音声情報の自動獲得機能を持つ分散型大規模音声データベース「K-DB」†

城風敏彦†† 牧野正三††† 城戸健一††††

音声研究には多数の話者の膨大な発声を収納し，効率よく編集，検索できる音声データベースが必要である．本研究では，従来から保存していた12Gbyteの音声データを6枚の光ディスクに収め，かつGP-IBネットワークとEthernetを併用することで，光ディスクファイルや各計算機の磁気ディスクファイルを，全計算機から利用できる音声データベース「K-DB」を構築した．K-DBは大量の音声データを，統一的かつ効率よく追加，編集，検索するため，関係型を用いている．また従来，人手に頼っていた音素ラベリング，ホルマントやピッチ周波数などの音声情報を，データベースの特徴である入力既知という情報を使って，高精度に行うシステムも備えている．さらに利用者が話者や音素環境による音素の特徴の違いを把握するためのデータベースエディタEDBを内蔵し，利用者がデータベースに関する知識がなくとも，分析を容易に行うことが可能である．EDBはネットワークやA/D, D/A変換装置などの周辺機器と結合して強力な編集，分析，表示機能を持ち，さらに音声認識システム，音声合成システムと有機的に結合しており，音声研究が容易に実行できるようになっている．

1. まえがき

音声は個人性，発声速度によって発声のたびに変動するため，音声研究において大量の音声データは必要不可欠である．大量の音声データを扱うことによって雑音的な要因に影響されない不変的な特徴を発見したり，あるいは体系的な変動モデルを構築することが可能になると考えられる．また音声分析や音声認識のアルゴリズムの有効性の確認のためには，多数話者の発声した大規模な音声データベースが必要である．

音声データベースは，従来からも電総研[1]，ATR[2]，大阪大[3]で構築されているが，小規模なものが多い．そのため単一の計算機と磁気ディスク装置を用いて構成され，構築ツールも既存のソフトウェアを使うことが多かった．また音声データを音声研究で利用しやすい形にするためには，入力した音声のどの部分がどの音素に対応しているかを示す音素ラベリングの作業や，音声の個人性や音韻性を表す基本周波数やホルマント周波数の抽出等の作業が必要である．しかし，これらの音声情報を自動的かつ高精度に得る方法が従来確立されていなかったり，不十分であったため，音声研究者が，視察によってこれらの情報を抽出していた．そのため，音声データベースの作成に膨大な時間と労力が必要であった．大規模音声データベースを作成するには，これらの音声情報を自動的に獲得するシステムを装備することが必要である．入力した音声のどの部分がどの音素に対応しているかを示すラベリング作業に関しては，いくつかの研究があるが[9)~11)]，十分なものではない．

我々は，従来から単語音声データを中心に音声資料の収集を続け，現在約12Gbyteの音声データを保有している．これらのデータ量は，従来の音声データベースに比べて格段に多く，従来のように単一の計算機や磁気ディスク装置に格納することは非常に困難である．また音声分析や音声認識では，分析結果や標準パタンとの尤度等，原データをはるかに超える2次的，3次的データが出力されるため非常に大容量の磁気ディスクが必要となる．したがってこのような大容量の音声データベースでは，単一の計算機や磁気ディスクを想定するのではなく，複数の計算機，磁気ディスクを想定した分散型音声データベースを構成し，統一的に管理することが必要となる．さらに磁気ディスク以外の安価な媒体や多種のネットワークを利用できることが望まれる．

さらに音声データベースを利用して音声研究を進めるためには，以下の機能が必要である．

1) 音声データの追加，削除が統一的に行えること．
2) 音声データが格納されている媒体の違いや場所を意識せずにアクセスできること．

† A Large Scale Distributed Speech Data Base "K-DB" with an Aquisition System of Speech Information by TOSHIHIKO SHIROKAZE (Systems Laboratories, Oki Electric Industry Co., Ltd.), SHOZO MAKINO (Research Center for Applied Information Sciences, Tohoku University) and KEN'ITI KIDO (Faculty of Information Engineering, Chiba Institute of Technology).

†† 沖電気工業(株)総合システム研究所
††† 東北大学応用情報学研究センター
†††† 千葉工業大学情報工学科

3) 音声データ検索のための種々の条件を指定できること.
4) ユーザのC言語やFORTRAN言語で書かれたプログラムからアクセスできること.
5) 音声データベース作成のための音声切り出し，音素の手動や自動のラベリングプログラム，種々の音声分析プログラムなどのユーティリティが整備されていること.

そこで我々は原データを光ディスクに格納し，2次的，3次的データを磁気ディスクに格納し，かつ全計算機からGP-IB, Ethernetを利用してアクセスでき，かつ上記の5つの機能を持つ分散型大規模音声データベース「K-DB」を作成した[4].

