Stephen McAdams et Emmanuel Bigand

PENSER LES SONS

Psychologie cognitive de l'audition

puf

Psychologie et sciences de la pensée

Penser les sons
Psychologie cognitive de l'audition

PSYCHOLOGIE ET SCIENCES DE LA PENSÉE

Collection dirigée par
Olivier Houdé

PENSER LES SONS

Psychologie cognitive de l'audition

Stephen McADAMS
Emmanuel BIGAND

Presses Universitaires de France

à Marie-Claire Botte

ISBN 2 13 046086 0

Dépôt légal — 1re édition : 1994, mars

108, boulevard Saint-Germain, 75006 Paris

LES COLLABORATEURS

Emmanuel Bigand, Laboratoire d'études des acquisitions et du développement (CNRS), Université de Bourgogne, 6, boulevard Gabriel, F-21000 Dijon, France.

Albert S. Bregman, Psychology Department McGill University, 1205 Dr. Penfield Ave., Montréal, Québec H3A 1B1, Canada.

Robert G. Crowder, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, Connecticut 06520, USA.

Mari Riess Jones, Department of Psychology, The Ohio State University, 142 Townsend Hall, 1885 Neil Ave., Columbus, Ohio 43210, USA.

Stephen McAdams, Laboratoire de psychologie expérimentale (CNRS), Université René-Descartes, 28, rue Serpente, F-75006 Paris and IRCAM, 1, place Stravinsky, F-75004 Paris, France.

Isabelle Peretz, Département de psychologie, Université de Montréal CP 6128, Succ. A, Montréal, Québec H3C 3J7, Canada.

Laurel J. Trainor, Center for Research in Human Development, University of Toronto, Erindale Campus, Mississauga, Ontario L5L 1C6, Canada.

Sandra E. Trehub, Center for Research in Human Development, University of Toronto, Erindale Campus, Mississauga, Ontario L5L 1C6, Canada.

Richard M. Warren, Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA.

William Yee, Department of Psychology, The Ohio State University, 142 Townsend Hall, 1885 Neil Ave., Columbus, Ohio 43210, USA.

Sommaire

Avant-propos

L'esprit voit et écoute, le reste est aveugle et sourd.

Epicharmus, 450 av. J.-C.*

A notre connaissance aucun ouvrage didactique ne présente en langue française l'ensemble des aspects cognitifs de l'audition humaine. C'est pour remédier à cela que des « journées didactiques » furent organisées par le groupe Audition de la Société française d'Acoustique. Des spécialistes des différents domaines de la perception et de la cognition auditive y furent invités et présentent dans les chapitres de ce livre les études réalisées dans leurs champs de recherches spécifiques.

L'audition révèle un paradoxe similaire à celui rencontré dans tous les secteurs de la perception : rien ne semble plus simple que de percevoir les sons de notre environnement et pourtant il s'agit là d'un phénomène particulièrement récalcitrant à l'analyse scientifique. Quelle difficulté par exemple à reconnaître son nom dans une conversation, à différencier le bruit d'une voiture de celui d'un avion à hélice, à percevoir le rythme entraînant d'un rock de Bill Haley, à reconnaître la voix de son enfant ou les bruits de pas d'une personne familière ? Il semble suffisant d'ouvrir grand ses oreilles pour cela.

Imaginez cependant, un seul instant, la quantité d'informations et le nombre de procédures qu'il serait nécessaire de donner à un ordinateur pour le rendre capable de distinguer un violon d'une flûte dans une polyphonie, de détecter un signal d'alarme dans le bruit de fond environnant, de saisir une relation entre un thème et ses variations, de repérer une anomalie de moteur à la seule écoute du bruit d'une voiture, de détecter un bruit inhabituel dans les battements de cœur d'un patient, etc. Malgré le très haut degré de sophistication technologique actuel, il est fort probable que ce type de question ne pourra être résolu qu'après de nombreuses années d'études en intelligence artificielle et en traitement du signal, tant la quantité d'informations nécessaires est grande et la façon dont elles doivent être combinées complexe.

* Cité par S. Coren et L. M. Ward, *Sensation and perception*, 3e éd., San Diego, Harcourt, Brace, Jovanovich, 1989.

Cette difficulté à analyser les processus d'écoute souligne la grande richesse de l'information présente dans le monde sonore qui nous entoure. L'objectif principal de ce livre est de rendre compte de notre connaissance actuelle des processus cognitifs qui traitent cette information chez l'homme. Il s'adresse non seulement aux étudiants en psychologie mais également aux étudiants en sciences cognitives et aux scientifiques spécialisés dans d'autres domaines que la psychologie de l'audition. Une connaissance élémentaire des termes usuels en acoustique, des concepts et des méthodes habituels en psychologie expérimentale est cependant préférable.

S. M.
E. B.
Paris, juin 1993

Remerciements

Cet ouvrage collectif fait suite aux quatrièmes journées didactiques organisées par le groupe Audition de la Société française d'Acoustique (SFA). Les trois précédentes ont porté sur la physiologie de la cochlée (Aran, Dancer, Dolmazon, Pujol et Tran Ba Huy, 1988), la psychoacoustique et la perception auditive (Botte, Canévet, Demany et Sorin, 1989) et le système nerveux auditif central (Roman, 1992). L'organisation de ces journées et la préparation de ce livre ont été rendues possibles grâce à l'aide financière et à l'infrastructure du groupe Audition de la SFA, du programme « Sciences de la cognition » du ministère de la Recherche et de la Technologie et du ministère de l'Education nationale, de la direction de recherches, études et techniques du ministère de la Défense et de l'Institut de recherche et coordination acoustique/musique (IRCAM).

Ces journées didactiques furent conçues et animées avec vigueur depuis plusieurs années par Marie-Claire Botte, alors présidente du groupe Audition et vice-présidente de la SFA. Grâce à ses efforts soutenus, le statut de la recherche française sur l'audition a suivi une évolution remarquable ces dernières années. C'est donc avec une grande reconnaissance et une profonde affection que nous lui dédions ce livre.

Les chapitres de cet ouvrage ont été soumis à un comité de lecture, chacun ayant été critiqué par deux à quatre spécialistes et par les directeurs de l'ouvrage (chacun d'eux assurant l'expertise du chapitre de l'autre). Nous souhaitons particulièrement remercier tous les auteurs pour leur coopération lors de cette tâche ardue ainsi que les experts suivants pour leur critique : Josiane Bertoncini, Jean-François Camus, Eric Clarke, Rachel Clifton, Christopher Darwin, Stanislas Dehaene, Jay Dowling, Francis Eustache, Stephen Handel, Alain Lieury, Brian Moore, Jose Morais, Bruno Repp, Ken Robinson, Earl Schubert, Karen Yankelovitch, Charles Watson, Robert Zatorre, et trois experts supplémentaires qui souhaitent rester anonymes. Nous remercions aussi Cécile Marin

pour la réalisation de l'index ainsi que Marie-Claire Botte, Carolyn Drake et les auteurs pour leur aide lors de la préparation du glossaire.

Ce livre est également publié en anglais sous le titre *Thinking in Sound : The Cognitive Psychology of Human Audition* par Oxford University Press.

CHAPITRE I

Introduction à la cognition auditive

Stephen McAdams et Emmanuel Bigand

QU'Y A-T-IL DE COGNITIF DANS L'AUDITION ?

L'étymologie du terme cognitif renvoie à la notion de connaissance. Ce terme a ensuite été utilisé dans un sens plus précis pour désigner les conditions qui permettent l'acquisition et le développement d'une connaissance du monde. Il va sans dire que la perception joue un rôle essentiel à ce niveau : aucune théorie de la connaissance ne serait complète sans une théorie de son acquisition et donc de la perception. L'étude des aspects cognitifs de l'audition a pour objectif de comprendre comment l'information auditive contribue fondamentalement au développement de la connaissance.

