

TR-I-0305

HMMとエキスパートシステムの手法を用いた  
音素セグメンテーションシステムと  
ラベリングワークベンチ

藤原 紳吾  
S. Fujiwara

杉山 雅英  
M. Sugiyama

1993.3

音声情報処理システムの研究開発に必須である音声データベースの効率的作成・保守のための音素セグメンテーションシステム、および音素ラベリングワークベンチの研究開発の成果を報告する。開発したワークベンチはユーザーインタフェース・モジュールと音素セグメンテーション・エンジンから構成される。ユーザ・インタフェース部は X Window System を利用しており、マウス操作により簡単にラベリングを実行できる。音素セグメンテーション実行部は HMM とエキスパートシステムの手法を用いており、その性能は、音素境界誤差の平均が 12 ms (単語データ)、14 ms (文データ) であった。また、ワークベンチを使用して作成したラベルデータの境界誤差は平均約 6 ms、標準偏差約 10 ms であり、従来の視察ラベリング作業で得られるデータとほぼ同じ精度であった。本報告では、ワークベンチの構成・操作性・性能について述べ、さらに、我々が実際に構成したシステムについて説明し、そのパフォーマンスと作成したラベルデータの精度について報告する。

# 目次

1	はじめに	1
2	音素ラベリングワークベンチのシステム概要	1
3	ユーザーインタフェース・モジュール	2
3.1	音素セグメンテーション・エンジンとの通信機能	2
3.2	音響パラメータの表示機能	3
3.3	ワークベンチの表示画面	3
3.4	階層音素ラベルの入力機能	6
3.5	カスタマイズ機能	8
3.6	ヒストリー機能	9
4	音素セグメンテーション・エンジン	9
4.1	構成	9
4.2	セグメンテーション性能評価	10
5	システム構成の実例	11
5.1	システム構成例	11
5.2	ワークベンチを用いたラベリング・テスト	12
6	おわりに	12
	参考文献	13
	付録 A 初期設定ファイルのキーワード	14
	付録 B 初期設定ファイルの例	16
	付録 C 対外研究発表 抄録	18
	C.1 Fujiwara91ASJ3	19
	C.2 Fujiwara91KOREA7	21
	C.3 Fujiwara91ASJ10	27
	C.4 Iwahashi91ASJ10	29
	C.5 Fujiwara92ISSPA92	31
	C.6 Fujiwara92ICSLP	35
	C.7 Fujiwara92ASJ10	39
	C.8 Fujiwara93IEICE1	41
	C.9 Fujiwara93ASJ3	49

## 1 はじめに

最近の音声認識・合成システムなど、いわゆる音声情報処理システムの研究開発には、整備された大規模な音声データベースが必要とされている [1,2]。特に、音素の位置情報を表す音素ラベルデータは重要な役割を果たしている。例えば、認識システムでは音素モデル (HMM, neural network 等) のパラメータの推定 (学習) に用いられ、合成システムでは合成用音素片データの作成に利用されており、多数の研究成果が報告されている。

音素ラベルデータは、ラベリングの専門家により手作業 (スペクトログラムの視察によるラベリング) で作成されている。このため、作業には膨大な時間を費し、また複数の作業員間でのデータ統一性を保つために多大の努力が必要となっている。従来の音素ラベリング作業におけるこのような問題点を克服するため、種々の自動ラベリング手法が提案されている [4-10]。また、音声処理ワークベンチも実際に構成されている [11,12]。しかし、操作性の良いユーザーインタフェース機能と自動ラベリング機能を備えた実用的システムは、ほとんどないのが現状である。

音素ラベルデータの効率的作成・保守のために、開発した音素ラベリングワークベンチは、ユーザーインタフェース・モジュールと音素セグメンテーション・エンジンから構成されている [16,17]。音素セグメンテーション実行部はHMMを用いたセグメンテーションと、エキスパートシステムの手法、すなわちスペクトログラムリーディング知識に基づくルール処理から構成され、2つの処理により高精度な音素セグメンテーションを実現している。また、ユーザ・インタフェース部は *X Window System* を利用しており、マウスあるいはキーボード操作により簡単にラベリング作業を行うことができる。

本報告では、音素ラベリングワークベンチのシステム構成、ユーザーインタフェース部の操作性、および音素セグメンテーション実行部の構成・性能について述べる。また、システム構成の実例として我々が実現しているシステムについて説明し、そのパフォーマンスと作成したラベルデータの精度について報告する。なお、音素セグメンテーションシステムの詳細に関しては、対外研究発表における予稿を付録として収録する。

## 2 音素ラベリングワークベンチのシステム概要

音素ラベリングワークベンチは音声データベースにおける音素ラベルデータの効率的作成・保守のための支援ツールであり、エンジニアリング・ワークステーション (EWS) 上で稼働する。図1にワークベンチのシステム構成を示す。システム全体はクライアント/サーバー形式

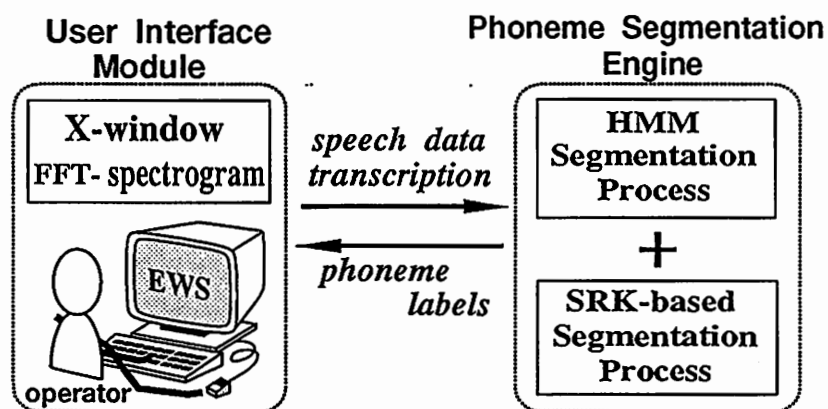


図1: 音素ラベリングワークベンチ

となっており、ユーザーインタフェース・モジュール(クライアント)と音素セグメンテーション・エンジン(セグメンテーション・サーバー)から構成される [19,20]。音素セグメンテーション部をサーバーとして分離することで、1サーバー対マルチクライアントでのシステム構成が可能になり、WSにおける処理効率の向上が期待される。

