

Title	振幅変調マスキングの知見に基づいた音源分離に関する研究
Author(s)	三浦, 慎
Citation	
Issue Date	1999-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1262
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

修 士 論 文

**振幅変調マスキングの知見に基づいた
音源分離に関する研究**

指導教官 赤木正人 助教授

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

三 浦 慎

1999 年 2 月 15 日

目次

1 はじめに	1
2 研究の背景	3
2.1 振幅変調知覚に関する心理学的知見	3
2.2 振幅変調マスキングの知見	5
2.2.1 振幅変調検出による周波数選択性	5
2.2.2 振幅変調音に対する順応	6
2.3 聴覚の情景解析	6
2.3.1 共通の立ち上がり・立ち下がりの規則（OLD-PLUS- NEW発見則）	7
2.3.2 漸近的变化に関する規則	7
2.3.3 調波関係に関する規則	8
2.3.4 1つの音響事象に生じる変化の規則	8
2.4 本研究の特色	9
3 変調—搬送周波数マップの検出とシミュレーション	10
3.1 変調—搬送周波数マップ検出モデルの概要	10
3.1.1 変調—搬送周波数領域における検出	10
3.1.2 フィルタバンク	14
3.1.3 フィルタバンクの式	16
3.2 変調—搬送周波数マップの検出のシミュレーション	17
3.2.1 実験データ	18
3.2.2 変調—搬送周波数マップの検出結果	18
3.2.3 考察	26
4 AM 音とマスキアーの混合信号の分離・抽出とシミュレーション	28
4.1 AM 音とマスキアーの分離・抽出	28

4.1.1	DC 成分の分離	28
4.2	分離・抽出のシミュレーション結果	29
4.2.1	考察	29
5	二重母音の分離・抽出とシミュレーション	31
5.1	二重母音の分離・抽出	31
5.1.1	調波関係を用いた音源分離	31
5.1.2	音の共通の立ち上がり／立ち下がりを用いた音源分離	33
5.2	二重母音の分離・抽出のシミュレーション	38
5.2.1	分離・抽出結果の評価法	38
5.2.2	シミュレーション結果	38
5.2.3	考察	45
6	おわりに	48
	謝辞	49

第 1 章

はじめに

機械による音声認識は、雑音が含まれていない状態でならば、ほぼ実用レベルに達している。しかし、認識される音声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かなり低下する。一方人間は、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分離して聞くことができるカクテルパーティ効果と呼ばれる機能を持っている。この音源の分離機能がモデル化されれば、前述の雑音下や複数音が存在する状況下において、目的とする音だけを選択的に取り出すことが工学的に可能となり、これを前処理とすることで雑音に強い音声認識や音声合成などの様々な音声情報処理へ応用することができると思われる。

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 [1]。このような様々な音における振幅変調を検出することが出来れば音源の識別や音声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る。

そこで本研究ではこの振幅変調情報に着目し、人間の聴覚情報処理機構を基にこのような振幅変調情報を検出する変調—搬送周波数マップのモデルを作成する。これによって音にどのような振幅変調がされているのかを調べその特徴を見ることで音源分離の手がかりとする。またここで、構築した変調—搬送周波数マップが本研究の基礎となっている振幅変調マスキングの知見 [3] [4] の工学的なモデルとなっているかについて検討する。さらに、この検出結果と Bregman によって提案された聴覚の情景解析 [2] の考え方の”調波関係に関する規則”と”共通の立ち上がり／立ち下がりに関する規則”を実装することで検出された変調—搬送周波数マップを用いてカクテルパーティ効果において人間の聴覚では、うまく分離が行える 2 つの母音に時間的な重なりがあるような二重母音（double vowel）の音源分離を行う。

以下第2章では本研究の背景となっている変調知覚に関する心理学的知見について述べ、特に本研究の基礎となっている振幅変調マスキングの知見及び音源分離の実装アルゴリズムに用いている聴覚の情景解析 [2] について述べる。第3章では前述の振幅変調マスキングの知見を基に構築した音の解析を行う変調—搬送周波数マップの検出モデルを示す。次に構築したモデルを用いてシミュレーション音による検出結果を示す。第4章では実装した変調—搬送周波数マップを用いて振幅変調マスキングの知見における実験データである AM 音とマスキング音の分離・抽出のシミュレーションを示す。第5章では検出した変調—搬送周波数マップを使って聴覚の情景解析の “調波関係を用いた音源分離”, “音の共通の立ち上がり／立ち下がりを用いた音源分離” に基づいた音源分離の手法について述べる。同様にシミュレーションによって二重母音の場合の音源分離の結果を示す。第6章では、本研究をまとめて今後の課題を考察する。

第 2 章

研究の背景

本章では本研究の考えの基本になっている振幅変調マスキングの知見とそれに関連する変調知覚に関する心理学的知見, Bregman によって提案された聴覚の情景解析 (Auditory Scene Analysis) について述べる. 最後に本研究の特色について述べる.

2.1 振幅変調知覚に関する心理学的知見

我々が日々耳にする音のほとんどは複雑に時間変化する振幅包絡, すなわち振幅変調 (AM) を伴っている. この振幅変調は音色にも大きな影響を与え音源の識別においても重要な情報となっていることが知られている [12].

このように振幅変調は聴覚の広範な側面において重要な役割を果たすため, 多くの研究者の関心を呼んでいるが, その処理メカニズムに関してまだ定説と呼ばれるものは得られていない.

AM 処理の代表的なモデルは次のようなものがある. 図 2.1 (a) に示されるように耳に入力された信号は聴覚フィルタの機能を反映した帯域通過フィルタを通過し, 半波整流を受ける. その出力が低域通過フィルタで平滑化され振幅包絡が検出される. 図 2.2 に示されるように AM 検出の心理物理実験から, この低域通過フィルタの時間的変調度伝達関数 (TMTF) は 10Hz までほぼ平坦で 50Hz で約 3dB 減衰し, それ以上の変調周波数ではおよそ 4-5 dB/oct の割合で減衰することが示されている [5]. この図で m は変調深さ (modulation depth) を示しており変調に対する感度が高いほど値は負となる.

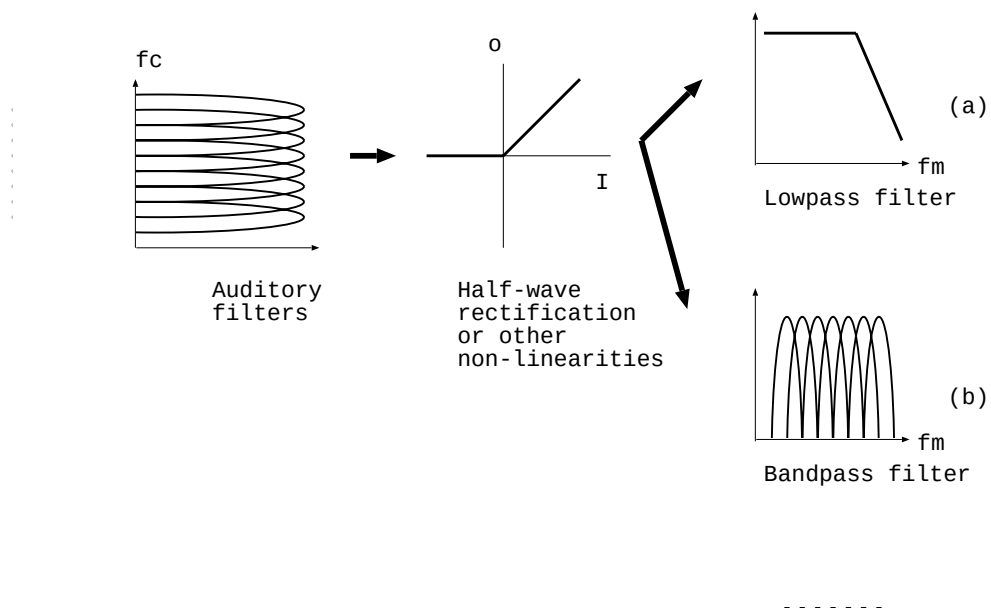


図 2.1: 聴覚の時間分析モデル

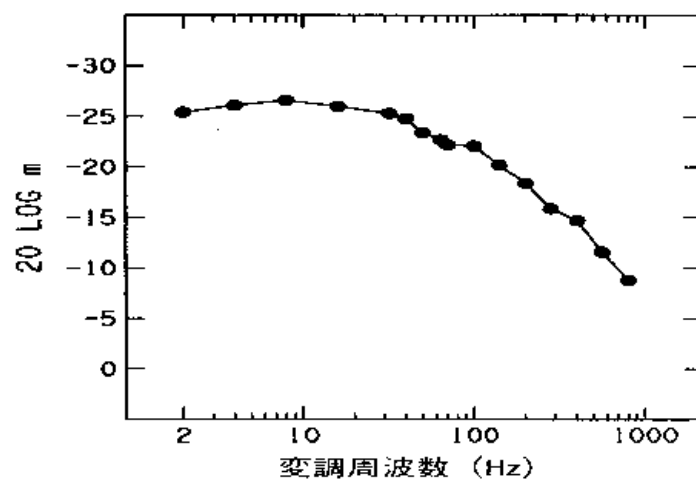


図 2. 2: 時間的変調度伝達関数 (TMTF) [5]

2.2 振幅変調マスキングの知見

しかし，この低域通過型モデルでは，目的音とマスキアーの搬送周波数が異なる正弦波でも変調周波数が近いほどマスキングが生じる変調検出妨害（modulation detection interference：MDI）をうまく説明できない．

ここで，このような単一の低域通過フィルタの代わりに図 2.1(b) に示すように，それぞれが異なった変調周波数に同調する複数の帯域通過フィルタ群（変調フィルタバンク）を仮定するモデルも提案されている．この変調フィルタバンクの存在を支持する知見について次に述べる．

2.2.1 振幅変調検出による周波数選択性

T. Hutgast, Sid P. Bacon, D. Wesley Grantham [4] は，信号と帯域雑音を同じ変調周波数で振幅変調することにより著しくマスキング量が増加することを示した．図 2.3 [3]

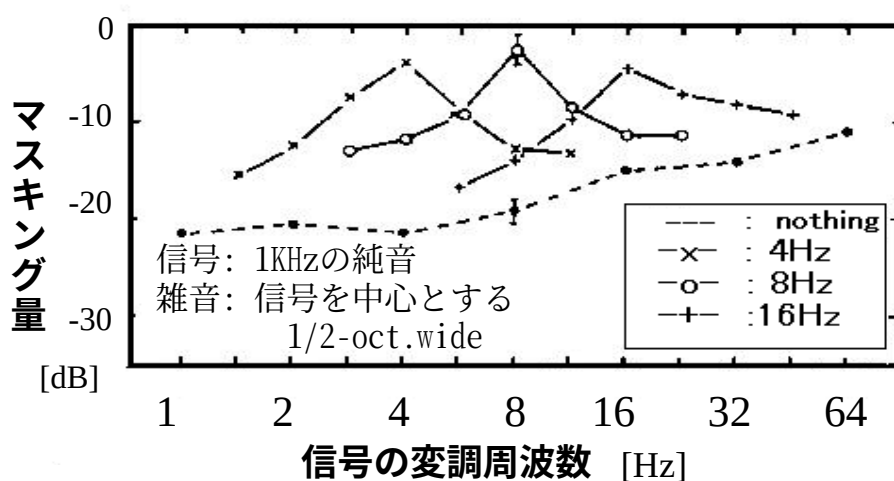


図 2.3: Hutgast の実験結果 [3]

に T. Hutgast による実験結果を示す．図では横軸が信号の変調周波数，縦軸がマスキング量を表している．結果は，4 種類の変調周波数を掛けた雑音に対して示されている．それぞれ破線（—）は変調なし，（—×—）は雑音に 4 kHz の変調，（—○—）は 8 kHz の変調，（—+—）は 16 kHz の変調を示している．

(— + —) は 16Khz の変調を掛けた場合の結果を示している。この結果より、信号と雑音の変調周波数が同じの時にマスキング量が増加していることが分かる。この振幅変調によるマスキング量の増加現象は、各聴覚フィルタからの出力の振幅包絡が信号と雑音で同じであるため、聴覚が信号を分離抽出できないことによって生じる。逆に信号と雑音の振幅包絡が異なる場合、マスキング量は減少する。このことは、聴覚が聴覚フィルタからの出力信号の振幅包絡をフィルタリングし、信号を分離抽出するような機能を持っていることを示している。

2.2.2 振幅変調音に対する順応

柏野・岩澤ら [1] [7] は振幅変調 (AM) の閾値が、順応によってどのように変化するかを、搬送波・変調波ともに正弦波を用いて測定している。その結果、(1) 順応音と検査音の変調周波数が同一のとき検出閾は上昇するが両者の差が増大すると検出閾上昇は小さくなる。(2) 順応音と検査音についても同様の選択性がある、(3) 順応音と検査音を反対耳に提示すると同時耳に提示した場合より順応による検出閾上昇が小さいことが示された。これらの結果は、聴覚系の中でも比較的初期の両耳融合以前の段階に、各搬送周波数ごとに、特定範囲の振幅変調周波数に選択性を持つ変調フィルタバンクが存在することを示している。

2.3 聴覚の情景解析

Bregman によって広く知られるようになった聴覚の情景解析 (Auditory Scene Analysis) [2] の提案は、それまで個別に研究されてきた聴覚的現象を、音を通じて環境を把握するための機能としてとらえ直すことにより統一的に理解しようとするものであった [6]。

聴覚はいともたやすくこの情景解析の問題を解いているように感じられる。しかし、日常我々が経験するような、様々な音源からの直接音や反射音、残響、周囲騒音などが混在する環境において、1次元 (片耳) もしくは2次元 (両耳) の信号から、周囲に何が起きてるかを把握することは、音源や環境に関する拘束がない限り一意的な答の出ない困難な問題である。このような困難な問題を解くために聴覚が利用している拘束条件の幾つかを、Bregman は以下の発見的規則としてまとめている。

- 規則 1
共通の立ち上がり／立ち下がりに関する規則
- 規則 2
漸近的变化に関する規則
- 規則 3
調波関係に関する規則
- 規則 4
1つの音響事象に生じる変化の規則

聴覚はこれらの規則を利用して、入ってきた音のかたまりを複数の音の系列（音脈:stream）に分け、外界の解釈を試みる。

以下にこの4つの規則の詳細について述べる。

2.3.1 共通の立ち上がり・立ち下がりの規則（OLD-PLUS- NEW発見則）

この発見的規則は”関連のない音と一緒に始まったり終わったりすることはない“ということを示している。言い換えると”ある1つの音源で生じた音は、同時に始まったり終わったりするが、別の音源で生じた音は前者の音と同時に始まったり終わったりしない”ことを意味している。この発見則の1つとして OLD-PLUS- NEW発見則が知られている。

