

平成 10 年度

修士論文

音の立ち上がりのずれが音源分離に与える影響

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

9751018 鈴木 基

指導教官

阪口 豊

出澤 正徳

岡本 秀輔

平成 11 年 2 月 3 日提出

目次

1 序論	3
1.1 研究の目的	3
1.1.1 本論文の構成	4
1.2 研究の背景	6
1.2.1 聴覚の情景分析	6
1.2.2 音源分離の物理的要因	8
1.2.3 音の立ち上がりの重要性	8
1.2.4 知覚的音源分離と知覚体制化	10
2 音の立ち上がりが音源分離に与える影響	13
2.1 実験装置	13
2.2 実験 1	14
2.2.1 実験の目的	14
2.2.2 川田の研究	14
2.2.3 刺激	15
2.2.4 実験方法	16
2.2.5 結果	16
2.2.6 考察	19
2.3 実験 2	20
2.3.1 実験の目的	20
2.3.2 刺激	20
2.3.3 実験方法	20
2.3.4 結果	20
2.3.5 考察	24

3	音の立ち上がりが音列の群化に与える影響	25
3.1	音列とその群化	25
3.1.1	音の系列の群化	25
3.2	実験 3	26
3.2.1	予備実験	26
3.2.2	実験の目的	28
3.2.3	刺激	28
3.2.4	実験方法	28
3.2.5	結果	29
3.2.6	考察	31
3.3	実験 4	33
3.3.1	実験の目的	33
3.3.2	刺激	33
3.3.3	実験方法	33
3.3.4	結果	34
3.3.5	考察	39
4	一般的考察	40
4.1	音の立ち上がりのずれが音源分離に与える影響	40
4.2	音の立ち上がりが音列の群化に与える影響	41
4.3	実験での問題点	42
4.4	今後の展開	43
5	結論	44

第 1 章

序論

1.1 研究の目的

我々は視覚や聴覚などの感覚器官を用いて、外界からの多くの情報を分析している。このことは普通にげなく自然に行われるために、我々はそれぞれの感覚器の特性に気づかないまま過ごしているが、感覚器官では外界からの様々な情報から行動に必要な情報を抽出、分析し、それに対応して身体を機能させている。これが、我々が周囲に何が起こっていて何が存在しているのかが「わかる」ということである。

このように、感覚器において複雑な情報処理がなされていると考えることができ、それがどのようにして外界からの情報を捉え、処理しているのかを解明することは非常に興味深い問題である。しかしながら、知覚心理学の研究は視覚を中心として進められ、聴覚の研究は心理学の本流から離れ、むしろ工学の分野と深く結びついた形で行われた。このことは、精密である反面、人間の知覚の本質を理解しようとする態度の乏しいものであった。そのような中、近年、Bregman 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きた [1]。多くの研究蓄積のある視覚研究の手法を聴覚の分野に取り入れることにより、聴覚心理学の研究はこれからより一層の発展を遂げることが期待されている。

本研究では、聴覚心理学の立場から、聴覚の情景分析の機能の一つである知覚的音源分離について、音の立ち上がりに着目して実験的研究を行い、音の物理的特性と知覚的な音の聞こえ方において定量的な関連づけを行い、人間の音源分離における聴覚の情報処理のメカニズムの解明に向けた新たな知見を得ることを目的とする。

1.1.1 本論文の構成

本論文は図 1.1 のような構成になっている。まず第 1 章で、聴覚の情景分析とその代表的な機能である知覚的音源分離を紹介し、本研究で着目している立ち上がり時刻の重要性について述べる。第 2 章では、立ち上がり時刻のずれが音源分離に与える影響について、音事象の刺激を用いて実験を行う。第 3 章では、立ち上がり時刻のずれが音源分離に与える影響について、音列の刺激を用いて実験を行う。第 4 章で、第 2 章と第 3 章の結果から一般的考察を行い、第 5 章で結論を述べる。

音事象と音脈

ここで、第 2 章で扱う「音事象」と第 3 章で扱う「音列」の定義をしておく。

我々の日常生活の中で、一つの音として知覚される事象は、立ち上がり、継続部、立ち下がりという三つの要素からなる時間的な構造を持っている。この三つの要素からなる音の始まりから終わりまでの単位を「音事象」と呼び、この音事象と空白時間との連鎖である音の系列を「音列」と呼ぶことにする。楽器音で例えるならば、音事象は音符一つの音であり、音列は音符が連なったメロディーと考えることができる。我々が普段聞いている日常的な音も、この音事象と音列から成り立っていると考える。

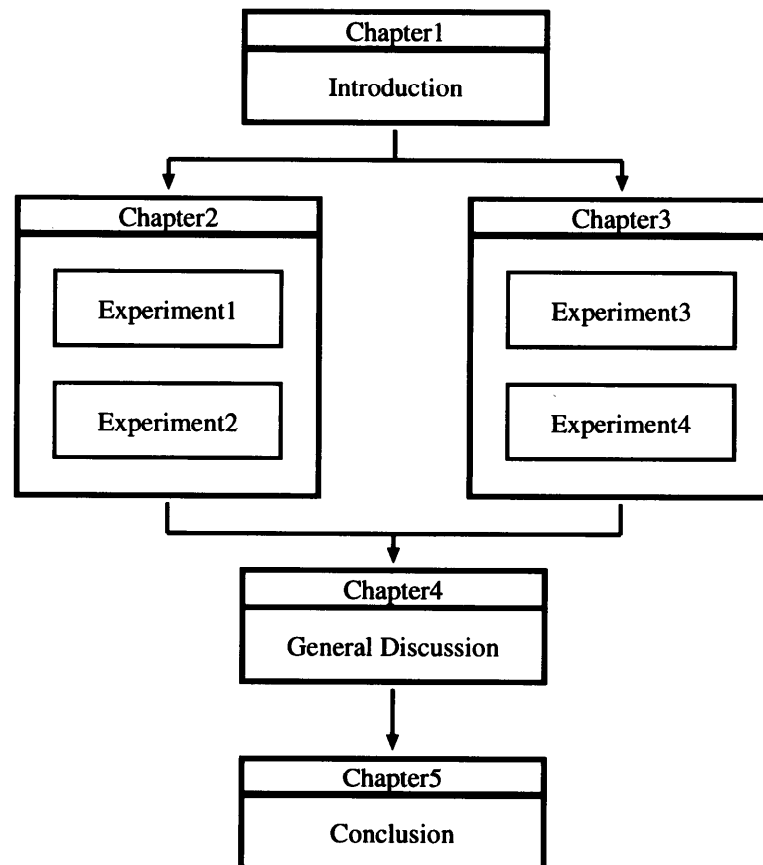


図 1.1: 本論文の構成

1.2 研究の背景

Bregman 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いて聴覚の情景分析の研究が盛んになり、これらの研究から知覚的音源分離や音の錯覚等の新たな知見が報告されている。本節では、聴覚の情景分析とはどのようなものを説明し、それに関連した知見を紹介し、本研究の目的であ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

人間が視覚や聴覚などの五感を用い、感覚器を通して、外界の物体の色、形、匂いや熱といったようなものから周囲で起きている事象を分析している。このような機能を、情景分析 (scene analysis) と呼び、特に音情報から情景分析を聴覚の情景分析 (auditory scene analysis) と呼んでいる。

我々の日常生活の中には様々な音が混在している。しかしながら耳に入ってくる音が、色々な情報が混ざった音であっても、我々は人々の話し声や自動車の走る音といったように聞き分けることができる。また、それらを雑踏という一つの音として捉えることもできる。このような現象は聴覚の情景分析の機能によりもたらされるものであり、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であっても、我々はそれを音源からの一つ一つの音の性質に基づいて、情報を分離、抽出し、同じ性質のものを一つの音のまとまりとして知覚することで処理していると考えられる。

聴覚の情景分析において、情景分析を可能にしている機能の一つに知覚的音源分離機能がある。ここで、音源波形そのものの分離ではなく、人間が一つのもので知覚または認識するような音響エネルギーのまとまりを「知覚的な音」と呼び、複数の音源から目的とする音を知覚的に分離する機能を「知覚的音源分離機能」と呼ぶことにする。この機能の代表的な例として次のようなものがある。

音脈分凝

高い周波数の純音 H と低い周波数の純音 L が、遅いテンポで HLHLHL と交互に提示された場合、人間は全体を H - L - H - L - というように全体を一つのまとまりとして知覚するが、速いテンポで同様に提示した場合には、H - H - H - という高い音のまとまりと L - L - L - という低い音のまとまりに分離して知覚される。この現象は、H と L の周波数の差が大きければ大きいほど音列が分離して知覚されやすくなり、音楽においても、多くの作曲家達がこのような現象の効果をうまく利用している。

Bregman & Campbell は、このように音の属性によって群化し、その要素が相互に知覚

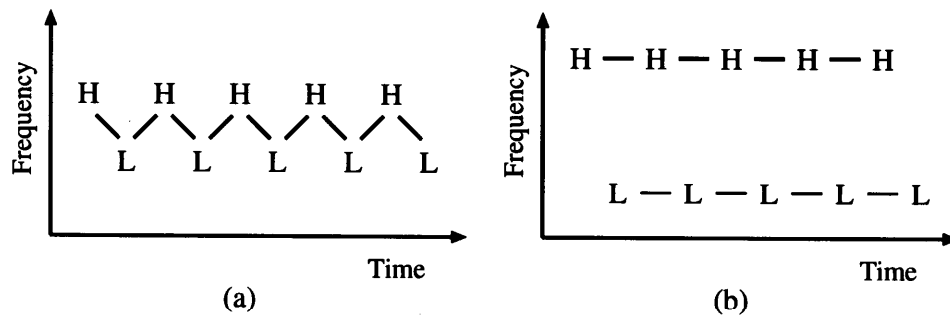


図 1.2: 音脈分凝

的に関連し合っているような系列を音脈 (auditory stream) と名付け、その分裂を音脈分凝 (auditory stream segregation) と呼んだ [3].

カクテルパーティー効果

カクテルパーティーなどの雑踏の中で、大勢の人の声などの様々な雑音が混じっているにも関わらず、我々は相手の声を聞き分けて話ができる。このような聴覚上の現象は「カクテルパーティー効果」と呼ばれている。カクテルパーティー効果は対象に対して注意を向けることで、雑踏の中から対象の音を分離していると考えられているが、カクテルパーティー効果において用いられる音源分離のメカニズムに関しては、左右の耳への音の到達時間差や両耳間の音の強度差から計算される音源定位による空間知覚が重要であるという部分的な説明が与えられているに過ぎず、昨日の全体的な理解をもたらすような知見は未だ得られていない。

しかし、モノラルのスピーカーからラジオやCDを聞いて人の声や楽器音などを分離できるように、音像定位の情報がない音に対してもカクテルパーティー効果は生じる。このようなことが起こる要因としては、音源のピッチの違いや音色の違いが物理的要因として考えられる。また音声の場合には言葉のつながり、言語的知識、経験といったような認知的手がかりも関係してくると考えられる。

このようなカクテルパーティー効果も、特定の音を聞き取るということにおいて聴覚の情景分析機能が用いられている現象の一つと考えられるであろう。

1.2.2 音源分離の物理的要因

ここでは音源分離を行う手がかりとして用いられる代表的な物理的要因について述べる。

まず、音源分離の能力の一つとして「音源定位」が上げられる。その能力によって、求める物体あるいは回避する物体の方向を決定し、また視覚的注意を向けるべき適切な方向を知ることができる。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、両耳に到達する信号の比較により得られ、方向定位をするための重要な手がかりとなる。これは、空間内に複数の音響事象が存在するときに効果を発揮する。

また、我々は方向情報のないモノラルな音を聞いたときでも音源を分離することができる。モノラルのスピーカーから聞こえるラジオなどの音源の空間的位置が一つの場合でも、そこから聞こえてくる人の声、楽器音を判別できるのがその例である。そのような音の分離に影響を与え得る物理的要因に、周波数成分の高調波関係のずれ、周波数成分の立ち上がり時刻のずれ、周波数成分の立ち下がり時刻のずれ、周波数成分の立ち下がりのずれなどが報告され、中でも周波数成分の高調波関係の有無と立ち上がり時刻の同時性が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 [4]。多くの楽器音は、近似的には基本周波数成分とその整数倍の周波数成分から成り立っており、擦弦楽器、打弦楽器、管楽器においては、周波数成分の立ち上がり時刻のずれは数十 ms 程度ずれていることが報告されている [5]。このほかに、2 音が同時に立ち上がる場合にも、音の立ち上がり時間の違いや、音の周波数変化、振幅変化も音源分離に大きな影響を与えていると報告されている [6,7]。

1.2.3 音の立ち上がりの重要性

前節までに、我々が音源分離を行っている際に様々な物理的手がかりを用いていることを述べた。本研究では、その中でも音の立ち上がりが音源分離において最も重要な役割を果たしていると考え、音源分離以外においても音の立ち上がりが音の認識に大きな影響を与えていることから、立ち上がり時刻に着目して研究を進めた。

音声、楽器音、車などの騒音から風の音、木々のざわめきに到るまで、我々の日常生活には実に様々な音が充満しているが、我々はこのような音を聞いて、その音がなんであるかを留意に聞き分けることができる。我々が音を聞き分ける場合、単に音の大きさや高さのみを判断の基準としているわけではなく、むしろその音の「音色」に着目している場合が多い。音色とは American Standards Association によって、『同じ大きさと同じ高さを持った2音が提示され、聴取者がこの2音が異なると判断するときのその相違に対応する聴覚の属性』と定義されている。音色を規定する物理的要因の一つとして立ち上がり、立ち下がり時の過渡的な時間変化が重要であることは古くからいわれている。

