



## 自動採譜システムの構築

メタデータ	言語: jpn 出版者: 公開日: 2013-12-10 キーワード (Ja): キーワード (En): 作成者: 梅本, 敏孝, 矢野, みのり, 西田, 雅恵 メールアドレス: 所属:
URL	<a href="https://doi.org/10.24729/00007729">https://doi.org/10.24729/00007729</a>

## 自動採譜システムの構築

梅本 敏孝\*・矢野 みのり\*\*・西田 雅恵\*\*\*

Construction of Transcription System

Toshitaka UMEMOTO\*, Minori YANO\*\* and Masae NISHIDA\*\*\*

## ABSTRACT

In order to develop an automatic music transcription system, we have to find peak frequency and a length of each tone precisely with a computer to find the musical scale. The length of each tone is generally found from the starting timing of each tone extracting from data's energy in a short time. But it is difficult to make the scale correct, because the power of energy differs from the piece and the tone of music. The peak frequency of each tone is extracted from short-time spectra analysis, and FFT is used in general. But it is necessary for this method to make corrections of the frequency. In this study, we first proposed the new process in order to find the length of each sound using a standard beat that was determined by self-correlative function of short time energy. Then measures, bars, rhythm and keys were found by extending the discussion to the process for extracting the length of each tone. Furthermore, the effectiveness of the proposed method for the transcription system was experimentally demonstrated by applying it to the sung songs.

**Key Words:** FFT, spectrum, tone, transcription

## 1. はじめに

採譜をコンピュータを使って行うには、精度良く各音の持続時間（音価）を抽出し、それに基づいて音階を同定しなくてはならない。従来、音価は、振幅の変化を調べるのに有利な表現である短時間エネルギーを用いて、音の立ち上がりを調べることによって決定していた。しかし曲や音によっては短時間エネルギーの振幅がはっきりしない場合があり、精度良く音価を抽出することができなかった。また、音階の同定では、各音素の基本周波数がどの音階に属するかを同定しなくてはならない。この音階同定法としては井口らによって提案された高速フーリエ変換による複素スペクトル面での内挿法<sup>1),2)</sup>がある。この方法において有効に周波数を補正するには隣合う周波数サンプル点間に存在する周波数成分は一つである必要性があり、そのような状態になるように高速フーリエ変換の点数を長くしなくてはならない。このため、音素の基本周波数が低く、リズムの速い曲の場合、音階の同定には不向きである。また、この方法を改良した隣合う

周波数成分の位相差を利用する方法<sup>3)</sup>も提案されている。しかし、この方法でも音素の基本周波数が低く、リズムの速い曲では位相差は少ないと考えられるので有効でない場合が多い。そこで我々は、求まる周波数を自由に選ぶことができ、周波数補正を行わずに各音素がどの音階に属するかを同定することができる周波数分析法として、適応技術を用いた周波数分析法を提案した<sup>4),5)</sup>。また、求まる周波数を自由に選ぶことができる方法として青島が提案した複素1次系を用いた短時間スペクトル推定法<sup>6)</sup>がある。そこで、本研究では、基準拍というものをまず抽出し、その基準拍と、上述の周波数分析によって得られた音階データを利用した各音素の持続時間の抽出方法を提案し、求まった各音素の持続時間毎に、周波数分析法によって得られたデータから音階を同定し、その精度の比較を行う。さらにそれらのシステムを拡張し、拍子、小節、調子といった各パラメータの同定を行う方法を提案し、実験によってその有効性を確かめる。

## 2. 音階の同定に用いる周波数分析法

採譜システムを構築するのに最も重要な処理の一つに音階の同定がある。音階の同定では、各音素の基本周波数がどの音階に属するかを決定しなくてはならない。この章では音階を同定するための周波数分析法について説明を行なう。

1998年4月9日受理

\* システム制御工学科 (Department of Systems and Control Engineering)

\*\* ロート製薬株式会社 (Rohto Pharmaceutical Co., Ltd.)

\*\*\* 三共製薬株式会社 (Sankyo Pharmaceutical Co., Ltd.)

