

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？

益富 恵子^{1,a)} Nicolas Barascud^{3,b)} Tobias Overath^{3,c)} 柏野 牧夫^{2,1,d)} Josh H. McDermott^{4,e)}
Maria Chait^{3,f)}

概要：聴覚系は、たとえ全く未知の音であっても、その音が繰り返し現れるならば、混合音の中からひとつの音のまとまりとして検出することが可能である [McDermott, *et al.* (2011) PNAS. 108(3):1188- 93]. 本研究ではこの「繰り返し」手がかりに基づく音の分離・検出が、音に対する選択的注意を向けない場合にも可能であるかを調べた。実験では「繰り返し」に基づく音の検出課題と、難易度の高い視覚課題との二重課題を設定し、実験参加者が視覚課題に集中した際の音の検出課題の成績を測定した。実験の結果、参加者は混合音の再生中に視覚課題に集中していたとしても、チャンスレベル以上の精度で対象の音を検出できることが分かった。この結果は「繰り返し」手がかりに基づく音の検出に、認知・判断に係る処理系以前の、自動性の強い処理段階が関与する可能性を示唆する。

キーワード：聴覚情景分析, カクテルパーティー効果, 同時音グルーピング, 選択的注意

1. はじめに

私たちは、パーティー会場のようなたくさんの音が同時に存在する環境で、友人の声やグラスの触れ合う音などの個々の音源から発せられた音に気づき、聞き分けることができる。しかし、耳に届く音は全ての音源からの音が混じり合った複雑な時系列信号であり、このような信号から個々の音源の情報を取り出すことは計算上非常に難しい。この問題は「カクテルパーティ問題」として知られている [1].

これまで、人間の聴覚系がどのような方法を用いて音信号の分離を実現しているのか、また信号中のどのような特徴が分離に重要であるかが調べられてきた [2], [3], [4]. あ

わせて、こうした音の分離処理に関する聴覚系の機能を働かせるにあたって、人間が「意識的に」音を聴く必要があるか、すなわち音に対して選択的注意を向ける必要があるかについて論じられてきた [6], [7], [8], [9].

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 [5]. 刺激音は、時間・周波数手がかりが非常に少ない人工的な音であった。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」という状況そのものが、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。

本研究の目的は、McDermott らが発見した「繰り返し」手がかりに基づく音の分離が行われる際、聴取者が音に対して選択的注意を払う必要があるかを調べることである。実験では McDermott らによるオリジナルの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。実験参加者は視覚による「おとり」課題と聴覚課題を同時に行った。「おとり」課題には、注意の負荷量が異なる「Low-load 課題（以下、Lo

¹ 東京工業大学
Tokyo Institute of Technology, Nagatsuta-cho 4259, Midori-ku, Yokohama-shi, 226-8503 Japan

² UCL Ear Institute
332 Gray's Inn Road, London, WC1X 8EE

³ NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corporation, Morinosato-Wakamiya 3-1, Atsugi-shi, 243-0198 Japan

⁴ Massachusetts Institute of Technology
77 Massachusetts Avenue, 46-4065, Cambridge MA 02139

a) masutomi@u.ip.titech.ac.jp

b) n.barascud@ucl.ac.uk

c) t.overath@ucl.ac.uk

d) makio.kashino@lab.ntt.co.jp

e) jhm@cns.nyu.edu

f) m.chait@ucl.ac.uk

課題と記載)「High-load 課題 (以下, Hi 課題と記載)」を用意し, これら二つの課題と同時に行われる聴覚課題の成績を比較することで, 注意の状況が聴覚課題の成績に与える影響を調べた. 参加者らは, 「おとり」課題に集中するように動機づけられ, 聴覚課題の回答に際して, 見当がつかない場合は推測を行うように求められた.

2. 実験手続 (実験 1 と 2 に共通)

2.1 音刺激の作成

音刺激の作成には, McDermott らと同様の手続を用いた [5]. 実験参加者が周波数配置の規則性の手がかりや発音同時性の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, これに自然音の持つ時間相関, 周波数相関の性質を持たせることで作成した. 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がみられる [10]. 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・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. なお, この方法によって作成した音刺激の例は, Web サイト http://www.cns.nyu.edu/~jhm/source_repetition/ にて確認できる.

