

Title	SPEECH-DB利用説明書
Author(s)	産業科学研究科
Citation	大阪大学大型計算機センターニュース. 1985, 58, p. 77-105
Version Type	VoR
URL	https://hdl.handle.net/11094/65661
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

S P E E C H—D B 利用説明書

1. ま え が き

SPEECH-DB (1)(2)は音声認識研究を支援する為に構築された音声データベースである。現在、特定話者・小語彙認識装置の商品開発が行われている。これを更に発展させて、不特定話者・大語彙向きシステムへ移行するには音韻単位の認識を基礎とする必要があることは自明なことであるが、音声中の音韻は話者による影響はもちろんのこと、調音結合の影響により大きく左右され、これ等の変動の程度及び性質を把握することは、不特定話者・大語彙向きシステムの開発研究を行う上で不可欠なこととなる。この為には大量の音声データの収集と解析が必要となるが、現実の音声研究の環境では各研究者が独自に音声データを収集し、しかも用途に応じて同一データを重複して保持したり、又検索手段がない為、希望するデータを入手するのに多くの労力を要する等の問題があった。このような既存の音声データ管理方式は上記の目的の為には不十分であると言わねばならない。SPEECH-DBはこのような今後の音声認識研究を遂行するにあたって必要な音声データの管理・検索機能を兼ね備えた本格的な音声データベースである。

SPEECH-DBは、任意の音韻環境の下での音韻の検索はもちろん、その他、様々な音声データを検索する機能、検索された波形のグラフィック表示、その簡単な分析機能、さらに検索データの利用者ファイルへ転送する機能等を持っているが、これ等は全てIQLと呼ばれるコマンドを用いることにより、利用者は容易に(最小限の予備知識で)行うことができる。

音声データには、「標準データ」と呼べるものは存在せず、従ってSPEECH-DBに格納されているデータは我々が独自の基準で選択したものである。それ故、各利用者が真に要求する音声データが必ずしも格納されているとは限らない。ことことを考慮して、SPEECH-DBはデータの格納に関しても利用者の参加を想定して設計されており、データの格納に必要な作業は全てコマンド化されている。

尚、本データベースは大阪大学大型計算機センターのACOS 1000上で、汎用DBMS, INQを用いて実現されている。

注意事項：現在INQのバージョンupによって、正常に動作しないコマンドがあります。

単語を検索項目とした場合、検索はできますが、表示のためのLISTコマンドが無限ループに陥って、表示しません。

2. SPEECH-DB入門

S：前章の説明でSPEECH-DBの大体のことはお分かりいただけましたか？

U：全然！ 普通の文献データベースとは大分性質が違うようなので、何となく分りにくいのですが。文献データベースですと、「このような問題を論じている文献」というように、

検索対象とそれが満足すべき条件のイメージがわくのですが、音声データベースと言われても何を条件にして、何を検索するのかよく分かりません。

S：音声で表現された文章は単語に分解され、単語は音節に分解され、そして音節は音韻に分解されます。このように、音声は色々な概念から構成されて……………。

U：音韻ってなんですか？

S：あっ、どうも。音韻とは、音声の基本的単位で、子声、母音等のことです。連続に発声された文章（連続音声）はマイクロに見れば子音、母音の連鎖として見ることができますし、マクロに見れば単語の系列と見るができるわけです。

U：それ位は言われなくても知っています。

S：このことを御存知なら、半分はお知りいただけただけのようなものです。

SPEECH-DBでは実質的な検索対象、即ち文献データベースの「文献」、はあくまでもサンプリングされた音声データそのものですが、概念的な検索対象は音声データそのものに、先程申しあげました音声の属性或いは見かたを考慮したもので、単語、音節、音韻というものになります。従って、「～のような単語(音節、音韻)音声データ」が検索対象なのです。

U：検索条件にはどのようなものがあるのですか？

S：検索条件となる属性（検索項目）は全部で76個あります。SPEECH-DBは基本的には音声認識研究を支援することを想定していますが、かなり幅広い検索要求に応じられるように設計されています。例えば、全検索について使えるものとして、発声者の性別、年齢、方言等があります。従って、「25才以上の関西方言の男性が発声した……」という検索ができます。単語を検索する場合には、その他に「～のような音節（音韻）を含む」と言った条件を付け加えられます。又、音韻を検索する場合ですと、「～という単語に含まれる」という条件も付け加えることができます。

U：「母音/a/と/i/には含まれた子音がほしい」というような検索もできるのですか？

S：できます。さらに、有声破裂音（*b, d, g*）、摩擦音（*s, c, z*）、鼻音（*m, n*）という様なグループ化されたカテゴリー名でも検索できるようになっています。

U：利用者は具体的にはどのような形態で音声データを操作できるのですか？

S：2つありまして、1個はINQのDMLという言語をFORTRANのサブルーチン形式で使うことにより、利用者の応用プログラムから直接アクセスする方法です。これにはDMLの知識及び、SPEECH-DBの論理構造に関する知識が必要で、まず初心者の方には困難だと思います。他の1つは、我々が開発したコマンド言語IQLを使って必要なデータを利用者のパーマネントファイルに転送する方法です。この方法では最小限の知識で必要な音声データを入手することができます。

3. SPEECH-DB 概説

3.1 音声データの階層性

音声は図 3.1 に示すような階層構造をなしていると考えられる。例えば、“今日の天気は曇り”という連続音声は、“今日”、“の”、“天気”、“は”、“曇り”と 5 つの単語に分けられ、単語“曇り”は“KU”、“MO”、“RI”と 3 つの音節になる。更に音節“RI”は“R”、“R-I”、“I”と音韻に分けることができる。但し、“R-I”は R から I への遷移部を意味している。VCV 音節（母音+子音+母音）も階層構造の中では音節と同レベルであると考えられる。

