

平成8年度

修士論文

知覚的音源分離に関する実験的研究

電気通信大学大学院 情報システム学研究科
情報ネットワーク学専攻
ヒューマン・インターフェース学講座

9551013 川田 晋

指導教官

阪口 豊

出澤 正徳

伊藤 秀一

平成9年2月4日提出

目次

1 序論	3
1.1 本研究の目的	3
1.1.1 本論文の構成	4
1.2 本研究の背景	5
1.2.1 聴覚の情景分析	5
1.2.2 知覚的音源分離を行う上での困難さ	8
1.2.3 知覚的音源分離に用いられる物理的な手がかり	9
1.2.4 知覚的音源分離と知覚体制化	12
1.2.5 立ち上がりにおける音源分離と音の認識	13
2 実験	15
2.1 装置	15
2.2 実験1：立ち上がりのずれと協和度との関係	16
2.2.1 目的	16
2.2.2 方法, 条件	16
2.2.3 結果	18
2.2.4 考察	23
2.3 実験2：立ち上がりのずれと整数倍音との関係	24
2.3.1 目的	24
2.3.2 条件, 方法	24
2.3.3 結果	24
2.3.4 考察	28

2.4 実験3：立ち上がりのずれと音圧比との関係	29
2.4.1 目的	29
2.4.2 条件, 方法	29
2.4.3 結果	30
2.4.4 考察	37
3 一般的考察	39
3.1 音の群化・分離と周波数関係	39
3.2 音圧を考慮した群化と分離	40
3.3 分離と音色の関係	41
3.4 基音と群化	42
3.5 問題点	43
3.6 今後の展開	45
4 結論	46
謝辞	48
参考文献	49

第 1 章

序論

1.1 本研究の目的

我々は視覚、聴覚などの感覚器官を通して、外界から多くの情報を摂取することができる。物を見ることや音を聞くことがきわめて容易に自然に行われるために、我々は、それらを行う上での感覚器官の性質に気がつかないままに過ぎてしまうことが多い。実際には外界からの入力情報は、様々な過程を経て処理されているのである。視覚においては、外界からの視覚情報が網膜上に投影され、まずこの投影像のエッジや点などの形状分析が行われている。また、聴覚情報は鼓膜で機械的な振動に変換された後、蝸牛にある基底膜を通じて周波数分析が行われている。人間の知覚系では、このようにして、みずからの行動を遂行するのに必要な外界の情報を抽出し、それをもとに高次のレベルで外界の構造および状態を分析している。つまり、周囲に何が存在しどのような事象が生じているかを理解しているのである。

このように、人間がどのように情報を捉え処理しているかを明らかにし、その情報表現と整合する分析処理の体系を確立することは、人間の知覚機構の本質を明らかにする面で意義のあることであり、その重要性も高まりつつある。しかしながら、このような情景分析の問題は視覚の研究が先行し、聴覚では残念ながら発展途上の段階であった。

そのような状況の中、近年、Bregman[1]らによって「聴覚の情景分析」が提唱され、聴覚心理学の分野における研究が盛んになって来た。多くの体系的な研究蓄積がある視覚研究の方法論を聴覚の分野に持ち込むことで、今後一層の発展が遂げられることと期待が大きい。例えば、知覚体制化の一般的な原理は、視覚、聴覚に共通して見られるものであるが、ほとんどが最初、視覚において見いだされたものである。

そこで本研究は、聴覚心理の立場から、聴覚の情景分析の基本的な機能の一つである知覚的音源分離に着目し、音の立ち上がりが音源分離に与える影響に関して心理実験を行う。聴覚心理とは、音の物理的性質と聴こえの心理的性質について定量的な関連とそれを裏付けている知覚過程を求めることを指す。音響情報が与えられたときに生じる知覚が、聴覚器官から脳にいたるまでの間でどのような処理によって起こるものかは未知なままにしておき、その入出力関係の法則性を見出すことが聴覚心理における大きな役目である(図 1.1)。本研究では、この聴覚心理を通して、人間の音源分離メカニズムの解明へ向けて新たな知見を得ることを目的とする。

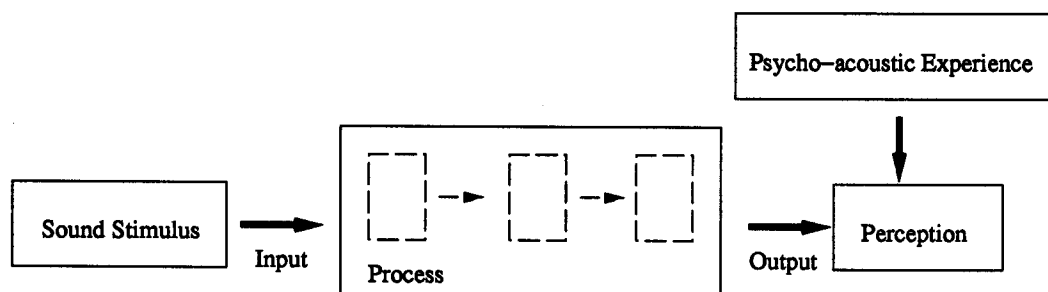


図 1.1: 聴覚心理学の構造

1.1.1 本論文の構成

本論文は、図 1.2のような構成になっている。まず第 1 章で、聴覚の情景分析の基本的な機能である知覚的な音源分離を紹介し、知覚的な音源分離を行う上で手がかりにしている物理的な手がかりの中から、立ち上がりの重要性について述べる。第 2 章では、立ち上がり音が音源分離に与える影響について、3つの実験を通して考察を行う。第 3 章では、実験から得られた結果を踏まえ、一般的考察を行い、第 4 章を結論とする。

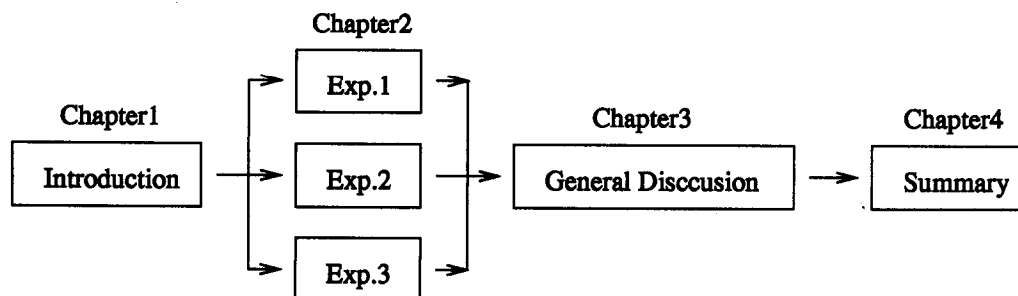


図 1.2: 本論文の構成

1.2 本研究の背景

近年、聴覚心理学の分野において聴覚の情景分析の研究が盛んになり、この研究から知覚的な音源分離や音の錯覚などの新たな知見が報告され、心理学の分野だけでなく、生理学や工学の分野でもさかんに取り組まれるようになった。本節では、聴覚の情景分析とはいかなるものかを説明し、関連する研究報告例の紹介を通し、本研究の目的である知覚的な音源分離における音の立ち上がりの重要性について述べる。

1.2.1 聴覚の情景分析

聴覚の情景分析とは、人間が聴覚を通して周囲に生じている事象を分析する機能を指し、人間が持ち合わせている優れた情報処理機能の一つである。この機能により、我々は複数の音源からの音が混ざり合い、一つの波形として耳に入力される音に対して、音源からの一つひとつの音の性質に基づいて適応的に分離、抽出し、同じ性質のものを群化し、状況に応じて補完して、入力した情報を処理していると考えられる。例えば、我々は、階層化された音全体を雑踏と言う一つの音としてとらえたり、あるいは、その中に含まれる複数の音源、ここでは車の音や人々の話し声と言ったものをそれぞれ一つの音としてとらえることができるが、これは情景分析の結果にはかならない。このように、聴覚の情景分析とは、人間が周囲の環境に応じて合理的に行動するために、必要不可欠な機能なのである。

ところで、この聴覚の情景分析の機能をさらに細かく見ていくと、情景分析を可能にしている基本的な機能の中に、知覚的な音源分離機能と音素修復機能とがある。知覚的な音源分離とは、複数の音響事象から目的とする音を知覚的に分離する機能を指し、また、音素修復とは、ある事象に関する情報が他の事象の干渉によって欠落している場合にそれを補完する機能を指す。これらは、環境内の音響事象の物理的な性質を反映した合理的な解釈が得られるように機能している。ここでは、具体的な説明は後述するが、双方の代表的な研究報告例を紹介する。

- 知覚的な音源分離機能

周波数の高い音 H と低い音 L からなる音列が、比較的遅いテンポで HLHLHL と提示された場合、人間は全体を一つのまとまりとして知覚するが、テンポを速めた場合には音の物理的な時間順序に対応した群化が起こるのではなく、高さの特性に従った群化が生じる。その結果として提示された音列は、高い音のまとまり HHH と低い音のまとまり LLL の二つの音のまとまりに分離して知覚される (図 1.3)。このように知覚的に分離された音のまとまりを「知覚的な音」と呼んでいる。この現象は、音列の提示速度が速くなるほど、また、周波数 L と H の差が離れているほど起こりやすくなる。これは、同じ楽譜でもテンポが異なると形成される知覚的な音の数が増えることを示しており [2]、多くの作曲家達は、このような現象の効果を演奏にうまく利用しているという [3]。

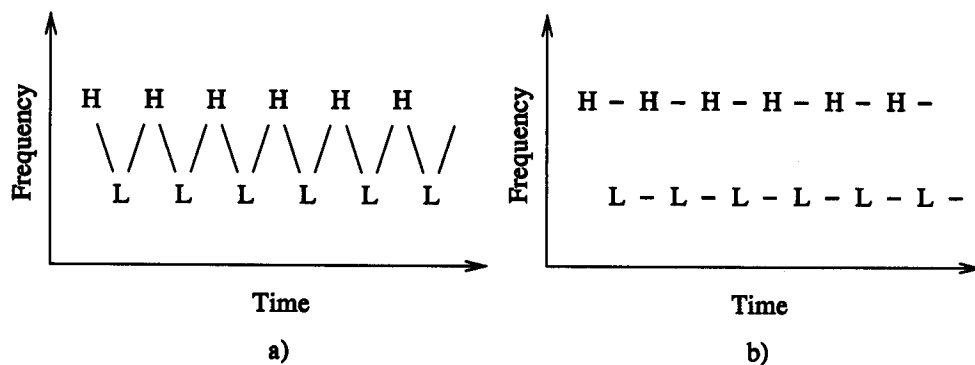


図 1.3: 知覚される音のまとまり

- a) テンポが遅い場合、全体を一つのまとまりに知覚
- b) テンポが早い場合、別々の二つのまとまりに知覚

同様に、周波数領域でオクターブ離れた 2 音の交替音列 HLHLHL を、同時に左右の耳に交互に入れ替わりながら、例えば、右耳には HLHLHL、左耳には LHLHLH と提示すると、右は H-H-H-、左は-L-L-L- (‘-’は無音を意味する) というような知覚像を得ることが多い (図 1.4)。これは、一つの音が耳から耳へと交互に動き、かつその移動と同期して音がオクターブ隔てて交代しているような錯覚が生じていることを指している。このようなオクターブ錯覚という錯覚現象も報告されている [3]。

Right Ear	H L H L H L H L H L
Left Ear	L H L H L H L H L H

a) Presented Stimulus Sequence

Right Ear	H - H - H - H - H -
Left Ear	- L - L - L - L - L

b) Perception

図 1.4: オクターブ錯覚

- 音素修復機能

これは、知覚的な音源分離と並んで情景分析を行う上で必要不可欠な機能である。例えば、ある人と会話をしているときに、別な人のドアの開閉によって、相手の音声の一部が雑音によってマスクされ、その部分の音響情報が音声の知覚に使えないことがある。このような中でも、我々は相手の音声を正しく聞き取ることができる。この機能は音素修復と呼ばれ、欠落した音響情報を知覚的に補完する機能を指している [2]。

音素修復には文脈による非音響的な情報もかかわっているが、単純な音を用いても補完されることがある。この場合は、音素修復というよりも聴覚的補完 (連続聴錯覚) といい、一方の音によって覆い隠された他方の音を知覚的に修復するという聴覚的誘導の過程に生じる現象をさす。二つの純音の間に雑音が存在する場合は、雑音が 300msec 程度であっても純音は雑音の下で連続して知覚される。この場合、純音と雑音との間に無音部分があってはならず、また、雑音の両側に補完される純音が存在していなければならないなどの複数の条件が満たされる必要がある。

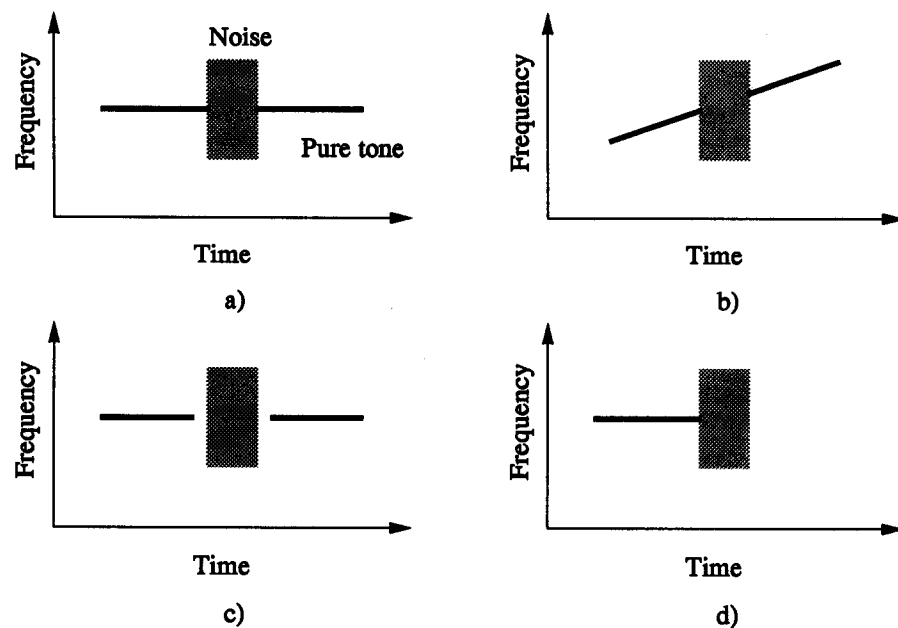


図 1.5: 連続聴錯覚

a), b) 聴覚的補完が可能な条件

c), d) 聴覚的補完が不可な条件

1.2.2 知覚的音源分離を行う上での困難さ

前節で述べたように、聴覚の情景分析に関する研究により、人間は同時に存在する複数音源の中から、耳に到来する混合音の各周波数成分を容易に各音源に振り分けることができる。その結果として、一連の音のまとまり (知覚的な音) を形成し、聞き分けていることがわかってきた。