Bien sûr, les processus de traitement de l'information acoustique dont il sera question ici pourront sembler très différents de ceux intervenant dans des activités intellectuelles plus abstraites telles que la pensée, la logique, le raisonnement, la prise de décision, l'imagination, etc. Deux remarques doivent cependant être faites sur ce point. Tout d'abord l'originalité du projet cognitiviste est de présenter une vision intégrée de l'ensemble des processus intellectuels en mettant en évidence la continuité existant entre les aspects les plus élémentaires de ces activités (traitement de l'information sensorielle) et les aspects les plus abstraits (traitement de l'information symbolique). Par conséquent, le projet cognitiviste dépasse la traditionnelle division en fonctions intellectuelles indépendantes : perception, mémoire, apprentissage, langage, intelligence, etc. Envisager des aspects cognitifs dans l'audition, c'est vouloir situer les processus de perception auditive à l'intérieur de cette continuité.

La seconde remarque concerne plus directement la nature des processus mentaux impliqués dans la perception auditive. L'importance accordée au terme cognitif dans cet ouvrage signifie que, par-delà les étapes élémentaires de traitement, des processus de traitement de haut niveau (représentations mentales, prise de décision, inférence,

interprétation) semblent indispensables au système auditif pour élaborer une représentation cohérente du monde sonore.

En effet, le postulat de base d'une approche cognitive de l'audition est que l'information sensorielle doit être interprétée pour donner naissance à une perception cohérente. Cette interprétation est nécessaire puisque l'information contenue dans les stimuli qui atteignent les organes sensoriels se révèle souvent incomplète ou ambiguë. Dans ces cas, le système perceptif doit représenter puis comparer des informations qui ne sont plus directement disponibles au niveau sensoriel. Ceci est d'autant plus frappant dans le cas de l'audition, puisque les événements sonores se succèdent dans le temps : l'élaboration d'une représentation mentale s'avère indispensable pour percevoir leur structure, c'est-à-dire pour établir des relations entre des événements séparés par plusieurs minutes ou même plusieurs heures. La musique est une structure sonore exemplaire de ce point de vue : comment percevoir l'unité d'un ensemble de sons qui se développe sur de très longs laps de temps (une heure et demie dans le cas de la 9e *Symphonie* de Beethoven) sans élaborer une représentation des sous-structures (thématiques par exemple) qui sont développées dans l'œuvre ?

Lorsque l'information immédiatement disponible au niveau sensoriel se révèle insuffisante, le système perceptif analyse la situation en considérant la connaissance acquise sur le monde sonore. Même chez le nouveau-né, l'information de l'environnement ne parvient pas dans un organisme complètement naïf. Les connaissances interagissent avec les données sensorielles actuelles dans l'interprétation des stimulations auditives. Imaginez-vous un instant en pleine forêt amazonienne : vous entendriez exactement les mêmes bruits que le guide qui vous accompagne, mais, étant donné votre manque de connaissance du milieu, vous seriez incapable d'extraire du fond sonore les sons correspondant aux cris de l'iguane, aux singes macaques, aux chants des ouistitis ou aux bruissements des arbres tropicaux. De ce fait, vous seriez dans l'incapacité d'attribuer une signification à l'ensemble de la structure sonore, ce qui pourrait être important pour votre survie dans l'environnement. De la même façon, les *oreilles d'or* de la marine française, qui ont été entraînées à la détection de sonar, parviennent à percevoir et à identifier des sources sonores comme des baleines, des marsouins, des bancs de poissons, des paquebots, les cliquetis des crevettes là où la plupart d'entre nous n'entendraient qu'un ensemble indifférencié de bruits sous-marins. Ils réussissent même à identifier les bruits provenant de bateaux de commerce ou de bateaux militaires, de bateaux en surface ou de sous-marins, propulsés par un

moteur nucléaire ou un moteur diesel, voire même jusqu'à la nationalité russe, américaine ou française du sous-marin. Dans les cas les plus dramatiques, l'incapacité à différencier les sources sonores et à identifier leur origine peut aboutir à ne pas déduire de ces signaux la présence d'un danger imminent : la présence d'un jaguar dans le premier exemple ou celle d'un sous-marin clandestin dans le dernier.

Dans le cas de structures sonores telles que la musique, dont l'organisation est hautement déterminée par des règles culturelles, une simple observation de l'information enregistrée au niveau sensoriel ne suffit pas à expliquer les très grandes différences perçues lorsqu'on passe d'un quatuor de Mozart à un quatuor de Beethoven par exemple.

Dans bien d'autres situations, l'information disponible au niveau sensoriel peut se révéler trop ambiguë pour donner lieu à une perception univoque de la situation. Imaginez-vous par exemple la nuit dans un vieux manoir isolé dans la campagne : les bruits perçus résultent-ils des pas d'un cambrioleur cherchant à s'introduire dans la maison, ou s'agit-il simplement des crissements du vent sur les vieux matériaux ? L'identification de la source sonore pourra être différente selon le type d'information privilégiée : régularité des bruits ou qualités timbrales des matériaux mis en vibration. Confronté à des stimuli ambigus de ce type, le système perceptif prendra des décisions inconscientes pour organiser la figure sonore. La connaissance de l'auditeur peut jouer un rôle important à ce niveau : dans l'exemple précédent, un enfant, un adulte ou le propriétaire de la maison arriveraient sans doute à des interprétations différentes.

Les illusions auditives constituent une autre situation illustrant le travail d'interprétation réalisé par le système perceptif : dans ce cas, l'inférence effectuée sur la base des informations disponibles est incorrecte et conduit à la perception d'un objet sonore fictif. Les compositeurs ont su depuis plusieurs siècles exploiter ces caractéristiques de la perception pour créer des figures sonores séduisantes. Les sonates et partitas pour violon seul de J.-S. Bach offrent de nombreux cas remarquables d'illusion auditive : l'auditeur perçoit assez clairement la présence de deux violons jouant dans des registres différents bien qu'il n'y ait en fait qu'un seul violoniste jouant très rapidement des notes alternativement graves et aiguës. De la même façon, en jouant une même mélodie en parallèle à des hauteurs différentes et dans plusieurs familles d'instruments, Ravel, dans le *Boléro*, réussit à détourner l'oreille de la perception d'un ensemble de timbres différents et à l'orienter vers la perception d'un timbre unique aux qualités chatoyantes et nouvelles qui provient d'une source sonore « virtuelle ».

ÉTAT ACTUEL DE LA RECHERCHE SUR LA COGNITION AUDITIVE

Notre perception du monde sonore dépasse ainsi largement les qualités de l'information sensorielle disponible à chaque instant : elle résulte d'un traitement de l'information. L'étude des aspects cognitifs du traitement effectué dans différentes modalités sensorielles est actuellement en grand progrès dans les sciences cognitives. Dans le domaine visuel, les processus de traitement de haut niveau sont l'objet d'un très grand nombre d'articles et d'ouvrages. Marr (1982), Pinker (1984) et Humphreys et Bruce (1989) proposent une importante synthèse des concepts théoriques et des données expérimentales fondamentales dans ce domaine. Dans le cas de l'audition, l'étude des processus cognitifs a été abordée dans des publications concernant essentiellement la perception et la compréhension du langage. Pour l'essentiel, les ouvrages généraux de psychologie traitent habituellement des thèmes suivants :

— perception des qualités, des attributs et des patterns de stimuli (pour toutes les modalités sensorielles) ;
— processus d'organisation perceptive (le plus souvent dans le seul domaine de la vision) ;
— catégorisation et constance perceptive (essentiellement dans les domaines de la vision et du langage) ;
— reconnaissance et identification des objets, des événements et des patterns d'événements (vision et langage) ;
— processus de mémorisation (vision et langage, avec quelques références occasionnelles à l'audition) ;
— processus attentionnels (principalement pour la vision, mais l'écoute dichotique est souvent abordée) ;
— développement, apprentissage, acquisition d'habiletés (rarement dans le domaine auditif) ;
— aspects grammaticaux des grandes structures temporelles (exclusivement dans le domaine du langage, avec de très rares références à la musique) ;
— résolution de problèmes et raisonnement (aucune référence aux situations de résolution de problèmes dans le domaine sonore comme ceux impliqués dans la composition musicale par exemple).

