

## 音脈のリセットに対する音源移動と頭部運動の影響

近藤 洋史<sup>†1</sup> Daniel Pressnitzer<sup>†2, †3</sup> 戸嶋 巖樹<sup>†1</sup> 柏野 牧夫<sup>†1, †4</sup>

聴覚情景分析では、入力された音響情報を知覚的な音脈として解析することが重要になる。これまでの研究で、音脈は瞬時に形成されるものではないことが示されてきた。音脈は、時間とともに形成され、情景の突然の変化によってリセットされる。本研究では、音脈の知覚変化に対して外因的な音源移動と内因的な頭部運動がどのように影響するかを検討した。耳元での音響信号の変化、知覚される音源位置の変化、および自己受容感覚の変化といった頭部運動で惹起される要因を分解するために、テレプレゼンス・ロボットを用いた。その結果、音源移動と同様に、頭部運動によって知覚が分凝から群化へとリセットした。さらに、運動に関係する要因よりも、音響的な要因がリセットに強く影響することが認められた。これは、上述した要因の変化が聴取者の頭部運動によって全て説明される場合であっても、聴覚系における低次の情報処理が音脈の形成に影響していることを意味する。したがって、我々の結果は分散した神経ネットワークが知覚の体制化に寄与していることを示している。

### Effects of Sound Motion and Head Motion on the Resetting of Auditory Streaming

Hirohito M. KONDO<sup>†1</sup>, Daniel PRESSNITZER<sup>†2, †3</sup>, Iwaki TOSHIMA<sup>†1</sup>,  
and Makio KASHINO<sup>†1, †4</sup>

Auditory scene analysis needs to parse the incoming flow of acoustic information into perceptual streams, such as distinct musical melodies or sentences from a single talker. Previous studies have demonstrated that the formation of auditory streams is not instantaneous; streaming builds up over time and can be reset by sudden changes in the acoustics of the scene. Here, we examined the effect of changes induced by exogenous sound motion and voluntary head motion on streaming. A telepresence robot in a virtual reality setup was used to disentangle all potential consequences of head motion: changes in acoustic cues at the ears, changes in apparent sound location, and changes in motor or attentional processes. The results showed that head motion, as well as sound motion, induced the resetting of two streams into one stream. An additive model analysis further revealed that resetting was strongly influenced by acoustic cues and apparent sound location rather than by non-auditory factors related to head motion. Thus, low-level changes in sensory cues can affect perceptual organization, even though those changes are fully accounted for by the head motion of the listener. We consider that our results reflect a widely distributed neural architecture for the formation of auditory streams.

#### 1. はじめに

知覚系の重要な機能は入力された感覚情報から知覚的な情景を構築することである。聴覚においては、これを聴覚情景分析 [1] あるいはカクテルパーティ問題 [2] と呼んでいる。これまでの研究で、聴覚情景分析に影響する音響的な要因が同定されている。しかし、感覚情報と自己運動情報との間の相互作用は見過ごされてきた。とくに、頭部運動は音源定位の際に重要となるが、それが聴覚情景分析にどのような影響を与えるかは明らかではない。

本研究 [3] では、音脈分凝課題 [4] を用いて、聴覚情景分析における感覚と運動の相互作用を検討した。この課題における典型的な刺激では、2種類のトーンで構成されるABA-パターンの音系列が用いられる。刺激提示直後は群化した知覚 (ABA-ABA-) が生じるが、数秒ほど経過すると知覚が2つの音脈 (A-A-A- and -B---B---) に分凝する(図1)。

そして、それ以降の知覚は双安定状態となる [5]。群化から分凝への最初の知覚変化をビルドアップ (build-up) と呼称する。興味深いことに、ビルドアップした分凝知覚の状態でも刺激に無音区間を挿入したり、刺激の提示場所を変化させたりすると、群化した知覚にリセットされてしまう。あたかも、新たに得られた手がかりによって、聴覚系が知覚の体制化を促しているかのようである。

