

## 特集 ■ VRの源流としての錯視・錯覚

## 錯覚からみた聴覚の「賢さ」



柏野牧夫

NTT コミュニケーション科学基礎研究所

KASHINO MAKIO

## 1. はじめに

聴覚というものは一般に過小評価されていると感じることがしばしばある。典型が、「人間は情報の大半を視覚から得ている」という類の文句である。「大半」のところに「95%」などと具体的な数字が入っていることもある。しかし、説得力のある根拠が示されているのは見たことがない。そもそも、感覚モダリティはそれぞれ性質が異なっているので、単純な量的比較など不可能なはずだ。例えば空間知覚に限ってみても、確かに視覚は視野内、とりわけ注視点付近の分解能は非常に高いが、視野は顔の前方に限られている。カバーする範囲の広さという尺度を持ち出せば、常時全方向を守備範囲とする聴覚に軍配が上がる。どちらが優れているかという問題ではなくて、相補的なのである。視覚が特定の対象を精査するのに適しているのに対して、聴覚は常時周囲の状況をモニターし、異変をいち早く察知するのに適している。そのため、自分が今どういう場にいるかという感覚は、聴覚によるところが大きい。さらに、解剖学的、生理学的に見ても聴覚と情動との結びつきは強く、周囲の異変を察知すればいち早く身体が反応する。したがって、バーチャルリアリティのように、その場にいるかのようなリアリティをつくり出す上では、聴覚の役割は想像以上に大きい。激しいカーチェイスも戦闘シーンも、音を消してしまえば全く迫力はない。

もう一つよく耳にするのが、「聴覚にも錯覚はあるのですか?」という質問である。「もちろんあります」と答えると驚かれることが多いのは、聴覚は視覚と違って単純で受動的な感覚だと思われている証拠だろう。知覚の内容が感覚器に与えられる刺激の物理的特性と単純に対応しないことを錯覚と呼ぶなら、聴覚も視覚同様錯覚に満ちている。その種類は数多く、とてもここで紹介しきれものではない。興味をお持ちの読者は、CDや

Webのデモを参照されたい[1-5]。本稿では、「存在しない音が聞こえる」「同じ音が違って聞こえる」「知覚的な時間や空間が歪む」という3種類の錯覚現象を取り上げて、その背後にある聴覚情報処理の原理について考えてみたい。一見錯覚だが、そこには適応的な意味がある。聴覚も思いの外「賢い」ということを伝えることができれば幸いである。

## 2. 存在しない音が聞こえる

## 2.1 隠されると聞こえる

まず取り上げるのは、物理的には存在しない音が聞こえるという錯覚である。文章を声に出して読み上げた音声を録音し、その文章の頭から100～200ms程度の間隔ごとに、音声を削除する。言い換えれば、一定間隔ごとに無音区間を挿入することになる。こうすると、何を言っているのか非常に聞き取りにくくなる。次に、音声を削除した部分に、広帯域の雑音を挿入する。雑音の強さは音声よりも大きくする必要がある。こうすると、雑音の背後で、削除されたはずの音声は滑らかにつながって聞こえる(図1)。雑音があろうがなかろうが、同じ分量の音声が入力された場合に限って、削除された部分が修復されるのである。これを音素修復(phonemic restoration)と呼ぶ[1]。

音素修復は、知覚される音が物理的な音のコピーではないということを如実に物語っている。しかし、これは知覚の誤りではない。それどころか、このような補完機能があるからこそ、日常生活をうまく送ることができる。というのも、日常的な環境は往々にして妨害音が多いからである。例えば人の話を聴いているとき、別の人が入浴をバタンと閉めて入ってくれば、その部分の音声はドアの音に隠されて検出できなくなってしまう。これをマスキングと呼ぶ。音素修復は、マスキングの影響を補償

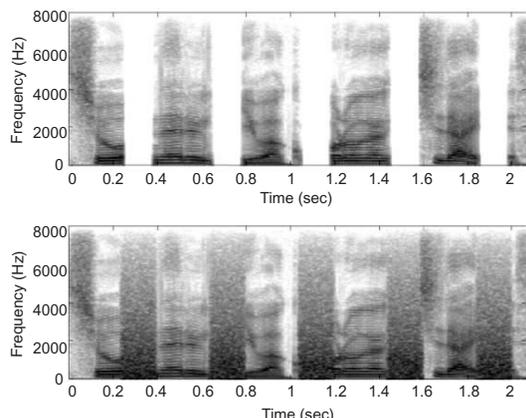


図 1 音声の知覚的補完現象 (音素修復)