音素セグメンテーション・エンジンは、音声データと発声内容を表すローマ字列を入力として自動音素セグメンテーションを行い、得られた音素ラベル情報をユーザーインタフェース・モジュールに返す。ユーザーインタフェース・モジュールには、波形・スペクトログラムなどのラベリング作業に必要な音響パラメータと音素セグメンテーション・エンジンで得られた音素ラベル情報(自動ラベルと呼ぶ)が表示される。

ラベリング作業者は、音響パラメータ、自動ラベルを観察しながら、マウスあるいはキーボードを使用して対話形式でラベリングを行うことができる。作業により作成(あるいは修正)された音素ラベルはラベルデータファイルとしてセーブされる。

### 3 ユーザーインタフェース・モジュール

ユーザーインタフェース・モジュールはXウィンドウシステムを使用して構成され、EWS、オペレーティング・システム(OS)あるいは動作環境(window managerなど)への依存性をできるだけ排除するためXlibのみを用いてプログラミング言語Cで記述されている。ユーザーインタフェース・モジュールは、

- 音素セグメンテーション・エンジンとの通信機能
- マルチプロセスによるスペクトログラム計算機能
- 正確な音響パラメータの表示機能
- 階層構造ラベルの入力機能(マウス、キーボード)
- カスタマイズ機能
- ヒストリー機能
- Portability — X11R4, R5 に対応。Xlibのみ使用

を持つように設計されている。

#### 3.1 音素セグメンテーション・エンジンとの通信機能

ユーザーインタフェース・モジュールでは、socket, bind, connect, selectなどのUNIX System Callを用いて、TCP/IPに基づくプロセス間通信を行い、音素セグメンテーション・エンジンとの通信機能を実現している [20]。通信時は2つの通信経路を設け、一方を波形データとローマ字表記の送信用、他方を音素ラベル情報の受信用としている。音素セグメンテーション・エンジンによる音素ラベル情報は、ウィンドウに表示可能である。

また、自動ラベリング部はバッチ処理が可能であり、あらかじめ音素セグメンテーション・エンジンにより音素ラベルファイルを作成しておき、そのラベルファイルから音素ラベル情報を読み込むこともできる。すなわち、自動ラベリングだけを夜間に実行しておき、作業時はその結果のみ参照すればよく、作業時間が短縮される。

## 3.2 音響パラメータの表示機能

従来のラベリング作業では、レーザービーム・プリンタにより紙面に印刷されたスペクトログラムなどの音響パラメータを観察して音素ラベリングを行っている [2]。したがって、印刷されたパラメータとほぼ同等の音響パラメータをワークベンチの画面上に表示できるように、パラメータ表示ルーチンが設計されている。また、スペクトログラムの計算は子プロセスで実行される。各種音響パラメータの分析方法は従来のラベリング作業時のものと同等にしてあり、音響パラメータとしては、

- 全波形
- 波形 (部分波形)
- ログパワー
- スペクトル平均変化率
- スペクトログラム
- スペクトログラム (濃淡表示)

が表示可能である。特に、スペクトログラム表示では、高精度スペクトログラム表示ルーチンを利用している [14]。上記パラメータ以外にも、

- 階層構造の音素ラベル (第1～5層)
- 自動ラベリングによる音素ラベル情報
- 時間軸
- コマンドパネル

が表示可能である。

## 3.3 ワークベンチの表示画面

ワークベンチの表示画面の例を図2に示す。メインウィンドウには複数のサブウィンドウが設けられ、コマンドパネルのほかに、音響パラメータ、音素ラベル、時間軸がページ単位で表示される。また、マウスマウスカーソルの位置にはサブウィンドウをまたがって縦線 (ガイドライン) が表示される。カスタマイズ機能 (後述) を利用して、ラベリング操作に必要なサブウィンドウのみを表示することができ、図2の例では、表示ページは0～1200 msであり、コマンドパネル、全波形、波形、ログパワー、スペクトル平均変化率、自動ラベル、スペクトログラム、時間軸、5層の音素ラベルが表示されている。

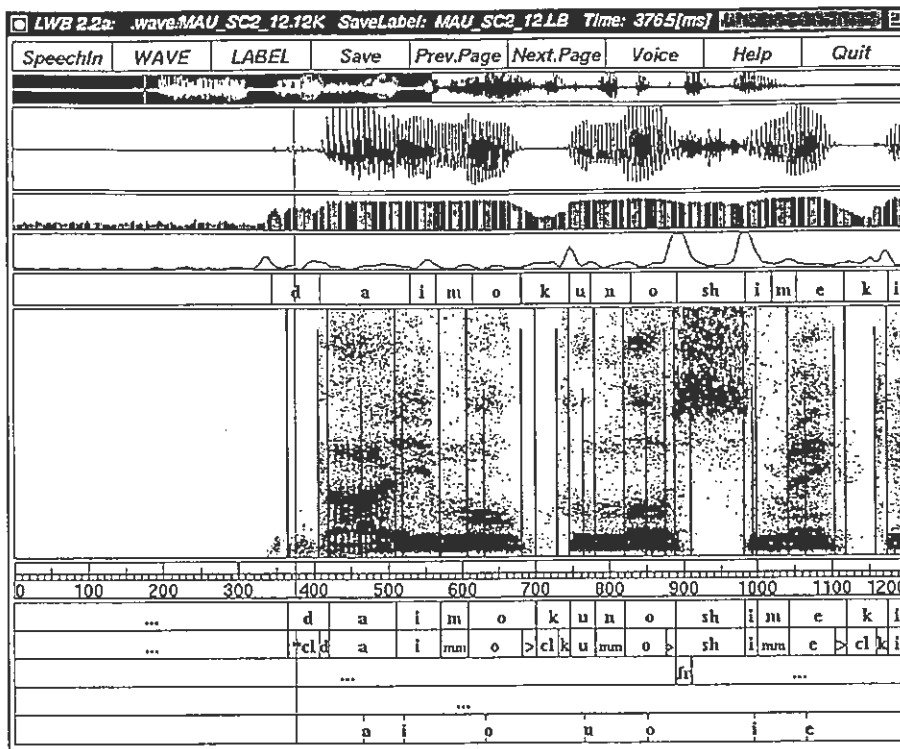


図 2: ラベリングワークベンチの操作ウィンドウ

ワークベンチの画面に表示可能な項目(サブウィンドウ)について、次に説明する。

### 全波形

波形全体を表示する。図 3 に示すように、現在の表示ページにあたる波形部分が反転表示される。



図 3: 全波形・サブウィンドウ

### 波形

波形を表示する。このサブウィンドウを使用して音声出力ができる。次の 3 つのいずれかの方法により、マウスの左ボタンで領域を指定した後、

- (1) 開始点、終点を指定
- (2) ダブルクリックで現ページを指定
- (3) 何も指定しない(→ 波形全体を指定)

