

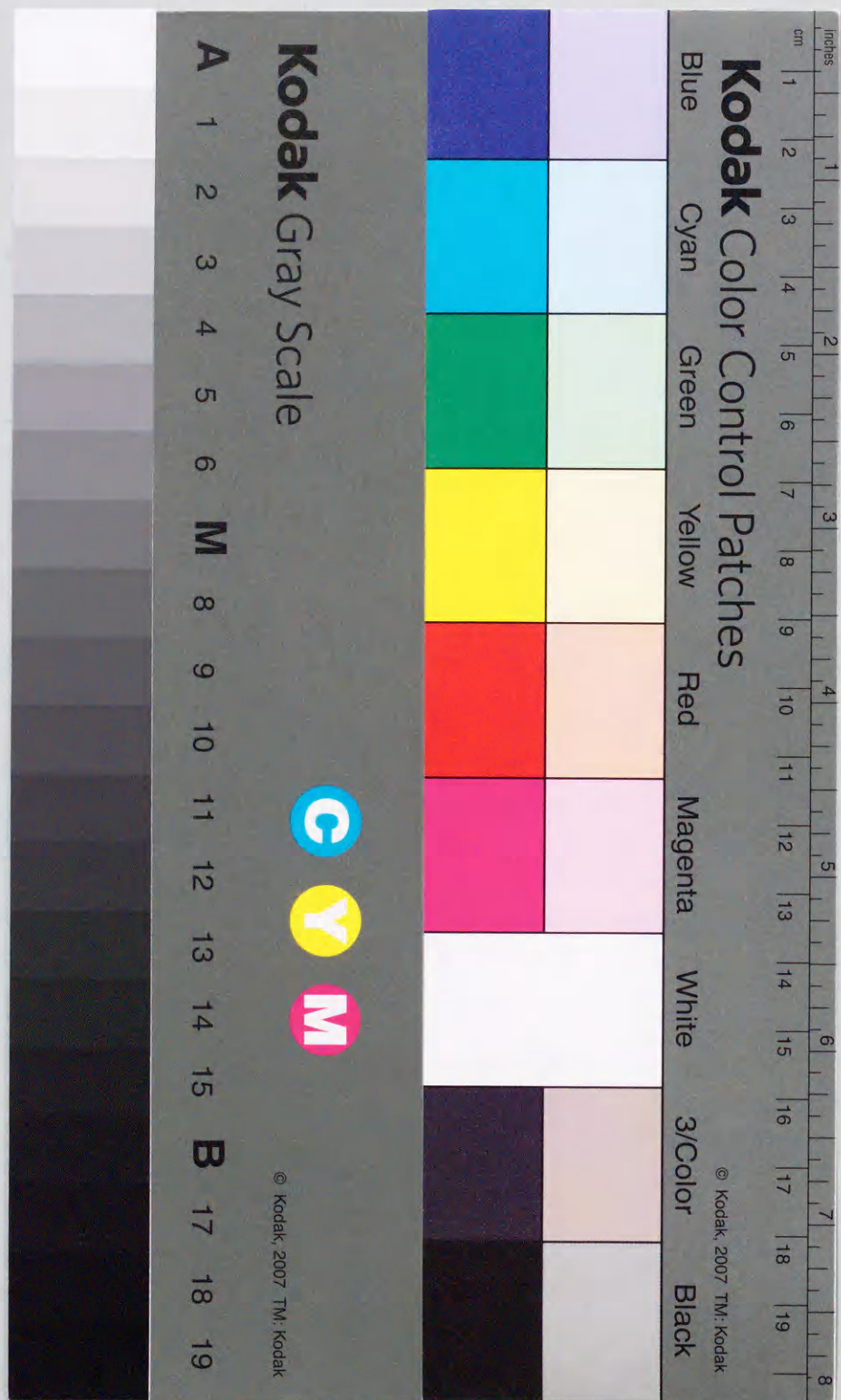
くし形フィルタによる
自動採譜システムに関する研究

2001年1月

博士(工学)

三輪多恵子

豊橋技術科学大学



くし形フィルタによる
自動採譜システムに関する研究

2001 年 1 月

博士（工学）

三輪 多恵子

豊橋技術科学大学

論文要旨

採譜とは、楽器で演奏された曲などの音情報から、予め定められた記号系列、即ち楽譜を作成する処理であり、近年、計算機による自動採譜システムの研究が盛んに行なわれている。計算機を用いた採譜においては、与えられた音楽信号（異種複数楽器音）から、音高（基本周波数）、音価（音の継続時間）、楽器の種類を推定する必要がある。しかし、平均率音階の持つ不等間隔な周波数配置により、楽器音に含まれる多数の高調波成分が複雑に重複するため、異種複数楽器音の解析は非常に困難とされているのが現状である。

本研究では、くし形フィルタ $H_{r,p}(z_p) = 1 - z_p^{-r}$ ($r = 2, 4, 8$: オクターブに関係, $P = C, C\#, \dots, B$: 音名) に基づく音楽信号の解析法を提案する。この $H_{r,p}(z_p)$ は、0[Hz] からサンプリング周波数までを r 等分する箇所に零点を持ち、楽器音に含まれる全周波数（基本周波数+高調波）成分を除去する性質を持つ。本論文では、 $H_{r,p}(z_p)$ による音高判別システム、及び、和音分離システムを説明し、さらに、スペクトル構造を利用した楽器推定法を示すことで、本方式による自動採譜システムの可能性を検討する。

本方式では、 $C, C\#, \dots, B$ 音に対応する $H_{r,p}(z_p)$ を縦続接続し、周波数成分が除去されたか否か、即ち、零・非零出力の検出により、和音に含まれる音名を判別する。その後、和音中の各音に対し、 $H_{r,p}(z_p)$ により注目音以外の音を除去することで、単音抽出（和音分離）を行なう。本方式は、 $H_{r,p}(z_p)$ の周波数成分除去の性質を利用しており、楽器音のスペクトル構造（楽器の種類）に依存せず上記の処理が可能であるという利点を持つ。さらに、前述の抽出音に対し、DFT によるスペクトル解析を行ない、各楽器の単音スペクトル（テンプレート）と比較することで演奏楽器を推定する。抽出音に含まれる周波数成分は基本周波数と高調波成分に制限されており、非常に少ないサンプル点数の DFT によりスペクトル解析が可能である。なお、本方式では、予め各スペクトルをノルム=1.0 に正規化することで、上記の比較処理を絶対値差分の最小値検出のみで実現している。

本論文では、採譜対象音域をオクターブ 3 から 5 の 36 音とし、電子楽器音 (MIDI: KOR-G 05R/W) 使用して本方式の性能評価を行なった。結果として、楽器音の非定常性の影響から線形補間や閾値を用いた判別が必要となるものの、従来法と比較して、本方式が非常に簡単なアルゴリズム（少ない演算量）で異種複数楽器音の音高及び音価推定が可能であることを示した。また、スペクトル構造を利用した楽器推定を行ない、同属楽器（弦楽器同士、木管楽器同士、等）和音、異属楽器（弦楽器+木管楽器、木管楽器+金管楽器、等）

和音ともに、本方式を用いて楽器推定可能であることを示した。これらの結果より、計算機による採譜システムに対する本方式の有用性を明らかにした。本研究の成果は、音楽認知や演奏解釈等の音楽学の分野に貢献できるものと信ずる。

A Study of Automatic Musical Transcription using Comb Filters

Abstract

It is called musical transcription to write a musical score from musical instrument sounds and songs. The musical transcription is necessary in the field of musicology and also one of the interesting targets as application of machine perception and artificial intelligence. In the automatic musical transcription using the computer, a musical pitch, its duration time, and a musical instrument must be estimated from an input signal of the different musical instrument sounds. However, those processings were very difficult in the previous methods that are based on the extraction of pitch frequency, since the frequency ratio of adjacent tones of an equal temperament scale is $2^{1/12}$ and a musical sound contains many harmonic frequencies.

We propose a method for analyzing a music signal of different musical instrument sounds using comb filters $H(z) = 1 - z^{-n}$ which can remove all the frequency components (fundamental frequency and harmonic frequencies) of a tone. In this paper, we show that we can estimate the musical pitch, its duration time using comb filters and determine the musical instrument based on spectrum structure. Then we prove the availability of the proposed method by the computer simulations.

In our method, the musical pitch of the different musical instrument sounds can be estimated by detecting the zero outputs of the cascade connection of the comb filters which correspond to $C, C\#, \dots, B$. And, by removing $m-1$ tones from m tones (polyphony) using comb filters, the different musical instrument sound is easily separated to each tone. Furthermore, for each separated tone, the played musical instrument is distinguished by the feature of spectrum structure. Since the frequency components contained in the separated tone is only the fundamental frequency and its harmonic frequencies, the frequency analysis by the DFT using the few sampling points is possible.

In this paper, we assume the conditions of computer simulations in which the range of the scale is from the tone name C of the octave number 3 to the tone name B of the octave number 5, i.e., 36 tones, and an analyzed signal is different musical instrument sounds of a MIDI sound source (KORG:05R/W). As a result, it is clear that the proposed method is useful for the detection of the musical pitch, its duration time and the played musical instrument for different musical instrument sounds (polyphonic music).

目次

1 緒言	1
1.1 背景	1
1.2 研究目的	3
1.3 論文概要	4
2 自動採譜に関する基礎的考察	7
2.1 自動採譜のタスク	7
2.2 楽器音の特徴	9
2.2.1 平均率音階における周波数配置	10
2.2.2 楽器音に含まれる高調波成分	12
2.2.3 時間軸上の振幅変化と波形	14
2.3 異種複数楽器音の採譜	15
2.3.1 音高の推定	16
2.3.2 音価の推定	16
2.3.3 楽器の推定	17
2.4 既存のアルゴリズム	17
2.4.1 音高推定法	17
2.4.2 楽器推定法	22
3 くし形フィルタによる音高、音価推定	25
3.1 くし形フィルタの原理	25
3.1.1 くし形フィルタの構成	26
3.1.2 オーバサンプリングによるくし形フィルタの近似的な実現	29
3.2 くし形フィルタの縦続接続による音高検出	32
3.2.1 基本的な音名判別システムの構成	33
3.2.2 基本的な音高検出アルゴリズム	33
3.2.3 周波数の重複を利用した音名判別システム	35
3.2.4 周波数の重複を利用した音高判別アルゴリズム	37

3.2.5	周波数のずれに対する音名判別システムの特性	41
3.3	くし形フィルタの零出力検出による音価推定	44
4	くし形フィルタとスペクトル構造を利用した楽器推定	47
4.1	楽器推定のためのテンプレート構造	47
4.2	くし形フィルタによる和音分離とスペクトル抽出	50
4.2.1	基本的な和音分離システムの構成	50
4.2.2	DFTによるスペクトル解析	51
4.2.3	テンプレートマッチングによる楽器推定	52
4.3	くし形フィルタによる和音分離の問題点	52
4.4	和音分離の問題点の解決	54
4.4.1	2音の周波数成分が独立な和音	56
4.4.2	2音の高調波成分の一部が重複する和音	57
4.4.3	1音の基本周波数が高調波成分に重複する和音	60
4.4.4	2音の基本周波数が重複する和音	62
5	実験結果と考察	65
5.1	実験条件	65
5.1.1	使用する音源	65
5.1.2	自動採譜システムの概要	66
5.2	音高、音価推定結果	67
5.2.1	単一楽器音の単音に対する音高推定	67
5.2.2	単一楽器音の単音に対する音価推定	71
5.2.3	単一楽器の和音に対する音高推定	72
5.2.4	単一楽器の和音に対する音高、音価推定	78
5.3	楽器推定結果	81
5.3.1	2音の周波数成分が独立な和音	81
5.3.2	2音の高調波の一部が重複する和音	82
5.3.3	1音の基本周波数が高調波に重複する和音	84
5.4	異種複数楽器音の採譜結果	86
5.4.1	単一楽器の単音の旋律に対する採譜	86
5.4.2	単一楽器の和音に対する採譜	88
5.4.3	異種複数楽器音に対する採譜	91

6	結言	93
6.1	まとめ	93
6.2	今後の課題	95

参考文献	99
------	----

付録 A	テンプレートスペクトル
------	-------------

付録 B	記号
------	----

第 1 章

緒 言

1.1 背景

音符や休符などの音楽記号を使用し、楽器で演奏した曲や歌唱等を楽譜に書き表す作業を採譜と呼ぶ。現代社会には、民族音楽や民謡のように、楽譜形式ではなく口伝え等の方式で継承されている様々な楽曲が存在する。このような楽曲は、時代の流れと共に失われてしまう可能性があるため、採譜を行なうことで正確に記録し、文化を保存することが強く要求されている [6]。また、音楽の演奏は奏者の感性に依存するため、その場の雰囲気や奏者の気分により、楽譜には存在しない即興演奏が行なわれることが多い [8]。さらに、作曲者はピアノやギター等を演奏しながら曲を作り出すため、演奏しながら楽譜上に音楽記号を書く必要があり、この作業には大変な労力が必要である。これらの理由から、演奏家や作曲者によって演奏された曲に対し、リアルタイムで採譜を行ない楽譜を作成することが求められている [47]。

人間が採譜を行なうためには、その楽器が何の音高を演奏しているかを聞き分ける聴覚（絶対音感）、曲調からのテンポの判別、楽器音の継続時間とテンポの相対的な関係からの音符の割り出し等が必要であり、そのため、音楽に関する知識や多くの経験が必要とされる。しかし、このような知識と経験を有していても、楽曲の採譜をリアルタイムで正確に行なうことは大変困難であり、特に複数楽器による多重音（和音）等に対するリアルタイムの採譜作業は、人間の能力では不可能に近いとされている [6]。そこで、特別な能力を持たない場合でも、簡単かつ正確な採譜作業を可能にするために、計算機を用いた自動採譜の研究が盛んに行なわれている [1]-[37]。従来の自動採譜の研究が単一楽器音の音高推定を主体としていたのに対し、近年では、異なる複数の楽器により演奏された和音（以後、異種複数楽器音とする）の音高推定及び楽器推定を行なうことへ、その研究対象が拡張される傾向にある [21]-[37]。

音楽の世界で一般的に用いられている平均率音階では、オクターブ番号 4 の A 音の基本周波数 ($A^4: 440.0[\text{Hz}]$) を基準とし、1 オクターブ中に含まれる 12 音 ($C, C\#, \dots, B$) の基本周波数が対数的に割り当てられている。従って、隣り合う音高の周波数比は $2^{1/12}$ 倍となり、各基本周波数は不等間隔な周波数配置を持つ [44]。また、楽器音には、音高を決定する基本周波数成分の他に、多数の高調波（倍音）成分が含まれるという特徴を持つ [45],[46]。

計算機を利用して採譜を行なうためには、入力された楽器音に含まれる複数の周波数成分を解析し、基本周波数成分を検出することで音高を推定する必要がある。このための周波数解析法として、多くの研究では高速フーリエ変換 (FFT : Fast Fourier Transform) が用いられている [1]-[8],[11]。しかし、FFT は解析周波数が基本波周波数 (サンプリング周波数 f_s [Hz] で N 点 FFT の場合には f_s/N [Hz]) の整数倍に限られるといった制約を持つため、前述の性質を持つ音楽信号に使用する場合には、窓関数の考慮が必要となる。また、井口らによる報告では、周波数補正等の特別な処理が必要とされている [4]。FFT のように解析周波数に制約を持たない手法として、梅元らにより適応フィルタを用いた周波数解析法 [9] が、また、門藤らにより拡張ノッチフーリエ変換 (E-NFT : Extended Notch Fourier Transform) を用いた音高判別法 [10] が提案されている。しかし、前述のように、楽器音には多数の高調波成分が含まれていることから、音高推定を行なう際には、単音の場合でも非常に多くの周波数成分を含む信号の解析が必要である。このため、適応フィルタや E-NFT を用いた周波数解析法では、計算量の増大や処理系の複雑化等が新たな問題となっている。

さらに、異種複数楽器音の解析を行なう場合には、音高の推定以外に、各楽器音の分離や楽器推定といった新たな処理が必要となる [41],[47],[48]。異種複数楽器音を単一楽器音に分離するためには、まず、和音に含まれる音高の数や種類を判別する必要がある。このため、FFT に基づく周波数 (スペクトル) 解析を主体とする方式では、和音全体のスペクトルのローカルピーク (ある基本周波数) を検出することで音高の判別を行ない、その後、単一楽器音に含まれる全ての周波数成分を分離するための高調波のグルーピング処理を必要とする [25],[29],[31]。しかし、異種複数楽器音に含まれる各音の周波数成分は複雑に重畳しているため、基本周波数の検出が非常に困難であり、高調波成分のグルーピングの精度がパワーの強弱に依存する問題が磯部らにより報告されている [29]。上記の問題に対して、白土らの報告では統計的検定のような特別な判断基準が設けられている [31]。さらに、楽器推定におけるマッチング処理において、小野らの方式では、予め用意したテンプレートに対する総当たりの検索が必要とされる等、計算量の増大が問題となっている [25]。

また、各楽器音 (音源) の分離のみを目的とした手法として、永田らによる複数センサを用いた音源抽出法 [23] や、植田らによるブラインドデコンポジションアルゴリズムによる音源分離手法 [24] が報告されているが、2 章で述べるように、手法が適用できる状況や音源に対して厳しい制約条件が必要とされるのが現状である。

1.2 研究目的

本研究では、くし形フィルタ $H_{r,P}(z_P) = 1 - z_P^{-r}$ ($P = C, C\#, \dots, B$, $r = 2, 4, 8$: オクターブに対応) を用いた音楽信号解析法について検討を行なう。このくし形フィルタは、周波数領域において 0 [Hz] からサンプリング周波数までを r 等分する箇所に零点を有し、サンプリング周波数を各音高の周波数に対応させることで、楽器音に含まれる全ての周波数 (基本周波数、高調波) 成分を除去する性質を持つ。これにより、問題とされていた高調波成分の影響を回避することができる。また、くし形フィルタの伝達関数は遅延器のみで構成されるため、時間領域においてはサンプル点間の差分をとることで実現可能であり、全体の処理に必要な演算量の削減、処理系の単純化が期待できる。

本研究では、 $C, C\#, \dots, B$ 音に対応するくし形フィルタを縦続接続することで音高推定システムを構成する。この音高推定システムでは、12 種類のくし形フィルタの縦続接続において、零出力を検出した段のフィルタを随時先頭に移動させることにより、入力音高の数、種類の判別を行なう。本方式の特徴は、楽器音に含まれる全ての周波数成分を除去することで音高の判別を行なうことにより、入力された各音に含まれる周波数 (高調波) 成分の数や振幅に依存せず音高推定が可能となる点である。このため、各楽器音の周波数成分が複雑な重複を持ち、基本周波数の判別 (音高の推定) が困難とされていた異種複数楽器音に対しても、単一楽器音と同様な処理で音高推定が可能であると考えられる。

次に、異種複数楽器音の採譜に必要な和音分離、楽器推定法を検討する。本方式では、前述と同様にくし形フィルタの周波数成分除去の性質を利用し、時間領域において和音分離を行なうシステムを構成する。このシステムでは、異種複数楽器音中の 1 音に注目し、それ以外の音に対応するくし形フィルタを利用して不必要な音高を除去することにより、単音 (単一楽器音) の抽出を行なう。抽出された各音に含まれる周波数成分は、その音の基本周波数成分と高調波に制限されているため、非常に少ないサンプル点数 (サンプリング周波数 54.054 [Hz] で C 音に対して 413 点 ~ B 音に対して 219 点程度) の DFT によりスペクトル解析が可能となる。本方式では、予め各楽器の単音のスペクトルをテンプレートとして計算機に保存し、上述の分離音のスペクトルと比較することで楽器推定を行なう。入力された異種複数楽器音を予め単音に分離しているため、各スペクトルの正規化が可能であり、これにより楽器推定のためのマッチング処理を単純化できるという利点を持つ。本方式では、正規化に各スペクトルのノルム=1.0 を使い、テンプレートスペクトルと分離音のスペクトルの絶対値差分を計算することで比較を行なう。同一楽器音のスペクトルは類似した形を持つため、差分の最小値を検出することで楽器推定が可能である。

本研究では、採譜対象音域をオクターブ3のC (C^3 : 130.81[Hz]) からオクターブ5のB (B^5 : 987.77[Hz]) の36音とし、複数の楽器が異なる音高を同時に演奏した場合の異種複数楽器音の解析を行なう。なお、この36音はオーケストラ等において主旋律を演奏する楽器に対応しており[44]、現段階では、この範囲に音域を制限することで提案法の性能評価を行なうことを目的とする。また、入力としては電子楽器音(MIDI:KORG05R/W)を用いる。このシステムは、パーソナルコンピュータのディスプレイ上に所望の楽譜を作成し、それに基づきシンセシス・システムからアンプ内蔵のスピーカに音を増幅して出力するものである[43]。この音源は、実際の楽器音をサンプリングして得られた波形をデータとして保存しているため、音量の時間的変化、音色等、実楽器音を忠実に再現している[43],[47]。このため、実楽器音に含まれる波形の立上り直後の強い非調和周波数成分、ビブラート等による周波数変動等[38]、周波数解析に不利となる条件についても忠実に再現されているため、より実用的なシステムの構築に役立つものと考えられる。

本論文では、上述のMIDI音源により、ピアノ音、バイオリン音、ビオラ音、フルート音、クラリネット音、ホルン音、トランペット音を発生させ、これらの楽器音に対する音高推定及び楽器推定を行なう。なお、ピアノは打弦楽器(弦をハンマーで叩いて発音する楽器)であるため、その波形は減衰を示し、バイオリン、ビオラは弓奏弦楽器であるため、その周波数成分は微量ながら変動する特徴的な楽器である。これらの弦楽器は、楽音に含まれている高調波成分の数が多く、一般に解析が難しいとされている楽器である。一方、フルート、クラリネットは木管楽器で高調波成分の数が少なく、また、トランペット、ホルンは金管楽器で高調波成分の数は多いながらも、各管楽器ともに波形はほぼ定常であるという性質を持つ[44]。このように、同じ仲間(弦楽器同士、木・金管楽器同士)の楽器による和音や異なる仲間(弦楽器+木管楽器、弦楽器+金管楽器、等)の楽器による和音を解析し、本方式の性能評価を行なう。

1.3 論文概要

以下に本報告の概要を示す。

第1章は緒言であり、本研究の背景及び目的を示す。

第2章では、自動採譜システムに関する基礎的考察について述べる。この章では、初めに自動採譜システムに必要なタスクについて述べ、次に、楽器音の特徴についての説明を行なう。さらに、異種複数楽器音の採譜に必要な処理について述べ、最後に、第3章で述べる本方式との比較のために、既存の音高(オクターブ及び音名)、楽器の推定方法を紹

介し、その問題点について述べる。

第3章では、くし形フィルタによる音高、音価(音の継続時間)推定の原理について説明する。初めに、縦続接続したくし形フィルタによる音高推定システムについて述べる。ここでは、本システムで用いるくし形フィルタの構成原理及びその性質について述べ、サンプリング周波数の異なるくし形フィルタを縦続接続する際の、オーバーサンプリング法について説明する。さらに、各音高に対応したくし形フィルタを縦続接続し、各段の零出力を検出することで音高検出を行なう方式について説明する。最後に、くし形フィルタの零・非零出力の検出による音価推定法について説明する。

第4章では、くし形フィルタとスペクトル構造を利用した楽器推定法について説明する。ここでは、初めに、楽器推定のためのテンプレート作成方法について説明する。次に、くし形フィルタによる和音分離システムについて述べ、DFTによる分離音のスペクトル解析法及びテンプレートマッチングを用いた楽器推定法について説明する。さらに、くし形フィルタによる和音分離の問題点について明らかにする。最後に、周波数成分の重複の形により和音をパターン分けし、各々の和音に対して楽器推定法(スペクトル抽出法)の検討を行なう。

第5章では、本方式を用いた計算機シミュレーション結果及び考察を述べる。初めに、シミュレーションにおける条件を明らかにし、本システムの概要を説明する。次に、入力が単一楽器音の単音、和音、異種複数楽器音の場合に分け、本方式を用いた音高、音価推定結果について述べる。さらに、様々な周波数重複パターンを持つ異種複数楽器音について、和音分離、楽器推定結果を示す。最後に、単一楽器音、及び、異種複数楽器音の旋律に対する採譜結果を示す。

第6章は結言であり、本研究のまとめと残された課題について述べる。

第 2 章

自動採譜に関する基礎的考察

本章では、自動採譜に関する基礎的な事柄についての説明を行なう。初めに、楽曲と音楽的な表記法及び採譜の関係について簡単に述べ、自動採譜を行なう際に要求されるタスクについて説明する。次に、本研究の処理対象である楽器音について、異種複数楽器音の解析を行なう際に注目すべき特徴を説明する。さらに、自動採譜システムにおける音高、音価、楽器の推定について、それぞれに必要な処理の説明を行なう。最後に、既存の異種複数楽器音に対する採譜アルゴリズムを説明し、従来法における問題点を明らかにする。

2.1 自動採譜のタスク

一般に音楽は、楽譜に記されている音符、階名に従って楽器を演奏することで表現される。ここで、音符とは四分音符、八分音符、十六分音符、等の音の長さを表す記号であり、階名とはド、レ、ミ等の、一般的にドで表現される主音からの相対的な音の位置（音の種類）である。これらは、楽譜の五線譜上における音符の種類、及び、その位置により表される。採譜とは演奏と逆の手順であり、様々な楽器で演奏された音の長さ、音の高さ、楽器の種類を聞き分け、五線譜上に楽譜の形式で書き表す作業である。図 2.1 に演奏と採譜の関係を示す。

計算機を用いて異種複数楽器音の採譜を行なうためには、入力された音楽信号に対して

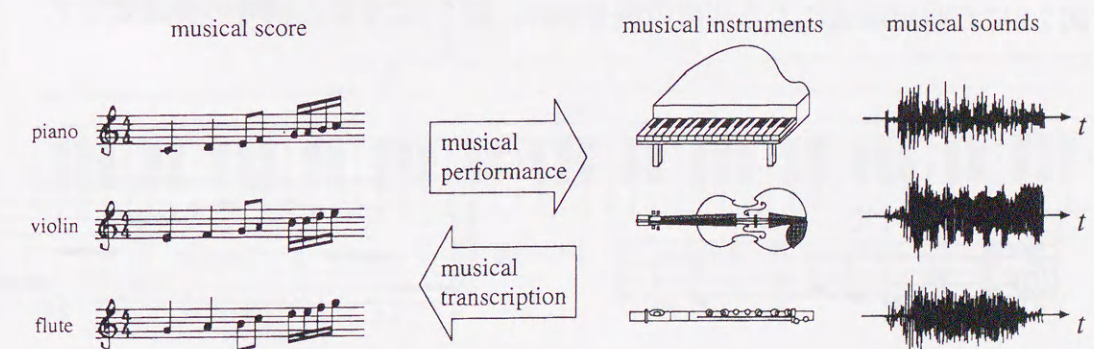


図 2.1 演奏 (musical performance) と採譜 (musical transcription) の関係

演奏された音符、階名、楽器の種類を推定した後、人間に分かる形式、即ち、楽譜形式に書き表す必要がある。なお、一般的な楽譜は、演奏される曲の速さや演奏楽器の種類により、以下のような異なる基準を持つ。

- テンポ：1 分間に演奏される基本の拍数で曲の速さを表したもの。例えば、4/4 拍子や 3/4 拍子の $tempo = 60$ ならば、四分音符（基本となる音符）が 1 分間に 60 個演奏される速さとなり、四分音符は 1.0[s] の長さを持つ。一方、2/2 拍子の $tempo = 60$ ならば基本となる音符は二分音符となり、四分音符の長さは 0.5[s] となる。
- 主音：階名表現上で基準となる（ドと表現する）音。楽器の種類などの条件により異なる場合がある。例えば、ハ調の楽器であるピアノにおいてドと呼ばれる音が C であるのに対し、トランペットのドの音は A# であり、フレンチホルンでは F である。異種複数楽器音を各楽器に対応する楽譜に書き直す場合には、主音の考慮が必要となる。

上記のような楽譜表記上の違いによる問題は、入力された音楽信号を解析し、音の高さ、音の長さ、演奏楽器の種類、等を推定した後、計算機においてソフトウェア的な処理を行なう必要がある。このため、一般的な採譜アルゴリズムにおいては、楽譜表記上の階名や四分音符、八分音符などの音符の代わりに、

1. 音高：各音に含まれる基本周波数。一般的に、音の種類を表現する固有名である音名（C, C#, ..., B）とオクターブ番号の 2 つを用いて表現される。例えば、オクターブ 4 の A（ピアノのラ）音は $A^4 = 440.0[\text{Hz}]$ である。

2. 音価：演奏された音の立ち上がりから立ち下がりまでの継続時間。

の 2 種類のパラメータ値を用いた表現方法が一般的である。この表記方式を奏譜という。図 2.2 に五線譜形式による楽譜の例を示し、図 2.3 にそれと同等の奏譜を示す。

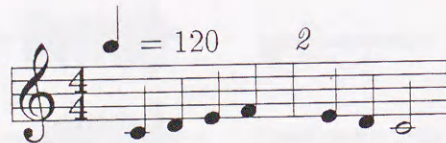


図 2.2 五線譜形式による音楽表現

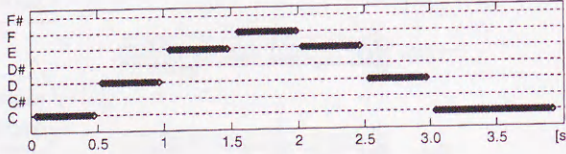


図 2.3 奏譜形式による音楽表現

2.2 楽器音の特徴

本節では、異種楽器という点に注目して、楽器音の様々な特徴について説明する。

各々の楽器は、発音機構、共鳴機構、音域、等の違いにより、周波数成分や振幅変化（振動、減衰、等）に異なる特徴を持つことが知られている [44]。人間が複数の楽器音を同時に耳にした際、楽器についての予備知識をほとんど持っていない場合でも、何種類の楽器による演奏であるかを判別することができるのは、人間の耳が無意識のうちに上記の特徴を識別しているためである [46], [47]。演奏された異種複数楽器音に対して、計算機を用いて楽器の判別を行なうためには、各音の成分を個別に認識し、それらの特徴を抽出することにより、どの音が何の楽器によるものであるかを判別する必要がある。

表 2.1 主な楽器の種類

種類	楽器名
弾奏弦楽器	ギター、ハープ、...
打弦楽器	ピアノ、...
弓奏弦楽器	バイオリン、ビオラ、チェロ、...
機械&エアリード	オルガン
木管楽器（単式リード）	クラリネット、サキソフォン、...
木管楽器（複式リード）	オーボエ、バスーン、...
木管楽器（エアリード）	フルート、ピッコロ、...
金管楽器（リップリード）	トランペット、ホルン、チューバ、...

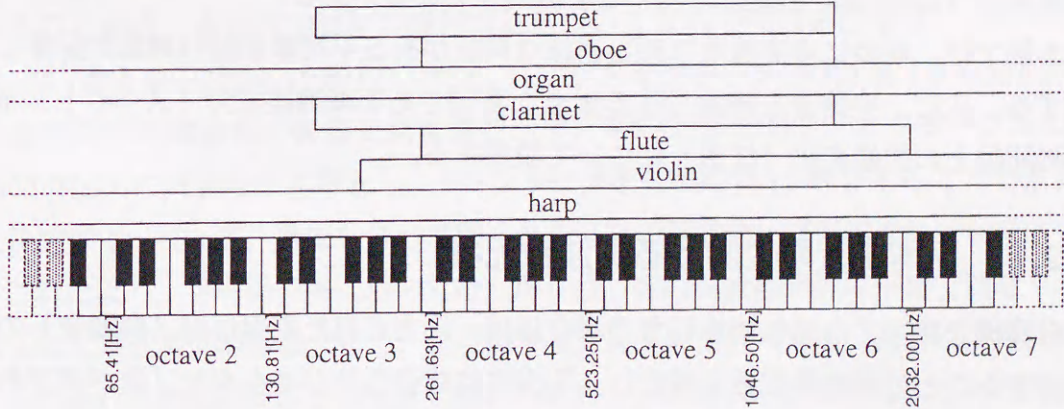


図 2.4 代表的な楽器の音域

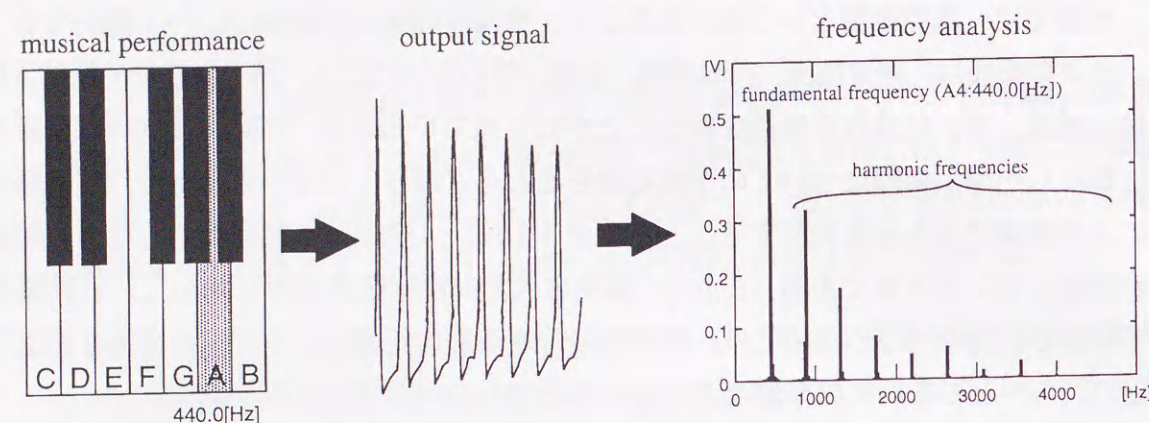


図 2.5 楽器音の性質

一般に知られている主な楽器について、楽器の構造（演奏方法）により大別したものを表 2.1 に示し、代表的な楽器の音域を図 2.4 に示す [44]。なお、5 章における計算機シミュレーションで使用する楽器を表 2.1 中に太字で示した。

楽器音を採譜システムの入力信号と考えた場合に、その波形及び周波数成分が持つ特徴について、図 2.5 に模式的に示す。時間軸上で楽器音を観察した場合、その振幅は時間により変化する非定常な特性を持ち、さらに、その形状は正弦波のように単純ではなく、複数の高調波成分による合成波となる。また、楽器の音高は基本周波数により決定されるため、周波数解析を行なう場合には、基本周波数成分（fundamental frequency）とそれ以外の高調波成分（harmonic frequencies）とを区別する必要がある。

本節では、初めに音高推定処理と密接な関係がある平均率音階の周波数配置について説明する。次に、楽器音の特徴付ける大きな要因である高調波成分（スペクトル構造）、及び時間軸上の振幅変化（波形）について説明する。

2.2.1 平均率音階における周波数配置

計算機を用いて自動採譜を行なう場合には、入力された音楽信号（楽器音）の周波数を分析することで基本周波数を推定し、入力された音名及びオクターブ数が何であるかを判別する必要がある。この基本周波数と音名の関係を表す音階率は、1 オクターブ 12 音の周波数配置によって純正律音階、平均律音階、等がある [44]。

純正律音階は、調和級数となる比較的小さな整数の比で表される周波数間隔を用いた音

表 2.2 純正律音階における各音階の周波数比（主音 C）

音階	C	D	E	F	G	A	B	C
周波数	f	$\frac{9}{8}f$	$\frac{5}{4}f$	$\frac{4}{3}f$	$\frac{3}{2}f$	$\frac{5}{3}f$	$\frac{15}{8}f$	$2f$

階である。つまり、1 オクターブ中の各音の周波数は、主音（基準となる音）の周波数に基づき、定められた分数比となるように決定される。例として、主音を C とした場合の純正律音階の周波数比を表 2.2 に示す。なお、ハ調（主音 C）の周波数は A^4 を 440.0[Hz] として計算されており、他の調の主音周波数はハ調の各音の周波数から決定される。

純正律音階では音楽で用いられる種々の和音（半音、長音、長三度、等）を単純な整数比で表現しているため、曲の調を決定する主調音が異なる場合には、楽譜上では同じ音でも実際の周波数が異なるという問題が生じる。例えば、ハ調（主音 C）のオクターブ 4 のラが $A^4=440.0$ [Hz] であるのに対し、ト調（主音 G）のオクターブ 4 のレ (A'^4) は

$$\begin{aligned}
 A'^4 &= \frac{9}{8} \times G^4 \\
 \text{ハ調の } G^4 &= \frac{3}{5} \times A^4 \times \frac{3}{2} = 396.0 \\
 &= \frac{9}{8} \times 396.0 = 445.5 \text{ [Hz]} \quad (2.1)
 \end{aligned}$$

となり、同じ A^4 でありながら異なる基本周波数を持つことが分かる。

オーケストラの演奏等では、和音の響きの美しさを求めるために意図的に純正律に近い和音演奏を行なう場合もあるが、ピアノ等の固定ピッチの楽器は一般的に平均律音階を用いていることから、本研究では平均律音階に基づいて音名の判別を行なう。

表 2.3 に、オクターブ 0 の C 音 ($C^0:16.35$ [Hz]) からオクターブ 7 の B 音 ($B^7:3951.07$ [Hz]) までの範囲の平均律音階の各音高周波数を示す。平均律音階では、オクターブ 4 の A 音 ($A^4:440.0$ [Hz]) の周波数を基準とし、1 オクターブを対数的に 12 等分することで他音の周波数を決定する。従って、半音も含めて隣り合う音高の周波数比は常に $2^{\frac{1}{12}} \cong 1.059 \dots$ 倍となる。これにより、平均律音階においては、隣り合う音の周波数間隔は一定ではなく、低音域になるほど周波数間隔は狭くなり、逆に高音域になるほど広がる。例えば、オクターブ 2 の C と C# の周波数間隔が 3.89[Hz] であるのに対し、オクターブ 6 の A# と B の周波数間隔は 110.87[Hz] となっており、その周波数間隔には 100[Hz] 以上の違いがあることがわかる。

表 2.3 国際標準周波数による平均律音階 [Hz]

ハ長音階 (音名)	オクターブ番号							
	0	1	2	3	4	5	6	7
ド (C)	16.35	32.70	65.41	130.81	261.62	523.25	1046.50	2093.00
ド# (C#)	17.32	34.65	69.30	138.59	277.18	554.37	1108.73	2217.46
レ (D)	18.35	36.71	73.42	146.83	293.66	587.33	1174.66	2317.46
レ# (D#)	19.45	38.89	77.78	155.56	311.13	622.25	1244.51	2489.02
ミ (E)	20.60	41.20	82.41	164.81	329.63	659.26	1318.50	2637.02
ファ (F)	21.83	43.65	87.31	174.61	349.23	698.46	1396.91	2793.83
ファ# (F#)	23.12	46.25	92.50	185.00	369.99	739.99	1479.98	2959.96
ソ (G)	24.50	49.00	98.00	196.00	392.00	783.99	1567.98	3135.96
ソ# (G#)	25.96	51.91	103.83	207.65	415.30	830.61	1661.22	3322.44
ラ (A)	27.50	55.00	110.00	220.00	440.00	880.00	1760.00	3520.00
ラ# (A#)	29.14	58.27	116.54	233.08	466.16	932.33	1864.66	3729.31
シ (B)	30.87	61.74	123.47	246.94	493.88	987.77	1975.53	3951.07

2.2.2 楽器音に含まれる高調波成分

前述のように、楽器音には音高を示す基本周波数成分 (fundamental frequency) の他に、多数の高調波成分 (倍音: harmonic frequencies) が含まれている。この高調波成分の数や振幅、位相 (スペクトル構造) は、楽器の発音機構、共鳴機構により決定されるため、楽器の種類により様々な特徴を持つことが知られている [44]。

楽器音のスペクトル構造の違いを示す例として、ピアノ、バイオリン、フルートのオクターブ 4 の A 音 ($A^4=440.0[\text{Hz}]$) に含まれる周波数成分 (パワースペクトル) を図 2.6 に示す。

一般に、ピアノやバイオリンのような弦楽器は高調波成分の数が多く、また、それらが比較的大きな振幅を持つことが知られている。特に、図 2.6(b) に示すバイオリン音は、ビブラートの影響から周波数変動が生じており、高調波成分のスペクトルが広がっていることが観察できる。また、同図 (a), (b) に示すピアノ、バイオリン共に、非調和周波数成分 (線スペクトルの形で現れる高調波成分以外の雑音成分) が含まれていることが分かる。一方、同図 (c) に示すフルート音は、高次の高調波成分の振幅が小さく、非調和周波数成分をほとんど含まないことが分かる。このように、各楽器音のスペクトル構造は、楽器音の特徴付ける重要な要素であると考えられる。

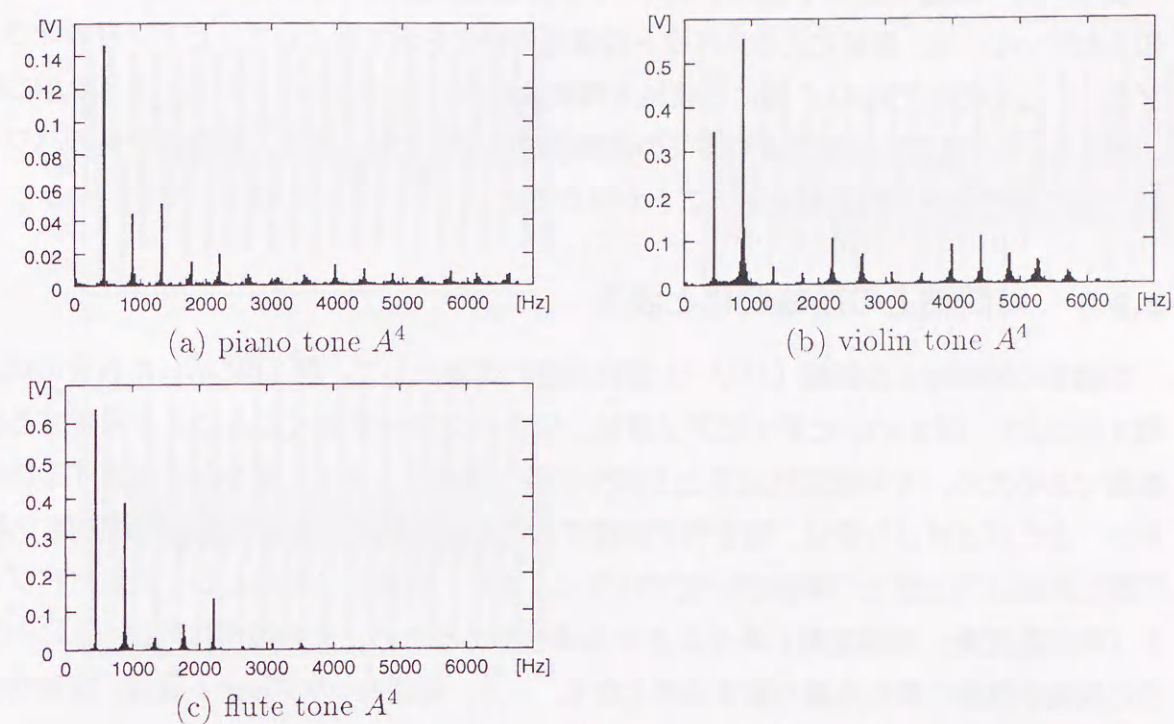


図 2.6 各楽器音のスペクトル構造の違い

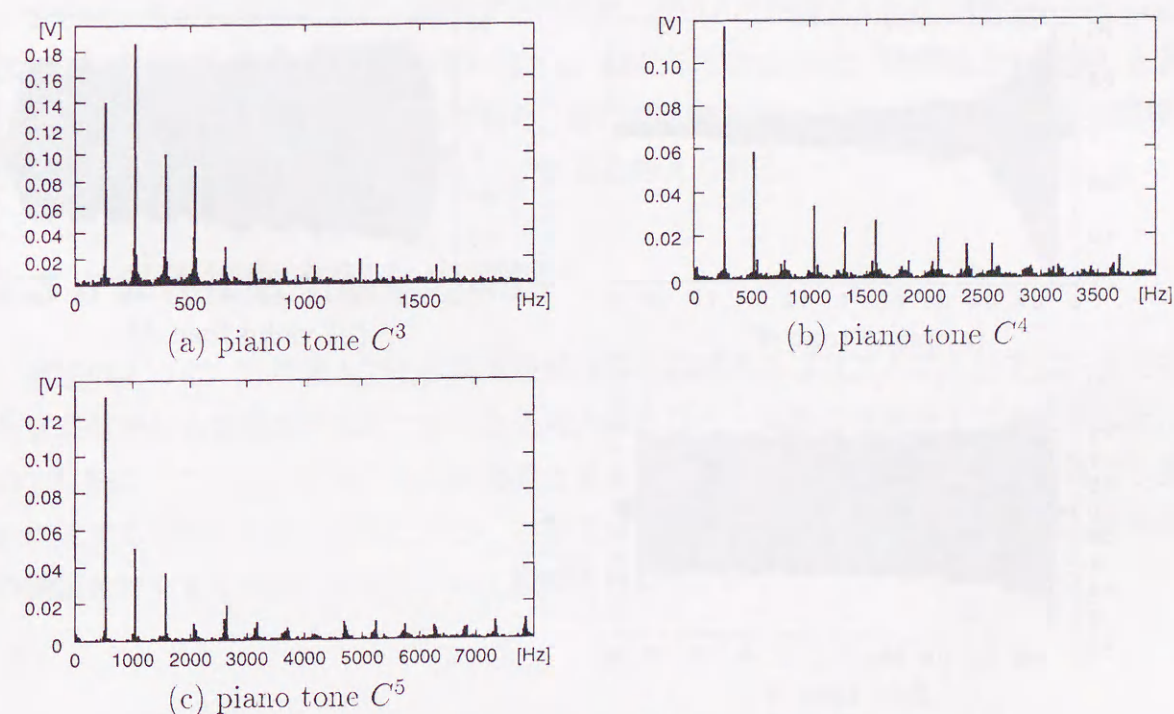


図 2.7 音域によるスペクトル構造の違い

また、同一楽器であっても、そのスペクトル構造は音域により若干の違いを持つことが知られている[44]。音域によるスペクトル構造の変化を示す例として、ピアノ音のオクターブ3, 4, 5のそれぞれのC音に含まれる周波数成分（パワースペクトル）を図2.7に示す。