2.2 聴覚課題

聴覚課題は, オリジナルの実験と同様の刺激・手続きである Segregation 課題 (以下 Sg 課題と表記) と, コントロール条件の Identification 課題 (以下 Id 課題と表記) の 2 種類であった. Id 課題は, Sg 課題と類似しているが「繰り返し」に基づく音の検出課題ではない点が異なる. Sg 課題では, 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 (図 1 A). 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. Id 課題では, 参加者はターゲット音と妨害音との混合音を繰り返し 10 回聴いた後, 検査音を聞き, 検査音が先に呈示された混合音と同じものか, 偽ターゲット音と妨害音との混合音かを回答した (図 1 B). 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.

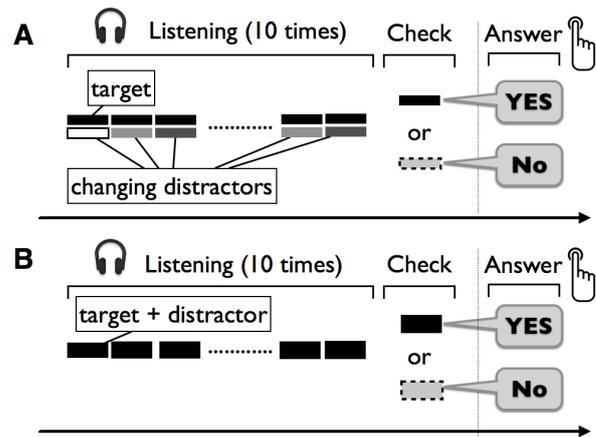


図 1 聴覚課題の説明図. A: Sg 課題 (1 試行) の流れ. 各長方形がひとつの音刺激をあらわす. B: Id 課題の流れ.

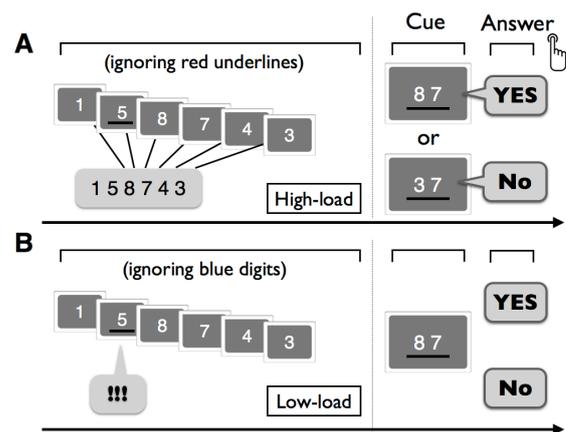


図 2 視覚課題の説明図. A: Hi 課題 (1 試行) の例. 参加者は数字列を記憶した後, cue 画面に表示される 2 つの数字が記憶した数字列の一部であるか否かを回答した. B: Lo 課題の例. 参加者は数字列の呈示中に赤い線が出現したか否かを, cue 画面の呈示時に回答した.

3. 実験 1: 数字記憶課題

実験 1 では「おとり」課題として, 記憶能力を必要とする視覚課題が用いられた.

3.1 視覚課題

視覚刺激列の作成・呈示方法は Hi 課題と Lo 課題に共通であった. コンピュータ画面には 1 から 8 までの数字が, 画面中央に 1 つずつ順に, 5 つあるいは 6 つ表示された. 呈示する数字と数字の順序は, 試行ごとにランダムに決定した. また全試行中 50% の割合で, ランダムに選ばれた数字の下に赤い線が 1 回出現した. Hi 課題では, 参加者は数字を順に記憶し, その後 cue 画面に表示される 2 つの数字が, 記憶した数字列の一部であるか否かを Yes・No のボタンで回答した (図 2A). Lo 課題では, 参加者は数字を全て無視し, 赤い線が数列中にあらわれたか否かを cue の表示中に回答した (図 2B). 聴覚刺激は視覚刺激の呈示中に再生された. 参

