

Sous la direction de
Serban Ionescu
et
Alain Blanchet

**Psychologie cognitive
et
bases neurophysiologiques
du fonctionnement cognitif**

Volume coordonné
par Daniel Gaonac'h



PRESSES UNIVERSITAIRES DE FRANCE

PAR SOPHIE DONNADIEU, ÉDOUARD GENTAZ
ET CHRISTIAN MARENDAZ

La perception désigne l'ensemble des mécanismes et procédures qui nous permettent de prendre connaissance du monde qui nous entoure sur la base des informations élaborées par nos différents sens. Ce chapitre traitera des sens visuel, auditif et haptique (tactilo-kinesthésique). La présentation de ces modalités sensorielles a été organisée en tenant compte, d'une part, de l'état des problématiques et des connaissances propres à chaque modalité et, d'autre part, du fait que, dans la vie de tous les jours, ces modalités interagissent. La section A (vision) est principalement consacrée aux modèles de reconnaissance des scènes et des objets. De tels modèles n'existent encore pas pour l'audition (section B) et le sens haptique (section C) ; le développement de ces sections est centré sur les dimensions et les structures perceptives qui sont extraites dans chacune de ces modalités. Nos perceptions ne sont pas le reflet en miroir de la structure physique du monde ; la section D est consacrée aux *illusions* perceptives car elles révèlent la complexité, et paradoxalement la *fiabilité*, des traitements engagés dans l'acte de percevoir. Enfin, nos sens interagissent en permanence ; la section E expose l'intégration intermodale.

A - LA PERCEPTION VISUELLE

La vision est un sens extrêmement rapide, automatique et fiable... comme peut nous le montrer aisément toute activité de *zapping* devant notre écran de télévision. En effet, malgré le défilement rapide des images des différentes chaînes, nous sommes capables de voir ce que présente chacune d'elles. Cette performance n'est pas due à l'uniformisation des programmes TV, mais à la rapidité de la perception visuelle. Par ailleurs, l'expérience télévisuelle permet de faire deux autres constats. Le premier est que la reconnaissance des images se fait sans effort et en quelque sorte « malgré nous ». Cela traduit le caractère relativement automatique de la perception visuelle. Le second

constat est que si des amis font l'expérience en même temps que nous, ils percevront approximativement la même chose, même s'ils l'interprètent différemment. Cela traduit la *fiabilité* de la perception visuelle. Cette *fiabilité* indique que les mêmes mécanismes de traitement sont à l'œuvre d'un individu à l'autre.

Cependant, lorsque l'on essaie de préciser la nature de ces mécanismes et leur fonctionnement d'ensemble, les difficultés théoriques apparaissent. En effet, la simplicité phénoménologique de la vision masque la très grande complexité des traitements qui la sous-tendent. Cette complexité transparaît avec force et drame dans le comportement et devenir des personnes qui recouvrent la vue après une *cécité* précoce. Si leurs yeux retrouvent leur fonctionnalité, ces personnes ne voient pas pour autant. Leur perception est souvent morcelée, fragmentaire, incompréhensible, très coûteuse au niveau attentionnel et anxiogène, comme si leur vision restait bloquée aux premières étapes du traitement cortical, les mécanismes qui permettent à tout un chacun de percevoir un environnement tridimensionnel structuré n'étant pas fonctionnels (le lecteur intéressé pourra se reporter aux cas emblématique de Virgil, un aveugle recouvrant la vue, décrit par Sacks, 1995).

Qu'est-ce que voir ? Dans la première section seront résumées les propriétés et contraintes (*rapidité*, *automatisme*, *fiabilité*, gestion des référentiels spatiaux) de la perception visuelle ; dans la seconde, les modèles (traitement du signal, extraction de caractéristiques primitives) dont dispose la psychologie pour rendre compte de notre capacité à percevoir une scène visuelle et les objets qui la composent.

I - Les propriétés et les contraintes de la perception visuelle

a. La rapidité

Pour déterminer le *temps de traitement* d'une image, Cattell (1886) puis, un siècle plus tard, notamment Intraub (1979) ont utilisé un paradigme de temps de réponse. Le sujet devait identifier (réponse verbale) le plus rapidement possible des lettres, des mots, des images d'objets ou de scènes présentés aléatoirement. Les temps de réponse obtenus (temps séparant le début de la présentation du début de la réponse) variaient de 500 à 900 ms selon le stimulus présenté. Ces temps, déjà brefs, surestimaient pourtant le *temps de traitement* de l'image car ils incluaient le temps de préparation et d'exécution de la réponse. Comment alors mesurer plus spécifiquement le *temps de traitement* ? La méthode expérimentale la plus pertinente consiste à calculer un seuil temporel d'identification d'images présentées en rafale. Comme dans l'exemple du *zapping* ci-dessus, le sujet doit identifier différentes images qui lui sont présentées successivement, le temps de présentation des images variant d'un essai à l'autre en respectant une procédure psychophysique de calcul de seuil (voir encadré, et Bagot, 1996 ; Bonnet, 1986). L'hypothèse est que le sujet ne pouvant traiter l'image que durant sa présentation (puisqu'ensuite une autre image apparaît, qui masque visuellement la première et doit être analysée), le temps de présentation minimum des images permettant leur reconnaissance correspond alors au *temps de traitement*. Le seuil obtenu est d'une centaine de millisecondes (Thorpe, 1988).

La notion de seuil

La notion de seuil est le concept central de la psychophysique. La psychophysique, initiée par Weber au XIX^e siècle (et à sa suite Fechner, Helmholtz, Wundt, Stevens, etc.) est un domaine d'étude de la psychologie expérimentale dont l'objet est de déterminer les relations quantitatives qui existent entre un stimulus physique et la perception qu'on en a. On distingue deux types de seuil : le « seuil absolu » (SA), qui est la « valeur minimale du stimulus pouvant provoquer une sensation juste perceptible », et le « seuil différentiel » (SD), qui est la « quantité minimale dont un stimulus doit varier pour produire une sensation différente » (par exemple plus lourd, plus lumineux, plus chaud, etc.). Par extension, chercher à déterminer le temps de présentation minimum d'une image permettant son identification s'apparente à un problème d'estimation d'un SA. La méthode la plus classique d'estimation de ce seuil est celle dite des « stimuli constants ». Le temps de présentation de l'image varie de manière aléatoire d'un essai à l'autre entre des bornes définies par des essais préliminaires (par exemple entre 10 ms et 500 ms). L'image est suivie de la présentation d'une image « masque » (par exemple, un ensemble de traits de taille et d'orientation aléatoires) destinée à stopper la rémanence rétinienne (persistance sensorielle de l'image alors que celle-ci n'est plus présentée). Lorsqu'on réalise l'histogramme du taux de reconnaissance (ordonnée) en fonction du temps de présentation (abscisse), on observe une fonction psychométrique de type sigmoïde (intégrale de la distribution normale des SA). La forme de cette fonction, très évasée dans ses limites supérieure et inférieure, ne permet pas d'estimer précisément le temps de présentation correspondant à 100 % de reconnaissance. C'est la raison pour laquelle on détermine habituellement le seuil sur la base d'un taux de réussite de 80 %. Cette valeur, très proche de 100 %, se situe dans une zone sensible de la fonction (pente élevée, qui traduit le fait que le taux de reconnaissance est très sensible à toute variation du temps de présentation) ; elle est donc plus aisée à estimer.

On trouvera aussi des illustrations des techniques d'évaluation sensorielle part. 3, chap. 17.

b. L'automatisme

Chez l'adulte la perception visuelle est automatique dans le sens où le processus de traitement est irrépressible : dès qu'on ouvre les yeux, on ne peut s'empêcher de voir quelque chose et de l'identifier. Cet *automatisme* transparait pleinement dans la tâche de Stroop, où le sujet doit dénommer la couleur de l'encre dans laquelle est écrit un nom de couleur. L'interférence du nom de la couleur sur l'identification de l'encre indique que le sujet a traité le mot jusqu'au niveau lexical, c'est-à-dire bien au-delà de la composante couleur requise par la consigne (cf. aussi part. 1, chap. 3, p. 154, et chap. 5, p. 203). Bien évidemment, cet effet n'existe que chez les personnes qui savent lire, c'est-à-dire qui ont appris et sur-appris à décoder et interpréter cette forme particulière, artificielle, qu'est un mot. Avec l'apprentissage, cette forme est devenue de plus en plus familière et son traitement de plus en plus automatisé. Il en est de même pour les objets que nous percevons quotidiennement, car l'environnement dans lequel nous vivons varie peu. Au fil du temps, nous avons appris à automatiser leur traitement.

c. La fiabilité et les illusions

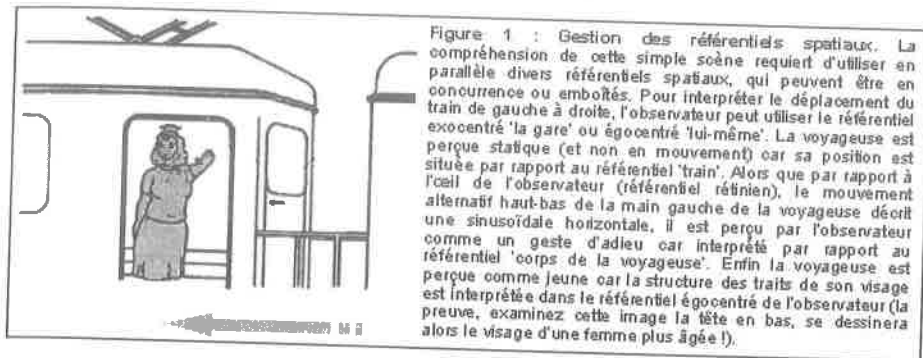
La *fiabilité* renvoie au fait qu'un individu donné perçoit un stimulus comme étant le même au temps t et $t+1$ (*fiabilité* intra individuelle) alors que le stimulus

peut être plus ou moins déformé et le contexte avoir quelque peu changé, ou que deux individus confrontés au même stimulus perçoivent globalement la même chose (*fiabilité* inter individuelle). Cette propriété est une condition *sine qua non* de la perception, quelle que soit la modalité sensorielle impliquée. Elle traduit une invariance dans les mécanismes de traitement et leur relative solidité face aux variations et des informations perceptives et de l'état interne de l'observateur (propriété de résistance au bruit). Les *illusions* perceptives sont un révélateur spectaculaire de ces mécanismes et de leur invariance. Elles font l'objet d'un développement particulier dans la section D.

d. La gestion des référentiels spatiaux

Pour maintenir notre équilibre, localiser et reconnaître les objets qui nous environnent et interagir avec eux, nous avons besoin de savoir où sont le haut, le bas, la droite, la gauche, le devant, le derrière. Autrement dit, nous utilisons tout un ensemble de repères spatiaux qui ne sont pas vrais dans l'« absolu » mais toujours « relatifs à un point de vue », c'est-à-dire à un système de référence ou *référentiel spatial*. En fait, à l'instant t , le sujet dispose de multiples référentiels spatiaux, qui peuvent être en concurrence, généraux ou locaux, et relever de différentes modalités sensorielles. Illustrons cette problématique en examinant la « Scène d'adieu » schématisée figure 1. Pour un observateur confronté à cette situation, une telle scène ne souffre d'aucune ambiguïté perceptive. Et pourtant, comme l'explique la légende de la figure, son interprétation requiert d'utiliser au même moment différents référentiels spatiaux, susceptibles de mettre en jeu des modalités sensorielles différentes. Comprendre comment cette prise et gestion des référentiels s'opérationnalisent reste une gageure scientifique. Les expériences de privation sensorielle sur l'animal (Held et Hein, 1963) ou d'adaptation de personnes à des lunettes prismatiques inversant par exemple le haut et le bas ou la droite et la gauche (expériences de Stratton, Erisman ou Kolher, pour une synthèse, cf. Gregory, 2000) montrent d'une part que l'observateur doit avoir une vision active et d'autre part que l'adaptation n'est jamais complète. Par exemple, en cas d'inversion droite-gauche, après adaptation, certains sujets percevaient les piétons du côté correct de la rue, mais leurs habits — et ce qui était écrit dessus — demeuraient inversés. Le changement de référentiel suite à l'adaptation n'était donc pas général mais local, lié à des contraintes perceptivo-motrices comme localiser correctement les personnes (et objets) de la rue afin de pouvoir y déambuler sans incident.

En résumé, la perception visuelle se marque d'une simplicité et d'une efficacité qui contraste avec la complexité des problèmes spatiaux à résoudre. Par ailleurs, la neurophysiologie et la neuropathologie révèlent que la vision est au départ très analytique, locale et massivement parallèle. Elle débute par une décomposition de la scène en traits et propriétés élémentaires qui sont alors à recomposer en ensembles cohérents. Pour les psychologues, cette conception pose un problème théorique majeur, car au niveau phénoménologique la vision ne garde que peu de traces de ces subdivisions et mécanismes fonctionnels réalisés par le système visuel (voir l'encadré).



Le cortex visuel primaire (cf. part. 2, chap. 9 et 11) décompose l'information rétinienne en primitives visuelles : à ce niveau, chaque cellule code une valeur sur l'une des dimensions physiques que sont l'orientation, la fréquence spatiale, le mouvement, la couleur ou la profondeur binoculaire. Or, lorsque nous percevons notre environnement, nous ne voyons pas ces caractéristiques physiques élémentaires, mais une scène globale structurée. Cela signifie que le cerveau reconstruit un tout global cohérent à partir d'un patchwork local de primitives visuelles. Comment procède-t-il ? Existe-t-il dans les aires supérieures des cellules qui vont coder les éléments de la vie de tous les jours ? Les recherches actuelles ne militent pas en ce sens. Les cellules des aires supérieures codent des caractéristiques ou des propriétés de plus en plus complexes par intégration des codages réalisés par les aires inférieures (contours illusoirs ; constance de la couleur ; mouvement global ; formes complexes, etc.), mais sans pour autant que le traitement visuel ne débouche sur des cellules représentant les formes, objets, etc., rencontrés dans notre environnement naturel, si ce n'est peut-être les visages, qui semblent bénéficier d'un codage plus spécifique (cf. part. 2, chap. 11, p. 365).

