ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES

THÈSE

pour l'obtention du grade de

DOCTEUR DE L'EHESS

Discipline : Sciences Cognitives, mention Psychologie Cognitive

présentée et soutenue publiquement

par

Caroline BEY

le 15 décembre 1999

RECONNAISSANCE DE MÉLODIES INTERCALÉES ET FORMATION DES FLUX AUDITIFS : ANALYSE FONCTIONNELLE ET EXPLORATION NEUROPSYCHOLOGIQUE

JURY

M. Stephen McADAMS, Directeur de thèse

M. Michel IMBERT, Président

M. Laurent DEMANY, Rapporteur

M. Robert ZATORRE, Rapporteur

Mme. Cécile MARIN, Examinateur

M. Christophe MICHEYL, Examinateur

Remerciements

Je tiens à remercier tout particulièrement mon directeur de thèse, Stephen McAdams, pour m'avoir communiqué sa passion de la recherche en particulier sur ce thème de l'organisation auditive. Son enthousiasme et son dynamisme ont rythmé mes années de thèse. Je le remercie de ses encouragements et de ses conseils qui m'ont été très précieux et je souhaite notre collaboration longue sur ce thème.

Je remercie Cécile Marin pour sa gentillesse et sa générosité. Je lui suis très reconnaissante de m'avoir confié les cours de perception auditive.

Je remercie très chaleureusement Clarisse Baruch pour m'avoir confié des travaux dirigés et m'avoir ainsi accompagné dans mes débuts d'enseignante. Je la remercie pour son soutien et ses conseils.

Je remercie Jean-Didier Bagot de m'avoir confié les travaux dirigés portant sur la perception à l'Université Paris V, ses enseignements m'ont beaucoup appris.

Je remercie Serge Nicolas de m'avoir permis d'enseigner la mémoire, avec qui il a été très sympathique de travailler.

Je remercie Anne-Christine Lhadky pour l'accueil chaleureux qu'elle m'a réservé chaque année à l'ISEN.

Je remercie Christian Lorenzi de m'avoir proposé de collaborer avec lui sur l'étude du patient L.B. Je le remercie vivement pour sa disponibilité et son aide.

Je remercie Séverine Samson pour m'avoir initié à la neuropsychologie et m'avoir permis de mener cette étude auprès de patients cérébro-lésés à l'hôpital de la Salpêtrière.

Je remercie Isabelle Peretz de son accueil à l'Université de Montréal et de l'intérêt qu'elle a manifesté pour mes recherches.

Je remercie Albert S. Bregman des suggestions qu'il a pu faire sur ce travail et de m'accueillir dans son laboratoire pour pouvoir y donner suite.

Je remercie mes amis sans qui cette aventure doctorale n'aurait pas pris cette dimension, Didine, Anne, Bennett, Olive, Patrick, Sophie, et Vince.

Je remercie également mes collègues de la 426 pour les moments sympathiques que nous avons partagés, Alix, Anne M., Céline, Dorine, Elsa, Juliette, Olivier C., Sophie S., Véronique et Xavier.

Je remercie les membres des deux équipes qui m'ont accueillis et donnés des conditions de travail propices à l'élaboration de cette thèse, l'équipe Perception Auditive du laboratoire de Psychologie Expérimentale de l'Université Paris V et l'équipe Perception et Cognition Musicale de l'IRCAM.

Je remercie Juan Segui de m'avoir accueilli au sein de son laboratoire pour réaliser cette thèse, ainsi que tous les membres du Laboratoire de Psychologie Expérimentale que j'ai pu côtoyés durant ces années.

Je remercie L. Demany, M. Imbert, C. Marin, C. Micheyl et R. Zatorre d'avoir accepté de faire partie du jury.

Je remercie toutes les personnes qui ont gentiment participé aux expériences que j'ai réalisées.

Enfin, je tiens à remercier Christophe pour son soutien, son optimisme et l'intérêt qu'il a manifesté pour mon travail. Je lui dédie ce document avec toute mon affection.

Résumé

Les processus impliqués dans la construction des flux auditifs ont été examinés autour de trois questions, définissant les trois parties expérimentales de la thèse : Comment mesurer ces processus ? Quels sont les mécanismes qui sous-tendent cette analyse de la scène auditive ? Et quelles sont leurs bases neurales ? Un paradigme de post-reconnaissance de mélodies intercalées a été élaboré afin de mesurer de façon objective les mécanismes impliqués dans l'analyse primaire de la scène auditive. Cette tâche consiste à présenter à l'auditeur une mélodie cible intercalée à une séquence distractrice suivie d'une mélodie test, et à lui demander de déterminer si la mélodie test était ou non présente dans la séquence composite. La reconnaissance de la mélodie intercalée s'améliorait à mesure que la différence de hauteur ou la dissemblance de timbre la séparant des sons distracteurs augmentait. Le seuil de reconnaissance de mélodies intercalées était de 10 demi-tons. Néanmoins, pour une différence moyenne de 2-3 demi-tons, correspondant au seuil de fission d'une séquence cyclique, les auditeurs ne répondaient plus au hasard suggérant que la mélodie s'était scindée partiellement. Les performances de post-reconnaissance ont été comparées à celles obtenues dans une tâche où la mélodie à reconnaître était présentée avant la séquence à organiser. Lorsque les auditeurs entendaient préalablement la mélodie qu'ils devaient extraire du "mélange", leurs performances de reconnaissance étaient globalement augmentées mettant en évidence le rôle des connaissances dans la construction des flux auditifs. En outre, leurs critères de réponse avaient été modifiés. Enfin, une investigation auprès de patients cérébro-lésés a été initiée. Une dissociation dans l'aptitude à reconnaître une mélodie intercalée lorsque celle-ci était ou non présentée préalablement a été mise en évidence chez un patient présentant une lésion temporale droite.

Mots-clés : flux auditif, mélodies intercalées, fission auditive, analyse primaire de scène auditive, analyse guidée par les connaissances.

Summary

The processes implicated in the formation of auditory streams were examined with respect to three questions that defined the three experimental parts of the thesis: How can the processes be measured? What are the mechanisms that underly auditory scene analysis? And what are their neural bases? A paradigm using post-recognition of interleaved melodies was developed in order to measure objectively the mechanisms implicated in primitive auditory scene analysis. This task consists of presenting a target melody interleaved with a distractor sequence to a listener, followed subsequently by a test melody. Listeners were required to decide whether the test melody was present or not in the composite sequence. The recognition of the interleaved target melody improved with increasing pitch difference or timbral dissimilarity between the target and the distractor. The interleaved melody recognition threshold was 10 semitones. Nevertheless for a mean difference of 2-3 semitones which corresponds to the fission boundary of a cyclic sequence, listeners didn't respond at chance any more suggesting that the melody was partially streamed. Post-recognition performance was compared to that obtained in a task in which the melody to be recognized was presented before the sequence to be organized. When listeners heard the melody to be extracted from the "mixture" beforehand, their recognition performance was globally higher, providing evidence for the role of knowledge in the formation of auditory streams. Moreover, their response biases were changed. Finally, an investigation with brain-lesioned patients was initiated. A dissociation in the ability to recognize an interleaved target melody when the test melody was presented before or after the mixture was revealed in a patient with a right temporal lesion.

Keywords: auditory stream, interleaved melodies, auditory fission, primitive auditory scene analysis, knowledge-driven analysis.

Table des matières

INTRODUCTION 10

PARTIE THÉORIQUE 15

1 Les multiples bases de la ségrégation auditive 16
1.1 Mesure des processus de formation des flux auditifs 17
1.1.1 Fission auditive et formation des flux auditifs 17
1.1.2 Paradigmes et stimuli utilisés pour mesurer la formation des flux auditifs ... 18
1.1.3 Méthode directe versus indirecte 20
1.2 Les paramètres affectant le groupement et la ségrégation des événements auditifs 21
1.2.1 La différence de fréquence des sons purs 21
1.2.2 Ségrégation perceptive des sons complexes : Rôle de la hauteur et du timbre 32
1.2.3 La différence d'intensité et de localisation spatiale 44
1.3 Principe général 49
1.3.1 Transformation progressive des événements émis par une même source sonore 49
1.3.2 La complémentarité psychophysique 51

2 Phénomène périphérique ou central, attentif ou pré-attentif : Considérations théoriques. 58
2.1 Origine périphérique de la formation des flux : La théorie des "canaux" 59
2.1.1 La cohérence temporelle et la fission : succès et échec de la détection de mouvement de hauteur 59
2.1.2 Modélisation de la formation des flux auditifs sur la base de la stimulation ou non d'un même filtre 64
2.2 Origine centrale de la formation des flux 67
2.2.1 Arguments à l'encontre d'une théorie périphérique de la formation des flux auditifs 67
2.2.2. Analyse des Scènes Auditives 74
2.2.3 Attentif / Pré-attentif 78
2.3 Bases neurales des processus de formation des flux auditifs 82

2.3.1 Troubles centraux et déficit de la formation des flux auditifs 82
2.3.2 Exploration neuropsychologique 83
2.3.3 Apport théorique de l'étude des bases neurales des processus de formation des flux auditifs 84

3 Reconnaissance de mélodies en situation monophonique et polyphonique : Composantes cognitives 87
3.1 Reconnaissance de mélodies présentées isolément 88
3.1.1 Mémoire : Considérations théoriques générales 88
3.1.2 Mémoire des mélodies : Indices encodés et durée de stockage 90
3.2 Déficit de la reconnaissance de mélodies : Apport théorique des études de neuropsychologie cognitive 100
3.2.1 Déficit dans le traitement du contour et des intervalles 101
3.2.2 Déficit de la reconnaissance de mélodies familières et non familières 106
3.2.3 Analyse fonctionnelle de la reconnaissance de mélodies : Modèle théorique 112
3.3 Reconnaissance de mélodies en présence d'une autre séquence : Spécificités attentionnelles et mnésiques 115
3.3.1 Focalisation attentionnelle sélective 116
3.3.2 Encodage en mémoire de deux mélodies 118

PARTIE EXPÉRIMENTALE 123

4 Post-reconnaissance de mélodies intercalées et formation des flux auditifs : Effet d'une différence de hauteur moyenne et d'une dissemblance de timbre 124
4.1 Introduction 124
4.2 Méthode Générale 126
4.2.1 Une tâche de post-reconnaissance de mélodies intercalées 126
4.2.2 Stimuli 127
4.2.3 Construction des mélodies et des distractrices 127
4.2.4 Procédure 128
4.2.5 Matériel 129
4.3 Expérience 1 : Post-reconnaissance de mélodies intercalées et nombre de flux perçus 130
4.3.1 Introduction 130
4.3.2 Méthode 130
4.3.3 Résultats et discussion 132

4.4 Expérience 2 : Post-reconnaissance de mélodies intercalées sur la base d'une différence de hauteur 134
4.4.1 Introduction..... 134
4.4.2 Méthode 135
4.4.3. Résultats et discussion..... 136
4.5 Expérience 3 : Post-reconnaissance de mélodies intercalées sur la base d'une dissemblance de timbre..... 142
4.5.1 Introduction..... 142
4.5.2 Méthode 143
4.5.3 Résultats et discussion..... 147
4.6 Expérience 4 : Effet de la durée de l'intervalle de rétention sur la post-reconnaissance de mélodies intercalées..... 151
4.6.1 Introduction..... 151
4.6.2 Méthode 152
4.6.3 Résultats et discussion..... 153
4.7 Discussion générale..... 155
4.7.1 Résumé des principaux résultats 155
4.7.2 Post-reconnaissance de mélodies intercalées : Une mesure des processus primaires d'organisation mais pas seulement... 156
4.7.3 Performances de post-reconnaissance de mélodies intercalées et nombre de flux perçus, pas d'équivalence logique 157
4.8 Conclusion..... 158

5 Reconnaissance de mélodies intercalées avec ou sans présentation préalable de la mélodie : Mise en évidence des processus d'analyse de scène auditive guidée par les connaissances 159
5.1 Introduction..... 159
5.2 Expérience 5 : Reconnaissance de mélodies intercalées avec ou sans présentation préalable de la mélodie 161
5.2.1 Introduction..... 161
5.2.2 Méthode 161
5.2.3 Résultats et discussion..... 164
5.3 Expérience 6 : Reconnaissance de mélodies intercalées lorsque la mélodie présentée préalablement est transposée de +12, 13 ou 14 demi-tons 171
5.3.1 Introduction..... 171
5.3.2 Méthode 172
5.3.3 Résultats et discussion..... 173
5.4 Expérience 7 : Nombre de sons séparés et progression des performances de post-reconnaissance de mélodies intercalées..... 178

5.4.1 Introduction..... 178
5.4.2 Méthode..... 179
5.4.3 Résultats et discussion..... 180
5.5 Discussion générale..... 184
5.5.1 Résumé des principaux résultats..... 184
5.5.2 Nouvelles hypothèses théoriques sur le fonctionnement de l'analyse de la scène auditive..... 186
5.6 Conclusion..... 188

6 Déficit de la reconnaissance de mélodies intercalées : Etude de patients cérébro-lésés..... 189
6.1 Introduction..... 189
6.2 Post-reconnaissance de mélodies intercalées : Étude d'un groupe de patients cérébro-lésés..... 191
6.2.1 Introduction..... 191
6.2.2 Méthode..... 191
6.2.3 Résultats et discussion..... 193
6.3 Reconnaissance de mélodies intercalées avec ou sans présentation préalable de la mélodie : Un cas de dissociation..... 197
6.3.1 Introduction..... 197
6.3.2 Méthode..... 198
6.3.3 Résultats et discussion..... 202
6.4 Discussion générale..... 204
6.4.1 Résumé des principaux résultats..... 204
6.4.2 Interférence en mémoire due à l'encodage de plusieurs informations simultanées..... 205
6.4.3 Dissociation entre les mécanismes impliqués dans les conditions pré et post..... 205
6.5 Conclusion..... 207

CONCLUSION ET PERSPECTIVES..... 208

ANNEXES..... 214
A Les mélodies..... 215
B Les séquences distractrices..... 221
C Consignes..... 229
D Plages du CD..... 250

BIBLIOGRAPHIE..... 256

Introduction

"[...] Le monde et le sujet percevant se déterminent l'un l'autre"

Varela, Thompson et Rosch (1993)

Perception et Organisation perceptive

La perception est classiquement définie en Sciences Cognitives comme étant l'activité qui nous permet de prendre connaissance de notre environnement sur la base des informations prélevées par nos sens (Streri, 1998). "Prendre connaissance" de notre environnement signifie dans une perspective cognitiviste, se construire un modèle interne adapté qui rende compte des modifications de notre environnement et qui nous permette d'adapter de manière appropriée notre comportement. Les processus d'organisation perceptive regroupent l'ensemble des mécanismes qui permettent à partir des informations sensorielles qui nous parviennent de construire des "objets" qui ont un sens (Pomerantz & Kubovy, 1986). Ces processus jouent donc un rôle clé dans la construction de ce modèle interne, car de la construction des objets va dépendre l'extraction de leurs propriétés et ainsi leur reconnaissance et leur identification.

En raison de l'importance de ces processus dans la construction d'une représentation cohérente du monde, l'étude des processus d'organisation perceptive constitue un thème privilégié pour aborder la perception et de façon plus générale le fonctionnement du système cognitif. Elle soulève des questions fondamentales sur le type de relation existant entre le système cognitif et l'environnement qui ont, nous le verrons, des implications philosophiques dans la façon d'envisager la cognition.

L'Analyse des Scènes Auditives

Notre environnement sonore est composé de nombreuses sources qui émettent simultanément ou dont les émissions se recouvrent au moins partiellement. Si l'on enregistre le

signal acoustique émis dans une période de temps donné par l'ensemble de ces sources sonores, on constate qu'une onde complexe résultant de la combinaison de leurs ondes respectives parvient à nos oreilles à un instant donné, et que ce signal composite évolue dans le temps. Dans une situation naturelle d'écoute, le système auditif est ainsi soumis à une "mixture sonore". Pourtant nous ne percevons pas un brouhaha confus et inintelligible mais au contraire nous distinguons chacune des sources sonores en présence, et nous sommes capables de suivre leurs émissions indépendamment. Nous pouvons ainsi suivre la conversation d'une personne alors que d'autres parlent en même temps, ce que l'on appelle l'effet "cocktail party" (Cherry, 1953). Et nous sommes capables d'isoler différentes voix mélodiques parallèles dans une pièce musicale lorsqu'elles sont jouées par différents instruments de musique ou bien par le même instrument mais dans différents registres de hauteur. Comment le système perceptif procède-t-il pour décomposer cette mixture sonore en images auditives distinctes représentant les différences sources sonores de l'environnement, et lier les composantes provenant des mêmes sources au cours du temps nous permettant ainsi de suivre leur évolution ?

Ces deux analyses qui consistent d'une part à regrouper ou séparer à un instant donné les composantes sensorielles, et d'autre part à établir un lien perceptif ou bien à séparer les événements sonores au cours du temps en fonction de leur appartenance à une ou plusieurs sources, définissent deux types de processus respectivement les processus d'organisation simultanée et les processus d'organisation séquentielle. Ces deux mécanismes interagissent dans la construction de descriptions distinctes des différents objets de l'environnement (Bregman & Pinker, 1978) permettant d'aboutir à une représentation cohérente de notre univers sonore. Leur distinction ne se justifie que par la nécessité de disséquer les divers processus impliqués dans l'organisation auditive, afin d'explorer de façon approfondie chacun des aspects d'un mécanisme unique que Bregman (1990) appelle l'Analyse des Scènes Auditives par analogie à l'analyse des scènes visuelles en termes d'objets.

Dans cette thèse sont étudiés les processus impliqués dans le liage et la ségrégation perceptive d'événements présentés successivement dans le temps qui conduisent à la formation des flux auditifs.

Construction des flux auditifs

Qu'est-ce qu'un flux auditif ? Lorsque nous écoutons une séquence sonore, nous établissons un lien entre les sons successifs qui composent cette séquence, ou certains d'entre eux, et formons ainsi un ou plusieurs flux auditifs (Bregman & Campbell, 1971). Une mélodie, un flot de parole constituent des exemples de flux. Un flux auditif est donc la représentation mentale que nous avons d'une suite d'événements. Il peut regrouper l'ensemble des événements successifs émis et préserver ainsi l'ordre dans lequel ils sont apparus ainsi que leur relation temporelle. Dans ce cas, la différence entre un flux et une séquence sonore n'apparaît pas

immédiatement. Cependant, il se peut que les événements qui apparaissent successivement dans le temps ne proviennent pas de la même source sonore. Sur la base de la différence de leurs caractéristiques (Bregman, 1990), et une fois que le système auditif a accumulé suffisamment d'informations (Bregman, 1978a ; Rogers & Bregman, 1993 ; Rogers & Bregman, 1998), plusieurs flux auditifs seront créés. La succession temporelle initiale des sons est alors perdue (Bregman & Campbell, 1971) et de nouvelles relations entre les événements s'établissent. Dans ce cas, la différence entre un flux auditif et une séquence sonore apparaît de façon plus manifeste. Néanmoins, il ne faut pas se fier aux apparences et même dans le cas où l'ordre des événements présentés initialement dans une séquence est préservé, un flux et une séquence sonore ne décrivent pas la même réalité. Alors qu'une séquence sonore désigne une succession d'événements sans lien entre eux, un flux est la représentation mentale d'une suite de sons liés perceptivement. Or le lien établi entre les différents événements fait émerger de nouvelles propriétés qui n'existaient pas initialement dans la séquence, comme le contour mélodique, les intervalles et le rythme.

Lorsque l'on étudie les processus impliqués dans la formation des flux auditifs plusieurs questions se posent : Quels sont les facteurs qui interviennent dans la construction des flux auditifs ? Comment le système perceptif procède-t-il pour analyser la scène auditive ? Comment mesurer ces processus et accéder ainsi à la représentation que l'auditeur se construit de son univers sonore ? Ces différentes questions sont abordées tout au long de ce document et constituent le fil conducteur de cette thèse.

Plan de la thèse

Les recherches qui ont été menées inscrivent les travaux de W. Jay Dowling sur la perception de mélodies intercalées (Dowling, 1973 ; Dowling, Lung & Herrbold, 1987), dans la perspective théorique de l'"Analyse des Scènes Auditives" développée par Albert S. Bregman (Bregman, 1990, 1994). Cette thèse constitue une microanalyse des processus impliqués dans la reconnaissance d'une mélodie intercalée à des sons distracteurs. Elle permet d'aborder différents points importants pour la compréhension des processus de formation des flux auditifs. Le premier concerne le problème de la mesure de ces processus et les conséquences de la méthode employée sur le type de représentation sondée, conditionnant les inférences faites sur le fonctionnement et le dysfonctionnement de ces processus. Le deuxième concerne l'étude de la nature des processus impliqués dans la formation des flux auditifs. Le paradigme de reconnaissance de mélodies intercalées permet d'étudier le rôle des connaissances dans l'analyse perceptive d'une séquence sonore, et ainsi d'examiner une question apparemment paradoxale qui est de savoir si la connaissance d'un "objet" intervient dans sa construction. Cette question permet de tester la théorie de Bregman (1990, 1994) qui postule l'existence de deux processus distincts impliqués dans l'analyse des scènes auditives, l'analyse primaire et l'analyse guidée par les connaissances, et de préciser leur contribution respective dans la reconnaissance d'une

mélodie entremêlée ainsi que les relations qu'ils entretiennent. Ces recherches révèlent donc l'apport théorique de ce paradigme de reconnaissance de mélodies intercalées dans l'étude des processus mis en jeu dans la construction des flux auditifs et constituent un premier pas dans l'étude des bases neurales de ces processus.

Ce document est organisé en deux grandes parties. Une partie théorique présente l'état des connaissances dans les domaines abordés et soulève les questions qui seront discutées par la suite. Une partie expérimentale décrit les recherches réalisées dans le cadre de cette thèse. Un compact disc est joint à ce document et permet d'écouter les séquences qui ont été présentées aux auditeurs dans les principales expériences réalisées. Les plages du CD correspondant aux exemples sonores appropriés à chaque expérience sont mentionnées dans les sections concernées, et la liste des extraits se trouvant sur le disque figure dans l'annexe D.

La partie théorique se compose de trois chapitres. Le premier relate les études qui ont examiné les facteurs intervenant dans la fission auditive et la construction des flux auditifs. La mise en évidence de la contribution de différents indices dans le groupement et la ségrégation des événements auditifs et la mesure de ces processus sont abordés ensemble, afin de tenter de mieux cerner la nature de la représentation sondée. Un principe général énoncé par Bregman (1994) qui rend compte de l'heuristique de ces mécanismes, est présenté. Les différentes théories proposées pour expliquer le fonctionnement de ces processus sont exposées dans le deuxième chapitre. Elles se différencient sur deux points : celui de l'origine périphérique ou centrale de ces mécanismes et celui du niveau préattentif ou attentif auquel ils interviennent. Enfin, dans la mesure où les études menées dans le cadre de cette thèse ont examiné les processus de formation des flux auditifs par une méthode indirecte utilisant un paradigme de reconnaissance de mélodies intercalées, une synthèse des travaux portant sur les processus perceptifs et mnésiques impliqués dans la reconnaissance d'une mélodie font l'objet du troisième chapitre.

Les réalisations expérimentales sont également organisées en trois parties et sont décrites dans les chapitres 4, 5 et 6. Le chapitre 4 décrit les études qui ont conduit à l'élaboration d'un paradigme de post-reconnaissance de mélodies intercalées. L'effet de la différence de hauteur et de la dissemblance de timbre sur les performances de post-reconnaissance a été testé. La notion de fission mélodique partielle et le seuil de reconnaissance de mélodies intercalées sont proposés pour établir un lien entre ce type de mesure indirecte et les méthodes directes mesurant le seuil de fission d'une séquence alternante (van Noorden, 1975). Dans le chapitre 5, le rôle des connaissances dans la construction des flux auditifs est étudié en contrastant le paradigme de reconnaissance de mélodies intercalées classiquement utilisé (Dowling, 1973 expérience 2) avec le paradigme de post-reconnaissance élaboré. La nature des processus d'analyse de scène auditive guidée par les connaissances (Bregman, 1990, 1994) est examinée ainsi que le lien

entre ces processus et l'analyse primaire de scène auditive. Le chapitre 6 aborde la question, encore peu étudiée, des bases neurales des processus de formation des flux auditifs. Les premiers résultats de deux études exploratoires examinant l'aptitude de patients cérébro-lésés à reconnaître une mélodie intercalée à des sons distracteurs, sont reportés.

Enfin, les opérations mentales impliquées dans la perception de mélodies intercalées déduites à partir des résultats obtenus dans les différentes expériences, sont présentées en conclusion. L'apport des recherches menées dans le cadre de cette thèse pour la compréhension de la perception auditive est souligné.

Partie Théorique

Chapitre 1

Les multiples bases de la ségrégation auditive

Les divers aspects du phénomène de formation des flux ont conduit les chercheurs à multiplier les moyens d'investigation de ces processus. Les différentes méthodes employées pour mesurer ces processus sont examinées dans ce chapitre, et une classification sur la base du type de représentation sondée est proposée. Les études mettant en évidence les paramètres affectant le groupement et la ségrégation des événements auditifs sont décrites. Un principe général formulé par Bregman (1994) pouvant rendre compte de l'heuristique déployée par le système perceptif pour construire des flux auditifs est présenté.

L'étude des processus de formation des flux auditifs conduit à se poser deux types de questions : Quels sont les paramètres qui interviennent dans le groupement et la ségrégation des événements auditifs ? Et quelles sont les conséquences perceptives de ces processus ? Ces deux questions sont intimement liées car pour évaluer le rôle des différents facteurs dans la construction des flux auditifs, il est nécessaire de pouvoir accéder à la représentation de la scène auditive de l'auditeur.

Les différents moyens d'investigation déployés par les chercheurs pour accéder à l'analyse perceptive d'une séquence sonore réalisée par un auditeur, sont donc présentés dans la première partie de ce chapitre. Ces mesures sont regroupées en deux classes selon la façon, directe ou indirecte, d'accéder à ces processus de formation des flux. La différence de nature de la représentation sondée par ces deux types de méthodes nous amènera à préciser la terminologie employée tout au long de ce document. Dans une deuxième partie, les différents paramètres, acoustiques et perceptifs, intervenant dans la construction des flux auditifs sont

examinés. Les recherches sont relatées en précisant le type de méthode utilisé, et les études recourant au paradigme de reconnaissance de mélodies intercalées sont particulièrement développées. Enfin dans une troisième partie, les règles générales gouvernant ces processus de formation des flux auditifs sont présentées et rapprochées de certaines lois d'organisation perceptive formulées par les psychologues Gestaltistes dans le domaine visuel. Un principe général énoncé par Bregman (1994) rendant compte de l'heuristique conduisant à la formation des flux est présenté. Ce principe repose sur la théorie de la "complémentarité psychophysique" développée par Shepard (1981) selon laquelle le système perceptif exploiterait certaines régularités de l'environnement. Les implications philosophiques d'une telle position dans la façon d'envisager la perception sont discutées.

1.1 Mesure des processus de formation des flux auditifs

1.1.1 Fission auditive et formation des flux auditifs

C'est la découverte du phénomène de fission auditive qui a lancé l'étude des processus mis en jeu dans la construction des flux auditifs. La fission auditive désigne notre expérience perceptive du bascule entre deux percepts, lorsqu'une séquence organisée en un seul flux se scinde en deux flux. On en fait l'expérience essentiellement en laboratoire où l'on peut manipuler des facteurs induisant un changement dans l'analyse perceptive, ou bien à l'écoute de certaines pièces musicales dans lesquelles le compositeur fait varier rapidement le registre d'un instrument pour produire cet effet. En situation d'écoute naturelle, la redondance des informations rend peu probable le changement soudain de l'analyse de la scène auditive et donc la perception de ce phénomène.

La mise en évidence de ce phénomène de fission a donc été le révélateur de l'existence de processus plus généraux de construction des flux auditifs. Ces derniers ne sont en effet pas seulement mis en jeu lorsque l'on perçoit une fission, mais ils interviennent de façon permanente dans notre vie quotidienne pour séparer les émissions des différentes sources sonores de l'environnement et suivre leur évolution au cours du temps (Bregman, 1990). Ainsi, si le phénomène de fission atteste de l'existence et des propriétés des mécanismes de construction des flux auditifs, il n'en est pas pour autant le représentant. Afin de ne pas les assimiler, j'adopterai une terminologie différente pour les désigner. J'emploierai le terme de fission auditive traduisant le mot anglais "streaming" pour désigner ce phénomène de séparation perceptive des sons, et je parlerai de formation ou construction des flux auditifs pour désigner l'ensemble des mécanismes d'organisation séquentielle. Nous verrons que cette terminologie a son importance pour catégoriser les différentes méthodes employées pour mesurer ces processus et pour les comparer.

1.1.2 Paradigmes et stimuli utilisés pour mesurer la formation des flux auditifs

1.1.2.1 Les séquences cycliques

Du fait de l'histoire de l'étude scientifique de ces processus, la majorité des recherches étudiant les processus de formation des flux auditifs ont examiné ce phénomène de fission. Ils ont pour cela principalement utilisé des séquences cycliques composées d'une alternance de deux sons A et B présentant des caractéristiques différentes (fréquence, contenu spectral, etc.). Les études étant dans la plupart des cas réalisées avec des sons purs, ces patterns cycliques étaient généralement composés d'une alternance de sons aigu et grave. La séquence présentée était soit de type ABAB (Dannenbring & Bregman, 1976a, b ; Miller & Heise, 1950 ; van Noorden, 1975, expérience 2), soit de type ABA-ABA (Rogers & Bregman, 1993 ; Rogers & Bregman, 1998 ; Rose & Moore, 1997 ; Singh & Bregman, 1997 ; van Noorden, 1975, expérience 1 ; Vliegen, Moore & Oxenham, 1999 ; Vliegen & Oxenham, 1999, expérience 1), produisant un pattern rythmique différent si la séquence est organisée perceptivement en un flux (galop entendu) ou en deux flux (deux flux réguliers perçus, un rapide et un lent) (figure 1.1).

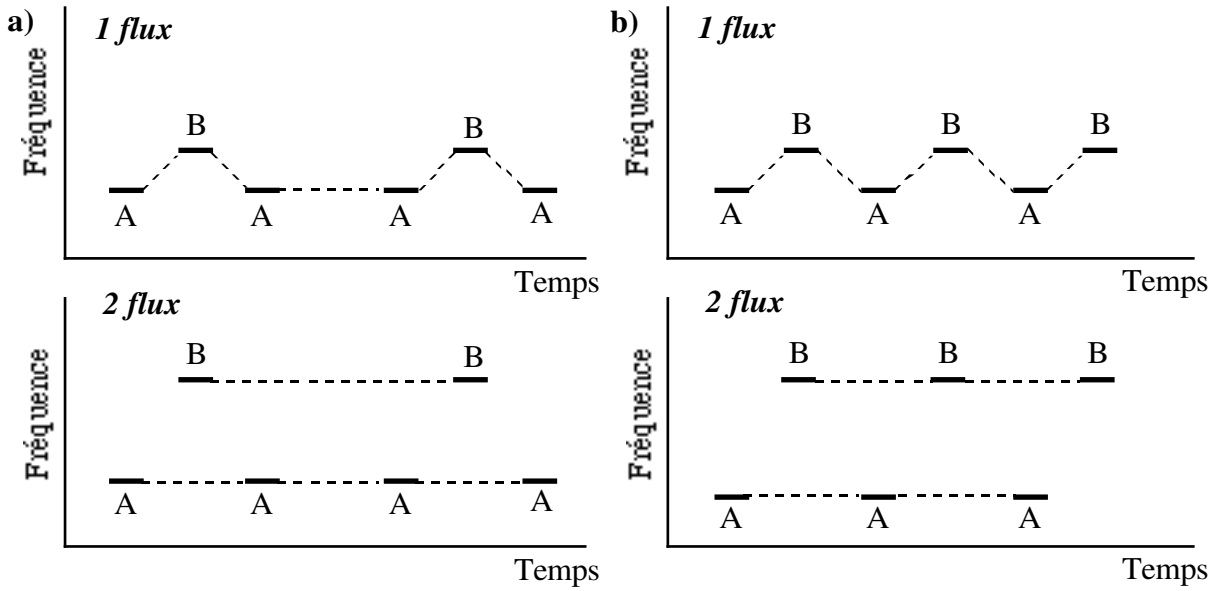


Figure 1.1. Représentation schématique des deux types de séquences cycliques couramment utilisées dans l'étude de la formation des flux auditifs : **a)** la séquence de type ABA-ABA et **b)** la séquence de type ABAB.

Des séquences cycliques dans lesquelles un pattern simple de type AB était intégré dans une séquence plus longue ont également été employées (Bregman, 1978b ; Bregman & Rudnicky, 1975), ainsi que des séquences composées de la répétition d'un pattern composé de plus de deux sons différents (Bregman, 1978a ; Bregman & Campbell, 1971 ; Bregman, Liao & Levitan, 1990 ; McAdams & Bregman, 1979 ; Singh, 1987).

Ces séquences cycliques sont fréquemment utilisées parce qu'elles présentent plusieurs avantages (Bregman, 1990, p. 53) : 1) il s'agit de séquences longues permettant d'amplifier le phénomène de fission qui est un processus cumulatif (Anstis & Saida, 1985 ; Bregman, 1978a ; Rogers & Bregman, 1998), 2) la structure reste constante ce qui permet de pouvoir en donner une description simple écartant ainsi la possibilité de développer des stratégies de réponse et enfin 3) ce type de pattern évite que les réponses soient influencées par le début et la fin de la séquence, ce qui est important quand la tâche est de rendre compte de l'ordre des sons par exemple (Bregman & Campbell, 1971). Ces stimuli ont cependant l'inconvénient majeur d'être prédictibles, permettant à l'auditeur d'élaborer un schéma de la séquence présentée qui pourrait intervenir dans la formation des flux auditifs (Bregman, 1990, chap. 4).

1.1.2.2 Les différentes méthodes

Les processus de formation des flux auditifs ont été souvent étudiés en mesurant le seuil de fission de séquences cycliques par une méthode d'ajustement (Bregman, 1978a ; Dannenbring & Bregman, 1976a ; Heise & Miller, 1951 ; Miller & Heise, 1950 ; van Noorden, 1975, expérience 2 ; van Noorden, 1977) ou bien par la méthode des limites (Rose & Moore, 1997 ; van Noorden, 1975, expérience 1 ; Vliegen & Oxenham, 1999, expérience 1). La méthode d'ajustement consiste à demander à l'auditeur de régler un paramètre physique quelconque de la séquence, par exemple la différence de fréquence des sons successifs ou le tempo, jusqu'à ce qu'il perçoive que la séquence se scinde en deux flux. Dans la méthode des limites, les caractéristiques physiques de la séquence varient progressivement et l'auditeur doit signaler le moment où son percept change, passant de un à deux flux ou l'inverse. Ces méthodes souffrent cependant du problème de l'inertie perceptive ("hysterésis"), c'est-à-dire cette tendance à préserver la même représentation perceptive alors que l'événement physique change. Les seuils mesurés s'avèrent donc extrêmement variables.

Compte tenu des multiples conséquences perceptives de la formation des flux auditifs, d'autres méthodes ont été également employées (Bregman, 1990, pp. 55-57) :

La mesure du temps d'intégration ou de fission d'une séquence. L'auditeur indique en maintenant le doigt appuyé sur une touche le temps durant lequel il perçoit la séquence intégrée ou scindée (Anstis & Saida, 1985), et il signale le changement du percept en arrêtant d'appuyer. La proportion de temps durant laquelle la séquence a été perçue comme scindée constitue un indice de la force de ségrégation.

Le jugement du nombre de flux perçus sur une échelle numérique. Le sujet indique s'il a perçu un ou deux flux en utilisant une échelle numérique rendant compte du degré de fission de la séquence et/ou de l'incertitude de sa réponse (Iverson, 1995, expérience 1).

La reconnaissance d'un pattern cible mélodique ou rythmique. L'auditeur doit extraire et reconnaître un pattern mélodique ou rythmique qui émerge de la formation des flux auditifs. Par exemple reconnaître une mélodie intercalée à des sons distracteurs ou à une autre mélodie (Deutsch, 1975b ; Dowling, 1973 ; Gregory, 1994 ; Hartmann & Johnson, 1991 ; Iverson, 1995 ; Vliegen & Oxenham, 1999) ou reconnaître un rythme de galop (Rogers & Bregman, 1993 ; Singh & Bregman, 1997 ; Vliegen et al., 1999).

Dessiner ce que l'on perçoit. Les auditeurs peuvent transcrire par écrit ce qu'ils ont entendu, reproduire les notes s'ils ont la formation musicale requise ou bien apparier un dessin au percept entendu (Singh, 1987).

Juger l'ordre des sons ou compter les sons. Déterminer l'ordre des sons présentés (Bregman, 1978b ; Bregman & Campbell, 1971 ; Bregman & Dannenbring, 1973 ; Bregman & Rudnicky, 1975 ; McNally & Handel, 1977) ou leur nombre, est une tâche rendue difficile lorsque la séquence est scindée en deux unités perceptives différentes. Les performances sont alors le reflet de l'organisation perceptive de la séquence.

1.1.3 Méthode directe versus indirecte

L'ensemble de ces paradigmes peut être regroupé en deux grandes classes de méthodes si on les distingue non plus sur la base de comment on mesure la formation des flux auditifs mais sur la base de ce que l'on mesure, le type de représentation sondée. En effet, certaines méthodes mesurent la fission perceptive d'une séquence, ce qui constitue un premier aspect observable de la formation des flux auditifs. Nous sommes d'une certaine façon "témoin" du changement de l'analyse de la scène auditive, passant d'une organisation en un flux à une organisation en deux flux. D'autres mesurent les propriétés qui émergent de la formation des flux auditifs, qui constituent un deuxième aspect observable de ces mécanismes. La mesure de ces deux aspects, la fission perceptive et la représentation des propriétés émergeant de la formation des flux auditifs, constitue deux accès respectivement direct et indirect à l'analyse de la scène auditive. On peut ainsi regrouper l'ensemble des méthodes utilisées pour mesurer la formation des flux auditifs sur la base de ce critère et de définir ainsi deux grandes catégories, les méthodes directes et les méthodes indirectes.

Les méthodes directes permettent, comme leur nom l'indique, d'accéder *directement* à l'analyse de la scène auditive en mesurant le phénomène de fission. C'est le cas lorsque l'expérimentateur demande à l'auditeur de signaler le moment où son percept change passant de 1 à 2 flux (ou l'inverse), ou d'ajuster un paramètre de la séquence jusqu'à ce qu'il perçoive cette fission. Ce type de méthode mesure l'expérience perceptive que l'auditeur a du changement de l'analyse de la scène auditive, une représentation de nature *phénoménologique*. Elle constitue

une mesure *subjective* car la réponse donnée par le sujet est basée uniquement sur la conscience perceptive qu'il a de ce phénomène. Cette mesure présente deux difficultés : elle ne permet d'accéder qu'à ce qui est consciemment accessible par l'auditeur et elle n'est absolument pas vérifiable.

Les méthodes indirectes mesurent *la nature et la précision de la représentation des propriétés d'un flux cible* en demandant à l'auditeur de reconnaître un pattern intra-flux (rythmique, mélodique, l'ordre des sons). Le nombre de flux construits par le système auditif est alors inféré sur la base des performances de reconnaissance de ce pattern cible. Ces méthodes accèdent ainsi de façon *indirecte* à la formation des flux auditifs. En raison de cet accès indirect, elles font également intervenir d'autres composantes cognitives qui n'ont pas nécessairement un lien avec les processus mesurés, comme des composantes mnésiques, attentionnelles, des connaissances préalables, par exemple la formation musicale s'il s'agit de reconnaître des mélodies. Par conséquent, il convient d'être prudent dans les inférences que l'on peut faire sur le nombre de flux perçus à partir des performances obtenues dans ce type de tâche (voir chap. 4 pour une discussion de cet aspect). Ces méthodes fournissent une mesure *objective* puisque la réponse donnée par l'auditeur peut être comparée à la réponse attendue. Dans une tâche de discrimination par exemple, la réponse "même" ou "différent" donnée par l'auditeur est rapportée à la différence physique des stimuli présentés permettant de calculer des réponses correctes et incorrectes et/ou des indices de sensibilité (d') et des critères décisionnels (ß) (Green & Swets, 1974 ; Macmillan & Creelman, 1991).

1.2 Les paramètres affectant le groupement et la ségrégation des événements auditifs

Les méthodes d'accès à la scène auditive ainsi définies, dans cette section sont exposés les résultats des études examinant les facteurs qui affectent la formation des flux auditifs. La plupart des études décrivent les indices qui induisent la fission perceptive d'une séquence (le plus souvent cyclique) en recourant donc à des méthodes directes. Cependant, les recherches examinant l'influence de certains paramètres sur la reconnaissance d'un pattern cible sont également relatées et les expériences sur la reconnaissance de mélodies intercalées sont particulièrement développées (Dowling, 1973 ; Hartmann & Johnson, 1991).

1.2.1 La différence de fréquence des sons purs

1.2.1.1 Interaction entre la différence de fréquence et le tempo

Miller et Heise (1950) ont été les premiers à étudier de façon systématique le rôle de la différence de fréquence de sons purs dans la fission perceptive d'une séquence. Ils avaient constaté de façon empirique que le trille entendu lorsque deux sons de fréquences différentes

alternaient n'était plus perçu si la différence de fréquence augmentait et si l'alternance était suffisamment rapide. Il existait un point de rupture, que le auteurs ont nommé "le seuil de trille", au-delà duquel ce n'était plus l'alternance entre les deux sons qui était perçue mais deux voix isotones indépendantes. Ils ont donc présenté à quatorze participants une séquence dans laquelle deux sons purs aigu et grave alternaient tous les 100 ms (séquence de type ABAB voir figure 1.1b), et leur ont demandé d'ajuster la fréquence d'un des deux sons jusqu'à ce qu'ils ne puissent plus entendre l'alternance entre les deux sons. Ils ont ainsi montré que le seuil de trille, c'est-à-dire la limite au-delà de laquelle cette séquence se scindait, était proportionnel à la fréquence de référence. Il était d'environ 15 %, soit 2.4 dt, pour une gamme de fréquences allant de 150 à 7000 Hz. Ce seuil avait tendance à diminuer pour les fréquences supérieures à 5000 Hz et la variabilité inter-sujet à augmenter.

La fission perceptive semblait donc dépendre du rapport de fréquences entre les sons successifs mais également du tempo de la séquence. En effet, Miller et Heise (1950) avaient remarqué que cet effet disparaissait si le tempo de la séquence était ralenti. C'est à van Noorden (1975) que revient le mérite d'avoir montré que non seulement le tempo intervenait dans ce phénomène de fission perceptive, mais qu'en plus l'effet du tempo interagissait avec l'effet de la différence de fréquence sur la ségrégation perceptive. Les deux expériences menées par l'auteur sur l'effet de la fréquence et du tempo sur la fission perceptive sont relatées dans le deuxième chapitre de sa thèse désormais célèbre. Dans la première expérience, l'auteur a présenté à trois auditeurs (dont lui-même) une séquence composée de l'alternance de deux sons purs de fréquences différentes de type ABA-ABA (figure 1.1a). Cette séquence a la propriété de faire émerger un pattern rythmique différent selon la façon dont elle est organisée perceptivement. Lorsque les sons sont liés perceptivement, ce que van Noorden appelle la cohérence temporelle, un rythme de galop dû au regroupement des trois sons ABA est entendu. En revanche, lorsque les sons A et B se séparent, deux flux isotones et isochrones sont perçus, un flux aigu rapide et un flux grave lent ou l'inverse si les sons A sont plus aigus ou plus grave que les sons B. Dans cette expérience, la fréquence des sons B était maintenue constante (1000 Hz) et la fréquence des sons A variait (le logarithme du rapport de fréquence entre les sons A et B variait linéairement avec le temps). L'auditeur entendait ainsi alternativement une fission (A-A-A-A et B---B), puis la cohérence temporelle (ABA-ABA), puis de nouveau une fission, etc. La séquence durait 80 secondes (figure 1.2).

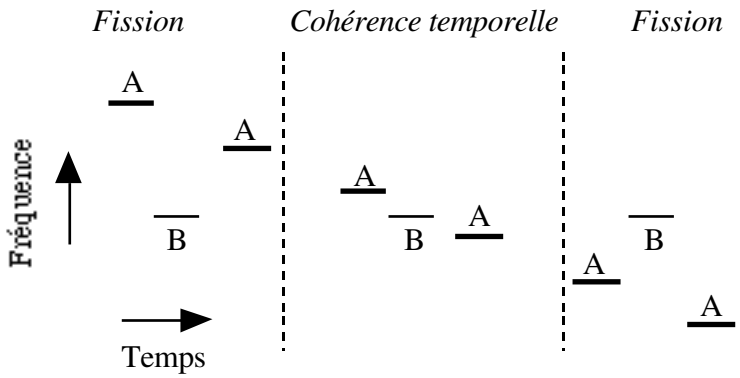


Figure 1.2. Méthode des limites ("sweep measurement") utilisée par van Noorden (1975, expérience 1) pour mesurer les seuils de fission et de cohérence temporelle. Une séquence cyclique ABA-ABA est présentée à l'auditeur durant 80 secondes. Alors que la fréquence du son B reste constante, celle du son A varie de telle sorte que l'auditeur perçoit alternativement une fission, la cohérence temporelle, une fission...[D'après van Noorden (1975, p. 9)].

Deux consignes ont été données successivement aux auditeurs : une consigne d'écoute sélective qui consistait à essayer d'entendre séparément les sons A et B, et une consigne d'écoute intégrative où l'expérimentateur demandait aux participants d'essayer d'entendre l'alternance ABA et le rythme galopant. Pour un tempo donné, les participants devaient donc essayer d'entendre le plus longtemps possible la cohérence temporelle durant huit présentations ininterrompues de la séquence, soit environ 10 minutes, et indiquer le moment où ils ne parvenaient plus à maintenir leur percept. Puis, durant de nouveau 10 minutes ils devaient essayer d'entendre le plus longtemps possible la fission et indiquer lorsque leur perception basculait. Cette procédure était répétée pour chacun des dix intervalles de temps séparant le début des sons A et B, compris entre 60 et 150 ms (Inter Onset Interval, IOI). L'auteur a ainsi mesuré la limite entre la cohérence temporelle et la fission lorsque l'auditeur adoptait une écoute intégrative, ce qu'il a appelé le seuil de cohérence temporelle, et lorsque l'auditeur essayait d'entendre séparément les sons A et les sons B, ce qu'il a nommé le seuil de fission. Les résultats obtenus par les trois auditeurs lorsque la fréquence de A était supérieure et inférieure à celle de B (les seuils obtenus étant symétriques), ont été moyennés et représentés sur le graphique ci-dessous (figure 1.3).

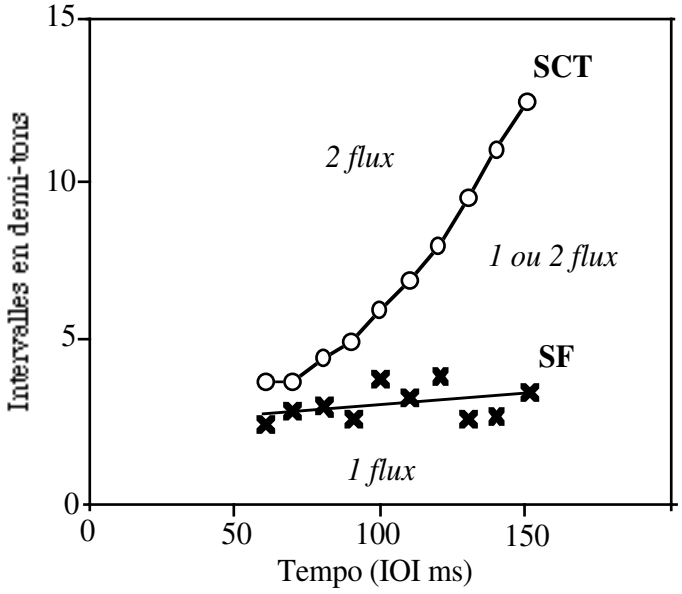


Figure 1.3. Seuil de fission (SF) et seuil de cohérence temporelle (SCT) obtenus par van Noorden (1975, expérience 1) en fonction de l'intervalle de temps qui sépare le début de deux sons successifs (IOI) exprimé en millisecondes et de l'intervalle en hauteur en demi-tons entre les sons A et B composant la séquence ABA-ABA.[D'après van Noorden (1975, p. 13)].

Pour un tempo donné, le seuil de cohérence temporelle est donc la différence de hauteur au-delà de laquelle il n'y a plus de cohérence temporelle et l'auditeur perçoit 2 flux. Le seuil de fission est la différence de hauteur en-deçà de laquelle l'auditeur ne parvient plus à scinder la séquence et il perçoit un flux. Les résultats de cette expérience ont révélé que la fission perceptive de la séquence était fonction de l'interaction entre la différence de fréquence des sons A et B et du tempo de la séquence. Plus la séquence était rapide plus la différence de fréquence nécessaire pour scinder la séquence était petite, et au contraire lorsque la séquence se ralentissait la différence de fréquence induisant une fission augmentait. Ce résultat s'applique essentiellement au seuil de cohérence temporelle qui augmente avec le tempo. Le seuil de fission semble relativement indépendant du tempo, il est environ de 2-3 dt, valeur identique à celle trouvée par Miller et Heise (1950).

Le résultat majeur de cette expérience est sans aucun doute d'avoir montré l'effet de la consigne sur la formation des flux auditifs, et ainsi l'existence de ces deux seuils différents qui définissent une zone ambiguë dans laquelle ce que l'auditeur perçoit est fonction de ce qu'il cherche à percevoir (1 ou 2 flux). A un tempo d'environ 7 sons/sec (soit un IOI de 150 ms), cette séquence ABA-ABA se scinde pour une différence de 3 dt entre les sons A et B, si les auditeurs cherchent à entendre séparément les sons aigus et graves. A ce même tempo, si les auditeurs cherchent au contraire à entendre l'alternance et le galop, une différence de 12 dt est nécessaire pour induire cette fission. Ce résultat met en évidence le rôle des connaissances dans la formation des flux auditifs. Ce point est discuté dans le chapitre suivant dans la section consacrée à l'analyse de la scène auditive guidée par les schémas.

Ces seuils obtenus par la méthode des limites ont été de nouveau mesurés par une méthode d'ajustement dans une deuxième expérience testant une gamme d'IOI plus étendue, comprise entre 48 et 800 ms (van Noorden, 1975, expérience 2). Cette fois la séquence présentée était une alternance régulière de sons aigu et grave, séquence cyclique de type ABAB (figure 1.1b). La fréquence du son B était toujours constante et égale à 1000 Hz et la fréquence du son A variait mais demeurait toujours supérieure à celle de B. Les deux types de consigne étaient donnés comme précédemment à deux auditeurs, dont l'auteur. Ils devaient donc ajuster la fréquence du son A de telle sorte qu'ils entendent juste la cohérence temporelle ou juste la fission selon la consigne. Les seuils mesurés étaient qualitativement comparables à ceux obtenus dans la première expérience, mais globalement abaissés. Le seuil de fission n'était plus que de 1 dt pour des IOI compris entre 48 et 400 ms, et la pente du seuil de cohérence temporelle était moins abrupte, révélant une tendance plus prononcée à scinder la séquence. En outre, la plus large gamme de tempi testée a permis de montrer que le seuil de fission était bien indépendant du tempo jusqu'à un IOI de 400 ms, mais augmentait ensuite avec l'IOI pour atteindre une valeur de 5 dt à 800 ms.

Les travaux réalisés par van Noorden (1975) ont donc révélé l'interaction entre deux facteurs influençant la formation des flux auditifs, la différence de fréquence et le tempo. Le système auditif semble lier perceptivement les sons proches en fréquence. Si la différence de fréquences entre les sons successifs augmente, il scinde la séquence en deux flux et cette tendance est d'autant plus prononcée que les sons sont proches temporellement.

1.2.1.2 Une différence de fréquence relative

Pour scinder une séquence, le système auditif se base-t-il sur une différence de fréquence fixe de 2-3 demi-tons, comme le révèlent les études de Miller et Heise (1950) et van Noorden (1975, expérience 1 (SF)), ou cette valeur est-elle sensible au contexte dans lequel les sons apparaissent ?

