

音脈の知覚変化に対する頭部運動の影響

戸嶋巖樹^{†1} 近藤洋史^{†1} Daniel Pressnitzer^{†2,†3} 柏野牧夫^{†1,†4}

2つの周波数のトーン(AとB)をABA_ABAの様に提示し、ABAをひとかたまりとするリズムとして知覚(群化)するか、AとBの二つの音列がある様に知覚(分凝)するかを判断する課題において、頭部運動の与える影響について分析した。この時、我々は、頭部運動を行うと音脈の知覚が分凝から群化に変わることを指摘し、頭部運動の効果を定量的に評価した(Kondo et.al. 2012)。これは頭部運動が聴覚情景分析に影響を与えることを示唆する結果である。本発表では、頭部運動を行うことが知覚に与える影響について、時間的変化を含めた分析を行った。その結果、頭部運動が知覚変化の回数に影響を与えることや、音像の移動など音響的变化による効果よりも早く知覚に影響が出ることを見出した。

The effect of head motion on auditory streaming

Iwaki Toshima^{†1} Hirohito M. Kondo^{†1}
Daniel Pressnitzer^{†2,†3} Makio Kashino^{†1,†4}

We analyzed effects of head movement on auditory streaming. The stimuli we used were consisted with two tones (A and B). They were presented as ABA ABA ..., and participants answered one stream or two streams. A previous study has demonstrated that self-motion induces some resetting of two streams to one stream, even when the acoustic scene itself has not changed (Kondo et al., 2012, PNAS). We analyzed time-series data of perceptual changes to further examine effects of head motion on auditory streaming. The results showed that the number of perceptual changes increased after head motion and effects of motor cues on streaming occurred earlier than those of acoustic cues. Thus, information processing and temporal resolution of binaural inputs may differ between the volitional control system and auditory system.

1. はじめに

急速に進歩するロボット技術を用いて、人間の性質を解明しようとする研究が多くなされている。そもそも人間の発達や知能を議論するにおいて、外界という未知なるものを如何にロボット(即ちシミュレートされた人間)に理解させるかということは大問題である。そこで、身体を通じた外界の理解に代表される、いわゆる「身体性」が提唱されて久しい[1]。これは、1990年代のBrooksらが主張し、ロボット研究で議論されてきた、知能は身体を必要とする、という議論の発展であると捉えられる[2]。このように、古くから知能そのもの、あるいはその獲得プロセスをシミュレーションするためには、物理的存在のある身体、すなわちロボットの身体が必要であると考えられてきた。

ところで、本研究で対象とする聴覚についても身体性の役割は重要である。例えば、人間は個々に異なる形状の頭部を持つ。従って、同一の環境において、同一の音源から、全く同一の音が発生したとしても、鼓膜に伝わる音響信号

は異なる。また、頭部運動は音響信号の測定点、即ち耳介の位置を変化させるため、音環境理解において、重要な要素となる。この場合、聞き手としての人間は、自らの頭部形状の影響、頭部運動の影響を事前に学習し、音環境を理解していると考えられる[3]。

一方、人間における聴覚の重要な機能として、聴覚情景分析(Auditory Scene Analysis)がある[4]。いつどこでどのような音が発生し、その音はいかなる文脈を持つのか。人間が動物の一種として、生きていく上で、欠かせない重要な能力の一つであると言える。この聴覚情景分析にも頭部運動は重要な役割を果たしている。我々はこれまで、音脈分凝課題に対する頭部運動の影響を対象として研究を行ってきた。特に、音脈分凝課題の中で、周波数の異なる2種類のトーンA、BをABA-ABA-の様に提示し、それをAとBの2つの音脈に分凝して知覚するか、全体として1つの音脈に群化して知覚するかを問う課題を設定した場合、途中で頭部運動を加えることにより、群化から分凝に知覚が変化することを指摘した。さらに、頭部運動によって生じる、音響的变化や音像の変化が、頭部運動によって生じる頸椎の運動指令等の、音響・音像以外の要素よりも大きな影響を与えていることを示した[5]。しかし、この時、評価としては、頭部運動後6秒間の積分的な影響を調べており、音響的变化に対する知覚と頭部運動そのものに対する知覚という別種の感覚を、一律の長い時間窓で比較することど

†1 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corporation.
†2 UMR 8158, CNRS and Université Paris Descartes
†3 Ecole normale supérieure
†4 東京工業大学大学院 総合理工学研究科
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