OLD- PLUS- NEW発見則とは図 2.4 の (a) に示すような2つの同じような周波数成分を持つ大きさの異なる音があり、このときこれを (b) に示すように連続的に提示すると人間の聴覚では (c) に示すように古い音に新しい音が足されたような形で知覚するという現象である。

2.3.2 漸近的变化に関する規則

この規則は、漸近的な変化：

- 単一の音はその性質上、ゆっくり滑らかに変化する傾向がある
- 同じ音源からくる一連の音は、その性質上ゆっくり変化する傾向がある

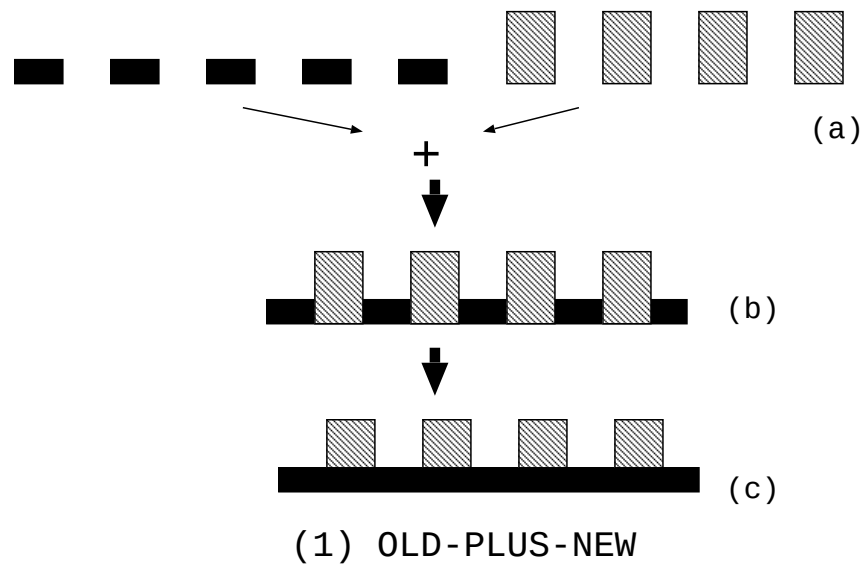


図 2.4: OLD-PLUS- NEW発見則

を示している．これは同じ音源から来る音は急激に変化せず，急激な変化が生じた場合に，別の音源で生じた音を含んでいることを意味している．また，漸近的变化は，時間軸上と周波数軸上の両方で考えられている．

2.3.3 調波関係に関する規則

この規則は，“繰り返し周期で一連の音が振動するとき，その変動は周波数成分が共通な基本周波数の整数倍になるような聴覚パターンを発生する”ということを示している．これは定常母音といった調波音に関する規則であり，調波関係を示している．

2.3.4 1つの音響事象に生じる変化の規則

この発見的規則は”一つの音響事象に生ずる多くの変化は，その音を構成する各成分に同時に同じような影響を与える．”ということを示している．

これを能動的に考えられている例として共変調マスキング解除（CMR：G-modulation masking release）が知られている．共変調マスキング解除とは図 2.5に示すように [6] 純音に帯域雑音を加えて妨害するとき，帯域雑音を振幅変調した場合には妨害音の帯域幅を拡大して音量を増加させることにより，逆に妨害量が減少することである．これを，この発見的規則：1つの音響事象に生じる変化の規則で捕らえると，1つの音源から発する音

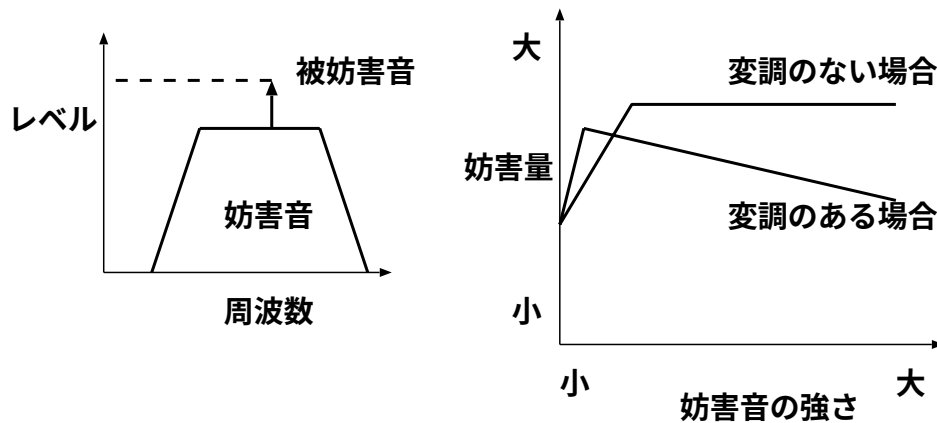


図 2.5: 共変調マスキング解除 (CMR)

に含まれる周波数成分は、同じ傾向で変化するという物理的世界における拘束条件を聴覚は利用しているという考えで説明される。

2.4 本研究の特色

本研究では、変調知覚の知見を基にして、工学的な変調フィルタバンクを作成する。これによって変調—搬送周波数領域における時間変化を検出することができる。

また、この構築した変調—搬送周波数領域における時間変化を、先に示した聴覚の情景解析の規則 1 の共通の立ち上がり／立ち下がりに関する規則と規則 3 の調波関係に関する規則の考え方を工学的に捕え直し実装することで音源分離を行う。

これによって従来は困難であった同じ周波数が重なるような場合の音源分離を試みる。ここでは特に音声等の認識において重要な時間的な振幅包絡の分離に重点をおいている。

第 3 章

変調—搬送周波数マップの検出とシミュレーション

本章では最初に変調—搬送周波数マップ検出モデルの概要について述べ、次に実装した変調—搬送周波数マップ検出モデルを用いて行ったシミュレーション結果を示す

3.1 変調—搬送周波数マップ検出モデルの概要

3.1.1 変調—搬送周波数領域における検出

振幅変調マスキングの知見から工学モデルへの展開

前章で述べた振幅変調マスキングの知見を基に工学的なモデルを構築する．工学的に展開していく上で先の Houtgast [3] [4] らの聴覚心理実験実験の結果を基にしている．図 3.1 にその考え方を示す．

図 3.1 左，右共に信号と雑音の搬送周波数は Houtgast の心理聴覚実験で示した 1kHz である．ここでの例で示している変調周波数は，図 3.1 左は信号，雑音共に 8Hz で変調したとき，図 3.1 右は信号を 16Hz で，雑音を 8Hz で変調したと考えたときのものである．このように考えると先の Houtgast の聴覚心理実験において，信号成分 (signal component) と雑音成分 (noise component) の搬送周波数成分に重なりがある時であっても，変調周波数が異なるとマスキング量が減少し (図 3.1 右)，変調周波数が近ければ近いほどマスキング量が増加する (図 3.1 左) という現象をうまく説明できる．

このような変調周波数を検出するために，図 3.2 に示すような変調—搬送周波数マップを考える．これによって例えば図 3.3 に示されような，互いの搬送周波数に重なりがある

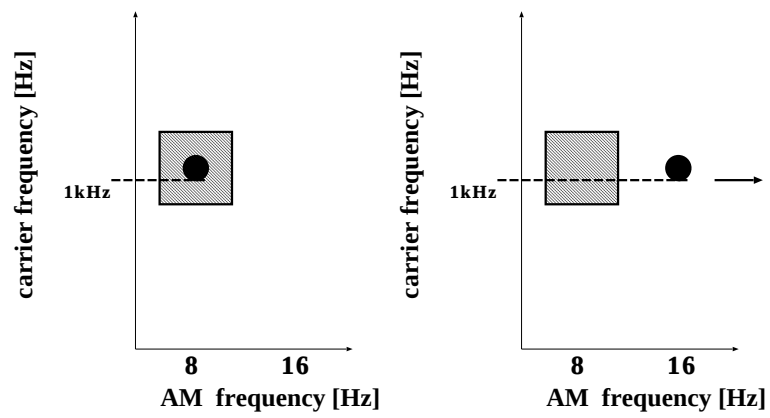


図 3.1: 振幅変調マスキングの概念図

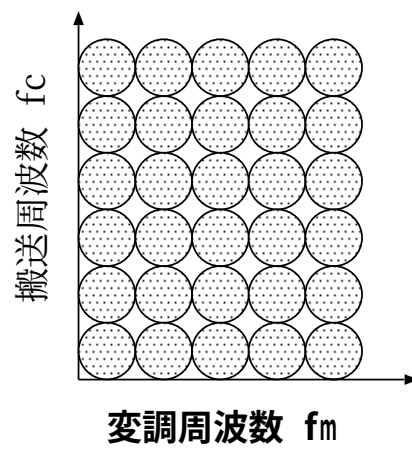


図 3.2: 変調—搬送周波数マップ

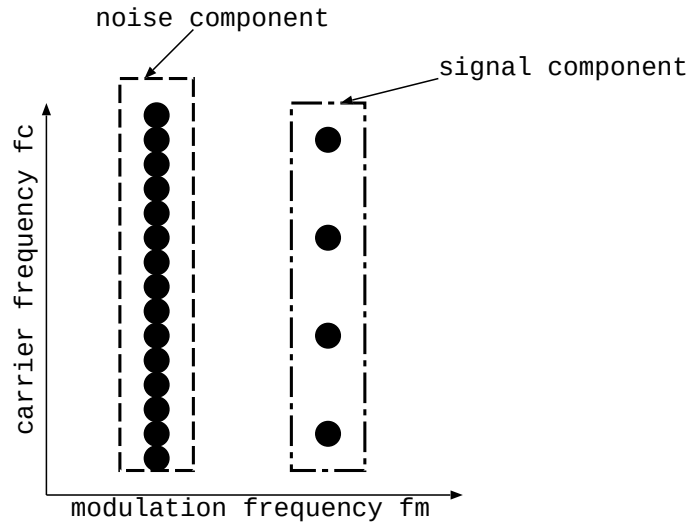


図 3.3: 変調—搬送周波数マップの例

ような2つの信号であっても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る場合に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ることが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. 4に構築した変調—搬送周波数検出モデルを示す。ここで FILTER BANK1 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 BANK2 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。

- 1. 入力信号を各搬送周波数 f_c に分けるために FILTER BANK1 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 BANK1 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

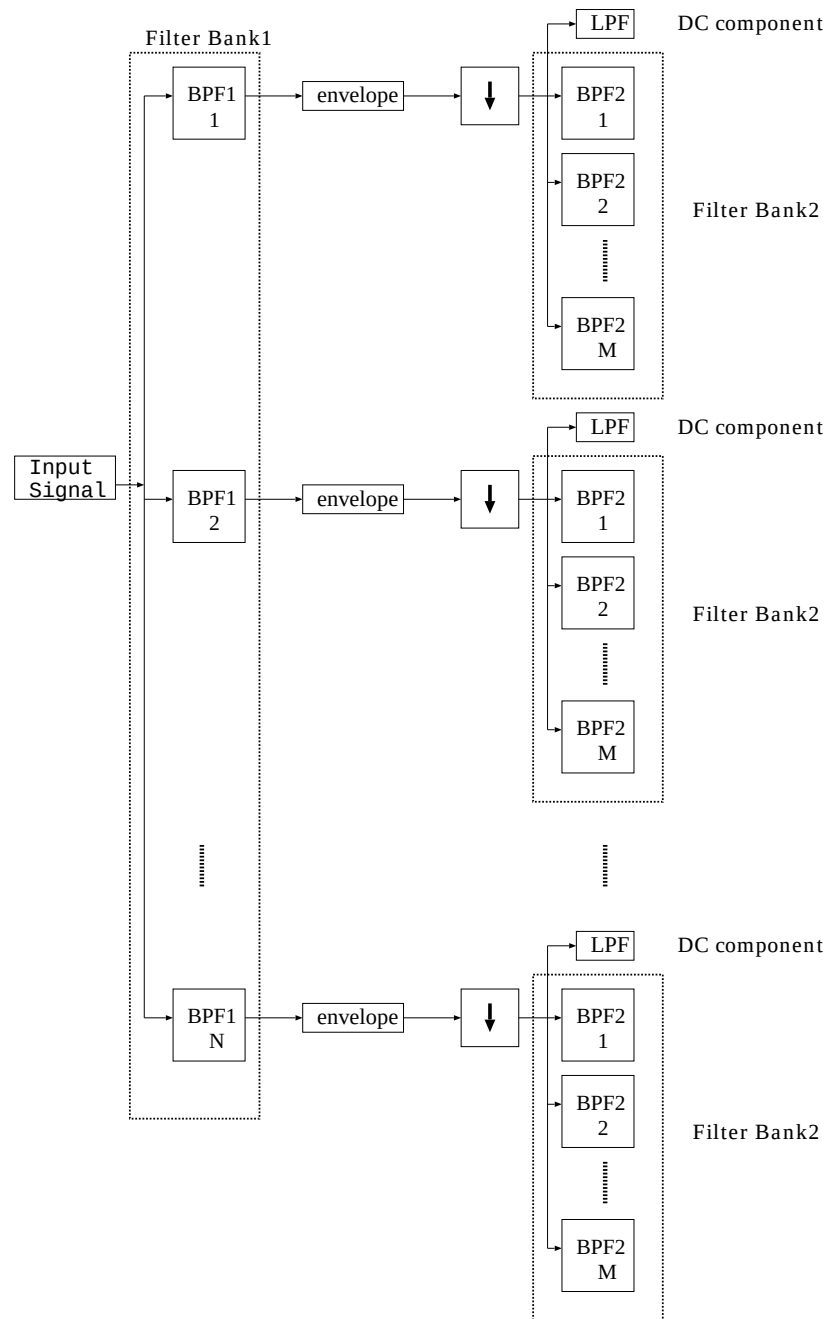


図 3.4: 全体の処理の概要図

- - 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 BANK2 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 BANK1 を通過する変調周波数成分が FILTER BANK1 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う.

3.1.2 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 について考察を行う. 搬送周波数のフィルタリングには FILTER BANK1, 変調周波数のフィルタリングには FILTER BANK2 の通過帯域や中心周波数などが問題となってくる. これについては

- 定帯域フィルタバンク
- 定Qフィルタバンク
- 定帯域フィルタバンクと定Qフィルタバンクの混合

3つ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.

