

[révisé 21 Decembre 2004]

Alain de Cheveigné

Colloque espaces de l'Homme, Collège de France, 14-15 Octobre 2003.

Espace et Son

Introduction

Ce chapitre est un essai sur les interactions cognitives entre son et espace. Sont abordés les indices sonores qui nous renseignent sur l'espace, les indices spatiaux qui facilitent la perception de la scène sonore, les métaphores spatiales du son, les liens entre la perception et l'action d'un organisme, et enfin les exemples qui montrent que la connaissance de ces phénomènes est utile pour guider notre propre action.

Dans ses grandes lignes, notre système auditif ressemble à celui des autres vertébrés: mammifères (Fay et Popper 1994; Au et al. 2000), oiseaux ou reptiles (Dooling et al. 2000), amphibiens ou poissons (Fay et Popper 1999). Ceci témoigne de son ancienneté phylogénétique. La diversité des oreilles des insectes fait penser que l'audition chez eux est apparue plusieurs fois de façon indépendante (Hoy et al. 1998), signe d'une forte pression sélective en faveur de l'évolution d'un système auditif. La sensibilité *spatiale* auditive s'observe dans tout le règne animal, depuis la rudimentaire sélectivité angulaire des cellules ciliées individuelles chez les poissons (Popper et Fay 1999) jusqu'à l'exceptionnelle sélectivité spatiale de la chouette effraie ou la fine résolution du biosonar des mammifères marins et aériens (Popper et Fay 1995; Dooling et al. 2000; Au et al. 2000). La diversité de ces stratégies et structures peut nous aider à comprendre la cognition spatiale auditive chez l'espèce qui nous intéresse le plus: la nôtre.

La psychophysique s'efforce généralement d'isoler les compétences purement auditives en minimisant les interactions avec d'autres modalités (Blauert 1997). Dans le monde animal, les compétences spatiales sont d'une grande variété (Heffner et Heffner 1992), mais chez de nombreuses espèces, dont l'homme, on constate que la discrimination spatiale auditive est médiocre comparée à la vision. En vision, notre acuité angulaire est de l'ordre de 5 secondes d'arc (Barlow et Mollon 1989), et en audition de 1 à 3° en azimut et 10 à 20° en élévation, au mieux. L'acuité auditive est particulièrement médiocre pour certains stimuli, et se dégrade en présence de réverbération (Blauert 1997). Quelle est l'utilité d'une modalité spatialement aussi

peu compétitive ? Pour le comprendre il faut tenir compte des complémentarités entre sens, des interactions intermodales, et des relations entre perception et action. Si l'évolution nous a dotés d'un système perceptif aussi peu précis, c'est certainement qu'il existe des situations cruciales où il est suffisant et nécessaire à guider (ou à spécifier un modèle qui guide) notre action et assurer notre survie.

Pour illustrer les liens entre perception et action, les physiologistes Szentágothai et Arbib (1975) ont donné l'exemple d'un organisme hypothétique, sorte de poisson muni de deux nageoires et de deux yeux, et d'un système nerveux primitif liant l'oeil gauche à la nageoire droite et vice-versa. Si une proie apparaît dans le champ du capteur gauche, l'ordre est transmis à la nageoire droite de se mouvoir et le poisson s'oriente alors vers la cible. Cette orientation se maintient par l'activité équilibrée des deux nageoires jusqu'à ce que la proie soit atteinte. Chez cet organisme simple, *perception* et *action* sont synonymes. Szentágothai et Arbib suggèrent qu'il en est de même pour des organismes plus évolués, à cela près que des niveaux successifs d'inhibition complexifient le comportement. La structure croisée de notre système nerveux actuel serait l'héritage de celle de cet organisme primitif.

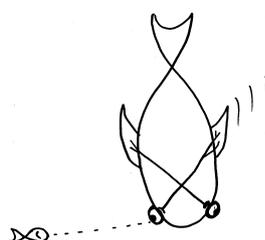


Figure 1. Szentágothai et Arbib (1975) ont utilisé l'exemple d'un organisme primitif tel que celui-ci pour illustrer la nature fondamentalement *active* de la perception. Chaque "oeil" est relié à la nageoire contralatérale, et la perception d'une proie équivaut ainsi à un ordre moteur pour le rejoindre.

La simplicité de cet organisme, qui paraît caricaturale, se retrouve chez certains papillons de nuit (Notodontidae) dont chaque oreille est dotée d'un neurone unique qui déclenche un réflexe de chute du papillon à l'approche d'une chauve-souris (Fullard, 1998; Pollack, 1998). L'information spatiale extraite est rudimentaire (essentiellement, deux "bits" d'information: présence vs absence, gauche vs droite), mais elle suffit dans la mesure où elle augmente les chances de survie. Il est à noter que l'action *complète* l'information perçue: l'orientation de l'organisme de la Figure 1 vers la proie détermine (à posteriori) son azimuth. Son déplacement détermine la distance. Ainsi l'action peut transformer une perception spatiale rudimentaire en coordonnées spatiales plus précises. Nous reviendrons plus loin sur cette idée.

Pour survivre dans un environnement qui comporte des prédateurs en plus de proies, un organisme doit disposer d'un répertoire de comportements plus riche que l'organisme de la Figure 1: tantôt s'approcher pour manger, tantôt s'éloigner pour

éviter d'être mangé (Pollack, 1998). Cela suppose aussi une compétence perceptive plus riche qu'on peut qualifier de *discrimination* (ou *reconnaissance*). Si, de plus, l'environnement est peuplé de façon dense (de prédateurs, proies ou obstacles) l'organisme devra en outre être capable de *ségrégation* perceptive, faute de quoi un organisme qui rencontre plusieurs proies risque d'adopter une trajectoire moyenne et de les rater toutes. L'importance de l'Analyse de Scène Auditive (ASA) n'a été pleinement reconnue que récemment (Bregman 1990), mais il est probable que la structure sophistiquée de notre système auditif résulte d'une pression sélective pour faire cette analyse de façon efficace (de Cheveigné 2001c).

Des sources distinctes occupent souvent des *positions spatiales* distinctes, et l'organisme met à profit cette structure spatiale pour faciliter l'organisation perceptive des sources qui composent la scène. Il le fait principalement de trois façons: (1) échantillonnage du champ acoustique au moyen des deux oreilles (complété par les propriétés directives des pavillons), (2) déplacement dans l'espace pour enrichir cet échantillonnage, (3) émission d'un champ acoustique pour sonder l'espace (écholocation). Les sections qui suivent explorent divers aspects de la perception spatiale auditive.

1. Rôle du son dans la perception de l'espace

1.1. Comportement et indices

Le champ acoustique issu d'une scène acoustique varie de point en point, et la nature de la variation dépend de la structure spatiale de la scène. Nos oreilles prélèvent deux échantillons de ce champ, et infèrent sur cette base la structure spatiale de la scène sonore. Pour connaître les propriétés acoustiques qui permettent la perception spatiale, une approche est de présenter aux oreilles de sujets expérimentaux des stimuli bien contrôlés, et de leur demander ce qu'ils entendent.