Ces aires sont reliées par des connexions latérales, ascendantes et descendantes qui s'organisent en deux voies principales (cf. part. 2, chap. 11, p. 358) :

- *la voie dorsale*, qui a pour objet de transmettre rapidement des informations spatiales utiles soit pour agir (par exemple saisir un objet, focaliser son attention sur un endroit précis de l'environnement), soit pour reconnaître l'agencement des éléments qui composent la scène visuelle : c'est la voie du « Où », ou du « Comment » ;
- *la voie ventrale*, qui a pour fonction de transmettre les informations de détail nécessaires pour différencier et identifier ces éléments : c'est la voie du « Quoi ».

A-t-on des arguments qu'une telle description anatomofonctionnelle, inférée des recherches neurophysiologiques sur l'animal, est valide chez l'homme ?

Zihl, Von Cramon et Mai (1983) ont décrit un cas d'*akinétopsie* (perte de la perception du mouvement) : la patiente voit le mouvement comme une succession d'images statiques, comme sous un éclairage stroboscopique, ce qui l'handicape très fortement pour se déplacer et interagir avec les personnes et les objets. Sacks (1995) décrit le cas d'un peintre qui perd la perception de la couleur : cet expert de la couleur non seulement ne la perçoit plus, mais ne peut plus l'imaginer, se la représenter, ce qui amène à penser que la mémoire de la couleur transite par les aires visuelles qui la traitent.

Goodale, Milner, Jakobson et Carey (1991) ont décrit une atteinte du cortex visuel altérant la voie ventrale. La patiente (DF) montrait une *agnosie visuelle* marquée : absence de reconnaissance des formes et des objets. Elle ne pouvait distinguer avec fiabilité une ligne verticale d'une ligne horizontale (traitement de l'orientation), ni un carré d'un rectangle (traitement de la forme et de la taille). Tout se passe comme si DF ne percevait que les compo-

santes très élémentaires et locales de la scène visuelle, sans recombinaison de ces dernières en formes ou objets globaux. En dépit de ce trouble visuel, DF était capable de saisir correctement les objets dont pourtant l'orientation ou la taille était variable : l'ouverture dynamique de sa pince (distance pouce-index) était similaire à celle observée chez les sujets contrôles. Autrement dit, DF avait perdu la « perception pour l'identification » mais conservé la « perception pour l'action » (due à la relative préservation de la voie dorsale ?).

Une dissociation inverse a été décrite par Perenin et Vighetto (1988) chez des patients souffrant d'*ataxie optique* suite à une atteinte pariétale (voie dorsale). Ces patients étaient capables de juger visuellement l'orientation d'une fente (rainure), mais incapables d'utiliser cette information pour guider leur main vers et dans la fente.

Dans la section suivante, nous exposerons les deux modèles principaux dont dispose actuellement la psychologie pour rendre compte de l'acte de reconnaissance visuelle. L'un, d'inspiration neurophysiologique, est fondé sur l'extraction des *fréquences spatiales* ; l'autre, issu de l'Intelligence Artificielle, repose sur l'extraction de *primitives visuelles*. Le premier modèle s'adresse davantage à la reconnaissance globale des scènes visuelles, le second davantage aux objets qui la composent.

II - La reconnaissance des scènes et des objets

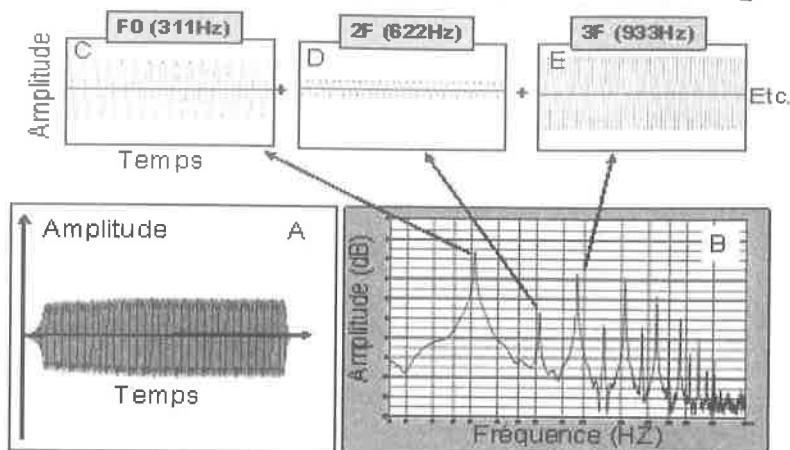
a. Les scènes : perception globale et fréquences spatiales

1 / Les lois de la Gestalt et décomposition de Fourier : le principe CtF

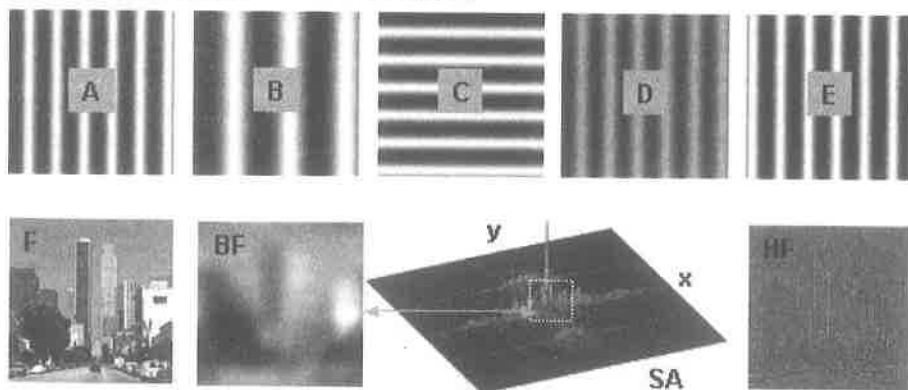
Au début du siècle dernier, les théories sur la perception visuelle étaient dominées par la psychologie de la Gestalt. Le concept de Gestalt renvoie à l'idée qu'une forme globale ou certains groupements d'éléments sont immédiatement perçus par l'observateur sans qu'il soit nécessaire d'analyser les éléments composant la forme perçue (cf. introduction de ce volume). Cette supériorité du tout sur les parties était expliquée en termes de lois d'organisation perceptive ou « lois de la Gestalt », à savoir : la proximité, la similitude, la continuité, la clôture, le destin commun et la « bonne forme » (cf. fig. 9 et 14 ; voir aussi part. 1, chap. 4, p. 179, à propos de la position de ce courant dans le domaine de la résolution de problème). Pour Wertheimer (1923), considéré comme le père fondateur de la théorie, ces lois étaient innées, dans le sens où elles ne nécessitaient aucun recours à une expérience antérieure. Néanmoins, l'auteur et ses contemporains n'ont jamais pu expliquer clairement comment de telles lois pouvaient être implémentées dans le cerveau.

Les avancées actuelles de la neuroanatomie fonctionnelle du système visuel permettent maintenant d'apporter des éléments de réponse. Nous insisterons ici sur l'idée que le système perceptif fonctionne comme une sorte d'analyseur de *Fourier* : il analyse le champ visuel en le décomposant en fréquences spatio-temporelles. L'encadré « Transformée de *Fourier* des sons et des images » illustre une telle décomposition dans le cadre des *fréquences spatiales* et donne des exemples d'images filtrées en basses (BF) ou

Transformée de Fourier des sons et des images



Fréquences temporelles et spectre d'amplitude d'un son. Exemple de décomposition d'un son complexe périodique dont la fréquence la plus basse (fréquence fondamentale) est de 311 Hz (A : représentation temporelle du signal acoustique) en plusieurs sinusoïdes élémentaires (C à E) de fréquences et d'amplitudes différentes. L'onde sonore (A) peut alors être représentée par son spectre d'amplitude (B) avec la fréquence (axe x) et l'amplitude (axe y) de chaque harmonique. Le spectre d'amplitude montre que les fréquences des différentes harmoniques sont des multiples entiers de la fréquence fondamentale (C). Le rang d'une harmonique est déterminé par le facteur par lequel est multipliée la fréquence fondamentale pour obtenir la fréquence de l'harmonique considérée ($3F=3^{\circ}1F$). Les fréquences respectives des différentes harmoniques sont donc 622 Hz pour l'harmonique de rang 2 (D), 933 Hz pour l'harmonique de rang 3 (E), etc. Pour les sons complexes non périodiques, le principe est identique mais le nombre de composantes est beaucoup plus important et leurs fréquences respectives ne correspondent plus à des multiples entiers de la fréquence fondamentale. Dans ce cas on parlera de partiels et non plus d'harmoniques.



Fréquences spatiales et spectre d'amplitude d'une image. Par rapport au réseau central A, le réseau B a une fréquence spatiale plus basse, le réseau C une orientation orthogonale, le réseau D une amplitude moindre, et le réseau E une phase différente. La figure SA est une représentation spatiale du spectre d'amplitude (densité spectrale) de la scène de ville (F). Les axes représentent les dimensions « fréquences spatiales » (au centre la fréquence est nulle). Un pic donné représente la densité d'énergie de toute l'image à une fréquence et une orientation données (plus il est élevé, plus l'énergie est grande). Le spectre d'amplitude indique que, quelle que soit la bande de fréquences spatiales, les orientations verticales (surtout) et horizontales dominent dans la scène de ville et qu'il y a beaucoup d'énergie dans les basses fréquences. La figure BF est un filtrage en basse-fréquence de la ville F : elle a été reconstruite (transformée de Fourier inverse) en ne prenant que les basses fréquences représentées à l'intérieur du rectangle pointillé blanc sur la figure SA. Inversement pour la figure HF (Nota : dans la figure SA, les axes x et y de l'image F sont inversés, les échelles sont logarithmiques, et le SA a une symétrie radiale.)

en hautes (HF) *fréquences spatiales*. Sur le plan neurophysiologique, cette décomposition pourrait être réalisée au niveau du cortex visuel primaire (cellules simples et complexes), et l'information BF, grossière, serait véhiculée rapidement par le système magnocellulaire, et celle HF, de détail, par la voie parvocellulaire plus lente (part. 2, chap. 9 et 11).

Se dessine alors le principe de traitement suivant : les *fréquences spatiales* seraient extraites en parallèle, mais avec une précedence temporelle des BF sur les HF. Nous parlerons du *principe ClF* (abréviation de l'expression anglaise « Coarse-to-Fine », qui littéralement signifie « du Grossier au plus Fin »). Appliquée à la théorie de la Gestalt, « la supériorité et précedence du tout sur les parties » pourrait alors s'expliquer sur le plan neurophysiologique par un tel principe. Voyons ce qu'il en est.

2 / La précedence globale, reconnaissance de scènes et des expressions faciales

Dans les années 1970-1980, la problématique de la supériorité du tout sur les parties a fréquemment été étudiée en utilisant le paradigme des figures hiérarchiques (Navon, 1977). Le stimulus est une forme globale déterminée par l'arrangement spatial de formes locales identiques, qui peuvent être semblables à la forme globale (congruence) ou non (non-congruence) (fig. 2). La consigne est d'identifier le plus rapidement possible la forme au niveau global ou local. Les résultats montrent que les temps d'identification de la forme globale sont rapides et indépendants du degré de congruence de la forme locale. Par contre, pour les formes locales, les temps sont plus longs et varient en fonction du degré de congruence : ils sont plus rapides dans les situations de congruence que dans celles de non-congruence.

Ce pattern de résultats est conforme à la théorie gestaltiste de la précedence du tout sur les parties : la forme globale est traitée plus rapidement que les parties (formes locales) et, en conséquence, interfère sur le traitement de ces dernières, en le facilitant si



Figure 2 : Formes hiérarchiques et images hybrides. **Bloc de gauche** : La lettre globale 'C' composée de lettres locales 'o' (stimulus congruent ; colonne de gauche) est filtrée passe-haut au centre et passe-bas en bas. Idem pour lettre globale 'C' composée de lettres locales 'o' (stimulus non congruent). Les deux types de stimuli apparaissent identiques en basses fréquences. Le filtrage passe-bas a gommé les détails nécessaires à la perception de la structure des formes locales mais a conservé la structure de la forme globale (Nota: les lettres globales ont été volontairement déformées dans cet exemple pour les distinguer en basses fréquences). **Bloc de droite** : exemples de figures hybrides: en haut, scène de plage HF superposée à une scène d'autoroute BF; en bas, visage de femme 'neutre' superposée à un visage d'homme 'en colère'. En regardant les images de loin, seules celles en basse fréquence sont visibles (Nota: pour des raisons d'impression, les contrastes des images HF ont été augmentés).

les formes locales et globales sont similaires (congruence) et en le ralentissant si elles sont différentes (non congruence).

Peut-on expliquer ce phénomène de *précédence globale* dans le cadre du *principe CtF*? Le phénomène de *précédence globale* disparaît lorsque les stimuli sont filtrés en hautes *fréquences spatiales* (fig. 2), c'est-à-dire lorsque les basses fréquences sont supprimées (Badcock *et al.*, 1990). En HF, le temps d'identification des formes globales augmente à la *hauteur* de celui des formes locales et il n'y a plus d'interférence du global sur le local. Ces résultats suggèrent que la *précédence* de la forme globale observée avec des stimuli hiérarchiques non filtrés est bien liée à l'information déduite de l'analyse des basses *fréquences spatiales* et à leur extraction plus rapide par le système perceptif. Notons que lorsqu'on simule cette information basse fréquence en réalisant un *filtrage* passe-bas (suppression des hautes fréquences) sur les stimuli hiérarchiques, la forme globale apparaît encore plus clairement alors que l'identité de la forme locale est perdue (fig. 2).

