Q-A System の発声リスト
4. FORTRAN によるリスト処理システム DLOP
5. 641都市名の発声リスト
6. 大語い単語音声認識実験における誤り
7. Voice Q-A System II の音韻変形規則
8. Voice Q-A System II - 1 における発声者別の認識実験結果
9. 音声データの LPC スペクトル包絡の系列およびセグメンテーションに用いる基本特徴量の図示
10. Voice Q-A System II の構文解析部の動作例
11. Voice Q-A System II - 2 における発声者別の認識実験結果

付録 1. Voice Q-A System I の言語情報

図 A 1-1 プラグマティックスのリスト表現

```

PRAGMATICS
( SEM ( YES-NO ) ( * YES-NO ) SEM ( VERB ) SEM ( LATTICE 4 20 ) ( * VERB ) SEM (
SYNTAX 1 ) SEM ( LATTICE 4 24 ) ( * DATE )
SEM ( SYNTAX 2 ) SEM ( LATTICE 6 26 ) ( * STARTING-STATION )
SEM ( SYNTAX 3 ) SEM ( LATTICE 8 26 ) ( * ARRIVING-STATION
) SEM ( SYNTAX 4 ) SEM ( LATTICE 12 64 ) ( * STARTING-TIME ) SEM ( SYNTAX 5 )
SEM ( LATTICE 10 64 ) ( * NAME-OF-TRAIN ) SEM ( SYNTAX 6 ) SEM ( LATTICE 2 18 ) (
* SEAT-CLASS ) SEM ( SYNTAX 7 ) SEM ( LATTICE 4 14 ) ( * NUMBER-OF-TICKETS ) )
    
```

図 A 1-2 構文のリスト表現

(Date , Starting Station , Arriving Station , Starting Time , Train Name ,
Seat Class , Number of Tickets , Verb and Yes-No)

```

o DATE
( ( OR ( * TSUITACHI ) ( * FUTSUKA ) ( * MIKKA ) ( * YOKKA ) ( * ITSUKA ) (
* NUIKA ) ( * NANOKA ) ( * YOKA ) ( * KOKONOKA ) ( * TOKA ) ( * HATSUKA )
( * KYOASU )
( ( OPT ( * 2 ) ) ( * 10 ) ( OR ( ( OR ( * 1 ) ( * 2 ) ( * 3 ) ( * 5 ) ( * 6 )
( * SHICHI ) ( * 7 )
( * 8 ) ( * KU ) ) ( * NICHU ) ( * YOKKA ) ) ) ( * 3 ) ( * 10 ) (
OPT ( * 1 ) ) ( * NICHU ) ) ( * AX ) )
o STARTING-STATION
( ( * EKINEI ) ( OPT ( * EKI ) ) ( OR ( * KARA ) ( * HATSU ) ( * YORI ) ) (
* AX ) )
o ARRIVING-STATION
( ( * EKINEI ) ( OPT ( * EKI ) ) ( OR ( * MADE ) ( * YUKI ) ( * IKI )
SEM ( LATTICE 3 5 ) ( * E ) ) ) ( * AX ) )
o STARTING-TIME
( ( * SUJ16-22 ) ( * JI ) ( OPT ( * PAUSE ) ) ( * SUJ10-59FUN ) ( OPT ( * HATSU
) ) ( * AX ) )
o NAME-OF-TRAIN
160
( ( OR ( ( * HIKARI ) ( OPT ( * PAUSE ) ) ( * SUJ11-199 ) ) ( ( * KODAMA ) ( OPT
( * PAUSE ) ) ( * SUJ200-299 ) ) ) ( * GO ) ( * AX ) )
o SEAT-CLASS
( ( OR ( * SHITEI ) ( * FUTSU ) ( * GREEN ) ) ( OPT ( OR ( * KEN ) ( * SEKI ) ) )
( * AX ) )
o NUMBER-OF-TICKETS
( ( * SUJ11-9 ) ( * MAI ) ( * AX ) )
o YES-NO
( OR ( * HAI ) ( * IIE ) ( ( * SO ) ( * DESS ) ) ( ( * CHICAI ) ( * MASS ) ) )
o VERB
( OR ( ( OR ( ( OR ( * YOYAKU ) ( * ONEGAI ) ) ( OPT ( * ITA ) ) ( *
SHI ) ) ( ( * MOSHIKOMI ) ( OPT ( ( * ITA ) ( * SHI ) ) ) ) ( * MASS )
) ( ( * ARI ) ( OR ( * MASU ) ( * MASEN ) ) ( * KA ) )
( * DESS ) )
EKINEI
( OR ( * TOKYO ) ( * SHINYOKOHAMA ) ( * ODAWARA ) ( * ATAMI ) ( * MISHIMA )
( * SHIZUOKA ) ( * HANAMATSU ) ( * TOYOHASHI ) ( * NAGOYA ) ( * GIFUHASHIMA ) (
* MAIBARA ) ( * KYOTO ) ( * SHINOSAKA ) ( * SHINKOBE ) ( * NISHIAKASHI ) ( *
HIMEJI ) ( * AIOI ) ( * OKAYAMA ) ( * SHINKURASHIKI ) ( * FUKUYAMA ) ( * MIHARA
) ( * HIROSHIMA ) ( * SHINIWAKUNI ) ( * TOKUYAMA ) ( * OGORI ) ( * SHINSHIMONOS
EKI ) ( * KOKURA ) ( * HAKATA ) )
AX
( OPT ( SEM ( LATTICE 1 6 ) OR ( * NO ) ( * O ) ( * DE ) ( * NOO ) ( * NODE ) (
* WA ) ( * GA ) ( * NOWA ) ) )
SUJ16-22
( OR ( * 6 ) ( * 7 ) ( * 8 ) ( * KU ) ( ( * 10 ) ( OPT ( OR ( * SUJ11-9 ) ( * YO
) ) ( * KU ) ) ) ) ( * 2 ) ( * 10 ) ( OPT ( OR ( * 1 ) ( * 2 ) ) ) )
SUJ11-9
( OR ( * 1 ) ( * 2 ) ( * 3 ) ( * 4 ) ( * 5 ) ( * 6 ) ( * 7 ) ( * 8 ) ( * 9 ) )
SUJ10-59FUN
( OR ( ( OR ( * 0 ) ( * RE ) ) ( * FUN ) ) ( OPT ( ( OPT ( * SUJ12-5 ) ) ( * 10 ) )
) ( OR ( ( OR ( * I ) ( * RO ) ( * HA ) ) ( * PPUN ) ) ( OR ( * 2 ) ( * 5 ) (
* 7 ) ( * 8 )
( * 9 ) ) ( * FUN ) ) ( OR ( * 3 ) ( * 4 ) ) ( * PUN ) ) ) ( OPT (
* SUJ12-5 ) ) ( OR ( * ZI ) ( * ZYU ) ) ( * PPUN ) ) )
SUJ12-5
( OR ( * 2 ) ( * 3 ) ( * 4 ) ( * 5 ) )
SUJ11-199
( OR ( * SUJ11-9 ) ( ( OR ( ( OPT ( OR ( OR ( * 2 ) ( * 9 ) ) ( ( * 100 ) (
OPT ( OR ( * 2 ) ( * 5 ) ( * 6 ) ( * 7 ) ( * 8 ) ( * 9 ) ) ) ) ( * 10 ) ) ( *
100 ) ) ( OPT ( * SUJ11-9 ) ) ) ) )
SUJ200-299
( ( * 2 ) ( * 100 ) ( OPT ( ( OPT ( OR ( * 2 ) ( * 3 ) ( * 6 ) ( * 5 ) ( * 8 ) (
* 9 ) ) ) ( * 10 ) ) ) ( OPT ( * SUJ11-9 ) ) ) )
KYOASU
( OR ( * KYO ) ( * HONJITSU ) ( * ASU ) ( * ASHITA ) ( * ASATTE ) )
MASS
( OR ( * MAS ) ( * MASU ) )
DESS
( OR ( * DES ) ( * DESU ) )
NICHU
( OR ( * NITI ) ( * NNCHI ) )
    
```

図A1-3 単語辞書のリスト表現

TSUITACHI (T U I T A T I)	MAIBARA (M A I B A R A)	GREEN (G U R I I N N)
FUTSUKA (H U T U K A)	KYOTO (K Y O O T O)	FUTSU (H U T U U)
MIKKA (M I K K A)	SHINOSAKA (S I N N O O S A K A)	SHITEI (S I T E I)
YOKKA (Y O K K A)	SHINKOBE (S I N N K O O B E)	KEN (K E N N)
ITSUKA (I T U K A)	NISHIAKASHI (N I S I A K A S I)	SEKI (S E K I)
MUIKA (M U I K A)	HIMEJI (H I M E Z I)	MAI (M A I)
NANOKA (N A N O K A)	AIOI (A I O I)	HAI (H A I)
YOKA (Y O O K A)	OKAYAMA (O K A Y A M A)	IIE (I I E)
KOKONOKA (K O K O N O K A)	SHINKURASHIKI (S I N N K U R A S I K I)	YOYAKU (Y O Y A K U)
TOKA (T O O K A)	FUKUYAMA (H U K U Y A M A)	ONEGAI (O N E G A I)
HATSUKA (H A T U K A)	MIHARA (M I H A R A)	ITA (I T A)
1 (I T I)	HIROSHIMA (H I R O S I M A)	SHI (S I)
2 (N I)	SHINIWAKUNI (S I N N I W A K U N I)	MASU (M A S U)
3 (S A N N)	TOKUYAMA (T O K U Y A M A)	ARI (A R I)
4 (Y O N N)	OGORI (O G O O R I)	MASEN (M A S E N N)
5 (G O)	SHINSHIMONOSEKI (S I N N S I M O N O S E K I)	KA (K A)
6 (R O K U)	KOKURA (K O K U R A)	DESU (D E S U)
7 (N A N A)	HAKATA (H A K A T A)	DES (D E)
8 (H A T I)	EKI (E K I)	MAS (H A)
9 (K Y U U)	NITI (N I T I)	PAUSE (**)
10 (Z Y U U)	KARA (K A R A)	NOGHIKOMI (H O O S I K O M I)
YO (Y O)	HATSU (H A T U)	NOWA (N O W A)
SHICHI (H I T I)	YORI (Y O R I)	KYO (K Y O O)
KU (K U)	NO (N O)	HONJITSU (H O N N Z I T U)
I (I)	O (O)	ASU (A S U)
RO (R O)	DE (D E)	ASHITA (A S I T A)
HA (H A)	NOO (N O O)	ASATTE (A S A K K E)
ZI (Z I)	NODE (N O D E)	NNCHI (N N T I)
ZYU (Z Y U)	WA (W A)	SO (S O O)
100 (H Y A K U)	GA (G A)	CHICAI (T I G A I)
0 (R E I)	MADE (M A D E)	
RE (R E E)	YUKI (Y U K I)	
TOKYO (T O O K Y O O)	IKI (I K I)	
SHINYOKOHAMA (S I N N Y O K O H A M A)	E (E)	
ODAWARA (O D A W A R A)	JI (Z I)	
ATAMI (A T A M I)	FUN (H U N N)	
MISHIMA (M I S I M A)	PUN (P U N N)	
SHIZUOKA (S I Z U O K A)	PPUN (P P U N N)	
HANAMATSU (H A M A M A T U)	HIKARI (H I K A R I)	
TOYOHASHI (T O Y O H A S I)	KODAMA (K O D A M A)	
NAGOYA (N A G O Y A)	GO (G O O)	
CIFUHASHIMA (C I H U H A S I M A)		

図A1-4 音韻変形規則のリスト表現 (Voice Q-A System I)

PHONEME REWRITING RULES FOR /A/
 (A - ((E - 1 B 1) (U M D 5 2) (N N H - 4 3) (E Y - 1 3 2)
 (O M - 7 4) (O H - 4 3) (E T T 5 2)
 (U H M 5 3) (O H M 6 3) (O D - 7 3) (O K - 4 3) (E N N 1 0 3) (O W R
 1 0 3) (E H Y K 7 2) (O M - 3 2) (O S K 4 3)) ((A 1 0
 1)) ((U 6 2) (E B 2) (O B 2) (I 6 3) (N N 5 3) (I 4 4))
 - ((R 1 0 2)
 (N 1 0 4) (W 1 0 4) (H 8 3) (K 1 0 4) (N 1 0 4) (Y 1 0 4))))
 PHONEME REWRITING RULES FOR /I/
 (I - ((E A - 9 1) (E - - 2 1)
 (E Z - 4 2) (U Z - 4 2) (E E I 4 2) (E
 - - 3 3) (N N N - 1 0 2) (N N S - 6 3) (U R - 6
 3) (U T - 7 2) (N N T - 5 3) (E S A 6 3) (U A - 7 3))
 (U S - 5 2) (N N - - 6 6) (E O - 7 4) (U N - 5 3))
 ((I 7 1) (N N 1 0 1) (U 1 2 2) (O 1 0 3) (T 7 3) (I 5 3)) ((E 1 0 2
) (N N 6 2)) - ((Z 8 1) (N N 9 3) (N 8 3))))
 PHONEME REWRITING RULES FOR /U/
 (U - ((N N Y - 6 2) (I - - 2 2) (N N K - 6 2) (I K - 3 2)
 (E - - 4 3) (N N - - 9 3
) (O - - 2 2) (O K - 8 2) (O H - 9 2) (O P - 1 1 2) (O P P -
 1 1 2) (O G - 7 2) (O M - 5 3) (O S - 9 4) (O T - 7 3) (E T
 - 6 4) (E K - 6 4))
 ((U 8 1) (N N 8 2) (I 8 3) (I 5 6))
 ((I 8 1) (N N 8 2)) - -))
 PHONEME REWRITING RULES FOR /E/
 (E ((N 6 2))
 ((I - - 2 2) (U - - 3 3) (N N - - 4 3
) (N N D - 1 2 2) (N N S - 7 3) (I I K 2 0 3) (A D - 6
 3) (U D - 1 2 2) (I A K 4 4) (I S K 4 4) (I M - 7 3)))
 ((E 8 1) (I 1 0 3)) ((U 6 2) (N N 6 2) (I 4 2)) - -))
 PHONEME REWRITING RULES FOR /O/
 (O -
 ((N N Y - 6 2) (A Y - 5 3) (E Y - 4 1) (U Y -
 1 2 2) (U K D 5 3) (U T - 8 3) (N N - - 3 2) (N N N - 5 3) (U
 - K 6 4) (U N - 1 0 3) (A R S 5 3) (E T - 5 4) (A N - 4 3)) ((O
 8 1)) ((N N 7 1) (U 6 2) (I 6 5)) - ((N 1 0 6))))
 PHONEME REWRITING RULES FOR /N/
 (N N ((M 6 1) (M 3 2) (N 3 2) (* 4 3) (G 3 2)))
 ((U I - 5 3) (I I - 9 3)
 (I I I - 1 0 3) (N I I 4 3)) - ((I 1 0 4)) ((M U 1) (N U 2))
 (M O 2) (O N N 3) (O U 3) (N N I 3) (N I 4) (H Y U 3) (N O 2))
 (Z Y U 4) (M I 4)) -))
 PHONEME REWRITING RULES FOR /II/
 (I I ((I 7 1) (I 4 3)))
 - - ((I N 3) (E I 3)) -))
 PHONEME REWRITING RULES FOR /UU/
 (U U ((U 6 0) (U 4 1) (N N 6 1) (N N 4 3) (M 6 1) (O 7 3)))
 ((O Z Y - 1 5 1) (I Z Y - 1 0 2) (O K Y - 1 0 2) (I K Y - 1 5 2) (E Z Y
 - 8 2)))
 - - ((U O 2) (U Z 1)) -))
 PHONEME REWRITING RULES FOR /EE/
 (E E ((E 6 0) (E 4 1)) - - - ((E H 3)) -))
 PHONEME REWRITING RULES FOR /OO/
 (O O ((O 8 0) (O 6 3)) - - - ((O M 2)) -))
 PHONEME REWRITING RULES FOR /K/
 (K ((G - 1 3) (** - 4 3) (C A 1 3)))
 ((H - 9 1) (S - 8 2) (P - 8 1) (T - 1 0 1) (K K - 1 2 3))
 (K Y E 8 2)) - ((U Y P I Y 2)) ((* 2)) -))
 PHONEME REWRITING RULES FOR /S/
 (S ((T - 2 1) (K - 2 1) (H Y I 2 1) (H Y - 2 2)
 (B I 3 2) (Z - 4 3) (Z E 3 3) (C I 2 3) (** - 4 3)))
 ((P - 8 1) (K Y - 8 2) (H - 8 2)))
 - - ((* 1)) -))
 PHONEME REWRITING RULES FOR /T/
 (T ((S - 4 1) (K - 2 1) (H - 4 2) (D - 3 2) (S - 1 2)))
 (** - 4 2) (H E 1 2) (G I 5 2)) ((P - 8 1) (H Y O 3 1))
 (H Y U 3 3)) - - ((* 1)) -))
 PHONEME REWRITING RULES FOR /H/
 (H ((* U 4 1) (T I 3 1) (S - 3 2) (W A 2 2) (B - 1 2) (M A 2 2
) (S A 1 3) (N A 1 3) (C - 1 2) (M U 1 2) (Z A 5 3)))
 ((K - 8 1) (P - 8 1) (N N A 7 3) (N U 6 2) (H Y U 5 3)) ((O A 1)
 (A A 2) (- U 2) (- A 3)))
 ((I R H Y U N 3) (I M H Y U M 3) (U N N M O N 4)))
 ((* 1)) -))
 PHONEME REWRITING RULES FOR /P/
 (P ((T - 5 1) (K - 5 1) (H - 5 1) (S - 5 2) (B - 5 2)))
 (* - 8 1)) ((P P - 1 0 1)))
 - - ((* 1)) -))
 PHONEME REWRITING RULES FOR /N/
 (N ((M - 1 1) (D A 1 2) (G I 1 3) (Z I 1 3) (R O 2 3)))
 (G O 2 3) (* I 1 2)))
 ((T U 5 2) (U O 6 3) (N N O 1 1 3) (Y A 3 3) (N N I 9
 2) (H A 3 3)) ((N N - 1) (U I 2) (- A 3)))
 (U U I 3) (I I 1)) ((E C * N N G 3) (I T * N N T 3)) ((* 3)) -))

図 A 1-4 音韻変形規則のリスト表現 (Voice Q-A System I) (続き)

PHONEME REWRITING RULES FOR /M/
 (M ((N - 1 1) (R A 3 3))
 ((G - 5 2) (B A 5 3) (N N A 7 3) (H A 4 3) (N N U 7 3) (U A 4 3)
 (Z I 4 3))
 ((N N A 5) (I A 3) (A A 3)) - - -)
 PHONEME REWRITING RULES FOR /Y/
 (Y ((W O 3 2) (N A 4 3) (R - 1 3) (N O 1 4)
 (E A 3 4) (I O 7 3) (I U 7 3))
 ((Z Y 0 1 1 1) (I A 6 3) (Z Y 0 0 1 1 1) (K Y 0 6 3) (K Y 0 0 1 0 3)
 (Z Y A 4 3) (N 0 0 1 4) (D O 4 3) (H A 4 3))
 ((O O 3)) - -
 ((R H * B H Y) (E N N I) (H
 N R M) 4 2))
 PHONEME REWRITING RULES FOR /R/
 (R - ((B - 4 1) (D A 4 1) (D - 3 2) (N - 6 2) (Y
 - 4 2) (Z - 3 2) (N N I 5 3) (M - 5 3) (G - 4 3) (N I I 6 4)))
 ((U A 2) (- I 2) (- A 3) (I - 3) (U - 3)
 (- O 3)) - - -)
 PHONEME REWRITING RULES FOR /W/
 (W ((Z A 2 3) (N A 1 3))
 ((U A 4 2) (H A 6 2) (M A 7 3) (O A 6 3))
 ((N N A 2) (A - 2)) - - -)
 PHONEME REWRITING RULES FOR /G/
 (G ((R - 1 1) (D - 1 2) (Z - 1 2) (N - 1 2) (M - 2 2)
 (B - 4 2) (K - 4 2) (K 0 0 6 3) (H 0 0 5 3)
 (N N 0 0 1 0 3) (H 0 5 3) (N N - 7 3) (U 0 0 6 2))
 ((- O 1) (U U - 2) (U - 3) (E A 3) (O - 3)
 (N N - 3))
 ((U R * N N R 3) (U R H Y U R 3))
 ((* 3)) -)
 PHONEME REWRITING RULES FOR /Z/
 (Z - ((G I 6 1) (R - 6 1) (B - 6 2) (N N I 8 3) (N I 8 3)
 (D - 6 2) (Y - 6 2) (B I 6 2) (H U 5 3) (E I 3 3) (Z Y - 3 3))
 ((- I 2) (N N - 2) (I U 3)) - - -)
 PHONEME REWRITING RULES FOR /D/
 (D ((R - 2 1) (* * E 2 1) (Z A 2 2) (* E 5 2))
 ((B - 4 2) (G - 4 2) (T - 5 2) (N N E 6 2))
 (U E 5 2) (H E 7 3) (U A 4 3) (O A 3)) - ((* 5)) -)
 PHONEME REWRITING RULES FOR /B/
 (B ((R - 2 1) (D - 2 1) (G - 1 2) (Z I 2 2) (N
 - 2 2) (M - 2 3)) ((Y - 6 2) (P - 6 3) (K E 1 3)) - - - -
)
 PHONEME REWRITING RULES FOR /KK/
 (K K ((K - 7 1) (T E 8 1))
 ((P P - 1 5 1)) - - - -)
 PHONEME REWRITING RULES FOR /PP/
 (P P ((K K - 1 0 1) (P - 7 2) (K - 8 3))
 - - - -)
 PHONEME REWRITING RULES FOR /KY/
 (K Y ((H Y - 6 1) (R 0 0 6 3) (Y 0 0 6 3) (Z Y - 6 3))
 - - - ((* 3)) ((T K S K K) (I E) (* G M Y R) 7 1)
)
 PHONEME REWRITING RULES FOR /ZY/
 (Z Y ((I U U 8 3)
 (Y U U 3 1) (Y U 3 3) (H Y U U 4 4)) - ((N N U 3))
 - - ((B N M Z G *) (I) (Z Y R Z Y E) 1 2 2))
 PHONEME REWRITING RULES FOR /HY/
 (H Y ((K Y - 1 1) (D A 5 2) (H A 5 1) (Z Y A 1 2) (Y A 3 3) (H A 1 2
))
 ((N A 4 2) (E A 4 2)) - ((A K N A K 2)
 (A K N A G 2) (A K N A P 2)) ((* 5)) ((H S) (I E
) (*) 4 2))
 PHONEME REWRITING RULES FOR /**/
 (* * ((T - 6 3) (P P - 1 2 1) (K K - 1 2 1) (K - 6 3))
 - - - -)
 PHONEME REWRITING RULES FOR INSERTION ERROR CORRECTION
 ((A (A - 4 1) (- A 4 1) (- - 2 3) (U R 4 3) (N O 3 3) (E N N 3
 2)) (I (T E 3 2) (I - 4 2) (I - 8 4) (- I 4 1) (I I - 1 2 0
) (U U - 5 2) (K E 4 2) (E - 3 3) (- Y 6 3) (- Z Y 6 1) (- - 2
 3) (U M 3 3) (E Z 4 3))
 (U (U - 4 1) (- U 4 1) (U U - 6 1) (O O - 6 1) (I H 5 3) (O
 Z Y 4 3) (I N 5 3) (I D 5 3) (- - 2 3) (O M 4 3) (* M 4 3)
 (S E 4 3)) (E (E -
 4 1) (- E 4 1) (E E - 4 1) (A I 3 3) (I N N 4 3) (- - 2 3)
 (I P P 3 3) (* A 3 3) (A S 3 3) (H Y A 3 1) (A T 3 3) (A Z 3 3
) (U U - 4 3) (K A 4 3)) (O (O - 4 1) (- O 4 1) (O O - 9
 1) (A G 4 3) (U - 3 3) (E N N 4 3) (- - 1 2) (U R 4 3
) (U U - 6 1) (M A 3 3)) (N N (- - 3 2) (* - 1 0 3) (I
 - 4 3) (- Z 4 3))
 (H (N N 0 4 1) (- - 1 2) (I E 2 3) (E - 3 3))
 (M (- N N 4 2) (- - 2 1) (U U - 3 3))
 (R (- I 2 2) (- - 3 3))
 (G (- O 3 2) (- A 3 2) (- - 2 3)) (Z (- O 4 2)
 (* - 6 2) (Z - 2 2)
) (D (- - 1 2) (* - 9 3))
 (B (- O 4 1) (* - 9 1) (- I 4 2)) (N
 (* - 1 0 1) (- - 4 3))
 (Y (* - 3 3)) (Z Y (* - 3 3) (- - 1 3)))
 PHONEME REWRITING RULES FOR CORRECTION OF DEVOCALIZED VOWELS
 ((K S T H P) (I U) (K S T H P -))
 LIST OF VOICED PHONEMES
 ((N M Y R W G Z D B Z Y * A I U E O N N) 6 2)

図 A 1 - 5 時刻表

SHINKANSEN JIKOKUHYO KUDARI

	HIKARI 19 GO	KODAMA 205 GO	KODAMA 213 GO	HIKARI 91 GO	KODAMA 217 GO	HIKARI 17 GO	KODAMA 221 GO	HIKARI 21 GO	KODAMA 201 GO	HIKARI 101 GO	HIKARI 151 GO	KODAMA 203 GO	HIKARI 121 GO	KODAMA 207 GO	HIKARI 1 GO
TOKYO								6:00	6:04	6:12	6:24	6:28	6:48	6:52	7:00
SHINYOKOHAMA									6:22			6:46		7:10	
ODAWARA									6:46			7:10		7:34	
ATAMI									6:57			7:21		7:45	
NISHIMA									7:11			7:35		7:59	
SHIZUOKA									7:39			8:04		8:28	
IHAMAMATSU									8:07			8:31		8:55	
TOYOHASHI									8:27			8:51		9:15	
NAGOYA						7:03		8:03	8:56	8:15	8:27	9:20	8:51	9:44	9:03
GFUBHASHIMA									9:15			9:39		10:03	
MAIBARA									9:33			9:57		10:21	
KYOTO									10:00	9:05	9:17	10:24	9:41	10:48	9:53
SHINOSAKA	6:02	6:06	6:49	7:02	7:41	8:12	8:36	9:12	10:18	9:24	9:37	10:42	10:00	11:06	10:12
SHINKOBE		6:22	7:05		7:57		8:52					9:53			
NISHIKASHI		6:35	7:17		8:09		9:04					10:05			
HIMEJI		6:50	7:35		8:24		9:19			10:04	10:20				
AIOI		7:02	7:47		8:35		9:30				10:31				
OKAYAMA	7:02	7:27	8:12	8:02	8:39	9:12	9:54	10:12		10:35	10:53		11:00		11:12
SHINKURASHIKI		7:41	8:26		9:13		10:09			10:49			11:15		
FUKUYAMA		7:56	8:41		9:28		10:26			11:04			11:36		
MIHARA		8:11	8:56		9:46		10:46			11:19			11:50		
HIROSHIMA	8:00	8:39	9:24	9:00	10:14	10:10	11:14	11:10		11:46			12:19		12:10
SHINJUKUNI		8:58	9:43		10:33		11:33			12:05			12:38		
TOKUYAMA	8:33	9:13	10:03	9:33	10:53		11:53			12:24			12:56		
OGORI		9:38	10:23		11:13		12:13	11:58		12:44			13:18		
SHIRSHIMONOSEKI	9:11	10:01	10:46	10:11	11:36		12:37			13:07			13:42		
KOKURA	9:24	10:14	10:57	10:24	11:49	11:24	12:51	12:29		13:19			13:56		13:24
HAKATA	9:55	10:46	11:31	10:56	12:21	11:56	13:20	13:01		13:51			14:30		13:56

	KODAMA 209 GO	HIKARI 153 GO	HIKARI 105 GO	KODAMA 211 GO	HIKARI 191 GO	HIKARI 3 GO	KODAMA 215 GO	HIKARI 155 GO	KODAMA 219 GO	HIKARI 107 GO	HIKARI 23 GO	KODAMA 223 GO	HIKARI 157 GO	KODAMA 225 GO	HIKARI 109 GO
TOKYO	7:16	7:24	7:36	7:40	7:48	8:00	8:16	8:24	8:40	8:48	9:00	9:16	9:24	9:40	9:48
SHINYOKOHAMA	7:34			7:58			8:34		8:58			9:34		10:22	
ODAWARA	7:53			8:22			8:58		9:22			9:58		10:33	
ATAMI	8:09			8:33			9:09		9:33			10:09		10:33	
NISHIMA	8:23			8:47			9:23		9:47			10:23		10:47	
SHIZUOKA	8:52			9:16			9:52		10:16			10:52		11:16	
IHAMAMATSU	9:19			9:43			10:19		10:43			11:19		11:43	
TOYOHASHI	9:39			10:03			10:59		11:03			11:39		12:03	
NAGOYA	10:08	9:27	9:39	10:32	9:51	10:03	11:08	10:27	11:32	10:51	11:03	12:08	11:27	12:32	11:51
GFUBHASHIMA	10:27			10:51			11:27		11:51			12:27		12:51	
MAIBARA	10:46			11:09	10:17		11:45		12:09			12:45		13:09	
KYOTO	11:12	10:17	10:29	11:36	10:44	10:53	12:12	11:17	12:36	11:41	11:53	13:12	12:17	13:36	12:41
SHINOSAKA	11:30	10:37	10:48	11:54	11:02	11:12	12:30	11:37	12:54	12:00	12:12	13:30	12:37	13:54	13:00
SHINKOBE		10:53	11:04					11:53		12:17			12:53		13:17
NISHIKASHI		11:03						12:03					13:03		
HIMEJI		11:20	11:20					12:20		12:45			13:20		13:45
AIOI		11:31						12:31					13:31		
OKAYAMA		11:53	11:59			12:12		12:53		13:17	13:12		13:53		14:17
SHINKURASHIKI			12:13							13:31					14:31
FUKUYAMA			12:28							13:46					14:46
MIHARA			12:46							14:01					15:01
HIROSHIMA			13:16			13:10				14:29	14:10				15:29
SHINJUKUNI			13:35							14:48					15:48
TOKUYAMA			13:55							15:08					16:08
OGORI			14:10							15:20	14:58				16:20
SHIRSHIMONOSEKI			14:30							15:51					16:51
KOKURA			14:51			14:24				16:04	15:29				17:04
HAKATA			15:25			14:56				16:36	16:01				17:36

	HIKARI 5 GO	HIKARI 127 GO	KODAMA 229 GO	HIKARI 159 GO	KODAMA 233 GO	HIKARI 111 GO	HIKARI 7 GO	KODAMA 237 GO	HIKARI 161 GO	KODAMA 239 GO	HIKARI 113 GO	HIKARI 9 GO	KODAMA 231 GO	HIKARI 163 GO	KODAMA 257 GO
TOKYO	10:00	10:12	10:16	10:24	10:40	10:48	11:00	11:16	11:24	11:40	11:48	12:00	12:16	12:24	12:40
SHINYOKOHAMA			10:34		10:58			11:34		11:58			12:34		12:58
ODAWARA			10:58		11:22			11:58		12:22			12:58		13:22
ATAMI			11:00		11:33			12:09		12:33			13:09		13:33
NISHIMA			11:23		11:47			12:23		12:47			13:23		13:47
SHIZUOKA			11:52		12:16			12:52		13:16			13:52		14:16
IHAMAMATSU			12:19		12:43			13:19		13:43			14:19		14:43
TOYOHASHI			12:39		13:03			13:39		14:03			14:39		15:03
NAGOYA	12:03	12:15	13:06	12:27	13:32	12:51	13:03	14:06	13:27	14:32	13:51	14:03	15:06	14:27	15:32
GFUBHASHIMA			13:27		13:51			14:27		14:51			15:27		15:51
MAIBARA			13:45		14:09			14:45		15:09			15:45		16:09
KYOTO	12:53	13:05	14:12	13:17	14:36	13:41	13:53	15:12	14:17	15:36	14:41	14:53	16:12	15:17	16:36
SHINOSAKA	13:12	13:24	14:30	13:37	14:54	14:00	14:12	15:30	14:37	15:54	15:00	15:12	16:30	15:37	16:54
SHINKOBE		13:40		13:53		14:17			14:53		15:17			15:53	
NISHIKASHI				14:05					15:05					16:05	
HIMEJI		14:04		14:20		14:45			15:20		15:45			16:20	
AIOI				14:31					15:31					16:31	
OKAYAMA	14:12	14:33		14:53		15:17	15:12		15:53		16:17	16:12		16:53	
SHINKURASHIKI						15:31					16:31				
FUKUYAMA						15:46					16:46				
MIHARA						16:01					17:01				
HIROSHIMA	15:10					16:29	16:10				17:29	17:10			
SHINJUKUNI						16:48					17:48				
TOKUYAMA						17:08					18:08				
OGORI						17:28					18:29				
SHIRSHIMONOSEKI						17:51					18:51				
KOKURA	16:24					18:04	17:24				19:04	18:24			
HAKATA	16:56					18:36	17:56				19:36	18:56			

図 A 1 - 5 時刻表 (続き)

	HIKARI 115 GO	HIKARI 11 GO	KODAMA 251 GO	HIKARI 165 GO	KODAMA 253 GO	HIKARI 123 GO	HIKARI 25 GO	HIKARI 181 GO	KODAMA 259 GO	HIKARI 167 GO	KODAMA 263 GO	HIKARI 117 GO	HIKARI 13 GO	HIKARI 97 GO	KODAMA 267 GO
TOKYO	12:48	13:00	13:16	13:24	13:40	13:48	14:00	14:12	14:16	14:24	14:40	14:48	15:00	15:12	15:16
SHINYOKOHAMA	-	-	13:34	-	13:58	-	-	-	14:34	-	14:58	-	-	-	15:34
ODAWARA	-	-	13:50	-	14:22	-	-	-	14:58	-	15:22	-	-	-	15:58
ATAMI	-	-	14:09	-	14:33	-	-	-	15:09	-	15:33	-	-	-	16:09
NISHINA	-	-	14:23	-	14:47	-	-	-	15:23	-	15:47	-	-	-	16:23
SHIZUOKA	-	-	14:52	-	15:16	-	-	-	15:52	-	16:16	-	-	-	16:52
HANAMATSU	-	-	15:19	-	15:43	-	-	-	16:19	-	16:43	-	-	-	17:19
TOYOHASHI	-	-	15:39	-	16:03	-	-	-	16:39	-	17:03	-	-	-	17:39
NAGOYA	14:51	15:03	16:08	15:27	16:32	15:51	16:03	16:10	17:08	16:27	17:32	16:51	17:03	17:15	18:08
CI FUDASHIMA	-	-	16:27	-	16:51	-	-	-	17:27	-	17:51	-	-	-	18:27
MAIBARA	-	-	16:45	-	17:09	-	-	-	17:45	-	18:09	-	-	-	18:45
KYOTO	15:41	15:53	17:12	16:17	17:36	16:41	16:53	17:05	18:12	17:17	18:36	17:41	17:53	18:05	19:12
SHIOSAKA	16:00	16:12	17:30	16:37	17:54	17:00	17:12	17:24	18:30	17:37	18:54	18:00	18:12	18:22	19:30
SHIKODE	16:17	-	-	16:53	-	-	-	17:40	-	17:53	-	18:17	-	-	-
NISHIAKASHI	-	-	-	17:05	-	-	-	-	-	18:05	-	-	-	-	-
HIMEJI	16:45	-	-	17:20	-	-	-	18:04	-	18:20	-	18:45	-	-	-
AIOI	-	-	-	17:31	-	-	-	-	-	18:31	-	-	-	-	-
OKAYAMA	17:17	17:12	-	17:53	-	18:00	18:12	18:33	-	18:53	-	19:17	19:12	-	-
SHINJURASHIKI	17:31	-	-	-	-	18:15	-	-	-	-	-	19:31	-	-	-
FUKUYAMA	17:46	-	-	-	-	18:36	-	-	-	-	-	19:46	-	-	-
NIHARA	18:01	-	-	-	-	18:30	-	-	-	-	-	20:01	-	-	-
HIROSHIMA	18:29	18:10	-	-	-	19:19	19:10	-	-	-	-	20:29	20:10	-	-
SHINJAKUNI	18:48	-	-	-	-	19:38	-	-	-	-	-	20:48	-	-	-
TOKUYAMA	19:08	-	-	-	-	19:58	-	-	-	-	-	21:08	-	-	-
OGORI	19:28	-	-	-	-	20:18	19:58	-	-	-	-	21:28	-	-	-
SHINSHIMONOSEKI	19:51	-	-	-	-	20:42	-	-	-	-	-	21:51	-	-	-
KOKURA	20:04	19:24	-	-	-	20:56	20:29	-	-	-	-	22:04	21:24	-	-
HAKATA	20:36	19:56	-	-	-	21:30	21:01	-	-	-	-	22:36	21:56	-	-

	HIKARI 169 GO	KODAMA 269 GO	HIKARI 119 GO	HIKARI 27 GO	HIKARI 171 GO	KODAMA 235 GO	HIKARI 173 GO	KODAMA 261 GO	HIKARI 175 GO	KODAMA 265 GO	HIKARI 15 GO	HIKARI 193 GO	KODAMA 255 GO	HIKARI 177 GO	KODAMA 295 GO
TOKYO	15:24	15:40	15:48	16:00	16:12	16:16	16:24	16:40	16:48	16:52	17:00	17:12	17:16	17:24	17:40
SHINYOKOHAMA	-	15:58	-	-	-	16:34	-	16:58	-	17:10	-	-	17:34	-	17:58
ODAWARA	-	16:22	-	-	-	16:58	-	17:22	-	17:34	-	-	17:58	-	18:22
ATAMI	-	16:33	-	-	-	17:09	-	17:33	-	17:45	-	-	18:09	-	18:33
NISHINA	-	16:47	-	-	-	17:23	-	17:47	-	17:59	-	-	18:23	-	18:47
SHIZUOKA	-	17:16	-	-	-	17:52	-	18:16	-	18:28	-	-	18:52	-	19:16
HANAMATSU	-	17:43	-	-	-	18:19	-	18:43	-	18:55	-	-	19:19	-	19:43
TOYOHASHI	-	18:03	-	-	-	18:59	-	19:03	-	19:15	-	-	19:39	-	20:03
NAGOYA	17:27	18:32	17:51	18:03	18:15	19:08	18:27	19:32	18:51	19:44	19:03	19:13	20:08	19:27	20:32
CI FUDASHIMA	-	18:51	-	-	-	19:27	-	19:51	-	20:03	-	-	20:27	-	20:51
MAIBARA	-	19:09	-	-	-	19:45	-	20:09	-	20:21	-	19:41	20:45	-	21:09
KYOTO	18:17	19:36	18:41	18:53	19:05	20:12	19:17	20:36	19:41	20:48	19:53	20:08	21:12	20:17	21:36
SHIOSAKA	19:00	19:54	19:12	19:25	19:25	20:30	19:37	20:54	20:01	21:06	20:12	20:26	21:30	20:37	21:54
SHIKODE	18:53	-	19:17	-	19:41	-	19:53	-	20:10	-	20:05	-	20:35	-	20:53
NISHIAKASHI	19:05	-	-	-	19:53	-	20:05	-	20:35	-	20:30	-	21:00	-	21:30
HIMEJI	19:20	-	19:45	-	20:08	-	20:20	-	20:49	-	21:00	-	21:20	-	21:50
AIOI	19:31	-	-	-	20:19	-	20:31	-	21:00	-	-	-	21:31	-	21:51
OKAYAMA	19:53	-	20:17	20:12	20:43	-	20:53	-	21:24	-	21:12	-	21:53	-	-
SHINJURASHIKI	-	-	20:31	-	20:57	-	-	-	21:38	-	-	-	-	-	-
FUKUYAMA	-	-	20:46	-	21:12	-	-	-	21:53	-	-	-	-	-	-
NIHARA	-	-	21:01	-	21:27	-	-	-	22:08	-	-	-	-	-	-
HIROSHIMA	-	-	21:29	21:10	21:53	-	-	-	22:34	-	22:10	-	-	-	-
SHINJAKUNI	-	-	21:48	-	-	-	-	-	-	-	-	-	-	-	-
TOKUYAMA	-	-	22:08	-	-	-	-	-	-	-	-	-	-	-	-
OGORI	-	-	22:28	21:58	-	-	-	-	-	-	-	-	-	-	-
SHINSHIMONOSEKI	-	-	22:51	-	-	-	-	-	-	-	-	-	-	-	-
KOKURA	-	-	23:04	22:29	-	-	-	-	-	-	23:24	-	-	-	-
HAKATA	-	-	23:36	23:01	-	-	-	-	-	-	23:56	-	-	-	-

	HIKARI 179 GO	KODAMA 291 GO	HIKARI 125 GO	HIKARI 195 GO	KODAMA 293 GO	HIKARI 183 GO	HIKARI 183 GO	KODAMA 297 GO	HIKARI 185 GO	HIKARI 29 GO	KODAMA 299 GO	HIKARI 93 GO	HIKARI 95 GO	HIKARI 99 GO
TOKYO	17:48	17:52	18:00	18:12	18:16	18:24	18:48	18:52	19:00	19:24	19:28	19:48	20:00	20:24
SHINYOKOHAMA	-	18:10	-	-	18:34	-	-	19:10	-	-	19:46	-	-	-
ODAWARA	-	18:34	-	-	18:58	-	-	19:34	-	-	20:10	-	-	-
ATAMI	-	18:45	-	-	19:09	-	-	19:45	-	-	20:21	-	-	-
NISHINA	-	18:59	-	-	19:23	-	-	19:59	-	-	20:35	-	-	-
SHIZUOKA	-	19:28	-	-	19:52	-	-	20:28	-	-	21:04	-	-	-
HANAMATSU	-	19:55	-	-	20:19	-	-	20:55	-	-	21:31	-	-	-
TOYOHASHI	-	20:15	-	-	20:39	-	-	21:15	-	-	21:51	-	-	-
NAGOYA	19:51	20:44	20:03	20:15	21:08	20:27	20:51	21:44	21:03	21:27	22:20	21:51	22:03	22:27
CI FUDASHIMA	-	21:03	-	-	21:27	-	-	22:03	-	-	22:39	-	-	-
MAIBARA	-	21:21	-	20:41	21:45	-	-	22:21	-	-	22:57	-	-	-
KYOTO	20:41	21:48	20:53	21:05	22:12	21:17	21:41	22:48	21:53	22:17	23:24	22:41	22:53	23:17
SHIOSAKA	21:01	22:06	21:13	21:26	22:30	21:37	21:58	23:06	22:13	22:34	23:42	22:58	23:10	23:34
SHIKODE	21:17	-	21:29	-	21:53	-	22:05	-	22:29	-	-	-	-	-
NISHIAKASHI	21:29	-	21:41	-	22:05	-	22:17	-	22:41	-	-	-	-	-
HIMEJI	21:44	-	21:56	-	22:20	-	22:31	-	22:56	-	-	-	-	-
AIOI	21:58	-	22:07	-	22:31	-	22:41	-	23:07	-	-	-	-	-
OKAYAMA	22:18	-	22:31	-	22:53	-	-	-	23:29	-	-	-	-	-
SHINJURASHIKI	-	-	22:45	-	-	-	-	-	-	-	-	-	-	-
FUKUYAMA	-	-	23:00	-	-	-	-	-	-	-	-	-	-	-
NIHARA	-	-	23:15	-	-	-	-	-	-	-	-	-	-	-
HIROSHIMA	-	-	23:41	-	-	-	-	-	-	-	-	-	-	-
SHINJAKUNI	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TOKUYAMA	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OGORI	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SHINSHIMONOSEKI	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KOKURA	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HAKATA	-	-	-	-	-	-	-	-	-	-	-	-	-	-

図 A 1 - 5 時刻表 (続き)