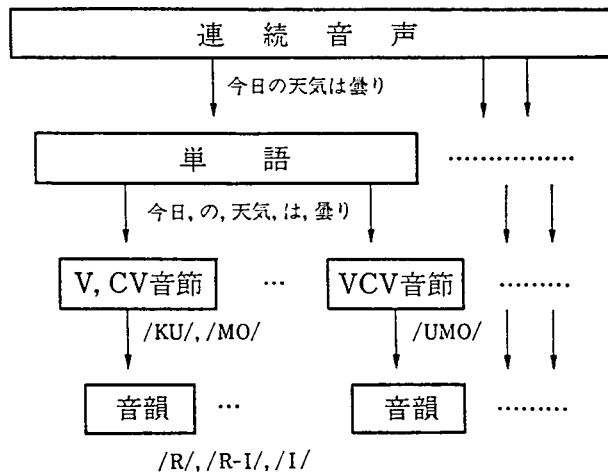


図 3.1 音声の階層構造

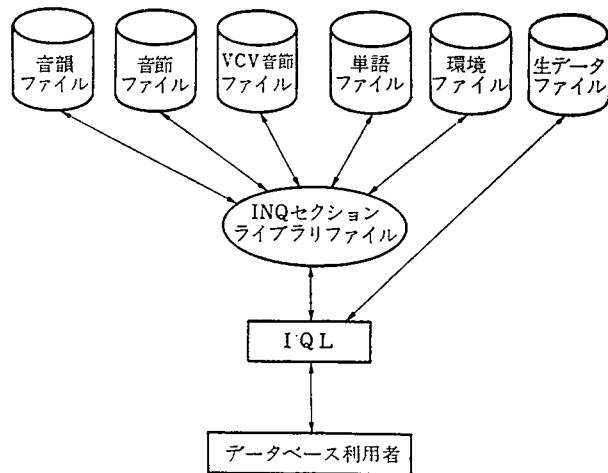


図 3.2 SPEECH-DBのファイル構成

本データベースはこのような音声の階層構造を暗黙に仮定して構成されている。このことは利用者が音声データを検索する際に知っていなければならない。

3.2 SPEECH-DBの基本構成

3.1節で述べたような階層構造を持つ音声の情報を効率よく格納する為に、図3.2に示すデータベース構成を行った。

現在本データベースは音韻、音節、V C V音節、単語、環境、音声データの6つの独立したファイルから構成されている。音声データファイルは普通のランダムファイルで、10 kHzでサンプリングした音声の生データを12ビットで量子化し、メモリの節約の為に1ワード(36ビット)に3サンプル点ずつ格納している。他の5つのファイルは、音声データの種々の属性値が格納されている。

3.3 格納項目

SPEECH-DBに格納されている項目の数は、音韻ファイルに14、音節ファイルに15、V C V音節ファイルに15、単語ファイルに14、環境ファイルに24の計82である。それぞれのファイルに格納されている項目の名前と省略名、意味を表3.1、3.2及び3.3に示す。表3.2において属という欄で、S、V、Wはそれぞれ音節ファイル、V C V音節ファイル、単語ファイルにその項目が格納されていることを意味する。環境ファイルの一部を除く76項目は全て検索の際の条件として使うことができる。

環境ファイルのSPEECH-CATEは音声のカテゴリーを格納する項目である。表3.4にSPEECH-CATEの項目値を示す。

3.4 音声データのセグメンテーション

音声情報のデータベースへの格納において一番問題となるのは、音韻のセグメンテーションとランク付けである。現在、SPEECH-DBではセグメンテーションを時間波形、ホルマント遷移の視察、更には実際に音声を耳で聞くことにより行っている。またランク付けに関しては、RANK1に定常(A)、非定常(B)という評価を格納している。各セグメントのラベル付けは概ね次の原則に従って行った。

(1)音韻表記は日本語モーラの記号を採用する。

(2) /M/と/N/は場合により母音と子音の両方に使い分ける。

例えば /GOHAN/, /EMPICU/のN/, /M/は母音、/NIMOCU/の
/N/, /M/は子音である。

(3)音韻間の遷移部分には特殊なラベルを用意する。

$/K A I / \rightarrow /K \cdot K - A \cdot A_{定} \cdot A - I \cdot I_{定} \cdot I_{非}/$

但し $\alpha - \beta$ は音韻 α から β への遷移部のラベル、 $\alpha_{定}$ 、 $\alpha_{非}$ は各々音韻 α の定常部、非定常部を示すラベルである。

一般的傾向として連続音声中の母音には非定常音が多くなり、またその多くが遷移部としてラベル付けされる。従って、検索に際して単にランクBの母音Vと指定するだけでは実際に存在する非定常音よりかなり少ないデータしか得られない。しかし、このような場合には、後述するIQLの中間一致機能を用いることによりV、 $V - \alpha$ 、 $\alpha - V$ とラベル付けされたすべての音韻を検索できるようになっている。表3.5にランクの決め方、図3.3にセグメンテーションの例を各々示す。

また、単語単位のセグメンテーションに関しては、厳密に文法的な意味では行っていない。“体言+附属語”の場合は“体言”、“附属語”と分けている。これは例えば“大阪という単語”といった検索が想定されるからである。しかし“用言+附属語”の場合には語幹だけを1つの単語として扱っている。中間一致機能を使って語幹により用言を検索することも可能である。

表 3.1 音韻ファイルの格納項目

項目名	省略名	意味
ID-FRAME	I F	フレームのIDコード
ID-SYLLABLE	I S	このフレームが属する音節のIDコード
ID-VVAL	I V	このフレームが属する VCV音節のIDコード
ID-WORD	I W	このフレームが属する単語のIDコード
ID	I D	このフレームが属する音声のIDコード
FNUM	F N	この音声内でのフレーム番号
PHONEME	P H	音韻表記
RANK1	RK1	定常、非定常の評価
RANK2	RK2	未使用
RMS	R M	エネルギー
ZERO-KOSA	Z K	零交差
PITCH	P T	ピッチ
DATA-FROM	D F	このフレームの開始点
DATA-TO	D T	このフレームの終点

表 3.2 音節, VCV 音節, 単語ファイルの格納項目

項目名	省略名	属	意味
ID-SYLLABLE ID-VCVSYLL ID-WORD	I S I V I W	S V W	この音節 (VCV音節, 単語) の IDコード
ID-WORD ID-WORD	I W I W	S V	この音節 (VCV音節) が属する単語の IDコード
ID ID ID	I D I D I D	S V W	この音節 (VCV音節, 単語) が属する音声の IDコード
SYLLABLE VCVSYLLABLE WORD	S Y VCV W D	S V W	音節名, VCV音節名, 単語名
SAV-RMS VAV-RMS WAV-RMS	SAR VAR WAR	S V W	この区間のエネルギーの平均
SMAX-RMS VMAX-RMS WMAX-RMS	SXR VXR WXR	S V W	この区間のエネルギーの最大値
SMIN-RMS VMIN-RMS WMIN-RMS	SNR VNR WNR	S V W	この区間のエネルギーの最小値
SAV-ZERO-KOSA VAV-ZERO-KOSA WAV-ZERO-KOSA	SAZ VAZ WAZ	S V W	この区間の零交差の平均
SMAX-ZERO-KOSA VMAX-ZERO-KOSA WMAX-ZERO-KOSA	SXZ VXZ WXZ	S V W	この区間の零交差の最大値
SMIN-ZERO-KOSA VMIN-ZERO-KOSA WMIN-ZERO-KOSA	SNZ VNZ WNZ	S V W	この区間の零交差の最小値
SAV-PITCH VAV-PITCH WAV-PITCH	SAP VAP WAP	S V W	この区間のピッチの平均
SMAX-PITCH VMAX-PITCH WMAX-PITCH	SXP VXP WXP	S V W	この区間のピッチの最大値
SMIN-PITCH VMIN-PITCH WMIN-PITCH	SNP VNP WNP	S V W	この区間のピッチの最小値
SDATA-FROM VDATA-FROM WDATA-FROM	SDF VDF WDF	S V W	音節 (VCV音節, 単語) の開始点
SDATA-TO VDATA-TO WDATA-TO	SDT VDT WDT	S V W	音節 (VCV音節, 単語) の終点