Le modèle CtF apparaît donc *heuristique*. Plus récemment, un certain nombre d'auteurs ont essayé d'éprouver sa validité dans le cadre de stimuli plus écologiques que les figures hiérarchiques : les scènes naturelles. Une scène naturelle (ville, montagne, plage, forêt, etc.) diffère d'une autre par sa structure spatiale. Cette structure ne peut pas être donnée par les composantes locales de la scène (par exemple pour une ville, le nombre d'immeubles ou de rues, leur emplacement particulier, etc.), qui varient d'un exemplaire de la catégorie à l'autre, mais par ses composantes globales. Par exemple, en termes d'orientation, une image de ville moderne contient beaucoup d'éléments verticaux, une image de plage une frontière mer/plage souvent horizontale, une image de montagne des éléments obliques, etc. Cette organisation est-elle perceptible en BF ?

Une manière élégante de le tester est de présenter au sujet une *image hybride*, c'est-à-dire composée de la superposition de deux scènes, l'une en BF, l'autre en HF (fig. 2), et de voir, en manipulant le temps de présentation de l'image, quelle scène est perçue. Schyns et Oliva (1994) observent que pour un temps de présentation rapide (30 ms), la catégorisation de l'*image hybride* se fait principalement sur la scène BF (63 % contre 28 % sur la scène HF), et inversement pour un temps de présentation plus long (150 ms : 86 % HF vs 18 % BF). Autrement dit, conformément au discours temporel CtF, lorsque le *temps de traitement* est contraint (ici par le temps de présentation de 30 ms), l'information la plus disponible est celle BF, mais lorsque le *temps de traitement* est moins contraint (nous avons vu au début de ce chapitre qu'une centaine de millisecondes étaient suffisantes pour reconnaître une image), les HF, qui fournissent une information plus précise, sont privilégiées par le système de traitement.

Par la suite, le modèle CtF a été appliqué à d'autres stimuli comme les expressions faciales. Nous réagissons par exemple très rapidement et très fortement à une expression de peur. En utilisant l'imagerie par résonance magnétique fonctionnelle, Vuilleumier *et al.* (2003) montrent que les expressions faciales de peur (par rapport à des visages neutres) activent directement l'amygdale par la voie sous-corticale magnocellulaire (cf. part. 2, chap. 11 et 12), qui est rapide et BF, ce qui expliquerait la vitesse de notre réaction émotionnelle et corrobore dans un autre registre fonctionnel le modèle CtF. Mais attention, il faut entendre le discours temporel CtF comme un « principe de traitement par défaut ». Des contraintes attentionnelles peuvent conduire le sujet à focaliser le traitement sur une gamme de fréquences donnée.

b. Les objets : la théorie RBC des « geons »

Lorsque nous examinons une scène, notre regard se porte sur les objets qui la composent. Comment, au niveau perceptif, s'opère leur reconnaissance ? En général, les objets se distinguent les uns des autres par une organisation structurale assez précise des traits et surfaces qui les composent. Partant des travaux de Marr (1982), qui mixent intelligence artificielle et neurosciences, et dont l'écho sera très grand dans les sciences de la vision des années 1980, Biederman (1987) propose la théorie RBC (« Recognition-By-Components ») de reconnaissance d'objets par ses composantes. Pour le moment, elle demeure la plus élaborée en psychologie cognitive.

Cette théorie postule qu'au niveau catégoriel, tout objet est représenté en mémoire sous une forme schématisée d'un arrangement spatial de primitives visuelles tridimensionnelles (3D) que l'auteur dénomme « geon » (contraction de « geometrical ion »). Les *geons* sont directement dérivés des traits et propriétés bidimensionnelles de l'objet, c'est-à-dire des segments de ligne composant les contours internes/externes de l'objet, et des propriétés (gestaltistes) liées à ces segments de ligne comme la collinéarité, la curvilinearité, la symétrie, le parallélisme et la co-terminaison (fig. 3). D'un point de vue mathématique, un geon est un cône. La théorie RBC définit un alphabet de 24 geons par croisement de quatre dimensions spatiales liées à l'axe du cône et sa section. Tout objet connu et à venir est censé pouvoir être représenté à partir de cet alphabet 3D.



Biederman et son équipe ont développé de nombreuses expériences comportementales pour valider le modèle RBC. Retenons-en simplement deux :

- Les auteurs vont tout d'abord montrer que la reconnaissance est aussi performante (exacte, rapide) que les objets soient présentés sous leur aspect naturel (photographie de haute qualité) ou épuré (dessin schématisé aux traits ; cf. pour exemple, le *lapin* figure 3). Autrement dit, pour être efficace, le système de reconnaissance visuel ne semble pas avoir besoin des informations de détails, de couleur, de texture, etc., mais simplement des contours schématisant la structure des composantes volumétriques de l'objet.

- Par contre, si ces composantes ne sont pas perceptibles parce qu'il manque par exemple les jonctions des contours, qui sont de puissants indices structuraux, la reconnaissance échoue ou est très difficile (fig. 3).

Les travaux ultérieurs de l'équipe de Biederman auront pour objet de montrer que ce modèle théorique possède les propriétés de rapidité, de résistance au bruit et de relative invariance spatiale observées au niveau phénoménologique. Le modèle sera également complété par des mécanismes supplémentaires pour qu'il puisse fonctionner non seulement au niveau des catégories, mais aussi à celui des exemplaires (pouvoir par exemple dire que la scène perçue est une ville, et plus précisément Grenoble).

Si la théorie *RBC* constitue pour l'instant la proposition la plus achevée en psychologie cognitive de la reconnaissance visuelle des objets, elle n'est pas sans défaut. Elle s'adapte bien à la reconnaissance des objets manufacturés qui, par construction, sont très géométriques, mais moins aux objets plus naturels pour lesquels l'extraction des volumes requiert très souvent, comme l'a montré la psychologie expérimentale, de prendre en compte les variations d'ombre, de texture, de mouvement... qui sont justement les indices non traités dans le modèle *RBC*.

B - LA PERCEPTION AUDITIVE

Étudier la perception auditive revient à s'interroger sur les relations qui existent entre notre environnement sonore et sa perception. C'est le domaine d'étude de la *psycho-acoustique*, qui étudie donc les conséquences perceptives des variations des paramètres physiques du son que sont l'amplitude, la fréquence et le contenu spectral ainsi que leurs interactions. Par ailleurs, une des spécificités de l'audition est que les signaux acoustiques qui proviennent du même objet et d'objets différents parviennent simultanément aux deux oreilles. Un des problèmes qui se pose alors est de comprendre comment le système auditif arrive à organiser son environnement sonore. Les études sur l'audition utilisent souvent des sons parfaitement contrôlés mais qui sont rarement rencontrés dans la vie de tous les jours. Les sons de notre environnement sont des sons complexes qui donnent lieu de façon instantanée à des « images auditives ». La voix d'un de nos proches est par exemple instantanément identifiée au téléphone. La question est alors de comprendre quels sont les processus en jeu dans la reconnaissance des sons de notre environnement sonore.

I - Le champ audible humain

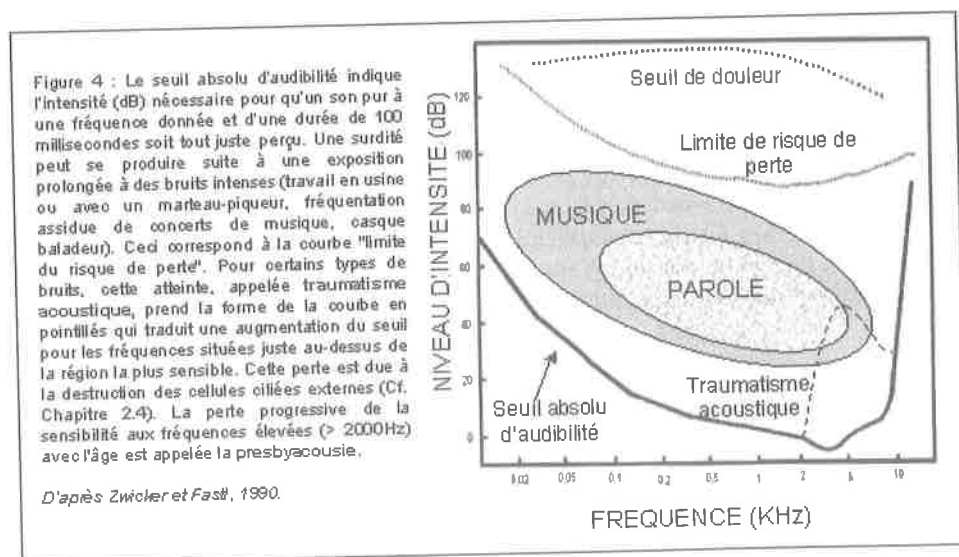
Un son est défini selon quatre paramètres :

- l'*intensité*, correspondant à la valeur de la pression acoustique (décibel (dB)). Lorsque vous augmentez ou diminuez le volume de votre téléviseur, vous modifiez l'amplitude du son ;

- la *fréquence*, correspondant au nombre de cycles d'oscillations effectuées pendant une seconde (Hertz (Hz)). Plus le nombre de cycles par seconde est élevé, plus la *fréquence* est élevée ;
- la *durée*, exprimée en secondes par exemple ;
- la composition du **spectre**, qui indique le niveau acoustique de chaque composante du son en fonction de la fréquence (voir plus haut l'encadré Transformée de Fourier des sons et des images).

La gamme de fréquences audibles par l'être humain va de 20 Hz à 16 000 Hz ; au-delà, on atteint les ultrasons dont les fréquences sont trop élevées pour être audibles et en deçà on se trouve dans le domaine des infrasons dont les fréquences sont trop basses pour être audibles.

Ces quatre paramètres donnent lieu à des sensations auditives de sonie, de hauteur, de durée et de timbre. Un des problèmes majeurs qui se pose lorsque l'on cherche à établir les seuils d'audibilité est celui de l'interaction des dimensions perceptives sur la sensation auditive résultante. En effet, si l'on peut faire varier indépendamment la fréquence et l'intensité d'un son, perceptivement ces deux dimensions physiques donneront lieu à des changements de hauteur et de sonie, mais qui interagiront l'une sur l'autre. Ainsi, il est impossible de savoir à quelle intensité physique il faut ajuster un son pour qu'il soit tout juste audible si l'on ne connaît pas sa fréquence ! Ainsi, si l'on rend plus aigu un son en augmentant sa fréquence on le rendra également perceptivement plus fort pour l'auditeur, alors même que son intensité physique n'aura pas été changée. Pour comprendre la gamme des sons audibles par l'être humain il faut donc tenir compte à la fois de la fréquence et de l'intensité physique des sons (fig. 4).



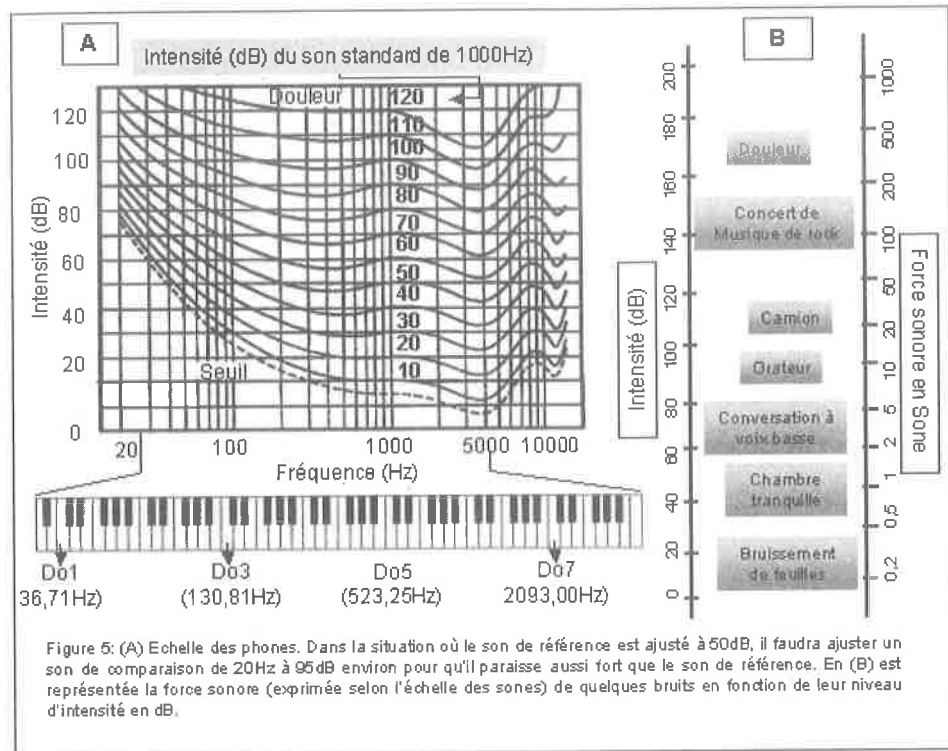
II - Les dimensions perceptives du son

Le premier but des psycho-acousticiens a été d'établir des échelles psychophysiques (voir plus haut à propos de la mesure des seuils). Pour établir de telles échelles dans le domaine sonore il faut distinguer les sons purs des sons complexes. Un son pur correspond à la variation sinusoïdale de la pression acoustique correspondant à l'alternance compression/raréfaction des molécules en mouvement. Les sons de notre environnement sont rarement des sons purs. Nous sommes généralement en présence de sons complexes. Parmi les sons complexes, on distingue les sons complexes périodiques et non périodiques. Les premiers présentent une variation complexe (non sinusoïdale) d'amplitude qui se reproduit, identique à elle-même à des intervalles de temps régulier (par exemple instruments de musique et parole (voyelles)). Pour les seconds la variation d'amplitude ne présente aucune périodicité (bruits de l'environnement). Nous avons vu que le système visuel analyse le champ visuel en le décomposant en une somme de fréquences spatio-temporelles. Il en est de même pour le système auditif. L'encadré « Transformée de Fourier » (voir plus haut) illustre la décomposition d'un son complexe périodique composé de plusieurs sons purs (sinusoïdes) de fréquences différentes appelées harmoniques. Dans le cas des sons complexes non périodique, on parlera de partiels.

a - La perception de l'intensité

La Sonie est la dimension perceptive principalement liée à l'intensité physique du son. L'intérêt d'établir une échelle psychophysique est qu'elle nous indique dans quelle proportion la sonie d'un son (sensation d'intensité sonore) est augmentée lorsque le niveau de ce même son augmente d'un nombre de décibels donné (cf. Botte, 1989, pour une revue).