定帯域フィルタバンク

定帯域フィルタバンクでは前章の TMTF (図 2. 2 より人間の聴覚において聞き分けやすい変調周波数が 100 Hz までであることを考慮し FILTER BANK2 の BPF の解析範囲を 100 Hz までとした, これに伴い FILTER BANK1 の BPF の帯域幅は中心周波数に関係なく 200 Hz まで一定とした. また, FILTER BANK2 の BPF の帯域幅は, ここでは処理時間及び信号の分離幅を考慮し 10 Hz とした. 以上のことをまとめると

- FILTER BANK1 (搬送周波数): BPF の帯域幅は一定で 200 Hz で中心周波数は BPF の帯域幅の重なりがないよう 200 Hz 間隔で解析範囲は 200 Hz - 4800 Hz
- FILTER BANK2 (変調周波数): BPF の帯域幅は一定 5 Hz で, 中心周波数は BPF の帯域幅の重なりがないよう 10 Hz 間隔で解析範囲は 100 Hz

となり, FILTER BANK1 の BPF の中心周波数に関係なく FILTER BANK2 における変調周波数の解析範囲は一定となる. すなわちここでは, すべての搬送周波数出力において 100 Hz までの変調周波数を解析する.

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た．

定Qフィルタバンクで構築された FILTER BANK2（変調周波数）における解析範囲は FILTER BANK1（搬送周波数）の帯域幅の半分となる．ここで FILTER BANK1 の帯域幅はある程度の幅がないと FILTER BANK2 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 BANK1（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 BANK2（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 BANK1 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 BANK2 の解析範囲も増加する．

本研究におけるフィルタバンク

フィルタバンクを形成する上で前節で述べた2つのBPFを考えたが，実際にはそれぞれのBPFは下記に示されるような欠点と利点がある．

- 定帯域フィルタバンク
 - － 利点：高い搬送周波数における解析が容易である．
 - － 欠点：低い中心周波数におけるBPFでの詳細な解析が行えず，信号の復元が困難である．
- 定Qフィルタバンク
 - － 利点：低い中心周波数における解析が容易である．
 - － 欠点：高い中心周波数における詳細な解析が容易である．信号の復元が容易である．

そこでもう1つのタイプのフィルタバンクとして，定帯域，定Qフィルタバンクの利点を活かし中心周波数が低い位置での周波数の高調波成分の解析は，定Qのフィルタバンクを

高い中心周波数では定帯域フィルタバンクを用いることを考える。ここでの定帯域フィルタバンクと定Qフィルタバンクの切替周波数は定Qフィルタバンクの帯域幅が定帯域フィルタバンクの帯域幅を越えるときの中心周波数とする。またここで定QフィルタバンクのBPF群はWavelet分析合成系で構築している。

3.1.3 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 [8] [9] [10] 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

次に以上の3つの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た。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Gabor関数を基にBPFを形成する。下記の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。

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

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 3.5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 20Hz で、フィルタの中心周波数は 400Hz である。また図 3.5 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

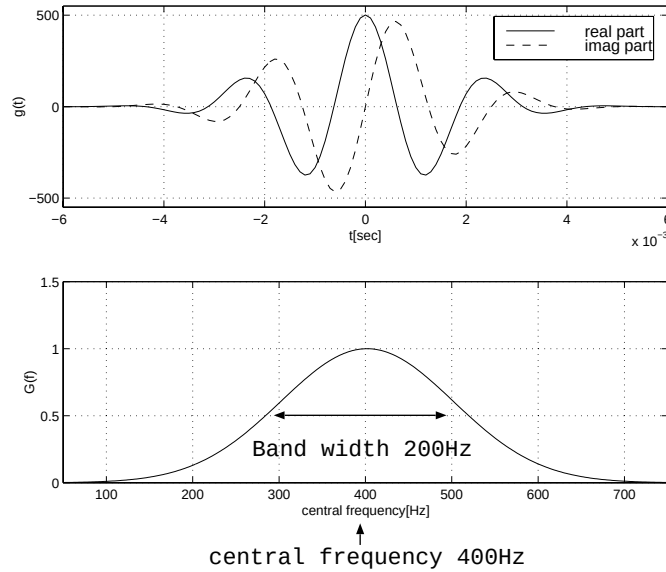


図 3.5: Gabor 関数による BPF（上：時間軸，下：周波数軸）

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が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．本研究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 [11] 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（チャンネル数）は 64 とした．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（チャンネル）は 64 とした．ここでフィルタバンク 2 の通過帯域の上限は先の前節 3.1.2 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．

3.2 変調—搬送周波数マップの検出のシミュレーション

本節では前説の変調—搬送周波数マップの検出シミュレーション結果を示す．

3.2.1 実験データ

本節では、実装した変調—搬送周波数検出マップにおいて振幅変調マスキングの知見に示されている、純音と雑音が同じ搬送周波数で、異なる変調周波数の時どのような結果になるかをシミュレーションを行った。

次に ATR 音声データベースを基に実際の音声の場合どのような変調がなされているかについてシミュレーションを行った。

AM 音とマスキング

実験データは次のものを用いた。信号として搬送周波数 1kHz の純音に変調深さ 0.5 で 50Hz の変調をかけ（図 3.6 上）、マスキングとして広帯域雑音（1Hz ～4800Hz）に変調深さ 0.5 で 30Hz で変調をかけたもの（図 3.6 中）についてのシミュレーションを行った。

図 3.6 下に解析した混合波形を示す。

音声

実験データとして ATR 音声データベースの母音の信号を用いた。それぞれの実験データを示す。図 3.7 に男性母音の /a/、図 3.8 に女性母音の /i/ である。また、図 3.9 はこの 2 つの母音を時間差 70msec で足し併せた信号である。すべての信号はサンプリング周波数 10kHz にダウンサンプリングした信号である。

3.2.2 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った。

ここで定Qフィルタバンクの変調周波数（ f_m ）、搬送周波数（ f_c ）の添字はそれぞれのフィルタバンクのチャンネル番号を表している。ここでは 1 が一番低い中心周波数の BPF（50Hz）で 64 が一番高い中心周波数の BPF（5000Hz）となっている。

AM 音とマスキングの検出結果

それぞれ図 3.10 が変調—搬送周波数マップの時間平均を定帯域フィルタバンクで求めた結果、図 3.11 が AM 音とマスキングの混合波形が定Qフィルタバンクで求めたものを示している。