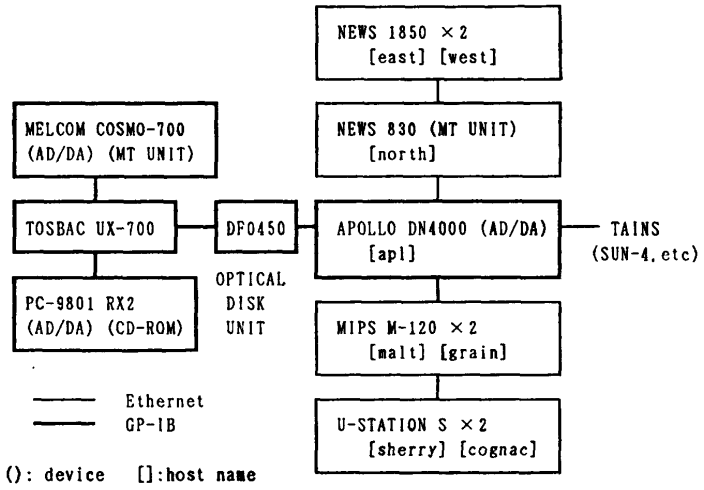


図1 計算機ネットワークの構成
(TAINS: 東北大学総合情報ネットワーク)
Fig. 1 The constitution of computer network.
(TAINS: Tohoku University's Advanced Information Network)

本データベースの大きな特色は，基本周波数やホルマント周波数の構成度抽出法と音素の自動ラベリングシステムを備えていることである．本研究では入力音声が既知であることを利用して，前述の音声情報の抽出の高精度化，自動化をはかった[5),6)]．また分析，表示において，従来のデータベースでは音素の特徴の追跡機能が不備であったが，K-DBにおいては，ホルマントの重ね書きを中心とする編集，分析，表示システムEDBを開発することで，これらの問題を改善した.

2. K-DB のハードウェア

2.1 計算機ネットワーク

図1にK-DBの計算機ネットワークのハードウェアの構成を示す．APOLLO DN 4000がGP-IBとEthernetの2つのインタフェースを持つため，DN 4000をGP-IBネットワークコントローラとする．図2にGP-IBネットワークのアルゴリズムを示す．コントローラがサービスリクエスト待ちをしている状態で，スレーブがサービスリクエストを送信してコントローラの待ちを終了する．今度はスレーブがコントローラのシリアルポーリング待ちを行い，コマンド実行の同期をとる．表1にGP-IBネットワークノード間のデータ転送速度を示す．光ディスクの動作時間と転送時間を分離して計測することができないため，この表の数値は，光ディスク，磁気ディスクのアクセス時間を含んでいる．表よりデータ転送速度が20〜70

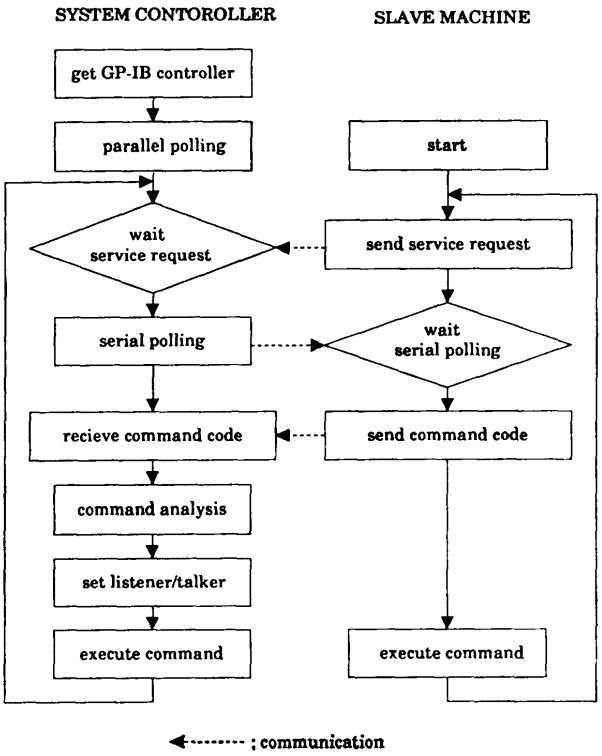


図2 GP-IBネットワークのアルゴリズム
Fig. 2 Algorithm of GP-IB network.

Kbyte/secであることがわかる.

DN 4000が光ディスクから読み込んだファイルを，全計算機はEthernetによってアクセスでき，それを分析した2次的，3次的データを各計算機の磁気ディ

表 1 GP-IB ネットワークノード間のデータ転送速度 (単位: Kbyte/sec)
Table 1 Speed of data transfer between nodes of GP-IB network. (Unit: Kbyte/sec)

source \ destination	DF 0450	APOLLO DN 4000	COSMO-700	UX-700	PC 9801 RX 2
DF 0450		30	60	50	20
APOLLO DN 4000	40		40	40	20
COSMO-700	30	30		30	20
UX-700	70	30	60		20
PC 9801 RX 2	20	20	20	20	

スクに格納することで，分散型の音声データベース K-DB を拡張していく．また東北大学総合情報ネットワーク「TAINS」と Ethernet によって接続してあり，外部からの利用も可能である．

2.2 光ディスク上のデータの管理

2.2.1 音声データ

我々がもつ音声データのうち，主なものは以下の 5 集合である．

a．274 単語集合
　男性 20 名，女性 20 名　10959 単語

b．212 単語集合
　男性 32 名，女性 40 名
　16800 単語

c．3000 単語集合
　男性 15 名，女性 20 名
　113480 単語

d．連続音声データ集合
　男性 5 名　1440 文章

e．ATR データベース[3]
　男性 2 名，女性 2 名
　23400 単語

表 2 データラベル
Table 2 Data label.