表 3.3 環境ファイルの格納項目

項目名	省略名	意味
ID	I D	この音声のIDコード
ID-SUBJECT	I B	話者コード
SUBJECT	S B	話者名
SUBJECT-SEX	S S	話者の性別
SUBJECT-AGE	SAG	話者の年齢
RECORD-DATE	R D	この音声を収録した年, 月, 日
SPEECH-CATE	S C	この音声のカテゴリ (連続音声, 単音節等)
SPEECH-DESC	S D	発声した内容
SAMPLING-FRQ	S F	サンプリング周波数
DIALECT	D L	話者の方言
FRAME-LENGTH	F L	対応する音韻ファイルのフレーム長
FRAME-SHIFT	F S	対応する音韻ファイルのフレームシフト間隔
FRAME-MAX	F M	この音声の最大フレーム数
EAV-RMS	EAR	この音声区間のエネルギーの平均
EMAX-RMS	EXR	この音声区間のエネルギーの最大値
EMIN-RMS	ENR	この音声区間のエネルギーの最小値
EAV-ZERO-KOSA	EAZ	この音声区間の零交差の平均
EMAX-ZERO-KOSA	EXZ	この音声区間の零交差の最大値
EMIN-ZERO-KOSA	ENZ	この音声区間の零交差の最小値
EAV-PITCH	EAP	この音声区間のピッチの平均
EMAX-PITCH	EXP	この音声区間のピッチの最大値
EMIN-PITCH	ENP	この音声区間のピッチの最小値
ADATA-FROM	ADF	この音声の開始点
ADATA-TO	ADT-	この音声の終点

表 3.4 SPEECH-CATE の項目値

カテゴリー	項目値
単音節	SY (SYLLABLE)
孤立単語	WD (WORD)
連続音声	CS (CONTINUOUS-SPEECH)

表 3.5 RANK I の決め方

	A	B	C	D
時間波形	定常	定常	非定常	非定常
ホルマント	定常	非定常	定常	非定常
RANK I	A	B	B	B

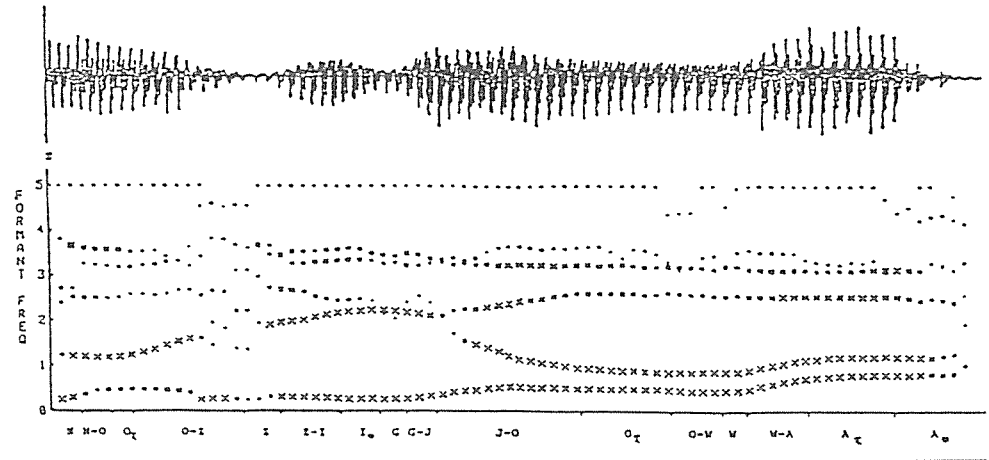


図 3.3 セグメンテーションの例

3.5 音声データ

現時点（1982年11月）においてSPEECH-DBには、表3.6に示される音声データが格納されている。

表 3.6

カテゴリー	時間表	発声者
5種類の連続音声	3.5秒/文	男性 3名 (女性1名:生データだけ)
525種類のVCV音節	0.8秒/音節	(男性1名:生データだけ)
単独発声された5母音	0.6秒/母音	男性 10名

連続音声の内訳を以下に示す。尚、これ等の連続音声は日本語の全子音を含んでいる。

- 1) 爆音が銀世界の高原に広がる、
- 2) 今日の天気は曇り後晴れるでしょう、
- 3) 朝御飯にパンと卵を食べました。
- 4) 午前の授業は地理と図工です。
- 5) 明日の試験にはペンと鉛筆を持参のこと。

4. IQLコマンド

表4.1にIQLコマンド及びその機能を示す。IQLにおける検索は、音声の生データファイルにおける開始点及び終点を知る為の操作と、その位置情報を用いた生データのグラフィック表示、或いは転送の為の操作との2段階の操作から成っている。さらに第1段の操作は検索条件を満たすレコードの検索とその項目（属性）値の利用者への提示の2段階に分れている。以下にコマンドの機能について述べる。

4.1 SEARCHコマンド

SEARCHは検索を行うコマンドである。詳細な説明に入る前に、具体例を用いてSEARCHコマンドの概略を述べる。

男の発声で /BA/を含む 単語が欲しい
SUBJECT-SEX=MALE SYLLABLE=BA CONDITION WORD;

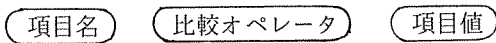
は次のように翻訳される。

CONDITION WORD; SUBJECT-SEX=MALE AND SYLLABLE = BA
省略形を使えば次のようにも書ける。

COND WD; SS = MALE AND SY = BA

初めのCONDITION (COND) は、次に検索条件が続くことを意味している。WORD (WD) は検索の単位を表わしている。IQLでは検索の単位として連続音声(CS), 単語(WD), CV音節(SY), VCV音節(VCV), 音韻(PH)の5種類を想定している。

“;”以後に検索対象が満足すべき条件が記述されるが、順序は任意であり、一般的には、



の形をしている。また前述の如く、IQLでは音声の階層性を仮定している為、 $PH < (SY, VCV) < WD < CS$ (但し $\alpha < \beta$ は β が α の上位にあることを示す)なる関係が暗黙に仮定されている。従って次のように、

WD ; SY = BA

と書けば、/BA/という音節を含む単語を意味し、

PH ; SY = DA

と書けば、音節/DA/に含まれる音韻を意味する。また検索対象よりも下のレベルの同じ項目名を次の例のようにANDで結んだ場合には、

WD ; SY = SA AND SY = KA

/SA/という音節を含みかつ/KA/という音節を含む単語と解釈される。

次にSEARCHコマンドの構文規則の詳細に関する説明を行う。図4.1にSEARCHコマンドの構文規則を示す。デリミターはすべてブランクとなっている。

検索は検索条件において“;”の前に書かれた項目の属するファイルの記録単位で行われる。即ち“IW”, “WD”等が書かれれば単語単位, “IS”, “SY”等の場合は音節単位で検索が行われる。また基本条件1において項目名、比較オペレータが等しい場合には省略することができる。

