

## 解説

## 計算機による採譜†



高澤 嘉光††

## 1. はじめに

採譜とは、「歌声や楽器演奏の音から楽譜を作る」ことである。音楽家にとって、音を聞き取り、それを楽譜に書き表せることは、非常にたいせつな資質である。しかし、採譜とは大変難しい技術であり、長年の訓練を積んでも、正しく楽譜を再現することはなかなかできない。われわれが聞き親しんでいる西洋音楽ならまだしも、民族音楽になると、人間の耳による採譜は不可能になってしまう。

そこで、コンピュータによる音声処理技術の進歩とともに、コンピュータによる採譜の研究が行われるようになった<sup>1),2)</sup>。次に、コンピュータによる採譜の目的あるいは効果について考えてみよう。

## (1) 民族音楽などの記録

この地球上には、たくさんの民族が住み、各民族は固有の音楽を有している。しかし、なんらかの形で記録されているものは少ないし、楽譜になっているものももっと少ない。このような音楽の研究のためには、音の記録とともに楽譜が必要である。日本にもいろいろな種類の音楽があるが、楽譜になっているものは少ない。各地に伝わる「わらべうた」もその一つであろう。これらをコンピュータで採譜して、そのデータ・ベースを作る試みが行われている<sup>3),4)</sup>。

## (2) 浄写あるいは写符

楽譜をきれいに書くことはなかなか面倒で、時間のかかる仕事である。録音テープから楽譜が作成できれば、作曲家にとっては、浄写の手間が省け、作曲に専念することができる。

現在、楽譜印刷はコンピュータ化されつつあり<sup>5)</sup>、採譜システムと連動すれば、人手を煩わせることなく楽譜データの入力ができる。

## (3) 演奏技術の解析、名演奏家の再現

楽譜には、演奏のための情報が100%含まれているわけではない。楽譜とは、絵に例えればスケッチであり、すなわち、構図だけができていて、色が無い。それに色を付けるのが演奏者である。どのように色を付けるかは、演奏者の責任であり、そこに演奏者の個性が表れる。最近、パーソナル・コンピュータなどで、簡単に楽器を演奏（正確に言えば、制御）できるようになったが<sup>6)</sup>、それを使って実験してみると、このことが、再認識させられる。人間が直接楽器を演奏すると、無意識のうちにいろいろな情報が追加されるが、それは当然のこととして受けとめられている。

名演奏とはいったいどんな演奏であろうか、これをコンピュータで解析する研究が行われるようになった<sup>7),8)</sup>。このためには、演奏された曲の譜面を再現する必要がある。このような演奏情報を含んだ譜面は、「奏譜」と呼ばれている。コンピュータによる採譜で、演奏の各種のデータが得られれば、いろいろな角度から演奏を解析することができる。もちろん、奏譜をコンピュータで正確に演奏すれば、名演奏を再現することもできる<sup>9)</sup>。

## (4) 音楽教育

これまで、演奏技術の修得は、耳がすべてであった。先生や名演奏者の演奏を耳で聞いて覚える。先生は、生徒の演奏を耳で聞き、間違いを再演してみるかあるいは、言葉で説明するがどうしても、説明が抽象的になる。また、演奏の評価にしても、現在は人間の感覚に頼り、主観的になってしまう。このようなとき、奏譜があれば話が具体的にになり、理解しやすくなるであろう。

コンピュータによる採譜の過程は大きく二つに分かれる。

- (1) 音から奏譜を作成する。
- (2) 奏譜から楽譜を作成する。

演奏技術の解析のためには、奏譜だけで用が足りる。奏譜とは音という物理データを五線譜に表したも

† Transcription with Computer by Yoshimitsu TAKASAWA  
(Department of Information and Mathematics, University of  
Electro-Communications).

