

TR-I-0078

連続音声の基本周波数データベース
Pitch frequency database on continuous speech

阿部 匡伸 桑原 尚夫
Masanobu ABE Hisao KUWABARA

(ATR自動翻訳電話研究所)
ATR Interpreting Telephony Labs.

1989.4

内容梗概

ATRは、各種研究目的に共同利用可能な音声データベースの構築を進めている。ATRが作成した音声データベースのうち、特に連続音声(文を読み上げた音声)データベースに対して、新たに音声の基本周波数を追加したので報告する。付録として、マニュアルによって基本周波数を修正するツールの使用方法を掲載した。

1.はじめに

ATRは、各種研究目的に共同利用可能な音声データベースの構築を進めている。これまでに階層的音韻ラベル付けを提案し、ラベル付きの音声データベースを提供してきた^{[1][2]}。これら音声データベースのうち、特に連続音声(文を読み上げた音声)データベース^{[3][4]}に対して、新たに基本周波数を追加したので報告する。付録として、マニュアルによって基本周波数を修正するツールの使用方法を掲載した。

2.基本周波数データベースの目的とタスク

音声の基本周波数を自動的にかつ誤りなく行うことは、大変困難である。基本周波数の自動抽出法に関しては、これまでも多くの検討がなされているが、最終的には人手による基本周波数抽出ミスの修正は避けられない。このような理由から、音声研究に基本周波数を利用することは、大変な労力を伴うものであった。本基本周波数データベースの目的は、抽出ミスのない基本周波数を提供することによって、音声研究を援助することにある。

基本周波数変化の現象には、微視的な特徴(例えば、子音から母音への渡りにおける、子音の種類と基本周波数変化との関係)から巨視的な特徴(例えば、アクセント型と基本周波数の変化パターンとの関係)までいくつかのレベルがある。本基本周波数データベースは、主として、巨視的な特徴の提供を目的として作成された。その理由は、

- ①基本周波数の微妙な変化を明らかにするためには、高精度の基本周波数抽出とそれに加えて視察による修正など莫大な手作業が必要となる。
- ②微視的な特徴を取り扱う場合には、音声信号から抽出された基本周波数だけでは不十分であり、声帯の振動との関係を明かにしておくことが必要であると考えられる。

以上の考えに沿って、ATRの音声データベースのうち連続音声(文を読み上げた音声)のデータベース^{[3][4]}に関して基本周波数データベースを作成した。

3.基本周波数抽出法と修正

基本周波数の自動抽出には誤りが避けられないことから、基本周波数の自動抽出を行った後に、視察による修正を加えた。基本周波数抽出法と抽出誤りの修正の方針を以下に述べる。

3.1基本周波数の抽出法

基本周波数の抽出は、次のように行った(ケプストラム法)。まず音声をフーリエ変換し、対数振幅スペクトルを求めた後に逆フーリエ変換してケプストラムを求める。次に、基本周波数に相当する探索区間でケプストラムの最大値を求め、最大値を与えるケフレンシーから基本周波数を決定する。分析の条件を表1に示す。

表1. 分析条件

音声データ	12KHzサンプリング 16bit
分析窓	ブラックマンウインドウ
分析窓長	384ポイント (32msec)
分析シフト	30ポイント (2.5msec)
FFTポイント数	512ポイント
ピッチ探索範囲	男性話者 50~350ポイント (34Hz ~240Hz) 女性話者 24~120ポイント (100Hz ~500Hz)

3.2抽出誤り修正の方針

自動抽出された基本周波数と音声波形をワークステーション上に同時に表示し(約2秒分)、人間が視察によって修正を加えた(基本周波数の修正に用いたツールに関しては、付録を参照)。修正は次の順序に従って行った。

- ①音声波形が存在しない部分で基本周波数が表示されていた場合には、これを削除する。
- ②基本周波数の概形を観察し、際だって不連続であった場合には、その基本周波数を整数倍または整数分の1を計算し、その結果概形に当てはまったらその基本周波数を正解とする。
- ③抽出された基本周波数が乱れている場合には、音声波形のピークからピークを指定して基本周波数を計算する。
- ④③の処理で音声波形から基本周波数を計算する場合において、連続する複数の基本周期を計算する必要があり、かつその基本周波数が緩やかに変化していると判断できる場合には、2点の基本周波数を計算後、その間の基本周波数を直線補間によって決定する。
- ⑤以上の方法によって基本周波数を決定できない場合には、自動抽出の値をそのまま保存する。

例として、図1にケプストラム法によって得られた基本周波数を、図2に修正後の基本周波数を示す。

4.データフォーマット

基本周波数データは、基本周期のポイント数で表現されており、c言語の `short integer` 形式で格納されている。3.2の④で直線補間によって修正された基本周期は、他と区別するため、修正後の基本周期にマイナス1を掛けてある。

基本周波数データベースのファイル名は、音声データベースの命名法に準拠している⁵⁾。命名のフォーマットを図3に示す。

話者名: 性別(M=男、F=女)、話者のイニシャル(YI)
 タスク: 連続音声データベース、自由発声(__SD__)
 文章番号: 文は10のグループに別れており(A~J),

各グループが50文(01~50)からなる。
識別子:基本周期データ(PIT)

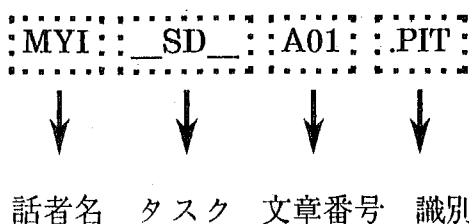


図3. ピッチデータファイル名のフィールド構成

5.まとめ

連続音声データベースに基本周波数データを追加した。その結果、連続音声データベースで基本周波数を容易に利用をできるようになった。

謝辞 本研究を進める機会を与えてくださった樽松社長に深謝致します。また日頃有益な討論を頂く鹿野室長をはじめとするATR自動翻訳電話研究所研究員の皆様に感謝致します。

参考文献

- [1]武田、匂坂、片桐”音声データベース構築のための音韻ラベリング”,音講論集,2-5-10(1987.3)
- [2]武田、匂坂、片桐、桑原”音韻ラベルを持つ日本語音声データベースの構築”,音声研資,sp87-19(1987)
- [3]磯、渡辺、桑原”音声データベース用文セットの設計”,音講論集,2-2-19(1988.3)
- [4]Kuwabara,Takeda,Sagisaka,Morikawa,Watanabe”Construction of a Large-Scale Japanese Speech Database and its Management System”,ICASSP’89(1989.5)
- [5]武田、匂坂、片桐、阿部、桑原”研究用日本語音声データベース利用解説書”,ATRテクニカルレポート、TR-I-0028(1988.5)

