

論文 / 著書情報
Article / Book Information

題目(和文)	聴覚情景分析における補完と分離のメカニズムに関する心理物理学的研究
Title(English)	
著者(和文)	益富恵子
Author(English)	Keiko Masutomi
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第10026号, 授与年月日:2015年12月31日, 学位の種別:課程博士, 審査員:柏野 牧夫,内川 恵二,小池 康晴,小林 隆夫,金子 寛彦,古川 茂人
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第10026号, Conferred date:2015/12/31, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

東京工業大学 大学院総合理工学研究科

学位論文

聴覚情景分析における補完と分離のメカニズムに
関する心理物理学的研究

指導教員 柏野 牧夫 連携教授

副指導教員 内川 恵二 教授

平成二十七年十二月

東京工業大学 大学院総合理工学研究科

物理情報システム専攻 柏野牧夫研究室

益富 恵子

要旨

聴覚は、外界で起きているものごとを生物が認識するための重要な感覚のひとつである。聴覚情報処理系 (聴覚系) が、複数の音が混在する音響信号を解析し、音源の数・位置・状態などの情報を引き出し、外界で何が起きているのかを知る働きのことを「聴覚情景分析」という。聴覚系は情景分析を行うにあたって、入力された音響信号のパターンを解析し、特定の音源や複数の音でできたハーモニーなどの何がしかの意味をもつ「知覚的まとまり」を生み出すと考えられる。このまとまりのことを「聴覚オブジェクト」という。また知覚系は外界の事象のある特定部分のみに選択的に意識を向け、その他の情報を意識の外へ追いやる働きを持つ。この働きを「選択的注意」という。本研究では、心理物理学的手法を用いて、人間の聴覚系がオブジェクトを形成する過程で、欠落した情報を補う「補完」をどのように行うか (研究 1)、また重畳した音響信号をそれぞれのオブジェクトに分ける「分離」に選択的注意がどのように関係するか (研究 2) を調べた。

研究 1 では、聴覚系の「補完」の働きを反映した現象である「連続聴錯覚」に関する実験を行った。連続聴錯覚とは、強い雑音が音響信号の一部に重なるとき、雑音の背後で信号が滑らかに繋がって知覚される現象である。聴覚オブジェクトの形成にあたっては、蝸牛で行われる周波数解析の後、音の大きさや位置、周波数変化などの基礎的な特徴が取り出される「特徴抽出段階」があると考えられている。研究 1 では、周波数が一方向に変化するスイープ音について、雑音が重畳したことで知覚される「錯覚のスイープ音」が「実際のスイープ音」と同様に周波数変化残効を引き起こすか否かを調べた。周波数変化残効とは、周波数が一方向に変化するスイープ音を聴き続けることで、直後に呈示される検査音の周波数変化方向が聴き続けた音の変化方向とは逆に聴こえやすくなる現象である。周波数変化残効には、周波数の変化を検知する特徴抽出段階が関与すると考えられている。実験の結果「錯覚のスイープ音」による残効量は「実際のスイープ音」と比較して有意に小さいことが示された。この結果から、連続聴錯覚によって知覚される「錯覚のスイープ音」は「実際のスイープ音」と同様には周波数変化残効を起こさないこと、また「錯覚のスイープ音」の聴覚オブジェクトは特徴抽出段階よりも後の段階で生成されることが示唆された。

研究2では、聴覚系の「分離」の働きのうち、聴覚系が行っているとされる「繰り返し」に基づく音響パターンの分離処理に聴取者の選択的注意が影響するか否かを調べた。近年の研究によって、人間の聴覚系はこれまでに入力のない全く未知の周波数-時間パターンを持つ音であっても、その音が混合音中で繰り返し出現するならば、ひとつの知覚的まとまりとすることが可能であることが示された (McDermott, Wroblewski, & Oxenham, 2011)。この研究で呈示された音刺激は、過去の研究から聴覚系が音をひとまとまりとするために用いていることが知られている「古典的手がかり」が極力排除された人工音であった。参加者は、ターゲット音と妨害音との混合音が、妨害音を変えながら2回以上繰り返し呈示されるとき、チャンスレベル以上の精度でターゲット音を検出することができた。この結果は「音が繰り返し呈示される」という条件のみで、混合音から音を分離する手がかりが得られることを示唆する。研究2では、McDermottらの発見した「繰り返し手がかり」による音の分離が行われるとき、聴取者が意識的に音を聴く必要があるか否かを調べた。実験では、McDermottらが行った聴覚実験と合わせて、集中が必要な注意課題を参加者に課し、参加者が注意課題に十分集中していたとしてもチャンスレベル以上の成績で聴覚課題に解答できるかを調べた。注意課題は、コンピュータの画面上に連続で呈示される数字列を記憶する課題、画面上を移動する点を目で追いつける課題、早い速度で呈示されるランダム音列中の音数を数える課題の3種類を用意した。実験の結果、注意課題を同時に行っていたとしても、参加者は「繰り返し手がかり」による音の分離が可能であることが示された。この結果は、混合音から未知の音を分離する聴覚系の機能が、聴取者の注意や意識の働きと関係する比較的高次の認知システムの影響を受けにくい、自動性のある情報処理段階に位置していることを示唆する。

以上の研究を総合して、聴覚情景分析の「補完」と「分離」には、聴覚情報処理の段階のうち、基本的な特徴抽出の段階よりは後であるが、注意や意識といった比較的高次の認知処理の段階よりも前である、中間レベルの処理段階が特に関与することを考察した。

目 次

要旨	2
第 1 章 序論	9
1.1 はじめに	9
1.2 聴覚情景分析	9
1.3 聴覚オブジェクト分析	10
1.4 聴覚系が持つ事前知識と音響的手がかり	11
1.5 聴覚の補完・分離機能を反映する現象	14
1.5.1 連続聴錯覚	14
1.5.2 音脈分凝と ABA 音ストーリーミング	16
1.6 聴覚系の構造と機能	20
1.6.1 聴覚末梢系の構造と機能	21
1.6.2 聴覚中枢系の構造と機能	24
1.7 音の特徴抽出	28
1.7.1 音高 (ピッチ) の表現	29
1.7.2 音色 (スペクトル構造) の表現	30
1.8 音響的要素のグルーピング	30
1.8.1 同時グルーピングに関する研究	30
1.8.2 経時グルーピングに関する研究	31
1.9 聴覚オブジェクトと選択的注意	33
1.9.1 聴覚オブジェクト形成への選択的注意の影響	33
1.9.2 注意機能の存在意義	34
1.9.3 注意機能の神経基盤	35

1.10	これまでの研究から分かっていないこと	36
1.11	本研究の目的とアプローチ	38
1.11.1	研究 1: 補完に関する研究	38
1.11.2	研究 2: 分離に関する研究	39
第 2 章	研究 1 : 聴覚情景分析における補完のメカニズムの研究	41
2.1	はじめに	41
2.2	実験 1: 実際のスweep音と錯覚のスweep音の残効量比較	44
2.2.1	刺激	44
2.2.2	課題	46
2.2.3	実験参加者	47
2.2.4	解析	47
2.2.5	結果と考察	47
2.3	実験 2: ノッチ雑音を重畳したスweep音の周波数変化残効量の測定	50
2.3.1	刺激	50
2.3.2	実験参加者	52
2.3.3	課題	53
2.3.4	解析	53
2.3.5	結果と考察	53
2.4	研究 1: 総合考察	56
第 3 章	研究 2 : 聴覚情景分析における分離のメカニズムの研究	59
3.1	はじめに	59
3.2	メイン聴覚課題	61
3.2.1	刺激	61
3.2.2	おとり課題	63
3.3	実験 1: 視覚入力による記憶・集中課題	65
3.3.1	手法	66
3.3.2	実験参加者	68

3.3.3	結果と考察	68
3.4	実験 2: 偽ターゲット音のスペクトル情報による影響の確認	71
3.4.1	方法	71
3.4.2	実験参加者	71
3.4.3	結果と考察	72
3.5	実験 3: 視覚入力による集中課題	75
3.5.1	方法	76
3.5.2	実験参加者	76
3.5.3	結果と考察	79
3.6	実験 4: 聴覚入力による記憶・集中課題	80
3.6.1	方法	81
3.6.2	実験参加者	81
3.6.3	結果と考察	84
3.7	研究 2: 総合考察	85
第 4 章	全体考察	87
	謝辞	91
	文献	92
	付録	107
	本研究に関する論文	107
	本研究に関する国際会議発表	107
	本研究に関する国内学会発表	108

第1章 序論

1.1 はじめに

生物は外界の状況を知ることによって、近づいてくる敵からの逃走や食料となる動物の発見、同種の生物の探索などの基本的な生命活動が可能となる。聴覚は、外界で起きているものごとを生物が認識するための重要な感覚のひとつである。聴覚は、視覚や触覚などの感覚が働きにくい状況であっても、突発的なイベントの発生を即座に検知し、生物に警告する役割を持つ。ヒトの生活では、聴覚は音楽や言語コミュニケーションなどの高度な精神活動にも深く関与する。

聴覚機能の実現に関わる身体器官・脳・神経システムを総称して聴覚系と呼ぶ。聴覚系は耳から取り込んだ音波を、意味のある情報へと変換する。耳に届く音波は、多くの場合、様々な音源から発せられた音が足し合わされた混合信号となっている。聴覚系はこの混合信号を解析し、音源の数や位置などの有用な情報を取り出している。また、聴覚系は数多くの音の中から今現在の行動にとって重要な音を選び取り、選択的に注意を向けることができる。本研究の大目的は、聴覚系がどのような機構を備えることで、またどのような情報処理を行うことで上記のような高度な認識機能を実現しているのかを明らかにすることである。

1.2 聴覚情景分析

視覚情報処理系が周辺の物体や状況の認識を行う過程を「情景分析 (scene analysis)」と呼ぶ (Marr, 1982)。Bregman は、視覚研究の分野で用いられていた「情景分析」の語を聴覚研究に適用し、聴覚情報処理の問題に対する独自のアプローチを展開した。聴覚情報処理系が、複数の音が混在する音響信号を解析し、音源の数・位置・状態などの情

報を引き出すことで外界で何が起きているのかを知る過程を「聴覚情景分析」(auditory scene analysis)と呼ぶ(総説は Bregman, 1990; Pressnitzer, Suied, & Shamma, 2011 を参照). Auditory scene analysis の語は Bregman による著書の題名でもある. 聴覚心理学は, その他の心理学研究の分野と比較して, 伝統的に要素還元主義の性質を強く持つ傾向があった. 音の知覚的な大きさ(ラウドネス)や周波数弁別能力などの音が備える個々の特徴に関しては精密な心理尺度や定量的なモデルの構成が盛んであった. 一方で「音が持つ個々の特徴から, どのようにして私達の知覚世界が作り上げられるのか」という問いに答えようとする研究は数少なかった. Bregman は, 聴覚研究の分野において「知覚のまとまり」を形作るために聴覚系が前提とする規則について, 音の刺激に対する実験参加者の知覚を分析するアプローチをとった. Bregman のアプローチには, 視覚研究におけるゲシュタルト心理学のアプローチとの類似性がみられる. ゲシュタルト心理学とは, 20 世紀初頭にドイツを中心として起こった心理学の一派である (M. Singh et al., 2012). ゲシュタルト心理学派の研究者らは, ヒトの知覚を個別な要素に還元できない「全体的なまとまり」としての観点から捉え, このまとまりが形成されるための普遍的な規則を記述しようと試みた.

1.3 聴覚オブジェクト分析

聴覚情景分析は聴覚オブジェクト分析 (auditory object analysis) とも呼ばれる (総説は Griffiths & Warren, 2004; Bizley & Cohen, 2013; Kubovy & Valkenburg, 2001; Winkler, Denham, & Nelken, 2009 を参照). ここで言う「オブジェクト」とは, ヒトや動物が知覚的に「同一のもの」や「同一の存在」また「同一の状態」として認識する「知覚上のまとまり」を指す. 聴覚オブジェクトの厳密な定義については, 視覚のオブジェクトの定義と同様に古くから議論が続いている (Bizley & Cohen, 2013; Griffiths & Warren, 2004; Kubovy & Valkenburg, 2001). 聴覚オブジェクトを「音源」と同義であるとみなす研究者がいる (Bizley & Cohen, 2013) 一方, 必ずしも聴覚オブジェクトが音源に相当するとは言えないと考える研究者もいる (Griffiths & Warren, 2004). 後者は同一のメロディを 2 つの楽器 (音源) が奏で, メロディが 1 つに融合する場合などを挙げて, 聴覚オブジェ

クトは音源と混同するべきではないと述べている。また、視覚世界のオブジェクトとは異なり、はっきりとした形を持つことのない音に対して「オブジェクト」の語を当てはめることは不適當であると考え、聴覚イベント (auditory event) という語を用いる研究者も存在する (Blauert, 1997)。

本稿では、聴覚オブジェクトの定義については、Griffiths と Warren らが記述した「オブジェクト分析の定義」(Griffiths & Warren, 2004) に則ることとする。Griffiths と Warren はオブジェクトを「知覚的な実体であり、感覚情報を表現し、分析するために役立てられる脳内メカニズムによって決定される」と表現した上で、「オブジェクト分析」の4つの性質を提案した。(1) オブジェクト分析は感覚世界の中の物事に対応する情報の分析に関与する。(2) オブジェクト分析は感覚世界の残りの部分と関係するオブジェクトや情報を分離することに関与する。(3) オブジェクト分析は感覚情報の抽象化に関与する。これによって、オブジェクトに関する情報は、他の感覚領域から得たある感覚経験との間で一般化可能である (例として、ある個人の顔は、見る角度や明るさに関わらず、同じ人物の顔として認識されうる)。(4) オブジェクト分析は、異なる感覚間での概念の一般化に関与する。例として、同じ個人の顔と声を結びつけを行う。ただし、Griffiths と Warren は (1), (2), (3) の要件は必須であるが (4) の条件が真であるか否かは依然明確ではないとしている。

1.4 聴覚系が持つ事前知識と音響的手がかり

感覚系による情景分析は、数学的には「不良設定の逆問題」として捉えられる (Marr, 1982)。良設定問題とは「数学的に解が存在すること」「解が一意であること」「連続的なパラメタ変化の際、解も連続的に変化すること」の三条件を満たす問題である。感覚系による情景分析では、解としてのオブジェクトが真に存在することも、オブジェクトが常に同じものであることも、刺激の変化に対応してオブジェクトが変化することも、いずれも保証されていない。このため、聴覚系が入力信号のみから妥当な解を導くためには、何らかの制約条件が必要となる (総説は Pressnitzer et al., 2011; McDermott, 2009 を参照)。「知覚系には周辺世界に関する事前知識が埋め込まれている」という考えは、視

覚の研究において広く浸透している (総説は Barlow, 1997; Gregory, 1997 を参照). 聴覚研究の分野においても, 周囲の環境や実際の音源の状態を反映した「妥当な知覚」を得るため, 聴覚系がどのような事前知識を利用しているかを調べる研究が行われている.

聴覚系は, 音そのものが一般的に持つ物理的な特性についての知識を有しており, この知識を利用して音の分離に役立てていることが知られている. (総説は Bregman, 1990; Darwin, Hukin, & Al-Khatib, 1995; Carlyon, 2004 を参照). 聴覚系が利用する, 音の物理的特性の手がかりを「ボトムアップ手がかり」と呼ぶ. 主なボトムアップ手がかりとして「周波数の調波構造」「立ち上がりの同期性」「振幅変動の共通性」「空間的位置」が知られている. (1) 周波数の調波構造の手がかり (図 1.1A): 環境中に存在する多くの音は周期性を持つため, そのスペクトル構造は基本周波数の整数倍の周波数成分を持つ調波構造となることが多い. 聴覚系はこの知識を利用し, 同じ基本周波数を持つ周波数成分同士をひとつの知覚的まとまりとして認識する傾向がある (Darwin, 2008). (2) 立ち上がりの同期性 (図 1.1B): 同じ音源から発生した音の周波数成分は, 同時に出現することが多い. 聴覚系はこの知識を利用し, 同時に出現する周波数成分同士をひとつのまとまりとして知覚する傾向がある (Hukin & Darwin, 1995). (3) 時間的変動の共通性 (図 1.1C): 同じ音源から発生した音の周波数成分は, 時間的変動が共通していることが多い. 聴覚系はこの知識を利用し, 変動が共通する成分同士をひとつの知覚的まとまりとして認識する傾向がある (Roberts & Bailey, 1996). (4) 空間的位置: 同じ音源から発生した音の周波数成分は, 同じ方向から耳に届くことが多い. 聴覚系は音源の位置情報を使うことで周波数軸上で離れた成分同士をひとつの知覚的まとまりとして認識する傾向がある. ただし空間的位置による手がかりは, 前述した 3 つの手がかりと比較して知覚的まとまりが作られにくいことが知られている (Darwin, 1997; Culling & Summerfield, 1995; Darwin & Hukin, 1997).

入力音が音声や音楽などのある特定の分類に属する場合, 聴覚系は, ボトムアップ手がかりの他にも, 言語的規則や音楽的規則を利用して認識を行うことが知られている (R. M. Warren, Obusek, & Ackroff, 1972). こうした規則の手がかりを「トップダウン手がかり」と呼ぶ (Carlyon, 2004; Bregman, 1990).

聴覚系がボトムアップ手がかりやトップダウン手がかりなどの事前知識をどのように

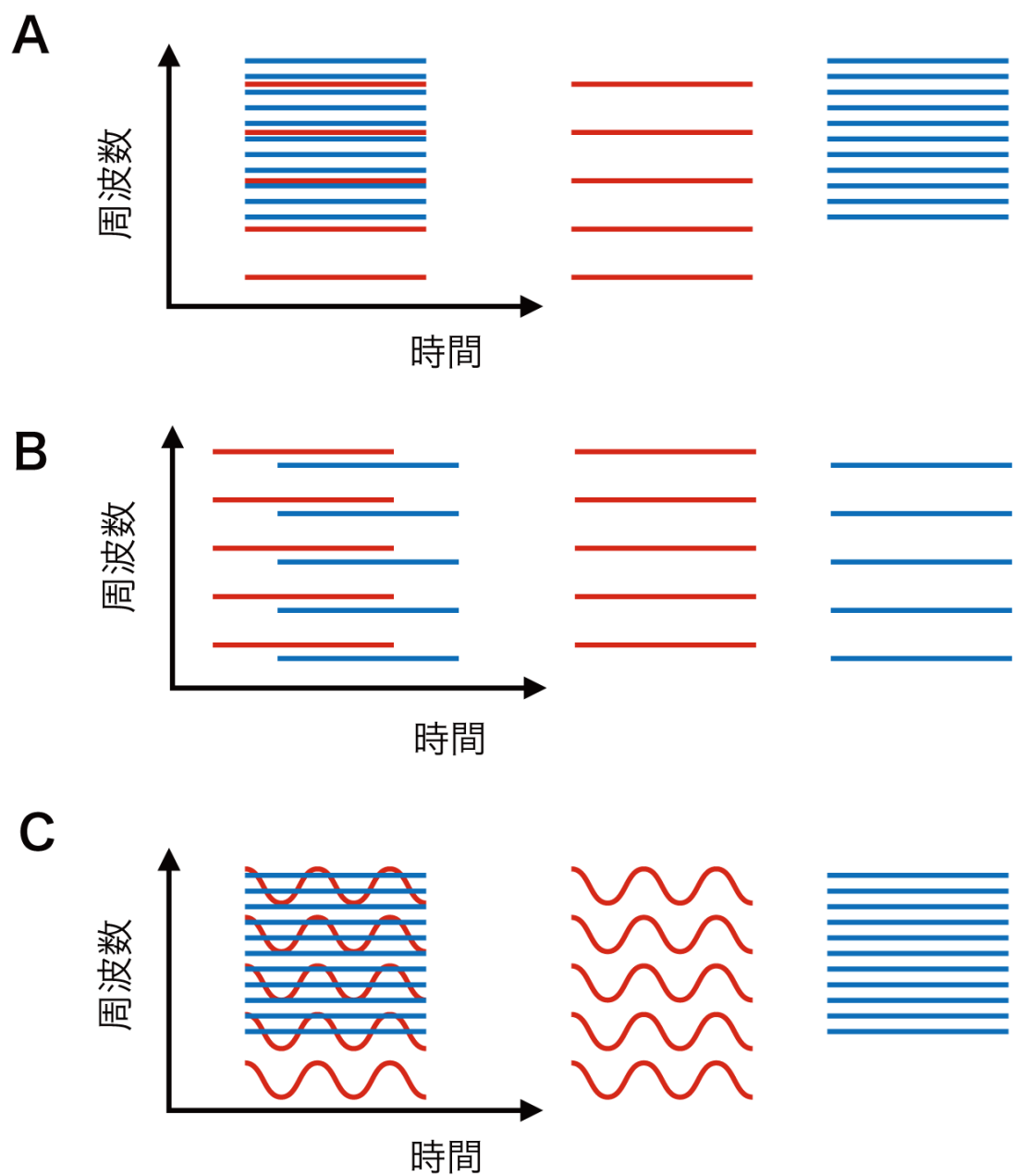


図 1.1: 聴覚系が音の分離に用いる「ボトムアップ手がかり」. A: 調波構造. B: 立ち上がりの同期性. C: 時間的変動.

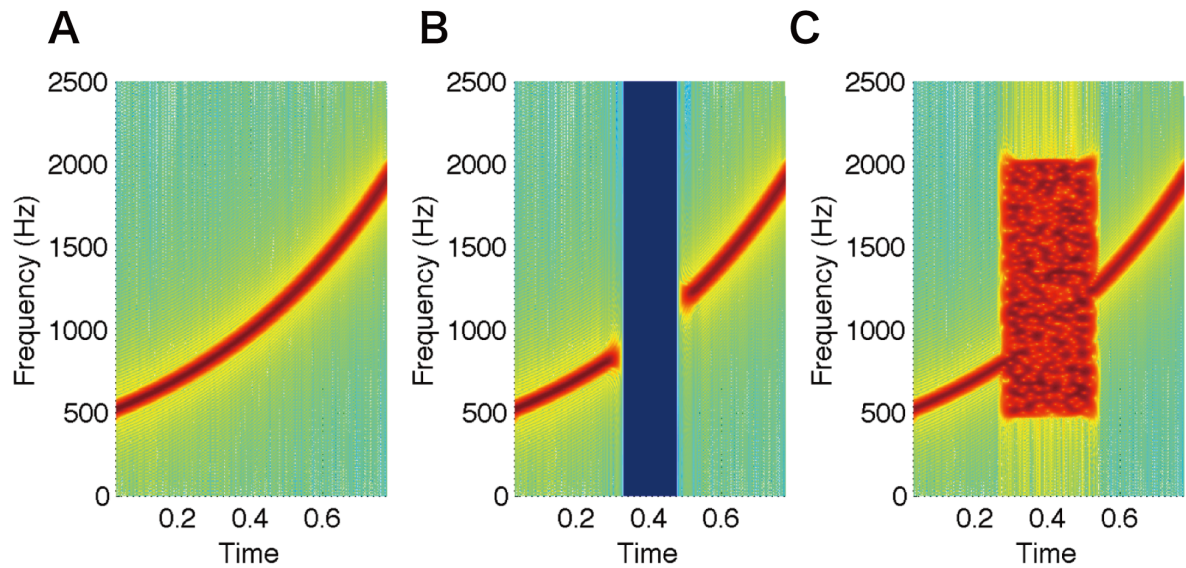


図 1.2: 連続聴錯覚. A: 周波数スイープ音のスペクトログラム. B: A の音の中間部 200 ミリ秒を空白で置き換えた音のスペクトログラム. C: B の空白部分に強い帯域通過雑音を挿入した音のスペクトログラム. この音では連続聴錯覚が生じ, A の音のような周波数スイープ音が雑音とともに聴こえる.

して得るのか, またこれらの手がかりがどのような神経基盤のもとに保存されているのか, この点に関して現在も議論が交わされている (McDermott et al., 2011).

1.5 聴覚の補完・分離機能を反映する現象

聴覚情景分析の一端をあらわす例として広く知られる 2 つの知覚現象 (連続聴錯覚・ABA 音ストリーム) について説明する. これらの現象は, 聴覚情景分析に関する多くの研究で取り上げられる.

1.5.1 連続聴錯覚

講演を聴いている最中に他者が咳をしたとしても, 講演者の話し声は咳の背後で滑らかに繋がって聞こえる. 一般に, 強い雑音が音響信号の一部に重なる場合, 聴取者は信号

が雑音の背後で連続的に聴こえるように感じる。この現象は、特定の状況が揃えば、もとの音が雑音と置き換えられ、物理的に存在しなかったとしても発生する (図 1.2)。この現象は、連続聴錯覚もしくは連続聴効果 (continuity illusion, auditory induction, amodal completion) と呼ばれる (総説は Petkov & Sutter, 2011; R. M. Warren, 2008; Bregman, 1990 を参照)。連続聴錯覚は言語音でも起こり、この場合特に音素修復、あるいは音韻修復 (phonemic restoration) と呼ばれる (R. M. Warren et al., 1972)。

これまでの研究によって、連続聴錯覚が生じるためにはいくつかの条件が必要であることが明らかとなっている (総説は Winkler et al., 2009; R. M. Warren, 2008 を参照)。条件の一つは、補完されるべき音に挿入される雑音が、補完されるべき音をマスキングするために十分な音圧を持つことである。この条件をマスキング可能性の法則と呼ぶ。また、連続して聴こえる音 (被誘導音) と誘導音 (雑音など、被誘導音の間に差し挟まれる音) の間に知覚的に検出可能な長さの無音区間が存在しないこと、誘導音の長さが 200 ミリ秒から 300 ミリ秒程度以下であること、誘導音と被誘導音の音響特性が不連続に変化していること、誘導音の前後で被誘導音の音響特性が著しく異ならないこと、などの法則が知られている。これらの法則から、連続聴錯覚は物理的に背後の音が隠されていると解釈して矛盾のない場合にのみ起こることが分かる。連続聴錯覚の存在は、聴覚系が、雑音の背後に存在すべき音をあらかじめ予測し、周囲の状況に合わせて適切に補完する機構を備えていることを示唆する。

視覚研究の分野においても、連続聴錯覚に類似する現象が知られている。図 1.3 の A では、青い領域は無意味な断片として見えるが、図 B では文字として認識される。図 A と図 B では青い領域の形状は共通している。図 B では丸い円が文字を隠すように配置されている一方、図 A では円は背景の白と同化している。この現象は、連続聴錯覚のマスキング可能性の法則と類似している。ある音の一部が削除されると「そこには何もない」ことが明らかとなる。しかし、削除された部分に、音を十分にマスキング可能な雑音などがある場合、聴覚系は「そこには何もない」という物理的な証拠が存在しないため、雑音の背後に音が存在すると解釈しやすくなる。

連続聴錯覚は視覚の補完現象である「カニッツアの三角形」(Kanizsa, 1955) という錯覚との類似性がある。「カニッツアの三角形」は一部が欠けた円とくさびの線のみで構成

される図形の中に、はっきりとした白い三角形の像が浮かび上がる現象である (図 1.4). 白い三角形を囲む仮想の輪郭は主観的輪郭 (illusory contour) と呼ばれる。「カニッツアの三角形」は、視覚系が周辺の図形の情報をもとに三角形の存在を補完するために現れると考えられている (Kalar, Garrigan, Wickens, Hilger, & Kellman, 2010; Kanizsa, 1976).

1.5.2 音脈分凝と ABA 音ストリーミング

聴覚系が音に含まれる要素を時間方向でまとめ、知覚的に独立したひとつのまとまりとして解釈する働きのことを音脈分凝 (auditory streaming) と呼ぶ (総説は Bregman, 1990 を参照). 音脈分凝のように、聴覚系が周波数軸上について距離のある音響的要素同士をまとめ、時間的な流れを作る働きを経時グルーピング (sequential grouping) と呼ぶ. 一方、時間的に同時に生じる音響的要素について、知覚上のまとまりを作る働きを同時グルーピング (simultaneous grouping) と呼ぶ.

音脈分凝については、「ABA 音ストリーミング刺激」と呼ばれる実験刺激を用いた研究が多く存在する (van Noorden, 1977; Micheyl, Tian, Carlyon, & Rauschecker, 2005; Shamma, 2011 など). ある短い正弦波音 (A 音) と、A 音と同じ時間長を持ち、また A 音よりも周波数が高いサイン波の音 (B 音) を、ABA, ABA... のように繰り返す音列を考える (図 1.5 A). この音列は「ABA」の 3 つの音の組をひとまとまりとして捉える知覚 (1 ストリーム知覚: 図 1.5 B) と、「A 音のみの系列のまとまり」「B 音のみの系列のまとまり」が独立して存在するように感じる知覚 (2 ストリーム知覚: 図 1.5 C) の 2 種類の知覚状態を生じさせる. A 音と B 音の周波数差が大きいほど、また、ABA 系列の呈示速度が速いほど、2 ストリーム知覚が生じやすいことが知られている (van Noorden, 1977).

ABA 音系列は、特定の条件下で「多義刺激」として振る舞う. 多義刺激とは知覚的な曖昧性を持つ刺激である. A 音と B 音の周波数差と、音の呈示速度がある一定範囲に設定されている場合、ABA 系列を連続して繰り返し聴くうちに、音の聴こえが 1 ストリーム知覚から、過渡的な知覚を経て、2 ストリーム知覚の状態へと移り変わる. この現象を「音脈分凝のビルドアップ」と呼ぶ (Bregman, 1990; Carlyon, 2004). 音を聴き続けると知覚の状態は 2 ストリームから 1 ストリームへと戻り、その後交替を繰り返す.