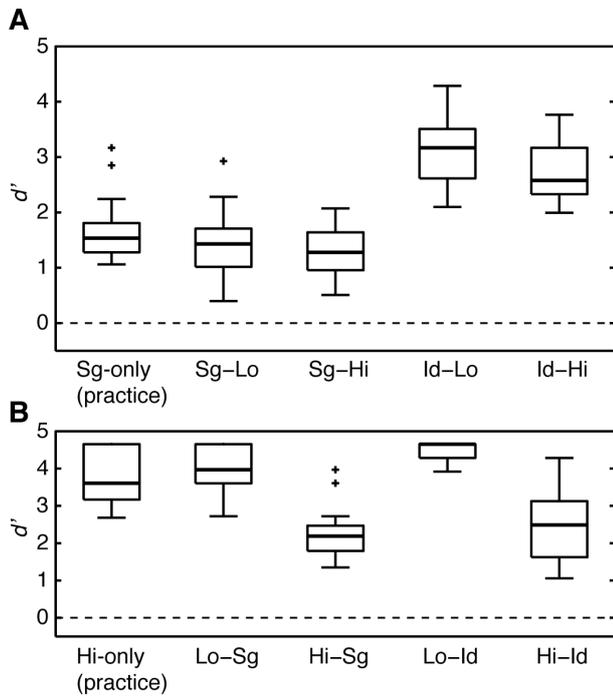


図3 実験1の結果。聴覚・視覚課題の各組み合わせにおける、聴覚課題の識別率 (d')。A: 聴覚課題の成績。B: 視覚課題の成績。箱内の横線は中央値、箱の端は第1四分位点と第3四分位点、ひげは最大値と最小値、+はひげの範囲外の値、グラフ下部の点線はチャンスレベルを示す。

加者はその後検査音を聴いて聴覚課題に回答し、続いて視覚課題に回答した。参加者は視覚と聴覚課題の全ての組み合わせ [Sg-Lo 条件, Sg-Hi 条件, Id-Lo 条件, Id-Hi 条件] を各2ブロックずつ、計8ブロック行った。1ブロックは40試行からなり、開始から終了までに約8分を要した。実験には17名の男女が参加した。全ての参加者は視覚・聴覚ともに正常であった。刺激の作成と呈示および参加者の応答の取得には MacBook Air (Apple Inc., USA), MATLAB 7.6.0 (The MathWorks Inc., USA), オーディオインタフェース EDIROL UA-25EX (Roland Corporation, Japan) とヘッドホン HD 595 (Sennheiser Electronic, Germany) を用いた。全ての音刺激は、60 - 70 dB SPL の範囲の、各参加者が課題を行いやすい音圧レベルで呈示した。

3.2 結果と考察

参加者らの聴覚・視覚課題の正解率を、信号検出理論 [11] における弁別力の指標 d' であらわした。図3に各聴覚・視覚課題の組み合わせにおける、聴覚課題と視覚課題の d' の値の中央値とそのばらつきを示す。

参加者が Sg 課題, Id 課題にチャンスレベル以上の精度で回答可能であったかを調べるため、Sg-Lo 条件と Sg-Hi 条件での d' の値について、それぞれ t 検定を行った。結果 Sg, Id 両課題の d' は、どちらの視覚課題との組み合わせについても全てチャンスレベル ($d' = 0$) より有意に高かった [Sg-Lo 条件: $t(17) = 9.9, p < 0.001$], [Sg-Hi 条件:

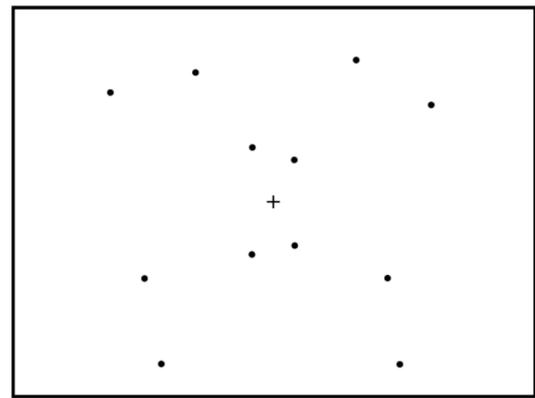


図4 実験2で用いた視覚課題(複数物体追跡課題)の開始画面の一例。(文献 [12] より引用)。中央の+マークは注視点。

$t(17) = 11.37, p < 0.001$], [Id-Lo 条件: $t(17) = 20.18, p < 0.001$], [Id-Hi 条件: $t(17) = 22.12, p < 0.001$].

