

OE4-1

脳から臨場感通信へ：タイミングの知覚メカニズム From the Brain to Tele-presence: Perceiving the Timing of Events

柏野 牧夫†
Makio Kashino

1. まえがき

地球の裏側の出来事であってもあたかもその場にいるように体験できる——臨場感通信は、生身の身体のもつ空間的制約から解放され、伸縮自在の新しい身体を手に入れる技術である。通信の広帯域化をはじめとする技術革新によって、その実現可能性は俄然高まってきたように思える。しかし忘れてはならないのは、臨場感通信の成否は脳が握っているということである。新しい身体、つまり人工的な感覚系と運動系によって通常とは異なる感覚情報や運動情報が脳に与えられた場合、リアリティのある知覚は成立しうるか。あるいは、成立するための要件は何か。どうすれば、最も効率よく、しかもリアルな知覚を生み出せるか。これらの問いは、臨場感通信の実現可能性に関わる根本問題であると同時に、「生身の」人間の認知メカニズムを解明する上でも貴重な手がかりをもたらすものである。本稿では、時間的側面、とりわけ事象のタイミングの知覚に絞って、認知神経科学の観点から問題点を指摘し、最近の知見を紹介する。

2. タイミング知覚における認知神経科学的問題

ひとつの思考実験として、究極の臨場感通信システムが完成したとしよう。体験者は東京にいて、その分身である人型ロボットが、地球の裏側リオデジャネイロにある。分身には視覚、聴覚、触覚等のセンサーが備わっており、それらから得られた感覚情報は超高速回線を経由して体験者の各感覚器に高精度で提示される。一方、体験者の手足の動きに関する運動情報も超高速回線によって分身に伝えられるので、体験者は分身の手足を忠実に動かすことができる。さて、この体験者は、リオのカーニバルの熱気を、現実にその場にいるのと同等のリアリティをもって体験できるだろうか。

第一の問題は、情報の伝送遅延である。地球の裏側まで地表を最短距離で結べば約 2 万 km であり、光速でも約 1/15 秒かかる。仮に超高速システムで遅延が片道 1/10 秒 (100 ms) に収まるとしても、分身からの感覚情報を得て、体験者が何か手足を運動させたとき、それが分身に反映されるまでには 200 ms の時間がかかる。200 ms といえば無視できない時間で、サンバのリズムにあわせて踊っても 1 拍ばかり遅れてしまうはずである。

第二の問題は、感覚モダリティ間の非同期である。分身からの視覚情報、聴覚情報など、センサーのチャンネルが異なる情報は、別々に処理され、伝送される。その間に、情報の同期性が失われる可能性がある。遠くの花火を見るように、映像と音はばらばらに観察者に届くことになる。

このような原理的問題にもかかわらず、リアリティのある知覚は成立するであろうか。その答えは、脳に求めな

ればならない。実は、ここで述べたのと同様の問題は現実の環境や生身の身体でも生じているのである。

まず遅延の問題を考える。例えばサンバの演奏を、演奏者から 34 m 離れた場所で聞いているとすると、音が耳に届くまで 100 ms かかる。耳に届いてから、内耳、聴神経、脳幹のいくつかの神経核を経て、大脳の第一次聴覚野に信号が最初に届くまで約 20 ms を要する。高次聴覚野、連合野と、さらに処理が続く。したがって、サンバの音が聞こえたとき知覚された時点では、実際に音が発生してから少なくとも百数十 ms 経過していることになる。

次に非同期の問題である。サンバからの光と音が目と耳に届いた時点で、この例なら既に 100 ms のずれが存在する。このずれは観察者と対象との距離に依存する。さらに、感覚系内部でのずれも無視できない。視覚の場合、目に光が入ってから第一次視覚野に信号が最初に届くまで約 50 ms を要し、そこからさらに高次視覚野に届くまで、部位にもよるが、50 ms 前後の時間がかかる。つまり、聴覚よりもかなり遅い。さらに、感覚系では、視覚であれば動きや形、色など、聴覚であれば音源定位やピッチなどの特徴が、並列の専門化した神経機構 (モジュール) によって処理されるが、それらのモジュール間でも伝送や処理に要する時間は異なる。