†† 電気通信大学情報工学科

のであり、楽譜とは楽譜を楽譜表記規則に基づいて書き直したものである。この二つの変換方法は全く異なる。前者は、主に物理的な情報によって変換されるが、後者の場合は、豊富な音楽的知識が必要である。これに対して、知識工学的手法を適用する研究が行われている<sup>10)</sup>。

音から楽譜を作成する過程で一番難しいことは、雑音も含む音の中から必要な音を抽出することである。一方、19世紀の後半から今世紀にかけて数多く作られた自動ピアノの紙ロールや、現代の自動ピアノのフロッピー・ディスクや、MIDI規格の楽器の出力データは、演奏情報そのものであり、これから楽譜を作成することは非常に簡単である。

ここでは、狭義の意味での採譜、すなわち音から楽譜を作成することに話を限定する。

## 2. 手法と技術

コンピュータによる採譜の手順は、次のようになる。ここで使用しているデータは、レコードに録音されたピアノ曲の音である。

(1) 図-1のように短い時間  $T$  (図では、1/8秒)の間の音を A/D コンバータから入力する。この音を入力する短い時間のことは、時間窓あるいは観測窓と呼ばれる。

(2) 入力した音響データをフーリエ変換して、 $T$ の間の音を周波数成分に分解し、パワースペクトルを求める。図-1の音響データに対して、図-2のパワースペクトルが得られる。

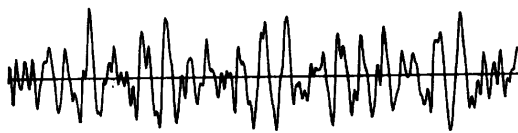


図-1 時間窓  $T$  の音

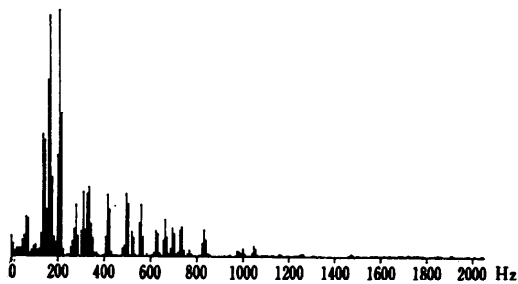


図-2 パワースペクトル

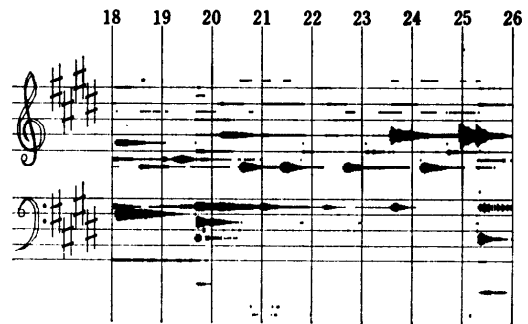


図-3 時間周波数マップ

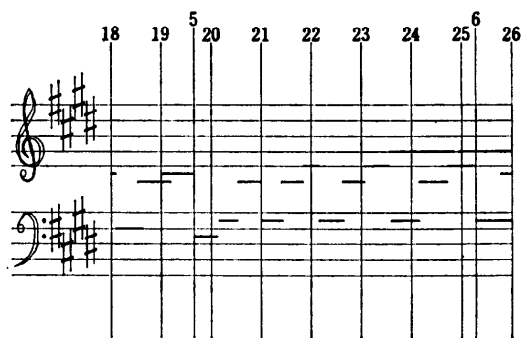


図-4 楽譜

(3) パワースペクトルから、ピークを見つけ、その周波数と強さを求める。一般に、ピークは一つとは限らない。

(4) 時間窓を少しずつ移動して、(1)~(3)を行う。これにより得られたピーク周波数とその強さから、時間周波数マップを作成する。図-3は、時間周波数マップを五線譜形式で表したものである。図の上部の数字は秒数を表す。