- Dans le cas des sons purs, l'échelle des phones (fig. 5) correspond à la perception de la force tout en tenant compte de l'interaction entre la fréquence et l'intensité du son. Les courbes d'isophonies décrivent les niveaux sonores pour lesquels les sons de différentes fréquences prennent pour l'auditeur une force identique à celle du son de référence. Pour chaque courbe, le son de référence a la même fréquence de 1 000 Hz, mais il varie en intensité. Le niveau en décibels du son de 1 000 Hz détermine le nombre de phones pour chaque courbe.
- Dans le cas des sons complexes, on utilisera l'échelle des sones, selon laquelle la force sonore d'un son est déterminée en comparaison à un son standard de 1 000 Hz. La valeur en sone pour laquelle un son de 1 000 Hz semble aussi fort qu'un son complexe constitue le niveau en sone de celui-ci. La figure 5 présente la valeur en sone de quelques sons complexes.



b. La perception de la hauteur

La hauteur est, de même que la sonie, un attribut de la sensation auditive selon lequel les sensations s'ordonnent suivant une échelle allant du grave à l'aigu. La hauteur dépend presque exclusivement d'une seule dimension physique du stimulus : la fréquence (cf. Demany, 1989, pour une revue).

La hauteur tonale d'un son pur est, selon l'échelle musicale (ou chroma), fonction du logarithme de la fréquence sonore. Chaque octave dans l'échelle musicale est exactement le double de la fréquence de l'octave précédente. Ainsi, le Do3 ayant une fréquence de 130 Hz, les Do4 et Do5 auront alors respectivement des fréquences de 260 et 520 Hz (fig. 5). Néanmoins, cette échelle ne rend pas compte de la réalité perceptive de la plupart des auditeurs. En effet, doubler ou diminuer de moitié la fréquence d'une note ne double ni ne diminue de moitié la hauteur perçue. L'échelle des mels, construite pour rendre compte de la perception de la hauteur des sons purs, est en fait une fonction monotone de la fréquence.

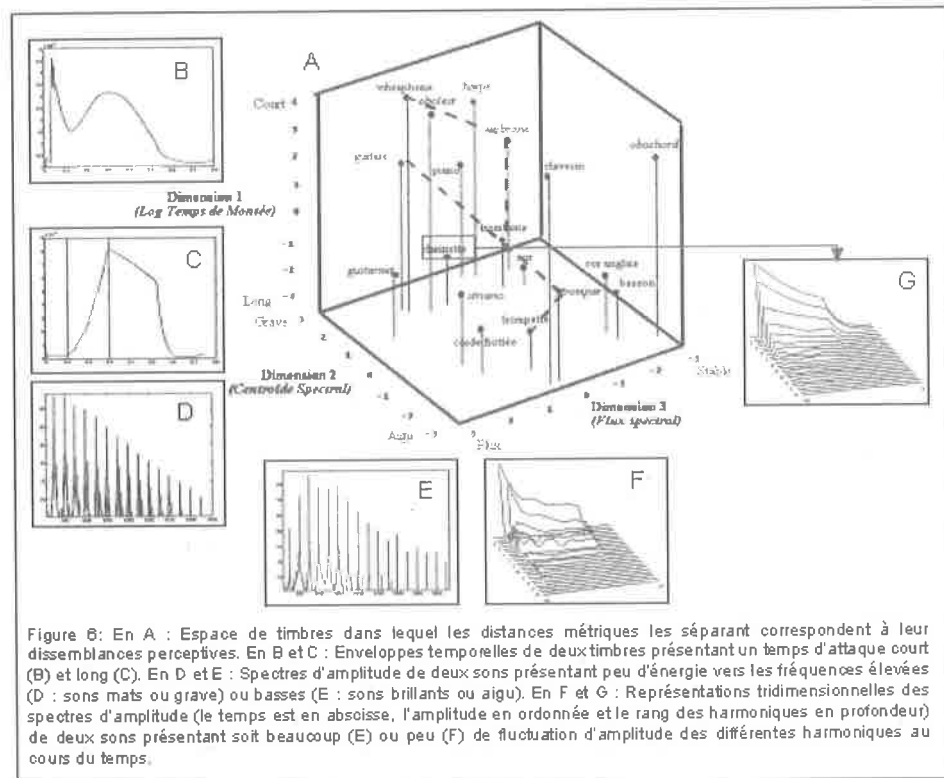
À l'écoute d'un son complexe périodique (par exemple une note de guitare), une hauteur unique est spontanément perçue égale à la composante la plus basse du son complexe, c'est-à-dire la fréquence fondamentale. Dans le cas d'un son complexe non

périodique (ou inharmonique), les partiels ne correspondent pas à des multiples entiers de la fréquence fondamentale ; la hauteur perçue est alors souvent moins précise et l'auditeur peut parfois percevoir plusieurs hauteurs en même temps.

c. La perception du timbre

Contrairement à la sonie et à la hauteur, le timbre n'est pas clairement défini. Selon l'*American Standards Association* (1960), « le timbre est l'attribut de la sensation auditive suivant lequel un auditeur peut différencier deux sons présentés dans les mêmes conditions et ayant la même sonie et la même hauteur ». Cette définition nous indique ce que n'est pas le timbre – c'est-à-dire ni la hauteur, ni la sonie, ni la durée –, mais il reste tout à dire sur ce qu'il est. Ce qui rend difficile toute définition du timbre est sans doute son aspect multidimensionnel. En effet, contrairement à la hauteur et à la sonie, le timbre ne peut se mesurer sur un seul continuum comme faible-intense (sonie) ou grave-aigu (hauteur).

La question a donc été de déterminer le nombre de dimensions perceptives du timbre et les corrélats physiques qui les sous-tendent. Pour étudier cette question, des sons d'instruments de musique sont présentés par paires, et les auditeurs doivent juger



le degré de dissemblance perceptive entre les sons (par exemple de 1 (similaire) à 9 (dissemblable)). Les sons sont égalisés en hauteur, sonie et durée de façon à ce que les auditeurs basent leurs jugements uniquement sur des différences de timbres. Les jugements sont ensuite soumis à une analyse multidimensionnelle dont l'objectif est de trouver l'espace géométrique dans lequel les distances métriques séparant les timbres représentés par des points reflètent le mieux possible leur dissemblance perceptive. Ainsi, deux timbres jugés dissemblables seront éloignés, et deux timbres jugés similaires seront proches dans l'espace. McAdams *et al.* (1995) ont observé que trois dimensions perceptives peuvent rendre compte de la perception d'un ensemble de 18 timbres d'instruments de musique. Des analyses acoustiques ont permis d'interpréter la nature des dimensions observées (fig. 6) :

- La première dimension, liée à la qualité d'attaque des sons, est bien corrélée au temps mis par l'enveloppe temporelle des sons pour atteindre un maximum d'énergie. Cette dimension oppose les instruments à vent (temps de montée lents) aux instruments à cordes frappées ou pincées (temps de montée rapide).
- La seconde dimension correspond à la « brillance » du son opposant les sons brillants ou « nasillards » aux sons « mats » ou « sourds ». Le paramètre acoustique corrélé à cette dimension est le centre de gravité du spectre qui mesure le nombre d'harmoniques présents et leurs amplitudes relatives à la fréquence fondamentale.
- La troisième dimension correspond à une mesure du degré de fluctuation du contenu spectral des sons au cours du temps, d'où l'expression flux spectral.

III - L'organisation perceptive en audition

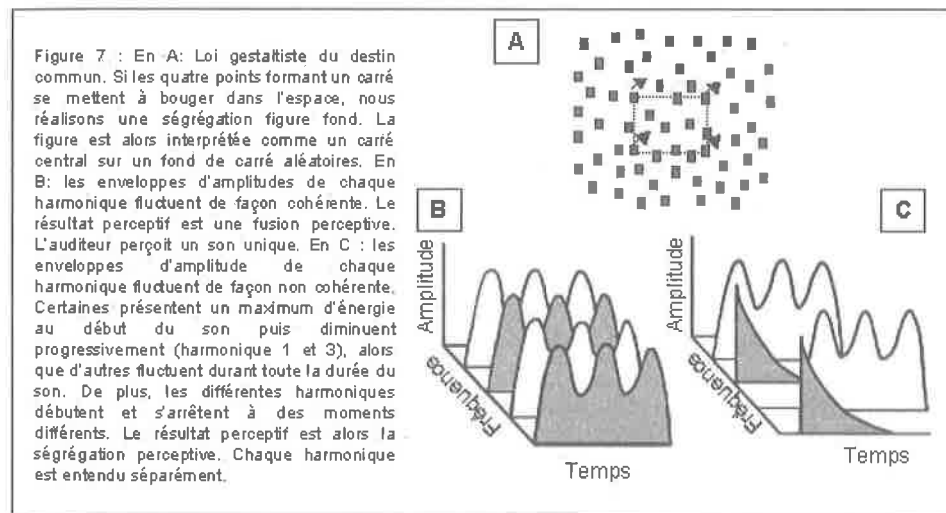
Il s'agit ici de comprendre comment le système auditif analyse les scènes auditives complexes de notre environnement sonore quotidien. En effet, les situations quotidiennes sont des situations « auditives » complexes où l'on peut par exemple entendre en même temps la sonnerie d'un portable, les rires d'enfants, tout en suivant le discours de son interlocuteur. Autrement dit, à un même moment arrive à nos oreilles un ensemble de bruits qui changent continuellement et qui résultent du mélange des sources sonores émises simultanément. Pourtant, nous identifions sans effort ces différentes sources sonores, même si tous ces sons sont présents simultanément et évoluent ensemble au cours du temps. Cela suggère que nous sommes capables d'organiser simultanément les ondes sonores qui proviennent à nos oreilles en regroupant les informations qui proviennent d'une même source (fusion perceptive) et en séparant celles qui proviennent d'une autre source (ségrégation perceptive). Parallèlement, nous pouvons lier perceptivement à travers le temps les événements successifs émis par la même source tout en les séparant d'événements provenant de sources indépendantes afin de suivre le « message ». Cela correspond au problème de l'organisation séquentielle.

Les études sur l'organisation simultanée et séquentielle ont montré que certaines lois gestaltistes (voir ci-dessus, p. 54) peuvent expliquer une part du fonctionnement du système auditif pour organiser les scènes auditives complexes. La loi du destin commun – selon laquelle les éléments évoluant en même temps au cours du temps vont

s'organiser en une même forme – permet par exemple d'expliquer les phénomènes de ségrégation figure/fond en audition. De même, les lois de similarité ou de proximité – selon lesquelles les éléments similaires ou proches dans l'espace vont s'organiser en une même forme – permettent d'interpréter comment nous arrivons par exemple à suivre la mélodie chantée par le chanteur tout en suivant l'accompagnement joué au piano.

a. L'organisation simultanée

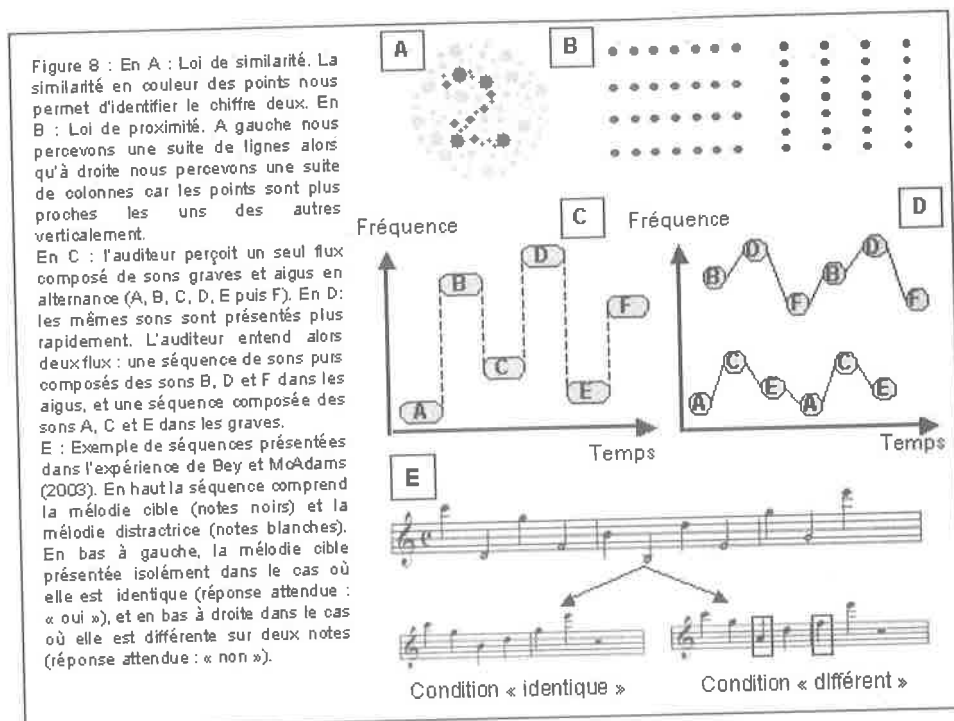
Les sons provenant de sources distinctes commencent et s'arrêtent rarement en même temps. Ainsi, les événements auditifs qui évoluent de la même manière au cours du temps seront alors considérés par le système auditif comme provenant d'une même source. En général, un changement brutal de l'un des paramètres du son (intensité, fréquence) sera interprété comme l'arrivée d'un autre événement auditif provenant d'une source sonore différente (Rasch, 1978). De même, si l'on synthétise un son dont les enveloppes d'amplitude de chaque partiel fluctuent de manière cohérente, c'est-à-dire exactement de la même façon, on entend un son fusionné. Cependant, lorsque ces fluctuations ne sont pas en phase ou sont complètement différentes, on entend les partiels séparés. La figure 7 illustre le résultat perceptif de la combinaison de ces deux paramètres que sont le synchronisme des attaques et l'évolution cohérente des changements d'amplitude sur l'organisation simultanée de sons complexes.



b - L'organisation séquentielle

Un flux sonore est perçu lorsque les sons qui le composent forment un tout présentant une continuité et qui est donc interprété comme provenant d'une même source sonore (Bregman, 1990). Le système auditif va interpréter deux événements

sonores évoluant au cours du temps comme appartenant à deux sources sonores distinctes si les sons appartiennent à deux régions fréquentielles éloignées ; alors qu'il ne percevra qu'une source sonore si les sons sont proches du point de vue de leur fréquence. Le tempo, c'est-à-dire la cadence à laquelle les sons vont être présentés à l'auditeur, peut également influencer l'organisation auditive séquentielle. Ainsi, lorsque six sons purs situés dans deux régions de fréquences sont joués lentement, on entendra un seul flux auditif de six notes. Si le tempo est suffisamment augmenté, une fission perceptive se produit et deux flux de trois sons purs sont perçus (Bregman et Campbell, 1971 ; voir fig. 8).



