

Modèles de traitement auditif dans le domaine temps

Alain de Cheveigné

29 avril 2000

Mémoire d'habilitation à diriger des recherches
Université Paris 6
Neurosciences

The Mind, that Ocean where each kind,
Does streight its own resemblance find ;
Yet it creates, transcending these,
Far other Worlds, and other Seas,
Annihilating all that's made
To a green Thought, in a green Shade.
Andrew Marvell (1621 - 1678)

Table des matières

Remerciements	5
Avertissement	7
Résumé	9
1 Introduction	11
2 Le système auditif	13
2.1 Le système auditif et son environnement	13
2.2 Systèmes périphérique et central	15
2.3 Système auditif périphérique	16
2.3.1 Innervation des cellules ciliées	23
2.3.2 Transduction	24
2.4 Le système auditif central	26
2.4.1 Une hiérarchie de niveaux	26
2.4.2 Subdivisions	28
2.4.3 Structure bilatérale	32
2.4.4 Parallélisme et sauts “trans-hiérarchiques”	34
2.4.5 Voies descendantes	36
2.4.6 Autres facteurs de complexité	40
2.5 Conclusion	43
3 Tonotopie et traitement temporel	45
3.1 La tonotopie dans le système auditif	45
3.1.1 Répartition de la tonotopie	47
3.1.2 Spéculations sur le sens de la tonotopie	51
3.2 L’information temporelle	52
3.2.1 Répartition dans le système auditif	52
3.2.2 Spécialisations en faveur du traitement temporel	58
3.2.3 Modèles	61
3.2.4 Vers des modèles plus complexes ?	67
3.3 Conclusion	69
4 Enjeux théoriques en Audition	71
4.1 Dimensions physiques, qualités sensorielles	71
4.2 Le rôle privilégié de la Hauteur Tonale	73

5	Séparation à base de F_0	79
5.1	La conjecture d'annulation harmonique	79
5.2	Expériences "voyelles doubles"	80
5.3	Filtre d'annulation neuronal	85
6	Perception de la hauteur	87
6.1	Modèle "annulation" de perception de la hauteur	87
6.2	Modèle de perception des hauteurs multiples	88
6.3	Modèles de décalage de hauteur des partiels désaccordés	89
6.4	Modèle de l'asymétrie entre pics et creux de modulation de fréquence	91
6.5	Modèle de discrimination des tons purs	92
7	Perception des voyelles	93
7.1	Identification des voyelles concurrentes	93
7.2	Modèle "données manquantes" du timbre des voyelles	95
8	Applications	97
8.1	Estimation de la F_0 de la parole	97
8.2	Séparation des sources	99
8.3	Harmonicité et estimation spectrale	100
8.3.1	Modèle source-filtre et échantillonnage	100
8.3.2	Lissage et échantillonnage	101
8.3.3	Périodicité et lissage optimal	103
9	Conclusion	107
	Annexe - Les limites de l'Analyse de Fourier	109
9.1	Systèmes linéaires invariants	109
9.1.1	Un descripteur de la périodicité?	111
9.2	Unicité et cohérence de la description	112
9.3	Conclusion	113
	Liste des sigles	115
	Liste des publications	119
	Bibliographie générale	125
	Tirés à part	149

Remerciements

De nombreuses personnes ont contribué à cet effort, parfois à leur insu. L'activité scientifique se distingue d'autres activités créatrices par l'importance que revêt la *transmission* du savoir. François Raveau m'a appris l'importance de la bibliographie, et Antoine Culioli celle de savoir terminer une thèse. Shuji Doshita m'a poussé à soumettre ma première communication, et de Yoh'Ichi Tohkura, j'ai appris la recette pour faire accepter un article : persévérer. Quentin Summerfield, reviewer du premier article ainsi accepté, m'a transmis son goût de la clarté.

“Rien n'est jamais acquis à l'homme, ni sa force, ni sa faiblesse. . .”. Je suis particulièrement reconnaissant à ceux qui ont fait preuve, tantôt de tolérance, tantôt d'exigence à des moments critiques. François Raveau et Shuji Doshita, déjà cités, mais aussi Jean-Sylvain Liénard et Charles Berthommier, dont la pièce d'angle dans une tour de Jussieu est une étape dans le destin de nombreuses personnes.

Mon intérêt pour l'Audition date d'une visite au Laboratoire de Adrian Fourcin, à Londres, en compagnie de Georges Boulakia. Fourcin avait inventé un casque muni de microphones sur de longues tiges (pour créer des retards interauraux supérieurs aux valeurs physiologiques), dont il s'est servi pour découvrir le phénomène de hauteur binaurale qui porte son nom. Ce dispositif biscornu a frappé mon imagination, et la lecture du *tiré à part* m'a donné envie d'expliquer un jour le phénomène. Bien plus tard, lors d'un long séjour à Kyoto, Kengo Ohgushi m'a nourri de photocopies d'articles sur la hauteur, tandis que Laurent Demany prêtait un sage scepticisme aux premières idées de modèle que je lui envoyais. Steve McAdams m'a initié, non sans mal, aux métiers de la psychoacoustique, et encouragé à passer de la hauteur à l'organisation perceptive. Les comparses de mes expériences ont été Jean Laroche, Muriel Rosenberg et Cécile Marin en France, et Hideki Kawahara, Minoru Tsuzaki, Kiyooki Aikawa et Rieko Kubo au Japon. John Culling a généreusement prêté son logiciel |WAVE indispensable à la synthèse. J'ai aussi bénéficié des logiciels de John Holdsworth et Roy Patterson, Lowel O'Mard et Ray Meddis, et Malcolm Slaney. Mon premier article s'est appuyé sur des données psychoacoustiques inédites de Andrew Lea, des enregistrements physiologiques prêtés par Alan Palmer, et un très utile théorème démontré par Jean Laroche.

Depuis le début, la lumière de spectacle a été le concurrent et le contrepoint de la recherche. La notion d'annulation, importante dans mes modèles, est la transposition de celle d'ombre, dont j'ai fait ma spécialité en lumière. Feu Hideki Yano a été mon maître dans ce domaine, et plus récemment Mark Tompkins a ressuscité l'éclairagiste après un long sommeil pour quelques aventures imprévues et inespérées.

Le Japon a joué un grand rôle dans mon cheminement. J'ai cité Shuji Doshita, qui a gentiment accueilli un gaijin ignorant, et Yoh'Ichi Tohkura, architecte des laboratoires ATR. C'est Kiyohiro Shikano qui m'a invité et accueilli durant mon premier séjour à ATR, pendant lequel j'ai bénéficié de la gentillesse de Hisao Kuwabara. Plus tard, Hideki Kawahara m'a fait profiter de son bouillonnement d'idées

et son enthousiasme presque enfantin. À chaque retour à ATR je prends plaisir à retrouver l'équipe : Minoru Tsuzaki, Erik McDermott, Eric Bateson, Alain Biem, Satoru Fujita, Hiroaki Kato, Hiroko Kato, et bien d'autres.

Avertissement

L'Habilitation à Diriger des Recherches est une invention récente, dont la forme est moins solidement enracinée que celle de la Thèse. Les précieux conseils de Umberto Eco (1995) sur la façon d'écrire une thèse sont d'un secours limité pour cerner les contours d'un mémoire d'habilitation. Le mémoire résume une activité dont la substance est éparpillée dans des articles, rapports ou communications. Un rapide travail de reliure est une option, mais on devine que le résultat serait indigeste. Les articles ont chacun leur Introduction, Argumentation, Revue de la Littérature, etc., ce qui ne favorise pas la cohérence de leur juxtaposition. Le problème est d'autant plus aigu que leur contenu, accumulé au fil d'un parcours scientifique, est hétéroclite. Individuellement, chaque article est (on l'espère) clair et lisible, mais collectivement ils représentent un dédale pour le lecteur.

Néanmoins il est hors de question de reprendre l'ensemble de cette matière dans un nouveau texte. Un mémoire d'habilitation n'est dit-on ni une thèse, ni un livre. L'énergie requise, à supposer qu'elle soit disponible, est mieux investie dans d'autres oeuvres. Il y a donc un compromis à trouver, une forme à inventer, une structure à créer dans laquelle les articles s'insèrent sans trop d'interférence ou de redondance. Il ne suffit pas de doubler chaque article d'un maigre résumé. Le lecteur ne devrait pas avoir à se plonger dans les articles pour trouver matière à comprendre, et en deçà d'un certain seuil un texte devient incompréhensible faute d'un développement suffisant des idées.

Ce mémoire représente un essai pour maîtriser cette "forme" qu'est le mémoire d'habilitation. Il comprend, outre une Introduction, une paire de chapitres consacrés à la description du système auditif, par "approximations successives". En partant d'une perspective macroscopique, la description progresse vers la complexité et le détail, pour ensuite se focaliser sur tout ce qui peut conforter (ou au contraire infirmer) les modèles décrits dans les articles. Ces chapitres servent à la fois de *revue bibliographique*, d'*examen critique* de la plausibilité des modèles, et de *contexte* physiologique et anatomique pour préciser leur description.

Comparés à ces chapitres descriptifs, les chapitres qui introduisent les articles sont succints, la matière étant contenue dans les articles eux-mêmes. De cet effort résulte une construction à l'équilibre incertain, dans laquelle mes résultats originaux occupent une part succinte, et la revue et la spéculation une part plus ample. J'espère que le lecteur verra dans ce déséquilibre un "contrepoint", et dans l'ensemble un tableau alliant lumière et ombre, enseignements et interrogations.

Résumé

Ce travail porte sur trois problèmes en Audition : la *séparation des sources*, la *hauteur tonale*, et le *timbre des voyelles*. Un fil les relie : le rôle joué par la périodicité, ou harmonicité de certaines sources sonores.

Il s'appuie sur trois conjectures. La première est que la *structure temporelle* fine des décharges du nerf auditif joue un rôle dans le traitement auditif, contrairement à la doctrine classique qui assimile la cochlée à un pur analyseur de Fourier. La deuxième est que le système auditif opère, dans certains cas, par *annulation*. L'annulation des corrélats d'une source peut permettre de la caractériser (les paramètres qui donnent la meilleure annulation reflètent sa structure). Elle facilite aussi la perception d'autres sources, après suppression de la première. La troisième conjecture est que le système auditif accorde un *poids variable* à différentes dimensions ou portions de données en fonction de leur fiabilité. Ces conjectures ont servi de guide à l'élaboration des modèles, mais elles ne sont pas la cible directe des modèles et expériences, et elles seront défendues sans dogmatisme.

Ce travail fait appel à trois méthodologies. La première est la *modélisation* : construction de modèles théoriques pouvant expliquer les données de physiologie et de psychoacoustique (souvent, mais ce n'est pas obligatoire, avec simulation sur ordinateur). La deuxième est l'*expérimentation psychoacoustique* : des expériences ont été conçues pour tester des hypothèses, départager les modèles, ou simplement en apprendre plus sur les phénomènes. La troisième est l'*application* de principes issus des modèles à des problèmes pratiques d'ingénierie tels que le traitement de la parole. En dehors de l'utilité éventuelle des résultats, cette démarche permet de jauger la valeur pragmatique des principes utilisés par les modèles.

De la modélisation est issue l'hypothèse d'*annulation harmonique*, que l'expérimentation a confirmé. En présence de sons masquants harmoniques, le système auditif utilise leur structure harmonique pour réduire le masquage. L'hypothèse rivale de *renforcement harmonique* n'a pas été confortée par l'expérience. De nombreux modèles ont été proposés dans la littérature, dont les plus récents sont le modèle de battements de Culling et Darwin (1993) et celui à base de sélection de canaux de Meddis et Hewitt (1992). Les expériences ont permis de les écarter au profit d'un nouveau modèle faisant intervenir un filtrage stochastique "neuronal" dans le domaine temps[14, 12]. Elles ont fourni d'autres renseignements : La ségrégation fonctionne lorsque la voyelle cible est faible (-25 dB), et lorsque les fréquences fondamentales (F_0) de la cible et du fond diffèrent d'à peine 0.4%. La modulation de fréquence (vibrato) ne joue pas de rôle direct, contrairement à ce que laissaient espérer les principes d'Analyse de Scènes Auditives (Bregman, 1990). La largeur de

bande des formants a un effet très marqué sur la ségrégation, alors qu'on sait depuis longtemps qu'elle n'affecte guère leur timbre ou identité. En revanche le spectre de phase (intra-voyelle et inter-voyelles) a très peu d'effet, etc..

Parallèlement à l'expérimentation, un ensemble de modèles a été développé pour en expliquer les résultats, ainsi que d'autres phénomènes apparentés observés en psychoacoustique. Parmi ces modèles :

- Un "filtre d'annulation neuronal", fonctionnant comme un détecteur d'anticoïncidences entre impulsions nerveuses[14].
- Un modèle d'identification de voyelles concurrentes fondé sur le précédent filtre[12].
- Un modèle de perception de la hauteur tonale, fondé lui aussi sur le filtre d'annulation neuronal[8].
- Un modèle de perception de hauteurs multiples[6].
- Un modèle qui explique les décalages de hauteur des partiels désaccordés décrits par Hartmann et Doty (1996)[9, 3].
- Un modèle de perception des voyelles basé sur la notion de "données manquantes". Ce modèle rend compte de la quasi-invariance du timbre des voyelles par rapport aux changements de fréquence fondamentale[5].

Ces modèles diffèrent par les phénomènes qu'ils visent et par le degré de détail de leurs hypothèses. Ils ont pour ambition de rendre compte des données expérimentales psychoacoustiques, tout en étant plausibles du point de vue de la physiologie du système auditif.

Enfin, parmi les applications abordées figurent l'*estimation de la fréquence fondamentale* de la parole (F_0), l'*estimation des F_0 de sources multiples concurrentes*, la *suppression de voix parasites* pour la reconnaissance de la parole, et l'*estimation d'enveloppe spectrale*.

1 Introduction

L' Audition est, avec la Vision, le Toucher, le Goût et l'Odorat, l'un des cinq sens qui nous permettent d'appréhender le monde. À ces sens classiques selon Aristote, il convient d'ajouter des sens somesthésiques qui renseignent sur la position du corps, la tension des muscles, ou l'état des viscères (Buser et Imbert, 1982). Les ondes acoustiques issues d'une source sont traduites en impulsions nerveuses par le système auditif périphérique, après modification par les propriétés acoustiques de l'oreille externe, de la chaîne des osselets de l'oreille moyenne, et du système actif que forment les fluides et structures de l'oreille interne. Les modulations des décharges du nerf auditif sont relayées vers le cortex par une succession de noyaux dont seuls les plus périphériques sont exclusivement spécialisés pour l'Audition, alors que les plus centraux intègrent des apports d'autres modalités.

Le système périphérique est assez bien connu, même si des données récentes suscitent de nouvelles interrogations sur le rôle des mécanismes cochléaires actifs, et le contrôle descendant qui s'exerce sur eux. Le système auditif central reste plus obscur du fait de son extrême complexité et des difficultés techniques pour sonder sa structure et son fonctionnement. Une partie de ces difficultés vient d'un manque de *modèles* de traitement auditif pour guider les études neuroanatomique ou neurophysiologiques et donner un sens aux données obtenues. L'objet du présent travail est de proposer quelques nouveaux modèles de traitement pour des tâches telles que la perception de la hauteur ou du timbre, ou la ségrégation perceptive de sources concurrentes.

Ces modèles ont la particularité de faire une part importante au traitement de la structure *temporelle* des patterns nerveux traduits par la cochlée et relayés par les niveaux nerveux plus centraux. Une place relativement moins importante est faite à la structure *spectrale* de ces informations. Ce travail hérite donc de la vieille controverse, datant de Ohm, Helmholtz et Seebeck, entre les tenants d'une représentation spectrale en termes d'activité en fonction de la position dans la cochlée (théorie "place"), et les tenants d'un codage par structure temporelle des décharges dans le nerf auditif (théorie "temps"). Cette controverse domine l'histoire de l'Audition depuis plus de 120 ans avec une virulence parfois stérile à laquelle j'espère échapper. Dans la Section 4 je parle de certains aspects de cette controverse, dont la compréhension est nécessaire à quiconque veut appréhender les enjeux théoriques de l'Audition.

Les phénomènes que j'ai abordés sont la *hauteur tonale*, en particulier la perception des *hauteurs de sons concurrents* (par exemple des ensembles d'instruments), la *ségrégation des voix* sur la base de leur structure har-

monique (un aspect du fameux “effet cocktail”, traduction de “cocktail party effect”, Cherry, 1953), et la *perception du timbre* des voyelles isolées ou concurrentes. Un lien existe entre ces phénomènes en apparence disparates, en ce qu’ils dépendent tous de la *structure périodique* (harmonique) du stimulus, et que leur traitement pourrait donc relever de mécanismes semblables. Le champ couvert est néanmoins divers, et un défi que doit relever ce document est de donner une cohérence à un ensemble hétéroclite de modèles et résultats expérimentaux.

Du point de vue méthodologique, je me suis appuyé sur l’*expérimentation psychoacoustique* et la *modélisation*, avec simulation informatique des modèles. J’ai aussi abordé quelques *applications*. Les résultats sont décrits dans une série d’articles qui sont introduits dans les Sections 5 (ségrégation), 6 (hauteur tonale), 7 (voyelles) et 8 (applications en traitement de la parole).

Ce travail se situe à la confluence entre trois domaines : la *psychoacoustique*, la *modélisation* et les *neurosciences*. J’ai pratiqué et contribué aux deux premiers, tout en m’appuyant sur les résultats du troisième. Les enjeux expérimentaux, théoriques et de modélisation sont traités dans les articles et les chapitres qui les introduisent, mais il manque à ces articles les données neuroanatomiques et neurophysiologiques qui seraient utiles à la compréhension de ces hypothèses. C’est pourquoi les deux sections sur la structure du système auditif (Sections 2 et 3) occupent paradoxalement la majeure partie de ce document. La Section 2 est une tentative de description du système auditif, qui cherche à donner une idée précise des différents niveaux de complexité tout en restant compréhensible. J’ai expérimenté pour cela une technique de description par *approximations successives*, différente des descriptions habituelles par niveaux ou ensembles anatomiques. La description est générale, et en même temps résolument orientée vers les indices qui peuvent confirmer ou infirmer mes modèles. Ces indices (tonotopie et traitement temporel) sont ensuite développés dans la Section 3.

La description du système auditif est une entreprise difficile et imprudente pour un non-biologiste. J’espère néanmoins avoir atteint les objectifs de (a) fournir un contexte anatomique et physiologique général aux modèles, (b) esquisser les indices en leur faveur (ou défaveur), (c) situer les structures qui pourraient les héberger, et (d) réunir un ensemble de références bibliographiques pertinentes.

2 Le système auditif

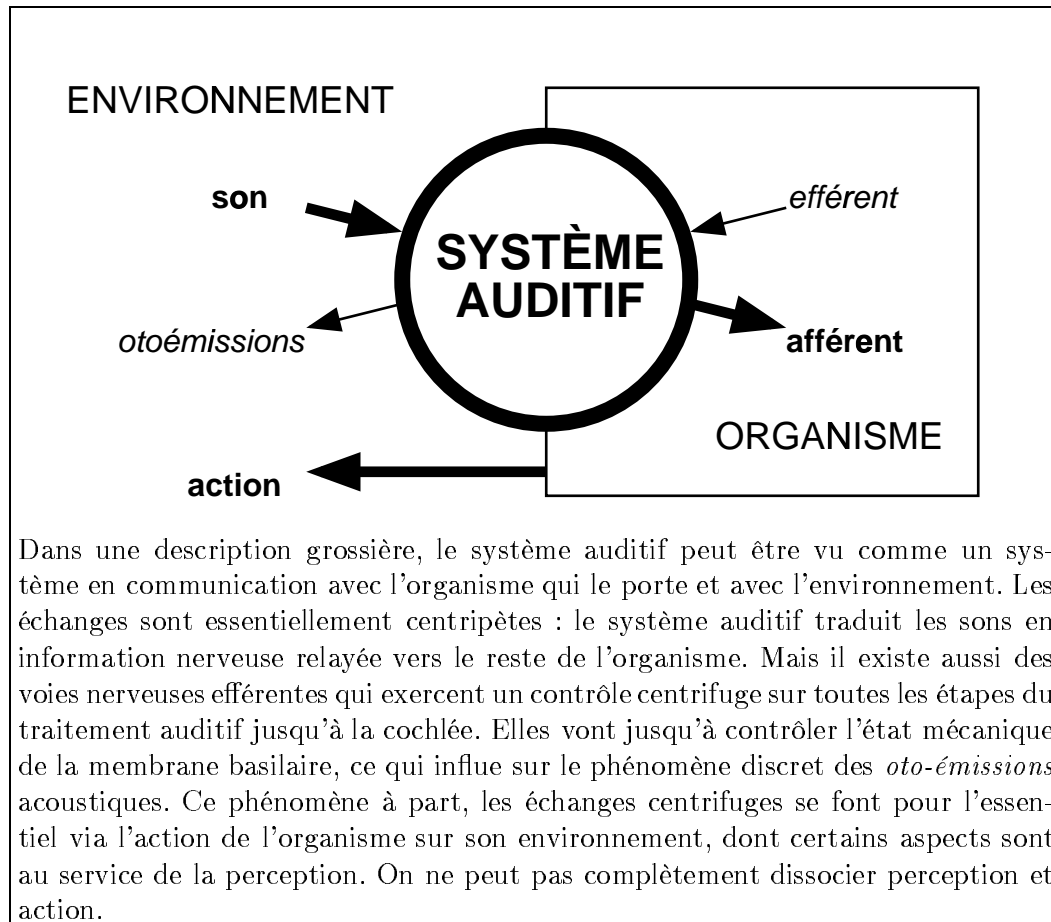
J'entreprends dans ce chapitre une description du système auditif. L'objectif est de fournir un contexte aux modèles "neuronaux" présentés plus loin, de diriger le lecteur vers les sources où il pourra satisfaire sa curiosité de détails plus approfondis, et de signaler quelques fronts d'évolution rapide de notre compréhension de ce système. Il s'agit d'un système d'une grande complexité, dont la compréhension est difficile quelle que soit manière qu'on s'y prenne pour la décrire. J'ai opté pour une démarche d'*approximation successive*, qui, partant d'une perspective macroscopique la moins détaillée possible, révèle l'un après l'autre des niveaux de détail progressivement plus complexes. Les chapitres suivants se focaliseront ensuite sur des détails plus précis.

2.1 Le système auditif et son environnement

Dans une description la moins détaillée possible, le système auditif peut être vu comme un ensemble insécable, une "boite noire", qu'une première frontière sépare de l'organisme qui le porte, et une deuxième de l'environnement.

L'oreille est un organe de perception sensible aux vibrations transmises par l'air ou, dans certains cas, conduites à travers les tissus du corps. Sa sensibilité extraordinaire est décrite en détail dans des ouvrages de base comme Moore (1997), Buser et Imbert (1987), ou Hartmann (1997). Retenons que la gamme de fréquences perçues par l'homme est de 16 à 16000 Hz, soit environ 10 octaves, et que la sensibilité en pression au seuil est de $20 \mu Pa$. La petitesse de cette valeur est difficile à imaginer : elle représente environ 10^{-10} fois la pression atmosphérique, et correspond à une onde plane dans l'air qui transmettrait une puissance d'à peine $10^{-12} W/m^2$ (dont une fraction, à peine le dix-millième, entrerait dans le canal de l'oreille...). Le déplacement de l'air au seuil de perception est environ $10^{-11} m$, soit moins que le rayon d'un atome d'hydrogène. La vitesse est de l'ordre de $5 \times 10^{-8} m/s$, soit 10^{10} fois moins que la vitesse cinétique des molécules de l'air à température ambiante (Hartmann, 1995). À souligner enfin, la gamme de fonctionnement (entre seuil de perception et seuil de douleur) d'environ 120 dB (un rapport de pression de 10^6 , ou de puissance de 10^{12}).

À côté du rôle passif qui lui est largement reconnu et décrit, on peut trouver à l'oreille un rôle *actif*. Passons sur l'usage de la motilité des pavillons chez certains animaux (dont quelques humains) comme moyen d'expression ou de communication visuelle (Darwin, 1872) ou pour chasser les mouches. La motilité des pavillons sert surtout à modifier les propriétés acoustiques



de l'oreille pour améliorer la réception d'une source désirée, ou atténuer une source non-désirée. Il s'agit d'une forme d'"écoute active", mais il en est d'autres : l'individu peut tourner la tête, déplacer son corps, tourner un bouton pour monter le son, chausser un stéthoscope, voire tapoter une cloche pour "voir comment elle sonne". Pour peu qu'on admette ces exemples, la notion d'écoute active a des limites très larges.

Dans cette même logique active, mais sans aller si loin, le système auditif efférent modifie la tension des os de l'oreille moyenne en contractant les muscles stapédius et tensor tympani, ce qui atténue la transmission dans les basses fréquences et modifie quelque peu la réponse fréquentielle (Møller, 1983). Cette contraction (réflexe stapédien) peut être déclenchée par un bruit intense, mais sa latence de (25-35 ms) est malheureusement trop longue pour protéger l'oreille d'un bruit impulsionnel. Elle est aussi déclenchée par divers stimuli non-auditifs (Buser et Imbert, 1987) et lors de la phonation, atténuant ainsi la voix du sujet dans ses propres oreilles. À un niveau légèrement

plus central, les cellules ciliées externes de l'oreille interne, qui participent à la mécanique cochléaire, sont sous contrôle nerveux efférent.

Ces considérations sur l'innervation efférente de l'oreille interne seraient déplacées dans une description de l'oreille comme une "boîte noire", si les mécanismes en question n'avaient pas des effets acoustiques mesurables, et parfois même audibles, à l'extérieur de l'oreille. Depuis Kemp (1978) on sait que l'oreille peut émettre de l'énergie acoustique après un stimulus, sous la forme d'un "écho auditif" (oto-émission évoquée), dont la latence est trop grande pour qu'on puisse l'attribuer à une simple réflexion acoustique par les structures de l'oreille. Depuis Zurek (1981) on sait que l'émission peut aussi être spontanée (oto-émission spontanée), sous la forme d'un petit nombre de composantes à bande étroite. Leur niveau dans le conduit auditif externe est généralement très faible, mesurable uniquement à l'aide de techniques de moyennage, mais des cas ont été rapportés d'oreilles dont le sifflement était audible à proximité (Klinke, 1986). L'émission est parfois audible par le sujet sous forme de tinnitus, mais le plus souvent elle n'est pas perceptible, et n'interfère avec l'audition que près du seuil (Zwicker, 1986). Inversement le tinnitus, ou "sifflement d'oreille", n'est pas nécessairement lié à une oto-émission évoquée (Moore, 1997). Son origine reste aussi mystérieux aujourd'hui qu'il y a trois siècles :

Le tintement des oreilles est un symptôme dont les causes sont difficiles à expliquer. (Du Vernay, 1683, p192)

Une belle expérience, rapportée par l'équipe de Montpellier (Puel et al., 1988), a montré que les oto-émissions évoquées sont modifiées au cours de tâches cognitives. Cela implique que l'état cognitif au niveau le plus central influe sur les propriétés mécaniques au niveau le plus périphérique, allant jusqu'à modifier l'environnement acoustique !

Pour résumer, le rôle le plus évident du système auditif est celui d'un récepteur acoustique d'une sensibilité exquise. Son rôle productif (oto-émissions) est plus discret, mais trahit les mécanismes actifs en jeu au niveau neuronal et sensitif, et sert à rappeler que la perception n'est pas entièrement passive, ni dissociable de l'activité de l'organisme.

2.2 Systèmes périphérique et central

Dans un premier pas vers la complexité, nous scindons le système auditif en *deux* boîtes noires. La première, périphérique, comprend l'oreille externe, moyenne et interne. La seconde, centrale, comprend l'ensemble du système

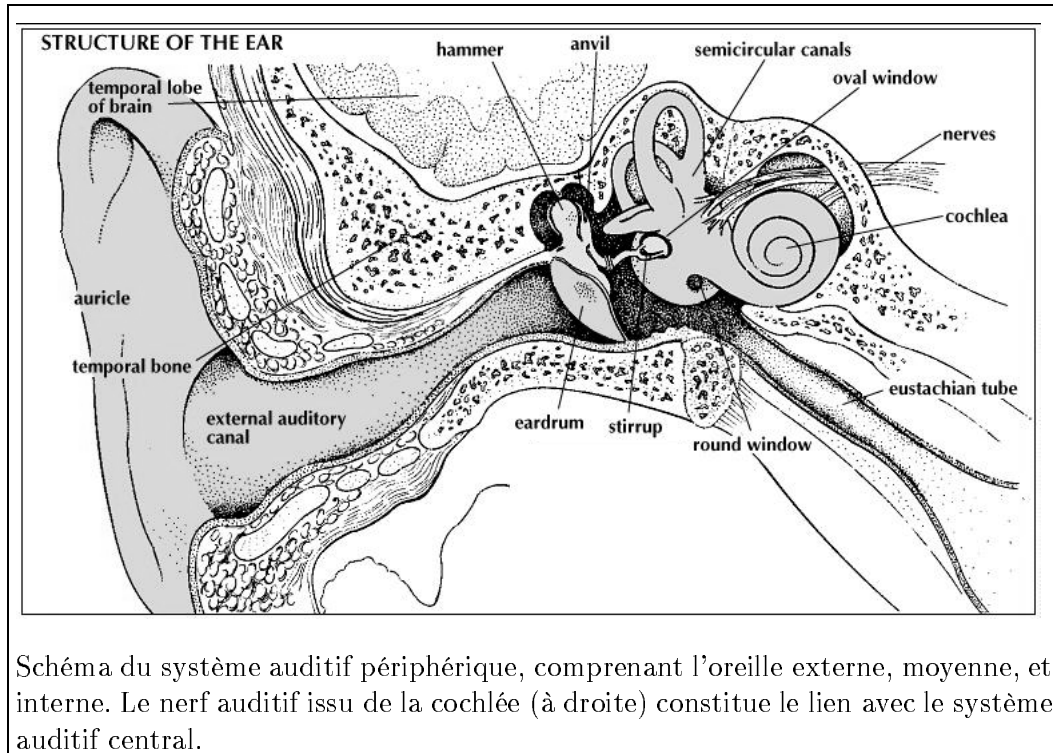


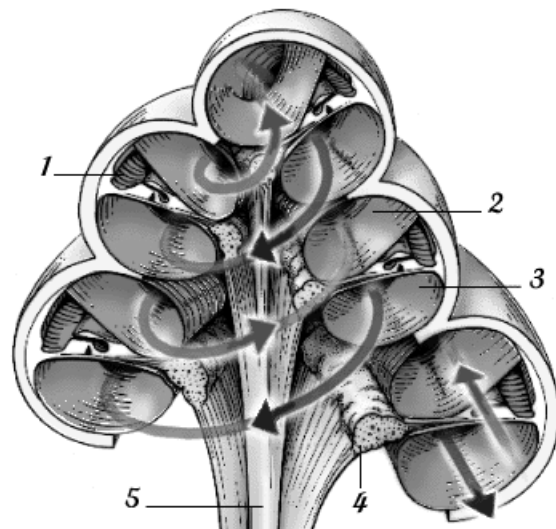
Schéma du système auditif périphérique, comprenant l'oreille externe, moyenne, et interne. Le nerf auditif issu de la cochlée (à droite) constitue le lien avec le système auditif central.

nerveux auditif, du ganglion spiral jusqu'au cortex. L'accent ici est sur l'interface entre systèmes périphérique et central, et la nature des informations échangées entre les deux. Cependant, pour que la présentation soit compréhensible, il nous faut transgresser notre convention et regarder brièvement l'intérieur de la première boîte noire, le système périphérique. La description de la deuxième boîte, le système auditif central, sera développée dans les sections suivantes.

2.3 Système auditif périphérique

Le système auditif périphérique est décrit en détail dans de nombreux ouvrages de base (Kahle et al., 1981 ; Møller, 1983 ; Buser et Imbert, 1987 ; Pickles, 1988 ; Yates, 1995 ; Moore, 1997). Les ondes acoustiques incidentes sont légèrement filtrées par réfraction sur le torse, la tête et le pavillon de l'oreille externe, filtrage qui dépend de l'azimut et de l'élévation de la source. Les ondes sont transmises le long du conduit auditif externe vers le tympan, qui met en vibration la chaîne des osselets de l'oreille moyenne, dont le dernier (étrier, ou stapédius) s'insère dans la fenêtre ovale de la cochlée.

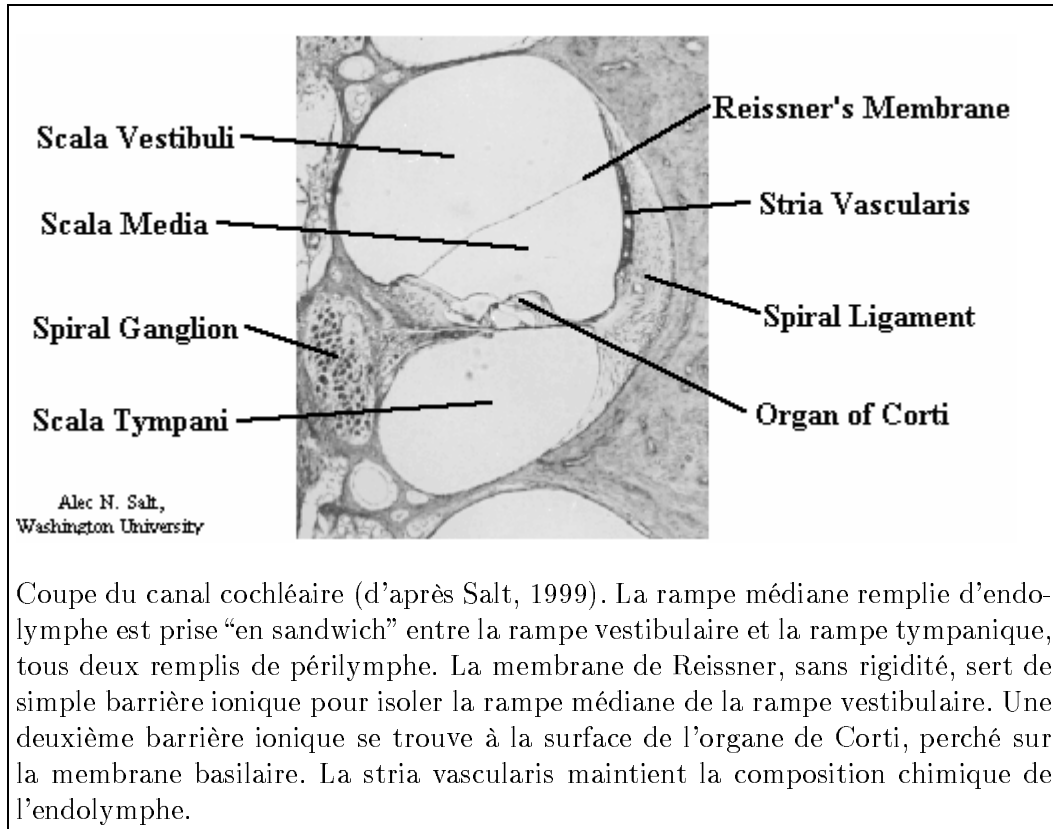
La cochlée est un tunnel de forme spirale creusé dans l'os temporal, composé de deux rampes, l'une montante (*rampe vestibulaire*) et l'autre descen-



Section axiale de la cochlée (d'après Pujol et al., 1999). Le sens des flèches permet de distinguer la rampe vestibulaire (flèche montante) de la rampe tympanique (flèche descendante). Ces rampes communiquent au sommet par l'hélicotrème. (1) canal cochléaire, (2) rampe vestibulaire, (3) rampe tympanique, (4) ganglion spiral, (5) nerf cochléaire.

dante (*rampe tympanique*). Sur presque toute leur longueur les rampes sont séparées par la lame spirale (osseuse) que prolonge la *membrane basilaire* (fibreuse). Au sommet de la cochlée, lame spirale et membrane basilaire s'effacent et les deux rampes communiquent par un trou appelé *hélicotrème*. A la base, la rampe montante est fermée par la fenêtré ovale (sur laquelle s'insère le stapédius) et la rampe descendante par la fenêtré ronde. Toutes deux donnent sur l'oreille moyenne.

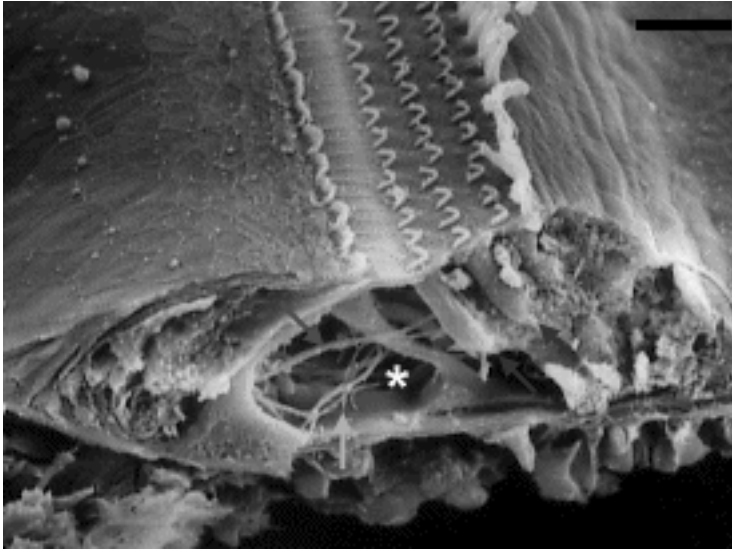
Les deux rampes sont remplies du même liquide, le *périmplympe*, à l'exception d'une poche (*scala media*) qui occupe une partie de la rampe vestibulaire immédiatement supérieure à la membrane basilaire. Elle est remplie d'un liquide de nature chimique différente, l'*endolympe*. Cette poche (de forme spirale comme le reste) est séparée de la rampe vestibulaire *stricto sensu* par une membrane souple et fine, la membrane de Reissner. Cette membrane n'a pas de rôle mécanique, elle sert simplement de barrière ionique entre les deux liquides afin que la membrane basilaire soit baignée sur une face de périmplympe, et sur l'autre d'endolympe. Le gradient électrochimique entre ces deux liquides est en effet la "pile" de l'oreille, la source d'énergie dans laquelle puisent les cellules sensibles de la cochlée, les *cellules ciliées*. Cette "pile" est rechargée par la stria vascularis, une bande de cellules fortement



vascularisée qui maintient la composition chimique de l'endolymphe.

Les cellules ciliées sont insérées dans l'organe de Corti, sorte de promontoire fait de cellules de support, qui chevauche la membrane basilaire à proximité de l'axe de la cochlée. Elles forment deux groupes. Le groupe le plus proche de l'axe comprend les *cellules ciliées internes* (CCI) en une rangée unique. Le groupe le plus éloigné de l'axe comprend trois à cinq rangées de *cellules ciliées externes* (CCE). Ces deux sortes de cellules ciliées diffèrent par leur morphologie, leur innervation, et leur rôle présumé. Les CCI sont innervées de façon massivement afférente, et ont un rôle sensitif au sens propre. Les CCE sont innervés de façon efférente, et, aussi étrange que cela puisse paraître, leur rôle est *moteur*.

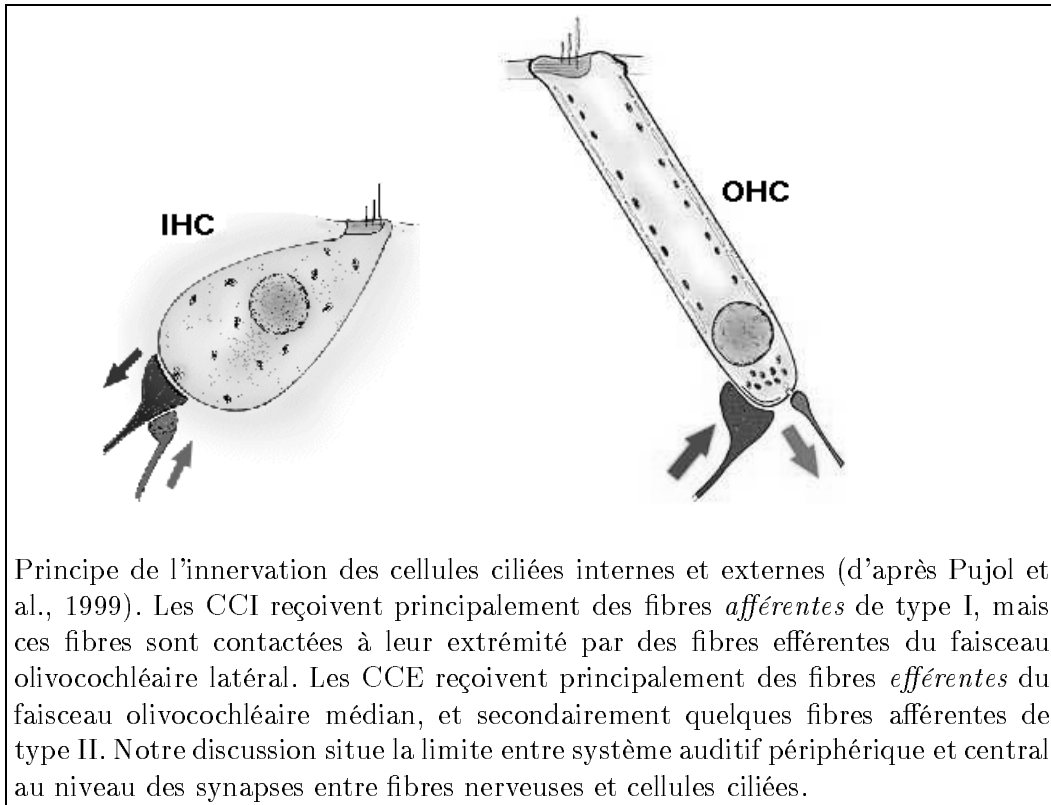
À leur extrémité ciliée, les CCI et CCE affleurent à la surface de l'organe de Corti. Ne dépassent de la surface que les cils. L'organe de Corti est renforcé intérieurement par deux rangées de piliers rigides, qui forment un tunnel de section triangulaire dont le sommet s'insère à la surface entre CCI et CCE, et dont la base s'appuie sur la membrane basilaire. Ce tunnel triangulaire (tunnel de Corti) donne une rigidité à l'ensemble, et rend la surface d'où émergent les cils strictement solidaire des mouvements de la membrane



Organe de Corti au microscope électronique à balayage (d'après Pujol et al., 1999). La barre représente 20 μm . Dans la partie supérieure on distingue les cils d'une rangée de CCI (à gauche) et de trois rangées de CCE (à droite), qui émergent de la surface de l'organe de Corti. Dans la partie inférieure l'organe de Corti est visible en coupe, avec la membrane basilaire en bas, et le tunnel de Corti (astérisque) entouré de piliers et traversé des fibres nerveuses qui desservent les CCE. On ne distingue en revanche ni le corps des CCI, ni les fibres qui les innervent. La membrane tectoriale, qui dans une cochlée normale vient se poser au dessus de l'organe de Corti et recouvrir les cils, est ici absente.

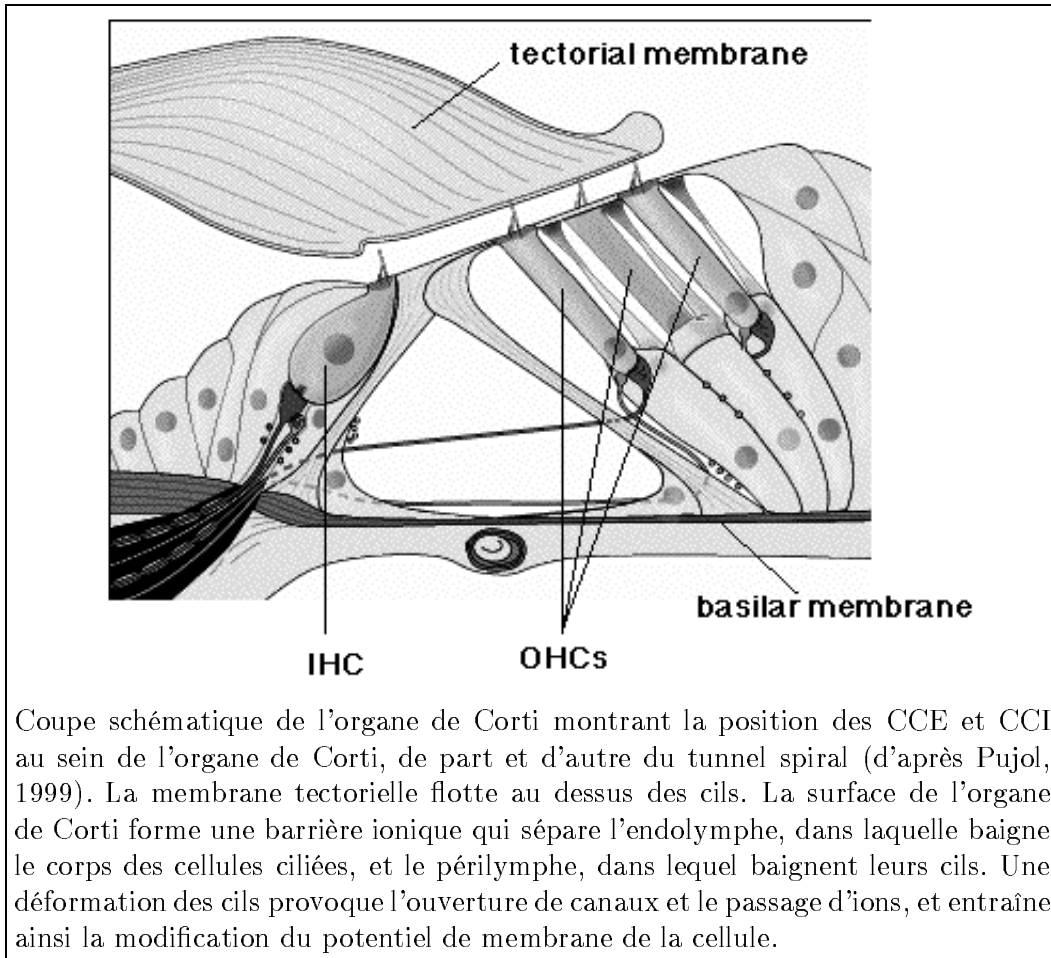
basilaire. Au dessus des cils flotte une membrane flasque, la membrane tectorielle, dans laquelle s'insèrent les cils des CCE. Les cils des cellules ciliées baignent dans l'endolymphe, et leur corps baigne dans la périlymphe, qui remplit le tunnel de Corti et les interstices entre cellules. Le potentiel à l'intérieur des cellules est intermédiaire entre les potentiels des deux liquides. Une déformation mécanique provoque l'ouverture de sortes de "trappes" sur la surface des cils, et laisse passer des ions entre l'endolymphe et le milieu intracellulaire. Il en résulte une modification du potentiel intracellulaire qui, dans le cas de la CCI, déclenche la sécrétion de neurotransmetteur qui excite les fibres nerveuses innervant la cellule.

Une surpression statique de l'air extérieur est transmise par les osselets du tympan à la fenêtre ovale, qui se déforme vers l'intérieur de la cochlée (rampe vestibulaire). Les rampes communiquant au sommet par l'hélicotrème, la colonne de liquide monte dans la rampe vestibulaire et descend dans la rampe



tympanique. La poussée du liquide déforme la fenêtre ronde vers l'extérieur. La membrane basilaire, elle, ne se déforme pas puisque l'hélicotreme égalise les pressions sur ses deux faces. Cela est vrai pour une poussée *statique*. Pour une poussée *variable* dans le temps, l'inertie du liquide tendra à s'opposer à son déplacement, et la membrane basilaire aura tendance à céder sous la pression. La position à laquelle elle se déforme dépend de la vitesse de variation de pression : près de l'apex si la variation est lente (pour une variation sinusoïdale : si la fréquence est basse), et près de la base pour des variations rapides (fréquences élevées). Cette dépendance fréquentielle des lieux de déformation maximale de la membrane est à l'origine de la sélectivité fréquentielle de l'oreille.

Le son est transmis du tympan à la cochlée par un enchaînement improbable d'osselets qui servent à adapter l'impédance acoustique du milieu extérieur (air) à celui du milieu liquide de la cochlée. Les *pressions* sont multipliées par un facteur considérable. Certains auteurs emploient le terme d'amplification pour qualifier cette multiplication de la pression : le terme d'atténuation conviendrait aussi bien, puisque la contrainte de conservation d'énergie exige que l'amplitude des *déplacements* soit diminuée dans un rap-



Coupe schématique de l'organe de Corti montrant la position des CCE et CCI au sein de l'organe de Corti, de part et d'autre du tunnel spiral (d'après Pujol, 1999). La membrane tectoriale flotte au dessus des cils. La surface de l'organe de Corti forme une barrière ionique qui sépare l'endolymphe, dans laquelle baigne le corps des cellules ciliées, et le périlymphe, dans lequel baignent leurs cils. Une déformation des cils provoque l'ouverture de canaux et le passage d'ions, et entraîne ainsi la modification du potentiel de membrane de la cellule.

port au moins égal à l'augmentation de pression. Le terme d'amplification ne convient donc pas à l'oreille moyenne, mécaniquement passive. En revanche il peut s'appliquer à l'oreille interne pour laquelle les CCE constituent une source d'énergie. La découverte du rôle actif des CCE (ou plus exactement la lente et encore incomplète acceptation de cette idée sous la pression d'un faisceau de preuves) a bouleversé les modèles de mécanique cochléaire et sélectivité fréquentielle, et aussi les modèles d'audition plus généraux dans la mesure où l'innervation des CCE, massivement efférente, met logiquement cette source d'énergie sous contrôle du système nerveux central.

Les mécanismes actifs confèrent à la réponse mécanique de la membrane basilaire une grande sélectivité fréquentielle. Les mesures récentes montrent qu'elle suffit à expliquer celle constatée au niveau du nerf auditif, et rendent inutile l'hypothèse d'un "second filtre" au delà la membrane basilaire (Evans, 1974). Les courbes d'accord (seuil de réponse d'une fibre en fonction de la

fréquence) ont la forme d'une vallée étroite, dont le flanc droit (vers les hautes fréquences) est extrêmement raide, avec une pente de l'ordre de 200 à 250 dB/octave. La distribution d'activité évoquée par un ton pur a donc une frontière apicale très nette, et n'empiète pas sur les zones plus proches de l'apex (de fréquence caractéristique plus basse). En revanche le flanc gauche des courbes d'accord est plus doux, ce qui fait que la distribution d'activité d'un ton pur peut facilement déborder sur les zones plus proches de la base. Cette asymétrie des patterns d'excitation se comprend dans la mesure où, avant d'atteindre son lieu de sensibilité maximale, une onde sinusoïdale doit se propager à travers des zones plus basales sensibles à des fréquences plus élevées.