図2.7より、低音には高音より多くの高調波成分が含まれ、また、楽器音全体のパワーに対する高調波成分の割合が大きいことが分かる。

2.2.3 時間軸上の振幅変化と波形

楽器音の時間軸上の振幅（パワー）変化の違いの例として、図2.6に示した各音の波形を図2.8に示す。図2.8(a)に示すピアノ音は、弦をハンマーで叩くことにより発音する打弦楽器であるため、その波形は立ち上がり部分が最も振幅が大きく、緩やかに減衰する特徴を持つ。またバイオリン音は、弦を弓で摩擦することにより発音する弓奏弦楽器の例である。弓奏弦楽器は弓と弦との摩擦が一定ではなく、また、演奏上の効果として故意にビブラート（周波数変動、振幅変動）を生じさせる場合があるため、その波形は同図(b)に示すように振幅が強弱の変化を繰り返す波形となる。一方、同図(c)のフルート音は、定常空気流により発音する木管楽器（正確にはエアリード楽器）の例である。フルート音は息を吹き込むことにより発音させるため、発音時のアタックの部分の振幅が若干大きい以外は、そ

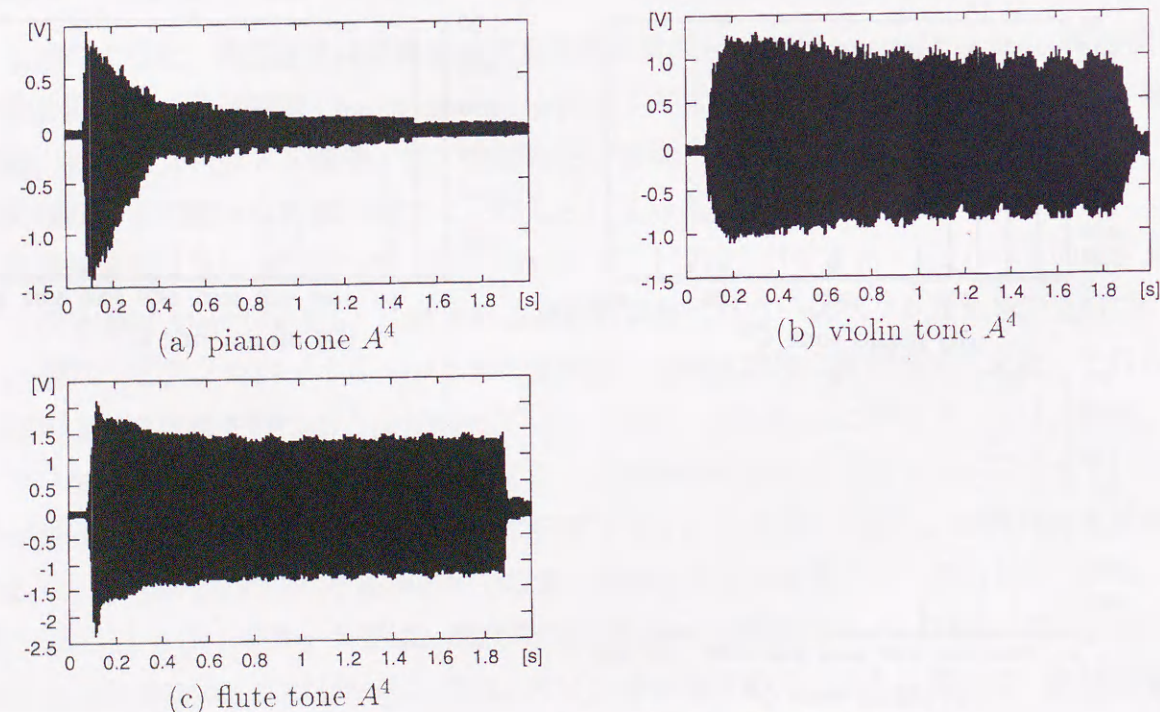


図2.8 各楽器音の時間軸上の振幅変化

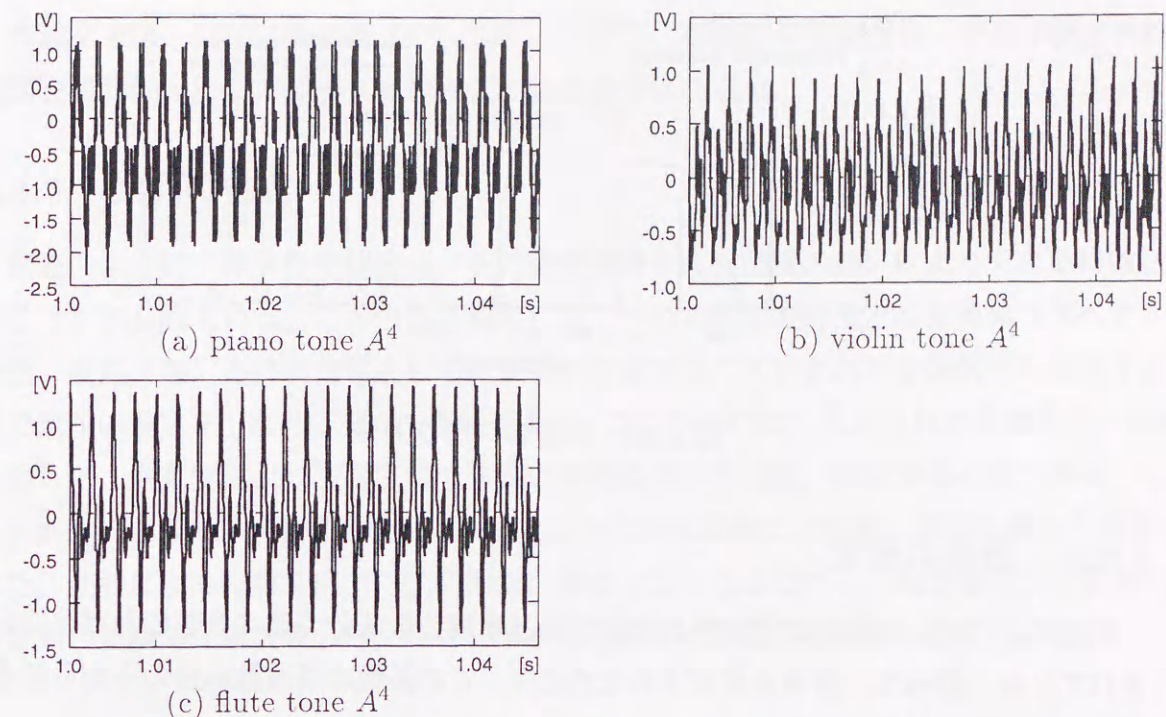


図2.9 各楽器音の波形

の波形はほぼ定常信号であるという特徴を持つ。

さらに、各々の波形を詳しく観察するために、図2.8における1.0[s]の時点から50[ms]の区間の波形を図2.9にそれぞれ示す。図2.8, 2.9に示したように、時間軸上で観察した波形は楽器ごとに異なった特徴を持っており、前述の高調波成分（スペクトル構造）と同様に、各楽器音を区別する重要な要素の1つであると考えられる。

2.3 異種複数楽器音の採譜

前述のように、計算機上での異種複数楽器音の採譜アルゴリズムにおいては、入力された音楽信号から奏譜の作成までが主な処理内容となり、実際の五線譜上での記号表記については別のソフトウェア的な処理を必要とする。従って、本論文では、音楽信号から奏譜を作成する作業を大きく採譜と呼び、本節では、このために必要な音高、音価、演奏楽器の種類を判別する処理の内容について説明を行なう。

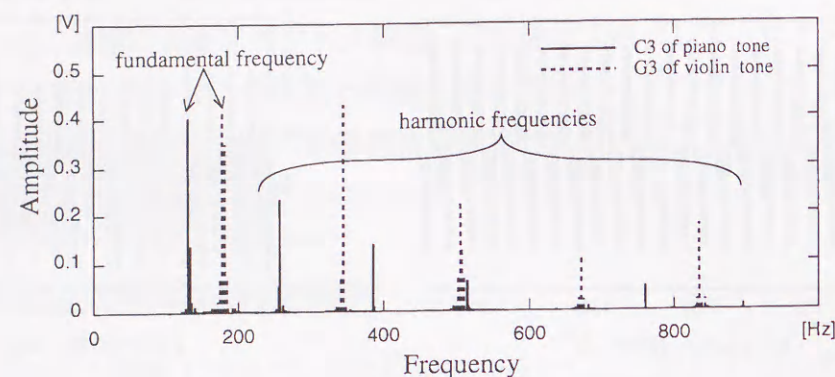


図 2.10 和音のスペクトル

2.3.1 音高の推定

前述のように、楽器音には音高を決定する基本周波数成分の他に多数の高調波成分が含まれている。従って、音高を推定するためには、この多数の周波数成分の中から基本周波数成分を検出する必要がある。

しかし、前述の平均率音階の周波数関係から、図 2.10 に示すように、音楽表現上の多重音（和音）に含まれる周波数成分は周波数軸上で非常に複雑な重なりを持つ。また、楽器により高調波成分の数や振幅が異なっていることから、和音全体のスペクトルにおいて、必ずしも基本周波数成分の振幅が突出して大きいとは限らないという特徴がある。このため、音高の推定（基本周波数の検出）は、計算機を用いた自動採譜システムの研究において、大きな課題の 1 つとされている。

2.3.2 音価の推定

入力された音楽信号に含まれる各音高成分について、立ち上がりから立ち下がりまでの継続時間を計測することで音価の推定を行なう。一般的には、閾値により音の有無を判別する方式が用いられる。

なお、前述のように五線譜上に楽譜として表記するためには、音楽信号における基本周波数成分の継続時間や出力分布等から、予め楽曲独自のテンポ推定しておく必要がある。このための手法として、これまでに、音楽信号のパワーを微分して各音の立ち上がりを抽出し、その間隔時間のヒストグラムを作成した上で人間が自然と感じる範囲の基準拍を決定する手法や [7]、あらかじめリズムパターンを作成し、曲中のどこに強拍がくるかというこ

とをトップダウン的に探ることにより、テンポの判別を行なう手法が提案されている [8]。

本論文では、縦軸に各音高 ($C^3, C\#^3, \dots, B^5$)、横軸に時間を取り、音高が検出された時間を時間軸上に示す方式（奏譜形式）により音価を示す。

2.3.3 楽器の推定

前述のように、様々な楽器音は、その発音機構や共鳴機構の違いにより異なる特徴を持つことが知られている。この性質を利用して、予め各楽器の特徴となる要素（スペクトル構造、波形、等）を事前情報として計算機内に保存し、入力された音楽信号と比較することで楽器の推定を行なう方式が一般的である。このためには、入力された音楽信号（和音）に対して、各楽器により演奏された音高の周波数成分を分離、抽出する必要がある。しかし、前述のように、和音に含まれる周波数成分は周波数軸上で非常に複雑な重なりを持ち、また、楽器により高調波成分の数や振幅が異なっていることから、混合音和音に含まれる各音高の認識や高調波成分のグルーピング等は非常に困難とされている。

2.4 既存のアルゴリズム

本節では、計算機を用いた自動採譜システムにおける既存のアルゴリズムについて簡単に説明し、各々の手法における問題点を明らかにする。

2.4.1 音高推定法

1. 離散フーリエ変換による音高推定

計算機で採譜を行なう際の周波数分析法として最も一般的に用いられている手法は、離散フーリエ変換 (DFT: Discrete Fourier Transform) の高速演算法として提案された高速フーリエ変換 (FFT: Fast Fourier Transform) である [1],[8],[11]。以下に、DFT の基本原理及び音楽信号に適用する際の問題点について述べる。

DFT では、時間領域上の N 個のサンプル点 $x(0), x(1), x(2), \dots, x(N-1)$ に対する周波数スペクトル $X(e^{j\omega_k})$ が式 (2.2) で与えられる。なお、 T はサンプリング周期であり $1/fs$ (fs : サンプリング周波数) である。

$$X(e^{j\omega_k}) = \sum_{n=0}^N x(n) e^{-j\frac{2\pi k}{N}n} \quad (2.2)$$

ここで、 ω_k は周波数軸上の N 点の離散周波数であり式 (2.3) を意味している。

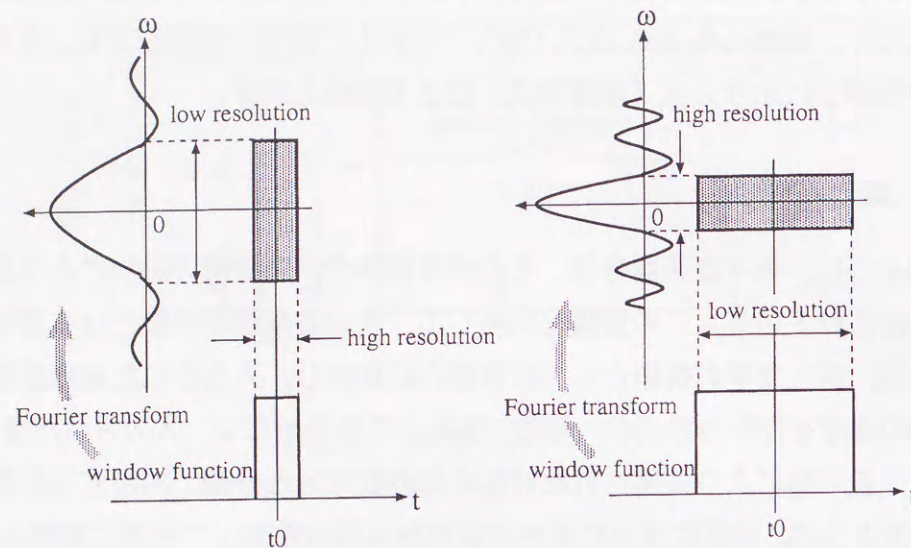


図 2.11 DFT の時間分解能と周波数分解能の関係

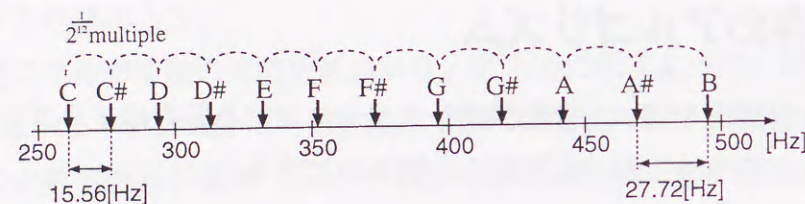


図 2.12 オクターブ 4 の周波数の周波数配置

$$\omega_k = \frac{2\pi k}{NT} = \frac{2\pi f_s}{N} k = \omega_1 k \quad (2.3)$$

式(2.3)より、DFTで解析可能な周波数は基本波周波数 ω_1 の整数倍に制限されていることが分かる。また、DFTの時間-周波数分解能は図2.11のようなトレードオフの関係を持ち、周波数分解能を上げるためには、大きな N を必要とする。

しかし、音楽表現上の短い音符(4/4拍子 tempo = 120における十六分音符で125[ms]程度)の場合には、十分な周波数分解能を得るためのデータ長が得られないことも考えられる。また、音楽の分野で用いられる平均律音階は、前述のように $2^{1/12} \cong 1.059 \dots$ 倍の周波数比を持ち、図2.12に示すような不等間隔な周波数配置を持つ。このような問題から、DFTを用いて音楽信号の周波数解析を行なう場合には、時間、周波数分解能の性質を考慮し、周波数補正等の特別な処理が必要となる[4]。

2. 適応フィルタによる音高推定

DFTのように解析周波数に制約を持たない音楽信号の解析法として、適応フィルタによる方式が提案されている[9]。一般的な適応フィルタは、図2.13に示すような、可変タップ重みを有するFIRフィルタとそれを制御する適応アルゴリズムから構成される。以下に、適応フィルタによる周波数解析法の原理を説明する。

サンプル値系列 $x(0), x(1), x(2), \dots, x(N-1)$ が $\omega < \omega_l$ と $\omega > \omega_h$ の区間で周波数成分を持たない帯域制限周期波形であると仮定し、フーリエ係数 $X(e^{j\omega_k})$ を用いて次式のように表現する。

$$x(n) = \sum_{k=l}^h [X(e^{j\omega_k})e^{j\frac{2\pi k}{N}n} - X(e^{-j\omega_k})e^{-j\frac{2\pi k}{N}n}] \quad (2.4)$$

ここで、 $W(k)$ を式(2.5)のように定義し、式(2.4)を式(2.6)のように書き改める。なお、 X は各フーリエ係数 $X(e^{j\omega_k T})$ を要素として持つ L 次元ベクトルであり、 \bar{W} 、 \bar{X} はそれぞれのベクトルの複素共役である。

$$W(k) = [e^{j\frac{2\pi l}{N}k}, \dots, e^{j\frac{2\pi i}{N}k}, \dots, e^{j\frac{2\pi h}{N}k}]^T \quad (2.5)$$

$$\begin{aligned} x(n) &= W^T(n) \cdot X - \bar{W}^T(n) \cdot \bar{X} \\ &= X \cdot W^T(n) - \bar{X} \cdot \bar{W}^T(n) \end{aligned} \quad (2.6)$$

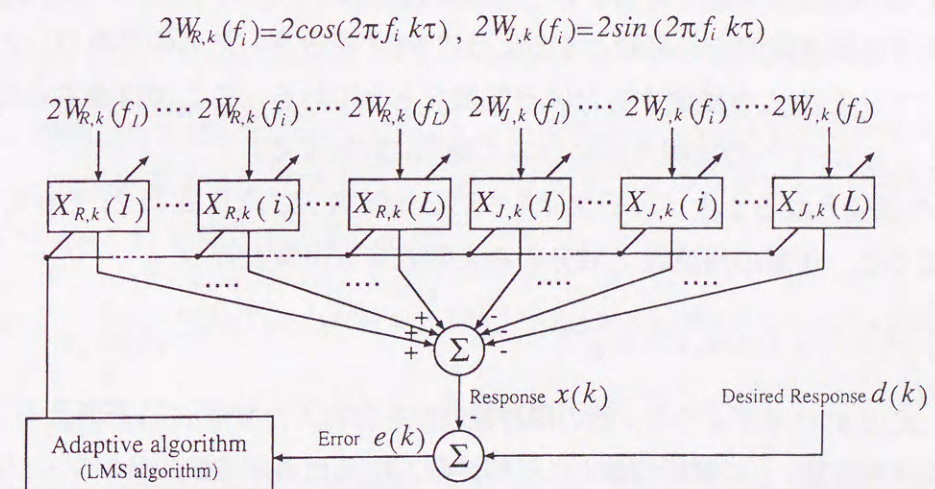


図 2.13 一般的な適応システム

適応フィルタを用いた周波数解析法では、式(2.6)における $\mathbf{W}(n)$ を入力、 \mathbf{X} を重みベクトル、 $x(n)$ を出力と考えることで解析を行なう。

解析信号を $d(n)$ とすると、瞬時誤差 $\epsilon(n)$ は、 $d(n)$ とフィルタ出力 $x(n)$ との差から次のように表すことができる。

$$\epsilon(n) = d(n) - x(n) = d(n) - \mathbf{X} \cdot \mathbf{W}^T(n) + \bar{\mathbf{X}} \cdot \bar{\mathbf{W}}^T(n) \quad (2.7)$$

この $\epsilon(n)$ による2乗平均誤差 $E[\epsilon(n)^2]$ を最小にするようにフィルタ係数を逐次更新することで、最適なフィルタ係数 \mathbf{X} を求める。ここでのフィルタ係数 \mathbf{X} はフーリエ係数を表していることから、この手法では、DFTのような周波数補正等の特別な処理を必要とせず、最適(任意)な周波数解析が可能であることを示している。

しかし、この手法では周波数成分が存在する帯域に制限があり、多数の高調波成分が重畳する和音の処理が困難であるという問題が生じる。また、2乗平均誤差 $E[\epsilon(n)^2]$ を最小にするために勾配法を用いたベクトル演算が必要であり、計算量の増大が問題となる。

3. 拡張ノッチフーリエ変換による音高推定

本研究室では、FFTのように解析周波数に対する制約がなく、少ないサンプル点から周波数解析が可能な拡張ノッチフーリエ変換(E-NFT: Extended Notch Fourier Transform)を用いた音高推定法を提案した[10]。E-NFTの基本アルゴリズムであるノッチフーリエ変換(NFT: Notch Fourier Transform)は、入力信号に含まれる周波数の数 r 及びそれぞれの周波数値 f_p ($p = 1, 2, \dots, r$) が既知である信号に対し、 $2r$ 元1次方程式を解くことで振幅、位相の推定が可能であるという特徴を持つ[39]。E-NFTは、 r と各 f_p が未知な入力信号に対する周波数解析を可能とするようにNFTを改良したものであり、ノッチフィルタに基づいて r と各 f_p を推定するという特徴をもっている。ここでは基本となるNFTの原理を示し、E-NFTへの拡張方法について簡単に説明する。

NFTの基礎となる2次ノッチフィルタの伝達関数 $H_p(z)$ は式(2.8)で与えられ、図2.14に示すように、任意の周波数 f_p 成分のみを除去する性質を持つ。

$$H_p(z) = 1 - 2 \cos \theta_p z^{-1} + z^{-2} \quad (2.8)$$

ここで、式(2.9)に示すような r 個の周波数成分を含む入力信号 $x(n)$ を考える。なお、 f_s はサンプリング周波数、 f_p は解析対象となる周波数、 a_p, b_p は各周波数成分のフーリエ係数である。

$$x(n) = \sum_{p=1}^r (a_p \cos \theta_p n + b_p \sin \theta_p n) \quad \text{ここで、} \quad \theta_p = \frac{2\pi f_p}{f_s} \quad (2.9)$$

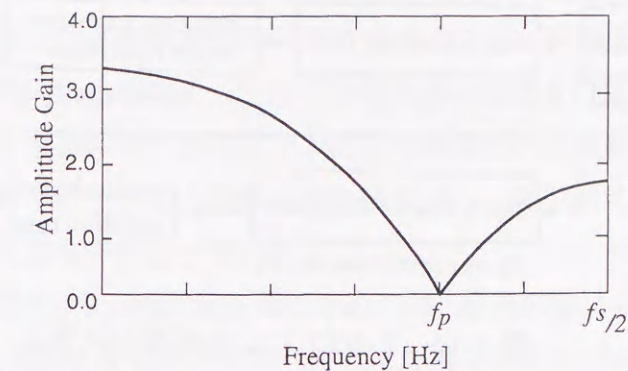


図2.14 ノッチフィルタの周波数特性

図2.14に示す周波数成分除去の性質を利用し、式(2.10)に示すように、 $H_p(z)$ ($p = 1, 2, \dots, r, \neq k$) を縦続接続することで、入力 $x(n)$ から所望の周波数 f_k 成分のみを抽出することが可能である。ここで、合成フィルタ $G_k(z)$ の出力 $y_k(n)$ は式(2.11)で表現される。

$$G_k(z) = \prod_{i=1, \neq k}^r H_i(z) = \prod_{i=1, \neq k}^r (1 - 2 \cos \theta_i z^{-1} + z^{-2}) \quad (2.10)$$

$$y_k(n) = A_k [a_k \cos \{n - (r-1)\} \theta_k + b_k \sin \{n - (r-1)\} \theta_k] \quad (2.11)$$

但し、 A_k は $G_k(z)$ の振幅利得であり、

$$A_k = 2^{r-1} \prod_{i=1, \neq k}^r (\cos \theta - \cos \theta_k) \quad (2.12)$$

である。

NFTは、連続した2点の出力 $y_k(n)$, $y_k(n+1)$ を利用し、式(2.13)に示すように、フーリエ係数 a_k , b_k を計算する手法である。なお、ノッチフィルタを $r-1$ 個縦続接続しているため、 $n \geq 2(r-1)$ を満たす必要がある。

$$\begin{aligned} a_k &= \frac{1}{A_k \sin \theta_k} [y_k(n) \sin \{(n-r+2)\theta_k\} - y_k(n+1) \sin \{(n-r+1)\theta_k\}] \\ b_k &= \frac{1}{A_k \sin \theta_k} [-y_k(n) \cos \{(n-r+2)\theta_k\} + y_k(n+1) \cos \{(n-r+1)\theta_k\}] \end{aligned} \quad (2.13)$$

上述のNFTを未知信号に対して拡張したものがE-NFTであり、NFTを実行する前に、ノッチフィルタ $H_p(z)$ に基づいて周波数推定を行なう必要がある。この周波数推定では、 $H_p(z)$ が $x(n)$ の周波数に対応している場合に出力 $y(n)$ が零となる性質に着目し、逆に、 $x(n)$

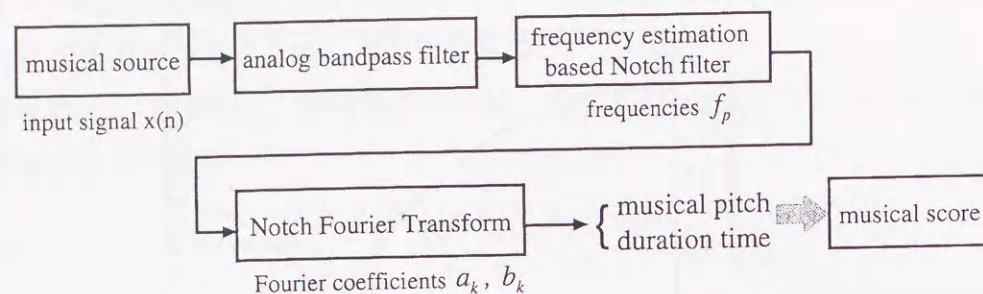


図 2.15 E-NFT による採譜システム

から出力を零にするようなフィルタ係数 $2 \cos \theta_p$ を算出することで、 θ_p から未知周波数の計算を行なう。ただし、このためには r 元連立 1 次方程式、及び、 r 次方程式を解く必要があるため、 r の増加に伴う計算量の増大、計算の複雑化、推定精度の劣化が問題となった。また、楽器音には多数の高調波成分が含まれているため、入力が単音の場合でも非常に多くの（未知周波数に対する）周波数推定を行なう必要があり、非効率的である。

このため、前処理として高次のアナログ型及びデジタル型のバンドパスフィルタを利用して音楽信号を各オクターブごとに分割し、並列に周波数推定処理を行なうことで上述の問題を回避している。しかし、前処理のバンドパスフィルタでは高調波成分を完全に除去することは困難であり、また、バンドパスフィルタによる前処理が加わったため、新たに処理系の複雑化といった問題が生じた。図 2.15 に E-NFT による採譜システムの構成を示す。

2.4.2 楽器推定法

異種複数楽器音の解析においては、和音の分離（音源分離）と楽器の推定（音源同定）は密接に関係しているが、処理的には明確に区別されている [42],[47],[48]。つまり、楽器固有の特徴を事前情報として必要とせず、各楽器を単に複数の音源とみなすことで、それらの分離のみを目的としたアプローチと、前述のように、楽器固有の特徴を事前情報として用意し、ある種のマッチング処理を行なうことで楽器推定を行なうことを目的としたアプローチの 2 種類が存在する。ここでは、上記の 2 種類の解析法について簡単に説明を行なう。

1. 複数センサによる音源分離

異種楽器和音の分離方法として、複数のマイクロフォンを用いて集音し、各音源（楽器）

の位置の情報を利用することによって目的の楽器音を抽出する方式が提案されている [23]。

この手法では、対象となる複数楽器の周辺にマイクロフォンを広く分布させてその焦点を一点に結ばせ、空間内の走査によって先ず目標楽器の位置を推定した後、目的の信号を抽出する。ここでは、複数のマイクロフォンにより集音された複数の楽器音から目標の楽器音を抽出するための手法として、加算平均を繰り返し、疑似的に対象となる楽器音が一つだけ存在するとした処理を行なっている。

しかし、この方法で十分な分解性能を得るためには、集音の際に多数のマイクロフォンを準備する必要があり、また、それらを特定の形状に配置することが必要であるため、システムの物理的規模が大きくなってしまったといった問題が生じる。また、各楽器及び雑音源の位置に関する制約が大きく、楽器の位置のずれに対するシステムの感度が、通常かなり高いという欠点がある。従って、音源の位置を利用する手法は、有効性のある方式ではあるが、実際上は、一般の場合にこれだけで十分な異種楽器音の分離を行なうことは難しいと考えられる。

2. 独立成分分析による音源分離

ある音源の各周波数成分の時間的変化が一定であると仮定し、独立成分分析（ICA : Independent Component Analysis）の一手法であるブラインドデコンポジションアルゴリズムを用いて各音源を分離する方式が提案されている [24]。これは、入力信号に含まれる全ての周波数成分の時間変化についての相関をとり、時間的に相似して振幅が変化する各周波数成分をグループとして 1 つの楽器とみなす方式である。以下に、ピアノ音とフルート音の和音を例に取り具体的に説明する。

前述のようにピアノ音は減衰波形であるため、この楽音に含まれる全ての周波数成分（基本周波数成分、高調波）は、時間軸上で観察した際にある一定の減衰を示す。一方フルート音は定常波形であるため、楽音に含まれる全周波数成分は時間的に定常である。この性質を利用し、短時間 FFT により和音に含まれる各周波数成分の時間変化についての相関を利用することで、各音源から発せられた高調波成分を分離することができる。この方式では、各楽器固有の特徴等の事前情報を必要としないという利点があるが、各音源の時間変動が等しい場合（単一楽器音の和音等）には適用できないという問題が生じる。

3. スペクトルのマッチングによる音源同定

前述のように、楽器音のスペクトルはそれぞれ異なる特徴を持っている。この性質を利用して、予め各楽器音の単音スペクトルをテンプレートとして用意し、マッチングを行なうことで楽器を推定する手法が提案されている [25],[29],[31]。これは、入力音に対して、楽

器，音程，発音時刻，位相等のパラメータセットに応じて複数の単音（テンプレート）を合成したものを比較に用いる．その結果，入力音との誤差が最小となるようなパラメータセットに含まれる楽器が演奏楽器であると判別される．

しかし，テンプレートの数や同時に演奏される音の数の増加に伴い，パラメータセットの組合せが爆発的に増加するため，莫大な計算量が必要となる問題がある．また，比較に用いるパラメータセットを決めるためには，音高や音量，各音の位相関係等の情報が必要となるが，前述のように，FFT等を用いる音高の推定には様々な問題が残されている．このため，和音中の各音の基本周波数を正確に抽出することが困難であり，結果的に，パラメータセットの総当たりの探索が必要となる場合が多い．

このための探索法（マッチング法）として，遺伝的アルゴリズムを用いる手法 [25] や，時間-周波数両方向のグルーピングを利用する手法 [29], [31] 等が提案されている．しかし，FFTを用いることによる周波数分解能の問題や，マッチング精度が基本周波数のパワーに依存する問題などが残されている．

4. 波形のマッチングによる音源同定

前述のスペクトルと同様に，楽器音の波形もまた，各々の特徴を持っている．この性質を利用し，波形情報をテンプレートとして利用する音源同定の手法が提案されている [26]．

しかし，波形を用いてマッチングを行なうためには位相情報が不可欠であり，異種複数楽器音が非常に複雑な性質を持つことから，このための位相トラッキングには複雑な処理が必要となる．また，楽器音の波形は非定常な性質を持つため，波形情報をテンプレートとする際には，その多様性に対応させるため適応的なフィルタリング処理が必要となる．

第 3 章

くし形フィルタによる音高，音価推定

本章では，縦続接続したくし形フィルタの各段の零出力検出による音高推定アルゴリズム [19],[20]，及び，音価推定アルゴリズムについて説明する．

なお，本研究では，採譜対象音域として，表 3.1 に示すオクターブ番号 3 の C 音 (C^3 : 130.81[Hz]) からオクターブ番号 5 の B 音 (B^5 : 987.77[Hz]) の 36 音について検討を行なう．

3.1 くし形フィルタの原理

ここでは，先ず，本方式で使用するくし形フィルタの構成を示し，くし形フィルタを利用して周波数成分を除去することにより，音名，及び，オクターブの判別が可能であることを説明する．次に，サンプリング周波数の異なるくし形フィルタを一定のサンプリング周波数で実現するためのオーバサンプリング法について述べ，これにより生じる零点誤差を軽減するために用いた線形補間法について説明する．

表 3.1 採譜対象音域の音高周波数 [Hz]

音 名	オクターブ番号		
	3	4	5
C	130.81	261.62	523.25
$C\#$	138.59	277.18	554.37
D	146.83	293.66	587.33
$D\#$	155.56	311.13	622.25
E	164.81	329.63	659.26
F	174.61	349.23	698.46
$F\#$	185.00	369.99	739.99
G	196.00	392.00	783.99
$G\#$	207.65	415.30	830.61
A	220.00	440.00	880.00
$A\#$	233.08	466.16	932.33
B	246.94	493.88	987.77

3.1.1 くし形フィルタの構成

本方式で使用するくし形フィルタの基本的な伝達関数を式(3.1)に、そのフィルタ構成を図3.1に、それぞれ示す。この $H(z)$ は、図3.2に示すように、周波数領域において0[Hz]からサンプリング周波数 f_s [Hz]までを r 等分する箇所に零点を持つ。また、式(3.2)に示すように、時間領域において、 $H(z)$ はサンプル点 $x(n)$ の減算のみで実現可能である。

$$H(z) = 1 - z^{-r} \quad (3.1)$$

$$y_r(n) = x(n) - x(n-r) \quad (3.2)$$

ここで、オクターブ番号 q 、音名 P の基本周波数を $f_{q,P}$ で表現した場合、隣り合うオクターブの同一音の基本周波数 $f_{q+1,P}$ 、 $f_{q,P}$ は式(3.3)の関係を持つ。また、楽器音に含まれる k 次高調波周波数は、式(3.4)で表現される。

$$f_{q+1,P} = 2 \times f_{q,P} \quad (P = C, C\#, \dots, B) \quad (3.3)$$

$$f_{q,P} \text{ の } k \text{ 次高調波周波数} = k \times f_{q,P} \quad [k \text{ は整数}] \quad (3.4)$$

式(3.3)、(3.4)より、採譜対象音域であるオクターブ3から5の P 音に含まれる全ての周波数（基本周波数、高調波）成分は、 $f_{3,P}$ の整数倍の周波数値をとることがわかる。

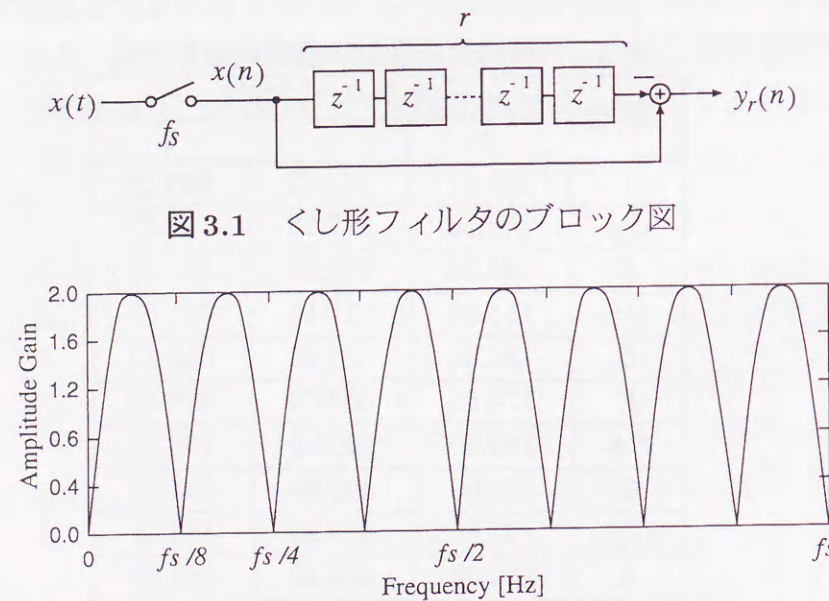


図3.1 くし形フィルタのブロック図

図3.2 8次くし形フィルタの周波数特性 ($r=8$)

前述の性質を考慮し、 $P = C, C\#, \dots, B$ の各音に対応させるため、 $f_{5,P}$ の2倍のサンプリング周波数を持つ12種類のくし形フィルタを構成する。この場合、 $f_{3,P}$ とサンプリング周波数 $f_{sP} = 2 \times f_{5,P}$ は、式(3.5)の関係を持つ。

$$\frac{f_{sP}}{f_{3,P}} = \frac{2 \times f_{5,P}}{f_{3,P}} = 8 \quad (3.5)$$

式(3.5)より、くし形フィルタを $r=8$ 次で構成することで、 $f_{3,P}$ の整数倍の箇所に零点が設置され、オクターブ3から5の P 音に含まれる全ての周波数成分の除去が可能となる。

ここで、各音に対応する12種類のくし形フィルタ $H_{r,P}(z_P)$ の伝達関数を式(3.6)に示す。なお、 z_P^{-1} はサンプル点間隔 $1/f_{sP}$ による単位遅延であり、 $x_P(n)$ はサンプリング周波数 f_{sP} によるサンプル点である。

$$H_{r,P}(z_P) = 1 - z_P^{-r} \quad (P = C, C\#, \dots, B) \quad (3.6)$$

$$y_{r,P}(n) = x_P(n) - x_P(n-r) \quad (3.7)$$

例として、ピアノの $A^3=220.0$ [Hz]音に含まれる周波数成分、及び、 A 音に対応する8次のくし形フィルタ $H_{8,A}(z_A)$ の周波数特性を図3.3に示す。 $H_{8,A}(z_A)$ の零点は f_{sA} と $r=8$ に対応する周波数成分（この場合は $f_{sA}/8 = f_{3,A}$ の整数倍）と一致する箇所に存在し、 A^3 音に含まれる全ての周波数成分を除去する性質を持つことが確認できる。

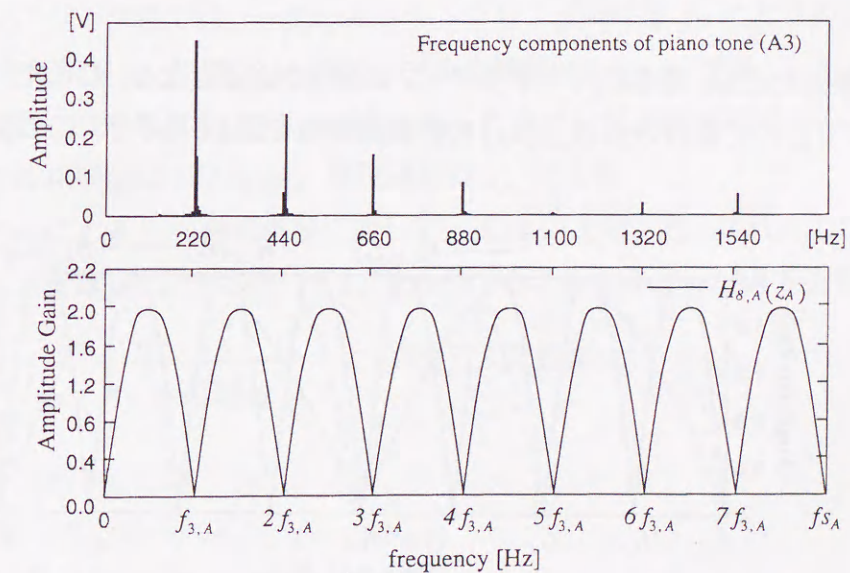


図3.3 くし形フィルタを利用した周波数成分除去の原理

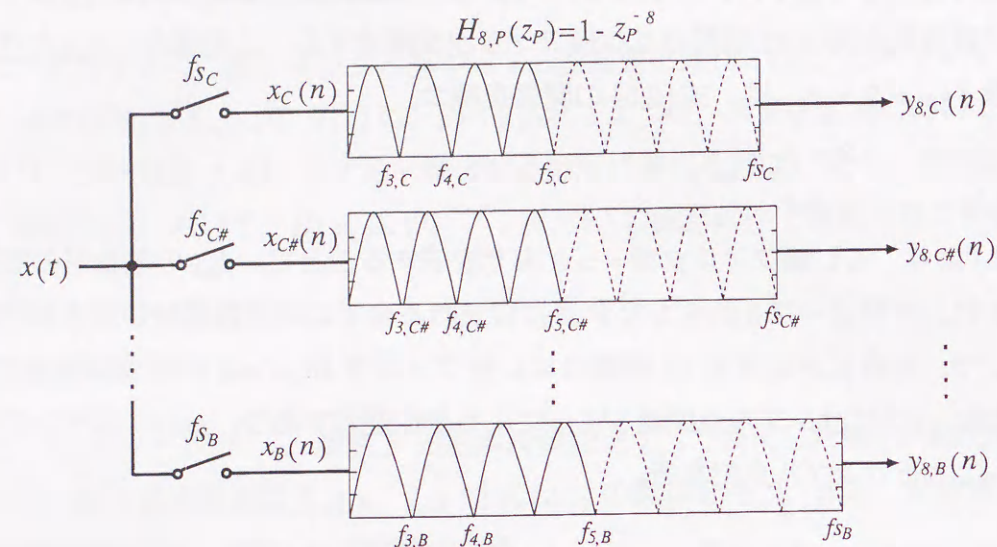


図 3.4 単音入力に対する音名判別システム

前述のような $H_{r,P}(z_P)$ の周波数成分除去の性質を利用し、図 3.4 に示すように、単音入力に対する音名判別システムを構成する。図 3.4 の各出力 $y_{8,P}(n)$ において、零・非零値を判別することにより、式 (3.8) に示すように入力音名 P を判別できる。

$$y_{8,P}(n) \begin{cases} = 0 & \rightarrow \text{detecting tone name } P, \quad P \in \{C, C\#, \dots, B\} \\ \neq 0 & \rightarrow \text{other tone} \end{cases} \quad (3.8)$$

さらに、楽譜化には、音名以外にオクターブの情報が必要となる。ここで、 $H_{r,P}(z_P)$ の次数を $r=8, 4, 2$ とした場合の各 $H_{r,P}(z_P)$ の周波数特性を図 3.5 に示す。この図より、 $H_{8,P}(z_P)$

