JAIST Repository
https://dspace.jaist.ac.jp/

Title	振幅変調情報を用いた楽器音の音源同定に関する研究
Author(s)	川平，勇
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1331
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

修 士 論 文

振幅変調情報を用いた
楽器音の音源同定に関する研究

指導教官　赤木正人　教授

北陸先端科学技術大学院大学
情報科学研究科　情報処理学専攻

川平 勇

2000 年 2 月 15 日

Copyright © 2000 by Isamu Kawahira

目 次

1 序論 **1**

1.1 はじめに 1

1.2 研究の背景 1

1.3 本研究の目的 2

1.4 本論文の構成 2

2 音源同定アルゴリズムの概要 **3**

3 変調—搬送周波数マップ検出モデル **5**

3.1 目的 5

3.2 変調 – 搬送周波数マップ検出モデルの概要 5

3.2.1 フィルタバンク 6

3.2.2 フィルタバンクの式 8

4 マップ検出のシミュレーション **11**

4.1 目的 11

4.2 分析する楽器音 11

4.3 変調 – 搬送周波数マップの検出結果 11

4.4 考察 27

5 マッチング **29**

5.1 目的 29

5.2 全体マッチング 29

5.2.1 目的 29

5.2.2 マッチング方法 29

5.2.3 マッチング結果 30

5.2.4 考察 40
5.3 部分マッチング 40
5.3.1 目的 40
5.3.2 マッチング方法 40
5.3.3 マッチング結果 42
5.3.4 考察 46

6 結論 **47**
6.1 本論文で明らかになったことの要約 47
6.2 今後の課題 47

謝辞 **48**

参考文献 **49**

図 目 次

2.1 全体の流れ 4

3.1 変調 – 搬送周波数マップ（左）とその例（右） 6
3.2 全体の処理の概要図 7
3.3 Gabor 関数による BPF（上：時間軸, 下：周波数軸） 9

4.1 Cello の波形（上：C♯1 音, 中：E1 音, 下：G1 音） 13
4.2 Cello の波形（上：A♯1 音, 中：C♯2 音, 下：E2 音） 13
4.3 Cello の波形（上：G2 音, 中：突発音, 下：足し合わせた音） 14
4.4 Cello の波形（上：A♯2 音, 中：突発音, 下：足し合わせた音） 14
4.5 Cello の波形（上：C♯3 音, 中：突発音, 下：足し合わせた音） 15
4.6 Violin の波形（上：G2 音, 中：突発音, 下：足し合わせた音） 15
4.7 Violin の波形（上：A♯2 音, 中：突発音, 下：足し合わせた音） 16
4.8 Violin の波形（上：C♯3 音, 中：突発音, 下：足し合わせた音） 16
4.9 Cello C♯1 音の変調 – 搬送周波数マップの検出結果 17
4.10 Cello E1 音の変調 – 搬送周波数マップの検出結果 17
4.11 Cello G1 音の変調 – 搬送周波数マップの検出結果 18
4.12 Cello A♯1 音の変調 – 搬送周波数マップの検出結果 18
4.13 Cello C♯2 音の変調 – 搬送周波数マップの検出結果 19
4.14 Cello E2 音の変調 – 搬送周波数マップの検出結果 19
4.15 Cello G2 音の変調 – 搬送周波数マップの検出結果 20
4.16 Cello A♯2 音の変調 – 搬送周波数マップの検出結果 20
4.17 Cello C♯3 音の変調 – 搬送周波数マップの検出結果 21
4.18 Violin G2 音の変調 – 搬送周波数マップの検出結果 21
4.19 Violin A♯2 音の変調 – 搬送周波数マップの検出結果 22
4.20 Violin C♯3 音の変調 – 搬送周波数マップの検出結果 23

4.21 突発音の変調 – 搬送周波数マップの検出結果 23
4.22 Cello G2 音 + 突発音の変調 – 搬送周波数マップの検出結果 24
4.23 Cello A♯2 音 + 突発音の変調 – 搬送周波数マップの検出結果 24
4.24 Cello C♯3 音 + 突発音の変調 – 搬送周波数マップの検出結果 25
4.25 Violin G2 音 + 突発音の変調 – 搬送周波数マップの検出結果 25
4.26 Violin A♯2 音 + 突発音の変調 – 搬送周波数マップの検出結果 26
4.27 Violin C♯3 音 + 突発音の変調 – 搬送周波数マップの検出結果 26

5.1 全体マッチングでのマッチング領域 30
5.2 Cello C♯1 音と各音階との類似度の時間変化 31
5.3 Cello E1 音と各音階との類似度の時間変化 31
5.4 Cello G1 音と各音階との類似度の時間変化 32
5.5 Cello A♯1 音と各音階との類似度の時間変化 32
5.6 Cello C♯2 音と各音階との類似度の時間変化 33
5.7 Cello E2 音と各音階との類似度の時間変化 33
5.8 Cello G2 音と各音階との類似度の時間変化 34
5.9 Cello A♯2 音と各音階との類似度の時間変化 34
5.10 Cello C♯3 音と各音階との類似度の時間変化 35
5.11 Violin G2 音と各音階との類似度の時間変化 35
5.12 Violin A♯2 音と各音階との類似度の時間変化 36
5.13 Violin C♯3 音と各音階との類似度の時間変化 36
5.14 Cello G2 音 + 突発音と各音階との類似度の時間変化 37
5.15 Cello A♯2 音 + 突発音と各音階との類似度の時間変化 37
5.16 Cello C♯3 音 + 突発音と各音階との類似度の時間変化 38
5.17 Violin G2 音 + 突発音と各音階との類似度の時間変化 38
5.18 Violin A♯2 音 + 突発音と各音階との類似度の時間変化 39
5.19 Violin C♯3 音 + 突発音と各音階との類似度の時間変化 39
5.20 部分マッチングでのマッチング領域 41
5.21 Cello G2 音 + 突発音 $\Longleftrightarrow$ Cello G2 音の各領域の類似度の時間変化 42
5.22 Cello A♯2 音 + 突発音 $\Longleftrightarrow$ Cello A♯2 音の各領域の類似度の時間変化 . . . 43
5.23 Cello C♯3 音 + 突発音 $\Longleftrightarrow$ Cello C♯3 音の各領域の類似度の時間変化 . . . 43
5.24 Violin G2 音 + 突発音 $\Longleftrightarrow$ Violin G2 音の各領域の類似度の時間変化 . . . 44
5.25 Violin A♯2 音 + 突発音 $\Longleftrightarrow$ Violin A♯2 音の各領域の類似度の時間変化 . . 44

5.26 Violin C♯3 音 + 突発音 ⟺ Violin C♯3 音の各領域の類似度の時間変化 . . 45

表 目 次

4.1 解析した楽器音の種類と音階 12
4.2 解析した楽器音の種類と音階およびその基本周波数と FB2 側から見たチャネル番号 28

第 1 章

序論

1.1 はじめに

計算機による音認識は, 雑音が含まれていない状況では, ほぼ実用レベルに達している. しかし, 認識される音響信号に雑音や他の音声等の複数の音が混在するような場合においては認識率は著しく低下する.

対照的に人間は, カクテルパーティ効果 [1] と呼ばれる複数の音源から到来する音の中から特定の音源の音だけを抽出して聞くことができる機能を持っている. この機能が実現できれば, 雑音で汚れていないきれいな音だけを取り出すことが可能となり, これを前処理とすることで雑音に強い音認識システムや, 複数種類の楽器演奏を対象とする自動採譜システムなどへ応用することができると期待される.

また, 実環境の多くの音においては各周波数成分の振幅が時間と共に変化する. このような振幅変調（AM）は聴覚情報処理において極めて重要な役割を果たしている [2]. このような様々な音における振幅変調を検出すれば音源の識別や音認識等における重要な情報と成り得るのではないかと考えられる.

1.2 研究の背景

音源分離問題を聴覚の情景解析（Auditory Scene Analysis：ASA）[3] の問題としてとらえ直し, それを計算論の立場から議論する研究がある. 代表的なモデルとしては, Brown と Cooke ら [4] による音響事象に基づいた分離モデル, Ellis [5] による発見的規則を階層的に実装した分凝モデル, 中谷ら [6] による聴覚の情景解析をマルチエージェントシステムにより実現した分凝モデルがある. また柏野ら [7, 8] による 2 つの周波数成分の分離知

覚に関して, スペクトログラム上の複数の特徴と分離知覚の生じる割合との定量的関係をモデル化している. しかし, これらのモデルのほとんどは音響的な特徴量として振幅スペクトルを用いている. そのため同じ周波数成分を含むような場合, 完全に分離できているとはいえない.

三浦 [9] は, 聴覚における音の分離の一要因と考えられる振幅変調マスキング現象に着目した音源分離モデルを提案している. このモデルは時間変動する振幅包絡である変調周波数の解析に主眼を置いている.

これら音源分離の研究は聴覚の情景解析問題の半分しか取り扱っていない. 残りの半分の問題とは, 音源同定のことであり, これは各音響事象の名前やラベルを認識することを意味する. 音源分離と比較すると音源同定を扱った論文の数は限られている. 音源を同定するシステムとして, Kinoshita ら [10] の研究があり, OPTIMA と呼ばれるシステムを提案している. しかしこのシステムも周波数成分が重複し, 物理的特徴が影響し合っていても, 重複していない場合と同じように実行するので, 重複成分を持つ音源を同定することができない.

1.3 本研究の目的

本研究では, 単一の楽器音と突発音を対象とし, 時間と共に変化する振幅情報に着目して, 変調 – 搬送周波数マップ検出モデルを用いて楽器音の特徴量を検出する. そして検出したマップ, その楽器名および音階をデータベースとして持っておき, 楽器音と突発音が足し合わされた音響信号のマップとマッチングを行なうことによって楽器音の音階と楽器名を同定する.

1.4 本論文の構成

本論文の構成を以下に示す.

第 1 章では, 過去の研究の現状と問題点を指摘し, 本論文の目的を明らかにする.

第 2 章では, 音源同定アルゴリズムについて述べる.

第 3 章では, 変調 – 搬送周波数マップ検出モデルの説明を行なう.

第 4 章では, 楽器音のマップ検出シミュレーションに結果について述べる.

第 5 章では, マッチングについて述べる.

第 6 章では, 本論文で得られた結果を要約し, 今後の課題を示す.

第 2 章

音源同定アルゴリズムの概要

図 2.1 に全体の処理の大まかな流れを示す.

1. 変調 – 搬送周波数マップ検出モデルにより, 特徴量を得る（Analysis）
2. 基本周波数が存在する場所を探索し, 音階を決定する（F0 Estimation）
3. その音階の楽器のマップと全体マッチングを行なう（Matching 1）
4. 類似度がある値以上ならその楽器音であると決定. 未満なら他の音あるいは突発音が入っていると仮定
5. 部分マッチングを行ない, 類似度がある値以上のものが多いものをその楽器音と決定する（Matching 2）
6. 残りの類似度が低い部分を DB のものと入れ換える（Replacement）
7. 音源分離の操作へともっていく

本論文では上記の 1 〜 6 を取り扱う.

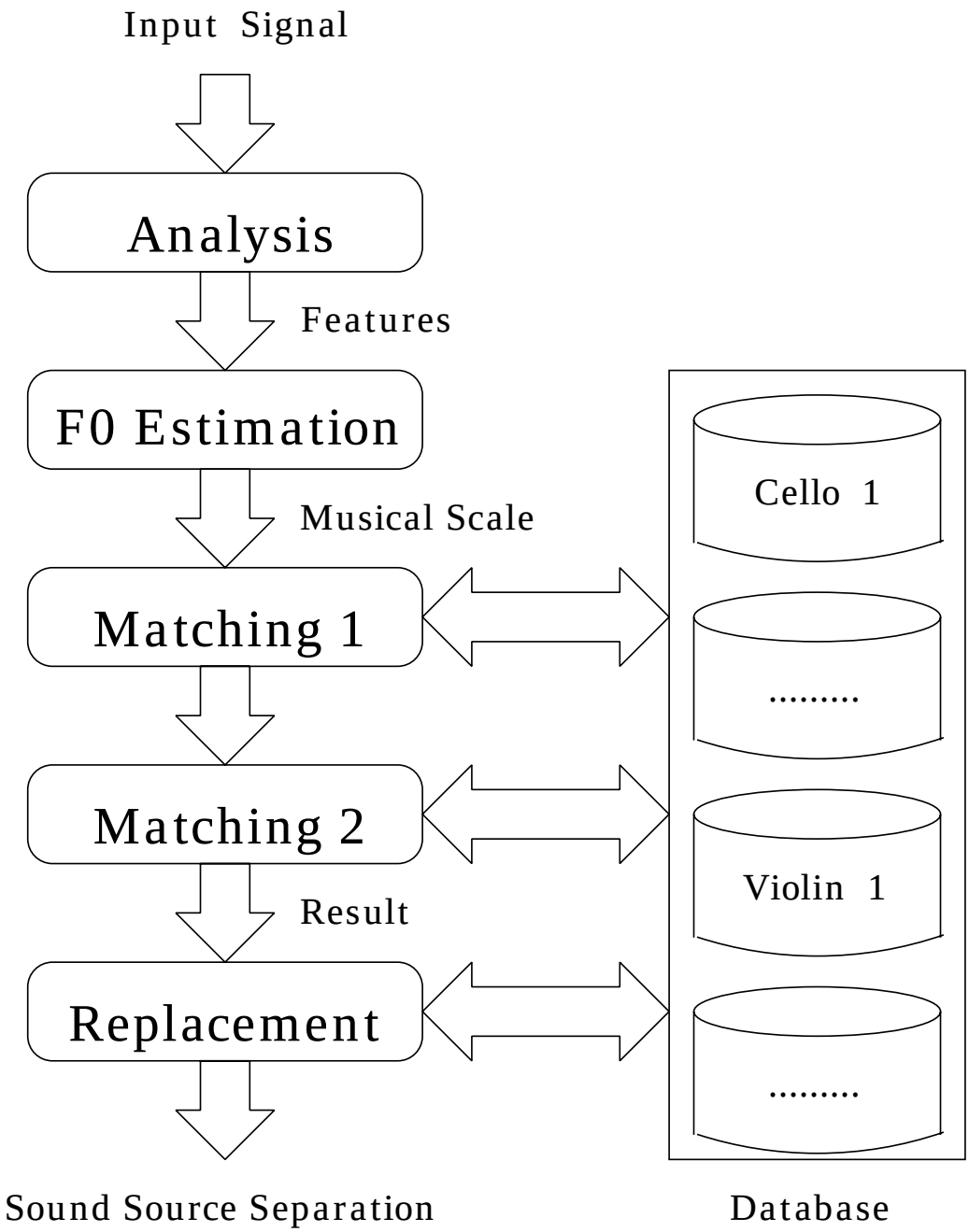
Input Signal
Analysis
Features
F0 Estimation
Musical Scale
Matching 1
Matching 2
Result
Replacement
Cello 1
.....
Violin 1
.....
Sound Source Separation
Database

図 2.1: 全体の流れ

第 3 章

変調—搬送周波数マップ検出モデル

3.1　目的

本研究では, 特徴量として変調—搬送周波数マップを取り扱う. そのため本節ではマップ検出モデルの概要について述べる.

3.2　変調 – 搬送周波数マップ検出モデルの概要

変調 – 搬送周波数マップ

図 3.1（左）に変調 – 搬送周波数マップを示す. これによって例えば図 3.1（右）に示されているような, 互いの搬送周波数に重なりがあるような 2 つの信号を考える. このとき, 雑音と信号成分のそれぞれが同じような変調周波数で変調されていれば, 求めたい信号成分とそれ以外の成分の変調周波数の差を見て, 求めたい信号の周波数成分をグルーピングし, 抽出することで音源分離が行えるのではないかと考える.

変調 – 搬送周波数検出モデル

先の考え方を基に図 3.2に構築されている変調—搬送周波数検出モデルを示す. ここで FILTER BANK 1 は, 搬送周波数を分けるためにある帯域幅を持つ BPF 群である. 同様に FILTER BANK 2 は変調周波数を分けるためにある帯域幅を持つ BPF 群である.

ここでの DC 成分とは LPF を使用しているため, 実際に求まっているのは DC 成分と非常に低い変調周波数成分になっている. これによって時間変化する振幅包絡より DC 成分が求められる.

