■S3 群（脳・知能・人間）－ 2 編（感覚・知覚・認知の基礎）

2 章 聴覚と音声

（執筆者：西村竜一）［2008 年 5 月 受領］

■概要■

人と人とがコミュニケーションを行う手段として獲得し，一般に最も頻繁に利用しているのが，音声による情報の発信と，聴覚による受信である．音声コミュニケーションに必要な道具は，すべて身体に具備されており，ほかに道具を必要としないため，あらゆる場面で利用が可能である．音声は空気の振動にほかならないが，どのようにして発声器官が様々な音声情報を空気の振動に乗せ，また，聴覚がその情報を解読しているのかを知ることは，聴覚・音声を理解する上での第一歩である．また，聴覚は，音声の知覚ばかりでなく，その広いダイナミックレンジと全方位性のために，危険回避にも重要な役割を果たしている．つまり自分を取り囲む空間の情報を常に収集しており，「目に見えるところ」，「手の届くところ」以外に広がっている部分も含めて空間の知覚を行っている．可聴音は，指向性が必ずしも高くないため，複数の情報が混ざり合うことがしばしばある．しかし，聴覚処理機構は，この混ざりあった情報を必要な情報に分解する優れた波動情報解析の機能を有している．これらは，生理学的に実現されるものもあれば，認知学的に実現されるものもあり，これに倣った多くの信号処理技術の研究開発も行われている．

【本章の構成】

本章では，聴覚機構（2-1 節），音の知覚（2-2 節），聴覚の諸特性（2-3 節），音声の発声（2-4 節），音響信号の解析（2-5 節），マルチモーダル知覚（2-6 節）について，聴覚生理学，聴覚心理学及び音響信号処理の観点から，現在までに明らかにされている基本的な知見について述べる．

■S3 群 － 2 編 － 2 章

2-1 聴覚機構

（執筆者：西村竜一）［2008 年 5 月 受領］

2-1-1 外　耳

俗に耳と読んでいる耳介（pinna）［耳殻，耳翼とも呼ばれる］，及び，耳孔から鼓膜（tympanic membrane）までを指す外耳道（External auditory meatus）からなる．耳介の大きさは，上下の長さがおよそ 53 mm，前後の幅がおよそ 27 mm である [1]．形状はおおむね円錐状をしているが細部は個人ごとに異なり，前方から到来する音の収音の役割や，ここでの音の反射に起因する周波数特性の変化が，音源の到来方向を判断するための手がかりとして利用される [2]．左右方向の判断は，両耳で聴取された音のレベル差や時間差を比べることによっても可能であるが，片耳が塞がっている場合や，正中面における仰角の変化は，左右差では判断が困難であり，耳介の反射による変化の利用が有効な手がかりとなる．また，前後方向の判断については，耳介での回折による周波数変化が大きな役割を果たしている．一方，成人の外耳道の長さはおよそ 33 mm 程度であるが，必ずしも直線ではなくややＳ字に近い形状をしている．耳介のなかでも耳孔付近の部位は特に耳甲介（concha）と呼ばれ，この空間と外耳道との組み合せが共鳴器として働き，ヒトの場合は 2.5 kHz 付近が共振により 10 dB から 20 dB 程度増幅される [3]．

2-1-2 中　耳

鼓膜及び鼓膜から蝸牛（cochlea）の間の伝音器官を指し，鼓膜側から槌（つち）骨（malleus），砧（きぬた）骨（incus），鐙（あぶみ）骨（stapes）の三つの耳小骨が連結して，鼓膜の振動を蝸牛の卵円窓（oval window）［前庭窓とも呼ばれる］へ伝達している．耳小骨が収まっている空間は，鼓室（tympanic cavity）と呼ばれ，耳管（Eustachian tube）により鼻の奥の空間である鼻咽腔と空気的につながっている．このため，鼓膜の内側の圧力も外側の大気圧と同じになり，鼓膜のインピーダンスは空気のそれとほぼ等しくなっている．通常，発声時には耳管が閉ざされるが，耳管開放症の場合は耳管が開いたままになるため，自声音が耳に響いてうるさく聞こえる症状が現れる．中耳は，空気中を伝搬してきた音の振動エネルギーを，空気よりもはるかに大きなインピーダンスを有する蝸牛内のリンパ液に導くためのインピーダンス整合器の役割を果たしている．このインピーダンス整合は，三つの耳小骨を組み合わせてテコの原理を利用している．また，鼓膜の面積が卵円窓に接続している鐙骨底板よりも約 17 倍大きいため，変換比によっても 25 dB 程度の利得が得られ，これによってもインピーダンス整合が行われる [1]．