(5) 時間周波数マップから、有効な音を抽出し、楽譜を作成する。図-4は、図-3から得られた楽譜である。図の上部の数字5,6は小節番号を表す。

採譜の基本は、フーリエ変換である。これにより、ある短い時間(時間窓)内の音の高さ(ピッチ)と強さが求まる。実際に、採譜を行うためには、上に示したように、この時間窓を微小時間ずつ移動する必要がある。したがって、採譜が実用的なものであるためには、このフーリエ変換の計算が高速であることが必須条件である。このために、FFT(Fast Fourier Transformation)と呼ばれる計算法が使われる。現在は、数値演算プロセッサがパーソナル・コンピュータに装着されるようになり、FFTの計算時間が飛躍的に短

表-1 国際標準周波数による平均律音階

音名	オクターブ番号							
	0	1	2	3	4	5	6	7
C	16.35	32.70	65.41	130.81	261.62	523.25	1046.50	2092.99
C#	17.32	34.65	69.30	138.59	277.18	554.36	1108.72	2217.45
D	18.35	36.71	73.42	146.83	293.66	587.33	1174.65	2349.30
D#	19.45	38.89	77.78	155.56	311.13	622.25	1244.50	2489.00
E	20.60	41.20	82.41	164.81	329.63	659.25	1318.50	2637.00
F	21.83	43.65	87.31	174.61	349.23	698.45	1396.90	2793.81
F#	23.12	46.25	92.50	185.00	369.99	739.98	1479.97	2959.93
G	24.50	49.00	98.00	196.00	391.99	783.99	1567.97	3135.94
G#	25.96	51.91	103.83	207.65	415.30	830.60	1661.21	3322.41
A	27.50	55.00	110.00	220.00	440.00	880.00	1759.99	3519.97
A#	29.14	58.27	116.54	233.08	466.16	932.32	1864.64	3729.28
B	30.87	61.74	123.47	246.94	493.88	987.76	1975.52	3951.04

くなった<sup>9)</sup>。

次に、採譜における重要な問題点について述べる。

(1) 音高の同定

音楽の世界で用いられる音階は、1オクターブ（周波数が2倍）をどう分解するかによって、純正音階、ピタゴラス音階、平均律音階などに分けられている。現在は、表-1の1953年に改正された平均律音階が用いられている。これは、1オクターブを対数で12等分したものであり、隣り合う音（半音）の周波数の比は、 $2^{1/12}$  = 約 1.06 であり、A<sub>4</sub> を 440 Hz にしたものである。半音の周波数の相対的な差は 6% であり、その絶対的な差は低音になればなるほど小さくなる。たとえば、ピアノの中央の C から 2 オクターブ低い C<sub>2</sub> では 3.9 Hz しかない。この二つの音を識別するには時間窓を約 0.26 秒にする必要がある。これに対して中央の C では約 0.06 秒、2 オクターブ高い C<sub>6</sub> では 0.016 秒でよい。一方、ピアノの演奏において、8 分音符をメトロノーム速度 120 で弾くと、その長さは 0.125 秒、さらに 16 分音符は 0.062 秒になる。この程度の短い音は、ピアノ演奏では頻繁に現れる。したがって、この短い音を抽出するためには、時間窓をそれより短くする必要がある。しかし、時間窓を短くすると、周波数分解能が悪くなる。

そこで、短い時間窓で FFT の計算を行い、その結果に補正を加えて精度を高める方法として、次の三つの補正法が考えられている。

1) 倍音の周波数から逆算する<sup>11)</sup>。

普通の楽器では必ず倍音が発生する。この倍音の周波数から基音の周波数を計算する。倍音が高ければ高いほど精度は良くなる。たとえば、4 倍音であれば元の分解能の 4 分の 1 の精度が得られる。しかし、倍音