これらの問題にもかかわらず、我々はサンバのリズムに合わせて踊ることもできるし (身体を動かすには運動のプログラミングや神経伝達等にまた時間がかかることに注意)、パーカッションの音と映像のタイミングがずれて知覚されることも、位置と音色がばらばらに知覚されることもない。つまり、脳は、日頃から、地球規模の臨場感通信とさして変わらない遅延や非同期に直面し、それらをうまく処理しているのである。以下の 3 節では、聴覚を中心に、一見奇妙な知覚現象を取り上げながら、脳の情報処理戦略について考察する。

3. まとまりとタイミング

聴覚におけるタイミング知覚の精度は、前後の文脈によって大幅に変化する。例えば、周波数の異なる二つの短い正弦波 A, B の時間順序を判断するという課題は、条件にもよるが、両者の立ち上がりの差が数 ms から十数 ms あれば可能である。しかし、A, B の前後に別の周波数の正弦波 X をつけて XABX の形にすると、順序識別閾はほぼ一桁上昇する。さらに、XX...XABX...XX という形にすると、また元通りの識別が可能となる [1]。この場合、AB がひとまとまりの事象として知覚的に群化されるか否かが重要である。

群化とタイミング知覚の関係を示す例は言語音や音楽などにも豊富に見られる。

例えば、連続発声中の音素の知覚は、高いタイミング判断の精度を要する。日常会話では 1 音素あたり数十 ms 程度であるし、各音素はさらに微細なスペクトル変化によって特徴付けられている。ところが、連続発声中の 1 音素を削除し、その部分に短い外来音 (雑音など) を挿入すると、

そのタイミングを判断するのは難しく、数音素分(数百ms)ずれることもある[2].

音楽でも、音域が離れた音符が速いテンポで提示されると高音域と低音域に音の流れ(音脈)が分凝し、音脈間ではタイミングの判断が困難になる。この音脈分凝現象を利用して、単旋律しか演奏できない木管楽器などで擬似的に複数の旋律を同時に聞かせるという手法はしばしば用いられてきた[1].

このように、タイミング知覚の精度が群化によって大幅に変わるのには、どのような神経メカニズムによるのであろうか。ここで興味深いのは、群化された場合に聴覚が示すmsオーダーの時間分解能は、大脳皮質聴覚野のほとんどのニューロンの時間分解能をはるかに超えているという事実である。筆者らは、群化された周波数成分の時間順序は、時間分解能の高い聴覚系の初期段階で単一の非時間的な神経符号に変換されるという仮説を提唱し、それを支持する実験結果を得ている[3].

4. 過去と未来の逆転

物理的には、時間は過去から未来へと、一方向に流れていく。しかし知覚的な時間は、それに対応しているとは限らない。

聴覚の補完現象の例を挙げる。周波数が連続的に上昇していく音の途中の200ms程度を無音にすると、そこは明らかに途切れて聞こえるが、その部分に適切な特性をもつ雑音を挿入すると、元通り連続して聞こえる。つまり、雑音部分に本来あるはずの音が知覚的に補完される。ここで、雑音の後の部分を、下降するように変えると、滑らかに上昇して下降する連続的な音が知覚される。さらに、雑音の後の部分を削除すると、周波数変化音は雑音の開始点で終わるように知覚される。

この現象は、時間の観点からすると奇妙である。なぜならば、聴取者は、雑音の後の部分を聞かない限り、雑音の部分をどのように補完すべきか、知る由もないからである。それにもかかわらず、まず周波数変化音が聞こえ、それが雑音の背後で鳴り続け、雑音の後も滑らかに変化しているように知覚される。これは、雑音の後に聞いた後で、後付け的に雑音部分の知覚が創り出されていると考える他はない。