Heise et Miller (1951) ont examiné l'effet de la configuration mélodique sur le seuil de fission. Ils ont présenté une séquence de 11 sons purs équidistants en rapport de fréquence. Celui-ci, exprimé en demi-tons, pouvait être égal à 0 si les sons avaient la même fréquence, 0.5, 1, 1.65, 1.96 ou 2.42. Les sons purs étaient compris dans une gamme de fréquences allant de 900 à 2900 Hz environ. Quatre patterns différents étaient présentés : un pattern ascendant, un pattern descendant ou bien deux combinaisons différentes des deux trajectoires. La séquence était composée d'un pattern descendant puis ascendant ou l'inverse, et avait ainsi respectivement la forme d'un V à l'endroit ou à l'envers. Les auditeurs devaient ajuster la fréquence du son qui se trouvait au milieu de la séquence, jusqu'à ce qu'il se sépare perceptivement des autres sons composant le pattern. Trois auditeurs ont ainsi réalisé quatre ajustements sur vingt-quatre

séquences issues de la combinaison des quatre patterns et des six rapports de fréquences différents. Les résultats ont confirmé que le seuil de fission était proportionnel au rapport de fréquences entre les sons (Miller & Heise, 1950), et ils ont révélé qu'il dépendait de la configuration du pattern présenté. En effet, il était plus élevé dans les patterns ascendant et descendant que dans ceux qui combinaient ces deux trajectoires. Pourtant la fréquence des sons qui précédaient celui à ajuster était identique entre le pattern ascendant et celui qui avait une forme de V renversé, ainsi qu'entre le pattern descendant et celui en forme de V. Ce résultat suggère que le calcul fait par le système auditif pour construire des flux auditifs ne dépend pas uniquement des caractéristiques des sons qui précèdent mais également de celles des suivants. Le système perceptif semble prendre en compte le pattern dans sa globalité.

Bregman et Rudnicky (1975) ont confirmé le rôle du contexte dans la formation des flux auditifs. Bregman et Campbell (1971) avaient montré que l'ordre des sons était accessible uniquement si les sons étaient regroupés dans un même flux auditif. Les auteurs ont donc utilisé cette méthode indirecte et ont demandé à des auditeurs de juger l'ordre de deux sons A et B dont les fréquences étaient respectivement de 2200 et 2400 Hz. Ils étaient présentés simultanément avec deux autres sons X de fréquence égale à 1460 Hz (soit une différence de 7 et 8.6 dt entre les sons X et les sons A et B respectivement). Les auditeurs entendaient tout d'abord la paire AB ou BA présentée isolément, puis après une seconde de silence une séquence de comparaison de type XABX ou CCCXABXCC. Les sons C dits "capteurs" avaient une fréquence identique à celle des sons X (1460 Hz), ou égale à 1030 ou 590 Hz soit une différence respectivement de 6 et 15.7 dt. La tâche consistait à juger si l'ordre des sons apparaissant dans la paire AB présentée isolément, était identique à celui présenté dans la séquence de comparaison. Les résultats ont révélé que les performances des auditeurs augmentaient avec la fréquence des capteurs. Plus elle s'approchait de celle des sons X (une différence inférieure à 6 dt), plus les auditeurs parvenaient à reconnaître l'ordre des sons AB. Ce résultat montre que lorsque le rapport de fréquences entre les sons C et X est inférieur à celui existant entre les sons X et AB, les sons C "captent" les sons X. Les sons C et X sont ainsi regroupés en un flux distinct de celui formé par les sons A et B, rendant alors possible l'accès à l'ordre de cette paire de sons.

Bregman (1978b) a également montré le rôle des caractéristiques des sons avoisinants dans le groupement et la ségrégation de deux événements auditifs. Dans son étude, une séquence standard composée de la répétition d'une paire de sons purs A et B de fréquences différentes, était présentée suivie d'une séquence de comparaison dans laquelle les deux sons A et B étaient diffusés en présence de deux autres sons X et Y dans une séquence cyclique de type ABXY. L'auteur a examiné par le biais de différentes méthodes comment le système auditif allait organiser perceptivement cette dernière séquence. Dans une première expérience, il a demandé à l'auditeur de juger si l'ordre des sons A et B présentés dans la séquence standard était identique ou différent à celui présenté dans la séquence de comparaison. Dans une

deuxième expérience il a demandé avec quelle facilité la paire AB pouvait être entendue séparément des autres sons, et également de rendre compte du rythme perçu. La différence de fréquences entre les quatre sons A, B, X et Y variait de façon à créer deux conditions expérimentales. Dans la condition dite "isolée", les sons A et B étaient proches en fréquence et éloignés des sons X et Y, induisant le groupement des sons AB dans un flux séparé des sons XY. Dans la condition dite "absorbée", les sons X et Y étaient présentés dans le même registre de fréquence que les sons A et B, produisant une affectation de ces sons dans deux flux différents, AX et BY (figure 1.4).

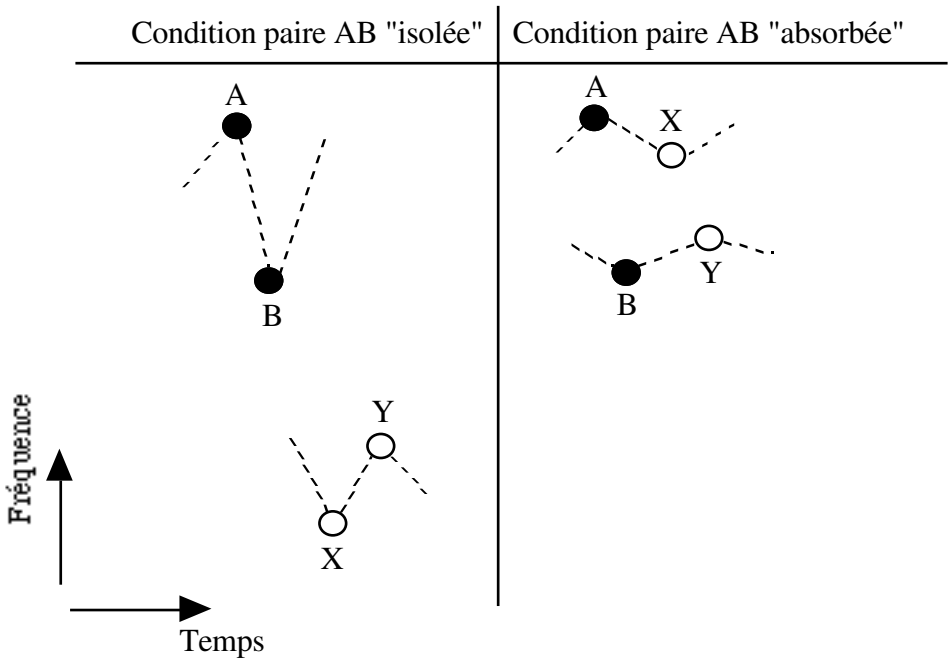


Figure 1.4. Les sons X et Y sont distants (condition "isolée") ou proches (condition "absorbée") en fréquence des sons A et B. Les lignes en pointillés représentent les liens perceptifs créés par le système auditif entre ces différents sons dans ces deux conditions expérimentales. [D'après Bregman (1978b, p. 393)].

Les différents types de jugements convergeaient pour montrer qu'une différence de fréquence fixe d'environ 10 dt entre les sons A et B présentés à un IOI de 110 ms, pouvait aussi bien conduire au groupement de ces sons en un flux (condition "isolée") qu'à leur séparation en deux flux distincts (condition "absorbée"), selon leur distance en fréquence des sons X et Y (respectivement de 27 et 1 dt). Ce résultat révèle que le système auditif ne se base pas uniquement sur le rapport de fréquence entre deux sons dans l'absolu pour les lier ou les séparer, mais prend en compte les caractéristiques des autres sons composant la séquence. La taille de la fenêtre d'intégration prise en compte par le système auditif pour réaliser cette analyse reste à déterminer.

1.2.1.3 Degré de recouvrement de la gamme de fréquences d'une mélodie

Nous avons vu que le rapport de fréquences entre les sons purs et la configuration de la séquence affectaient la fission perceptive d'une séquence sonore relativement simple. Quel(s) indice(s) le système auditif utilise-t-il pour extraire un pattern mélodique entremêlé ? Se base-t-il sur la différence de hauteur des sons successifs composant le pattern mélodique, une différence de hauteur moyenne ou un degré de recouvrement entre les gammes de fréquences occupées par les mélodies ?

Dowling (1973, expériences 1 et 2) a examiné l'effet de la différence de hauteur moyenne séparant deux mélodies intercalées sur leur reconnaissance. Dans une première expérience, l'auteur a présenté à six auditeurs deux mélodies familières intercalées. Il s'agissait pour la plupart d'extraits de comptines (Frères Jacques, Ah ! vous dirais-je maman, Joyeux Anniversaire...) composés de 16 notes, joués de façon isochrone à un tempo de 4 sons/sec et dont l'ambitus variait entre 7 et 12 dt. Des séquences composites formées par l'alternance des notes appartenant à deux comptines différentes étaient ainsi élaborées. Elles étaient composées de 32 notes et jouées à un tempo de 8 sons/sec. Les séquences étaient diffusées par haut-parleur. Un exemple de deux mélodies intercalées, Frères Jacques et Joyeux Anniversaire, présentées dans le même registre est représenté dans la figure ci-dessous (figure 1.5).

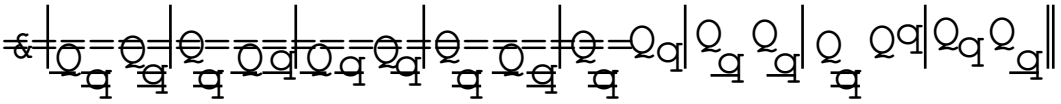


Figure 1.5. Exemple de deux mélodies familières intercalées, Frères Jacques et Joyeux Anniversaire, présentées dans le même registre et jouées à un tempo de 8 sons/sec (Dowling, 1973, expériences 1 et 3). Les notes impaires dont la hampe est dirigée vers le bas sont celles de la comptine Frères Jacques, et les notes paires dont la hampe est dirigée vers le haut sont celles appartenant à l'air de la chanson Joyeux Anniversaire.

Après une tâche préalable d'identification de l'ensemble des huit mélodies familières utilisées dans l'expérience, l'expérimentateur en présentait deux intercalées. Elles étaient jouées dans le même registre de hauteur durant les quatre premiers essais, puis tous les quatre essais une des deux mélodies était transposée de +1 dt. La différence de hauteur moyenne était ainsi augmentée progressivement par pas de 1 dt, jusqu'à ce que l'auditeur puisse identifier successivement les deux mélodies familières présentées. L'auteur a ainsi mesuré la différence de hauteur moyenne nécessaire pour identifier les deux mélodies familières intercalées (la deuxième identification suivant toujours la première). Ces deux identifications successives avaient lieu dans 50 % des cas au bout de 32 essais, soit pour une différence de 10 dt. Cependant, il est

difficile de distinguer le rôle de la différence de hauteur moyenne de celui du nombre d'essais, car dans cette étude les deux facteurs étaient confondus. L'augmentation du nombre d'essais a en effet pu favoriser l'intervention des connaissances dans l'organisation perceptive de cette séquence, d'autant plus qu'il s'agissait de mélodies familières. Néanmoins, W. Jay Dowling a constaté que la moitié des mélodies était identifiée au moment où la différence de hauteur entre la note la plus haute de la mélodie présentée dans le registre bas et la note la plus basse de la mélodie transposée, était de 1.2 dt. En outre, 75 % des identifications correctes survenaient lorsque le recouvrement entre les deux mélodies était nul. La mélodie Joyeux Anniversaire était par exemple reconnue une fois transposée d'un peu plus d'une octave (pour une différence comprise entre 12 et 14 dt), c'est-à-dire lorsque sa plus basse note était juste au-dessus de la note la plus haute de Frères Jacques (figure 1.6).

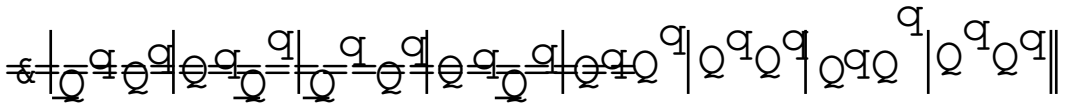


Figure 1.6. Transcription des deux mélodies familières intercalées, Frères Jacques et Joyeux Anniversaire, lorsque la mélodie Joyeux Anniversaire est transposée de +12 dt.

D'après Dowling (1973), la variable pertinente pour la fission mélodique serait le degré de recouvrement entre les ambitus des deux mélodies entremêlées, c'est-à-dire la séparation en hauteur entre la note la plus haute d'une des mélodies et la note la plus basse de l'autre. Dans l'étude qu'il a menée, la différence de hauteur moyenne entre les deux mélodies et le degré de recouvrement étaient deux variables confondues. Cependant, l'auteur s'appuie sur le fait que la différence de hauteur moyenne conduisant à l'identification des mélodies était plus variable que le taux de recouvrement.

Afin de distinguer l'effet du nombre d'essais de celui de la différence de hauteur sur la fission mélodique, Dowling (1973, expérience 2) a conduit une deuxième expérience dans laquelle chaque mélodie n'était présentée qu'une seule fois. L'auteur a ainsi mis en place un paradigme de reconnaissance immédiate de mélodies non familières intercalées (figure 1.7). Deux mélodies successives composées de cinq notes étaient présentées : une mélodie standard suivie après une pause de 2 secondes de la mélodie à comparer intercalée avec des sons distracteurs. Les mélodies étaient présentées à un tempo de 3 sons/sec, soit un IOI de 330 ms et la séquence composite à un tempo deux fois plus rapide de 6 sons/sec. Les deux mélodies étaient identiques ou bien la mélodie à comparer était une nouvelle mélodie dont le contour, c'est-à-dire la direction du changement de hauteur, différait de celui de la mélodie standard (voir la section 3.1.2 pour une définition de cet indice et une discussion de son importance dans la reconnaissance de mélodies non familières). Les mélodies étaient différentes à chaque essai.

Elles étaient construites à partir de petits intervalles (intervalles de ± 1 dt avec une probabilité de 0.5 et de ± 2 ou ± 3 dt avec une probabilité de 0.25), de telle sorte que pour une différence de hauteur moyenne de 6 dt le recouvrement soit faible. L'affectation des notes paires et impaires pour la mélodie ou la distractrice était choisie aléatoirement. Deux séquences distractrices différentes ont été élaborées. Elles étaient soit présentées dans le même registre de hauteur que celui de la mélodie à comparer (différence de hauteur moyenne de 0 dt), ou elles s'éloignaient vers les hautes fréquences de 6 ou 12 dt en moyenne. Après une phase de familiarisation composée de 18 essais, les séquences étaient diffusées par haut-parleur à douze participants qui devaient juger si les deux mélodies présentées étaient identiques ou différentes, et donner un degré de confiance dans leur jugement.

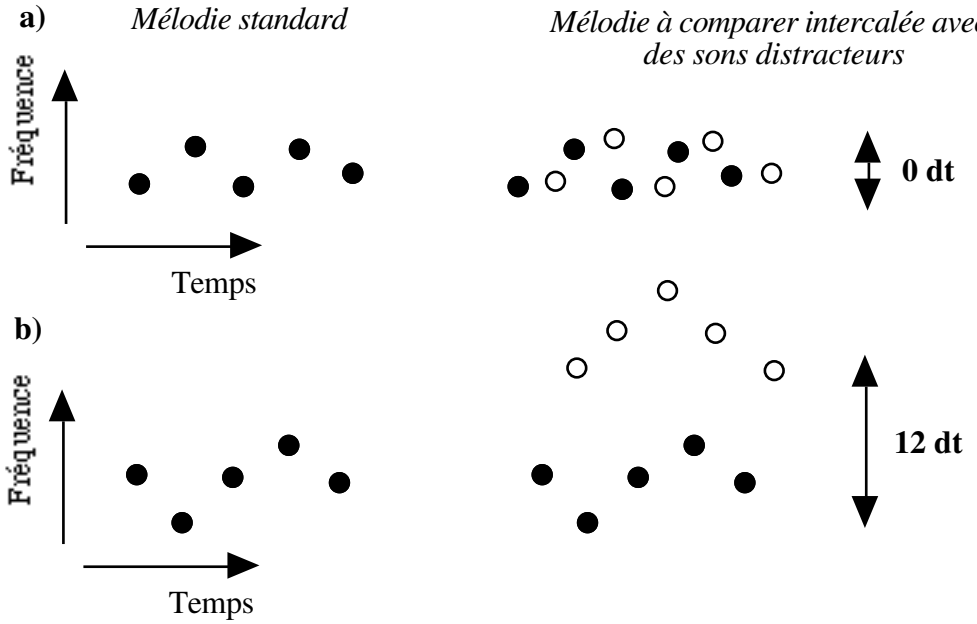


Figure 1.7. Paradigme de reconnaissance de mélodies intercalées utilisé par Dowling (1973, expérience 2). Une mélodie standard isolée est présentée suivie après une pause de 2 secondes d'une mélodie à comparer (points noirs) intercalée avec des sons distracteurs (points blancs). L'auditeur doit juger si les deux mélodies présentées sont identiques ou différentes. Les sons distracteurs sont présentés dans le même registre que la mélodie ou éloignés de +6 ou +12 dt.

L'aire sous la courbe MOC (Memory Operating Characteristic), homologue de la courbe ROC (Receiver Operating Characteristic) pour la mémoire, a été calculée, donnant une estimation non biaisée de la proportion de réponses correctes (Norman & Wickelgren, 1965 ; Swets, 1973). Les résultats figurent dans le tableau ci-dessous (Tableau 1.1).

	Différence de hauteur moyenne en demi-tons			
	0	6	12	Moyenne
Session 1 (60 essais)	0.57	0.72	0.69	0.66
Session 2 (60 essais)	0.62	0.65	0.69	0.65
Moyenne	0.59	0.69	0.69	

Tableau 1.1. Aire sous la courbe MOC donnant une estimation non biaisée de la proportion de reconnaissances correctes obtenues pour les différentes conditions de séparation en hauteur moyenne entre la mélodie et la séquence distractrice, pour les 60 premiers essais (session 1) ainsi que les 60 suivants (session 2). [D'après Dowling (1973, p. 330)].

Les résultats ont révélé que les proportions de reconnaissances correctes augmentaient avec la différence de hauteur moyenne qui séparait la séquence distractrice et la mélodie. Les performances obtenues lorsque la séquence distractrice était présentée dans le même registre que celui de la mélodie étaient faibles, mais supérieures au hasard ($p < .05$). Elles ont eu tendance à augmenter au fur et à mesure des essais, suggérant un effet d'apprentissage dans cette tâche. Les performances obtenues lorsque la différence de hauteur était de 12 dt étaient équivalentes à celles recueillies pour une séparation de 6 dt. Ce résultat conforte l'hypothèse formulée par W. Jay Dowling, selon laquelle la fission mélodique serait liée au degré de recouvrement de la gamme de fréquences des mélodies. En effet, compte tenu de la construction des mélodies, à 6 dt il n'y a plus de recouvrement entre la mélodie et la séquence distractrice, ce qui pourrait expliquer le fait que les performances plafonnent. Les performances de reconnaissance de mélodies intercalées demeurent cependant inférieures à celles obtenues dans une simple tâche de discrimination de mélodies non familières (.98) (Dowling & Fujitani, 1971). Ceci suggère que la présence de la distractrice interfère dans cette tâche de reconnaissance (ce point est discuté dans la section 3.3.2).

Hartmann et Johnson (1991) ont reproduit la première expérience de Dowling (1973) en étendant le corpus de mélodies familières utilisé (34 mélodies différentes au lieu de 8). Les mélodies étaient comprises dans une gamme allant de 5 à 12 dt. Elles avaient la même note moyenne, le la3 (440 Hz) afin de maximiser le croisement entre les mélodies lorsqu'elles étaient intercalées ainsi que leur contact, c'est-à-dire le nombre de fois où les notes appartenant aux deux mélodies étaient identiques. Les auteurs ont examiné le rôle de divers facteurs dans la fission mélodique, la différence de fréquence moyenne séparant les deux mélodies, la différence de latéralisation, la différence de timbre des sons appartenant aux deux mélodies (dimensions spectrale, temporelle et spectro-temporelle), la différence d'intensité, de durée, de rythme, de

rugosité, de réverbération, etc. Nous nous focaliserons ici sur l'effet d'une transposition de 12 dt (Hartmann & Johnson, 1991, expérience 1 conditions 0 et 1), l'influence des autres facteurs est relatée dans les sections correspondantes. Alors que les performances de double identification des mélodies intercalées étaient aléatoires lorsque les mélodies avaient la même hauteur moyenne, le pourcentage moyen de reconnaissance correcte atteignait presque 90 % lorsque la différence de hauteur moyenne qui les séparait était de 12 dt. L'ambitus maximal des mélodies étant de 12 dt, la transposition avait pour effet de supprimer le recouvrement fréquentiel entre les mélodies.

Récemment Vliegen et Oxenham (1999, expérience 2 avec des sons purs) ont reproduit le paradigme de reconnaissance immédiate de mélodies non familières intercalées utilisé par Dowling (1973, expérience 2). Ils ont également trouvé qu'une différence de hauteur moyenne de 11 dt conduisait à une proportion moyenne de reconnaissance correcte de .85 chez cinq musiciens et .64 chez cinq non musiciens. Les auditeurs qui n'avaient pas bénéficié de formation musicale avaient des performances globalement plus faibles, et parmi eux deux répondaient au hasard.

Les résultats des études décrites ci-dessus montrent donc qu'une différence de hauteur moyenne d'au moins une octave (12 dt soit un rapport de fréquence égal à 2), est nécessaire pour que les auditeurs puissent extraire une mélodie intercalée (Dowling, 1973 ; Hartmann & Johnson, 1991 ; Vliegen & Oxenham, 1999). Le paramètre pertinent pourrait être le degré de recouvrement entre les gammes de fréquences occupées par les mélodies (Dowling, 1973), ce qui renforcerait l'idée que le système auditif se base sur la configuration d'un pattern plutôt que sur le rapport de fréquences des sons successifs pour former des flux auditifs. Cependant, dans la plupart des études la différence de hauteur moyenne et le taux de recouvrement sont confondus, ne permettant pas de distinguer clairement l'effet de ces deux facteurs.

1.2.2 Ségrégation perceptive des sons complexes : Rôle de la hauteur et du timbre

La plupart des recherches étudiant les processus de formation des flux auditifs ont utilisé des sons purs. Certains auteurs se sont cependant intéressés à la fission perceptive de sons complexes permettant ainsi de distinguer le rôle que pouvaient jouer la hauteur et le timbre dans ces processus. Le rôle de la fréquence fondamentale et de la composition spectrale des sons complexes dans la fission perceptive et la formation des flux auditifs sont examinés dans cette section. Les recherches évaluant le rôle du timbre dans sa multidimensionnalité (Grey, 1977 ; Krumhansl, 1989 ; McAdams, Winsberg, Donnadieu, De Soete & Krimphoff, 1995), non seulement sa dimension spectrale mais aussi ses dimensions temporelle et spectro-temporelle, sont également relatées.

1.2.2.1 Hauteur et dimension spectrale du timbre

Van Noorden (1975, expériences 1 et 2) a montré le rôle de la différence de fréquences des sons dans la fission perceptive d'une séquence alternante. Cependant, ces études étant menées sur des sons purs, elles ne permettaient pas de déterminer si le groupement perceptif s'effectuait sur la base de la contiguïté fréquentielle des sons ou sur la base de la contiguïté de leur hauteur. Dans le but de distinguer ces deux alternatives, van Noorden (1975, chap. 3) a examiné la cohérence temporelle de séquences composées de sons complexes ayant la même hauteur mais des composantes fréquentielles non contiguës ou ayant des hauteurs différentes mais présentant une contiguïté fréquentielle.

Deux séquences cycliques différentes composées de sons de même hauteur mais dont les fréquences n'étaient pas contiguës, ont été élaborées. L'une était composée de l'alternance d'un son pur et d'un son complexe dont la fréquence fondamentale absente était identique à la fréquence du son pur (figure 1.8a). L'autre était composée de l'alternance de deux sons complexes ayant la même fréquence fondamentale absente, mais composés d'harmoniques de rangs différents (figure 1.8b). Lorsque les sons étaient présentés à un IOI de 100 ms, ces deux types de séquences se scindaient. Deux flux étaient perçus, l'un formé par les sons purs A l'autre par les sons complexes C pour la séquence ACAC, et l'un formé par les sons complexes C et l'autre par les sons complexes C' pour la séquence CC'CC'. La seule façon d'entendre un flux était de ralentir la séquence à un IOI de 150 ms.

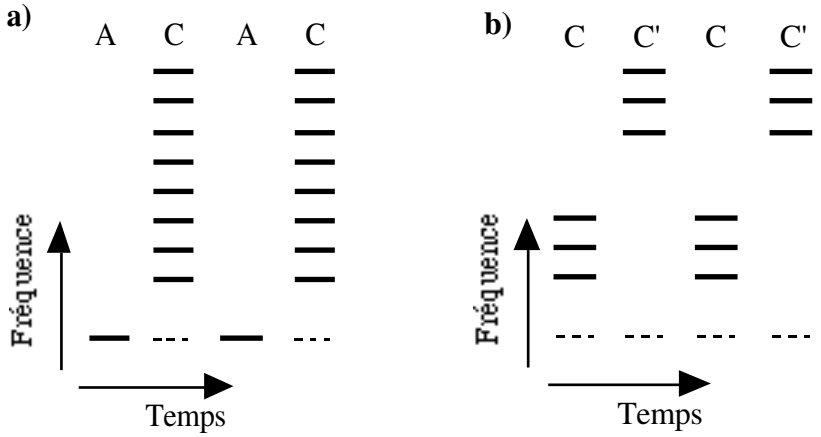


Figure 1.8. Séquences construites par van Noorden (1975, chap. 3) : **a)** alternance d'un son pur A de fréquence égale à 200 Hz et d'un son complexe C composé des harmoniques de rang 3 à 10 (comprises dans une gamme allant de 600 à 2000 Hz) et de fréquence fondamentale absente égale à la fréquence du son pur ; **b)** alternance de deux sons complexes de même fréquence fondamentale absente mais de composition spectrale différente, un son C composé des harmoniques de rang 3, 4 et 5 et un son C' composé des harmoniques de rang 8, 9 et 10.

Par ailleurs, une séquence composée de sons présentant une contiguïté fréquentielle mais des hauteurs différentes, a été construite. Il s'agissait d'une séquence dans laquelle deux sons complexes alternaient au même tempo que précédemment, l'un de fréquence fondamentale de 140 Hz et composé des harmoniques de rang 8, 9 et 10, et l'autre de fréquence fondamentale différente qui variait mais filtré de sorte que le maximum d'énergie spectrale puisse être modifié. Van Noorden a alors observé qu'il percevait une cohérence temporelle lorsque le filtrage était centré sur 1250 Hz, fréquence qui correspondait à celles des harmoniques du premier son, et une fission lorsque le maximum d'énergie spectral était déplacé, que la hauteur des deux sons soit identique ou non. Il en a déduit que le système auditif semblait regrouper les sons sur la base de la contiguïté fréquentielle plutôt que la contiguïté de la hauteur. Néanmoins, compte tenu des stimuli utilisés on ne peut pas exclure la possibilité de l'intervention du timbre de ces sons dans cette fission perceptive.

McAdams (1977), dans une étude décrite par McAdams et Bregman (1979), a en effet montré qu'une séquence composée de quatre sons purs, deux sons aigus et deux sons graves de différentes fréquences présentés en alternance, initialement perçue comme un seul flux pouvait se scinder en deux flux si les deux sons les plus aigus étaient enrichis de leur troisième harmonique. Si au contraire le même enrichissement harmonique était appliqué aux deux sons les plus graves, la fission s'en trouvait réduite. Cette étude met en évidence le rôle des facteurs spectraux dans la formation des flux auditifs ainsi que leur interaction avec la hauteur des sons. En effet, l'augmentation de la brillance du son combinée à la différence de hauteur entre les sons successifs, a eu pour conséquence d'augmenter la distance entre les événements induisant une fission dans le premier cas, alors qu'elle a au contraire réduit cette distance dans le second cas.

Wessel (1979) a conforté l'idée que la composition spectrale d'un son pouvait intervenir dans la formation des flux auditifs en mettant en évidence le rôle de la dimension spectrale du timbre dans l'organisation perceptive d'un pattern mélodique. Il a en effet montré qu'un pattern mélodique ascendant formé par une triade de notes répétée, était perçu comme un pattern descendant si les sons étaient joués alternativement par des timbres distants sur la dimension spectrale du timbre (figure 1.9). Cette dimension est physiquement liée à la répartition des harmoniques dans le spectre et correspond perceptivement à la "brillance" du son, la trompette et le cor d'harmonie constituent deux exemples d'instruments situés aux antipodes de cette dimension (Grey, 1977).

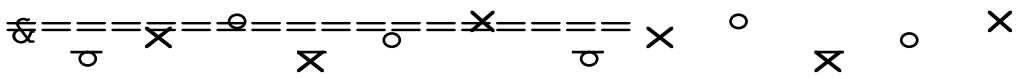


Figure 1.9. Représentation schématique du pattern mélodique utilisé par Wessel (1979). Des triades de notes formant un pattern ascendant sont présentées et jouées alternativement par deux instruments différents symbolisés par les signes "O" et "X". Si les sons sont groupés perceptivement sur la base de leur proximité en hauteur, une ligne mélodique ascendante est perçue. S'ils sont groupés sur la base de leur similarité timbrale, deux lignes mélodiques descendantes sont entendues.[D'après Wessel (1979, p. 49)].

Plus récemment, Hartmann et Johnson (1991) ont montré que la différence de composition spectrale des sons induisait une fission mélodique équivalente à celle observée pour une différence de fréquence moyenne entre les sons purs de 12 dt. A l'aide du paradigme de double identification de mélodies familières intercalées décrit précédemment, ils ont constaté que deux mélodies intercalées présentées dans le même registre de hauteur étaient identifiées si l'une était jouée par des sons purs et l'autre par des sons complexes. Alors que les performances étaient aléatoires si les deux mélodies étaient jouées par des sons purs. Il s'agissait de sons complexes harmoniques filtrés passe-bas avec une fréquence de coupure de 5000 Hz dont les 9 premières harmoniques étaient à leur amplitude maximale (à l'exception de la cinquième).

La contribution respective de la hauteur et du contenu spectral dans la formation des flux auditifs ainsi que l'interaction de ces deux facteurs, ont été examinées de façon systématique par Singh (1987). L'auteur a présenté une séquence cyclique composée de la répétition de quatre sons complexes dont la hauteur "virtuelle" (fondamentale absente) et le timbre (série harmonique) variaient. Quarante-neuf séquences de type T2H1-TmH1-T2Hn-TmHn, résultant de la combinaison de 7 différences de timbre (T2-Tm) et 7 intervalles de hauteur (H1-Hn), ont été élaborées. Le premier son avait une hauteur (do3 dont la fondamentale est égale à 262 Hz) et un timbre (harmoniques de rang 2 à 5) constants, le deuxième son avait la même hauteur fixe que le premier mais un timbre qui variait (7 séries de 4 harmoniques différentes), le troisième son avait le même timbre fixe que le premier son mais une hauteur qui variait (7 hauteurs différentes do3, mi3, sol3, do4, mi4, sol4, do5) et enfin le quatrième son avait à la fois une hauteur et un timbre qui variaient. Une telle séquence générait un pattern rythmique et mélodique différent en fonction de l'indice choisi par le système auditif pour grouper et séparer les sons. Si les sons était groupés sur la base de la similarité de leur hauteur, l'auditeur regroupait les sons successifs par paire et percevait une alternance de paire de sons graves et aigus. S'ils étaient groupés sur la base de la similarité de leur timbre, les sons paires et impaires formaient respectivement 2 flux réguliers composés d'une alternance de sons aigu et grave. Les

auditeurs devaient désigner l'illustration visuelle qui correspondait au percept entendu (figure 1.10).

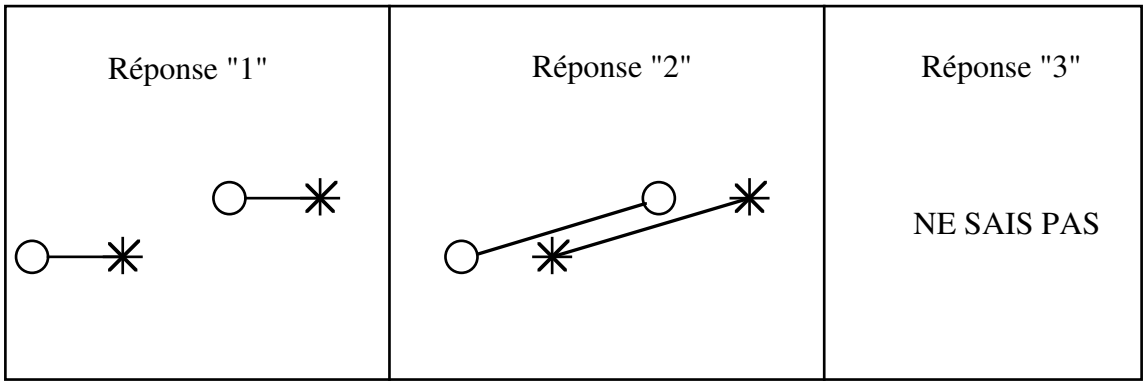


Figure 1.10. Illustrations visuelles proposées aux participants dans l'expérience menée par Singh (1987). Les auditeurs indiquent le type de percept entendu, celui produit par un groupement auditif sur la base de la hauteur (réponse "1"), sur la base du timbre (réponse "2") et lorsque la réponse est incertaine (réponse "3"). [D'après Singh (1987, p. 889)].

Trois conditions expérimentales ont été élaborées. Dans la première condition, la différence de hauteur (H1-Hn) était constante et la différence de timbre (T2-Tm) variait. Dans la deuxième condition, la différence de timbre (T2-Tm) était constante alors que la différence de hauteur (H1-Hn) variait. Dans la troisième condition, les deux facteurs hauteur et timbre variaient simultanément, soit dans le même sens, soit dans le sens inverse. Lorsque la différence de hauteur était constante, la différence de timbre suffisante pour induire un changement de percept passant d'un groupement sur la base de la hauteur à un groupement sur la base du timbre, augmentait avec l'intervalle de hauteur. Pour une différence de hauteur nulle (lorsque les quatre sons avaient la même hauteur), les auditeurs regroupaient les sons sur la base du timbre si les séries harmoniques étaient décalées de deux harmoniques (série 2, 3, 4, 5 et 4, 5, 6, 7). A mesure que l'intervalle de hauteur augmentait, cette différence de timbre conduisant à un changement d'organisation perceptive, était accrue. A partir d'une différence de hauteur de 19 dt, la séquence était toujours organisée sur la base de la hauteur. Lorsque la différence de timbre demeurait constante, le groupement se faisait sur la base de la hauteur au-delà d'une différence de 7 dt. Si les sons avaient le même timbre, ils étaient regroupés sur la base de la hauteur. Enfin, lorsque ces deux indices variaient dans le même sens le groupement se faisait sur la base d'une différence de hauteur, et lorsqu'ils variaient dans le sens inverse le groupement sur la base du timbre tout d'abord perçu était abandonné au profit d'un groupement sur la base de la hauteur pour une différence de hauteur supérieure à 7 dt. Ces résultats révèlent que la hauteur (même en l'absence de la fréquence fondamentale) et le timbre (contenu spectral) interviennent dans l'organisation perceptive d'une séquence et qu'ils interagissent. Par défaut, le système auditif semble regrouper les sons successifs sur la base de la hauteur, mais si la

différence de timbre s'accroît le paramètre pris en compte pour organiser perceptivement la séquence est alors le timbre.

Bregman, Liao et Levitan (1990) ont cherché à étendre les résultats obtenus par Singh (1987) sur des sons complexes dont la composition harmonique était cette fois identique et où seule la position du maximum d'énergie spectral, ce que les auteurs ont nommé le "pic spectral", fluctuait. L'idée était de déterminer si le regroupement perceptif sur la base du contenu spectral observé par Singh (1987), était dû à l'appartenance des sons à la même bande de fréquences ou bien à la similarité du pic "formantique". Les sons présentés dans la première expérience étaient des sons complexes filtrés, de telle sorte que la fréquence du pic spectral puisse être déplacée. Le paradigme s'apparentait à celui utilisé par Singh (1987). Une séquence cyclique formée par la répétition de quatre sons présentant une similarité dans leur fréquence fondamentale ou dans leur pic spectral était présentée. Le premier son composant cette séquence, le son "pivot", avait une fréquence fondamentale fixe de 128 Hz et un pic spectral égal à 1000 Hz. Le deuxième son, le son "contraste", était choisi parmi les 25 sons élaborés résultant de la combinaison de 5 fréquences fondamentales et 5 pics spectraux, en faisant en sorte qu'il diffère du son pivot sur ces deux dimensions. Les rapports entre les différentes fréquences fondamentales et les différents pics spectraux étaient identiques. Le troisième son avait la même fréquence fondamentale que le son pivot et un pic spectral identique au son contraste. Et enfin le quatrième son avait au contraire une fréquence fondamentale identique au son contraste et le même pic spectral que le son pivot. Selon l'indice utilisé par le système auditif pour grouper perceptivement les sons de cette séquence, l'agencement des sons perçu était différent. S'il se basait sur la similarité de la fréquence fondamentale, les sons 1 et 3 étaient groupés et séparés des sons 2 et 4. Si en revanche il se basait sur la similarité du pic spectral, les sons 2 et 3 étaient liés perceptivement et séparés des sons 1 et 4. Cette séquence cyclique était précédée d'une séquence standard composée des deux premiers sons du cycle, et la tâche de l'auditeur était de juger la facilité avec laquelle il pouvait entendre le pattern mélodique et rythmique créé par cette paire standard dans la séquence cyclique. Le jugement était reporté sur une échelle numérique allant de 1 (difficile) à 5 (facile). La différence de fréquence fondamentale et de pic spectral variaient indépendamment, et les auteurs examinaient l'effet du facteur testé par rapport au facteur dit "perturbateur". Les résultats ont révélé que les deux facteurs, la fréquence fondamentale et le pic spectral, induisaient une fission perceptive. A mesure que la différence de fréquence fondamentale augmentait, la ségrégation perceptive sur la base de ce facteur était accrue, de même pour l'augmentation de la différence de pic spectral. Ce dernier facteur est apparu plus puissant pour des rapports de fréquences équivalents. Pour une différence donnée de fréquence fondamentale ou de pic spectral, la variation du facteur perturbateur a eu tendance à contrecarrer l'effet du facteur testé, suggérant une interaction entre ces deux paramètres. Les résultats de cette expérience montrent que la fréquence fondamentale

et le pic spectral interviennent et interagissent dans la formation des flux auditifs, confirmant les résultats obtenus par Singh (1987) avec cette fois des sons dont les spectres se recouvrent.

Dans une deuxième expérience, les auteurs ont examiné le rôle que pouvait jouer l'acuité de ce pic spectral dans la formation des flux auditifs. L'acuité du pic spectral correspond à la différence d'intensité relative entre le pic et les autres composantes fréquentielles. Pour un son ne présentant qu'un seul pic spectral (ce qui est le cas dans cette étude), ce facteur est corrélé au centre de gravité spectral qui est la moyenne des fréquences des harmoniques pondérée par leurs amplitudes relatives. A l'aide d'une méthode identique à celle adoptée dans la première expérience, mais en générant les sons par synthèse additive afin de s'assurer qu'ils étaient tous composés des mêmes harmoniques, les auteurs ont fait varier l'acuité du pic spectral dont les valeurs étaient de 0 (spectre plat), 6, 12, 18 ou 24 dB. Ils ont ainsi constaté que pour un rapport de fréquence constant, la fréquence fondamentale constituait un facteur de ségrégation plus puissant que le pic spectral, contrairement à ce qui avait été observé dans la première expérience. En outre, l'augmentation de l'acuité de ce pic spectral compensait cet effet et l'inversait, donnant finalement plus de puissance au pic spectral. Ce résultat suggère que le pic spectral ou le centre de gravité spectral (les deux paramètres étant ici confondus) affecte la formation des flux auditifs.

Toutes ces études suggèrent donc que les facteurs spectraux jouent un rôle important dans la formation des flux auditifs.

Récemment Vliegen et ses collaborateurs (Vliegen et al., 1999 ; Vliegen & Oxenham, 1999) ont examiné le rôle du codage temporel de la hauteur dans la formation des flux auditifs. Ils ont mené deux études successives évaluant l'effet de la différence de période de sons complexes sur la fission perceptive d'une séquence de type ABA-ABA (SF) et d'une mélodie intercalée à des sons distracteurs (Vliegen & Oxenham, 1999), ainsi que sur l'échec de l'intégration perceptive d'une séquence ABA-ABA (SCT) (Vliegen et al., 1999).

Dans la première étude, Vliegen et Oxenham (1999) ont mesuré la fission auditive par deux types de méthode : une méthode directe consistant à demander à l'auditeur de signaler le moment où il ne parvenait plus à scinder une séquence alternante, et une méthode indirecte de reconnaissance immédiate de mélodies non familières intercalées (Dowling, 1973, expérience 2). Dans la première expérience, les auteurs ont mesuré le seuil de fission d'une séquence cyclique de type ABA-ABA composée de 12 triplets. La fréquence du son A était de 100 Hz et celle du son B variait, l'intervalle entre les sons A et B était de 1, 2, 3, 4, 5, 7 ou 11 dt. Trois types de sons étaient présentés : des sons purs, des sons complexes composés d'harmoniques basses résolues sur la membrane basilaire (filtre passe-bande 500 à 2000 Hz), ou des sons complexes composés d'harmoniques élevées non résolues sur la membrane basilaire (filtre passe-bande 2000 à 8000 Hz). L'intervalle de temps qui séparait les sons successifs était de 110 ms ainsi que le silence entre les triplets. Les auditeurs devaient donc essayer de séparer perceptivement les sons B (aigus) des sons A et signaler lorsqu'ils n'y parvenaient plus. Le

seuil de fission obtenu était équivalent pour les trois types de sons, purs, complexes composés d'harmoniques résolues ou non, et comparable à celui trouvé par van Noorden (1975, expérience 1). En l'absence d'indices spectraux, les auditeurs sont donc parvenus à séparer les deux sons sur la base uniquement de leur périodicité. Afin de contrôler si les harmoniques élevées non résolues n'avaient pas créé des distorsions dans les basses fréquences qui auraient pu induire la ségrégation, les auteurs ont mené une expérience contrôle dans laquelle ils ont ajouté un bruit masquant. Les résultats étaient comparables à ceux obtenus précédemment, suggérant que la ségrégation observée pour les sons composés d'harmoniques élevées non résolues ne pouvait pas être expliquée par la présence de distorsion dans les basses fréquences. Les auteurs ont confirmé ce résultat dans une deuxième expérience avec une tâche de reconnaissance de mélodies intercalées, identique à celle utilisée de l'expérience 2 de Dowling (1973) (figure 1.7). Deux mélodies non familières atonales composées de 5 notes étaient présentées successivement, elles étaient identiques ou différentes sur deux notes (la troisième et la quatrième) de ± 2 dt. La deuxième mélodie était intercalée à des sons distracteurs distants de la mélodie de 1 ou 11 dt en moyenne vers les hautes fréquences. L'auditeur devait déterminer si les deux mélodies entendues étaient identiques ou différentes. Deux types de sons étaient proposés, des sons purs ou bien des sons complexes dont les harmoniques n'étaient pas résolues (sons filtrés entre 4000 et 8000 Hz). La séquence intercalée était présentée à un IOI de 110 ms. Dix auditeurs, cinq musiciens et cinq non musiciens ayant participé à l'expérience précédente (à l'exception de deux) ont été soumis à cette tâche. Les performances de reconnaissance augmentaient entre 1 et 11 dt de la même façon pour les deux types de son pur et complexe (composés d'harmoniques non résolues) surtout chez les musiciens suggérant que la mélodie cible pouvait être extraite des sons distracteurs sur la base de la périodicité du son.

Cette première étude a donc révélé que les auditeurs parvenaient à séparer perceptivement des sons de hauteurs différentes malgré le fait que les harmoniques ne soient pas résolues sur la membrane basilaire, suggérant que la période des sons pouvait être utilisée par le système auditif pour scinder une séquence.

Vliegen et ses collaborateurs (Vliegen et al., 1999) ont examiné dans une deuxième étude, si ce paramètre pouvait conduire à la fission irrésistible des sons d'une séquence que les auditeurs tentent d'intégrer (SCT). Ils ont présenté à huit auditeurs deux séquences : une séquence de référence composée de quatre triplets ABA- induisant un rythme de galop, et une séquence de comparaison dans laquelle le son B avait été décalé ou non du son A dans le temps. Ce décalage temporel conduisait à la perception d'une irrégularité rythmique si la séquence était intégrée perceptivement (galop irrégulier). Les auditeurs devaient indiquer dans quelle séquence le son B avait été décalé. Trois conditions expérimentales ont été élaborées. Les sons A et B étaient des sons purs (la fréquence de A était de 300 Hz et celle de B variait), ou des sons complexes de fréquences fondamentales différentes (la fréquence fondamentale de A était de 100 Hz et celle de B variait) et filtrés de telle sorte que seules les harmoniques élevées non

résolues sur la membrane basilaire soient présentes, ou bien des sons complexes de même fréquence fondamentale mais de contenus spectraux différents. La différence de fréquence pour les sons purs, de hauteur pour les sons complexes dont la fondamentale variait et de contenu spectral pour les sons de même fondamentale était de 1, 4, 8, 13 ou 18 dt. Les expérimentateurs ont mesuré pour l'ensemble de ces conditions le seuil de discrimination temporelle par une méthode adaptative. Les résultats ont révélé que les six auditeurs parvenant à réaliser la tâche (sur les huit participants), voyaient leur seuil de discrimination temporelle croître à mesure que la différence de hauteur tonale ou spectrale augmentait. Ce résultat suggère que l'augmentation de cette différence de hauteur induisait une fission irrésistible de la séquence empêchant l'auditeur de juger si le son B avait été décalé ou non. Les performances de discrimination temporelle se détérioraient particulièrement dans le cas où les sons A et B étaient des sons purs ou des sons complexes de contenu spectral différent, et dans une moindre mesure lorsque les sons A et B différaient uniquement par leur périodicité (harmoniques non résolues).

Cette deuxième étude confirme donc le rôle de la période dans la formation des flux auditifs, mais révèle que la puissance de ce facteur temporel pour induire une ségrégation irrésistible est moindre que celle des paramètres spectraux.

Outre les facteurs spectraux, le système auditif semble pouvoir utiliser des informations temporelles pour construire des flux auditifs. Il s'agissait dans les deux études précédentes de la périodicité des sons, mais nous verrons dans la section suivante que les informations temporelles liées au timbre affectent également l'analyse de la scène auditive.

1.2.2.2 Dimensions temporelle et spectro-temporelle du timbre

Récemment, des études se sont intéressées au rôle que le timbre pouvait jouer dans la formation des flux auditifs. Si le groupement et la séparation des événements sont opérés en fonction de leur appartenance ou non à la même source sonore (Bregman, 1990), le timbre, qui permet d'identifier les sources, devrait intervenir dans cette analyse perceptive. Cet attribut complexe a été défini par l'American Standards Association en 1960 comme étant "ce qui distingue deux sons provenant du même endroit, ayant les mêmes hauteurs tonales, sonies et durées subjectives" [ma traduction] (McAdams et al., 1995, p. 177). Des études ont tenté de mieux le définir et ont mis en évidence son caractère multidimensionnel (Grey, 1977 ; Krumhansl, 1989 ; McAdams et al., 1995). Ces recherches montrent que les dimensions perceptives pouvant rendre compte des jugements de similarité/dissemblance effectués sur des paires de timbre, sont liées à des variations physiques spectrales, temporelles et spectro-temporelles. Le rôle de la composition spectrale du son sur la formation des flux auditifs ayant été examiné dans la section précédente, nous nous intéresserons ici à la contribution des dimensions temporelle (attaque du son), et spectro-temporelle (le flux spectral correspondant au synchronisme des attaques des harmoniques ainsi qu'à la fluctuation de leurs amplitudes au

cours du temps) dans la formation des flux auditifs et également au rôle global de la dissemblance du timbre.

Gregory (1994) a initié cette recherche en examinant le rôle de la dissemblance du timbre sur la formation des flux auditifs. Son étude s'est déroulée en deux étapes. Dans un premier temps, il a demandé à des auditeurs d'estimer sur une échelle numérique la dissemblance de sons présentés par paire. Il s'agissait de sons de synthèse imitant les instruments de musique suivants : la clarinette, la flûte, le cor d'harmonie, le piano, la trompette, le violon et le violoncelle. L'analyse multidimensionnelle a révélé un espace de timbre en trois dimensions, corrélées respectivement aux variations spectrale (énergie située au niveau des partiels 1 à 3), temporelle (chute du son) et spectro-temporelle (rapport des harmoniques paires et impaires) des sons. Dans un deuxième temps, l'auteur a présenté des gammes ascendantes et descendantes dont les notes simultanées ou intercalées étaient jouées ou non par le même instrument. Les auditeurs devaient indiquer le percept entendu parmi trois réponses possibles : une gamme entière ascendante ou descendante, une demi gamme alternativement ascendante puis descendante ou l'inverse, ou bien une mélodie alternante. Lorsque deux gammes entières respectivement ascendante et descendante étaient intercalées dans le temps et jouées par le même instrument de musique, les auditeurs percevaient une demi gamme révélant qu'ils avaient organisé la séquence sur la base de la similarité en hauteur des sons successifs. Si la dissemblance de timbre des deux instruments jouant ces deux gammes augmentait, alors l'auditeur rapportait percevoir des gammes entières suggérant que les sons avaient été regroupés sur la base de la similarité de timbre. La dissemblance de timbre devait cependant être plus importante pour contrecarrer l'effet de la hauteur dans le cas où un instrument jouait une demi gamme ascendante (ou descendante) une note sur deux intercalée avec une demi gamme descendante (ou ascendante) également une note sur deux. Ce résultat pourrait s'expliquer par un conflit entre l'activation de schémas (connaissance des gammes) et la fission perceptive due à des dissemblances de timbre (voir la section 2.2.2 pour la distinction entre les deux processus d'analyse de la scène auditive). Globalement, cette étude conforte l'idée d'une interaction entre ces deux facteurs hauteur et timbre dans l'organisation séquentielle d'une séquence (Bregman et al., 1990 ; Singh, 1987). Cependant, certains résultats laissaient penser que la dissemblance globale de timbre sur les trois dimensions perceptives n'était peut être pas le facteur déterminant la formation des flux auditifs. En effet, deux instruments comme le violon et le violoncelle étaient plus séparés perceptivement que le violon et la trompette pourtant plus distants dans l'espace. Il se trouve que le violoncelle est plus distant sur la dimension spectrale du violon que ne l'est la trompette. Selon l'auteur, certaines dimensions perceptives en particulier la dimension spectrale auraient une plus grande importance dans ces processus de formation des flux auditifs.

Iverson (1995) s'est proposé de déterminer quelles dimensions perceptives du timbre intervenaient dans la formation des flux auditifs et si les jugements de similarité/dissemblance

pouvaient constituer de bon indicateurs de la métrique développée par le système auditif pour grouper et séparer perceptivement les sons. Dans une première expérience, il a présenté une séquence cyclique dans laquelle alternaient deux sons de même hauteur mais de timbres différents choisis parmi un ensemble de 16 timbres d'instruments de musique réels. Les auditeurs devaient juger à l'aide d'un curseur s'ils percevaient plutôt un flux composé de l'alternance de deux timbres différents, ou bien plutôt deux flux monotimbraux. Par une technique d'analyse multidimensionnelle (Shepard, 1962a, b), ces jugements estimant le degré de fission ont été transformés en distance dans un espace géométrique euclidien. Un espace bidimensionnel a été trouvé dans lequel se trouvaient les différents instruments plus ou moins éloignés selon leur degré de "séparabilité". Les deux dimensions étaient corrélées perceptivement à l'attaque et à la brillance des sons, montrant l'intervention des facteurs spectraux mais aussi temporels dans la formation des flux auditifs. La différence de centre de gravité spectral (CGS) des sons affectait la fission perceptive d'autant plus que les sons avaient un CGS élevé. L'attaque des sons intervenait également mais uniquement lorsque la séquence était isochrone physiquement et non perceptivement, suggérant que l'irrégularité rythmique pouvait affecter la formation des flux auditifs. Enfin, la différence de flux spectral semblait induire une fission surtout pour les sons en ayant peu. Cet espace s'apparentait à celui obtenu à partir de jugements de dissemblance sur le même ensemble de timbres (Iverson & Krumhansl, 1993). Les jugements de fission perceptive étaient corrélés aux jugements de dissemblance ($r = .81$, $df = 118$, $p<.001$), suggérant que les jugements de dissemblance de deux timbres constituaient de bons prédicteurs de leur degré de fission.