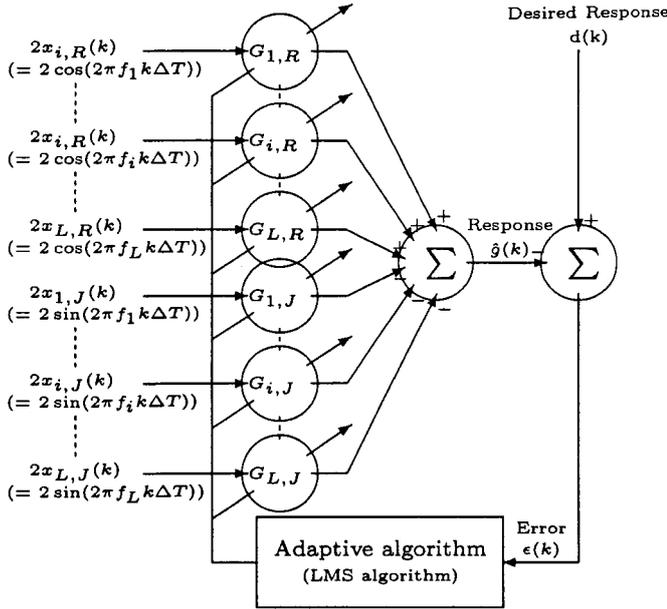


Fig. 1 Spectra analysis system with LMS algorithm.

2.1 高速フーリエ変換による複素スペクトル面での内挿法

この方法は、井口らによって提案されたものであり、以下に示すような方法によって音の周波数を求めるものである。まず、高速フーリエ変換によってパワースペクトル分布のピーク部を検出し、ピークを挟む二つの成分  $z_m, z_{m+1}$  から次に示すような単位ベクトルを求める。

$$u = \frac{z_{m+1} - z_m}{|z_{m+1} - z_m|} \quad (1)$$

$u$  と  $z_{m+1}, z_m$  の内積から音の周波数  $\hat{f}$  が次式により求められる方法である。

$$\hat{f} = m + \frac{(u, z_{m+1})}{(u, z_{m+1}) - (u, z_m)} \quad (2)$$

2.2 適応技術を用いた周波数分析法

この方法は適応フィルタを用いてその係数の収束性から周波数分析を行なう方法である。まず、フーリエ係数  $G(k)$  を重みベクトルと考え、次式で定義された  $X(k)$  をフィルタの入力すると Fig. 1 に示すようなシステムになる。

$$X(k) = [\exp(j2\pi f_1 kT), \dots, \exp(j2\pi f_i kT), \dots, \exp(j2\pi f_L kT)] \quad (3)$$

フーリエ変換を行ないたい信号とフィルタ出力との誤差を  $\epsilon(k)$  すれば、係数を更新する式は次のようになる。

$$G(k+1) = G(k) + 4\mu\epsilon(k) \cdot \bar{X}(k) \quad (4)$$

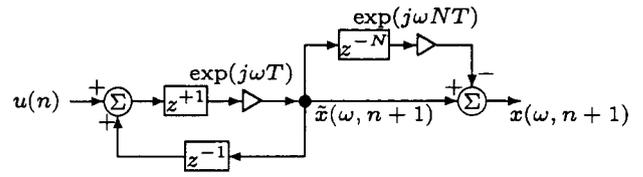


Fig. 2 Spectra analysis system with CFOS.

ここで  $\mu$  は、ステップサイズパラメータと呼ばれる正定値であり、文献 5) から適切に選定できる値である。この式は再帰式であり、適応を繰り返すことによって  $\epsilon(k)$  は 0 に近づき、結果として周波数分析が行なえる方法である。

2.3 複素 1 次系を用いた周波数分析法

複素一次系の応答  $x(n)$  は次式によって計算することができる。

$$x(n+1) = x(n) \exp\{(\sigma + j\omega)T\} + \frac{c_R + c_J}{\sigma + j\omega} u(n) [\exp\{(\sigma + j\omega)T\} - 1] \quad (5)$$

ここで  $T$  はサンプリング周期であり、 $x(n) = x(nT)$  である。初期値  $x(-0) = 0$  とし、 $n = 0$  で大きさ 1 のインパルスが加わり、それ以後入力  $u(n) = 0$  であるとする、 $n = d - 1$  における値は次のようになる。

$$x(d-1) = \frac{c_R + c_J}{\sigma + j\omega} [\exp\{(\sigma + j\omega)T\} - 1] \cdot \exp\{(d-1)(\sigma + j\omega)T\} \quad (6)$$