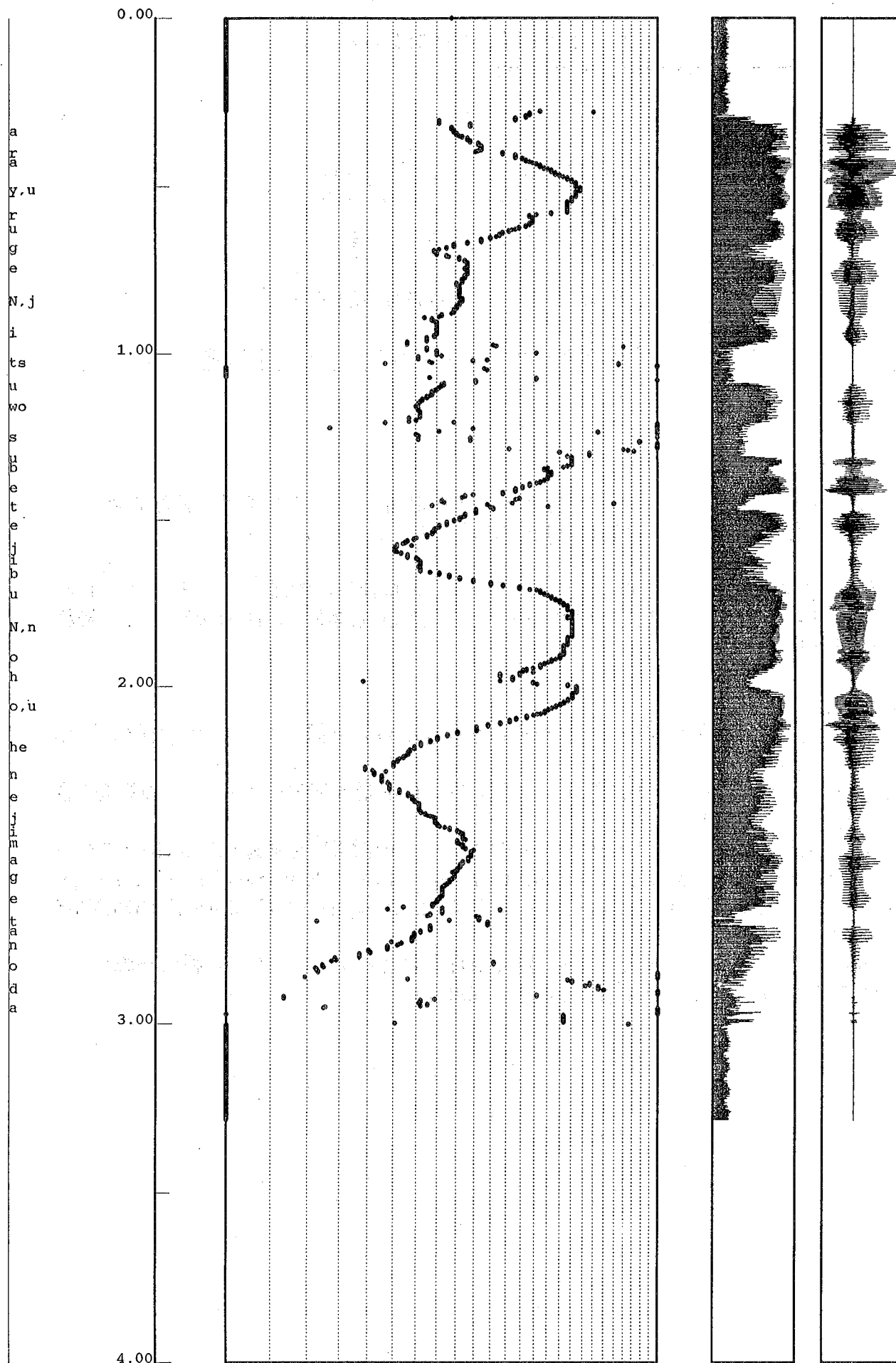


図1 ケプストラム法によって抽出された基本周波数

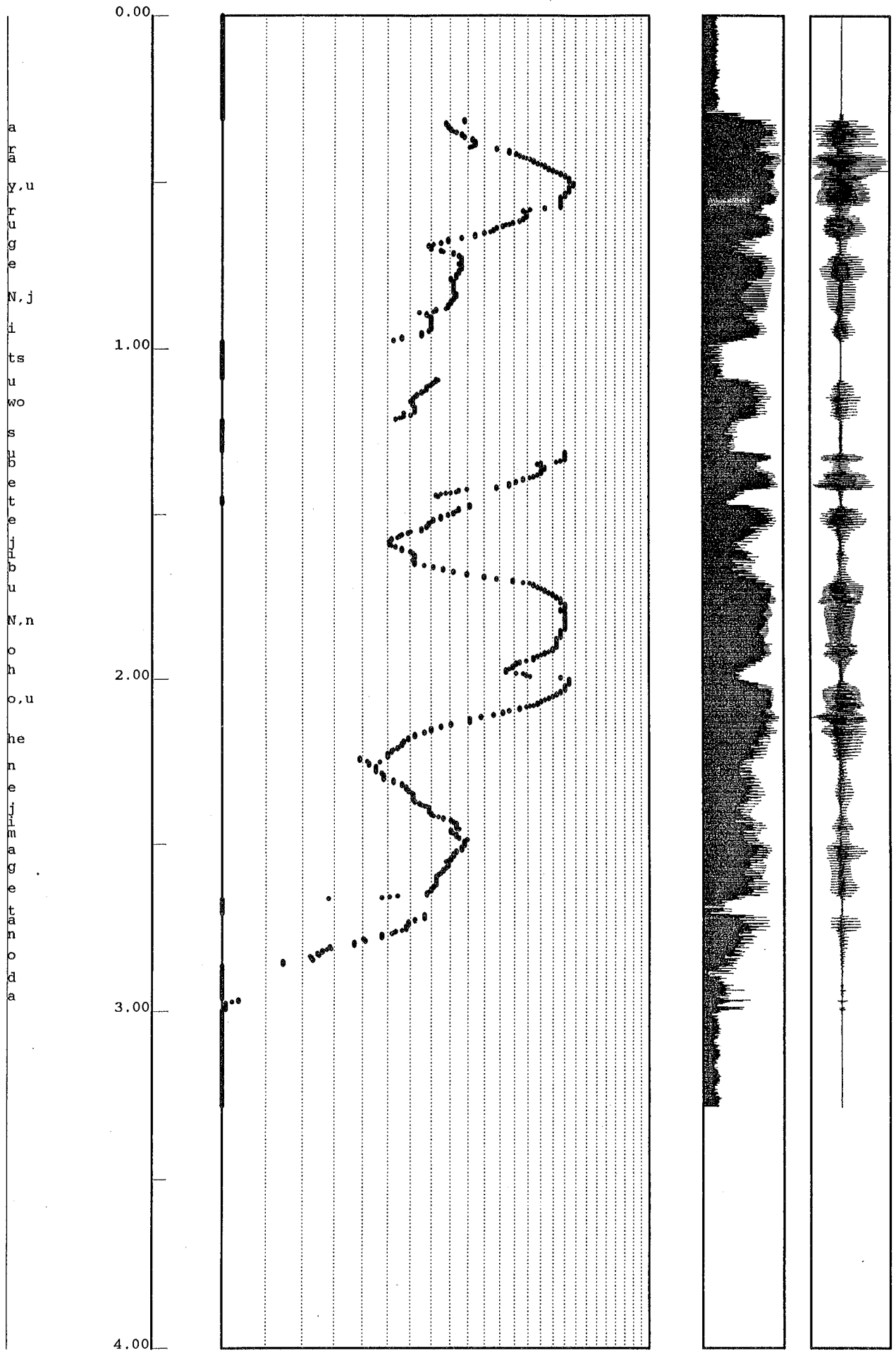


図2 マニュアル修正後の基本周波数

付録

音声の基本周波数修正ツール「ped」

○ ピッチ修正プログラム 'ped' 使用説明

§ 1. 機能

ピッチデータファイル (short in) の内容を、ワークステーション (以下、WSと記す) 上に表示させ、マニュアル操作によりデータ修正を行う。

§ 2. 対象データ

- ・ピッチデータファイル (必須)
- ・音声データファイル
- ・ラベルデータファイル

但し、表示用音声データファイルは、分析に用いたものと同じでなければならない。

§ 3. セットアップ

1. atr-sp: morikawa から以下のファイルを copy する。

- ・ pedrc (パラメータファイル)
- ・ peder (エラーメッセージファイル)

2. 環境変数 'DISPLAY' を設定する。

(例) setenv DISPLAY atrv04:0

3. 使用するワークステーションで、'xhost' を実行して、プログラムを実行するマシンが含まれているかを見る。
もし、含まれていなければ、'xhost ホスト名' を実行する。

§ 4. パラメータファイル '.pedrc' について

```
csh>cat .pedrc
```

```
FREQUENCY      12000
SHIFT_POINTS    30
LINE_MODIFY     MINUS
PITCH_SCALE     LOG
```

キーワード説明

- ・ FREQUENCY
分析、表示用波形データサンプリング周波数
- ・ SHIFT_POINTS
分析時のフレームシフトポイント数
- ・ LINE_MODIFY

直線補間で修正されたデータを正值、または負値にする。
' PLUS ' または、' MINUS ' に設定

- PITCH_SCALE
ピッチ表示のスケールの設定
' LINEAR ' または ' LOG ' に設定

§ 5. 表示画面フォーマット

Fig. 1 に表示画面のフォーマットを示す。図中に記述した名称を以下の説明で用いる。

☆表示画面に関して

- ・スクリーンに表示されているデータ長は 2.0 sec である。
- ・ウインドウに表示されているデータは 100 フレームである。

§ 6. 実行方法

1. コマンド syntax

```
ped <音声データファイル名> <ピッチデータファイル名1>  
    <ラベルデータファイル名> <ピッチデータファイル名2>  
    <オプション>
```

- ・ピッチデータファイル 1 は必ず存在しなければならない。
- ・音声データ、ラベルデータは存在すれば表示する。
- ・修正した結果はピッチデータファイル 2 に出力される。
- ・オプション：ピッチ表示エリアの指定。

```
-m ..... 男性 ( 50 - 300 Hz)  