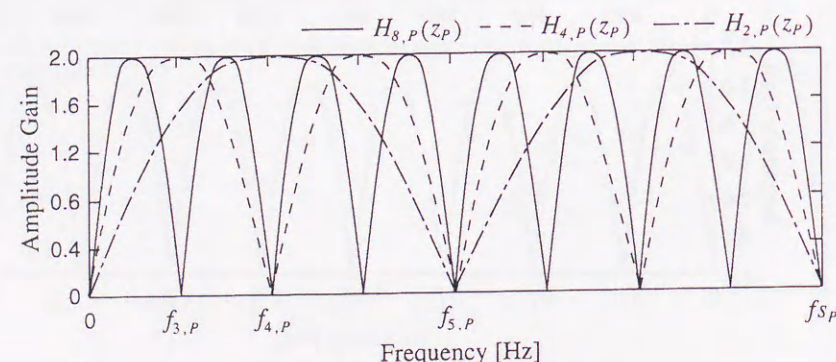


図 3.5 オクターブ推定の原理

ではオクターブ 3, 4, 5 の P 音に含まれる全ての周波数（基本周波数，高調波）成分が除去されるが、 $H_{4,P}(z_P)$ ではオクターブ 3 の P 音の奇数次高調波成分が除去されないことがわかる。この性質を利用することで、 $y_{8,P}(n) = 0$ （音名判別時点）かつ $y_{4,P}(n) \neq 0$ である場合、入力は P 音のオクターブ 3 であることが判別できる。同様にオクターブ 4, 5 に関しても、 $H_{r,P}(z_P)$ の次数を 8 次→4 次、4 次→2 次に変化させ、非零出力を検出することで、式 (3.9) に示すようにオクターブ判別が可能である。

$$\begin{aligned} y_{8,P}(n) = 0 \quad \& \quad y_{4,P}(n) \neq 0 & \Rightarrow \text{octave number 3} \\ y_{8,P}(n), y_{4,P}(n) = 0 \quad \& \quad y_{2,P}(n) \neq 0 & \Rightarrow \text{octave number 4} \\ y_{8,P}(n), y_{4,P}(n), y_{2,P}(n) = 0 & \Rightarrow \text{octave number 5} \end{aligned} \quad (3.9)$$

本節で述べた $H_{r,P}(z_P)$ の性質は、入力音が f_{s_P} に対応している場合には、その周波数成分の割合に依存せず成立する。従って、本研究の採譜対象である異種複数楽器音に対して、各楽器の周波数成分（スペクトル構造）についての事前情報を必要とせず、音名及びオクターブの判別が可能であるという利点を持つ。

3.1.2 オーバサンプリングによるくし形フィルタの近似的な実現

前節で述べたように、 $P = C, C\#, \dots, B$ に対応したくし形フィルタ $H_{r,P}(z_P)$ は、各々サンプリング周波数 f_{s_P} が異なっているため、各 $H_{r,P}(z_P)$ を直接接続する場合に不都合が生じる。この問題を解決する方法として、本方式ではオーバサンプリング法を用いる。

オーバサンプリング法では、一定の高いサンプリング周波数 f_s により得られるサンプル値系列 $x(n)$ を使用する。この $x(n)$ のサンプル点間隔は $1/f_s$ [s] であり、単位遅延を z^{-1} で表現する。なお、前述のように、 f_{s_P} により得られるサンプル値系列は $x_P(n)$ であり、この場合のサンプル点間隔は $1/f_{s_P}$ [s]、単位遅延は z_P^{-1} である。

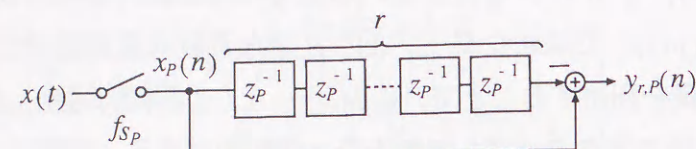
ここで、 $P = C, C\#, \dots, B$ の各音に対して、式 (3.5) で示した r の代わりに式 (3.10) を満たす整数値 r_P を計算し、 z^{-r_P} を用いて z_P^{-r} を近似することで各 $H_{r,P}(z_P)$ を近似的に実現する。

$$r_P \cong \frac{f_s}{f_{3,P}} \quad (\text{四捨五入で整数化}) \quad (3.10)$$

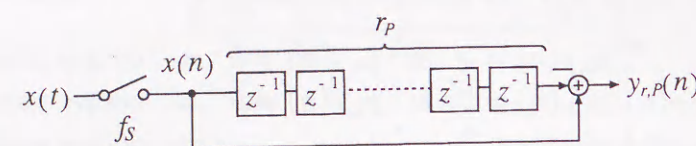
$$H_{r,P}(z) \cong 1 - z^{-r_P} \quad (3.11)$$

$$y_{r,P}(n) \cong x(n) - x(n - r_P) \quad (3.12)$$

式 (3.11) より、図 3.6(a) に示す理想的な $H_{r,P}(z_P)$ は近似的に同図 (b) の $H_{r,P}(z)$ で表現される。これにより、 $P = C, C\#, \dots, B$ に対応する各 $H_{r,P}(z_P)$ は、全て一定のサンプリング



(a) $H_{r,p}(z_P) = 1 - z_P^{-r}$



(b) $H_{r,p}(z) = 1 - z^{-r_P}$

図 3.6 オーバサンプリング法による $H_{r,p}(z_P)$ の近似的実現

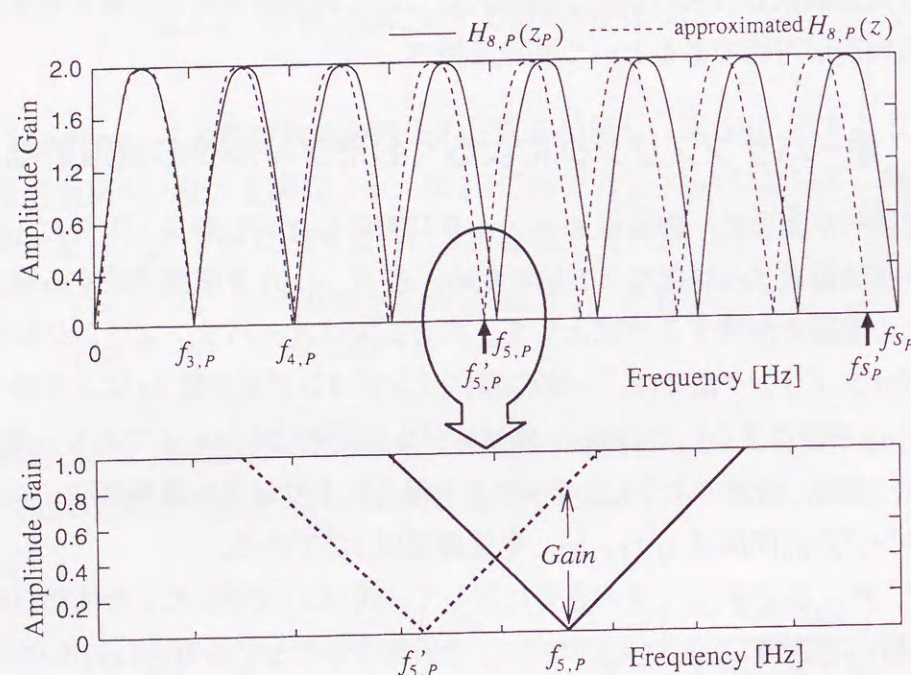


図 3.7 オーバサンプリングによる零点のずれ

周波数 f_s [Hz] により近似的に実現可能となる。但し、前述の近似により、式 (3.11) に示す $H_{r,p}(z)$ の零点周波数は、理想的な零点周波数（平均率音階の規定周波数）と完全には一致せず、若干のずれを持つ。このため、 P 音が入力された場合の $H_{r,p}(z)$ の出力は厳密には零値をとらず、各周波数成分の振幅に図 3.7 の $Gain$ に示す利得が乗じられた値となる。

表 3.2 $f_s = 54.054$ [kHz] により近似されたサンプリング周波数とその誤差

音名 P	f_{s_P} [Hz]	f'_{s_P} [Hz]	Δf_{s_P} [Hz]
C	1046.50	1047.05	0.55
$C\#$	1108.73	1108.80	0.07
D	1174.66	1175.09	0.43
$D\#$	1244.51	1246.20	1.69
E	1318.50	1318.39	0.11
F	1396.91	1394.94	1.97
$F\#$	1479.98	1480.93	0.95
G	1567.98	1566.78	1.20
$G\#$	1661.22	1663.20	1.98
A	1760.00	1757.86	2.14
$A\#$	1864.66	1863.93	0.73
B	1975.53	1974.58	0.95

ここで、近似的な $H_{r,p}(z)$ の各サンプリング周波数 f'_{s_P} 、及び、理想的なサンプリング周波数 f_{s_P} との誤差 Δf_{s_P} は、式 (3.13) で計算される。なお、本方式では、実際に使用する A/D 変換ボード（詳しくは 5 章で説明）の都合上、計算機シミュレーションで使用するオーバサンプリング周波数 f_s を 54.054 [kHz] に定めている。この $f_s = 54.054$ [kHz] に対する各 $P = C, C\#, \dots, B$ 音の f'_{s_P} 、 Δf_{s_P} を、表 3.2 に示す。

サンプリング周波数の誤差 Δf_{s_P} により、 $H_{r,p}(z)$ の各零点（ $r = 8$ 次の場合、オクターブ 3 の基本周波数 $f_{3,p}$ 、及び、 k 次高調波 $k \times f_{3,p}$ ）には、式 (3.14) に示す周波数誤差が生じる。

$$f'_{s_P} = \frac{r \times f_s}{r_P}, \quad \Delta f_{s_P} = |f'_{s_P} - f_{s_P}| \quad (3.13)$$

$$\text{零点 } k \times f_{3,p} \text{ の周波数誤差} = \frac{k}{8} \times \Delta f_{s_P} \quad [k \text{ は整数}] \quad (3.14)$$

ここで、近似的な $H_{r,p}(z)$ の振幅特性は式 (3.15) で計算されるため、上述の各周波数成分 $n \times f_{3,p}$ に生じる利得は式 (3.16) となる。

$$|H_{r,p}(e^{j\omega})| = |1 - \exp(-j \cdot r_P \times \omega)| = \sqrt{2 - 2 \cos(r_P \times \omega)} \quad (3.15)$$

ここで, $\omega = \frac{2\pi f}{f_s}$

$$k \times f_{3,p} \text{ に生じる利得} = \sqrt{2 - 2 \cos \left(r_P \times \frac{2\pi \cdot \frac{k}{8} \times \Delta f_{s_P}}{f_s} \right)} \quad (3.16)$$

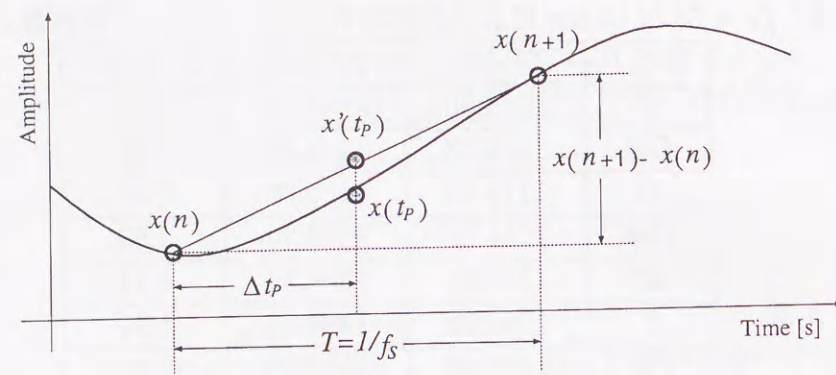


図 3.8 サンプル点の線形補間

なお、表 3.2 に示した各 $\Delta f s'_p$ に対して生じる利得は、最大で A 音の $\Delta f s'_A (= 8 \times f_{3,A}) = 2.14$ [Hz] の場合であり、入力された振幅の 4% 程度が除去しきれない成分として出力される。

前述のような零点のずれに対して、本方式では、図 3.8 に示すような線形補間法を用いることで、除去しきれない成分による出力誤差の軽減を図る。

オーバーサンプリング周波数 f_s により得られるサンプル点を $x(n)$ とした場合、 $x(n)$ と $x(n+1)$ の間のサンプル点間隔は $T = 1/f_s$ [s] である。一方、所望のサンプル点を $x(t_p)$ 、 $nT < t_p < (n+1)T$ とした場合、 nT と t_p の間には $\Delta t_p = t_p - nT$ [s] の時間差が生じる。この関係より、式 (3.17) に示すように、線形補間による近似的なサンプル点 $x'(t_p)$ を計算する。

$$x'(t_p) = x(n) + \{x(n+1) - x(n)\} \times \frac{\Delta t_p}{T} \quad (3.17)$$

本方式では、所望のサンプル点 $x(t_p)$ の代わりに、式 (3.17) における $x'(t_p)$ を用いることで零点誤差の軽減を図る。これにより、除去しきれない成分の振幅は、線形補間を行なう前と比較して 1/50 倍程度（最終的なずれ成分に生じる利得は、最大で 0.2% 程度）にまで軽減できることを確認した。

3.2 くし形フィルタの縦続接続による音高検出

本章では、くし形フィルタを利用した音高検出原理について説明する。

初めに、1 オクターブ中の 12 音 ($C, C\#, \dots, B$) に対応するくし形フィルタを縦続接続する基本的な音名判別システムの構成、及び、そのアルゴリズムについて説明する。さらに、平均率音階における周波数の重複関係を利用し、くし形フィルタの縦続数を削減する方式について示す。

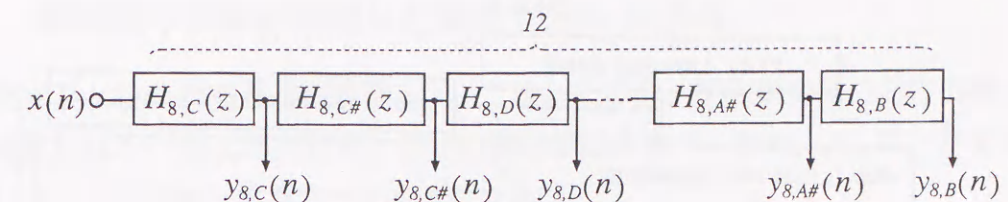


図 3.9 $H_{8,p}(z)$ の縦続接続によるシステム構成

3.2.1 基本的な音名判別システムの構成

本方式では、図 3.9 に示すように、 $P = C, C\#, \dots, B$ の各音に対応する 12 種類のくし形フィルタ $H_{8,p}(z)$ を縦続接続することで、音名判別システムを構成する。

このシステムでは、 $H_{8,C}(z), H_{8,C\#}(z), \dots, H_{8,B}(z)$ を用いることで、入力信号 $x(n)$ に含まれる音を $C, C\#, \dots, B$ の順に 1 音ずつ除去し、全ての音を除去し終わった段階で初めて零値が出力される。本方式では、この零出力を検出することで音名の判別を行なう。なお、本システムでは、各音の基本周波数成分が他音と重複を持たないことを入力条件とする。

3.2.2 基本的な音高検出アルゴリズム

図 3.9 のシステムによる音高検出アルゴリズム（フローチャート）を図 3.10 に示し、以下にその説明を行なう。なお、ここでは、入力として m 種類の音を含む和音（及び、単音） $x(n) = \sum \text{tone } P_i$ ($P_i \in \{C, C\#, \dots, B\}$, $1 \leq i \leq m$, $m \leq 12$) を扱う。

図 3.9 のシステムに対して上記の信号 $x(n)$ を入力した場合、最も高い音 P_m に対応した $H_{8,P_m}(z)$ の出力 $y_{8,P_m}(n)$ で最初に零値が検出される。この結果から、 $x(n)$ には少なくとも P_m 音が含まれていることが判別できる。ここで、 $H_{8,P_m}(z)$ の次数を $r = 8 \rightarrow 4 \rightarrow 2$ と変化させ、 $y_{r,P_m}(n)$ の零・非零出力を判別することにより、オクターブ推定を行なう。

step 1. $y_{8,P_m}(n) = 0 \rightarrow$ detecting tone P_m

step 2. detecting octave of tone P_m

次に、 $H_{8,P_m}(z)$ を $H_{8,C}(z)$ の前段（縦続の先頭）に移動する。

step 3. placing $H_{8,P_m}(z)$ in front of $H_{8,C}(z)$

$x(n)$ が単音の場合、先頭に移動した $H_{8,P_m}(z)$ の出力 $y'_{8,P_m}(n)$ において、再び零値が検出されるため、処理を終了する。一方、 $x(n)$ が和音の場合、移動した $H_{8,P_m}(z)$ のみでは和

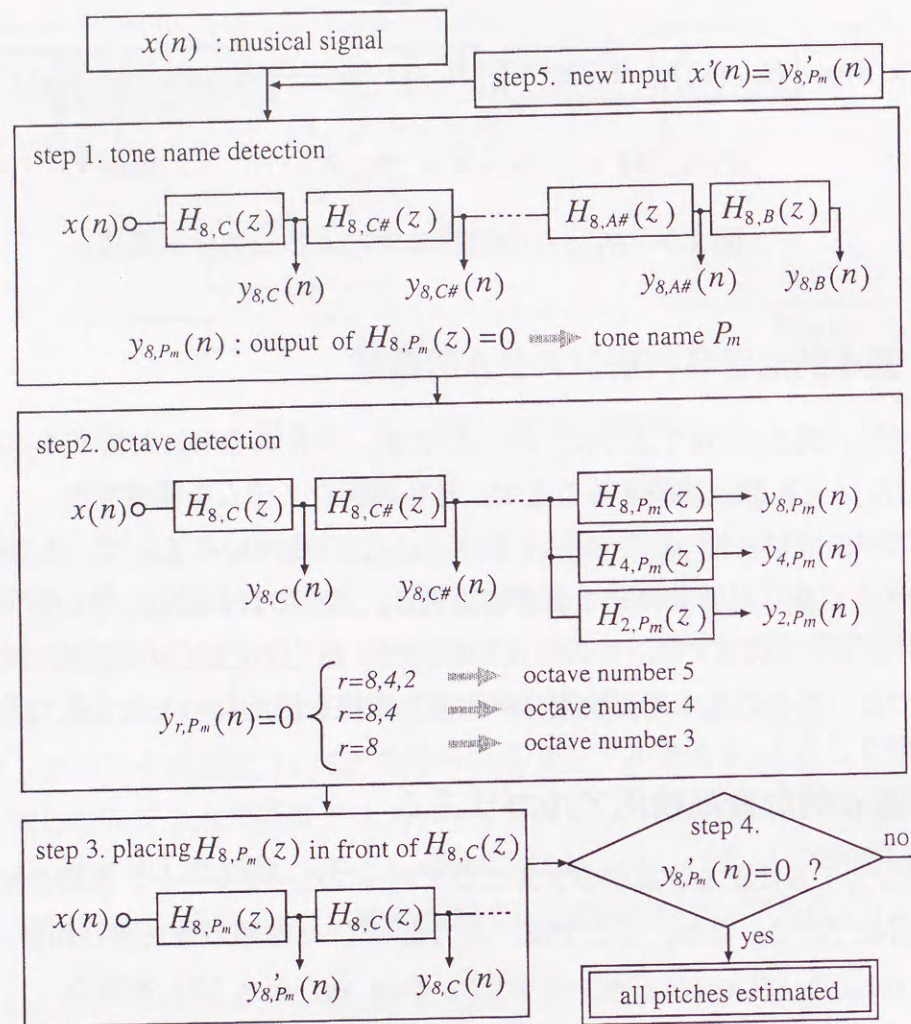


図 3.10 基本的な音名判別アルゴリズム

音に含まれる他の音の周波数成分を除去しきれないため、非零値が検出される。

step 4. $y'_{P_m}(n) \begin{cases} = 0 \rightarrow \text{monophony } P_m \\ \neq 0 \rightarrow \text{polyphony} \end{cases}$

step 4. において非零出力 $y'_{8,P_m}(n) \neq 0$ が検出された場合、 $x(n)$ に含まれる他の音高を判別するため、再び $H_{8,C}(z)$ から順に同様な処理を行なう。なお、 $H_{8,P_m}(z)$ を先頭に移動させたことで、 $x'(n) = y'_{8,P_m}(n) = x(n) - \text{tone } P_m$ がシステムにおける $H_{8,C}(z)$ の新たな入力となり、 $x'(n)$ に含まれる音は $m-1$ 種類となる。

step 5. $y'_{8,P_m}(n) \rightarrow \text{new input } x'(n) \Rightarrow \text{go to step 1.}$

上述の処理を繰り返すことで、 $x(n)$ を構成する音高を 1 音ずつ検出可能である。

3.2.3 周波数の重複を利用した音名判別システム

オクターブ q の P 音の基本周波数を $f_{q,P}$ で表した場合、平均率音階における隣り合う音高の周波数比は $2^{1/12}$ 倍であるため、以下の関係が成り立つ。なお、 P_{+d} は、 P 音を基準として高い方へ d 音の隔たりを持つ音を意味する。

$$\frac{f_{q,P+1}}{f_{q,P}} = 2^{1/12} \cong 1.059463 \quad \text{より,}$$

$$\frac{f_{q,P+19}}{f_{q,P}} = 2^{19/12} \cong 2.996614 \cong 3.0$$

$$\frac{f_{q,P+28}}{f_{q,P}} = 2^{28/12} \cong 5.039684 \cong 5.0$$

$$\frac{f_{q,P+31}}{f_{q,P}} = 2^{31/12} \cong 5.993228 \cong 6.0$$

上記の周波数関係より、採譜対象音域である C^3 から B^5 の 36 音で考えた場合、 P 音の高調波と P_{+19} 音、 P_{+28} 音、 P_{+31} 音の基本周波数が非常に近い値となる（重複する）ことが分かる。なお、 P_{+19} 、 P_{+31} 音はそれぞれオクターブ 4、5 の P_{+7} 音であり、 P_{+28} 音はオクターブ 5 の P_{+4} 音である。このような重複関係を持つ P 、 P_{+7} 、 P_{+4} 音について、それぞれの周波数配置を図 3.11 に示す（なお、このような周波数の重複関係については、次章で詳しく説明する）。

図 3.11 に示すような重複関係を持つ $H_{8,P}(z)$ と $H_{8,P+7}(z)$ 、及び、 $H_{8,P}(z)$ と $H_{8,P+4}(z)$ の零点は非常に近い箇所に存在し、 P_{+19} 音、 P_{+28} 音、 P_{+31} 音の周波数成分は $H_{8,P}(z)$ の対応する零点で除去可能となる。この性質より、 $P = C, C\#, \dots, B$ の各音に対応する 12 種類の

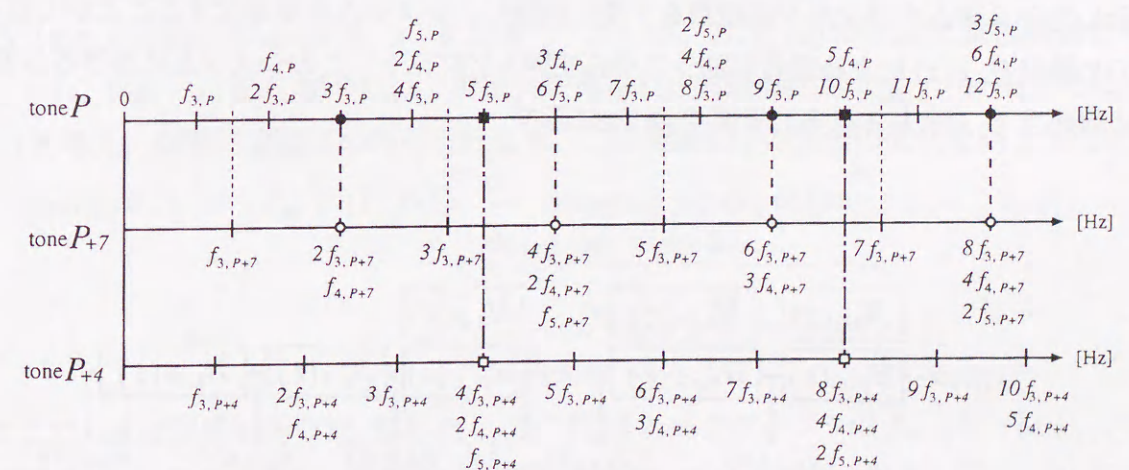


図 3.11 P 、 P_{+7} 、 P_{+4} 音の周波数配置

表 3.3 各 $H_{8,P}(z)$ で除去可能な音高

	$H_{8,C}(z)$	$H_{8,C\#}(z)$	$H_{8,D}(z)$	$H_{8,D\#}(z)$
Pitch	C^3, C^4, C^5 G^4, G^5 E^5	$C\#^3, C\#^4, C\#^5$ $G\#^4, G\#^5$ F^5	D^3, D^4, D^5 A^4, A^5 $F\#^5$	$D\#^3, D\#^4, D\#^5$ $A\#^4, A\#^5$
	$H_{8,E}(z)$	$H_{8,F}(z)$	$H_{8,F\#}(z)$	$H_{8,G}(z)$
Pitch	E^3, E^4 B^4, B^5	F^3, F^4	$F\#^3, F\#^4$	G^3
	$H_{8,G\#}(z)$	$H_{8,A}(z)$	$H_{8,A\#}(z)$	$H_{8,B}(z)$
Pitch	$G\#^3$	A^3	$A\#^3$	B^3

$H_{8,P}(z)$ は、それぞれ、表 3.3 に示すような音高の周波数成分を除去することが可能である。

表 3.3 より、 $H_{8,G}(z), H_{8,G\#}(z), \dots, H_{8,B}(z)$ の 5 種類のくし形フィルタは、 $G^3, G\#^3, \dots, B^3$ の各 1 音のみを除去するために使用されていることが分かる。ここで、図 3.11 に示した周波数関係に注目すると、 $P = C, C\#, \dots, F\#$ の 7 音に対して $H_{8,P}(z)$ の零点を 2 分割する点に新たな零点を設けることで、即ち、 $H_{8,P}(z)$ の代わりに $H_{16,P}(z)$ を使用することで、 $G^3, G\#^3, \dots, B^3$ の周波数成分が除去可能となることが分かる。なお、 $H_{16,P}(z)$ 場合、式 (3.10) で示したオーバーサンプリング法によるくし形フィルタの次数 r_P は、

$$r_P \cong \frac{fs}{f_{3,P} \div 2} \quad [\text{四捨五入で整数化}] \quad (3.18)$$

となる。

この性質を利用することで、 $H_{8,G}(z), H_{8,G\#}(z), \dots, H_{8,B}(z)$ が不要となるため、図 3.9 に示したシステムのフィルタ縦続数を 7 段に削減し、システムを簡単化することができる。この 7 段の $H_{r,P}(z)$ による音名判別システムを図 3.12 に示し、このシステムにおける各 $H_{r,P}(z)$ に対応する（除去される）音高を表 3.4 に示す。

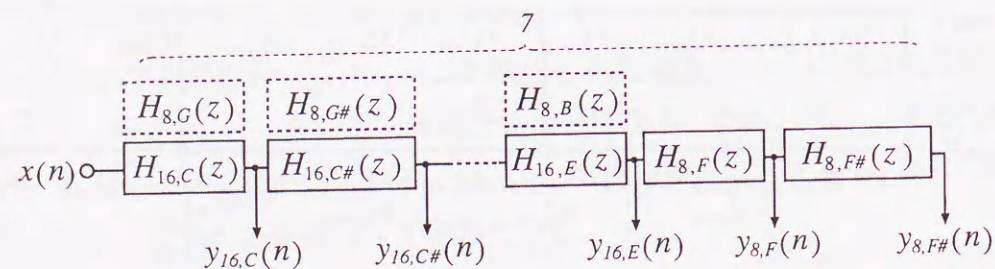


図 3.12 周波数の重複を利用した音名判別システム

表 3.4 各 $H_{8,P}(z)$ で除去可能な音高

	$H_{16,C}(z)$	$H_{16,C\#}(z)$	$H_{16,D}(z)$	$H_{16,D\#}(z)$
Pitch	C^3, C^4, C^5 G^3, G^4, G^5 E^5	$C\#^3, C\#^4, C\#^5$ $G\#^3, G\#^4, G\#^5$ F^5	D^3, D^4, D^5 A^3, A^4, A^5 $F\#^5$	$D\#^3, D\#^4, D\#^5$ $A\#^3, A\#^4, A\#^5$
	$H_{16,E}(z)$	$H_{8,F}(z)$	$H_{8,F\#}(z)$	
Pitch	E^3, E^4 B^3, B^4, B^5	F^3, F^4	$F\#^3, F\#^4$	

なお、表 3.4 より、重複音からなる和音、例えば、 C^5 と E^5 と G^5 からなる入力音に対しては、縦続接続の初段の $H_{16,C}(z)$ の出力が零となり、前述のアルゴリズムでは和音の判別が不可能となる。図 3.12 のシステムに対する新たな重複音高判別アルゴリズムについては、次節で説明する。

3.2.4 周波数の重複を利用した音高判別アルゴリズム

まず、3.2.2 で述べた基本的な音高検出アルゴリズムと同様に、図 3.12 のシステムにおいて零出力検出を行ない、入力信号 $x(n)$ の全ての周波数成分を除去するために必要な $H_{r,P_i}(z)$ ($r=16$ or 8) を明らかにする。

step 1. $y_{r,P_m}(n) = 0 \rightarrow$ detecting zero output of $H_{r,P_m}(z)$

ここで、 $H_{r,P_m}(z)$ が他音と重複関係を持たないもの（表 3.4 より、 $P_m \in \{F, F\# \}$ ）である場合、前述と同様に P_m 音のオクターブ判別を行ない、 $H_{r,P_m}(z)$ を縦続の先頭に移動させる（3.2.2 における step 3, 4, 5 に従う）。一方、 $H_{r,P_m}(z)$ が他音と重複関係を持つ（ $P_m \in \{C, C\#, \dots, E\}$ ）場合には、零出力が得られた前段の出力 $y_{r,P_{m-1}}(n)$ を新たな入力 $x'(n)$ と考え、個別に重複音の判別を行なう。この段階までの処理の流れを図 3.13 に示す。

step 2'. if $P_m \in \{F, F\# \} \rightarrow$ detecting octave of tone P_m
 \Rightarrow go to step 3.

else \rightarrow new input $x'(n) = y_{r,P_{m-1}}(n)$
 \Rightarrow go to step 6.

ここで、重複関係を持つ P_m 音と P_{m+7} 音、 P_m 音とオクターブ 5 の P_{m+4} 音（以下、 P_{m+4}^5 で表現する）に注目すると、前節と逆の重複は成立しないことが分かる。即ち、 P_m 音の入力に対して、 P_m, P_{m+7}, P_{m+4} 音に対応する各くし形フィルタは、以下の出力値をとる。なお、 $y_{2,P_{m+4}}(n)$ は P_{m+4}^5 音に対応する $H_{2,P_{m+4}}(z)$ の出力である。

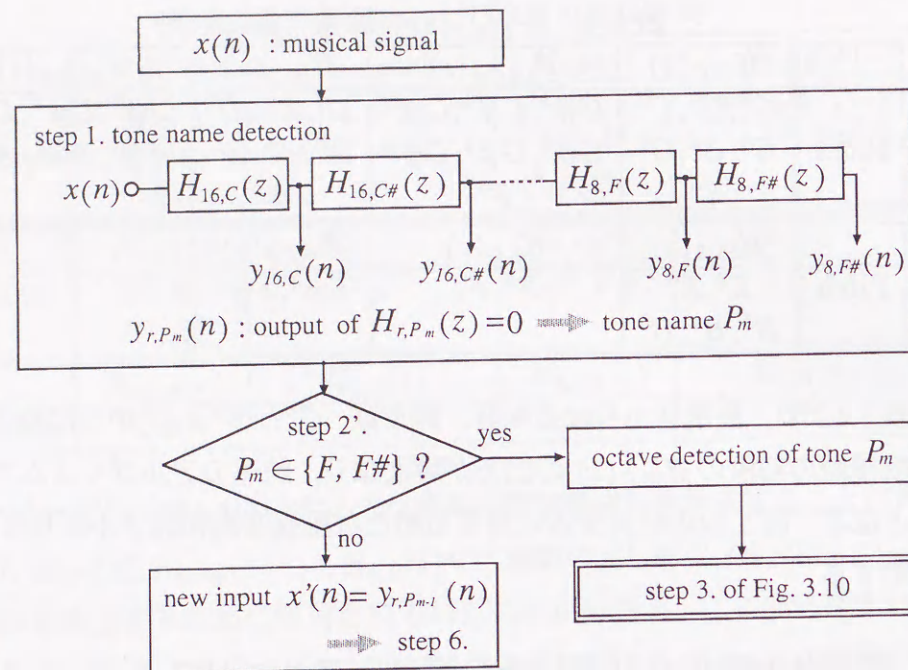


図 3.13 周波数の重複を利用した音名判別アルゴリズム 1

$$\text{input tone } P_m \rightarrow \begin{cases} y_{16,P_m}(n) = 0 \\ y_{8,P_{m+7}}(n) \neq 0 \\ y_{2,P_{m+4}}(n) \neq 0 \end{cases}$$

また、この重複に関係する P_{m+7} 音と P_{m+4}^5 音の間にも、同様に重複関係は成り立たない。

$$\begin{aligned} \text{input tone } P_{m+7} &\rightarrow \begin{cases} y_{8,P_{m+7}}(n) = 0 \\ y_{2,P_{m+4}}(n) \neq 0 \end{cases} \\ \text{input tone } P_{m+4}^5 &\rightarrow \begin{cases} y_{8,P_{m+7}}(n) \neq 0 \\ y_{2,P_{m+4}}(n) = 0 \end{cases} \end{aligned}$$

この性質を利用し、図 3.14 に示すように、 $x'(n)$ に対して P_{m+7} , P_{m+4}^5 音に対応する $H_{8,P_{m+7}}(z)$, $H_{2,P_{m+4}}(z)$ のみを使用して零出力検出を行なう。

この段階で $y_{8,P_{m+7}}(n) = 0$ ないし $y_{2,P_{m+4}}(n) = 0$ が得られた場合、対応する P_{m+7} 音あるいは P_{m+4}^5 音が単音で入力されたと判断し、前述と同様に、3.2.2 における step 3, 4, 5 に従って処理を行なう。なお、 $y_{2,P_{m+4}}(n) = 0$ の場合には、オクターブ 5 の P_{m+4} 音であることが明らかであるが、 $y_{8,P_{m+7}}(n) = 0$ の場合には、前述と同様に、 P_{m+7} 音のオクターブ推

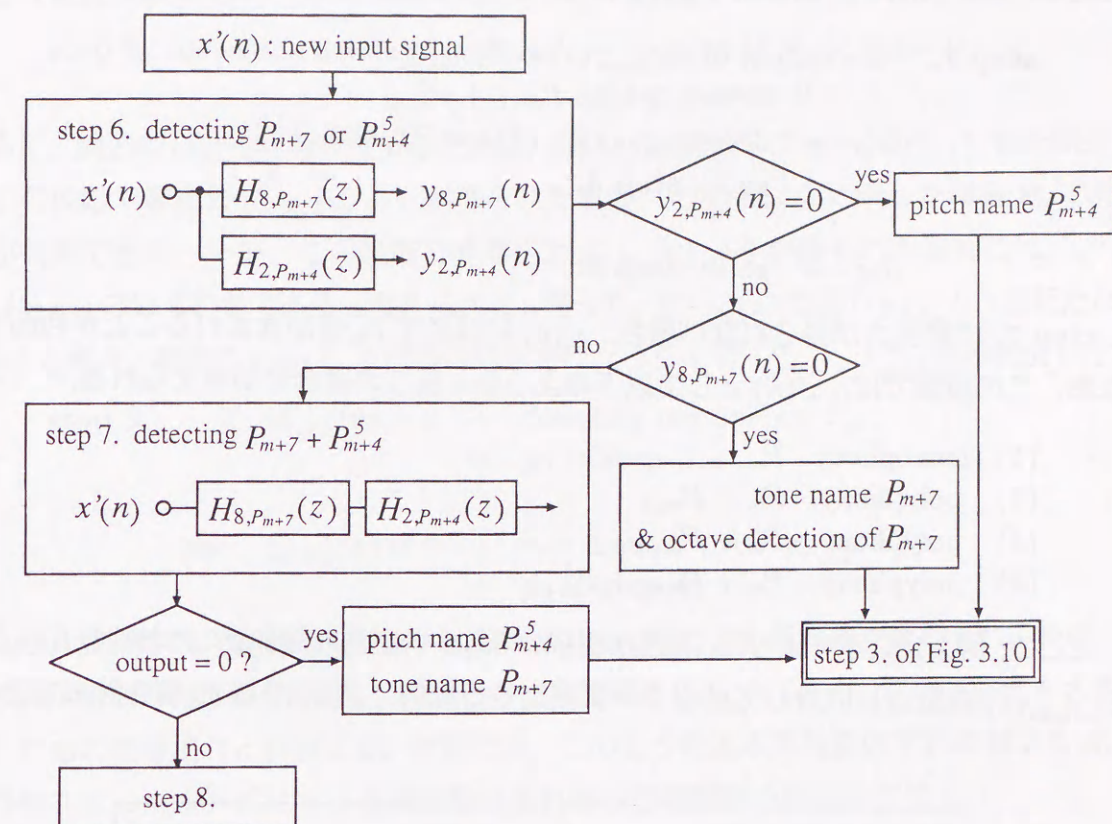


図 3.14 周波数の重複を利用した音名判別アルゴリズム 2

定を行なう必要がある。

step 6. if $y_{8,P_{m+7}}(n) = 0 \rightarrow$ detecting tone P_{m+7}
 \rightarrow detecting octave of tone P_{m+7}
 \Rightarrow go to step 3.

else if $y_{2,P_{m+4}}(n) = 0 \rightarrow$ detecting tone P_{m+4}^5
 \Rightarrow go to step 3.

else \Rightarrow go to step 7.

step 6. で零出力が得られない場合には、 $H_{8,P_{m+7}}(z)$ と $H_{2,P_{m+4}}(z)$ の 2 種類を縦続接続して周波数成分除去を行なう。この段階で零出力が得られた場合、入力 P_{m+7} , P_{m+4}^5 の和音

であることが判別できるため、 P_{m+7} 音のオクターブ推定を行ない2音の音高推定を終了する。

step 7. if output of $H_{8,P_{m+7}}(z) \times H_{2,P_{m+4}}(z) = 0$
 \rightarrow detecting tone $P_{m+7} + P_{m+4}^5$
 \rightarrow detecting octave of tone P_{m+7}
 \Rightarrow go to step 3.

else \Rightarrow go to step 8.

step 7. で零出力が得られない場合、 $x'(n)$ には必ず P_m 音が含まれることが判別できる。なお、この段階では、 $x'(n)$ として以下のような4通りの組合せが考えられる。

- (1) monophony P_m
- (2) polyphony $P_m + P_{m+7}$
- (3) polyphony $P_m + P_{m+4}^5$
- (4) polyphony $P_m + P_{m+7} + P_{m+4}^5$

ここで、図3.15に示すように、存在が明らかである単音 P_m (step 7.における $H_{8,P_{m+7}}(z) \times H_{2,P_{m+4}}(z)$ の出力) に対してオクターブ推定を行ない、 $x'(n)$ から P_m 音を除去するために

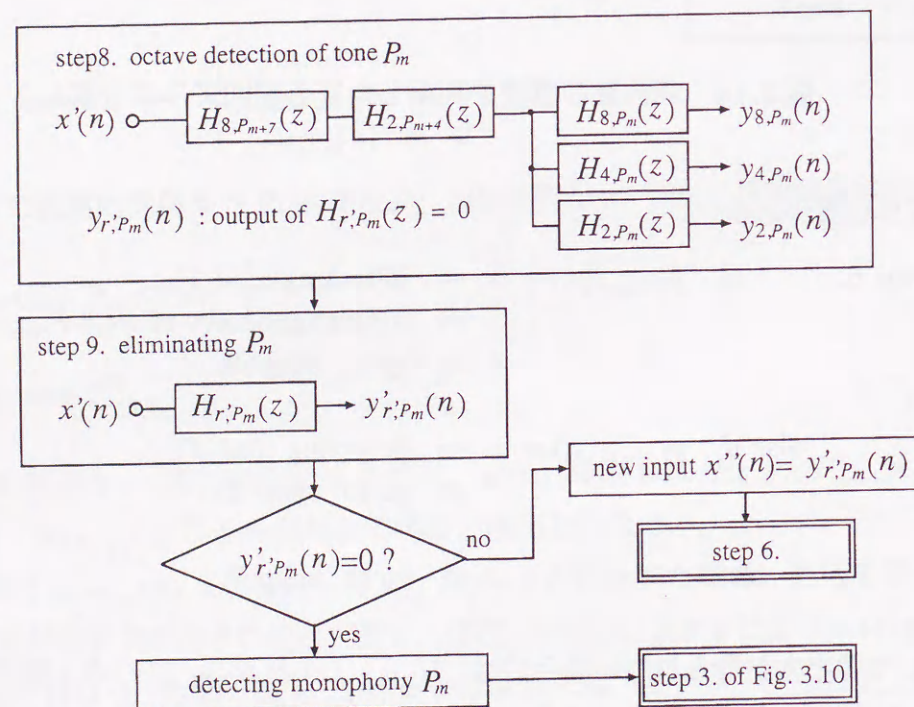


図3.15 周波数の重複を利用した音名判別アルゴリズム3

必要な最低次数の $H_{r',P_m}(z)$ を決定する。

step 8. detecting octave of tone $P_m \rightarrow$ output of $H_{r',P_m}(n) = 0$
 \Rightarrow go to step 9.

さらに、前述の(1)~(4)を判別するために、 $H_{r',P_m}(z)$ を用いて $x'(n)$ から P_m 音を除去する。

この段階で零出力 $y'_{r',P_m}(n) = 0$ が得られた場合、 $x'(n)$ は(1)に示した単音 P_m であることが判断できる。一方、この段階で非零出力 $y'_{r',P_m}(n) \neq 0$ が得られた場合には、 $x'(n)$ は(2),(3),(4)のいずれかであると考えられる。従って、 $H_{r',P_m}(z)$ の出力 $y'_{r',P_m}(n)$ を新たな入力 $x''(n)$ と考え、前述のstep 6. と同様な処理を行なうことで、 P_{m+7} 、 P_{m+4}^5 音の判別を行なう。

step 9. if $y'_{r',P_m}(n) = 0 \rightarrow$ detecting monophony P_m
 \Rightarrow go to step 3.

else $y'_{r',P_m}(n) \neq 0 \rightarrow$ new input $x''(n) = y'_{r',P_m}(n)$
 \Rightarrow go to step 6.

