

Title	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
Author(s)	溝口, 理一郎; 角所, 収
Citation	大阪大学大型計算機センターニュース. 46 P.71-P.81
Issue Date	1982-08
Text Version	publisher
URL	http://hdl.handle.net/11094/65537
DOI	
rights	

Osaka University Knowledge Archive : OUKA

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

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 —

大阪大学産業科学研究所

溝 口 理一郎・角 所 収

1. はじめに

昭和56年度の本学計算センターにおける開発研究の援助を受けて開発された音声データベース「SPEECH-DB」⁽¹⁾⁽²⁾についてその概略を述べたいと思います。最近、音声認識装置とか音声合成用LSIとかいう言葉を耳にする機会が多くなり、特にマイコンが好きな方の中には「音声」というものに興味をお持ちの方も沢山いらっしゃると思います。どんな研究をするにしてもデータを管理することは大切ですが、特に音声認識の研究では重要な事となりつつあります。現在、特定話者用の数10単語程度の単語音声認識装置が実用化されていますが、なんと言っても理想は話者を特定しない連続音声認識です。しかし、話者が変わると音声の特徴も大きく変わる上に、同じ話者であっても連続音声中の音韻（母音、子音等の音声の単位）の特徴が前後の音韻の影響を受けて変化します。そこで、それらの変動の様子を詳細に分析する必要があり、その為に大量のデータの収集とその管理が必要となるわけです。従来から小規模な音声データ管理システムは各音声研究者が個々に持っていました。SPEECH-DBのように検索機能がある本格的な音声データベースは日本で初めてのものです。SPEECH-DBは一応音声認識研究を支援することを念頭において設計されていますが、後で述べますようにその他色々な研究にも利用できるようになっています。

データベースを苦勞して作っても使ってもらえなければなんにもなりません。その為、SPEECH-DBではデータベースをあまり知らない人でも簡単に使えるように色々と工夫されています。先ず、欲しいデータが満足すべき条件を適当に並べるだけで検索ができます。このことは当然のように思う方も多いと思いますが、実は通常のデータベースでは条件を入力する前に、利用者が考えているデータベースの論理構造をシステムに教えてやる必要があります、これが結構面倒なことなのです。次に、検索も含めて全部で5個のコマンドと31個のサブコマンド全てを音声で入力することができます。（残念ながら、事前に自分の声を登録しておかなければなりません）そればかりでなく、入力を指示するシステムからの音声による応答があるので。

このように、SPEECH-DBは単に音声データを格納管理するだけでなく、知的アクセス機能を持つ、使い易いデータベースとなる様に設計されていますが、以下ではその中身について簡単に述べてみようと思います。

2. SPEECH-DBの利用法

2.1 音声データの利用

音声データの検索とは大体どういう風にするのか、検索の際にどんな条件を指定できるのか、といった疑問をお持ちの方の為に、次に5種類の検索要求例をお見せします。

- Q 1. 男性の発声した連続音声で破裂音を含むものが欲しい。
- Q 2. 20才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. 女性の発声で鼻音と母音/e/からなるCV音節が欲しい。
- Q 5. 東京方言の20才以上35才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。

ここで、破裂音とは子音のp, t, k, b, d, gを意味し、後の3つを有声破裂音と言います。鼻音とはmとnのことです。又、Cは子音ConsonantのCで、Vは母音VowelのVです。色々な検索ができることがお分かりいただけたと思いますが、音声の研究に意味があるような条件付けは大部分が可能です。

SPEECH-DBはTSS端末からグラフィックディスプレイを介した会話的な方法で使うことも、バッチ等の応用プログラムからデータファイルのように使うこともできます。応用プログラムで直接データベースからデータを検索するにはデータベースに関する知識が相当必要ですが、TSS端末から使う場合には我々が開発した知的コマンドIQL(Intelligent Query Language)が使えるので極めて簡単に検索ができます。次にそのIQLコマンドについて述べます。

2.2 IQL

表1にIQLコマンド及びその機能を示します。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ます。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段階に分れています。ここでは検索に用いられるSEARCHコマンドについて述べます。次の検索要求を例にとります。

「男の発声で無声破裂音を含む単語が欲しい」

これは次のように翻訳されます。

```
CONDITION WORD ; SUBJECT-SEX = MALE AND  
(次の条件の)単語が欲しい           男の発声で
```

```
SYLLABLE = VOICELESS-PLOSIVE + $  
           無声破裂音を含む
```

表 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。

省略形を使うと次のようにもかけます。

COND WD; SS=MALE AND SY=VLP+\$

「COND」は次に検索条件が続くことを意味しています。「WD」は検索対象を表わしています。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ます。「;」以後に検索対象が満足すべき条件が示されますが、順序は任意であり、一般的には「項目名=項目値」の形をしています。ところで、音声には図1に示す様な階層構造があると一般に考えられていますが、IQLでもやはりこの階層構造を想定している為、 $PH < (SY, VCV) < WD < CS$ (但し、 $\alpha < \beta$ は β が α の上位にあることを表わしています)なる関係が暗黙に仮定されています。従って、