-f ..... 女性 (150 - 500 Hz)
```

2. 操作方法

プログラム起動後の操作は全て、マウスの移動、中央ボタンクリックにより行なう。

3. 操作手順

プログラムを起動するとワークステーション上に、波形、パワー、ピッチが表示され、次にスクリーンメニューが表示される (Fig 1)。

☆スクリーン操作

この段階では、以下の二つの操作のどちらかが可能である。

○スクリーンメニュー選択

- ・NextScreen
次のページに進む。
- ・PreviousScreen
前のページに戻る。
- ・RefreshScreen
画面の再表示
- ・Savefile&Exit
修正結果をファイルに書きだし終了する。

• **Q u i t**

プログラムから抜ける。修正結果はファイルに書き出さない。

- ピッチ表示スクリーン上でピックアップしたい範囲を指定。
(始点、終点でマウス中央ボタンをクリック。)

ピックアップされたデータは、ピックアップウィンドウに表示される。
データ修正はここでピックアップした範囲のデータに関してのみ可能。
ピックアップ出来る最大の範囲はスクリーン一画面分である。
この後、"ウィンドウ操作"に移る。

☆ウインドウ操作

ここでの処理は以下のメニューより選択する。

- **N e x t**
次のページに進む。
- **P r e v i o u s**
前のページに戻る。
- **M o d i f y**
選択されたポイント値を修正する。これを選択すると、
”修正モードメニュー”が表示され、そこで選択されたモードに
従ってポイントの値が修正される。
- **R e g i o n**
選択されたポイント間の値を修正する。これを選択すると、
”修正モードメニュー”が表示され、そこで選択されたモードに
従ってポイントの値が修正される。
選択されたポイントが2点でなければ処理されない。
選択された2点も処理対象になる。
- **L i n e**
選択された2点間のデータを直線補間する。選択されたポイントが
2点でなければ処理されない。また、0以下の値を持つポイントが
選択された場合も処理は行なわれない。
- **F o c u s**
選択された2点間のデータを拡大表示し、波形から手作業でピッチ
を計算する。選択されたポイントが2点でなければ処理されない。
- **C r e a t e**
選択されたポイントの値を、マウスがクリックされた位置の値に
修正する。選択されたポイントは、1点でなければならない。
- **E n d**
修正結果をsaveして、”スクリーン操作”に戻る。
- **Q u i t**
修正結果を無効にして、”スクリーン操作”に戻る。

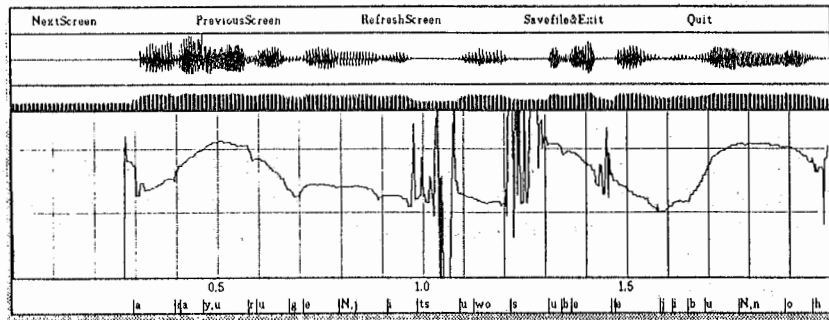
○処理の対象となるポイント選択

ピックアップウインドウ内にマウスを移動すると、クロスバーが
現れる。選択したいポイント付近にパーティカルバーを合わせて、
マウス中央ボタンをクリックする。選択されたポイントは、
' x ' で表示される。