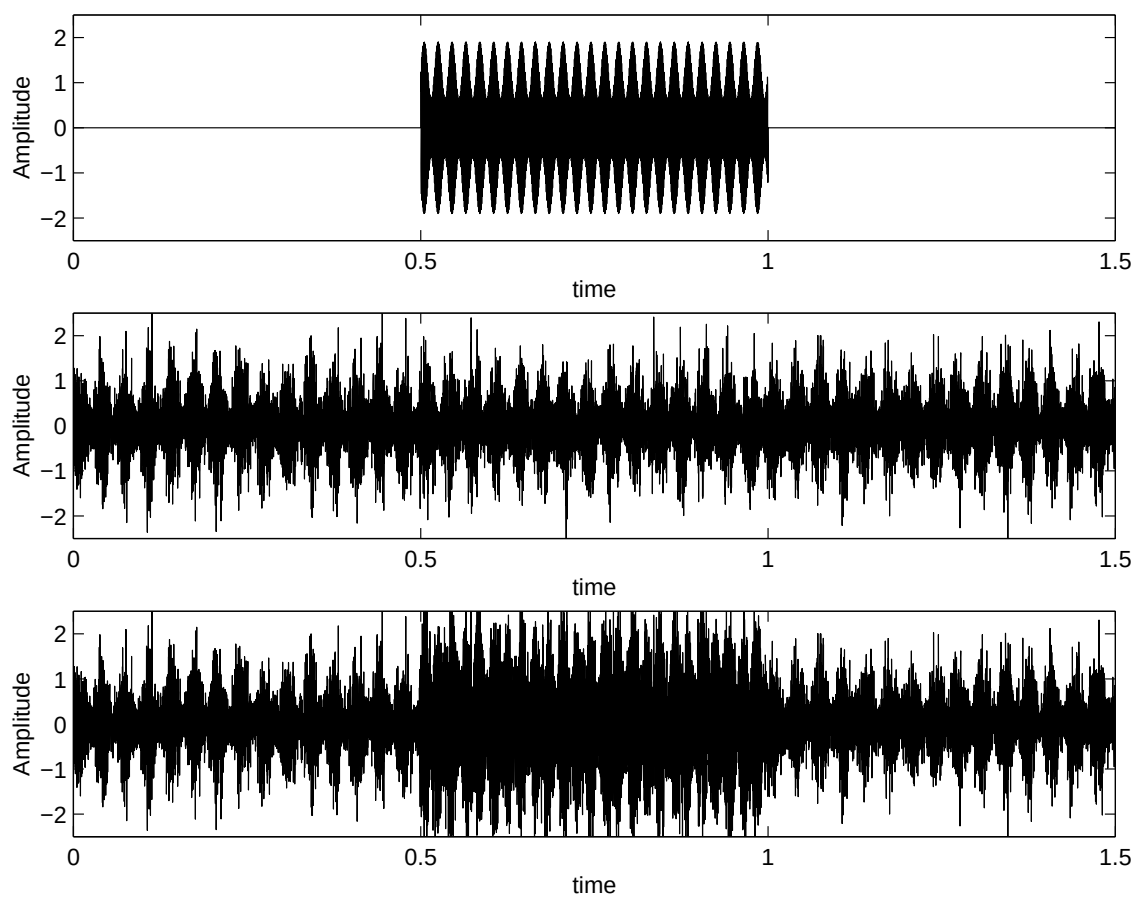


図 3.6: AM 音とマスクの波形（上：AM 音，中：マスク，下：混合波形）

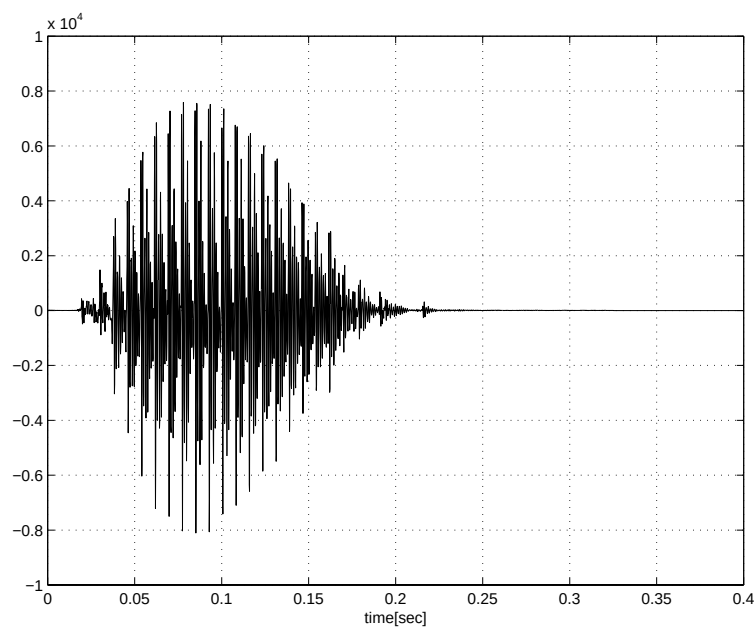


図 3.7: ATR 音声データベースの男性母音 mms の/a/

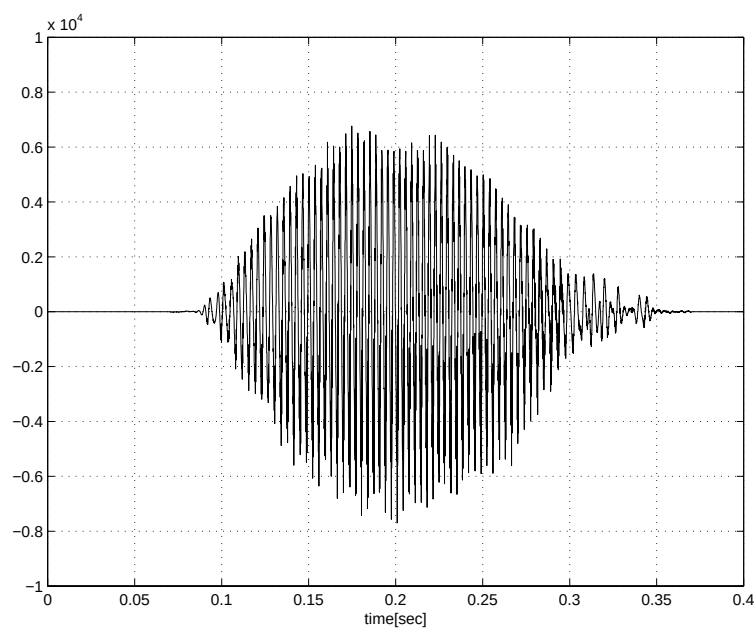


図 3 . 8ATR 音声データベースの女性母音 fsu の/i/

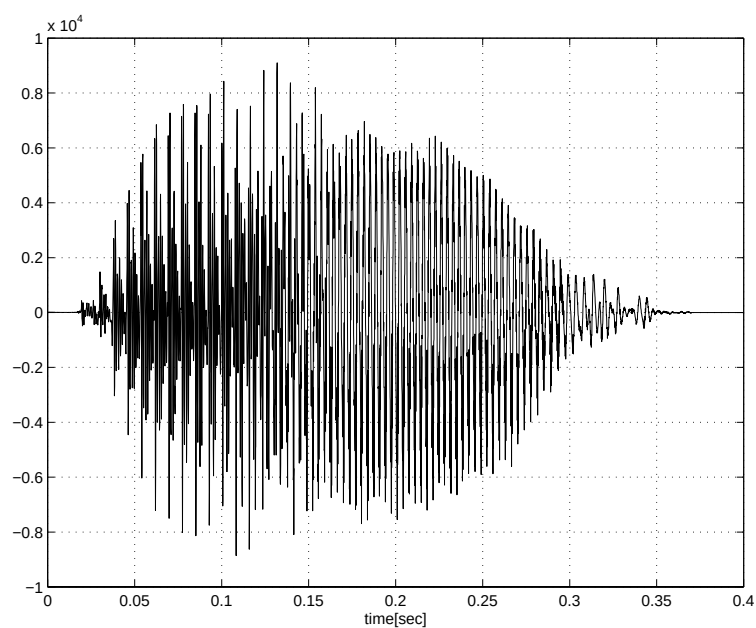


図 3.9: mms の/a/と fs uの iの混合波形

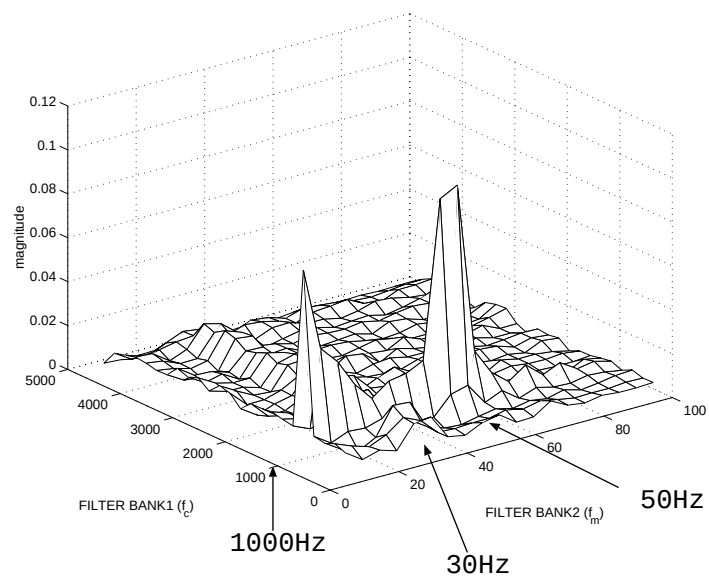


図 3. 10: AM 音とマスキャーの変調—搬送周波数マップの検出結果（定帯域フィルタバンク）

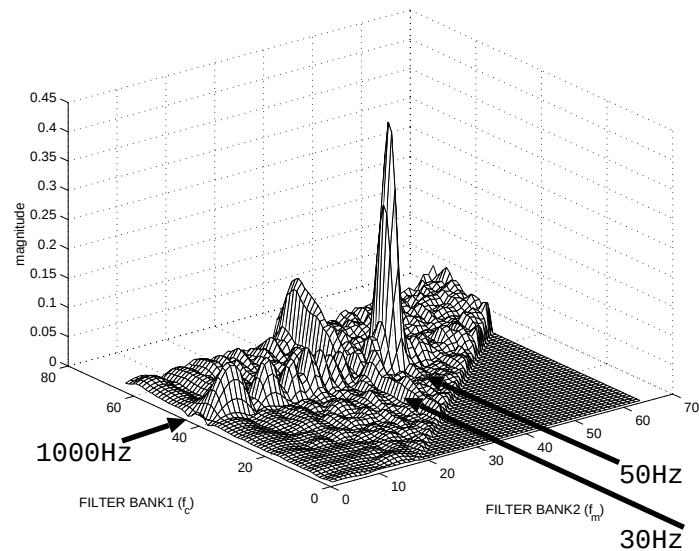


図 3.11: AM 音とマスキャーの変調—搬送周波数マップの検出結果（定Qフィルタバンク）

音声の検出結果

先の3つの音声信号の変調—搬送周波数マップの検出結果を示す．ここでは検出した変調—搬送周波数マップの時間変化を平均したものを示している．

mms の/a/の解析

それぞれ図 3.12が mms の/a/ の変調—搬送周波数マップの時間平均を定帯域フィルタバンクで求めたもの、図 3.13が定Qフィルタバンクを示している．

fsu の/i/の解析結果

図 3.14は fsu の/i/の変調—搬送周波数マップの時間平均を定帯域フィルタバンクで求めたもの、図 3.15が定Qフィルタバンクを示している．

mms の/aとfsuの/iの混合信号の解析

図 3.16が mms の/a/と fsuの/i/の混合信号の変調—搬送周波数マップの時間平均を定帯域フィルタバンクで求めたもの、図 3.17が定Qフィルタバンクの結果を示している．

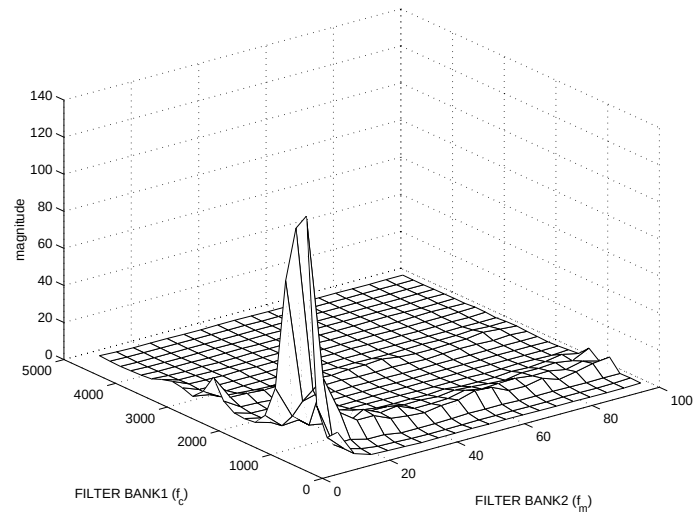


図 3.12: mms の/a/の変調—搬送周波数マップの検出結果（定帯域フィルタバンク）

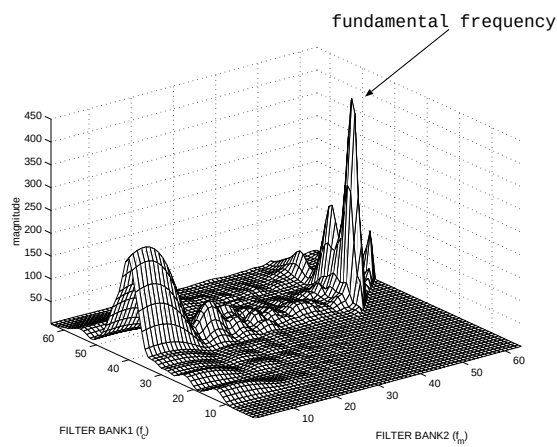


図 3.13: ms の/a/の変調—搬送周波数マップの検出結果（定Qフィルタバンク）

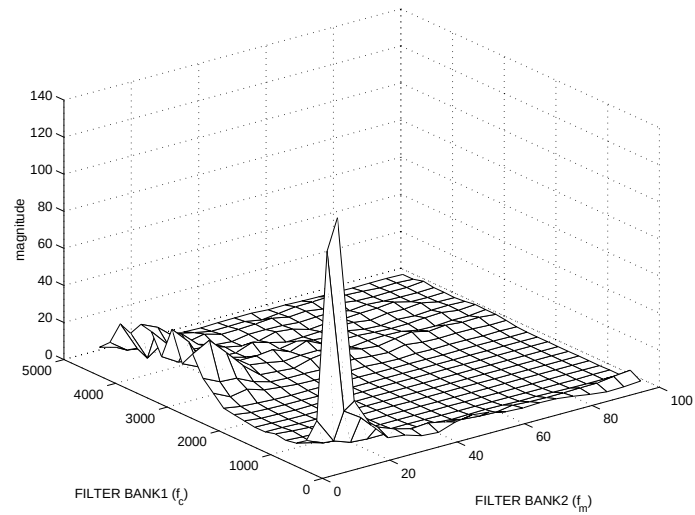


図 3.14: fsu の/i/の変調—搬送周波数マップの検出結果（定帯域フィルタバンク）

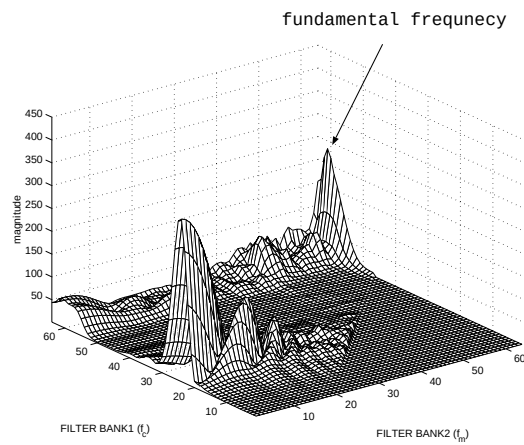


図 3.15: fsu の/i/の変調—搬送周波数マップの検出結果（定Qフィルタバンク）

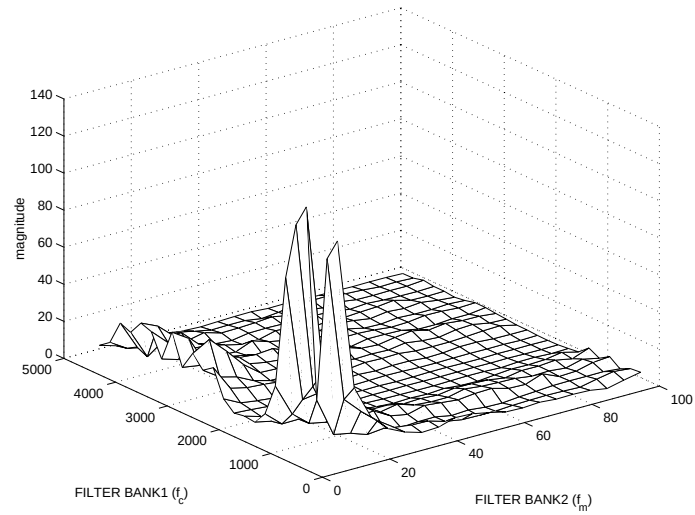


図 3.16: mms の/a/と fs uの/i/の混合信号の変調—搬送周波数マップの検出結果（定帯域フィルタバンク）

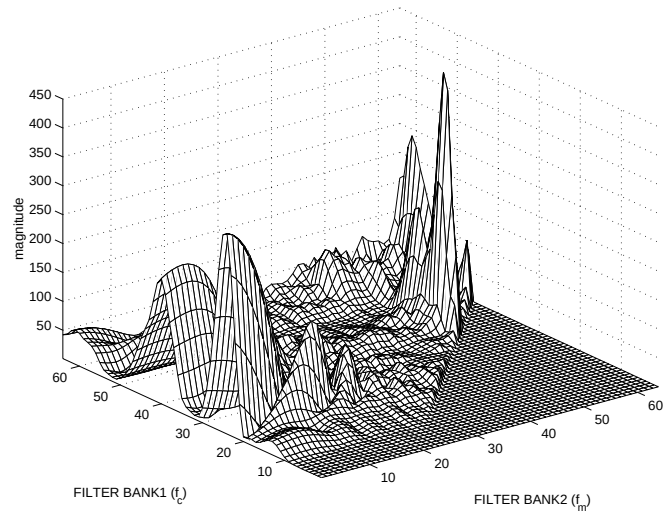


図 3.17: mms の/a/と fsuの/i/の混合信号の変調—搬送周波数マップの検出結果（定Qフィルタバンク）

3.2.3 考察

AM 音とマスキングの検出結果の考察

この検出結果より、図 3.10 の定帯域フィルタバンク、図 3.11 の定 Q フィルタバンク共に 2 つのピークが得られた。2 つのピークのうち 1 つは信号成分である FILTER BANK1 (f_c) の 1kHz, FILTER BANK2 (f_m) の 50Hz の位置に、もう 1 つのピークは雑音成分である FILTER BANK1 (f_c) の全体, FILTER BANK2 (f_m) 30 Hz の位置に得られた。

この結果は作成したシミュレーション波形の搬送周波数 1 KHz の純音を変調周波数 50Hz で変調した信号、帯域雑音 (1Hz ~ 4800Hz) に変調周波数 30Hz で変調したマスキングに一致しており、先の節で述べた振幅変調マスキングの知見を工学的に捕え直した図 3.3 のような形で検出されることが確認できた。

音声の検出結果の考察

今回の変調—搬送周波数マップの検出結果より、定帯域フィルタバンクの解析においては変調周波数の変化はあまり得られなかった。また、帯域幅と中心周波数が等間隔であるため音声など比較的低い中心周波数の搬送波成分に情報が多いためでは十分な解析が行えなかった。

一方、定 Q フィルタバンクの変調—搬送周波数マップの検出結果からは、高い中心周波数の変調周波数 f_m 軸にはそれぞれの音声の基本周波数成分が表れている。

図 3.18 に示しているように一般的に音声の基本周波数の分布は男声、女声それぞれ対

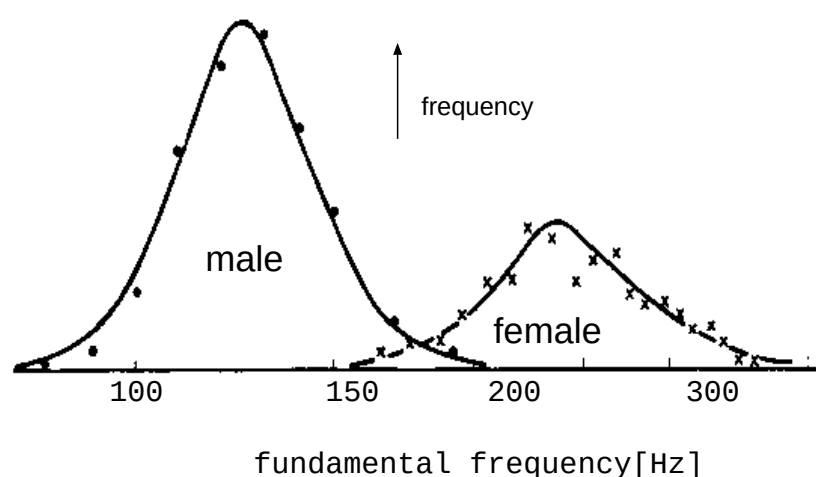


図 3.18: 発声者による基本周波数の分布

数周波数軸上で正規分布となり，男声の基本周波数の平均値と標準偏差はそれぞれ 125Hz および 205Hz となっている [13].

そこで定Qフィルタバックを使って変調—搬送周波数マップを検出することにより，図 3.18の範囲の変調周波数成分を検知することで音声の基本周波数成分を調べることができ，検出した基本周波数成分を基にこれを整数倍することで調波倍音成分を検出することが可能になる.

以上のことより今後の解析においては定Qフィルタバンクの変調—搬送周波数マップの検出を行い，先の節で示したように搬送周波数成分の高い高調波成分の解析には定帯域フィルタバンクを用いることとする.