さらに  $n = d - 1$  において大きさ  $A$  (複素数) のインパルスが加わったとすると  $n = d$  における値は次のようになる。

$$x(d) = \frac{c_R + c_J}{\sigma + j\omega} [\exp\{(\sigma + j\omega)T\} - 1] \cdot \exp\{d(\sigma + j\omega)T\} + \frac{c_R + c_J}{\sigma + j\omega} A [\exp\{(\sigma + j\omega)T\} - 1] \quad (7)$$

ここで、

$$A = -\exp\{d(\sigma + j\omega)T\} \quad (8)$$

とすると、 $x(d)$  は 0 となり、入力がその後 0 であるならば  $x(n)$  は  $n \geq d$  で恒等的に 0 になる。そこで、Fig. 2 のシステムを考え、 $n = 0$  において  $u(n)$  としてインパルスを加えると、 $n = d$  以後は出力  $x(n)$  は 0 になる。 $\sigma = 0$  の場合、複素 1 次系のインパルス応答は正弦波になるから Fig. 2 にインパルスでない一般の波形を入力すると、ある瞬間の出力はその時刻の直前の長さ  $d$  の波形に正弦波を乗算し、積分したものになりフーリエ係数を

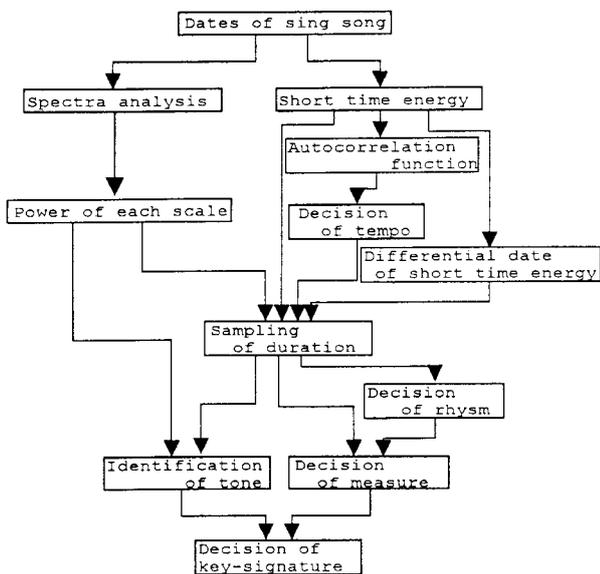


Fig. 3 Flow chart of the transcription system.

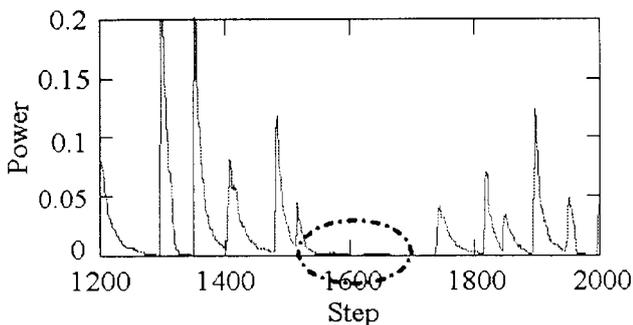


Fig. 4 Short time energy.

計算したものと等しくなる。このことを利用して Fig. 2 で示されるシステムを用いて短時間スペクトルを求める方法である。

### 3. 採譜処理の過程

採譜処理全体の流れを Fig. 3 に示す。以下に各過程について説明と歌唱「むすんでひらいて」を用いた実験結果を示す。

#### 3.1 各音素の持続時間抽出

上述の周波数分析を用いて精度の良い音階の同定を行うには、各音素の持続時間を精度良く抽出する必要がある。一般的に、音の終了時点というのは認識が困難であるので、音の立ち上がり時点から次の音の立ち上がり時点までを各音素の持続時間と考え、振幅の変化を調べるのに有利な表現である短時間エネルギーが用い、音の立ち上りを調べることによって各音素の持続時間を求めてられていた。この方法では、短時間エネルギーの差分データがしきい値以上の時に、音が立ち上がりがあると