Cette première expérience a révélé l'intervention des facteurs temporels dans la formation des flux auditifs. Or Wessel (1979) ainsi que Hartmann et Johnson (1991, expérience 1 condition 5), avaient trouvé que l'attaque des sons n'affectait pas la fission mélodique. La divergence entre ces deux résultats peut être attribuée à deux grandes différences : 1) une différence dans les stimuli utilisés—Iverson (1995, expérience 1) utilise une séquence isotone alors que Wessel (1979) et Hartmann et Johnson (1991) ont eu recours à des mélodies, et 2) une différence dans la méthode employée—directe pour Iverson (1995, expérience 1) qui a mesuré ces processus en interrogeant les auditeurs sur la fission perceptive d'une séquence et indirecte pour Wessel (1979) et Hartmann et Johnson (1991) qui accédaient à la scène auditive en testant la reconnaissance d'un pattern mélodique. Dans une deuxième expérience, Iverson (1995) a donc testé si les facteurs temporels intervenaient dans l'organisation perceptive d'une séquence dans laquelle les hauteurs fluctuaient. Un pattern cible de quatre notes intercalé à des sons distracteurs, était présenté quatre fois en boucle. Les auditeurs devaient déterminer si le pattern cible était un pattern mélodique ascendant ou descendant et mentionner le degré de confiance dans leur jugement sur une échelle numérique. Les sons composant le pattern cible et les sons distracteurs étaient choisis dans l'espace de timbre (Iverson & Krumhansl, 1993) de telle sorte que l'attaque variait alors que la composition spectrale demeurait relativement constante. Les résultats révèlent que la différence d'attaque entre les sons distracteurs et les sons

composant le pattern cible, permet de reconnaître le pattern mélodique d'autant plus que l'attaque des sons cibles est brève. S'agissant de sons naturels, des fluctuations spectrales étaient corrélées à ces variations d'attaque, ce qui ne permet pas de distinguer l'effet des différences temporelles de celui des différences spectro-temporelles dans la formation des flux auditifs. Néanmoins, cette étude suggère que des facteurs dynamiques, attaque et flux spectral, interviennent dans la formation des flux auditifs même pour une séquence dont la hauteur varie.

La contribution respective des facteurs spectraux et temporels dans la formation des flux auditifs a été examinée de façon systématique par Singh et Bregman (1997) avec des sons de synthèse permettant de contrôler ces paramètres indépendamment. Les auteurs ont présenté une séquence de type ABA-ABA (figure 1.1a) à un IOI de 110 ms et un silence entre les triplets de 120 ms. Ils ont demandé aux auditeurs d'essayer d'entendre le galop le plus longtemps possible et de signaler lorsque la séquence se scindait, mesurant ainsi le seuil de cohérence temporelle (van Noorden, 1975). Les sons A et B avaient la même enveloppe temporelle et le même nombre d'harmoniques, ou se différenciaient sur une des deux caractéristiques ou bien les deux. L'enveloppe présentait une montée abrupte et une décroissance progressive ou l'inverse, correspondant respectivement à des temps de montée de 5 ms et de chute de 95 ms et inversement. Le nombre d'harmoniques était de 2 ou 4. Quatre conditions expérimentales issues de la combinaison de ces deux facteurs, ont été élaborées. Dans chacune de ces conditions, la différence de hauteur entre les sons A et B augmentait ou diminuait progressivement par pas de 1/4 de tons, dans une gamme comprise entre 262 Hz et 524 Hz. Le SCT était d'environ 2-2.5 dt lorsque les sons A et B avaient la même enveloppe temporelle et le même nombre d'harmoniques, valeur inférieure à celle trouvée par van Noorden (1975, expérience 1) qui était de 7 dt. Il diminuait lorsque l'enveloppe (1.5-2 dt), ou le nombre d'harmoniques différait (1 dt) et lorsque les deux variaient (0.9 dt). Ce résultat montre l'effet des deux dimensions du timbre, temporelle et spectrale, dans la fission perceptive. Le SCT était également affecté par la direction du changement de hauteur. Il était plus élevé lorsque la différence de hauteur entre les sons A et B augmentait, que lorsqu'elle diminuait. Ceci suggère que la tendance à la fission est plus importante lorsqu'on commence par entendre la séquence scindée (voir la section 2.2.2.2 pour une discussion de cet aspect). Dans la condition où les sons avaient la même enveloppe et le même nombre d'harmoniques, les auteurs ont constaté un effet du type d'enveloppe. Le SCT était plus élevé lorsque l'attaque du son était rapide. Iverson (1995, expérience 2) avait constaté que ce facteur était important, mais dans son expérience qui consistait au contraire à essayer de séparer les sons composant la séquence, la rapidité de l'attaque favorisait au contraire la fission. Par ailleurs, ils n'ont pas observé d'effet du nombre d'harmoniques.

En résumé, cette étude montre qu'une différence de nombre d'harmoniques entraîne une diminution du seuil de cohérence temporelle de 1.5 dt, révélant une fois encore l'importance des facteurs spectraux dans la fission auditive (Bregman et al., 1990 ; Hartmann & Johnson, 1991 ; McAdams & Bregman, 1979 ; Singh, 1987 ; Wessel, 1979). En outre, elle confirme la

contribution de l'enveloppe temporelle dans la formation des flux auditifs (diminution du SCT d'environ 1 dt). Ce résultat converge avec celui obtenu par Iverson (1995). Hartmann et Johnson (1991, expérience 1 condition 5) n'avaient pas trouvé que ce facteur (une attaque de 74 ms et une chute de 8 ms et l'inverse) permettait d'identifier deux mélodies familières intercalées. Cependant, les auteurs avaient ajusté le tempo des mélodies de telle sorte que le mélange soit perçu isochrone malgré la différence d'attaque des sons successifs. Or, Iverson (1995, expérience 1) a montré que l'effet de l'attaque n'apparaissait que pour des séquences isochrones physiquement, suggérant que l'effet de l'attaque sur la fission auditive serait lié à l'irrégularité rythmique induite.

1.2.3 La différence d'intensité et de localisation spatiale

Si l'importance des facteurs spectraux dans la fission auditive semble incontestable, cela ne signifie pas pour autant qu'ils soient les seuls facteurs à intervenir dans la formation des flux auditifs, comme nous avons pu le voir avec l'influence des facteurs temporels. Selon Bregman (1990), le système auditif regroupe perceptivement les sons provenant d'une même source sonore de l'environnement. Par conséquent, des indices comme la localisation spatiale et l'intensité des événements doivent renseigner le système perceptif sur la provenance de ces sons et donc leur appartenance ou non à une même source. Les études décrites ci-dessous examinent la contribution de ces paramètres dans la formation des flux auditifs.

1.2.3.1 La différence de localisation spatiale

La localisation spatiale des événements sonores peut être appréhendée par la différence interaurale de temps d'arrivée des signaux, par la différence de signaux qui arrivent aux deux oreilles (écoute dichotique) ou bien par la différence interaurale d'intensité. Si ce facteur intervient dans la formation des flux auditifs, ces indices devraient affecter l'analyse perceptive d'une séquence.

Van Noorden (1975, chap. 3) a constaté qu'une séquence de type ABAB (figure 1.1b), dans laquelle les sons A et B étaient présentés à des oreilles différentes et alternaient tous les 100 ms, ne permettait pas de percevoir une réelle cohérence temporelle quel que soit l'intervalle de hauteur qui séparait les deux sons. De même, Hartmann et Johnson (1991, expérience 1 condition 2) ont constaté que deux mélodies familières jouées dans le même registre de fréquence mais présentées aux deux oreilles différentes étaient identifiées aussi bien que si elles avaient été séparées d'une octave (environ 90 % d'identifications correctes). Une présentation binaurale de ces mêmes mélodies (présentées dans le même registre) ne permettaient pas de les séparer, le pourcentage d'identification avoisinant les 55 %. En outre, l'ajout d'un décalage

temporel entre les deux mélodies (condition 7) ne s'avérait pas aussi efficace qu'une présentation dichotique, le pourcentage d'identification étant de l'ordre de 80 %.

Cette ségrégation perceptive sur la base de l'oreille de présentation est-elle liée à la stimulation de "canaux" différents comme le postulent ces auteurs (Hartmann & Johnson, 1991 ; van Noorden, 1975) (voir la section 2.1 pour une discussion de cette position théorique) ? Ou bien au fait que la présentation dichotique induit une différence de localisation spatiale des signaux, qui est interprétée par le système auditif comme le signe de la présence de deux sources sonores (Bregman, 1990) ?

Deutsch (1975b) a montré que la présentation dichotique de deux séquences sonores n'était pas une condition suffisante pour séparer perceptivement les sons composant ces séquences. L'auteur a présenté deux gammes simultanées, l'une ascendante l'autre descendante. Les notes de chacune de ces gammes alternaient entre les deux oreilles toutes les 250 ms (figure 1.11a).

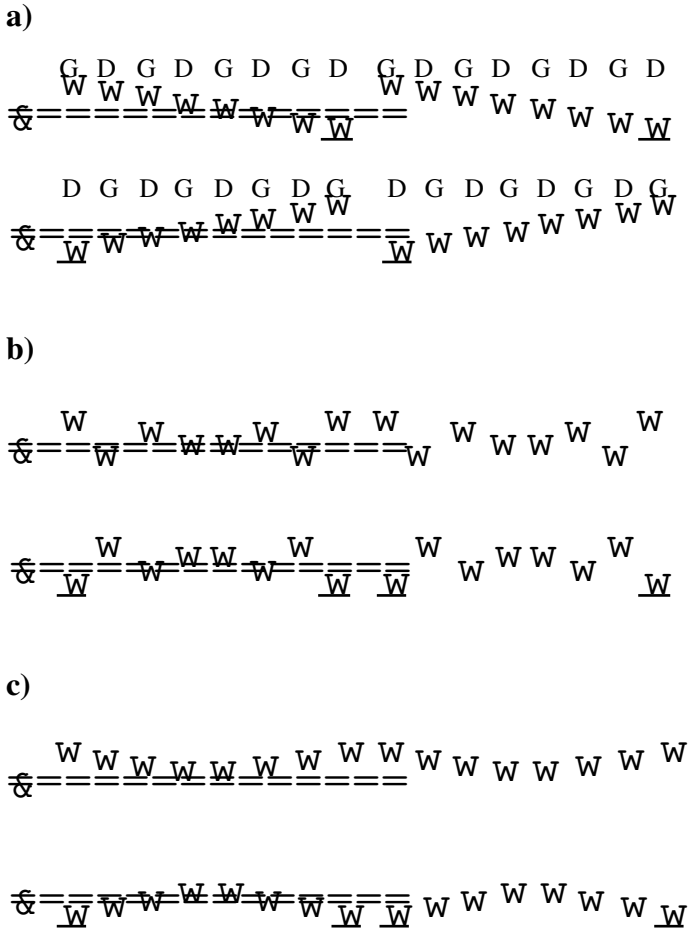


Figure 1.11. Patterns mélodiques présentés par Deutsch (1975b). Les lettres D et G symbolisent respectivement l'oreille de présentation Droite et Gauche. **a)** Gammes ascendantes et descendantes présentées simultanément. **b)** Patterns perçus si les sons sont groupés sur la base de l'oreille de présentation. **c)** Patterns perçus si les sons sont groupés sur la base de la proximité fréquentielle.[D'après Bregman (1990, p. 77)].

Si les auditeurs avaient groupé les sons sur la base de l'oreille de présentation un pattern mélodique du type de celui représenté sur la figure 1.11b aurait été perçu. Or, les auditeurs ont reporté entendre une demi-gamme ascendante puis descendante ou l'inverse selon la condition (figure 1.11c), suggérant qu'ils avaient groupé les sons sur la base de leur proximité fréquentielle. Le groupement perceptif sur la base de la proximité fréquentielle semble donc primer sur celui de l'oreille de présentation, laissant penser que la latéralisation différente des signaux ne suffit pas à séparer les sons. Ce résultat a été obtenu également dans une étude faisant varier la localisation spatiale des sons à l'aide de haut-parleurs (Butler, 1979).

La présentation plus lente des sons dans l'étude de Deutsch (1975b) (IOI = 250 ms) que dans celles de van Noorden (1975, chap. 3) (IOI = 100 ms) et Hartmann et Johnson (1991) (IOI = 125 ms), pourrait expliquer la divergence des résultats. Cependant, Bregman (1990, pp. 76-77) a reproduit les séquences présentées par Deutsch (1975b), et a constaté qu'à un tempo rapide ces séquences étaient également organisées sur la base de la proximité fréquentielle. Une autre différence pourrait rendre compte de la divergence des résultats. Dans l'étude de Deutsch (1975b), les signaux sont présentés simultanément aux deux oreilles alors que van Noorden (1975, chap. 3) ainsi que Hartmann et Johnson (1991) ont fait alterner les signaux entre les deux oreilles. La diminution de la différence d'intensité entre les deux oreilles induite par la présentation simultanée des signaux aux deux oreilles, a pu diminuer la différence de localisation perçue et ainsi favoriser le groupement par proximité fréquentielle.

Une étude ultérieure de Deutsch (1979) a effectivement montré qu'un pattern mélodique présenté avec un IOI de 130 ms, dont les notes étaient présentées à une oreille puis à l'autre de façon aléatoire, ne pouvait pas être intégré. La fission perceptive des sons composant ce pattern était cependant supprimée si un son grave continu était présenté simultanément à l'oreille contralatérale. L'ajout de ce "bourdon" avait pour effet de diminuer la différence d'intensité entre les deux oreilles, et ainsi d'affaiblir les indices spatiaux qui conduisaient à la ségrégation perceptive.

Récemment, Rogers et Bregman (1998) ont confirmé le rôle des indices spatiaux dans la formation des flux auditifs, et ils ont révélé en particulier le rôle de la progression des changements de localisation spatiale. Les auteurs ont examiné l'effet cumulatif de la fission (Bregman, 1978a), c'est-à-dire l'augmentation de la tendance qu'une séquence dans laquelle alternent deux sons aigu et grave se scinde lors d'une écoute prolongée. Ils ont présenté aux auditeurs deux séquences cycliques de type ABA-ABA, une séquence dite inductive et une séquence test, de façon continue. Ils ont étudié l'effet des changements de localisation spatiale des sons entre la séquence inductive et la séquence test sur l'augmentation de la tendance que manifestait la séquence test à se scinder. Ces changements étaient induits par des différences interaurales de temps d'arrivée (expérience 1), d'intensité (expérience 2), des différences de position des haut-parleurs (expérience 3). La quatrième expérience évaluait l'effet de

changements globaux d'intensité. Les auteurs ont formulé l'hypothèse que des changements brusques de localisation spatiale réinitialiseraient ce phénomène cumulatif, alors que des changements de même ampleur mais graduels ne devraient pas le contrecarrer.

Dans les quatre expériences menées, une séquence inductive composée de 12 cycles ABA- était présentée suivie sans pause d'une séquence test composée de 3 cycles. La durée des sons était de 100 ms ainsi que la durée du silence séparant les triplets. La séquence inductive durait donc 4.8 sec et la séquence test 1.2 sec. La tâche de l'auditeur consistait à essayer d'entendre le galop (cohérence temporelle), et de signaler si à la fin de la séquence ils y parvenaient toujours. Leurs jugements étaient effectués sur une échelle numérique en 8 points, "1" signifiant qu'ils avaient perçu une fission de façon certaine et "8" qu'ils avaient perçu le galop de façon sûre. Différentes conditions expérimentales ont été élaborées variant le degré de changement de localisation entre la séquence inductive et la séquence test. Soit il n'y avait pas de changement, dans les deux types de séquence inductive et test les sons étaient localisés à gauche. Soit il y avait un changement graduel, les sons de la séquence inductive localisés à droite étaient progressivement "déplacés" vers la gauche. Ce même changement pouvait être soudain, les sons de la séquence inductive étaient localisés à droite alors que les sons de la séquence test demeuraient à gauche. Il pouvait être toujours aussi soudain mais de moins grande importance si la séquence inductive était perçue au centre. Le changement de localisation pouvait également être aléatoire. Enfin une condition contrôle dans laquelle la séquence inductive était remplacée par du bruit était également proposée, Bregman (1978a) ayant montré que l'effet du silence et du bruit était équivalent et réinitialisait l'effet cumulatif observé. La différence de fréquence des sons aigu et grave variait dans une gamme allant de 5 à 14 dt. La variation affectée à chaque essai dépendait de la réponse donnée par l'auditeur. S'il répondait avoir perçu le galop, la différence de hauteur entre les sons était augmentée de 1 dt, si au contraire il répondait ne plus pouvoir percevoir le galop cette différence était diminuée de 1 dt. La différence de fréquence conduisant à la fission de la séquence en fonction des conditions expérimentales, constituait un indicateur de la tendance de la séquence à se scinder.

Dans la première expérience la différence de localisation était induite par un changement interaurale de temps d'arrivée de la séquence aux deux oreilles. Un décalage de 0.7 ms entre l'oreille droite et l'oreille gauche était appliqué de façon à ce que la séquence inductive soit latéralisée à gauche. Les résultats ont révélé que la différence de fréquence conduisant à la fission de la séquence alternante était plus importante lorsque le changement était brutal que lorsqu'il était graduel. Ce qui signifie qu'un changement brutal de latéralisation a "réinitialisé" le processus de cumulation. La différence de fréquence était de 11.5 dt valeur proche de celle obtenue dans la condition contrôle où la séquence inductive avait été remplacée par du bruit (13.3 dt).

La deuxième expérience faisant varier cette fois la différence interaurale d'intensité (10 dB) a conduit à des résultats similaires. Un changement soudain avait pour conséquence de contrecarrer l'effet cumulatif de la fission, alors qu'une même variation mais graduelle n'avait

pas cet effet. Un changement brusque d'intensité globale testé dans la quatrième expérience ne contrecarrait pas le phénomène cumulatif, sauf pour un accroissement brutal pouvant être interprété par le système auditif comme la survenue de nouveaux événements. Des données récentes ont par ailleurs montré qu'une augmentation d'intensité était surestimée par rapport à une diminution de même ampleur (Canévet & Scharf, 1990 ; Neuhoff, 1998).

Les résultats des expériences 2 et 3 simulant à l'aide de paradigmes d'écoute dichotique une différence de localisation spatiale, ont été confortés par une troisième expérience dans laquelle les changements de localisation étaient de réels changements de position des haut-parleurs dans l'espace. Cette troisième expérience a donc permis de conclure que des changements soudains de localisation spatiale supprimaient l'augmentation de la tendance qu'une séquence alternante se scinde au cours du temps, alors que des changements graduels ne contrecarraient pas cet effet cumulatif.

1.2.3.2 La différence d'intensité

Dowling (1968) a constaté que la séparation perceptive de deux mélodies intercalées était favorisée si les mélodies étaient diffusées à différents endroits de l'espace, mais aussi si elles étaient jouées à des niveaux sonores différents.

Hartmann et Johnson (1991, expérience 1 condition 4) ont quantifié la différence de niveau sonore nécessaire pour identifier deux mélodies familières intercalées. Ils ont montré qu'une différence de 8 dB entre les deux mélodies intercalées permettait aux auditeurs d'identifier 80 % des mélodies. Ces performances suggèrent que la différence d'intensité permet de séparer perceptivement ces mélodies, mais elles demeurent cependant inférieures à celles obtenues pour une différence de fréquence moyenne de 12 dt, une différence d'oreille de présentation, ou une différence de timbre.

Van Noorden (1977) a mesuré par une méthode d'ajustement la différence de niveau en décibel qui conduisait à la fission d'une séquence cyclique de type ABAB (figure 1.1b). Dans cette séquence, les sons A et B avaient une fréquence identique de 1000 Hz et le son B était présenté à 35 dB SL (35 dB au-dessus du seuil absolu). L'auditeur devait ajuster le niveau du son A de telle sorte que la séquence se scinde, et qu'il entende séparément les sons B sur lesquels il focalisait son attention. Deux auditeurs dont l'auteur ont réalisé six ajustements par tempo (trois pour un niveau de A supérieur à celui de B, et trois lorsque c'était l'inverse). Dix IOI étaient testés dans une gamme comprise entre 43 et 800 ms. Trois zones ont été distinguées en fonction de l'IOI. Lorsqu'il était compris entre 100 et 400 ms, le seuil de fission était de 3-4 dB, valeur inférieure au niveau nécessaire à une fission mélodique (Hartmann & Johnson, 1991). Il était indépendant du tempo et symétrique (même valeur lorsque A est plus fort que B ou l'inverse), comme ce qui avait été trouvé pour une différence de fréquence (van Noorden, 1975, expérience 2). Pour des IOI supérieurs à 400 ms le seuil de fission augmentait avec l'IOI

et demeurait symétrique. Pour des IOI inférieurs à 100 ms, il augmentait avec la rapidité de la séquence quand A était plus fort que B, alors qu'il diminuait lorsque le niveau de A était plus faible que celui de B.

La différence d'intensité des sons successifs conduit donc à une fission perceptive qui semble avoir des propriétés proches de celles observées dans le cas d'une fission liée à une différence de fréquence (van Noorden, 1975, 1977). Ces études suggèrent donc que la différence d'intensité des sons successifs constitue également un indice sur lequel se base le système auditif pour organiser perceptivement une séquence.

1.3 Principe général

1.3.1 Transformation progressive des événements émis par une même source sonore

Les études décrites précédemment révèlent la contribution de nombreux paramètres acoustiques et perceptifs dans la ségrégation auditive : la différence de fréquence de deux sons purs en interaction avec la rapidité avec laquelle ils alternent (Miller & Heise, 1950 ; van Noorden, 1975), la différence de hauteur et de composition spectrale de sons complexes (Bregman et al., 1990 ; Hartmann & Johnson, 1991 ; McAdams & Bregman, 1979 ; Singh, 1987 ; Vliegen & Oxenham, 1999), la dissemblance de timbre sur les dimensions spectrale mais aussi temporelle et spectro-temporelle (Iverson, 1995 ; Singh & Bregman, 1997) ainsi que la différence d'intensité et de localisation qui constituent également des facteurs de ségrégation même si leur contribution est moindre (Deutsch, 1975a, b ; Hartmann & Johnson, 1991 ; van Noorden, 1977). Dans l'ensemble, le système auditif semble se baser sur la similarité ou la dissemblance des sons successifs pour lier ou séparer perceptivement ces événements.

Ce phénomène rappelle les lois d'organisation perceptive énoncées par les psychologues Gestaltistes dans le domaine visuel, et en particulier le groupement par proximité et par similarité (Koffka, 1935 ; Köhler, 1964). La loi de proximité postule que des éléments proches spatialement sont groupés perceptivement. Ce principe transposé dans la modalité auditive peut être appliqué soit au domaine temporel, soit au domaine fréquentiel, si l'on considère le temps ou la fréquence comme l'analogue de l'espace en vision. On regroupe en effet des sons proches temporellement, mais également des sons purs de fréquences proches (Bregman, 1990). Cependant, on peut considérer que le regroupement sur la base de la proximité fréquentielle constitue un cas particulier de regroupement des événements sur la base de leur similarité, similarité fréquentielle pour des sons purs et de façon plus générale similarité spectro-temporelle, s'appliquant à la fois à des sons purs et à des sons complexes.

La loi de bonne continuité pourrait également rendre compte de ces mécanismes de formation des flux auditifs. Cette loi prédit que des événements qui suivent la même trajectoire

sont regroupés en une unité perceptive. Or, Bregman et Dannenbring (1973) ont montré que le fait d'introduire un glissement de fréquence entre des sons successifs avait pour conséquence de contrecarrer la ségrégation perceptive opérée sur la base de la dissemblance fréquentielle de sons espacés temporellement. L'ajout d'un glissando même partiel pointant la fréquence d'un son vers celle du suivant, réduisait également la tendance de la séquence dont les fréquences aiguës et graves alternaient à se scinder. La continuité du signal acoustique affecte donc la formation des flux auditifs, suggérant que le système auditif serait en mesure de suivre la trajectoire fréquentielle des sons et que la présence d'un pointeur suffirait à prévoir la fréquence de l'événement suivant. L'interprétation des résultats de cette étude a été par la suite sujette à caution. Bregman (1990, p. 135) souligne en effet le fait que l'ajout d'un glissement de fréquence partiel a également eu pour conséquence de rapprocher la fréquence des sons successifs. Il est donc impossible de distinguer l'effet de la continuité de celui de la proximité fréquentielle. McPherson, Ciocca et Bregman (1994) ont récemment examiné la part respective de ces deux facteurs. Ils ont montré que deux glissements de fréquences qui s'entrecroisaient étaient perçus comme tel si leur taux de changement en fréquence (pente des glissandi) différait. Au contraire, si ces patterns présentaient le même taux de variation, le système auditif segmentait la séquence sur la base de la proximité en fréquence en formant deux flux juxtaposés, l'un regroupant la gamme de fréquences supérieures, l'autre la gamme de fréquences inférieures.

Plusieurs lois peuvent donc décrire ce phénomène de formation des flux auditifs : le groupement par proximité, similarité et continuité. Elles présentent l'avantage d'être suffisamment générales pour pouvoir s'appliquer au groupement perceptif visuel et auditif. Néanmoins, elles demeurent descriptives et ne fournissent aucune explication du phénomène. De plus, prises indépendamment ces lois ne peuvent rendre compte de la globalité et de la dynamique de ce mécanisme. En effet, prenons l'exemple des lois de proximité et de similarité. Si la proximité temporelle des sons renforce le groupement perceptif lorsque ces sons sont similaires sur la dimension spectro-temporelle, son action est inverse dans le cas où les sons sont dissemblables. Deux sons proches temporellement vont donc avoir tendance à être groupés d'autant plus qu'ils sont similaires en fréquence ou en timbre, mais à l'inverse deux sons dissemblables sur ces mêmes dimensions seront d'autant plus séparés qu'ils sont proches temporellement. Ces deux facteurs, la proximité temporelle et la similarité spectro-temporelle interagissent dans la formation des flux auditifs (van Noorden, 1975).

Bregman (1994) a proposé un principe général prenant en compte l'ensemble du phénomène. Ce principe a en outre une valeur explicative puisqu'il repose sur l'idée formulée par Shepard (1981) selon laquelle le système perceptif exploiterait les régularités de son environnement pour organiser perceptivement les informations sensorielles dont il dispose. Selon Bregman (1994, p. 21), pour construire des flux auditifs le système auditif exploiterait le fait que "les propriétés d'une séquence de sons issue de la même source tendent à se modifier

lentement". Ainsi, des événements présentant des caractéristiques proches ou des propriétés qui changent progressivement vont être liés perceptivement et regroupés en un seul flux. C'est le cas d'une succession de pas, d'une série de pépiements d'oiseaux, où les caractéristiques des événements pris isolément peuvent s'avérer différentes mais dont l'évolution est relativement progressive. En revanche, une transformation soudaine des caractéristiques de ces événements est interprétée par le système auditif comme le signe de la présence d'une autre source sonore. Ceci le conduit à former plusieurs flux afin d'adapter notre représentation mentale du monde à l'existence probable de plusieurs sources distinctes. La similarité/dissemblance de deux sons successifs est donc complètement liée à la proximité de ces sons. Une même dissemblance perceptive entre deux sons successifs peut conduire, soit à la ségrégation perceptive si ces sons sont proches temporellement, soit au groupement si ces sons sont plus espacés dans le temps. L'étude menée par Rogers et Bregman (1998), décrite précédemment, a d'ailleurs montré que le changement soudain de localisation spatiale des événements contrecarrait l'augmentation de la tendance d'une séquence alternante de type ABA-ABA à se scinder (Bregman, 1978a), alors qu'une modification de même ampleur mais graduelle ne réinitialisait pas cet effet cumulatif. Ce résultat conforte l'idée que le système auditif se base sur le taux de changement des événements successifs pour inférer le nombre de sources en présence : *"[...] sudden changes, in general, tell the auditory system that new events are occurring and it should start-up a new analysis rather than update an old one"* (Rogers & Bregman, 1998, p. 1227). Ainsi, la gradualité du changement des propriétés des sons successifs le conduit à inférer la présence d'une seule source sonore qui évolue dans le temps, alors qu'un changement soudain signale la survenue d'une nouvelle source.

1.3.2 La complémentarité psychophysique

Ce principe énoncé par Bregman (1994) repose sur une hypothèse séduisante formulée par Shepard (1981) selon laquelle le système perceptif se serait réglé sur les régularités de notre environnement. Notre système perceptif est le résultat d'une longue évolution biologique, soumis aux contraintes de la sélection naturelle. Cette évolution aurait abouti au développement d'un système dont la structure (l'architecture neuronale) nous conférerait une adaptation optimale à notre environnement, ce qui signifie que les propriétés fonctionnelles de ce système (les processus perceptifs) s'ajusteraient aux régularités des modifications physiques de l'environnement. Shepard (1981) propose ainsi l'idée que le système cognitif entretient avec le monde une relation fonctionnelle de complémentarité. Cette "complémentarité psychophysique" se serait établie entre les règles de formation et de transformation des processus internes et les invariances de l'environnement. Selon Shepard (1981, p. 76) la perception consiste donc à projeter une structure innée sur un monde qui a fait naître cette structure. L'auteur répond ainsi à l'interrogation formulée par un des acteurs de la théorie de la Gestalt, Koffka (1935) qui se demandait si le monde nous apparaissait de cette façon parce qu'il était physiquement comme

cela ou parce que nous le percevions ainsi, en deux points : 1) le monde est ce qu'il est parce que nous sommes ce que nous sommes, et 2) nous sommes ce que nous sommes parce que nous avons évolué dans un monde qui est ce qu'il est.

Cette théorie de la complémentarité psychophysique rejoint l'idée de couplage avancée par Varela (Varela, Thompson & Rosch, 1993 ; Varela, 1989, 1996). Considérer que le système cognitif est dans une relation de complémentarité avec l'environnement a des implications philosophiques dans la façon d'envisager la perception. Deux courants philosophiques sont classiquement distingués. La position matérialiste considère que le monde tel qu'on le perçoit existe indépendamment du fait qu'on le perçoive. Les représentations mentales des objets sont donc le reflet partiel et déformé de la réalité matérielle, position adoptée généralement dans l'étude de la perception. La position idéaliste au contraire considère que le système cognitif crée son propre monde, c'est-à-dire que l'objet n'existe que parce qu'un sujet le perçoit. Cette conception met l'accent sur l'action structurante du système cognitif soulignant le fait que la perception est un acte de création. Ces deux courants, matérialiste et idéaliste, même s'ils adoptent des positions opposées, admettent tous les deux qu'il existe une distinction entre nous et l'univers. Varela (Varela et al., 1993 ; Varela, 1996) propose une voie moyenne qui s'affranchit de cette séparation. Il considère que le système cognitif et l'environnement font partie intégrante d'un même système, et que par conséquent "le monde et le sujet percevant se déterminent l'un l'autre". La cognition est alors envisagée comme l'émergence d'un couplage entre notre organisme et le monde. Le système cognitif donne forme à son environnement en même temps qu'il est façonné par lui-même. Ainsi les différents organismes vivant sur terre peuvent développer des aptitudes perceptives et cognitives communes puisqu'ils sont soumis aux mêmes variations physiques. Cependant, dans la mesure où ils forment autant de couplages différents, ils génèrent également des représentations distinctes de ce même monde en fonction de l'utilité qu'ils ont de ces percepts pour guider leur comportement.

Ces idées de complémentarité psychophysique ou de couplage entre le sujet percevant et le monde, suggèrent que le système perceptif serait doté d'une architecture neuronale permettant la construction d'une représentation cohérente de notre environnement. L'aptitude à organiser perceptivement les informations sensorielles qui nous parviennent et en particulier la capacité de former des flux auditifs, devrait donc être présente à la naissance. On peut également imaginer que d'autres organismes, pour qui l'information acoustique est pertinente pour la survie, soient dotés de cette compétence permettant de distinguer les différentes sources sonores. Plusieurs recherches ont effectivement révélé la précocité de ces processus de formation des flux auditifs dans l'échelle à la fois ontogénétique et phylogénétique.

1.3.2.1 Une compétence innée

Demany (1982) a examiné si des bébés de 7 à 15 semaines organisaient leur environnement sonore de la même façon que les adultes, et en particulier s'ils liaient

perceptivement des sons successifs pour former des flux auditifs. Pour cela il a présenté à de jeunes adultes et à des enfants de 1.5 à 3.5 mois, une séquence cyclique composée d'un pattern de quatre notes a, b, c, d qui se répétait. Les quatre notes étaient présentées dans le même registre, ou les notes b et d s'éloignaient vers les hautes fréquences. La tâche consistait à discriminer la séquence originale de sa version inversée. La distinction de ces deux patterns était possible si l'auditeur regroupait les sons en un 1 flux (abcdabcd/cbadcbad), mais pas si la séquence était organisée en deux flux (bdbd/dbdb et acac/caca). L'intensité de cette séquence cyclique étant diminuée au début et à la fin de la séquence afin d'éviter que les sujets se basent sur le premier et/ou le dernier son pour réaliser la tâche.

Ces séquences ont donc été présentées dans un premier temps à 12 jeunes adultes. Dans un paradigme de type AAX, les auditeurs devaient déterminer si la troisième séquence présentée était identique ou différente à la première présentée deux fois. Les résultats ont montré, conformément aux prédictions, que les performances des auditeurs étaient significativement inférieures dans la condition où la séquence était organisée en 2 flux. Ces mêmes stimuli ont été ensuite présentés à 44 bébés de 7 à 15 semaines. Basée sur le principe d'habituation/réaction à la nouveauté, la technique consistait à enregistrer la durée durant laquelle les enfants fixaient une lampe. La fixation de la lampe provoquait l'émission de la séquence cyclique, et le détournement du regard de l'enfant entraînait son arrêt. Après la présentation répétitive de cette séquence durant une phase dite d'habituation, la version inversée de la séquence était présentée. La durée de fixation de la lampe durant les différentes périodes d'observation de la phase d'habituation a été comparée à celle enregistrée immédiatement après cette phase lors de la première présentation de la séquence inversée, fournissant ainsi un indice de la détection d'un changement. Elle augmentait dans le cas où les quatre sons étaient présentés dans le même registre de hauteur, attestant de l'aptitude des bébés à discriminer deux séquences sonores dont l'ordre des sons avait été inversé. Cependant, dans le cas où les sons b et d étaient éloignés vers les hautes fréquences, le temps de fixation s'est avéré équivalent voire même inférieur à celui enregistré durant la phase d'habituation. Ce résultat suggère que les enfants n'ont pas pu différencier les deux séquences dans ce cas, laissant penser que comme les adultes, les enfants de 1.5 à 3.5 mois organisent cette séquence en deux flux distincts sur la base de la différence de hauteur des sons successifs.

McAdams et Bertoncini (1997) ont examiné si cette aptitude à organiser perceptivement les événements sonores successifs était présente dès la naissance. Les auteurs ont testé si des nouveau-nés de 3-4 jours pouvaient organiser une séquence sonore en deux flux, sur la base d'une dissemblance de timbre et d'une différence de localisation spatiale. Cette aptitude a été mesurée à l'aide d'une tâche de discrimination d'un pattern mélodique. Deux types de séquences ont été présentés, une première séquence ayant une configuration dite 3/1 et une deuxième de configuration dite 2/2. Il s'agissait dans les deux cas de séquences cycliques composées de la répétition d'un pattern de quatre notes dont la hauteur montait. Dans la

configuration 3/1, les trois premières notes étaient jouées par un vibraphone et diffusées par le premier haut-parleur, et la quatrième note était jouée par une trompette et diffusée par le deuxième haut-parleur. Dans la configuration 2/2, deux notes étaient jouées par le vibraphone et diffusées par le premier haut-parleur, la première et la troisième, et les deux autres, la deuxième et la quatrième, étaient jouées par la trompette et diffusées par le deuxième haut-parleur. D'après les résultats des études examinant la contribution de différents facteurs sur la formation des flux auditifs (Bregman, 1990), les auteurs ont formulé l'hypothèse que les auditeurs devaient organiser ces séquences en deux flux sur la base de la dissemblance de timbre des instruments choisis (McAdams et al., 1995), et sur la base de l'origine spatiale différente des sons. Ainsi, ils devaient percevoir dans la configuration 3/1 un flux composé des 3 premières notes montantes et un autre de la répétition de la quatrième note, et dans la configuration 2/2 un flux composé de l'alternance de la première et la troisième note et un autre composé de l'alternance de la deuxième et la quatrième note. Les expérimentateurs ont mesuré la capacité des auditeurs à discriminer ces séquences de leur version rétrograde, c'est-à-dire un pattern mélodique ascendant d'un pattern descendant. Si les séquences présentées étaient organisées conformément à l'hypothèse des auteurs, cette discrimination devait être possible à l'écoute de la configuration 3/1, sur la base de la différence de contour du flux joué par le vibraphone. En revanche, elle ne devait pas pouvoir se faire dans la configuration 2/2 dans la mesure où les deux patterns mélodiques, le pattern ascendant et la version inversée, étaient alors parfaitement symétriques au sein de chaque flux.

La technique utilisée par les auteurs pour tester la capacité des nouveau-nés à discriminer ces deux patterns, était le paradigme de succion non nutritive. Basée sur le principe d'habituation/réaction à la nouveauté, elle consiste à enregistrer le taux de succion des nouveau-nés pendant la présentation du premier pattern jusqu'à ce que ce taux de succion atteigne une certaine valeur déterminant le critère d'habituation, puis lors de la présentation du nouveau pattern. Les taux de succion mesurés avant et après l'habituation sont alors comparés. Une augmentation de l'activité de succion non nutritive est interprétée comme une réaction à la nouveauté, attestant de la détection du changement de pattern par le nouveau-né.

La première étape de l'étude a consisté à s'assurer de l'aptitude de nouveau-nés de 3-4 jours à différencier un pattern mélodique ascendant d'un pattern mélodique descendant dans la configuration 3/1. Les deux premières expériences ont donc été menées à cette fin, et ont conduit à réviser les caractéristiques des séquences présentées. Le tempo de la séquence a été ralenti, passant de 10 à 5 sons/sec, et les intervalles de hauteur ont été augmentés, passant de 2 à 5 dt, révélant les limites des nouveau-nés dans la résolution temporelle et/ou fréquentielle pour discriminer un pattern mélodique. Ensuite, les taux de succion de vingt nouveau-nés à qui le pattern mélodique inversé avait été présenté après la phase d'habituation (groupe expérimental), et de 16 nouveau-nés à qui le même pattern était présenté durant les deux phases (groupe contrôle), ont été enregistrés (expérience 3). Ils étaient en moyenne significativement supérieurs dans le groupe expérimental par rapport au groupe contrôle lorsque la configuration

3/1 était présentée. Cette différence ne s'observait pas en revanche pour la configuration 2/2. Par ailleurs une expérience menée auprès de 12 adultes qui devaient juger si les patterns ascendants et descendants des deux types de séquences étaient identiques ou différents (expérience 4), a confirmé que cette discrimination était possible dans la configuration 3/1 et pas dans la configuration 2/2. Cette étude suggère que les nouveau-nés de 3-4 jours organisent perceptivement une séquence sur la base de la dissemblance de timbre et de la différence de localisation spatiale des événements successifs, comme le font les adultes.

Ces processus de formation des flux auditifs sont donc opérationnels à la naissance (Demany, 1982 ; McAdams & Bertoncini, 1997). Le nouveau-né utilise des indices tels que la différence de fréquence, de timbre, de localisation spatiale et peut être d'autres restant à déterminer, pour séparer perceptivement les sons n'appartenant pas aux mêmes sources sonores de l'environnement. Dès la naissance notre système perceptif semble ainsi être "configuré" pour exploiter les régularités de l'environnement, ici le fait que des sons émis par un même objet en vibration partagent des caractéristiques communes.

1.3.2.1 Une compétence adaptative

L'aptitude à organiser perceptivement les informations sensorielles qui nous parviennent est une compétence particulièrement importante pour la survie d'un organisme dans la mesure où elle conditionne la reconnaissance et l'identification des objets. Dans le cas de l'audition, pouvoir séparer par exemple le bruit des pas d'un prédateur arpentant une forêt du bruissement créé par le mouvement des arbres, et suivre ce flux pour évaluer si l'animal s'approche ou s'éloigne, permet à l'organisme d'évaluer le danger et d'adopter un comportement de fuite si la situation le nécessite. Construire une image auditive des sources sonores et suivre leur évolution dans le temps est un avantage adaptatif qui laisse penser que de nombreuses espèces en soient dotées.

Hulse et MacDougall-Shackleton (Hulse, MacDougall-Shackleton & Wisniewski, 1997 ; MacDougall-Shackleton & Hulse, à paraître) ont examiné cette compétence chez l'étourneau (Sturnus vulgaris). La première étude a été menée par Hulse, MacDougall-Shackleton et Wisniewski (1997) qui ont étudié l'effet "cocktail party" (Cherry, 1953) chez les oiseaux. Ils ont examiné la capacité des étourneaux à extraire un chant d'oiseau appartenant ou non à la même espèce dans une mixture composée de la superposition d'échantillons de 2 à 4 chants d'oiseaux d'espèces différentes. A l'aide d'une technique de conditionnement opérant, les expérimentateurs ont appris aux oiseaux à discriminer deux mixtures sonores composées chacune de deux chants d'oiseaux différents superposés. Puis, ils ont étudié la façon dont cet apprentissage allait être transféré sur d'autres types de mixtures. Les étourneaux se sont avérés capables de discriminer de nouveaux mélanges, suggérant que les oiseaux ne s'étaient pas basés

sur des indices spécifiques aux mixtures présentées durant la phase d'apprentissage mais que des indices plus généraux étaient extraits pour parvenir à réaliser cette tâche. Afin de déterminer plus précisément les indices sur lesquels les étourneaux se basaient pour discriminer et catégoriser les mélanges de chant d'oiseaux, les auteurs ont examiné comment ils allaient discriminer et catégoriser les chants des oiseaux présentés isolément. Encore une fois les étourneaux ont été capables de catégoriser ces extraits, suggérant qu'ils avaient pu avoir accès à ces chants isolés à l'écoute des mixtures. Une expérience supplémentaire a montré que ces mêmes extraits présentés dans un "bruit de fond", s'apparentant à un concert matinal d'oiseaux, étaient également discriminés. Ces résultats suggèrent que les étourneaux ont pu extraire les chants d'oiseau présentés dans la mixture sonore et se baser sur ces chants isolés pour discriminer et catégoriser les mélanges présentés. Ceci signifie que les étourneaux seraient capables d'extraire un signal dans une mixture sonore, indiquant que les processus d'analyse de scène auditive étudiés chez l'être humain s'étendraient à d'autres espèces comme les oiseaux.

Dans une deuxième étude, MacDougall-Shackleton et Hulse (à paraître) ont examiné la capacité de sept étourneaux à organiser une séquence composée de l'alternance de deux sons purs de fréquences différentes en deux flux, un composé des sons aigus et un autre des sons graves. Les chants des oiseaux utilisés dans l'étude précédente ont ainsi été remplacés par des stimuli synthétiques, plus simples, qui malgré leur non validité écologique présentent l'avantage de permettre l'étude systématique des paramètres affectant ces processus de formation des flux auditifs. L'étude s'est déroulée en deux phases, une phase d'apprentissage et une phase test. Durant la phase d'apprentissage, les étourneaux ont appris à discriminer la structure temporelle d'une séquence isotone. Cette séquence pouvait avoir un rythme de galop, ou bien être isochrone avec un tempo lent ou rapide. Elle était composée de sons ayant une fréquence de 1000, 1050, 1710 ou 4538 Hz. Les oiseaux devaient donc indiquer s'il s'agissait du galop en donnant des coups de bec sur un premier levier, ou s'il s'agissait d'une séquence isochrone lente ou rapide en donnant des coups de becs sur un autre levier. Ils apprenaient ainsi à distinguer le rythme de la séquence indépendamment du tempo et de la fréquence des sons. Une fois que leurs performances de discriminations correctes étaient supérieures à 80 %, les séquences tests ont été présentées. Il s'agissait de séquences cycliques composées d'une alternance de sons aigus et graves qui avaient un rythme de galop si les sons étaient groupés en un flux, et induisaient la perception de deux flux isochrones et isotones si elles se scindaient (van Noorden, 1975) (figure 1.1a). Trois séquences tests ont été élaborées. La fréquence du son le plus grave était constante et égale à 1000 Hz, celle du son le plus aigu variait et était égal à 1050, 1710 ou 4538 Hz, produisant une différence de hauteur respectivement de 1, 9 et 26 dt. Les oiseaux ont "répondu" dans 70 % des cas que la séquence test était une séquence isochrone lorsque la différence de fréquence était supérieure ou égale à 9 dt, alors que cette réponse n'était observée que dans 25 % des cas lorsque la différence de hauteur était inférieure à 1 dt. Leurs performances de discrimination rythmique d'une séquence isotone étaient par

ailleurs maintenues durant cette phase test. Ce résultat clair et étonnant, montre que les étourneaux organisent une séquence composite en flux sonores distincts et que la ségrégation perceptive est fonction de la différence de fréquence des sons successifs comme on peut l'observer chez l'être humain. Présente chez les oiseaux, cette compétence est probablement une aptitude commune à de nombreuses autres espèces pour qui l'information acoustique est pertinente. Du fait de sa valeur adaptative, on peut imaginer que ces processus impliqués dans l'analyse de la scène auditive sont présents précocement dans l'échelle phylogénétique.

La formation des flux auditifs est donc une compétence innée et adaptative, confortant la thèse développée par Shepard (1981) de l'existence d'une complémentarité psychophysique entre le système perceptif et le monde.

L'ensemble des études présentées dans ce chapitre révèlent les multiples paramètres qui interviennent dans l'organisation perceptive d'une séquence sonore. Ces processus automatiques qui nous permettent de distinguer les différentes sources sonores de l'environnement, ont été explorés essentiellement en examinant la fission auditive d'une séquence cyclique. Cette méthode présente l'avantage d'accéder directement à l'analyse perceptive réalisée par le système auditif. Cependant, elle utilise des séquences dont la prédictibilité pourrait modifier l'analyse réalisée, et elle se base sur l'expérience que l'auditeur rapporte du bascule entre deux percepts, réponse subjective qu'il est impossible de vérifier. Les quatre premières expériences conduites dans le cadre de cette thèse et décrites dans le chapitre 4, abordent cette question de la mesure des processus primaires impliqués dans la formation des flux auditifs.

Chapitre 2

Phénomène périphérique ou central, attentif ou pré-attentif : Considérations théoriques

Ce chapitre présente les théories avancées pour expliquer le fonctionnement des mécanismes impliqués dans la formation des flux auditifs. La nature des indices affectant le groupement et la ségrégation des événements auditifs ainsi que leur puissance respective, ont amené certains auteurs à proposer que ce mécanisme pouvait être réalisé par le système auditif périphérique sur la base de la stimulation ou non d'un même "canal" (filtres auditifs et/ou oreille de présentation). D'autres au contraire attribuent une origine centrale à ces processus, mettant en avant le rôle du contexte et l'implication des connaissances dans l'analyse de la scène auditive. Leur caractère attentif et/ou pré-attentif est discuté. Et les bases neurales de ces mécanismes, peu étudiées pour le moment, sont également examinées.

Nous avons examiné dans le chapitre précédent l'influence de différents facteurs sur la formation des flux auditifs : la fréquence des sons purs, la hauteur et la composition spectrale des sons complexes, l'enveloppe d'amplitude déterminant l'attaque et la chute du son, les fluctuations spectro-temporelles, l'intensité et la localisation spatiale des sons. Dans une situation naturelle d'écoute, l'information dont dispose le système auditif est redondante, et la plupart de ces indices sont présents et convergent vers une même analyse. Par exemple, lorsqu'un violoniste et un pianiste jouent une pièce musicale, le système auditif peut se baser à la fois sur la différence de hauteur, de timbre (sur les trois dimensions perceptives), de localisation spatiale et d'intensité, pour isoler les parties jouées par chacun de ces deux

instruments. Tous ces facteurs vont le conduire à inférer la présence de deux sources sonores différentes, et l'amener à construire deux flux auditifs. En laboratoire, nous avons la possibilité d'isoler chacun de ces paramètres et d'évaluer leur contribution respective. Comme nous avons pu le constater dans le chapitre précédent, ces différents paramètres n'ont pas tous la même puissance pour induire la fission perceptive d'une séquence. Alors que l'intensité, l'enveloppe temporelle, la durée des sons n'interviennent que de façon mineure dans cette analyse perceptive, d'autres facteurs comme les indices spectraux et l'oreille de présentation semblent jouer un rôle prépondérant.

La prégnance des indices spectraux et de l'oreille de présentation a conduit certains auteurs à postuler que la construction des flux auditifs était réalisée par le système auditif périphérique (Anstis & Saida, 1985 ; Hartmann & Johnson, 1991 ; van Noorden, 1975). Cette conception selon laquelle les sons seraient liés perceptivement s'ils stimulent les mêmes "canaux" cochléaires (filtres auditifs et oreille), est exposée dans la première partie de ce chapitre. Les arguments expérimentaux confortant ce point de vue sont examinés, ainsi que les modèles issus de ce courant théorique. A cette conception périphérique de la formation des flux, s'oppose la théorie de l'Analyse des Scènes Auditives proposée par A. S. Bregman (1990, 1994) qui attribue une origine centrale à ces processus. L'exposé de cette théorie qui distingue les processus primaires d'analyse de scène auditive et les processus guidés par les connaissances, fait l'objet de la deuxième partie de ce chapitre. Le caractère attentif et/ou pré-attentif de ces processus est également discuté (Bregman, 1990 ; Bregman, 1994 ; Jones, 1976 ; Jones, Kidd & Wetzel, 1981 ; Jones & Yee, 1994). Enfin, la troisième partie de ce chapitre relate les études, encore peu nombreuses, qui ont exploré les corrélats neuronaux de ces mécanismes.

2.1 Origine périphérique de la formation des flux : La théorie des "canaux"

D'inspiration physiologique, la théorie des "canaux" postule que la formation des flux auditifs résulterait d'un processus périphérique. L'auditeur regrouperait les sons stimulant les mêmes "canaux" sensoriels, c'est-à-dire les mêmes filtres auditifs et la même oreille, et séparerait perceptivement ceux qui activeraient des canaux différents.

2.1.1 La cohérence temporelle et la fission : succès et échec de la détection de mouvement de hauteur

La théorie "périphérique" de la formation des flux auditifs a été initialement proposée par van Noorden (1975, 1977) à la suite d'observations empiriques concernant le rôle de la contiguïté fréquentielle des sons et de l'oreille de présentation, sur la cohérence temporelle d'une séquence cyclique de type ABAB (figure 1.1b). Van Noorden (1975, chap. 3) a examiné

la cohérence temporelle de séquences composées de sons complexes qui avaient la même hauteur mais dont les composantes fréquentielles n'étaient pas contiguës, et de sons ayant des hauteurs différentes mais présentant une contiguïté fréquentielle. Il a constaté que les sons étaient groupés perceptivement s'ils étaient contigus en fréquence plutôt qu'en hauteur (voir §1.2.2.1, pour plus de détails). Il en a alors déduit que la perception de la cohérence temporelle dépendait du recouvrement de l'activité des cellules ciliées au niveau de la cochlée : *"[...] overlapping groups of hair cells have to be excited if temporal coherence is to be heard"* (van Noorden, 1975, p. 21). De plus, il a constaté que la présentation dichotique des deux sons A et B qui alternaient tous les 100 ms, ne permettait pas de percevoir une réelle cohérence temporelle, quel que soit l'intervalle de hauteur qui les séparait. Cette nouvelle observation l'a conduit à postuler qu'une séquence alternante était organisée en un flux si deux conditions étaient satisfaites : les sons devaient être contigus en fréquence et présentés à la même oreille (van Noorden, 1975, p. 24). Selon l'auteur, ces deux propositions confortent l'idée que la condition *nécessaire* pour percevoir la cohérence temporelle d'une séquence est la contiguïté de l'activité des cellules ciliées au niveau de la cochlée. Néanmoins, cette condition est nécessaire mais pas suffisante, puisque van Noorden (1975, 1977) a constaté que deux sons de même fréquence et présentés à la même oreille, se séparaient si l'intensité des sons différait.