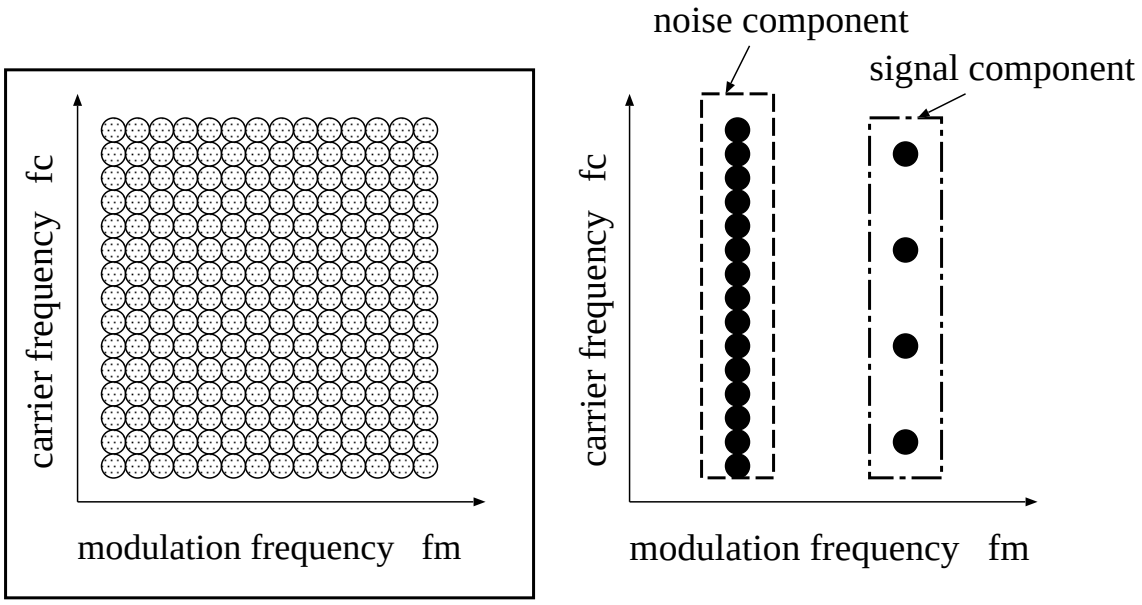

図 3.1: 変調 – 搬送周波数マップ（左）とその例（右）

- 1.　入力信号を各搬送周波数 f_c に分けるために FILTER BANK 1 に通す.
- 2. 各搬送周波数に対する振幅包絡（envelope）を求めるために FILTER BANK 1 の出力より瞬時振幅を求める.
- 3. 求めた瞬時振幅をダウンサンプリングする
- – 4-1. 各搬送周波数成分より LPF を用いて DC 成分を求める
  – 4-2. 各搬送周波数成分を FILTER BANK 2 に通して各搬送周波数成分の変調周波数成分 f_m を求める.

この処理過程 3 でのダウンサンプリングは, 処理過程 1 で FILTER BANK 1 を通過する変調周波数成分が FILTER BANK 1 の BPF の帯域幅までの周波数成分しか持たないということや処理時間等を考慮して行っている.

3.2.1　フィルタバンク

ここではフィルタリング処理における BPF について説明する. 搬送周波数のフィルタリングには FILTER BANK 1, 変調周波数のフィルタリングには FILTER BANK 2 の通過帯域や中心周波数などが問題となってくる. ここでは

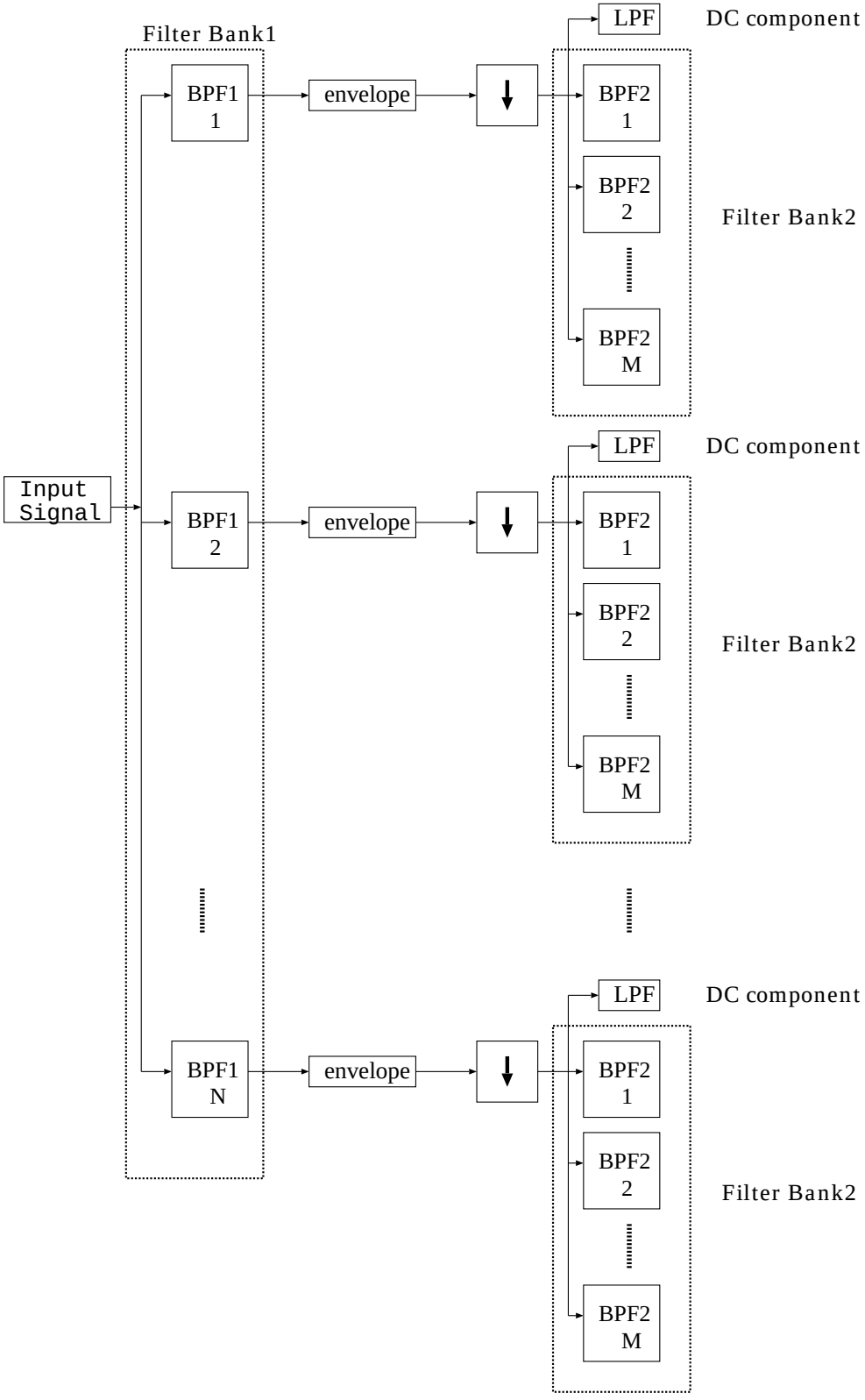

図 3.2: 全体の処理の概要図

- 定Qフィルタバンク

のタイプのフィルタバンクを考える. 以下にその詳細について述べる.

定Qフィルタバンク

音声などの搬送周波数成分は主として比較的低い周波数に主要な情報が多い. そこで低い中心周波数ほどより詳細に解析でき, かつ復元を考える上でも有益と考えられる定Qフィルタバンク考えている.

定Qフィルタバンクで構築された FILTER BANK 2（変調周波数）における解析範囲は FILTER BANK 1（搬送周波数）の帯域幅の半分となる. ここで FILTER BANK 1 の帯域幅はある程度の幅がないと FILTER BANK 2 の解析範囲が狭くなり, 変調周波数の十分な解析が行えない. しかし, 帯域幅が広すぎると搬送周波数におけるフィルタリング結果が, ぼやけてしまい, 音源の分離が困難になってしまう. 以上のことをまとめると

- FILTER BANK 1（搬送周波数）: BPF の中心周波数と帯域幅は対数周波数で等間隔
- FILTER BANK 2（変調周波数）: BPF の中心周波数と帯域幅は, 対数周波数で等間隔で解析範囲は FILTER BANK 1 のそれぞれの BPF の帯域幅の半分

となり, FILTER BANK1 の BPF の中心周波数の増加に従い（対数周波数で等間隔）FILTER BANK 2 の解析範囲も増加する.

またここで定Qフィルタバンクの BPF 群は Wavelet 分析合成系で構築されている.

3.2.2 フィルタバンクの式

本節ではフィルタバンクに用いている Gabor 関数 [11] [12] [13] を用いた BPF の式について説明する. また定Qフィルタバンクではこれを基底関数とした wavelet 変換を用いた分析合成系を用いているのでこれについて説明する.

BPF の式

次にフィルタバンクで用いている BPF の式について述べる. ここで用いている BPF は周波数的にも時間的にも局在し, さらに瞬時振幅を求める上で都合のよいヒルベルト変換対を持つものを考えている. この条件に合う関数として周波数領域での不確定性と時間領域での不確定性を最小にする関数として知られている [12] Gabor 関数を基に BPF を

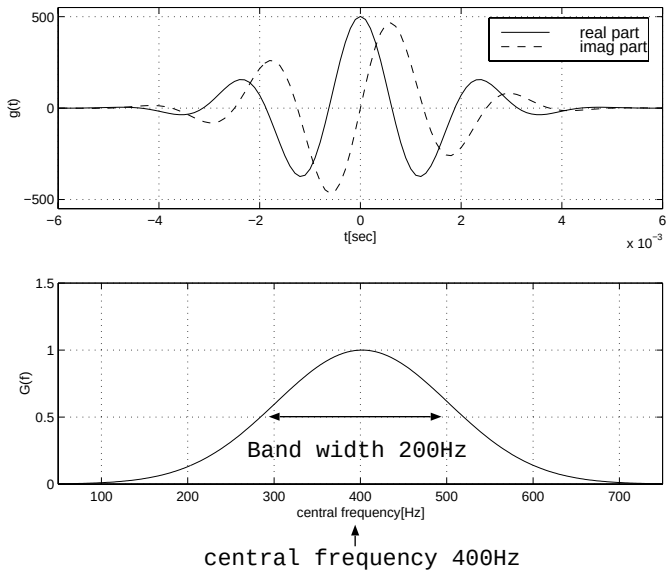

図 3.3: Gabor 関数による BPF（上：時間軸, 下：周波数軸）

形成している. 下記に Gabor 関数の式を示す. 上式が時間領域における式で下式が周波数領域における式である.

$$
\begin{aligned}
f(t) &= Aexp(-Bt^2) \cdot exp(-j\omega t) \\
G(\omega) &= A\sqrt{\frac{\pi}{B}} \exp(-\frac{\omega^2}{4B})
\end{aligned}
$$

ここで帯域幅は Gabor 関数の振幅が $1/\sqrt{2}$ の範囲を帯域幅としている. 図 3.3 に上式を基に作成した BPF の例を示す. ここでフィルタの帯域幅は 200Hz で, フィルタの中心周波数は 400Hz である. また図 3.3の上図は時間軸, 下図は周波数軸（振幅特性）における BPF を示している.

Gabor 関数を基底関数とした Wavelet 変換

従来の信号解析では, 短時間フーリエ変換（STFT）が一般的に用いられていた. これは, 中心周波数とその周波数帯域を動かしても, 時間周波数の窓はその幅が変化しないという意味で固定化されたものであった. また, これは時間と周波数の不確定性原理によって束縛されたものであり, STFT は非常に高いかあるいは低い周波数を持つ信号解析に対

して適切なものとは言えなかった. これに対して Wavelet 変換は, 時間解析窓が自動的に, 高周波数現象を調べるときに狭まり, 低周波数現象を調べるときに広くなるという性質を備えている. また, 定Qフィルタバンクによる分析を数学的に記述したものと考えられている. 本モデルでは人間の内耳における信号処理と Wavelet 変換に基づく信号処理に類似性 がある [14] ことや解析後の復元が容易であるという利点を基に Wavelet 変換を利用している.

Wavelet 分析合成系

本研究で作成した Wavelet 分析合成系について述べる. ここでフィルタバンク 1 は中心周波数 $f_0 = 500$[Hz], 通過帯域を 50-5000[Hz] , フィルタ数（チャネル数）は 64 となっている. また, フィルタバンク 2 は中心周波数 $f_0 = 30$[Hz], 通過帯域を 30-300[Hz], フィルタ数（チャネル）は 64 となっている. ここでフィルタバンク 2 の通過帯域の上限は先の前節 3.2.2 で述べたようにフィルタバンク 1 の BPF の帯域幅によって変化する.

第 4 章

マップ検出のシミュレーション

4.1 目的

本章では前章で述べた変調 – 搬送周波数マップ検出モデルを用いて楽器音のマップを得て, 各音の特徴を調べる.

4.2 分析する楽器音

実験データとして, 市販されているサンプリング CD の音響信号を用いた. 解析した楽器音の種類および音階を表 4.1 に示す.

すべての信号は CD 上では WAV 形式のファイルで収録されているので, 取り扱える形式に変換を行ない, サンプリング周波数を 44.1 kHz から 11 kHz へダウンサンプリングした. 次に Cello, Violin の信号については, 処理時間を考慮して 4 秒の長さの信号から 2 〜 3 秒間にわたる 1 秒間を切り出したものを入力に用いた. 図 4.1 〜 図 4.8 に解析した音響信号を示す. 図 4.3 〜 図 4.8 は各音階（上）, 突発音（中）および突発音が 0.25 秒から始まるように足し合わせた信号（下）である.

4.3 変調 – 搬送周波数マップの検出結果

以上の実験データを基に変調 – 搬送周波数マップの検出を行なった. ここで搬送周波数（f_c）, 変調周波数（f_m）の添字はそれぞれのフィルタバンクのチャネル番号を示している. ここでは, 1 が一番低い中心周波数の BPF（f_c：50 Hz, f_m：3 Hz）で 64 が一番高い中心周波数の BPF（f_c：5000 Hz, f_m：300 Hz）となっている.

表 4.1: 解析した楽器音の種類と音階

Cello	Violin
C♯1	–
E1	–
G1	–
A♯1	–
C♯2	–
E2	–
G2	G2
A♯2	A♯2
C♯3	C♯3

各音の検出結果

図 4.9 〜 図 4.21 に Cello, Violin の G2 音, A♯2 音, C♯3 音, および突発音の解析結果の FILTER BANK 1, 2 出力の時間平均を示す.

楽器音と突発音の混合信号の検出結果

図 4.22 〜図 4.27 に Cello, Violin の G2 音, A♯2 音, C♯3 音と, 突発音の混合信号の解析結果の FILTER BANK 1, 2 出力の時間平均を示す.