頭を動かしただけでは周囲の情景が変化しないことを我々は知っている。そのような状況で、頭部運動によって音脈のリセットは生じるのであろうか。聴取者の自己運動によって耳元での音響信号が変化する。すなわち、音源と耳元がなす角度は頭の向きで決定されるので、音脈のリセットは入力される音響信号の影響を受けるかもしれない。その一方で、自己運動は聴取者の随意運動であるから、リセットは外界の変化に依存しないと考えることもできる。この場合、頭部運動は聴覚情景分析に影響せず、リセットが抑制されるという可能性もある。

通常の場合では、聴覚の手がかりと自己運動の手がかりは完全に相関しているため、両者を分離することは本質的に難しい。そこで、我々はテレヘッド・システム [6] というバーチャル・リアリティ技術を用いて、この問題を解決

†1 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所  
NTT Communication Science laboratories, NTT Corporation

†2 Centre National de la Recherche Scientifique (CNRS)

†3 Ecole normale supérieure (ENS)

†4 東京工業大学

Tokyo Institute of Technology

した。このシステムでは、テレヘッドが聴取者の頭の動きとリアルタイムに同調して動くという特徴を備えていた。

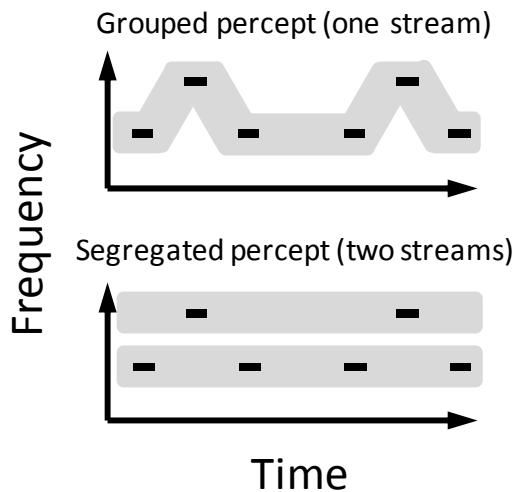


図 1 音脈の群化と分凝

Figure 1 Grouped and segregated percepts in auditory streaming

## 2. 方法

### 2.1 聴取者

10名（男女各5名，平均年齢25.3歳）が実験に参加した。

### 2.2 実験装置

聴取者は実験室の中央に着座して，ヘッドフォンから音刺激を聴取した（図2）。音刺激は無響室に設置されたテレヘッドの前方1mのスピーカから提示された。刺激はテレヘッドの耳介内にあるマイクロフォンで集音され，ヘッドフォンに伝達された。頭部運動のための視覚手がかりとして，2つの発光ダイオード（LED）が使用された。それらは聴取者の目の高さに合わせて，斜め前方2m（正中線から視角にして60°）の場所に配置された。

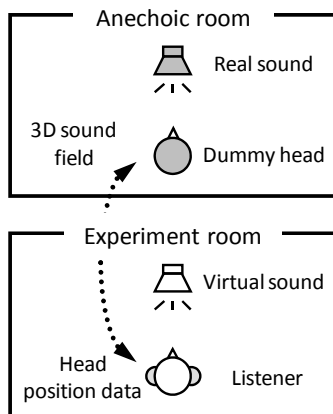


図 2 テレヘッド・システムの概要

Figure 2 Illustration of the Telehead robotic system

### 2.3 刺激と手続き

刺激は ABA-パターンの狭帯域ピンクノイズで，50 回の繰り返し（1 試行は 20 s）で構成されていた。ノイズの帯域幅は 4 半音とし，ノイズ間の中心周波数は 1 kHz，ノイズ間には 6 半音の周波数差を設けた。それぞれのノイズの持続時間は 62.5 ms，ノイズの SOA を 100 ms とした。これらの刺激特性によって，テレヘッド・システムのラインノイズ（24 dB SPL 以下）をマスクするとともに，聴取者の音源定位感が向上した。刺激の音圧は 65 dB SPL であった。