はパワーが小さく、識別が困難なこともあり、ときには役に立たないこともある。

2) スペクトルのピークの短い時間変化における位相化速度から計算する<sup>12)</sup>。

短い時間  $\Delta\tau$  だけ離れた時間窓で得られた周波数  $F_0$  の同一ピークの位相を  $\theta_1$  と  $\theta_2$  とすると、

$$F = F_0 + \frac{\theta_1 - \theta_2}{2\pi\Delta\tau}$$

で、周波数を補正することができる。これを求めるには、2 回の FFT の計算が必要である。

3) 複素スペクトル面で内挿する<sup>13)</sup>。

FFT によって求められた複素スペクトルのピークを挟む二つの成分を  $z_m$  と  $z_{m+1}$  とする。ここで、

$$u = (z_{m+1} - z_m) / |z_{m+1} - z_m|$$

という単位ベクトルを定義すると、

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

で、周波数を補正することができる。ここで、

( $\cdot, \cdot$ ) は内積を表す。

三つの補正法のうち、3) の内挿法が一番効率が良い。ここで、それを具体例で示す。時間窓を 1 とし、 $\sin(2\pi at)$  という波形データに対して、256 点の FFT の計算を行った。表-2(a) には、 $a$  に対するピーク値と内挿法による補正值を示した。このことから次のことが分かる。もし、時間窓が 0.1 秒であれば、

表-2(a)  $\sin(2\pi at)$  の場合の補正值

a	ピーク値	補正值	誤差
1.4	1	1.33	0.07
2.4	2	2.36	0.04
4.4	4	4.36	0.02
8.4	8	8.39	0.01

表-2(b)  $\sin(2\pi at) + \sin(2\pi bt)$  の場合の補正值

$a$	$b$	$a$ の補正值	$b$ の補正值
1.4	3.4	1.14	3.45
2.4	5.4	2.25	5.44
5.5	9.4	5.40	9.44
5.5	14.5	5.44	14.51

分解能は 10 Hz である。これに補正を施すと、14 Hz の周波数を 0.7 Hz の精度で求めることができる。さらに、44 Hz の周波数は、0.1 Hz の精度で得られる。ここで、小さい  $a$  の値に対して、精度が悪くなるのは、二つのピークが近づくためである。数値の上ではピークが一つであるが、理論的には 0 を対象軸として負の側にも同じものがある。したがって、 $a$  の値が小さくなると自分自身との干渉で誤差が大きくなる。内挿法は、複数のピークをもつ波形データにも適用できる。表-2(b) には、 $\sin(2\pi at) + \sin(2\pi bt)$  の波形に対して適用した結果を示す。この表より、二つのピークが離れるほど、補正值の精度が良くなっていることが分かる。

しかし、内挿法が有効である条件は、周波数を  $F$  ( $=1/T$ ) で区間分けを行ったとき、その区間の中にピークが一つしか存在しないときである。たとえば、 $C_3$  は 130.81 Hz であり、半音高い  $C\#_3$  は 138.59 Hz であるので、時間窓を 0.1 秒とした場合には、この二つの音は同一の区間に入り、識別は不可能になる。このときには、時間窓を長くして、識別可能な周波数区間幅を狭くする必要がある。

## (2) 時間窓の移動間隔

FFT の計算は効率が良いといっても、パーソナル・コンピュータではかなりの時間を要するので、移動間隔を小さくすると楽譜の作成のための時間が長くなってしまふ。一方、演奏技術の解析のためには、音の発生と消滅の時点を、できるだけ精度を高めて計測する必要がある。曲の内容にもよるが、ピアノ曲では、一般に 0.01 秒単位が要求される。これには、時間窓を 0.01 秒単位で移動する必要がある、毎秒 100 回の FFT の計算を行うことを意味している。もし、1 回の FFT の計算に 0.1 秒の時間を要するならば、1 秒の演奏の解析のために、FFT の計算だけに 10 秒の時間を要する。1 回が 0.01 秒になって、ようやく 1 秒になり、同時間になる。このように FFT を高速計算できれば、リアル・タイムで楽譜を作成することができる。しかし、FFT の計算をこの速度で行うことは、パーソナル・コンピュータでは不可能である。ま