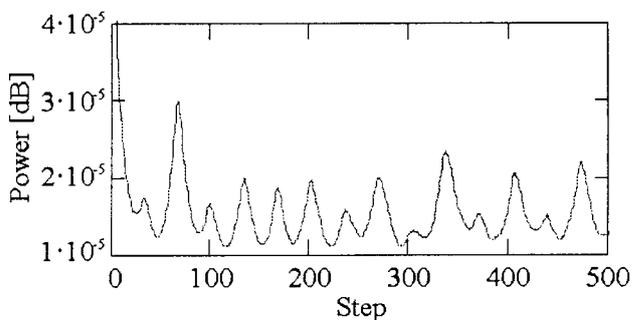


Fig. 5 Autocorrelation function.

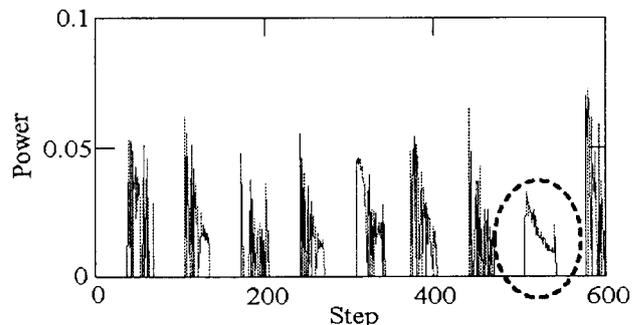


Fig. 6 Power spectrum of 'C4'.

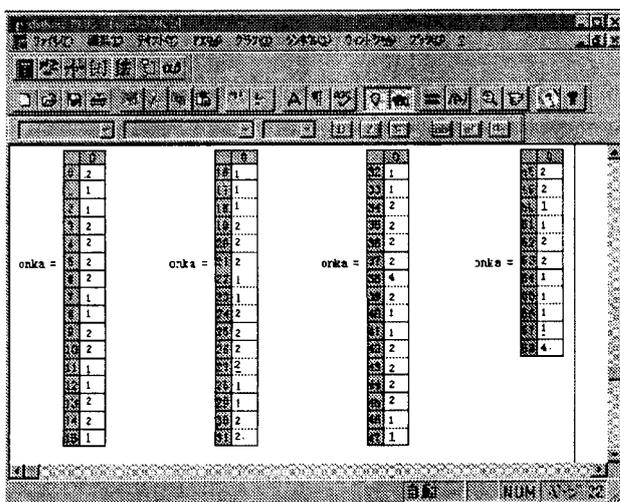


Fig. 7 Results of duration. '1' is an eighth note. '2' is a quarter note. '4' is a half note.

したが、Fig. 4 に示すように、短時間エネルギーの差分のみでは音の立ち上がりと判別できない部分が存在するため、精度上の問題があった。そこで、本論文では各音素の持続時間を抽出するための時間的な基準として基準拍という考え方を採用することにした。ここでいう基準拍とは、各音素の持続時間の最小の単位となるものであり、上述のデータの短時間エネルギーの自己相関関数 (Fig. 5) によって求めた値の山の頂点間のステップ数から求めたものである。この基準拍と差分データを用いて求めた各音素の持続時間が基準拍の2倍以上になる場合に、演奏

データを1章の方法によって周波数分析して求めた各音階のパワー (Fig. 6) から, 音の立ち上がりの再確認を行う方法を用いた. この方法を用いた結果を Fig. 7 に示す.

### 3.2 音階の同定

#### 3.2.1 周波数分析

採譜処理の場合, 各音素の基本周波数を求めることが目的であるので, サンプリング周波数を高くする必要性はあまりない. そこで DAT を用いて録音したサンプリング周波数 48kHz のデジタルデータを 12kHz にダウンサンプリングし, コンピュータに取り込む. そのデータに対して各周波数分析法を行い, 音階の同定を行なう. 高速フーリエ変換を用いる場合は観測窓は 50 点づつソフトし, 1kHz までの周波数帯を周波数補正し, ピーク周波数を求め, 各音階に当てはめて音階の同定を行なう. また, 適応システムを用いた周波数分析と複素 1 次系による周波数分析を行なう場合は, 第 2 オクターブから第 5 オクターブまでの 4 オクターブ, つまり 48 点の周波数成分で周波数分析を行ない, 音階の同定を行う.