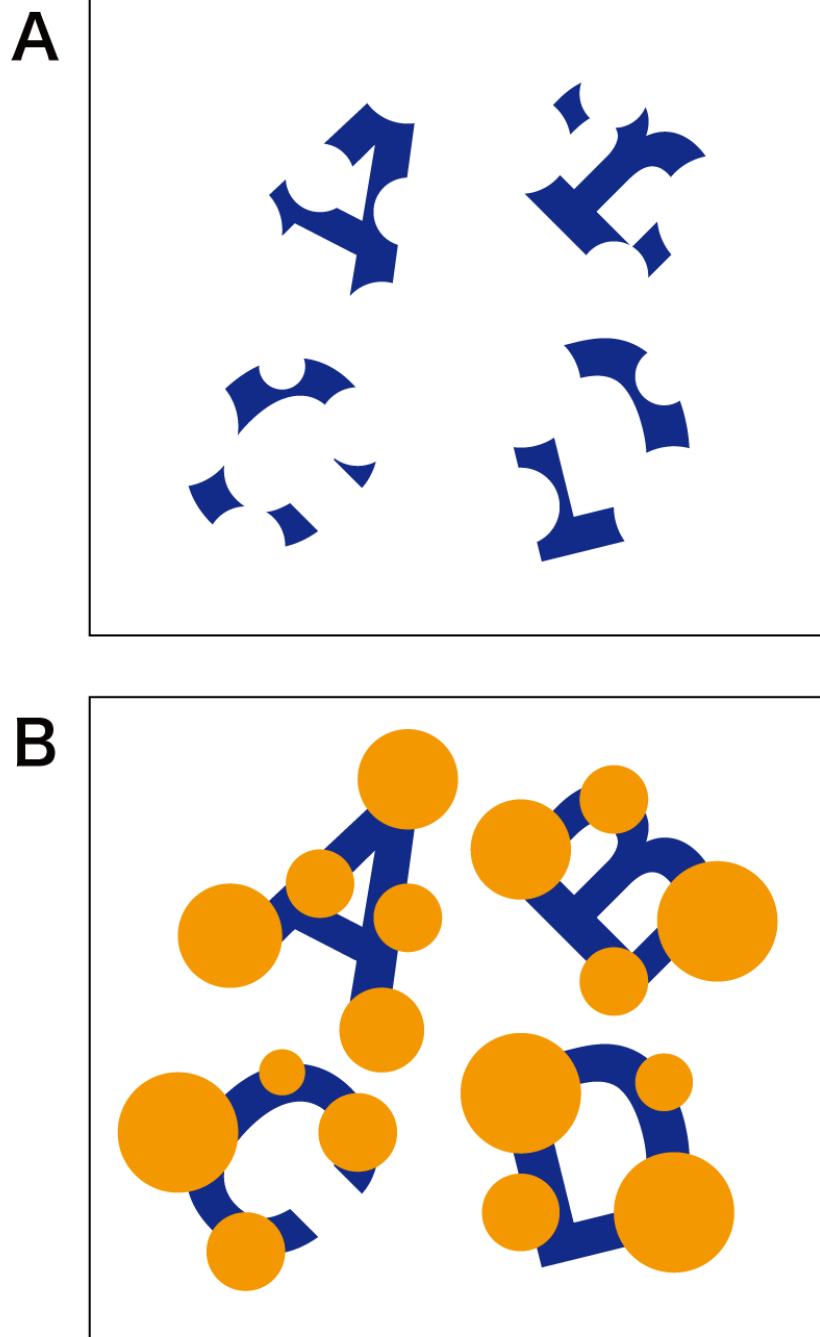


図 1.3: 連続聴錯覚と類似する視覚現象. A では断片化された青い領域の意味は読み取れないが, B のように欠落部分に色が存在する場合, 青い領域が文字として認識されやすくなる. ウェブサイト「イリュージョンフォーラム」(<http://www.kecl.ntt.co.jp/IllusionForum/>) より一部改変して引用.

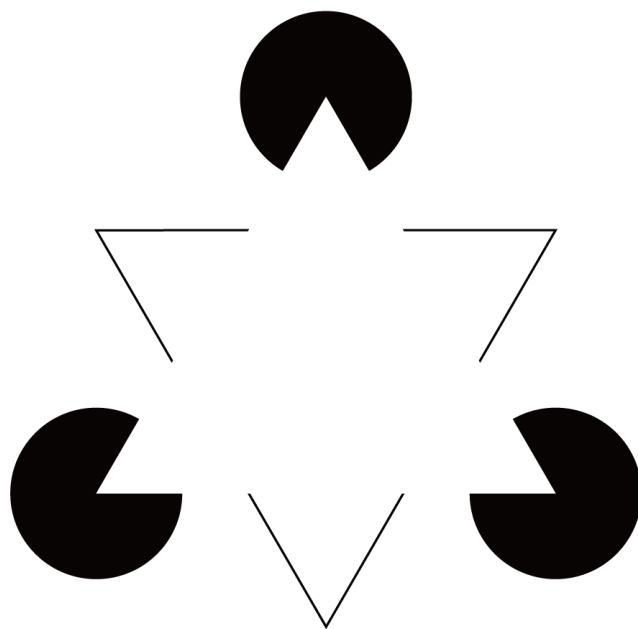


図 1.4: カニッツアの三角形 (Kanizsa,1955). 物理的には存在しない白い三角形の輪郭が知覚される.

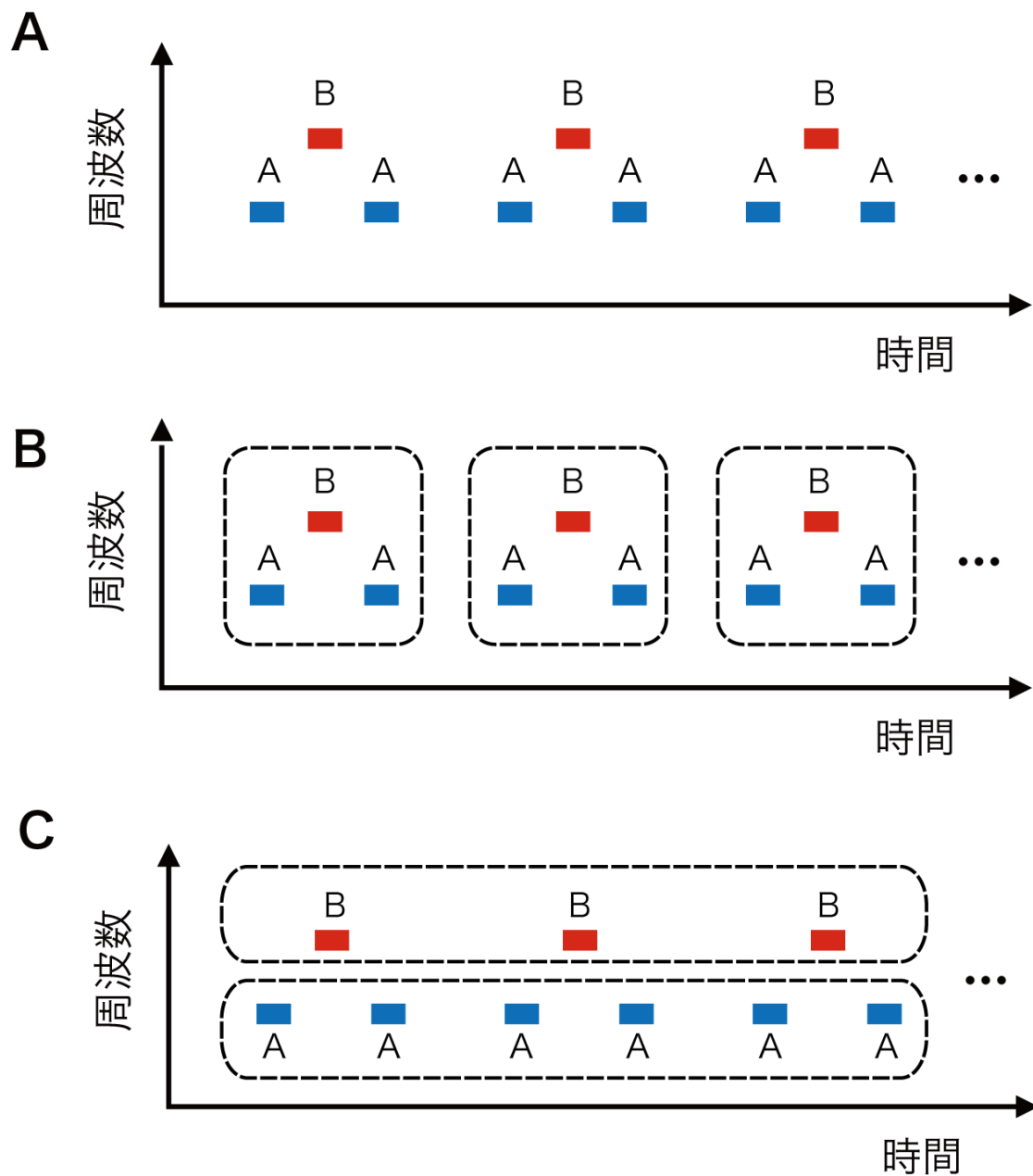


図 1.5: ABA 音ストリーミングの模式図. 赤と青の四角は周波数の異なる 2 つの正弦波音を表す. A: ABA 音系列. B: 1 ストリーム知覚の模式図. 「ABA」の三つ組の音がひとまとまりとして聴こえる. C: 2 ストリーム知覚の模式図. A 音のみのまとまりと B 音のみのまとまりが独立して存在するように聴こえる.

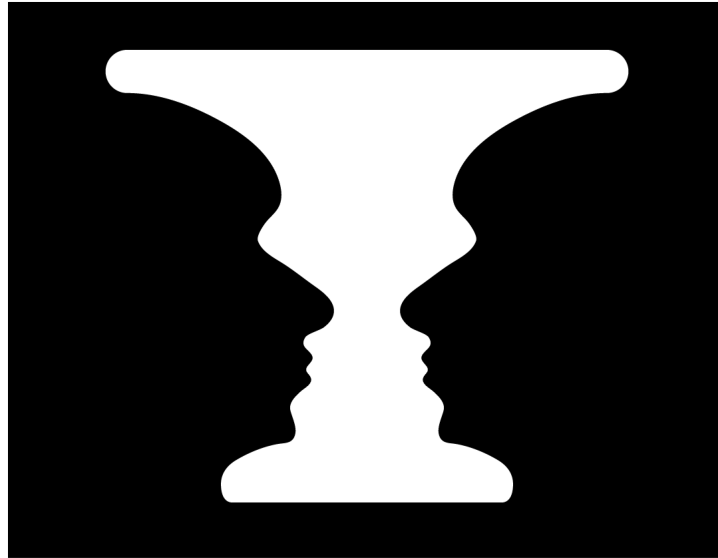


図 1.6: ルビンの壺.「向かい合う2つの顔」の知覚と「白い壺」の知覚が入れ替わる.

多義刺激は、知覚と意識のメカニズムを調べる材料として、主に視覚研究の分野において広く取り上げられてきた (総説は Logothetis & Leopold, 1999; Long & Toppino, 2004; Sterzer, Kleinschmidt, & Rees, 2009 を参照). 「ルビンの壺」(図 1.6) は広く知られる多義刺激である. 「ルビンの壺」を長時間見つめると、黒い背景の中に白い壺がある図と、白い背景の中で人間の顔が向かい合う図の知覚が交互にあらわれる. 聴覚研究の領域では、ABA 音の他に音声の知覚交替現象 (verbal transformation) が知られている (R. M. Warren & Gregory, 1958). これは「バナナ」などの単語の発話を連続して聴くうちに「ナッパ」「ナンバ」など、元々の音声の単語とは異なる単語が聴こえる現象である. 多義刺激を用いることにより、知覚の切り替わりを担う脳の部位を調べることができる (Tong, Meng, & Blake, 2006). ABA 音列を用いた知覚交替の研究については、1.8.2 節にてさらに詳細を述べる.

1.6 聴覚系の構造と機能

この節では、音波の入口である聴覚末梢系から大脳の聴覚皮質に至るまでの経路 (聴覚経路) について、過去の研究から得られた知見を概説する. これまでの解剖学的・生

理学的・心理物理学的知見から、聴覚経路は階層性を持つ複雑な系であり、様々な脳の領域が動的に協調することで機能を実現していることが分かっている (Plack, 2010; Rees & Palmer, 2010).

1.6.1 聴覚末梢系の構造と機能

聴覚末梢系は外耳・中耳・内耳から構成される (図 1.7). 外耳は耳介と外耳道からなる. 耳介は複雑な形状を持つため、高い周波数の音波は耳介で干渉を起こす. 個々人で異なる耳介と頭部の形状によって、音源から外耳道入口までの音の伝達特性が決まる. これを頭部伝達関数と呼ぶ. 音源から耳に届く音波は、頭部伝達関数と外耳道の伝達特性によって、スペクトルが修飾された状態で鼓膜へ届く.

中耳を構成する主な構造は鼓膜、耳小骨、中耳腔である. 鼓膜は薄い膜状の器官であり、外耳道と接する位置に存在する. 耳小骨は鼓膜の振動を耳に伝えるつち骨、きぬた骨、あぶみ骨の総称である.

内耳は蝸牛と前庭器官と三半規管からなる. 内耳は骨に囲まれた空洞の内部に膜で隔てられた管を持つ. 蝸牛は聴覚の受容器であり、前庭器官と三半規管は平衡感覚の受容器である. 蝸牛はカタツムリの殻のように巻かれた形状をしている. 蝸牛内部はライスネル膜、基底膜という 2 つの膜状の組織によって、前庭階、中央階、鼓室階に分けられる. 基底膜は中央階、鼓室階を隔てる弾力性を持つ組織である. コルチ器官は、音波を神経信号へと変換する役割を持つ器官であり、基底膜の上に載っている (図 1.8). 前庭階は前庭窓と呼ばれる穴であぶみ骨と接し、音の振動を受け取る.

音波が蝸牛に伝わると基底膜が振動する. 基底膜は前庭窓に近い基部ほど狭く厚く、先端部に進むほど薄く広い構造を持つ. このため、基底膜の振動は卵円窓に近い基部から先端部へ進む進行波となる. 波の振幅は、高周波数の音では基底膜上の基部に近い場所で最大となり、低い周波数では先端部に近い場所で最大となる. 複合音では、高周波成分が基底膜の基部近辺を振動させたのち、順次奥の基底膜を振動させる (図 1.9). 蝸牛の基底膜は、音波の周波数情報を基底膜上の場所情報に変換する役割を持つ. 聴覚系は、ある音波がある瞬間にどのような周波数成分を持つかを、基底膜上のどの位置で振動が起こっ

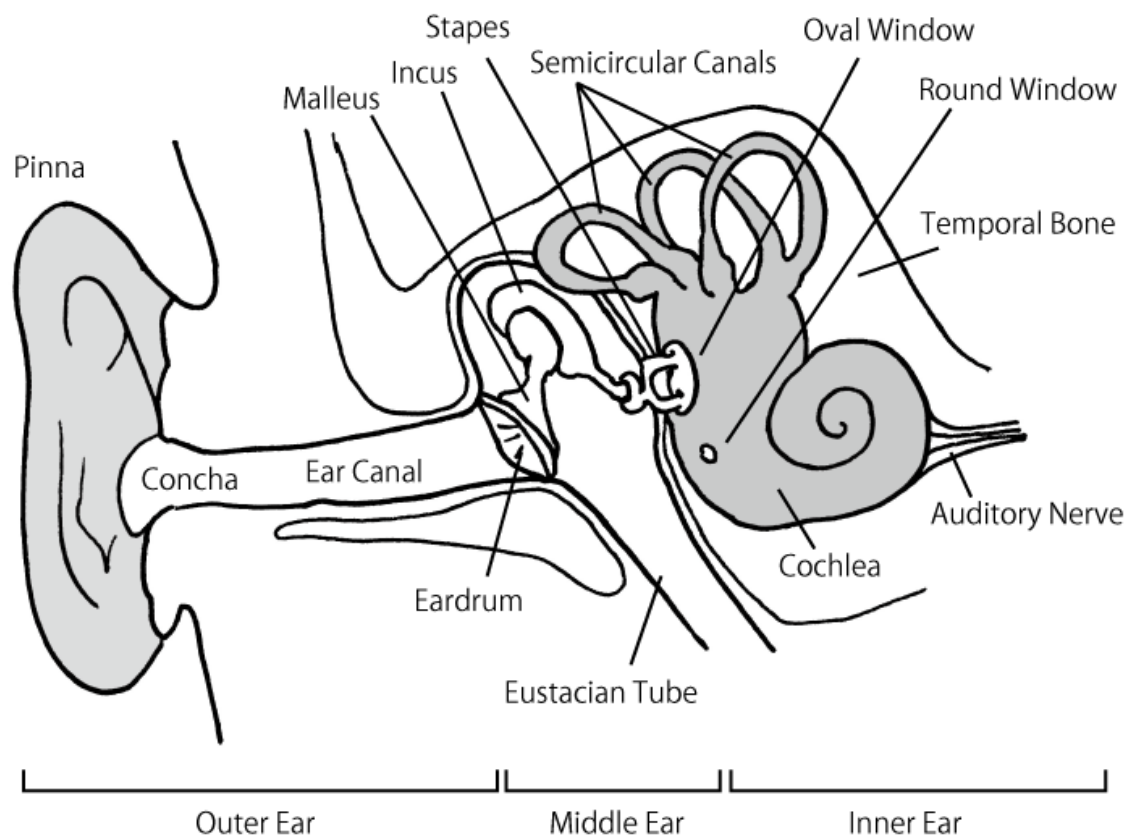


図 1.7: 聴覚末梢系の模式図. Outer Ear: 外耳, Middle Ear: 中耳, Inner Ear: 内耳, Pinna: 耳介, Concha: 耳甲介, Ear Canal: 外耳道, Eardrum: 鼓膜, Eustacian Tube: エウスタキ管, Malleus: つち骨, Incus: きぬた骨, Stapes: あぶみ骨, Semicircular Canals: 三半規管, Oval Window: 卵円窓, Round Window: 正円窓, Temporal Bone: 側頭骨, Cochlea: 蝸牛, Auditory Nerve: 聴神経. Plack (2010) を参考に作成.

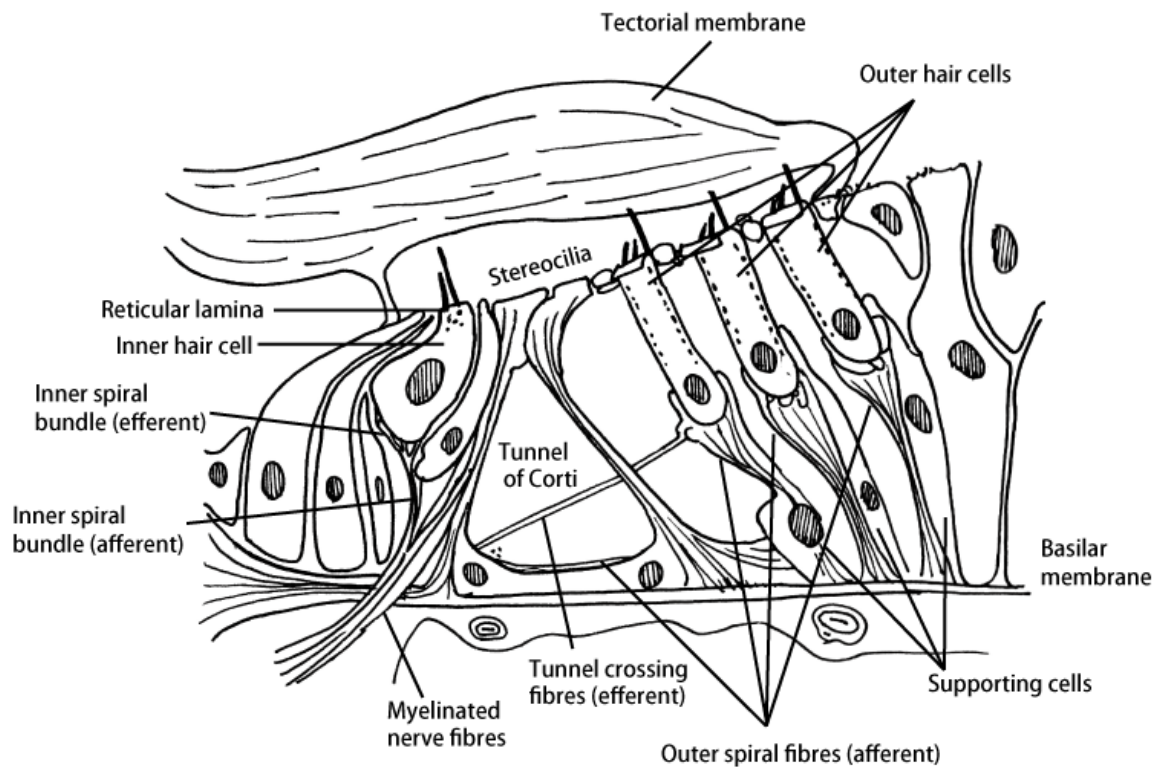


図 1.8: コルチ器官の概略図。蓋膜 (Tectorial membrane) の下に 1 列の内有毛細胞 (Inner hair cell) と 3 列の外有毛細胞 (Outer hair cell) が並ぶ。 Moore (2012) を参考に作成。

たを捉えることによって知る。すなわち、蝸牛の基底膜は音の周波数成分を解析する帯域フィルタの集合体として機能している。このフィルタの通過帯域幅は低域ほど狭く、高域ほど広い。また応答特性は非線形時変である。

基底膜上には有毛細胞が内側に 1 列、外側におよそ 3 列で並ぶ (図 1.8)。内側に並ぶ有毛細胞を内有毛細胞、外側に並ぶ有毛細胞を外有毛細胞と呼ぶ。入力音により基底膜と蓋膜が振動することで 2 つの有毛細胞の毛も振動する。毛がある一定方向に傾くとき、細胞内に電位が発生し、聴神経に発火情報が伝達される。ヒト内有毛細胞には複数個の一次聴神経が接続する一方、外有毛細胞にはほとんど接続しない。外有毛細胞には中枢から末梢の方向へ信号を送る、遠心性の神経が接続する。ひとつの内有毛細胞に接続する聴神経には、発火の閾値が低い種類のものと高い種類のものが混在する。聴神経は基底膜の振動を発火頻度と発火タイミングの情報として符号化する。ある内有毛細胞はある特定の

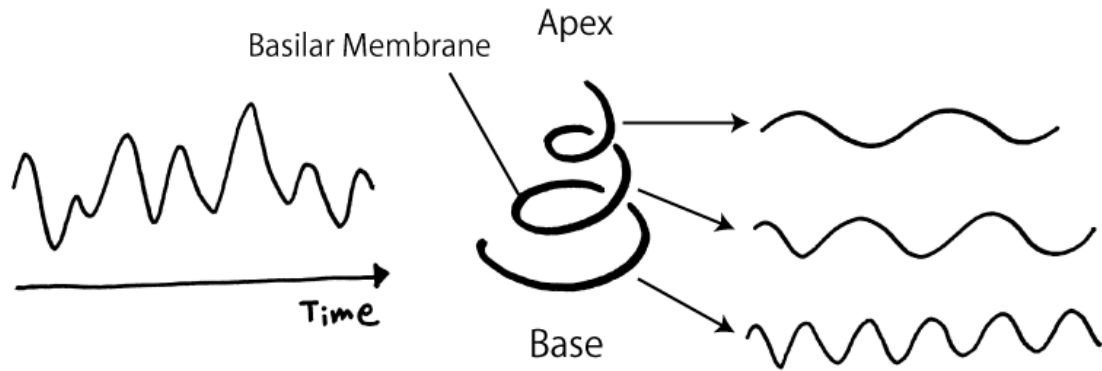


図 1.9: 蝸牛での周波数分析の模式図. 基底膜 (Basilar Membrane) の基部 (Base) 側に近い場所は音の高周波数成分を, 先端部 (Apex) に近い場所は低周波数成分を神経符号化する. Plack (2010) を参考に作成.

領域の周波数成分に応答する. 耳に入力された音は蝸牛内で周波数分析され, 周波数スペクトル情報と波形の周期情報は聴神経の発火パターンとして符号化される.

1.6.2 聴覚中枢系の構造と機能

聴覚系を構成する蝸牛神経核以降の神経核群を聴覚中枢系と呼ぶ. 聴覚中枢系では音の持つ特徴が強調され抽出されていると考えられているが, 聴覚末梢系と比較して機能の面においては未だ不明な点が多い. 図 1.10 に聴覚中枢系の主要な神経投射の上向経路を示す.

蝸牛神経核は聴覚中枢系の最初の神経核である. 聴神経からの投射の空間配置は, 周波数軸的に連続となっていることが多い. この周波数軸での神経投射配置の連続性をトノトピーと呼ぶ. トノトピーは蝸牛神経核から聴覚皮質まで, 聴覚経路の多くの領域で保存されている. 蝸牛神経核の前腹側核は聴神経の発火間隔の変動を減らし, 両側の上オリーブ複合体に中継する.

上オリーブ複合体 (superior olivary nucleus; SO) は両側の蝸牛神経核からの投射を受けて, 両側の耳に到達した音の音圧レベル差 (両耳間レベル差) と到達時間差 (両耳間時

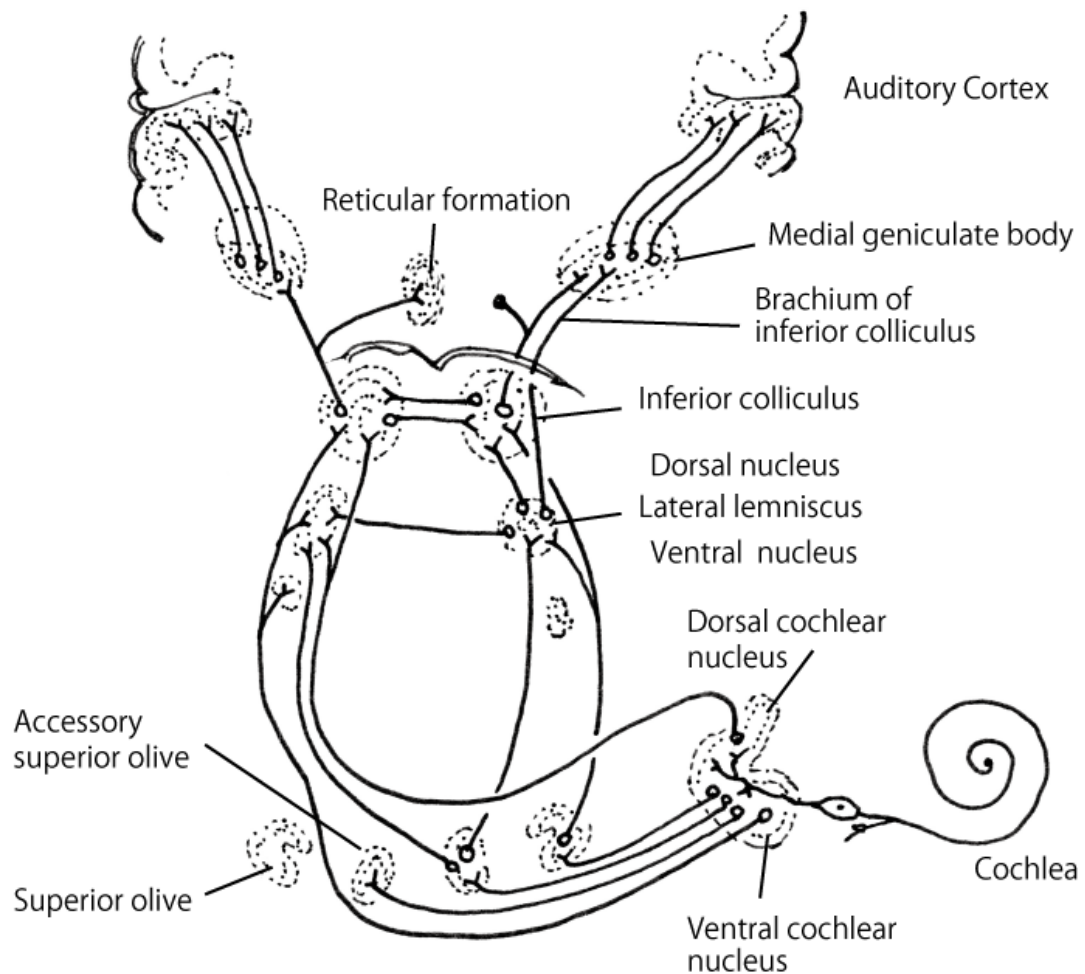


図 1.10: 聴覚上向 (求心性) 経路の模式図. 代表的な経路のみを記す. Cochlea: 蝸牛, Ventral cochlear nucleus: 腹側蝸牛神経核, Dorsal cochlear nucleus: 背側蝸牛神経核, Superior olive: 上オリーブ核, Lateral lemniscus: 外側毛帯核, Inferior colliculus: 下丘, Medial geniculate body: 内側膝状体, Auditory cortex: 聴覚皮質. Warren (2008) を参考として作成.

