

知的アクセス機能をもつ音声データベース 「SPEECH-DB」†

溝口 理一郎†† 前田 直孝†††* 浜口 理彦†††**
芥子 育雄††† 柳田 益造†† 角 所 収††

近年、音声認識に関する研究はハードウェア技術の発達とともに急速に進歩し、単語単位の認識に基づく特定話者用認識装置が商品化されるに至っている。さらに現在では音韻単位の認識に基づく不特定話者・大語彙認識装置の開発へと研究が進められている。音韻単位の認識に基づく音声認識を行う際には、話者の相違および調音結合の影響による音韻の変動の解明等基礎的研究を行うために大量の音声データの分析が必要となり、その蓄積・管理が重要な問題となると予想される。本論文で述べる音声データベース SPEECH-DB は新しい音声データ管理システムであり、今後の音声認識の基礎研究を支援することを目的として開発された。SPEECH-DB では任意の音韻環境の下での音韻あるいは音節の検索はもちろん、その他種々の音声データの容易な検索が可能である。また、本データベースは知的アクセス機能をもっていることも特徴の一つとしている。通常のデータベースでは利用者はデータベースの構造に関する知識をもっていなければならないが、利用に際しての障害となっているが、SPEECH-DB では知的コマンド IQL を用いることによりたんに検索条件を羅列するだけで検索が可能となっている。さらに、すべてのコマンドの音声入力が可能であり、また、システムからの音声による応答があるなどマン・マシンインタフェースにおける多くの工夫がなされている。

1. ま え が き

単語単位の認識に基づく特定話者小語彙音声認識装置が実用化された現在、大語彙を扱える音韻単位の不特定話者音声認識装置の開発を目差した研究が盛んに行われている。連続音声の中の音韻は話者の違いによる影響はもちろんのこと、前後に発声された音韻の影響(調音結合)により大きく左右され、それらのばらつき程度および質を把握することは音韻認識を行う上で不可欠なこととなる。このためには大量の音声データの収集とその分析が必要となるが、検索機能をもたない従来のデータ管理方式では不十分であると考えられる。このことから今後の音声認識研究を支援するための新しい音声データ管理システムの必要性が理解されるが、管理システムがもつべき機能として次の4点が考えられる。

- 1) 大量データの蓄積およびその管理
- 2) 蓄積データの容易な検索と利用
- 3) 検索データの簡単な分析とグラフィック表示
- 4) 全国の音声研究者の共同利用

そこでわれわれは既存のデータベース管理システムを利用することにより、上記の四つの機能を持ち、かつ任意の音韻環境の下での音韻の検索およびその他音声認識研究を遂行する際に必要な種々の音声データの検索を容易にする音声データベース SPEECH-DB¹⁾⁻⁶⁾を開発した。

一方、現在多くのデータベースが開発されているが、一般利用者はデータベースの論理構造に関する予備知識を要求されるばかりでなく、不慣れたキーボードにおいてデータベース固有の複雑なコマンドを用いることを余儀なくされている。SPEECH-DB では利用者とのマン・マシンインタフェースの円滑化を目差して、知的アクセス機能をもつ検索言語 IQL (Intelligent Query Language) が作成されている⁴⁾。IQL を用いることにより、利用者はデータベースの論理構造を意識することなしに検索条件を羅列するだけで検索が行える。また、IQL のコマンドおよびサブコマンドすべてを音声により入力することができ、さらには、システムからの入力促進等の音声応答機能もあり初心者にも容易に利用できるように配慮されている。利用

† Speech Data Base with an Intelligent Access Mechanism—SPEECH-DB by RIICHIRO MIZOGUCHI (Institute of Scientific and Industrial Research, Osaka University), NAOTAKA MAEDA, MICHIOKI HAMAGUCHI, IKUO KESHI (Faculty of Engineering, Osaka University), MASUZO YANAGIDA and OSAMU KAKUSHO (Institute of Scientific and Industrial Research, Osaka University).

†† 大阪大学産業科学研究所

††† 大阪大学工学部電子工学科

* 現在 日本電気株式会社

** 現在 神戸製鋼株式会社

者は TSS 用コマンドとしての IQL を使う以外に、FORTRAN で書かれた応用プログラムから通常のファイルへのアクセスとほぼ同様にして音声データを利用することもできる。このようにさまざまな機能をもつ SPEECH-DB は、たんに音声という数値データを格納・管理するだけでなく、検索、分析およびグラフィック表示等のソフトウェアが有機的に結合された総合的なデータベースシステムとなっている。

次章において SPEECH-DB の概略について述べる。3 章では IQL について、また 4 章では実働化について述べる。また、付録において IQL の使用例を示す。

2. 音声データベース

2.1 音声認識研究における音声データの利用

まえがきにおいても述べたように、音声認識研究を遂行するにあたって大量の音声データの解析が必要となると予想されるが、以下に音声データベースの利用法を概観する。

連続音声の中の音韻は、調音結合の影響を受けて、単独に発声される場合の特性から大きく変化するため、その変動を把握することは連続音声認識研究の最も重要な課題となっている。このような音声認識研究において、前後の音韻環境を条件とした検索機能をもつ SPEECH-DB では、たとえば 5 母音間の遷移の状態あるいは子音 /b/ の後続母音の相違による特性の差等を調べるためのデータの提供が可能となる。さらに、有声音・無声音・無音の判定アルゴリズム、セグメンテーション（単位音声区間の自動ラベル付け）等の研究にも各種の音韻の存在を条件とした音声データを提供する。また、すべての音声データは開発された各認識システムの性能評価のための標準的なデータ*として用いることもできる。