Si le liage perceptif des sons est dû à la stimulation des mêmes filtres auditifs, alors des sons qui stimulent des filtres différents devraient être séparés. Van Noorden (1975, expérience 1) a établi que le seuil de fission d'une séquence alternante de type ABA-ABA était de 2-3 dt environ, lorsque la fréquence du son B était constante et égale à 1000 Hz. Miller et Heise (1950) ont par ailleurs montré qu'il était proportionnel à la fréquence de référence et d'environ 15 % pour une gamme de fréquence allant de 150 à 7000 Hz, soit 2.4 dt pour une fréquence de 1000 Hz. Ces deux études s'accordent donc pour évaluer le seuil de fission d'une séquence alternante dont les sons oscillent autour de 1000 Hz, à 2-3 dt. Il se trouve que cette valeur correspond effectivement à la largeur d'un filtre auditif centré sur une fréquence de 1000 Hz, qui est de 133 Hz soit 2.2 dt (Glasberg & Moore, 1990). Celle-ci est approximée par la largeur d'un filtre rectangulaire équivalent (FRE),

$$FRE = 24.7\ (4.37\ Fc + 1)$$ (Glasberg & Moore, 1990)

où Fc désigne la fréquence centrale du filtre exprimée en kHz.

Cependant, van Noorden (1975, expériences 1 et 2) a également constaté que la fission d'une séquence alternante dépendait de ce que l'auditeur cherchait à entendre. Lorsqu'il essayait de percevoir l'alternance des sons aigu et grave, la fission irrésistible de la séquence dépendait de la différence de fréquence entre les sons en interaction avec le tempo. Pour rendre compte de l'intervention de ces deux facteurs et de leur interaction, l'auteur a postulé l'existence de

"détecteurs de mouvement de hauteur" par analogie aux détecteurs de mouvement en vision. La cohérence temporelle d'une séquence s'apparente en effet au phénomène de mouvement apparent (Bregman & Achim, 1973). Lorsque deux spots lumineux spatialement distincts s'allument en alternance, un mouvement de va et vient est perçu entre ces deux lampes. Ce mouvement apparent s'estompe au cours du temps, pour laisser la place à la perception de deux spots isolés qui s'allument en alternance. Cette fission s'observe également lors de l'écoute prolongée d'une séquence alternante (Anstis & Saida, 1985 ; Bregman, 1978a). Le mouvement apparent est expliqué en psychophysique par l'existence de détecteurs de mouvement qui suivent les changements de direction spatiale. Ce postulat est appuyé par la présence de neurones qui répondent de façon spécifique à certaines directions de mouvement. Ainsi selon van Noorden, de la même façon que le système visuel détecte ces mouvements dans l'espace, le système auditif suivrait des transitions fréquentielles. La fission auditive proviendrait alors de la limite de ces détecteurs de mouvement de hauteur à suivre des changements trop rapides (van Noorden, 1975 pp. 48-51).

Cette idée selon laquelle la cohérence temporelle s'expliquerait par l'existence de détecteurs de mouvement de hauteur et que la fission serait liée à la limite de ces détecteurs, a également été adoptée par Anstis et Saida (1985). Les auteurs ont formulé l'hypothèse que l'augmentation de la tendance qu'une séquence alternante se scinde suite à une présentation prolongée, ce que Bregman (1978a) a appelé l'effet cumulatif, était liée à l'adaptation sensorielle de ces détecteurs.

Dans une série de cinq expériences, Anstis et Saida (1985) ont examiné la diminution au cours du temps de la cohérence temporelle d'une séquence. Ils ont présenté des sons modulés en fréquence par une onde carrée ou sinusoïdale, en écoute binaurale (sauf dans l'expérience 5 où ils ont également présenté les sons en écoute dichotique). La tâche des auditeurs consistait dans les expériences 1 et 3 à signaler le nombre de flux perçus (un ou deux) en appuyant sur le bouton approprié. Dans les autres expériences (2, 4 et 5), les participants devaient ajuster le tempo de la séquence de telle sorte qu'ils perçoivent toujours la cohérence temporelle.

Dans la première expérience, les auteurs ont examiné la diminution progressive de la cohérence temporelle d'une séquence dans laquelle alternaient deux sons de fréquences égales à 800 et 1200 Hz (soit une différence de 7 dt) présentée à trois IOI différents, 63, 125 et 250 ms. Il s'agissait de sons purs modulés en fréquence soit par une onde carrée, soit par une onde sinusoïdale rendant les changements de fréquence plus progressifs. Durant les trente secondes de présentation de la séquence, les auditeurs signalaient en appuyant sur le bouton approprié s'ils percevaient une cohérence temporelle ou une fission. La probabilité d'entendre la cohérence temporelle diminuait progressivement au cours du temps. Cette diminution était d'autant plus importante que l'alternance était rapide. La cohérence temporelle était plus fréquemment perçue dans le cas d'une modulation par une sinusoïde plutôt qu'une onde carrée. Ce dernier résultat est cohérent avec celui obtenu par Bregman et Dannenbring (1973) qui ont

montré qu'un glissement de fréquence contrecarrait la fission perceptive d'une séquence alternante. Anstis et Saida (1985) ont ainsi établi que la diminution de la probabilité de percevoir une cohérence temporelle, était proportionnelle au logarithme du tempo de la séquence. Cette diminution suite à une stimulation constante et prolongée, confortait leur hypothèse selon laquelle la fission perceptive résulterait de l'adaptation sensorielle de détecteurs de modulations de fréquences.

Dans une deuxième expérience, les auteurs ont étendu les résultats à d'autres intervalles de hauteur (4, 7, 12 et 19 dt), et ont demandé aux auditeurs cette fois d'ajuster le tempo de la séquence de sorte de percevoir toujours la cohérence temporelle. Lorsque la séquence se scindait, ils pouvaient ainsi diminuer la rapidité de l'alternance initialement fixée à un IOI de 63 ms jusqu'à de nouveau percevoir la cohérence temporelle. Les résultats ont confirmé ceux de l'expérience précédente quant au rôle du tempo dans l'adaptation sensorielle, et ont montré également le rôle de l'intervalle de hauteur (ou profondeur de modulation) entre les deux sons en accord avec les résultats de van Noorden (1975).

L'effet de l'intervalle de hauteur sur l'adaptation sensorielle a été examiné de façon systématique dans la troisième expérience. Un son de 1000 Hz était présenté et la profondeur de la modulation de fréquence variait (1, 2, 4, 6, 8, ou 12 dt). Le taux de modulation était de 4 cycles par secondes correspondant à un IOI de 125 ms. Durant la présentation de la séquence, l'auditeur signalait s'il percevait une cohérence temporelle ou une fission comme dans la première expérience. Les auteurs ont ainsi montré que la probabilité de percevoir une cohérence temporelle déclinait linéairement avec le logarithme des intervalles exprimés en demi-tons. Le seuil était de 4 dt, valeur proche du seuil de fission de van Noorden (1975), mais il était fonction du tempo comme le seuil de cohérence temporelle.

Enfin, les deux dernières expériences ont examiné si cette adaptation sensorielle était périphérique ou centrale. Pour cela, ils ont testé si elle concernait spécifiquement une fréquence ou une bande critique donnée, ou bien si elle se généralisait à d'autres fréquences (expérience 4), et si l'adaptation à une oreille donnée se transmettait à l'autre oreille (expérience 5). Dans l'expérience 4, les auteurs ont donc procédé à l'adaptation des auditeurs à une modulation de fréquence autour d'une fréquence centrale donnée, avant de les tester sur d'autres fréquences. Une première modulation autour de 1000 Hz (alternance de deux sons de 944 et 1060 Hz respectivement, inclus dans un filtre centré sur 1000 Hz) était présentée aux auditeurs pendant 4 secondes, suivie d'une autre modulation de fréquence autour de fréquences choisies dans une gamme comprise entre 500 et 2000 Hz. Les auditeurs ajustaient le tempo de cette séquence afin de toujours percevoir la cohérence temporelle. Les résultats ont révélé qu'ils avaient ralenti la séquence au moment où la fréquence test se trouvait à proximité de celle à laquelle ils avaient été adaptés (1000 Hz) dans une gamme de 2 dt. Les auteurs ont représenté le seuil de cohérence temporelle en fonction du tempo (taux de modulation) et de l'intervalle entre les deux sons, et ont trouvé une sorte de courbe d'accord comparable à celle des filtres auditifs. D'après les auteurs, ce résultat révèle que l'adaptation est spécifique d'une fréquence dans une gamme

correspondant à la largeur d'un filtre (Glasberg & Moore, 1990). Dans l'expérience 5, le même paradigme a été reconduit mais cette fois le transfert interaural de ce processus d'adaptation était testé. Trois conditions expérimentales ont été élaborées : deux conditions d'écoute dichotique et une condition contrôle d'écoute binaurale. Dans les conditions d'écoute dichotique, les expérimentateurs procédaient d'abord à l'adaptation d'une oreille (droite ou gauche) pour un son modulé en fréquence, puis testaient ensuite le degré d'adaptation de l'autre oreille (gauche ou droite) pour ce même son. Les auteurs ont constaté une adaptation en écoute binaurale mais pas en écoute dichotique, suggérant qu'il n'y avait pas eu de transfert interaural de l'adaptation sensorielle. Ils en ont conclu que le processus d'adaptation des détecteurs de modulation de fréquence conduisant à fission perceptive, opérait probablement au niveau du système auditif périphérique. Cependant, ce résultat ne permettait pas de distinguer l'effet dû à l'oreille de présentation, de celui dû à la différence de localisation spatiale qui pourrait être plus central.

En résumé, cette étude montre que la probabilité de percevoir une cohérence temporelle diminue progressivement au cours du temps suite à une stimulation prolongée, en accord avec les résultats de Bregman (1978a). Cette diminution est proportionnelle au logarithme du tempo et à la profondeur de modulation (intervalles exprimés en demi-tons). Elle ne s'observe que pour la gamme de fréquences précédemment entendue correspondant à la largeur des filtres auditifs cochléaires, et ne se transmet pas à l'oreille controlatérale.

Cette conception périphérique de la formation des flux a été de nouveau avancée récemment par Hartmann & Johnson (1991). Les auteurs ont présenté aux auditeurs des mélodies familières intercalées et leur ont demandé de les identifier. Douze conditions expérimentales ont été élaborées afin d'examiner la puissance de différents indices dans la fission mélodique : les facteurs spectraux, temporels, et la latéralisation (écoute dichotique). Les résultats ont révélé que les performances d'identification des deux mélodies intercalées étaient supérieures ou égales à 80 % de réponses correctes lorsque les deux mélodies étaient transposées de 12 dt (condition 1), lorsqu'elles étaient jouées par des sons de compositions harmoniques différentes (conditions 3 et 10), et lorsqu'elles étaient présentées à des oreilles différentes (condition 2). Les facteurs temporels, comme la différence de l'enveloppe d'amplitude, la différence de durée des sons, la réverbération ainsi que la différence d'intensité, ne conduisaient qu'à des performances de l'ordre de 70 % de réponses correctes. La plus grande puissance des indices spectraux et des indices de latéralisation à induire une fission mélodique, a conduit les auteurs à en déduire que la fission était favorisée par la stimulation de "canaux" sensoriels distincts (bande critique ou oreille de présentation différente). En accord avec la position de van Noorden (1975) et celle de Anstis et Saida (1985), ils ont ainsi postulé que le système auditif liait perceptivement les sons stimulant les mêmes "canaux" auditifs et séparait ceux stimulant des "canaux" distincts.

Ce courant théorique considère donc que la formation des flux auditifs est un mécanisme réalisé par le système auditif *périphérique*. Le système auditif lierait perceptivement les sons stimulant les mêmes "canaux" sensoriels. L'intégration perceptive des événements serait liée au degré de recouvrement de l'activité des cellules ciliées au niveau de la cochlée. La fission d'une séquence dans laquelle deux sons aigu et grave alternent serait due, à la *limite* des détecteurs de transition fréquentielle à suivre des changements trop rapides (donc à la limite d'un processus normal de regroupement), et dans le cas d'une stimulation prolongée à l'adaptation sensorielle de ces détecteurs.

2.1.2 Modélisation de la formation des flux auditifs sur la base de la stimulation ou non d'un même filtre

L'idée selon laquelle la cohérence temporelle et la fission perceptive d'une séquence pourraient s'expliquer par le fonctionnement du système auditif périphérique (Anstis & Saida, 1985 ; van Noorden, 1975) a été exploitée par Beauvois et Meddis (1991, 1996) dans leur modèle computationnel de la formation des flux auditifs. Les auteurs ont développé un modèle d'inspiration physiologique qui permet de rendre compte de la cohérence temporelle et de la fission perceptive d'une séquence dans laquelle alternent deux sons purs aigu et grave (Miller & Heise, 1950 ; van Noorden, 1975), ainsi que de l'effet cumulatif du phénomène de fission (Anstis & Saida, 1985 ; Bregman, 1978a).

Le fonctionnement du système peut être décrit de la façon suivante. Le signal, une onde sinusoïdale, est traité par une série de filtres passe-bande. Le nombre de filtres créés correspond au nombre de fréquences différentes présentées au système, deux s'il s'agit d'une séquence de type ABAB (seule séquence testée), plus un filtre supplémentaire centré sur la moyenne géométrique des fréquences présentées (moyenne géométrique en raison de la configuration logarithmique des filtres sur la membrane basilaire). Ces filtres ont les caractéristiques des filtres cochléaires (Glasberg & Moore, 1983 ; Glasberg & Moore, 1990 ; Patterson & Moore, 1986) et représentent ainsi l'analyse tonotopique réalisée au niveau de la membrane basilaire. Ce signal filtré va alors être transmis à deux voies différentes : une voie dans laquelle il est préservé tel quel, que les auteurs nomment "la voie du signal filtré" et une voie où il va subir différentes transformations, appelée "la voie du niveau d'excitation". L'ensemble de ces deux voies converge vers la sortie du système où est calculée l'activation globale produite.

Dans "la voie du niveau d'excitation" la valeur du signal filtré subit diverses modifications. Tout d'abord, un seuil est appliqué, par analogie avec les mécanismes de transduction cochléaire. Ces mécanismes à l'origine de la transformation d'une onde mécanique en signal bioélectrique, sont en effet basés sur un principe de seuil. L'onde propagée sur la membrane basilaire provoque une variation de potentiel membranaire au niveau des cellules ciliées internes, et si l'amplitude de cette dépolarisation est suffisante, elle déclenche l'émission d'un potentiel d'action. Une quantité aléatoire proportionnelle à l'activation précédente du filtre

est ensuite ajoutée à ce niveau d'activation. Ce paramètre est ajustable, il constitue une modélisation simple de l'aspect stochastique des décharges neuronales. L'ajout de cette quantité a pour conséquence d'amplifier l'activation d'un filtre déjà activé, modélisant l'effet cumulatif de la fission (Anstis & Saida, 1985). Cette activation est à la fois cumulée et dissipée dans le temps, par ce que les auteurs appellent une "intégration temporelle avec fuite". En effet, lorsque le filtre est activé, les différents niveaux d'activation sont sommés dans le temps. En revanche, lorsqu'il n'est pas activé (silence), son taux d'activation chute de façon exponentielle. La constante de temps constitue le deuxième paramètre ajustable du système. Le niveau d'activation de chacun des filtres est alors comparé et celui présentant la plus forte activité est dit "filtre dominant". Le niveau d'activation de ce filtre dominant est gardé constant, alors que celui des autres filtres est divisé par deux.

La sortie du système correspond à la somme globale des activités de tous les filtres, le niveau d'activité des filtres issu de cette "voie du niveau d'excitation", ainsi que celui issu de la "voie du signal filtré". Le rapport entre cette activité globale obtenue en réponse au son de fréquence A, et celle obtenue suite à la stimulation du son B, est calculé. Si ce rapport excède une valeur critique, Z_{crit} (une valeur supérieure à 1.5 est en général choisie par les auteurs), cela signifie que l'activation produite dans les filtres par le son A est plus grande que celle produite par le son B, et donc que la séquence se scinde en deux flux. Au contraire si ce rapport est inférieur à cette valeur critique et proche de 1, cela signifie que l'activation produite dans les filtres respectifs est comparable ou bien que le filtre correspondant à la moyenne géométrique de la fréquences des sons A et B est dominant. Dans ce cas, la cohérence temporelle est alors inférée. Cette valeur critique constitue le troisième paramètre ajustable du système. Dans ce modèle, le nombre de flux perçus est donc fonction du rapport de l'activation des différents filtres auditifs produit par les différents sons. Si l'activation des filtres induite par la présentation de chacun des deux sons est équivalente ou bien si l'activation du filtre centré sur la moyenne géométrique des fréquences des deux sons est dominante, le système en déduit qu'il y a cohérence temporelle. Si au contraire l'activation d'un des filtres correspondant à un des deux sons est dominante, alors cela signifie que la séquence est organisée en deux flux distincts.

Les auteurs ont pu ainsi retrouver les seuils de fission et de cohérence temporelle trouvés par van Noorden (1975) et répliqués par eux-mêmes (Beauvois & Meddis, 1991), ainsi que l'effet cumulatif de la fission (Anstis & Saida, 1985 ; Beauvois & Meddis, 1997 ; Bregman, 1978a). Le résultat de cette simulation suggère que ces phénomènes pourraient s'expliquer par des mécanismes simples réalisés au niveau du système auditif périphérique. Cependant, le modèle parvient à inférer le nombre de flux perçus à partir du rapport d'activation entre deux filtres parce qu'il s'agit de sons purs, et que par conséquent un son active un seul filtre. Néanmoins, il ne devrait pas pouvoir réaliser cette même analyse dans le cas de sons complexes composés de plusieurs fréquences différentes où cette fois un son activerait plusieurs canaux, et où un canal serait activé par plusieurs sons différents en raison de leur recouvrement spectral. Par conséquent, si ce modèle laisse entrevoir la possible contribution de l'analyse sensorielle

périphérique dans la formation des flux auditifs, il ne convainc pas sur la suffisance de cette analyse pour rendre compte de la globalité de ces mécanismes. Dans la mesure où ce modèle computationnel implémente les postulats de la théorie "périphérique" de la formation des flux auditifs, il laisse finalement entrevoir la limite de cette conception pour expliquer l'heuristique déployée par le système auditif pour construire des flux auditifs.

Un modèle mathématique des mécanismes mis en jeu dans la formation des flux auditifs, qui à l'inverse de celui de Beauvois et Meddis (1991, 1996) n'est pas intégré dans un contexte physiologique défini, a été proposé par McCabe et Denham (1997). Dans ce modèle, l'organisation perceptive de la séquence en un ou deux flux est également liée à la comparaison de l'activation des différents filtres auditifs. Cependant, le niveau d'activation du filtre dominant émerge de l'inhibition latérale de l'activité des filtres voisins. Ainsi, aucun coefficient arbitraire n'est appliqué, contrairement à l'opération proposée par Beauvois et Meddis (1991) qui consiste à diviser par 2 l'activité des filtres non dominants. Le signal acoustique est représenté de façon tonotopique sur deux réseaux de neurones en interaction. L'activation dans les filtres est calculée de façon similaire à la méthode employée par Beauvois et Meddis (1991). Le pattern d'activité produit sur ces deux réseaux de neurones liés par des connexions excitatrices et inhibitrices, est comparé au pattern d'entrée. La corrélation entre l'activité produite par un son donné sur une des deux couches et sa représentation tonotopique initiale, permet de déterminer sur quelle couche le son est représenté. Une mesure du degré de cohérence est alors obtenue sur la base de la comparaison des corrélations successives des différents sons avec ces deux couches de neurones. Ce modèle fonctionnel se base sur la représentation tonotopique du signal et sur les connexions excitatrices et inhibitrices entre les différentes unités, pour rendre compte de la formation des flux auditifs. Il reproduit la cohérence temporelle et la fission perceptive de séquences cycliques dans lesquelles alternent deux sons A et B, comme le modèle de Beauvois et Meddis (1991, 1996), mais aussi l'effet de contexte dans la formation des flux auditifs (Bregman & Rudnicky, 1975).

Ces deux systèmes, l'un d'inspiration physiologique modélisant le fonctionnement du système auditif périphérique (Beauvois & Meddis, 1991, 1996), et l'autre fonctionnel et plus abstrait (McCabe & Denham, 1997), reposent essentiellement sur le postulat selon lequel la cohérence temporelle est due à l'activation des mêmes filtres et la fission à l'activation de filtres distincts. La simulation s'avère satisfaisante pour rendre compte de la cohérence temporelle ou de la fission d'une séquence simple composée de deux sons purs. Néanmoins, ces deux systèmes devraient trouver leurs limites pour prévoir l'organisation perceptive de séquences composées de sons complexes, situation qui n'a d'ailleurs pas été testée. Ils devraient également présenter des difficultés pour rendre compte de l'analyse perceptive de séquences constituées de plus de deux sons de fréquences différentes, à l'exception peut être du modèle de McCabe et

Denham (1997) qui par le jeu d'interactions entre les couches de neurones parvient à intégrer les sons avoisinants dans la construction des flux auditifs.

En implémentant les assertions sur lesquelles les tenants de la théorie périphérique de la formation des flux se basent (Anstis & Saida, 1985 ; Hartmann & Johnson, 1991 ; van Noorden, 1975), le modèle de Beauvois et Meddis (1991, 1996) permet de montrer les limites de cette conception théorique. Les différents arguments remettant en cause la plausibilité de l'explication des mécanismes de formation des flux par une simple analyse périphérique sont présentés ci-après.

2.2 Origine centrale de la formation des flux

2.2.1 Arguments à l'encontre d'une théorie périphérique de la formation des flux auditifs

L'explication avancée par les acteurs de la théorie sur l'origine périphérique de la formation des flux auditifs (Anstis & Saida, 1985 ; Hartmann & Johnson, 1991 ; van Noorden, 1975) peut être résumée en quatre propositions :

1- Le groupement perceptif des sons se produit lorsque les mêmes canaux sont stimulés, c'est-à-dire les mêmes filtres auditifs et la même oreille.

2- La fission irrésistible d'une séquence dans laquelle alternent deux sons aigu et grave (le seuil de cohérence temporelle) est due à la limite des détecteurs de mouvement de hauteur à suivre des changements fréquentiels rapides (intégration par défaut).

3- L'augmentation de la tendance d'une séquence alternante à se scinder suite à une présentation prolongée, l'effet cumulatif, est liée à l'adaptation de ces détecteurs de mouvement de hauteur.

4- Le seuil de fission, c'est-à-dire la limite au-delà de laquelle l'auditeur ne peut plus séparer perceptivement deux sons, est lié à la largeur des filtres auditifs.

Dans cette section, nous allons voir comment trois de ces propositions ont été remises en cause par les résultats des travaux menés par A. S. Bregman et ses collaborateurs. La plupart des études ayant été relatées dans le chapitre précédent, elles sont évoquées ici de façon succincte.

2.2.1.1 Réfutation de la proposition 1 : La stimulation des mêmes filtres auditifs n'est ni une condition nécessaire ni une condition suffisante pour la cohérence temporelle

La stimulation de canaux différents n'empêche pas la cohérence temporelle. La stimulation des mêmes oreilles, donc des mêmes cochlées, n'est pas une condition nécessaire pour que le système auditif lie perceptivement les sons successifs. En effet, des sons présentés à des oreilles différentes (écoute dichotique), ne se séparent pas nécessairement (Deutsch, 1975b, 1979). Deusch (1975b) a montré que des sons étaient regroupés sur la base de leur proximité fréquentielle plutôt que sur la base de l'oreille de présentation. Dans une étude ultérieure, elle a précisé que la fission perceptive induite par l'écoute dichotique était contrecarrée par l'ajout d'un son grave continu dans l'oreille controlatérale (Deutsch, 1979). L'ajout de ce "bourdon" avait pour effet de diminuer la différence d'intensité entre les deux oreilles, et ainsi d'affaiblir les indices de localisation spatiale. Cette étude suggérait donc que la fission n'était pas liée à l'activation de populations de cellules situées dans des "canaux" différents (l'oreille de présentation ici), mais à l'origine spatiale distincte des sons présentés.

Rogers et Bregman (1998) ont récemment confirmé le rôle de la localisation spatiale dans la formation des flux auditifs. Les auteurs ont montré que la différence de localisation perçue, induite par une différence interaurale de temps d'arrivée, d'intensité ou par le positionnement différent des haut-parleurs, contrecarrait l'effet cumulatif de la fission (Bregman, 1978a). Plus précisément, ils ont mis en évidence le rôle de la gradualité du changement dans la construction des flux auditifs. Un changement brutal de localisation réinitialisait l'effet cumulatif de la fission perceptive d'une séquence de type ABA-ABA, alors qu'un changement graduel de même ampleur ne contrecarrait pas l'augmentation de la tendance à la ségrégation. Le fait que le système auditif sépare perceptivement des sons localisés à des endroits différents, laisse penser que la construction des flux auditif est au contraire un processus plus central.

La stimulation des mêmes filtres peut induire une fission. Van Noorden lui même avait reconnu que la stimulation des mêmes filtres n'était pas une condition suffisante pour percevoir une cohérence temporelle (van Noorden, 1975). En effet, l'auteur a montré qu'une séquence composée de l'alternance de deux sons de même fréquence mais de niveaux sonores différents se scindait (van Noorden, 1975, 1977). Le rôle de la différence d'intensité des sons successifs sur la ségrégation perceptive a depuis été confirmée (Hartmann & Johnson, 1991 ; Rogers & Bregman, 1998, expérience 4).

Bregman, Liao et Levitan (1990) ont également montré que le recouvrement de l'activité des filtres auditifs n'empêchait pas la fission perceptive d'une séquence. En effet, leur étude a révélé que deux sons complexes présentant un recouvrement spectral, donc un recouvrement d'activité au niveau de la membrane basilaire, se séparaient sur la base de leur différence de pic spectral ou centre de gravité spectral.

En outre, des études récentes ont suggéré que la représentation tonotopique n'était pas la seule information prise en compte par le système auditif pour organiser perceptivement une séquence sonore en flux auditifs. Elles ont mis en évidence le rôle des facteurs temporels dans la ségrégation perceptive. Les dimensions temporelle (attaque) et spectro-temporelle (flux spectral) du timbre favorisaient la fission (Iverson, 1995 ; Singh & Bregman, 1997), ainsi que la période des sons (codage temporel de la hauteur) (Vliegen et al., 1999), suggérant que la stimulation des mêmes "canaux" fréquentiels ne suffit pas à assurer la cohérence temporelle d'une séquence.

2.2.1.2 Réfutation de la proposition 3 : L'effet cumulatif de la fission perceptive n'est pas lié à l'adaptation de détecteurs de mouvement de hauteur

Bregman (1978a) a montré qu'une séquence composée d'une alternance de deux sons aigu et grave perçue initialement comme un seul flux, avait tendance à se scinder lors d'une écoute prolongée. Les expériences de l'auteur ont révélé que la fission de la séquence augmentait avec le nombre d'informations dont disposait l'auditeur sur cette séquence, ce qu'il a appelé l'effet cumulatif. Anstis et Saida (1985) ont interprété cette augmentation de la tendance d'une séquence alternante à se scinder au cours du temps comme le reflet de l'adaptation de détecteurs de transitions fréquentielles, détecteurs dont l'existence avait été postulée par van Noorden (1975). Les auteurs ont établi que cette adaptation était circonscrite à une gamme de fréquences donnée (largeur du filtre centré sur la fréquence qui module) et à l'oreille à laquelle l'alternance était présentée.

Or, Rogers et Bregman (1993, expérience 1 conditions 1-3) ont constaté que l'effet cumulatif de la fission s'observait même dans le cas où la séquence inductive était une séquence isotone. En effet, ils ont montré qu'une séquence inductive composée uniquement des sons aigus du triplet ABA, soit les sons B, favorisait la fission perceptive d'une séquence test de type ABA-ABA. Ceci suggérait que la répétition de l'alternance des deux fréquences n'était pas nécessaire pour augmenter la tendance d'une séquence alternante à se scinder. Ce résultat s'oppose donc à l'idée que la fission résulterait de l'adaptation de détecteurs de transitions fréquentielles.

De plus, les auteurs ont montré que l'effet cumulatif n'était pas provoqué par la stimulation prolongée de la même oreille (Rogers & Bregman, 1993, expérience 3). Ils ont en effet constaté que la présentation d'un galop (séquence ABA-ABA) à l'oreille droite avec des sons distracteurs proches en fréquences présentés à l'oreille gauche, ne suffisait pas à promouvoir la fission perceptive de la séquence test ABA-ABA également présentée à l'oreille droite (figure 2.1).

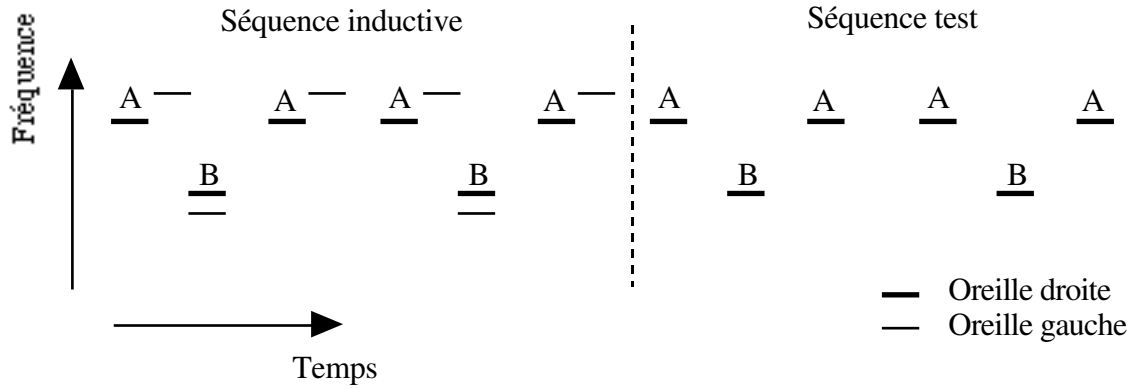


Figure 2.1. Effet de la présentation d'un galop "camouflé" sur l'augmentation de la tendance de la séquence test à se scinder (Rogers & Bregman, 1993, expérience 3). La séquence inductive est composée d'un galop ABA-ABA présenté à l'oreille droite mélangé à des sons distracteurs proches en fréquence présentés à l'oreille gauche. Si l'effet cumulatif était lié à la stimulation de la même oreille (Anstis & Saida, 1985), ce type de séquence inductive devrait induire une fission de la même façon que s'il n'y avait pas de sons distracteurs. Or le degré de fission de cette séquence est inférieur à celui constaté lorsqu'il n'y a pas de sons distracteurs et il est comparable à celui observé dans le cas où un bruit blanc est présenté à la place de la séquence inductive.

Ce type de séquence inductive conduisait en effet à un degré de ségrégation comparable à celui obtenu dans la condition contrôle dans laquelle un bruit blanc précédait la séquence test. L'effet cumulatif de la fission était donc "réinitialisé" malgré la présentation de l'alternance des sons aigu et grave à la même oreille.

L'effet cumulatif ne semble donc pas pouvoir être interprété comme le résultat d'un mécanisme périphérique d'adaptation sensorielle de détecteurs de changements de fréquence, comme le postulent Anstis et Saida (1985). Selon Bregman (1978a) le système auditif lierait perceptivement les sons successifs par défaut, et une fois qu'il a accumulé suffisamment d'informations lui permettant d'inférer l'existence de deux sources sonores, il organiserait cette séquence en deux flux. Ce type de résultat conduit une fois encore à envisager la formation des flux auditifs comme un processus central.

2.2.1.3 Réfutation de la proposition 4 : Le seuil de fission varie selon le contexte et n'est pas lié à la largeur des filtres auditifs

Une même différence de fréquence peut induire une fission ou une cohérence temporelle. L'hypothèse selon laquelle le seuil de fission serait lié à la largeur des filtres auditifs, est fortement remise en cause par le résultat de trois études ayant montré les effets de

contexte dans la formation des flux auditifs (Bregman, 1978b ; Bregman & Rudnicky, 1975 ; Heise & Miller, 1951).

Dans le chapitre précédent, nous avons vu en effet que le groupement perceptif de deux sons n'était pas seulement lié au rapport de leurs fréquences, mais également à la configuration mélodique dans laquelle ils apparaissaient. Heise et Miller (1951) ont montré que le seuil de fission ajusté par un auditeur variait en fonction du pattern présenté, unidirectionnel (ascendant ou descendant) ou bidirectionnel (ascendant puis descendant ou l'inverse), bien que les fréquences des sons qui précédaient celui à ajuster étaient identiques pour l'ensemble de ces séquences.

Bregman et Rudnicky (1975) ont également montré que les auditeurs parvenaient à extraire la paire AB et discriminer l'ordre dans lequel ces deux sons étaient présentés dans une séquence XABX, uniquement si des sons C dits "capteurs" encadrant les quatre sons (CCCXABXCC) avaient une fréquence proche de celle des sons X. Ce résultat suggérait que la présence de ces sons capteurs modifiait l'organisation perceptive de la séquence XABX présentée isolément.

Enfin, la preuve la plus flagrante a été apportée par Bregman (1978b) qui a montré qu'une même différence de fréquence entre deux sons présentés à un tempo fixe pouvait conduire, soit à la cohérence temporelle de la séquence, soit à la fission en fonction de la fréquence des sons voisins (§1.2.1.2).

Ces différentes études révèlent donc que le seuil de fission varie en fonction du contexte, et n'est pas toujours égal à 2-3 dt ou 15 % de la fréquence de référence (Miller & Heise, 1950 ; van Noorden, 1975). Par conséquent, il ne correspond pas nécessairement à la largeur d'un filtre auditif.

Pas de relation directe entre le seuil de fission et la largeur des filtres auditifs. Une étude menée par Beauvois et McAdams (1996) a montré qu'une augmentation de niveau sonore de 40 à 90 dB SPL n'affectait ni la cohérence temporelle, ni la fission d'une séquence alternante de type ABAB, malgré l'élargissement des patterns d'excitation sur la membrane basilaire induit pas ces changements.

En outre, l'étude de surdités cochléaires n'a pas permis d'établir de relation systématique entre l'augmentation de la largeur des filtres suite à cette pathologie, et un éventuel accroissement de la valeur du seuil de fission.

De Laat et Plomp (1985) ont étudié les conséquences d'une surdité cochléaire sur la perception de plusieurs mélodies présentées simultanément. Quarante cinq personnes ont participé à l'étude, trente présentaient une surdité cochléaire et quinze étaient normo-entendants. Les malentendants avaient une perte comprise entre 40 et 70 dB HL ("hearing level"), soit un seuil absolu de 40 à 70 dB au-dessus des valeurs standards établies pour chaque fréquence en

audiométrie. Les auteurs ont présenté à ces auditeurs trois séquences successives : 1) une mélodie test présentée seule, 2) une séquence A comprenant trois mélodies simultanées (une mélodie cible et deux mélodies "distractrices" présentées respectivement dans un registre de fréquence plus élevé et plus bas que la mélodie cible), et 3) une séquence B comprenant également trois mélodies superposées. La tâche de l'auditeur consistait à déterminer dans quelle séquence, A ou B, la mélodie test était présentée. La mélodie cible présentée dans les séquences A et B était soit identique à la mélodie test, soit deux notes avaient été inversées. Les auteurs ont fait varier la distance en fréquence séparant la mélodie cible des deux autres mélodies. Elle était de 45 dt au départ, puis diminuait progressivement par pas de 2 dt. Ils ont ainsi mesuré un seuil de reconnaissance de la mélodie test, c'est-à-dire la plus petite différence de fréquence nécessaire pour extraire la mélodie cible des deux autres mélodies "distractrices" et la comparer à la mélodie test présentée. Les seuils obtenus par les personnes présentant une surdité cochléaire étaient de l'ordre de 27 dt, alors que les normo-entendants parvenaient à extraire la mélodie cible pour une différence de 5 dt. Cependant, les seuils étaient très variables à l'intérieur de chacun de ces deux groupes. Les auteurs ont tenté de corréler ces seuils de reconnaissance de mélodies intercalées à la largeur des filtres auditifs de chacun des participants atteints d'une surdité cochléaire, mais leur augmentation n'était pas corrélée à la largeur des bandes critiques. Le fait que les seuils soient supérieurs chez les déficients cochléaires pourraient donc provenir d'un problème de masquage, puisque les mélodies étaient présentées simultanément. Par conséquent, cette étude ne permet pas d'établir un lien entre la largeur des filtres auditifs et le seuil de fission. Elle ne fournit donc pas d'éléments laissant penser que la formation des flux auditifs dépende de l'analyse auditive périphérique.

Grose et Hall (1996, expérience 2) ont répliqué l'expérience menée par de Laat et Plomp (1985) avec neuf personnes présentant une surdité cochléaire dont les seuils absolus étaient compris entre 30 et 50 dB HL pour une gamme de fréquences allant de 250 à 8000 Hz, et huit normo-entendants. Afin de favoriser la ségrégation perceptive des sons de la mélodie de ceux des séquences distractrices superposées, la mélodie et les séquences distractrices étaient présentées simultanément comme c'était le cas dans l'étude de de Laat et Plomp (1985), ou bien décalées de 40 ms. De plus, les sons composant ces séquences étaient modulés ou non. Les résultats n'ont pas révélé d'effet de l'asynchronisme de la mélodie cible avec les mélodies distractrices superposées, ni d'effet de la modulation des sons de la mélodie, pour l'ensemble des auditeurs. Le seuil de reconnaissance de la mélodie cible était en moyenne plus élevé chez les déficients cochléaires que chez les témoins d'environ 10 dt. Il était de 12 dt chez les sujets contrôles et de 22 dt chez les malentendants. Cependant, il y avait une grande variabilité interindividuelle à la fois chez les malentendants et chez les normo-entendants. De même que Laat et Plomp (1985), les auteurs n'ont pas pu établir de relation simple entre la baisse des performances de reconnaissance d'une mélodie cible présentée simultanément avec deux séquences distractrices, et le degré de sélectivité fréquentielle des déficients cochléaires. Si l'on

peut supposer que l'encodage dégradé de l'information auditive affecte le traitement ultérieur de ces sons et en particulier le groupement perceptif, le lien entre la largeur des filtres et la fission mélodique ne semble pas direct. En outre, la première expérience des auteurs avait par ailleurs montré que ces patients, présentant une surdité cochléaire, organisaient perceptivement une séquence alternante en deux flux pour réaliser une tâche de discrimination de durée de silence de la même façon que les témoins. Ils palliaient ainsi à la baisse d'acuité temporelle constatée pour discriminer la durée d'un silence borné par deux sons de fréquences éloignées (Grose & Hall, 1996, expérience 1). Ce dernier résultat suggère que ces mécanismes de formation des flux auditifs demeurent intacts chez les auditeurs présentant une surdité cochléaire.

Rose et Moore (1997) ont également examiné ces processus chez quatre personnes présentant une surdité cochléaire unilatérale, dix ayant une surdité bilatérale et douze normo-entendants. Ils ont mesuré le seuil de fission d'une séquence ABA-ABA dans laquelle la fréquence de A était fixe et celle de B variait, avec une procédure identique à celle utilisée par van Noorden (1975, expérience 1) (figure 1.2). Les auditeurs avaient pour consigne d'essayer d'entendre séparément le son B, dont la fréquence variait, le plus longtemps possible. Ils signalaient en appuyant sur un bouton lorsqu'ils n'y parvenaient plus, et que le son B formait avec le son A un pattern galopant. Les auteurs ont constaté que le seuil de fission n'était pas plus élevé chez les personnes présentant une surdité cochléaire unilatérale que chez les témoins, malgré une augmentation de la largeur de leurs filtres auditifs. Chez les auditeurs présentant une surdité bilatérale, la moitié avait un seuil de fission comparable à celui des normo-entendants, bien que leur perte d'audition s'élevait à 70 dB HL. Ils en ont conclu que l'augmentation de la largeur des filtres auditifs ne pouvait expliquer l'augmentation du seuil de fission, suggérant que la fission perceptive ne dépendait pas de la largeur des filtres auditifs.

Tous ces arguments convergent vers l'idée que la construction des flux auditifs est un processus plus *central*. Selon Bregman (1978b), le système auditif n'intègre pas les sons par défaut, mais ceux provenant d'une même source sonore. Les propriétés des sons, par exemple leur similarité spectro-temporelle, fourniraient des informations au système auditif quant au nombre de sources en présence. La fission perceptive ne serait donc pas la conséquence d'un mécanisme d'intégration défaillant, mais un mécanisme de *construction*, un processus à part entière de décomposition d'une mixture sonore en événements signifiants. Bregman (1978b, p. 392) définit son approche en disant : *"This approach views stream formation as an accomplishment, not a breakdown; it factors an input into streams that probably arose from different sources"*.

2.2.2. Analyse des Scènes Auditives

Selon Bregman (1990), la formation des flux auditifs s'intègre dans une analyse générale qu'il appelle l'Analyse des Scènes Auditives. Cette analyse comprend "l'ensemble des mécanismes regroupant en une unité perceptive l'ensemble des signaux provenant d'une seule source sonore de l'environnement" (Bregman, 1994, p. 12). Elle serait sous-tendue par deux processus psychologiques distincts : l'analyse primaire de la scène auditive et l'analyse guidée par les schémas.

2.2.2.1 L'analyse primaire de scène auditive

L'analyse primaire de la scène auditive est un processus *"ascendant"*. Il s'agirait d'une méthode générale et automatique de *partition* de l'entrée sensorielle en entités perceptives distinctes, effectuée sur la base d'indices acoustiques généraux. Le système auditif exploiterait les régularités acoustiques partagées par un ensemble large de sons comme, l'harmonicité d'un grand nombre de sons de notre environnement, l'asynchronisme des sources indépendantes, le changement progressif des propriétés des sons provenant d'une seule source, afin d'inférer le nombre de sources en présence (voir §1.2 pour l'exposé détaillé de la contribution des différents indices acoustiques dans l'organisation primaire). A. S. Bregman (Bregman, 1990, 1994) adopte ainsi la conception proposée par Shepard (1981) selon laquelle le système perceptif des êtres vivants se serait réglé sur les régularités du monde physique au cours de l'évolution. Le système perceptif entretiendrait ainsi une relation de "complémentarité psychophysique" avec son environnement (voir §1.3.2 pour le développement de cette idée).

Cette aptitude à séparer les informations provenant de sources sonores distinctes ferait donc partie de notre héritage génétique. Plusieurs études ont apporté des éléments en faveur de l'hypothèse d'une aptitude innée et adaptative. Demany (1982) et plus récemment McAdams et Bertoncini (1997) ont montré que la capacité de former des flux était présente très tôt chez l'Homme. MacDougall-Shackleton et Hulse (à paraître) ainsi que Hulse et ses collaborateurs (1997) ont montré que d'autres espèces comme les oiseaux, en particulier l'étourneau, en était pourvue. Récemment, une étude électrophysiologique menée par Sussman, Ritter et Vaughan (1999) a fourni une preuve de l'existence d'une composante automatique, pré-attentive dans le processus de formation des flux auditifs, étayant l'existence de cette analyse primaire de scène auditive. Cette étude est relatée dans la section suivante (voir §2.2.3).

2.2.2.2 L'analyse guidée par les schémas

L'analyse guidée par les schémas est un processus *"descendant"*. Alors que l'analyse primaire de la scène auditive a été définie par A. S. Bregman comme une méthode de *partition*

sensorielle, l'analyse guidée par les schémas serait en revanche un processus de *sélection* de l'information. Un schéma mental est une structure de connaissance abstraite construite à partir de notre expérience sur notre environnement sonore (Neisser, 1967). Cette analyse consisterait donc à sélectionner l'information auditive sur la base de l'activation de connaissances spécifiques concernant les sons et les séquences signifiantes pour l'individu (musique, parole, bruits de machine, cris d'animaux etc.). L'activation de ces connaissances pourrait être automatique. Cette situation se produit par exemple lorsque nous nous trouvons dans un environnement bruyant et que nous avons l'impression d'entendre notre nom. L'image auditive de notre nom est activée par la représentation sensorielle d'un agglomérat de sons qui lui ressemble approximativement. Néanmoins, les connaissances pourraient également être activées de façon volontaire, c'est le cas lorsque nous "cherchons à entendre" un son ou une séquence dans une "mixture" sonore.

Nous disposons actuellement de peu d'informations sur la contribution de ces processus d'analyse de scène auditive guidée par les connaissances dans la construction de la scène auditive (Bregman, 1990, chap. 8). Pourtant son importance est probablement très grande dans la perception, l'expérience perceptive d'un auditeur étant composée majoritairement de sons familiers.

Van Noorden (1975, expériences 1 et 2) a le premier mis en évidence que ce que l'on entendait dépendait en partie de ce que l'on cherchait à entendre. Il a en effet montré que la différence de fréquence conduisant à la fission d'une séquence composée de l'alternance de deux sons aigu et grave, variait selon la consigne donnée aux auditeurs. Si les auditeurs essayaient d'entendre séparément les sons aigu et grave, la séquence se scindait lorsque la différence de fréquence entre les deux sons était de 2-3 dt (expérience 1). Ce seuil de fission était considérablement augmenté si les auditeurs essayaient de percevoir l'alternance entre les deux sons. Dans ce cas, la séquence se scindait pour une différence d'environ 12 dt à un IOI de 150 ms (expérience 1). L'auteur a appelé cette limite au-delà de laquelle les auditeurs ne parvenaient plus à entendre l'alternance entre les deux sons, le seuil de cohérence temporelle. Il a ainsi défini une zone ambiguë dans laquelle pour une même différence de fréquence entre deux sons successifs et un même tempo, la séquence était organisée en un ou deux flux selon ce que l'auditeur cherchait à entendre (figure 1.3).

Dowling (1973, expérience 3) a montré que des auditeurs entraînés, à qui on présentait des mélodies familières intercalées avec des sons distracteurs présentés dans le même registre de fréquence, parvenaient à les reconnaître si le titre de la mélodie leur était donné préalablement. Ces mélodies, des comptines pour la plupart, étaient intercalées avec des sons distracteurs changeant à chaque présentation pour maximiser le contraste figure/fond. Les six auditeurs ayant participé à cette expérience devaient signaler si oui ou non la mélodie annoncée par

l'expérimentateur était présente dans le mélange. Le titre annoncé correspondait à la mélodie intercalée aux sons distracteurs (amorçage congruent) ou non (amorçage non congruent). Le mélange était présenté jusqu'à ce que la mélodie soit reconnue (17 présentations maximum). Lorsque le titre de la mélodie effectivement présentée dans le mélange était donné, les auditeurs sont parvenus à reconnaître la mélodie au bout de 3.6 présentations du mélange en moyenne. En revanche, lorsque le titre annoncé ne correspondait pas à la mélodie présentée, les auditeurs ne parvenaient pas à reconnaître la mélodie mais ne faisaient pas non plus de fausses reconnaissances. La préactivation de la mélodie familière par l'amorçage verbal de son titre a donc permis d'extraire cette mélodie du mélange alors que les sons distracteurs étaient présentés dans le même registre de hauteur.

Cette étude princeps a été reproduite par Dowling, Lung et Herrbold (1987, expérience 1) avec des sujets moins entraînés et une présentation unique de la séquence composite. Comme dans l'expérience de Dowling (1973), les auditeurs devaient juger si la mélodie entendue était ou non celle qui avait été annoncée en appuyant respectivement sur deux touches différentes du clavier. En moyenne les auditeurs reconnaissaient 78 % des mélodies intercalées lorsqu'elles étaient présentées dans le même registre que les sons distracteurs.

Les résultats de ces deux études suggèrent donc qu'il est possible d'extraire une mélodie familière intercalée à des sons distracteurs présentés dans le même registre, si l'auditeur connaît la mélodie qui est présentée dans le mélange. Cependant, sur la base de ces résultats il est difficile d'affirmer que cette extraction a été réalisée en l'absence d'analyse primaire de scène auditive. En effet, même si la différence de fréquence moyenne entre les sons distracteurs et les sons de la mélodie est nulle, cela ne signifie pas pour autant qu'il n'y ait pas de différences de fréquences locales qui permettent une fission partielle de la mélodie par une analyse primaire. Cette fission partielle pourrait alors permettre l'émergence de portions de chacune des mélodies suffisant à activer en mémoire le schéma de celle-ci.

Ces deux types d'analyse, l'analyse primaire et l'analyse guidée par les connaissances, opèrent probablement conjointement dans l'organisation perceptive d'une séquence, et leur relation (leur éventuelle interaction) demande à être précisée. Il est par conséquent difficile de les isoler totalement.

Selon Bregman (1994 p. 27), la fission irrésistible d'une séquence alors que les auditeurs cherchent à grouper les sons (SCT), résulterait de processus primaires d'analyse de scène auditive. Cependant, le caractère automatique de la fission, ne nous assure pas de l'unique implication de ce processus ascendant. Nous avons en effet souligné précédemment que les connaissances pouvaient également intervenir de façon automatique. En outre, même si la fission est involontaire et opposée aux intentions de l'auditeur, celui-ci cherche à entendre un pattern spécifique, par exemple l'alternance des sons ou le rythme de galop. Par conséquent, cette fission pourrait faire également intervenir des processus d'analyse de scène guidée par les connaissances. Singh et Bregman (1997) ont d'ailleurs constaté que le seuil de cohérence

temporelle était plus élevé lorsque la différence de hauteur entre les sons A et B augmentait, que lorsqu'elle diminuait. Quand l'auditeur commence par entendre le galop, le schéma de ce pattern rythmique est activé et permet de conserver ce percept plus longtemps que dans le cas contraire où la fission est d'abord entendue. Cette observation conforte l'idée que les connaissances de l'auditeur, que ce soit la consigne ou ce qu'il a entendu précédemment, pourraient intervenir dans la mesure du seuil de cohérence temporelle.

De façon générale, que ce soit pour déterminer le seuil de fission ou le seuil de cohérence temporelle d'une séquence alternante, l'auditeur a une consigne précise. Il tente respectivement d'essayer de séparer ou de grouper les sons. Ceci l'incite à chercher à entendre un pattern spécifique, les sons aigus ou graves ou bien l'alternance. Par conséquent, l'analyse guidée par les connaissances intervient probablement dans la mesure de ces deux seuils.

Lorsque la différence de fréquence entre les sons A et B augmente, l'analyse primaire tend vers la ségrégation de la séquence. Si l'on mesure le seuil de fission, les intentions de l'auditeur sont également de séparer les sons. Par conséquent les deux types d'analyse, primaire et guidée par les connaissances, induisent une fission. Si en revanche, on mesure le seuil de cohérence temporelle, l'auditeur cherche à grouper les sons. L'analyse guidée par les connaissances est alors en conflit avec l'analyse primaire. L'issue de ce conflit correspond au seuil de cohérence temporelle, c'est-à-dire cette limite au-delà de laquelle quelles que soient les intentions de l'auditeur, l'analyse primaire prend le dessus.

Maintenant, examinons le cas où la différence de fréquence diminue. L'analyse primaire conduit cette fois au liage perceptif des sons de la séquence. Si l'on mesure le seuil de fission, les intentions de l'auditeur qui sont de séparer les sons vont alors à l'encontre de cette analyse. L'analyse guidée par les connaissances et l'analyse primaire sont en conflit. Celui-ci s'achève par l'échec du processus d'analyse guidée par les connaissances, cette limite déterminant le seuil de fission. En revanche, lorsque l'on mesure le seuil de cohérence temporelle, les deux types d'analyse conduisent à grouper les sons.

Selon ce raisonnement, les deux seuils, les seuils de fission et de cohérence temporelle, seraient le reflet du conflit entre les processus d'analyse primaire et les processus d'analyse guidée par les connaissances. La zone ambiguë que délimitent ces deux seuils (figure 1.3) (van Noorden, 1975), correspondrait alors à une zone de flexibilité dans laquelle l'analyse guidée par les connaissances pourrait changer l'interprétation perceptive induite par l'analyse primaire. Ces deux seuils représenteraient donc les limites de cette plasticité.