2-1-3 内　耳

蝸牛，前庭（vestibule），半規管（semicircular canals）からなる聴覚及び平衡感覚器官である．ヒトの蝸牛は，直径がおよそ 10 mm，高さがおよそ 5 mm で，約二巻半の渦巻状の形をしており，巻いている管を延ばすとおよそ 35 mm 程度になる．管のなかは，基底膜（basilar membrane）とライスネル膜（Reissner’s membrane）により，管に沿って三つの区間に隔てられ，中央に位置する中央階（scala media）［蝸牛管とも呼ばれる］がライスネル膜によって前

庭階（scala vestibuli）と分かれている．更に，基底膜によって鼓室階（scala tympani）とに分かれる．前庭階は前庭に始まり，蝸牛の頭頂部において鼓室階ともつながっている．また，その内部は外リンパ液で満たされている．中耳に接続している卵円窓は，前庭階の基底部にあたる．一方，鼓室階の基底部は，正円窓（roundwindow）［蝸牛窓とも呼ばれる］ により中耳の鼓室に面し，正円窓膜が張られている．中央階の内部には，有毛細胞（haircell）と神経終末を蔵するコルチ器（organ of Corti）が基底膜上にあり，有毛細胞の上に蓋膜（tectorial membrane）が覆い被さっている [4]．卵円窓から伝えられた音の振動は，前庭階の外リンパ液を振動させ，基底膜に進行波を生じさせる．中央階を満たしている内リンパ液は，外リンパ液よりも高い電位を有しており，内有毛細胞の毛と蓋膜との偏位によりイオンチャネルが開き，内有毛細胞の電位が変化することで，接続している聴神経に興奮活動が引き起こされる．基底膜の幅は，基底部から頭頂部へ向けて太くなっており，硬さは，基底部から頭頂部へ向けて柔らかくなっている．そのため，共振周波数が部位により異なる．その結果，到来音を構成する各周波数成分が，蝸牛において最も大きく振動するそれぞれの部位で電気信号に変換される．すなわち，蝸牛は機械 - 電気変換と周波数解析を同時に行う役割を果たしている．一方，平衡感覚器官である前庭は，直線運動の受容器である球形嚢と卵形嚢を格納する空間を指す．球形嚢は，結合管で中央階と，卵形嚢とは球形嚢管及び卵形嚢管を介して接続されており，内リンパ液で満たされている．また，上半規管（Superior semicircular canals），外半規管（Lateral semicircular canals），後半規管（Posterior semicircular canals）は，三半規管と総称され，回転運動の知覚をつかさどっている．

■S3 群 － 2 編 － 2 章

2-2 音の知覚

（執筆者：西村竜一）［2008 年 5 月 受領］

2-2-1 聴覚域値

音の物理的な大きさに相当する音圧レベルは，20 μ Pa を基準となる 0 dB として定義される．この基準となる音圧レベルは，1000 Hz の純音に対する聴覚域値を目安として定められている．音圧レベルを下げたときに聞こえなくなる限界が，聴覚域値（最小可聴値とも呼ばれる）である．聴覚域値は，ISO において規格化されており，反射のない空間である自由音場で測定した最小可聴音場（minimum audible field : MAF）とヘッドホンを用いて測定した最小可聴音圧（minimum audible pressure : MAP）がある．ヒトの聴覚は，2 kHz から 4 kHz 付近が最も聴覚域値が低く，音圧レベルにしておよそ－5 dB 程度である．外耳や中耳における伝達特性が，これらの周波数帯域で最もエネルギー損失が少ないことに起因していると考えられる．聴覚域値は，周波数によって大きく変化し，低い周波数や高い周波数では域値が上昇する．一方，音圧レベルを上げたときに，音以外の触覚や痛覚が耳に生じ始めるレベルを最大可聴値という．最大可聴値の周波数依存性はあまり高くなく，音圧レベルで 120 dB から 130 dB 程度である．その結果，ダイナミックレンジが周波数によって大きく変化する．また，加齢とともに聴覚域値は上昇するが，最大可聴値はほとんど変化しない．そのため，ダイナミックレンジが狭まり，音圧レベルを上昇させてもなかなか音が聞こえ始めないが，いったん聞こえ始めると急激に音が大きく聞こえるリクルート現象を引き起こす．一般に，加齢に伴う聴覚域値の上昇は，高い周波数のほうが低い周波数よりも顕著に現れる．

