

Internal Use Only

非公開

TR - H - 187

0008

音声・楽音の基本周波数
データ修正ツール
- xdpitch -

加藤 宏明

森中 由香里

(TSG, ユーニス)

1996. 2. 20

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2 ☎ 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-774-95-1011
Facsimile: +81-774-95-1008

©(株)ATR人間情報通信研究所

音声・楽音の基本周波数データ修正ツール

— xdpitch —

加藤 宏明 (ATR 人間情報通信研究所第1研究室)

森中 由香里 (TSG, ユーニス)

要旨

基本周波数データ修正ツールをリリースします。これは、自動抽出アルゴリズムなどにより計算された音声・楽音の基本周波数データに対して後処理として手動で修正を加える作業を支援するソフトウェアです。入力された基本周波数データを対応する音声波形や音韻ラベルと画面上で同期させながら、対話的に修正を進めることができます。

もくじ

I	リリースノート編	4
1	はじめに	4
2	開発の背景と経緯	4
3	主な特徴と仕様	5
3.1	主な特徴	5
3.2	主な仕様	6
4	現行版の問題と今後の課題	7
4.1	現行版の問題	7
4.2	今後の課題	7
II	マニュアル編	8
5	環境設定	8
6	プログラムの起動	8
6.1	options	8
6.1.1	-in, -wave, -label	9
6.1.2	-out	9
6.1.3	-period, -sf, -max, -min	9
6.1.4	-sp, -sw	10
6.1.5	-pp	10
7	プログラムの終了	10
8	処理の進め方	10
8.1	画面構成と表示範囲の変更	10
8.1.1	on-map データ	10
8.1.2	on-screen データ	11
8.2	Buttons	11
8.2.1	File...	11
8.2.2	Scale...	12
8.2.3	Region...	12
8.2.4	Undo	12
8.2.5	Magnet	12
8.3	F0 データの修正	13
8.3.1	F0 データ表示領域から	13
8.3.2	時間波形データ表示領域から	13

8.3.3 Region 指定によって	13
-------------------------------	----

図一覧

1 xdpitch の表示画面	5
2 メッセージウィンドウ	8
3 初期画面	11
4 Region によるデータ修正	14
5 数値の入力	14

表一覧

1 オプション	9
-------------------	---

第 I 部

リリースノート編

1 はじめに

本ツールは音声言語情報生成機構の研究において知覚実験に供する音刺激作成の効率化を図る目的で開発された。音声言語情報の知覚実験では、高品質の音声刺激を手軽に得るために実音声を元にした分析合成を利用する機会が多い。この音声の分析合成には一般に基本周波数 (F0) の自動抽出および再利用の手順が含まれる。ところが、実音声の F0 抽出では、明瞭に発話された母音などの場合を除き、何らかの抽出誤りが入ることが多い。人間はごく少量、短時間の抽出誤差でもそれが自然に発話されるものから逸脱している場合には非常に敏感に反応する。抽出誤りを修正しておかなければ、再合成された音声の歪みばかりが気になって知覚実験自体がうまくいかないこともある。この人間にとって「気にならない」F0 パターンをいかに実現するか... これには決まったやり方が存在するわけではない、あるいはそのようなやり方はまだ知られていない。現段階ではカット・アンド・トライすなわち人間が手作業で修正し、再合成された音を聴いて確認する以外に確実な方法は存在しない。その作業を本ツールは強力的にサポートし、介在する人手を軽減する。

本ツールは言語音声に限らず歌声、楽器音などの基本周波数の表示および修正にも利用できる。

2 開発の背景と経緯

本ツールの開発以前にATR人間情報通信研究所においてF0修正に使用できるソフトウェアには、ESPS & waves+ と SoundDesk があった。しかし、以下のような(我々にとっての)問題点があったためいずれも満足できるものではなかった。