Les ouvrages comparables dans l'audition non verbale (Moore, 1982 ; Warren, 1982 ; Botte, Canévet, Demany et Sorin, 1989) concernent essentiellement les aspects physiologiques et psychoacous-

tiques de l'audition. Les publications récentes de Sloboda (1985), Dowling et Harwood (1986), Handel (1989), Dooling et Hulse (1989) et Bregman (1990), marquent cependant un tournant important puisque l'audition y est abordée dans des situations d'écoute et avec des stimuli sonores très différents de ceux habituellement rencontrés dans les laboratoires de psychoacoustique depuis plus d'un siècle.

On constatera que le développement de la psychologie de la musique a fortement contribué à l'étude des processus cognitifs dans le domaine de l'audition non verbale. Depuis les vingt à trente dernières années, la problématique et les méthodes d'étude de ce qu'il est convenu d'appeler « l'esprit musicien » se sont particulièrement enrichies. La plupart des auditeurs possèdent et utilisent des processus perceptifs et cognitifs hautement sophistiqués pour comprendre, apprécier et participer aux activités musicales. Le fait que les systèmes musicaux observés dans toutes les cultures du monde aient atteint un degré de complexité structurelle (ou grammaticale) comparable à ceux rencontrés dans le langage et qu'ils puissent eux aussi être acquis sans apprentissage spécifique dès un très jeune âge est particulièrement intéressant pour les sciences cognitives. La place occupée par la psychologie de la musique dans la plupart des chapitres suivants témoigne de son importance pour la cognition auditive.

Un thème reste cependant peu étudié chez l'adulte : celui de l'intégration des processus cognitifs auditifs et des processus intervenant dans d'autres modalités sensorielles ou dans des systèmes cognitifs généraux impliqués dans la vie quotidienne. L'utilisation des indices sonores pour se déplacer en toute sécurité dans l'environnement urbain, pour actionner et conduire des machines sophistiquées ou pour évaluer l'identité, et la signification d'événements de l'environnement qui ne se trouvent pas actuellement dans le champ de vision sont des situations qui illustrent cette intégration. L'étude de ces situations montrera probablement que l'audition joue dans la vie quotidienne un rôle plus important que ce que l'on a considéré jusqu'à présent.

PRÉSENTATION DE L'OUVRAGE

Il y a de nombreuses façons de subdiviser un champ de recherche. Celle adoptée dans cet ouvrage comprend les différentes spécialisations suivantes, certaines pouvant se chevaucher partiellement : orga-

nisation perceptive (Bregman, chap. II), perception des caractéristiques globales des séquences acoustiques (Warren, chap. III), processus attentionnels (Jones et Yee, chap. IV), mémoire (Crowder, chap. V), reconnaissance (McAdams, chap. VI), neuropsychologie (Peretz, chap. VII), perception de la musique (Bigand, chap. VIII) et psychologie du développement (Trehub et Trainor, chap. IX).

Une représentation schématique des principaux processus auditifs étudiés dans cet ouvrage et de leurs interactions est proposée dans la figure 1. Les vibrations sonores parvenant à l'oreille interne sont analysées et transformées en impulsions nerveuses qui progressent vers le cerveau par le nerf auditif. Ce processus de *transduction* est brièvement abordé par McAdams (chap. VI).

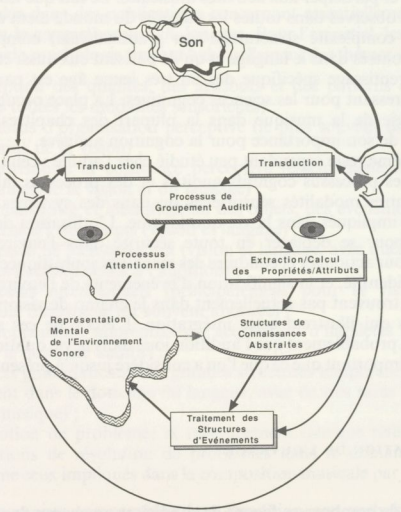

Figure 1. — Principaux processus de traitement de l'information auditive et leurs interactions.

Les processus de groupement auditif effectuent la fusion et la ségrégation des éléments sonores simultanés en événements auditifs et organisent les événements successifs en flux auditifs. Ces processus interviennent quotidiennement dans différentes situations. Par exemple, si nous regardons la télévision et qu'une voiture klaxonne soudainement dans la rue, notre système perceptif parviendra sans aucune hésitation à interpréter ce nouveau bruit comme un objet sonore distinct, se surimposant à ceux provenant au même moment de la télévision et ne se confondant pas avec eux. Ces processus de groupement auditifs sont amplement présentés par Bregman (chap. II) et sont également discutés dans les chapitres de Warren (chap. III), Jones et Yee (chap. IV), Bigand (chap. VIII) et Trehub et Trainor (chap. IX).

Conformément à la théorie de l'analyse des scènes auditives, le groupement auditif précède généralement *l'extraction ou le calcul des propriétés ou attributs perceptifs.* Ces attributs constituent des propriétés perceptives qui sont dérivées des éléments qui ont été préalablement groupés. L'œuvre musicale de Rimsky-Korsakov intitulée *Le vol du bourdon* permet d'illustrer la nature de ces processus : les notes s'y suivent à une vitesse si rapide qu'elles forment un flux sonore unique et ne peuvent pas être entendues séparément. L'auditeur perçoit essentiellement la forme globale de la ligne mélodique rythmée par ses pics et par les notes accentuées par le trompettiste. La perception globale de la structure et de l'organisation temporelle des séquences sonores est abordée dans les chapitres de Warren (chap. III) et Jones et Yee (chap. IV).

Une fois représentées dans le système perceptif, les qualités perceptives peuvent être interprétées en fonction des *structures de connaissances abstraites* évoquées. L'événement sonore, ou la séquence d'événements, est alors reconnu, identifié, et reçoit une signification qui dépend du contexte et de l'expérience antérieure de l'auditeur. Imaginez-vous en train de préparer le repas dans la cuisine. Tout d'un coup se produit un important vacarme dans la salle à manger. L'analyse du bruit perçu vous permet d'identifier le bruit d'assiettes qui se brisent, le son des fourchettes et des cuillères rebondissant sur le carrelage, le son feutré du saladier s'écrasant sur le sol et au milieu de tout cela le miaulement plaintif du chat. L'identification de ces bruits vous permet alors de donner une signification à l'ensemble de la scène : le chat en jouant avec le coin de la nappe a fait tomber tout ce qui se trouvait sur la table. Les chapitres de Crowder (chap. V), McAdams (chap. VI) et Peretz (chap. VII) présentent

les processus psychologiques et neuropsychologiques impliqués dans l'identification des sources sonores, des sons et des séquences de sons. Le chapitre de Crowder porte plus précisément sur les aspects spécifiques de l'audition impliqués dans la mémoire. La perception des relations entre les événements musicaux est elle aussi partiellement déterminée par les connaissances acquises par des auditeurs adultes au contact d'une culture musicale précise. La structure des connaissances musicales est abordée par Bigand dans le cadre de la musique occidentale tonale/métrique (chap. VIII). L'existence chez le bébé d'aptitudes précoces permettant leur développement est abordée tout au long des travaux expérimentaux présentés par Trehub et Trainor (chap. IX). Peretz (chap. VII) discute plusieurs cas de patients cérébro-lésés chez qui l'accès à de telles structures est sérieusement altéré.

La perception des relations entre les événements sonores induit un cadre interprétatif général qui peut influencer la perception des événements ultérieurs. Ce cadre conditionne fortement l'élaboration de larges réseaux de relations structurelles (de nature hiérarchique ou associative). Les processus de *traitement de la structure des événements* sont responsables de ce niveau de perception et de compréhension (fig. 1). Imaginez-vous assis dans une salle de concert attendant l'arrivée du chef d'orchestre. Les musiciens sont en train de s'échauffer créant ainsi un magma sonore de forte intensité. Même en écoutant attentivement, vous ne parvenez pas à saisir de relation entre ce que chacun joue. Vous en concluez que probablement chaque musicien répète différents passages de la partition. Puis le concert commence. Cette fois, les événements semblent se suivre naturellement et vous percevez clairement que l'œuvre progresse et se développe dans le temps : l'exposition des thèmes vient juste de finir, et vous entendez des accords modulants, nettement différents de ceux de la section précédente. Ces changements de tonalités indiquent que le développement des thèmes est commencé. Saisir cette progression suppose que chaque nouvel événement sonore puisse être relié aux précédents. Lorsque ces nouveaux événements sont complètement inconnus de l'auditeur, ou ne correspondent à aucun système de relations connues, des processus de structuration plus élémentaires, telle la segmentation en groupes d'événements, pourraient opérer directement sur la base des relations existant entre les attributs de surface.