Cette asymétrie vaut pour les filtres de fréquence caractéristique (CF) élevées, au delà de 1 kHz. Pour les CF plus basses les courbes d'accord sont plus symétriques. La profondeur de la courbe d'accord et son asymétrie diminuent lorsque l'état physiologique de l'animal se dégrade, lorsqu'on administre des poisons spécifiques des CCE, ou lorsqu'on stimule électriquement le système efférent olivocochléaire, ce qui suggère que la forte sélectivité fréquentielle est bien le fruit de l'activité motrice des CCE.

Sélectivité et asymétrie diminuent avec l'amplitude. Il en résulte que, pour des tons près de la fréquence caractéristique, la réponse de la membrane basilaire n'est pas linéaire. Cette non-linéarité se trahit aussi par la production de certains produits de distorsion (en particulier le terme $2f_1 - f_2$). La non-linéarité implique que l'on ne peut pas strictement prédire la réponse à une somme de tons purs par superposition des réponses à ces tons isolés. Du coup, la notion de *fonction de transfert* (rapport entre les transformées de Fourier de la sortie et de l'entrée du filtre) perd de son sens. Le pouvoir séparateur d'un canal (exclusion des composantes de fréquence autre que sa fréquence caractéristique) n'est pas aussi bon que le laisseraient prévoir les courbes d'accord. Il se dégrade à amplitude élevée, et la dégradation se traduit par le "recrutement" progressif par une composante intense des canaux de CF voisine. Malgré ces divers aspects non-linéaires, l'hypothèse de linéarité des réponses mécaniques est utile en première approximation.

Quelques évolutions récentes concernant le filtrage périphérique

Une approximation linéaire très utilisée pour la modélisation est celle du filtre "gammatone". Elle a été proposée par Johannesma (1972), et Carney et Yin (1988) ont montré qu'elle représente bien la forme des réponses impulsionnelles mesurées par la méthode revcor (corrélation entre un stimulus bruit blanc et la réponse d'une fibre nerveuse). Sur le plan pratique, plusieurs implémentations logicielles sont disponibles (Holdsworth et al., 1988 ;

O'Mard, Hewitt et Meddis, 1997 ; Slaney, 1993). Une évolution récente est l'apparition de versions non-linéaires du filtre gammatone (Lyon, 1997 ; Irino, 1997 ; Meddis, O'Mard et Poveda, 1999). Sur le plan psychoacoustique, Rosen (1998) a démontré que la non-linéarité dépend du niveau à la sortie des filtres, plutôt qu'à leur entrée. Cette constatation permet de simplifier la description des effets de la non-linéarité sur la sélectivité (si ce n'est que des décennies d'expériences se sont appuyées sur l'hypothèse inverse de contrôle par niveau à l'entrée du filtre). Les conséquences de la non-linéarité, des phénomènes actifs, et surtout de leur contrôle par le système efférent n'ont pas fini d'être explorées.

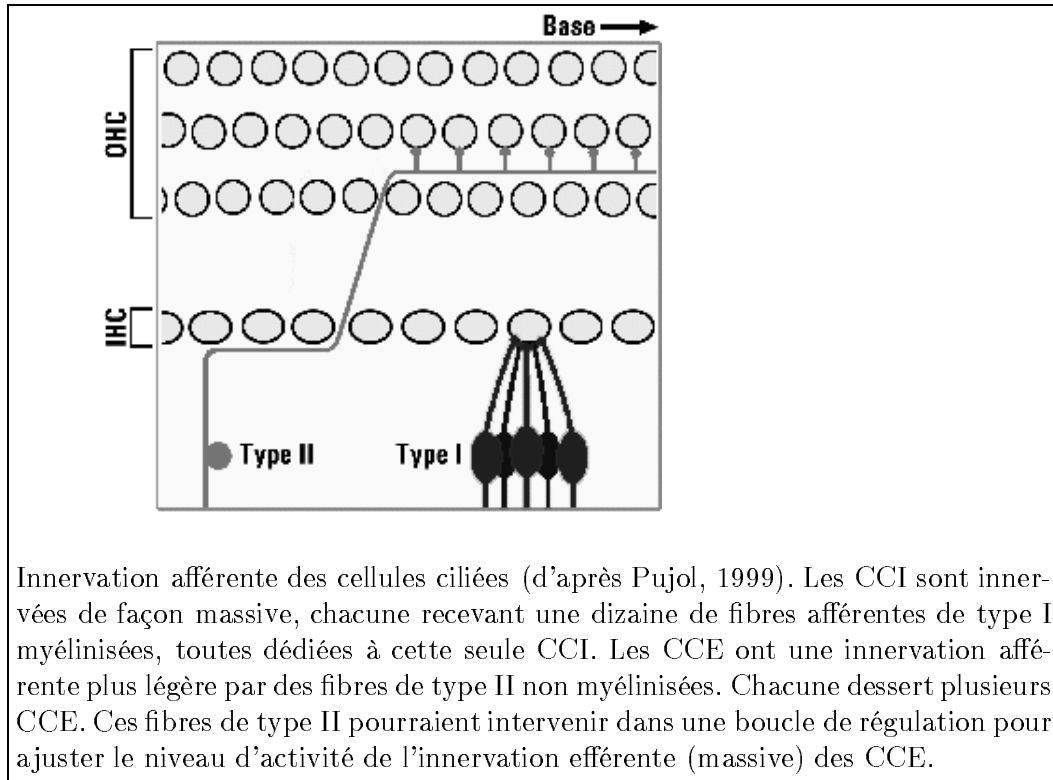
2.3.1 Innervation des cellules ciliées

Les cellules ciliées internes sont innervées par les dendrites de neurones du ganglion spiral. Chaque neurone dessert une seule CCI, mais chaque CCI reçoit des synapses de 10 à 30 neurones. L'information produite par une cellule est ainsi codée par une population de 10 à 30 neurones, qui envoient chacun un axone vers les niveaux centraux via le nerf auditif. Cette population n'est pas tout à fait homogène : on distingue les fibres à taux spontané élevé ($> 18/s$), dont le seuil est faible, des fibres à taux spontané faible, dont le seuil est plus élevé. Les neurones correspondants diffèrent aussi par leur morphologie et les détails de leur connectivité (Ruggero, 1992 ; Ryugo, 1992).

Le faisceau de 30000 à 50000 fibres afférentes provenant des CCI s'enroule autour de son axe pour former le nerf auditif. Les cellules sont ordonnées en fonction de leur fréquence caractéristique (CF), ce que l'on appelle *tonotopie*. L'ordonnancement tonotopique des CCI de l'apex (basses fréquences) à la base (hautes fréquences) est conservé dans le nerf (HF au centre, BF à la périphérie), et restitué plus centralement à de multiples étapes.

À côté des fibres provenant des CCI (fibres de type I) qui constituent l'essentiel des fibres afférentes, il y existe une deuxième population (fibres de type II), dont les terminaisons se trouvent sur les fibres efférentes qui desservent les CCE. Les enregistrements faits dans le nerf auditif proviennent des fibres de type I, qui sont pour la plupart myélinisées (Ryugo, 1992). Personne n'a réussi à enregistrer à partir des fibres de type II, non myélinisées, qui représentent 5 à 10% de la population efférente. Il pourrait s'agir de la partie afférente d'une boucle de contrôle du système de commande des CCE (Ryugo, 1992).

Outre ces fibres afférentes, il existe deux populations de fibres *efférentes* : la voie olivocochléaire médiane (MOC) qui dessert les CCE, et la voie olivocochléaire latérale (LOC) qui se termine sur les CCI (ou sur les dendrites afférentes qui les desservent). Cette dernière pourrait avoir un rôle d'ajuste-



ment de la sensibilité des fibres afférentes de type I.

2.3.2 Transduction

Le mouvement de la membrane basilaire déforme les cils des CCI et provoque l'émission de neurotransmetteurs qui modulent la probabilité de décharge des fibres afférentes. La transduction est asymétrique : une déformation positive (dans le sens rampe tympanique - rampe vestibulaire) accroît la probabilité de décharge. La déformation inverse produit une dépression par rapport au taux de décharge spontané, limitée par le fait qu'une probabilité ne peut devenir négative. Le taux *moyen* de décharge augmente avec l'amplitude, mais sature assez rapidement. Néanmoins, grâce à un effet de "contrôle automatique de gain", le taux *instantané* suit assez fidèlement le déplacement de la membrane dans son alternance positive. Ces propriétés - redressement simple alternance et contrôle automatique de gain - sont bien représentées par le modèle de cellule ciliée de Meddis (1988).

La déformation de la membrane détermine, via la cellule ciliée, la *probabilité* de décharge de la fibre de type I qui l'innerve. Les décharges elles-mêmes se produisent au hasard, selon un processus stochastique qui est bien repré-

senté par un processus de Poisson. Le processus est dit *inhomogène*, car son taux est modulé par la CCI. L'information neuronale paraît donc à la fois désordonnée (puisque les décharges se produisent au hasard) et structurée (puisque la probabilité est précisément contrôlée).

La probabilité de décharge suit un processus de Poisson inhomogène et et à *effets réfractaires* car un deuxième facteur contrôle la probabilité de décharge à un instant donné : l'histoire des décharges précédentes. En pratique, cette histoire peut être réduite à la durée de l'intervalle depuis la décharge précédente. Cette dépendance peut se représenter par une fonction $r(t - t_0)$. Cette fonction est nulle pendant la *période réfractaire absolue* (d'environ 0.7 ms), puis elle croit rapidement pendant la *période réfractaire relative*, avant d'attendre sa valeur asymptotique. Les deux facteurs se combinent de façon multiplicative :

$$p(t) = s(t)r(t - t_0) \tag{1}$$

ou $p(t)$ est le taux du processus, $s(t)$ est la fonction de probabilité produite par la CCI, t_0 est l'instant de la décharge précédente, et $r(t - t_0)$ s'appelle fonction de récupération (Johnson et Swami, 1983, [63, 8, 3]).

En l'absence de stimulus, le taux moyen de décharge prend une valeur constante appelée *taux spontané*. Comme noté plus haut, le taux spontané varie selon la fibre. On distingue trois populations de fibres : à faible, moyen ou fort taux spontané (de 0 à $0.5s^{-1}$ pour les faibles, de 0.5 à $18 s^{-1}$ pour les moyens, et de 18 et $140 s^{-1}$ pour les forts). La première se distingue des deux autres par la morphologie des fibres, la position de leurs synapses sur la CCI, et la position de leur projection dans le noyau cochléaire (Ruggero, 1992). Pour un stimulus soutenu, le taux moyen sature à une valeur de l'ordre de 100 à 300 /s, mais pendant des intervalles plus courts le taux instantané peut être plus élevé.

Pour des stimuli de fréquence suffisamment basse, le taux instantané suit la forme d'onde des vibrations : la décharge est *synchronisée*. Pour des fréquences supérieures à 1 kHz, la qualité de la synchronisation diminue. Elle disparaît au delà de 5 kHz pour le chat, ou 9-10 kHz pour la chouette. Cette perte de synchronisation a plusieurs causes. Elle peut s'expliquer en partie par un "bruit" temporel (imprécision de définition de l'instant de décharge), dont une partie pourrait être d'origine instrumentale, liée aux contraintes techniques de l'enregistrement des décharges. Mais elle a aussi une origine plus profonde, dans la caractéristique passe-bas de la cellule ciliée due à la capacité de membrane (Ruggero, 1992).

Au vu de ces données, on peut envisager deux types de codage de l'information. Un premier codage serait *temporel*, porté par un faisceau parallèle de fibres qui chacune porte une part de l'information dans la statistique tem-

porelle de ses décharges. Un deuxième serait de type *spectral*, représenté par le taux de décharge d'une fibre ou neurone en fonction de sa position sur un axe tonotopique.

Pour résumer : La cochlée se comporte comme un banc de filtres dont la sélectivité, extrêmement fine, est le résultat d'interactions entre les propriétés mécaniques passives de la membrane basilaire et des fluides qui l'entourent, et les propriétés actives des cellules ciliées externes. Les vibrations sont traduites en un ensemble de décharges de la population de fibres du nerf auditif. La population est organisée tonotopiquement, et la probabilité de décharge de chaque fibre varie dans le temps. La décharge d'ensemble peut donc se concevoir, soit comme une succession de spectres à court terme, soit comme un ensemble parallèle de signaux temporels qui diffèrent par le filtrage qu'ils ont subi. L'échange d'information entre périphérie et centre n'est pas à sens unique : le système nerveux central contrôle, via les CCE, l'état mécanique du système périphérique. Les propriétés et effets de ce système de contrôle sont mal connus.

2.4 Le système auditif central

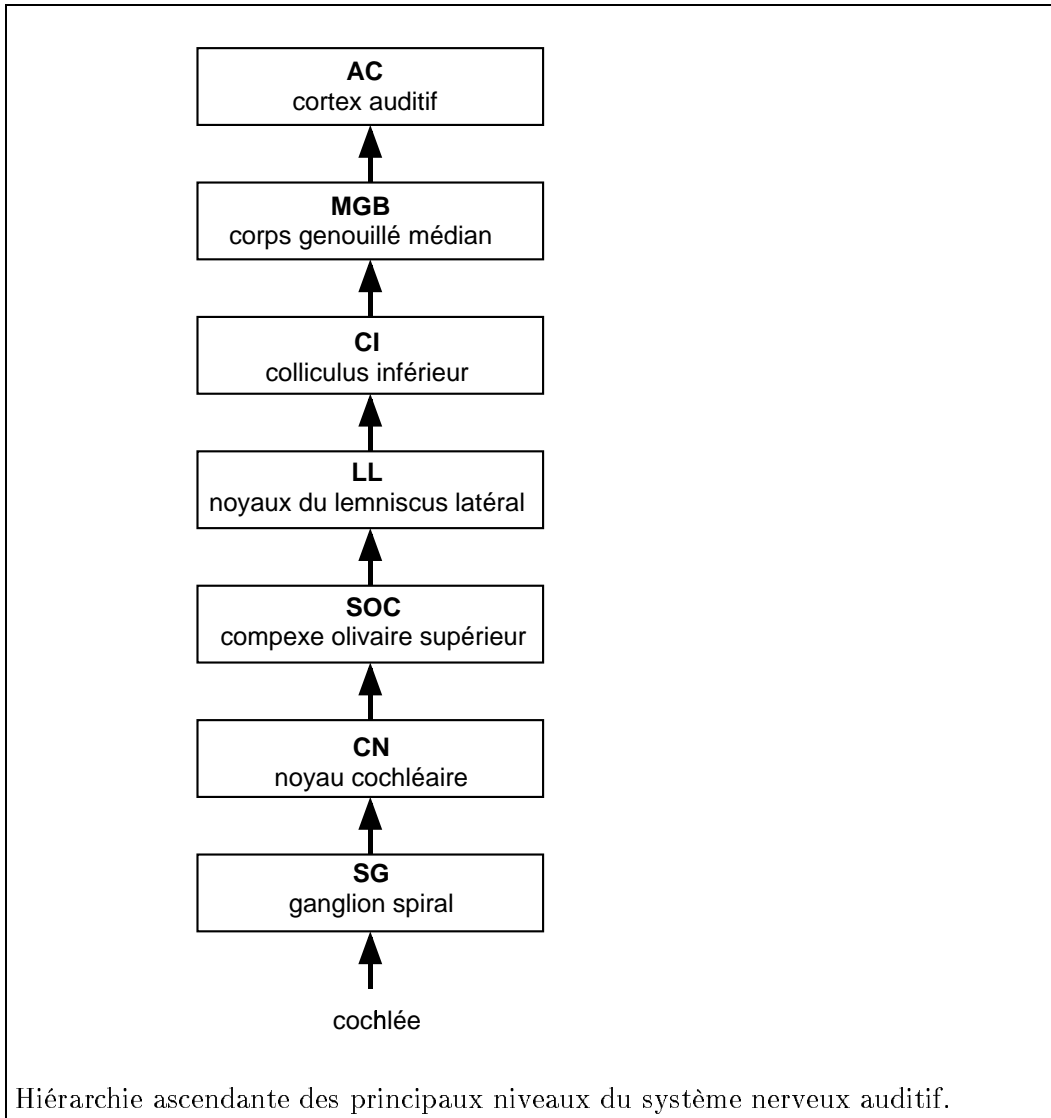
La section précédente décrivait la partie périphérique et son interface avec le nerf auditif, en insistant sur la nature tant efférente qu'afférente de cette interface. Cette section aborde une description du système auditif central, toujours par approximations successives, en allant du simple vers le complexe. Dans le chapitre suivant la discussion se concentrera sur des questions particulières, en particulier les indices d'une organisation *tonotopique* et de circuits spécialisés dans le *traitement temporel*.

2.4.1 Une hiérarchie de niveaux

Le système auditif central peut se décrire comme une hiérarchie ascendante de noyaux ou niveaux de traitement, allant du ganglion spiral jusqu'au cortex auditif.

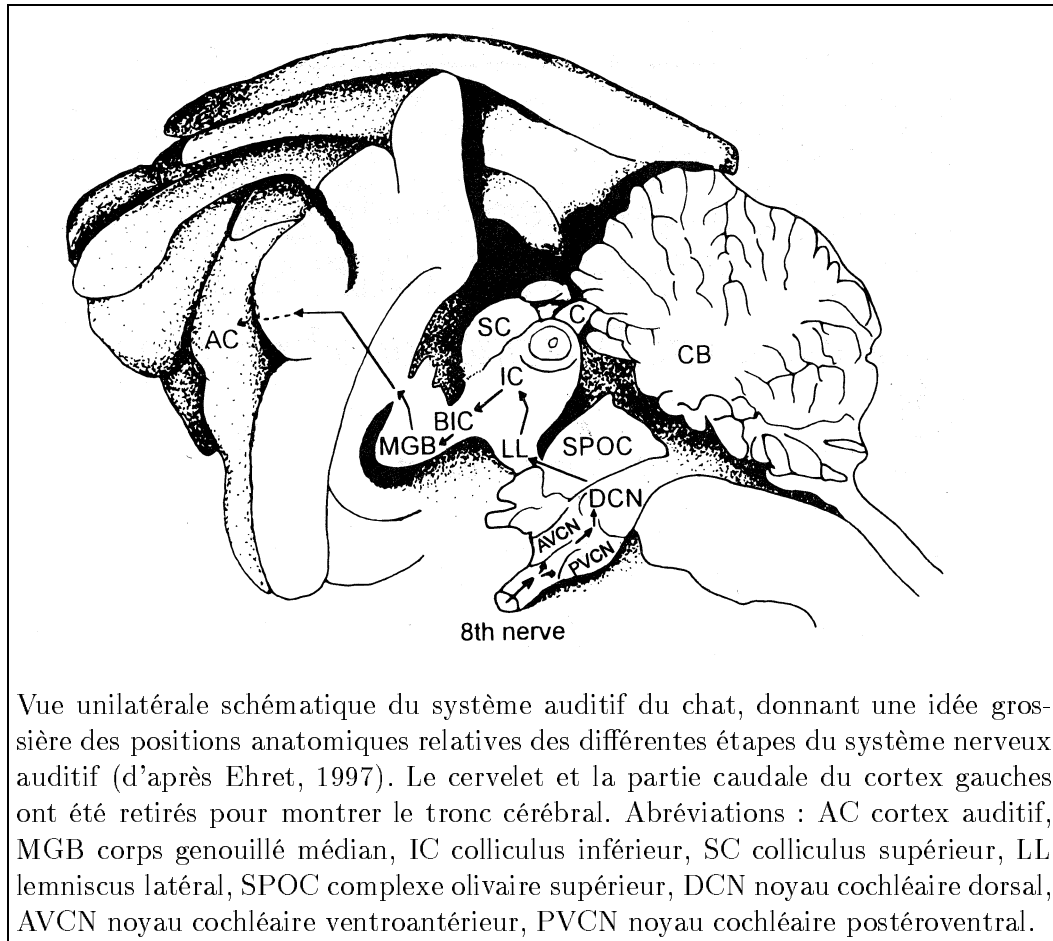
Les fibres afférentes du nerf auditif sont les axones des neurones du *ganglion spiral* (SG), dont les dendrites font synapse avec les cellules ciliées. Les étapes suivantes sont le *noyau cochléaire* (CN), le *complexe olivaire supérieur* (SOC), les *noyaux du lemnicus latéral* (LL), le *colliculus inférieur* (IC), le thalamus et en particulier le *corps genouillé médian* (MGB), et enfin le *cortex auditif* (AC).

Le ganglion spiral est logé dans une cavité osseuse près de l'axe de la cochlée. Noyau cochléaire et complexe olivaire se trouvent dans le *tronc cérébral*.



Le colliculus occupe le *mésencéphale*, le thalamus le *diencéphale*, et le cortex auditif le *téleencéphale*. Les noyaux du lemniscus latéral sont intermédiaires entre tronc cérébral et mésencéphale.

Les étapes les plus périphériques (CN, SOC, LL) sont pour l'essentiel spécialisées dans le traitement auditif, mais il existe néanmoins des apports non-auditifs dès le CN (Romand et Avan, 1997) qui se renforcent au fur et à mesure qu'on monte dans la hiérarchie (IC, MGB, cortex). La variété et la complexité des réponses aux sons augmente de la périphérie vers le centre, et on peut parler de *hiérarchie* des structures et des traitements.



2.4.2 Subdivisions

La description précédente était extrêmement simplifiée. Un premier pas vers une description plus fidèle est de considérer que chaque "étape" de la hiérarchie possède elle-même sa structure, faite de distinctions tant macroscopiques (divisions anatomiques) que microscopiques (distinctions de morphologie et connectivité neuronale, neurotransmetteurs, etc.). On a coutume de distinguer les divisions suivantes (Rouiller, 1997) :

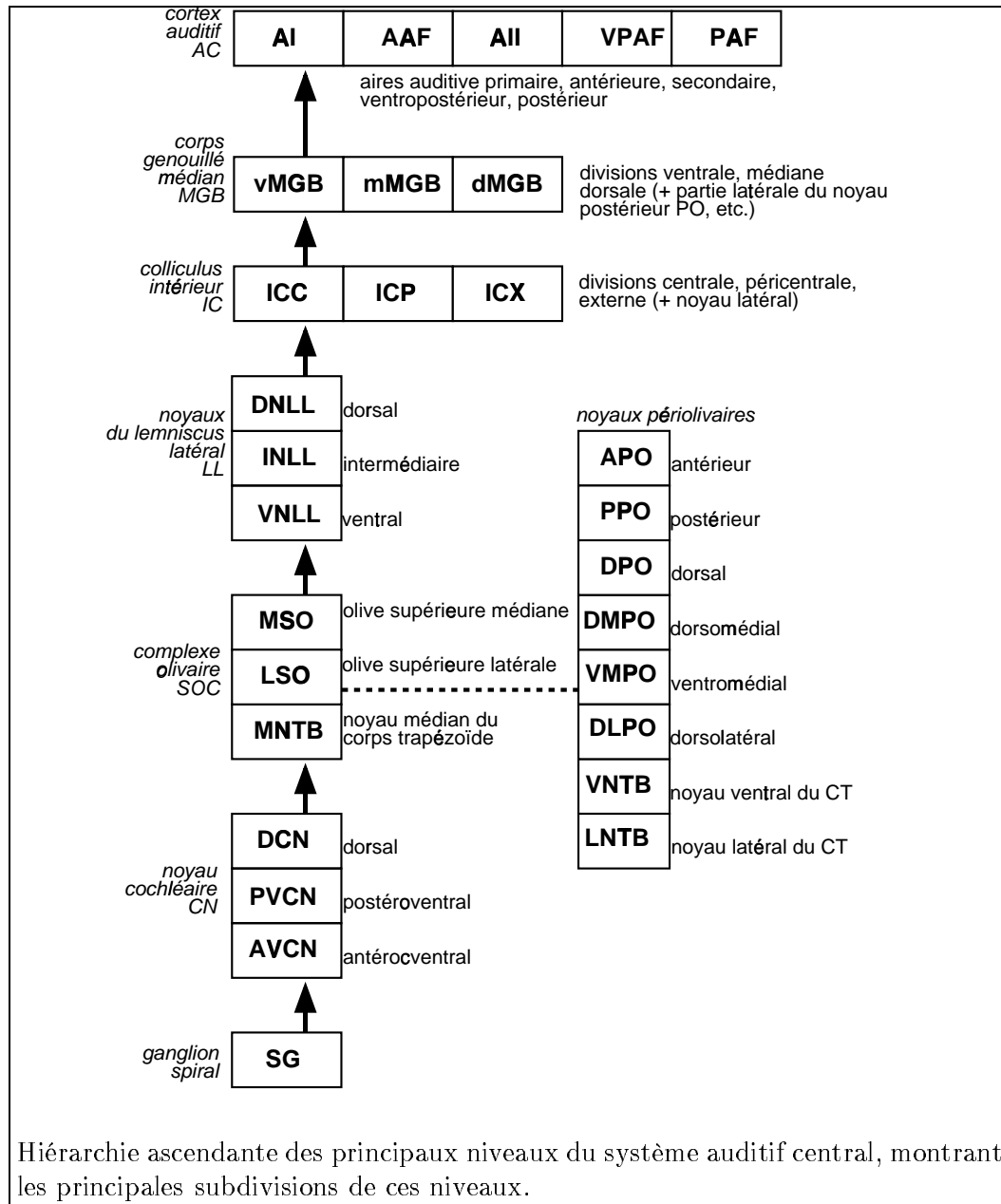
- Le noyau cochléaire (CN) comprend une division antéroventrale (AVCN), postéroventrale (PVCN) et dorsale (DCN).
- Le complexe olivaire supérieur (SOC) comprend trois noyaux principaux : le noyau latéral (LSO), le noyau médian (MSO), et le noyau médian du corps trapézoïde (MNTB). Outre ces noyaux principaux bien définis, on compte une demi-douzaine de noyaux périphériques dont les contours sont plus ou moins nets selon les espèces (Helfert et

Aschoff, 1997).

- Le noyau du lemniscus latéral peut être divisé en deux zones principales : une zone dorsale (DNLL) et une zone ventrale (VNLL), et on distingue aussi un noyau intermédiaire (INLL) (Helfert et Aschoff, 1977).
- Le colliculus inférieur (IC) comprend une division centrale (ICC) à la structure laminaire très marquée, entourée d'un noyau dorsomédian, d'un cortex dorsal, et d'un noyau latéral.
- Dans le thalamus auditif, le corps genouillé médian (MGB) comprend une division ventrale (vMGB), une division médiane, et une division dorsale. On peut lui associer deux autres noyaux : la partie latérale du groupe postérieur des noyaux thalamiques, et la partie auditive du noyau réticulaire du thalamus (de Ribaupierre, 1997).
- Le cortex auditif (AC) a une structure qu'on peut concevoir comme le produit cartésien d'une division en *couches* et d'une division en *aires*. La structure en couches (numérotées de I à VI depuis la surface) du cortex auditif ressemble à celle d'autres parties du cortex, avec des spécificités dans l'importance relative de certaines couches. Orthogonalement à cette structure en couches, les aires se répartissent selon des critères anatomiques (cortex auditif primaire, secondaire, tertiaire, etc.) ou électrophysiologiques. On distingue ainsi quatre champs auditifs (organisés tonotopiquement chez l'animal anesthésié) : le champ antérieur (AAF), primaire (AI), postérieur (PAF) et ventropostérieur (VPAF). Ils sont entourés de champs dont l'organisation est moins nette ou moins bien étudiée que les aires ventral, temporal ou dorsopostérieur (de Ribaupierre, 1997).

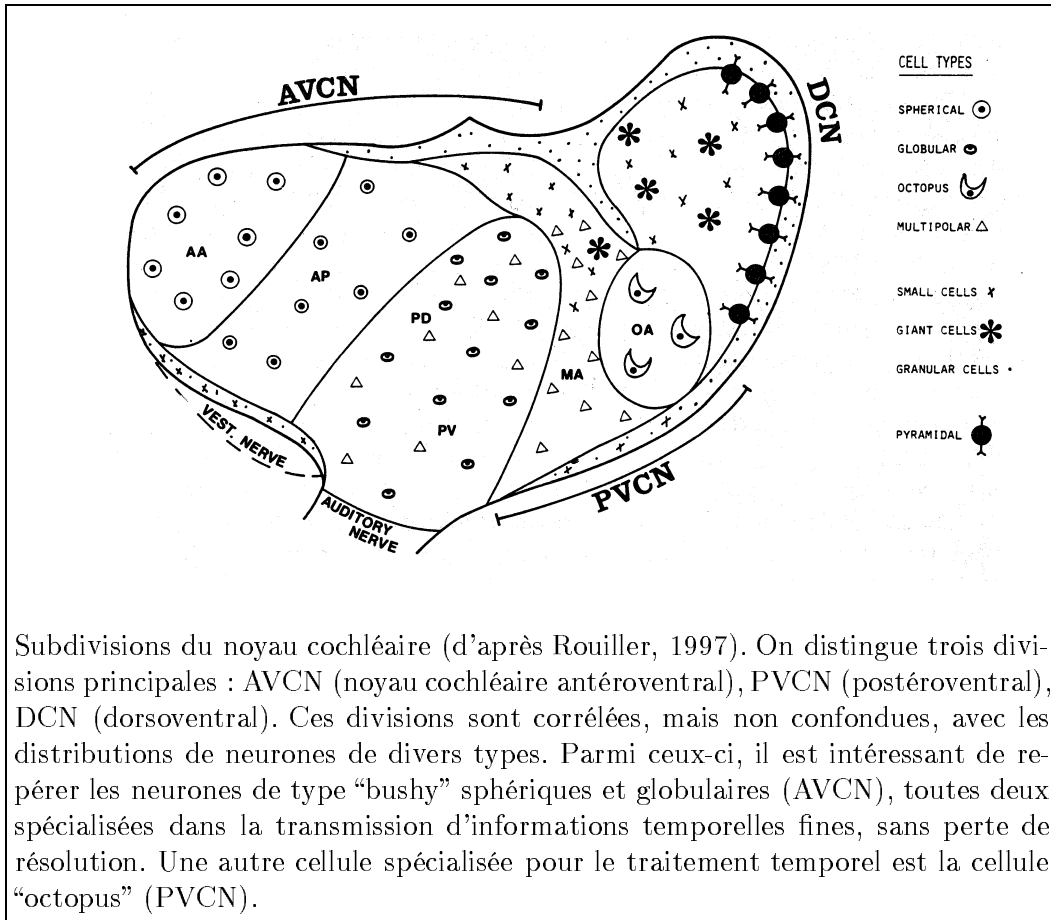
Des divisions plus fines peuvent bien entendu être faites, et selon les auteurs, les critères, et/ou l'espèce étudiée, les frontières de divisions et leur noms peuvent changer. À côté d'une organisation selon des lignes anatomiques, la population de neurones de chaque noyau peut être classée selon des critères morphologiques (taille et forme des soma et arborescences axonales et dendritiques), de connectivité, électrophysiologiques (forme des réponses), ou neuropharmacologiques (nature des neurotransmetteurs). Les frontières de ces classes peuvent être cohérentes avec les divisions anatomiques, voire les sous-tendre lorsqu'une division est définie par le type de cellules qu'elle contient. Mais le plus souvent un type de cellule est distribué sur plusieurs divisions anatomiques, et une division contient plusieurs types cellulaires.

La variété de cellules croît de la périphérie vers le centre. Dans le ganglion spiral on ne distingue que trois types de neurones. Les neurones de type II sont petits, à axones non-myélinisés, et leurs dendrites contactent les fibres efférentes qui contrôlent les CCE. Les neurones de type I sont plus grosses,



leurs axones sont souvent myélinisés (selon l'espèce), et ils desservent chacun une CCI. Les neurones de type I existent en deux variétés qui se distinguent par leur taux de décharge spontané et par des détails morphologiques et de connectivité.

En revanche au niveau cortical Winer (1992) distingue un total de 47 sortes de neurones (dont certaines, il est vrai, sont communes à plusieurs



Subdivisions du noyau cochléaire (d'après Rouiller, 1997). On distingue trois divisions principales : AVCN (noyau cochléaire antéroventral), PVCN (postéroventral), DCN (dorsoventral). Ces divisions sont corrélées, mais non confondues, avec les distributions de neurones de divers types. Parmi ceux-ci, il est intéressant de repérer les neurones de type "bushy" sphériques et globulaires (AVCN), toutes deux spécialisées dans la transmission d'informations temporelles fines, sans perte de résolution. Une autre cellule spécialisée pour le traitement temporel est la cellule "octopus" (PVCN).

divisions). À la variété des neurones, on peut ajouter celle des synapses, qui diffèrent par leur morphologie, les neurotransmetteurs qu'elles emploient, les propriétés de membrane, etc..

Parmi les types cellulaires, il sera intéressant pour la suite de distinguer les cellules qui sont spécialisées dans la transmission avec une résolution temporelle fine. Ce sont par exemple les cellules de type "bushy" de la division antéro-ventrale du noyau cochléaire (AVCN), qui relaient les informations du nerf auditif avec une grande fidélité vers les niveaux supérieurs, ou les cellules de type "octopus" de PVCN qui répondent avec des impulsions isolées précisément synchronisées avec l'attaque ou la période d'un stimulus.

Pour résumer, chaque niveau de la structure hiérarchique (SG -> CN -> SOC -> LL -> MGB -> cortex) peut être subdivisé à son tour, de façon diverse selon les critères adoptés. Une stratification fine produit une description plus proche de la réalité, mais plus difficile à retenir et plus lourde à manier.

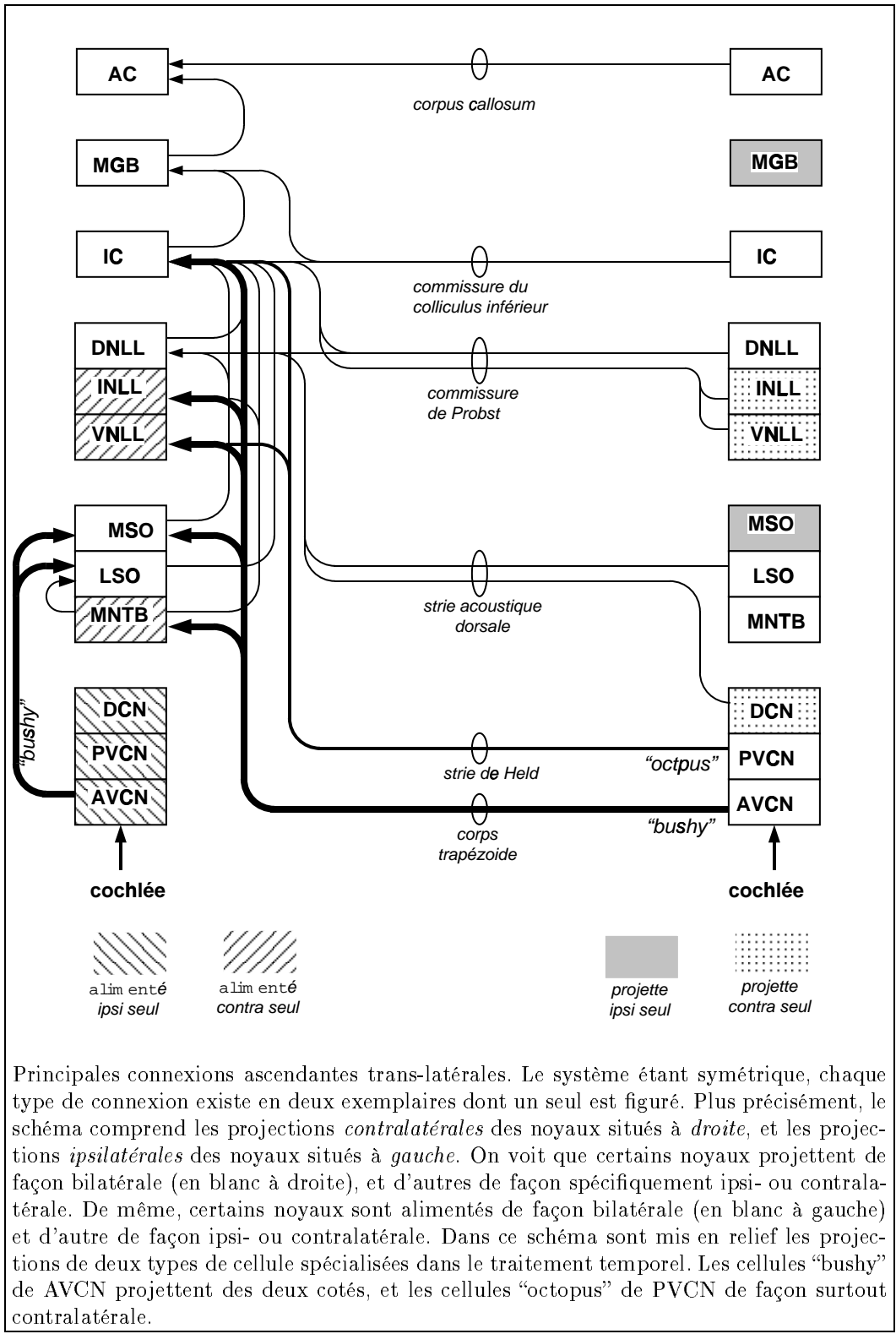
2.4.3 Structure bilatérale

Jusqu'ici la discussion a ignoré le fait que chaque niveau existe en *deux exemplaires*, depuis les oreilles jusqu'au cortex. Une symétrie règne entre les côtés gauches et droits, du moins aux niveaux les plus périphériques (oreille, SG, CN, SOC, LL, IC) et sauf spécialisation exceptionnelle, tels les pavillons asymétriques de la chouette effraie (*Tyto alba*, en anglais *barn owl*). Le cortex en revanche montre des signes de spécialisation hémisphérique, et il serait logique qu'il en soit de même de MGB qui lui est étroitement associé. L'hémisphère gauche est censé, chez la plupart des individus, traiter la parole et le langage, et l'hémisphère droit les aspects intonatifs ou musicaux (de Ribaupierre, 1997).

Des connexions croisées gauche-droite existent à plusieurs niveaux. Le premier niveau important d'intégration binaurale est le complexe olivaire supérieur (SOC), dont deux noyaux (LSO et MSO) sont réputés spécialisés respectivement dans le traitement des disparités binaurales d'intensité et de temps d'arrivée. Des projections depuis le côté opposé existent cependant dès le noyau cochléaire (CN), voire même dès la cochlée, puisque les voies olivocochléaires latérale et médiane ont chacune une composante ipsilatérale et contralatérale. Aux niveaux plus centraux que SOC, des échanges gauche-droite ont lieu à de multiples niveaux. La projection de LSO sur IC est ipsilatérale pour les fréquences basses, et contralatérale pour les fréquences hautes (pour cette raison elle a été comparé au chiasma optique, Rouiller, 1997). Des connexions croisées massives existent à la sortie de LL (via la commissure de Probst) et de IC (via la commissure du colliculus inférieur), et au niveau du cortex auditif via le corps calleux qui relie les deux hémisphères. C'est dire que les connexions bilatérales existent à la sortie de pratiquement tous les niveaux principaux supérieurs SOC, sauf un : MGB (Ehret, 1997).

Malgré ce brassage bilatéral massif, chaque niveau garde un certain caractère monaural de par la connectivité qui le relie de façon plus dense à une oreille, et de par sa sensibilité plus grande aux sons venant d'un côté plutôt que de l'autre. Ce qui complique un peu cette image, c'est que la dominance est tantôt ipsilatérale, tantôt contralatérale, selon le niveau. Ainsi l'oreille droite alimente exclusivement le CN droit, et préférentiellement le LSO droit et les MNTB, LL, et IC gauches. Au niveau du thalamus (MGB) et du cortex les neurones à sensibilité monaurale sont plus rares, mais on note cependant une certaine dominance contralatérale (de Ribaupierre, 1997). Le MSO est alimenté de façon assez équilibrée par les deux oreilles. Il en est de même du LSO, mais la source contralatérale, via le MNTB, est surtout inhibitrice.

En résumé, des échanges translatéraux existent à de nombreux niveaux. Certains s'expliquent par les traitements binauraux nécessaires pour la lo-



Principales connexions ascendantes trans-latérales. Le système étant symétrique, chaque type de connexion existe en deux exemplaires dont un seul est figuré. Plus précisément, le schéma comprend les projections *contralatérales* des noyaux situés à *droite*, et les projections *ipsilatérales* des noyaux situés à *gauche*. On voit que certains noyaux projettent de façon bilatérale (en blanc à droite), et d'autres de façon spécifiquement ipsi- ou contralatérale. De même, certains noyaux sont alimentés de façon bilatérale (en blanc à gauche) et d'autre de façon ipsi- ou contralatérale. Dans ce schéma sont mis en relief les projections de deux types de cellule spécialisées dans le traitement temporel. Les cellules "bushy" de AVCN projettent des deux cotés, et les cellules "octopus" de PVCN de façon surtout contralatérale.

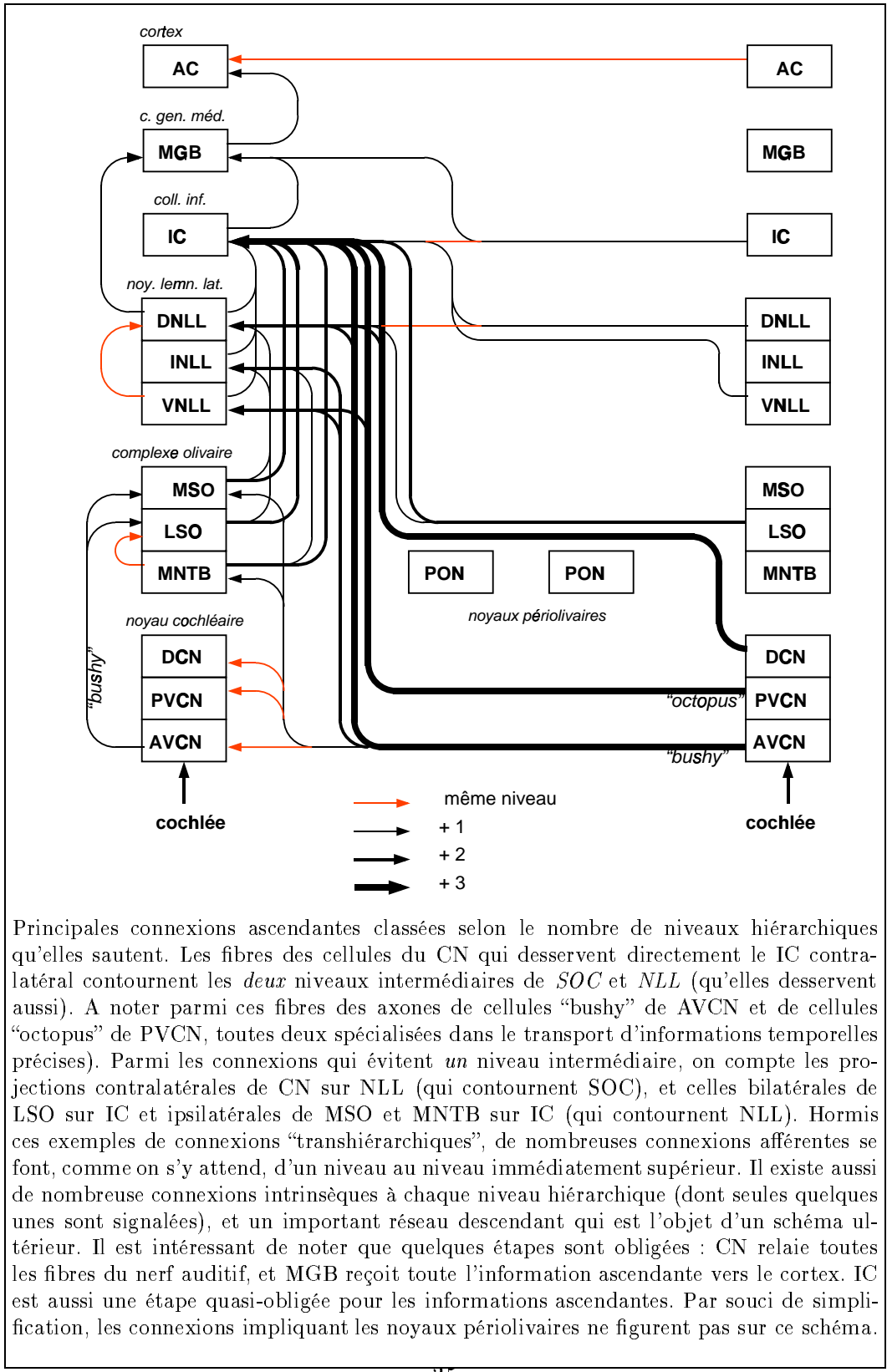
calisation et la séparation des sources, mais il est peu probable que ce soit le cas de toutes les convergences bilatérales. Il semble plutôt qu'il s'agisse d'un aspect du parallélisme de la connectivité entre niveaux (voir section suivante). La multiplication des circuits possibles qui résulte de ce parallélisme pourrait contribuer à la sophistication des traitements, qu'ils soient ou non binauraux. Elle pourrait aussi contribuer à la robustesse, grâce à la redondance des circuits.

2.4.4 Parallélisme et sauts “trans-hiérarchiques”

Jusqu'ici la discussion a présenté la succession de niveaux comme s'il s'agissait d'une hiérarchie stricte, chaque niveau alimentant le niveau immédiatement supérieur. En fait, il existe de nombreuses voies parallèles ascendantes, dont certaines sautent des étapes dans leur progression vers le centre. C'est un facteur de complexité en ce sens qu'il interdit la récapitulation au niveau de chaque étape. Il faut souligner que l'abstraction et l'encapsulation de détails ("data hiding"), qui sont des techniques familières de maîtrise de la complexité en ingénierie informatique, nécessitent que soit respectée une hiérarchie stricte. Elles ne peuvent s'appliquer à un système qui comporte des sauts de niveau.

Des exemples de sauts de niveau sont les connexions de CN vers LL et IC (sans passer par SOC), de SOC vers IC (sans passer par LL), et de LL vers MGB (sans passer par IC), etc.. La combinatoire est cependant limitée par le fait que certains noyaux sont des étapes obligées. Ce sont le *noyau cochléaire* où se terminent toutes les fibres ascendantes du nerf auditif, et le *colliculus inférieur* qui reçoit des fibres de presque tous les noyaux de niveau inférieur (des deux côtés) et que ne contournent que quelques fibres destinées au MGB en provenance du noyau de la voie acoustique centrale (NCAT) (Schwartz, 1992) ou de VNLL (Schofield et Cant, 1997). Une troisième étape obligée est le *thalamus* (essentiellement MGB) qui filtre toute information ascendante vers le cortex auditif (AC). Outre les voies qui montent en sautant une ou plusieurs étapes, on note bien sûr de nombreuses et parallèles voies d'un niveau vers le niveau supérieur, ainsi qu'à l'intérieur de chaque niveau.

Pour résumer, le parallélisme des liaisons ascendantes est un élément supplémentaire de complexité, auquel il faut ajouter les liaisons *intrinsèques* à un noyau ou niveau donné, et les connexions *descendantes* qui sont décrites dans la section suivante.



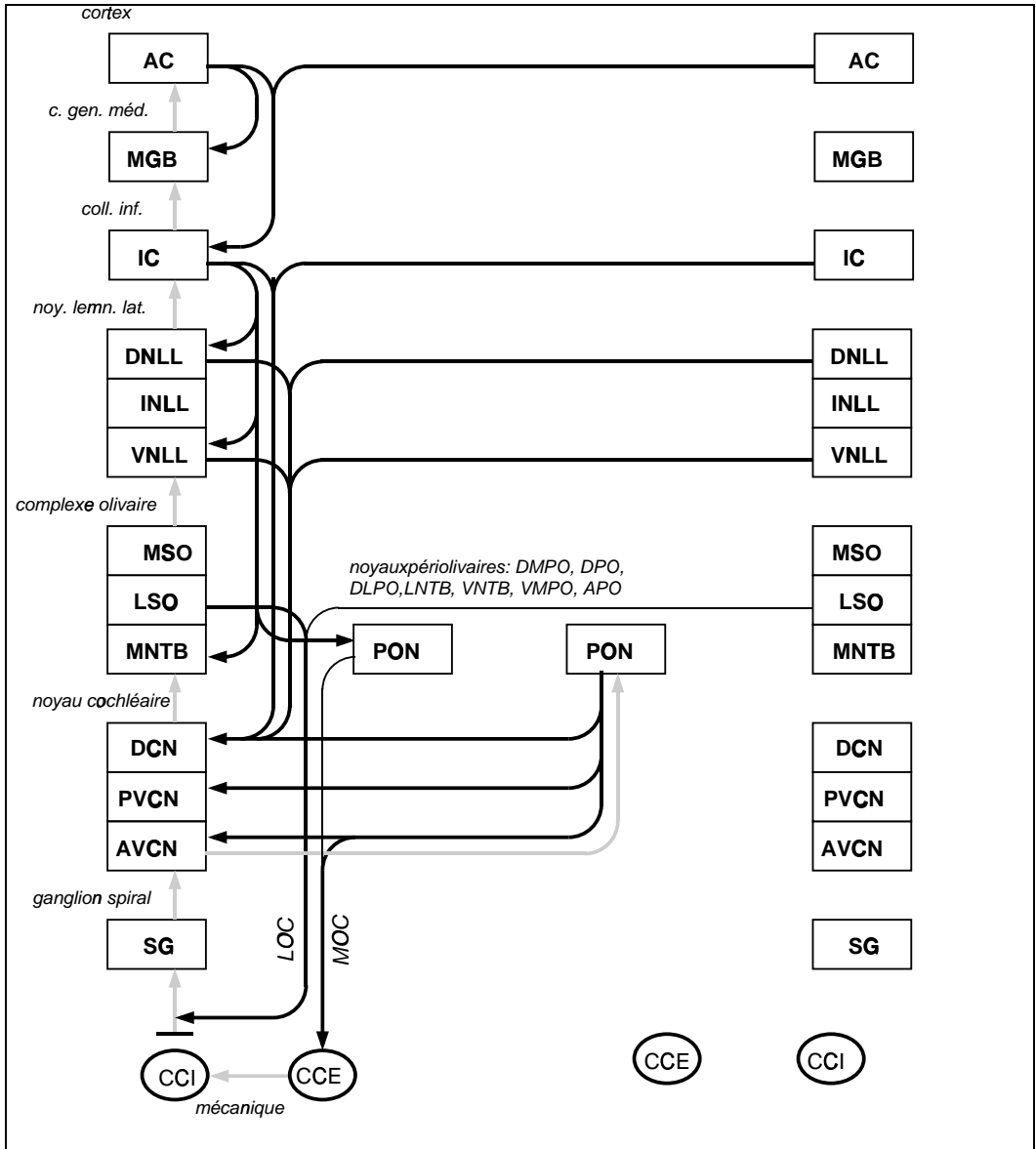
2.4.5 Voies descendantes

L'existence de voies descendantes est un cauchemar pour la caractérisation fonctionnelle d'un système. Les chemins possibles pour aller d'un niveau à un autre se démultiplient à l'infini (à la manière des miroirs en vis-à-vis que l'on trouve dans certains restaurants). On ne peut donc préciser le niveau hiérarchique dans un échange d'information, ni même compter le nombre d'"étapes" menant d'un niveau à un autre. La rétroaction laisse craindre un risque d'instabilité, qui est difficile à apprécier car le système est non-linéaire et non-invariant dans le temps (du fait de changements d'état, contexte, éveil, etc.). Les voies descendantes sont pourtant nombreuses, presque autant que les voies ascendantes, et dans certains cas aussi massives. Notre seule consolation est de noter les quelques cas où elles sont absentes ou particulièrement simples (boucle réciproque).