SN	LABEL
1	speaker number
2	word number
3～5	recording date
6～8	speaker's birthday
9	speaker's age
10	speaker's native place
11	utterance condition
12	flag of dialect
13	dialect number
14	flag of comment
15～30	comment
31～32	analysis type (ASCII)
33	record length
34	block length
35～40	analysis conditions
41～60	speaker's name (ASCII)
61～80	word's name
81～120	phoneme table (ASCII)
121	number of frames
122	number of phonemes
123～442	labeling data
443～480	comment

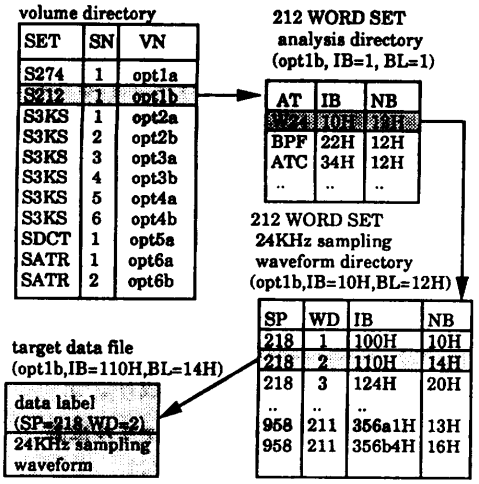
我々は BPF 分析を主としているため 24 kHz 波形データと 29 ch BPF データを基本としてデータベースを構築している．それぞれの波形データ，BPF データなど各分析データの先頭に 960 byte のデータラベルが付加してある．これは音素ラベリングファイルとして単独でも存在するが，データの独立性維持のため個々のデータにヘッダとして付加している．**表 2** はその内容を示す．なお，同一人，同一単語の異なる発声は，SN (serial number) 11 の utterance condition (雑音環境など) で区別する．発声条件が全く同じ場合でも，発声順に utterance condition を異なるものにすることで曖昧さを除く．ただし，現在のところ K-DB は同一人，同一単語で異なる発声を含んでいないので，本論文では utterance condition を省略して解説する．

2.2.2 光ディスク上のデータの管理

光ディスク（東芝 DF-0450 用）の媒体仕様は以下のとおりである[4),7)]．ブロック数 6c000 H は 16 進数の 6c000 の意味であり，本論文では以後，有効数字の後に 'H' をつけて 16 進数を表す．

a．データ容量	1.8 Gbyte/face
b．データ転送速度	2.5 Mbit/sec
c．ランダムアクセス速度	1.0 sec (mean)
d．ブロックあたりデータ容量	4096 byte
e．ブロック数	6 c 000 H block
f．媒体寿命	10 年
g．インタフェース	IEEE-488 bus (GP-IB)
h．直径	30 cm
i．記憶型	追記型

光ディスクへ波形データを移植する際，1 単語ごとに光ディスク上での先頭ブロック番号と，格納に要したブロック数を得る．これに話者番号と単語番号を加えて 1 レコードとして単語数だけ記憶装置に蓄えて単



SN: serial number, IB: initial block number, VN: volume number, NB: number of blocks, AT: analysis type, SP: speaker number, WD: word number.

図 3 光ディスク上のデータ管理

Fig. 3 Data management on optical disks.

語ディレクトリを構成する．単語ディレクトリの長さは，構築した単語数に等しい．

次の単語の先頭ブロック番号は，前の単語の先頭ブロック番号に光ディスク上で占めるブロック数を加えたものにする．これは光ディスクファイル間に未使用ブロックを残すと，ディレクトリの再構成が極度に困難になるので，物理的に連続に書き込むことが必要なためである．

光ディスクでは，表裏の2面に記録できるが，ここではその片面をボリュームと呼ぶ．ボリュームの独立性のためには，光ディスク自体にディレクトリ情報があることが重要であり，まず単語ディレクトリを1ファイルとして光ディスクのディレクトリ領域に改めて書き込む．次にその先頭アドレスと大きさを記憶し，ボリューム作成が完了した後，光ディスクの第1ブロックにデータ集合の情報を書き込んで完成する．