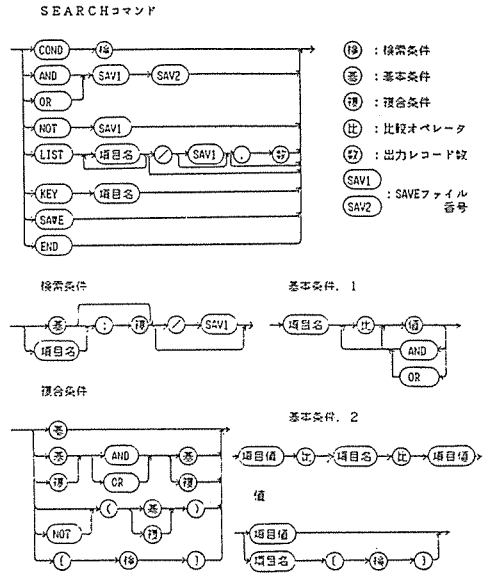


図4.1 SEARCHコマンドの構文規則

表 4.1 IQL コマンドとその機能

コマンド名	サブコマンド名	機 能
SEARCH	CONDITION	条件を満足するレコードを検索する
	AND OR NOT	SAVEファイル間の論理演算を行う
	LIST	検索されたレコードの項目値を表示する
	KEY	検索の対象となる項目値とその件数を表示する
	SAVE	レコードをSAVEファイルに格納する
	END	SEARCHコマンドを終了する
DISPLAY*	SIGNAL	時間波形を表示する
	SPECENV	スペクトル包絡を表示する
	FORMANT	ホルマント遷移を表示する
	END	DISPLAY コマンドを終了する
MOVE*		検索された区間、切り出された区間の生音声データを指定されたファイルへ転送する
ANALYSIS	SIGNAL	時間波形を表示する
	SPECENV	スペクトル包絡を表示する
	AREA	声道断面積を表示する
	FORMANT	ホルマント遷移を表示する
	ALPHA	α パラメータを求める
	KAPPA	κ パラメータを求める
	FFT	FFT によるスペクトル分析を行う
	CEPS	ケプストラム分析を行う
	RMS	エネルギーを求める
	ZERO	零交差分析を行う
	PITCH	ピッチを求める
	END	ANALYSIS コマンドを終了する
HELP*	SEARCH	SEARCH コマンドを説明する
	DISPLAY	DISPLAY コマンドを説明する
	MOVE	MOVE コマンドを説明する
	ANALYSIS	ANALYSIS コマンドを説明する
	HELP	HELP コマンドを説明する
	ITEM	項目名とその意味を説明する
	END	HELP コマンドを終了する
LOAD-RD LOAD-PH LOAD-SY :		データベースへのデータのロードを支援する

* MOVEはSEARCHの、DISPLAYはSEARCHとMOVEの、HELPはすべてのコマンドのサブコマンドとして使える。

即ち、

a) SY = SA OR SY = KA

b) SY = SA OR KA

a) , b) は同じ意味である。

更に、音声研究における専門用語を項目値として使うことも許されている。例えば有声破裂音 (Voiced Plosive) を使うと、

c) SY = VOICED-PLOSIVE+A

d) SY = BA OR DA OR GA

c) , d) は同じ意味となる。(日本語では有声破裂音はバ行, ダ行, ガ行だけである。) ここで “ VOICED-PLOSIVE ” の代わりにその省略形 “ VP ” を使うこともできる。また “ +A ” は後続母音を /A/ と指定することを意味する。後続母音を問題にしない場合には、 “ + \$ ” と入力すればよい。SEARCHコマンドで使用できる専門用語の一覧表を表 4.2 に示す。かっこの中は省略名である。

音韻記号を用いた項目値 (/BA/, /U/等) に関して、音韻の一般名 (PLOSIVE, FRICATIVE等) を用いて、その後続母音を指定する場合には、

SY = PL + A

VCV = A + FR + \$

のように “ + ” 記号を用いる。しかしながら音韻に直接対応する記号を用いる場合には、

SY = BA, SY = B\$, VCV = AG\$

のように書く。

更に、SEARCHコマンドでは検索条件に話者数 (NSUB)、レコード数 (NUM) を指定することができる。

COND SY = SA ; NSUB = 3 AND NUM = 5

上記の例は「音節 /SA/ を話者 3 人につき 5 レコードずつ欲しい」を示している。

また、SEARCHコマンドでは多重検索も行える。“ [“ , ”] ” で囲まれた区間の条件が一回の検索の対象となる条件である。“ [” の前が項目名の場合には、この検索で見つかったレコードの項目名に対応する項目値が次の検索の条件に加えられる。項目名以外の場合には見つかったレコードの ID コードが次の検索の条件に入れられる。

4.2 その他の IQL コマンド

DISPLAY コマンドは簡単なグラフィック表示を行うコマンドで、SEARCH コマンドのサブコマンドとしても使える。MOVE コマンドは音声の生データを利用者のパーマネン

トファイルに転送するコマンドである。ANALYSIS コマンドは音声の生データを分析し、結果をグラフィック表示したり利用者に転送したりするコマンドである。HELPコマンドはシステムに不慣れな利用者に、コマンドの使い方を説明する為に設けられている。

表 4.2 専門用語

専門用語 (日本語)	専門用語 (英語)	対応する音韻
母音	VOWEL (VW)	A, I, U, E, O
半母音	SEMIVOWEL (SV)	J, W
破裂音	PLOSIVE (PL)	B, D, G, P, T, K
有声破裂音	VOICED-PLOSIVE (VP)	B, D, G
無声破裂音	VOICELESS-PLOSIVE (VLP)	P, T, K
摩擦音	FRICATIVE (FR)	Z, C, S
有声摩擦音	VOICED-FRICATIVE (VF)	Z
無声摩擦音	VOICELESS-FRICATIVE (VLF)	C, S
鼻音	NASAL (NS)	M, N

5. 使用例 *

5.1 起動及び終了方法

```
① SYSTEM ?SPEECH-DB
_____
WELCOME TO SPEECH-DB
COMMAND ?
_____
```

① SYSTEM レベル或いはビルドモードでSPEECH-DBとキーインする。

```
② =SEARCH ↓
TYPE IN SUBCOMMAND
=COND SV = BS; SS = MALE ↓
_____
5 RECORDS FOUND.
```

② このレベルでIQLコマンドが使える。

⋮

TYPE IN SUBCOMMAND	
=END ↵	
COMMAND ?	
= ↵	③ キャリッジリターンのみで全処理が終了する。
GOOD-BYE	
SYSTEM ?	

* 例において下線部は利用者の入力を示す。

5.2 IQLの使用例

5.2.1 SEARCHコマンド

検索コマンドの例を示す。

Q 1 : 破裂音を含む連続音声が良い。

COND CONTINUOUS-SPEECH ; PHONEME=PLOSIVE 或いは
COND CS ; PH = PL

Q 2 : 男性が発声した単語で、母音 / A / と / I / を含むものが欲しい。

COND WORD ; PHONEME=A AND PHONEME = I AND SUBJECT-SEX
= MALE
COND WD ; PH = A AND PH = I AND SS = MALE

Q 3 : 26才以上の関西方言の男性の発声で、平均ピッチが7.5 ms 以上であり、かつ有声破裂音を含むVCV音節が良い。

COND VCV = \$ + VOICED-PLOSIVE + \$; SS = MALE AND SUB -
JECT-AGE GE 26 AND DIALECT = KANSAI AND VAV-PITCH
GE 7.5 或いは
COND VCV = \$ + VP + \$; SS = MALE AND SAG GE 26 AND VAP GE 7.5
或いは
COND ↵
= VCV = \$ + VP + \$; SS = MALE AND SAG GE 26 ↵
= AND DL = KANSAI AND VAP GE 7.5 ↵
= ↵