2.2.2.3 Un modèle Computationnel de l'Analyse des Scènes Auditives (CASA)...

Si la théorie périphérique de la formation des flux possède un modèle computationnel implémentant les principes sur la base desquels repose cette conception (Beauvois & Meddis, 1991, 1996), il n'en n'est pas de même pour la théorie fonctionnelle proposée par Bregman (1990, 1994). A. S. Bregman esquisse lui-même les propriétés auxquelles doit satisfaire un

modèle de l'Analyse des Scènes Auditives (Bregman, 1995). Un modèle computationnel de l'Analyse des Scènes Auditives doit rendre compte du principe général selon lequel le système auditif regrouperait les sons dont les propriétés changent progressivement. La probabilité étant grande que deux événements ayant des caractéristiques spectro-temporelles très différentes proviennent de sources sonores distinctes. Ce mécanisme ascendant fournirait des contraintes au système pour réaliser l'analyse de la scène auditive. L'analyse primaire serait toutefois flexible et permettrait différentes interprétations. L'analyse finale émergerait alors du couplage entre cette analyse primaire et l'analyse guidée par les schémas, mécanisme descendant qui déterminerait en fonction du contexte, et des schémas activés l'interprétation la plus appropriée. A. S. Bregman souligne à quel point ces deux mécanismes ascendant et descendant doivent être conçus comme un tout intégré avec des échanges, et non comme deux systèmes distincts indépendants.

Un premier pas vers ce type de modèle a été récemment proposé par Ellis (1999). L'architecture "guidée par les prédictions" qu'il propose intègre un module de traitement représentant les connaissances que l'auditeur possède de son environnement sonore. Dans ce type d'approche, le signal est analysé par des filtres passe-bande. Ses caractéristiques spectrales (représentation tonotopique) et temporelles (représentation périodotopique) sont ainsi codées. Sur la base de cette analyse sensorielle, plusieurs hypothèses quant au nombre de sources et à leur nature sont formulées. Les propriétés des sons de l'environnement représentées dans un module dans lequel sont stockées ces connaissances, vont alors prédire et choisir l'interprétation la plus adéquate.

Si le principe de l'architecture se rapproche des idées développées par Bregman (1990), l'implémentation de ce type de modèle nécessite de formuler explicitement les connaissances de notre environnement sonore au système, et de ce fait reste très spécifique (signaux de parole, musique...). Alors que peut-être le fait que l'auditeur exploite les régularités de son environnement pour créer des entités perceptives est inhérente au fonctionnement d'un type d'architecture donné.

2.2.3 Attentif / Pré-attentif

La question de savoir si la partition sensorielle conduisant à la formation des flux auditifs est un processus attentif ou pré-attentif, constitue un point de désaccord entre la théorie développée par A. S. Bregman (Bregman, 1990, 1994) et l'approche plus générale du traitement de l'information auditive proposée par M. R. Jones (Jones, 1976 ; Jones & Yee, 1994) (voir Brochard, 1997, pour une confrontation entre ces deux conceptions théoriques).

Nous avons vu que selon Bregman (1990, 1994), l'analyse de la scène auditive serait le résultat de l'action conjointe de deux processus : un processus pré-attentif—l'analyse primaire,

et un processus pouvant être pré-attentif ou attentif—l'analyse guidée par les connaissances (les schémas de différentes natures pouvant être activés de façon volontaire ou non). A. S. Bregman considère donc que le mécanisme de partition sensorielle, conduisant à la construction des entités perceptives que sont les flux auditifs, est pré-attentif. L'attention permettrait de sélectionner les flux formés, c'est-à-dire des événements déjà organisés (Neisser, 1967).

M. R. Jones en revanche envisage l'attention comme un processus structurant et organisateur (Jones, 1976). Selon la théorie de "l'attention dynamique" qu'elle a développée, le traitement des événements auditifs serait fonction d'un rythme interne qui "orchestrerait" la prise d'informations. Cette conception appliquée à l'organisation auditive conduit à envisager l'intégration des événements sonores, comme une synchronisation entre cette rythmicité interne et la cadence des événements survenant dans l'environnement. Ainsi, si la périodicité attentionnelle se synchronise avec l'arrivée des événements sonores successifs, l'auditeur perçoit une séquence intégrée. Au contraire, l'asynchronisme induit la fission perceptive. L'idée qu'une synchronisation et une désynchronisation d'activité puissent être à l'origine de la formation des flux auditifs, a été exploitée dans un modèle computationnel de l'analyse primaire de la scène auditive basé sur des oscillateurs (Wang, 1996).

Cette conception envisage donc la fission auditive comme un échec de suivi des événements successifs, rappelant la limite temporelle des détecteurs de changement de fréquence postulée par les acteurs de la théorie périphérique (Anstis & Saida, 1985 ; van Noorden, 1975). Cependant, il s'agit dans le cas présent d'une limite temporelle émergeant de l'interaction entre l'attention (un processus central) et l'arrivée des événements. Cette approche théorique ne distingue pas de processus primaires et de processus guidés par les connaissances, mais intègre ces étapes de traitement dans un mécanisme global et interactif de synchronisation. La construction des flux auditifs émanerait donc de processus attentifs de nature rythmique présentant une certaine périodicité.

Les prédictions concernant l'organisation séquentielle des événements sonores, découlant spécifiquement de cette position théorique (les prédictions identiques à celles émanant de la théorie de A. S. Bregman ne sont pas présentées) sont les suivantes : 1) des séquences présentant une certaine prédictibilité temporelle ou fréquentielle, comme les séquences isochrones, isotones ou alternantes, seraient mieux intégrées que des séquences dont la rythmicité et la composition fréquentielle ne sont pas prévisibles, et 2) des sons sur lesquels l'auditeur ne porte pas son attention ne devraient pas être organisés perceptivement.

Si le rôle de la proximité temporelle des événements intervient dans la formation des flux auditifs (Jones et al., 1981), Rogers et Bregman (1993, expérience 1) ne sont cependant pas parvenus à montrer d'effet de la prédictibilité temporelle d'une séquence. Une séquence inductive de même tempo (isochrone) et dont les sons avaient la même durée que ceux composant la séquence test, ne favorisaient pas plus sa ségrégation qu'une séquence de tempo

irrégulier et dont la durée des sons variait. La régularité et la durée des sons ne semblaient donc pas promouvoir l'effet cumulatif de la fission. Les auteurs ont par ailleurs mis en évidence dans une deuxième expérience, l'effet de la densité des événements sur cet effet cumulatif. L'augmentation du nombre d'événements dans la séquence isotone augmentait la tendance de la séquence test à se scinder. Cette densité était similaire à celle de la séquence test, suggérant que l'indice pertinent était la similarité entre la séquence inductive et la séquence test.

Van Noorden (1975, chap. 5) a également constaté que le seuil de cohérence temporelle mesuré à partir d'une séquence dans laquelle les fréquences des sons successifs étaient choisies aléatoirement, était équivalent à celui trouvé à l'écoute d'une séquence cyclique de type ABA-ABA ou ABAB. Ce résultat suggère que la suppression des anticipations fréquentielles que pouvait effectuer le système auditif, n'a pas affecté l'intégration des événements sonores.

Ces deux résultats vont donc à l'encontre d'une des prédictions issue de la théorie de "l'attention dynamique" de M. R. Jones, selon laquelle la prédictibilité fréquentielle et/ou temporelle favoriserait la formation des flux auditifs (1). En outre, une étude électrophysiologique récente a révélé que des sons non focalisés semblaient être organisés perceptivement. Le résultat de cette étude relatée ci-dessous, infirme une autre prédiction de cette théorie qui postule que l'attention serait à l'origine de la formation des flux auditifs (2).

En effet, une étude menée par Sussman, Ritter et Vaughan (1999) a fourni une preuve électrophysiologique de l'existence d'une composante pré-attentive dans la formation des flux auditifs. Les auteurs ont présenté une séquence dans laquelle étaient intercalées deux triades de notes formant un pattern ascendant, l'une aiguë composée de sons purs dont les fréquences étaient respectivement de 1150, 1250 et 1350 Hz et l'autre grave composée de sons de fréquences 400, 450 et 500 Hz. Le tempo de la séquence était lent, l'auditeur entendait alors un flux composé de l'alternance de sons aigus et graves, ou rapide induisant la fission des deux triades aiguë et grave organisées dans deux flux distincts (figure 2.2a). Dans 16 % des essais, le triplet ascendant grave ou aigu était remplacé par une séquence déviante formant un pattern descendant (figure 2.2b). Les séquences étaient ainsi présentées aux auditeurs pendant qu'ils lisaient un ouvrage (écoute inattentive). Les auteurs ont enregistré grâce la technique des potentiels évoqués une composante appelée MMN *(mismatch negativity)* ou négativité de discordance, apparaissant 100 ms après la présentation du stimulus et associée à la détection automatique de changements de caractéristiques d'un stimulus auditif, au moment de la présentation du déviant et lorsque la séquence était présentée à un tempo rapide, c'est-à-dire lorsqu'elle se scindait en deux flux. En revanche, cette composante n'était pas observée lorsque la séquence était présentée à un tempo plus lent, c'est-à-dire lorsque l'auditeur ne percevait qu'un seul flux. La détection automatique du déviant attestée par l'enregistrement de cette onde de discordance suggère que malgré leur écoute inattentive, les auditeurs ont organisé

perceptivement la séquence en deux flux. Ce résultat étaye la théorie développée par Bregman (1990, 1994) qui postule que la construction des flux auditifs serait un mécanisme pré-attentif.

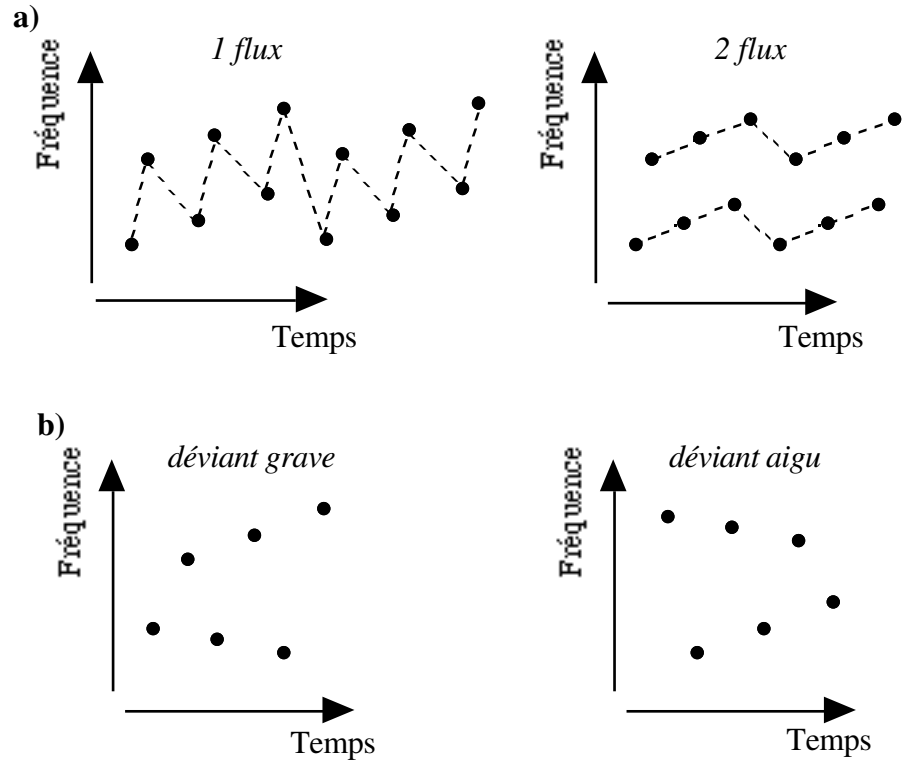


Figure 2.2. a) Séquence composée de triades de notes ascendantes intercalées, organisée en 1 flux lorsque les sons alternent tous les 750 ms, en 2 flux lorsqu'ils alternent tous les 100 ms. **b)** Dans 16 % des essais le triplet ascendant aigu ou grave est remplacé par une séquence déviante descendante. [D'après Sussman, Ritter et Vaughan (1999, p. 24)].

Le fait que la partition sensorielle soit pré-attentive ne signifie pas pour autant que des processus attentifs n'interviennent pas du tout dans l'analyse de la scène auditive. Bregman (1990, 1994) postule en effet l'existence de mécanismes d'analyse guidée par les connaissances pouvant être automatiques, mais aussi volontaires et attentifs. Des études électrophysiologiques ont d'ailleurs montré que des composantes attentionnelles intervenaient dans la formation des flux auditifs.

Sussman et ses collaborateurs (1998) ont proposé à des auditeurs deux conditions expérimentales successives : une condition d'écoute "inattentive" identique à celle mentionnée précédemment, et une condition dans laquelle les auditeurs devaient focaliser leur attention sur les sons aigus de la séquence alternante. Leur tâche était alors de détecter une séquence déviante, c'est-à-dire une séquence dont le pattern de hauteur était descendant plutôt qu'ascendant. Celle-ci était présentée à un tempo auquel on n'observait pas de fission automatique (van Noorden, 1975). Dans cette situation où la fission dépendait de ce que l'auditeur cherchait à entendre, l'enregistrement électroencéphalographique a révélé la présence

d'une MMN uniquement dans la condition "attentive". Ce résultat suggère que l'attention a orienté l'organisation perceptive de la séquence (zone ambiguë mise en évidence par van Noorden) et provoqué la détection automatique de la séquence déviante. Cette étude révèle que les processus attentionnels pourraient intervenir précocement dans le traitement, et modifier l'analyse primaire de la scène auditive, comme l'atteste la présence de cette onde de discordance à la fois pour le flux focalisé et non focalisé. Alain et Woods (1994) ont également étudié les interactions entre les processus attentifs et pré-attentifs mis en jeu dans l'organisation perceptive d'une séquence sonore. Ils ont observé que les signaux électrophysiologiques corrélés à la variable attentionnelle étaient modulés par les processus automatiques d'analyse perceptive.

2.3 Bases neurales des processus de formation des flux auditifs

2.3.1 Troubles centraux et déficit de la formation des flux auditifs

Si aucune étude n'a permis de mettre en évidence un lien entre des troubles d'origine périphérique et des déficits de la formation des flux auditifs, une recherche finlandaise récente vient en revanche de montrer qu'un trouble central du langage, la dyslexie, pourrait être lié à un déficit de l'organisation séquentielle des sons.

La dyslexie développementale se définit comme un déficit sévère et durable de l'apprentissage de la lecture (Eustache & Faure, 1996, p. 166). Plusieurs formes différentes de cette pathologie existent. L'une d'elles la dyslexie phonologique développementale, se caractérise par le fait que les enfants sont incapables de lire des non-mots alors que des mots familiers sont lus sans difficulté. Ce trouble pourrait provenir d'une déficience dans le traitement phonologique, due à un dysfonctionnement plus général de la perception de stimuli auditifs rapides. Helenius, Uutela et Hari (1999) ont examiné le lien qu'il pourrait y avoir entre ce trouble de la lecture et l'organisation perceptive des sons successifs. Deux hypothèses ont été formulées par les auteurs :

1) Ce trouble pourrait provenir d'un déficit de la perception des transitions formantiques. L'absence ou la dégradation du codage de ces transitions qui permettent de lier les segments phonétiques présentés à une cadence rapide, produirait une fission des sons successifs de la parole. Bregman et Dannenbring (1973) ont en effet montré que le fait d'introduire un glissement de fréquence permettait de contrecarrer la fission auditive d'une séquence. La fission des segments de la parole entraînerait alors une perte de l'ordre dans lequel ils ont été prononcés (Bregman & Campbell, 1971), et expliquerait ainsi les inversions de phonèmes observées chez les dyslexiques.

2) Une deuxième hypothèse, serait que la fenêtre temporelle d'intégration prise en compte par le système auditif pour organiser perceptivement les événements sonores, serait plus importante chez les personnes atteintes de ce trouble que chez les normo-lecteurs. Nous avons

vu dans le chapitre précédent que le système auditif prenait en compte les caractéristiques des deux sons pour les lier ou les séparer, mais également de celles des sons voisins dans une limite qui reste à définir (§1.2.1.2). Cette fenêtre pourrait être plus large chez les dyslexiques. Ceci les conduirait à prendre en compte un ensemble plus important d'événements, plus distants dans le temps. Ainsi, si ces événements éloignés temporellement sont distants en fréquence, le système auditif tiendrait compte de cette différence de fréquence dans l'analyse perceptive de la séquence, et cela le conduirait à séparer ces sons.

Helenius et ses collaborateurs (1999) ont donc examiné les processus de formation des flux auditifs chez treize adultes dyslexiques et 18 sujets contrôles. Ils ont présenté une séquence cyclique de type ABAB dans laquelle deux sons purs de 1000 et 400 Hz (soit une différence de 16 dt environ) d'une durée de 50 ms, alternaient. Les auditeurs devaient essayer de maintenir cette séquence intégrée (1 flux perçu), et signaler en pressant sur une touche lorsqu'ils n'y parvenaient plus et que la séquence se scindait (2 flux perçus). L'intervalle de temps qui séparait le début des sons variait dans une gamme allant de 50 à 800 ms (IOI). Les auteurs ont constaté que le seuil de cohérence temporelle était plus élevé chez les dyslexiques que chez les normo-lecteurs. La séquence se scindait en effet pour un tempo plus lent d'environ 100 ms chez les dyslexiques par rapport aux témoins. Ce seuil était corrélé au temps de dénomination des mots chez les dyslexiques. Plus la difficulté de lecture était importante, plus la tendance à scinder une séquence augmentait.

Ces résultats suggèrent que des personnes dyslexiques manifesteraient une tendance plus prononcée à séparer des événements sonores successifs que des normo-lecteurs. Cette fission accrue pourrait provenir de la prise en compte d'une fenêtre d'intégration temporelle plus importante pour l'analyse perceptive (conformément à la deuxième hypothèse formulée par les auteurs). Elle aurait pour conséquence de modifier l'ordre des phonèmes perçu et gênerait ainsi l'apprentissage de la lecture. Les difficultés que rencontrent les dyslexiques dans le traitement phonologique de la parole pourraient donc être liées au moins en partie à un déficit d'organisation séquentielle des sons.

2.3.2 Exploration neuropsychologique

Les structures cérébrales impliquées dans ces mécanismes de formation des flux auditifs sont actuellement totalement inconnues. Seules deux recherches ont, à ma connaissance, exploré l'aptitude de patients cérébro-lésés à séparer les sons provenant de sources sonores distinctes.

Efron et al. (1983) mentionnent la difficulté qu'ont des patients épileptiques ayant subi une lobectomie temporale antérieure droite ou gauche, à extraire et localiser un son de l'environnement familier présenté controlatéralement à la résection lorsqu'il est mélangé à quatre

autres sons. La résection effectuée pour pallier une épilepsie intraitable par voie médicamenteuse, épargnait les gyri de Heschl où siège le cortex auditif primaire (Galaburda & Sanides, 1980 ; Liégeois-Chauvel, Musolino & Chauvel, 1991). Cette étude suggère que l'intégrité du lobe temporal droit ou gauche est nécessaire pour isoler un son dans une scène sonore composée de plusieurs sources, lorsque celui-ci est localisé perceptivement dans l'hémi-espace controlatéral. Cependant les résultats de cette étude menée sur 31 patients ne permettent pas de déterminer si le déficit observé est lié à l'extraction et l'identification du son ou bien à sa localisation.

Clarke et ses collaborateurs (1996) ont exploré les compétences auditives non verbales de vingt patients cérébro-lésés dont le site de la lésion ainsi que son origine (accident cérébro-vasculaire, résection à des fins thérapeutiques...), étaient très variées. En plus de tests de discrimination perceptive, de reconnaissance et d'identification de sons familiers, les auteurs ont procédé à une évaluation sommaire des processus de ségrégation auditive. La tâche des auditeurs, patients cérébro-lésés et population témoin, consistait à juger du nombre de sons ou de flux perçus selon le test. Sur les vingt patients étudiés, huit présentaient un déficit dans la capacité de séparer perceptivement les sons. Parmi ces huit patients, six avaient une lésion incluant le lobe temporal gauche. Les résultats ne permettent pas globalement d'établir un lien clair entre les performances de ces patients dans les différentes tâches proposées, à savoir la tâche de reconnaissance, d'identification et de ségrégation auditive. De plus, les auteurs mentionnent la difficulté de mesurer ces processus d'organisation auditive de façon directe, en raison de problèmes de compréhension des instructions. Ils soulèvent ainsi la nécessité d'élaborer un test clinique pour que l'exploration neuropsychologique de ces processus puisse être réalisable.

Ces deux études ne nous apportent pas d'indications claires sur les bases neurales de ces processus de formation des flux auditifs. De plus, à ma connaissance aucune étude d'imagerie fonctionnelle cérébrale n'a encore jamais été réalisée sur ce sujet. Dans l'état des connaissances actuelles, il est donc délicat de formuler des hypothèses quant aux structures cérébrales impliquées dans la formation des flux auditifs ; l'heure est à l'exploration...

2.3.3 Apport théorique de l'étude des bases neurales des processus de formation des flux auditifs

Peu d'études ont donc exploré les bases neurales des processus de formation des flux auditifs. Pourtant, la mise en évidence de déficits de l'analyse de la scène auditive suite à des lésions ou résections cérébrales, et l'étude des corrélats neuronaux de ces processus par des techniques de neuroimagerie, présentent un intérêt à différents niveaux.

Tout d'abord, l'exploration neurophysiologique combinée à l'analyse fonctionnelle des processus de formation des flux auditifs, conduirait à une compréhension exhaustive de ces mécanismes (Bregman, 1991, 1996). Mettre en évidence une architecture neuronale qui manifesterait des propriétés fonctionnelles obéissant aux principes énoncés par Bregman (1994), permettrait de faire le lien entre deux niveaux d'interprétation du phénomène, symbolique et sub-symbolique. Même à un niveau plus intégré, connaître les structures activées dans différentes tâches de ségrégation auditive impliquant ou non des connaissances grâce à des techniques de neuroimagerie, permettrait de tester l'hypothèse formulée par Bregman (1990) de l'existence de deux processus fonctionnellement distincts : l'analyse primaire et l'analyse guidée par les connaissances.

En outre l'intérêt ne s'arrête pas à la connaissance approfondie de ces mécanismes. Les modèles actuels du traitement de l'information auditive considèrent que ces processus d'analyse de scène auditive jouent un rôle clé dans la perception. S'agissant de processus de construction d'"objets" sonores, ils conditionnent l'émergence des propriétés de ces objets, ainsi leur reconnaissance et leur identification. Un déficit de ces processus pourrait donc être à l'origine des agnosies auditives. En effet, l'agnosie auditive se caractérise par l'incapacité de reconnaître des événements sonores que ce soit dans le domaine du langage, de la musique ou des sons de l'environnement (Peretz, 1994). Ces troubles de reconnaissance ont été attribués jusqu'à présent soit à une atteinte de l'analyse perceptive, définissant l'agnosie aperceptive, soit à des troubles de l'activation des représentations en mémoire, définissant l'agnosie associative (figure 2.3).

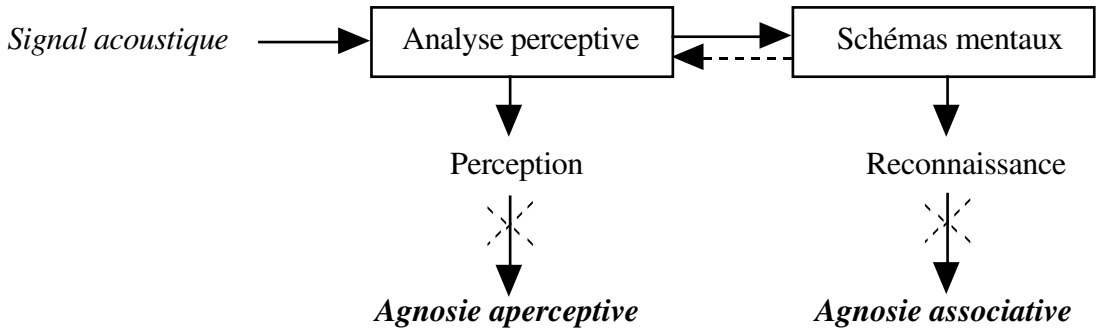


Figure 2.3. Représentation du traitement de l'information auditive aboutissant à la reconnaissance d'un événement sonore. Deux étapes sont distinguées : une étape d'analyse perceptive du signal, et une étape d'association du percept avec les connaissances antérieures de l'auditeur. De cette conception en deux étapes résulte la distinction de deux types d'agnosie. La perturbation de l'étape d'analyse perceptive conduit à une agnosie aperceptive ; le patient ne parvient plus à discriminer des événements sonores. L'incapacité à reconnaître un événement alors que le patient ne présente pas de trouble perceptif, définit l'agnosie associative. [Adapté de Peretz (1994, p. 224)].

Néanmoins, cette distinction reste très globale. En particulier dans le cas du déficit perceptif, il pourrait s'agir d'un déficit dans la construction des objets (construction des flux auditifs) ou dans le traitement ultérieur des propriétés de ces objets (traitement mélodique ou rythmique). Examiner les éventuels déficits de l'analyse de la scène auditive permettrait donc d'apporter des précisions fonctionnelles dans l'étude des agnosies auditives. En outre, cette exploration apporterait des éléments de réponse à la question de savoir si musique et langage sont deux fonctions autonomes et traitées indépendamment dès les premières étapes d'analyse de l'information sensorielle (cas d'amusie sans aphasie).

Dans ce chapitre, les différentes théories formulées pour expliquer les mécanismes impliqués dans la formation des flux auditifs ont été examinées. Les travaux menés dans le cadre de cette thèse, s'inscrivent dans l'approche développée par Bregman (1990, 1994). L'analyse primaire de la scène auditive et l'analyse guidée par les connaissances, sont étudiées respectivement dans les chapitres 4 et 5. La recherche d'éventuels déficits de ces processus d'organisation auditive, question qui nous l'avons vu reste encore peu explorée, est abordée dans le chapitre 6.

Chapitre 3

Reconnaissance de mélodies en situation monophonique et polyphonique : Composantes cognitives

Les expériences menées dans le cadre de cette thèse mesurent les processus de formation des flux auditifs par une méthode indirecte utilisant un paradigme de reconnaissance de mélodies intercalées. Une synthèse des travaux examinant les processus perceptifs et mnésiques impliqués dans la reconnaissance d'une mélodie font donc l'objet de ce chapitre. Les questions discutées par la suite dans les réalisations expérimentales sont abordées. Elles concernent la représentation des mélodies en mémoire, plus précisément les indices mélodiques encodés en mémoire à court et à long terme, et les processus attentionnels et mnésiques spécifiques à la perception et à la mémorisation de mélodies présentées simultanément avec une autre séquence.

Les recherches menées dans le cadre de cette thèse examinent les processus de formation des flux auditifs par le biais d'un paradigme de reconnaissance de mélodies intercalées. Les performances de reconnaissance d'une mélodie cible intercalée à des sons distracteurs constituent un indicateur de la façon dont le système auditif organise perceptivement une séquence (voir §1.1.2.2). En effet, un auditeur parvient à accomplir cette tâche s'il réussit à extraire la mélodie du "mélange", donc s'il organise la séquence composite en deux flux distincts. Cependant, cette méthode indirecte fait également intervenir d'autres composantes cognitives spécifiques à la tâche qui n'ont pas nécessairement un lien avec l'opération mentale

étudiée, à savoir la formation des flux auditifs. Ce point justifie la présence de ce chapitre dans lequel sont examinés les processus perceptifs et mnésiques impliqués dans la reconnaissance de mélodies, en situation monophonique lorsqu'une mélodie isolée est présentée à l'auditeur (situation expérimentale la plus répandue) et en situation polyphonique c'est-à-dire lorsque plusieurs voix mélodiques sont jouées simultanément et que l'auditeur réalise une tâche sur une de ces lignes mélodiques. Cette dernière situation a pour le moment été très peu étudiée, essentiellement par W. Jay Dowling (Dowling, 1973 ; Dowling et al., 1987) dont les études sont relatées dans le détail. Les connaissances théoriques que nous avons sur la nature de la représentation des mélodies dans ces deux types de situations sont présentées, et les spécificités attentionnelles et mnésiques de l'écoute polyphonique sont soulignées.

3.1 Reconnaissance de mélodies présentées isolément

3.1.1 Mémoire : Considérations théoriques générales

3.1.1.1 Encodage, stockage et récupération

Se souvenir de ses expériences passées suppose avoir enregistré certaines informations, les avoir conservées un certain temps et pouvoir les retrouver. Ces trois postulats définissent les trois opérations mnésiques classiquement distinguées dans l'étude de la mémoire : l'encodage, le stockage et la récupération (Baddeley, 1992). L'encodage est l'opération par laquelle une information est enregistrée sous un certain format. Concernant la mémorisation des mélodies, la question est donc de savoir sous quel format le système cognitif "enregistre" une mélodie. Le stockage désigne le processus de conservation des informations. Et la récupération est le mécanisme qui nous permet de réactualiser nos connaissances. Certaines connaissances sont accessibles par simple rappel ou reconnaissance, on parle alors de mémoire explicite. Cette terminologie renvoit à l'acte intentionnel de récupération, c'est le cas par exemple lorsque l'on retrouve l'air d'une chanson à partir de son titre. D'autres au contraire ne sont pas accessibles directement mais interviennent de façon incidente dans des traitements ultérieurs, on parle alors de mémoire implicite. Cette situation est typiquement illustrée par les paradigmes d'amorçage dans lesquels un stimulus préalablement présenté appelé l'amorce, pour lequel aucune tâche ni rétention n'est demandée, affecte le traitement ultérieur d'un stimulus test. La mémoire de l'harmonie, c'est-à-dire la connaissance des relations entre les différentes notes d'une gamme et de leur importance relative, constitue un bon exemple de connaissance musicale implicite. En effet, des auditeurs n'ayant suivie aucune formation musicale génèrent des attentes harmoniques qui guident leur écoute musicale (Bigand, 1994).

Ces trois opérations, l'encodage, le stockage et la récupération ne sont pas indépendantes. En effet, la nature de l'information encodée conditionne la durée du stockage et par conséquent sa récupération. Comme nous le verrons dans le cas de la mémorisation des

mélodies, l'auditeur n'encode pas les mêmes indices si la mélodie est stockée à court ou à long terme (Dowling, 1978 ; Dowling & Fujitani, 1971 ; Dowling & Harwood, 1986).

3.1.1.2 Modèle discret versus continu de la mémoire

La mémoire est-elle un système unitaire ? Deux conceptions théoriques classiques s'affrontent : l'une d'inspiration structurale représentée par le modèle de Atkinson et Shiffrin (1968) envisage la mémoire comme un ensemble de "registres mnésiques" distincts, l'autre fonctionnelle considère la mémoire comme le résultat d'un continuum dans le traitement de l'information ; il s'agit du modèle des "niveaux de traitement" développé par Craik et Lockart (1972), conception partagée par Crowder (1994) pour la mémoire auditive.

La mémoire compartimentée. Sur la base de la différence de durée et de capacité de stockage, Atkinson et Shiffrin (1968) ont distingué trois systèmes de mémoire différents : la mémoire sensorielle, la mémoire à court terme (MCT) et la mémoire à long terme (MLT) (figure 3.1).

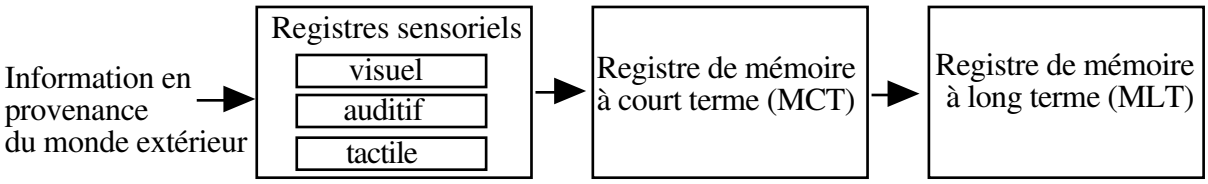


Figure 3.1. Représentation schématique du modèle modal de Atkinson et Shiffrin (1968) postulant l'existence de différents registres mnésiques par lesquels l'information transite.

Selon cette conception théorique, l'information en provenance du monde extérieur serait tout d'abord traitée en parallèle par différents registres sensoriels (visuel, auditif, tactile etc.), qui maintiendraient de façon transitoire cette information. Elle serait ensuite conduite dans le registre de mémoire à court terme présentant une capacité limitée (empan de 7 unités ± 2), et une durée de stockage relativement brève. Puis elle serait transférée dans le registre de la mémoire à long terme, qui ne présente aucune limite en termes de durée et de capacité. Baddeley (1992) a par la suite précisé la nature de la mémoire à court terme. Il a proposé le concept de mémoire de travail, consistant en un espace de mémoire de capacité limitée pouvant être alloué soit au traitement des informations, soit à leur stockage.

La conception d'une mémoire composée de registres distincts repose sur deux arguments principaux. Le premier concerne l'existence de deux types d'amnésie observés suite à des lésions cérébrales : l'amnésie antérograde se traduisant par l'incapacité de former de nouveaux souvenirs, et l'amnésie rétrograde consistant en une incapacité à se rappeler les souvenirs anciens. Cette double dissociation de déficits suggère l'existence de ces deux systèmes mnésiques distincts : une mémoire à court terme et une mémoire à long terme. Le

deuxième argument s'appuie sur les résultats d'expériences désormais classiques portant sur les effets sériels de l'oubli, c'est-à-dire le fait que les derniers et les premiers items présentés soient les mieux rappelés. Ces effets appelés respectivement, effet de récence et effet de primauté, sont expliqués selon ce modèle par le stockage des derniers items présentés dans le registre de mémoire à court terme, et des premiers dans le registre de mémoire à long terme.

La mémoire comme épiphénomène du traitement de l'information. A cette conception structurale de la mémoire, Craik et Lockhart (1972) ont préféré une conception plus fonctionnelle, et ont proposé le modèle des "niveaux de traitement". Selon ces auteurs, la durée de stockage et la récupération d'un souvenir sont directement liées à la profondeur du traitement effectué sur l'information, et donc à la nature du codage. Plus un item est traité en profondeur, plus persistante sera la trace mnésique, et meilleur sera le rappel. Par exemple, l'accès au sens d'un mot (codage sémantique) améliorera la rétention de celui-ci par rapport à la mémorisation de l'image auditive de ce même mot (codage phonologique). Les auteurs expliquent ainsi les effets de récence et de primauté par des différences dans les stratégies d'encodage de l'information par les sujets. Selon ce courant théorique, la mémoire est envisagée comme la résultante d'un processus continu de traitement de l'information plutôt que comme une série d'étapes discrètes.

3.1.2 Mémoire des mélodies : Indices encodés et durée de stockage

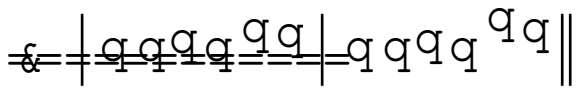
Sous quel format une mélodie est-elle mémorisée ? Ce format varie-t-il en fonction de la nature du traitement effectué sur cette mélodie et la durée de sa rétention ?

Mémoriser une mélodie ne se réduit pas à mémoriser une succession de notes sans lien entre elles. Une mélodie se définit en effet par des liens entre différentes notes desquels émergent certaines propriétés comme le contour, les intervalles, la tonalité. La contribution de ces différents indices dans la mémorisation à court et à long terme des mélodies est examinée dans les études relatées ci-dessous. Ces études explorent la mémoire mélodique explicite à l'aide de paradigmes de reconnaissance, technique la plus utilisée en raison de l'expertise musicale nécessaire pour le rappel (Crowder, 1994).

3.1.2.1 Prégnance du contour et des notes en mémoire à court terme, des intervalles en mémoire à long terme

Le contour mélodique est défini comme étant "la suite de montées et de descentes créées par les variations de hauteur des notes successives indépendamment de la taille des intervalles" [ma traduction] (Dowling & Fujitani, 1971, p. 524). On le représente par une suite de signes + et -, symbolisant respectivement la montée et la descente. Lorsque deux notes identiques se suivent, on le mentionne par le signe =, ou par un 0, ou bien on peut également ne pas le noter.

Prenons l'exemple de la chanson "Joyeux Anniversaire". La partition des deux premières phrases musicales sans le rythme original, ainsi que la formalisation correspondante du contour et des intervalles (taille absolue et relative), sont représentées ci-dessous :



Contour	[= + - + -] - [= + - + -]
Intervalles	
Taille absolue (en demi-tons)	[0 +2 -2 +**5** -**1**] -4 [0 +2 -2 +**7** -**2**]
Taille relative	[< = < >] [< = < >]

Dans cet exemple, on voit que les deux premières phrases musicales de "Joyeux Anniversaire" ont le même contour mélodique, mais que la taille absolue des intervalles est différente (notons que les intervalles contiennent l'information relative au contour). La taille relative des intervalles exprimant la relation d'ordre entre les tailles des intervalles conjoints, est néanmoins préservée dans ces deux phrases.

Dowling et Fujitani (1971, expérience 1) ont examiné le rôle du contour mélodique dans la reconnaissance immédiate de mélodies non familières atonales (stockées en mémoire à court terme). Ils ont présenté à deux groupes d'auditeurs deux mélodies atonales de 5 notes, une mélodie standard suivie après une pause de 2 secondes d'une mélodie de comparaison, et leur ont demandé de juger si ces deux mélodies étaient identiques ou différentes. Lorsqu'elles étaient différentes, la mélodie de comparaison était composée de notes et d'intervalles distincts et le contour était modifié pour le premier groupe d'auditeurs (1), alors que pour le second groupe il était identique à celui de la mélodie standard (2). La moitié des mélodies à comparer était présentée dans le même registre que celui de la mélodie standard (commençait sur la même note) et l'autre moitié était transposée dans une gamme de ± 7 dt dans le but d'examiner les stratégies de comparaison de notes. Les jugements étaient exprimés sur une échelle en 4 points "même sûr" / "même" / "différent" / "différent sûr", afin de calculer l'aire sous la courbe MOC (Memory Operating Characteristic), homologue de la courbe ROC (Receiver Operating Characteristic) pour la mémoire, donnant une estimation non biaisée de la proportion de réponses correctes (Norman & Wickelgren, 1965 ; Swets, 1973).

Les résultats reportés dans le tableau 3.1 montrent que la reconnaissance des mélodies est affectée lorsque les auditeurs ne peuvent se baser sur le contour (0.98 lorsque le contour de la mélodie à comparer est différent de celui de la mélodie standard et 0.91 lorsqu'il est identique). La diminution des performances (différence de 0.07) est amplifiée si la mélodie à

comparer est transposée (différence de 0.36), les performances sont alors considérablement altérées et proches du hasard (0.53). Ces résultats suggèrent que les auditeurs se basent à la fois sur les notes et sur le contour mélodique pour réaliser cette tâche de reconnaissance immédiate de mélodies non familières atonales, et non sur les intervalles (qui demeurent constants après transposition).

	Groupes de sujets	
	contour préservé (2)	contour non préservé (1)
Mélodie non transposée	0.91	0.98
Mélodie transposée	0.53	0.89

Tableau 3.1. Résultats extraits de l'expérience 1 de Dowling et Fujitani (1971) mentionnant les performances (aire sous la courbe MOC, 0.50 correspond au hasard) de reconnaissance immédiate de mélodies atonales non familières obtenues lorsque la mélodie à comparer à le même contour que celui de la mélodie originale (2), ou non (1), et lorsqu'elle a été présentée dans le même registre de hauteur ou bien transposée de ± 7 dt.

La dominance du contour sur les intervalles dans la reconnaissance de mélodies non familières atonales, n'est-elle pas le reflet de la difficulté qu'éprouvent les auditeurs à accéder aux intervalles de mélodies atonales pour lesquelles ils ne disposent pas de connaissances préalables ? En d'autres termes, le fait que les intervalles ne soient pas extraits, ne s'explique-t-il pas par l'atonalité des mélodies plutôt que par le fait qu'elle soit stockée en mémoire à court terme ?

Pour tester cette possibilité, Dowling (1978) a étendu l'étude précédente à des mélodies tonales. La mélodie à comparer était soit une transposition exacte d'une mélodie standard tonale soit inexacte. Lorsqu'elle était inexacte, c'est-à-dire lorsque la taille des intervalles était modifiée, la mélodie à comparer était : 1) présentée dans la même tonalité que celle de la mélodie standard et préservait le contour, 2) atonale mais préservait toujours le contour, 3) atonale et avait un contour différent.

Les performances de reconnaissance de mélodies non familières tonales étaient aléatoires (0.49) si la tonalité et le contour étaient préservés, même si les intervalles étaient modifiés (tableau 3.2). Ce résultat suggère que les auditeurs n'ont pas accès en mémoire à court terme aux intervalles, même pour des mélodies tonales. Elles s'amélioraient si les deux mélodies à comparer n'avaient pas la même tonalité, si l'une était tonale et l'autre atonale, même si le contour était préservé, suggérant que la tonalité fournit un indice pour discriminer des mélodies non familières. Ce résultat était plus prononcé chez les musiciens, pour qui les performances augmentaient de 0.31 contre 0.10 pour les non musiciens. Le fait que la tonalité soit un indice plus discriminant pour les musiciens que pour les non musiciens, est probablement lié à leurs

connaissances plus précises du système tonal. Enfin, lorsque le contour différait entre les deux mélodies à comparer en plus de la tonalité, les performances augmentaient encore de 0.05 pour les musiciens et de 0.22 pour les non musiciens, démontrant la contribution de cet indice malgré la présence d'une différence de tonalité.

	Type de mélodie de comparaison		
	(1) tonale même tonalité contour préservé	(2) atonale contour préservé	(3) atonale contour non préservé
Musiciens	0.48	0.79	0.84
Non musiciens	0.49	0.59	0.81
Dowling et Fujitani (mélodie atonale et transposée)		0.53	0.89

Tableau 3.2. Performances (aire sous la courbe MOC) obtenues dans une tâche de reconnaissance de mélodies non familières tonales lorsque la mélodie à comparer est transposée de façon inexacte mais 1) préserve la tonalité et le contour de la mélodie standard, 2) préserve le contour mais est atonale, et 3) ne préserve ni la tonalité ni le contour. Les performances des auditeurs ayant reçu ou non une formation musicale sont mentionnées. Les résultats obtenus dans des conditions expérimentales similaires par Dowling et Fujitani (1971, expérience 1) pour des mélodies atonales, figurent également dans le tableau pour comparaison. [D'après Dowling et Harwood (1986, p. 135)].

Ces résultats obtenus pour des mélodies atonales (Dowling & Fujitani, 1971, expérience 1), et étendus aux mélodies tonales (Dowling, 1978), montrent qu'à l'écoute d'une mélodie non familière nous encodons les notes, le contour mélodique et la tonalité, mais pas les intervalles.

Qu'en est-il des mélodies familières stockées en mémoire à long terme ? Dowling et Fujitani (1971, expérience 2) ont testé le rôle du contour et de la taille absolue et relative des intervalles sur la reconnaissance de mélodies familières tonales, stockées en mémoire à long terme. Ils ont présenté des mélodies familières jouées approximativement au même tempo, dans leur version originale et dans une version modifiée. Ils ont demandé aux auditeurs de désigner le titre de la mélodie parmi un ensemble de titres figurant sur une liste qui leur était proposée. Trois types de distorsion ont été appliqués à ces mélodies : 1) la taille des intervalles était modifiée mais le contour et la relation entre les tailles des différents intervalles étaient préservés, 2) les tailles absolue et relative des intervalles étaient modifiées mais le contour était préservé,

3) le contour et les intervalles étaient modifiés mais les premières notes étaient maintenues ainsi que l'harmonie implicite de la mélodie originale.

Les résultats de cette étude ont révélé que les mélodies familières présentées dans leur version originale ont été parfaitement reconnues (0.99). En revanche, les performances de reconnaissance de mélodies ayant juste les premières notes et l'harmonie implicite de la mélodie d'origine étaient aléatoires (0.28, le hasard étant 0.20). L'altération de la taille des intervalles (1) a affecté les performances de reconnaissance (0.66), et l'ajout d'une distorsion supplémentaire modifiant la relation entre la taille des intervalles des notes adjacentes (2) a encore fait chuter les performances (0.59), suggérant le rôle à la fois de la taille absolue et relative des intervalles pour reconnaître une mélodie familière. Comme pour les études précédentes (Dowling, 1978 ; Dowling & Fujitani, 1971), le contour a joué un rôle dans la reconnaissance de mélodies ici familières (différence de performances de 0.31 avec la condition dans laquelle seules les premières notes et l'harmonie étaient maintenues).

En résumé, à la question de savoir sous quel format une mélodie est stockée en mémoire, les recherches menées par W. Jay Dowling (Dowling, 1978 ; Dowling & Fujitani, 1971 ; Dowling & Harwood, 1986) permettent d'apporter deux éléments de réponse :

Le premier est la démonstration du rôle de divers indices dans la reconnaissance de mélodies : le contour, la taille absolue et relative des intervalles, et la différence de tonalité dont l'importance relative semble varier selon les connaissances préalables de l'auditeur.

Le deuxième est que la nature du schéma mélodique diffère selon la durée de conservation de la mélodie. Les notes et le contour d'une mélodie sont encodés en mémoire à court terme, et les intervalles en mémoire à long terme. Notes, contour et intervalles constituent ainsi des codes de plus en plus abstraits, résultant d'un traitement de plus en plus élaboré.

3.1.2.2 Effet de la durée de l'intervalle de rétention dans l'encodage du contour et des intervalles

Si l'indice encodé diffère selon la durée de rétention d'une mélodie, l'encodage d'une nouvelle mélodie devrait varier avec la durée de l'intervalle de rétention. Le contour d'une mélodie devrait pouvoir être extrait dans un intervalle de rétention bref, alors que l'extraction des intervalles devrait nécessiter un intervalle de temps plus long. C'est ce que Dowling et Bartlett (1981, expérience 4) ont proposé d'examiner à l'aide d'un paradigme "externe/interne" (figure 3.2).

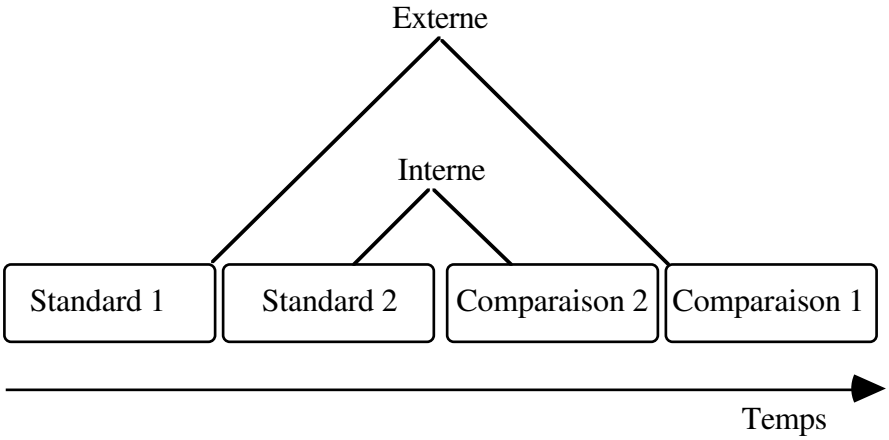


Figure 3.2. Représentation schématique de la procédure "externe/interne". L'auditeur réalise deux jugements même/différent sur deux paires de mélodies. Le premier porte sur la paire "interne", dans laquelle les mélodies (2) sont séparées par un délai court (5 sec). Le deuxième est effectué sur la paire "externe", dans laquelle les mélodies (1) sont séparées par un délai plus long (25 sec).

Dans cette expérience les auditeurs réalisaient à chaque essai deux tâches de discrimination mélodique respectivement sur deux paires de mélodies atonales : une paire "interne" et une paire "externe". Chaque paire de mélodies était composée d'une mélodie standard suivie d'une mélodie de comparaison. Dans la paire "interne" (2), la mélodie à comparer était séparée de la mélodie standard par un délai court (5 sec). Dans la paire "externe", la mélodie à comparer (1) était séparée par un délai plus long (25 sec). La mélodie à comparer pour les deux paires, "interne" et "externe", était soit une transposition exacte de la mélodie standard, soit une transposition inexacte, préservant ou non le contour. Les participants étaient informés de la procédure et invités à maintenir le plus longtemps possible en mémoire la première mélodie entendue en la chantant dans leur tête.

Lorsque la durée de l'intervalle de rétention était court, les auditeurs ont répondu un même nombre de fois que les mélodies étaient identiques lorsque la mélodie à comparer était une transposition exacte et lorsqu'elle ne l'était pas mais que le contour était préservé. Cette erreur était moins fréquente lorsque l'intervalle de rétention était plus long. Ce résultat a été interprété comme le reflet d'un changement dans l'encodage mélodique selon la durée de l'intervalle de rétention. Les auditeurs n'auraient accès qu'au contour dans un intervalle de temps court, alors qu'ils pourraient extraire les intervalles si la durée de rétention était plus longue.

Est-ce la durée de l'intervalle de rétention qui a transformé le format stocké de la mélodie ou le fait que les auditeurs se la répétaient mentalement ? Dans cette expérience, les auditeurs avaient en effet pour consigne de chanter la mélodie dans leur tête, par conséquent apprentissage et intervalle de rétention étaient deux facteurs confondus. De plus, les participants n'ont-ils pas

changé de stratégie d'encodage pour les intervalles de rétention longs, puisqu'ils étaient informés de la procédure employée ?

Après avoir répliqué dans une première expérience les résultats de Dowling et Bartlett (1981, expérience 4), Dewitt et Crowder (1986, expérience 2) ont mené une deuxième expérience visant à répondre à ces deux questions. Afin d'éviter que l'auditeur ne puisse prévoir la durée des intervalles de rétention (de 1 sec pour l'intervalle court et de 30 sec pour l'intervalle long), les auteurs ont présenté les essais dits "interne" et "externe" indépendamment, et aléatoirement à chaque essai. De plus, durant l'intervalle de rétention long, une tâche interférente consistant à compter à rebours de 3 en 3 était proposée aux auditeurs pour éviter la répétition interne des mélodies, et ainsi distinguer l'effet de la durée de l'intervalle de rétention de celui attribué à l'apprentissage. Les résultats obtenus étaient comparables à ceux observés précédemment par les auteurs (Dewitt & Crowder, 1986, expérience 1) et par Dowling et Bartlett (1981, expérience 4) pour des mélodies composées de 7 notes, mais pas pour des mélodies de 5 notes pour lesquelles les performances étaient globalement augmentées. Cette étude confortait donc l'idée que la durée de l'intervalle de rétention pouvait déterminer la nature de l'encodage mélodique, l'auditeur n'accédant aux intervalles que s'il disposait d'un temps de rétention suffisamment long.

Cependant, Croonen et Kop (1989) ont montré que les intervalles pouvaient être extraits dans un intervalle de temps court (1 seconde) dans le cas de mélodies tonales. Dans leur étude, les auditeurs reconnaissaient en effet mieux une transposition exacte qu'une transposition inexacte préservant le contour, et cette différence persistait malgré le déclin global des performances dû à l'allongement de l'intervalle de rétention. Les auteurs ont alors formulé l'hypothèse que la tonalité plus ou moins marquée de la mélodie pourrait jouer un rôle dans l'extraction des intervalles, quelle que soit la durée de l'intervalle de rétention.

Dowling (1991) a reconduit une nouvelle étude dans laquelle il a introduit ce facteur. Il a proposé aux auditeurs une tâche de mémorisation mélodique en continu s'apparentant au paradigme "interne/externe" (figure 3.2) utilisé dans l'étude précédente (Dowling & Bartlett, 1981). Néanmoins, dans ce nouveau paradigme les participants n'avaient pas connaissance de la manipulation du délai de rétention, ni des items tests. Une mélodie était présentée à chaque essai, et l'auditeur devait juger à l'aide d'une échelle numérique exprimant le degré de confiance de son jugement s'il avait déjà entendu cette mélodie (figure 3.3). Les mélodies standard et test à comparer étaient espacées d'un délai court (11 sec en moyenne) ou long (39 sec en moyenne). La mélodie test était soit une transposition exacte de la mélodie standard jouée dans une tonalité proche, soit une transposition inexacte. Ce leurre préservait ou non le contour original. Il était fortement tonal (commençait et terminait sur une note importante de la gamme, il s'agissait ici de la tonique, premier degré de la gamme), faiblement tonal (commençait et terminait sur une note de faible importance, le troisième degré de la gamme) ou atonal.