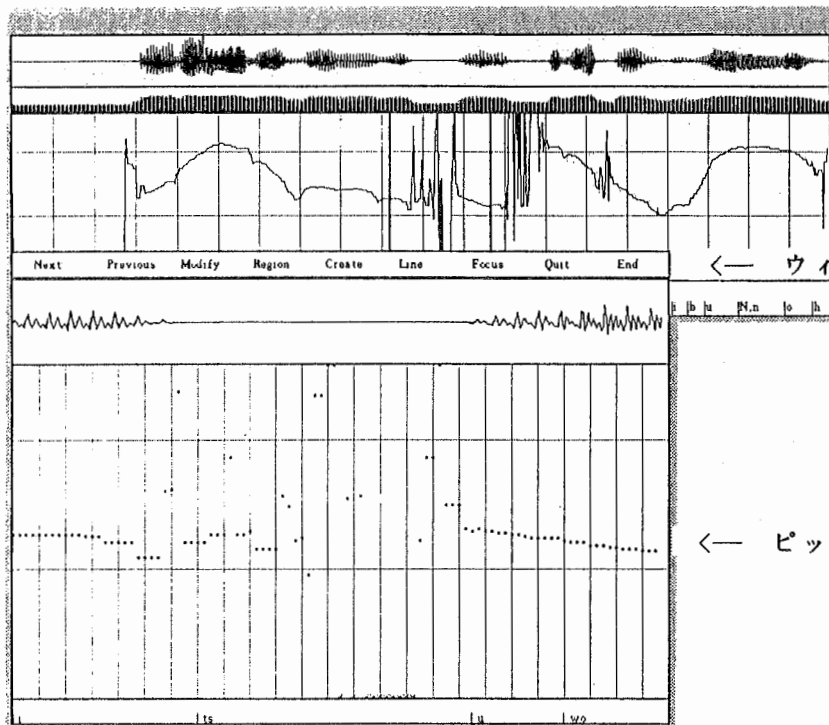
○修正モードメニュー選択 (M o d i f y , R e g i o n)

- **O r g** ピッチデータファイル1の値に戻す。
- **M o d** ピッチデータファイル2の値に戻す。
- **N o p** 何も行なわない。
- **1 / 2** 1 / 2 倍
- **2** 2 倍
- **1 / 3** 1 / 3 倍
- **3** 3 倍
- **1 / 4** 1 / 4 倍
- **4** 4 倍
- **Z e r o** 対象データを0にする。

修正モードに従って変更されたデータは ' . ' で、変更前のデータは ' * ' で表示される。また、0に修正されたデータは ' x ' で直線補間で修正されたデータは ' * ' で表示される。