2.2 INQ (DBMS) の概要

SPEECH-DB は DBMS として日本電気の INQ¹⁰⁾ を用いて実現されている。INQ の論理構造の記述法は次の二つのステップより成る。

1) 概念スキーマの記述

INQ では、データベースは INQ ファイルと呼ばれる複数の独立したファイルの集合として構成される。各 INQ ファイルの論理構造は階層型であり FDL (File Description Language) と呼ばれる言語で記述され、これらの記述の集合としてデータベースの概念

スキーマが構成される。

2) 外部スキーマ記述

外部スキーマは INQ では INQ セクションと呼ばれる。INQ セクションは四つ以下のファイルの結合関係を記述し、さまざまな論理構造を仮想的に構成する。結合関係の論理構造はやはり階層型である。通常利用者は、自分で INQ セクションを作成するか、すでに作成されたもののなかから自分の検索要求に適したものを選択する。

2.3 SPEECH-DB の論理構造

データベースの論理構造の設計にあたって、実際に現れると予想される数 10 個の検索要求の検討を行い、格納項目*の決定、コマンドの設計等に役立てた。以下に検討された検索要求例を示す。

- Q 1 : 男性の発声で破裂音を含む連続音声が欲しい。
 Q 2 : 20 歳以上の人の発声で母音 /a/ と /i/ を含む単語が欲しい。
 Q 3 : 関西方言の男性の発声で平均ピッチが 150Hz 以上でありかつ有声破裂音を含む VCV 音節が欲しい。
 Q 4 : 女性の発声で鼻音を含み後続母音が /e/ の CV 音節が欲しい。
 Q 5 : 東京方言の 20 歳以上 35 歳以下の男性の発声で母音 /a/ から /i/ への遷移部の音韻が欲しい。
- 音声は図 1 に示すような階層構造**をなしていると一般に考えられている。ここで C, V はおのおの子音, 母音を表す。この音声の階層性を効率よく実現

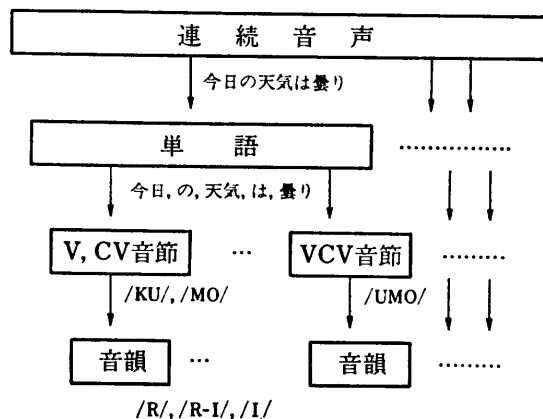


図 1 音声の階層構造

Fig. 1 Hierarchical structure of speech.

* 音声データの標準化は、手書き文字のデータベースと同様音声認識装置の評価のためには必要なことであるが、困難な点が多く残されており本論文では触れないことにする。

* INQ では関係データベースで言う属性は項目と呼ばれる。

** この階層構造が利用者に要求されるデータ構造に関する唯一の予備知識であり、事実上無視できるものと考えられる。

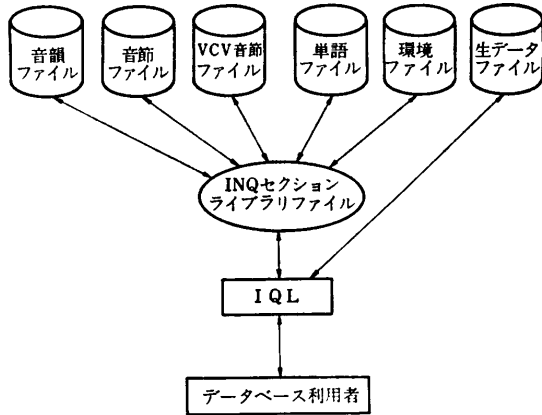


図 2 SPEECH-DB の基本構成

Fig. 2 Basic architecture of SPEECH-DB.

するために SPEECH-DB を図 2 に示すように、音韻、CV 音節、VCV 音節、単語、環境、生データの 6 個の独立したファイルで構成し、階層構造は INQ セクション記述を用いて仮想的に実現することとした。SPEECH-DB に格納されている項目の数は、音韻ファイルに 14、CV 音節ファイルに 15、VCV 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。項目の数が比較的多くなっているが、利用者の多様な検索要求に応じるために必要かつ十分なものと考えている。各 INQ ファイルには音声データを検索するために必要な情報が格納されている。表 1 に音韻ファイルの項目を、表 2 に環境ファイルの項目を示すが他のファイルも同様である。生データファイルは通常のランダムファイルであり、遮断周波数

表 1 音韻ファイルの格納項目

Table 1 Items contained in phoneme file.