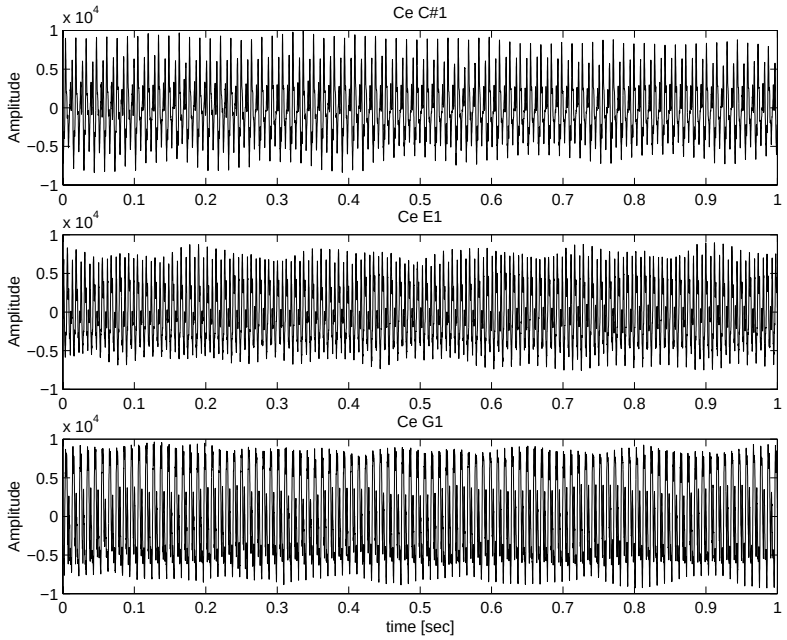

図 4.1: Cello の波形（上：C♯1 音, 中：E1 音, 下：G1 音）

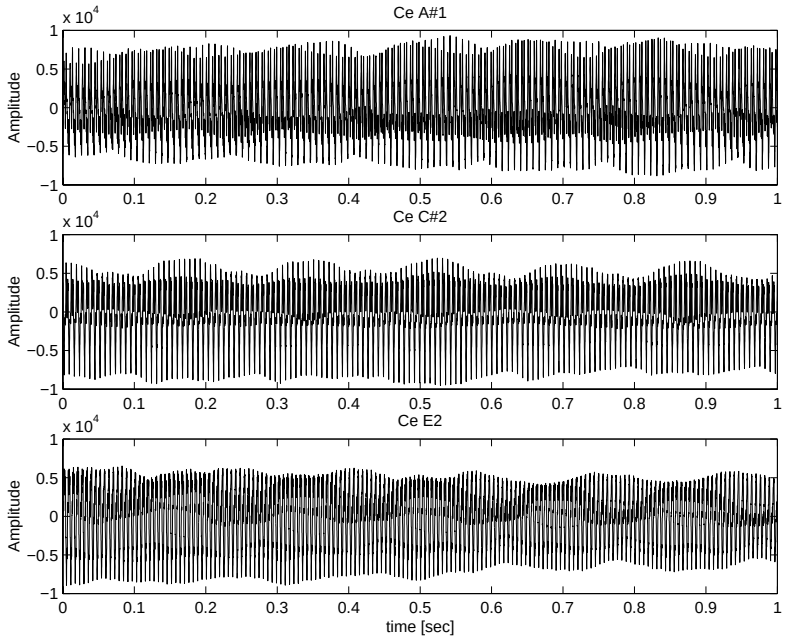

図 4.2: Cello の波形（上：A♯1 音, 中：C♯2 音, 下：E2 音）

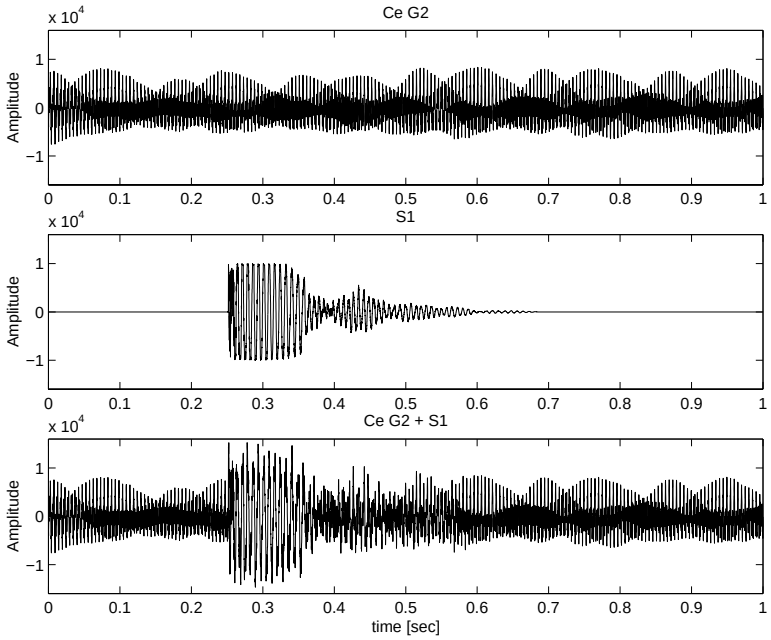

図 4.3: Cello の波形（上：G2 音, 中：突発音, 下：足し合わせた音）

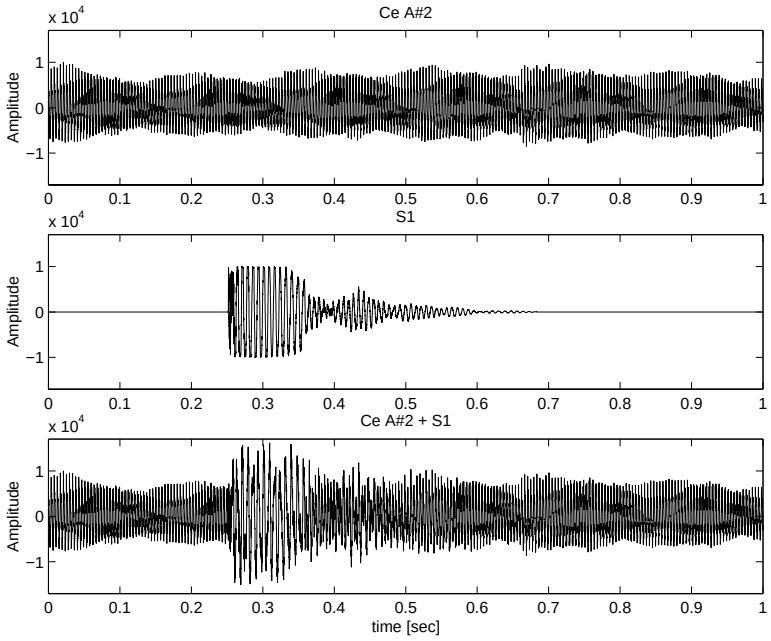

図 4.4: Cello の波形（上：A♯2 音, 中：突発音, 下：足し合わせた音）

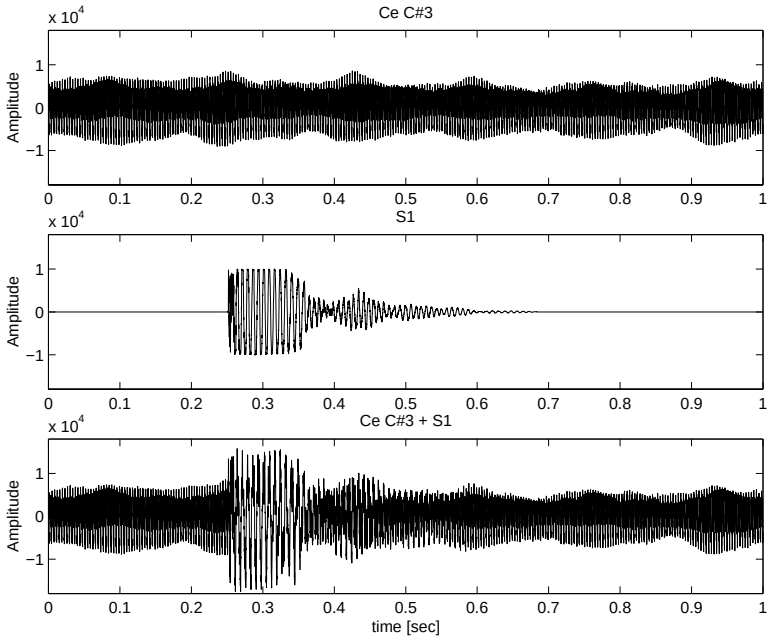

図 4.5: Cello の波形（上：C♯3 音, 中：突発音, 下：足し合わせた音）

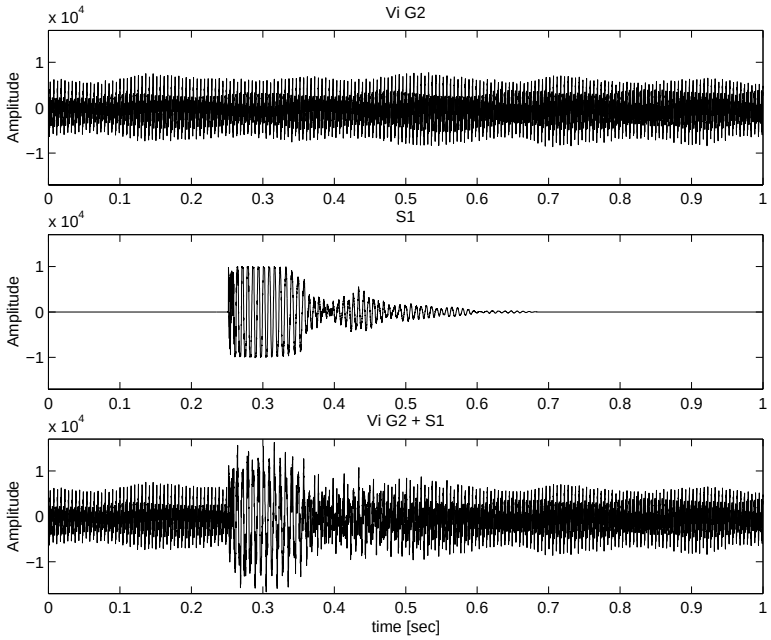

図 4.6: Violin の波形（上：G2 音, 中：突発音, 下：足し合わせた音）

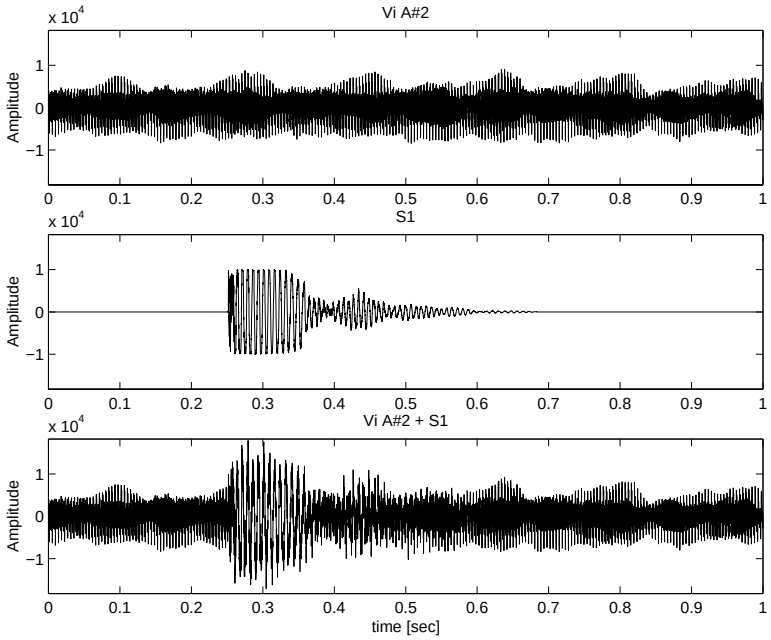

図 4.7: Violin の波形（上：A♯2 音, 中：突発音, 下：足し合わせた音）

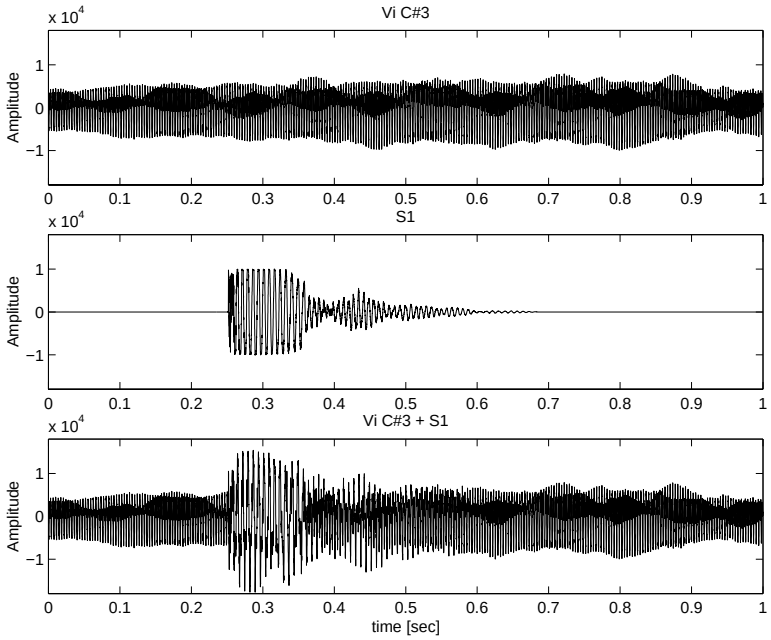

図 4.8: Violin の波形（上：C♯3 音, 中：突発音, 下：足し合わせた音）

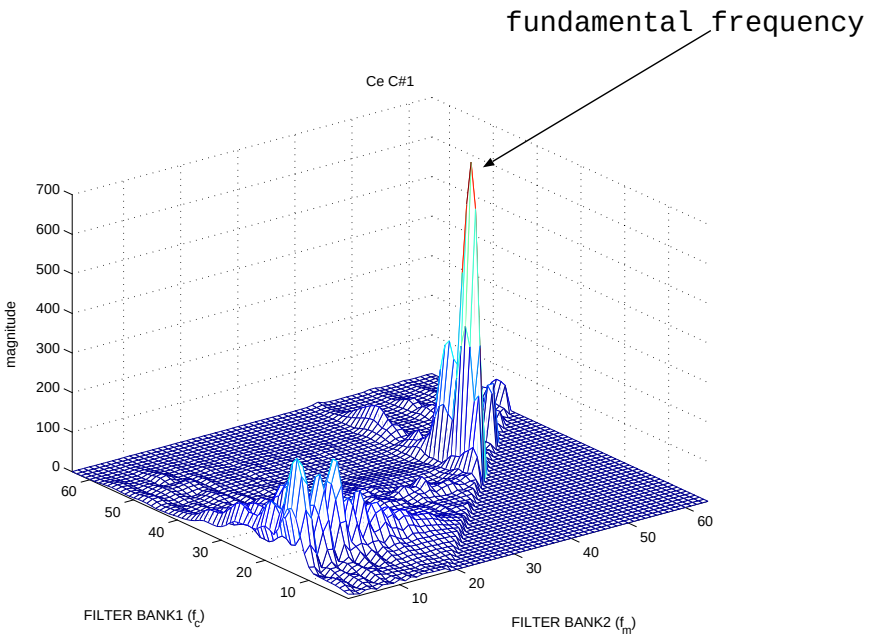

図 4.9: Cello C♯1 音の変調 – 搬送周波数マップの検出結果

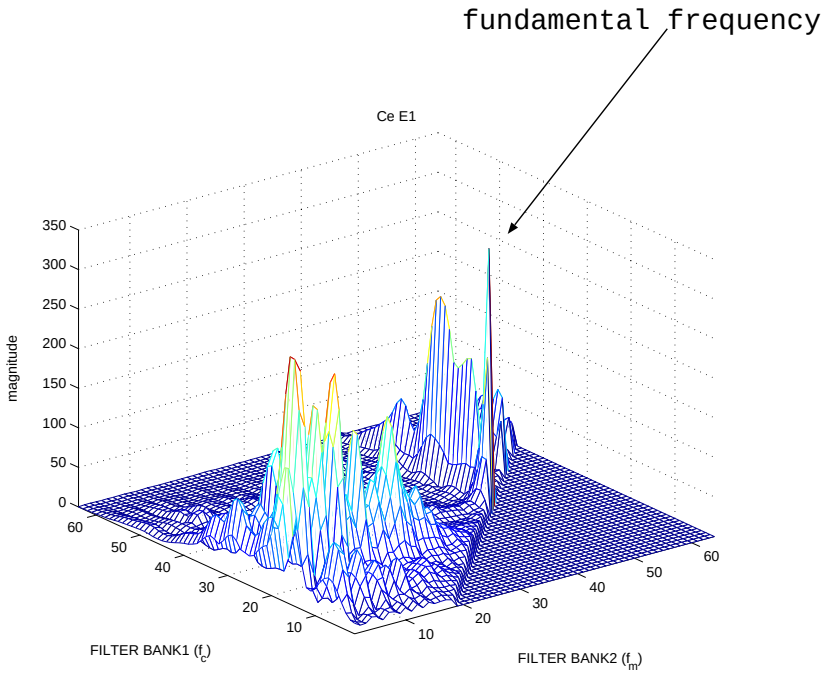

図 4.10: Cello E1 音の変調 – 搬送周波数マップの検出結果

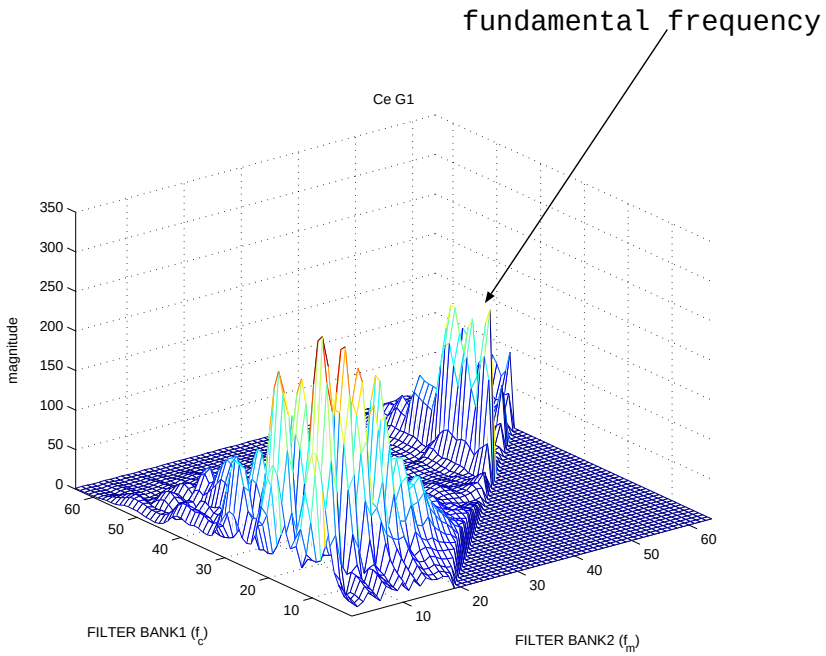

図 4.11: Cello G1 音の変調 – 搬送周波数マップの検出結果

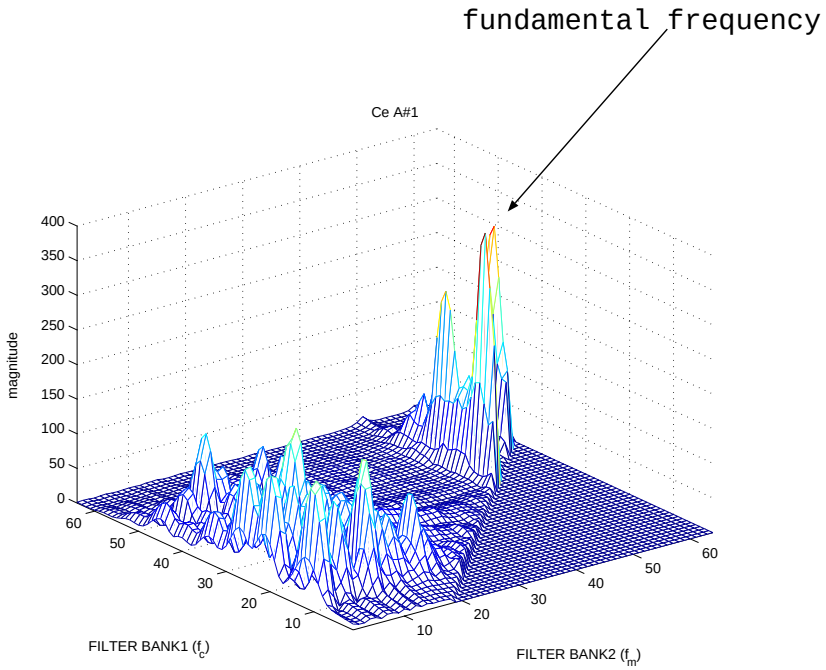

図 4.12: Cello A♯1 音の変調 – 搬送周波数マップの検出結果

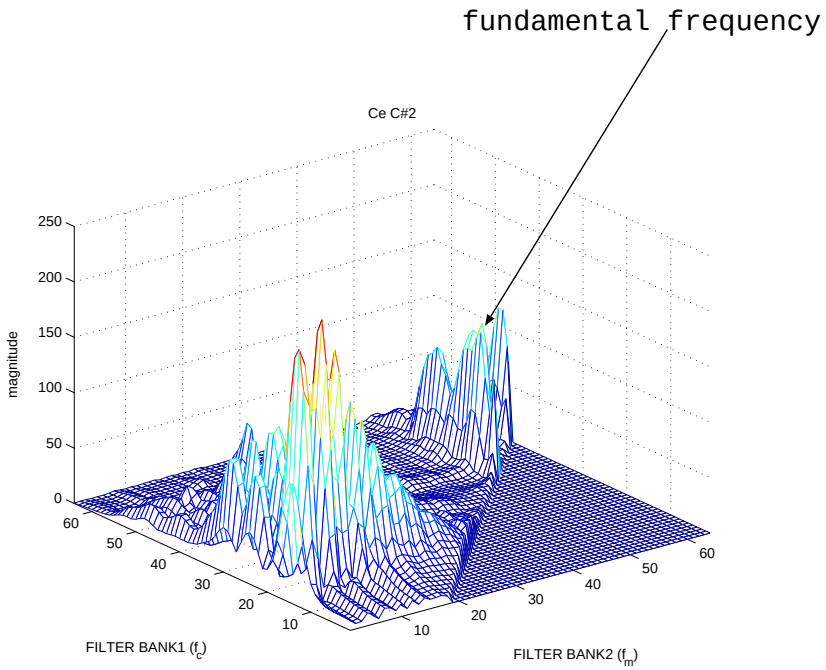

図 4.13: Cello C♯2 音の変調 – 搬送周波数マップの検出結果

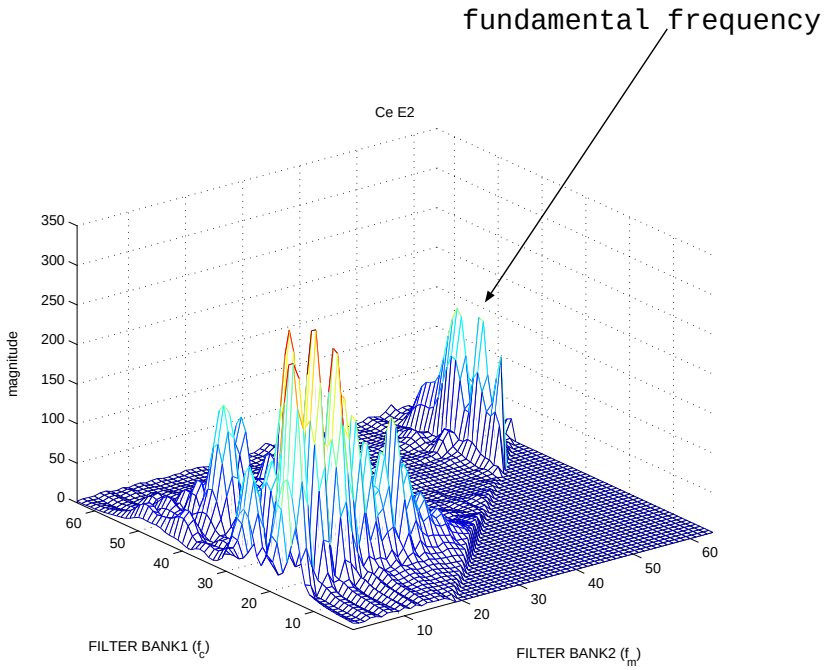

図 4.14: Cello E2 音の変調 – 搬送周波数マップの検出結果

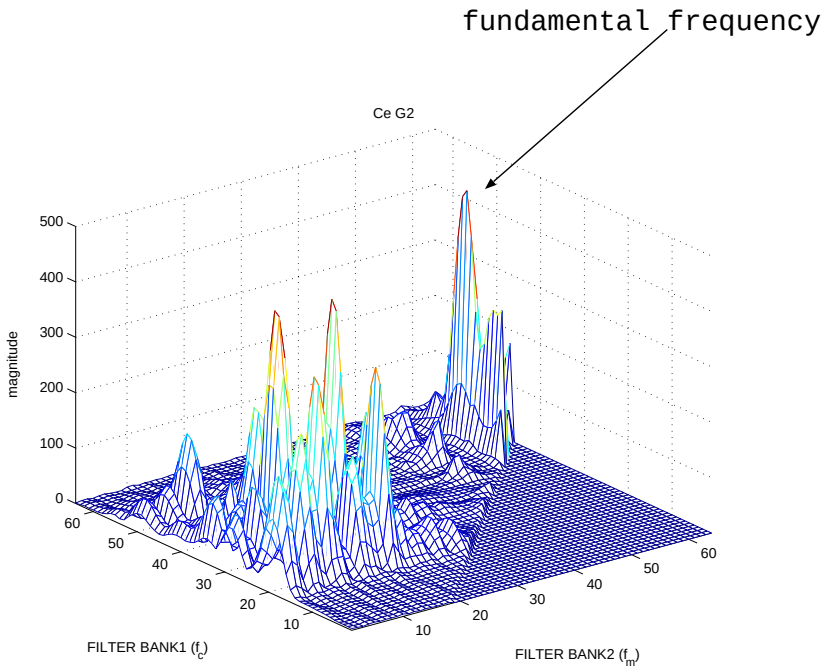

図 4.15: Cello G2 音の変調 – 搬送周波数マップの検出結果

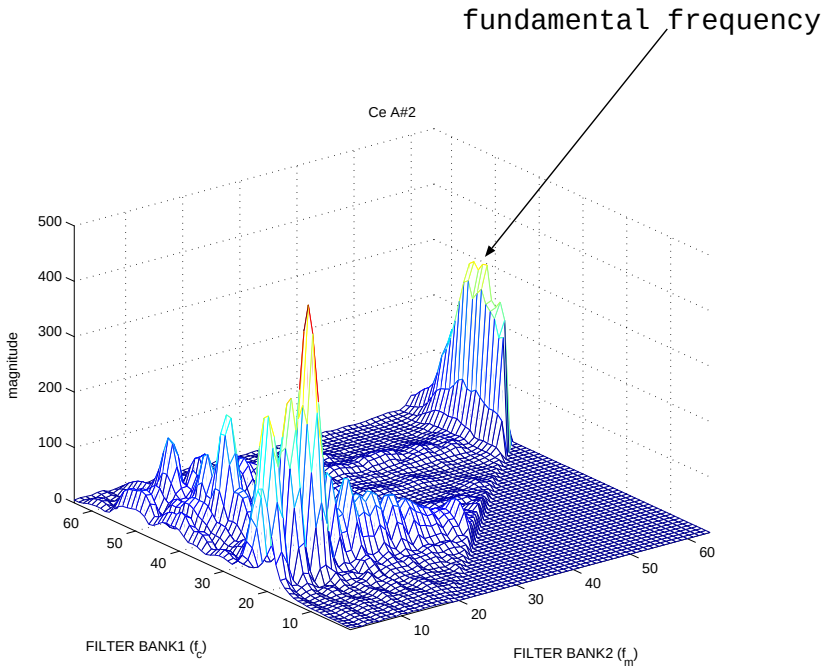

