

# 琉球大学学術リポジトリ

## 即時応答型音程トレーナーの為の周波数解析手法について

|       |   |
|-------|---|
| メタデータ | 言語:<br>出版者: 琉球大学<br>公開日: 2007-08-08<br>キーワード (Ja):<br>キーワード (En):<br>作成者: 仲間, 正浩, Nakama, Masahiro<br>メールアドレス:<br>所属: |
| URL   | <a href="http://hdl.handle.net/20.500.12000/1374">http://hdl.handle.net/20.500.12000/1374</a>                         |

# 即時応答型音程トレーナーの為の 周波数解析手法について

仲間正浩

Real-Time Response Pitch Discrimination Training System :  
A Pitch Computing Method

Masahiro NAKAMA\*  
(Received April 28, 1994)

Abstract

It is meaningful to develop a Real-Time Response Pitch Discrimination System by utilizing a computing sound pitch. There are many approaches in computing sound pitch. But, no approach was seen in developing enough response-time and accuracy of frequency discrimination for pitch discrimination training. Therefore new methodology is considered to improve response-time and accuracy of frequency discrimination. In FFT (Fast Fourier Transformation), there exist mathematical relations such that

(1)  $\text{time\_range\_of\_analysis} \cdot \text{frequency\_discrimination} = 1$  and

(2)  $\text{response\_time} = \text{time\_range\_of\_analysis} / 2 \cdot \text{computing\_time}$ ,

which shows no improvement in both response-time and accuracy of frequency discrimination. Fortunately, there exist the fact such that many peaks of power spectrum :  $f$  [Hz],  $f \cdot 2$  [Hz],  $f \cdot 3$  [Hz], …… as the results of FFT from a single  $f$  [Hz] pitch sound.

To improve both response-time and accuracy of frequency discrimination, the following new method is considered in this paper. Step1 : Compute pitch roughly from  $f$  [Hz] power spectrum peak. Step2 : Compute pitch accurately from  $f \cdot n$  [Hz].

## 1. はじめに

コンピュータの音楽への応用について、様々な試みがなされてきた。しかし、これまで行われてきた研究開発は、そのほとんどがコンピュータそのものを楽器として取り扱うもの（シンセサイザー）、作曲を支援するもの（楽譜の作成支援）、自動演奏（ロボットで演奏するなど）等である。また、音楽教育に関するものでは、リズム、楽典や、曲の知識等に関するものがほとんどであり、音楽の実践や教育の際に最も多くの時間を要する

“練習”を支援するようなコンピュータソフトは、これまで、ほとんど開発されていない。

本研究では、一人でも練習できる音程トレーナーのパーソナルコンピュータ上での実現を試みている。音程トレーナーをパーソナルコンピュータ上に実現する場合、効率的な練習を可能にするためには、十分な周波数精度と、練習者のリズム感を崩さない程度の応答時間を持つようなシステム化をすることが望まれる。以下では、この条件を満たすようなシステムを実現する為の周波数解析手法の考察について報告する。

---

\*Department of Computer Science, College of Education, University of the Ryukyus, Nishihara, Okinawa 903-01, Japan.

## 2. 本研究で制作する音程トレーナーの基本的要件

効率的な練習を可能とするような音程トレーナーシステムを実現する為に考慮すべき要件を挙げると以下の通りとなる。

- (1) トレーニングする際に扱う楽器や声の音域を扱えること
- (2) 十分な音程精度を持つこと
- (3) 十分に短い応答時間であること  
(リズム感を崩さない程度に)
- (4) 効率的にトレーニングできるコースウェアが設定されていること
- (5) どのような音色の音でも正しく音程の認識ができること

楽器には様々なものがあるが、これらは、個々にその音域が異なる。例えば、バスチューバの音域は41.20～261.63Hz、ピッコロの音域は440～3729.31Hzであり<sup>1)</sup>、これらの周波数範囲を扱える様にすることで(1)の要件が満たされる。

ある2つの音  $f_1$ 、 $f_2$  を発生させた場合、その音程は、通常、以下の式で表現される<sup>2)</sup>。

音程 =  $300 / \log 2 * \log (f_1 / f_2)$  [サヴァール (savart)] …… (1)