マウスの中ボタンを押す(あるいは、音声出力コマンドを実行する)と指定した領域が音声出力される。また、マウスの右ボタンはリセット用に使用される。

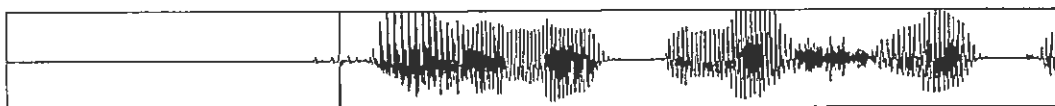


図 4: 波形・サブウィンドウ

### ログパワー

音声のログパワーを表示する。



図 5: ログパワー・サブウィンドウ

### スペクトル平均変化率

スペクトルの平均変化率を表示する。平均変化率として、 $\Delta$ ケプストラムのノルムに相当するパラメータを用いている [13]。LPCケプストラムを使用している。



図 6: スペクトル平均変化率・サブウィンドウ

### スペクトログラム

短時間FFT分析で求めたスペクトログラムをビットマップ表示する [14]。スペクトログラム強度は、ランダムドットの密度に変換される。このサブウィンドウを使用して音素ラベル境界を入力・削除・移動できる。

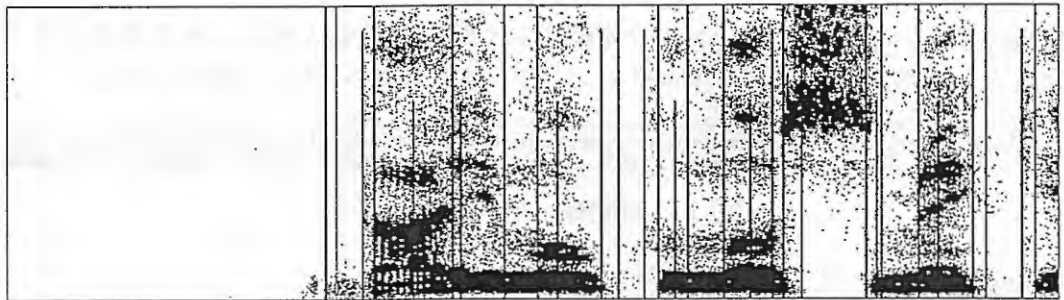


図 7: スペクトログラム・サブウィンドウ

### スペクトログラム (濃淡)

短時間FFT分析で求めたスペクトログラムを濃淡表示する。スペクトログラム強度は、濃淡 (グレイスケール) に変換される。このサブウィンドウを使用して音素ラベル境界を入力・削除・移動できる。

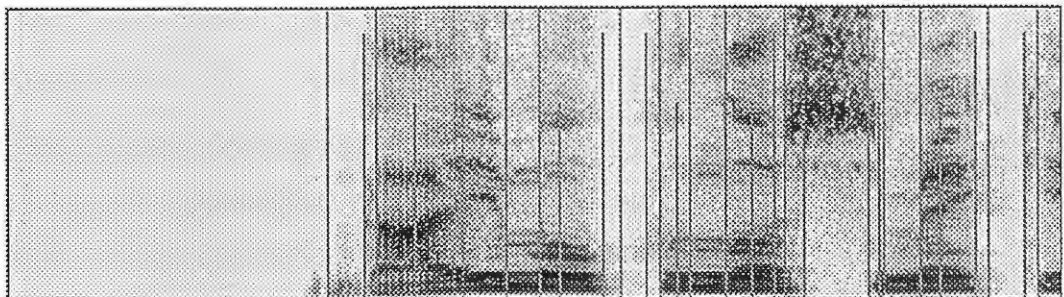


図 8: 濃淡スペクトログラム・サブウィンドウ

### ラベル

作成(あるいは修正)中の音素ラベルデータを表示する。この際、ラベル記号は読みやすいように大きさが自動スケールされる。このサブウィンドウを使用して、音素ラベルシンボルを入力・修正できる。

...	d	a	i	m	o	k	u	n	o	sh	i	m	e	k	i			
...	>cl	a	i	mm	o	>	cl	u	mm	o	>	sh	i	mm	e	>	cl	i
	fr																	
	...																	
	a	i	o	u	o	i	e											

図 9: ラベル・サブウィンドウ

### 自動ラベル

音素セグメンテーション・エンジンによる音素ラベル情報を表示する。

	d	a	i	m	o	k	u	n	o	sh	i	m	e	k	i
--	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---

図 10: 自動ラベル・サブウィンドウ

### 時間軸

時間軸を表示する。このサブウィンドウを使用して時刻を読みとれる。マウスの左ボタンをクリックすると、その点の時刻がメインウィンドウのタイトル・バーに表示される。

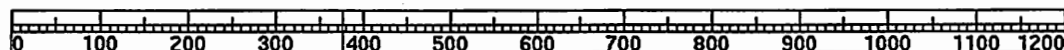


図 11: 時間軸・サブウィンドウ

### コマンドパネル

コマンド実行用メニューを表示する。メニュー項目をクリックすることで、波形データロード・セーブ、ラベルデータロード・セーブ、次ページへのスクロール、前ページへのスクロール、音声入力・出力、終了などのコマンドを実行できる。