| 項目名 | 省略名 | 意味 |
|-------------|------|---------------------------|
| ID-FRAME | IF | フレームの ID コード |
| ID-SYLLABLE | IS | このフレームが属する音節の ID コード |
| ID-VVAL | IV | このフレームが属する VCV 音節の ID コード |
| ID-WORD | IW | このフレームが属する単語の ID コード |
| ID | ID | このフレームが属する音声の ID コード |
| FNUM | FN | この音声内でのフレーム番号 |
| PHONEME | PH | 音韻表記 |
| RANK 1 | RK 1 | 定常, 非定常の評価 |
| RANK 2 | RK 2 | 未使用 |
| RMS | RM | エネルギー |
| ZERO-KOSA | ZK | 零交差 |
| PITCH | PT | ピッチ |
| DATA-FROM | DF | このフレームの開始点 |
| DATA-TO | DT | このフレームの終点 |

* 環境ファイル内の 6 項目はたんに表示のためだけであるので、検索の際の条件となりうる項目は 76 項目である。

表 2 環境ファイルの格納項目

Table 2 Items contained in environment file.

| 項目名 | 省略名 | 意味 |
|----------------|-----|------------------------|
| ID | ID | この音声の ID コード |
| ID-SUBJECT | IB | 話者コード |
| SUBJECT | SB | 話者名 |
| SUBJECT-SEX | SS | 話者の性別 |
| SUBJECT-AGE | SAG | 話者の年齢 |
| RECORD-DATE | RD | この音声を収録した年, 月, 日 |
| SPEECH-CATE | SC | この音声のカテゴリ (連続音声, 単音節等) |
| SPEECH-DESC | SD | 発声した内容 |
| SAMPLING-FRQ | SF | サンプリング周波数 |
| DIALECT | DL | 話者の方言 |
| FRAME-LENGTH | FL | 対応する音韻ファイルのフレーム長 |
| FRAME-SHIFT | FS | 対応する音韻ファイルのフレームシフト間隔 |
| FRAME-MAX | FM | 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値 |
| 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値 |
| 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値 |
| 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 |

4.5 kHz で低域濾波された後、10 kHz, 12 bit で AD 変換された音声データがメモリ節約のために 1 word (36 bit) に 3 サンプル点ずつ格納されている。

音声データはフレームと呼ばれる 20ms 長のデータ (200 サンプル) 単位に音韻等の属性が定められ、音韻ファイルの 1 レコードは 1 フレームの属性値の集合に対応する。したがって、音節ファイルの 1 レコードの下位レコードとして通常 20 個程度の音韻ファイルのレコードが対応している。

音声の階層構造はあくまで図 1 に示した 1 とおりのものであるが、検索する側から見る場合、たとえば検索の単位として連続音声をとれば (ある条件を満たす連続音声があれば)、一つの発声の一つのレコードになり、また、音節を検索の単位とすれば、一つの発声につき各音節を一つのレコードとして数 10 個のレコードが存在することになる。このことを実現するために数種類の INQ セクション記述が用意されている。INQ セクションについては 4.3 節において再び論じられる。

2.4 SPEECH-DB の利用環境

図 3 に SPEECH-DB の利用環境を示す。SPEECH-DB の計算機環境は、INQ を用いた数値データベースとしてすでに大阪大学大型計算機センターにおいて

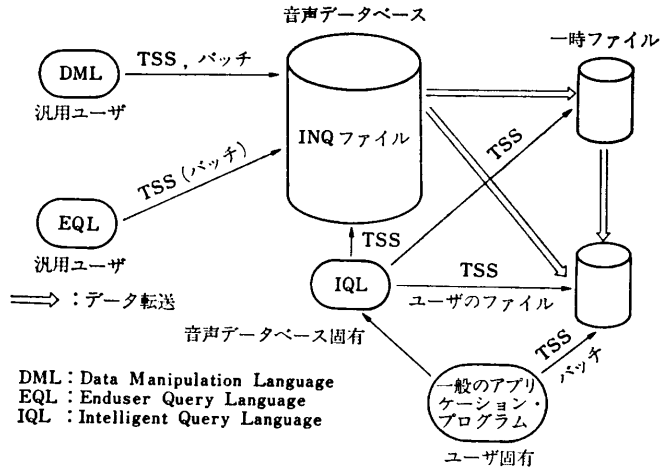


図 3 SPEECH-DB の利用環境

Fig. 3 Computer environment of SPEECH-DB.

サービスされている蛋白質データベース PROTEIN-DB¹¹⁾ と同一であるのでここではその詳細については省略する。INQ には汎用のデータベース操作言語 DML と TSS 利用者用言語 EQL が用意されているが、これだけでは音声研究支援のためのソフトウェアとしては不十分である。そこで、SPEECH-DB では音声データベース固有の会話型検索言語 IQL が開発されている。SPEECH-DB は大阪大学大型計算機センターの ACOS 1000 上で作成されているが、7 大学大型計算機センター間ネットワークを通して各地の TSS 端末からアクセスすることができる。なお、IQL では分析およびグラフィック表示コマンドも整備されており、実際に波形等を確認しながら利用者固有のファイルへ音声データを転送できる。