例えば、楽器音の認識は立ち上がりの過渡部分、および包絡の時間構造にかなり大きく

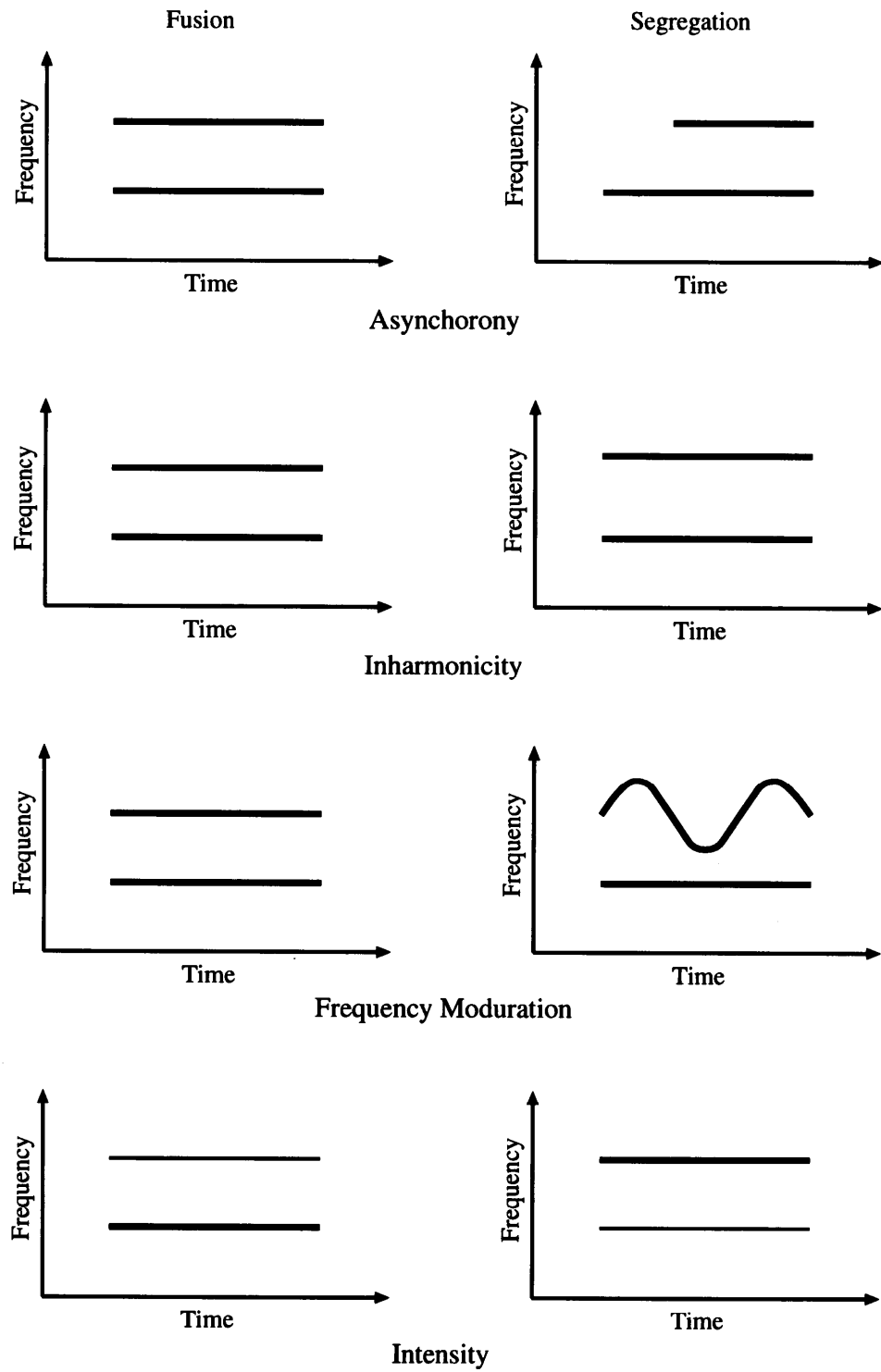


図 1.3: 音源分離の物理的要因

依存しており、楽器音の実音をその場で聞かせたものと、立ち上がり部分を削って聞かせ、それぞれどの楽器音に聞こえるかという識別実験を行ったところ、立ち上がりの過渡部分を削った刺激は識別率が低下するという知見が得られている。また、立ち上がり部分を削った場合には識別率が下がるが、立ち下がり部分を削った場合は識別率はほとんど変化しない [8]。

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、また、立ち上がりのスペクトル構造の過渡的变化が音声の音色を聞き分け、群化し認識する過程で極めて重要である場合が多い。

このように、音の立ち上がりが音の認識と密接に関係しており、音源分離が音の認識と相互に関係しあっていることを考えると、音の立ち上がりが音源分離において重要な要因となっていることが考えられる。そこで、本研究では音の立ち上がりに着目し、立ち上がり時刻のずれが音源分離に与える影響について、聴覚心理学実験を行い、定量的に関連づけていく。

1.2.4 知覚的音源分離と知覚体制化

Bregman と Pinker (1978) は、音源と音の流れとを区別することが有用であると述べている [9]。彼らは音響信号の要素分けを「解剖 parsing」と呼んだ。すなわち、視覚において網膜に映る映像情報が背景と対象とに分解されるのと同様に、音響情報も分解されていくつかの音響エネルギーのまとまりを形成するのである。音の分解には二つの側面が含まれており、一つめの側面は、ある瞬間に同一音源から発せられた周波数成分が全て同じグループに帰属するように群化されるというもので、二つめの側面は、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いでいくというものである。これらの群化の結果として、知覚的な音が構成される。

ゲシュタルト心理学者たちは、知覚体制化を支配する様々な要因について述べている。彼らの説明や彼らが見出した原理は、音響情報の入力を分解するために様々な物理的手がかりが用いられている様子とうまく一致する。つまり、知覚体制化の法則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ことを示している。

以下に知覚的体制化の主要な原理、要因について述べる。これらの原理は一般に視覚、聴覚に共通してみられるもので、ほとんどが最初、視覚において見出されたものである。

近接の要因

他の条件が一定ならば、近い距離にあるものがまとまり、群を形成する。聴覚では周波数近接の要因と、時間軸近接の要因が代表的なものである。周波数近接の要因の例として、

二つの純音成分 B, C からなる複合音と, B に極めて近い周波数を持つ純音 A とが交互に提示される場合, ある適切な条件においては, 純音 A が複合音の成分 B 結びつき, もう一つの成分 C が孤立して, 一つの純音として聞こえるという報告がある [1]. 時間軸近接の要因の例としては, 三三七拍子が挙げられる. 時間軸において, 近接する三つの音と, 三つの音と七つの音が群化することは, 一般的によく知られている.

共通運命の要因

一つの音源で生じた各周波数成分は, 通常同期して変化する. これらの成分は同時になり始め, なり終り, 強さや周波数がともに変化する傾向を持っている. この事実は知覚において利用され, このことによって共通運命の原理が形成される. すなわち, 二つまたはそれ以上の成分が複合音の中で同時に同様の变化をおこすとき, これらは同じ音源の一部として知覚されるのである.

調波性の要因

聴覚系は, スペクトル成分の調波性に対して敏感であり, 調波性を持ったいくつかの成分が同時に提示されると, 一つの音に聞こえるという現象は日常頻繁に見られる. この傾向を「調波性の原理」として, 知覚的融合の要因の一つに数えられている. この原理は同時にならされる成分の融合としては, 多くの場合, 周波数近接の原理よりも強力である. この原理によって知覚的に融合された成分は, しばしば知覚上の独立性を失って融合してしまい, 融合された音が全体として一つの高さを持つ. これは聴覚特有の知覚体制化の要因だが, 臨界帯域ごとに時間周期が検出されるとの考え方もあり, この場合は, 類同の原理, あるいは共通運命の原理により, 共通の時間的周期を示す臨界帯域が結びつけられると解釈される.

類同の要因

この原理は, 類似した要素が一つに群化するというものである. 聴覚における類似性とは, 通常, 音色, 高さ, 大きさ, または定位が近いことを意味する. 前述したように, 音脈分凝の研究においてこの原理の例を見ることができる. 比較的速いテンポで音列が提示された場合, 二つの純音の周波数が近い, すなわち類似しているときには一つの流れが知覚され, 周波数が遠いときには分離した流れが知覚されることがこれにあたる.

図と地の現象および注意

我々は、一般に、聴覚入力全ての側面に意識的な注意を向けることはできず、むしろ特定の部分を選択して意識的な分析をしていると考えることができる。複雑な音は、いくつかの流れに分解され、我々はそのときそのときに、これらの流れのうちの一つに注意を向けていると考えられる。このとき注意を向けた流れは浮かび上がって知覚され、その他の音とは目立たなくなってしまう。ゲシュタルト心理学者たちは、注意が向けられる流れと注意が向けられない流れとの分離を「図と地の現象」と呼んだ。

我々は混雑したカクテルパーティーにおいて、そのときそのときに一つの会話だけに注意を向け、その他の会話がある主の背景として聞いている。もちろん、注意を向けていない流れにもいくぶん意識をしていると考えられるが、完全な意識的な分析を行うのは、一度に一つの流れだけであると考えられる。

第 2 章

音の立ち上がりが音源分離に与える影響

2.1 実験装置

実験では， Visual C++ で作成した刺激音を， Sound Blaster を内蔵した PC-AT 互換機から出力し， 音声出力はオーディオアンプで増幅し， ヘッドフォンを通して提示した． なお， 刺激音のサンプリング周波数は， 44.1kHz とした．

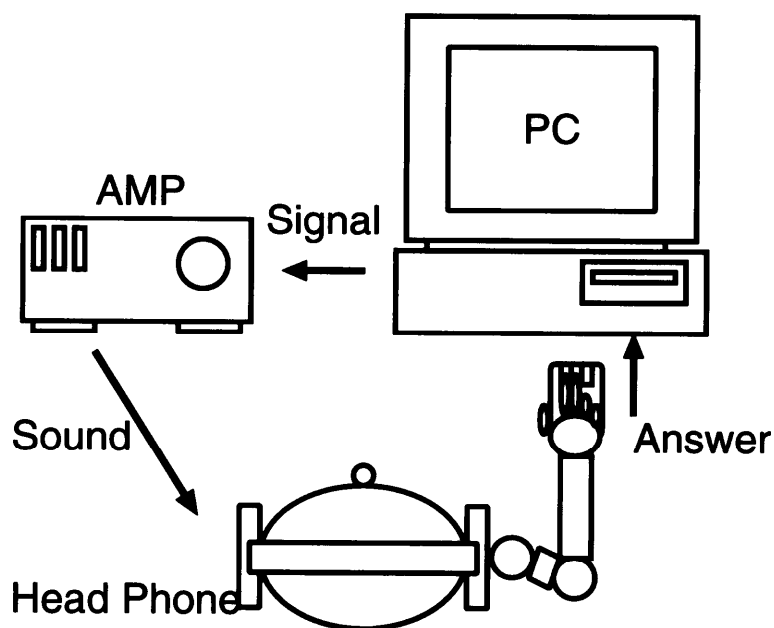


図 2.1: 実験装置の構成

表 2.1: 実験装置

装置	実験装置名称
Computer	DELL 社 OptiPlex GXi5166M
Audio Amplifier	STAX 社 SRM-T1S
Headphone	STAX 社 Lambda Nova Signature

2.2 実験 1

2.2.1 実験の目的

二つの音の周波数比が高調波関係にあり、かつ、同時刻に音が立ち上がる場合、それらは融合して一つの音として知覚されることが知られている [4]。しかし、二つの音が高調波関係にあったとしても、立ち上がり時刻のずれが大きくなると融合せず二つの音として知覚されると考えられる。

実験 1 では高調波関係にある二つの純音を用いて、立ち上がり時刻のずれが音源分離に与える影響を調べた。

2.2.2 川田の研究

川田 (1996) は二つの純音を立ち上がり時刻をずらして提示し、2 音に分離して知覚されるか、1 音に融合して知覚されるかを被験者に答えさせる実験を行い、2 音の協和度が高い場合、倍音成分が先行して立ち上がった方が、基本周波数成分が先行して立ち上がった場合に比べて 2 音に分離して知覚されやすいという報告をしている [2]。

この実験の実験方法は恒常法を用いて実験を行っており、刺激の提示方法は 2 音の立ち上がりの時刻差を 15 段階に設定したものをランダムに提示している。そのために、試行回数が多くなり、被験者の負担が大きくなり、刺激音の比較対象がないために、提示された刺激音が 2 音に分離されたのか、あるいは音の立ち上がりが変わったことで音の音色が変化したのかの判断に迷いが生じる結果となった。

そこで、比較対象となる同時に立ち上がる複合音を、立ち上がりのずれた複合音と連続して聞かせ、どちらがより一つの音に聞こえるかというタスクに変え、被験者の負担を軽減するために、閾値の求め方も恒常法から階段法にすることによって試行回数を減らした。

2.2.3 刺激

試行で提示される刺激には二つの純音からなる複合音を用いた。複合音は基音とその2倍音で構成される。基本周波数は 200 Hz, 400 Hz, 800 Hz の3種類を設定した。各部分音の音圧レベルは 60 dB SPL(sound pressure level) とした。

実験では、立ち上がりのずれた複合音と同時に立ち上がる複合音をランダムな順番に連続して聞かせた。立ち上がりのずれた複合音は先にどちらか一方の部分音を提示し、遅れてもう一方の部分音を提示する。立ち上がり時刻のずれ SOA (Stimulus Onset Asynchrony) は、5 ms 刻みで変化させた。部分音の立ち下がりは同時刻とする。音の立ち上がりと立ち下がりと同じ傾きで、2 dB/ms, 1 dB/ms の2種類を設定し、それぞれ 30 ms, 60 ms で完全に立ち上がる。刺激音の提示時間を 800 ms とし、最初の刺激が提示されてから次の刺激が始まるまでの時間 ISI (Inter Stimulus Interval) は 1000 ms とした。