なお、実際には、重複関係を持つ P_m 音と P_{m+7} 、 P_{m+4}^5 音の基本周波数のずれ、及び、楽器音の非定常性等の影響を考慮し、前段との出力振幅比 $|y_{r,P_m}(n)|/|y_{r,P_{m-1}}(n)|$ が最も大きく減衰したものを零出力と判別する。次節では、このような基本周波数のずれに対する $H_{r,P}(z)$ の特性を示し、本方式における周波数のずれ幅の許容範囲を明らかにする。

3.2.5 周波数のずれに対する音名判別システムの特性

入力音のピッチがずれを持つ場合、前節の音名判別アルゴリズムで最も厳しい条件は隣接音の検出である。ここで、オクターブ番号 q の P_m 音と隣接する P_{m+1} 音の和音を想定し、対応する $H_{r,P_m}(z)$ 、 $H_{r,P_{m+1}}(z)$ の出力特性を考察する。なお、 P_m 音の基本周波数を f_{q,P_m} とし、 P_{m+1} 音の基本周波数が規定された $f_{q,P_{m+1}}$ から $-\Delta f_{P_{m+1}}$ のずれを持つと仮定する。

この場合、 P_{m+1} 音の k 次高調波と $H_{r,P_{m+1}}(z)$ の零点の周波数間隔は式(3.19)となる。

$$\underbrace{P_{m+1} \text{ 音の } k \text{ 次高調波}}_{k \times (f_{q,P_{m+1}} - \Delta f_{P_{m+1}})} - \underbrace{\text{零点周波数}}_{k \times f_{q,P_{m+1}}} = k \times \Delta f_{P_{m+1}} \quad (3.19)$$

ここで、 P_{m+1} 音の k 次高調波の振幅を $s_{m+1,k}$ とし、各成分の位相が一致していると仮定した場合、 $H_{r,P_{m+1}}(z)$ の出力振幅 $|y_{r,P_{m+1}}(n)|$ は式(3.20)で表現される。なお、 K は P_{m+1} 音に含まれる高調波の数である。

$$|y_{r,P_{m+1}}(n)| = \sum_{k=1}^K s_{m+1,k} \times |H_{r,P_{m+1}}(e^{j\omega_k})|$$

なお、 $\omega_k = \frac{k \times (f_{q,P_{m+1}} - \Delta f_{P_{m+1}})}{f_s}$ (3.20)

一方、 P_{m+1} 音の k 次高調波と $H_{r,P_m}(z)$ の対応する零点 $k \times f_{q,P_m}$ の周波数間隔 $k \times \Delta f_{P_m}$ を式 (3.21) に、 P_{m+1} 音を入力した場合の $H_{r,P_m}(z)$ の出力を式 (3.22) に、それぞれ示す。

$$k \times (f_{q,P_{m+1}} - \Delta f_{P_{m+1}}) - k \times f_{q,P_m} = k \times \overbrace{(f_{q,P_{m+1}} - \Delta f_{P_{m+1}} - f_{q,P_m})}^{\Delta f_{P_m}} \quad (3.21)$$

$$|y_{r,P_m}(n)| = \sum_{k=1}^K s_{m+1,k} \times |H_{r,P_m}(e^{j\omega_k})| \quad (3.22)$$

ここで、零点付近の $|H_{r,P_{m+1}}(z)|$ を傾き α_{m+1} の直線で近似し、同様に、 $|H_{r,P_m}(z)|$ を傾き α_m の直線で近似した場合、図 3.16 に示すように、 $|H_{r,P_{m+1}}(e^{j\omega_k})|$ 、 $|H_{r,P_m}(e^{j\omega_k})|$ はそれぞれ式 (3.23) で表現される。

$$\begin{aligned} |H_{r,P_{m+1}}(e^{j\omega_k})| &= \alpha_{m+1} \cdot k \times \Delta f_{P_{m+1}} \\ |H_{r,P_m}(e^{j\omega_k})| &= \alpha_m \cdot k \times \Delta f_{P_m} \end{aligned} \quad (3.23)$$

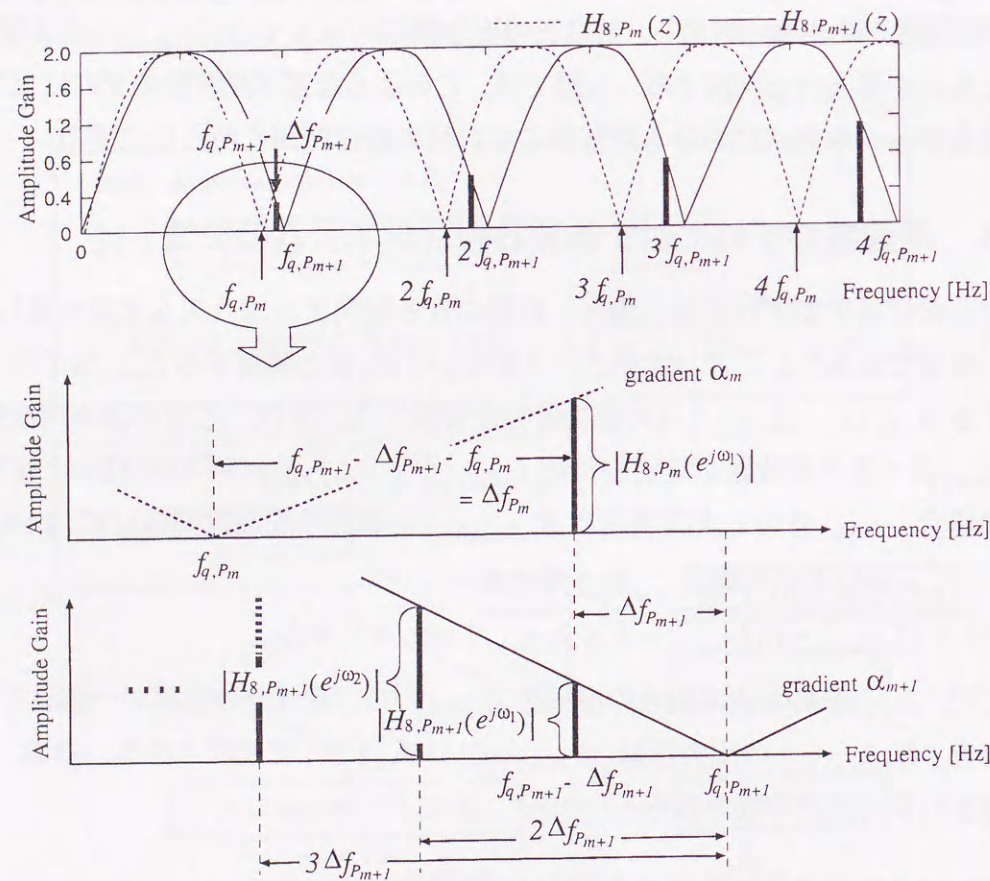


図 3.16 周波数のずれ Δf_{m+1} に対する $H_{r,P}(z)$ の特性

ここで、くし形フィルタが縦続接続されたシステムに、前述の P_m 、 P_{m+1} 音からなる和音 $x(n)$ が入力されたと仮定する。この場合、 P_m 音の基本周波数 f_{q,P_m} は $H_{r,P_m}(z)$ の零点周波数と一致しているため、入力音 $x(n)$ に対する各くし形フィルタの出力は、式 (3.23) より、次のように変化する。なお、 P_m 音の k 次高調波の振幅を $s_{m,k}$ で表現する。

$$\begin{aligned} |x(n)| &= \sum_{k=1}^K s_{m,k} + \sum_{k=1}^K s_{m+1,k} \\ &\Downarrow \\ |y_{r,P_m}(n)| &= \sum_{k=1}^K 0 \cdot s_{m,k} + \sum_{k=1}^K \alpha_m \cdot k \times \Delta f_{P_m} \cdot s_{m+1,k} \\ &\Downarrow \\ |y_{r,P_{m+1}}(n)| &= \sum_{k=1}^K \alpha_{m+1} \cdot k \times \Delta f_{P_{m+1}} \cdot \alpha_m \cdot k \times \Delta f_{P_m} \cdot s_{m+1,k} \end{aligned}$$

前述のように、 P_{m+1} 音を検出するためには、 $|y_{r,P_{m+1}}(n)|$ において、前段の出力時以上の減衰が必要となる。従って、 P_{m+1} 音検出のための条件は、式 (3.24) となる。

$$\frac{|y_{r,P_m}(n)|}{|x(n)|} > \frac{|y_{r,P_{m+1}}(n)|}{|y_{r,P_m}(n)|} \quad (3.24)$$

ここで、 $H_{r,P_m}(z)$ と $H_{r,P_{m+1}}(z)$ の傾き、及び、両音の振幅がほぼ等しいと仮定し、式 (3.24) を次のように近似的に表現する。

$$|\alpha_m| \cong |\alpha_{m+1}| = \alpha, \quad \sum s_{m,k} \cong \sum s_{m+1,k} \text{ より,}$$

$$\frac{\sum \alpha \cdot k \times \Delta f_{P_m}}{2} > \frac{\sum \alpha^2 \cdot k^2 \times \Delta f_{P_{m+1}} \cdot \Delta f_{P_m}}{\sum \alpha \cdot k \times \Delta f_{P_m}}$$

上記の関係より、

$$\overbrace{f_{q,P_{m+1}} - \Delta f_{P_{m+1}} - f_{q,P_m}}^{\Delta f_{P_m}} > 2\Delta f_{P_{m+1}}$$

従って、

$$\Delta f_{P_{m+1}} < \frac{f_{q,P_{m+1}} - f_{q,P_m}}{3} \quad (3.25)$$

式 (3.25) より、ピッチのずれが半音の周波数間隔の 1/3 程度ならば、各くし形フィルタの出力の減衰により音名判別が可能である。しかし、振幅 $s_{m+1,k}$ が小さい場合や、 P_{m+1} 音の周波数成分を極端に減衰させるくし形フィルタを通過した場合、許される変動幅は式 (3.25) より狭くなる。

3.3 くし形フィルタの零出力検出による音価推定

くし形フィルタ $H_{r,P}(z)$ の持つ周波数成分除去の性質を利用し、前節で示した音高推定システムにおける各 $H_{r,P}(z)$ の零・非零出力を比較することで、注目音の継続時間の検出が可能である。以下にその原理を説明する。

入力音とくし形フィルタが対応している場合、オーバーサンプリングによる近似、及び、楽器音の非定常性の影響から、完全に零出力にはならないものの、図3.17, 3.18に示すように、出力の振幅には大きな減衰が見られる。なお、立ち上がり直後の大きな雑音は、くし形フィルタの過渡現象によるものである。一方、くし形フィルタが入力音に対応していない場合、図3.19に示すように、一致している場合と比較して大きな振幅（非零値）が出力される。

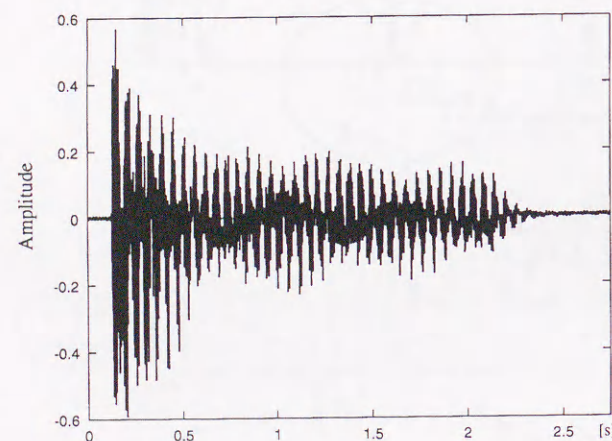


図 3.17 入力信号 piano C^3

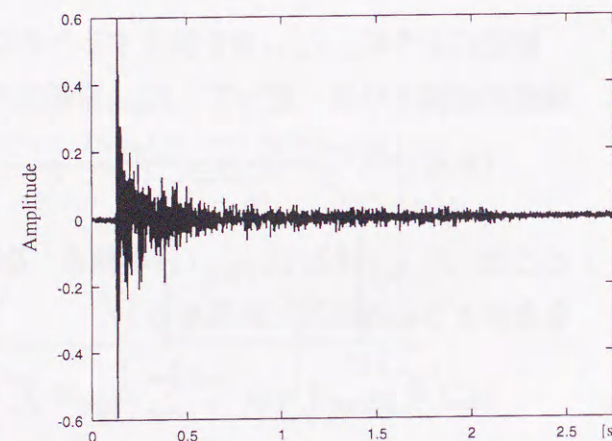


図 3.18 $H_{8,C}(z)$ の出力

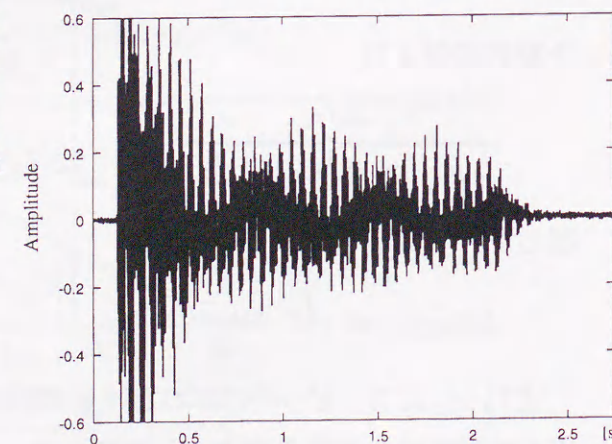


図 3.19 $H_{8,D}(z)$ の出力

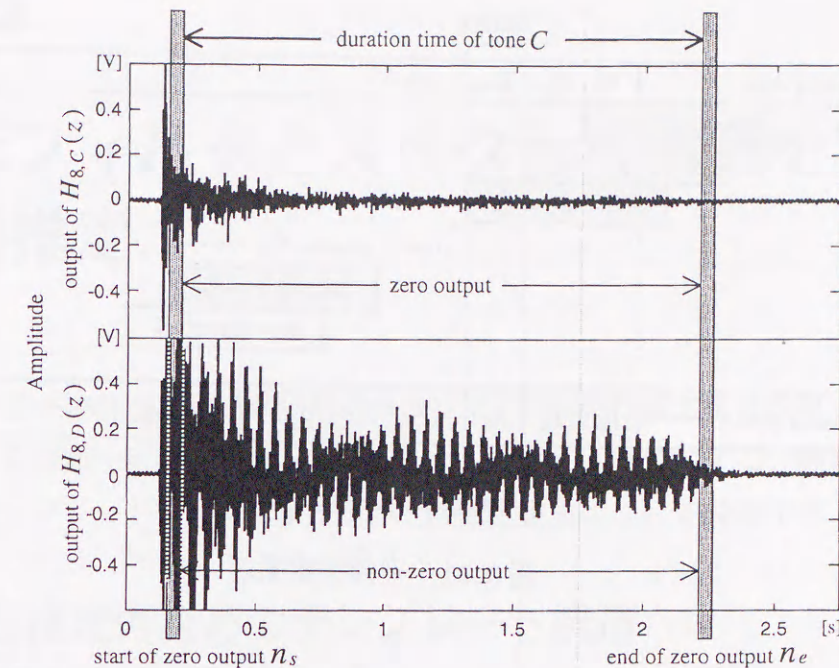


図 3.20 図 3.17 の入力に対する音価検出

従って、前述のシステムにおいて、 P_m 音に対応するくし形フィルタ $H_{r,P_m}(z)$ の出力が $y_{r,P_m}(n) \cong 0$ であり、それ以外の音に対応する $H_{r,P_i}(z)$ の出力が $y_{r,P_i}(n) \gg 0$ である間、 P_m 音が継続していることが判別可能である。即ち、図3.20に示すように、零出力の立ち下がり n_s から立ち上がり n_e を検出することで、以下のように音価を判別する。なお、 f_s は前述と同様、オーバーサンプリング周波数である。

$$\left. \begin{array}{l} y_{r,P_m}(n_s \rightarrow n_e) \cong 0 \\ y_{r,P_i}(n_s \rightarrow n_e) \gg 0 \end{array} \right\} \Rightarrow \text{duration time of tone } P_m : \frac{n_e - n_s}{f_s} [\text{s}]$$

実際には、零・非零出力の判別に一定のデータ長が必要であるため、入力信号を時間軸上で w サンプル間隔に分割し、各ブロック W_1, W_2, \dots において零出力検出処理を繰り返すことで音価推定を行なう。つまり、あるブロック W_i において音高 P_m を推定 ($y_{r,P_m}(z) \cong 0$ かつ $y_{r,P_i}(n) \gg 0$) し、一定の間隔 (R ブロック) で継続して推定された場合に、 $w \times R / f_s [\text{s}]$ を P_m 音の音価として検出する。

即ち、以下のように音価を推定する。

$$\left. \begin{array}{l} y_{r,P_m}(W_i \rightarrow W_{i+R'}) \cong 0 \\ y_{r,P_i}(W_i \rightarrow W_{i+R'}) \gg 0 \end{array} \right\} \Rightarrow \text{duration time of tone } P_m : \frac{w \times R}{f_s} [\text{s}]$$

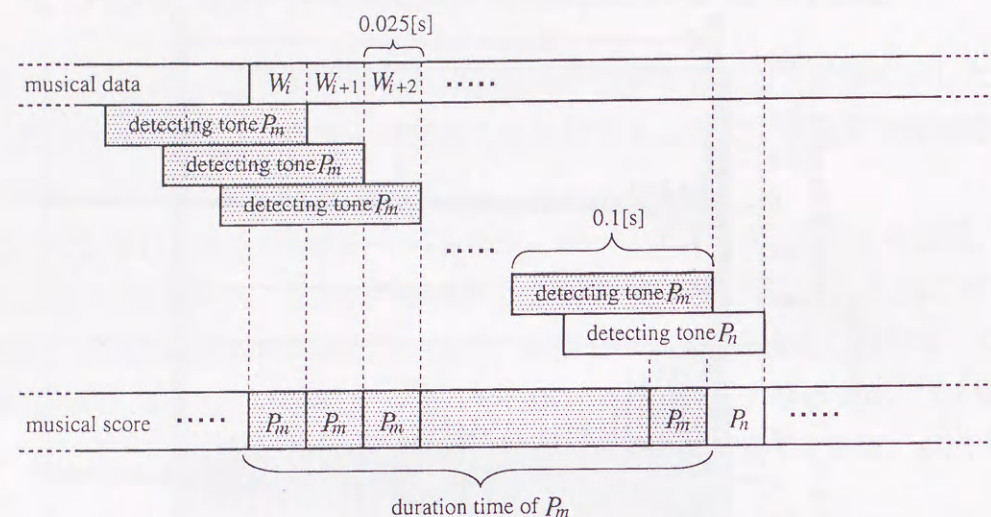


図 3.21 音価推定原理

なお、分割の最小単位 w 、最小継続ブロック数 R_{min} は、楽曲に含まれる最短音符 ($tempo = 120$ の 16 分音符で $0.125[s]$) が検出できる程度の長さが必要である。本方式では、 $f_s = 54.054[kHz]$ に対して $w = 1351$ サンプル点 ($0.025[s]$) を用い、 $0.125 \div 0.025 = 5$ より、最小継続ブロック数 $R_{min} = 5$ とする。なお、 $R < 5$ となる結果が得られた場合には検出不能であると判断する。

但し、本方式では、雑音や前述の $H_{r,p}(z)$ の過渡状態の影響を考慮し、図 3.21 に示すように、分割ブロック W_i より前の $0.75[s]$ を含む $0.1[s]$ (5405 サンプル点) の区間で音高推定処理を行ない、最後の $0.025[s]$ における結果を用いて音価を推定する。

第 4 章

くし形フィルタとスペクトル構造を利用した楽器推定

本章では、初めに演奏楽器推定のためのテンプレートの作成方法を説明する。次に、くし形フィルタを用いた和音の分離システムについて述べ、DFT による分離音のスペクトル解析法及びテンプレートマッチングを用いた楽器推定法について説明する。

4.1 楽器推定のためのテンプレート構造

2 章で述べたように、楽器音のスペクトル構造は楽器の発音、共鳴機構の違い、及び音域の違いにより異なる特徴を持つことが知られている。本方式では、スペクトル構造の特徴に注目し、各楽器音のスペクトルをテンプレートとして利用することで楽器推定を行なう。

なお、前述のように、楽器音は振幅が時間的に変化する非定常信号であり、また、楽器音の立ち上がりには多くの非調和周波数成分が含まれるため [44]、これらの影響を考慮する必要がある。例として、MIDI ピアノ C^4 音に対し、時間軸上で窓関数をずらすことによ

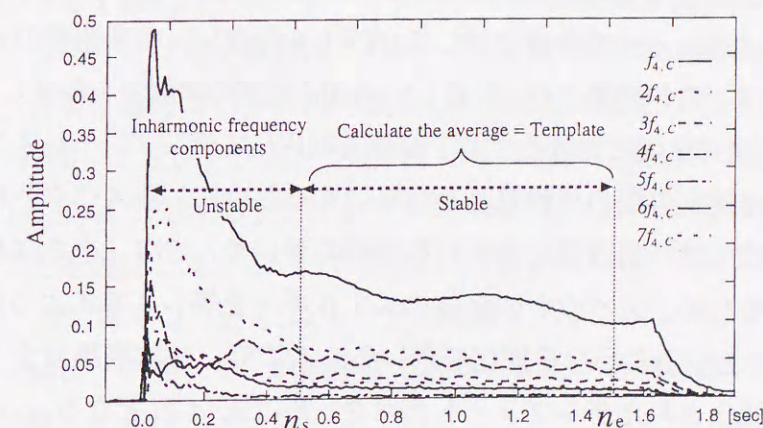


図 4.1 ピアノ音 C^4 の STFT 結果

表 4.1 各楽器音の n_s/fs

楽器	piano	violin	viola	flute	clarinet	horn	trumpet
n_s/fs [s]	0.50	0.10	0.35	0.20	0.15	0.10	0.10

り短時間フーリエ変換 (STFT) を行ない, 各高調波 $k \times f_{4,C}$ (k は整数) の振幅の時間変化を測定した結果を図 4.1 に示す. なお, サンプリング周波数 54.054 [kHz] に対し, 207 点の方形窓 (周波数分解能 261.13 [Hz] $\cong C^4 : 261.62$ [Hz]) を用いた.

図 4.1 より, 立ち上がり直後の $0 \sim n_s$ (約 0.5 [s]) 間において, 各高調波成分の振幅が不規則に大きく変動していることが分かる. また, 各高調波成分の振幅が比較的安定している状態 ($n_s \sim n_e \cong 1.0$ [s] 間) においても, 各々の成分は緩やかな振幅変動を持つことが分かる. このような非定常性の影響を考慮し, 楽器推定のためのテンプレートスペクトルは, 図 4.1 における安定区間 $n_s \sim n_e$ の各スペクトル振幅を平均した値を用いる.

ここで, 楽器 M の P^q 音 (オクターブ q , P 音) について, 時間軸上の各サンプル点で得られる k 次高調波のスペクトルを $S_{M,P^q}(n, k)$ とした場合, 平均化されたテンプレートスペクトル $\hat{T}_{M,P^q}(k)$ は, 式 (4.1) で計算される.

$$\begin{aligned} M &\in \{\text{piano, violin, } \dots \text{ etc.}\} \\ P^q &\in \{C^3, C^3\#, \dots, B^5\} \\ \hat{T}_{M,P^q}(k) &= \frac{1}{n_e - n_s} \sum_{n=n_s}^{n_e} S_{M,P^q}(n, k) \end{aligned} \quad (4.1)$$

なお, 安定状態に至るまでの時間 n_s/fs [s] は楽器により異なるため, 予め, 個別に調べる必要がある. ここで, 5 章の計算機シミュレーションで使用する piano, violin, viola, flute, clarinet, horn, trumpet の各音に対して STFT を行ない, 各楽器音の n_s/fs を計算した結果を表 4.1 に示す. この結果より, 予め 2.5 ~ 3.0 [s] 程度の楽器音を録音し, 各楽器音に対して n_s より 1.5 [s] 程度の区間における STFT 結果を利用してテンプレートスペクトルを作成した.

さらに, 楽器推定の際の比較処理を容易にするために, 各スペクトルに対して正規化を行なう. 本方式では, 楽器音に含まれる振幅の著しく大きな (小さな) 高調波成分による悪影響を避けるため, スペクトル全体のエネルギーを等しくするような, 即ち, $\hat{T}_{M,P^q}(k)$ のノルムを 1.0 とするような正規化を用いる. ここで, K 次高調波までを楽器推定に使用する場合, 正規化されたテンプレートスペクトルは式 (4.2) となる.

$$T_{M,P^q}(k) = \frac{\hat{T}_{M,P^q}(k)}{\sum_{j=1}^K \hat{T}_{M,P^q}(j)} \quad (4.2)$$

表 4.2 グループ化したテンプレートスペクトルに含まれる音高 P^q

M	piano	violin	viola	flute
$T_{M,1}(k)$	$C^3 \sim D^3$	C^3	$C^3 \sim C\#^3$	$C^3 \sim C\#^4$
$T_{M,2}(k)$	$D\#^3 \sim F\#^3$	$C\#^3 \sim G^3$	$D^3 \sim A^3$	D^4
$T_{M,3}(k)$	$G^3 \sim D^4$	$G\#^3 \sim D^4$	$A\#^3 \sim G^4$	$D\#^4 \sim F\#^4$
$T_{M,4}(k)$	$D\#^4 \sim A^4$	$D\#^4 \sim A^4$	$G\#^4 \sim B^5$	$G^4 \sim A\#^4$
$T_{M,5}(k)$	$A\#^4 \sim C\#^5$	$A\#^4 \sim F^5$		$B^4 \sim G^5$
$T_{M,6}(k)$	$D^5 \sim A^5$	$F\#^5 \sim B^5$		$G\#^5 \sim B^5$
$T_{M,7}(k)$	$A\#^5 \sim B^5$			
M	clarinet	horn	trumpet	
$T_{M,1}(k)$	$C^3 \sim D^3$	$C^3 \sim C\#^3$	$C^3 \sim C\#^3$	
$T_{M,2}(k)$	$D\#^3 \sim G^3$	$D^3 \sim F\#^3$	$D^3 \sim G\#^3$	
$T_{M,3}(k)$	$G\#^3 \sim C^4$	$G^3 \sim A\#^3$	$A^3 \sim C\#^4$	
$T_{M,4}(k)$	$C^4 \sim G^4$	$B^3 \sim C\#^4$	$D^4 \sim B^4$	
$T_{M,5}(k)$	$G\#^4 \sim C^5$	$D^4 \sim A^4$	$C^5 \sim G^5$	
$T_{M,6}(k)$	$C\#^5 \sim G^5$	$A\#^4 \sim B^5$	$G\#^5 \sim B^5$	
$T_{M,7}(k)$	$G\#^5 \sim B^5$			

なお, 楽器音は音域によりスペクトルの形が若干変化するため, 1 種類の楽器に対するテンプレートには非常に多くのスペクトルが必要となる問題がある. 近年の一般的な電子楽器は, 各楽器に対して, 予め幾つかの音高波形をサンプリングし, その波形を変調することで, 幅広い音域の音を合成している [41]. この性質を利用し, 本方式では $C^3, C\#^3, \dots, B^5$ の 36 音に対して, 類似したスペクトル構造を持つ近隣の音をグループ化し, 同一グループに含まれる複数の音を 1 種類のテンプレートスペクトルで表現することで, 上述の問題を解決する. 実際の楽器音には電子楽器ほどの同一性はないものの, 演奏条件に著しい違いが無い場合には, 近隣音域のスペクトルは類似した形になることが知られている [44].

なお, 5 章の計算機シミュレーションで使用する電子楽器の piano, violin, viola, flute, clarinet, horn, trumpet に対するテンプレートスペクトルは, 表 4.2 に示すようにグループ化可能であることを確認した. ここで, $T_{M,1}(k), T_{M,2}(k), \dots$ はテンプレートスペクトルのグループ 1, 2, \dots を意味する. 表 4.2 より, 予め用意するテンプレートスペクトルの数は, 最大でも piano, clarinet の場合で 7 種類程度である. また, 各楽器音に対するテンプレートスペクトルを付録 A に示す.

4.2 くし形フィルタによる和音分離とスペクトル抽出

本節では、くし形フィルタ $H_{r,p}(z)$ を利用した基本的な和音分離システムについて述べる。さらに、分離された各音のスペクトル解析を行ない、前述のテンプレートスペクトルを比較することで楽器推定を行なう方式について説明する。

なお、3章で示したように、平均率音階には周波数成分に重複関係を持つ2音間隔があり、このような関係にある和音が入力された場合、 $H_{r,p}(z)$ による和音分離は困難となる（この問題に関しては、4.3で詳しく説明する）。

4.2.1 基本的な和音分離システムの構成

2章で述べたくし形フィルタの周波数成分除去の性質を利用し、 m 種類の音を含む異種複数楽器音 $x(n) = \sum \text{tone } P_i$ ($P_i \in \{C, C\#, \dots, B\}$, $1 \leq i \leq m$, $m \leq 12$) に対して、式(4.3)に示す合成フィルタ $G_{r,P_i}(z)$ を使用することで、 P_i 音の抽出が可能である。

$$G_{r,P_i}(z) = \prod_{j=1, j \neq i}^m H_{r,P_j}(z) \quad (4.3)$$

この $G_{r,P_i}(z)$ は、入力 $x(n)$ に含まれる m 種類の音から、注目する P_i 音以外の $m-1$ 音を除去する性質を持つ。この性質を利用し、図4.2に示すように、 $x(n)$ に含まれる全ての音に対して $G_{r,P_i}(z)$ を構成することで、 $x(n)$ が m 個の単音に分離される。

なお、 $G_{r,P_i}(z)$ の周波数特性は式(4.4)となるため、抽出音の各周波数成分の振幅は、 $|G_{r,P_i}(e^{j\omega})|$ で計算される利得が乗じられた値に変化する。ここで、 f_s はオーバーサンプリング

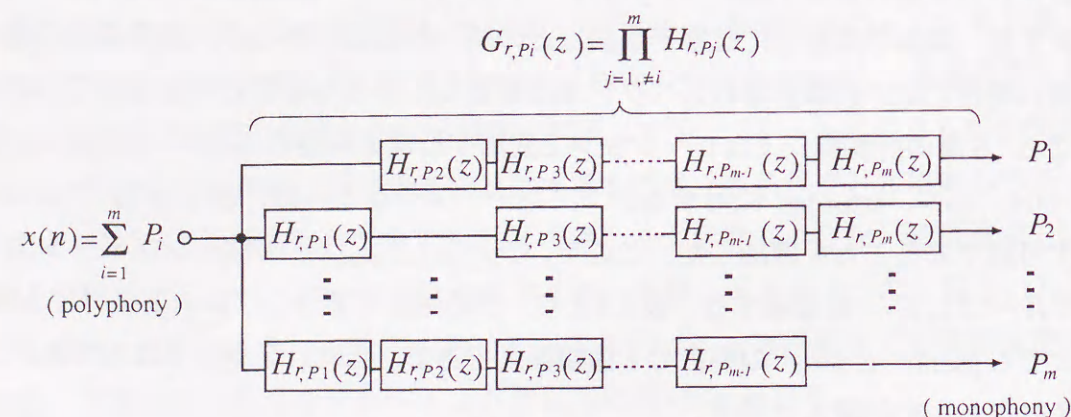


図4.2 和音分離システム

周波数であり、 r_{P_j} は式(3.10)に示したオーバーサンプリング法くし形フィルタの次数である。

$$\begin{aligned} |G_{r,P_i}(e^{j\omega})| &= \prod_{j=1, j \neq i}^m |H_{r,P_j}(e^{j\omega})| = \prod_{j=1, j \neq i}^m \left| 1 - \exp\left(-j \frac{r_{P_j} \times 2\pi f}{f_s}\right) \right| \\ &= \prod_{j=1, j \neq i}^m \sqrt{2 - 2 \cos\left(\frac{r_{P_j} \times 2\pi f}{f_s}\right)} \end{aligned} \quad (4.4)$$

4.2.2 DFT によるスペクトル解析

和音から抽出された各 P 音に対して、本方式では、DFT を用いてスペクトル解析を行なう。

前述のように、 P 音の周波数成分は $f_{3,P}$ の整数倍に制限されているため、オーバーサンプリング周波数 f_s に対し、式(4.5)に示す N_P 点の DFT を用いることで周波数解析が可能となる。ここで、 P 音のサンプル値系列 $u_P(n)$ のスペクトルは式(4.6)で計算される。

$$N_P = \frac{2\pi f_s}{\omega_1} = \frac{f_s}{f_{3,P}} \quad [\text{四捨五入で整数化}] \quad (4.5)$$

$$\hat{U}_P(e^{j\omega_k}) = \sum_{n=0}^{N_P} u_P(n) \exp\left(-j \frac{2\pi k}{N_P} n\right) \quad (4.6)$$

ここで、後述の計算機シミュレーションで使用する $f_s = 54.054[\text{kHz}]$ に基づいて計算した各 N_P 、及び、周波数分解能 f_s/N_P [Hz] を表4.3に示す。この表より、219点~413点程度

表4.3 各音に対応した N_P と周波数分解能 f_s/N_P

音名	N_P	$f_{3,P}$ [Hz]	f_s/N_P [Hz]
C	413	130.81	130.88
C#	390	138.59	138.60
D	368	146.83	146.89
D#	347	155.56	155.78
E	328	164.81	164.80
F	310	174.61	174.37
F#	292	185.00	185.12
G	276	196.00	195.85
G#	260	207.65	207.90
A	246	220.00	219.73
A#	232	233.08	232.99
B	219	246.94	246.82

のDFTにより各音の周波数解析が可能であることが分かる。また、各 N_P における周波数分解能と基本周波数の差は、最大で A^3 における $0.27[\text{Hz}]$ 程度である。

なお、前述のように、合成フィルタ $G_{r,P}(z)$ の利得の影響から、抽出音の各 $\hat{U}_P(e^{j\omega_k})$ は元信号と異なる値に変化している。従って、楽器推定を行なう前に、各周波数成分 $k \times f_{3,P}$ に対応する $|G_{r,P}(e^{j\omega_k})|$ で除算し、元信号の振幅を復元する必要がある。さらに、前述のテンプレートスペクトル $T_{M,Pq}(k)$ と同様に、ノルム = 1.0 となるように正規化することで、楽器推定のための比較に使用する抽出音のスペクトル $U_{Pq}(k)$ は、式(4.7)となる。なお、 K は前述と同様にテンプレートスペクトルとの比較に使用する高調波成分の数である。

$$U_{Pq}(k) = \frac{1}{|G_{r,Pq}(e^{j\omega_k})|} \times \frac{\hat{U}_{Pq}(e^{j\omega_k})}{\sum_{i=1}^K \hat{U}_{Pq}(e^{j\omega_k})} \quad (4.7)$$

4.2.3 テンプレートマッチングによる楽器推定

本方式では、2章で述べたような同一楽器音におけるスペクトル構造の類似性を利用し、テンプレートスペクトルと抽出音スペクトルを比較することで演奏楽器の推定を行なう。

前述のような正規化を行なうことで、式(4.8)に示すように、分離音のスペクトル $\{U_{Pq}(k)\}$ と各楽器のテンプレートスペクトル $\{T_{M,Pq}(k)\}$ の差分 $E_{M,P}$ を計算し、 $E_{M,P}$ が最小となる楽器 M を見つけることで演奏楽器が推定される。

なお、前述と同様に $M \in \{\text{piano, violin, } \dots \text{ etc.}\}$, $P^q \in \{C^3, C\#^3, \dots, B^5\}$ であり、比較には K 次高調波までを使用する。

$$E_{M,Pq} = \sum_{k=1}^K |U_{Pq}(k) - T_{M,Pq}(k)|$$

$$\min_M [E_{M,Pq}] \rightarrow \text{discrimination of instrument } M \quad (4.8)$$

4.3 くし形フィルタによる和音分離の問題点

4.2で初めに述べたように、図4.2に示す和音分離システムは、くし形フィルタの周波数成分除去の性質を利用し、和音から1音のみを抽出する（他音を除去する）ことで入力音の分離を行なっている。このため、2章で述べたような周波数成分が重複する和音に対しては、ある音の高調波（及び基本周波数）が他音に対応するくし形フィルタで除去されてしまい、各音の周波数成分を全て抽出することが困難となるため、その後の楽器推定に問題が生じる。本節では、このような周波数成分の重複により生じる問題点について明らかにする。

オクターブ q の P 音の基本周波数を $f_{q,P}$ で表すと、平均率音階では、隣り合う音の基本周波数は $2^{1/12}$ 倍であり、式(4.9)の関係が成り立つ。ここで、 P_{+d} は基準とする音 P から高い方へ d 音の隔たりの持つ音を意味する。

$$\left. \begin{aligned} \frac{f_{q,P+1}}{f_{q,P}} &= 2^{\frac{1}{12}} \cong 1.059463 \\ \frac{f_{q,P+12}}{f_{q,P}} &= 2^{\frac{12}{12}} = 2.0 \quad (1 \text{ オクターブ}) \\ \frac{f_{q,P+19}}{f_{q,P}} &= 2^{\frac{19}{12}} \cong 2.996614 \cong 3.0 \\ \frac{f_{q,P+24}}{f_{q,P}} &= 2^{\frac{24}{12}} = 4.0 \quad (2 \text{ オクターブ}) \\ \frac{f_{q,P+28}}{f_{q,P}} &= 2^{\frac{28}{12}} \cong 5.039684 \cong 5.0 \\ &\dots \text{etc.} \end{aligned} \right\} \quad (4.9)$$

式(4.9)より、基本周波数が 2^i 倍となる $12i$ (i は整数であり、 i オクターブ違いの同音を意味する) 音間隔以外に、整数倍に非常に近い周波数比を持つ2音間隔 d が存在することが分かる。このような関係を持つ $H_{r,P}(z)$, $H_{r,P+d}(z)$ の零点は近接する箇所が存在し、和音分離の際に、 $H_{r,P}(z)$ で P_{+d} 音の重複零点に対応する周波数成分が非常に小さな値に減衰してしまう（また、その逆も有り得る）。さらに、分離に用いる合成フィルタの振幅 $|G_{r,P}(e^{j\omega})|$ が非常に小さな値となるため、式(4.7)に示した振幅補正における除算の際に、誤差や雑音成分の影響を無視できなくなる。

$H_{r,P}(z)$ の零点が近接するためスペクトル抽出に不都合が生じる例として、異種複数楽器音 (piano C^4 + flute E^4 + violin G^4) から抽出した C^4 のSTFT結果 ($|G_{r,P}(e^{j\omega})|$ による

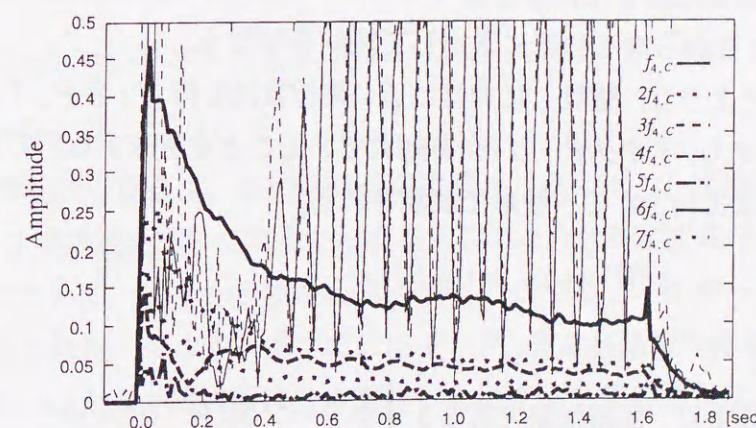


図4.3 抽出音 C^4 のSTFT結果

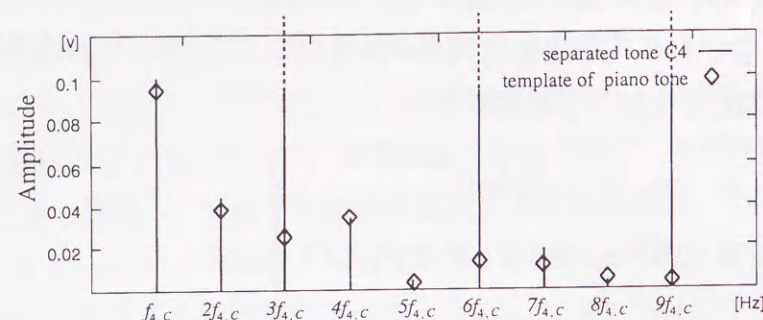


図 4.4 抽出音 C^4 のスペクトル

補正後)を図 4.3に, そのスペクトル抽出(復元)結果を図 4.4に示す. また, 比較として piano のテンプレートスペクトルを図 4.4に点◇で示す.

この和音中では, C^4 と $2 \times G^4$ が式(4.9)における $f_{q,P+12i}/f_{q,P} \cong 3.0$ の関係となり, $f_{4,C}$ の $3k$ 次高調波(k は整数)に対する $|G_{r,P}(e^{j\omega})|$ が微小な値となる. これにより, 図 4.3に示すように, $3k$ 次高調波成分($3k \times f_{4,C}$)に発振を伴う大きな誤差が生じ, 正しくスペクトルが復元されないことが分かる. また, オクターブ違いの同音では $H_{r,P}(z)$ の零点が完全に重複するため, $H_{r,P}(z)$ を用いた和音分離は実質不可能であり, 同様に楽器推定が困難となる. この問題の解決法について, 次節で説明を行なう.

4.4 和音分離の問題点の解決

本節では, 前述の問題点を解決するために, 周波数成分の重複により和音を4種類に分類し, 各々の和音に適した和音分離(スペクトル抽出)法を検討する. なお, 説明の簡単化のために, 2音からなる和音に限定して説明を行なう.

前節で述べたような, 整数 L に非常に近い周波数比を持つ2音 P, P_{+d} の間隔 d を $12i + \hat{d}$ (i, \hat{d} は整数であり, i オクターブと \hat{d} 音離れていることを意味する) で表現すると, 式(4.9)に示した関係は式(4.10)の一般形で表現される.

$$\begin{aligned} \frac{f_{q,P+12i+\hat{d}}}{f_{q,P}} &= 2^{\frac{12i+\hat{d}}{12}} \cong L \text{ より,} \\ f_{q,P+12i+\hat{d}} &\cong L \times f_{q,P} \end{aligned} \quad (4.10)$$

式(4.10)の関係より, 採譜対象である3オクターブ36音($C^3, C\sharp^3, \dots, B^5$)において, 基本周波数に重複(近接)が生じる2音間隔 d を表 4.4に示す. なお, ここではオクターブ q の P 音を基準とする. この表では, 例えば $d = 12i + 7$ の場合, 基本周波数 $f_{q,P}, f_{q,P+12i+7}$

表 4.4 基本周波数が重複する2音の間隔 d と周波数比

2音間隔 d	周波数比 $f_{q,P+d} \div f_{q,P}$	基本周波数の重複
$12i$	$2^{12i/12} = 2.0^i$	$2^i f_{q,P} = f_{q+i,P}$
$12i + 7 \ (i \geq 1)$	$2^{(12i+7)/12} \cong 2^{i-1} \times 3.0$	$2^{i-1} \times 3 f_{q,P} \cong f_{q+i,P+7}$
$12i + 4 \ (i \geq 2)$	$2^{(12i+4)/12} \cong 2^{i-2} \times 5.0$	$2^{i-2} \times 5 f_{q,P} \cong f_{q+i,P+4}$

表 4.5 高調波のみが重複する2音の間隔 k と周波数比

2音間隔 d	周波数比 $f_{q,P+d} \div f_{q,P}$	2音の周波数関係
$12i + 7 \ (i < 1)$	$2^{(12i+7)/12} \cong 2^{i-1} \times 3.0$	$3 f_{q,P} \cong 2 f_{q,P+7}$
$12i + 4 \ (i < 2)$	$2^{(12i+4)/12} \cong 2^{i-2} \times 5.0$	$5 f_{q,P} \cong 4 f_{q,P+4}$
$12i + 2 \ (i \geq 3)$	$2^{(12i+2)/12} \cong 2^{i-3} \times 9.0$	$9 f_{q,P} \cong 8 f_{q,P+2}$
$12i + 1 \ (i \geq 4)$	$2^{(12i+1)/12} \cong 2^{i-4} \times 17.0$	$17 f_{q,P} \cong 16 f_{q,P+1}$
$12i + 3 \ (i \geq 4)$	$2^{(12i+3)/12} \cong 2^{i-4} \times 19.0$	$19 f_{q,P} \cong 16 f_{q,P+3}$

の比が $2^{i-1} \times 3.0$ となり, オクターブ $q+i$ の P_{+7} 音の基本周波数がオクターブ q の P 音の $2^{i-1} \times 3$ 次高調波と非常に近い周波数値を持つことを示している.

ここで, 式(4.10)における2音 $P, P_{+12i+\hat{d}}$ を同一オクターブで考えると, その周波数関係は式(4.11)となる.

$$\begin{aligned} f_{q,P+12i+\hat{d}} &= f_{q+i,P+\hat{d}} = 2^i \times f_{q,P} \\ 2^i \times f_{q,P+\hat{d}} &\cong L \times f_{q,P} \end{aligned} \quad (4.11)$$

さらに, j オクターブ違いの周波数関係を考慮することで式(4.12)が導かれ, 式(4.10)の関係を持つ2音 P, P_{+d} は, そのオクターブに依らず高調波の一部に重複を持つことが分かる.

$$\begin{aligned} 2^i \times f_{q+j,P+\hat{d}} &= 2^j \times L \times f_{q,P} \text{ あるいは,} \\ 2^j \times 2^i \times f_{q,P+\hat{d}} &= L \times f_{q+j,P} \end{aligned} \quad (4.12)$$

このような高調波間の重複は, その整数倍の高調波(例えば, L 次高調波が重複するならば, $2L, 3L, \dots$ 次高調波)成分にも当てはまる. このような高調波成分の一部が重複する2音 $P, P_{+12i+\hat{d}}$ について, 主なものを表 4.5に示す. ここでは, オクターブ q の P 音を基準とし, 周波数関係には同一オクターブにおける P, P_{+d} 音の高調波の重複状態のみを示す.

ここで, 表 4.4, 4.5から, 周波数成分の重複状態による2音の関係は, 次の4種類に分類することができる.