聴取者は個別に実験に参加した。まず，刺激の時間構造と音脈分凝の概念が説明された。続いて，群化の知覚（ABA-ABA-）あるいは分凝の知覚（A-A-... and -B--B--）を判断し，それをボタン押しで回答するという練習をおこなった。最後に，LED の点灯に応じて，聴取者は頭部運動をおこなった。

各試行が始まる直前に，聴取者の頭部位置の較正が実施された。試行開始とともに左右どちらかの LED が点灯し，聴取者はその方向に顔を向けた。試行の途中で LED の点灯位置が変わることもあり，聴取者は点灯した方向にできる限り素早く頭を動かすように求められた。このように教示したのは，試行間における頭部運動速度の変動を最小化し，音脈のリセットに対する効果を最大化するという目的のためであった。

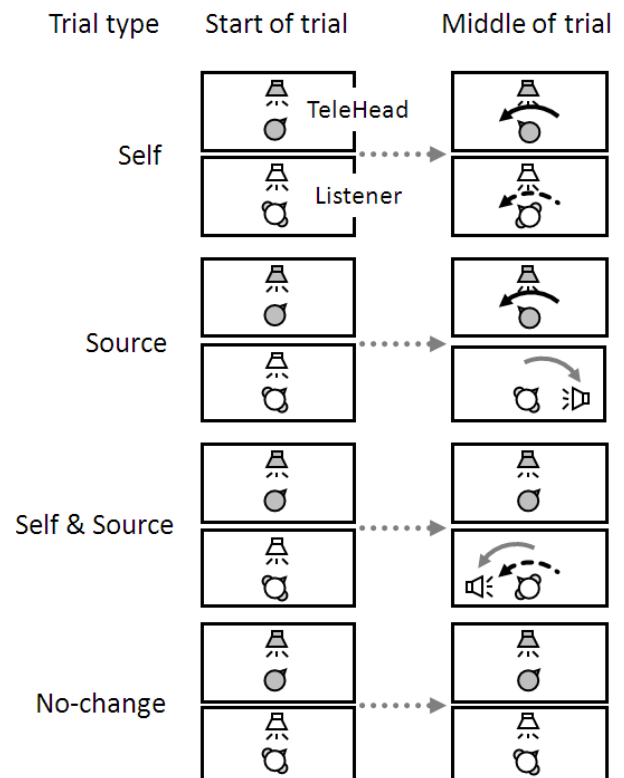


図 3 実験条件の模式図

Figure 3 Schematic representation of trial types

4つの実験条件を図3および表1に示す。Self条件では刺激提示後10sに対側のLEDが点灯した。テレヘッドは聴取者の動きと同調していた。Source条件ではLEDが同側で点灯したままで、テレヘッドだけが頭部運動をおこなった。そのため、音源が移動したように知覚された。Self & Source条件では最初だけテレヘッドが聴取者の動きに追従し、それ以降は同調しなかった。したがって、運動成分はSelf条件と同じであったが、耳元での音響的な手がかりは変化しなかった。No-change条件ではLEDの点灯位置は変化せず、聴取者もテレヘッドも動かなかった。

表1 実験条件に含まれる要因  
 Table 1 Structure of the trial types

Trial type	$\Delta A$	$\Delta S$	$\Delta H$
Self	1	0	1
Source	1	1	0
Self & Source	0	1	1
No-change	0	0	0

### 3. 結果と考察

LEDによる視覚手がかりの提示から頭部運動を終えるまでの平均時間は1.40s (SE = 0.07s)であった。運動速度に換算すると133°/sec (14°/sec)で、総じて聴取者は教示に従っていたものと考えられる。図4に分凝知覚を報告した比率の時間変化を示す。刺激提示後10sにおいて、いずれの条件でも分凝の比率は60%前後であり、条件間で有意な差は認められなかった。この結果から、視覚手がかりが与えられるまでは、遂行中の試行がどの条件に該当するのかを聴取者は推測できなかったと考えられる。