た、実際に楽譜を作る場合には、FFT のパラメータを状況に応じて変える必要があり、現在の段階では、楽譜をリアル・タイムで作ることはできない。そのため、AD 変換された音のデータをいったんディスクに蓄えておく必要がある。このとき、音のサンプリング時間が短くて、変換されたデータがすべてメモリ内に記憶しておくことができればよいが、時間が長くなる場合には工夫が必要である<sup>6)</sup>。

## (3) 音価の決定

音価とは音の持続時間であり、これを決めることも採譜において難しい問題である。音価を決定するには、音の発生した時点と消滅した時点を決めなければならない。もし、音が短形波的に発生するならば、この二つの時点を見つけることは非常に簡単であるが、現実の楽器ではそうはいかない。たとえば、ピアノのような打弦楽器では、叩いた瞬間は強い音が出るが音は減衰して消える。その減衰のカーブは、2本の直線に分けることができる<sup>10)</sup>。初めの短くて速い減衰は打音であり、後の長いゆっくりとした減衰は余韻である。したがって、ピアノ音の場合、発音の時点は比較的簡単に決めることができる。これに対して、消滅の時点を見つけることは難しい。短い音であれば、鍵盤を離した瞬間、減衰カーブより急な傾きで消滅する。長く続く音は、レベルが小さくなり、カーブの変化点を見つけることが困難になる。

## (4) 倍音

採譜で問題になるのは、倍音である。この倍音は、解析しようとする音が単音 (monophony) である場合には比較的簡単であるが、複音 (polyphony) になると、難しくなる。単音とは、目的とする音素が常に 1 個しか含まれていない音であり、たとえば、伴奏のない歌とか、旋律 (メロディ) だけのピアノ曲などがこれにあたる。複音とは、目的とする音素が何個含まれているか分からない音である。たとえば、伴奏付きの歌唱とかコーラス、伴奏付きのピアノ曲、シンフォニーなどである。

単音であれば、FFT の結果から得られるパワースペクトルで一番高いピークを見つけ、その周波数を求める音の高さとすることができる。これに対して、複音の場合には、抽出するピークの数不明である。また、多くの場合、楽器には倍音が現れるが、その音が倍音であるか、実際に弾かれた音であるかを判定することは、パワースペクトルの段階では難しい。そこで、時間周波数マップで、音の減衰カーブから判定す

ることになる。

複音の採譜においても一つ問題になるのは、非常に近いピーク周波数の音が含まれる場合である。表-1から分かるように、低音では半音の間の周波数の差は非常に小さい。したがって、低音域で近接した周波数の音があった場合には、時間窓を長く取り周波数分解能を高める必要がある。

### 3. 知的採譜システムと採譜支援システム

歌唱などの単音に対して、コンピュータによる採譜はかなり実用的なレベルに達している<sup>15),16)</sup>。しかし、複音に対しては、研究が始まったばかりといえる<sup>17),18)</sup>。単音と複音の違いは、物理的には、そこに含まれる音(素)の数の違いである。しかし、ここで取り扱う音は、音楽の音であり、一つ一つの音は音楽的に意味をもち、音楽的に結ばれている。単音の場合には、この関係は前と後ろという時間的なものである。ところが、複音の場合には、時間的な関係に加えて、音どうしの関係も入る。したがって、複音の採譜が難しいのは、音に音楽的な意味があり、単に物理的な情報だけで判断できないからである。それでどうしても、音楽の知識が必要になる。この処理に対して、次の二つの試みがある。

#### (1) AI の手法を用いる方法

この一つとして、ピアノ曲を対象にした知的採譜システム<sup>19)</sup>を紹介する。そこでは、人間の“聞く”という知識処理過程をシミュレートしている。そのシステムは、制御モジュール、処理モジュール、音楽解析モジュールによって構成されており、各モジュールは次の働きをする。