間差) を算出する。両耳間レベル差は外側核 (lateral superior olive; LSO) にて発火頻度情報から計算され、両耳間時間差は内側核 (medial superior olive; MSO) にて発火タイミングの情報から計算されると考えられている。

外側毛帯 (lateral lemniscus; NLL) は蝸牛神経核と上オリーブ複合体からの投射を受け下丘へ投射する。下丘 (inferior colliculus; IC) は聴覚系のほぼ全ての神経核から投射を受ける。下丘は中心核 (ICC)、背側核 (ICD)、周辺核 (ICX) の3領域で構成される。ICC は下丘の主な求心性の経路である。ICX は ICC と体性感覚の情報を受け、多感覚の情報処理を行う。ICD は主に聴覚皮質からの遠心性投射を受ける。

内側膝状体 (medial geniculate body; MGB) は視床に位置し、解剖学的に腹側核 (vMGB)、背側核 (dMGB)、内側核 (mMGB) に分けられる。vMGB は ICC からの投射を受けて一次聴覚皮質へ投射する聴覚の主経路である。dMGB は ICD から投射を受け一次聴覚野の周辺領域へ投射する。mMGB は ICX からの投射を受け、多感覚情報を統合する。mMGB には、情動情報の処理に関係する扁桃体への投射経路が存在する。

聴覚皮質は、霊長類では側頭葉上側頭回の皮質上に位置し、シルビウス溝の内側に存在する (図 1.11)。聴覚皮質は解剖学的に中心領域 (core) 中心領域を取り囲む帯状領域 (belt)、さらに側方の傍帯状領域 (parabelt) に分けられる。一次聴覚野は中心領域に存在し、vMGB からの投射を受ける。一次聴覚野はヒトの脳ではヘシュル回と呼ばれる領域に相当する。一次聴覚野のニューロンは正弦波音によく応答し、各領域内にトノトピーがみられる。また、一次聴覚野には音の時間・周波数・空間的特徴に対して特異的な応答を示すニューロンが存在する。その応答特性は単純ではなく、被験体・実験参加者への麻酔の有無、過去に入力された音信号、実験参加者の課題の内容などによって変化する。帯状領域は中心領域と密接な相互連絡を持つ。帯状領域には MGB から連絡があるが、主経路である vMGB からの投射は傍帯状領域と比較して少ない。傍帯状領域は MGB やその他の視床の領域から投射を受ける。傍帯状領域は側頭葉の多感覚情報を統合する領域や前頭前野の領域と接続している。

聴覚皮質から大脳皮質の各領域には、機能が異なる複数の情報処理経路があると考えられている。一次聴覚野から帯状領域の尾側を経由し頭頂葉へ向かう経路は、音源の空間的位置の情報を処理する 'where' 経路、一次聴覚野から帯状領域の吻側と傍帯状領域

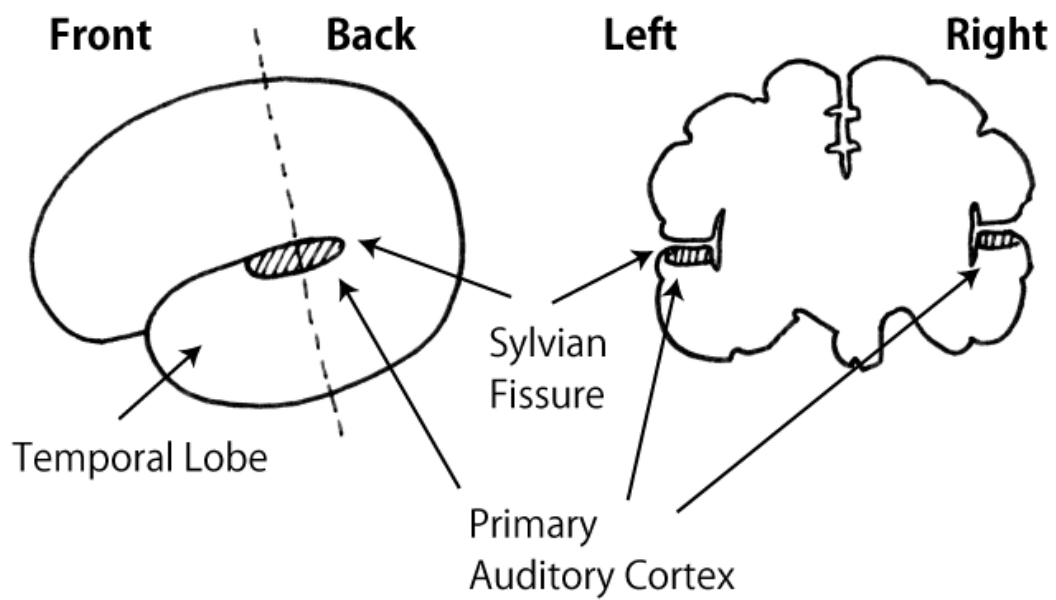


図 1.11: 大脳皮質における一次聴覚皮質の位置. (左図) 側面からの図. (右図) 左図の点線を断面としたときの図. Temporal Lobe: 側頭葉, Primary Auditory Cortex: 一次聴覚皮質, Sylvian Fissure: シルビウス溝. Plack (2010) を参考に作成.

を經由して前頭前野に向かう経路は、音源が何であることを特定する‘what’経路であると考えられている (総説はCloutman, 2013; Kaas & Hackett, 1999 を参照). また頭頂葉を經由して前頭前野に至る経路を、音響信号を身体運動と結びつける‘do’経路と意味付ける研究もある (Belin & Zatorre, 2000). こうした「経路の意味付け」に関する仮説は、視覚情報の処理経路に関する仮説との類似性がみられる (Rust & Stocker, 2010; Ison & Quiroga, 2008; Riesenhuber & Poggio, 2002). ただし、‘what’経路とされる前頭前野の経路であっても空間情報処理に関連する神経応答が見つかるほか、‘where’経路とされる頭頂葉の経路についても、入力音が何であることを特定するような処理に関連する神経応答が見られる (総説は Rauschecker, 2012; Bizley, Walker, Silverman, King, & Schnupp, 2009; S. D. Miller, Litovsky, & Kluender, 2009; Stecker & Middlebrooks, 2003 を参照).

聴覚経路は、鼓膜から皮質の方向へ情報を伝達する求心性経路だけではなく、上位の神経核から下位の神経核へ情報を伝達する遠心性の経路も数多く持つ. 遠心性経路と求心性経路は、各段階で情報伝達経路のループを形成する. 聴覚皮質と MGB には特に密接な双方向連絡が存在する. SOC から両側の蝸牛へは、オリーブ蝸牛束と呼ばれる神経繊維束による連絡がある. オリーブ蝸牛束は、突発的な聴覚入力から蝸牛を保護する機能を持つと考えられている. また、選択的注意 (1.9 節) に関する機能を持つと主張する研究もある (Guinan Jr., 2006; Delano, Elgueta, Hamame, & Robles, 2007; De Boer & Thornton, 2007).

1.7 音の特徴抽出

鼓膜の振動として入力された音の信号は、蝸牛にて周波数解析を受け、その後いくつかの神経核を中継される中で様々な処理を受ける. 本節では、蝸牛基底膜での周波数解析の後、聴覚オブジェクトの構成要素である「音高 (ピッチ)」「音色」の特徴が抽出される過程についての知見をまとめる.

1.7.1 音高 (ピッチ) の表現

音高 (ピッチ) は、音の高さの主観的感覚であり、音の基本的な特徴のひとつである。音楽の旋律はピッチの上昇下降によって作られる。また、音声においてもピッチの変化パターンによって意味が伝達される。

一般的に、時間的周期性を持つ音のピッチは、音の周波数成分のうちもっとも低いものによって決定される。この周波数成分は基本周波数と呼ばれる。ただし、スペクトル包絡の形状の異なる音響信号であっても同一のピッチを与える場合があることが知られている (Moore, 2012)。また音の基本周波数が存在しなくとも、その連続倍音が残されているとき基本周波数のピッチが知覚される場合がある。この現象はミッシング・ファンダメンタル (missing fundamental) と呼ばれている。基本周波数を雑音でマスクした場合にもピッチが知覚されることから、ピッチの知覚は蝸牛の基底膜ではなく、聴覚系の中枢にて生成されていると考えられている。ピッチに関する研究の歴史は比較的長く、基底膜上の場所情報がピッチを決定するという説 (場所説) と聴神経の発火パターン情報がピッチ知覚に重要であるとする説 (時間説) の対立が存在していた。現在では、基底上の場所の情報、聴神経の発火パターンの情報の2つ共にピッチ決定に重要な役割を果たすとする折衷案が立てられている (総説は Moore, 2012 を参照)。

近年の研究では、ピッチ知覚に対応する神経活動が聴覚皮質の特定の領域において発見されている。霊長類では、聴覚皮質の中心領域や帯状領域にて、ピッチに対して選択的に応答するニューロンが発見されている (Bendor & Wang, 2005)。また、ヒトの研究についても、ヘシュル回前部においてピッチ感覚に対する応答を示す領域が発見されている (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Penagos, Melcher, & Oxenham, 2004; J. D. Warren & Griffiths, 2003)。ただし、ピッチ知覚に関与する神経系は聴覚皮質上の広範囲にネットワークとして分布しているとする研究 (Garcia, Hall, & Plack, 2010; Griffiths et al., 2010; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009; Hall & Plack, 2009) があり、皮質上のある特定の領域だけがピッチ知覚に関係すると捉えることは難しい。

1.7.2 音色 (スペクトル構造) の表現

音色は、音のスペクトル形状およびその時間変化のパターンなどによって決定される音の質である。音色の特徴をとらえることにより、聴取者はヴァイオリンやギターなどの楽器が同じピッチの旋律を奏でていたとしても楽器の種類を聴き分けることができる。音色は「大きい, 小さい」「高い, 低い」などの1次元の軸ではなく、複合的かつ多次元的に表現される。

音色の神経表現もまた、聴覚皮質に広く分布していることが知られている。単一細胞記録 (Bizley et al., 2009; Schebesch, Lingner, Firzlaff, Wiegrebe, & Grothe, 2010; Versnel & Shamma, 1998) や脳画像イメージング (Kumar, Stephan, Warren, Friston, & Griffiths, 2007; Formisano, De Martino, Bonte, & Goebel, 2008) の研究にて、音色に特異的に応答するニューロンやニューロン群が発見されている。ただし、音色の神経表現は一定不変のものではなく、同じ音が持つその他の特徴 (ピッチ, 空間的位置) の影響を受けて変動することが知られている (Bizley et al., 2009)。

1.8 音響的要素のグルーピング

聴覚皮質にて構成要素であるピッチや音色などの特徴が表現された後には、個々の特徴を繋ぎ合わせて聴覚オブジェクトを形成する段階が存在すると考えられる。これまでに、聴覚オブジェクトが形成される段階について、同時グルーピングと経時グルーピング (1.5.2 節を参照) が聴覚系内でどのように行われるかが調べられてきた。本節では、抽出された音の特徴がひとつの知覚的まとまりとして構成され、聴覚オブジェクトとなるまでの過程について、これまでに知られている知見をまとめる。

1.8.1 同時グルーピングに関する研究

いくつかの報告から、連続聴錯覚 (1.5.1 節参照) の知覚と一次聴覚皮質の働きには関係があると考えられるようになってきた。連続聴錯覚を引き起こす音刺激について、音が連続に聴こえたか不連続に聴こえたかを判断する実験を行い、脳画像イメージングを

行った研究 (Riecke, Esposito, Bonte, & Formisano, 2009; Riecke, Micheyl, & Oxenham, 2012) から、物理的には全く同じ音であっても、聴取者が「連続である」と知覚した場合と「連続ではない」と知覚した場合では、一次聴覚皮質において異なる神経活動のパターンがみられることが明らかとなった。この発見から、単純な周波数時間情報ではなく、聴覚オブジェクトが聴取者の知覚を決定するとする説が提案された (Riecke et al., 2012)。また、マカクザルの一次聴覚皮質からは、連続聴錯覚によって補完された音の表現に対応する神経活動が見つかっている (Petkov, O'Connor, & Sutter, 2007)。ただし、この研究では神経活動のデータと行動実験のデータが同時には集められていないため、実験結果が聴覚オブジェクトの知覚と対応するか否かは不明である (Bizley, Walker, Nodal, King, & Schnupp, 2013)。

1.8.2 経時グルーピングに関する研究

ABA 音系列を聴くことで知覚される代表的な 2 つの知覚状態は、グルーピングを行う際に聴覚系が用いる規則の違いに対応すると考えられるため、ABA 音系列を使って知覚状態の切り替わりに関係する脳部位を特定する研究が行われている (Micheyl et al., 2007; Kondo & Kashino, 2009; Kashino & Kondo, 2012)。動物での電気生理実験とヒトの脳画像イメージングの研究から、音脈の知覚が 1 ストリーム知覚であるときと 2 ストリーム知覚であるときに応答が変化する脳領域は、聴覚皮質 (Shamma & Micheyl, 2010; Gutschalk, Micheyl, Melcher, Scherg, & Oxenham, 2005)、頭頂間溝 (Cusack, 2005) などの高次聴覚情報処理系に存在するとともに、内側膝状体 (Kondo & Kashino, 2009)、下丘 (Schadwinkel & Gutschalk, 2011)、さらに蝸牛神経核 (Pressnitzer, Sayles, Micheyl, & Winter, 2008) にも存在することが明らかとなった。

近年の知覚交替に関する研究から、音声の知覚交替では、視覚の知覚交替で活動する脳領域と重複する領域が活動することが判明した (Sato, Basirat, & Schwartz, 2007; Kondo & Kashino, 2007)。視覚の知覚交替に関連する脳の領域はネットワークとして偏在していると考えられている (総説は Tong et al., 2006; Sterzer et al., 2009 を参照)。同様に、聴覚の知覚交替についても複数の領域の活動量の動的変化が知覚を決定するとする説が

ある (Kashino & Kondo, 2012).

これまでに、ABA 音系列による音脈分凝を説明するためのモデルがいくつか提案されている。ある説では、皮質上におけるニューロン群の発火領域が重複しているか否かが、知覚の状態を決定するとしている (Fishman, Reser, Arezzo, & Steinschneider, 2001; Kanwal, Medvedev, & Micheyl, 2003; Fishman, Arezzo, & Steinschneider, 2004)。この説では、聴覚皮質で起こる神経活動の位置を読み取る機構が、聴覚皮質よりも後の情報処理段階に存在すると仮定している。聴覚皮質においてもトノトピー (1.6.2 節を参照) は保存されているため、A 音と B 音の周波数差が小さい場合、神経活動が活発な位置は限局され、神経活動のピークは一カ所に集中する。このとき、ピーク位置を読み取る仮説の機構は、ピークが一カ所のみに集中していることを理由に「1 ストリーム知覚」の判断を下すことができる。A 音と B 音の周波数差が大きくなるにつれて、神経活動は位置的に離れた 2 つのピークを形成するようになり、仮説の機構は「2 ストリーム知覚」の判断を下す。しかし、音脈分凝は正弦波音のみで構成される ABA 音系列だけではなく、ノイズや調波複合音でも起こることが知られている (Moore & Gockel, 2002)。上記の仮説では、こうした複雑な刺激で音脈分凝が起こる理由を説明できない。

近年発表された「Temporal Coherence モデル」(Shamma, 2011; Shamma & Micheyl, 2010; Elhilali, Ma, Micheyl, Oxenham, & Shamma, 2009) では、A 音と B 音の周波数差が大きい場合であったとしても、呈示されるタイミングが同時である場合には 1 ストリーム知覚が起こるという研究結果 (Elhilali et al., 2009) をもとに、神経活動の 2 つのピークの活動量変化のタイミングが揃っている場合には、聴取者は 1 ストリームを知覚し、タイミングがずれている場合には、2 ストリームを知覚すると説明する。ただし、さらに最近の研究から、時間的に同期する A 音と B 音であっても 2 ストリーム知覚が起こる場合があることが分かり (Micheyl, Kreft, Shamma, & Oxenham, 2013)、モデルの見直しが行われている。

1.9 聴覚オブジェクトと選択的注意

聴覚系はある周囲の事物や事象の特定部分、あるいは精神活動の特定の側面に対して、無関連な情報からの侵入を防ぎ、選択的に意識を向ける脳の働きを持つ (Driver, 2001; Lavie, 2005). 例えば騒がしいパーティ会場のように、音楽やグラスの触れ合う音が溢れる環境であっても、人は知人が自分を呼ぶ声を聴くことができる. 選択的聴取は、多くの音が存在する中で、そのとき聴取者が聴く必要のある音や、聴こうとする音が意識にのぼる現象である (総説は Shinn-Cunningham, 2008 を参照). Cherry による先駆的研究 (Cherry, 1953) では、2つの音声为重なりあった場合に片方の声が聴き取られる条件を考察している. パーティの情景は「騒がしい環境」の具体例として多くの論文に登場する.

1.9.1 聴覚オブジェクト形成への選択的注意の影響

オブジェクトの形成過程に注意はどのような形で影響するのかについて、決定的な結論は導かれていない. ある瞬間に利用可能な脳の計算資源には限りがあることから、選択的注意は、オブジェクトの効率的な形成を助けると考えられている (Shinn-Cunningham, 2008). 視覚研究の分野には、オブジェクトの解釈に対立が生じる理由は、外界に存在するすべてのオブジェクトを脳内で表現することが非効率であるためとする説がある (Desimone & Duncan, 1995; Kastner & Ungerleider, 2000).

Bregman は自著の中で、聴覚情景分析は「原初的な過程 (primitive scene analysis)」と「図式による過程 (schema-based analysis)」という2つの枠組みからなると主張した (Bregman, 1990). 「原初的な過程」とは、刺激そのものが含む物理的の手がかり (ボトムアップ手がかり) をもとに、関連性の深い音の成分同士をまとめ、意味や関係のある知覚的な音のまとまりを形成する過程である. 「図式による過程 (schema-based analysis)」とは、形成された音脈に対して、記憶などの事前知識や注意による編集が施される過程である. Bregman は「図式による過程」の段階で、分離された音のストリームが積極的な注意の影響を受けると考えた. Bregman の見解が発表されて以降「注意は聴覚情景分析のどの過程に影響するか」について、現在も活発な議論が続いている.

同時グルーピングと注意の関係を調べた研究では「注意は同時グルーピングの処理に

影響を与えない」と結論づけるものが多くみられる。知覚される聴覚オブジェクトの数が異なる際に発生する陰性の電位変化である「オブジェクト由来陰性電位 object-related negativity (ORN)」を測定した研究の結果、ORN は聴取者の注意の状態とは独立に発生することが明らかになった (Alain, Snyder, He, & Reinke, 2007)。

また、聴取者が刺激の中の特定の特徴の変化を捉えようとするときにも、注意は必要ではないことが知られている。音系列の中に一定割合で稀な音を挿入する「オドボール課題」を使って測定されるミスマッチ陰性電位 (mismatch negativity, MMN) の研究では、MMN は聴取者が音に対する注意を行わなくても起こることが知られている (Näätänen & Picton, 1987; Kujala, Tervaniemi, & Schröger, 2007; Picton, Alain, Otten, Ritter, & Achim, 2000; Alain, Woods, & Ogawa, 1994)。連続聴錯覚を刺激として、オドボール課題のパラダイムを適用した実験から、聴取者が音に注意をしなくても、連続性に関する判断は聴覚系内部で自動的に行われることが示唆されている (Micheyl et al., 2005)。

ABA 音列のストリーミング現象については、ストリーミングに注意が必要か否かについて研究者の意見が分かれている。ある研究では、音に注意を向けるか否かによって、聴取者の知覚が変わりうるということが明らかとなった (Carlyon, Cusack, Foxton, & Robertson, 2001; Cusack, Deeks, Aikman, & Carlyon, 2004)。この研究では、注意を別の課題に向けることによって、2 ストリーム知覚が 1 ストリーム知覚へ戻ることが示された。ただし、この結論には実験パラダイムに由来する問題点も指摘されており、Sussman らによる研究では、ストリーミングには必ずしも音に対する注意が必要ではないと結論づけられている (Sussman, Horváth, Winkler, & Orr, 2007)。

1.9.2 注意機能の存在意義

注意は、聴覚情景が曖昧性を持つとき、曖昧性を解決し、ただひとつの解釈を決定するためにあると考えられる。上記の解釈を裏付けるように、聴取者が ABA 音系列を呈示された場合には、2 つの異なる種類の事象関連電位が異なる時間遅れを伴って生じることが知られている (Winkler, Takegata, & Sussman, 2005; Snyder, Alain, & Picton, 2006; Snyder, Carter, Hannon, & Alain, 2009)。時間的に早く出現する 1 番目の電位変化は、2

つの知覚的解釈 (1 ストリーム知覚, 2 ストリーム知覚) を反映し, 続く 2 番目の電位変化は聴取者の知覚の選択と決定を反映すると解釈されている (Winkler et al., 2005). また, 音信号が大まかにそれぞれの構成要素である聴覚オブジェクトに分解された後, 選択的注意がさらなる処理の促進を行い, 曖昧で対立した知覚解釈を状況に合わせて修正していくとの見方もある (Knudsen, 2007; Shinn-Cunningham & Best, 2008). 注意は, 聴覚オブジェクトを選び取る段階にて影響を及ぼす (Alain & Arnott, 2000; Desimone & Duncan, 1995; Zatorre, Mondor, & Evans, 1999) とともに, 処理段階としてはさらに下位である音の特徴抽出の段階に対しても影響を及ぼすと考えられている. また, ある音源の特徴に対して注意が向けられているならば, 同じ音源のその他の特徴に対しても検出の感度が上がると考えられている (Duncan, 2006). 事実, オブジェクトの形成に齟齬が生じた場合, 音源を特定する能力が低下することを示す研究が存在する (例として A. K. C. Lee & Shinn-Cunningham, 2008; Darwin & Hukin, 1998; Best, Gallun, Carlile, & Shinn-Cunningham, 2007). 日常環境は実験室とは異なり, 音源が数多く存在するため, 聴取者が目的に応じた妥当な知覚を選び取る能力はさらに重要であると考えられる (Shinn-Cunningham, 2008).

1.9.3 注意機能の神経基盤

注意は, 単純な一方向の神経接続によって実現されているのではなく, 空間的な注意と非空間的な注意とで異なるネットワークを持つ, 複雑な総体によって実現されていると考えられている (総説は Hill & Miller, 2010; A. K. C. Lee et al., 2012 を参照). fMRI の研究によって上側頭回 (superior temporal gyrus) の聴覚関連領域から, 注意の働きに関係する領域が見つかっている (Zatorre et al., 1999; Degerman, Rinne, Salmi, Salonen, & Alho, 2006; Salmi, Rinne, Degerman, Salonen, & Alho, 2007; Ahveninen et al., 2006). これに加えて上側頭溝 (superior temporal sulcus) や下前頭溝 (inferior parietal sulcus) についても注意に関係するとみられる神経活動が見つかっている. これらの領域は, 複雑で実際の環境の音に近い環境の音に注意を向けるとき, 関係性の高い神経活動が見られる (Hill & Miller, 2010) ことから, こうした高次の領域が下位領域にフィードバックを

返すために、下位領域でも注意に関係する応答が現れると主張する研究がある (Buffalo, Fries, Landman, Liang, & Desimone, 2010).

聴覚の注意の処理を反映すると考えられる神経活動は、聴覚皮質の様々な領域で観測されている。一次聴覚皮質では、注意はニューロンの選択性の特性を変化させるという報告がある (Fritz, Shamma, Elhilali, & Klein, 2003; Atiani, Elhilali, David, Fritz, & Shamma, 2009; Niwa, O'Connor, Sutter, & Johnson, 2012; C.-C. Lee & Middlebrooks, 2011). ヒトの研究において、注意が事象関連電位や脳イメージングの信号強度を増加させるという報告もある (Alain & Woods, 1997; Woods, Alho, & Algazi, 1993, 1992; Petkov et al., 2004; Rinne et al., 2007; Woldorff et al., 1993). 後横側頭領域 (posterior auditory cortex) などの高次の聴覚皮質では、神経活動がある特定のオブジェクトの知覚に対応していると主張する研究もある。ある研究では、実験参加者が重複する2つの音声のうちどちらかだけに注意を向けるよう指示されているとき、参加者が注意を向けた音声は、聴覚皮質の特定の領域の神経活動を優先的に変調させることがわかった (Ding & Simon, 2012). 近年、ヒトの聴覚皮質に表面電極を取り付けて行う測定実験の結果、注意を向けていない音に関係する神経活動が、注意を向けた音に関係する活動と比較して抑制されたという報告も発表されている (Mesgarani & Chang, 2012).

1.10 これまでの研究から分かっていないこと

聴覚系は複雑な構造と無数の下位構造を内部に持つ大規模な情報処理系である。それぞれの構造は双方向に連携しあい、同時並行で情報を伝送し、処理しあっている。また、注意機能を実現するシステムは、一般的に聴覚に関係すると考えられている脳の領域を超え、広範囲に分布して存在する。聴覚情景分析と知覚のグルーピングに関しては、現在も解かれていない問題が多くあるとともに、問題の範囲は多岐に渡る。聴覚情景分析の過程を解明する問題は、二つの下位問題に分けられる。ひとつは聴覚系の「どこで (where)」処理が行われているのかという問題、もうひとつは「どのように (how)」処理が行われているのかという問題である (図 1.12)。以下には、Shamma and Micheyl (2010) と Carlyon (2004) にて示された未解決の問題を示す。

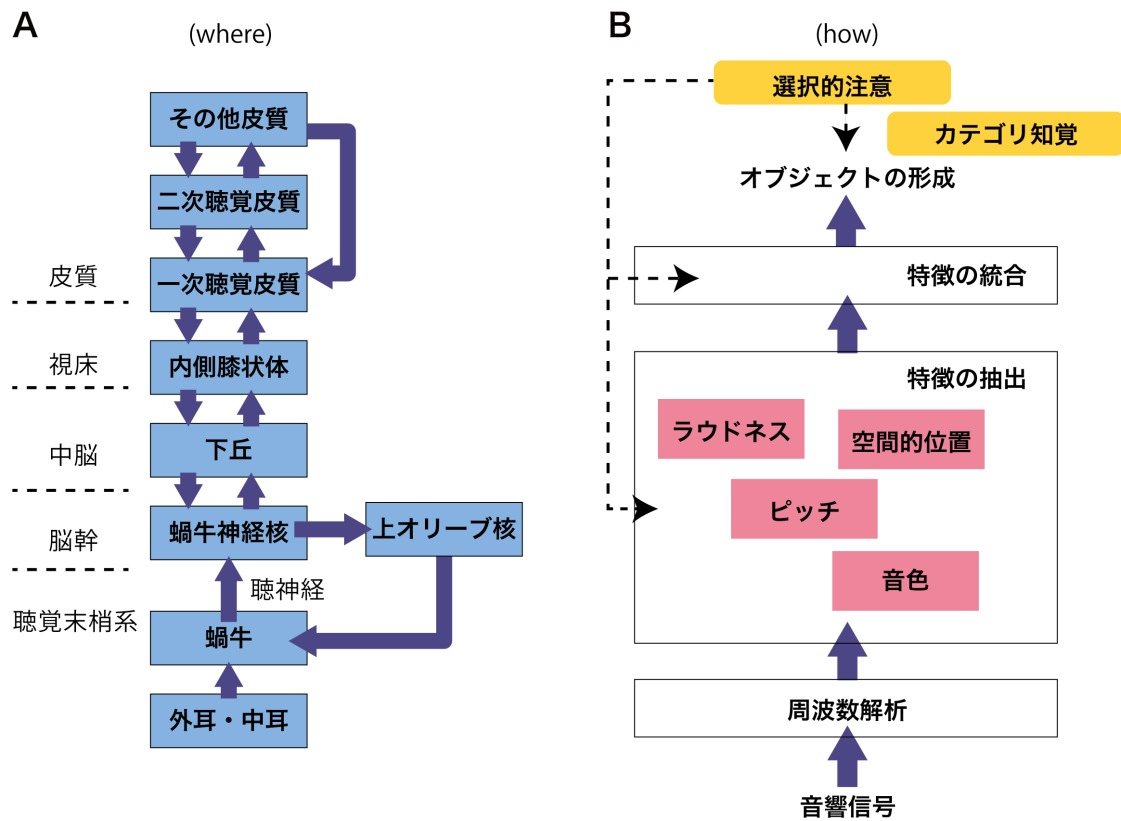


図 1.12: 聴覚情景分析の処理に関するふたつの問題. A: どこで処理は行われるのか (where). B: どのように処理は行われるのか (how).