No-change条件を除く全ての条件で音脈分凝のリセットが認められた(図4)。ビルドアップ以降にリセットが最大となる時間を求めたが、いずれの条件でも13s前後となり、有意な差は認められなかった。そこで、刺激提示後10sに分凝知覚が報告された試行のみを選択して、データを標準化した(分析の詳細は[3]を参照)。そのうえで、No-change条件をベースラインとしてリセット量( $R$ )を算出した。このとき、分析対象となる時間窓を10sから16sまでの6秒間に設定した(データの時間変化を詳細に検討した結果は[7]を参照)。リセット量は、重みづけ( $K$ )された $\Delta A$ ,  $\Delta S$ , および $\Delta H$ の総和として次のように表された。

$$R = K_A \Delta A + K_S \Delta S + K_H \Delta H$$

リセットは頭部運動速度に影響を受ける可能性がある。そこで、Self試行において刺激提示後10sから13sの間にリセットが生じた試行を対象として分析をおこなった。リセットが生じた試行は66% (5.5%)、生じなかった試行は34% (5.5%)であった。しかし、両者の頭部運動速度の間に有意な差は認められなかった。

リセット量に対して、(1)音響的な手がかりの変化、(2)音像位置の変化、あるいは(3)運動に伴う自己受容感覚の変化がどの程度寄与しているのかを検討した。これらの3つの要因を含む線形モデルを仮定して(表1)、聴取者ごとに計算をおこなった。その結果、 $\Delta A = 15\%$  (2.8%),  $\Delta S = 14\%$  (2.2%),  $\Delta H = 7\%$  (3.4%)となった。リセット量に対する運動の要因の寄与率は他の要因と比較して小さかった [ $F(2,18) = 7.01, P < 0.01$ ]。この知見は先行研究[8][9][10]とも合致すると言える。

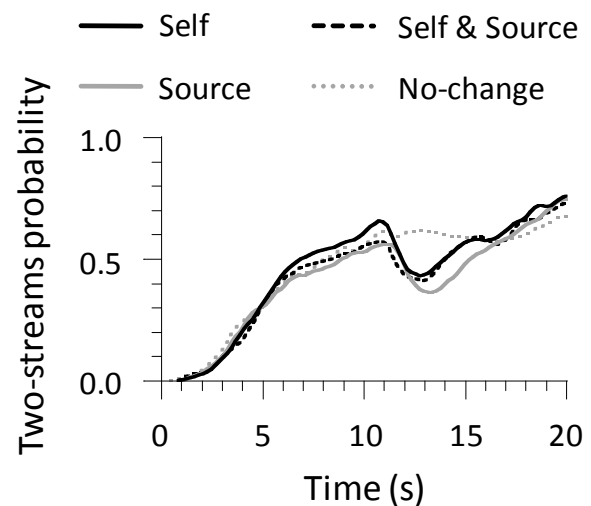


図4 音脈分凝の時間変化  
 Figure 4 Probability of two-streams responses ( $N = 10$ )

### 4. 考察

我々は、音源の急速な移動が分凝知覚のリセットをもたらすことを示した。また、耳元での音響手がかりの変化が自己運動に帰属される場合であっても、リセットは生じた。これらの結果に対して、いくつかの仮説が考えられる。

まず、LEDによる視覚手がかりによって、音刺激への注意が低下した可能性がある。しかし、補足実験によって[3]、この効果は無視できるほど小さいことが確認された。

次に、頭部運動だけが聴覚知覚の何らかの変化をもたらした可能性がある。急速な頭部運動が生じる直前に、聴覚空間が圧縮されるという報告もある[11]。しかし、本研究において、頭部運動速度はリセットに大きく影響していなかった。

音源位置を把握するために、頭部の位置情報と両耳情報は常に保持されていると考えられる。情景分析の観点からすると、頭部の位置情報は音響手がかりの変化を説明するために用いられ、結果としてリセットは抑制されるであろう。しかし、素早い頭部運動の際には、この機構がうまく働かないとすれば、音脈分凝のリセットを説明できるかもしれない。

分散した神経ネットワークが聴覚情景分析に関与していると考えれば、上述した説明に合致する。近年、音脈の神経相関は聴覚野 [12]、頭頂間溝 [13]、視床 [14]、下丘 [15]、あるいは蝸牛神経核 [16] などの多くの段階で見出されている。このネットワークの一部は、自己運動中の頭部位置情報による調整を受けていないかもしれない。したがって、聴覚系の低次の段階からリセット信号が発せられているという可能性がある。