- 制御モジュールは、システムの中心的なモジュールであり、OPS 83 言語によって記述され、制御用ルール、推論エンジンから成り立っている。

- 処理モジュールは、採譜処理の中心になり、次の処理を行う。

- 1) 基準拍の長さの推定
- 2) 基本音の抽出
- 3) 打鍵音であるか連続音であるかの判別
- 4) 基準拍より短い音の認識

- 音楽解析モジュールは、音楽理解のためのシンタックス解析を目標として構成されたモジュールであり、メロディ、リズム、コード進行、調などの抽出が可能になっている。

このシステムでは、音楽解析部からの情報の採譜へ

のフィードバックや、内部パラメータの自動修正などが可能になっている。そのため、認識率が 90~100% と非常に高くなっている。

#### (2) 対話形式による方法

一般に、レコードや CD (コンパクト・ディスク) などの曲の場合には、演奏に緩急があるので対話型コンピュータ・システムを使って、人間のもっている音楽的な知識を生かしながら、採譜を進めていくという採譜支援システムが効果的である<sup>20)</sup>。このシステムが有効である理由として、次のことが考えられる。

1) パラメータの変更が容易である。

FFT の計算を行うには、点の数や時間窓の幅や移動間隔など、いろいろのパラメータを設定しなければならない。これらを固定してしまうよりも、時と場合によって変更した方が良い。たとえば、高音域の解析に対しては時間窓を短くして時間分解能を高め、低音域に対しては時間窓を長くして周波数分解能を高めたほうが効果的である。また、周波数が近接した音が現れた場合には、時間窓を長くする必要がある。

2) 情報のフィードバックが可能である。

採譜の結果得られた音楽的な情報を利用する。たとえば、その曲が何調であるかが分かれば、それ以降の解析に大きな手がかりを与えてくれる。

3) 耳による確認が可能である。

採譜した結果をコンピュータで自動演奏させ、原曲と比較することができる。

### 4. 採譜の一応用例

図-5 は、ベートーヴェン作曲ピアノ・ソナタ「月

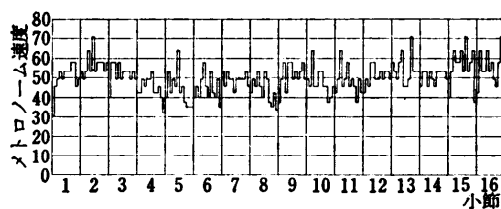


図-5(a) バックハウスの演奏

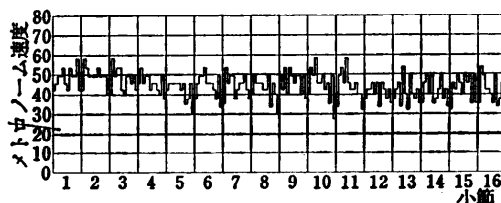


図-5(b) バレンボイムの演奏



図-6 「月光」の第1小節

光」作品 27 の 2 の、二人のピアニストによる演奏の局所メトロノーム速度であり、最初の 16 小節分が示してある。図-5(a) はバックハウス<sup>21)</sup>であり、図-5(b) はバレンボイム<sup>22)</sup>である。この図は、採譜によって得られた楽譜から一つ一つの音符の長さ(演奏時間)を測定し、それから計算したメトロノーム速度をグラフ化したものである。この曲は、図-6 のように三連符が最初から最後まで続く。しかし、これらの図から分かることは、同じ三連符でもその演奏速度が非常に異なっていることである。両者を比較すると、バックハウスの方が緩急が激しい。その差は、2 倍以上にもなっている。

このように、採譜から得られた楽譜データを図式化すれば、演奏の特徴を明確に把握することができる。

## 5. おわりに

音は、たとえ、それを録音しても、聞くとときには一瞬のうちに消えてしまう。採譜とは、音を図示することであり、これによって音を眺めることができる。また、それから得られた情報によって演奏のさまざまな解析ができる。採譜が手軽にかつ正確にできるようにすれば、音楽教育などいろいろの方面で利用されるであろう。そのためには、FFT の計算が高速になり、さらに、AI の手法を使った対話型のシステムの開発が必要である。おそらく、採譜は、計算機の音楽への応用の中で、将来一番期待されている分野であろう。