- (1) 2音の周波数成分がほぼ独立

- (2) 2音の高調波成分の一部が重複 (表 4.5)
- (3) 一方の音の基本周波数成分が他方の高調波成分に重複 (表 4.4)
- (4) 2音の基本周波数成分及び全ての高調波成分が重複 (同一音高)

上記の各和音について、従来法による解析の問題点とその解決法を以下に説明する。

4.4.1 2音の周波数成分が独立な和音

図 4.5 に示すように、2音の基本周波数成分及び全高調波がほぼ独立である場合には、前章で述べた音高推定が可能であり、前述の和音分離、スペクトル抽出 (楽器推定) が可能となる。この原理について、piano C^3 + violin A^3 の和音に対して C^3 音の楽器推定を行なう様子を図 4.6 に示す。

ここで、図 4.6 には、

- (a) 入力音 piano C^3 + violin A^3 のスペクトル
- (b) $H_{8,A}(z)$ による A^3 音の除去 (C^3 音の抽出) の様子
- (c) 抽出音 C_3 に対する $|H_{8,A}(e^{j\omega})|$ による振幅補正とテンプレートスペクトル (piano ●, violin ○ で表現) との比較の様子

を、それぞれ示している。この場合、分離された C^3 音のスペクトル ($|H_{8,A}(e^{j\omega})|$ による補正後) は piano のテンプレート●とほぼ一致しているため、演奏楽器は piano であることが

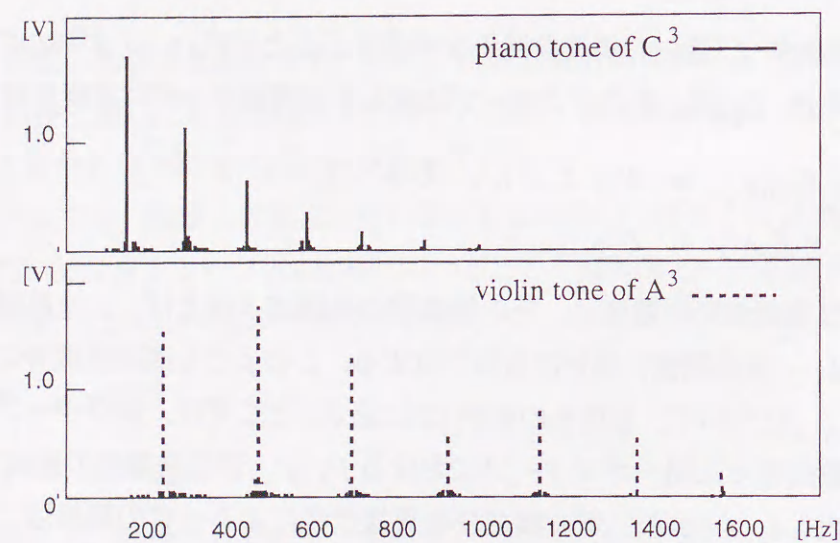


図 4.5 周波数成分が独立な和音の例

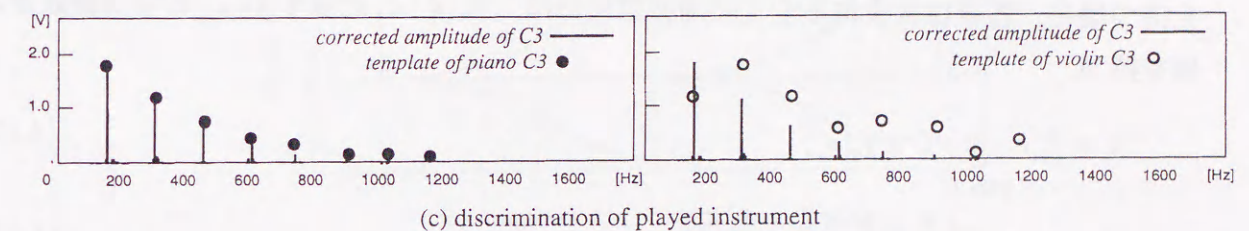
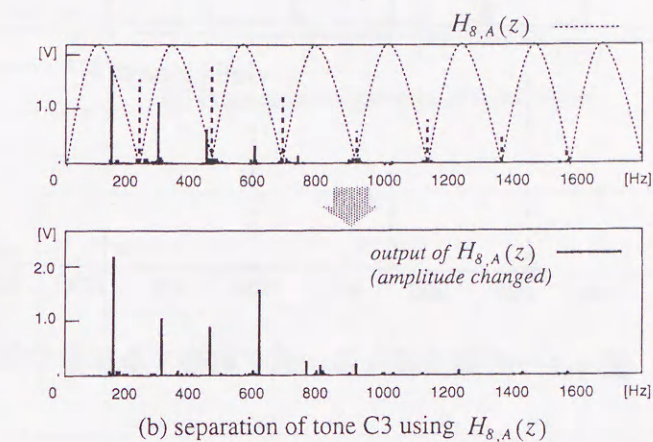
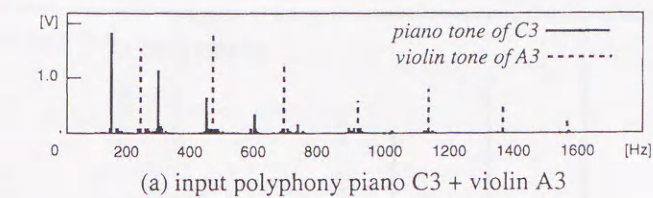


図 4.6 周波数成分が重複しない2音の分離の例

判別できる。

4.4.2 2音の高調波成分の一部が重複する和音

表 4.5 に示したような和音は、2音の高調波成分の一部に重複関係を持つものの、図 4.7 のように基本周波数成分が互いに独立であるため、予め音高の判別が可能である。しかし、重複成分に対応する互いの $|H_{r,P}(e^{j\omega})|$ が微小な値となり、スペクトル復元の際の振幅補正 ($|H_{r,P}(e^{j\omega})|$ による除算) で誤差が増大するため、復元したスペクトルの精度に問題が生じる。この問題を回避するために、重複成分を合成された1つの成分とみなし、和音全体に対して DFT を行なうことでスペクトルを解析する。なお、このような和音を解析する場合の DFT のサンプル点数は、以下のように求められる。

式 (4.13) に示すように、 P 音の k 次高調波と P_{+d} 音の L 次高調波が非常に近い周波数値

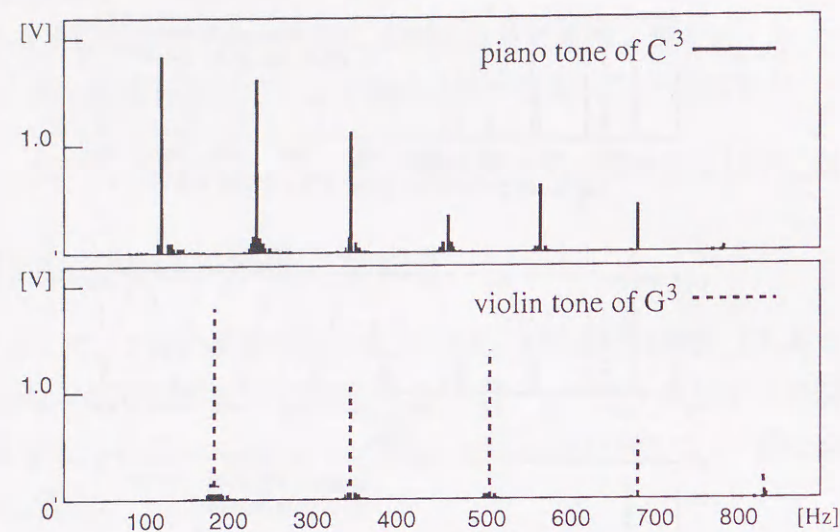


図 4.7 高調波成分の一部が重複する和音の例

を持つ場合、各音に含まれる全ての周波数成分は、式 (4.14) に示す $f_{P,P+d}$ のほぼ整数倍の値を持つ。

$$k \times f_{q,P} \cong L \times f_{q,P+d} \quad (4.13)$$

$$f_{P,P+d} = \frac{k \times f_{q,P} + L \times f_{q,P+d}}{2Lk} \quad (4.14)$$

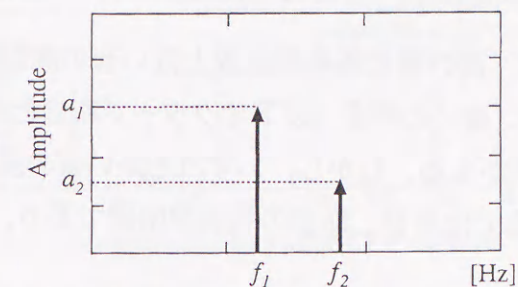
なお、

$$\begin{aligned} L \times f_{P,P+d} &\cong f_{i,P} \\ k \times f_{P,P+d} &\cong f_{q,P+d} \end{aligned} \quad (4.15)$$

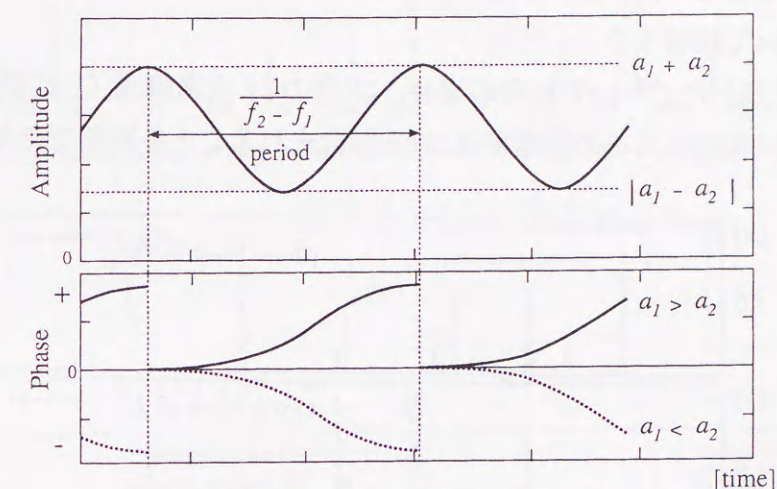
である。この性質を利用し、式 (4.5) に示す $N_P = 2\pi f_s / f_{3,P}$ [四捨五入で整数化] の代わりに、式 (4.16) による $N_{P,P+d}$ 点 DFT を用いることで和音全体の DFT が可能となる。なお、 f_s は前述と同様にオーバーサンプリング周波数である。

$$N_{P,P+d} = \frac{f_s}{f_{P,P+d}} \quad [\text{四捨五入で整数化}] \quad (4.16)$$

但し、上述の重複成分 $k \times f_{q,P}$ 、 $L \times f_{q,P+d}$ は周波数に若干のずれを持つため、合成成分の振幅は時間軸上で増減を繰り返すビートの特性を持つ。このようなビート特性について、例として近接する 2 成分 f_1 、 f_2 ($f_1 < f_2$) からなる合成信号の振幅変化、及び、その位相変化を図 4.8 に示し、以下にその性質を説明する。



(a) Frequency components of input signal



(b) Amplitude and phase of input signal(a)

図 4.8 近接成分 f_1 、 f_2 によるビート特性

図 4.8 より、合成信号の振幅変動の最大値は各成分の振幅の和 $a_1 + a_2$ 、最小値は振幅の差 $|a_1 - a_2|$ 、振幅変動（ビート）の周期は $1/|f_1 - f_2|$ となることが分かる。また、振幅が最大値をとる時点を基準とした場合、合成信号の位相は $a_1 > a_2$ ならば増加、 $a_1 < a_2$ ならば減少する性質を持つ。この性質を利用し、上述の重複成分 $k \times f_{q,P}$ 、 $L \times f_{q,P+d}$ についても同様に各成分の振幅、位相を推定することが可能であると考えられる。

さらに、得られた全てのスペクトルに対し、それぞれの音に属する成分をグルーピングすることで和音分離を行なう。分離された各音成分に対して、前述と同様にテンプレートスペクトルとの比較を行ない、楽器を推定する。但し、このためには解析データにある程度の長さ、例えば、 $3 \times C^3$ と $2 \times G^3$ の場合には $1/(3 \times 130.81 - 2 \times 196.0) \cong 2.33[s]$ が必要となる。

4.4.3 1音の基本周波数が他音の高調波成分に重複する和音

表4.4に示したような、高い音の基本周波数と低い音の高調波成分が重複する和音には、図4.9のようなオクターブ違いの同音（以下オクターブ和音と呼ぶ）と、図4.10のような音名の異なる和音の2種類がある。しかし、いずれも高い音の基本周波数が重複を持つため、前章で述べた音高判別法では単音、和音の判別が困難であり、音高推定の時点では低い音のみが検出される。

このような和音に対しては、和音中に含まれる他音と重複を持たない（独立な）スペクトルに注目することで、1音ずつ順にスペクトルの分離、楽器推定を行なう[36]。以下に、この原理について説明する。

図4.10(a)に示した $C^3 + G^4$ 和音の場合、 C^3 音の $3k$ 次高調波（ k は整数）と G^4 音の全ての周波数成分が重複し、この段階では G^4 音が含まれることを判別できない。ここで、和音

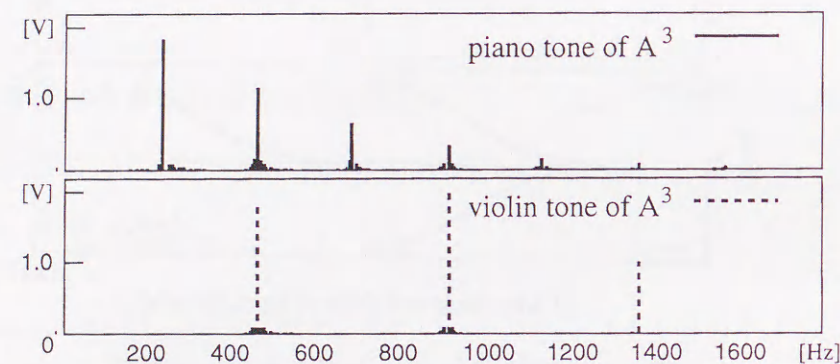


図4.9 オクターブ和音の周波数成分の重複例

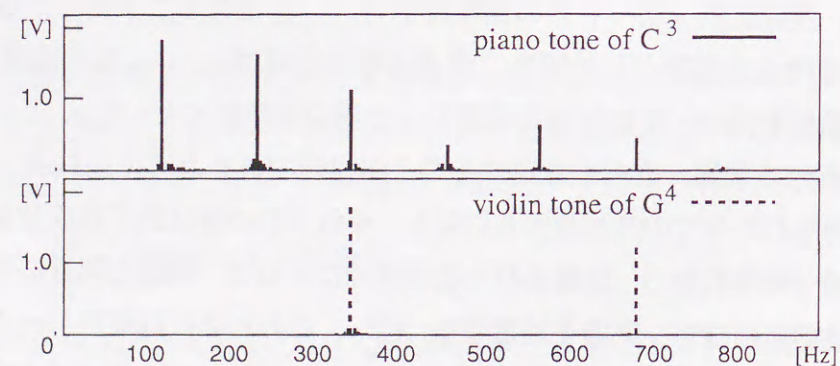
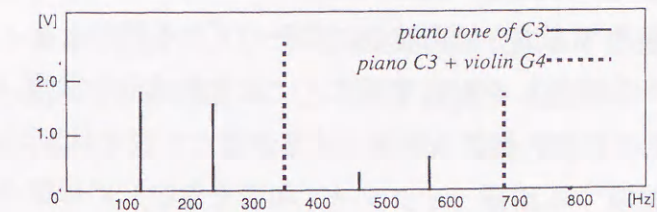
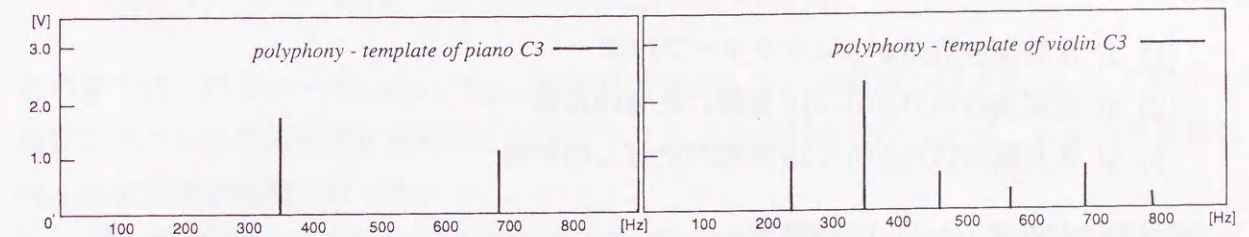


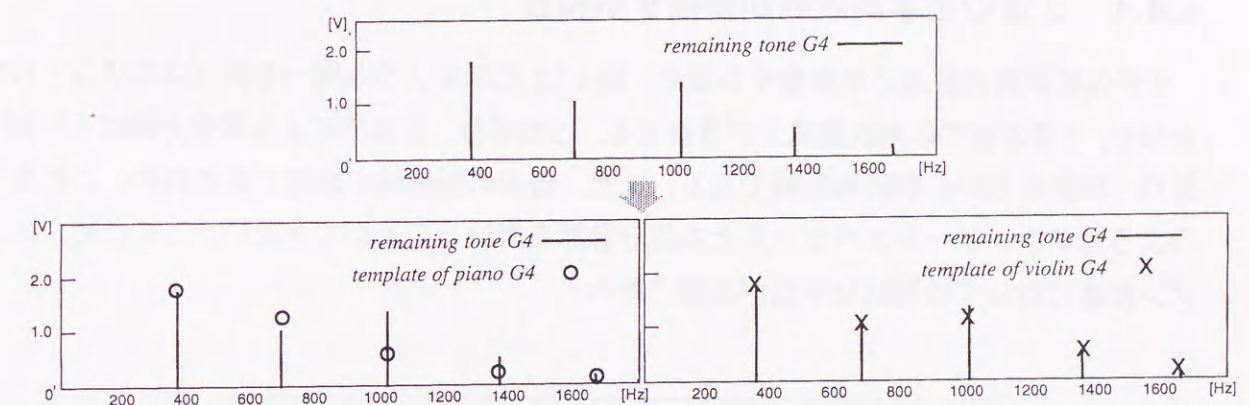
図4.10 基本周波数が他音の高調波成分に重複する和音の例



(a) input polyphony piano C3 + violin G4



(b) discrimination of playing instrument of C3 and detection of remaining tone G4



(c) discrimination of playing instrument of G4

図4.11 基本周波数が他音に重複する和音の分離

中の最も低い周波数成分 $f_{3,C}$ に注目し、図4.11(b)に示すように、和音全体の複素スペクトルとの C_3 音に対応するテンプレートスペクトルの差分を計算する。なお、図4.11(b)には、差分結果のパワースペクトルを示した。この場合、pianoのテンプレートスペクトルとの差分値の $3k$ 次高調波以外がほぼ0であることから、 C^3 音の演奏楽器がpianoであることが判別できる。また、この結果として、 C^3 の3次高調波 $\cong G^4$ であることから、もう1種類の音が G^4 音であることが判別できる。

さらに、図4.11(c)に示すように、残された G^4 音成分と対応するテンプレートスペクトルとの比較を行なう。この場合、 \times で示すviolinのテンプレートスペクトルと G^4 音のスペ

クトルがほぼ一致しているため、 G^4 音の楽器が violin であることが判別される。

なお、和音中の重複する成分は周波数に若干のずれを持つため、前述のように、時間軸上の振幅変動（ビート特性）を観察することにより各成分の振幅、位相を求める必要があり、解析データにある程度の長さを必要とする。また、表 4.4 に示したように、基本周波数が他音の高調波と重複する和音（オクターブ和音を含む）には様々な種類がある。このため、和音のスペクトルとテンプレートスペクトルの差分を計算する際に、

- (1) $2k$ 次高調波以外が $0 \rightarrow$ オクターブ和音
- (2) $3k$ 次高調波以外が $0 \rightarrow 19$ 音離れた音の和音
- (3) $5k$ 次高調波以外が $0 \rightarrow 28$ 音離れた音との和音

等の判別が必要となる（ k は整数）。

4.4.4 2 音の基本周波数が重複する和音

2 音の基本周波数成分が重複する場合、図 4.12 に示すような同一音高（ユニゾン）和音となり、2 音の全ての周波数成分が重複する。この場合、従来法による和音分離は不可能であり、和音か否かの判別も困難である。また、各音の振幅値が未知であるため、これまでのようなテンプレートスペクトルとの差分計算を用いることができない。このため、ユニゾン和音に対しては特別な手法が必要である。

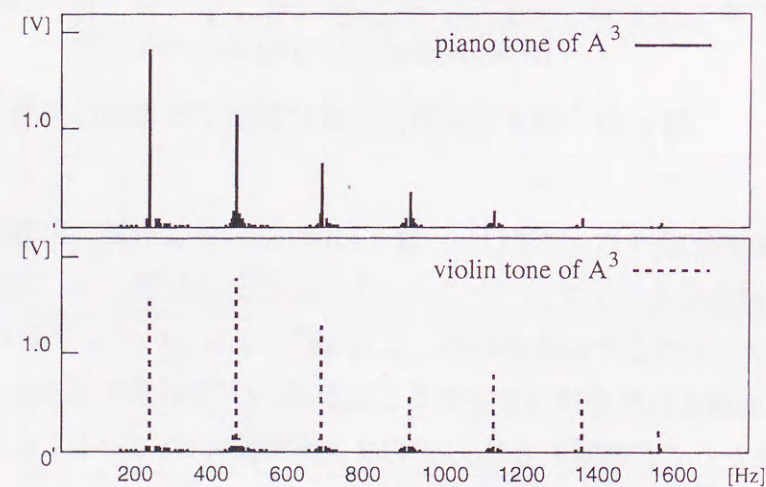


図 4.12 ユニゾン和音の周波数成分の重複例

ここで、ある時刻（サンプル時点 n ）における楽器 $M \in \{\text{piano, violin, ...etc.}\}$ の P^q 音の k 次高調波のスペクトルを $S_{M,P^q}(n, k)$ 、楽器 M の振幅（音量）を a_{M,P^q} 、合成時の位相を θ_{M,P^q} とする。このとき、入力されたユニゾン和音 P^q の k 次高調波のスペクトル $S_{P^q}(n, k)$ は式 (4.17) で表現される。ここで、演奏楽器の推定には各パラメータ a_{M,P^q} 、 θ_{M,P^q} の計算が必要である。

$$S_{P^q}(n, k) = \sum_M a_{M,P^q} \cdot S_{M,P^q}(n, k) \cdot \exp(j \cdot k \times \theta_{M,P^q}) \quad (4.17)$$

これまでに、パラメータ a_{M,P^q} 、 θ_{M,P^q} の計算方法として、非線形方程式を用いる手法 [32]、複素スペクトルの時間変化を利用する手法 [33] などの検討を行なったが、いずれも多数の極小解の存在が問題となった。

このため、ユニゾン和音に対する和音分離、楽器推定法については、未だ有効な手法が確立されておらず、現在も検討中である。

第 5 章

実験結果と考察

電子楽器音に対して、本方式を用いて解析を行なった結果について報告する。

本章では、始めに本研究の実験条件を述べる。次に、採譜システムの入力として、単一楽器単音、単一楽器和音、異種複数楽器音の3種類を用いた場合の音高、音価推定結果について、それぞれ説明する。さらに、異種複数楽器音に対して本方式を用いて単一楽器単音成分への分離を行ない、楽器推定を行なった結果について説明する。

5.1 実験条件

5.1.1 使用する音源

本方式では、解析対象として電子楽器（MIDI：KORG05R/W）音を用いる。この音源は、パーソナルコンピュータのディスプレイ上に所望の楽譜を作成し、それに基づきシンセシス・システムからアンプ内臓のスピーカに音を増幅して出力するものである。この音源の特徴としては、楽器音の再現性が高いという点が挙げられる。

楽器音の音の構成要素には、音高、音量、音色の3種類がある。音高とはその楽器が演奏した音階の基本周波数に依存するものであり、前述のように、音名 $C, C\#, \dots, B$ 、及びそのオクターブにより決定される。また、音量とは楽器音の強度を意味するが、2.2.3 で

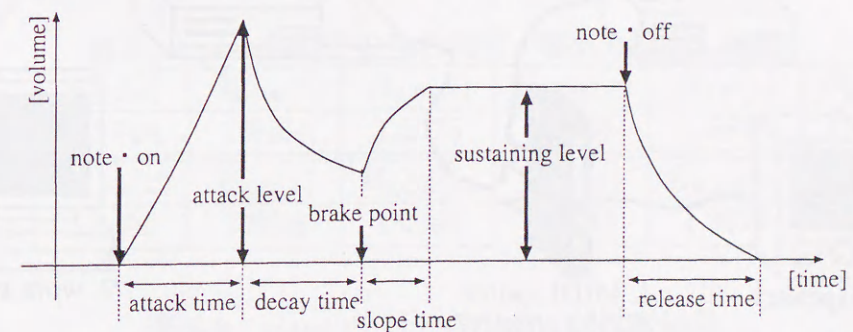


図 5.1 音量の構成パラメータ

述べたように、各楽器音が持つ音量の時間変化には固有の性質があり、それが各楽器音を区別する特徴の1つとなっている。この音源(MIDI: KORG05R/W)では、音量の構成要素として図5.1に示すような複数のパラメータを持ち、実際の楽器音の音量変化を忠実に再現している。また音色とは、空気の振動具合、つまり楽器音の1周期を切り出して観察した際の波の形により決定される。この波形は、楽器音に含まれる基本周波数成分の振幅と、各高調波成分の数及びその振幅(スペクトル構造)により決定されるため、楽器ごとに異なる特徴を持っている。この音源では、各楽器音の音色を周波数成分の割合ではなく、実際の楽器音をサンプリングした波形情報として保持しているため、周波数の変化(ビブラート、ビート)や図5.1におけるアタックレベルからサスティングレベルにいたる音色変化、音域による音色変化等を忠実に再現可能である[43]。

このような音源を用いて計算機シミュレーションを行なうことで、将来的に実楽器音の解析を目指すための、より実用的なシステムの構築に役立つものと考えられる。

5.1.2 自動採譜システムの概要

本研究のシステム構成を図5.2に示し、以下に各部所の説明を行なう。

1. 入力楽器音として、前述のMIDI音源(KORG:05R/W)を用いる。
2. 1.のMIDI音源は、パーソナルコンピュータで制御する。

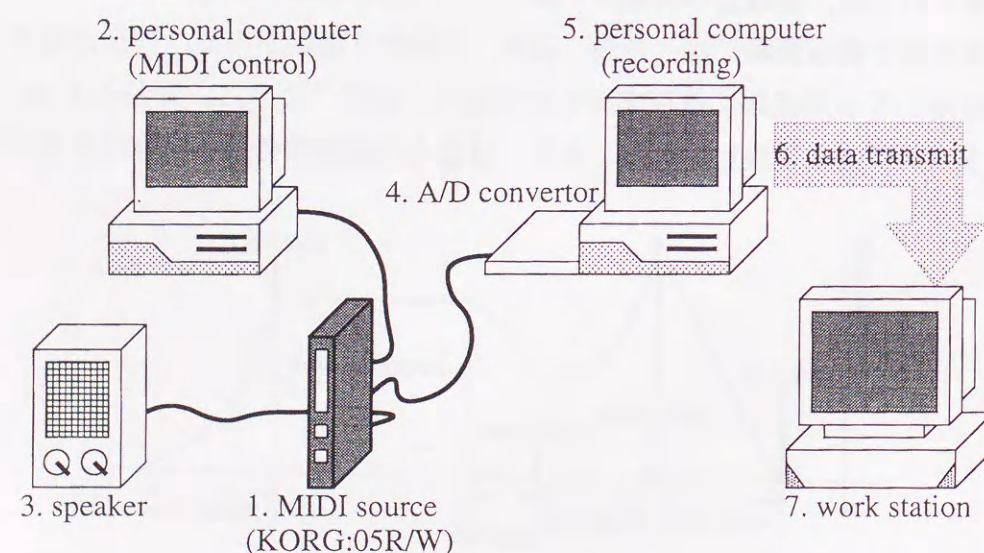


図5.2 システム概観図

3. MIDI音源からの楽器音は、MIDI付属のスピーカを通して確認する。
4. A/D変換ボード(UAD-ATKR)を使用して、アナログデータである楽器音をデジタルデータに変換する。
5. 4.のA/D変換ボードはパーソナルコンピュータ(FMV-DESK POWER SP)上のC言語プログラムで制御され、得られたデジタルデータはこのパーソナルコンピュータ上に格納される。なお、このA/Dボードは入力範囲として最大 $\pm 5[V_{p-p}]$ まで使用可能であり、最小サンプリング周期は $8\mu[s]$ であり、本方式で使用するオーバーサンプリング周波数は $54.054[kHz]$ である。
6. パーソナルコンピュータからワークステーション上へ、楽器音データを転送する。
7. くし形フィルタによる音高検出及び和音分離の具体的な処理は、使用メモリ等の都合により全てワークステーション上のC言語プログラムで行なう。

5.2 音高、音価推定結果

5.2.1 単一楽器音の単音に対する音高推定

piano C^3 音, violin D^4 音, flute F^5 音に対して、図3.4に示した単音に対する音名判別システムを用いて音高推定を行なった結果について説明する。

音名判別の結果として、表5.1に入力 $x(n)$ と各くし形フィルタの出力 $y_{s,p}(n)$ の振幅比 $(|y_{s,p}(n)|/|x(n)|)$ を計算した結果を示し、各くし形フィルタの出力例として、図5.3にpiano C^3 音、及び $H_{s,B}(z)$, $H_{s,C}(z)$, $H_{s,C\#}(z)$ の出力を示す。

表5.1及び図5.3の結果より、 $H_{s,C}(z)$ の出力が最も減衰(太字で記入)していることから、入力がC音であることが分かる。また、オクターブ推定の結果として、 $x(n)$ と $H_{s,C}(z)$,

表5.1 piano C^3 音の音名判別(零出力検出)結果

	$H_{s,C}(z)$	$H_{s,C\#}(z)$	$H_{s,D}(z)$	$H_{s,D\#}(z)$	$H_{s,E}(z)$	$H_{s,F}(z)$
$y_{s,p}/x$	0.3748	1.1484	1.5575	1.7682	1.5828	1.6178
	$H_{s,F\#}(z)$	$H_{s,G}(z)$	$H_{s,G\#}(z)$	$H_{s,A}(z)$	$H_{s,A\#}(z)$	$H_{s,B}(z)$
$y_{s,p}/x$	1.6100	1.5908	1.6095	1.4193	1.3679	1.3787

表5.2 piano C^3 音のオクターブ推定結果

	$H_{s,C}(z)$	$H_{A,C}(z)$	$H_{2,C}(z)$
$y_{r,p}/x$	0.3748	1.4089	1.6345

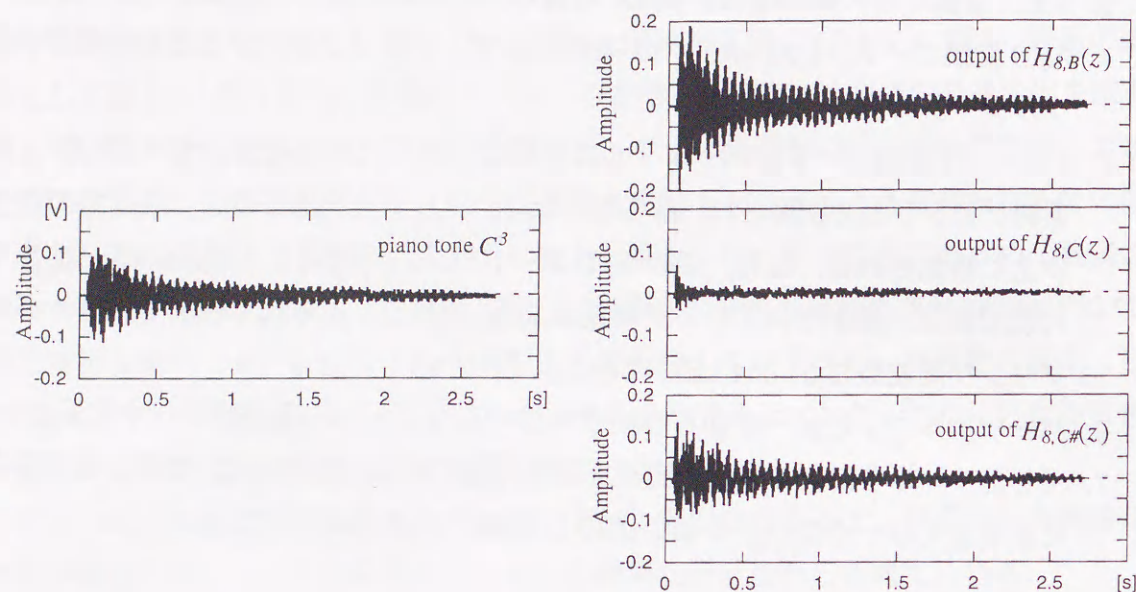


図 5.3 piano C^3 音の音名判別（零出力検出）結果

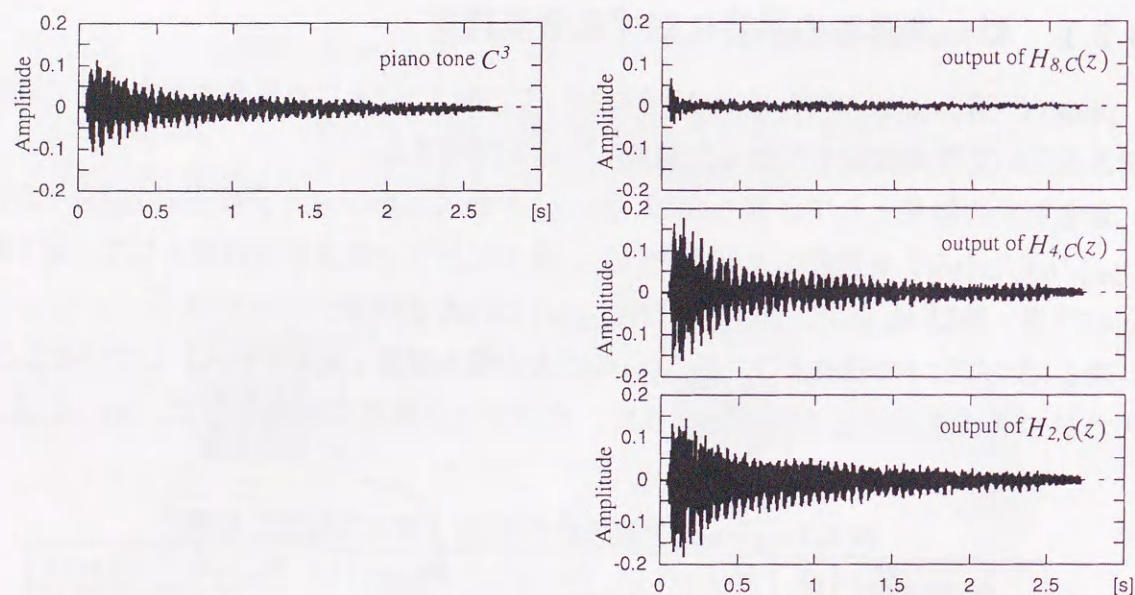


図 5.4 piano C^3 音のオクターブ推定結果

$H_{4,C}(z)$, $H_{2,C}(z)$ の出力 $y_{r,P}(n)$ の振幅比 ($|y_{r,P}(n)|/|x(n)|$) を計算した結果を表 5.2 に示し、各フィルタの出力波形を図 5.4 に示す。図 5.4 及び表 5.2 の結果より、4 次の $H_{4,P}(z)$ の出力が $H_{8,P}(z)$ の出力と比較してかなり大きい（太字で記入）ことから非零出力が検出されていると判断でき、入力音はオクターブ 3 であることが判別できる。

表 5.3 violin D^4 音の音名判別（零出力検出）結果

	$H_{8,C}(z)$	$H_{8,C\#}(z)$	$H_{8,D}(z)$	$H_{8,D\#}(z)$	$H_{8,E}(z)$	$H_{8,F}(z)$
$y_{8,P}/x$	1.7176	1.3659	0.2460	1.2836	1.5364	1.6917
	$H_{8,F\#}(z)$	$H_{8,G}(z)$	$H_{8,G\#}(z)$	$H_{8,A}(z)$	$H_{8,A\#}(z)$	$H_{8,B}(z)$
$y_{8,P}/x$	1.8112	1.8623	1.8415	1.6925	1.7395	1.3642

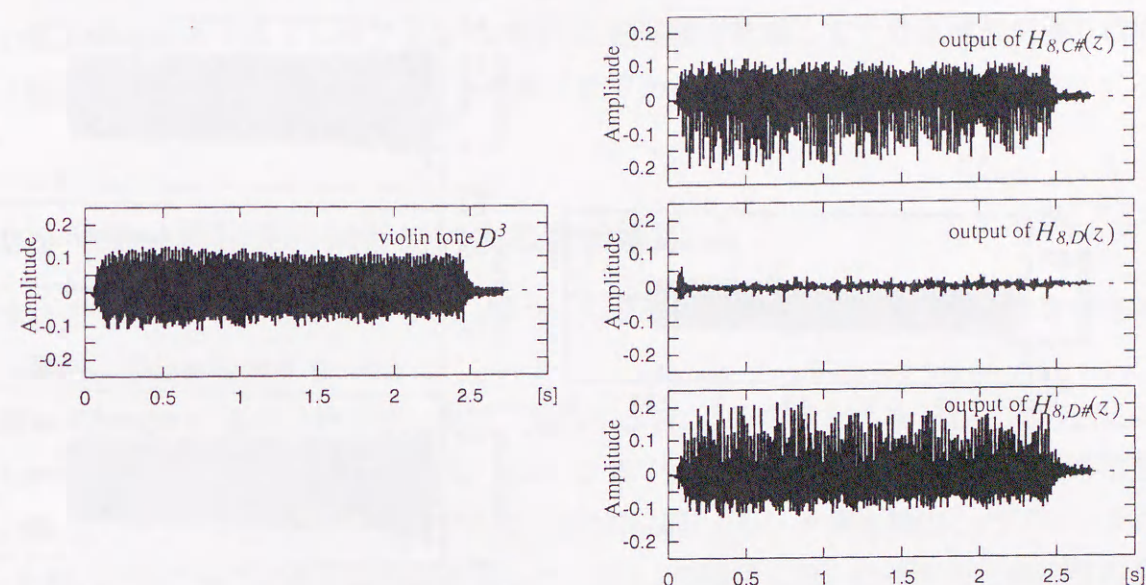


図 5.5 violin D^4 音の音名判別（零出力検出）結果

表 5.4 violin D^4 音のオクターブ推定結果

	$H_{8,D}(z)$	$H_{4,D}(z)$	$H_{2,D}(z)$
$y_{r,P}/x$	0.2460	0.1549	1.8556

同様に、violin D^4 , flute F^5 に対する 12 種類の $H_{8,P}(z)$ の入出力振幅比を表 5.3, 表 5.5 に、また、各入力に対するくし形フィルタの出力波形を図 5.5, 図 5.6 にそれぞれ示す（なお、零出力検出結果については、両隣の音に対応する $H_{8,P}(z)$ の出力のみを示した）。さらに、piano C^3 音の場合と同様に、フィルタ次数を変化させることでオクターブ推定を行った結果を表 5.4, 5.6 に示す。なお、前述と同様に、音名判別における零出力、及び、オクターブ検出における非零出力の値をそれぞれ太字で示した。

これらの結果より、演奏楽器の種類に関わらず、 $H_{8,P}(z)$ の零出力を検出することで入力音名を判別できることが確認できる。なお、表 5.5 に示す flute F^5 音の音名判別結果では、 F

表 5.5 flute F^5 音の音名判別（零出力検出）結果

	$H_{8,C}(z)$	$H_{8,C\#}(z)$	$H_{8,D}(z)$	$H_{8,D\#}(z)$	$H_{8,E}(z)$	$H_{8,F}(z)$
$y_{8,P}/x$	1.4874	0.5410	1.7509	1.7224	1.7950	0.1279
	$H_{8,F\#}(z)$	$H_{8,G}(z)$	$H_{8,G\#}(z)$	$H_{8,A}(z)$	$H_{8,A\#}(z)$	$H_{8,B}(z)$
$y_{8,P}/x$	1.8054	1.6457	1.4937	1.7270	0.1693	1.7520

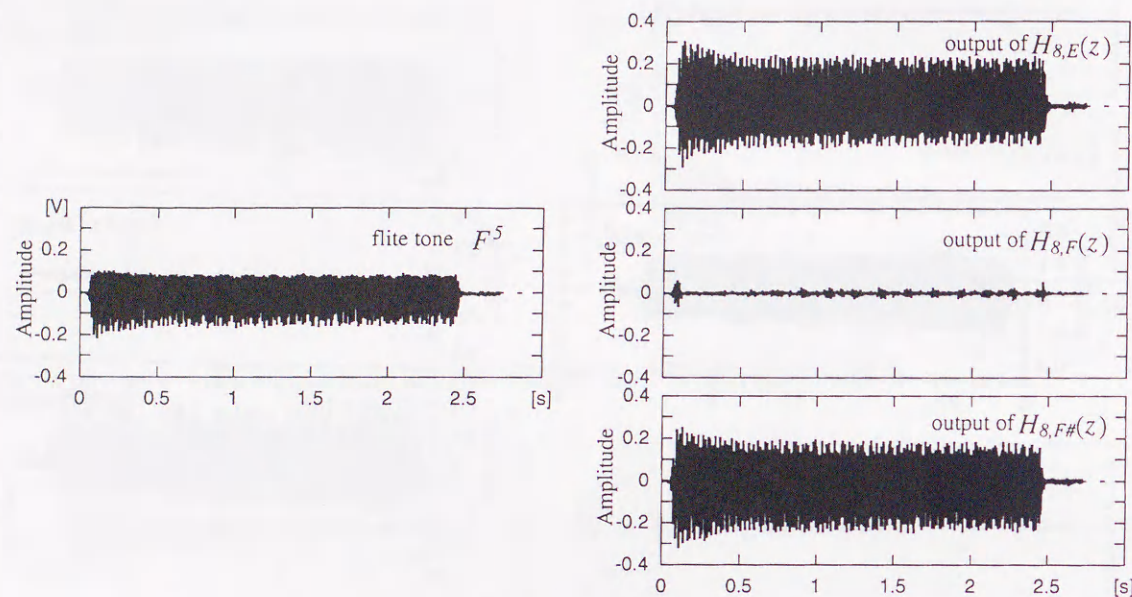


図 5.6 flute F^5 音の音名判別（零出力検出）結果

表 5.6 flute F^5 音のオクターブ推定結果

	$H_{8,F}(z)$	$H_{4,F}(z)$	$H_{2,F}(z)$
$y_{r,P}/x$	0.1279	0.0986	0.0867

に対応する $H_{8,F}(z)$ 以外にも、 $H_{8,C\#}(z)$ 、 $H_{8,A\#}(z)$ の出力に減衰が見られる。これは、前述の周波数成分の重複関係より、 $f_{5,F} \cong 5.0 \times f_{3,C\#}$ 、 $f_{5,F} \cong 3.0 \times f_{3,A\#}$ の関係が成り立つためである。しかし、このような減衰を考慮しても、単音入力に対する音名判別システムにおいては、入力に対応した $H_{r,P}(z)$ の出力が最も減衰しており、他音に対応する $H_{8,P}(z)$ の出力振幅と比較することで十分音名判別可能であることを確認した。

なお、図 5.3、5.5、5.6 における入力音 P に対応する $H_{8,P}(z)$ の出力は完全に零ではなく、雑音（除去しきれない）成分が含まれていることが分かる。これは、楽器音の非定常性の影響と考えられる。特に、piano 音はハンマーで弦を叩くことで発音するという楽器の

構造上、図 5.1 の音量の構成要素におけるアタックレベルからブレイクポイントまでの減衰率が大きく、また、その部分に多くの非調和周波数成分が含まれていると考えられる。また、violin 音については、音を継続させる際のビブラートの影響から、波形の後半部分には周波数（特に高次の高調波）成分に周波数変動が現れることが分かっている。このように $H_{r,P}(z)$ では除去しきれない非調和周波数成分が、零出力の誤差として生じるためと考えられる。しかし、このような誤差を考慮しても、入力音名に対応した $H_{r,P}(z)$ の出力には、他の音に対応するくし形フィルタの出力の振幅値と比較して十分な減衰が得られており、閾値を設けることにより十分音名判別可能である。

5.2.2 単一楽器音の単音に対する音価推定

本方式では、入力音に対応するくし形フィルタの零出力を検出することで、音の継続時間、即ち、音価を推定する。

図 5.7 に flute F^5 音の信号波形、及び、対応する $H_{8,F}(z)$ の零出力を示し、本方式により得られた F 音の音価推定結果を示す。なお、3 章で述べたように音価検出のための時間軸上の最小ブロック単位は 1351 サンプル点（約 25[ms]）とし、 F 音を検出したブロックには 1.0 を出力することで音高継続時間を示した。図 5.7 の結果より、ほぼ入力と同程度の長さの音価が正しく検出されていることが分かる。

同様に、図 5.8 に violin D^4 音、図 5.9 に piano C^3 音に対して音価推定を行なった結果を示す。図 5.8、5.9 の結果より、violin D^4 音、piano C^3 音ともに flute F^5 音の場合と同様に、入力