SHINKANSEN JIKOKUHYO NODORI PAGE

	KODAMA 228 GO	KODAMA 204 GO	HIKARI 94 GO	KODAMA 212 GO	KODAMA 200 GO	HIKARI 98 GO	KODAMA 202 GO	HIKARI 102 GO	KODAMA 206 GO	HIKARI 196 GO	KODAMA 208 GO	HIKARI 180 GO	KODAMA 210 GO	HIKARI 152 GO	KODAMA 214 GO
HAKATA															
KOKURA															
SHINSHIMONOSEKI															
OGORI															
TOKUYAMA															
SHINJIAKASHI															
HIROSHIMA															
MIHARA															
FUKUYAMA															
SHINKURASHIKI															
OKAYAMA															
AIOI												6:03			6:27
HIMEJI												6:28			6:49
NISHIAKASHI												6:36			7:00
SHINKOBE												6:51			7:15
SHINOSAKA			6:00		6:14	6:22	6:38	6:46	7:02	7:10	7:14	7:22	7:38	7:46	8:02
KYOTO			6:19		6:33	6:41	6:57	7:05	7:21	7:23	7:33	7:41	7:57	8:05	8:21
MAIBARA			-		6:59	-	7:23	-	7:47	-	7:59	-	8:23	-	8:47
GIFUHASHIMA			-		7:20	-	7:44	-	8:08	-	8:20	-	8:44	-	9:09
NAGOYA	6:20		7:09	7:17	7:36	7:30	8:00	7:55	8:24	8:19	8:36	8:31	9:00	8:55	9:24
TOYOHASHI	6:47		-	7:43	8:07	-	8:31	-	8:55	-	9:07	-	9:31	-	9:55
HAMAMATSU	7:04		-	8:00	8:24	-	8:48	-	9:12	-	9:24	-	9:48	-	10:12
SHIZUOKA	7:33	8:00	-	8:30	8:54	-	9:18	-	9:42	-	9:54	-	10:18	-	10:42
MISHIMA	7:57	8:25	-	8:59	9:23	-	9:47	-	10:11	-	10:23	-	10:47	-	11:11
ATAMI	8:07	8:37	-	9:10	9:34	-	9:58	-	10:22	-	10:34	-	10:58	-	11:22
ODAWARA	8:18	8:48	-	9:24	9:48	-	10:12	-	10:36	-	10:48	-	11:12	-	11:36
SHINYOKOHAMA	8:39	9:08	-	9:44	10:08	-	10:32	-	10:56	-	11:08	-	11:32	-	11:56
TOKYO	9:00	9:28	9:10	10:04	10:28	9:32	10:52	9:56	11:16	10:20	11:28	10:32	11:52	10:56	12:16

HIKARI HIKARI HIKARI KODAMA HIKARI KODAMA HIKARI HIKARI HIKARI KODAMA HIKARI KODAMA HIKARI HIKARI KODAMA
190 GO 156 GO 158 GO 220 GO 160 GO 224 GO 162 GO 164 GO 2 GO 230 GO 192 GO 234 GO 168 GO 170 GO 239 GO

HAKATA															
KOKURA															
SHINSHIMONOSEKI															
OGORI															
TOKUYAMA															
SHINJIAKASHI															
HIROSHIMA			6:03				6:51	7:36							
MIHARA			6:30				7:18	-							
FUKUYAMA			6:44				7:32	-							
SHINKURASHIKI			7:00				7:48	-							
OKAYAMA		7:03	7:13		7:26		7:51	8:03	8:33			8:58	9:15		
AIOI		7:25	7:37		7:40		8:13	8:25	-			9:22	9:37		
HIMEJI		7:36	7:48		7:59		8:24	8:36	-			9:37	9:49		
NISHIAKASHI		7:51	8:03		8:14		8:39	8:51	-			9:52	10:03		
SHINKOBE		8:03	8:15		8:26		8:51	9:03	-			10:04	10:15		
SHINOSAKA	8:06	8:22	8:34	8:30	8:46	9:02	9:10	9:22	9:34	9:38	9:42	10:02	10:22	10:34	10:38
KYOTO	8:25	8:41	8:53	8:57	9:05	9:21	9:29	9:41	9:53	9:57	10:01	10:21	10:41	10:53	10:57
MAIBARA	8:52	-	-	9:23	-	9:47	-	-	-	10:23	10:28	10:47	-	-	11:23
GIFUHASHIMA	-	-	-	9:44	-	10:08	-	-	-	10:45	-	11:08	-	-	11:44
NAGOYA	9:19	9:31	9:43	10:00	9:55	10:24	10:19	10:31	10:43	11:00	10:55	11:24	11:31	11:43	12:00
TOYOHASHI	-	-	-	10:31	-	10:55	-	-	-	11:31	-	11:55	-	-	12:31
HAMAMATSU	-	-	-	10:40	-	11:12	-	-	-	11:48	-	12:12	-	-	13:48
SHIZUOKA	-	-	-	11:10	-	11:42	-	-	-	12:10	-	12:42	-	-	13:10
MISHIMA	-	-	-	11:47	-	12:11	-	-	-	12:47	-	13:11	-	-	13:47
ATAMI	-	-	-	11:58	-	12:22	-	-	-	12:58	-	13:22	-	-	13:58
ODAWARA	-	-	-	12:12	-	12:36	-	-	-	13:12	-	13:36	-	-	14:12
SHINYOKOHAMA	-	-	-	12:32	-	12:56	-	-	-	13:32	-	13:56	-	-	14:32
TOKYO	11:20	11:32	11:44	12:52	11:56	13:16	12:20	12:32	12:44	13:52	12:56	14:16	13:52	13:44	14:52

HIKARI HIKARI KODAMA HIKARI HIKARI KODAMA HIKARI KODAMA HIKARI HIKARI KODAMA HIKARI KODAMA HIKARI KODAMA HIKARI HIKARI
100 GO 20 GO 232 GO 104 GO 4 GO 254 GO 172 GO 256 GO 106 GO 6 GO 250 GO 174 GO 252 GO 108 GO 8 GO

HAKATA	6:06	6:19		6:44	7:24				7:44	8:24				8:44	9:24
KOKURA	6:39	6:52		7:17	7:57				8:17	8:57				9:17	9:57
SHINSHIMONOSEKI	6:51	-		7:29	-				8:29	-				9:29	-
OGORI	7:14	7:21		7:52	-				8:52	-				9:52	-
TOKUYAMA	7:40	-		8:12	-				9:12	-				10:12	-
SHINJIAKASHI	8:00	-		8:32	-				9:32	-				10:32	-
HIROSHIMA	8:21	8:12		8:53	9:12				9:53	10:12				10:53	11:12
MIHARA	8:48	-		9:20	-				10:20	-				11:20	-
FUKUYAMA	9:02	-		9:34	-				10:34	-				11:34	-
SHINKURASHIKI	9:10	-		9:50	-				10:50	-				11:50	-
OKAYAMA	9:33	9:09		10:05	10:09		10:27		11:05	11:09		11:27		12:05	12:09
AIOI	-	-		-	-		10:49		-	-		11:49		-	-
HIMEJI	10:03	-		10:37	-		11:00		11:37	-		12:00		12:37	-
NISHIAKASHI	-	-		-	-		11:18		-	-		12:15		-	-
SHINKOBE	10:20	-		11:03	-		11:27		12:03	-		12:27		13:03	-
SHINOSAKA	10:46	10:10	11:02	11:22	11:10	11:38	12:02	12:22	12:10	12:38	12:46	13:02	13:22	13:10	13:10
KYOTO	11:05	10:29	11:21	11:41	11:29	11:57	12:05	12:21	12:41	12:29	12:57	13:05	13:21	13:41	13:29
MAIBARA	-	-	11:47	-	-	12:23	-	-	-	-	13:23	-	13:47	-	-
GIFUHASHIMA	-	-	12:08	-	-	12:44	-	-	-	-	13:44	-	14:08	-	-
NAGOYA	11:55	11:19	12:24	12:31	12:19	13:00	12:55	13:24	13:31	13:19	14:00	13:55	14:24	14:31	14:19
TOYOHASHI	-	-	12:55	-	-	13:31	-	13:55	-	-	14:31	-	14:55	-	-
HAMAMATSU	-	-	13:12	-	-	13:48	-	14:12	-	-	14:48	-	15:12	-	-
SHIZUOKA	-	-	13:42	-	-	14:18	-	14:42	-	-	15:18	-	15:42	-	-
MISHIMA	-	-	14:11	-	-	14:47	-	15:11	-	-	15:47	-	16:11	-	-
ATAMI	-	-	14:22	-	-	14:58	-	15:22	-	-	15:58	-	16:22	-	-
ODAWARA	-	-	14:56	-	-	15:12	-	15:36	-	-	16:12	-	16:36	-	-
SHINYOKOHAMA	-	-	-	-	-	15:32	-	15:56	-	-	16:32	-	16:56	-	-
TOKYO	13:56	13:20	15:16	14:32	14:20	15:52	14:56	16:16	15:32	15:20	16:52	15:56	17:16	16:32	16:20

付録2. Voice Q-A System I の構文解析部の動作例

```

RECOGNITION PROCESS ( SPEAKER KI )
                     ( SYSTEM VOICE Q-A SYSTEM I )
( TSUITACHINO SHINOSAKAKARA HAKATAMADE 6JI2FUNHATSUNO
  HIKARI19GONO SHITEKENO 9MAI ONEGAISHIMASU )

( TSUITACHINO )
PHONEME LATTICE
( 1 ( ( S R H B T ) 4 2 ) )
( 2 ( ( I E ) 7 3 ) )
( 3 ( ( K T S ) 6 4 ) ( ( ZY KY HY ) 7 5 ) )
( 4 ( ( A ) 4 6 ) )
( 5 ( ( A ) 3 6 ) )
( 6 ( ( K H D T S ) 7 7 ) )
( 7 ( ( I ) 4 8 ) )
( 8 ( ( N G R M ) 2 9 ) )
( 9 ( ( O ) 11 END ) )

**** RECOGNITION PROCESS ****
( T U I T A T I )
( TSUITACHI ( 1 ) 8 9 )
( N O )
( NO ( 8 9 ) END 9 )
( H U T U K A )
( M I K K A )
( Y O K K A )
( I T U K A )
( ITSUKA ( 1 ) 6 )
( N O )
( O )
( D E )
( N O O )
( N O D E )
( W A )
( G A )
( N O W A )
( M U I K A )
( N A N O K A )
( Y O O K A )
( K O K O N O K A )
( T O O K A )
( H A T U K A )
( KY O O )
( H O N N Z I T U )
( A S U )
( A S I T A )
( A S A K K E )
( N I )
( 2 ( 1 ) 3 2 )
( ZY U U )
( ZY U U )
( S A N N )

RECOGNITION RESULT
(( ( 1 ) N I C H I ) ( --- K A R A ) ( --- M A D E ) ( --- J I --- F U N H A T S U ) ( --- --- G O ) ( --- S E K I
 ) ( --- M A I ) )
( S H I N O S A K A K A R A )

PHONEME LATTICE
( 1 ( ( S H R ) 6 2 ) ( ( S ) 10 4 ) ( ( ZY KY HY ) 21 6 ) )
( 2 ( ( I ) 2 3 ) )
( 3 ( ( R S H ) 1 4 ) ( ( Y HY ZY ) 12 6 ) )
( 4 ( ( I ) 8 5 ) )
( 5 ( ( * R ) 1 6 ) )
( 6 ( ( U ) 4 7 ) ( ( U O ) 4 8 ) ( ( U O N N ) 6 9 ) )
( 7 ( ( Y * R ) 2 10 ) )
( 8 ( ( Y * R ) 2 11 ) )
( 9 ( ( * ) 1 10 ) )
( 10 ( ( * ) 1 11 ) ( ( O E ) 8 12 ) )
( 11 ( ( O E ) 7 12 ) )
( 12 ( ( T K S ) 6 13 ) )
( 13 ( ( A ) 2 14 ) )
( 14 ( ( K T D S ) 5 15 ) )
( 15 ( ( A ) 3 16 ) )
( 16 ( ( T K D ) 5 17 ) )
( 17 ( ( A ) 3 18 ) ( ( A ) 6 20 ) ( ( A ) 12 END ) )
( 18 ( ( * R ) 2 19 ) )
( 19 ( ( A ) 6 END ) )
( 20 ( ( * ) 1 21 ) )
( 21 ( ( A ) 5 END ) )

**** RECOGNITION PROCESS ****
( T O O K Y O O )
( S I N N Y O K O H A M A )
( O D A W A R A )
( A T A M I )
( M I S I M A )
( S I Z U O K A )
( SHIZUOKA ( 1 ) 14 )

```

継続時間 (×15msec)
 セグメント・ポインター
 マッチングを試みた単語辞書
 マッチングが成功した結果を示すリスト

(E K I)
 (K A R A)
 (H A T U)
 (Y O R I)
 (H A M A M A T U)
 (T O Y O H A S I)
 (N A G O Y A)
 (G I H U H A S I M A)
 (M A I B A R A)
 (KY OO T O)
 (S I NN OO S A K A)
 (SHINOSAKA (1) 16)
 (E K I)
 (K A R A)
 (KARA (16) END 18 19 20 21)
 (S I NN K OO B E)
 (N I S I A K A S I)
 (H I M E Z I)
 (A I O I)
 (O K A Y A M A)
 (S I NN K U R A S I K I)
 (H U K U Y A M A)
 (M I H A R A)
 (H I R O S I M A)
 (S I NN I W A K U N I)
 (T O K U Y A M A)
 (O C OO R I)
 (S I NN S I M O N O S E K I)
 (K O K U R A)
 (H A K A T A)

RECOGNITION RESULT

(((1) NICH I)) ((SHINOSAKA) KARA) (--- MADE) (--- JI --- FUN HATSU) (--- --- GO) (--- SEKI) (--- MAI)

(HAKATAMADENO)

PHONEME LATTICE

(1 ((B S R D M) 4 2))
 (2 ((A) 2 3))
 (3 ((K T D S) 5 4))
 (4 ((A) 3 5))
 (5 ((S K T H) 5 6))
 (6 ((A) 3 7))
 (7 ((M N Z G) 4 8))
 (8 ((A) 5 9))
 (9 ((G Z) 3 10))
 (10 ((E U O) 4 11))
 (11 ((M N C R Z) 4 12))
 (12 ((O) 9 END))

**** RECOGNITION PROCESS ****

(T OO KY OO)
 (S I NN Y O K O H A M A)
 (O D A W A R A)
 (A T A M I)
 (M I S I M A)
 (S I Z U O K A)
 (H A M A M A T U)
 (T O Y O H A S I)
 (N A G O Y A)
 (G I H U H A S I M A)
 (M A I B A R A)
 (KY OO T O)
 (S I NN OO S A K A)
 (S I NN K OO B E)
 (N I S I A K A S I)
 (H I M E Z I)
 (A I O I)
 (O K A Y A M A)
 (S I NN K U R A S I K I)
 (H U K U Y A M A)
 (M I H A R A)
 (H I R O S I M A)
 (S I NN I W A K U N I)
 (T O K U Y A M A)
 (O C OO R I)
 (S I NN S I M O N O S E K I)
 (K O K U R A)
 (H A K A T A)
 (HAKATA (1) 7 8)
 (E K I)
 (M A D E)
 (MADE (7 8) 11)
 (N O)
 (NO (11) END 12)

RECOGNITION RESULT

(((1) NICH I)) ((SHINOSAKA) KARA) ((HAKATA) MADE) (--- JI --- FUN HATSU) (--- --- GO) (--- SEKI) (--- MAI)

```

( ROKUJINIHUNHATSUNO )
PHONEME LATTICE
( 1 ( ( P R H * B ) 1 2 ) ( ( T N M R B ) 4 4 ) )
( 2 ( ( U ) 1 3 ) ( ( O ) 6 5 ) )
( 3 ( ( R M ) 1 4 ) )
( 4 ( ( O ) 3 5 ) )
( 5 ( ( T K D ) 6 6 ) )
( 6 ( ( U E ) 1 7 ) )
( 7 ( ( C Z N D M ) 3 8 ) )
( 8 ( ( I ) 4 9 ) )
( 9 ( ( M N R C Z ) 4 10 ) )
( 10 ( ( I ) 3 11 ) )
( 11 ( ( K S H ) 7 12 ) )
( 12 ( ( U O NN ) 11 13 ) )
( 13 ( ( Z H M ) 6 14 ) )
( 14 ( ( A ) 2 15 ) )
( 15 ( ( T S K P ) 9 16 ) )
( 16 ( ( NN ) 2 17 ) )
( 17 ( ( R M ) 1 18 ) )
( 18 ( ( O ) 12 END ) )
**** RECOGNITION PROCESS ****
( R O K U )
( 6 ( 1 ) 7 8 )
( Z I )
( JI ( 7 8 ) 9 10 )
( ** )
( R E I )
( R EE )
( N I )
( 2 ( 9 10 ) 11 )
( ZY UU )
( S A NN )
( Y O NN )
( 4 ( 9 10 ) 14 13 )
( ZY UU )
( G O )
( ZY UU )
( I )
( I ( 9 10 ) 11 )
( PP U NN )
( R O )
( H A )
( H U NN )
( FUN ( 11 ) 14 15 13 )
( H A T U )
( HATSU ( 14 15 13 ) 17 18 16 )
( N O )
( NO ( 17 18 16 ) END 18 )
( N A N A )
( H A T I )
( KY UU )
( P U NN )
( Z I )
( ZY U )
( Z I )
( ZY U )
( Z I )
( ZI ( 9 10 ) 11 )
( ZY U )
( N A N A )
( H A T I )
( 8 ( 1 ) 7 )
( Z I )
( JI ( 7 ) 9 10 )
( K U )
( KU ( 1 ) 3 4 5 2 )
( Z I )
( JI ( 3 4 5 2 ) 7 )
( ** )
( R E I )
( R EE )
( N I )
( 2 ( 7 ) 9 10 )
( S A NN )
( Y O NN )
( G O )
( ZY UU )
( I )
( I ( 7 ) 9 )
( PP U NN )
( R O )
( H A )
( H U NN )
( FUN ( 9 10 ) 14 13 )
( H A T U )
( HATSU ( 14 13 ) 17 18 16 )
( N A N A )
( H A T I )
( KY UU )

```

(Z I)
 (ZI (7) 9 10)
 (PP U NN)
 (ZY U)
 (ZY UU)
 (N I)
 (2 (1) 3)
 (ZY UU)

RECOGNITION RESULT

(((1) NICH I) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6 9 8) JI (2) FUN HATSU) ((
 --- GO) (--- SEKI) (--- MAI))

(HIKARIZYUKYUGONO)

PHONEME LATTICE

(1 ((H S * P R) 2 2) ((S) 10 4))
 (2 ((I) 1 3))
 (3 ((K T D) 5 4))
 (4 ((A) 5 5))
 (5 ((M R G Z N) 3 6))
 (6 ((I) 4 7))
 (7 ((D Z H) 5 8) ((ZY KY) 13 10))
 (8 ((E) 5 9) ((E U A) 11 11))
 (9 ((Y *) 1 10))
 (10 ((U E) 3 11))
 (11 ((K T) 8 12) ((ZY KY) 16 14))
 (12 ((E) 5 13) ((U) 11 15))
 (13 ((Y * R) 2 14))
 (14 ((U O) 3 15))
 (15 ((G Z M N H) 4 16))
 (16 ((O) 13 17))
 (17 ((C Z N M H) 3 18))
 (18 ((O) 11 END))

**** RECOGNITION PROCESS ***

(H I K A R I)
 (HIKARI (1) 7)
 (**)
 (I T I)
 (N I)
 (S A NN)
 (Y O NN)
 (G O)
 (R O K U)
 (N A N A)
 (H A T I)
 (8 (7) 12)
 (G OO)
 (KY UU)
 (HY A K U)
 (100 (7) 15 12)
 (N I)
 (2 (15 12) 16)
 (ZY UU)
 (G O)
 (5 (15 12) 17 18 16)
 (ZY UU)
 (R O K U)
 (N A N A)
 (H A T I)
 (KY UU)
 (ZY UU)
 (ZY UU)
 (10 (7) 11)
 (I T I)
 (N I)
 (S A NN)
 (Y O NN)
 (G O)
 (R O K U)
 (N A N A)
 (H A T I)
 (KY UU)
 (9 (11) 15 16)
 (G OO)
 (GO (15 16) 17 18)
 (N O)
 (NO (17 18) END 18)
 (G OO)
 (I T I)
 (G OO)
 (S A NN)
 (Y O NN)
 (G OO)
 (GO (17 18 16) END 17)
 (G OO)
 (GO (15 12) 17 18)
 (K O D A M A)

RECOGNITION RESULT

(((1) NICH I) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6 9 8) JI (2) FUN HATSU) ((
 HIKARI) (100 105 19) GO) (--- SEKI) (--- MAI))

(SHITEKENO)
 PHONEME LATTICE
 (1 ((S H) 4 2) ((S) 12 4))
 (2 ((I U) 5 3))
 (3 ((K T S) 5 4))
 (4 ((E) 11 5))
 (5 ((K T) 6 6) ((ZY KY) 10 7))
 (6 ((E) 6 8))
 (7 ((E U) 2 8))
 (8 ((*) 1 9) ((*) 8 11))
 (9 ((NN) 7 10))
 (10 ((*) 1 11))
 (11 ((O) 13 END))

**** RECOGNITION PROCESS ****
 (S I T E I)
 (SHITEI (1) 5)
 (K E NN)
 (KEN (5) 10 11 8 9)
 (N O)
 (NO (10 11 8 9) END 11)
 (H U T UU)
 (FUTSU (1) 3)
 (K E NN)
 (KEN (3) 8 9 11 5)
 (N O)
 (NO (8 9 11 5) END 11)
 (C U R II NN)

RECOGNITION RESULT
 ((1) NICHII) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6 9 8) JI (2) FUN HATSU) ((
 HIKARI) (100 105 19) GO) ((FUTSU) SEKI) (--- MAI)

(KYUMAI)
 PHONEME LATTICE
 (1 ((S T K H R) 5 2) ((Y KY HY) 12 5))
 (2 ((E) 5 3) ((E I U) 6 4) ((U) 11 6))
 (3 ((* R) 1 4))
 (4 ((*) 1 5) ((U O) 4 6))
 (5 ((U O NN) 4 6))
 (6 ((M N Z H) 6 7))
 (7 ((A) 7 8))
 (8 ((Y * R G) 3 9))
 (9 ((E I U) 3 END))

**** RECOGNITION PROCESS ****
 (I T I)
 (N I)
 (S A NN)
 (Y O NN)
 (4 (1) 7 6)
 (M A I)
 (MAI (7 6) END 8 9)
 (C O)
 (R O K U)
 (N A N A)
 (H A T I)
 (KY UU)
 (9 (1) 6 7)
 (M A I)
 (MAI (6 7) END 8 9)

RECOGNITION RESULT
 (1) NICHII) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6 9 8) JI (2) FUN HATSU) ((
 HIKARI) (100 105 19) GO) ((FUTSU) SEKI) ((9 4) MAI)

(ONEGAISHIMASU)
 PHONEME LATTICE
 (1 ((T * B R G) 2 2))
 (2 ((O) 3 3))
 (3 ((C R N M Z) 3 4) ((Y HY ZY) 11 6))
 (4 ((E) 5 5))
 (5 ((* R) 1 6) ((* R) 2 7))
 (6 ((A O) 12 8))
 (7 ((A O) 11 8))
 (8 ((S HY) 7 9))
 (9 ((E I U) 1 10))
 (10 ((M R N G Z) 3 11))
 (11 ((A) 7 END))

**** RECOGNITION PROCESS ****
 (Y O Y A K U)
 (YOYAKU (1) 10 9)
 (I T A)
 (S I)
 (SHI (10 9) 10)
 (M A)
 (MAS (10) END)

RECOGNITION RESULT (WITHOUT INFERENCE)
 ((1) NICHII) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6 9 8) JI (2) FUN HATSU) ((
 HIKARI) (100 105 19) GO) ((FUTSU) SEKI) ((9 4) MAI)

RESPONSE SENTENCE
 9 MAI DESUKA

FINAL RESULT AFTER INFERENCE
 ((1) NICHII) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6) JI (2) FUN HATSU) ((HIKARI
) (19) GO) ((FUTSU) SEKI) ((9 4) MAI)

PENALTY SCORES
 ((6)) ((0)) ((4)) ((4)) JI (4) ((100) (100)) ((3)) ((1 1))

付録3. Voice Q-A System の発声リスト

(グループ1.) 予約項目の発声順序自由

- (1) 1日の、新大阪から、博多までの、6時2分発の、ひかり19号の、指定券を、9枚、お願いします。
- (2) 新神戸発の、6時22分の、グリーン券を、小倉まで、8枚。
- (3) 3日の、こだま213号で、普通席を、西明石から、新下関まで、7枚、予約します。
- (4) 岡山から、広島への、4日の、ひかり91号の、指定席を、6枚。
- (5) 姫路駅より、小郡駅までの、グリーン券で、5日の、こだま217号を、5枚。
- (6) 6日の、ひかり17号で、京都から、博多までの、指定を、4枚。
- (7) 相生発の、9時30分の、徳山までの、グリーン券は、ありますか。
- (8) 東京から、名古屋までの、6時0分発の、グリーン券で、8日のは、ありますか。
- (9) 9日ので、新横浜発の、6時22分の、こだま201号の、米原ゆきの、指定券は、ありませんか。
- (10) 新倉敷から、10時49分発の、新岩国までの、指定を、1枚。
- (11) ひかり151号で、11日のを、名古屋から、相生まで、2枚、グリーン券を、予約します。
- (12) 12日、小田原発の、7時10分の、岐阜羽島までの、3枚の、指定券を、申し込みます。
- (13) ひかり121号で、京都から、三原までの、指定席で、13日の、9時41分発は、ありませんか。
- (14) 14日の、熱海から、7時45分発の、こだま207号で、5枚の、グリーン券は。
- (15) 15日の、ひかり1号で、東京から、小倉までの、普通席を、6枚。
- (16) 三島から、豊橋への、指定を、7枚。
- (17) 17日、ひかり153号、名古屋発の、9時27分の、姫路への、グリーン券を、8枚。
- (18) 18日ので、新大阪駅から、博多駅までの、ひかり105号の、グリーンを、9枚。
- (19) 静岡から、京都までの、指定券で、19日の、9時16分発のは、ありますか。
- (20) 20日の、ひかり191号で、東京から、7時48分発、名古屋いきの、指定を、8枚。

(グループ2.) 予約項目の発声順序自由(続き)

- (21) 21日、ひかり3号、京都から、広島まで。
- (22) 浜松から、米原まで、22日の、こだま215号の、指定を、6枚。

- (23) ひかり 153 号で，10 時 53 分発のを，相生まで，5 枚。
- (24) 24 日の，ひかり 105 号で，姫路から，11 時 28 分発の，指定を，4 枚，徳山まで。
- (25) 25 日で，こだま 211 号で，豊橋から，新大阪まで，グリーンを，3 枚。
- (26) 26 日の，7 時 48 分，東京発の，ひかり 191 号の，普通を，2 枚，米原まで。
- (27) 27 日の，指定券で，岡山から，小倉まで，12 時 12 分のを。
- (28) 新横浜から，浜松への，28 日の，こだま 215 号の，グリーンを，1 枚。
- (29) 29 日の，10 時 27 分，名古屋発の，西明石ゆきの，グリーンを。
- (30) 熱海から，岐阜羽島駅まで，指定券を，3 枚，30 日の，9 時 33 分発の，こだま 219 号で。
- (31) 31 日の，ひかり 107 号で，新神戸駅から，小郡駅への，グリーンを，4 枚。
- (32) 9 時 0 分発，東京からで，博多までの，1 日の，グリーンを，5 枚，予約します。
- (33) 2 日の，こだま 223 号で，三島から，豊橋まで，6 枚，お願い致します。
- (34) 名古屋から，新大阪まで，3 日の，指定は，ありませんか。
- (35) こだま 225 号で，静岡より，新大阪までの，指定で，8 枚，4 日のは，ありませんか。
- (36) 5 日の，新倉敷から，小郡まで，14 時 31 分発の，ひかり 109 号は，ありませんか。
- (37) 6 日ので，京都より，小倉いきの，グリーン券で，12 時 53 分発の，ひかり 5 号で，9 枚。
- (38) 7 日の，ひかり 127 号で，東京から，岡山まで，グリーンを，8 枚。
- (39) 浜松から，新大阪まで，8 日の，こだま 229 号で，指定券を，7 枚，お願いします。
- (40) 東京から，西明石ゆきの，9 日の，指定券は，ありますか。

(グループ 3) 予約項目の発声順序固定

- (1) 1 日の，新大阪から，博多までの，6 時 2 分発の，ひかり 19 号の，指定券を，9 枚，お願いします。
- (2) 2 日の，新神戸発の，小倉まで，6 時 22 分の，こだま 205 号の，グリーン券を，8 枚，お願いします。
- (3) 3 日の，西明石から，新下関まで，7 時 17 分の，こだま 213 号で，普通席を，7 枚，予約します。
- (4) 4 日の，岡山から，広島への，8 時 2 分発，ひかり 91 号の，指定席を，6 枚，予約します。
- (5) 5 日の，姫路駅より，小郡駅までの，8 時 24 分の，こだま 217 号で，グリーン券を，5 枚，お願いします。

- (6) 6日の、京都から、博多までの、7時53分発の、ひかり17号で、指定を、4枚、予約します。
- (7) 7日の、相生発の、徳山までの、9時30分の、こだま221号の、グリーン券は、3枚、ありますか。
- (8) 8日の、東京から、名古屋まで、6時0分発の、ひかり21号の、グリーン券は、2枚、ありますか。
- (9) 9日ので、新横浜発の、米原ゆきの、6時22分の、こだま201号の、指定券は、1枚、ありませんか。
- (10) 10日の、新倉敷から、新岩国までの、10時49分発の、ひかり101号の、指定を、1枚、お願いします。
- (11) 11日の、名古屋から、相生まで、8時27分の、ひかり151号の、グリーン券を、2枚、予約します。
- (12) 12日の、小田原発の、岐阜羽島までの、7時10分の、こだま203号の、指定券を、3枚、申し込めます。
- (13) 13日の、京都から、三原までの、9時41分発の、ひかり121号の、指定券は、4枚、ありませんか。
- (14) 14日の、熱海から、米原までの、7時45分発の、こだま207号の、グリーン券は、5枚、ありますか。
- (15) 15日の、東京から、小倉までの、7時0分発の、ひかり1号の、普通券を、6枚、お願いします。
- (16) 16日の、三島から、豊橋への、8時23分の、こだま209号の、指定を、7枚、予約します。
- (17) 17日の、名古屋発の、姫路への、9時27分の、ひかり153号の、グリーン券を、8枚、お願いします。
- (18) 18日の、新大阪駅から、博多駅までの、10時48分の、ひかり105号の、グリーンを、9枚、予約します。
- (19) 19日の、静岡から、京都までの、9時16分発の、こだま211号の、指定券は、9枚、ありますか。
- (20) 20日の、東京から、名古屋いきの、7時48分発、ひかり191号で、指定を、8枚、お願いします。

(グループ4) 予約項目の発声順序固定, 構文制限

- (1) 1日の, 新大阪から, 博多まで, 6時2分の, ひかり19号の, 指定を, 9枚。
- (2) 2日の, 新神戸から, 小倉まで, 6時22分の, こだま205号の, グリーンを, 8枚。
- (3) 3日の, 西明石から, 新下関まで, 7時17分の, こだま213号の, 普通を, 7枚。
- (4) 4日の, 岡山から, 広島まで, 8時2分の, ひかり91号の, 指定を, 6枚。
- (5) 5日の, 姫路から, 小郡まで, 8時24分の, こだま217号の, グリーンを, 5枚。
- (6) 6日の, 京都から, 博多まで, 7時53分の, ひかり17号の, 指定を, 4枚。
- (7) 7日の, 相生から, 徳山まで, 9時30分の, こだま221号の, グリーンを, 3枚。
- (8) 8日の, 東京から, 名古屋まで, 6時0分の, ひかり21号の, グリーンを, 2枚。
- (9) 9日の, 新横浜から, 米原まで, 6時22分の, こだま201号の, 指定を, 1枚。
- (10) 10日の, 新倉敷から, 新岩国まで, 10時49分の, ひかり101号の, 指定を, 1枚。
- (11) 11日の, 名古屋から, 相生まで, 8時27分の, ひかり151号の, グリーンを, 2枚。
- (12) 12日の, 小田原から, 岐阜羽島まで, 7時10分の, こだま203号の, 指定を, 3枚。
- (13) 13日の, 京都から, 三原まで, 9時41分の, ひかり121号の, 指定を, 4枚。
- (14) 14日の, 熱海から, 米原まで, 7時45分の, こだま207号の, グリーンを, 5枚。
- (15) 15日の, 東京から, 小倉まで, 7時0分の, ひかり1号の, 普通を, 6枚。
- (16) 16日の, 三島から, 豊橋まで, 8時23分の, こだま209号の, 指定を, 7枚。
- (17) 17日の, 名古屋から, 姫路まで, 9時27分の, ひかり153号の, グリーンを, 8枚。
- (18) 18日の, 新大阪から, 博多まで, 10時48分の, ひかり105号の, グリーンを, 9枚。
- (19) 19日の, 静岡から, 京都まで, 9時16分の, こだま211号の, 指定を, 9枚。
- (20) 20日の, 東京から, 名古屋まで, 7時48分の, ひかり191号の, 指定を, 8枚。

付録4. FORTRANによるリスト処理システムDLOP

計算機が得意とするのは、その名が示すように算術計算であると思われてきた。しかし現在では、算術計算以外の分野にも多いに適用されている。ごく身近な例では、OSとかエディターなどがあり、これらは、計算という概念からほど遠いものである。このような利用形態を考えると、計算機は、むしろ、汎用記号処理機械 (a general-purpose symbol manipulating machine) とでも呼ばれるべきものであると考えられる。さらに、将来は、身かけ上、人に命令されたことを忠実に実行するだけではなくなると思われる。この徴候は、自然言語処理やパターン認識などの人工知能の研究の分野に表われはじめており、そこでは、計算機に個々の知識を手続きとして与え、それらの利用は、問題解決の手法で、適当に、計算機が組み合わせて動作するような方法がとられている。このような場合では人が計算機に、直接、動作を命令していないと考えることもできる。

ところで、我々パターン認識の研究者の大部分は、FORTRANのみを使って、認識のアルゴリズムをプログラミング化している。この主たる原因は、プログラム言語として、FORTRANが身近にあること、入力データの処理にはある程度の計算が必要で、かつ、配列型のデータ・ベースが便利であることがあげられる。しかしながら、パターンの構造の記述などの問題になると配列型データでは困難でより柔軟なデータ・ベースであるトリー型のデータ・ベースも必要となってくる。DLOPは、著者が作成したFORTRANによるリスト処理システムである。DLOPはData List Oriented Processorの略であり、文献(73)(108)の流れをくむリスト処理システムである。DLOPの特徴は、FORTRANでトリー構造のデータを扱えることと、プッシュダウン・スタックを用いて帰納的なプログラムを記述できるようにしている点である。現在、DLOPは、U-400上に作成されており、会話音声認識系の言語処理の作成や音響処理の評価プログラムの作成に利用されている。

以下DLOPの特徴を述べ、さらに、DLOPの基本的な構成を述べる。最後にDLOPのユーティリティ・ルーチンとリスト処理ルーチンを示す。

DLOPの特徴

DLOPはミニコンU-400上に作成されている。U-400は、セグメント方式の主記憶をもつ特異なミニコンである。セグメント方式は、図A4-1に示すような主記憶の配置からなる。主記憶は、コモン・ブロック(CB)とセグメント・ブロック(SB)からなり、OSはコモ

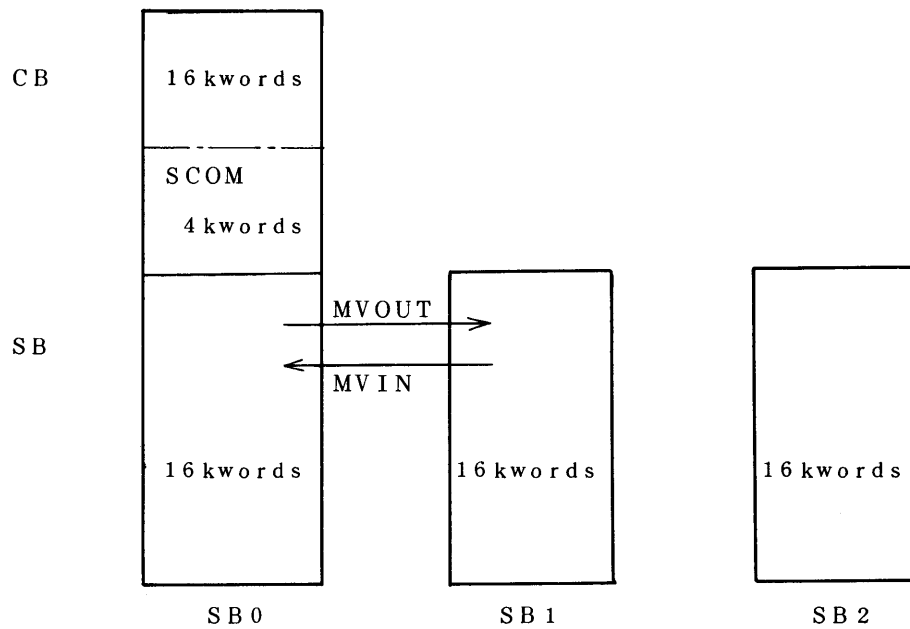


図 A 4-1 セグメント方式

ン・ブロックに常駐している。ユーザの利用可能領域は、コモン・ブロックのシステム・コモン領域 (SCOM) の 4 kwords と 3 枚のセグメント・ブロック (各々 16 kwords) の計 52 kwords である。さらに、セグメント・ブロック間の特殊な命令として、セグメント・ブロック 0 (SB0) とセグメント・ブロック 1 (SB1) あるいは 2 (SB2) の間のデータ転送命令 MVIN, MVOUT がある。我々は、セグメント・ブロック 1 (SB1) をリスト領域として、命令 MVIN, MVOUT を利用して、リスト処理のためのユーティリティ・ルーチンを作成した。

DLOP の特徴として、次の 7 項目があげられる。

- (1) FORTRAN プログラムでリスト処理がおこなえる。

FORTRAN によって、トリー型のデータを容易に扱うことができ、リスト処理や記号列処理も可能である。

- (2) 帰納的アルゴリズムをプログラミングするために、プッシュダウン・スタックとそれに関するユーティリティをユーザに開放している。

パターン・マッチングの際に利用するトリーサーチの depth-first method などの帰納的なアルゴリズムをプログラミングするのが容易である。

- (3) FORTRAN を主言語としているため、機能の拡張が容易である。

他の計算機と接続した場合のデータ形式の変更 (例えば、配列→リスト表現) のプログラ

ムも FOTRAN で容易に書くことができる。

- (4) 不要なセルは、自動的なガーベージ・コレクションによって回収される。

リスト型変更の宣言で、必要なリストを登録することにより、必要なセルと不必要なセルを区別している。よって、自由リスト領域の空セルがなくなるとガーベージ・コレクターが起動し、不要なセルを回収して、空セルとすることができる。

- (5) リストに名前をつけることができる。

リストに名前（記号列，整数）をつけることができ、その名前を使って、リストを検索するプログラムなどが作成できる。

- (6) 豊富な入出力ルーチンが用意してある。

すべての入出力機器に対して、リスト表現の入出力ができる。

- (7) 固定リスト領域（プログラム中で変更の必要のないデータ・ベース用）と自由リスト領域とがあり、かつ、両者の交換も可能であり、効率よくリスト領域が利用できる。

変更の必要がないリストと変更の必要があるリストを別々に記憶することにより、ガーベージ・コレクションの速度が短縮できる。

リストの内部表現は、LISP とほとんど同じ簡潔な表現方法をとっている。現在、アトムとしては、記号列と整数値（-999～9239）をとることができる。

DLOP の基本構成

U-400 上の DLOP では、リストを表わすための領域として、セグメント・ブロック 1 (SB1) の 16kwords とシステム・コモン領域 (SCOM) の 2kwords が使われる。表 A 4-1 に、DLOP でリスト処理のためにとられる領域を示す。さらに、図 A 4-2 にリスト処理のための領域の SB1 および SCOM 上の配置を示す。

アトムとしては、記号列と整数と NIL を許している。記号列アトムはアトムテーブルに蓄えられる。同じ記号列は、アトムテーブルの同じ場所に蓄えられる。整数と NIL は、リスト領域のセルの中に蓄えられる。

リストを内部表現にして蓄えるために、リスト領域 (list area) とアトムテーブル (atom table) を使う。リスト領域は、リスト自由領域 (free list area) とリスト固定領域 (fixed list area) とに分けられている。変更することのない知識のようなリスト・データは、リスト固定領域に蓄えられる。このリスト固定領域は、ユーティリティによって、ディスクとの転送も可能になっている。一方リスト自由領域は、通常のリスト処理に利用され、ガ

表 A 4 - 1 DLOP のリスト処理のための領域

リスト領域名	ワード数	テーブル数	主記憶の位置
free list area	4 kwords	2 kcells	SB1
fixed list area	8 kwords	4 kcells	SB1
atom table	2 kwords	512 tables	SB1
push-down stack	2 kwords	—	SB1
list name table	956 words	478 tables	SCOM
list variable table	64 words	32 tables	SCOM
input and output buffer	1 kwords	—	SCOM
garbage pointer 1 *	1 words	—	SCOM
garbage pointer 2 **	1 words	—	SCOM
push-down stack pointer	1 words	—	SCOM
list variable	1 words	—	SCOM
total	18 kwords	SB1 16kwords SCOM 2kwords	

* garbage pointer in free list area

**garbage pointer in fixed list area

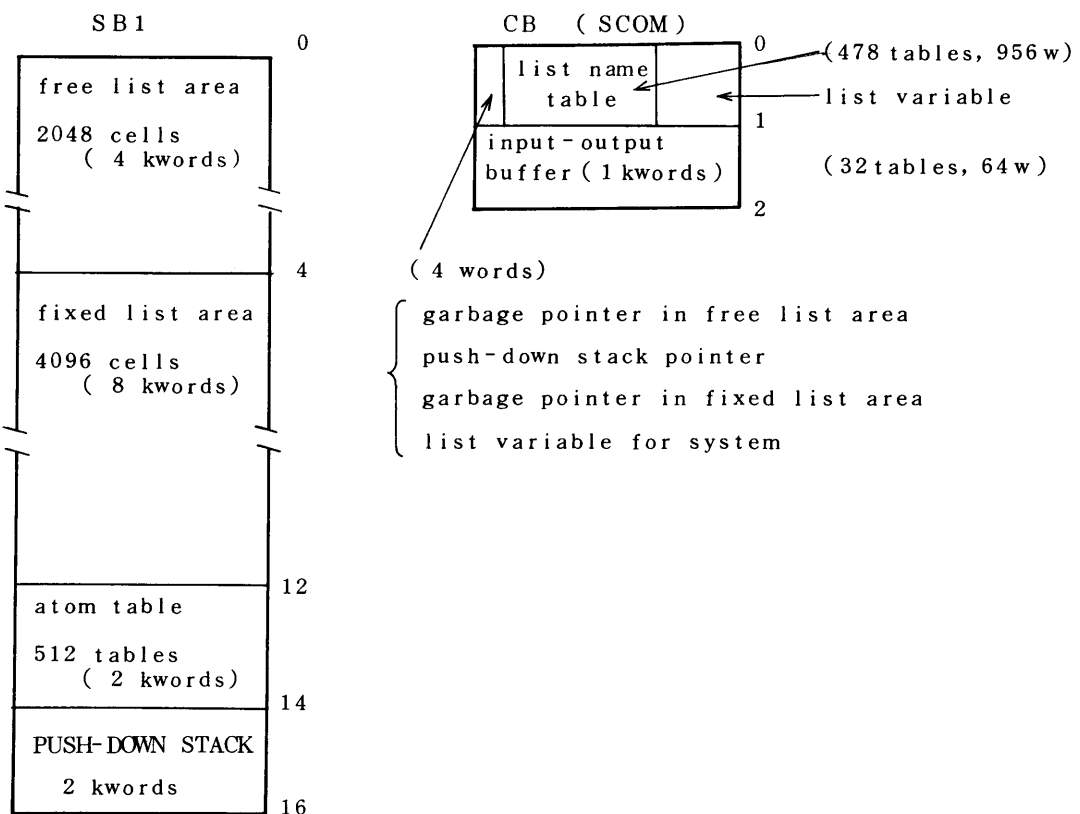
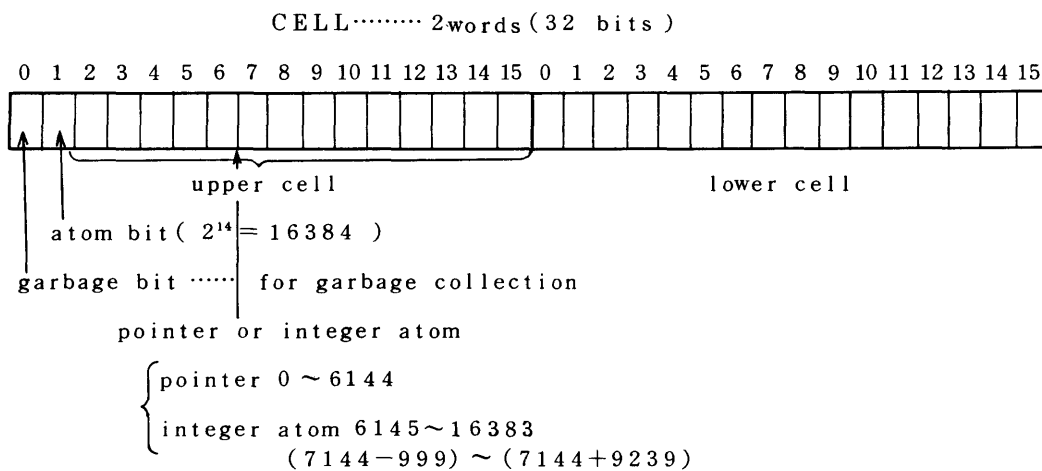