しかしながら、このような音源分離問題は、左右の耳に与えられた混合音の情報から周囲で生じている事象を復元するいわば逆問題であって、一般には一意的な解が存在しない非常に困難な問題である。あらかじめ幾つの音源に割り振れば、入力された音響情報を適切に分析することができるか未知なため、成分の分け方、つまり、知覚的な音源は無数に存在する。このような音源分離問題において、解を一意的に定めるためには、何らかの制約条件が必要になる。そこで知覚上での分離を行う際、我々は個々の音源の様々な物理的な手がかりを用いていると考えられる。なぜなら、一つの音響事象から生じた成分であれば、それらの変化の仕方は互いに何らかの類似性があり、成分間で相関をもって変化して

いると考えられ、また、別々の音源から発生した場合、それぞれの成分間にはそのような制約はないと考えられるからである。したがって、これらの物理的な手がかりを音源分離の判断に用いれば、合理的な音響事象の解釈ができる可能性が高い。つまり、制約条件として音響事象の物理的な手がかりを用いることにより、複数の音響事象が混ざりあった混合音を分析し、その結果、知覚的に音源を分離するという逆問題を解くことができる。

1.2.3 知覚的音源分離に用いられる物理的な手がかり

ここでは、実際に知覚的な音源分離をする際に用いられている代表的な物理的な手がかりを紹介する。

まず、ステレオ音を用いた場合、左右の耳に到達する音の時間差 (両耳間時間差) や強度差 (両耳間強度差) は、音源位置の情報を与えるため音源を分離する上で大きな手がかりになる。音源を定位する上で両耳聴取の機能は重要であり、特定の方向から到来する音に注意を向けることで、他の音を効果的に排除し、目的とする音を分離することができる。これは、環境内で複数の音源が競合するときにおいてその能力を発揮する。また、この他に音源位置が音の分離に貢献していることを示す例には、両耳間でのマスキングレベル差がある [4]。この現象は、信号とマスキャーを両耳に同位相、同レベルで入力した場合に比べて、信号の両耳間での位相差、レベル差をつけると信号を検知しやすいという現象である。このことは、背景音の中での目的とする信号音の分離に際して、片耳聴取に比べて両耳聴取の方が、明らかに分離に有利であることを示している。

また、人間はモノラルの音を用いた場合でも、ある程度の分離を行うことができる。二つの音の間に周波数成分の高調波関係のずれや立ち上がりのずれが生じると、分離して知覚されることが報告され、これらの要因は、人間の音源分離知覚に大きな影響を与えていると言われている [1, 5]。多くの楽器音は、近似的に基本周波数とその整数倍の周波数からなっており、また、それらの周波数成分の立ち上がり時刻のずれは数十 msec 程度であって、ほぼ同時ながらずれが生じて立ち上がっていることが知られている [6]。これらの性質は、楽器音だけでなく世の中の自然な音において広くあてはまるものである。

このほか、2音が同時に立ち上がる場合でも、音の立ち上がり時間が異なったり、一方の音の周波数を変調したりすると分離が起こりやすいことも報告されている [8, 9]。さら

に、2音の基本周波数が音源分離を行う上でも重要であることも知られている [7, 10](図 1.6参照).

上記のように、異なる音源から生じる成分を知覚上で分離する際に用いられる物理的な手がかりは複数存在するが、これらのうちのある一つの手がかりがどのような場面でもいつも使われているのではなく、これらの手がかりを統合して入力された音響情報を分離しているのである.

本論文の目的とは異なるが、これらの手がかりが音源分離にとって重要であることを示す例は、工学的な研究の立場からも知ることができる. 雑音の中から目的とする音声を抽出したり、あるいは、ある音声と別の音声を分離すると言ういわゆる典型的な信号分離問題を解決する際、従来は音声信号処理的な手法が用いられてきた. しかしながら、情景分析の研究より物理的な手がかりによる音源分離の心理学的な知見をとりいれて、楽器音を対象とした周波数成分のクラスタリングにおける評価統合モデルを提案した例 [15] や、音声を対象としたマルチエージェントシステムも提案されている [16]. その他にも、これらの手がかりを用いて多くの音源分離モデルも報告されるようになった [17, 18, 19, 20]. このような事実は、いずれも人間の心理学的な知見を取り入れたものであり、音源分離モデルを構築する際に、これらの手がかりが有益な情報になり得ることを示している.

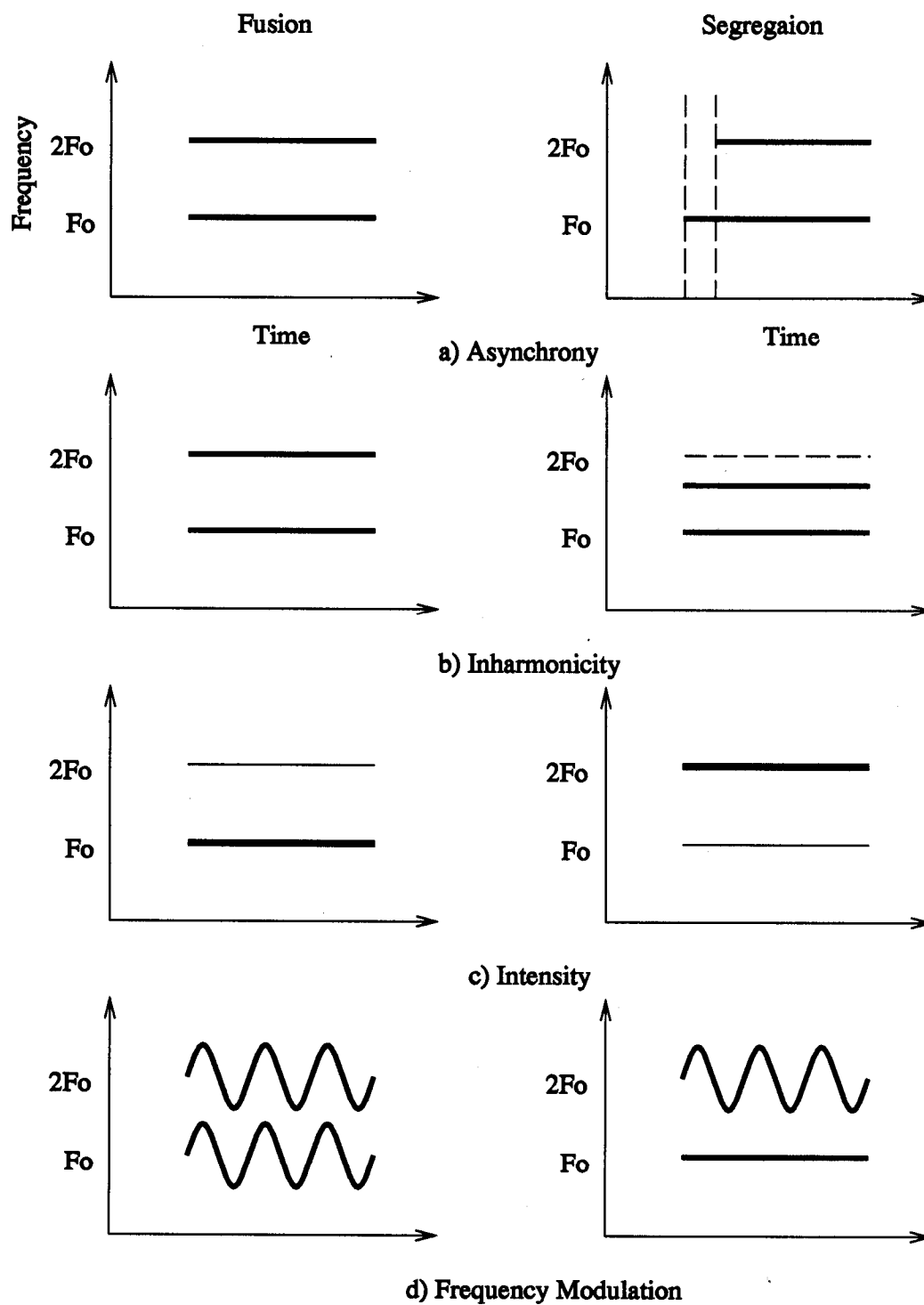


図 1.6: 物理的な手がかりの例

1.2.4 知覚的音源分離と知覚体制化

我々の耳に到来する音が一つの音源だけから発せられたものではなく、通常、複数の異なる音源から発せられたものであることは既に述べた通りである。この際、複数の音源からの混合波形を示す鼓膜の振動パターンが、耳小骨の機械的な振動になり、それが内耳の蝸牛の中にある基底膜に伝えられ、ここで入力された音の周波数分析が行われている。分解された各成分は、より高次の過程において、様々な特性を手がかりにそれらが発せられた音源ごとに群化され、入力された音響情報を幾つかの知覚的な音に分離されると考えられる。

ところで、このような知覚的な音源分離をする際の音の群化過程には、二つの側面がある。一つは、ある瞬間に同一音源から同時に発せられた周波数成分がすべて同じグループに帰属するように群化するという周波数方向の群化であり、もう一つは、同一音源から次々に発せられた時間的に変化する周波数をつなぐように群化する時間方向の群化である。これらの群化の結果として、知覚的な音が構成されているのである。

そこで心理学的な群化、つまり知覚体制化という点から知覚的な音源分離との関連について簡単に触れておく。なぜなら、ゲシュタルト心理学者が述べている知覚体制化を支配する一般的な原理は、知覚的な音源分離をする際の物理的な手がかりが用いられる様子とうまく合致しているからである。

彼らは、これらの原理を用いることにより、同じ音源で発せられる部分を群化し、そうでない部分から分離できるように述べている。例えば、類同の原理とは、高さ、音色、大きさ、または定位が類似している音が一つのまとまりへと群化されるというものである。良い連続の原理とは、ある音の周波数、強さ、定位、またはスペクトルが滑らかに変化する場合には、これは一つの音源における変化と知覚され、急激に変化する場合には、音源が変わったことを示すというものである。共通運命の原理とは、二つの成分が同時に同様な変化をするとき、これらが一つの音源から発せられた成分だと知覚されるというものである。この原理は、音声や音楽の知覚分析において重要な役割を果たしていると考えられる。閉合の原理とは、ある音の一部がマスクされたりふさがれたとき、この音が続いているように知覚されるというものである。これらの例で、彼らが述べた群化を支配する原理は、音響情報をいくつかのまとまりのある形態としてとらえ群化していると考えることが

できる。

上記の原理は、そのいずれかだけですべて説明されるものではなく、通常、これらの複数の原理が共同して機能していると考えられる。それゆえ、これらの原理が複雑に組み合わせられ、それにしただった群化は、我々が環境を最も効率的に解釈することを可能にするのであろう。

1.2.5 立ち上がりにおける音源分離と音の認識

前節までに、我々は音源分離をする際には様々な物理的な手がかりを用いていることを述べたが、本研究ではその中から立ち上がり時刻のずれに着目した。その理由は、立ち上がりのずれが最も音源分離に影響を与える要因の一つであり、音源分離以外においても、音の立ち上がりが音の認識において重要な役割を果たしているからである。

立ち上がりが楽器音への識別に与える影響を調べた研究は多く、例えば、人間は楽器音を途中ですりかえてもそのことに気づかないこと、また、楽器音の立ち下がり部を削った場合は楽器音の識別率がほとんど低下しないのに対して、立ち上がり部を削るとその識別率が著しく低下する現象が報告されている [3, 11]。また、音の立ち上がりは、楽器音の音色にも影響を与え、楽器音の立ち上がり時の過渡的な時間変化が、その音色を特徴づける上で重要であることは、古くから知られている [11]。例えば、ピアノの音の強さは、立ち上がりが急に大きくなり、その後はゆるやかに減少しているが、オルガンの音は、立ち上がりがゆるやかであって、鍵盤を離すまで一定の強さが持続する。また、音の強さが時間的に変化するばかりでなく、成分音の立ち上がり方も一様ではなく、それがまたその楽器音への音色を特徴づけてもいる [12]。これらは、楽器音の種類にもよるが、音の立ち上がりが楽器音の音色を決定し、それゆえ、その楽器音の識別をする上でも重要な役割を演じていることを示している。

さらに、音声の場合でも、立ち上がりのずれが母音の音色にも影響を与え [21]、また、立ち上がりのスペクトル構造の過渡的な変化が、音声の音色を聞き分け、群化し、認識する過程できわめて重要である場合が多い [13]。

上記のように、音の立ち上がりが音の認識と密接な関係にあること、また、音の認識と音源分離は相互に関係していることを考えると、音の立ち上がりは音源分離においても重

要な意味をもっている可能性が強いと言える (図 1.7)。しかし、従来の研究では、音の立ち上がりの影響が単独で議論されており、音の周波数構造を含めて考慮した研究はみられない。

そこで本研究では、まず第一に二つの音源の周波数関係に着目して、第二にそれらの相対的な音圧を操作し、音の立ち上がりのずれが音源分離にどのような影響をもたらすかを実験的に検討を行った。

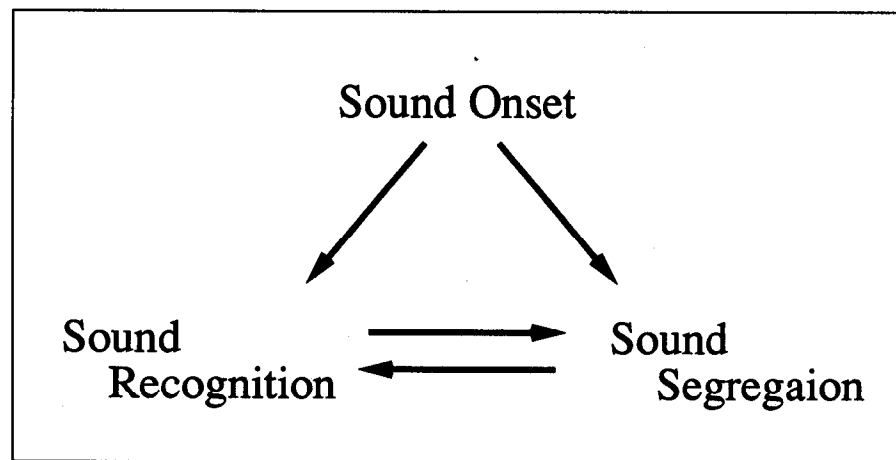


図 1.7: 音の立ち上がりの重要性

第 2 章

実験

2.1 装置

本実験では、図 2.1 のような装置を用いた。図に示したように、この装置は、Sound Blaster を内蔵した IBM-PC 互換機を用いて、刺激音を作成した。実験中、音声出力は、オーディオアンプで増幅し、個々の被験者へは、防音設備のない実験室において周囲を暗幕で取り囲み、ヘッドフォンを通して提示した。個々の実験装置の仕様を表 2.1 に示す。