SpeechIn	WAVE	LABEL	Save	Prev.Page	Next.Page	Voice	Help	Quit
----------	------	-------	------	-----------	-----------	-------	------	------

図 12: コマンドパネル・サブウィンドウ

## 3.4 階層音素ラベルの入力機能

階層音素ラベルは、第1～5層ラベルまで入力でき、それぞれ次のように用いられる [2]。

- 第1層(音声記号層) — 発声のローマ字表記を対応する音声区間に記述する。
- 第2層(イベント層) — 音声現象を出来るだけ詳細に記述するために、各種の記号を付与する。
- 第3層(異音化層) — 異音化現象を記述する。
- 第4層(融合化層) — スペクトログラム上で区分化不可能な複数の音声事象を記述する。



- 第5層(母音中心層) — 母音区間で最も安定した箇所を明示する。

各層のラベルは、スペクトログラム(または濃淡スペクトログラム)・サブウィンドウ上で音素境界線、ラベル・サブウィンドウ上で音素ラベル記号を入力することで作成される。音素ラベルの入力手順を以下に示す(図13参照)。

1. ラベル層の選択 — 入力したいラベル層に対応するラベル・サブウィンドウをマウスの右ボタンでクリックする。選択されたサブウィンドウは背景色が変わる。
2. 音素境界線の入力(挿入) — スペクトログラム・サブウィンドウ上で音素境界を設定したい場所で、マウスの左ボタンをクリックすると境界線が描かれる。選択しているラベル・サブウィンドウにも音素区間の境界線が描かれる。
3. 音素境界線の移動 — スペクトログラム・サブウィンドウ上に描かれている音素境界線をマウスの左ボタンでクリックし(境界線の色が変わる)、移動先でもう一度マウスの左ボタンをクリックする。
4. 音素境界線の削除 — スペクトログラム・サブウィンドウ上に描かれている音素境界線をマウスの左ボタンでダブル・クリックする。
5. 音素ラベル記号の入力(ポップアップメニュー使用) — 選択しているラベル・サブウィンドウ上の音素区間をダブルクリックすると、ラベル記号入力用のポップアップ・メニューが現れる。ポップアップ・メニュー上の記号をクリックすると、ラベル記号として入力される。
6. 音素ラベル記号の入力(キーボード使用) — 選択しているラベル・サブウィンドウ上の音素区間をクリックした後、キー入力する。リターンキーで確定すると入力対象の音素区間は右に移動する。入力対象の区間は矢印キー(←→↑↓)でも移動できる。

音素ラベル記号を入力したあとでも、新たな音素境界線の入力、既存の音素境界線の移動・削除が可能である。

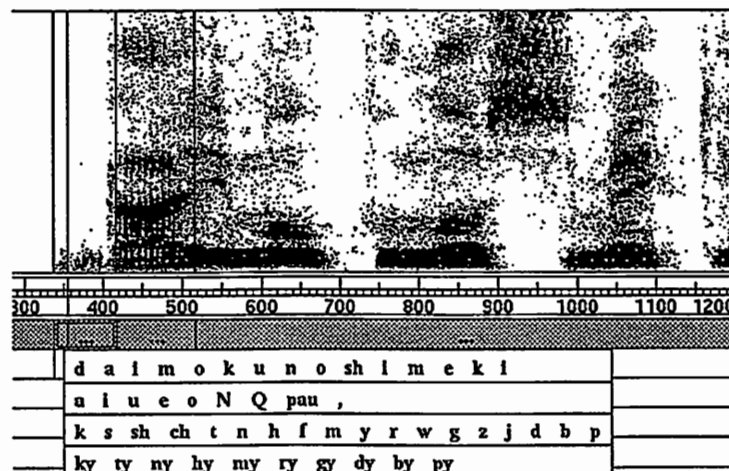


図13: ポップアップメニューによる音素ラベル記号の入力

### 3.5 カスタマイズ機能

ワークベンチにおいて、作業に必要な音響パラメータの選択、任意の音響分析条件への対応などを可能にするため、カスタマイズ機能を持つ。動作環境の設定ファイル(.lwbrc)を記述することで、作業者ごとにワークベンチのカスタマイズが可能である。設定ファイルでは、

- メインウィンドウの位置・大きさ
- 表示サブウィンドウとその位置・大きさ
- 音響パラメータの分析条件
- 1 ページの時間幅とページ間隔
- ラベル入力メニューで用いるラベル記号
- 文字フォント、背景色

などを指定できる。表4から表10に、設定ファイルにおけるキーワードと設定内容を示す。

#### サブウィンドウの配置

サブウィンドウは、設定ファイルで記述された順番にメインウィンドウの上方から配置される。この際、サブウィンドウ間には *WindowSeparate* で指定したピクセル値分のスペースが取られる。また、各サブウィンドウの大きさはサブウィンドウ名(表5)に与える数値で決められる。ただし、数値が1未満の時は  $\langle \text{MainWindowHeight} \times \text{数値} \rangle$  の値が大きさとなる。例えば、

MainWindowWidth:	1000
MainWindowHeight:	700
WaveWindow:	100
PowerWindow:	0.05
TimeScaleWindow:	40
SpecWindow:	0.5
Label1Window:	50
Label2Window:	50
WindowSeparate:	5

と記述されていれば、

高さ 100 の波形サブウィンドウ、  
 高さ 35 のログパワーサブウィンドウ、  
 高さ 40 の時間軸サブウィンドウ、  
 高さ 350 のスペクトログラムサブウィンドウ、  
 高さ 50 の第1層ラベルサブウィンドウ、  
 高さ 50 の第2層ラベルサブウィンドウ

がメインウィンドウ (1000×700) 内に上から5ピクセル間隔で配置される。

## ラベル記号入力メニュー

設定ファイルで、

```
LabelMenu1: a i u e o \\ N \\ k s t n h m y r w.
LabelMenu2: < > cl *cl tr dj mm \\ a i u e o
LabelMenu3: fr dv
```