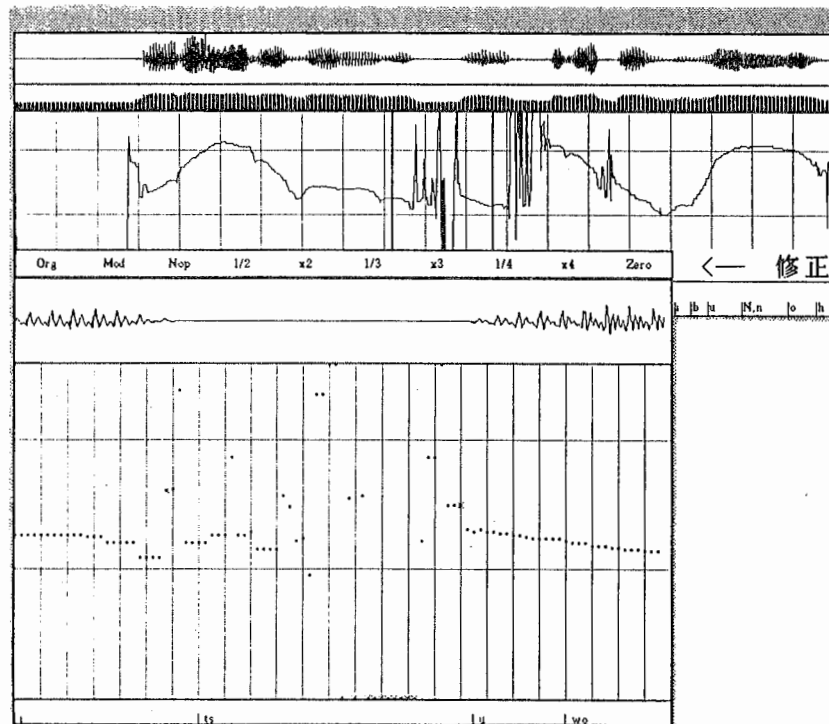


- <— スクリーンメニュー
- <— 波形表示スクリーン
- <— パワー表示スクリーン
- <— ピッチ表示スクリーン



<— ウィンドウメニュー

<— ピックアップウィンドウ

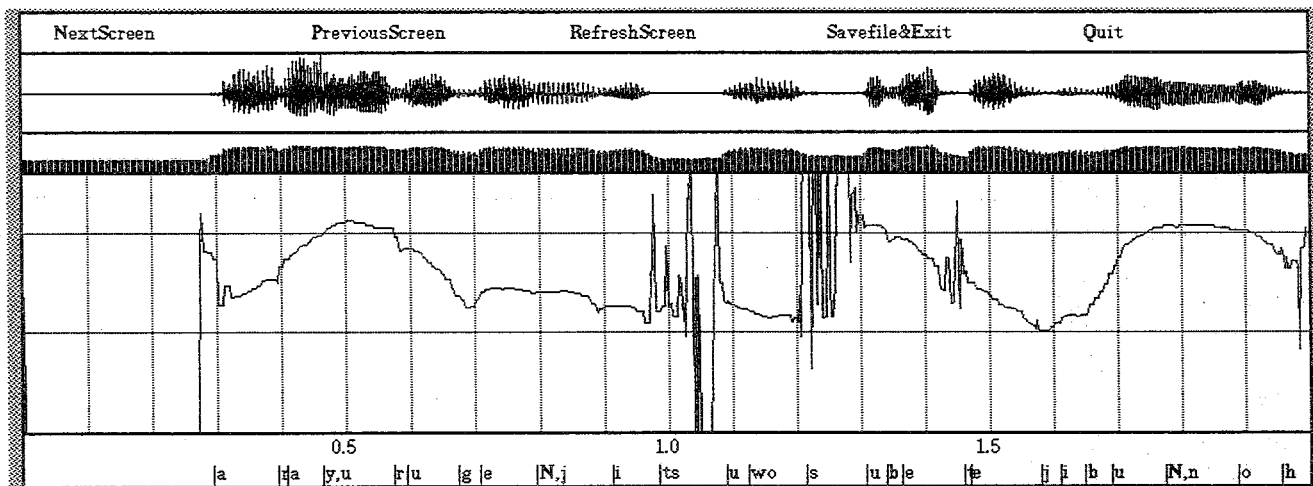


<— 修正モードメニュー

Fig. 1 表示画面

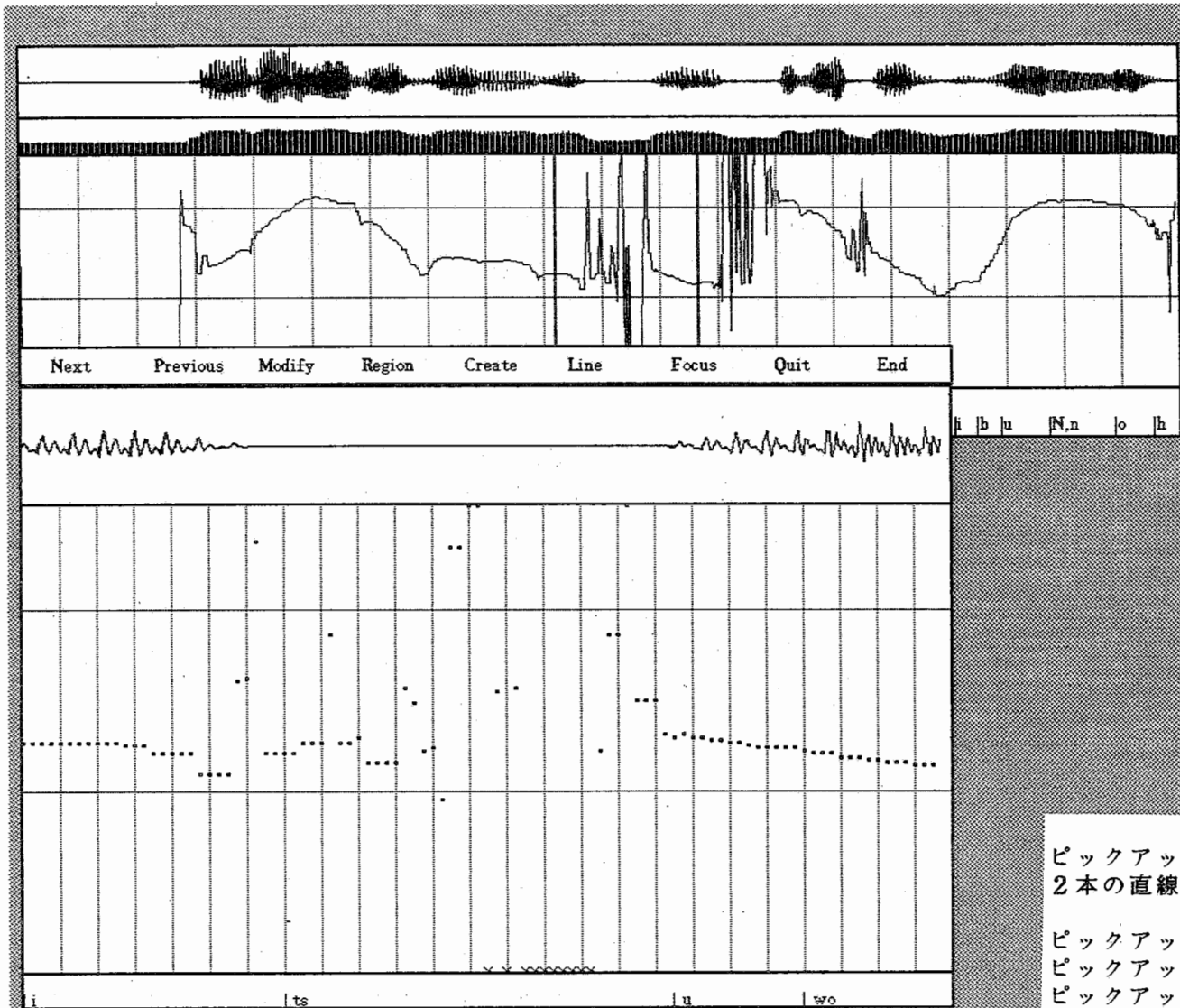
◎ 実行サンプル例

1. プログラムを起動する。
2. ピックアップウィンドウの2ポイントをクリックする。
ピックアップされた区間はピッチ表示スクリーン上で2本の直線に囲まれた部分である。
ピックアップされた始めの250msec分のデータがピックアップウィンドウ上に表示される。ピックアップウィンドウ上の縦の点線の間隔は10msecである。
3. 修正したいポイントを選択する。
選択されたポイントを表示するマーカーが変化する。
4. ウィンドウメニューで"Region"を選択する。
5. 修正モードメニューで"Zero"を選択する。
選択された2点間のデータが変更される。修正前のデータが"*"で、修正後のデータが"."で表示される。
6. ウィンドウメニューで"End"を選択する。
修正結果がピッチ表示スクリーンに反映される。