- 扱えるデータ長が十分ではない (ESPS & waves+)
- 入出力のデータ形式に制限がある (ESPS & waves+, SoundDesk)
- 画面上での修正方法が限られており (フリーハンドによる輪郭のトレース), 十分な精度が得られない (ESPS & waves+, SoundDesk)
- ATR 音声データベースの音韻ラベルが表示できない (ESPS & waves+, SoundDesk)

また、F0 は対応する時間波形上の基本周期の逆数に対応するので、時間波形を参照しながら波形上で直接基本周期に該当する範囲を指定するという方法を用いれば精度の良い修正が期待できる。しかしそれを簡単に可能にするツールも知る限りでは存在しなかった。そこで、新たに F0 修正専用のツールを開発し、社内での利用に広く供することにした。機能仕様を加藤宏明が設計し、開発には森中由香里があたった。開発環境は SunOS + X-windows であった。

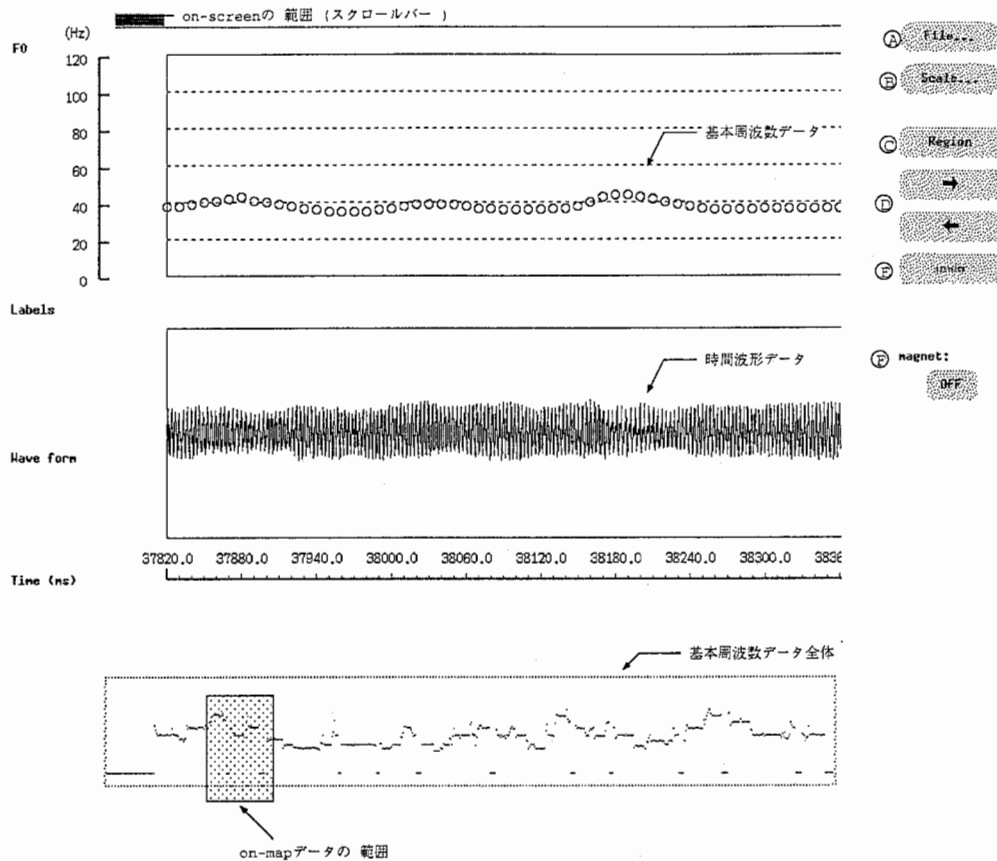


図 1: xdpitch の表示画面

3 主な特徴と仕様

本ツールでは図 1 に示すような操作画面により対話的に F0 データの修正を行なう。以下に主な特徴と仕様を概説する。詳細はマニュアル編を参照してもらいたい。

3.1 主な特徴

表示機能

- F0 データを修正可能な状態で周波数 (Hz)- 時間 (ms) 軸上に表示する (図 1 上段)
- 参照のため、音韻ラベルデータ、波形データを同じ時間軸上に表示する (図 1 中段)
- F0 データ全体の概形を別枠で表示する (図 1 下段)。現在修正中の部分 (図 1 上段) が全体の中で占める範囲を強調表示する (図 1 下段の網掛け部分)
- 時間軸の拡大・縮小・移動はマウス操作により行なう
- F0 データの表示周波数範囲は任意に設定できる