Q 4 : 鼻音と母音 / I / からなるCV音節が良い。

COND SY = NASAL + I 或いは
COND SY = NS + I

Q 5 : 男性の発声で、母音 / A / の定常部が欲しい。

COND PH = A ; RK1 = A AND SS = MALE

つぎに、他のサブコマンドの使用例を示す。

```

①
② TYPE IN SUBCOMMAND
③ =COND
   =WD ; ( PH = A AND RK1 = A )
   =AND ( PH = N AND RK1 = A )
   =AND WAP < 7.0
   =↓
                                     5 RECORDS FOUND.
TYPE IN SUBCOMMAND
④ =LIST IW WD WDF WDT ←'84年4月現在 動かない
⑤ HOW MANY RECORDS TO LIST?
   =↓
-----
ID-WORD          1
WORD              BAKUON
WDATA-FROM       51
WDATA-TO         6700
-----
ID-WORD          16
WORD              ASAGOHAN
WDATA-FROM       74501
WDATA-TO         81400
-----
ID-WORD          42
WORD              BAKUON
WDATA-FROM       185451
WDATA-TO         192200
-----
ID-WORD          44
WORD              GINSEKAI
WDATA-FROM       196051
WDATA-TO         202700
-----
ID-WORD          58
WORD              PAN
WDATA-FROM       263801
WDATA-TO         266550
-----
TYPE IN SUBCOMMAND
⑥ =SAVE
   SAVED FILE ID NUMBER =          1

```

(WD) , 単語の開始点 (WDF) , 終点 (WDT) を表示させた。

⑤出力レコード数を全検索レコード (キャリジリターンにより) と指定した。

⑥後で使うのでSAVEサブコマンドでセーブファイルにセーブした。

①SEARCHコマンドは既に起動されているものと仮定する。

②サブコマンドの入力待ち状態である。

③CONDサブコマンドを入力して検索を行っている。条件は「 / A / という音韻 (PH) でのランク (RK1) が定常 (A) のものを含み、かつ / N / でランクが定常のものを含む単語 (WD) で、その平均ピッチ (WAP) が 7.0 m sec より小さいもの」である。

④③で検索したレコードについてLISTサブコマンドで単語のIDコード (IW) , 単語名

TYPE IN SUBCOMMAND
 ⑦ =COND SY = PLOSIVE ; SS = MALE AND SAG GE 20
 ⑧ FOLLOWING VOWEL ?
 =A

16 RECORDS FOUND.

TYPE IN SUBCOMMAND
 ⑨ =LIST SY /,

SYLLABLE	BA
SYLLABLE	GA
SYLLABLE	KA
SYLLABLE	GA
SYLLABLE	PA
SYLLABLE	TA
SYLLABLE	TA
SYLLABLE	TA
SYLLABLE	BA
SYLLABLE	GA
SYLLABLE	KA
SYLLABLE	GA
SYLLABLE	PA
SYLLABLE	TA
SYLLABLE	TA
SYLLABLE	TA

TYPE IN SUBCOMMAND
 ⑩ =COND WD ; SY = UP + A AND DL = DL IB = 1 J
 6 RECORDS FOUND.

⑦再び検索を行っている。

条件は「話者の性別 (SS) が男 (MALE) で話者の年齢 (SAG) が20才以上の、破裂音 (PLOSIVE) の音節 (SY)」である。

⑧⑦で後続母音の指定がなかった為に I Q L が聞き返している。

⑨“ /, ”と指定すると全検索レコードを出力する。

⑩今度は2重検索を行っている。条件は「後続母音が / A / の有声破裂音 (VP) を含み、話者コード (IB) が1の話者と同じ方言 (DL) で発声された単語」である。

⑪MOVEサブコマンドでセーブファイル1の音声の生データを利用者のファイルへ転送する。

⑫転送に先だってDISPLAYコマンドで波形のチェックを行うかどうか聞いている。

⑬DISPLAYコマンドのサブコマンドSIGNALで波形を表示させる。

⑭波形表示の為の条件を設定する。

⑮カーソルにより1, 2の点のポイント数を表示させた。

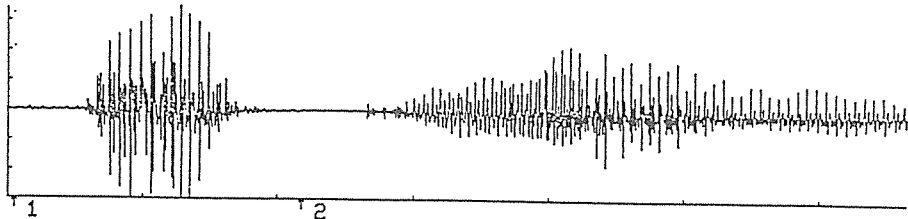
⑯SIGNALサブコマンドを終了する。

⑰DISPLAYコマンドを終了する。

```

TYPE IN SUBCOMMAND
⑪ =MOVE /1
⑫ RAW DATA SIGNAL CHECK (YE OR NO)?
  =YE
TYPE IN SUBCOMMAND
⑬ =SIGNAL
  DISPLAY FORMAT DECISION!!!
  HARD COPY ?
  =NO
  MURGIN MODE? [C]=30; (NORMAL=          50) ]
  = /
  AXIS MODE? [ <0; UNDER AXIS , >0; CENTER AXIS :ABS=SCALE ]
  NORMAL AXIS MODE=          -1000
  = /
  DETAIL OR GROBAL(NORMAL=GROB)?
  = /
⑭ START POINT, END POINT ?
  =51,6700
  SIGNAL MAX, MIN?
  =0.0
  HOW MANY FLOOR ? (NORMAL=          1)
  = /
  DISPLAY PREPARANCE OK!
  IF YOU SCALE BY CURSOR, INPUT ( C ) AFTER BELL!!!
  IF YOU SEE THE MEAN, TYPE ANY KEY!
  TINPUT SUB END !!

```



```

⑮ { ***POSI 1= 107 CURSOR, DISPLAY, NEXT?(FIRST 2CHR)
    *##POSI 2= 2235 CURSOR, DISPLAY, NEXT?(FIRST 2CHR)
    =NEXT
    *****
    * DISP NO. I          1 *
    * DISP START POINT    51 *
    * DISP END POINT      6700 *
    * MAX.MIN U.I        1736.-1488.*
    *****
    NEXT, FORM, STOP, HELP
⑯ =STOP
    ***SIGNAL DISPLAY END***
    TYPE IN SUBCOMMAND
⑰ =END
    DISPLAY COMMAND END!!

