5R-2

ICOTone — 音楽情報認識ユニットninoru —

吉田 実, 下山 健, 小池 汎平, 田中 英彦
東京大学*

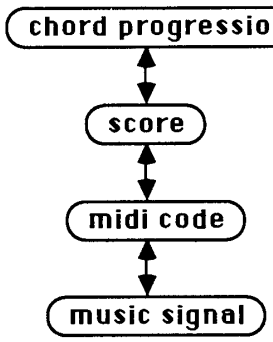

図 1: the Music Information Hierarchy

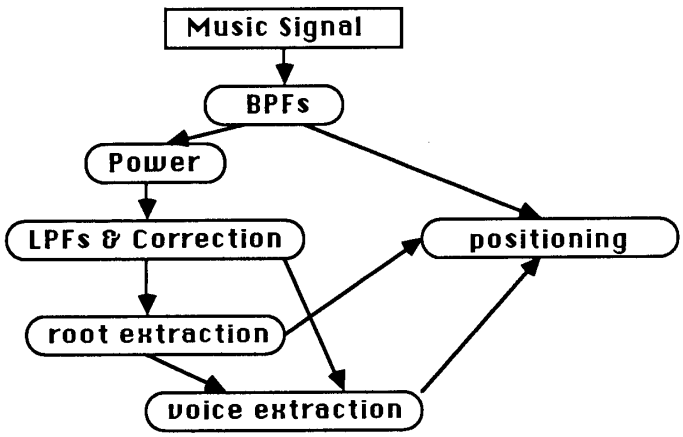

図 2: the Structure of ninoru

1 はじめに

われわれは、新世代音楽システム ICOTone[1] の開発をすすめている。 ICOTone では、現段階では、図1のように音楽をコード進行、楽譜、MIDI コード、音楽信号の各階層でとらえ、その相互間の変換をはかっている。 ninoru は、そのうちで最も下位の音楽信号レベルから MIDI コードへの変換を行うユニットである。従来、この部分は AckII[2] が行なってきたが、処理の精度と可能性の拡大をねらってって、新方式を導入した。 ninoru は、人間の耳の処理方法をまねていて、抽出する周波数ごとに最適な時間分解能と周波数分解能がえられ、また、抽出したデータは安定している。例えば、半音違いで音がなっているような時でも、十分抽出できることがわかった。基音の抽出だけでなく、楽器の決定、定位の決定への応用の可能性もある。

2 ninoru の構成

ninoru では、音楽信号は、中心周波数が非線形に並べられた帯域の狭いバンドパスフィルタ群(BPF 群)を通る。その出力は、正弦波に近い波形が得られる。その後、その出力の 2 乗をとってローパスフィルタ群 (LPF 群) を通す。さらに、BPF 群の特性を補正する処理を行い、各キーごとのパワーが得られる。これに、n($n \geq 1$) オクターブ下の音からは抑制性の、n オクターブ上からの音からは興奮性の働きをえる基音抽出層をへて基音の結果をえる。楽器の決定は、基音抽出の結果とキーごとのパワーを使う。定位抽出は、さらに、BPF 群の結果も使う。その構成は、図2のようになっている。

*ICOTone - the music signal recognition unit "ninoru", Yoshida Minoru, Tanaka Hidehiko, University of Tokyo

3 BPF 群をつかうことのメリット

FFT を使うのにくらべ、1) サンプルをとった時の位相の状態による影響をほとんど受けない、2) 周波数分解能が一定でなく、自由に設定できる、3) 時間分解能が各抽出周波数での最適値になる、などがあり、MEM にくらべると、上の 3) の他に、4) 高調波に強い、ということがいえる。また、サンプルした場所により変な値を出さない、ということが強みであり、安定した音楽信号の認識を可能とする。

4 BPF 群

BPF は、2 次の IIR でも実現が可能である。1 オクターブに12 × k個つける。ただし、kは半音当りのフィルタの個数である。今回の目的のためには、k = 1で十分である。しかし、周波数の連続な変化を追跡するためにはk = 2 ～ 4程度、さらに、弦楽器の周波数のゆらぎを検出するためには、$k \geq 4$以上必要であろう。次に、 BPF のバンド幅であるが、あまりゆるい特性だと周波数分解能が低くなってしまうが、きつくしすぎると時間分解能がさがり、また、わずかに周波数がずれただけで、抽出できなくなる。今回は、半音もしくは 1/4 音でカットオフになるようにしている。