La présentation par casque est un moyen commode d'assurer que les stimuli sont contrôlés, mais dans ces conditions, les sujets décrivent en général une image auditive *interne* à leur tête. L'image est centrée si les signaux sont identiques; si l'un des signaux est *plus intense*, ou temporellement *en avance*, l'image se déplace vers l'oreille correspondante, tout en restant subjectivement interne à la tête. Cet ersatz de perception spatiale, appelé *latéralisation*, a fait l'objet de nombreuses expériences dans lesquelles la situation des sujets (sommés de choisir "gauche" ou "droite") n'était pas sans rappeler celle du poisson de la Figure 1. Malgré la pauvreté spatiale de l'impression subjective, ces expériences ont révélé les indices les plus importants de la perception spatiale: la *différence interaurale de temps d'arrivée* (ITD) et la *différence interaurale d'intensité* (IID).

Avec des techniques plus sophistiquées (écoute en champ libre, restitution synthétique au casque) les sujets consentent à décrire des sources dans l'espace extérieur à leur tête. Dans ces conditions d'écoute plus naturelles, on a pu montrer que les deux indices précités (ITD et IID) déterminent tous deux principalement *l'azimut* de la source subjective (Thompson, 1882; Blauert, 1997). Cela est logique puisque le son provenant d'une source située hors du plan médian parvient à l'oreille opposée avec retard (du fait d'un chemin plus long) et atténuation (du fait de l'effet

d'ombre de la tête). Chez l'homme et la plupart des animaux (Heffner et Heffner, 1992), ITD et IID déterminent conjointement l'azimut, le premier étant plus utile aux fréquences basses et le second aux fréquences élevées. Il y a des exceptions, notamment la chouette effraie (*Tyto Alba*). Ses oreilles sont asymétriques, et elle utilise les IID produits par cette asymétrie pour déterminer l'*élévation* d'une source. La chouette forme ainsi une carte bidimensionnelle précise de la direction d'incidence d'une source (équivalente, pour un oiseau en vol, à la position d'une proie sur le sol). Cette précision est exceptionnelle. Chez l'homme, la perception d'élévation est bien moins précise (la perception s'appuie sur la variation, en fonction de l'élévation, des propriétés de filtrage fréquentiel des pavillons, de la tête et du torse). La perception de la *distance* est encore plus aléatoire. Elle dépend d'indices tels que l'intensité (pour les sources dont la puissance habituelle est connue), le rapport d'intensité entre fréquences hautes et basses (les premières étant plus atténuées que les secondes par la propagation dans l'air), le rapport entre son direct et son réverbéré dans un espace clos, etc. Chacun de ces indices nécessite, pour être exploitable, une certaine familiarité avec la situation. Il s'agit donc d'*inférences* auxquelles contribuent, outre les indices acoustiques, la connaissance préalable des sources ou des lieux, les indices visuels, etc. Il est plus juste de parler de la construction, au sein de l'organisme, d'un *modèle* du monde environnant sur la base d'apports multiples dont seuls certains sont acoustiques. Le modèle, une fois construit, permet de faire des inférences qui dépassent la simple localisation d'une source.

Les indices cités sont ceux de la perception *passive*. L'action les complète si les circonstances le permettent. En tournant la tête nous distinguons une source située devant nous d'une source située à l'arrière (ITD et IID sont ambigus sur ce point). En penchant la tête de côté nous pouvons réinterpréter ces indices pour juger de l'élévation. Un déplacement latéral permet d'estimer la distance par triangulation, et un déplacement jusqu'à la source enlève bien sûr toute incertitude concernant sa position (Durlach et al. 2003). Nous reviendrons plus en détail sur la perception active de l'espace.

1.2. Anatomie et physiologie

Les vertébrés disposent en général de deux oreilles placées de part et d'autre de la tête (chez les insectes leur nombre et implantations sont plus variables). Chaque oreille prélève dans le champ acoustique un signal unique, combinaison d'ondes de provenance spatiale diverse. La forme et l'orientation (parfois motile) du pavillon pondèrent chaque apport de façon différentielle selon l'angle d'incidence, mais la somme est un signal unique pour chaque oreille (Bonnier 1896 pensait que l'incidence était conservée jusqu'au sacculus, qui formait ainsi une sorte de "rétine acoustique". Il n'en est rien.). Ce signal unique est transmis par les osselets de l'oreille moyenne à la cochlée où il est analysé comme par un banc de filtres. Plus précisément, le déplacement de chaque point de la membrane basilaire est une version filtrée de la pression acoustique au tympan, elle-même somme d'ondes incidentes. La base de la cochlée répond aux fréquences élevées, et l'apex aux fréquences basses. Le déplacement en chaque point est traduit par les cellules

ciliées internes en impulsions nerveuses qui sont transmises au cerveau par le nerf auditif. Les patterns nerveux transmis reflètent deux aspects des ondes acoustiques: leur *contenu fréquentiel* (codé par le degré d'activité de chaque fibre en fonction de sa provenance dans la cochlée) et leur *structure temporelle* (reflétée par la répartition temporelle des décharges).

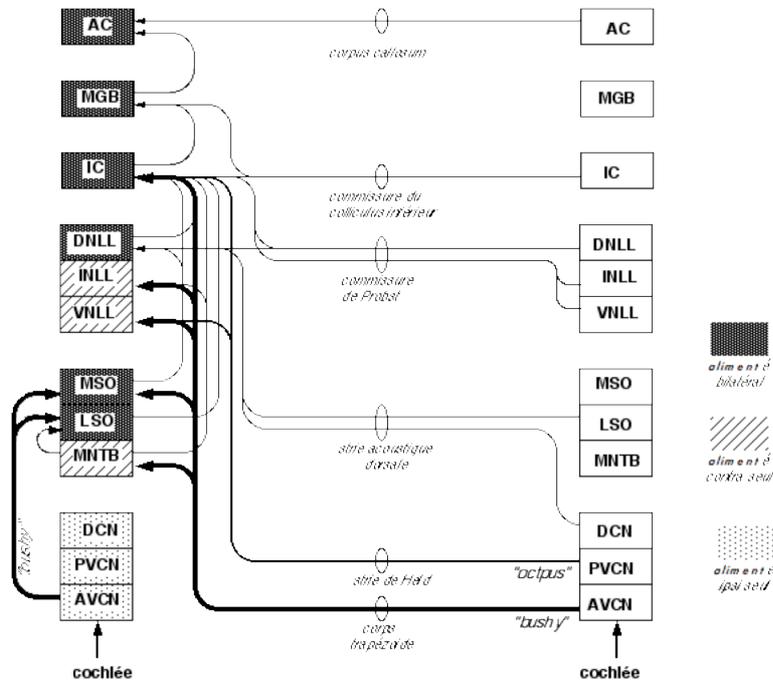


Figure 2. Représentation schématique des principaux relais auditifs, et des voies ascendantes qui permettent l'interaction binaurale. Les noyaux en sombre reçoivent des informations des deux oreilles: ce sont des sites possibles de traitement des indices spatiaux par interaction binaurale. Le noyau cochléaire se subdivise en parties étiquetées DCN, PVCN et AVCN, le complexe olivaire supérieur en parties étiquetées MSO, LSO et MNTB, et les noyaux du lemniscus latéral en parties étiquetées DNLL, INLL et VNLL. Le colliculus inférieur est étiqueté IC, le corps genouillé médian MGB, le cortex auditif AC.

