知的アクセス機能を持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
（大阪大学 産業科学研究所）

1 まえがき

不特定話者・大語彙を対象とした音声認識装置の開発には音韻単位の認識を基礎とする必要があることは周知のことであるが，連続音声中の音韻は話者の違いによる影響は勿論のこと，調音結合の影響により大きく左右され，それ等のバラツキの程度及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．この為には大量の音声データの収集と解析が必要となるが，検索機能を持たない従来のデータ管理方式では不十分であった．このことから，今後の音声認識研究を支援する為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，管理システムが満足すべき性質として次の4点が考えられる．

1) 大量データの蓄積及びその管理
2) 蓄積データの検索
3) 検索データの簡単な分析及びグラフィック表示
4) 全国の音声研究者の共同利用

そこで我々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4つの性質を満足し任意の音韻環境での音韻の検索は勿論，その他音声認識研究を遂行する際に必要とされる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-DB[(1)-(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知識を要求されるばかりでなく，不慣れな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持つ検索言語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でき，又IQLのコマンド及びサブコマンド全てを音声により入力することができる．[(5)] さらに，システムから入力促進等の音声応答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．TSSコマンドとしてのIQLの他に，FORTRANで書かれた応用プログラムから音声データを通常のデータファイルへのアクセスとほぼ同様にして利用することもできる．このような様々な機能を持つSPEECH-DBは，単に音声という数値データを格納・管理するだけではなく，知的アクセス機能を持ち，分析，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．本稿ではSPEECH-DBの概略について述べる．

2 音声データベース

2.1 SPEECH-DBの環境

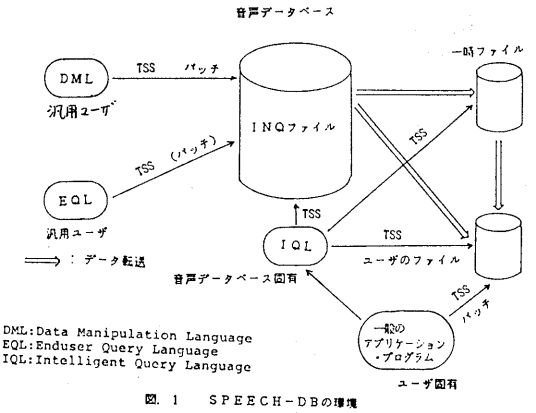


図. 1　SPEECH-DBの環境

図1にSPEECH-DBの利用環境を示した．本データベースはデータベース管理システムとして，日本電気提供のINQ[(10)]を用いている．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為のソフトウェアとしては不十分であり，特にDMLを使用する場合にはデータベースの論理構造に関する知識が必要である．これ等の高度な予備知識を持たない一般の利用者は我々が開発したIQLを用いて会話的にデータベースを利用することが可能となっている．SPEECH-DBは大阪大学大型計算機センターのACOS900上で作成されているが，大学内コンピュータネットワークを通じて7大学のTSS端末からSPEECH-DBにアクセスできる．尚，IQLでは分析・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．

2.2 音声データの階層性

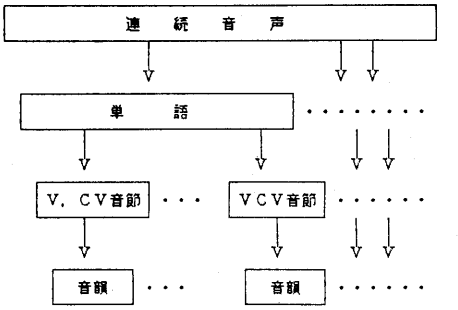
音声は図2に示すような階層構造をなしていると一般に考えられている．ここでC,Vは各々子音，母音を表わす．VCV音節の導入により音韻がCV音節とVCV音節の両方に属する場合が起り，厳密な意味での階層構造が成立しなくなる為処理が多少複雑になるが，利用者の便宜を考慮して敢えて導入した．この階層構造が利用者に要求されるデータ構造に関する唯一の予備知識であり，事実上無視できるものと考えられる．

2.3 SPEECH-DBの基本構成

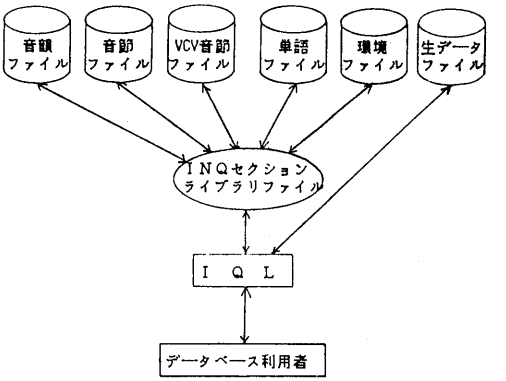
本データベースは図3に示すような音韻，音節，VCV音節，単語，環境，生データの6つの独立したファイルから構成されている．生データファイルは通常のランダムファイルで，10KHz，12bitでAD変換された音声データを メモリ節約の為に1ワード(36bit)に3サンプル点ずつ格納している．他の5つのファイルはINQファイルと呼ばれるINQシステム専用のファイルである．