#### 3.2.2 ノイズの除去とビブラート吸収

上述の周波数分析で求めた各音階の周波数成分はインパルス状のピークが生じることがある. 音階を精度良く同定するためにはこのようなインパルス状のノイズやビブラートを抑制する必要がある. 本論文では, 求めた周波数成分に対して前節の方法で求めた音価単位ごとに求めた周波数成分を積分し, 隣合う音階の周波数成分で比較することや, ピークとして出現した回数を比較することによってインパルス状のノイズの除去とビブラート吸収を行なう.

#### 3.2.3 高調波成分の除去

楽器が奏でる音は基本波成分と高調波成分からなる. この楽器が奏でる音の基本波成分と高調波成分のパワーの比は楽器の種類や基本周波数によって異なる. このことから基本波成分と高調波成分のパワーの比のテーブルを用意しておき, 基本波成分のパワーをその基本波成分が持つ高調波成分のパワーを加えたものと考ええることによって高調波成分の除去を行なう.

#### 3.2.4 音階の同定

音階の同定は上述のようにして求められた周波数成分に対して楽器の種類に応じたノイズレベルまで行なう. 同定結果を Fig. 8 に, また, 各周波数分析法を用いて分析した比較を Table 1 に示す. Fig. 8 において, 同定された音階は次のような 2 つの種類に分けて考える. 第一はその音価内にその周波数の音の立ち上がりがある場合であり, 図では "1" の数字で表されている. 第二は前音価からの継続する音である場合であり, "2" の数字で表

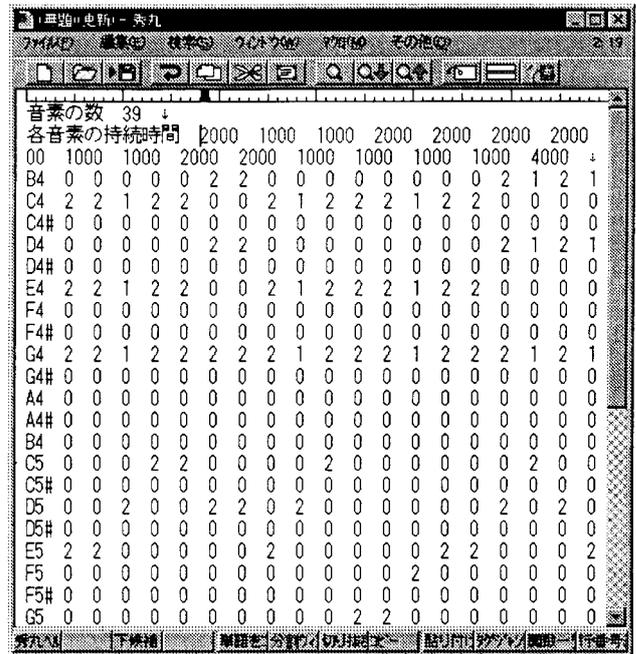


Fig. 8 Idification of tone.

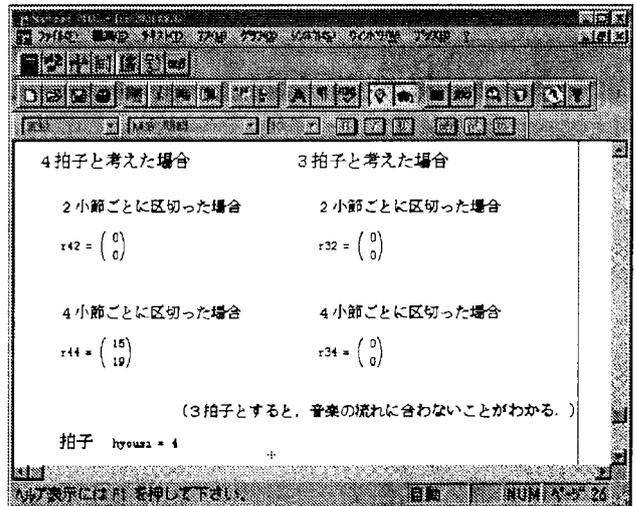


Fig. 9 Results of rhythm.