2-2-2 大きさの知覚

音の感覚的な大きさは，ラウドネス（loudness）と呼ばれる．音圧レベルが音の物理量であるのに対し，音の大きさは主観量である．そのため，同じ音圧レベルであってもラウドネスは個人ごとに異なり，また，周波数によっても変化することが知られている．異なる周波数の純音に対して同じ大きさに聞こえる平均的な音圧を結んだ等ラウドネス（レベル）曲線（等感曲線とも呼ばれる）は，工学的な応用でも利用されることから ISO で規定されている [5]．ラウドネスの単位として用いられる sone は，音圧レベルが 40 dB の 1 000 Hz の純音を基準となる 1 sone と定めている．比率尺度であるため，ラウドネスが 2 倍になるものが 2 sone となる．また，音の大きさのレベルを表す単位として phon が用いられ，その音と同じ大きさに知覚される 1000 Hz 純音の音圧レベルの値で与えられる．騒音は人間が感じるうるささであることから，物理的な音圧よりも音の感覚的な大きさとの相関が高いため，音圧レベルに対して等ラウドネス曲線の特性で重み付けを行い，聴感補正して算出される．聴感補正を行う特性はいくつか提案されており，そのなかで A-特性と呼ばれる特性を用いて算出したレベルは，騒音レベル（A-weighted sound pressure level）と呼ばれる．騒音レベルの単位は dB であり，A-特性の重み付けがされていることを明示するために dB(A)と記されることもある．日本では，騒音レベルの単位を「ホン」（カタカナ書き）と表記する場合もあるが，音の大きさのレベルを表す単位である phon とは別物である．

2-2-3 高さの知覚

音を音階上に順序づけることができる聴覚の属性が，音の高さ（pitch）である．単一周波数である純音の場合には，主に周波数がこれを決める．一方，複数の周波数成分からなる周期的な複合音においては，主に最も低い周波数（基本周波数）に対応する高さが知覚される．また，基本周波数も含めて低次の周波数が欠けている高調波複合音においては，存在しない基本周波数の高さが知覚される．信号の包絡線が基本周波数の逆数の周期で変化しており，それを手がかりに音の高さを知覚していると考えられている[6,7]．音の高さも音の大きさと同じく主観量であるため，直接的に測定することはできない．また，周波数だけではなく，音圧レベルにも依存して変化する．音の高さを表す単位には，mel が使用される．音圧レベルが 40 dB の 1 000 Hz の純音を 1 000 mel と定義し，音の高さが 2 倍に感じられると，数値も 2 倍になるように実験的に定められた比率尺度である[8]．

2-2-4 音色空間

音色とは，音の属性のなかで大きさと高さが共に等しくても，その 2 音が違って聞こえるときのその相違の部分に相当する属性である．音色の知覚は，大きさや高さのように一次元的ではなく多次元的であることから音色空間と呼ばれる．音色空間の測定には，対象刺激音間の（非）類似度を基にして空間を構築する多次元尺度構成法（multidimensional scaling: MDS）や，SD（semantic differential）法と因子分析（factor analysis）を組み合わせた手法が用いられる．後者のほうが，一般に様々な形容詞対を用いて評価が行われるため，得られた空間の各軸の意味づけは多次元尺度構成法よりも容易である．これまでの研究により，「明るさ因子」，「やわらかさ因子」，「美的・叙情的因子」，「量的・空間的因子」などの因子が提案されていたり，そのうちのいくつかに替えて，「迫力因子」，「金属性因子」などの因子も提案されている[9-11]．因子の決め方には多様性があるため複数の因子が提案されているが，次元数については多くの研究者で一貫性があり，3～4 程度と考えられている．