修正機能 以下の 3 種類の方法で F0 データの修正ができる

- フリーハンドの修正. マウスによるクリック&ドラッグで F0 データを 1 個毎に任意の周波数へ移動させる (図 1 上段の F0 データ窓内).
- 波形データと連動した修正. 時間波形の窓 (図 1 中段) で, ある区間を指定すると, 対応する F0 (その区間長の逆数) が計算され, 即時に F0 データの表示に反映される. 区間の始点, 終点を波形のピーク, デイップ, あるいは零点に精密に一致させることもできる.
- 数値入力を使った修正. 1 個あるいは連続した複数のデータを選択し, それらに対して簡単な算術演算による修正を加えることができる. 任意の 2 点間のデータの線形補間, 対数補間もできる.

その他の特徴

- 多くのデータ形式に対応している. F0 データ, 波形データは, 'short', 'float', 'double', 'ascii' の各形式に対応している. 'short' の場合はバイト逆順で読むことも可能である. さらに, F0 データは周波数 (Hz) 以外に, 周期 (時間) として読み込むことも可能である.
- 歌声や楽器音も想定しているので, 扱えるデータ長が長い (例: 10 ms 間隔で 10 分程度. ESPS & waves+ では, Zoom-in/out を行ないながら修正する場合, 10 ms 間隔で 10 秒程度が限界)
- 歌声や楽器音も想定しているので, 扱える F0 の範囲が広い (0Hz ~ 5kHz)

3.2 主な仕様

表示できるデータ: F0, 波形, 音韻ラベルの内の 1 つあるいは任意の組み合わせ

修正できるデータ: F0

入出力フォーマット:

波形, F0¹: short (byte swap も可能), float, double, ascii

音韻ラベル: ATR 音声データベース²に準拠したフォーマット

表示する F0 の範囲: 0Hz - 5kHz の任意の範囲 (デフォルト: 0 - 700 Hz)

サンプリング周波数: 0.1 - 100 kHz (デフォルト: 12 kHz)

F0 データの時間間隔: 0.01 - 1000 ms (デフォルト: 2.5 ms)

¹F0 データの単位は周波数の場合は Hz, 周期の場合はサンプル数 (= サンプリング周波数の逆数)

²例えば以下を参照: 武田, 匂坂, 片桐, 桑原, "音声データベース構築のための視察に基づく音韻ラベリング," ATR テクニカルレポート, TR-I-0019/TR-A-0019, 1988.

4 現行版の問題と今後の課題

4.1 現行版の問題

当初予定していた機能はほぼ盛り込むことができたが、満足のいかない部分も多少残った。その多くはスピード面での不十分さなので、計算機の高速化に伴い解消されるものと思われる。その他では以下のような点を改善したい。

- 起動時のデフォルト値が固定である（ユーザ毎に変更できた方が良い）
- 起動後に周波数軸の表示範囲変更ができない
- 修正結果が全体表示窓には反映されない

一方、以下のような機能は、あれば便利だが少なくとも本ツールが開発された時点でのATR人間情報通信研究所の環境では他の手段によって代替可能なものであったので、システムの簡素化を優先し、あえて組み込むことは避けた。

- 波形のD/A……例えば波形窓をクリックするとその波形が音として再生される
- ラベルデータの編集……音韻ラベリングの作業を並行して行なう
- オプションデータの並列表示……例えば波形のパワー、有声／無声の情報などを同一時間軸で参照しながら作業を行なう
- F0の自動抽出……与えられたF0を修正するだけでなく、入力波形から自動的にF0を抽出する