L'information en provenance de la cochlée traverse une succession de relais dont les plus importants sont (chez les mammifères) le noyau cochléaire (CN), le complexe olivaire supérieur (SOC), les noyaux du lemniscus latéral (NLL), le colliculus inférieur (IC), le corps genouillé médian (MGB) et le cortex auditif (AC). La répartition des fréquences le long de la cochlée (tonotopie) se trouve projetée dans certaines parties de chacun de ces relais: à chaque relais, on trouve des neurones qui répondent préférentiellement aux composantes du stimulus proches de leur *fréquence caractéristique* (CF). Les voies nerveuses traversent le plan médian à plusieurs niveaux, ce qui permet l'interaction des signaux provenant des deux oreilles (Figure 2).

Le principal niveau d'interaction binaurale est le complexe olivaire supérieur. Au sein de ce complexe, l'*olive supérieure médiane* (MSO) reçoit bilatéralement de CN des axones de neurones spécialisés dans la transmission de l'information temporelle fine (Smith et al. 1993, 1998). Les apports des deux oreilles sont comparés pour déterminer un éventuel décalage temporel (ITD), reflet de la position latérale de la source (Yin et Chan 1990). Plus précisément: un axone d'une cellule spécialisée de CN rencontre un axone provenant du CN contralatéral. Les deux font synapse sur un neurone "détecteur de coïncidences" qui répond uniquement si des impulsions arrivent simultanément sur ses deux entrées. Si les délais de propagation des axones sont égaux, le neurone répond uniquement à une source centrée. S'ils diffèrent, il répond si la différence de délai *externe* (ITD) est égale et opposée à la différence de délai *interne*. Un pic d'activité au sein de la population de neurones signale ainsi la position de la source. MSO comprend, pour chaque CF, un réseau de cellules sensibles à une plage de ITD. Chez certaines espèces leur répartition est spatialement régulière (carte d'ITD), chez d'autres la régularité est moins évidente (Schwartz 1992; Helfert et Aschoff 1997; Carr et Code 2000). Cette description est conforme au modèle de Jeffress (1948), dont la valeur prédictive s'est révélée plusieurs décennies après sa publication (Joris et al 1998; mais voir aussi McAlpine et al 2001).

Le complexe olivaire supérieur contient deux autres noyaux importants pour le traitement des indices spatiaux: le *noyau médian du corps trapézoïdal* (MNTB) et l'*olive supérieure latérale* (LSO). Le premier rend inhibitrice l'activité provenant du CN contralatéral, et la présente ainsi "inversée" à LSO. LSO reçoit par ailleurs, comme MSO, un apport excitateur du CN ipsilatéral, et opère ainsi une sorte de "soustraction" des patterns provenant des deux oreilles. Ainsi serait traité le deuxième indice majeur de localisation spatiale, la différence d'intensité interaurale (IID).

Les calculs faits par MSO et LSO sont transmis au colliculus inférieur (IC). IC fait office de "gare de triage" pour les résultats des traitements du tronc cérébral, avant leur transmission vers le thalamus (MGB) et le cortex (AC). On y trouve de nombreuses cellules sensibles aux disparités interaurales, certaines avec une réponse semblable à celle observée dans MSO, d'autres avec des réponses plus complexes. Il y a notamment une convergence des cellules de MSO de même ITD caractéristique, mais de CF différente. IC intègre aussi l'apport des cellules de LSO sensibles à l'IID (Kuwada et al. 1980, 1997; Yin et al. 1987; Ehret 1997).

Chez la chouette effraie, on observe dans le noyau extérieur (ICx) une représentation bidimensionnelle de la sensibilité à ITD et IID (Klump, 2000; Carr et Code, 2000) qui devient, après projection dans le tectum (colliculus supérieur), et mise en correspondance avec les apports visuels, une véritable *carte spatiale* représentant azimuth et élévation. Chez d'autres espèces, on trouve bien des cellules sensibles aux indices spatiaux, mais leur répartition au sein du tissu nerveux est moins systématique.

IC projette sur MGB (thalamus) qui projette sur AC (cortex), où l'on retrouve des cellules spatialement sensibles mais dont les réponses sont plus complexes et souvent labiles (les propriétés du neurone varient au cours du temps) (Nelken et al. 2003; Klump 2000). On ne trouve pas de carte spatiale dans le thalamus ou le

cortex (Clarey et al. 1994), et la nature des traitements et représentations de l'espace reste obscure. L'information auditive se combine avec celle provenant d'autres modalités (vision, vestibule, etc.) dans des aires polysensorielles ou d'association, et aussi dans des aires primaires. Par exemple, le cortex auditif projette dans la partie de l'aire visuelle V1 sensible à la périphérie du champ visuel (Falchier et al. 2002).

2. Rôle de l'espace dans la perception du son

Les sons qui se propagent de la source jusqu'à l'auditeur sont affectés par l'espace qu'ils traversent. Ici, un seul aspect sera évoqué: l'avantage qui en résulte pour la perception de sources concurrentes dans une scène complexe. Les ondes provenant de plusieurs sources forment aux oreilles un mélange complexe, que le système auditif interprète pourtant en termes d'une scène auditive peuplée de sources distinctes. C'est ce que Bregman (1990) appelle l'*Analyse de Scène Auditive* (ASA). Parmi les indices exploités, il en est qui découlent de la structure spatiale de la scène. Ainsi, la perception d'une source est facilitée lorsqu'elle est spatialement distincte par rapport aux sources concurrentes. Les indices et les mécanismes sont de trois types: *acoustiques* (géométrie de la propagation), *neuraux* (interaction binaurale dans le tronc cérébral) et *cognitifs* (modèles spatiaux facilitant l'interprétation).

La géométrie des ondes incidentes peut réduire l'apport de la source perturbante à l'une ou l'autre oreille, grâce à l'effet d'ombre acoustique de la tête, ou grâce aux propriétés directives des pavillons. Il suffit alors de prêter une attention privilégiée à l'oreille la plus favorable (voire ajuster son orientation). Ces indices géométriques sont particulièrement utiles aux hautes fréquences, pour lesquelles la longueur d'onde est faible par rapport aux dimensions de la tête et des pavillons, en particulier chez les espèces munies de pavillons directifs et/ou motiles (chauve-souris, chouette effraie, etc.). L'oreille de certains insectes comporte un tympan dont les deux faces sont accessibles aux ondes incidentes (Hoy et al. 1998). L'oreille fonctionne alors comme capteur de *différence de pression*. Pour un certain angle d'incidence, les ondes provenant d'une source arrivent également aux deux faces, et ont un effet nul. L'orientation de la direction de sensibilité minimale vers la source perturbante facilite la perception d'une source cible. Un principe semblable a été proposé chez les oiseaux, dont les oreilles moyennes communiquent par un canal central rempli d'air (Klump, 2000).

Chez l'homme, les pavillons sont peu directifs et (sauf exception) peu motiles. L'ombre de la tête atténue une source située du côté opposé à l'oreille, mais cet effet n'est appréciable qu'aux fréquences élevées, et n'est utile que si les sources cible et interférentes sont situées de part et d'autre du plan médian. Notre tympan n'est exposé que sur sa seule face extérieure, et la stratégie d'annulation acoustique offerte (en théorie) aux insectes et oiseaux est donc inaccessible. Un bénéfice équivalent serait possible si les signaux acoustiques, traduits en impulsions nerveuses, étaient soustraits *centralement*, les apports de l'oreille droite étant retranchés à ceux de l'oreille gauche et vice-versa. C'est de fait ce que propose le modèle d'*égalisation-annulation* (equalization-cancellation, EC) de Durlach (1963).