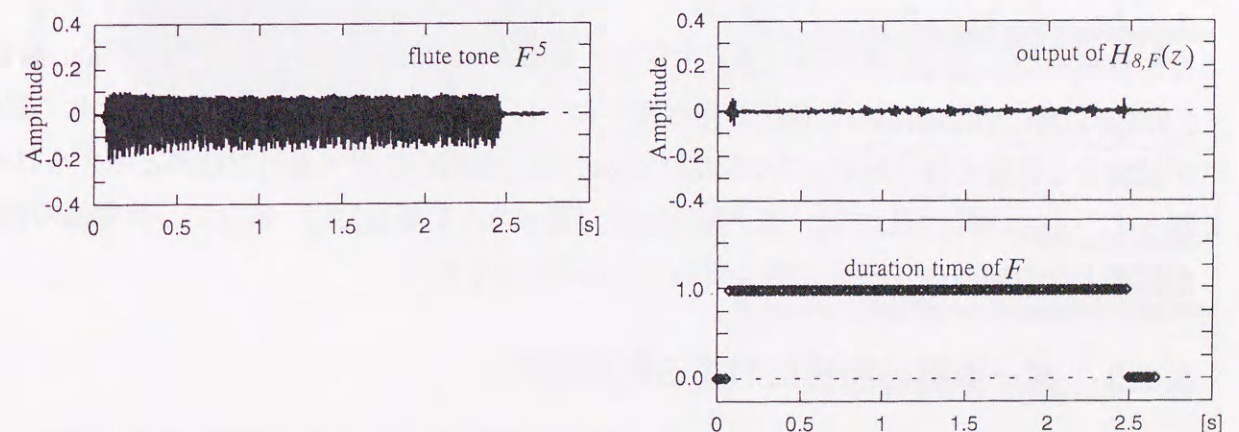


図 5.7 flute F^5 の音価（音の継続時間）推定結果

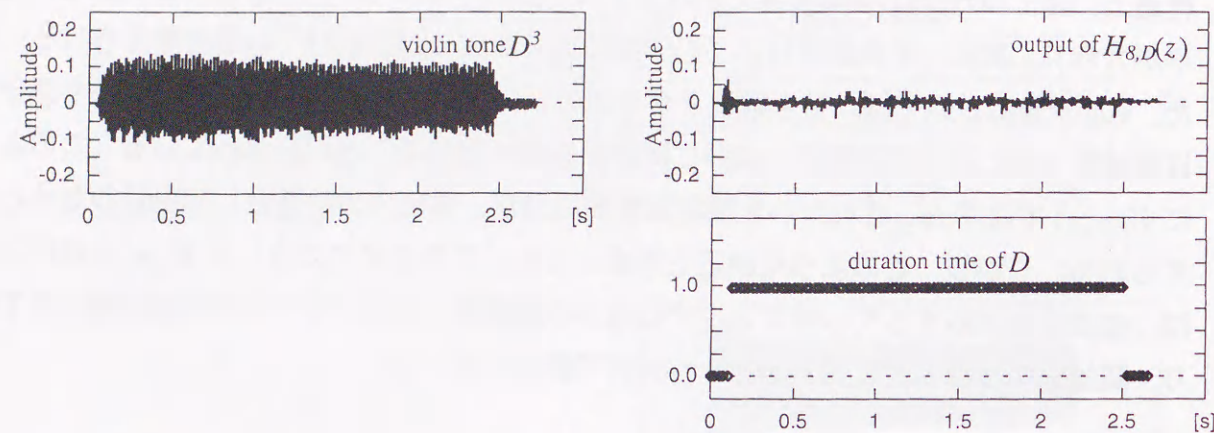


図 5.8 violin D^4 の音価（音の継続時間）推定結果

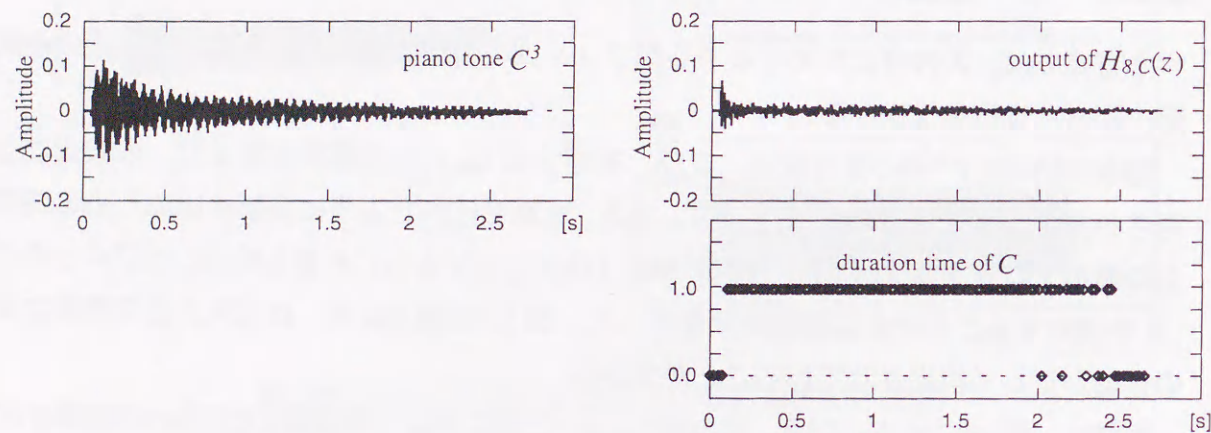


図 5.9 piano C^3 の音価（音の継続時間）推定結果

音の継続時間がほぼ正確に検出されていることが分かる。但し、図 5.9 に示す piano 音は、2.025[s] 以降で数箇所誤った検出結果が得られている。これは、piano 音が時間と共に振幅が減衰する性質を持つため、この区間での入力音の振幅が小さくなったためと考えられる。従って、piano 音に対しては、音の継続時間が長くなった場合には、零ないし非零値の判別が困難となり実際の音価より短く検出される可能性がある。

5.2.3 単一楽器の和音に対する音高推定

本節では flute, violin, piano による単一楽器和音に対して、3.2.3 で示した音名判別システムを用いて音高（音名及びオクターブ）推定を行なった結果について以下に説明する。

まず、入力として violin F^4+G^4 の和音を用いた場合の音名判別結果を図 5.10 に示し、こ

の結果について考察する。ここで、 G 音は C 音と周波数の重複関係を持ち、 $H_{16,C}(z)$ で除去される性質を持つ。なお、図 5.10 に以下の (1) から (6) までの波形を示した。

(1) violin F^4+G^4 の入力波形。

(2) F 音の検出前の状態 ($y_{16,E}(n) \gg 0$)。

フィルタ接続状態: $x(n) \rightarrow H_{16,C}(z) H_{16,C\#}(z) \cdots H_{16,E}(z) \rightarrow$ 出力

なお、この段階までは零出力は一度も検出されていない。

(3) F 音の検出結果 ($y_{8,F}(n) \cong 0$)。

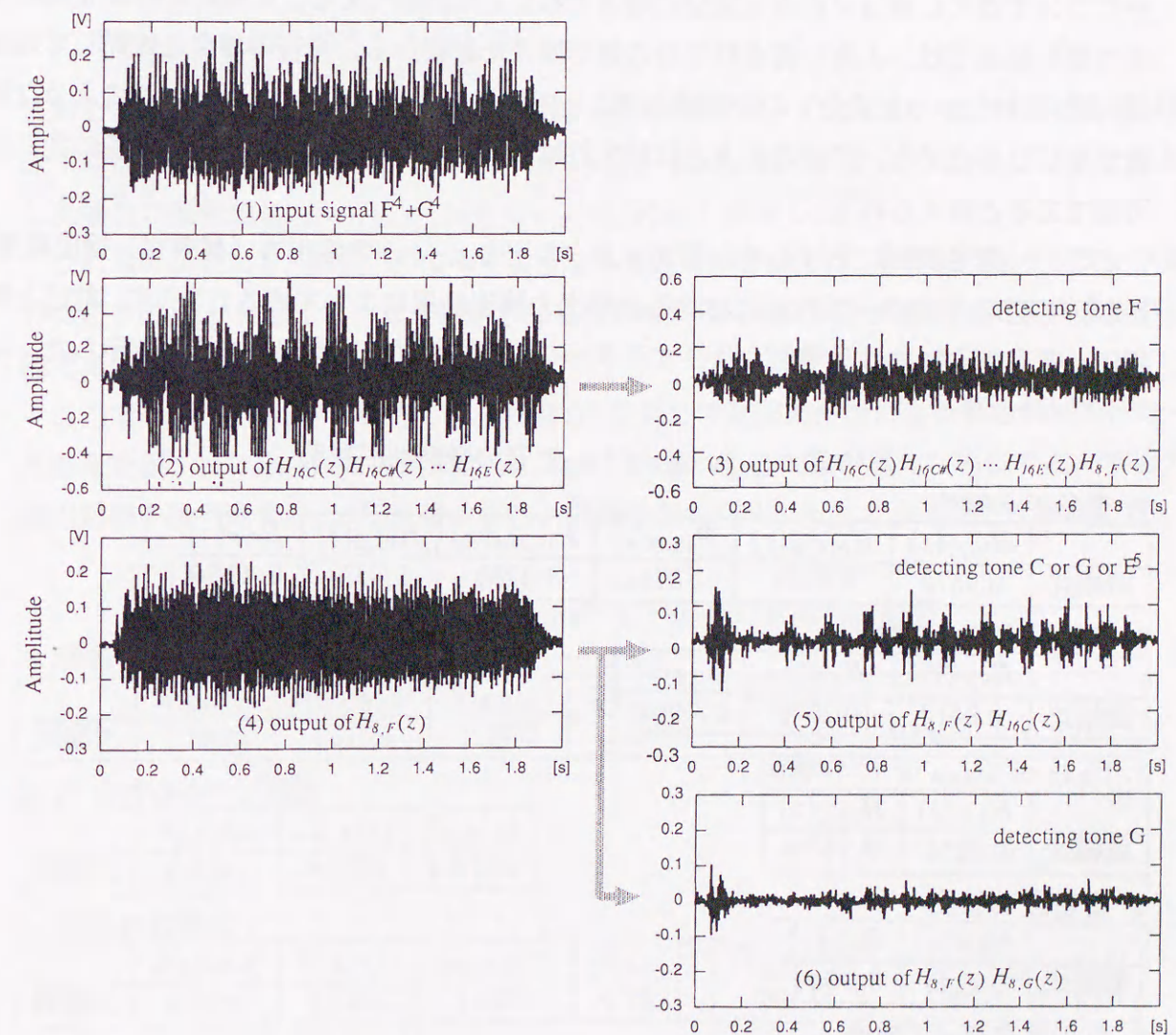


図 5.10 violin F^4+G^4 に対する音名判別（零出力）検出結果

フィルタ接続状態： $x(n) \rightarrow H_{16,C}(z) H_{16,C\#}(z) \cdots H_{8,F}(z) \rightarrow$ 出力
 (4) F^4 音を含む和音であることを判別 (only $y_{8,F}(n) \gg 0$) .

フィルタ接続状態： $x(n) \rightarrow H_{8,F}(z) \rightarrow$ 出力
 (5) $C^3 \sim C^5$, $G^3 \sim G^5$, E^5 のいずれかを含むことを判別 ($y_{16,C}(n) \cong 0$) .

フィルタ接続状態： $x(n) \rightarrow H_{8,F}(z) H_{16,C}(z) \rightarrow$ 出力
 (6) G 音の検出結果 ($y_{8,G}(n) \cong 0$) .

フィルタ接続状態： $x(n) \rightarrow H_{8,F}(z) H_{8,G}(z) \rightarrow$ 出力

前述のように, violin 音は周波数変動や振幅変動等を含んでおり, このような非定常成分についてはくし形フィルタで完全に除去することは困難である. このため, 図 5.10 の零出力検出結果では, 入力に含まれる音の数が増えた影響から, 単音の場合と比較して雑音 (除去しきれなかった成分) の振幅値が増大していることが分かる. しかし, このような雑音が生じた場合でも, 前段の出力と比較すれば十分な減衰が得られており, 音名の判別は可能であると考えられる.

ここで, 音名判別におけるくし形フィルタの出力 $y_{r,p}(n)$ の零出力 (減衰率) 検出結果, 及び, 各音のオクターブ判別における非零出力検出結果について表 5.7 に示す. 前述と同様に, 音名判別における零出力及びオクターブ判別における非零出力を太字で示した. な

表 5.7 violin $F^4 + G^4$ 音の音高判別結果

1. 音名 F の検出

	$H_{16,C}(z)$	$H_{16,C\#}(z)$	$H_{16,D}(z)$	$H_{16,D\#}(z)$	$H_{16,E}(z)$	$H_{8,F}(z)$
振幅比	0.9573	0.9107	1.3714	1.3289	1.6073	0.4180

2. F のオクターブ検出

	$H_{8,F}(z)$	$H_{4,F}(z)$	$H_{2,F}(z)$
振幅比	0.3514	0.2579	0.7966

3. $C_{3,4,5}$ or $G_{3,4,5}$ or E^5 の検出

	$H_{8,F}(z)$	$H_{16,C}(z)$
振幅比	0.9255	0.2579

4. 音名 G の検出

	$H_{8,F}(z)$	$H_{8,G}(z)$
振幅比	0.7143	0.2047

5. G のオクターブ検出

	$H_{8,G}(z)$	$H_{4,G}(z)$	$H_{2,G}(z)$
振幅比	0.2047	0.1946	0.7649

お, 前述の単音に対する音名判別システムにおいては, 零出力検出のために入出力振幅比 $|y_{r,p}(n)/x(n)|$ を用いていたが, 和音に対する音名判別システムでは, その構成上, 前段の出力 $y_{r,p-1}(n)$ との振幅比 $|y_{r,p}(n)/y_{r,p-1}(n)|$ を利用している.

表 5.7 の結果より, 縦続接続した $H_{r,p}(z)$ の零出力 (出力の大幅な減衰) を検出することで, 入力音名の判別が可能であることを確認した. また, 単音の場合と同様に, $H_{r,p}(z)$ の次数を変化させ, 非零出力を検出することで, オクターブの判別が可能であることを確認した.

同様な方式で, 単一楽器による 3 種類の和音

(a) violin $F^4 + F\#^4$

(b) piano $C^3 + F^4$

(c) flute $C^4 + E^4 + G^4$

に対して, 本方式を用いて音高推定を行なった結果について, 表 5.8 ~ 5.10 に示す. なお, 表中には, 前述と同様に, 縦続接続した $H_{8,p}(z)$, 及び, 前段の出力 (ないし入力) と比較した場合の振幅比 $|y_{r,p}(n)/y_{r,p-1}(n)|$ ($|y_{r,p}(n)/x(n)|$) を示した.

これらの結果より, 和音 (a) に示すような半音違い和音, (b) に示すようなオクターブ違いの和音の, (c) に示すような 3 音からなる和音に対しても, 本方式を用いて音高推定が可能であることを確認した.

ここで, 3 和音の場合の例として, 和音 (c) に対して零出力検出による音名判別を行なった結果を図 5.11 に示す. この結果より, 和音に含まれる音の数が増えた場合でも, 零出力検出処理を繰り返すことで, 2 音の場合と同様に音名判別が可能であることを確認した.

表 5.8 和音 (a) violin $F^4 + F\#^4$ 音の音高判別結果

1. 音名 $F\#$ の検出

	$H_{16,C}(z)$	$H_{16,C\#}(z)$	$H_{16,D}(z)$	$H_{16,D\#}(z)$	$H_{16,E}(z)$	$H_{8,F}(z)$	$H_{8,F\#}(z)$
振幅比	1.0829	1.1029	1.4982	1.0933	0.9434	0.9814	0.3256

2. F^4 のオクターブ検出

	$H_{8,F\#}(z)$	$H_{4,F}(z)$	$H_{2,F}(z)$
振幅比	0.2341	0.3958	1.0104

3. 音名 F の検出

	$H_{8,F\#}(z)$	$H_{16,C}(z)$	$H_{16,C\#}(z)$	$H_{16,D}(z)$	$H_{16,D\#}(z)$	$H_{16,E}(z)$	$H_{8,F}(z)$
振幅比	0.8929	1.0193	1.2932	1.2573	0.8434	0.9002	0.2416

4. F のオクターブ検出

	$H_{8,F}(z)$	$H_{4,F}(z)$	$H_{2,F}(z)$
振幅比	0.1968	0.2094	0.9989

表 5.9 和音 (b) piano $C^3 + F^4$ 音の音高判別結果

1. 音名 F の検出

	$H_{16,C}(z)$	$H_{16,C\#}(z)$	$H_{16,D}(z)$	$H_{16,D\#}(z)$	$H_{16,E}(z)$	$H_{8,F}(z)$
振幅比	0.9785	1.2255	1.1546	1.0640	1.0545	0.4876

2. F 音のオクターブ検出

	$H_{8,F}(z)$	$H_{4,F}(z)$	$H_{2,F}(z)$
振幅比	0.1315	0.2458	0.6654

3. $C_{3,4,5}$ or $G_{3,4,5}$ or E^5 の検出

	$H_{8,F}(z)$	$H_{16,C}(z)$
振幅比	0.9574	0.2414

4. C^3 音の検出

	$H_{8,F}(z)$	$H_{8,C}(z)$
振幅比	0.9574	0.2347

表 5.10 和音 (c) flute $C^4 + E^4 + G^4$ 音の音高判別結果

1. $E^{4,5}$ (or B 音) の検出

	$H_{16,C}(z)$	$H_{16,C\#}(z)$	$H_{16,D}(z)$	$H_{16,D\#}(z)$	$H_{16,E}(z)$
振幅比	0.7758	1.0246	1.5428	1.0640	0.3584

2. E^4 の検出

	$H_{16,C}(z)$	$H_{16,C\#}(z)$	$H_{16,D}(z)$	$H_{16,D\#}(z)$	$H_{4,E}(z)$
振幅比	0.7758	1.0246	1.5428	1.0640	0.4483

3. C or G or E^5 の検出

	$H_{4,E}(z)$	$H_{16,C}(z)$
振幅比	1.0025	0.3457

4. C^4 音の検出

	$H_{4,E}(z)$	$H_{8,G}(z)$	$H_{8,G}(z)$	$H_{4,C}(z)$
振幅比	0.9574	0.8726	0.9884	0.2547

5. G^4 音の検出

	$H_{8,G}(z)$	$H_{4,G}(z)$	$H_{2,G}(z)$
振幅比	0.2518	0.3365	1.0957

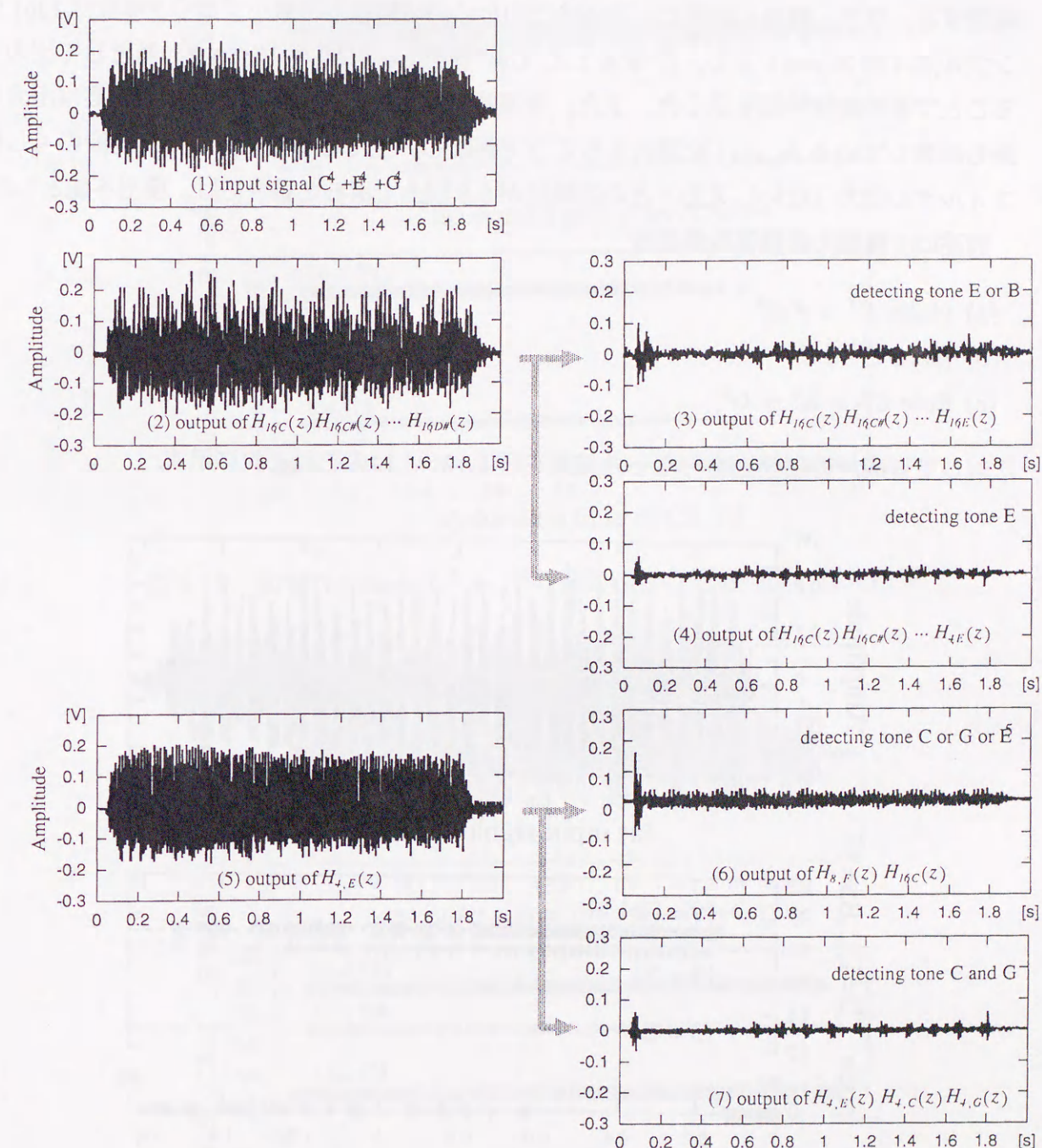


図 5.11 flute $C^4 + E^4 + G^4$ に対する音名判別 (零出力) 検出結果

5.2.4 単一楽器の和音に対する音高、音価推定

本方式を用いて単一楽器による和音に対する音価推定を行なった結果について、以下に報告する。なお、前述と同様に、音価検出のための時間軸上の最小ブロック単位は1351 サンプル点（約25[ms]）とし、 C^3 音を1.0、 $C\sharp^3$ 音を2.0、 \dots 、 B^5 を36.0の値として出力することで音高継続時間を示した。また、音高推定アルゴリズムに従い、入力音高は出力が最も減衰している $H_{r,p}(z)$ を選択することで判断し、全ての音高に対して、前段のくし形フィルタの出力（ないし入力）との振幅比が0.6以上であった場合には、検出不能とした。

前述の3種類の異種複数楽器音

- (a) violin $F^4 + F\sharp^4$
- (b) piano $C^3 + F^4$
- (c) flute $C^4 + E^4 + G^4$

に対して音高、音価検出を行なった結果を図5.12～5.14のそれぞれに示す。

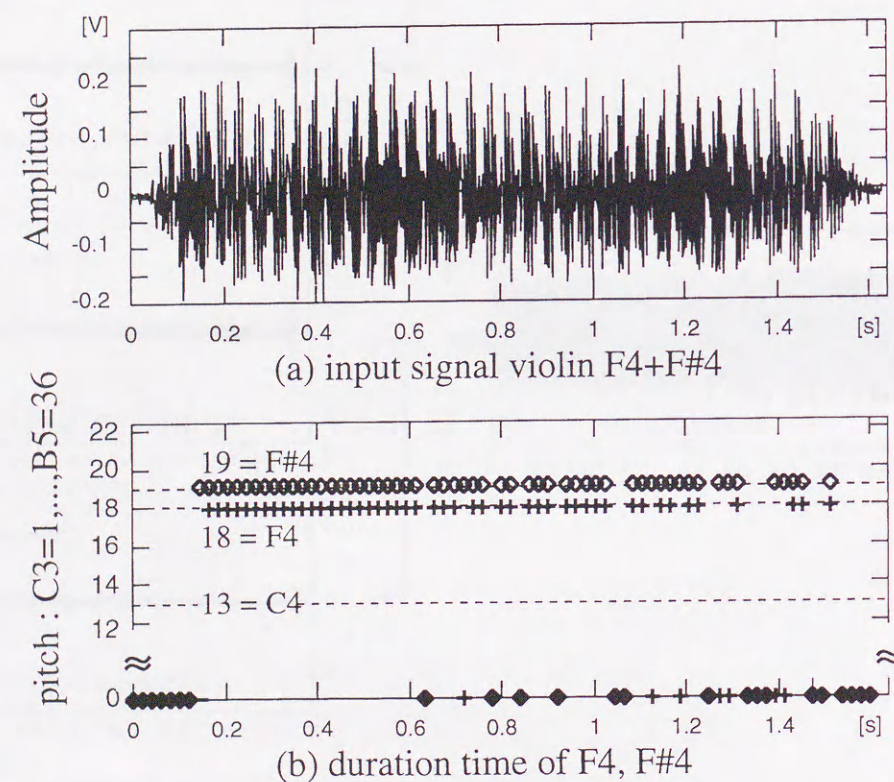


図5.12 和音(a) violin $F^4 + F\sharp^4$ の音価（音の継続時間）推定結果

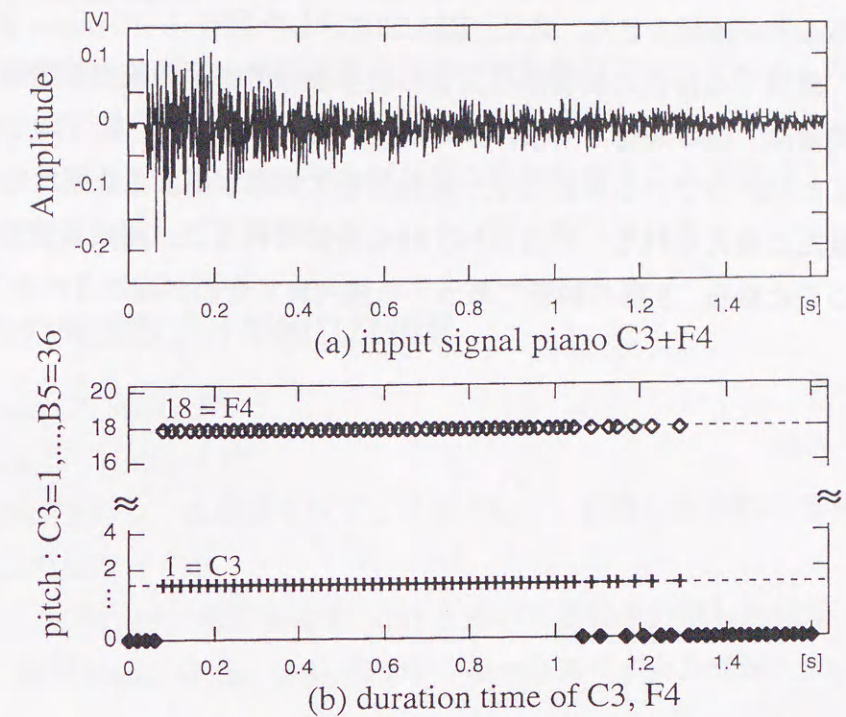


図5.13 和音(b) piano $C^3 + F^4$ の音価（音の継続時間）推定結果

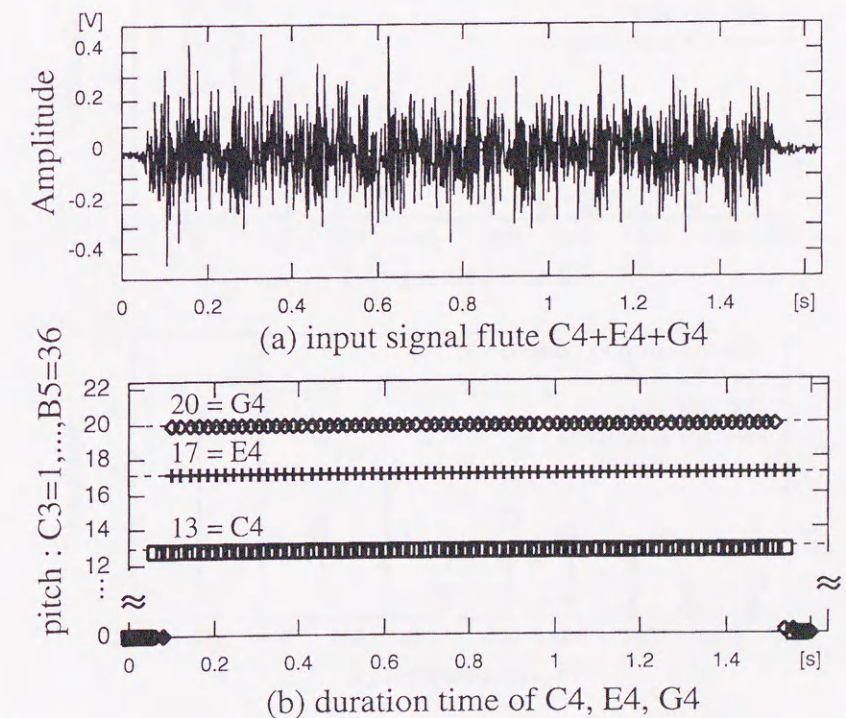


図5.14 和音(c) flute $C^4 + E^4 + G^4$ の音価（音の継続時間）推定結果

図5.12より、和音(a)のviolin音については、前述の周波数変動の影響から、若干の測定不能点があることが確認できた。また、図5.13に示した和音(b)のpiano音については、前述と同様に、減衰する特性の影響から、音の後半部分の音高の検出が困難であることが分かる。この問題は、他の楽器と同様の音高の判別条件（閾値等）を用いているためであり、入力がpianoと分かっている場合には、減衰信号に対応するような新たな条件を設けることで改善可能だと考えられる。和音(c)のflute音については、周波数変動が少なく、振幅も定常であることから、3音の和音であっても精度良く音価が推定されることを確認した。

5.3 楽器推定結果

本節では、異種複数楽器音の楽器推定を行なった結果について述べる。なお、ここで解析を行なう和音は、前章においてパターン分けした4種類のうち、ユニゾン和音を除く3種類とする。なお、今回は本方式のアルゴリズムを検証することを目的とし、和音に含まれる音の数を2音に制限して計算機シミュレーションを行なった。

5.3.1 2音の周波数成分が独立な和音

和音(1) piano C^3 , violin F^3

和音(2) horn D^4 , trumpet G^4

に対して楽器推定を行なった結果を以下に示す。なお、和音に含まれる各音高は、前述の音高判別法により既知とする。

和音(1)から、 C^3 と F^3 に対応する $H_{r,p}(z)$ を用いて各音を分離した結果($|H_{r,p}(e^{j\omega})|$ による補正後)、及びpiano, violin, violaのテンプレートスペクトルを図5.15に示す。図5.15

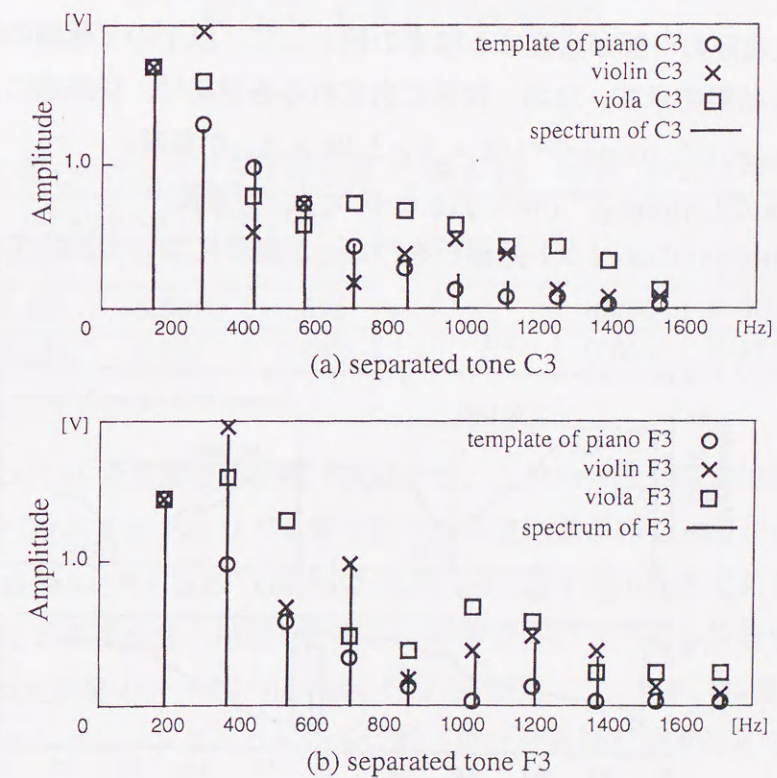


図5.15 和音(1)の分離結果

表 5.11 和音 (1) のスペクトル差分値

	instrument of template		
tone	piano	viola	violin
C^3	0.026	0.217	0.312
F^3	0.442	0.392	0.037

表 5.12 和音 (2) のスペクトル差分値

	instrument of template		
tone	horn	trumpet	piano
D^4	0.052	0.471	0.421
G^4	0.683	0.077	0.469

より, C^3 と F^3 に対応するテンプレート (piano, violin) との差が最も小さく, 楽器判別が可能であることが分かる. なお, 計算機で楽器の自動判別を行なうために, 和音 (1), (2) から分離した各音のスペクトルとテンプレートスペクトルとの差分の絶対値和を計算した結果を表 5.11, 5.12に示す. ここで, 差分には基本周波数成分から 9 倍音までのスペクトルを用いた. 表 5.11, 5.12の結果より, 演奏楽器に対応するテンプレートスペクトルとの差分値が最も小さく (表中に太字で記入), 楽器推定が可能であることを確認した.

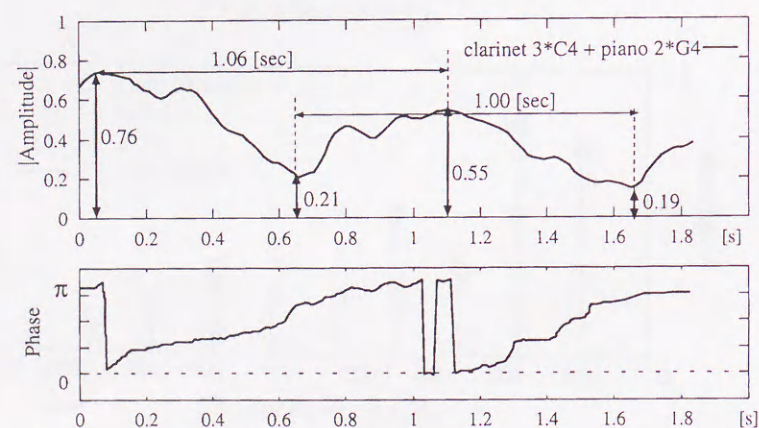
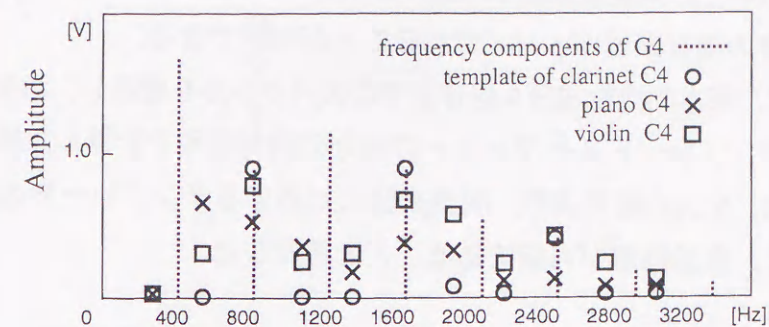
5.3.2 2 音の高調波の一部が重複する和音

2 音の高調波成分の一部が重複する和音の例として, 以下の 2 種類の和音に対して楽器推定を行なった結果を示す. なお, 和音に含まれる各音高は, 従来法により既知とする.

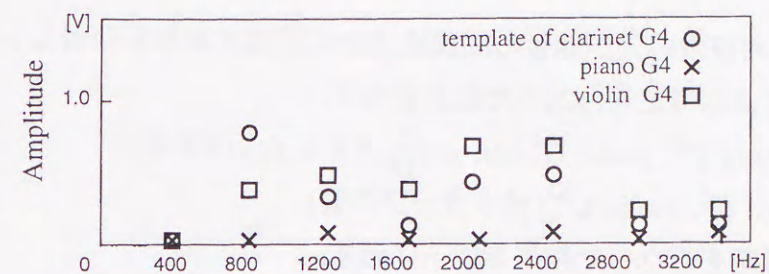
和音 (3) clarinet C^4 , piano G^4 ($3k \times f_{4,C}$ と $2k \times f_{4,G}$ が重複)

和音 (4) viola C^3 , violin E^3 ($5k \times f_{3,C}$ と $4k \times f_{3,E}$ が重複)

例として, 和音 (3) のスペクトル解析を行なった結果について説明する. ここで, C^4 の

図 5.16 重複成分 $3k \times f_{4,C} + 2k \times f_{4,G}$ の振幅変化 (STFT 結果), 及び位相変化

(a) polyphony - template of clarinet C4 = G4



(b) G4 - template of piano G4 = all 0

図 5.17 和音 (3) の分離結果

表 5.13 和音 (3) のスペクトル差分値

	instrument of template		
tone	piano	clarinet	violin
C^4	0.460	0.096	0.502
G^4	0.101	0.532	0.738

表 5.14 和音 (4) のスペクトル差分値

	instrument of template		
tone	piano	violin	viola
C^3	0.298	0.083	0.430
E^3	0.611	0.273	0.022

$3k$ 次高調波と G^4 の $2k$ 次高調波は重複 (近接) し, これらの合成成分は図 5.16に示すような振幅変動, 及び位相変動を持つ. 4 章で述べた手法に基づいて, 図 5.16の振幅変動の最大値より, 重複成分 $3nC^4 + 2nG^4$ の振幅は $(0.76 + 0.55) \div 2 = 0.655$ であることが判別できる. 同様に, 全ての重複成分の振幅を計算し, 和音全体のスペクトルを求める.