2-2-5 音響空間の知覚

音の到来方向は，水平面については両耳間時間差（interaural time difference: ITD）［両耳間位相差（interaural phase difference: IPD）とも呼ばれる］及び両耳間レベル差（interaural level difference: ILD）［両耳間強度差（interaural intensity difference: IID）とも呼ばれる］が主な判断要素であると考えられている．両耳への音の到来時間差は，経路差のために音が正面あるいは真後から到来する場合には小さく，横方向から到来する場合には大きくなる．低い周波数の音については，聴神経の発火が位相に同期するため，両耳での位相差を基にして音の到来方向を判断することができる[12]．一方，高い周波数については頭部での遮蔽の効果が大きいため，左右耳におけるレベル差が横方向から到来する音で大きくなり，音の到来方向の判断が可能となる．これらに加えて，耳介による回折に起因するスペクトル形状の変化により，前後や仰角方向の判断が行われる．特に両耳を結ぶ直線を中心にして円錐を描いたときの底面の縁に相当する矢状面の音源については，両耳への到来時間差が等しくなる．そのため，周波数スペクトルの外形の情報を手がかりにして，音源方向を推定する必要がある[13,14]．これ以外にも，反射音のある空間や複数の音源が存在する場合には，直接音と反射音の時間差や残響，両耳間相関なども音の広がりの知覚に影響を及ぼしていると考えられている．

■S3 群 － 2 編 － 2 章

2-3 聴覚の諸特性

（執筆者：西村竜一）［2008 年 5 月 受領］

2-3-1 マスキング現象

ある音信号の存在が別の音信号の知覚を抑制する現象であり，周波数マスキング（frequency masking）と継時マスキング（temporal masking）の二つが知られている．ここで，抑制するほうの信号はマスカ（masker），抑制されるほうの信号はマスキ（maskee）と呼ばれる．周波数マスキングは，マスカとマスキが時間的に同時に発生しているときに観測されることから，同時マスキング（simultaneous masking）とも呼ばれる．マスキングの起源の一端は蝸牛の構造に見ることができ，音は基底膜に進行波を生じさせるが，卵円窓から遠いところに低い周波数の共振点があるため，低い周波数成分が存在していると，その手前の高い周波数成分に感応する部位にも振動が伝搬し，この部位での感度が低下する [15]．そのため，周波数マスキングは，一般に低い周波数成分がそれよりもやや高い周波数成分をマスクすることが多い．一方，継時マスキングは，時間的にマスカとマスキが重なって発生していないため，非同時マスキングとも呼ばれる．また，マスカとマスキの時間関係によって二つに分けられ，マスカがマスキよりも時間的に先行する場合を順向性マスキング（forward masking）［前方性マスキングとも呼ばれる］，その逆を逆向性マスキング（backward masking）［後方性マスキングとも呼ばれる］と呼ばれる [16, 17]．大きな音が蝸牛に到来すると聴神経の発火頻度を増大させるが，その神経インパルスが脳へ到達するのに要する時間には広がりがあり，その結果，先行あるいは後続する神経インパルスと重なって感度が低下すると考えられている．これらの聴覚生理学的な理由に端を発するマスキング現象以外にも，より高次の脳活動に起因して発生するマスキングもあり，それらは情報マスキング（information masking）と呼ばれる．

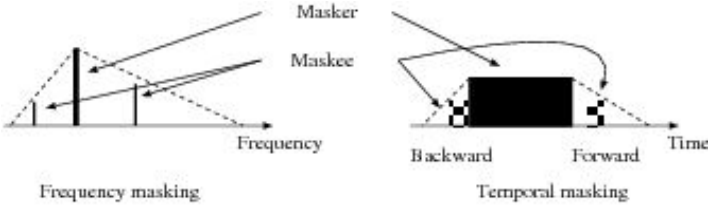

図 2・1 周波数マスキングと継時マスキング

2-3-2 共変調マスキング解除

ある周波数帯域でマスキング現象が生じているときに，別の周波数帯域にマスカと同じような振幅変調を行った信号を提示すると，マスキング量が低下する．この現象を，共変調マスキング解除（co-modulation masking release）という．聴覚は，耳に到来した音信号を周波数的に近傍に中心周波数をもつ臨界帯域（critical band）ごとにまとめて聞いている．したがって，マスキを純音，マスカをマスキの周波数を中心とする帯域雑音として帯域幅を徐々に広げると，臨界帯域を超えるまではマスキング量も増加するが，臨界帯域を超えるとそれ以上にマスキング量は増加しない．しかし，マスカにゆっくりとした振幅変動を与えていると，