表 2.2: 刺激の周波数の組合せ

部分音の周波数の組 [Hz]	200 + 400	400 + 800	800 + 1600
立ち上がりの傾き [dB/ms]	2	1	

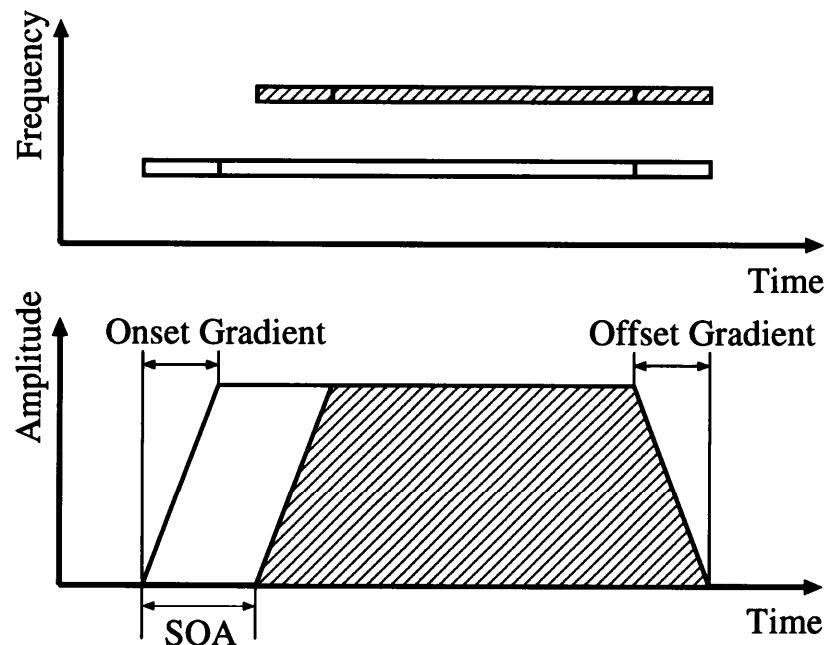


図 2.2: 刺激の概要

2.2.4 実験方法

それぞれの試行において、被験者に立ち上がり時刻のずれた複合音と同時刻に立ち上がる複合音をランダムな順番に連続して与え、より一つの音として聞こえる方を答えさせた。

実験は、予想される閾値よりやや大きい立ち上がり時刻のずれを設定して始める。3回連続して正解すると SOA を 1 段階小さくし、間違えると SOA を 1 段階大きくする。SOA の変化が増加から減少への移り変わったときと SOA の変化が減少から増加への移り変わったときを折り返し点とし、12 回折り返したところで試行が終了する。折り返し点の後半 8 回の平均から閾値を求めた。

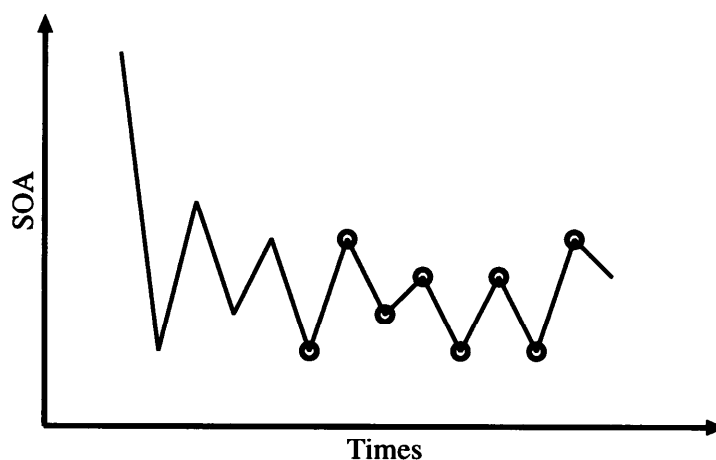


図 2.3: 階段法の閾値の求め方

2.2.5 結果

実験 1 の結果を以下に示す。このグラフは、被験者 4 人の結果を合わせたものである。縦軸は 2 音に分離すると知覚される立ち上がり時刻のずれの閾値を示し、横軸は複合音の基本周波数ごとに、基本周波数成分が先に立ち上がった場合と、倍音成分が先に立ち上がった場合の順番で配置する。図 2.4 は音の立ち上がり時間が 2 dB/ms の場合の結果で、図 2.5 は音の立ち上がり時間が 1 dB/ms の場合の結果を示している。

この結果からは、倍音成分が先に立ち上がった場合と、基本周波数成分が先に立ち上がった場合では、2 音に分離されると知覚される立ち上がり時刻のずれの閾値に、有意な差が見られなかった。

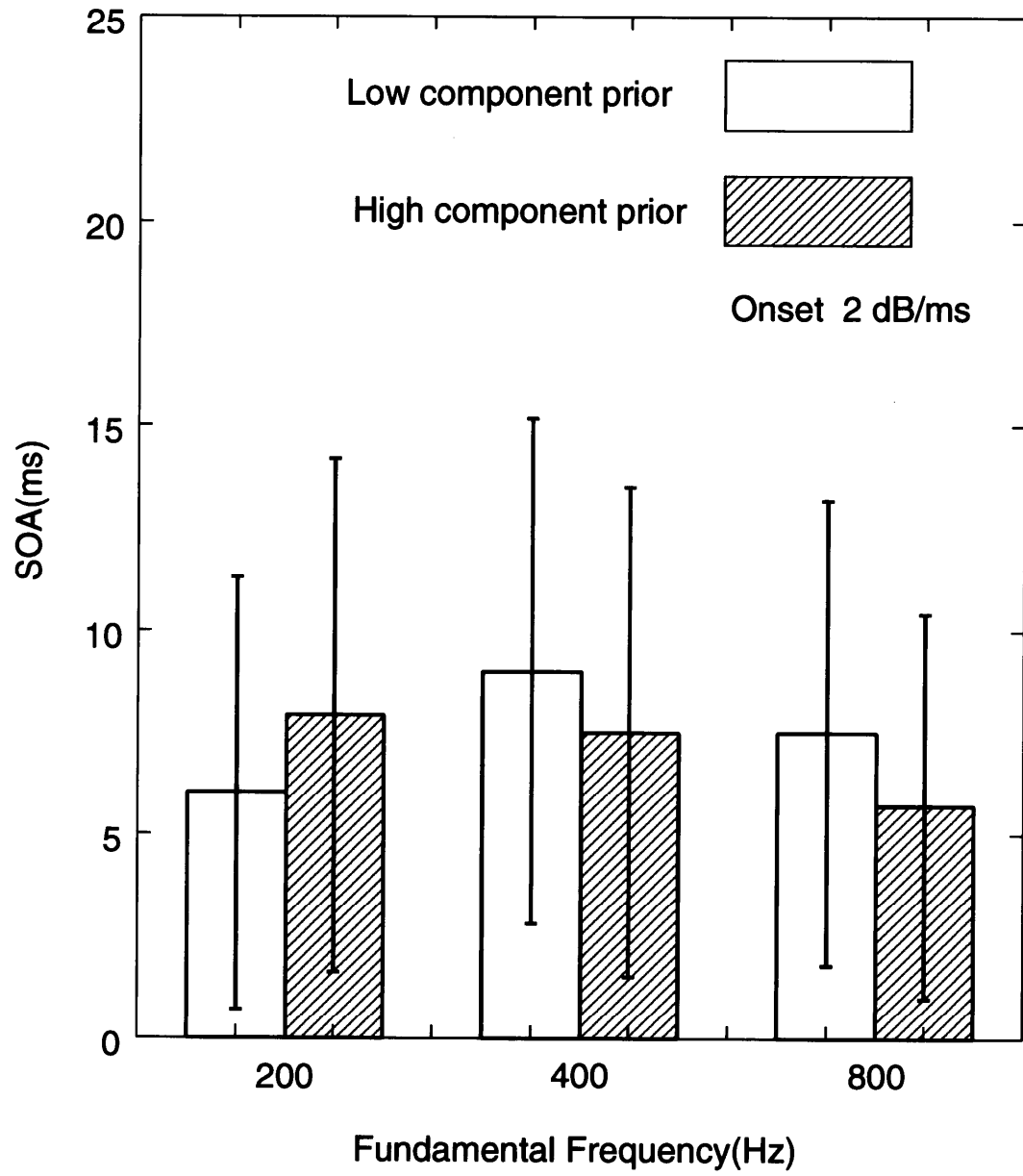


図 2.4: 実験 1 の結果

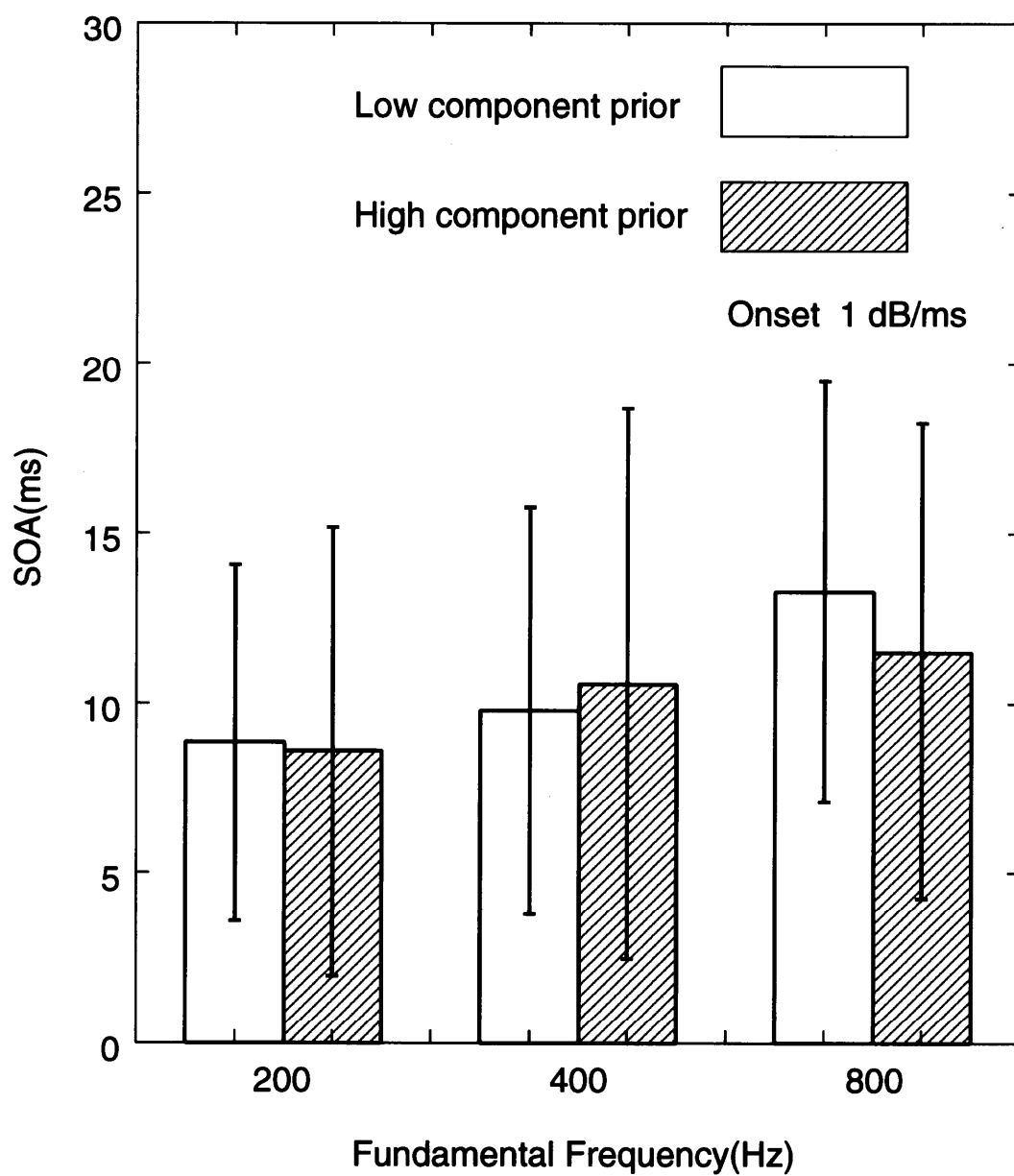


図 2.5: 実験 1 の結果

2.2.6 考察

結果から、実験1では、倍音成分が先に立ち上がった場合と、基本周波数成分が先に立ち上がった場合では、2音に分離されると知覚される立ち上がり時刻のずれの閾値に、有意な差が見られなかった。

これに対する考察として、実験方法に問題があったことが考えられる。川田の実験では、立ち上がりがずれた音のみを提示して一つに聞こえるかどうかを聞いたために、2音に分離したかどうかという判断が難しいという報告があった。そのため、今回の実験では比較対象となる同時に立ち上がる複合音を立ち上がりのずれた複合音と連続して聞かせ、どちらがより一つの音として聞こえるかを被験者に答えさせた。しかしながら、この実験では同時に立ち上がる複合音を探すタスクになってしまう可能性がある。実験開始時の立ち上がり時刻のずれは予想されるずれの閾値よりも大きい値のずれを設定しているため、実験開始時にはどちらが一つの音かが確実に判断できることになってしまい、そのために、実験では、一つに聞こえる同時に立ち上がる複合音を学習し、あとはそれと同じ音を探すだけのタスクになってしまう。立ち上がりがずれていることにより、一つの音として知覚されたとしても同時に立ち上がる音とは異なった音質が知覚されるために、このような結果になったものを考えられる。