次に、聴覚課題の成績について聴覚刺激と視覚刺激の種類を要因とする分散分析を行った結果、聴覚課題の種類の主効果 [$F(1, 16) = 174.4, p < 0.001$] と視覚課題の難易度による主効果がみられた [$F(1, 16) = 12.75, p < 0.01$]. 一方、聴覚課題と視覚課題の種類の変異作用はみられなかった [$F(1, 16) = 1.83, p = 0.2$]. なお、Huynh-Feldt の ϵ による補正を自由度が1より大きい反復測定の場合に用いた。

以上の結果から、聴覚課題の成績は、Hi, Lo どちらの視覚刺激と同時に進行される際にもチャンスレベルよりも有意に高いことが分かった。Hi 課題を行っている際には、Lo 課題よりもわずかに成績値が低い傾向がみられるものの、減少量はわずかであることがグラフから分かる。このことから、Sg 課題は、負荷量の大きい Hi 課題を行っているときでも十分に可能であることが示唆された。

4. 実験2: 物体追跡課題

実験2では、注意を必要とするが音による記憶を必要としない、複数物体の追跡課題 [12] を「おとり」課題として用いた。

4.1 課題

図4に実験で用いた物体追跡課題の一例を示す。まず12個のドットが画面にあらわれ、うち4個のドットが「ターゲット」として赤で示された。1秒後全てのドットは画面上をランダムに動き始め、参加者は聴覚課題の音が再生される3.7秒の間、最初に赤で示されていたドットを追跡した。赤いドットの色は1秒後黒に戻った。その後全てのドットは停止し、参加者は最初に「ターゲット」として指定されていたドットをマウスで選択した。その他ドットの動作に関する規則は文献 [12] 内の各実験手続きと同様とした。ドットのスピードは Hi 条件では5フレーム毎秒、Lo 条件では2フレーム毎秒に設定した。

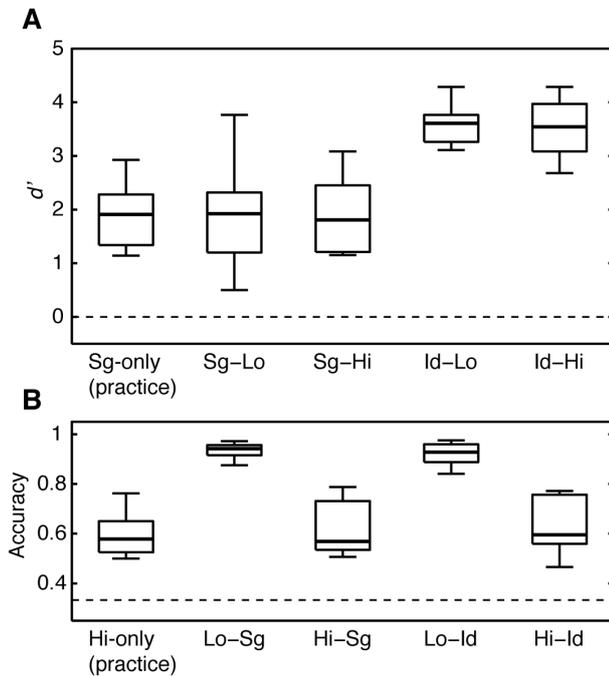


図5 実験2の結果。A: 聴覚課題の成績。B: 視覚課題の成績。

参加者は視覚と聴覚課題の全ての組み合わせを各2ブロックずつ、計8ブロック行った。1ブロックは40試行からなり、開始から終了までに約10分を要した。実験には正常聴力を持つ10名の男女が参加した。その他の手続きは実験1と同様とした。

4.2 結果と考察

図5に各聴覚・視覚課題の組み合わせにおける、聴覚課題と視覚課題の d' の値の中央値とそのばらつきを示す。

参加者が Sg 課題, Id 課題にチャンスレベル以上の精度で回答可能であったかを調べるため、Sg-Lo 条件と Sg-Hi 条件での d' の値について、それぞれ t 検定を行った。結果 Sg, Id 両課題の d' は、どちらの視覚課題との組み合わせについても全てチャンスレベル ($d' = 0$) より有意に高かった [Sg-Lo 条件: $t(9) = 6.13, p < 0.001$], [Sg-Hi 条件: $t(9) = 8.49, p < 0.001$], [Id-Lo 条件: $t(9) = 32.93, p < 0.001$], [Id-Hi 条件: $t(9) = 18.31, p < 0.001$].