上段:「省エネルギーは心がけ次第です」という文音声を一時間ごとに削除(無音置換)したもの。下段:上段の無音置換の部分に雑音を挿入したもの。縦軸は周波数、横軸は時間、濃いところはエネルギーが大きい部分を表す。上下とも、元の音声の残存量は同一であるが、上段は内容が聞き取れず、下段ははっきりと聞き取れる。

するのに役立つ。

このような補完現象が生じるのは言語音声に限らない。対象が音楽でも環境音でも、あるいは純音でも雑音でも、特定の条件さえ満たされていればどのような音でも生じる。その条件とは、ひとことで言えば、対象音を削除した部分に挿入する雑音が、本来の音がそこにあったとしても十分マスキングすることができる特性があるということである。すなわち、挿入する音のスペクトルが対象音に近く、強さが対象音より強く、残存した対象音と挿入する音との間に検知できる時間的な隙間がなければよい(図2)。このような条件は、日常的な環境を考えれば、きわめて理に適っている。補完が必要なのは、本来存在している音が別の音でマスキングされた場合だけであって、もともと切れている音を補完する必要はない。聴覚には、マスキングされた可能性が高い場合に限って補完するという選択性が備わっているのである。

### 2.2 音声信号の冗長性

さて、欠落部分を補完するか否かはそこに挿入される音のマスキング能力によって決まるとしても、そこをどのように補完するかはそれだけでは決められない。では、音素修復の場合、なぜ元の音素が知覚されるのだろうか。そのヒントとなる錯覚を紹介しよう。

前提として、ある内容を発声して録音したテープを逆転再生しても文字を逆から読んだようには聞こえない。それどころか、普通は日本語にすら聞こえない。ところが、次のような逆転再生では様子が違う。まず、録音した音声を 50 ms 程度の区間に区切る。次に、それぞれの区間の音声を時間的に逆転させる。しかし区間同士は入れ替えない。つまり、それぞれの区間を仮に A, B, C, D,

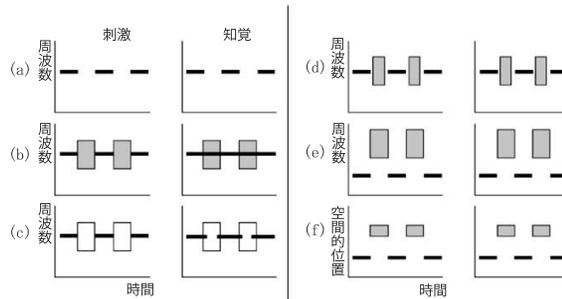


図 2 知覚的補完現象の生起条件

(a) 断続的な純音は断続的に知覚される。(b) 断続的な純音の間に適切な周波数・音圧レベルの雑音を挿入すると、純音が連続しているように知覚される。(c) 雑音の周波数は適切でも音圧レベルが低すぎると、完全に連続しては知覚されない。(d) 純音と雑音の間に検知できるギャップがあると、連続的に知覚されない。(e) 純音と雑音の周波数帯域が異なると音圧レベルが高くても連続的に知覚されない。(f) 周波数と音圧レベルが(b)の条件でも純音と雑音の空間的位置が異なると連続的に知覚されない。

E, ... と名付けるなら、A を逆転したもの、B を逆転したもの、C を逆転したもの、... というように並べ、つなぎ合わせる(図3)。これを再生してみると、意外にも、もとの発話内容がそのまま聞き取れる。若干エコーがかかったような感じにはなるが、少なくとも何かが逆転しているような印象はない[6]。

このような現象が生じるのは、音声信号の生成過程に原因がある。音声信号は顎、唇、舌などの調音器官の動きによって生成される。例えば「いた(/ita/)」と発声するとしよう。母音は、特定の顎の開き方、唇の形、舌の位置などによって共鳴管を形成し、声帯の振動によって生じるブザーのような音のスペクトルをその共鳴管で整形することによって生成される。/ita/ という発声では、/i/ から /a/ へ、顎、唇、舌などが連続的に動いていく。その途中に、子音 /t/ を生成するための動作が挿入される。すなわち、舌の先が上の歯の後ろの部分に一瞬接触して呼気をせき止め、すぐに離れて呼気が勢いよく流れ出る。呼気がいったんせき止められる直前から声帯振動は止まっており、呼気が解放されるやいなや声帯振動が再び始まる。/ita/ というきわめて単純な発声でさえ、複数の調音器官が精妙にタイミングを合わせて動いている。