図 A 4 - 2 リスト処理のための領域

ページ・コレクションの対象にもなる。また、ユーティリティによって、両リスト領域の交換も可能である。

DLOPのセル(Cell)は、2語からなり、1語は16ビットである。16ビットのうち、下の14ビット(2ビット目から15ビット目まで)がポインタ(pointer)として用いられる。実際には、1~6144までの値がポインタとして用いられ、6145~16383の値は、整数値アトムを表わすのに用いられる。0ビット目は、ガーベージ・コレクションのためのガーベージ・ビット(garbage bit)として、1ビット目は、記号列アトムとNILを示すアトム・ビット(atom bit)として使われる。このセルのビット配分の様子を図A4-3に示す。



図A4-3 セルのビット配分

セルの中でNILは、 $2^{14} (= 16384)$ の値で表わされ、整数値(-999~9239)アトムは、

$$7144 + \text{〔 整数値 〕}$$

で表わされる。記号列アトムは、

$$2^{14} + \text{〔 アトム・テーブルのアドレス 〕}$$

として表わされる。

記号列アトムを蓄えるアトムテーブルは、4語を1組とする512個のテーブル(table)からなっている。1語目は、記号列の長さを表わす。ここでは、1語に2文字入るので、1テーブルに6文字まで格納できる。1語目の値は、

$$\left\{ \frac{\text{〔 記号列の長さ 〕} + 1}{2} \right\}$$

で表わされており、その値が4以上(記号列の長さが7以上)のときは、4語目に、続きのテ

アドレス	1 語目	2 語目	3 語目	4 語目
1	1~3			
2	4~30			続きのテーブル のアドレス
3	-1	-	-	-

} 記号列
空テーブル

図 A 4 - 4 アトム・テーブルの表現

1	2	4		
2	16385	3		
3	16386	16384		
4	5	9		
5	16387	6		
6	7	16384		
7	16385	8		
8	16388	16384		
9	10	20		
10	16389	11		
11	12	14		
12	16385	13		
13	16390	16384		
14	15	17		
15	16385	16		
16	16391	16384		
17	18	16384		
18	16385	19		
19	16392	16384		
20	21	16384		
21	16385	22		
22	16393	16384		

1	1	*		
2	3	EK	IM	EI
3	2	OP	T	
4	2	EK	I	
5	1	OR		
6	2	KA	RA	
7	3	HA	TS	U
8	2	YO	RI	
9	1	AX		

((*EKIMEI) (OPT (*EKI))
(OR (*KARA) (*HATSU)
(*YORI)) (*AX))

図 A 4 - 5 リスト表現の内部表現

テーブルのアドレスが入れられる。記号列の長さは、60 まで許される。また、1 語目の値が-1 であるときは、そのテーブルが空であることを示しており、新しい記号列を格納することができる。この様子を図 A 4 - 4 に示す。

内部表現の例として、リスト表現 ((*EKIMEI) (OPT (*EKI)) (OR (*KARA) (*HATSU) (*YORI)) (*AX)) を図 A 4 - 5 に示す。

また、リストに名前をつけるために、リスト名テーブル (list name table) を用意している。これは、478 テーブルからなり、1 テーブルは 2 語からなっている。このテーブルの 1 語目にはリスト名の内部表現が、2 語目には、リストのアドレスが蓄えられる。このテーブル

を用いることにより、リスト間の検索を容易に行うことができる。

ガーベージ・コレクションを自動的に行うには、必要なリストと不必要なリストとの区別を行う必要がある。DLOPでは、必要なリストは、前もってリスト型変数であると宣言する方法をとった。そのために、リスト変数テーブル(list variable table)を用意した。これは、32テーブルからなり、1テーブルは2語からなっている。第1語目に、主記憶上のリスト型変数の絶対アドレスを、第2語目に、配列の大きさを蓄える。また、リスト処理関数LCONSなどで、リスト作成中にガーベージ・コレクションが生じても、その途中結果を保持できるようにするために、システムのリスト変数(list variable)が1語用意されている。

ガーベージ・コレクションは、ガーベージ・ポインタ(garbage pointer)が0になったときに動作し、リスト型変数によって束ばくされていないリスト自由領域のセルを集め再利用できるようにする。

プッシュダウン・スタック(push-down stack)の深さは、2048語で、プッシュダウン・ポインタ(push-down stack pointer)をもっている。このスタックは、DLOPのユーティリティ・ルーチンでも利用されているし、また、ユーザにも開放されている。ユーザは、このスタックを利用して、帰納的アルゴリズムをプログラミングすることができる。

DLOPのルーチン

DLOPのルーチンは、基本的なユーティリティ・ルーチンとリスト処理ルーチンとにわけることができる。Sと示されたのはサブルーチンを、Fは関数であることを示す。また、*は実行用アセンブラルーチンもあることを示す。

ユーティリティ・ルーチン

DLOP	……	リスト処理の初期設定(S)
SETLST	……	リスト型変数の宣言(S)
GCOLL	……	ガーベージ・コレクター(S)
PDOWN	……	プッシュダウン(S*)
POPUP	……	ポップ・アップ(S*)
KEEP	……	スタックに区切りを入れる(S)
RELES	……	スタックを区切りまでポップ・アップ(S)
LCONS	……	L, Mからリスト(L, M)をつくる(F*)
LCAR	……	(L, M)からLを取り出す(F*)

LCDR (L, M) から M を取り出す (F*)
 RPLCAR (L, M) の L を取りかえる (S*)
 RPLCDR (L, M) の M を取りかえる (S*)
 IATOM 整数値アトムをつくる (F*)
 INTATM 整数値アトムから整数を得る (F*)
 KATOM 記号列アトムをつくる (F)
 LREAD リスト表現を内部表現に変換 (S)
 LWRITE 内部表現をリスト表現に変換 (S)
 WRTATM アトムを内部表現から S 式に変換 (S)
 NGPTCH リスト自由領域とリスト固定領域を交換 (S)
 NLSTRD 名前付きのリストをカード・リーダーから読み込む (S)
 LV リスト名からリストのアドレスを求める (F*)
 LVIN リストのアドレスからリスト名を求める (F*)
 DKATOM アトム・テーブルから記号列アトムを取り除く (S)
 DLSTV リスト型変数の宣言を取り消す (S)
 ADRES 変数名からその絶対アドレスを求める (S)
 LVADRS 絶対アドレスからその値を求める (S)
 PDS, NPDS, ATOM1, ATOM2, NATOM1, NATOM2, NLST1, NLST2, NLIST1,
 NLIST2 S B 1 のリスト処理のための領域からのデータの読み出し, 書き込み
 のアセンブラ・ルーチン

リスト処理ルーチン

LCAAR LISP の CAAR (F*)
 LCDDR LISP の CDDR (F*)
 LCADR LISP の CADR (F*)
 LCDAR LISP の CDAR (F*)
 LCDDDR LISP の CDDDR (F*)
 LST 2 つの要素のリストをつくる (F)
 LST 3 3 つの要素のリストをつくる (F)
 LST 4 4 つの要素のリストをつくる (F)
 LST 5 5 つの要素のリストをつくる (F)

LAST	……	リストの最後の要素を求める (F)
MEMBER	……	リストの中に要素として含まれているか否か (F)
LEQUAL	……	2つのリストが等しいか否か (F)
LCOPY	……	同じリストをつくる (S)
PPRNT, LPRNT, FPRNT, TPRNT, DPRNT, LLPRNT, FFPRNT	……	リストを出力機器に出力 (S)
LWATM, FWATM	……	アトムを出力機器に出力 (S)
RREAD, CREAD, FREAD, TREAD, DREAD	……	リスト表現を入力機器から読み込み, 内部表現をつくる (S)
CORED R	……	リスト処理領域とディスクとの転送 (S)
DSTFL	……	リスト固定領域をディスクに転送する (S)
DLDFL	……	リスト固定領域をディスクから読み出す (S)
APPEND	……	リストに他のリストの要素を加える (S)
AP	……	リストに他のリストを加える (S)
LDELET, IDELET, LINSRT, RPLCE, ICHNG, DELETA, DELET	……	リスト中の要素の変更
UNION	……	2つのリストのユニオンをとる (S)
INTERS	……	2つのリストのインタセクションをとる (S)
FLATTN	……	リストの深さをとり去ったリストをつくる (S)
REVLST	……	リスト中の要素の順序を逆にする (S)
PAIRLS	……	参照リストにつけ加える (S)
LASSOC	……	参照リストから特定のペアをとり出す (S)
GRAF	……	トリーをグラフに変換 (S)
REVS	……	リストの構造と要素を逆にしたリストをつくる (S)

以上, ユーティリティ・ルーチンと現在使用可能なリスト処理ルーチンをあげた。ユーザは, これらのユーティリティ・ルーチンと必要なリスト処理ルーチンを用いてプログラムすればよい。

付録6. 大語い単語音声認識実験における誤り

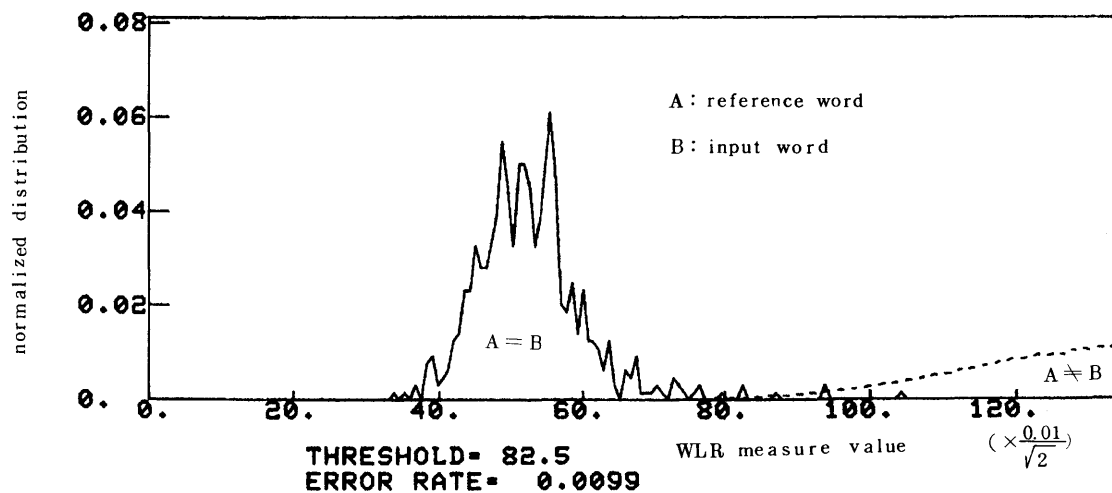
(8msec. ステップ分析, 10次LPC)

LPCスペクトル・マッチング尺度: WLR (単語認識率 96.4%)

発声単語		第1位	第2位	第3位	第4位	第5位
????	19 MON-BETSU 88888.000	NOBORIBETSU 89.009	NIHON-MATSU 89.370	GO:TSU 91.543	JO:ETSU 96.366	KUDAMATSU 99.260
????	39 MUTSU 88888.000	FUTTSU 80.894	SETTSU 85.884	BUZEN 90.898	FUCHU: 92.074	GOSEN 92.526
***	65 NOSHIRO 72.212 ←スコア	TOCHIO 66.238	KUSHIRO 70.732	GOJO: 71.324	NOSHIRO 72.212	NISHIO 75.000
*	69 OGA 53.465	KOGA 52.674	OGA 53.465	TOBA 83.239	TODA 87.218	SAGA 93.320
*	72 KAZUNO 75.415	TATSUNO 57.974	KAZUNO 75.415	HAN-NO 77.128	SANO 78.651	KAZO 80.215
*	92 SUKAGAWA 56.632	FUKAGAWA 50.622	SUKAGAWA 56.632	TAGAWA 57.298	SUNAGAWA 61.060	SAWARA 75.645
*	93 KITAKATA 54.512	SAKATA 54.204	KITAKATA 54.512	ASAKA 70.610	HIRAKATA 72.000	HIRATA 77.382
*	99 KOGA 46.465	OGA 45.431	KOGA 46.465	TODA 69.094	TOBA 79.600	OBAMA 81.951
*	199 KIMIZU 57.942	SHIMIZU 55.934	KIMIZU 57.942	NI:TSU 72.120	NI:ZA 85.253	IZUMO 88.171
*	226 AKIKAWA 49.500	TAKIKAWA 44.435	AKIKAWA 49.500	TACHIKAWA 65.611	KAKEGAWA 71.327	GUSHIKAWA 73.617
*	246 SAN-JO: 47.601	AN-JO: 45.986	SAN-JO: 47.601	HON-JO: 61.843	SAIJO: 62.693	HON-JO: 65.926
????	275 KOMATSU 93.660	KARATSU 82.150	NAKATSU 85.576	O:TSU 86.100	OWASE 91.960	NUMAZU 92.011
*	277 SUZU 72.156	TSU 58.875	SUZU 72.156	TSURU 78.086	UTO 90.770	MUTSU 94.192
**	288 KO:FU 73.388	O:BU 57.129	CHO:FU 68.166	KO:FU 73.388	HO:FU 73.885	KOKUBU 86.230
*	305 NAKANO 59.971	NAGANO 54.688	NAKANO 59.971	NAGATO 67.164	FURANO 73.598	TAMANO 78.365
????	311 SAKU 88888.000	TAKU 52.877	OTARU 89.655	SANO 92.939	SAGA 93.656	ESASHI 93.966
*	318 MINO 47.918	MINOO 47.179	MINO 47.918	NAGO 65.578	EBINO 69.391	TATSUNO 70.046
****	369 O:BU 86.833	HO:FU 75.438	KO:FU 79.852	CHO:FU 80.758	KOKUBU 86.273	O:BU 86.833
**	370 CHITA 103.811	TAMA 96.052	TOSA 100.286	CHITA 103.811	KOGA 112.117	HITA 112.610
????	436 KATANO 88888.000	SANO 56.678	HADANO 63.086	NAKANO 75.322	TAMANO 76.093	TATSUNO 80.932
*	492 SO:JA 57.312	O:DA 57.195	SO:JA 57.312	O:TA 64.899	O:MUTA 72.240	SO:KA 73.414
***	518 HO:FU 72.780	O:BU 43.867	CHO:FU 66.994	KO:FU 70.467	HO:FU 72.780	KOKUBU 83.840
*	533 UWAJIMA 50.819	WAJIMA 49.187	UWAJIMA 50.819	MASUDA 79.426	KUWANA 81.117	IWANUMA 83.597
****	550 SUKUMO 93.125	KAMO 84.172	SANO 87.655	ONO 88.229	YONO 89.172	SUKUMO 93.125
*	556 I:ZUKA 63.872	IIDA 59.806	I:ZUKA 63.872	NI:ZA 66.335	KAIZUKA 77.330	IRUMA 80.831
????	574 TOSU 88888.000	MUTSU 88.780	TOCHIO 95.118	TSU 97.360	TOSA 97.622	TOKI 99.194
*	633 ISHIKAWA 60.773	ICHIKAWA 59.990	ISHIKAWA 60.773	TAKIKAWA 69.934	TACHIKAWA 70.414	AKIKAWA 73.672
ORDER	1 2 3 4 5 10			20 50	100 200	200
SAMPLES	614 15	2 2	2 1	0 0	0 0	0 5
RATE(%)	0.9579 0.0234	0.0031 0.0031	0.0031 0.0016	0. 0.	0. 0.	0. 0.0078
ACCUM.RATE	0.9579 0.9813	0.9844 0.9875	0.9906 0.9922	0.9922 0.9922	0.9922 0.9922	0.9922 1.0000

????	19	MON-BETSU	88888.000	O:ME	95.929	KO:BE	100.140	GOSE	101.205	HON-DO	105.289	UBE	105.475
????	39	MUTSU	88888.000	UOZU	83.218	TOSU	88.780	SUZU	94.192	GO:TSU	97.855	O:TSU	97.944
*	69	OGA	53.465	KOGA	45.431	OGA	53.465	SO:MA	65.561	OBAMA	70.244	TOBA	70.542
*	174	HASUDA	60.526	MASUDA	55.091	HASUDA	60.526	KAKUDA	70.013	KASUGA	72.020	HASHIMA	74.007
*	177	ICHIKAWA	60.021	ISHIKAWA	59.990	ICHIKAWA	60.021	GUSHIKAWA	71.475	TACHIKAWA	77.088	TAKIKAWA	80.022
????	275	KOMATSU	93.660	NIHON-MATSU	86.415	NUMAZU	89.333	KUDAMATSU	89.890	KO:NOSU	90.249	O:MACHI	91.159
*	288	KO:FU	73.388	HO:FU	70.467	KO:FU	73.388	O:BU	79.852	O:FUNATO	87.287	CHO:FU	90.222
***	297	UEDA	82.459	IKEDA	76.013	FUJIEDA	80.321	UENO	81.542	UEDA	82.459	FUSSA	93.363
*	308	CHINO	60.580	HINO	57.131	CHINO	60.580	MINOO	81.742	MINO	82.782	UTO	85.110
????	311	SAKU	88888.000	KUSATSU	81.727	SAKURAI	91.601	KARATSU	95.580	TAKATSUKI	96.343	YAMAGUCHI	97.780
*	334	FUJI	57.030	UJI	56.145	FUJI	57.030	KUJI	70.057	TSUKUMI	80.450	ZUSHI	80.703
????	369	O:BU	86.833	HO:FU	43.867	KO:FU	57.129	CHO:FU	59.043	O:ZU	68.020	O:MUTA	69.087
????	370	CHITA	103.811	HITA	76.531	CHIBA	80.695	USA	85.677	HIRATA	94.697	INA	95.911
*	371	CHIRYU:	65.703	KIRYU:	61.528	CHIRYU:	65.703	CHICHIBU	83.171	SHIMIZU	96.588	FUCHU:	97.109
**	376	TSU	76.279	SUZU	58.875	TSURU	67.645	TSU	76.279	SUZUKA	90.508	USA	96.716
????	436	KATANO	88888.000	TAKASAGO	85.537	NAKANO	87.055	HARAMACHI	88.111	HANAMAKI	88.601	KASAOKA	90.141
????	550	SUKUMO	93.125	IZUMO	82.052	MINOO	86.270	NAGO	86.980	YONO	87.042	GOBO:	87.783
*	557	TAGAWA	57.331	SUKAGAWA	57.298	TAGAWA	57.331	KAKOGAWA	67.674	SHIRAKAWA	67.773	FUKAGAWA	67.994
????	574	TOSU	88888.000	O:TSU	68.953	GO:TSU	69.112	RYO:TSU	73.747	UOZU	79.170	KUSATSU	83.101
ORDER		1	2	3	4	5	10	20	50	100	200	200	
SAMPLES		622	8	1	1	0	4	0	0	0	0	0	5
RATE(%)		0.9704	0.0125	0.0016	0.0016	0.	0.0062	0.	0.	0.	0.	0.	0.0078
ACCUM. RATE		0.9704	0.9828	0.9844	0.9860	0.9860	0.9922	0.9922	0.9922	0.9922	0.9922	0.9922	1.0000

AVERAGED RECOGNITION RATE													
ORDER	1	2	3	4	5	10	20	50	100	200	200	200	10
SAMPLES	1236	23	3	3	2	5	0	0	0	0	0	0	0
RATE(%)	0.9641	0.0179	0.0023	0.0023	0.0016	0.0039	0.	0.	0.	0.	0.	0.	0.0078
ACCUM. RATE	0.9641	0.9821	0.9844	0.9867	0.9883	0.9922	0.9922	0.9922	0.9922	0.9922	0.9922	0.9922	1.0000

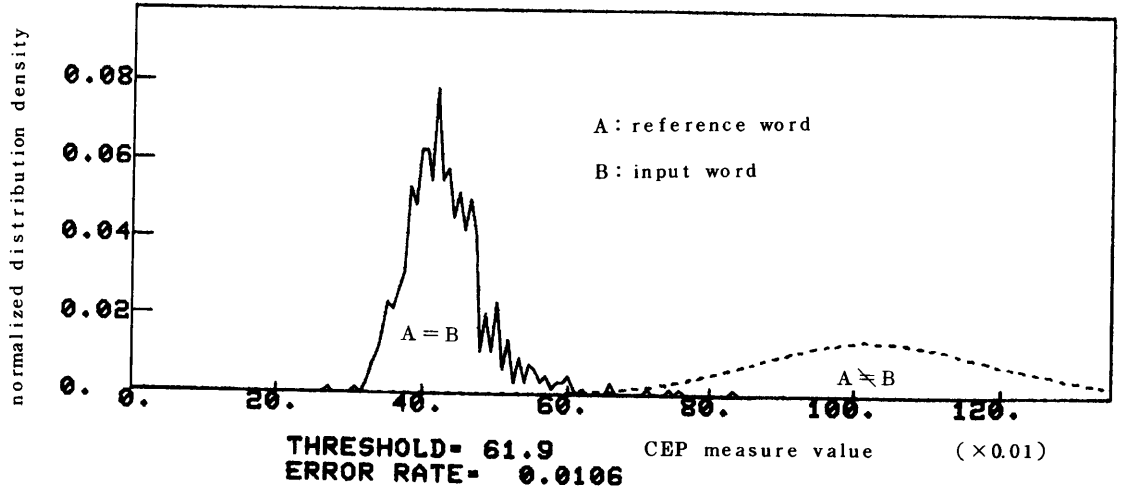


LPC スペクトル・マッチング尺度：CEP (単語認識率：95.4%)

発声単語		第 1 位	第 2 位	第 3 位	第 4 位	第 5 位						
????	19 MON-BETSU 88888.000	NOBORIBETSU 69.828	NIHON-MATSU 69.856	KOMATSU 70.113	FUTTSU 70.877	JO:ETSU 73.728						
????	39 MUTSU 88888.000	FUTTSU 58.187	SETTSU 60.906	GOSEN 66.355	BUZEN 68.388	NOSHIRO 69.544						
*	51 ESASHI 58.843 ←スコア	HITACHI 56.836	ESASHI 58.843	TANASHI 60.480	ISESAKI 63.118	HISAI 66.154						
*	64 AKITA 48.929	KATSUTA 47.549	AKITA 48.929	TAKETA 52.033	TAKIKAWA 58.673	SAKATA 59.223						
**	65 NOSHIRO 55.868	TOCHIO 51.437	KUSHIRO 54.993	NOSHIRO 55.868	GOJO: 57.011	NISHIO 64.701						
*	68 HON-JO: 52.788	HON-DO 49.429	HON-JO: 52.788	AN-JO: 54.500	SAN-JO: 60.797	TEN-DO 67.240						
***	72 KAZUNO 61.896	TATSUNO 47.421	HADANO 59.722	HAN-NO 61.890	KAZUNO 61.896	KANUMA 64.692						
*	92 SUKAGAWA 46.556	FUKAGAWA 42.017	SUKAGAWA 46.556	TAGAWA 50.205	SUNAGAWA 51.479	KAGA 61.979						
*	168 KUKI 54.095	TOKI 49.435	KUKI 54.095	MIKI 64.154	FUJI 65.122	KUJI 65.617						
*	199 KIMIZU 45.424	SHIMIZU 42.592	KIMIZU 45.424	NI:TSU 56.188	NI:ZA 64.285	IZUMO 67.414						
*	212 HINO 47.376	CHINO 46.010	HINO 47.376	MINO 61.230	UTO 66.431	NAGO 70.258						
*	226 AKIKAWA 37.810	TAKIKAWA 36.005	AKIKAWA 37.810	TACHIKAWA 47.543	GUSHIKAWA 50.848	KAKEGAWA 52.924						
*	246 SAN-JO: 43.081	AN-JO: 42.404	SAN-JO: 43.081	HON-JO: 56.660	SAIJO: 59.507	TAGAJO: 59.510						
*	251 KAMO 59.074	SANO 53.409	KAMO 59.074	KAZO 71.892	NACO 76.094	YONO 78.117						
***	275 KOMATSU 70.613	KARATSU 64.776	NAKATSU 67.797	NUMAZU 68.530	KOMATSU 70.613	KUMANO 72.238						
*	277 SUZU 53.859	TSU 41.854	SUZU 53.859	TSURU 62.160	MUTSU 72.481	UTO 73.960						
*	288 KO:FU 52.064	O:BU 46.747	KO:FU 52.064	HO:FU 52.311	CHO:FU 58.224	KOKUBU 67.829						
*	305 NAKANO 48.389	NAGANO 48.309	NAKANO 48.389	HADANO 60.514	NAGATO 61.028	FURANO 61.360						
????	311 SAKU 88888.000	TAKU 39.833	SANO 63.439	ESASHI 68.787	SAGA 73.326	OTARU 73.336						
**	313 O:CAKI 58.440	O:KUCHI 52.350	O:MACHI 52.419	O:CAKI 58.440	O:TSUKI 59.179	KOMAKI 62.318						
*	334 FUJI 42.091	KUJI 40.627	FUJI 42.091	UJI 44.221	KUKI 54.158	TOKI 62.008						
????	369 O:BU 74.126	HO:FU 58.611	KO:FU 63.690	CHO:FU 68.804	O:NO 71.296	HO:JO: 72.217						
**	370 CHITA 82.622	TAMA 77.156	HITA 80.351	CHITA 82.622	TOSA 83.403	ESASHI 83.507						
*	371 CHIRYU: 43.783	KIRYU: 42.869	CHIRYU: 43.783	CHICHIBU 55.397	FUCHU: 64.841	GIFU 68.250						
????	436 KATANO 88888.000	SANO 48.550	HADANO 51.781	NAKANO 57.178	KUMANO 64.194	TAMANO 64.237						
*	513 HO:FU 59.978	O:BU 44.000	HO:FU 59.978	KO:FU 61.489	CHO:FU 61.969	O:NO 64.980						
*	517 HIKARI 46.760	HITACHI 46.750	HIKARI 46.760	KITAKAMI 53.922	HISAI 54.546	KITAMI 54.794						
***	550 SUKUMO 75.727	ONO 66.679	KAMO 67.959	YONO 73.057	SUKUMO 75.727	SANO 75.982						
*	556 I:ZUKA 45.775	IIDA 43.829	I:ZUKA 45.775	NI:ZA 49.657	KAIZUKA 56.163	IYAMA 58.678						
????	574 TOSU 88888.000	MUTSU 64.520	TOSA 66.389	TOCHIO 74.609	TSU 75.101	USA 78.172						
*	587 KUMAMOTO 57.143	KITAMOTO 56.297	KUMAMOTO 57.143	NARUTO 69.011	MUROTO 69.272	KUMANO 69.969						
ORDER	1	2	3	4	5	10	20	50	100	200	200	5
SAMPLES	610	19	3	3	0	1	0	0	0	0	0	5
RATE (%)	0.9516	0.0296	0.0047	0.0047	0.	0.0016	0.	0.	0.	0.	0.	0.0078
ACCUM. RATE	0.9516	0.9813	0.9860	0.9906	0.9906	0.9922	0.9922	0.9922	0.9922	0.9922	0.9922	1.0000

????	19	MON-BETSU 88888.000	O:ME	GOSE	UBE	KO:BE	BUZEN							
????	39	MUTSU 88888.000	TOSU	UOZU	SUZU	GOSE	USA							
*	168	KUKI 54.095	TOKI	KUKI	FUJI	TSUKUMI	KIKUCHI							
*	174	HASUDA 46.103	MASUDA	HASUDA	KATSUTA	KAKUDA	MATSUBARA							
*	199	KIMIZU 45.424	SHIMIZU	KIMIZU	NI:TSU	I:ZUKA	CHICHIBU							
*	205	O:ME 45.745	KO:BE	O:ME	O:DATE	HIKONE	UBE							
*	261	SHIRONE 60.250	HIKONE	SHIRONE	CHITOSE	YOKOTE	HIROSAKI							
**	267	UOZU 55.780	O:ZU	O:TSU	UOZU	GO:TSU	RYO:TSU							
????	275	KOMATSU 70.613	NIHON-MATSU	KUDAMATSU	NUMAZU	NUMATA	NAKATSU							
*	297	UEDA 59.886	IKEDA	UEDA	FUJIEDA	UENO	UBE							
*	308	CHINO 55.367	HINO	CHINO	UTO	MINO	SETO							
????	311	SAKU 88888.000	KUSATSU	KARATSU	SAKURAI	YAMAGUCHI	HATOGAYA							
*	334	FUJI 42.091	UJI	FUJI	KUJI	ZUSHI	KIKUCHI							
*	358	AN-JO: 42.967	SAN-JO:	AN-JO:	HON-JO:	SHIN-JO:	HON-JO:							
????	369	O:BU 74.126	HO:FU	KO:FU	CHO:FU	O:MUTA	O:ZU							
????	370	CHITA 82.622	HITA	CHIBA	HIRATA	USA	INA							
*	376	TSU 55.407	SUZU	TSU	TSURU	SUZUKA	USA							
*	426	MINOO 46.067	MINO	MINOO	MITO	ITO:	TEN-DO							
????	436	KATANO 88888.000	NAKANO	TAKASAGO	HARAMACHI	KARATSU	HANAMAKI							
*	441	NISHINOMIYA 41.603	FUJINOMIYA	NISHINOMIYA	ICHINOMIYA	UTSUNOMIYA	ISHINOMAKI							
*	456	SAN-DA 45.772	HAN-DA	SAN-DA	AN-NAKA	SAKATA	MASUDA							
**	513	HO:FU 59.978	KO:FU	O:BU	HO:FU	CHO:FU	O:MURA							
*	522	TOKUSHIMA 52.758	FUKUSHIMA	TOKUSHIMA	KUSHIMA	KACOSHIMA	TSUSHIMA							
????	550	SUKUMO 75.727	YONO	IZUMO	KOKUBU	SUMOTO	CHIKUCO							
????	574	TOSU 88888.000	GO:TSU	RYO:TSU	O:TSU	UOZU	FUTTSU							
*	589	HITOYOSHI 65.591	KURAYOSHI	HITOYOSHI	MIYOSHI	YU:KI	TOTTORI							
*	590	ARAO 65.262	NANAO	ARAO	AKO:	WAKO:	HADANO							
*	615	KUSHIMA 41.400	TSUSHIMA	KUSHIMA	MISHIMA	HASHIMA	FUKUSHIMA							
ORDER		1	2	3	4	5	10	20	50	100	200	200	200	5
SAMPLES		613	17	2	0	0	3	1	0	0	0	0	0	0
RATE(%)		0.9563	0.0265	0.0031	0.	0.	0.0047	0.0016	0.	0.	0.	0.	0.	0.0078
ACCUM.RATE		0.9563	0.9828	0.9860	0.9860	0.9860	0.9906	0.9922	0.9922	0.9922	0.9922	0.9922	0.9922	1.0000

AVERAGED RECOGNITION RATE													
ORDER	1	2	3	4	5	10	20	50	100	200	200	200	10
SAMPLES	1223	36	5	3	0	4	1	0	0	0	0	0	0
RATE(%)	0.9540	0.0281	0.0039	0.0023	0.	0.0031	0.0008	0.	0.	0.	0.	0.	0.0078
ACCUM.RATE	0.9540	0.9821	0.9860	0.9883	0.9883	0.9914	0.9922	0.9922	0.9922	0.9922	0.9922	0.9922	1.0000

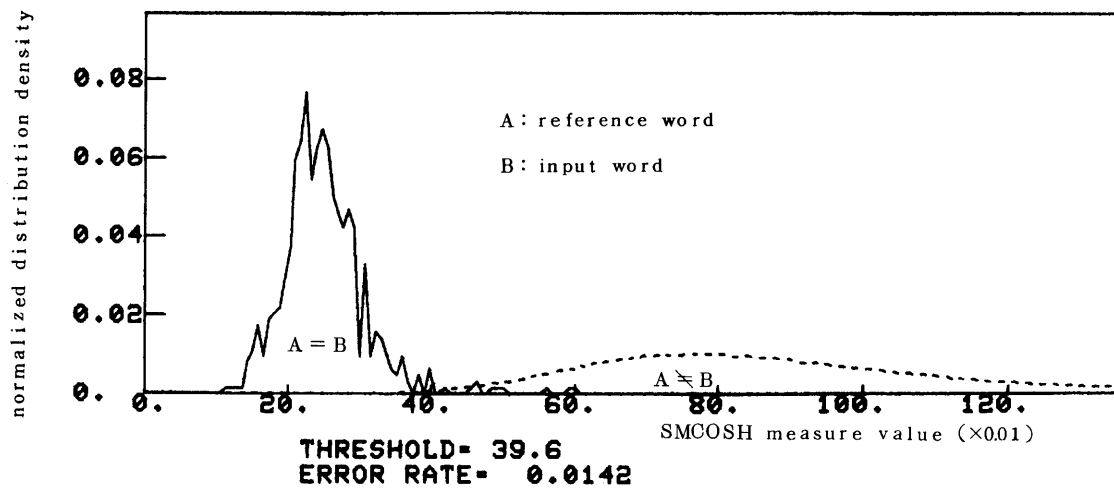


スペクトル・マッチング尺度：SMCOSH(単語認識率：94.3%)

発声単語	第1位	第2位	第3位	第4位	第5位							
???? 16 ASHIBETSU 56.093	TOSASHIMIZU 46.511	MATSUE 48.093	SAKAIDE 49.016	HASHIMOTO 49.327	TARUMIZU 49.948							
???? 19 MON-BETSU 88888.000	GO:TSU 41.062	JO:ETSU 44.686	NOBORIBETSU 47.113	KUDAMATSU 47.631	UOZU 47.878							
???? 39 MUTSU 88888.000	FUCHU: 48.807	TEN-RYU: 52.123	FUTTSU 52.253	FUCHU: 52.294	RYU:GASAKI 52.639							
???? 65 NOSHIO 39.180 ←スコア	GOJO: 30.680	KUSHIRO 33.477	NISHIO 33.982	TOCHIO 34.821	MUROTO 36.290							
* 69 OGA 21.713	KOGA 20.977	OGA 21.713	TOBA 33.526	NODA 38.918	TODA 39.918							
** 72 KAZUNO 33.476	TATSUNO 30.325	HAN-NO 32.762	KAZUNO 33.476	KAZO 33.594	MINOO 36.693							
** 92 SUKACAWA 25.109	FUKACAWA 24.023	TACAWA 24.532	SUKACAWA 25.109	SUNACAWA 26.836	SAWARA 31.374							
* 93 KITAKATA 32.268	SAKATA 29.322	KITAKATA 32.268	SAWARA 35.275	ASAKA 38.688	IWATA 41.792							
* 99 KOCA 20.790	OGA 18.578	KOGA 20.790	TODA 27.607	TOBA 30.576	SO:MA 37.090							
* 145 HAN-NO 20.779	AN-JO: 20.607	HAN-NO 20.779	SAN-NO 21.284	KAZO 23.257	HON-DO 28.534							
* 174 HASUDA 34.769	KATSUTA 33.962	HASUDA 34.769	KASUGA 41.567	KASHIWA 42.205	KAMEYAMA 43.520							
* 187 TO:GANE 30.000	TOKONAME 27.952	TO:GANE 30.000	KOMAGANE 31.760	O:DATE 35.362	KOGANEI 36.326							
* 199 KIMIZU 25.099	SHIMIZU 24.544	KIMIZU 25.099	GIFU 38.247	IZUMO 38.993	CHICHIBU 39.776							
* 208 CHO:FU 25.546	O:BU 20.901	CHO:FU 25.546	UOZU 38.822	O:ZU 39.053	KOKUBU 41.466							
* 226 AKIKAWA 23.680	TAKIKAWA 21.613	AKIKAWA 23.680	TACHIKAWA 36.032	ARIDA 38.155	OKINAWA 39.210							
* 246 SAN-JO: 18.570	AN-JO: 17.950	SAN-JO: 18.570	HON-JO: 26.690	HON-JO: 27.404	SAIJO: 27.877							
* 264 TOYAMA 28.780	OYAMA 27.513	TOYAMA 28.780	SAYAMA 34.275	YOKOHAMA 35.808	TSUYAMA 36.069							
???? 275 KOMATSU 58.588	O:TSU 41.735	KARATSU 42.557	NAKATSU 44.831	OTARU 46.131	NUMAZU 47.828							
** 277 SUZU 39.367	TSU 28.927	TSURU 33.190	SUZU 39.367	GIFU 40.102	UBE 42.220							
** 288 KO:FU 32.734	O:BU 24.624	CHO:FU 29.268	KO:FU 32.734	HO:FU 35.879	KOKUBU 37.156							
* 305 NAKANO 30.154	NAGANO 29.663	NAKANO 30.154	FURANO 34.384	NAGATO 38.583	HIRADO 39.585							
???? 311 SAKU 88888.000	TAKU 29.864	OTARU 47.933	SANO 50.367	SACA 56.295	TOSU 56.731							
* 318 MINO 20.896	MINOO 18.779	MINO 20.896	NACO 25.459	TATSUNO 31.795	URAWA 32.493							
* 343 SHIMODA 14.099	NODA 13.262	SHIMODA 14.099	GYO:DA 27.032	ONODA 29.216	ZAMA 30.896							
**** 344 SUSONO 32.069	TO:NO 20.965	ONO 23.937	O:NO 26.741	YONO 31.450	SUSONO 32.069							
* 354 TSUSHIMA 27.054	KUSHIMA 25.928	TSUSHIMA 27.054	MISHIMA 33.075	KASHIMA 36.885	IRUMA 39.277							
???? 369 O:BU 49.843	KO:FU 38.452	HO:JO: 38.792	CHO:FU 39.515	HO:FU 40.330	KOKUBU 42.156							
???? 370 CHITA 59.486	TAMA 49.571	KOGA 49.662	TOSA 55.416	KANO 56.868	TOBA 58.312							
* 376 TSU 46.128	TSURU 42.977	TSU 46.128	IYO 53.686	UTO 64.407	MUTSU 64.547							
* 407 SAKAI 37.924	AKASHI 36.205	SAKAI 37.924	TANASHI 39.727	TAKAHASHI 43.691	HITACHI 44.877							
* 430 SETTSU 45.530	EBETSU 45.437	SETTSU 45.530	TOSU 52.725	BEPPU 52.875	KUSATSU 54.207							
???? 436 KATANO 88888.000	SANO 28.178	HADANO 32.271	KAZO 37.504	TAMANO 38.818	TAGAWA 39.530							
* 446 TOYOOKA 29.702	TOYOKAWA 29.616	TOYOOKA 29.702	TOMIOKA 34.840	TOYOSAKA 36.418	O:KAWA 36.543							
* 484 GO:TSU 26.201	RYO:TSU 23.441	GO:TSU 26.201	UOZU 28.710	O:TSU 32.298	TOSU 38.179							
* 492 SO:JA 24.541	O:DA 19.185	SO:JA 24.541	GYO:DA 27.568	SO:MA 28.711	ONODA 29.770							
???? 513 HO:FU 41.577	O:BU 14.475	CHO:FU 25.025	KO:FU 33.077	O:ZU 33.928	KOKUBU 34.596							
* 533 UWAJIMA 21.955	WAJIMA 18.538	UWAJIMA 21.955	IWANUMA 34.040	NODA 38.470	MASUDA 39.011							
???? 550 SUKUMO 49.203	YONO 34.115	ONO 35.257	MINOO 39.594	SANO 41.265	MINO 43.042							
???? 556 I:ZUKA 38.199	IIDA 29.301	NI:ZA 30.414	SUZUKA 36.695	TSURUGA 37.313	IKEDA 37.682							
???? 574 TOSU 88888.000	MUTSU 43.350	TAKU 50.047	TOCHIO 50.491	TOKI 57.243	TSU 58.101							
ORDER	1	2	3	4	5	10	20	50	100	200	200	5
SAMPLES	601	22	4	0	1	6	1	1	0	0	0	0
RATE(%)	0.9376	0.0343	0.0062	0.	0.0016	0.0094	0.0016	0.0016	0.	0.	0.	0.0078
ACCUM.RATE	0.9376	0.9719	0.9782	0.9782	0.9797	0.9891	0.9906	0.9922	0.9922	0.9922	0.9922	1.0000

**	16	ASHIBETSU	TOSASHIMIZU	HIGASHIKURUME	ASHIBETSU	KITAKYU:SHU:	NOBORIBETSU					
????	19	MON-BETSU	NINOHE	KO:BE	O:ME	UBE	YOKOTE					
????	39	MUTSU	UOZU	TOSU	GO:TSU	NOSHIRO	O:TSU					
*	69	OCA	KOCA	OCA	SO:MA	O:DA	TOBA					
*	72	KAZUNO	HAN-NO	KAZUNO	AN-JO:	TATSUNO	SAN-JO:					
*	93	KITAKATA	HIRAKATA	KITAKATA	SAKATA	SAWARA	HIRATA					
*	142	GYO:DA	O:DA	GYO:DA	NODA	SHIMODA	SO:JA					
*	160	TODA	SO:MA	TODA	KOCA	O:DA	TOBA					
***	174	HASUDA	MASUDA	HASHIMA	KASUGA	HASUDA	KAKUDA					
*	177	ICHIKAWA	ISHIKAWA	ICHIKAWA	GUSHIKAWA	ISHIOKA	NISHIO					
*	182	NODA	SHIMODA	NODA	ZAMA	YUZAWA	SHIMADA					
*	208	CHO:FU	HO:FU	CHO:FU	KO:FU	KOKUBU	O:ZU					
*	261	SHIRONE	HIKONE	SHIRONE	TSUBAME	NINOHE	YOKOTE					
????	275	KOMATSU	O:MACHI	NIHON-MATSU	KO:NOSU	KUDAMATSU	O:DATE					
*	276	WAJIMA	UWAJIMA	WAJIMA	IWANUMA	IYOMISHIMA	HIROSHIMA					
????	297	UEDA	UENO	FUJIEDA	IKEDA	YONEZAWA	OKEGAWA					
*	305	NAKANO	NAGANO	NAKANO	KUMANO	NAGATO	FURANO					
????	311	SAKU	HAKUI	KUSATSU	SAKURAI	ATSUGI	TAKAISHI					
*	318	MINO	MINOO	MINO	HINO	CHINO	MITO					
*	328	SHIMIZU	KIMIZU	SHIMIZU	CHICHIBU	SHIN-GU:	FUCHU:					
*	335	IWATA	NUMATA	IWATA	SHIBATA	SUZAKA	IBARA					
????	369	O:BU	HO:FU	O:ZU	CHO:FU	KO:FU	O:MIYA					
????	370	CHITA	HITA	CHIBA	USA	TOSA	INA					
*	371	CHIRYU:	KIRYU:	CHIRYU:	SHIN-GU:	CHICHIBU	KIMIZU					
**	376	TSU	TSURU	SUZU	TSU	TSUKUMI	UBE					
**	430	SETTSU	FUTTSU	EBETSU	SETTSU	NI:TSU	O:TSU					
????	436	KATANO	TAKASAGO	KAWANOE	HANAMAKI	HARAMACHI	KANAZAWA					
****	513	HO:FU	KO:FU	O:BU	KOKUBUN-JI	O:MIYA	HO:FU					
*	532	IMABARI	NABARI	IMABARI	ISHINOMAKI	NAGAI	YU:BARI					
????	550	SUKUMO	IZUMO	MINOO	NAGO	YONO	GOBO:					
*	553	O:MUTA	O:MURA	O:MUTA	SO:JA	O:DA	O:TA					
????	574	TOSU	O:TSU	GO:TSU	UOZU	KUSATSU	RYO:TSU					
**	626	O:KUCHI	KO:CHI	ONOMICHI	O:KUCHI	O:TSUKI	CHO:SHI					
ORDER	1	2	3	4	5	10	20	50	100	200	200	5
SAMPLES	608	17	4	1	1	1	1	3	0	0	0	0
RATE(%)	0.9485	0.0265	0.0062	0.0016	0.0016	0.0016	0.0016	0.0047	0.	0.	0.	0.0078
ACCUM.RATE	0.9485	0.9750	0.9813	0.9828	0.9844	0.9860	0.9875	0.9922	0.9922	0.9922	0.9922	1.0000

AVERAGED RECOGNITION RATE												
ORDER	1	2	3	4	5	10	20	50	100	200	200	10
SAMPLES	1209	39	8	1	2	7	2	4	0	0	0	0
RATE(%)	0.9431	0.0304	0.0062	0.0008	0.0016	0.0055	0.0016	0.0031	0.	0.	0.	0.0078
ACCUM.RATE	0.9431	0.9735	0.9797	0.9805	0.9821	0.9875	0.9891	0.9922	0.9922	0.9922	0.9922	1.0000