Ni le ganglion spiral ni le noyau cochléaire n'ont de projections efférentes. Du complexe olivaire supérieur, deux voies distinctes se projettent vers la cochlée. La première est la voie olivocochléaire latérale (LOC) qui part des cellules marginales de LSO et aboutit sur des synapses au contact des fibres afférentes des CCI. Elle suit une organisation tonotopique précise : une fibre efférente contacte le lieu cochléaire d'où parvient la fibre qui nourrit, via une cellule multipolaire de AVCN, son neurone d'origine dans le SOC. Il pourrait s'agir d'une boucle réciproque à trois neurones très simple, dont le but serait d'ajuster la sensibilité des fibres afférentes.

La deuxième voie olivocochléaire est le faisceau médian (MOC) qui assure, via des fibres myélinisées, une innervation efférente massive des CCE. Elle trouve son origine dans le MNTB et les noyaux périolivaires du SOC contralatéral (Warr, 1992 ; Schwartz, 1992). La plupart des cellules d'origine sont situés dans le corps trapézoïdal, zone où passent de nombreuses fibres issues de SOC, ainsi que des fibres descendantes de IC. Cela fait dire à Warr (1992) que les neurones d'origine du MOC sont "les neurones les mieux informés du tronc cérébral". Ils sont alimentés de façon dominante par des fibres venant du CN contralatérale, c'est à dire du côté où ils projettent. On penserait à une simple boucle réciproque (à laquelle participeraient les neurones de type II du SG ; Ryugo, 1992), si n'était la présence massive d'autres apports, notamment descendant de niveaux supérieurs.

Le MOC contrôle les CCE, et donc les propriétés mécaniques actives de la cochlée. L'hypothèse d'un traitement sophistiqué avec modifications dynamiques du filtrage en temps réel est attrayant, mais il est difficile d'en faire la preuve. Les hypothèses avancées sont généralement moins ambitieuses : contrôle automatique de gain, ajustement (mécanique) du point de fonctionnement de la CCI, etc.. Il est intéressant de noter que l'innervation efférente



Principales connexions descendantes. Il est intéressant de noter les lacunes : aucune connexion efférente n'aboutit au ganglion spiral, aucune n'est issue du ganglion ni du noyau cochléaire, et aucune ne contourne le colliculus inférieur (à quelques exceptions près), ce qui fait que IC est une étape quasi-obligée des voies descendantes comme ascendantes. Parmi les voies efférentes vers la cochlée, la voie olivocochléaire latérale (LOC) issue de LSO se termine par des synapses sur les dendrites des fibres afférentes de type I (à proximité de leur synapse sur la CCI dans la cochlée). Il pourrait s'agir d'une boucle de rétroaction chargée d'ajuster la sensibilité des fibres afférentes. La voie olivocochléaire médiane (MOC) projette sur les CCE motrices de la cochlée en provenance des noyaux périolivaires contralatéraux. Ces derniers sont alimentés depuis le CN ipsilatéral : le MOC contrôle donc de la cochlée qui l'alimente. Il peut s'agir ici encore d'une boucle de rétroaction, dont le but serait d'ajuster le point de fonctionnement mécanique de la cochlée.

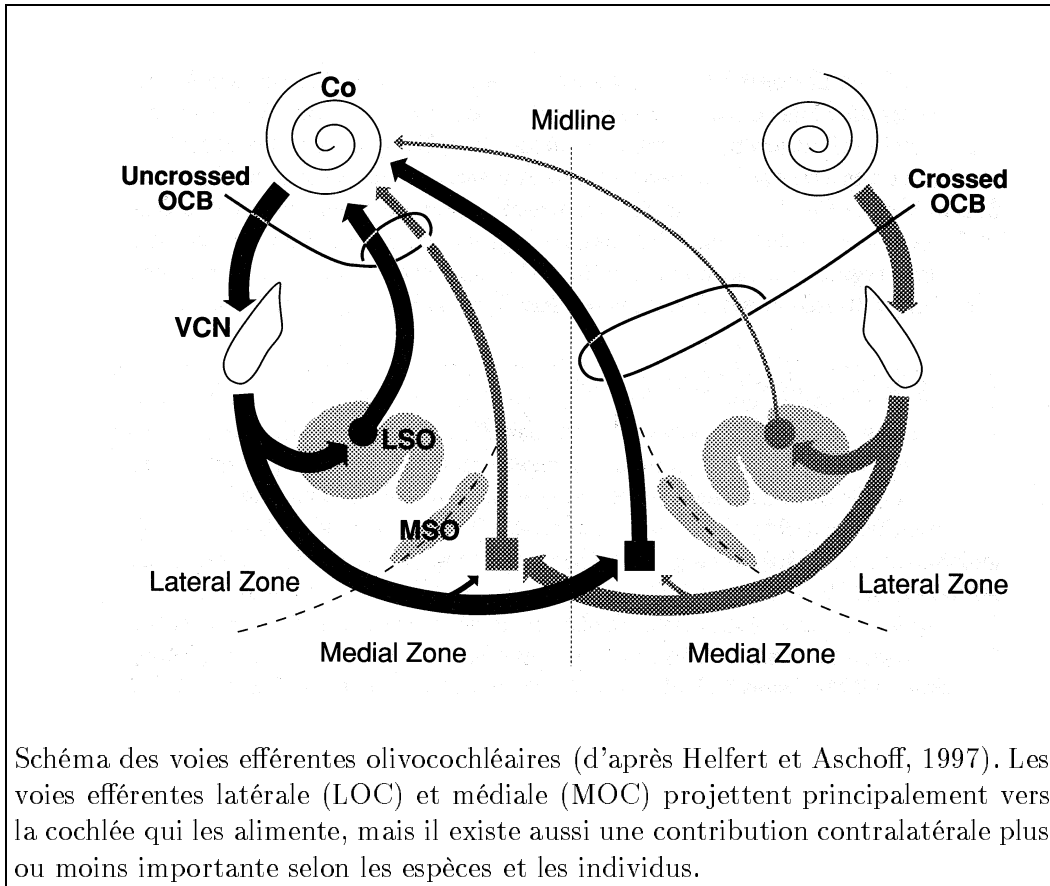
est phylogénétiquement aussi ancienne que les cellules ciliées elles-mêmes, et précéderait la différenciation des vertébrés amniotes en une lignée menant aux mammifères, et d'autres lignées menant aux tortues, reptiles ou oiseaux, il y a environ 300 millions d'années (Manley et Köppl, 1998).

Une hypothèse hautement spéculative est celle d'une interaction sur la membrane basilaire entre sons venus de l'environnement et *vibrations endogènes*, produites par la transduction de signaux nerveux à structure temporelle fine (synthétisés centralement, ou provenant de l'oreille opposée). Ce qui rend l'hypothèse attrayante c'est qu'elle mettrait à profit la relative linéarité de la membrane basilaire, "dernière étape linéaire avant le système auditif central" (Møller, 1983). Certains indices sont en sa faveur, dont le fait que les fibres efférentes du MOC sont myélinisées, et que certains des neurones qui leur donnent naissance reçoivent des synapses d'un type s'apparentant aux calices de Held (Warr, 1992). Autre donnée en sa faveur, une expérience psychoacoustique a mis en évidence des battements binauraux entre un stimulus supraliminaire d'un côté, et *subliminaire* (de 10 à 20 dB) de l'autre (Ayres et Clack, 1984). Il est difficile d'imaginer qu'un stimulus subliminaire soit suffisamment bien représenté centralement pour produire des battements. En revanche les battements pourraient s'expliquer si on admettait l'idée d'une interaction mécanique entre stimulus subliminaire et vibrations transmises d'une oreille à l'autre via un codage nerveux.

Malheureusement pour cette hypothèse, (a) ces données psychoacoustiques n'ont pas été confirmées par des expériences plus récentes (Gu, Wright et Green, 1995), et (b) les enregistrements dans les fibres du MOC n'ont révélé que des réponses régulières du type "chopper", et non une structure temporelle corrélée avec le stimulus. Cette hypothèse attrayante est donc à écarter.

La majorité des fibres du système olivocochléaire proviennent de neurones alimentés depuis le côté qu'ils desservent. Une minorité dessert l'autre oreille. Le nombre de neurones impliqués et les proportions (LOC vs MOC, ipsilatéral vs contralatéral) est très variable d'une espèce à l'autre, et d'un individu à l'autre (Warr, 1992). Certaines fibres olivocochléaires desservent aussi CN, et il existe aussi des efférences SOC->CN propres (Schwartz, 1992). La cochlée ne reçoit pas de connexions descendantes directes depuis des niveaux supérieurs à SOC.

En ce qui concerne les noyaux de niveau supérieur au complexe olivaire, la plupart des projections de LL sont ascendantes, mais il existe néanmoins des efférences du noyau dorsal (DNLL) vers le CN. Le colliculus inférieur (IC) projette bilatéralement vers le noyau dorsal du CN (DCN), et ipsilatéralement vers les noyaux périolivaires et vers LL (Ehret, 1997). Le MGB n'a apparemment pas de projections descendantes (Winer, 1992). Le cortex



auditif (AC) en revanche projette vers MGB et IC. Les connexions entre cortex et MGB sont massives et ordonnées, et sauf exception chaque division de MGB est connectée avec une aire particulière du cortex par une projection réciproque. Le cortex ne semble pas projeter directement vers LL, SOC (Schwartz, 1992), ou CN (Cant, 1992), ni bien sûr vers SG ou la cochlée). Une exception, citée par Romand et Avan (1997), est une projection directe du cortex vers CN.

En résumé, les voies descendantes sont nombreuses. Tous les noyaux importants du système auditif (sauf le ganglion spiral, SG) sont, soit la source, soit l'aboutissement d'une voie centrifuge. L'absence de certaines combinaisons n'en est que plus remarquable. Le thalamus (MGB) n'a aucune projection descendante, et la plupart de celles du cortex ne dépassent pas IC. Le colliculus inférieur joue ainsi un rôle de "verrou" pour les voies descendantes, semblable à celui qu'il joue pour les voies ascendantes. Ce rôle (et peut-être aussi la régularité structurelle de son noyau central ?) lui a valu d'être qualifié

de "gare de triage" du traitement de l'information (Ehret, 1997). IC projette à tous les niveaux qui lui sont inférieurs sauf la cochlée (SG et cellules ciliées). Les cellules ciliées quand à elles ne reçoivent d'efférences que du complexe olivaire. L'absence d'une projection directe entre deux niveaux donnés n'interdit bien sûr pas une influence *indirecte*. Ainsi, il suffit de trois étapes pour passer du cortex à la cochlée (via IC et SOC). On peut expliquer ainsi l'effet, cité plus haut, que peut avoir une tâche cognitive sur les oto-émissions évoquées (Puel et al., 1988).

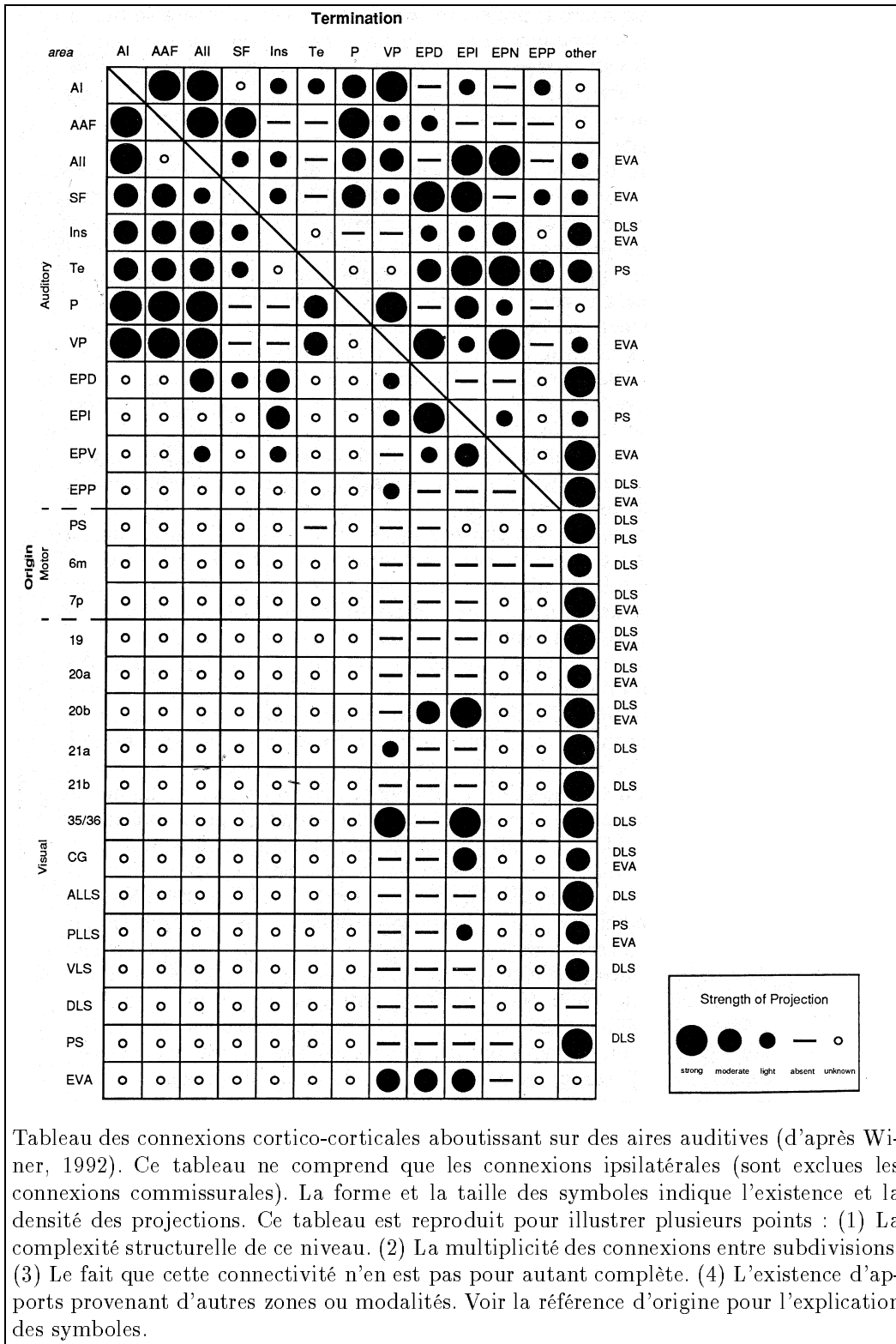
2.4.6 Autres facteurs de complexité

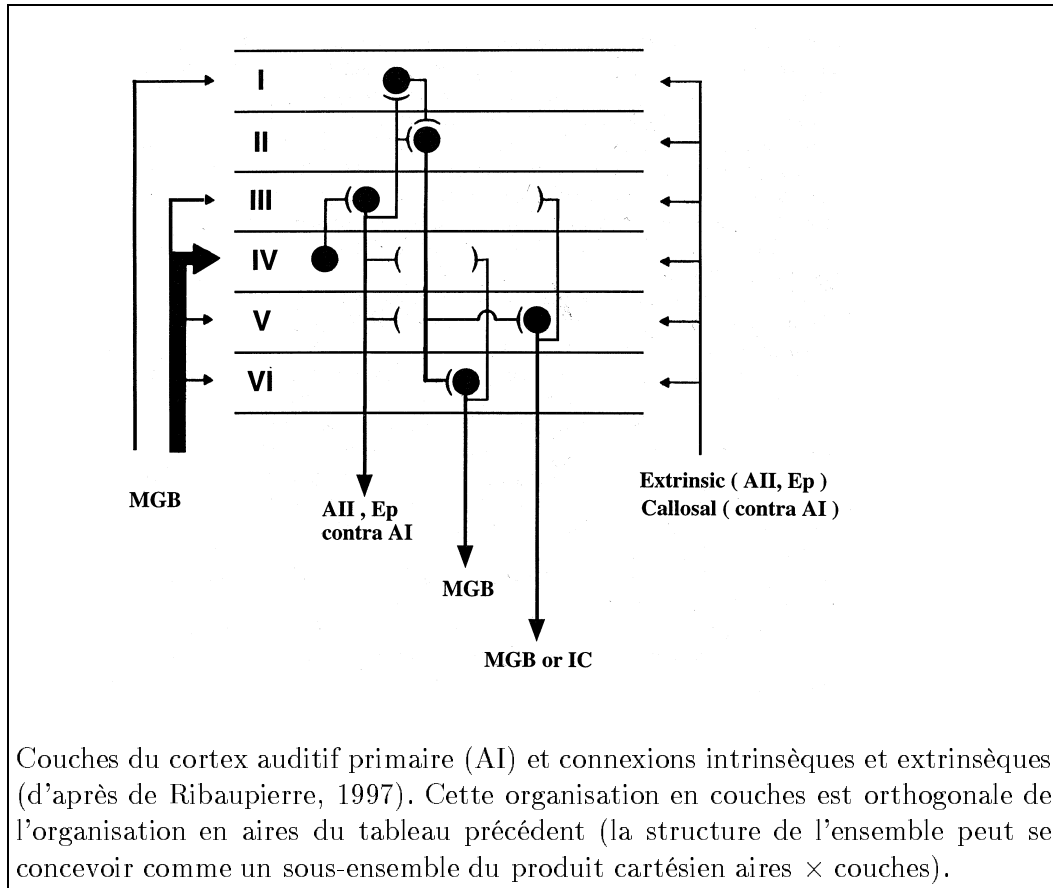
Nous avons parlé jusqu'ici des connexions ascendantes ou descendantes, mais il en existe aussi qui sont intrinsèques à un niveau hiérarchique. Elles sont nombreuses, et d'autant plus importantes qu'une deuxième hiérarchie de traitements semble exister à chaque niveau. Par exemple dans le noyau cochléaire on peut établir une hiérarchie AVCN -> PVCN -> DCN, les circuits et traitements au niveau de DCN étant plus complexes que ceux de VCN. Autre exemple, la distinction au sein de l'ensemble thalamus-cortex entre un système tonotopique (relativement "simple"), un système diffus, et un système multimodal, (Winer, 1992, Rouiller, 1997)].

Autre facteur de complexité, les échanges avec d'autres modalités sensorielles ou motrices, voire avec l'ensemble du système nerveux. Ils sont nombreux dès le colliculus inférieur (en relation étroite avec le colliculus supérieur), et massifs dans le thalamus et le cortex.

Des facteurs méthodologiques contribuent à brouiller notre compréhension du système. Ce sont par exemple les effets des anesthésiques utilisées dans les études physiologiques, la diversité des modèles animaux, et bien sûr l'impossibilité de faire certaines expériences chez l'espèce qui nous intéresse le plus, l'homme. Certains aspects sont constants chez tous les animaux étudiés (ou à l'intérieur de groupes plus restreints tels que les animaux à sang chaud, les mammifères, les carnivores, ou les primates). D'autres sont d'une grande variabilité, parfois entre individus d'une même espèce. Ainsi, chez l'homme le LSO, spécialisé dans le traitement des disparités binaurales d'intensité (et de temps d'attaque, Joris, 1998) serait-il peu développé, et les noyaux périolivaires seraient diffus (Kahle et al., 1979; Schwartz, 1992) [mais Covey, Vater et Casseday (1991) assurent que LSO est au contraire d'une grande stabilité phylogénétique. . .].

Un dernier facteur important de la complexité du système est le nombre d'éléments microscopiques en jeu à chaque étape. De Ribaupierre (1997) donne les chiffres suivants pour le néocortex : 40000 neurones/mm², 20000 synapses par neurone, dont 2000 inhibitrices, 9000 excitatrices locales, et





9000 excitatrices lointaines. À noter que la complexité des connexions ne se mesure pas seulement au nombre de synapses : il faut aussi tenir compte du fait qu'elles ont été choisies parmi un nombre plus important encore de connexions *possibles*. On tend parfois à voir dans chaque élément neuronal un élément de circuit au comportement assez simple. Cependant si on les considère selon une perspective évolutionnaire, comme les descendants des organismes unicellulaires qui, à un stade de l'évolution, ont convenu de se grouper, puis de se spécialiser pour assurer un rôle particulier au service du groupe (Gould, 1985), on peut s'attendre à un comportement sophistiqué de chaque élément. Une partie de la sophistication de comportement de leurs ancêtres unicellulaires, individus autonomes puis coopérateurs, a pu être conservé par les neurones, notamment lors de leur croissance.

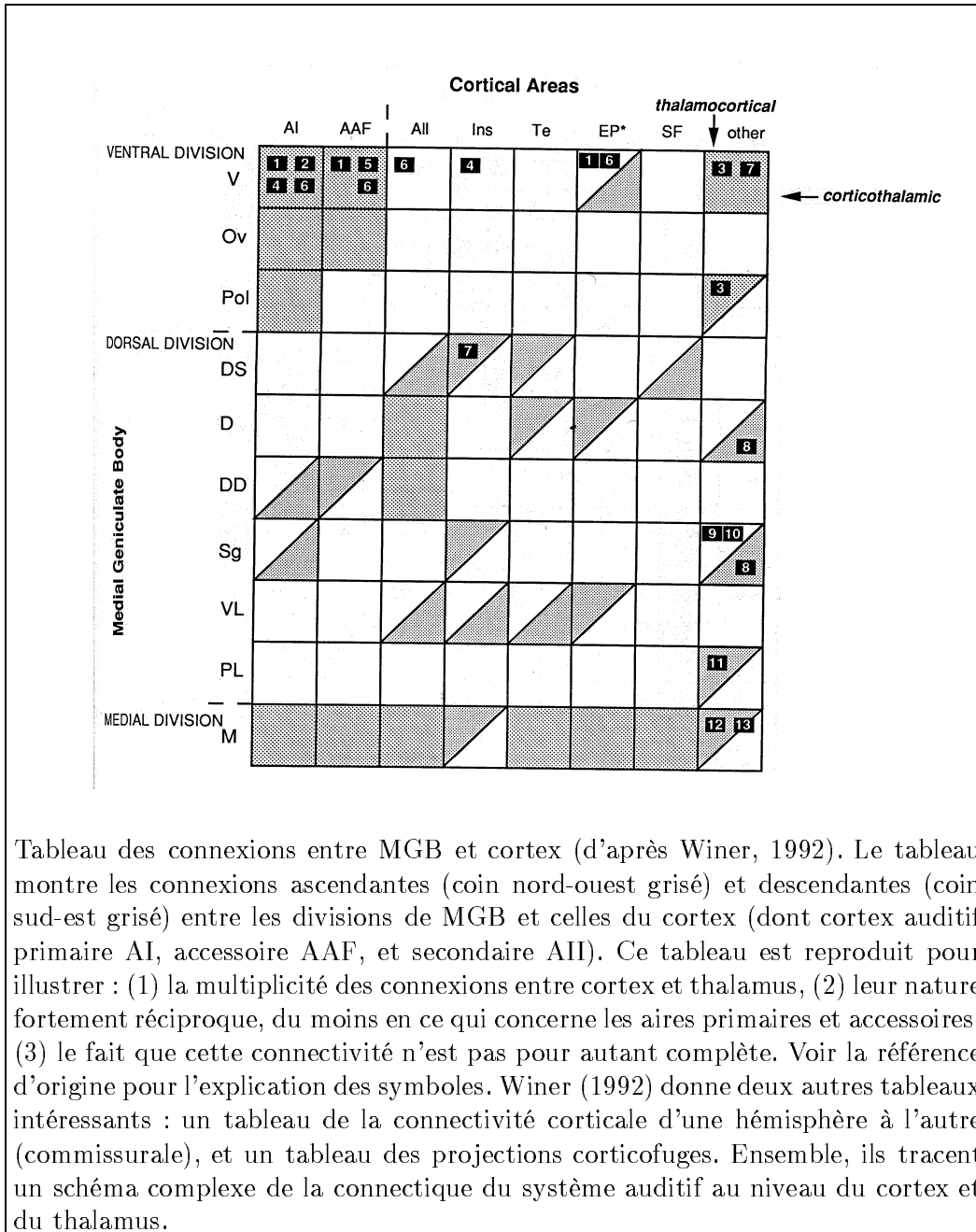
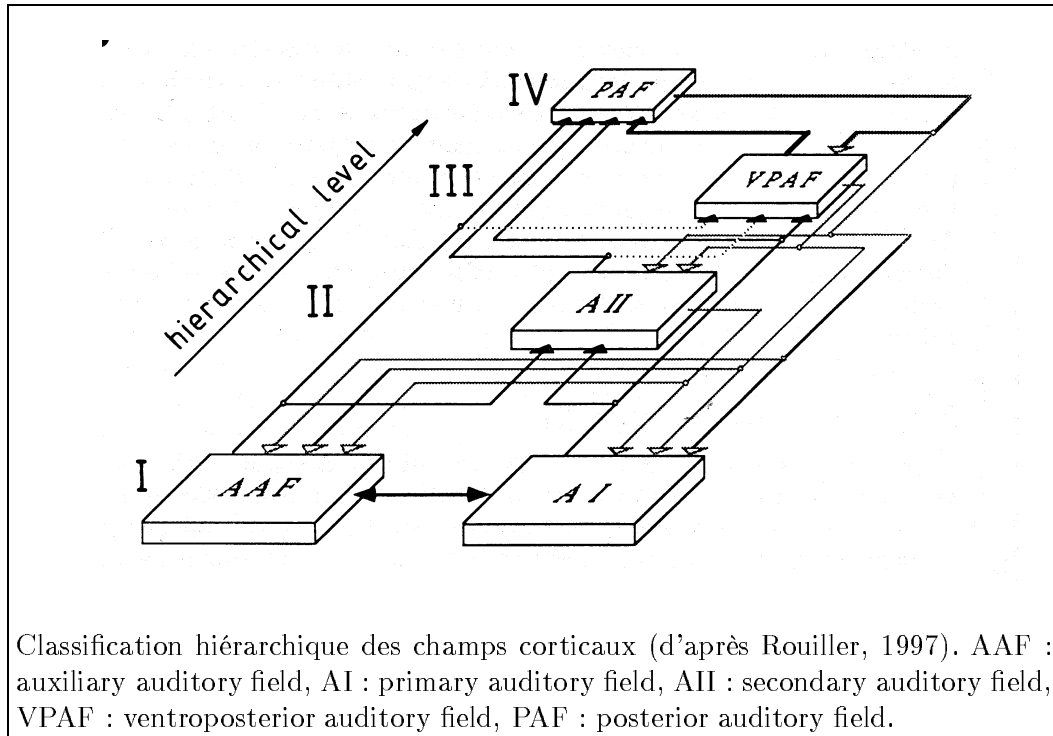


Tableau des connexions entre MGB et cortex (d'après Winer, 1992). Le tableau montre les connexions ascendantes (coin nord-ouest grisé) et descendantes (coin sud-est grisé) entre les divisions de MGB et celles du cortex (dont cortex auditif primaire AI, accessoire AAF, et secondaire All). Ce tableau est reproduit pour illustrer : (1) la multiplicité des connexions entre cortex et thalamus, (2) leur nature fortement réciproque, du moins en ce qui concerne les aires primaires et accessoires, (3) le fait que cette connectivité n'est pas pour autant complète. Voir la référence d'origine pour l'explication des symboles. Winer (1992) donne deux autres tableaux intéressants : un tableau de la connectivité corticale d'une hémisphère à l'autre (commissurale), et un tableau des projections corticofuges. Ensemble, ils tracent un schéma complexe de la connectique du système auditif au niveau du cortex et du thalamus.

2.5 Conclusion

Pour résumer, les étapes successives de notre description ont révélé une complexité considérable à chaque niveau de description. De nombreuses régularités existent, mais elles ont chacune de nombreuses exceptions, dont le



recensement fait obstacle à la compréhension du système. Il faut rendre hommage aux biologistes d'avoir mis au point des techniques d'exploration et de description capables d'appréhender une telle complexité. Dans les sections suivantes je reviens sur certains aspects de cette structure, à la recherche d'éléments qui confortent, ou au contraire affaiblissent, les hypothèses qui sous-tendent mes modèles. Je suis spécifiquement à la recherche d'indices de l'existence de mécanismes *temporels* de traitement du signal, en particulier utilisant l'inhibition. Je commencerai néanmoins par examiner les indices en faveur d'un traitement dans le domaine spectral, pour deux raisons. La première est que les modèles spectraux représentent une alternative qu'il faut garder présente à l'esprit. La deuxième est que, si l'essentiel du traitement se fait dans le domaine temps, il faut trouver une explication à l'ubiquité des représentations tonotopiques.

3 Tonotopie et traitement temporel

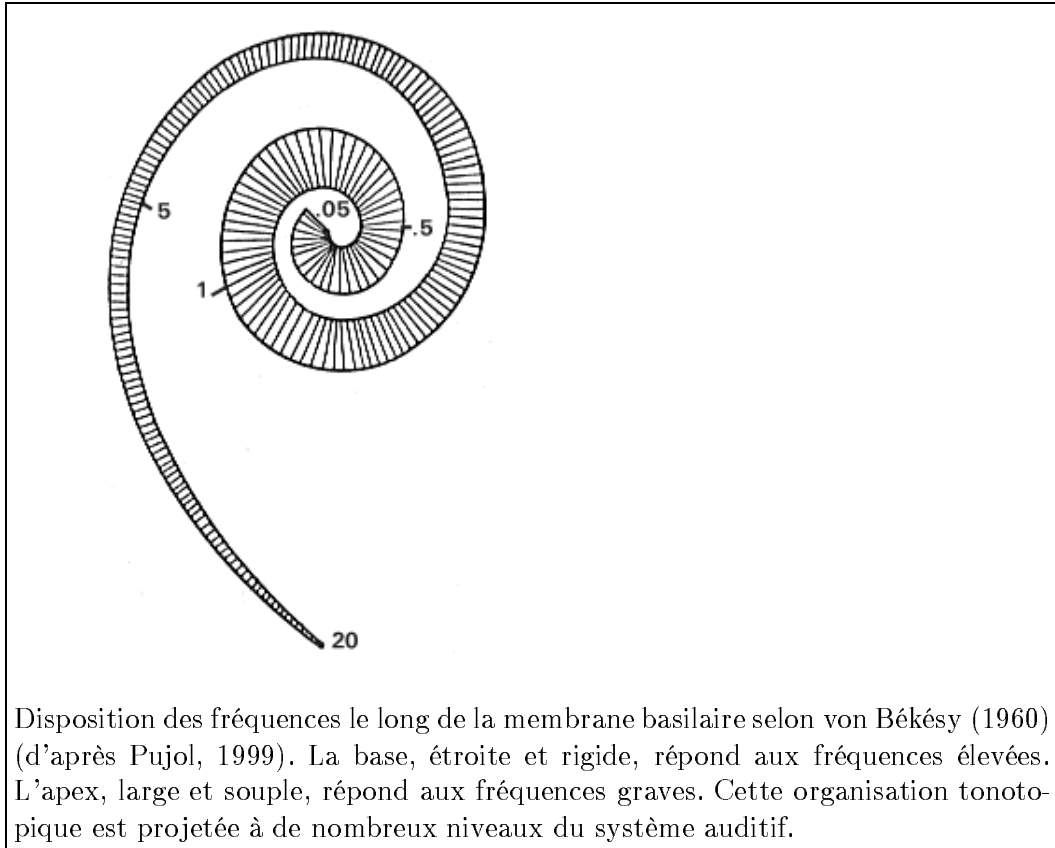
Le chapitre précédent décrivait le système auditif en allant du simple vers le complexe. Le présent chapitre recherche plus particulièrement les indices révélateurs de la nature des traitements auditif. Deux classes d'hypothèse sont en lice : traitements dans le domaine *fréquence*, s'appuyant sur une représentation tonotopique, et traitements dans le domaine *temps*, s'appuyant sur la structure temporelle des trains de décharges nerveuses.

3.1 La tonotopie dans le système auditif

L'idée de *tonotopie* (répartition spatiale des fréquences dans la cochlée, et reproduction de cette organisation à des niveaux centraux) est souvent attribuée à Helmholtz. Elle remonte en fait au moins à Du Vernay (1683) :

Enfin cette lame [*la lame spirale osseuse de la cochlée, à laquelle Du Vernay attribuoit les qualités résonnantes qu'on attribue aujourd'hui à la membrane basilaire*] n'est pas seulement capable de recevoir les tremblemens de l'air, mais sa structure doit faire penser qu'elle peut répondre à tous leurs caractères differens ; car estant plus large au commencement de la premiere revolution qu'à l'extremité de la derniere, où elle finit comme en pointe, & ses autres parties diminuant proportionnellement de la largeur, on peut dire que les parties les plus larges pouvant estre ébranlées sans que les autres le soient, ne sont capables que de fremiffemens plus lents qui répondent par consequent aux tons graves ; & qu'au contraire ses parties les plus étroites estant frappées, leurs fremiffemens sont plus viftes, & répondent par consequent aux tons aigus , de mesme les parties les plus larges d'un ressort d'acier sont des fremiffemens plus lents, & répondent aux tons graves , et que les plus étroites en sont de plus frequens & de plus viftes, & répondent par consequent aux tons aigus ; de sorte qu'enfin selon les differens ébranlemens de la lame spirale , les esprits du nerf , qui se répand dans la substance reçoivent différentes impressions qui representent dans le cerveau les diverses apparences des tons.

Les cellules ciliées répondent à des fréquences différentes en fonction de leur position le long de la cochlée. Cette organisation tonotopique est reproduite à tous les niveaux du système auditif central jusqu'au cortex avec un degré variable de précision. Pour qu'il y ait tonotopie, deux conditions doivent être remplies : (a) chaque cellule doit avoir une fréquence caractéristique bien définie, et (b) ces cellules doivent être disposées régulièrement dans l'espace en fonction de cette fréquence. La première condition est nécessaire pour que la seconde ait un sens, mais on peut parfaitement imaginer un



ensemble désordonné de cellules sélectives en fréquence, pour lequel le terme de tonotopie ne s'appliquerait pas.

La tonotopie est de qualité variable selon les niveaux, et il est intéressant de considérer de quelles façons cette qualité se dégrade. Une première façon est par élargissement de la courbe d'accord, par exemple par convergence (via une ou plusieurs étapes synaptiques) des activités d'une population de neurones de fréquences caractéristiques diverses. La convergence peut être justifiée par un traitement nécessitant une intégration large-bande, mais elle pourrait aussi être le résultat de simples imperfections de la connectique. Dans les cas extrêmes on observe des courbes d'accord à plusieurs vallées, voire de forme non-fonctionnelle (en forme de courbe fermée, etc.). Dans ces cas la notion de fréquence caractéristique n'a pas de sens univoque. Une deuxième forme de dégradation de la tonotopie est par perte de l'ordonnement linéaire, sans perte de sélectivité des cellules individuelles. La perte d'ordonnement linéaire ne signifie pas forcément perte de tout ordre spatial. Licklider (1959) souligne qu'un ensemble de cellules désordonné peut être remis en correspondance avec une dimension linéaire par auto-organisation

Hebbienne. À noter que la perte d'ordre spatial au niveau N risque d'accroître la perte de *sélectivité* au niveau $N + 1$, pour peu que la projection de N à $N + 1$ fasse intervenir une sommation de neurones proches.

D'autres facteurs peuvent affecter la qualité de la tonotopie. Aux niveaux supérieurs, la représentation de la fréquence entre en concurrence avec les représentations émergentes d'autres paramètres (intensité, disparités binoculaires, etc.), et le degré de sélectivité observé peut différer selon la valeur de ces paramètres. Il dépend de la présence d'anesthésie, et chez les animaux non-anesthésiés il dépend de l'état de veille (Clarey, 1992). Ainsi, la tonotopie régulière observée dans les MGB et cortex de l'animal anesthésié n'est plus évidente chez l'animal éveillé (de Ribaupierre, 1997).

La simple tonotopie ne suffit pas à la représentation précise d'un spectre. Il faut encore que l'activité puisse représenter l'*amplitude* précise du spectre à chaque fréquence. On a beaucoup glosé sur le fait que les réponses des fibres du nerf auditif ont une dynamique réduite, et saturent aux amplitudes élevées (Sachs et Young, 1979). Néanmoins, l'existence de populations de fibres avec des taux spontanés élevés et une dynamique large, associée à la redondance d'information offerte par le nombre élevé de fibres, pourrait permettre le codage précis de l'amplitude de chaque canal fréquentiel par le nerf auditif sur une large plage de niveaux (Delgutte, 1987). L'information spectrale pourrait alors être extraite à un niveau plus central. Au niveau du noyau cochléaire un type particulier de cellules, les cellules-étoile (stellate), semble spécialisé dans le codage précis de l'amplitude. Leur réponse, de type "chopper", a la forme de trains d'impulsions séparées par des intervalles de durée constante et indépendante de la structure temporelle du stimulus. La régularité des intervalles permet le codage précis d'une valeur scalaire. Une cellule de ce type pourrait permettre de coder par exemple une valeur du spectre. Ces cellules projettent de façon contralatérale sur VNLL, sur VNTB, LNTB et autres noyaux périolivaires, et sur IC (Oertel, 1999).

3.1.1 Répartition de la tonotopie

La cochlée est organisée tonotopiquement avec fréquences élevées à la base et basses à l'apex. Le nerf auditif a une structure qu'on peut interpréter comme résultant de l'enroulement sur lui-même d'un ruban de fibres organisé tonotopiquement. Les fibres de basse fréquence (provenant de l'apex de la cochlée) sont au centre, celles de haute fréquence (provenant de la base) à la périphérie.

L'organisation tonotopique se retrouve dans la projection du nerf auditif dans le noyau cochléaire (CN) : fréquences basses représentées ventralement, et hautes dorsalement (Cant, 1992). Ceci vaut tant pour les nombreuses

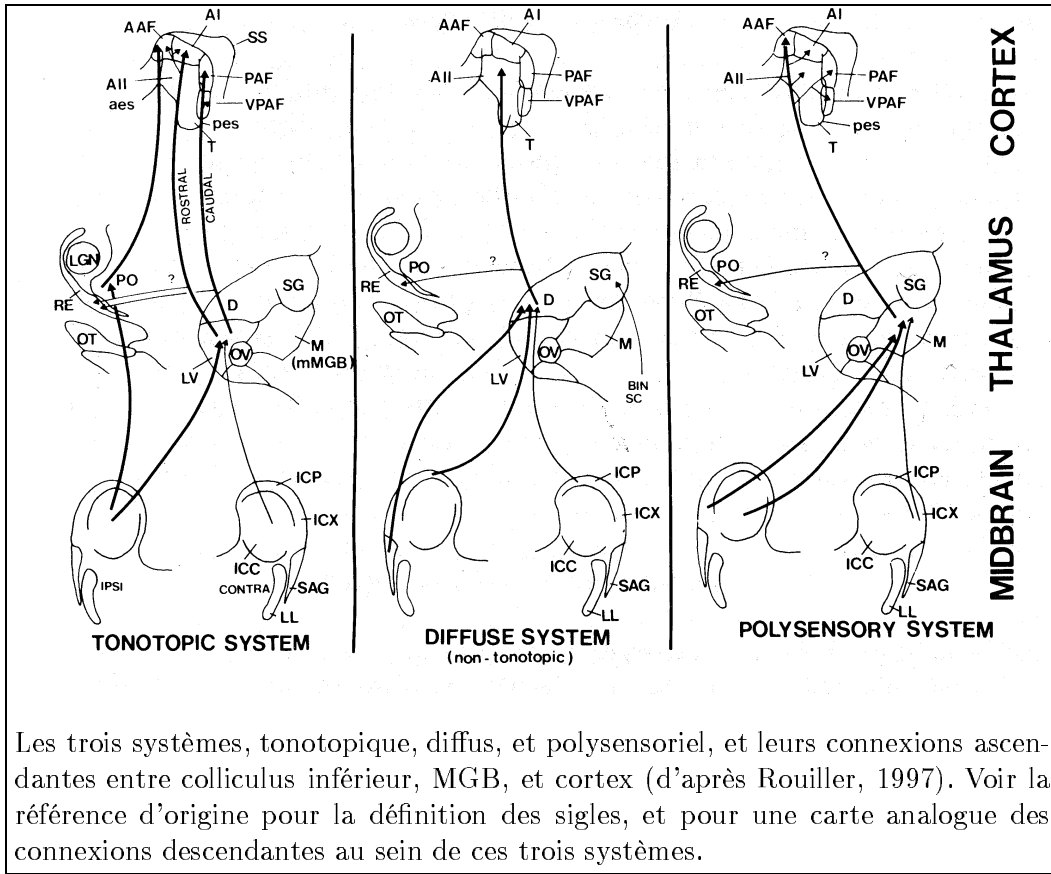
fibres de type I en provenance des CCI, que pour les quelques fibres de type II en provenance des CCE (Ryugo, 1992). La tonotopie est nette dans toutes les divisions de CN (AVCN, PVCN, DCN), quoique dans le DCN la complexité des traitements, et la complexité des courbes d'accord qui en résultent, rendent la définition d'une fréquence caractéristique plus difficile.

Au niveau du complexe olivaire supérieur, MNTB, LSO et MSO sont tous trois organisés tonotopiquement. Les hautes fréquences sont relativement mieux représentées dans MNTB et LSO (réputés spécialisés dans l'exploitation de différences binaurales d'intensité), et les basses fréquences mieux représentées dans MSO (spécialisé dans les disparités binaurales de temps d'arrivée). L'importance relative de ces différents noyaux varie d'ailleurs selon les espèces, en particulier selon que la taille de la tête favorise l'utilisation de l'un ou l'autre type de disparité binaurale (Heffner et Heffner, 1992).

Parmi les noyaux périolivaires, VNTB et DMPO sont organisés tonotopiquement (Helfert et Aschoff, 1997), et il en est de même des voies olivocochléaires efférentes médiane (MOC) et latérale (LOC), aussi bien en ce qui concerne leur projection vers la cochlée que celle vers CN. La tonotopie est absente ou inconnue dans les autres noyaux périolivaires.

Au niveau du lemniscus latéral (LL), le noyau dorsal (DNLL) est tonotopiquement organisé, mais non les noyaux ventraux (VNLL) et intermédiaires (INLL) dont les cellules ont des courbes d'accord larges (Helfert et Aschoff, 1997) (voir cependant Malmierca, 1998).

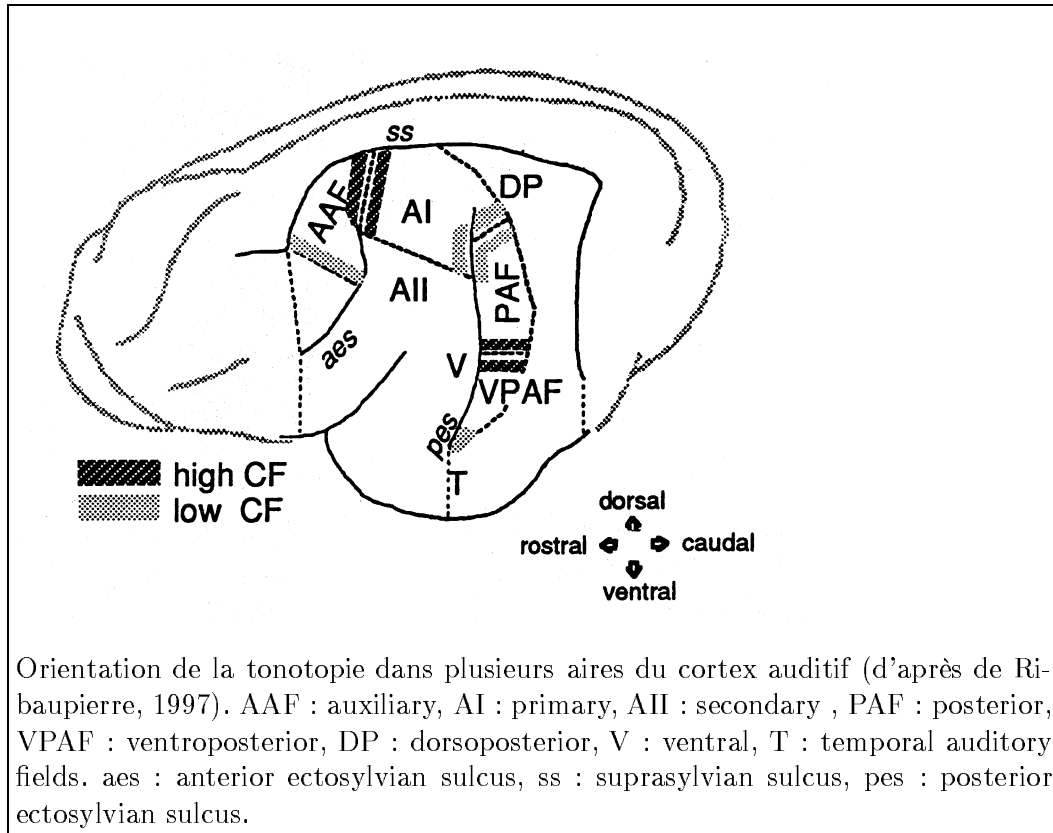
Au niveau du colliculus inférieur, le noyau central (ICC) est organisé tonotopiquement, avec une structure laminaire faite de lames isofréquence superposées, avec les fréquences basses situées de façon dorsolatérale, et les fréquences élevées de façon ventromédiane. Dans le plan d'une lame les cellules ont la même fréquence caractéristique, mais leur sélectivité serait meilleure au centre et progressivement moins bonne dans les zones concentriques qui l'entourent. Les lames étant trop épaisses pour permettre une représentation fréquentielle fine, il a été proposé qu'il existe en outre un gradient tonotopique fin dans le plan de chaque lame. La disposition en zig-zag des fréquences dans cet ensemble rappellerait le balayage d'un tube cathodique. Si cela est le cas, cette représentation tonotopique fine est en concurrence avec celle d'autres dimensions censées être codées selon des axes perpendiculaires à l'axe tonotopique : temps de latence, sensibilité, fréquence de modulation préférentielle (BMF), disparités binaurales, durée (Ehrlich et al, 1997), modulation de fréquence (Casseday et al., 1997), variations temporelles de disparité binaurale (Spitzer et Semple, 1998), etc., ce qui implique soit des corrélations entre ces dimensions, soit une sorte de "multiplexage" des représentations dont on comprend mal la nature (Irvine, 1992). Outre le noyau central ICC, le cortex dorsal du colliculus inférieur et la projection depuis le cortex auditif AC



Les trois systèmes, tonotopique, diffus, et polysensoriel, et leurs connexions ascendantes entre colliculus inférieur, MGB, et cortex (d'après Rouiller, 1997). Voir la référence d'origine pour la définition des sigles, et pour une carte analogue des connexions descendantes au sein de ces trois systèmes.

qui l'alimentent seraient organisés tonotopiquement (Oliver et Huerta, 1992). La tonotopie est absente (ou inconnue) dans les autres noyaux du colliculus inférieur.

Au niveau du thalamus et du cortex, Rouiller (1997) et de Ribaupierre (1997) distinguent trois systèmes : tonotopique, diffus, et polysensoriel. Il y a une ségrégation assez nette entre les trois : les connexions réciproques entre cortex et thalamus, ou commissurales d'un hémisphère cortical à l'autre, se font de préférence à l'intérieur du même système. Le premier système comprend, comme son nom l'indique, des zones organisées tonotopiquement dans le thalamus : la division ventrale du corps genouillé médian (vMGB) et la partie latérale du noyau postérieur (PO). Dans le cortex auditif il comprend les champs antérieur (AAF), primaire (AI) et postérieur (PAF). Chez l'animal anesthésié, l'organisation tonotopique est relativement stricte pour PO et la moitié antérieure de vMGB dans le thalamus, ainsi que pour AAF et AI dans le cortex. Elle est moins stricte pour PAF et la moitié postérieure de vMGB. Chez l'animal éveillé les courbes d'accord sont plus complexes et la disposition plus désordonnée. Le système diffus, par définition, n'est pas



tonotopique, mais dans le système polysensoriel on distingue une organisa-
tion tonotopique dans la partie antérieure de la division médiane du MGB
(mMGB).

Au niveau du cortex on rencontre des "cartes tonotopiques" multiples.
Leur nombre varie selon l'espèce, et aussi peut-être selon l'acharnement des
chercheurs à les trouver (Clarey, 1992). Une particularité est que des cartes
voisines sont généralement aboutées par la même extrémité de l'échelle des
fréquences, de sorte que la fréquence varie de façon continue lorsqu'on passe
d'une carte à l'autre, en changement simplement de sens de variation.

Lorsqu'on progresse de la périphérie au centre on rencontre donc des
marques d'une organisation tonotopique de façon répétée. Dans certains cas
la tonotopie est sous-tendue par une structure laminaire. Par exemple, dans
ICC le champ dendritique des cellules principales est en forme de disque
orienté parallèlement à la lame qui contient les cellules de même fréquence
caractéristique. L'épaisseur de la lame (70 à 150 μm) fait que l'ICC (épais-
seur 4 mm) ne peut contenir qu'un nombre limité de lames. Ehret (1997)
établit un parallèle entre les lames de ICC et les bandes critiques de la psy-
choacoustique, mais Brugge (1992) met en garde contre une interprétation

aussi simpliste. La structure laminaire révèle des bandes contiguës distinctes, ce qui ne correspond pas à la phénoménologie de la bande critique en psychoacoustique.

La tonotopie existe à de nombreux niveaux, de la cochlée au cortex. Cette ubiquité doit néanmoins être relativisée. Un, il existe de nombreuses zones non-tonotopiques, surtout aux niveaux supérieurs. Deux, une zone donnée peut comprendre divers types de cellules, dont certaines seulement sont organisées tonotopiquement (selon leurs degrés d'activité relatifs, ou la facilité d'enregistrement des différentes cellules, la zone pourrait apparaître comme strictement tonotopique ou au contraire dénuée de toute tonotopie). Trois, le degré de tonotopie dépend des critères adoptés, et peut varier selon les conditions expérimentales (amplitude des stimuli, anesthésie, etc.). Le chercheur est soumis à une pression qui l'encourage à constater la présence d'une tonotopie. Il explorera patiemment - c'est tout à son honneur - l'espace des paramètres jusqu'à en trouver les indices.

3.1.2 Spéculations sur le sens de la tonotopie

La multiplicité des cartes tonotopiques à travers le système auditif peut s'interpréter en faveur d'une théorie spectrale du traitement auditif. On peut aussi au contraire s'étonner de la répétition de cette représentation à des niveaux multiples : n'y a-t-il donc pas progrès d'une étape à l'autre, de récapitulation ? On ne s'attend pas à ce que cette représentation relativement primitive persiste jusqu'au cortex, au lieu de céder la place à des représentations plus abstraites. Il est intéressant de rechercher d'autres explications à l'ubiquité de la tonotopie dans le système nerveux auditif.