L'élaboration d'une *représentation mentale de l'environnement sonore* actuellement perçu par l'auditeur constitue le stade final du traitement de la structure des événements. Son étude est essentielle pour la psychologie cognitive. En effet, cette représentation mentale

est cruciale pour une modalité sensorielle où les stimulations sont totalement transitoires. Comme il n'y a pas d'objet persistant dans l'audition, l'information relative aux événements acoustiques doit être stockée dans le temps, anticipée activement et traitée en fonction des expériences antérieures. Ainsi par exemple dans une symphonie, la saisie du développement des différents événements musicaux nécessite d'avoir élaboré une représentation mentale de ce qui s'est produit au début de l'œuvre. De cette façon, on peut réaliser par exemple que, à ce moment précis, la musique s'oriente progressivement vers la phase de réexposition des thèmes, ce qui indique la fin prochaine du mouvement. Dans le chapitre VIII, Bigand présente plusieurs processus impliqués dans le traitement des structures d'événements et dans l'élaboration d'une représentation mentale de l'organisation des sons de l'environnement. Ces processus déterminent vraisemblablement la capacité de l'auditeur à établir des relations sur de grandes échelles de temps et ainsi à apprécier la forme et le développement des idées musicales.

La perception de l'organisation des événements a des conséquences importantes pour le traitement de l'information auditive puisqu'elle constitue un cadre de référence qui va fortement influencer la structuration des informations sonores à venir : des anticipations et des attentes perceptives vont être mises en œuvre orientant l'attention de l'auditeur sur des événements à venir bien particuliers ou sur des moments spécifiques de la séquence. Ces *processus attentionnels* pourraient aussi être affectés par des régularités perçues dans la structure du stimulus. Il est possible en retour que l'attention influence des processus d'organisation de bas niveau, voire même influence les processus de transduction sensorielle. L'attention dans le domaine auditif et ses implications pour la perception et l'action sont traitées dans le chapitre de Jones et Yee (chap. IV).

CHAPITRE II

L'analyse des scènes auditives : l'audition dans des environnements complexes

Albert S. Bregman

INTRODUCTION

L'approche de la perception auditive qui sera présentée ici s'écarte des thèmes abordés en psychophysique traditionnelle. Cette dernière cherche à répondre à des questions comme celles-ci : Quelle est la quantité minimale d'énergie dont le système perceptif peut saisir la présence ? Quel doit être l'écart entre les fréquences de deux sons purs pour qu'ils puissent être discriminés lorsqu'ils sont émis successivement ou simultanément ? Dans quelle mesure la sonie croît-elle lorsqu'on augmente l'intensité physique du son ? Comment les différences entre les patterns acoustiques captés par les deux oreilles nous renseignent-elles sur la localisation de la source sonore ?

Grâce au développement de l'intelligence artificielle au cours des dernières années, nous savons que répondre à ce type de questions n'apporte qu'une compréhension partielle du fonctionnement du système auditif. Un exercice utile, pouvant aider à cerner les problèmes, consiste à imaginer un robot doté d'un système auditif et à se demander ensuite en quoi ses capacités auditives (telles qu'on les conçoit aujourd'hui) pourraient l'aider.

Imaginons que nous équipions le robot d'une paire de mécanismes sensoriels tenant lieu d'oreilles et que nous y incorporions toutes les propriétés de l'audition humaine. Le robot éprouvera néanmoins une grande difficulté à utiliser l'information reçue, en particulier les mixtures sonores[1]. N'importe quel signal sera enregistré en tant que somme de l'ensemble des sources sonores activées simultanément au moment de la captation. Supposons que le robot ait en mémoire une définition du timbre d'une voix énonçant un mot particulier. Il n'est pas sûr pour autant qu'il pourra l'identifier, la présence

1. *N.d.T.* — La mixture est un agglomérat de sons distincts.

d'autres sons pouvant générer une situation où aucun des segments perçus n'est suffisamment proche de la définition en question. Pire même, la perception d'un mot pourra résulter de la combinaison fortuite de deux voix.

Voici un autre exemple du type de difficulté que pourrait rencontrer le robot. Selon les psychophysiciens, l'intensité des sons parvenant aux deux oreilles peut servir d'indice pour inférer la localisation d'une source sonore. Mais comment le robot saura-t-il, lorsqu'il compare l'intensité aux deux « oreilles », qu'il analyse l'énergie dérivée d'une source sonore unique ? S'il y a deux sources sonores situées en des endroits distincts, la stratégie simple consistant à évaluer l'intensité aux deux oreilles ne fonctionne plus : il faut alors comparer les intensités dérivées de chacune des sources sonores. Comment peut-il connaître la quantité d'énergie provenant de chaque source sonore à chacune des oreilles ?

Pour pouvoir identifier les sons qui composent la mixture parvenant aux oreilles, le système auditif doit créer des descriptions isolées, fondées uniquement sur les composants issus d'une même source de l'environnement. Pour cela, il utilise un processus appelé l' « analyse de scènes auditives » (Bregman, 1990).

Forgé par des informaticiens dans le domaine de la vision, le terme d' « analyse de scènes » fait référence à la résolution par un ordinateur du problème suivant : dans une photographie d'une scène de complexité normale, il arrive souvent que les parties visibles d'un objet soient discontinues, l'objet ayant été masqué dans le champ de vision de l'appareil-photo par la présence d'un autre objet, situé entre l'objet photographié et l'appareil-photo (par exemple, Guzman, 1969). L' « analyse de scènes » désigne la stratégie par laquelle l'ordinateur regroupe l'ensemble des propriétés visibles — pourtours, textures de surface, couleurs, distances, etc. — d'un même objet. Ce processus permet de déterminer la forme globale et les propriétés correctes d'un objet. Par analogie, l'analyse de scènes *auditives* est le processus réunissant en une unité perceptive l'ensemble des signaux provenant, dans une période de temps donnée, d'une seule source sonore de l'environnement. Ce chapitre décrira les méthodes adoptées par le système auditif ainsi que certaines des recherches qui les ont mises à jour.

L'ANALYSE DE SCÈNES EN AUDITION

La plupart des individus à qui l'on décrit le problème des mixtures affirment le résoudre « en portant simplement attention à l'un des sons à la fois ». En d'autres termes, les composants d'un même son forment un tout cohérent à partir duquel le processus attentionnel effectue une sélection. On rappellera cependant ici que l'oreille ne perçoit qu'un pattern de changements temporels de la pression acoustique : rien dans la représentation graphique de la forme d'onde d'une mixture sonore ne permet d'étiqueter le son comme une mixture ou de savoir comment le décomposer.

On sait que le système auditif humain effectue le premier stade de l'analyse sonore dans la cochlée : le son y est décomposé en patterns neuraux distincts, représentant approximativement les différentes fréquences du signal (cf. Moore et Patterson, 1986, pour des considérations sur les limites de l'aptitude de l'oreille à séparer des fréquences très proches). L'analyse des fréquences composant le son s'utilise aussi dans l'étude de la compréhension de l'audition. Les résultats sont présentés sur un sonagramme, l'abscisse symbolisant le temps et l'ordonnée, la fréquence. Les points foncés indiquent l'intensité du son à un moment et à une fréquence particuliers. Comme en témoigne le sonagramme, l'utilité de la décomposition du signal en fréquences est réduite.