Bey et McAdams (2003) ont confirmé le rôle de la hauteur dans la formation de flux sonore pour des sons complexes d'instrument de musique. Les sujets devaient dire si oui ou non une mélodie cible était identique à une mélodie de référence présentée entremêlée avec une mélodie distractive (fig. 8). La mélodie cible était parfois identique parfois différente (sur deux notes) de la mélodie de référence. L'écart entre la hauteur moyenne des 6 notes de la mélodie de référence et de la mélodie distractive variait de 0 à 25 demi-tons. La performance des auditeurs augmente avec le degré de séparation en hauteur, ce qui suggère que, plus les 2 mélodies sont éloignées du point de vue de la hauteur, plus l'auditeur réussit à séparer les deux mélodies perceptivement.

IV - La reconnaissance des objets sonores

Que se passe-t-il entre le moment où un son est produit et son identification ? Contrairement à la vision, peu de modèles de la reconnaissance des objets sonores ont été proposés. Celui proposé par McAdams (1994) permet cependant de poser des hypothèses sur les processus en jeu lors de la reconnaissance et/ou identification d'un objet sonore.

a. Les étapes en jeu dans le processus de reconnaissance auditive

- La *première étape* correspond à la transduction sensorielle du signal acoustique au niveau du système auditif périphérique pour ensuite arriver au niveau du système auditif central (voir partie 2, chap. 9).
- Lors de la *seconde étape*, des processus de groupement de l'information acoustique (cf. ici § III) interviennent afin de relier les informations provenant du même objet sonore et de séparer les informations provenant de deux sources sonores différentes.
- Lors de la *troisième étape*, des processus de calcul des attributs perceptifs (calcul de la hauteur, intensité, durée, timbre) permettraient la construction d'une représentation perceptive de l'objet sonore. Le timbre serait un attribut essentiel dans le processus d'identification auditive puisqu'il véhiculerait sans doute l'identité des sources sonores, comme en témoigne l'expérience de Risset et Matthieu (voir encadré).

L'expérience de Risset et Matthieu (1969) montre que l'analyse de l'enveloppe temporelle de l'onde sonore est un paramètre acoustique essentiel lors de l'identification d'un son. Le principe consiste à enregistrer une pièce musicale jouée au piano. Chaque note de piano est ensuite inversée de façon à ce que les sons (correspondants aux notes) soient entendus de la fin du signal acoustique au début tout en respectant l'ordre de présentation des notes composant la mélodie (la mélodie reste identique). À l'écoute de l'extrait, les auditeurs peuvent parfaitement identifier la mélodie, mais la plupart indiqueront que cette mélodie est jouée par un accordéon, voire un orgue, c'est-à-dire un instrument soufflé, alors que l'instrument à partir duquel la mélodie a été exécutée reste le piano. Le fait d'avoir simplement inversé l'enveloppe temporelle de chaque son change l'identité de la source sonore. Cet exemple suggère que la dimension perceptive du timbre liée à la qualité d'attaque (voir p. 63) joue un rôle essentiel dans le processus de reconnaissance et d'identification auditive des sources sonores complexes. Si l'on modifie certains aspects de cette dimension on touche à l'identité même du son, ce qui appuie l'hypothèse selon laquelle le timbre véhiculerait bien l'identité des sources sonores complexes.

- La *dernière étape* consiste à appairer la représentation perceptive ainsi construite à celles stockées en mémoire à long terme (lexique de sons). Si l'appariement se fait, l'événement sonore est alors reconnu, voire identifié.

b. L'agnosie auditive : une incapacité à reconnaître les objets sonores

Le terme d'agnosie auditive désigne un trouble de la reconnaissance des objets sonores qui apparaît suite à une lésion cérébrale. Elle se différencie de la surdité, puisque le patient reste capable de percevoir des changements de fréquence, d'intensité et de durée dans les événements sonores qui lui sont présentés. Le patient se plaint généralement d'entendre tous les sons comme des bruits sans signification (les sons sont perçus comme des frottements ou des grincements). Il confond les bruits d'un moteur avec l'aboiement d'un chien, et paraît surpris par le caractère cacophonique, voire désagréable, de la musique. Souvent les sons de l'environnement et musicaux sont perturbés (Eustache *et al.*, 1990) sans troubles de la perception des sons verbaux. Parfois, la perception des sons de l'environnement et du langage est perturbée malgré une perception musicale parfaite (Mendez, 2001). Les cas d'agnosies auditives concernant spécifiquement les sons de l'environnement sont rarement décrits. Takayuki *et al.* (2000) ont cependant décrit le cas d'une patiente présentant un trouble spécifique de la reconnaissance des sons de l'environnement en cours d'évolution d'une agnosie auditive globale touchant les trois domaines (verbal, sons de l'environnement et musicaux). Lorsque seul le domaine musical est perturbé on parlera d'*agnosie musicale* (voir l'encadré Perception de la musique et agnosies musicales). Ces études suggèrent que parole, musique et sons de l'environnement sont des catégories bien distinctes, qui pourraient correspondre à trois systèmes de reconnaissance différents, cependant régis de la même manière, mais qui seraient spécialisés suivant la nature des événements sonores.

c. L'agnosie auditive aperceptive versus associative

Comme pour l'agnosie visuelle, on distingue deux types d'agnosie auditive : *aperceptive* versus *associative*. Si le trouble est aperceptif, il correspond alors à une atteinte des étapes 2 et 3 du processus de reconnaissance auditive présentées ci-dessus (p. 67). Dans ce cas, le patient ne peut plus discriminer des patrons acoustiques différents, soit parce que les processus de ségrégation des objets sonores (étape 2) sont déficients, soit parce que la représentation perceptive est lésée (étape 3). Si le trouble est associatif, il correspondra à une atteinte au niveau des représentations mnésiques (étape 4). Le patient ne peut plus associer une signification aux événements sonores malgré une analyse adéquate de leurs caractéristiques acoustiques. Cette dissociation entre agnosie auditive aperceptive versus associative a été mise en évidence par Clarke *et al.* (1996). Plus précisément, les auteurs montrent trois doubles dissociations chez des patients présentant des troubles de la reconnaissance auditive pour des tâches impliquant des sons de l'environnement. Une double dissociation s'observe entre l'identification sémantique (apparier un son à une image) et la reconnaissance asémantique (indiquer si deux extraits sonores différents sont produits ou non par un même objet), entre l'identification et la ségrégation d'objets sonores (repérer un objet dans un environnement bruité en augmentant son intensité jusqu'à sa détection) et enfin entre la reconnaissance sémantique et la ségrégation d'objets sonores. Ces résultats vont dans le sens d'une séparation entre les trois étapes du processus de reconnaissance auditive décrit ci-dessus.

Perception de la musique et agnosies musicales

Peretz (1994) propose que la reconnaissance d'un air musical repose sur l'analyse en parallèle des composantes en jeu dans l'organisation mélodique (intervalles de hauteur entre chaque note, contour et tonalité) et temporelle (rythme et métrique). Un tel modèle du traitement de l'information musicale permet d'expliquer certains troubles spécifiques du traitement de l'information musicale d'origine neurologique, appelés « agnosies musicales ». Ces troubles sont très variés et peuvent concerner soit le pôle expressif soit le pôle perceptif.

Lorsqu'il concerne le pôle expressif, des cas de musiciens présentant une *agraphie musicale* (impossibilité d'écrire la musique) ont été rapportés. Maurice Ravel (1875-1937) se trouva durant les quatre dernières années de sa vie dans l'incapacité de composer et de jouer de la musique alors qu'il avait gardé ses capacités d'écoute et d'interprétation musicales, la mémoire de ses propres œuvres et ses connaissances musicologiques. On parlera d'*amusie motrice* dans le cas de patients qui se trouvent dans l'incapacité de chanter, siffler ou fredonner une mélodie.

La plupart des agnosies musicales décrites dans la littérature concernent le versant perceptif. Ainsi, pour ce qui concerne la perception de la mélodie et du rythme, Peretz (1994) a pu montrer que ces deux composantes pouvaient être sélectivement touchées suite à une lésion cérébrale. Dans le cas où c'est la composante mélodique qui est perturbée, il s'agit d'une *amélie* (impossibilité de discriminer ou d'identifier des mélodies différentes), alors que l'on parlera d'*arythmie* (impossibilité de discriminer des rythmes différents) si c'est la composante temporelle qui est lésée. Di Pietro et coll. (2004) ont récemment observé le cas d'un musicien professionnel (chanteur) qui après un infarctus cérébral affectant le lobe pariéto-temporal gauche souffre d'une arythmie sans amélie. Il est tout à fait capable de discriminer deux mélodies dont le contour ou les intervalles diffèrent, mais il échoue lorsque seul le rythme entre les deux mélodies diffère. Lorsqu'on lui demande de réaliser une dictée musicale, ses productions (Figure ci-dessous) témoignent qu'il reste parfaitement capable de reproduire les intervalles et le contour d'une mélodie, mais le rythme n'est plus présent. Le trouble observé pour le traitement temporel des mélodies n'affecte que la perception du rythme tout en épargnant la métrique, puisqu'il identifie parfaitement cette dernière (il reste ainsi capable de dire si il s'agit d'une valse ou d'une marche militaire, par exemple).

Originale



Production de D.L.

"Au Clair de la Lune"



Originale



Production de D.L.



"Frère Jacques"

En conclusion, nous souhaitons souligner l'aspect de plus en plus dynamique des recherches relatives à la perception auditive. Les questions théoriques sont en pleine révolution aujourd'hui. En effet, comme nous l'avons déjà évoqué en introduction, la modalité auditive a au départ suscité moins d'intérêt que la modalité visuelle. Certes, les recherches sur la perception de la musique que nous n'avons pas abordées ont permis de comprendre une part des mécanismes en jeu dans le traitement d'une information auditive non verbale. Néanmoins, ces études montrent que les sons musicaux sont des stimuli qui mettent en jeu des processus cognitifs spécifiques et par conséquent différents des bruits de l'environnement. Aujourd'hui, l'audition non verbale portant sur des sons de l'environnement suscite de plus en plus d'études, et de nouveaux modèles de reconnaissance des objets sonores complexes devraient donc voir le jour dans un futur proche.

C - LA PERCEPTION HAPTIQUE MANUELLE

Nous pouvons distinguer deux types de perception tactile manuelle : la perception cutanée et la perception haptique. La perception cutanée ou passive résulte de la stimulation d'une partie de la peau, alors que le segment corporel qui la porte est totalement immobile. Tel est le cas lorsque le dos de la main repose sur une table et qu'un objet pointu est déplacé sur sa paume. Dans ce cas, comme seule la couche superficielle de la peau est soumise à des déformations mécaniques, le traitement perceptif ne concerne que les informations cutanées liées au stimulus appliqué sur la main. Comme la perception cutanée est rarement mise en œuvre dans la vie quotidienne, elle sera peu abordée ici.

La perception haptique (tactilo-kinesthésique) résulte de la stimulation de la peau provenant des mouvements actifs d'exploration de la main entrant en contact avec des objets (Gibson, 1962). C'est ce qui se produit quand, par exemple, la main et les doigts suivent le contour d'un objet pour en apprécier la forme. Dans ce cas, il s'ajoute nécessairement à la déformation mécanique de la peau celle des muscles, des articulations et des tendons qui résulte des mouvements d'exploration. Des processus très complexes sont impliqués ici car ils doivent intégrer simultanément les informations cutanées et les informations proprioceptives et motrices liées aux mouvements d'exploration cutanée pour former un ensemble indissociable appelé perceptions haptiques.

I - Les mouvements d'exploration manuelle

Des mouvements d'exploration volontaires, variant en fonction des caractéristiques de ce qu'il faut percevoir, doivent être produits par la personne pour compenser l'exiguïté du champ perceptif cutané (limité à la zone de contact avec les objets) et appréhender les objets dans leur intégralité. La perception du stimulus va donc dépendre de la façon dont il est exploré. Il en résulte une appréhension morcelée, plus ou moins cohérente, parfois partielle et toujours très séquentielle, qui charge lourdement la mémoire de travail (cf. part. 1, chap. 2) et qui nécessite, en fin d'exploration, un travail mental d'intégration et de synthèse pour aboutir à une représentation unifiée de l'objet (Revesz, 1950).