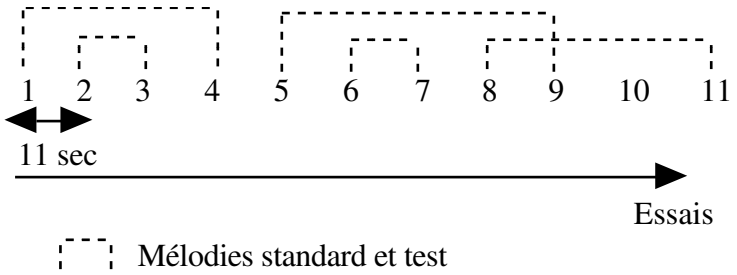


Figure 3.3. Tâche de mémorisation en continu employée par Dowling (1991). Les mélodies standard et test à comparer sont présentées à différents délais (court ou long). L'auditeur, n'ayant pas connaissance de ces différentes conditions de temps de rétention, indique à chaque essai s'il a déjà entendu la mélodie qui lui est présentée. [D'après Dowling (1991, p. 308)].

Les résultats ont révélé que les fausses reconnaissance de mélodies transposées de façon inexacte préservant le contour étaient plus importantes lorsque le délai de rétention était court que lorsqu'il était long. Ce résultat confirmait ainsi les précédents (Dewitt & Crowder, 1986 ; Dowling & Bartlett, 1981, expérience 4) en montrant que l'auditeur semblait accéder au contour dans un intervalle de rétention court, et aux intervalles pour un intervalle de rétention plus long. L'auteur a également observé un effet global du degré de saillance de la tonalité. Les mélodies qui établissaient de façon claire la tonalité étaient mieux retenues dans un délai court que les mélodies faiblement tonales et atonales, suggérant que la tonalité affectait la mémoire du contour mélodique.

Plusieurs différences expérimentales peuvent rendre compte de cette divergence entre les études de W. Jay Dowling (Dowling, 1991 ; Dowling & Bartlett, 1981), Dewitt et Crowder (Dewitt & Crowder, 1986) et celles de Croonen (Croonen, 1994a ; Croonen & Kop, 1989). La première est que dans l'étude de Croonen et Kop (1989) malgré un intervalle de rétention court, les mélodies présentées sont répétées au cours des différents essais. Par conséquent, les auditeurs ont pu les apprendre et les stocker en mémoire à long terme, l'intervalle de rétention court n'étant pas un garant du stockage de la mélodie en mémoire à court terme. La deuxième est liée aux paradigmes expérimentaux ("externe/interne" et mémorisation en continu) utilisés par W. Jay Dowling (Dowling, 1991 ; Dowling & Bartlett, 1981) et Dewitt et Crowder (1986, expérience 1). Dans ces situations expérimentales, l'intervalle de rétention long est occupé par la présentation de mélodies et de jugements même/différent, ou d'une tâche distractrice. Or ces stimuli et/ou ces tâches pourraient provoquer des interférences en mémoire de travail (Deutsch, 1970) (voir §3.3.2 pour une discussion de cet aspect).

3.1.2.3 Effet de la tonalité, de la longueur des mélodies et de la complexité du contour

Si les notes et le contour semblent constituer des indices prépondérants pour accéder à des mélodies stockées en mémoire à court terme, qu'en est-il des autres indices comme la tonalité, la longueur des mélodies et la complexité du contour ?

Nous avons vu précédemment que la différence de tonalité entre deux mélodies, le fait que l'une soit atonale et l'autre tonale, était un facteur de discrimination mélodique utilisé par des musiciens et mais également dans une moindre mesure par des non musiciens (Dowling, 1978). De même, la prégnance de la tonalité, le fait qu'une mélodie suscite de façon plus ou moins forte sa tonalité d'origine, affectait la mémorisation d'une mélodie à court terme (Croonen, 1994a, b ; Croonen & Kop, 1989 ; Dowling, 1991), suggérant que la tonalité et le contour n'agissent pas indépendamment dans la mémorisation mélodique. Les auditeurs semblent donc pouvoir déterminer de façon relativement immédiate l'appartenance ou non de deux mélodies à un même système tonal, probablement en détectant une déviation perceptive par rapport à un système auquel ils sont exposés le plus souvent. Si cet indice est accessible rapidement, qu'en est-il des différences de tonalité (différence de gammes) à l'intérieur du système tonal ?

La distance entre deux tonalités est définie dans la théorie de la musique en fonction du nombre de notes qu'elles partagent. Par exemple, la tonalité de do majeur (dont les notes sont do, ré, mi, fa, sol, la, si) est proche de la tonalité sol majeur (sol, la, si, do, ré, mi, fa#) parce qu'elles ont en commun six notes sur sept. En revanche, la tonalité de do majeur est éloignée de la tonalité de si majeur (dont les notes sont si, do#, ré#, mi, fa#, sol#, la#) car elles ne possèdent plus que deux notes communes, le si et le mi. Bartlett et Dowling (1980) ont examiné l'effet de la distance des tonalités dans la discrimination de mélodies familières et non familières. Dans une première expérience, ils ont présenté deux mélodies successives : une mélodie de référence et une mélodie de comparaison, séparées par un silence de 4 secondes. Les auditeurs, musiciens et non musiciens, devaient juger si les deux mélodies étaient identiques ou non. Il s'agissait de mélodies tonales inédites composées de 5 notes. La mélodie à comparer était transposée de façon exacte ou inexacte. Lorsque la transposition était inexacte, elle préservait le contour original de la mélodie de référence ou non. Dans les conditions où le contour était préservé, la mélodie était jouée dans la même tonalité que celle de la mélodie de référence, dans une tonalité proche, éloignée, ou bien elle était atonale. Les performances de discrimination obtenues dans ces cinq conditions expérimentales ont révélé un effet de la distance entre les tonalités des deux mélodies à comparer. Lorsque la mélodie à comparer était transposée de façon inexacte, les fausses reconnaissances diminuaient avec l'augmentation de la distance entre les tonalités. La discrimination apparaissait donc plus aisée lorsque les mélodies étaient jouées dans des tonalités éloignées plutôt que dans des tonalités proches, voire identiques. Les performances étaient en effet aléatoires si la mélodie était transposée de façon

inexacte mais jouée dans une tonalité identique ou proche de celle de la mélodie de référence, alors qu'elles s'amélioraient lorsque la distance entre les tonalités était accrue (0.60, aire sous la courbe MOC), lorsque la mélodie à comparer était atonale (0.72) ou bien lorsque le contour différait (0.82). Cet effet de la différence de tonalité permettant de discriminer une transposition exacte d'une transposition inexacte, a également été trouvé pour des mélodies familières (expériences 2 et 3). Cependant, les auteurs n'ont pas pu mettre en évidence un effet de la distance entre les tonalités dans le cas de transposition exacte, en raison de performances globalement très élevées (effet plafond). Cette étude révèle donc la contribution de la distance entre les tonalités dans la discrimination de mélodies familières et non familières, chez des auditeurs ayant ou non bénéficié d'une formation musicale.

Les recherches menées par Croonen (1994a, b ; Croonen & Kop, 1989) ont également révélé l'intervention d'autres facteurs dans la mémoire mélodique immédiate : la longueur des mélodies et la complexité du contour, c'est-à-dire le nombre de changements de directions de hauteur. L'auteur a en effet montré que des mélodies tonales inédites composées de 7 notes étaient mieux reconnues que celles constituées de 10 notes. Il est probable que le résultat inverse s'observerait pour des mélodies familières dans la mesure où un extrait plus long donnerait plus d'informations à l'auditeur pour récupérer la mélodie en mémoire. Cet effet de la longueur des mélodies en mémoire à court terme interagissait avec l'effet de la prégnance de la tonalité et également avec la complexité du contour. Le fait que la tonalité soit plus ou moins marquée améliorait d'autant plus les performances que les mélodies étaient courtes (7 notes plutôt que 10), et qu'elles avaient un contour simple (2-3 changements de directions de hauteur plutôt que 4-5), donc probablement d'autant plus qu'elles étaient aisément mémorisables.

Il semble donc que les notes, le contour, la taille absolue et relative des intervalles, la tonalité, la longueur de la mélodie et la complexité du contour interviennent dans la perception et la reconnaissance de mélodies, et qu'ils n'affectent pas indépendamment leur mémorisation. En outre, la contribution de ces indices diffère selon la nature du traitement opéré sur ces mélodies et la durée de leur rétention. Les notes et le contour mélodique constituent des caractéristiques prépondérantes pour reconnaître une mélodie stockée en mémoire à court terme, alors que les intervalles constituent des indices prégnants lorsque les mélodies sont familières et stockées en mémoire à long terme. Dowling (1978) a ainsi proposé un modèle à deux composantes de la mémoire des mélodies dans lequel il postule une distinction et une hiérarchie fonctionnelle entre le contour et les intervalles. Les auditeurs encoderaient le contour mélodique en mémoire à court terme et cette information temporaire servirait d'ancrage au codage des intervalles qui seraient stockés en mémoire à long terme.

3.2 Déficit de la reconnaissance de mélodies : Apport théorique des études de neuropsychologie cognitive

Si les modèles théoriques élaborés sur la base des performances de sujets sains éclairent l'interprétation des déficits observés chez des patients cérébro-lésés. De façon réciproque, la mise en évidence de déficits sélectifs suite à une lésion ou une résection cérébrale, constitue un apport considérable pour notre compréhension du fonctionnement normal et par conséquent pour la validation des modèles existants. Dans cette section, sont donc relatées les études examinant la capacité de patients présentant une résection ou une lésion cérébrale à reconnaître des mélodies familières et non familières. Les patients qui ont participé aux études décrites ci-dessous sont soit des patients ayant eu une lésion cérébrale focale suite à un accident cérébro-vasculaire, soit des patients ayant subi une excision d'une partie du cerveau (régions temporales et/ou frontales) pour le traitement de crises d'épilepsie rebelles à tout traitement médicamenteux. Ces patients épileptiques sont le plus souvent testés après l'opération chirurgicale mais certains peuvent également être testés avant, afin d'étudier l'effet des atrophies et des lésions occasionnées par la maladie (l'idéal étant de disposer d'une évaluation pré- et post-opératoire afin de distinguer l'effet de la résection de l'effet de l'épilepsie). L'intervention chirurgicale des patients ayant participé aux études menées par R. J. Zatorre et S. Samson, dont beaucoup sont décrites dans cette partie, a été réalisée à l'Institut Neurologique de Montréal. Elle consiste en une lobectomie temporale antérieure droite ou gauche (excision de 4 à 6 cm de la partie latérale du lobe temporal à partir du pole temporal, le long de la scissure de Sylvius) comprenant l'amygdale, l'uncus et des quantités variables de l'hippocampe et du gyrus parahippocampique. Le cortex auditif primaire situé au niveau des gyri transversaux de Heschl (Galaburda & Sanides, 1980 ; Liégeois-Chauvel et al., 1991) est parfois inclus en partie ou en totalité dans l'excision.

Dans un premier temps, nous examinerons dans quelle mesure les données neuropsychologiques confortent le modèle à deux composantes proposé par Dowling (Dowling, 1978 ; Dowling & Harwood, 1986) qui postule une distinction et une hiérarchie fonctionnelle entre deux indices, le contour et les intervalles, dans la mémorisation des mélodies. Existe-il des déficits affectant sélectivement le traitement de ces indices mélodiques ? Un déficit dans le traitement du contour est-il toujours associé à un déficit dans le traitement des intervalles ? Le contour et les intervalles étant stockés respectivement en mémoire à court terme et à long terme, nous examinerons ensuite si les déficits observés dans le traitement de ces indices pourraient s'inscrire dans une déficience plus globale respectivement de la mémoire mélodique à court et à long terme. Enfin, à la lumière des résultats de ces études et des observations de patients présentant une amusie (agnosie auditive dans le domaine musical), une analyse fonctionnelle des processus impliqués dans la reconnaissance de mélodies élaborée par Peretz (1994) sera présentée.

3.2.1 Déficit dans le traitement du contour et des intervalles

L'idée que la perception mélodique pourrait faire intervenir des régions cérébrales distinctes selon la nature du traitement effectué (global ou analytique) est née à la suite d'études montrant une latéralisation différente selon l'expertise musicale. Bever et Chiarello (1974) ont en effet montré que des auditeurs ayant bénéficié d'une formation musicale présentaient un avantage de l'oreille droite donc de l'hémisphère gauche (compte tenu de la dominance des connexions controlatérales sur les connexions ipsilatérales), pour discriminer deux mélodies (tâche dite "globale") alors que pour les non musiciens la prédominance inverse s'observait. Par ailleurs, seuls les musiciens parvenaient à effectuer une tâche dite "analytique" consistant à indiquer si un intervalle présenté était ou non extrait de la mélodie entendue préalablement. Cette latéralisation hémisphérique différente selon l'expertise musicale s'expliquerait, selon les auteurs, par une différence dans la nature du traitement mélodique réalisé par des auditeurs musiciens et non musiciens. Les musiciens auraient une écoute analytique leur permettant d'accéder aux intervalles, alors que les non musiciens adopteraient une écoute plus globale leur donnant accès à des propriétés telles que le contour. L'hypothèse formulée par les auteurs est donc que les indices mélodiques globaux comme le contour seraient traités par l'hémisphère droit, alors que l'accès à des indices plus analytiques comme les intervalles dépendrait de l'hémisphère gauche. Ce résultat obtenu en écoute monaurale n'a pas été confirmé par la suite. Zatorre (1979) a examiné en utilisant un paradigme d'écoute dichotique si les systèmes impliqués dans la perception mélodique étaient latéralisés différemment chez les auditeurs musiciens et non musiciens, mais aucune différence n'est apparue. Les deux groupes se comportaient de la même façon montrant une prédominance de l'oreille gauche donc de l'hémisphère droit dans la tâche de reconnaissance de mélodies, et une prédominance de l'oreille droite donc de l'hémisphère gauche pour traiter de la parole. La difficulté de définir des catégories "musicien" et "non musicien" homogènes, basées sur des aptitudes musicales déterminées, ainsi que les différences méthodologiques rendent délicate la comparaison entre ces études et limitent la portée des conclusions.

Cette hypothèse d'une spécialisation hémisphérique différente selon la nature du traitement mélodique effectué, a été testée de façon systématique plus tard par Zatorre (1985). Il a examiné la contribution respective de deux indices, la tonalité et le contour, dans la discrimination de mélodies non familières chez soixante-huit patients épileptiques ayant subi une résection unilatérale temporale droite (30) ou gauche (28) ou bien frontale droite (10). Certaines excisions temporales incluaient au moins en partie le gyrus de Heschl (12 à droite et 10 à gauche). L'ensemble de ces patients avaient le langage localisé à gauche d'après le test de l'injection intracarotidienne d'amytal sodique (Wada & Rasmussen, 1960), barbiturique anesthésiant de façon temporaire un hémisphère cérébral, et les réponses à un questionnaire de préférence manuelle. Vingt témoins droitiers appariés aux patients sur différents critères, âge,

sexe, nombre d'années d'éducation et QI, ont également participé à cette étude. Les patients et les témoins étaient tous non musiciens.

Deux tâches étaient proposées, une tâche de discrimination mélodique et une tâche de reconnaissance différée des mélodies présentées durant le test, examinant la mémoire incidente des mélodies. La tâche de discrimination mélodique consistait à juger si deux mélodies tonales composées de 6 notes présentées successivement espacées par un silence de 2 secondes, étaient identiques ou différentes. Lorsqu'elles étaient identiques, la deuxième mélodie était une répétition de la première. Lorsqu'elles différaient, la quatrième note était modifiée en moyenne de ± 4 dt : 1) préservant le contour de la mélodie originale ainsi que sa tonalité, 2) préservant le contour mais pas la tonalité, 3) modifiant le contour sans changer la tonalité, ou 4) modifiant le contour et la tonalité. Dans le test de reconnaissance différée des mélodies, une transposition des mélodies utilisées dans la tâche de discrimination ainsi qu'une transposition de mélodies nouvelles construites en inversant l'ordre des notes des mélodies existantes, étaient présentées aux auditeurs. Ces derniers devaient dire si oui ou non les mélodies avaient déjà été entendues préalablement.

Les résultats ont révélé que seul le groupe de patients ayant une résection temporale droite avait des performances de discrimination mélodique (d') inférieures à celles obtenues par la population témoin, ainsi qu'un critère de réponse (bêta) différent. Le pattern de résultats observé dans les quatre conditions expérimentales était identique pour les trois groupes de patients et pour la population témoin, montrant la contribution à la fois de la tonalité et du contour dans la discrimination mélodique (et probablement des notes puisque les mélodies étaient présentées dans le même registre) (Dowling, 1978 ; Dowling & Fujitani, 1971 ; Dowling & Harwood, 1986). Les performances de discrimination augmentaient lorsque la tonalité entre les deux mélodies différait, ou bien lorsque le contour variait, et cette amélioration était plus importante lorsque les mélodies différaient à la fois sur la tonalité et sur le contour. Dans la tâche de reconnaissance différée, les deux groupes de patients dont la résection était temporale (droite et gauche) étaient déficitaires, soulignant les difficultés mnésiques conséquentes aux résections temporales touchant des parties de l'amygdale et de l'hippocampe. Les performances de discrimination mélodique des patients présentant une résection temporale droite et gauche ont ensuite été analysées en distinguant les résections incluant au moins en partie ou non le gyrus de Heschl, et elles ont été comparées à celles des témoins. L'auteur a alors constaté que les patients dont la résection (droite et gauche) incluait au moins en partie le gyrus de Heschl, étaient globalement affectés dans leur aptitude à discriminer deux mélodies non familières. Ce déficit était cependant plus prononcé pour des résections droites. La même analyse réalisée sur les performances de reconnaissance différée ne révèle pas d'effet de la résection au moins partielle du gyrus de Heschl. Ce dernier résultat a conduit Zatorre (1985, 1988a) à formuler l'hypothèse que le gyrus de Heschl pourrait être impliqué dans la mémoire mélodique à court terme (cette question est abordée dans la section suivante §3.2.2).

L'implication des structures temporales droites et des gyri de Heschl droit et gauche dans la discrimination d'un pattern mélodique, a été confirmée par la suite par Samson et Zatorre (1988) dans une tâche plus simple consistant à comparer des patterns de trois notes. L'étude a été réalisée auprès de soixante-dix-sept patients cérébro-lésés ayant subi une excision unilatérale temporale droite (26 dont 12 incluant au moins partiellement le gyrus de Heschl) ou gauche (28 dont 13 incluant au moins partiellement le gyrus de Heschl), frontale droite (14) et fronto-temporale droite (9), ainsi que vingt témoins appariés. Les auditeurs étaient soumis à deux tâches : une tâche de discrimination mélodique consistant à comparer deux patterns de trois notes, et une tâche de discrimination harmonique résidant dans la comparaison de deux accords de trois notes. Les patients ayant subi une résection temporale droite et fronto-temporale droite, sont apparus déficitaires dans la tâche de discrimination mélodique. Une analyse effectuée en distinguant les patients sur la base de l'inclusion ou non du gyrus de Heschl dans la résection, a révélé que des patients présentant une résection temporale gauche avaient des performances de discrimination significativement altérées si le gyrus de Heschl était au moins partiellement excisé. Par ailleurs, aucun déficit de discrimination harmonique n'a été mis en évidence pour l'ensemble des patients.

Ces deux études révèlent le rôle des régions temporales droites ainsi que la contribution du gyrus temporal supérieur droit et gauche dans la discrimination de mélodies non familières. Cependant, elles ne mettent pas en évidence de déficit affectant de façon différentielle la perception des intervalles, du contour ou de la tonalité, qui suggérerait l'implication de structures cérébrales distinctes dans le traitement de ces diverses propriétés mélodiques. Deux études récentes ont néanmoins apporté des éléments laissant entrevoir cette possibilité.

Peretz (1990) a examiné la contribution respective du contour et des intervalles dans la discrimination de mélodies tonales non familières chez vingt patients présentant des lésions unilatérales droite (10) ou gauche (10) suite à un accident cérébro-vasculaire (le site exacte de la lésion n'est pas mentionné), et vingt témoins appariés. Les auditeurs devaient juger si deux mélodies tonales non familières présentées successivement étaient identiques ou différentes. Lorsqu'elles étaient identiques, la version originale de la mélodie était répétée ou la mélodie était transposée selon la condition. Lorsqu'elles différaient, une seule note dont la position variait à chaque essai avait été modifiée, préservant ou non le contour mélodique. Trois tests successifs composés chacun de 24 essais (8 essais "même" et 16 essais "différent"), ont été proposés. Dans le premier, la mélodie à comparer était soit identique (version originale répétée), soit elle différait par une note modifiant le contour. Dans le second, la mélodie était soit identique comme précédemment, soit une note avait été modifiée en préservant le contour de la mélodie originale. Dans le troisième, la mélodie à comparer était soit une transposition exacte de la première mélodie, soit une transposition dans laquelle une note avait été changée altérant dans la moitié des essais le contour et le préservant dans l'autre moitié.

Les sujets contrôles parvenaient à réaliser cette tâche de discrimination mélodique et leurs performances étaient accrues lorsque le contour des deux mélodies à comparer différait, confirmant la prégnance de cet indice pour reconnaître une mélodie non familière (Dowling, 1978 ; Dowling & Fujitani, 1971 ; Dowling & Harwood, 1986). Seuls les patients présentant une lésion dans l'hémisphère droit se sont avérés déficitaires dans cette tâche par rapport aux témoins, que le contour soit ou non modifié, et que la mélodie soit ou non transposée. Ceci révélait leur incapacité à discriminer une mélodie sur la base de son contour et de ses intervalles. Il est intéressant de noter que leurs performances obtenues dans le cas où la mélodie à comparer était transposée (40 % de réponses correctes), étaient inférieures à celles qu'ils obtenaient lorsque les deux mélodies étaient présentées dans le même registre (58 % de réponses correctes). Ce résultat peu discuté par I. Peretz (1990, 1994) suggère que les patients présentant une lésion dans l'hémisphère droit semblent encore capables de se baser sur les notes pour discriminer deux mélodies non familières (Dowling & Fujitani, 1971), alors qu'ils manifestent des difficultés à extraire le contour mélodique et à accéder aux intervalles. Cette proposition est cohérente avec le résultat de l'étude de Zatorre (1988b) qui montre que des patients présentant une résection du lobe temporal droit incluant ou non le gyrus de Heschl, sont encore capables d'extraire la hauteur de sons isolés pour réaliser un jugement de changement de direction de hauteur lorsque la fréquence fondamentale des sons est présente (condition contrôle).

La variabilité intra-groupe importante a conduit Peretz (1990) à examiner pour chacune des conditions les performances des patients considérés comme réellement déficitaires, c'est-à-dire des patients dont les performances se situaient en deçà de la plus petite performance obtenue par les témoins. Elle a ainsi pu constater que les cinq patients cérébro-lésés droits qui manifestaient une déficience dans l'utilisation du contour, étaient également déficients dans l'utilisation des intervalles. En revanche, l'association de ces deux déficits n'était pas observé chez les trois patients cérébro-lésés gauches. Ils ne parvenaient pas à extraire les intervalles mais avaient par ailleurs des performances équivalentes à celles obtenues par les témoins lorsque les deux mélodies différaient par leur contour. L'association et la dissociation de déficits dans le traitement du contour et des intervalles observés respectivement chez les cérébro-lésés droits et gauche, suggèrent que ces deux indices seraient traités par des systèmes distincts latéralisés dans des hémisphères différents (Peretz, 1990, 1994). Ce résultat étaye le modèle à deux composantes développé par Dowling (1978), qui postule que l'abstraction du contour est une étape nécessaire à l'extraction des intervalles.

En reprenant la batterie de tests employée par Peretz (1990), Liégeois-Chauvel et ses collaborateurs (1998) ont récemment confirmé ce pattern d'association et de dissociation de déficits dans le traitement du contour et des intervalles. L'étude a porté sur soixante-cinq patients ayant subi une cortectomie unilatérale (22 à gauche et 43 à droite) et vingt-quatre témoins appariés. Trois groupes de patients ont été distingués selon le site de la résection : 1) le groupe "T1 épargné" comprenait des patients dont l'excision incluait les gyri

temporaux médian (T2) et inférieur (T3) ainsi que les structures limbiques (chez cinq patients la résection incluait le pole temporal), 2) le groupe "T1 antérieur inclus" comprenait les patients dont la partie antérieure du gyrus temporal supérieur (T1) était incluse dans la résection (aire 38 et partie antérieure de l'aire 22 selon la nomenclature de Brodmann) en plus des structures citées précédemment, 3) le groupe "T1 postérieur inclus" était composé de patients dont la résection incluait la partie postérieure du gyrus temporal supérieur. L'étendue de la résection variait selon le côté, il s'agissait d'une excision temporo-pariétale à droite (réalisée uniquement à droite en raison de la présence de l'aire de Wernicke dans les régions homologues à gauche) comprenant les aires 22, 42, 39, 40, et d'une résection postérieure de T1 à gauche en plus des structures excisées dans le cas (1). Ainsi, trente-trois patients avaient une résection "T1 épargné" (19 à droite et 14 à gauche), treize patients avaient une résection "T1 antérieure inclus" (8 à droite et 5 à gauche), et huit patients avaient une résection "T1 postérieure inclus" (5 à droite et 3 à gauche).

La tâche de discrimination mélodique était donc identique à celle mentionnée ci-dessus, à l'exception du troisième test. Il était remplacé par un test dans lequel lorsque la mélodie à comparer différait de la mélodie originale, la tonalité était modifiée. Ainsi, la mélodie à comparer était identique à la mélodie de référence ou bien différait par une note dont la position variait à chaque essai modifiant : 1) le contour de la mélodie originale en préservant la tonalité, 2) la tonalité en préservant le contour, 3) les intervalles en préservant le contour et la tonalité. Ces trois modifications constituaient les trois tests successifs comprenant 30 essais, avec un même nombre d'essais identiques et différents.

Une première analyse contrastant les patients selon le côté de la résection et l'inclusion ou non de T1, a été effectuée. Elle a révélé que les patients ayant une résection incluant T1 à droite avaient des performances de discrimination mélodique inférieures à celles des témoins lorsque la discrimination était effectuée sur la base du contour (différence de performances de 5 à 10 %). En outre, l'ensemble des patients ayant une résection temporale droite et gauche incluant ou non T1, étaient déficients pour discriminer les mélodies sur la base des intervalles. Aucun déficit n'a été observé quand la discrimination se basait sur la différence de tonalité. Une deuxième analyse contrastant les patients selon la partie antérieure ou postérieure de T1 incluse, a ensuite montré que les patients présentaient un déficit de discrimination mélodique dans l'ensemble des tests lorsque la partie postérieure de T1 était excisée (comprenant pour les résections droites, c'est-à-dire pour 5 patients sur 8, les aires 22, 42, 39 et 40). Cinq patients sur les huit (3 droits et 2 gauches) avaient des performances de deux écart-types inférieures aux témoins lorsque les mélodies différaient par le contour ou par la tonalité et 7 sur 8 lorsque la discrimination nécessitait l'extraction des intervalles. Un déficit de discrimination sur la base du contour était toujours associé à un déficit de discrimination sur la base des intervalles, mais pas l'inverse. Ce résultat portant sur des patients épileptiques est cohérent avec ceux obtenus par Peretz (1990) sur des patients dont la lésion était due à un accident cérébro-vasculaire. Cette étude confirme en outre l'importance du gyrus temporal supérieur droit (T1) dans la

reconnaissance de patterns mélodiques (Samson & Zatorre, 1988 ; Zatorre, 1985 ; Zatorre, Evans & Meyer, 1994 ; Zatorre & Samson, 1991). Elle met également en évidence le rôle de la partie postérieure du gyrus temporal supérieur dans la discrimination mélodique, suggérant que la résection des aires auditives secondaires (aires 22 et 42 correspondant respectivement à la partie latérale du gyrus de Heschl et au planum temporale) laissant intacte l'aire auditive primaire (aire 41, partie dorsoposteromédiane du gyrus de Heschl) est suffisante pour affecter les performances de discrimination mélodique. Ce dernier résultat est cohérent avec les études montrant qu'une lobectomie temporale plus postérieure affecte d'avantage les performances de discrimination mélodique (Samson & Zatorre, 1988 ; Zatorre, 1985, 1988a).

3.2.2 Déficit de la reconnaissance de mélodies familières et non familières

Les études de Peretz (1990) et de Liégeois-Chauvel et ses collaborateurs (1998) révèlent l'existence d'une dissociation fonctionnelle et anatomique dans le traitement du contour et des intervalles. Des patients présentant une lésion ou une résection temporale gauche manifestent des difficultés pour traiter les intervalles, alors qu'ils conservent l'aptitude à extraire le contour d'une mélodie. Cette aptitude est en revanche altérée chez des patients présentant une lésion ou résection temporale droite. Cette dissociation s'inscrit-elle dans un déficit plus général respectivement de la mémoire auditive à court terme et à long terme ? L'hémisphère gauche serait-il impliqué dans la mémoire des mélodies familières et l'hémisphère droit dans la reconnaissance de mélodies non familières ?

3.2.2.1 Mélodies non familières

La plupart des auteurs s'accordent pour penser que les régions temporales et frontales droites sont impliquées dans la discrimination et la rétention de mélodies en mémoire à court terme.

Plusieurs études menées auprès de patients cérébro-lésés révèlent qu'une lésion ou une résection temporale droite induit des déficits dans la discrimination de mélodies non familières, et ce d'autant plus que le gyrus de Heschl est partiellement touché (Liégeois-Chauvel et al., 1998 ; Peretz, 1990 ; Samson & Zatorre, 1988 ; Zatorre, 1984, 1985, 1988a). La contribution des structures temporales et frontales droites dans la discrimination de mélodies non familières a également été mise en évidence chez des sujets sains dans une étude d'imagerie cérébrale fonctionnelle. Zatorre et al. (1994) ont montré, grâce à la technique de Tomographie par Émission de Positons (TEP), que le gyrus temporal supérieur droit était significativement plus activé dans une tâche d'écoute passive de mélodies tonales non familières que dans une tâche d'écoute passive de séquences de bruit. En soustrayant l'activité cérébrale induite par une tâche de discrimination mélodique (consistant à comparer la hauteur des deux premières notes ou de la première et de la dernière note de la mélodie) de celle produite par l'écoute passive de ces

mêmes mélodies, ils ont également mis en évidence une augmentation significative de l'activité cérébrale des régions frontales droites. Ce résultat suggère que cette tâche de discrimination mélodique ferait intervenir des réseaux impliqués dans la mémoire de travail.

Néanmoins, Zatorre (1985) ainsi que Samson et Zatorre (1988) ont constaté que, indépendamment du côté de la résection, ce qui semblait être crucial dans la possibilité de discriminer deux mélodies non familières était l'intégrité des gyri de Heschl, puisque des patients ayant une résection au moins partielle de cette structure à droite comme à gauche s'avéraient déficitaires dans ce type de tâche. Par ailleurs, ils ont constaté que l'inclusion ou non de cette région n'altérait pas la mémoire incidente différée des mélodies, c'est-à-dire la mémoire implicite à long terme. Zatorre (1985, 1988a) a donc formulé l'hypothèse que le gyrus de Heschl pourrait être impliqué dans le maintien temporaire d'une information tonale.

Zatorre et Samson (1991) ont donc examiné le rôle du gyrus de Heschl dans la rétention temporaire d'une hauteur tonale. Pour cela, ils ont proposé à des patients ayant subi une lobectomie temporale unilatérale incluant partiellement ou non le gyrus de Heschl et à une population témoin appariée, une tâche consistant à discriminer la hauteur de deux sons présentés successivement en présence ou non de sons interférents durant l'intervalle de rétention (Deutsch, 1970). Soixante et onze patients épileptiques ayant subi une excision unilatérale temporale droite (26), temporale gauche (26), frontale droite (12) et fronto-temporale droite (7), ainsi que dix-huit témoins appariés, ont participé à l'expérience. Parmi les patients ayant subi une résection temporale droite ou gauche, la moitié présentait une excision au moins partielle du gyrus de Heschl, alors que pour l'autre moitié cette structure était épargnée. Dans tous les cas l'amygdale et l'uncus étaient excisés. Une partie limitée des structures hippocampiques étaient incluses dans la résection pour dix-huit patients à droite et vingt-deux à gauche, et une partie plus étendue pour huit patients à droite et quatre à gauche.

Deux sons complexes étaient donc présentés successivement séparés par un intervalle de rétention de 1650 ms. La tâche des participants consistait à juger si leur hauteur était identique ou différente. Elle était identique dans la moitié des essais, et dans l'autre moitié la hauteur du son à comparer était augmentée de 2, 4 ou 6 dt. Deux conditions étaient proposées : une condition contrôle dans laquelle les deux sons cibles étaient séparés par du silence, et une condition "interférente" où six sons distracteurs étaient présentés durant l'intervalle de rétention. Le premier son distracteur débutait 325 ms après le premier son cible et le dernier son distracteur se terminait également 325 ms avant le deuxième son à comparer. La gamme occupée par ces sons distracteurs était d'une octave. La hauteur du premier son distracteur variait comme dans l'expérience de Deutsch (1972). Elle était identique au premier son cible pour 1/3 des essais, ou bien éloignée de ± 2 dt ou ± 4 dt pour les deux autres tiers respectivement. Les auditeurs avaient pour consigne d'ignorer les sons distracteurs dans la condition "interférente".

Les résultats ont révélé que les témoins et l'ensemble des patients cérébro-lésés étaient capables de discriminer la hauteur de deux sons avec moins de 10 % d'erreurs (5 % chez les témoins) (Zatorre, 1988b). Cependant, en présence de sons distracteurs, les patients ayant subi une résection cérébrale droite (temporale, frontale et fronto-temporale) étaient plus affectés que les sujets contrôles pour réaliser cette tâche. Leur pourcentage d'erreurs était en moyenne de 20 %, contre 6-7 % chez les témoins. Par ailleurs, les patients ayant une résection temporale gauche ne différaient pas du groupe témoin. Une analyse contrastant les patients selon l'inclusion ou non du gyrus de Heschl ou bien selon l'excision limitée ou étendue de l'hippocampe, n'a pas révélé de déficits supplémentaires. Il semble donc qu'une résection temporale antérieure suffise à produire une interférence accrue induite par la présence de sons distracteurs, infirmant l'hypothèse de l'implication spécifique du gyrus de Heschl dans la mémoire tonale à court terme. Par ailleurs, l'effet de la proximité en hauteur entre les sons cibles et les sons distracteurs sur l'amplitude de ce phénomène d'interférence s'observait non seulement chez les témoins (Deutsch, 1972) mais aussi chez les patients suggérant que l'encodage de la hauteur n'était pas altéré.

Les structures temporales, frontales et fronto-temporales droites semblent donc jouer une rôle crucial dans la mémoire tonale à court terme (Samson & Zatorre, 1991 ; Zatorre et al., 1994). Ceci pourrait expliquer en partie les troubles des patients présentant des lésions ou résections temporales droites dans des tâches de discrimination mélodique (Liégeois-Chauvel et al., 1998 ; Peretz, 1990 ; Samson & Zatorre, 1988 ; Zatorre, 1985, 1988a). La résection du gyrus de Heschl ne produit pas de déficits mnésiques supplémentaires, alors que cette structure apparaissait critique dans la discrimination de mélodies non familières (Samson & Zatorre, 1988 ; Zatorre, 1985, 1988a). Compte tenu du fait que dans l'ensemble des études décrites les mélodies à comparer étaient présentées dans le même registre de hauteur (Liégeois-Chauvel et al., 1998 ; Samson & Zatorre, 1988 ; Zatorre, 1985, 1988a), on peut formuler l'hypothèse que le gyrus de Heschl serait impliqué dans la détection d'un changement de note dans la comparaison de deux mélodies non familières (Dowling & Fujitani, 1971). Cette région où siège l'aire auditive primaire serait ainsi impliquée plus spécifiquement dans des mécanismes d'extraction et de comparaison de hauteurs, alors que les structures temporales adjacentes (aires auditives secondaires ou associatives) seraient impliquées dans le maintien temporaire d'une information tonale.

3.2.2.2 Mélodies familières

Si les structures temporales et frontales droites semblent être impliquées dans la rétention des mélodies à court terme, qu'en est-il des mélodies stockées en mémoire à long terme ?

Zatorre (1984, 1988a) a constaté que des patients présentant une résection temporale gauche avaient des difficultés à reconnaître une mélodie familière ayant subi des distorsions. Dans cette étude, la tâche des participants consistait à reconnaître un air musical familier en désignant le titre de la mélodie parmi une liste qui leur était proposée. Cinq mélodies familières étaient présentées dans leur version originale (à l'exception du rythme), et selon 3 types de déformations qui consistaient respectivement : 1) à multiplier les intervalles de la mélodie originale par 2 (préservant la taille relative des intervalles mais pas leur valeur absolue), 2) à transposer la mélodie à l'octave (préservant ainsi la taille absolue des intervalles et les notes), ou bien 3) à transposer la mélodie de 10 ou 11 dt (préservant la taille absolue des intervalles mais pas les notes). Toutes ces déformations préservaient le contour, mais seule la transposition à l'octave permettait de garder les mêmes notes. Les résultats ont révélé que cette dernière version modifiée était la mieux reconnue après la version originale, suivie de la version dans laquelle les intervalles étaient multipliés par deux. Enfin, la moins bien reconnue était la transposition de la mélodie de 10-11 dt. Ce pattern de résultats s'observait chez les témoins et l'ensemble des patients cérébro-lésés, mais les patients ayant subi une résection temporale gauche avaient des performances de reconnaissance globalement plus faibles que les témoins. L'excision des gyri de Heschl n'engendrait pas de déficit supplémentaire. Selon l'auteur, le fait qu'une résection temporale gauche soit plus critique pour identifier une mélodie familière qu'une résection temporale droite, pourrait provenir des difficultés de mémoire verbale et de dénomination rencontrées à la suite de ce type d'excision (Zatorre, 1984). En effet, la représentation en mémoire des mélodies familières utilisées était fortement associée à un contenu verbal implicite. Il est par conséquent difficile de distinguer dans cette étude la contribution des structures temporales gauches dans le traitement de l'information verbale, de leur éventuelle implication dans le stockage des mélodies en mémoire à long terme.

Samson et Zatorre (1991) ont donc examiné l'implication des régions temporales gauches dans le traitement verbal et mélodique dans un contexte musical, en étudiant la mémorisation respective des paroles et de l'air de chansons nouvelles. Quarante-trois patients ayant subi une résection temporale unilatérale gauche (22) ou droite (21) ainsi que vingt témoins appariés, ont participé à la première expérience dans laquelle les paroles et l'air de la chanson étaient présentés simultanément. Une première phase a consisté à présenter 24 extraits de chansons que les auditeurs devaient essayer de retenir. Puis durant la phase de reconnaissance, 20 extraits étaient proposés. Il s'agissait soit du même extrait composé des mêmes paroles et du même air, soit d'extraits différents composés : 1) des anciennes paroles associées à un nouvel air, 2) de l'ancien air avec de nouvelles paroles, 3) de paroles et d'un air nouveau, ou 4) de paroles et d'un air déjà présentés antérieurement mais ne provenant pas de la même chanson. L'auditeur devait juger après chaque extrait si la chanson (c'est-à-dire à la fois les paroles et l'air) avait été présentée préalablement, si les paroles avaient déjà été entendues indépendamment de l'air, et si l'air avait déjà été entendu indépendamment des paroles.

Globalement, les résultats ont révélé que les témoins et les patients étaient plus performants pour reconnaître les paroles plutôt que l'air. Les patients ayant une résection temporale gauche apparaissaient déficients dans les deux types de reconnaissance par rapport aux témoins, alors que les patients ayant une résection temporale droite ne différaient pas du groupe contrôle. L'analyse des performances obtenues dans la reconnaissance d'un air ancien associé à de nouvelles paroles, a révélé un déficit pour les deux groupes de patients par rapport au groupe témoin. Ce résultat confirme le rôle des structures temporales gauches dans la mémorisation de l'air d'une chanson lorsque des paroles même nouvelles sont associées. En outre, elle met en évidence un déficit chez des patients ayant une résection temporale droite pour reconnaître un air lorsqu'ils ne peuvent se baser sur les paroles. La résection ou non des gyri de Heschl ou bien des structures hippocampiques n'affectaient pas les performances. Cette expérience suggère que les structures temporales gauches seraient impliquées dans la mémorisation d'une mélodie en présence de paroles, même si elles sont étrangères à la chanson d'origine. En revanche, les structures temporales droites n'interviendraient que lorsque les paroles ne permettent pas de réaliser la tâche.

Si les structures temporales droites interviennent dans l'encodage en mémoire de l'air musical et les structures temporales gauches dans l'encodage des paroles, alors des patients présentant une résection temporale droite devraient être affectés dans l'encodage en mémoire de la mélodie mais pas des paroles, et les patients présentant une résection temporale gauche devraient présenter le déficit inverse. Une deuxième expérience a donc été menée pour examiner la mémoire des paroles et de l'air d'une chanson lorsque ces deux composantes étaient présentées séparément. L'étude a été conduite auprès de vingt-quatre patients présentant une excision temporale droite (12) ou gauche (12), et dix témoins. Les extraits musicaux étaient identiques à ceux proposés dans la première étude mais cette fois les paroles (récitées comme une histoire) et l'air (chanté en remplaçant les paroles par des "lalala") étaient présentés dans deux conditions séparées. Conformément à l'hypothèse, les patients dont l'excision temporale était localisée à gauche étaient affectés dans la mémorisation des paroles, et ceux dont l'excision temporale se situait à droite apparaissaient déficients dans la reconnaissance des airs. Cette étude souligne la spécificité mnésique des lobes temporaux droit et gauche. Les structures temporales droites seraient impliquées de façon prédominante dans la mémoire musicale mélodique, alors que les structures temporales gauches seraient impliquées dans la mémoire musicale verbale. Cependant, cette latéralisation franche contraste avec les résultats des études précédentes montrant l'intervention des deux lobes temporaux dans la mémorisation à long terme de mélodies (Zatorre, 1984, 1988a). Cette différence peut être attribuée au fait que dans cette étude il s'agit de mélodies inédites. Or stocker ces mélodies en mémoire à long terme, nécessite de les encoder préalablement en mémoire à court terme. Par conséquent, le déficit des patients présentant une résection temporale droite dans la reconnaissance de ces extraits mélodiques, pourrait être liée à leur difficulté d'encoder ces nouvelles mélodies (Zatorre, 1985, 1988a ; Zatorre & Samson, 1991).

Afin de mieux comprendre l'implication des structures temporales gauche et droite dans l'apprentissage et la rétention de mélodies nouvelles, Samson et Zatorre (1992) ont proposé à quarante patients épileptiques ayant subi une résection temporale droite (20) ou gauche (20) et à quinze témoins, d'apprendre des mélodies non familières et des non-mots (des mots n'existant pas dans le lexique et ne ressemblant pas à un mot pouvant évoquer un contenu sémantique). Six tests de reconnaissance ont été proposés : cinq le jour de l'apprentissage et un test de reconnaissance différé après 24 heures. Les résultats ont montré que les performances de reconnaissance des non-mots et des mélodies non familières de l'ensemble des patients étaient inférieures à celles obtenues par les témoins, mais présentaient une progression comparable. Un résultat particulièrement intéressant concernait les performances obtenues lors du test de reconnaissance proposé 24 h après l'apprentissage. Les patients présentant une résection temporale gauche amélioraient leurs performances de reconnaissance de mélodies par rapport au dernier test passé la veille, alors que les performances des patients présentant une résection temporale droite chutaient. Ce résultat pourrait conforter l'hypothèse selon laquelle le déficit des patients présentant une résection temporale droite pour mémoriser des mélodies nouvelles proviendrait d'un déficit dans l'encodage de la mélodie en mémoire à court terme. Les patients présentant une résection temporale gauche auraient pu encoder les nouvelles mélodies mais présenteraient des difficultés par rapport aux témoins pour les stocker en mémoire à long terme. Le pattern inverse s'observait pour la reconnaissance des non-mots. Les patients présentant une résection temporale droite amélioraient leurs performances de reconnaissance après 24 h, alors que celles des patients présentant une excision temporale gauche chutaient, confirmant ainsi la spécificité du lobe temporal gauche dans la mémoire verbale. Les performances des témoins plafonnaient dans ces deux tests de reconnaissance différée.

Même s'il est difficile au vu des résultats, de se prononcer de façon certaine sur l'implication des régions temporales gauches dans la mémoire mélodique, il semble que ces structures interviennent dans la mémoire à long terme de mélodies nouvellement apprises (Samson & Zatorre, 1992), ainsi que dans la reconnaissance de mélodies familières (Zatorre, 1984, 1988a). La contribution du lobe temporal gauche dans la mémorisation à long terme des mélodies demande donc à être plus amplement explorée afin de pouvoir établir un lien entre la dissociation fonctionnelle et anatomique observée dans le traitement du contour et des intervalles (Liégeois-Chauvel et al., 1998 ; Peretz, 1990), et la spécialisation des structures temporales droite et gauche respectivement dans le stockage mélodique à court et à long terme.

3.2.3 Analyse fonctionnelle de la reconnaissance de mélodies : Modèle théorique

Les études portant sur des groupes de patients cérébro-lésés (Zatorre, 1984, 1988a, pour des revues) dont certaines ont été relatées ci-dessus, ainsi que l'observation de cas d'amusie (Peretz, 1994, pour une revue), apportent des éléments précieux non seulement pour la connaissance des bases neurales des processus impliqués dans la perception et la reconnaissance de mélodies, mais également pour la compréhension du fonctionnement du système cognitif exécutant ces tâches. Les connaissances dont nous disposons actuellement sur les étapes de traitement mises en jeu dans la reconnaissance d'un air musical, que ce soit une mélodie ou bien une chanson combinant des informations verbales à l'information musicale, ont été synthétisées et analysées par I. Peretz (1994) dans un modèle fonctionnel dont la représentation schématique est reproduite dans la figure 3.4. Nous résumerons ici les étapes de traitement impliquées dans la reconnaissance d'une mélodie.

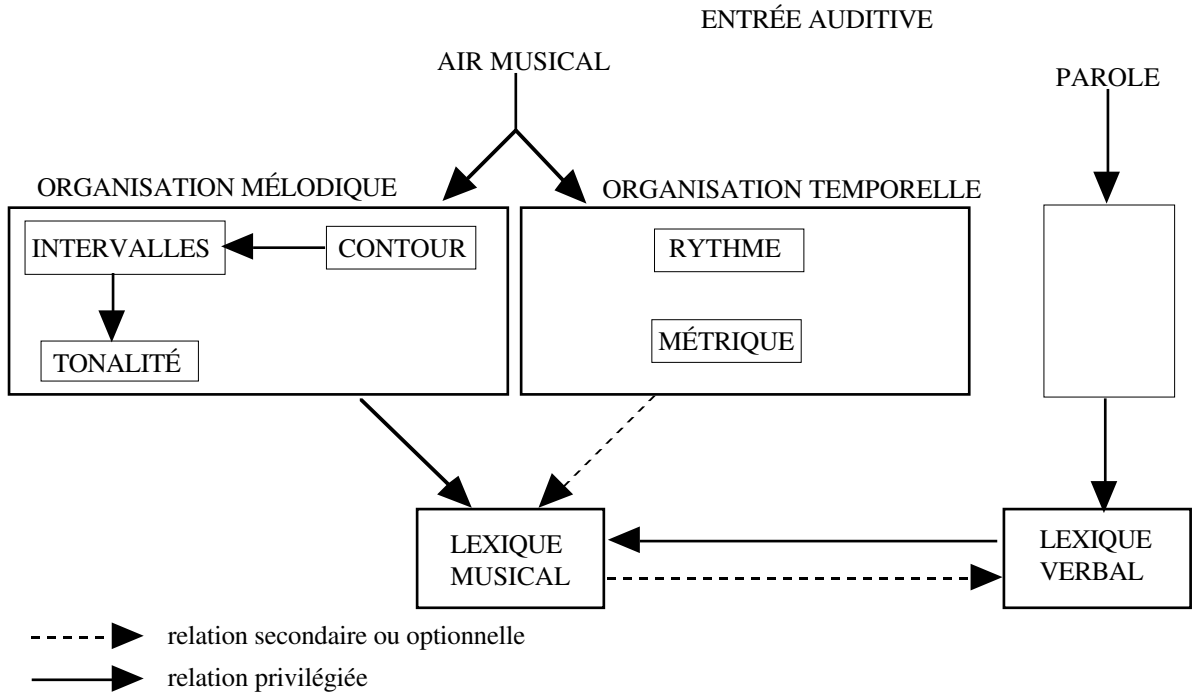


Figure 3.4. Schéma fonctionnel de la reconnaissance d'un air musical établi sur la base de l'observation des déficits de patients cérébro-lésés. [Reproduit à partir de Peretz (1994, p. 233)].

3.2.3.1 Structure mélodique et structure temporelle : Deux systèmes distincts

Concernant les étapes de traitement mises en jeu dans la perception et la reconnaissance de mélodies, ce modèle postule l'existence de deux modules distincts pour l'analyse respectivement de la dimension mélodique, définie par les variations séquentielles de hauteur, et

de la dimension temporelle, définie par les variations de durée. Cette distinction s'appuie sur l'observation de cas de double dissociation dans des tâches de discrimination mélodique et rythmique chez des patients présentant respectivement des lésions ou résection temporale droite et gauche.

En effet, dans l'étude relatée précédemment Peretz (1990) avait non seulement examiné l'aptitude des patients présentant une lésion cérébrale unilatérale droite ou gauche à discriminer des mélodies, mais aussi à traiter des indices temporels tels que la métrique et le rythme. Deux tests temporels avaient été proposés : une tâche métrique dans laquelle il s'agissait de reconnaître si l'extrait musical était une valse ou une marche, et une tâche de discrimination rythmique consistant à juger si deux rythmes étaient identiques ou différents. L'auteur a examiné les performances obtenues par chacun des patients dans les tâches mélodiques et rythmiques, et a mis en évidence deux cas de double dissociation. Deux patients présentant une lésion cérébrale à droite étaient déficitaires dans la tâche mélodique mais pas dans la tâche rythmique, et à l'inverse deux patients qui avaient une lésion au niveau de l'hémisphère cérébral gauche étaient déficients dans la tâche rythmique mais pas dans la tâche mélodique. A l'appui de ces observations, Peretz (1994) a formulé l'hypothèse que ces deux dimensions, mélodiques et rythmiques seraient traitées par deux systèmes indépendants. Cette hypothèse a été confortée par la suite par Peretz et Kolinsky (1993) qui ont révélé que des patients présentant des lésions temporales bilatérales présentaient une amélodie sans arythmie.

Néanmoins, la distinction anatomique de ces deux systèmes c'est-à-dire une éventuelle asymétrie hémisphérique, la dimension temporelle traitée par l'hémisphère gauche et la dimension mélodique par l'hémisphère droit, n'apparaît pas clairement dans les différentes études. Liégeois-Chauvel et ses collaborateurs (1998) ont repris la batterie de test employée par Peretz (1990), et ont trouvé des déficits dans la tâche métrique chez des patients ayant une résection incluant la première circonvolution temporale, que ce soit à droite ou à gauche. En outre, Penhume, Zatorre et Feindel (1999) ont récemment mis en évidence des troubles dans la reproduction du rythme d'une séquence chez des patients ayant subi une résection temporale droite incluant au moins en partie le gyrus de Heschl.

De plus, les résultats des études comportementales ne sont pas toujours compatibles avec les données neuropsychologiques, certaines suggérant une interaction entre ces deux dimensions dans la reconnaissance d'une mélodie. Selon Peretz (1994, pour une discussion), mélodie et rythme seraient donc traités dans des modules distincts dans les premières étapes du traitement, puis intégrés plus tardivement. Ceci est représenté dans le schéma par deux voies de traitements parallèles, la voie mélodique et la voie temporelle.

3.2.3.2 Les composantes du système mélodique

A la lumière des travaux réalisés par W. J. Dowling (Dowling & Harwood, 1986, pour une revue) et des dissociations de déficits observées (Liégeois-Chauvel et al., 1998 ; Peretz,

1990), ce modèle mentionne la contribution de trois indices mélodiques dans la reconnaissance d'un air musical : le contour, les intervalles et la tonalité. L'étude de patients cérébro-lésés a révélé une dissociation entre le traitement de deux de ces propriétés mélodiques : le contour et les intervalles. Les patients cérébro-lésés droits présentaient en effet des difficultés à la fois dans le traitement du contour et des intervalles, alors que les patients cérébro-lésés gauche montraient une dissociation dans le traitement de ces deux indices puisqu'ils parvenaient à discriminer les mélodies sur la base de leur différence de contour, mais pas sur la base de leur différents intervalles. Cette association et dissociation de déficits étaye le modèle de Dowling (1978) qui postule une hiérarchie fonctionnelle dans le traitement de ces indices. Elle est symbolisée par le sens de la flèche allant du contour vers les intervalles puis vers la tonalité.