La figure 1 (à gauche) présente le sonagramme d'une mixture sonore. On pourrait supposer que ce genre de représentation permet d'identifier les sources sonores distinctes dans des mixtures. Ce serait le cas si les sources étaient des sons purs, de hauteur fixe et de fréquence bien distincte : chaque bande horizontale représenterait alors un son de l'environnement distinct qui se prolonge pendant une certaine période de temps. La figure 1 (à gauche) illustre un cas plus naturel : chaque son de l'environnement est composé de nombreuses fréquences variant dans le temps (il s'agit d'un mélange sonore composé d'une voix d'homme énonçant le mot « choux », d'une autre chantonnant et d'une musique de fond instrumentale étrangère aux autres sons). Un auditeur n'éprouvera aucune difficulté à entendre le mot « choux » dans l'enregistrement représenté par ce sonagramme. La figure 1 (à droite) montre le mot prononcé isolément. C'est lui qui doit être extrait du mélange sonore. C'est faisable visuellement après avoir vu le pattern isolé. Des auditeurs y parviennent aussi, même

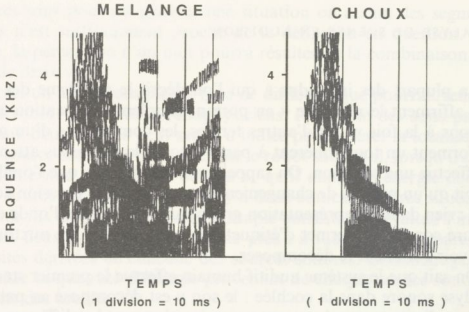

Figure 1. — Sonagramme d'une mixture sonore (à gauche) et d'un des composants de l'ensemble : le mot « choux » (à droite)

s'ils ignorent quel est le mot imbriqué dans le mélange sonore. Le problème auquel ils sont confrontés est illustré dans le sonagramme : dans le mélange sonore, les bandes représentant les composants du mot s'entrelacent avec celles qui représentent les composants des autres sons et les recouvrent partiellement. Une seule et même bande peut représenter la somme de deux ou de plusieurs composants de fréquences identiques, ou quasi identiques, dérivés de sons différents. Comme l'illustre la figure 1 (à gauche), la discrimination des fréquences par le système nerveux et leur répartition temporelle ne fournissent pas des touts cohérents pouvant être sélectionnés par l'attention. De toute évidence, la décomposition spectrale du signal ne livre pas les sources sonores distinctes mentionnées par les auditeurs. L'objectif de l'analyse de scènes auditives est de les découvrir.

A mon sens, trois processus sont mis en œuvre dans la décomposition des mélanges auditifs par l'auditeur humain. Le premier est l'activation purement automatique de schémas appris. Nous savons tous qu'on croit parfois entendre son nom dans un environnement bruyant, par exemple à un coin de rue en ville. L'intervention fortuite de plusieurs sons peut, semble-t-il, activer la représentation mentale du son de notre nom. Il est vraisemblable que cette hypersensibilité et cette activation automatique découlent de l'habitude que nous avons

d'entendre notre nom : son schéma se trouvant dans un état de haute potentialité, l'activation se produirait chaque fois que le son perçu correspond, même approximativement, à la définition acoustique. Pour autant que le pattern sonore reconnu par le schéma n'ait pas été complètement dénaturé, cette méthode d'analyse permettrait donc d'analyser les mixtures.

Le deuxième processus est l'utilisation volontaire de schémas : par exemple, lorsque l'on cherche à entendre son nom lors de l'appel des rendez-vous dans un bureau où règne une grande animation. Le fait de « chercher à entendre » atteste l'implication de l'attention volontaire. On notera cependant que le schéma du nom est aussi impliqué, même s'il ne s'agit pas ici d'une activité automatique. D'une manière générale, chaque fois que l'écoute cherche à se focaliser sur certains sons ou certaines classes de sons, elle utilise un critère de reconnaissance des cibles. Je qualifie ce critère de schéma puisqu'il est une représentation mentale d'un ensemble particulier de caractéristiques.

Ces deux méthodes — la reconnaissance automatique et la reconnaissance volontaire — supposent que les schémas (la connaissance de la structure de sons particuliers ou de classes de sons importants pour un individu) *ont été formés* par une écoute *antérieure*. Si les méthodes fondées sur un schéma étaient les seules mises en œuvre dans la décomposition des mixtures sonores, nous éprouverions des difficultés à former des schémas des sons importants avant de les avoir fréquemment rencontrés isolément. Il faudrait alors disposer de méthodes générales de partition d'une mixture sonore en sources acoustiques distinctes, utilisables préalablement à toute connaissance spécifique des sons importants de l'environnement.

Il semble que nous disposions de telles méthodes. J'y ai fait référence collectivement sous le terme d' « analyse primitive de scènes auditives » (Bregman, 1990, p. 38). En les qualifiant de primitives, je sous-entends qu'elles ne sont pas liées à la connaissance de types spécifiques de sons, comme les voix, les instruments de musique ou les machines, mais dépendent de propriétés acoustiques générales utilisables dans la décomposition de n'importe quel type de mélange sonore.

L'utilisation de régularités acoustiques générales

Il existe certaines relations entre les propriétés acoustiques d'un son. Elles pourront être mises en œuvre dans l'analyse de scènes, même si l'auditeur n'est pas familiarisé avec le signal. Ces relations

découlent des propriétés générales des sons et ne sont pas spécifiques des voix, de la musique, des sons d'animaux ou de toute autre classe particulière de sons.

On trouve un exemple de ces propriétés générales dans la nature *harmonique* de nombre de sons de notre environnement. C'est-à-dire que les fréquences de leurs composants sont (approximativement) toutes des multiples d'une fréquence fondamentale. Lorsqu'on imprime une force à certains types de corps, celle-ci déclenche en eux un mouvement oscillatoire. Les subdivisions de ces corps se mettent également à osciller. Des subdivisions, constituant des moitiés, des tiers, des quarts (et ainsi de suite) du corps entier, forment des oscillations dont les fréquences sont respectivement le double, le triple, le quadruple (et ainsi de suite) de la fréquence d'oscillation du corps entier. Les vocalisations animales (y compris les cordes vocales humaines), les instruments à cordes et certaines parties de n'importe quelle machine en sont des exemples.

Le système auditif peut utiliser une stratégie exploitant cette régularité. Supposons un signal dont l'analyse des fréquences indique que tous les composants de sons purs sont des multiples d'une fréquence fondamentale commune. Est-il composé de plusieurs sons ou s'agit-il d'un son unique ? En quoi serait-il faux de supposer qu'on a affaire à plusieurs sons, ayant chacun un ensemble différent d'harmoniques qui seraient tous des multiples d'une fréquence fondamentale commune ? De toute évidence, un tel cas est fort improbable. Il serait donc préférable de supposer que le signal est composé d'un son unique générant des harmoniques. Voici un autre exemple : que doit parier le système auditif lorsqu'il détecte deux sous-ensembles de composants dont les membres sont tous des multiples d'une fréquence fondamentale différente ? De toute évidence, il devrait conclure que le signal est un mélange de deux sons harmoniques. L'harmonicité est donc la mise en œuvre d'une régularité générale de l'univers pour statuer sur le nombre des sons entendus.

Aucune régularité n'est cependant totalement fiable. Il n'est pas impossible que les harmoniques de deux sons puissent brièvement entretenir une relation apparemment régulière. Mais, selon toute probabilité, on découvrira qu'ils émanaient de localisations spatiales différentes ou qu'ils ont démarré et se sont arrêtés à des moments différents. Puisque les fréquences des composants d'un signal acoustique unique tendent à provenir d'une même localisation et à démarrer et s'arrêter *grosso modo* au même moment, des différences dans ces propriétés empêcheront de considérer des composants fortuitement har-

moniques comme des parties d'un seul et même son. Pour parvenir à la bonne réponse, il faudra donc exploiter quantité de régularités simultanément. Selon Shepard (1981), le système perceptif des animaux aurait appris à tirer parti des régularités du monde dans lequel ils ont évolué. Autrement dit, les propriétés de leur système perceptif se seraient réglées sur les régularités du monde physique au fur et à mesure de leur évolution. On appelle cette équivalence la « complémentarité psychophysique ». Si l'hypothèse de Shepard est exacte, pour découvrir les lois de l'organisation auditive, il faut rechercher les relations systématiques entre les composants émanant de différentes sources sonores de l'environnement du son perçu. Ensuite, on vérifiera expérimentalement si le système auditif humain exploite ces relations dans la décomposition de mixtures.