次に、聴覚課題の成績について聴覚刺激と視覚刺激の種類を要因とする分散分析を行った結果、聴覚課題の種類の主効果 [$F(1, 9) = 59.3, p < 0.001$] がみられた。一方、視覚課題の難易度による主効果と、聴覚課題と視覚課題の種類との交互作用はみられなかった [視覚課題の主効果 $F(1, 9) = 0.02, p = 0.88$; 交互作用 $F(1, 9) = 0.18, p = 0.68$]. Huynh-Feldt の ϵ による補正を自由度が1より大きい反復測定の場合の F 値の検定に用いた。

以上の結果から、高負荷の物体追跡課題を聴覚課題と同時に課した場合であっても、聴覚課題の成績は低負荷の視覚課題を同時に行っているときと大きな差異がないことが分かった。

5. まとめ

実験1, 2の結果から、参加者は難易度の高い視覚課題を同時に行っている場合であっても、呈示された混合音の繰り返しからターゲット音を検出可能であることが示された。視覚課題の成績について着目したとき、参加者は Lo, Hi の両視覚課題ともにチャンスレベル以上の高い正解率を示していた。また、Lo 課題の正解率はほぼ全問正解に近い値が得られている。選択的注意の研究では、参加者の注意が完全に対象の課題からそらされていることを保証することは難しい。しかし前述の結果は、参加者が十分に視覚刺激に対して注意を向け課題を行っていたことの間接的な保証となる。

この結果は、McDermott らが発見した「繰り返し」手がかりに基づく音の検出を行う聴覚系の機能は、参加者が視覚課題に集中し、音に対する注意をそらされていたとしても活動可能であること、またこの機能が認知・判断や多感覚との統合に関係する比較的高次の情報処理系ではなく、情報の入力側に近い、自動性の高い比較的低次の段階にて実現されていることを示唆する。

参考文献

- [1] E. Cherry, "Some experiments on the recognition of speech, with one and with two ears", *J. Acoust. Soc. Am.*, vol.25, no.5, pp.975-979, 1953.
- [2] A. S. Bregman, "Auditory Scene Analysis: The Perceptual Organization of Sound", MIT Press, 1990.
- [3] R. Carlyon, "How the brain separates sounds", *Trends Cogn. Sci.*, vol.8, no.10, pp.465-471, 2004.
- [4] J. H. McDermott, "The cocktail party problem", *Curr. Biol.*, vol.19, no.22, pp.1024-1027, 2009.
- [5] J. H. McDermott, D. Wroblewski and A. J. Oxenham, "Recovering sound sources from embedded repetition", *Proc. Natl. Acad. Sci.*, vol.108, no.3, pp.1188-1193, 2011.
- [6] R. P. Carlyon et al., "Effects of attention and unilateral neglect on auditory stream segregation", *J. Exp. Psychol. Human.*, Vol.27, no.1, pp.115-127, 2001.
- [7] C. Alain and A. Izenberg, "Effects of attentional load on auditory scene analysis", *J. Cognitive Neurosci.*, Vol. 15, no.7, pp.1063-1073, 2003.
- [8] W. J. Macken et al., "Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory", *J. Exp. Psychol. Human.*, Vol.29, no.1, pp.43-51, 2003.
- [9] E. S. Sussman et al., "The role of attention in the formation of auditory streams", *Percept. Psychophys.*, Vol.69, no.1, pp.136-152, 2007.
- [10] R. F. Voss and J. Clarke, "'1/f noises' in music and speech", *Nature*, vol.258, Nov. pp.317-318, 1975.
- [11] D. M. Green and J. A. Swets, "Signal Detection Theory and Psychophysics", Wiley, 1966.
- [12] S.L. Franconeri, S.V. Jonathan, and J.M. Scimeca, "Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity", *Psychol. Sci.*, Vol.21, no.7, pp.920-925, 2010.