図 4.16: Cello A♯2 音の変調 – 搬送周波数マップの検出結果

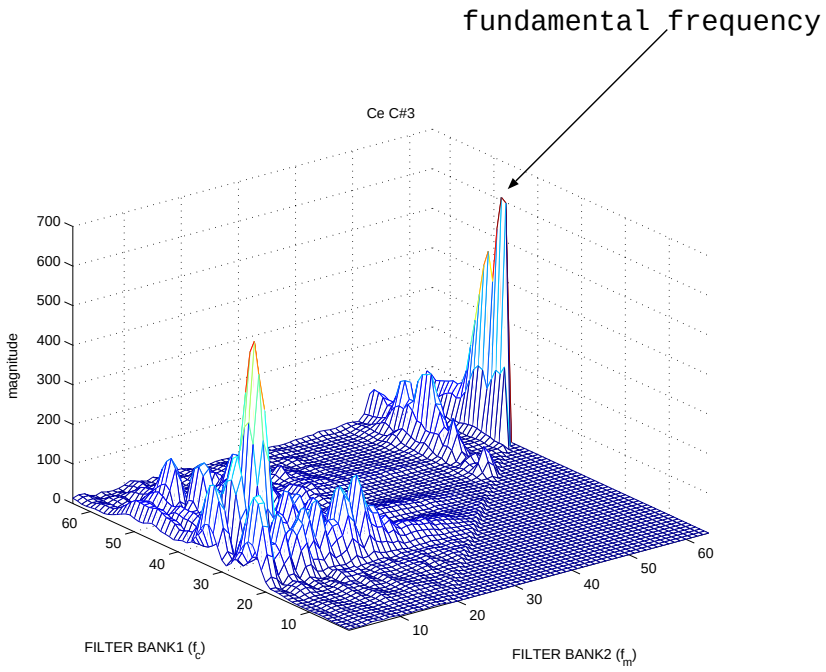

図 4.17: Cello C♯3 音の変調 – 搬送周波数マップの検出結果

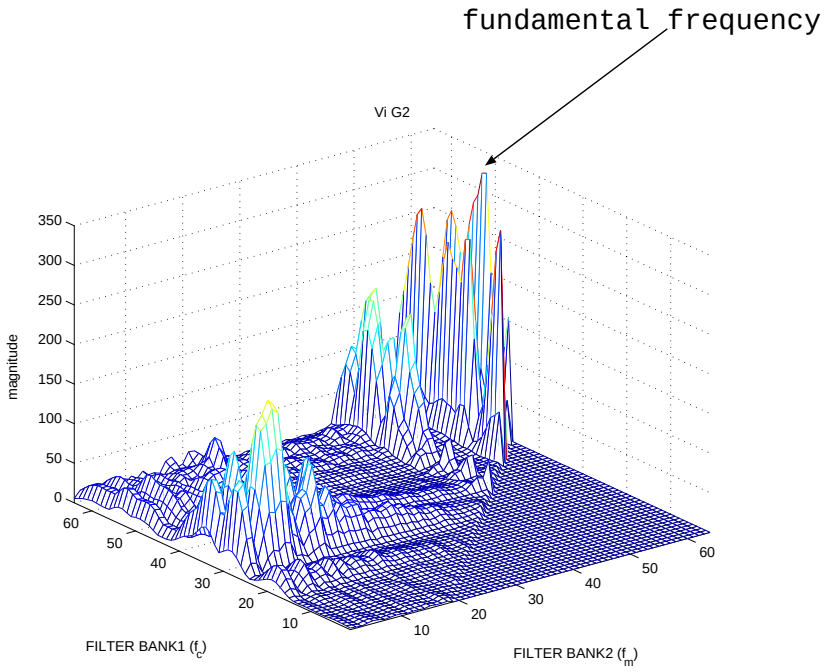

図 4.18: Violin G2 音の変調 – 搬送周波数マップの検出結果

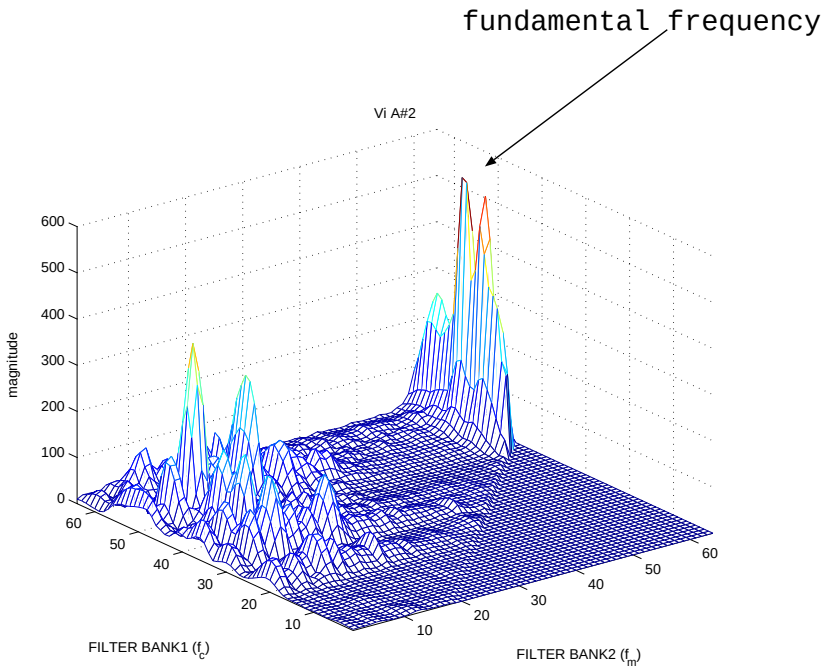

図 4.19: Violin A♯2 音の変調 – 搬送周波数マップの検出結果

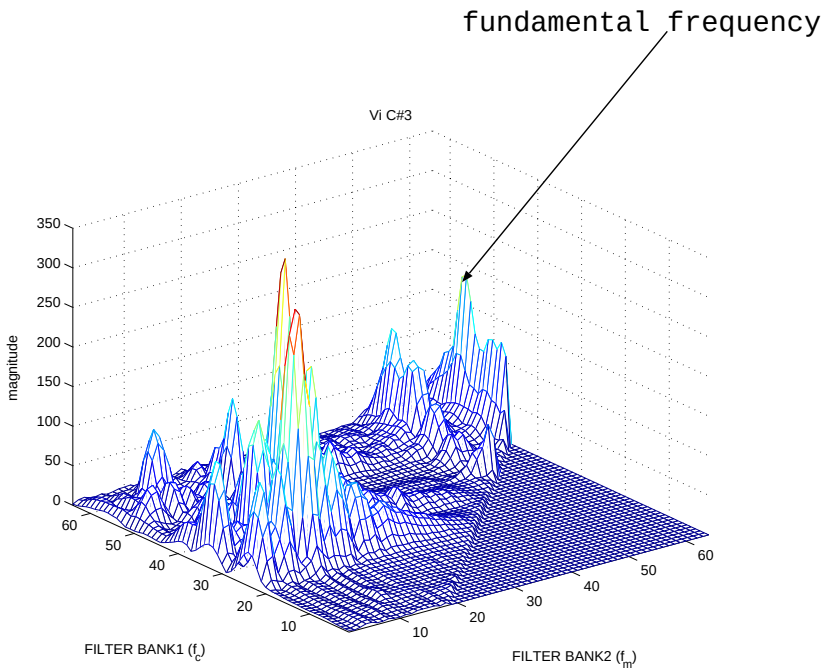

図 4.20: Violin C♯3 音の変調 – 搬送周波数マップの検出結果

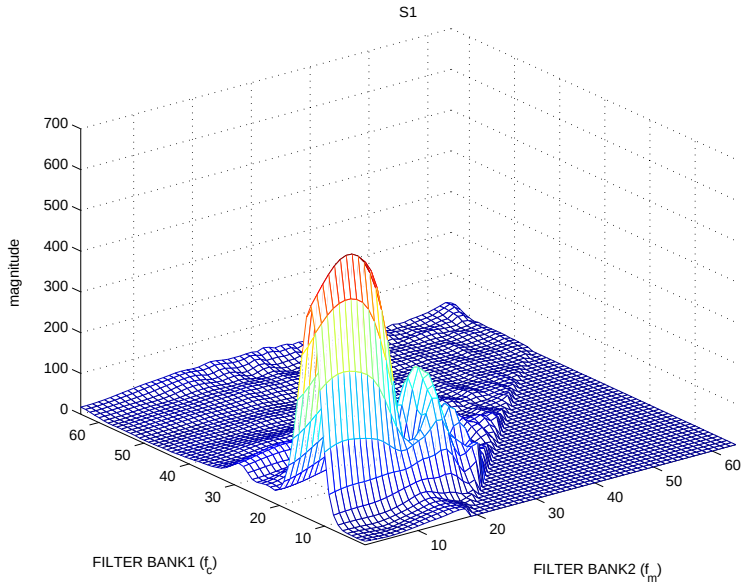

図 4.21: 突発音の変調 – 搬送周波数マップの検出結果

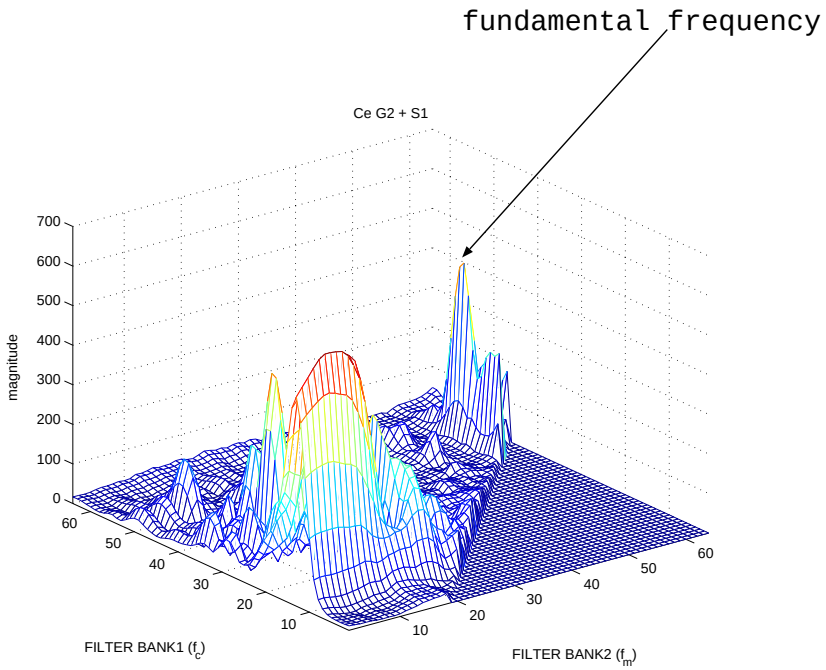

図 4.22: Cello G2 音 + 突発音の変調 – 搬送周波数マップの検出結果

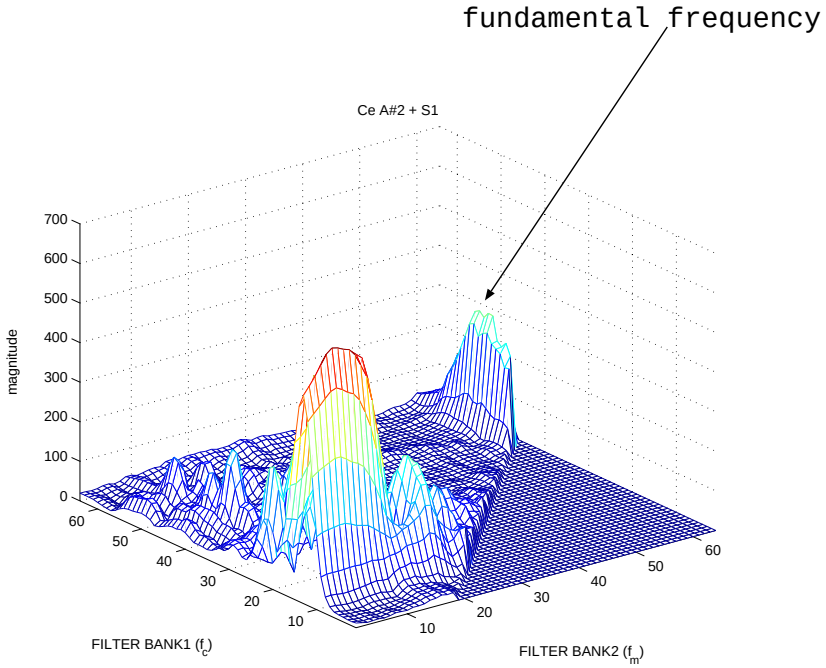

図 4.23: Cello A♯2 音 + 突発音の変調 – 搬送周波数マップの検出結果

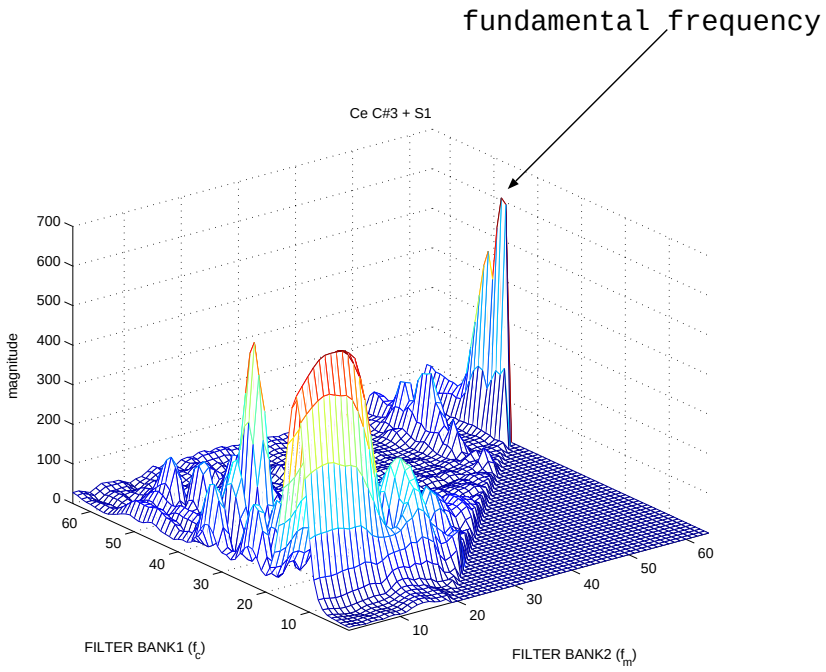

図 4.24: Cello C♯3 音 + 突発音の変調 – 搬送周波数マップの検出結果

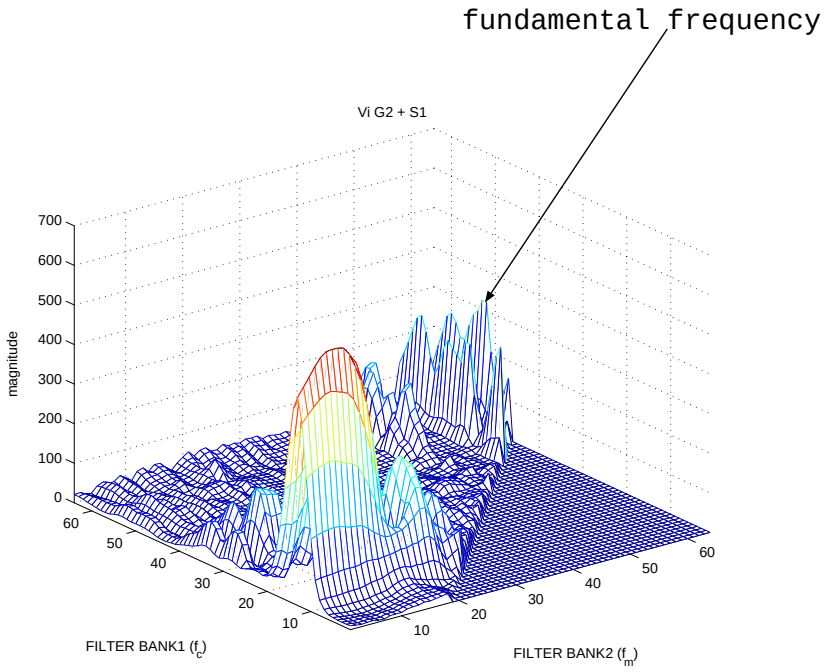

図 4.25: Violin G2 音 + 突発音の変調 – 搬送周波数マップの検出結果

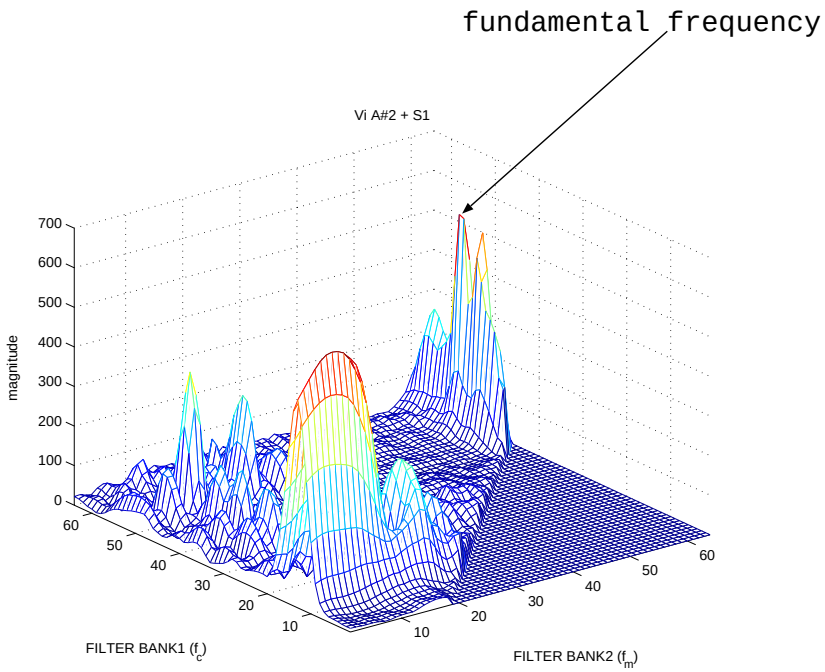

図 4.26: Violin A♯2 音 + 突発音の変調 – 搬送周波数マップの検出結果

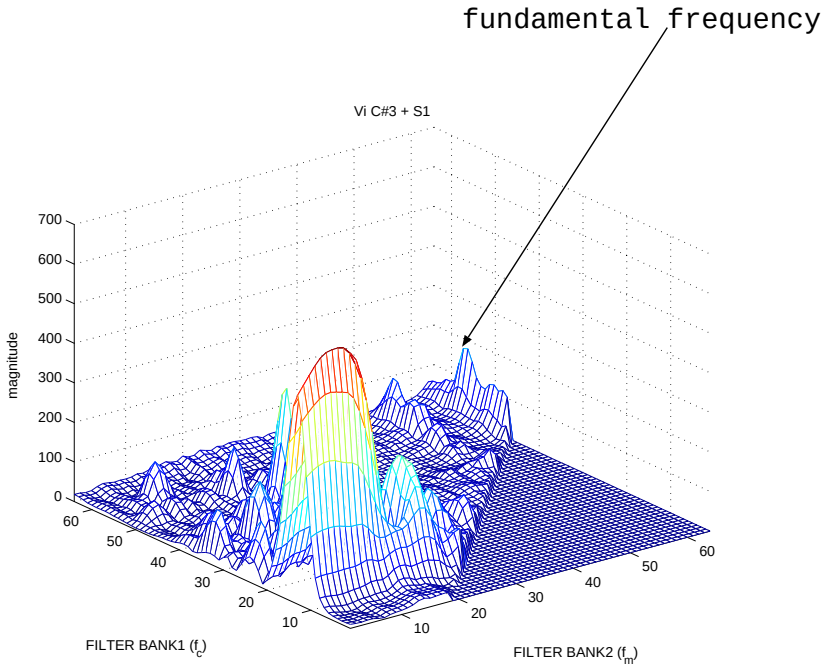

図 4.27: Violin C♯3 音 + 突発音の変調 – 搬送周波数マップの検出結果

4.4 考察

変調 – 搬送周波数マップの検出結果より, 変調周波数 f_m 軸の高い位置にそれぞれの音階の基本周波数成分が表れて, この場所は変動が激しい. 表 4.2 に各音階の基本周波数および中心周波数のチャネル番号を示す. また, 搬送周波数 f_c 軸の全域および変調周波数 f_m 軸の番号 1 から番号 40 にわたる領域（64×40 の領域）は, 基本周波数が表れている位置の変動と比較すると時間的変化がほとんどなく, 搬送周波数 f_m 軸側から見ると基本周波数のほぼ整数倍の値の中心周波数の部分に成分が表れている.

突発音を解析した時間平均のマップを見ると, f_c 軸および f_m 軸の低い部分に成分が表れているが, 高い周波数成分（f_m 軸：64 付近, f_c 軸：64 付近）の付近は, 時間軸上で突発音が表れると急激に成分が出現し, すぐに減衰していく.