付録7. Voice Q-A System II の音韻変形規則

PHONEME REWRITING RULES FOR /A/
 (A - ((E - 1 8 1) (U M D 5 2) (O H - 6 2) (E Y - 1 3 2)
 (O M - 7 4) (N N H - 4 3) (N N K - 4 3) (U K - 4 3) (E T T 5 2)
 (U H - 5 3) (O D - 7 3) (O K - 4 3) (E N N 1 0 3) (O W R
 1 0 3) (E H Y K 7 2) (O S K 4 3) (U H Y K 5 2) (E R - 4 3) (A 8
 1)) ((U 6 2) (E 8 2) (O 8 2) (I 6 3) (N N 5 3) (I 4 4))
 - ((R 1 0 2)
 (M 1 0 4) (W 1 0 4) (H 8 4) (N 1 0 3) (Y 1 0 4)))
PHONEME REWRITING RULES FOR /I/
 (I ((Y 7 3)) (E A - 9 1) (E - - 3 1) (E K - 6 2) (E T - 7 2)
 (E Z - 8 2) (U Z - 4 2) (U S - 8 2) (U N - 8 2)
 (N N S - 6 1) (U T - 7 1)
 (E S - 1 3 2) (U A - 7 3) (N N - - 5 2) (U - - 4 3)
 (E - - 8 3) (N N - - 7 3))
 ((I 7 1) (N N 7 2) (U 1 2 2) (O 1 0 3) (T 4 3) (I 5 3) (A 6 3) (E 8 2))
 ((E 1 0 2) (N N 6 2) (I 6 2)) - ((N N 1 4 3) (N 8 3)))
PHONEME REWRITING RULES FOR /U/
 (U - ((I - - 2 2) (I K - 3 2) (I T - 4 3)
 (E - - 4 3) (N N - - 1 1 2
) (O - - 2 2) (O K - 8 3) (O H - 1 1 1) (O P - 1 1 1) (O P P -
 1 1 1) (O G - 7 2) (O M - 5 2) (O S - 9 4) (O T - 7 2) (E T
 - 6 4) (E K - 6 3)))
 ((U 8 1) (N N 8 1) (O 6 3) (I 6 2) (U 4 3))
 ((I 8 1) (N N 8 2)) - -)
PHONEME REWRITING RULES FOR /E/
 (E -
 ((I - - 3 2) (N N D - 1 2 2) (U D - 1 2 2)
 (U N C 1 0 3) (N N S - 7 3) (I I K 2 0 3) (A D - 6 3) (U M - 7 3)
 (I A K 4 4) (I S K 4 4) (I M - 7 3) (U - - 3 4) (N N - - 4 4))
 ((E 8 1) (I 1 0 3)) ((U 6 2) (N N 5 2) (I 4 2)) - -)
PHONEME REWRITING RULES FOR /O/
 (O - ((A - - 1 1) (E Y - 4 1) (A - - 2 2)
 (N N Y - 6 2) (A Y - 7 2) (U Y - 1 2 2) (A K - 5 3)
 (U K D 5 3) (U T - 8 3) (E G Y 5 3) (A G Y 7 3) (U -
 K 6 4) (U N - 1 0 3) (A R S 5 3) (E T - 5 4) (A N - 4 3)) ((O
 8 1)) ((N N 7 1) (U 6 2) (I 6 5)) - ((N 1 0 4)))
PHONEME REWRITING RULES FOR /NN/
 (N N ((M 6 1) (M 3 2) (N 3 2) (G 3 2) (* 5 3))
 ((U I - 4 4) (U I I - 1 2 2) (I I - 9 3)
 (I I I - 1 0 3) (U E - 5 3)) - ((I 1 0 4)) ((M U 2) (N U 2)
 (M O 2) (O N N 3) (O U 3) (N N I 3) (N I 4) (N O 2)
 (Y U 2) (N N P 4) (R E 4)) -)
PHONEME REWRITING RULES FOR /II/
 (I I ((I 7 0) (I 3 2) (E 8 2) (N N 5 3)) ((U R N N 5 4))
 ((N N 9 2) (U 8 2)) - - - ((I * I 3) (I * U 4)))
PHONEME REWRITING RULES FOR /UU/
 (U U ((U 6 0) (U 3 2) (U 1 4) (N N 4 3))
 ((O Z Y - 1 5 1) (O K Y - 1 0 2) (U Z Y - 4 4) (E Z Y - 6 3))
 - - - -)
 ((U * U 2) (U * O 4) (I * U 4) (E * U 4)))
PHONEME REWRITING RULES FOR /EE/
 (E E ((E 6 0) (E 4 1)) - - - - ((E * E 2)))
PHONEME REWRITING RULES FOR /OO/
 (O O ((O 8 0) (O 4 3)) - - - -)
 ((O * O 2) (O * U 2)) (U * O 2))
PHONEME REWRITING RULES FOR /K/
 (K ((H A 4 2) (H - 3 2) (S - 3 2) (D - 4 3)
 (G - 3 3)) ((P - 1 4 1) (T - 1 4 1) (K K - 1 4 3)
 (K Y E 8 2)) - ((U Y P I Y 2)) ((* 3)) ((K T) (I U) (R *) 2 3))
PHONEME REWRITING RULES FOR /S/
 (S ((K - 2 1) (T - 2 2) (H Y - 2 2)
 (B I 3 2) (Z - 4 3) (Z E 3 3) (G I 2 3) (R I 2 4))
 ((P - 8 1) (K Y - 8 2) (H - 5 2))
 - - ((* 1)) ((S * T) (E U) (R) 8 4))
PHONEME REWRITING RULES FOR /T/
 (T ((K - 5 1) (H - 3 2) (K - 2 3) (D - 3 3)
 (S - 3 2) (H E 1 2)) ((P - 8 1)
) - - ((* 1)) ((T K) (I U) (Y * R) 2 2))
PHONEME REWRITING RULES FOR /H/
 (H ((* U 3 1) (N U 3 1) (W A 2 1) (B - 1 2) (M A 2 2
) (N A 1 3) (G - 1 2) (M U 1 1) (Z A 5 3) (* - 1 3))
 ((K - 8 2) (P - 8 2) (S - 8 2) (T I 6 2) (H Y U 8 3)) ((O A 1)
 (N N A 1))
 ((U N N H O Z 3) (U N N M O N 3))
 ((* 1)) -)
PHONEME REWRITING RULES FOR /P/
 (P ((T - 5 1) (K - 5 1) (H - 5 1) (S - 5 2) (B - 5 2)
) ((P P - 1 0 1))
 - - ((* 1)) -)
PHONEME REWRITING RULES FOR /N/
 (N ((M - 1 1) (D A 1 2) (G - 1 2) (Z I 1 3)
 (R - 1 2) (* - 1 3))
 ((T U 5 2) (N N - 1 1 3) (Y A 3 3)
 (H A 3 3) (T A 1 0 3)) ((N N - 1) (U I 2) (- A 3)
 (U U I 3) (I I 1)) ((E C * N N C 3)) ((* 3)) ((R) (N N) (N M) 2 3))

PHONEME REWRITING RULES FOR /M/
 (M ((N - 1 1) (R - 1 3) (W A 4 3) (* A 4 3))
 ((G - 5 2) (B - 5 2) (N N - 7 3) (H A 4 3))
 (Z 1 4 3))
 ((- A 3)) - - ((R B) (N N) (N M) 2 3))
 PHONEME REWRITING RULES FOR /Y/
 (Y ((W O 3 2) (N A 4 3) (R - 1 2) (N O 1 4))
 (* - 1 2) (E A 3 4))
 ((Z Y 0 1 1 1) (I A 6 3) (Z Y 0 0 1 1 1) (K Y 0 6 3) (K Y 0 0 1 0 3))
 (Z Y A 4 3) (N O 0 1 4) (D O 4 3) (H A 4 3) (U A 5 3))
 ((O O 3)) - -
 ((R H * B H Y) (E N N I) (* N R Y) 7 2))
 PHONEME REWRITING RULES FOR /R/
 (R - ((B - 4 2) (* O 5 2) (D - 5 2) (N - 6 2) (Y - 4 2) (Z - 3 2) (M - 5 3) (G - 4 3)))
 ((- - 3))
 - - -)
 PHONEME REWRITING RULES FOR /W/
 (W ((Z A 2 3) (N A 1 3) (* A 1 3)))
 ((U A 4 2) (H A 6 2) (H A 7 3) (O A 6 3))
 ((- A 2) (A - 2)) - - ((*) (O U) (*) 6 3))
 PHONEME REWRITING RULES FOR /G/
 (G ((R - 1 2) (D - 1 2) (Z - 1 2) (N - 1 2) (M - 2 1))
 (* - 3 4)) ((B - 8 2) (K - 6 2))
 (N N 0 0 1 0 3) (H - 5 3) (N N - 7 3) (U 0 0 6 2))
 ((N N - 3))
 ((U R * N N R 3) (U R H Y U R 3)))
 ((* 3)) -)
 PHONEME REWRITING RULES FOR /Z/
 (Z ((* 1 4 4)) ((G - 6 1) (R - 6 3) (B - 8 2) (N - 8 2))
 (M - 8 2) (D - 6 2) (H U 5 3) (S U 6 3))
 ((N N - 2)) - - -)
 PHONEME REWRITING RULES FOR /D/
 (D ((R - 2 2) (Z A 2 2) (M E 3 3)))
 ((B - 6 2) (G - 6 2) (T - 6 2))
 (H E 7 3)) - - ((* 5)) -)
 PHONEME REWRITING RULES FOR /B/
 (B ((R - 2 1) (D - 2 1) (G - 1 2) (Z I 2 2) (N N I L 2 2) (M - 2 3)) ((Y - 6 2) (P - 6 3) (K E 1 3)) - - - -)
)
 PHONEME REWRITING RULES FOR /KK/
 (K K ((K - 7 1) (T E 8 1)))
 ((P P - 1 5 1) (S - 1 0 4)) - - - -)
 PHONEME REWRITING RULES FOR /PP/
 (P P ((K K - 1 0 1) (P - 7 1) (K - 8 2)))
 - - ((U N N P O Z 3)) - -)
 PHONEME REWRITING RULES FOR /KY/
 (K Y ((H Y - 6 1) (S O 6 3) (Y 0 0 6 3) (K - 5 3) (Z Y - 6 3)))
 - - - ((* 3)) ((K Y T K S K K) (I E) (* G Y H Y R) 7 2))
)
 PHONEME REWRITING RULES FOR /ZY/
 (Z Y ((H Y U U 4 2) (K Y - 3 2) (G U U 4 3) (Z U U 4 3) (D U U 3 3))
 (Y - 3 3)) - -
 - - ((B N M Z C R) (I E N N) (Z Y R *) 7 2)))
 PHONEME REWRITING RULES FOR /HY/
 (H Y ((K Y - 1 1) (D A 5 2) (H A 1 2) (Z Y A 1 2) (Y A 3 3))
 (S A 4 3) (C A 3 3))
 ((N A 3 2) (E A 4 2)) - ((A K N A K 2))
 (A K N A G 2) (A K R A K 2)) - ((H S) (I E))
) (*) 4 2))
 PHONEME REWRITING RULES FOR /**/
 (** ((T - 6 3) (P P - 1 2 1) (K K - 1 2 1) (K - 6 3)))
 - - - -)
 PHONEME REWRITING RULES FOR INSERTION ERROR CORRECTION
 ((A (A - 4 2) (- A 4 2) (- - 2 3) (U R 4 3) (N O 3 3) (E N N 3 2)) (I (T E 3 2) (I - 4 2) (I - 8 4) (- 1 4 2) (I I - 1 2 0)) (U U - 5 2) (K E 4 2) (E - 3 3) (- Y 6 3) (- Z Y 6 1) (- - 2 3) (U M 3 3) (E Z 4 3)) (U (U - 4 2) (- U 4 2) (U U - 6 1) (O O - 6 1) (I H 5 3) (O Z Y 4 3) (I N 5 3) (I D 5 3) (- - 2 3) (O M 4 3) (* M 4 3) (N O 4 3) (S E 4 3)) (E (E - 4 2) (- E 4 2) (E E - 4 1) (A I 3 3) (I N N 4 3) (- - 2 3)) (I P P 3 3) (* A 3 3) (A S 3 3) (H Y A 3 1) (A T 3 3) (A Z 3 3) (U U - 4 3) (K A 4 3)) (O (O - 4 2) (- O 4 2) (O O - 9 1) (A G 4 3) (U - 3 3) (E N N 4 3) (- - 1 2) (U R 4 3) (U U - 6 1) (M A 3 3)) (N N (* - 2 1) (- - 3 3) (I - 4 3) (- Z 4 3)) (H (N N 0 4 1) (- - 1 2) (I E 2 3) (E - 3 3) (* - 4 3)) (M (- N N 4 2) (- - 2 1) (U U - 3 3)) (R (- I 2 1) (- - 4 2)) (G (- O 3 2) (- A 3 2) (- - 2 3)) (Z (- O 4 2)) (* - 6 2) (Z - 2 2)) (D (- - 1 2) (* - 9 3)) (B (- O 4 1) (* - 9 2) (- I 4 2)) (N (* - 1 0 1) (- - 4 3)) (Y (* - 3 3))))
 PHONEME REWRITING RULES FOR CORRECTION OF DEVOCALIZED VOWELS
 ((K S T H P) (I U) (K S T H P R -))
 LIST OF VOICED PHONEMES
 ((N M Y R W G Z D B Z Y * A I U E O N N) 5 3)

付録8. Voice Q-A System II-1における発声者別の認識実験結果

(単独に区切って発声した母音から推定した母音標準パターン使用)

(発声データ：付録3のグループ1の20文章)

発声者；RN

(予約項目)

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	O	-	X
1	Q	0 2	0 1	0 1	0 1	Q	0 2	2	5	0	0	0
2	Q	Q	0 1	Q		Q	Q	4	1	0	0	0
3	0 1	Q	Q		Q	Q	Q	5	1	0	0	0
4	0 1	0 2	Q		0 1	Q	0 1	2	4	0	0	0
5	Q	Q	0 1		Q	0 1	Q	4	2	0	0	0
6	0 2	0 1	0 1		0 1	Q	0 1	1	5	0	0	0
7		0 1	0 1	0 2		Q		1	3	0	0	0
8	0 1	0 3	0 1	0 1		Q		1	4	0	0	0
9	0 1	Q	0 1	0 1	0 1	Q		2	4	0	0	0
10		0 1	Q	0 1 3		Q	Q	3	2	0	0	0
11	0 1	0 1	Q		Q	Q	0 1	3	3	0	0	0
12	0 1	0 1	0 1	0 1		Q	Q	2	4	0	0	0
13	Q	Q	Q		Q	Q		6	0	0	0	0
14	0 1	Q		Q	Q	0 1	0 1	3	3	0	0	0
15	0 1	0 3	Q		Q	Q	0 1	3	3	0	0	0
16		0 1	0 2		Q	Q		2	2	0	0	0
17	0 1	0 1	Q	0 2	0 2	Q	Q	3	4	0	0	0
18	0 3	Q	0 1		0 1	0 1	0 1	1	5	0	0	0
19	0 1	0 3	0 1	0 2		Q		1	4	0	0	0
20	0 2	0 2	0 1	0 1	0 4	Q	Q	2	5	0	0	0

(発声文章番号)

TOTAL	(115)							TOTAL	
CORRECT	16	20	19	12	13	20	15	115	100.0%
UNIQUE	(3)	(7)	(7)	(3)	(6)	(17)	(8)	(51)	< 44.3 % >
=1	(10)	(7)	(11)	(5)	(5)	(3)	(6)	(47)	< 73.4 % >
>1	(3)	(6)	(1)	(4)	(2)	(0)	(1)	(17)	< 90.6 % >
REJECT	0	0	0	0	0	0	0	0	0.0 %
ERROR	0	0	0	0	0	0	0	0	0.0 %

Q：認識結果が一意で，結果が正しい。

O_i：認識結果が複数個あり，

その i 番目の候補が正しい。

-：拒絶

×：認識誤り

(Q + O) / TOTAL の百分率

Q / (Q + O) の百分率

O₁ / ΣO_i の百分率

(O₁ + O₂) / ΣO_i の百分率

(O₁ + O₂ + O₃) / ΣO_i の百分率

拒絶率

誤り率

発声者：HN

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	Q	Q	0 1	0 4	0 1	Q	4	3	0	0	0
2		0 1	0 1	0 1			Q	1	4	0	0	0
3	0 2	0 1	Q		0 1	Q	0 1	2	4	0	0	0
4	0 1	0 1	0 1		0 2	Q	0 1	1	5	0	0	0
5	0 1	0 2	0 1		Q	Q	0 1	2	4	0	0	0
6	0 2	0 1	0 1		Q	Q	0 1	2	4	0	0	0
7		Q	0 1	0 1		Q		2	2	0	0	0
8	0 1	0 1	0 1	0 1		Q		1	4	0	0	0
9	0 1	0 1	0 1	0 1	0 1	Q		1	5	0	0	0
10		Q	Q	0 3		Q	0 1	3	2	0	0	0
11	0 1	0 1	0 1		0 1	0 1	Q	1	5	0	0	0
12	Q	0 1	Q	0 1		Q	0 1	3	3	0	0	0
13	0 1	0 1	0 1	0 4	-	Q		1	4	1	0	0
14	Q	Q		Q		0 1	0 1	4	2	0	0	0
15	0 1	0 2	0 1		0 1	Q	Q	2	4	0	0	0
16		Q	0 1			Q	0 1	2	2	0	0	0
17	0 3	0 1	0 1	0 3	0 2	0 1	0 1	0	7	0	0	0
18	Q	Q	Q		0 1	0 1	0 1	3	3	0	0	0
19	0 1	0 1	0 2	0 1		Q		1	4	0	0	0
20	0 3	0 2	0 1	0 2	X	Q	0 1	1	5	0	1	0

TOTAL (115)								TOTAL	Q	0	-	X
CORRECT	16	20	19	12	13	20	15	113	98.3%			
UNIQUE	(4)	(6)	(5)	(1)	(3)	(15)	(3)	(37)	< 32.7 % >			
=1	(8)	(11)	(13)	(7)	(5)	(5)	(12)	(61)	< 80.3 % >			
>1	(4)	(3)	(1)	(4)	(3)	(0)	(0)	(15)	< 92.1 % >			
REJECT	0	0	0	0	1	0	0	1	0.9 %			
ERROR	0	0	0	0	1	0	0	1	0.9 %			

発声者：KI

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	Q	0 1	0 1	0 5	Q	0 1	3	4	0	0	0
2		Q	0 5	Q		0 1	0 1	2	3	0	0	0
3	0 1	0 2	Q		Q	Q	Q	4	2	0	0	0
4	0 1	0 1	Q		0 1	Q	0 1	2	4	0	0	0
5	0 1	Q	Q		Q	Q	0 1	4	2	0	0	0
6	Q	0 1	Q		Q	Q	0 1	4	2	0	0	0
7		0 1	0 1	0 1		Q		1	3	0	0	0
8	0 2	0 1	Q	0 1		Q		2	3	0	0	0
9	Q	Q	0 1	Q	Q	Q		5	1	0	0	0
10		0 2	0 2	0 5		Q	Q	2	3	0	0	0
11	0 1	0 1	0 1		0 1	0 1	0 1	0	6	0	0	0
12	0 1	0 1	Q	0 1		Q	0 1	2	4	0	0	0
13	0 1	0 1	Q	0 2	0 1	Q		2	4	0	0	0
14	0 1	Q		Q	Q	0 1	0 1	3	3	0	0	0
15	Q	0 2	0 1		Q	Q	0 1	3	3	0	0	0
16		Q	0 1			Q	Q	3	1	0	0	0
17	0 1	Q	Q	Q	Q	0 1	Q	5	2	0	0	0
18	0 1	Q	0 1		0 1	0 1	0 1	1	5	0	0	0
19	0 1	Q	0 2	0 1		Q		2	3	0	0	0
20	0 1	0 2	Q	0 1	0 2	Q	Q	3	4	0	0	0

TOTAL (115)								TOTAL	Q	0	-	X
CORRECT	16	20	19	12	13	20	15	115	100.0%			
UNIQUE	(4)	(9)	(9)	(4)	(7)	(15)	(5)	(53)	< 46.1 % >			
=1	(11)	(7)	(7)	(6)	(4)	(5)	(10)	(50)	< 80.6 % >			
>1	(1)	(4)	(3)	(2)	(2)	(0)	(0)	(12)	< 95.2 % >			
REJECT	0	0	0	0	0	0	0	0	0.0 %			
ERROR	0	0	0	0	0	0	0	0	0.0 %			

発声者：KS

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	0 1	Q	0 1	Q	Q	Q	0 1	4	3	0	0	0
2		Q	0 1	Q		Q	Q	4	1	0	0	0
3	0 1	Q	0 2		Q	Q	0 1	3	3	0	0	0
4	0 1	0 1	0 2		Q	Q	0 1	2	4	0	0	0
5	0 1	Q	Q		Q	Q	0 1	4	2	0	0	0
6	0 2	0 2	0 1		0 2	Q	0 1	1	5	0	0	0
7		Q	0 2	0 1		0 1		1	3	0	0	0
8	0 4	0 2	0 1	0 2		0 1		0	5	0	0	0
9	Q	Q	Q	0 4	0 2	Q		4	2	0	0	0
10		0 1	0 1	0 4		Q	Q	2	3	0	0	0
11	0 1	0 1	0 1		Q	0 2	0 1	1	5	0	0	0
12	0 1	0 2	0 2	0 2		Q	Q	2	4	0	0	0
13	0 1	0 2	0 1	0 7	0 2	Q		1	5	0	0	0
14	0 3	Q		-	0 1	Q	0 1	2	3	1	0	0
15	0 1	0 2	0 1		Q	Q	0 1	2	4	0	0	0
16		0 1	Q			Q	Q	3	1	0	0	0
17	0 4	0 1	Q	0 3	0 2	0 1	0 1	1	6	0	0	0
18	-	0 1	Q		0 1	Q	0 1	2	3	1	0	0
19	0 1	Q	0 1	0 1		Q		2	3	0	0	0
20	0 1	0 2	0 1	0 3	0 2	Q	0 1	1	6	0	0	0

TOTAL (115)								TOTAL	
	16	20	19	12	13	20	15	113	98.3%
CORRECT	15	20	19	11	13	20	15	(42)	< 37.2 % >
UNIQUE	(1)	(8)	(5)	(2)	(6)	(16)	(4)	(44)	< 62.0 % >
=1	(10)	(6)	(10)	(2)	(2)	(3)	(11)	(27)	< 88.7 % >
>1	(4)	(6)	(4)	(7)	(5)	(1)	(0)		< 93.0 % >
REJECT	1	0	0	1	0	0	0	2	1.7 %
ERROR	0	0	0	0	0	0	0	0	0.0 %

発声者：SA

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	0 1	Q	Q	Q	Q	Q	0 1	5	2	0	0	0
2		Q	0 1	0 2		Q	0 1	2	3	0	0	0
3	0 1	0 2	Q		0 1	Q	0 1	2	4	0	0	0
4	0 1	0 1	0 1		0 2	Q	0 1	1	5	0	0	0
5	0 1	X	X		Q	Q	0	3	1	0	2	0
6	0 1	0 2	Q		0 1	Q	0 2	2	4	0	0	0
7		Q	0 1	Q		Q		3	1	0	0	0
8	0 1	0 2	0 1	0 1		Q		1	4	0	0	0
9	Q	0 1	0 1	0 1	Q	Q		3	3	0	0	0
10		0 1	0 1	-		Q	Q	2	2	1	0	0
11	0 3	0 1	X		0 1	Q	0 1	1	4	0	1	0
12	0 2	0 2	Q	0 1		Q	0 1	2	4	0	0	0
13	0 1	Q	0 1	X	0 2	Q		2	3	0	1	0
14	0 1	0 1		Q	Q	Q	0 1	3	3	0	0	0
15	0 2	0 2	0 1		Q	Q	Q	3	3	0	0	0
16		Q	-			Q	Q	3	0	1	0	0
17	X	Q	Q	0 1	0 1	0 1	0 1	2	4	0	1	0
18	0 2	Q	0 1		0 1	0 1	0 1	1	5	0	0	0
19	0 1	Q	0 1	0 1		Q		2	3	0	0	0
20	0 2	0 2	X	X	0 1	Q	Q	2	3	0	2	0

TOTAL (115)								TOTAL	
	16	20	19	12	13	20	15	106	92.2%
CORRECT	15	19	15	9	13	20	15	(45)	< 42.5 % >
UNIQUE	(1)	(8)	(5)	(3)	(5)	(18)	(5)	(46)	< 75.4 % >
=1	(9)	(5)	(10)	(5)	(6)	(2)	(9)	(15)	< 98.4 % >
>1	(5)	(6)	(0)	(1)	(2)	(0)	(1)		< 100.0 % >
REJECT	0	0	1	1	0	0	0	2	1.7 %
ERROR	1	1	3	2	0	0	0	7	6.1 %

発声者：MK

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	0 2	Q	0 1	0 3	Q	0 1	3	4	0	0	0
2		0 2	0 1	0 2		0 1	0 1	0	5	0	0	0
3	0 1	0 1	0 1		Q	Q	0 1	2	4	0	0	0
4	0 1	0 1	-		-	Q	0 1	1	3	2	0	0
5	Q	Q	0 1		Q	Q	0 1	4	2	0	0	0
6	0 1	0 1	0 1		Q	Q	0 1	2	4	0	0	0
7		Q	0 3	Q		-		2	1	1	0	0
8	0 1	0 3	0 1	0 2		Q		1	4	0	0	0
9	Q	Q	0 1	0 4	0 1	Q		3	3	0	0	0
10		Q	Q	0 1		Q	Q	4	1	0	0	0
11	0 4	0 1	Q		Q	0 1	0 1	2	4	0	0	0
12	Q	0 2	0 1	0 1		Q	0 1	2	4	0	0	0
13	0 1	0 1	0 1	0 3	0 2	Q		1	5	0	0	0
14	0 1	Q		0 1	Q	Q	0 1	3	3	0	0	0
15	0 2	0 2	0 1		Q	Q	-	2	3	1	0	0
16		0 3	0 1			0 1	0 1	0	4	0	0	0
17	X	0 1	0 1	0 2	0 1	Q	0 1	1	5	0	1	0
18	0 1	0 1	Q		0 1	Q	0 1	2	4	0	0	0
19	X	Q	Q	0 1		Q		3	1	0	1	0
20	0 1	0 2	0 1	0 1	0 5	Q	0 1	1	6	0	0	0

TOTAL	(115)							TOTAL	
	16	20	19	12	13	20	15	109	94.8%
CORRECT	14	20	18	12	12	19	14	(39)	< 35.8 % >
UNIQUE	(4)	(6)	(5)	(1)	(6)	(16)	(1)	(52)	< 74.3 % >
=1	(8)	(7)	(12)	(6)	(3)	(3)	(13)	(18)	< 88.6 % >
>1	(2)	(7)	(1)	(5)	(3)	(0)	(0)		< 95.7 % >
REJECT	0	0	1	0	1	1	1	4	3.5 %
ERROR	2	0	0	0	0	0	0	2	1.7 %

発声者：NA

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	0 1	0 1	0 1	0 3	Q	Q	3	4	0	0	0
2		Q	0 1	0 1		Q	0 1	2	3	0	0	0
3	0 2	0 2	Q		Q	Q	0 1	3	3	0	0	0
4	0 1	0 1	0 1		0 1	Q	Q	2	4	0	0	0
5	0 1	Q	Q		Q	Q	X	4	1	0	1	0
6	0 1	0 1	0 1		0 1	Q	0 1	1	5	0	0	0
7		Q	0 3	0 1		-		1	2	1	0	0
8	X	0 1	Q	0 1		0 1		1	3	0	1	0
9	0 1	Q	0 1	Q	Q	Q		4	2	0	0	0
10		0 1	Q	0 2		Q	Q	3	2	0	0	0
11	0 1	0 2	0 1		0 1	0 1	Q	1	5	0	0	0
12	0 1	0 1	Q	0 1		Q	Q	3	3	0	0	0
13	0 2	0 3	0 1	0 4	0 2	Q		1	5	0	0	0
14	0 1	Q		0 3	Q	Q	0 2	3	3	0	0	0
15	0 1	0 2	0 1		0 1	Q	Q	2	4	0	0	0
16		0 2	0 1			Q	0 1	1	3	0	0	0
17	0 1	Q	Q	0 1	X	Q	0 1	3	3	0	1	0
18	0 3	0 1	0 1		0 1	0 1	Q	1	5	0	0	0
19	0 1	X	0 1	0 1		Q		1	3	0	1	0
20	0 2	0 1	Q	-	0 2	Q	0 1	2	4	1	0	0

TOTAL	(115)							TOTAL	
	16	20	19	12	13	20	15	109	94.8%
CORRECT	15	19	19	11	12	19	14	(42)	< 38.5 % >
UNIQUE	(1)	(6)	(7)	(1)	(4)	(16)	(7)	(50)	< 74.6 % >
=1	(10)	(8)	(11)	(7)	(5)	(3)	(6)	(17)	< 91.0 % >
>1	(4)	(5)	(1)	(3)	(3)	(0)	(1)		< 98.5 % >
REJECT	0	0	0	1	0	1	0	2	1.7 %
ERROR	1	1	0	0	1	0	1	4	3.5 %

発声者：KH

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	0 1	Q	0 1	0 1	Q	0 1		3	4	0	0
2		Q	0 1	Q		0 1	0 1		2	3	0	0
3	0 1	0 2	0 1		Q	Q	Q		3	3	0	0
4	0 1	0 1	0 1		0 1	Q	0 1		1	5	0	0
5	0 1	X	X		Q	0 1	0 1		1	3	0	2
6	0 1	0 1	0 1		0 1	Q	0 1		1	5	0	0
7		0 1	0 1	0 1		Q			1	3	0	0
8	0 2	0 3	0 1	0 1		Q			1	4	0	0
9	0 1	Q	Q	0 2	0 4	Q			3	3	0	0
10		0 1	-	0 2		Q	Q		2	2	1	0
11	X	Q	0 1		0 1	0 2	0 1		1	4	0	1
12	0 1	0 1	Q	0 1		Q	0 1		2	4	0	0
13	Q	0 1	Q	0 4	0 2	Q			3	3	0	0
14	0 1	Q		Q	Q	0 2	0 1		3	3	0	0
15	0 1	0 3	0 1		0 1	Q	0 1		1	5	0	0
16		Q	-			Q	Q		3	0	1	0
17	0 2	0 1	Q	0 1	0 2	0 2	0 1		1	6	0	0
18	0 1	Q	Q		0 3	0 1	0 1		2	4	0	0
19	0 3	0 3	0 2	0 1		Q			1	4	0	0
20	0 2	0 2	0 2	0 2	0 2	Q	0 1		1	6	0	0

TOTAL (115)								TOTAL	Q	0	-	X
	16	20	19	12	13	20	15	110	95.7%			
CORRECT	15	19	16	12	13	20	15	(36)	< 32.7 % >			
UNIQUE	(2)	(6)	(6)	(2)	(3)	(14)	(3)	(51)	< 68.9 % >			
=1	(9)	(8)	(8)	(6)	(5)	(3)	(12)	(23)	< 90.5 % >			
>1	(4)	(5)	(2)	(4)	(5)	(3)	(0)		< 97.3 % >			
REJECT	0	0	2	0	0	0	0	2	1.7 %			
ERROR	1	1	1	0	0	0	0	3	2.6 %			

発声者：MS

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	0 3	X	0 1	0 2	0 2	Q	0 1		1	5	0	1
2		0 2	0 2	0 1		Q	0 1		1	4	0	0
3	0 1	0 3	Q		Q	Q	Q		4	2	0	0
4	0 1	0 1	0 1		0 1	Q	0 1		1	5	0	0
5	0 1	0 2	X		Q	Q	Q		3	2	0	1
6	0 2	0 2	X		0 1	Q	0 1		1	4	0	1
7		Q	0 3	Q		0 1			2	2	0	0
8	0 1	0 2	0 1	0 1		0 1			0	5	0	0
9	Q	0 1	0 1	0 3	0 1	Q			2	4	0	0
10		0 1	Q	0 5		Q	Q		3	2	0	0
11	0 2	0 1	0 1		0 1	0 1	Q		1	5	0	0
12	0 1	0 1	0 6	0 1		Q	0 2		1	5	0	0
13	0 1	0 1	0 1	0 3	0 1	Q			1	5	0	0
14	0 1	0 1		0 6	Q	Q	0 1		2	4	0	0
15	0 2	0 2	0 1		0 1	Q	0 1		1	5	0	0
16		X	0 1			Q	Q		2	1	0	1
17	0 1	0 1	0 1	0 2	-	0 2	0 1		0	6	1	0
18	0 4	0 3	0 3		0 1	0 1	0 1		0	6	0	0
19	0 1	0 2	0 2	0 3		Q			1	4	0	0
20	0 1	0 2	0 1	0 2	-	Q	0 1		1	5	1	0

TOTAL (115)								TOTAL	Q	0	-	X
	16	20	19	12	13	20	15	109	94.8%			
CORRECT	16	18	17	12	11	20	15	(28)	< 25.7 % >			
UNIQUE	(1)	(1)	(2)	(1)	(3)	(15)	(5)	(51)	< 63.0 % >			
=1	(10)	(8)	(10)	(3)	(7)	(4)	(9)	(30)	< 85.2 % >			
>1	(5)	(9)	(5)	(8)	(1)	(1)	(1)		< 95.1 % >			
REJECT	0	0	0	0	2	0	0	2	1.7 %			
ERROR	0	2	2	0	0	0	0	4	3.5 %			

付録9. 音声データのLPCスペクトル包絡の系列および セグメンテーションに用いる基本特徴量の変化の図示

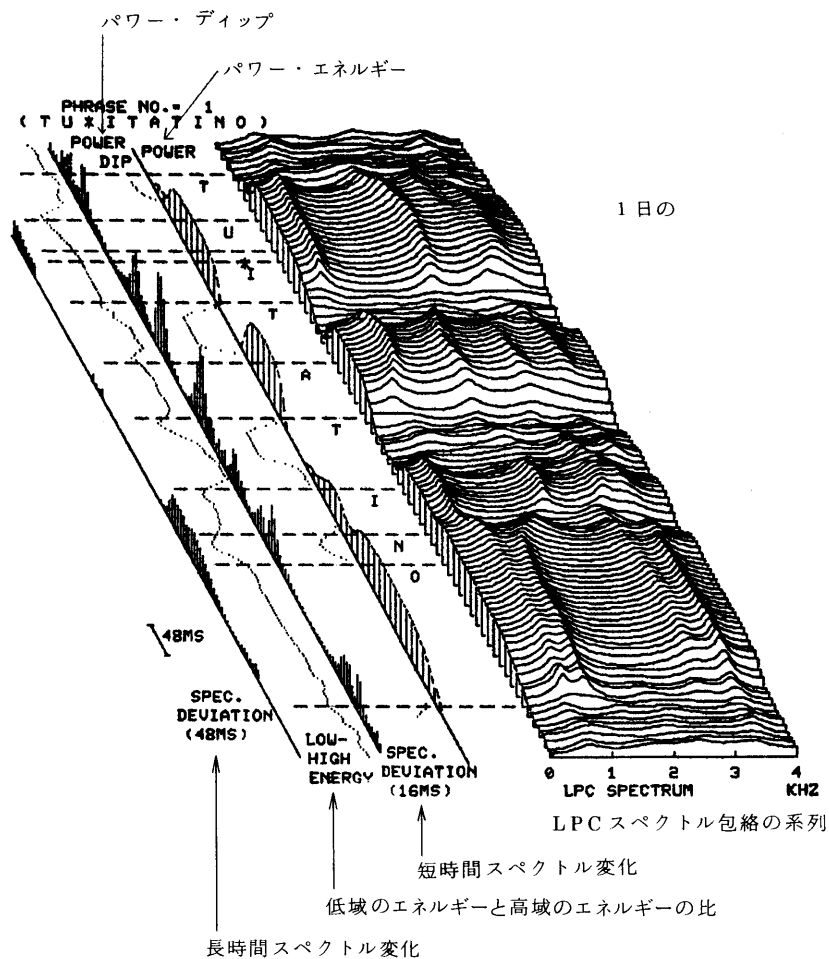
発声者：男性4名(KI, KS, RN, KH)

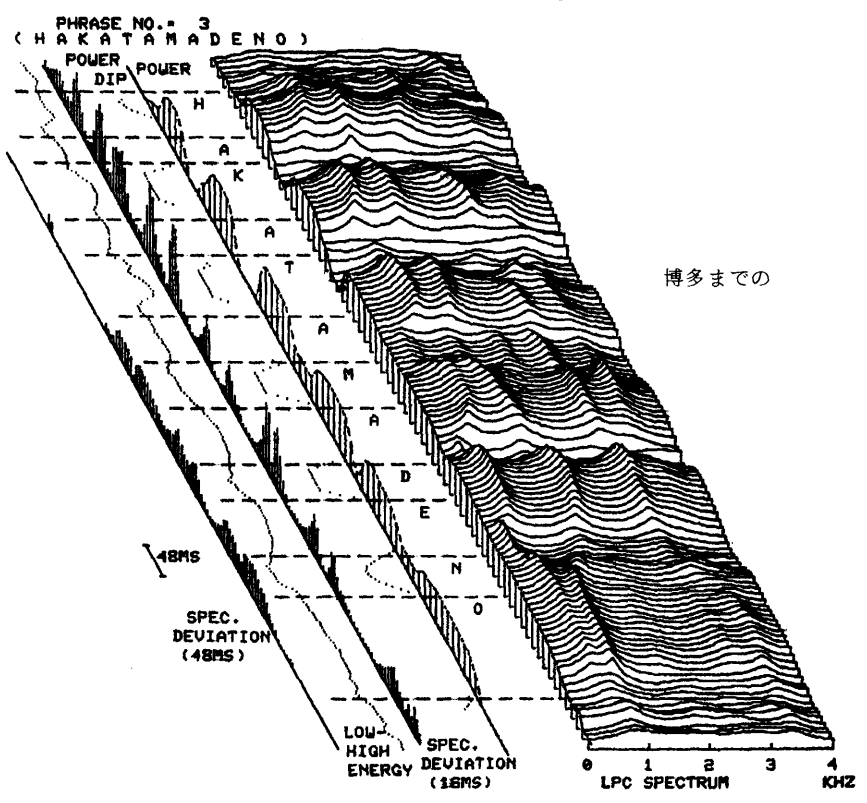
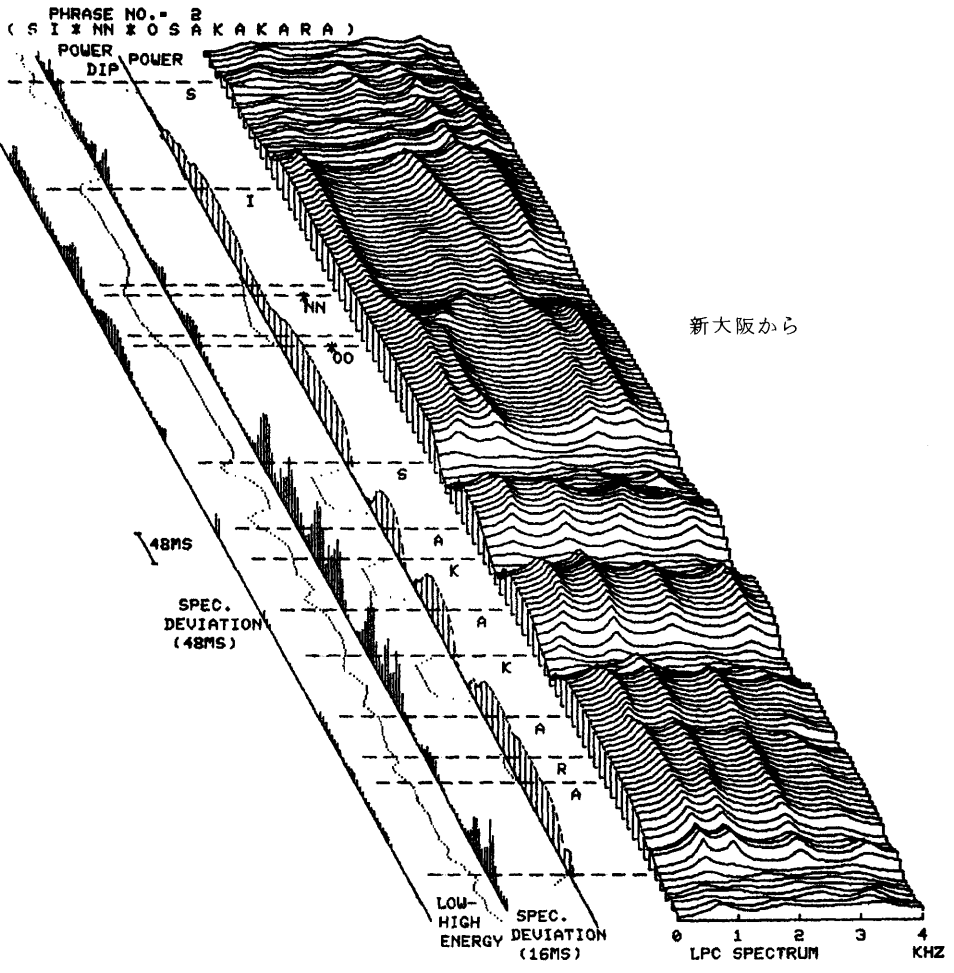
発声文章：○1日の, 新大阪から, 博多までの, 6時2分発の,
ひかり19号の, 指定券を, 9枚, お願いします。

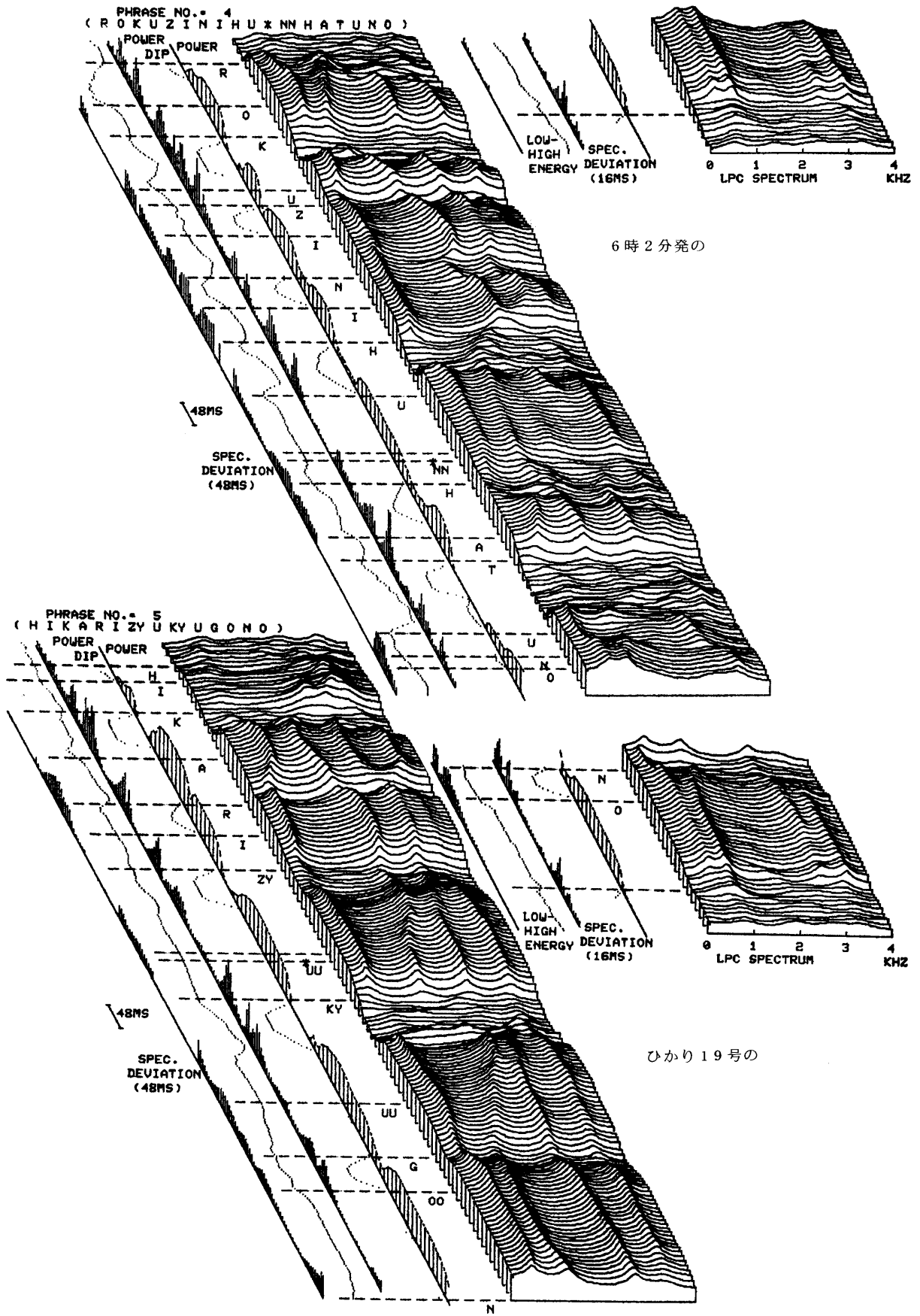
○新神戸発の, 6時22分の, グリーン券を, 小倉まで, 8枚。

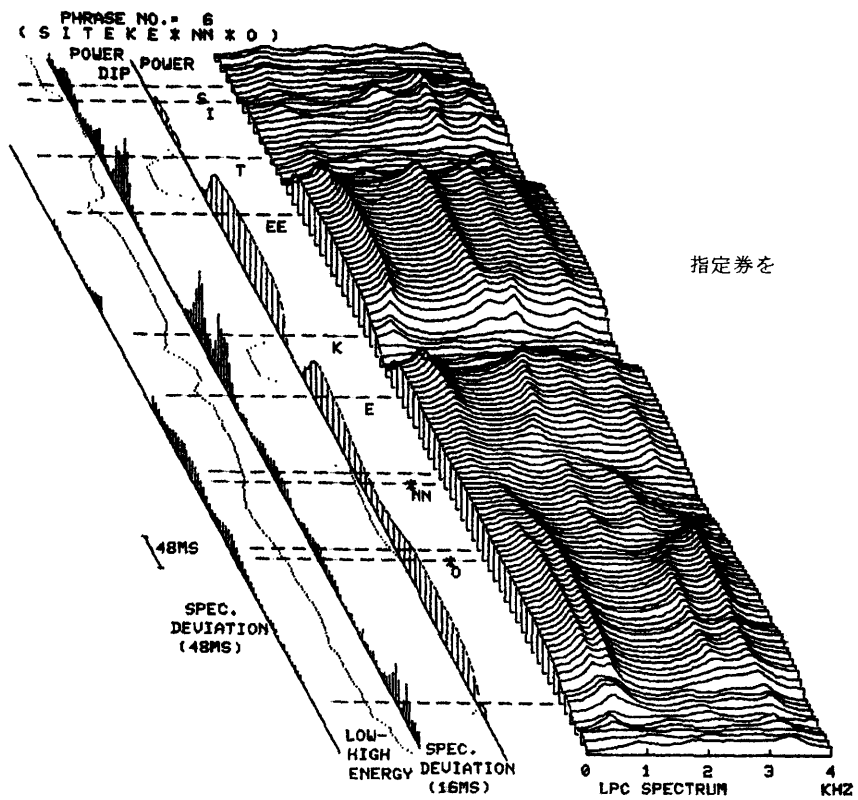
分析：4KHz低域フィルタ, 8KHzサンプリング, 32msecハミング窓,
スペクトルの平坦化, 8msecステップの10次のLPC分析。