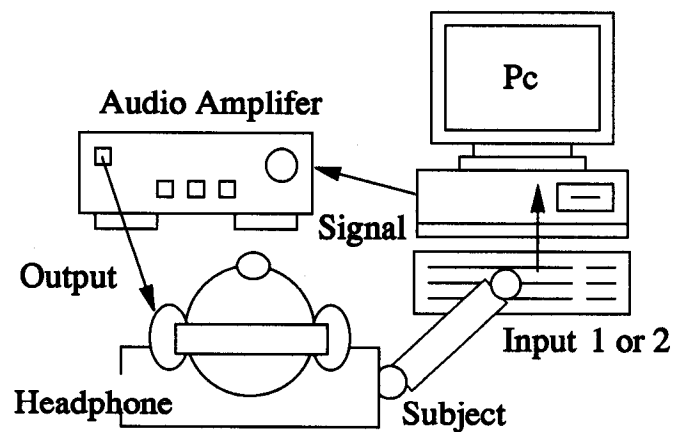


図 2.1: 実験装置の構成

表 2.1: 実験装置の仕様

装置	名称
Computer	Dell Optiplex XL 575
Audio Amplifier	SONY TA-F510R
Headphone	SONY MDR-Z600

2.2 実験1：立ち上がりのずれと協和度との関係

2.2.1 目的

従来の研究より、二つの音の周波数比が簡単な整数比で、かつ、立ち上がり時刻のずれがない時、それらは融合して一つの音として知覚され、また、協和度が高いほど良く融けあった音として知覚されることが知られている [12, 14]. しかし、このように協和度の高い2音を用いても、それらの立ち上がり時刻のずれが大きくなると、二つの音として知覚されるようになると考えられる。

そこで本実験では、二つの音の立ち上がり時刻のずれを変化させたとき、それらの音が分離して聞こえるずれの閾値を測定し、それが二つの音の協和度によってどのように変化するかを調べた。

2.2.2 方法、条件

● 刺激

それぞれの試行で提示される刺激には2つの純音を用い、低い音の周波数を固定して、2音の周波数比を1:2～5:8までの7種類(表2.2)、低い音を400 Hz, 1 kHzの2種類設定した。

表 2.2: 提示刺激音 (Hz)

音程名	周波数比	1kHz の組	400Hz の組
完全8度	1:2	1k+2.00k	400+800
完全5度	2:3	1k+1.50k	400+600
完全4度	3:4	1k+1.33k	400+533
長6度	3:5	1k+1.67k	400+667
長3度	4:5	1k+1.25k	400+500
短3度	5:6	1k+1.20k	400+480
短6度	5:8	1k+1.60k	400+640

刺激音の継続時間を500 msec、二つの周波数成分の振幅は同じとし、音の立ち上がり・立ち下がり、それぞれ1 dB/msecの傾きで立ち上げ、-1 dB/msecで立ち下げた。実際には、音が完全に立ち上がるまで約100 msecかかる(図2.2)。各立ち上が

り時刻のずれは、もう一方の音に対して、0 msec から 100 msec まで全部で (-100, -70, -50, -40, -30, -20, -10, 0, 10, 20, 30, 40, 50, 70, 100) の 15 段階で変化させた。また、音の立ち下がり、は、常に同時にした。なお、刺激音のサンプリング周波数は、44.1 kHz とした。

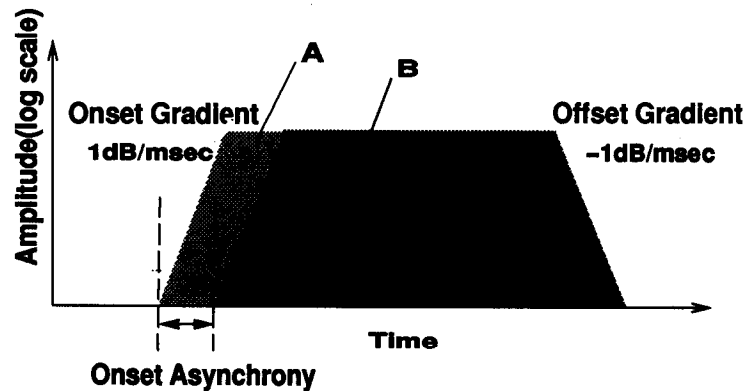


図 2.2: 実験の概要

- 手順

被験者へは、刺激音が提示された時に、2 音を知覚することができたら、分離とし、そうでないならば融合として判断し、分離と判断した場合は、キーボードで 2 を、融合と判断した場合は 1 を押すように教示した。

提示方法は、一つの周波数比の組み合わせに対して、15 段階設定した立上り時刻のずれをランダムにそれぞれ 3 回ずつ被験者へ提示し、それを 1 セットとし、それを 5 セット繰り返した。つまり、被験者は一つの刺激音に対して 15 回聴取し、それを 2 種類の音に対してそれぞれ 7 種類の周波数比の組み合わせについて聴取した。試行回数は全てで 70 セットになり 5 セット毎に休憩を挟んで実験を行った。この刺激系列を図 2.3 に示す。

また、刺激として用いた 400 Hz, 1 kHz は、10 msec の間に、整数個の周期を持つので、各ずれにおける 2 音の位相はすべて同位相になっている。

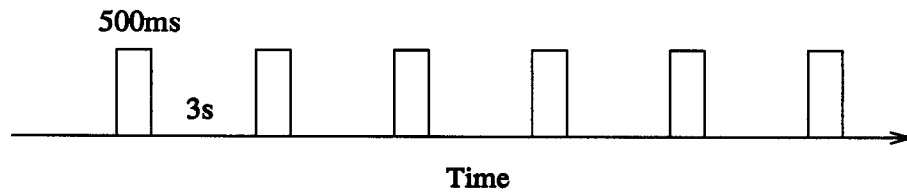


図 2.3: 刺激音の系列

2.2.3 結果

図 2.4に 1 kHz と 2 kHz の純音を組み合わせたときに得られた psychometric function の一例を示す。横軸は 2 音の立ち上がり時刻のずれの絶対値を表し、縦軸は 2 音に分離して知覚された割合を指している。図中の実線は 1 kHz が 2 kHz よりも先行して立ち上がった場合、点線はその逆を示している。この図より、2 kHz の成分が先に立ち上がる方が、閾値が小さくなることが読みとれる。本実験では、このグラフが一人につき 14 枚得られることになる。また、被験者 3 人分の実験結果まとめたものを図 2.5に示す。図中の Higher component prior, Lower component prior とは、先行して立ち上がる音の高低を指し、2 音に分離して知覚される立ち上がり時刻の閾値の絶対値をグラフで示している。また、図中で×印の部分は、同時に立ち上げても融合して知覚されなかったことを示している。図に示す通り、低音成分、高音成分それぞれ 2 つの場合の閾値の差と標準偏差の関係を考えると、被験者間で差はあるものの周波数比が 1 対 2 の時、高い成分が先行すると分離しやすいことが明らかになった。つまり、高い成分が先行する方が低い成分が先行する場合に比べ、閾値の差が大きくなった。また、その他の周波数比では、2 音の立ち上がり時刻の差がほとんど同じであった。400~800 Hz の周波数範囲で行った実験によると、400 Hz の音が先行すると分離する傾向が見られた。

なお、図中の*は、閾値が 100 msec を超えていることを示している。実際の実験では、立ち上がり時刻のずれは 0~100 msec の範囲であるが、あくまでも psychometric function から得られた値であり、測定値ではない。

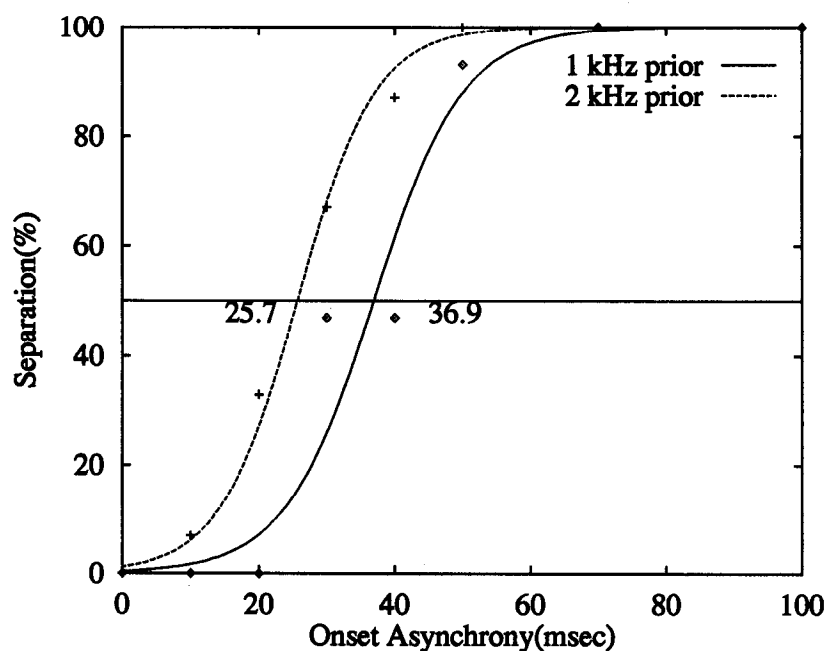
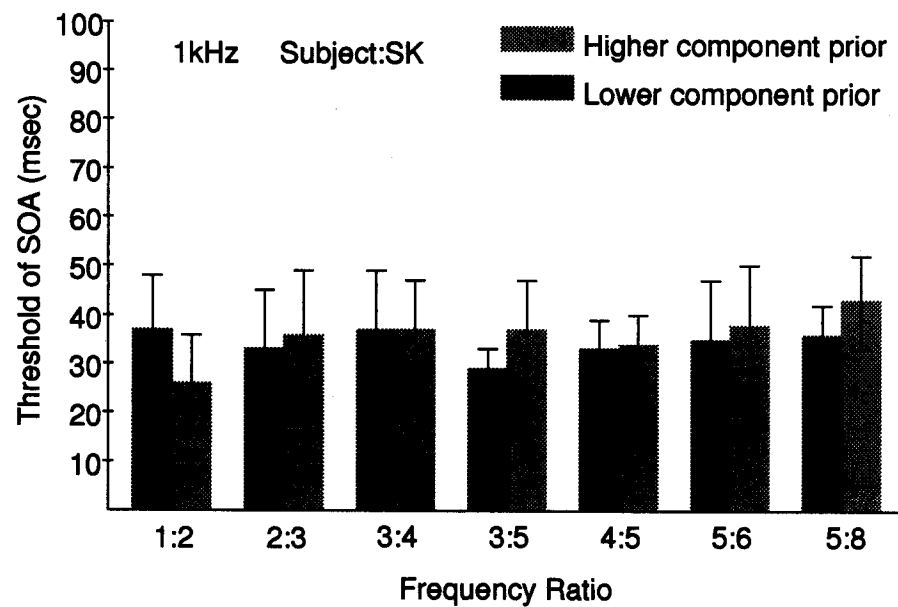
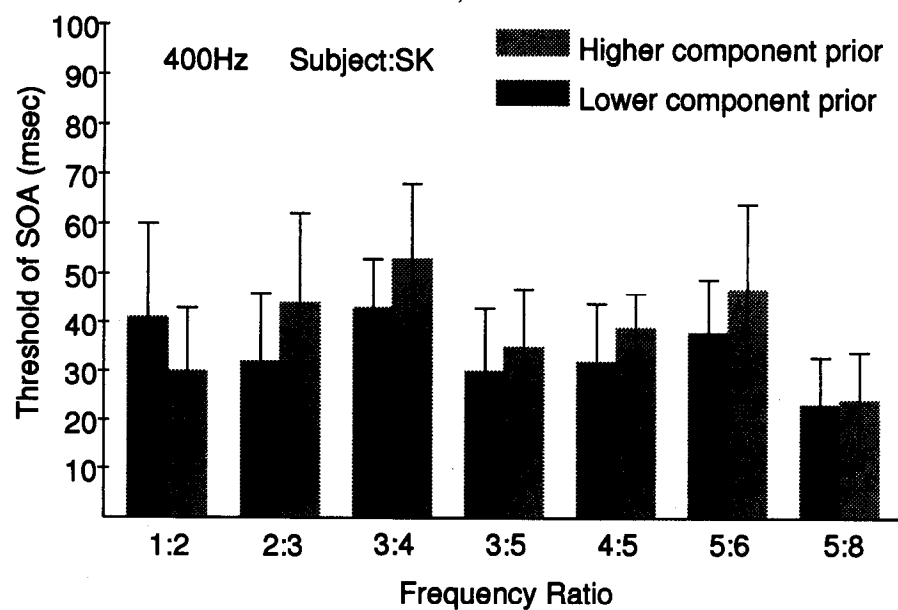


図 2.4: 実験結果 (psychometric function)

ここで psychometric function とは、被験者に多数回刺激を提示して判断を求め、各刺激に対して2音が分離して知覚される確率 (P) と、 P を刺激値 (刺激の物理量) S の関数としたとき、この S - P 関係を指す。経験的に得られた P を、横軸を刺激値としてプロットし曲線で結ぶと、どの感覚様相においてもこの psychometric function は S 字型の増加関数となる。このとき弁別閾は、この関数における $P=0.50$ にあたる S の値、つまり刺激に対する判断が半々になる刺激値となり、標準偏差は、 $P=0.84$ に対する S の値と弁別閾との差として表される [22]。

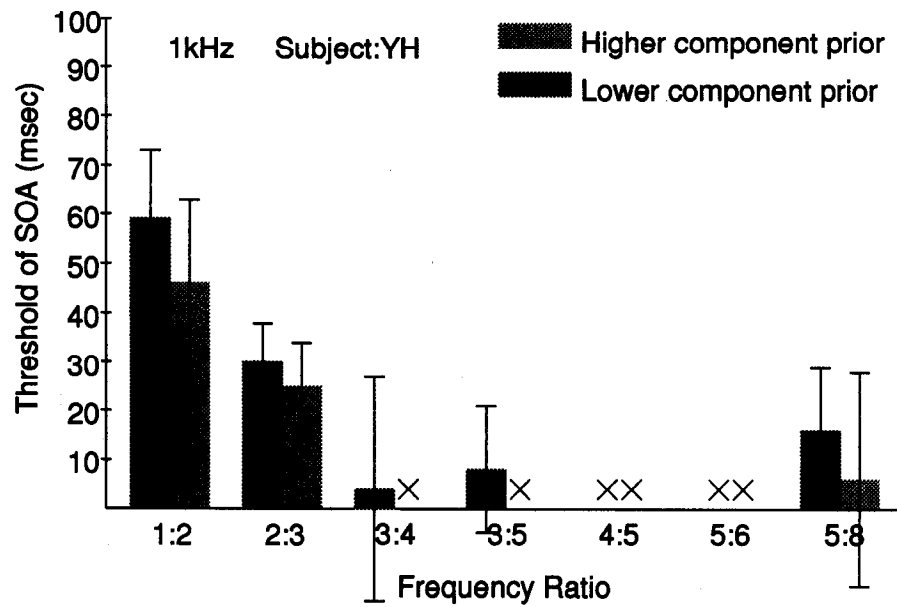


a)