4.2 今後の課題

「はじめに」で述べたようにF0の修正はそれ単独で完結する類の作業ではなく、多くの場合、修正後のデータを基に信号を再合成し、修正の知覚的（聴感上の）効果を確認する作業とリンクする。修正-再合成のサイクルにかかる時間が作業全体の効率を決定する要因であると云ってもよい。しかし再合成では、特に音声合成の場合など、F0以外にも多くのパラメータの制御を必要とするので、知覚的な確認作業は大きな負担を伴いがちである。そこで、手軽にF0を“試聴”できるような機能を備えれば作業の能率は向上するだろう。具体的には、修正作業を続けながら簡単な波形（純音／鋸歯状波／母音など）で修正したF0を再生する機能を想定している。

また、本ツールが修正の対象として想定しているF0データは、自動抽出アルゴリズムなどにより基本的には正しいものが計算されており、時々おかしいものが紛れ込んでいる、という比較的cleanなデータである。正しいものの割合の方が少ないようなnoisyなデータを相手にする場合には修正というよりも“作成”の作業になり、補間機能として現バージョンが持つ算術的な補間だけでは不十分であろう。F0生成のモデル（輪郭のモデル（臨界制動モデル etc.）とゆらぎ成分のモデル（線形予測モデル etc.））による補間機能を備えれば修正対象の可能性は広がる。

第 II 部

マニュアル編

5 環境設定

1. プログラムへのパスを確認してください。

```
% which xdpith
/usr/hearing/bin/xdpitch
```

上記のようにパスが表示されなければ、パスを設定してください。

```
% setenv PATH "${PATH}:/usr/hearing/bin"
```

2. 環境変数 XAPPLRESDIR で定義されたディレクトリに resource ファイルをコピーしてください。XAPPLRESDIR が設定されていなければ、設定してください。³ 必要な resource ファイルは、 /usr/hearing/xdpitch/Resource/XDpitch です。

```
例：自分のホームディレクトリ ($HOME) 直下に作る場合
% mkdir $HOME/Resource
% setenv XAPPLRESDIR $HOME/Resource
% cp /usr/hearing/xdpitch/Resource/XDpitch $HOME/Resource
```

6 プログラムの起動

プログラム⁴は、 xdpitch options で起動します。ファイルの指定をせずに起動すると、図 2 のウィンドウが表示され、ファイル名の入力を促します。

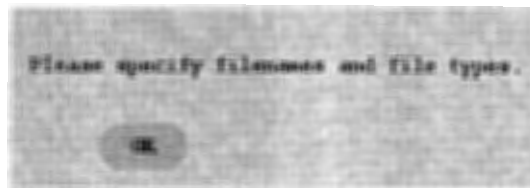


図 2: メッセージウィンドウ

6.1 options

プログラムのオプションは表 1 のとおりです。

³XAPPLRESDIR が設定されているかどうかは、 env | grep XAPPLRESDIR で確認出来ます。

⁴バージョン 1.0

-in	<i>file name</i>	<i>type</i>
-wave or -w	<i>file name</i>	<i>type</i>
-label or -l	<i>file name</i>	
-out or -o	<i>file name</i>	
-period or -p	<i>decimal (ms)</i>	
-sf or -s	<i>decimal (kHz)</i>	
-max or -x	<i>decimal (Hz)</i>	
-min or -n	<i>decimal (Hz)</i>	
-sp		
-sw		
-pp		
-help or -h		

表 1: オプション

6.1.1 -in, -wave, -label, -out

`-in (-wave, -label, or -out) file-name [format]`

これらのオプションを指定するには、ファイル名の指定が必要です。-in と -wave ではさらにデータ形式の指定が可能です。指定が可能なデータ形式は、short, float, double, ascii のいずれかです。