第 4 章

AM 音とマスキアーの混合信号の分離・抽出とシミュレーション

本章では，振幅変調マスキングの知見の聴覚心理実験で用いられている純音と帯域雑音にそれぞれ異なる変調周波数を掛けた AM 音マスキアー信号の工学的な分離・抽出法について説明する．

4.1 AM 音とマスキアーの分離・抽出

信号の分離・抽出に必要な成分としては，それぞれの信号の搬送周波数，AM 周波数，DC 成分である．ここで搬送周波数は Filter Bank1 の出力位置から，AM 周波数は Filter Bank2 の出力位置から求める．しかし DC 成分は図 4.1(a) に示すように同じ搬送周波数上に 2 つの変調周波数（図中：O，X）が存在すると先の LPF で求めた DC 成分上にはこの 2 つの大きさが足されたものになってしまう．そこで，この 2 つの信号の DC 成分の分離をする必要がある．

従ってそれぞれの信号の抽出を行うには 2 つの信号の DC 成分を分離する必要がある．本研究では 2 章で述べた OLD-PLUS- NEW 発見則の知見を用いてこの 2 つの信号の DC 成分の分離を行っている．次に DC 成分の分離について説明する．

4.1.1 DC 成分の分離

ここでは先の OLD- PLUS- NEW 発見則の”古い音が新しい音に足されたように知覚する”という考え方を基に 2 つの成分の分離を行う．

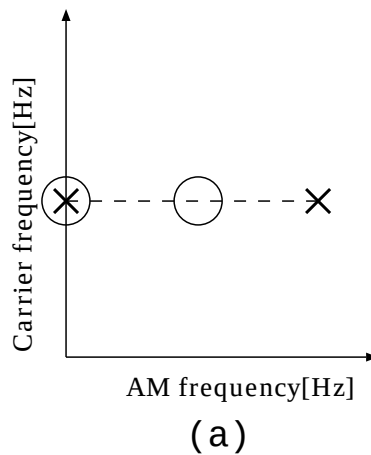


図 4.1: LPF 出力（重なりのある DC 成分）

図 4.2にその概要を示す。(a) に示すように2つの DC 成分が存在する場合，新しく入ってきた音（new sound）は最初に入ってきた古い音（old sound）に足されたように知覚されるということを利用して，図中 (b)，(c) に示すように最初に存在する信号の大きさを最初に入ってきた信号の DC 成分（A2）とし，新しく入ってきた音の DC 成分の大きさを新しく入った音の DC 成分と古い音の DC 成分から引いたものとする（A1）。

4.2 分離・抽出のシミュレーション結果

以上の考え方をを用いて先の 図 3.6の AM 音とマスキ어의混合信号の分離・抽出を行った。分離・抽出結果を図 4.3に示す。

4.2.1 考察

このシミュレーションの結果，若干の立ち上がりに誤差があるものの AM 音とマスキ어의分離が行えた。これによって搬送周波数成分に重なりがあっても変調周波数の差を見ることが分離が行えることが確認できた。

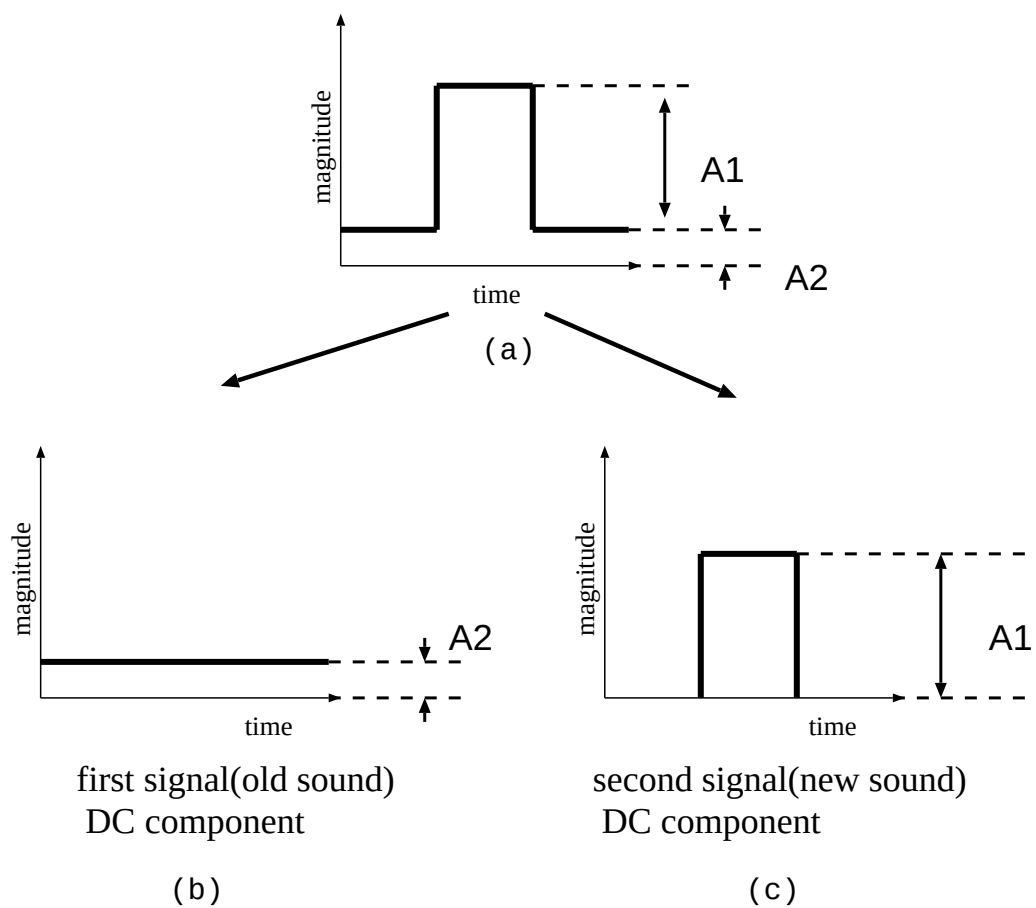


図 4.2: DC 成分の分離

EPS File ./thesis-~~lekka~~/new-
chap/si n50- noi 30-
separate. eps. psnot
found

図 4. 3: AM 音とマスキング混合信号の分離・抽出結果

第 5 章

二重母音の分離・抽出とシミュレーション

本章では検出された変調—搬送周波数マップの時間変化を基に二重母音の分離抽出法について説明する．さらに実際にシミュレーション音声を作成し，2つの音の分離抽出結果を示す

5.1 二重母音の分離・抽出

本研究では先の変調—搬送周波数マップで検出された時間変化と基本周波数成分を次の聴覚の情景解析の考え方を基にした3つの手法を用いて音源分離を行っている．これについて次に先の図 3.17の例を用いながら述べる．

- 調波関係を用いた音源分離
- 音の共通の立ち上がり／立ち下がりを用いた音源分離
- OLD-PLUS- NEW発見則を用いた DC 成分の分離

5.1.1 調波関係を用いた音源分離

この手法は，第2章で述べた”繰り返し周期で音が振動するとき，その変動は周波数成分が共通な基本周波数の整数倍になるような聴覚パターンを発生する”するという規則に基づいている．この考え方を基に先の変調—搬送周波数マップの検出結果より基本周波数を求め，調波倍音の関係より harmonic sieve を行って音を分離する．

例えば，図 5.1(a) に示すように2つの調波関係にある2つの音が存在するとき，図 (a) に示すようにそれぞれの音の基本周波数成分 f_0 を求めることが出来れば，それぞれの基

本周波数 f_0 を整数倍することで，ここでは (b)，(c) に示すようにそれぞれの基本周波数 f_0 を持つ成分に分離することができる．

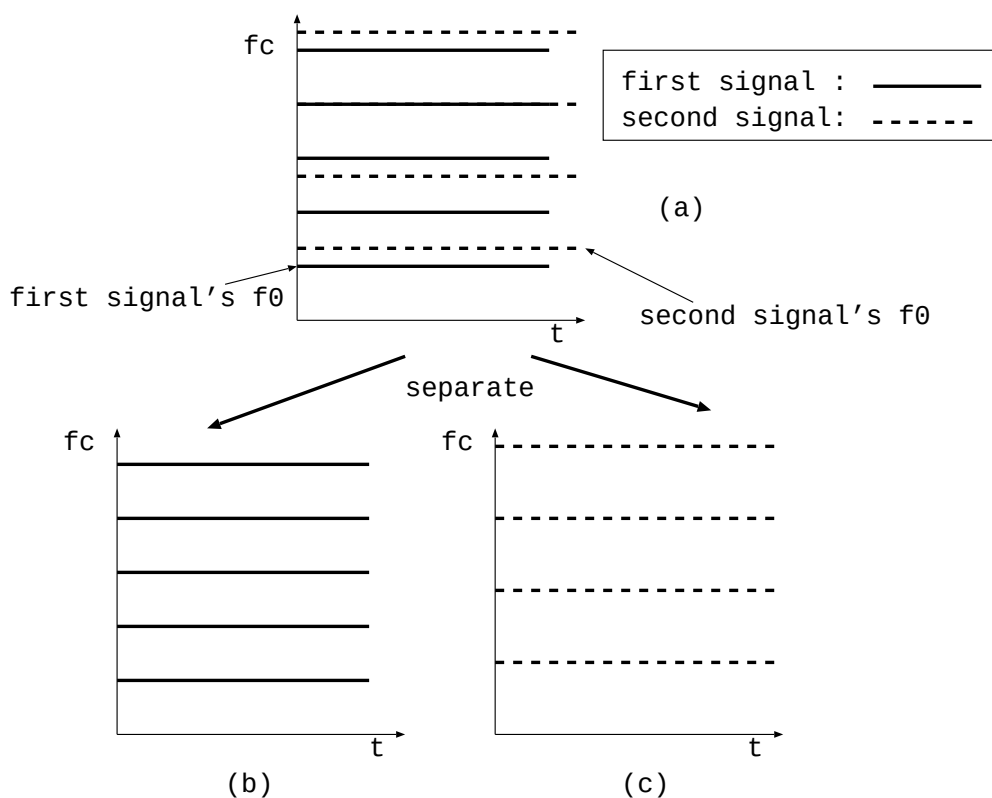


図 5.1: 調波関係を用いた音源分離の概念図

図 5.2 は図 3.17 の結果である．図 5.2 に示すように，ある時間 Δt において変調—搬送周波数マップよりそれぞれの基本周波数成分を求める．次にこの結果を基に harmonic sieve を行う．ここで先のフィルタバンクの節で述べたように，harmonic sieve を行うのに定 Q フィルタバンクのみでは高い中心周波数の BPF では BPF 出力がぼやけてしまい正確な harmonic sieve を行うことが出来ない．そこで，harmonic sieve を行うために高い中心周波数では定帯域フィルタバンクを，低い中心周波数では定 Q フィルタバンクを用いて f_c 軸に対して harmonic sieve を行う．

図 5.3 に harmonic sieve の具体例を示す．

図 5.3 (b) (c) に示すように抽出した基本周波数成分より調波倍音の関係を調べ，ある時間 Δt における (a) の混合波形より (d), (e) の 2 つの成分に分離する．

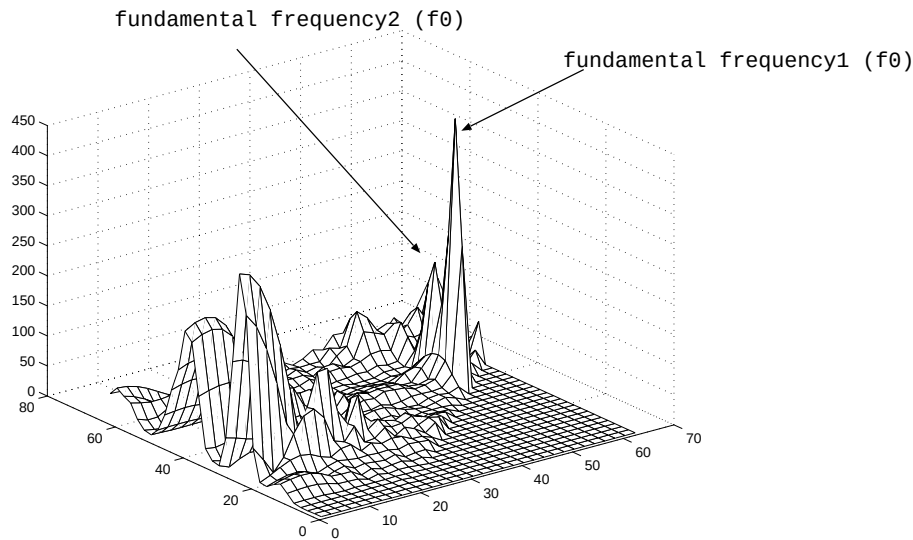


図 5.2: 基本周波数成分の推定

5.1.2 音の共通の立ち上がり／立ち下がりを用いた音源分離

ここでは、先の背景で述べた”無関係な音は、同時に始まったり終わったりしない”という考えを基に音源分離を行う。

搬送周波数

図 5. 4に搬送周波数における時間変化の概念図を示す。 (a) に示されるように2つの立ち上がり／立ち下がりの異なる音が存在する場合に、その立ち上がり／立ち下がりの情報を利用して、それぞれ共通の立ち上がり／立ち下がりを持つ成分図 (b), (c) に示すように分離する。

変調周波数

前節は搬送周波数領域での立ち上がり／立ち下がりを利用した音源分離であったが、同様に変調—搬送周波数マップ上でも同じように立ち上がり／立ち下がりを見て共通の立ち上がり／立ち下がりを持つ2つの成分に分離する。

図 5. 6は変調—搬送周波数マップの時間変化検出の概念図を示す。 図での立ち上がり／立ち下がり、(a) の部分の立ち上がり、(b) の部分の立ち上がり、(a) の部分の立ち下がり、(b) の部分の立ち下がりという時間変化をしている。 このような場合にこの立ち上がり／立ち下がりを利用して (a) と (b) という2つの成分に分離する。