実行サンプル
1. プログラム起動直後の画面

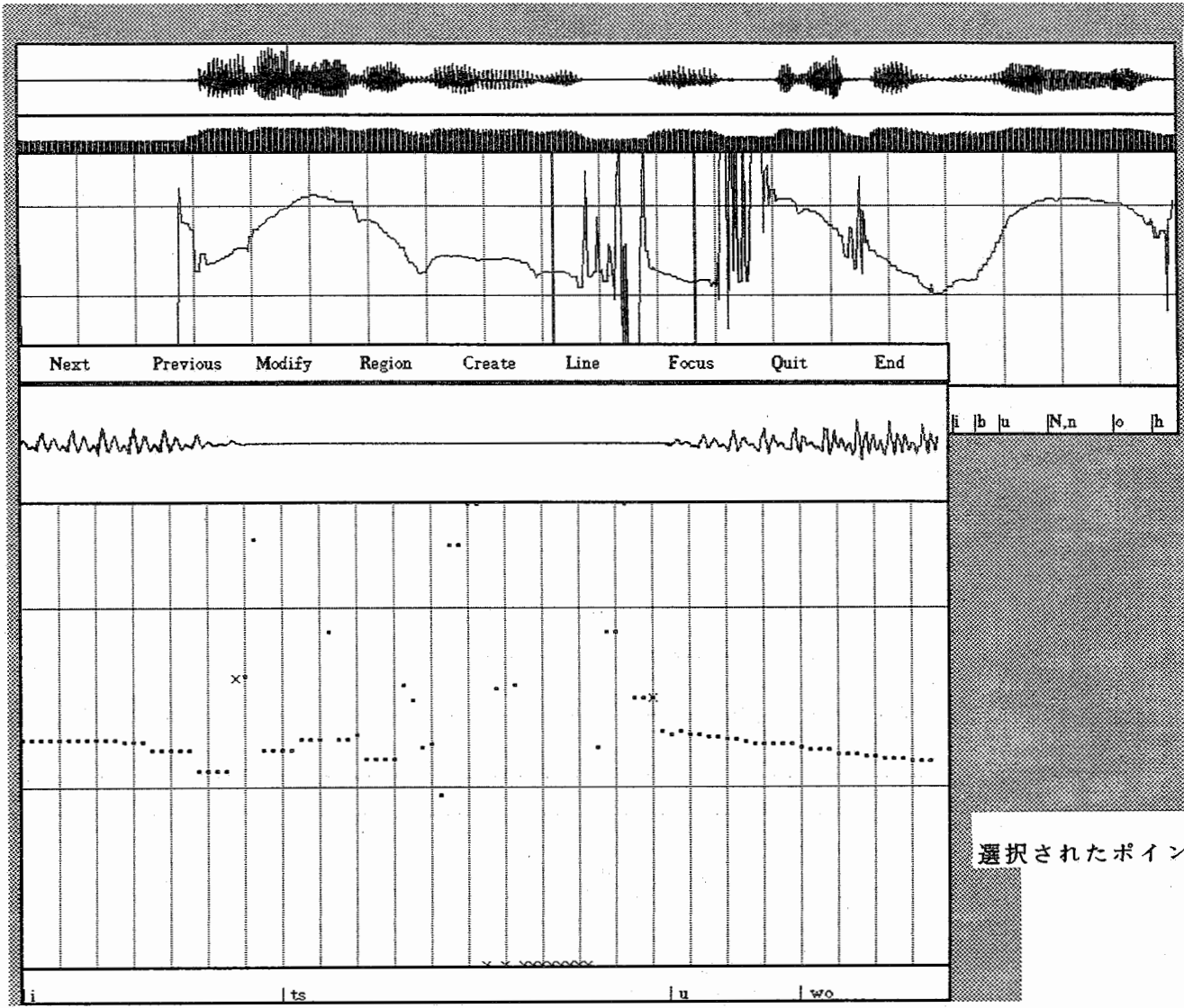
実行サンプル
2. ピックアップ直後の画面



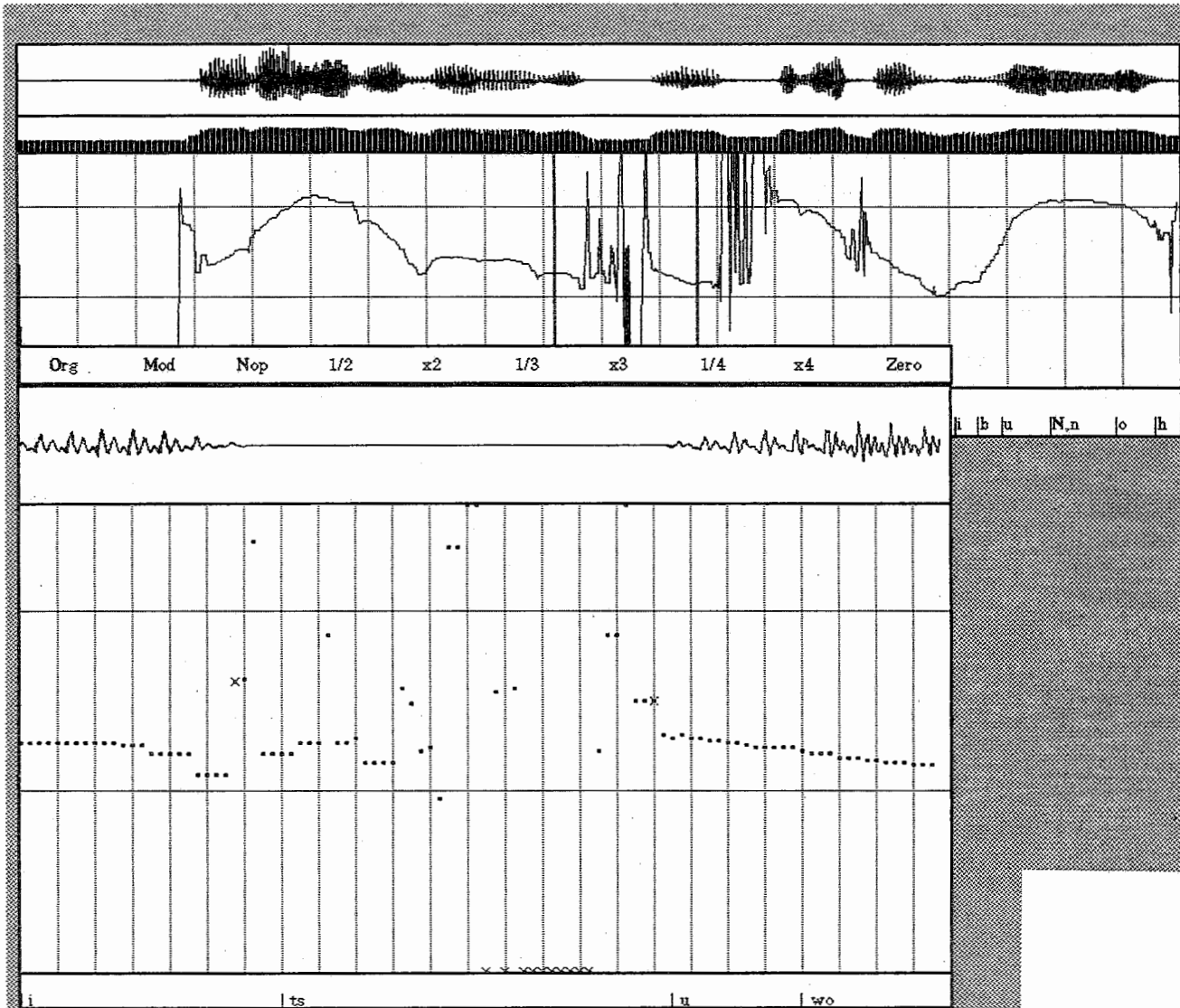
ピックアップされた区間はピッチ表示スクリーン上で
2本の直線に囲まれた部分。

ピックアップされた始めの250msec分のデータが
ピックアップウィンドウ上に表示されている。
ピックアップウィンドウ上の縦の点線の間隔は10msec。

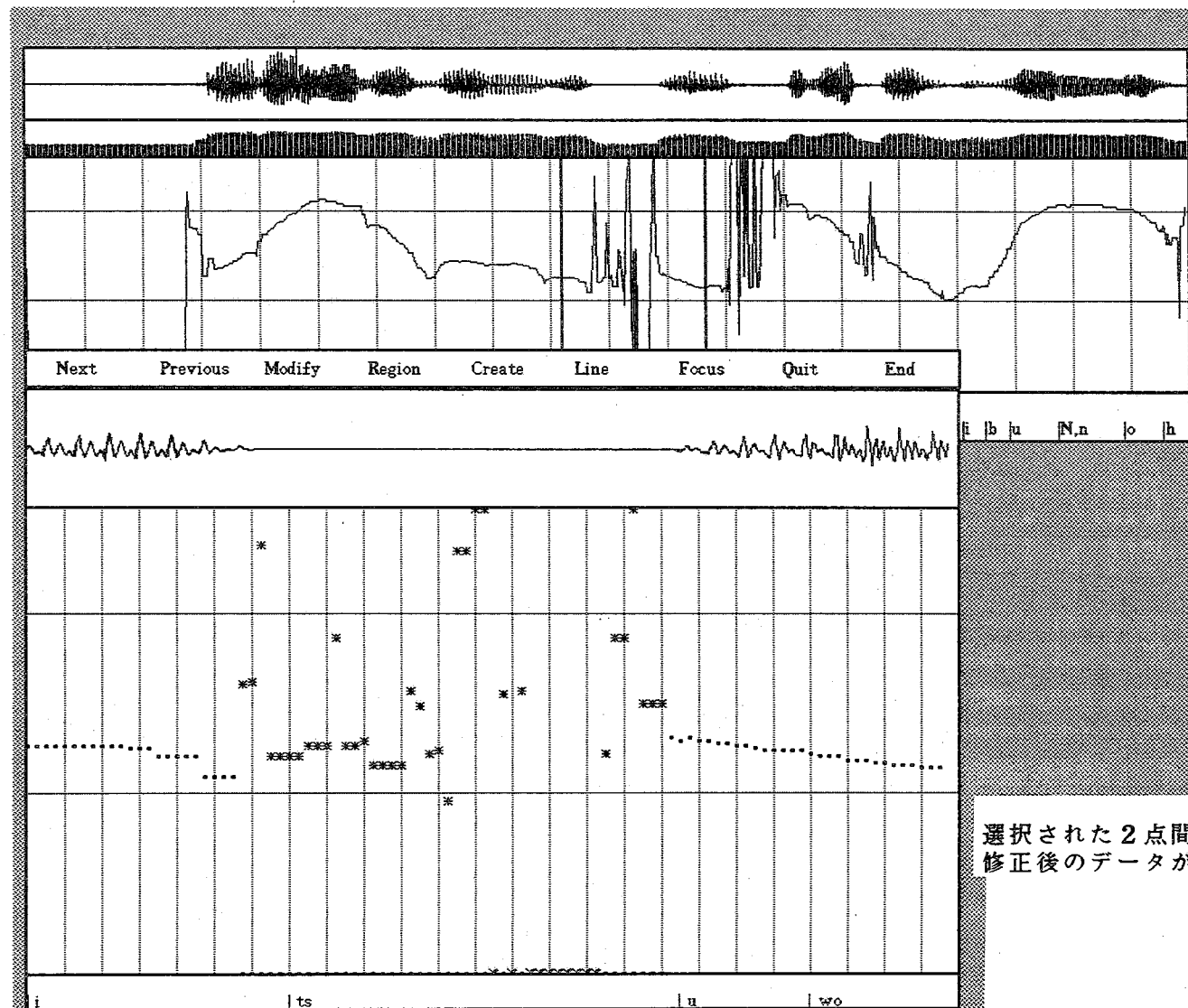
実行サンプル
3. ポイント選択後の画面



選択されたポイントを表示するマーカーが変化している。