-in は、修正したい基本周波数 (F0) データのファイル名を指定します。デフォルトのデータ形式は short 型のバイナリデータですので、short 型のデータを使用する場合は、データ形式の指定は省略できます。

たとえばファイル foo.pit(short 型, binary データ) を指定する場合は、-in foo.pit になり、ファイル foo2.pit(float 型, binary データ) を指定する場合は、-in foo2.pit float になります。

-wave(or -w) は、時間波形データのファイル名を指定します。指定方法とデフォルトのデータ形式については F0 データと同じです。

-label(or -l) は、音韻ラベルのファイル名を指定します。データ形式は、ascii のみです。

-out(or -o) は、修正後の F0 データを出力するファイル名を指定します。データ形式は F0 データの入力ファイルと同一のものが選択されます。

6.1.2 -period, -sf, -max, -min

`-period (-sf, -max, or -min) number`

これらのオプションを指定するには、数値の引数が必要です。

-period(or -p) は、F0 データの時間間隔を指定します。デフォルトは 2.5ms です。

-sf(or -s) は、時間波形のサンプリング周波数を指定します。デフォルトは 12kHz です。

-max(or -x) は、F0 を何 Hz まで表示するかを指定します。デフォルトは 700Hz です。

-min(or -n) は、F0 を何 Hz から表示するかを指定します。デフォルトは 0Hz です。

6.1.3 -sp, -sw

-sp or -sw

-sp (Swap-byte Pitch) は、F0 のデータ形式が short 型のときに限り有効なオプションで、バイト逆順で F0 データを読み込みます。

-sw (Swap-byte Wave) は、時間波形のデータ形式が short 型のときに限り有効なオプションで、バイト逆順で時間波形データを読み込みます。

6.1.4 -pp

-pp

-pp (Pitch-data by Period) は、F0 データが、周波数 (時間の逆数) ではなく周期 (時間) で与えられる場合に指定します。単位は対応する時間波形のサンプル数です。

6.1.5 -help

-help or -h

簡単な使い方を表示して終了します。

7 プログラムの終了

11ページの 8.2.1 のとおり、処理をしてください。

8 処理の進め方

8.1 画面構成と表示範囲の変更

必要な事項を指定してプログラムを起動すると、ウィンドウが開き、図 3 のような画面が表示されます。中央上段に入力された F0 データ (の初めの部分)、その下に対応する時間波形が表示されます。最下段には、入力された F0 データ全体の概形が表示されます。

8.1.1 on-map データ

現在バッファ上にロードされているデータが on-map データです。修正は on-map データに対してのみ行うことができます。入力データのサイズが大きい場合、一部のデータしかバッファ上にロードできないことがあります。on-map データの範囲は、全体表示のウィンドウ上でブルーの部分として示されます。