このことを考慮し、比較対象をなくして、立ち上がりのずれた複合音のみを提示した場合に、どのような影響が出るかを次の実験2で行うことにする。

2.3 実験 2

2.3.1 実験の目的

実験 1 の結果から，比較対象をおいてどちらが 1 音に聞こえるかを聞くタスクでは，同時に立ち上がる音をさがすタスクになってしまうことが考えられるため，2 音に分離されると知覚される立ち上がり時刻のずれの閾値を求める実験方法としては適切ではなかった。

そこで，この実験では同時に立ち上がる比較対象をなくして，立ち上がり時刻がずれた複合音のみを与えて実験を行う。

2.3.2 刺激

実験 2 でも実験 1 と同様，試行で提示される刺激には二つの純音からなる複合音を用いた。複合音は基音とその 2 倍音で構成される。基本周波数は 200 Hz, 400 Hz, 800 Hz の 3 種類を設定した。基音の音圧レベルを 60dB SPL とした。立ち上がり時刻のずれ (SOA) は 5 ms 刻みで変化させた。部分音の立ち下がりは同時刻とした。音の立ち上がりとしち下がりとは同じ傾きで，2 dB/ms, 1 dB/ms の 2 種類を設定し，それぞれ 30 ms, 60 ms で完全に立ち上がる。音の立ち上がりとしち下がりとは同じ傾き 2 dB/ms で，30 ms で完全に立ち上がる。刺激音の提示時間は 800 ms とした。

2.3.3 実験方法

実験 2 では，立ち上がり時刻のずれた複合音を提示し，融合して一つの音として知覚されるか，分離して二つの音として知覚されるかを答えさせた。

実験は，同時に立ち上がる一つの音として聞こえる状態から開始した。一つの音に聞こえると答えたら SOA を 1 段階小さくし，二つの音に聞こえると SOA を 1 段階大きくする。SOA の変化が増加から減少への移り変わったときと SOA の変化が減少から増加への移り変わったときを折り返し点とし，12 回折り返したところで試行が終了する。折り返し点の後半 8 回の平均から閾値を求めた。

2.3.4 結果

実験 2 の結果を以下に示す。このグラフは，被験者 4 人の結果を合わせたものである。実験 1 の結果と同様に，縦軸は 2 音に分離すると知覚される立ち上がり時刻のずれの閾値を示し，横軸は複合音の基本周波数ごとに，基本周波数成分が先に立ち上がった場合と，倍音成

分が先に立ち上がった場合の順番で配置する。図 2.6 は音の立ち上がり時間が 2 dB/ms の場合の結果で、図 2.7 は音の立ち上がり時間が 1 dB/ms の場合の結果を示している。

この結果から、わずかの差ではあるが、基本周波数成分が先に立ち上がる複合音の方が、倍音成分が先に立ち上がる複合音に比べ 2 音に分離しにくいという傾向が得られた。

この結果から分散分析を行ったところ、基本周波数成分が先に立ち上がる場合と倍音成分が先に立ち上がる場合では、2 音に分離して知覚される立ち上がり時刻のずれに有意な差は得られなかった。これは、今回の実験で用いた被験者数が十分でなかったためであると考えられる。今後、被験者数を増やすことにより統計的に有意であることを証明していきたい。

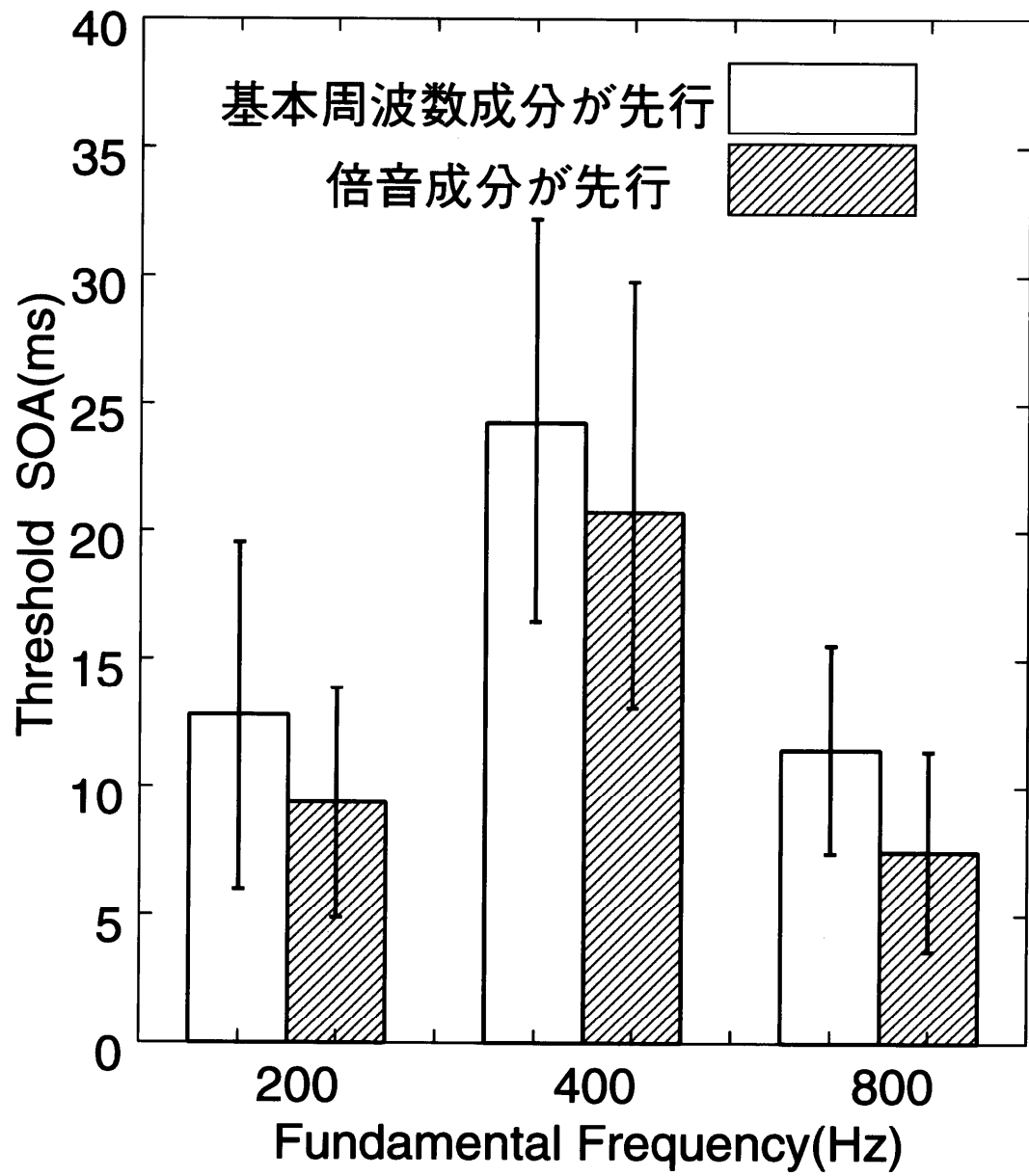


図 2.6: 実験 2 の結果

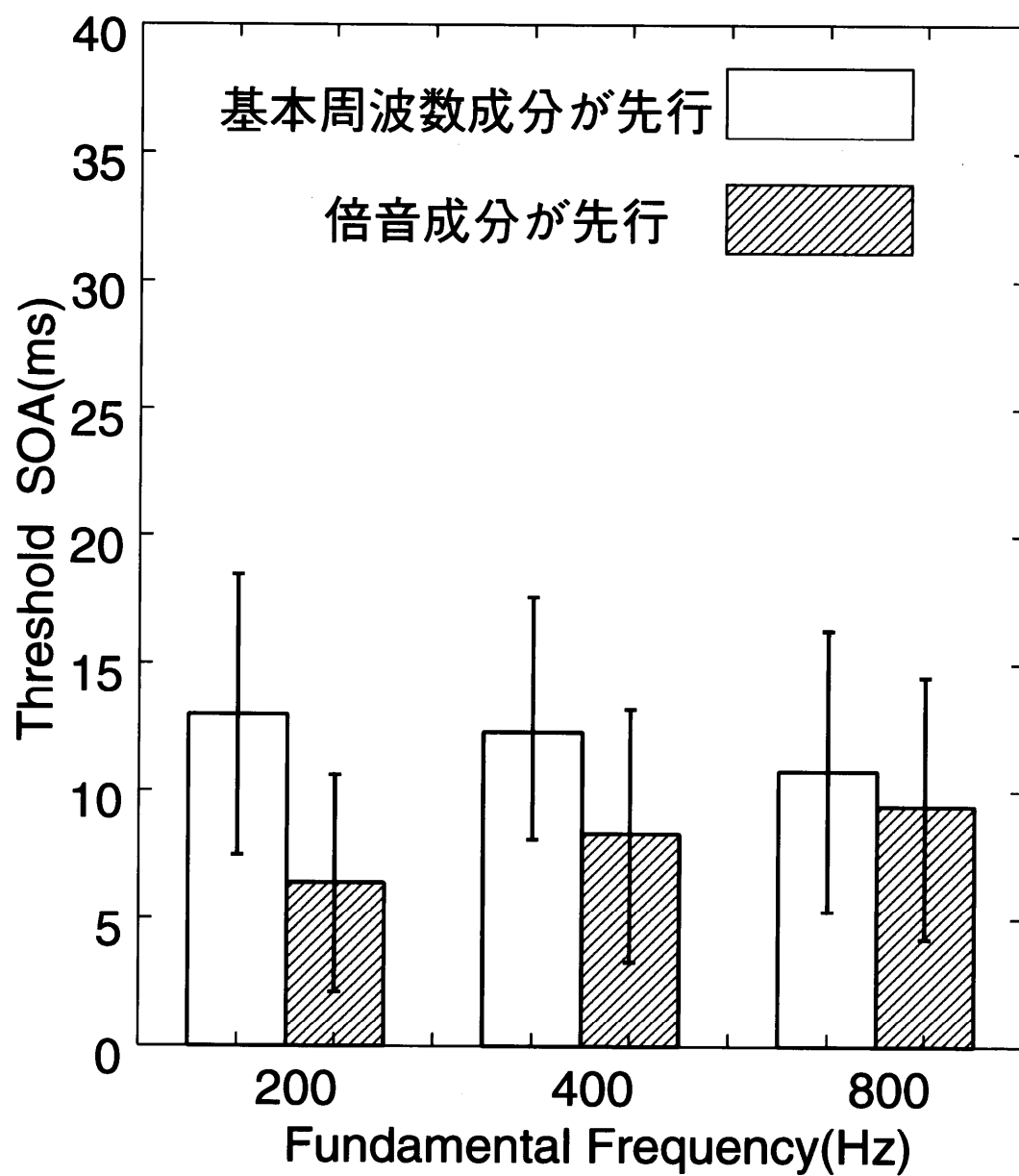


図 2.7: 実験 2 の結果

2.3.5 考察

以上の結果から、基本周波数成分が先に立ち上がる方が、倍音成分が先に立ち上がる音に比べて分離して知覚されにくい傾向が得られた。このことは、倍音成分が先に立ち上がる場合は複合音の群化が弱まるのに対し、基本周波数成分が先に立ち上がる場合は複合音の群化が保たれることを示唆している。これに対する、詳しい考察は第4章で述べる。

この実験では、各純音の音圧レベルを一定にして 60 dB SPL で実験を行ったが、人間の音の大きさの心理的尺度は周波数によって異なる。今回の実験で用いた音の周波数では、倍音成分の方が同じ音圧レベルでも大きい音として知覚されることが知られている [10]。そのため、本実験ではそのような周波数間の感度の違いが結果の非対称性を生み出した可能性が残っている。

そのため、音の大きさの心理的尺度を求めておく必要がある。そのような尺度構成法の一つは、マグニチュード推定法を用いて、音の物理的な強さと音の大きさの判断との関係を求めるものである。測定方法としては、1000 Hz を基準として音のレベルは固定し、大きさを測定する音 (検査音) と連続して聞かせ、被験者は検査音の大きさが 1000 Hz と等しい大きさに聞こえるように検査音の大きさを調整する。これにより、周波数ごとに等しい大きさに聞こえる音の強さが測定でき、周波数ごとに等感度の刺激を用いることができる。今後、このような測定を行い等感度の大きさの刺激を使って、立ち上がりのずれが主要な原因であることを厳密に証明したい。

第 3 章

音の立ち上がりが音列の群化に与える影響

3.1 音列とその群化

実験 1 と 2 では、二つの純音を用いて立ち上がり時刻のずれが音源分離に与える影響について調べた。これらの実験は、「音事象」に対する実験である。

しかしながら、我々は普段情景分析を行う際に、会話や音楽などの音事象が連なっている音に対してどの相手が話した言葉か、またはどの楽器が弾いた音かを判断することができる。これは、同じ性質を持つ音を分離、抽出し、音事象のつながりを一つの音響エネルギーのまとまりとして群化しているためと考えられる。

3.1.1 音の系列の群化

聴覚刺激系列の群化は、時間的要因、ピッチ、ラウドネスなど、系列を構成する音の知覚的属性の影響を強く受ける。それらの間の知覚的差異が小さいとき、その系列は一つのまとまった系列として知覚されるが、知覚的差異が大きい場合には分裂して別々の系列を形成する。

前述したように、高い周波数を持つ純音 H と低い周波数を持つ純音 L が、遅いテンポで HLHLHL と交互に提示された場合、人間は全体を H - L - H - L - というような一つのまとまりとして知覚するが、速いテンポで同様に提示した場合には、H - H - H - という高い音のまとまりと L - L - L - という低い音のまとまりに分離して知覚される。