この様子を**図3**に示す．212単語集合は opt1bにあり，その24kHz波形データのディレクトリは10Hブロックを先頭とし12Hブロックを占めている．

3. K-DB のソフトウェア

3.1 ソフトウェアの構成

図4にソフトウェアの構成を示す．本データベースでは，統一辞書によって光ディスクや複数の計算機に付属した磁気ディスク中のファイルを管理する．また辞書を参照，編集しながら分析ライブラリを用いて音声データを分析し，磁気ディスクに分析ファイルを作成する．また編集，分析，表示システム EDB によって，音声データを検索，表示する．EDB の機能は，利用者の C や FORTRAN のプログラムからも利用できる．

データの検索方法は，データベース統一辞書を用いた音声データ集合→分析の種類→話者→単語と検索するトップダウン検索と，音素逆引き辞書を用いた音素→単語のボトムアップ検索の2つからなる．音素単位の検索，表示は，音声の静的，動的特徴を捉えるために有効であり，そのためには，ボトムアップ検索が必要である．トップダウン検索のための音素逆引き辞書は，音素ラベリングファイルから作成する．音素逆引き辞書はレコードを話者順，音素順にソートし，ハッシュ関数を用いて検索を高速化している．

3.2 辞書の構造

3.2.1 関係型統一辞書の構造

データベースを作成する際，辞書の構造はデータの追加，更新の際の一貫性が他の型に優っているため関係型[8]を導入するのが統一的である．辞書は与えられたキーから希望のデータの光ディスクと磁気ディスクへのポインタを与える．

キーは次の4つである．

1） 集合名 (SET)

2） 分析タイプ (AT) (w 24：24 kHz 波形データ，

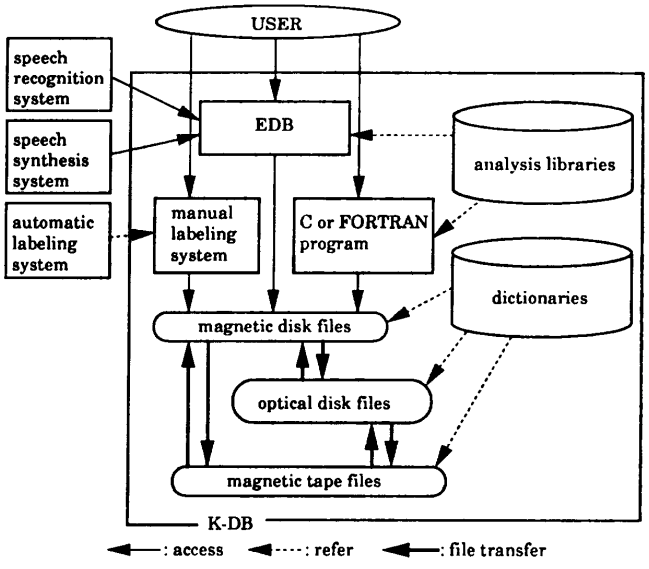


図 4 K-DB の構造

Fig. 4 The architecture of K-DB.

BPF など)

3) 話者番号 (SP) (集合に無関係，全体で共通)

4) 単語番号 (WD) (集合によって異なる単語を示す)

磁気ディスクへのポインタは次の3つ組である.

5) 磁気ディスクファイル名 (FN)

6) 磁気ディスクファイル中での先頭レコード番号 (IR)

7) 磁気ディスクファイル中でのレコード数 (NR)

光ディスクへのポインタは次の3つ組である.

8) 光ディスクボリューム名 (VN)

9) 光ディスク上での先頭ブロック番号 (IB)

10) 光ディスク上でのブロック数 (NB)

この 10 個のデータの組が辞書での1レコードとなり，これを全単語の全分析分だけならべて統一辞書を作成する．この統一辞書は長大なため，実際には，集合，分析について別々に辞書を作成する.

また以下の補助的なデータが，4つのキーのいくつかに従属に存在し，それに対応した表を作成する.

a) 話者番号→話者名など話者に関する情報.

b) 集合名，単語番号→単語名.

c) 集合名，分析タイプ→分析条件.

3.2.2 統一辞書の第三正規形への分解

統一辞書は第一正規形であり表内部に従属関係を含んでおり，光ディスクや磁気ディスク上のデータをアクセスするためには，10 個よりも少ないキーでアクセスが可能である．例えば以下の関係式の左辺がわかれば右辺が一意に求まる.

d) SET, AT→FN.

e) SET, SP, WD→IR, NR.

f) SET, AT, SP, WD→VN, IB, NB.

d), e), f)では，左辺の値が決まると右辺のすべての値が一意に定まり，左辺の属性集合はその値を求めるための最小属性集合である．このような関係を第三正規形の関係と呼ぶ．統一辞書を第三正規形に分解することで辞書のサイズを小さくし，検索を高速化することができる.

上記の関係式は，光ディスクをマルチボリュームと仮定している．しかし実際には光ディスクでマルチボリュームであるほどの巨大な集合の分析が，現状の磁気ディスクのシングルボリュームに収まることはないので，光ディスクごとに集合を分割すると以下のようになる.

g) SET, VN, AT→FN.

h) SET, VN, SP, WD→IR, NR.

i) SET, VN, AT, SP, WD→IB, NB.