2.4 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，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．表1に音韻ファイルの項目，表2に環境ファイルの項目を示す．他のファイルも同様である．



図．2　音声の階層構造



図．3　SPEECH-DBの基本構成

表　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 2　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ー（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

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易さの2点において決まると言われている。データベースへの問合わせ言語としては，QBE[9]，SEQUEL[7]，SQUARE[8] 等数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利用者がデータベースの論理構造に関する予備知識を持っていることが仮定されており，利用者にとって大きな負担となっている。SPEECH-DBは使い易さにも重点を置いて設計されており，一般の音声研究者がデータベースに関する知識なしに容易に利用できるような知的コマンドIQLが容易されている。

3.1 SEARCHコマンド

図4にSEARCHコマンドの構文規則を

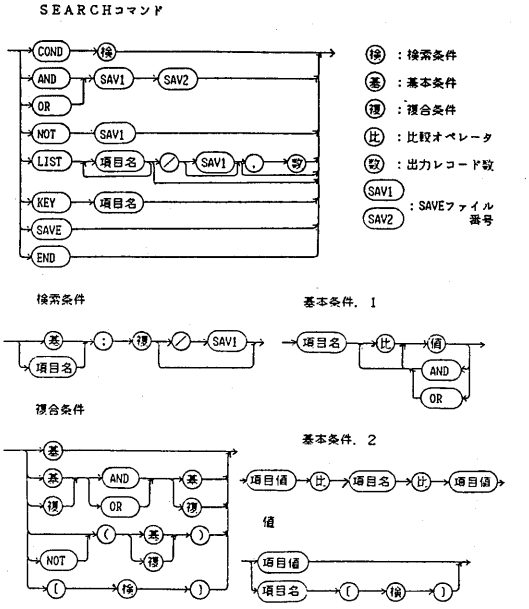


図 4　SEARCHコマンドの構文規則

示す。デリミタはブランクとなっている。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
（単語が欲しい）　（男の発声で）

AND SYLLABLE = VOICELESS-PLOSIVE + ＊
（無声破裂音を含む）

省略形を使えば次のようにも書ける。

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る。WORDは検索対象を表わしている。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"項目名＝項目値"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 α < β 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

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 /A/ 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。例えば，有声破裂音（Voiced Plosive）を使うと

SY = VOICED-PLOSIVE + A

SY = BA OR DA OR GA

の2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"を使うことができる。又"+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"+#"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けられ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検索されたレコードのIDコードが次の検索の条件に入れられる。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り，専門語を直接使えることから，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分る。

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のグラフィック表示，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。表3にIQLのコマンド及びその機能を示す。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為に設けられている。又データロード用コマンドも用意されている。

表．3　IQLの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	12種類	検索されたデータの分析・表示を行う．
HELP*	7種類	コマンドの説明を行う．
LOAD-RD LOAD-PH ⋮		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は全てのコマンドのサブコマンドとして使える．

3.3 音声コマンド入力・応答サブシステム(5)

SPEECH-DBでは利用者の負担をさらに軽減する為に音声によってシス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-IIと，SPEECH-DBを格納しているホストコンピュータ（ACOS 900）とサブシステムとのインタ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っている。使用例を以下に示し，その注釈

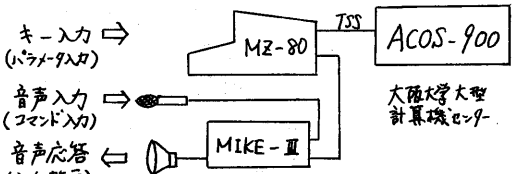


図5 音声コマンド入力・応答サブシステムのブロック図

表4. 利用者とシステムの一般的な対話.

① 'SPEECH-DBへようこそ' (音声応答)
② 'コマンドを入力して下さい' (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下さい' (音声応答)
⑤ サブコマンドを入力 (音声入力)
　パラメータが必要であれば⑥へ, なければ⑧へ
⑥ 'パラメータをタイプインして下さい' (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)

<例>

SYSTEM ?①/SEARCH/
TYPE IN SUBCOMMAND ④
=COND⑤SY = SA ; SS = MALE/
　　4 RECORDS FOUND.
TYPE IN SUBCOMMAND ④
=

番号は表 4 の音声応答
—は音声入力
—はキー入力

を表4に示す. MIKE-IIIは特定話者・認識語句登録方式で同時に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が階層的であること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計られている.