(1) 聴覚情景分析の神経基盤について: 一次聴覚皮質よりも先のさらに高次の皮質や、聴覚関連領域ではない皮質領域は聴覚情景分析においてどのような役割を果たしているのか。また、皮質領域に至るより前の神経核では、どのような処理が行われているのか。ある知覚が起こったことをはっきりと表すような、神経的な信号は存在するなのか。

(2) 聴覚情景分析への選択的注意の影響について: 選択的注意のプロセスは、聴覚情景分析のどの段階から関与するのか。選択的注意は音の特徴のグルーピングする過程に対して、どの段階でどのような影響を与えるのか。

1.11 本研究の目的とアプローチ

本研究では、聴覚情景分析において重要な役割を果たす2つのプロセスの特性とメカニズムについて、心理物理学的手法を用いて検討した。心理物理学とは、人間の感覚系をひとつのシステムと見なし、物理的な刺激を入力した際にどのような反応が得られるかを定量的に記述する学問領域である。また、ヒトを対象とした実験では、ヒトに知覚の状態を直接尋ねられるという利点がある。動物を被験体とした研究では主観的な知覚については知ることが出来ない。研究1では、連続聴錯覚による補完のプロセスに関する検討を行った。研究2では、重畳する音から音を取り出す分離のプロセスと選択的注意プロセスの関連性に関する検討を行った。

研究1, 研究2ともに、これまで聴覚研究の分野で多く用いられてきた刺激よりも、やや複雑な特徴を持った刺激を用いて実験を行っている。多くの研究では、ABA 音列パターンのように、正弦波音や白色雑音で構成される非常に単純な音響特性の音刺激が用いられる。単純な音響特性の音を刺激とすることで実験条件の統制は容易となる。しかし、自然界で発生する音信号は複雑で変化に富んでいるため、単純な音刺激を用いた実験では聴覚情景分析の重要な側面を捉え損ねる可能性があると考えた。

1.11.1 研究1: 補完に関する研究

研究1では、連続聴錯覚 (1.5.1 節参照) を引き起こす音を刺激として用い、雑音によって隠された音の内容を補完して聴覚オブジェクトを作成する情報処理段階が、特徴抽出

の段階よりも先の処理段階に位置するか否かを調べた。過去の連続聴錯覚の研究では、連続聴錯覚の原因が聴覚末梢系での聴神経の発火パターンが分解される過程のみで説明可能であるという見方が示されていた (McAdams, Botte, & Drake, 1998)。しかし、音声や音楽のような複雑な音が被誘導音となって起こる連続聴錯覚では、背後の音を滑らかに繋げて聴くには、より高次の予測の過程が必要と考えられる (Bregman, 1990)。過去の研究の多くで用いられていた音刺激は、時間的な変化の少ない単純な正弦波音と雑音の組み合わせであったために、蝸牛基底膜の段階での処理のみで説明できた可能性がある。研究1では、連続聴錯覚の被誘導音 (雑音の背後で補完される音) として、周波数が一方向に連続的に変化するスイープ音 (周波数スイープ音) を用いた。そして周波数変調を検知する段階の反映として知られている、周波数変化残効 (2.1 節を参照) が、雑音を含む「知覚上連続的に聴こえるスイープ音」の入力を受けて起こるか否かについて調べた。

1.11.2 研究2: 分離に関する研究

研究2では、聴覚オブジェクトの形成処理と選択的注意の関係について、自然音を持つ統計的性質に似た性質を持つ雑音 (McDermott et al., 2011) を入力とすることで調べた。聴覚系が聴覚情景分析のために必要な「音源の特徴に関する事前知識」をどのようにして獲得するのかは過去議論の対象となってきた (1.4 節を参照)。研究2では、この事前知識を得るための機構である「繰り返し」手がかりに基づく音の分離機構が、音に対する選択的な注意を必要とするか否かについて、注意に関わるネットワークへの負荷の種類を変えながら検討した。

第2章 研究1：聴覚情景分析における補完のメカニズムの研究

2.1 はじめに

連続聴錯覚は聴覚の補完現象である (1.5.1 節参照)。連続聴錯覚はヒトのみならず、トリ、ネコおよびサルのような動物でも起こることが報告されている (Petkov et al., 2007; Sugita, 1997; C. T. Miller, Dibble, & Hauser, 2001)。こうした報告は、連続聴錯覚を引き起こす機能が聴覚の基礎的な機能として、種々の動物に実装されていることを示唆する。またこの現象は、マスキング可能性の法則を満たす場合、すなわち信号から雑音へ移り変わりが起こるときに、神経の発火レベルが減少しない場合のみ生じることが知られている。これらの事実、聴覚末梢系が連続聴効果を発現させる上で重要な役割を担うことを示唆する。

連続聴錯覚による補完が聴覚系のどこでどのように生じるかについては、いくつかの説がある (1.8.1 節参照)。ある説では、蝸牛において聴神経が表現する興奮パターンが分解される過程で連続聴錯覚が生じると説明する。具体的には、誘導音 (雑音バーストのような振幅の大きい音) によって生み出された神経の興奮が、被誘導音 (錯覚の連続音) とその残りの成分 (residue) に分解される過程で生じるという説である (総説は Bregman, 1990 を参照)。こうした「分解理論」による説明は連続聴が起こるとき誘導音の知覚上の大きさが減少するという観察結果と整合性がある (R. M. Warren, Bashford, Healy, & Brubaker, 1994; McAdams et al., 1998)。しかし、近年の研究において、音声などの複雑な信号が雑音の背後にある場合の連続聴錯覚では、連続的な知覚の原因が興奮パターンの分解のみでは説明できないことが指摘されている。Lyzenga らは振幅変調 (AM) 音、周波数変調 (FM) 音を被誘導音とした場合の連続聴効果を検討し、たとえ実験参加者が

被誘導音が連続的と知覚した場合であっても、AM と FM の位相変化の連続性を検知することは難しいことを明らかにした (Lyzenga, Carlyon, & Moore, 2005). この結果は、雑音の最中では周波数変調や振幅変調の詳しい特徴が神経表現として保存されていないことを示唆している。また、Darwin は、調波複合音と、正弦波音を被誘導音としたとき、連続聴錯覚が起こるか否かが、それらの成分が調波複合音の特定成分であるか否かに依存すること、すなわち、聴覚オブジェクトとしての解釈が妥当であるか否かが関係することを報告した (Darwin, 2005). さらに、誘導音のオンセットと同時に発生した視覚フラッシュが錯覚の連続性に影響することも示されている。 (Kobayashi, Osada, & Kashino, 2007). これらの結果は、音が複雑な特徴を持つ場合には、聴覚オブジェクトとしての解釈が行われる処理段階が錯覚の連続性の決定に関与することを示唆する。

本章で説明する研究 1 では、神経の興奮パターンの「分解」が連続聴錯覚のただ一つの原因ではないという説に対して、新たな証拠を付け加えるため、時間的な変化を持つ複雑な音を被誘導音とした連続聴錯覚について調べた。

この研究では、錯覚の周波数スイープ音の神経表現が周波数変化検出のような特徴抽出段階の前に作成されるか否かを評価するため、選択的順応パラダイムを用いた (Demany & Ramos, 2005). 選択的順応パラダイムは、特徴抽出器を研究するため視覚研究の領域で広く使用される (総説は Webster, 2011 を参照). 聴覚系についても、周波数スイープ音を繰り返し聴いた後、直後に聴いた周波数スイープ音の検知閾が上昇することが知られている (Gardner & Wilson, 1979). また、周波数スイープ音の直後に周波数変化のないトーンを呈示すると、このトーン音はあたかも順応したスイープ音の周波数変化方向と反対方向に上昇もしくは下降するかのように知覚される (Kayahara & Sato, 2001). 図 2.1) に周波数変化残効の模式図を示す。この順応現象は、周波数変化に敏感な神経機構によって起こると考えられている。実際、上方や下方へ変化する周波数変動に選択的に応答するニューロンは、蝸牛神経核から聴覚皮質までの様々な領域で発見されている (総説は Palmer, 1995 を参照)。

もしも錯覚の連続性が周波数変動検知の処理段階の前に生成される場合、「中心に空白のあるスイープ音」ではなく「実際のスイープ音」および、中央に雑音を持つことによって生じる「錯覚のスイープ音」の両方が周波数変化残効を引き起こす。一方、錯覚の連

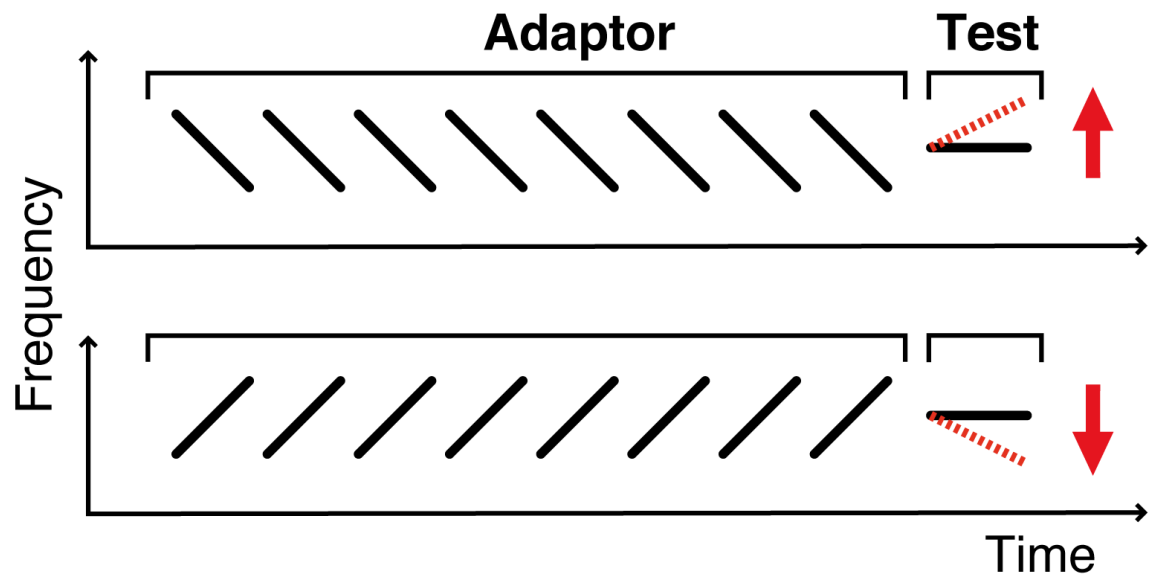


図 2.1: 周波数変化残効の模式図. 下降音に対する順応が起こった場合, 直後に呈示された音は上昇音として聴こえやすくなる. また, 上昇音に対する順応が起こった場合, 直後に呈示された音は下降音として聴こえやすくなる.

続性が周波数変化検出の段階より後に生成される場合「実際のスイープ音」のみが残効を引き起こすと考えられる.

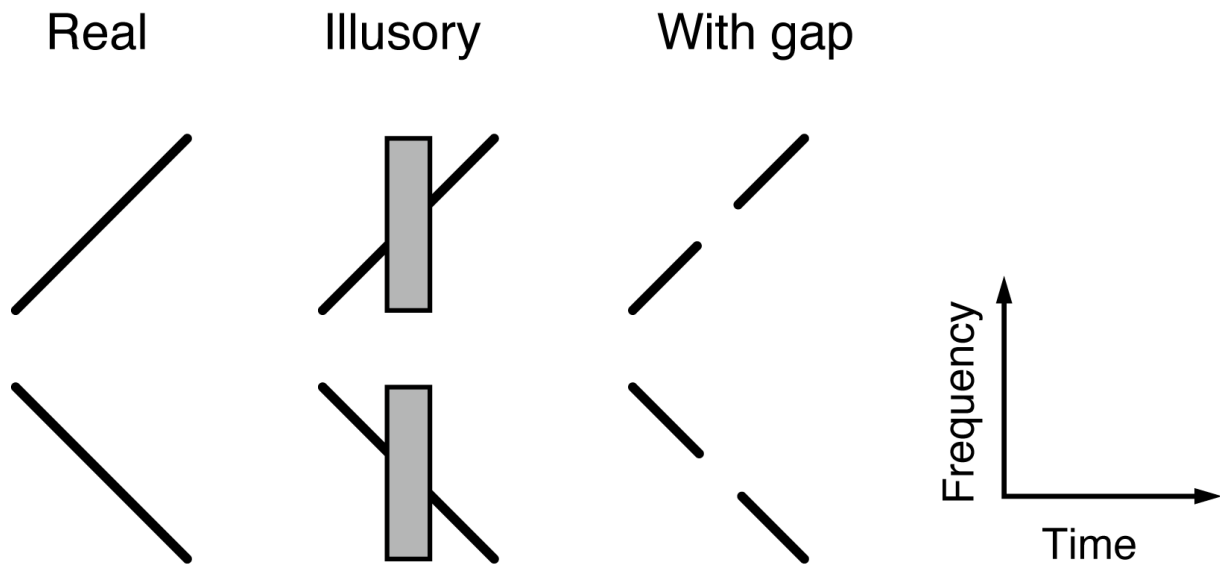


図 2.2: 順応刺激の模式図.

2.2 実験 1: 実際のスイープ音と錯覚のスイープ音の残効量比較

2.2.1 刺激

図 2.2 に、順応刺激の模式図を示す。実験では 3 種類の順応刺激を用意した。刺激は「実際のスイープ音 (Real)」(800 ミリ秒の時間長をもち、400 ミリ秒の時点で 1000 Hz を通過し、1/2 オクターブを横断する連続的なスイープ音)、「錯覚のスイープ音 (Illusory)」(中間の 200 ミリ秒の信号が帯域通過雑音と置換された「実際のスイープ音」)、そして、中間の 200 ミリ秒の信号が無音に置き換えられた「空白を持つスイープ音 (With gap)」であった。それぞれの順応刺激の種類について、上向きと下向きの両方を用意した。スイープ音の周波数は 800 ミリ秒の時間中、841 - 1189 Hz 間を対数比で変化した。

実験に用いた雑音サンプルは、841 - 1189 Hz の帯域を持つ 240 ms の帯域通過雑音であった。実験の開始前にあらかじめ 12 の雑音サンプルを作成し、うち 6 つを実験中の各試行において無作為に選び使用した。雑音は、白色雑音を高速フーリエ変換によって周波数領域での表現に変換した後、矩形の周波数領域の窓をかけ、時間領域へ再変換する

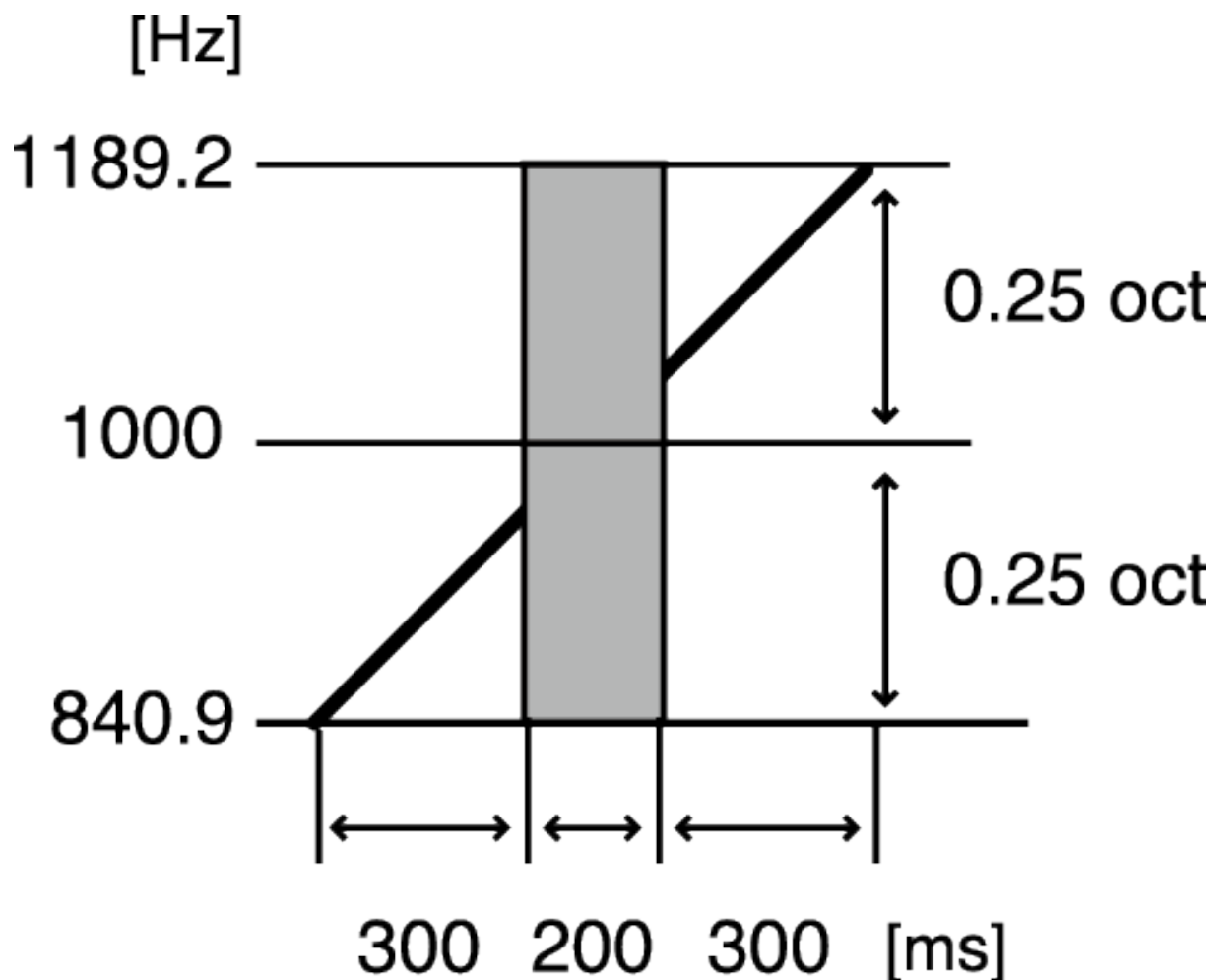


図 2.3: Illusory 刺激の詳細. 黒線は周波数スイープ音, 灰色の領域は帯域雑音を表す.

ことにより作成した. 雑音の始点と終点には 20 ミリ秒の長さのコサイン型の窓を掛け, 立ち上がりと立ち下がりをおやかにした. Illusory 音中の雑音とスイープ音は 20 ミリ秒の範囲でクロスフェードした.

検査刺激には, 800 ミリ秒の周波数スイープ音を使用した. 検査刺激の始点と終点には 20 ミリ秒のコサイン型の窓を掛けた. 検査刺激の中心部分 (400 ミリ秒時点) での周波数は 1000 Hz とした. 検査刺激の周波数スイープ音傾斜は 7 つの値を取った ($-15/96$, $-10/96$, $-5/96$, 0 , $5/96$, $10/96$, $15/96$ オクターブ毎秒). 負の値は下方向への変化, 正の値は上方向への変化を表す.

音刺激はすべてコンピュータ (MacBook, Apple Inc., USA) 上にて MATLAB 7.6 を使用して生成した. 作成と呈示の際のサンプリング周波数は 44100 Hz, 解像度は 16 ビット

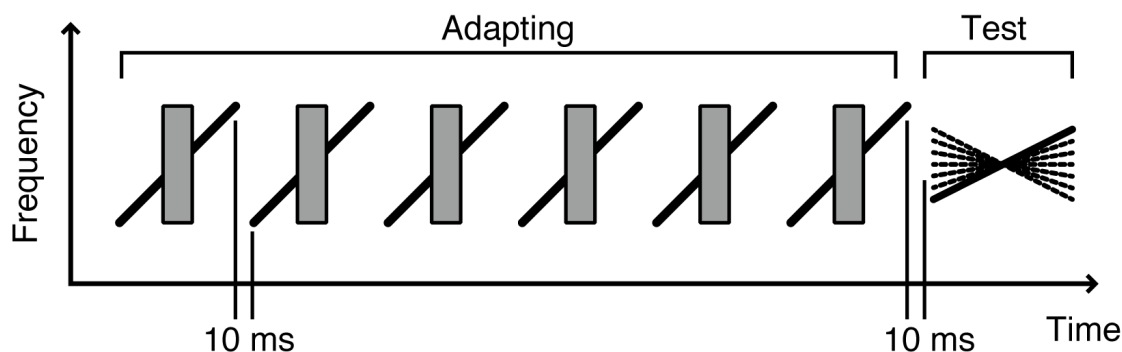


図 2.4: 1 試行の流れ. Adapting: 順応段階. Test: テスト段階. 参加者はこの後検査音が上向きに聴こえたか下向きに聴こえたかを回答した.

とした. 音刺激は, デジタル・オーディオ・インターフェース (EDIROL UA-20, Roland Corp., Japan) およびヘッドホン (HDA 200, Sennheiser Electronic Corp., Germany) を通して呈示した.

2.2.2 課題

実験では恒常法を用いて, 実験参加者が刺激への順応後にどの傾きを持つ検査刺激付近を「周波数変化のない平坦な音」として捉えたかを評価した. 各試行では, 実験参加者は, 10 ミリ秒 (順応段階; 図 2.4 を参照) の刺激間間隔を備えた順応刺激を 6 回聴いた. その後, 実験参加者は 7 つの検査刺激のうちひとつを聴き, 検査刺激の周波数変化の方向が上向きと下向きのどちらに聴こえたかを判断した (テスト段階).

順応刺激の種類および順応刺激の周波数変化の方向は 1 つのセッションの間固定した. 各実験参加者は 6 つのセッションを行った. 各試行について, 検査音の傾斜は, 7 つの値からランダムに選択された. 検査音は, 7 種類それぞれが 1 つのセッションの中で 25 回ずつ呈示された. 1 回のセッションには終了まで約 30 分を要した. 実験参加者は, 各セッションの開始前に練習として 3 回の試行を行った. スイープ音の音圧レベルは 57 dB, 雑音の音圧レベルは 70 dB に設定した. 順応音および検査音はすべて右耳に呈示した. 実験は防音室にて行われた.

2.2.3 実験参加者

12 名 (男性 8 名, 女性 4 名; 平均年齢 27.7 歳; 年齢の標準偏差 4.5) が実験に参加した。参加者は全員正常聴力 (500, 1000, 2000, 4000 Hz の純音について, 20 dB HL 以上の聴力) を持っていた。実験参加者は全員書面による実験の説明を受けた。また, 実験は NTT コミュニケーション科学基礎研究所の倫理委員会の承認を得て行った。

2.2.4 解析

各実験参加者ごと, 各順応刺激 (3 種類, 2 方向) ごと, 各検査刺激 (7 種類) ごとに, 実験参加者が検査刺激を上向きだと判断した回数の割合を算出した。その後, 心理測定関数としてロジスティック関数を当てはめた。曲線上にて値が 50% に相当するときの横軸の値を, 主観的な周波数変化が起こらないスイープ音の傾斜の値 (主観的無変化点) として定義した。1 つ以上の心理測定関数について, 傾きが著しく小さい曲線を描いていた 2 名の実験参加者のデータを分析から除外した。

それぞれの順応刺激による残効強度は, 下向きと上向きの順応刺激の主観的無変化点間の差として定義した。残効強度の値について, 順応刺激の種類を要因とする 1 要因の繰り返しのある分散分析を行った。解析の際には, 必要に応じて, Greenhouse-Geisser のイプシロンによる自由度の調整を行った。また, 残効強度が有意に 0 とは異なっていたかどうかを確かめるため, 各順応刺激の残効強度について, 1 サンプルの t 検定を行った。

2.2.5 結果と考察

図 2.5 に, ある実験参加者 1 名の結果を示す。図 2.6 は, 「実際のスイープ音 (Real)」, 「錯覚のスイープ音 (Illusory)」および「空白をもつスイープ音 (With gap)」それぞれによって生じた残効の強さの平均値である。分散分析の結果, 順応刺激の種類の主効果が見られた [$F(2, 18) = 11.13, p < 0.05$]。下位検定の結果, 「実際のスイープ音 (Real)」による残効の強度が, 他の 2 種類の刺激よりも有意に高いことが示された [$p < 0.05$, Shaffer の方法 (Shaffer, 1995) による]。また t 検定の結果, 「実際のスイープ音」による残効の

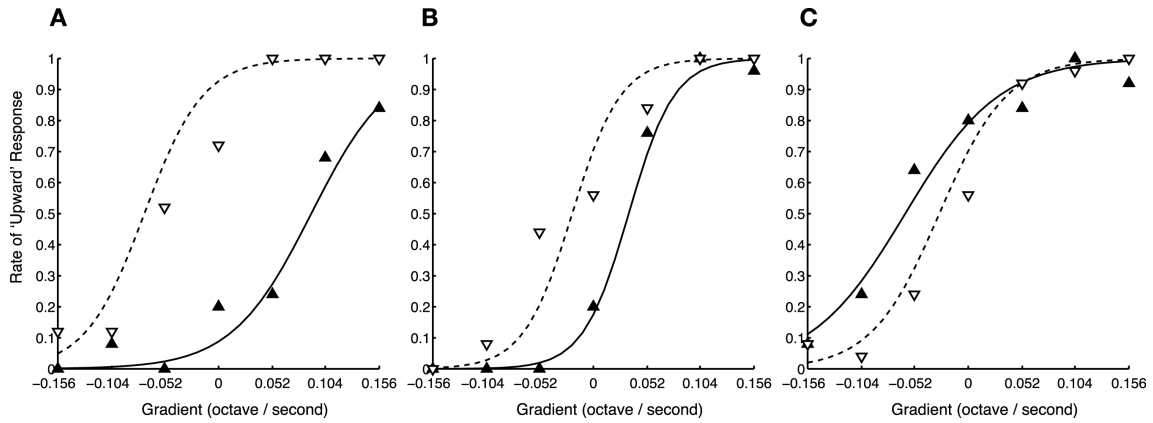


図 2.5: 実験参加者 1 名から得られた典型的な心理測定関数の例。横軸は各検査音の傾き、縦軸は試行中、ある検査音を「上昇音」と答えた試行の割合を示す。上向きの黒三角と実線で示すグラフは、順応音が上昇音であるときの結果、下向きの白三角と点線で示すグラフは、下降音の結果を表す。A: Real 順応音による結果。B: Illusory 順応音による結果。C: With gap 順応音による結果。

強度のみが、有意に 0 とは異なる値をとることが示された [Real: $t(9) = 5.02$, $p < 0.05$, Illusory: $t(9) = 0.95$, $p = 0.36$, With gap: $t(9) = -0.13$, $p = 0.89$].

実験の結果、順応刺激の種類によって残効の強度が大きく異なることが分かった。また、実験参加者が「実際のスイープ音」順応刺激の影響を明白に示すことが分かった。「錯覚のスイープ音」および「空白を持つスイープ音」刺激による残効の量は「実際のスイープ音」と比較して有意に小さく、また、0 と有意に異なるとはいえないということが示された。以上の結果から、錯覚の周波数スイープ音に対する順応があったとしても、周波数変化検出器に対する影響は、実際の周波数スイープ音への順応のように起こらないことが示唆された。

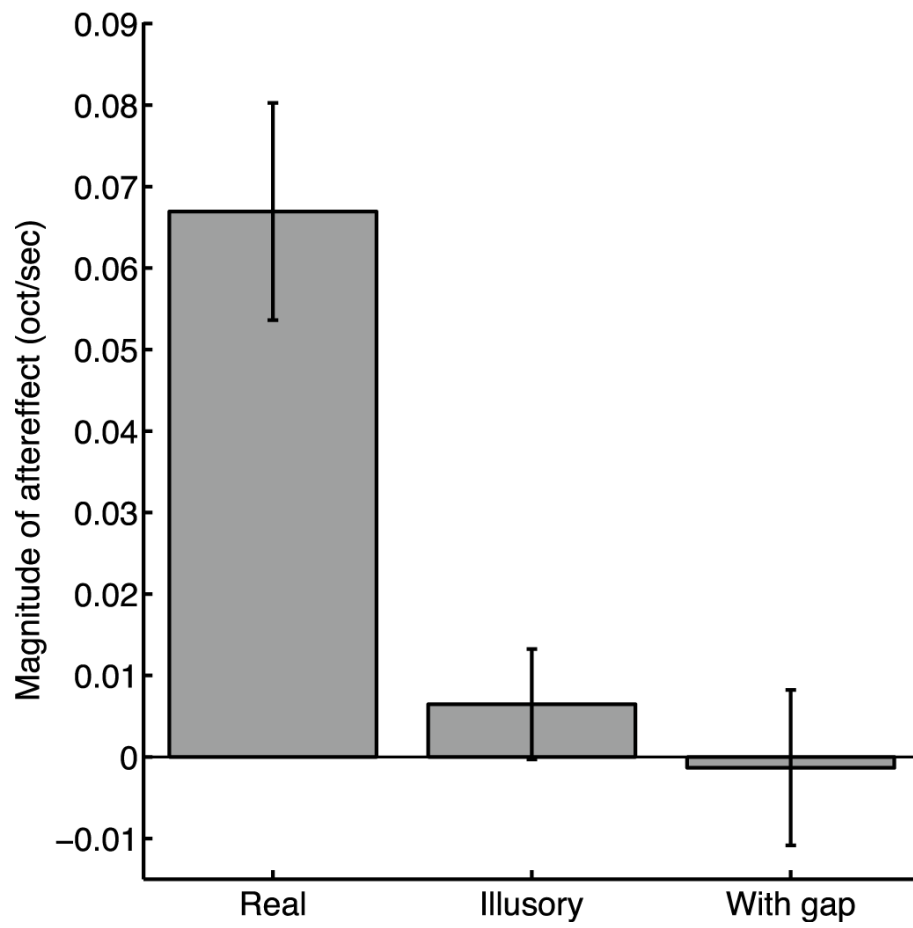


図 2.6: 実験 1: 周波数変化残効量の参加者平均 (9 名). エラーバーは標準誤差をあらわす.

2.3 実験2: ノッチ雑音を重畳したスイープ音の周波数変化

残効量の測定

実験1の結果は「錯覚のスイープ音」刺激に含まれていた雑音が周波数変化検出器に影響したためであるとも解釈できる。「錯覚のスイープ音」による残効の強度と、中間に連続性限界 (continuity limit) を越えた強度を持つ「実際のスイープ音」の残効強度とを比較することで、上記の可能性を検討できる。実験2では、「実際のスイープ音」の刺激にノッチ雑音を加えた刺激を作成し、実際のスイープ音との周波数変化残効量を比較した。

2.3.1 刺激

図2.7に順応刺激の概要を示す。実験では、6種類、2方向、計12種類の順応刺激を用意した。「実際のスイープ音 (Real)」, 「空白をもつスイープ音 (With gap)」は実験1と同様である。実験2では上記の刺激に加えて、実際のスイープ音にノッチ雑音を足し合わせた「実際のスイープ音+ノッチ雑音 (Real+NN)」, 中間の200ミリ秒の信号がノッチ雑音と置換された「実際のスイープ音」である、「空白のあるスイープ音とノッチ雑音 (With gap+NN)」 「実際のスイープ音」から「空白をもつスイープ音」を切り取った「ピップ (Pip)」 「ピップ」にノッチ雑音を加えた「ピップ+ノッチ雑音 (Pip+NN)」を用意した。スイープ音の周波数は800ミリ秒の間、841 - 1189 Hzの間を対数比で変化した。図2.8に With gap+NN の模式図を示す。

実験に用いた雑音サンプルは、841 - 978.6 Hz, また、1021.9 - 1189 Hzの帯域間に信号を持つように作成した240ミリ秒のノッチ雑音であった。実験の開始前にあらかじめ12の雑音サンプルを作成し、うち6つを各試行にて無作為に選び呈示した。雑音は、広帯域白色雑音を高速フーリエ変換によって周波数領域での表現に変換した後、2つの矩形の周波数領域の窓をかけ、時間領域へ再変換することにより作成した。雑音の始点・終点には、20ミリ秒分のコサイン型の窓を掛け、立ち上がり立ち下がりをつなげた。雑音とスイープ音は20ミリ秒の範囲でクロスフェードした。

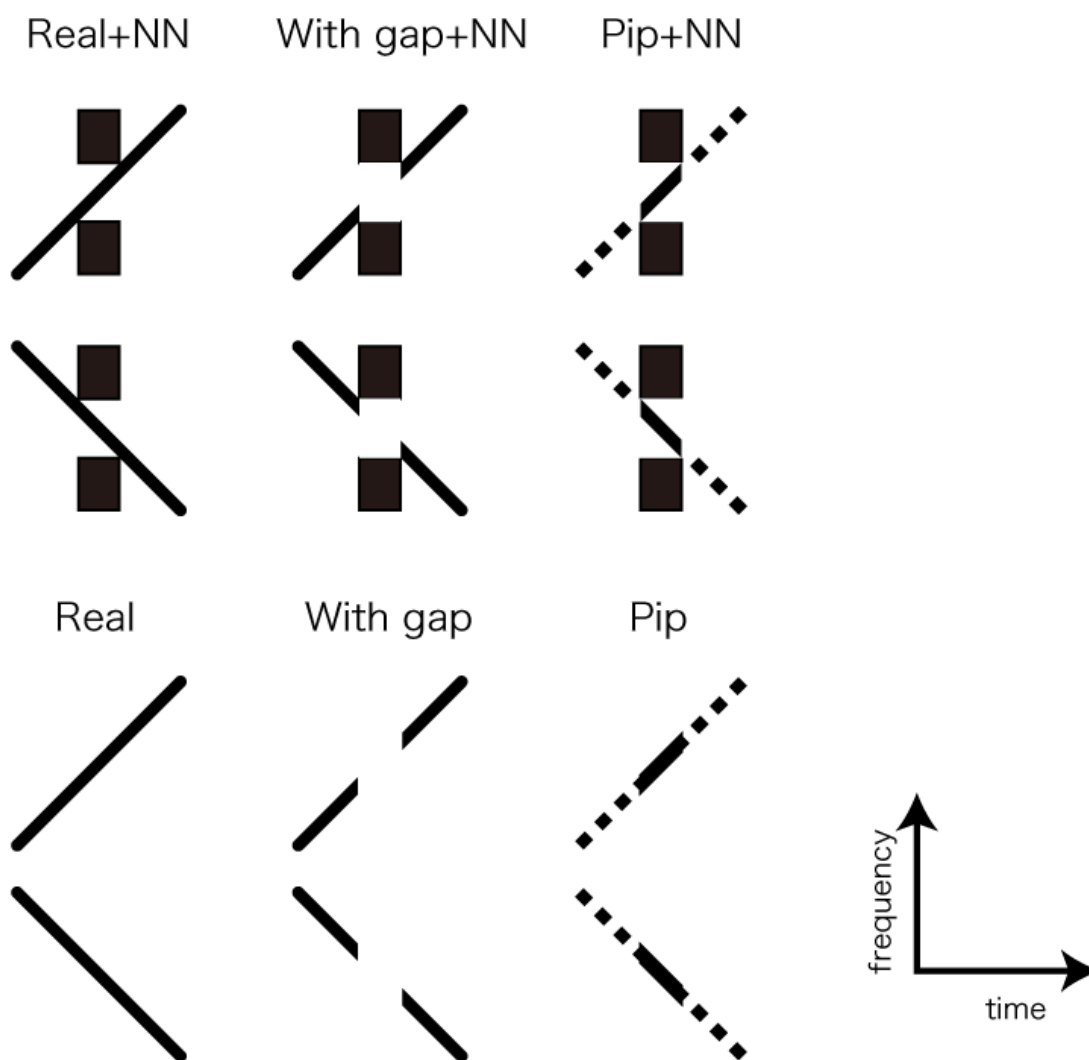


図 2.7: 実験 2: 順応刺激の種類. 黒線はスイープ音, 黒色の領域は雑音の成分が存在する領域を表す. 点線部分は無音区間をあらわす.