3. IQL

一般にデータベースの価値はデータの質の高さとその使いやすさの 2 点において決まるといわれている。データベースへの問合せ言語としては、SEQUEL⁷⁾, SQUARE⁸⁾, QBE⁹⁾ 等多くのものが提案されているが、そのいずれの言語においても汎用を目差しているため、利用者のデータベースの構造に関する予備知識が仮定されている。たとえば QBE の場合では、利用者が検索に必要なと考える関係を複数個指定し、どの関係同士をどの属性に関して結合するかを陽に指定しなければならない。そのためには、どの属性がどの関係に含まれているかを知っていなければならない。初心者にとっては大きな負担となっている。本章では、一般の音声研究者がデータベースの論理構造に関する知

識なしに容易に利用できるように開発された SPEECH-DB 専用の知的コマンド IQL について述べる。

3.1 SEARCH コマンド

表 3 に IQL コマンドおよびその機能を示す。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および終点を知るための操作とその位置情報を用いた生データの表示、あるいは転送のための操作との 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値の利用者への提示の 2 段階に分かれている。ここでは検索に用いられる最も重要な SEARCH コマンドについて述べる。

表 3 IQL コマンドとその機能

Table 3 IQL commands and their functions.

| コマンド名 | サブコマンド名 | 機能 |
|------------------------------------|---------------------------------|--|
| SEARCH | CONDITION AND, OR, NOT LIST KEY | 条件を満たすレコードを検索する SAVE ファイル間の論理演算を行う 検索されたレコードの項目値を表示する 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。

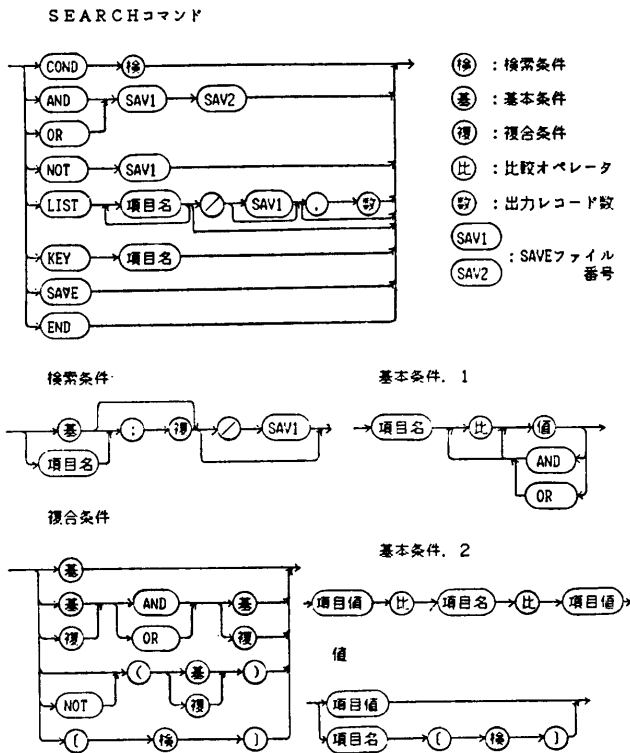


図 4 SEARCH コマンドの構文規則
Fig. 4 Syntax of SEARCH command.

図 4 に SEARCH コマンドの構文規則を示す。デリミタは空白となっている。具体例を用いて SEARCH コマンドの概略について述べる。

「男の発声で無声破裂音を含む単語が欲しい」は次のように翻訳される。

```
CONDITION WORD; SUBJECT-SEX=MALE
(次の条件の)単語が欲しい 男の発声で
AND SYLLABLE=VOICELESS-PLOSIVE+s
無声破裂音を含む
```

また、省略形を使えば次のようにも書ける。

```
COND WD; SS=MALE AND SY=VLP+s
```

“COND” は次に検索条件が続くことを意味している。“WD” は検索対象を表している。IQL では検索対象として連続音声 (CS), 単語 (WD), CV 音節 (SY), VCV 音節 (VCV), 音韻 (PH) の 5 種類が想定されている。“;” 以後に検索対象が満足すべき条件が示されるが、順序は任意であり、一般的には“項目名=項目値”の形をしている。また前述のごとく、IQL では音声の階層性を仮定しているため、PH<(SY, VCV)<WD<CS (ただし $\alpha < \beta$ は β が α の上位にあることを表す) なる関係が暗黙に仮定されている。したが

って、

WD; SY=VLP+s

と書けば無声破裂音を含む単語を意味し、

PH; SY=BA

と書けば音節 /BA/ に含まれる音韻を意味する。音声の専門用語を項目値として使うことも許されている。たとえば、有声破裂音 (Voiced Plosive) を使うと

SY=VOICED-PLOSIVE+A

SY=BA OR DA OR GA

の二つの条件は同じ意味となる。ここで

“VOICED-PLOSIVE”の代りにその省略形“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“+s”を入力すればよい。さらに多重検索も行え、“[”, “]”で囲まれた区間の条件が1回の検索に対応する。“[”の前が項目名の場合にはこの検索で見つめられたレコードの項目名に対応する項目値がその値となり次の検索における条件に加えられる。項目名以外の場合には検索されたレコードの ID コードが値となる。

以上のように、SEARCH コマンドでは単に検索条件を羅列するだけで検索可能であること、また、専門用語を直接使うことができ、埋込み文を多重検索機能を用いて容易に表現できることから、かなり自然語に近く使いやすいコマンドとなっていることがわかる。