発声者：KI

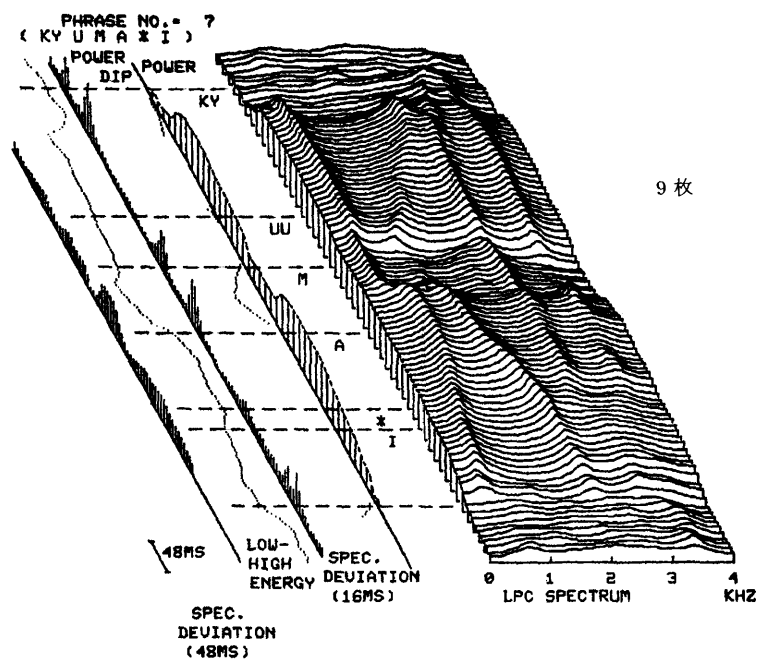




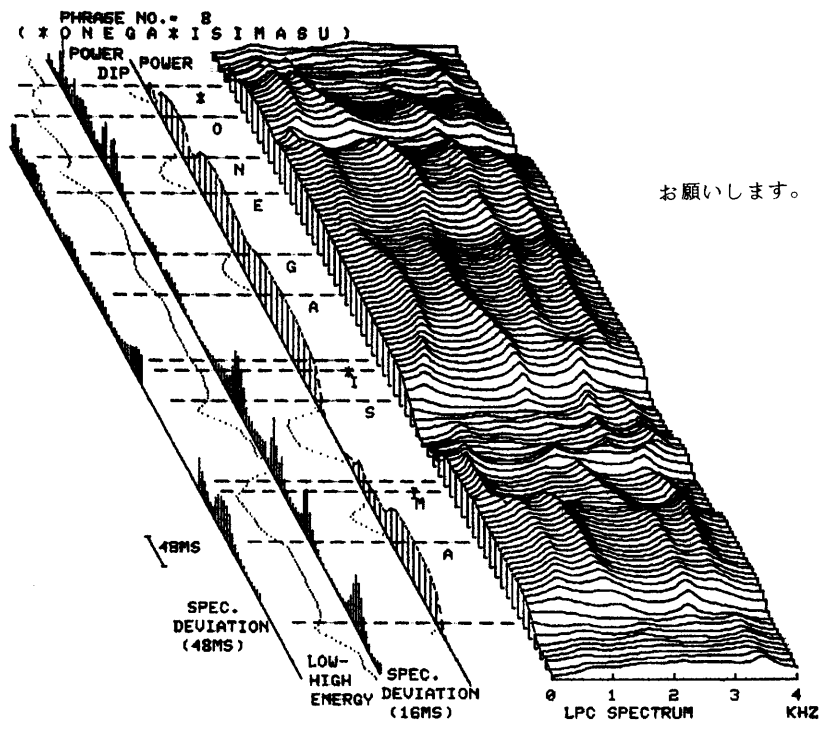




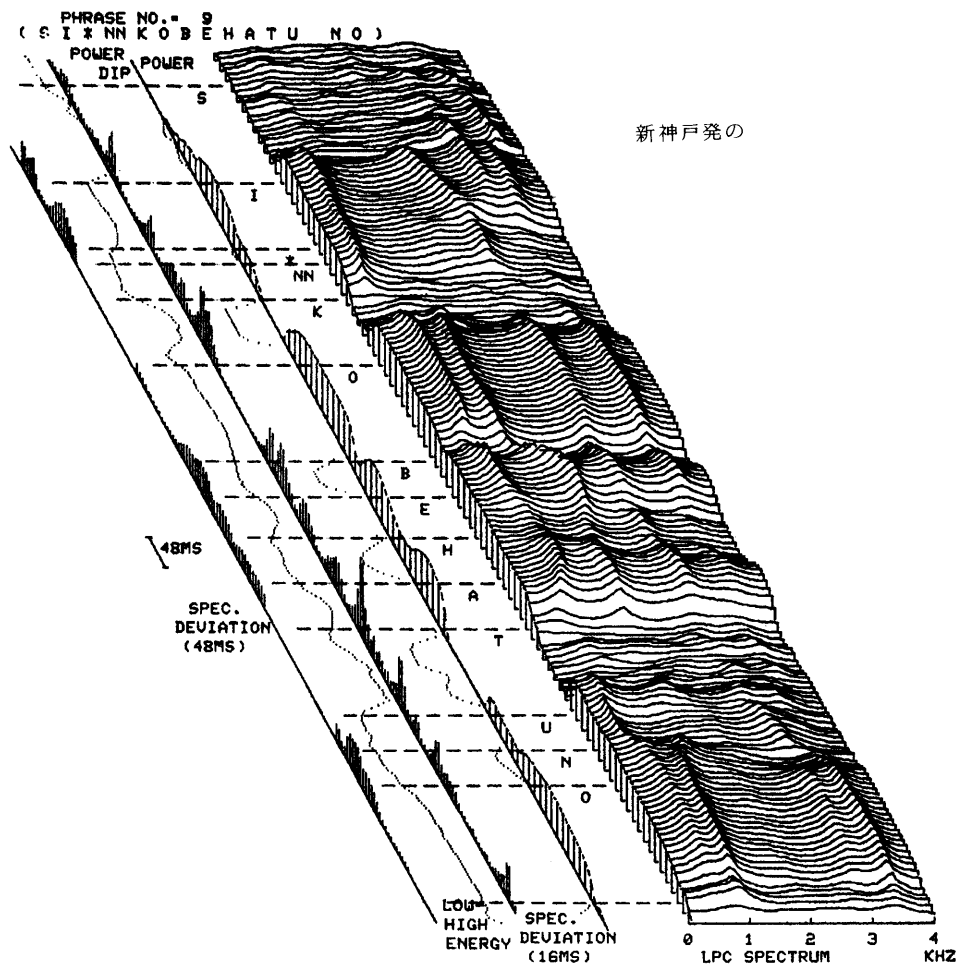
指定券を



9枚

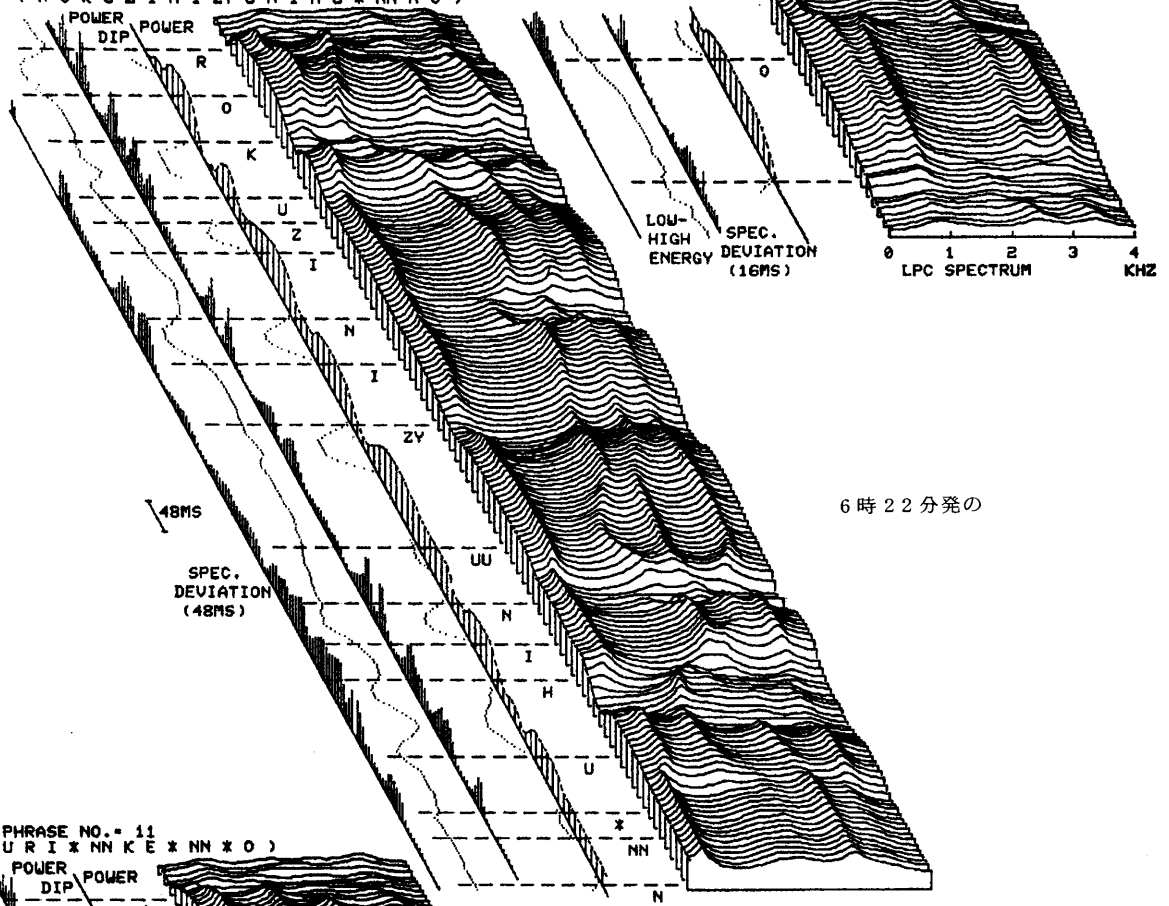


お願いします。

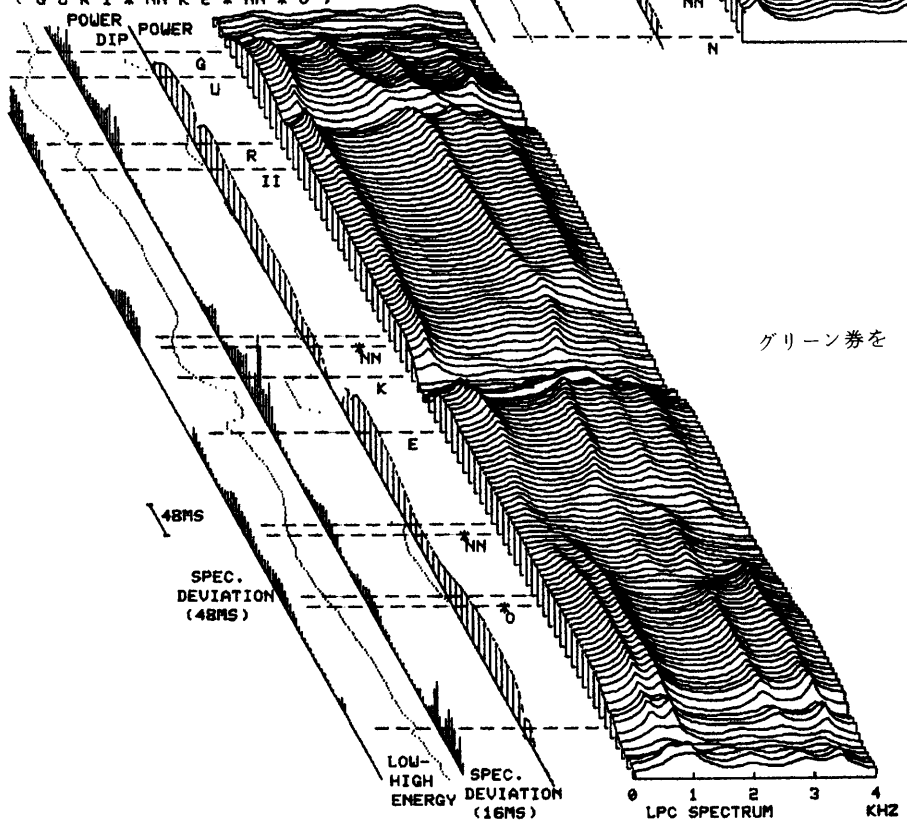


新神戸発の

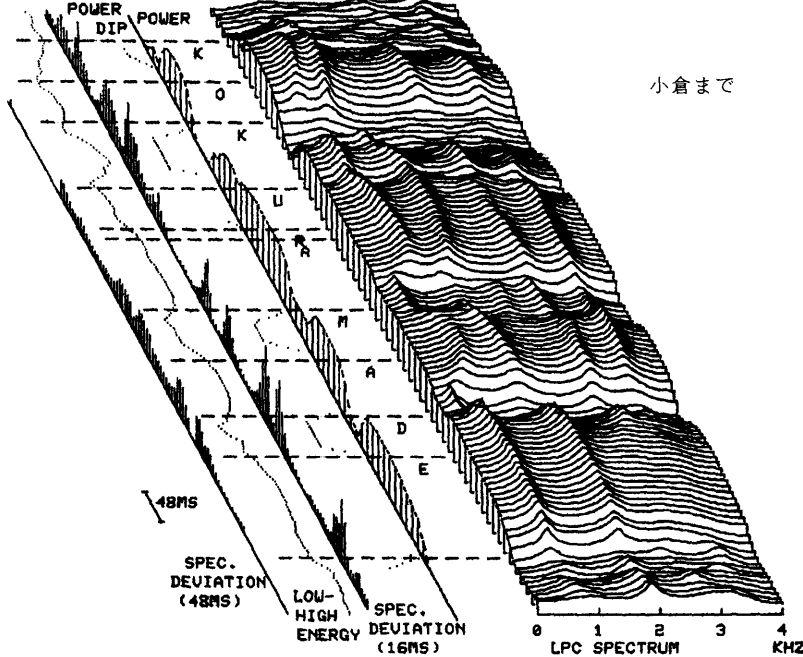
PHRASE NO. 10
 (R O K U Z I N I Z Y U N I H U * N N O)



PHRASE NO. 11
 (G U R I * N N K E * N N * O)

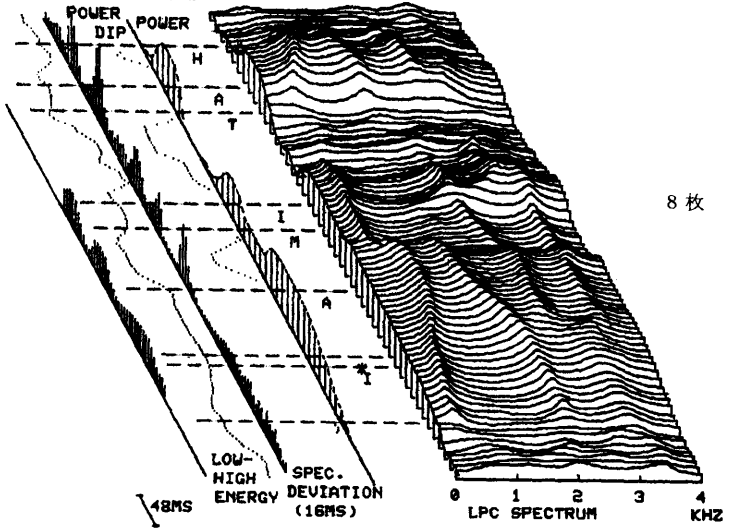


PHRASE NO. = 12
(K O K U R A M A D E)



小倉まで

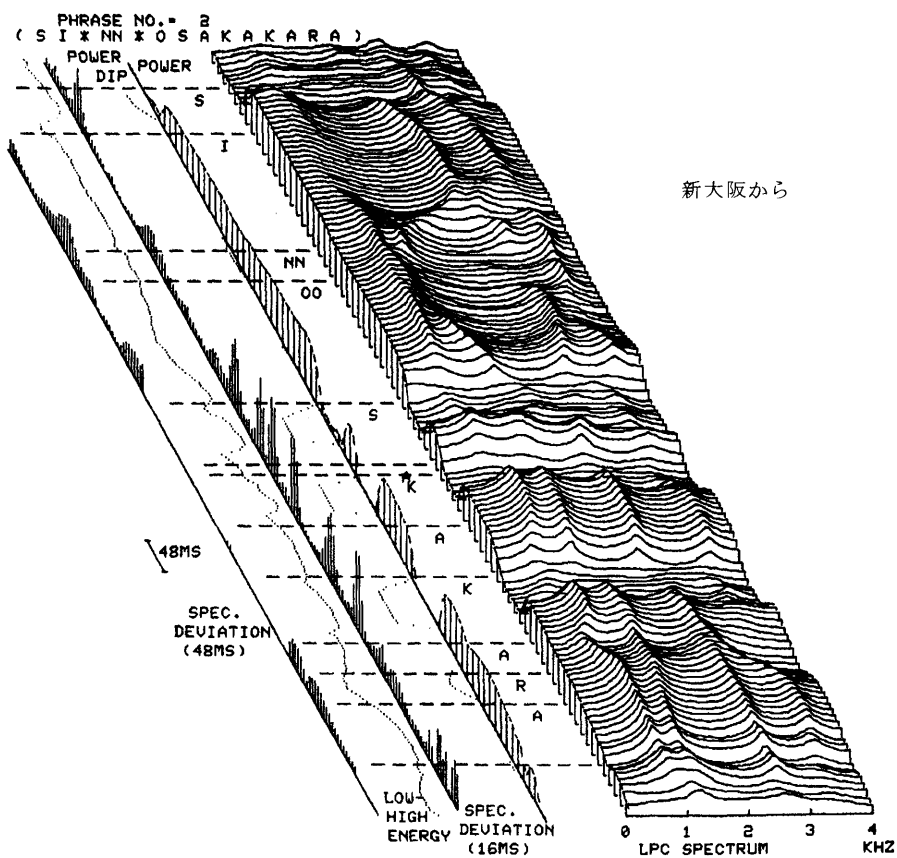
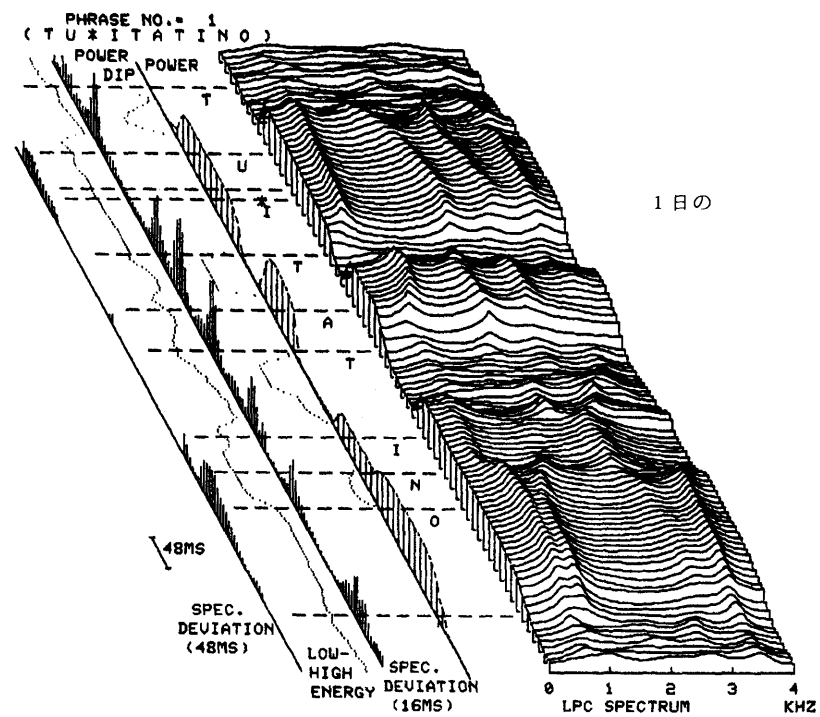
PHRASE NO. = 13
(H A T I M A X I)

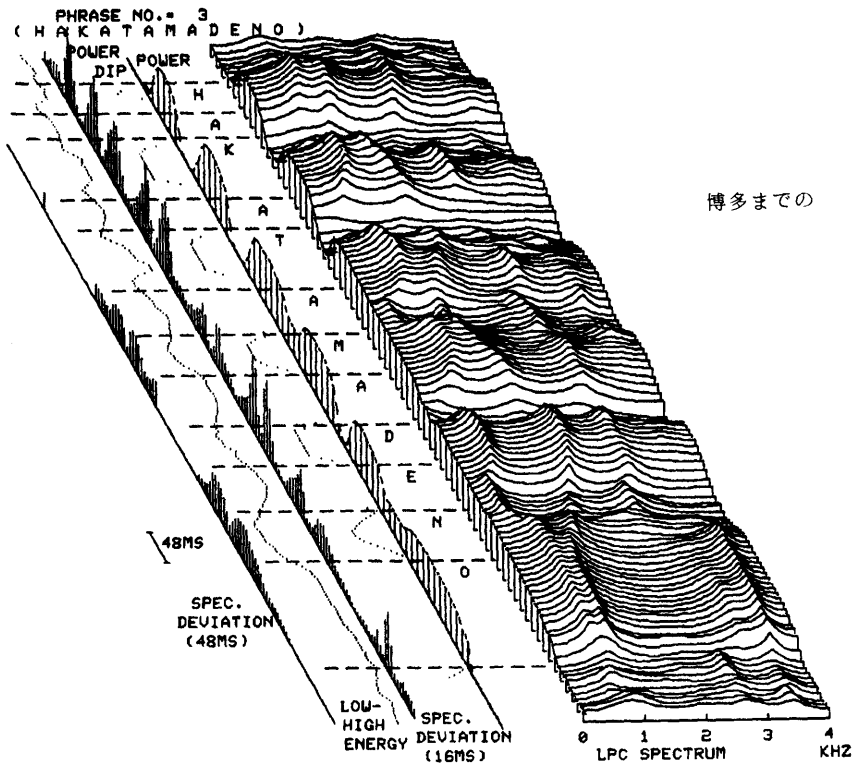


8枚

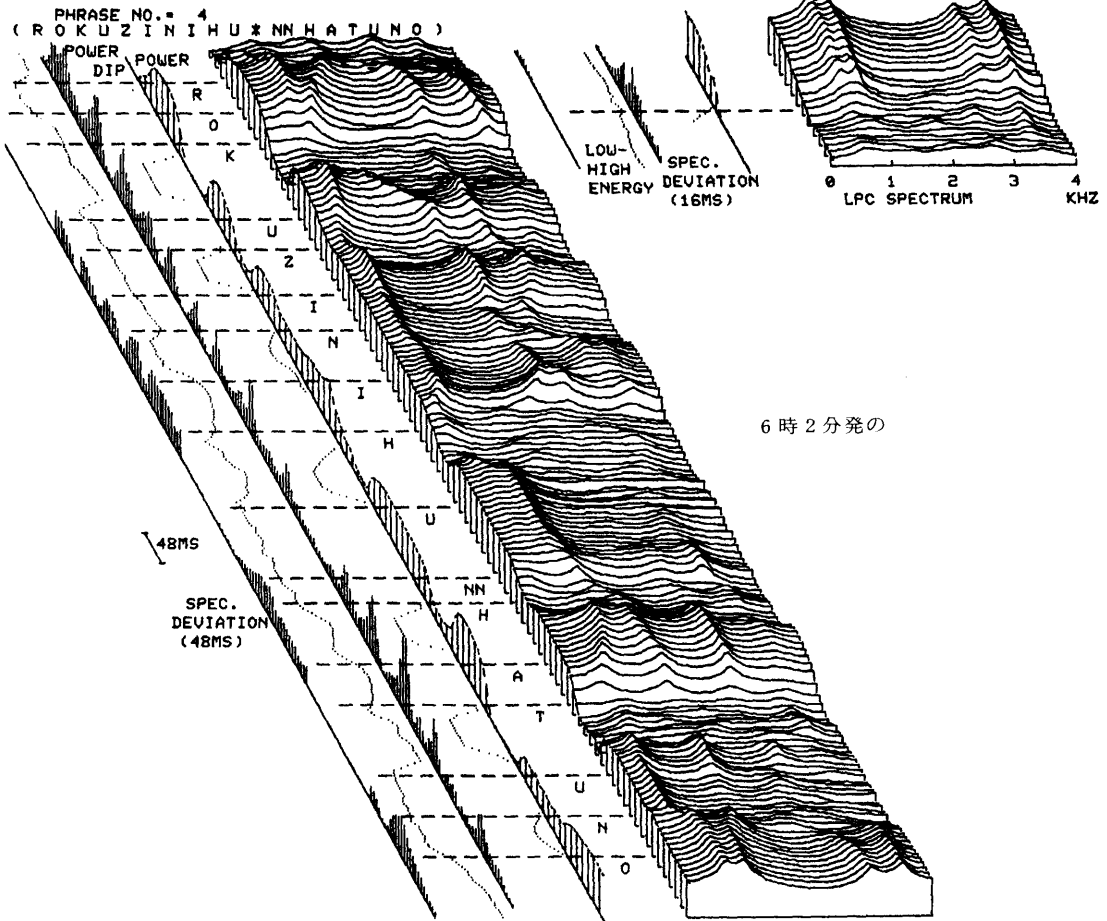
SPEC. DEVIATION (48MS)