したがって, 次のようにマップを用いることにする.

- f_c 軸の全域および f_m 軸の番号 1 〜 40

  ⇒ 時間平均をデータベースとして登録しておく領域

  　マッチングを行なう領域

- f_m 軸の番号 41 〜 64

  ⇒ 音階を同定する基本周波数を探索する領域

表 4.2: 解析した楽器音の種類と音階およびその基本周波数と FB2 側から見たチャネル番号

Cello	Violin	F0 [Hz]	FB2 のチャネル番号
C♯1	–	69.295	43 - 44
E1	–	82.405	46 - 47
G1	–	98.000	48 - 49
A♯1	–	116.54	50 - 51
C♯2	–	138.59	53 - 54
E2	–	164.81	55 - 56
G2	G2	196.00	58 - 59
A♯2	A♯2	233.08	60 - 61
C♯3	C♯3	277.18	62 - 63

第 5 章

マッチング

5.1 目的

本章では, 前章で得られた各音階の変調 – 搬送周波数マップを類似度法を用いてマッチングを行なう. マッチングは, 全体マッチングと部分マッチングの 2 種類を行なう.

5.2 全体マッチング

5.2.1 目的

各音階ごとに全体マッチングを行なって, どのくらいの類似度が得られるのかを確認し, 同定のための評価尺度を考察する.

5.2.2 マッチング方法

全体マッチングは, 前章で述べた 64×40 の領域の各点の大きさを, パターンベクトルの要素として取り扱い, そのベクトルを用いて類似度を計算する. 図 5.1 に全体マッチングの領域を示す.

類似度は, 以下に示す式で計算している.

$$s^{(l)} = \cos\theta^{(l)} = \frac{\left(\boldsymbol{x}, \boldsymbol{r}^{(l)}\right)}{\|\boldsymbol{x}\| \cdot \|\boldsymbol{r}^{(l)}\|} \qquad (0 \leq s^{(l)} \leq 1)$$

$\boldsymbol{x}$ ： 入力マップ

$\boldsymbol{r}^{(l)}$ ：データベースにあるマップ

$s^{(l)}$ ： 類似度 $(1 \leq l \leq 12)$

l は 1 〜 9 が Cello の各音階, 10 〜 12 が Violin の各音階を表している.

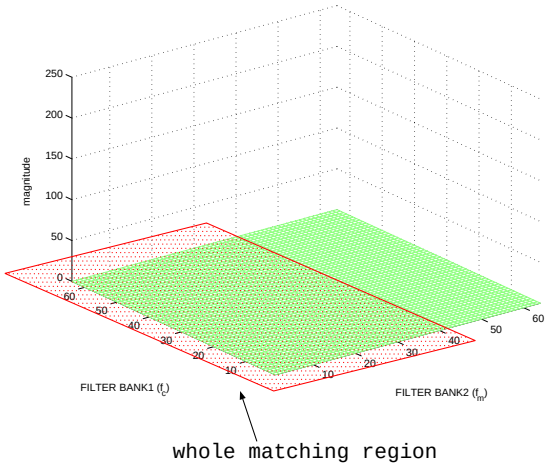

図 5.1: 全体マッチングでのマッチング領域

5.2.3 マッチング結果

個々の音階どうしのマッチング結果

図 5.2 〜 図 5.13 に, 個々の音階どうしの類似度の時間変化を示す.

突発音を足し合わせた場合のマッチング結果

図 5.14 〜 図 5.19 に突発音を足し合わせた場合の類似度の時間変化を示す.