まった。そこで、本研究では、頭部運動の効果をいくつかの要素に分けた後、さらにそれらの効果量の時間的遷移に注目して分析した。

2. 実験手法

(1) 音響テレプレゼンスロボット「テレヘッド」

テレヘッドは、使用者と同形状の頭部を持ち、聴取者の頭部運動に追従動作することで、ロボットの設置場所における音環境を頭部形状と頭部運動の効果を伴って伝達出来るロボットである(詳しくは[6, 7])。このテレヘッドを用いることで、頭部運動の効果を制御することができる。聴取者の頭部運動に追従動作する以外にも、追従動作しないことで、頭部運動によって生じるはずの音響的な変化を無効にすることが可能である。また、テレヘッドだけを動かす事で、音響的な効果を、聴取者の頭部運動無しで実現することもできる[5]。

(2) 頭部運動効果の要素

実験条件を設定するにあたり、頭部運動の効果が3つの要素から構成されると仮定した。以下にその3要素を列挙する。

1. 音響信号の変化: Acoustic cue (ΔA)
2. 音像の移動: Sound localization cue (ΔS)
3. 頭部運動自体: Head movement cue (ΔH)

ΔA は音響信号の変化であるが、これは、頭部と音源の、相対的な位置関係によって変化する。相対座標系での音源の位置とも解釈できる。 ΔS は音像の変化であるが、これは、音響信号の変化から、頭部運動の効果を減じたものとなる。絶対座標系での音源の知覚位置とも解釈できる。 ΔH は狭義の頭部運動の効果であり、頭部を動かしたことによって生じる、筋の動き、筋を動かす指令が脳で発生することなどが含まれる。以降、頭部運動の効果と記述した場合、この狭義の効果 ΔH のことを指す。

(3) 実験条件

ここでは、実験条件について簡単に触れる。詳細な実験条件・環境の記述は、本講演会の別発表の原稿等を参照されたい[5,8]。聴取者は音脈刺激を聴き、分凝か群化のどちらかが知覚されるかを聴きながら常にボタン押しで回答した。その際、頭部運動を行う場合は、開始10秒後に、頭部運動をするように指示を出した。

この時、頭部運動を構成すると仮定した3要素を直接単独で変化させることは難しい。例えば、頭部を動かしている(ΔH がある)状態では、相対座標系での音源の位置が変化する。この時、音源が相対座標系の変化に追従して移動すれば、相対座標系での変化(即ち ΔA)は発生しない。しかし、これを絶対座標系での音源位置を変化させずに(即

ち ΔS なしで) 実現することは不可能である。そこで、3要素のうちの任意の2つの要素だけが存在する様な実験条件を用意し、結果を比較した。テレヘッドを用い、聴取者の頭部運動とテレヘッドの頭部運動を独立に制御することで、3要素の中から2要素に絞って影響をコントロールすることは容易に可能である。

- 実験条件 1) Self 条件: 聴取者が頭部運動し、テレヘッドがその運動に追従動作する条件。 $\Delta A + \Delta H$
- 実験条件 2) Source 条件: 聴取者は頭部運動しないが、テレヘッドは頭部運動する条件。 $\Delta A + \Delta S$
- 実験条件 3) Self & Source 条件: 聴取者は頭部運動するが、テレヘッドは動作しない条件。 $\Delta S + \Delta H$
- コントロール条件) No change 条件: 聴取者もテレヘッドも動作しない条件。

以上の4条件に対して、群化および分凝の時間変化を測定し、分析した。

3. 結果・考察

各実験条件での群化・分凝の時間変化などの生データは、本研究会の別発表等を参照されたい[5,8]。ここでは、頭部運動によって生じる3つの効果の関係を式(1)の線型モデルに従うと仮定した。

$$R = K_A \Delta A + K_S \Delta S + K_H \Delta H. \quad (1)$$

R は各瞬間における群化の比率。 $K_{A,S,H}$ は各要素 $\Delta A, S, H$ の寄与率である。 $\Delta A, S, H$ は各実験条件で効果として含まれるか否かで1または0の値を取る。例えばSelf条件では、 $\Delta A=1, \Delta S=0, \Delta H=1$ となる。 R は各実験条件、各時刻に対して存在する。音脈分凝課題では、刺激が無くても、群化や分凝の知覚が自然に交代するため、いずれもコントロール条件での各時刻における群化の確率を減じて計算した。

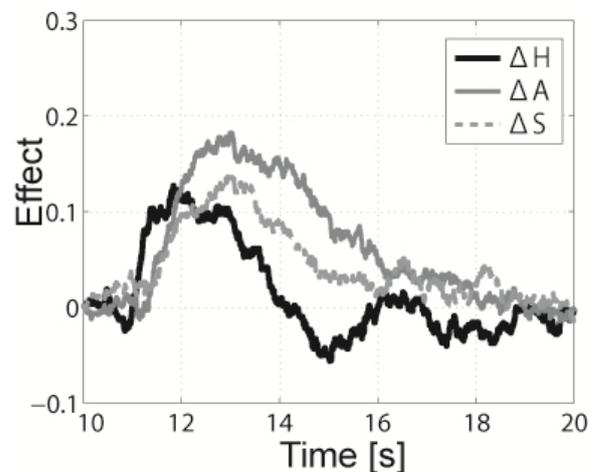


図1 $\Delta H, \Delta A, \Delta S$ の寄与率の時間変化
 Figure 1 Time course of the effects ($\Delta H, \Delta A,$ and ΔS).