すなわち g)の左辺の値を FN に入れることで g)の関係を表せる．h)の関係は磁気ディスク辞書として SET と VN をファイル名に入れ，SP, WD, IR, NR を1レコードとして並べることで表す．i)の関係も SET, VN, AT をファイル名中に含め，SP, WD, IB, NB を1レコードとして並べたファイルとする．これは 2.2 節でのディレクトリである.

また VN を集合における光ディスクのボリューム番号としても g), h), i)の関係は成り立つ.

3.2.3 分散型データベースのファイル位置

磁気ディスク用辞書のファイル名は，

dic.(集合名)

とし，光ディスクのディレクトリファイル名は，

表 3 データファイル管理表
Table 3 The management table of data file.

```
###   dbtable   ###
odb=apl:/users/speech/kdb
mdb=west:/usr4/bigtmp3/speech

274_WORDS_SET.274
optical_disk.1
opt1a.2   w24.80.15 odb/dir.w24.274
          bpf.a0.15 odb/dir.bpf.274
magnetic_disk.0

212_WORDS_SET.212
optical_disk.2
opt1b.2   w24:10.21 odb/dir.w24.212
          bpf:50.1d odb/dir.bpf.212
opt5a.2   w10:20.9  odb/dir.w10.212
          atc:29.9  odb/dir.atc.212
magnetic_disk.3
   mdb/w10.212.100*2
   mdb/bpf.212.29*4.1000
   mdb/fps.212.8*4

3000_WORDS_SET.3ws
optical_disk.6
opt2a.1   w24:10.28 odb/dir.w24.3ws.1
opt2b.1   w24:10.27 odb/dir.w24.3ws.2
opt3a.1   w24:10.27 odb/dir.w24.3ws.3
opt3b.1   w24:10.27 odb/dir.w24.3ws.4
opt4a.1   w24:40.26 odb/dir.w24.3ws.5
opt4b.1   w24:10.20 odb/dir.w24.3ws.6
magnetic_disk.0

DICTATION_SYSTEM_DATA_SET.dct
optical_disk.1
opt5b.0
magnetic_disk.4
   mdb/w24.dct.240*2
   mdb/w10.dct.100*2
   mdb/bpf.dct.29*4.1000
   mdb/fps.dct.8*4
```

dir.(分析名).(集合名)
とする．例えば212単語集合の磁気ディスク用辞書はdic. 212であり，3000単語集合(3 ws)の光ディスクの2面目の24 kHz波形(w 24)の辞書はdir. w 24.3 ws. 3である．

分析データのファイル名は
(分析名).(集合名)
とする．

K-DBは分散型データベースであり，ネットワークでの光ディスクディレクトリ，辞書と分析ファイルの位置を示す表が必要であり，これをdbtableとする．

現在のdbtableを**表3**に示す．dbtableの様式は以下のとおりである．

(集合),(集合名)
optical_disk,(光ディスクのボリューム数)
ボリューム名，分析数
分析名 1：IB, NB host_name：(辞書ファイル名)
…
magnetic_disk,(分析ファイル数)
(分析ファイル名) (レコード長)*(ワード長),
(倍率)
…

分析ファイルの1レコードは1フレーム＝10 msecに相当するデータである．データは計算機による浮動小数点実数の形式の不統一をさけるためすべて整数化した．この際精度が不足する場合は実数を定数倍した後整数化し，このときの定数（倍率）を記している．

4. 入力音声からの音声情報の自動獲得

4.1 ピッチとホルマントの高精度自動抽出

K-DBでは音声の音韻性，個人性による変動を主にピッチとホルマントで把握する．音声データベースでは入力音声の音素系列や発声者の性別などが既知であり，また実時間性を必ずしも要求しない．そこで，ピッチは，入力音声の連続性を考慮することによって高精度に抽出する[5]．またホルマントは，入力音声のピッチから予測される話者の性別情報あるいは入力音声の音素系列情報を用いて，ホルマント周波数の出現領域を予測し，これらの制約を組み込んだ動的計画法に基づく手法を用いて，高精度に抽出する[5]．5母音の認識率で評価した場合，従来の分析法に比べ，ピッチ情報から予測したホルマントの出現領域の拘束を用いると，男声で4.1%，女声で13.6%，認識率が改善された[5]．さらに入力音声の音素系列情報で予測したホルマントの出現領域の予測を用いると，従来の分析法に比べ，男声で7.3%，女声で18.4% 認識率が向上した[5]．**表4**に本システムで用意している音声分析法を示す．音声分析で用いられてきた従来の分析法をほとんど網羅している．この表でbsc (basical analysis)は対数パワー，零交差数などの基本的な分析である．K-DBでは3.2.1項での関係c)での分析条件でのファイルを作成して，分析種類，標本化周波数，窓長，窓関数，フレーム速度，線形予測次数などを記録する．また分析では，分析種類，標本化周波数，窓長，窓関数の種類，フレーム速度，線形予測次数などを変えることができる．

表4 音声分析タイプ
Table 4 Type of speech analysis.