Une première hypothèse est que l'organisation tonotopique est un épiphénomène, manifestation remarquable d'un principe banal, par exemple le résultat de contraintes de croissance (Thompson, 1917). La tonotopie périphérique est liée à l'analyse fréquentielle de la cochlée, mais celle des niveaux supérieurs existerait simplement parce qu'il "eut été plus compliqué qu'il en soit autrement", et seul un Bernardin de Saint Pierre y chercherait un sens plus profond.¹ Cet argument est symétrique de celui de Whitfield (1978) pour qui c'est la structure *temporelle* des décharges nerveuses qui est épiphénomène.

Une deuxième hypothèse est que la tonotopie est utilisée pour guider l'organisation lors de l'ontogénèse, ou la réorganisation après un traumatisme (Clarey, 1992). La régularité spectrale de nombreux sons, associée à la résolution spectrale limitée de la cochlée, entraînent une corrélation de l'activité de

¹Bernardin de Saint Pierre voyait dans les marques sur la peau des melons, pratiques pour les partager en quartiers, la preuve de l'existence d'un Dieu bienveillant

canaux voisins qui pourrait faciliter le repérage des défauts d'ordre. Dans un système bien organisé un stimulus à bande étroite excite une population localisée, et toute exception à cette règle signale un défaut qui peut être corrigé par auto-organisation. Selon cette hypothèse, la tonotopie serait l'analogie de l'étiquetage des lignes dans un central téléphonique, ou de l'ordre alphabétique d'un dictionnaire. Une fois le câblage fait, le fonctionnement du système suivrait d'autres principes, mais la tonotopie pourrait subsister dans certains états particuliers. Cela pourrait expliquer par exemple la labilité de la tonotopie de vMGB chez l'animal éveillé. On sait en tout cas que la qualité de la tonotopie s'affine avec le développement (Sanes et Rubel, 1988).

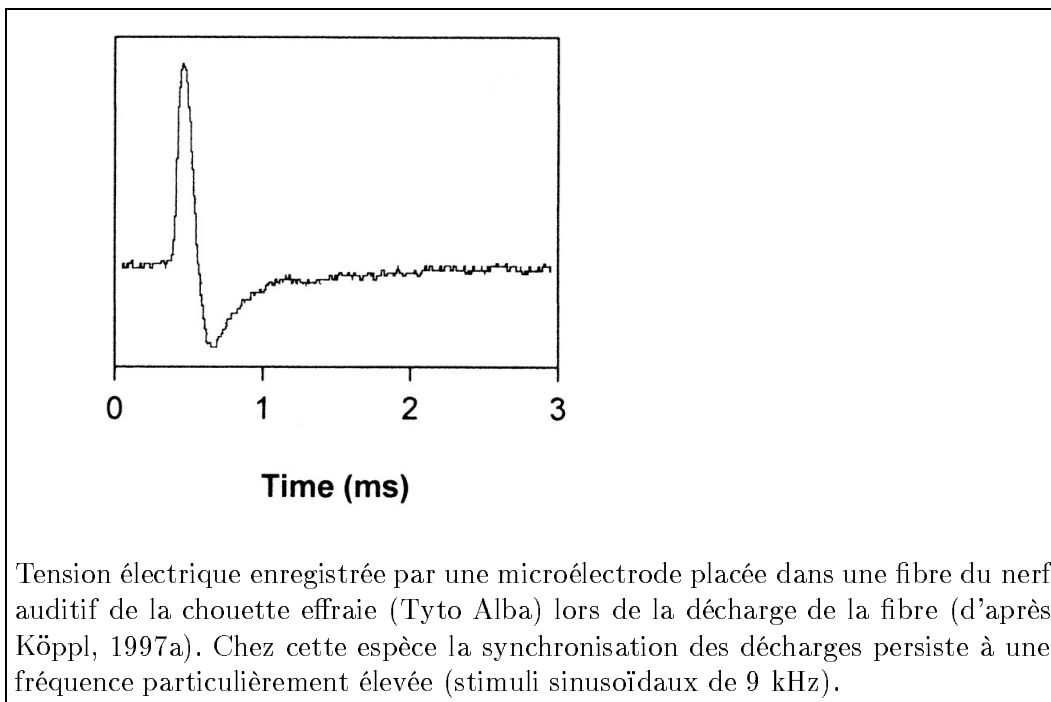
Une troisième hypothèse est que le système auditif traite essentiellement la *structure temporelle* de l'activité nerveuse issue de la cochlée, mais qu'il opère ce traitement en parallèle sur différents canaux fréquentiels. L'avantage, dans une optique de ségrégation de sources, est que les divers canaux diffèrent dans leur sensibilité à des bruits parasites (supposés concentrés dans une partie restreinte du spectre). Chaque canal peut alors être pondéré selon le degré de bruit qui l'affecte, c'est à dire selon la fiabilité des informations qu'il porte. Cette idée est développée dans la Sect. 7.2. La sélectivité observée jusqu'au cortex serait le support de ce mécanisme de ségrégation. Cette hypothèse n'implique pas forcément une disposition *régulière* (tonotopie), mais celle-ci pourrait s'expliquer soit comme épiphénomène (conséquence d'une croissance régulière), soit par l'utilité qu'il y aurait à pouvoir grouper des canaux de fréquence voisine à l'issue des traitements intra-canal.

Pour résumer, la tonotopie s'observe à tous les niveaux du système auditif, mais il n'est pas obligatoire d'y voir forcément la preuve d'une représentation purement spectrale de l'information auditive. Elle peut s'expliquer aussi comme servant à la ségrégation des informations provenant de sources concurrentes par la pondération différentielle de différents canaux fréquentiels. Néanmoins, l'hypothèse selon laquelle le traitement auditif opère sur des représentations spectrales reste largement acceptée.

3.2 L'information temporelle

3.2.1 Répartition dans le système auditif

Les décharges naissent dans les dendrites qui innervent les cellules ciliées internes, et se propagent le long des axones qui constituent le nerf auditif. Chaque impulsion a une durée totale de l'ordre de la milliseconde, mais sa forme est stéréotypée et son instant d'apparition peut être défini (sinon mesuré) avec une résolution bien plus fine, de l'ordre de la dizaine de mi-



crosecondes. La probabilité de décharge suit approximativement une version redressée (simple alternance) des mouvements de la membrane basilaire, ce qui imprime à la statistique des trains d'impulsions une structure temporelle qui reflète celle du son.

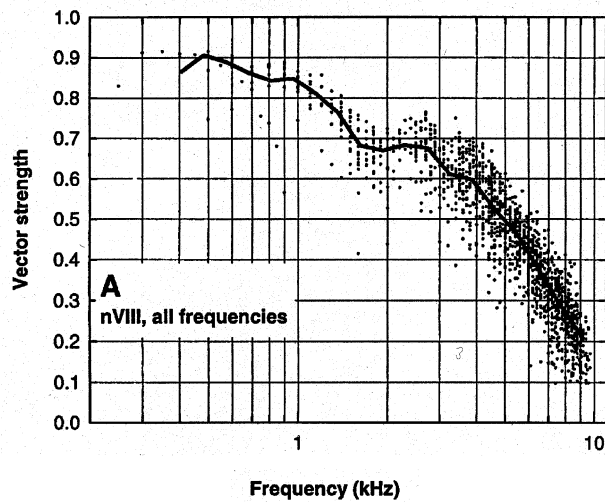
Chaque CCI est innervé par 10 à 20 fibres. Chez le mammifère la population n'est pas homogène (60-65% ont un taux de décharge spontané élevé, un seuil faible et une dynamique réduite, et les 35-40% qui restent un taux de décharge spontané moyen ou faible, un seuil élevé, et une dynamique large). On peut considérer que le mouvement en chaque point de la membrane est traduit de façon assez précise par un codage d'ensemble, version stochastique du codage "en volées" imaginé par Wever et Bray (1930a,b). Pour autant qu'on puisse imaginer ou mesurer (Johnson et Kiang, 1976), les processus stochastiques de fibres différents sont statistiquement indépendants, la seule corrélation étant celle résultant d'une excitation mécanique commune. On ne peut pas (ou difficilement) enregistrer ces fibres simultanément, mais on peut à la place enregistrer la même fibre à plusieurs stimuli identiques. Si on fait l'hypothèse raisonnable d'indépendance statistique entre présentations, le résultat obtenu est le même qu'on enregistre (a) une population de N fibres ou (b) la même fibre avec N répétitions du stimulus (Delgutte, 1996). La réponse d'une fibre à plusieurs stimuli répétés mesurée par le physiologiste remplace celle, quasi-impossible à mesurer, de plusieurs fibres semblables à

un stimulus unique.

La précision temporelle peut s'apprécier de plusieurs façons. La mesure la plus courante est le *coefficient de synchronisation* en réponse à une sinusoïde. Ce coefficient est mesuré à partir de l'histogramme de période, qui compte le nombre d'impulsions qui tombent à chaque phase de la période du stimulus. Un histogramme plat indique que les impulsions tombent au hasard, un histogramme modulé indique une bonne synchronisation. Le coefficient de synchronisation est le coefficient fondamental de la série de Fourier normalisé par division par le taux moyen. Il vaut 0.0 si l'histogramme est plat, et 1.0 s'il a la forme d'une impulsion. Dans le nerf auditif, le coefficient est élevé (0.8) pour les fréquences inférieures à 1 kHz. À partir de 1 ou 2 kHz le coefficient diminue, et tend vers zéro au delà de 4 à 6 kHz (Johnson, 1980). La perte de synchronie a plusieurs origines. Une première est une imprécision temporelle dans la production des décharges ("jitter"), ou leur transport (variation de vitesse de conduction). L'imprécision peut aussi avoir une composante instrumentale (variations de l'instant de déclenchement). Néanmoins l'origine la plus importante semble être le filtrage passe-bas opéré dans la cellule ciliée du fait de la capacité de membrane (Russel et Palmer, 1986). Ce filtrage diminue l'amplitude de la composante alternative du potentiel intracellulaire relativement à la composante continue, ce qui diminue la composante variable de la probabilité de décharge, relative à sa composante constante.

De la perte de synchronie au delà de 4-6 kHz, il résulte que les fibres dont la CF est plus élevée ne répondent pas de façon *synchronisée* aux fréquences auxquelles elles sont le plus sensibles. Elles peuvent néanmoins se synchroniser à des tons de fréquence plus basse, moyennant un niveau sonore plus élevé. Elles peuvent aussi se synchroniser à des *modulations d'amplitude* d'une porteuse dépassant les 4-6 kHz. La fréquence limite de réponse à ce type de modulation est néanmoins plus faible (d'environ 0.8 octave) que pour la synchronie à un son pur.

La perte de synchronie au delà de 4-5 kHz pose un problème aux modèles temporels. Si pour traiter des fréquences supérieures il faut faire appel à des mécanismes spectraux, l'hypothèse temporelle perd une partie de son intérêt. À cette critique on peut opposer deux sortes de réponse. La première est que le codage temporel est utilisable de 16 à 4000 Hz, soit huit octaves sur les dix que nous pouvons entendre. Ne manquent que deux octaves : le problème n'est pas si grave. Précisément, on observe en psychoacoustique une transition vers 4-5 kHz pour les phénomènes liés à la hauteur tonale (Demany, 1987, 1988). Au delà de cette limite, les seuils de discrimination s'élèvent et la perception du chroma se dégrade. D'autres phénomènes ont des limites plus basses. Hartmann (1990) a ainsi observé que la détection d'inharmonicité d'un partiel se dégrade rapidement entre 2 et 3.5 kHz. Quant



Coefficient de synchronisation en fonction de la fréquence mesurée chez la chouette effraie (d'après Köppl, 1997b). On observe un "plateau" entre 2 et 3 kHz qui est spécifique de cette espèce, et pourrait refléter les effets d'une spécialisation pour la synchronie aux fréquences élevées. On ne voit pas de plateau analogue dans les données pour les mammifères tels que le chat, pour lesquels la synchronie n'est pas mesurable au delà de 4-6 kHz. (Johnson, 1980).

aux phénomènes d'interaction binaurale, tels que le démasquage binaural (BMLD) ou les phénomènes de hauteur binaurale, ils ont une limite de l'ordre de 1 à 1.5 kHz (Hafter et Trahiotis, 1997).

La deuxième réponse fait valoir que la perte de synchronie au delà de 5 kHz n'entraîne pas nécessairement que la période correspondante (200 μ s) soit la limite de résolution temporelle. L'écart-type de la latence de réponse de certaines cellules de type "onset" du noyau cochléaire est inférieure à 100 μ s (Rhode et Greenberg, 1992). Chez l'homme, le seuil de disparité temporelle binaurale de 6 μ s (Irvine, 1992) laisse à penser que la limite de synchronie pourrait être due, non pas à une imprécision temporelle, mais à une difficulté de concilier précision et répétition régulière. On pourrait aussi spéculer que la synchronie s'affine par un mécanisme d'adaptation active (par exemple par contrôle efférent des CCE), qui pourrait ne pas être opérationnelle chez l'animal anesthésié. On observe chez certains oiseaux une limite de synchronie plus élevée, par exemple 9 kHz chez la chouette effraie (Carr et Konishi, 1990, Köppl, 1997a,b).

La limite de synchronie s'abaisse lorsqu'on s'élève dans le système auditif. Dans le noyau cochléaire, la limite est d'environ 3 kHz pour les cellules "primarylike" et "onset", et 1-2 kHz pour les "onset choppers" (Romand et

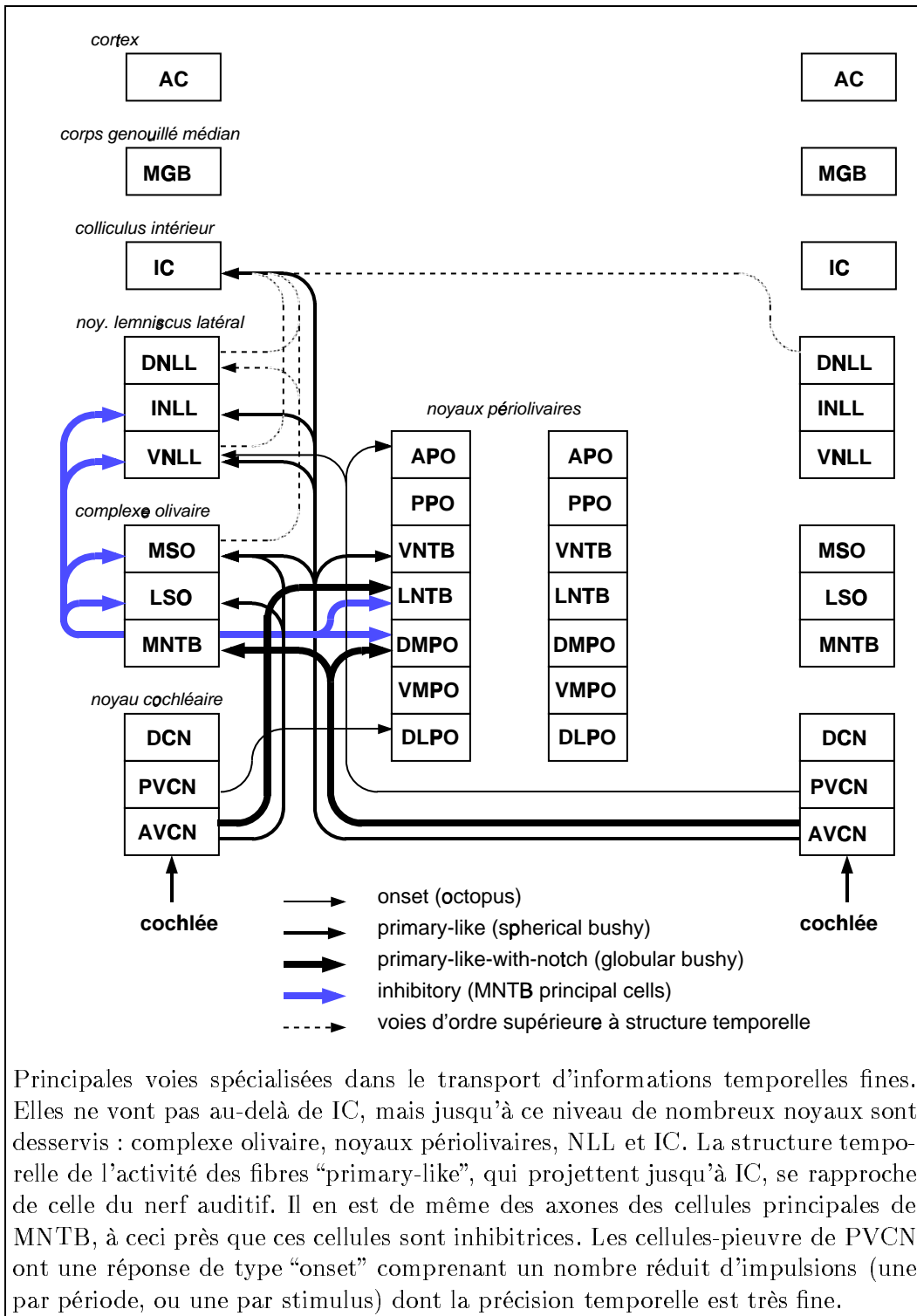
Avan, 1997).

Les cellules-pieuvre du PVCN ont une réponse de type "onset" qui a la forme simplifiée d'une impulsion unique par période. Cela se traduit par des coefficients de synchronisation proches de 1.0, plus élevés que dans le nerf auditif. On en déduit parfois qu'il y a "amélioration de la synchronie", mais il s'agit là d'une illusion puisque la concentration de la réponse en une impulsion unique traduit la perte de la structure fine intra-période. Les cellules-pieuvre du PVCN projettent via des axones myélinisés et synapses de type "calices de Held" sur le VNLL contralatéral, et noyaux périolivaires (Oertel, 1999). La projection se fait sur une sous-région de VNLL qui est particulièrement bien représentée chez l'homme. Les cellules cibles présumées (cellules globulaires) y représentent 38 % des cellules de VNLL, contre 4% chez le chat (Adams, 1997).

La synchronie des cellules du complexe olivaire (MNTB, LSO, MSO) est semblable à celle de leurs afférents du noyau cochléaire (cellules de type "bushy" du AVCN). Ces cellules opèrent un traitement de signal de nature instantanée, dont le résultat conserve généralement une structure temporelle fine. MNTB sert de simple relais qui transforme le signal exciteur des cellules "bushy" globulaires de AVCN en signal inhibiteur. MSO opère une détection de coïncidence (à base d'excitation), et LSO une détection d'anti-coïncidence (à base d'inhibition). Le résultat de ces traitements est lissé lorsqu'il est relayé dans le colliculus inférieur, où on constate la même forme de sensibilité à des stimuli binauraux qu'au niveau de SOC, mais sans la synchronie fine (Joris et Yin, 1998 ; Joris, 1996 ; Batra et al., 1989, 1997a,b). Dans le meilleur des cas, la synchronie qu'on observe ne dépasse pas 1.2 kHz dans le colliculus inférieur, et 1 kHz dans le thalamus (Rouiller, 1997). Dans le cortex la limite est plus basse, de l'ordre de 100 Hz (Clarey, 1992). Comme pour la tonotopie, les limites de synchronie dépendent des critères employés. Ceci est particulièrement vrai des niveaux supérieurs où les réponses sont complexes, diverses, labiles, et sensibles aux anesthésiques (Kuwada et al., 1989).

Comme pour la tonotopie, il existe une pression professionnelle pour "trouver de la synchronie" qui peut conduire à un biais en sa faveur. Inversement, plusieurs facteurs peuvent avoir conduit à une sous-représentation de la synchronie :

- L'anesthésie utilisée dans la plupart des expériences modifie les réponses temporelles, et peut masquer les effets d'inhibition, soit en l'augmentant comme dans IC (Kuwada et al., 1989 ; Batra et al., 1997), soit en la diminuant comme dans CN (Romand et Avan, 1997).
- La facilité d'enregistrement varie selon l'emplacement, la morphologie et les caractéristiques électriques des neurones ou fibres. La présence



- de grosses cellules sans synchronie peut masquer la présence de petites cellules ou fibres synchronisées dans une étude de population, ou lors d'un enregistrement de fibres multiples avec la même électrode. Dans certains cas des champs électriques parasites sont un obstacle, par exemple dans MSO ou CN (Yin et Chan, 1988 ; Joris et al., 1994).
- La synchronie présente à un niveau peut être projetée (donc présente) au niveau suivant sans forcément y être visible. Cela sera le cas par exemple si cette projection se fait via des fibres fines. On en conclue que la synchronie est *potentiellement disponible* aux étages immédiatement supérieurs à celui où elle est *constatée*.

En résumé, le système nerveux auditif comprend des voies spécialisées dans la transmission fiable de la structure temporelle fine. Ces voies desservent de nombreux niveaux entre la cochlée et le colliculus inférieur. La synchronie de la réponse à un son périodique est limitée à 4-6 kHz dans le nerf auditif, et cette limite diminue lorsqu'on s'élève dans le système. Cependant la période correspondante (200 μ s pour le nerf auditif) ne définit pas forcément la limite de résolution temporelle d'événements non-répétitifs. La section suivante détaille certaines spécialisations qui semblent motivées par la nécessité d'un transport ou traitement précis des informations temporelles.

3.2.2 Spécialisations en faveur du traitement temporel

Sur le plan anatomique, certains traits semblent être associés à une spécialisation pour la transmission temporelle. Les *calices de Held* et synapses assimilées assurent une transmission sûre et temporellement précise de chaque impulsion. Ces synapses sont observées dans AVCN entre fibres du nerf auditif et cellules de type "bushy", et entre les fibres issues de ces cellules et les cellules-relais du MNTB (Joris, 1996). Les cellules de MSO et LSO reçoivent aussi des synapses sûres de ces fibres, et les cellules de LSO reçoivent en outre des synapses sûres des cellules de MNTB (Schwartz, 1992). Cette spécialisation synaptique pour le traitement temporel permet aux noyaux du complexe olivaire supérieur d'exploiter les disparités temporelles binaurales.

Des synapses de type calice sont signalées aussi dans le noyau ventral du lemnicus latéral (VNLL), à l'extrémité des fibres provenant des cellules "pieuvre" (octopus cells) du noyau cochléaire (PVCN) et des cellules-relais du MNTB (Schwartz, 1992). Des cellules spécialisées dans la transmission temporelle fine (morphologie semblable aux cellules "bushy", innervation par une synapse de grande surface) sont aussi signalées dans le sous-noyau postéro-ventral du noyau latéral du corps trapézoïdal (pvLNTB) (Spirou et Berrebi, 1996 ; Spirou, Rowland et Berrebi, 1997 ; Rowland et Spirou, 1998 ; Helfert

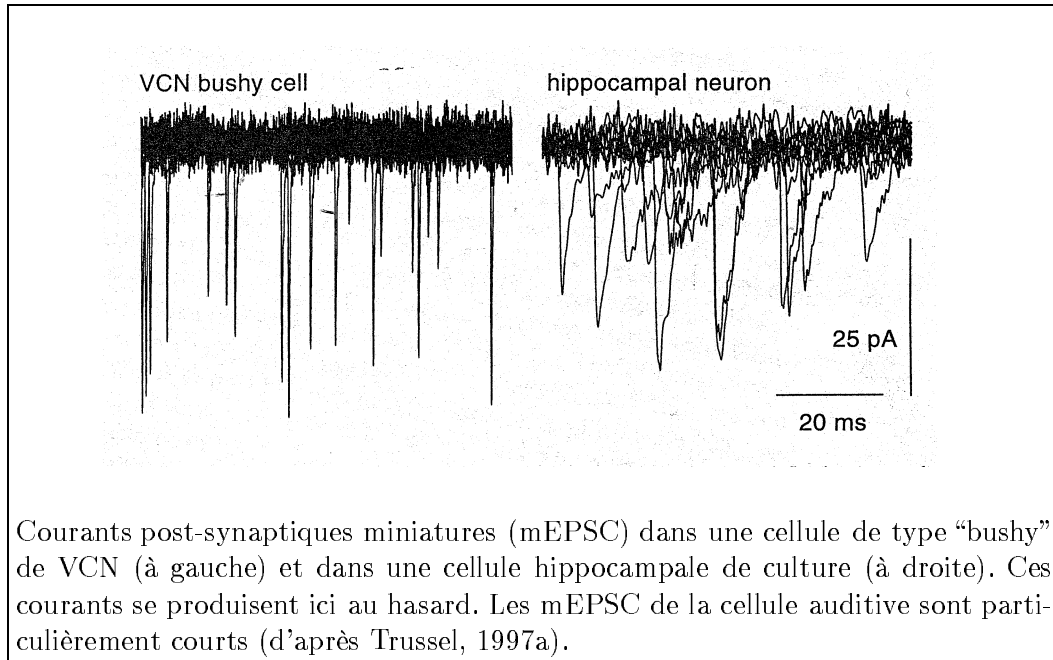
et Aschoff, 1997).

Les axones myélinisés et de fort diamètre transportent les impulsions avec un temps de latence court (et donc sans doute peu variable). C'est le cas d'au moins quatre groupes de fibres :

- Les fibres de type I du nerf auditif projettent sur les cellules de type "bushy" du noyau cochléaire.
- Les fibres issues des cellules de type "bushy" projettent vers MSO, LSO, et surtout MNTB (Joris, 1996), comme noté ci-dessus. Après branchement ces mêmes fibres projettent sur certains noyaux périolivaires (LNTB, VNTB, NCAT) (Schwartz, 1992, Romand et Avan, 1997), le noyau ventral du lemniscus latéral (VNLL), et le noyau central du colliculus inférieur (ICC).
- Les cellules principales de MNTB projettent sur LSO, comme noté ci-dessus, et aussi sur VNLL et les noyaux périolivaires DMPO, VMPO (Schwartz, 1992), LNTB, VNTB, et MSO (Helfert et Aschoff, 1997). La voie efférente olivocochléaire médiane qui part de MNTB vers le CN et la cochlée emprunte aussi des fibres myélinisées (Warr, 1992).
- Les fibres issues des cellules "pieuvre" du PVCN qui projettent dans VNLL sont myélinisées (Schwartz, 1992). Elles projettent aussi dans les noyaux périolivaires APO, DMPO et DLPO (Helfert et Aschoff, 1997).

La transmission précise d'informations temporelles fait intervenir des mécanismes membranaires post-synaptiques spécialisés qui (a) limitent dans le temps la durée d'un EPSP supraliminaire de façon qu'il ne produise qu'une seule et unique décharge, (b) stabilisent le temps de latence entre l'événement pré-synaptique et la décharge, pour éviter l'estompage des patterns temporels, et (c) raccourcissent le temps de récupération de la cellule pour que des décharges puissent se succéder à une cadence rapide (Brew et Forsythe, 1995, Sabatini et Regehr, 1999).

Ces mécanismes membranaires s'appuient sur des canaux à potassium (K^+) spécialisés. Des canaux vulnérables au 4-aminopyridine (4-AP) assurent à la membrane des propriétés non-linéaires dont notamment une forte augmentation de la conductivité en réponse à une dépolarisation. Ces canaux sont de deux sortes au moins, qui se distinguent par leur cinétique et leur sensibilité à des toxines spécifiques (Brew and Forsythe, 1995; Smith, 1995; Banks et Smith, 1992). Les fortes conductivités assurent des constantes de temps réduites, mais il en découle que des courants élevés sont nécessaires pour que soit atteint le seuil (en potentiel) de décharge. Cela expliquerait la taille exceptionnelle des synapses spécialisées du type calice de Held, capables de lâcher une dose massive de neurotransmetteur (Oertel, 1983). Ces caractéristiques membranaires post-synaptiques se rencontrent dans les cellules de type "bushy" et "octopus" du noyau cochléaire (Wu et Kelly, 1992),



dans les cellules principales de MNTB (Brew and Forsythe, 1985, Fonseca et Hallows, 1998) et de MSO (Smith, 1995), et dans les cellules globulaires de VNLL (Wu, 1999).

Banks et Smith (1992) rapportent que, dans les cellules principales de MNTB, les constantes de temps varient avec le potentiel de repos : une dé-polarisation raccourcit l’événement post-synaptique, tout en rapprochant la membrane de son seuil de décharge. Le potentiel de repos dépend de l’équi-libre des afférences. Hormis les calices de Held, le corps et les dendrites des cellules principales de MNTB reçoivent d’autres synapses excitateurs et in-hibiteurs. On pourrait spéculer que la résolution temporelle est affinée par un mécanisme qui ajuste le potentiel de membrane à la limite de la dé-charge spontanée, dans la zone où la cinétique est la plus rapide (Kopp, 1998 ; Köppl, 1997b ; Yang et al., 1999). La sensibilité éventuelle de ce mécanisme à l’anesthésie diminuerait les chances d’observer des patterns de coïncidence normaux.

Outre les caractéristiques post-synaptiques, la cinétique des canaux à calcium dans la cellule présynaptique est importante pour une transmission synaptique rapide et temporellement précise (Sabatini et Regehr, 1999). Le neurotransmetteur des synapses excitatrices rapides est probablement le glu-tamate, et les récepteurs correspondants de type AMPA (α -amino-3-hydroxy-5-methyl-4-isoazolepropionate), en particulier l’isoforme D (GluR-D_{flop}), la plus rapide (Trussel, 1997a,b). Le neurotransmetteur des synapses inhibi-

trices rapides est probablement la glycine (Trussel, 1997b). Le neurotransmetteur est lâché sous forme de vésicules élémentaires, qui produisent chacune un potentiel postsynaptique “miniature” (mEPSP). Dans la limite d’une approximation de sommation linéaire, l’EPSP peut se concevoir comme la convolution du mEPSP avec un train d’impulsions décrivant les temps d’arrivée des vésicules sur la membrane post-synaptique (Sabatini et Regehr, 1997 ; Trussel, 1997a). Pour que le EPSP soit temporellement précis, il faut que le mEPSP soit court *et* que les vésicules arrivent en groupe serré (Chuhma et Ohmori, 1998).

En résumé, au traitement et le transport d’informations temporelles précises sont associées des propriétés particulières des cellules, axones et synapses. La rapidité et la fiabilité de transmission synaptique fait intervenir des propriétés particulières des membranes pré- et post-synaptiques. Ces propriétés se rencontrent à des niveaux multiples dans le tronc cérébral, et débordent largement les limites traditionnellement assignées au traitement temporel binaural (circuit AVCN-MSO du modèle de Jeffress). Malgré la présence massive de ces structures spécialisées, le neurophysiologiste hésite sur le sens à accorder à l’information temporelle dans le système auditif. Il est tributaire pour cela de l’existence de *modèles*. La prochaine section recense plusieurs modèles qui exploitent la structure temporelle, ainsi que les données anatomo-physiologiques éventuelles qui les confortent.

3.2.3 Modèles

Coïncidence et traitement corrélatif Deux modèles se distinguent dans l’histoire des modèles de traitement auditif temporel : le modèle de localisation de Jeffress (1948) et le modèle de hauteur tonale de Licklider (1951, 1956, 1959, 1961). Tous deux s’appuient sur le calcul de *fonctions de corrélation*, à l’aide de réseaux de lignes à retard et détecteurs de coïncidence. Le premier de ces deux modèles calcule la fonction de corrélation binaurale, le second la fonction d’autocorrélation monaurale. Chacun a donné naissance à une famille de modèles dérivés, enrichis plus tard de modèles de *ségrégation de sources* binaurale (Lyon, 1983) ou monaurale (Weintraub, 1985, Meddis et Hewitt, 1992) qui ont inspiré à leur tour des modèles récents d’“analyse de scène auditive computationnelle” (CASA) (Cooke, 1991 ; Brown, 1992, [22]). Il a aussi été proposé que l’autocorrélation serve de support à la reconnaissance des timbres de voyelles (Meddis et Hewitt, 1992 ; [12]). La corrélation est donc un mécanisme d’une grande importance théorique.

Le modèle de Jeffress (1948) a été largement conforté par les découvertes de l’anatomie et de la physiologie. Sur le plan de l’anatomie, un réseau régulier

de "lignes à retard" alimentant des "compteurs de coïncidence" a été mis en évidence dans le nucléus laminaris (homologue du MSO) de plusieurs espèces de hibou (Konishi et al., 1988 ; Carr et Konishi, 1990 ; Irvine, 1992 ; Volman et Konishi, 1989), ainsi que dans le MSO du chat (Schwartz, 1992 ; Yin et Chan, 1990 ; Smith, Joris et Yin, 1993). Sur le plan physiologique, des enregistrements faits dans MSO ont révélé des réponses tout à fait conformes à celles d'un compteur de coïncidences (Goldberg et Brown, 1992 ; Yin et Chan, 1990 ; Irvine, 1992 ; Batra et al., 1997).

Les cellules principales de MSO reçoivent bilatéralement des fibres en provenance des SBC des noyaux cochléaires. Chez le chat, la voie contralatérale est disposée de façon à introduire un gradient de retard, croissant selon l'axe rostral-caudal, le retard ipsilatéral étant fixe (Smith, Joris et Yin, 1993). La différence de retard interaurale suit donc un axe anatomique qu'on peut assimiler à la variable τ de la fonction de corrélation croisée du modèle de Jeffress :

$$CCF(\tau) = \int_{-\infty}^{+\infty} i(t)c(t - \tau)dt \quad (2)$$

La réponse des cellules principales de MSO présente une structure temporelle fine (Goldberg et Brown, 1969 ; Spitzer et Semple, 1998) qui reflète la valeur instantanée du produit $i(t)c(t - \tau)$ des probabilités de décharge des voies afférentes ipsilatérale $i(t)$ et contralatérale $c(t - \tau)$. L'étape d'intégration temporelle qui suit ce produit (l'intégrale dans Eq. 2) n'est pas complète au niveau de MSO, mais intervient semble-t-il au niveau de IC où les cellules principales de MSO projettent. Le modèle de Jeffress a été confirmé chez de nombreuses espèces, directement au niveau de MSO (Goldberg et Brown, 1969 ; Yin et Chan, 1990 ; Smith, Joris et Yin, 1993) ou indirectement au niveau de IC (Yin, Kuwada et Sujaku, 1984 ; Chan, Yin et Musicant, 1987 ; Yin, Chan et Carney, 1987 ; Konishi et al., 1988 ; Yin et Chan, 1988 ; Kuwada, Batra et Stanford, 1989 ; Carney et Yin, 1989 ; Volman et Konishi, 1989 ; Carr et Konishi, 1990 ; Kuwada et al., 1990).

Il faut cependant noter quelques aspects qui s'éloignent du modèle de Jeffress. L'innervation bilatérale en provenance des SBC du noyau cochléaire se complète d'une innervation inhibitrice, bilatérale elle aussi, via MNTB (alimenté par les GBC contralatéraux) et LNTB (alimenté par les GBC ipsilatéraux). Quoique de nombreuses cellules répondent de manière conforme au modèle de Jeffress, d'autres reflètent des contributions inhibitrices (Grothe, 1994 ; Grothe et Park, 1998).

Contrairement au modèle de Jeffress, le modèle de Licklider a reçu peu de confirmation directe. Il serait cependant prématuré de le rejeter pour les raisons suivantes : (a) Il n'existe pas d'autre modèle qui soit aussi plausible physiologiquement. (b) Phénoménologiquement, il rend compte d'une

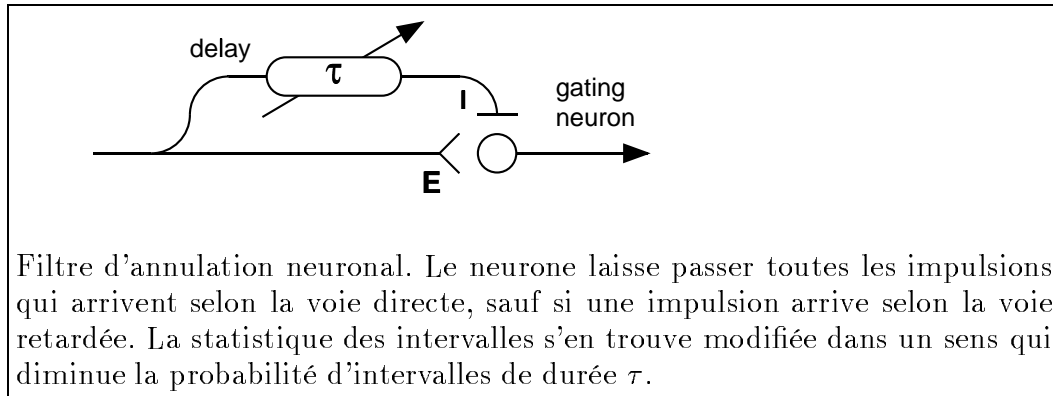
large classe de phénomènes (Meddis et Hewitt, 1990a,b ; Cariani et Delgutte, 1996 ; Yost, 1996 ; [8]). (c) Les difficultés d'exploration du système auditif font qu'il ne faut pas exagérer le poids de l'*absence* d'indices en faveur d'un modèle, surtout si on admet ce dernier puisse prendre dans son implémentation une forme inattendue (par exemple remplacement de la coïncidence par l'anticoïncidence[8]). Casseday et al. (1995) soulignent que chez la chauve-souris les structures de VNLL ressemblent à celles requises par le modèle de Licklider.

Anticoïncidence et inhibition L'inhibition se manifeste partout dans le système auditif. On lui attribue souvent un rôle diffus d'ajustement de seuils (Yang et al, 1999), de renforcement de contraste spectral ou temporel, de suppression de réponse, bref un rôle de "faire-valoir" d'un signal excitatif. Il n'est pas courant de voir attribuer à une activité inhibitrice un rôle de signal, et à ses effets (suppression d'activité) un rôle de pattern précis dans le temps ou l'espace. Des mesures physiologiques in vitro et in vivo montrent que l'interaction entre excitation et inhibition au sein d'une cellule peut être très rapide (Joris et al., 1995 ; Joris, 1996 ; Batra et al., 1997 ; Zacksenhouse et al., 1995 ; Finlayson et Caspary, 1991 ; Wu et Kelly, 1991, 1992 ; Sanes, 1990, Tsuchitani, 1997). Les propriétés de membrane assurent une dépendance fine de l'intégration d'impulsions excitatives et inhibitrices sur leur timing relatif (Sanes, 1990 ; Wu et Kelly, 1991 ; Zacksenhouse et al., 1995 ; Batra et al., 1997).

Dans [8] je propose un modèle de hauteur tonale à base d'annulation, version "en négatif" du modèle de Licklider. La coïncidence est remplacée par l'*anticoïncidence*, et les maxima cèdent la place aux *minima* dans le rôle d'indice de la périodicité. Les deux modèles sont formellement identiques en première approximation (en deuxième approximation le modèle annulation offre certains avantages).

Le modèle précédent est construit à partir d'un "filtre d'annulation neuronal", qui s'est révélé efficace pour expliquer la ségrégation des sons sur la base de leur structure harmonique [14, 12].

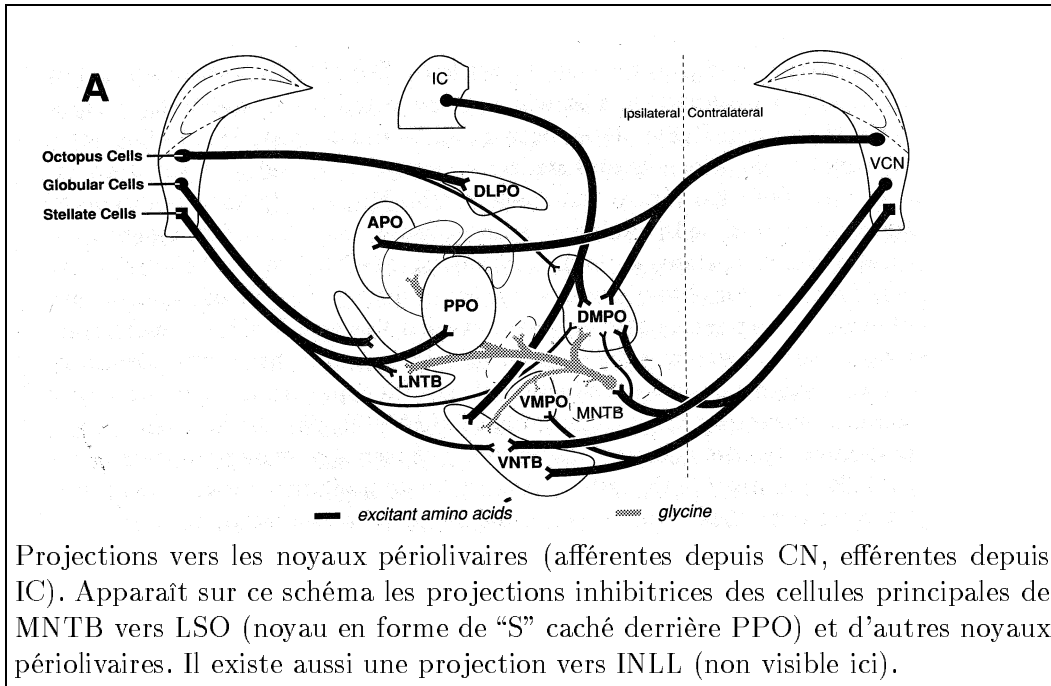
L'inhibition peut aussi être invoquée pour les interactions binaurales. Le rôle de LSO dans l'exploitation des disparités binaurales d'amplitude est bien connu. A priori on imaginerait que ce mécanisme puisse se contenter d'une interaction lente, mais paradoxalement, comme noté plus haut, les circuits qui alimentent LSO (via AVCN et MNTB) sont spécialisés pour une transmission temporelle fiable, avec fibres myélinisées épaisses et calices de Held (Joris et al., 1994a,b, 1995 ; Schwartz, 1992). Cette spécialisation temporelle peut s'expliquer si on étend les compétences de LSO à la détection de disparités in-



teraurales d'*instant d'attaque* (Goldberg et Brown, 1969 ; Joris et Yin, 1998), qu'on peut interpréter comme une disparité d'intensité pendant un temps très court. Cette hypothèse explique la spécialisation temporelle du circuit qui alimente LSO, tout en préservant son rôle classique dans la comparaison d'intensités. Certaines cellules de type IE de LSO (et aussi MSO : Goldberg et Brown, 1969 ; Finlayson et Caspary, 1991 ; Joris et al. 1995 ; Joris, 1996 ; Batra et al., 1997 ; Sanes et Rubel, 1988) sont sensible à des disparités de *phase*. On peut attribuer cette sensibilité à un mécanisme qui calculerait une sorte de "fonction d'anticorrélation", version en négatif du modèle de Jeffress, avec des creux à la place des pics. On peut formuler un tel modèle de façon à ce qu'il soit formellement équivalent à celui de Jeffress, en remplaçant l'opération de multiplication par une opération de soustraction (comme je l'ai fait pour le modèle de Licklider [8]). Il est cependant difficile d'expliquer la présence de deux mécanismes, l'un "positif" dans MSO, et l'autre "négatif" dans LSO, formellement équivalents et dont la fonction serait la même.

Autre modèle classique, le modèle "Equalization and Cancellation" de Durlach (1963) explique la réduction de masquage binaurale (BMLD) par une interaction soustractive. Dans ce modèle, les signaux provenant des deux oreilles sont *égalisés* en appliquant un retard et un facteur multiplicatif, puis *annulés* par soustraction. La présence éventuelle d'un résidu d'annulation traduit alors la présence d'une cible. Le LSO a été proposé comme site de cette interaction (Itoh et al., 1982 ; Itoh, 1983, 1984). Dans le modèle EC de Durlach, les paramètres d'égalisation sont déterminés globalement, et appliqués de façon uniforme à tous les canaux fréquentiels. Un progrès récent est la démonstration par Culling et Summerfield (1995 ; Culling et al., 1998a,b) que le système auditif peut déterminer et appliquer les paramètres de façon *indépendante dans chaque canal fréquentiel*.

Le principe d'anticorrélation est d'un grand attrait théorique et il est intéressant de lui chercher un corrélat physiologique associant les ingrédients



retard, coïncidence et inhibition.

Sources d'inhibition temporellement fine Le MNTB semble spécialisé dans la production d'un signal inhibiteur d'une grande précision temporelle. Il est exploité par LSO dans le mécanisme binaural cité plus haut, mais cela ne lui interdit pas d'autres rôles du même type, par exemple dans un mécanisme monaural tel que le filtre d'annulation [14, 12]. Joris et Yin (1995) notent que chez le chat dans MNTB les cellules de basse fréquence caractéristique existent, mais la densité de leur projection sur LSO est faible, ce qui suppose qu'elles sont utilisées ailleurs. Comme nous l'avons noté plus haut, MNTB projette sur de nombreux noyaux. Dans le complexe olivaire, outre LSO et lui-même, il projette sur MSO, DMPO, VMPO, VNTB et LNTB (Helfert et Aschoff, 1997). Centralement il projette sur INLL et VNLL, et périphériquement sur VCN et les cellules ciliées externes de la cochlée, via la voie efférente médiane (MOC) (Spangler, Warr et Henkel, 1985). L'existence de projections vers la cochlée justifie une conjecture attrayante mais improbable : que l'interaction soustractive se fasse mécaniquement dans la cochlée. Attrayante par la relative linéarité, et surtout la dynamique pratiquement illimitée d'une interaction mécanique. Improbable, car, comme noté plus haut, les enregistrements faits dans les fibres du système MOC n'ont jusqu'à présent révélé que des décharges de type "chopper", avec des ISI ré-

gouliers, et non une structure temporelle liée au signal (Rouiller, 1998, Brown et al., 1998).

Comme MNTB, LNTB semble être un relais source d'inhibition (Spirou et Berrebi, 1996). Alors que MNTB est desservi par les cellules globulaires (GBC) du VCN contralatéral, LNTB est desservi par ceux du noyau ipsilatéral. LNTB dessert toutes les divisions du CN ipsilatéral (Spirou, Walker et Berrebi, 1995). LNTB et MNTB desservent aussi le MSO homolatéral, qui reçoit donc, en plus de l'innervation bilatérale excitatrice postulée par le modèle de Jeffress, une innervation inhibitrice elle aussi bilatérale (Smith, 1995 ; Grothe, Park et Schuller, 1997 ; Grothe et Park, 1998).

Pour ces derniers auteurs, une innervation mixte inhibitrice-excitatrice (IE) provenant de la même oreille aurait pour "fonction" de conférer à la cellule une caractéristique "passe-bas" en réponse aux modulations d'amplitude. Ce comportement serait le résultat de la coïncidence de deux apports synchronisés avec la modulation : l'une excitatrice, et l'autre inhibitrice retardée. En réponse à un stimulus modulé en fréquence, la coïncidence des apports inhibiteurs et excitateurs annulerait l'activité du neurone lorsque la période est égale au retard, seule survivant la première impulsion. En principe, un tel mécanisme devrait produire une succession de zéros aux multiples de l'inverse du retard, plutôt qu'une simple caractéristique passe-bas. La suppression de réponses au-delà du premier zéro serait le résultat d'une persistance des effets inhibiteurs (Grothe, 1994, Covey et al., 1991). Le même principe a été invoqué pour expliquer les propriétés semblables observées dans DNLL (Yang et Pollak, 1997), ou (avec moins de succès) IC (Burger et Pollak, 1998).

Les noyaux du lemniscus latéral, DNLL (binaural) et VNLL et INLL (monauraux, collectivement connus sous la désignation de VCLL, ventral complex of the lateral lemniscus) sont à prédominance inhibitrice. Comme MNTB, ils constituent une source d'inhibition à structure temporelle fine qui projette sur IC (Yang et Pollak, 1997).

Autre source d'inhibition à la structure temporelle fine, les cellules onset-chopper (O_c) du noyau cochléaire, qui desservent les DCN ipsi- et contralatéraux (via la strie acoustique dorsale) (Joris et Smith, 1998). Les neurotransmetteurs les plus souvent cités pour cette inhibition rapide sont la glycine (cellules O_c du CN, principales de MNTB, LNTB) et dans une moindre mesure l'acide gamma-aminobutyrique (GABA).

Sites d'inhibition temporelle fine LSO est connu pour être le site d'une interaction binaurale inhibitrice, sensible à la structure temporelle fine des apports excitateurs et inhibiteurs. Parmi les sites possibles pour une interaction *monaurale*, on peut citer les suivants.

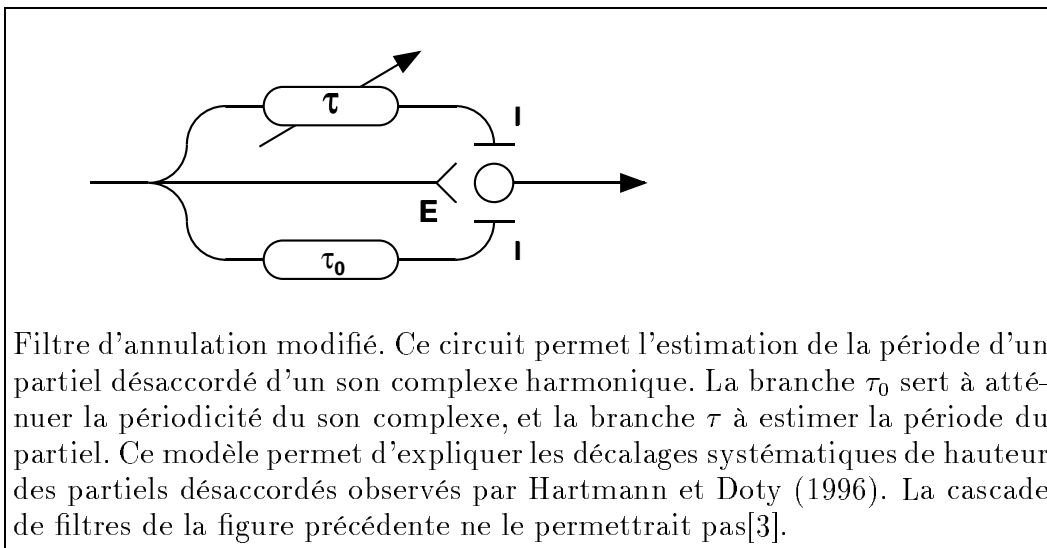
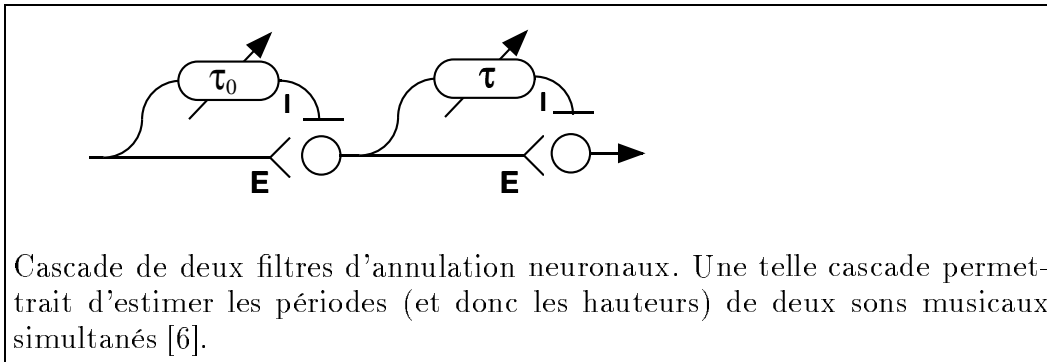
1. VNLL est systématiquement desservi par les fibres ascendantes des SBC et GBC contralatéraux, ainsi que par celles du MNTB homolatéral (Smith, Joris et Yin, 1993, 1998). Il reçoit aussi des fibres myélinisées en provenance des cellules-pieuve du CN contralatéral, via des synapses de type calice de Held (Schwartz, 1992). L'essentiel de cette information provient de l'oreille contralatérale, ce qui suppose que VNLL a un rôle monaural (voir cependant Batra et Fitzpatrick, 1997).
2. Le noyau périolivaire dorsomédian (DMPO) chez le chat, ou le noyau paraolivaire supérieur (SPN) chez le rat reçoivent aussi une innervation excitatrice et inhibitrice essentiellement contralatérale (Smith, Joris et Yin, 1993 ; Banks et Smith, 1992 ; Kuwada et Batra, 1999, Dehmel et al., 1999).
3. Dans DCN on note une interaction entre afférences excitatrices du nerf auditif, et inhibitrices via des cellules de type onset-chopper (O_c).
4. Si la vocation binaurale de MSO et LSO n'est pas exclusive, on peut les inclure parmi les sites possibles d'interaction monaurale.
5. IC reçoit de nombreuses projections à structure temporelle fine, excitatrices depuis AVCN et inhibitrices depuis VNLL (Schofield et Cant, 1997). L'activité dans IC n'a pas une synchronisation précise, mais il se peut qu'elle dépende de l'interaction temporelle des afférences qui projettent sur les dendrites de ses cellules.