ここで重要なのは、調音器官は自由自在には動けな

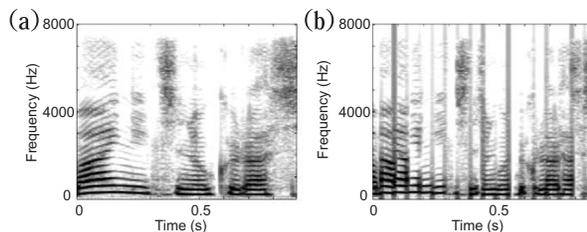


図 3 部分逆転音声

(a):「まいにちのくらし」という原音声のサウンドスペクトログラム。(b):(a)を 50ms ごとに区切り、それぞれを時間的に逆転してつなぎ合わせたもの。

いということである。舌であれ顎であれ、質量があるので慣性が働き、ある状態から別の状態へは滑らかにしか移れない。また、動ける範囲にも自ずと限界があり、他の調音器官とまったく独立に動けるわけでもない。したがって、例えば /ita/ という発声において、/a/ を調音するための準備は既に /i/ を発声している時点から行われているし、/a/ を発声している時点には /i/ の調音の影響が尾を引いている。当然これに /i/ を調音する動作の影響が重なる。このような調音運動の特性を反映して、各音素についての情報が互いに重なりながら数百 ms 程度の範囲に分散することになる。だから、素朴な内観と裏腹に、音声信号の上でここからここまでがこの音素、などということとはできない。無理に音声信号の一部を切り出して聞いてみると、本来の音素とはまったく違ったものに聞こえることが多い。機械による音声認識や音声合成が難しいのはこのためだ。活字を並べて文章ができるというようなわけにはいかないのである。

しかし裏を返せば、これは音声信号が時間的に冗長であるということに他ならない。部分逆転再生の場合には、実は破壊されているのは逆転させる区切りの長さ(上記の例では 50 ms)よりも細かい構造だけで、調音動作に対応するような緩やかな変化は比較的保存されている。音素修復に話を戻せば、100 ms 程度の信号が欠落しても、やはりそれより緩やかな変化の情報は残されている。だからこそ、当て推量ではなく、適切な補完が可能なのである。もちろん信号の欠落が上記の例よりひどくなれば、音声信号に含まれる情報だけでは不十分である。実際には、その場合でも直ちに音素修復ができなくなるわけではない。文の意味や文法など、より高次の情報も効果的に使われるからである。

### 2.3 過去と未来の逆転

前節で、一つの音素の情報が数百 ms の範囲に分散していると述べたが、物理的に考えて、知覚に利用できるのは欠落部分よりも前にある情報だけであるはずだ。ところが奇妙なことに、欠落部分よりも後にある情報も補完に影響するのである。話を明確にするために、より単純な例を挙げてみよう。周波数が連続的に上昇していく音の途中の 200 ms 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(図 4a)、その部分に適切な特性を持つ雑音を挿入すると、元通り連続して聞こえる(図 4b)。雑音部分に本来あるはずの音が知覚的に補完されたわけである。ここで、雑音の後の部分を、下降するように変えてみると、滑らかに上昇して下降する連続的な音が知覚される(図 4c)。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っているように聞こえる(図 4d)。この例では、聴取者は、雑音の後の

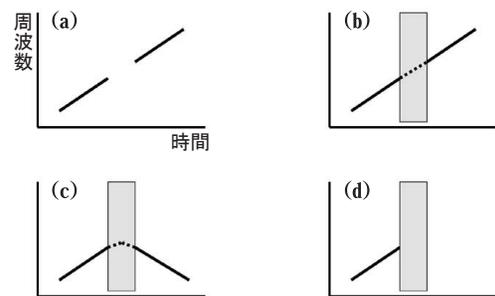


図 4 補完現象における過去と未来の逆転

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。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、雑音の後にも滑らかに変化しているように知覚される。音素修復の場合も同様である。音声の一部を雑音で置換し、雑音の前後の音声を系統的に変化させてみると、補完される内容は、雑音の前だけでなく後に現れる音声の特徴にも依存して変化する [7]。これらの現象は、雑音の後を聞いた後で、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。つまり、物理的な時間は、過去から未来へ、一方向に流れているのに対して、知覚的な時間は、それに対応しているとは限らない。