3.2.3.3 L'existence d'un lexique musicale

Les indices mélodiques (contour, intervalles et tonalité) et temporels (métrique et rythme) permettent ainsi d'accéder au lexique musical qui regroupe l'ensemble des mélodies stockées en mémoire. De cet accès dépendent la reconnaissance et l'identification des pièces musicales. S'il est probable que l'unité d'accès au lexique combine ces deux types d'information, mélodique et rythmique, les données comportementales et neuropschologiques (cas de deux amusiques ayant une amélodie sans arythmie) montrent le rôle privilégié de la voie mélodique dans l'accès au lexique (flèche en trait plein sur le schéma) (Peretz, 1994).

Ce lexique musical entretient probablement un lien avec le lexique verbal dans le cas de mélodies ayant un contenu verbal implicite (air d'une chanson sans les paroles) ou explicite (chansons comprenant la musique et les paroles). Néanmoins, l'existence de cas de dissociation dans la reconnaissance de l'air et des paroles (patient C. N) décrits par Peretz (1994), suggèrent que paroles et musique sont probablement traitées dans des systèmes distincts avant d'être intégrées.

Dans ce lexique la mémoire mélodique à court et à long terme n'est pas différenciée. Pourtant à la lumière des études menées par Dowling et ses collaborateurs (Dowling, 1978 ; Dowling & Fujitani, 1971 ; Dowling & Harwood, 1986), Peretz (1990) et Liégeois-Chauvel et al. (1998), on pourrait postuler l'existence de deux registres distincts à l'intérieur du lexique correspondant respectivement à la mémoire à court et à long terme. L'accès à ces deux systèmes mnésiques serait déterminé par des indices mélodiques distincts. Le contour permettrait d'accéder à une mélodie stockée en mémoire à court terme, et les intervalles à une mélodie stockée à plus long terme.

Il faut souligner que ce corpus théorique a été élaboré à partir d'études comportementales et neuropsychologiques se basant uniquement sur des tests dans lesquels une mélodie isolée est présentée (situation monophonique). Les indices sur lesquels le système auditif se base dans la reconnaissance d'une mélodie sont probablement identiques à l'écoute d'une ou de plusieurs

voix mélodiques présentées simultanément (situation polyphonique). Cependant, les processus perceptifs et mnésiques engagés dans ces deux types de tâches sont quant à eux certainement différents. Les spécificités attentionnelles et mnésiques de la reconnaissance de mélodies présentées simultanément avec une autre séquence, sont envisagées dans la section suivante.

3.3 Reconnaissance de mélodies en présence d'une autre séquence : Spécificités attentionnelles et mnésiques

Nous avons examiné jusqu'à présent les processus perceptifs et mnésiques impliqués dans la reconnaissance d'une mélodie familière et non familière présentée isolément. Or dans une scène auditive complexe, nous sommes la plupart du temps en présence de plusieurs flux sonores simultanés. La musique illustre bien cette situation dans laquelle plusieurs voix mélodiques sont jouées simultanément (polyphonie), que ce soit par un seul instrument (par exemple le piano) ou par l'ensemble des instruments de musique composant un orchestre symphonique. L'auditeur peut alors focaliser son attention sur une des voix mélodiques ou bien sur la partie jouée par un des instruments de l'orchestre, et la suivre.

Nous allons nous intéresser à une situation expérimentale qui s'apparente à celle dans laquelle se trouve un auditeur à l'écoute d'une polyphonie. Il s'agit du paradigme de reconnaissance de mélodies intercalées dans lequel l'auditeur doit reconnaître une mélodie intercalée à une autre séquence également "mélodique" mais que l'on a coutume d'appeler "séquence distractrice" dans la mesure où la tâche de l'auditeur porte sur la mélodie cible et qu'il doit "ignorer" cette autre séquence. Lorsque ces deux séquences, la mélodie et la distractrice, sont présentées dans des registres de hauteur suffisamment éloignés, l'auditeur perçoit deux lignes mélodiques superposées dans le temps. Ils focalise son attention sur la mélodie cible sur la base de ses attentes concernant ses caractéristiques fréquentielles et temporelles, encode la mélodie (il extrait les notes et le contour s'il s'agit d'une nouvelle mélodie, les intervalles si cette mélodie est familière), et la stocke en mémoire de façon à pouvoir la reconnaître. Même si la séquence distractrice est probablement traitée de façon moins "approfondie" que la mélodie cible, il est raisonnable de penser que cette séquence est également encodée et stockée en mémoire.

Nous proposons dans les deux sections qui suivent de discuter des processus attentionnels et mnésiques impliqués spécifiquement dans le reconnaissance d'une mélodie intercalée. La focalisation attentionnelle guidée par les attentes de l'auditeur est examinée, ainsi que les phénomènes mnésiques propres à cette situation où deux mélodies sont encodées simultanément en mémoire.

3.3.1 Focalisation attentionnelle sélective

Reconnaître une mélodie lorsqu'elle est présentée simultanément avec une autre séquence nécessite que l'auditeur puisse focaliser son attention de façon sélective sur la mélodie cible. La présence simultanée de deux voix indépendantes suppose que les sons appartenant à ces deux mélodies appartiennent à des registres de fréquence distincts, ou bien que les deux séquences soient jouées par des instruments présentant des timbres différents (Dowling, 1973 ; Dowling et al., 1987 ; Hartmann & Johnson, 1991) (voir §1.2 pour un exposé des différents paramètres qui interviennent dans la formation des flux auditifs). Par conséquent, pour sélectionner la mélodie cible, l'auditeur doit orienter son attention sur la région fréquentielle occupée par la mélodie cible ou bien sur les caractéristiques spectro-temporelles des sons composant cette mélodie.

Mondor et Bregman (1994) ont montré que l'auditeur pouvait orienter son attention sur une fréquence spécifique, et que cette focalisation sélective de l'attention améliorait le traitement d'un attribut de l'événement focalisé autre que la fréquence. Il s'agissait dans le cadre de cette étude de l'identification de la durée d'un son pur (court ou long) présenté isolément, dont la fréquence avait été indiquée préalablement par un son ("cue"). Les temps de réponse ainsi que les proportions d'erreurs d'identification de la durée étaient abaissés lorsque le son cible sur lequel portait le jugement était présenté avec la même fréquence que l'indice. Lorsque la fréquence de la cible s'éloignait de celle de l'indice, les temps de réponse et la proportion d'erreurs augmentaient progressivement, suggérant que le gain produit par la focalisation sélective sur une fréquence déclinait selon un gradient à mesure que la cible s'éloignait de la fréquence attendue. Mondor et ses collaborateurs ont étendu ces résultats à d'autres attributs comme la localisation spatiale (Mondor & Zatorre, 1995), ainsi qu'à la conjonction de deux attributs comme la fréquence et la localisation spatiale (Mondor, Zatorre & Terrio, 1998). La conclusion des auteurs est que l'attention se focaliserait sur un événement défini par une conjonction d'attributs plutôt que sur des attributs isolés, ce qui signifie que l'attention sélectionnerait des événements déjà organisés perceptivement et non des dimensions sensorielles (Neisser, 1967). L'étude de Botte et ses collaborateurs (1997) confirme cette interprétation. Les auteurs montrent que comme dans le cas de sons isolés, un flux auditif non focalisé est atténué d'environ 15 dB. Cette position est en accord avec la conception théorique de Bregman (1990) qui considère que le système auditif construit de façon pré-attentive des "objets", un événement sonore isolé ou bien un flux auditif, sur lesquels l'attention va pouvoir se focaliser.

Si l'attention peut être orientée de façon sélective sur un "objet auditif", que ce soit un événement isolé ou bien sur une suite d'événement liés perceptivement formant un flux auditif, il doit donc être possible de suivre une mélodie présentée simultanément avec une autre

séquence. Les études sur la perception de mélodies intercalées confirment ce fait et montrent que les auditeurs parviennent à reconnaître une mélodie lorsqu'elle est présentée simultanément avec une séquence distractrice distante d'environ 12 dt (Dowling, 1973 ; Dowling et al., 1987 ; Hartmann & Johnson, 1991 ; Vliegen & Oxenham, 1999, expérience 2). Cette aptitude à focaliser de façon sélective son attention sur une mélodie intercalée avec une séquence distractrice, se construit progressivement au cours de l'ontogenèse. Andrews et Dowling (1991) montrent en effet que des enfants de 5-6 ans éprouvent des difficultés à reconnaître une mélodie familière intercalée à des sons distracteurs présentés dans un registre de hauteur éloigné et joués par des instruments différents, alors que des enfants de 9-10 ans parviennent à réaliser cette tâche de la même manière que les adultes.

Même si les performances de reconnaissance de mélodies intercalées demeurent inférieures à celles obtenues lorsque la mélodie est présentée isolément (Dowling, 1973 expérience 2 ; Dowling & Fujitani, 1971, expérience 1) (ce point est discuté dans la section suivante), ces recherches montrent que les auditeurs parviennent à focaliser leur attention de façon sélective sur une certaine gamme de fréquences occupées par la mélodie. Dowling et ses collaborateurs (1987) ont tenté de caractériser la taille de la fenêtre d'attentes, c'est-à-dire l'étendue de la gamme de fréquences sur laquelle les auditeurs portaient leur attention lors de la perception de mélodies intercalées. Ils ont pour cela mené trois expériences dans lesquelles ils ont examiné le rôle des attentes fréquentielles dans des tâches de discrimination de contour et de hauteur lorsqu'une mélodie était présentée intercalée avec des sons distracteurs (expérience 2) et lorsqu'elle était présentée isolément (expériences 3 et 4 pour les jugements de contour et de hauteur respectivement).

Dans l'expérience 2, l'expérimentateur présentait une mélodie "amorce" de 5 notes puis une mélodie cible intercalée avec des sons distracteurs distants de 2-3 dt de la mélodie vers les basses ou les hautes fréquences. La mélodie cible était identique à l'amorce à l'exception de la quatrième note qui était plus ou moins éloignée en fréquence de la note attendue. Les auditeurs réalisaient deux tâches dans deux sessions différentes. Dans une première session, ils jugeaient si le contour de la mélodie cible était identique ou différent de celui de l'amorce. Dans une deuxième session, ils devaient déterminer si la note modifiée de la mélodie était égale, plus basse ou plus haute que la hauteur d'un son isolé présenté après le mélange. Les résultats ont révélé que les auditeurs étaient plus performants pour discriminer le contour mélodique lorsque celui-ci restait à l'intérieur du registre fréquentiel attendu plutôt que lorsqu'il survenait à l'extérieur. Un petit effet était également observé pour les jugements de hauteur, mais les performances étaient globalement proches du hasard. Ces faibles performances pourraient s'expliquer par le fait que la hauteur d'une note est probablement difficile à extraire de son contexte mélodique, d'autant plus que la présence de sons distracteurs devaient interférer dans ce jugement (Deutsch, 1970). Cette expérience suggère ainsi que les événements survenant à l'extérieur de la fenêtre d'attente, c'est-à-dire dans une région fréquentielle différente de celle

amorcée par la première mélodie, sont difficiles à percevoir. Les auteurs ajoutent que certains participants ont même rapporté avoir eu l'impression qu'une note avait disparue.

Les expérience 3 et 4 ont testé respectivement l'effet des attentes fréquentielles dans les jugements portant sur le contour et la hauteur lorsque la mélodie cible était cette fois présentée sans séquence distractrice. La procédure expérimentale était identique à celle employée dans l'expérience précédente. Les résultats ont révélé que les performances de discrimination de contour étaient équivalentes que la note de la mélodie cible soit présentée dans le registre fréquentiel attendu ou non. Et de façon surprenante, les jugements portant sur la hauteur s'amélioraient lorsque celle-ci était présentée à l'extérieure du registre fréquentiel attendu. Ces deux résultats montrent que contrairement à ce que l'on observait dans l'expérience 2, les événements tombant à l'extérieur de cette fenêtre d'attente ont cette fois été traités de façon relativement équivalente à ceux qui étaient attendus.

Cette étude suggère que la taille de la fenêtre d'attentes mise en jeu lors de la perception d'une mélodie varie en fonction de la présence ou non de sons distracteurs. Elle est plus étroite en présence de distracteurs qu'en leur absence. Focaliser sélectivement son attention sur une mélodie lorsqu'elle est présentée simultanément avec une autre séquence, revient donc à réduire la taille de la fenêtre d'attentes fréquentielles.

3.3.2 Encodage en mémoire de deux mélodies

Comme nous venons de le voir, la focalisation de l'attention induit une amélioration du traitement des attributs de l'événement focalisé, et ce gain décroît à mesure que l'on s'éloigne de cet événement (Dowling et al., 1987 ; Mondor & Bregman, 1994). Par conséquent dans une tâche de reconnaissance de mélodies intercalées, la séquence présentée simultanément à la mélodie qui n'est pas focalisée est probablement traitée de façon moins "approfondie" que ne l'est la mélodie, et elle est probablement atténuée perceptivement (Botte et al., 1997). Cependant, l'altération des performances de reconnaissance constatée lorsque la mélodie est présentée en présence d'une séquence distractrice par rapport à une situation dans laquelle elle est présentée seule (Dowling, 1973), suggère que la distractrice est tout de même encodée et stockée en mémoire avec la mélodie cible. En effet, dans l'étude de Dowling (1973, expérience 2) les performances de reconnaissance sont de 0.69 (aire sous la courbe MOC) lorsque la séquence distractrice est séparée de 12 dt de la mélodie cible, alors qu'elles sont de 0.98 dans une étude similaire menée par Dowling et Fujitani (1971, expérience 1) lorsque la mélodie est présentée isolément. Vliegen et Oxenham (1999, expérience 2) ont utilisé le même paradigme que Dowling (1973, expérience 2) et ont trouvé des proportions moyennes de réponses correctes de 0.89 chez les musiciens et 0.64 chez les non musiciens (conditions sons purs et sons complexes moyennées) lorsque la mélodie était séparée de 11 dt de la séquence distractrice. Ces performances demeurent encore une fois inférieures à celles obtenues sans

séquence distractrice (si l'on compare avec les résultats de Dowling et Fujitani, 1971, expérience 1, les auteurs n'ayant pas réalisé cette condition contrôle).

Deux hypothèses peuvent être formulées pour expliquer cette dégradation des performances de reconnaissance d'une mélodie en présence d'une autre séquence présentée simultanément :

1) L'altération des performances de reconnaissance pourrait être la conséquence d'un effet de distractibilité attentionnelle. L'attention de l'auditeur alternerait entre la mélodie et la distractrice, manifestant des difficultés à focaliser son attention de façon permanente sur un flux, de sorte que la détection d'un changement dans la mélodie cible puisse lui échapper.

2) Elle pourrait également s'expliquer par l'interférence en mémoire causée par la présence de la séquence distractrice. Dowling (1973, p. 330) en comparant les performances de reconnaissance obtenues avec et sans distractrice, évoque cette éventualité : *"This comparison shows that even at one octave separation the interleaved background interferes considerably with recognition"*.

L'existence d'un phénomène d'interférence en mémoire immédiate a été mise en évidence par Deutsch (1970), et concernait la mémoire de la hauteur tonale de sons isolés. Le paradigme désormais classique consistait à présenter deux sons successifs séparés par un intervalle de temps de 5 secondes, et à demander aux auditeurs de juger si leur hauteur était identique ou différente. Dans cette expérience princeps, ils s'agissait de sons purs et lorsque la hauteur de ces deux sons était différente, la différence était de ± 1 dt. Les auditeurs sélectionnés sur la base de leurs performances parfaites dans cette tâche de discrimination de hauteur en l'absence d'items interférents, passaient ensuite quatre conditions expérimentales dans lesquelles des items distracteurs étaient présentés pendant l'intervalle de temps séparant la présentation des deux sons tests. Dans les deux premières conditions, la tâche des auditeurs était de juger si la hauteur des deux sons était identique ou différente en ignorant les items interférents présentés, qui étaient respectivement des sons purs de différentes hauteurs ou des chiffres énoncés verbalement. Dans les deux autres conditions expérimentales, les items interférents étaient des chiffres énoncés verbalement, mais cette fois l'auditeur devait prêter attention à ces chiffres pour pouvoir les rappeler ultérieurement dans l'ordre dans lequel ils avaient été présentés. Ils réalisaient en plus la tâche de discrimination de hauteur dans la première condition, mais pas dans l'autre condition. Les participants étaient informés dans ce dernier cas que les deux sons tests présentés avaient une hauteur identique, et qu'ils devaient simplement noter leur réponse "même" à chaque essai. Les résultats de cette étude montrent que les auditeurs font plus d'erreurs de discrimination de hauteur des deux sons tests lorsque les items interférents sont des sons plutôt que des chiffres énoncés verbalement, que ceux-ci soient

ignorés ou non. Ce résultat suggère que la mémoire de la hauteur tonale des sons est affectée par la présence pendant l'intervalle de rétention de sons de même nature (sons purs plutôt que sons verbaux), appuyant l'idée de l'existence de modules de traitement distincts pour les sons verbaux et non verbaux.

Deutsch (1972) a mis en évidence dans une étude ultérieure que l'interférence était accrue lorsque la hauteur des sons interférents était proche de celle des sons tests (pour une différence de hauteur inférieure à 1 ton). Semal et Demany (1991, 1993) ont poursuivi cette direction de recherche plus récemment, et ont montré que cet effet d'interférence dépendait exclusivement de la proximité en hauteur entre les sons interférents et les sons tests. En effet, la diminution des performances de discrimination de hauteur des deux sons tests en présence des sons interférents présentés durant l'intervalle de rétention, persistait malgré des différences de timbre, d'intensité ou d'enveloppe d'amplitude entre les sons interférents et les sons tests. Ce résultat suggère que la mémoire de la hauteur tonale serait indépendante des variations portant sur une toute autre dimension.

Dans l'objectif de tester si l'absence d'effet du timbre sur cet effet d'interférence pouvait être généralisé à des sons de parole, Semal et ses collaborateurs (1996) ont reconduit l'étude menée par Deutsch (1970) qui avait montré que la parole n'interférait pas dans la mémoire de la hauteur de sons purs. Les auteurs ont contrôlé la hauteur des signaux verbaux présentés, de sorte que la différence de hauteur entre les items interférents et les sons tests étaient équivalents qu'il s'agisse de sons de parole ou de sons purs. Les résultats ont révélé que la proximité de la fréquence fondamentale des sons interférents avec celle des sons tests altérait les performances de discrimination de hauteur des sons tests, à la fois pour les sons verbaux et non verbaux.

Ces différentes recherches montrent donc que la mémoire de la hauteur tonale d'un son présenté isolément est affectée par la présentation d'autres sons durant l'intervalle de rétention, lorsque la hauteur de ces sons interférents est proche de celle du son mémorisé quelle que soit la nature de ce son, qu'il ait ou non le même timbre et que ce soit ou non un son de parole. A ma connaissance, aucune recherche analogue à celles réalisées par Deutsch (1970, 1972) ou par Sémal et ses collaborateurs (Semal & Demany, 1991, 1993 ; Semal et al., 1996) examinant de façon systématique l'effet de la présentation d'items interférents pendant l'intervalle de rétention sur la mémoire immédiate des mélodies, n'a été entreprise. Seuls Dowling, Kwak et Andrews (1995) ont montré à l'aide d'une tâche de mémorisation mélodique en continu (figure 3.3) que la présence de mélodies pendant un intervalle de rétention supérieur à 7 secondes affectait la mémorisation du contour mélodique. Dans une première expérience, les auteurs ont présenté à cinquante-sept auditeurs, des mélodies de 7 notes tonales et atonales successivement séparées par un silence de 7 secondes. Les participants devaient juger à chaque essai si la mélodie avait déjà été présentée précédemment. Les paires de mélodies test étaient soient identiques, la mélodie à comparer était alors une transposition exacte de la mélodie standard, soient différentes, dans ce cas la mélodie à comparer était une transposition inexacte, préservant ou

non le contour original. Elles étaient présentées successivement, c'est-à-dire après un silence de 7 sec, ou bien espacées par un intervalle de 17, 27, 47 ou 87 sec soit par 1, 2, 4, ou 8 mélodies respectivement. Les auteurs ont constaté que les auditeurs parvenaient mieux à distinguer une transposition exacte d'une transposition inexacte si le contour de la mélodie était modifié après un silence de 7 secondes, résultat en accord avec les recherches précédentes (Dowling, 1978 ; Dowling & Fujitani, 1971). En revanche et de façon surprenante, ils ont remarqué que ce résultat s'inversait pour les intervalles de rétention de 17 secondes et plus, occupés par les items interférents. Les auditeurs parvenaient mieux à distinguer les leurres qui imitaient le contour de la mélodie standard plutôt que ceux dont le contour avait été modifié. Afin de distinguer l'effet de la durée de l'intervalle séparant la présentation des deux mélodies à comparer et l'effet d'interférence provoquée par la présence de mélodies durant ces intervalles, les auteurs ont examiné les performances de reconnaissance pour des délais comparables mais cette fois en l'absence de mélodies (expérience 4). La tâche consistait donc à comparer deux mélodies par paires séparées par un silence de 7, 12 et 33 sec. Les performances obtenues déclinaient avec l'augmentation de la durée de l'intervalle de rétention, mais les auditeurs demeuraient plus performants pour discriminer une transposition inexacte d'une transposition exacte lorsque le contour était modifié quel que soit ce délai. Par conséquent, l'inversion constatée dans la première expérience dans laquelle des mélodies étaient présentes pendant les intervalles supérieurs à 7 secondes, était due à la présence de ces items interférents. Cette étude suggère que la présence de mélodies durant l'intervalle de rétention provoque une interférence dans la mémorisation mélodique à court terme, c'est-à-dire pour l'encodage du contour mais pas pour celui des intervalles. Elle met ainsi en évidence l'existence d'une interférence dans la mémoire des mélodies, et révèle que le contour et les intervalles ne sont pas affectés de la même façon par la présence d'items interférents. Ce résultat confirme une fois de plus le fait que le contour soit encodé en mémoire de travail (sensible à l'interférence), et les intervalles en mémoire à long terme (Dowling, 1978 ; Dowling & Harwood, 1986).

L'interférence en mémoire constatée dans les tâches de reconnaissance de mélodies intercalées due à la présence de la distractrice pourrait donc être analogue à celle décrite par Deutsch (1970, 1972), ou Dowling et ses collaborateurs (1995). La différence résiderait dans le fait qu'il s'agirait d'une interférence en mémoire qui surviendrait pendant l'encodage de la mélodie, et non au cours du stockage. En effet, l'encodage des propriétés de la séquence distractrice pourrait interférer avec l'encodage de celles de la mélodie cible, puisque ces indices sont probablement traités dans le même module. Cette interférence se produirait donc uniquement dans le cas où les deux séquences sont encodées simultanément. Par conséquent, si la mélodie cible est déjà familière à l'auditeur, c'est-à-dire si elle a déjà été encodée et stockée en mémoire antérieurement, la présence de la distractrice ne devrait pas affecter sa représentation en mémoire. Si en revanche l'altération des performances induite par la présence des sons distracteurs était due à un phénomène de distractibilité attentionnelle, les performances de

reconnaissance ne devraient à priori pas être sensibles au fait que la mélodie soit familière ou non. Comme nous l'avons vu précédemment, la dégradation des performances de reconnaissance a été constatée dans le cas de mélodies nouvelles (Dowling, 1973, expérience 2 ; Dowling & Fujitani, 1971, expérience 1 ; Vliegen & Oxenham, 1999, expérience 2). En revanche, les mélodies familières sont reconnues lorsqu'elles sont séparées d'une octave en moyenne de façon parfaite ou quasi parfaite (Dowling, 1973, expérience 1 ; Dowling et al., 1987, expérience 1 ; Hartmann & Johnson, 1991), et même lorsqu'elles sont présentées dans le même registre que la distractrice (Dowling, 1973, expérience 3 ; Dowling et al., 1987, expérience 1). Il est malheureusement difficile de tirer des conclusions sur la base de ces études qui demeurent trop peu nombreuses et qui ne relatent pas toujours de façon précise les performances de base des auditeurs dans une tâche de reconnaissance de mélodies présentées sans séquence distractrice (condition contrôle souvent omise).

Toutes les études décrites dans ce chapitre apportent des éléments de réponse à la question de savoir sous quel format une mélodie est stockée en mémoire, et mettent en évidence les spécificités attentionnelles et mnésiques de l'écoute polyphonique. Les opérations mentales impliquées dans la perception d'une mélodie présentée simultanément avec une autre séquence, sont encore peu explorées. Elles ont été examinées dans le cadre de cette thèse, et les résultats des recherches menées sont relatés dans la partie expérimentale suivante. Ces travaux conduits auprès de sujets sains et de patients cérébro-lésés, nous amènent à proposer une analyse fonctionnelle des différentes composantes cognitives impliquées dans la reconnaissance d'une mélodie intercalée à des sons distracteurs.

Partie Expérimentale

Chapitre 4

Post-reconnaissance de mélodies intercalées et formation des flux auditifs : Effet d'une différence de hauteur moyenne et d'une dissemblance de timbre

Afin de mesurer de façon objective les processus primaires impliqués dans la formation des flux auditifs, nous avons élaboré un paradigme de post-reconnaissance de mélodies intercalées. Quatre expériences ont été conduites afin de tester cette nouvelle situation expérimentale. L'effet de la différence de hauteur moyenne et de la dissemblance de timbre sur la post-reconnaissance d'une mélodie cible intercalée, est examiné. La notion de fission mélodique partielle et le seuil de reconnaissance de mélodies intercalées (SRM) sont proposés pour établir un lien entre ce type de méthode indirecte et les méthodes directes mesurant le seuil de fission d'une séquence alternante.

4.1 Introduction

Dans ce chapitre, nous proposons d'aborder la question de la mesure des processus impliqués dans la formation des flux auditifs. Nous avons vu dans le premier chapitre de ce document, que les processus primaires impliqués dans la formation des flux auditifs avaient été jusqu'à présent principalement examinés par des méthodes mesurant le seuil de fission d'une séquence cyclique (voir Bregman, 1990, chap. 2 pour une synthèse). Ces méthodes qui présentent l'avantage d'accéder directement aux processus d'analyse de scène auditive, ont cependant deux inconvénients majeurs : 1) elles utilisent des séquences prédictibles faisant probablement intervenir des processus d'analyse de scène guidée par les connaissances

(Bregman, 1990, chap. 4), et 2) elles fournissent une réponse subjective puisque l'auditeur rapporte l'expérience qu'il a du changement de l'analyse perceptive (nombre de flux perçus), sans qu'il soit possible de vérifier sa réponse.

Quatre expériences ont donc été menées dans l'objectif : 1) de mettre en place une méthode indirecte mesurant de façon objective les processus primaires impliqués dans la formation des flux auditifs, et 2) : d'établir un lien entre les résultats obtenus par ce type de méthode sondant la précision de la représentation des propriétés d'un flux cible (performances de reconnaissance) et les résultats obtenus par des méthodes directes examinant le phénomène de fission perceptive (seuil de fission).

Nous avons élaboré un paradigme de post-reconnaissance de mélodies intercalées. Cette situation expérimentale est issue du paradigme de reconnaissance immédiate de mélodies non familières intercalées proposé par Dowling (1973, expérience 2). Dans le paradigme original, les auditeurs devaient comparer deux mélodies successives, une mélodie standard présentée seule suivie d'une mélodie cible intercalée à des sons distracteurs. Dans ce paradigme de post-reconnaissance, la mélodie standard est présentée après le "mélange" afin de minimiser les connaissances préalables dont dispose l'auditeur pour réaliser cette tâche et ainsi faire intervenir majoritairement des processus primaires d'organisation.

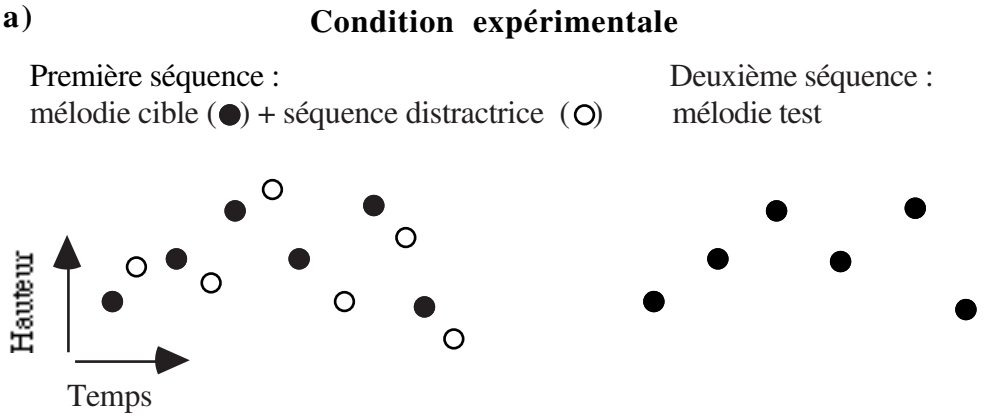
La première expérience a été conduite dans le but de tester ce paradigme. Il fallait en particulier vérifier le pouvoir distracteur des séquences intercalées, afin de s'assurer que la tâche ne soit réalisable qu'à la condition que l'auditeur ait séparé perceptivement au moins partiellement les sons de la mélodie. Cette expérience vise à établir un lien entre l'organisation perceptive d'une séquence en 1 ou 2 flux et les performances de post-reconnaissance d'une mélodie intercalée. Les deux expériences suivantes ont examiné respectivement l'effet de la différence de hauteur moyenne séparant la mélodie cible de la séquence distractrice sur les performances de post-reconnaissance de la mélodie cible lorsque les séquences étaient jouées par des sons purs et complexes (composés de 6 harmoniques de même amplitude) (expérience 2), et l'effet de la dissemblance de timbre (expérience 3). Un lien entre les performances de post-reconnaissance de mélodies intercalées et le seuil de fission d'une séquence (c'est-à-dire la valeur au-delà de laquelle une séquence alternante organisée initialement en un flux se scinde en deux flux), a été établi et un seuil de reconnaissance de mélodies intercalées a été défini. Enfin, pour s'assurer que les performances obtenues étaient bien le reflet de la représentation immédiate que l'auditeur avait de la séquence composite et non une reconstruction a posteriori (extraction des indices mélodiques après répétition mentale de la séquence durant l'intervalle de silence), nous avons mené une quatrième expérience dans laquelle nous avons réduit l'intervalle de temps séparant la séquence composite et la mélodie.

4.2 Méthode Générale

Nous avons mesuré la capacité des auditeurs à organiser une séquence en deux flux à l'aide d'une tâche de reconnaissance immédiate de mélodies non familières intercalées (Dowling, 1973, expérience 2). Afin que l'auditeur ne connaisse pas la mélodie qu'il doit extraire du mélange, la mélodie à reconnaître a été présentée *après* le "mélange" composé de la mélodie cible intercalée à une séquence distractrice. Cette tâche de *post*-reconnaissance de mélodies intercalées a été élaborée dans le but de minimiser les connaissances mises à la disposition de l'auditeur pour organiser perceptivement la séquence composite en deux flux, et ainsi faire intervenir majoritairement des processus primaires d'analyse de scène auditive (processus ascendants) (Bregman, 1990).

4.2.1 Une tâche de post-reconnaissance de mélodies intercalées

Deux mélodies de six sons ont été présentées : 1) une mélodie cible intercalée à une séquence distractrice, et 2) une mélodie test. La mélodie test était identique à la mélodie cible ou en différait par deux notes. La tâche de l'auditeur consistait à déterminer si les mélodies cible et test étaient identiques ou différentes. Les notes impaires de la séquence composite correspondaient aux notes de la mélodie et les notes paires à celles de la séquence distractrice. Le "mélange" débutait donc par la première note de la mélodie et s'achevait par la dernière note de la séquence distractrice. La mélodie cible était toujours présentée dans le même registre de hauteur et avec le même timbre, seule la distractrice variait en hauteur ou en timbre selon l'expérience. Cette condition expérimentale était suivie d'une condition contrôle dans laquelle la mélodie cible était présentée sans séquence distractrice. Il s'agissait alors d'une simple tâche de reconnaissance immédiate de mélodies (Dowling & Fujitani, 1971). Ces deux conditions expérimentale et contrôle sont illustrées sur la figure 4.1. Les performances de post-reconnaissance de mélodies intercalées ont été comparées aux performances de simple reconnaissance de mélodies dans chacune des expériences.



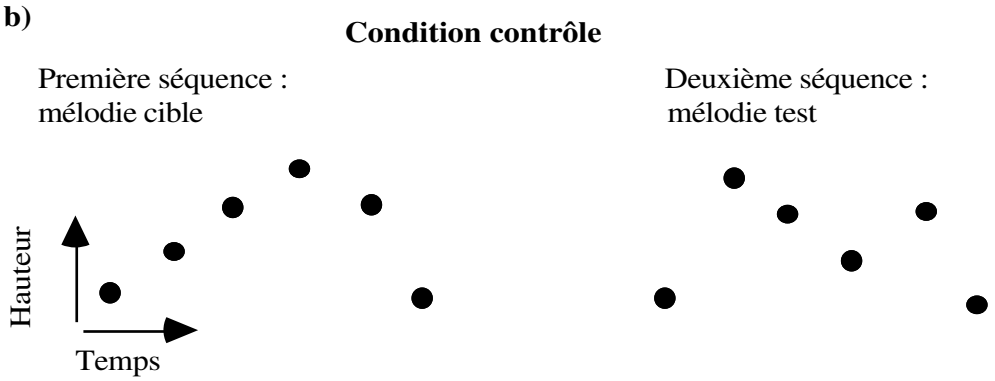


Figure 4.1. Illustration visuelle des deux types de conditions présentées aux auditeurs : **a)** une condition expérimentale dans laquelle la mélodie cible est présentée intercalée à une séquence distractrice, un exemple d'essai où les mélodies cible et test sont identiques est représenté ; **b)** une condition contrôle dans laquelle la mélodie cible est présentée seule, un exemple d'essai où les mélodies cible et test diffèrent par deux notes, la deuxième et la quatrième, est représenté.

4.2.2 Stimuli

Les sons composant ces séquences étaient des sons de synthèse d'une durée de 110 ms. L'intervalle de temps séparant le début de deux sons successifs (*Inter Onset Interval*, IOI) était de 165 ms pour la première séquence composée de 12 notes (les 6 notes de la mélodie cible intercalées avec les 6 notes de la séquence distractrice), et de 330 ms pour la deuxième séquence composée des 6 notes de la mélodie test. Les deux séquences étaient séparées par un silence d'une durée de 1870 ms.

4.2.3 Construction des mélodies et des distractrices

36 mélodies et 180 distractrices composées chacune de 6 notes ont été élaborées. Les intervalles sont fixes mais la hauteur moyenne à laquelle ces séquences ont été présentées fluctuait à chaque essai dans une gamme allant de -3 à +2 dt.

4.2.3.1 Les mélodies

Les mélodies sont composées d'intervalles ascendants et descendants dont la taille varie dans une gamme allant de 0 à 8 dt, soit un rapport entre les fréquences fondamentales des sons successifs compris entre 1 (sept mélodies ont des notes répétées) et 1.6. Toutes les mélodies sont jouées à l'intérieur d'une octave. Leur ambitus varie dans une gamme allant de 5 à 11 dt. La note moyenne des mélodies est le la4 (note MIDI 81), soit une fréquence fondamentale de 880 Hz dans les expériences où la différence de hauteur varie (expériences 1, 2 et 4), et le mib3

(note MIDI 63), soit une fréquence fondamentale de 311 Hz dans l'expérience 3 où l'indice de ségrégation est le timbre. Chacune des 36 mélodies élaborées présente une version originale et une version dite modifiée dans laquelle deux notes ont été changées dans une gamme de ± 4 dt, la deuxième et la quatrième ou la troisième et la cinquième note. Pour l'ensemble des mélodies, le changement de ces deux notes altère le contour mélodique original c'est-à-dire modifie la direction du changement de hauteur. Le contour mélodique constitue un indice prégnant dans la reconnaissance immédiate de mélodies non familières, qu'elles soient tonales ou atonales (Dowling, 1978 ; Dowling & Fujitani, 1971). Ces versions, originale et modifiée, sont affectées aux mélodies cible et test de façon équivalente. Au total sur les 72 mélodies différentes, 46 sont tonales et 26 sont atonales. La liste des mélodies et leurs caractéristiques respectives se trouvent en annexe (annexes A1 et A2 respectivement).

4.2.3.2 Les distractrices

Cinq séquences distractrices différentes ont été construites pour chacune des 36 mélodies (annexe B1). Il s'agit de séquences atonales qui répondent à deux contraintes : 1) elles alternent entre des fréquences élevées et basses afin de chevaucher le plus possible la mélodie dans ses deux versions (Hartmann & Johnson, 1991), et 2) elles encadrent la mélodie cible lorsqu'elles se trouvent dans le même registre de hauteur (annexe B2 et figure B2.1). Cette dernière contrainte a été ajoutée à la suite d'expériences préliminaires qui ont montré que lorsque la mélodie cible et la séquence distractrice étaient présentées dans le même registre et que les mélodies à comparer différaient par deux notes, les auditeurs avaient des performances supérieures au hasard (la proportion moyenne de reconnaissance correcte était de 0.60). Ce résultat suggérait que les auditeurs parvenaient à extraire un indice leur permettant de réaliser la tâche alors que la mélodie n'était pas séparée perceptivement de la séquence distractrice. Une écoute informelle nous a conduit à émettre l'hypothèse qu'ils avaient pu se baser sur le contour global de la séquence composite pour inférer celui de la mélodie intercalée. Afin d'éviter que l'auditeur puisse développer ce type de stratégie, nous avons donc appliqué cette deuxième contrainte faisant en sorte que les sons les plus aigus et les sons les plus graves de la séquence composite appartiennent à la séquence distractrice.

4.2.4 Procédure

Les expériences se sont déroulées en deux phases, une phase expérimentale suivie d'une phase contrôle. Chacun des tests était précédé de 8 essais de familiarisation de difficulté croissante durant lesquels la réponse correcte était donnée aux auditeurs une fois leur jugement effectué. Pour chacune des expériences réalisées, une consigne écrite était remise aux participants (annexe C1-7 pour les expériences 1 à 7), et à la fin de la passation un questionnaire leur était donné à remplir afin de connaître la formation musicale de chacun (annexe C10). Les

auditeurs étaient placés dans une salle insonorisée et les séquences étaient diffusées sous écouteurs (Sennheiser HD 520 II). Les participants donnaient leur réponse en appuyant sur deux touches du clavier, "m" s'il s'agissait des mêmes mélodies et "d" si elles leur semblaient différentes. L'essai suivant était lancé une fois leur réponse donnée.

Les mélodies et les séquences distractrices correspondantes étaient choisies aléatoirement à chaque essai. Dans chaque condition de séparation en hauteur ou en timbre, un nombre égal d'essais "même" et "différent" était présenté. Dans les essais "même" on affectait aux mélodies cible et test la même version, originale ou modifiée. Dans les essais "différents", on leur affectait des versions différentes (originale et modifiée ou modifiée et originale). Au total, quatre types d'essais différents ont été présentées (oo, mm, om, mo) en nombre égal dans chaque condition de séparation.

Des analyses de variance au seuil de significativité de .05 ont été effectuées sur l'arcsinus de la racine carrée de la proportion moyenne de réponses correctes, afin de remédier à la violation du postulat d'égalité des variances intra-groupes. Une correction de Greenhouse-Geisser (1959) a été appliquée si nécessaire en raison de l'inhomogénéité de la covariance. Nous avons également calculé les proportions de détections correctes (proportions de réponse "différentes" quand les deux mélodies sont différentes) et de fausses alarmes (proportions de réponse "différentes" lorsque les deux mélodies sont identiques) (Green & Swets, 1974 ; Macmillan & Creelman, 1991). Nous avons ainsi déterminé pour chaque condition de séparation le d' de chaque auditeur (expérience 1), ou un d' de groupe (Macmillan & Kaplan, 1985) lorsque le nombre d'observations par individu était insuffisant pour calculer ce paramètre individuellement (expériences 2, 3 et 4).

4.2.5 Matériel

Les sons ont été générés par un synthétiseur Yamaha TX802 FM Tone Generator par une technique de modulation de fréquence (Chowning, 1973). Le TX802 était piloté par un micro-ordinateur Macintosh SE/30 via une interface MIDI. Les programmes des différentes expériences ont été écrits en LISP.

4.3 Expérience 1 : Post-reconnaissance de mélodies intercalées et nombre de flux perçus

4.3.1 Introduction

Cette première expérience a été conduite dans le but de tester si ce paradigme de post-reconnaissance de mélodies intercalées permettait de mesurer les processus impliqués dans la construction des flux auditifs. Nous avons donc examiné si les performances de post-reconnaissance de la mélodie cible dépendaient de l'organisation perceptive de la séquence composite formée par la mélodie cible intercalée avec la séquence distractrice. La séquence distractrice a été présentée dans le même registre que la mélodie cible et éloignée de 24 dt en moyenne vers les fréquences basses. Les recherches antérieures ont montré que le seuil de fission d'une séquence dans laquelle deux sons purs alternaient était d'environ 2-3 dt (Miller & Heise, 1950 ; van Noorden, 1975) et que les auditeurs parvenaient à extraire une mélodie intercalée pour une différence de hauteur moyenne comprise entre 6 et 12 dt (Dowling, 1973, expériences 1 et 2 ; Hartmann & Johnson, 1991 ; Vliegen & Oxenham, 1999, expérience 2). Par conséquent, la mélodie cible ne devrait pas pouvoir être extraite du mélange pour une différence de hauteur moyenne nulle, mais elle devrait être séparée perceptivement de la séquence distractrice pour une différence moyenne de 24 dt. Ainsi, si la réalisation de cette tâche dépend des processus primaires d'organisation, les performances devraient être aléatoires (50 % de réponses correctes) lorsque cette différence est nulle et comparables à celles obtenues en l'absence de séquence distractrice (condition contrôle) lorsque la différence est de 24 dt.

L'objectif de cette expérience est donc de tester le pouvoir distracteur de chacune des séquences intercalées à la mélodie cible, ceci afin de s'assurer que l'auditeur ne peut réaliser cette tâche que s'il a organisé perceptivement la séquence composite en deux flux. Cette expérience permet également de contrôler l'équivalence de la discriminabilité de l'ensemble des mélodies élaborées. En effet, les trente-six mélodies construites présentent des caractéristiques différentes du point de vue du contour, de la position des notes changées, de l'ambitus du changement et surtout de leur appartenance ou non au système musical diatonique, certaines étant tonales d'autres atonales (annexe A2). Or ces caractéristiques s'avèrent importantes dans la perception et la mémoire des mélodies (Dowling & Harwood, 1986).

4.3.2 Méthode

4.3.2.1 Participants

Dix-sept personnes ont participé à l'expérience. Seules quatorze ont été intégrées à l'analyse, neuf femmes et cinq hommes d'un âge moyen de 26.5 ans [17 ; 36] dont neuf avaient reçu une formation musicale. Les résultats de trois participants ont été écartés. Deux répondaient

au hasard en présence de séquence distractrice même lorsqu'elle était distante de la mélodie de 24 dt (proportions correctes de 0.57 en moyenne), soit un taux d'échec de 12 %. Et un auditeur avait des performances supérieures au hasard (0.59 ; $p < .000001$) dans la condition où la mélodie et la distractrice étaient présentées dans le même registre fréquentiel. Il se trouve que cet auditeur avait l'oreille absolue et a rapporté avoir eu recours à une autre stratégie pour réaliser la tâche. Cette stratégie a consisté à mémoriser l'ensemble des notes de la séquence composite, à restituer une note sur deux (correspondant aux notes de la mélodie cible) et à comparer ces notes à celles de la mélodie test. Notre intérêt étant de faire le lien entre la formation des flux et la perception de mélodies intercalées, nous avons écarté tout auditeur ayant l'oreille absolue dans la mesure où leurs performances ne dépendaient pas seulement de leur aptitude à séparer perceptivement la mélodie cible de la séquence distractrice.

4.3.2.2 Stimuli

Nous avons testé les cinq séquences distractrices différentes (versions a, b, c, d et e) associées à chaque mélodie construites sur le même principe (annexe B2). Deux conditions de séparation en hauteur ont été présentées : la distractrice était présentée dans le même registre que la mélodie cible (la différence de hauteur moyenne était nulle) ou s'en éloignait en moyenne de 24 dt vers les fréquences basses (CD plages 1 et 9 respectivement), ainsi qu'une condition contrôle dans laquelle la mélodie cible était présentée sans séquence distractrice (CD plages 10 et 11). Les sons composant les séquences étaient des sons purs.

4.3.2.3 Procédure

Les participants ont passé cinq conditions expérimentales dans un ordre contrebalancé, testant respectivement les cinq séquences distractrices, puis la condition contrôle. Ces six tests ont été proposés respectivement au cours de six sessions différentes à raison d'une par jour. Chaque séance durait une heure environ et une pause était aménagée à la moitié. La consigne écrite remise aux participants figure dans l'annexe C1. Chaque condition expérimentale comportait 216 essais. Les quatre structures d'essais possibles (les deux essais identiques composés des versions originales ou des versions modifiées des mélodies, et les deux essais différents composés des deux versions dans les deux ordres) combinées aux 36 mélodies étaient présentés à chaque participant dans la condition où la séquence distractrice et la mélodie cible étaient présentées dans le même registre de hauteur, soit 144 essais. Dans la condition de séparation en hauteur de 24 dt, seuls 2 types d'essai (un identique, un différent) étaient combinés aux 36 mélodies afin de ne pas alourdir l'expérience, soit 72 essais mais les quatre structures d'essai étaient équireprésentées sur l'ensemble des participants. Chacun de ces essais était choisi aléatoirement. La condition contrôle comportait 144 essais. Dans l'ensemble de ces tests, un nombre égal d'essais "même" et "différent" était présenté.

4.3.3 Résultats et discussion

4.3.3.1 Les performances de post-reconnaissance de mélodies intercalées dépendent de la construction des flux auditifs

Nous avons calculé les proportions moyennes de réponses correctes obtenues par les 14 auditeurs pour l'ensemble des séquences distractrices testées dans les trois conditions, les deux conditions expérimentales dans lesquelles respectivement la séquence distractrice et la mélodie étaient séparées en moyenne de 0 et 24 dt et la condition contrôle. Lorsque la mélodie et la séquence distractrice étaient présentées dans la même gamme de fréquences, les performances moyennes obtenues étaient proches du hasard (0.52, valeur légèrement supérieure à 0.51 qui est la proportion au-delà de laquelle la probabilité qu'elle appartienne à une distribution normale pour n = 10080 est inférieure à .01). Lorsque la séquence distractrice était éloignée de 24 dt en moyenne de la mélodie cible, les performances (0.94) se rapprochaient de celles obtenues dans la condition contrôle (0.97) mais demeuraient légèrement inférieures (figure 4.2).

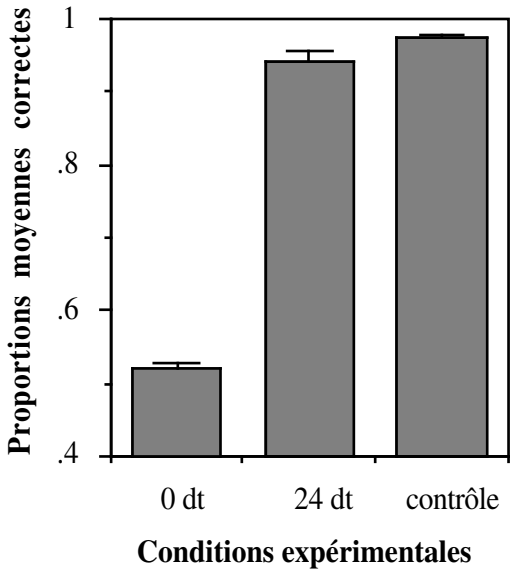


Figure 4.2. Proportions moyennes de réponses correctes obtenues par 14 auditeurs pour l'ensemble des séquences distractrices dans les différentes conditions de séparation de hauteur entre la mélodie cible et la séquence distractrice (0 et 24 dt) et dans la condition contrôle. Les moyennes ont été calculées à partir de 10080 observations dans la condition 0 dt (144 essais * 5 distractrices * 14 participants), 2016 (144 essais * 14 participants) dans la condition contrôle, et de 5040 (72 essais * 5 distractrices * 14 participants) dans la condition 24 dt. Les barres verticales indiquent ± 1 erreur-type.

Une analyse de variance (plan d'analyse S14*H3) évaluant l'effet des conditions (les deux conditions de séparation en hauteur 0 et 24 dt et la condition contrôle) sur les performances de reconnaissance, a révélé que les proportions de réponses correctes obtenues dans ces trois conditions étaient globalement significativement différentes, $F(2, 26) = 207.11$, $p < .0001$. L'analyse planifiée des contrastes a confirmé que les performances obtenues dans la condition de séparation 24 dt étaient significativement inférieures à celles obtenues dans la condition contrôle, $F(1, 26) = 5.42$, $p < .05$. Ce résultat suggère que même si l'auditeur est parvenu à séparer perceptivement la mélodie cible de la séquence distractrice, la présence de la séquence distractrice affecte légèrement la reconnaissance de la mélodie cible.

Nous avons également calculé les proportions de détections correctes et de fausses alarmes obtenues par chaque participant pour l'ensemble des distractrices dans les deux conditions expérimentales et dans la condition contrôle afin de déterminer le d' de chaque auditeur (Green & Swets, 1974 ; Macmillan & Creelman, 1991). La moyenne des d' obtenus par les 14 participants confirme les résultats observés à partir des proportions de réponses correctes (tableau 4.1). Lorsque la séquence composite est organisée perceptivement en 1 flux, la sensibilité pour discriminer les deux mélodies cible et test est quasi nulle alors que lorsque l'auditeur parvient à extraire la mélodie cible du mélange (24 dt), le d' a une valeur proche de celle obtenue en l'absence de distractrice qui elle-même est proche de la valeur maximale que peut prendre cet indice dans un modèle même/différent (6.9).

	Moyenne d'	**Ecart-types**	**Erreur-types**
condition 0 dt	0.6	0.4	0.1
condition 24 dt	5.5	1.3	0.4
condition contrôle	6.1	0.8	0.2

Tableau 4.1. Moyenne des d', écart-types et erreur-types obtenus par 14 auditeurs dans les deux conditions de séparation en hauteur entre la mélodie cible et la séquence distractrice (0 et 24 dt) et la condition contrôle.

Les performances obtenues dans cette tâche de post-reconnaissance de mélodies intercalées dépendent donc de la façon dont les auditeurs organisent perceptivement la première séquence composée de la mélodie cible intercalée à la séquence distractrice, à l'exception des auditeurs ayant l'oreille absolue. Les performances sont aléatoires lorsque cette séquence composite est perçue comme un seul flux et proches de celles obtenues dans une simple tâche de reconnaissance de mélodies lorsque la séquence composite est organisée en deux flux.

4.3.3.2 Efficacité des distractrices

Contrairement à ce qui avait été trouvé dans des expériences préliminaires réalisées avec d'autres séquences distractrices, lorsque la mélodie cible et la séquence distractrice étaient présentées dans le même registre, les auditeurs ne sont pas parvenus à reconnaître la mélodie cible. Une analyse détaillée des proportions de réponses correctes obtenues par l'ensemble des auditeurs pour chacune des 180 associations de mélodies et de distractrices différentes présentées, a révélé que toutes ces combinaisons conduisaient à des performances aléatoires. Ce résultat révèle l'efficacité des séquences distractrices construites par cet algorithme pour "brouiller" la mélodie cible (annexe B2). La contrainte ajoutée dans cette expérience, faisant en sorte que les séquences distractrices encadrent la mélodie lorsqu'elles se trouvent dans le même registre de hauteur, a donc permis d'éviter que les auditeurs ne développent d'autres stratégies pour accomplir cette tâche. Elle permet ainsi de s'assurer que la tâche n'est réalisable qu'à condition de séparer au moins partiellement les sons de la mélodie cible du mélange.