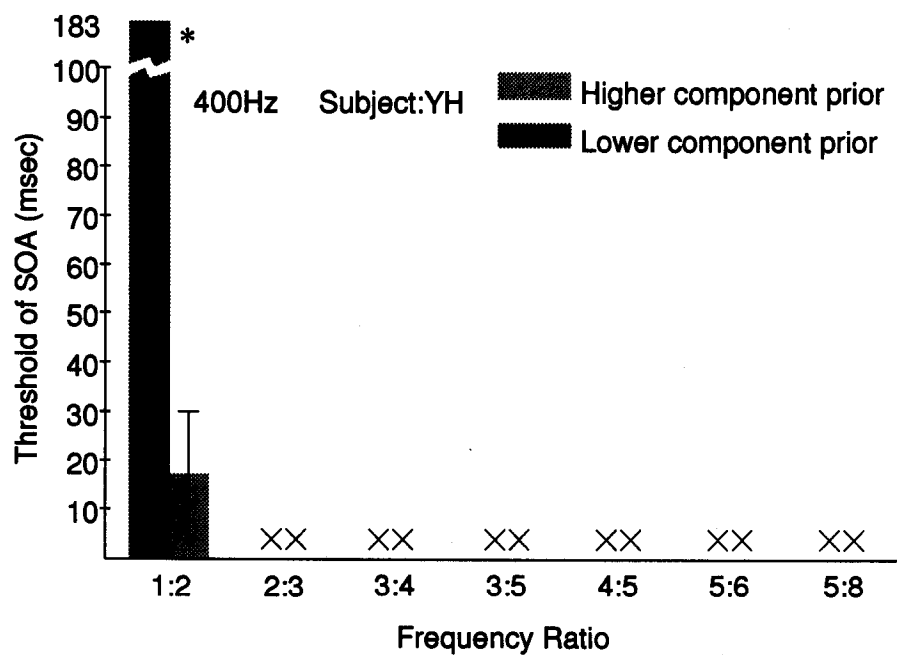


b)

図 2.5: 実験結果

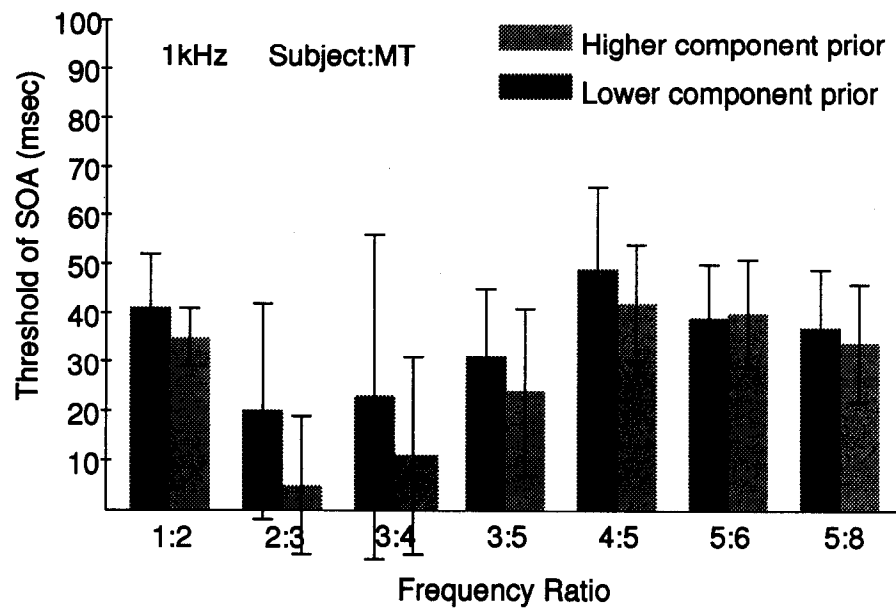


c)

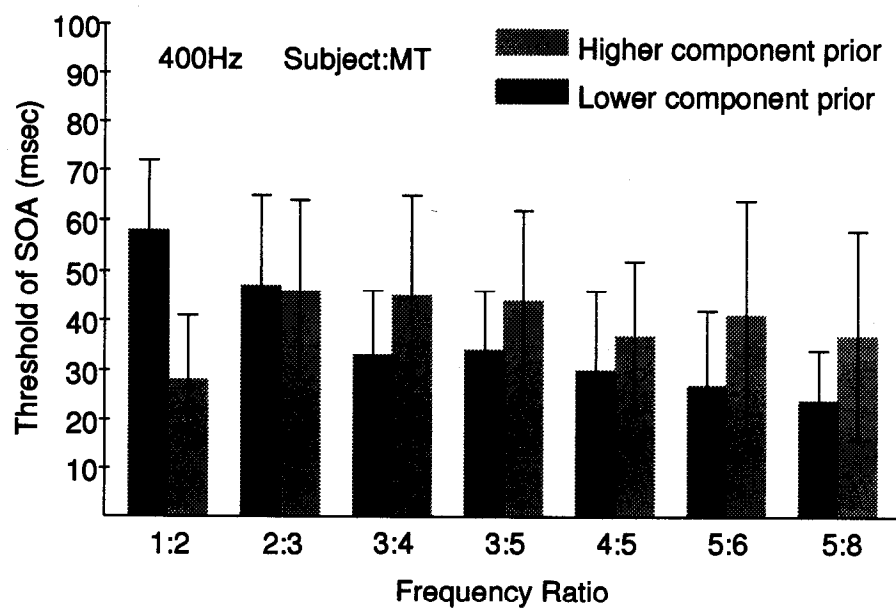


d)

図 2.5: 実験結果



e)



f)

図 2.5: 実験結果

2.2.4 考察

上の結果から、二つの成分の協和度の違いにより2音が分離して知覚される立ち上がりのずれの閾値に違いがあることがわかった。特に周波数比が1:2の閾値のときは他の協和度とは異なり、高い音が先行すると分離して知覚されやすい傾向が見出された。すなわち、低い音が先行する場合と高い音が先行する場合で、ずれの閾値が非対称になり、音の順序により音源分離の仕方に違いがあることを示している。

この結果の一つの解釈として、低い音を基本音としたとき、基音と高い音(第2倍音)との間で群化が働いたことが考えられる。それゆえ、第2倍音が先行して立ち上がると、その群化から解き放たれ分離しやすくなると考えられる(図2.6参照)。この点については、後でさらに考察する。

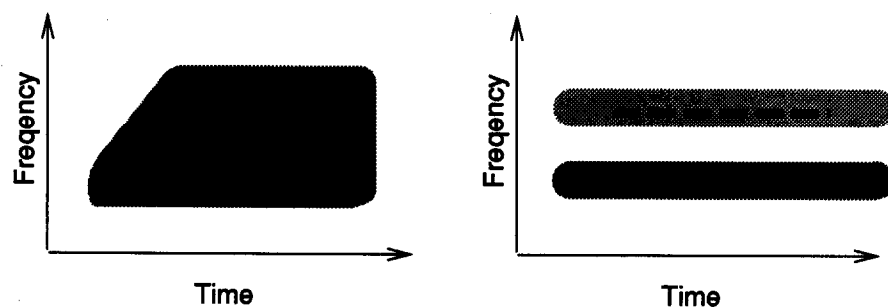


図 2.6: 知覚される音のまとめ

2.3 実験2：立ち上がりのずれと整数倍音との関係

2.3.1 目的

実験1より、周波数比が1：2のとき倍音が先に立ち上がると、2音に分離して知覚される傾向が明らかになった。このようなオクターブ関係にある2音は、立ち上がりが同時のとき、最も良く融合して一つの音として知覚されやすい組み合わせであった。

そこで本実験では、2音の周波数比を整数倍音にし、基音と高次倍音との組み合わせに対して同様な実験を行い、立ち上がりが音源分離に与える影響を考察する。

2.3.2 条件、方法

400 Hz と 1 kHz を基本周波数として、それぞれの整数倍音を2, 3, 4, 5倍音まで組み合わせた4種類、合計で8種類(表2.3)を設定し実験を行った。このとき、試行回数は40セットになるが、その他の実験条件と方法は、実験1と同じである。

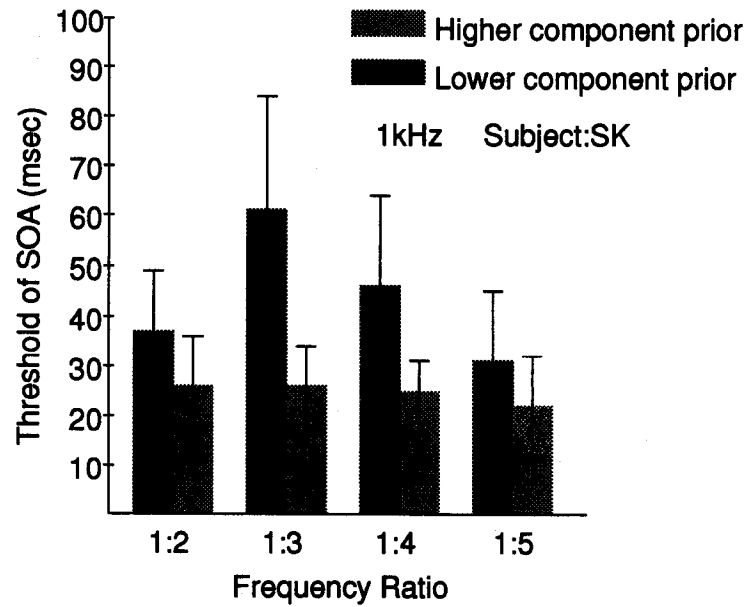
表 2.3: 提示刺激音 (Hz)

周波数比	1kHz の組	400Hz の組
1:2	1k+2k	400+800
1:3	1k+3k	400+1200
1:4	1k+4k	400+1600
1:5	1k+5k	400+2000

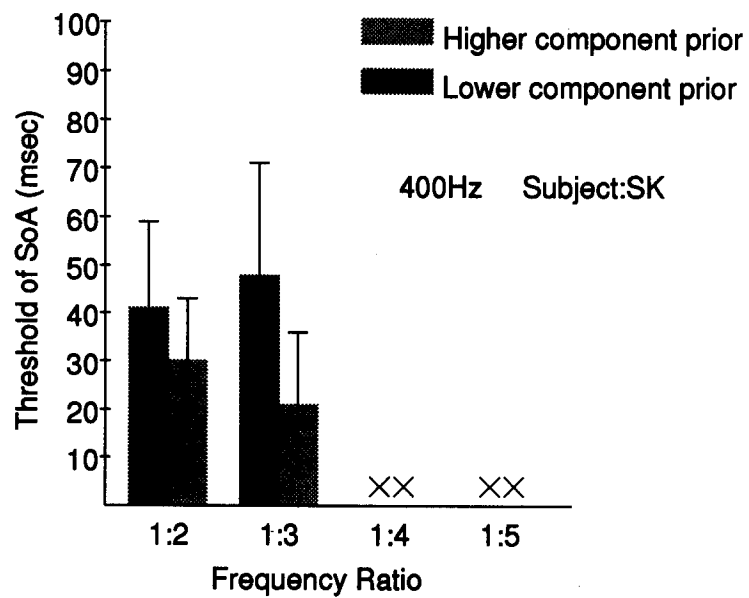
2.3.3 結果

本実験から得られた結果を図2.7に示す。横軸は倍音関係にある2音の周波数比を指し、その他の構成は図2.5と同じである。本結果より、整数倍音が基音より先行して立ち上がると、2音に分離しやすい傾向が得られた。基音が1 kHz の場合、周波数比が1：3のとき、閾値の差が最大となって得られ、400 Hz の場合では、周波数比が1：2のとき、閾値の差が開いた。

また、双方の基音の組み合わせにおいて周波数比が1：4以上に広がると、同時に立ち上げて分離して二つの音として知覚される傾向が得られた。

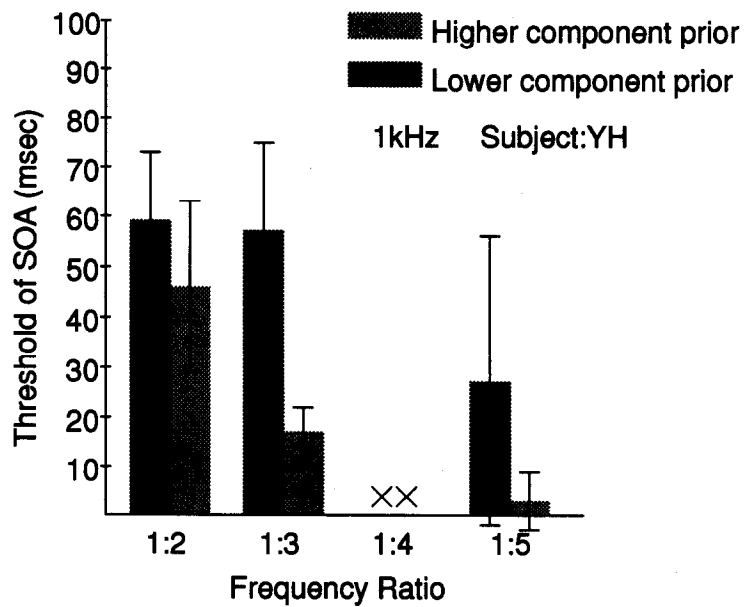


a)