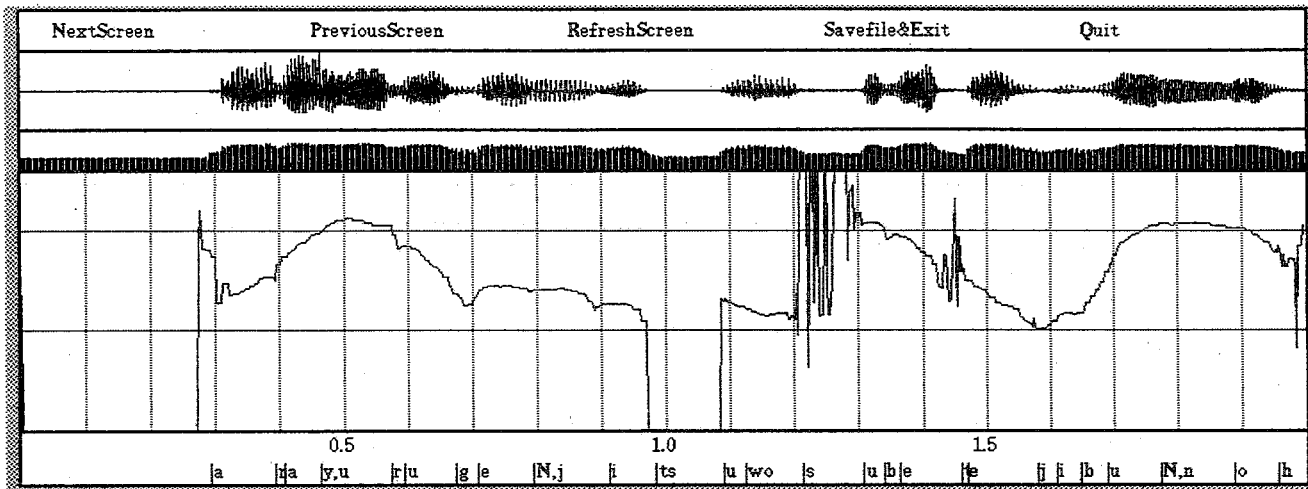


実行サンプル
 4. ウィンドウメニューで"Region"
 を選択した後の画面



実行サンプル
 5. 修正モードメニューで"Zero"
 を選択した後の画面

選択された2点間のデータが変更された。修正前のデータが"*"で、修正後のデータが"."で表示されている。



実行サンプル
 6. ウィンドウメニューで”End”
 を選択した後の画面

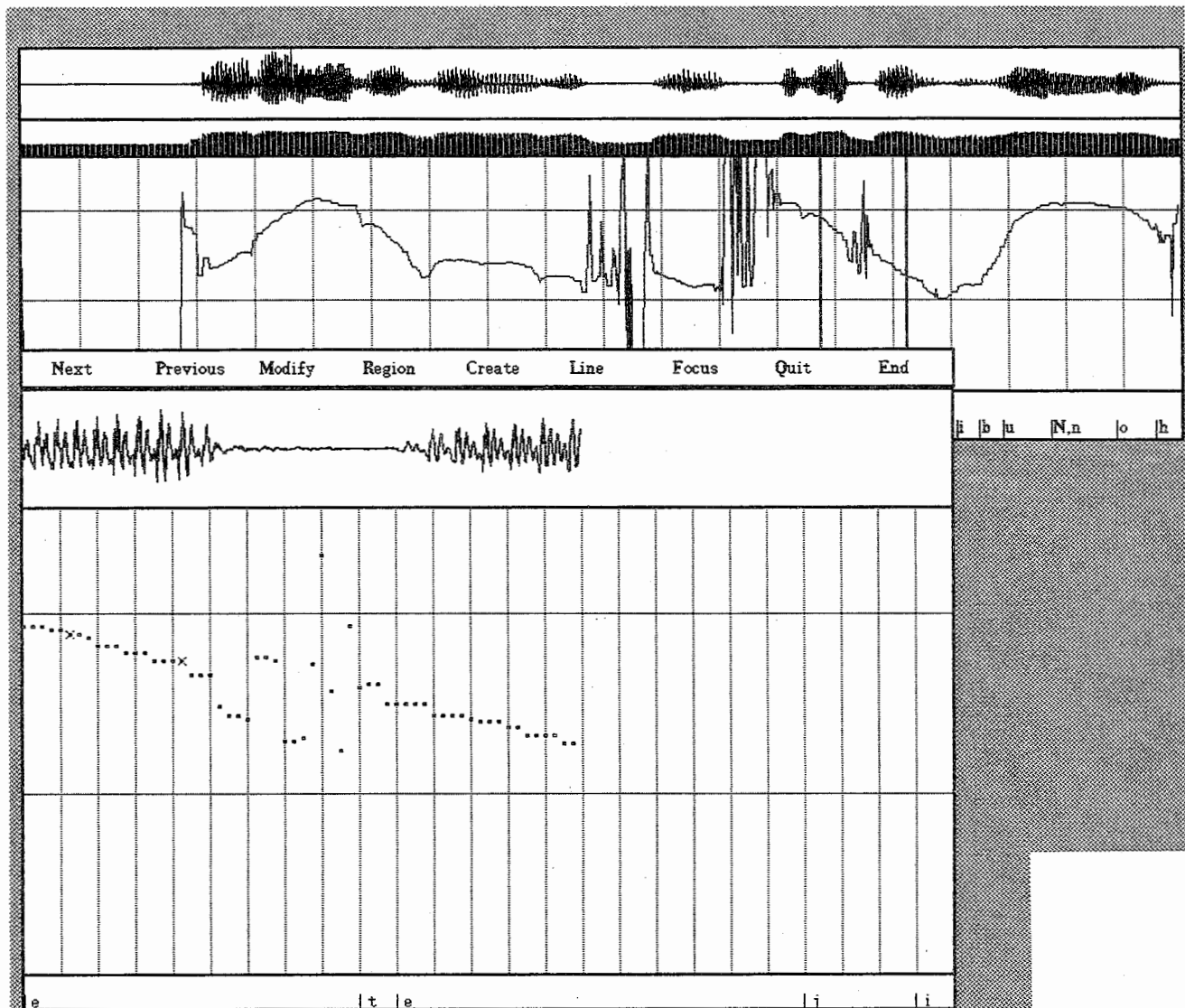
修正結果がピッチ表示スクリーンに反映されている。

◎ Focusの使用例

1. ピックアップウィンドウの2ポイントをクリックする。
2. ウィンドウメニューの'Focus'を選択する。