3.2 その他のコマンド

SPEECH-DB は一般の文献データベースとは異なり「音声」を格納する数値データベースであることから、検索されたデータの確認のためのグラフィック表示、ならびにその属性を知るための標準的な分析等の機能が必要となる。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ある。MOVE 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、ANALYSIS 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、利用者のファイル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ために設けられている。また、データロード用コマンドも用意されている。

3.3 音声コマンド入力・応答サブシステム⁵⁾

SPEECH-DB では利用者とのマン・マシンインタフェースをさらに改善するために、音声によってシス

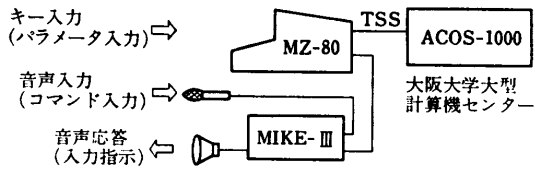


図5 音声コマンド入力・応答サブシステムのブロック図
Fig. 5 Block diagram of the subsystem for spoken command input and voice response.

テムと対話できるように配慮されている。本サブシステムは図5に示すように、Centigram Corp. 製の音声認識・応答装置 MIKE-III と、SPEECH-DB を格納しているホストコンピュータ ACOS 1000 とサブシステムとのインターフェースとしてのパーソナルコンピュータ MZ-80 によって構成されている。入力操作はキーと音声を用いし、音声によってコマンドおよびサブコマンドを認識し、パラメータをキー入力する形式をとっている。動作の概略を以下に示す。

- ① 『SPEECH-DB へようこそ』(音声応答)
 - ② 『コマンドを入力してください』(音声応答)
 - ③ コマンドを入力 (音声入力)
 - ④ 『サブコマンドを入力してください』(音声応答)
 - ⑤ サブコマンドを入力 (音声入力)
パラメータが必要ならば⑥なければ⑥
 - ⑥ 『パラメータをタイプインしてください』(音声応答)
 - ⑦ パラメータを入力 (キー入力)
 - ⑧ コマンド実行
1分ごとに『しばらくお待ち下さい』(音声応答)
 - ⑨ 結果の出力 コマンドにより②または④へ行く
- MIKE-III は特定話者・認識語句登録方式であり同時に99個の認識語句を指定できるが、本サブシステムではコマンドの階層性を利用して同時に認識すべき語句を最小限に止めることにより認識率の向上が図られている*。

4. 実働化

4.1 INQ ファイルの実現

各 INQ ファイルの論理構造を表1,2に示した項目を基にして実現した。図6に音韻ファイルの論理構造をFDLで記述した例を示す。他の INQ ファイルも同様に実現された。

4.2 音声データ

音声データはフレーム単位に定められた音韻をもと

```
FDL FRAME, 1.
DATABASE SPEECH-DB.
02 ID-FRAME PIC CP6 PKY.
02 ID-SYLLABLE PIC CP6.
02 ID-VCV (N).
03 ID-VVAL PIC CP6.
02 ID-WORD PIC CP6.
02 ID PIC CP6.
02 FNUN PIC CP6.
02 PHONEME PIC X(4).
02 RANK1 PIC X(4).
02 RANK2 PIC X(4).
02 RMS PIC FB UNIT V.
02 ZERO-KOSA PIC FB UNIT 1/RS.
02 PITCH PIC FB UNIT MS.
02 DATA-FROM PIC CP6.
02 DATA-TO PIC CP6.
02 VUS-MANUAL PIC X(4).
02 VUS-ALGORITHM PIC X(4).
```

図6 音韻ファイルの FDL 記述

Fig. 6 FDL description of phoneme file.

にして音節、単語等の区切りを求め対応する INQ ファイルへ格納した。現在格納されているデータは成人男性5名により発声された次の五つの連続音声*、成人男性10名により単独に発声された5母音、および成人男性1名により発声された525個の VCV 音節である。

- 1) 爆音が銀世界の高原に広がる。
- 2) 今日の天気は曇りのち晴れるでしょう。
- 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。

図7に音韻名のラベル付けの例を示す。上の部分は「…の授業は…」の部分の時間波形、下はホルマント周波数(声道の共振周波数)の時間的遷移を示している。/N-O/, /A 定/, /A 非/ はおのおの /N/ から /O/ への遷移部分、/A/ の定常部分、/A/ の非定常部分を示す。このように音韻間の遷移部分および音韻の定常性に関する情報を格納し、それらを条件とした音声データの検索を可能とすることにより SPEECH-DB の有用性が高められるものと思われる。音声データを格納するにはセグメンテーション(フレーム単位のラベル付け)を行わねばならない。現在、1秒の連続音声のセグメンテーションに1~2時間要しており、音声データ作成の障害となっている**、このことは5章において再び論じられる。