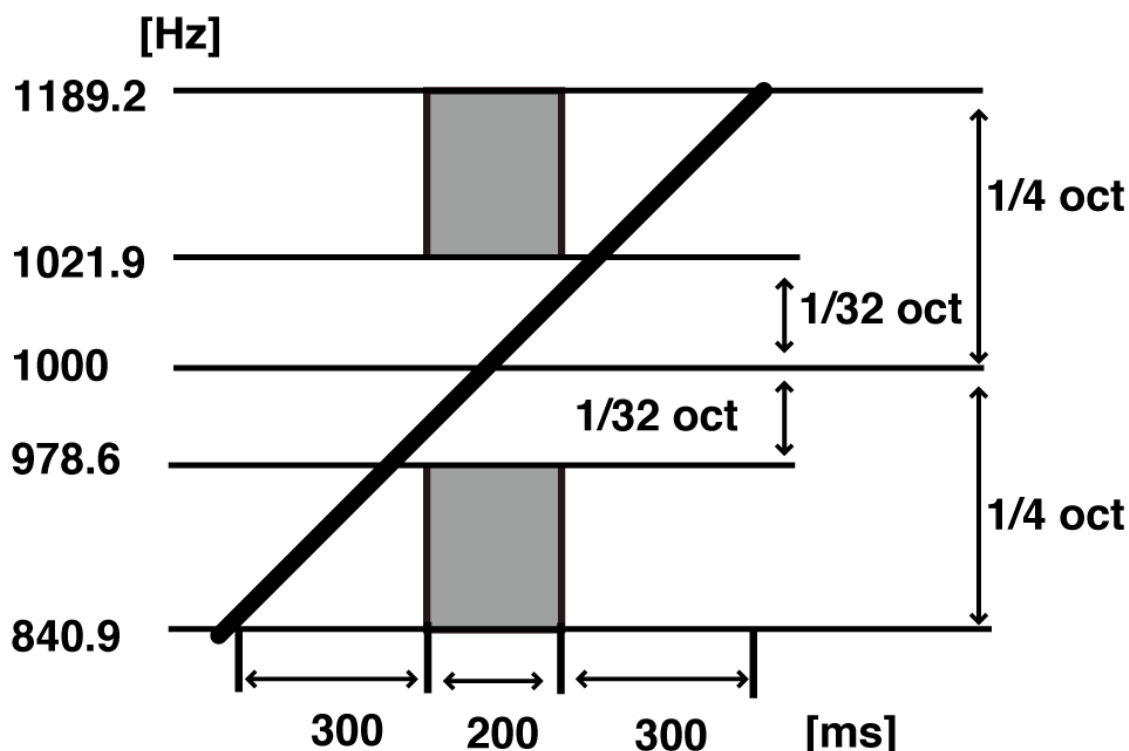


図 2.8: Real+NN (実際のスイープ音にノッチ雑音を加えた刺激) の模式図. 黒線はスイープ音, 灰色の領域は雑音の成分が存在する領域を表す.

検査刺激は, 実験 1 と同様に 800 ミリ秒の周波数スイープ音とした. 検査刺激の始点と終点にはコサイン型の窓を掛けた. 検査刺激の中心部分 (400 ミリ秒時点) での周波数は 1000 Hz に固定した. 検査刺激の周波数スイープ音傾斜は 9 つの値を取った ($-4/76$, $-3/76$, $-2/76$, $-1/76$, 0 , $1/76$, $2/76$, $3/76$, $4/76$ オクターブ毎秒). 負の値は下方向への変化, 正の値は上方への変化を表す. その他の条件は実験 1 と同様とした.

2.3.2 実験参加者

4 名 (男性 4 名; 平均年齢 28.8 歳; 年齢の標準偏差 4.8) が実験に参加した. 参加者は全員正常聴力 (500, 1000, 2000, 4000 Hz の純音について, 20 dB HL 以上の聴力) を持っていた. 実験参加者は全員書面による実験の説明を受けた. また, 実験は NTT コミュニケーション科学基礎研究所の倫理委員会の承認を得て行った.

2.3.3 課題

順応刺激の種類および周波数変化の方向は1つのセッションの間固定した。各試行について、検査音の傾斜は、9つの値からランダムに選択された。1つの順応刺激について、2つのセッションを用意した。1セッションは90試行で構成された。1回のセッションには終了まで約20分を要した。検査音9種類は1つのセッションの中で各10回、2つのセッションを合わせて、各20回呈示された。実験参加者は、各セッションの開始前に練習として3回の試行を行った。スイープ音の音圧レベルは57 dB、ノッチ雑音の音圧レベルは70 dBに設定した。順応音および検査音はすべて右耳に呈示した。

2.3.4 解析

各実験参加者ごと、各順応刺激(6種類、2方向)ごと、各検査刺激(9種類)ごとに、実験参加者が検査刺激を上向きだと判断した回数の割合を算出した。その後、心理測定関数としてロジスティック関数を当てはめた。曲線上の値が50%に相当するときの横軸の値を主観的無変化点として定義した。

それぞれの順応刺激による残効強度は、下向きと上向きの順応刺激の主観的無変化点間の差として定義した。残効強度の値について、9名の参加者の順応刺激の種類を要因とする、繰り返しのある分散分析を行った。解析の際には、必要に応じてGreenhouse-Geisserのイプシロンによる自由度の調整を行った。また、残効強度が有意に0とは異なっていたかどうかを確かめるため、各順応刺激の残効強度について1サンプルの t 検定を行った。

2.3.5 結果と考察

図2.9に、各順応刺激によって生じた周波数変化残効量の参加者平均値を示す。分散分析の結果、順応刺激の種類の主効果は見られなかった [$F(1, 3) = 3.6, p = 0.11$]。また t 検定の結果、「実際のスイープ音 (Real)」と「空白のあるスイープ音+ノッチ雑音 (With gap+NN)」による残効の強度とが、有意に0とは異なる値をとることが示された [Real+NN: $t(3) = 3.28, p < 0.05$, With gap+NN: $t(3) = 0.79, p = 0.49$, Pip+NN: $t(3)$

$= 0.87, p = 0.45$, Real: $t(3) = 3.36, p < 0.05$, With gap: $t(3) = 1.46, p = 0.24$, Pip: $t(3) = 1.67, p = 0.2$].

実験2の結果、実験1と同様に、順応刺激の種類によって残効の強度が大きく異なることが分かった。また、実験参加者が「実際のスイープ音」順応刺激による影響を受けることが分かった。一方でその他の刺激による残効の量は「実際のスイープ音」と比較して有意に小さく、また、0と有意に異なるとはいえないということが示された。

サンプル数の少なさのため統計的有意差はみられないものの、グラフからは Real+NN 順応音と With gap+NN 順応音の残効量は Real とは異なる傾向がみられた。この予備的実験の解釈からは、連続聴錯覚が生じているか否かではなく、音圧の高い雑音が存在していることが周波数変化検出器の動作を妨げている可能性を否定できない。また、ノッチ雑音の間の領域が、聴覚末梢系の周波数解析フィルタの重なりと比較して狭かったため、末梢の神経発火表現の段階で実際のスイープ音にノッチ雑音を足し合わせた刺激と、中間部を切り取ったスイープ音にノッチ雑音を足し合わせた刺激の表現には差が存在しなかった可能性がある。

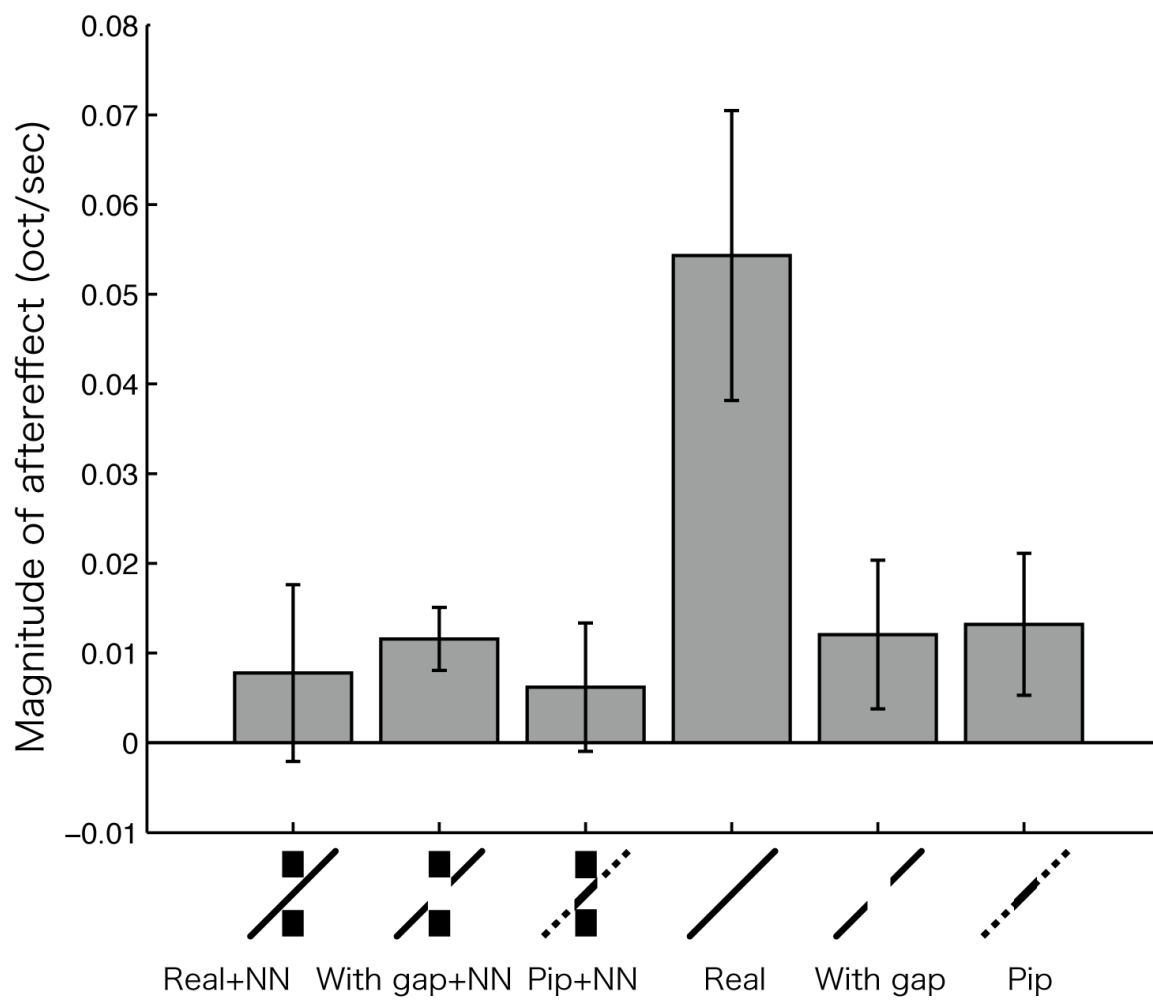


図 2.9: 実験 2: 周波数変化残効量の参加者平均 (4 名). エラーバーは標準誤差をあらわす.

2.4 研究1: 総合考察

実験1の結果、順応刺激の種類によって周波数変化残効の強度が大きく異なることが分かった。また、実験参加者が「実際のスイープ音」順応刺激の影響を明白に示すことが分かった。「錯覚のスイープ音」および「空白を持つスイープ音」の刺激による残効量は「実際のスイープ音」と比較して有意に小さく、また、0と有意に異なるとはいえないことが示された。以上の結果から、錯覚の周波数スイープ音に対する順応があったとしても、仮想的な周波数変化検出器に対する影響は、実際の周波数スイープ音に対する順応があった場合のように起こらないことが示唆された。

実験2での予備的検討からは、雑音の存在そのものが周波数変化残効を起こりにくくする可能性が示唆された。しかし、いずれにせよ今回の発見は「神経の興奮パターンが周波数スイープ音の連続聴錯覚を構成する」という説明が十分ではないという見解を支持する。基底膜振動情報の「分解」が連続聴錯覚の原因であるとする説によれば、錯覚の連続音に対応する神経の興奮パターンと、実際の連続音によるパターンは基底膜での神経表現の時点で同一の信号であり、「錯覚のスイープ音」に対する順応の結果は、実際のスイープ音に対する順応の結果と同一になるためである。

最近普及している説では、連続聴効果は自動的に起こり、聴取者が集中して音を聴き取ろうとしなくても生じていると考えられている(1.9節参照)。連続聴錯覚の構成には中央のレベルの処理(末梢のフィルタリングおよび局所的な特徴検出段階より後であり、認識や注意に関係する段階よりも前の処理)が重大な役割を担うようである。

今回の研究にて仮定した周波数変化残効の検出段階が、聴覚系の情報処理経路のどこに位置するのかについては考察の必要がある。また、この検出段階での処理がオブジェクトとして独立したスイープ音の知覚に対して適用されるのか、あるいはスイープ音のオブジェクトと雑音のオブジェクトの入力を同時に受け、同時に処理するのかについては不明である。後者の場合には研究2の結果が、雑音のオブジェクトによる妨害によって残効がリセットされたためと説明できる。

近年の神経生理学的研究(Petkov et al., 2007) および脳神経イメージングの研究(Riecke, van Opstal, Goebel, & Formisano, 2007; Riecke et al., 2009) から、連続聴錯覚の構成に

は聴覚皮質が関与することが示されている (1.8.1 節参照). 聴覚皮質は聴覚オブジェクトの構成に重要な役割を果たすことが示唆されている (Micheyl et al., 2005). こうした研究もまた, 連続聴錯覚の生成の際には, 聴覚オブジェクトとしての解釈を行うための処理段階の関与が重要であり, 錯覚の連続性を知覚するために, 末梢系の段階で「分解」することによってオブジェクトに相当する実効的な信号を生み出す必要はないという説に, 新たな証拠を付け加える.

第3章 研究2：聴覚情景分析における分離のメカニズムの研究

3.1 はじめに

聴覚系は、音響信号が一般的に持つ物理的な特性についての知識を有しており、この知識を利用して音の分離に役立っていることが知られている (1.4 節参照). これらの手がかりは「ボトムアップ手がかり」と呼ばれる. さらに聴覚系は音声の言語的知識や、音楽の規則の知識などの「トップダウン手がかり」を合わせて、混合音から特定の音を分離していると考えられる. しかし、聴覚系がこうした手がかりを生得的に持っているのか、後天的な学習によって得ているのかについては明らかになっていなかった. たとえ聴覚系が後天的な学習によって事前知識を得るとしても、多くの場合耳に届く音は既に混合音であるため、聴覚系が混合音から「分離された」状態である特定の音を聴く機会是非常に稀である. 聴覚系がいかにしてある特定の音の音響特性を学習するかについては不明であった.

近年、McDermott らは「たとえ全く未知の音であっても、混合音中でその音が繰り返し現れるならば、人間はその音をひとつのまとまりとして検出することが可能である」ことを発見した (McDermott et al., 2011). 刺激音は、一般に知られている手がかりが非常に少ない (Carlyon et al., 2001) 人工音であった. 実験参加者は、ある音 (ターゲット音) と妨害音との混合音が妨害音の種類を変えながら2回以上繰り返し呈示されるとき、チャンスレベル以上の精度でターゲット音を検出することができた. この結果は「音が繰り返しあらわれる」という条件そのものが、混合音から音を分離するための手がかりとして利用可能であることを示唆する.

McDermott らの発見は、聴覚系が時間-周波数パターンをある程度蓄積し、加算平均

することによって、一貫性を保って数度出現する時間・周波数パターンをひとつの知覚的なまとまりとして分離する機構を備えていると考えることで説明できる。この機構はボトムアップ手がかりやトップダウン手がかりなどの事前知識が得られない状況でも、聴覚オブジェクトを生成する有効な手法として働くと考えられる。

研究2の一連の実験では、この「繰り返し手がかりに基づく音の分離」が音への選択的注意によって影響を受けるか、あるいは注意や集中による影響に関係なく可能であるか否かを調べた。実験では McDermott らによる実験を二重課題デザインに応用し、実験参加者らが音を聞くことに集中していなくとも「繰り返し」手がかりに基づく音の分離が可能であるかを検討した。

選択的注意に関係する処理系は分散しており、課題の内容によって負荷が与えられる領域は変化すると考えられる(1.9.3節参照)。そこで、研究2では注意課題を複数用意し、負荷を与える注意機能の内容を変化させた。実験1では、注意系への負荷が視覚入力を通じて音響的短期記憶(エコイックメモリ)に負荷を与える際の聴覚課題の成績を調べた。実験2では、実験1にて参加者が偽ターゲット音の非常に単純な音響的要素のみに基づいて回答していたか否かを検討した。実験3では、エコイックメモリが必要とされないが集中を必要とする課題を同時に行った際の聴覚課題の成績を調べた。実験4では、突発的な音を検出する聴覚課題と共に、聴覚課題を行ったときの能力を調べた。注意の負荷量の操作は、異なるブロックでそれぞれ「低負荷」および「高負荷」課題を行なうよう参加者に指示することで行われた。参加者はおとり課題に集中するよう指示を受け、聴覚課題について自信を持って回答することが難しい場合、ランダムに回答するように指示された。

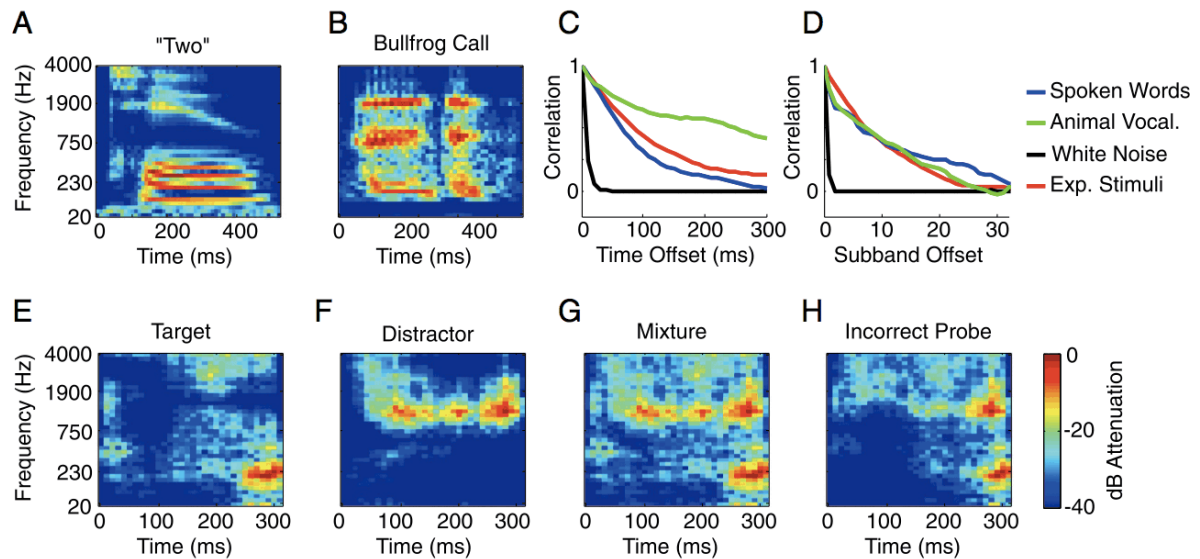


図 3.1: 研究 2 にて用いた音刺激. A, B: 音声 (A) と カエルの鳴き声 (B) のスペクトログラムの例. 周波数の調波構造や, 成分の立ち上がりと立ち下りの同期性が見られる. C, D: 時間軸上の距離の関数 (C) と周波数軸上の距離の関数 (D) として描いた時間・周波数スペクトログラムのセルの値の相関. E, F: 実験に用いた刺激のスペクトログラムの一例. G: E と F の刺激を加算して作成した「混合音」のスペクトログラム. H: G の混合音から一部成分を抽出して作成した偽ターゲット音の一例. McDermott et al., (2011) Figure 1 の一部を抜粋.

3.2 メイン聴覚課題

3.2.1 刺激

音刺激の作成には, McDermott らと同様の手続きを用いた (McDermott et al., 2011). 図 3.1 に刺激の概要を示す. 実験参加者がボトムアップ手がかりを利用して音を分離することを避けるため, 全ての音は白色雑音を基礎として作成した. ただし, ある程度 of 自然音の特徴を刺激に反映させるため, 刺激には自然音の持つ時間相関, 周波数相関の性質を組み込んだ. 一般に, 自然音には時間-周波数スペクトログラム上のある 2 点について, 点同士の距離が短いほど振幅値の相関が高く, 距離が離れるにつれて低下する傾向

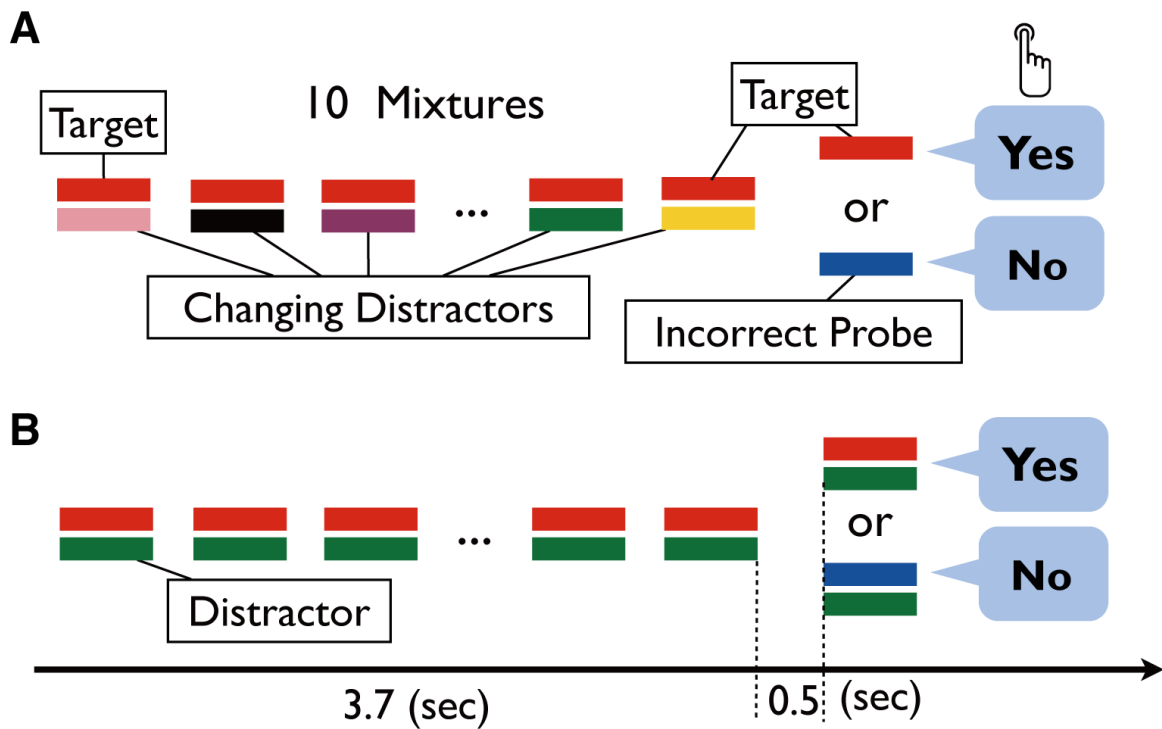


図 3.2: メイン聴覚課題の模式図。 図中の線分は刺激を表す。 時間的に同時に呈示された 2 つの刺激は、垂直方向に重なった線分として、また異なる刺激は異なる色で表す。 A: 分離課題 (Sg). B: 弁別課題 (Ig).