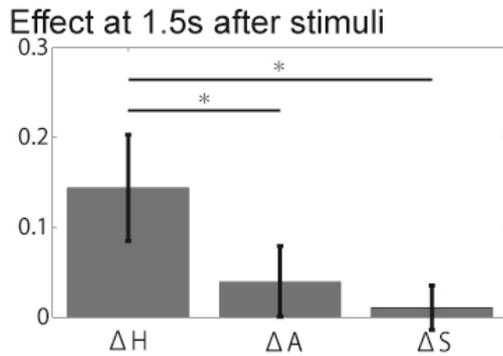


図2 時刻 11.5 秒における寄与率
Figure 2 Effects at 11.5 s

ΔH , ΔA , ΔS の寄与率の時間変化を図 1 に示す. 図 1 で, 太線は ΔH , 即ち頭部運動の効果の寄与率を表している. また, 灰色の実線は ΔA , 破線は ΔS の寄与率の時系列変化である. 時刻 12 秒, 即ちイベント発生後 2 秒までの間, ΔH の寄与率は他よりも大きいことが読み取れる. また, イベント発生後 1 秒間はほとんど効果が現れないが, これは, 刺激を認識し, 頭部運動を行う等し, 知覚変化を回答するのに時間がかかるためである. 先行研究[5]では, イベント発生後 6 秒間のトータルの効果として寄与率を計算していたため, その他の要素に比べて ΔH が小さくなったが, 分析対象とする時間によっては, ΔH の効果が大きくなる場合がある. これは, 時系列変化のパターンとしては, ΔH が時刻によってはマイナスの効果にもなる多峰性の形状であったのに対し, ΔA および ΔS は 13 秒周辺をピークとする単峰性の形状となったことによる.

イベント後 1.5 秒後における寄与率を図 2 に示す. ΔH の寄与率が ΔA , ΔS の寄与率よりも有意に大きくなった (T 検定で, それぞれ $P=0.025, 0.036$). また, ΔH は 0 よりも有意に大きくなった (同じく $p=0.03$). つまり, イベント後 1.5 秒後の時点では, 頭部運動の効果が音響情報や定位知覚の変化よりも大きな影響を与えていた.

さらに, 知覚変化の回数を図 3 に示す. ΔH の効果を含む条件である Self や Self & Source の条件で知覚変化の回数が増えていることが分かる. この結果は図 1 において ΔH の時系列変化が正負をまたぐ形での多峰性のグラフになっていることと合致する.

音脈分凝課題に与える影響として, 例えば, ABA 音の間隔を長くすることや, ノイズを加えるなど, 何らかの操作をした場合, 操作の瞬間に変化が起こることは経験的に知られていたが, 長期的に知覚変化を増やす様な影響が残る条件については, ほとんど指摘されていなかった. このため, 我々は先行研究[5]において, 各効果について, 単峰性

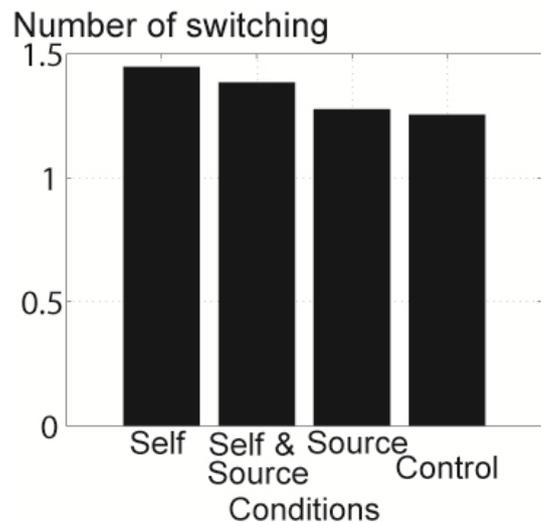


図3 各実験条件における知覚変化の回数
Figure 3 Number of switching in each experimental condition.

の時系列変化を無意識に予想していた. この仮定の下では, 効果の収束が観察されるまでの十分長い時間での積分的効果を検証することが重要である. なぜなら, 頭部運動と音響情報, 音像定位という効果時間の異なる可能性がある知覚入力について比較する試みであるため, 全ての効果が完全に含まれる時間を対象とするのが公平な比較と考えられるからである.