4.3 IQL の実現

SEARCH コマンドは約5,000ステップから成る FORTRAN プログラムで構成されており、INQ の

* 日本語のすべての子音が含まれている。

** 概算であるが、上記5種類の連続音声1人分の格納に約5日間要した。

* 現在、認識率は約95%程であり実用上問題は無い。

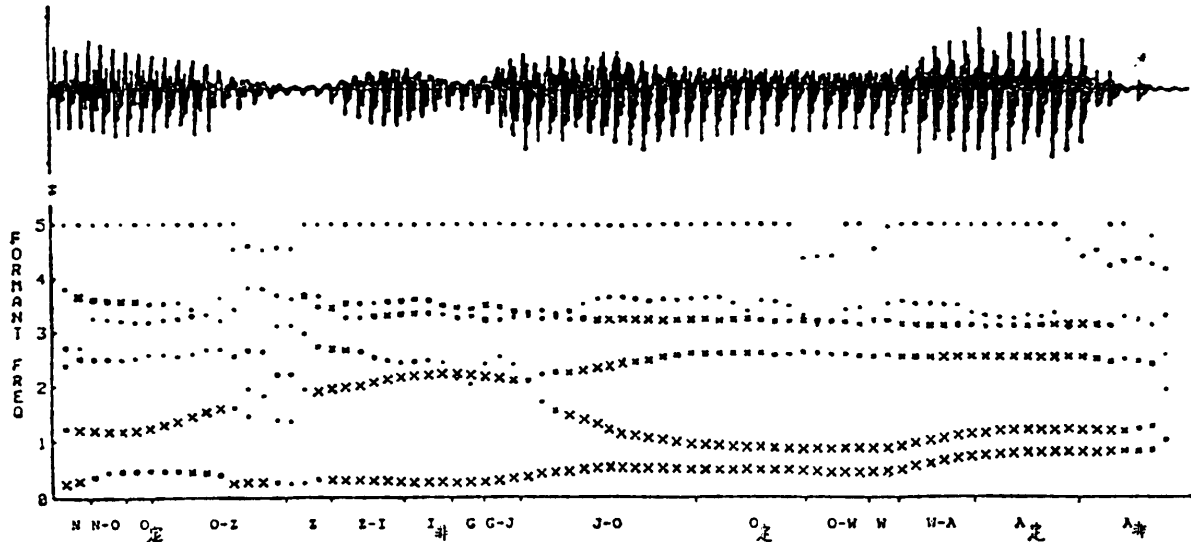


図7 セグメンテーションの例 (…の授業は…).

Fig. 7 An example of segmentation of speech (…NO ZIGJOWA…).

データ操作言語 DML を用いて SPEECH-DB を操作している。IQL では外部スキーマを定義するという負担が利用者から取り除かれているが、そのためには利用者が入力した検索条件から暗黙に仮定されている外部スキーマを推定する必要がある。IQL では3.1節で述べた検索対象と VCV 音節ファイルの参照の必要性

に応じて8種類の外部スキーマ (INQ セクション) が用意されており、検索条件に適したものを求め適宜使い分けている。INQ では同時に四つ以下のファイルの結合が可能であるが、音声は図2に示したように4レベルからなる階層構造をなしていることから不都合なく INQ セクションが構成できる。INQ では検

| 環境 | | 単語 | | CV音節 | | | 音韻 | | | | | | |
|------|-------|------|------|-------|------|------|------|-------|------|------|------|------|-------|
| ID=1 | ----- | ID=1 | IW=1 | ----- | ID=1 | IW=1 | IS=1 | ----- | ID=1 | IW=1 | IS=1 | IF=1 | ----- |
| ID=1 | ----- | ID=1 | IW=1 | ----- | ID=1 | IW=1 | IS=1 | ----- | ID=1 | IW=1 | IS=1 | IF=2 | ----- |
| ID=1 | ----- | ID=1 | IW=1 | ----- | ID=1 | IW=1 | IS=1 | ----- | ID=1 | IW=1 | IS=1 | IF=3 | ----- |
| | . | | . | | | . | | | | . | | . | |
| | . | | . | | | . | | | | . | | . | |
| | . | | . | | | . | | | | . | | . | |

(a)

| 環境 | | 単語 | | CV音節 | | | 音韻 | | | | | | |
|------|-------|------|------|-------|------|------|------|-------|------|------|------|-------|-------|
| ID=1 | ----- | ID=1 | IW=1 | ----- | ID=1 | IW=1 | IS=1 | ----- | ID=1 | IW=1 | IS=1 | IF=1 | ----- |
| ID=1 | ----- | ID=1 | IW=1 | ----- | ID=1 | IW=1 | IS=1 | ----- | ID=1 | IW=1 | IS=1 | IF=2 | ----- |
| | . | | . | | | . | | | | . | | . | |
| | . | | . | | | . | | | | . | | . | |
| | . | | . | | | . | | | | . | | . | |
| ID=1 | ----- | ID=1 | IW=1 | ----- | ID=1 | IW=1 | IS=1 | ----- | ID=1 | IW=1 | IS=1 | IF=8 | ----- |
| ID=1 | ----- | ID=1 | IW=1 | ----- | ID=1 | IW=1 | IS=2 | ----- | ID=1 | IW=1 | IS=2 | IF=9 | ----- |
| ID=1 | ----- | ID=1 | IW=1 | ----- | ID=1 | IW=1 | IS=2 | ----- | ID=1 | IW=1 | IS=2 | IF=10 | ----- |
| | . | | . | | | . | | | | . | | . | |
| | . | | . | | | . | | | | . | | . | |

(b)

図8 INQ セクションにより記述された仮想ファイルの例

Fig. 8 Two examples of virtual files described by INQ section.