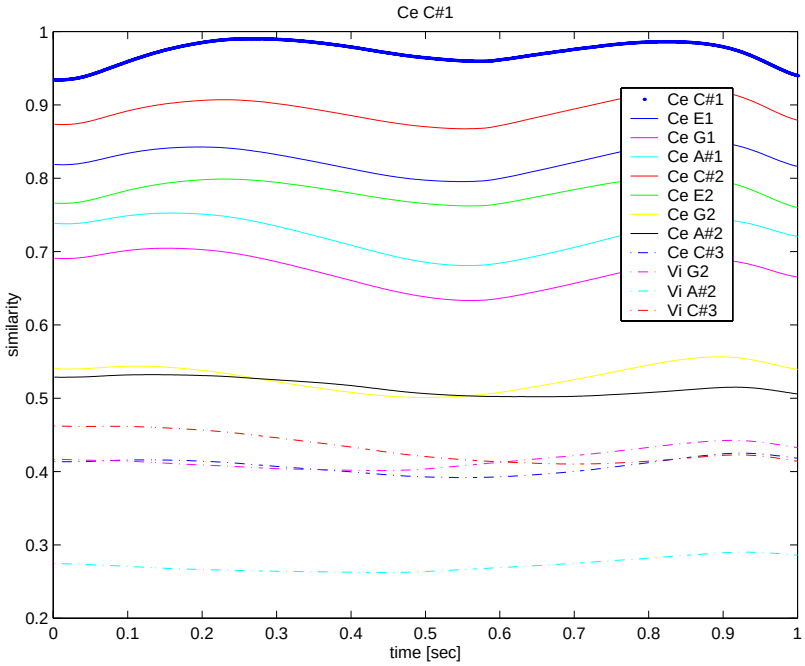

図 5.2: Cello C♯1 音と各音階との類似度の時間変化

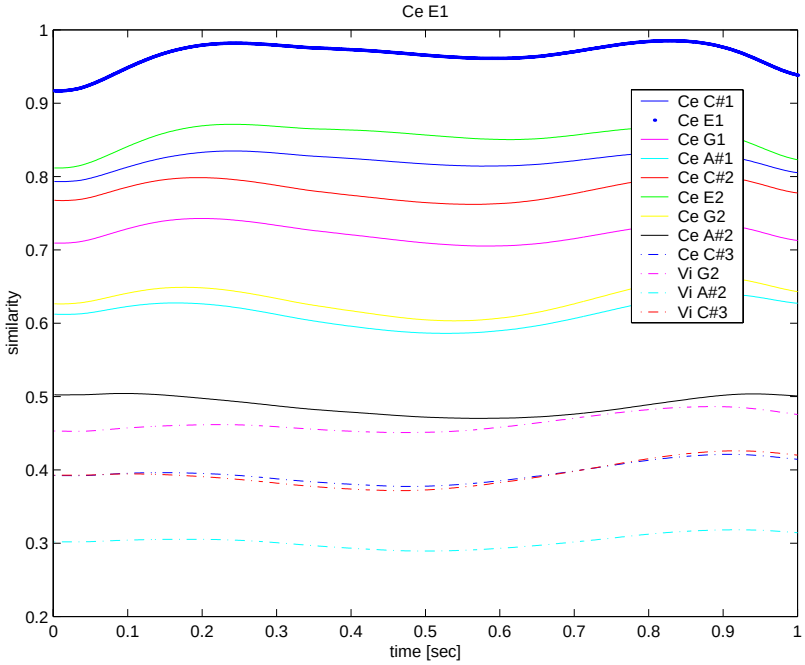

図 5.3: Cello E1 音と各音階との類似度の時間変化

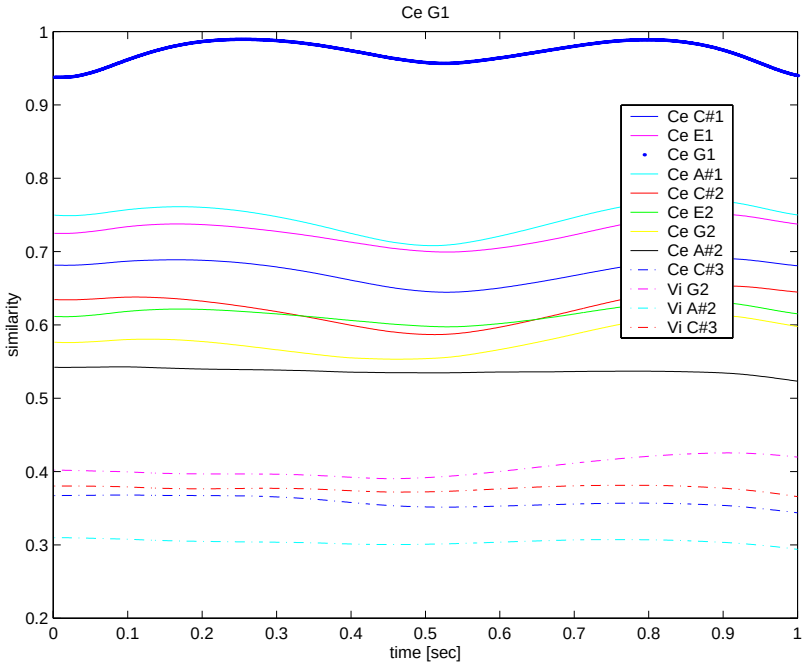

図 5.4: Cello G1 音と各音階との類似度の時間変化

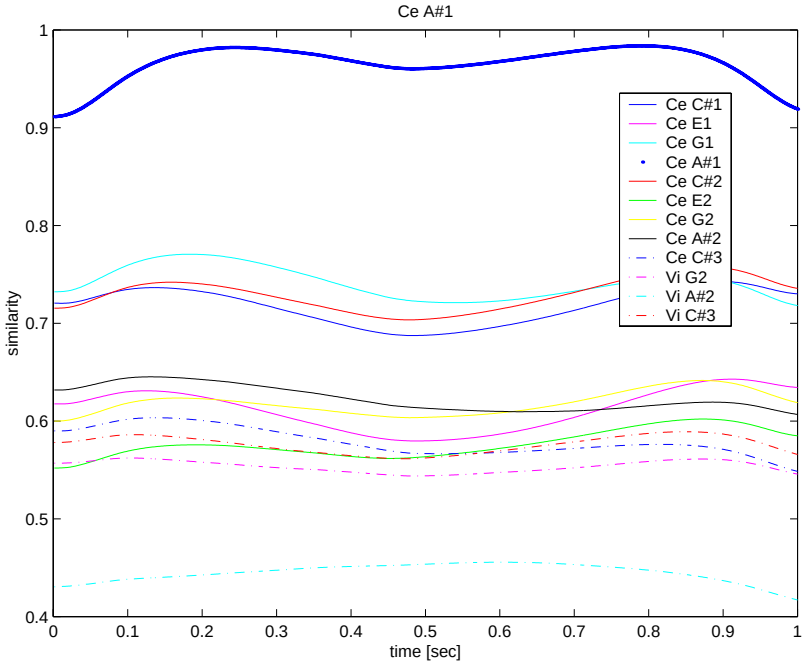

図 5.5: Cello A♯1 音と各音階との類似度の時間変化

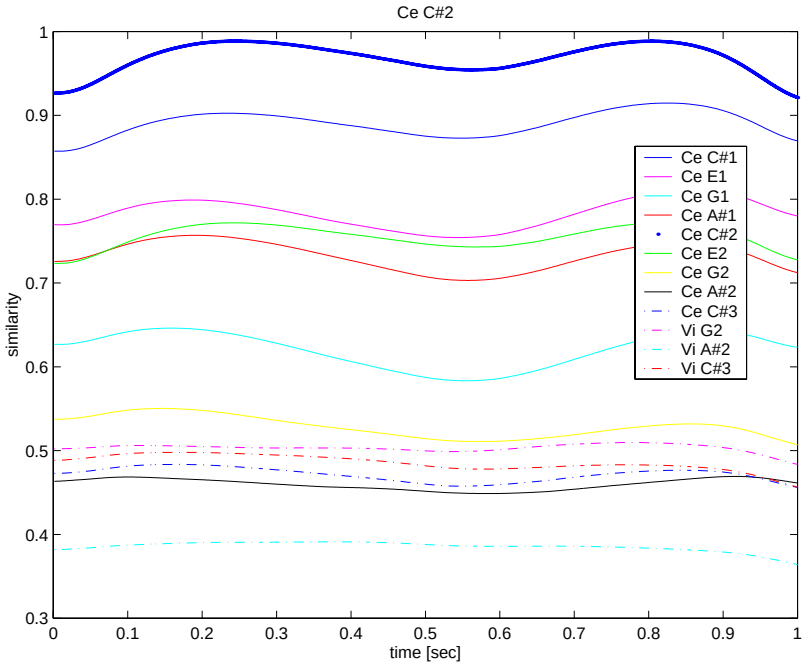

図 5.6: Cello C♯2 音と各音階との類似度の時間変化

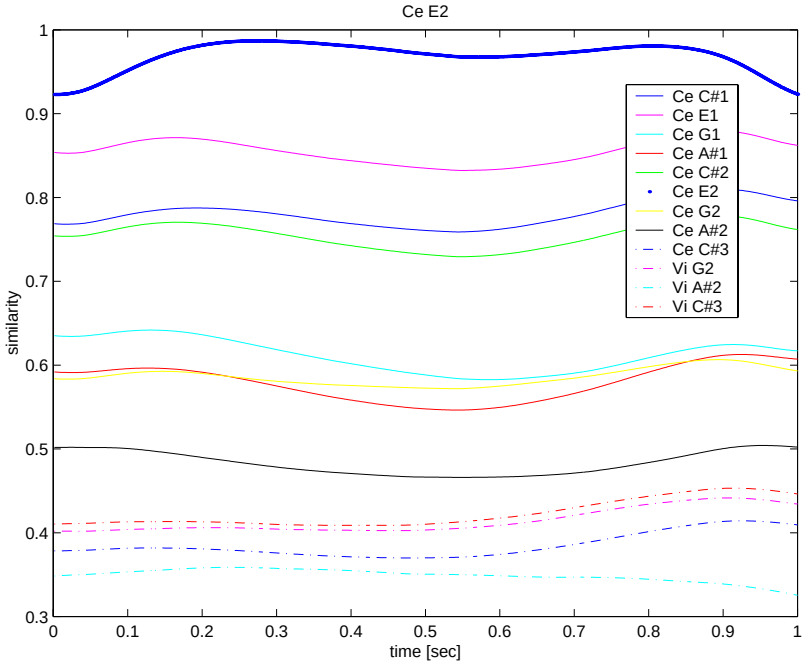

図 5.7: Cello E2 音と各音階との類似度の時間変化

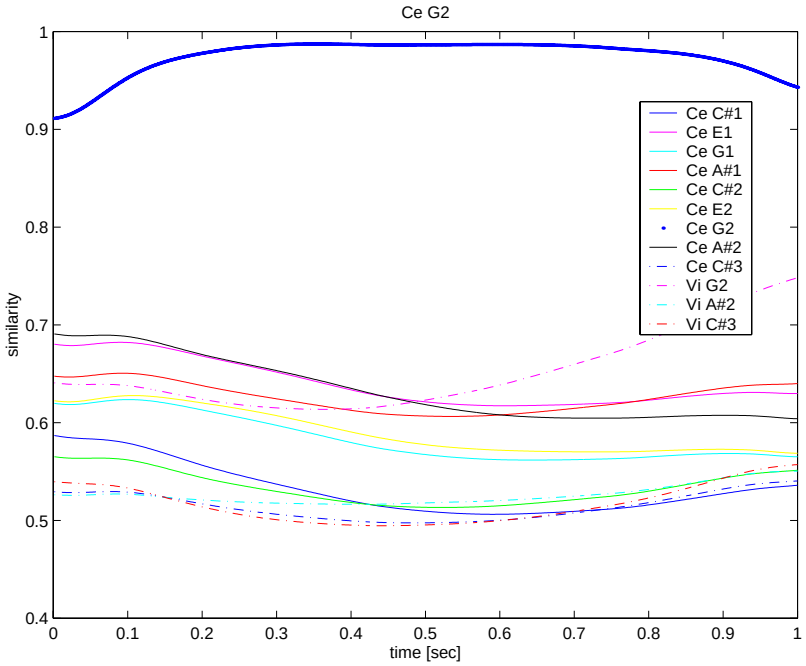

図 5.8: Cello G2 音と各音階との類似度の時間変化

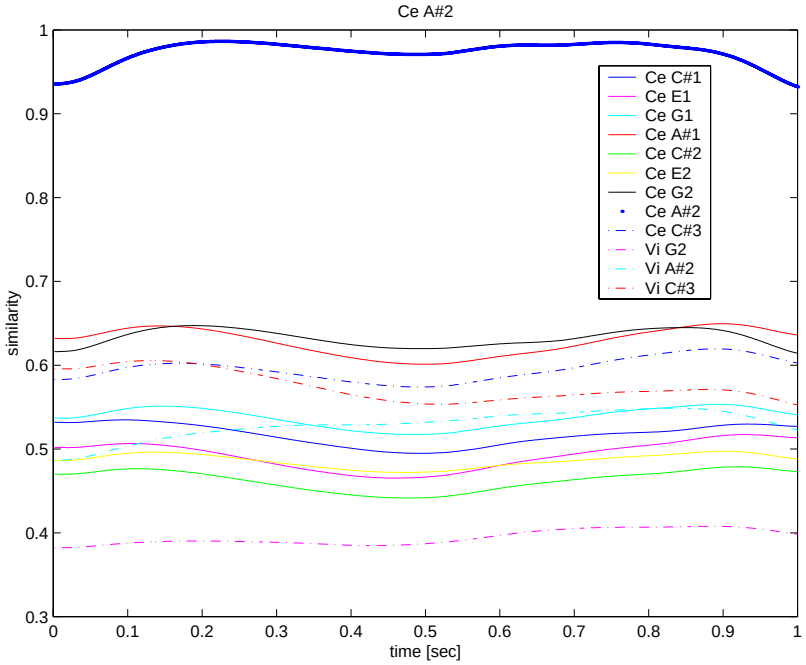

図 5.9: Cello A♯2 音と各音階との類似度の時間変化

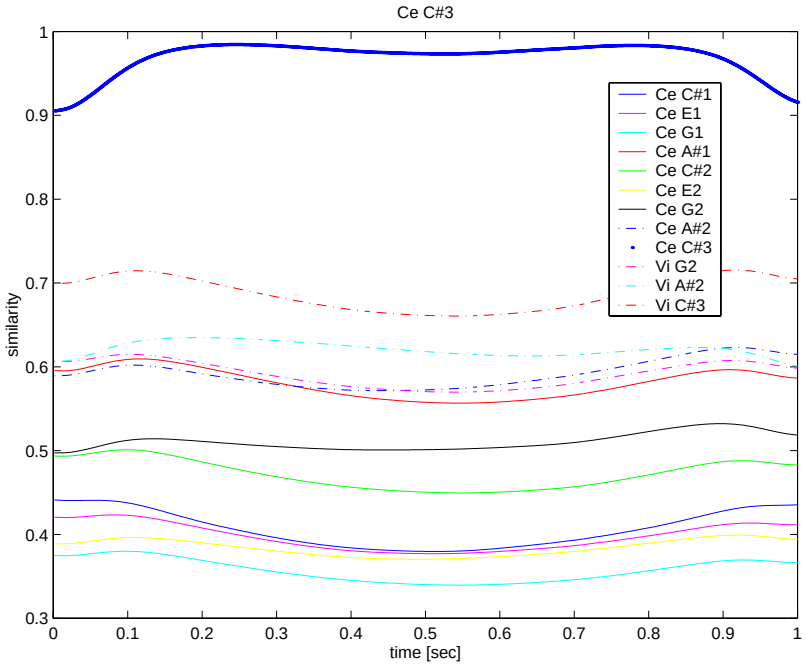

図 5.10: Cello C♯3 音と各音階との類似度の時間変化

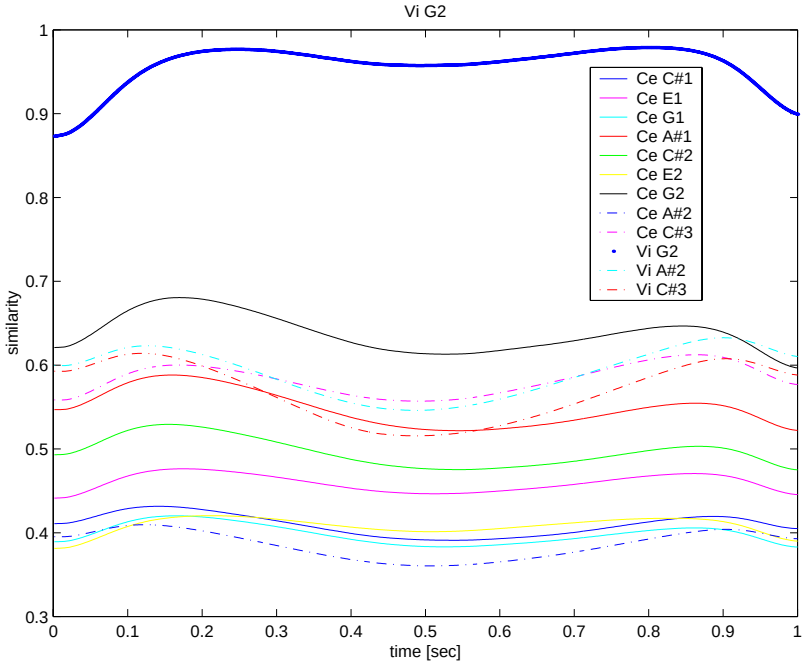

図 5.11: Violin G2 音と各音階との類似度の時間変化

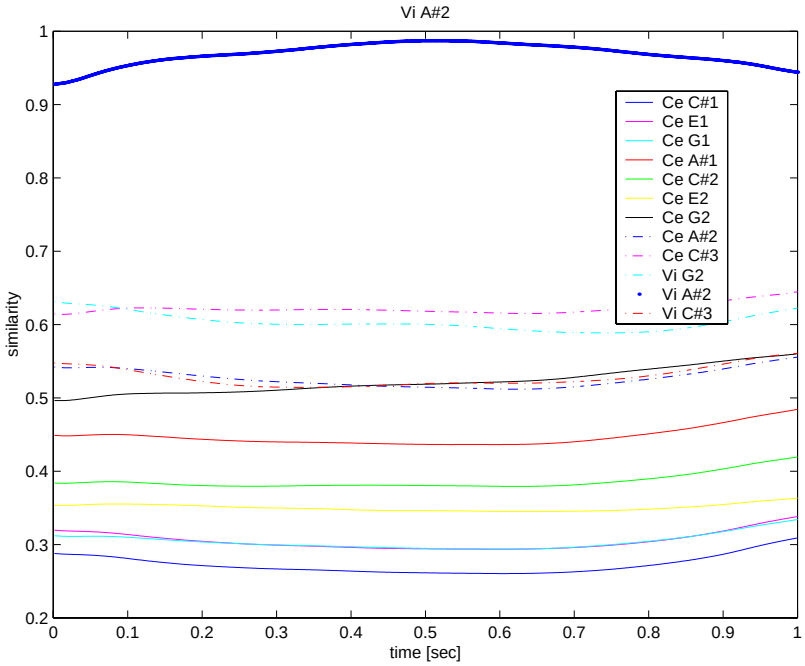

図 5.12: Violin A♯2 音と各音階との類似度の時間変化

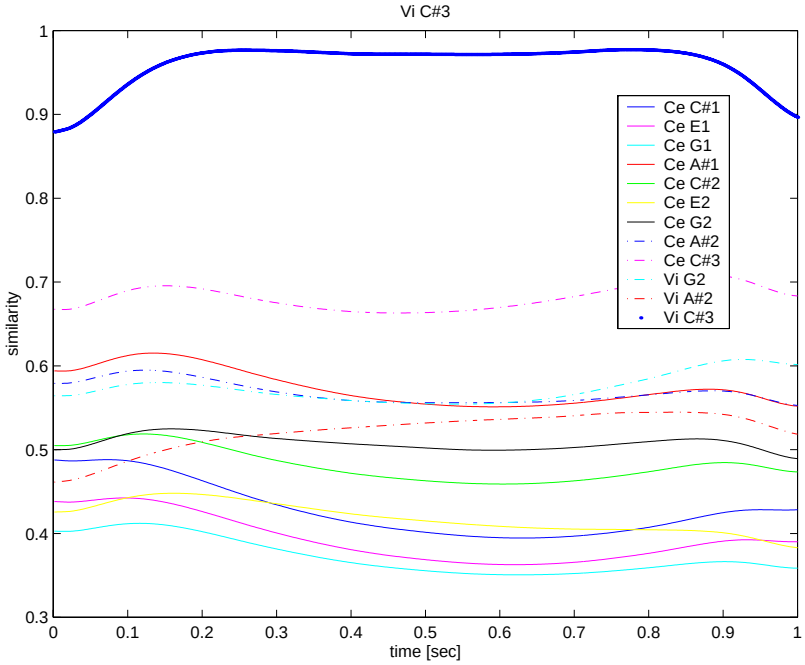

図 5.13: Violin C♯3 音と各音階との類似度の時間変化

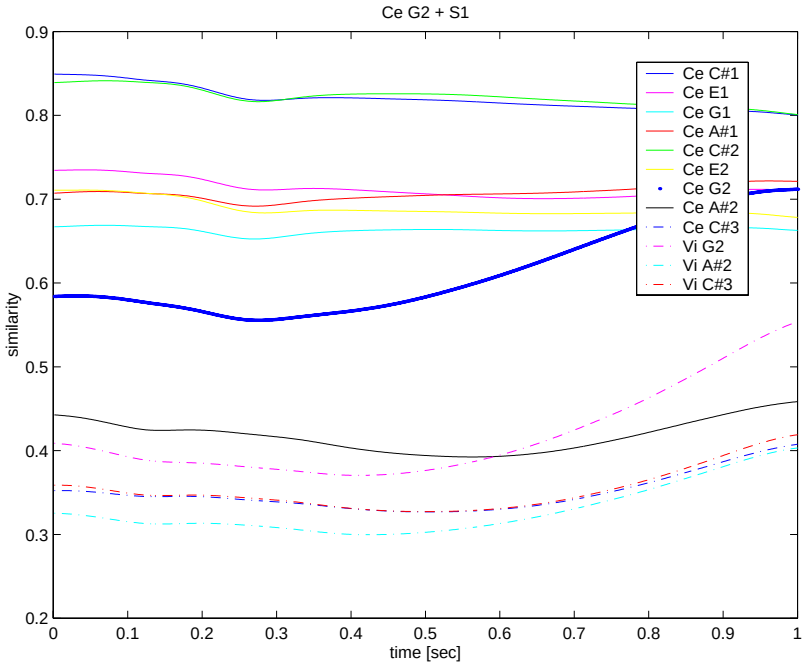

図 5.14: Cello G2 音 + 突発音と各音階との類似度の時間変化

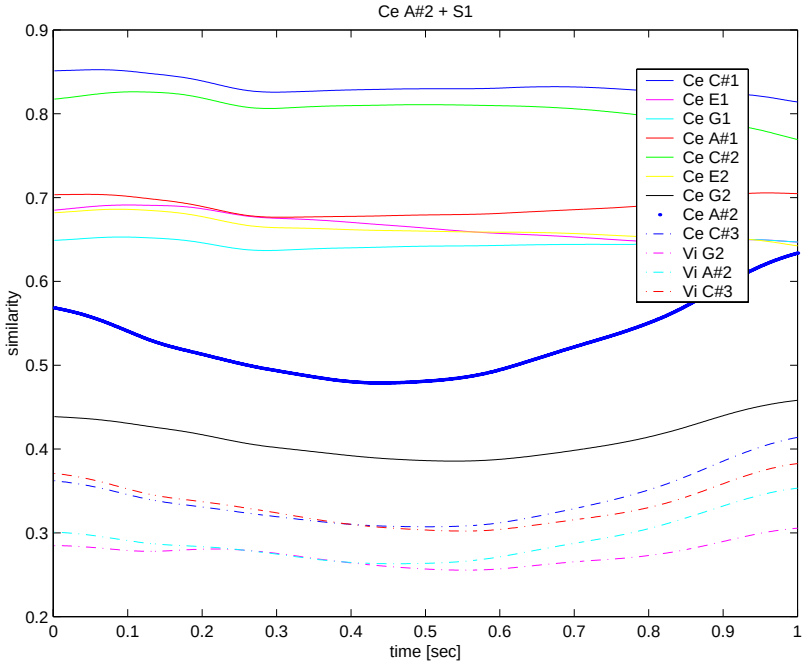

図 5.15: Cello A♯2 音 + 突発音と各音階との類似度の時間変化

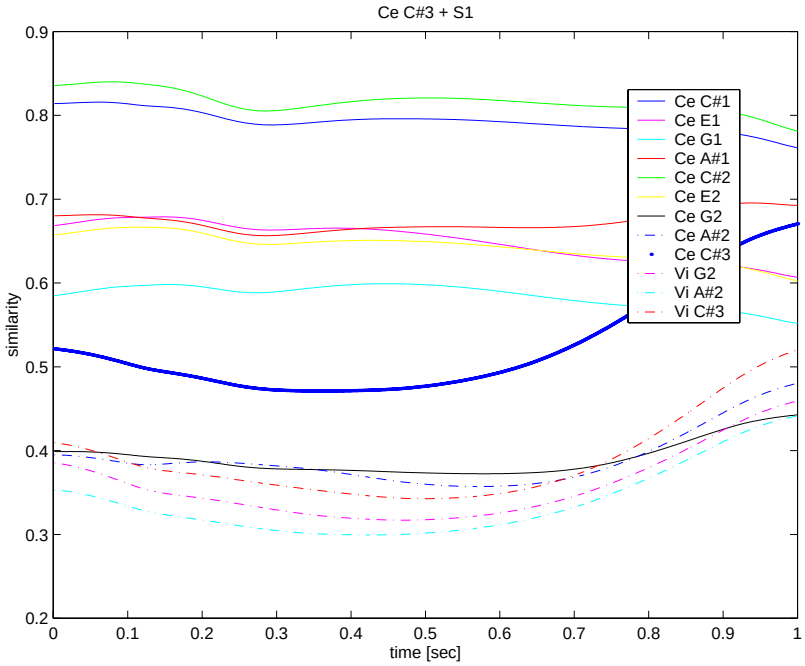

図 5.16: Cello C♯3 音 + 突発音と各音階との類似度の時間変化

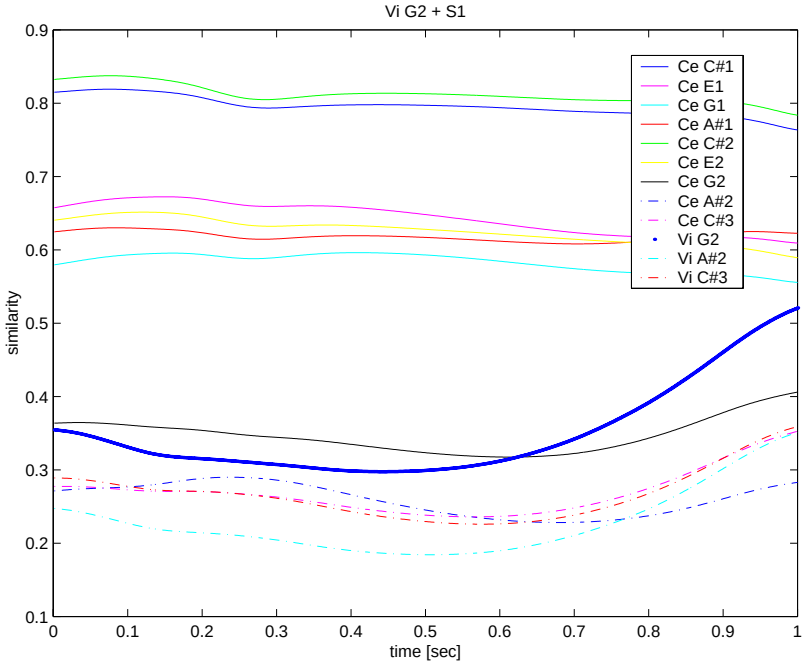

図 5.17: Violin G2 音 + 突発音と各音階との類似度の時間変化

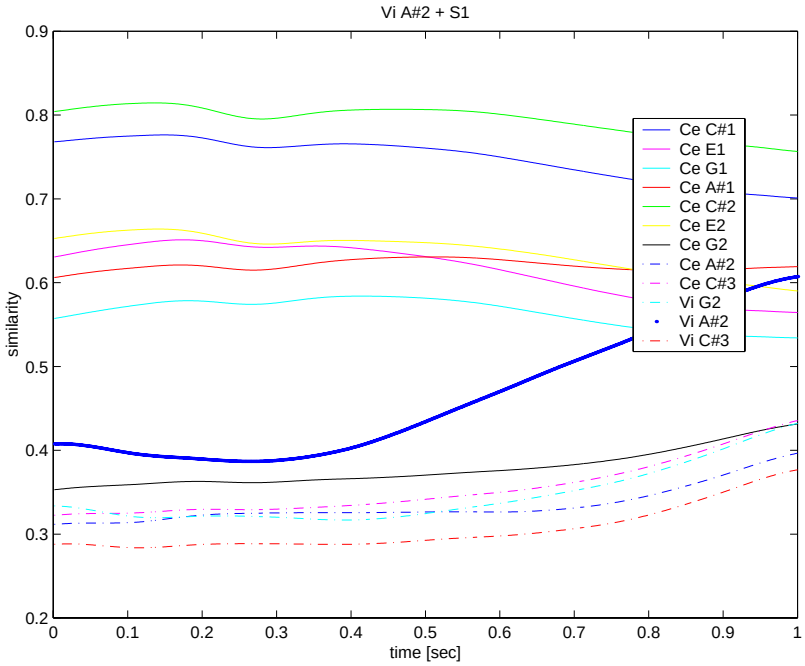