この他にも、物理的には後で起きた事象がそれ以前の事象の認知に影響するという例は、聴覚に限らず数多い。我々が知覚している事象の順番は、物理的な順番そのものではなく、ある程度——概ね2~300ms——の情報に基づいて、脳内で再構成されたものであると考えられる。

5. モダリティ間の相互作用

事象の解釈が時間知覚において決定的な役割を果たすのは、視覚なら視覚、聴覚なら聴覚という単一の感覚モダリティ内だけにとどまらない。最近、下條信輔(カリフォルニア工科大学・NTTコミュニケーション科学基礎研究所)らの研究グループは、聴覚に与えられた情報が視覚の知覚内容を変えるという例を相次いで報告した。

例えば、黒い背景に、白い円盤がフラッシュ提示され、それに短い音が伴うとする。観察者は、何回フラッシュが見えたかを報告する。観察者は、単一のフラッシュに二つ以上のピープ音が伴う時はいつも、複数のフラッシュが見えると誤って報告した[4]。ここで興味深いのは、フラッ

シュが光っている時間は視覚系の時間分解能に比べて非常に短い(この実験では13ms)ので、その間に物理的に2回フラッシュが光っても、融合して一つにしか見えないということである。ごく短時間内での視聴覚の事象は対応しているとの前提のもとに、時間分解能の高い聴覚系からの情報を用いて、視覚系の時間分解能を超える視覚知覚が生じたと考えられる。

この現象は、聴覚が視覚の「見え方」自体を変容させるという例であるが、下條らはその他にも、聴覚が視覚の時間分解能を変化させたり、視覚の多義性を解消したりする現象を報告している。これらの現象がこれまでの常識に反する点は、聴覚情報が視覚の知覚内容を変えているという点である。これまでも視聴覚の相互作用を示す錯覚現象は数々知られていたが、ほとんどは視覚情報が聴覚の知覚内容を変えるというものであった。腹話術の人の口の口から声が聞こえるという腹話術師効果や、口の画像が「が」、音声は「ば」のとき「だ」や「が」が聞こえるというMcGurk効果がその代表例である。そのため、人間は視覚優位の動物であると言われることも多かった。しかし下條らの発見は、時間知覚に関しては視覚よりも聴覚が優位であることを示している。空間なら視覚、時間なら聴覚というように、より確かな情報をもたらすモダリティがモダリティ間の情報統合において優先されるというのが正しい見方であると考えられる。

6. おわりに

以上のように、事象のタイミングの知覚は、物理的な時間順序そのもので決まるわけでも、脳の特定部位に情報が到達したり処理が完了したりしたタイミングで決まるわけでもない。

より適切な考え方は、事象のタイミングは、脳が、さまざまなモダリティからの時間的に相前後する情報を総合的に勘案して、外界の事象を解釈した結果形成されるというものである。つまり、タイミング知覚も、他の知覚と同様、不完全な情報に基づいた一種の推定問題を解く過程と見ることができる。各モダリティ、各モジュールで断片的にサンプリングされた情報に、処理系内部での変動が加わったものから、最も尤もらしい事象のタイミングを求めているわけである。その過程の詳細は今後解明すべき点が多い。

最後に臨場感通信の話に戻るならば、ある程度の伝送遅延や非同期があること自体は、リアリティを創り出す上で致命傷とはならないと考えられる。脳は元来そのような問題に対処するメカニズムを備えており、しかもきわめて可塑的なものだからである。ただし、おのずと限界はあるはずなので、それを明らかにすることが必要である。さらに、脳が事象を解釈しリアリティを創り出す要件およびそのメカニズムを解明することは、認知神経科学、臨場感通信の両方の観点から重要であり、技術と科学の互恵的關係が期待される。

文献

- [1] Bregman, A. S. (1990). *Auditory Scene Analysis*. MIT Press.
- [2] Warren, R. M. (1999). *Auditory Perception: A new analysis and synthesis*, Cambridge University Press.
- [3] Kashino, M. & Okada, M. (2002). *Forum Acusticum Sevilla 2002* (in press).
- [4] Shams, L, Kamitani, Y., & Shimojo, S. (2000). *Nature* 408, 788.