索の単位となるレコードをもつファイルはベースファイル（一つの INQ セクションに一つ）と呼ばれ、他のファイルは結合ファイルと呼ばれる。図 8 (a) は音韻ファイルをベースファイルとした INQ セクション記述により生成された仮想的なファイル（関係）の例を示している。ここで、ID, IW, IS, IF はおのおの、環境、単語、音節、音韻ファイルの ID 番号である。これは音韻を検索単位とした検索を想定しており、階層構造は分解されて音韻ファイルのレコード（1 フレームの情報）ごとにすべての項目が展開されている。図 8 (b) は CV 音節単位の検索用であり、環境、単語、CV 音節ファイル間の関係は(a)と同一で、音韻ファイルだけが CV 音節ファイルの下位のファイルとして位置づけられている。8種類の INQ セクションの内訳は、ベースファイルとして環境、単語、音韻の各ファイルを選んだとき CV 音節と VCV 音節のどちらを用いるかで6通り、および CV 音節、VCV 音節の各ファイルをベースファイルとするときの2通りとなっている。用意された上述の8種類の外部スキーマでほとんどすべての有意な検索要求に応じることができることが確かめられている。このように SPEECH-DB への検索要求という限定された状況においては比較的容易に外部スキーマの自動切り換え能力を有するコマンドが実現され得ることがわかる。したがって利用者は、SPEECH-DB が常に考えうるすべての項目を含み、かつ各自の要求に適した論理構造をもっているとみなして検索を行うことができる。なお、他のコマンドも同様に約 2,000 ステップの FORTRAN プログラムで実現され、音声コマンド入力・応答サブシステムは約 1,500 ステップのアセンブリ言語で実現された。最後に、付録において SPEECH-DB の使用例を示す。

5. む す び

知的アクセス機能をもつ音声データベース SPEECH-DB について述べた。音声データベースはその必要性は早くから認識されながらも音声データの選択およびセグメンテーションの困難さ等も原因して、実現されることがあっても検索機能をもたない個々の研究者用の小規模なものであった。ところが、近年音声認識装置の実用化、さらにはデータベース管理システムの進歩によって音声データベース構築の気運が高まり、いくつかの報告がなされている^{12), 13)}。なかでも電総研¹³⁾は音声データの収集とその標準化に力を注い

でおり、その成果は期待されている。このようなときに SPEECH-DB のような本格的な音声データベースが構築されたことは意義深いものと考えられる。

今後の課題としてはデータ量の拡大が挙げられる。現在、筆者らによりデータ格納の作業が継続的に行われているが、データの選択に関しては必ずしもすべての利用者にとって満足のいくように行われているとは限らない。このことは音声データの標準化の困難さを端的に表している。しかしながら、SPEECH-DB は DBMS として INQ を用いていることから基本的に開かれた性格をもっているため、各研究者が必要と思われる音声データを選択し格納することによりデータベースのすべてのコマンドを利用することができる。したがって、蓄積すべき音声データの選択およびその格納に関して、筆者らが単独に行うのではなく、利用者自身が参加して行うことが可能である。また、そうすることによって初めて多くの利用者にとって有用なデータをもつデータベースが完成されるものと思われる。上述の目的のためには、音声データの格納作業の簡易化が必要となるが、SPEECH-DB の IQL では 3.2 節において述べたようにデータ格納の手続きがコマンド化されている³⁾。データ格納用コマンドでは、音声データのセグメンテーションという最も煩雑な作業をグラフィックディスプレイと各種の分析・表示コマンドを採用したインタラクティブな方法で行えるようになっている。さらに、ロード用ファイルの作成とロードの実行もコマンド化されているため、データを格納する際に必要となる詳細な知識をもたない一般利用者にもデータの格納が行える。データの質と量および利用環境のすべてにおいて優れたデータベースは 1 日にして成るものではなく、今後他の音声研究者の利用およびデータの格納の両面における協力を得て、SPEECH-DB の拡張を進めていきたいと考えている。

謝辞 音声コマンド入力・応答サブシステムのインプリメントに協力していただいた関西大学工学部電気工学科 4 年生木村景三君（現在日本電気株式会社）ならびに、INQ に関して有益な助言を賜った大阪大学大型計算機センター磯本征雄講師に深謝する。

参 考 文 献

- 1) 前田, 浜口, 溝口, 柳田, 角所: 音声データベース (SPEECH-DB) の試作—設計方針及びデータモデルについて—, 音響学会講演論文集, 3-1-8 (1981).