Ces caractéristiques peuvent engendrer des traitements perceptifs haptiques spécifiques. Ainsi par exemple, à cause du caractère séquentiel de l'exploration et de la possibilité de modifier à volonté la taille du champ perceptif tactile, la modalité haptique est moins sensible que la vision aux lois gestaltistes d'organisation de la configuration spatiale (cf. ci-dessus p. 54), car le doigt peut exclure de son champ les éléments induisant les « effets de champ » (c'est-à-dire les interactions entre éléments d'une figure). Cette possibilité de décomposition de la figure en ses éléments explique pourquoi, contrairement à ce qu'on observe chez les adultes, « la loi gestaltiste de proximité » (cf. fig. 8 B) n'a pas, à 5 ans, l'effet de « groupement » qu'on lui connaît dans la modalité visuelle. En effet, avec des mains de petite taille, des doigts très fins, une exploration partielle et peu active, les jeunes enfants perçoivent isolément des points qui, visuellement, sont groupés en une unité structurale difficilement sécable (une suite de lignes ou de colonnes). Au contraire, chez les adultes qui explorent la totalité des points de la figure, la proximité des points a les mêmes effets structuraux que dans la vision (Hatwell, 1995 ; Hatwell, Orliaguet et Brouty, 1990). Ces caractéristiques expliquent aussi pourquoi certaines illusions géométriques visuelles n'apparaissent pas dans la modalité haptique (cf. section D).

Les mouvements exploratoires des adultes ont été classés en « procédures exploratoires » (PE), c'est-à-dire en ensembles spécifiques de mouvements volontaires qui se caractérisent par la quantité d'information qu'ils peuvent apporter et donc par l'éventail des propriétés auxquelles ils sont adaptés (Lederman et Klatzky, 1987). Certaines procédures sont très spécialisées, d'autres plus générales. Ainsi, le « Frottement latéral » est adapté seulement à la texture, le « Soulèvement » au poids, la « Pression » à la dureté du matériau. Le « Contact statique » informe principalement sur la température et, plus approximativement, sur la forme, la taille, la texture et la dureté. L'« Enveloppement » donne aussi des informations globales sur ces propriétés, tandis que le « Suivi des contours » donne une connaissance précise de la forme et de la taille, et une connaissance plus floue de la texture et de la dureté.

Ces différentes procédures sont soit nécessaires (obligatoires pour une propriété), soit suffisantes, et certaines sont optimales, c'est-à-dire ont une efficacité maximale pour une propriété. Ainsi, le frottement latéral est optimal pour la texture, tandis que le soulèvement est nécessaire et optimal pour le poids. Comme ces mouvements sont volontaires, le sujet ne percevra pas ou percevra mal la propriété en question s'ils ne sont pas produits. On voit donc l'implication de facteurs cognitifs dans l'exploration perceptive manuelle, qui est beaucoup plus marquée que dans l'exploration visuelle.

Lederman et Klatzky (1993) ont observé une stratégie d'exploration en deux temps : d'abord sont produites des procédures non spécialisées, mobilisant toute la main et apportant des informations peu précises sur plusieurs propriétés, ce qui donne une connaissance globale de l'ensemble. Puis les procédures spécifiques sont mises en œuvre. Par exemple, pour la forme, les adultes commencent par l'enveloppement, puis passent au suivi des contours.

Nous allons voir maintenant que le caractère spécialisé des procédures d'exploration manuelle a des conséquences sur le traitement des propriétés des objets. Ainsi, la modalité haptique est très performante dans la perception de la texture et la dureté des matériaux, alors qu'elle l'est moins dans celle des propriétés spatiales des

objets. Cette spécialisation s'explique sans doute par la simplicité des procédures exploratoires optimales pour percevoir la texture ou la dureté, alors que celles propres aux propriétés géométriques exigent des mouvements coordonnés dans le temps.

II - La perception des propriétés spatiales : l'exemple de l'orientation spatiale

L'accès aux propriétés de forme, taille, orientation, distance, etc., nécessite des procédures exploratoires lentes et coûteuses en raison de la taille réduite du champ perceptif tactile. C'est pourquoi les performances dans ce domaine sont significativement inférieures à celles de la vision du point de vue quantitatif : seuils de discrimination (voir plus haut l'encadré sur les seuils), nombre d'erreurs, durée du processus. C'est ce qui a été montré dans un grand nombre de travaux ces vingt dernières années (cf. Gentaz, Hatwell et Hennion, 2006). De plus, les processus eux-mêmes ne sont pas toujours les mêmes dans la modalité visuelle et la modalité haptique. La perception des orientations spatiales d'un stimulus dans ces deux modalités illustre parfaitement ces caractéristiques.

a. Des processus haptiques moins efficaces que les processus visuels correspondants

À tous les âges, la perception des orientations est plus précise en vision qu'en haptique, et cette précision s'améliore au cours du développement dans les deux modalités. Dans la perception haptique des orientations, on ne retrouve pas toujours une anisotropie (une perception qui varie selon la valeur de l'orientation) qui est presque systématique dans la perception visuelle. Un « effet de l'oblique » est présent à tous les âges dans les tâches visuelles de comparaison et de reproduction (Appelle, 1972). Il est mis en évidence en faisant la différence entre les performances observées pour les verticale et horizontale d'une part (qui ne se différencient pas le plus souvent et sont regroupées) et les obliques d'autre part (regroupées aussi quand elles ne se différencient pas). Comme il s'agit d'une différence, cet effet de l'oblique peut apparaître quelle que soit par ailleurs la précision globale des réponses. Par exemple, dans une tâche de reproduction, les mêmes sujets font des erreurs moyennes en vision et en haptique respectivement de 1° et de 3,8° pour l'orientation verticale et de 3,6° et 5,7° pour les orientations obliques (Gentaz *et al.*, 2001).

b. Des processus haptiques similaires et originaux

En haptique, les études montrent l'existence, seulement sous certaines conditions, d'un effet de l'oblique intrinsèque à la perception haptique, puisque cet effet est observé chez les aveugles précoces (Gentaz et Hatwell, 1998). Ensuite, ces résultats montrent que la perception haptique des orientations dépend des conditions d'exploration, de maintien et de reproduction du stimulus (en général une baguette). Ainsi, l'effet de l'oblique est

absent lorsque les adultes travaillant sans voir explorent et après cinq secondes reproduisent avec la même main l'orientation d'une baguette dans le plan horizontal en prenant appui avec leur avant-bras/poignet/main sur le plateau qui supporte la baguette (condition « avec appui »). Dans cette condition, les sujets déploient peu de forces antigravitaires pour explorer la baguette, et reçoivent donc peu d'informations sur l'orientation de la verticale gravitaire. En revanche, l'effet de l'oblique est présent lorsque les sujets maintiennent leur avant-bras/poignet/main en l'air (condition « sans appui »). Dans cette condition, des forces antigravitaires sont nécessairement produites car la baguette stimulée se trouve à 8 cm au-dessus du plateau (Gentaz et Hatwell, 1996).

La présence ou l'absence d'un effet de l'oblique haptique suggère que différents types de processus sont impliqués dans chaque cas (Gentaz, 2000). Ainsi, la présence de l'effet témoignerait d'un processus cognitif de haut niveau, et plus précisément d'une tendance à catégoriser (inconsciemment) un espace de travail circulaire en quadrants délimités par les axes vertical et horizontal (Gentaz, 2005). Ces deux axes principaux auraient un statut particulier, et seules les orientations suffisamment proches de l'un de ces axes seraient identifiées de façon efficiente (comme c'est le cas pour les orientations verticale ou horizontale). Par défaut, toutes les autres orientations seraient assimilées à une classe d'orientations obliques moins bien définie, dont le prototype serait la diagonale la plus proche, entraînant une identification moins efficiente. Ce type de processus, relativement stable dans le temps, a l'avantage de ne pas nécessiter beaucoup de ressources attentionnelles (insensibilité aux interférences), mais a l'inconvénient de générer une perception anisotropique.

L'absence d'un effet de l'oblique témoignerait, quant à lui, de la mise en œuvre de processus haptiques originaux fondés sur un codage des orientations par la dynamique même du mouvement d'exploration réalisé et mémorisé. Chaque orientation serait codée par une séquence kinesthésique de mouvements. Dans ce cas, le fait que toutes les orientations ne soient pas encodées relativement à des normes de référence mais plutôt par des séquences de mouvements expliquerait l'absence d'anisotropie dans la perception haptique des orientations. Ces processus haptiques, eux aussi relativement stables dans le temps, ont l'avantage de ne pas générer une perception anisotropique des orientations, mais ils présentent l'inconvénient de nécessiter des ressources attentionnelles conséquentes (sensibilité aux interférences).

III - La perception des propriétés matérielles des objets : l'exemple de la texture

Au sens large, toutes les propriétés physiques définissant la microstructure d'une surface sont englobées dans le terme « texture » : rugosité, dureté et élasticité, etc. Mais seules la rugosité et, dans une moindre mesure, la dureté, ont été bien étudiées. Pour faire varier expérimentalement la rugosité, on utilise des tissus différents, des papiers abrasifs dont la densité des grains est contrôlée, ou des stries rectilignes dont la profondeur et l'espacement sont variés.

a. La perception haptique est aussi précise que la perception visuelle

La perception tactile de la texture est aussi performante que la perception visuelle, et parfois même, pour les textures extrêmement fines des papiers abrasifs (allant des n° 1000 à 6000, selon les normes standard), elle la surpasse. C'est ce qu'a trouvé Heller (1989) chez des adultes voyants travaillant sans voir, des aveugles précoces et des aveugles tardifs, le statut visuel n'ayant pas d'effet sur les performances. La question de savoir si la perception bimodale visuelle et tactile (V + T) améliore ou non la discrimination est en discussion. Heller (1983) trouve une supériorité de la condition bimodale par rapport aux conditions unimodales (V seule, et T seul). Cependant, cette supériorité est due non à la perception de la texture elle-même, mais à la vision du mouvement d'exploration des mains. Dans d'autres études (Heller, 1985), il n'y a pas de différence entre les présentations uni- et bimodale du stimulus : les deux modalités participent également à la réponse donnée en présentation bimodale, mais les réponses sont plus rapides avec la présentation visuelle.

b. Les deux composantes (vibratoire et spatiale) de la perception haptique

Katz (1925/1989) est le premier à avoir étudié la perception haptique de la texture. Il a d'abord souligné la nécessité du mouvement d'exploration et, surtout, il a suggéré que la discrimination des textures relève d'un processus complexe comportant une composante spatiale, impliquée dans la perception des textures rugueuses, et une composante vibratoire à l'œuvre dans les textures plus fines. Les travaux contemporains ont approfondi ces différents points.

Le nécessaire changement dans la stimulation qui résulte du mouvement peut provenir indifféremment de l'observateur qui frotte la surface avec ses doigts, ou de la surface qui est déplacée sous le doigt immobile de l'observateur. Quand on compare la discrimination de papiers abrasifs dans une condition totalement passive (sans mouvement) et en condition de déplacement passif de la surface sous le doigt du sujet, nous observons que les textures les plus fines sont bien discriminées en présence de mouvement, mais ne le sont plus en condition purement statique. Au contraire, pour les textures les plus rugueuses, il n'y a pas de différence entre les conditions statique et avec déplacement. Selon Hollins et Risner (2000), c'est parce que la condition statique supprime la composante vibratoire de la stimulation qu'elle dégrade la discrimination des textures fines alors que pour les textures rugueuses, la composante spatiale (la seule préservée en condition passive) est suffisante pour permettre la discrimination.

Pour valider la présence des deux composantes vibratoire et spatiale dans la perception haptique des textures, Hollins, Fox et Bishop (2001) montrent que si, à l'insu des participants, on imprime une légère vibration (150 Hz) à une surface texturée, celle-ci sera jugée de plus en plus rugueuse à mesure que l'amplitude de cette vibration augmente jusqu'à 400 Hz. Cet effet apparaît même chez les personnes qui se sont aperçues de la présence d'une vibration de la surface. C'est donc bien que la composante vibratoire joue un rôle dans la perception de la texture, en plus de la composante spatiale déjà connue.

On sait que la perception haptique ne se fait pas uniquement par contact cutané direct avec une surface, mais aussi en contactant cette surface avec un instrument rigide tenu par la main (comme par exemple la canne utilisée par les aveugles). Klatzky et Lederman (1999) montrent que la composante vibratoire est préservée en grande partie quand l'exploration de la surface se fait par l'intermédiaire d'un instrument rigide. Elles ont ainsi comparé l'estimation subjective de la rugosité et la discrimination de la rugosité de surfaces comportant des petits points en relief selon que les participants utilisent leur doigt nu, ou leur doigt portant un gant, ou avec un gant comportant une petite plaque rigide sous l'extrémité de l'index, ou enfin en utilisant un bâtonnet cylindrique terminé soit par une base arrondie de 4 mm de diamètre, soit par une base conique de 2 mm de diamètre. Dans toutes les conditions, les discriminations sont au-dessus du niveau du hasard, et la rugosité estimée varie en fonction de l'espacement entre les éléments de la surface. Les meilleures performances sont observées avec le doigt, et, quand la perception se fait à travers un objet rigide, l'impression de rugosité augmente pour les surfaces les plus fines.

IV - L'identification des objets usuels

Lorsqu'il s'agit d'identifier de petits objets usuels ayant une signification pratique (fourchette, clé, bouchon de bouteille, etc.), la modalité haptique se montre très performante, voire aussi performante que la modalité visuelle. L'identification est en effet très rapide et pratiquement sans erreurs (Klatzky *et al.*, 1993). Le processus de reconnaissance haptique a des éléments communs avec la reconnaissance visuelle, comme la décomposition de l'objet en ses traits, la comparaison en mémoire, et la contribution des traitements descendants aussi bien qu'ascendants.