Selon ce modèle, les signaux provenant des deux oreilles sont "égalisés" (par atténuation pour compenser l'IID de la source interférente, et retard pour compenser son ITD) puis soustraits. Le reliquat de la soustraction reflète alors uniquement les sources dont l'IID et/ou l'ITD diffèrent de la source annulée. L'annulation neuronale offre, sur l'annulation acoustique, plusieurs avantages. L'ajustement des paramètres est interne et ne nécessite pas l'orientation de la tête. Le traitement se fait dans des canaux fréquentiels issus du filtrage cochléaire, et des paramètres différents peuvent donc être appliqués à chaque bande de fréquence. Ainsi le système peut-il compenser les irrégularités inévitables d'ITD et IID en fonction de la fréquence, dues à la dispersion. Il peut aussi annuler simultanément plusieurs sources si elles occupent des régions spectrales distinctes. Le modèle mEC ("modified EC") de Culling et al (1998) propose un tel mécanisme d'annulation canal-par-canal. Il est intéressant de remarquer la similitude de ce dernier modèle avec les techniques récentes d'*analyse en composantes indépendantes* (ICA) dans le domaine fréquence (Annemüller 2001; de Cheveigné 2001a).

Un substrat neuronal possible du modèle EC de Durlach (ou sa version modifiée de Culling et al.) est le noyau LSO que nous avons évoqué plus haut. L'anatomie de ce circuit suggère une spécialisation pour le transport des informations temporelles fines (synapses de Held dans le CN et MNTB, gros axones myélinisés) que les mesures physiologiques confirment (Schwartz 1992; Joris et Yin 1998). Une difficulté avec cette proposition est que, chez la plupart des animaux étudiés, LSO traite principalement les hautes fréquences alors que les données psychophysiques en faveur de ce modèle (chez l'homme) concernent plutôt les fréquences basses. D'autres sites d'interaction binaurale du type IE (inhibiteur exciteur) existent, dans les noyaux du lemnicus latéral, les noyaux périolivaires, et même MSO (Finlayson et Caspary 1991; Grothe 1994). Par ailleurs, l'analyse théorique montre que les fonctionnalités d'une interaction IE peuvent être obtenues à partir d'une interaction de type EE, telle qu'il existe aux basses fréquences dans MSO (de Cheveigné 2001b). L'hypothèse EC reste donc attrayante.

Un autre mécanisme de séparation spatiale a été proposé. En appliquant le modèle de Jeffress (1948) dans chaque canal fréquentiel indépendamment on peut, en principe, grouper les canaux en fonction de la direction d'incidence qui les domine. Cette idée, proposée à l'origine par Lyon (1983) et redécouverte de temps à autre (e.g. Patterson et al. 1996), reste populaire dans les modèles dits d'"analyse de scène auditive computationnelle" (Cooke 1991; Brown 1992; Ellis 1996). L'étiquetage spatial des canaux pourrait intervenir par exemple dans MSO, en appliquant le modèle de Jeffress dans chaque canal. La sélection des canaux pourrait intervenir à ce niveau ou à un niveau supérieur, par exemple le thalamus ou le cortex (voir de Cheveigné 2001c pour une revue de ces stratégies de ségrégation).

Les mécanismes cités jusqu'à présent exploitent des capacités de "traitement de signal" d'un niveau relativement bas, et ne suffisent pas à rendre compte de tous les aspects des apports spatiaux à l'analyse de scènes. Le modèle EC rend bien compte de l'apport binaural à la *détection*, mais non à l'*intelligibilité* (Shackleton et al 1994). Il semble que le bénéfice d'une séparation spatiale pour l'intelligibilité de la parole soit pour partie de nature *cognitif*, en ce qu'il permet au sujet de former un *modèle spatial* de la voix cible et/ou concurrente (Darwin et Hukin 1999, 2000a,b).

Les apports visuels ou autres concourent certainement à la formation de tels modèles, mais le détail de ces mécanismes de haut niveau est mal connu (Shinn-Cunningham 2002).

En dépit de ces mécanismes, les capacités d'analyse spatiale d'une scène acoustique complexe sont limitées. La localisation auditive est fortement dégradée en présence de sources multiples (Lorenzi et al. 1999; Durlach et al. 2003), alors que la localisation simultanée de nombreux détails visuels est possible. Ceci pose encore une fois la question de l'importance de l'apport de la modalité auditive à la cognition spatiale. La section suivante suggère que ce qui compte est l'adéquation de cet apport à l'action de l'organisme.

3. Perception active de l'espace

Une hypothèse est que des performances qui semblent rudimentaires sont néanmoins suffisantes, voire optimales, pour guider l'organisme dans son action. Pour évaluer cette hypothèse il faut rapprocher la perception de l'action au lieu de l'en isoler (comme aurait tendance à le faire la psychoacoustique classique).

Le poisson primitif de Szentágothai et Arbib illustre trois idées. La première est qu'il y a continuité entre systèmes sensitifs et moteurs. La seconde est que la qualité de la perception se juge à l'efficacité de l'action (qu'elle soit immédiate ou différée). La troisième est que l'action elle-même peut participer à la perception et l'enrichir. La résolution spatiale de l'organisme de la Figure 1 paraît pauvre, mais il lui suffit de *s'orienter* jusqu'à l'équilibre des excitations bilatérales pour déterminer l'azimut de la source, et ensuite de *s'avancer* pour déterminer, a posteriori, sa distance. L'action supplée ainsi à la perception.

Imaginons que le poisson de la Figure 1 puisse atténuer et/ou retarder l'activité d'un de ses capteurs par rapport à l'autre. Il pourrait alors déterminer l'azimut par simple ajustement de ces paramètres internes, *sans déplacer son corps*. Supposons de plus que, au lieu d'un seul "comparateur" muni d'un jeu unique de paramètres, il dispose d'une gamme parmi laquelle il peut choisir. Il pourrait alors déterminer l'azimut *sans ajustement*. Pour l'essentiel c'est ainsi que fonctionne MSO, dont le réseau de cellules comparatrices se comporte comme une population de "poissonculi", chacun attentif à une orientation particulière. L'avantage est évident: la détermination est rapide, peu coûteuse en énergie, et peu susceptible d'attirer l'attention.

L'avantage du système évolué est donc, non pas la précision spatiale de l'information glanée, mais le fait qu'elle est obtenue sans action (ou avant l'action). Le résultat d'un ensemble d'actions potentielles est en quelque sorte *prédit*. Ce rôle rappelle celui des mécanismes de contrôle proactifs ("feedforward") du cervelet par lesquels un modèle prédictif guide un mouvement complexe et rapide, en s'affranchissant des contraintes de temps et stabilité d'un contrôle rétroactif (Miall et al. 1993). Ainsi, après un son de seulement 75 ms, trop court pour qu'intervienne un ajustement rétroactif, la chouette effraie oriente sa tête avec une précision de l'ordre de 2° en azimut et élévation (dans un champ de ±30°). La grenouille, elle, opère une succession de sauts en zigzag guidés par le cri de son futur partenaire pour se

rapprocher de lui (Feng et Schellart 1999). L'avantage fonctionnel du système plus évolué de la chouette est évident.