LES EFFETS DE L'ANALYSE DE SCÈNES AUDITIVES SUR LA PERCEPTION

L'organisation des données sensorielles à laquelle procède l'analyse de scènes auditives affecte des aspects de la perception auditive qui dépassent le cadre traditionnel de l'organisation perceptive. On pense souvent que l'organisation est un groupement de propriétés auditives grossières, *consécutif* à leur création par la perception. Des propriétés apparemment aussi élémentaires que la sonie sont cependant affectées par l'organisation.

Warren (1982) décrit le concept de « continuité homophonique » (cf. aussi Warren 1984 ; Warren, Obusek et Ackroff, 1972) : supposons que, après s'être maintenu pendant quelques secondes à une intensité déterminée, un son devienne soudainement plus fort et se maintienne brièvement à cette nouvelle intensité avant de retourner au niveau original où il reste quelque temps. Si la période de plus forte intensité est courte, elle ne sera pas perçue comme une transformation du signal original, mais interprétée comme l'intervention momentanée d'un deuxième son identique au premier. On perçoit le premier signal se prolongeant à un niveau identique derrière le son ajouté. Lorsqu'elle est au niveau le plus élevé, l'intensité n'est pas interprétée comme un seul son fort, mais comme un ensemble de deux sons plus faibles. En d'autres termes, l'information relative à l'intensité est partagée entre deux sons perçus. Par ailleurs, les expériences

menées à partir de ce stimulus montrent que cette interprétation n'a pas lieu si l'augmentation et la baisse d'intensité ne sont pas soudaines (Bregman, 1991). On perçoit une transformation de la sonie du son original : au niveau d'intensité le plus élevé, l'intégralité de l'information est utilisée pour dériver la sonie d'un son unique. En résumé, les processus d'organisation décideront de ce qui est entendu : un son fort ou deux sons d'intensité plus faible.

En variant cet exemple, on peut démontrer que l'organisation perceptive modifie aussi la perception de la localisation spatiale d'un son (Bregman, 1991). Supposons que le signal que l'on vient de décrire est présenté uniquement à l'oreille droite. Simultanément, l'oreille gauche entend un son à tous égards identique au signal original, sauf qu'il se maintient constamment à une intensité plus faible. Au cours de la phase initiale, lorsque les deux oreilles reçoivent le même signal d'intensité faible, un seul son de hauteur fixe est entendu au milieu. Ensuite, lorsque le signal de l'oreille droite augmente subitement en intensité, reste brièvement à ce niveau, puis redescend, on perçoit toujours le son original au milieu, mais il est brièvement accompagné d'un son à l'extrême droite. On identifie donc deux sons, provenant chacun d'une localisation distincte.

Supposons, maintenant, que l'augmentation et la diminution de l'intensité à l'oreille droite soient moins subites. On entendra alors le son se déplacer vers la droite puis revenir au centre, la perception spatiale évoluant en fonction des modifications de l'équilibre des intensités aux deux oreilles. On identifie donc un seul son, provenant d'une localisation unique.

Que la transformation soit soudaine ou progressive, l'équilibre de l'intensité entre les deux oreilles est identique au cours de la phase centrale (lorsque l'intensité du son est plus forte à l'une des deux oreilles). Si l'évolution temporelle du son n'entraînait pas de modification de la localisation perçue, le percept serait identique dans les deux cas. Mais il n'en est rien. La localisation perçue n'est donc pas celle d'un « son » dans l'abstrait, mais celle d'un son particulier. L'ensemble des localisations perçues est fonction du nombre des sons créés par l'organisation perceptive. Sans les modifier, il est donc impossible d'appliquer aux environnements acoustiques complexes les règles simples régissant la perception spatiale, telles qu'elles ont été mises à jour par la psychophysique classique en testant des auditeurs dans des environnements simples et calmes.

L'analyse de scènes auditives influence aussi la perception d'un son de timbre riche ou celle d'un certain nombre de sons de timbre

plus pur (Bregman et Pinker, 1978). Elle peut modifier l'identité perçue d'un son parlé par la ségrégation de composants acoustiques : n'étant plus perçus comme des éléments constitutifs du son en question, ces composants sont affectés à un autre signal perçu (Ciocca et Bregman, 1989 ; Darwin, 1984). Elle permet aussi de déterminer si deux mélodies mélangées sont entendues séparément ou si elles donnent naissance à une nouvelle mélodie formée de toutes les notes (Dowling, 1973). Elle permet enfin de déterminer si, dans un ensemble, on identifie les propriétés particulières des instruments ou si celles-ci fusionnent en un timbre global (Bregman, 1990, chap. 5). En résumé, l'analyse de scènes modifie n'importe quelle expérience auditive dans des environnements naturels.

La plupart des recherches sur l'analyse de scènes auditives ont été examinées ailleurs (Bregman, 1990). Je cherche ici à faire comprendre la base de cette approche. La présentation s'organise autour des différentes régularités de l'environnement exploitées par le système auditif.

Régularité 1. Il est extrêmement rare que des sons n'ayant aucun rapport entre eux démarrent et s'arrêtent exactement au même moment.

Les composants acoustiques dérivés de sons distincts de l'environnement tendent à ne pas démarrer ni s'arrêter au même moment. L'un est donc susceptible d'être déjà actif au moment où un autre prend naissance. Le système auditif exploite cette vérité générale dans la stratégie que je qualifie d' « ancien-plus-nouveau » : lorsqu'un spectre devient subitement plus complexe, tout en conservant ses composants fréquentiels initiaux, le système nerveux l'interprète comme une continuation d'un signal ancien auquel s'ajoute un signal nouveau. Le son ancien continue à être entendu pendant un court moment en même temps que le nouveau. Si le spectre redevient simple — autrement dit, s'il n'est constitué que des composants du son ancien —, la perception de la présence continue du son ancien est renforcée. Les qualités perçues du son ajouté sont dérivées des composants du spectre complexe après soustraction des composants du premier spectre, plus simple.

Les divers phénomènes de continuité perçue (cf. Warren, 1984) découlent de cette stratégie. L'exemple le plus connu est un son long et pur présenté en alternance avec une brève bouffée de bruits plus forts stimulant des canaux neuraux comprenant celui qui a été stimulé par le son. On entend le son se prolonger à travers le bruit.

Autrement dit, on perçoit une présence continue du signal. Le système auditif n'a donc relevé aucune incompatibilité entre l'activité neurale et la continuation du son au cours de l'explosion sonore. Ceci le conduit à un percept où le son se prolonge derrière le bruit. Il est vraisemblable que le rejet de l'interprétation d'une alternance — un son se muant en bruit — découle de l'exploitation par le système auditif d'une autre régularité de l'environnement, qui sera examinée plus loin : les propriétés dérivées d'un seul son de hauteur fixe ne tendent pas à se transformer subitement.

L'exemple précédent ne permet pas de déterminer aisément si la partie de l'activité neurale interprétée comme la continuation du son joue un rôle dans la perception du bruit : même si l'on retranche la bande de fréquence étroite assignée au son, on ne percevra pas de modification de la bande de bruits large. On a donc cherché un meilleur stimulus pour analyser le partage de l'information totale fournie par le bruit : il s'agit d'un signal dans lequel une bande de bruits, disons de 0 à 1 kHz (« kHz » étant le symbole de « milliers de cycles par seconde »), alterne avec une bande plus large (disons de 0 à 2 kHz) et est de plus longue durée qu'elle. Les composants communs aux deux sons (s'échelonnant de 0 à 1 kHz) sont d'intensité égale. On entend le bruit de la bande étroite se prolonger à travers celui de la bande large (Warren, 1982, chap. 2). On a en outre l'impression que les composants utilisés par le système auditif pour créer le percept du bruit continu de la bande étroite n'entrent pas dans la composition du son court intermittent. Les composants de 0 à 1 kHz ont donc été extraits et le bruit « ajouté » est perçu comme étant de qualité identique à un bruit de 1 à 2 kHz.