Pendant, les reconnaissances haptique et visuelle d'objet divergent quand on considère la nature des propriétés traitées. Dans la vision, cette identification se fait principalement par l'analyse de l'arrangement spatial des arêtes, constitutive de la forme (Biederman, 1987 ; cf. ci-dessus p. 58). Or, nous avons vu que la modalité haptique est peu performante dans le domaine spatial et en particulier dans la discrimination des formes complexes. Klatzky et Lederman (2000) répondent que si la forme intervient effectivement dans l'identification haptique des objets familiers, leurs propriétés matérielles (texture, rigidité, température, poids) jouent un rôle important et se combinent à la forme dans l'identification de l'objet.

En étudiant la dynamique temporelle de l'intégration de ces propriétés, Lederman et Klatzky (1997) montrent que les propriétés matérielles des objets (texture et dureté) s'intègrent très précocement et en tous cas avant les propriétés spatiales. Pour cela, les auteurs ont repris en haptique le célèbre paradigme de détection de cible parmi un nombre variable de distracteurs développé en vision par Treisman et ses collègues (Treisman et Gormican, 1988). Voici sa logique : si le temps de réponse pour identifier la cible augmente linéairement avec le nombre de distracteurs, c'est qu'un processus d'attention sélective est à l'œuvre (comme par exemple, identifier un « Q » parmi des

« O »). Au contraire, si ce temps reste stable, c'est que la cible « saute aux yeux » sans mobiliser l'attention (comme par exemple, identifier un « O » parmi des Q). En haptique, Lederman et Klatzky (1997) trouvent que les propriétés matérielles sont accessibles très tôt, alors que les propriétés spatiales d'orientation et de contour continu sont d'un accès tardif.

En résumé, l'efficacité de la modalité haptique dans la perception des propriétés de matériaux et la combinaison de ces perceptions aux propriétés de forme auxquelles elle accède (bien que de façon peu efficace) expliquent la très bonne expertise de cette modalité dans l'identification d'objets usuels.

L'agnosie haptique : une incapacité à reconnaître les objets avec ses mains


Les troubles de la perception haptique peuvent se situer à plusieurs niveaux. Il faut d'abord distinguer ceux liés aux caractéristiques spatiales ou matérielles des objets de ceux relatifs à leur intégration (Delay, 1935). Ainsi, l'*amorphognosie* est une incapacité à percevoir correctement la taille et la forme des objets, et l'*ahylognosie* concerne le poids ou la texture. L'*agnosie haptique* est un défaut d'identification des objets en l'absence des deux troubles précédents.

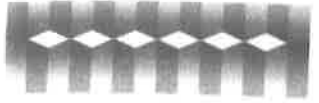
L'agnosie haptique doit être ensuite distinguée de l'*aphasie haptique*, qui recouvre un trouble de la dénomination des objets perçus (Endo *et al.*, 1992 ; Geschwind, 1965). L'aphasie haptique est généralement observable chez des patients ayant subi une section chirurgicale ou accidentelle du corps calleux (cf. part. 2, chap. 7), et concerne principalement la main gauche lorsque les objets tactiles utilisés dans les tests sont simples. Ce défaut de dénomination des objets palpés de la main gauche serait dû à la déconnexion hémisphérique qui empêcherait la mise en relation de la représentation de l'objet tactile et le nom correspondant. En effet, les aires corticales impliquées dans la construction de la représentation spatiale des objets palpés de la main gauche sont situées dans l'hémisphère droit qui, suite à la section du corps calleux, n'est plus en contact avec les zones du langage situées dans l'hémisphère gauche (Baynes, Tramo, Reeves et Gazzaniga, 1997). Cependant, en présence de matériel plus difficilement identifiable que les objets communs (lettres et formes avec ou sans signification découpées dans du papier de verre), des troubles de la reconnaissance haptique de la main droite peuvent être observés (Badan et Caramazza, 1997). Ces résultats soulignent les limites de l'hémisphère gauche pour le traitement des informations haptiques, et mettent l'accent sur la nécessité d'une contribution de l'hémisphère droit pour la construction des représentations spatiales haptiques.

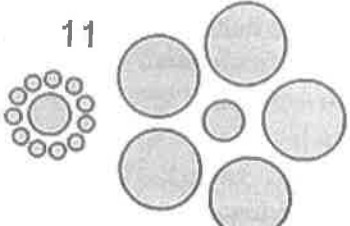
D - LES ILLUSIONS PERCEPTIVES

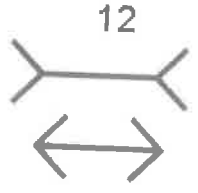
Les illusions constituent une voie royale pour saisir la nature, la diversité, la complexité des traitements perceptifs ainsi que leur invariance d'une personne à l'autre, puisque la plupart d'entre nous y sommes sensibles. Si elles sont un objet d'intérêt fort ancien (Antiquité), leur caractérisation est récente. En effet, on doit aux pionniers de la psychologie expérimentale (Wundt, Titchener, Helmholtz, Müller-Lyer ; moitié du XIX^e siècle) la démonstration des premières illusions géométriques, et à Mach (1868) la première explication neurophysiologique de certaines d'entre elles (contraste de lumi-

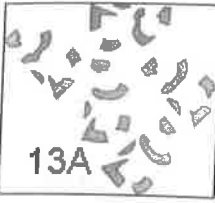
Illusions perceptives

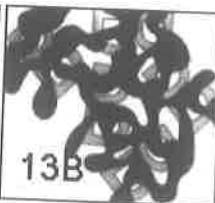
9 

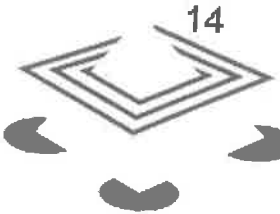
10 

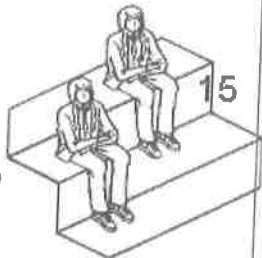
11 

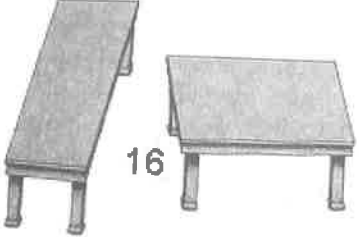
12 


13A 

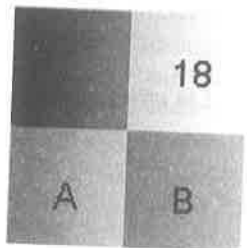
13B 

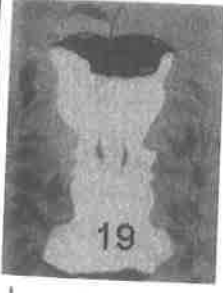
14 

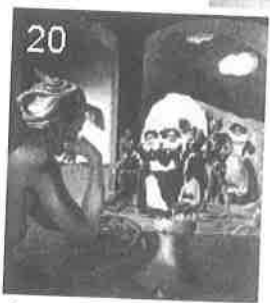
15 


16 


17 


18 

19 

20 

21 

22 

23 

nance, cf. fig. 9 et 10). Parmi les autres dates marquantes, retenons les travaux de la théorie de la Gestalt (cf. section 1) et ceux de Julesz (1960). Ce chercheur va expliquer la vision 3D stéréoscopique en réalisant sur ordinateur les premiers stéréogrammes à points aléatoires qui, dans les années 1980, donneront naissance aux fameux autostéréogrammes vendus dans le commerce. Si la très grande majorité des illusions perceptives concerne la vision, il en existe toutefois de très marquantes au niveau auditif, mais très peu en haptique.

Nous rappelons ci-dessous les principaux types d'illusion observés pour chaque modalité sensorielle, en resituant brièvement les mécanismes perceptifs les sous-tendant (voir Ninio, 1998, pour une revue). De nombreuses et puissantes illusions visuelles incluent le mouvement. Elles ne peuvent donc être illustrées ici. La consultation des différents sites web qui se développent sur ce thème permet d'en avoir un aperçu (cf. par exemple le site de M. Bach, un des plus exhaustifs du domaine : <http://www.michaelbach.de/ot/index.html>).

I - Les illusions visuelles

a. Les effets de contraste

Pour déterminer les contours et surfaces qui composent notre environnement, le système visuel analyse les différences continues de luminosité ou de couleur par le biais de cellules qui fonctionnent en interaction inhibitrice latérale (cf. part. 2, chap. 9). Parfois, parce que les contrastes sont forts et les éléments trop rapprochés, des *distorsions visuelles* apparaissent comme dans les figures 9 et 10 : la ligne de démarcation entre les carrés noirs n'est pas inclinée mais horizontale, et les losanges ne sont pas plus blancs que le reste de la page. Plus surprenant, des contours illusoire peuvent même être extraits : la pyramide blanche n'est en fait pas dessinée en 14.

b. Les effets de contexte

L'illusion 12 (Müller-Lyer) relève à la fois d'un effet de contraste (les penes selon leur ouverture masquent plus ou moins la perception des extrémités de lignes et conséquemment la lecture de leur longueur) et d'un effet plus cognitif liée à la perspective (la ligne avec les penes ouvertes pourrait évoquer un angle de mur en creux (donc à une vue de loin), celle avec les penes fermées à un angle en coin (vue de plus près)). L'illusion 11 (Ebbinghaus/Titchener) renvoie à l'estimation des surfaces du cercle central en fonction d'une référence spatiale fournie par le contexte des cercles extérieurs ; du coup, alors que les cercles centraux sont de taille identique, celui de droite apparaît plus petit.

c. Les effets de perspective et d'ombrage

Nous percevons le monde en trois dimensions alors que l'image rétinienne est bidimensionnelle. Différents indices sont utilisés par le système visuel pour reconstruire la profondeur :

- des indices proprioceptifs comme l'accommodation du cristallin ou la convergence oculaire ;
- des indices rétinien comme la disparité binoculaire ou la parallaxe du mouvement, qui renvoient au fait que les projections rétiniennes sur les deux yeux diffèrent selon que l'objet est en deçà ou au-delà du point de fixation (disparité binoculaire) ou proche ou lointain (parallaxe du mouvement) ;
- des indices plus cognitifs de type géométrique (perspective, variation de texture) ou de masquage des objets entre eux ou d'ombrage.

Ces derniers indices sont particulièrement puissants et peuvent générer de fortes illusions. Les deux tables (16) ont les mêmes longueur et largeur, mais la perspective fait que celle de gauche est perçue plus étroite et plus longue. La variation de luminance de la figure 17 est interprétée en termes de volumes concentriques, alternant bosses et creux. Cette perception du relief repose sur l'hypothèse (non consciemment formulée) que la lumière vient du haut (soleil). Mais attention, dans cette illusion, le « haut » est « ce qui est au-dessus de notre tête », même si le dessus peut être le bas au niveau terrestre. En effet, creux et bosses s'inversent si l'on regarde la figure 17 en se retournant, la tête en bas basculée entre les jambes (cf. la gestion des référentiels spatiaux évoqués dans la section A). C'est également l'interprétation non consciente de l'ombre dans la figure 18, ici de l'ombre « portée » qui provient de la gauche de l'image, qui fait apparaître les carreaux A et B comme étant de luminosité différente alors qu'ils sont strictement identiques. La variation de l'ombre portée d'un ballon au dessus du sol aide par exemple le système perceptif à estimer sa hauteur et trajectoire. La figure 13 B renvoie à deux mécanismes : le masquage par la tache noire fait apparaître les lettres B comme situées en arrière de la tache (le masquage est alors un indice de profondeur), mais, comme l'illustre la figure 13 A, la tache permet surtout aux lettres B de se révéler. Le masquage illustre alors deux lois de la Gestalt : la loi de clôture, présente en 13 A au niveau de chaque forme locale et les faisant apparaître comme des éléments isolés, et la loi de continuité, qui s'exerce en 13 B et permet de relier les éléments locaux en formes globales (ici les lettres B). Dans ce dernier cas, la figure 13 renvoie au problème de la ségrégation figure-fond.

d. Les figures ambiguës et la ségrégation figure-fond

Interpréter une image requiert de déterminer ce qui, dans ses composantes, va constituer le fond de l'image sur lequel se détacheront ses figures. Comme l'illustrent les figures « ambiguës » 19 et 20 (extrait du « Marché aux esclaves » de S. Dali), la séparation figure/fond peut parfois être problématique : est-ce Adam et Ève de profil ou la pomme croquée (19), « le buste de Voltaire » ou « deux nonnes se promenant » (20) ?

Remarquons que nous percevons l'une ou l'autre possibilité mais jamais les deux en même temps. La résolution de l'ambiguïté va dépendre de nos attentes par rapport à l'image. Elle dépendra également du niveau d'échelle spatiale que nous utilisons pour analyser l'image : vu de loin (*i.e.* avec une résolution spatiale basse : cf. section A), seul le buste de Voltaire est visible. De nombreux peintres ont incorporé ce genre d'ambiguïté dans leur œuvre, tels Giuseppe Arcimboldo au XVI^e siècle (avec ses séries de tableaux dits des « quatre saisons » dont la plupart étaient des portraits formés d'une juxtaposition de fruits, légumes, végétaux, symbolisant les saisons) et plus récemment Magritte, Dalí, Escher, Vasarely, l'école du pop-art, etc.

c. Les figures impossibles

Les figures 15 et 21 (extrait d'une lithographie d'Escher) illustrent des aberrations de structure spatiale : en 16 les personnages sont-ils assis sur la même marche ? En 21 les escaliers du sommet du château sont-ils montant et descendant ? Le fait que chacun perçoive ces aberrations montre que nous utilisons tous les mêmes mécanismes d'analyse et de structuration de l'image et en tirons les mêmes conclusions cognitives. Soulignons également que la lithographie d'Escher regroupe nombre d'indices de perception de la profondeur décrits plus haut (perspective, ombrage, masquage).