臨界帯域を超えても帯域が広がるにつれて，マスキング量が低下する現象が観測される [18]．これは，複数の臨界帯域の出力信号に対して，変調の情報を用いて関連性を評価し，関連性のあるものをつなげてひとまとまりに知覚している可能性を示唆している．このマスキング量の低下は，最大で 10 dB 程度に及ぶことがある．また，マスカの帯域を広げるのではなく，全く別の周波数帯域に提示した場合や，一方の耳にマスカとマスキを提示し，もう一方の耳に別の周波数の信号を提示した場合にも，この現象は観測される [19, 20]．

2-3-3 先行音効果

数ミリから数十ミリ秒の時間差で到来した二つの音源の音像が，最初に到来した音の音源方向にまとめて知覚される現象を先行音効果（precedence effect）という [21]．この効果は，ハース効果（Haas effect）や第一波面の法則（the law of the first wave front）と呼ばれることもある [22]．ただし，音像の定位方向は，先行音と後続音の音響的な特徴の関係に依存して決まる．また，数ミリ秒よりも時間差が短い場合には，複数の音源の平均的な方向から到来しているかのように知覚され，加法定位と呼ばれる．先行音効果の生起には，先行音と後続音の音響的な性質が似ていることが必要である．音響エコーは，同じ音が時間的にずれて到来するため，最も先行音効果が起こりやすい例である．一般的には，先行音が後続音の知覚を抑圧することに起因していると考えられている．したがって，音像定位への影響だけでなく，先行音の存在により後続音の音響的な弁別域が低下する現象も先行音効果の範疇とする考え方もある．

2-3-4 カクテルパーティ効果

複数の音のなかから，ほかの音をできるだけ無視して，着目している音のみを聞き取ることができ，このような現象をカクテルパーティ効果という．カクテルパーティのような雑踏のなかにおいても，会話をしている相手の声を聞き分けることができることから，このような名前で呼ばれている．これを実現する一つの方法は，二つの耳を利用して特定の方向から到来する音を聞き取り，そのほかの方向から到来する音を抑制する指向性合成の考えに基づく方法である．しかし，片耳であってもある程度のカクテルパーティ効果が観測されることから，特定の音響パターンを抽出する脳の高次機能を利用しているとも考えられている．また，カクテルパーティ効果は，雑踏のなかにあっても自分の名前などの特定の言葉や音が発せられた際に，その音が聞き取れる能力を指す場合もある．

■S3 群 － 2 編 － 2 章

2-4 音声の発声

（執筆者：西村竜一）［2008 年 5 月 受領］

2-4-1 声　帯

声帯は，食道と気道が分かれてすぐの気道の中（喉頭）にある，弁のような役割を果たす器官である．呼吸をしているときは，通常開放した状態を保っており，母音などの有声音を発声する際や嚥下時に閉鎖される．声帯を閉鎖した状態で肺から空気を押し出そうとすると内圧が高くなり，閉じておく力よりも強くなると声帯が開いて空気が抜け，内圧が下がることで再び声帯が閉じる．音声を発声する際には，この現象を繰り返すため，声帯で生成される音圧の変化は，のこぎり波状となる．大きな声を出すときには，声帯を閉鎖している期間が長くなり，これにより内圧も上がり大きな圧力変化が生じる．このため，強い発声ではのこぎり波状になり，弱い発声では正弦波に近くなる．また，声帯の長さは，男性のほうが女性よりも長く，子供は短い．この差により振動の周期に違いが生じ，日本人男性の基本周波数の平均値は 100 Hz から 150 Hz 程度，女性では 200 Hz から 300 Hz 程度，子供では 250 Hz から 400 Hz 程度となる [23].

2-4-2 声道モデル

声帯から口までを断面積が連続的に変化する音響管とみなし，この音響管の伝達特性によって，声帯で生成した信号源波形が口や鼻孔から放射されるまでにフィルタリングされて音声が生成されていると考えるモデルである．実際には，区分的に断面積が一定の音響管をいくつも接続してモデル化することで，音響管内の音波の伝搬方程式に基づいて音声波形をシミュレーションすることができる．逆に音声波形から声道の形状をある程度推測することも可能になる．分岐を無視すると母音の場合には伝達関数が全極型となり，鼻音の場合には極零型となる．また，子音についても，声道の途中に音源を設けることにより，同様のモデル化が可能である．