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設計した。図6に音韻ファイルの論理構造をFDL (INQで用意されているファイル記述言語) で記述した例を示す。他のINQファイルも同様にして定義された.

4.2 音声データ

音声データはフレームと呼ばれる20m sec. 長のデータ (200サンプル) 単位に音韻名を定め, それを基にして音節, 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2名により発声された次の5つの連続音声*と, 成人男性10名による

* 日本語の全ての子音が含まれている.

```
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 FNUM              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/MS.
 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.

1) 単独に発声された5母音である.
1) 爆音が銀世界の高原に広がる.
2) 明日の天気は曇り後ち晴れるでしょう.
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. 上が「…の授業は…」の部分の時間波形, 下はホルマント周波数の遷移を示している. /N-O/, /A定/, /A非/ は各々/N/と/O/の遷移部分, /A/の定常部分, /A/の非定常部分を示す. このような音韻間の遷移部分及び各音韻の定常性に関する情報までも格納し, それ等に基づく音声データの検索を可能にすることによりSPEECH-DBの有用性が高められている.

4.3 IQLの実現

SEARCHコマンドは約5000ステップから成る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, その為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

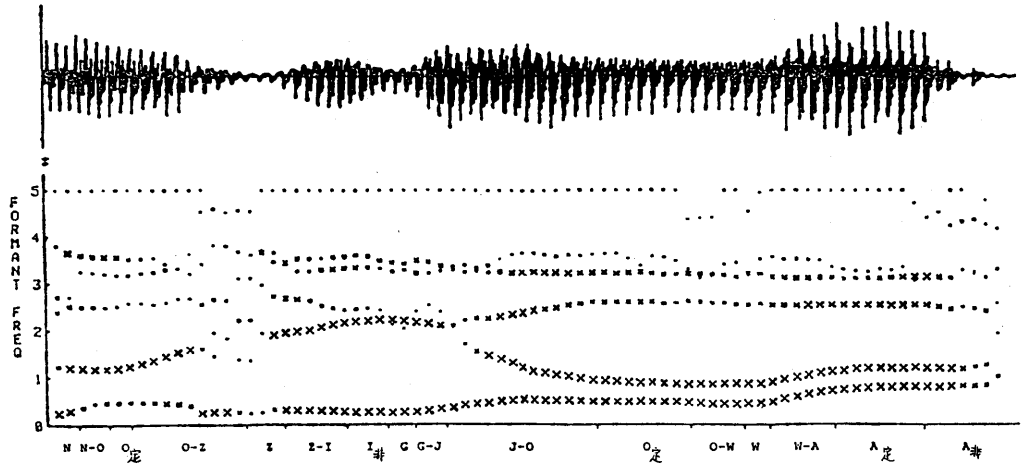


図7. セグメンテーションの例 (…の授業は…).

照の必要性とに応じて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切り換えている．用意された8種類の外部スキーマでほとんど全ての有意味な検索要求に応じることができる．このようにSPEECH-DBへの検索要求という限定された状況であれば比較的容易に外部スキーマの自動切り換え能力のある知的コマンドが実現され得ることが分る．INQにおけるデータは実際には階層構造で格納されている．従って，検索されたデータへのアクセスにはやはりその構造に関する知識が必要である．そこでLISTサブコマンドでは，単に項目名を指定するだけで現在使用されている外部スキーマを基にしてその項目値を出力することができるように工夫されている．従って利用者は，SPEECH-DBは常に考え得る項目全てを含み各自の要求に適した論理構造を持っているとみなして検索を行うことができ，高いインテリジェンスが実現されている．尚，他のコマンドも同様にして約2000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1000ステップのアセンブリ言語で実現された．

5 むすび

知的アクセス機能を持つ音声データベースSPEECH-DBについて述べた．今後，データ量の拡大と知的アクセス機能の強化を行い，近々大阪大学大型計算機センターにおいてサービスを開始する予定である．

<文献>

(1) 前田他："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 —"，音講論文集，3-1-8 (昭56. 5).
(2) 浜口他："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 —"，音講論文集，3-1-9 (昭56. 5).
(3) 溝口他："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"，音講論文集，3-1-7 (昭56. 10).
(4) 前田他："音声データベース(SPEECH-DB)の知的コマンドについて"，情処講論集，2F-2 (昭56. 10).
(5) 木村他："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 (昭56. 11).
(6) 前田他："音声データベース"SPEECH-DB"，信学技報 EA81-56 (昭57. 1).
(7) Chamberlin, D. D. et al.: "SEQUEL: A structured English Query Language," Proc. of ACM-SIGFIDET WORKSHOP, Ann Arbor, PP. 249-264, (1974).
(8) Boyce, R.F. et al.: "Specifing 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 (昭56).