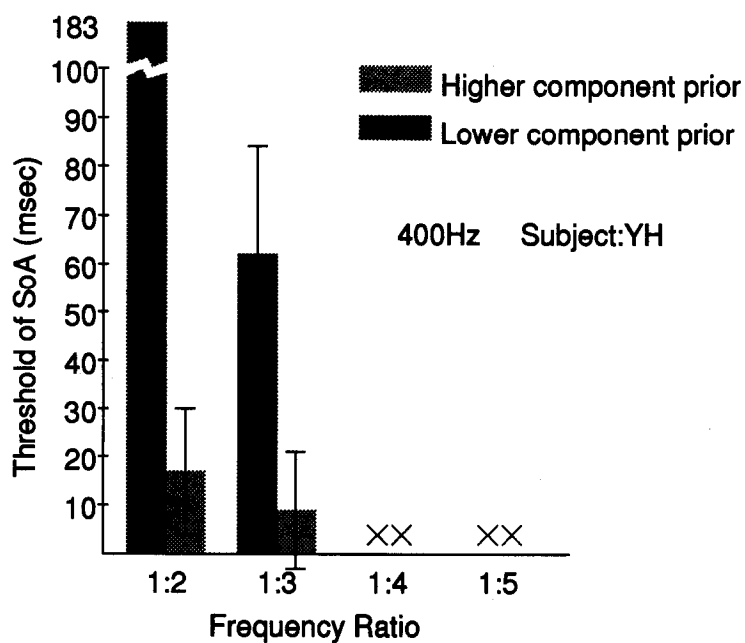


b)

図 2.7: 実験結果

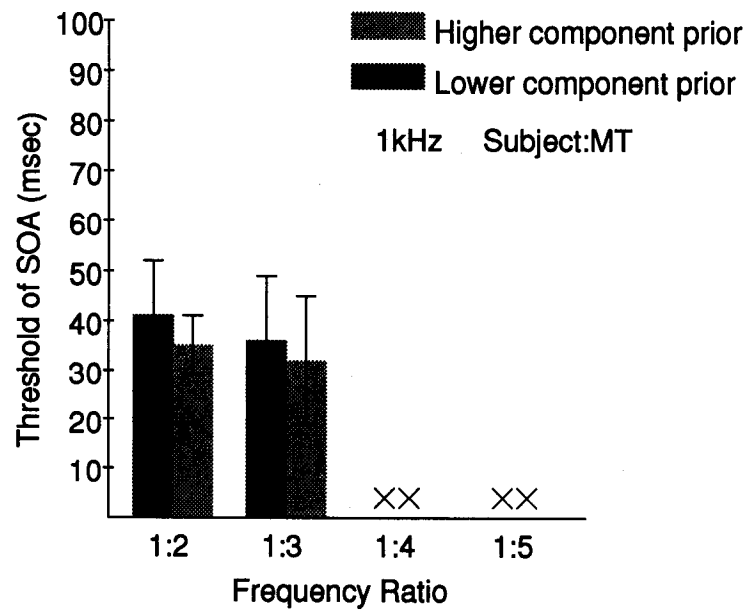


c)

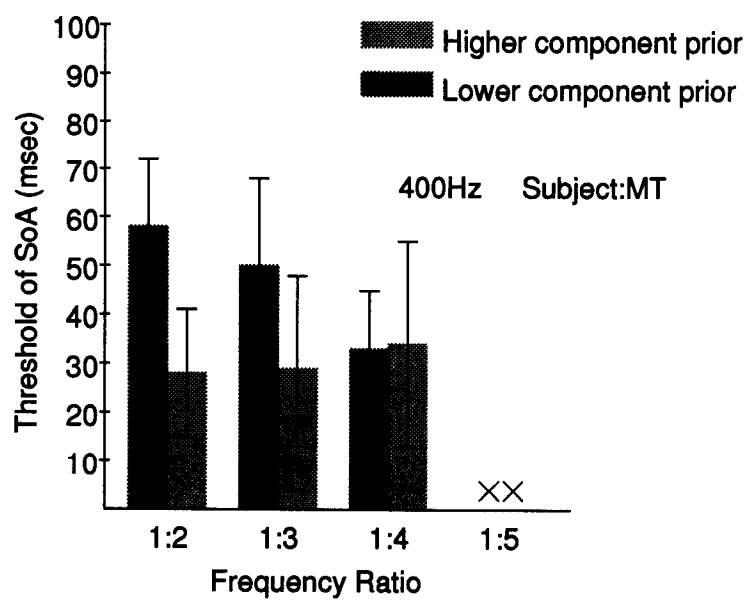


d)

図 2.7: 実験結果



e)



f)

図 2.7: 実験結果

2.3.4 考察

実験結果から、整数倍音が基音よりも先行して立ち上がると分離しやすい傾向にあることが確かめられた。本実験では、基音に対して第5倍音まで組み合わせた2音について評価を行ったが、周波数比が1:2,あるいは,1:3のとき,閾値の差が最大になり,それ以降ではその差が減少することが明らかになった。基音が1kHzのときは周波数比が1:3,400 Hzのときは1:2になると閾値の差が最大になっているが,これは基音の高さによって,群化作用が働く周波数比の広さが異なっている可能性があることを示している。このことは,基音と整数倍音との群化が,第2倍音から第3倍音あたりまでにおいて最も強く作用していることを示唆している。また,本結果では,2音の周波数比が1:4以上になると,同時に立ち上げても分離して知覚される傾向にあった。これは,たとえ2音が倍音関係にあっても,必ずしも融合して一つの音として知覚されないことを示している(図2.8)。このような結果が得られた理由としては,融合して一つの音に知覚するには,周波数比が大きすぎるためだと考えられるが,基音とそれぞれの倍音との音圧比も,分離に影響を与えていると考えられよう。なぜなら,我々は,音圧が等しい基音と高次の倍音とを一つの良く融け合った音として知覚することは難しいからである。少なからず,基音に対する倍音の音圧の方が低いと考えられる。事実,楽器音には,そのようなスペクトルパターンを有しているものが多い[6]。

次節では,この2音の相対的な音圧を考慮した音源分離について実験を行う。

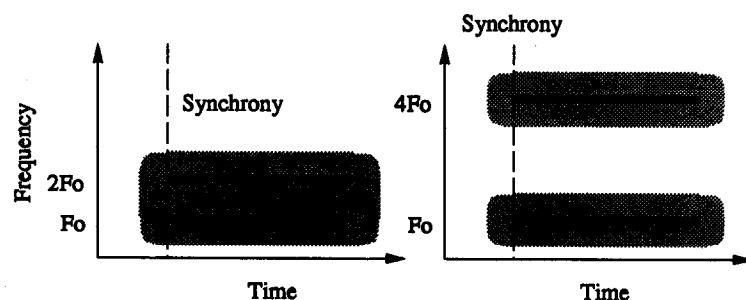


図 2.8: 同時に立ち上げても分離される例

2.4 実験3：立ち上がりのずれと音圧比との関係

2.4.1 目的

前節までに、2音の周波数が倍音関係にある場合、倍音が先行して立ち上がると分離しやすい傾向が得られた。前の実験ではいずれの場合も、2音の音圧は同じ値に設定して行っていた。そこで本実験では、周波数比が1：2である2つの純音を用いて、二つの音の音圧比をコントロールしたときに、立ち上がりのずれが分離に与える影響を考察した。

2.4.2 条件、方法

100, 200, 300, 400, 600, 800, 1000, 1200, 1400, 1600 Hz を基本周波数として、それぞれ周波数比が1：2にある10個の組合せ(表2.4)に対して、実験を行った。このとき2音の音圧は、基音の音圧を固定にし、倍音の音圧を基音のそれに対して0 dB から-10 dB まで2 dB ずつ6段階に変化させた。なお、基音と倍音の振幅値をそれぞれ p_0 , p としたとき、相対的な音圧は、

$$\alpha = 20 \log \frac{p}{p_0}$$

と表される。試行回数が300セットになる点を除き、実験条件と方法は実験1と同じである。

表 2.4: 提示刺激音 (Hz)

周波数比 1:2	
100 + 200	800 + 1600
200 + 400	1000 + 2000
300 + 600	1200 + 2400
400 + 800	1400 + 2800
600 + 1200	1600 + 3200

2.4.3 結果

結果を図 2.10, 2.9 に示す。これらのグラフは、いずれも一人の被験者が一連の試行を 2 回行って得られた結果の平均をとったものである。

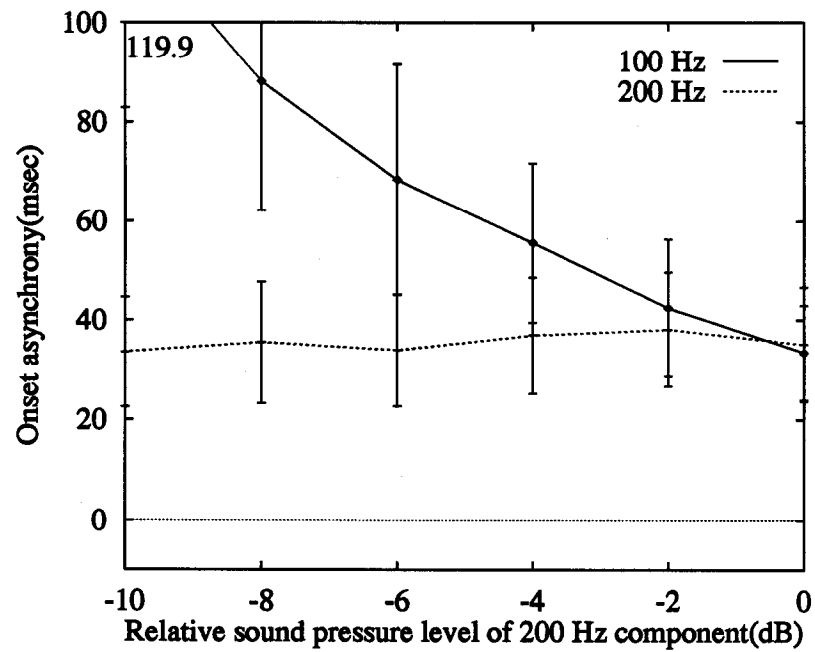
図 2.9 は、条件の節で述べた周波数を用い、それぞれの基音に対して、倍音との音圧比をコントロールしたときの、ずれの閾値を示している。グラフの横軸を基音に対する倍音の相対的な音圧、縦軸は 2 音に分離して知覚される立ち上がり時刻差の閾値を指している。

本実験より、基音の周波数が 100, 200 Hz の場合、倍音の相対的な音圧が下がるにつれて、2 音に分離して知覚される立ち上がり時刻のずれの閾値の差が大きく広がる傾向が得られた。特に、基本周波数が 100 Hz のとき、基音と倍音との閾値の差は大きく開いた。これに対し、基音の周波数が 400 Hz 以上に高くなると、そのような傾向は見られず、閾値の差は音圧比に依存せず、ほとんど同じ差で変移した。また、周波数が 600 Hz の場合、一部に閾値の差がなく、それぞれの閾値の値はほぼ同じである結果が見られた。

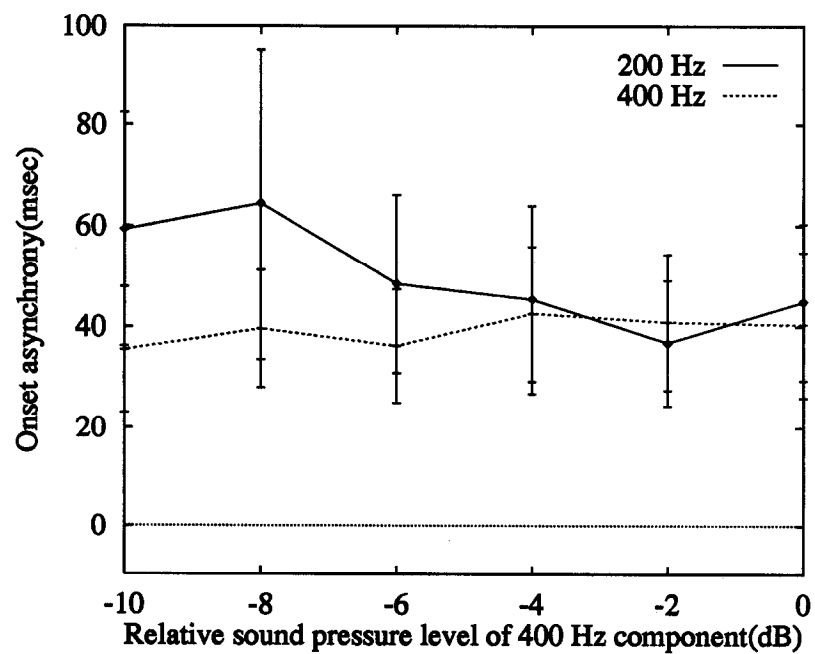
また図 2.10 は、2 音の音圧比が 0 dB ときの 2 音に分離して知覚される立ち上がり時刻差の閾値を 100~1600 Hz にわたって、閾値の変移を調べたものである。横軸は基音の周波数を、縦軸はずれの閾値を表している。また、実線は基音を、点線は倍音の閾値を指す。音圧比が 0 のとき、100 Hz を除いては、基音の閾値よりも倍音のそれの方が小さくなっている。また、基音が 1600 Hz 近くの高周波数になると、閾値の差が広がる傾向が得られた。

上記の結果をまとめ、全てのパラメータに対して 3 次元的に表示したものが図 2.11 になる。ここでは、倍音が先に立ち上がる場合の閾値と、基音のそれとの差の大きさをプロットした。すなわち、倍音の閾値が基音よりも大きければ正の値をとる。

本結果からも、前節までの結果と同じように倍音の相対的な音圧を下げて、それが基音よりも先行して立ち上がると 2 音に分離して知覚される結果が得られた。また、基本周波数が 100 Hz の場合、音圧比が 0 dB のとき、200 Hz の場合、-10 dB のとき、基音と倍音のどちらが先行して立ち上っても、閾値はほぼ同じになった。

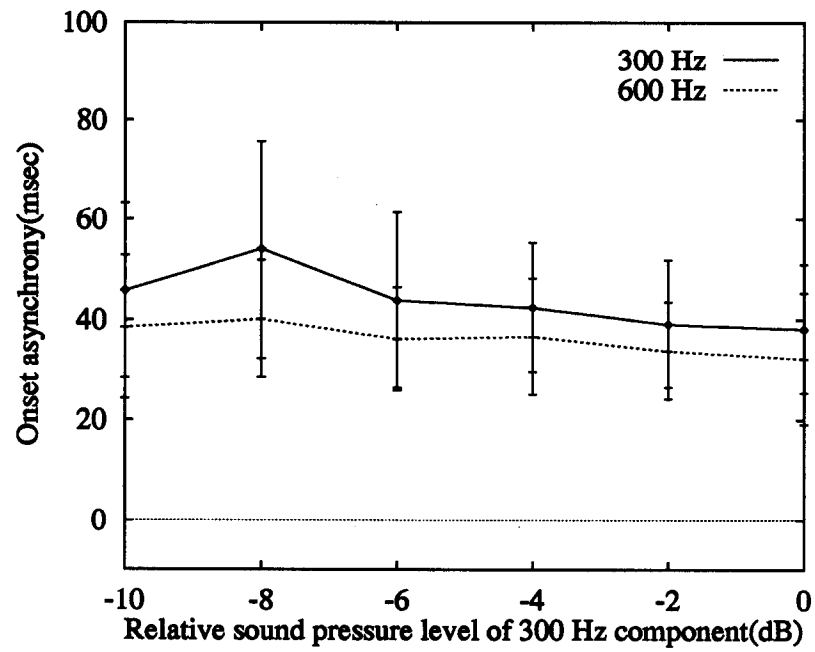


a)