平均律の半音が25 [サヴァール]、全音が50 [サヴァール]、純正のオクターヴが300 [サヴァール]と定義されている。従って、ある一定の音程に対

して考えれば、高音程の領域では、その周波数差は、大きくなり、逆に低音程領域では、周波数差は、小さくなる。この関係を、各音程ごとに示したのが図1である。

楽器演奏前のチューニングは、通常440Hzの周波数を基本にしてうなり(うなりの回数 =  $|f_1 - f_2|$ ) を聞くことによって行われる。この時、通常少なくとも  $1/2$  Hz (うなりの回数が2秒間に一回) 以下の周波数精度で行われると考えられるので、全音の  $1/128$  即ち0.39 [サヴァール] 以下の音程精度を持たなければ(2)の要件を満足することはできないであろう。

合奏や合唱を行う際きれいなハーモニーを奏する為には、曲の演奏中に自分の発している音と周りから発せられる音の音程の違いを聞き分け、瞬時に音程の修正を行う能力を養成しておく必要がある。従って、我々の作成するシステムは、(3)を満足させることによって、十分にこの要求を満足させることができるようにしておく必要がある。

音程練習の方法は、まず、音程のずれを認識して、それを修正する能力を身に付け、次いで、協和音になる音の音程を覚え、更に、音階を覚えていくという様に行うのが効果的である<sup>3)</sup>。従って、解説や、トレーニング方法、練習曲などを練習者の能力に応じて段階的にトレーニングできるようなコースウェアの設定が不可欠である。

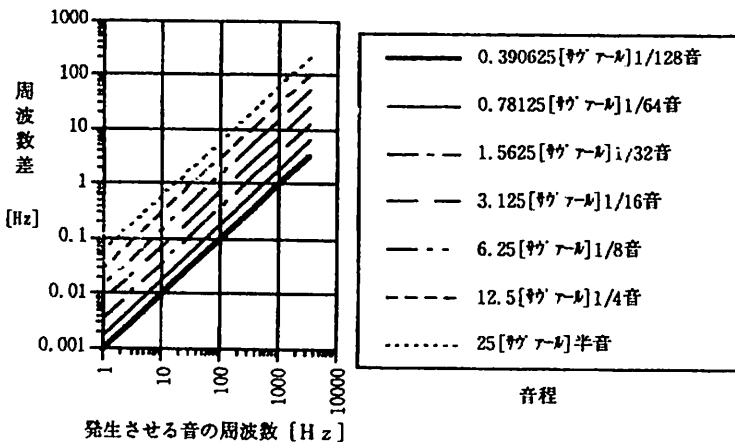


図1 音程と周波数差との対応

### 3. 音程トレーナーのパーソナルコンピュータ上での実現

本研究で実現する機械のモデルを図示すると、図2の通りとなる。コンピュータは、ある一定の高さを持った音の発生や練習曲の演奏を行う。練習者はこれらの音と同一音高の音をコンピュータに接続されたマイクに送り込む。コンピュータは、

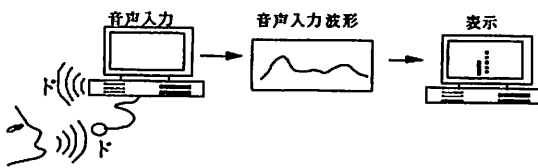


図2 音程トレーナーのイメージ

マイクから入ってきたアナログ音声データをA/D変換し、そのデータの周波数を計算し、マイクから取り込まれた音の音高と音程のずれを画面上に表示する。練習者は、画面上に表示された音程のずれを見ながら、自分の発生する音を正しい音高に修正する。この機械を使用することによって音程感の不確かな初心者でも簡単に視覚的に音程のずれを認識することができるので、効果的な練習が可能となると考えている。

### 4. 音程トレーニングのための周波数分析機の実現

本研究でこれまでに試作した周波数分析機の動作の概要と周波数計算方法をそれぞれ、図3、図4に示す。この方法は、練習に使用できる音源を限定してしまう（きれいな正弦波に近い音色の音しか扱えない）という欠点を持つが音程認識能力の訓練には十分に使用可能である。