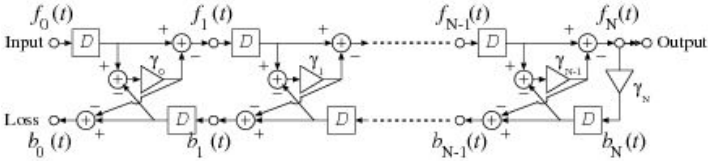

図2・2 声道モデルのシグナルフロー

2-4-3 調　音

呼気をエネルギー源とする信号源波形に対し，言語的に意味のある，より多様な情報を乗せるために，唇，舌，顎などを動かして声道中に閉鎖，狭め，分岐を形成し，音声波形に変化を加えて言語音を生成することを指す．調音の仕方を調音様式（manner of articulation）［調音方式とも呼ばれる］といい，調音様式の違いにより，母音，半母音，摩擦音，破裂音，鼻音などに音声が分類される．一方，調音を行う位置によっても子音は分類が可能であり，唇

音，歯音，硬口蓋音，軟口蓋音，声門音などに分類される．調音を行う各器官は，瞬時に該当する動きを実行することができないために，時間的に前後する調音の影響を大きく受ける．この影響は調音結合（coarticulation）と呼ばれ，連続音声の聞き取りに深く関与している．

■S3 群 － 2 編 － 2 章

2-5 音響信号の解析

（執筆者：西村竜一）［2008 年 5 月 受領］

2-5-1 スペクトル解析

フーリエ変換により，解析対象信号を複素正弦波が基底系をなす空間に写像することで，時間信号である音声信号を周波数表現で観測するのに用いられる．人間の聴覚系も周波数分析を行って音を知覚していることから，スペクトル解析により，音情報が脳の聴覚野に導かれる際の信号に類似した情報が得られる．信号をフーリエ変換すると複素数が得られるため，振幅の周波数特性である振幅スペクトルと，位相の周波数特性である位相スペクトルの二つで表現される．また，振幅を 2 乗して対数を取り 10 倍することで，デシベルで表現したパワースペクトルが用いられる場合も多い．パワースペクトルには振幅の情報しかなく，時間波形に戻すためには位相スペクトルの情報も必要になる．したがって，パワースペクトルが同一であっても位相スペクトルが違っていれば異なる時間波形となり，違った音色に聞こえる．ただし，音声の知覚において位相情報はほとんど利用されていないため，音声の解析には主に振幅スペクトルやパワースペクトルが重要となる．

2-5-2 オクターブ分析

人間の聴覚は，近い周波数の複数の成分をまとめて知覚しており，この周波数帯域は臨界帯域（critical band）と呼ばれる．この臨界帯域の周波数幅は，周波数が上昇するにつれて周波数に比例して広がることが知られている．そこで，周波数を対数に変換してオクターブ間隔の成分でまとめて解析すると，聴覚系における処理に近い信号の解析結果が得られる．試験用広帯域雑音として用いられるピンク雑音は，高い周波数へ向かって－3 dB/octave で減衰する低周波数比あたりのパワーが一定な信号であるため，この信号をオクターブ分析すると平坦なスペクトルとなる．オクターブ分析は，音の大きさと密接に関係する騒音の評価などにおいて用いられることが多く，人の臨界帯域幅を考慮して 1/3 オクターブ分析や，1/1 オクターブ分析がしばしば用いられる．

2-5-3 ケプストラム分析

信号をフーリエ変換して得られた複素数の絶対値を対数に変換してから逆フーリエ変換する解析手法をケプストラム分析（cepstrum analysis）という．また，絶対値を取らずに，複素数のまま対数変換して逆フーリエ変換して得られる係数は，複素ケプストラムと呼ばれる．出力は時間と同じ次元であるが，時間と区別するためにケフレンシー（quefrency）と呼ばれる．周波数領域で信号操作することを一般にフィルタリングと呼ぶのに対し，ケフレンシー上で何らかの信号処理を施すことはリフタリング（liftering）と呼ばれる．音声は，声帯で生成した音源波形に声道の特性に対応するインパルス応答を畳み込むことで表現される．これらの特性は時間軸では畳み込みで表現されるが，周波数領域では乗算で表されるため，ケプストラム分析を行うと乗算が加算演算で表現される．これを利用して，声帯での生成信号のピッチ特性と声道におけるフィルタリングのスペクトル包絡特性に分離することが可能となる．このようにケプストラム分析は音声の特徴をよく表すことから，二つの信号をケプスト