音素修復が生じているときには、ほぼ自動的に言葉が聞こえてくるように感じられる。しかしその背後には、かなり複雑で巧妙な情報処理が行われているのである。

## 3. 同じ音が違って聞こえる

### 3.1 反復に伴う知覚の変化

次に紹介するのは、物理的には同一の音であるのに、反復すると聞こえ方がどんどん変わっていくという現象である。まず、短い単語を録音する。例えば「バナナ」とやや早口で発声する。次に、それを切れ目なくループ(反復)再生する。すると、「バナナ」のはずが、「ナッパ」になったり、「ハナ」になったり、さらには 2 人の声に分かれたり、機械的な音が聞こえてきたり、人によって中身は様々だが、普通は 1 分間も聞けばかなりの変化が体験できる。次々と違う内容が聞こえる人もいれば、比較的少数の聞こえ方が交互に現れるという人もいる。時には全く何も変化しないという人もいるが、別に異常ではない(ただし、加齢につれて変化が起りにくくなるというデータもある)。この現象自体は古くから知られていて、単語変形効果(verbal transformation)と呼ばれている [1]。視覚でも、同一の刺激に対して見ている間に知覚が変わる、いわゆる多義的知覚と呼ばれる現象は、ネッカー・キューブ、両眼視野闘争などいろいろある。単語変形効果は、それらと似ているところもあるし、違うと

ころもある。「バナナ」をループすると、「ナバナ」、「ナナバ」など音節のまとまり方に多義性ができるが、聞こえてくる内容はそれよりもバリエーションが多く、もともと含まれていない音が聞こえることも珍しくない。

### 3.2 仮説の生成と検証

なぜこのような奇妙な現象が生じるのか、そのメカニズムはよくわかっていない。筆者の考えでは、ボトムアップ、トップダウンのダイナミックな相互作用に原因がある。

知覚は一般に、感覚器に入力される情報だけでボトムアップ的に決まるものではない。環境内の事物一般、あるいは特定の対象に関する知識(内部モデル)が脳内に蓄えられており、そこから生成される予測がトップダウン的に働くという側面も重要である。知覚は無自覚的、自動的ではあるが、犯罪捜査のような推理の過程と形として似た部分がある。犯罪捜査では、現場に残された断片的な手がかりだけでは犯人を特定できないことが多い。そこから、事件の中身や犯人に関する仮説を立て、過去の事件データや関連情報も参考にしつつ、犯人像を絞り込んでいく。仮説に矛盾する証拠(アリバイなど)がなければ、犯人である可能性が高い。見込み捜査が行き過ぎて証拠不十分だと冤罪になるが、仮説もなく捜査しても迅速な解決は望めない。知覚の場合、犯人=外界の事物、証拠=感覚情報、事件データや関連情報=内部モデルというわけである。刻々と入力される感覚情報は不完全で断片的であることが多い。そこで、内部モデルを用いてオンラインでどんどん外界の事物に関する仮説を生成し、感覚情報と照らし合わせて検証していく。仮説に合致する情報が感覚器から得られれば知覚が成立する。むしろ、仮説に合致しない情報が得られさえしなければそれでよいとさえ言えるかもしれない。例えば前章の音素修復の例で、欠落部分が隠される場合のみ補完が生じるというのは、「そこが実際に無音になっている」という感覚情報(証拠)が「ない」からだとみることができる。無音というのは重要な音声特徴の一つである。欠落部分が他の音で隠されていないと、「無音である」という強い証拠があるので、調音器官の動きの制約や文の意味から生成された仮説が棄却され、補完が生じない。このような、内部モデルに基づく予測(仮説)と感覚情報に基づく検証というループは、実環境で効率よく、しかも適切な知覚を行う上できわめて有効であろう。

単語変形効果に話を戻すと、感覚情報という観点では、少なくとも2種類の不完全性がある。一つは先に述べた、ループによる音節のまとまり方の多義性である。もう一つは、特徴抽出過程の順応による感覚情報の変形である。聴覚系の各段階には、音の様々な特徴、例えば周波数の変化や振幅の変化などに対して選択的に応答する

ニューロン群が存在し、特徴の分析器として機能している。同一の単語を反復提示すると、その単語に含まれる音響特徴に対して応答するニューロン群が順応して、反応量が低下する。その結果、物理的には同一の単語であっても、知覚系の上位段階に送られる感覚情報は時々刻々形を変えていくことになる。それに対して、トップダウンの過程は、もっとも当てはまりのよい音素列に関する仮説を生成されるわけであるが、感覚情報の不完全性のゆえに、仮説が一つに絞りきれない。有力な仮説が複数ある場合、その時点における感覚情報とより整合するものが採用されることによって、知覚が変化していくと考えられる。通常の状況では適切な知覚につながるこのボトムアップ、トップダウンのダイナミックな相互作用が、同じ単語が反復されるという特殊な状況では錯覚を生み出すのではないだろうか。