この図に示した方法を説明すると次の通りである。音声データを一定の時間幅（サンプリング間隔 \* サンプル数）で取得し、次いで周波数の計算とその結果の表示をする。コンピュータは、この動作をある一定の時間間隔（=周波数の計算及び表示時間）で繰り返す（図3）。周波数の計算方法は、取得された音声データの平均値の軸を上から下に何回交差したかを数えることによって、発生された音の周期を調べると同時に、一番最初の交差点から最後の交差点まで、どれくらいの時間の経過があったかをサンプル数を数えることによって行う（図4）。このとき、求められる周波数  $f$  は、

$$f = 1 / (\text{サンプル数} * \text{サンプリング間隔} * \text{周波数}) \dots\dots (2)$$

となる。また、データの入力から表示完了までの平均遅延時間は、

$$\text{遅延時間} = T / 2 + T_c \dots\dots (3)$$

となる。周波数の計算は、上記(2)式で示した一回の計算時間で済むので、

$$T_c \propto T$$

の関係となり、結局、(3)式は

$$\text{遅延時間} \propto T \dots\dots (4)$$

となる。

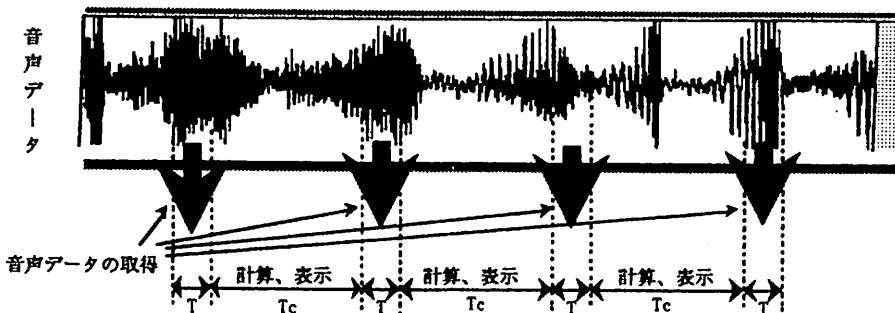


図3 音声データの取得、計算、表示のタイミング

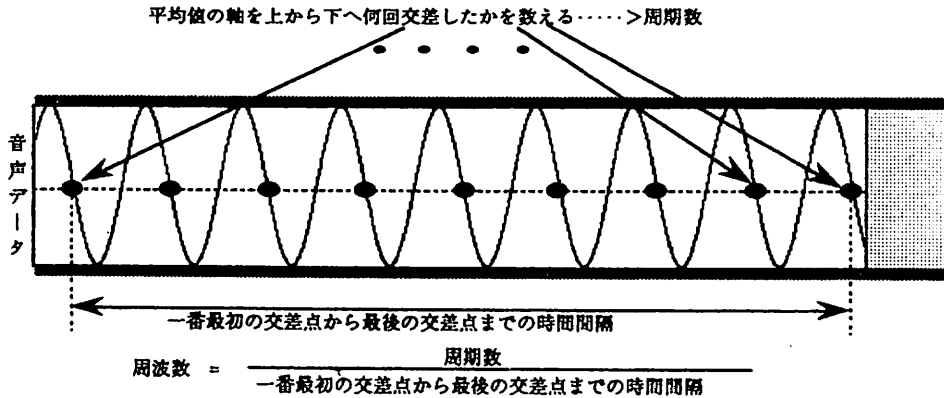


図4 周波数の計算方法

### 5. 音程誤差を考慮した各種パラメータの決定

本研究で制作するシステムは、遅延時間、音程認識誤差ともに小さい程良い。しかし、遅延時間を小さくするには一回の解析に使用するデータの区間Tを小さくすればよいのだが、残念ながら、この遅延時間とデータの区間Tは、反比例の関係にある。また、扱える周波数の範囲は、できるだけ広いほうが望ましいが、周波数の上限を高くすれば、計算するサンプル数を増やすことになり、周波数の下限を低くすれば、計算の対象となるデータの時間幅を広くする必要が出てきて、結果的に、遅延時間を増やすことになってしまう。結局、システムを実現する上で、各種のパラメータをどのように設定するのが、非常に重要な問題になってくる。以下では、パラメータ設定のための分析結果について述べる。