En résumé, l'interaction inhibitrice rapide requise par des modèles d'annulation binaurale ou monaurale est présente à plusieurs niveaux du système auditif. Le MNTB se distingue comme apte à jouer un rôle de "source d'inhibition". Ses nombreuses projections laissent un grand choix pour rechercher les sites d'interaction requis par les modèles. Pour l'instant le seul exemple qui ait été mis en évidence de façon claire l'interaction binaurale dans LSO.

3.2.4 Vers des modèles plus complexes ?

Une chose frappante est le décalage entre la complexité du système nerveux auditif, et la *simplicité* relative des modèles. Les modèles de Jeffress ou de Licklider supposent, certes, des réseaux de nombreux neurones opérant en parallèle avec des paramètres différents. Cependant, chaque élément de ces circuits est d'une grande simplicité : un neurone muni de deux synapses. L'anatomie est bien plus complexe.

Le "filtre d'annulation" neuronal que je propose [14, 12] est un pas dans la direction de topologies de traitement plus complexes, puisqu'il est censé préparer à d'autres traitements. On peut imaginer des filtres multiples mis en cascade, comme je l'ai proposé pour l'estimation des hauteurs de sons



multiples[6]. J'ai pu montrer, en modélisant les données psychoacoustiques de Hartmann et Doty (1996), la présence plausible d'une version un peu plus complexe du "filtre d'annulation", comprenant deux voies inhibitrices au lieu d'une.

On peut imaginer des combinaisons plus complexes faisant intervenir de multiples lignes à retard et synapses excitatrices ou inhibitrices. Ces circuits pourraient intervenir dans les interactions binaurales, l'élimination de bruit, l'estimation de la périodicité (hauteur) ou du timbre.

Des modèles temporels relativement complexes ont été proposés pour la discrimination chez la chauve-souris des durées d'écho ou de modulation de fréquences (Casseday et al., 1995), ou la caractérisation par effet doppler des vitesses de cible (Suga, 1994).

Maas (1997a,b) a suggéré que par combinaison additive de potentiels post-synaptiques excitateurs et inhibiteurs (EPSP et IPSP) on peut synthétiser des filtres de réponse impulsionnelle arbitraire. Plus généralement, il a

montré que des réseaux de neurones formels traitant des informations codées par *instant* de décharge sont aussi puissants que les réseaux classiques “sigmoïdes”, qui supposent que l’information est codée par les taux moyens de décharge dans les différents neurones. Dans certains cas ils sont même plus puissants, dans le sens où certaines opérations peuvent se faire avec un nombre de neurones bien moindre qu’avec des réseaux classiques. Le codage temporel a été suggéré pour des systèmes autres que auditif, par exemple le système visuel (Thorpe, 1996 ; Usrey et Reid, 1999).

3.3 Conclusion

On trouve des indices d’organisation tonotopique à tous les niveaux du système auditif, de la cochlée au cortex. Il est habituel d’y voir autant de représentations spectrales, servant de substrat à des traitement auditifs dans le domaine fréquentiel. C’est ce que supposent la plupart des modèles classiques de traitement auditif. Parallèlement, on trouve des indices de structure temporelle fine, et de spécialisations cellulaires et synaptiques pour son transport et traitement. Leur présence s’explique en partie par les besoins de comparaison des temps d’arrivée binaurale, selon le modèle classique de Jeffress (1948), mais leur répartition est bien plus large que ne le requiert ce modèle. On les retrouve en particulier dans VNLL et divers noyaux périolivaires qui ne sont pas habituellement impliqués dans le traitement binaural. On trouve aussi une présence massive de circuits *inhibiteurs* spécialisés dans le traitement de la structure temporelle fine. Le manque de modèles nécessitant ces éléments est un obstacle à la compréhension des données anatomiques et physiologiques, et aussi à leur récolte. Faute de modèle qui leur donne un sens, la description des propriétés temporelles est parfois omise, parfois incomplète.

4 Enjeux théoriques en Audition

4.1 Dimensions physiques, qualités sensorielles

L'ensemble des sons forme un espace (au sens mathématique) vaste et complexe. Il en est de même de l'ensemble des sensations, perceptions, voire actions que ces sons peuvent évoquer ou déclencher. Décrire ces ensembles dans toute leur complexité est impossible. La théorie de l'Audition s'est donc efforcée de les structurer, de tailler chacune selon des dimensions, et d'établir entre dimensions des relations simples et générales. Il faut souligner le caractère *réducteur* de cette approche. Sont écartés les interactions entre l'audition et les autres sens, les particularités individuelles, l'état du sujet, les effets du contexte, des contraintes pragmatiques, etc.. Sont retenues les caractéristiques communes à tout sujet, reproductibles, pouvant être projetées sur une dimension simple. Sont écartés l'émotion unique produite par telle note de musique à la première écoute, le pouvoir évocateur d'un air (l'équivalent sonore de la "madeleine" de Proust), la richesse du monde sonore qui transparaît dans le discours du musicien ou du mélomane. Cette aspect réducteur de la théorie de l'Audition explique qu'elle déçoive souvent ceux qui aiment le son : il manque ce qui relève de l'unique, pour eux essentiel.

La théorie de l'Audition s'est ainsi intéressée à des "qualités sensorielles" ou attributs perceptifs : sonie, hauteur tonale, timbre, localisation, durée subjective, etc.. La *hauteur tonale* d'un son, par exemple, est la qualité qui permet de classer certains sons - ceux qui sont périodiques - sur une échelle qui va du grave à l'aigu (ANSI, 1960 ; AFNOR, 1977). La hauteur est une abstraction : chaque hauteur est une *classe d'équivalence*, au sens mathématique, contenant un ensemble de sons qui peuvent différer par leurs caractéristiques physiques (autres que la période) et les sensations qu'ils évoquent (autres que la hauteur). Il est remarquable que les hauteurs s'ordonnent, en première approximation, le long d'une unique "dimension" sensorielle. Remarquable aussi que cette dimension sensorielle corresponde, en première approximation, à une dimension physique (la période ou son inverse, la Fréquence Fondamentale ou F_0). Remarquable, enfin, que la hauteur d'un son soit peu affectée par d'autres dimensions physiques : durée, intensité, enveloppe spectrale (Demany, 1987, 1988).

Il ne faut donc pas trop s'étonner de ce qu'une analyse plus fine révèle une structure plus complexe. A la dimension linéaire de la hauteur tonale on peut superposer une structure cyclique de *chroma* (correspondant aux notes de l'échelle musicale), qu'on peut voir comme des classes d'équivalence groupant les hauteurs de stimuli différant d'une octave (Bachem, 1950 ; Shepard, 1964 ; Demany et Armand, 1984 ; Ueda et Ohgushi, 1984 ; Demany, 1986). La

hauteur paraît alors comme une quantité bidimensionnelle. Une dimension, circulaire, différencie les notes différentes à l'intérieur de la gamme chromatique. L'autre, linéaire différencie les notes de même nom mais séparées d'une ou plusieurs octaves. L'ensemble prend la forme d'une spirale. Shepard (1982) propose même des représentations de dimensionalité plus élevée. L'intérêt de ces représentations est d'expliquer les relations privilégiées entre certaines notes (octaves, tierces, etc.). Selon Sethares (1998), cependant, ces relations privilégiées dépendent étroitement de la structure spectrale des notes, harmonique pour la plupart des instruments occidentaux, mais inharmonique pour des instruments d'autres cultures. Il n'est donc pas certain que ces dimensions aient toutes une valeur universelle. Même au sein de notre culture, les individus diffèrent par le poids qu'ils accordent aux différentes dimensions, et par leur capacité à discriminer, classifier ou identifier les sons (ou les intervalles entre sons) selon ces dimensions.

Autre facteur de complexité : la relation entre hauteur et période n'est pas parfaitement biunivoque : des stimuli de même période (voire identiques) peuvent évoquer des hauteurs différentes. La différence est le plus souvent d'une octave, mais dans certains cas (ton complexe réduit à deux ou trois partiels) d'autres ambiguïtés sont possibles. La dimension physique de fréquence fondamentale (F_0) n'a de sens rigoureux que dans le cas des sons périodiques, et à condition d'aménager la définition pour accommoder le fait que tout son est limité dans le temps. La hauteur dépend étroitement de la périodicité, mais il faut pourtant rappeler qu'un vaste ensemble de stimuli non-périodiques (sons inharmoniques, bruits filtrés, etc.) évoquent aussi une hauteur tonale, plus ou moins claire.

D'autres qualités sonores admettent une description "dimensionnelle", qu'on peut mettre en correspondance avec une ou plusieurs dimensions physiques. Comme la hauteur, la *sonie* prend la forme simple d'une échelle unidimensionnelle. La dimension physique correspondante est l'*intensité sonore* (Buser et Imbert, 1987 ; Hartmann, 1997), mais la sonie est affectée aussi par la durée et le contenu spectral du stimulus. Le *timbre* résiste plus à une caractérisation simple, du fait de l'ambiguïté de ce terme dans le langage courant et de sa définition formelle même, qui, en caractérisant le timbre comme ce qui différencie deux sons de même hauteur et sonie, lui abandonne tout ce qui s'écarte des deux premières dimensions. Ceci dit, des expériences d'analyse multidimensionnelle ont permis de dégager, selon les études, deux ou trois dimensions qui semblent sous-tendre cette qualité (McAdams et Cunnibile, 1992), auxquelles on peut faire correspondre approximativement des dimensions physiques (Krimphoff et al. 1994 ; McAdams et al., 1995).

4.2 Le rôle privilégié de la Hauteur Tonale

La relative simplicité de la dimension sensorielle de hauteur tonale, le lien étroit avec la dimension physique (F_0), l'importance de son rôle dans la musique, etc., ont fait que la théorie de l'Audition a longtemps été la théorie de la hauteur tonale. Malgré la simplicité relative de la relation entre les dimensions sensorielle et physique, la nature exacte de la hauteur, et les mécanismes physiologiques qui la sous-tendent, font l'objet depuis un siècle d'une controverse parfois violente (Schouten, 1970; de Boer, 1976; Evans, 1978; Moore, 1989; Houtsma, 1995) dont les braises sont encore chaudes (pour un exemple récent voir le commentaire de Carlyon, 1998 en réponse à Meddis et O'Mard, 1997).

L'origine de la controverse est une application sans doute trop zélée du Théorème de Fourier par le physicien Ohm (1843, cité par de Boer, 1976 et Houtsma, 1995). Le théorème dit qu'une fonction périodique peut s'analyser comme une somme de fonctions sinusoïdales, dont les fréquences sont toutes multiples d'une même fréquence, la fréquence fondamentale F_0 , inverse de la période (série de Fourier). Le théorème ne dit *pas* que tous les termes de la série doivent être présents. En particulier, le coefficient du terme de fréquence F_0 peut très bien être nul, sans que cela compromette la périodicité de la fonction analysée. Néanmoins, Ohm choisit de faire de la présence physique de cette composante la condition de la perception de la hauteur correspondant à la période. Seebeck (1841, 1843, cité par de Boer, 1976 et Houtsma, 1995) avait pourtant montré expérimentalement que cette présence n'était pas nécessaire à cette perception. Helmholtz (1877, 1954) intervint aux côtés de Ohm, et son autorité suffit à mettre fin au débat, en laissant la controverse dans un état peu satisfaisant. Les observations et arguments de Seebeck étaient pourtant irréfutables, et furent confirmées dans le siècle suivant par les expériences de Schouten (1940, cité par Schouten, 1970 et de Boer, 1976), ou de Licklider (1956).

La controverse, *dans sa forme d'origine* est aujourd'hui éteinte, dans un sens qui donne raison à Seebeck, et tort à Ohm et Helmholtz. Il est largement admis que la présence d'une composante sinusoïdale à la fréquence F_0 n'est *pas* nécessaire à la perception de la hauteur correspondante, même si ce "paradoxe de la fondamentale absente" continue d'être redécouvert de temps à autre. Ce qui compte est la périodicité du signal, dont la décomposition de Fourier peut, ou peut ne pas, comporter de coefficient fondamental non-nul. Dans des formes dérivées, cependant, la controverse reste bien vivante. Schouten (1970) et de Boer (1976) font d'excellentes (et divertissantes!) revues de ses avatars successifs, que Evans (1978) a complété de son point de vue de physiologiste. Des exposés plus récents sont ceux de Demany (1986),

Moore (1997) ou Houtsma (1995). La lecture de ces histoires est indispensable à quiconque veut trouver un sens à la théorie de l'Audition. Elle éclaire le choix parfois étrange des expériences, la schizophrénie qu'on devine parfois chez les acteurs, leur extrême prudence (nul n'ayant envie de subir le sort d'un Seebeck), les échanges parfois violents, et surtout la multiplication d'expériences qui s'attaquent inlassablement aux mêmes hypothèses, les réfutant coup après coup sans que rien ne semble diminué de leur vigueur.

Subsistent aujourd'hui de ce débat deux questions intéressantes, l'une générale et l'autre spécifique à la hauteur tonale. La première concerne les parts respectives des représentations fréquentielles (tonotopiques) et temporelles dans les traitements effectués par le système auditif. La deuxième concerne l'hypothèse selon laquelle la hauteur tonale d'un ton résulterait d'un traitement algorithmique des fréquences de ses partiels ("pattern matching"), plutôt que d'une estimation directe de sa périodicité fondamentale.

La première question prend racine dans la controverse Ohm-Seebeck, et la position prise par Helmholtz dans ce débat. Il s'agit de dire si l'information est transmise par au cerveau sous une forme *temporelle* représentant les mouvements de la cochlée, ou une forme *spectrale* résultant d'une analyse fréquentielle, chaque fréquence étant représentée par sa "place" sur l'épithélium sensoriel. La première hypothèse correspond à la théorie "téléphone" de Lord Rutherford (1886, cité par Pickles, 1988). La seconde obéit au principe des énergies spécifiques de Johannes Müller (Brugge, 1992). Helmholtz (1877) proposa que la cochlée contient un banc de résonateurs, accordés chacun à une fréquence de la gamme à laquelle l'oreille est sensible. L'ensemble opère une analyse spectrale : à chaque composante du signal correspond un maximum de vibration à un lieu spécifique. En particulier si on suppose que le son est périodique *et contient une composante sinusoïdale de fréquence F_0* , on pourrait déduire la hauteur de la position du maximum correspondant. On aurait ainsi un mécanisme physiologique simple et élégant pour expliquer la *perception* de la hauteur. C'est peut-être ce qui poussa Helmholtz à soutenir la thèse de Ohm, qui garantissait la présence, pour tout son doué d'une hauteur, d'un partiel sur lequel son modèle puisse opérer.

L'hypothèse de Helmholtz fut à la fois confirmée et infirmée par les expériences de von Békésy (1960). Confirmée en ce que la cochlée fonctionne bien comme un banc de résonateurs, un ton pur produisant un pic de vibration à une place qui dépend de sa fréquence. Infirmée en ce que ce pic était, dans les mesures de von Békésy, extrêmement large, impossible à assimiler à la raie d'une transformée de Fourier. Ce problème incita von Békésy (1960) à proposer des mécanismes de renforcement de contraste (qu'il étudia expérimentalement sur la peau), Licklider (1959) à introduire une deuxième dimension d'analyse plus fine, et les physiologistes à rechercher des mécanismes de

suppression latérale ou de "second filtre" (Evans, 1978). Dans chaque cas, il s'agissait de rapprocher de la sélectivité parfaite d'une transformation de Fourier la sélectivité médiocre mise en évidence expérimentalement. Des expériences plus récentes ont mis en évidence une sélectivité bien plus fine (Evans, 1974; Ruggero, 1992), mais qui reste loin de l'idéal d'une transformée de Fourier : (a) la sélectivité la plus fine s'observe pour des intensités proches du seuil, et se dégrade à intensité plus élevée, (b) elle s'observe essentiellement pour des tons purs isolés (courbes d'accord) : pour des sons complexes, les patterns de réponse sont très différents d'un spectre.

L'hypothèse rivale de Rutherford fut elle aussi confirmée expérimentalement par Wever (1930b), qui mesura dans le nerf auditif un signal électrique avec une structure temporelle semblable au signal acoustique. Depuis, les mesures électrophysiologiques de plus en plus fines ont confirmé la structure temporelle du pattern porté par les fibres du nerf auditif (Kiang, 1965; Rose et al., 1967; Evans, 1986; Cariani et Delgutte, 1996a,b). L'expérimentation physiologique n'a donc pas permis de trancher de façon nette entre codage tonotopique ou temporel. C'est une situation commune dans ce domaine, que deux théories rivales soient toutes deux supportées par des données expérimentales.

Si, historiquement, la question codage-temporel/codage-spectral est associée à la controverse entre Ohm et Seebeck au sujet de la hauteur, elles sont néanmoins logiquement distinctes. Le codage spectral est en accord avec l'importance donnée par Ohm à la composante fondamentale, mais rien n'empêche que cette composante soit codée de façon temporelle. Inversement, le codage temporel s'accorde bien de l'*absence* d'une composante fondamentale, mais il en est de même d'un codage spectral si l'on fait appel aux mécanismes de "reconnaissance des formes" proposés par les théories récentes.

La deuxième question intéressante concerne justement l'hypothèse de "reconnaissance des formes". La forme à reconnaître (et à associer à la hauteur) est la série des fréquences des partiels présents dans le son. Les partiels d'un son périodique forment une série régulière de partiels (série harmonique) de fréquences multiples de la fondamentale². La série détermine la fondamentale de façon très redondante, et cette dernière peut donc en être dérivée même si la série est incomplète. Il suffit (a) que l'ensemble des partiels présents admette cette période comme PGCD (plus grand commun diviseur), et (b) qu'on ait l'assurance qu'il en est ainsi. En particulier, on peut dériver la *fréquence* fondamentale même si la *composante* fondamentale est absente. C'est le principe des modèles "reconnaissance des formes" (pattern matching) de

²ce sont les périodes, et non les fréquences, qui forment une série harmonique au sens mathématique

perception de la hauteur (Goldstein, 1973 ; Wightman, 1973 ; Terhardt, 1974 ; de Boer, 1977), qui dérivent la hauteur d'un son complexe de l'ensemble des fréquences de ses partiels.

L'idée est séduisante, mais on peut se demander pourquoi se donner tant de mal pour abstraire du spectre une caractéristique immédiatement évidente dans l'onde : sa période. Périodicité et structure harmonique du spectre sont mathématiquement équivalentes. Pourquoi postuler une analyse de Fourier suivie d'un calcul algorithmique, alors que le résultat est évident dans la forme d'onde ? La raison de la popularité de ces modèles est à rechercher dans un ensemble de facteurs.

- On a reproché aux modèles temporels rivaux des insuffisances (réelles ou imaginées) : sensibilité à l'enveloppe (la hauteur est dans certains cas dissociée de la périodicité de l'enveloppe), sensibilité à la phase (la hauteur est généralement indépendante de la phase), mais aussi paradoxalement une *insensibilité* excessive à la phase (la hauteur est parfois sensible à la phase). En fait ces caractéristiques dépendent simplement de l'implémentation.
- La valeur et la saillance d'une hauteur sont dominées par des composantes situées dans une région bien déterminée, dans laquelle les partiels sont relativement bien "résolus". Cette dominance s'explique si l'on admet l'hypothèse de reconnaissance de formes. Inversement, le poids des partiels d'ordre élevé est faible, alors qu'ils produisent dans les canaux périphériques une activité clairement dominée par la période fondamentale, à laquelle un modèle temporel devrait être sensible.
- La reconnaissance des formes permet de "sauver" l'hypothèse de Helmholtz en la dissociant de celle de Ohm, démentie par l'expérience. Plutôt qu'un pic de vibration à la fondamentale, on postule un ensemble de pics à chacun des partiels
- À ces facteurs s'en ajoutent d'autres plus difficiles à cerner, qui tiennent de l'"écologie" des idées. de Boer (1976) décrit le mouvement des idées entre "période" et "place" comme celui d'un pendule. L'analyse de Fourier fait partie du bagage de nombreux scientifiques, et la tendance est forte de voir dans le passage par l'espace dual la certitude de faire un pas vers l'explication des phénomènes.

Récemment, les modèles classiques de "reconnaissance des formes" cèdent la place au modèle "autocorrélation". La popularité récente du modèle "autocorrélation" peut s'interpréter comme un nouveau mouvement du balancier de de Boer (1976), cette fois dans le sens "période". La fonction d'autocorrélation peut se concevoir comme une version du signal dans laquelle toutes les composantes seraient alignées sur une même phase cosinus. C'est une fonction du temps, comme le signal, mais invariante par rapport aux manipulations de

phase, comme le spectre d'amplitude. Elle présente un maximum à l'origine, produite par la sommation des composantes en phase cosinus, et, si le signal est périodique, un pic d'égale amplitude à la période. Ce pic sert d'indice de périodicité. La fonction d'autocorrélation peut être implémentée physiologiquement d'une manière assez simple et plausible, selon un mécanisme de coïncidence neuronale proposé par Licklider (1956). Par contraste, nul mécanisme physiologique plausible n'a été proposé pour implémenter les modèles classiques "reconnaissance des formes" (à noter cependant une proposition récente de Shamma et Klein, 2000, et la théorie d'apprentissage de Terhardt, 1974).

Le modèle de Licklider a été repris sous des formes diverses (Moore, 1977 ; van Noorden, 1982 ; Meddis et Hewitt, 1991a,b, [8]). Récemment, il a été conforté par une série de mesures physiologiques utilisant des stimuli identiques à ceux qui ont servi en psychoacoustique pour mettre évidence les effets importants liés à la hauteur (Cariani et Delgutte, 1996a,b). C'est un excellent modèle de premier ordre : simple, plausible, et qui rend compte de la grande majorité des phénomènes. Ces qualités étaient pour l'essentiel évidentes il y a cinquante ans, lors de l'introduction de ce modèle, et on peut s'interroger sur les facteurs qui ont maintenu ce modèle en marge de la théorie de l'audition pendant si longtemps.

Malgré ses qualités, le modèle d'autocorrélation souffre de deux handicaps. Le premier est qu'on n'a pas encore réussi à lui trouver un substrat physiologique, contrairement au modèle de localisation de Jeffress (1948) (Section 3). Le deuxième handicap est que le modèle d'autocorrélation ne rend pas bien compte des phénomènes de dominance d'une certaine partie du spectre, celle occupée par les partiels "résolus" par le filtrage périphérique. On entend par "résolus" les partiels dont la séparation est suffisante pour que, quel que soit le partiel, il existe certains canaux périphériques qui ne répondent qu'à ce seul partiel, en excluant les autres. Par "non-résolus" on entend des partiels qui, quel que soit le canal, sont mêlés à d'autres partiels avec lesquels ils battent. Il semble que les partiels résolus et non-résolus jouent des rôles différents dans la hauteur tonale, les premiers ayant une importance plus grande (Carlyon et Shackleton, 1994 ; Houtsma, 1995). L'autocorrélation est capable d'exploiter aussi bien les indices provenant des partiels résolus que ceux provenant des non-résolus, et il n'est pas évident à priori de lui faire rendre compte de la différence de nature ou de saillance des hauteurs que ces deux sortes de partiels évoquent. Des simulations (Meddis et Hewitt, 1991a,b ; Meddis et O'Mard, 1997) et enregistrements physiologiques (Cariani et Delgutte, 1996a,b) ont permis de répondre à cette objection en partie, mais en partie seulement (Carlyon, 1998).

La perception de "dimensions" perceptives telles que la hauteur tonale

occupe toujours une place centrale en Théorie de l'Audition. Récemment de nouvelles préoccupations ont surgies, concernant le rôle écologique de la perception, sa relation avec l'action, les interactions entre les sens (perception multimodale), l'Analyse de Scènes, la reconnaissance des formes (au sens général), la Cognition. Ces développements ont été influencés par les applications (robotique, reconnaissance de la parole, etc.) qui ont servi d'inspiration mais aussi créé une demande pour des résultats et modèles. En même temps que cet élargissement des problématiques, on assiste à un approfondissement à la faveur des progrès techniques. Les techniques d'exploration anatomique et physiologique : électrophysiologie, imagerie, biochimie, etc. L'informatique qui permet le traitement de ces données complexes et leur modélisation.

L'accumulation des données et la complexification des problématiques renforce le besoin de modèles. Le présent travail se place dans cette perspective. Sa contribution est de proposer des modèles de traitement temporel pour la perception de la hauteur, séparation perceptive des sources multiples ("cocktail party effect"), perception du timbre. La question des rôles respectifs de la tonotopie et de l'information temporelle est cruciale : il est évident qu'une erreur de choix à ce niveau compromet la validité de l'édifice de modélisation. Le manque de modèles de traitement temporel pose problème face à l'accumulation de données favorables à ce type de traitement.

Pour résumer, les préoccupations modernes de l'Audition vont au delà des dimensions perceptives classiques de la hauteur, sonie, etc.. La hauteur tonale continue pourtant à occuper une position centrale, sur laquelle viennent se cristalliser des questions théoriques plus larges, en particulier le débat entre représentations fréquentielles et temporelles. L'accumulation des données expérimentales n'a pas permis de trancher ce débat vieux de plus d'un siècle. Il s'agit pourtant d'un choix crucial, qui engage la validité de l'édifice théorique. L'obstacle est moins l'insuffisance de données que le manque de modèles. Le présent travail propose des modèles de perception de la hauteur tonale et du timbre des voyelles, et de ségrégation perceptive des source sonores. Il s'agit de modèles ponctuels, mais qui font appel à des mécanismes et principes (coïncidence, inhibition, pondération sélective) qui peuvent servir de base à une gamme de modèles plus large.

5 Séparation à base de F_0

Cette section traite de la ségrégation des sources sonores sur la base de leur structure harmonique.

5.1 La conjecture d'annulation harmonique

J'ai abordé pour la première fois la problématique de l'*organisation perceptive de l'environnement sonore* en 1991, à la faveur d'un projet de recherche en collaboration avec Stephen McAdams. Il s'agissait d'expliquer comment l'information acoustique provenant de sources multiples est partitionnée entre sources (ségrégation perceptive), et comment les composantes appartenant à une source sont regroupées (fusion perceptive). Parmi les facteurs qui favorisent la ségrégation, le mieux étudié est la *structure harmonique*, exploitée par exemple pour séparer et rendre intelligibles deux voix concurrentes ("cocktail party effect"). L'analyse de la littérature a révélé que les nombreux modèles qui exploitent ce facteur pouvaient se classer en deux catégories selon la stratégie qu'ils adoptent : *renforcement* ou *annulation* harmonique. Le renforcement harmonique exploite la structure harmonique d'une cible (par exemple une voix) pour la renforcer par rapport aux bruits qui l'accompagnent. L'annulation harmonique exploite au contraire la structure harmonique des sons parasites pour les éliminer.

Cette classification a permis de mieux comprendre le fonctionnement des modèles et de déterminer dans quelles conditions chaque type de modèle est efficace. Renforcement et annulation harmoniques ayant des propriétés différentes, on peut concevoir des expériences pour dire à quelle classe appartiennent les mécanismes employés par le système auditif. Cette stratégie est décrite en détail dans [14]. Renforcement et annulation ont chacun ses qualités et défauts théoriques ou pratiques, qui sont explorés dans plusieurs articles [14, 38, 13, 34]. En particulier :

- Dans [14] je présente un théorème (dû à Jean Laroche) qui fixe une limite au degré d'amélioration du rapport signal-sur-bruit qu'on peut attendre du renforcement harmonique (l'annulation permettant une amélioration théoriquement infinie).
- Dans [61, 38, 39, 34] je rapporte les résultats d'une expérience avec un système de reconnaissance de la parole, dans laquelle les deux stratégies furent employées pour réduire les effets d'une voix parasite. La stratégie d'annulation s'est révélée la plus efficace.
- Dans [13, 11] sont rapportées des expériences de psychoacoustique dont l'objectif était de déterminer la ou les stratégies utilisées par le système auditif. Ces expériences utilisaient comme stimuli des paires de voyelles

synthétiques concurrentes, harmoniques et inharmoniques, et mesurait les taux d'identification de chaque voyelle de la paire. Il en est ressorti que le système auditif utilise la stratégie d'annulation, mais non celle de renforcement. Une conclusion semblable découle des expériences de Lea (1992) et Summerfield et Culling (1992), ainsi que d'autres expériences plus récentes [10, 4, 2].

Ainsi, l'hypothèse d'annulation est confortée par un faisceau convergent d'arguments, aussi bien théoriques qu'expérimentaux. Sa rivale, l'hypothèse de renforcement des cibles harmoniques, trouve peu en sa faveur. C'est pourtant sur cette stratégie-là que s'appuient nombre d'algorithmes de traitement de signal proposés pour la séparation des voix, ainsi que des théories perceptives du type "analyse de scènes auditives". Selon ces théories, la structure harmonique d'une cible jouerait un rôle de "motif" ou "texture" permettant de délimiter les plages spectro-temporelles appartenant à une cible, les distinguer du fond, et grouper ensemble des plages disjointes appartenant à une même source. Cette attente n'est *pas* confirmées par l'expérience.

L'hypothèse de renforcement harmonique restera néanmoins sans doute populaire, car l'idée que la structure harmonique d'une cible sert à sa ségrégation est attrayante et difficile à abandonner. Il se peut d'ailleurs que des expériences futures révèlent sa pertinence, par exemple dans des tâches nécessitant une organisation de plus haut niveau. Pour l'instant ce n'est pas le cas.

5.2 Expériences "voyelles doubles"

Cette série d'expériences psychoacoustiques sonde de façon générale les mécanismes de ségrégation employés par le système auditif. Au départ elles s'adressaient à la conjecture précédente, en posant deux questions : (1) le système auditif exploite-t-il la structure harmonique d'une cible pour la séparer d'un fond, (2) exploite-t-il la structure harmonique d'un fond pour l'éliminer ? Pour obtenir des réponses fines, des modifications ont été apportées aux méthodes expérimentales classiques. Les nouvelles méthodes ont permis à leur tour d'explorer d'autres questions, et d'apporter un nouvel éclairage à l'ensemble des phénomènes de ségrégation.

Le paradigme classique d'identification de voyelles concurrentes est dû à Scheffers (1983). Il a été employé depuis par de nombreux auteurs (Zwicker, 1984 ; Assmann et Summerfield, 1990, 1994 ; Summerfield et Assmann, 1991 ; McKeown, 1992 ; Culling et Darwin, 1993, 1993). Des stimuli constitués de la somme de deux voyelles sont présentés à des sujets, leur tâche étant de les identifier toutes les deux. La ségrégation perceptive des voyelles rend leur identification plus facile, ce paradigme permet donc de mesurer l'effet des

facteurs qui favorisent cette ségrégation, tels que la *différence de fréquence fondamentale* (ΔF_0) entre voyelles. Classiquement, la mesure de ségrégation est le taux d'identification "deux voyelles correctes", c'est à dire la proportion de stimuli pour lesquels les deux voyelles sont correctement identifiées. On a pu montrer ainsi que la structure harmonique des voyelles facilite leur ségrégation, à condition que les F_0 s soient différentes.

Malheureusement, ce paradigme ne permet pas de dire si le mécanisme responsable est du type renforcement ou annulation. Pour répondre à cette question la méthode doit être modifiée pour utiliser (a) des stimuli qui associent voyelles harmoniques et inharmoniques, et (b) une mesure qui soit sensible à l'identification de chaque voyelle séparément (et non les deux à la fois comme dans la méthode classique). Cette modification fut faite par Lea (1992), qui employa des mélanges de voyelles voisées et chuchotées et mesura séparément les taux d'identification pour les deux voyelles. Cette nouvelle mesure "cible-correcte" fut employée dans l'ensemble de mes expériences [4, 10, 11, 13, 2, 60, 55, 40, 37, 36, 35].

La première expérience de cette série[13] fut menée en collaboration avec Steve McAdams, Jean Laroche et Muriel Rosenberg. Elle employa comme stimuli des paires de voyelles qui pouvaient être chacune harmonique (voisée normale) ou inharmonique (petite perturbation aléatoire de chaque partiel). L'objectif était de répliquer les expériences de Lea (1992) avec des stimuli mieux définis que les voyelles chuchotées (dont le spectre continu est difficile à apparier avec le spectre de raies d'une voyelle voisée). Nous avons confirmé les résultats de Lea : le système auditif emploie bien une stratégie d'annulation harmonique. En revanche ni Lea, ni nous mêmes, n'avons trouvé de résultat en faveur de l'hypothèse de renforcement harmonique. Summerfield et Culling (1992) avaient abouti à la même conclusion avec une procédure différente (adaptative).

La mesure indépendante de l'identification de chaque voyelle de chaque paire est un premier progrès par rapport à la procédure classique. Une deuxième modification a consisté à introduire systématiquement une *différence d'amplitude* entre les voyelles. Lorsque deux voyelles sont de même niveau, il peut arriver qu'elles soient toutes deux identifiées avec un taux près de 100%, même en l'absence d'indices de ségrégation. L'expérience est alors peu sensible (effet plafond). Une différence d'amplitude rend la tâche plus difficile pour la voyelle la plus faible, et du coup l'expérience devient sensible à tout ce qui facilite la ségrégation. Une différence d'amplitude rapproche l'expérience des conditions naturelles, puisque dans la vie courante il est rare que deux voix concurrentes soient de même niveau à chaque instant. Une première expérience [10] a utilisé les amplitudes relatives (-20, -10, 0, 10, 20 dB), croisées avec deux conditions de ΔF_0 (0 et 6%). Les effets de ΔF_0 se sont effecti-

vement révélés bien plus grands pour des cibles faibles (-10 et -20 dB) que lorsque les deux voyelles avaient la même amplitude ou lorsque la cible était la plus forte. Une deuxième expérience [10] a utilisé une gamme plus large de conditions : (-35, -25, -15, -5, 5, 15, 25, 35 dB) croisés avec ($\Delta F_0 = 0, 3, 6, 12\%$). Ces expériences ont permis de conclure qu'une différence d'amplitude de l'ordre de 15 dB permet d'obtenir la meilleure sensibilité. Il s'agit d'une deuxième amélioration importante par rapport à la procédure classique.

Au passage, ces expériences ont fourni un argument supplémentaire en faveur de l'hypothèse d'annulation harmonique. L'annulation harmonique requiert une connaissance de la F_0 de la voyelle concurrente. Celle-ci est facile à estimer lorsque la cible est faible, et c'est précisément dans ce cas que les effets les plus grands sont observés. À l'inverse, le renforcement harmonique requiert une connaissance de la F_0 de la cible. Il est difficile d'imaginer comment celle-ci pourrait être estimée lorsque la cible est à -20 ou -25 dB par rapport à son concurrent. L'existence d'effets importants de ΔF_0 pour des cibles faibles permet donc d'écarter cette hypothèse.

Une troisième modification de la procédure classique a été de permettre aux sujets de répondre *une ou deux* voyelles, plutôt que de les obliger à faire deux réponses pour chaque stimulus. Cette modification se justifie de la façon suivante :

- La tâche est plus facile, ce qui fait plaisir aux sujets et permet de leur infliger en échange des jeux de stimuli plus nombreux.
- La tâche est caractéristique de situations naturelles où le nombre des sources n'est pas connu à l'avance.
- L'expérience devient sensible à des indices qui signalent précisément la multiplicité des sources en présence. En obligeant les sujets à répondre deux voyelles, la tâche classique limitait sévèrement l'expression de tels indices, pourtant utiles à la ségrégation.
- En pratique, les effets des indices de ségrégation sont effectivement plus grands, ce qui facilite leur mise en évidence avec un niveau de confiance statistique satisfaisant. Le *nombre moyen de réponses par stimulus* est une mesure intéressante.

Les inconvénients de la tâche modifiée sont, premièrement la difficulté de comparer avec des résultats d'études utilisant la procédure classique, et deuxièmement sa sensibilité à des "indices de multiplicité" en plus d'indices de ségrégation plus "classiques". Cette sensibilité multiple introduit une incertitude sur l'origine précise des effets observés. Ces inconvénients sont mineurs par rapport aux avantages.

Pour résumer, les trois modifications apportées à la procédure classique (taux "cible-correcte", différence d'amplitude de 15 dB, tâche "une réponse ou deux") aboutissent à un outil très sensible pour l'étude des indices de

ségrégation. J'ai appliqué cet outil à une gamme assez large d'indices :

Phase. Plusieurs théories laissent supposer un effet de la phase sur l'identification, aussi bien le spectre de phase de chaque voyelle (phase "intra-voyelle") que le spectre des phases relatives (phase "intervoyelles"). Ce sont la théorie PPA (pitch period asynchrony) de Summerfield et Assmann (1991), et la théorie des battements de Culling et Darwin (1993). La dernière, notamment, était jusqu'à présent l'explication "acceptée" des effets de petites différences de F_0 (moins de 3%). Au moyen de deux expériences [11, 2], j'ai fait la démonstration que les effets de phase sont soit inexistantes, soit de très petite taille. Ils sont en tout cas largement insuffisants pour expliquer les effets de ΔF_0 . Ces résultats infirment la théorie PPA et celle des battements, au moins en tant qu'explication des effets de ségrégation liés aux différences de F_0 .

Petites valeurs de ΔF_0 . La différence de F_0 étant un indice important pour la ségrégation, il est intéressant d'étudier la limite des ΔF_0 très faibles. Les études classiques montrent que l'essentiel de l'effet de la ΔF_0 s'observe entre 0% et la valeur non-nulle la plus faible utilisée dans chaque étude (6%, 3% ou 1.5%). Pourtant, la zone des ΔF_0 inférieures à 1.5% n'a jamais été explorée. Son exploration pose une difficulté méthodologique, du fait des fluctuations du spectre qui interviennent lorsque la période de battements (inverse de la ΔF_0) est plus longue que le stimulus. Tout en contrôlant de tels effets, j'ai pu mesurer des effets statistiquement fiables jusqu'à $\Delta F_0 = 0.4\%$ [2]. Traduite en Hz (0.5 Hz) cette valeur paraît extrêmement faible comparée à la largeur de bande des filtres cochléaires (de l'ordre de 60 à 300 Hz dans la plage des formants F1-F2). Elle est aussi très faible en termes de résolution temporelle nécessaire pour différencier les périodes des fondamentales des deux voyelles (de l'ordre de 30 μs).

FM. La modulation de fréquence (FM) a été proposée comme indice de ségrégation. L'hypothèse est séduisante car elle correspond au principe Gestalt de "destin commun" (Bregman, 1990). La modulation de fréquence d'un son confère à ses partiels des tracés fréquentiels communs, qui devraient, selon ce principe, les marquer comme appartenant au même son et distincts des autres. Pourtant des expériences récentes ont montré que la FM n'a guère d'effet (en dehors des effets liés à l'éventuelle ΔF_0 induite) (McAdams, 1989; Marin et McAdams, 1991; Demany et Semal, 1990; Carlyon, 1991, 1994). Une expérience menée en collaboration avec Cécile Marin a sondé cette question, avec un jeu de stimuli conçu pour tester une large gamme de mécanismes hypothétiques. La sensibilité de nos méthodes devait révéler d'éventuels effets

qui auraient pu échapper à d'autres auteurs. Les résultats furent essentiellement négatifs : nous n'avons trouvé aucun effet qu'on puisse attribuer de façon sûre à la FM [36, 36, 55].

Largeur de bande des formants. La largeur de bande des formants d'une voyelle isolée est connue pour avoir peu d'effet sur son timbre ou son identité. Une expérience a révélé qu'elle a un cependant un effet important sur l'identification de paires de voyelles concurrentes. Les stimuli comprenaient des paires de voyelles dont chacune avait une largeur de formants double ou moitié de la largeur "standard". Ces conditions étaient croisées avec des différences d'amplitude (de -25 à 25 dB par pas de 10) et de F_0 ($\Delta F_0 = 0, 6\%$). Il s'est avéré qu'une voyelle à formants étroits est (a) plus résistante au masquage, et (b) un meilleur masqueur qu'une voyelle à formants larges de même amplitude RMS [20, 55].

Effets spécifiques à des voyelles ou paires de voyelles. Un problème pour l'interprétation de ce type d'expérience est qu'on connaît très peu de choses du *mécanisme d'identification* des paires de voyelles concurrentes. Une expérience (citée plus haut) utilisait des paires de voyelles avec une large gamme de rapports d'amplitude (-35 à 35 dB, par pas de 10 dB). En construisant les stimuli, j'ai pris soin de donner aux deux voyelles le même spectre de phase, de façon à ce que l'enveloppe spectrale du stimulus puisse se calculer facilement à partir des enveloppes de chaque voyelle, en prenant la moyenne pondérée, sans que les effets complexes de sommation vectorielle ne viennent compliquer la structure du stimulus. J'ai ensuite analysé, pour chaque paire, les réponses obtenues à chaque rapport d'amplitude en les comparant au spectre. Cette analyse a permis de déterminer les indices sur lesquels s'appuie le système auditif pour juger que deux voyelles sont présentes dans un stimulus complexe et les identifier [4, 55].

D'autres expériences ont exploré l'identification de *triplets* de voyelles, les effets de continuité de F_0 , etc. [55], ou les effets de disparités binaurales (non publié).

Pour résumer, les améliorations apportées au paradigme classique d'identification de voyelles doubles ont produit un outil très sensible pour la mesure des phénomènes de ségrégation. Les expériences ont montré que système auditif utilise la stratégie d'annulation harmonique pour la ségrégation de sons harmoniques concurrents, mais ne semble pas utiliser la stratégie de renforcement harmonique, pourtant postulée par les théories d'analyse de scène auditive (Bregman, 1990) et utilisée dans divers systèmes computationnels.

Les mécanismes de ségrégation et d'identification ont été clarifiés, ainsi que les effets de nombreux paramètres qui les déterminent. Plusieurs modèles de ségrégation ont pu être écartés : la théorie PPA de Summerfield et Assmann (1991) [11], la théorie des battements de Culling et Darwin (1993, 1994) [2] et la théorie de sélection des canaux de Meddis et Hewitt (1992) [11, 2].

5.3 Filtre d'annulation neuronal

Le filtre d'annulation neuronal est constitué d'un neurone muni de deux synapses. L'une, excitatrice, est alimentée directement, et l'autre, inhibitrice, est alimentée via une ligne à retard. Les caractéristiques de ce neurone sont telles qu'il transmet toute impulsion provenant du circuit direct, *sauf* si elle coïncide avec une impulsion arrivant selon le circuit retardé. En supprimant certaines impulsions, le filtre modifie la statistique des intervalles inter-impulsions, en diminuant la densité des intervalles de longueur égale au retard.

Le filtre fut testé à l'aide de données enregistrées dans le nerf auditif du Cobaye en réponse à des stimuli faits d'un mélange de deux voyelles doubles (Palmer, 1990). Il s'est révélé efficace pour supprimer les corrélats de l'une ou l'autre voyelle [14]. Lorsque le paramètre retard du filtre est égal à la période de l'une des voyelles du mélange, la structure temporelle à la sortie du filtre ne reflète plus cette voyelle, mais uniquement sa concurrente. Il peut paraître étonnant qu'un circuit aussi simple ait une telle efficacité.

Une version simplifiée opérant sur des *probabilités* de décharge a été utilisée pour construire un modèle quantitatif permettant de prédire les résultats d'une expérience de perception de voyelles concurrentes [12]. Ce modèle est celui qui rend compte le mieux des résultats expérimentaux actuellement disponibles. Par rapport à celui de Meddis et Hewitt (1992), jusqu'ici le plus plausible, il a l'avantage de fonctionner même lorsque la voyelle-cible a une amplitude beaucoup plus faible que son concurrent. Rappelons que le modèle de Meddis et Hewitt opère une partition de la population de canaux périphériques (ou fibres du nerf auditif), selon qu'elles répondent ou non avec la périodicité de la voyelle dominante. Lorsque le rapport d'amplitude est grand, il peut arriver que cette voyelle domine *tous* les canaux. Dans ce cas la partition n'est plus possible, et le modèle de Meddis et Hewitt ne fonctionne plus. Le modèle à base de filtre d'annulation n'a pas cette faiblesse.

Le même filtre d'annulation sert de base au modèle de perception de la hauteur de la Section 6.1 [8], au modèle de perception de hauteurs multiples de la Section 6.2 [6, 29, 14], au modèle de décalage des partiels désaccordés (Hartmann et Doty, 1996) de la Section 6.3 [3]. La Section 3 a examiné la plausibilité physiologique d'un tel filtre, et les niveaux du système où il

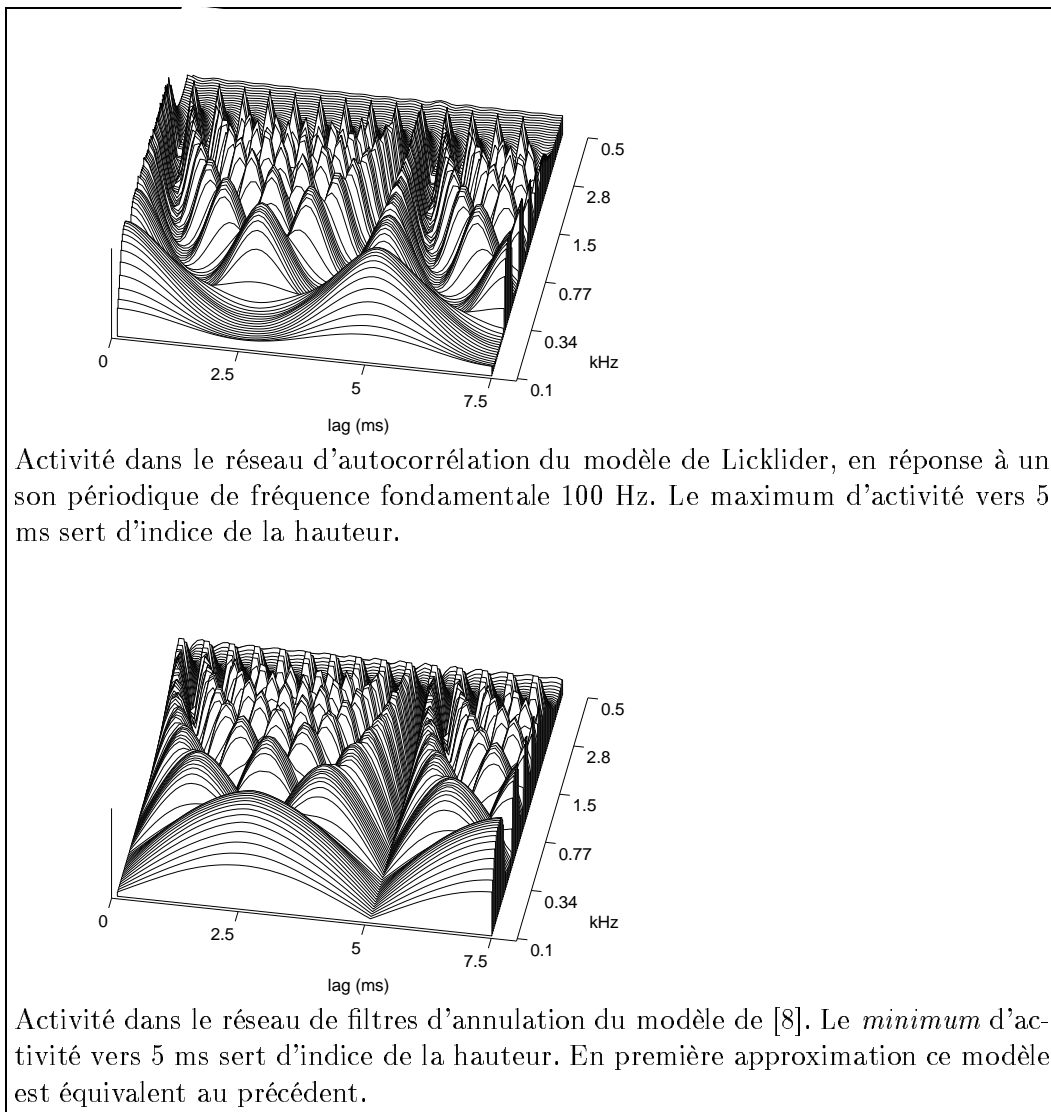
pourrait se trouver implémenté.

Le filtre d'annulation est donc une hypothèse plausible et applicable de façon assez large à des aspects multiples du traitement auditif.

6 Perception de la hauteur

6.1 Modèle "annulation" de perception de la hauteur

Dans [8] je propose un modèle de perception de la hauteur fondé sur le filtre d'annulation neuronal (Sect. 5.3. Le modèle est formellement équivalent au modèle d'autocorrélation de Licklider (1953), repris sous des formes diverses par divers auteurs (par exemple Moore, 1977 ou Meddis et Hewitt, 1991a,b).



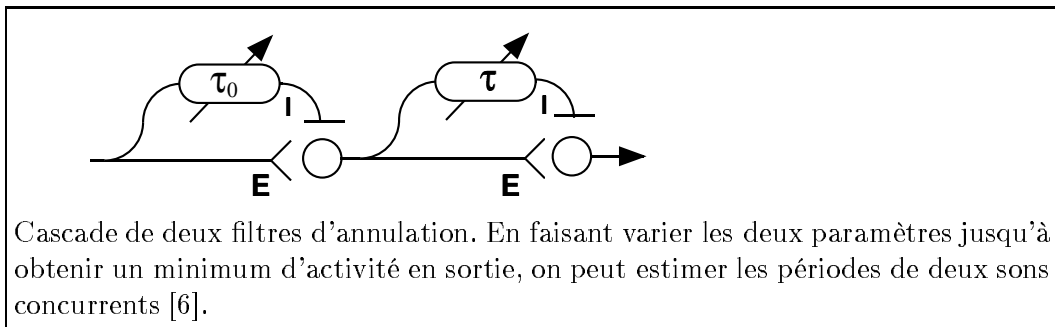
L'opération de multiplication qui intervient dans l'autocorrélation est

remplacée par une opération de soustraction, et la période est estimée par la recherche d'un *minimum* plutôt qu'un maximum. Malgré cette équivalence formelle, le modèle d'annulation présente par rapport à celui de Licklider quelques avantages qui sont détaillés dans [8], dont le plus intéressant est le lien qu'il entretient avec les modèles de ségrégation de la Sect. 5, et le fait qu'il peut être généralisé pour expliquer les hauteurs multiples de sons concurrents.