さらに, 得られた和音のスペクトル分離を行なう (図 5.17). 先ず, 和音中の C^4 音 ($3nC^4$ 以外) の成分とテンプレートスペクトルを比較 (差分を計算) することで, C^4 音の演奏楽器が clarinet であることが分かる. 次に, 和音の複素スペクトルから clarinet のテンプレート (複素) スペクトルを減算し, 残された G^4 音の成分に対しても同様な比較を行なう. こ

の結果、 G^4 音の全ての周波数成分が piano G^4 のテンプレートとほぼ一致し、差分値が零となるため、 G^4 音の演奏楽器が piano であることが判別できる。

前節と同様に、基本周波数成分9倍音までのスペクトルを使用し、和音(3)、(4)の各音のスペクトルとテンプレートスペクトルとの差分の絶対値和を計算した結果を表5.13、5.14に示す。表5.13、5.14の結果より、演奏楽器に対応するテンプレートスペクトルとの差分値が最も小さく、楽器推定が可能であることを確認した。

5.3.3 1 音の基本周波数が他音の高調波に重複する和音

2 1 音の基本周波数成分が他音の高調波成分に重複する和音の例として、以下の2種類の和音に対して楽器推定を行なった結果を示す。

和音(5) clarinet C^3 , piano G^4 ($3k \times f_{3,C}$ と $k \times f_{4,G}$ が重複)

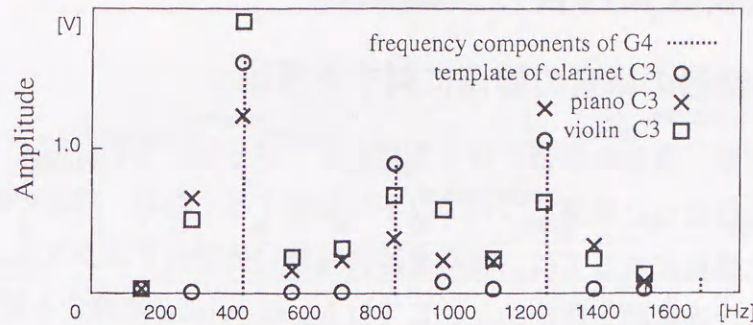
和音(6) piano C^3 , violin C^4 (オクターブ和音)

に対して楽器推定を行なった結果を以下に示す。

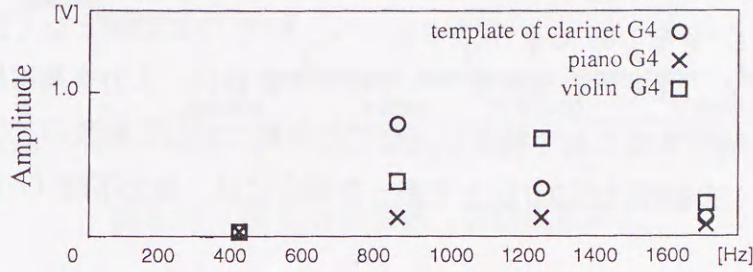
なお、この段階では、和音中の高い音は検出されていないため、和音全体のスペクトルの振幅変化をある程度の時間観察し、増減を繰り返すならば協和音の可能性があり、ほぼ定常な振幅が得られればオクターブ和音の可能性があると判断する。振幅の増減を繰り返すような周波数成分の振幅については、前述の手法で振幅値を求める必要がある。このようにして求めた和音(5)のスペクトルに対し、楽器推定、和音の分離を行なう。

まず、図5.18に示すように、入力音のスペクトルとテンプレートスペクトルを比較することで、 C^3 音の演奏楽器が clarinet であることを判別する。また、この比較において、 C^3 音の $3n$ 倍音成分が clarinet のテンプレートスペクトルと一致しないため、入力音に G^4 音が含まれることが分かる。次に、入力の実素スペクトルから clarinet のテンプレート(複素)スペクトルを減算し、残された G^4 音の成分に対しても同様な比較を行なう。この結果、 G^4 音の全ての周波数成分が piano G^4 のテンプレートとほぼ一致するため、 G^4 音の演奏楽器が piano であることが判別できる。

前節と同様に、基本周波数成分9倍音までのスペクトルを使用し、和音(5)、(6)の各音のスペクトルとテンプレートスペクトルとの差分の絶対値和を計算した結果を表5.15、5.16に示す。表5.15、5.16の結果より、演奏楽器に対応するテンプレートスペクトルとの差分値が最も小さく、楽器推定が可能であることを確認した。



(a) polyphony - template of clarinet C3 = G^4



(b) G^4 - template of piano G^4 = all 0

図5.18 (5)の分離結果

表5.15 和音(5)のスペクトル差分値

tone	instrument of template		
	piano	clarinet	violin
C^3	0.228	0.082	0.261
G^4	0.131	0.406	0.517

表5.16 和音(6)のスペクトル差分値

tone	instrument of template		
	piano	clarinet	violin
C^3	0.041	0.259	0.288
C^4	0.419	0.226	0.060

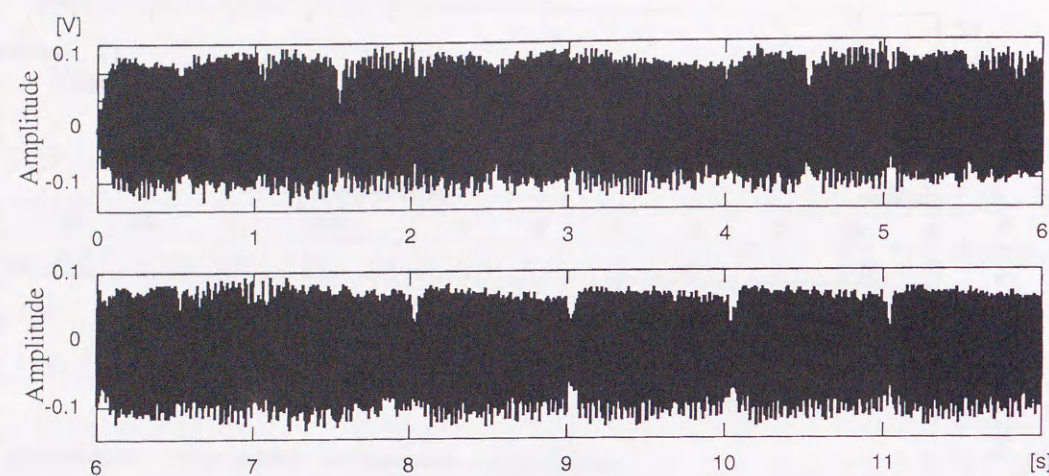


図 5.22 「カエルの歌」を violin で演奏した場合の出力波形

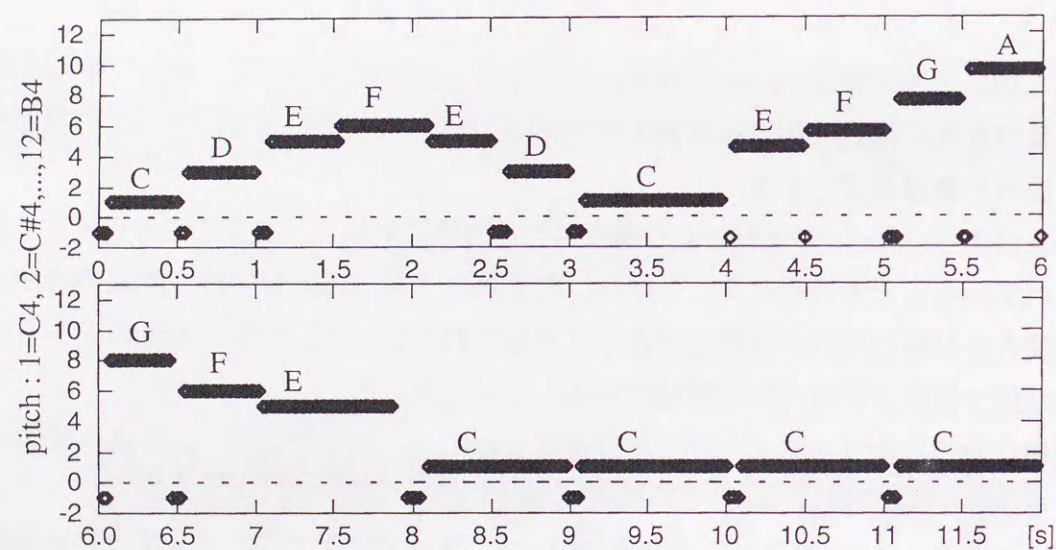


図 5.23 図 5.22 の音高、音価推定結果

5.4.2 単一楽器の和音に対する採譜

本システムの単一楽器の和音に対する性能を明らかにするために、図 5.24 に示す旋律に対して採譜を行なった結果について以下に説明する。

図 5.24 に示す楽譜を piano, violin で演奏した際の出力波形を図 5.25, 5.26 に、また、本方式を用いて音高及び音価の推定を行なった結果を、図 5.27, 5.28 に、それぞれ示す。

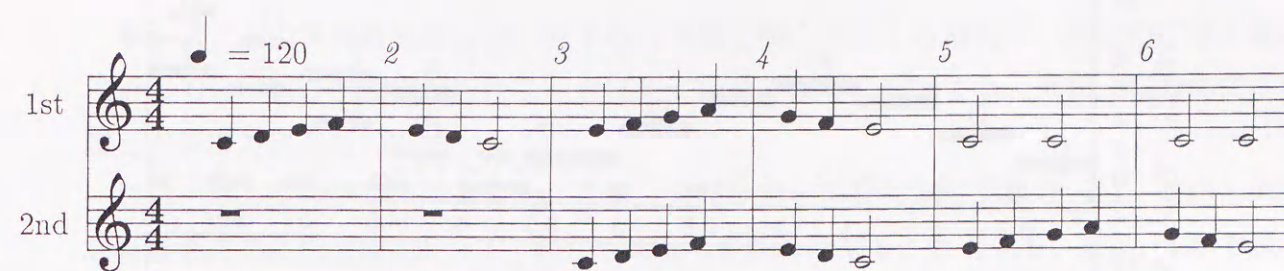


図 5.24 和音の採譜対象楽譜「カエルの歌」の輪唱

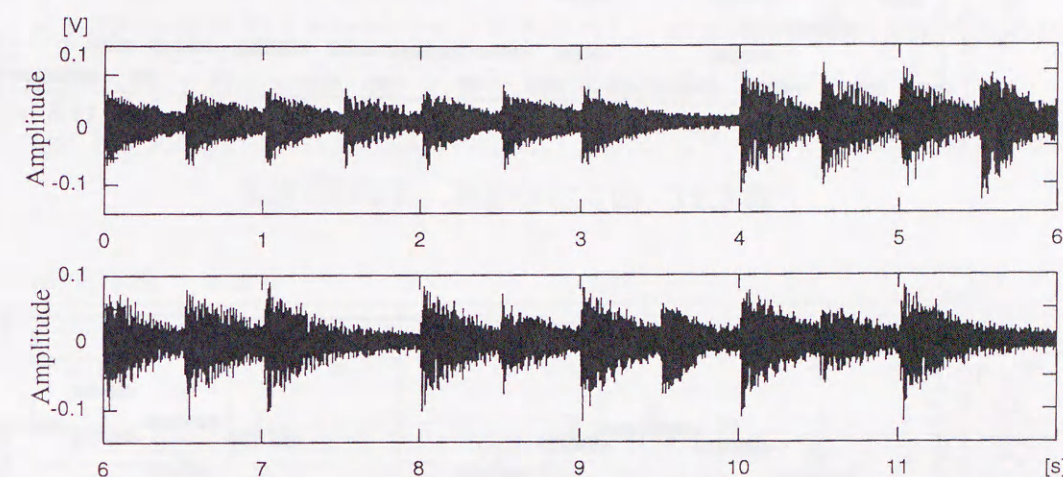


図 5.25 「カエルの歌」の輪唱を piano で演奏した場合の出力波形

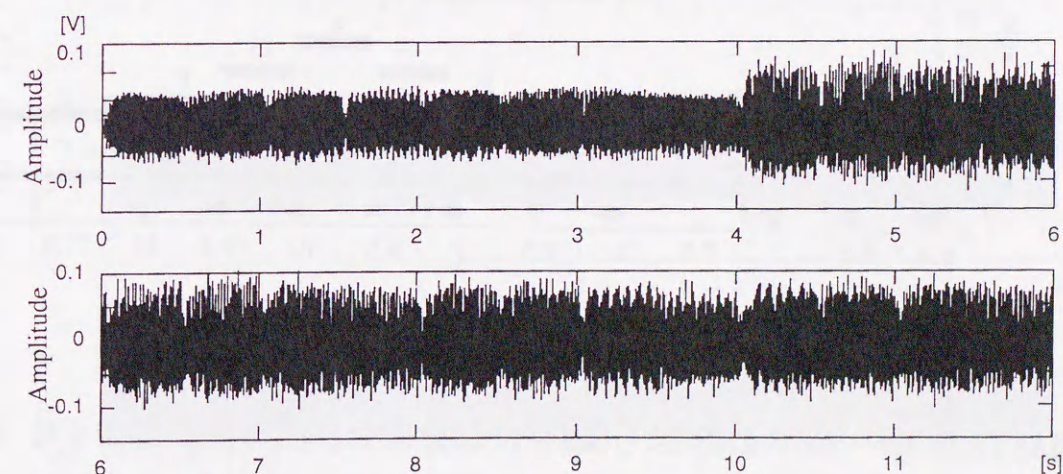


図 5.26 「カエルの歌」の輪唱を violin で演奏した場合の出力波形

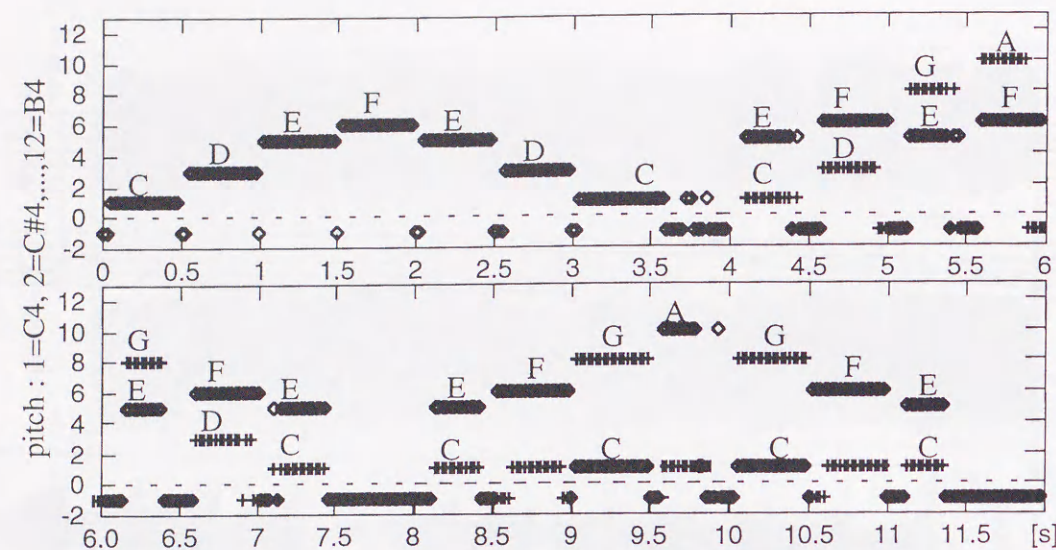


図 5.27 図 5.25 の音高、音価推定結果

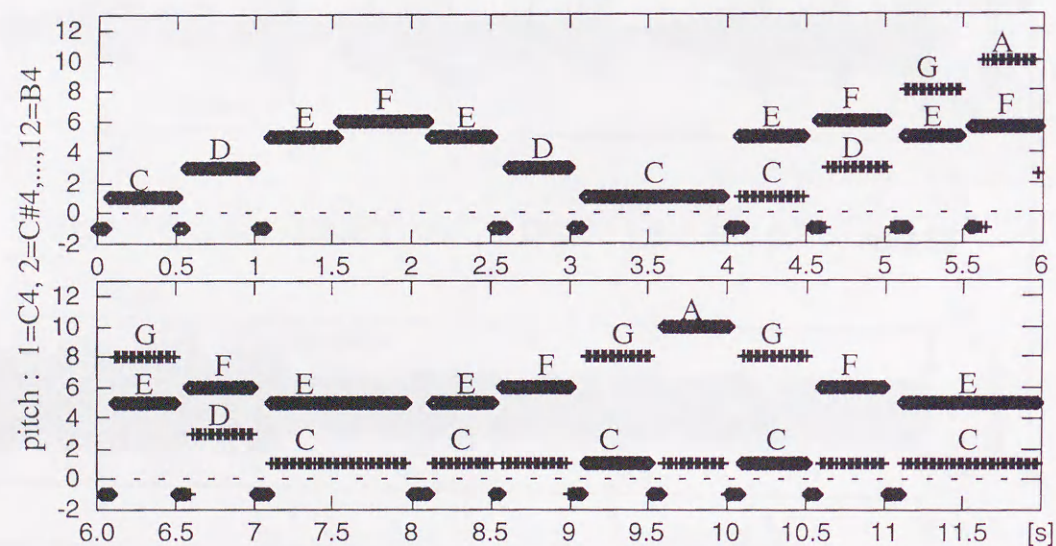


図 5.28 図 5.26 の音高、音価推定結果

図 5.27 より、単一楽器の和音により構成される旋律に対して、本方式により音高、音価が正しく検出されることを確認した。但し、単音の場合と同様に、piano 音の持つ減衰の特性より、継続時間の長い音については、音価が正しく検出されないという問題があることが分かった。

5.4.3 異種複数楽器音に対する採譜

本システムの異種複数楽器音に対する有用性を明らかにする目的で、図 5.29 に示す旋律に対して音高及び音価推定を行なった結果について以下に説明する。まお、奏譜化には前述と同様な方式を用いた。

なお、本方式のアルゴリズム上、音高、音価推定と楽器推定は別処理とし、前述と同様な奏譜形式による出力に対して、後から楽器名を割り当てることとする。また、4 章で説明した和音分離及び楽器推定法は、基本周波数が他の音と重複を持つ場合、その音高を推定するために非常に長いデータ長を必要とする。このため、今回の異種複数楽器音に対しては、このような和音が入力されないことを条件とし、楽器推定時間の短縮を図るため、重複成分のスペクトルは楽器推定のための差分計算に使用しない。

図 5.30 に上記の楽譜を演奏した際の波形を、図 5.31 に本方式で音高及び音価の推定を行

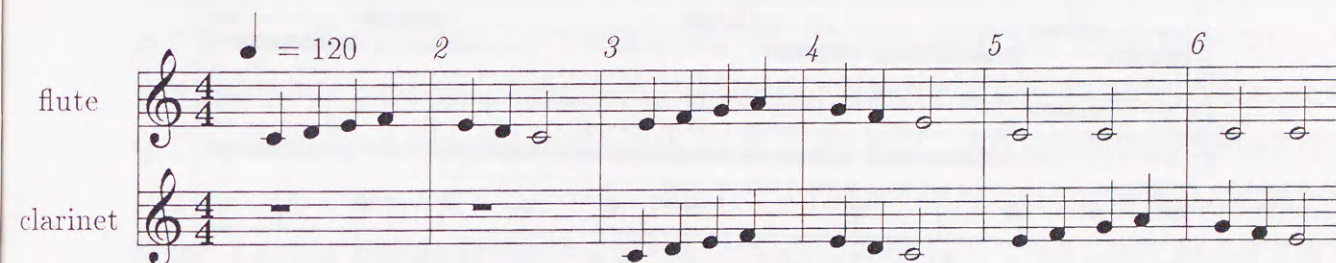


図 5.29 異種複数楽器の採譜対象楽譜「カエルの歌」の輪唱

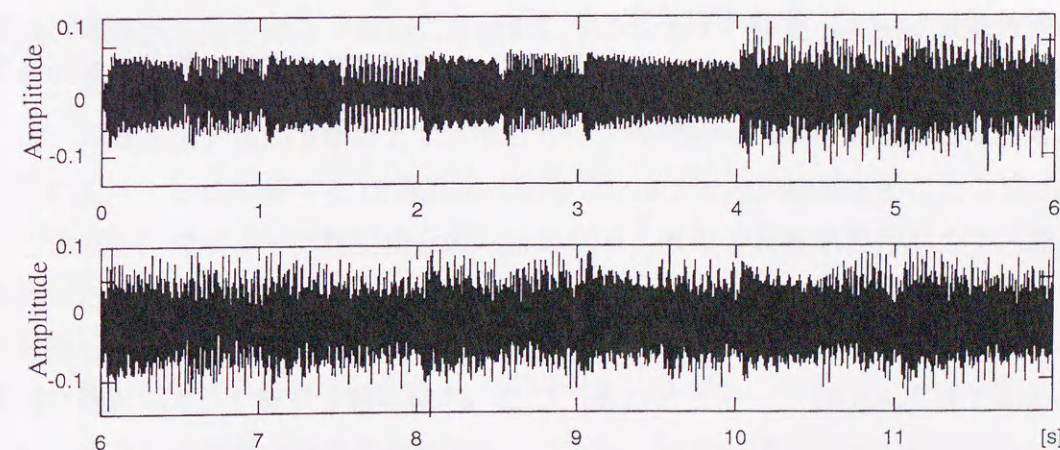


図 5.30 図 5.29 の演奏波形

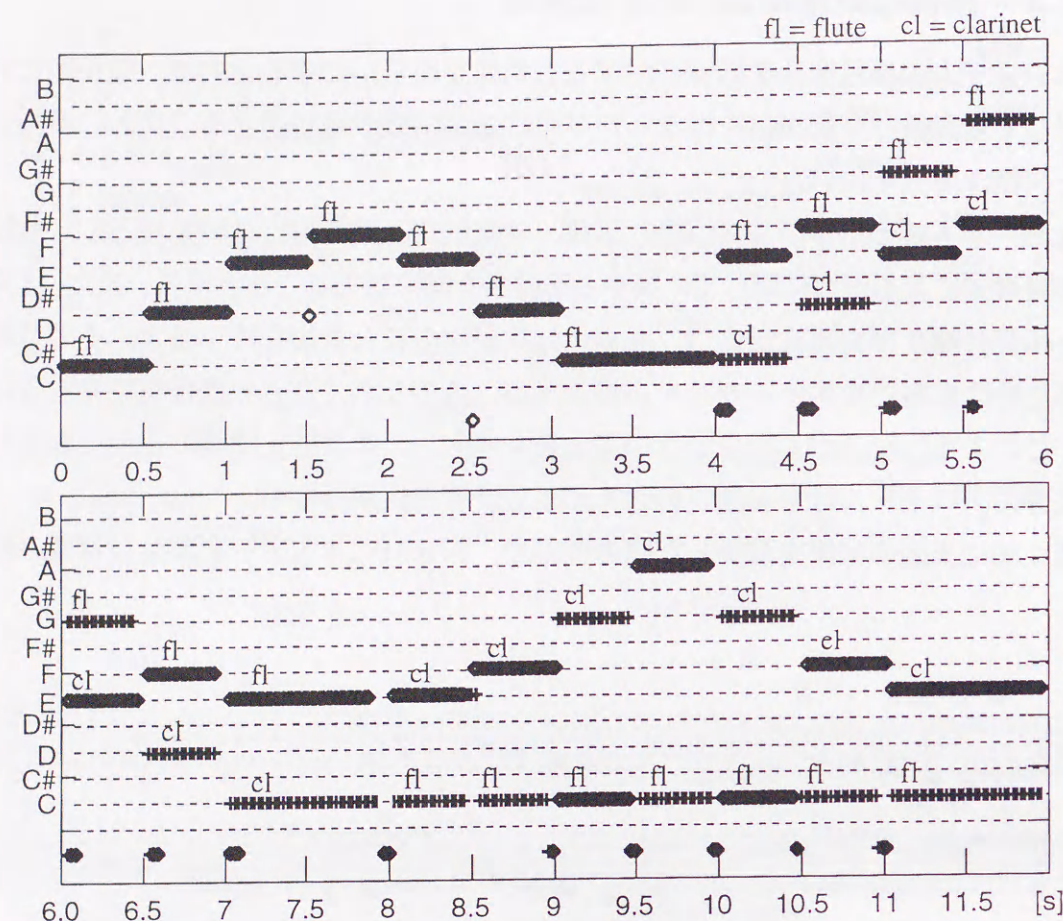


図 5.31 図 5.30 の音高、音価推定結果

なった結果について、それぞれ示す。

図 5.31 より、単一楽器の単音の場合と同様に、異種複数楽器音の旋律に対しても、本方式により音高及び音価が正しく検出されていることを確認した。

第 6 章

結 言

6.1 まとめ

楽器音や歌唱音などの音楽信号は、一般に、音高（音名、オクターブ）を決定する基本周波数の他に多数の高調波を含んでいる。また、演奏される楽曲が平均率音階に従う場合、各音高は非整数 ($2^{1/12}$) 倍の周波数比を持つ。このため、音楽信号は非常に複雑なスペクトル構造を持ち、その解析は、一般に、非常に困難とされてきた。

本研究では、上述の音楽信号の性質を考慮して、くし形フィルタによる自動採譜システムを提案し、その可能性について検討を行なった。本研究で使用するくし形フィルタは、周波数領域において等間隔に零点を持ち、音楽信号に含まれる基本周波数成分、及び、その全ての高調波成分を除去する性質を持つ。また、くし形フィルタは時間領域においてサンプル点の減算で実現できることから、採譜システム全体の処理に必要な演算量の削減が可能である。本論文では、採譜対象音域をオクターブ 3 の C (C^3 : 130.81[Hz]) からオクターブ 5 の B (B^5 : 987.77[Hz]) の 36 音とし、複数の楽器が異なる音高を同時に演奏した場合の異種複数楽器音に対して解析を行なうことで、提案法の妥当性について検証を行なった。

本論文では、初めに、くし形フィルタの等間隔な零点を利用し、入力音に含まれる基本周波数及び高調波成分を除去（零出力を検出）することで、音名の判別が可能であることを示した。また、くし形フィルタの次数と各基本周波数の関係から、フィルタ次数を変化させ、零・非零出力を検出することで、オクターブの判別が可能であることを示した。さらに、C, C#, ..., B に対応する 12 種類のくし形フィルタを縦続接続することで音名判別システムを構成し、オクターブ判別法と組み合わせることで音高推定を行なう手法を提案した。提案した音高推定アルゴリズムでは、上記のシステムに対して音楽信号を入力し、零出力を検出した段のくし形フィルタを随時先頭に移動させることで、和音に含まれる音の数、種類に関わらず、1 音ずつ順に音名、及び、そのオクターブが判別可能である。さらに、くし形フィルタの零点の重複を利用することで、音名判別システム全体のフィルタの縦続数を 7 段に削減し、重複音判別のための別処理を付け加えることで、様々な和音に対応可能なシステムを構築した。

従来の方式が、入力音に含まれる全ての周波数成分から、各音の基本周波数成分を抽出することで音高推定を行なうのに対し、本方式では、各音の全周波数成分（基本周波数＋全高調波）を除去することで音高推定を行なう。このため、従来法で問題とされていた高調波成分の重畳による処理の複雑さを回避することができ、多数の周波数成分からの基本周波数の検索を必要としないという利点を持つ。

次に、くし形フィルタの周波数成分除去の性質を利用し、時間領域において和音分離を行なうシステムを提案した。このシステムでは、 m 種類の音を含む異種複数楽器音から、くし形フィルタを利用して $m-1$ 音を除去することで、単音（単一楽器音）の抽出が可能である。このため、本方式では、従来法で必要とされていた高調波成分のグルーピング処理を必要とせず、予め用意した 12 種類のくし形フィルタを組み合わせることで、簡単に異種複数楽器音の分離が可能であるという利点を持つ。

さらに、各楽器の単音のスペクトルをテンプレートとして用意し、前述の抽出音のスペクトルと比較することで楽器推定を行なう手法について検討した。抽出された各音に含まれる周波数成分は、その音の基本周波数成分と高調波に制限されているため、非常に少ないサンプル点数の DFT により周波数解析（スペクトル抽出）が可能である。また、入力された異種複数楽器音を予め単音に分離しているため、各スペクトルの正規化が可能であり、これにより、楽器推定のためのマッチング処理が簡単化できるという利点を持つ。実際には、各スペクトルのノルム=1.0 とする正規化を行なった。同一楽器音のスペクトルは類似した形を持っているため、上記の正規化により、絶対値差分の最小値を検出することで楽器推定が可能であることを示した。

最後に、入力として電子楽器音（MIDI:KORG05R/W）を用い、計算機シミュレーションにより提案法の性能評価を行なった。MIDI 音源は、一般に、音量の時間的変化、音色（スペクトル構造）等、実際の楽器音を忠実に再現しており、より実楽器音に近い演奏データを得ることが可能であるとされている。それと同時に、発音時のアタック（波形の立上り時点の強い雑音成分）、ビブラートなどの周波数変動等、周波数解析に不利となる条件についても忠実に再現されているため、より実用的なシステムの構築に役立つものと考えられる。

本論文では、上述の MIDI 音源により、ピアノ音、バイオリン音、ビオラ音、フルート音、クラリネット音、ホルン音、トランペット音を発生させ、これらの楽器音に対する音高推定及び楽器推定を行なった。なお、ピアノ、バイオリン、ビオラは弦楽器であり、その波形は非定常性が強く、楽器音に含まれる高調波成分の数も多いことから、一般に解析が難しいとされている楽器である。一方、フルート、クラリネットは木管楽器で高調波成分の数が少なく、また、トランペット、ホルンは金管楽器で高調波成分の数は多いながら

も、各管楽器ともに波形はほぼ定常であるという性質を持つ。このように、同じ仲間（弦楽器同士、木・金管楽器同士）の楽器による和音や異なる仲間（弦楽器＋木管楽器、弦楽器＋金管楽器、等）の楽器による和音を解析し、本方式の性能評価を行なった。

計算機シミュレーションでは、まず、くし形フィルタの性能を明らかにするために、単一楽器音に対する音高、音価推定を行なった。これにより、入力音に対応したくし形フィルタの出力がほぼ零値をとり、使用楽器の種類に依存せず入力音名、及び、その音価が検出されることを確認した。しかし、この零出力の検出結果より、ピアノ音の減衰率が大きい部分（アタックレベルからブレイクポイントの間）には非調和周波数成分が含まれ、出力には比較的大きな雑音が生じることが明らかになった。また、バイオリン音の波形の後半部分では、ビブラートによる周波数変動が生じているため、出力にはくし形フィルタでは除去しきれない雑音（振動成分）が含まれることが分かった。しかし、このような雑音成分が含まれている場合でも、各くし形フィルタの出力の減衰率を計算し、最も大きく減衰しているものを零出力と判断することで、音名判別が可能であることを確認した。さらに、くし形フィルタの次数を変化させ、零・非零出力検出によりオクターブ推定が可能であることを確認した。

次に、上述の楽器を組み合わせた異種複数楽器音に対して音高、音価推定を行ない、本方式が楽器の種類に依存せず、音高、音価を推定できることを確認した。さらに、入力された異種複数楽器音に対して楽器推定を行ない、様々な周波数の重複パターンを持つ和音に対して、予め用意したテンプレートスペクトルと分離音のスペクトルの比較により楽器推定が可能であることを確認した。

さらに、異種複数楽器音の簡単な旋律に対して採譜を行ない、奏譜形式での表現が可能であることを示した。

6.2 今後の課題

本論文における計算機シミュレーションでは、電子楽器音を用い、単一楽器により演奏された単音及び和音、異種楽器により演奏された和音に対して、音高推定、音価推定、楽器推定を行なった。しかし、今回の実験では、採譜対象楽器の種類が少なく、また、計算機シミュレーションに使用した和音の音の組合せも限定されている。従って、現時点では、くし形フィルタによる音高、音価推定、及びスペクトル構造を利用した楽器推定について、その原理を確かめた段階であり、今後は採譜対象を拡大してさらなる実験を行なう必要があると考えられる。特に、楽器推定における協和音の解析には、1.0[s]～2.0[s] 程度の音価が必要であり、実際の採譜アルゴリズムとしては現実的ではないと言える。また、現段階

では実現できていないユニゾン和音に対するアルゴリズムを確立する必要がある。

さらに、現在はいくし形フィルタを利用した採譜により得られた結果について、時間軸上で各音高の継続時間を示す楽譜の形式を用いているが、実際に楽譜化するためには音符形式により表示する方が望ましいと考えられる。しかし、一般的に、同じ音符でもテンポが異なれば音価に違いが生じるため、入力された楽音に対して、その音が楽譜上の何音符であるのかを正確に判断するためには、予め楽曲のテンポを推定しておく必要がある。テンポの推定方法としては、ペースとなる楽器音を選び、その音高成分の変化から推定する手法や、旋律の変化のパターン（同じ旋律の繰り返し等）から推定する手法などが考えられる。また、楽曲に含まれる最短音符を十六分音符として、他の音符を相対的に決定する手法なども考えられる。今後は、入力された音楽信号から、テンポを判別するためのアルゴリズムを確立し、音価との相対的な関係から音符を判別する手法についても検討を行なう必要がある。

しかし、民族楽器や民謡等を採譜対象として考えた場合、一般的に、伝統的な音楽は平均率音階で規定されている音高とは異なる独自の音階を持つことが多く、五線譜形式による記録は困難である。自動採譜システムの目的として、このような歴史的、民族的な文化としての音楽を記録、保存することを考えるならば、五線譜形式ではなく、より拡張性の高い表記方法を検討する必要があると考えている。

最後に、音楽というのは演奏家の感性に依存する部分が大きく、同じ楽曲を演奏した場合でも奏者や指揮者によりその表現は大きく異なっている。この感性は、テンポやピッチ（基本周波数）の揺らぎ、音の立ち上がりや立ち下がり具合等、音楽信号特有の非定常な性質として現れる。しかし、現在の採譜の研究では、このような演奏家に依存する不確定さを無視し、周波数補正や音符の長さ等の均一化を図った上で、規格化された楽譜に書き直すことが一般的である。上述のような感性を楽譜化することができれば、一般に名演奏と呼ばれる演奏を紙面に写すことが可能となり、音楽の分野では大きな意味を持つと考えられる。このような感性については、現在の楽譜形式では完全に表現することは不可能であるが、五線譜という形式にこだわらずに、将来的には演奏家の感性を含んだ採譜の実現を目標にしたいと考えている。

謝辞

本論文は、豊橋技術科学大学情報工学系情報システム大講座信号処理研究室において行なった研究の成果をまとめたものである。

本研究がこのような成果を出すことができたのは、終始、懇切な御指導と暖かい御助言を頂いた本学情報システム大講座田所嘉昭教授の賜であり、ここに深く感謝致申し上げます。

また、御多忙な時間を割いて御討論頂き、本研究に関して有益なる御助言、御指導を賜わった本学計算機大講座中川聖一教授、ならびに情報システム大講座宮崎保光教授、横山光雄教授に深く感謝の意を表します。

さらに、本研究を進めるにあたり、有益な御意見ならびに御指導を頂いた静岡大学電子工学研究所川人祥二教授に厚く御礼申し上げます。

最後に、日頃の研究生生活において、数々の御支援、御助力を頂いた信号処理研究室の諸氏に感謝致します。

参考文献

- [1] Martin Piszczalski, Bernard A. Galler, "Automatic Music Transcription", Computer Music Journal, Box E, CA 94025, pp.24-31, 1977
- [2] James A. Moore, "signal Processing Aspects of Computer Music : A Suvey , "Proceeding of the IEEE, Vol.65, No.8, pp.1108-1137, 1977
- [3] 井口 征士, "音楽情報の処理-電子計算機を用いた自動採譜-", 計測制御, Vol.19, No.3, pp.314-319, 1980
- [4] 原 裕一郎, 井口 征士, "複素スペクトルを用いた周波数同定, " 計測制御, Vol.19, No.9, pp.718-723, 1983
- [5] Chris Chafe, David Jaffe, Kyle Kashima, Bernard Mont-Reynaud, Julius Smith, "Techniques for Note Identification in Polyphonic Music, "ICMC, pp.399-405, 1985
- [6] 高沢 嘉光, "計算機による採譜, " 情報処理学会誌, Vol.29, No.6, pp.593-598, 1988
- [7] 片寄 晴弘, 井口 征士, "知的採譜システム, " 人工知能学会誌, Vol.5, No.1, pp.59-66, 1988
- [8] 片寄 晴弘, 今井 正和, 井口 征士, "音楽における感性情報抽出の試み, " 人工知能学会誌, Vol.3, No.6, pp.748-754, 1988
- [9] 梅本 敏孝, 青島 伸治, "適応技術を用いた採譜に有効な周波数分析法, " 計測制御, Vol.28, No.5, pp.619-625, 1992
- [10] 門藤 博, 田所 嘉昭, "拡張ノッチフーリエ変換による採譜のための音高抽出, " 計測制御, Vol.32, No.12, pp.205-210, 1995
- [11] William j. Pielemeier, Gregory H. Wakefield, Mary h. Simoni, "Time-Frequency Analysis of Musical Signals, "IEEE, Vol.8, No.9, pp.1216-1230, 1996
- [12] 三輪 多恵子, 田所 嘉昭, "ノッチフーリエ変換による採譜の検討, " 電気関係学会東海支部連合大会, 550, 1995

- [13] 本多 英基, 斎藤 努, 田所 嘉昭, “倍音処理に適したくし型フィルタに基づく自動採譜システムの検討,” 信学技報, DSP95, No.82, pp.25-30, 1995
- [14] 斎藤 努, 本多 英基, 田所 嘉昭, “自動採譜のためのくし型フィルタによる和音処理の検討,” 信学ソサイエティ大会, A-102, 1995
- [15] William j. Pielemeier, Gregory H. Wakefield, “A High-Resolution Time-Frequency Representation for Musical Instrument Signals,” J.Acoust.Soc.Am., Vol.99, No.4, pp.2382-2396, 1996
- [16] William J. Pielemeier, Gregory H. Wakefield, Mary H. Simoni, “Time-Frequency Analysis of Musical Signals,” Proc. of IEEE, Vol. 84, No. 9, pp.1216-1230, 1996
- [17] Tsutomu Saito, Takashi Matsui, Hideki Honda, Yoshiaki Tadokoro, “Real-time Realization of Transcription Based on Comb Filters Using DSPs,” ICSPAT96, pp.1594-1598, 1996
- [18] Yoshiaki Tadokoro, Miori Ozawa, Miho Wadamori, Tsutomu Saito, “A Transcription System Based on Synchronous Addition and Subtraction Processing,” ICSPAT96, 1996
- [19] 三輪 多恵子, 田所 嘉昭, 斎藤 努, “零出力に注目したオーバーサンプリング法くし型フィルタによる音階検出,” 電気関係学会東海支部連合大会, 527, 1996
- [20] 三輪 多恵子, 田所 嘉昭, 斎藤 努, “零出力に注目したくし型フィルタによる音階検出,” 電学論 C, vol.118-C, no.1, pp.57-64, 1998
- [21] 後藤 真孝, 村岡 洋一, “打楽器音を対象とした音源分離システム,” 信学論 (D-II), Vol.J77-D-II, No.5, pp.901-911, 1994
- [22] 柏野 邦夫, 木下 智義, 中臺 一博, 田中 英彦, “音楽情景分析の処理モデル OPTIMA における和音の認識,” 信学論 (D-II), Vol.J79-D-II, No.11, pp.1762-1770, 1996
- [23] 永田 仁史, 安部 正人, 城戸 健一, “多数センサによる音源波形の推定,” 日本音響学会誌, Vol.47, No.4, pp.268-273, 1995
- [24] 植田 護, 橋本 周司, “音源分離のためのブラインドデコンポジションアルゴリズム,” 情報処理学会誌, Vol.38, No.1, pp.146-156, 1997

- [25] 小野 徹太郎, 斎藤 英雄, 小沢 慎治, “自動採譜のための GA を用いた混合音推定,” 計測制御学, Vol.33, No.5, pp.417-423, 1997
- [26] 柏野 邦夫, 村瀬 洋, “適応型混合テンプレートを用いた音源同定,” 信学技報, SP96-117, pp.21-26, 1997
- [27] Ramamurthy Mani, S. Hamid Mawab, “Knowledge-based processing of multicomponent signals in a musical application,” Signal Processing 74, pp.47-69, 1999
- [28] 三輪 多恵子, 田所 嘉昭, 斎藤 努, “くし型フィルタによる採譜のための音源分離,” 信学技報, DSP97-79, pp.61-66, 1997
- [29] 磯部 隆俊, 山村 毅, 大西 昇, “複数楽器のモノラル混合信号からの音源の分離,” 信学技報, pp.9-16, 1998
- [30] 柏野 邦夫, 村瀬 洋, “アンサンブル実演奏の自動アンミキサ,” 信学技報, SP97-104, pp.23-40, 1998
- [31] 白土 保, “混合音からの基本周波数分離抽出,” システム制御情報学会論文誌, Vol.12, No.1, pp.56-62, 1999
- [32] 西川 英一, 三輪 多恵子, 田所 嘉昭, “採譜のための同一音高の複数楽器音からの楽器推定,” 信学総合大会, A-4-53, 1999
- [33] 三輪 多恵子, 田所 嘉昭, 斎藤 努, “異楽器による同一音高和音の楽器推定法の検討,” 信学ソサイエティ大会, A-46, 1999
- [34] Taeko Miwa, Yoshiaki Tadokoro, Tsutomu Saito, “Musical pitch estimation and discrimination of musical instruments using comb filters for transcription,” The 42nd Midwest Symposium on Circuits and Systems (MWSCAS), vol.1, pp.105-108, Aug.1999
- [35] Taeko Miwa, Yoshiaki Tadokoro, Tsutomu Saito, “Musical pitch estimation of different musical instruments sound using comb filters for transcription,” The International Conference on Signal Processing Application and Technology (ICSPAT'99), Specialized Applications, no.7, Nov.1999

- [36] 山口 満, 三輪 多恵子, 田所 嘉昭, “異なるオクターブの同一音和音に対する楽器推定法の検討,” 信学総合大会, A-4-31, 2000
- [37] Yoshiaki Tadokoro, Taeko Miwa, "Musical pitch and instrument estimation of polyphony using comb filters for transcription," 4th World Multiconference on Circuits, Systems, Communications and Computers (CSCC2000), Advances in Physics, Electronics and Signal Processing Applications, pp.315-319, July 2000
- [38] 山口 公典, 安藤 繁雄, “短時間スペクトル分析法の自然楽器音への適用,” 日本音響学会誌, Vol.33, No.6, pp.291-300, 1977
- [39] Yoshiaki Tadokoro and Kenchi Abe, "Notch Fourier transform," Trans. Aconst., Speech, and Signal Processing, vol.ASSP-35, no.9, pp.1282-1288, 1987
- [40] 樋口龍雄, “ディジタル信号処理の基礎,” 昭光堂, 1986
- [41] “コンピュータと音楽,” bit 別冊, 共立出版, 1987
- [42] “複数種類の楽器演奏を対象とする自動採譜システムの調査研究報告書,” 財団法人 機械システム振興協会, システム技術開発調査研究 3-R-1(6), 1991
- [43] “05R/W Quick Guide,” KORG, 1994
- [44] H. F. オルソン, “音楽工学,” 誠文堂新光社
- [45] J. R. Poerce, “音楽の化学,” 日系サイエンス社, 1989
- [46] 安藤 由典, “楽器の音響学,” 音楽之友社, 1996
- [47] “コンピュータと音楽の世界-基礎からフロンティアまで,” bit 別冊, 共立出版, 1998
- [48] “文字と音の情報処理,” マルチメディア情報学 4, 岩波書店, 2000

研究業績

投稿論文

- 1: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “零出力に注目したくし形フィルタによる音階検出,” 電気学会論文誌 C, vol.118-C, no.1, pp.57-64, Jan.1998
- 2: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “くし形フィルタを利用した採譜のための異楽器音中のピッチ推定,” 電子情報通信学会論文誌 DII, vol.81-DII, no.9, pp.1965-1974, Sept.1998

国際会議

- 1: Taeko Miwa, Yoshiaki Tadokoro, Tsutomu Saito, "Musical pitch estimation and discrimination of musical instruments using comb filters for transcription," The 42nd Midwest Symposium on Circuits and Systems (MWSCAS), vol.1, pp.105-108, Aug.1999
- 2: Taeko Miwa, Yoshiaki Tadokoro, Tsutomu Saito, "Musical pitch estimation of different musical instruments sound using comb filters for transcription," The International Conference on Signal Processing Application and Technology (ICSPAT'99), Specialized Applications, no.7, Nov.1999
- 3: Yoshiaki Tadokoro, Taeko Miwa, "Musical pitch and instrument estimation of polyphony using comb filters for transcription," 4th World Multiconference on Circuits, Systems, Communications and Computers (CSCC2000), Advances in Physics, Electronics and Signal Processing Applications, pp.315-319, July 2000

口頭発表

- 1: 三輪 多恵子, 田所 嘉昭, “ノッチフーリエ変換による採譜の検討,” 電気関係学会東海支部連合大会, 550, Sept.1995
- 2: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “零出力に注目したくし形フィルタによる音階検出,” 電子情報通信学会ソサイエティ大会, A-111, Sept.1996

- 3: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “零出力に注目したオーバーサンプリング法くし形フィルタによる音階検出,” 電気関係学会東海支部連合大会, 527, Oct.1996
- 4: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “零出力に注目したくし形フィルタによるオクターブ2の音階検出,” 電子情報通信学会総合大会, A-458, Mar.1997
- 5: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “くし形フィルタを用いた採譜のための音源分離,” 電子情報通信学会技術研究報告, DSP97-79, pp.57-64, July 1997
- 6: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “くし形フィルタによる混合音和音からの演奏楽器の推定,” 情報処理学会全国大会, 5C-04, Oct.1998
- 7: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “くし形フィルタによる異楽器音中のピッチ推定と演奏楽器推定,” 情報処理学会研究報告, 98-MUS-27, pp.15-20, Oct.1998
- 8: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “くし形フィルタによる実楽器音を対象とした演奏楽器推定,” 電子情報通信学会技術研究報告, DSP98-138, pp.15-22, Feb.1999
- 9: 西川 英一, 三輪 多恵子, 田所 嘉昭, “採譜のための同一音高の複数楽器音からの楽器推定,” 電子情報通信学会総合大会, A-4-53, Apr.1999
- 10: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “異楽器による同一音高和音の演奏楽器推定法の検討,” 電子情報通信学会ソサイエティ大会, A-4-6, Nov.1999
- 11: 山口 満, 三輪 多恵子, 田所 嘉昭, “異なるオクターブの同一音和音に対する楽器推定法の検討,” 電子情報通信学会総合大会, A-4-31, Mar.2000
- 12: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “周波数成分が重複する和音に対するスペクトル抽出法の検討,” 電気関係学会東海支部連合大会, 442, Sept.2000
- 13: 山口 満, 三輪 多恵子, 田所 嘉昭, “くし形フィルタによる歌唱音のピッチ推定の基礎検討,” 電気関係学会東海支部連合大会, 441, Sept.2000
- 14: 三輪 多恵子, 田所 嘉昭, 斎藤 努, “くし形フィルタによる楽器音の採譜の問題点とその解決法の検討,” 電子情報通信学会技術研究報告, Sept.2000
- 15: 山口 満, 三輪 多恵子, 田所 嘉昭, “並列構成くし形フィルタによる歌唱の採譜の基礎検討,” 電子情報通信学会技術研究報告, Sept.2000

- 16: 山口 満, 三輪 多恵子, 田所 嘉昭, “くし形フィルタによる歌唱のピッチ推定の実験結果,” 電子情報通信学会ソサイエティ大会, A-4-29, Oct.2000

賞罰

- 1: 零出力に注目したオーバーサンプリング法くし形フィルタによる音階検出,” 平成8年度 東海支部連合大会 奨励賞, 電子情報通信学会 東海支部, Feb.1997
- 2: “くし形フィルタを利用した採譜のための異楽器音中のピッチ推定,” 平成10年度 学生研究奨励賞, 電子情報通信学会 東海支部, May 1999

付録 A

テンプレート スペクトル

各楽器のテンプレートスペクトルを以下に示す.