### 3.3 錯覚を生み出す脳活動

このような考えを検証するために、筆者らは、機能的磁気共鳴画像法(fMRI)を用いて、反復単語を聴取しているときの脳活動を計測した[8]。ここで興味があるのは、「聞こえ方が変化する」という事象に関連した脳活動である。単純に音を聞いていることによる活動や、「バナナ」や「ナッパ」など特定の聞こえ方に対応する活動ではない。そこで、聞こえ方が変化するたびに実験参加者にボタンを押してもらい、それに関連した脳活動の部位を推定した。さらに、対照実験として、同じ反復単語に加えてときどき提示される短い純音を検出してボタンを押すという課題を別のセッションで行い、その脳活動も計測した。純音の提示タイミングは、同一の被験者の、同一反復単語の聞こえ方が変化する頻度に基づいて設定したが、聞こえ方の変化とは同期していない。

実験の結果、聞こえ方の変化を検出する課題で有意な活動が見られた部位は、左右聴覚野、左右前頭前野、左前頭葉腹側部(ブローカ領域)、左島皮質、前帯状皮質、右頭頂間溝、視床前部などである(図5\*口絵カラーページ参照)。一方、反復単語中の純音を検出する課題では、左右の聴覚野から後方に頭頂葉にかけての部位、左島皮質、視床前部の一部などに活動が見られた。詳細な分析は省くが、両条件で顕著な違いが見られたのは前頭葉である。特に、ブローカ領域の活動は、知覚の変化を頻繁に報告した人の方がそうでない人比べて顕著であった。ブローカ領域は、伝統的に、言語音声の生成に関与すると考えられてきた部位である。しかしこの実験の課題はもっぱら知覚課題であって、生成課題ではない。そこで思い出されるのが、「音声の知覚は、音響信号から、それを生成した発話者の調音ジェスチャーを推定することに基づいてなされる」という音声知覚の運動理論である[9]。運動理論

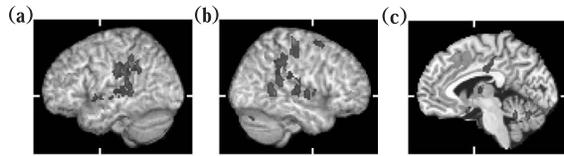


図5 「バナナ」の反復を聴取中の脳活動(N=12) \*口絵カラーページ参照  
(a) 左半球, (b) 右半球, (c) 正中面断面図. オレンジ色は聞こえ方が変化したときのみ, 青色は純音検出時のみ, 緑色はその両方に有意な活動が見られる部位.

は今日なお賛否両論あり,直接的な証拠は得られていない. 筆者らの実験結果は, 音声の知覚に生成モデルが関与するという運動理論の一つの証拠を提供するものである. 同時に, 単語変形効果に内部モデルによる予測生成が関与するという説にも合致する. また, 他の証拠とも合わせて, 前帯状皮質では, 聴覚野からの感覚情報とブローカ領域からの予測とが比較されているのではないかと考えている.

錯覚と脳活動計測をうまく組み合わせて, 今後さらに詳細な分析をすれば, 意識に上る知覚がどのような脳内メカニズムによって生じるかを明らかにするための手がかりが得られるに違いない.

#### 4. 時空が歪む

##### 4.1 音源定位の順応

直前の刺激によって後続の刺激の知覚が影響されるというのは, どの感覚モダリティでも, かなり一般的に見られることである. 例えば, 冷たい水に手をつけてからぬるま湯に手をつけると, 始めからぬるま湯に手をつけた場合よりも熱く感じられる. 流れ落ちる滝をしばらく見つめると, 止まっているはずの隣の岩がゆっくり上に動いているように見える. 野球でも, 豪速球であっても同じような球を何球も続けて投げられるとタイミングが合いやすいが, スローカーブの後だと速球がより速く感じられ, さほど速くない球でも振り遅れてしまう. これらも物理的な刺激と知覚の内容とがずれているわけであるから, 錯覚と呼ぶことができるだろう. スポーツ選手や料理人, 音楽家など, 人間の感覚を相手にするプロフェッショナルは, このような知覚の特性を暗黙のうちに熟知してうまく利用している. 聴覚でももちろんこのような文脈効果や順応効果は各種見られる. ここでは, 音源の空間的位置の知覚(音源定位)に関する現象を紹介しよう.

ある音を提示した直後に, いろいろな位置に別の音を提示すると, 後に提示した音は, 単独で提示した場合に比べて, 直前の音から遠ざかる方向にずれて知覚される[10, 11]. これを聴覚定位残効(auditory localization aftereffect)と呼ぶ. 水平面では, このずれは角度にして40度ほど両者が離れているとき最大となり, その量は

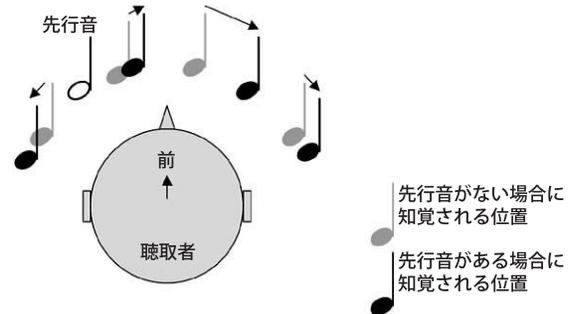


図6 聴覚定位残効の模式図  
後続音の空間的位置(定位)が, 先行音から遠ざかる方向にずれて知覚される. ずれの量は先行音から少し離れたところで最大となる.