発声者：KS

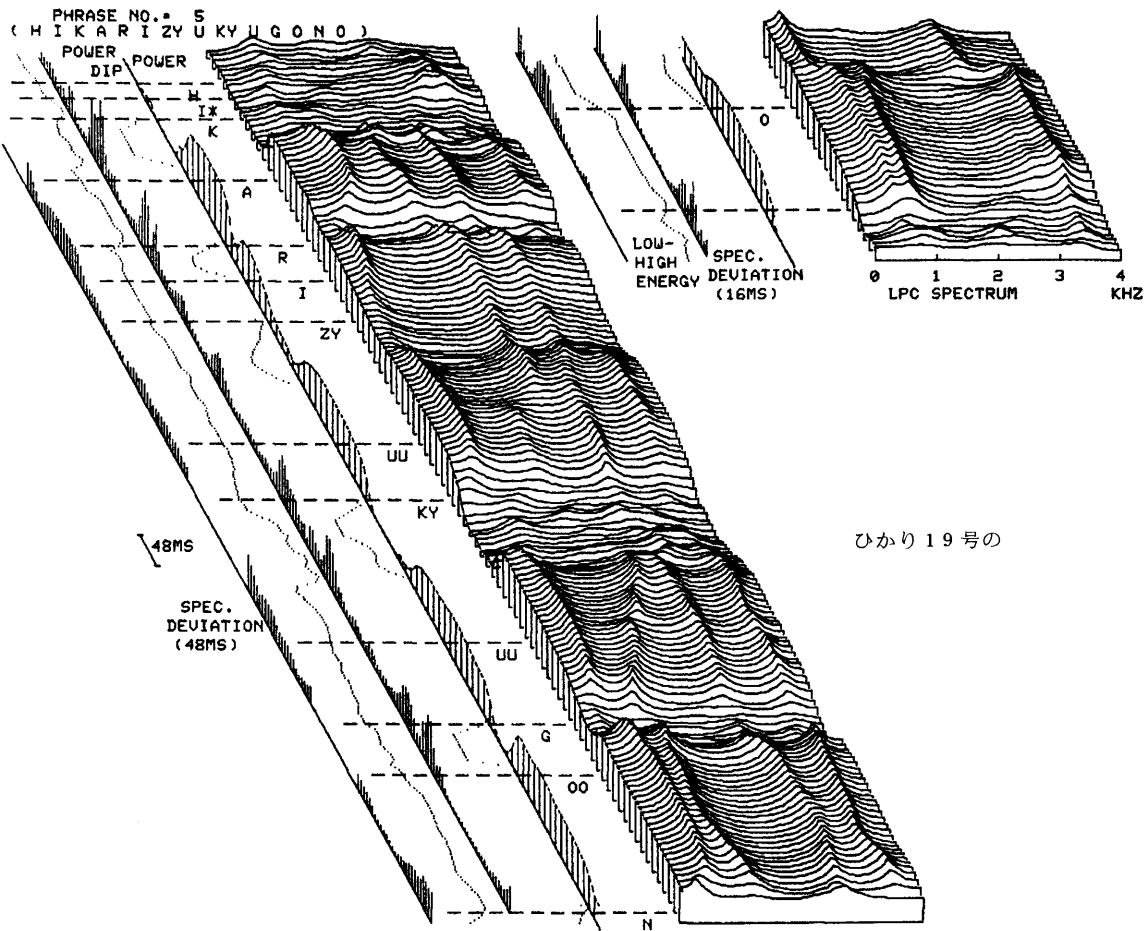




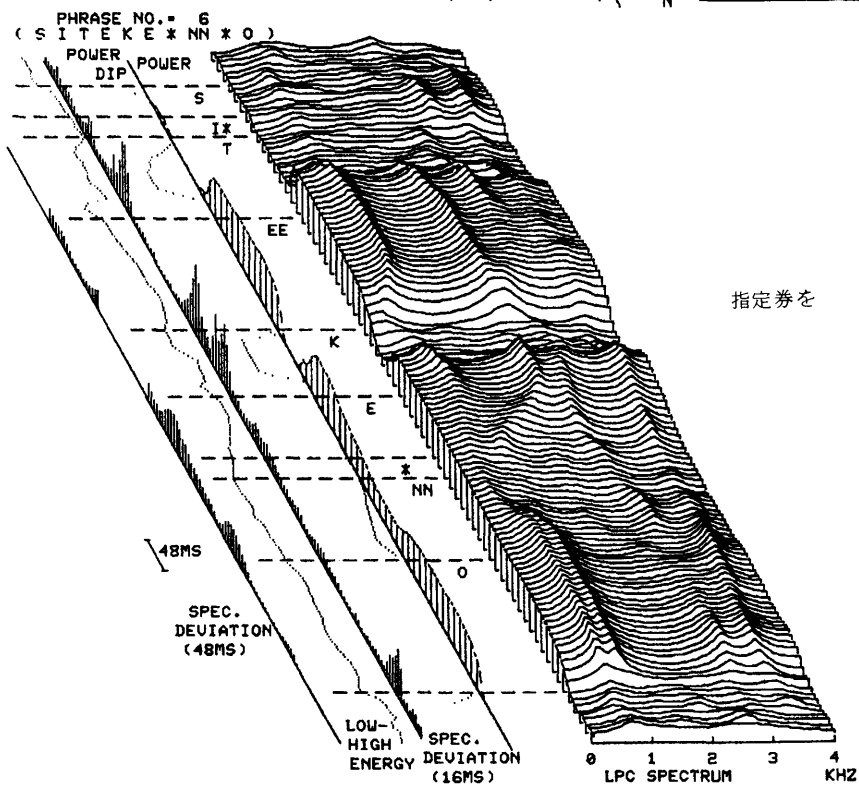
博多までの



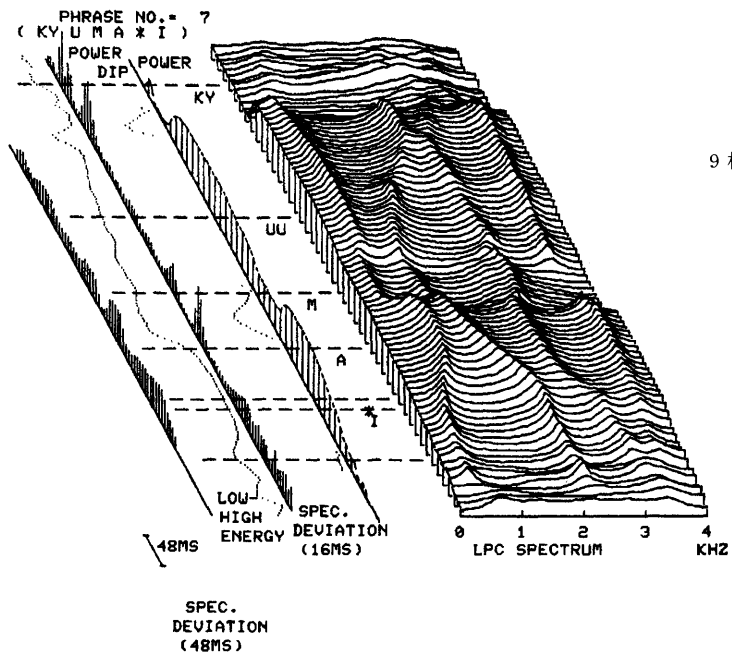
6時2分発の



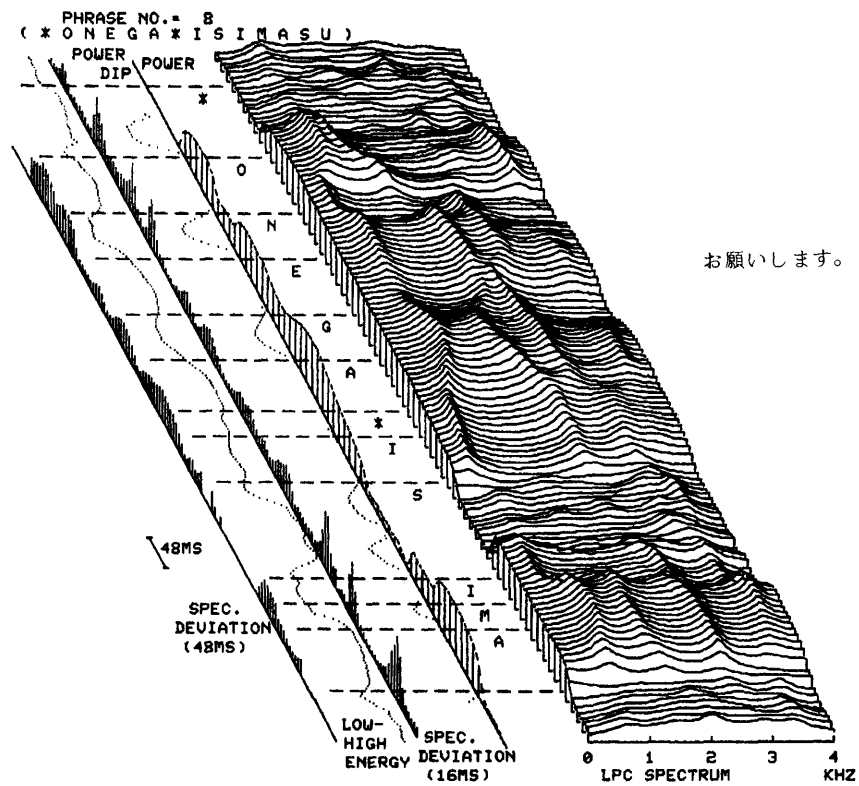
ひかり19号の



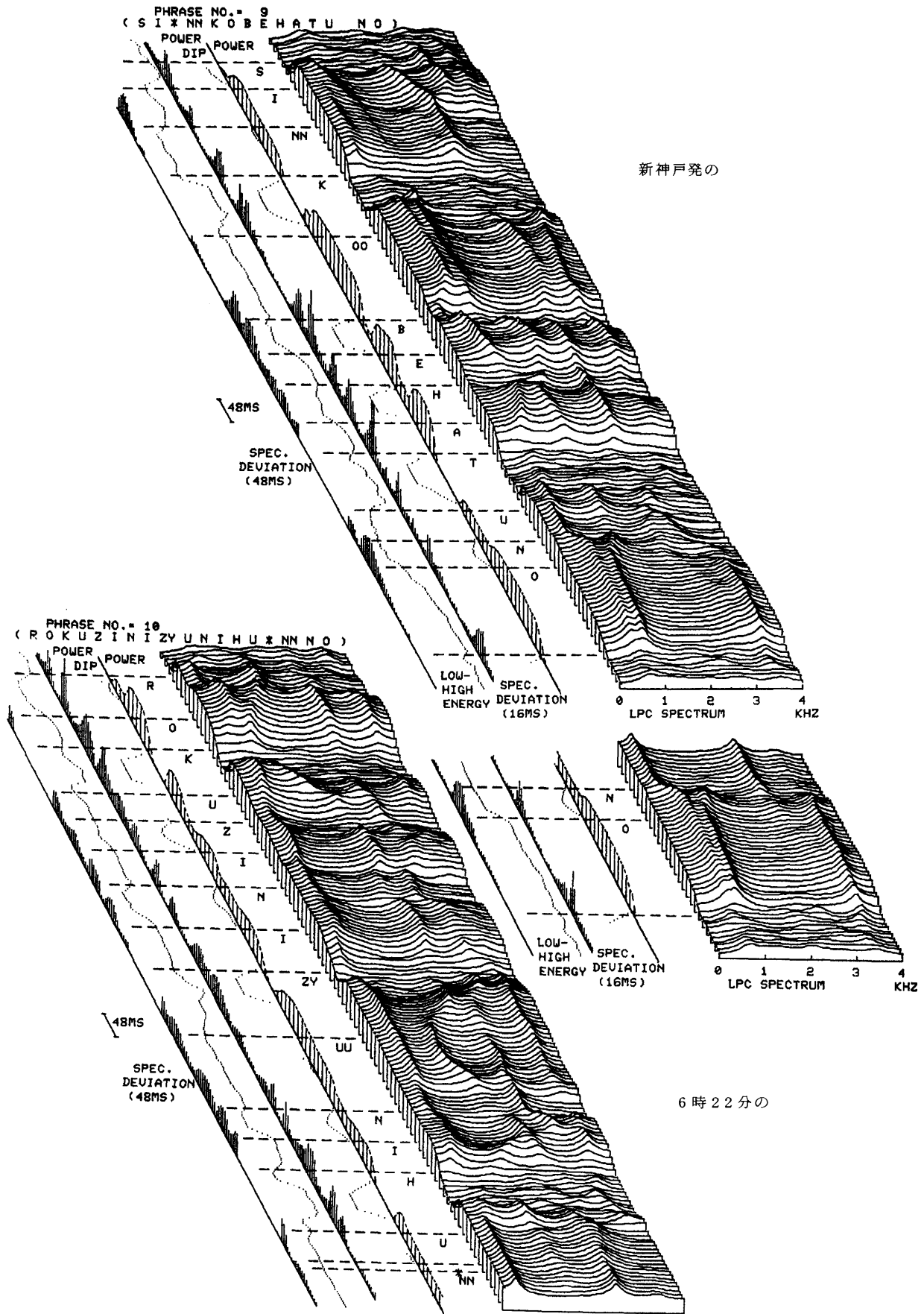
指定券を



9枚

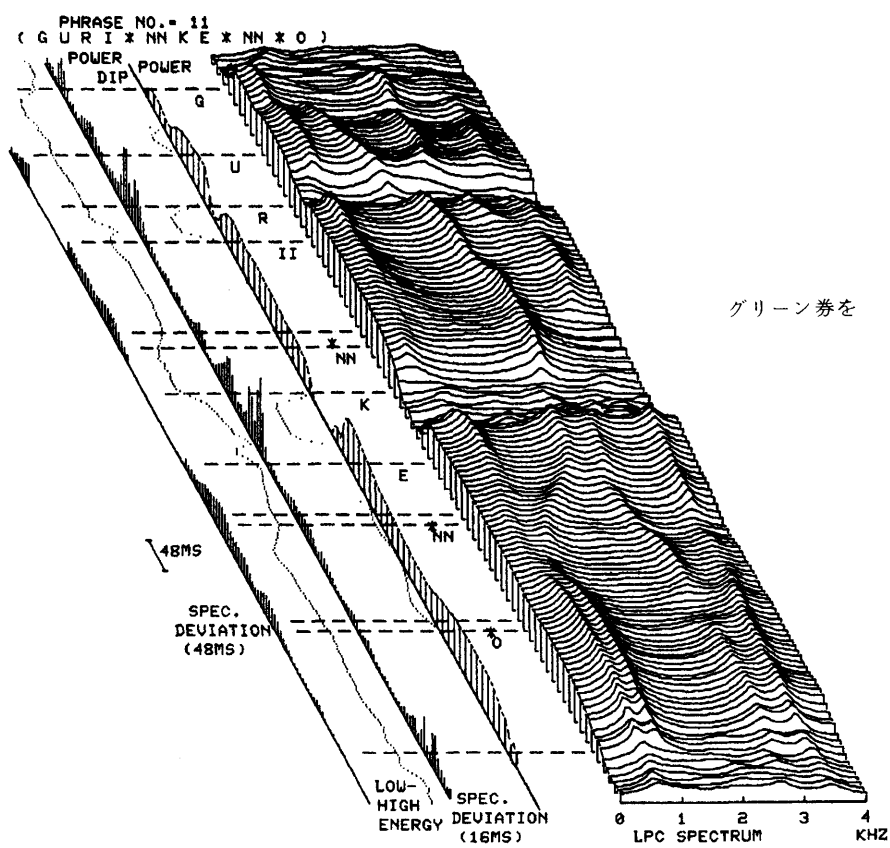


お願いします。

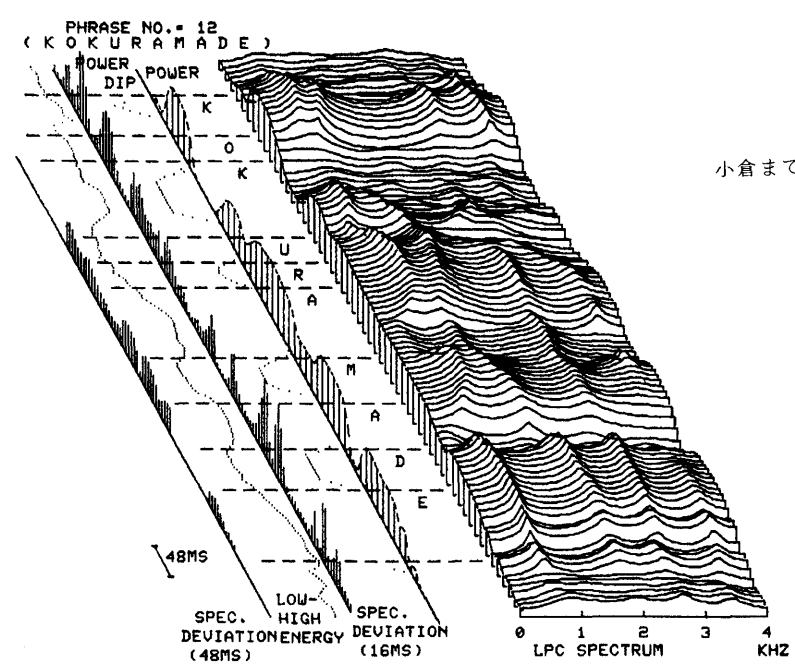


新神戸発の

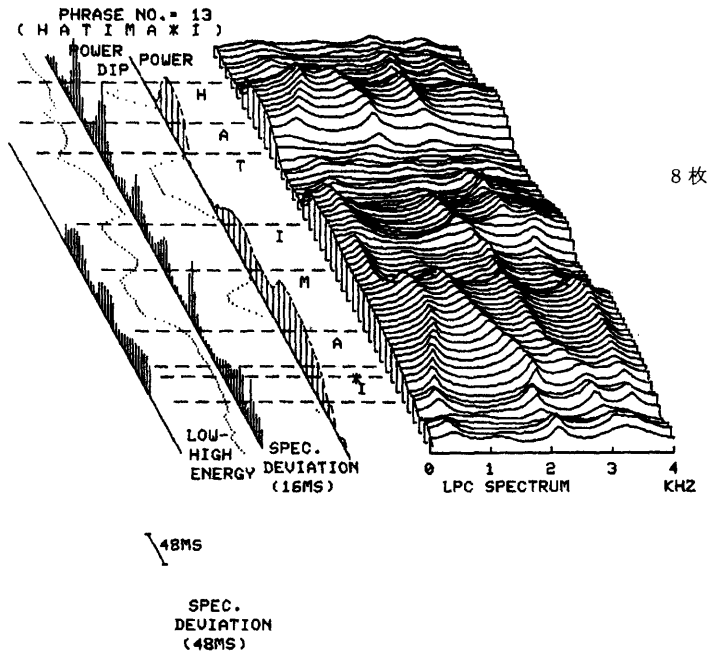
6時22分の



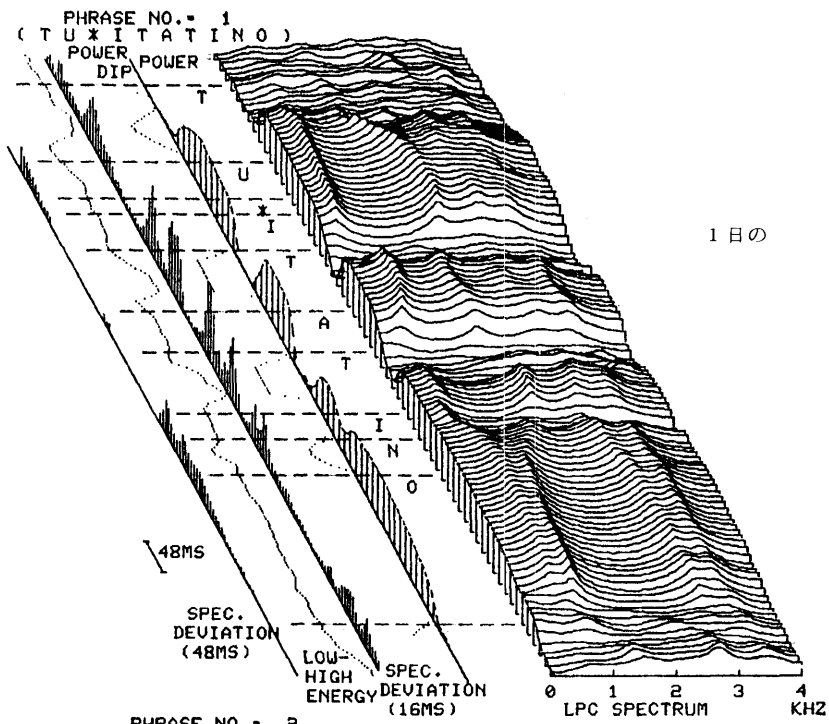
グリーン券を



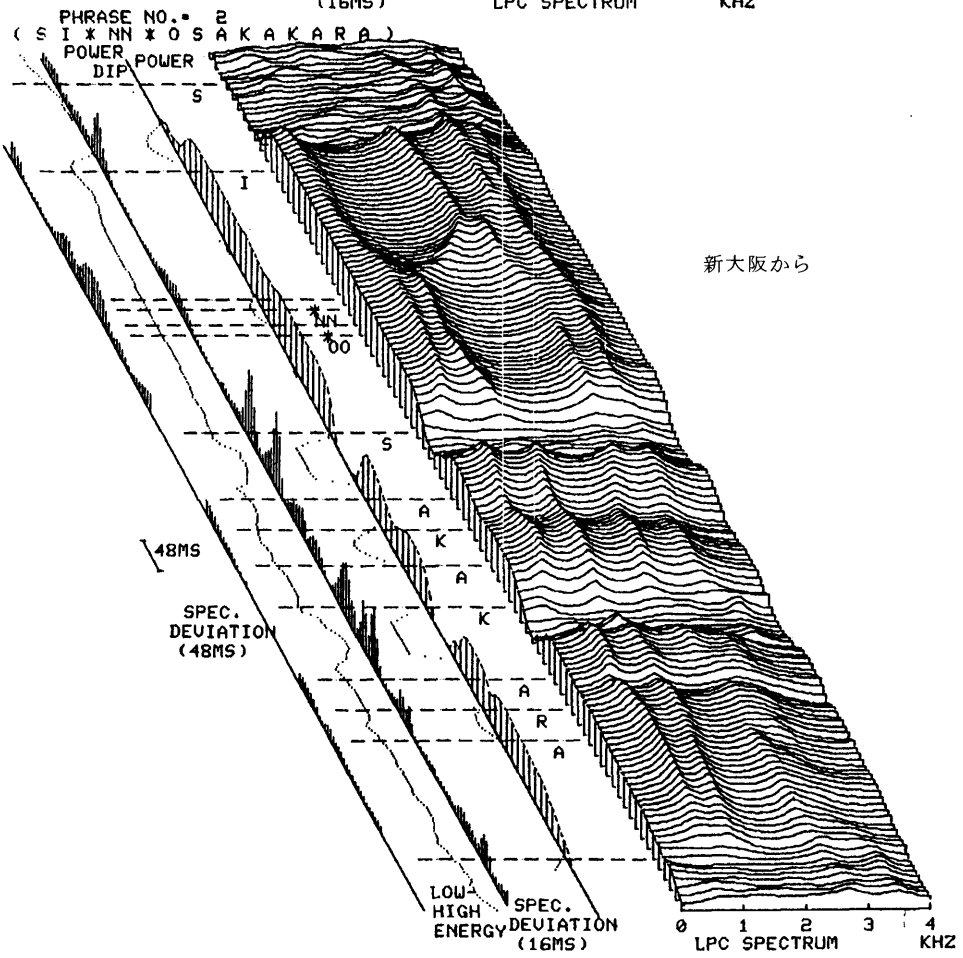
小倉まで



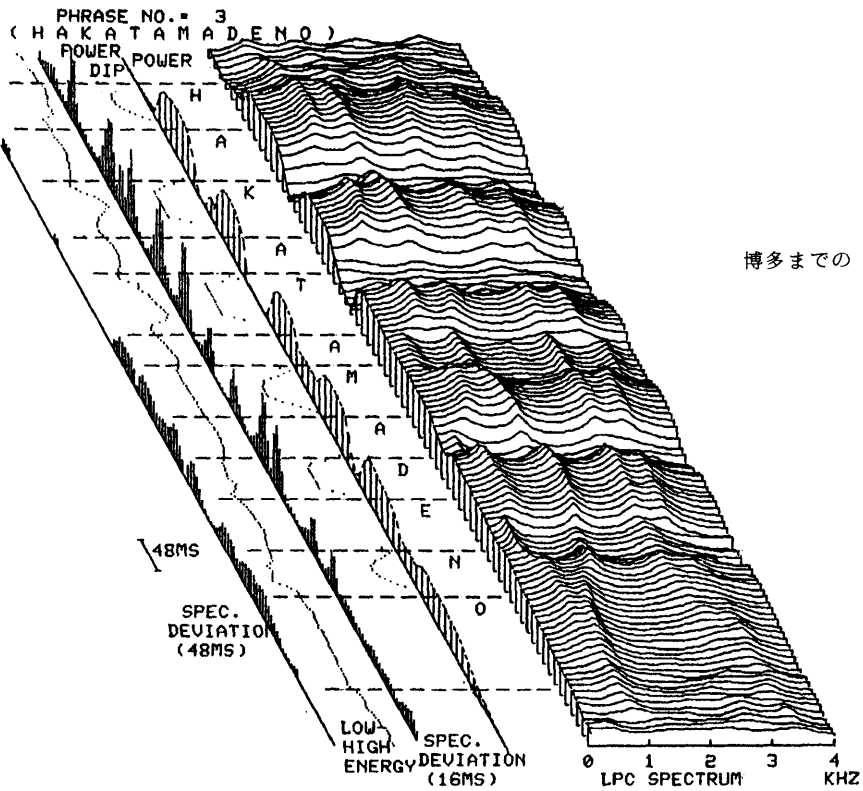
発声者：RN



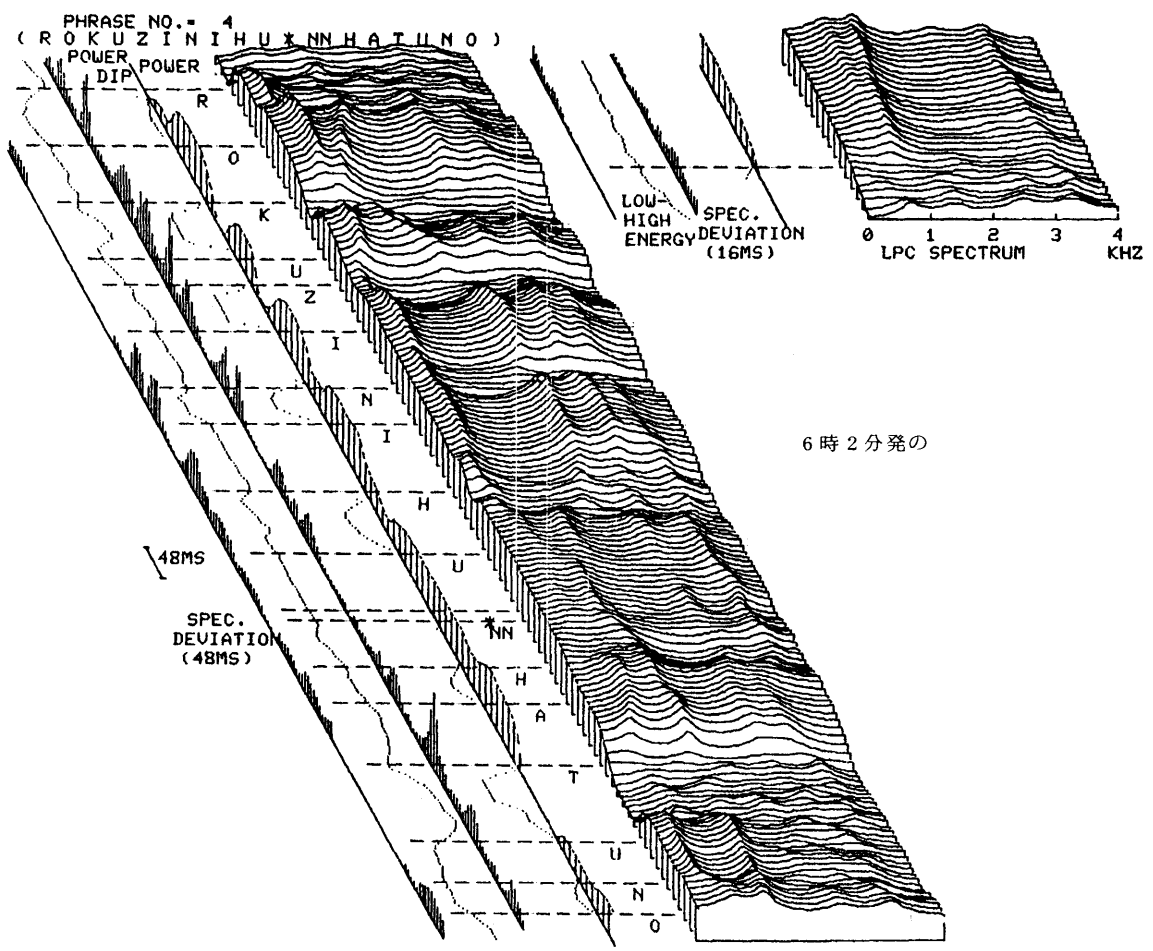
1日の



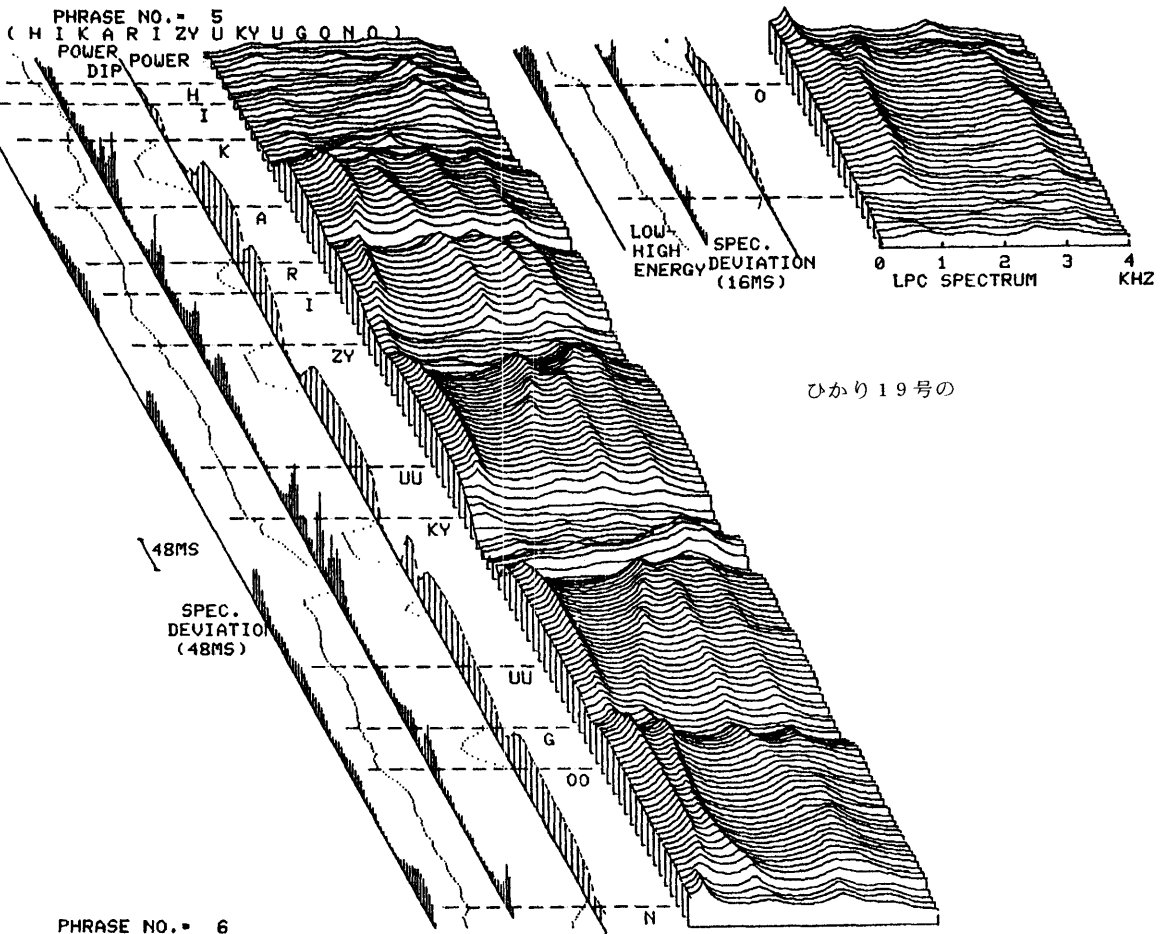
新大阪から



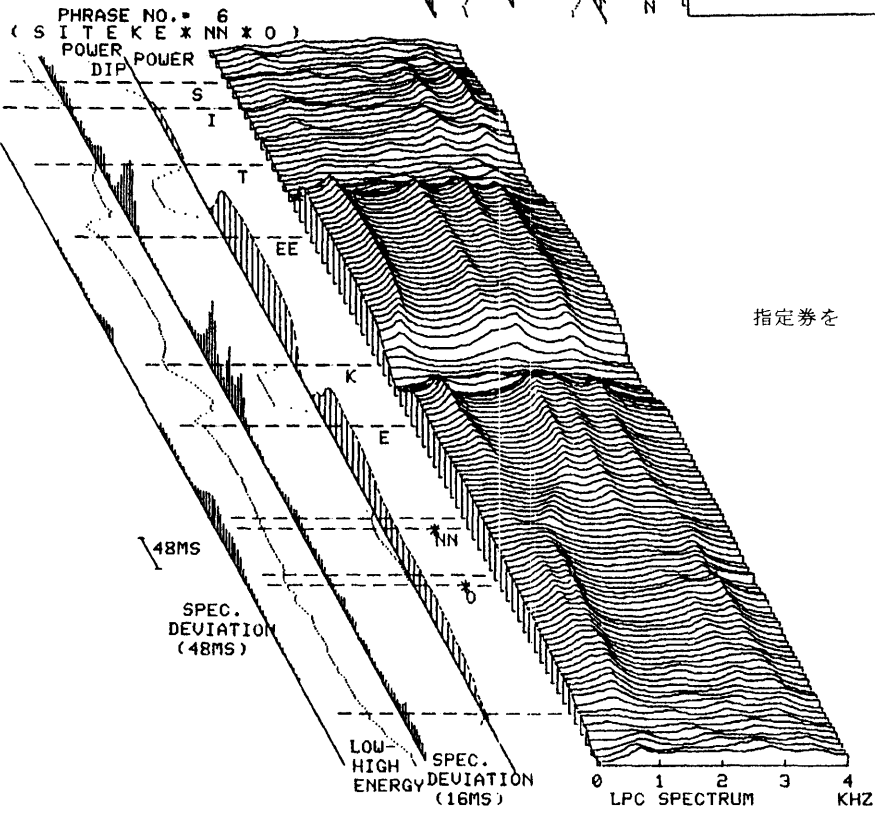
博多までの



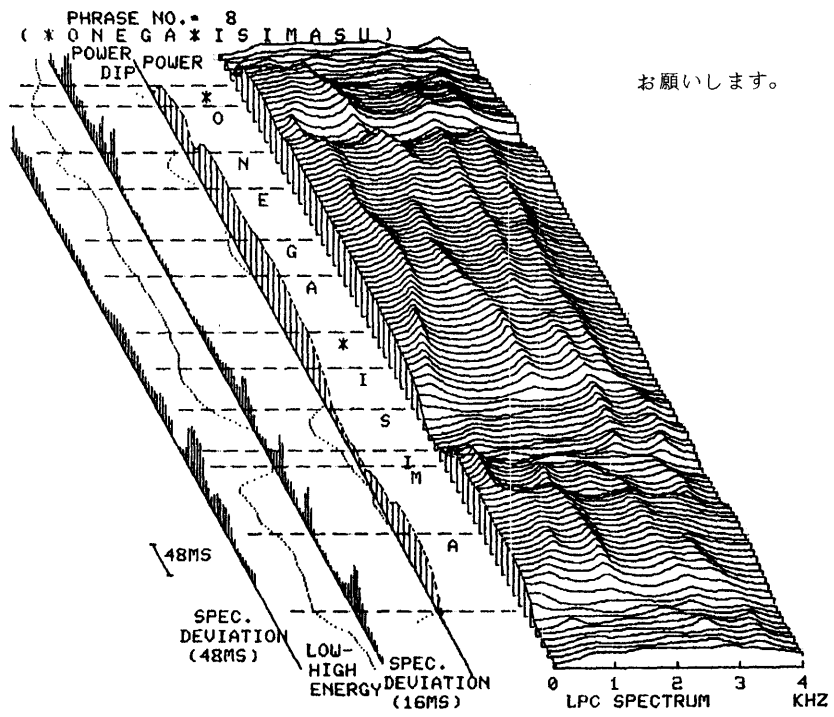
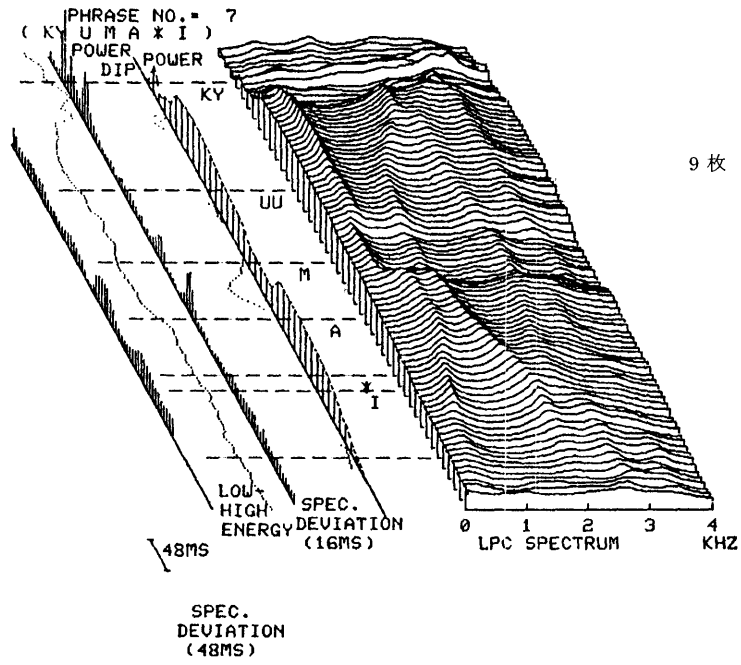
6時2分発の



ひかり19号の

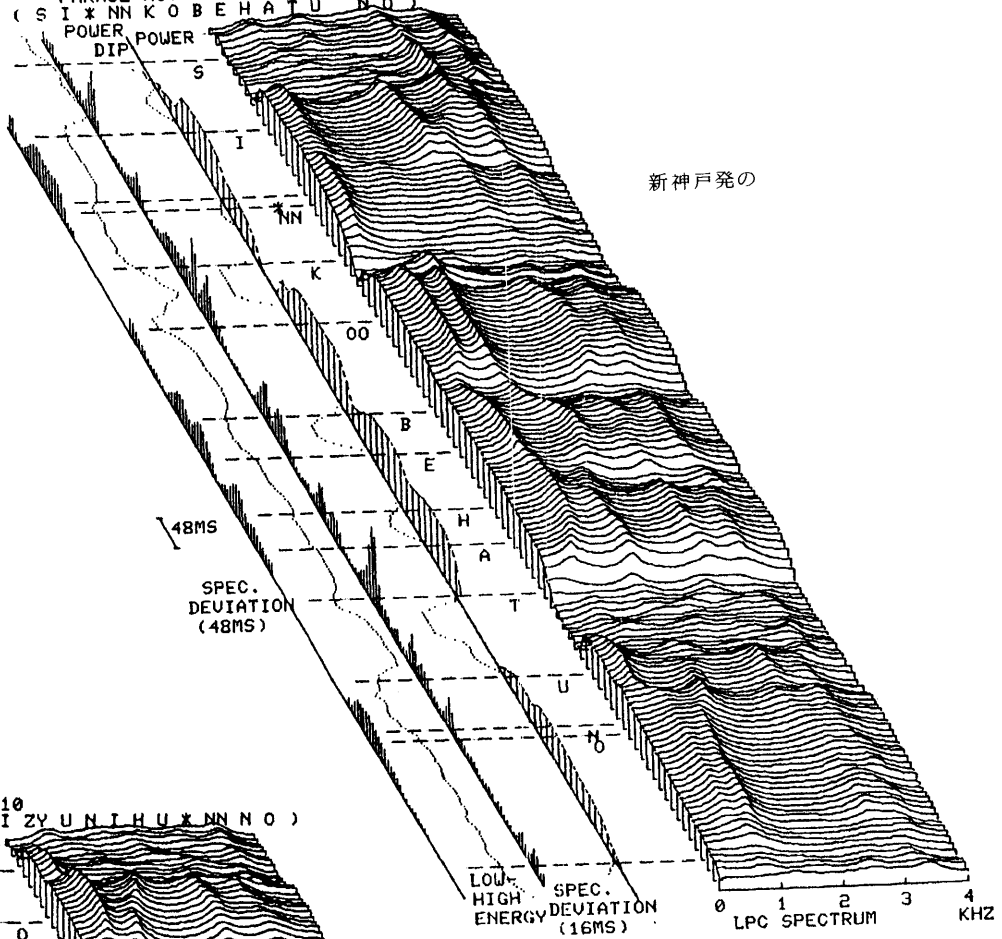


指定券を

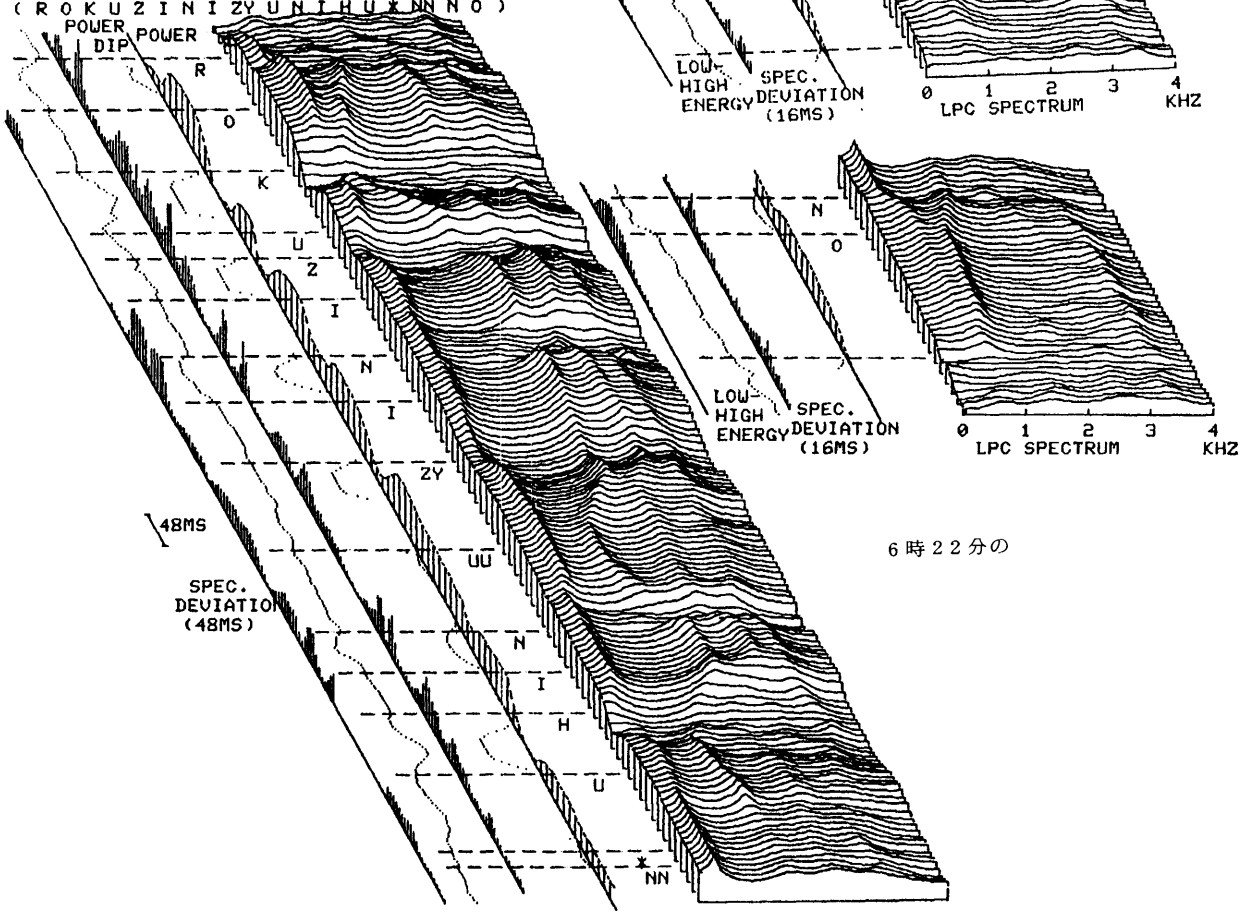


PHRASE NO. = 9
 (S I * NN K O B E H A T U N O)
 POWER POWER
 DIP

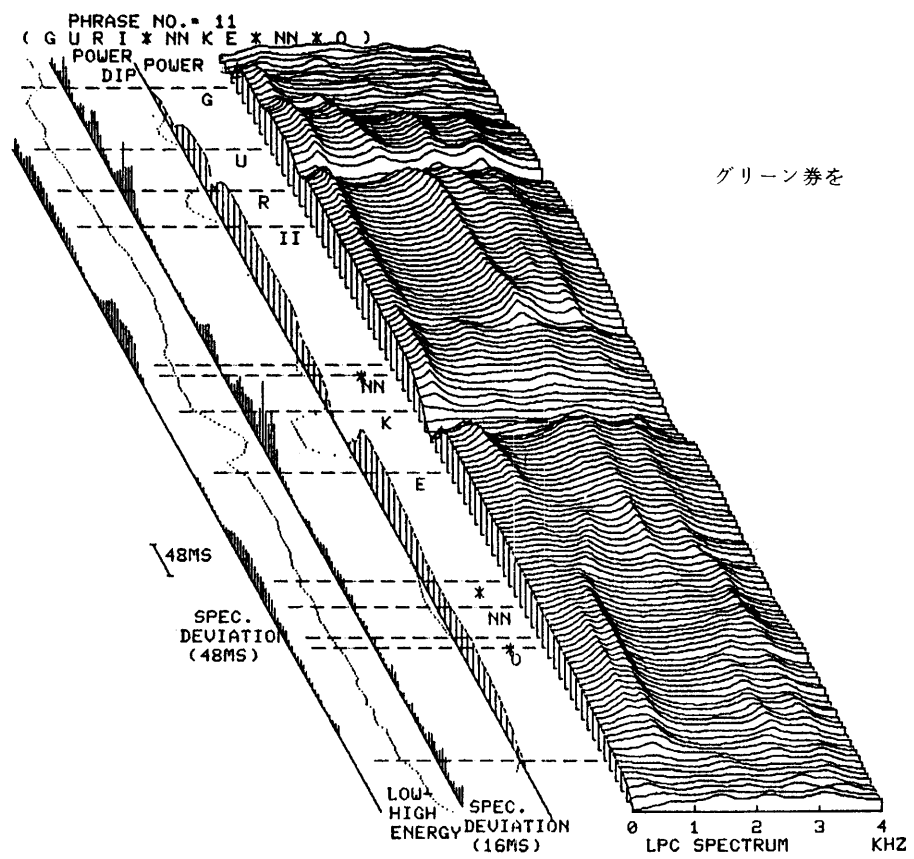
新神戸発の



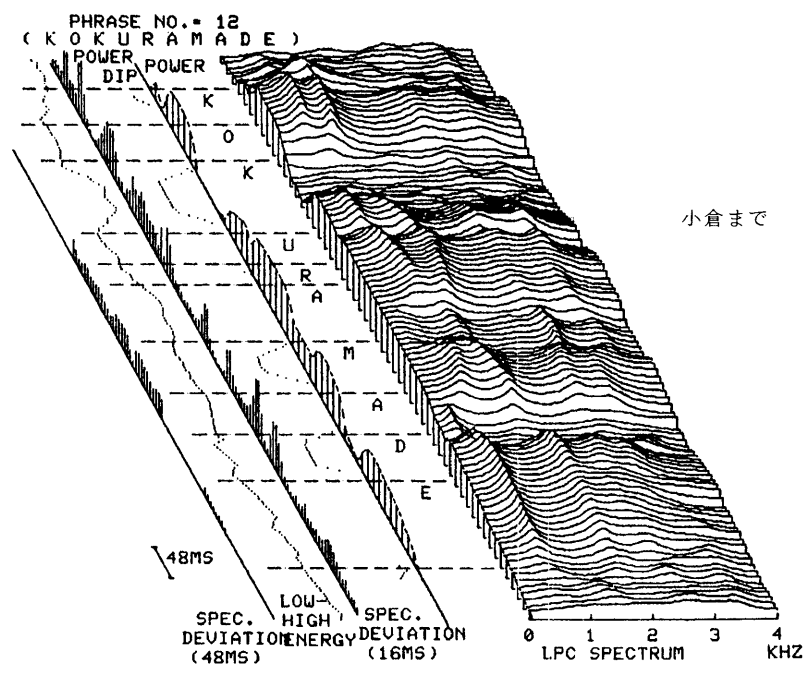
PHRASE NO. = 10
 (R O K U Z I N I Z Y U N I H U * NN N O)
 POWER POWER
 DIP



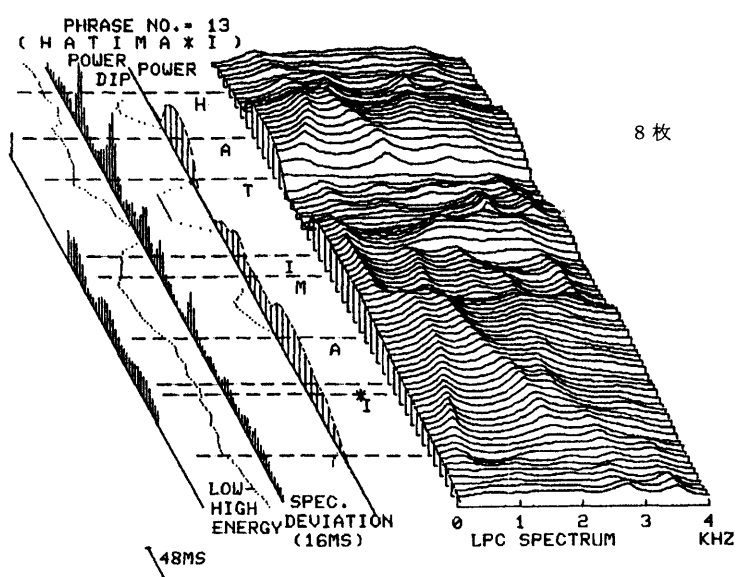
6時22分の



グリーン券を



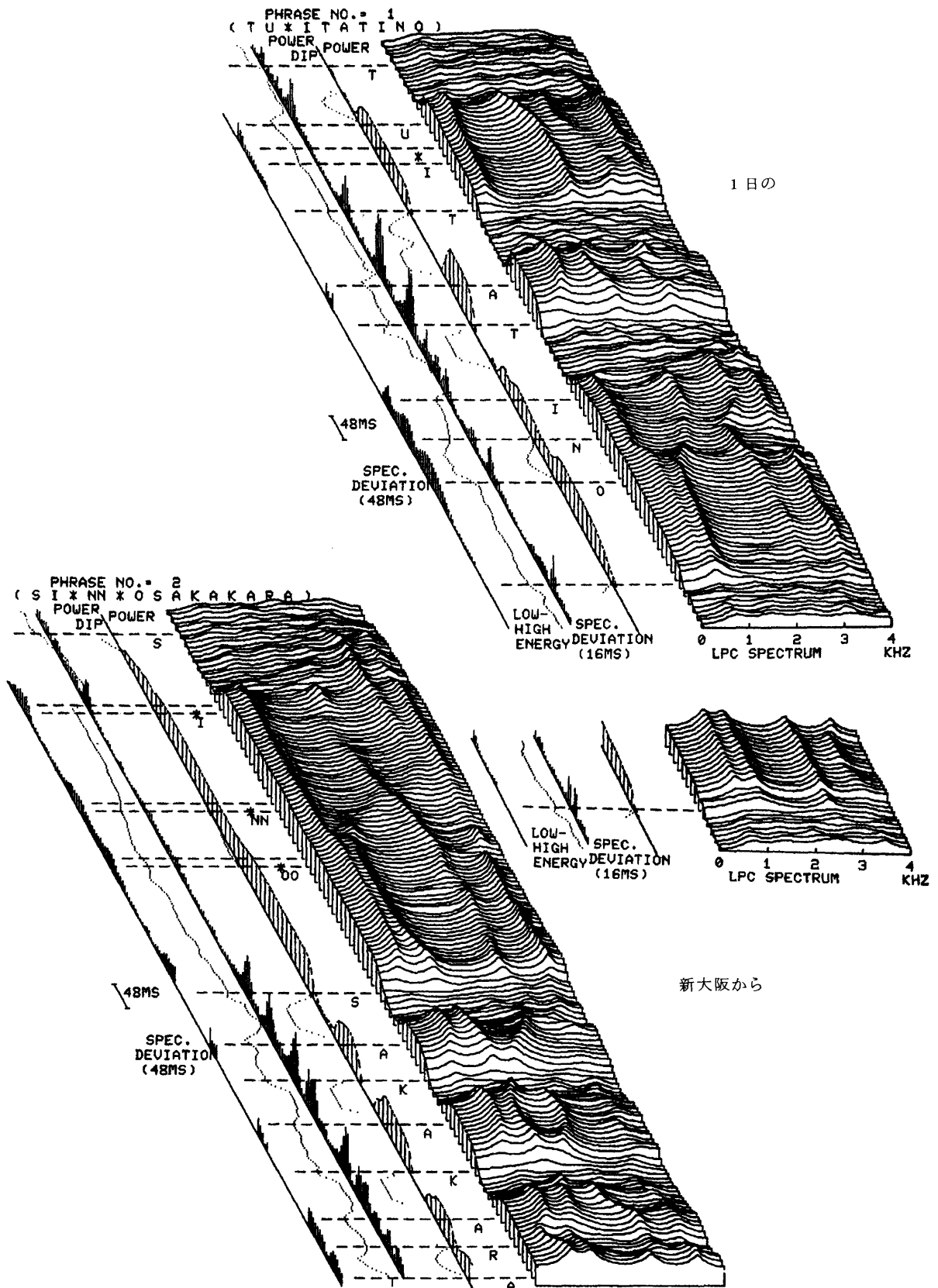
小倉まで

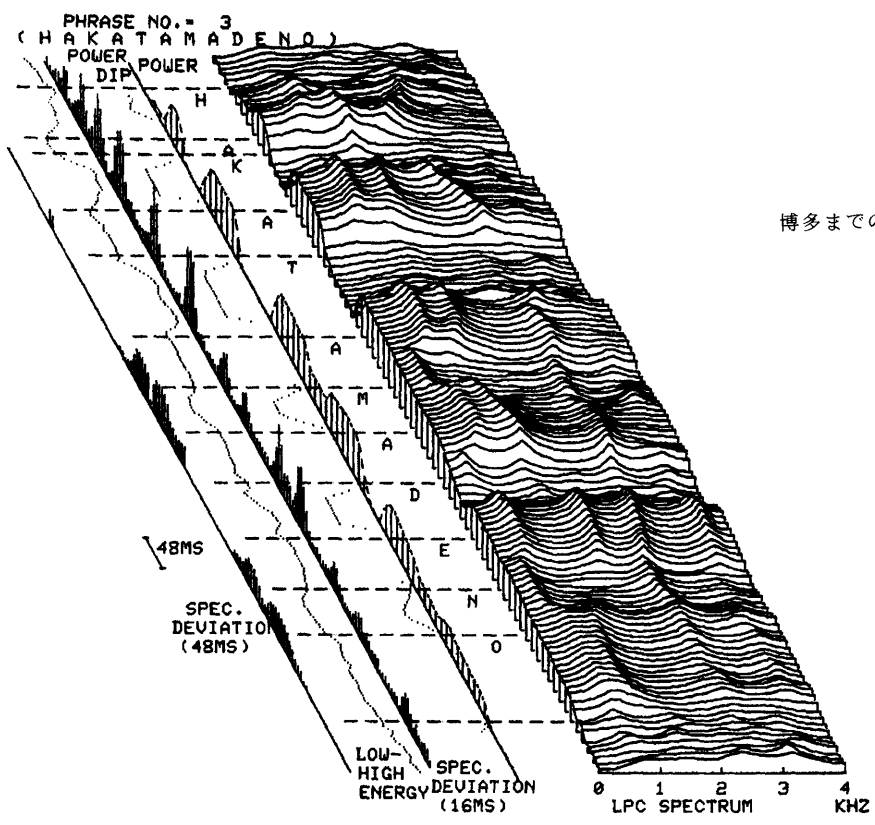


8枚

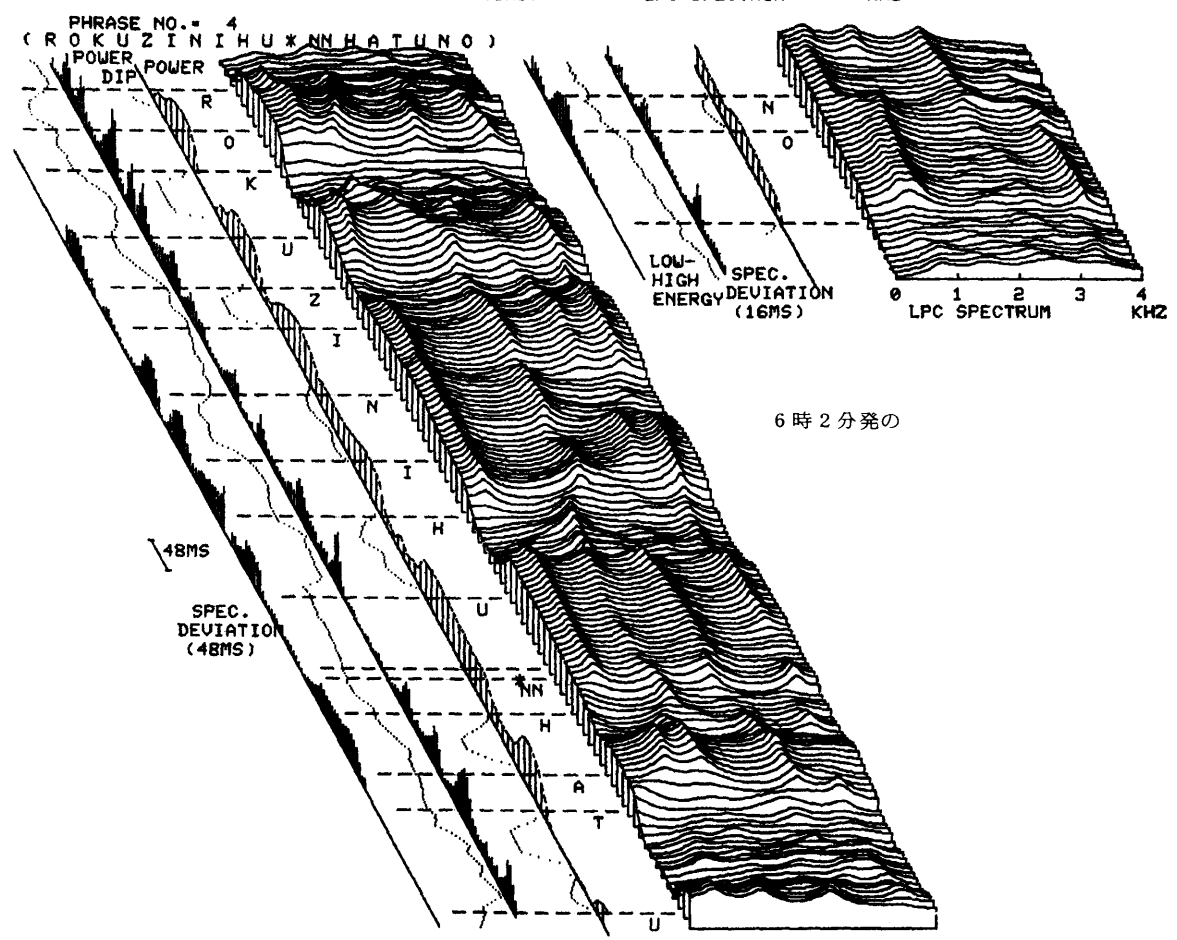
SPEC. DEVIATION (48MS)