6.2 Modèle de perception des hauteurs multiples

Le filtre d'annulation sur lequel se fonde le modèle précédent peut être utilisé comme base d'un modèle de perception des hauteurs multiples qu'évoquent des sources périodiques multiples (par exemple des instruments jouant ensemble). Cela peut se faire de deux façons :

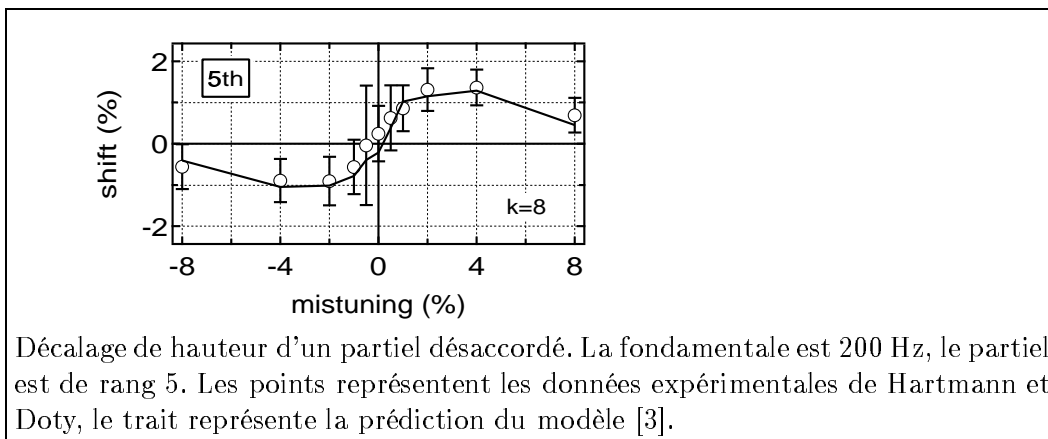
1. On peut associer un *modèle d'estimation de période* (celui du paragraphe précédent ou un autre) au filtre d'annulation de la Sect. 5. Une première estimation de période sert à ajuster le filtre d'annulation pour éliminer une des sources. Une deuxième estimation, faite à la sortie du filtre, permet d'obtenir la deuxième période. Ces étapes peuvent être répétées pour affiner les estimations, ou pour estimer les périodes d'autres sources.
2. On peut disposer deux (ou plusieurs) filtres d'annulation en cascade, et faire une recherche exhaustive de leurs paramètres de retard pour obtenir un minimum d'activité à la sortie du dernier filtre. Les paramètres correspondants fournissent les estimations des périodes. Ce modèle est en quelque sorte la généralisation multidimensionnelle du modèle du paragraphe précédent.



Ces principes sont décrits en détail dans [6, 29]. Les résultats d'une expérience d'évaluation de l'estimation de la F_0 avec des paires de voix naturelles sont décrits dans [14].

6.3 Modèles de décalage de hauteur des partiels désaccordés

Lors d'une expérience visant à explorer la fusion et ségrégation de partiels d'un son complexe, Hartmann, McAdams et Smith (1986, 1990) ont mis en évidence un phénomène inattendu. Lorsqu'on demande aux sujets d'apparier la hauteur d'un partiel désaccordé à celle d'un ton pur isolé, ils surestiment systématiquement le désaccord. Plus récemment, Hartmann et Doty (1996) puis Lin et Hartmann (1997) et Roberts et Brunstrom (1998; Brunstrom et Roberts, 1998) ont étudié le phénomène en détail. La tâche est plus facile pour une gamme de partiels de rang faible (2 à 5), mais des décalages peuvent aussi être mesurés à la fondamentale et pour des partiels de rang plus élevé (jusqu'à 13). La fonction qui relie décalage perceptif et désaccord physique a la forme d'un "S" : les décalages ont le même signe que le désaccord, tendent à être maximaux pour un désaccord de l'ordre de 4%, et décroissent au delà. L'amplitude de ce décalage varie entre sujets, mais la forme de la fonction reste plus ou moins la même.



L'effet est faible, mais intéressant car il paraît être la signature du mécanisme par lequel la hauteur du partiel est perçue. Nous disposons de nombreux modèles de perception de hauteur, qui rendent tous également bien compte des effets majeurs, et qu'il est donc difficile de départager par l'expérience. C'est pourquoi de tels effets mineurs prennent, de façon paradoxale, une grande importance théorique, car eux seuls permettent de trancher entre les modèles.

Pour tenter d'expliquer le phénomène de Hartmann et collègues, j'ai procédé en deux étapes.

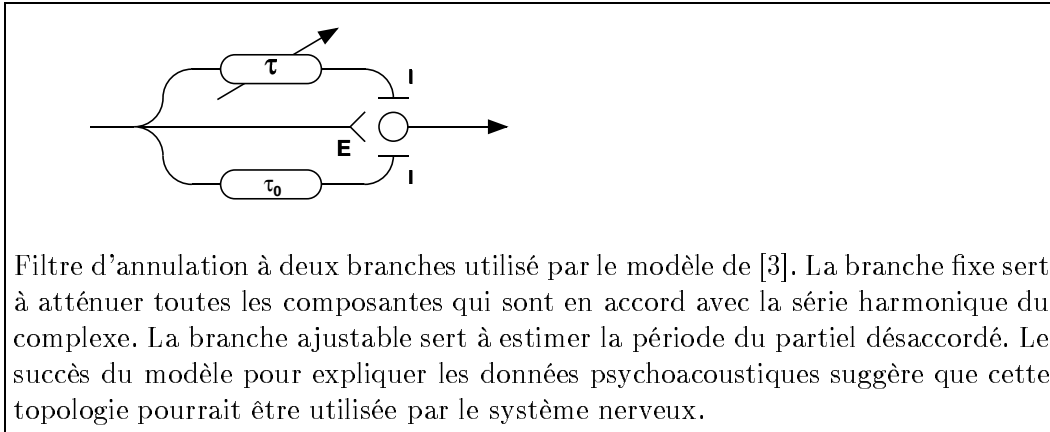
1. Dans une première étape, j'ai proposé un modèle *abstrait* probabiliste [9]. Ce modèle avait pour ambition d'expliquer le phénomène à partir

d'hypothèses les plus générales possibles : (1) l'existence d'une *représentation interne* de la fréquence du partiel, (2) un *bruit* affectant cette représentation, (3) un *peigne* ou *passoire harmonique* déterminant la tendance du partiel à fusionner avec le son complexe. Le décalage de hauteur était produit par une interaction probabiliste entre la représentation interne bruitée et le peigne harmonique. Aucune hypothèse n'était faite concernant la nature de la représentation fréquentielle, de la source de bruit qui l'affecte, ou du peigne harmonique. Le modèle rendait très bien compte des décalages observés [9, 30].

2. Dans une deuxième étape, j'ai proposé un modèle *physiologique* faisant au contraire des hypothèses précises sur la nature de la représentation interne (intervalles inter-impulsions), du bruit (distribution stochastique des instants des impulsions nerveuses) et du "peigne" (filtre d'annulation neuronal) [3]. Il s'agit en quelque sorte d'une instanciation du modèle abstrait précédent.

Le deuxième modèle postule un circuit nerveux dérivé du filtre d'annulation. Comme la version standard de ce filtre, il comprend un neurone "porte" muni d'une synapse excitatrice alimentée directement (depuis le nerf auditif ou depuis un relais plus central), et une synapse inhibitrice alimentée indirectement via une ligne à retard. Ce qui le distingue de la version standard du filtre d'annulation, c'est qu'il comprend aussi une deuxième synapse inhibitrice, alimentée par une deuxième ligne à retard. Comme dans le modèle "annulation" de perception de la hauteur[8], la période du partiel est estimée par la recherche d'un minimum d'activité en fonction de la valeur du retard de la *première* branche. Le retard de la *deuxième* branche est supposé fixe et égal à la période du complexe harmonique. Cette deuxième branche joue en quelque sorte le même rôle que la "passoire harmonique" de Duifhuis (1982), mais en opérant dans le domaine temps plutôt que fréquentiel. L'effet est d'atténuer les corrélats des composantes accordées du complexe harmonique, et d'accentuer ainsi la représentation du partiel désaccordé. Sans un mécanisme de ce type, la tâche d'appariement serait difficile, car rien ne distingue a priori le partiel désaccordé de ses voisins accordés. De fait, la tâche d'appariement est difficile lorsque tous les partiels sont accordés.

Comme son parent abstrait, le modèle rend très bien compte des données expérimentales. Ce qui est remarquable, c'est que ses paramètres sont moins nombreux que pour le premier modèle, et qu'un ensemble unique de paramètres rend compte (à un facteur d'échelle près) de l'ensemble des données de Hartmann et Doty, là où le modèle abstrait nécessitait plusieurs valeurs de paramètres pour accommoder les effets aux différents partiels. De plus, les prédictions du modèle physiologique varient très peu avec ses paramètres. Le



fait qu'elles s'accordent avec les données expérimentales malgré ce manque de "souplesse" est un argument supplémentaire en sa faveur.

Les effets de décalage s'obtiennent avec la forme très particulière de filtre d'annulation à deux retards décrit plus haut. Ils ne s'obtiennent pas avec d'autres topologies, à priori aussi plausibles. Ce modèle permet donc, en s'appuyant sur des données psychoacoustiques, de faire des inférences sur la topologie des circuits nerveux.

6.4 Modèle de l'asymétrie entre pics et creux de modulation de fréquence

Demany et ses collègues (Demany et McAnally, 1994 ; Demany et Clément, 1995a,b, 1997, 1998) ont mis en évidence une asymétrie perceptive entre *pics* et *creux* de modulation de fréquence. Un stimulus comprenant une succession de pics et creux de modulation est perçu comme une succession de tons correspondant aux seuls *pics* de modulation. Les pics ont ainsi une saillance perceptive plus grande que les creux. La discrimination de la fréquence extrême d'un pic est aussi bien meilleure que celle d'un creux, d'un facteur qui peut aller jusqu'à 5 ou 6 pour certaines conditions. Cette asymétrie s'observe pour une large gamme de paramètres (durée, fréquence d'extremum), et pour des types de stimuli assez différents (sinusoïdes, complexes, sons de Shepard, bruit modulé).

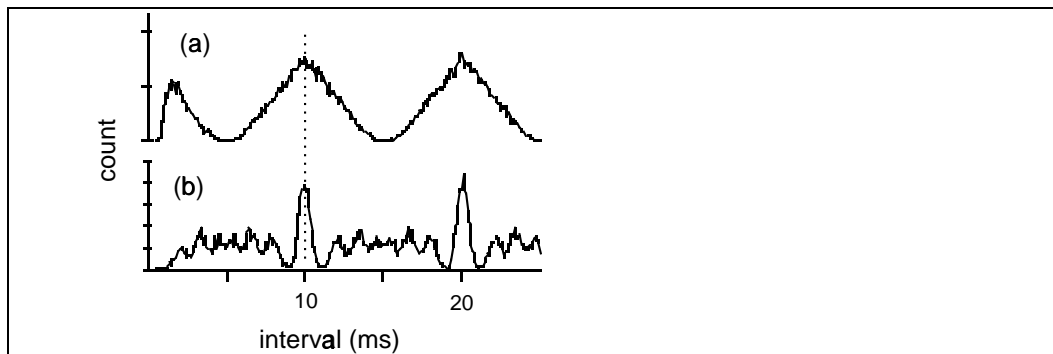
Il s'agit donc d'un phénomène robuste, mais pour lequel il n'existe guère d'explication. Les théories classiques de la hauteur ne prédisent pas d'asymétrie. Il est bien sûr possible de postuler un module dont le seul rôle serait d'introduire une telle asymétrie (Clément, 1996), mais une telle démarche *ad hoc* n'est pas entièrement satisfaisante. Il est plus attrayant de voir dans cette asymétrie la "signature" d'un mécanisme physiologique. L'émergence

d'une telle asymétrie dans un modèle serait alors un gage de sa plausibilité.

Dans leur forme standard, les modèles classiques de hauteur ne prédisent pas d'asymétrie, mais il est possible que celle-ci surgisse comme la conséquence d'un aspect de leur *implémentation*, qu'on peut alors modéliser. J'ai cherché à déterminer à quelles conditions cela pouvait être le cas pour un modèle à base de traitement temporel de type autocorrélation (Licklider, 1953; Meddis et Hewitt, 1992a,b). Il s'avère qu'une asymétrie adéquate émerge à condition de faire un certain nombre d'hypothèses qui sont détaillées dans [1].

6.5 Modèle de discrimination des tons purs

Le modèle autocorrélation rend bien compte de la plupart des phénomènes de hauteur, mais il est difficile de lui faire expliquer la précision de la discrimination de la hauteur, qui peut atteindre 0.2% vers 1 kHz, et aussi sa dépendance sur la durée du stimulus (Moore, 1973). En effet, le pic de la fonction d'autocorrélation qui sert d'indice de hauteur est relativement large, et sa position est difficile à préciser du fait du bruit stochastique. J'ai proposé un modèle inspiré de la notion de *fonction d'autocorrélation affinée* (narrowed autocorrelation function, Brown et Puckett, 1989). Ce modèle explique assez bien la finesse de la discrimination de la hauteur de tons purs, et aussi sa dépendance sur la durée et la fréquence [46, 43].



Fonction d'autocorrélation (a) et fonction d'autocorrélation affinée (b), calculés à partir de trains d'impulsions stochastiques produites par un modèle de nerf auditif. Le pic de la fonction d'autocorrélation est trop large pour rendre compte de la finesse de la discrimination de hauteur d'un ton pur (0.2% à 1 kHz). La fonction affinée rend bien mieux compte de cette discrimination, et aussi de la façon dont elle dépend de la durée et de la fréquence du stimulus (Moore, 1973) [46, 43].

7 Perception des voyelles

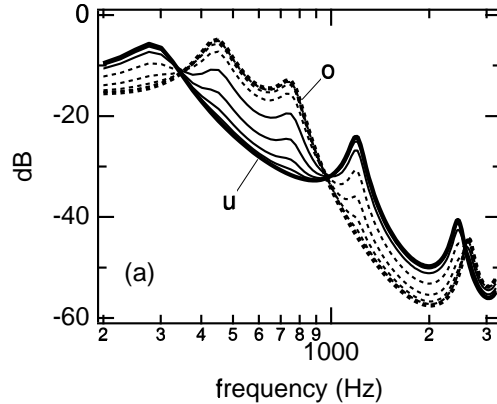
J'ai abordé en fait deux problèmes assez différents : l'*identification des voyelles doubles*, et l'*invariance perceptive* des voyelles simples face aux variations de la hauteur de voisement.

Le premier est issu des expériences sur la ségrégation des voyelles concurrentes décrites dans la Sect. 5. L'interprétation fine des résultats suppose une compréhension des processus d'identification des voyelles. Pour comprendre comment la ségrégation renforce les traits utiles à l'identification, il faut savoir quels sont ces traits. Or, on connaît mal les processus d'identification d'une voyelle lorsqu'elle est isolée, et à fortiori lorsqu'elle est en concurrence avec une autre voyelle. En fait, les expériences de perception de voyelles doubles peuvent être une bonne source d'information sur la perception de voyelles en général.

Le deuxième problème s'est imposé de façon assez fortuite lors d'un travail de collaboration avec l'équipe Reconnaissance de la Parole des laboratoires ATR, au Japon. Nous étions en train d'explorer deux idées : l'utilisation de l'information sur la fréquence de voisement (F_0) pour améliorer la qualité des indices acoustiques, et l'application de la Théorie des Données Manquantes (Missing Data Theory) pour gérer les données incomplètes produites par un module de ségrégation. Il est devenu apparent que ces deux idées pouvaient être combinées entre elles, et appliquées à un vieux problème de la Phonétique Acoustique, l'invariance de la qualité des voyelles face aux variations de la F_0 .

7.1 Identification des voyelles concurrentes

Quels traits sont utilisés pour identifier une voyelle, en particulier lorsqu'elle est en concurrence avec une autre voyelle ? Cette question a été abordée en marge d'une expérience dont le but était autre : l'étude des effets des différences de F_0 et d'amplitude relative. Le jeu de stimuli comprenait, parmi d'autres, des conditions où les F_0 des deux voyelles étaient les mêmes, et le stimulus avait donc un spectre harmonique similaire à celui d'une voyelle normale, mais avec une enveloppe spectrale composite résultant de la sommation des deux voyelles. Les stimuli étaient construits à partir de voyelles ayant toutes le *même spectre de phase*, de façon à ce que l'enveloppe spectrale du stimulus puisse se calculer facilement comme la moyenne des enveloppes des voyelles composantes, pondérées en fonction du rapport d'amplitude entre voyelles.



Enveloppes spectrales des voyelles /o/ et /u/ du japonais (traits épais), et enveloppes spectrales des stimuli obtenus en additionnant ces deux voyelles avec des rapports d'amplitude allant de -35 dB à +35 dB (traits fins). En rapprochant ces formes spectrales avec les réponses des sujets pour les stimuli correspondants, on peut déterminer quels indices spectraux sont nécessaires à l'identification d'une voyelle [4].

Ces conditions étaient répétées pour une large gamme de rapports d'amplitude (-35, -25, -15, -5, 5, 15, 25, 35 dB), et on disposait ainsi d'un large ensemble de stimuli "hybrides" entre deux voyelles, dont l'enveloppe spectrale était soit proche de l'une (à -35 dB) ou de l'autre (à 35 dB), soit intermédiaire. En observant les taux de réponse pour chacune de ces conditions et en les comparant avec l'enveloppe spectrale correspondante, il était possible de déterminer les traits qui déterminent l'identification de chaque voyelle[4].

Lorsque les F_0 des deux voyelles sont identiques, il apparaît que l'identification d'une voyelle est possible dès que ses formants F1 et F2 sont saillants dans l'enveloppe spectrale du stimulus (sans qu'on puisse écarter un rôle éventuel des formants d'ordre supérieur). La saillance éventuelle des formants de la voyelle concurrente n'est pas un obstacle, tant qu'elle ne diminue pas celle des formants de la voyelle cible. Lorsqu'il y a coïncidence entre des fréquences formantiques (F1 ou F2) des deux voyelles, le pic formantique commun est "partagé" par les deux voyelles. Il n'y a pas d'allocation exclusive à l'une ou l'autre voyelle.

Lorsque les F_0 sont différentes, l'amélioration de l'identification d'une voyelle s'interprète comme résultant de l'augmentation de la saillance de ses formants F1 et/ou F2. Encore une fois, l'augmentation de la saillance d'une voyelle ne se fait pas au détriment de l'autre : toutes deux peuvent profiter

de la ΔF_0 . Un formant bénéficie d'une ΔF_0 même s'il culmine à un niveau plus faible que celui de l'enveloppe spectrale de la voyelle concurrente à la même fréquence. Toutefois il ne faut pas qu'il soit dominé de plus de 5 à 10 dB : un formant plus faible ne profite pas de la ΔF_0 .

7.2 Modèle "données manquantes" du timbre des voyelles

On sait depuis longtemps que l'identité d'une voyelle (qualité, ou timbre) est fonction de son *enveloppe spectrale*, et en particulier de la position des deux ou trois premiers pics formantiques. En première approximation, elle ne dépend pas de la fréquence fondamentale. Mais lorsqu'on calcule le spectre à court terme à partir de l'onde, celui-ci présente au contraire une structure harmonique qui varie fortement avec la F_0 . Il en est de même des représentations de type fonction d'excitation qui pourraient être calculées par le système auditif. Comment le système auditif dérive-t-il une qualité invariante (timbre) à partir de représentations qui ne le sont pas ?

Ce problème est connu depuis longtemps. La solution généralement admise est de postuler un mécanisme hypothétique d'extraction d'un "invariant". On a par exemple proposé qu'existe une sorte de lissage (le chiffre de 3,5 Bark a été proposé comme largeur de fenêtre de lissage). Sur la base d'un raisonnement appuyé sur la Théorie de l'Échantillonnage, j'ai pu montrer qu'une telle hypothèse n'est pas suffisante : le lissage ne suffit pas à produire une représentation invariante, surtout si la F_0 est élevée. À la place, j'ai proposé un modèle fondé sur la Théorie des Données Manquantes.

Cette théorie a été proposée dans le cadre de la reconnaissance automatique de la parole (Cooke et al., 1996, 2000 ; Lippmann et Carlson, 1997). Elle a pour objet de permettre la reconnaissance de formes à partir de données incomplètes ou corrompues. Cela peut être le cas de parole masquée par du bruit, dont il manque certaines portions temporelles ou spectrales. On applique, dans l'étape de reconnaissance des formes, une fonction de pondération qui affaiblit ou annule l'apport des dimensions corrompues ou manquantes, en espérant que la partie qui reste suffira à la reconnaissance.

On peut appliquer la même idée dans le cas d'une voyelle voisée. L'information disponible, celle qu'on peut extraire de l'onde, se réduit aux seules valeurs de l'enveloppe spectrale échantillonnée aux harmoniques de F_0 . L'information nécessaire pour décrire la forme de l'enveloppe entre ces points est absente. Le théorème d'échantillonnage dit que l'enveloppe peut néanmoins être *reconstruite* parfaitement si la densité d'échantillonnage est suffisante par rapport à la complexité de la forme de cette enveloppe. La reconstruction se fait par interpolation ou simple lissage. Malheureusement si la F_0 est trop élevée (voix de femmes ou d'enfants, voix chantée) la densité est insuffisante.

Le lissage produit alors une distorsion (dite "par repliement du spectre" ou "aliasing") qui peut être sévère.

Plutôt que de procéder à une interpolation entre les points disponibles, comme le ferait un lissage, la Théorie des Données Manquantes restreint la reconnaissance de formes ("pattern matching") à ces seuls points. La reconnaissance se fait donc sur l'enveloppe spectrale *échantillonnée* aux seules fréquences des harmoniques. Dans le domaine spectral, cela se fait avec une fonction de pondération en forme de "peigne harmonique". Dans le domaine temps on peut faire une opération équivalente en tronquant la fonction d'autocorrélation [5].

Pour résumer, ce modèle explique l'invariance perceptive du timbre des voyelles en fonction de la F_0 , et apporte ainsi une réponse plausible à une énigme classique de la Phonétique (Klatt, 1982). Le modèle s'appuie sur la structure harmonique de la voyelle, et se rapproche ainsi des modèles de hauteur et de ségrégation. La notion de "données manquantes" qu'il exploite est utile dans les applications telles que la reconnaissance de la parole, mais c'est aussi un ingrédient intéressant pour les modèles perceptifs de tout ordre.

8 Applications

Je me suis intéressé aux applications par goût (le plaisir de construire), et par respect du principe selon lequel la validité d'une connaissance se juge à l'efficacité de l'*action* qu'elle guide. On peut y voir la transposition, dans le domaine des idées, du principe selon lequel tout système perceptif est à l'origine un système *actif* (Szentágothai et Arbib, 1975). La valeur d'un système perceptif se juge à la qualité de l'action qu'elle guide, qui se mesure à son tour à la survie de l'individu ou de l'espèce. En transposant ce critère écologique aux idées, on peut les juger à la survie de la théorie, du chercheur ("publish or perish") ou de la société qui l'accueille. Quoi qu'il en soit, il est évident que se frotter aux applications est une bonne façon de découvrir les faiblesses d'une théorie.

Il importe néanmoins de faire une distinction claire entre *modèles* et *méthodes*, dont les objectifs et critères de validation sont différents. Trop souvent on voit défendre un modèle peu plausible sous prétexte qu'il est "computationnellement efficace", ou une méthode inefficace sous prétexte que "c'est ainsi que procède le système auditif". Les critères d'évaluation ne sont pas les mêmes pour les modèles et les méthodes, et rien ne dit qu'une même construction intellectuelle puisse satisfaire aux deux. Ceci dit, rien n'empêche d'établir des *parallèles* fructueux entre modèle et méthode, mais ils sont souvent à établir à un niveau assez abstrait.

Les applications que j'ai abordées sont l'estimation de la Fréquence Fondamentale (F_0) de la parole, l'estimation de F_0 s de voix ou instruments multiples, la suppression de sources parasites (pour la reconnaissance de la parole), et l'estimation spectrale en tenant compte de la périodicité.

8.1 Estimation de la F_0 de la parole

La fréquence fondamentale (F_0) est la fréquence de vibration des cordes vocales. Si la vibration est périodique et soutenue, on peut lui associer le corrélat perceptif de *hauteur tonale*. Cependant, dans le cas de la parole naturelle les portions soutenues peuvent être de courte durée, et il n'est pas évident que le voisement donne toujours lieu à une hauteur perçue en tant que telle. Il est plus sage de parler d'"intonation", dont la F_0 est un ingrédient en association avec l'intensité et la durée. L'estimation de la F_0 est un problème réputé difficile, sur lequel ont travaillé des générations de chercheurs. Je l'ai abordé épisodiquement depuis de nombreuses années, en commençant par un stage de DEA au Laboratoire de Phonétique de l'Université Paris 7.

A l'époque j'ai exploré des méthodes dites "temporelles" consistant à extraire de la forme d'onde des "marqueurs", supposés liés à un événement

précis du cycle de vibration glottique (instant de fermeture glottale). La F_0 s'estime alors comme inverse des intervalles inter-marqueur. Mes efforts n'ont pas abouti à des résultats plus satisfaisants que des méthodes précédentes du même type. Outre une certaine frustration, il m'est resté (a) la conviction de l'utilité d'une évaluation formelle, (b) une méfiance envers l'ajustement de paramètres "jusqu'à ce que ça marche" (tinkering), et (c) un doute sur l'existence de ces fameux "marqueurs temporels", et sur la possibilité de les extraire de façon invariante par rapport à la variabilité de l'onde de parole³.

La difficulté est que la position des marqueurs, voire leur nombre, est susceptible de varier de façon "catastrophique" (au sens mathématique qu'une variation infime du signal résulte en une modification importante du pattern de marqueurs). Le déplacement intempestif d'un marqueur modifie la valeur des intervalles qui le précèdent et le suivent, et donc perturbe l'estimation de F_0 . Pour gagner en robustesse, on a intérêt à appuyer la mesure d'intervalle sur des repères *distribués*, plutôt que concentrés en un point, et à l'extrême à utiliser chaque valeur du signal comme repère de comparaison. C'est ainsi que fonctionnent une classe de méthodes dites "décalage et comparaison"[53, 52], parmi lesquelles on compte la méthode de l'*autocorrélation* et celle de l'*AMDF* (Average Magnitude Difference Function, Ross et al. 1974). L'estimation se faisant par comparaisons le long de l'axe temps, ces méthodes s'apparentent aux méthodes temporelles. Cependant, la mesure de comparaison doit, pour être stable, être moyennée sur une fenêtre. C'est ce qui explique leur classification par Hess (1983) parmi les méthodes d'*analyse à court terme*, à côté d'autres méthodes dont la plupart sont spectrales.

J'ai porté mes efforts sur la méthode AMDF, et en particulier :

- L'évaluation [62, 57].
- Des améliorations apportées à cette méthode[62, 57, 45], en particulier l'application d'un banc de filtres inspiré du modèle de Licklider (1953)[45].
- L'extension de la méthode AMDF à l'estimation des F_0 de voix multiples [65, 42, 14, 29, 6].
- L'exploration de méthodes fondées sur la fréquence instantanée, en collaboration avec H. Kawahara[59, 32, 31, 56, 58, 26, 27, 28, 7].

Le choix de méthodes de type "décalage et comparaison" est cohérent avec les modèles perceptifs de type temporel, par exemple le modèle auto-corrélation de perception de la hauteur (Licklider, 1953; Meddis et Hewitt,

³C'est une des raisons de ma réticence envers le modèle "STI" (strobed temporal integration) de Patterson et al. (1992), qui postule l'utilisation de marqueurs de ce type par le système auditif

1991a,b). Le choix plus particulier de la méthode de l'AMDF est cohérent avec le modèle "annulation" de la Section 6[8].

Pour résumer, j'ai abordé la question de l'estimation de la F_0 à plusieurs reprises, et selon des angles d'approche assez différente. Ce travail est encore incomplet : il reste à faire une évaluation formelle des méthodes les plus nouvelles, et à insérer ces méthodes dans des outils informatiques utilisables dans la pratique. Ces objectifs ont été repoussés par manque de temps.

8.2 Séparation des sources

J'ai appliqué les deux principes d'annulation et renforcement à la tâche de *suppression d'une voix parasite* dans un système de reconnaissance de la parole[61, 39, 34]. Le système de reconnaissance était rudimentaire (reconnaissance de mots par programmation dynamique) mais suffisant pour tester l'efficacité des méthodes de ségrégation de voix. Bien qu'il s'agisse d'une tâche appliquée, l'objectif était surtout de comparer les deux principes théoriques exposés dans la Sect. 5. De cette expérience j'ai pu conclure :

- La stratégie d'annulation harmonique est efficace.
- Celle de renforcement harmonique est généralement peu efficace, et sa combinaison avec celle d'annulation n'apporte guère d'amélioration par rapport à l'annulation seule.
- L'annulation harmonique introduit néanmoins une distorsion spectrale, du fait qu'elle élimine toutes les composantes du signal qui appartiennent à la série harmonique de l'interférence, y compris celles qui appartiennent à la cible.

C'est ce dernier point qui a motivé mon intérêt pour la Théorie des Données Manquantes (TDM). En effet, les effets néfastes de la distorsion spectrale produite par le processus d'annulation harmonique peuvent être atténués en appliquant la *même distorsion* aux patrons de référence. Cette opération peut s'interpréter comme l'application d'une pondération non-uniforme dans le processus de reconnaissance de formes [61]. Elle préfigure les techniques "données manquantes" proposées par Cooke, Morris et Green (1997).

La TDM est un ingrédient utile pour la reconnaissance des formes, car elle permet de pondérer chaque partie d'un pattern en fonction de sa *fiabilité*, à condition que cette fiabilité puisse être estimée au moment de la reconnaissance⁴. L'annulation harmonique, la séparation aveugle, la décomposition en composantes indépendantes (ICA) supposent (dans le cas convolutif) une matrice convolutive qui introduit une distorsion spectrale. La TDM permet de

⁴Par hypothèse le pattern de fiabilité est supposée différent entre la phase d'apprentissage et la phase de reconnaissance.

gérer cette distorsion, et d'autres qui peuvent surgir du fait de l'application de modèles d'Analyse de Scènes Auditives. Elle permet aussi la combinaison de sources disparates d'information (par exemple audio-visuelle). Ces idées, qui restent à exploiter de façon plus approfondie, sont développées dans [34, 22, 24].

8.3 Harmonicité et estimation spectrale

8.3.1 Modèle source-filtre et échantillonnage

De nombreux sons de la parole ou d'instruments de musique sont produits par l'excitation d'un résonateur par une source, qui varie chacun de façon indépendante. De la forme du résonateur dépend la *fonction de transfert* et donc l'*enveloppe spectrale* qui décrit l'amplitude des partiels en fonction de leur fréquence⁵. À *résonateur constant*, des sources de F_0 différente produisent des sons dont le timbre est approximativement le même, mais la hauteur différente. À *source constante*, des résonateurs de forme différente produisent des sons dont la hauteur est la même, mais le timbre différent. Il arrive souvent qu'une information importante soit associée à la forme du résonateur, et que pour l'extraire du signal on doive estimer son enveloppe spectrale. Par exemple, un système d'analyse-synthèse a besoin d'extraire l'enveloppe pour pouvoir l'exciter par une nouvelle source, afin par exemple de manipuler l'intonation de la voix. Autre exemple, un système de reconnaissance de la parole doit estimer des "traits acoustiques" qui soient invariants par rapport aux variations de F_0 .

Pour certains sons, dont la parole, la forme du résonateur varie au cours du temps. En toute rigueur, cette variation compromet la validité du modèle linéaire invariant (LTI) sur laquelle repose la notion de fonction de transfert, et donc d'enveloppe spectrale. Néanmoins, si ces variations sont lentes par rapport à la durée de la réponse impulsionnelle du résonateur, on peut parler d'*enveloppe spectrale variable en fonction du temps*, qui peut se concevoir comme une surface définie sur les deux dimensions de fréquence et temps.

L'estimation de l'enveloppe spectrale présente des difficultés de plusieurs ordres. Pour commencer, si la production de l'onde acoustique par le résonateur excité par la source est un processus sans ambiguïté, il n'en est pas de même du problème inverse. Plusieurs combinaisons de source et résonateur peuvent produire la même onde acoustique. L'estimation du résonateur à partir du signal est donc ambiguë, et ne peut se faire que si la source est connue.

⁵Pour simplifier je suppose que la source a une enveloppe spectrale plate, et qu'on peut donc assimiler l'enveloppe spectral du signal à la fonction de transfert du résonateur qui l'a formé

Même dans ce cas, la qualité de l'estimation dépend de facteurs tels que la nature de la source, et la durée observable du signal. Ces problèmes, qui existent déjà pour un résonateur fixe, sont exacerbés si le résonateur est variable. La difficulté de la tâche est intuitivement évidente si l'on songe qu'elle revient à estimer l'enveloppe spectrale, une fonction de deux variables (temps et fréquence), à partir d'un signal fonction d'une seule variable (temps).

Dans la pratique, l'estimation spectrale se fait par des techniques qui exploitent le signal sur un intervalle restreint (fenêtre) pour fournir une estimation propre à cet intervalle. Le déplacement de l'intervalle d'analyse dans le temps permet alors de développer une surface spectro-temporelle censée approximer l'enveloppe spectrale. Le résultat dépend du choix de la méthode (transformée de Fourier à court terme, LPC, etc.) et de ses paramètres (taille et forme de la fenêtre notamment). Outre le problème des limites de la qualité de l'approximation obtenue, se pose celui de son manque d'*unicité*. Le fait qu'on puisse dériver plusieurs estimations de formes différentes à partir du même signal est un facteur de confusion.

La taille et la forme de la fenêtre d'analyse déterminent la résolution d'analyse : une analyse temporellement fine est spectralement grossière et vice-versa. Cette relation inverse (principe d'incertitude) est souvent perçue comme un obstacle à la qualité de l'estimation spectrale. En fait, nous souffrons surtout du problème contraire d'*excès* de résolution. Une fenêtre trop grande produit un spectre trop détaillé (par rapport à l'enveloppe), alors qu'une fenêtre trop petite produit une estimation trop variable au cours du temps. Aucun choix ne permet d'avoir une estimation stable et lisse dans les deux dimensions (comme on suppose être l'enveloppe spectrale). Ce problème n'est résolu ni par l'analyse en ondelettes, ni par un filtrage "auditif" (Klatt, 1982).

8.3.2 Lissage et échantillonnage

Face à cet excès de résolution, on a coutume d'opérer des lissages supplémentaires : moyennage temporel de spectres consécutifs (méthode de Welch), moyennage spectral, lissage cepstral (en limitant le cepstre à ses coefficients d'ordre faible), etc.. Ces méthodes, et d'autres telles que la LPC, permettent de réduire la résolution spectro-temporelle pour obtenir une enveloppe plus lisse, et donc, on l'espère, plus proche de l'enveloppe d'origine. Une fois de plus, le résultat dépend du choix de la méthode de lissage et de ses paramètres. On peut se poser la question de l'adéquation pour un signal donné, et de l'*unicité* de l'estimation produite.

La question du lissage est étroitement liée à celle de l'*échantillonnage*. On sait qu'une fonction temporelle peut-être reconstruite à partir de ses

échantillons si elle obéit à la condition de Nyquist, c'est à dire qu'elle ne contient pas de composantes de fréquence supérieure à la moitié de la fréquence d'échantillonnage. La reconstruction se fait par interpolation entre les échantillons, autrement dit, par lissage. La même idée vaut pour des fonctions d'autres variables que le temps, et en particulier pour une enveloppe spectrale fonction de la fréquence. La seule difficulté est alors d'éviter la confusion lorsqu'on remplace "temps" par "fréquence" et "fréquence" par "délai" ou "qu'éfrencé"⁶. "Fréquence de Nyquist" devient alors "délai de Nyquist", etc. Une enveloppe spectrale peut être échantillonnée, puis reconstruite à partir de ses échantillons, si sa forme n'a pas de composantes au delà du délai de Nyquist. Pour dire la même chose de façon simple : l'enveloppe est parfaitement représentée par la série d'échantillons sauf si sa forme est trop compliquée.

Dans la production de sons tels que la parole l'échantillonnage intervient à plusieurs niveaux :

1. Lors de la production, l'enveloppe spectrale est échantillonnée spectralement par les raies d'une source périodique, et/ou temporellement par les impulsions temporelles qui la composent.
2. Lors d'une analyse (par exemple spectrale), afin de diminuer le coût de calcul les opérations sont faites à intervalles fixes le long de l'axe temporel (par exemple tous les 25 ms). Il s'agit là aussi d'une forme d'échantillonnage. L'axe fréquentiel est lui aussi échantillonné par la représentation spectrale discrète.
3. Lors d'une éventuelle resynthèse, l'enveloppe spectrale estimée sera *ré-échantillonnée* par la nouvelle source.

Le niveau (1) ne pose pas de problème si l'enveloppe spectrale est suffisamment simple, et/ou son échantillonnage suffisamment dense. Dans le cas contraire il y a sous-échantillonnage, donc perte d'information et risque de distorsion. Comme je montre dans [5] cela peut être le cas pour la parole. Lorsque l'enveloppe est sous-échantillonnée lors de la production, il est impossible de l'estimer de façon exacte à partir du signal, à moins de disposer de contraintes supplémentaires (par exemple celles d'un modèle de production).

Le niveau (2) peut se concevoir comme l'échantillonnage du résultat d'une analyse *courante* (running analysis)⁷. L'analyse courante produit à chaque instant une série de coefficients (par exemple spectraux), qui chacun peut être vu comme une fonction du temps. Une analyse trame-par-trame est

⁶Verlan de l'analyse cepstrale : "spectre" devient "cepstre", "fréquence" devient "qu'éfrencé", "filtre" devient "lifre", etc.

⁷Une analyse courante est calculée à chaque instant, ou en pratique à chaque échantillon du signal, plutôt qu'à une périodicité plus grande

équivalente à un échantillonnage (à une périodicité plus grande) de ces fonctions. Son adéquation dépend de l'étendue du spectre de chaque fonction par rapport à la fréquence de Nyquist.

L'étendue du spectre d'un paramètre dépend de la façon dont il a été calculé. L'analyse fait généralement intervenir un moyennage sur une fenêtre qui limite l'étendue spectrale et tend donc à satisfaire la condition de Nyquist. Cependant la rejection des composantes de fréquence élevée n'est pas absolue, et la condition de Nyquist n'est donc pas remplie de façon parfaite. Un échantillonnage insuffisant a pour conséquence l'introduction d'une *distorsion due au repliement du spectre* (aliasing) lorsque les échantillons sont interpolés ou rééchantillonnés.

Le niveau 3 correspond à un rééchantillonnage pour une resynthèse. Supposons qu'une estimation spectrale faite sur un signal a produit une enveloppe spectrale fonction du temps (comme au niveau 2). On utilise alors cette enveloppe pour synthétiser un nouveau son (comme au niveau 1). Pour cela on doit échantillonner l'enveloppe, et pour que cet échantillonnage soit adéquat il faut que l'enveloppe estimée satisfasse à la condition de Nyquist, c'est à dire qu'elle n'ait pas de composantes au delà de la fréquence (ou délai) de Nyquist. Or, on l'a vu au niveau 2, il n'est pas certain que les lissages faits au cours de l'estimation suffisent à supprimer toutes les composantes de fréquence élevée. La conséquence de leur présence est une distorsion du son resynthétisé (qu'on peut concevoir comme due à un phénomène de "moiré" entre la périodicité du d'origine et la périodicité de resynthèse). C'est probablement un facteur de la mauvaise qualité de la parole synthétique.

Pour résumer, lissage, échantillonnage, et estimation spectrale sont étroitement liés. Le lissage sert à reconstruire l'enveloppe après échantillonnage, et aussi à limiter sa composition spectrale (selon les deux axes de temps et fréquence) avant rééchantillonnage. Un lissage insuffisant peut résulter en une distorsion par repliement du spectre. Inversement, un lissage trop important élimine une partie du détail de l'enveloppe spectral et aboutit à une perte d'information.

8.3.3 Périodicité et lissage optimal

Les questions posées plus haut sont pertinentes quel que soit le type d'excitation, mais elles se posent d'une manière particulière lorsque l'excitation est *périodique*, et cela pour deux raisons. La première est que les distorsions peuvent prendre une forme systématique (effets de moiré) qui pourrait augmenter leur saillance perceptive. La seconde est que des solutions particulières existent pour traiter le cas périodique et obtenir un lissage "optimal". Ces solutions sont l'objet des travaux sur la gestion de la périodicité que j'ai

menés avec Hideki Kawahara (Kawahara, 1997) [5, 54, 19, 32, 58, 31, 7].

Si le signal est périodique il est possible d'obtenir un lissage *parfait* des paramètres qu'on en extrait. Pour comprendre pourquoi, il suffit de se rappeler que toute transformation déterministe d'un signal périodique est périodique et de même période (ou d'une période sous-multiple). C'est le cas de chaque paramètre produit par une analyse "courante". Un signal périodique comprend une composante continue et une série de composantes de fréquences multiples de la fondamentale F_0 (spectre harmonique). Seule nous intéresse la composante continue (ou qui varie lentement). Toutes les autres sont superflues, voire gênantes dans la mesure où elles peuvent contribuer à violer la condition de Nyquist en cas de rééchantillonnage. Heureusement, elles peuvent être éliminées si on connaît la fréquence fondamentale, par convolution avec une fenêtre temporelle de forme et taille adéquates. Une fenêtre carrée de durée égale à la période (ou son multiple) convient, de même que toute fenêtre obtenue par convolution avec cette fenêtre carrée. Des remarques analogues s'appliquent aussi au domaine spectral pour éliminer sa structure périodique (spectre de raies). Ce lissage est adaptatif, ses paramètres s'ajustant en fonction de la F_0 .

Lissage spectral et temporel peuvent se faire en une fois par convolution 2D avec un "noyau" dont la forme s'adapte à la périodicité du signal. C'est le principe du système d'analyse-synthèse STRAIGHT de Kawahara. L'excellente qualité acoustique de ce système tient probablement à l'élimination des effets de "moiré" mentionnés plus haut.

Des problèmes analogues se posent lors de l'estimation d'indices acoustiques en reconnaissance de la parole. On a coutume de calculer ces indices trame par trame, ce qui implique un échantillonnage, mais on se préoccupe rarement de savoir si le contenu spectral des indices estimés (considérés comme des séries temporelles) est suffisamment limité pour supporter un tel échantillonnage. Un échantillonnage insuffisant produit une erreur dans les estimations qui risque d'être gênante en particulier pour des *paramètres différentiels* tels que le delta-cepstre, ou les coefficients de Fourier des séries de coefficients (exploités dans la méthode RASTA, Hermansky et Morgan, 1994 ; Kanadera, Hermansky et Arai, 1998). Le problème est d'autant plus sévère que la F_0 est faible. Des problèmes analogues de sous-échantillonnage selon l'axe spectral se posent lorsque F_0 est élevé. Le lissage adaptatif permet de minimiser ces problèmes. L'application de la F_0 à l'estimation des paramètres de reconnaissance de la parole reste à développer.

Le lissage adaptatif fournit le meilleur lissage possible en termes d'élimination des résidus de périodicité. Cependant il ne permet pas l'estimation fidèle de l'enveloppe d'origine si celle-ci a été sous-échantillonnée lors de la production, c'est à dire si l'enveloppe contient des composantes de fréquence

“spatiale” (ou “qu’éfrentielles”) au delà des limites de Nyquist. Dans ce cas il est impossible d’estimer l’enveloppe de façon exacte, à moins de disposer de contraintes complémentaires telles que fournies par un modèle de production. Du sous-échantillonnage il résulte une *perte d’information*, et aussi éventuellement une *distorsion* comme souligné plus haut, dû au repliement du spectre lorsque les échantillons sont interpolés. Dans une application de *reconnaissance des formes* les conséquences néfastes de ce dernier facteur peuvent être limitées en faisant appel aux techniques de TDM. Cette idée est exposée dans [5] où elle a été appliquée à un modèle de perception de voyelles (Sect. 7). Elle est susceptible d’une application plus large, par exemple en Reconnaissance de la Parole.

À signaler enfin la possibilité d’utiliser la TDM pour contraindre un *modèle de production*. Le modèle est contraint, dans le domaine spectral, par les seules valeurs aux harmoniques de F_0 de l’enveloppe spectrale qu’il produit.

Pour résumer, l’estimation d’une enveloppe spectrale est rendue difficile par l’échantillonnage qui a lieu lors de la production. Une distorsion par repliement du spectre (aliasing) peut se produire lors de l’estimation si l’enveloppe spectrale d’origine n’était pas assez lisse pour obéir aux conditions de Nyquist. Une distorsion supplémentaire peut se produire lorsque l’enveloppe estimée est rééchantillonnée, par exemple lors d’une resynthèse. Cette distorsion dépend du contenu spectral des paramètres fournis par l’analyse, et donc de leur lissage. Le lissage doit être suffisant pour éviter cette distorsion, mais un lissage trop important ferait perdre de l’information.

Dans le cas d’un *signal périodique* il est possible d’obtenir un excellent compromis. La convolution avec une fenêtre de taille adaptée en fonction de la F_0 élimine toute fluctuation due à la périodicité. Ce lissage est optimal, mais il ne permet cependant pas de reconstruire l’enveloppe d’origine si celle-ci a été sous-échantillonnée lors de la production. Néanmoins, dans une application de *reconnaissance des formes* on peut éviter les effets néfastes de ce sous-échantillonnage en appliquant les techniques de la Théorie des Données Manquantes. La TDM peut permettre aussi la reconstruction exacte de l’enveloppe spectrale si on dispose d’un *modèle de production*.

9 Conclusion

L'Introduction a rappelé l'importance historique de l'hypothèse de traitements auditifs dans le domaine temps, notamment pour la perception de la hauteur. Les Sections 2 et 3 en ont esquissé les possibles substrats physiologiques. La Section 5 a présenté des expériences qui sondent le rôle de la périodicité dans la ségrégation perceptive des sources, et un modèle de traitement pouvant en rendre compte : le filtre d'annulation neuronal. La Section 6 s'est concentré sur des modèles dérivés du modèle précédent, qui rendent bien compte de la hauteur tonale. La Section 7 a abordé la question de la perception du timbre des voyelles, et le rôle que peut jouer la périodicité dans ce cas. Enfin, la Section 8 a passé en revue des tentatives d'application de ces modèles pour des applications pratiques.

J'espère avoir réussi à imposer une certaine unité à cet ensemble de travaux hétéroclites. J'espère surtout que les idées développées ici seront utiles à l'avenir pour améliorer notre compréhension du système auditif.

Annexe - Les limites de l'Analyse de Fourier

L'Analyse de Fourier est un outil puissant qui rend de grands services en Audition. Cependant son succès même fait qu'on en fait un usage systématique et parfois imprudent. Cela a suscité en retour des critiques, exprimées parfois de façon confuse. L'objectif de cette section est de formuler, de façon la plus claire possible, certaines des critiques qu'on peut adresser à l'utilisation de l'Analyse de Fourier en Audition.

9.1 Systèmes linéaires invariants

Le domaine d'application par excellence de l'Analyse de Fourier est celui des *systèmes linéaires invariants dans le temps* (LTI).

La transformée de Fourier décompose tout signal en une somme d'exponentielles complexes (sinusoïdes). Ces fonctions forment une base de l'espace des signaux, et la transformée de Fourier projette le signal sur cette base. L'utilité d'une telle projection est que les éléments de la base sont *vecteurs propres* des systèmes LTI. Un vecteur propre par définition ne "change pas de direction" quand on le transforme : l'effet de la transformation est de multiplier le vecteur par un scalaire, appelé *valeur propre*. Dans le cas présent cela signifie que la transformée d'une sinusoïde est une sinusoïde de même fréquence mais dont l'amplitude a changé d'un certain facteur. L'ensemble des valeurs propres pour toutes les fréquences constitue la *fonction de transfert* du système.

Cette simplicité de comportement des sinusoïdes dans un système LTI fait la force de l'Analyse de Fourier. Pour connaître l'effet du système sur un signal quelconque, il suffit de faire les gestes suivants : (1) décomposer le signal sur la base (= Transformée de Fourier), (2) pour chaque dimension (= chaque fréquence), multiplier chaque coordonnée du signal (= chaque valeur de son spectre) par la valeur propre (= valeur de la fonction de transfert) correspondant à cette dimension, et enfin (3) appliquer la transformée de Fourier inverse. Le résultat est exactement le signal qui sortira du système. Cette succession d'opérations peut paraître complexe, mais c'est de loin la façon la plus facile d'obtenir ce résultat.

Ce rôle privilégié de l'Analyse de Fourier est spécifique aux systèmes LTI. Si au contraire le système *varie dans le temps*, ou s'il n'est *pas linéaire*, les exponentielles complexes ne sont plus vecteurs propres et la décomposition perd une partie de son intérêt. Qu'en est-il du système auditif ? En ce qui concerne l'invariance, il existe divers phénomènes d'adaptation, et l'état du système est susceptible de changer surtout aux étages supérieurs. Cependant l'hypothèse d'invariance dans le temps est souvent raisonnable en régime

stationnaire pour les parties périphériques. En ce qui concerne la linéarité, on peut faire deux réponses, une simple et l'autre plus compliquée.

Commençons par la réponse compliquée. La linéarité est une propriété d'une transformation d'un *espace vectoriel* sur un autre *espace vectoriel*. Cette qualité d'espace vectoriel pour les ensembles de départ et d'arrivée est nécessaire pour qu'ils soient complets pour les opérations de *somme* et de multiplication par un scalaire qui interviennent dans la définition de la linéarité⁸. Il ne fait pas de doute que l'ensemble de départ des *vibrations acoustiques* dans l'air forme un espace vectoriel. Il en est de même des vibrations des osselets et de chaque point de la membrane basilaire. Les transformations entre elles peuvent être ou ne pas être linéaires mais la notion de linéarité a certainement un sens. La réponse est moins évidente en ce qui concerne les *signaux nerveux*. Certes, on peut décrire les potentiels d'excitation post-synaptique (EPSP et IPSP) comme des fonctions numériques du temps, qui forment bien un espace vectoriel. En revanche des difficultés surgissent pour les décharges dans une fibre. Comment définir la "somme de deux décharges", ou l'"opposé d'une décharge"? La question est encore plus ardue pour les patterns complexes dans une population de neurones. Enfin, dans le cas de *sensations* ou *percepts*, parler d'espace vectoriel n'a guère de sens. Quel sens donner à la "somme de deux percepts", ou l'"opposé d'un percept" (c'est à dire un percept qui, "ajouté" à un autre percept, produirait le "percept nul")? Faute d'espace vectoriel à l'arrivée, il est difficile de donner une réponse sensée à la question "le système auditif est-il linéaire".