L'action reste l'objectif ultime de la perception, et les exemples où perception et action sont intégrées sont multiples. L'organisme de la Figure 1 n'est pas le plus primitif qu'on puisse imaginer. Un simple détecteur de présence (produisant un bit d'information au lieu de deux) permettrait de localiser une proie par déplacement aléatoire de type mouvement Brownien. C'est ainsi que H. Sapiens trouve son téléphone portable, objet notoirement difficile à localiser. Le déplacement angulaire de notre tête permet de lever une ambiguïté d'azimut (par exemple avant vs arrière) ou d'élévation, et un déplacement latéral nous fait quitter un noeud ou ventre du champ réverbérant d'un espace clos, l'ombre acoustique d'un obstacle, ou l'influence d'une source parasite. Il a été suggéré que, chez le pigeon, l'effet doppler induit par le vol pourrait permettre de déduire l'orientation d'une source (Klump, 2000).

Un organisme doté de pavillons pourra les orienter. C'est le cas par exemple de chauves-souris dont les pavillons bougent (l'un vers l'arrière l'autre vers l'avant) à chaque impulsion émise (Popper et Fay 1995). Certains oiseaux possèdent des "caches" amovibles à l'entrée de leurs conduits auditifs (Klump 2000) dont le rôle pourrait être d'ajuster la sensibilité aux indices spatiaux. Le rôle des muscles de l'oreille moyenne est plus probablement d'atténuer des bruits intenses prévisibles, tels ceux produits par la voix. En revanche il a été proposé que les *cellules ciliées externes* de la cochlée, sous contrôle efférent du système olivocochléaire, permettent d'ajuster finement la sensibilité aux IID (Sahey et al. 1997). Ces cellules, dont le rôle est moteur, interviennent pour compenser les pertes dissipatives et ainsi augmenter la sensibilité et la sélectivité de la mécanique cochléaire. Ce système est phylogénétiquement ancien puisqu'on le retrouve chez tous les vertébrés et certains invertébrés (Warr, 1992; Guinan, 1996; Carr et Code, 2000; Sahey et al. 1997). La voie olivocochléaire est l'une des nombreuses voies centrifuges du système nerveux auditif, qui s'étendent depuis le cortex jusqu'à la cochlée, et confèrent au système auditif une nature fondamentalement *active*. On connaît mal le rôle fonctionnel du système efférent.

Mieux compris sont les mécanismes d'écholocation des chauves-souris et mammifères marins (Popper et Fay 1995; Au et al. 2000). La gamme de fréquences employée est large (jusqu'à 150 kHz). Les divers types de modulations temporelles exploitent à la fois le retard de l'écho (qui renseigne sur la profondeur) et l'effet doppler (qui renseigne sur la vitesse). L'écholocation sert à localiser et caractériser des proies, mais aussi à éviter les obstacles et sonder la structure spatiale d'un lieu (par exemple une grotte), stratégie utilisée aussi par certains oiseaux cavernicoles (Klump, 1998). Truax (1984) rapporte le cas de marins qui utilisent des sifflets pour se diriger à l'écho à l'entrée du port. Les aveugles caractérisent les espaces et s'orientent en leur sein en utilisant l'écho du bruit de leur canne, de leurs pas, ou de clics produits par leur bouche. Reznikoff (1987) a émis l'hypothèse selon laquelle l'homme préhistorique se dirigeait à la voix dans les cavernes en utilisant comme points de repère des résonances particulières. La voix est une source puissante, relativement directive, et dont la structure spectrotemporelle peut être finement contrôlée.

La portée de l'écholocation est limitée par l'atténuation du double chemin d'aller et retour, et par la nature aléatoire de la réflexion (même si certaines fleurs ont pris la forme de catadioptrés acoustiques pour attirer les chauves souris...). La portée est étendue si l'écho est remplacé par la *réponse* d'un autre organisme. L'échange acoustique sert, soit à rejoindre un conspécifique (partenaire, parent ou petit), soit à s'en éloigner ou le maintenir à distance (rival), soit encore à se situer spatialement par rapport à un troupeau ou une nuée. Un exemple remarquable d'Analyse de Scène Auditive est la localisation par les pingouins de leurs petits parmi des milliers d'autres (Aubin et Jouventin 1998). Un phénomène analogue (quoique moins spectaculaire) s'observe les dimanches au bois de Fontainebleau.

Il est intéressant de noter que l'"espace" qui sous-tend ces exemples diffère de l'espace tridimensionnel habituel. D'abord par l'*anisotropie* qui découle de la résolution perceptive inégale des paramètres d'azimut, élévation ou distance. Ensuite par la *granularité* qui découle des seuils de discrimination élevés (extrême dans le cas d'une simple discrimination gauche/droite, ou présence/absence). Enfin, par la topologie induite par le répertoire des comportements possibles (chute pour le papillon, fuite pour le poisson, orientation puis saut pour la grenouille, orientation de la tête pour la chouette, navigation, approche et capture pour la chauve souris, etc.). La perception servant l'action, la topologie de l'action prévaut en dernière analyse. Cela dit, des *modèles cognitifs*, intermédiaires entre perception et action, peuvent bénéficier d'une topologie plus riche.

4. Métaphores spatiales du son

L'expérience subjective du son est complexe, et l'on utilise souvent des métaphores spatiales pour la comprendre ou la représenter. Le *temps*, dimension inexorable du son, est traduit par l'axe horizontal de la partition musicale, du sonagramme (représentation graphique du spectre en fonction du temps) ou de l'oscillogramme (représentation de l'onde en fonction du temps). Sauf exceptions rares, l'axe temporel est orienté de gauche à droite. On serait tenté d'y voir un universel, si les partitions de tradition chinoise ou japonaise n'étaient orientées verticalement de haut en bas, en colonnes qui s'accumulent de droite à gauche.

Dans de nombreuses langues, les termes "bas" et "haut" servent à comparer deux notes musicales. La *hauteur* tonale est représentée par l'axe vertical de la partition, la fréquence par celui du sonagramme. On pourrait ici aussi penser à un universel, mais il semble qu'il s'agisse d'une convention récente, datant de l'apparition du chant grégorien aux IXe et Xe siècles (Duchez 1989). Les Grecs utilisaient plutôt des termes signifiant "tendu" ou "relâché", conformes à leur expérience des instruments à cordes (Hunt 1990). Pour concilier la dimension linéaire de la hauteur tonale avec la structure circulaire des *chroma* (identité des notes dans chaque octave) on a proposé des représentations en forme de hélice (Shepard 1964, 1982), qui captent tant bien que mal les relations de proximité de note à note, révélées par les techniques d'analyse multidimensionnelle (MDS) (Ueda et Ohgushi 1987). D'autres représentations spatiales ont été proposées, qui

tiennent compte de la consonance ou du rôle de chaque intervalle dans une tonalité (Krumhansl 1990).