図 5.18: Violin A♯2 音 + 突発音と各音階との類似度の時間変化

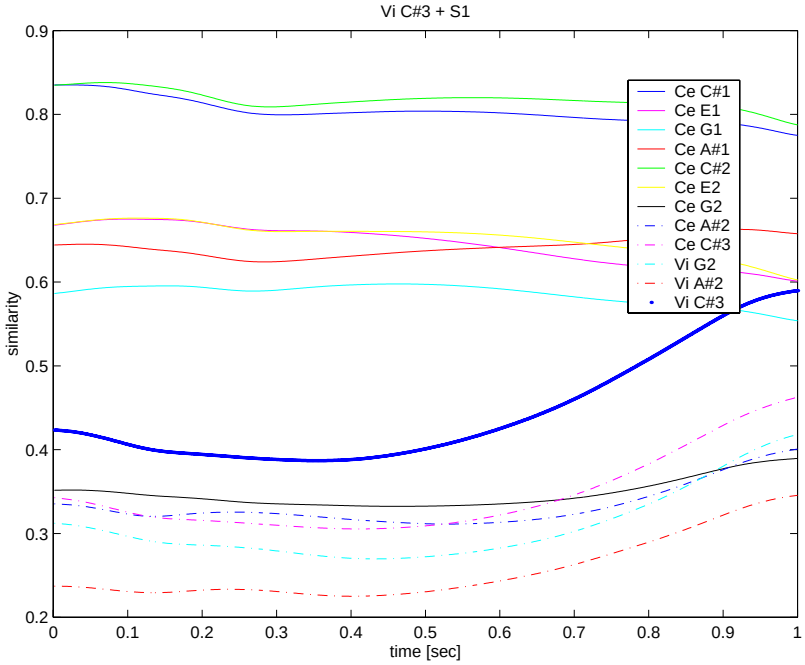

図 5.19: Violin C♯3 音 + 突発音と各音階との類似度の時間変化

5.2.4 考察

各音階とのマッチングでは, 正しい楽器音の音階との類似度はすべて 0.9 以上であった. 突発音を足し合わせた場合では, 類似度は 0.9 以上のものはなかった. したがって, 全体マッチングで類似度が 0.9 以上であればその楽器の音であると同定することができる. また, 正しい音階との類似度よりも他の音階との類似度の方が高くなっているものがある. 特に Cello の C♯1 と C♯2 が高い価を示している. これは, 突発音およびこの 2 つの音階のマップをマッチング部分の範囲内で見てみると, 突発音の成分が存在している部分と上記の 2 つの音階の成分が存在している部分とが比較的同じ場所にあり, また突発音の成分がかなり大きいので, おおまかな形状が似てしまっていることが原因ではないかと考えられる.

したがって, 全体マッチングでは, 何も足し合わされていない場合の音しか同定することができないので, 部分的なマッチングを行なうことによって, 足し合わされている場合でも目的の音の成分がマップ上に存在することを調べる必要がある.

5.3 部分マッチング

5.3.1 目的

全体マッチングでは, 突発音が足し合わされた状態では何の楽器音が存在しているのかを同定することができなかったので, 部分的にマッチングを行なうことで存在しているのかを調べる.

5.3.2 マッチング方法

部分マッチングは, 全体マッチングでの領域を f_c 軸で 64 分割し, 分割した領域の各点の大きさを, パターンベクトルの要素として取り扱い, そのベクトルを用いて類似度を計算する. 図 5.20 に部分マッチングの領域を示す. 類似度は前節と同じ式を用いて計算している.

$$s_k{}^{(l)} = \cos\theta_k{}^{(l)} = \frac{(\boldsymbol{x}_k, \boldsymbol{r}_k{}^{(l)})}{\|\boldsymbol{x}_k\| \cdot \|\boldsymbol{r}_k{}^{(l)}\|} \qquad (0 \leq s_k{}^{(l)} \leq 1)$$

$\boldsymbol{x}_k$ ： 入力マップ

$\boldsymbol{r}_k{}^{(l)}$ ： データベースにあるマップ

$s_k{}^{(l)}$ ： 類似度 $(1 \leq l \leq 12)$

l は 1 〜 9 が Cello の各音階, 10 〜 12 が Violin の各音階を表していて, k は f_c 軸 の各番号を表している.