- 2) 浜口, 前田, 溝口, 柳田, 角所: 音声データベース (SPEECH-DB) の試作—コマンド体系及びその使用例について, 音響学会講演論文集, 3-1-9 (1981).
- 3) 溝口, 前田, 芥子, 柳田, 角所: 音声データベース SPEECH-DB におけるデータロード支援サブシステム, 音響学会講演論文集, 3-1-7 (1981).
- 4) 前田, 浜口, 溝口, 柳田, 角所: 音声データベース (SPEECH-DB) の知的コマンドについて, 情報処理学会全国大会講演論文集, 2F-2 (1981).
- 5) 木村, 西山, 前田, 溝口, 角所: 音声データベース (SPEECH-DB) における音声コマンド入力・応答システムについて, 電気関係学会関西支部連合大会講演論文集, G15-6 (1981).
- 6) 前田, 浜口, 芥子, 溝口, 柳田, 角所: 音声データベース “SPEECH-DB”, 信学技報, EA81-56 (1982).
- 7) Chamberlin, D. D. and Boyce, R. F.: SEQUEL: A Structured English Query Language, Proc. of ACM-SIGFIDET Workshop, Ann Arbor, pp. 249-264 (1974).
- 8) Boyce, R. F., Chamberlin, D. D., King W. E. and Hammer, M. M.: Specifying Queries as Relational Expressions: SQUARE, IBM Report RJ 1291 (1973).
- 9) Zloof, M. M.: Query-By-Example, AFIPS Conf. Proc., Vol. 44, pp. 431-438 (1975).
- 10) 日本電気(株): INQ 概説書, INQ 文法説明書, INQ 運用説明書, INQ エンドユーザ言語 (EQL) 説明書 (1981).
- 11) 磯本, 定岡, 田中, 松浦, 角戸: 学術研究用たんぱく質データベース PROTEIN-DB, 情報処理, Vol. 21, No. 1, pp. 15-22 (1979).
- 12) 小玉, 高橋: INQ による音声情報データベース試作, 情報処理学会データベース管理システム研究会資料, 22-1 (1980).
- 13) 板橋, 大村, 三国, 速水: 音声データベースの構成について, 一標準化周波数変換方式の検討, 音響学会講演論文集, 3-1-8 (1981).

付 録 IQL の使用例

図9に IQL の使用例を示す。下線部が利用者によってタイプされた部分である。

① COND 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) でそのランク (RKI) が定常 (A) のものを含み, かつ定常な音韻 /N/ を含む単語 (WD) で, その平均ピッチ (WAP) が 7ms より小さいもの」である。

② 2 回目の検索であり条件は「話者の性別 (SS) が男 (MALE) で年齢 (SAG) が 20 歳以上の人の発声で, 破裂音 (PLOSIVE) の音節 (SY)」である。

③ ②で破裂音の後続母音の指定がなかったので IQL が聞き返している。

④ 2重検索の例であり条件は「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。

⑤ セーブファイル1にセーブされた1回目の検索結果を利用者のファイルに転送する。

⑥ 転送の前にデータのチェックをするために DISPLAY コマンドに入った。

⑦ 検索されたデータすべてを転送するかどうか聞いている。

⑧ 転送される各データの開始点と終点を変更するかどうか聞いている。

⑨ /BAKUON/ の /BA/ の部分のみを転送すると指定している。

(昭和 57 年 4 月 21 日受付)

(昭和 57 年 11 月 8 日採録)

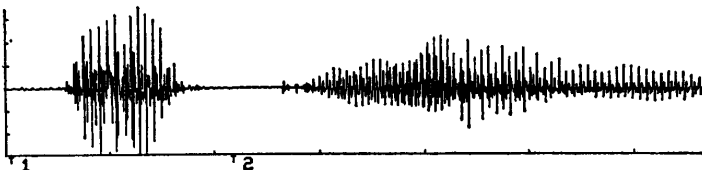
```

SYSTEM ?/SEARCH/
TYPE IN SUBCOMMAND
COND/
① WD : ( PH = A AND RK1 = A )/
  AND ( PH = N AND RK1 = A )/
  AND WAP < 7.0/
  /
5 RECORDS FOUND.
TYPE IN SUBCOMMAND
LIST IW WD WDF WDT/
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
SAVE/
SAVED FILE ID NUMBER = 1
TYPE IN SUBCOMMAND
COND SY = PLOSIVE : SS = MALE AND SAG GE 20/
② FOLLOWING VOWEL ?
A/
16 RECORDS FOUND.
TYPE IN SUBCOMMAND
LIST SY /, /
-----
SYLLABLE        BA
-----
SYLLABLE        GA
-----
SYLLABLE        KA
-----

```

⋮

```

TYPE IN SUBCOMMAND
④ =COND WD ; SY = UP + A AND DL = DLC IB = 1 J J
      6 RECORDS FOUND.
TYPE IN SUBCOMMAND
⑤ =MOVE /1 J
RAW DATA SIGNAL CHECK (YE OR NO)?
=YE J
⑥ TYPE IN SUBCOMMAND
=SIGNAL J
DISPLAY FORMAT DECISION!!!
      :
START POINT,END POINT ?
=51,6700 J
      :

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

⑦ MOVE ALL DATA(YE OR NO)?
=NO J
TYPE IN ID-WORD
=1 J
=J

⑧ CHANGE FROM,TO(YE OR NO)?
=YE J
TYPE IN ID-WORD TO BE CHANGED
=1 J
TYPE IN FROM,TO(ID-WORD= 1)

⑨ =107,2235 J
TYPE IN ID-WORD TO BE CHANGED
=J
OUTPUT USER FILE NAME
ASSIGNED USR ID/CATALOG/---/FILE ?
=6088050333/DEMO J
ASSIGNED FILE =6088050333/DEMO
FILE OK (YE OR NO) ?
=YE J
**FILE-FORMAT
1) PACKED RANDOM FILE
2) NORMAL RANDOM FILE
SELECT FILE FORMAT (1 OR 2)?
=1 J
TYPE IN SUBCOMMAND
=END J

```

図 9 IQL の使用例

Fig. 9 Dialog between a user and SPEECH-DB using IQL.