ラム分析して得られた係数（ケプストラム係数）の差の二乗平均はケプストラム距離（cepstral distance）と呼ばれ，音声認識や音声の音質劣化の指標として用いられる．ケプストラム分析は，畳み込みに対しての準同形変換とみなすことができ，反射波や基本周波数の検出などにも用いられる．

2-5-4 音声知覚

音声の言語としての情報は，声帯の周期的信号の有無，調音による声道の形状変化に伴う共振・反共振の特性，及び，持続と遷移に関する時間の三つで主に特徴づけられる．これらによって生成された音声信号の振幅スペクトル包絡の時間変化が，言語的な意味を形成するのに重要な役割を果たす．特に定常な母音は，いくつかの特定の周波数領域上にエネルギーの大きな成分をもち，この部位はホルマントと呼ばれる [24]．このホルマントの中心周波数の分布が，母音の種類の識別に利用される．非音声と音声の知覚は異なる面が少なくない．脳における音声の言語情報の処理の多くは左大脳半球で行われる人が多いが，旋律の同定や比較については右大脳半球が基本的な役割を果たしていることが多い [25]．また，音声はカテゴリー的な知覚が行われており，音声の音響的特徴をもった信号を知覚するときにのみ働く，音声モード（speech mode）の存在が議論の的になっている．

■S3 群 － 2 編 － 2 章

2-6 マルチモーダル知覚

（執筆者：西村竜一）［2008 年 5 月 受領］

2-6-1 聴覚と視覚のマルチモーダル知覚

視覚的な刺激の有無や内容によって，聴覚刺激の知覚は影響を受ける．代表的なものに，腹話術効果（ventriloquism effect）とマガーク効果（McGurk effect）があげられる．腹話術効果は，音声が実際には人形の口からは発声されていないのに，視野中で音声と同期して動くものがほかにないため，実際にその音を発声している腹話術師ではなく，人形から発声されているように知覚される現象である [26]．一方，マガーク効果は，例えば，視覚的に“ga”と発話する映像を見せつつ，それと同期して“ba”という音声を聞かせると，“da”という視覚的にも聴覚的にも与えていない別の音声を知覚する現象である [27]．腹話術効果は，空間知覚における視覚情報の聴覚情報に対する優位性を示唆しているが，マガーク効果は，音声知覚における聴覚情報と視覚情報の融合知覚の一例を示している．このほかにも，聴覚情報が視覚情報に対して優位性を示すものもあり，時間的な知覚がそれにあたる．例えば，何回点滅したのかの判断がやや困難なぐらい短時間で数回点滅する映像を見せ，それと同時に連続するクリック音を提示すると，クリック音の回数で点滅しているかのように見えてくる．

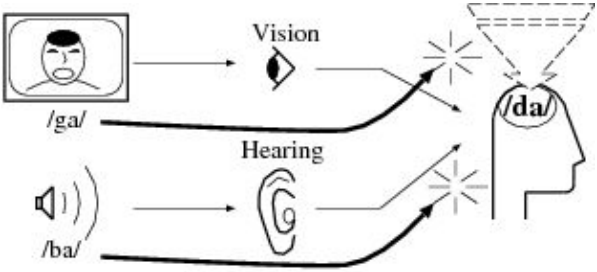

図 2・3 マガーク効果の例

2-6-2 聴覚と体性感覚のマルチモーダル知覚

視覚と体性感覚のマルチモーダル知覚としてベクションが知られているが，聴覚と体性感覚の間にもマルチモーダル知覚の存在が確認されている．ベクションは，例えば，列車の窓から隣のホームの列車を見ていたときに，隣の列車が動き出したのに自分の列車が動き出したと錯覚する現象を指す．この場合，視覚刺激の動きの方向と，自身が知覚する動きの方向は逆方向になる．一方，聴覚の場合は，例えば，前後に周期的に体を動かしながら聴覚刺激を左右に振らすと，斜め方向に体が移動しているように知覚され，その振動は聴覚刺激と同期する．このように，視覚の場合と聴覚の場合とで振る舞いが異なり，その知覚モデルについてはいまだ解明されていない．