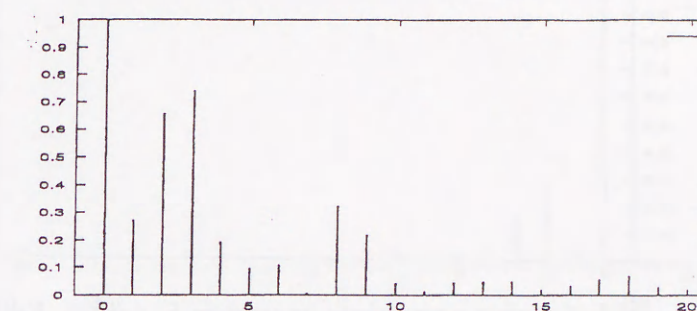


図 A.1 piano template spectrum 1 $C^3 \sim D^3$

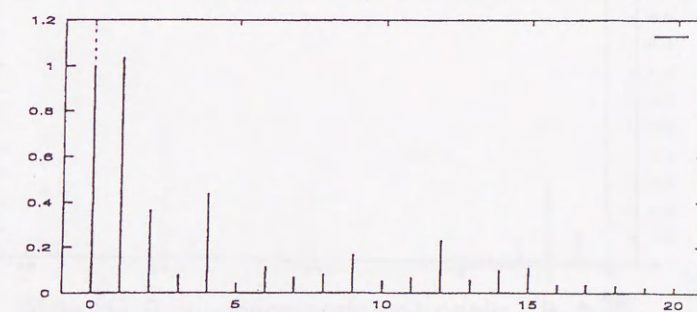


図 A.2 piano template spectrum 2 $D\#^3 \sim G^3$

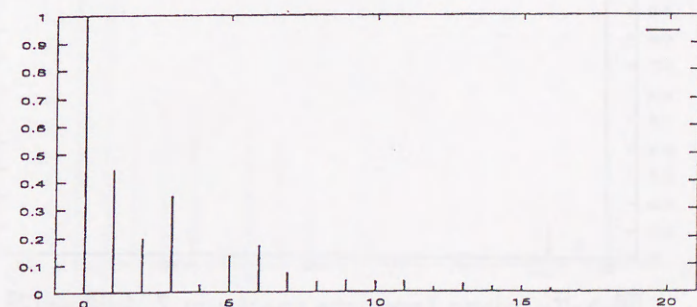
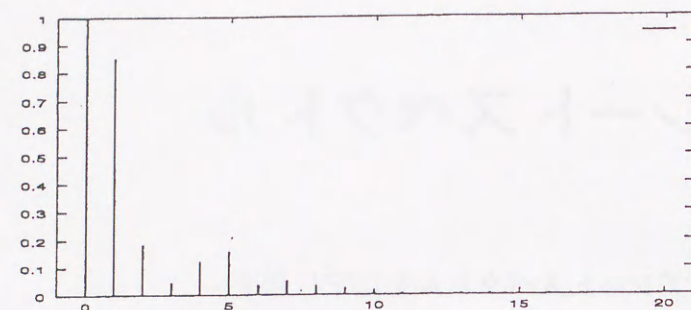
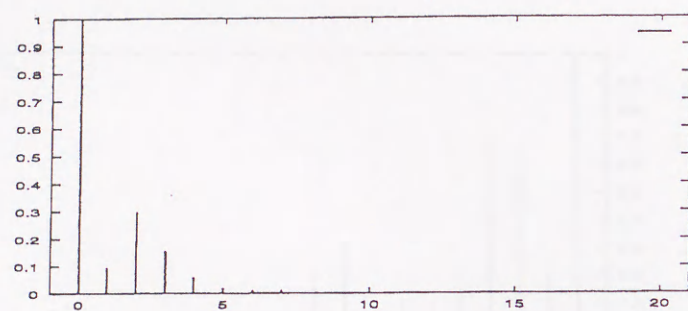


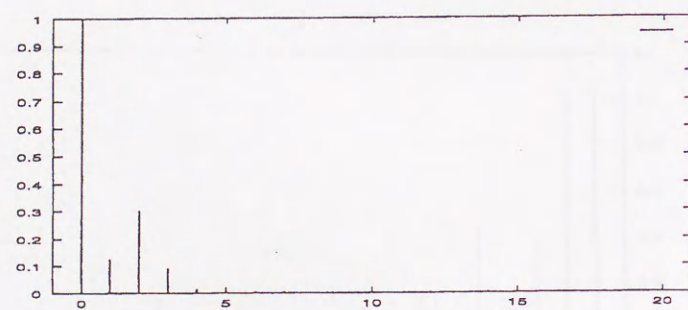
図 A.3 piano template spectrum 3 $G^3 \sim C^4$



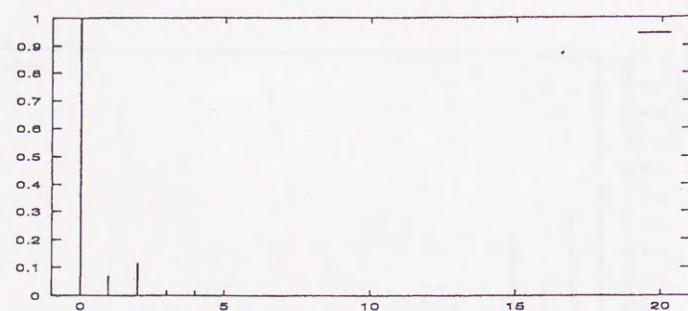
☒ A.4 piano template spectrum 4 $D\#^4 \sim D^4$



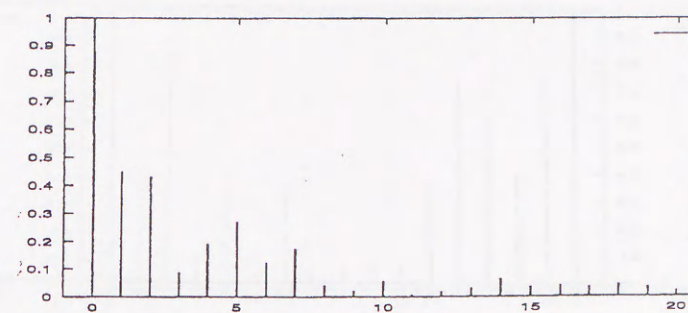
☒ A.5 piano template spectrum 5 $A\#^4 \sim F\#^4$



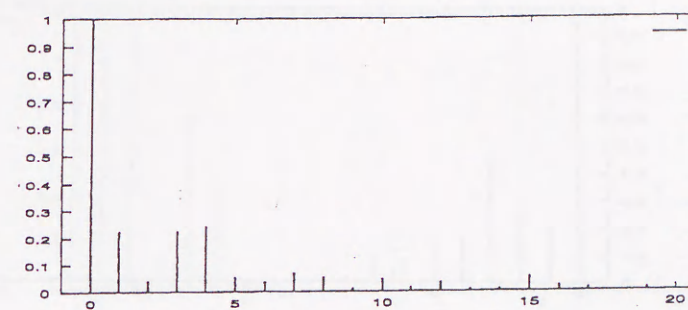
☒ A.6 piano template spectrum 6 $D^5 \sim B^4$



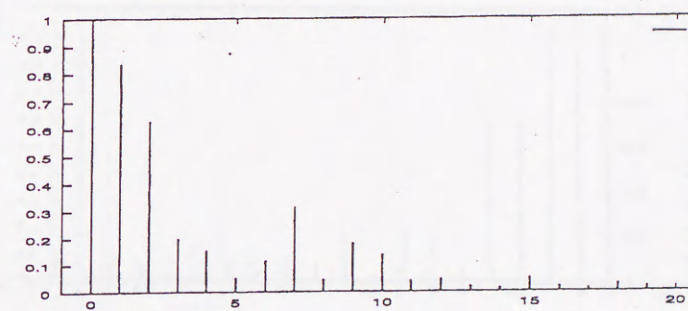
☒ A.7 piano template spectrum 7 $A\#^5 \sim G^5$



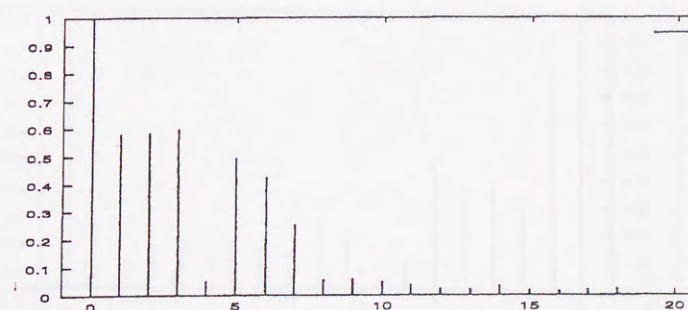
☒ A.8 violin template spectrum 1 C^3



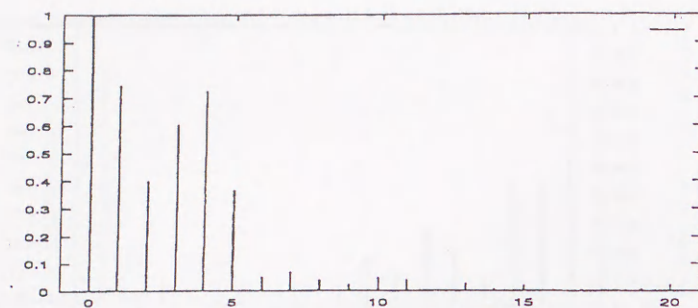
☒ A.9 violin template spectrum 2 $C\#^3 \sim G^3$



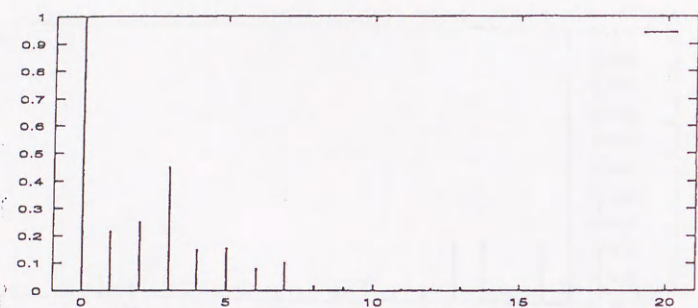
☒ A.10 violin template spectrum 3 $G\#^3 \sim D^4$



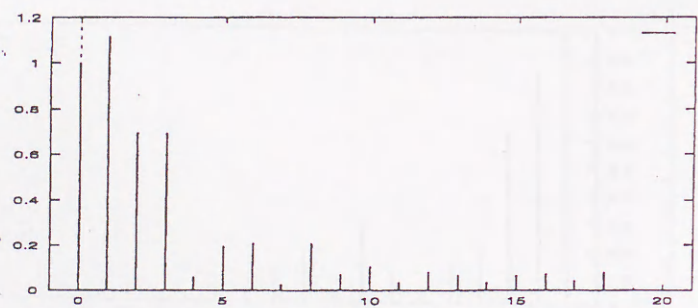
☒ A.11 violin template spectrum 4 $D\#^4 \sim A^4$



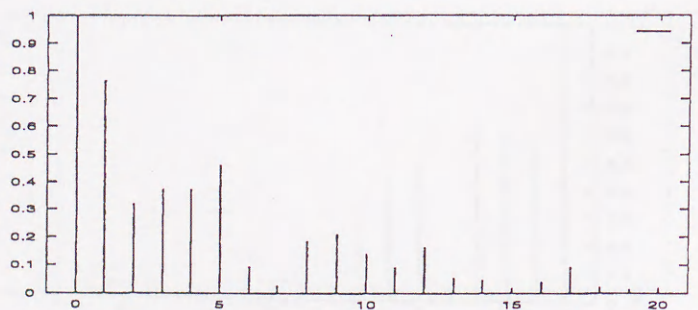
☒ A.12 violin template spectrum 5 $A\#^4 \sim F^5$



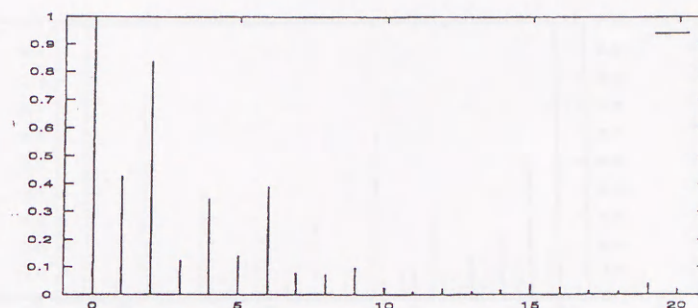
☒ A.13 violin template spectrum 6 $F\#^5 \sim B^5$



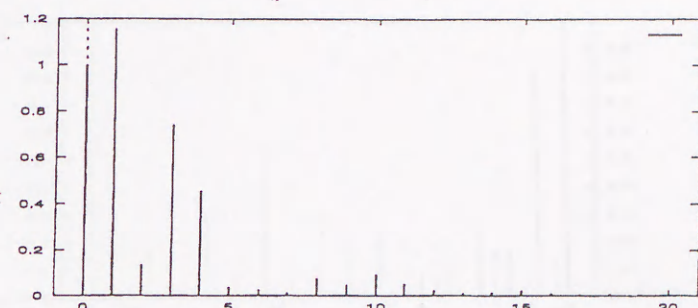
☒ A.14 viola template spectrum 1 $C^3 \sim C\#^3$



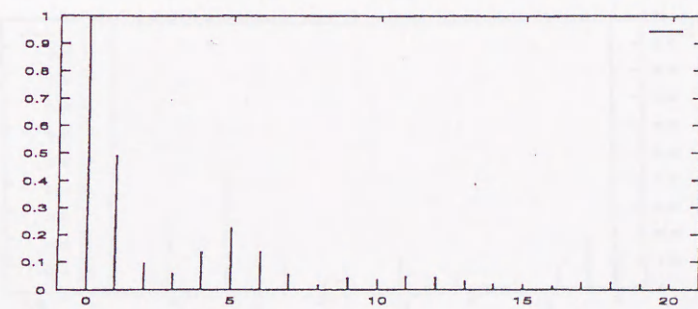
☒ A.15 viola template spectrum 2 $D^3 \sim A^3$



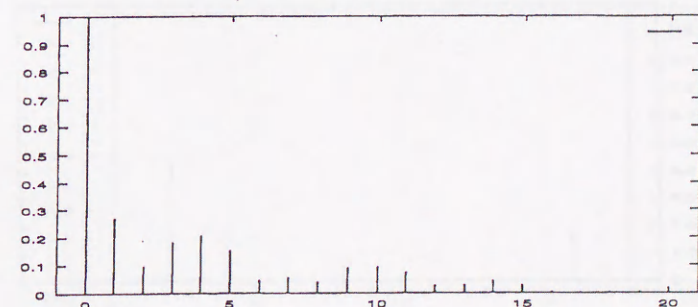
☒ A.16 viola template spectrum 3 $A\#^3 \sim G^4$



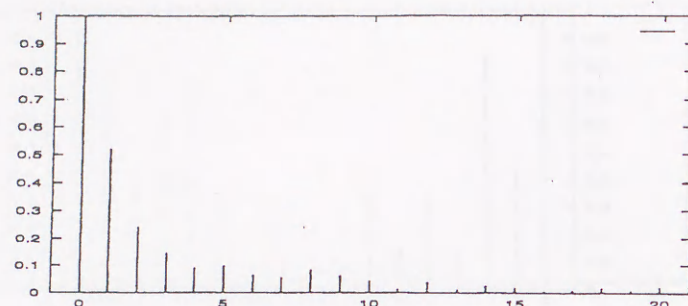
☒ A.17 viola template spectrum 4 $G\#^4 \sim B^5$



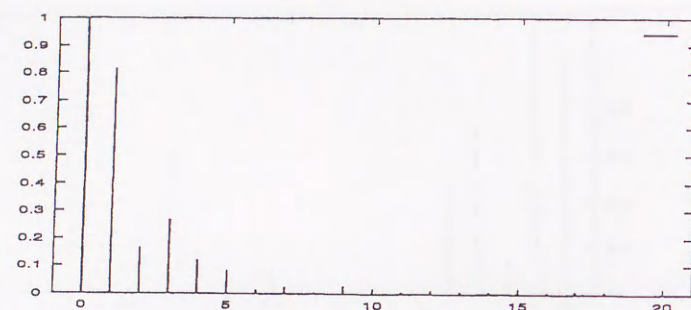
☒ A.18 flute template spectrum 1 $C^3 \sim C\#^4$



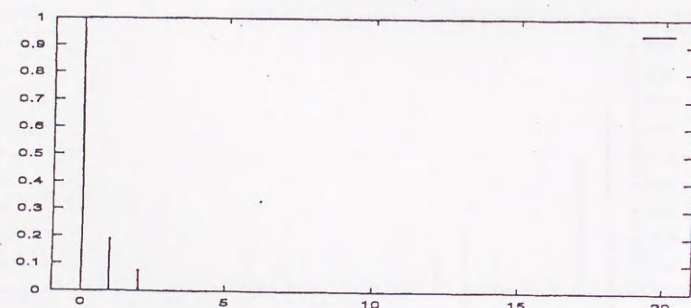
☒ A.19 flute template spectrum 2 D^4



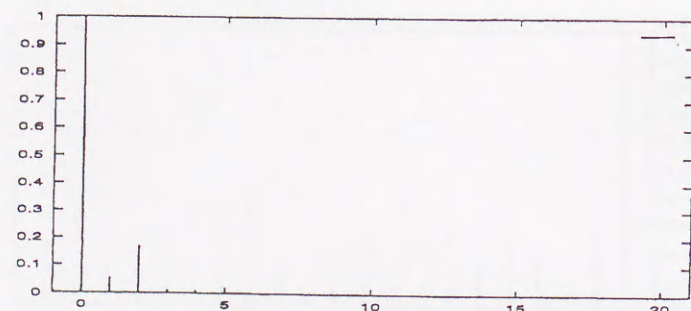
☒ A.20 flute template spectrum 3 $D\#^4 \sim F\#^4$



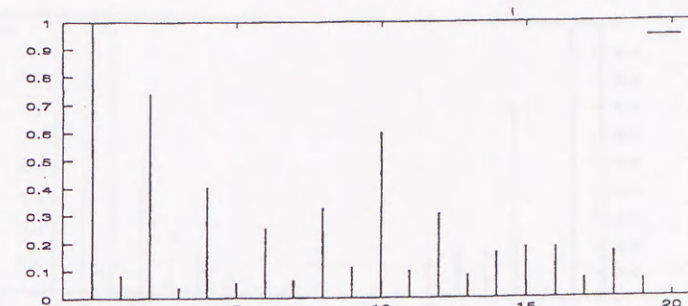
☒ A.21 flute template spectrum 4 $G^4 \sim A\#^4$



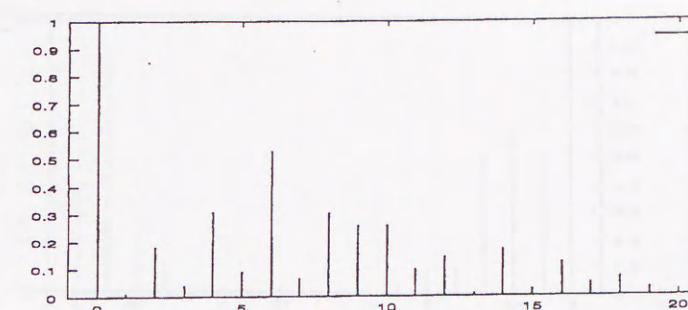
☒ A.22 flute template spectrum 5 $B^4 \sim G^5$



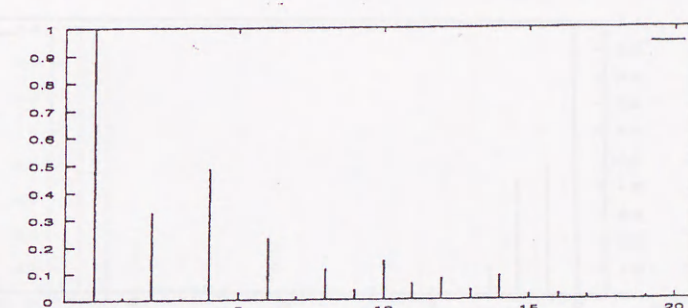
☒ A.23 flute template spectrum 6 $G\#^5 \sim B^5$



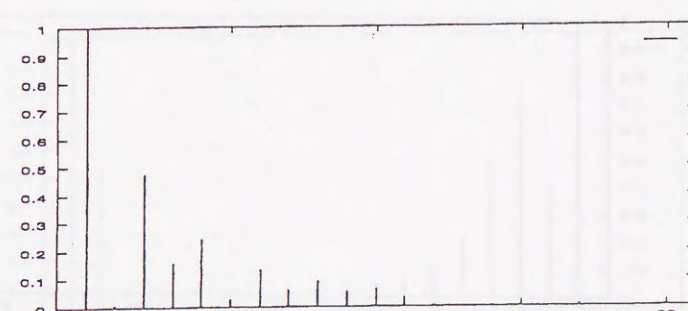
☒ A.24 clarinet template spectrum 1 $C^3 \sim D^3$



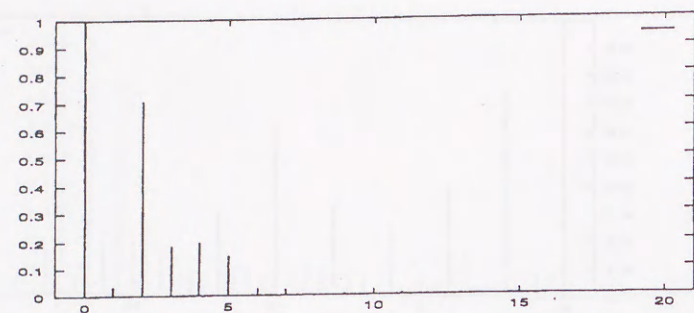
☒ A.25 clarinet template spectrum 2 $D\#^3 \sim G^3$



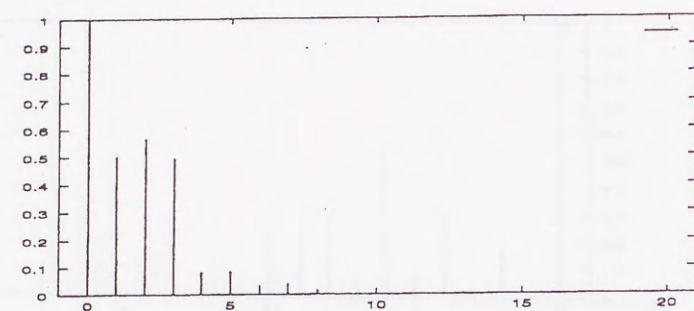
☒ A.26 clarinet template spectrum 3 $G\#^3 \sim C^4$



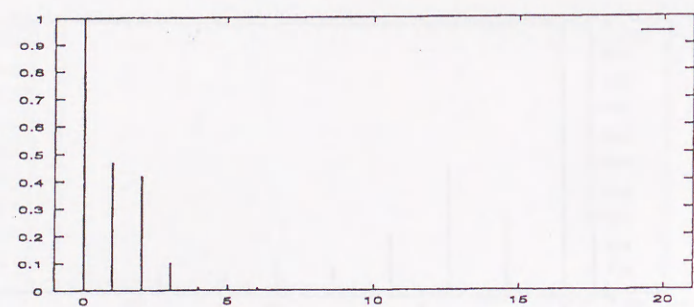
☒ A.27 clarinet template spectrum 4 $C\#^4 \sim G^4$



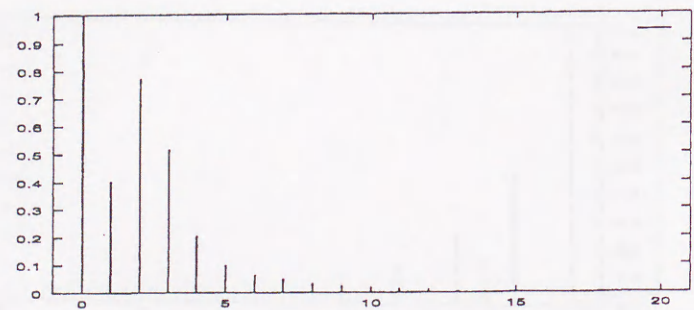
☒ A.28 clarinet template spectrum 5 $G\#^4 \sim C^5$



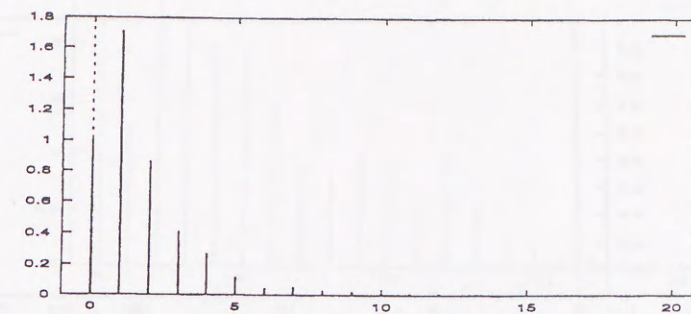
☒ A.29 clarinet template spectrum 6 $C\#^5 \sim G^5$



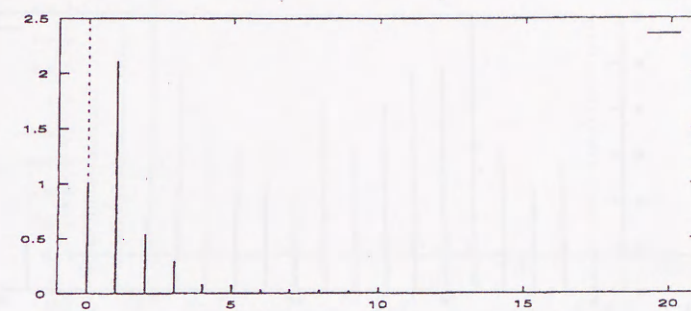
☒ A.30 clarinet template spectrum 7 $G\#^5 \sim B^5$



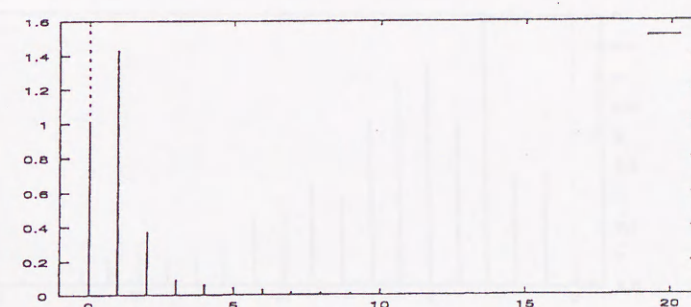
☒ A.31 horn template spectrum 1 $C^3 \sim C\#^3$



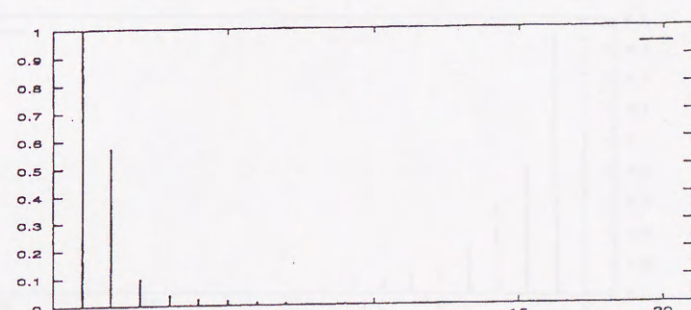
☒ A.32 horn template spectrum 2 $D^3 \sim F\#^3$



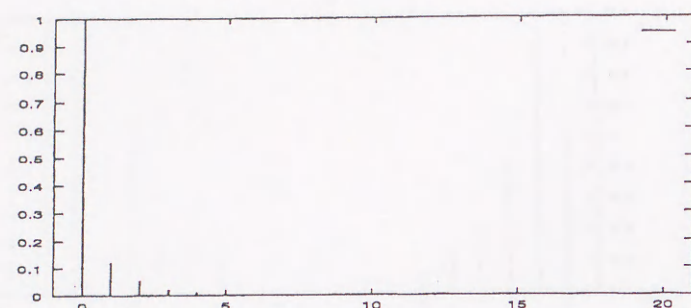
☒ A.33 horn template spectrum 3 $G^3 \sim A\#^3$



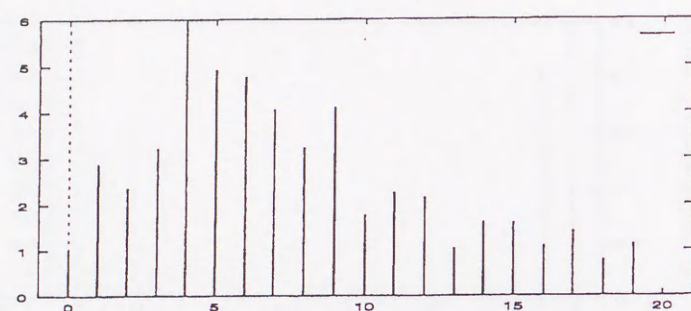
☒ A.34 horn template spectrum 4 $B^3 \sim C\#^4$



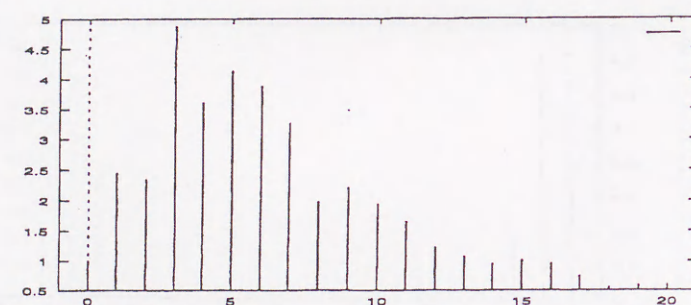
☒ A.35 horn template spectrum 5 $D^4 \sim A^4$



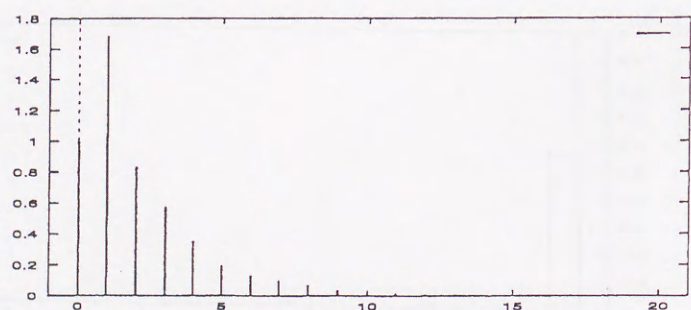
☒ A.36 horn template spectrum 6 $A\#^4 \sim B^5$



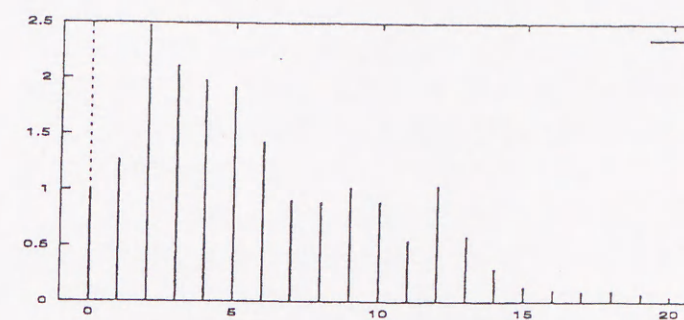
☒ A.37 trumpet template spectrum 1 $C^3 \sim C\#^3$



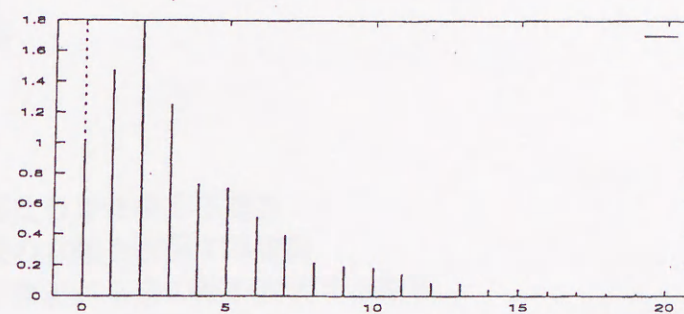
☒ A.38 trumpet template spectrum 2 $D^3 \sim G\#^3$



☒ A.39 trumpet template spectrum 3 $A^3 \sim C\#^4$



☒ A.40 trumpet template spectrum 4 $D^4 \sim B^4$



☒ A.41 trumpet template spectrum 5 $C^5 \sim B^5$

付録 B

記号

3 章から 5 章において使用した記号、及び、その説明を以下に示す。

基本的な整数記号

- i, j : 整数
- d, \hat{d} : 音の隔たりを表現する整数
- k, K : 高調波の次数を表現する整数
- m : 和音に含まれる音の数を表現する整数
- n : サンプル点を表現する整数
- q : オクターブ番号を表現する整数

その他の基本的な記号

- P : 音名を表す記号. $P \in \{C, C\#, \dots, B\}$
- P_{+d} : P 音を基準として高い方へ d 音離れた音
- P^q : オクターブ番号 q の P 音. $P^q \in \{C^3, C\#^3, \dots, B^5\}$
- M : 楽器を表す記号. $M \in \{\text{piano, violin, } \dots, \text{etc.}\}$

くし形フィルタの原理に関する記号

- $H(z)$: 基本的なくし形フィルタ
- $y_r(n)$: 基本的なくし形フィルタの出力
- $f_{q,P}$: オクターブ番号 q の P 音の基本周波数
- f_{sP} : P 音に対応するくし形フィルタのサンプリング周波数. $f_{sP} = 2 \times f_{5,P}$
- f_s : オーバサンプリング周波数
- $f_{s'_P}$: オーバサンプリング法で近似した f_{sP}
- $\Delta f_{s'_P}$: 理想的な f_{sP} と近似的な $f_{s'_P}$ の差

$x_P(n)$: サンプリグ周波数 f_{sP} による入力サンプル値系列
 $x(n)$: オーバサンプリグ周波数 f_s による入力サンプル値系列
 z_P^{-1} : サンプリグ周波数 f_{sP} による単位遅延
 z^{-1} : オーバサンプリグ周波数 f_s による単位遅延
 r : くし形フィルタの次数
 r_P : オーバサンプリグ法により近似的に実現された $H_{r,P}(z_P)$ の次数
 $H_{r,P}(z_P)$: サンプリグ周波数 f_{sP} による r 次のくし形フィルタ
 $H_{r,P}(z)$: オーバサンプリグ法により近似的に実現された $H_{r,P}(z_P)$
 $y_{r,P}(n)$: P 音に対応する r 次のくし形フィルタの出力系列

$|H_{r,P}(e^{j\omega})|$: $H_{r,P}(z)$ の利得

線形補間に関する記号

T : オーバサンプリグ周波数 f_s によるサンプル点間隔
 t_P : 理想的な f_{sP} によるサンプル時刻
 Δt_P : 理想的なサンプル点と実際のサンプル点の時間差
 $x'(t_P)$: 線形補間法により近似されたサンプル点

音高推定に関する記号

$s_{m,k}$: P_m 音の k 次高調波の振幅
 α_m : $|H_{r,P_m}(e^{j\omega})|$ の零点付近の勾配
 Δf_{P_m} : P_m 音のピッチのずれ

音価推定に関する記号

n_s : 目的とする波形の開始を示すサンプル点
 n_e : 目的とする波形の終了を示すサンプル点
 w : 音価検出のための単位ブロックに含まれるサンプル数
 W_i : 音価検出のための各単位ブロック
 R : 継続ブロック数
 R_{min} : 最小継続ブロック数

テンプレート構造に関する記号

$S_{M,Pq}(n,k)$: サンプル時点 n における楽器 M の P^q 音の k 次高調波スペクトル
 $\hat{T}_{M,Pq}(k)$: $S_{M,Pq}(n,k)$ の平均値
 $T_{M,Pq}(k)$: $\hat{T}_{M,Pq}(k)$ の正規化された値. テンプレートスペクトル
 $T_{M,i}(k)$: グループ化されたテンプレートスペクトル. i はグループ番号

和音分離に関する記号

$G_{r,P}(z)$: 入力音から P 音を分離抽出するための合成フィルタ
 $|G_{r,P}(e^{j\omega})|$: $G_{r,P}(z)$ の利得

DFT によるスペクトル解析に関する記号

$u_P(n)$: 抽出音 P
 N_P : P^3 音 1 周期に相当するサンプル点数. DFT の窓関数として使用
 $\hat{U}_P(e^{j\omega_k})$: 抽出音 P のスペクトル
 $U_P(e^{j\omega_k})$: 振幅補正及び正規化された $\hat{U}_P(e^{j\omega_k})$

テンプレートマッチングによる楽器推定に関する記号

$E_{M,P}$: P 音のスペクトルとテンプレートスペクトルとの絶対値差分

くし形フィルタによる和音分離の問題点に関する記号

L : 協和音における整数倍に近い周波数間隔
 $f_{P,P+d}$: P 音と $P+d$ 音の基本周波数を整数倍で表現し得る最大の周波数
 $N_{P,P+d}$: $f_{P,P+d}$ の 1 周期分に相当するサンプル点数. DFT の窓関数として使用
 f_i, a_i : 疑似的な周波数とその振幅
 $a_{M,Pq}$: 楽器 M の P^q 音の音量 (振幅)
 $\theta_{M,Pq}$: 楽器 M の P^q 音の合成時の位相
 $S_{Pq}(n,k)$: 時刻 t におけるユニゾン和音 (P^q 音) の k 次高調波スペクトル

目 次

2.1 演奏 (musical performance) と採譜 (musical transcription) の関係	7
2.2 五線譜形式による音楽表現	8
2.3 奏譜形式による音楽表現	8
2.4 代表的な楽器の音域	9
2.5 楽器音の性質	10
2.6 各楽器音のスペクトル構造の違い	13
2.7 音域によるスペクトル構造の違い	13
2.8 各楽器音の時間軸上の振幅変化	14
2.9 各楽器音の波形	15
2.10 和音のスペクトル	16
2.11 DFT の時間分解能と周波数分解能の関係	18
2.12 オクターブ 4 の周波数の周波数配置	18
2.13 一般的な適応システム	19
2.14 ノッチフィルタの周波数特性	21
2.15 E-NFT による採譜システム	22
3.1 くし形フィルタのブロック図	26
3.2 8 次くし形フィルタの周波数特性 ($r = 8$)	26
3.3 くし形フィルタを利用した周波数成分除去の原理	27
3.4 単音入力に対する音名判別システム	28
3.5 オクターブ推定の原理	28
3.6 オーバサンプリング法による $H_{r,p}(z_p)$ の近似的実現	30
3.7 オーバサンプリングによる零点のずれ	30
3.8 サンプル点の線形補間	32
3.9 $H_{8,p}(z)$ の縦続接続によるシステム構成	33
3.10 基本的な音名判別アルゴリズム	34
3.11 P , P_{+7} , P_{+4} 音の周波数配置	35
3.12 周波数の重複を利用した音名判別システム	36
3.13 周波数の重複を利用した音名判別アルゴリズム 1	38

3.14 周波数の重複を利用した音名判別アルゴリズム 2	39
3.15 周波数の重複を利用した音名判別アルゴリズム 3	40
3.16 周波数のずれ Δf_{m+1} に対する $H_{r,p}(z)$ の特性	42
3.17 入力信号 piano C^3	44
3.18 $H_{8,C}(z)$ の出力	44
3.19 $H_{8,D}(z)$ の出力	44
3.20 図 3.17 の入力に対する音価検出	45
3.21 音価推定原理	46
4.1 ピアノ音 C^4 の STFT 結果	47
4.2 和音分離システム	50
4.3 抽出音 C^4 の STFT 結果	53
4.4 抽出音 C^4 のスペクトル	54
4.5 周波数成分が独立な和音の例	56
4.6 周波数成分が重複しない 2 音の分離の例	57
4.7 高調波成分の一部が重複する和音の例	58
4.8 近接成分 f_1, f_2 によるビート特性	59
4.9 オクターブ和音の周波数成分の重複例	60
4.10 基本周波数が他音の高調波成分に重複する和音の例	60
4.11 基本周波数が他音に重複する和音の分離	61
4.12 ユニゾン和音の周波数成分の重複例	62
5.1 音量の構成パラメータ	65
5.2 システム概観図	66
5.3 piano C^3 音の音名判別（零出力検出）結果	68
5.4 piano C^3 音のオクターブ推定結果	68
5.5 violin D^4 音の音名判別（零出力検出）結果	69
5.6 flute F^5 音の音名判別（零出力検出）結果	70
5.7 flute F^5 の音価（音の継続時間）推定結果	71
5.8 violin D^4 の音価（音の継続時間）推定結果	72
5.9 piano C^3 の音価（音の継続時間）推定結果	72
5.10 violin F^4+G^4 に対する音名判別（零出力）検出結果	73
5.11 flute $C^4+E^4+G^4$ に対する音名判別（零出力）検出結果	77

5.12 和音 (a) violin $F^4 + F\sharp^4$ の音価（音の継続時間）推定結果	78
5.13 和音 (b) piano $C^3 + F^4$ の音価（音の継続時間）推定結果	79
5.14 和音 (c) flute $C^4 + E^4 + G^4$ の音価（音の継続時間）推定結果	79
5.15 和音 (1) の分離結果	81
5.16 重複成分 $3k \times f_{4,C} + 2k \times f_{4,G}$ の振幅変化（STFT 結果），及び位相変化	82
5.17 和音 (3) の分離結果	83
5.18 (5) の分離結果	85
5.19 単音の採譜対象楽譜「カエルの歌」	86
5.20 「カエルの歌」を piano で演奏した場合の出力波形	86
5.21 図 5.20 の音高，音価推定結果	87
5.22 「カエルの歌」を violin で演奏した場合の出力波形	88
5.23 図 5.22 の音高，音価推定結果	88
5.24 和音の採譜対象楽譜「カエルの歌」の輪唱	89
5.25 「カエルの歌」の輪唱を piano で演奏した場合の出力波形	89
5.26 「カエルの歌」の輪唱を violin で演奏した場合の出力波形	89
5.27 図 5.25 の音高，音価推定結果	90
5.28 図 5.26 の音高，音価推定結果	90
5.29 異種複数楽器の採譜対象楽譜「カエルの歌」の輪唱	91
5.30 図 5.29 の演奏波形	91
5.31 図 5.30 の音高，音価推定結果	92
A.1 piano template spectrum 1 $C^3 \sim D^3$	1
A.2 piano template spectrum 2 $D\sharp^3 \sim G^3$	1
A.3 piano template spectrum 3 $G^3 \sim C^4$	1
A.4 piano template spectrum 4 $D\sharp^4 \sim D^4$	2
A.5 piano template spectrum 5 $A\sharp^4 \sim F\sharp^4$	2
A.6 piano template spectrum 6 $D^5 \sim B^4$	2
A.7 piano template spectrum 7 $A\sharp^5 \sim G^5$	2
A.8 violin template spectrum 1 C^3	3
A.9 violin template spectrum 2 $C\sharp^3 \sim G^3$	3
A.10 violin template spectrum 3 $G\sharp^3 \sim D^4$	3
A.11 violin template spectrum 4 $D\sharp^4 \sim A^4$	3

A.12 violin template spectrum 5 $A\#^4 \sim F^5$	4
A.13 violin template spectrum 6 $F\#^5 \sim B^5$	4
A.14 viola template spectrum 1 $C^3 \sim C\#^3$	4
A.15 viola template spectrum 2 $D^3 \sim A^3$	4
A.16 viola template spectrum 3 $A\#^3 \sim G^4$	5
A.17 viola template spectrum 4 $G\#^4 \sim B^5$	5
A.18 flute template spectrum 1 $C^3 \sim C\#^4$	5
A.19 flute template spectrum 2 D^4	5
A.20 flute template spectrum 3 $D\#^4 \sim F\#^4$	6
A.21 flute template spectrum 4 $G^4 \sim A\#^4$	6
A.22 flute template spectrum 5 $B^4 \sim G^5$	6
A.23 flute template spectrum 6 $G\#^5 \sim B^5$	6
A.24 clarinet template spectrum 1 $C^3 \sim D^3$	7
A.25 clarinet template spectrum 2 $D\#^3 \sim G^3$	7
A.26 clarinet template spectrum 3 $G\#^3 \sim C^4$	7
A.27 clarinet template spectrum 4 $C\#^4 \sim G^4$	7
A.28 clarinet template spectrum 5 $G\#^4 \sim C^5$	8
A.29 clarinet template spectrum 6 $C\#^5 \sim G^5$	8
A.30 clarinet template spectrum 7 $G\#^5 \sim B^5$	8
A.31 horn template spectrum 1 $C^3 \sim C\#^3$	8
A.32 horn template spectrum 2 $D^3 \sim F\#^3$	9
A.33 horn template spectrum 3 $G^3 \sim A\#^3$	9
A.34 horn template spectrum 4 $B^3 \sim C\#^4$	9
A.35 horn template spectrum 5 $D^4 \sim A^4$	9
A.36 horn template spectrum 6 $A\#^4 \sim B^5$	10
A.37 trumpet template spectrum 1 $C^3 \sim C\#^3$	10
A.38 trumpet template spectrum 2 $D^3 \sim G\#^3$	10
A.39 trumpet template spectrum 3 $A^3 \sim C\#^4$	10
A.40 trumpet template spectrum 4 $D^4 \sim B^4$	11
A.41 trumpet template spectrum 5 $C^5 \sim B^5$	11

表 目 次

2.1 主な楽器の種類	9
2.2 純正律音階における各音階の周波数比（主音 C ）	11
2.3 国際標準周波数による平均律音階 [Hz]	12
3.1 採譜対象音域の音高周波数 [Hz]	25
3.2 $f_s = 54.054$ [kHz] により近似されたサンプリング周波数とその誤差	31
3.3 各 $H_{8,P}(z)$ で除去可能な音高	36
3.4 各 $H_{8,P}(z)$ で除去可能な音高	37
4.1 各楽器音の n_s/f_s	48
4.2 グループ化したテンプレートスペクトルに含まれる音高 P^q	49
4.3 各音に対応した N_P と周波数分解能 f_s/N_P	51
4.4 基本周波数が重複する 2 音の間隔 d と周波数比	55
4.5 高調波のみが重複する 2 音の間隔 k と周波数比	55
5.1 piano C^3 音の音名判別（零出力検出）結果	67
5.2 piano C^3 音のオクターブ推定結果	67
5.3 violin D^4 音の音名判別（零出力検出）結果	69
5.4 violin D^4 音のオクターブ推定結果	69
5.5 flute F^5 音の音名判別（零出力検出）結果	70
5.6 flute F^5 音のオクターブ推定結果	70
5.7 violin $F^4 + G^4$ 音の音高判別結果	74
5.8 和音 (a) violin $F^4 + F\#^4$ 音の音高判別結果	75
5.9 和音 (b) piano $C^3 + F^4$ 音の音高判別結果	76
5.10 和音 (c) flute $C^4 + E^4 + G^4$ 音の音高判別結果	76
5.11 和音 (1) のスペクトル差分値	82
5.12 和音 (2) のスペクトル差分値	82
5.13 和音 (3) のスペクトル差分値	83
5.14 和音 (4) のスペクトル差分値	83
5.15 和音 (5) のスペクトル差分値	85