この現象を用いて、この章では音事象と空白時間との連鎖である「音列」について、立ち上がり時刻のずれが音の系列の群化に与える影響について調べていく。

3.2 実験 3

3.2.1 予備実験

実験 3 を始める前に，HLHLHL と交互に提示した場合，どのくらい遅いテンポで提示したら一つのまとまりとして知覚し，どのくらい速いテンポで提示したら二つのまとまりに分離して知覚されるのかを調べる必要がある．そこで，予備実験として群化の起こる音の提示時間の閾値を求めた．

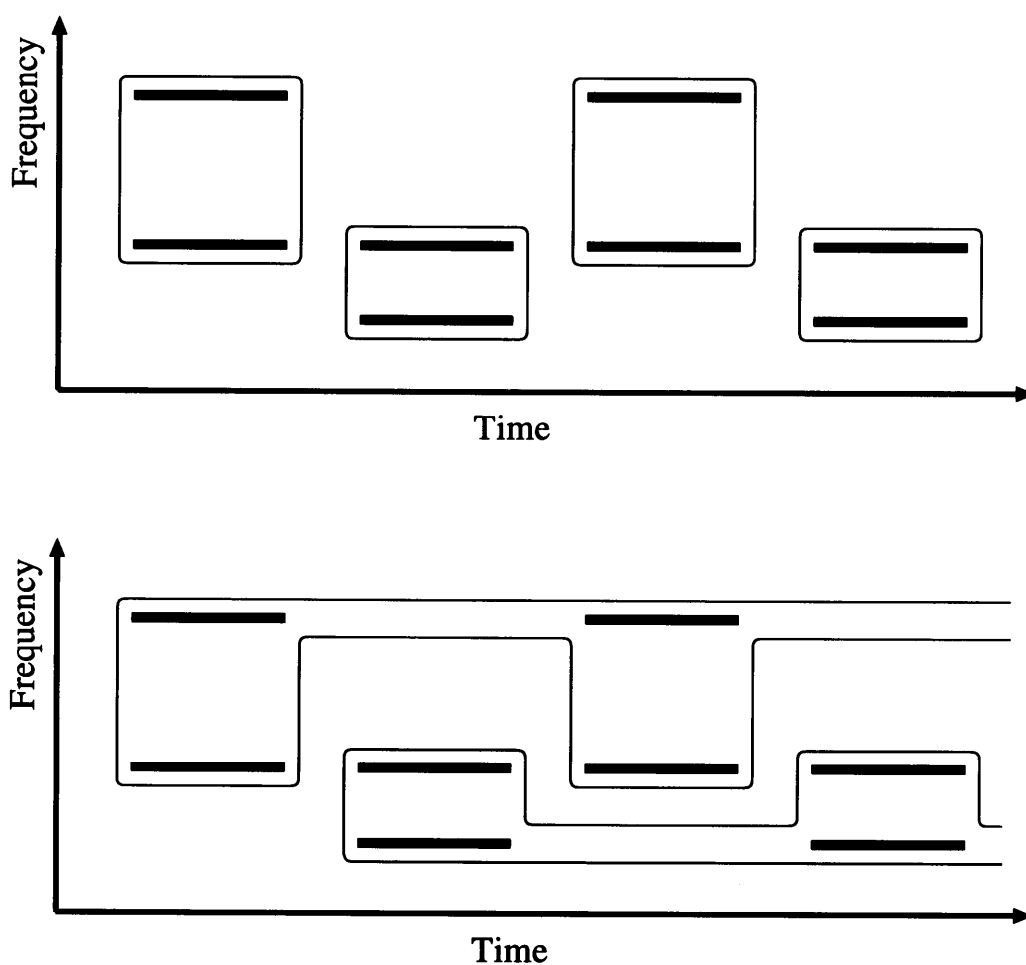


図 3.1: 音列の分離

刺激

交互に提示される刺激には二つの純音からなる複合音を用いた．複合音は基音をその 2 倍音で構成される．低い複合音の基本周波数は 400 Hz，高い複合音の基本周波数は 800 Hz を

設定した。各部分音の音圧レベルは 60 dB SPL。複合音の各部分音は同時に同じ傾きで立ち上がり、同時に同じ傾きで立ち下がる。音の立ち上がりと立ち下がるの傾きは 2 dB/ms に設定した。交互に提示される各複合音の間隔は 50 ms で、音列の総提示時間は 4000 ms とした。

実験方法

十分交互に聞こえるほど長い各複合音の提示時間から実験を開始した。

高い複合音 H と低い複合音 L を交互に提示し、被験者は、交互に聞こえるか、音脈が分離するかを答える。交互に聞こえると答えたら複合音の提示時間を 1 段階短くし、音脈が分離すると答えたところで試行を終了させる。これを 10 回繰り返して、分離したと答えた提示時間の平均を求めた。提示時間の変化は 1 段階で 5 ms とした。

結果

予備実験の結果を以下の表に示す。結果の提示時間の単位は ms である。

表 3.1: 音脈が分離する提示時間の閾値

被験者	結果の最大値	平均値	標準偏差
ms	350	336	15.2
km	370	333	26.0
ky	340	300	21.6
mw	290	221	60.5
ya	280	235	44.3

この結果をもとに、実験 3 で用いる刺激の複合音の提示時間を予備実験の結果の最大値とした。これは、実験 3 では交互に聞こえる音列を提示し、その立ち上がり時刻を少しずつずらして音脈が分離したと聞こえる立ち上がり時刻の閾値を求めることを目的をしているために、ちょうど交互に聞こえる提示時間の値を設定する必要があるためである。

また、予備実験の結果から、被験者により音脈が分離する閾値が異なることがわかる。ここで、予備実験と同じ実験方法で、複合音の提示時間が音脈が分離していると聞こえるのに十分短い状態から、交互に聞こえると答えたまで提示時間を長くした場合の実験を行ったところ、表 3.2 のような結果が得られた。この予備実験の二つの結果から、音脈分離が生じるか生じないかの曖昧な領域が存在すると考えることができる。この曖昧な領域を複合音の立ち上がり時刻をずらすことにより、解剖していくことも興味深い問題だと考えている。

表 3.2: 交互に聞こえる提示時間の閾値

被験者	結果の最大値	平均値	標準偏差
ms	190	100	22.4
km	240	231	7.8
ky	240	186	26.0
mw	270	178	51.0

3.2.2 実験の目的

実験3では、音の立ち上がり時刻のずれが音列の群化に与える影響について調べることが目的とし、交互に聞こえる音列の立ち上がり時刻をずらすことにより、音脈の分離する立ち上がり時刻のずれの閾値を求めた。

この実験は、予備実験のように複合音の提示時間を変化させるのではなく、複合音の提示時間は予備実験で求めた条件で固定し、立ち上がり時刻をずらすことにより複合音を分離させて、新たな音脈の形成を生じさせることを意図して構成している。

3.2.3 刺激

予備実験と同様に、交互に提示される刺激には二つの純音からなる複合音を用いた。複合音は基音とその2倍音で構成される。基本周波数が400 Hzである低い複合音と、基本周波数が800 Hzである高い複合音を設定した。各部分音の音圧レベルは60 dB SPL。複合音の各部分音は同じ傾きで立ち上がり、同時に同じ傾きで立ち下がる。立ち上がり時刻のずれ(SOA)は5 ms刻みで変化させた。複合音の提示時間は予備実験で求めた結果の最大値を用い、交互に提示される各複合音の間隔は50 msで、音列の総提示時間は4000 msとした。

3.2.4 実験方法

実験では、交互に高い複合音と低い複合音の音列を提示し、複合音の立ち上がり時刻をずらして、被験者に交互に聞こえるか音脈が分離して聞こえるかを解答させた。複合音の分離を答えてもらうのではなく、あくまでも音脈が分離して聞こえるかどうかを判断させた。

実験は、同時に立ち上がる交互に聞こえる音列の状態から開始した。被験者が、交互に聞こえると答えたらSOAを1段階大きくし、二つの音に聞こえるとSOAを1段階小さくする。SOAの変化が増加から減少への移り変わったときとSOAの変化が減少から増加へ

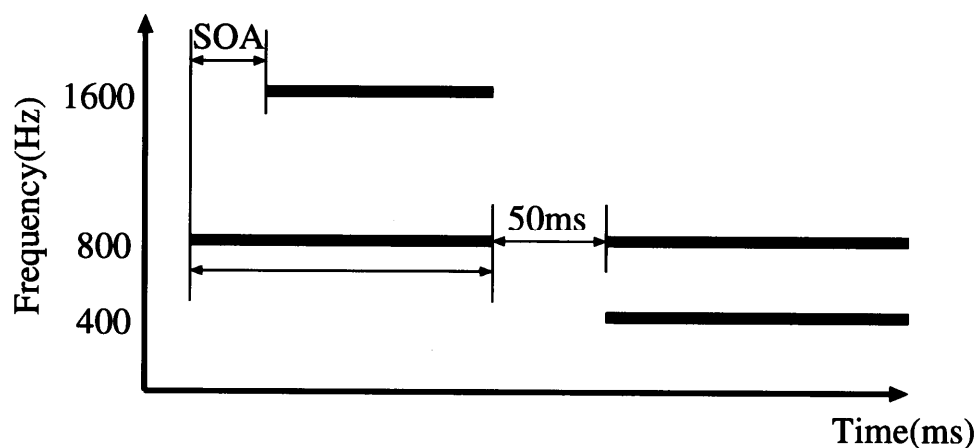


図 3.2: 実験 3 の刺激

の移り変わったときを折り返し点とし、12 回折り返したところで試行が終了する。折り返し点の後半 8 回の平均から閾値を求めた。

立ち上がり時刻をずらす複合音は低い複合音か高い複合音のどちらか一方とする。

3.2.5 結果

実験 3 の代表的な結果を以下に示す。図 3.3 は階段法の実験の経過を示したものである。縦軸は複合音の立ち上がり時刻のずれを表し、横軸は階段法での折り返し点を表している。

この結果から、一度音脈が分離してしまうと、その後の試行にこの知覚が影響を及ぼし、同時に立ち上がったとしても分離して知覚されるという結果が得られた。

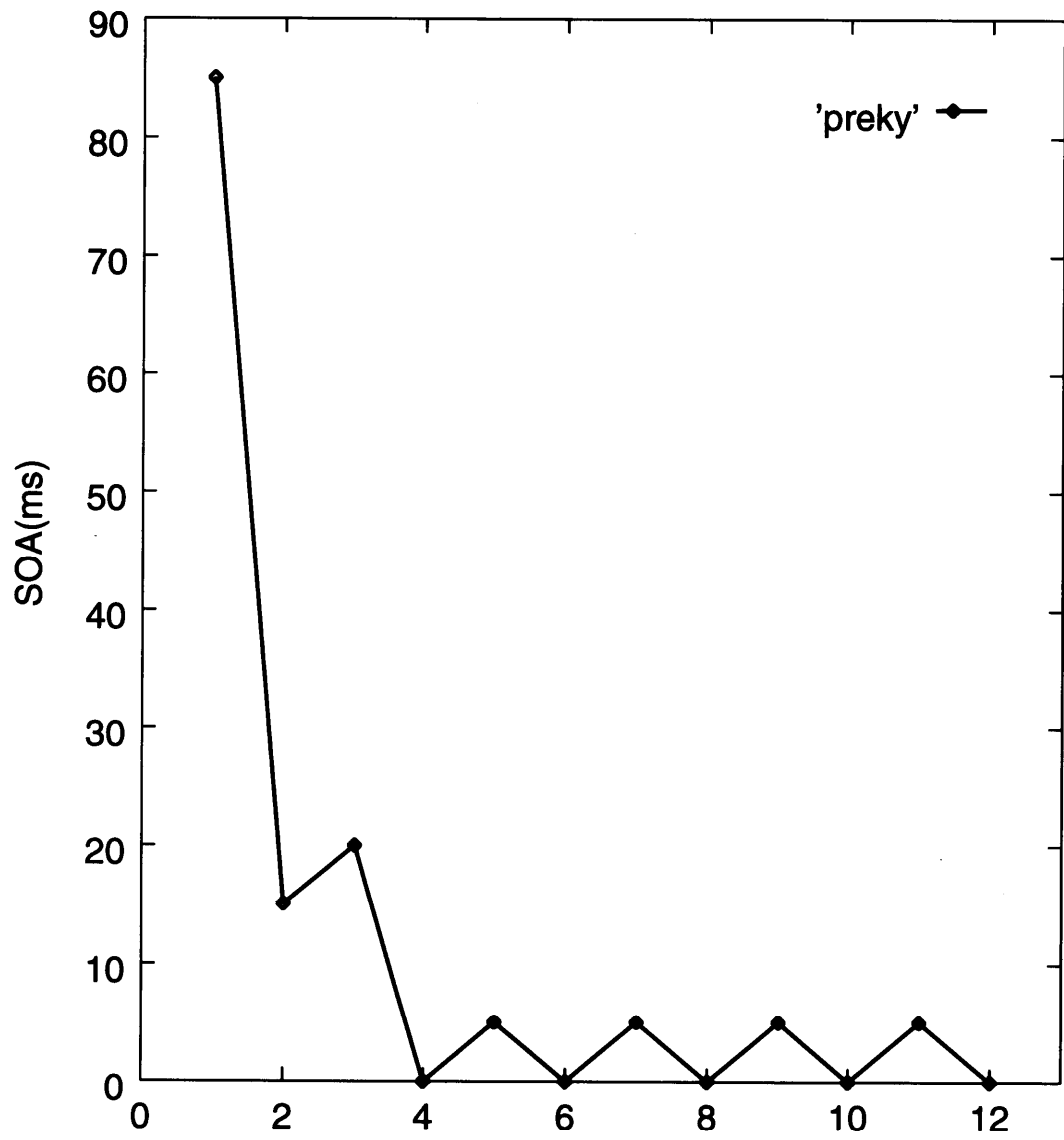


図 3.3: 実験 3 での実験の経過

3.2.6 考察

この実験では、立ち上がり時刻をずらすことにより複合音を分離させ、複合音の周波数成分のそれぞれの要因により、新たな音脈の形成を生じさせることを意図して構成している。そのために、被験者にどのような音脈が知覚されたかを図に書いてもらった。その結果、図 3.4 のような音脈の群化が報告された。