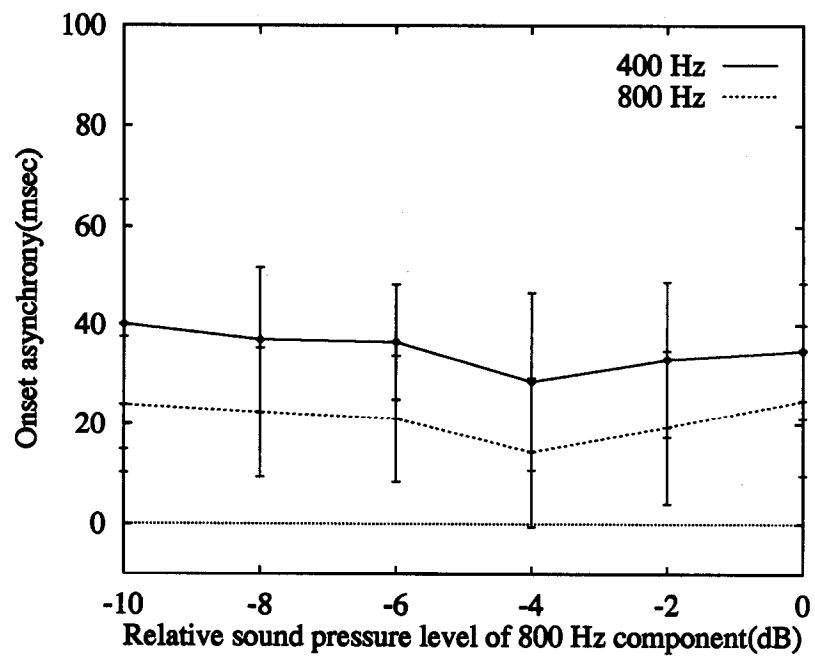


b)

図 2.9: 実験結果

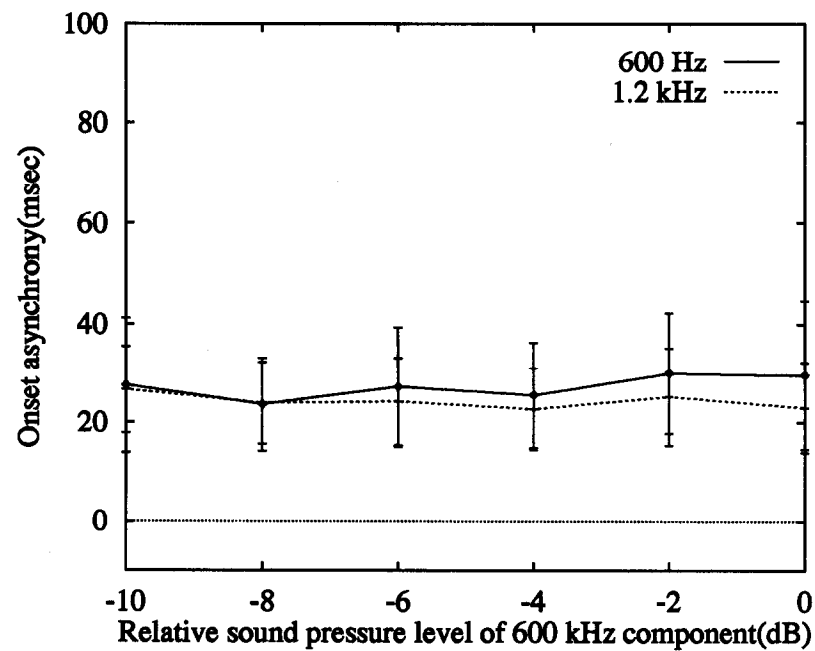


c)

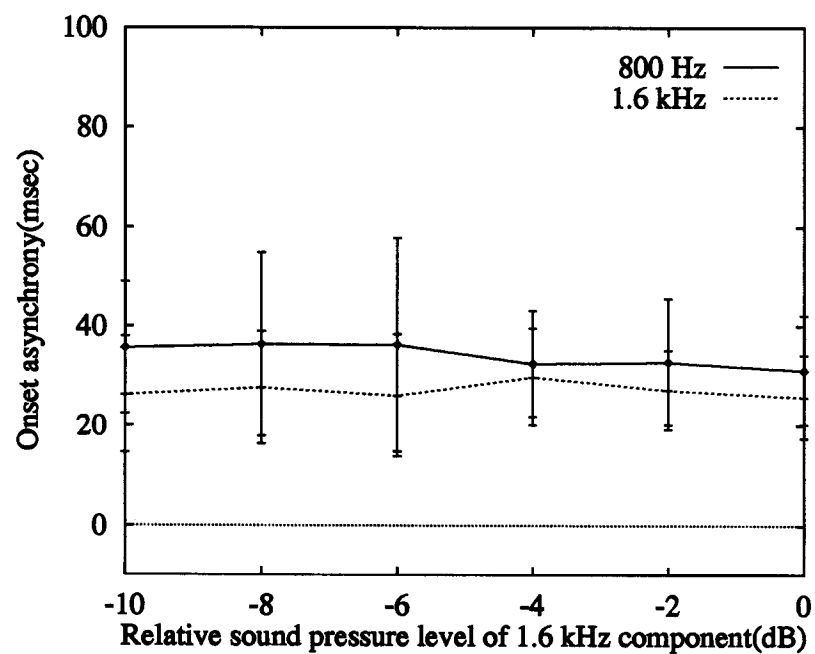


d)

図 2.9: 実験結果

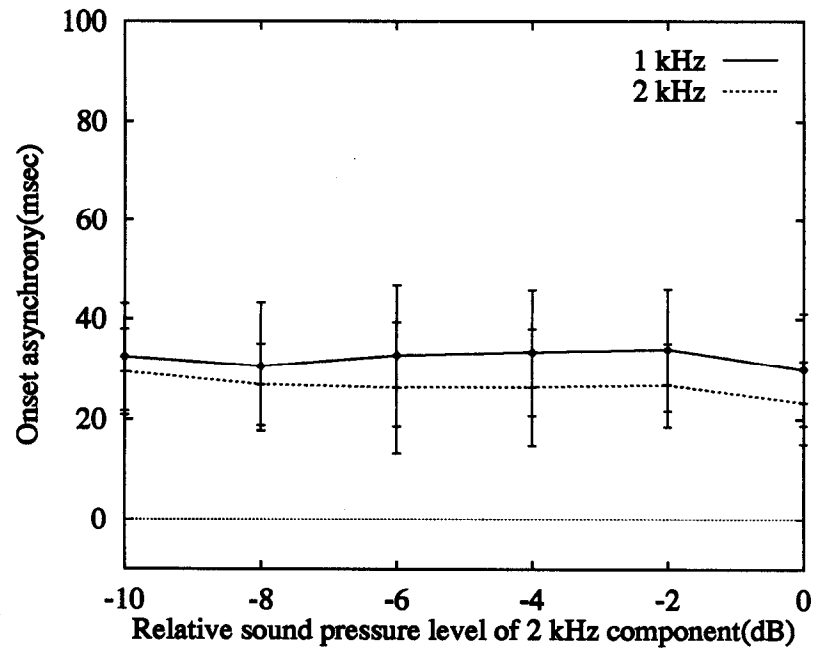


e)

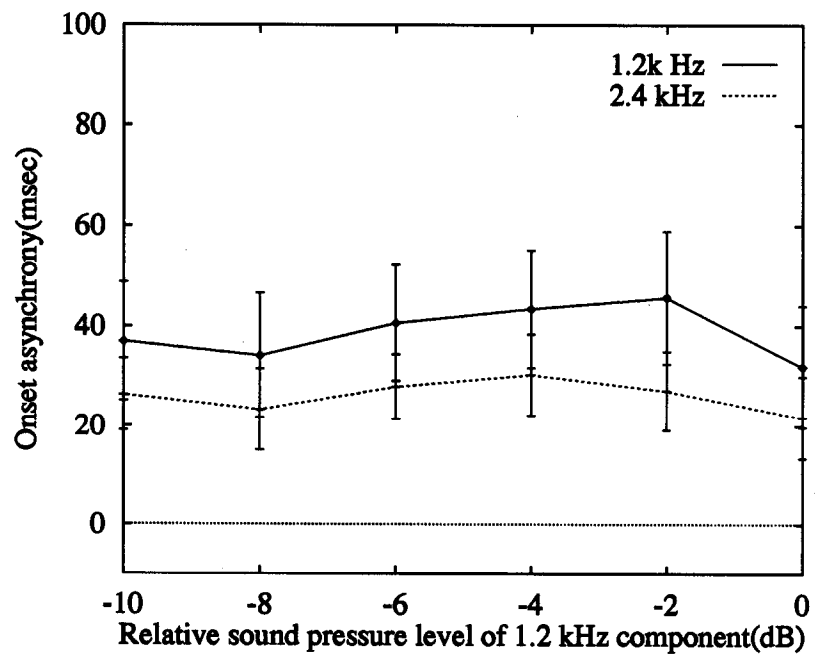


f)

図 2.9: 実験結果

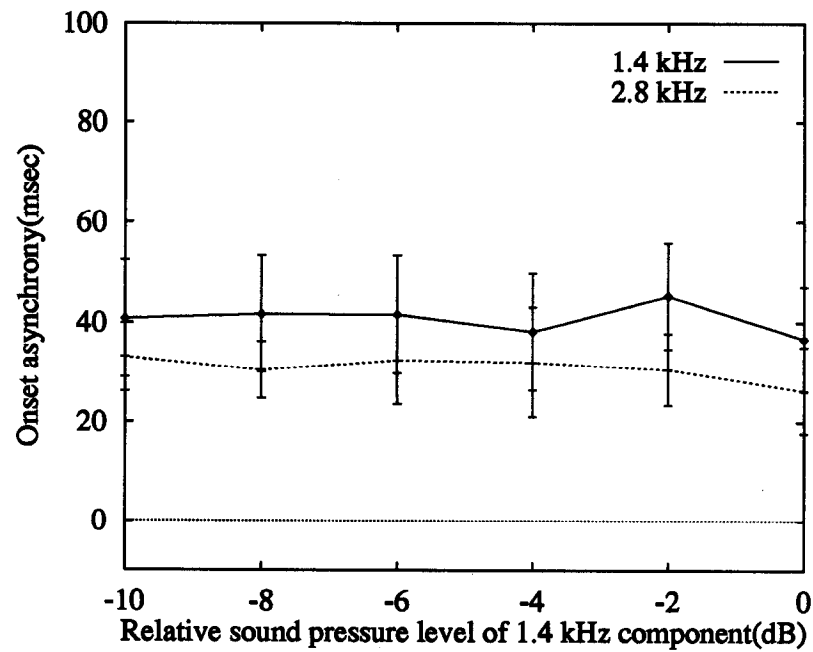


g)

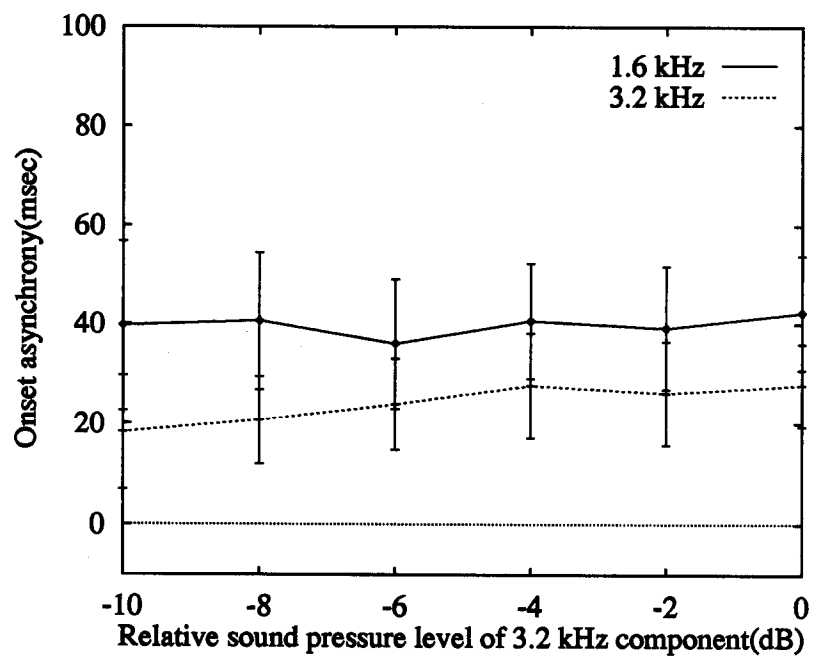


h)

図 2.9: 実験結果



i)



j)

図 2.9: 実験結果

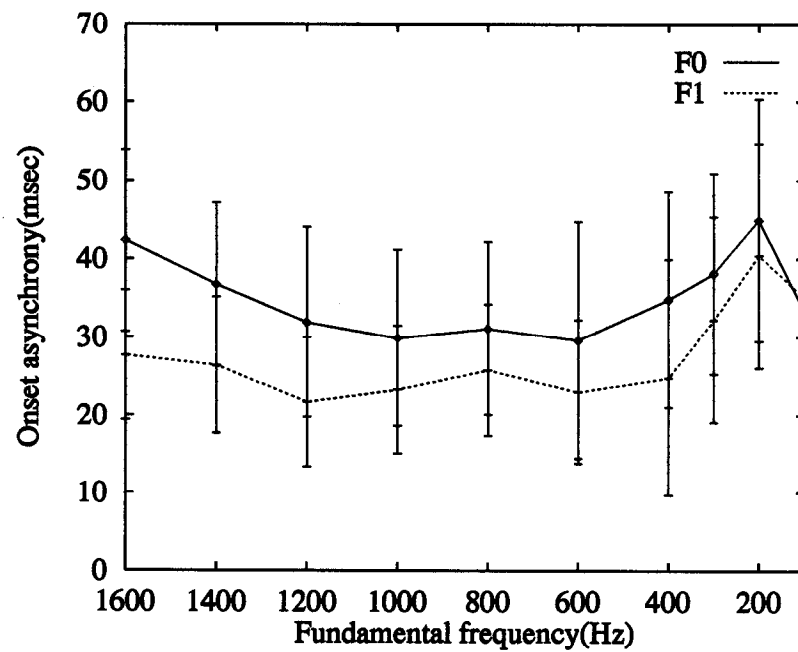


図 2.10: 実験結果 (0 dB の場合)