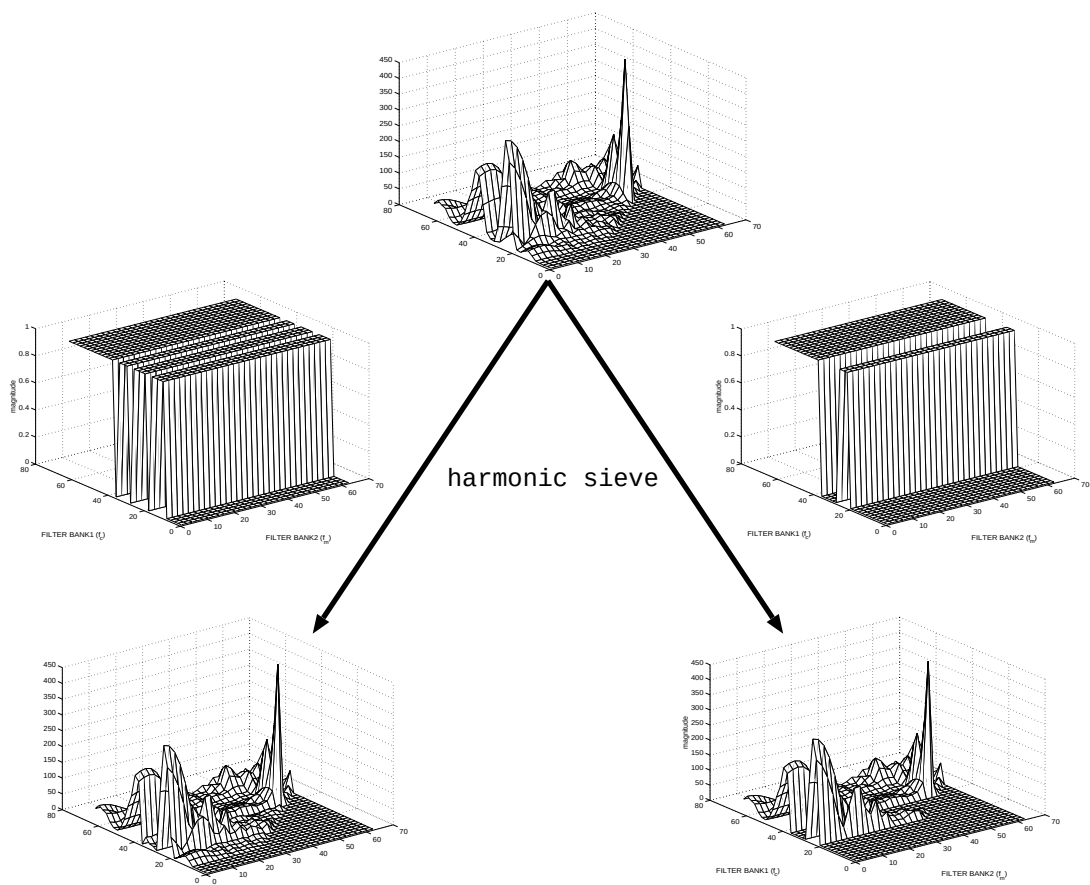


图 5.3: harmonic sieve

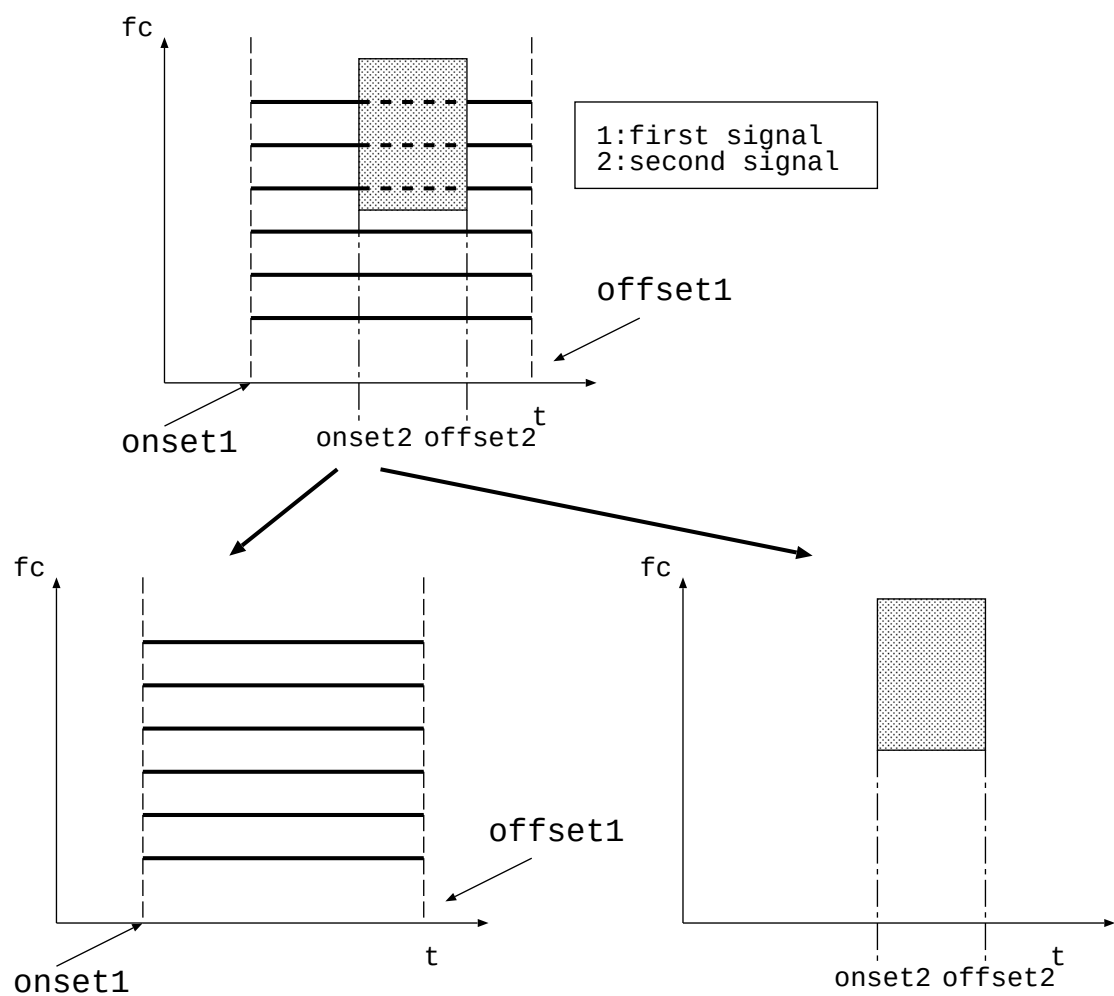


図 5.4: 共通の立ち上がり立ち下がりの音源分離の概念図

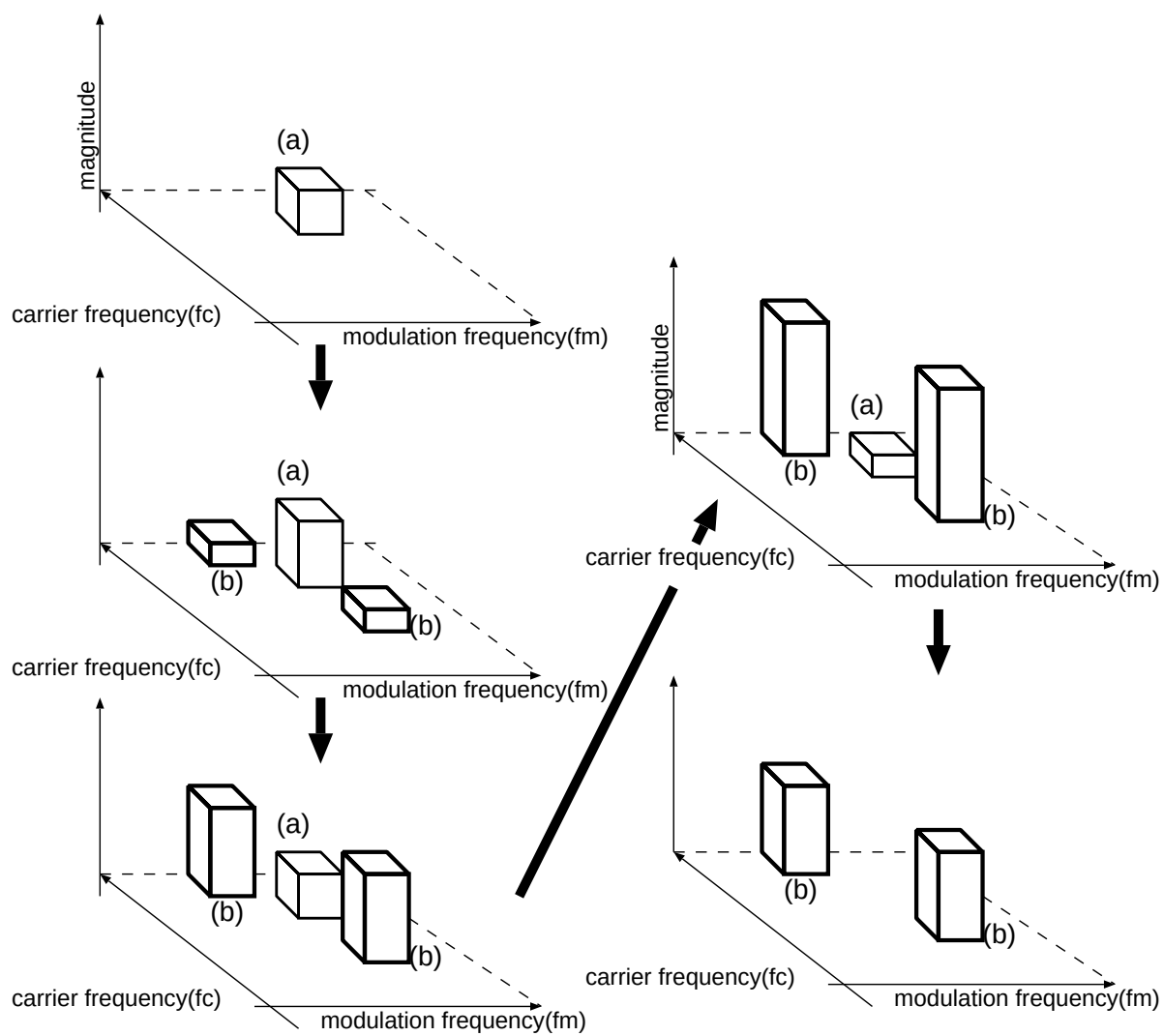


図 5.5: 変調—搬送周波数マップでの立ち上がり／立ち下りの概念図

図 5.6は先の 3.17 の時間変化を示している．図では (a) — (f) に時間変化している．この図の中で，まず図 5.6(b) において，ある音（図中：NO1 ONSET）が立ち上がり，(d) において立ち下がってきている（図中：NO1 OFFSET）．また，(d) において，別のある音（図中：NO2 ONSET）が立ち上がってきて (e) において立ち下がっている（図中：NO2 OFFSET）．このような2つの別の立ち上がり／立ち下がりが存在する場合に共通の立ち上がり／立ち下がりを持つ物に分離する．ここでは NO1 と NO2 の2つの成分に分離することになる．

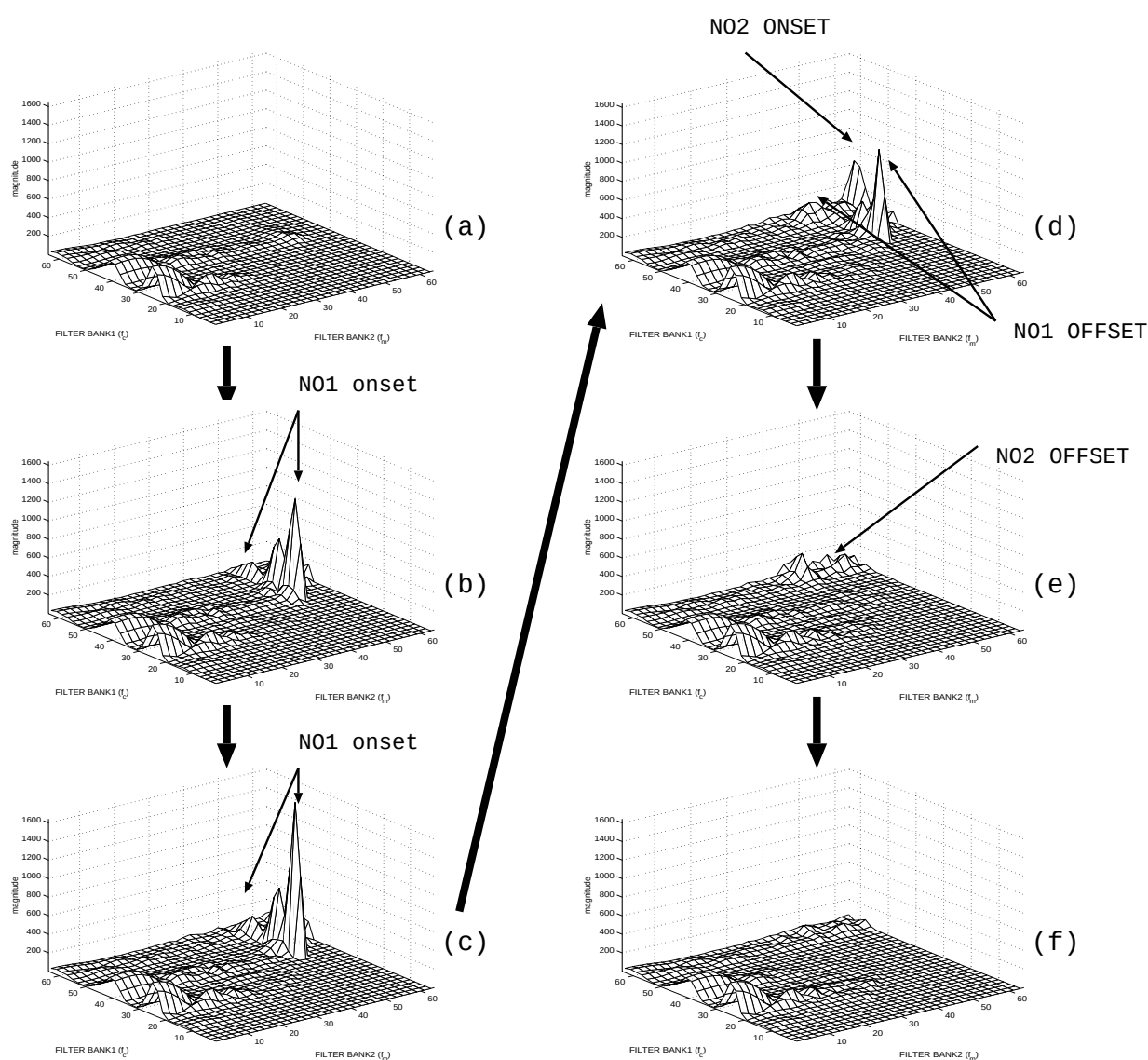


図 5. 6: mms の/a/と fsuの/i/の時間変化

5.2 二重母音の分離・抽出のシミュレーション

本節では前説で述べた調波関係を用いた音源分離，共通の立ち上がり／立ち下がりを用いた音源分離，DC成分の分離の実装したものを基に先の時間変化する検出された変調—搬送周波数マップの結果を使って，音源分離のシミュレーション結果を示す．

5.2.1 分離・抽出結果の評価法

ここでシミュレーションを行うために以下に示すSD法（スペクトル歪み）を用いてSD値で測定している．ここではフレーム長 366ms でハニング窓で切り出した波形を以下に示す式で計算している．またここでフレーム周期は 188ms である

$$SD = 10 \log \frac{1}{\omega} \int_0^{\omega} (|S(\omega)| - |\hat{S}(\omega)|)^2 d\omega$$

$S(\omega)$ ：原信号のスペクトル

$\hat{S}(\omega)$ ：解析信号のスペクトル

5.2.2 シミュレーション結果

実験に用いたのは先ほどの図 3.9の mms の/a/と fsu の/i/の混合波形である．図 5.7 に mms の/a/ と fsu の/i/ のそれぞれのスペクトル，図 5.8 に mms の/a/ と fsu の/i/ を混合した波形のスペクトル，図 5.9 に分離したあとのスペクトル，図 5.10に分離した波形を示す．分離する前と分離した後の平均SD値を示す．