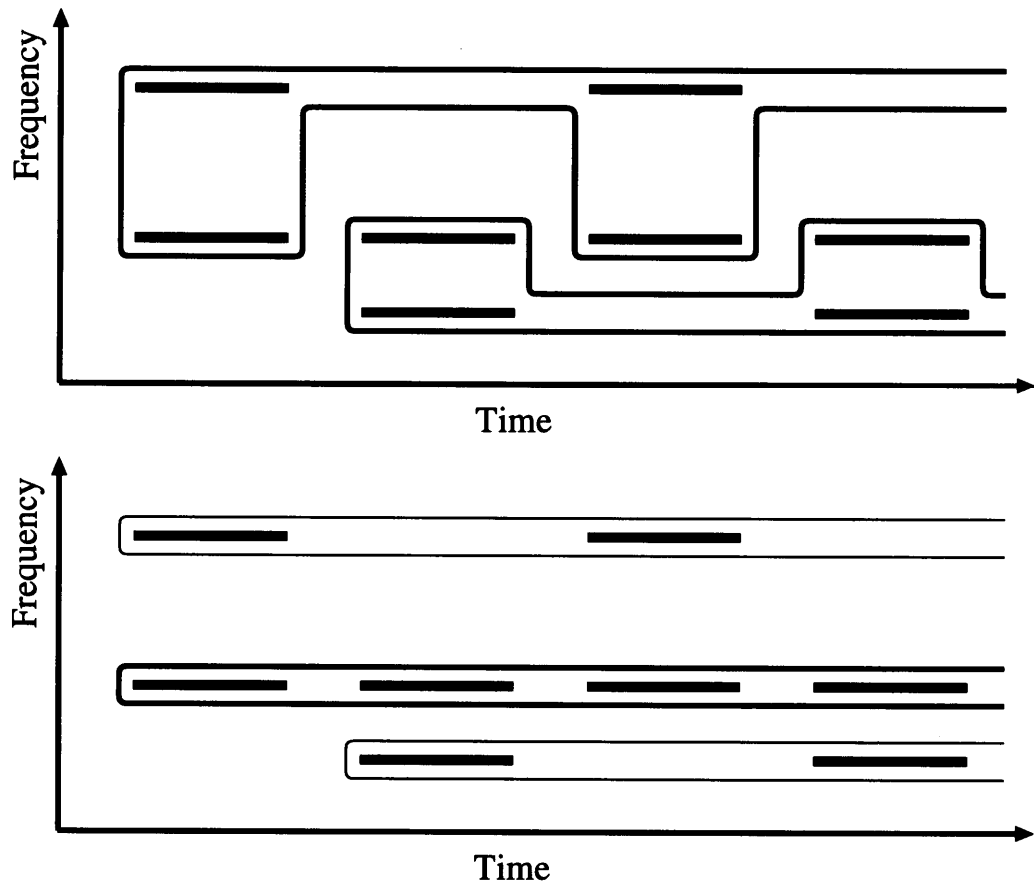


図 3.4: 群化される音脈

従来の音脈分凝の研究における音脈の群化の様子は上図の群化であるが、実験 3 においては下図のような群化の傾向が得られた。中でも 800 Hz の群化が特に強く感じられ、この音脈の群化がその後の試行にも残り、図 3.3 のような結果をもたらしたものと考えられる。

音列の群化と知覚体制化

この群化の現象を、知覚体制化の法則を用いて考えてみると、複合音は、高調波関係にあり、同時に立ち上がり、同時に立ち下がるために、「調波性の要因」と「共通運命の要因」

により一つの音に融合して知覚されると考えられる。そして、調波性の要因は「周波数近接の要因」よりも強いために多くの場合、実験3の提示時間では交互に聞こえるが、複合音の立ち上がり時刻をずらすことによって、共通運命の要因が弱まり、周波数近接の要因により低い複合音の 400 Hz の倍音成分である 800 Hz と、高い複合音の基本周波数成分である 800 Hz が音脈を形成すると考えられる。

また、この実験では、予備実験で求めた複合音の提示時間を用いて実験を行っている。この提示時間は同時に立ち上がる高い周波数を持つ複合音と低い周波数を持つ複合音が交互に提示された時、音脈が分離して聞こえる閾値から求めたものであるために、同じ高さの複合音どうしの間隔は分離して聞こえる提示時間よりも少し長く設定してある。ここで、音脈が分離する要因は音の周波数の差が十分開いていることと、音が提示されるテンポが比較的速いことが挙げられている、前者は「類同の要因」で後者が「時間軸近接の要因」によるものである。つまり、立ち上がり時刻をずらして複合音を分離させた場合、時間軸近接の要因が、高い複合音どうしよりも 800 Hz の周波数成分のほうが強くなるために、他の音脈の群化よりも強く知覚されることが考えられる。

また階段法での実験のため、刺激の変化が少ないので、似た刺激音を聞き続ける試行となる。そのために、聴覚においてこの音脈の影響が消えないうちに、同じような音が入力されるためにこのような結果になったことが考えられる。

3.3 実験 4

3.3.1 実験の目的

実験 3 では、800 Hz の周波数成分が音脈を形成することと、その群化の強度が強いことから、交互に聞こえる音列の立ち上がり時刻をずらすことにより、音脈の分離する立ち上がり時刻のずれの閾値を求めることができなかった。また、実験方法が階段法であったために試行間で音列の群化の印象が残ってしまい、試行ごとの知覚に影響がでてしまった。

そこで、実験 4 では実験方法を恒常法に変えて、実験 3 よりも長い提示時間を用いて実験を行うことにした。これは、複合音の提示時間を長くすることで、共通運命の要因を強め、時間軸近接の要因を弱める効果を意図している。

また、この実験では、立ち上がり時刻のずれを一定にし、提示時間をランダムに変化させて実験を行う。これは、立ち上がり時刻のずれを一定にすることにより複合音の音色の変化をなくし、提示時間をランダムにすることで 800 Hz の周波数成分の音脈の群化の影響を次の試行に残さないようにするためである。

3.3.2 刺激

これまでの実験と同様に、交互に提示される刺激には二つの純音からなる複合音を用いた。複合音は基音をその 2 倍音で構成される。低い複合音の基本周波数は 400 Hz、高い複合音の基本周波数は 800 Hz を設定した。各部分音の音圧レベルは 60 dB SPL。複合音の各部分音は同じ傾きで立ち上がり、同時に同じ傾きで立ち下がる。立ち上がり時刻のずれ (SOA) は 20 ms とした。複合音の提示時間は本実験の直前に立ち上がり時刻のずれた音列を聞かせ、まず提示時間を短くして 800 Hz の周波数成分の音脈の群化を起こさせ、それから徐々に提示時間を長くして、群化が弱まり、交互に聞こえると判断される提示時間を中心に 20 ms 刻みで、7 段階設定した。交互に提示される各複合音の間隔は 50 ms で、音列の総提示時間は 4000 ms とした。

3.3.3 実験方法

被験者は、高い複合音と低い複合音を交互に提示された刺激に対して、交互に聞こえるか音列が分離して聞こえるかを解答した。立ち上がり時刻がずれているので複合音が分離して聞こえることもあり得るが、実験 3 と同様に、複合音の分離ではなく、あくまでも音脈が分離したかどうかで判断するように指示した。

実験は 7 段階に複合音の提示時間を設定した刺激をランダムにそれぞれ 20 回ずつ提示

し、その結果から精神測定関数を導き、交互に聞こえると答えた複合音の提示時間の閾値を求めた。

3.3.4 結果

以下に実験4の結果を示す。図は実験で得られた精神測定関数 psychometric function を示す。

ここで、精神測定関数とは、被験者に多数回刺激を提示して判断を求め、各刺激に対して音列が交互に聞こえると知覚される確率 (P) と、 P を刺激値 S の関数とした時、この $S-P$ 関係を指す。縦軸に得られた P を、横軸に刺激値をとりプロットし曲線で近似すると、どの感覚様相においてもこの精神測定関数は S 字型の増加関数となる。このとき、弁別閾は $P = 0.5$ にあたる S の値、つまり刺激に対する判断が 50 % となる刺激値となる。

図は、縦軸は交互に聞こえると判断した割合を示し、横軸は複合音の提示時間を示している。実線は基本周波数成分が先に立ち上がった場合の関数を、点線は倍音成分が先に立ち上がった場合の関数を示している。これらの曲線が 50 % の値を横切る提示時間を音列の分離の閾値として求めた。

表1は、音列の分離の閾値を示したものである。それぞれ、HL は立ち上がりをずらした複合音、HT は倍音成分が先に立ち上がった場合の分離の閾値、FT は基本周波数成分が先に立ち上がった場合の分離の閾値を示す。

この結果から、基本周波数成分が先に立ち上がった場合の方が、倍音成分が先に立ち上がった場合よりも、短い提示時間で音脈が分離する傾向が見られる。また、基本周波数成分が 800 Hz の複合音の立ち上がり時刻をずらした場合と基本周波数成分が 400 Hz の複合音の立ち上がり時刻をずらした場合では、前者よりも後者の方が基本周波数成分が先に立ち上がった場合の閾値と、倍音成分が先に立ち上がった場合の閾値の差が大きいという傾向が得られている。

表 3.3: 分離の閾値

被験者	HL	HT(ms)	FT(ms)	閾値の差 (ms)
MS	Low	361	345	16
YA	Low	241	228	13
MS	High	370	370	0
MW	High	260	254	6