Le terme de *volume* a été utilisé pour caractériser une dimension perceptive aujourd'hui tombée en désuétude, car covariante avec une combinaison de sonie et hauteur (Boring 1926). D'autres métaphores spatiales sont d'usage courant, même si le sens de chacun est parfois difficile à cerner: compacité, largeur, profondeur, enveloppement, présence, etc. Le *timbre*, lui, à une définition précise mais pas de connotation spatiale évidente. Des tentatives expérimentales pour cerner sa nature ont abouti à des représentations spatiales (Plomp 1970, Grey, 1977). L'analyse multidimensionnelle des réponses des sujets montrent qu'ils traitent les sons musicaux comme s'ils faisaient partie d'un espace euclidien avec un petit nombre de dimensions (de un à quatre, selon les ensembles de sons). La réalité *psychophysique* (à défaut de *psychologique*) de ces dimensions est étayée par le fait qu'elles peuvent être prédites par des "dimensions physiques" (descripteurs calculés à partir des stimuli) comme le centre de gravité du spectre, le temps d'attaque, etc. Ces dimensions n'ont pas de lien évident avec les dimensions spatiales du monde, il s'agit donc d'un usage métaphorique dont le but est d'aider à comprendre la chose complexe qu'est le timbre.

Le physiologiste est friand de représentations spatiales. Le concept de *tonotopie*, selon lequel une représentation spatiale du spectre est formée dans la cochlée et projetée centralement, date de Du Vernay (1683). Des représentations tonotopiques existent à tous les niveaux depuis la cochlée jusqu'au cortex (même si une part importante de chaque niveau n'est pas organisée tonotopiquement). Pour Helmholtz (1877), le fait d'attribuer chaque composante à un lieu distinct constituait une *explication* de la perception du son. Cette conception a eu une influence profonde et durable, même si elle pose de nombreux problèmes (de Cheveigné 2005). Son attrait tient peut-être à sa nature *spatiale*.

Le noyau central de IC récapitule l'ensemble des traitements du tronc cérébral. On a mis en évidence dans ce noyau des gradients de sensibilité à plusieurs indices acoustiques. Ainsi, sur la surface de chaque plan isofréquence (perpendiculaire à l'axe tonotopique) seraient représentés des indices multiples: ITD, IID, variations fines de fréquence, largeur de bande, période fondamentale d'un son complexe, fréquence de modulation d'amplitude, direction de modulation de fréquence, etc. (Ehret, 1997). Le nombre de ces indices est supérieur au nombre (deux) des dimensions du plan, et il ne peut donc s'agir toutes de "cartes" distinctes. Plus précisément: le profil d'activité en fonction de la position dans le plan est ambigu, et ne suffit pas à la représentation de la valeur d'un indice. On peut alors se demander si la distribution spatiale des cellules sélectives pour un indice participe à la *représentation* de cet indice. Il y a des exemples où cela est le cas: chez la chouette effraie les indices spatiaux ITD et IID ont une répartition elle-même spatiale dans le noyau extérieur ICx, qui projette dans le tectum pour former une carte visuo-auditive d'azimut et élévation. Cet exemple est exceptionnel: chez d'autres espèces, on trouve des représentations spatialement sélectives mais non ordonnées spatialement.

L'*imagerie cérébrale* permet de localiser des fonctions de traitement auditif, voire mettre en évidence des cartes de tonotopie, "périodotopie", "chronotopie",

"timbrotopie", etc. (Drake et al. 1999; Fishman et al. 2000; Zatorre et Krumhansl 2002; Janata et al. 2002). Nous devons nous méfier de deux risques. Le premier est que, par sa nature, l'imagerie joue un rôle de filtre spatial. Elle choisit, parmi les parts d'activité cérébrale, celles dont la répartition est spatiale au détriment d'activités plus diffuses mais potentiellement aussi importantes. Le deuxième risque est que l'usage que nous faisons de métaphores spatiales pour "comprendre le son" nous encourage à accorder du poids aux représentations spatiales fournies par l'imagerie. La découverte d'une représentation spatiale donne au chercheur le sentiment agréable de progresser vers la compréhension. Il faut prendre garde que cette "compréhension" agréable n'en masque une autre plus utile, dans l'hypothèse où les représentations seraient non-spatiales.

5. À quoi cette connaissance est-elle utile?

La perception se jauge à l'aune de l'action qu'elle guide. Prenant la perception comme métaphore de la connaissance, nous pouvons jauger notre savoir à l'aune de ce que nous en tirons d'utile. À quoi peut servir de développer notre connaissance de la cognition spatiale auditive ?

Dans le domaine médical, la connaissance des fonctions et mécanismes guide les stratégies de substitution ou compensation, et permet de concevoir des appareillages qui pallient une insuffisance auditive. Dans le cas d'un handicap visuel, qui force le sujet à s'appuyer particulièrement sur la modalité auditive, elle permet de s'assurer que les indices auditifs nécessaires sont préservés, notamment dans l'hypothèse d'une prothèse par substitution sensorielle.

Dans le domaine architectural, la qualité acoustique a une forte influence sur notre capacité à organiser des scènes sonores complexes. On sait qu'une réverbération excessive nuit à l'intelligibilité de la parole, mais il est intéressant de savoir qu'une dose plus faible lui est propice (Shinn-Cunningham 2002, 2003). La réverbération est un indice utile de distance de la source. Un "effet de salle" bien dosé est essentiel à la qualité d'une salle de concert ou une simulation. La réverbération des espaces clos transforme toute pièce monodique en polyphonie ce qui, selon Sabine (1907), expliquerait que des échelles musicales propres à l'harmonie sont apparues avant la polyphonie.

La spatialisation offre des dimensions musicales supplémentaires dont ont usé par exemple Giovanni Gabrielli au XVI^e siècle et de nombreux compositeurs contemporains, notamment à l'Ircam. L'effet sur le public d'une disposition spatiale particulière dépend des limites perceptives des auditeurs, et des contraintes acoustiques de la situation d'écoute. La connaissance des limites et stratégies perceptives permet au compositeur d'ajuster son écriture pour obtenir une efficacité maximale.

À l'aide d'algorithmes et transducteurs (haut-parleurs ou casque) il est possible de simuler l'effet d'un champ acoustique naturel. Une simulation précise est coûteuse, mais en connaissant les limites de la sensibilité on peut trouver des approximations dont l'erreur est inaudible. Inversement, la connaissance des indices importants pour l'analyse de scène permet de les préserver, voire de les renforcer et

ainsi améliorer la "transparence" de la scène auditive simulée. Ce savoir est utilisé pour le codage (par exemple dans la norme MPEG 4), la diffusion, la réalité virtuelle, les jeux informatiques, l'"auralisation" de données complexes, etc.

Les capacités d'analyse de scène de l'homme ou de l'animal sont inégalées par les systèmes de traitement de signal. Les biosonars des mammifères marins ou aériens dépassent en efficacité les radars ou sonars artificiels (Popper et Fay 1995; Au et al. 2000), et l'homme comprend la parole dans des conditions qui mettent en défaut les systèmes de reconnaissance de la parole (Lippmann 1997; Hermansky 1998). Les mécanismes de traitement de signal biologique, qu'ils soient acoustiques ou neuronaux, sont une source d'inspiration pour les algorithmes. Inversement, les systèmes artificiels servent de modèles pour les processus naturels. Il y a par exemple une forte interaction entre méthodes de traitement de signaux multicapteurs (synthèse d'antennes, "beam-forming", ICA) et modèles de perception binaurale. Autre interaction fructueuse, les méthodes de traitement de signal spatial (réseaux de microphones, etc.) peuvent servir à l'étude des comportements bioacoustiques sur le terrain. À titre d'illustration, la Figure 3 représente sous la forme d'un "réseau conceptuel" quelques-unes des interactions entre domaines d'études et applications.