WD; SY=VLP+\$

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します。音声の専門用語を使うことも許されています。例えば、有声破裂音(Voiced Plosive)を使うと

SY=VOICED-PLOSIVE+A

SY=VP+A

SY=BA OR DA OR GA

の3つの条件は同じ意味となります。又、「+A」は後続母音を/A/と指定することを意味し、後続母音を問題にしない場合には「+\$」と入力します。さらに多重検索も行え、「」,「[」で囲まれた区間の条件が1回の検索に対応します。「[」の前が項目名の場合にはこの検索で見つけ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ます。項目名以外の場合は検索されたレコードのIDコードが値となります。以上のことから、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、又、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易いコマンドとなっていることがお分かりいただけると思います。詳しいことは付録の使用例を参照して下さい。

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為にグラフィック表示、並びにその属性を知る為の標準的な分析等の機能が必要となる為、検索性以外にもいくつかのコマンドが用意されています。DISPLAY コマンドは音声波形等の簡単なグラフィック表示を行うコマンドです。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す。その他、システムに不慣れた利用者にコマンドの使い方を説明する為にHELPコマンド、データの格納用の

ロードコマンド等が用意されています。

2.3 音声コマンド入力・応答サブシステム

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為に、音声によってシステムと対話ができるように配慮されています。本サブシステムは、Centigram Corp 製の音声認識・応答装置MIKE-IIIとパーソナルコンピュータMZ-80 によって構成されています。入力動作はキーと音声を併用し、音声によってコマンド及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ります。動作の概略を以下に示します。

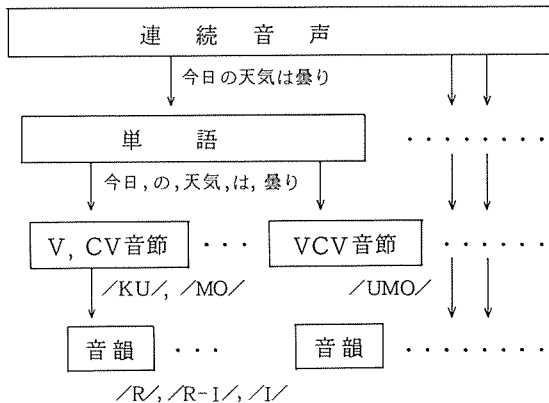


図 1. 音声の階層構造

- 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⑧
- 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分毎に「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

3. SPEECH-DBの実現

3.1 基本構成

データベース管理システムとしては INQ (INformation Query) を用いました。SPEECH-DBは図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つの独立したファイルから構成されており、音声の階層構造はINQセクション記述により仮想的に実現されています。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、メモリ節約の為に1WORD(36bit)

に3サンプル点ずつ格納されています。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となっています。項目の数が比較的多くなっていますが、利用者の多様な検索要求に応じる為に必要かつ十分なものと考えています。

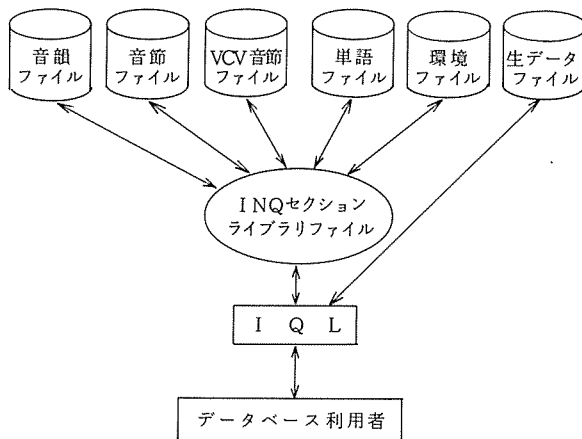


図2. SPEECH-DBの基本構成

3.2 IQLの実現

IQLのSEARCHコマンドは約5300ステップからなるFORTRANプログラムで構成されており、INQのデータ操作言語DMLを用いて検索を行っています。FORTRANは言語としては好きではないのですが、INQはFORTRANかCOBOLでしか使えないことも含めて、ACOS-6の優れた機能を最大限に利用するには今のところFORTRANに勝る言語はないようです(WESTRANはその点良くできていますが)。他のコマンドも約2000ステップのFORTRANで、音声コマンド入力・応答サブシステムは約1500ステップのアセンブリ言語で実現されています。SEARCHコマンドが知的であるというのは利用者が羅列した検索条件から利用者自身の頭にある音声データの論理構造を推定して、データベースを仮想的にそのような構造として実現することにより検索を実行するからです。この様なことが簡単にできるのには一つにはINQが優れたDBMSであることも原因となっています。