されている. また, "0" はその音階の音がないことを示している.

### 3.3 拍子・小節の決定法

わらべうたは, 2小節または4小節でひと区切りとなり, 第一メロディが冒頭で2回繰り返されるか, 最後に現れて曲が締めくくられるかの, 必ずどちらかであると決まっている. 本論文ではこの特徴を用いて, 曲が4拍子であるか3拍子であるかを決定する. つまり曲を4拍子または3拍子と考え, 音価を2小節および4小節ごとに区切ったものをつくる. この区切ったものをそれぞれメロディのかたまりだと考え, 第一メロディと第二メロディ, 第一メロディと最後のメロディの音価の一致率を

Table 1 Accuracy of transcription

Music	1024 points FFT method	Spectra analysis with CFOS	Spectra analysis with LMS algorithm
むすんでひらいて	90.0%	86.4%	90.7%
泉のほとり	79.9%	-	91.5%
谷川の流れ	80.0%	-	95.1%
メリーさんの羊	85.3%	80.9%	92.6%
夜汽車	81.0%	73.1%	91.0%
Average	83.2%	80.1%	92.2%

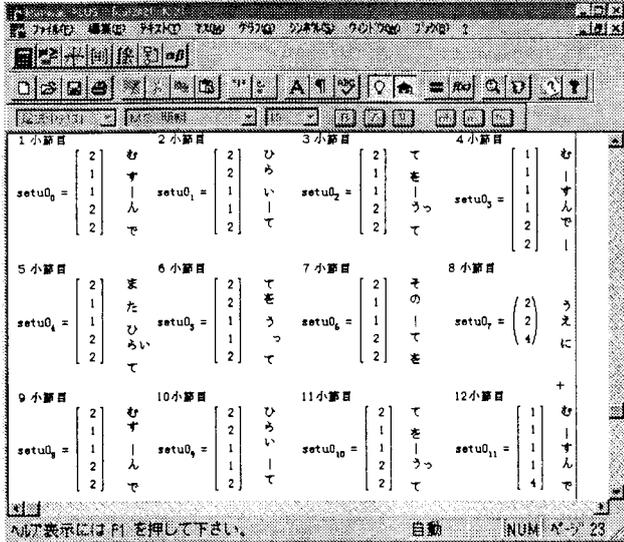


Fig. 10 Results of measure. '1' is an eighth note. '2' is a quarter note. '4' is a half note.

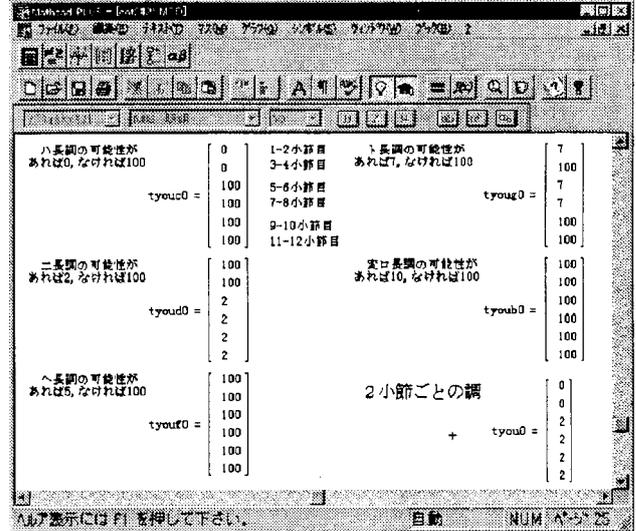


Fig. 12 Decision of key-signature.