15~20度にもおよぶことがある. また, 先行音と後続音の周波数が近いときにしか残効が生じない.

このような現象が生じる理由として, 筆者らは, 空間の位置と周波数とに選択性を持つニューロン群の感度調節を想定した. 空間の位置の手がかりとして, とりえず両耳間の時間差を取り上げるならば, モデルの概略は次のようになる. 聴覚系には, 末梢の周波数分析を経た後, 周波数帯域ごとに, 様々な両耳間時間差に同調したニューロンが多数存在している. 両耳間時間差は, それらのニューロン全体の興奮パターン(例えばその重心)で符号化される. そして, 個々のニューロンは, それぞれの直前の反応の強さに応じて感度を変化させる. 入力が大きければ感度を下げ, 小さければ逆に上げるといふ具合である(図6).このようなモデルを計算機でシミュレーションしてみると, たしかに心理物理学実験で得られたような残効の特性が再現できた.

次に, 実際にそのような現象が神経系で起きているかどうかを調べるために, スナネズミという動物を用いて, 下丘と呼ばれる脳幹の部位からニューロン活動を記録する実験を行った[12]. この実験に用いた刺激の時間特性は人間の心理物理学実験に用いたものよりかなり短いなどの違いはあるものの, 基本的には上記のモデルと一致する結果が得られた.

聴覚定位残効は, 先行音の存在によって空間が伸び縮みすると表現することもできる. これは, 聴覚系が不正確であることを意味しているのであろうか. 実は, このような処理特性には情報処理上の利点がある. わずかに位置の異なるふたつの音を弁別する実験を行うと, 先行音がある場合には, ない場合に比べ, 先行音の付近だけ位置の弁別力が向上するのである. そのかわり, 先行音から離れたところでは弁別力は低下する[13]. 先行音の付近の空間が伸びるということは, ちょうどそこを虫眼鏡で拡大して見ているようなもので, 新しい音と先行音との差分にはきわめて敏感になる.

空間の伸び縮みの機能的な意義について, すこし詳し

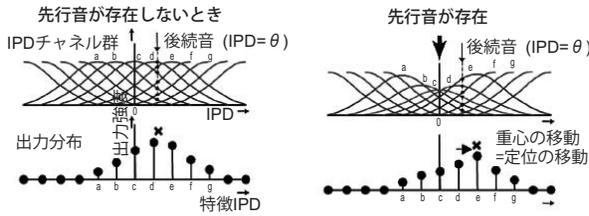


図 7 下丘ニューロンの特性に基づく聴覚定位残効のモデル

下丘には特定範囲の両耳間位相差 (IPD; この場合両耳間時間差と等価) に選択的に反応するニューロン群 (IPD チャンネル群) が存在する。IPD は、このニューロン群の反応強度分布の重心として符号化される (左)。先行音が存在すると、先行音の IPD を中心に感度が低下する。その状態で後続音が入力されると、反応強度分布の重心は先行音から遠ざかる方向にずれることになる (右)。これが知覚される定位残効に相当する。

く考えてみよう。一般的に言って、個々のニューロンは、ある範囲内では入力の値が大きくなると単位時間あたりのスパイク数が増えることによって入力値を表現している。しかしその範囲はあまり広くなく、上限付近では飽和し、下限付近では雑音に埋もれる。ところが、我々のモデルのように、入力値が大きいときに感度を下げ、小さいときに上げると、実効的には広い範囲の入力に対応してその差を表現することが可能となる。個々のニューロンがそのように振る舞えば、その全体としては、直前の入力の近くでわずかに異なった入力に対する興奮パターンの差が大きくなり、弁別力が向上することになる。これは知覚系一般に見られる順応現象に適用できる情報処理原理であろう。

以上のように、錯覚の分析から、聴覚における空間情報処理がダイナミックで適応的なものであることが明らかになった。聴取者自身も気づかぬうちに、入力される音に応じてあちこち処理の焦点を動かして、効率のよい処理を実現しているのである。

#### 4.2 同時性の順応

どの感覚モダリティでも従来あまり研究されてこなかったことだが、時間についても順応効果が生じることが最近わかってきた。聴覚でも音の時間順序判断が先行音の影響を受けるのだが [14]、ここでは視覚と聴覚にまたがった同時性判断における順応の例を紹介する [15]。