II - Les illusions auditives

a. L'illusion de continuité

L'illusion de continuité peut conduire soit à percevoir un stimulus qui n'existe pas, soit, à l'inverse, à ne pas percevoir un stimulus qui pourtant existe physiquement. Cette illusion est bien connue dans le domaine de la perception visuelle (fig. 13).

Dans le domaine de la perception de la parole, cette illusion est connue sous le nom de restauration phonémique. Ainsi, un discours interrompu par des silences de quelques centièmes de seconde est perçu de façon tellement hachée qu'il en devient inintelligible pour l'auditeur. Si les silences sont remplacés par des bruits, non seulement le discours semble normal (même si les bruits sont bien perçus) mais surtout l'intelligibilité du discours n'est pas perturbée. L'auditeur restaure donc les phonèmes manquants de façon à pouvoir suivre le message.

Cette illusion est également très puissante dans le domaine de la perception auditive non verbale. Si l'on fait entendre un son pur d'une fréquence et d'une amplitude donnée qui reste stable pendant deux secondes, interrompu par un silence de quelques centièmes de seconde, l'auditeur entendra deux sons successifs. Si on remplace le silence par un son complexe (un bruit) d'amplitude plus importante que le son pur et de durée strictement égale à celle du silence, l'auditeur aura alors l'impression d'entendre de façon continue le son pur se poursuivre derrière le bruit plus fort. Inversement, si le son pur se poursuit pendant la présentation du bruit mais s'arrête au même moment que le bruit, l'auditeur déclare entendre un son pur qui s'arrête au moment où le bruit commence !

b. L'illusion de localisation

Les illusions auditives donnent souvent lieu à des erreurs d'interprétation dans la localisation des sons. L'illusion de la gamme en est un exemple. Les sons présentés aux oreilles droite et gauche (fig. 22 A) forment une gamme ascendante et une gamme descendante. Le résultat perceptif est la perception par l'oreille droite des sons aigus qui descendent une partie de la gamme puis remontent, et la perception par l'oreille gauche des sons graves qui montent une partie de la gamme puis descendent (fig. 22 B). Ce qui est surprenant dans cette illusion est que non seulement l'auditeur ignore les sons graves transmis par l'oreille droite et les sons aigus transmis par l'oreille gauche, mais qu'il déclare entendre dans l'oreille gauche un son grave pourtant présenté dans l'oreille droite. Cette illusion vient appuyer l'idée que le système auditif tout comme le système visuel mettent en jeu deux types de processus distinguant l'identification et la localisation des objets.

c. L'illusion de hauteur

Comme pour l'escalier illustré dans le tableau d'Escher (fig. 21), on peut construire des sons qui donnent l'illusion de monter ou de descendre indéfiniment (fig. 23). Cette illusion, créée par Shepard (1964), consiste en une suite de notes qui montent (ou qui descendent) par pas de demi-ton (*Do, Do#, Ré, Ré#, Mi, Fa, Fa#, etc.*). Dans la première mesure (fig. 23), la montée chromatique se joue en clé de *sol* avec un *decrescendo* (diminution progressive de l'intensité des sons), alors qu'elle se fait en clé de *fa* en *crescendo* (augmentation progressive de l'intensité des sons) et ainsi de suite pour les mesures suivantes. La hauteur du son peut donc ainsi continuer à monter ou à descendre indéfiniment.

III - Les illusions haptiques géométriques

L'illusion de Müller-Lyer (fig. 12) est observée dans la modalité haptique, avec une intensité similaire chez les adultes travaillant sans voir et des aveugles tardifs (ayant eu de l'expérience visuelle), et même chez des aveugles de naissance (sans aucune expérience visuelle). Trois facteurs sont responsables de la présence de l'illusion de Müller-Lyer et des variations de son intensité en vision et en haptique : 1 / la répétition, avec une illusion qui diminue à mesure que le nombre de présentations de la figure augmente) ; 2 / l'angle, avec une illusion qui est d'autant plus forte que l'angle aigu formé par les pennures et le segment à évaluer est petit ; et 3 / les consignes, avec une illusion qui disparaît lorsqu'on demande aux participants d'ignorer les pennures et d'utiliser leur corps comme référence spatiale.

En revanche, les illusions du type de Titchener (fig. 11) existent seulement dans la vision. Ces résultats montrent que le sens haptique est parfois aussi sensible et parfois moins sensible que la vision aux illusions perceptives (Gentaz et Hatwell, 2004). Il semble donc être un sens un peu moins « trompeur » que la vision.

E - L'INTÉGRATION MODALE

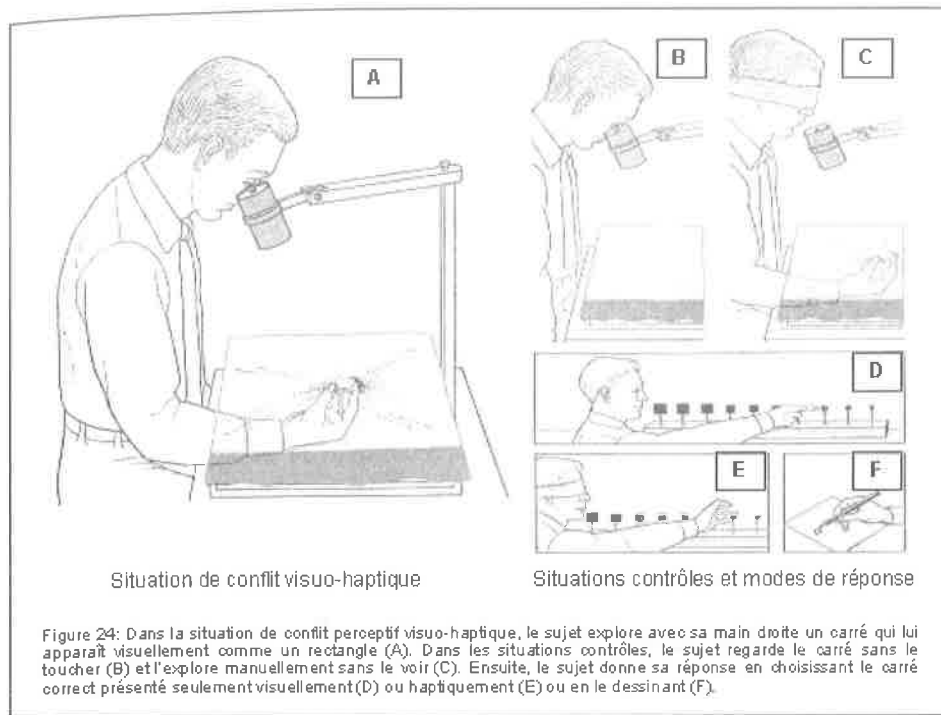
I - L'intégration des informations apportées par les modalités visuelle et haptique

Le plus souvent, les modalités visuelle et haptique fonctionnent en même temps. Dans certains cas, elles peuvent informer sur des propriétés distinctes des objets, comme par exemple lorsque l'œil perçoit la forme d'une éponge, et la main perçoit sa souplesse. Il faut alors combiner ces deux données pour former une représentation complète de l'objet. Dans d'autres cas, qui ont été étudiés plus souvent, les deux modalités accèdent à la même propriété, comme lorsque nous percevons en même temps, par la vision et le toucher, la forme, ou la taille, ou la texture d'un objet. Il est important alors que les deux systèmes perceptifs transmettent des paramètres identiques concernant la propriété en question, pour assurer la cohérence des représentations. Mais comment cette cohérence peut-elle être atteinte alors que les modalités visuelle et haptique n'ont pas les mêmes capacités discriminatives et ne mettent pas toujours en œuvre les mêmes processus de traitement ?

Pour étudier l'intégration visio-haptique, on utilise souvent les *situations bimodales conflictuelles*. On dit qu'il y a un conflit perceptif quand les données visuelles et haptiques sont discordantes. Par exemple, Rock et Victor (1964) ont demandé à des adultes de regarder un carré en bois, inséré dans un trou aménagé au centre d'une tablette, à travers un cylindre doté d'un prisme qui rétrécit la dimension horizontale des objets de 50 % (les sujets ignorent évidemment la présence de ce prisme). Visuellement donc, le carré apparaît comme un rectangle ayant une largeur inférieure de moitié à sa hauteur. En même temps, les sujets explorent sans voir ce même carré sous la tablette, qui est protégée par un rideau (fig. 24). Haptiquement, le carré n'a pas subi de déformation physique et devrait donc apparaître comme un carré. Après une phase de présentation de ces données conflictuelles visuelles et haptiques, le sujet devait reconnaître (soit par la vision seule, soit par la main seule) l'objet vu et palpé parmi un ensemble de carrés et rectangles. La question est évidemment de savoir quel sera le percept final du sujet dans cette situation conflictuelle.

Les résultats montrent que quand le conflit porte, comme ici, sur une propriété spatiale, les sujets manifestent une tendance à la « capture visuelle », c'est-à-dire qu'ils ne tiennent pratiquement pas compte des données haptiques et basent leur jugement perceptif sur les données apportées par la vision, surtout lorsque le test est lui-même visuel. Cela signifie qu'à la phase test, ils désignent le rectangle qu'ils ont vu (ou un rectangle très proche) comme étant celui qu'ils ont vu et palpé pendant la présentation. Quand le test est haptique, on trouve parfois des réponses de « compromis » où la réponse du sujet est une sorte de moyenne entre les valeurs visuelle et haptique du stimulus.

En revanche, quand le conflit porte non plus sur une propriété spatiale, mais sur la texture de papiers abrasifs de différentes rugosités, on observe soit des réponses de



compromis entre les deux valeurs conflictuelles (Lederman et Abbott, 1981), soit une tendance à une capture haptique lorsque la consigne demande d'évaluer la rugosité proprement dite. Par contre, lorsque la consigne demande d'évaluer la densité spatiale des grains de papiers abrasifs, une tendance à la capture visuelle est observée (Lederman, Thorne et Jones, 1986). La modalité haptique semble donc bien spécialisée dans l'appréhension des propriétés matérielles des objets, alors que la vision l'est dans le domaine spatial.

II - L'intégration des informations apportées par les modalités visuelle et auditive : l'exemple de l'effet McGurk

En général, la vision l'emporte sur l'audition. Le cas le plus impressionnant est celui de la ventriloquie, qui donne véritablement l'impression que c'est la marionnette qui parle et non le marionnettiste. Si le marionnettiste ne mettait pas en mouvement sa marionnette, l'erreur de localisation de la source auditive disparaîtrait.

Dans une série d'expériences, McGurk et MacDonald (1976) montrent que la perception d'une syllabe présentée auditivement peut être affectée par la présentation

visuelle simultanée d'un geste articulatoire. Ainsi, si la syllabe auditive est /ba/ et que le sujet voit en même temps sur un écran une bouche effectuant le geste articulatoire correspondant à la production de la syllabe /ga/, sa réponse sera généralement de dire qu'il a entendu la syllabe /da/. Cette syllabe /da/ occupe une position intermédiaire du point de vue articulatoire entre /ba/ et /ga/. Cela indique clairement que les informations auditives et visuelles ont été « intégrées » par le sujet, la réponse de ce dernier traduisant un compromis entre les deux sortes d'informations contradictoires. Cet effet McGurk pourrait d'ailleurs expliquer pourquoi un mauvais doublage d'un film engendre une baisse de l'intelligibilité du discours.

Le but de ce chapitre était de dresser une brève synthèse de l'état des connaissances sur les trois sens les plus utilisés par l'homme dans sa vie de tous les jours. Comme nous l'avons vu, cet état diffère d'une modalité à l'autre, reflétant la densité des recherches consacrées à chacune d'elles, densité induite elle-même par le poids respectif de ces modalités dans l'organisation du comportement perceptif. La vision domine chez le primate. Mais comme le souligne la dernière section de ce chapitre consacrée à l'intégration intermodale, chaque modalité a en fait son domaine de compétence. Il faut donc penser l'interaction entre les modalités sensorielles en termes de complémentarité plutôt que de concurrence, et l'enjeu des recherches à venir sera d'expliquer plus précisément comment fonctionne le dialogue intermodalitaire. C'est un enjeu théorique, mais également technologique et sociétal dans ce monde de plus en plus gagné par la réalité virtuelle plurisensorielle.

LECTURES CONSEILLÉES

- Bagot, J. D. (1996). *Information, sensation et perception*. Paris : Armand Colin/Masson.
- Botte, M.-C., Canévet, G., Demany, L., & Sorin, C. (1989). *Psycho-acoustique et perception auditive*. Paris : Éditions INSERM/EMI/SFA/CENT.
- Chokron, S., & Marendaz, C. (2005). *Comment voyons-nous ?* Paris : Le Pommier.
- Gentaz, E., Hatwell, Y., & Hennion, B. (2006). Le sens haptique manuel. In P. Fuch (Ed.), *Traité de réalité virtuelle 3* (pp. 265-289). Paris : École des Mines (version électronique en accès libre pour les étudiants via le site de l'École des Mines de Paris).
- Gregory, R. L. (2000). *L'œil et le cerveau*. Bruxelles : De Boeck.
- Hatwell, Y., Streri, A., & Gentaz, E. (Eds.) (2000). *Toucher pour connaître*. Paris : PUF.
- Marendaz, C., Rousset, S., & Charnallet, A. (2003). Reconnaissance des scènes, des objets et des visages. In A. Delorme & M. Flückiger (Eds.), *Perception et Réalité* (pp. 247-273). Montréal : Gaëtan Morin, et Bruxelles : De Boeck.
- McAdams, S., & Bigand, E. (1994). *Penser les sons : Psychologie cognitive de l'audition*. Paris : PUF.
- Ninio, J. (1998) *La science des illusions*. Paris : Odile Jacob.
- Sacks, O. (1996). *Un Anthropologue sur Mars*. Paris : Le Seuil.