扱える周波数の範囲は、標準化定理および、上で示した計算の方法より、次の式のとおりとなる。

$$\text{扱える周波数の上限} = f_s / 2 \dots\dots (5)$$

$$\text{扱える周波数の下限} = 1 / T \dots\dots (6)$$

一方、先に示した計算の方式を考慮して、音程誤差、音声周波数f、サンプリング周波数 $f_s$ 解析区間Tの関係を計算すると、次式が得られる。

$$\text{音程誤差} = (300 / \log 2) \log (1 + 1 / ((f_s + f)T)) \dots\dots (7)$$

楽器で取り扱う周波数の範囲は、27.5~4186.0Hzであるので、(5)、および(6)式より、

$$f_s > 2 * 4186.0 = 8372 \text{Hz} \dots\dots (8)$$

$T > 0.037 \text{s} \dots\dots (9)$ の結果を満足するようにして、Tを仮に $T=0.05$ として、(7)式より、幾つかのサンプリング周波数にたいして音程誤差fとの関係をグラフ化すると、図5の通りとなる。このグラフより、サンプリング周波数を22kHz以上にすれば、(8)式も満足することになる。

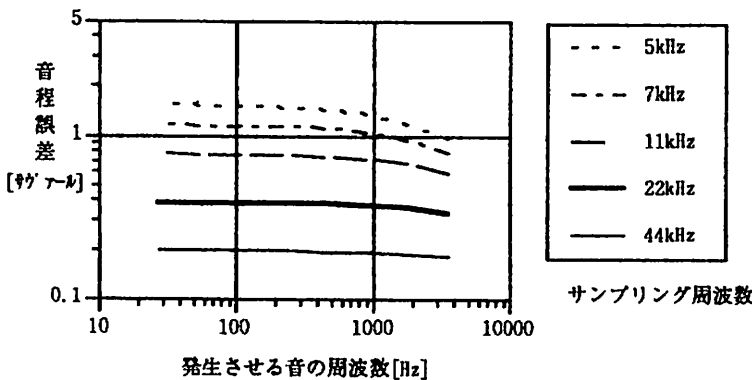


図5 各サンプリング周波数ごとの周波数-音程誤差関係 (T=0.05 [s])

先に述べた方式のソフトウェアを上で決定したパラメータ ( $T=0.05$ 、 $f_s=22\text{kHz}$ ) で周波数計算の結果を示すだけの機能に絞っての試作を行った。これをパソコン (Macintosh IIVX : CPU68030、FPU68882 32MHz) 上で実行させた結果、平均応答時間0.175秒と、十分実用に耐えうるシステムが構築可能なことが確認できた。

## 6. 複雑な音声波形に対応する周波数解析方法

上で述べた周波数解析方法で、十分な平均応答時間をもつシステムを実現可能なことが確認できた。しかし、この方法では、波形が複雑になれば正確な音程の認識が出来なくなるという欠点が残っている。周波数を正確に計算するだけであれば、FFTが有効であると考えられるが、これを採用すれば、

周波数精度 \* 解析区間 = 1

となり、応答時間、周波数精度のいずれかを犠牲にすることになる。短い解析区間のデータでも細かく周波数を計算する方法としては最大エン

ロピー法による方法がある。これは、現代我々が使用しているパソコンでは、計算に時間がかかりすぎ、結果的に応答時間を犠牲にしまうおそれがあるので、この方法を採用することはできない。

では、どのような手法であれば、複雑な波形を持つ音声でも (どのような音色の音でも) 扱え、かつ上で述べたような要求を十分に満たす性能を持つ周波数解析機が実現できるのであろうか。以下では、これまで検討した複雑な波形を持つ音色を扱える音程トレーナーを実現するための平均応答時間の短縮化の手法について述べる。