on-map データの範囲を変更するには次の 2 通りの方法があります

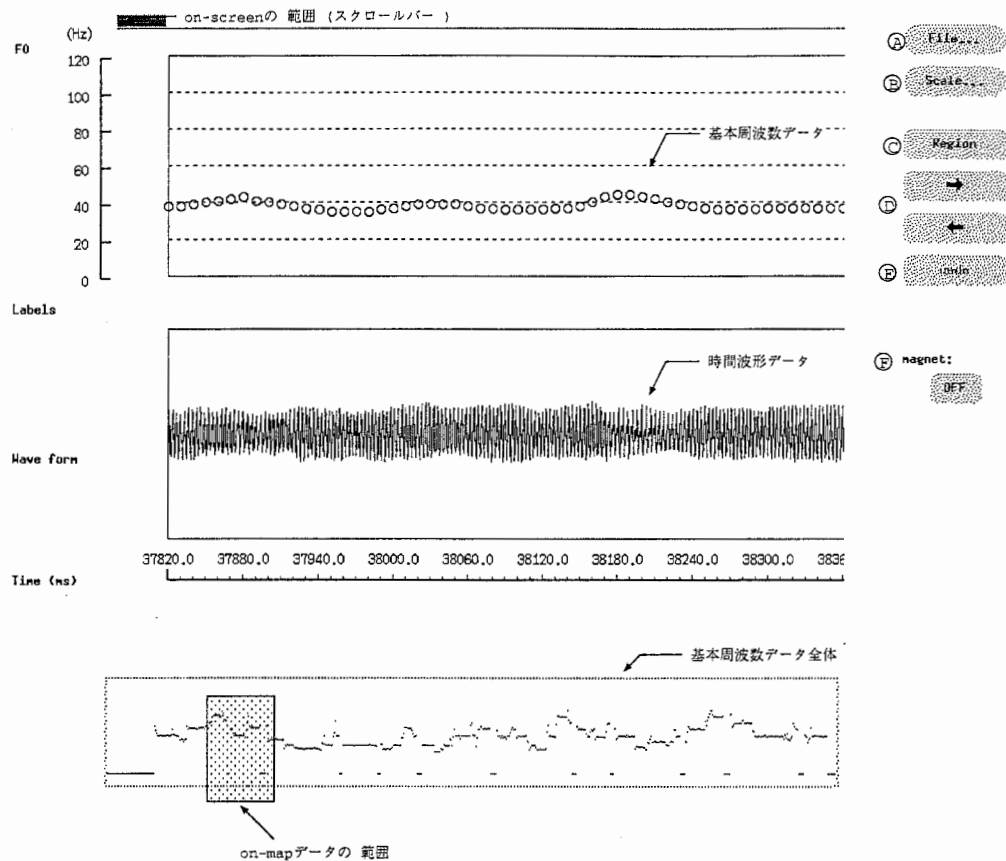


図 3: 初期画面

- F0データの全体表示のウィンドウ上でロードしたい部分の開始時点をクリックする方法
- ←または→のボタンを使って直前または直後の範囲をロードする方法

8.1.2 on-screen データ

現在表示されている on-map データが on-screen データです。on-screen データの on-map データ全体に対する範囲は、水平スクロールバーのグレーの部分として表示されます。

on-screen データの範囲を変更するには水平スクロールバーを使用します。

8.2 Buttons

8.2.1 File...

このボタン(図3のA)を押すとさらにウィンドウが開き、ファイルの Load, Save, プログラムの Quit という処理に進むことができます。

Load は、現在のデータセットをプログラムから消去し、新しいデータセットをロードします。Save は、修正されたデータをセーブします。Load または Save を選択すると、別のウィンドウが開き、ファイル名やデータ形式の入力を受け付けます。

Quit を選択すると、プログラムを終了します。Load または Quit が選択された際にセーブ済みでないデータが残っている場合には、別のウィンドウが開きセーブするかどうかの確認が求められます。

8.2.2 Scale...

時間軸の拡大／縮小を行ないます。このボタン (図 3 の B) を押すとさらにウィンドウが開き、ZoomIn, ZoomOut のどちらかを選択することができます。

ZoomIn はデータの表示区間を半分に、ZoomOut は 2 倍にします。

8.2.3 Region...

このボタン (図 3 の C) は複数個の F0 データを一括して処理する場合に使います。詳細は次節で説明します。

8.2.4 Undo

このボタン (図 3 の E) は、直前に行なわれた修正を無効にします。

8.2.5 Magnet

このボタン (図 3 の F) は、時間波形データの表示領域から基本周波数データを修正する場合に使用します (8.3.2 参照)。時間波形上の特定の特徴 (ピーク, ディップ, ゼロ交叉) を探索し、マウスで選択された区間の開始／終了を探索によって求められた時点へ吸着させることができます。これを Magnet 機能と呼びます。ボタンを押すごとに OFF → PEAK → DIP → ZERO → OFF のように切り替わります。

- OFF

Magnet 機能を off にします。

時間波形データの表示領域においてマウス (左ボタン) が押された座標に該当する時間波形データを区間の開始時点とし、マウスが放された座標に該当する時間波形データを区間の終了時点とします。