と記述すれば (\\ はメニューにおける改行を指示する)、第1層、第2層、第3層のラベル記号入力時のポップアップメニューは、それぞれ次のようになる (第4、5層も同様)。

a i u e o
N
k s t n h m y r w

< > cl *cl tr dj mm
a i u e o

fr dv
-------

## 3.6 ヒストリー機能

ラベリング操作履歴の保存、異常終了時の入力ラベルデータの復旧、あるいは、ログファイルからの自動デモンストレーションを行うためヒストリー機能を持つ。

### 操作履歴の保存

ワークベンチ起動時に、コマンド・オプション -e を指定すると、マウスカーソルの移動・ボタンクリック、キーボード入力など、すべての操作の履歴がログファイルに保存される。

### ログファイルからの起動

ワークベンチ起動時に、コマンド・オプション -E を指定すると、ログファイルの履歴を読み込んで、ワークベンチが操作される。

## 4 音素セグメンテーション・エンジン

ワークベンチ・システムにおいて、自動ラベリング・サーバーとして機能する音素セグメンテーション・エンジンの構成について説明する。さらに、単語音声・連続音声データに対する音素セグメンテーション性能を示す。

### 4.1 構成

音素セグメンテーション・エンジンの構成を図14に示す。HMMによる音素セグメンテーション (HMM-process) とスペクトログラムリーディング知識に基づくルール処理 (SRK-process) により音素セグメンテーションが実行される [15,16]。また、セグメンテーション・エンジンはサーバーとして動作し、クライアント (ユーザーインタフェース・モジュール) から音声データとローマ字列を受け取り、セグメンテーション結果をクライアントに返す。

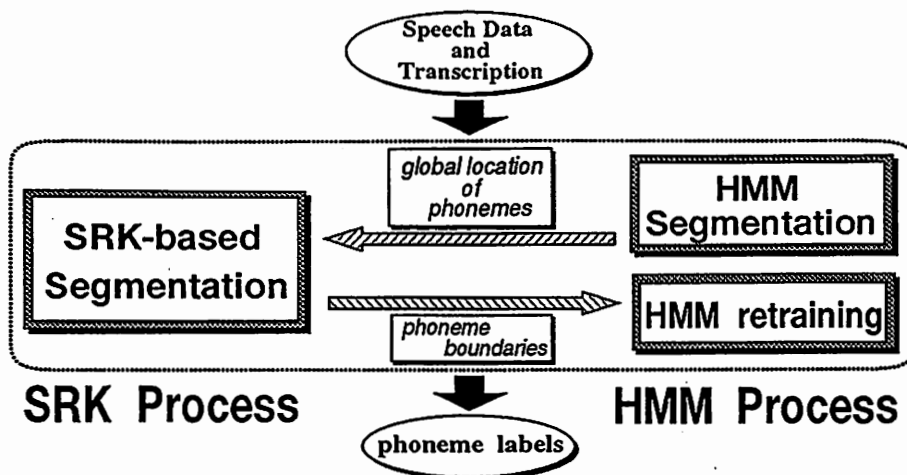


図 14: 音素セグメンテーション・エンジンの構成

## 4.2 セグメンテーション性能評価

単語音声データ [2] および連続音声データ [3] を用いて実験を行い、音素セグメンテーション・エンジンのセグメンテーション性能を評価した。

単語音声データ(3名の男性話者、各 5240 語)を用いて、音素セグメンテーション実験を行った [15,16]。HMMセグメンテーション部には、25 個の音素モデルを用い、ラベルなし連結学習で各話者の音素HMMを学習し、これを初期モデルとした。また、1名の男性話者(話者MAU)の単語音声データを参照して、約 300 のセグメンテーション・ルールを記述し、ルール処理部で使用した。

初期モデルを用いた音素セグメンテーション実験の結果、および再学習したモデルを用いて再度セグメンテーションした結果を表 1 に示す。HMMの再学習により境界誤差の平均 11.5 ms、音素セグメンテーション率 96.1% (50ms 以内)、89.1% (30ms 以内) まで性能が向上している。

表 1: 単語音声データ(特定話者)の実験結果

	境界誤差 (ms)	セグメンテーション率 (%)	
	平均	50 ms	30 ms
初期 HMM	19.8	92.5	83.5
再学習 HMM	11.5	96.1	89.1

連続音声データ(5名の男性話者、各 503 文)をテストデータとして、音素セグメンテーション実験を行った。HMMセグメンテーション部のHMMには、テストデータの発声者とは異なる5名の男性話者の単語データ(各 1264 [= 216 + 5240 ÷ 5] 語)とその音素ラベルデータを用いて学習した音素モデルを用いた。

表 2 に示すように、境界誤差の平均 13.7 ms、音素セグメンテーション率 92.1% (50ms 以内)、82.8% (30ms 以内) という、結果が得られた [18]。

表 2: 連続音声データ (不特定話者) の実験結果

境界誤差 (ms)	セグメンテーション率 (%)	
平均	50 ms	30 ms
13.7	92.1	82.8

## 5 システム構成の実例

我々が、実際に構成したシステムについて説明する。また、ワークベンチを使用してラベリング作業を行い、システムのパフォーマンス、作成されたラベルデータの正確さの評価結果を示す [16,17]。

### 5.1 システム構成例

ワークベンチ・システム構成の実例を図 15 に示す。ワークベンチのユーザーインタフェース・モジュールは、作業用ワークステーション WS<sub>1</sub>~<sub>3</sub> (HP/UX 9000) で動作し、各々 TCP/IP ネットワークを介してセグメンテーション・エンジンと接続される。