がみられる (Voss & Clarke, 1975; Attias & Schreiner, 1997; N. C. Singh & Theunissen, 2003). 刺激音の作成にあたっては、この関係性を近似した単調減少の時間相関関数と周波数相関関数を用い、スペクトログラム上の各分割セル間の共分散行列を作成した。この共分散行列をもとにランダムガウス分布を生成し、白色雑音のサンプルにかけあわせた。

参加者は、1つの実験ブロック中で、2種類の聴覚課題のうちの1つを行った。(1) 分離課題 (以後 Sg と表記) : McDermott らによる研究で提案されたものと同内容の課題であり、混合音中に繰り返し出現する「ターゲット音」を検出する課題。参加者はターゲット音と妨害音との混合音を 10 回聴いた後、検査音をひとつ聴き、検査音がターゲット音であったか否かを Yes・No のボタンで回答した。10 回の混合音のターゲット音は常に同じであった一方、妨害音はそれぞれ異なる音であった。検査音はターゲット音あるいは

ターゲット音と妨害音の混合音から、ターゲット音の音圧レベルを越えない範囲で成分を抜き出して作成した「偽ターゲット音」のいずれかであった。参加者がターゲット音についての事前知識を得ることを防ぐため、ターゲット音は実験を通して1度のみ呈示された。

もう一方の課題は (2) 弁別課題 (以後 Id と表記) であった (図 3.2B)。Id 課題は、Sg 課題に限らず、同時に行われる聴覚課題の成績であれば注意の影響を受けるか、あるいは Sg 課題のみが注意の影響を受けるかを確かめるためのコントロール条件として用意した。Id 課題は Sg 課題と似た構造となっているが、Sg 課題とは異なり音の分離を行う必要はない。Id 課題では、あるターゲット音とある妨害音足して混合音を作成し、このひとつの混合音のみを繰り返し 10 回呈示した後、検査音として、最初に呈示したものと同じ混合物あるいは異なる妨害音から生成された偽ターゲット音と妨害音を足した音のいずれかを呈示した。参加者は、繰り返し再生された音が検査音として現れたか否かのみを答えることを要求された。音刺激は、実験 1, 2 および 3 では両耳から呈示され、実験 4 では片耳からのみ呈示された。

3.2.2 おとり課題

参加者は、聴覚課題と平行して、脳の計算資源を奪う「おとり」課題を行った。「おとり」課題は2つの負荷量で用意された。(1) 高負荷課題 (High-load, 以下 Hi と表記) は、強い集中および特定の認識能力を要求する課題であり (2) 低負荷課題 (Low-load, 以下 Lo と表記) は、刺激は Hi 課題と同じであるが、最低限の資源しか必要としない課題とした。刺激の内容は2つの条件で全く同じであり、負荷の量は、参加者に与えられる課題の内容の違いによって調整された。聴覚課題とおとり課題の組み合わせは、1回の実験ブロック内で固定された。

Hi 課題の負荷が十分に軽く、メイン聴覚課題に参加者が集中する事態を避けるため、また、負荷が重すぎて参加者が課題の遂行を諦める事態を避けるため、あらかじめ Hi 課題の難易度を調整する目的で予備実験を行った。各予備実験では、参加者は、同時に聴覚課題を行うことなく数段階の難易度の Hi-load 課題に取り組んだ。予備実験の参加者

は、本実験の参加者とは異なっていたが、一部数名が両方の実験に参加した。各予備実験の詳細は各実験セッション内にて述べる。

メイン聴覚課題の成績は、各参加者の d' スコア (Green & Swets, 1966) を計算することによって定量化した。検査音がターゲットであった場合に「はい」と解答したとき‘検出’を記録し、疑似ターゲット音に「はい」と解答したとき誤検出を記録した。 d' スコアはおとり課題の成績評価でも使用した。ただし、MOT 課題を行った実験3では、正答率をおとり課題の遂行指標として利用した。

おとり課題の負荷量が Sg 課題の結果を下げるかどうかを調べるため、主に Hi 課題を行っている際の Sg 課題のパフォーマンスと Lo 課題を行っている際の Sg 課題のパフォーマンスの差に焦点を当てた。もしも参加者の、おとり課題のない、Sg 課題単体での成績が非常に低い場合には、Sg-Hi と Sg-Lo との差を見ることは困難となる。この測定上の制限を回避するため、本実験に参加するために必要とされる実験参加者の成績の基準を設定し、事前にすべての実験参加候補者について Sg 課題のパフォーマンスを測定した。

参加の d' のしきい値は、Sg 課題については d' が1以上であること、Hi 課題については d' が2.5以上であることとした。Sg 課題については、20 試行を含む8つの実験ブロックを繰り返したのち、最後の40 試行について d' が1を超えた参加者のみが本課題に進んだ。事前実験の参加者と、本課題へ通過した参加者の人数については、各実験の節で述べる。

音刺激はコンピュータ (MacBook, Apple Inc., USA) 上で MATLAB 7.6(The Mathworks, Natick, MA) を使用して生成した。作成・呈示の際のサンプリング周波数は24000 Hz、解像度は16ビットとした。音刺激は、デジタル・オーディオ・インターフェース (EDIROL UA-25EX, Roland Corp., Japan) およびヘッドホン (HD555, Sennheiser Electronic Corporation, Germany) を通して、60 dB から 70 dB の間の、参加者が課題を行いやすい音圧レベルで呈示した。参加者は実験の前に刺激のデモを聴き、音を聞き取りやすい音圧レベルに調整した。実験は防音室で行われた。統計解析は、R 2.13.1 と MATLAB Statistics Toolbox 上の解析パッケージを使用して行った。統計解析の結果については、必要に応じて、Greenhouse-Geisser のイプシロンによる自由度の調整を行った。

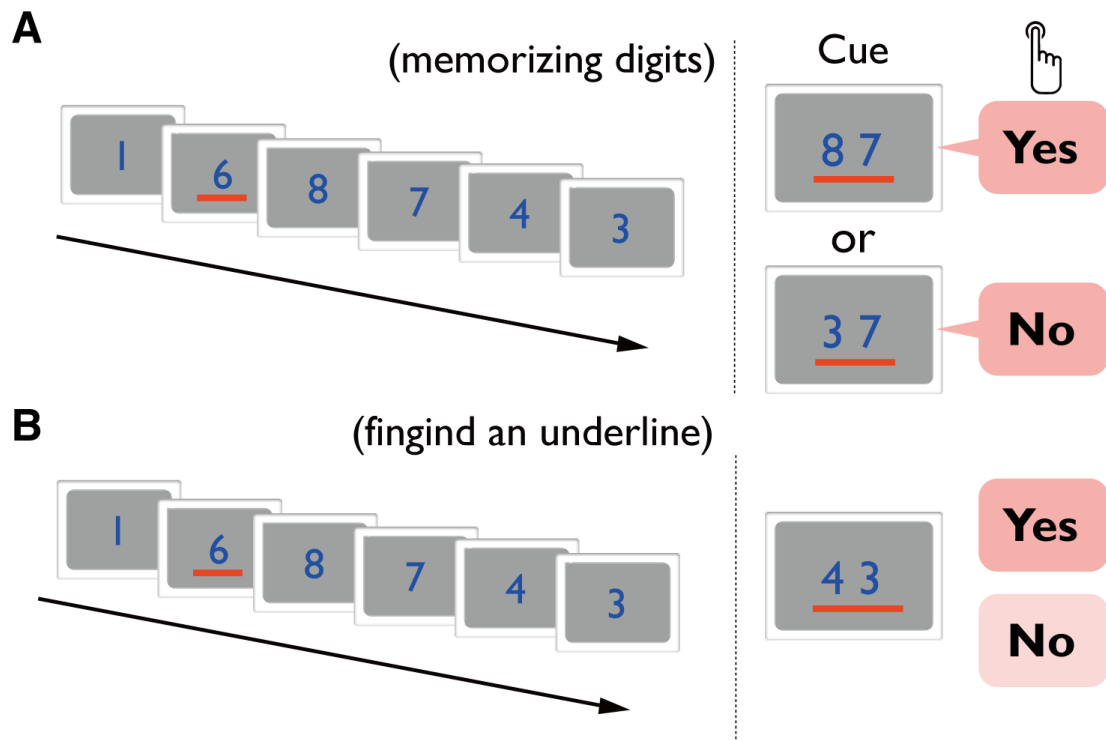


図 3.3: 実験 1 のおとり課題 (記憶課題) の模式図。灰色の四角はコンピュータの画面を表す。A: 高負荷 (Hi) 課題。参加者は画面中央にあらわれる数字の列を記憶し、赤い下線は無視するよう指示された。B: 低負荷 (Lo) 課題。参加者は数字を無視し、赤い下線が出現したか否かのみを答えるように指示された。

3.3 実験 1: 視覚入力による記憶・集中課題

実験 1 では、メイン聴覚課題を実行している間、参加者は、おとり課題として連続的に提示される視覚的刺激列を使用した。このような種類の課題では、視覚からの入力を通して、負荷が聴覚情報処理からの干渉を受けることなくエコイックメモリに負荷を与えと考えられる (Lavie, 2005)。

3.3.1 手法

参加者は、課題の [Sg-Lo Sg-Hi, Id-Lo, Id-Hi] の 2 回ずつ、8 実験のブロックの合計のすべての組み合わせに取り組んだ。1 ブロックは 40 試行で構成され、完了するまで約 8 分を要した。メイン聴覚課題の刺激が視覚刺激の提示と同時に提示された。参加者はまずメイン聴覚課題のプローブ音を聞いて解答し、その後おとり課題の解答のために画面を見て、おとり課題に答えた。実験時間は約 150 分であった。

図 3.3 はおとり課題の概略図である。両方の課題とも、1 から 8 の数字のうちの 하나가、順次画面の中央にあらわれた。数列が含む数字の数は、後述する事前テストにて測定された各参加者の成績に応じて、6 桁または 5 桁に設定した。一連の数字は 3.2 秒間呈示された。これは、メイン聴覚刺激の 10 の混合音を再生するための時間長である。各数字は 6 桁の条件ではひとつあたり 0.53 秒、5 桁の場合は 0.62 秒呈示された。半数の試行には、呈示される数字のうちひとつと同時に赤い下線が呈示された。数字と下線の呈示タイミングは各試行でランダム化した。参加者は、数字や下線が呈示されない間は、画面中央の注視点をを見るように指示された。Hi 条件では、参加者はメイン聴覚課題に答えた後、一連の数字を暗記するように求められた。また、試行を通じて、下線が呈示された場合には無視するように指示された。その後、参加者は画面上に表示された 2 桁の数字を確認し、それが記憶した 5 または 6 桁の数字列の一部であるかを「はい」または「いいえ」の応答ボタンを使用して解答した。Lo 条件では、それらは下線が「はい」または「いいえ」の応答ボタンを使って、数字のひとつの中で登場するかどうかを報告するように求められた。

事前に Hi 条件における負荷の適切な量を決定するために、予備実験を行った。8 名（平均年齢 26.3 歳、女性 6 名）が予備実験に参加した。各条件 [4, 5, 6, 7, 8 桁] は、別々のブロック内で呈示された。また、ブロックの順序は参加者間でランダム化した。各条件について 2 回のブロックが用意された。ひとつのブロックには 20 試行が含まれた。結果は図 3.4 に示す。このグラフの結果から、成績の天井効果がみられなくなる値である、6 桁の条件を本実験にて適用した。また、実験参加者の成績のばらつきも考慮し、6 桁の条件にて測定された d' の成績が 2.5 を超えない参加者については、5 桁の条件を適用した。

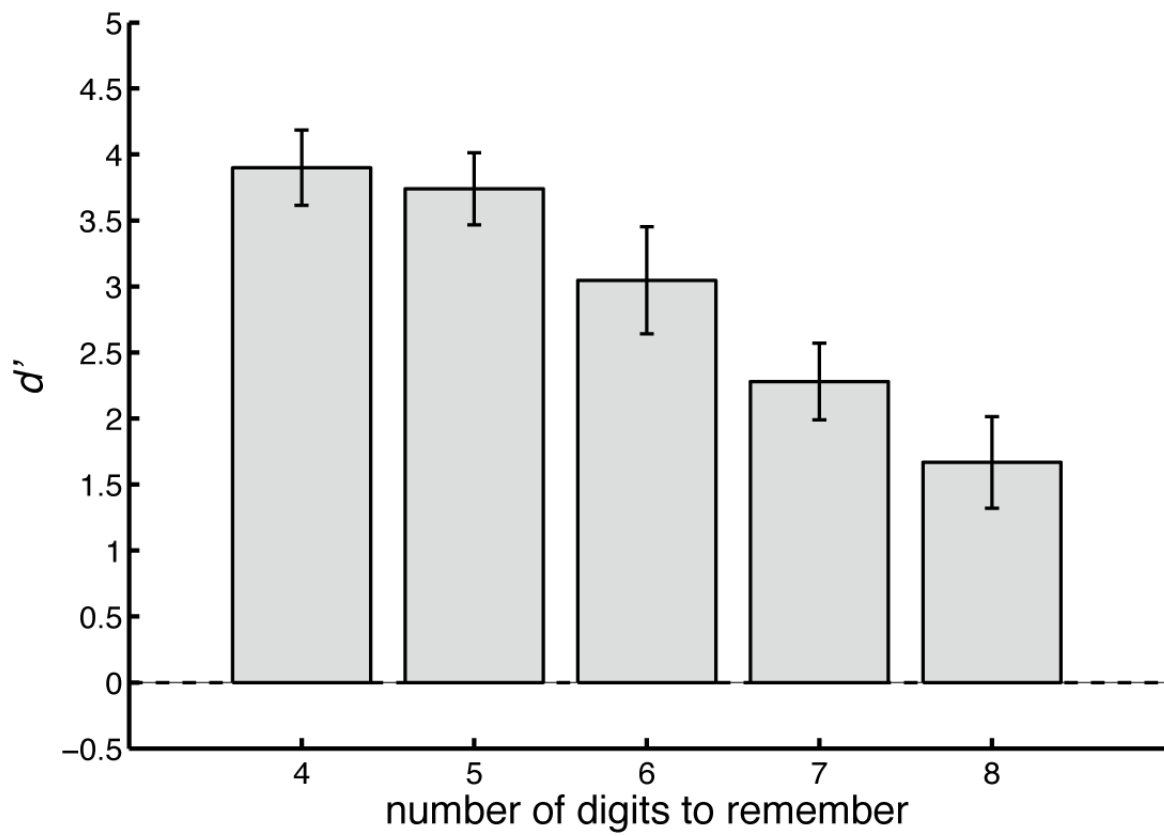


図 3.4: 実験 1: Hi 課題の予備実験の結果 (8 名平均). 記憶する数字の数を横軸に表す. エラーバーは標準誤差を表す.

3.3.2 実験参加者

全ての参加者は正常聴力を示し、神経疾患の既往歴を持たなかった (実験 2, 3, 4 についても同様)。彼らは実験への参加によって金銭を受け取った。実験手順はロンドン大学・ユニバーシティカレッジ (UCL) の研究倫理委員会によって承認された。また書面による誓約書を各参加者から受け取った。

事前実験として、すべての潜在的な参加者について Sg 課題および Hi 課題の成績を測定した。Sg 課題の事前実験では、20 試行ずつを含む 8 ブロックで課題を繰り返し練習したのち、最後の 2 ブロック 40 試行にて d' が 1 を超えた参加者のみが Hi 課題の事前実験を受けた。その後参加者は数字 6 の数の条件で Hi 課題に挑戦した。2 ブロック終了後、40 試行について d' が 2.5 を超えた参加者は、Hi 課題を数字 6 桁の条件で行った。残りの参加者は、数字 5 桁の条件について 2 ブロックを行い、40 試行について d' が 2.5 を超えた場合は数字 5 桁の条件で本課題を行った。

事前実験には 20 名が参加した。Sg 課題と Hi 課題の成績を測定した後、17 名 (平均年齢 28.8 歳、女性 5 名) が本実験に参加した。事前実験の後、6 人は 5 桁、11 人は 6 桁の条件で Hi 課題に参加した。

3.3.3 結果と考察

3.3.3.1 おとり課題の成績

実験 1 のおとり課題の結果を図 3.5 に示す。t 検定の結果、参加者はすべてのおとり課題の条件について、チャンスレベルよりも有意に高い成績をおさめたことが示された。[Sg-Lo: $t(16) = 30.3, p < 0.001$, Sg-Hi: $t(16) = 13.3, p < 0.001$, Id-Lo: $t(16) = 65.7, p < 0.001$, Id-Hi: $t(16) = 9.8, p < 0.001$]。 「メイン聴覚課題の種類」と「おとり課題の種類」を要因として、繰り返しのある分散分析を行った結果、メイン聴覚課題による主効果がみられた。 [$F(1, 16) = 11.5, p < 0.01$] また、おとり課題による主効果もみられた [$F(1, 16) = 103.3, p < 0.001$]。 2 つの要因の交互作用は見られなかった [$F(1, 16) = 0.51, p = 0.49$]。期待したとおり、参加者はおとり課題に十分注意していたことが示された。

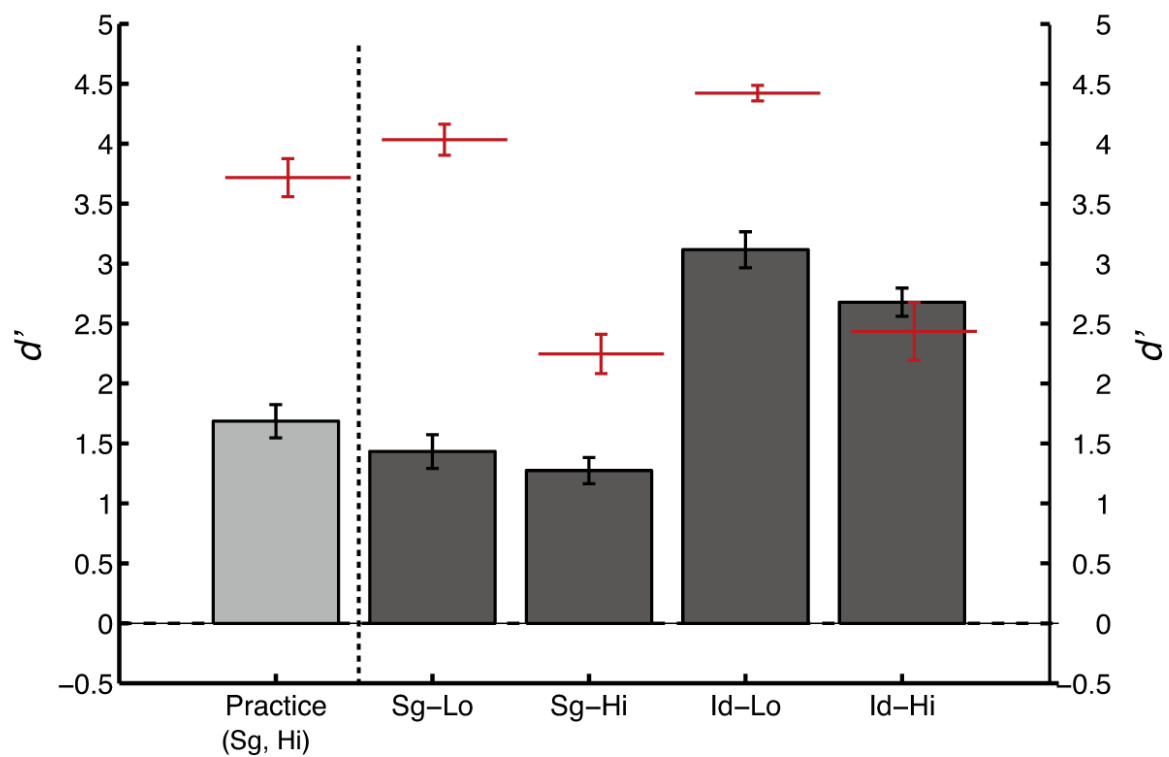


図 3.5: 実験 1: 結果 (17 名平均) . 左側の縦軸と棒グラフで示される値はメイン聴覚課題の成績値を表し、右側の縦軸と赤線で示される結果は対応するおとり課題の成績値を表す。エラーバーは標準誤差を表す。

3.3.3.2 メイン聴覚課題の成績

図 3.5 A に示す通り，Sg-Hi 課題と Sg-Lo 課題の成績は有意にチャンスレベルより高かった [Sg-Lo: $t(16) = 9.9, p < 0.001$, Sg-Hi: $t(16) = 11.4, p < 0.001$]. また Id-Lo 課題 [$t(16) = 20.2, p < 0.001$] Id-Hi 課題 [$t(16) = 22.1, p < 0.001$] も有意にチャンスレベルより高かった.

「メイン聴覚課題の種類」と「おとり課題の種類」を要因として，繰り返しのある分散分析を行った結果，メイン聴覚課題による主効果がみられた． $F(1, 16) = 174.4, p < 0.001$]. また，おとり課題による主効果もみられた [$F(1, 16) = 12.7, p < 0.01$]. 2つの要因の交互作用は見られなかった [$F(1, 16) = 1.8, p < 0.2$] Sg 課題，Id 課題ともに，Lo 課題と Hi 課題について有意な差がみられたが，その差はグラフと偏差でみられるかぎり比較的小さく，成績はチャンスレベルより十分に高い．このため，注意課題による影響は二重課題の実験パラダイムでよくみられる傾向として説明が可能である．

Sg 課題の成績は Sg-Hi, Sg-Lo 条件ともに有意にチャンスレベルより高かった．Lo 課題を行っている最中の Sg 課題の成績と，Hi 課題を行っている最中の Sg 課題の成績は非常に似通っている．両方の聴覚課題の同時高負荷視作業に若干の成績が低下する傾向がある．こうした Sg 課題への影響は，二重課題の一般的効果によって説明することができると考えられる．以上の実験 1 の結果は，Sg 課題の成績は最小限のみ高負荷課題の影響を受けていることを示唆している．

3.4 実験2: 偽ターゲット音のスペクトル情報による影響の確認

実験2は、実験参加者がスペクトル情報のみに基づいて検査音がターゲット音か偽ターゲット音かを判断していないことを確認するために行った。McDermott らの実験パラダイムにおいて、ターゲット音と偽ターゲット音を区別するために、検査音の平均スペクトルが使用されていた可能性がある。実験2では、偽ターゲット音として実験1で利用した偽ターゲット音 (以下、OR 検査音と表記) に代えて、ターゲット音と平均スペクトルが等しい時間反転ターゲット音 (以下 TR 検査音と表記) を用いた。

3.4.1 方法

実験の基本的な手順は実験1とほぼ同様であった。ただし、Sg-Hi, Sg-Lo, Id-Hi, Id-Loそれぞれの条件ブロックについて、実験1とは倍の数の試行が用意された (1 ブロックについて 80 試行)。OR 検査音および TR 検査音は1つのブロック内で半数ずつランダムに呈示された。80 の OR 検査音、80 の TR 検査音と 160 のターゲット音が、検査音の呈示段階でランダムに呈示された。試行の合計数は、各条件合計して 320 であった。参加者は、2 種類の検査音の違いについて知らされず、ただ検査音が混合物中のものと同じ音だったかどうかを判断するように指示された。実験には参加者 1 名あたり約 180 分を要した。

3.4.2 実験参加者

事前実験の方式は、実験1と同様であった。17名の参加候補者について、Sg課題とHi課題の成績を確認し、それぞれ参加基準のしきい値以上の成績を出した12名 (平均年齢 27.8 歳、女性 4 名) が本実験に参加した。12名のうち10名がHi課題を数字6桁の条件で行い、2名が数字5桁の条件で行った。

3.4.3 結果と考察

3.4.3.1 おとり課題の成績

実験結果を図 3.6 に示す。t 検定の結果、参加者はすべてのおとり課題の条件について、チャンスレベルよりも有意に高い成績をおさめたことが示された。[OR-Sg-Lo: $t(11) = 18.4, p < 0.001$, OR-Sg-Hi: $t(11) = 5.8, p < 0.001$, OR-Id-Lo: $t(11) = 24.0, p < 0.001$, OR-Id-Hi: $t(11) = 6.8, p < 0.001$, OR-Sg-Lo: $t(11) = 20.5, p < 0.001$, OR-Sg-Hi: $t(11) = 6.8, p < 0.001$, OR-Id-Lo: $t(11) = 29.6, p < 0.001$, OR-Id-Hi: $t(11) = 7.6, p < 0.001$] 「検査音の種類」「メイン聴覚課題の種類」と「おとり課題の種類」の3要因による、繰り返しのある分散分析を行った結果「検査音の種類」による主効果はみられず [$F(1, 11) = 1.7, p < 0.21$] , 「おとり課題の種類」による主効果がみられた [$F(1, 11) = 31.0, p < 0.001$]。また、メイン聴覚課題の種類による主効果は、有意ではなかったものの、有意傾向がみられた [$F(1, 11) = 3.3, p < 0.09$]。3つの要因による交互作用はみられなかった [$F(1, 11) = 0.008, p = 0.93$]。期待したとおり、参加者はおとり課題に十分注意していたことが示された。

3.4.3.2 メイン聴覚課題の成績

実験結果を図 3.6 に示す。参加者の成績は、全条件について有意にチャンスレベルより高かった [OR-Sg-Lo: $t(11) = 6.8, p < 0.001$, OR-Sg-Hi: $t(11) = 5.7, p < 0.001$, OR-Id-Lo: $t(11) = 16.5, p < 0.001$, OR-Id-Hi: $t(11) = 12.0, p < 0.001$; TR-Sg-Lo: $t(11) = 5.8, p < 0.001$, TR-Sg-Hi: $t(11) = 4.9, p < 0.001$, TR-Id-Lo: $t(11) = 24.1, p < 0.001$, TR-Id-Hi: $t(11) = 13.0, p < 0.001$]。 「検査音の種類」「メイン聴覚課題の種類」と「おとり課題の種類」の3要因による繰り返しのある分散分析を行った結果「検査音の種類」による主効果がみられた [$F(1, 11) = 20.1, p < 0.001$]。また「メイン聴覚課題」による主効果がみられた [$F(1, 11) = 109.4, p < 0.001$]。また「おとり課題」の主効果もみられた [$F(1, 11) = 8.7, p < 0.05$]。また「検査音の種類」と「メイン聴覚課題」の間に交互作用がみられた [$F(1, 11) = 6.7, p < 0.05$]。Shaffer の方法による多重比較を行ったところ、Sg 課題の成績について、OR 検査音条件と TR 検査音条件の間に有意な差がみられた (p

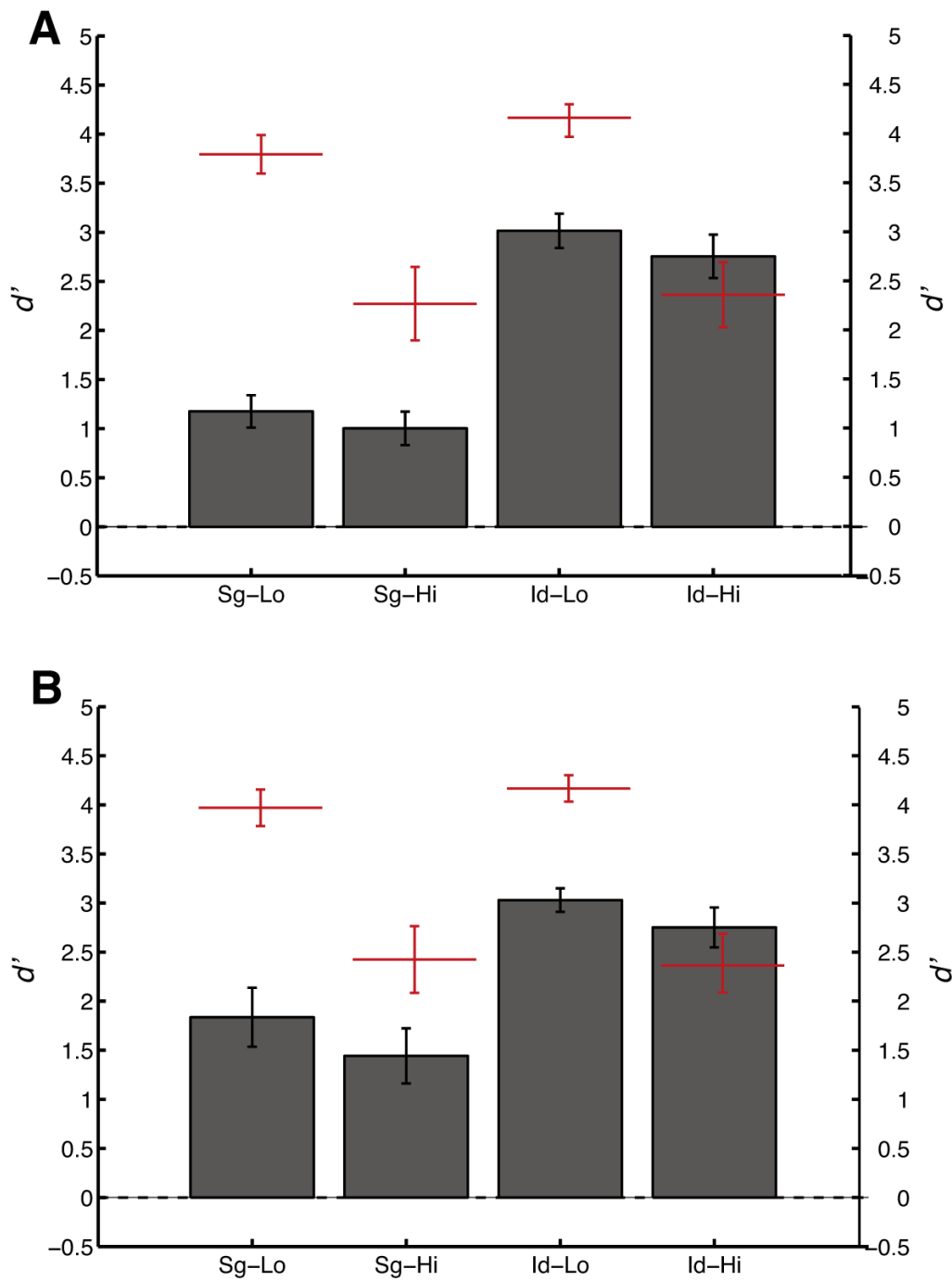


図 3.6: 実験 2: 結果 (10 名平均). 左側の縦軸と棒グラフで示される値はメイン聴覚課題の成績値を表し, 右側の縦軸と赤線で示される結果は対応するおとり課題の成績値を表す. A: OR 検査音での結果. B: TR 検査音での結果.