Passons à la réponse simple. La transmission acoustique dans l'oreille externe et moyenne est linéaire. Le filtrage cochléaire est non-linéaire, mais il peut souvent être décrit comme approximativement linéaire. La transduction au niveau de la cellule ciliée est réputée non-linéaire (la relation entre déplacement ou vitesse et *probabilité de décharge* est celle d'un redresseur simple alternance). Au-delà on ne peut plus parler de linéarité, non pas que les transformations soient "non-linéaires", mais parce qu'elles se font entre espaces qui n'ont rien de vectoriel. Des techniques d'analyse linéaire peuvent donner des résultats utiles jusqu'au cortex (Shamma, 2000), mais il faut garder à l'esprit la fragilité de l'hypothèse de linéarité qui les sous-tend.

On peut cependant apporter une nuance. Pour établir un modèle fidèle de l'environnement, le système auditif est amené à tenter d'associer à chaque source distincte un percept distinct ("écoute analytique"). Lorsqu'il y parvient, son comportement évoque celui d'un système "linéaire", puisque les corrélats de chaque source paraissent superposés dans le modèle interne du

⁸Une transformation T est linéaire si pour toute paire d'éléments a et b de l'espace de départ, et tout scalaire λ , on a $T(a + b) = T(a) + T(b)$, et $T(\lambda a) = \lambda T(a)$

monde. Il faut voir dans cette superposition (qui fait penser à une superposition linéaire) le résultat de mécanismes sophistiqués d'analyse de scène auditive, plutôt que l'indication que les représentations qui sous-tendent son fonctionnement sont elles-mêmes linéaires.

La "non-linéarité" du système auditif restreint l'utilité de la transformée de Fourier comme outil de description. Si le système n'est pas linéaire, on ne peut pas prédire sa réponse à un stimulus quelconque à partir des réponses aux sinusoïdes qui le composent. On peut donc s'interroger sur l'emploi intensif qui a été fait de *tons purs* comme stimuli, et surtout sur le statut qu'on leur prête de "stimuli élémentaires". En fait, pour l'étude de certains phénomènes (localisation, hauteur tonale), le ton pur est un stimulus particulièrement mal adapté.

Pour résumer, l'Analyse de Fourier est utile pour caractériser les systèmes linéaires et invariants dans le temps, car elle décompose le signal sur une base de vecteurs propres de ces systèmes. Malheureusement le système auditif comprend de nombreuses non-linéarités. Pire, la notion même de linéarité est difficile à lui appliquer car ni l'espace d'arrivée (percepts), ni certains espaces intermédiaires (patterns nerveux) n'ont la structure d'espace vectoriel requise. Ceci limite l'intérêt de l'Analyse de Fourier. Cet outil est néanmoins utile pour caractériser les étapes linéaires ou quasi-linéaires du système auditif, et au delà il peut être servir d'outil descriptif à condition de bien comprendre ses limites.

9.1.1 Un descripteur de la périodicité?

L'utilisation de l'Analyse de Fourier pour caractériser la périodicité procède d'un malentendu.

Un théorème de Fourier dit que tout signal périodique peut se décomposer en une série, la *série de Fourier*, faite de composantes sinusoïdales de fréquences multiples d'une fréquence commune, inverse de la période. Il ne dit *pas* que toutes les composantes de cette série doivent être présentes, et notamment pas celle de fréquence fondamentale. Il ne fournit pas non plus une méthode pour décider qu'un signal est périodique ou non.

En fait, pour caractériser la périodicité d'un signal on utilise plutôt la *transformée de Fourier*. Pour cela il faut (1) calculer la transformée, (2) vérifier qu'elle est constituée d'impulsions de Dirac, et (3) s'assurer que leurs fréquences sont toutes multiples d'une fondamentale commune. La F_0 se calcule alors comme le plus grand commun diviseur de ces fréquences. En d'autres termes on recherche, dans le résultat de la transformée, une forme rappelant la série de Fourier. Le moins qu'on puisse dire est que la démarche

est indirecte (une démarche directe est de vérifier directement la périodicité du signal...). Elle n'est pas non plus sans embûches. Une erreur classique (faite par Ohm et reprise par Helmholtz) est de lier la périodicité du signal à la présence d'une composante à la fondamentale. Une autre est de lier périodicité du signal et périodicité du spectre. La méthode pose aussi des problèmes pratiques d'implémentation : choix d'une fenêtre, résolution limitée, interpolation, identification de pics, etc.

Pour résumer, l'Analyse de Fourier n'est pas un outil particulièrement bien adapté pour caractériser la périodicité. Celle-ci est souvent plus simple à mettre en évidence directement sur le signal, par des méthodes telles que l'AMDF ou l'autocorrélation.

9.2 Unicité et cohérence de la description

Un signal est entièrement caractérisé par sa transformée de Fourier (stricto sensu), et on peut déduire l'un de l'autre sans ambiguïté. Malheureusement, dans la pratique on utilise des transformées de Fourier à *court terme* et autres représentations "spectro-temporelles". Leur emploi est nécessaire pour obtenir une description spectrale qui conserve une certaine localisation temporelle. Se pose alors le problème de l'*unicité* de cette description. Le fait qu'on puisse donner au même objet des descriptions multiples est un facteur de confusion.

Il existe toute une faune de représentations spectro-temporelles. On les voit parfois comparées selon leurs aptitudes à "représenter" tel ou tel signal. En général c'est que la représentation de ce signal est particulièrement simple, ou proche de la façon dont on l'a synthétisé. Le problème est que ce qui est "simple" pour un type de signal ne l'est pas forcément pour un autre. De plus, le même signal peut être défini ou synthétisé de façons multiples. Par exemple s'il est synthétisé par sommation d'un petit nombre de sinusoïdes une représentation par somme de sinusoïdes sera idéale. Mais le *même* signal peut être synthétisé par modulation d'amplitude et/ou fréquence d'un partiel unique. Une représentation spectrale à court terme donnerait une image plus proche de cette "réalité". Quelle représentation choisir ?

En Audition (et particulièrement en Analyse de Scènes Auditives) on a coutume de raisonner comme si un stimulus "contenait" un certain nombre de sinusoïdes. Ces composantes sont éventuellement modulées en fréquence ou amplitude, elles démarrent et terminent à des instants précis, etc., et on attend du système auditif qu'il les "groupe". C'est oublier que le système auditif n'a pas connaissance de cette définition particulière du signal, et n'a donc aucune raison de manipuler ces "composantes". Le système auditif étant

généraliste, il n'a pas de raison d'utiliser la représentation particulière qui a servi pour la synthèse du stimulus.

En résumé, dans la pratique une analyse spectrale se fait à court terme de façon à obtenir une représentation localisée dans le temps. Pour un signal donné, il existe de nombreuses variantes d'analyse selon la méthode ou les paramètres, qui donnent des représentations différentes. Toutes ne sont pas également adaptées à tous les signaux, ce qui donne lieu à des palabres pour décider quelle est la "meilleure". Le manque d'unicité des représentations est un facteur de confusion, surtout lorsque dans un même raisonnement on invoque plusieurs représentations, ou lorsqu'on projette sur le système auditif la représentation particulière qui a servi à la synthèse du signal.

9.3 Conclusion

L'Analyse de Fourier est un outil puissant, particulièrement bien adapté à l'étude de systèmes linéaires et invariants dans le temps. Le système auditif n'est qu'en partie linéaire, ce qui enlève à cet outil une partie de son intérêt. Ni l'activité neuronale ni les percepts n'ont pas une structure d'espace vectoriel, et il ne sert donc à rien de parler de "fonction de transfert" vers ces domaines, sauf dans un sens restreint qu'il convient chaque fois de définir. Le statut de "stimulus élémentaire" conféré au ton pur est discutable.

L'Analyse de Fourier fait partie du bagage intellectuel du scientifique et de l'ingénieur, qui ont tendance à l'employer de façon systématique. La complexité des relations de dualité entre le domaine physique et le domaine spectral peut éventuellement masquer son inadéquation, et donner à tort l'illusion d'un progrès vers la compréhension des phénomènes, alors qu'ils ont simplement été transformés. L'Analyse de Fourier est ainsi souvent utilisée pour caractériser la périodicité, alors qu'elle n'est pas particulièrement bien adaptée à cette tâche.

L'analyse spectrale doit être à court terme pour être utile. La multiplicité des méthodes et paramètres d'analyse fait qu'un signal donné admet des représentations spectro-temporelles multiples. Cela réduit en partie son intérêt comme outil descriptif, puisqu'à chaque description il faut associer une description de la méthode (sans parler de l'effort qu'il faut investir pour comprendre les multiples représentations et leurs propriétés). La multiplicité des représentations est facteur de confusion lorsqu'elles sont juxtaposées sans précaution, ou lorsqu'on attribue un statut injustifié à une représentation particulière (par exemple celle utilisée pour synthétiser le stimulus).

L'Analyse de Fourier est un outil précieux en Audition mais il faut prendre des précautions dans son emploi et dans l'interprétation de ses résultats.

Faute de précautions l’outil peut générer des problèmes artificiels, et on se demande parfois si certains problèmes théoriques en Audition (on songe à la “fondamentale absente”) n’ont pas une telle origine.

Sigles employés

- AAF** - Anterior Auditory Field - *Champ auditif antérieur*
- AC** - Auditory Cortex - *Cortex auditif*
- ACF** - AutoCorrelation Function - *Fonction d'autocorrélation*
- AI** - Primary Auditory Field - *Champ auditif primaire*
- AII** - Secondary Auditory Field - *Champ auditif secondaire*
- aes** - Anterior Ectosylvian Sulcus
- AMDF** - Average Magnitude Difference Function - *Fonction utilisée dans l'estimation de la F_0*
- AMPA** - α -amino-3-hydroxy-5-methyl-4isoxazole propionic acid
- APO** - Anterior Periolivary Nucleus - *Noyau périolivaire antérieur*
- AVCN** - Anteroventral Cochlear Nucleus - *Noyau cochléaire antéroventral*
- BIC** - Brachium of the Inferior Colliculus
- BF** - Best frequency - *Fréquence de réponse optimale*
- BMF** - Best modulation frequency - *Fréquence de modulation optimale*
- BMLD** - Binaural masking level difference - *Taux de démasquage binaural*
- CASA** - Computation Auditory Scene Analysis - *Analyse de scènes auditives computationnelle*
- CCI** - Cellule Ciliée Interne
- CCE** - Cellule Ciliée Externe
- CF** - Characteristic frequency - *Fréquence caractéristique*
- CN** - Cochlear Nucleus - *Noyau cochléaire*
- DCN** - Dorsal cochlear nucleus - *Noyau cochléaire dorsal*
- DLPO** - Dorsolateral periolivary nucleus - *Noyau périolivaire dorsolatéral*
- dMGB** - Dorsal division of the MGB
- DMPO** - Dorsomedial periolivary nucleus - *Noyau périolivaire dorsomédian*
- DNLL** - Dorsal Nucleus of the Lateral Lemniscus - *Noyau dorsal du lemniscus latéral*
- DPO** - Dorsal periolivary nucleus - *Noyau périolivaire dorsal*
- EE** - Excitatory-excitatory - *Excitateur/excitateur (interaction binaurale)*
- EI** - Excitatory-inhibitory - *Excitateur/inhibiteur (interaction binaurale)*
- Ep** - Posterior ectosylvian field
- EPSP/EPSC** - Excitatory post-synaptic potential/current
- F_0 - Fundamental Frequency - *Fréquence fondamentale*
- GABA** - Gamma-aminobutyric acid - *Acide γ -aminobutyrique*
- GBC** - Globular bushy cells - *Cellules globulaires (du CN)*
- IC** - Inferior Colliculus - *Colliculus Inférieur*
- ICA** - Independent Component Analysis - *Analyse en composantes indépendantes*

ICC - Central nucleus of th Inferior Colliculus - *Noyau centrale du colliculus Inférieur*
ICP - Posterior nucleus of th Inferior Colliculus - *Noyau postérieur du colliculus Inférieur*
ICX - External nucleus of th Inferior Colliculus - *Noyau externe du colliculus Inférieur*
IE - Inhibitory-excitatory - *Inhibiteur/excitateur (interaction binaurale)*
IHC - Inner Hair Cell - *Cellule ciliée interne*
INLL - Intermediate nucleus of the Lateral Lemniscus - *Noyau intermédiaire du lemniscus latéral*
IPSP/IPSC - Inhibitory post-synaptic potential/current
ISI Interspike Interval
LL - Lateral lemniscus - *Lemniscus latéral*
LPC - Linear Predictive Coding - *Codage prédictif linéaire*
LNTB - Lateral nucleus of the Trapezoidal body - *Noyau latéral du corps trapézoïde*
LOC - Lateral olivocochlear bundle - *Faisceau croisé olivocochléaire latéral*
LTI - Linear Time Invariant (System) - *Système linéaire et invariant dans le temps*
mEPSP/mEPSC - miniature EPSP/EPSC
mMGB - Medial division of the MGB - *Division médiane du corps genouillé médian*
MGB - Medial Geniculate Body - *Corps genouillé médian*
MOC - Medial olivocochlear bundle - *Faisceau croisé olivocochléaire médian*
MNTB - Medial Nucleus of the Trapezoidal Body - *Noyau médian du corps trapézoïde*
MSO - Medial Superior Olive - *Olive supérieur médiane*
MDT - Missing Data Theory - *Théorie des Données Manquantes*
NCAT - Nucleus of the Central Acoustic Tract - *Noyau du tractus acoustique central*
NMDA - N-methyl-D-aspartate
OCB - Olivocochlear bundle - *Faisceau olivocochléaire*
O_c - Onset chopper
OHC - Outer Hair Cell - *Cellule ciliée externe*
PAF - Posterior Auditory Field - *Champ auditif postérieur*
pes - Posterior Ectosylvian field
PO - Posterior nucleus of the MGB - *Noyau postérieur du corps genouillé médian*
PON - Periolivary nuclei - *Noyaux périolivaires*
PPO - Posterior Periolivary Nucleus - *Noyau périolivaires postérieur*
PVCN - Posteroventral cochlear nucleus - *Noyau cochléaire postéroventral*
pvLNTB - Posterior division of the LNTB - *Division postérieure du noyau latéral du corps trapézoïde*
SBC - Spherical Bushy Cells - *Cellules sphériques (du CN)*
SC - Superior Colliculus - *Colliculus supérieur*
SG - Spiral Ganglion - *Ganglion spiral*
SOC - Superior Olivary Complex - *Complexe olivaire supérieur*

SPO - Superior periolivary nucleus - *Noyau supérieur périolivaire*
SPN - Superior paraolivary nucleus
VCLL - Ventral complex of the Lateral Lemniscus - *Complexe ventral du lemniscus latéral*
VCN - Ventral cochlear nucleus - *Noyau cochléaire ventral*
vMGB - Ventral division of the Medial Geniculate Body - *Division ventrale du corps genouillé médian*
VMPO - Ventromedial periolivary nucleus - *Noyau périolivaire dorsoventral*
VNLL - Ventral Nucleus of the Lateral Lemniscus - *Noyau ventral du lemniscus latéral*
VNTB - Ventral Nucleus of the Trapezoidal Body - *Noyau ventral du noyau trapézoïde médian*
VPAF - Posteroventral Auditory Field - *Champ auditif postéroventral*

Publications

Les articles marqués d'un signe \diamond sont reproduits à la fin du document.

Articles parus ou à paraître dans une revue à comité de lecture :

- [1] \diamond de Cheveigné, A. (2000). "A model of the perceptual asymmetry between peaks and troughs of frequency modulation," *J. Acoust. Soc. Am.*, in press.
- [2] \diamond de Cheveigné, A. (1999). "Waveform interactions and the segregation of concurrent vowels," *J. Acoust. Soc. Am.* 106, 2959-2972.
- [3] \diamond de Cheveigné, A. (1999). "Pitch shifts of mistuned partials : a time-domain model," *J. Acoust. Soc. Am.* 106, 887-897.
- [4] \diamond de Cheveigné, A. (1999). "Vowel-specific effects in concurrent vowel identification," *J. Acoust. Soc. Am.* 106, 327-340.
- [5] \diamond de Cheveigné, A., and Kawahara, H. (1999). "Missing data model of vowel perception," *J. Acoust. Soc. Am.* 105, 3497-3508.
- [6] \diamond de Cheveigné, A., and Kawahara, H. (1999). "Multiple period estimation and pitch perception model," *Speech Communication* 27, 175-185.
- [7] Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A. (1999). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction : Possible role of a repetitive structure in sounds," *Speech Communication* 27, 187-207.
- [8] \diamond de Cheveigné, A. (1998). "Cancellation model of pitch perception," *J. Acoust. Soc. Am.* 103, 1261-1271.
- [9] \diamond de Cheveigné, A. (1997). "Harmonic fusion and pitch shifts of inharmonic partials," *J. Acoust. Soc. Am.* 102, 1083-1087.
- [10] \diamond de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (1997). "Concurrent vowel identification I : Effects of relative level and F0 difference," *J. Acoust. Soc. Am.* 101, 2839-2847.
- [11] \diamond de Cheveigné, A., McAdams, S., and Marin, C. (1997). "Concurrent vowel identification II : Effects of phase, harmonicity and task," *J. Acoust. Soc. Am.* 101, 2848-2856.
- [12] \diamond de Cheveigné, A. (1997). "Concurrent vowel identification III : A neural model of harmonic interference cancellation," *J. Acoust. Soc. Am.* 101, 2857-2865.
- [13] \diamond de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (1995). "Identification of concurrent harmonic and inharmonic vowels : A test of the theory of harmonic cancellation and enhancement," *J. Acoust. Soc. Am.* 97, 3736-3748.
- [14] \diamond de Cheveigné, A. (1993). "Separation of concurrent harmonic sounds : Fundamental frequency estimation and a time-domain cancellation model of auditory processing," *J. Acoust. Soc. Am.* 93, 3271-3290.
- [15] de Cheveigné, A., Abe, M., and Doshita, S. (1985). "The human interface of a speech work-station," *Studia Phonologica* XIX, 18-26.

Chapitres de livre :

- [16] de Cheveigné, A. (2000). "L'analyse de scènes auditives computationnelle," in "La parole, des modèles cognitifs aux machines communicantes - Développement," Edited by J. Mariani, Paris, Hermès, en préparation.

Conférences :

- [17] ◇ de Cheveigné, A. (2000). "The auditory system as a separation machine.", Proc. International Symposium on Hearing, august 2000.
- [18] de Cheveigné, A. (1999). "Time-domain models for segregation, pitch, and multiple pitch perception.", Proc. NIPS workshop on Neural Mechanisms of Music Perception.
- [19] de Cheveigné, A. (1999). "A "missing-data" model of F0-invariance in vowel perception," J. Acoust. Soc. Am. 105, 1187 (abstract).
- [20] de Cheveigné, A. (1999). "Formant bandwidth affects the identification of competing vowels.", Proc. ICPHS, 2093-2096.
- [21] Kawahara, H., Katayose, H., Patterson, R. D., and de Cheveigné, A. (1999). "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity," Proc. Eurospeech.
- [22] de Cheveigné, A. (1998). "Analyse de scènes auditives computationnelle," École d'été - La parole, des modèles cognitifs aux machines communicantes (Grenoble).
- [23] de Cheveigné, A. (1998). "Periodicity and missing feature theory in audition," Proc. CREST Workshop on Stable representations for periodic sounds, Nagoya.
- [24] de Cheveigné, A. (1998). "The auditory system as a separation machine," Proc. ATR workshop on events and auditory temporal structure, 1-7.
- [25] de Cheveigné, A. (1998). "Time domain processing in the auditory system," Proc. ICONIP (International Conference on Neural Information Processing), 1327-1332.
- [26] Kawahara, H., de Cheveigné, A., and Patterson, R. D. (1998). "An instantaneous-frequency-based pitch extraction method for high quality speech transformation : revised TEMPO in the STRAIGHT-suite," Proc. ICSLP, (Dec. 1998).
- [27] Kawahara, H., Katayose, H., Patterson, R. D., and de Cheveigné, A. (1998). "Equilibrium points of frequency-to-instantaneous-frequency mapping and its application to accurate F0 extraction," Proc. CREST workshop on "Stable representations for periodic sounds", Nagoya.
- [28] Kawahara, H., Katayose, H., Patterson, R. D., and de Cheveigné, A. (1998). "Highly accurate F0 extraction using instantaneous frequencies," Proc. ASJ special interest group on hearing, H-98-XX/MA-98-XX.
- [29] de Cheveigné, A., and Kawahara, H. (1997). "Modeling the perception of multiple pitches," Proc. IJCAI workshop on Computational Auditory Scene Analysis, Nagoya, 11-18.
- [30] de Cheveigné, A., and Tsuzaki, M. (1997). "A model of the pitch shifts of mistuned partials," Proc. Acoust. Soc. Japan spring meeting, 415-416.
- [31] Kawahara, H., Masuda, I., and de Cheveigné, A. (1997). "STRAIGHT : A speech analysis-modification-synthesis method and TEMPO : A pitch extraction method," Proc. Spring meeting of the ASJ, 233-234.

- [32] Kawahara, H., Masuda, I., and de Cheveigné, A. (1997). "Restructuring speech representations using STRAIGHT-TEMPO : possible role of a repetitive structure in sounds," Proc. IJCAI workshop on CASA, Nagoya, 103-112.
- [33] Marin, C., and de Cheveigné, A. (1997). "Rôle de la modulation de fréquence dans la séparation de voyelles," Proc. SFA, 527-530.
- [34] de Cheveigné, A. (1996). "A neural cancellation model of F0-guided sound separation," Proc. ESCA Workshop on the Auditory Basis of Speech Perception, Keele, 180-185.
- [35] de Cheveigné, A. (1996). "Ségrégation de voyelles simultanées : effets du niveau relatif et de la différence de F0," Proc. Journées d'Étude de la Parole, 55-58.
- [36] de Cheveigné, A., and Marin, C. (1996). "The segregation of frequency-modulated concurrent harmonic sounds," J. Acoust. Soc. Am. 100, 2718.
- [37] de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (1995). "Sensitive experimental techniques for the study of sound segregation," Proc. ASJ autumn meeting, 373-374.
- [38] de Cheveigné, A. (1994). "Strategies for voice separation based on harmonicity," Proc. ICSLP, Yokohama, 1071-1074.
- [39] de Cheveigné, A., Kawahara, H., Aikawa, K., and Lea, A. (1994). "Speech separation for speech recognition," Journal de Physique IV 4, C5-545-C5-548.
- [40] de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (1994). "Identification de voyelles simultanées harmoniques et inharmoniques," Journal de Physique IV 4, C5-553-C5-556.
- [41] Lea, A., and de Cheveigné, A. (1993). "Effects of an interfering complex tone on a vowel continuum," Proc. Autumn meeting of the ASJ, 195-196.
- [42] de Cheveigné, A. (1991). "A mixed speech F0 estimation algorithm," Proc. ESCA (Eurospeech)-91, Genova, 445-448.
- [43] de Cheveigné, A. (1991). "A model for the discrimination of pure tone pitch," Proc. ICPhS, Aix en Provence, 74-77.
- [44] de Cheveigné, A. (1991). "L'éditeur de signal 'à distance' MapSignal," Proc. ICPhS, 490-493.
- [45] de Cheveigné, A. (1991). "Speech f0 extraction based on Licklider's pitch perception model," Proc. ICPhS, 218-221.
- [46] de Cheveigné, A. (1989). "Pitch and the narrowed autocoincidence histogram," Proc. Proc. ICMPC, Kyoto, 67-70.
- [47] de Cheveigné, A. (1989). "The MapSignal remote speech editor," Proc. Proc. Acoust. Soc. Japan Autumn meeting, 383-384.
- [48] de Cheveigné, A., and Kuwabara, H. (1989). "Narrowed autocoincidence of nerve spike patterns and pure tone pitch," Proc. Proc. Acoust. Soc. Japan Autumn meeting, 383-384.
- [49] de Cheveigné, A. (1986). "A pitch perception model," Proc. Proc. IEEE ICASSP, 897-900.
- [50] de Cheveigné, A. (1985). "Phase sensitivity and pitch shifts : the behavior of three temporal pitch models," Proc. Proc. Acoust. Soc. Japan Autumn meeting, 277-278.
- [51] de Cheveigné, A. (1985). "A nerve fiber discharge model for the study of pitch," Proc. Trans. Acoust. Soc. Japan Committee on Speech/Hearing Research, 279-286.

- [52] de Cheveigné, A. (1984). "A shift and comparison model of pitch perception," Proc. Trans. Acoust. Soc. Japan Committee on Speech Research,, 295-302.
- [53] de Cheveigné, A. (1984). "A signal processing approach via shift and comparison," Proc. Proc. Acoust. Soc. Japan Spring meeting, 5-6.

Rapports techniques ATR et IEICEJ :

- [54] de Cheveigné, A., and Kawahara, H. (1998), "A model of vowel perception based on missing feature theory," ATR-HIP technical report, TR-H-252.
- [55] de Cheveigné, A. (1997), "Ten experiments in concurrent vowel segregation," ATR Human Information Processing Research Labs technical report, TR-H-217.
- [56] Kawahara, H., and de Cheveigné, A. (1997), "F0 extraction method free from errors and its evaluation," IEICEJ (Institute of Electronics, Information and Communication Engineers) technical report, SP-69.
- [57] de Cheveigné, A. (1996), "Speech fundamental frequency estimation," ATR-HIP technical report, TR-H-195.
- [58] Kawahara, H., Masuda, I., and de Cheveigné, A. (1997), "Restructuring speech representations using STRAIGHT-TEMPO : possible role of a repetitive structure in sounds," ATR-Human Information Processing Research Laboratories technical report.
- [59] Kawahara, H., and de Cheveigné, A. (1996), "Error free F0 extraction algorithm and its evaluation," ATR technical report, TR-H-206.
- [60] de Cheveigné, A. (1995), "Experiments in vowel segregation," ATR-HIP technical report, TR-H-154.
- [61] de Cheveigné, A. (1993), "Time-domain comb filtering for speech separation," ATR Human Information Processing Laboratories technical report, TR-H-016.
- [62] de Cheveigné, A. (1990), "Experiments in pitch extraction," ATR Interpreting Telephony Res. Labs technical report technical report, TR-I-0138.
- [63] de Cheveigné, A. (1990), "Auditory nerve fiber spike generation model," ATR Interpreting Telephony Res. Labs. technical report technical report, TR-I-0139.
- [64] de Cheveigné, A. (1990), "The MapSignal remote speech editor," ATR Interpreting Telephony Res. Labs. technical report technical report, TR-I-0137.
- [65] de Cheveigné, A. (1990), "F0 estimation from mixed speech," ATR Auditory and visual perception research labs technical report, TR-A-0097.

Rapports MPEG-7 :

- [66] de Cheveigné, A. (2000), "Core experiment on the scale tree," ISO/IEC JTC1/SC29/WG11, MPEG00/m5910 technical report.
- [67] de Cheveigné, A., and Peeters, G. (2000), "Core set of audio signal descriptors," ISO/IEC JTC1/SC29/WG11, MPEG00/m5885 technical report.
- [68] de Cheveigné, A., Pfeiffer, S., and Casey, M. (2000), "Framework for audio descriptors," ISO/IEC JTC1/SC29/WG11 N3242 technical report.
- [69] de Cheveigné, A. (1999), "Scale tree update," ISO/IEC JTC1/SC29/WG11, MPEG99/m5443 technical report.

- [70] de Cheveigné, A., and Peeters, G. (1999), "A set of scalable audio descriptors," ISO/IEC JTC1/SC29/WG11, MPEG99/m5444 technical report.
- [71] de Cheveigné, A., and Peeters, G. (1999), "Scale tree," ISO/IEC JTC1/SC29/WG11, MPEG99/m5076 technical report.
- [72] de Cheveigné, A., Peeters, G., and Herrera, P. (1999), "Issues on the core experiment for melody description," ISO/IEC JTC1/SC29/WG11, MPEG99/m5199 technical report.
- [73] de Cheveigné, A., and Smith, B. (1999), "A 'Sound Transparency' descriptor," ISO/IEC JTC1/SC29/WG11, MPEG99/m5198 technical report.

Thèse 3e cycle :

- [74] de Cheveigné, A. (1982), "Mémoire digitale de transitoires : contribution à la phonétique instrumentale," Paris 7 unpublished doctoral dissertation.

Bibliographie

- Ayres, T. J., and Clack, T. D. (1984). "Detection of interaural phase shift between a subaudible and an audible tone," *J. Acoust. Soc. Am.* 76, 411-413.
- Adams, J. C. (1996). "Neural circuits in the human auditory brainstem.", *Proc. Workshop on the auditory basis of speech perception*, Keele, 39-44.
- Adams, J. C. (1997). "Projections from octopus cells of the posteroventral cochlear nucleus to the ventral nucleus of the lateral lemniscus in cat and human," *Aud. Neurosci.* 3, 335-350.
- Assmann, P. F., and Summerfield, Q. (1988). "Pitch-pulse asynchrony and the perceptual segregation of competing voices.", *Proc. Speech 88 Conference (7th FASE)*, Edinburgh, 531-538.
- Assmann, P. F., and Summerfield, Q. (1989). "Modeling the perception of concurrent vowels : Vowels with the same fundamental frequency," *J. Acoust. Soc. AM* 85, 327-338.
- Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels : Vowels with different fundamental frequencies," *J. Acoust. Soc. AM* 88, 680-697.
- Assmann, P. F., and Summerfield, Q. (1994). "The contribution of waveform interactions to the perception of concurrent vowels," *J. Acoust. Soc. AM* 95, 471-484.
- Assmann, P. F., and Paschall, D. D. (1998). "Pitches of concurrent vowels," *J. Acoust. Soc. AM* 103, 1150-1160.
- Assmann, P. F. (1999). "Fundamental Frequency and the Intelligibility of Competing Voices.", *Proc. ICPhS*, .
- Bachem, A. (1950). "Tone height and tone chroma as two different pitch qualities," *Acta psychologica* 7, 80-88.
- Banks, M. I., and Smith, P. H. (1992). "Intracellular recordings from neurobiotin-labeled cells in brain slices of the rat medial nucleus of the trapezoid body," *J. Neurosc.* 12, 2819-2837.
- Batra, R., and Fitzpatrick, D. C. (1997). "Neurons sensitive to interaural temporal disparities in the medial part of the ventral nucleus of the lateral lemniscus," *J. Neurophysiol.* 78, 511-515.
- Batra, R., Kuwada, S., and Fitzpatrick, D. C. (1997). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. II. Coincidence detection," *J. Neurophysiol.* 78, 1237-1247.
- Batra, R., Kuwada, S., and Fitzpatrick, D. C. (1997). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. I. Heterogeneity of responses," *J. Neurophysiol.* 78, 1222-1236.
- Bregman, A. S. (1977). "Perception and behavior as compositions of ideals," *Cognitive Psychology* 9, 250-292.
- Bregman, A. S. (1990). "Auditory scene analysis," Cambridge, Mass., MIT Press.
- Brew, H. M., and Forsythe, I. D. (1995). "Two voltage-dependent K⁺ conductances with complementary functions in postsynaptic integration at a central auditory synapse," *J. Neurosci.* 15, 8011-8022.

- Broadbent, D. E., and Ladefoged, P. (1957). "On the fusion of sounds reaching different sense organs," *J. Acoust. Soc. AM* 29, 708-710.
- Brokx, J. P. L., and Nootboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," *Journal of Phonetics* 10, 23-36.
- Brown, J. C., and Puckette, M. S. (1989). "Calculation of a "narrowed" autocorrelation function," *J. Acoust. Soc. AM* 85, 1595-1601.
- Brown, G. J. (1992), "Computational auditory scene analysis : a representational approach," Sheffield, Department of Computer Science unpublished doctoral dissertation.
- Brown, M. C., Kujawa, S. G., and Duca, M. L. (1998). "Single olivocochlear neurons in the guinea pig. I. Binaural facilitation of responses to high-level noise," *J. Neurophysiol.* 79, 3077-3087.
- Brugge, J. F., Anderson, D. J., Hind, J. E., and Rose, J. E. (1969). "Time structures of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds," *J. Neurophysiol.* 32, 386-401.
- Brugge, J. F. (1992). "Central auditory processing," in "The mammalian auditory pathway : neurophysiology," Edited by A. N. Popper and R. R. Fay, New York, Springer-Verlag, 1-33.
- Brunstrom, J. M., and Roberts, B. (1998). "Profiling the perceptual suppression of partials in periodic complex tones : further evidence for a harmonic template," *J. Acoust. Soc. AM* 104, 3511-3519.
- Burger, R. M., and Pollak, G. D. (1998). "Analysis of the role of inhibition in shaping responses to sinusoidally amplitude-modulated signals in the inferior colliculus," *J. Neurophysiol.* 80, 1686-1701.
- Buser, P., and Imbert, M. (1982). "Neurophysiologie fonctionnelle II : Psychophysiologie fonctionnelle," Paris, Hermann.
- Buser, P., and Imbert, M. (1987). "Neurophysiologie fonctionnelle III : Audition," Paris, Hermann.
- Cant, N. B. (1992). "The cochlear nucleus : neuronal types and their synaptic organization," in "The mammalian auditory pathway," Edited by D. B. Webster, A. N. Popper and R. R. Fay, New York, Springer Verlag, 66-116.
- Cariani, P. A., and Delgutte, B. (1996). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," *J. Neurophysiol.* 76, 1698-1716.
- Cariani, P. A., and Delgutte, B. (1996). "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate-pitch and the dominance region for pitch," *J. Neurophysiol.* 76, 1717-1734.
- Carlyon, R. P., and Shackleton, T. M. (1994). "Comparing the fundamental frequencies of resolved and unresolved harmonics : evidence for two pitch mechanisms?," *J. Acoust. Soc. AM* 95, 3541-3554.
- Carlyon, R. P. (1998). "Comments on "A unitary model of pitch perception" [*J. Acoust. Soc. Am.* 102, 1811-1820 (1997)]," *J. Acoust. Soc. AM* 104, 1118-1121.
- Carney, H., and Yin, T. C. T. (1988). "Temporal coding of resonances by low-frequency auditory nerve fibers : single fiber responses and a population model," *J. Neurophysiol.* 60, 1653-1677.

- Carney, L. H., and Yin, T. C. T. (1989). "Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks : excitatory and inhibitory components," *J. Neurophysiol.* 62, 144-161.
- Carney, L. (1990). "Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents," *J. Neurophysiol.* 64, 437-456.
- Carney, L. H. (1993). "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. AM* 93, 401-417.
- Carr, C. E., 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.
- Casseday, H., and Covey, E. (1995). "Mechanisms for analysis of auditory temporal patterns in the brainstem of echolocating bats," in "Neural representation of temporal patterns," Edited by E. Covey, H. L. Hawkins and R. F. Port, New York, Plenum, 25-51.
- Casseday, J. H., Covey, E., and Grothe, B. (1997). "Neural selectivity and tuning for sinusoidal frequency modulations in the inferior colliculus of the big brown bat, *Eptesicus fuscus*," *J. Neurophysiol.* 77, 1595-1604.
- Chan, J. C. K., Yin, T. C. T., and Musicant, A. D. (1987). "Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. II. Responses to band-pass filtered noises," *J. Neurophysiol.* 58, 543-561.
- Cherry, E. C. (1953). "Some experiments on the recognition of speech with one, and with two ears," *J. Acoust. Soc. AM* 25, 975-979.
- Chuhma, N., and Ohmori, H. (1998). "Postnatal development of phase-locked high-fidelity synaptic transmission in the medial nucleus of the trapezoidal body of the rat," *J. Neuroscience* 18, 512-520.
- Clarey, J. C., Barone, P., and Imig, T. J. (1992). "Physiology of the thalamus and cortex," in "The mammalian auditory pathway : neurophysiology," Edited by A. N. Popper and R. R. Fay, New York, Springer Verlag, 232-334.
- Clément, S. (1996), "Modélisation de la perception auditive des modulations de fréquence," Université Bordeaux 2, FRANCE unpublished doctoral dissertation.
- Colburn, H. S., and Durlach, N. I. (1978). "Models of binaural interaction," in "Handbook of perception," Edited by E. C. Carterette and M. P. Friedman, New York, Academic Press, IV, 467-518.
- Colburn, H. S., Han, Y., and Culotta, C. P. (1990). "Coincidence model of MSO responses," *Hearing Research* 49, 335-346.
- Colburn, H. S. (1995). "Computational models of binaural processing," in "Auditory Computation," Edited by H. Hawkins, T. McMullin, A. N. Popper and R. R. Fay, New York, Springer-Verlag, 332-400.
- Cooke, M. P. (1991), "Modeling auditory processing and organisation," Sheffield, Department of Computer Science unpublished doctoral dissertation.
- Cooke, M., and Brown, G. (1994). "Separating simultaneous sound sources : issues, challenges and models," in "Fundamentals of speech synthesis and speech recognition," Edited by E. Keller, John Wiley & Sons, 295-312.
- Cooke, M., Morris, A., and Green, P. (1996). "Recognising occluded speech.", *Proc. Workshop on the Auditory basis of Speech Perception, Keele*, 297-300.

- Cooke, M., Morris, A., and Green, P. (1997). "Missing data techniques for robust speech recognition.", Proc. ICASSP, 863-866.
- Cooke, M., and Ellis, D. P. W. (2000). "The auditory organization of speech and other sources in listeners and computational models," Speech Comm. in press,
- Cooke, M., Green, P., Josifovski, L., and Vizinho, A. (2000). "Robust automatic speech recognition with missing and unreliable acoustic data," Speech Communication (submitted)
- Cooke, M., and Green, P. (2000 (in press)). "Auditory organization and speech perception," in "Listening to speech : an auditory perspective," Edited by S. Greenberg and W. Ainsworth, Oxford, Oxford University Press, .
- Covey, E., Vater, M., and Casseday, J. H. (1991). "Binaural properties of single units in the superior olivary complex of the mustache bat," J. Neurophysiol. 66, 1080-1094.
- Covey, E., and Casseday, J. H. (1999). "Timing in the auditory system of the bat," Ann. Rev. Physiol. 61, 457-476.
- Culling, J. F. (1990). "Exploring the conditions for the perceptual segregation of concurrent voices using F0 differences," Proc. of the Institute of Acoustics 12, 559-566.
- Culling, J. F., and Darwin, C. J. (1993). "Perceptual separation of simultaneous vowels : Within and across-formant grouping by F0," J. Acoust. Soc. AM 93, 3454-3467.
- Culling, J. F., and Darwin, C. J. (1994). "Perceptual and computational separation of simultaneous vowels : Cues arising from low frequency beating," J. Acoust. Soc. AM 95, 1559-1569.
- Culling, J. F., Summerfield, Q., and Marshall, D. H. (1994). "Effects of simulated reverberation on the use of binaural cues and fundamental frequency differences for separating concurrent vowels," Speech Comm. 14, 71-95.
- Culling, J. F., and Summerfield, Q. (1995). "The role of frequency modulation in the perceptual segregation of concurrent vowels," J. Acoust. Soc. AM 98, 837-846.
- Culling, J. F., and Summerfield, Q. (1995). "Perceptual segregation of concurrent speech sounds : absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. AM 98, 785-797.
- Culling, J. F. (1996). "Signal processing software for teaching and research in psycholinguistics under UNIX and X-windows," Behavior Research Methods, Instruments, and Computers 28, 376-382.
- Culling, J. F., and Summerfield, Q. (1997). "Measurement of the binaural temporal window using a detection task," J. Acoust. Soc. AM in preparation,
- Culling, J. F., Marshall, D., and Summerfield, Q. (1998). "Dichotic pitches as illusions of binaural unmasking II : the Fourcin pitch and the Dichotic Repetition Pitch," J. Acoust. Soc. AM 103, 3525-3539.
- Culling, J. F., Summerfield, Q., and Marshall, D. H. (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, C. (1872). "The expression of emotions in man and animals," New York, Oxford University Press (1998 edition).
- Darwin, C. J. (1981). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," QJEP A 33, 185-207.

- Darwin, C. J., and Sutherland, N. S. (1984). "Grouping frequency components of vowels : when is a harmonic not a harmonic," *QJEP* 36A, 193-208.
- Darwin, C. J., and Gardner, R. B. (1985). "Which harmonics contribute to the estimation of first formant frequency?," *Speech Comm.* 4, 231-235.
- Darwin, C. J., and Gardner, R. B. (1986). "Mistuning of a harmonic of a vowel : Grouping and phase effects on vowel quality," *J. Acoust. Soc. AM* 79, 838-845.
- Darwin, C. J., and Gardner, R. B. (1987). "Perceptual separation of speech from concurrent sounds," in "The psychophysics of speech perception," Edited by M. E. H. Schouten, Dordrecht, Martinus Nijhoff, 112-124.
- Darwin, C. J., McKeown, J. D., and Kirby, D. (1989). "Perceptual compensation for transmission channel and speaker effects on vowel quality," *Speech Comm.* 8, 221-234.
- Darwin, C. J. (1992). "Listening to two things at once," in "The auditory processing of speech," Edited by M. E. H. Schouten, Berlin, Mouton de Gruyter, 133-147.
- Darwin, C. J., Buffa, A., Williams, D., and Ciocca, W. (1992). "Pitch of dichotic complex tones with a mistuned frequency component," in "Auditory physiology and perception," Edited by L. D. Y. Cazals, K. Horner, Oxford, Pergamon press, 223-229.
- Darwin, C. J., and Ciocca, V. (1992). "Grouping in pitch perception : effects of onset asynchrony and ear of presentation of a mistuned component," *J. Acoust. Soc. AM* 91, 3381-3390.
- Darwin, C. J., and Sandell, G. J. (1995). "Absence of effect of coherent frequency modulation on grouping a mistuned harmonic with a vowel," *J. Acoust. Soc. AM* 97, 3135-3138.
- Darwin, C. J., and Hukin, R. W. (1999). "Auditory objects of attention : the role of interaural time differences," *Journal of Experimental Psychology : Human perception and performance* 25, 1-13.
- de Boer, E. (1976). "On the "residue" and auditory pitch perception," in "Handbook of sensory physiology," Edited by W. D. Keidel and W. D. Neff, Berlin, Springer-Verlag, 479-583.
- de Boer, E. (1977). "Pitch theories unified," in "Psychophysics and physiology of hearing," Edited by E. F. Evans and J. P. Wilson, London, Academic, 323-334.
- Dehmel, S., Doerrscheidt, G. J., and Ruebsamen, R. (1999). "Electrophysiological characterization of neurons in the superior paraolivary nucleus of the gerbil (*Meriones unguiculatus*)," *ARO abstracts #279*.
- Delgutte, B. (1984). "Speech coding in the auditory nerve : II. Processing schemes for vowel-like sounds," *J. Acoust. Soc. AM* 75, 879-886.
- Delgutte, B. (1987). "Peripheral auditory processing of speech information : implications from a physiological study of intensity discrimination," in "The psychophysics of speech perception," Edited by M. E. H. Schouten, Dordrecht, Martinus Nijhoff, 333-353.
- Delgutte, B. (1996). "Physiological models for basic auditory percepts," in "Auditory computation," Edited by H. L. Hawkins, T. A. McMullen, A. N. Popper and R. R. Fay, New York, Springer-Verlag, 157-220.
- Demany, L. (1986). "La perception de la hauteur tonale," Paris 6 - Pierre et Marie Curie unpublished doctoral dissertation.

- Demany, L., and Armand, F. (1984). "The perceptual reality of tone chroma in early infancy," *J. Acoust. Soc. Am.* 76, 57-66.
- Demany, L. (1988). "Perception de la hauteur tonale," in "Psychoacoustique et perception auditive," Edited by M. C. Botte, G. Canévet, L. Demany and C. Sorin, Paris, Inserm, 43-81.
- Demany, L., and Semal, C. (1988), "Identification perceptive de voyelles isolées : l'effet d'une modulation de la fréquence fondamentale," Convention CNET technical report, .
- Demany, L., and Semal, C. (1989). "Detection thresholds for sinusoidal frequency modulation," *J. Acoust. Soc. AM* 85, 1295-1301.
- Demany, L., and Semal, C. (1990). "The effect of vibrato on the recognition of masked vowels," *Percept. & Psychophys.* 48, 436-444.
- Demany, L., and McAnally, K. I. (1994). "The perception of frequency peaks and troughs in wide frequency modulations," *J. Acoust. Soc. AM* 96, 706-715.
- Demany, L., and Clément, S. (1995). "The perception of frequency peaks and troughs in wide frequency modulations. II Effects of frequency register, stimulus uncertainty, and intensity," *J. Acoust. Soc. AM* 97, 2454-2459.
- Demany, L., and Clément, S. (1995). "The perception of frequency peaks and troughs in wide frequency modulations. III. Complex carriers," *J. Acoust. Soc. AM* 98, 2515-2523.
- Demany, L. (1997). "Asymétrie pics/creux : effet de la durée," .
- Demany, L., and Clément, S. (1997). "The perception of frequency peaks and troughs in wide frequency modulations. IV. Effects of modulation waveform," *J. Acoust. Soc. Am.* 102, 2935-2944.
- Demany, L., and Clément, S. (1998). "The perceptual asymmetry of frequency modulation," in "Psychophysical and Physiological Advances in Hearing," Edited by A. R. Palmer, A. Rees, A. Q. Summerfield and R. Meddis., London, Whurr, .
- de Ribaupierre, F. (1997). "Acoustical information processing in the auditory thalamus and cerebral cortex," in "The central auditory system," Edited by G. Ehret and R. Romand, New York, 1997, 317-397.
- Duifhuis, H., Willems, L. F., and Sluyter, R. J. (1982). "Measurement of pitch in speech : an implementation of Goldstein's theory of pitch perception," *J. Acoust. Soc. AM* 71, 1568-1580.
- Duquesnoy, A. J. (1983). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," *J. Acoust. Soc. Am.* 74, 739-743.
- Durlach, N. I. (1963). "Equalization and cancellation theory of binaural masking-level differences," *J. Acoust. Soc. AM* 35, 1206-1218.
- Du Vernay, J. G. (1683). "Traité de l'organe de l'ouïe, contenant la structure, les usages et les maladies de toutes les parties de l'oreille," Paris, .
- Eco, U. (1995). "Come si fa una tesi di laurea (Japanese translation)," Tokyo, Jiritsu Shobokan.
- Ehret, G. (1997). "The auditory midbrain, a "shunting yard" of acoustical information processing," in "The central auditory system," Edited by G. Ehret and R. Romand, New York, Oxford University Press, 259-316.

- Ehrlich, D., Casseday, J. H., and Covey, E. (1997). "Neural tuning to sound duration in the inferior colliculus of the big brown bat, *Eptesicus fuscus*," *J. Neurophysiol.* 77, 1360-2372.
- Ellis, D. (1996), "Prediction-driven computational auditory scene analysis," MIT unpublished doctoral dissertation.
- Evans, E. F. (1974). "Auditory frequency selectivity and the auditory nerve," in "Facts and models in hearing," Edited by E. Zwicker and E. Terhardt, Berlin, Springer-Verlag, 118-129.
- Evans, E. F. (1977). "Frequency selectivity at high signal levels of single units in cochlear nerve and cochlear nucleus," in "Psychophysics and physiology of hearing," Edited by E. F. Evans and J. P. Wilson, London, Academic Press, 185-196.
- Evans, E. F. (1978). "Place and time coding of frequency in the peripheral auditory system : Some physiological pros and cons," *Audiology* 17, 369-420.
- Evans, E. F. (1982). "Representation of complex sounds at cochlear nerve and cochlear nucleus levels," in "The representation of speech in the peripheral auditory system.," Edited by R. Carlson and B. Granström, Amsterdam, Elsevier, 27-42.
- Evans, E. F. (1983). "Pitch and cochlear nerve fibre temporal discharge patterns," in "Hearing-Physiological bases and psychophysics," Edited by R. Klinke and R. Hartmann, Berlin, Springer-Verlag, 140-146.
- Evans, E. F. (1986). "Pitch and cochlear nerve fiber discharge patterns," in "Auditory frequency selectivity," Edited by B. C. J. Moore and R. D. Patterson, Plenum Press, 253-264.
- Finlayson, P. G., and Caspary, D. M. (1991). "Low-frequency neurons in the lateral superior olive exhibit phase-sensitive binaural inhibition," *J. Neurophysiol.* 65, 598-605.
- Fonseca, R. C., and Hallows, J. L. (1998). "Localization of Kv channels in the auditory system," ARO abstract #849.
- Fourcin, A. J. (1970). "Central pitch and auditory lateralization," in "Frequency analysis and periodicity detection in hearing," Edited by R. Plomp and G. F. Smoorenberg, Sijthoff, 319-328.
- Goldberg, J. M., and Brown, P. B. (1969). "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli : some physiological mechanisms of sound localization," *J. Neurophysiol.* 22, 613-636.
- Goldstein, J. L. (1973). "An optimum processor theory for the central formation of the pitch of complex tones," *J. Acoust. Soc. AM* 54, 1496-1516.
- Goldstein, J. L., and Sruлович, P. (1977). "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," in "Psychophysics and physiology of hearing," Edited by E. F. Evans and J. P. Wilson, London, Academic Press, 337-347.
- Goldstein, J. L., Gerson, A., Sruлович, P., and Furst, M. (1978). "Verification of the optimal probabilistic basis of aural processing in pitch of complex tones," *J. Acoust. Soc. AM* 63, 486-497.
- Gould, S. J. (1985). "A most ingenious paradox," in "The flamingo's smile," Edited by S. J. Gould, London, Penguin, 78-95.