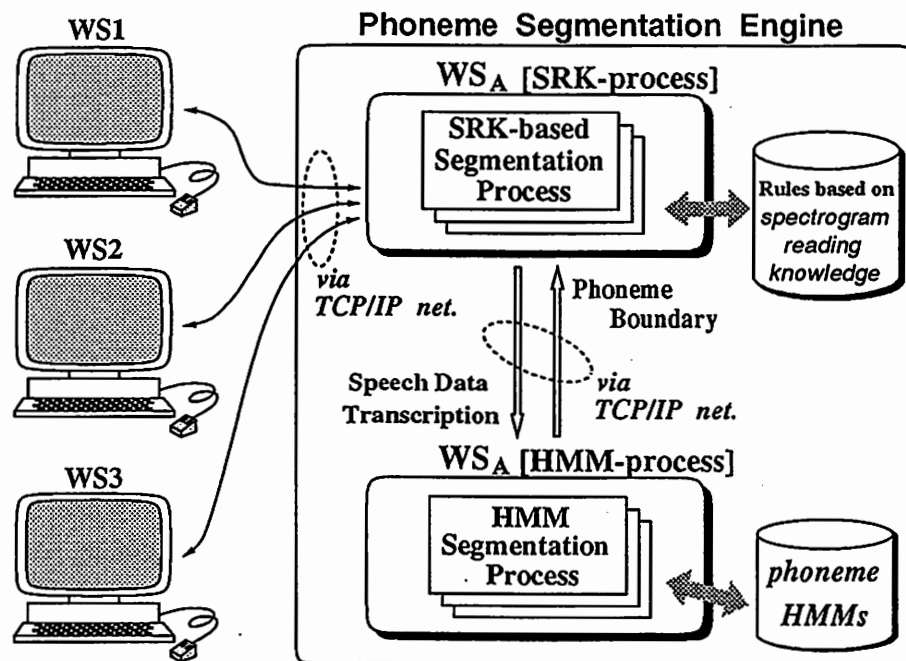


図 15: システム構成の実例

ホストマシン WS<sub>A</sub> (DECstation 5000) では、音素セグメンテーション・エンジンが稼働している。ルール処理部 (SRK-process) は、プログラミング言語 C をベースに ART-IM というエキスパートシステム構築ツールを用いて実現しており、HMMセグメンテーション部 (HMM-process) はプログラミング言語 C で作成されている。

作業者は、作業用 WS のいずれかを使用して音声データのラベリングを実行することができる。また、作業用 WS に X 端末装置を接続すれば、X 端末上でのラベリング作業も可能である。

## 5.2 ワークベンチを用いたラベリング・テスト

ワークベンチを用いて音声データを実際にラベリングして、ワークベンチのパフォーマンス、作成したラベルの正確さを評価した。テストデータとしては、男性話者1名(話者MSH)のバランス単語データ(216語)を使用し、2人の作業者によってラベリングを行った。

作成されたラベルデータを、視察ラベリングで得られているラベルデータと比較した結果を表3に示す。音素境界誤差の平均は6.1ms、標準偏差は10.2msであり、従来の視察ラベリングで得られるラベルデータの精度<sup>[2]</sup>と、ほぼ同程度の精度を持つラベルデータが得られた。

表3: ワークベンチを用いたラベリング結果

境界誤差 (ms)		セグメンテーション率 (%)		
平均	標準偏差	50ms	30ms	10ms
6.1	10.2	99.2	96.7	73.4

パフォーマンス に関しては、ユーザーインタフェース部において、最初の1ページの表示に約5sec、子プロセスを利用した2ページ以降の表示には約2secの時間を要した。処理時間のほとんどは、スペクトログラム表示に用いるビットマップ・イメージの作成に費された。また、音素セグメンテーション・エンジンの応答時間は、単語音声データで約10secであった。

従来の視察によるラベリング作業の工程<sup>[2]</sup>:

- 1) スペクトログラム印刷
- 2) ラベリング — 音素境界付け
- 3) 境界時刻読みとり ~ データファイル作成

と比較すると、1) 3) に要する時間が短縮され、2) のラベリングも操作性の良いユーザーインタフェースにより作業効率が向上し、全体としての作業量はかなり削減できると予想される。

## 6 おわりに

音声データベースの効率的作成・保守のための音素ラベリングワークベンチの研究開発の成果を報告した。ワークベンチの構成、ユーザーインタフェースの機能・操作性について述べ、自動ラベリング・サーバーとして動作する音素セグメンテーション・エンジンのセグメンテーション性能の評価結果を示した。さらに、我々が実際に構成したシステムについて説明し、そのパフォーマンス・作成したラベルデータの正確さの評価結果を示した。ワークベンチの処理時間は十分実用に耐えられるものであり、また、ワークベンチを用いて作成したラベルデータは、従来の視察ラベリングによるラベルとほぼ同じ精度であった。

今後は、ワークベンチの処理速度による性能評価を行う必要がある。また、ユーザーインタフェースをより操作性の良いもの・目の疲労が少ないなど使い易いものに改善し、音素セグメンテーション・エンジンについては、連続音声データのより正確な自動ラベリングの実現を目指す。

**謝辞** 研究の機会を与えて下さったATR自動翻訳電話研究所 樽松明社長、嵯峨山室長に感謝します。また、自動ラベリングの手法についてご指導を頂いた小森康弘氏(現キャノン)、システムのインプリメントとラベリング・テストにご協力頂いた下村賢、木田浩子、北野やよいの各氏に感謝します。さらに、適切などご指導、ご助言を頂いたATRの皆様へ感謝します。