## 参考文献

- 1) Chafe, C., Mont-Reynaud, B. and Rush, L.: Toward an Intelligent Editor of Digital Audio: Musical Construct Recognition, *Computer Music Journal*, Vol. 6, No. 1 (30-41) (Mar. 1982).
- 2) Foster, S., Rockmore, J. and Schloss, W.: Toward an Intelligent Editor of Digital Audio: Signal Processing Methods, *Computer Music Journal*, Vol. 6, No. 1 (42-61) (Mar. 1982).
- 3) 山本, 中山: 音楽データ・ベースとその検索, 音楽音響研究会資料 MA 83-3 (1983).

- 4) 山本: 民族音楽とデータベース, bit 別冊「コンピュータと音楽」 pp. 105-111, 共立出版 (1987).
- 5) 岡: 楽譜印刷, bit 別冊「コンピュータと音楽」 pp. 112-123, 共立出版 (1987).
- 6) 高澤: パソコンによる音楽演奏, bit, Vol. 16, No. 11, pp. 4-17, 共立出版 (1984).
- 7) 田口: 自動演奏と演奏モデル, bit 別冊「コンピュータと音楽」 pp. 42-53, 共立出版 (1987).
- 8) 太田, 田口, 馬道: ショパンのワルツとマズルカの演奏リズムについて, 音楽音響研究会資料 MA 87-14 (1987).
- 9) 高澤: 採譜から自動演奏へ, bit 別冊「コンピュータと音楽」 pp. 131-141, 共立出版 (1987).
- 10) Chowning, J., Rush, L., Mont-Reynaud, B., Chafe, C., Schloss, W. and Smith, J.: Intelligent Systems for the Analysis of Digital Acoustic Signals (final report), CCRMA, Dept. of Music, Stanford Univ., Report No. STAN-M-15 (Jan. 1984).
- 11) 北原, 清水, 末武: コンピュータによる楽音認識の試み, 日本音響学会誌, Vol. 33, No. 9, pp. 464-469 (Sep. 1977).
- 12) 上野, 中村, 井口: 音楽情報の入力法一採譜一, 電子通信学会技報, PRL 78-21 (1978).
- 13) 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, Vol. 19, No. 9, pp. 36-41 (Sep. 1983).
- 14) G. ウェインリーチ: ピアノの弦の物理学, 楽器の科学, 日経サイエンス社 (1987).
- 15) 新原, 今井, 井口: 歌唱の自動採譜, 計測自動制御学会論文集, Vol. 20, No. 10, pp. 68-73 (Oct. 1984).
- 16) 井口: 採譜と音楽情報処理, bit 別冊「コンピュータと音楽」, pp. 30-41, 共立出版 (1987).
- 17) Chafe, C. and Jaffe, D.: Source Separation and Note Identification in Polyphonic Music, ICASSP 86 (1986).
- 18) 今井, 高見, 辻本, 井口: 人間のピアノ演奏をシミュレートするシステム, 第 33 回情報処理学会全国大会予稿集 3N-10 (1986).
- 19) 片寄, 今井, 井口: 多重音を対象とした知的採譜システム, 音楽音響研究会資料 MA 87-18 (1988).
- 20) 高澤: パソコンによる採譜支援システム, 音楽音響研究資料 MA 86-8 (1986).

## 音楽資料

- 21) ベートーヴェン: ピアノ・ソナタ第 14 番嬰ハ短調作品 27 の 2 <月光>演奏ウイエルム・バックハウス SLC 1444
- 22) ベートーヴェン: ピアノ・ソナタ第 14 番嬰ハ短調作品 27 の 2 <月光>演奏ダニエル・バレンボイム F 35 G 20088

(昭和 63 年 2 月 16 日受付)