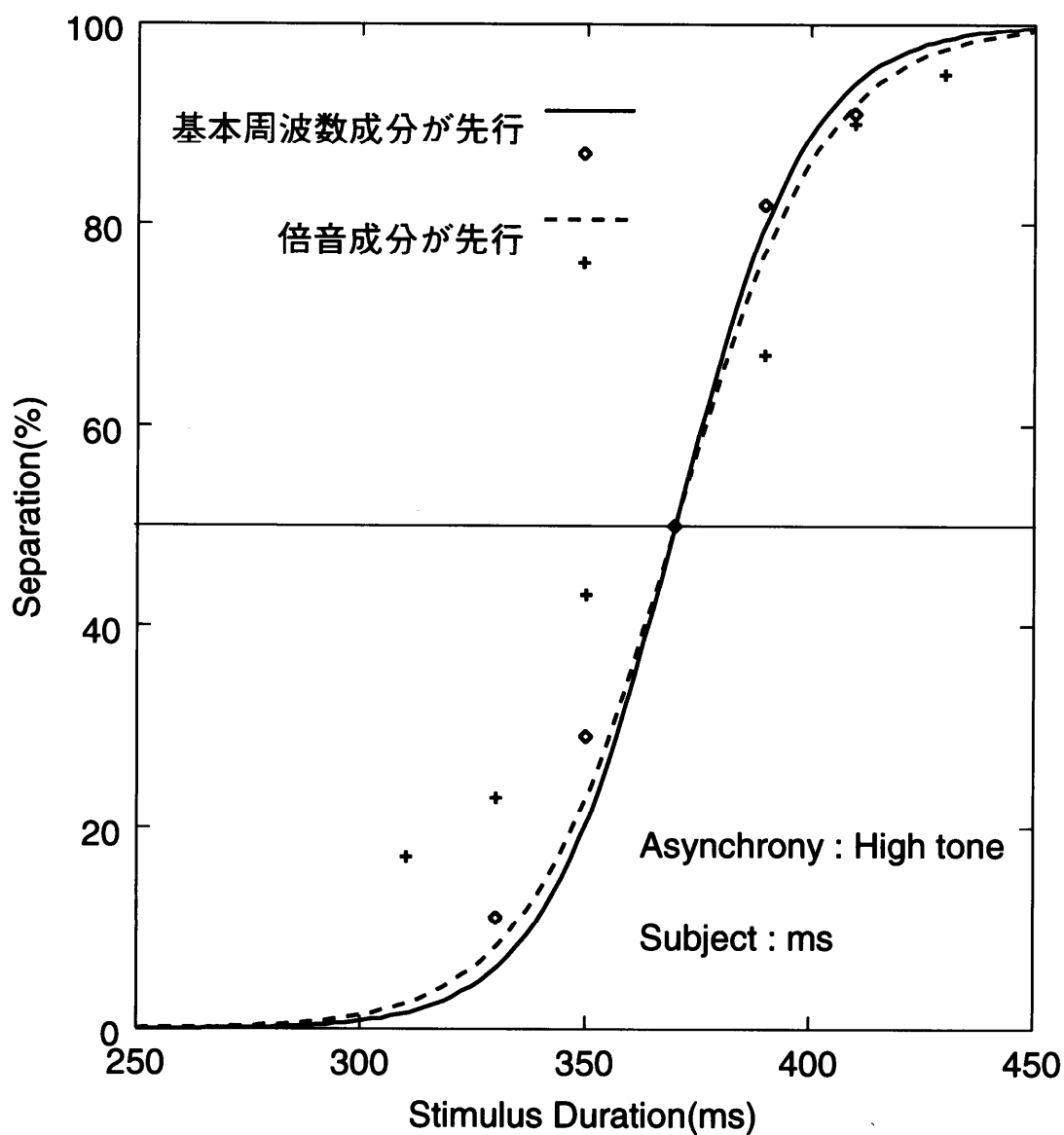


図 3.5: 実験 4 の結果

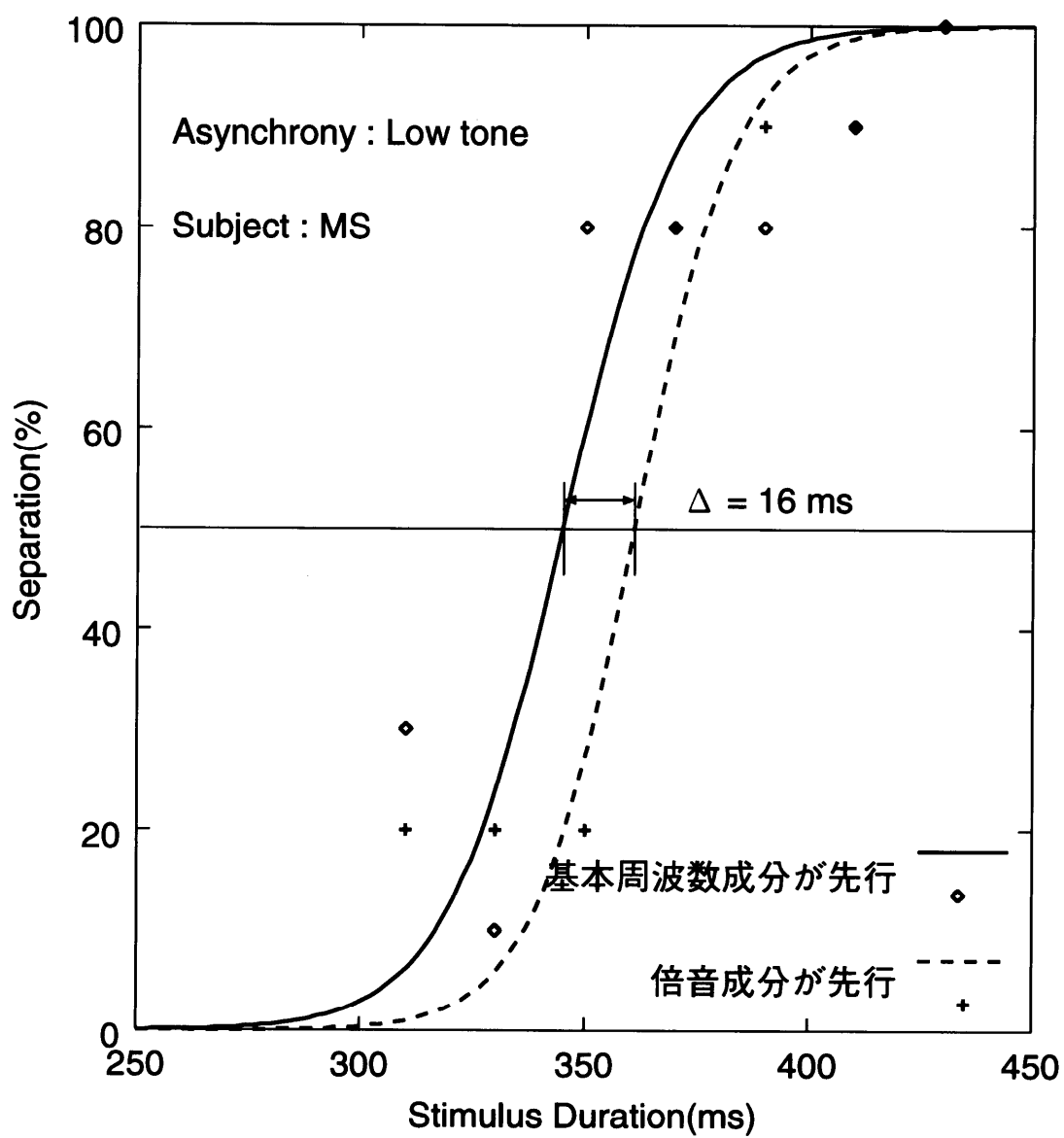


図 3.6: 実験 4 の結果

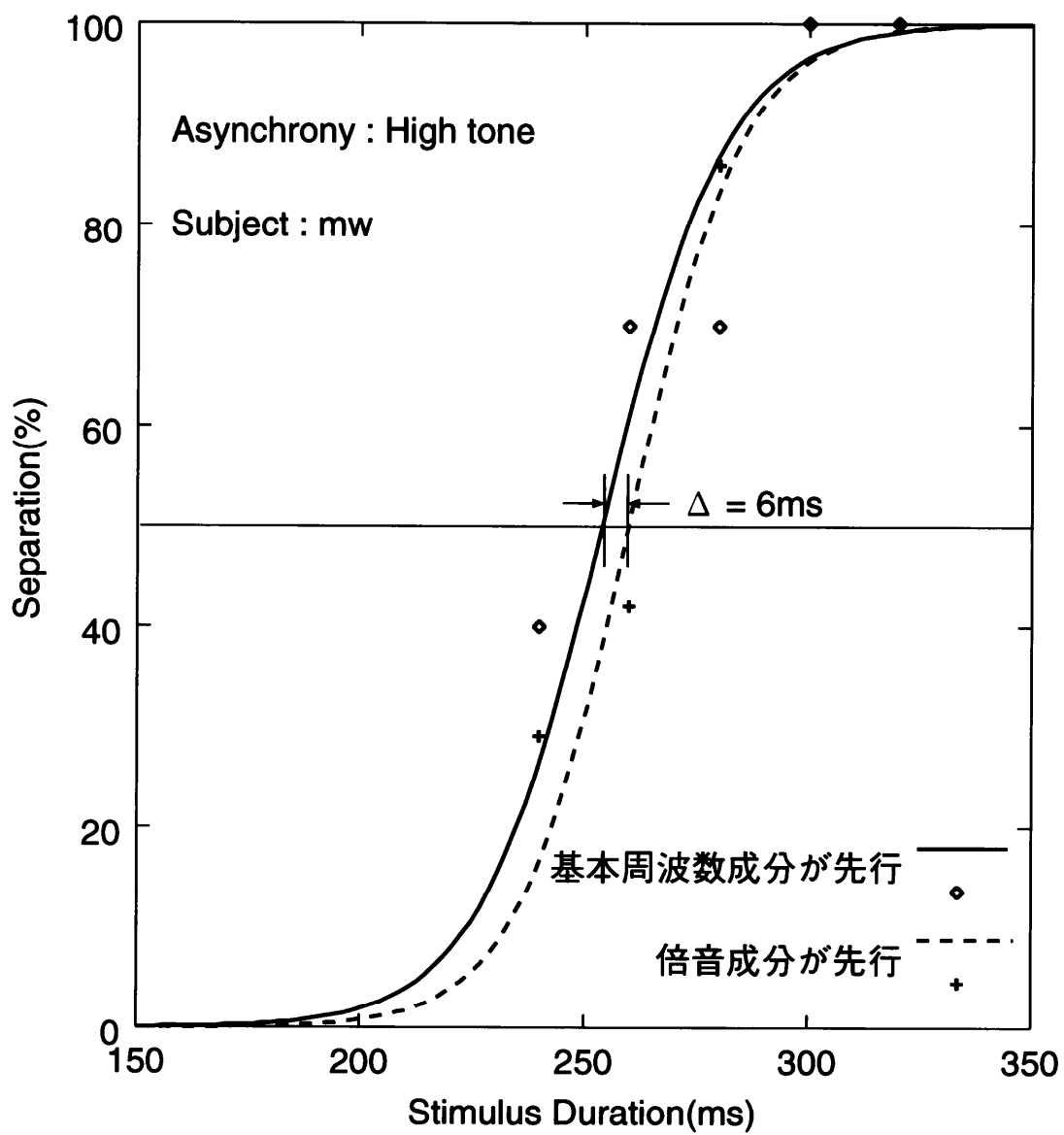


図 3.7: 実験 4 の結果

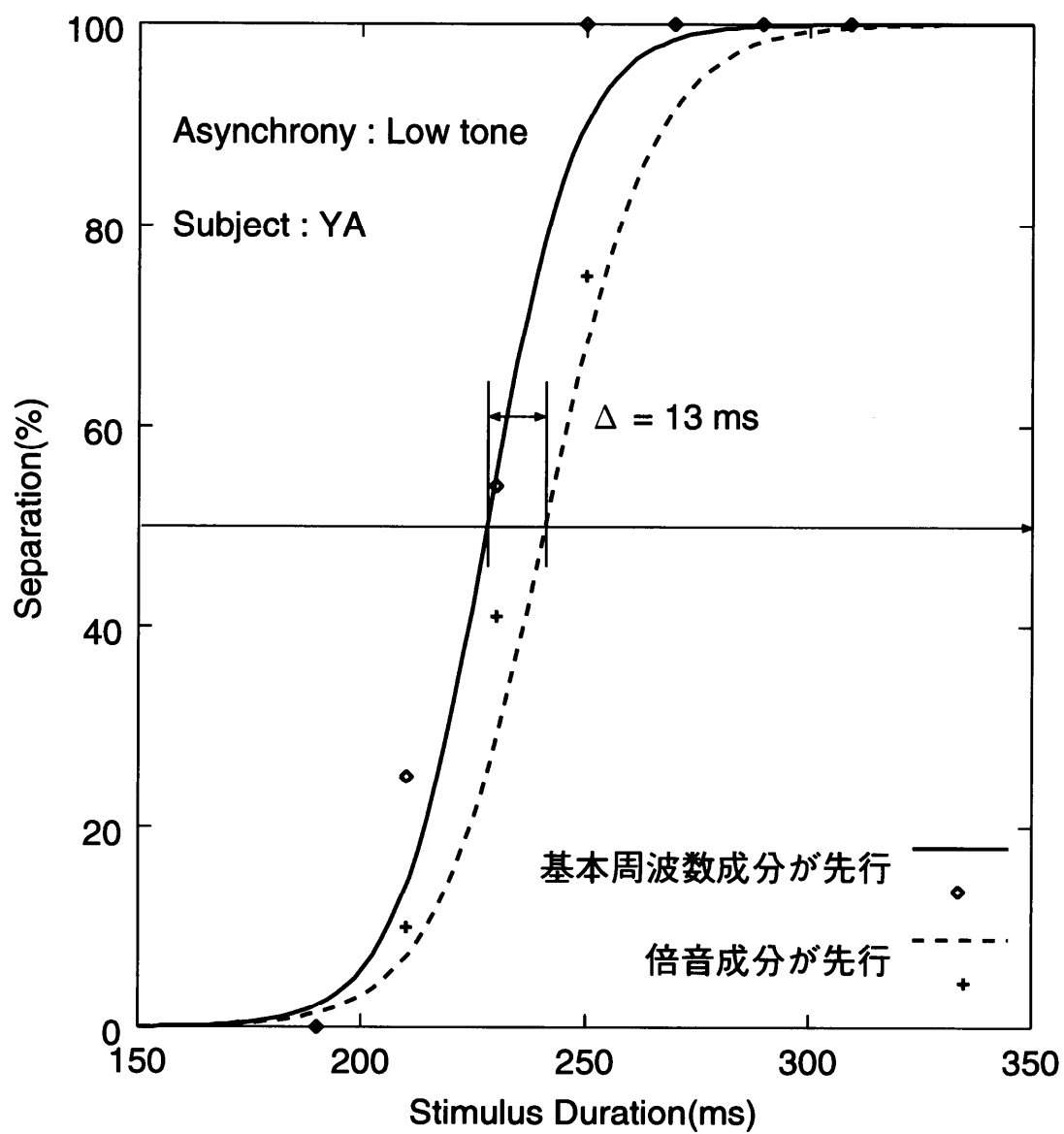


図 3.8: 実験 4 の結果

3.3.5 考察

実験結果から、基本周波数成分が先に立ち上がった場合の方が、倍音成分が先に立ち上がった場合よりも、短い提示時間で音脈が分離する傾向が見られた。それでは、複合音の提示時間が短いということは、知覚体制化の要因にどのような影響を与え、どのようなことが考えられるのかを考察してみる。

まず、SOA を大きくした場合には、共通運命の要因が弱くなり時間軸近接の要因が強まるので、複合音が分離しやすくなるために音列が分離しやすく、SOA を小さくした場合には、共通運命の要因が強くなり時間軸近接の要因が弱まるので、複合音が融合しやすくなるために交互に知覚される。次に、複合音の提示時間を短くした場合には共通運命の要因が弱くなり時間軸近接の要因が強まるので、複合音内部の融合が弱いために音列が分離しやすく、複合音の提示時間を長くした場合には共通運命の要因が強くなり時間軸近接の要因が弱まるので、複合音内部の融合が強いために交互に聞こえやすくなる。このことは、「提示時間が短くても交互に聞こえるということは複合音の融合が強い」ことを示し、「提示時間が長くないと交互に聞こえないということは複合音の融合が弱い」ことを示している。これにより、基本周波数成分が先に立ち上がった場合よりも、倍音成分が先に立ち上がった場合の方が複合音が分離して知覚されやすいと考えられる。

また、基本周波数成分が 800 Hz の複合音の立ち上がり時刻をずらした場合と基本周波数成分が 400 Hz の複合音の立ち上がり時刻をずらした場合では、前者よりも後者の方が長い提示時間で音脈が分離するという傾向が得られた。また、前者の方が交互に聞こえる提示時間の閾値の差が大きいという傾向も見られる。これについては、次のように考えられる。

高い複合音の立ち上がり時刻をずらして分離させた場合、複合音は 800 Hz の基本周波数成分と 1600 Hz 倍音成分にわかれる。そのために、低い複合音が分離しても 800 Hz の音列のみが分離するのに対して、1600 Hz 倍音成分は他の刺激音と比べて周波数差が大きい。これは、周波数の差が大きいほど音列が分離して知覚されやすいことからこのような傾向が得られたと考えられる。