- Gray, P. R. (1967). "Conditional probability analyses of the spike probability of single neurons," *Biophysical J.* 7, 759-777.
- Greenberg (1997). "Understanding speech understanding : towards a unified theory of speech perception.", *Proc. ESCA Workshop on the auditory basis of speech perception*, Keele, 1-8.
- Greenwood, D. D. (1990). "A cochlear frequency-position function for several species - 29 years later," *J. Acoust. Soc. Am.* 87, 2592-2605.
- 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.
- Grothe, B., Park, T., and Schuller, G. (1997). "Medial superior olive in the free-tailed bat : response to pure tones and amplitude-modulated tones," *J. Neurophysiol.* 77, 1553-1565.
- Grothe, B., and Park, T. J. (1998). "Sensitivity to interaural time differences in the medial superior olive of a small mammal, the mexican free-tailed bat," *J. Neurosci.* 15, 6608-6622.
- Gu, X., and Wright, B. A., Green, D.M. (1995). "Failure to hear binaural beats below threshold," *J. Acoust. Soc. AM* 97, 701-703.
- Guiguère, C., and Woodland, P. C. (1993). "A wave digital filter model of the entire auditory periphery.," *Proc. IEEE-ICASSP*, 708-711.
- Guinan, J. J. (1996). "Physiology of olivocochlear efferents," in "The cochlea," Edited by P. Dallos, A. N. Popper and R. R. Fay, New York, Springer Verlag, 435-502.
- Hafer, E. R., and Trahiotis, C. (1997). "Functions of the binaural system," in "Encyclopedia of Acoustics," Edited by M. J. Crocker, John Wiley and sons, 1461-1479.
- Hartmann, W. M., McAdams, S., and Smith, B. K. (1986), "Matching the pitch of a mistuned harmonic in a complex tone," *IRCAM technical report*, .
- Hartmann, W. M. (1988). "Pitch perception and the segregation and integration of auditory entities," in "Auditory function - neurological bases of hearing," Edited by G. M. Edelman, W. E. Gall and W. M. Cowan, New York, Wiley, 623-645.
- Hartmann, W. M., McAdams, S., and Smith, B. K. (1990). "Hearing a mistuned harmonic in an otherwise periodic complex tone," *J. Acoust. Soc. AM* 88, 1712-1724.
- Hartmann, W. H. M. (1991). "On the integration and segregation of musical tones.", *Proc. International workshop on recent trends in speech, music and allied signal processing (IWSMSP-91)*, New Delhi, India, 1-10.
- Hartmann, W. M. (1992). "Pitch and the segregation and integration of tones.", *Proc. International conference on music perception and cognition*, 1-8.
- Hartmann, W. M. (1993). "On the origin of the enlarged melodic octave," *J. Acoust. Soc. Am.* 93, 3400-3409.
- Hartmann, W. H. (1995). "The physical description of signals," in "Hearing," Edited by B. C. J. Moore, San Diego, Academic Press, 1-40.
- Hartmann, W. M. (1996). "Pitch, periodicity, and auditory organization," *J. Acoust. Soc. Am.* 100, 3491-3502.
- Hartmann, W. M., and Doty, S. L. (1996). "On the pitches of the components of a complex tone," *J. Acoust. Soc. Am.* 99, 567-578.

- Hartmann, W. M. (1997). "Signals, sound and sensation," Woodbury, N.Y., AIP.
- Hebb, D. O. (1949). "The organization of behavior," New York, Wiley.
- Heffner, R. S., and Heffner, H. E. (1992). "Evolution of sound localization in mammals," in "The evolutionary biology of hearing," Edited by D. B. Webster, R. R. Fay and A. N. Popper, New York, Springer-Verlag, 691-715.
- Helfert, R. H., and Aschoff, A. (1997). "Superior olivary complex and nuclei of the lateral lemniscus," in "The central auditory system," Edited by G. Ehret and R. Romand, New York, Oxford University Press, 193-258.
- Helmholtz, H. v. (1877). "On the sensations of tone (English translation A.J. Ellis, 1954)," New York, Dover.
- Hermansky, H., Hanson, B., and Wakita, H. (1985). "Low-dimensional representation of vowels based on all-pole modeling in the psychophysical domain," *Speech Comm* 4, 181-187.
- Hermansky, H., and Morgan, N. (1994). "RASTA processing of speech," *IEEE trans Speech and Audio Process.* 2, 578-589.
- Hermansky, H. (1998). "Should recognizers have ears?," *Speech Comm.* 25, 3-27.
- Hess, W. (1983). "Pitch determination of speech signals," Berlin, Springer-Verlag.
- Hind, J. E., Anderson, D. J., Brugge, J. F., and Rose, J. E. (1967). "Coding of information pertaining to paired low-frequency tones in single auditory nerve fibers of the squirrel monkey," *J. Neurophysiol.* 30, 794-816.
- Hirahara, T., and Kato, H. (1992). "The effect of F0 on vowel identification," in "Speech perception, production and linguistic structure," Edited by Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka, Tokyo, Ohmsha, 89-112.
- Hirahara, T. (1993). "On the role of relative harmonics level around the F1 of high vowel identification.", *Proc. ARO abstracts (ISSN 0742-3152)*, abstract #258, 65.
- Hirahara, T., Cariani, P., and Delgutte, B. (1996). "Representation of low-frequency vowel formants in the auditory nerve.", *Proc. ESCA Workshop on the auditory basis of speech perception*, 83-86.
- Holdsworth, J., Nimmo-Smith, I., Patterson, R. D., and Rice, P. (1988), "Implementing a GammaTone filter bank," *MRC Applied Psychology Unit technical report, SVOS final report, annex C.*
- Horst, J. W., Javel, E., and Farley, G. R. (1986). "Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms," *J. Acoust. Soc. Am.* 79, 398-416.
- Horst, J. W., Wit, H. P., and Segenhout, H. M. (1988). "Single fiber responses to complex sounds in the pigeon," in "Basic issues in hearing," Edited by H. Duifhuis, J. W. Horst and H. P. Wit, London, Academic Press, 235-243.
- Horst, J. W., Javel, E., and Farley, G. R. (1992). "Coding of fundamental frequency in auditory nerve fibers : effects of signal level and phase spectrum," in "The auditory processing of speech - From sounds to words," Edited by M. E. H. Schouten, Berlin, Mouton de Gruyter, 29-36.
- Houtsma, A. J. M. (1995). "Pitch perception," in "Hearing," Edited by B. C. J. Moore, London, Academic Press, 267-295.

- Huggins, W. H. (1953). "A theory of hearing," in "Communication theory," Edited by W. Jackson, London, Butterworth, 363-380.
- Hukin, R. W., and Darwin, C. J. (1995). "Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel," *J. Acoust. Soc. Am.* 98, 1380-1387.
- Hukin, R. W., and Darwin, C. J. (1995). "Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification," *Perception & Psychophysics* 57, 191-196.
- Irino, T. (1997). "Time-domain, level dependent auditory filter : the gammachirp," *J. Acoust. Soc. Am.* 101, 412-419.
- Irvine, D. R. F. (1992). "Physiology of the auditory brainstem," in "The mammalian auditory pathway : neurophysiology," Edited by A. N. Popper and R. R. Fay, New York, Springer Verlag, 153-231.
- Javel, J. B., Mott, J. B., Rush, N. L., and Smith, D. W. (1988). "Frequency discrimination : evaluation of rate and temporal codes," in "Basic issues in hearing," Edited by H. Duifhuis, J. W. Horst and H. P. Wit, London, Academic Press, 224-234.
- Jeffress, L. A. (1948). "A place theory of sound localization," *J. Comp. Physiol. Psychol.* 41, 35-39.
- Johannesma, P. I. M. (1972). "The pre-response stimulus ensemble of neurons in the cochlear nucleus.", *Proc. Symposium on hearing theory, IPO Eindhoven, .*
- Itoh, K., and Kikkawa, S. (1982). "Temporal factors in dominancy for sound image space - A neural model of binaural interaction," *Ann. Bull. RILP* 16, 125-142.
- Itoh, K. (1983). "A neuro-synaptic model of the bilateral auditory nervous system - image lateralization and binaural unmasking," *Ann. Bull. RILP* 17, 129-146.
- Itoh, K. (1984). "A neuro-synaptic model of the masking and unmasking process in the bilateral auditory system," *Ann. Bull. RILP* 18, 91-105.
- Johnson, D. H., and Kiang, N. Y.-S. (1976). "Analysis of discharges recorded simultaneously from pairs of auditory nerve fibers," *Biophysical Journal* 16, 719-734.
- Johnson, D. H. (1980). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," *J. Acoust. Soc. AM* 68, 1115-1122.
- Johnson, D. H., and Swami, A. (1983). "The transmission of signals by auditory-nerve fiber discharge patterns," *J. Acoust. Soc. AM* 74, 493-501.
- Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. T. (1994). "Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency," *J. Neurophysiol.* 71, 1022-1036.
- Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. T. (1994). "Enhancement of neural synchronization in the anteroventral cochlear nucleus. II. Responses in the tuning curve," *J. Neurophysiol.* 71, 1037-1051.
- Joris, P. X. (1995). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," *J. Neurophysiol.* 73, 1043-1062.
- Joris, P. X. (1996). "Envelope coding in the Lateral Superior Olive. II. Characteristic delays and comparison with responses in the Medial Superior Olive.," *J. Neurophysiol.* 76, 2137-2156.

- Joris, P. X., and Smith, P. H. (1998). "Temporal and binaural properties in dorsal cochlear nucleus and its output tract," *J. Neuroscience* 18, 10157-10170.
- Joris, P. X., Smith, P. H., and Yin, T. C. T. (1998). "Coincidence detection in the auditory system : 50 years after Jeffress," *Neuron* 21, 1235-1238.
- Joris, P. X., and Yin, T. C. T. (1998). "Envelope coding in the lateral superior olive. III. Comparison with afferent pathways," *J. Neurophysiol.* 79, 253-269.
- Joseph, A. W., and Hyson, R. L. (1993). "Coincidence detection by binaural neurons in the chick brain stem," *J. Neurophysiol.* 69, 1197-1211.
- Kaernbach, C., and Demany, L. (1998). "Psychophysical evidence against the autocorrelation theory of pitch perception," *J. Acoust. Soc. Am.* 104, 2298-2306.
- Kahle, W., Leonhardt, H., and Platzer, W. (1979). "Anatomie. III. Système nerveux et organes des sens," Paris, Flammarion.
- Kanadera, N., Hermansky, H., and Arai, T. (1998). "On properties of the modulation spectrum for robust automatic speech recognition.", *Proc. IEEE-ICASSP*, 613-616.
- Kashino, M., and Hirahara, T. (1995). "How many concurrent talkers can we hear out?," *Proc. ASJ Automn meeting (in Japanese)*, 467-468.
- Kashino, M., and Nishida, S. (1996). "Auditory localization aftereffects," *J. Acoust. Soc. Am.* 99, 2596.
- Kashino, M., and Nishida, S. (1997). "Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect," *J. Acoust. Soc. Am.*, submitted
- Kawahara, H. (1997). "Speech representation and transformation using adaptive interpolation of weighted spectrum : vocoder revisited.", *Proc. ICASSP*, 1303-1306.
- Kelly, J. B., and Kidd, S. A. (2000). "NMDA and AMPA receptors in the dorsal nucleus of the lateral lemniscus shape binaural responses in rat inferior colliculus," *J. Neurophysiol.* 83, 1403-1414.
- Kemp, D. T. (1978). "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. AM* 64, 1386-1391.
- Kemp, D. T., and Souter, M. (1988). "The dynamics of cochlear perturbations following brief acoustic and efferent stimulation - otoacoustic and CM data," in "Basic issues in hearing," Edited by H. Duifhuis, J. W. Horst and H. P. Wit, London, Academic Press, 116-123.
- Kiang, N. Y.-S. (1965), "Discharge patterns of single fibers in the cat's auditory nerve," MIT technical report, MIT Research Monograph 35.
- Kidd, S., and Kelly, J. B. (1996). "Contribution of the dorsal nucleus of the lateral lemniscus to binaural responses in the inferior colliculus of the rat : interaural time delays," *J. Neurosci.* 15, 7390-7397.
- Kim, D. O., Rhode, W. S., and Greenberg, S. R. (1986). "Responses of cochlear nucleus neurons to speech signals : neural encoding of pitch, intensity and other parameters," in "Auditory frequency selectivity," Edited by B. C. J. Moore and R. D. Patterson, New York, Plenum Press, 281-282.
- Kim, D. O., and Leonard, G. (1988). "Pitch-period following response of cat cochlear nucleus neurons to speech sounds," in "Basic issues in hearing," Edited by H. Duifhuis, J. W. Horst and H. P. Wit, London, Academic Press, 252-260.

- Klatt, D. H. (1982). "Speech processing strategies based on auditory models," in "The representation of speech in the peripheral auditory system," Edited by R. Carlson and B. Granström, Amsterdam, Elsevier, 181-196.
- Klinke, R. (1986), commentaire à Zwicker, 1986 (p. 57)
- Koch, U., and Grothe, B. (2000). "Interdependence of spatial and temporal coding in the auditory midbrain," *J. Neurophysiol.* 83, 2300-2314.
- Kohlrausch, A. (1988). "Masking patterns of harmonic complex tone maskers and the role of the inner ear transfer function," in "Basic issues in hearing," Edited by H. Duifhuis, J. W. Horst and H. P. Wit, London, Academic Press, 339-350.
- Kohlrausch, A., and Sander, A. (1995). "Phase effects in masking related to dispersion in the inner ear. II Masking period patterns of short targets," *J. Acoust. Soc. Am.* 97, 1817-1829.
- Konishi, M., Takahashi, T. T., Wagner, H., Sullivan, W. E., and Carr, C. E. (1988). "Neurophysiological and anatomical substrates of sound localization in the owl," in "Auditory function - neurobiological bases of hearing," Edited by G. M. Edelman, W. E. Gall and W. M. Cowan, New York, Wiley, 721-745.
- Kopp, C., Doerrscheidt, G. J., and Ruebsamen, R. (1998). "The MNTB is more than a relay," *ARO abstracts #359*.
- Köppl, C. (1997). "Frequency tuning and spontaneous activity in the auditory nerve and cochlear nucleus magnocellularis of the barn owl *Tyto alba*," *J. Neurophysiol.* 77, 364-377.
- Köppl, C. (1997). "Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl *Tyto alba*," *J. Neuroscience* 17, 3312-3321.
- Krimphoff, J., McAdams, S., and Winsberg, S. (1994). "Caractérisation du timbre des sons complexes. II Analyses acoustiques et quantification psychoacoustique," *Journal de Physique IV C5-625-628*.
- Kubovy, M. (1979). "Concurrent pitch-segregation and the theory of indispensable attributes," in "Perceptual organization," Edited by M. Kubovy and J. Pomeranz, Hillsdale, N.J., Lawrence Erlbaum, .
- Kuwada, S., Yin, T. C. T., Haberly, L. B., and Wickesberg, R. E. (1980). "Binaural interaction in the cat inferior colliculus : physiology and anatomy," in "Psychophysical, physiological and behavioral studies in hearing," Edited by G. v. d. Brink and F. A. Bilsen, Delft University Press, 401-411.
- Kuwada, S., Batra, R., and Stanford, T. R. (1989). "Monaural and binaural response properties of neurons in the inferior colliculus of the rabbit : effects of sodium pentobarbital," *J. Neurophysiol.* 61, 269-282.
- Kuwada, S., and Batra, R. (1999). "Coding of sound envelopes by inhibitory rebound in neurons of the superior olivary complex in the unanesthetized rabbit," *J. Neurosc.* 19, 2273-2287.
- Lamoré, P. J. J. (1977). "Pitch and masked thresholds in octave complexes in relation to interaction phenomena in two-tone stimuli in general," *Acustica* 37, 250- 257.
- Langner, G. (1981). "Neuronal mechanisms for pitch analysis in the time domain," *Exp. Brain Res.* 44, 450-454.

- Langner, G. (1983). "Neuronal mechanisms for a periodicity analysis in the time domain," in "Hearing-Physiological bases and psychophysics," Edited by R. Klinke and R. Hartmann, Berlin, Springer-Verlag, 334-341.
- Langner, G. (1983). "Evidence for neuronal periodicity detection in the auditory system of the guinea fowl : implications for pitch analysis in the time domain," *Exp. Brain Res.* 52, 333-355.
- Langner, G., and Schreiner, C. E. (1988). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," *J. Neurophysiol.* 60, 1799-1822.
- Langner, G. (1992). "Periodicity coding in the auditory system," *Hearing Research* 60, 115-142.
- Lazzaro, J., and Mead, C. (1989). "Silicon modeling of pitch perception," *Proc. Natl. Acad. Sci. USA* 86, 9597-9601.
- Lea, A. (1992), "Auditory models of vowel perception," Nottingham unpublished doctoral dissertation.
- Levitt, H., and Rabiner, L. R. (1967). "Binaural release from masking for speech and gain in intelligibility," *J. Acoust. Soc. Am.* 42, 601-608.
- Licklider, J. C. R. (1951). "A duplex theory of pitch perception," *Experientia* 7, 128-134.
- Licklider, J. C. R. (1956). "Auditory frequency analysis," in "Information theory," Edited by C. Cherry, London, Butterworth, 253-268.
- Licklider, J. C. R. (1959). "Three auditory theories," in "Psychology, a study of a science," Edited by S. Koch, New York, McGraw-Hill, I, 41-144.
- Licklider, J. C. R. (1962). "Periodicity pitch and related auditory process models," *International Audiology* 1, 11-36.
- Lin, J. L., and Hartmann, W. M. (1997). "The pitch of mistuned harmonics : evidence for a template model," *J. Acoust. Soc. AM* 103, 2608-2617.
- Lippmann, R. P., and Carlson, B. A. (1997). "Using missing feature theory to actively select features for robust speech recognition with interruptions, filtering, and noise.," *Proc. ESCA Eurospeech*, KN-37-40.
- Litowsky, R. Y., and Yin, T. C. T. (1998). "Physiological studies of the precedence effect in the inferior colliculus of the cat. II. Neural mechanisms," *J. Neurophysiol.* 80, 1302-1316.
- Loeb, G. E., White, M. W., and Merzenich, M. M. (1983). "Spatial cross-correlation - A proposed mechanism for acoustic pitch perception," *Biol. Cybern.* 47, 149-163.
- Lyon, R. F. (1983-1988). "A computational model of binaural localization and separation," in "Natural computation," Edited by W. Richards, Cambridge, Mass, MIT Press, 319-327.
- Lyon, R. (1984). "Computational models of neural auditory processing.," *Proc. IEEE ICASSP*, 36.1.(1-4).
- Lyon, R. (1991). "Automatic gain control in cochlear mechanics," in "Mechanics and biophysics of hearing," Edited by P. Dallos, C.D.Geisler, J. W. Mathews, M. A. Ruggero and C.R.Steele, New York, Springer-Verlag, .
- Lyon, R. F. (1997). "All-pole models of auditory filtering," in "Diversity in auditory mechanics," Edited by L. e. al., Singapore, World scientific publishing, 205-211.

- Maas, W. (1997). "Networks of spiking neurons : the third generation of neural network models," *Neural Networks* 10, 1659-1671.
- Maas, W. (1997). "Fast sigmoidal networks via spiking neurons," *Neural Computation* 9, 279-304.
- Malmierca, M. S., Leergard, T. B., Bajo, V. M., Bjaale, J. G., and Merchan, M. A. (1998). "Anatomic evidence of a three-dimensional mosaic pattern of tonotopic organization in the ventral complex of the lateral lemniscus in cat," *J. Neurosci.* 18, 10603-10618.
- Manley, G. A., and Köppl, C. (1998). "Phylogenetic development of the cochlea and its innervation," *Current Opinion in Neurobiology* 8, 468-479.
- Marin, C., and McAdams, S. (1991). "Segregation of concurrent sounds. II : Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width," *J. Acoust. Soc. Am.* 89, 341-351.
- Marin, C., and McAdams, S. (1996). "The role of auditory beats induced by frequency modulation and polyperiodicity in the perception of spectrally embedded target sounds," *J. Acoust. Soc. Am.* 100, 1736-1753.
- Marr, D. (1982). "Representing and computing visual information," in "Artificial Intelligence : an MIT perspective," Edited by P. H. Winston and R. H. Brown, Cambridge, Mass, MIT Press, 2, 17-82.
- Massaro, D. W. (1987). "Speech perception by ear and by eye : a paradigm for psychological inquiry," Hillsdale NJ, Erlbaum.
- Massaro, D. W. (1990). "Models of integration given multiple sources of information," *Psychological review* 97, 225-252.
- McAdams, S. (1982). "Spectral fusion and the creation of auditory images," in "Music, mind and brain," Edited by M. Clynes, Plenum Press, 279-298.
- McAdams, S., and X., R. (1988). "The role of FM-induced AM in dynamic spectral profile analysis," in "Basic issues in hearing," Edited by H. Duifhuis, J. Horst and H. Wit, London, Academic Press, 359-369.
- McAdams, S. (1989). "Segregation of concurrent sounds. I : Effects of frequency modulation coherence," *J. Acoust. Soc. Am.* 86, 2148-2159.
- McAdams, S., Winsberg, S., Donnadiou, S., De Soete, G., and Krimphoff, J. (1995). "Perceptual scaling of synthesized musical timbres : common dimensions, specificities, and latent subject classes," *Psychological Research* 58, 177-192.
- McAlpine, D., Jiang, D., Shackleton, T. M., and Palmer, A. R. (1998). "Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus," *J. Neurosci.* 18, 6026-6039.
- McFadden, D. (1973). "'Precedence effects and auditory cells with long characteristic delays," *J. Acoust. Soc. Am.* 54, 528-530.
- McKeown, J. D., and Darwin, C. J. (1991). "Effects of phase changes in low-numbered harmonics on the internal representation of complex sounds," *The quarterly journal of experimental psychology* 43A, 401-421.
- McKeown, J. D. (1992). "Perception of concurrent vowels : The effect of varying their relative level," *Speech Communication* 11, 1-13.
- McKeown, J. D., and Patterson, R. D. (1996). "The time course of auditory segregation : Concurrent vowels that vary in duration," *J. Acoust. Soc. Am.* 98, 1866-1877.

- Meddis, R. (1988). "Simulation of auditory-neural transduction : further studies," *J. Acoust. Soc. AM* 83, 1056-1063.
- Meddis, R., and Hewitt, M. J. (1991). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I : Pitch identification," *J. Acoust. Soc. AM* 89, 2866-2882.
- Meddis, R., and Hewitt, M. J. (1991). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II : phase sensitivity," *J. Acoust. Soc. AM* 89, 2883-2894.
- Meddis, R., and Hewitt, M. J. (1992). "Modeling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. AM* 91, 233-245.
- Meddis, R., and O'Mard, L. (1997). "A unitary model of pitch perception," *J. Acoust. Soc. Am.* 102, 1811-1820.
- Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (1999). "A computational algorithm for computing non-linear auditory frequency selectivity," *JASA* (submitted)
- Mellinger, D. K. (1991), "Event formation and separation in musical sound," Stanford Center for computer research in music and acoustics unpublished doctoral dissertation.
- Møller, A. R. (1983). "Auditory physiology," New York, Academic Press.
- Moore, B. C. J. (1977,1982,1989, 1996, 1997). "An introduction to the psychology of hearing," London, Academic Press.
- Moore, B. C. J. (1980). "Neural interspike intervals and pitch," *Audiology* 19, 363-365.
- Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (1984). "Frequency and intensity difference limens for harmonics within complex tones," *J. Acoust. Soc. Am.* 75, 550-561.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1985). "Thresholds for the detection of inharmonicity in complex tones," *J. Acoust. Soc. Am.* 77, 1861-1867.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1986). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.* 80, 479-483.
- Moore, B. C. J. (1987). "The perception of inharmonic complex tones," in "Auditory processing of complex sounds," Edited by Y. a. Watson, Hillsdale, Laurence Erlbaum Associates, 180-189.
- Nordmark, J. (1963). "Some analogies between pitch and lateralization phenomena," *J. Acoust. Soc. Am.* 35, 1544-1547.
- Nordmark, J. O. (1978). "Frequency and periodicity analysis," in "Handbook of perception, vol. IV - Hearing," Edited by E. C. Carterette and P. P. Friedman, New York, Academic Press, 243-282.
- Oertel, D. A. (1983). "Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus," *J. Neurosc.* 3, 2043-2053.
- Oertel, D. (1999). "The role of timing in the brain stem auditory nuclei of vertebrates," *Ann. Rev. Physiol.* 61, 497-519.
- Oliver, D. L., and Huerta, M. F. (1992). "Anatomy of the colliculi," in "The mammalian auditory pathway : neuroanatomy," Edited by D. B. Webster, A. N. Popper and R. R. Fay, New York, Springer Verlag, 168-221.

- O'Mard, L., Hewitt, M. J., and Meddis, R. (1997), "Luteal manual"
- Palmer, A. R., Winter, I. M., and Darwin, C. J. (1986). "The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons," *J. Acoust. Soc. Am.* 79, 100-113.
- Palmer, A. R., Winter, I. M., Gardner, R. B., and Darwin, C. J. (1987). "Changes in the phonemic quality and neural representation of a vowel by alteration of the relative phase of harmonics near F1," in "The psychophysics of speech perception," Edited by M. E. H. Schouten, Dordrecht, Martinus Nijhoff, 371-376.
- Palmer, A. R. (1988). "The representation of concurrent vowels in the temporal discharge patterns of auditory nerve fibers," in "Basic issues in hearing," Edited by H. Duifhuis, J. W. Horst and H. P. Wit, London, Academic Press, 244-251.
- Palmer, A. R. (1990). "The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers," *J. Acoust. Soc. Am.* 88, 1412-1426.
- Palmer, A. R. (1992). "Segregation of the responses to paired vowels in the auditory nerve of the guinea-pig using autocorrelation," in "Audition speech and language," Edited by M. E. H. Schouten, Berlin, Mouton-DeGruyter, 115-124.
- Palmer, A. R., and Winter, I. M. (1992). "Cochlear nerve and cochlear nucleus responses to the fundamental frequency of voiced speech sounds and harmonic complex tones," in "Auditory physiology and perception," Edited by Y. Cazals, L. Demany and K. Horner, Oxford, Pergamon Press, 231-239.
- Park, T. J., Grothe, B., Pollak, G. D., Schuller, G., and Koch, U. (1996). "Neural delays shape selectivity to interaural intensity differences in the lateral superior olive," *J. Neurosci.* 15, 6554-6566.
- Parsons, T. W. (1976). "Separation of speech from interfering speech by means of harmonic selection," *J. Acoust. Soc. Am.* 60, 911-918.
- Patterson, R. D. (1986). "Spiral detection of periodicity and the spiral form of musical scales," *Psychology of music* 14, 44-61.
- Patterson, R. D. (1987). "A pulse-ribbon model of monaural phase perception," *J. Acoust. Soc. Am.* 82, 1560-1586.
- Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (1992). "Complex sounds and auditory images," in "Auditory physiology and perception," Edited by Y. Cazals, K. Horner and L. Demany, Oxford, Pergamon Press, 429-446.
- Patterson, R. D. (1994). "The sound of a sinusoid : time-domain models," *J. Acoust. Soc. Am.* 94, 1419-1428.
- Patterson, R. D. (1994). "The sound of a sinusoid : spectral models," *J. Acoust. Soc. Am.* 96, 1409-1418.
- Patterson, R. D., Allerhand, M., and Guiguère, C. (1995). "Time-domain modeling of peripheral auditory processing : a modular architecture and a software platform," *J. Acoust. Soc. Am.* 98, 1890-1894.
- Patterson, R., Anderson, T. R., 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.

- Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (1983). "Pitch of components of complex tones," *J. Acoust. Soc. Am.* 73, 924-924.
- Perec, G. (1991). "Experimental demonstration of the tomatopic organization in the soprano (*Cantatrix sopranica L.*)," in "*Cantatrix sopranica L.*," Edited by G. Perec, Paris, Seuil, 11-33.
- Pickles, J. O. (1988). "An introduction to the physiology of hearing," London, Academic Press.
- Plomp, R. (1967). "Beats of mistuned consonances," *J. Acoust. Soc. Am.* 42, 462-474.
- Plomp, R. (1967). "Pitch of complex tones," *J. Acoust. Soc. Am.* 41, 1526-1533.
- Puel, J.-L., Bonfils, P., and Pujol, R. (1988). "Selective attention modifies the active micromechanical properties of the cochlea," *Brain Research* 447, 380-383.
- Pujol, R., Blatrix, S., Pujol, T. (1999) Promenade autour de la cochlée,
<http://www.iurc.montp.inserm.fr/cric/audition/index.htm>
- Rhode, W. S., and Greenberg, S. (1992). "Physiology of the cochlear nuclei," in "*The mammalian auditory pathway : physiology*," Edited by A. N. Popper and R. R. Fay, New York, Springer Verlag, 94-152.
- Rhode, W. S. (1995). "Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus," *J. Acoust. Soc. Am.* 97, 2414-2429.
- Ritsma, R. J. (1968). "Pitch : result of an autocorrelation process.," *Proc. 6th ICA*, Tokyo, A101-A104.
- Roberts, B., and Brunstrom, J. M. (1998). "Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes," *J. Acoust. Soc. AM* 104, 2326-2338.
- Robinson, K., and Patterson, R. D. (1995). "The stimulus duration required to identify vowels, their octave, and their pitch chroma," *J. Acoust. Soc. Am.* 98, 1858-1865.
- Robinson, K., and Patterson, R. D. (1995). "The duration required to identify the instrument, the octave, or the pitch chroma of a musical note," *Music Perception* 13, 1-15.
- Romand, R., and Avan, P. (1997). "Anatomical and functional aspects of the cochlear nucleus," in "*The central auditory system*," Edited by G. Ehret and R. Romand, New York, Oxford University Press, 97-191.
- Rose, J. E., Gross, N. B., Geisler, C. D., and Hind, J. E. (1966). "Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source," *J. Neurophysiol.* 29., 288-314.
- Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1967). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," *J. Neurophysiol.* 30, 769-793.
- Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1968). "Patterns of activity in single auditory nerve fibres of the squirrel monkey," in "*Hearing mechanisms in vertebrates*," Edited by A. V. S. De Reuk and J. Knight, London, Churchill, 144-168.
- Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1969). "Some possible neural correlates of combination tones," *J. Neurophysiol.* 32, 402-423.

- Rose, J. E., Kitzes, L. M., Gibson, M. M., and Hind, J. E. (1974). "Observations on phase-sensitive neurons of anteroventral cochlear nucleus of the cat : nonlinearity of cochlear output," *J. Neurophysiol.* 37, 218-253.
- Rosen, S., Baker, R. J., and Kramer, S. (1991). "Characterising changes in filter bandwidth as a function of level.", *Proc. Carcans meeting*, ???
- Rosen, S., Baker, R. J., and Darling, A. (1998). "Auditory nonlinearity at 2 kHz in normal hearing listeners," *J. Acoust. Soc. Am.* 103, 2539-2550.
- Rosenblith, W. A., Miller, G. A., Egan, J. P., Hirsh, I. J., and Thomas, G. J. (1947). "An auditory afterimage?," *Science* 103, 333-335.
- Ross, M. J., Shaffer, H. L., Cohen, A., Freudberg, R., and Manley, H. J. (1974). "Average magnitude difference function pitch extractor," *IEEE Trans. ASSP* 22, 353-362.
- Rouiller, E. M. (1997). "Functional organization of the auditory pathways," in "The central auditory system," Edited by G. Ehret and R. Romand, New York, Oxford University Press, 3-96.
- Rowland, K. C., and Spirou, G. A. (1998). "The adherens organelle, a specialized structure of large nerve terminals in the auditory brainstem.", *Proc. ARO abstract #852*,
- Ruggero, M. A. (1992). "Physiology of the auditory nerve," in "The mammalian auditory pathway : neurophysiology," Edited by A. N. Popper and R. R. Fay, New York, Springer Verlag, 34-93.
- Russel, I., and Palmer, A. (1986). "Filtering due to the inner hair-cell membrane properties and its relation to the phase-locking limit in cochlear nerve cells," in "Auditory frequency selectivity," Edited by B. C. J. Moore and R. D. Patterson, New York, Plenum Press, 199-207.
- Ryugo, D. K. (1992). "The auditory nerve," in "The mammalian auditory pathway : neuroanatomy," Edited by D. B. Webster, A. N. Popper and R. R. Fay, New York, Springer-Verlag, 23-65.
- Sabatini, B. L., and Regehr, W. G. (1999). "Timing of synaptic transmission," *Ann. Rev. Physiol.* 61, 521-542.
- Sachs, M. B., and Young, E. D. (1979). "Encoding of steady-state vowels in the auditory nerve : representation in terms of discharge rate," *J. Acoust. Soc. AM* 66, 470-479.
- Salt, A.N. (1999) Cochlear Fluids Research Laboratory,
<http://oto.wustl.edu/cochlea/>
- Sanes, D. H. (1990). "An in vitro analysis of sound localization mechanisms in gerbil lateral superior olive," *J. Neuroscience* 10, 3494-3506.
- Scheffers, M. T. M. (1983), "Sifting vowels," Gröningen unpublished doctoral dissertation.
- Schofield, B. R., and Cant, N. B. (1997). "Ventral nucleus of the lateral lemniscus in guinea pigs : cytoarchitecture and inputs from the cochlear nucleus," *J. Comp. Neurol.* 379-385, 363-385.
- Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (1962). "Pitch of the residue," *J. Acoust. Soc. AM* 34, 1418-1424.
- Schouten, J. F. (1970). "The residue revisited," in "Frequency analysis and periodicity detection in hearing," Edited by R. Plomp and G. F. Smoorenburg, Sijthoff, 41-58.

- Schreiner, C. E., and Langner, G. (1988). "Coding of temporal patterns in the central auditory nervous system," in "Auditory function - Neurobiological bases of hearing," Edited by G. M. Edelman, W. E. Gall and W. M. Cowan, New York, Wiley, 337-361.
- Schreiner, C. E., and Langner, G. (1988). "Periodicity coding in the inferior colliculus of the cat. II. Topographical organization," *J. Neurophysiol.* 60, 1823-1840.
- Schroeder, M. R. (1968). "Period histogram and product spectrum : new methods for fundamental-frequency measurement," *J. Acoust. Soc. Am.* 43, 829-834.
- Schroeder, M. R., and Hall, J. L. (1974). "Model for the mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am.* 55, 1055-1060.
- Schwartz, I. R. (1992). "The superior olivary complex and lateral lemniscal nuclei," in "The mammalian auditory pathway : neuroanatomy," Edited by D. B. Webster, A. N. Popper and R. R. Fay, New York, Springer-Verlag, 117-167.
- Servière, J. (1986), "La tonotopie du colliculus inférieur chez trois espèces de mammifères (chat, cobaye, singe). Étude anatomo-fonctionnelle par le 14C-2-desoxyglucose," Paris 6 unpublished doctoral dissertation.
- Sethares, W. A. (1997). "Tuning, timbre, spectrum, scale," London, Springer-Verlag.
- Shackleton, T. M., and Meddis, R. (1992). "The role of interaural time differences and fundamental frequency differences in the identification of concurrent vowel pairs," *J. Acoust. Soc. Am.* 91, 3579-3581.
- Shackleton, T. M., and Carlyon, R. P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.* 95, 3529-3540.
- Shackleton, T. M., Meddis, R., and Hewitt, M. J. (1994). "The role of binaural and fundamental frequency difference cues in the identification of concurrently presented vowels.," *QJEP A* 47, 545-563.
- Shamma, S. A. (1985). "Speech processing in the auditory system I : The representation of speech sounds in the responses of the auditory nerve," *J. Acoust. Soc. Am.* 78, 1612-1621.
- Shamma, S. A. (1985). "Speech processing in the auditory system II : Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," *J. Acoust. Soc. Am.* 78, 1622-1632.
- Shamma, S. (1988). "The acoustic features of speech sounds in a model of auditory processing : vowels and voiceless fricatives," *J. of Phonetics* 16, 77-91.
- Shamma, S., N., S., and Gopalaswamy, P. (1988). "Binaural processing without neural delays," in "Basic issues in hearing," Edited by H. Duifhuis, J. W. Horst and H. P. Wit, London, Academic Press, 135-143.
- Shamma, S. A., Shen, N., and Gopalaswamy, P. (1989). "Stereoausis : binaural processing without neural delays," *J. Acoust. Soc. Am.* 86, 989-1006.
- Shamma, S., and Klein, D. (2000). "The case of the missing pitch templates : how harmonic templates emerge in the early auditory system," *J. Acoust. Soc. Am.* 107, 2631-2644.
- Shamma, S. A. (2000). "Physiological basis of timbre perception," in "The new cognitive neurosciences," Edited by M. S. Gazziniga, Cambridge, Mass., MIT Press, 411-423.

- Shepard, R. N. (1964). "Circularity in judgments of relative pitch," *J. Acoust. Soc. Am.* 36, 2346-2353.
- Shepard, R. N. (1982). "Structural representations of musical pitch," in "The psychology of music," Edited by D. Deutsch, Orlando, Academic Press, 343-390.
- Shofner, W. P. (1991). "Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla," *J. Acoust. Soc. Am.* 90, 2450-2466.
- Shofner, W. P. (1998). "Evidence that primarylike units are the major CN subsystem which encodes pitch related information in their temporal discharge.", *Proc. ARO abstract #382*, .
- Siebert, W. M. (1970). "Frequency discrimination in the auditory system : place or periodicity mechanisms," *Proc. IEEE* 58, 723-730.
- Slaney, M. (1990). "A perceptual pitch detector.", *Proc. ICASSP-90*, 357-360.
- Slaney, M. (1993), "An efficient implementation of the Patterson-Holdsworth auditory filter bank," *Apple Computer technical report*, 35.
- Smith, P. H., Joris, P. X., and Yin, T. C. T. (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, P. H. (1995). "Structural and functional differences distinguish principal from nonprincipal cells in the guinea pig MSO slice," *J. Neurophysiol.* 73, 1653-1667.
- Smith, P. H., Joris, P. X., and Yin, T. C. T. (1998). "Anatomy and physiology of principal cells of the medial nucleus of the trapezoid body (MNTB) of the cat," *J. Neurophysiol.* 79, 3127-3142.
- Spangler, K. M., Warr, B., and Henkel, C. K. (1985). "The projections of principal cells of the medial nucleus of the trapezoid body in the cat," *J. Comp. Neurol.* 238, 249-262.
- Spirou, G. A., Rice, J. J., and Young, E. D. (1992). "Interneurons of the dorsal cochlear nucleus shape the responses of principal cells to complex sounds," in "Auditory physiology and perception," Edited by Y. Cazals, L. Demany and K. Horner, Oxford, Pergamon Press, 389-396.
- Spirou, G. A., Walker, M. P., and Berrebi, A. S. (1995). "Connectivity of the lateral nucleus of the trapezoid body in cats," *ARO abstracts #155*.
- Spirou, G. A., Rowland, K. C., and Berrebi, A. S. (1997). "Ultrastructural features of neurons in the cat lateral nucleus of the trapezoid body," #645.
- Spirou, G. A., and Berrebi, A. S. (1996). "Glycine immunoreactivity in the lateral nucleus of the trapezoid body in cats," #465.
- Spitzer, M. W., and Semple, M. N. (1998). "Transformation of binaural response properties in the ascending auditory pathway : influence of time-varying interaural phase disparity," *J. Neurophysiol.* 80, 3062-3076.
- Srulovicz, P., and Goldstein, J. L. (1983). "A central spectrum model : a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," *J. Acoust. Soc. Am.* 73, 1266-1276.
- Stubbs, R. J., and Summerfield, Q. (1988). "Evaluation of two voice-separation algorithms using normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* 84, 1236-1249.

- Stubbs, R. J., and Summerfield, Q. (1990). "Algorithms for separating the speech of interfering talkers : Evaluations with voiced sentences, and normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* 87, 359-372.
- Suga, N. (1990). "Cortical computational maps for auditory imaging," *Neural Networks* 3, 3-21.
- Suga, N. (1994). "Processing of auditory information carried by species-specific complex sounds," in "The cognitive neurosciences," Edited by M. S. Gazzaniga, Cambridge, Mass., MIT Press, 295-313.
- Summerfield, Q., Lea, A., and Marshall, D. (1990). "Modelling auditory scene analysis : strategies for source segregation using autocorrelograms," *Proc. Institute of Acoustics* 12, 507-514.
- Summerfield, Q., and Assmann, P. F. (1991). "Perception of concurrent vowels : Effects of harmonic misalignment and pitch-period asynchrony," *J. Acoust. Soc. Am.* 89, 1364-1377.
- Summerfield, Q. (1992). "Roles of harmonicity and coherent frequency modulation in auditory grouping," in "The auditory processing of speech : from sounds to words," Edited by M. E. H. Schouten, Berlin, Mouton de Gruyter, 157-166.
- Summerfield, Q., and Culling, J. F. (1992). "Periodicity of maskers not targets determines ease of perceptual segregation using differences in fundamental frequency.", *Proc. 124th meeting of the ASA*, 2317(A).
- Summerfield, Q., and Culling, J. F. (1992). "Auditory segregation of competing voices : absence of effects of FM or AM coherence," *Phil. Trans. R. Soc. Lond. B* 336, 357-366.
- Szentágothai, J., and Arbib, M. A. (1975). "Conceptual Models of Neural Organization," Cambridge, MA., The MIT Press.
- Terhardt, E. (1974). "Pitch, consonance and harmony," *J. Acoust. Soc. AM* 55, 1061-1069.
- Terhardt, E. (1979). "Calculating virtual pitch," *Hearing Research* 1, 155-182.
- Terhardt, E., Stoll, G., and Seewann, M. (1982). "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *J. Acoust. Soc. AM* 71, 679-688.
- Thompson, D. A. (1917, reedited 1961). "On growth and form," Cambridge, Cambridge University Press.
- Thorpe, S., Fize, F., and Marlot, C. (1996). "Speed of processing in the human visual system," *Nature* 381, 520-522.
- Trussel, L. O. (1997). "Cellular mechanisms for preservation of timing in central auditory pathways," *Curr. Opin. Neurobiol.* 7, 487-492.
- Trussel, L. O. (1999). "Synaptic mechanisms for coding timing in auditory neurons," *Ann. Rev. Physiol.* 61, 477-496.
- Tsuchitani, C. (1997). "Input from the medial nucleus of the trapezoid body to an interaural level detector," *HR* 105, 211-224.
- Ueda, K., and Ohgushi, K. (1984). "Spatial representations of two components of pitch using multidimensional scaling technique [in Japanese].", *Proc. Proc. ASJ special interest group in hearing*, 798-804.

- Usrey, W. M., and Reid, R. C. (1999). "Synchronous activity in the visual system," *Ann. Rev. Physiol.* 61, 435-456.
- van Noorden, L. (1982). "Two channel pitch perception," in "Music, mind, and brain," Edited by M. Clynes, London, Plenum press, 251-269.
- Volman, S. F., and Konishi, M. (1989). "Spatial selectivity and binaural responses in the inferior colliculus of the great horned owl," *J. Neurophysiol.* 9, 3083-3096.
- von Békésy, G. (1960). "Experiments in hearing (translated and edited by E.G. Wever)," McGraw-Hill.
- Warr, W. B. (1992). "Organization of olivocochlear efferent systems in mammals," in "The mammalian auditory pathway : neuroanatomy," Edited by D. B. Webster, A. N. Popper and R. R. Fay, New York, Springer Verlag, 410-448.
- Weintraub, M. (1985), "A theory and computational model of auditory monaural sound separation," Stanford unpublished doctoral dissertation.
- Wever, E. G., and Bray, C. W. (1930). "Present possibilities for auditory theory," *The psychological review* 37, 365-380.
- Wever, E. G., and Bray, C. W. (1930). "The nature of acoustic response : the relation between sound frequency and frequency of impulses in the auditory nerve," *Journal of experimental psychology* 13, 373-387.
- Whitfield, I. C. (1978). "The neural code," in "Handbook of perception, vol IV," Edited by E. C. Carterette and M. P. Friedman, New York, Academic Press, 163-183.
- Whitfield, I. C. (1979). "Periodicity, pulse interval and pitch," *Audiology* 18, 507-512.
- Whitfield, I. C. (1980). "Theory and experiment in so-called pulse-interval pitch," *Audiology* 20, 86-88.
- Wickesberg, R. E., and Oertel, D. (1988). "Tonotopic projection from the dorsal to the anteroventral cochlear nucleus of mice," *J. Comp. Neurol.* 268, 389-399.
- Wickesberg, R. E., and Oertel, D. (1990). "Delayed, frequency-specific inhibition in the cochlear nuclei of mice : a mechanism for monaural echo suppression.," *J. Neuroscience* 10, 1762-1768.
- Wightman, F. L. (1971). "A model of perceived pitch based on weighted autocorrelation.," *Proc. 7th ICA, Budapest*, 361-364.
- Wightman, F. L. (1973). "The pattern-transformation model of pitch," *J. Acoust. Soc. AM* 54, 407-416.
- Wightman, F. L. (1973). "Pitch and stimulus fine structure," *J. Acoust. Soc. AM* 54, 397-406.
- Winer, J. A. (1992). "Auditory forebrain functional anatomy," in "The mammalian auditory pathway : neuroanatomy," Edited by D. R. Webster, A. N. Popper and R. R. Fay, New York, Springer-Verlag, 222-409.
- Winter, I. W., and Palmer, A. R. (1990). "Temporal responses of primarylike anteroventral cochlear nucleus units to the steady-state vowel /i/," *J. Acoust. Soc. Am.* 88, 1437-1441.
- Wu, S. H., and Kelly, J. B. (1991). "Physiological properties of neurons in the mouse superior olive : membrane characteristics and postsynaptic responses studied in vitro," *J. Neurophysiol.* 65, 230-246.

- Wu, S. H., and Kelly, J. B. (1992). "Binaural interaction in the cat superior olive : time difference sensitivity studied in mouse brain slice," *J. Neurophysiol.* 68, 1151-1159.
- Wu, S. H. (1997). "Physiological properties of neurons in the ventral nucleus of the lateral lemniscus of the rat : intrinsic membrane properties and synaptic responses," *J. Neurophysiol.* 81, 2862-2874.
- Yang, L., and Pollak, G. D. (1997). "Differential response properties to amplitude modulated signals in the dorsal nucleus of the lateral lemniscus of the mustache bat and the roles of GABAergic inhibition," *J. Neurophysiol.* 77, 324-340.
- Yang, L., Monsivais, P., and Rubel, E. W. (1999). "The superior olivary nucleus and its influence on nucleus laminaris : a source of inhibitory feedback for coincidence detection in the avian auditory brainstem," *J. Neuroscience* 15, 2313-2325.
- Yates (1995). "Cochlear structure and function," in "Hearing," Edited by B. C. J. Moore, San Diego, Academic Press, 41-74.
- Yin, T. C. T., Kuwada, S., and Sujaku, Y. (1984). "Interaural time sensitivity of high-frequency neurons in the inferior colliculus," *J. Acoust. Soc. AM* 76, 1401-1410.
- Yin, T. C. T., Chan, J. C. K., and Carney, L. H. (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.
- Yin, T. C. T., and Chan, J. C. K. (1988). "Neural mechanisms underlying interaural time sensitivity to tones and noise," in "Auditory function - Neurological bases of hearing," Edited by G. M. Edelman, W. E. Gall and W. M. Cowan, New York, Wiley, 385-430.
- Yin, T. C. T., and Chan, J. C. K. (1990). "Interaural time sensitivity in medial superior olive of cat," *J. Neurophysiol.* 64, 465-488.
- Yost, W. A. (1996). "Pitch of iterated rippled noise," *J. Acoust. Soc. AM* 100, 511-518.
- Yost, W. A. (1996). "Pitch strength of iterated rippled noise," *J. Acoust. Soc. AM* 100, 3329-3335.
- Yost, W. A., Patterson, R., and Scheft, S. (1996). "A time domain description for the pitch strength of iterated rippled noise," *J. Acoust. Soc. AM* 99, 1066-1078.
- Zackenhause, M., Johnson, D. H., and Tsuchitani, C. (1995). "Transient effects during the chopping response of LSO neurons," *J. Acoust. Soc. AM* 98, 1410-1422.
- Zurek, P. M. (1981). "Spontaneous narrowband acoustic signal emitted by human ears," *J. Acoust. Soc. AM* 69, 514-423.
- Zwicker, E. (1986). "Spontaneous oto-acoustic emissions, threshold in quiet, and just noticeable amplitude modulation at low levels," in "Auditory frequency selectivity," Edited by B.C.J. Moore and R. D. Patterson, New York, Plenum Press, 49-59.
- Zwicker, U. T. (1984). "Auditory recognition of diotic and dichotic vowel pairs," *Speech Communication* 3, 256-277.

Tirés à part d'articles

Le reste du document comprend les copies d'une sélection d'articles, pour la plupart parus dans le Journal of the Acoustical Society of America. Les numéros sont ceux de la liste des publications.

- [17] de Cheveigné, A. (2000). "The auditory system as a separation machine.", Proc. International Symposium on Hearing, august 2000.
- [1] de Cheveigné, A. (2000). "A model of the perceptual asymmetry between peaks and troughs of frequency modulation," J. Acoust. Soc. Am., in press.
- [2] de Cheveigné, A. (1999). "Waveform interactions and the segregation of concurrent vowels," J. Acoust. Soc. Am. 106, 2959-2972.
- [3] de Cheveigné, A. (1999). "Pitch shifts of mistuned partials : a time-domain model," J. Acoust. Soc. Am. 106, 887-897.
- [4] de Cheveigné, A. (1999). "Vowel-specific effects in concurrent vowel identification," J. Acoust. Soc. Am. 106, 327-340.
- [5] de Cheveigné, A., and Kawahara, H. (1999). "Missing data model of vowel perception," J. Acoust. Soc. Am. 105, 3497-3508.
- [6] de Cheveigné, A., and Kawahara, H. (1999). "Multiple period estimation and pitch perception model," Speech Communication 27, 175-185.
- [8] de Cheveigné, A. (1998). "Cancellation model of pitch perception," J. Acoust. Soc. Am. 103, 1261-1271.
- [9] de Cheveigné, A. (1997). "Harmonic fusion and pitch shifts of inharmonic partials," J. Acoust. Soc. Am. 102, 1083-1087.
- [10] de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (1997). "Concurrent vowel identification I : Effects of relative level and F0 difference," J. Acoust. Soc. Am. 101, 2839-2847.
- [11] de Cheveigné, A., McAdams, S., and Marin, C. (1997). "Concurrent vowel identification II : Effects of phase, harmonicity and task," J. Acoust. Soc. Am. 101, 2848-2856.
- [12] de Cheveigné, A. (1997). "Concurrent vowel identification III : A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. 101, 2857-2865.
- [13] de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (1995). "Identification of concurrent harmonic and inharmonic vowels : A test of the theory of harmonic cancellation and enhancement," J. Acoust. Soc. Am. 97, 3736-3748.
- [14] de Cheveigné, A. (1993). "Separation of concurrent harmonic sounds : Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. 93, 3271-3290.

Comment les lire Commencez par [17], sans trop vous attarder sur la physiologie et les modèles. Passez ensuite à [8] en sautant la section IV. L'article [14] est à l'origine des idées les plus importantes, mais il est difficile à lire. C'est en fait trois articles en un. Sautez les détails les plus roboratifs.

Arrivé(e) à ce stade vous avez assez d'éléments pour lire le reste des articles. L'expérimentation psychoacoustique se trouve dans [10], [11], [2], [4], [13] (à lire dans cet ordre). Si les détails expérimentaux vous ennuyent, lisez l'analyse bibliographique et la discussion finale de chaque article. Le modèle de ségrégation à base d'annulation harmonique est décrit dans [14], puis dans [12]. Les modèles de hauteur tonale sont dans [8] (hauteur simple), [6] (hauteurs multiples), [9] et [3] (effet de Hartmann), et [1] (effet de Demany). Un modèle d'invariance perceptive du timbre des voyelles est dans [5]. Vous pouvez sauter le modèle (difficile à lire) et passer à la discussion qui contient une citation bibliographique rare.