4. むすび

音声データベースSPEECH-DBについてその概略を述べて来ました。現在、枠組としては完成しており試験的に稼動しています。データに関しては我々が判断した「標準的なデータ」の格納を継続的に行っています。又、将来は音響学会の方で進められている音声データの標準化の結果を参考にしてデータの格納を行うことを計画しています。しかし、何年経ったとしても全ての利用者に満足していただけるデータを用意することは不可能だと思います。そこで読者の皆様をお願いしたいことがあります。「こういう音声データが入っていれば使うのだが……」とお考えの方は是非御自分でデータを格納していただきたいのです。格納しさえすればSPEECH-DBの全機能が使えるようになります。「データベースのことは何にも知らないのに格納なんてとても出来やしない」と

言われる方の為に、SPEECH-DBではデータ格納の為の操作をインタラクティブにやってくれるコマンドが整備されています。従って、コマンドからの質問に答えるだけで、データベースに関する予備的知識無しで御自分の音声データが簡単に格納できるようになっています。言語学の研究をされている方であればアクセントパターン及びピッチパターン等の分析が容易になりますし、人間以外の動物の声等を格納しても面白い(特に人間科学部には興味をお持ちの方もいらっしゃると思います)でしょう。このように、我々はSPEECH-DBの開かれた利用形態を想定していますので、データベースの利用からデータの格納まで、できるだけインテリジェント化して、「データベースはあれば便利だが作るのは勿論、使うのも面倒だ」という方にも参加していただけるように工夫を凝らした次第です。現在、今秋からの正式サービスを目差してマニュアル作り等の整備を行っています。付録に使用例とその注釈を示しておきました。御興味のお有りの方は溝口(内線3566)まで御連絡下さい。

〈 参 考 文 献 〉

- (1) 溝口, 他: “知的アクセス機能を持つ音声データベース「SPEECH-DB」,” 情報処理学会データベース管理システム研究会資料, 29-6 (1982)
- (2) 前田, 他: “音声データベースSPEECH-DB”, 信学研資, EA 81-56 (1982)

下線部が利用者によってタイプされた部分です。

```
① SYSTEM ?/SEARCH
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
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
```

レコード(キャリジリターンにより)と指定しました。

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ます。

①SEARCH コマンドを起動しています。

②サブコマンドの入力待ち状態です。

③COND サブコマンドを入力して検索を行います。条件は「/A/という音韻(PH)でそのランク(RK I)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 msecより小さいものです。

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(WD), 単語の開始点(WDF), 終点(WDT)を表示させます。

⑤出力レコード数を全検索

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
-----

```

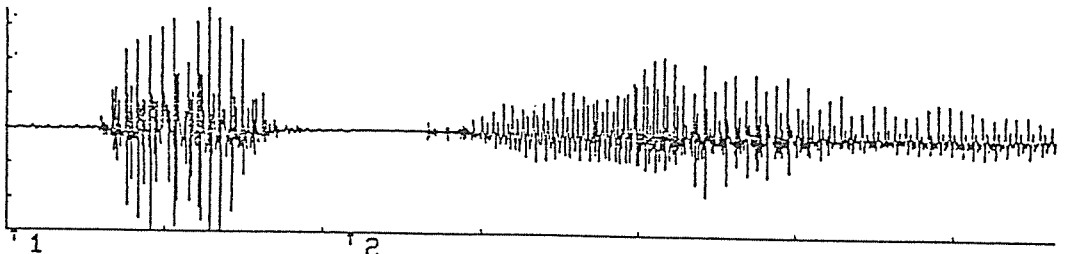
TYPE IN SUBCOMMAND
 ⑩ =COND WD ; SY = UP + A AND DL = DL IB = i]
 6 RECORDS FOUND.

- ⑦再び検索を行っています。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)です。
- ⑧⑦で後続母音の指定がなかった為にIQLが開き返しています。
- ⑨“/,”と指定すると全検索レコードを出力します。
- ⑩今度は2重検索を行っています。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」です。

```

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
  = /
⑭ DETAIL OR GROBAL(NORMAL=GROB)?
  = /
  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!!

```

- ⑪ MOVE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します。
- 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ます。
- ⑬ DISPLAYコマンドのサブコマンドSIGNALで波形を表示させます。
- ⑭ 波形表示の為の条件を設定しています。

⑮カーソルにより1, 2の点のポイント数を表示させます。

⑯SIGNALサブコマンドを終了します。

⑰DISPLAYコマンドを終了します。

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

㉕ TYPE IN SUBCOMMAND
=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ます。

⑲転送するレコードのID-WORDを聞いています。

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ます。

㉑変更するレコードのID-WORDを聞いています。

㉒変更した点を入力します。

㉓利用者の転送ファイルのカタログファイル名を入力します。

㉔転送ファイルの形式を指定します。IQL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ます。

㉕ENDサブコマンドでSEARCHコマンドを終了しました。

1)