Darwin (1984) fournit un exemple de la stratégie « ancien-plus-nouveau » dans le domaine de la perception de la parole. Grâce à la position d'un certain nombre de formants (les sommets d'intensité) à l'intérieur du spectre, on peut distinguer chacune des voyelles d'une langue. Chaque formant correspond à une plus forte intensité d'un certain nombre d'harmoniques. Darwin rallonge l'un des harmoniques du formant inférieur d'une voyelle de façon à ce qu'il démarre avant la voyelle et se prolonge tout au long de la durée de celle-ci. Dans ce type de stimulus, l'harmonique correspond au son « ancien » et la voyelle au son « nouveau ». La voyelle ressemble davantage comme une autre voyelle, puisque l'harmonique entendu isolément se voit assigner une identité en tant que son distinct. Ensuite, lors de l'association de l'harmonique au reste de la voyelle, l'ensemble de la stimulation neurale activée par la voyelle est décomposé. Une partie

de celle-ci est interprétée comme la continuation du son (ancien). Après soustraction de cette première partie, la partie restante est interprétée comme la voyelle. Les données sensorielles restantes montrent que la position du sommet du spectre diffère de celle que lui assignerait l'ensemble du corpus des données de la voyelle initiale : ce déplacement modifie l'identité de la voyelle perçue.

Régularité 2. Progression de la transformation :

a) Les propriétés d'un son isolé tendent à se modifier de façon continue et lentement.
b) Les propriétés d'une séquence de sons issue de la même source tendent à se modifier lentement.

La régularité de l'environnement portant sur la ressemblance séquentielle de sons provenant d'une source identique comporte deux formes. La première s'apparente à un *signal sonore isolé* étendu dans le temps. Par exemple (à des échelles temporelles différentes) : une note de violon isolée, la prononciation d'une syllabe isolée, le rugissement d'un lion ou le bruit continu d'un moteur. Les propriétés d'un son issu de ce type de signal tendent à se transformer de manière continue plutôt qu'abrupte pendant toute la durée du signal.

La deuxième forme de cette régularité séquentielle concerne une *succession de sons* provenant d'une même source. Les exemples sont nombreux : une succession de pas, une série de pépiements dans l'appel d'un oiseau ou d'une grenouille, une série de coups de bec d'un pic, et ainsi de suite. Des sons dérivés successivement de la même source acoustique tendent à se ressembler et à ne présenter que des modifications progressives des propriétés de la série.

Parfois, le caractère répétitif ne découle pas du son en lui-même, mais de sa perception intermittente : dans un mélange de sons, dont chacun croît et décroît en intensité, le système auditif peut être soumis à une exposition répétée aux propriétés particulières de l'un d'entre eux. Les échantillons spectraux furtivement saisis au cours de ces « aperçus » successifs sont susceptibles de se ressembler.

La transformation des sons est un élément important de la ressemblance séquentielle. Puisque les modifications tendent à être progressives, plus les échantillons d'un même signal acoustique ont été prélevés à des moments temporellement proches, plus leur ressemblance est étroite.

Deux règles suivies par le système auditif résument ces données relatives à la modification séquentielle des sons de l'environnement.

La première est la règle de la « transformation soudaine » : le système auditif traite une transformation soudaine des propriétés acoustiques comme le début d'un nouveau signal. On en a donné des exemples plus haut, notamment la continuité homophonique et l'absence de modification du son perçu par une oreille lorsqu'on élève subitement l'intensité envoyée à l'autre oreille. Dans ces deux cas, la soudaineté de la transformation conduit à l'interprétation d'un son ajouté. Ceci fixe le cadre dans lequel la stratégie « ancien-plus-nouveau » décompose le spectre en composants appartenant au son ancien ou au son ajouté.

Pour utiliser cette stratégie, il faut donc définir la soudaineté. Il n'existe cependant pas de frontière précise entre les transformations progressives et les transformations soudaines, comme en témoigne une étude récente non publiée de Kim et Bregman : une séquence rapide de quatre sons d'une seconde est présentée à des sujets dont la tâche consiste à juger l'ordre de démarrage des sons. Les sons se recouvrent largement : par exemple, un intervalle aussi bref que 0,15 s peut séparer l'attaque de chaque son successif. L'augmentation soudaine de l'intensité d'un son quelconque facilite son extraction de la mixture. Lorsque le temps d'attaque dure 0,04 s, les sons sonnent distinctement, et leur ordre est passablement bien jugé. En revanche, lorsque l'attaque dure 0,64 s, les sons fusionnent en un magma impénétrable, et tout jugement d'ordre devient quasi impossible. Le passage du progressif au soudain a cependant été progressif : un temps d'attaque intermédiaire, de 0,16 s, fournit un résultat intermédiaire. La soudaineté de la transformation spectrale agit comme n'importe quel autre indice dans l'analyse de scènes. Plus elle est importante, plus elle affecte le groupement.

Cet éventail des durées (de 0,04 à 0,64 s), qualifié de large gamme de soudaineté, pourrait n'être valide que pour les transformations d'intensité. On ne dispose pas de quantification précise des valeurs tenues pour soudaines et progressives dans d'autres caractéristiques du son, comme la localisation spatiale, le timbre ou la hauteur.

Il arrive que le système auditif ne décèle pas une équivalence suffisamment étroite entre les spectres précédant et suivant la transformation soudaine pour imposer l'interprétation de l'ancien-plus-nouveau : au lieu d'interpréter le changement comme l'ajout d'un son à la mixture, ou comme une modification de qualité de la continuation d'un son isolé, il conclut alors au remplacement du premier son par un second. On en trouve un exemple en perception de la

parole chez Darwin et Bethell-Fox (1977). Il faut souligner ici que la fréquence fondamentale et celle des formants peuvent être variées indépendamment l'une de l'autre. La fréquence fondamentale est celle dont tous les composants spectraux sont des multiples ou harmoniques. Les formants sont les parties du spectre dont les harmoniques ont été amplifiés en intensité. Ces chercheurs synthétisent un échantillon d'un énoncé continu, consistant en transitions continues de formants (les modifications progressives de la position des sommets spectraux) en amont et en aval de deux voyelles. Les voyelles sont représentées par les fréquences des formants restant inchangés pendant 0,06 s. Au milieu de chaque transition, Darwin et Bethell-Fox introduisent une transformation subite de la fréquence fondamentale du locuteur, de manière à abaisser la fréquence de l'une des voyelles, 101 Hz, et à hausser celle de l'autre, 178 Hz (« Hz » signifiant « cycles par seconde »). Ils obtiennent une ségrégation de deux locuteurs apparemment distincts, l'un parlant à une hauteur plus basse et l'autre à une hauteur plus élevée. En outre, malgré la présence d'une transformation continue de la fréquence des formants, la perception tend à isoler les parties de la transition qui précèdent et suivent le changement de hauteur abrupt : on a l'impression que chacune des voix est silencieuse pendant que l'autre parle. Enfin, les syllabes construites en cours de perception ne tendent *pas* à se fonder sur les patterns de formant qui restent constants malgré la modification de la fréquence fondamentale. Tout se passe comme si le système auditif cherche à éviter de commettre une erreur en créant des syllabes étrangères à partir des relations entre les sons des deux locuteurs distincts.

Le « groupement par similarité » est la deuxième règle exploitant la probabilité de ressemblance séquentielle entre des sons issus d'une même source. Pour le moment, nous ignorons les divers aspects que peut revêtir la similitude entre des sons. On sait cependant que des similitudes au niveau des fréquences des sons purs, de la localisation spatiale ou du contenu spectral et de la fréquence fondamentale des sons complexes affectent le groupement (Bregman, 1990, chap. 2). Nous ne savons pas non plus comment mesurer la similarité. S'agissant de fréquences de sons purs, la différence qui affecte le groupement est la différence entre les logarithmes de leurs fréquences. Mais la recherche n'a pas mis au jour de quantifications précises du degré de similarité dans les autres dimensions : elle a seulement montré l'effet de ces similitudes sur le groupement.

En un certain sens, la similarité a déjà été examinée ici : on peut

soutenir que la discontinuité est simplement l'absence de similarité séquentielle. Si l'on mentionne explicitement la règle de similarité, c'est en fonction des sons présentés par échantillons ou en épisodes discrets : une succession de pas ou une série de perceptions furtives d'un des sons d'une mixture à chaque diminution de l'intensité des autres signaux. Ces échantillons successifs sont modélisés en laboratoire par des successions de sons, des bouffées de bruits, de glissements de la fréquence d'un seul son, etc.