5 LPF 群と BPF 群の補正

BPF 群の出力の 2 乗をとって、それに LPF 群をかけると各キーのパワーがでてくる。このパワーは、BPF 群の特性の影響をうけていいるので、その特性の逆行列を使って、補正

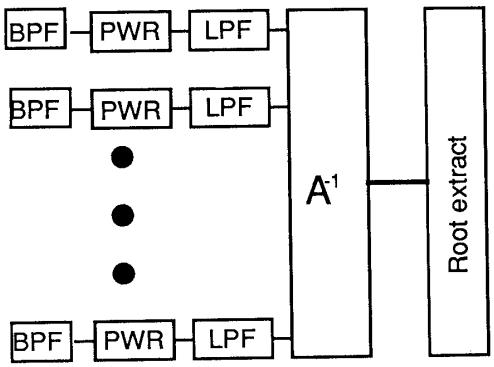

図 3: Processing for the Root Extraction

する。LPF 群は BPF 群の出力が、その周波数の正弦波なので、周波数ごとに特性を変えることができ、最適な時間分解能がえられる。

6 基音抽出

BPF 群の補正の出力は、まず、各時間ごとに上のオクターブの音からは興奮性の影響を受け、下のオクターブの音からは抑制性の影響を受ける。その後、時間軸方向に非線形フィルタをとおして、その出力が基音とみなされる。ここまでの処理は、ninoru の基本的な部分で図3のようである。

7 楽器抽出

基音が決定されたら、その結果と BPFs の補正結果を使って、楽器を抽出する。抽出の方法としては、パワーの時間変化、倍音構造とその時間変化、周波数の揺らぎを使ったものを計画している。

8 定位

人間は、音楽を主として 2 つの耳できく。その場合、定位はかなり重要な要素である。モノラルできくと、マスキング効果などできこえない音も、ステレオならきこえるということはよくある。また、楽器などの決定においても、定位の情報があると、精度を高めることができる。人間の耳は左右の耳の位相差はわかるといわれる。 ninoru では、基音を見つけたら、BPF 群の出力がほぼ正弦波なのを利用して、その音の左右の時間差を見ることによって、定位を調べることができる。

9 計算量

ninoru は、ニューラルネットもしくは並列マシン上での走らせることを仮定しているので、並列度が高ければ計算量は少々多くても構わないともいえるが、とりあえずは普通のハードウェア上で走らすことになるので、FFT を使った方法とくらべてそれほど変わらない計算量ですむことについて言及したい。FFT の場合は、例えば、1024 点なら乗算回数は$1024 \times \log 1024 \times c$である。 ninoru では、 n オクターブで半音あたり k 個の BPF をならべ、BPF が 2 次、LPF が 1 次の時$n \times 12 \times k \times 3$である。$n = 3, k = 1$とすると、108 回 / サンプルであるが、図4のように 1 オクターブ下がるごとに、サンプルを 1/2 に間引くなどという方法も使える。

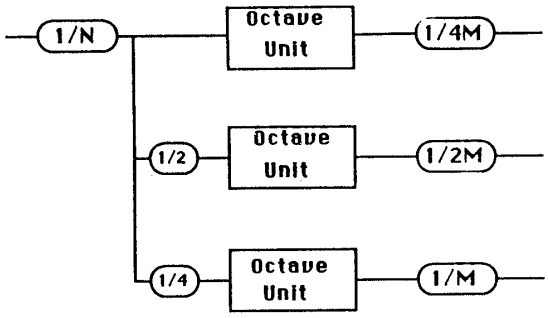

図 4: a structure of ninoru using octave units

10 まとめ

BPF を並べる方法は、時間分解能・周波数分解能の最適な安定した結果が得られた。今後は、後段の処理を充実させて、使えるものにしていきたい。

参考文献

[1] 平田他、“新世代音楽システム ICOTone の全貌”、第 33 回情処全大 5N-5
[2] 阿久津他、“ICOTone on PSI - AckII”、第 35 回情処全大 5Ff-8