< 0.05). (McDermott et al., 2011) の実験では、OR 検査音による実験の結果は、TR 検査音による結果よりもわずかに成績が良いことが分かっている。この結果は、本実験によって得られた方法とは逆となっている。

実験結果は、実験1の場合とほぼ同様であった。おとり課題はほぼSg課題の成績に影響を与えていないと考えられる。この結果は時間反転のターゲット音を検査音とする場合でも、参加者が高負荷の視覚課題を同時に行いつつSg課題を遂行することができることを示す。さらにこの結果は同時課題の実行中に参加者がターゲット音のスペクトル時間構造を回復することができることを示している。

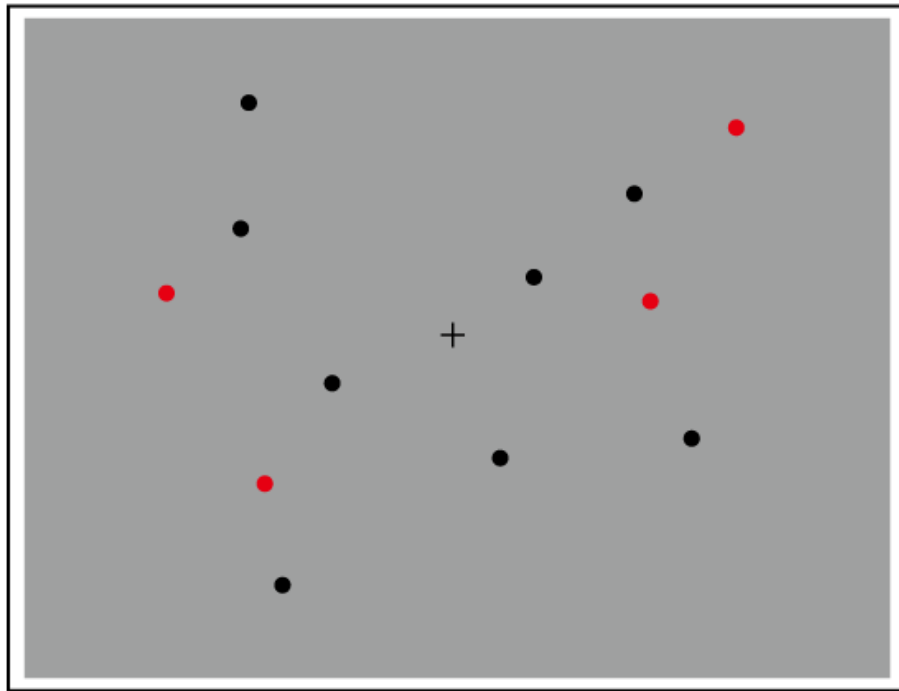


図 3.7: 移動物体追跡課題の開始画面の一例。赤いドットが追跡対象となる。中央の十字は注視点を表す。試行開始から 1 秒後に赤いドットは黒色と変わる。参加者は赤で表示されていたドットを追跡するよう指示された。

3.5 実験 3: 視覚入力による集中課題

実験 3 では注意的に要求しているが、エコイックメモリに依存しない視覚的な課題である、オブジェクト追跡課題 (moving object tracking: MOT 課題) をおとり課題として使用した。おとり課題の実験パラダイムは Franconeri, Jonathan, and Scimeca (2010) で用いられたオブジェクト追跡課題に基づいており、実験の目的にあわせて適宜追跡対象のオブジェクトの数やオブジェクトの移動速度などのパラメタを調整した。

3.5.1 方法

図 3.7 は刺激の画面の例を表す。実験 1, 2 とは異なり, Hi 課題と Lo 課題の指示内容は同一であった。参加者は, 刺激が出ていない間は画面上の注視点を見るように指示された。試験の開始時に, 12 個の黒丸がコンピュータの画面上に表示された。その 2 秒後, 12 個のうち 4 個の丸が赤で 2 秒間表示された。その 2 秒後, 12 個のドットは画面上をランダムに動き回った。最初の 1 秒が経過したのち, 赤で表示されていた丸が黒に戻った。参加者は, 3.7 秒の間, ターゲットとして表示された丸を追いつけた。その後, すべてのオブジェクトが停止し, 参加者はターゲットをマウスでクリックするよう指示された。

ドットの速度は Hi 条件では毎秒 5 フレーム, Lo 条件で毎秒 2 フレームとした。課題の成績は, それぞれの試行について, 4 ドットの正答率を合計することによって算出した。その他の手続きは Franconeri et al. (2010) と同様とした。実験 3 は 1 人あたりおよそ 150 分で終了した。

Hi 条件として用いるのに適切となるよう, MOT 課題の移動スピードを調節するため, 事前実験を行った。本実験参加者とは大部分が異なる 10 名の参加者 (平均年齢 28.5 歳, 女性 2 名) が事前実験に参加した。ドットのスピードを変えた 6 条件 [1, 2, 3, 4, 5, 6 フレーム毎秒の移動] が異なるブロックで呈示された。ひとつのブロックは 20 試行で構成された。ブロックの順序はランダム化した。事前実験の結果を図 3.8 に示す。5 フレーム毎秒のとき, およそ 60% の正答率が得られたことから, 最終的に Hi 課題のドット速度として 5 フレーム毎秒を選択した。

3.5.2 実験参加者

13 名が Sg 課題と Hi 課題の事前実験に参加し, うち 10 名 (平均年齢 29.2 歳, 女性 2 名) が本実験に参加した。全ての参加者が Hi 課題として 5 フレーム毎秒の条件にて実験を行った。

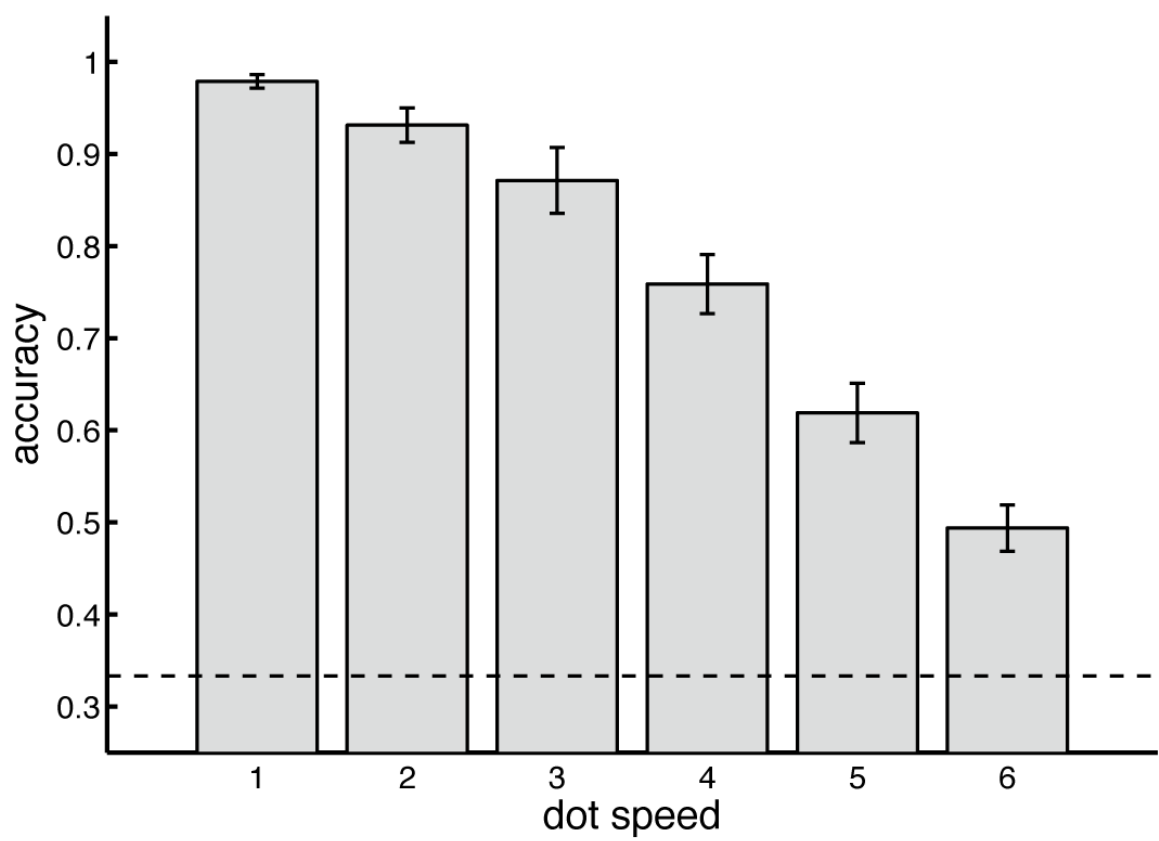


図 3.8: 実験 3: Hi 課題の予備実験の結果 (10 名平均). ドットの移動スピードのパラメタを横軸に表す. エラーバーは標準誤差を表す.

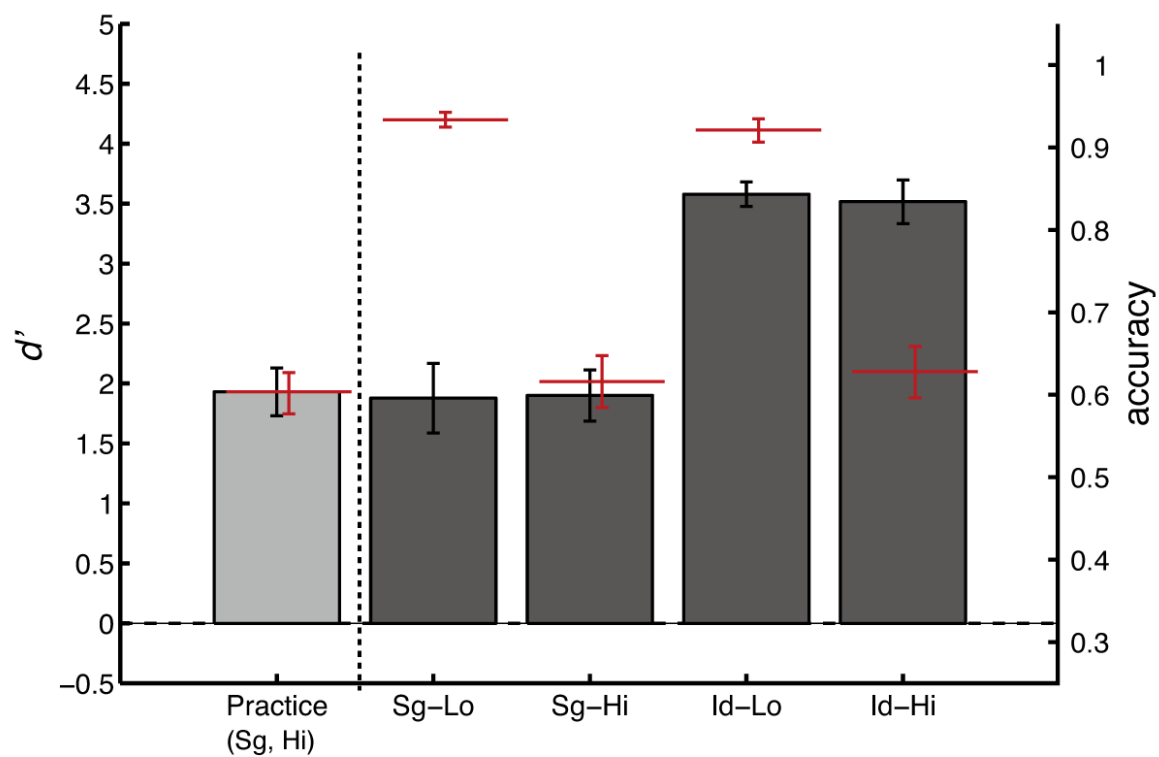


図 3.9: 実験 3: 結果 (10 名平均). 左側の縦軸と棒グラフで示される値はメイン聴覚課題の成績値を表し, 右側の縦軸と赤線で示される結果は対応するおとり課題の成績値を表す. エラーバーは標準誤差を表す.

3.5.3 結果と考察

3.5.3.1 おとり課題の成績

図 3.9 に結果を示す。t 検定の結果、参加者はすべてのおとり課題の条件について、チャンスレベルよりも有意に高い成績をおさめたことが示された。[Sg-Lo: $t(9) = 97.6, p < 0.001$, Sg-Hi: $t(9) = 18.6, p < 0.001$, Id-Lo: $t(9) = 62.2, p < 0.001$, Id-Hi: $t(9) = 19.1, p < 0.001$]. 「メイン聴覚課題の種類」と「おとり課題の種類」を要因として、繰り返しのある分散分析を行った結果、おとり課題の種類による主効果がみられた [$F(1, 9) = 160.8, p < 0.001$] また、2つの要因による交互作用の有意傾向がみられた [$F(1, 9) = 3.8, p = 0.08$]. メイン聴覚課題による主効果はみられなかった [$F(1, 9) = 0.01, p = 0.91$]. 期待したとおり、参加者はおとり課題に十分注意していたことが示された。

3.5.3.2 メイン聴覚課題の成績

図 3.9 にメイン聴覚課題の成績を示す。Sg-Hi 課題と Sg-Lo 課題の成績は有意にチャンスレベルより高かった [one-sample t test differed from 0, Sg-Lo: $t(9) = 6.1, p < 0.001$, Sg-Hi: $t(9) = 8.5, p < 0.001$] また、Id-Hi 課題と Id-Lo 課題の成績も有意にチャンスレベルよりも高かった [Id-Lo: $t(9) = 33.0, p < 0.001$; Id-Hi: $t(9) = 18.3, p < 0.001$]. 「メイン聴覚課題の種類」と「おとり課題の種類」を要因として、繰り返しのある分散分析を行った結果、メイン聴覚課題による主効果がみられた [$F(1, 9) = 59.3, p < 0.001$]. おとり課題による主効果はみられなかった [$F(1, 9) = 0.02, p = 0.89$]. また2つの要因の交互作用もみられなかった [$F(1, 9) = 0.18, p = 0.68$]. 上記の結果は実験1の結果と類似する。この結果は、高負荷の MOT 課題を同時に行っているときであっても、Sg 課題が可能であることを示唆する。

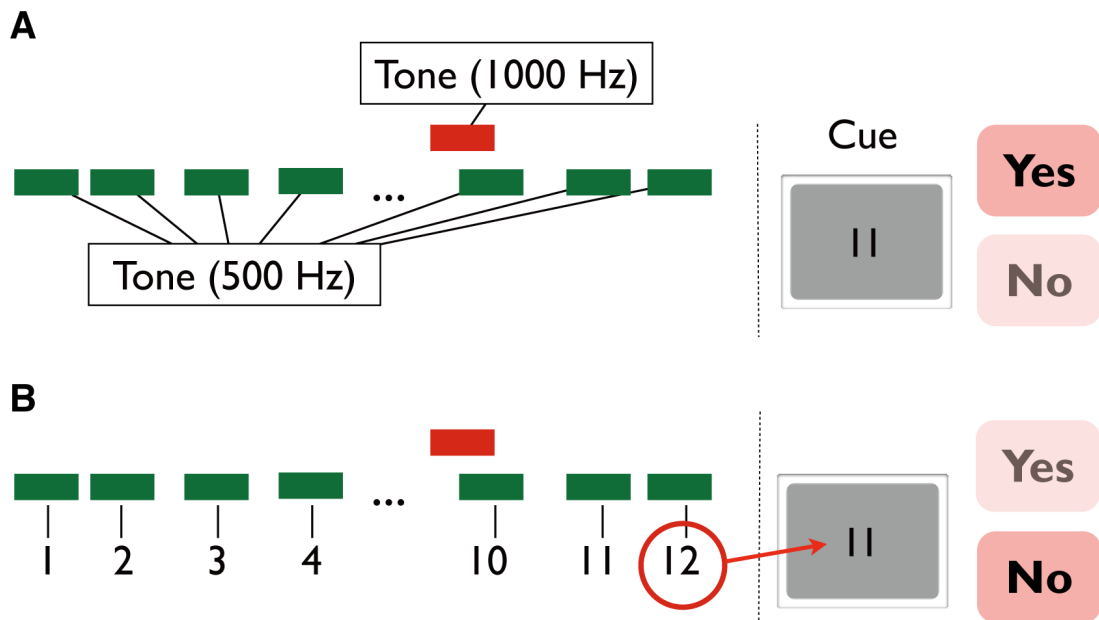


図 3.10: 実験 4 で用いた聴覚記憶課題の概念図。緑、赤の四角は正弦波音のトーンを表す。A: 低負荷 (Lo) 課題。参加者は音の再生中に 1000Hz の高い音が出現したか否かのみを回答した。参加者は低音は無視するよう指示を受けた。B: 高負荷 (Hi) 課題。参加者はランダムなタイミングで呈示されるトーン音を数え、cue の画面で表示される数と一致するか否かを回答した。参加者は高音は無視するよう指示を受けた。

3.6 実験 4: 聴覚入力による記憶・集中課題

実験 4 では、Hi 課題として高速で呈示されるトーン音の数を数える聴覚課題、Lo 課題として配列中の特異な音を検出する聴覚課題をおとり課題とした。実験 4 では、片側の耳からおとり課題が呈示され、もう片側の耳からメイン聴覚課題が呈示される両耳分離聴課題であった。

3.6.1 方法

図 3.10 におとり課題の内容を示す。刺激は Hi 課題と Lo 課題で同様であった。10, 11, 12, もしくは 13 のトーンピップが 3.2 秒の間に呈示された。トーンとトーンの間隔はランダム化されていた。トーンピップは 100 ミリ秒, 500 Hz での持続時間を有していた。トーンとトーンの最小間隔は 200 ミリ秒に設定した。また, トーン列には, 50% の割合の試行で 1000 Hz のトーンピップが 1 回, 試行ごとランダムな時点に呈示された。メイン聴覚課題の検査音が呈示されたのち, スクリーン上に 9 から 14 までの数字のうちいずれかが呈示された。呈示される数字は, 試行中呈示された低周波数のトーンピップの数か, その数から 1 上下にずれた数であった。

Hi 条件では, 参加者は低いトーンピップの数を数え, 画面上の数字を確認し, 数字が音の数と同じであったかどうかを答えるように要求された。参加者は低音のトーンピップを無視するように求められた。Lo 条件では, 参加者は音列の中から高音を検出し, 画面上の数字の内容を無視して, 音列の中で高い音を聞いたかどうかを答えるように要求された。呈示する耳の左右の条件は, 2 ブロックで交互に変更された。その他の手続きは実験 1 と同様であった。参加者は [Sg-Lo Sg-Hi, Id-Lo, Id-Hi] 課題を各 2 回ずつ, 8 ブロック行った。実験には約 150 分を要した。

実験 1, 2, 3 と同様に, 事前に Hi 課題の負荷量を決めるための事前実験を行った。10 名 (平均年齢 28.4 歳, 女性 6 名) が事前実験に参加した。各条件 ([5, 6, 7, 8], [8, 9, 10, 11], [11, 12, 13, 14], [14, 15, 16, 17], [17, 18, 19, 20] のトーンピップ数) が, 参加者間でランダム化されたブロックの順序で, 別々のブロックとして呈示された。各ブロックには 16 試行が含まれ, 参加者は各条件を 2 ブロックずつ行った。結果は図 3.11 に示す。最終的に d' で 2.5 近辺を与える [8, 9, 10, 11] 条件と [11, 12, 13, 14] 条件の間である [10, 11, 12, 13] のトーンの数を実験に採用した。

3.6.2 実験参加者

全ての参加者は事前実験として Sg 課題と Hi 課題を単独で行った。Sg 課題, Hi 課題の成績は片側の耳にて確かめられた。呈示する耳は 2 つのブロックで左右逆とした。お

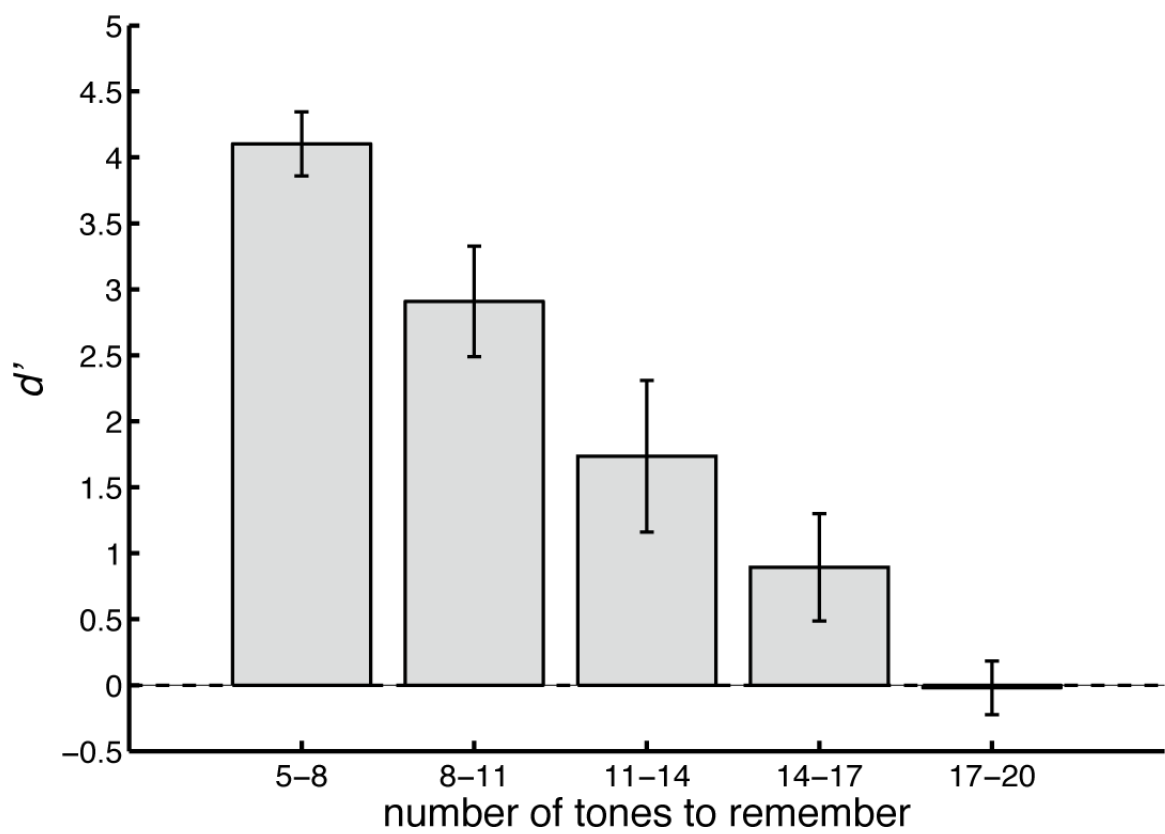


図 3.11: 実験 4: Hi 課題の予備実験の結果 (10 名平均). カウントするトーン音の数の組み合わせを横軸に表す. エラーバーは標準誤差を表す.

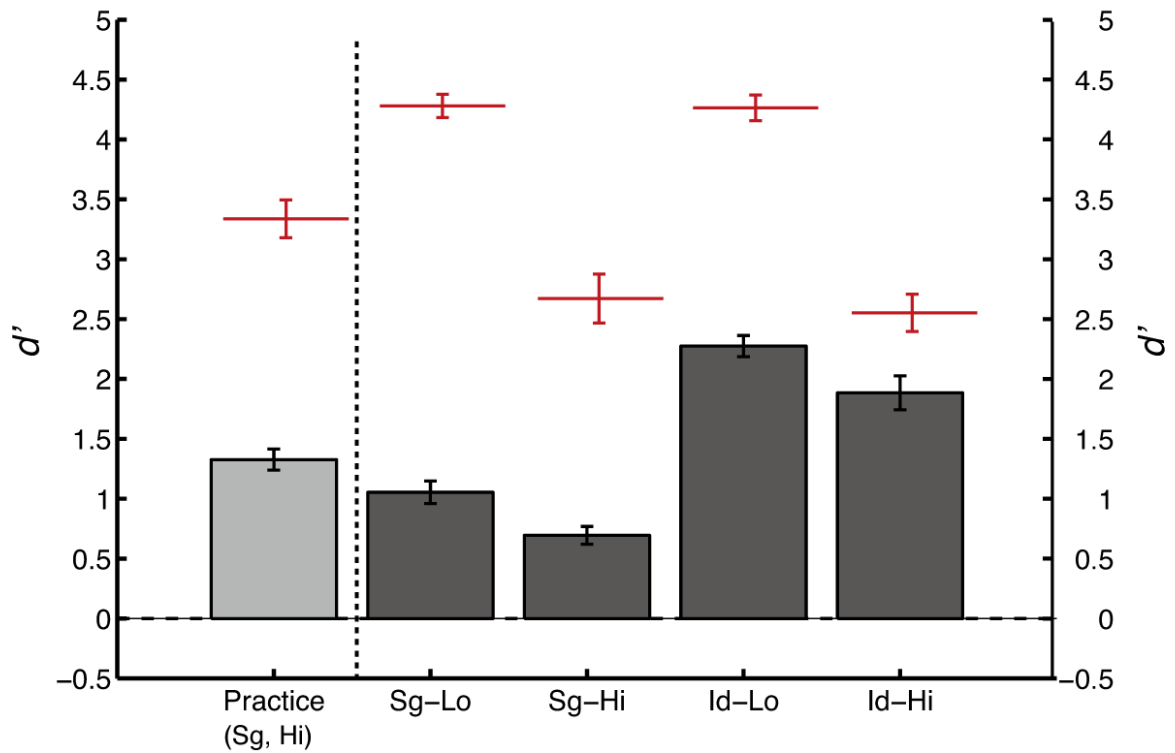


図 3.12: 実験 4: 結果 (17 名平均). 左側の縦軸と棒グラフで示される値はメイン聴覚課題の成績値を表し, 右側の縦軸と赤線で示される結果は対応するおとり課題の成績値を表す. エラーバーは標準誤差を表す.

とり課題については, [8, 9, 10, 11] 条件について, 20 試行を含む 2 ブロックを繰り返した後, d' が 2.5 を超えた参加者は, よりトーンの多い [10, 11, 12, 13] で 2 ブロックを行い, 成績を確かめられた. [10, 11, 12, 13] 条件について d' が 2.5 を超えた参加者は [10, 11, 12, 13] にて本実験の Hi 課題を行い, 2.5 を下回った参加者は [8, 9, 10, 11] にて本実験を行った. 44 名が事前実験に参加し, 本実験参加のための規定の成績値を満たした 17 名 (平均年齢 26.7 歳, 女性 8 名) が本実験に参加した.

3.6.3 結果と考察

3.6.3.1 おとり課題の成績

図 3.12 におとり課題の成績を示す。t 検定の結果、参加者はすべてのおとり課題の条件について、チャンスレベルよりも有意に高い成績をおさめたことが示された [Sg-Lo: $t(16) = 42.8, p < 0.001$, Sg-Hi: $t(16) = 12.7, p < 0.001$, Id-Lo: $t(16) = 38.7, p < 0.001$, Id-Hi: $t(16) = 16.0, p < 0.001$]。 「メイン聴覚課題の種類」と「おとり課題の種類」を要因として、繰り返しのある分散分析を行った結果、おとり課題の主効果がみられた [$F(1, 16) = 70.7, p < 0.001$]。 メイン聴覚課題の主効果はみられなかった [$F(1, 16) = 0.31, p = 0.58$]。 また 2 つの要因の交互作用もみられなかった [$F(1, 16) = 0.44, p = 0.52$]。 期待したとおり、参加者はおとり課題に十分注意していたことが示された。

3.6.3.2 メイン聴覚課題の成績

図 3.12 にメイン聴覚課題の成績を示す。Sg-Hi 課題と Sg-Lo 課題、Id-Lo 課題と Id-Hi 課題の成績はともに有意にチャンスレベルより高かった [Sg-Lo: $t(16) = 10.9, p < 0.001$, Sg-Hi: $t(16) = 9.0, p < 0.001$; Id-Lo: $t(16) = 25.0, p < 0.001$; Id-Hi: $t(16) = 13.0, p < 0.001$]。 「メイン聴覚課題の種類」と「おとり課題の種類」を要因として、繰り返しのある分散分析を行った結果、メイン聴覚課題による主効果がみられた [$F(1, 16) = 97.0, p < 0.001$]。 また、おとり課題の種類による主効果もみられた [$F(1, 16) = 33.6, p < 0.01$]。 2 つの要因の交互作用は見られなかった [$F(1, 16) = 0.03, p = 0.87$]。 Sg 課題、Id 課題ともに、Lo 課題と Hi 課題について有意な差がみられたが、その差はグラフと偏差で見られるかぎり比較的小さく、成績はチャンスレベルより十分に高い。 このため、注意課題による影響は二重課題の実験パラダイムでよくみられる傾向として説明が可能である。 以上の実験 4 の結果は、Sg 課題の成績は最小限のみ高負荷課題の影響を受けていることを示唆している。

3.7 研究2: 総合考察

本研究では、近年発見された「繰り返し」手がかりに基づく混合音の分離機能に対する注意の負荷の効果を調べた (McDermott et al., 2011). この研究での主な発見は、参加者が注意を要する困難な作業を実行している間でも、混合物からの音を識別するために、繰り返し手がかりを使用できることであった。実験で採用したおとり課題は「繰り返し」手がかりに基づく分離課題を実施するための能力に対して強い影響力を持たない。これまでのところ、採用したおとり課題への注意によって分離課題の成績が損なわれるという証拠は得られていない。実験1,2,3,4のすべてについて、結果に類似した傾向が見られた。研究2の結果は、参加者の注意が他の場所に向けられているときであっても、これらの混合物から新規な音を分離するために「繰り返し」手がかりが使用できることを示唆している。

McDermott らの論文で述べられているように、この新たに発見された聴覚分離手がかりは音響特性の事前知識を必要としない。この音響的手がかりは、対象の音が混合音の中で一貫して同じ音響的特性を維持するという状況であれば働きうる。繰り返しに基づく分離は、しばらくの間、聴覚入力を保持するための一時的な「記憶」を必要とする経時的な情報や聴覚システムを統合することにより、聴覚系で行われると考えられている。実験1および2においては、繰り返しに基づく分離の処理が、短期記憶を使用した数字の暗唱を実施した場合であっても十分に可能であることが示された。実験3においては、繰り返しに基づく分離の処理が、連続した集中刺激の呈示の最中に行われた場合でも、十分に可能であることが示された。これらの結果は、繰り返しに基づく分離の機能は、認知処理や、視覚などの他の感覚情報の統合に関し、比較的上位の情報処理システムの近くではなく、中間的で比較的低レベルの聴覚処理の段階に位置していることを示唆する。

この研究の結果は生態学的な観点から見て妥当性がある。生物にとって、捕食者の足音など、危険な兆候との関係性が深いリズムミカルな音を背景雑音から検出し、学習することは生命活動を維持するうえで重要である (Wiley & Richards, 1978). 常時の集中や高度な認知過程が音源の基本的な知識を取得するために必要であるとは考えにくい。繰り返しに基づく分離機能が、視覚や触覚などの他のモダリティの処理によって干渉されな

いという事実は、音源の学習過程の重要性を考えるとともっともらしい。

実験4では、結果の傾向がある程度実験1, 2, および3と異なっていた。両耳統合のプロセスがSg課題の最初の場所で一定の影響を持っている可能性がある。また実験4で正常にSg単体条件の事前実験に合格した参加者の割合が、他の条件と比べて劇的に減少しているが、この減少は課題を両耳分離聴形式に変更することによって引き起こされている可能性がある。

注意の客観的な定量化が難しいため、参加者の注意が完全にターゲットからそらされることを保証することは困難である。ただし、今回の研究では、事前にHi課題の難易度を確認し、事前実験を行っている。この際 $d' > 2.5$ を「適度な難易度」の指標として用いている。その結果、参加者はHi課題で高い正解率を達成した。これは、参加者がおとり刺激について、あきらめて聴覚課題に注意を向けることなく集中したことの間接的な保証となる。

Bregman は「原初的な過程」と「図式による過程」の二つの過程からなるモデルを提案した (Bregman, 1990)。Bregman は前者の過程は自動的に起こり、注意を必要としないと考えた (1.9 節参照)。その後 Carlyon et al. (2001) は ABA ストリームの聴覚情景分析のビルドアップ現象が注意を必要とすることを主張した。また、Sussman et al. (2007) は、注意はストリーミングのビルドアップに関与しないと主張した。議論は現在も継続されており、決定的な結論は未だ導かれていない (1.9 節を参照)。研究2の結果は、繰り返しに基づく分離が注意のための認知的資源を必要としないことを示唆している。

本研究で使用した McDermott らによる刺激は、高調波複合音のような同時分離の刺激に近い機能を備えていると考えられる。この刺激は、ABA 音や調波複合音よりも複雑であり、自然な音の統計的な二次機能を反映している。ただし、刺激呈示の時間スケールは経時分離に似ている。また、同じ時間内に2つの音が重なった点で同時分離の特徴を有している。過去のストリーミングの研究で使われていた刺激は、自然ではなく単純な音 (Roberts & Brunstrom, 1998; Carlyon, 1991; McDermott & Oxenham, 2008) か音声のような実際の環境の音 (Culling & Darwin, 1993) かのどちらかであることが多い。この新たな刺激と今回の調査結果は、聴覚情景分析と注意との関係性についての議論のために有用であると考えられる。