## 5. 結語

多くの先行研究は、頭部運動によって正確な音源定位が可能になることを示してきた。それに対して、本研究は自己運動と聴覚情景分析の関係性に焦点を当てた。音脈のリセットに対する効果は、運動に伴う変化よりも音響手がかりの変化のほうが大きかった。これは、聴覚系の低次での情報処理が知覚の形成に関与していることを示唆する。

## 参考文献

- 1) Bregman, A. S.: Auditory Scene Analysis: The Perceptual Organization of Sound, MIT Press, Massachusetts (1990).
- 2) Cherry, E. C.: Some Experiments on the Recognition of Speech, with One and Two Ears, *J. Acoust. Soc. Am.*, Vol.25, pp.975-979 (1953).
- 3) Kondo, H. M., Pressnitzer, D., Toshima, I., and Kashino, M.: Effects of Self-Motion on Auditory Scene Analysis, *Proc. Natl. Acad. Sci. U. S. A.*, Vol.109, pp.6775-6780 (2012).
- 4) van Noorden, L. P. A. S.: Temporal Coherence in the Perception of Tone Sequences, Eindhoven University of Technology (PhD thesis), Eindhoven, The Netherlands (1975).
- 5) Pressnitzer, D., and Hupé, J. M.: Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization, *Curr. Biol.*, Vol.16, pp.1351-1357 (2006).
- 6) Toshima, I., Aoki, S., and Hirahara, T.: Sound Localization Using an Auditory Telepresence Robot: TeleHead II. *Presence-Teleoper. Virtual Env.*, Vol.17, pp.392-404 (2008).
- 7) 戸嶋巖樹, 近藤洋史, Daniel Pressnitzer, 柏野牧夫: 音脈の知覚変化に対する頭部運動の影響, 情報処理学会研究報告 (印刷中)
- 8) Anstis, S., and Saida, S.: Adaptation to Auditory Streaming of Frequency-Modulated Tones, *J. Exp. Psychol. Hum. Percept. Perform.*, Vol.11, pp.257-271 (1985).
- 9) Rogers, W. L., and Bregman, A. S.: Cumulation of the Tendency to Segregate Auditory Streams: Resetting by Changes in Location and Loudness, *Percept. Psychophys.*, Vol.60, pp.1216-1227 (1998).
- 10) Roberts, B., Glasberg, B. R., and Moore, B. C. J.: Effects of the Build-Up and Resetting of Auditory Stream Segregation on Temporal Discrimination, *J. Exp. Psychol. Hum. Percept. Perform.*, Vol.34, pp.992-1006 (2008).

- 11) Leung, J., Alais, D., and Carlile, S.: Compression of Auditory Space during Rapid Head Turns, *Proc. Natl. Acad. Sci. U. S. A.*, Vol.105, pp.6492-6497 (2008).
- 12) Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., Tian, B., and Wilson, E. C.: The Role of Auditory Cortex in the Formation of Auditory Streams, *Hear. Res.* Vol.229, pp.116-131 (2007).
- 13) Cusack, R.: The Intraparietal Sulcus and Perceptual Organization, *J. Cogn. Neurosci.*, Vol.17, pp.641-651 (2005).
- 14) Kondo, H. M., and Kashino, M.: Involvement of the Thalamocortical Loop in the Spontaneous Switching of Percepts in Auditory Streaming, *J. Neurosci.*, Vol.29, pp.12695-12701 (2009).
- 15) Schadwinkel, S., and Gutschalk, A.: Transient Bold Activity Locked to Perceptual Reversals of Auditory Streaming in Human Auditory Cortex and Inferior Colliculus, *J. Neurophysiol.*, Vol.105, pp.1977-1983 (2011).
- 16) Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M.: Perceptual Organization of Sound Begins in the Auditory Periphery, *Curr. Biol.*, Vol.18, pp.1124-1128 (2008).