## 参考文献

- [1] W. Fisher, V. Zue, J. Bernstein and D. Pallett, "An Acoustic-phonetic Data Base," J. Acoust. Soc. Amer. Suppl.(A), 81(S92) (1987-5).
- [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," 音響学会誌, 44(10), pp.747-754 (1988-10).
- [3] 阿部, 匂坂, 梅田, 桑原, "研究用日本語音声データベース利用解説書(連続音声データ編)," ATR Technical Report(公開), TR-I-0166 (1990-9).
- [4] 田中, 速水, 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," 音響学会誌, 42(11), pp.860-868 (1986-11).
- [5] 小森, 小坂, 大洞, 新中, "統計的手法およびトップダウン情報を用いた男女声の自動ラベリング," 電学論(C), 108-C(10), pp.795-802 (1988-10).
- [6] N. Carbonell, J.-P. Darnes, D. Forh, J.-P. Haton, and F. Lonchamp, "APHONDEX, design and implementation of an acoustic decoding expert system," Proc. of ICASSP 86, pp.1201-1204 (1986).
- [7] 中川, 橋本, "HMM法とベイズ確率を用いた連続音声のセグメンテーション," 信学論(D), J72-D-2(1), pp.1-10 (1989-1).
- [8] 畑崎, 小森, 川端, 鹿野, "スペクトログラムリーディング知識を用いた音韻セグメンテーションエキスパートシステム," 信学論(D) J73-D, (1), pp.1-9 (1990).
- [9] 荒井, 他, "知識処理に基づく音声自動ラベリングシステム," 信学論(D), J74-D-2(2), pp.130-141 (1991-2).
- [10] M.S. Schmidt and G.S. Watson, "The Evaluation and Optimization of Automatic Speech Segmentation," Proc. of Eurospeech 91, pp.701-704 (1991-9).
- [11] 丸山, 川端, "Xウィンドウ上に実現した音声処理ワークベンチ," 音響学会講論集, pp.91-92 (1988-3).
- [12] H.G. Tillmann, M. Hadersbeck, H.G. Piroth and B. Eisen, "Development and Experimental Use of Phonwork a New Phonetic Workbench," Proc. of ICSLP 90, pp.1009-1012 (1990-10).
- [13] 嵯峨山, 板倉, "音声の動的尺度に含まれる個人性情報," 音響学会講論集, pp.589-590 (1979-6).
- [14] 丸山, 村田, 川端, "高精度スペクトログラム表示ルーチン," ATR Technical Report(非公開), TR-I-0021 (1988-2).
- [15] S. Fujiwara, Y. Komori, and M. Sugiyama, "An Integrated System for Automatic Labelling Base on HMM and Spectrogram Reading Knowledge," Proc. of ISSPA 92, pp.275-278 (1992-8).
- [16] S. Fujiwara, Y. Komori, and M. Sugiyama, "A Phoneme Labelling Workbench using HMM and Spectrogram Reading Knowledge," Proc. of ICSLP 92, pp.791-794 (1992-10).
- [17] 藤原, 小森, 杉山, "音素ラベリングワークベンチ," 音響学会講論集, pp.227-228 (1992-10).
- [18] 藤原, 杉山, "HMMとスペクトログラムリーディング知識に基づく不特定話者音素セグメンテーションの連続音声における評価," 音響学会講論集, 発表予定 (1993-3).
- [19] 村井, 井上, 砂原, "プロフェッショナル UNIX," アスキー出版, pp.220-221 (1986).
- [20] W. R. Stevens, "UNIX Network Programming," Prentice-Hall (1990).

## 付録 A 初期設定ファイルのキーワード

表 4: メインウィンドウ表示関連

MainWindowX	メインウィンドウの左上X座標
MainWindowY	メインウィンドウの左上Y座標
MainWindowWidth	メインウィンドウの幅
MainWindowHeight	メインウィンドウの高さ
MainWindowBackground	メインウィンドウの背景色
PopupMenuBackground	ポップアップメニューの背景色

表 5: サブウィンドウの種類

MenuWindow	メニューサブウィンドウ
WaveWindow	波形サブウィンドウ
WavePositionWindow	全波形サブウィンドウ
PowerWindow	ログパワーサブウィンドウ
DeltaCepWindow	△ケプストラムサブウィンドウ
AutoLabelWindow	自動ラベルサブウィンドウ
SpecWindow	スペクトログラムサブウィンドウ
ColorSpecWindow	濃淡スペクトログラムサブウィンドウ
TimeScaleWindow	時間軸サブウィンドウ
LabelWindow1	ラベル (第1レイヤ) サブウィンドウ
LabelWindow2	ラベル (第2レイヤ) サブウィンドウ
LabelWindow3	ラベル (第3レイヤ) サブウィンドウ
LabelWindow4	ラベル (第4レイヤ) サブウィンドウ
LabelWindow5	ラベル (第5レイヤ) サブウィンドウ

表 6: サブウィンドウ表示関連

WindowSeparate	サブウィンドウ間のセパレート幅
SubWindowBackground	サブウィンドウの背景色
LayerHighlight	ラベルサブウィンドウ選択時の背景色
BDLineColorName	音素境界線の色
BDRemarkColorName	音素境界線のリマーク時の色
BDLineWidth	音素境界線の幅
DisplayTimeRange	1 ページに表示する時間 (m 秒)
DisplayScrollTime	スクロール時間幅 (m 秒)
DisplayWaveRange	波形の表示レンジ
DisplayPowerRange	パワーの表示レンジ
DisplayDeltaCepRange	△ケプストラムの表示レンジ
DisplaySpecFreqRange	スペクトログラムの表示周波数レンジ
DisplaySpecCoeff	濃淡スペクトログラムの強度変換係数
DisplaySpecRange	濃淡スペクトログラムの表示レンジ



表 7: 音響分析関連

AnaSamplingFreq	サンプリング周波数 (kHz)
AnaFrameWidth	分析フレームサイズ (サンプル数)
AnaFrameShift	分析フレームシフト長 (サンプル数)
AnaFFTLengh	FFT 分解点数
AnaLPCOrder	LPC 分析次数
AnaCepOrder	CEP 次数
AnaDeltaCepDelta	$\Delta$ フレーム数 (平均変化率)

表 8: ラベルシンボル入力メニュー関連

LabelMenu1	第1層ラベル入力メニューのシンボル
LabelMenu2	第2層ラベル入力メニューのシンボル
LabelMenu3	第3層ラベル入力メニューのシンボル
LabelMenu4	第4層ラベル入力メニューのシンボル
LabelMenu5	第5層ラベル入力メニューのシンボル

表 9: フォント

FontNormal	デフォルトのフォント
FontMenu	メニューで使われるフォント
FontAutoLabel	自動ラベル表示用フォント
FontLabel	ラベル表示用フォント
FontLabelSmall	ラベル表示用フォント (小)
FontLabelTiny	ラベル表示用フォント (極小)
FontTimeScale	時間軸表示用フォント
FontWaveList	波形ファイル名入力メニューのフォント
FontLabelMenu	ラベルシンボル入力メニューのフォント