NAME	ANALYSIS
w24	24 kHz sampling waveform
w20	20 kHz sampling waveform
w10	10 kHz sampling waveform
bpf	band pass filter data
atc	auto-correlation
lpc	linear prediction coefficient
fpr	formant by PARCOR
fab	formant by A-b-S
fps	formant by PSLPA
bsc	basical analysis
cep	cepstrum
cpm	mel cepstrum
cpl	LPC cepstrum
clm	LPC mel cepstrum
lbl	phoneme labeling data

(PSLPA: pitch synchronized linear prediction analysis)

4.2 音素の自動ラベリングシステム

音声合成や音声認識の研究では，音声データの量とともに，音素区間情報がきわめて重要であり，これらの情報を持っているか否かが，音声データベースの価値を決める．

本研究では，前述のホルマントとピッチの抽出システムで得たホルマントを主特徴量とし，ピッチや零交差数などの特徴を副特徴量とした自動ラベリングシステムを作成した．このシステムでは，入力音声の音素系列から単語のホルマントパタンを予測し，入力音声から得たホルマントパタンと比較することによって，入力音声を入力音素系列に対応した区間に分割することができる．男声10人，女声10人の発声した212単語で評価したところ，2フレーム（1フレームは10 msec）以内の誤差で97.3%の正当率を得た．従

来の自動ラベリングシステム[9)-11)]の結果である2フレーム以内で 92～93% に比較して良好な結果を得た．これは標準パタンが前後の音素環境を考慮していることと，ラベリングを安定な部分から変化の激しい部分へと階層的に進めているためと考えられる．このシステムは，データベースにおける音素区間データの作成に用いられている[6)]．なお音素は以下の 30 種を設定し，外来者は /f/ のみ用いているため厳密な音素表記ではない．

母音	/a/, /o/, /u/, /i/, /e/
長母音	/aL/, /oL/, /uL/, /iL/, /eL/
無声化母音	/u−/, /i−/
半母音	/j/, /w/
鼻音	/m/, /n/, /ŋ/
流音	/r/
有声破裂音	/b/, /d/, /g/
無声破裂音	/p/, /t/, /k/
有声摩擦音	/z/
無声摩擦音	/s/, /h/
破擦音	/c/, /dʒ/
外来音	/f/

図5に男声 /nukiuci/ のラベリング例を示す．下部のラベルは上段が視察による結果，下段が自動ラベリングの結果である．視察による補正が必要なのは，図5のように，/i/ や /u/ と鼻音の境界である．有声音と無声音の境界については自動ラベリングで十分であり，音声研究者による音素ラベリングの作業を大幅に効率化できる．自動ラベリング結果のうち間違う可能性のある部分は視察によってチェックを行っている．

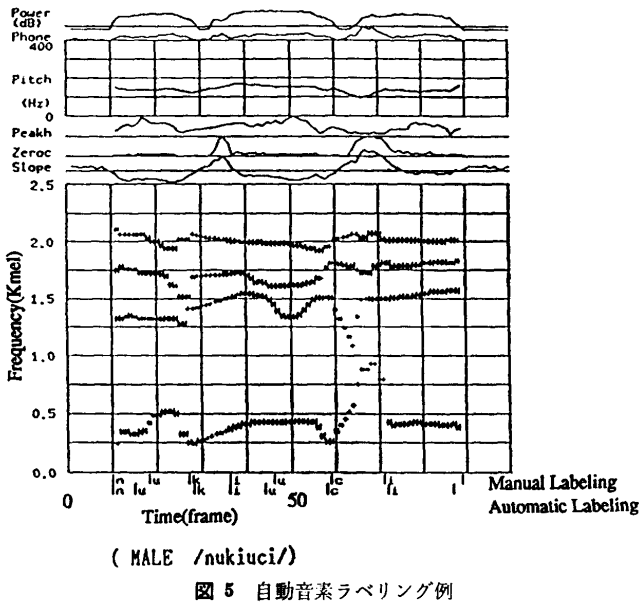


図5 自動音素ラベリング例
Fig. 5 An example of automatic phoneme labeling.

5. 音声データ編集，分析，表示システム EDB

5.1 EDB の概要

K-DB の目的の1つは音声の特徴の把握である．そのためには，使用者が，データベースの構造についての知識がなくとも，データにアクセスして，簡単に分析できることが望まれる．そこで原波形の D/A 変換による確認や，調音結合による音韻性の動的変化の分析のために，音素ラベリング，音声認識，音声合成などのシステムを呼び出して利用できるデータベースエディタ EDB を作成した．EDB は基本的に編集，分析，表示の機能を持ち，ラベリング，認識，合成と併せて6つの機能を利用できる．

編集機能は，A/D, D/A 変換を用いたデータの作成と確認，それに伴う辞書の更新，そして光ディスク，磁気ディスク，磁気テープ間のファイルの転送の3つである．分析機能は分析ライブラリによる音声データの分析と，分析ファイルの作成である．これにはラベリング，認識，合成のための標準パタン作成を含む．表示機能は分析ファイルを読み込んで，必要なデータを検索して表示することである．