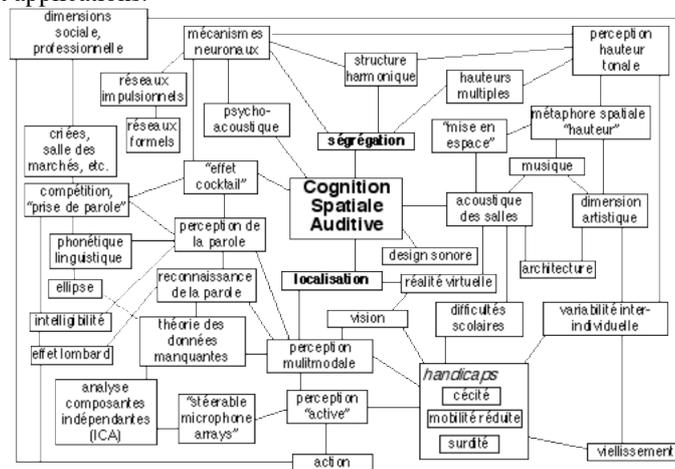


Figure 3. Réseau conceptuel utilisé dans la phase de "brainstorming" d'un projet récent sur la perception spatiale auditive. Chaque rectangle représente une problématique ou application, et les traits représentent des liens de parenté ou de dépendance.

Conclusion

L'audition travaille de concert avec les autres modalités, les connaissances, et l'activité exploratrice, pour guider l'action de l'organisme. Chez l'organisme primitif, les informations auditives sont étroitement associées au contrôle moteur de l'organisme. Chez l'organisme évolué, la dissociation entre perception et l'action permet l'élaboration de modèles prédictifs qui supportent une action plus fine,

moins coûteuse et plus efficace. La connaissance de ces mécanismes de la perception et cognition spatiales peut contribuer à l'efficacité de notre activité d'organismes évolués.

Bibliographie

- Anemüller J (2001) Across-frequency processing in convolutive blind source separation. PhD Thesis Oldenberg.
- Au WWL, Popper AN, and Fay RR (2000) Hearing by whales and dolphins. New York: Springer.
- Aubin T, and Jouventin P (1998) Cocktail-party effect in king penguin colonies. *Proc. R. Soc. Lond. B* 265:1665-1673.
- Barlow HB, and Mollon JD (1989) The senses. Cambridge: Cambridge University Press.
- Blauert J (1997) Spatial hearing. Cambridge, Mass.: MIT Press.
- Boring EG (1926) Auditory theory with special reference to intensity, volume and localization. *Am. J. Psych.* 37:157-188.
- Bregman AS (1990) Auditory scene analysis. Cambridge, Mass.: MIT Press.
- Brown GJ (1992) Computational auditory scene analysis: a representational approach. PhD Thesis Sheffield, Department of Computer Science.
- Carr CE, and Konishi M (1990) A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neuroscience* 10:3227-3246.
- de Cheveigné A (2001a) Fundamental frequency, correlation network, ICA, and hearing. *Proc. CREST workshop, Tokyo.*
- de Cheveigné A (2001b) Correlation Network model of auditory processing. *Proc. Workshop on Consistent & Reliable Acoustic Cues for Sound Analysis, Aalborg (Denmark).*
- de Cheveigné A (2001c) The auditory system as a separation machine. In Breebaart J, Houtsma AJM, Kohlrausch A, Prijs VF and Schoonhoven R (eds) *Physiological and Psychophysical Bases of Auditory Function*. Maastricht, The Netherlands: Shaker Publishing BV, 453-460.
- de Cheveigné A (2005) Pitch perception models. In Plack C and Oxenham A (eds) *Pitch*. New York: Springer Verlag.
- Clarey JC, Barone P, and Imig TJ (1994) Functional organization of sound direction and sound pressure level in primary auditory cortex of the cat. *J. Neurophysiol.* 72:2383-2405.
- Cooke MP (1991) Modeling auditory processing and organisation. PhD Thesis Sheffield, Department of Computer Science.
- Culling JF, Summerfield Q, and Marshall DH (1998) Dichotic pitches as illusions of binaural unmasking I: Huggin's pitch and the "Binaural Edge Pitch". *J. Acoust. Soc. Am.* 103:3509-3526.
- Darwin CJ, and Hukin RW (1999) Auditory objects of attention: the role of interaural time differences. *Journal of Experimental Psychology: Human perception and performance* 25:617-629.
- Darwin CJ, and Hukin RW (2000a) Effectiveness of spatial cues, prosody and talker characteristics in selective attention. *J. Acoust. Soc. Am.* 107:970-977.
- Darwin CJ, and Hukin RW (2000b) Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. *J. Acoust. Soc. Am.* 108:335-342.
- Dooling RJ, Fay RR, and Popper AN (2000) Comparative hearing: birds and reptiles. New York: Springer.

- Drake C, Adenier M, Brochard R, Rivenez M, Ducorps A & Ragot R (1999). La perception du tempo des séquences sonores : une Étude MEG indique une chronotopie. *Cahiers de l'Audition*, 12, 37-38.
- Duchez M-E (1989) La notion musicale d'élément <<porteur de forme>>. Approche épistémologique et historique. In McAdams S and Deliège I (eds) *La musique et les sciences cognitives*. Liège: Pierre Mardaga, 285-303.
- Durlach NI (1963) Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am.* 35:1206-1218.
- Durlach NI, Shinn-Cunningham BG, and Held RM (2003) Supernormal auditory localization. *Presence (MIT)* 2:89-103.
- Ehret G (1997) The auditory midbrain, a "shunting yard" of acoustical information processing. In Ehret G and Romand R (eds) *The central auditory system*. New York: Oxford University Press, 259-316.
- Ellis D (1996) Prediction-driven computational auditory scene analysis. PhD Thesis MIT.
- Falchier A, Clavagner S, Barone P, and Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci.* 22:5749-5759.
- Finlayson PG, and Caspary DM (1991) Low-frequency neurons in the lateral superior olive exhibit phase-sensitive binaural inhibition. *J. Neurophysiol.* 65:598-605.
- Fishman YI, Reser DH, Arezzo JC, and Steinschneider M (2000) Complex tone processing in primary auditory cortex of the awake monkey. II. Pitch versus critical band representation. *J. Acoust. Soc. Am.* 108:247-262.
- Fullard JH (1998) The sensory coevolution of moths and bats. In Hoy RR, Popper AN and Fay RR (eds) *Comparative hearing: insects*. New York: Springer-Verlag, 279-326.
- Fay RR, and Popper AN (1994) *Comparative hearing: mammals*. New York: Springer-Verlag.
- Fay RR, and Popper AN (1999) *Comparative hearing: fish and amphibians*. New York: Springer.
- Feng AS, and Schellart AM (1999) Central auditory processing in fish and amphibians. In Fay RR and Popper AN (eds) *Comparative hearing: fish and amphibians*. New York: Springer, 218-268.
- Grey JM (1977) Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.* 61:1270-1277.
- Grothe B (1994) Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustache bat. *J. Neurophysiol.* 71:706-721.
- Guinan JJ (1996) Physiology of olivocochlear efferents. In P. Dallos, Popper AN and Fay RR (eds) *The cochlea*. New York: Springer Verlag, 435-502.
- Heffner RS, and Heffner HE (1992) Evolution of sound localization in mammals. In Webster DB, Fay RR and Popper AN (eds) *The evolutionary biology of hearing*. New York: Springer-Verlag, 691-715.
- Helfert RH, and Aschoff A (1997) Superior olivary complex and nuclei of the lateral lemniscus. In Ehret G and Romand R (eds) *The central auditory system*. New York: Oxford University Press, 193-258.
- von Helmholtz H (1877) *On the sensations of tone* (English translation A.J. Ellis, 1885, 1954). New York: Dover.
- Hermansky H (1998) Should recognizers have ears? *Speech Comm.* 25:3-27.
- Hoy RR, Popper AN, and Fay RR (1998) *Comparative hearing: insects*. New York: Springer-Verlag.
- Hunt FV (1992 (original: 1978)) *Origins in acoustics*. Woodbury, New York: Acoustical Society of America.
- Janata P, Birk JL, Van Horn JD, Leman M, Tillmann B, and Barucha JJ (2002) The Cortical Topography of Tonal Structures Underlying Western Music. *Science* 298:2167-2170.