表 10: 拡張子・コマンド名

ExtensionLabelFile	ラベルファイル名の拡張子
VoiceCommand	音声出力コマンド名

## 付録 B 初期設定ファイルの例

```

#
#
#       LWB customize file
#
#
MainWindowX:          0
MainWindowY:          0
MainWindowWidth:      0.7
MainWindowHeight:     0.7
MainWindowBackground: #cb7955
SubWindowBackground: #d2ad8c
PopupMenuBackground: #f6e7c2
LabelLayerHighlight: #f7af79
BDLineColor:         #606060
BDRemarkColor:       red
BDLineWidth:         2
#
MenuWindow:           0.040
WavePositionWindow:  0.040
WaveWindow:           0.115
PowerWindow:         0.05
DeltaCepWindow:      0.05
AutoLabelWindow:     0.042
SpecWindow:          0.345
#ColorSpecWindow:    0.345
TimeScaleWindow:     0.05
Label1Window:        0.04
Label2Window:        0.04
Label3Window:        0.04
Label4Window:        0.04
Label5Window:        0.04
WindowSeparate:      5      [pixel]
#
DisplayTimeRange:    1200   [ms]
DisplayScrollTime:   1000   [ms]
DisplayWaveRange:    ( -10000 10000 )
DisplayPowerRange:   ( 0.0 80.0 )
DisplayDeltaCepRange: ( 0.0 6.0 )
DisplaySpecFreqRange: 6000   [Hz]
DisplaySpecRange:    ( 17.0 95.0 )
DisplaySpecCoeff:    -10.0
#
# for 12K
AnaSamplingFreq:     12      [kHz]
AnaFrameWidth:       120     [sample]
AnaFrameShift:       30      [sample]
AnaFFTLenght:        256     [sample]
AnaLPCOrder:         16
AnaCepOrder:         16
AnaDeltaCepDelta:    8       [frame]
#
# for 20K
#AnaSamplingFreq:    20      [kHz]
#AnaFrameWidth:      200     [sample]
#AnaFrameShift:      50      [sample]
#AnaFFTLenght:       256     [sample]
#AnaLPCOrder:        17
#AnaCepOrder:        17
#AnaDeltaCepDelta:   8       [frame]
#
#
#
LabelMenu1: AUTO \\ a i u e o N Q pau , \\ k s sh ch t n h f m y r w g z j d b p \\ ky ty ny hy my ry gy dy by py
LabelMenu2: AUTO \\ a i u e o N Q pau , \\ < > cl *cl tr dj mm \\ k s sh ch t n h f m y r w g z j d b p
LabelMenu4:      a i u e o N Q pau , \\ < > cl *cl tr dj mm \\ k s sh ch t n h f m y r w g z j d b p
LabelMenu5: a i u e o
LabelMenu3: fr dv
#
MenuSpeechIn:       SpeechIn
MenuLoadWave:       WAVE
MenuLoadLabel:      LABEL

```

```
MenuPreviousPage:      Prev.Page
MenuNextPage:         Next.Page
MenuSaveLabel:        Save
MenuHelp:             Help
MenuQuit:             Quit
MenuVoice:            Voice
MenuPrint:            Print
#
FontNormal:           -adobe-times-bold-r-normal--18-180-75-75-p-99-iso8859-1
FontMenu:             -adobe-helvetica-bold-o-normal--18-180-75-75-p-104-iso8859-1
FontAutoLabel:        -adobe-times-bold-r-normal--18-180-75-75-p-99-iso8859-1
FontAutoLabel:        -adobe-times-bold-r-normal--24-240-75-75-p-132-iso8859-1
FontLabel:            -adobe-times-bold-r-normal--18-180-75-75-p-99-iso8859-1
FontLabelSmall:       -adobe-times-bold-r-normal--14-140-75-75-p-77-iso8859-1
FontLabelTiny:        -adobe-times-bold-r-normal--12-120-75-75-p-67-iso8859-1
FontTimeScale:        -adobe-helvetica-medium-r-normal--18-180-75-75-p-98-iso8859-1
FontWaveList:         -adobe-courier-bold-r-normal--14-140-75-75-m-90-iso8859-1
FontLabelMenu:        -adobe-times-bold-r-normal--18-180-75-75-p-99-iso8859-1
#
ExtensionLabelFile:   .LB
VoiceCommand:         daout
#
```

## 付録 C 対外研究発表抄録

1. [Fujiwara91ASJ3] 藤原, 小森, 杉山: “HMMとスペクトログラムリーディング知識に基づくハイブリッド音素セグメンテーションシステムの構想,” 音学講論, 2-5-16, pp. 85-86 (1991.3).
2. [Fujiwara91KOREA7] S.Fujiwara, Y.Komori, M.Sugiyama: “A Hybrid System for Phoneme Labelling based on HMM and Spectrogram Reading Knowledge,” Proc. of Korea-Japan Joint Workshop, pp. 121-126 (1991.7).
3. [Fujiwara91ASJ10] 藤原, 岩橋, 小森, 杉山: “HMMとスペクトログラムリーディング知識に基づくハイブリッド音素セグメンテーションシステム,” 音学講論, 2-5-20, pp. 87-88 (1991.10).
4. [Iwahashi91ASJ10] 岩橋, 藤原, 小森, 杉山, 匂坂: “自動セグメンテーションによる音声合成単位の作成,” 音学講論, 1-6-21, pp. 231-232, (1991.10).
5. [Fujiwara92ISSPA92] S.Fujiwara, Y.Komori and M.Sugiyama: “An Integrated System for Automatic Labelling Based on HMM and Spectrogram-Reading Knowledge,” Proc. of ISSPA92, pp. 275-278 (1992.8)
6. [Fujiwara92ICSLP] S.Fujiwara, Y.Komori and M.Sugiyama: “A Phoneme Labelling Workbench using HMM and Spectrogram Reading Knowledge,” Proc. of ICSLP92, pp. 791-794 (1992.10)
7. [Fujiwara92ASJ10] 藤原, 小森, 杉山: “音素ラベリングワークベンチ,” 音学講論, 2-Q-28, pp. 227-228 (1992.10).
8. [Fujiwara93IEICE1] 藤原, 杉山: “HMMとエキスパートシステムの手法を用いた音素ラベリングワークベンチ,” 信学技報, SP92-132, pp.73-80 (1993.1).
9. [Fujiwara93ASJ3] 藤原, 杉山: “HMMとスペクトログラムリーディング知識に基づく不特定話者音素セグメンテーションの連続音声における評価,” 音学講論, 2-Q-12 (1993.3).