発声者：KH

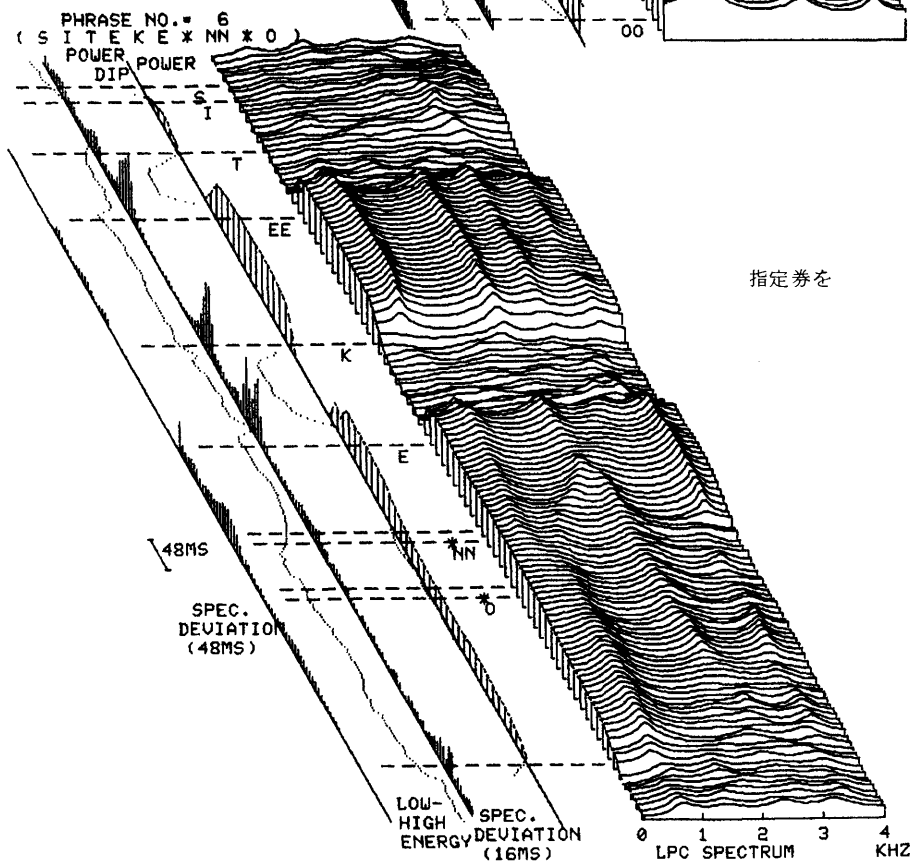
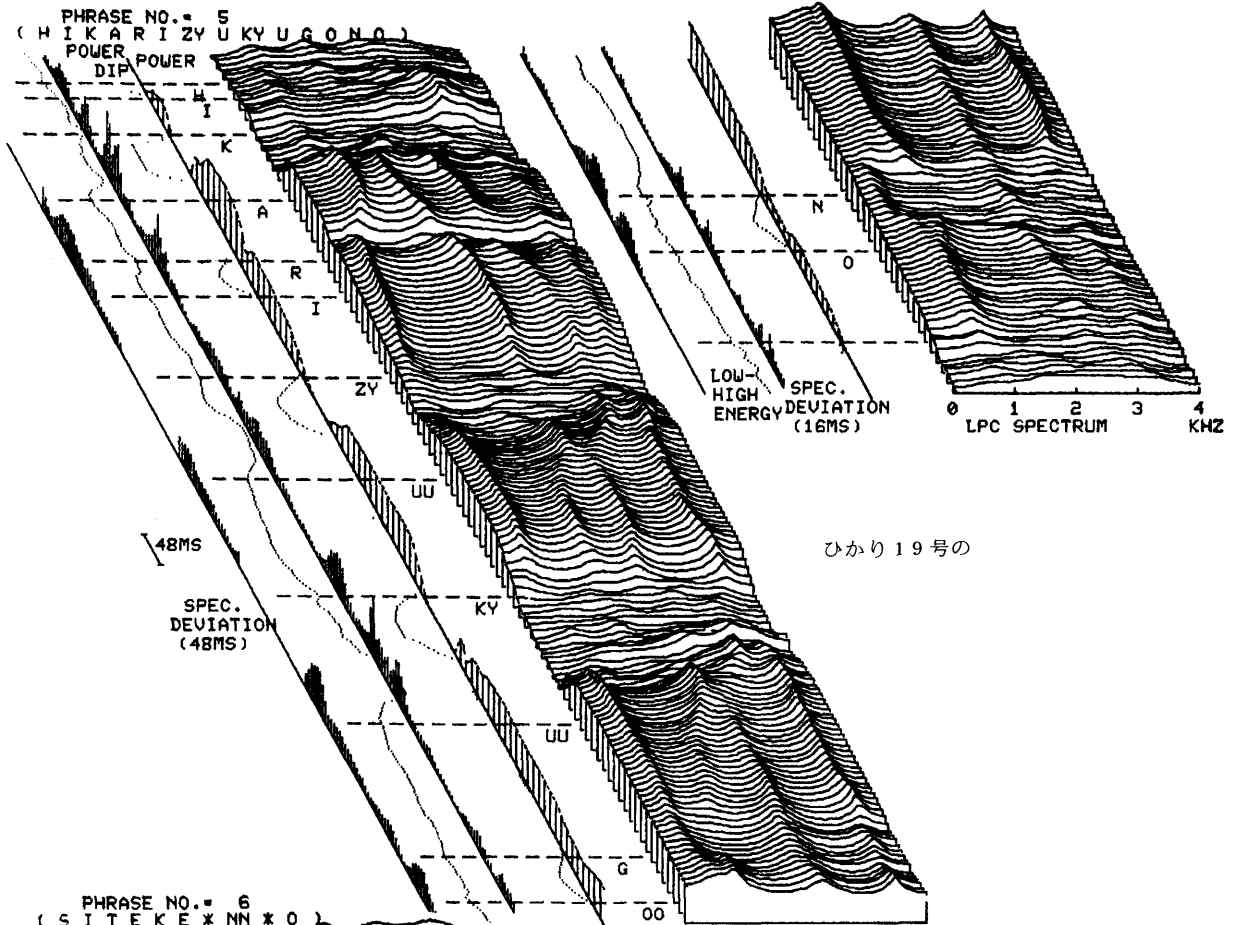


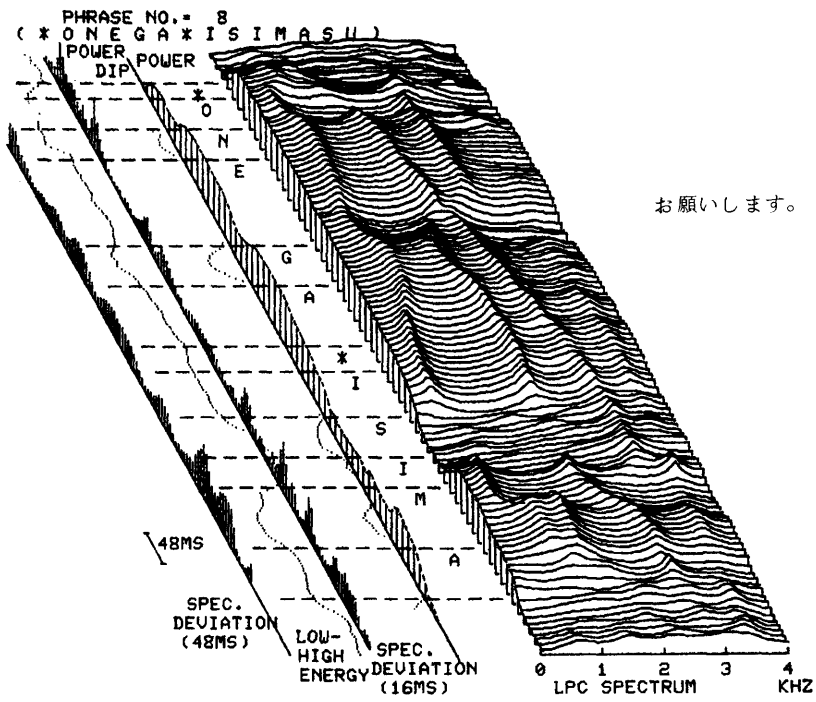
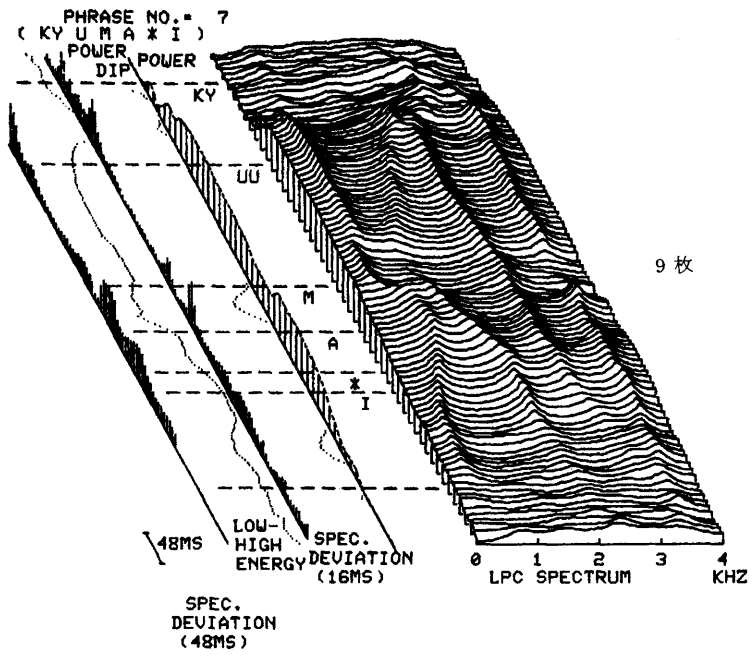


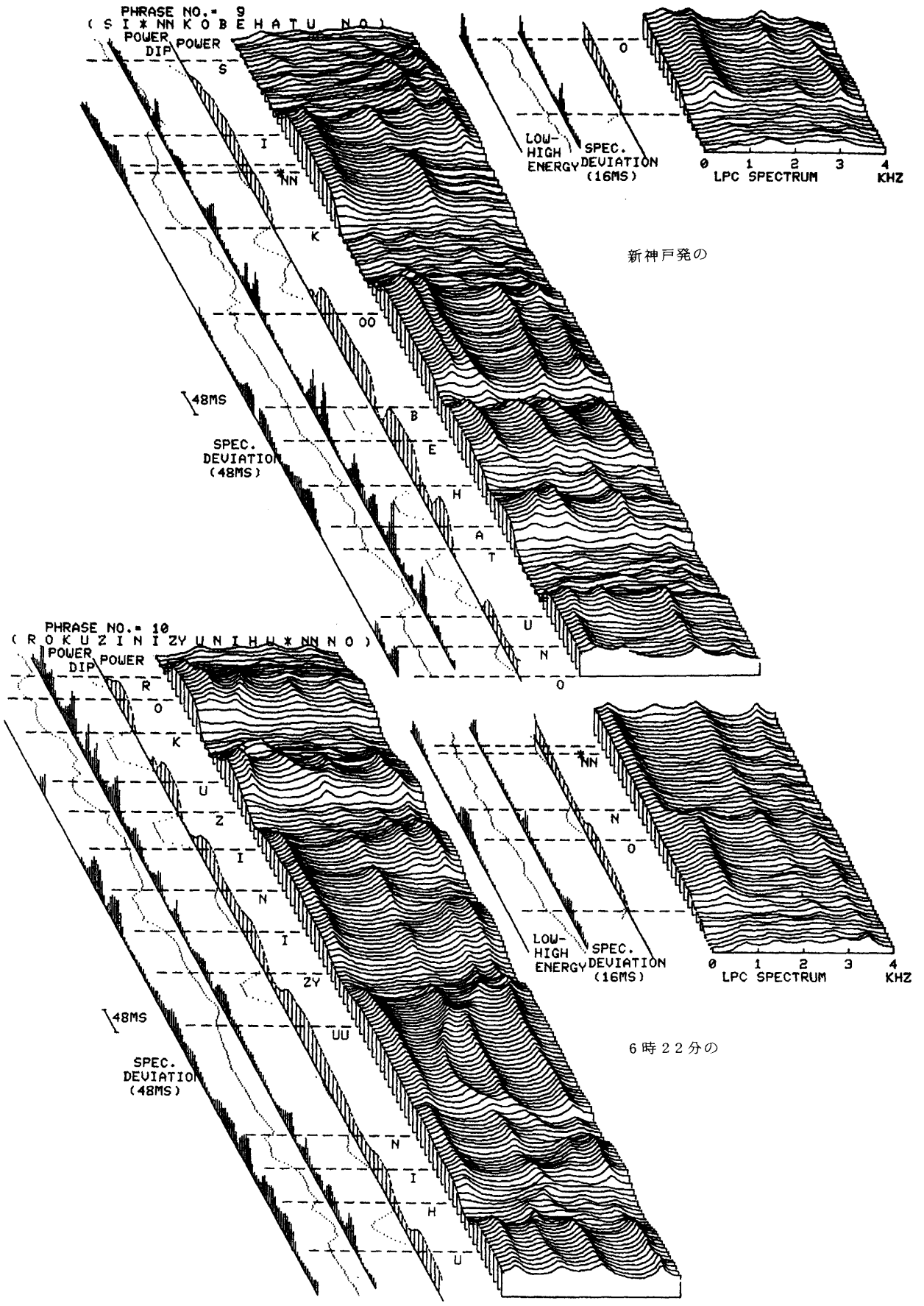
博多までの

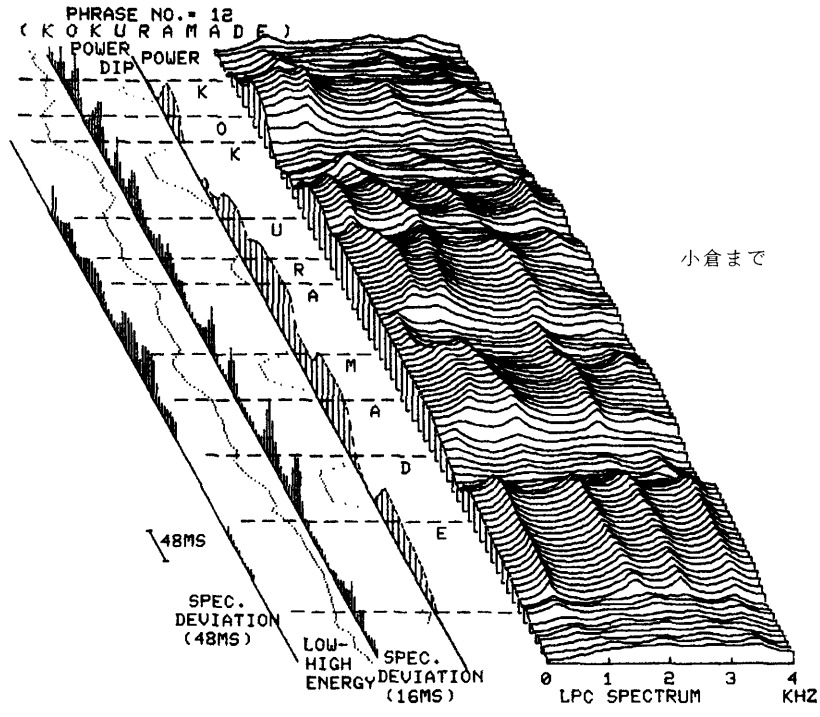
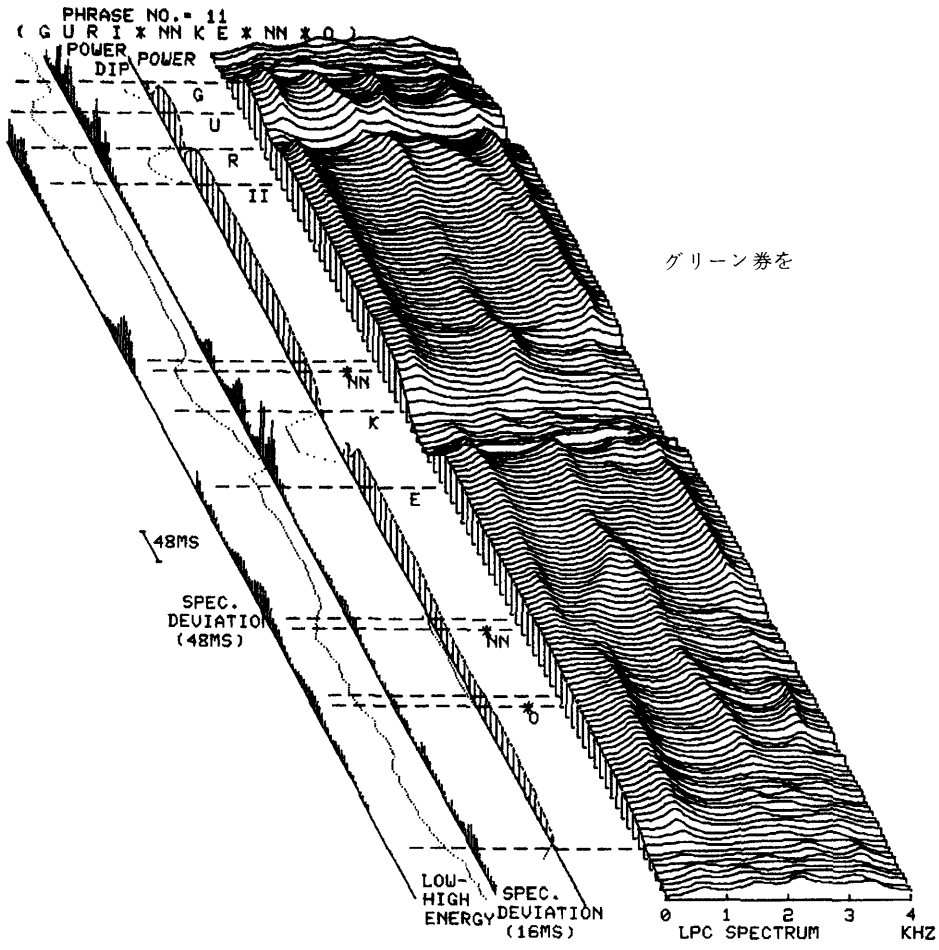


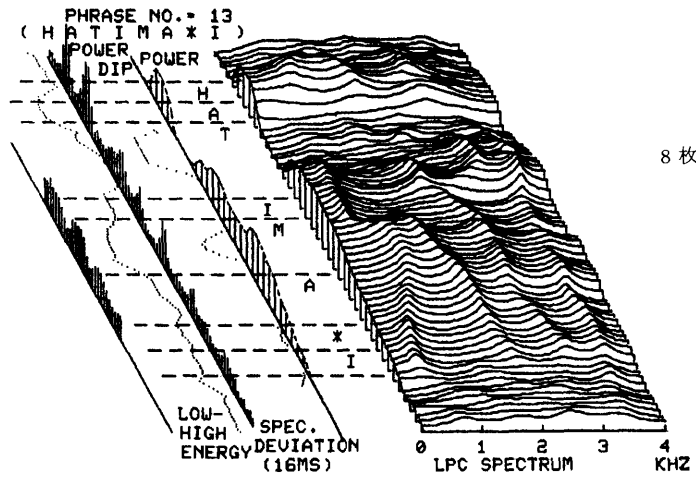
6時2分発の











8枚

48MS

SPEC. DEVIATION (48MS)

付録10. Voice Q-A System II の構文解析部の動作例

発声者：KS

```

RECOGNITION PROCESS ( SPEAKER KS )
                     ( SYSTEM VOICE Q-A SYSTEM II-2 )
( TSUITACHINO SHINOSAKAKARA HAKATAMADE 6J12FUNHATSUNO
  HIKARI19GONO SHITEKENO 9MAI ONECAISHIMASU )
( TSUITACHINO )
PHONEME LATTICE
( 1 ( ( S T ) 6 2 ) ( ( ZY KY HY ) 10 3 ) )
( 2 ( ( E U I ) 3 3 ) ( ( I ) 8 5 ) )
( 3 ( ( * ) 1 4 ) ( ( I E ) 4 5 ) )
( 4 ( ( I ) 4 5 ) )
( 5 ( ( K T S ) 6 6 ) )
( 6 ( ( A ) 3 7 ) )
( 7 ( ( S HY ) 8 8 ) )
( 8 ( ( U ) 2 9 ) )
( 9 ( ( M R N G Z ) 3 10 ) )
( 10 ( ( O ) 10 END ) )
**** RECOGNITION PROCESS ****
( TSUITACHI ( 1 ) 9 10 8 )
( NO ( 9 10 8 ) END 10 )
( FUTSUKA ( 1 ) 7 )
( YOKKA ( 1 ) 7 )
( 2 ( 1 ) 1 )
( 10 ( 1 ) 5 )
( 8 ( 5 ) 9 8 )
( KU ( 5 ) 6 )
( 10 ( 1 ) 5 )
( SHINKURASHIKI ( 1 ) 9 )
( SHITEI ( 1 ) 5 )
( GA ( 5 ) 7 )
( FUTSU ( 1 ) 3 4 5 )
( GA ( 3 4 5 ) 7 )
( 1 ( 1 ) 3 4 5 159 )
RECOGNITION RESULT
(( ( 1 ) NICHU ) ( --- KARA ) ( --- MADE ) ( --- JI --- FUN HATSU ) ( --- --- GO ) )
) ( --- MAI ) )

( SHINOSAKAKARA )
PHONEME LATTICE
( 1 ( ( S H R T B ) 4 2 ) )
( 2 ( ( NN U I ) 10 3 ) )
( 3 ( ( R * ) 2 4 ) )
( 4 ( ( O ) 11 5 ) )
( 5 ( ( K S H ) 5 6 ) )
( 6 ( ( A ) 1 7 ) )
( 7 ( ( T K S ) 4 8 ) )
( 8 ( ( A ) 4 9 ) )
( 9 ( ( K T S ) 6 10 ) )
( 10 ( ( A ) 2 11 ) )
( 11 ( ( * R N G M ) 2 12 ) )
( 12 ( ( A ) 5 END ) )
**** RECOGNITION PROCESS ****
( 2 ( 1 ) 3 4 2 1 )
( SHIZUOKA ( 1 ) 7 )
( HATSU ( 7 ) 10 )
( O ( 10 ) 11 12 )
( WA ( 10 ) 11 12 )
( GA ( 10 ) 11 12 )
( SHINOSAKA ( 1 ) 9 )
( KARA ( 9 ) END )
( KU ( 1 ) 3 4 2 )
( SHITEI ( 1 ) 3 4 2 )
( FUTSU ( 1 ) 3 4 5 )
( GREEN ( 1 ) 3 4 )
( 1 ( 1 ) 3 4 2 )
( 5 ( 1 ) 5 4 )
( 9 ( 1 ) 3 4 )
RECOGNITION RESULT
(( ( 1 ) NICHU ) ( SHINOSAKA ) KARA ) ( --- MADE ) ( --- JI --- FUN HATSU ) ( --- --- GO ) )
--- SEKI ) ( --- MAI ) )

```


(HAKATAMADENO)
 PHONEME LATTICE
 (1 ((* S P T) 3 2))
 (2 ((A) 1 3))
 (3 ((K T D) 5 4))
 (4 ((A) 2 5))
 (5 ((K T D S) 6 6))
 (6 ((A O) 4 7))
 (7 ((R M N G Z) 3 8))
 (8 ((A) 4 9))
 (9 ((C Z D H) 3 10))
 (10 ((E) 4 11))
 (11 ((M N G) 4 12))
 (12 ((O) 10 END))

**** RECOGNITION PROCESS ****
 (ASHITA (1) 7)
 (NO (7) 9)
 (DE (7) 11)
 (NODE (7) 11)
 (WA (7) 9)
 (GA (7) 9)
 (HAKATA (1) 7 8 9)
 (MADE (7 8 9) 11 12)
 (NO (11 12) END 12)
 (7 (1) 5)
 (8 (1) 5)
 (KU (1) 2)
 (5 (1) 3)

RECOGNITION RESULT
 ((1) NICHU)((SHINOSAKA) KARA)((HAKATA) MADE) (--- JI --- FUN HATSU) (--- --- GO)
 (--- SEKI) (--- MAI))

(ROKUJINIHUNHATSUNO)
 PHONEME LATTICE
 (1 ((B T * M R) 3 2))
 (2 ((O) 2 3))
 (3 ((T K) 7 4))
 (4 ((U) 2 5))
 (5 ((Z D H M G) 4 6))
 (6 ((I) 2 7))
 (7 ((M N Z H) 6 8))
 (8 ((I) 3 9))
 (9 ((D H K Z) 7 10))
 (10 ((O U) 9 11))
 (11 ((H D Z K S) 5 12))
 (12 ((A) 3 13))
 (13 ((K T S) 6 14))
 (14 ((U) 3 15))
 (15 ((R M N G Z) 3 16))
 (16 ((O) 9 END))

**** RECOGNITION PROCESS ****
 (FUTSUKA (1) 5)
 (MUIKA (1) 5)
 (6 (1) 5 4)
 (JI (5 4) 7)
 (2 (7) 9)
 (10 (9) 11)
 (HA (11) 13)
 (8 (11) 15 16 14)
 (FUN (15 16 14) 16 END)
 (NO (16 END) END 16)
 (10 (7) 11)
 (FUN (9) 11 12)
 (HATSU (11 12) 15 16 14)
 (NO (15 16 14) END 16)
 (ZI (7) 9)
 (ZYU (7) 11 10)
 (8 (1) 5 4)
 (JI (5 4) 7)
 (KU (1) 3 2)

RECOGNITION RESULT
 ((1) NICHU)((SHINOSAKA) KARA)((HAKATA) MADE) (6 8) JI (2 18 28) FUN HATSU)
 (--- --- GO) (--- SEKI) (--- MAI))

```

( HIKARIZYUKYUGONO )
PHONEME LATTICE
( 1 ( ( H S R ) 4 2 ) ( ( S ) 9 4 ) )
( 2 ( ( I ) 2 3 ) )
( 3 ( ( T K ) 5 4 ) )
( 4 ( ( A ) 6 5 ) )
( 5 ( ( C R Z ) 3 6 ) )
( 6 ( ( I U ) 3 7 ) )
( 7 ( ( H Z ) 6 8 ) ( ( ZY KY ) 11 9 ) )
( 8 ( ( I ) 4 9 ) ( ( I ) 9 11 ) )
( 9 ( ( Y * ) 1 10 ) ( ( U 0 ) 5 11 ) )
( 10 ( ( U 0 ) 4 11 ) )
( 11 ( ( T P K S ) 9 12 ) ( ( ZY KY ) 13 13 ) )
( 12 ( ( I E ) 4 14 ) ( ( U E ) 11 16 ) )
( 13 ( ( U 0 ) 7 16 ) )
( 14 ( ( * ) 1 15 ) )
( 15 ( ( U ) 6 16 ) )
( 16 ( ( D K T ) 5 17 ) )
( 17 ( ( O ) 13 18 ) )
( 18 ( ( M N Z H G ) 5 19 ) )
( 19 ( ( O ) 9 END ) )
**** RECOGNITION PROCESS ****
(FUTSUKA ( 1 ) 5 6 )
(ITSUKA ( 1 ) 5 6 )
(MUIKA ( 1 ) 5 )
( 2 ( 1 ) 3 )
( 3 ( 1 ) 6 5 )
(10 ( 6 5 ) 11 )
( 1 ( 11 ) 14 15 12 )
(HIKARI ( 1 ) 7 )
( 2 ( 7 ) 9 10 11 8 )
( 9 ( 7 ) 11 )
(10 ( 9 10 11 8 ) 16 )
( 5 ( 16 ) 18 17 )
(GO ( 18 17 ) END )
(GO ( 16 ) 18 )
(NO ( 18 ) END 19 )
(10 ( 11 ) 16 )
( 100 ( 7 ) 16 12 )
( 2 ( 16 12 ) 14 15 12 )
( 5 ( 16 12 ) 18 17 )
(10 ( 18 17 ) END )
(10 ( 16 12 ) 18 )
( 4 ( 18 ) END )
( 5 ( 18 ) END 19 )
(GO ( 18 ) END )
(10 ( 7 ) 11 )
( 9 ( 11 ) 16 )
(GO ( 16 12 ) 18 )
(SHITEI ( 1 ) 3 )
RECOGNITION RESULT
(( 1 ) NICH I)(( SHINOSAKA ) KARA)(( HAKATA ) MADE)(( 6 8 ) JI ( 2 18 28 ) FUN HATSU )
(( HIKARI )( 19 90 95 100 105 20 25 110 ) GO )( --- SEKI )( --- MAI ))

( SHITEKENO )
PHONEME LATTICE
( 1 ( ( S H ) 4 2 ) ( ( S ) 12 4 ) )
( 2 ( ( I U ) 3 3 ) )
( 3 ( ( T K P ) 7 4 ) )
( 4 ( ( E I U ) 10 5 ) )
( 5 ( ( K T S ) 6 6 ) ( ( ZY KY ) 10 7 ) )
( 6 ( ( E ) 3 7 ) ( ( E I ) 8 9 ) )
( 7 ( ( Y * R ) 1 8 ) ( ( U 0 ) 4 9 ) )
( 8 ( ( U 0 ) 3 9 ) )
( 9 ( ( * ) 1 10 ) )
(10 ( ( NN ) 5 11 ) )
(11 ( ( * ) 1 12 ) )
(12 ( ( O ) 12 END ) )
**** RECOGNITION PROCESS ****
( 2 ( 1 ) 3 )
(SHITEI ( 1 ) 5 4 9 10 11 12 7 8 6 3 )
(KEN ( 5 4 9 10 11 12 7 8 6 3 ) 9 10 11 12 5 )
(NO ( 9 10 11 12 5 ) END 12 )
(FUTSU ( 1 ) 5 9 10 3 )
(KEN ( 5 9 10 3 ) 9 10 11 12 5 )
RECOGNITION RESULT
(( 1 ) NICH I)(( SHINOSAKA ) KARA)(( HAKATA ) MADE)(( 6 8 ) JI ( 2 18 28 ) FUN HATSU )
(( HIKARI )( 19 90 95 100 105 20 25 110 ) GO )( ( FUTSU ) SEKI )( --- MAI ))

```

```

( KYUMAI )
PHONEME LATTICE
( 1 ( ( S ) 6 2 ) ( ( ZY KY HY ) 12 4 ) )
( 2 ( ( I ) 4 3 ) ( ( U O ) 13 6 ) )
( 3 ( ( Y * ) 3 5 ) )
( 4 ( ( U O ) 7 6 ) )
( 5 ( ( U O ) 5 6 ) )
( 6 ( ( M N ) 5 7 ) )
( 7 ( ( A ) 5 8 ) ( ( A ) 9 10 ) )
( 8 ( ( * ) 1 9 ) )
( 9 ( ( I ) 3 END ) )
**** RECOGNITION PROCESS ****
( KYO ( 1 ) 6 )
( WA ( 6 ) 8 9 END 7 )
( 10 ( 1 ) 6 )
( 2 ( 6 ) 7 )
( 7 ( 6 ) 8 9 END 7 )
( 4 ( 1 ) 7 6 )
( MAI ( 7 6 ) END 7 8 9 )
( 9 ( 1 ) 6 )
( MAI ( 6 ) END 7 8 9 )
RECOGNITION RESULT
(( 1 ) NICH I )(( SHINOSAKA ) KARA )(( HAKATA ) MADE )(( 6 8 ) JI ( 2 18 28 ) FUN HATSU )
(( HIKARI ) ( 19 90 95 100 105 20 25 110 ) GO )(( FUTSU ) SEKI )(( 9 4 ) MAI )

( ONEGAISHIMASU )
PHONEME LATTICE
( 1 ( ( S T H R B ) 7 2 ) )
( 2 ( ( O ) 4 3 ) )
( 3 ( ( M N G ) 4 4 ) )
( 4 ( ( E ) 5 5 ) )
( 5 ( ( * G R M N ) 4 6 ) )
( 6 ( ( A O ) 4 7 ) ( ( A E ) 7 8 ) ( ( A E ) 11 11 ) )
( 7 ( ( * R G ) 3 10 ) )
( 8 ( ( * ) 1 9 ) )
( 9 ( ( I NN ) 3 11 ) )
( 10 ( ( I U NN ) 3 11 ) )
( 11 ( ( S K ) 9 12 ) )
( 12 ( ( U I ) 2 13 ) )
( 13 ( ( R G W ) 2 14 ) )
( 14 ( ( A ) 5 END ) )
**** RECOGNITION PROCESS ****
( ONEGAI ( 1 ) 11 8 9 6 7 10 )
( ITA ( 11 8 9 6 7 10 ) END )
( SHI ( 11 8 9 6 7 10 ) 13 14 12 11 )
( MAS ( 13 14 12 11 ) END )
( KYO ( 1 ) 3 )
( 2 ( 1 ) 1 )
( HIROSHIMA ( 1 ) 7 10 8 9 11 6 )
( EKI ( 7 10 8 9 11 6 ) 13 14 12 )
( KARA ( 7 10 8 9 11 6 ) END )
( KOKURA ( 1 ) 7 10 8 9 11 6 )
( E ( 13 14 12 ) 13 14 )
( WA ( 13 14 ) END )
RECOGNITION RESULT
(( 1 ) NICH I )(( SHINOSAKA ) KARA )(( HAKATA ) MADE )(( 6 8 ) JI ( 2 18 28 ) FUN HATSU )
(( HIKARI ) ( 19 90 95 100 105 20 25 110 ) GO )(( FUTSU ) SEKI )(( 9 4 ) MAI )
PENALTY SCORES FOR RECOGNIZED ITEMS
(( 3 )(( 1 10 11 ) )(( 1 9 10 ) )(( 2 6 ) JI ( 2 7 8 ) )(( 3 ) ( 3 4 4 5
5 5 5 7 ) )(( 1 ) )(( 0 2 ) ) )
RESPONSE SENTENCE
9 MAI DESUKA
FINAL RESULT AFTER INFERENCE
(( 1 ) NICH I )(( SHINOSAKA ) KARA )(( HAKATA ) MADE )(( 6 ) JI ( 2 ) FUN HATSU
) )(( HIKARI ) ( 19 ) GO )(( FUTSU ) SEKI )(( 9 4 ) MAI )
PENALTY SCORES
(( 3 ) )(( 1 ) )(( 1 ) )(( 2 ) JI ( 2 ) )(( 100 ) ( 100 ) )(( 1 ) )(( 0 2 ) ) )

```

発声者：RN

```
RECOGNITION PROCESS ( SPEAKER RN )
                      ( SYSTEM VOICE Q-A SYSTEM II-2 )
( TSUITACHINO SHINOSAKAKARA HAKATAMADE 6JI2FUNHATSUNO
  HIKARI19GONO SHITEKENO 9MAI ONECAISHIMASU )
( TSUITACHINO )
PHONEME LATTICE
( 1 ( ( S T K ) 7 2 ) )
( 2 ( ( E I ) 8 3 ) )
( 3 ( ( K T S ) 5 4 ) ( ( ZY KY HY ) 6 5 ) )
( 4 ( ( A E ) 3 6 ) )
( 5 ( ( A E ) 2 6 ) )
( 6 ( ( Z K H D S ) 5 7 ) )
( 7 ( ( I E ) 2 8 ) )
( 8 ( ( G N R M ) 3 9 ) )
( 9 ( ( O ) 13 END ) )
***** RECOGNITION PROCESS *****
( TSUITACHI ( 1 ) 8 9 )
( NO ( 8 9 ) END 9 )
( 2 ( 1 ) 1 )
( SHINKURASHIKI ( 1 ) 9 )
( SHITEI ( 1 ) 3 2 )
( SEKI ( 3 2 ) 8 9 )
( 1 ( 1 ) 3 2 )
RECOGNITION RESULT
(( 1 ) NICH I )( --- KARA )( --- MADE )( --- JI --- FUN HATSU )( --- --- GO )( FUTSU
) SEKI )( --- MAI )

( SHINOSAKAKARA )
PHONEME LATTICE
( 1 ( ( S T H ) 8 2 ) )
( 2 ( ( I ) 7 3 ) )
( 3 ( ( Y * ) 1 4 ) )
( 4 ( ( U E ) 3 5 ) ( ( U E ) 4 6 ) ( ( O ) 11 9 ) )
( 5 ( ( * R ) 3 8 ) )
( 6 ( ( * R ) 2 7 ) )
( 7 ( ( O ) 4 9 ) ( ( O ) 10 11 ) )
( 8 ( ( O ) 3 9 ) )
( 9 ( ( * ) 1 10 ) )
( 10 ( ( O ) 4 11 ) )
( 11 ( ( T K S ) 6 12 ) )
( 12 ( ( A O E ) 2 13 ) )
( 13 ( ( C W R M N ) 3 14 ) )
( 14 ( ( A E U ) 3 15 ) )
( 15 ( ( T K S ) 6 16 ) )
( 16 ( ( A ) 2 17 ) )
( 17 ( ( R * N C W ) 2 18 ) )
( 18 ( ( A ) 9 END ) )
***** RECOGNITION PROCESS *****
( YOKA ( 1 ) 13 14 15 )
( NO ( 13 14 15 ) 15 )
( DE ( 13 14 15 ) 15 )
( WA ( 13 14 15 ) 15 )
( GA ( 13 14 15 ) 15 17 18 )
( KYO ( 1 ) 9 10 11 )
( 2 ( 1 ) 1 )
( KYOTO ( 1 ) 13 14 )
( EKI ( 13 14 ) 16 )
( KARA ( 16 ) END 18 )
( SHINOSAKA ( 1 ) 15 17 18 )
( KARA ( 15 17 18 ) END 18 )
( HIROSHIMA ( 1 ) 15 )
( KARA ( 15 ) END 18 )
( MADE ( 13 14 ) 17 18 )
( WA ( 17 18 ) END 18 )
( KU ( 1 ) 2 )
( FUTSU ( 1 ) 5 8 )
( KEN ( 5 8 ) 14 15 )
( NO ( 14 15 ) 15 )
( WA ( 14 15 ) 15 )
( GA ( 14 15 ) 15 17 18 )
RECOGNITION RESULT
(( 1 ) NICH I )( HIROSHIMA SHINOSAKA KYOTO ) KARA )( KYOTO ) MADE )( --- JI --- FUN HATSU
) ( --- --- GO )( FUTSU ) SEKI )( --- MAI )
```

(HAKATAMADENO)
 PHONEME LATTICE
 (1 ((S B * P T) 3 2))
 (2 ((A) 4 3))
 (3 ((S K) 5 4))
 (4 ((A) 2 5))
 (5 ((K T S) 6 6))
 (6 ((A) 2 7))
 (7 ((M Z N H) 5 8))
 (8 ((A) 3 9))
 (9 ((C D Z) 3 10))
 (10 ((E A) 4 11))
 (11 ((R M N C Z) 4 12))
 (12 ((O) 6 END))

**** RECOGNITION PROCESS ****

(HATSUKA (1) 5)
 (ASU (1) 4)
 (ASHITA (1) 7 5)
 (NO (7 5) 9)
 (NODE (7 5) 11 12)
 (WA (7 5) 9)
 (GA (7 5) 9 7)
 (KOKURA (1) 5)
 (KARA (5) 9)
 (DE (9) 11 12)
 (WA (9) 11)
 (GA (9) 11 12)
 (HAKATA (1) 7)
 (KARA (7) 11)
 (NO (11) END)
 (MADE (7) 11 12)
 (NO (11 12) END)
 (8 (1) 4)
 (KU (1) 2)

RECOGNITION RESULT

((1) NICHII)((HIROSHIMA SHINOSAKA KYOTO) KARA)((HAKATA KYOTO) MADE)((--- JI --- FUN
 HATSU)((--- --- GO)((FUTSU) SEKI)((--- MAI))

(ROKUJINIHUNHATSUNO)
 PHONEME LATTICE
 (1 ((D B N R M) 5 2))
 (2 ((O) 4 3))
 (3 ((K D T Z) 5 4))
 (4 ((U) 2 5))
 (5 ((D Z K C H) 4 6))
 (6 ((I E) 4 7))
 (7 ((C N R M) 3 8))
 (8 ((I) 6 9))
 (9 ((Z M H N) 5 10))
 (10 ((O) 10 11))
 (11 ((R M C W N) 3 12))
 (12 ((A) 5 13))
 (13 ((P S T) 7 14))
 (14 ((NN) 3 15))
 (15 ((N R * G M) 2 16))
 (16 ((O A) 10 END))

**** RECOGNITION PROCESS ****

(FUTSUKA (1) 5)
 (MUIKA (1) 5)
 (10 (1) 3)
 (1 (3) 5)
 (KU (3) 5)
 (KOKURA (1) 7)
 (6 (1) 5)
 (JI (5) 7 8)
 (2 (7 8) 9 8)
 (10 (9 8) 11)
 (HA (11) 13)
 (PPUN (13) END)
 (8 (11) 15 16 14)
 (FUN (15 16 14) END)
 (4 (7 8) 12 11)
 (10 (7 8) 11)
 (I (7 8) 9 8)
 (FUN (9 8) 12)
 (HATSU (12) 15 16 14)
 (NO (15 16 14) END 16)
 (ZI (9 8) 10)
 (ZI (7 8) 9 8)
 (8 (1) 5)
 (JI (5) 7 8)
 (KU (1) 3 2)
 (JI (3 2) 5)
 (0 (5) 9 7)
 (FUN (9 7) 12)
 (WA (12) 13)
 (GA (12) 13)

(RE (5) 7)
 (2 (5) 7 8)
 (ZI (5) 7 8)
 (ZYU (5) 8)
 (JI (3) 5)
 (FUTSU (1) 5)
 (KEN (5) 8 9)
 (SEKI (5) 9)

RECOGNITION RESULT

(((1) NI CHI)) ((HIROSHIMA SHINOSAKA KYOTO) KARA)) ((HAKATA KYOTO) MADE)) ((6 19 8) JI
 (2 18 28) FUN HATSU)) ((NIL NIL GO)) ((FUTSU) SEKI)) ((NIL MAI))

(HIKARIZYUKYUGONO)

PHONEME LATTICE

(1 ((S H R) 4 2) ((S) 11 4))
 (2 ((I) 4 3))
 (3 ((T K) 5 4))
 (4 ((A) 5 5))
 (5 ((R G S Z) 3 6))
 (6 ((I) 4 7))
 (7 ((D Z K H) 5 8))
 (8 ((I E) 4 9) ((U E) 11 11))
 (9 ((Y *) 1 10))
 (10 ((U E) 4 11))
 (11 ((D K T B H) 5 12) ((S T K) 9 14) ((ZY KY) 16 16))
 (12 ((I) 1 13))
 (13 ((R G K Z S) 3 14) ((Y HY ZY) 9 16))
 (14 ((E) 5 15) ((E) 10 17))
 (15 ((Y *) 1 16))
 (16 ((U O) 3 17))
 (17 ((R M N C Z) 4 18))
 (18 ((O) 12 19))
 (19 ((Z M H N G) 5 20))
 (20 ((O) 7 END))

**** RECOGNITION PROCESS ****

(FUTSUKA (1) 5 6)
 (ITSUKA (1) 5 6)
 (2 (1) 3)
 (3 (1) 6 5)
 (10 (6 5) 11)
 (1 (11) 13 14 15 16)
 (8 (1) 7)
 (JI (7) 9 10 8)
 (0 (9 10 8) 11 9 10)
 (FUN (11 9 10) 14 17)
 (NO (14 17) 19 18)
 (O (14 17) 19 18)
 (NOO (14 17) 19 18)
 (RE (9 10 8) 11 9 10)
 (FUN (11 9 10) 14 17)
 (2 (9 10 8) 9 10 8)
 (I (9 10 8) 11 9 10 8)
 (KU (1) 2 4)
 (HIKARI (1) 7)
 (2 (7) 9 10)
 (9 (7) 11)
 (10 (11) 17 18)
 (2 (17 18) 18)
 (4 (17 18) 20 END 19)
 (GO (20 END 19) END)
 (5 (17 18) 19 18)
 (GO (19 18) END)
 (6 (17 18) END 20)
 (GO (17 18) 19)
 (NO (19) END)
 (10 (7) 11)
 (GO (13 14 15 16) 19)
 (9 (11) 17 18)

RECOGNITION RESULT

(((1) NI CHI)) ((HIROSHIMA SHINOSAKA KYOTO) KARA)) ((HAKATA KYOTO) MADE)) ((6 19 8) JI
 (2 18 28) FUN HATSU)) ((HIKARI)) (90 19 95 11 94) GO)) ((FUTSU) SEKI)) ((--- MAI))

(SHITEKENO)

PHONEME LATTICE

(1 ((S H) 4 2) ((S) 13 4))
 (2 ((I) 6 3))
 (3 ((K T S) 6 4))
 (4 ((E O) 11 5))
 (5 ((K T D) 6 6) ((ZY KY) 17 9))
 (6 ((E) 6 7) ((E U) 8 8))
 (7 ((* G M R N) 4 10))
 (8 ((* R) 1 9))
 (9 ((O U NN) 12 END))
 (10 ((O U) 11 END))