第4章 全体考察

本研究では、聴覚情景分析において重要な役割を果たす「補完」と「分離」の二つのプロセスの特性とメカニズムについて、心理物理学的手法を用いて検討した。

研究1では、欠落した音の補完が聴覚系のどの処理段階で行われるかを調べるため、周波数スイープ音の連続聴錯覚を引き起こす音刺激を用いて、雑音によって隠された音の内容を補完して知覚的まとまりを作成する情報処理段階が、聴覚オブジェクトの特徴抽出段階よりも先の処理段階に位置するか否かを調べた。実験1では、実際のスイープ音の残効量と錯覚のスイープ音の残効量を比較した。結果、錯覚のスイープ音の残効量は実際のスイープ音と比較して非常に小さいことが明らかになった。実験2では、雑音が存在すること自体が残効を抑制する可能性について検討するため、実際のスイープ音にノッチ雑音を足し合わせた刺激と、中間部を切り取ったスイープ音にノッチ雑音を足し合わせた刺激の残効量と、雑音のない実際のスイープ音の残効量を比較した。結果、ノッチ雑音の重畳がある刺激の残効量は、実際のスイープ音の残効量と比較して非常に小さいことが明らかになった。研究1の結果、錯覚の周波数スイープ音に対する順応があったとしても、仮想的な周波数変動検知器に対する影響は、実際の周波数スイープ音への順応の場合と同様には起こらないことが分かった。この結果から、連続聴錯覚は聴覚オブジェクト生成の特徴抽出の段階よりも先の処理段階にて生成されていることが示唆された。先行研究では、連続聴錯覚には注意が必要とされないという見方が示されている。研究1の結果と先行研究の知見を総合すると、連続聴錯覚の知覚には、特徴抽出の段階よりは上位であり、注意の影響が及ぶ段階よりも自動性の強い段階が関与すると考えられる。

研究2では、ボトムアップ手がかりやトップダウン手がかりを必要としない「繰り返し手がかりに基づく音の分離」の処理が音への選択的注意によって影響を受けるか、ある

いはそれが部分的に注意や集中の影響を受けるか否かを調べた。実験は二重課題デザインとし、実験参加者の集中を必要とするおとり課題の種類を実験ごとに変えて、参加者が音を聞くことに集中していなくとも「繰り返し」手がかりに基づく音の分離が可能であるかを検討した。実験1では、視覚入力を通じてエコイックメモリに負荷が与えられた際の聴覚課題の成績を調べた。実験2では、実験1にて参加者が偽ターゲット音の単純な音響的要素(平均スペクトル)のみに基づいて解答していたか否かを確かめた。実験3では、エコイックメモリが必要とされないが集中を必要とする課題を同時に行った際の聴覚課題の成績を調べた。実験4では、聴覚から入力されるおとり課題に答えつつ、聴覚課題を行ったときの能力を調べた。研究の結果、3種のおとり課題と同時に課題が行われたとしても、「繰り返し手がかり」に基づく分離の成績の低下は限定的であることが示された。この結果から「繰り返し手がかり」に基づく音の分離過程は、選択的注意の影響を受けにくい段階に位置する、自動的なプロセスであることが示唆された。

以上の研究から、聴覚情景分析における「補完」と「分離」の二つのプロセスには、聴覚系の特徴抽出レベルよりは上位であり、かつ注意を必要とせず自動的に作動する中間レベルのメカニズムが関与していることが示唆された。

現在提案されている聴覚情景分析のモデルの多くは、聴覚情景分析の問題を極度に単純化することが多いと言える。ABA音パターンを実験刺激として用いた経時グルーピングの研究は数多く存在する。しかし、時間的に重複のある音のグルーピング(同時グルーピング)に関する研究は数が少ない。ABA音パターンを用いた実験の結果は聴覚情景分析研究の結果の代表例として扱われることが多い。しかし、ABA音パターンはあくまでも実験環境で用いられる単純な音刺激であり、現実の環境の聴覚情景分析の問題を必ずしも反映しない。聴覚系に入力される音のパターンは、自然界で起こりうる音の組み合わせの数だけ無数に存在する。そして、ひとつひとつの情景分析は、ひとつひとつの入力パターンに対して、適宜行われる。聴覚情景分析によって立ちあらわれる知覚は、入力のパターンに応じて毎度変化する。ShammaやMicheyl, Elhilaliらのグループは、音脈分凝の神経基盤について説明するモデルを提案している(Elhilali et al., 2009)。彼らは神経細胞群の発火タイミングの一致性が、聴覚系が「知覚のまとまり」を作る際の手がかりとなると主張し、統一的説明を試みている。しかし、ABA音パターンを入力として

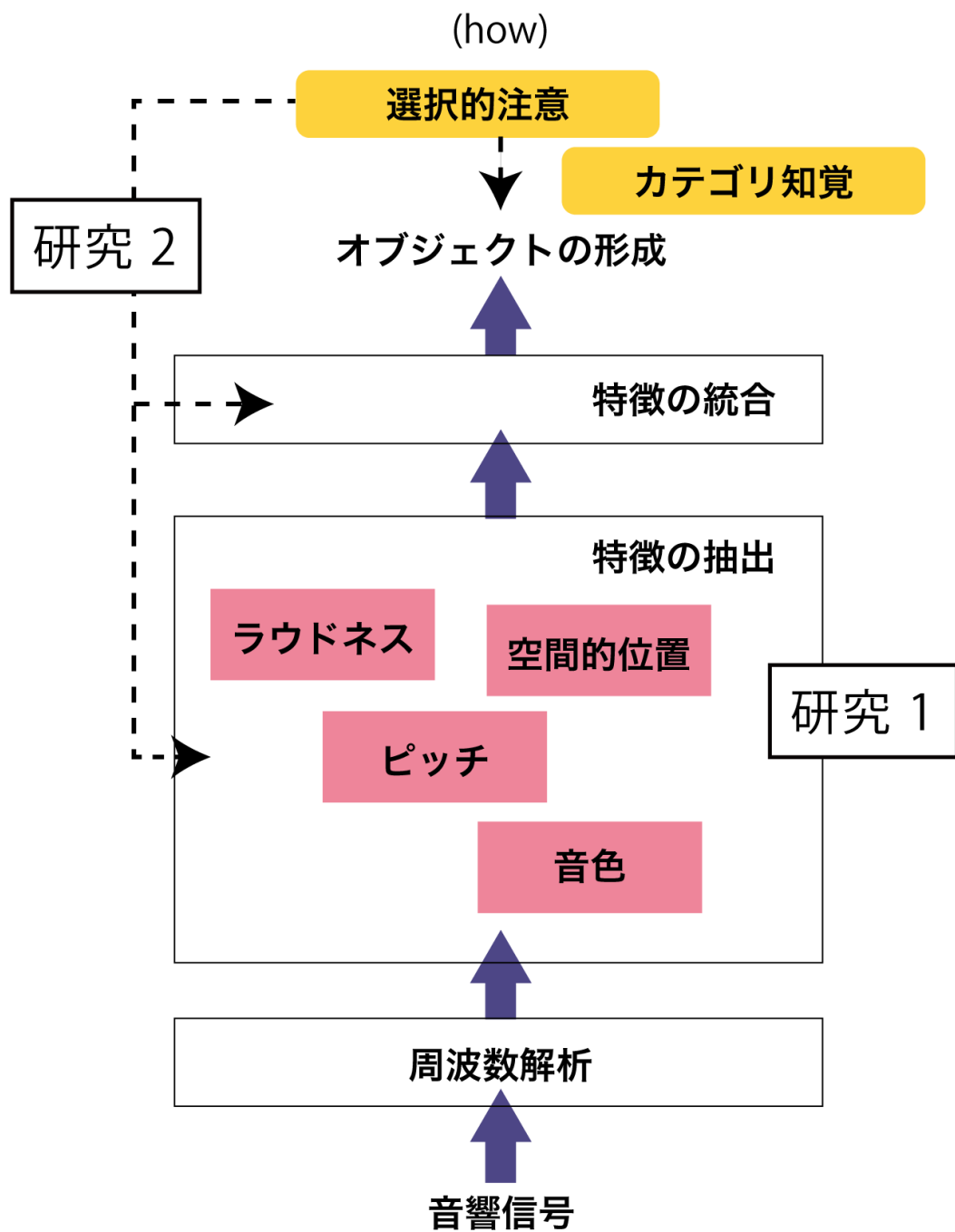


図 4.1: 聴覚情報処理過程の模式図上における, 研究 1 と研究 2 の関心領域.

得た聴覚系の出力は、ABA 音パターンが入力であったために起きた出力であるとも考えられる。限られた入力によって得た知見のみから、情景分析すべてに対する統一的な規則を導くことはできないと考える。これまでの研究の歴史の中で用いられてきた、純音や白色雑音などの単純な刺激のみでは、単純な出力結果しか得ることができない。

聴覚情景分析を実験室で検証しようとするとき、対象とするシーンや刺激を単純にすれば、実際に現実に行き起きている状態の再現度が低くなり、一方実験室で再現するシーンや刺激を複雑にすれば、音の含む要素が増え、普遍的な主張や発見が難しくなる。自然に生じる音の重要な特性を切り出すことができ、かつ、実験刺激として用いることが容易な音刺激を多く見つけていく必要がある。

本研究は、聴覚系という複雑な情報処理システムがどのような処理系を内部に持つのかについて、重要な知見を得たと考える。脳撮像や神経記録などの研究方法と、心理物理学の知見を組み合わせることで、聴覚系が知覚を生み出す過程に関する新しい発見が得られることだろう。

謝辞

本研究の遂行にあたり、長年に渡りご指導・ご鞭撻を賜りました、指導教官の柏野 牧夫教授、副指導教官の内川 恵二教授に心より感謝致します。また、研究の方向性と今後の展開について数多くの御助言を頂きました、NTT コミュニケーション科学基礎研究所の古川 茂人様、大石 悠貴様、米家 敦様、首都大学東京の林 意凡様 (当時 NTT コミュニケーション科学基礎研究所)、そして NTT コミュニケーション科学基礎研究所の皆様に感謝致します。

日々の研究活動の中で幾度となく相談に乗って頂き、またご自身の研究でお忙しい中多くの実験にお付き合い頂いた、NTT コミュニケーション科学基礎研究所の大塚 翔様 (当時 東京大学)、柏野研究室のメンバー、山岸 慎平さん、渡辺 謙さん、箕谷 啓太さん、修了生の芦原 孝典さん、鷺澤 史歩さん、山口 陽平さんに感謝致します。

博士一貫コースの海外研修中、研究2の遂行にあたってご指導を賜りました、University College London Ear Institute の Maria Chait 先生、Tobias Overath 様、École normale supérieure の Alain de Cheveigné 様、Massachusetts Institute of Technology の Josh H. McDermott 様に心より感謝致します。また、英語が不得手で会話もままならず、異郷の地でふて腐れていた私を慰め、ときに叱咤してくれた、University College London Ear Institute の Nicolas Barascud さん、Andreou Lefkothea-Vasiliki さん、Anahita Mehta さん、Sundeep Teki さん、そしてロンドンでお世話になった皆様にこの場を借りてお礼申し上げます。

最後に、今日に至るまで私を支えてくれた故郷の父、母、祖父、祖母と、いつも穏やかに見守ってくれた夫、洋平さんに感謝します。

この論文を最後まで書き上げることができたのは、ひとえに皆様のご支援とご指導の賜物です。心よりお礼申し上げます。

文献

- Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., . . . Belliveau, J. W. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. *Proceedings of the National Academy of Sciences*, *103*(39), 14608–14613.
- Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, *5*, D202–D212.
- Alain, C., Snyder, J. S., He, Y., & Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex*, *17*(May), 1074–1084.
- Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, *34*(5), 534–546.
- Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *NeuroReport*, *6*(1), 140–144.
- Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron*, *61*(3), 467–480.
- Attias, H., & Schreiner, C. E. (1997). Temporal low-order statistics of natural sounds. In *Advances in neural information processing systems* (Vol. 9, pp. 27–33). Cambridge, Massachusetts: MIT Press.
- Barlow, H. B. (1997). The knowledge used in vision and where it comes from. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*, 1141–1147.
- Belin, P., & Zatorre, R. J. (2000). ‘What’, ‘where’ and ‘how’ in auditory cortex. *Nature Neuroscience*, *3*(10), 965–966.
- Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*(7054), 1161–1165.
- Best, V., Gallun, F. J., Carlile, S., & Shinn-Cunningham, B. G. (2007). Binaural inter-

- ference and auditory grouping. *The Journal of the Acoustical Society of America*, *121*, 1070–1076.
- Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews: Neuroscience*, *14*(10), 693–707.
- Bizley, J. K., Walker, K. M. M., Nodal, F. R., King, A. J., & Schnupp, J. W. H. (2013). Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Current Biology*, *23*(7), 620–625.
- Bizley, J. K., Walker, K. M. M., Silverman, B. W., King, A. J., & Schnupp, J. W. H. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *The Journal of Neuroscience*, *29*(7), 2064–2075.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (Vol. 77). Cambridge, Massachusetts: MIT Press.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, Massachusetts: MIT Press.
- Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences*, *107*(1), 361–365.
- Carlyon, R. P. (1991). Discriminating between coherent and incoherent frequency modulation of complex tones. *The Journal of the Acoustical Society of America*, *89*, 329–340.
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*(10), 465–471.
- Carlyon, R. P., Cusack, R., Foxton, J., & Robertson, I. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 115–127.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979.
- Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams:

- Where, when and how? *Brain and Language*, 127(2), 251–263.
- Culling, J. F., & Darwin, C. J. (1993). Perceptual separation of simultaneous vowels: within and across-formant grouping by F0. *The Journal of the Acoustical Society of America*, 93, 3454–3467.
- Culling, J. F., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay. *The Journal of the Acoustical Society of America*, 98(2 Pt 1), 785–797.
- Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience*, 17(4), 641–651.
- Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of the Association for Research in Otolaryngology*, 30(4), 643–656.
- Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, 1(9), 327–333.
- Darwin, C. J. (2005). Simultaneous grouping and auditory continuity. *Perception & Psychophysics*, 67(8), 1384–1390.
- Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363, 1011–1021.
- Darwin, C. J., & Hukin, R. W. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *The Journal of the Acoustical Society of America*, 102(4), 2316–2324.
- Darwin, C. J., & Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *The Journal of the Acoustical Society of America*, 103(2), 1080–1084.
- Darwin, C. J., Hukin, R. W., & Al-Khatib, B. Y. (1995). Grouping in pitch perception: evidence for sequential constraints. *The Journal of the Acoustical Society of America*, 98(2 Pt 1), 880–885.
- De Boer, J., & Thornton, A. R. D. (2007). Effect of subject task on contralateral

- suppression of click evoked otoacoustic emissions. *Hearing Research*, 233, 117–123.
- Degerman, A., Rinne, T., Salmi, J., Salonen, O., & Alho, K. (2006). Selective attention to sound location or pitch studied with fMRI. *Brain Research*, 1077(1), 123–134.
- Delano, P. H., Elgueda, D., Hamame, C. M., & Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *The Journal of Neuroscience*, 27(15), 4146–4153.
- Demany, L., & Ramos, C. (2005). On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *The Journal of the Acoustical Society of America*, 117(2), 833.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Reviews in Neuroscience*, 18, 193–222.
- Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, 107(1), 78–89.
- Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, 92, 53–78.
- Duncan, J. (2006). EPS Mid-Career Award 2004: Brain mechanisms of attention. *Quarterly Journal of Experimental Psychology*, 59(1), 2–27.
- Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., & Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*, 61(2), 317–329.
- Fishman, Y. I., Arezzo, J., & Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *Journal of Acoustical Society of America*, 116, 1656–1670.
- Fishman, Y. I., Reser, D., Arezzo, J., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research*, 151(14), 167–187.

- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science*, *322*(5903), 970–973.
- Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, *21*(7), 920–925.
- Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, *6*(11), 1216–1223.
- Garcia, D., Hall, D. A., & Plack, C. J. (2010). The effect of stimulus context on pitch representations in the human auditory cortex. *NeuroImage*, *51*(2), 808–816.
- Gardner, R. B., & Wilson, J. P. (1979). Evidence for direction-specific channels in the processing of frequency modulation. *Journal of the Acoustical Society of America*, *66*(3), 704–709.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley and Sons.
- Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*, 1121–1127.
- Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., ... Howard, M. A. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology*, *20*(12), 1128–1132.
- Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Neuroscience*, *5*(November), 887–892.
- Guinan Jr., J. J. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear and Hearing*, *27*(6), 589–607.
- Gutschalk, A., Micheyl, C., Melcher, J. R., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *The Journal of Neuroscience*, *25*(22), 5382–5388.

- Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, *19*(3), 576–585.
- Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, *20*(March), 583–590.
- Hukin, R. W., & Darwin, C. J. (1995). Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Perception & Psychophysics*, *57*(2), 191–196.
- Ison, M. J., & Quiroga, R. Q. (2008). Selectivity and invariance for visual object perception. *Frontiers in Bioscience*, *13*, 4889–4903.
- Kaas, J. H., & Hackett, T. A. (1999). ‘What’ and ‘where’ processing in auditory cortex. *Nature Neuroscience*, *2*(12), 1045–1047.
- Kalar, D., Garrigan, P., Wickens, T., Hilger, J., & Kellman, P. (2010). A unified model of illusory and occluded contour interpolation. *Vision Research*, *50*(3), 284–299.
- Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea. *Rivista di Psicologia*, *49*(1), 7–30.
- Kanizsa, G. (1976). Subjective contours. *Scientific American*, *234*, 48–52.
- Kanwal, J., Medvedev, A., & Micheyl, C. (2003). Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat’s natural environment. *Network*, *14*(413-435), 15.
- Kashino, M., & Kondo, H. M. (2012). Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1591), 977–987.
- Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Reviews in Neuroscience*, *23*, 315–341.
- Kayahara, T., & Sato, T. (2001). Two types of representation for uni-directional frequency change revealed by the changing frequency aftereffect. *Acoustical Science and Technology*, *22*(1), 53–55.

- Knudsen, E. I. (2007). Fundamental components of attention. *Annual Reviews in Neuroscience*, 30, 57–78.
- Kobayashi, M., Osada, Y., & Kashino, M. (2007). The effect of a flashing visual stimulus on the auditory continuity illusion. *Perception & Psychophysics*, 69(3), 393–399.
- Kondo, H. M., & Kashino, M. (2007). Neural mechanisms of auditory awareness underlying verbal transformations. *NeuroImage*, 36(1), 123–130.
- Kondo, H. M., & Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *Journal of Neuroscience*, 29(40), 12695–12701.
- Kubovy, M., & Valkenburg, D. V. (2001). Auditory and visual objects. *Cognition*, 80, 97–126.
- Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biological Psychology*, 74(1), 1–19.
- Kumar, S., Stephan, K. E., Warren, J., Friston, K., & Griffiths, T. (2007). Hierarchical processing of auditory objects in humans. *PLoS Computational Biology*, 3(6), e100.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, 9(2), 75–82.
- Lee, A. K. C., Rajaram, S., Xia, J., Bharadwaj, H., Larson, E., Hämäläinen, M. S., & Shinn-Cunningham, B. G. (2012). Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Frontiers in Neuroscience*, 6, 190.
- Lee, A. K. C., & Shinn-Cunningham, B. G. (2008). Effects of reverberant spatial cues on attention-dependent object formation. *Journal of the Association for Research in Otolaryngology*, 9(1), 150–160.
- Lee, C.-C., & Middlebrooks, J. C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. *Nature Neuroscience*, 14(1), 108–114.

- Logothetis, N. K., & Leopold, D. A. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences*, 3(7), 254–264.
- Long, G. M., & Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychological Bulletin*, 130(5), 748–768.
- Lyzenga, J., Carlyon, R. P., & Moore, B. C. J. (2005). Dynamic aspects of the continuity illusion: Perception of level and of the depth, rate, and phase of modulation. *Hearing Research*, 210, 30–41.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, Massachusetts: MIT Press.
- McAdams, S., Botte, M.-C., & Drake, C. (1998). Auditory continuity and loudness computation. *The Journal of the Acoustical Society of America*, 103(3), 1580–1591.
- McDermott, J. H. (2009). The cocktail party problem. *Current Biology*, 19(22), R1024–1027.
- McDermott, J. H., & Oxenham, A. J. (2008). Spectral completion of partially masked sounds. *Proceedings of the National Academy of Sciences*, 105(15), 5939–5944.
- McDermott, J. H., Wroblewski, D., & Oxenham, A. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences*, 108(3), 1188–1193.
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485, 233–236.
- Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., ... Wilson, E. C. (2007). The role of auditory cortex in the formation of auditory streams. *Hearing Research*, 229(1-2), 116–31.
- Micheyl, C., Kreft, H., Shamma, S., & Oxenham, A. J. (2013). Temporal coherence versus harmonicity in auditory stream formation. *The Journal of the Acoustical Society of America*, 133(3), EL188–194.
- Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual organiza-

- tion of tone sequences in the auditory cortex of awake macaques. *Neuron*, 48(1), 139–148.
- Miller, C. T., Dibble, E., & Hauser, M. D. (2001). Amodal completion of acoustic signals by a nonhuman primate. *Nature Neuroscience*, 4(8), 783–784.
- Miller, S. D., Litovsky, R. Y., & Kluender, K. R. (2009). Predicting echo thresholds from speech onset characteristics. *The Journal of the Acoustical Society of America*(April), 134–140.
- Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing* (6th ed.). Bingley, UK: Emerald Group Publishing.
- Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United With Acustica*, 88, 320–332.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425.
- Niwa, M., O'Connor, K. N., Sutter, M. L., & Johnson, J. S. (2012). Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *The Journal of Neuroscience*, 32(27), 9323–9334.
- Palmer, A. R. (1995). Neural signal processing. In B. C. J. Moore (Ed.), *Hearing* (pp. 75–121). San Diego: Academic Press.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767–776.
- Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *The Journal of Neuroscience*, 24(30), 6810–6815.
- Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, 7, 658–663.

- Petkov, C. I., O'Connor, K. N., & Sutter, M. L. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, 54(1), 153–165.
- Petkov, C. I., & Sutter, M. L. (2011). Evolutionary conservation and neuronal mechanisms of auditory perceptual restoration. *Hearing Research*, 271, 54–65.
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch Negativity: Different water in the same river. *Audiology & Neuro-Otology*, 5, 111–139.
- Plack, C. J. (2010). *The Oxford handbook of auditory science: The auditory brain* (A. R. Palmer & A. Rees, Eds.). New York: Oxford University Press.
- Pressnitzer, D., Sayles, M., Micheyl, C., & Winter, I. M. I. (2008). Perceptual organization of sound begins in the auditory periphery. *Current Biology*, 18(15), 1124–1128.
- Pressnitzer, D., Suied, C., & Shamma, S. A. (2011). Auditory scene analysis : the sweet music of ambiguity. *Frontiers in Human Neuroscience*, 5(December), 1–11.
- Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience*, 4, 7.
- Rees, A., & Palmer, A. (2010). *The Oxford Handbook of Auditory Science: The Auditory Brain* (1st ed.). Oxford University Press, USA.
- Riecke, L., Esposito, F., Bonte, M., & Formisano, E. (2009). Hearing illusory sounds in noise: The timing of sensory-perceptual transformations in auditory cortex. *Neuron*, 64, 550–561.
- Riecke, L., Micheyl, C., & Oxenham, A. J. (2012). Global not local masker features govern the auditory continuity illusion. *The Journal of Neuroscience*, 32(13), 4660–4664.
- Riecke, L., van Opstal, A. J., Goebel, R., & Formisano, E. (2007). Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *Trends in Cognitive Sciences*, 27(46), 12684–12689.
- Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, 12, 162–168.

- Rinne, T., Stecker, G. C., Kang, X., Yund, E. W., Herron, T. J., & Woods, D. L. (2007). Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *NeuroReport*, 18(13), 1311–1314.
- Roberts, B., & Bailey, P. J. (1996). Regularity of spectral pattern and its effects on the perceptual fusion of harmonics. *Perception & Psychophysics*, 58, 289–299.
- Roberts, B., & Brunstrom, J. M. (1998). Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *The Journal of the Acoustical Society of America*, 104, 2326–2338.
- Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: two fundamental challenges for visual processing. *Current Opinion in Neurobiology*, 20(3), 382–388.
- Salmi, J., Rinne, T., Degerman, A., Salonen, O., & Alho, K. (2007). Orienting and maintenance of spatial attention in audition and vision: an event-related brain potential study. *The European Journal of Neuroscience*, 212(12), 3725–3733.
- Sato, M., Basirat, A., & Schwartz, J.-L. (2007). Visual contribution to the multistable perception of speech. *Perception & Psychophysics*, 69, 1360–1372.
- Schadwinkel, S., & Gutschalk, A. (2011). Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *Journal of Neurophysiology*, 105, 1977–1983.
- Schebesch, G., Lingner, A., Firzlaff, U., Wiegrebe, L., & Grothe, B. (2010). Perception and neural representation of size-variant human vowels in the Mongolian gerbil (*Meriones unguiculatus*). *Hearing Research*, 261(1-2), 1–8.
- Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46, 561–584.
- Shamma, S. A. (2011). Hearing impairments hidden in normal listeners. *Proceedings of the National Academy of Sciences*, 108(37), 16139–16140.
- Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, 20(3), 361–366.

- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(March), 182–186.
- Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, 12(4), 283–299.
- Singh, M., Wagemans, J., von der Heydt, R., Peterson, M. A., Kubovy, M., Palmer, S. E., & Elder, J. H. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138, 1172–1217.
- Singh, N. C., & Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America*, 114(6), 3394–3411.
- Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of cognitive neuroscience*, 18(1), 1–13.
- Snyder, J. S., Carter, O. L., Hannon, E. E., & Alain, C. (2009). Adaptation reveals multiple levels of representation in auditory stream segregation. *The Journal of the Association for Research in Otolaryngology*, 35(4), 1232–1244.
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, 19(6), 498–502.
- Stecker, G. C., & Middlebrooks, J. C. (2003). Distributed coding of sound locations in the auditory cortex. *Biological Cybernetics*, 89, 341–349.
- Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences*, 13(7), 310–318.
- Sugita, Y. (1997). Neuronal correlates of auditory induction in the cat cortex. *NeuroReport*, 8(5), 1155–1159.
- Sussman, E. S., Horváth, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, 69(1), 136–152.

- Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, 10, 502–511.
- van Noorden, L. P. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. *The Journal of the Acoustical Society of America*, 61, 1041–1045.
- Versnel, H., & Shamma, S. A. (1998). Spectral-ripple representation of steady-state vowels in primary auditory cortex. *The Journal of the Acoustical Society of America*, 103(5 Pt 1), 2502–2514.
- Voss, R. F., & Clarke, J. (1975). ‘1/f noises’ in music and speech. *Nature*, 258(November), 317–318.
- Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *The Journal of Neuroscience*, 23(13), 5799–5804.
- Warren, R. M. (2008). *Auditory Perception: An Analysis and Synthesis* (3rd ed.). Cambridge, England: Cambridge University Press.
- Warren, R. M., Bashford, J. A., Healy, E. W., & Brubaker, B. S. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Perception & Psychophysics*, 55(3), 313–322.
- Warren, R. M., & Gregory, R. L. (1958). An auditory analogue of the visual reversible figure. *The American Journal of Psychology*, 71(3), 612–613.
- Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176(39), 1149–1151.
- Webster, M. A. (2011). Adaptation and visual coding. *Journal of Vision*, 11(5), 1–23.
- Wiley, R. H., & Richards, D. G. (1978). Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations. *Behavioral Ecology and Sociobiology*, 3, 69–94.
- Winkler, I., Denham, S., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*,

13(12), 532–540.

- Winkler, I., Takegata, R., & Sussman, E. S. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research*, 25(1), 291–299.
- Woldorff, M. G., Gallen, C. C., Hampson, S., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, 90(18), 8722–8726.
- Woods, D. L., Alho, K., & Algazi, A. (1992). Intermodal selective attention I: Effects on event-related potentials to lateralized auditory and visual stimuli. *Electroencephalography and Clinical Neurophysiology*, 82(5), 341–355.
- Woods, D. L., Alho, K., & Algazi, A. (1993). Intermodal selective attention: Evidence for processing in tonotopic auditory fields. *Psychophysiology*, 30(3), 287–295.
- Zatorre, R. J., Mondor, T. A., & Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage*, 10(5), 544–554.

付録

本研究に関する論文

1. Masutomi, K., & Kashino, M. (2013). Frequency-change aftereffect produced by adaptation to real and illusory unidirectional frequency sweeps. *The Journal of the Acoustical Society of America*, 134, EL14-EL18.
doi: <http://dx.doi.org/10.1121/1.4807304>
2. Masutomi, K., Barascud, N., Kashino, M., McDermott, J. H., & Chait, M. (2015). Sound Segregation via Embedded Repetition Is Robust to Inattention. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: <http://dx.doi.org/10.1037/xhp0000147>

本研究に関する国際会議発表

1. Masutomi, K., Kashino, M. (2011). Frequency Change Aftereffect by Adaptation to Real and Illusory Sweeps. Association for Research in Otolaryngology MidWinter Research Meeting, 969.
2. Masutomi, K., Barascud, N., Overath, T., Kashino, M., McDermott, J. H., Chait, M. (2012). Does Recovering Sound Sources from Embedded Repetition Require Directed Attention?. Association for Research in Otolaryngology MidWinter Research Meeting, 1077.
3. Masutomi, K., Barascud, N., Overath, T., Kashino, M., McDermott, J. H., Chait, M. (2013). Recovering Sound Sources from Embedded Repetition And Directed

Attention: Effect of Spectral Information on Sound Segregation. Association for Research in Otolaryngology MidWinter Research Meeting, 197.

本研究に関する国内学会発表

1. 益富恵子, 柏野牧夫. (2011). 選択的順応を用いた連続聴錯覚形成過程の検討. 日本音響学会 2011 年春期研究発表会, 1-6-9.
2. 益富 恵子, Nicolas Barascud, Tobias Overath, 柏野 牧夫, Josh H. McDermott, Maria Chait. (2012). 混合音中で複数回出現する新規音の検出に選択的注意は必要か?. 日本音響学会 聴覚研究会 (電子情報通信学会 音声研究会との共同開催), 16.
3. 益富 恵子, Nicolas Barascud, Tobias Overath, 柏野 牧夫, Josh H. McDermott, Maria Chait. (2013). 混合音中で複数回出現する未知の音の検出に選択的注意は必要か?. 27, 情報処理学会 音楽情報科学研究会 音学シンポジウム 2013.
4. 益富 恵子, Nicolas Barascud, Tobias Overath, 柏野 牧夫, Josh H. McDermott, Maria Chait. (2013). 混合音中で複数回出現する未知の音の検出に選択的注意は必要か. 日本神経科学大会 (日本神経化学会大会, 日本神経回路学会大会と合同開催) .