3. 波形表示エリア内の2ポイントをクリックする。
修正前のデータが"*"で、修正後のデータが"."で表示される。それを
確認後、OKなら'Save'を選択する。NOなら'Quit'を選択する。

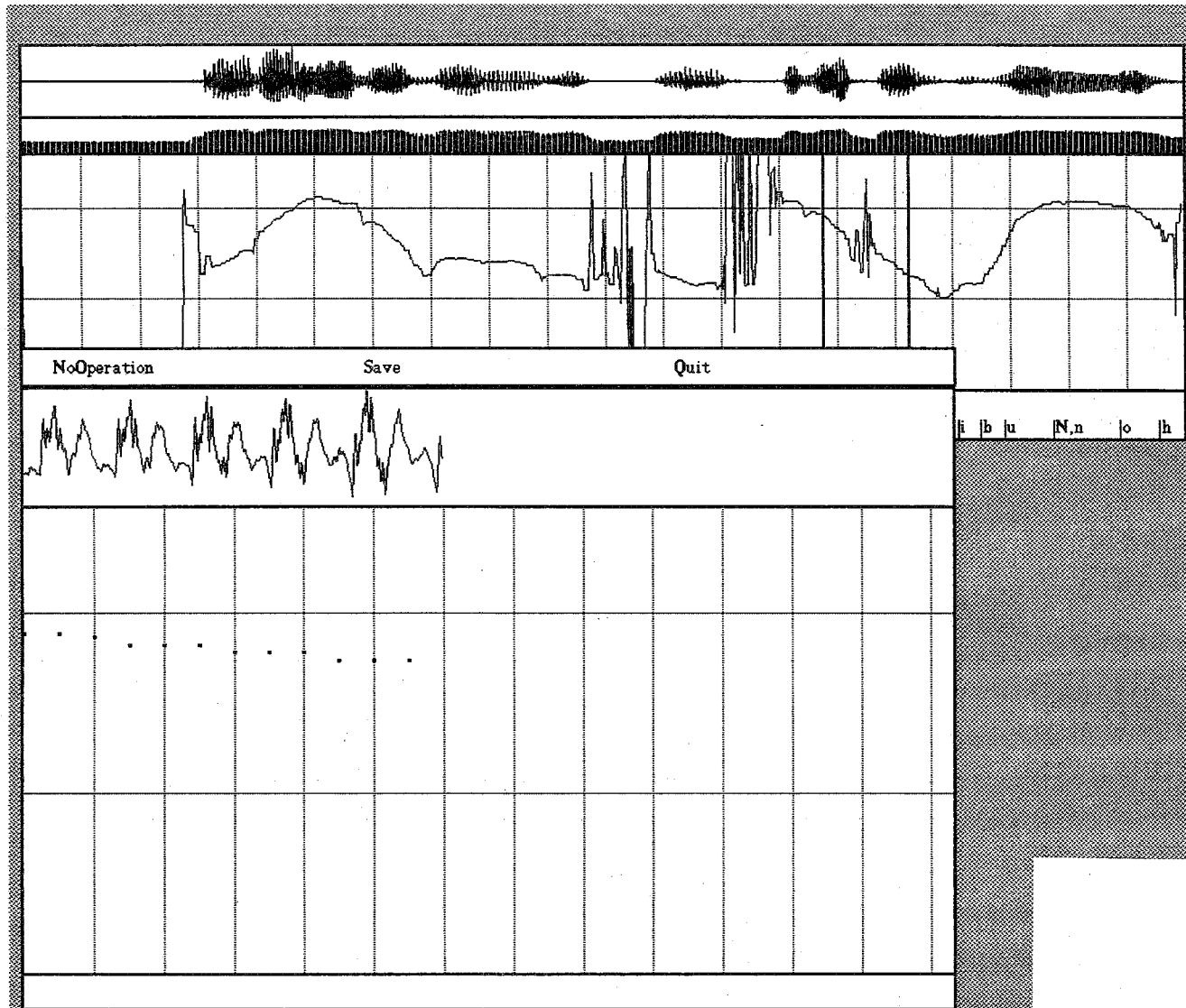
Focusの使用例

1. ピックアップウィンドウの2ポイントを
クリックする。



Focusの使用例

2. ウィンドウメニューの'Focus'を選択する。



F o c u s の使用例

3. 波形表示エリア内の2ポイントを
クリックする。修正前のデータが
"*"で、修正後のデータが"."で
表示される。それを確認後、
OKなら'Save'を選択する。
NOなら'Quit'を選択する。