**** RECOGNITION PROCESS ****

(KYO (1) 5)
 (DE (5) 7 10 8 9 6)
 (HONJITSU (1) 8 9 6)
 (NO (8 9 6) END 9)

(2 (1) 164)
 (TOKYO (1) END)
 (KYOTO (1) 8 9)
 (SHINKOBE (1) 7 10 8 9 6)
 (E (7 10 8 9 6) 7 10 8 9 6)
 (NO (7 10 8 9 6) END 10 9)
 (SHITEI (1) 5 4 10 9 7 8 6)
 (KEN (5 4 10 9 7 8 6) 10 END 7 8 9)
 (NO (10 END 7 8 9) END 10 9)
 (FUTSU (1) 8 9 END)
 (NO (8 9 END) END 9)
 (1 (1) 4)

RECOGNITION RESULT

((1) NICHU)((HIROSHIMA SHINOSAKA KYOTO) KARA)((HAKATA SHINKOBE KYOTO) MADE)((6 19
 8) JI (2 18 28) FUN HATSU)((HIKARI) (90 19 95 11 94) GO)((FUTSU) SEKI) (--- MAI)

(KYUMAI)
 PHONEME LATTICE

(1 ((P S *) 1 2) ((S) 4 4) ((Y KY HY) 8 5))
 (2 ((I) 1 3))
 (3 ((S R) 2 4) ((ZY KY) 8 6))
 (4 ((E) 6 7) ((E) 7 8) ((E) 12 11))
 (5 ((U E) 3 8))
 (6 ((U O) 6 11))
 (7 ((* R) 1 9))
 (8 ((* R) 1 10))
 (9 ((U O) 4 11))
 (10 ((NN O U) 3 11))
 (11 ((M N Z C H) 5 12))
 (12 ((A) 6 13) ((A) 11 END))
 (13 ((* R) 1 14))
 (14 ((E U) 3 END))

**** RECOGNITION PROCESS ****

(FUTSUKA (1) 13 14 END 12)
 (WA (13 14 END 12) 13 14 END 12)
 (2 (1) 3 4 5)
 (10 (3 4 5) 11)
 (2 (11) 12)
 (3 (11) 13 14 END)
 (7 (11) 13 14 END 12)
 (10 (1) 8 10 11)
 (2 (8 10 11) 12 11)
 (3 (8 10 11) 13 14 END)
 (5 (8 10 11) 11)
 (7 (8 10 11) 13 14 END 12)
 (HIMEJI (1) 12)
 (FUKUYAMA (1) 13 14 END)
 (HIKARI (1) 11)
 (2 (11) 12)
 (3 (11) 13 14 END)
 (7 (11) 13 14 END 12)
 (SHITEI (1) 11 7 9 8 10 4)
 (NO (11 7 9 8 10 4) 11)
 (O (11 7 9 8 10 4) 11)
 (WA (11 7 9 8 10 4) 13 14 END 12)
 (FUTSU (1) 11 8 10)
 (NO (11 8 10) 11)
 (O (11 8 10) 11)
 (WA (11 8 10) 13 14 END 12)
 (4 (1) 11 12)
 (MAI (11 12) END 12 13 14)
 (9 (1) 8 10 11)
 (MAI (8 10 11) END 12 13 14)

RECOGNITION RESULT

((1) NICHU)((HIROSHIMA SHINOSAKA KYOTO) KARA)((HAKATA SHINKOBE KYOTO) MADE)((6 19
 8) JI (2 18 28) FUN HATSU)((HIKARI) (90 19 95 11 94) GO)((FUTSU) SEKI) (4 9
) MAI)

(ONEGAISHIMASU)
 PHONEME LATTICE

(1 ((H * R B P) 1 2))
 (2 ((O) 4 3))
 (3 ((G R N M Z) 3 4) ((Y HY ZY) 10 6))
 (4 ((E U) 4 5) ((E) 8 7))
 (5 ((*) 1 6))
 (6 ((A O) 7 9) ((A O) 8 10) ((A E) 14 13))
 (7 ((*) 1 8))
 (8 ((A O) 5 9))
 (9 ((*) 2 12))
 (10 ((*) 1 11))
 (11 ((E U I) 4 13))
 (12 ((I E U) 3 13))
 (13 ((S H) 7 14))
 (14 ((I NN U) 1 15))
 (15 ((R M N C Z) 3 16))
 (16 ((A) 6 END))

```

**** RECOGNITION PROCESS ****
( YOYAKU ( 1 ) 15 16 14 )
( SHI ( 15 16 14 ) 15 16 )
( MAS ( 15 16 ) END )
( KYO ( 1 ) 3 )
( TOKYO ( 1 ) 10 11 13 )
( EKI ( 10 11 13 ) 15 16 14 )
( KARA ( 10 11 13 ) END )
( TOYOHASHI ( 1 ) 15 16 14 )
( OGORI ( 1 ) 13 )
( KARA ( 13 ) END )
( KOKURA ( 1 ) 9 12 10 11 13 6 )
( EKI ( 9 12 10 11 13 6 ) 15 16 14 )
( KARA ( 9 12 10 11 13 6 ) END )
( E ( 15 16 14 ) 15 16 )
( WA ( 15 16 ) END )
( 6 ( 1 ) 5 6 4 )
( KU ( 1 ) 3 4 2 )
( JI ( 3 4 2 ) 5 6 7 8 4 )
( 0 ( 5 6 7 8 4 ) 13 5 6 7 8 )
( FUN ( 13 5 6 7 8 ) 16 )
( WA ( 16 ) END )
( GA ( 16 ) END )
( RE ( 5 6 7 8 4 ) 13 5 6 7 8 )
( FUN ( 13 5 6 7 8 ) 16 )
( 4 ( 5 6 7 8 4 ) 9 12 10 11 )
( I ( 5 6 7 8 4 ) 6 5 7 8 4 )
( RO ( 5 6 7 8 4 ) 9 12 10 11 6 )
( HA ( 5 6 7 8 4 ) 9 12 10 11 13 6 )
( 8 ( 5 6 7 8 4 ) 15 )
( ZI ( 9 12 10 11 ) 13 )
( HIKARI ( 1 ) 4 )
( 100 ( 4 ) 15 14 )
( 2 ( 15 14 ) 16 15 )
( 4 ( 1 ) 4 5 6 )
( MAI ( 4 5 6 ) 13 )
( 5 ( 1 ) 3 4 )
( MAI ( 5 6 4 ) 13 )
( 9 ( 1 ) 3 )
RECOGNITION RESULT ( WITHOUT INFERENCE )
((( 1 ) NICHU )(( HIROSHIMA SHINOSAKA KYOTO TOKYO KOKURA OGORI ) KARA )(( HAKATA SHINKOBE TOYOHASHI
KYOTO TOKYO KOKURA ) MADE )(( 6 19 8 ) JI ( 2 18 28 ) FUN HATSU )(( HIKARI )( 90 19 95
11 94 ) GO )(( FUTSU ) SEKI )(( 4 9 ) MAI ))
PENALTY SCORES FOR RECOGNIZED ITEMS
((( 3 ))(( 5 6 9 10 12 12 ))(( 1 3 7 8 11 13 ))(( 6 10 11 ) JI ( 6 9 10 ))
( 4 )( 4 5 7 10 10 ))(( 2 ))(( 2 3 ))
RESPONSE SENTENCE
HIROSHIMA KARA DESUKA
FINAL RESULT AFTER INFERENCE
((( 1 ) NICHU )(( HIROSHIMA SHINOSAKA KYOTO TOKYO ) KARA )(( HAKATA SHINKOBE TOYOHASHI ) MADE )
( 6 19 8 ) JI ( 2 18 28 ) FUN HATSU )(( HIKARI )( 90 19 95 11 94 ) GO )(( FUTSU ) SEKI )
( 4 9 ) MAI ))
PENALTY SCORES
((( 3 ))(( 5 6 9 10 ))(( 1 3 7 ))(( 6 10 11 ) JI ( 6 9 10 ))( 4 )( 4 5
7 10 10 ))(( 2 ))(( 2 3 ))

```


発声者：KH

```
RECOGNITION PROCESS ( SPEAKER KH )
( SYSTEM VOICE Q-A SYSTEM 11-2 )
( TSUITACHINO SHINOSAKAKARA HAKATAMADE 6JI2FUNHATSUNO
HIKARI19GONO SHITEKENO 9MAI ONEGAISHIMASU )
( TSUITACHINO )
PHONEME LATTICE
( 1 ( ( S T ) 7 2 ) )
( 2 ( ( I ) 10 3 ) )
( 3 ( ( S T K ) 9 4 ) )
( 4 ( ( A ) 4 5 ) )
( 5 ( ( T K S ) 7 6 ) ( ( T K S ) 9 8 ) ( ( ZY KY ) 12 9 ) )
( 6 ( ( I ) 1 7 ) ( ( E I U ) 7 10 ) )
( 7 ( ( R S ) 1 8 ) ( ( Y HY ZY ) 5 10 ) )
( 8 ( ( E I U ) 5 10 ) )
( 9 ( ( E U ) 1 10 ) )
( 10 ( ( M N Z G R ) 4 11 ) )
( 11 ( ( O ) 11 END ) )
**** RECOGNITION PROCESS ****
( TSUITACHI ( 1 ) 7 8 10 11 6 )
( NO ( 7 8 10 11 6 ) END 11 )
( 2 ( 1 ) 1 )
( SHINKURASHIKI ( 1 ) 11 7 8 10 6 )
( E ( 11 7 8 10 6 ) 10 11 7 8 )
( NO ( 10 11 7 8 ) END 11 )
( HIKARI ( 1 ) 5 )
( PAUSE ( 5 ) 6 8 )
( 10 ( 6 8 ) END )
( 1 ( 5 ) 7 8 10 11 6 )
( GO ( 7 8 10 11 6 ) END )
( 9 ( 5 ) 10 )
( GO ( 10 ) END )
( 10 ( 10 ) END )
( 10 ( 5 ) 10 )
( 2 ( 10 ) 11 )
( 4 ( 10 ) END )
( 5 ( 10 ) END 11 )
( SHITEI ( 1 ) 3 2 )
( 1 ( 1 ) 159 152 )
RECOGNITION RESULT
(( 1 ) NICHU ) ( --- KARA ) ( --- MADE ) ( --- JI --- FUN HATSU ) ( --- --- GO ) ( --- SEKI )
( --- MAI )
```

```
( SHINOSAKAKARA )
PHONEME LATTICE
( 1 ( ( H S * P ) 3 2 ) ( ( S ) 8 4 ) ( ( ZY KY HY ) 15 5 ) )
( 2 ( ( I ) 3 3 ) )
( 3 ( ( R S ) 1 4 ) ( ( Y HY ZY ) 8 5 ) )
( 4 ( ( E NN I ) 12 6 ) ( ( E NN I ) 15 8 ) )
( 5 ( ( NN E ) 4 6 ) )
( 6 ( ( * R ) 1 7 ) )
( 7 ( ( U NN E ) 1 8 ) )
( 8 ( ( * G M N R ) 3 9 ) )
( 9 ( ( O ) 15 10 ) )
( 10 ( ( S K H ) 6 11 ) )
( 11 ( ( A ) 3 12 ) )
( 12 ( ( T K S ) 5 13 ) )
( 13 ( ( A ) 5 14 ) )
( 14 ( ( T K S ) 5 15 ) )
( 15 ( ( A ) 3 16 ) )
( 16 ( ( R G W N Z ) 3 17 ) )
( 17 ( ( A ) 8 END ) )
**** RECOGNITION PROCESS ****
( 2 ( 1 ) 3 4 )
( 10 ( 3 4 ) 6 7 8 9 )
( 2 ( 6 7 8 9 ) 9 8 )
( NNCHI ( 9 8 ) 11 )
( 5 ( 6 7 8 9 ) 10 9 )
( 6 ( 6 7 8 9 ) 11 )
( 10 ( 1 ) 6 7 8 9 )
( SHIZUOKA ( 1 ) 12 )
( HATSU ( 12 ) 15 )
( WA ( 15 ) 16 17 )
( GA ( 15 ) 16 )
( SHINOSAKA ( 1 ) 14 )
( KARA ( 14 ) END 17 )
( HIMEJI ( 1 ) 8 9 )
RECOGNITION RESULT
(( 1 ) NICHU ) ( SHINOSAKA ) KARA ) ( --- MADE ) ( --- JI --- FUN HATSU ) ( --- --- GO ) ( --- SEKI ) ( --- MAI )
```

(HAKATAMADENO)
 PHONEME LATTICE
 (1 ((S B * R H) 3 2))
 (2 ((A) 1 3))
 (3 ((K T S) 5 4))
 (4 ((A) 4 5))
 (5 ((S K T) 7 6))
 (6 ((A) 4 7))
 (7 ((R M N G Z) 4 8))
 (8 ((A) 4 9))
 (9 ((K D) 5 10))
 (10 ((E U) 4 11))
 (11 ((Z M N G H) 4 12))
 (12 ((O) 3 END))

**** RECOGNITION PROCESS ****
 (YOKKA (1) 5)
 (HATSUKA (1) 5)
 (ASU (1) 4)
 (ASHITA (1) 5)
 (KOKURA (1) 5)
 (HAKATA (1) 7 8 9)
 (MADE (7 8 9) 11 12)
 (NO (11 12) END)

RECOGNITION RESULT
 (((1) NICH I) ((SHINOSAKA) KARA) ((HAKATA) MADE) (--- JI --- FUN HATSU) (--- --- GO) (--- SEKI) (--- MAI))

(ROKUJINIHUNHATSUNO)
 PHONEME LATTICE
 (1 ((B R M N H) 3 2))
 (2 ((U A O) 3 3))
 (3 ((T K S) 7 4))
 (4 ((U) 4 5))
 (5 ((S H Z R) 4 6))
 (6 ((E I) 5 7))
 (7 ((Z M H N R) 7 8))
 (8 ((I E) 4 9))
 (9 ((S K) 7 10))
 (10 ((U) 11 11))
 (11 ((Z H S M) 5 12))
 (12 ((A) 3 13))
 (13 ((T K S P) 9 14))
 (14 ((U) 4 15))
 (15 ((R G M Z N) 3 16))
 (16 ((O) 7 END))

**** RECOGNITION PROCESS ****
 (FUTSUKA (1) 5 6)
 (MUIKA (1) 5 6)
 (HATSUKA (1) 5 6)
 (ASU (1) 5 6 4)
 (2 (1) 3)
 (6 (1) 5 6 4)
 (JI (5 6 4) 7 6)
 (0 (7 6) 9 7)
 (FUN (9 7) 12 11)
 (HATSU (12 11) 15 16 14)
 (NO (15 16 14) END)
 (RE (7 6) 9 7)
 (FUN (9 7) 12 11)
 (2 (7 6) 9 7 6)
 (I (7 6) 7 6)
 (FUN (9 7 6) 12 11)
 (ZI (9 7 6) 9)
 (ZI (7 6) 9)
 (8 (1) 5 6 7 4)
 (JI (5 6 7 4) 7 6 9)
 (PAUSE (7 6 9) 10)
 (0 (7 6 9) 9 7)
 (RE (7 6 9) 9 7)
 (2 (7 6 9) 9 7 6)
 (I (7 6 9) 7 6)
 (9 (7 6 9) 11)
 (ZI (7 6 9) 9)
 (KU (1) 3 2)
 (HIKARI (1) 7)
 (2 (7) 9)
 (100 (7) 11 10)
 (8 (11 10) 15 16 14)
 (GO (15 16 14) END)
 (FUTSU (1) 5 6 7 3)
 (KEN (5 6 7 3) 8 7)
 (SEKI (5 6 7 3) 9)

RECOGNITION RESULT
 (((1) NICH I) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6 8) JI (2 0) FUN HATSU) (--- --- GO) (--- SEKI) (--- MAI))

(HIKARIZYUKYUGONO)
 PHONEME LATTICE
 (1 ((P H S * R) 1 2) ((S) 10 4))
 (2 ((I) 2 3))
 (3 ((K T S) 6 4))
 (4 ((A) 5 5))
 (5 ((R Z G M N) 3 6))
 (6 ((E I) 3 7) ((I E) 7 9))
 (7 ((*) 1 8))
 (8 ((I) 3 9))
 (9 ((Z H S *) 5 10) ((ZY KY) 13 13))
 (10 ((I E) 4 11) ((U) 14 14))
 (11 ((Y * R) 1 12))
 (12 ((U O) 8 14))
 (13 ((U O) 6 14))
 (14 ((T K) 7 15) ((T K S) 12 17) ((ZY KY) 17 20))
 (15 ((I) 3 16))
 (16 ((R S) 1 17) ((Y HY) 6 20))
 (17 ((E) 2 18) ((E U) 5 21) ((U) 13 23))
 (18 ((Y *) 1 19))
 (19 ((U O) 10 23))
 (20 ((U O) 8 23))
 (21 ((* R) 1 22))
 (22 ((U) 6 23))
 (23 ((* G M R N) 4 24))
 (24 ((O) 15 25))
 (25 ((Z H * G M) 5 26))
 (26 ((O) 1 END))
 **** RECOGNITION PROCESS ****
 (FUTSUKA (1) 5 6)
 (MIKKA (1) 5)
 (ITSUKA (1) 5 6)
 (MUIKA (1) 5)
 (2 (1) 3)
 (3 (1) 6 7 8 5)
 (10 (6 7 8 5) 14)
 (1 (14) 16 17 18 19 21 22)
 (KU (1) 3 2 4)
 (HIKARI (1) 7 8 9 10 11 12 6)
 (1 (7 8 9 10 11 12 6) 11 12 10)
 (GO (11 12 10) 14)
 (2 (7 8 9 10 11 12 6) 9 10 11 12 14 7 8 6)
 (GO (9 10 11 12 14 7 8 6) 14)
 (4 (7 8 9 10 11 12 6) 14)
 (5 (7 8 9 10 11 12 6) 14 12)
 (6 (7 8 9 10 11 12 6) 21 22 23 24 17 16 18 19 15)
 (GO (21 22 23 24 17 16 18 19 15) 25 26 END)
 (NO (25 26 END) END)
 (9 (7 8 9 10 11 12 6) 14)
 (10 (9 10 11 12 14 7 8 6) 14 23 24 25 26)
 (1 (14 23 24 25 26) 16 17 18 19 21 22)
 (GO (16 17 18 19 21 22) 25 26 END)
 (2 (14 23 24 25 26) 24)
 (4 (14 23 24 25 26) 26 END 25)
 (5 (14 23 24 25 26) 25 26 END 24)
 (6 (14 23 24 25 26) END)
 (9 (14 23 24 25 26) 23 24 25 26 21 22)
 (GO (23 24 25 26 21 22) 25 26 END)
 (GO (14 23 24 25 26) 25 26 END)
 (10 (14) 23 24 25 26)
 (2 (23 24 25 26) 24)
 (4 (23 24 25 26) 26 END 25)
 (5 (23 24 25 26) 25 26 END 24)
 (6 (23 24 25 26) END)
 (GO (23 24 25 26) 25 26 END)
 (10 (7 8 9 10 11 12 6) 14)
 (9 (14) 23 24 25 26 21 22)

RECOGNITION RESULT
 ((1) NICH I) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6 8) JI (2 0) FUN HATSU) ((
 HIKARI) (19 90 6 20 29 11 21) GO) (--- SEKI) (--- MAI))

(SHITEKENO)
 PHONEME LATTICE
 (1 ((S H) 4 2) ((S) 12 4))
 (2 ((I) 5 3))
 (3 ((K T S) 6 4))
 (4 ((E) 11 5))
 (5 ((T K S) 9 6))
 (6 ((E) 8 7))
 (7 ((*) 1 8) ((*) 6 9) ((*) 10 11))
 (8 ((NN) 9 10))
 (9 ((NN U) 3 10) ((O) 10 END))
 (10 ((*) 1 11))
 (11 ((NN O) 6 END))
 **** RECOGNITION PROCESS ****
 (2 (1) 164)
 (SHITEI (1) 5 4 8 10 11 9 END 7 6)
 (KEN (5 4 8 10 11 9 END 7 6) 10 11 END 9 7 8)
 (NO (10 11 END 9 7 8) END 10 11 9)
 (1 (1) 4)
 RECOGNITION RESULT
 ((1) NICH I) ((SHINOSAKA) KARA) ((HAKATA) MADE) ((6 8) JI (2 0) FUN HATSU) ((
 HIKARI) (19 90 6 20 29 11 21) GO) ((FUTSU) SEKI) (--- MAI))

```

( KYUMAI )
PHONEME LATTICE
( 1 ( ( S T H K ) 6 2 ) ( ( Y KY HY ) 11 3 ) )
( 2 ( ( U E ) 12 4 ) )
( 3 ( ( U O ) 7 4 ) )
( 4 ( ( M N R G ) 4 5 ) )
( 5 ( ( A ) 4 6 ) ( ( A ) 9 END ) )
( 6 ( ( * ) 1 7 ) )
( 7 ( ( E I ) 3 END ) )
**** RECOGNITION PROCESS ****
( FUTSUKA ( 1 ) 6 7 END 5 )
( WA ( 6 7 END 5 ) 6 7 END 5 )
( ITSUKA ( 1 ) 6 7 END 5 )
( YOKA ( 1 ) 6 7 END 5 )
( KYO ( 1 ) 4 5 )
( NO ( 4 5 ) 6 7 )
( WA ( 4 5 ) 6 7 END 5 )
( 2 ( 1 ) 3 1 )
( 10 ( 3 1 ) 4 5 )
( 2 ( 4 5 ) 5 )
( 7 ( 4 5 ) 6 7 END 5 )
( 10 ( 1 ) 4 5 )
( SHITEI ( 1 ) 2 4 5 )
( NO ( 2 4 5 ) 6 7 )
( WA ( 2 4 5 ) 6 7 END 5 )
( FUTSU ( 1 ) 4 5 )
( NO ( 4 5 ) 6 7 )
( WA ( 4 5 ) 6 7 END 5 )
( 1 ( 1 ) 2 )
( 4 ( 1 ) 5 4 )
( MAI ( 5 4 ) END 5 6 7 )
( 9 ( 1 ) 4 5 )
( MAI ( 4 5 ) END 5 6 7 )
RECOGNITION RESULT
(( 1 ) NICH I (( SHINOSAKA ) KARA ) (( HAKATA ) MADE ) (( 6 8 ) JI ( 2 0 ) FUN HATSU ) ((
HIKARI ) ( 19 90 6 20 29 11 21 ) GO ) (( FUTSU ) SEKI ) (( 9 4 ) MAI ) )
( ONEGAISHIMASU )
PHONEME LATTICE
( 1 ( ( P H * R B ) 1 2 ) )
( 2 ( ( O U ) 2 3 ) )
( 3 ( ( R N G ) 2 4 ) )
( 4 ( ( E A ) 8 5 ) )
( 5 ( ( R G W M N ) 3 6 ) )
( 6 ( ( A O ) 6 7 ) )
( 7 ( ( * ) 1 8 ) )
( 8 ( ( E ) 4 9 ) )
( 9 ( ( S H ) 6 10 ) )
( 10 ( ( E I ) 1 11 ) )
( 11 ( ( M N R G Z ) 4 12 ) )
( 12 ( ( A ) 4 END ) )
**** RECOGNITION PROCESS ****
( ONEGAI ( 1 ) 9 7 8 )
( SHI ( 9 7 8 ) 11 12 10 )
( MAS ( 11 12 10 ) END )
RECOGNITION RESULT ( WITHOUT INFERENCE )
(( 1 ) NICH I (( SHINOSAKA ) KARA ) (( HAKATA ) MADE ) (( 6 8 ) JI ( 2 0 ) FUN HATSU ) ((
HIKARI ) ( 19 90 6 20 29 11 21 ) GO ) (( FUTSU ) SEKI ) (( 9 4 ) MAI ) )
PENALTY SCORES FOR RECOGNIZED ITEMS
(( 2 ) (( 0 ) (( 0 ) (( 4 5 ) JI ( 4 5 ) ) (( 1 ) ( 1 1 1 2 2 6 7 ) ) ((
3 ) (( 0 1 ) ) )
RESPONSE SENTENCE
9 MAI DESUKA
FINAL RESULT AFTER INFERENCE
(( 1 ) NICH I (( SHINOSAKA ) KARA ) (( HAKATA ) MADE ) (( 6 ) JI ( 2 ) FUN HATSU ) (( HIKARI
) ( 19 ) GO ) (( FUTSU ) SEKI ) (( 9 4 ) MAI ) )
PENALTY SCORES
(( 2 ) (( 0 ) (( 0 ) (( 4 ) JI ( 4 ) ) (( 100 ) ( 100 ) ) (( 3 ) (( 0 1 ) ) )

```

付録 11. Voice Q-A System II-2 における発声者別の 認識実験結果

(単独に区切って発声した母音から推定した母音標準パターン使用)
(発声データ : 付録 3 のグループ 1 の 20 文章)

発声者 : RN

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	0 2	0 1	0 1	0 2	Q	0 2	2	5	0	0	0
2		Q	0 1	0 1		Q	Q	3	2	0	0	0
3	0 1	0 1	0 1		0 2	Q	Q	2	4	0	0	0
4	0 1	0 2	0 1		0 1	Q	Q	2	4	0	0	0
5	0 1	Q	0 1		Q	Q	Q	4	2	0	0	0
6	0 1	Q	Q		Q	Q	0 1	4	2	0	0	0
7		0 1	0 1	0 2		Q		1	3	0	0	0
8	0 1	0 2	0 1	0 2		Q		1	4	0	0	0
9	0 1	Q	0 1	0 1	0 1	Q		2	4	0	0	0
10		0 2	Q	0 7		Q	Q	3	2	0	0	0
11	X	Q	Q		0 3	X	Q	3	1	0	2	0
12	0 1	0 2	0 1	0 1		Q	Q	2	4	0	0	0
13	0 1	Q	0 1	Q		Q	Q	4	2	0	0	0
14	0 1	Q		Q	Q	0 1	Q	4	2	0	0	0
15	Q	0 2	Q		Q	Q	0 1	4	2	0	0	0
16		Q	0 1			Q	Q	3	1	0	0	0
17	0 4	Q	0 1	Q	Q	Q	Q	5	2	0	0	0
18	0 1	0 1	Q		0 1	0 1	0 1	1	5	0	0	0
19	0 1	0 2	Q	0 1		Q		2	3	0	0	0
20	0 2	0 2	0 1	0 2	0 3	Q	0 1	1	6	0	0	0
TOTAL (115)												
	16	20	19	12	13	20	15	TOTAL				
CORRECT	15	20	19	12	13	19	15	113	98.3%			
UNIQUE	(2)	(9)	(6)	(3)	(6)	(17)	(10)	(53)	< 46.9 % >			
=1	(11)	(3)	(13)	(5)	(3)	(2)	(4)	(41)	< 68.3 % >			
>1	(2)	(8)	(0)	(4)	(4)	(0)	(1)	(19)	< 93.3 % >			
									< 96.7 % >			
REJECT	0	0	0	0	0	0	0	0	0.0%			
ERROR	1	0	0	0	0	1	0	2	1.7%			

(表の見方は付録 8 を参照)

発声者：HN

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	0 2	Q	Q	Q	Q	Q	-	5	1	1	0	0
2		Q	0 1	0 1	Q	0 1	0 1	1	4	0	0	0
3	0 1	0 1	0 1		0 1	Q	Q	2	4	0	0	0
4	Q	0 1	0 1		Q	Q	Q	4	2	0	0	0
5	0 2	0 2	-		Q	Q	0 1	2	3	1	0	0
6	0 1	0 1	0 1		Q	Q	0 1	2	4	0	0	0
7	0 2	0 2	Q		Q	Q		3	1	0	0	0
8	0 1	0 1	0 1	0 2		0 1		0	5	0	0	0
9	0 1	Q	0 2	Q	Q	Q		4	2	0	0	0
10		Q	Q	0 1	Q	Q	Q	4	1	0	0	0
11	0 2	0 2	Q		0 1	X	0 2	1	4	0	1	0
12	Q	0 1	0 1	0 1	Q	Q	0 1	2	4	0	0	0
13	0 1	0 1	0 1	0 4	0 1	Q		1	5	0	0	0
14	0 1	Q		Q	Q	Q	0 1	4	2	0	0	0
15	Q	0 2	0 1		Q	Q	Q	4	2	0	0	0
16		Q	Q		Q	Q	0 1	3	1	0	0	0
17	0 2	0 1	Q	0 5	0 1	0 1	0 1	1	6	0	0	0
18	0 2	Q	0 1		X	Q	0 1	2	3	0	1	0
19	Q	Q	0 1	0 1		Q		3	2	0	0	0
20	0 3	0 2	Q	0 2	0 1	-	Q	2	4	1	0	0

TOTAL (115)								TOTAL	
	16	20	19	12	13	20	15	110	95.7%
CORRECT	16	20	18	12	12	18	14	(50)	< 45.5 % >
UNIQUE	(4)	(9)	(6)	(4)	(7)	(15)	(5)	(43)	< 71.7 % >
=1	(6)	(7)	(10)	(4)	(5)	(3)	(8)	(17)	< 95.0 % >
>1	(6)	(4)	(2)	(4)	(0)	(0)	(1)		< 96.7 % >
REJECT	0	0	1	0	0	1	1	3	2.6 %
ERROR	0	0	0	0	1	1	0	2	1.7 %

発声者：KI

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	Q	0 2	Q	Q	Q	0 2	5	2	0	0	0
2		Q	0 1	Q	Q	Q	0 1	3	2	0	0	0
3	0 1	Q	Q		Q	Q	Q	5	1	0	0	0
4	0 1	0 1	Q		0 3	Q	0 1	2	4	0	0	0
5	0 1	Q	Q		Q	Q	0 1	4	2	0	0	0
6	0 1	0 1	Q		Q	Q	0 1	3	3	0	0	0
7		Q	0 1	Q		Q		3	1	0	0	0
8	0 3	0 1	Q	0 1		Q		2	3	0	0	0
9	Q	Q	0 1	0 1	0 1	Q		3	3	0	0	0
10		Q	0 2	X		Q	0 1	2	2	0	1	0
11	0 4	0 1	Q		Q	Q	Q	4	2	0	0	0
12	Q	Q	Q	Q		Q	0 1	5	1	0	0	0
13	0 1	Q	0 1	0 2	0 1	Q		2	4	0	0	0
14	0 1	Q		0 1	Q	Q	0 1	3	3	0	0	0
15	0 1	0 2	0 1		Q	Q	0 1	2	4	0	0	0
16		Q	Q		Q	Q	0 1	3	1	0	0	0
17	0 3	Q	Q	Q	Q	Q	0 1	5	2	0	0	0
18	0 1	Q	Q		0 1	0 1	0 1	2	4	0	0	0
19	0 1	Q	0 1	0 1		Q		2	3	0	0	0
20	0 1	0 2	0 1	0 1	0 4	Q	Q	2	5	0	0	0

TOTAL (115)								TOTAL	
	16	20	19	12	13	20	15	114	99.1%
CORRECT	16	20	19	11	13	20	15	(62)	< 54.4 % >
UNIQUE	(3)	(14)	(10)	(5)	(8)	(19)	(3)	(41)	< 78.8 % >
=1	(10)	(4)	(7)	(5)	(3)	(1)	(11)	(11)	< 90.4 % >
>1	(3)	(2)	(2)	(1)	(2)	(0)	(1)		< 96.2 % >
REJECT	0	0	0	0	0	0	0	0	0.0 %
ERROR	0	0	0	1	0	0	0	1	0.9 %

発声者：KS

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	Q	Q	Q	Q	Q	0 1		6	1	0	0
2		Q	Q	Q		0 1	0 1		3	2	0	0
3	0 1	0 2	0 2		0 1	Q	0 1		1	5	0	0
4	0 1	0 1	0 2		0 1	Q	0 1		1	5	0	0
5	0 1	Q	Q		Q	Q	Q		5	1	0	0
6	0 1	0 1	0 1		Q	Q	0 3		2	4	0	0
7		Q	0 2	Q		0 1			2	2	0	0
8	0 3	0 2	Q	0 4		Q			2	3	0	0
9	0 1	0 1	-	0 1	Q	Q			2	3	1	0
10		0 1	Q	0 2		Q	0 3		2	3	0	0
11	0 2	0 1	Q		Q	Q	0 1		3	3	0	0
12	0 2	0 2	0 1	0 1		Q	Q		2	4	0	0
13	0 1	Q	Q	0 4	0 1	Q			3	3	0	0
14	0 2	Q		Q	Q	Q	0 1		4	2	0	0
15	0 1	0 2	Q		0 1	Q	0 1		2	4	0	0
16		0 1	Q			Q	0 1		2	2	0	0
17	0 4	0 1	Q	0 3	0 2	Q	Q		3	4	0	0
18	0 3	Q	Q		0 1	0 1	0 1		2	4	0	0
19	0 1	0 1	Q	0 1		Q			2	3	0	0
20	0 1	0 2	0 1	0 3	0 4	Q	0 1		1	6	0	0

TOTAL (115)								TOTAL	Q	0	-	X
	16	20	19	12	13	20	15	114	99.1%			
CORRECT	16	20	18	12	13	20	15	114	99.1%			
UNIQUE	(1)	(7)	(12)	(4)	(6)	(17)	(3)	(50)	< 43.9 % >			
=1	(9)	(8)	(3)	(3)	(5)	(3)	(10)	(41)	< 64.1 % >			
>1	(6)	(5)	(3)	(5)	(2)	(0)	(2)	(23)	< 84.4 % >			
REJECT	0	0	1	0	0	0	0	1	0.9 %			
ERROR	0	0	0	0	0	0	0	0	0.0 %			

発声者：SA

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	0 1	Q	Q	Q	Q	Q	0 1		5	2	0	0
2		Q	0 1	0 1		Q	0 1		2	3	0	0
3	0 1	Q	Q		Q	Q	0 1		4	2	0	0
4	0 1	0 1	0 1		Q	Q	0 1		2	4	0	0
5	0 1	0 2	0 1		0 1	Q	0 1		1	5	0	0
6	0 1	0 2	Q		Q	Q	0 1		3	3	0	0
7		Q	0 1	Q		Q			3	1	0	0
8	0 2	0 2	X	0 1		Q			1	3	0	1
9	0 1	0 2	0 2	0 1	0 1	Q			1	5	0	0
10		Q	Q	0 1 1		Q	0 1		3	2	0	0
11	0 3	0 1	X		0 1	Q	Q		2	3	0	1
12	0 2	0 1	Q	-		Q	0 1		2	3	1	0
13	0 1	Q	0 1	0 6	0 3	Q			2	4	0	0
14	0 1	Q		0 1	Q	Q	0 1		3	3	0	0
15	0 2	0 2	0 1		0 1	Q	0 1		1	5	0	0
16		Q	-			Q	Q		3	0	1	0
17	0 3	0 1	Q	0 1	0 2	0 1	Q		2	5	0	0
18	0 1	X	Q		0 1	0 1	0 1		1	4	0	1
19	0 1	0 1	Q	0 1		Q			2	3	0	0
20	0 2	0 2	-	X	0 2	Q	Q		2	3	1	1

TOTAL (115)								TOTAL	Q	0	-	X
	16	20	19	12	13	20	15	108	93.9%			
CORRECT	16	19	15	10	13	20	15	108	93.9%			
UNIQUE	(0)	(8)	(8)	(2)	(5)	(18)	(4)	(45)	< 41.7 % >			
=1	(10)	(5)	(6)	(6)	(5)	(2)	(11)	(45)	< 71.4 % >			
>1	(6)	(6)	(1)	(2)	(3)	(0)	(0)	(18)	< 92.1 % >			
REJECT	0	0	2	1	0	0	0	3	2.6 %			
ERROR	0	1	2	1	0	0	0	4	3.5 %			

発声者：MK

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	0 1	0 1	Q	0 1	0 5	Q	0 1	2	5	0	0	0
2		0 2	0 1	0 2		Q	0 1	1	4	0	0	0
3	0 1	0 2	Q		Q	Q	Q	4	2	0	0	0
4	0 1	0 1	0 1		0 1	Q	0 1	1	5	0	0	0
5	0 1	Q	Q		Q	Q	Q	5	1	0	0	0
6	0 2	0 1	Q		Q	Q	0 1	3	3	0	0	0
7		Q	0 4	Q		-		2	1	1	0	0
8	0 1	0 2	0 1	0 3		0 1		0	5	0	0	0
9	Q	Q	0 1	0 3	0 1	Q		3	3	0	0	0
10		Q	0 1	0 3		Q	Q	3	2	0	0	0
11	0 2	0 1	Q		Q	0 1	Q	3	3	0	0	0
12	0 1	0 1	0 1	0 1		Q	0 1	1	5	0	0	0
13	Q	0 1	0 1	X	0 2	Q		2	3	0	1	0
14	0 1	Q		Q	Q	0 2	Q	4	2	0	0	0
15	X	0 2	0 1		0 1	Q	Q	2	3	0	1	0
16		0 1	0 1			Q	0 1	1	3	0	0	0
17	0 3	0 2	0 1	0 3	Q	0 1	0 1	1	6	0	0	0
18	Q	0 1	0 1		0 1	Q	0 1	2	4	0	0	0
19	0 2	Q	Q	0 1		Q		3	2	0	0	0
20	0 1	0 2	0 1	0 1	-	Q	Q	2	4	1	0	0

TOTAL (115)								TOTAL	
	16	20	19	12	13	20	15	111	96.5%
CORRECT	15	20	19	11	12	19	15	(45)	< 40.5 % >
UNIQUE	(3)	(6)	(6)	(2)	(6)	(15)	(7)	(47)	< 71.2 % >
=1	(8)	(8)	(12)	(4)	(4)	(3)	(8)	(19)	< 89.4 % >
>1	(4)	(6)	(1)	(5)	(2)	(1)	(0)		< 97.0 % >
REJECT	0	0	0	0	1	1	0	2	1.7 %
ERROR	1	0	0	1	0	0	0	2	1.7 %

発声者：NA

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	Q	0 1	Q	Q	Q	Q	6	1	0	0	0
2		Q	0 1	Q		Q	0 1	3	2	0	0	0
3	0 2	0 1	Q		Q	Q	Q	4	2	0	0	0
4	0 1	0 1	Q		Q	Q	Q	4	2	0	0	0
5	0 1	Q	Q		0 1	Q	0 1	3	3	0	0	0
6	0 1	0 1	Q		0 1	Q	Q	3	3	0	0	0
7		0 2	0 2	0 1		Q		1	3	0	0	0
8	0 2	0 2	Q	0 1		Q		2	3	0	0	0
9	0 1	Q	-	Q	Q	Q		4	1	1	0	0
10		0 3	0 1	0 4		Q	Q	2	3	0	0	0
11	0 1	0 1	Q		0 1	0 1	Q	2	4	0	0	0
12	-	0 2	Q	0 1		0 1	Q	2	3	1	0	0
13	0 1	0 1	0 1	0 4	0 1	Q		1	5	0	0	0
14	0 1	Q		0 4	Q	Q	0 1	3	3	0	0	0
15	0 1	0 2	0 1		0 1	Q	0 1	1	5	0	0	0
16		Q	Q			Q	Q	4	0	0	0	0
17	0 1	Q	Q	Q	Q	Q	0 1	5	2	0	0	0
18	0 1	0 1	Q		0 1	0 1	0 1	1	5	0	0	0
19	0 1	0 3	0 1	0 1		Q		1	4	0	0	0
20	0 2	0 2	0 1	0 1	Q	Q	0 1	2	5	0	0	0

TOTAL (115)								TOTAL	
	16	20	19	12	13	20	15	113	98.3%
CORRECT	15	20	18	12	13	20	15	(54)	< 47.8 % >
UNIQUE	(1)	(7)	(10)	(4)	(7)	(17)	(8)	(45)	< 76.3 % >
=1	(11)	(6)	(7)	(5)	(6)	(3)	(7)	(14)	< 89.8 % >
>1	(3)	(7)	(1)	(3)	(0)	(0)	(0)		< 93.2 % >
REJECT	1	0	1	0	0	0	0	2	1.7 %
ERROR	0	0	0	0	0	0	0	0	0.0 %

発声者：KH

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	Q	Q	Q	Q	Q	0 1	6	1	0	0	0
2	Q	Q	0 1	Q	Q	0 2	Q	3	2	0	0	0
3	0 1	0 1	Q	Q	Q	Q	Q	4	2	0	0	0
4	0 1	0 1	0 1	Q	0 1	Q	0 1	1	5	0	0	0
5	0 1	X	X	Q	Q	Q	0 1	2	2	0	2	0
6	0 1	0 1	Q	0 1	Q	Q	0 1	2	4	0	0	0
7	Q	0 1	0 3	0 1	Q	Q	Q	1	3	0	0	0
8	0 1	0 3	0 1	0 1	Q	Q	Q	1	4	0	0	0
9	Q	Q	Q	0 2	0 2	Q	Q	4	2	0	0	0
10	Q	0 1	0 1	0 2	Q	Q	Q	2	3	0	0	0
11	-	0 1	X	Q	0 1	0 2	0 1	0	4	1	1	1
12	0 1	X	Q	0 1	Q	Q	0 1	2	3	0	1	1
13	Q	0 1	Q	0 6	0 2	Q	Q	3	3	0	0	0
14	0 1	Q	Q	Q	Q	0 2	0 1	3	3	0	0	0
15	0 3	0 2	0 1	Q	0 1	Q	0 1	1	5	0	0	0
16	Q	Q	-	Q	Q	Q	Q	3	0	1	0	0
17	0 1	0 1	Q	0 1	0 1	0 2	0 1	1	6	0	0	0
18	X	0 2	0 1	Q	0 3	0 1	Q	1	4	0	1	1
19	X	Q	0 2	0 1	Q	Q	Q	2	2	0	1	1
20	0 2	0 2	0 2	0 2	0 1	Q	0 1	1	6	0	0	0

TOTAL (115)								TOTAL			
CORRECT	16	20	19	12	13	20	15	107	93.0%		
UNIQUE	(3)	(6)	(7)	(3)	(4)	(15)	(5)	(43)	< 40.2 % >		
=1	(8)	(8)	(6)	(5)	(6)	(1)	(10)	(44)	< 68.8 % >		
>1	(2)	(4)	(3)	(4)	(3)	(4)	(0)	(20)	< 92.2 % >		
REJECT	1	0	1	0	0	0	0	2	< 98.4 % >		
ERROR	2	2	2	0	0	0	0	6	1.7 %		
									5.2 %		

発声者：MS

	DATE	START	ARRIVE	TIME	TRAIN	SEAT	TICKET	TOTAL	Q	0	-	X
1	Q	0 2	0 2	0 1	0 1	0 1	0 1	1	6	0	0	0
2	Q	0 4	0 1	0 4	Q	Q	0 1	1	4	0	0	0
3	0 1	X	Q	Q	Q	Q	Q	4	1	0	1	1
4	0 1	0 1	0 2	Q	Q	Q	0 1	2	4	0	0	0
5	Q	Q	Q	Q	Q	Q	0 1	5	1	0	0	0
6	0 1	Q	0 3	0 5	Q	Q	0 2	2	4	0	0	0
7	Q	Q	0 3	Q	Q	0 1	Q	2	2	0	0	0
8	0 2	0 2	0 1	0 1	Q	Q	0 1	0	5	0	0	0
9	Q	Q	0 1	Q	Q	Q	Q	5	1	0	0	0
10	Q	0 1	Q	0 3	Q	Q	0 1	2	3	0	0	0
11	0 6	0 1	Q	Q	0 1	Q	0 1	2	4	0	0	0
12	0 1	0 3	0 2	0 1	Q	Q	0 1	1	5	0	0	0
13	0 1	Q	0 1	Q	Q	Q	Q	4	2	0	0	0
14	0 3	0 1	Q	0 4	Q	0 1	0 1	1	5	0	0	0
15	0 1	0 3	0 1	Q	0 1	Q	0 1	1	5	0	0	0
16	Q	0 1	0 1	Q	Q	Q	Q	2	2	0	0	0
17	0 2	0 2	0 1	0 1	X	0 1	0 1	0	6	0	1	1
18	0 7	0 1	0 1	Q	0 1	0 1	0 1	0	6	0	0	0
19	0 4	0 2	0 2	0 2	Q	Q	Q	1	4	0	0	0
20	0 2	0 2	Q	0 4	0 3	Q	0 1	2	5	0	0	0

TOTAL (115)								TOTAL			
CORRECT	16	20	19	12	13	20	15	113	98.3%		
UNIQUE	(3)	(5)	(5)	(3)	(6)	(14)	(2)	(38)	< 33.6 % >		
=1	(6)	(6)	(8)	(4)	(4)	(6)	(12)	(46)	< 61.3 % >		
>1	(7)	(8)	(6)	(5)	(2)	(0)	(1)	(29)	< 80.0 % >		
REJECT	0	0	0	0	0	0	0	0	< 89.3 % >		
ERROR	0	1	0	0	1	0	0	2	0.0 %		
									1.7 %		