### 6.1. 周波数解析の並列化による処理時間の短縮化

上で述べた周波数解析の方法は、全ての処理を逐次型に処理する方法であった。それを図示すると図6の通りとなる。この処理の流れは、

- (1) データの取得
- (2) 計算処理
- (3) 表示処理

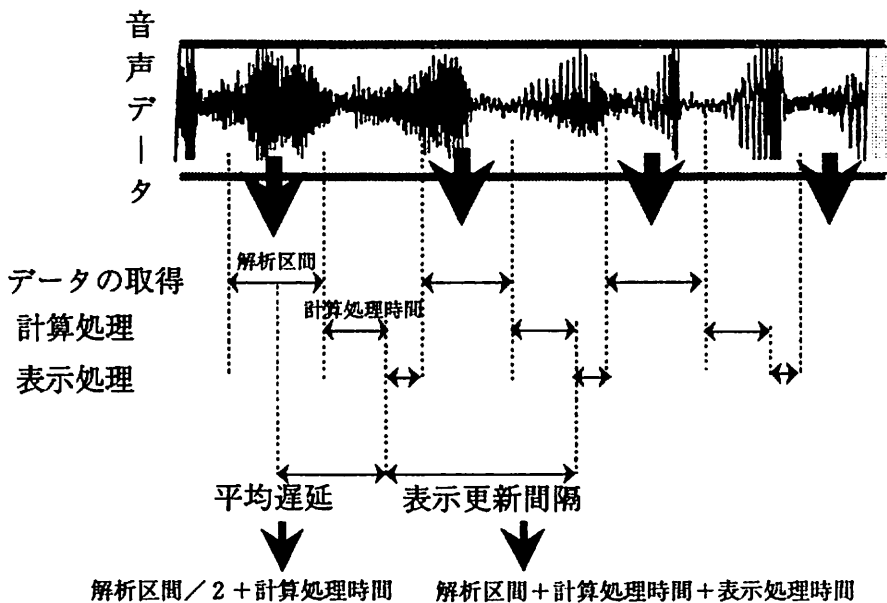


図6 逐次型の周波数解析処理

となる。この処理では平均遅延時間（マイクから入力されたデータが画面に表示されるまでの時間）、表示更新間隔（表示が新しく更新されるまでにかかる時間）はそれぞれ、

$$\begin{aligned} \text{平均遅延時間} &= \text{解析区間} / 2 + \text{計算処理時間} \\ \text{表示更新間隔} &= \text{解析区間} + \text{計算処理時間} + \text{表示処理時間} \end{aligned}$$

である。まず、最初に考えられる処理時間短縮の方法は、データの取得の処理を並列化する事であるが、これを図示すると、図7の通りとなる。この方法での平均遅延時間と表示更新間隔はそれぞれ、

$$\begin{aligned} \text{平均遅延時間} &= \text{解析区間} / 2 + \text{計算処理時間} \\ \text{表示更新間隔} &= \text{計算処理時間} + \text{表示処理時間} \end{aligned}$$