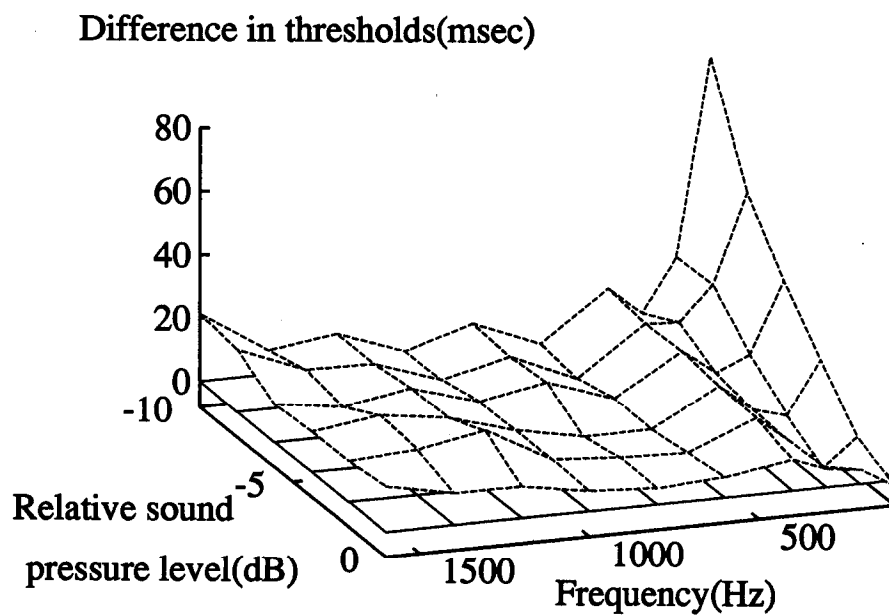


図 2.11: 実験結果 (3次元表示)

2.4.4 考察

本実験で用いた刺激音の組み合わせは、すべて周波数比が1：2であるので、倍音が先行すると分離して2音に知覚されやすいことは既に前節までに述べた通りである。

本結果において、基音の周波数が100, 200 Hzの場合、倍音が先行する際の閾値は、倍音の音圧が減少しているにもかかわらず変化がないのに対して、基音が先行する場合は、その閾値が音圧に依存して大きく増加する結果が得られた。このことから、基音が先行すると一つに知覚されやすく、また、このとき倍音の音圧が基音よりも低いと、2音が同じ音圧であるときよりも、基音と倍音の群化が一層強く作用するという仮説が得られる。本結果は、基音に対する倍音の相対的な音圧を減少させていくと、基音の周波数により分離の仕方に違いがあることを示している。

また本結果は、前節までの結果と同様に、音圧比がマイナスであっても、倍音が先行して立ち上がると2音に分離して知覚されやすいことを示している。この結果は、複合音を使用した類似した研究 [23] からも支持されている。それによると、2つの複合音のうち基音の高い音の音圧を、もう一方の音の音圧を基準として-60 dBまで低下させても、基音の高い音が低い音より、30 msec先に立ち上がると、その基音の高い音を知覚できるというものである(図 2.12)。これらは、立ち上がりのずれが音源分離に強く影響を与えていることを裏付けている。この音圧を考慮した音源分離については、後の章でさらに考察する。

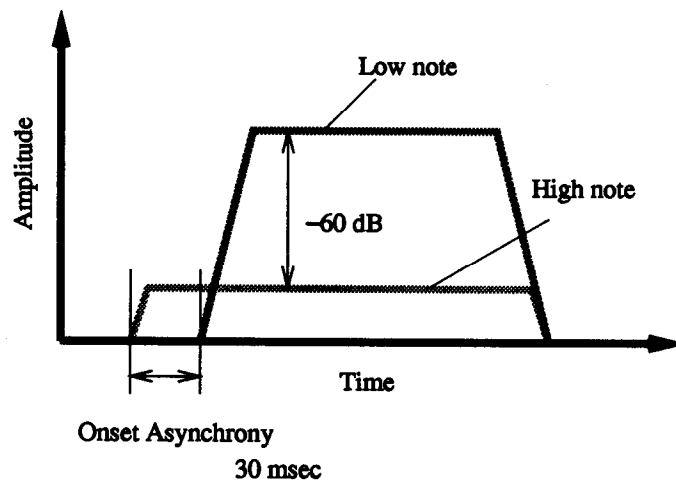


図 2.12: 刺激の概念図

ところで、基音の周波数が、100 Hz, 200 Hz になると、音圧比が0, -10 dB のとき、また、周波数が600Hz で音圧比が-8, -10 dB のとき、基音と倍音、どちらが先行して立ち上がっても、2音に分離して知覚される時刻差の閾値は、ほぼ同じであった。基音と倍音の群化が働いていれば、倍音が先行すれば分離しやすいはずである。しかしながら、そのような影響は見られなかった。それぞれの結果において、基音あるいは倍音のどちらが先行して立ち上がっても分離の仕方に違いはないことを示している。さらに、基音が1.6 kHz になると、ずれの差が再び大きく開いている。後述するように、周波数が1.6 kHz のように高い成分から立ち上がる音は、自然の音にはほとんど存在せず、普段、耳慣れない音である。それにもかかわらず、ずれの差が大きく開いていることは、このような高周波数域でも群化が作用していることを示唆している。現在のところ、これらの結果を説明する解釈は得られていない。

第 3 章

一般的考察

3.1 音の群化・分離と周波数関係

以上の実験では、2つの音が倍音関係にある場合、高い周波数成分から立ち上がる方が分離しやすい結果が得られている。これは逆にとらえれば、低い成分から立ち上がる方が一つの音として知覚されやすいことを示している。楽器音は、基本周波数や第一倍音といった低い周波数成分から立ち上がることが、実験的に知られているが[3, 11]、楽器音に限らず世の中のほとんどの音も(人工的な音をのぞけば)、同様に低い周波数成分から立ち上がっていると考えられる。また、多くの場合、楽器音と同じく協和構造を持っていると考えられる。このことは、人間は協和構造をもつ、低い周波数成分から立ち上がる音を日常的に耳にしていることを示している。したがって、頻繁に耳にしている音の構造を脳が学習によって獲得した副作用として、低い音が先に立ち上がると音の群化が促進され、このような非対称な結果を生んだものと考えられる。

ところで、ある音が立ち上がれば、基音、2倍音、3倍音と言った順番で立ち上がるため、2倍音、3倍音がある時間ずれで一緒に耳に入ってくることは頻繁にある。よって、周波数比が2:3の場合も周波数比が1:2のときと同じように群化すると考えられる。これは、3倍音、4倍音でも同様であろう。しかし、実験1によれば、特に、低い周波数成分を400Hzとした場合には、低い音が先行して立ち上がった方が2音に分離されやすいという結果が得られた。この場合、上で述べた群化が作用しているという考えがあてはまらない。この非対称性は、なぜ生じるのであろうか。

その理由の一つとして、後節で述べるように、被験者が音の分離と音色の変化を混同して知覚していることが考えられる。つまり、低い音が先に立ち上がる場合は、被験者が音

の分離そのものではなく、音色の変化をもって音が分離したと判断してしまうために、低い音が先行すると分離しやすいという結果が得られたと考えられる。しかし、この問題に関しては、適切な実験方法を考慮した上で、今後の実験で明らかにしていく必要がある。

3.2 音圧を考慮した群化と分離

2音が倍音関係にあるとき、倍音が先に立ち上がると2音に分離して知覚されやすい。また、倍音の音圧が低下していても、人間は二つの音に分離して知覚する傾向があることは既に述べた。これは、基音による倍音の群化が働いているものと考えられ、2音が同じ音圧でなく倍音の音圧が低い場合でも、倍音が先行して立ち上がると、基音からの群化から解かれ、分離して知覚しやすいことを示す一方で、音源分離において立ち上がりが重要な役割を演じていることを意味している。

ところで、先に述べたように、楽器音や世の中の自然の音は、基音あるいは低次倍音から立ち上がることが多いが、それと同様に高次倍音ほどその音圧が小さいことも知られている [6]。また、音の立ち上がりが同じで倍音の音圧が基音に比べ低いと、音を一つの音と知覚しやすいことも報告されている [5]。つまり、倍音関係にある2音の音圧が同じであること自体が、2音が同時に立ち上がっても分離を促す要因の一つになり得ると考えられる。

この問題は、実験2で直面したように、周波数比が1:4, 1:5の倍音関係にあり、同時に立ち上がっても分離して知覚された報告に関連していると考えられる。なぜなら、周波数構造が整数倍音になっており、立ち上がりが同時である音は、一つの音として知覚されやすいと考えられていたのに対し、本研究から得られた結果では、分離して2音に知覚されたからである。そこで、本研究では、あくまでも一つの解釈であるが、倍音関係にある2音の音圧比がその原因の一つであると捉え、その音圧比を操作し、このパラメータが音源分離に与える影響を調べることにした。前節2.4で述べたように、基音の周波数が200 Hz 以下になると、基音が先行する場合と倍音が先行する場合との閾値の差は、音圧が減少するにつれて大きくなる傾向が得られた。その一方、基音が400 Hz 以上になると、閾値の差は音圧比に関係なくほとんど変化しなかった。この結果の一つの解釈としては、脳が始終耳にしている音と同じ構造を持つように音を群化して知覚しているからであり、

ある意味で世の中の音の構造を反映している結果であると考えられる。

以上により、基本周波数が高い音よりも低い音の方が、群化の作用が強く働き、それゆえ、倍音の音圧が下がるにつれてより一層結び付きが強くなり2音を一つの音として知覚されやすくなる。一方、基本周波数がある程度の高さを持つと、2音に分離する際の閾値が音圧に依存しないようになるのであろう。その結果、基音が先行して立ち上がった場合と、逆に倍音が先行する場合の閾値との差が広がると考えられる。

しかしながら、ここでも重要なことは、倍音の音圧が基音よりも減少していても倍音が先行すると2音に分離しやすい結果は、立ち上がりのずれが音源分離に強く影響を与えている証拠であるといえよう。

3.3 分離と音色の関係

本実験では、2音が分離して知覚される時の立ち上がり時刻の差の閾値を実験を通して測定した。これは、刺激音をランダムな順序で被験者に呈示し、それぞれが融合して聞こえたか分離して聞こえたか強制的に解答 (AFC: Alternative Forced Choice) させて得られた値である。ところが、この実験を通して、先行して立ち上がる音の高・低によって、その知覚のされ方も異なって感じられることがわかった。被験者の報告によれば、高い音が先行する場合、立ち上がり時刻の差によって違いはあるものの、2音は明らかに分離して知覚されるのに対し、低い音が先行する場合は、かなり大きな時刻のずれがないかぎり、2音を明確に知覚することができず、呈示された刺激音が、2音に分離したのか、あるいは単に音色が変化したのか判断に迷うことがあったという。このように、低い成分から立ち上がる時に分離したのか音色が変化したのか分からないという事実は、被験者が低い音が先行して立ち上がる音を、一つの音として受け入れやすいことを表している。しかし、同時にこの種の実験を行う時の手続きが難しいことを示している。

本研究の目的は、音の立ち上がりが音源分離に与える影響を調べることであるので、被験者には「分離したかどうか」だけの判断をするように指示し、分離と音色の変化とは別なものとして扱った。とはいえ、この判断の基準は、被験者ごとに違っている可能性が大きいいため、得られた結果を絶対的な分離の閾値として解釈することは危険である。しかし、周波数比が整数倍の場合においては、全ての被験者に対して、一貫して、高い音が立

ち上がる方が分離の時刻差が大きくなったことは、本報告で述べた結果が普遍的なものであることを示しているといえよう。

3.4 基音と群化

本論文では、これまで、基音が先行して立ち上がると、基音と倍音との結びつきが増し、2音が一まとまりに知覚される一方、倍音が先行して立ち上がると、2音に分離しやすくなると考えてきた。また、この群化が、倍音の音圧にも依存すると考えられることも述べた。ここでは、周波数方向の群化において、基音がその役割を果たす重要さを、音の周波数構造といった物理的性質を勘案して考察する。

基音とは、文字通り周期的複合音に含まれている複数の正弦波(部分音)の基礎となるものであり、各倍音はその基本周波数の整数倍の周波数を持っている。また、この基音は、複合音の波形全体の繰り返し率に等しい周波数を持つものである。つまり、基音の整数倍の周波数をもつ倍音成分の数がいくら増えても、複合音全体の波形は大きく変化せず、複合音の高さは基本周波数の高さと関連づけられている。このような性質は、人間がその複合音を知覚する際、基音と倍音を群化して知覚していることを指していると考えられる。知覚的な音源分離の物理的な手がかりとして基本周波数があることは紹介したが、このように基本周波数の異なる楽器音や母音といった二つの調波複合音を聞いたとき、我々はどの倍音成分がどちらの音の成分であるかわからないことはない[10]。また、調波複合音中のある一つの倍音の周波数が倍音関係にある周波数成分から十分にずれたときは、複合音の中からその倍音の周波数にあたる純音が分離して知覚され[24]、そのとき、倍音成分を他の成分よりも音圧を下げて分離して知覚することができるという[25]。これらの事実も、基音と倍音の群化という点から捉えることができよう。基音との群化が働いているため、2音の基本周波数が異なれば、その倍音成分どうしを混同することなくそれぞれの音に割り振ることができ、また、倍音関係から周波数成分がずれたときに、その成分を分離して知覚することできると考えられる。また、基音が倍音より先行した場合に、それが2音に分離したのか、あるいは、音色が変化したのか判断に迷うと言う被験者からの報告も、基音と倍音との群化のため、後から立ち上がった倍音成分を包含し、全体を一つのまとまりへと知覚すると考えられる。これらは、自然界に存在する周期性を持った音が調波

構造を持つことが多いということからも、基音と倍音成分との群化が作用していることを示している。

3.5 問題点

本節では、本実験を通して浮き彫りになった問題点を取り上げ、その改善を含め検討を行う。

- 刺激音の音圧レベル。

音の絶対的なレベルを記述するためには、音圧がある基準の音圧 p_0 よりも何 dB だけ上である、あるいは下であると述べる必要がある。求める音の音圧を p としたとき、この音の音圧レベル α は、

$$\alpha = 20 \log \frac{p}{p_0}$$

となる。ただし、一般に $p_0 = 20 \mu\text{Pa}$ としている。しかしながら、現在の実験環境では、実験機器が不足しているため、刺激音の音圧レベルを測定できない状態である。音圧レベルを測定する際は、騒音計やブローブマイクといったものを利用して測定するが、受聴方法により測定方法も異なり、また音響環境にも依存するので、簡単には測定できない。