La règle a pour effet de prélever les sons qui présentent des propriétés semblables, de les réunir en groupes perceptifs et d'effectuer la ségrégation de ces groupes. Chaque groupe ainsi formé est considéré par le système auditif comme provenant d'une source sonore de l'environnement. C'est pourquoi les processus de reconnaissance de patterns tendent à traiter les sons groupés comme des « ensembles » de traits à l'intérieur desquels il y a lieu de rechercher des patterns familiers.

Le phénomène de la formation de flux auditifs illustre cette règle (Bregman et Campbell, 1971 ; Van Noorden, 1975, 1982). Prenons deux ensembles sonores, l'un composé de sons aigus et l'autre de sons graves, dont les registres aigu et grave sont séparés par un intervalle de hauteur important, alors que celui-ci est minime entre les sons d'un même ensemble. Supposons que nous alternions les sons aigus (A) et graves (G) (par exemple, AGAGAG...) pour construire une longue séquence que nous présenterons ensuite à des auditeurs. Pour mémoire, dans l'univers en général, plus deux échantillons issus de la même source acoustique sont proches temporellement, plus leurs propriétés sont susceptibles d'être semblables. La stratégie de « groupement par similarité » tient compte de cette régularité. Aussi longtemps que la succession des sons est lente (autrement dit, aussi longtemps que les échantillons restent éloignés temporellement), le système auditif interprète la totalité des ensembles sonores comme provenant d'une source unique. En revanche, dès que la séquence s'accélère, deux « ensembles » perceptifs, ou flux auditifs, sont formés : on perçoit deux sources distinctes approximativement actives en même temps. Les patterns formés d'une combinaison de sons aigus et de sons graves disparaissent au profit de patterns constitués soit des seuls sons aigus, soit des seuls sons graves : par exemple, on entend deux mélodies, l'une formée des notes aiguës et l'autre des notes graves. En outre, si la séquence combinatoire originale est irrégulière, on entend des patterns temporels distincts dans les registres aigu et grave.

Par exemple, la séquence de sons aigus (A) et graves (G) que voici :

... A A G A G G G A G A A G G A G...

formera les flux suivants :

... A A – A – – A – A A – – A – ...
et ... – – G – G G – G – – G G – G...

Dans ces diagrammes de lettres, les tirets symbolisent la présence de silences dans chacun des flux. Ceux-ci proviennent de l'absence des sons affectés à l'autre flux par le groupement perceptif. Ainsi des patterns temporels, comprenant des rythmes (dans les cas de séquences répétitives), peuvent-ils être créés ou modifiés par la ségrégation en flux auditifs.

Le groupement est sensible à la différence de fréquences entre les sons aigus et graves : la ségrégation est plus forte lorsqu'on augmente l'intervalle de fréquence. Une diminution de cette différence entre les fréquences peut compenser l'accélération de la séquence. Plus la différence de fréquences est minime, plus il faut accélérer la séquence pour produire la ségrégation (Van Noorden, 1975).

On peut penser que l'influence du tempo sur la ségrégation provient de la sensibilité au degré d'alternance entre les sons aigus et graves. Plus l'on passe rapidement de l'un à l'autre, moins la séquence a de chance d'être interprétée comme provenant d'une même source de l'environnement. Cette propriété du processus perceptif s'explique par une régularité de notre environnement : une série de sons émanant de la *même* source acoustique tend à ne se transformer que lentement.

La ségrégation des sons de deux classes (du moins lorsque les classes sont définies par un registre de fréquences) a un aspect cumulatif. Même à un tempo d'exécution rapide, les premiers sons (aigus et graves) sont intégrés en un flux unique. La ségrégation intervient lorsque le système auditif continue à percevoir deux populations de sons distinctes. Il semble que la tendance à la ségrégation nécessite au moins quatre secondes pour se manifester. Il lui faut de même au moins quatre secondes pour se dissiper. Une interruption de deux secondes ne supprime pas la tendance à la ségrégation : lorsqu'on redémarre la séquence, la ségrégation intervient même plus rapidement que la première fois (Bregman, 1978). Ce principe auditif correspond, semble-t-il, à une donnée de l'environnement : le fait d'avoir entendu un son juste un instant auparavant est le meilleur prédicteur de son occurrence. Selon toute vraisemblance, le système auditif conserve un « créneau vide » à la disposition du retour du son.

On peut penser que les jugements perceptifs simples sont aussi affectés par la ségrégation des composants en registres de fréquences différents. Il est plus difficile de juger la distance temporelle exacte entre deux sons, dont l'un est de fréquence plus élevée que l'autre, lorsqu'on augmente l'intervalle de hauteur (cf. Bregman, 1990, p. 159-163, pour un compte rendu de cette recherche).

LES BASES MULTIPLES DE LA SÉGRÉGATION

La fréquence des sons n'est pas la seule caractéristique qui affecte leur organisation séquentielle. Des sons peuvent être groupés sur la base de similitudes de leurs positions spatiales (décrites dans Bregman, 1990, p. 73-83). D'autres facteurs interviennent même pour les sons complexes. Lorsque les sons comportent un grand nombre d'harmoniques, comme c'est le cas des sons vocaux et musicaux, le groupement peut tout autant dépendre de similitudes de leur fréquence fondamentale que de leur timbre, si par timbre l'on entend le poids relatif des harmoniques dans le spectre (Bregman, Liao et Levitan, 1990 ; Singh, 1987). Des sons dont les formes d'onde se répètent par cycle se séparent souvent des bruits (Dannenbring et Bregman, 1976). Des sons qui glissent d'une fréquence à une autre s'assemblent mieux séquentiellement lorsque leurs pentes et leurs registres sont identiques (Steiger et Bregman, 1981).

Ces divers types de différences entrent en compétition ou se renforcent dans l'établissement du groupement. Lorsque divers facteurs favorisent des groupements contradictoires des sons, le vainqueur est le groupement obéissant au plus grand nombre de facteurs ou celui qui répond aux facteurs privilégiés par le système auditif (Bregman, 1990, p. 165-171, 218, 335).

Toutes les différences acoustiques n'ont donc pas une importance égale dans l'établissement du groupement. Hartmann et Johnson (1991) demandent à leurs sujets de reconnaître des paires de mélodies dont ils alternent les notes pour construire une nouvelle séquence de sons. Dans chacune des séquences, les notes des deux mélodies sont agencées de manière à n'affecter qu'une seule caractéristique acoustique : la ségrégation des deux mélodies est nettement facilitée lorsqu'elles sont envoyées dans des oreilles différentes, lorsqu'on transpose les notes d'une des mélodies à l'octave supérieure ou lorsqu'on

La psychologie cognitive de l'audition étudie les processus permettant au cerveau de traiter avec rapidité et efficacité les structures sonores de notre environnement. Elle s'affirme aujourd'hui comme un champ de recherche à part entière, encore mal connu de l'ensemble de la communauté scientifique et d'autres secteurs d'activités directement concernés par l'environnement sonore : secteurs industriels et technologiques, artistiques. L'objectif des journées organisées par le groupe Audition de la Société française d'acoustique en 1991 à l'IRCAM était de présenter de façon didactique l'ensemble des thèmes qui y sont abordés. Des spécialistes nord-américains et européens de l'organisation perceptive, de la mémoire, de l'attention, de la psychologie de la musique, de la neuropsychologie et de la psychologie du développement y ont présenté les principales recherches effectuées dans leurs domaines respectifs. *Penser les sons* rassemble ces contributions et fait ainsi le point sur les connaissances actuelles relatives à la cognition auditive. Il s'agit d'une introduction aisée à cette problématique pour les étudiants et chercheurs des différentes disciplines des sciences cognitives et, plus largement, pour tout lecteur intéressé par l'environnement sonore.

Stephen McAdams est chargé de recherches au CNRS, membre du Laboratoire de psychologie expérimentale de l'Université René Descartes (Paris) et responsable du projet « Processus perceptifs et cognitifs » à l'IRCAM.
Emmanuel Bigand est maître de conférences à l'Université de Bourgogne (Dijon) et membre du Laboratoire d'études des acquisitions et du développement, associé au CNRS.

198 FF

22409977/3/94