5.2 EDB の表示機能

EDB での分析データ表示のための構文は以下の様式で指定する．

MODE -sp SP -wd WD -gr GN -nt NT

MODE：検索形式，SP：話者番号，WD：単語番号

GN：グラフィック番号，NT：表示回数

検索は単語単位と音素単位で行うことができ，音素単位では調音結合を見るために前後の音素を考慮した三連接まで指定できる．また音素を弁別素性により大分類しそれを表す記号を導入している．

話者番号と単語番号は辞書へのサーチポインタの初期値を与える．回数は同じ環境のデータ（例えば前後が無声音の環境にある /i/）を重ね書きするとき用いる．実際に重ね書きして意味があるのはホルマントであり，ホルマントが前後の音素環境によってどのように変化するかを見ることにより，調音結合について考察する

```
****** MENU ******   E : end   M : menu   Z : clear
edb> Search_Mode -sp Speaker -wd Word -gr Graphic -nt NT
Search_Mode (ex. kVC-)
 W : WORD
 # : NON PHONEME    A : ANY PHONEME
 V : VOWEL         C : CONSONANT    N : NAZAL  L : LIQUID
 P : PLOSIVE       H : PLOSIVE FRICATIVE       F : FRICATIVE
   + : voiced      - : unvoiced     : : long
Graphic (ex. 03)
 0 : Formants      1 : Digital Sonagraph
 2 : Digital Sonagraph with High Band Emphasis
 3 : Local Peaks   4 : Local Peaks over LSFL   5 : over LSFL
NT (Repeat Number for Overwrite)
```

図 6 EDB メニュー画面
Fig. 6 Menu screen of EDB.

ことができる.

すべての要素はデフォルト値を持ち，何も指定しない場合は単語単位でデータベースの先頭から表示する.

例えば話者番号 958 で，語頭で後続が無声破裂音である /a/ のホルマント軌跡を 10 パタン表示するには，以下のように指定する.

EDB>#aP− −sp 958 −gr 0 −nt 10

このシステムのメニュー画面を**図 6** に示す．検索様式の説明が上部にあり，単語単位の検索の際は "W" を指定し，音素単位の検索は音素記号を 3 つまで並べて連接条件を指定し，/a/ や /m/ などの音素記号と同様に任意の破裂音 "P" などの記号を用いる．例えば母音は "V" で示し，さらに無声化母音は "V−"，長母音は "V:"，有声破裂音は "P+" などの指定ができる．下部にグラフィック画面の選択肢を示す，0 がホルマント軌跡，3 がスペクトルのローカルピークを単独で表示するオプションであり，03 と指定するとホルマントとローカルピークを重ね書きする.

図 7，**図 8** は EDB を用いた表示例であり，図の下部に EDB での指定を示している．図 8 中の数字は，29 ch BPF から抽出したローカルピークの，低周波数側から数えた順序である.

EDB は編集と分析機能のほかに，必要であれば音素ラベリング，音声認識，音声合成のプログラムを呼び出して利用することができる.

EDB は，約 5 千行の RATFOR プログラム（部分的に C 言語）として作成し，各計算機で実動している.

K-DB は，音声データとデータラベルを用意しさえすれば容易にデータの追加，修正が行えるシステムである．分析，認識，合成についても実験条件や実験結果を記憶するファイルの形式を決めてあり，これらを記したマニュアルを作成し，利用を容易にしている.

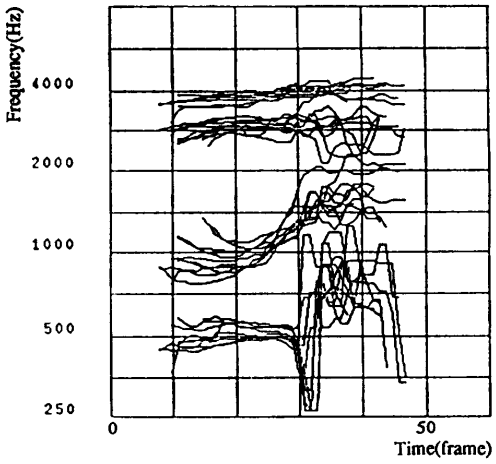


図 7 EDB 分析例（その 1）
Fig. 7 Examples of analysis using EDB (Part 1).

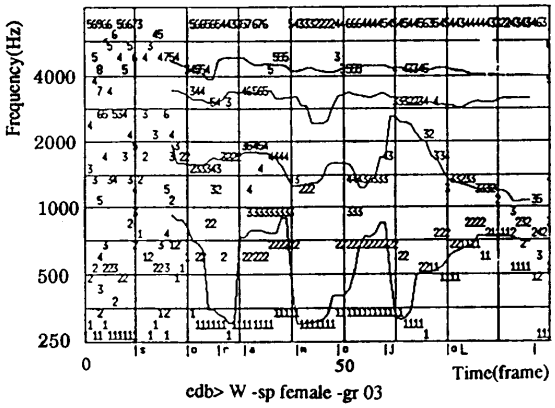


図 8 EDB 分析例（その 2）
Fig. 8 Examples of analysis using EDB (Part 2).