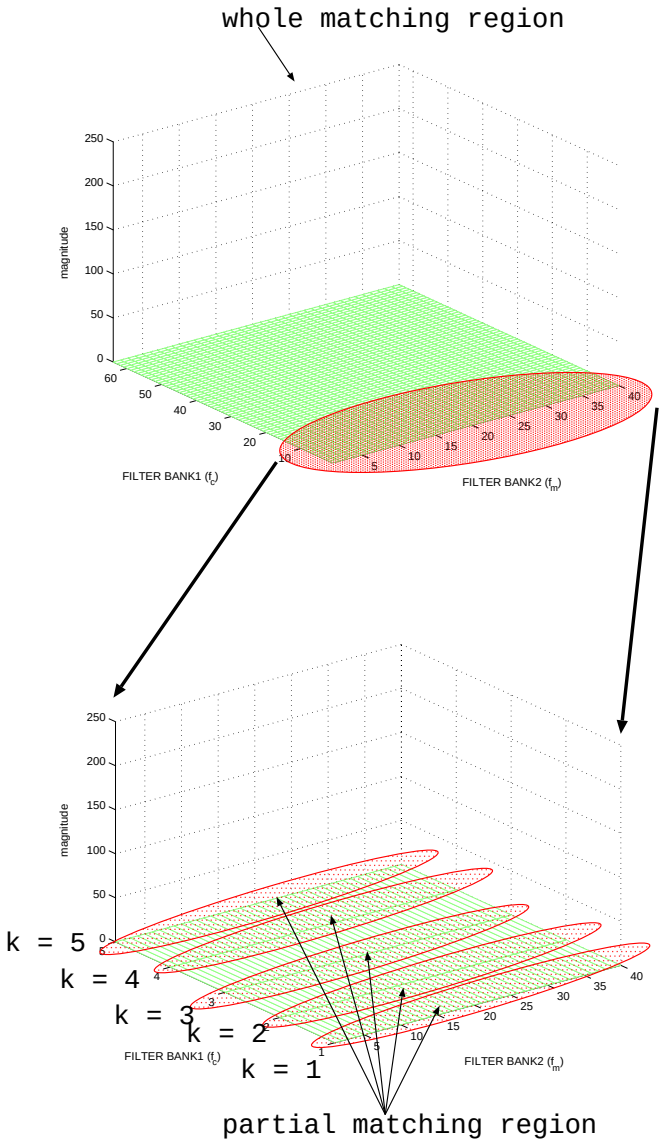

図 5.20: 部分マッチングでのマッチング領域

5.3.3　マッチング結果

5.3.2 節で述べた各領域ごとの類似度の計算を行なった．ここで各グラフの縦軸は類似度の値を表し，横軸は時間を表している．また各グラフの上方の添字は搬送周波数（f_c）のフィルタバンクのチャネル番号を表している．例えば，”33 - 34” はチャネル番号 33 から 34 のフィルタバンクを表している．

図 5.21 〜 図 5.26 に部分マッチングの結果を示す．

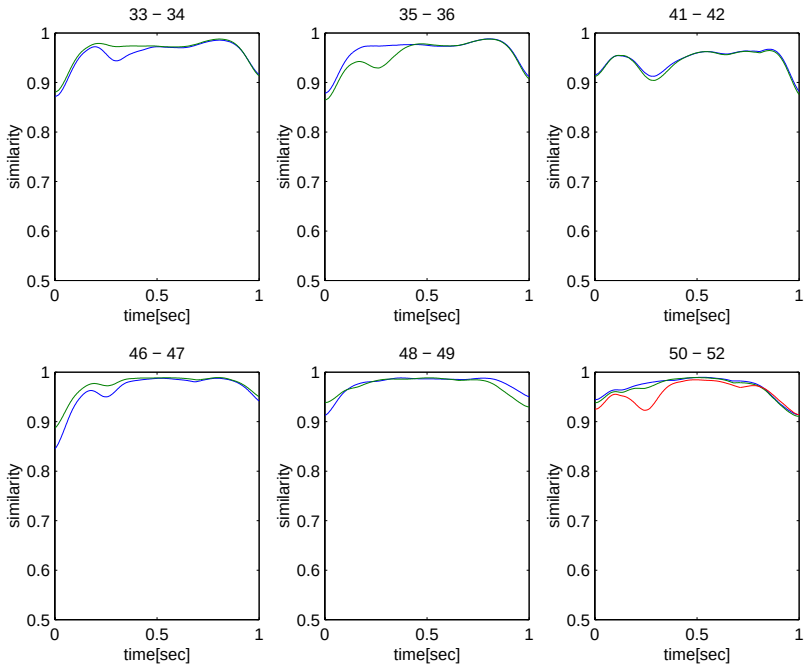

図 5.21: Cello G2 音 + 突発音 $\Longleftrightarrow$ Cello G2 音の各領域の類似度の時間変化

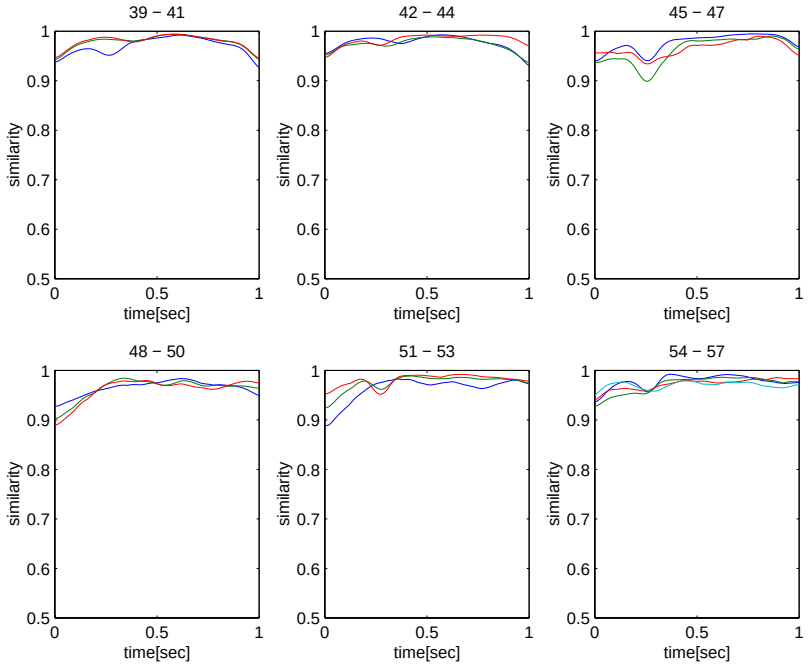

図 5.22: Cello A♯2 音 + 突発音 ⟺ Cello A♯2 音の各領域の類似度の時間変化

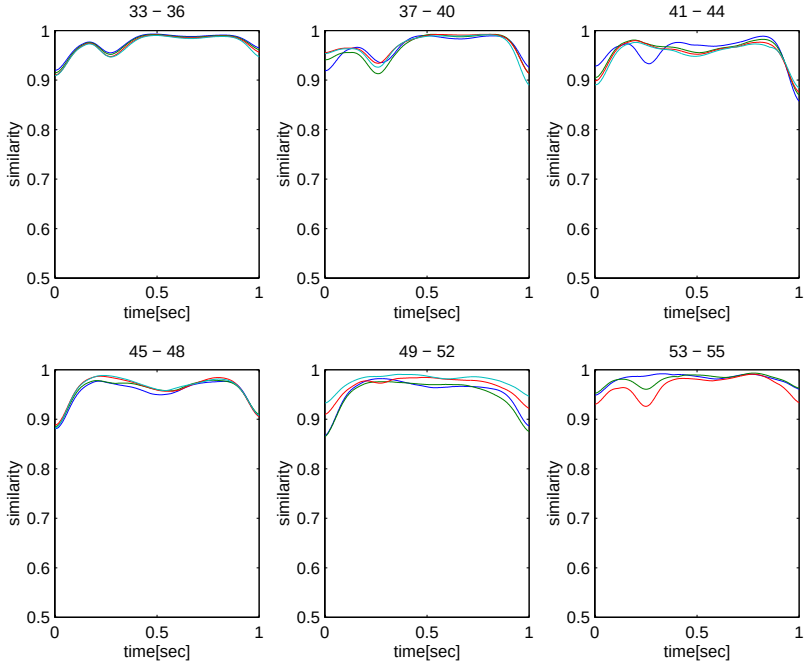

図 5.23: Cello C♯3 音 + 突発音 ⟺ Cello C♯3 音の各領域の類似度の時間変化

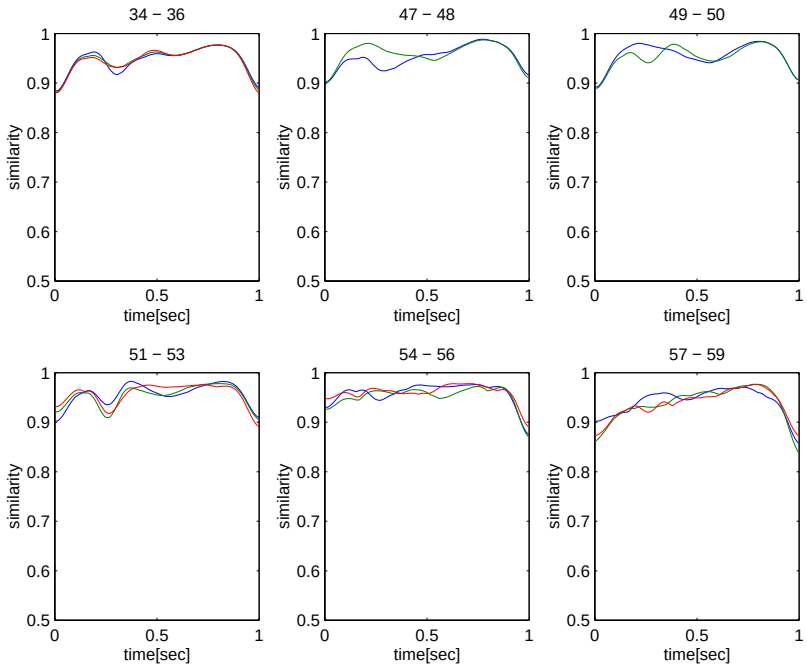

図 5.24: Violin G2 音 + 突発音 $\Longleftrightarrow$ Violin G2 音の各領域の類似度の時間変化

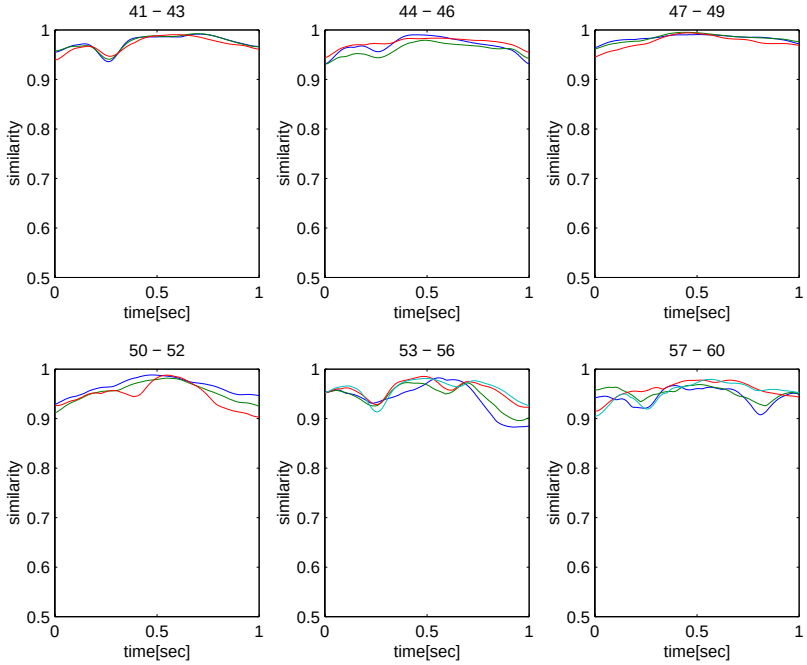

図 5.25: Violin A♯2 音 + 突発音 $\Longleftrightarrow$ Violin A♯2 音の各領域の類似度の時間変化

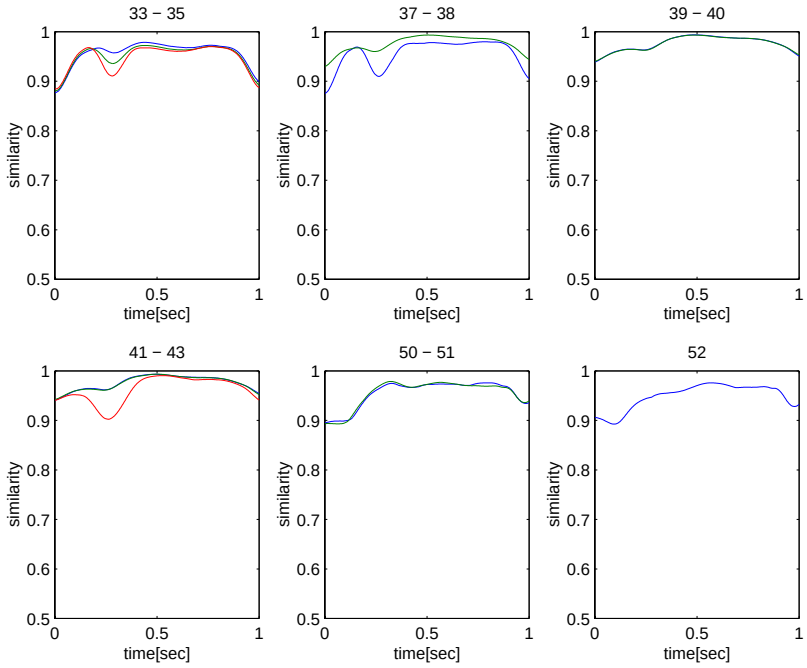

図 5.26: Violin C♯3 音 + 突発音 ⟺ Violin C♯3 音の各領域の類似度の時間変化

5.3.4 考察

今回グラフで表したものは, 搬送周波数（f_c）軸側のチャネル番号の 33 から 64 の中で, 類似度が全時間にわたってほぼ 0.9 以上の値を示したものである. これは, 突発音による影響が他のチャネルと比較してほとんどなく, 楽器音の成分が変調 – 搬送周波数マップ上に存在することを意味する. したがって, このような領域を手がかりして音源の同定が行なえるのではないかと考える. 例えば, 今回は f_c 軸側で分割して類似度を計算したが, これに加えて f_m 軸側で分割して類似度を計算し, 両軸ともに類似度が高い領域を部分マッチングすることも考えられる. このようにすれば, マップ上に楽器音の成分が存在することが確認され, 同定することが可能になるのではなかろうか. 同定されれば, 突発音によって影響を受けている領域を, データベースのものと入れ替えることによって, 楽器音が抽出できる可能性は高まる.

また, 同じ楽器音・音階との部分マッチングでは, 各分割領域の類似度は全時間にわたって 0.9 以上のものがほぼ 100 % であった. したがって, 同じ楽器音・音階どうしのマッチング結果と, 突発音が足し合わされた場合のものとのマッチング結果を比較すれば, どの成分のどの時間帯に突発音の影響が表れているのかを知ることができ, 突発音がどの時間帯に存在しているかという情報がおおまかではあるが得られる.

しかし, これまで述べたことは, 用いた突発音に限定されている. 今回用いた突発音は, 搬送周波数（f_c）軸側の 1 から 32 の領域に影響が著しく表れているが, 例えば, マップの全領域に影響を与えるような音が長時間足し合わされていれば, 本方法では音源を同定することが困難となり, 他の方法を考慮する必要がある.

第 6 章

結論

6.1 本論文で明らかになったことの要約

本論文では, 楽器音および突発音が付加された楽器音を変調 – 搬送周波数マップ検出モデルを用いて周波数分解し, 振幅変調情報に着目し, またデータベースとのマッチングを行なうことで, 楽器音を同定する方法を提案した. 本方法では, 楽器音の音階を, 変調 – 搬送周波数マップを f_m 軸側において探索することで同定できることがわかった. 全体マッチングでは, 突発音が足し合わされていると楽器音を同定することができないので, 部分マッチングによって同定への手がかりを得た. また, 部分マッチングの結果を利用すれば突発音が時間軸上のどの部分に存在しているのかをおおまかにではあるが検出できるのではないかと考えられる.

6.2 今後の課題

以下に今後の課題を列挙する.

- 他の音階への適応

  本研究で用いた変調 – 搬送周波数マップ検出モデルは 300 Hz の基本周波数までしか対応していないので, FILTER BANK 1, FILTER BANK 2 の通過帯域幅とフィルタ数（チャネル数）を他の音階へ適応させる必要がある.

- 他の楽器音の解析

  本研究では, 用いた楽器音はチェロとヴァイオリンの弦楽器のみであった. 管楽器などの他の楽器音も解析する必要がある.

謝辞

本研究を進めるにあたり, 多大な御助言を頂き熱心な御指導をして頂きました赤木 正人教授に心から感謝致します. 本研究に関して, 多大な御助言をして頂いた博士後期課程の水町 光徳氏, 伊藤 一仁氏に心より感謝致します. また, 本研究を進めるにあたり有意義な討論並びに, 有益な助言を賜った赤木研究室の皆様方に心より感謝します.

最後に, 2 年間の研究生活を支えて下さったすべての皆様に厚く感謝致します.

参考文献

[1] 三浦 種敏監修, " 新版・聴覚と音声", 信学会編 (1980).

[2] 柏野 牧夫, 岩澤 秀紀, " 振幅変調音に対する順応", (1997) 聴覚研資料, H-97-72 (1997).

[3] Albert S. Bregman, "Auditory scene analysis: hearing in complex environments, in Thinking sounds", pp.10-36, Oxford University Press, (1993).

[4] Cooke, M.P. and Brown, G.J., "Computational auditory scene analysis", Computer Speech and Language, pp.297-336, 8 (1994).

[5] D.P.W., Ellis, "Prediction-driven computational auditory scene analysis", Ph.D. Thesis, MIT Media Lab (1996).

[6] 中谷 智広, 川端 豪, 奥乃 博, " 計算論的アプローチによる音響ストリームの分離", 音響学会聴覚研資料, H-93-83 (1993).

[7] 柏野 邦夫, 田中 英彦, " 二つの周波数成分の分離知覚に関する工学的モデル", 電子情報通信学会論文誌, J77-A, 5, pp.731-740 (1994).

[8] 柏野 邦夫, " 計算機による聴覚の情景解析 – はじめの一歩 –", 日本音響学会誌, vol.50, No.12, pp.1023-1028 (1994).

[9] 三浦 慎, " 振幅変調マスキングに基づいた音源分離に関する研究", JAIST 修士論文 (1999).

[10] Tomoyoshi Kinoshita, Shuichi Sakai and Hidehiko Tanaka, "Musical Sound Source Identification Based on Frequency Component Adaptation", IJCAL-99 CASA'99 (1999).

[11] チャールズ K. チェウイ著／桜井 明・新井 勉共訳, " ウェーブレット入門", 第 3 章, 東京電機大学出版会 (1993).

[12] 佐藤 雅昭, ” ウエーブレット理論の数学的基礎 第 I 部 ”, 日本音響学会誌, vol47, No.6, pp.405-415 (1991).

[13] 佐藤 雅昭, ” ウエーブレット理論の数学的基礎 第 II 部 ”, 日本音響学会誌, vol47, No.6, pp.416-423 (1991).

[14] 河原 英紀, ” ウエーブレット解析の聴覚研究への応用”, 日本音響学会誌, vol47, No.6, pp.424-429 (1991).

[15] 長嶋 洋一, 橋本 周司, 平賀 譲, 平田 圭二共編, ”bit 8 月号別冊 コンピュータと音楽の世界 – 基礎からフロンティアまで –”, 共立出版 (1998).