```

```

⑱ MOVE ALL DATA(YE OR NO)?
  =NO
⑲ TYPE IN ID-WORD
  =1
  =↓
⑳ CHANGE FROM,TO(YE OR NO)?
  =YE
㉑ TYPE IN ID-WORD TO BE CHANGED
  =1
㉒ TYPE IN FROM,TO(ID-WORD=
  =107,2235
  TYPE IN ID-WORD TO BE CHANGED
  =↓
㉓ OUTPUT USER FILE NAME
  ASSIGNED USR ID/CATALOG/----/FILE ?
  =6088050333/DEMO
  ASSIGNED FILE =6088050333/DEMO
  FILE OK (YE OR NO) ?
  =YE
  **FILE FORMAT
  1) PACKED RANDOM FILE
  2) NORMAL RANDOM FILE
㉔ SELECT FILE FORMAT (1 OR 2)?
  =1
  TYPE IN SUBCOMMAND
㉕ =END

```

⑱すべての音声の生データを転送するかどうか聞いている。

⑲転送するレコードの ID-WORD を聞いている。

⑳転送するデータの開始点、終点を変更するかどうか聞いている。

1) ㉑変更するレコードの ID-WORD を聞いている。

㉒変更した点を入力する。

㉓利用者の転送ファイルのカタログファイル名を入力する。

㉔転送ファイルの形式を指定する。I Q Lでは3サンプル点ずつパックして転送する形式と1サンプル点のみをパックせずに転送する形式を用意している。

㉕ ENDサブコマンドでSEARCHコマンドを終了した。

5.2.2 DISPLAYコマンド

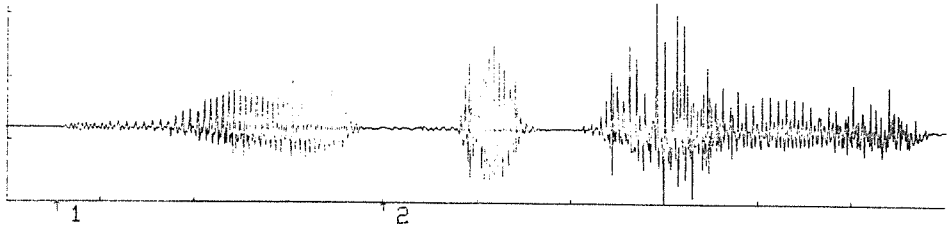
DISPLAYコマンドは表 4.1 に示す様に3個のサブコマンドから成り、グラフィック表示機能をもつ。

表 5.1 表示に関する補足事項

サブコマンド名	補 足 事 項
S I G N A L	音声の生データの時間波形を表示する。カーソルにより、任意の位値のポイント数を知ることができる。
S P E C E N V	スペクトル包絡を2次元あるいは3次元で表示することができる。表示のための分割に関してはデフォルト値が設定されているが、利用者が任意に設定することもできる。
F O R M A N T	FORMANT周波数をバンド幅に反比例した大きさとプロットする。縦軸のFORMANT周波数は0～5000Hzである。

以下に使用例を示す。

```
COMMAND ?
① =DISPLAY
TYPE IN SUBCOMMAND
② =SIGNAL
DISPLAY FORMAT DECISION!!!
START POINT,END POINT ?
③ =594000,604000
HOW MANY FLOORS ? (NORMAL=          1)
④ = 2
IF YOU WANT TO KNOW THE SEQUENCE NUMBER OF A POSITION,
INPUT ( C ) AFTER ADJUSTING CURSOL!
```



```
⑤ {***POSI 1=594550 CURSOL,DISPLAY,NEXT?(FIRST 2CHR)
*#POSI 2=598004 CURSOL,DISPLAY,NEXT?(FIRST 2CHR)
=DI
*****
* DISP NO. I          1 *
* DISP START POINT594000 *
* DISP END POINT 604000 *
* MAX.MIN U.I 1006. -593.*
*****
NEXT,FORM,STOP,HELP
⑥ =STOP
***SIGNAL DISPLAY END***
TYPE IN SUBCOMMAND
⑦ =END
DISPLAY COMMAND END!!

COMMAND ?
=
```

- ① DISPLAYコマンドを指定する。
- ② SIGNALサブコマンドを指定する。
- ③ 表示開始点、終了点を入力する。
- ④ 波形をいくつに分けて表示するか聞いている。
- ⑤ カーソルにより1, 2の点のポイント数を表示する。
- ⑥ SIGNALサブコマンドを終了する。
- ⑦ DISPLAYコマンドを終了する。

5.2.3 ANALYSIS コマンド

ANALYSIS コマンドは表 4.1 に示した様に 12 のサブコマンドから成る。各サブコマンドにより分析結果をグラフィック表示することができ、利用者の用意したファイルにデータとして格納することもできる。

○グラフィック表示方法について

表 5.2 グラフィック表示に関する補足事項

サブコマンド名	補 足 事 項
SIGNAL	DISPLAY コマンドの SIGNAL サブコマンドと同様
SPECENV	“ “ SPECENV “ “
AREA	表示法に 2 次元表示と 3 次元表示が用意されている。管面の分割に関しては、デフォルト値が設定されているが、利用者が任意に設定することもできる。
FORMANT	DISPLAY コマンドの FORMANT サブコマンドと同様
ALPHA	グラフィック表示機能を持たず
KAPPA	“
FFT	AREA サブコマンドと同様
CEPS	“
RMS	特になし
ZERO	零交差の閾値は利用者が任意に設定できる。
PITCH	表示グラフの縦軸は 50～350 Hz に設定している。

(注) ANALYSIS の各サブコマンドについての補足

各サブコマンドとも分析窓の長さ、窓の移動距離を任意に設定できるが PITCH サブコマンドは、SIFT アルゴリズムを用いているため分析窓長を 400 points (40 msec)、移動距離を 100 points (10 msec) に固定している。

○データ格納方法について

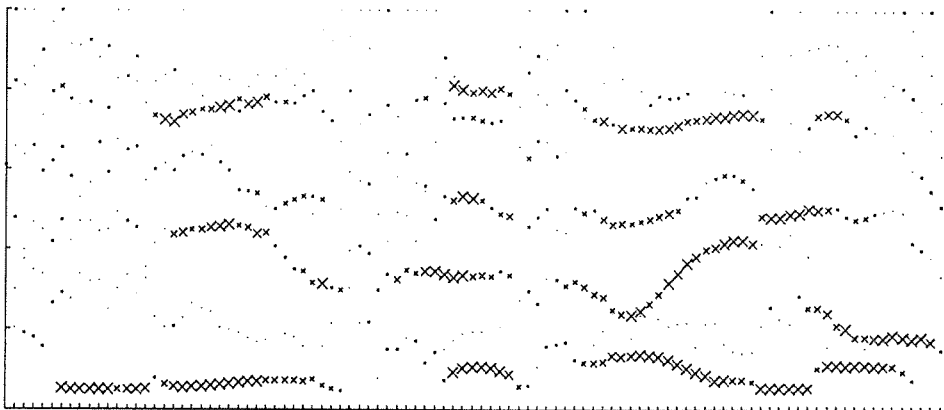
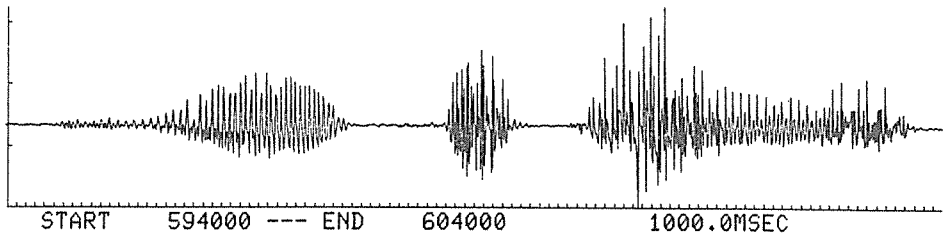
利用者の用意したファイルを直接編成ファイルとし、表 5.3 の様にデータを格納する。