視覚刺激として黒い背景上の白いリング状のフラッシュ、聴覚刺激として短い純音を用い、両者を様々なタイミングで提示して、同時か否かを観察者に判断してもらおう (図 8a)。この際、その前に数分間、視覚刺激よりも聴覚刺激の方が早いペアを観察すると、聴覚刺激の方が先に提示されたときに同時と知覚されるようになる。逆に聴覚刺激よりも視覚刺激の方が早いペアを観察した後では、視覚刺激の方が先に提示されたときに同時と知覚されるようになる。このような主観的同時点の移動量は、

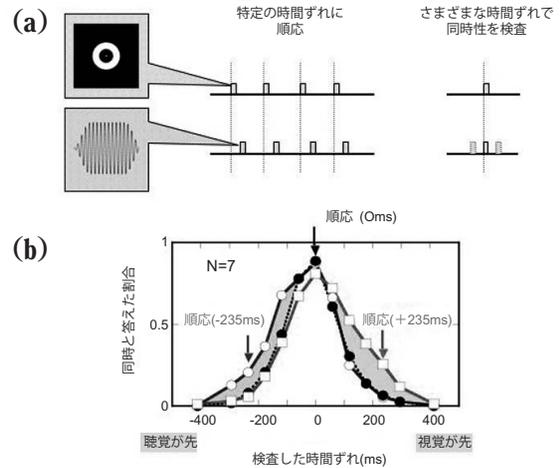


図 8 視聴覚の時間ずれに対する順応

(a) 刺激の模式図。(b) 順応による、視聴覚が同時と知覚される時間ずれの範囲の変化。

種々の条件にもよるが、順応した時間ずれの 15% 程度になることもある。また、視聴覚が同時と判断される検査刺激の時間ずれの範囲も、順応刺激の時間ずれの方向に広がる (図 8b)。つまり、一定の視聴覚の時間ずれを経験すると、そのずれを小さくする方向に同時性の判断基準が移動するのである。

順応効果は、直接的に視聴覚の同時性を判断しない課題でも観察できる。言葉で書くとちょっとややこしいが、まずは順応の話の前に、基本となる視聴覚現象を説明しよう。二つの小円が左右から近づいてきて、中央で交差するという視覚パターンを提示すると、ほとんどの場合小円が直進して通り過ぎてるように見える。ところが、交差する瞬間に短い音を鳴らすと、小円が衝突して左右に反発しているように見える率が高まる [16]。この現象も、他の視聴覚統合現象と同様、短い音を鳴らすタイミングと小円が交差するタイミングとが合っていないと効果がなくなる。つまり、視聴覚情報の同時性は、両者を結びつける強い手がかりとなる。さて、ここからが順応である。小円が交差する瞬間と、音が鳴る瞬間とが一定時間ずれているパターンをしばらく観察すると、反発効果が最大となる視聴覚のタイミングがそちらの方向にずれるのである [15]。このときのずれのパターンは、上述した、リングと純音の同時性を直接的に判断する場合の結果と非常によく似ている。つまり、視聴覚の時間ずれに対する順応効果は、主観的同时性判断だけではなく、知覚上の機能的な同時性にも影響することがわかる。

このような順応の機能的意義を考える上で重要なのは、視聴覚情報の同時性を判断するのは、脳にとっては原理的に非常に難しい情報処理課題だということである。第 1 に、光と音では空気中を伝わる速さが全く違う。光は秒速約 30 万 km、外界のイベントから目に到達

するまでにかかる時間は無視できる。一方音は秒速 340 m 程度なので、対象が 10 m ばかり先にあれば、光よりも 30 ms 程度遅れて耳に到達することになる。もちろん、この時間差は対象の距離によって変化する。第 2 に、仮に目と耳に同時に刺激が届いたとしても、そこから神経の活動が視聴覚それぞれの経路を通過して大脳皮質の視野と聴覚野に到達するまでの時間は無視できないくらい異なる。マカクザルの脳に電極を挿入して計測したデータや、人間の脳波や脳磁界の計測データによれば、神経活動が大脳皮質に到達する時間は聴覚の方が視覚よりも数十 ms も速い。この時間は注意の向け方や刺激の強さなどによっても変化する。これらの 2 種類の要因によって、外界のイベントから光と音が同時に発せられたとしても、脳内では、視聴覚情報の間に単純に予測できない時間のずれが生じることになる。視聴覚の同時性判断が原理的に難しいといったのはこのためだ。