第 4 章

一般的考察

4.1 音の立ち上がりのずれが音源分離に与える影響

本研究で、二つの倍音構造を持つ複合音を用いて、立ち上がり時刻のずれが音源分離に与える影響を調べる実験において、基本周波数成分が先に立ち上がる場合よりも、倍音成分が先に立ち上がる場合の方が分離して知覚されにくいという傾向が得られた。すなわち、基本周波数成分が先に立ち上がる場合と倍音成分が先に立ち上がる場合では、分離したと知覚される立ち上がり時刻のずれの閾値に非対称性が見られた。このことは、音源分離における立ち上がり時刻の効果は、今まで考えられてきたような「ずれの絶対値」で決まるものではなく、周波数構造との関係で決まることを示唆している。ここでは、この点について、複合音の知覚という観点から考察する。

我々が日常生活の中で聞いている音の多くは、基本周波数成分とその整数倍の周波数を持つ倍音成分から成り立っている。1章で述べたように、そのような、同時に立ち上がり、周波数成分が高調波関係にある複合音は一つの音として知覚されるが、このとき知覚される音の高さは基本周波数成分の音の高さである。また、各倍音成分はその複合音の基本周波数の整数倍の周波数を持っているが、この周波数構造が維持される限り、倍音成分が増えようとも複合音の全体の波形としては大きく変化せず、知覚的にも大きく変わることはない。このことは、基本周波数が音の知覚を決定づける重要な要素であることを示している。これは、高調波関係にある複合音のうちの一つの倍音成分の周波数を倍音関係からずらした場合、複合音の中からその倍音成分が分離して知覚されるが、基本周波数成分は他の倍音に比べて周波数を比較的大きくずらした場合でも複合音から分離しないという報告[11]からも基本周波数成分の重要性を見ることができる。

これにより、基本周波数成分が先に立ち上がる場合は倍音成分が遅れて立ち上がっても音の波形にあまり大きな影響を与えず、知覚される音の高さが変わらないために融合しや

すいのに対し、倍音成分が先に立ち上がる場合では、基本周波数成分が遅れて立ち上がると音の波形が大きく変わり、知覚される音の高さが変わるために融合できずに分離してしまうと考えられる..

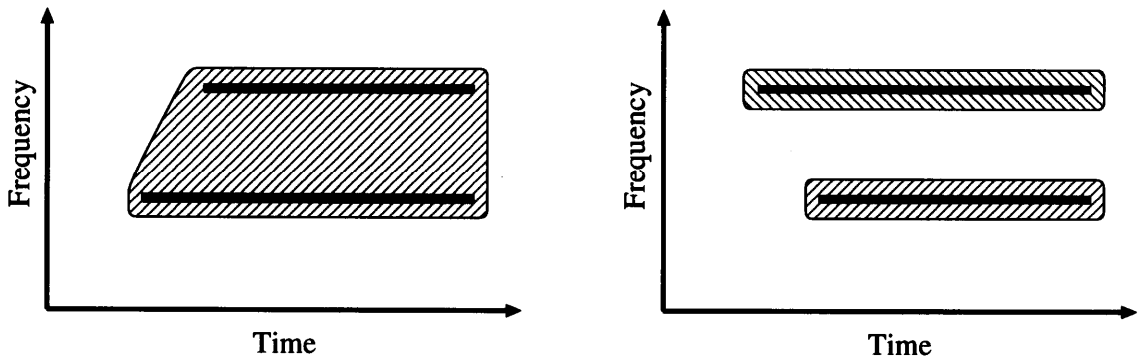


図 4.1: 周波数構造と音源分離

また、我々が日常生活の中で聞いている音の多くは、基本周波数成分とその整数倍の周波数を持つ倍音成分から成り立っている。管楽器、弦楽器などの楽器音は、基本周波数成分などの低い倍音成分から遅れて立ち上がっていることが実験的に調べられており [5]、楽器音に限らず、我々の身の回りで生じる多くの音が、このように低い倍音成分から遅れて立ち上がっていると考えられる。このため、我々が頻繁に耳にしている低い倍音成分が先に立ち上がる音の構造を脳が学習によって獲得し、このような結果を生んだものとも考えられる。

これらの理由により、複合音の周波数構造が高調波関係になっていることが立ち上がりずれの閾値に非対称性を生んだということが考えられる。

4.2 音の立ち上がりが音列の群化に与える影響

知覚的音源分離機能を用いた代表的な現象である音脈分凝の問題を使って、音の立ち上がり時刻のずれが音列の群化に与える影響を調べた実験では、基本周波数成分が先に立ち上がった場合の方が、倍音成分が先に立ち上がった場合よりも、短い提示時間で音脈が分離する傾向が見られた。

知覚体制化の法則の観点から考察してみると、まず、SOA を大きくした場合には、共通運命の要因が弱くなり時間軸近接の要因が強まるので、複合音が分離しやすくなるために音列が分離しやすく、SOA を小さくした場合には、共通運命の要因が強くなり時間軸近接の要因が弱まるので、複合音が融合しやすくなるために交互に知覚される。次に、複合音の提示時間を短くした場合には共通運命の要因が弱くなり時間軸近接の要因が強まる

ので、複合音内部の融合が弱いために音列が分離しやすく、複合音の提示時間を長くした場合には共通運命の要因が強くなり時間軸近接の要因が弱まるので、複合音内部の融合が強いために交互に聞こえやすくなる。このことは、「提示時間が短くても交互に聞こえるということは複合音の融合が強い」ことを示し、「提示時間が長くないと交互に聞こえないということは複合音の融合が弱い」ことを示している。これにより、基本周波数成分が先に立ち上がった場合よりも、倍音成分が先に立ち上がった場合の方が複合音が分離して知覚されやすいという結果が得られている。

基本周波数成分が先行する場合の方が倍音成分が先行するよりも分離して知覚されにくいことに対する考察は、前節で述べたとおりである。

また、音の立ち上がり時刻をずらして音列を分離させるときに、800 Hz の音が音列となって分離して知覚された。これは、一方の複合音を立ち上がり時刻をずらすことにより複合音が分離し、もう一方の複合音の分離を促進することを示唆している。

4.3 実験での問題点

本実験では、実験方法の選択と刺激の選択がうまくいかず、何度も実験をやり直すことになった。ここでは、特に苦勞した3章で行った実験の刺激と実験方法について、より一層の改善を目的とし、検討を行う。

実験3でも述べたように、周波数の差、刺激音の提示するテンポなど、音列の群化を促進させるパラメータは多数存在するが、音列が群化したかどうかの判断基準は被験者の刺激音への注意の仕方、聞き方によって、知覚に大きく差が出てしまう。実際、前日取った分離して聞こえる刺激時間の閾値が次の日の実験では大きく値が変わるために、実験を行うたびに予備実験を行ってパラメータを設定する必要があった。また、音事象の音源分離とは違い様々な要素が関わってくために、どのパラメータにしばって実験をするべきかは考えさせられるところである。音列の群化は被験者にある程度の知識を必要とする上に、聞こえ方が様々なので、どのような判断基準で実験をさせるかを考えて刺激と実験方法の設定をする必要がある。

本実験では、立ち上がり時刻のずれの変化による音列の群化のされ方を統一するために、立ち上がり時刻のずれを一定にして、音脈の群化の強さを制御するために複合音の提示時間を変化させて実験を行った。今回は立ち上がり時刻のずれを 20 ms のみの設定で行ったが、この設定を変えることにより、音列の群化における、立ち上がりのずれと提示時間との関係が表すことができると考えている。

4.4 今後の展開

今回の実験において明らかになった課題として次のようなものが挙げられる。

- 刺激の強度

この実験では、各純音の音圧レベルを一定にして 60 dB SPL で実験を行ったが、人間の音の大きさの心理的尺度は周波数によって異なる。今回の実験で用いた音の周波数では、倍音成分の方が同じ音圧レベルでも大きい音として知覚されることが知られている [10]。そのため、本実験ではそのような周波数間の感度の違いが結果の非対称性を生み出した可能性が残っている。それには、音の大きさの心理的尺度を求めるために、尺度構成法を用いて、周波数ごとに等しい大きさに聞こえる音の強さを測定し、周波数ごとに等感度の刺激を用いる必要がある。今後、このような測定を行い等感度の大きさの刺激を使って、立ち上がりのずれが主要な原因であることを厳密に証明できる。

- 周波数構造

複合音の知覚のされ方から、複合音の周波数構造が高調波関係になっていることが立ち上がりのずれの閾値に非対称性を生んだと考えたが、本研究では、複合音の部分音の周波数比を 1:2 として実験を行った。これは、周波数比を協和関係にすることにより、立ち上がり時刻のずれが少ないときに融合して一つの音として知覚されやすいためである。周波数比を不協和にした場合に、立ち上がり時刻のずれの閾値に非対称性が見られないことを示せば、ここでの主張をより明確にすることができるが、周波数比を不協和にすることにより複合音が一つの音として聞こえなくなってしまうために、今回と同様の実験方法では閾値を求めることはできない。今後、周波数構造が協和関係の場合と不協和関係の場合では、立ち上がり時刻のずれが分離に与える影響がどのように変わるのかを調べる必要がある。

我々が日常耳にする音はより複雑な倍音構造を持っており、音を聞き分ける場合、単に音の大きさや高さのみを判断の基準としているわけではなく、むしろその音の音色に着目している場合が多い。音色を規定する物理的要因の一つとして、立ち上がり、立ち下がり時の過渡的な時間変化が重要であることは古くからいわれている [8]。また、周波数成分ごとの音圧比も音の性質として重要な要素である。周波数構造、各倍音成分の音圧、周波数成分の過渡的な時間変化などの問題を一つ一つ明かにすることによって、人間の知覚的音源分離機能のメカニズムが解明ができるものと考えている。

第 5 章

結論

人間は知覚的音源分離を行う際、物理的要因を手がかりとして用いることにより、音情報から必要な情報を抽出し、周囲で起きている事象を分析している。本研究では、物理的
手がかりのうち、音の立ち上がりに着目して、立ち上がり時刻のずれが音源分離に与える
影響について、音事象の音源分離と音列の群化に関して心理学実験を行った。

実験 1 と 2 では、倍音構造を持つ二つの純音からなる複合音を用いて、立ち上がり時刻
のずれが複合音の音源分離に与える影響を調べた。その結果、基本周波数成分が先行して
立ち上がる場合よりも、倍音成分が先行して立ち上がる場合の方が分離して知覚されやす
いという傾向が得られた。

実験 3 と 4 では、高い周波数を持つ複合音と低い周波数を持つ複合音を交互に提示して、
複合音の各成分の立ち上がり時刻をずらすことにより、音列の群化にどのような影響を与
えるのかを調べた。その結果、複合音の倍音成分が先行して立ち上がる場合よりも、基本
周波数成分が先行して立ち上がる場合の方が複合音の融合を強めるために、音脈が分離し
にくくなるという結果が得られた。

以上の結果から、基本周波数成分が先に立ち上がる場合と倍音成分が先に立ち上がる場
合では、分離したと知覚される立ち上がり時刻のずれの閾値に非対称性が見られることが
明らかになった。この非対称性の原因として、高調波関係にある複合音の知覚は基本周波
数成分の性質に大きく依存していることから、基本周波数成分が先行し倍音成分が遅れて
立ち上がる場合は音の知覚に大きな変化がなく融合されるのに対して、倍音成分が先行し
基本周波数成分が遅れて立ち上がる場合は音の知覚が変わるために分離されることが考え
られる。このことは、音源分離における立ち上がり時刻のずれの効果は、今まで考えられ
てきたようにずれの絶対値で決まるものではなく、音の周波数構造により分離が変化する
ということを示唆している。

謝辞

本論分の作成にあたり，主任指導教官である阪口 豊 助教授に終始丁寧かつ適切な御指導を頂いた。先生からは，聴覚心理学における知識から，研究に対する姿勢，考え方まで幅広い面で得たものが多かった。出澤正徳教授には，研究を進めていく上で指針となるべき有形無形の指導をして頂いた。また，関根道昭助手には知覚体制化や心理学実験における様々な問題において助言して頂いた。その他，私と私の研究を支えてくれたヒューマンインターフェース学講座の皆様に深く感謝する。

参考文献

- [1] Bregman,A.S.(1990)Auditory Scene Analysis,MIT Press
- [2] 川田 晋 (1997)" 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 修士論文.
- [3] Bregman,A.S.and Campbell,J.(1971)"Primary auditory stream segregation and perception of order in rapid sequences of tones,"Journal of Experimental Psychology,89,244-249.
- [4] Handel,S.(1989)Listening,MIT Press
- [5] 山口 公典, 安藤 繁男 (1977)" 短時間スペクトル分析法の自然楽器音への適応," 日本音響学会誌,33,6,291-300.
- [6] McAdams,S.(1989)"Segregation of concurrent sounds. I:Effects of frequency modulation coherence,"J.Acoust.Soc.Am.,89,1364-1377.
- [7] McAdams,S.(1991)"Segregation of concurrent sounds. II:Effects of spectral envelope tarcing,frequency modulation coherence,and frequency modulation width,"J.Acoust.Soc.Am.,89,341-351.
- [8] 難波精一郎編 (1984) 聴覚ハンドブック, ナカニシヤ出版.
- [9] Bregman,A.S.and Pinker,S.(1978)" Auditory streaming and the building of timbre," Can.J.Pshychol.32,19-31.
- [10] Moore,B.C.J.(大串健吾監訳)(1995) 聴覚心理学概論, 誠信書房.
- [11] Moore,B.C.J.,Glasberg,B.R.and Peters,R.W.(1985)"Thresholds for the detection of in-harmonicity in complex tones,"J.Acoust.Soc.Am.77,1861-1867.

- [12] Moore, B.C.J. and Peters, R.W. (1986) "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.* 80, 479-483.