となる。次に考えられるのは、個々の計算処理を

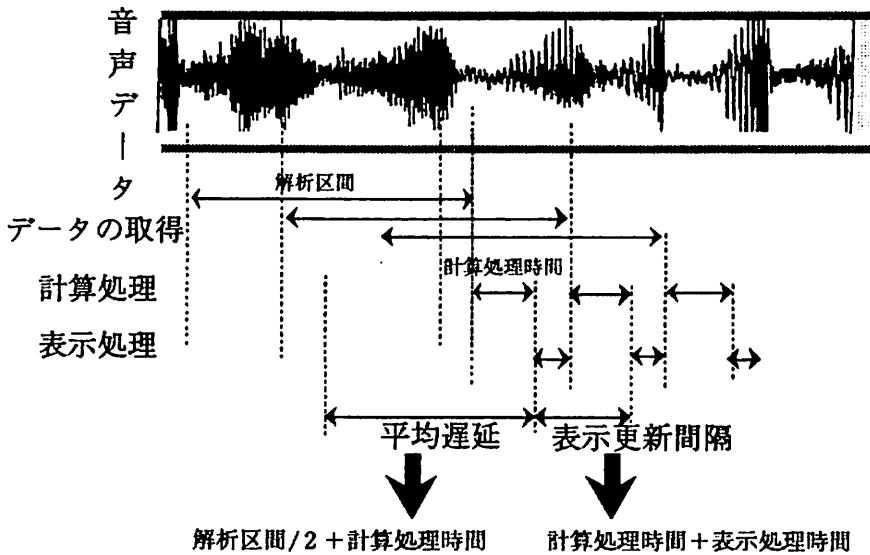


図7 データの取得と計算表示処理を並列化した場合の周波数解析処理

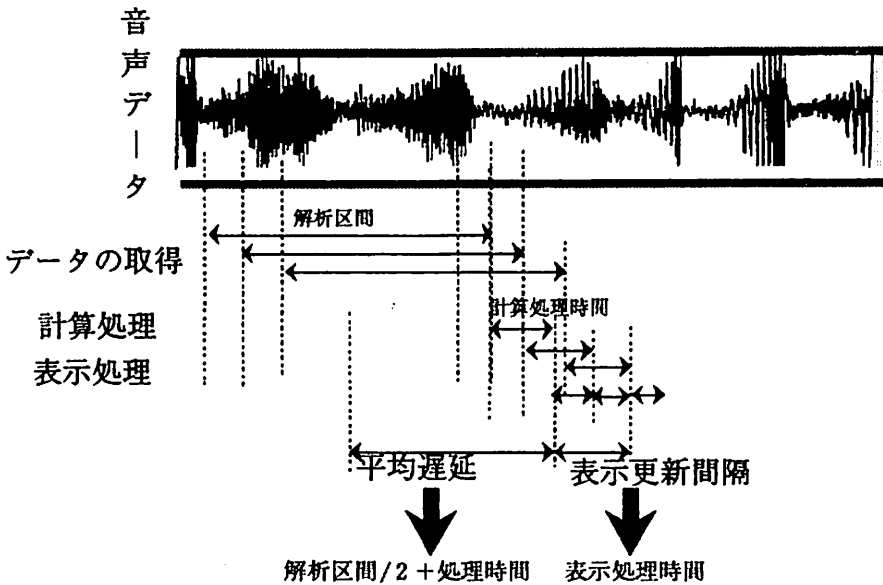


図8 個々の計算処理を並列化した場合の周波数解析処理

並列化する事であり、これを図示すると、図8の通りとなる。この方法での平均遅延時間と表示更新間隔はそれぞれ、

平均遅延時間＝解析区間／2＋計算処理時間

表示更新間隔＝表示処理時間

となる。従って、それぞれの処理を徹底的に並列化すれば、表示時間間隔を表示処理時間まで短くすることができる。しかし、データの取得から表示までの遅延時間は、

解析区間／2＋計算処理時間

となり、上に挙げたどの様な並列化を行っても全て同じ結果になる。この結果が示す意味を述べると、まず、表示更新間隔については、この時間が短いとコンピュータから表示される画面は、スムーズに流れるような動きになってくる。平均遅延時間は、マイクからデータが入力された後でその処理結果が画面表示に現れるまでにかかる時間である。この時間が長すぎるとかなり遅いテンポでの音程トレーニングでは、練習者が表示をみながら音程を直すのはかなり時間がかかってしまう。また、速いテンポでのトレーニングは事実上不可能になるばかりか、リズム感が悪くなってしまいうという副作用が起きてしまう。従って、音楽の練習を行うためのシステムを構築するにあたっては、この時間を短くしておくことはかなり重要な意味を持つことになる。

## 6.2. マイクから表示までの遅延時間を短くする手法

デジタルサンプリングされた音声データを解析する場合、解析区間内の数を  $n$  とすれば、線形予測法では  $O(n^2)$ 、フーリエ変換 (FFT) では  $O(n \log_2 n)$  の処理が、その計算に必要な。線形予測法では、周波数の精度は、任意に設定可能であるが、処理に時間がかかりすぎるため、この手法を採用すれば、平均遅延時間を短くすることは不可能となる。フーリエ変換では、周波数の精度は、

周波数精度＝1／解析区間

で決まるが、

平均遅延時間＝解析区間／2＋計算処理時間

であるので、周波数精度を良くしていけば、平均遅延時間が長くなり、逆に、平均遅延時間を短くすると周波数精度が悪くなり、その両方を良くすることは出来ない。従って、何らかの工夫が必要になってくる。

ここで述べた、フーリエ変換、線形予測法などの音声波形からパワースペクトルへの変換を行うと、その結果、図9に示す様なパワースペクトルが得られる。音声波の性質をもう少し細かく見ると、周波数変換を行う前の音声波形が、単一音高を持つものでありその波形が、きれいな正弦波