- 分離前

- mms の/a/ 平均SD値：37.3[dB]

- fsuの/i/ 平均SD値：30.9[dB]

- 分離後

- mms の/a/ 平均SD値：18.9[dB]

- fsuの/i/ 平均SD値：18.3[dB]

- 改善量

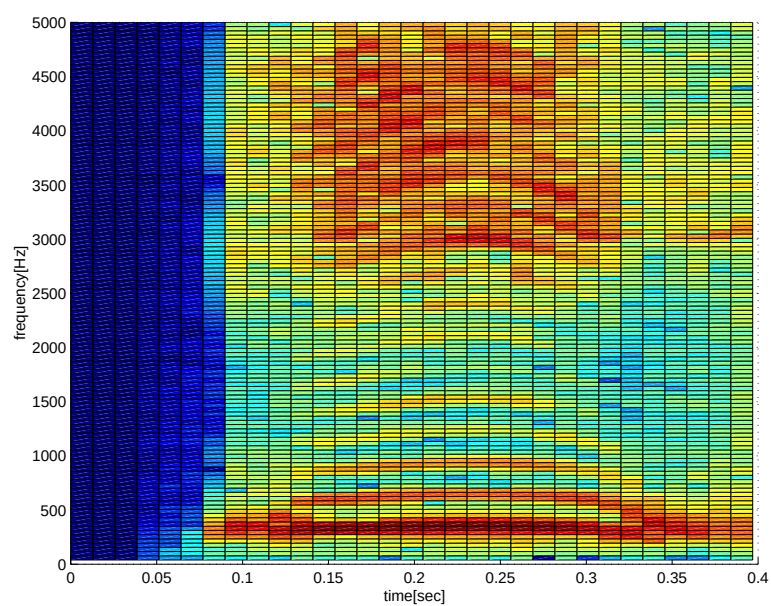
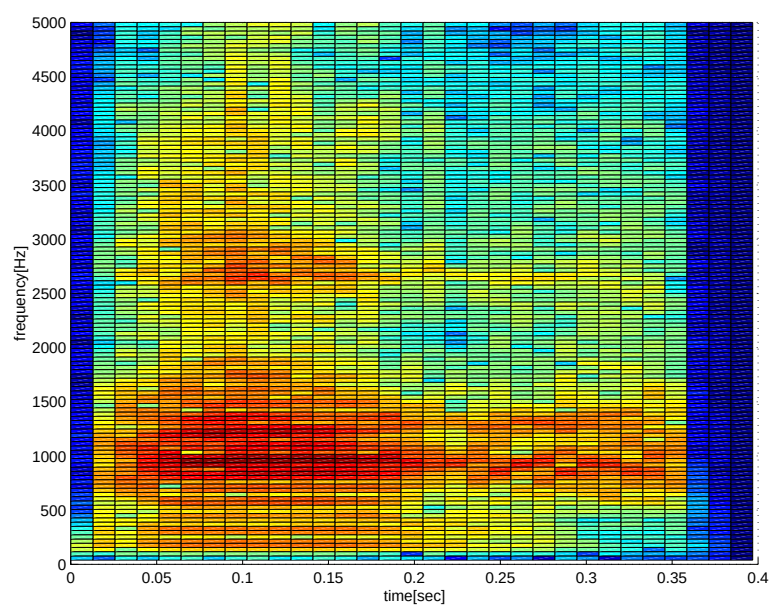


図 5.7: mms の/a/と fs uの/i/のスペクトル (上: mms の/a,/ 下: f s uの/i/)

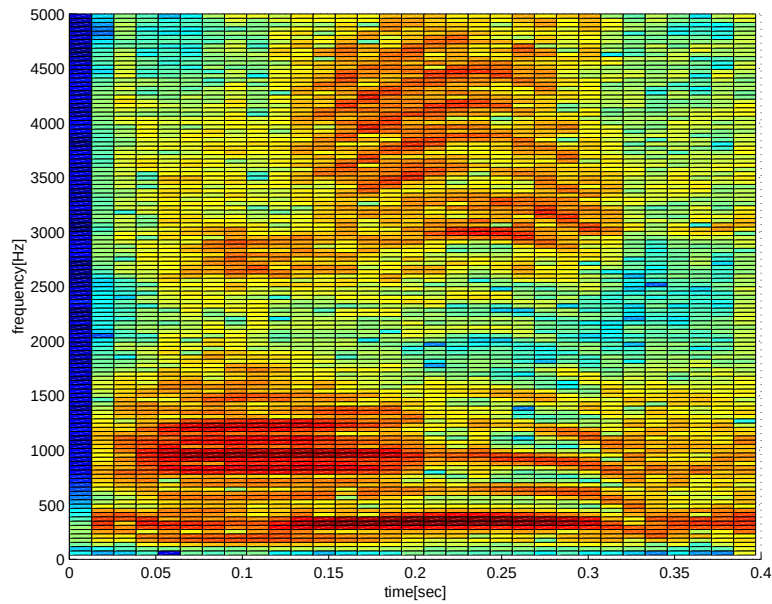


図 5.8: mms の/a/と fsu の/i/の混合波形のスペクトル

- mms の/a/ 平均 SD 値：18.4[dB]
- fsu の/i/ 平均 SD 値：12.6[dB]

次に図 5.11に示す mms の/a/と fsu の/u/を時間差 50ms で混合した波形について解析を行った。

図 5.7に mms の/a/と fsu の/u/のそれぞれのスペクトル，図 5.13に mms の/a/と fsu の/u/を混合した波形のスペクトル，図 5.14に分離したあとのスペクトル，図 5.15に分離した波形を示す。

- 分離前
 - mms の/a/ 平均 SD 値：25.3[dB]
 - fsu の/u/ 平均 SD 値：33.4[dB]
- 分離後
 - mms の/a/ 平均 SD 値：16.3[dB]
 - fsu の/u/ 平均 SD 値：17.7[dB]
- 改善量

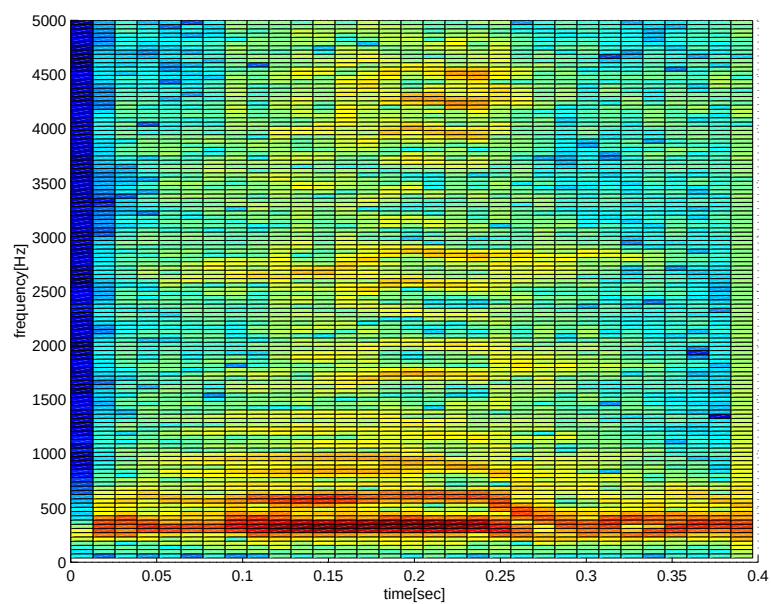
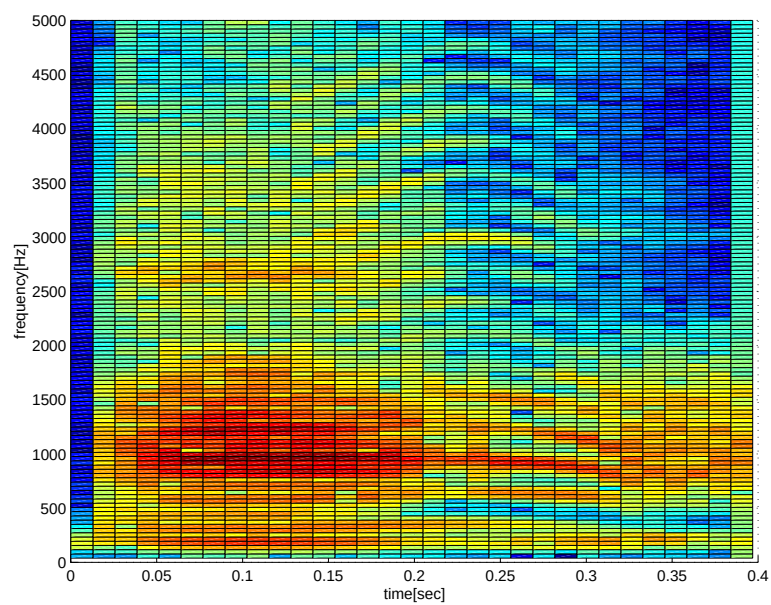


図 5.9: mms の/a/と fsu の/i/の分離したスペクトル (上: mms の/a/, 下: fsu の/i/)