表 5.3 データ格納方法

サブコマンド名	REC. No.	格 納 内 容
SPECENV	1 2 ~	開始点, 終了点, 窓長, 移動距離, 分析次数, 分析区 間数 (SEQ. No., スペクトラム (125))
AREA	1 2 ~	SPECENVと同様 (SEQ. No., 声道断面積 (分析次数 + 1))
FORMANT	1 2 ~	SPECENVと同様 (SEQ. No., FORMANT 周波数 (分析次数 / 2))
ALPHA	1 2 ~	SPECENVと同様 (SEQ. No., α パラメータ (分析次数 + 1))
KAPPA	1 2 ~	SPECENVと同様 (SEQ. No., κ パラメータ (分析次数 + 1))
FFT	1 2 ~	開始点, 終了点, 窓長, 移動距離, FFTの分析長, 分析 区間数 (SEQ. No., FFT (分析長))
CEPS	1 2 ~	開始点, 終了点, 窓長, 移動距離, CEPSの分析長, 分 析区間数 (SEQ. No., ケプストラム (分析長))
RMS	1 2 ~	開始点, 終了点, 窓長, 移動距離, 分析区間数 (SEQ. No., エネルギー値)
ZERO	1 2 ~	開始点, 終了点, 窓長, 移動距離, 閾値, 分析区間数 (SEQ. No., 零交差数)
PITCH	1 2 ~	開始点, 終了点, 分析区間数 (SEQ. No., ピッチ周波数)

以下に使用例を示す

- ① COMMAND ?
=ANALYSIS
- ② TYPE IN SUBCOMMAND
=FORMANT
- ③ FORMANT SUBCOMMAND START!!!
RESULT EXPRESSION (GRAPH OR DATA)?
=GRAPH
- ④ ANALYSIS METHOD (AUTO OR COVA)? (DEFAULT=AUTO)
=✓
- ⑤ ANALYSIS ORDER? (DEFAULT= 12)
=✓
- ⑥ WINDOW SELECTION (RECT,HANN,HAMM,BLUM)? (DEFAULT=HAMM)
=✓
- ⑦ PREPROCESSING OF SIGNAL (6-DB,ADAP,NOPR)? (DEFAULT=6-DB)
=✓
- ⑧ PREEMPHASIS COFF? (DEFAULT= 0.1000000E+01)
=✓
- ⑨ START POINT,END POINT?
⑨ =594000,604000
WINDOW LENGTH,SHIFT LENGTH (POINTS)?
⑩ =200,100
- ⑪ DO YOU WANT SIGNAL?
=YES
- ⑫ DETAIL DRAWING?
=NO



- ⑬ DISP OVER NEW PAGE?
=YES
- ⑭ OTHER FORMANT DISP?
=NO
- ⑮ FORM CHANGE?
=NO
*** FORMANT SUBCOMMAND END ***
TYPE IN SUBCOMMAND
=

- ① ANALYSIS コマンドを指定する。
- ② ANALYSIS のサブコマンド FORMANT を入力する。
- ③ グラフィック表示かデータ格納か聞いている。
- ④ 分析方法が自己相関か共分散か聞いている。
- ⑤ 分析次数を聞いている。
- ⑥ 分析窓の種別を聞いている。
- ⑦ 高域強調の方法を聞いている。
- ⑧ 高域強調が一次差分の場合、差分係数を聞いている。
- ⑨ 分析開始点と終了点を入力する。
- ⑩ 窓の長さ、移動距離を入力する
- ⑪ シグナル表示が必要か聞いている。
- ⑫ シグナル波形を詳細に書くか聞いている。
- ⑬ 表示面をクリアするか聞いている。
- ⑭ 他の位置を分析、表示するか聞いている。
- ⑮ 分析の条件を変更して表示するか聞いている。

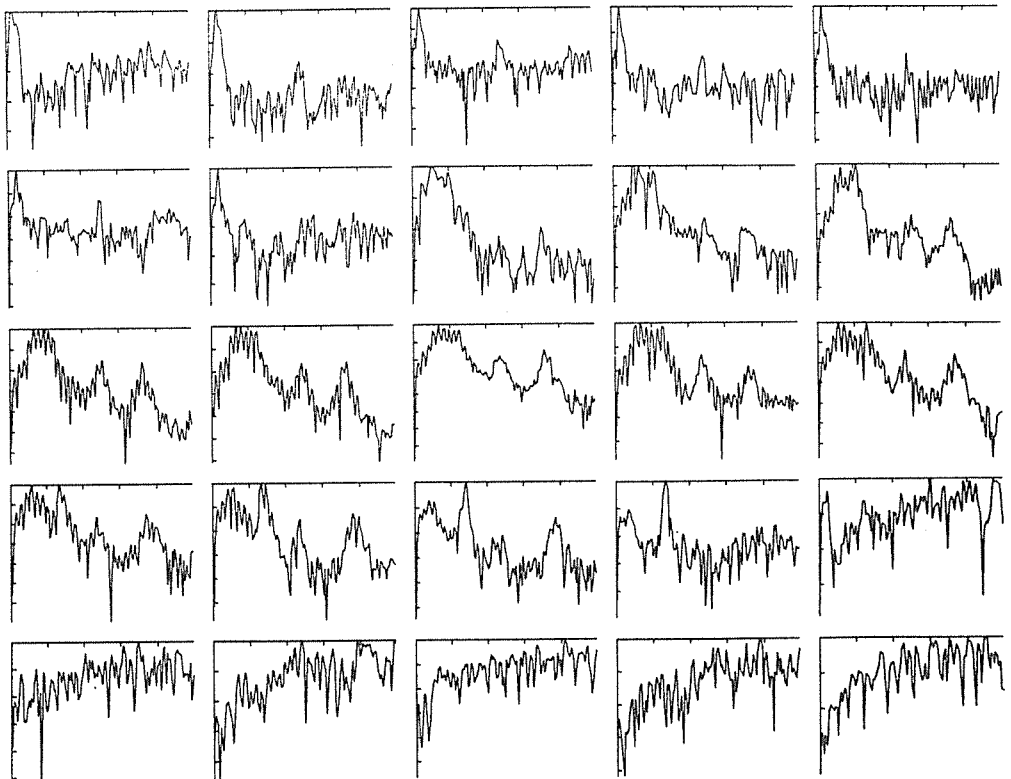
```

TYPE IN SUBCOMMAND
⑯ =ALPHA
ALPHA SUBCOMMAND START!!!
⑰ ANALYSIS METHOD (AUTO OR COVA)? (DEFAULT=AUTO)
=✓
⑱ ANALYSIS ORDER? (DEFAULT= 12)
=✓
⑲ WINDOW SELECTION (RECT,HANN,HAMM,BLUM)? (DEFAULT=HAMM)
=✓
⑳ PREPROCESSING OF SIGNAL (6-DB,ADAP,NOPR)? (DEFAULT=6-DB)
=✓
㉑ PREEMPHASIS COFF? (DEFAULT= 0.10000000E+01)
=✓
START POINT,END POINT?
㉒ =584200,586000
WINDOW LENGTH,SHIFT LENGTH (POINTS)?
㉓ =200,100
OUTPUT USER FILE NAME
ASSIGNED USER ID/CATALOG/----/FILE
㉔ =6000820408/DATA
FILE OK (YE OR NO) ?
㉕ =YE
*** ALPHA SUBCOMMAND END ***
TYPE IN SUBCOMMAND
=

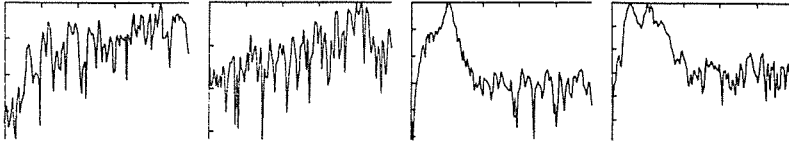
```