- Jeffress LA (1948) A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41:35-39.
- Joris PX, and Yin TCT (1998) Envelope coding in the lateral superior olive. III. Comparison with afferent pathways. *J. Neurophysiol.* 79:253-269.
- Joris PX, Smith PH, and Yin TCT (1998) Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron* 21:1235-1238.
- Klump GM (2000) Sound localization in birds. In Dooling RJ, Fay RR and Popper AN (eds) *Comparative hearing: birds and reptiles*. New York: Springer-Verlag, 249-307.
- Krumhansl KL (1990) *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- Kuwada S, Yin TCT, Haberly LB, and Wickesberg RE (1980) Binaural interaction in the cat inferior colliculus: physiology and anatomy. In Brink Gvd and Bilsen FA (eds) *Psychophysical, physiological and behavioral studies in hearing*. Delft University Press, 401-411.
- Kuwada S, Batra R, Yin TCT, Oliver D, Haberly LB, and Stanford TR (1997) Intracellular recordings in response to monaural and binaural stimulation of neurons in the inferior colliculus of the cat. *J. Neurosci.* 17:7565-7581.
- Lippmann RP (1997) Speech recognition by machines and humans. *Speech Comm.* 22:1-16.
- Lorenzi C, Gatehouse S, and Lever C (1999) Sound localization in noise in normal-hearing listeners. *J. Acoust. Soc. Am.* 105:1810-1820.
- Lyon RF (1983-1988) A computational model of binaural localization and separation. In Richards W (ed) *Natural computation*. Cambridge, Mass: MIT Press, 319-327.
- McAlpine D, Jiang D, and Palmer A (2001) A neural code for low-frequency sound localization in mammals. *Nature Neuroscience* 4:396-401.
- Miall RC, Weir DJ, Wolpert DM, and Stein JF (1993) Is the cerebellum a Smith Predictor? *Journal of Motor Behavior* 25:203-216.
- Michelson A (1998) Biophysics of sound localization in insects. In Hoy RR, Popper AN and Fay RR (eds) *Comparative hearing: insects*. New York: Springer, 18-62.
- Nelken I, Fishbach A, Las L, Ulanovsky N, and Farkas D (2003 (in press)) Primary auditory cortex of cats: feature detection or something else? *Biol. Cybern.*
- Patterson R, Anderson TR, and Francis K (1996) Binaural auditory images and a noise-resistant, binaural auditory spectrogram for speech recognition. *Proc. Workshop on the auditory basis of speech perception, Keele*, 245-252.
- Plomp R (1970) Timbre as a multidimensional attribute of complex tones. In Plomp R and Smoorenburg GF (eds) *Frequency analysis and periodicity detection in hearing*. Leiden: Sijthoff, 397-414.
- Pollack GS (1998) Neural processing of acoustic signals. In Hoy RR, Popper AN and Fay RR (eds) *Comparative hearing: insects*. New York: Springer Verlag, 139-196.
- Popper AN, and Fay RR (1995) *Hearing by bats*. New York: Springer Verlag.
- Popper AN, and Fay RR (1999) The auditory periphery in fishes. In Fay RR and Popper AN (eds) *Comparative hearing: Fish and amphibians*. New York: Springer, 43-100.
- Reznikoff I (1987) Sur la dimension sonore des grottes à peintures du paléolithique. *C.R. Acad. Sci. Paris* 304 - II:153-156.
- Rogers P (1986) What do fish listen to? (abstract). *J. Acoust. Soc. Am.* 79:S22.
- Sabine WC (1907) Melody and the origin of the musical scale. In Hunt FV (ed) *Collected papers on acoustics by Wallace Clement Sabine (1964)*. New York: Dover, 107-116.
- Sahey TL, Nodar RH, and Musiek FE (1997) Efferent auditory system. San Diego: Singular.
- Schwartz IR (1992) The superior olivary complex and lateral lemniscal nuclei. In Webster DB, Popper AN and Fay RR (eds) *The mammalian auditory pathway: neuroanatomy*. New York: Springer-Verlag, 117-167.

- Shackleton TM, Meddis R, and Hewitt MJ (1994) The role of binaural and fundamental frequency difference cues in the identification of concurrently presented vowels. *QJEP A* 47:545-563.
- Shepard RN (1964) Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.* 36:2346-2353.
- Shepard RN (1982) Structural representations of musical pitch. In Deutsch D (ed) *The psychology of music*. Orlando: Academic Press, 343-390.
- Shinn-Cunningham BG (2002) Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays. *Proc. ICAD 2002, Kyoto*, 419-423.
- Shinn-Cunningham BG, and Ram S (2003) Identifying where you are in room: Sensitivity to room acoustics. *Proc. ICAD, Boston*, 21-24.
- Smith PH, Joris PX, and Yin TCT (1993) Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: evidence for delay lines to the medial superior olive. *J. Comp. Neurol.* 331:245-260.
- Smith PH, Joris PX, and Yin TCT (1998) Anatomy and physiology of principal cells of the medial nucleus of the trapezoid body (MNTB) of the cat. *J. Neurophysiol.* 79:3127-3142.
- Szentágothai J, and Arbib MA (1975) *Conceptual Models of Neural Organization*. Cambridge, MA.: The MIT Press.
- Thompson SP (1882) On the function of the two ears in the perception of space. *Phil. Mag.* (S5) 13:406-416.
- Truax B (1984) *Acoustic communication*. Norwood, N.J.: Ablex.
- Ueda K, and Ohgushi K (1987) Perceptual components of pitch: spatial representation using a multidimensional scaling technique. *J. Acoust. Soc. Am.* 82:1193-1200.
- Du Verney JG (1683) *Traité de l'organe de l'ouïe, contenant la structure, les usages et les maladies de toutes les parties de l'oreille*. Paris.
- Warr WB (1992) Organization of olivocochlear efferent systems in mammals. In Webster DB, Popper AN and Fay RR (eds) *The mammalian auditory pathway: neuroanatomy*. New York: Springer Verlag, 410-448.
- Yin TCT, and Chan JCK (1990) Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* 64:465-488.
- Yin TCT, Chan JCK, and Carney LH (1987) Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. III. Evidence for cross-correlation. *J. Neurophysiol.* 58:562-583.
- Zatorre RJ, and Krumhansl KL (2002) Mental models and musical minds. *Science* 298