■参考文献■

1) (社) 日本音響学会編, “音響用語辞典,” コロナ社, 1988.
2) M. B. Gardner and R. S. Gardner, “Problem of localization in the median plane: effect of pinna cavity occlusion,” J. Acoust. Soc. Am., vol.53, no.2, pp.400-408, 1973.
3) F. M. Wiener and D. A. Ross, “The pressure distribution in the auditory canal in a progressive sound field,” J. Acoust. Soc. Am., vol.18, no.2, pp.401-408, 1946.
4) J. O. Pickles, 谷口郁雄監訳, “聴覚生理学,” 二瓶社, 1995.
5) ISO 226, “Acoustics ? Normal equal-loudness-level contours,” 2003.
6) G. F. Smoorenburg, “Pitch perception of two-frequency stimuli,” J. Acoust. Soc. Am., vol.48, No.4(2), pp.924-942, 1970.
7) R. Plomp, “Pitch of complex tones,” J. Acoust. Soc. Am., vol.41, no.6, pp.1526-1533, 1967.
8) S. S. Stevens and J. Volkmann, “The relation of pitch to frequency: a revised scale,” Am. J. Psychol., vol.53, no.3, pp.329-353, 1940.
9) 三戸左内・北村音壱・難波精一郎・松本倫平, “音色の研究 II：再生音の音質に関する因子的研究,” 音響学会講演論文集, p.55, 1961.
10) 曽根敏夫・城戸健一・二村忠元, “音の評価に使われることばの分析,” 音響誌, vol.18, no.6, pp.320-326, 1962.
11) J. M. Grey, “Multidimentional perceptual scaling of musical timbres,” J. Acoust. Soc. Am., vol.61, no.5, pp.1270-1277, 1977.
12) J. E. Rose, J. F. Brugge, D. J. Anderson and J. E. Hind, “Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey,” J. Neurophysiol., vol.30, no.4, pp.769-793, 1967.
13) J. Hebrank and D. Wright, “Spectral cues used in the localization of sound sources on the median plane,” J. Acoust. Soc. Am., vol.56, no.6, pp.1829-1834, 1974.
14) R. A. Butler and K. Belendiuk, “Spectral cues utilized in the localization of sound in the median sagittal plane,” J. Acoust. Soc. Am., vol.61, no.5, pp.1264-1269, 1977.
15) R. L. Wegel and C. E. Lane, “The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear,” Phys. Rev., vol.23, no.2, pp.266-285, 1924.
16) L. L. Elliot, “Backward and forward masking of probe tones of different frequencies,” J. Acoust. Soc. Am., vol.34, no.8, pp.1116-1117, 1962.
17) J. H. Patterson, “Additivity of forward and backward masking as a function of signal frequency,” J. Acoust. Soc. Am., vol.50, no.4(2), pp.1123-1125, 1971.
18) J. W. Hall and M. A. Fernandes, “The role of monaural frequency selectivity in binaural analysis,” J. Acoust. Soc. Am., vol.76, no.2, pp.435-439, 1984.
19) M. F. Cohen and E. D. Schubert, “Influence of place synchrony on detection of a sinusoid,” J. Acoust. Soc. Am., vol.81, no.2, pp.452-458, 1987.
20) G. P. Schooneveldt and B. C. J. Moore, “Comodulation masking release as a function of masker bandwidth, modulator bandwidth and signal duration,” J. Acoust. Soc. Am., vol.85, no.1, pp.273-281, 1989.
21) H. Wallach, E. B. Newman and M. R. Rosenzweig, “The precedence effect in sound localization,” Am. J. Psychol., vol.62, no.3, pp.315-336, 1940.
22) J. Blauert, “Spatial hearing,” MIT Press, Cambridge, 1983.
23) 中田和男, “音声,” コロナ社, 1977.
24) J. L. Flanagan, “Speech analysis,” synthesis and perception, Springer-Verlag, 1965.
25) D. E. Broadbent and M. Gregory, “Accuracy of recognition for speech presented to the right and leftears,” Q. J. Exp. Psychol., vol.16, no.4, pp.359-360, 1964.
26) H. A. Witkin, S. Wapner and T. Leventhal, “Sound localization with conflicting visual and auditorycues,” J. Exp. Psychol., vol.43, no.1, pp.58-67, 1952.
27) H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” Nature, vol.264, pp.746-748, 1976.