またシステム全体のマニュアルも整備しており，初心者にも利用が容易となっている.

6. むすび

分散型大規模音声データベース K-DB について述べた．従来の音声データベースは，小規模でかつ単一計算機や単一ディスクを想定したものであり，また音素ラベリング，ホルマントやピッチの分析などでかなりの人手を必要とするため，拡張性に乏しいものであった.

そこで本論文では，安価で大容量の補助記憶媒体である光ディスクを採用し，12 Gbyte の音声データを 6 枚の光ディスクに格納して音声データベースを構築した．これにより以前に比べて大量の音声データを容易かつ統一的に用いることができるようになった．またデータベース辞書の構成は関係型とした.

さらに，従来人手に頼っていたため，音声データベースの大規模化の最大の問題点であった，音素ラベリングやホルマント，ピッチなどに代表される音声情報を自動的に獲得するための機能を，本データベースでは備えている．

また調音結合による音声の変動をみるため，音声データ編集，表示システム EDB を作成した．前後の音素環境を指定した表示ができるため，音声の特徴量の静的，動的特徴についての理解が得られやすくなった．

以上の結果，音声研究の補助として，ハード的に光ディスク，ソフト的に関係データベースを導入した分散型大規模音声データベースを完成した．大量のデータに裏打ちされた音声認識システムや特徴量がこの中から現れることを期待して，今後もデータベースの拡張と改良を行っていきたい．

謝辞 GP-IB プログラミングの指導をしていただいた元山形大学工学部，故青木伴至氏に感謝します．

参考文献

1) 田中，速見，太田：音声の音素片ネットワーク表現と時系列のセグメント化法を用いた自動ラベリング手法，日本音響学会誌，Vol. 42, No. 11, pp. 860-868 (1986).
2) 武田，匂坂，片桐，桑原：研究用日本語音声データベースの構築，日本音響学会誌，Vol. 44, No. 10, pp. 747-754 (1988).
3) 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 (1982).
4) 城風，牧野，城戸：光ディスク上の大規模音声データベースを中核とした GP-IB ネットワーク，音声研資料，SP 87042 (1987).
5) 城風，牧野，城戸：種々の知識を用いたピッチとホルマントの抽出システム，電気音響研究会資料，EA 88-46 (1988).
6) 城風，牧野，城戸：ホルマントを主特徴量とした音素ラベリング，音響学会講演論文集，2-p-14 (1989).
7) 東芝(株)：光ディスク装置 DF-0450 インターフェイス使用書 (1984).
8) 植村：データベースシステムの基礎，オーム社 (1979).
9) 田中，速見，太田：既知音声の半自動ラベリングと疑似音素系列の生成，日本音響学会研究会資料，S 84-28 (1986. 6).
10) 相川，杉山，鹿野：音声の自動ラベリング，日本音響学会講演論文集，2-1-3 (1983. 3)
11) 辻堂，板橋，西野：動的計画法を利用した音声のセグメンテーション，日本音響学会講演論文集，2-2-9 (1983. 3).

(平成2年 4 月 9 日受付)
(平成2年 10 月 9 日採録)

城風 敏彦（正会員）

昭和 58 年東北大学工学部応用物理卒業．平成 2 年同大大学院情報工学専攻博士課程修了．工学博士．現在，沖電気工業(株)．音声信号処理，ヒューマン・インタフェースの研究に従事．電子情報通信学会，日本音響学会各会員．

牧野 正三（正会員）

昭和 44 年東北大学工学部電子卒業．昭和 49 年同大大学院博士課程修了．同年同大電気通信研究所助手．昭和 56 年同大応用情報学研究センター助手．現在同所助教授．工学博士．昭和 59～61 年アメリカ合衆国 STL 客員研究員．言語情報を利用した音声認識の研究，音声データベース，音声信号処理，音響信号処理，画像信号処理，文字認識の研究に従事．電子情報通信学会，日本音響学会各会員．

城戸 健一（正会員）

昭和 23 年東北大学工学部電気卒業．昭和 38 年同大電気通信研究所教授，昭和 51 年同大応用情報学研究センター教授，平成 2 年 4 月千葉工業大学情報工学科教授，現在に至る．音響機器，建築音響，騒音制御，心理音響の研究から始まり，現在は音声自動認識，ディジタル信号処理，特にその音響工学への応用に関する研究に従事．著書「音響工学」(電子通信学会編，コロナ社)，「ディジタル信号処理入門」(丸善)，「電子計算機概論上・下」(丸善)，「過渡現象論」(朝倉書店) 等．工学博士．日本音響学会，電子情報通信学会，電気学会，計測自動制御学会，韓国音響学会，IEEE, AES など各会員．アメリカ音響学会フェロー．