- PEAK

時間波形データの表示領域においてマウス (左ボタン) が押された座標に該当する時間波形データから前または後ろに波形データのピークを探し、区間の開始時点とします。同じくマウスが放された座標に該当する時間波形データから前または後ろにデータのピークを探し、区間の終了時点とします。

- DIP

時間波形データの表示領域においてマウス (左ボタン) が押された座標に該当する時間波形データから前または後ろに波形データのディップを探し、区間の開始時点とします。同じくマウスが放された座標に該当する時間波形データから前または後ろにデータのディップを探し、区間の終了時点とします。

- ZERO

時間波形データの表示領域においてマウス (左ボタン) が押された座標に該当する時間波形データから前または後ろに波形データのゼロ交叉点を探し、区間の開始時点とします。同じくマウスが放された座標に該当する時間波形データから前または後ろにデータのゼロ交叉点を探し、区間の終了時点とします。

8.3 F0 データの修正

以下の3通りの方法でF0データの修正を行なうことができます。

8.3.1 F0 データ表示領域から

個々のF0データは○で表現されています (図3参照)。マウスを使ってこれらを直接動かすことができます。マウスの左ボタンを押しながら修正したいデータを周波数軸上で移動させ、好みの場所で放します。

8.3.2 時間波形データ表示領域から

マウスで時間波形上の区間を指定すると対応する周波数が計算され、区間の開始時点に最も近いF0データが修正されます。区間はマウスの左ボタンを押しながら指定します。範囲の指定は、左から右へ向かって行ってください。

8.3.3 Region 指定によって

マウスによりF0データの範囲を指定して、複数個のデータをまとめて修正します。

Regionのボタン (図3のC) を押すとカーソルの形状が変化するので、マウスの左ボタンを押しながら範囲を指定します。範囲の指定は、左上から右下へ向かって行ってください。ウィンドウが開き、処理の選択を求められます (図4参照)。constant, add, multiply を選択した場合は、その後ウィンドウが開くので、数値を入力します (図5参照)。

それぞれの処理は、以下のような計算式を実行します。ここで、 n は指定された範囲に含まれるデータ数を、 $F0_i$ は指定された範囲の i 番目のデータの値を示します。左辺値が修正後のF0です。右辺のF0には修正前の値が代入されます。 a は使用者が入力した数値です。

1. constant : 一定値に揃える

$$F0_i = a$$

2. add : 定数の加算

$$F0_i = F0_i + a$$

(a は負でも可)

3. multiply : 定数の乗算

$$F0_i = F0_i \times a$$

(a は1以下でも可)

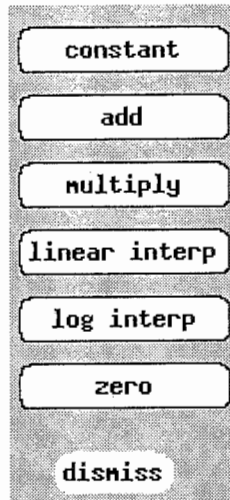


図 4: Region によるデータ修正

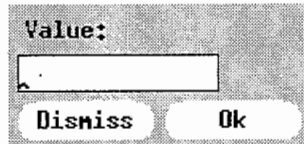


図 5: 数値の入力

4. linear interp : 指定範囲の始点 ($F0_1$) と終点 ($F0_n$) を直線補間

$$F0_i = F0_1 + (F0_n - F0_1) \left(\frac{i-1}{n-1} \right)$$

5. log interp : 指定範囲の始点 ($F0_1$) と終点 ($F0_n$) を対数補間

$$F0_i = F0_1 \times \left(\frac{F0_n}{F0_1} \right)^{\frac{i-1}{n-1}}$$

6. zero : ゼロに揃える⁵

$$F0_i = 0$$

⁵constant メニューで $a = 0$ を入力するのと等価ですが, ゼロ化は頻度が高いと思われるので別メニューを用意しました.