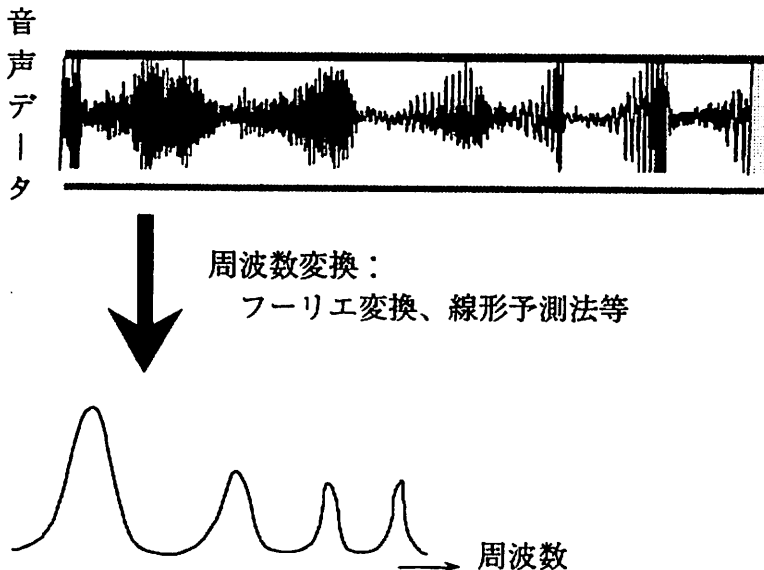


図9 音声波形からパワースペクトルへの変換



でなければその周波数変換を行った後に得られるパワースペクトルは、入力された音の基本周波数の他に、その周波数の整数倍の周波数の領域で極大値が存在する図10。例えば、2 Hz精度（解析区間0.5s）のフーリエ変換を行った場合、50Hzの音が入力されれば、その周波数領域においては半音程の精度の解析結果しか得られないが、50Hzの21倍の領域（約1000Hz付近）では、1/64音程の精度での解析結果が得られることになる（図1参照）。従って、大まかな周波数判定を基底周波数領域で行い、細かい判定は、倍音領域で行えば、平均遅延時間を短縮でき、音程精度を良くすることが可能になるはずである。

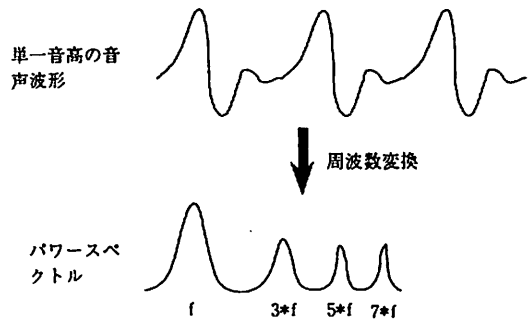


図10 音声波の性質

## 7. むすび

これまでの検討の結果、練習者がマイクから入力した音声データを、波形の符号が変わる点の時間間隔（＝周期）より、周波数を計算して音程を導き出すというかなり簡単な方法を採用することによって、十分な周波数精度と応答時間を持った音程トレーニングシステムを実現可能なことを確認した。しかし、この方法では、波形が、複雑になれば、正確な音程の認識ができなくなるという欠点がまだ残っている。よって、この欠点を克服する方法として、音声データの取得と周波数解析処理の並列化を行うことによって、コンピュータ側の表示処理の周期を短くして、画面上の音程の提示が滑らかに動くようにする事が可能であることを示した。また、周波数解析の方法として、フーリエ変換（FFT）で大まかな周波数判定を基底周波数領域で行い、細かい判定は倍音領域で行うという方法をとることによって平均遅延時間を短縮でき、音程精度を良くすることができ可能性があることを明らかにした。今後の課題としては、ここで述べたシステム設計案の具体的な実現および、その有効性の確認が残されている。

## 謝 辞

日頃、本研究の検討に御協力頂いている本学科の学生各位に感謝の意を表する。

## 文 献

- (1) アレクサンダー・ウッド 石井信生：  
“音楽の物理”、音楽之友社、1976.
- (2) John R. Pierican：“The science of musical sound”, Science American Books, 1983.
- (3) シェンドル、羽仁協子（訳）：  
“ハンガリーの音楽教育”、音楽之友社、1968.
- (4) 電子情報通信学会：  
“ディジタル信号処理ハンドブック”、オーム社、1993.
- (5) 仲間正浩、新田保秀：  
“即時応答型音程トレーナーの構想と実現方式について”、信学技報、ET93-105、1993-12.