本実験では、刺激音の絶対的な音のレベルに着目して実験を行っているわけではないので、厳密なレベルの測定は必ずしも必要ではないといえるが、実験的に研究を行う以上、環境を整備し、実験条件を客観的に記述できるようにすることは不可欠である。精密な実験については今後の研究を待ちたい。

- 刺激音の位相による影響

従来、人間は位相の変化に対して鈍感であると考えられてきたが、最近の研究により、位相の変化が音色に影響を及ぼすことが明らかになった [11]。このことは、実験を行う上で2つの音の位相関係を考慮する必要があることを示している。しかしながら、本実験では、一方の音が立ち上がったのち、わずかな時間遅れでもう一方の音が立ち上がる、その時点の2音の位相をそろえる程度にとどめた。具体的には、

本研究で行った実験は、すべて2つの純音を用い、一方の音を固定し、他方の立ち上がり時刻をずらした。この際、ずれの大きさは10 msec 単位にしており、また、固定した純音の周波数は、すべて10 msec の中に整数個の周期を含むものである。各ずれでもう一方の純音が立ち上がる時の位相が同じになっていることになる。しかしながら、これ以上の位相についての議論は、一層刺激のパラメータを複雑なものとし、本来の研究目的からはずれる危険があるため、本研究では、これ以上の位相の操作は行わなかった。

- 刺激音の提示方法

本実験では、恒常法を用いて刺激音を2音に分離して知覚される閾値を測定した。実際の提示方法は、2音の立ち上がり時間の時刻差をパラメータとして、15個のずれをランダムに3回ずつ提示したものを1セットとし、それを5セット繰り返す方法である。この試行回数について、通常、psychometric function を得るには、一つの刺激に対して少なくとも25～30回聴取することが望ましいといわれている。しかしながら本実験は、被験者の負担を軽減するために、その半分の回数としたが、音源を分離する傾向をみることはできたといえる。

提示方法においては、本論文の考察部分でも述べているが、提示された刺激音が2音に分離されたのか、あるいは単に音色が変化したのか判断に迷いが生じる結果となった。この回避策としては、立ち上がり時刻のない標準刺激とそのずれがある比較刺激との対を被験者に提示する方法や、また判断を強制的に「分離」「融合」のどちらかを選択させるのではなく、「分離」と「融合」の間に段階的な選択肢を与え、その中から判断を行わせる方法もある。

- 音響環境

これは、聴覚の心理実験を行う上で、重要な問題である。現在の環境は、計算機やその他の周辺機器があり、また人の話し声、車の音も聞こえる普通の部屋である。実験中にドアの開閉、廊下の話し声だけで実験がやり直しになる場合もある。HDDから生じる音や、ディスプレイから発せられる高周波も実験に影響を与える可能性がある。理想的には、実験は被験者を他の騒音から遮断された音響室内で行う必

要がある。また、受聴条件には、スピーカ等を用いた音場受聴とヘッドフォン等を用いた受話器受聴の2通りがある。本実験は後者の方法を用いている。この選択には実験の性格や実験条件から来る制約、提示音圧の校正法等によって決められるが、ヘッドフォンは人間にとって不自然な受聴方法なので、スピーカを用いた方が望ましいということも考えられる。

3.6 今後の展開

実験3では、基音に対して倍音の相対的な音圧を0 dB から-10 dB までとして実験を行った。この範囲において、基音が100, 200 Hzと言った低い音であると2音が分離して知覚される立ち上がり時刻の閾値の差が広がる傾向が得られたが、その音圧比の範囲をもっと広く、例えば-20 dB までにすれば、閾値の差に変化が見られなかった周波数においてもある程度の傾向が得られる可能性も考えられる。逆に、高周波成分の音圧を大きくした場合についても調べる必要がある。

また、本実験では周波数比が1:2の2音の組み合わせに対して考察を行ったが、2音の周波数比をさらに広げてみることや倍音関係にない組み合わせに対して同様な実験を行うことも興味がある。例えば、音圧比が1で群化が作用しない2音に対しても、一方の音の音圧が減少することにより他方の音との群化が作用し、2音間の結び付きを増幅させるものと考えられるからである。

楽器音や世の中の音の多くは、倍音構造をなしているといえるが、完全な倍音構造をなしていない等の複雑な構造を有している音が存在していることも確かである。また、複合音中の各倍音成分の音圧も人間の知覚に影響を与えている。そのような構造を考慮すると、上記の通り、本研究においては、まだ多くの課題が残されている。

本研究は、聴覚心理の立場から、音の立ち上がりが音源分離に与える影響を三人の被験者に対して心理実験を行い、そのメカニズムについて考察を行ってきた。被験者の数は三人ながら、前述してきたような音源分離の傾向が得られたが、心理実験から得られた結果について統計的な評価を行うためには、膨大なデータを必要とするものである。本研究で得られた傾向を、一層強く支持する結果を得るべく、一人でも多くの被験者に対して実験を重ねていく必要があるといえる。

第 4 章

結論

人間は知覚的な音源分離を行う際、様々な手がかりを用いて、複数の音源から発せられた音を含む混合音から、周囲で生じている事象を復元する逆問題を解くことができる。本研究は、そのような手がかりの一つである音の立ち上がりに着目した。音の立ち上がりは、人間の音源分離知覚に大きな影響を与えるだけでなく、音色や音の認識においても重要な役割を果たしており、音源分離と音の認識は相互に関係していると考えられる。そこで、本研究では、2つの純音を用い、音の立ち上がりのずれが音源分離にどのような影響を与えるかを実験的に検討し、人間の音源分離メカニズムを解明するべく、音の群化と周波数構造の視点から考察を行った。

まず、音源分離における立ち上がりと協和度との関係を調べた。2音の音圧が等しい場合、2音の協和度が高いとき、特に周波数比が1:2であるとき、高い音(倍音)が低い音(基音)よりも先に立ち上がると分離しやすいことが明らかになった。協和度の低いその他の周波数比では、どちらの音が先に立ち上がっても、分離の仕方に大きな違いは生じなかった。

次に、立ち上がりのずれと整数倍音との関係を調べた。この場合、整数倍音が先行すると2音に分離しやすく、基音が先行すると一つの音に知覚される傾向が見出された。また、基音と倍音それぞれが先に立ち上がった時に2音に分離して知覚されるずれの閾値の差は、周波数比が1:2あるいは1:3程度になると顕著であった。しかし、周波数比が1:4以上に大きく開くと、同時に立ち上げても分離して知覚されるという結果が得られた。

最後に、周波数比が1:2である2音について、2音の音圧比をコントロールしたとき

に、閾値がどのように変化するかを調べた。その結果、倍音の音圧を基音に対して下げていくと、基音の周波数が低い場合に限り、基音が先行して立ち上がると一つの音として知覚する傾向が得られた。また、逆に倍音が先行して立ち上がる場合は、倍音の音圧が基音より低くても分離して知覚されることが明らかになった。

また、全体の実験に共通して得られた知見として、基音が倍音より先行して立ち上がると、被験者は、その2音が分離したのか、あるいは、単に音色が変化しただけなのか判断に迷いが生じることが挙げられる。

以上の結果から、倍音が先行すると基音と倍音の群化が働かなくなり2音に分離して知覚され、逆に基音が先行して立ち上がると全体を包含して一つの音にまとまりやすいという結論が導かれる。その例外として、周波数比が離れた場合、倍音関係にあり、かつ、同時に立ち上げても、2音に分離して知覚される傾向が見られた。その原因としては、2音の一つに知覚される周波数の差に制限があること、また、2音の音圧が同じであることが考えられる。このほか、倍音の音圧が基音より低く、基音から立ち上がると一つの音に知覚されやすいという事実は、基音と倍音との群化が、両者の音圧比に依存することを示唆している。

これらの考察は、世の中の大部分の音の周波数構造が、1) 高調波関係で構成されている。2) 低い周波数成分から立ち上がっている。3) 高次倍音へいくほど音圧が下がっているという事実と整合する。このことから、日常我々が耳にしてきた音の構造を脳が学習し、その結果として音を過去の経験に基づいて群化していると考えられる。

本研究は、2つの純音に対して実験を行ってきたが、その目的は、音源の分離メカニズムを解明することにある。本研究では、たとえ実験的な音であろうと、評価、検討を行いやすい純音から研究に着手し、いずれ複合音、楽器音へと展開していくための橋渡しになることを狙った。そういう意味では、知覚的な音源分離に関する研究の初期段階に位置しているが、人間の音源分離メカニズムの解明を立ち上がりという切り口から攻めていくという点では、有効な研究であると考え。今後、刺激として用いている音を複合音や楽器音へと広げて実験を行うとともに、音源分離の計算理論を構築し、また、その神経メカニズムを解明する研究へと発展させていきたい。

謝辞

本論文の作成にあたり、主任指導教官である阪口 豊助教授に終始懇切なご指導を頂いた。先生には、ご自身がアメリカに滞在し研究活動に励んでおられるところ、Fax や Mail を通し筆者の研究を支援して頂いた。時間、空間的な障害を感じさせないそのご配慮に、大変心を打たれるものがあつた。研究に取り組む姿勢、考え方を筆を通して教えて頂いた。

本研究を進めていくうえで指針となるべき有形無形の指導を出澤正徳教授に頂いた。本年度は、大変お忙しいところ、折に触れ、参考になる文献を多数紹介して頂いた。

施衛富助手にはコンピュータ・プログラミングについて、また、研究を進めていく上で多くのご教示頂いた。

また、本学情報工学科の高澤嘉光教授には、研究環境や実験手順について指導を頂いた。お忙しいところ、研究室まで見学させて頂けた。聴覚研究の難しさを教えて頂いた上、関連する文献まで紹介して頂いた。

外部の研究者の方々にも貴重のご助言を頂いた。NTT基礎研究所の柏野牧夫氏、柏野邦夫氏、植松 尚氏、NHK技術研究所の小野一穂氏。これら多くの研究者の方々からの助言がなければ、本研究を進めることはできなかったと言っても過言ではない。深く感謝している。

研究室は異なるが、研究活動を通して生じた不安や悩みの多くを聞き入れて頂いた弓場研究室博士1年の山口佳紀氏、専攻事務室の平野氏、岩瀬氏に深く感謝する。研究という枠を超えて、多くのご助言を頂いた。

最後に、常に活発な研究を共にし、折に触れてご討論、ご討議頂いた出澤研究室、阪口研究室の皆様、他の研究室の友人達に厚くお礼申し上げます。

参考文献

- [1] Bregman, A.S.: *Auditory Scene Analysis*, MIT Press(1990)
- [2] 柏野牧夫: “音の流れを聞きとる,” 科学, **62**, 6, 374-379(1992)
- [3] Deutsch, D.(寺西立年, 大串健吾, 宮崎謙一訳): 音楽の心理学(上), 西村書店(1987)
- [4] Kubovy, M., Cutting, J.E. and McGuire, R.M.: “Hearing with the third ear: Dichotic perception of a melody without monaural familiarity cues,” *Science*, **186**, 272-274(1974)
- [5] Handel, S.: *Listening*, MIT Press(1989)
- [6] 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, **33**, 6, 291-300(1977)
- [7] Summerfield, Q. and Assman, P.: “Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony,” *J. Acoust. Soc. Am.*, **89**, 1364-1377(1991)
- [8] McAdams, S.: “Segregation of concurrent sounds. I: Effects of frequency modulation coherence,” *J. Acoust. Soc. Am.*, **86**, 2148-2159(1989)
- [9] McAdams, S.: “Segregation of concurrent sounds. II: Effects of spectral envelope tarc-ing, frequency modulation coherence, and frequency modulation width,” *J. Acoust. Soc. Am.*, **89**, 341-351(1991)
- [10] Broadbent, D.E. and Ladefoged, P.C.: “On the fusion of sounds reaching different sense organs,” *J. Acoust. Soc. Am.*, **29**, 708-710(1957)

-
- [11] 難波精一郎編：聴覚ハンドブック, ナカニシヤ出版 (1984)
- [12] 境 久雄：聴覚と音響心理, コロナ社 (1978)
- [13] Moore, B.C.J.(大串健吾監訳)：聴覚心理学概論, 誠信書房 (1995)
- [14] 大串健吾：“音の協和性について,” *NHK技研月報*, **20**, 280–285(1977)
- [15] 柏野邦夫, 田中英彦：“2つの周波数成分からの分離知覚に関する工学的モデル ～複数の要因の評価と統合～,” *信学論 (A)*, **J77-A**, 5, 731–740(1994)
- [16] 中谷智広, 奥野 博, 川端 豪：“音環境理解のためのマルチエージェントによる調波構造ストリームの分離,” *人口知能学会誌*, **10**, 2, 232–241(1995)
- [17] Assmann, P. F. and Summerfield, Q.: “Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,” *J. Acoust. Soc. Am.*, **88**, 680–697(1990)
- [18] 永田仁史, 安倍正人, 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*, **47**, 268–273(1991)
- [19] 黄 捷, 大西 昇, 杉江 昇：“音源の方位情報を用いた複数音源の分離,” *日本ロボット学会誌*, **9**, 4, 409–414(1991)
- [20] 赤木正人：“カクテルパーティー効果とそのモデル化,” *信学誌*, **78**, 5, 450–453(1995)
- [21] Hukin, R. W. and Darwin, C.J.: “Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification,” *Perception & Psychophysics*, **57**, 191–196(1995)
- [22] 竹内 啓編：“統計学辞典,” 東洋経済新報社 (1990)
- [23] Rasch, R.A.: “The perception of simultaneous notes such as in polyphonic music,” *Acustica*, **40**, 21–33(1978)
-

-
- [24] Moore, B. C. J. and Glasberg, B. R.: "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.*, **80**, 479–483(1986)
- [25] Martens, J. P.: "Comment on "Algorithm for extraction of pitch and pitch salience from complex tonal signals" [J. Acoust. Soc. Am. 71, 679–688(1982)]," *J. Acoust. Soc. Am.*, **75**, 626–628(1984)
-

研究発表

- [1] 川田 晋, 阪口 豊, 出澤正徳 “ 知覚的音源分離に関する検討,” 電気通信大学大学院, 第 3 回 IS シンポジウム, Sensing and Perception, 1-4(1996)
- [2] 川田 晋, 阪口 豊 “ 立ち上がりのずれが音源分離に与える影響,” 日本音響学会聴覚研究会発表資料, H-96-77(1996)
- [3] 川田 晋, 阪口 豊, 出澤正徳 “ 知覚的音源分離に関する実験的研究,” 電気通信大学大学院, 第 4 回 IS シンポジウム, Sensing and Perception, (発表予定)