- ⑯ ANALYSIS のサブコマンド ALPHA を指定する。
- ⑰～㉓ 分析の為の条件を設定する。
- ㉔ 書き込むファイル名を入力する。
- ㉕ ファイル名を確認する。

- ⑥ TYPE IN SUBCOMMAND
=FFT
FFT SUBCOMMAND START!!!
- ⑦ RESULT EXPRESSION (GRAPH OR DATA)?
=GRAPH
- ⑧ DISPLAY METHOD (2-D OR 3-D)?
=2-D
- ⑨ WINDOW SELECTION (RECT,HANN,HAMM,BLUM)? (DEFAULT=HAMM)
=
- ⑩ PREPROCESSING OF SIGNAL (6-DB,ADAP,NOPR)? (DEFAULT=NOPR)
=
START POINT,END POINT?
⑪ =584200,587000
WINDOW LENGTH,SEIFT LENGTH?
⑫ =200,100
FFT POINT NO.?
⑬ =256
⑭ LOG SCALE?
=YES
2-DIMENSINAL DISPLAY!
- ⑮ SCREEN DIVISION?(IX,IY,MURG)
=



FULL!



```

③ OTHER FFT DISP?
  =NO
  2-DIMENSIONAL DISPLAY END!
④ FORM CHANGE?
  =NO
  *** FFT SUBCOMMAND END ***
  TYPE IN SUBCOMMAND
  =

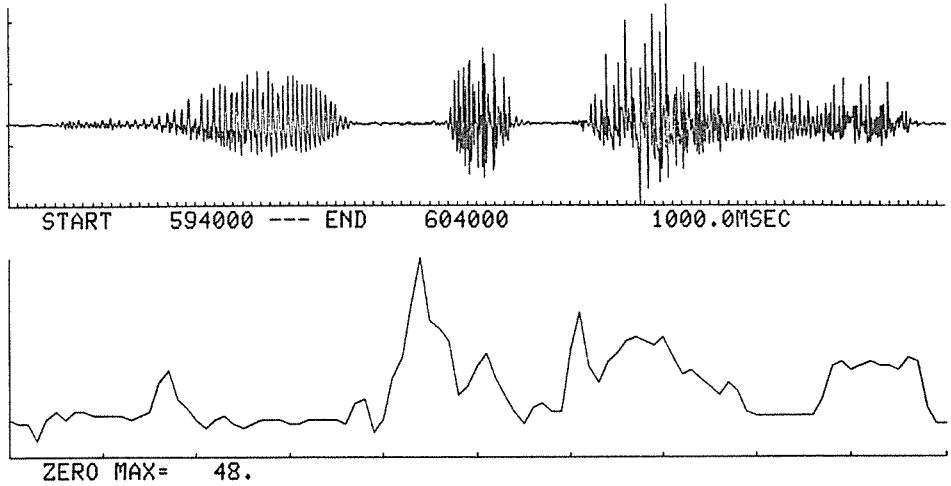
```

- ③ ANALYSISのサブコマンドFFTを指定する。
- ④ グラフィック表示かデータ格納か聞いている。
- ⑤ 表示方法（2次元か3次元）を聞いている。
- ⑥～⑧ 分析の為の条件を設定する。
- ⑨ FFTの分析長（2の巾乗）を入力する。
- ⑩ logスケールで表示するか聞いている。
- ⑪ 表示の分割方法を聞いている。（デフォルト値が設定されている。）
- ⑫ 他の位置の分析、表示をするか聞いている。
- ⑬ 分析方法を変更するか聞いている。

```

TYPE IN SUBCOMMAND
⑭ =ZEROCROSS
ZEROCROSS SUBCOMMAND START!!!
⑮ RESULT EXPRESSION (GRAPH OR DATA)?
  =GRAPH
⑯ THRESHOLD? (DEFAULT= 0.30000000E+01)
  = 1
START POINT,END POINT?
⑰ =594000,604000
WINDOW LENGTH,SHIFT LENGTH (POINTS)?
⑱ =200,100
⑲ DO YOU WANT SIGNAL?
  =YES
⑳ DETAIL DRAWING?
  =NO

```



```

⑤ DISPLAY OTHER ZEROCROSS?
  =NO
⑥ CHANGE THRESHOLD?
  =NO
  *** ZEROCROSS SUBCOMMAND END ***
  TYPE IN SUBCOMMAND
⑦ =END
  ANALYSIS COMMAND END!!

  COMMAND ?
  =

```

- ③ ANALYSISのサブコマンド ZEROを指定する。
- ③ グラフィック表示かデータ格納かを聞いている。
- ④ 零交差の閾値を聞いている。
- ④～⑥ 分析の為の条件を設定する。
- ③～④ シグナル表示について聞いている。
- ⑤ 他の位置の零交差を分析、表示するか聞いている。
- ⑥ 閾値を変更するか聞いている。
- ⑦ ANALYSISコマンドを終了する。

6. む す び

音声データベース SPEECH-DBの概略とその利用方法について述べた。本データベースは音声認識研究を支援する為の音声データベース構築の第一歩に過ぎず、その完成の為には、今後さらに検討を加える必要がある。

先ず第一に、音声データの選択及びその格納の問題がある。音声の「標準データ」は今のところ存在しないので独自に行わざるを得ない。この点に関しては、まえがきにおいて述べたように利用者のデータ格納面における参加を期待している。

2番目の問題としてSPEECH-DBと利用者とのマン・マシンインタフェースの改善がある。既に開発されたIQLコマンドは、利用者にデータベースの論理構造を意識させないという点では優れているが、やはり利用者が事前に持つべき知識及び構文上の制約がある。現在、全くの初心者が容易に使えるように日本語による検索要求を受け付けるシステム及びIQLの使用中に発生するエラーを診断するシステムを作成中であり、これ等2つのシステムが完成すれば利用者とのマン・マシンインタフェースはかなり改善されるものと思われる。

今後、他の音声研究者の利用及びデータの格納の両面における協力を得て、SPEECH-DBの拡張を進めて行きたいと考えている。

<参 考 文 献>

- (1) 溝口, 前田, 浜口, 芥子, 柳田, 角所: “知的アクセス機能を持つ音声データベース「SPEECH-DB」,” 情報処理学会論文誌, Vol. 24, No. 3, PP. 271-280.
- (2) 溝口, 角所: “知的アクセス機能を持つ音声データベース「SPEECH-DB」—研究開発計画の成果—,” 大阪大学大型計算機センター・ニュース, Vol. 12, No. 2 PP. 71-81 (1982)。