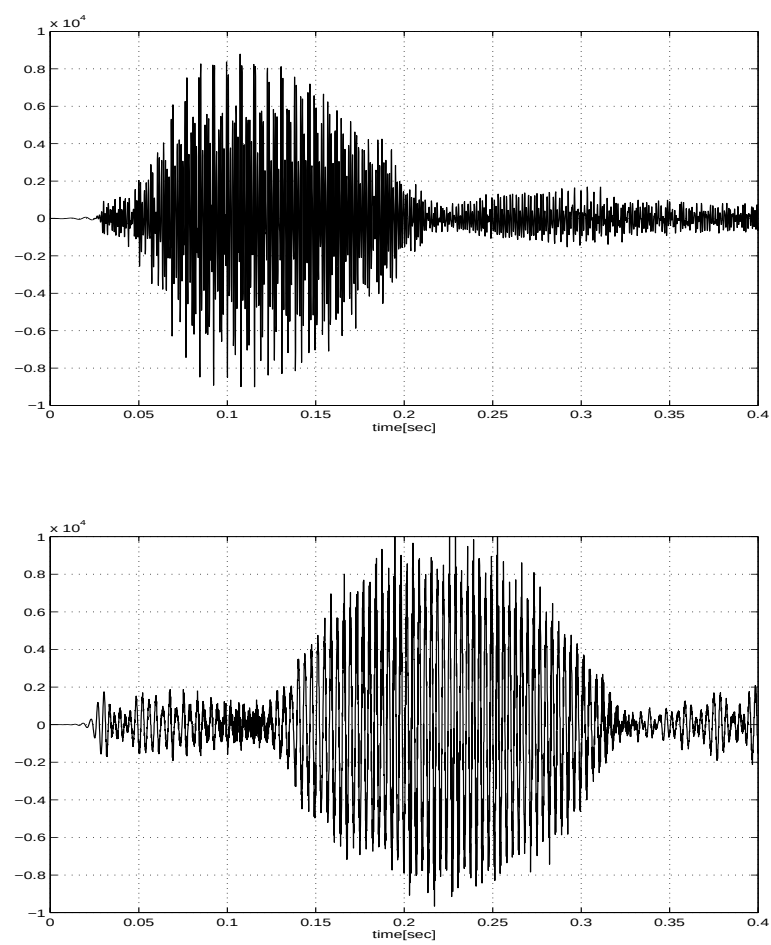


図 5.10: mms の/a/と fs uの/i/の分離結果波形（上：mms の/a/, 下：f s の/i/）

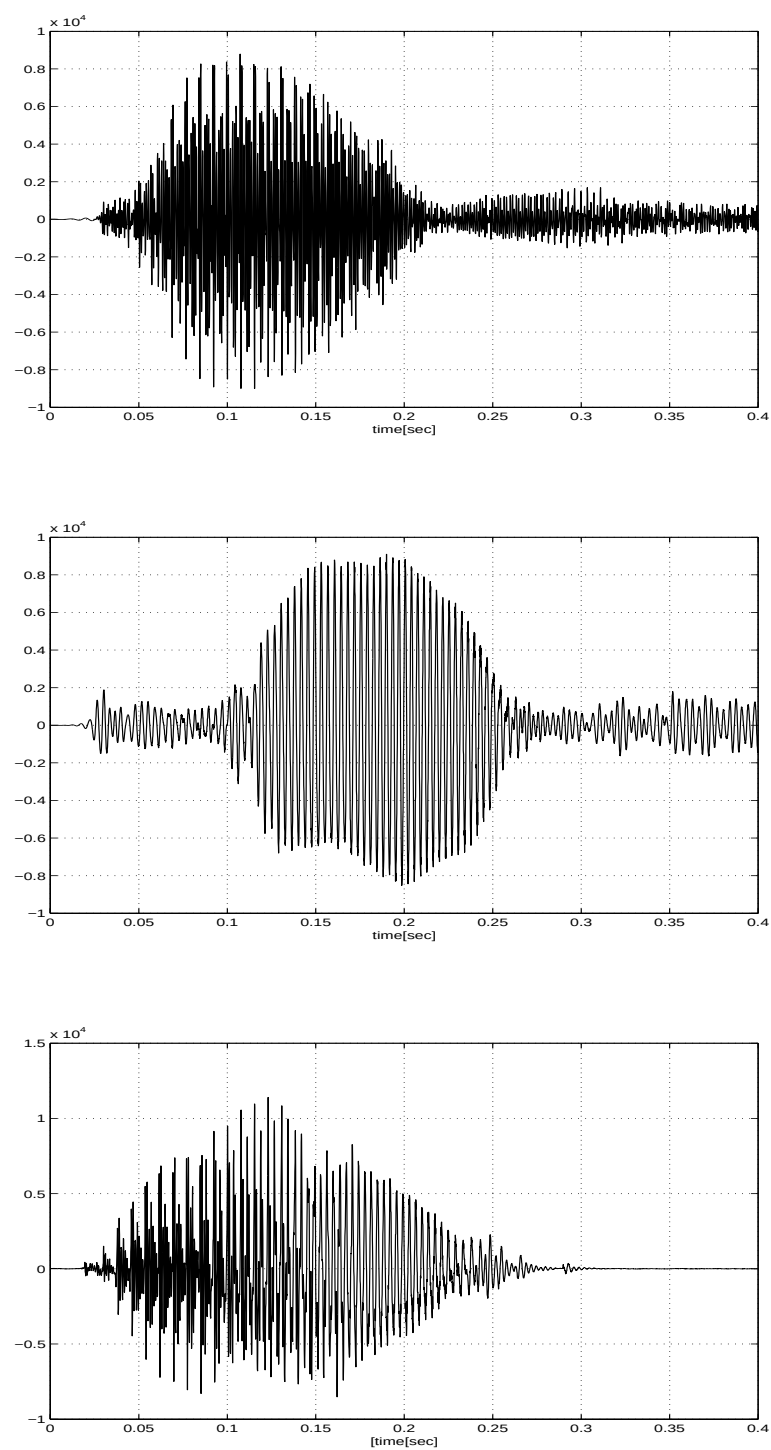


図 5.11: mms の/a/と fsu の/u/の波形（上： mms の/a/, 中： fsu の/u/, 下：混合波形

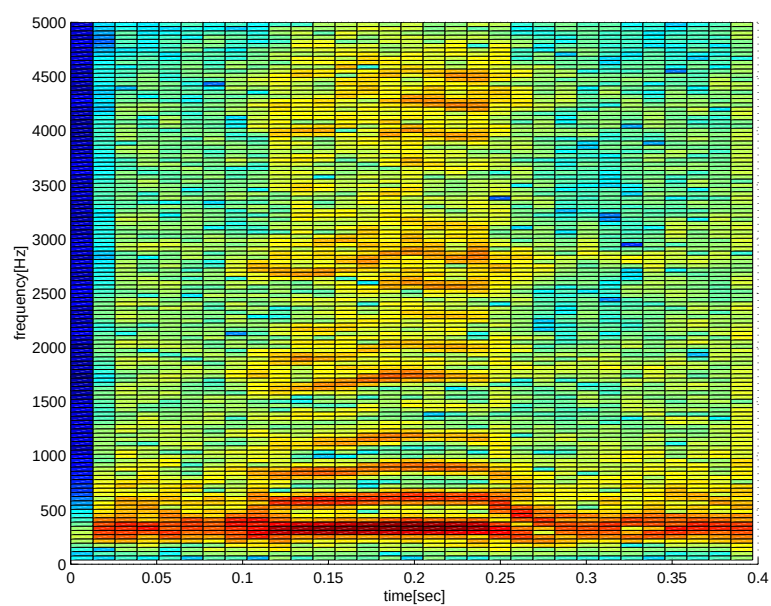
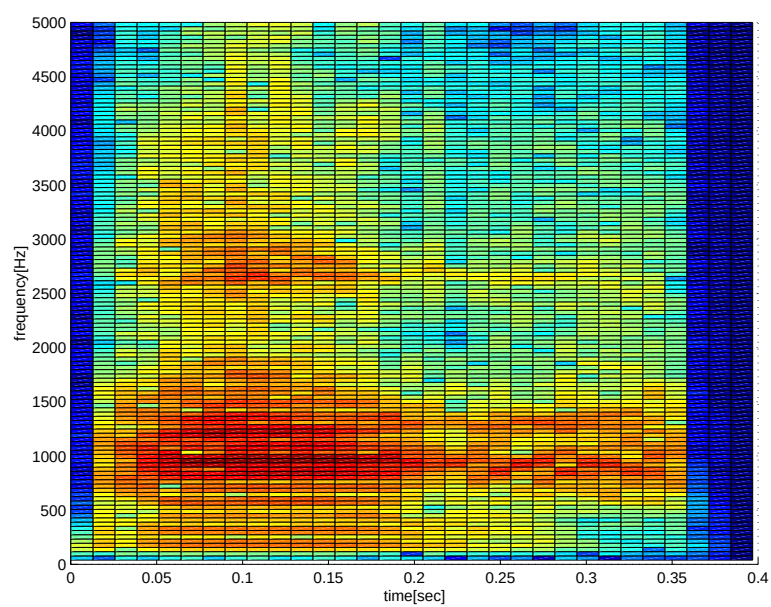


図 5.12: mms の/a/と fsu の/u/のスペクトル (上: mms の/a/, 下: fsu の/u/)

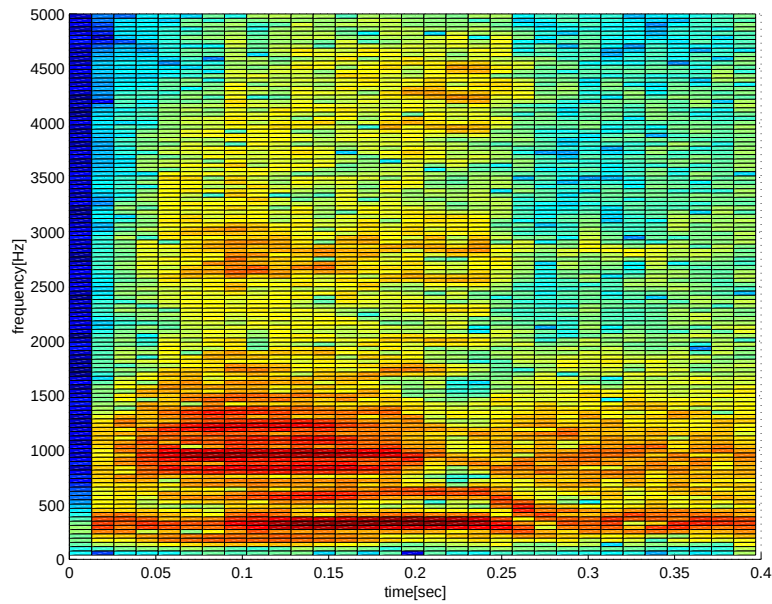


図 5.13: mms の/a/と fsu の/u/ の混合スペクトル（上：mms の/a/, 下：fsu の/u/）

- mms の/a/ 平均 SD 値：9[dB]
- fsu の/u/ 平均 SD 値：16.1[dB]

5.2.3 考察

このシミュレーションによって変調—搬送周波数マップ及び Bregman の規則を用いることで二重母音の分離が行えることが確認された。この結果において改善量に差があるのは変調—搬送周波数領域においても混合した信号の波形にかなりの重なりがあったもので、改善量が良いものは変調—搬送周波数領域において重なりの少ないものである。

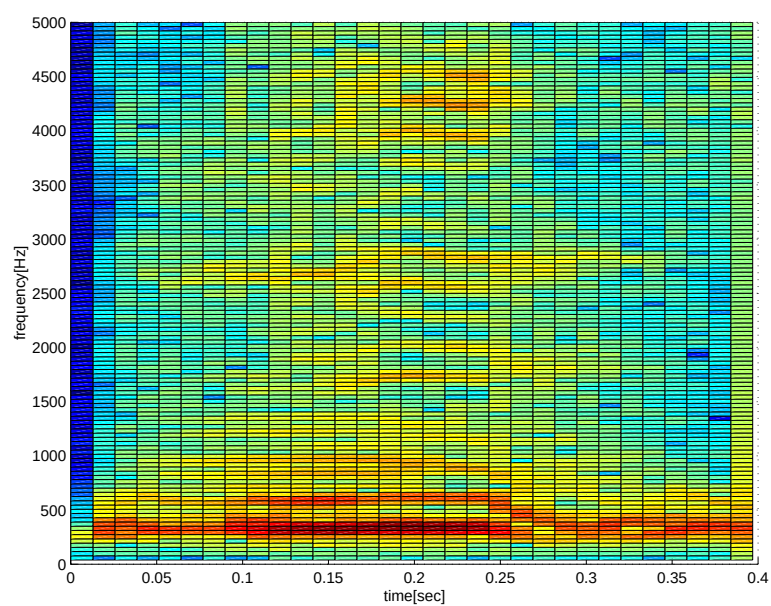
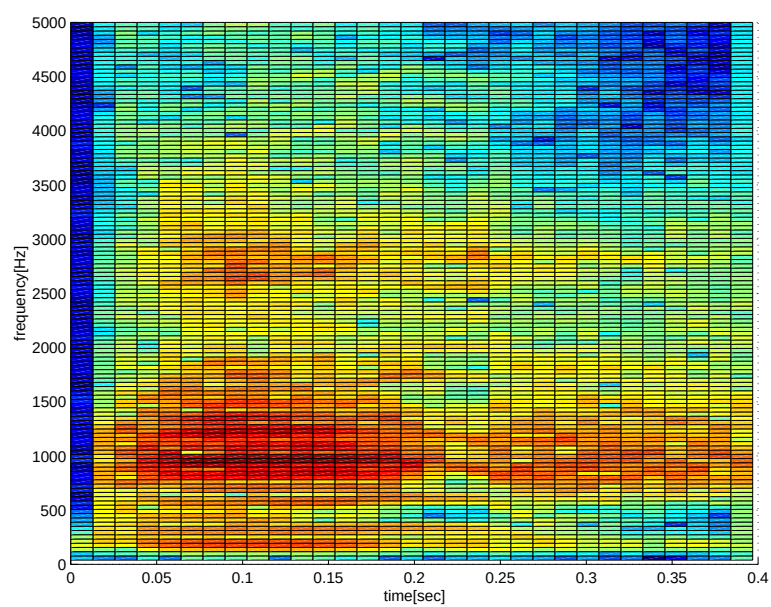


図 5.14: mms の/a/と fsu の/u/の分離したスペクトル (上: mms の/a/, 下: fsu の/u/)

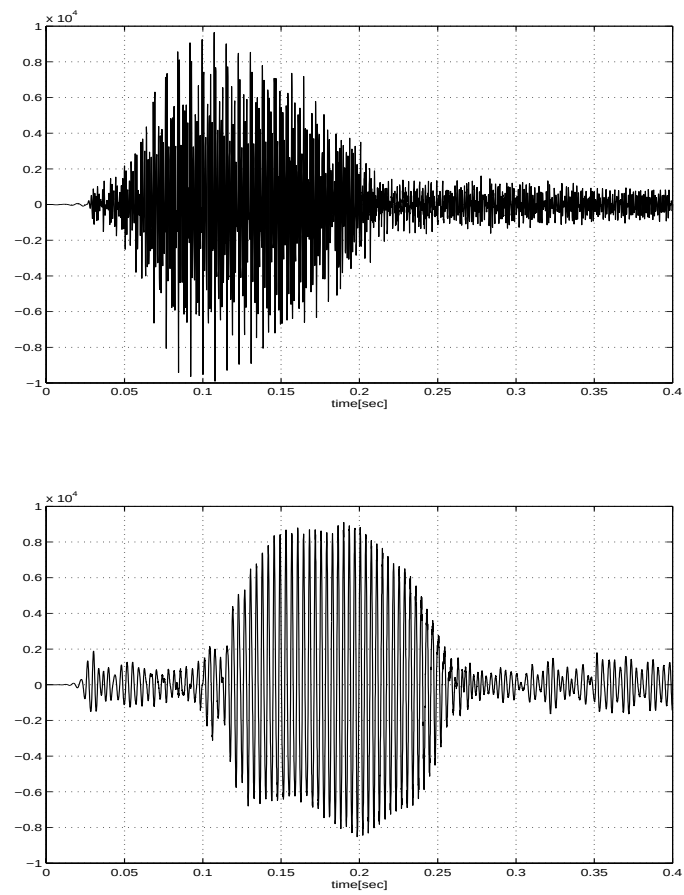


図 5.15: mms の/a/と fsu の/u/の分離結果（上：mms の/a/ 下：fsu の/u/）

第 6 章

おわりに

本研究では振幅変調マスキングの知見を基にして信号の搬送周波数とそれに対する変調周波数で信号の解析を行う変調—搬送周波数マップを定Qフィルタバンクを用いて構築した。

構築した変調—搬送周波数のマップ検出を用いて解析した結果、振幅変調マスキングの知見のような AM 音と帯域雑音を変調した雑音の場合搬送周波数に重なりがあっても変調周波数が異なっていれば分離できることが確認された。この解析によって純音とマスクの搬送周波数が同じ場合でも変調周波数が異なるほど分離が行える振幅変調マスキングの工学的な説明としても解釈できる。

また変調—搬送周波数マップの検出結果より音声等では高い変調周波数の位置にそれぞれの音声の基本周波数成分が表れることが確認された。これによって音声等の調波倍音の関係をこの求めた基本周波数で検出することが可能となる。

この変調—搬送周波数マップの検出結果を Bregman によって提案された聴覚の情景解析の” 調波関係に関する規則” と” 共通の立ち上がり／立ち下がりに関する規則” を用いることで二重母音の分離シミュレーションを行い分離できることが確認された。この際、定Qフィルタバンクの高周波数では帯域幅が広く出力がぼやけてしまうところは定帯域フィルタバンクを用いることで解決した。

最後に本研究では、音源や音声の重要な情報と考えられている時間変動する振幅包絡である変調周波数成分の解析に主眼を置いて行っており、搬送周波数成分の位相成分の分離は行っていないが、今後本モデルで提案した変調周波数成分における音源分離を音声情報処理へ応用することができれば、音声認識などの高精度化や、聴覚の情景解析等にも活用できると考えられる。

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助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，御討論をして頂いた新潟大学 岩城護助教授に心より感謝致します。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鶴木 祐史氏，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。

参考文献

- [1] 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)
- [2] Albert S. Bregman, “Auditory scene analysis: hearing in complex environments, in *Thinking sounds*”, pp. 10-36, Oxford University Press (1993)
- [3] T. Houtgast, “Frequency selectivity in amplitude-modulation detection”, *J. Acoust. Soc. Am.* 85(4), April 1989
- [4] Sd P. Bam and D. Wesley Grantham, “Modulation masking: Effects of modulation frequency, depth, and phase”, *J. Acoust. Soc. Am.* 85(6), June 1989
- [5] Brian C. J. Moore (大串 健吾訳), “聴覚心理学概論“, 誠心書房 (1994)
- [6] 河原 英紀, “聴覚の工学的表現”, 信学誌, vol. 76, pp. 1197-121 (1993)
- [7] 岩澤 秀紀, 柏野 牧夫, “2成分からなる振幅変調に対する順応“, (1998) 聴覚研資料, H-98-9 (1998)
- [8] 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)
- [9] 佐藤 雅昭, “ウェーブレット理論の数学的基礎 第I部“, 日本音響学会誌, vol 47, No. 6, pp. 405-415 (1991)
- [10] 佐藤 雅昭, “ウェーブレット理論の数学的基礎 第II部“, 日本音響学会誌, vol 47, No. 6, pp. 416-423 (1991)
- [11] 河原 英紀, “ウェーブレット解析の聴覚研究への応用“, 日本音響学会誌, vol 47, No. 6, pp. 424-429 (1991)

[12] Handel. S, “He a rin g” e d i te b y B. C. J. Moo r (Aca d e mi p r e ss)

[1 3]古井 貞熙, “デ ィ ジ タ ル 音 声 処 理 “, 東 海 大 学 出 版 会 (1995)