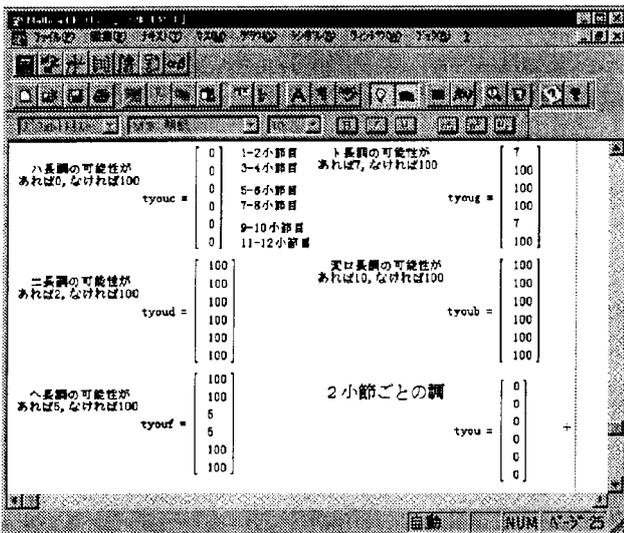


Fig. 11 Decision of key-signature.

調べ、一致率が高い方の拍子をその曲の拍子とする。その結果を Fig. 9 に示す。さらに、決定した拍子で音価を区切り、小節とする。求まった結果を Fig. 10 に示す。これらの図において'1'は8分音符、'2'は4分音符、'4'は2分音符を示す。

### 3.4 調子の決定法

調子を決定する目的を、譜表表示に用いられる臨時記号を少なくすることであると考えることもできる。本研究ではこのように考えて、2小節単位ごとに臨時記号を最も少なくできる調子を、その小節単位の調子として決定することにする。「むすんでひらいて」から求まった結果を Fig. 11 に、この曲を5小節目でへ長調からニ単調に転調するようにアレンジしたものを示した結果を Fig. 12 に示す。この図において、'0'はへ長調の可能性を示す。同様に'2'はニ長調、'5'はへ長調、'7'はト長調、'10'は変ロ長調の可能性を示す。また、'100'はその長調の可能性がない場合に示される数字である。

## 4. 結果

### 4.1 音階の同定精度の比較

適応技術を用いた周波数分析法は文献 5) から選定した適切なステップサイズパラメータを用いることによって信号に対する追随性が良く、結果として最も音階の同定精度が良かったと考えられる。つまり、高速フーリエ変換を用いて周波数補正を行なう場合には観測窓を必要とするため過渡状態を観測窓に含む時間が長くなり、結

果として基本周波数が低く、リズムの速い場合には音階の同定精度が劣化すると考えられる。また、複素1次系による場合には、この方法がもともと窓関数として方形窓を用いたデータの周波数成分を行うように設計されているので、サイドロープの影響によって周波数の漏れが大きいため、音階の同定精度の劣化になったと考えられる。このため、複素1次系による場合には比較的音の変化が速い場合においては、上述の影響が大きく、音階抽出精度が極度に悪く、表には記さなかった。

#### 4.2 採譜処理の過程について

従来は、短時間エネルギーの差分データを用い、その値があるしきい値以上である場合を音の立ち上がりとしていたが、Fig. 4のような音の立ち上がりが小さい場合には抽出できないという問題点があった。しかし、音に持続時間の基準となる基準拍と各音階のパワーデータを用いることによってこのような問題が解決できることが示された。また、拍子についても、従来のような音響パワーの変化パターンを用いて2拍子系か3拍子系か判断する方法ではなく、メロディーの周期性からも求められることが示された。さらに、音階データの転調の有無に関わらず、調子を決定することが可能な方法を提案することもできた。

#### 参 考 文 献

- 1) 原, 井口: 複素スペクトルを用いた周波数同定, 計測自動制御学会論文集,19-9,718/723 (1983)
- 2) 片寄, 井口: 知的採譜システム, 人工知能学会誌, Vol.5, No.1, 59/66(1990)
- 3) 山本, 他: 位相特性を利用した二つの隣接周波数の分離同定, 計測自動制御学会論文集,21-1,57/62 (1985)
- 4) 梅本, 青島: 適応技術を用いた採譜に有効な周波数分析, 計測自動制御学会論文集,28-5,619/625(1992)
- 5) 梅本, 青島: 適応フィルタを用いた周波数分析法におけるステップサイズパラメータの選定, 計測自動制御学会論文集,28-10,1257/1262(1992)
- 6) N.Aoshima: Short Time Spectrum Analysis by Complex First Order System, Proceedings of the 31st SICE Annual Conference, 1156/1160(1992)