しかし, 実際には, 頭部運動の効果だけが, 特にマイナスの効果が発生する時間帯が存在するほど, 振幅のはっきりした多峰性になっていた. これは, 頭部運動の効果が, 他の効果とは別のプロセスに作用をもたらしていることを示唆している. 頭部運動には, 筋肉の収縮や, そのための脳からの指令が含まれる. 本分析結果は頭部運動という主体的な動作を行うことで, 聴覚情景分析における解釈の部分で, より探索的な解釈が行われるようになった可能性を示唆すると考える.

聴覚情景分析とは異なるが, 音像定位も聴覚の重要な基礎的機能の一つであり, どこで音がしているかという知覚は, 聴覚情景分析に使われる重要な情報の一つである. この音像定位実験においては, 頭部運動が定位感を向上すること[9]から, 主体的な頭部運動が重要であるということ[10], 主体的な頭部運動における頸椎の体性感覚情報が音像定位精度を向上させること[11]等, 頭部運動の効果が幅広く指摘されている. 筆者もテレヘッドを使った定位実験で, 頭部運動を行うことで頭部形状の簡略化可能性が広がる等のことを確認している[12]. このように, 頭部運動は, 少なくとも定位感において, 強い影響がある. しかも, 他の音響的な情報とは異なり, 主体性の有無等が影響していることが知られている. 主体的かどうかということは, 高

次の情報処理に影響があるということを示唆していると考えられる。本研究で示唆された音脈分凝課題における頭部運動の効果が、他と異なる性質を示したことは、上記の一連の先行研究の結果と合致し、頭部運動が高次の情報処理に影響を与えたことを示唆する結果であると考えられる。

4. おわりに

聴覚情景分析における頭部運動の効果を考察するために、音脈分凝課題を用いて、群化および分凝の知覚変化と頭部運動の関係を調査した。特に頭部運動の効果が時系列的にどのように変化するかを分析した。その結果、以下の結論を得た。

1. 頭部運動の効果は音響情報や音像の変化の効果よりも早く現れる。
2. 頭部運動を行った場合、その後の群化および分凝の回数が増加する。
3. これは聴覚情景分析における頭部運動に対する情報処理が、他の音響情報や音像知覚の変化に対する情報処理と異なるプロセスとなっている可能性を示唆する。

参考文献

- 1) R. Pfeifer and C. Scheler, “知の創成”, 共立出版, 2001.
- 2) R. A. Brooks, “Elephants don't play chess”, *Robotics and Autonomous Systems*, vol.6 pp3-15, 1990.
- 3) J. Blauert, “Spatial hearing”, MIT Press, 1982.
- 4) A. S. Bregman, “Auditory Scene Analysis”, *The Perceptual Organization of Sound*, MIT Press, 1990.
- 5) H. M. Kondo, D. Pressnitzer, I. Toshima, M. Kashino, “Effect of self-motion on the resetting of auditory scene analysis”, *Proceedings of National Academy of Sciences (PNAS)* vol. 109, pp. 6775-6780, 2012.
- 6) I. Toshima, H. Uematsu, and T. Hirahara, “A steerable dummy head that tracks three-dimensional head movement: TeleHead”, *Acoustical Science and Technology*, vol. 24, no. 5, pp. 327-329, 2003.
- 7) I. Toshima, S. Aoki, and T. Hirahara, “Sound localization using an auditory telepresence robot: TeleHead II”, *Presence*, MIT Press, vol. 17, no. 4, pp.392-404, 2008.
- 8) 近藤洋史, Daniel Pressnitzer, 戸嶋巖樹, 柏野牧夫, “音脈のリセットに対する音源移動と頭部運動の影響”, *音学シンポジウム 2013*, 2013.
- 9) H. Wallach, “The role of head movements and vestibular and visual cues in sound localization,” *J.Exp.Psychol.*, vol. 27, pp. 339-368, 1940.
- 10) F. Wightman and D. J. Kistler, “Resolution of front-back ambiguity in spatial hearing by listener and source movement,” *J. Acoust. Soc. Am.*, vol. 105, pp. 2841-2853, 1999
- 11) 吉崎大輔, 平原達也 頭頸部の体性感覚情報が水平面音像定位に及ぼす影響 日本音響学会 2011 年秋期研究発表会 講演論文集, pp.479-480, 2011
- 12) I. Toshima, and S. Aoki, “Possibility of head-shape simplification for an acoustical telepresence robot: TeleHead”, *Journal of Robotics and Mechatronics*, Japan Society of Mechanical Engineers. vol. 21, no. 2, pp. 223-228, 2009.