では、この難しい情報処理課題を、脳はいかに解決しているのだろうか。ここで紹介した視聴覚の時間ずれに対する順応効果の意味するところは、入力された視聴覚情報間の時間のずれがある程度一定であれば、そのずれを小さくする方向に同時性の基準が移動するということである。つまり、脳は、固定的な同時性の基準を持っているのではなく、直近に経験した感覚情報をもとに、同時性の基準を適応的に変化させているのである。これが、物理的および神経的な非同期にもかかわらず、適切に視聴覚情報の同時性を判断するための脳の戦略であると考えられる。

## 5. おわりに

ここまで述べてきたように、聴覚の知覚世界は、時間も、空間も、そして対象事物の記述自体も、耳に入力された音響信号の物理的な特性とは単純に対応しない。つまり錯覚に満ちている。しかし、その錯覚は、聴覚の「賢さ」の現れとも言える。すなわち、実際の環境および脳の双方に不可避的に存在する種々の情報処理上の困難を克服し、適切にかつ効率よく環境内の事物の把握を行うための脳の戦略が、人工的な状況で顕在化したものが錯覚なのである。バーチャルリアリティをはじめ、聴覚に関わるメディアを創る際には、この聴覚の「賢さ」を十分に理解し、利用することが大切である。

## 参考文献

- [1] Warren, R. M. :Auditory perception:A new analysis and synthesis. Cambridge,Cambridge University Press(1999)[CD 付]
- [2] Bregman, A. S. :Auditory scene analysis:The perceptual organization of sound. Cambridge,MIT Press(1990)[CD 別売]
- [3] Deutch, D.:Musical illusions and paradoxes. Philomel Records(1995) [CD]

- [4] 内藤誠一郎, 柏野牧夫: イリュージョン・フォーラム (<http://www.brl.ntt.co.jp/IllusionForum/index2.html>), NTT コミュニケーション科学基礎研究所 (1998)
- [5] 中島祥好, 佐々木隆之, & ten Hoopen, G. : 聴覚デモンストラーション (<http://www.kyushu-id.ac.jp/~ynhome/JPN/Demo/illusions2nd.html>)(2000)
- [6] Saberi, K. & Perrott, D. R. : Cognitive restoration of reversed speech. *Nature*, 398, 760(1999)
- [7] 柏野牧夫: 閉鎖区間の前後に分散する手がかりに基づく日本語中閉鎖子音の知覚. *日本音響学会誌* 48, pp.76-86 (1992)
- [8] Kondo, H. & Kashino, M. :Distributed brain activation involved in the changes of auditory perceptual organization, an fMRI study on the verbal transformation illusion. The 27th MidWinter Meeting of ARO (2005)
- [9] Liberman, A. M. & Mattingly, I. G. :The motor theory of speech perception revisited. *Cognition*, 21, pp.1-36(1985)
- [10] Kashino, M. & Nishida, S. :Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect. *Journal of the Acoustical Society of America*, 103, pp.3597-3604(1998)
- [11] Carlile, S., Hyams, S., & Delaney, S. : Systematic distortions of auditory space perception following prolonged exposure to broadband noise. *Journal of the Acoustical Society of America*, 110, pp.416-424(2001)
- [12] Furukawa, S., Maki, K. Kashino, M. & Riquimaroux, H. (in press). Dependence of the interaural phase difference sensitivities of inferior collicular neurons on a preceding tone and its implications in neural population coding. *Journal of Neurophysiology*.
- [13] Kashino, M. : Adaptation in sound localization revealed by auditory after-effects. In A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Eds.), *Psychological and physiological advances in hearing*. London: Whurr Publishers, pp. 322-328(1998)
- [14] Okada, M. and Kashino, M. : The role of frequency-change detectors in auditory temporal order judgment. *NeuroReport* 14, pp.261-264(2003)
- [15] Fujisaki, W., Shimojo, S., Kashino, M., and Nishida, S. : Recalibration of audio-visual simultaneity. *Nature Neuroscience* 7, pp.773-778(2004)
- [16] Sekuler, R., Sekuler, A. B., & Lau, R.: Sound alters visual motion perception. *Nature*, 385, 308(1997)

## 【略歴】

柏野牧夫 (KASHINO Makio)

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 人間情報研究部 主幹研究員(感覚運動研究グループリーダー) 1987年東京大学文学部心理学科卒業。1989年同大学大学院修士課程修了、同年日本電信電話株式会社 NTT 基礎研究所入所。1998年同所主任研究員。2001年NTT コミュニケーション科学基礎研究所 (NTT 基礎研究所から組織改編) 企画担当研究総括担当課長。2002年同担当部長。2003年より現職。JST ERATO 下條潜在脳機能プロジェクト 潜在聴覚処理研究グループリーダー兼務。博士(心理学)。