4.3.3.3 Discriminabilité des mélodies

Les erreurs de discrimination mélodique effectuées par les auditeurs dans la condition contrôle et dans la condition où le degré de séparation était en moyenne de 24 dt étaient peu nombreuses et distribuées de façon équivalente sur l'ensemble des mélodies. Les caractéristiques différentes des mélodies utilisées, en particulier le fait que certaines soient tonales d'autres atonales, n'ont pas induit de différences de discriminabilité entre ces mélodies. Ceci peut s'expliquer par le fait que dans cette tâche, la discrimination mélodique est relativement aisée comme l'attestent les performances élevées obtenues par les auditeurs dans la condition contrôle. De nombreux indices sont en effet mis à la disposition de l'auditeur pour comparer ces mélodies : deux notes sont changées sur six, le contour est modifié et parfois même la tonalité (annexe A2).

4.4 Expérience 2 : Post-reconnaissance de mélodies intercalées sur la base d'une différence de hauteur

4.4.1 Introduction

La première expérience a donc montré que la proportion moyenne de post-reconnaissance de mélodies intercalées était de 0.52 lorsque la séquence composite était organisée perceptivement en un flux et de 0.94 lorsqu'elle était organisée en deux flux, laissant ainsi une gamme de variation de performances d'environ 0.42 pour mesurer les processus de fission perceptive à mesure que la différence de hauteur moyenne s'accroît. L'objectif de la deuxième expérience est précisément d'évaluer l'effet de l'augmentation de la différence de

hauteur moyenne séparant la mélodie cible de la séquence distractrice sur les performances de post-reconnaissance de la mélodie cible. Les performances obtenues pour les différents degrés de séparation en hauteur testés (0, 6, 12 et 24 dt) sont comparées au seuil de fission établi par des méthodes directes (Miller & Heise, 1950 ; van Noorden, 1975). La notion de fission mélodique partielle et le seuil de reconnaissance de mélodies intercalées sont proposés pour établir un lien entre les résultats obtenus par des méthodes indirectes, évaluant les performances de reconnaissance d'un pattern cible émergeant de la formation des flux auditifs, et les méthodes directes mesurant le seuil de fission d'une séquence alternante.

Les séquences sont jouées par des sons purs et des sons complexes composés de 6 harmoniques de même amplitude, afin d'examiner l'effet de la nature du son sur les performances à cette tâche. Nous avons ainsi testé si un plus grand recouvrement spectral entre les sons complexes de la mélodie et ceux de la séquence distractrice, allait affecter la fission mélodique. Certains auteurs considèrent en effet que la fission perceptive des sons successifs est due à la stimulation de filtres auditifs distincts (Anstis & Saida, 1985 ; Beauvois & Meddis, 1991 ; Hartmann & Johnson, 1991 ; van Noorden, 1975). Selon cette conception périphérique de la formation des flux, la stimulation des mêmes filtres devraient donc réduire la fission perceptive. Le degré de recouvrement de l'activité des filtres étant plus important lorsque les séquences sont jouées par des sons complexes que par des sons purs, la différence de hauteur moyenne nécessaire pour extraire la mélodie cible devrait par conséquent être plus élevée dans le cas où la séquence est composée de sons complexes.

4.4.2 Méthode

4.4.2.1 Participants

Vingt et une personnes ont participé à l'expérience. Les performances de cinq participants n'ont pas été prises en compte dans l'analyse. Une personne a signalé des problèmes auditifs après la passation. Quatre autres ont échoué au test, soit un taux d'échec de 19 %. Les performances de ces dernières étaient comprises entre 72 et 91 % de réponses correctes dans la condition contrôle, mais leurs réponses étaient aléatoires en présence de la séquence distractrice même lorsqu'elle était distante de la mélodie de 24 dt (pourcentages de réponses correctes compris entre 44 et 63 %). L'analyse a donc porté sur les résultats de seize auditeurs, dix femmes et six hommes d'un âge moyen de 29 ans [23 ; 52] dont dix avaient reçu une formation musicale.

4.4.2.2 Stimuli

Nous avons testé quatre degrés de séparation de hauteur moyenne. La distractrice était présentée dans le même registre que la mélodie cible ou s'en éloignait en moyenne de 6, 12 ou

24 demi-tons (dt) vers les fréquences basses. Les sons utilisés étaient soit des sons purs, soit des sons complexes harmoniques constitués de 6 fréquences de même amplitude. La sonie a été égalisée entre les deux types de sons. Le niveau physique maximum atteint par chacun des sons a été mesuré à l'aide d'un sonomètre Bruel & Kjaer 2209 (pondération A, intégration rapide). Il était de 76.2 dBA pour les sons purs et de 67 dBA pour les sons complexes.

Un extrait des séquences diffusées dans les différentes conditions de séparation en hauteur (0, 6, 12 et 24 dt) se trouve sur le disque respectivement au niveau des plages 1, 6, 8 et 9 pour les séquences jouées avec des sons purs et au niveau des plages 12, 13, 14 et 15 pour les séquences jouées par des sons complexes. Les plages 10 et 11 présentent deux exemples d'essais "même" et "différent" présentés dans la condition contrôle lorsque les séquences sont jouées par des sons purs et la plage 16 présente un exemple avec des sons complexes.

4.4.2.3 Procédure

Les participants ont passé deux conditions expérimentales, une avec des sons purs et une avec des sons complexes en ordre contrebalancé, puis la condition contrôle avec des sons purs pour la moitié des auditeurs et des sons complexes pour l'autre moitié. Chacun de ces trois tests comportait 32 essais (choix aléatoire de 32 mélodies parmi les 36). Ainsi chaque mélodie et distractrice associée (choisie aléatoirement par les 5) était présentée une seule fois dans chacun des tests. La passation durait 30 minutes environ. La consigne écrite délivrée aux participants se trouve dans l'annexe C2.

4.4.3. Résultats et discussion

4.4.3.1 Effet de la nature du son sur la fission mélodique

Les performances obtenues lorsque la mélodie cible et la séquence distractrice étaient jouées par des sons complexes étaient équivalentes à celles obtenues pour des sons purs (tableau 4.2).

Une analyse de variance évaluant l'effet de la nature du son pour les quatre degrés de séparation en hauteur (plan d'analyse S*N2*H4), n'a pas révélé d'effet de ce facteur, $F(1, 15) < 1$. Un plus grand recouvrement spectral entre la mélodie et la séquence distractrice n'a donc pas affecté la ségrégation perceptive de la mélodie cible. Ce résultat suggère que la stimulation ou non des mêmes filtres auditifs ne peut rendre compte de l'organisation perceptive d'une séquence composée de sons complexes, allant à l'encontre des modèles périphériques de la formation des flux auditifs (Beauvois & Meddis, 1991). En outre, des données récentes (Vliegen et al., 1999 ; Vliegen & Oxenham, 1999) ont révélé que des séquences composées de sons complexes dont les harmoniques n'étaient pas résolues sur la membrane basilaire, se scindaient sur la base de la différence de période des sons. Le recouvrement des patterns

d'activation sur la membrane basilaire n'est donc pas un facteur déterminant dans la formation des flux auditifs.

Nature du son	Différence de hauteur moyenne (dt)	Moyennes	Ecart-types	Erreur-types
Son pur	0	0.50	0.20	0.05
	6	0.64	0.14	0.04
	12	0.81	0.15	0.04
	24	0.92	0.09	0.02
Son complexe	0	0.54	0.17	0.04
	6	0.68	0.17	0.04
	12	0.81	0.18	0.04
	24	0.90	0.11	0.03

Tableau 4.2. Proportions moyennes de réponses correctes, écart-types, et erreur-types obtenus pour les différentes conditions de séparation de hauteur entre la mélodie cible et la séquence distractrice en fonction de la nature des sons (pur et complexe).

4.4.3.2 Post-reconnaissance de mélodies intercalées et fission auditive : Fission mélodique partielle et seuil de reconnaissance de mélodies intercalées (SRM)

En raison de l'absence d'effet de la nature du son sur les performances, nous avons moyenné les proportions correctes obtenues avec les sons purs et les sons complexes pour chaque degré de séparation en hauteur. Comme on peut le constater sur la figure 4.3, les performances de post-reconnaissance de mélodies intercalées s'amélioraient à mesure que la différence de hauteur moyenne entre la mélodie cible et la séquence distractrice augmentait.

Lorsque la mélodie cible et la distractrice étaient présentées dans le même registre de hauteur, les performances étaient aléatoires. La proportion moyenne de réponses correctes était égale à 0.52 comme dans l'expérience précédente, valeur inférieure à 0.57 qui est la fréquence atypique d'une distribution binomiale pour $n = 256$ à $p < .01$ (8 observations par condition de séparation * 16 participants * 2 types de son). Les performances progressaient encore pour une séparation de plus d'une octave. Et pour un degré de séparation de 24 dt, la proportion moyenne de réponses correctes (0.91) se rapprochait de celle obtenue dans la condition contrôle (0.98) mais demeurait inférieure.

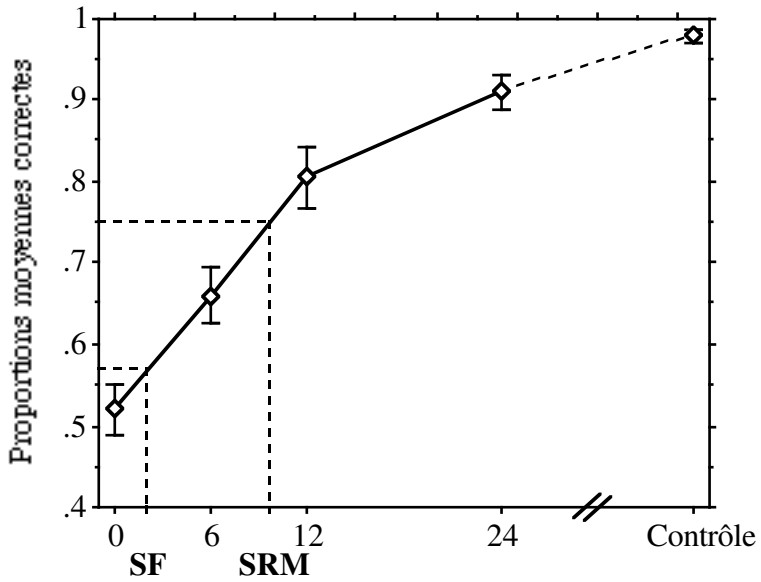


Figure 4.3. Proportions moyennes de réponses correctes obtenues par 16 auditeurs dans les différentes conditions de séparation en hauteur (0, 6, 12 ou 24 dt) et dans la condition contrôle. Le seuil de fission d'une séquence cyclique (SF) (Miller & Heise, 1950 ; van Noorden, 1975) et le seuil de reconnaissance de mélodies intercalées (SRM) sont représentés sur le graphique. Les barres verticales représentent ± 1 erreur-type.

Une analyse de variance évaluant l'effet des différentes conditions de séparation en hauteur incluant la condition contrôle (les auditeurs ayant passé cette condition avec des sons purs ou complexes) sur les performances de post-reconnaissance (plan d'analyse S*H5), a révélé que l'effet des conditions était globalement significatif, $F(4, 60) = 53.14$, $p < .0001$. Les différences de performances observées entre les conditions de séparation 0 et 6 dt, $F(1, 60) = 8.22$, $p < .01$, 6 et 12 dt, $F(1, 60) = 15.03$, $p < .001$ ainsi que 12 et 24 dt, $F(1, 60) = 9.11$, $p < .01$, étaient significatives. Les performances obtenues dans la condition contrôle étaient également significativement différentes de celles obtenues dans la condition de séparation 24 dt, $F(1, 60) = 9.45$, $p < .01$.

Les performances ont également été analysées selon la théorie de la détection du signal (Green & Swets, 1974) appliquée au paradigme même/différent (Macmillan & Creelman, 1991). Les proportions de détections correctes et de fausses alarmes ont été calculées sur l'ensemble des participants en raison du faible nombre d'observations par auditeur et un d' de groupe a été évalué pour chaque condition (tableau 4.3). Ne pouvant estimer le d' et ß de chaque individu, il n'est pas possible de savoir si ce d' de groupe est sous-estimé ou non (Macmillan & Kaplan, 1985). Cependant, les valeurs obtenues sont comparables à la moyenne des d' calculés individuellement dans la première expérience (conditions 0 dt et contrôle) à

l'exception de la condition dans laquelle la distractrice était distante de 24 dt en moyenne où le d' de groupe s'avère inférieur.

Différence de hauteur moyenne (dt)	d' groupe
0	0.6
6	1.9
12	3.1
24	4.5
Condition contrôle	6.3

Tableau 4.3. Sensibilité de l'ensemble des auditeurs pour chaque degré de séparation en hauteur. La séquence distractrice est présentée dans le même registre de hauteur (0 dt) ou séparée en moyenne de 6, 12, 24 dt ou bien elle n'est pas présente (condition contrôle).

L'augmentation des performances de post-reconnaissance de mélodies intercalées avec l'accroissement de la différence de hauteur moyenne entre la mélodie cible et la séquence distractrice montre que cette tâche permet de mesurer des processus primaires d'organisation sensibles aux caractéristiques du signal. L'augmentation de la différence de hauteur moyenne entre la mélodie cible et la séquence distractrice induit la fission perceptive des sons de la mélodie permettant ainsi à l'auditeur d'accéder aux propriétés de la mélodie cible et de la reconnaître. Les mélodies ont été reconnues dans 75 % des cas lorsque les sons distracteurs étaient éloignés en moyenne de 10 dt, correspondant au seuil de reconnaissance de mélodies intercalées (SRM) (figure 4.3). Le SRM est défini comme étant la valeur sur une dimension physique (ici la différence de hauteur moyenne) au-delà de laquelle l'auditeur peut reconnaître les mélodies dans 75 % des cas. Le niveau de performances à partir duquel on définit ce seuil est fixé à 0.75, valeur intermédiaire dans la gamme de variation de performances allant de 0.50 à 1. Un SRM de 10 dt a également été trouvé par Dowling (1973, expérience 1).

Cette fission mélodique présente cependant certaines particularités qui la distinguent de la fission d'une séquence dans laquelle deux sons aigu et grave alternent. En effet, une mélodie est par définition composée de plus de deux fréquences différentes donc de plus d'un intervalle (dans notre étude le nombre d'intervalles différents peut aller jusqu'à cinq). Par conséquent, une même différence de fréquence (il s'agit dans le cas d'une mélodie d'une différence de fréquence moyenne) peut induire la fission des deux sons composant une séquence alternante et ne provoquer qu'une fission partielle de la mélodie, du fait de l'existence d'intervalles de fréquence plus petits que la différence de fréquence moyenne. Ainsi, une différence de fréquence de 2-3 dt entre deux sons composant une séquence cyclique conduit à la formation de deux flux, l'un composé de la répétition du premier son, l'autre de la répétition du second son (van Noorden, 1975). Cette même différence moyenne dans le cas de mélodies conduit à la fission perceptive des sons formant un intervalle supérieur au égal à cette valeur alors que les

sons formant un intervalle plus petit demeurent liés perceptivement. Ainsi on peut constater que pour une différence de fréquence moyenne de 2-3 demi-tons, correspondant au seuil de fission d'une séquence dans laquelle deux sons purs de fréquences différentes alternent (Miller & Heise, 1950 ; van Noorden, 1975), les auditeurs ne répondaient plus au hasard (figure 4.3). Dans cette expérience et pour cette distribution, on peut effectivement considérer au seuil .01 que 0.57 correspond à une proportion moyenne de réponses correctes supérieure au hasard. Ce résultat suggère que les auditeurs disposent d'indices qui leur permettent d'effectuer la tâche, laissant penser que la mélodie s'est scindée en partie. Le seuil de fission semble ainsi correspondre à une fission mélodique partielle.

Les performances continuent de progresser au-delà du seuil de reconnaissance de mélodies intercalées et au-delà même d'une différence moyenne de 12 dt, valeur correspondant à l'ambitus maximal des mélodies élaborées (annexe A2). Selon Dowling (1973), la reconnaissance d'une mélodie intercalée pourrait dépendre du degré de recouvrement entre la gamme occupée par la mélodie et celle occupée par une autre mélodie ou des sons distracteurs, c'est-à-dire de la séparation en hauteur entre la note la plus haute d'une des mélodies et la note la plus basse de l'autre. Nous avons donc examiné l'ambitus des séquences distractrices et nous avons constaté que la gamme occupée par les séquences distractrices pouvait s'étendre jusqu'à 15 dt. Par conséquent, on peut considérer que dans cette expérience la mélodie et la séquence distractrice ne se recouvrent plus au-delà d'une différence moyenne de 15 dt, ce qui expliquerait que les performances progressent encore pour une différence moyenne supérieure à 12 dt.

Malgré cette progression, les performances obtenues pour un degré de séparation de 24 dt demeurent inférieures à celles obtenues dans la condition contrôle, comme dans l'expérience précédente. Ce résultat suggère que la présence de la séquence distractrice interfère dans la reconnaissance de la mélodie cible. Dowling (1973, expérience 2) avait également constaté que les performances de reconnaissance immédiate de mélodies non familières intercalées demeuraient inférieures à celles obtenues en l'absence de sons distracteurs (Dowling & Fujitani, 1971). Cette dégradation des performances de reconnaissance d'une mélodie en présence d'une autre séquence présentée simultanément pourrait être due à un effet de distractibilité attentionnelle. L'auditeur manifesterait des difficultés à focaliser son attention sur le flux cible, son attention alternerait donc entre la mélodie et la distractrice de sorte que la détection d'un changement dans la mélodie cible puisse lui échapper. Mais elle pourrait également s'expliquer par l'interférence en mémoire causée par la présence de la séquence distractrice (Deutsch, 1970). Cette interférence attentionnelle et/ou mnésique suggère que cette tâche de post-reconnaissance de mélodies intercalées, mesurant les processus primaires impliqués dans la formation des flux auditifs, fait également probablement intervenir d'autres processus. En effet, pour accomplir cette tâche l'auditeur doit non seulement organiser la séquence composite en deux flux, mais également focaliser son attention sur la mélodie cible

c'est-à-dire sélectionner un des deux flux en présence, accéder aux propriétés de ce flux (notes, contour mélodique, rythme...) et stocker ces informations en mémoire de travail.

4.4.3.3 Effet de l'expertise musicale

Les performances des musiciens étaient globalement supérieures à celles des non musiciens. Une analyse de variance évaluant l'effet de l'expertise musicale sur les performances obtenues pour les quatre conditions de séparation en hauteur (plan d'analyse S<M2>*H4), a révélé que la différence entre musiciens et non musiciens était statistiquement significative, $F(1, 14) = 7.1$, $p < .05$. La supériorité des performances des musiciens dans cette tâche ne semblait pas dépendre du degré de séparation en hauteur de la mélodie cible et de la séquence distractrice, l'interaction entre la formation musicale et la séparation en hauteur n'était pas significative, $F(3, 42) < 1$. Leurs performances étaient par ailleurs comparables dans la tâche contrôle. La proportion moyenne de réponses correctes était de 0.97 pour les musiciens et de 0.99 pour les non musiciens. L'équivalence de leurs performances dans cette condition est peut-être due à un effet plafond chez les musiciens.

L'effet de l'expertise musicale suggère que les performances de post-reconnaissance de mélodies intercalées varient selon les connaissances préalables des auditeurs. Nous avons voulu savoir si elles étaient également sensibles à l'apprentissage. Les auditeurs ayant passé deux tests successifs testant les deux types de son pur et complexe dans un ordre contrebalancé, nous avons donc examiné si les performances s'amélioraient lors du deuxième test quelle que soit la nature du son. Les proportions moyennes de réponses correctes avaient globalement tendance à augmenter dans le deuxième test (0.74) par rapport au premier (0.71). Une analyse de variance testant l'effet de l'ordre du test sur les performances en fonction du degré de séparation en hauteur et de la formation musicale des auditeurs (plan d'analyse S‹M2›*B2*H4), a révélé que cet effet était juste au-dessous du seuil de significativité, $F(1, 14) = 4.54$, $p = .051$. Cette tendance ne s'observait en fait que chez les musiciens mais l'interaction entre l'effet de l'ordre du test et la formation musicale n'était pas non plus significative, $F(1, 14) = 3.5$, $p = .08$. La supériorité des résultats des auditeurs ayant bénéficié d'une formation musicale et la tendance de leurs performances à augmenter lors du deuxième test par rapport au premier malgré le faible nombre d'essais de chaque test (32) témoignent de l'effet de l'entraînement dans cette tâche. Dowling (1973, expérience 2) a également constaté que les performances de reconnaissance immédiate de mélodies non familières intercalées avaient tendance à augmenter dans une deuxième session expérimentale.

4.5 Expérience 3 : Post-reconnaissance de mélodies intercalées sur la base d'une dissemblance de timbre

4.5.1 Introduction

Plusieurs études ont démontré le rôle du timbre dans la formation des flux auditifs. Des différences sur les dimensions spectrale, temporelle et spectro-temporelle de cet attribut multidimensionnel (Grey, 1977 ; Krumhansl, 1989 ; McAdams et al., 1995) semblent promouvoir la fission perceptive d'une séquence cyclique (Bregman et al., 1990 ; McAdams & Bregman, 1979 ; Singh, 1987 ; Singh & Bregman, 1997) et favoriser la fission mélodique (Gregory, 1994 ; Hartmann & Johnson, 1991 ; Iverson, 1995, expérience 2 ; Wessel, 1979). Iverson (1995, expérience 1) a montré que des jugements estimant la dissemblance de deux timbres constituaient de bons indicateurs de la métrique développée par le système auditif pour séparer perceptivement les sons. Il a présenté une séquence cyclique dans laquelle alternaient deux sons de même hauteur mais de timbres différents, choisis parmi un ensemble de 16 timbres d'instruments de musique réels. Et il a demandé aux auditeurs de juger à l'aide d'un curseur s'ils percevaient plutôt un flux composé de l'alternance de deux timbres différents ou bien plutôt deux flux monotimbraux. Par une technique d'analyse multidimensionnelle (Shepard, 1962a, b), ces jugements estimant le degré de fission ont été transformés en distance dans un espace géométrique euclidien. Un espace bidimensionnel dont les dimensions étaient respectivement corrélées à l'attaque et à la brillance des sons a été trouvé. Cet espace dans lequel se trouvaient les différents instruments plus ou moins éloignés selon leur degré de "séparabilité", s'apparentait à celui obtenu à partir de jugements de dissemblance sur le même ensemble de timbres (Iverson & Krumhansl, 1993). Les jugements de fission perceptive étaient corrélés aux jugements de dissemblance ($r = .81$, $df = 118$, $p < .001$), suggérant que les jugements de dissemblance de deux timbres constituaient de bons prédicteurs de leur degré de fission.

Nous proposons dans cette expérience de tester l'effet du degré de dissemblance des timbres de la mélodie cible et de la séquence distractrice sur les performances de post-reconnaissance de mélodies intercalées. Si ces jugements de dissemblance constituent de bons prédicteurs du degré de fission d'une séquence, les performances de post-reconnaissance de mélodies intercalées devraient augmenter à mesure que ce degré de dissemblance timbrale s'accroît. L'expérience précédente a suggéré que la reconnaissance d'une mélodie intercalée pourrait également dépendre de l'aptitude de l'auditeur à focaliser son attention sur le flux cible, nous avons donc également examiné si l'effet de la dissemblance de timbre sur la post-reconnaissance de mélodies intercalées était symétrique. Nous avons pour cela fait varier non seulement le degré de dissemblance des timbres de la mélodie cible et de la séquence distractrice, mais aussi le timbre de la mélodie cible afin de déterminer si les performances

dépendaient uniquement de la distance perceptive séparant les timbres des sons ou bien également du timbre spécifique avec lequel était jouée la mélodie recherchée.

4.5.2 Méthode

4.5.2.1 Participants

Trente personnes ont participé à l'expérience. Les résultats de deux participants ont été écartés, l'un ayant répondu au hasard en présence de la séquence distractrice (38 % de réponses correctes pour un degré de séparation en timbre maximal), l'autre ayant échoué dans la condition contrôle (44 % de réponses correctes), soit un taux d'échec de 7 %. Ce taux d'échec est inférieur à ceux observés dans les expériences précédentes, 12 % dans la première expérience et 19 % dans la seconde, pour une tâche de post-reconnaissance de mélodies intercalées sur la base d'une différence de hauteur. Les résultats de vingt-huit auditeurs ont donc été intégrés à l'analyse, neuf femmes et dix-neuf hommes d'un âge moyen de 30 ans [19 ; 51] dont dix-sept avaient reçu une formation musicale.

4.5.2.2 Stimuli

Nous avons utilisé des sons de synthèse imitant le timbre d'instruments de musique (Wessel, Bristow & Settel, 1987). Il s'agissait du vibraphone (VBS), de la guitare (GTR), du trombone (TBN) et du basson (BSN). Ces sons ont été choisis parmi les 18 sons de l'espace de timbre révélé par l'étude de McAdams et al. (1995), sons précédemment employés par Krumhansl (1989) (figure 4.4).

La mélodie cible et la séquence distractrice étaient présentées dans le même registre de hauteur (autour du mib3). La mélodie cible était jouée par le même instrument durant toute la durée de la condition expérimentale, seul le timbre de la séquence distractrice variait. Les différents timbres, le vibraphone (VBS), la guitare (GTR), le trombone (TBN) et le basson (BSN), variaient sur les trois dimensions perceptives : 1) la dimension liée aux caractéristiques spectrales du son et plus précisément au centre de gravité du spectre (CGS) correspondant à la moyenne des fréquences pondérées par leurs amplitudes relatives (dimension associée au caractère "brillant" du son) ; 2) la dimension liée à des caractéristiques temporelles corrélée au logarithme du temps de montée de l'enveloppe d'amplitude (Krimphoff, McAdams & Winsberg, 1994), correspondant perceptivement à l'attaque du son ; et 3) la dimension liée à des caractéristiques spectro-temporelles, associée selon les auteurs au flux spectral c'est-à-dire aux variations temporelles de l'enveloppe spectrale (McAdams et al., 1995) ou à la structure fine du spectre, c'est-à-dire à l'irrégularité spectrale qui est une mesure de la déviation des composantes spectrales par rapport à l'enveloppe moyenne (Krimphoff et al., 1994 ; Krumhansl, 1989).

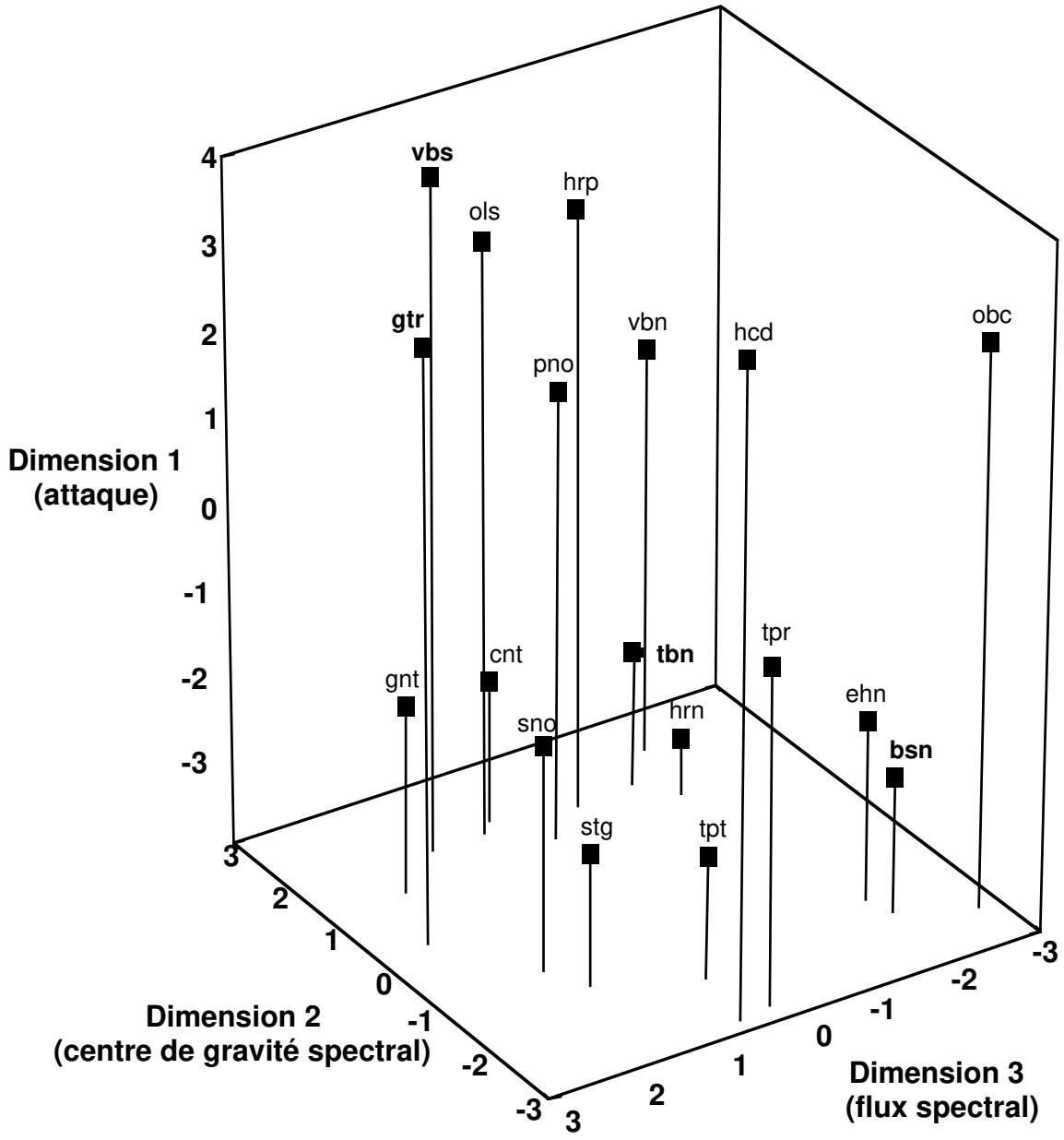


Figure 4.4. Espace de timbres à trois dimensions. Les corrélats acoustiques des dimensions perceptives sont indiqués entre parenthèses. La dimension 1 correspond à l'attaque (au logarithme du temps de montée de l'enveloppe d'amplitude), la dimension 2 à la "brillance" (au centre de gravité spectral), et la dimension 3 au flux spectral (à la variation de l'enveloppe spectrale au cours du temps). Les timbres utilisés dans cette expérience (VBS, GTR, TBN et BSN) sont notés en gras. [D'après McAdams et al., 1995, p. 185].

Nous avons donc considéré quatre degrés de séparation de timbre. La séquence distractrice avait le même timbre que la mélodie cible (S0) ou elle s'en éloignait de 1, 2 ou 3 "pas" dans l'espace (S1, S2, S3) sur les trois dimensions perceptives. Afin de tester la symétrie de l'effet du timbre sur la fission mélodique, nous avons élaboré quatre conditions expérimentales dans lesquelles la mélodie cible était jouée respectivement par les quatre

instruments. Ainsi les degrés de séparation S1, S2 et S3 pouvaient correspondre à différents instruments selon le timbre de la mélodie cible. Par exemple, si la mélodie cible était jouée par le vibraphone, la distractrice était également jouée par le vibraphone pour le degré de séparation S0, la guitare pour S1, le trombone pour S2 et le basson pour S3. En revanche, si la mélodie était jouée par le basson, la distractrice était jouée par le basson pour le degré de séparation S0, le trombone pour S1, la guitare pour S2 et le vibraphone pour S3.

Le nombre de paires différentes créées à partir de ces quatre timbres étant égal à 6, la combinaison des différents timbres produit 6 distances, 7 avec le degré 0. Ces distances ont été calculées selon le modèle euclidien étendu (McAdams et al., 1995),

$$d_{jj'} = \left[\sum_{r=1}^{R} \left(x_{jr} - x_{j'r} \right)^2 + s_j + s_{j'} \right]^{1/2}$$

où $d_{jj'}$ est la distance séparant le point j du point j', x_{jr} et $x_{j'r}$ sont les coordonnées des stimuli j et j' respectivement sur la dimension r et s_j et $s_{j'}$ sont les coordonnées des stimuli j et j' respectivement sur leur dimension spécifique.

Ces distances (tableau 4.4b) prennent en compte non seulement les coordonnées des points sur les axes, c'est-à-dire les dimensions continues partagées par l'ensemble des stimuli, mais aussi une coordonnée supplémentaire correspondant aux traits discrets spécifiques de certains stimuli appelés spécificités (tableau 4.4a). Puisque la relation établie entre les dissemblances jugées et les distances dans ce modèle spatial est linéaire, ces distances correspondent à une échelle d'intervalles.

Les distances entre les quatre timbres utilisés dans cette expérience, ont été obtenues à partir de jugements portant sur un ensemble plus large de timbres, les 18 sons utilisés par McAdams et ses collaborateurs (1995). Cependant, une étude menée par Donnadieu (1997) a révélé que ces jugements relatifs ne semblaient pas changer dans un contexte plus réduit. De plus, les jugements sur la base desquels sont calculées ces distances sont des jugements de dissemblance effectués sur des paires de sons et dans cette étude l'analyse réalisée par le système auditif prend en compte également deux timbres, le timbre des sons de la mélodie cible et celui des sons de la distractrice.

a)

Timbres	Dimension 1 (attaque)	Dimension 2 (centre de gravité spectral)	Dimension 3 (flux spectral ou irrégularité du spectre)	Spécificités$^{1/2}$
VBS	3.8	1.8	1.3	1.9
GTR	2.9	0.2	2.4	0
TBN	-2.4	1.7	-1.2	1.4
BSN	-2.4	-1.8	-2	1.4

b)

Paires de timbres	distance
VBS/GTR	2.9
TBN/BSN	4.1
GTR/TBN	6.7
VBS/TBN	7.1
GTR/BSN	7.3
VBS/BSN	8.2

Tableau 4.4. a) Coordonnées des différents timbres, le vibraphone (VBS), la guitare (GTR), le trombone (TBN) et le basson (BSN) sur les trois dimensions perceptives, extrait de McAdams et al. (1995). **b)** Distances calculées selon la formule de distance du modèle euclidien étendu.

La sonie et la durée subjective des sons ont été égalisées. La sonie des différents timbres a été égalisée pour la note mib3 (f0 = 311 Hz) en faisant varier la vélocité MIDI. Ce paramètre contrôle l'intensité et le spectre du son en modifiant la vitesse avec laquelle la touche est enfoncée (dynamique de jeu). La durée subjective des sons a été ajustée pour pallier aux variations subjectives dues à la présence d'enveloppes temporelles variées. Le niveau physique maximal atteint par chacun des sons a été mesuré à l'aide d'un sonomètre Bruel & Kjaer 2209 (pondération A, intégration lente). Le vibraphone avait un niveau maximum de 57.5 dBA, la guitare 53 dBA, le trombone 62 dBA et le basson 65 dBA.

4.5.2.3 Procédure

Les participants ont passé les quatre conditions expérimentales dans lesquelles la mélodie cible était jouée respectivement par les quatre instruments (VBS, GTR, TBN, BSN) puis la condition contrôle avec un des quatre timbres (les timbres ont été choisis de façon égale parmi les quatre proposés sur l'ensemble des participants) (CD plages 17 à 36, annexe D).

Chacune de ces conditions était précédée par une phase de familiarisation. Les effets de rang et d'ordre ont été contrebalancés sur l'ensemble des participants par la technique du carré balancé. Chaque condition occupait chaque position un même nombre de fois et afin de compenser l'éventuelle influence de la focalisation du timbre précédent sur le suivant, chaque condition précèdait chaque autre condition un même nombre de fois (tous les enchaînements possibles, 12 doublets possibles parmi 4 éléments, étaient représentés un même nombre de fois). Quatre ordres de passation des différentes conditions expérimentales correspondant aux différents timbres de la mélodie cible ont ainsi été présentés aux auditeurs (tableau 4.5). Ces ordres sont équireprésentés sur l'ensemble des participants.

Ordre 1	VBS	GTR	TBN	BSN
Ordre 2	TBN	VBS	BSN	GTR
Ordre 3	GTR	BSN	VBS	TBN
Ordre 4	BSN	TBN	GTR	VBS

Tableau 4.5. Ordres de présentation des quatre conditions expérimentales, la condition où la mélodie est jouée par le vibraphone (VBS), la guitare (GTR), le trombone (TBN) et le basson (BSN). Chaque ordre a été présenté à 7 auditeurs.

Chacun de ces cinq tests, les quatre conditions expérimentales et la condition contrôle, comportait 32 essais (choix aléatoire de 32 mélodies parmi les 36). Une consigne écrite était donnée aux participants (annexe C3). La passation durait 1 heure environ. Les auditeurs étaient indemnisés pour leur participation.

4.5.3 Résultats et discussion

4.5.3.1 Effet du degré de séparation en timbre sur la post-reconnaissance de mélodies intercalées et symétrie de l'effet du timbre

Les performances augmentaient avec le degré de séparation des timbres de la mélodie et de la séquence distractrice, et ne dépendaient pas du timbre de la mélodie cible. Pour un même degré de séparation entre les timbres, les performances étaient dans l'ensemble équivalentes quel que soit le timbre de la mélodie cible (courbes confondues) à l'exception de deux instruments, la guitare et le trombone (figure 4.5). En effet, les proportions moyennes de réponses correctes obtenues lorsque la mélodie cible était jouée par la guitare et la distractrice par le vibraphone (courbe points blancs au degré de séparation S1) (0.64) étaient inférieures à celles obtenues lorsqu'à l'inverse la mélodie cible était jouée par le vibraphone et la séquence distractrice par la guitare (courbe points noirs au degré de séparation S1) (0.75). Une même asymétrie s'observait entre le vibraphone et le trombone ; les performances étaient inférieures

quand la mélodie cible était jouée par le trombone (courbe carrés noirs, degré de séparation S3) (0.75) plutôt que par le vibraphone (courbe points noirs, degré de séparation S2) (0.84).

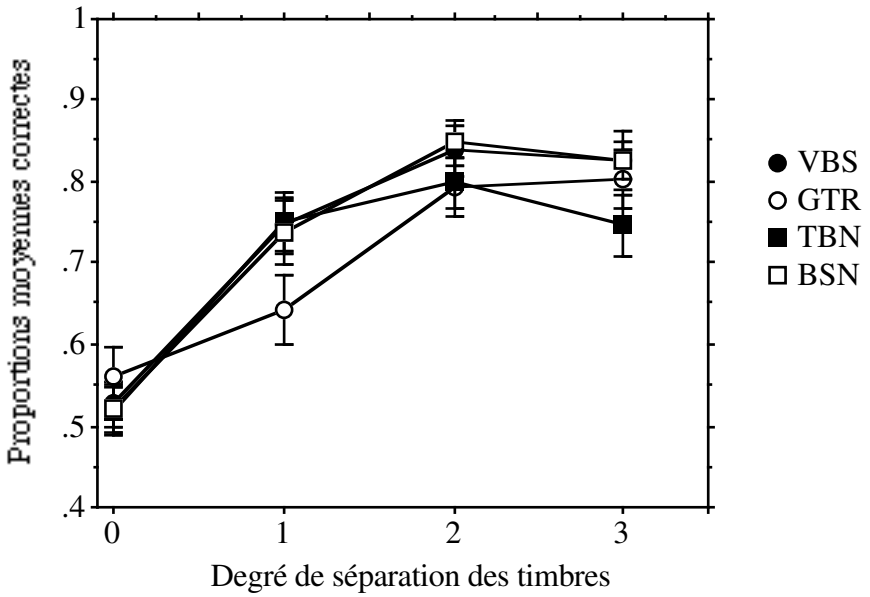


Figure 4.5. Proportions moyennes de réponses correctes obtenues par les 28 auditeurs lorsque la mélodie cible était jouée par le vibraphone (VBS), la guitare (GTR), le trombone (TBN) et le basson (BSN) et que la distractrice était jouée par le même timbre que la mélodie cible (degré de séparation 0) ou s'en éloignait de 1, 2 ou 3 "pas" dans l'espace de timbre (voir texte). Les barres verticales indiquent ± 1 erreur-type.

Une analyse de variance évaluant l'effet du timbre de la mélodie cible sur les performances en fonction du degré de séparation (plan d'analyse S*C4*T4), n'a pas révélé d'effet de ce facteur, $F(3, 81) = 1.21$, N.S. L'interaction entre l'effet du timbre de la mélodie cible et l'effet du degré de séparation en timbre sur les performances n'était pas non plus significative, $F(9, 243) = 1.79$, N.S. Cependant l'analyse planifiée des contrastes a révélé la significativité des deux asymétries. Pour une même différence de timbre, les performances étaient significativement meilleures lorsque la mélodie cible était jouée par le vibraphone (0.75) plutôt que la guitare (0.64), $F(1, 243) = 5.67$, $p < .05$, et le vibraphone (0.84) plutôt que le trombone (0.75), $F(1, 243) = 8.19$, $p < .01$. Nous ne sommes pas parvenus à expliquer ces deux asymétries locales. Une hypothèse qui peut être avancée est que le fait d'intervertir les timbres de la mélodie et de la séquence distractrice change non seulement le timbre qui est focalisé mais aussi le timbre des notes paires et impaires de la séquence composite. En effet, les notes impaires de la séquence composite correspondent toujours aux notes de la mélodie et les notes paires à celles de la distractrice. Or Dowling et ses collaborateurs (1987) ont montré dans une tâche de reconnaissance de mélodies intercalées que lorsque la mélodie était jouée par les notes impaires de la séquence composite sa reconnaissance était accrue. Cet effet combiné à des

différences de caractéristiques entre les timbres (par exemple sur la dimension liée à l'attaque) pourrait expliquer ces asymétries.

En raison de l'absence d'effet global du timbre de la mélodie cible, nous avons moyenné les proportions de reconnaissances correctes obtenues quel que soit le timbre de la mélodie cible pour chacun des degrés de séparation (S0, S1, S2 et S3). Une analyse de variance évaluant l'effet du degré de séparation des timbres sur les performances en intégrant la condition contrôle et en prenant en compte la formation musicale des auditeurs (plan d'analyse S<M2>*T5), a révélé que l'effet du degré de séparation en timbre était globalement significatif, $F(4, 104) = 65.96$, $p < .0001$. Les différences observées pour les degrés de séparation S0 et S1, $F(1, 104) = 40.54$, $p < .0001$, et S1 et S2, $F(1, 104) = 9.83$, $p < .01$ étaient significatives, mais pas S2 et S3, $F(1, 104) < 1$. Lorsque la mélodie cible et la séquence distractrice avaient le même timbre, les performances étaient aléatoires. La proportion moyenne de réponses correctes était égale à 0.53, valeur inférieure à 0.54 fréquence atypique d'une distribution binomiale pour $n = 896$ à $p < .01$ (8 observations par condition de séparation * 4 timbres de la mélodie cible * 28 auditeurs). Lorsque la différence de timbre était maximale, la proportion moyenne de reconnaissance correcte atteignait 0.80 mais demeurait inférieure à celle obtenue dans la condition contrôle (0.95). Cette différence était statistiquement significative, $F(1, 104) = 53.6$, $p < .0001$. Comme dans les expériences précédentes de post-reconnaissance de mélodies intercalées sur la base d'une différence de hauteur, la reconnaissance de la mélodie cible en présence de la séquence distractrice était altérée par rapport à la condition contrôle où la mélodie était présentée isolément.

Les auditeurs ont réalisé la tâche de façon équivalente qu'ils aient ou non bénéficié d'une éducation musicale. La différence de performances entre les musiciens et les non musiciens n'était pas significative, $F(1, 26) = 1.16$, N.S. Le faible taux d'échec et le fait qu'il n'y ait pas de différences de performances liées à l'expertise musicale, contrairement à ce qui a été trouvé dans l'expérience 2, suggèrent que cette même tâche de post-reconnaissance de mélodies intercalées est plus facile à réaliser lorsque la mélodie et la séquence distractrice sont jouées par des timbres différents plutôt que dans des registres distincts.

4.5.3.2 Le degré de dissemblance de timbre : Un bon prédicteur de la fission mélodique

Les combinaisons créées à partir de ces quatre timbres définissent sept distances dans l'espace de timbre (tableau 4.4b). Nous avons moyenné les proportions correctes obtenues pour chacune de ces distances indépendamment du timbre spécifique de la mélodie cible, afin d'établir un lien entre les performances de post-reconnaissance de mélodies intercalées et le degré de dissemblance entre le timbre de la mélodie cible et celui de la séquence distractrice (figure 4.6).

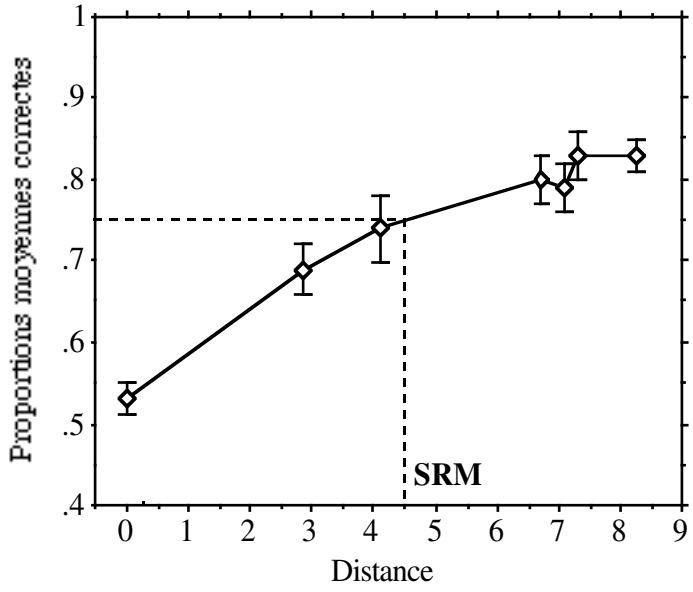


Figure 4.6. Proportions moyennes de réponses correctes obtenues pour chacun des degrés de dissemblance entre les timbres de la mélodie cible et de la distractrice représentés par des distances dans l'espace euclidien (McAdams et al., 1995). Les barres verticales indiquent ± 1 erreur-type.

Nous avons également calculé un d' de groupe (Macmillan & Kaplan, 1985) sur la base du nombre de détections correctes et de fausses alarmes obtenus par l'ensemble des auditeurs pour chacune de ces distances (tableau 4.6).

Distances entre le timbre de la mélodie et celui de la distractrice	**d' groupe**
0	0.7
2.9	2.1
4.1	2.5
6.7	3.0
7.1	3.0
7.3	3.2
8.2	3.2
Condition contrôle	5.2

Tableau 4.6. Sensibilité de l'ensemble des auditeurs pour chacune des distances entre les timbres.

Les performances de post-reconnaissance de la mélodie cible intercalée (proportions moyennes correctes et d') augmentaient avec la dissemblance des timbres de la mélodie et de la distractrice. Les d' de groupe obtenus dans cette expérience étaient corrélés (corrélation linéaire)

positivement à la distance séparant les timbres des sons successifs ($r = .97$, $p < .0001$). Pour une différence inférieure à une unité de distance dans l'espace de timbre, les performances de post-reconnaissance de mélodies intercalées sont supérieures au hasard. La distance nécessaire pour reconnaître la mélodie dans 75 % des cas (SRM) est de l'ordre de 4 unités de distance (4.5) (figure 4.6). L'échelle de sensibilité (d') permet de mettre en correspondance les performances de post-reconnaissance obtenues pour une dissemblance de timbre et celles obtenues pour une différence de hauteur moyenne (expérience 2). Par exemple, un d' d'une valeur approximativement de 3 est obtenue pour une différence de hauteur d'environ 12 dt et pour une distance dans l'espace de 7. Dans cette étude, le nombre de points est insuffisant pour pouvoir évaluer les deux fonctions et établir une équivalence de séparabilité perceptive entre la différence de hauteur moyenne et la distance des timbres.

Les résultats de cette expérience montrent que le degré de dissemblance des timbres, indice qui prend en compte la multidimensionnalité de cet attribut, constitue un bon prédicteur de la fission mélodique (Iverson, 1995). Ceci suggère que les jugements de dissemblance et l'analyse réalisée par le système auditif pour organiser perceptivement une séquence sonore pourraient reposer sur une représentation sensorielle commune.

4.6 Expérience 4 : Effet de la durée de l'intervalle de rétention sur la post-reconnaissance de mélodies intercalées

4.6.1 Introduction

Un des objectifs de cette étude est de mettre en place un paradigme qui puisse mesurer de façon objective les processus primaires mis en jeu dans la formation des flux auditifs. Les expériences précédentes ont montré que les performances de post-reconnaissance de mélodies intercalées augmentaient avec la différence de hauteur moyenne et la dissemblance de timbre entre la mélodie cible et la séquence distractrice, suggérant que cette tâche est effectivement sensible à des processus primaires d'organisation. Cependant, on peut se poser la question de savoir si les performances mesurées sont le reflet d'une représentation immédiate de la mélodie qui émerge de l'organisation perceptive de la séquence composite ou d'une reconstruction a posteriori. En effet, on peut imaginer que durant l'intervalle de temps qui sépare la séquence composite de la mélodie test, l'auditeur, engagé dans une tâche de reconnaissance de mélodies, essaie de se rejouer mentalement la première séquence afin d'extraire a posteriori les notes ou d'autres indices qui lui permettent d'accomplir cette tâche.

Cette quatrième expérience se propose donc d'examiner l'éventuel effet de la répétition mentale de la première séquence sur les performances de post-reconnaissance de mélodies intercalées. Une tâche de post-reconnaissance de mélodies intercalées sur la base d'une différence de hauteur moyenne est proposée à deux groupes d'auditeurs. Dans un premier groupe, l'intervalle de temps séparant la mélodie cible intercalée à la séquence distractrice et la

mélodie test a été réduit à 990 ms. Cet intervalle de silence ne permet plus à l'auditeur de se répéter mentalement la séquence composite ou la mélodie cible au tempo d'origine. Et dans un deuxième groupe, la durée de ce silence n'était plus que de 330 ms de telle sorte que la répétition mentale ne puisse plus s'opérer du tout. Les performances obtenues par ces deux groupes d'auditeurs sont comparées à celles recueillies auprès d'un sous-groupe de participants de l'expérience 2 dans laquelle la durée de l'intervalle de rétention était de 1870 ms.

4.6.2 Méthode

4.6.2.1 Participants

Vingt-quatre personnes ont participé à cette expérience. Il s'agissait pour la plupart d'étudiants en Psychologie de l'Université Paris V. Huit participants (tous non musiciens) ont échoué à ce test, soit un taux d'échec de 33 %. Leurs performances dans la condition contrôle étaient comprises entre 75 et 97 % de réponses correctes, mais leurs réponses étaient aléatoires en présence de la séquence distractrice même lorsqu'elle était distante de 24 dt en moyenne de la mélodie (pourcentages de réponses correctes compris entre 38 et 63 %). Le taux d'échec observé ne dépendait pas de la durée de l'intervalle de rétention choisie.

Seuls seize auditeurs ont donc été pris en compte dans l'analyse. Ils étaient répartis en deux groupes de composition homogène, un groupe soumis à un intervalle de rétention de 330 ms (âge moyen = 25 ans [22 ; 27], 7 femmes et 1 homme, autant de musiciens que de non musiciens) et un groupe pour qui la durée de l'intervalle inter-stimulus était de 990 ms (âge moyen = 25 ans [22 ; 37], 6 femmes et 2 hommes, autant de musiciens que de non musiciens). Les résultats obtenus par ces deux groupes d'auditeurs ont été comparés aux performances d'un troisième groupe composé d'un sous-ensemble de participants de l'expérience 2 (âge moyen = 31 ans [23 ; 52], 5 femmes et 3 hommes, autant de musiciens que de non musiciens). Ces auditeurs avaient passé la tâche de post-reconnaissance de mélodies intercalées sur la base d'une différence de hauteur avec un intervalle de rétention de 1870 ms. Les participants choisis pour cette analyse sont ceux qui avaient passé le premier test avec des séquences composées de sons purs (seules les performances de ce premier test sont incluses dans l'analyse), et la condition contrôle avec des sons pur ou complexe.

4.6.2.2 Stimuli

Les stimuli étaient identiques à ceux présentés dans l'expérience 2. La distractrice était présentée dans le même registre que la mélodie cible ou s'en éloignait en moyenne de 6, 12 ou 24 dt vers les fréquences basses. L'effet de trois intervalles de rétention de durée différente a été testé dans les conditions contrôle et expérimentale. L'intervalle de temps séparant la séquence composite et la mélodie test était de 330, 990 ou 1870 ms. Les séquences étaient

jouées par des sons purs. Pour la moitié des auditeurs du troisième groupe, les mélodies avaient été jouées dans la condition contrôle par des sons complexes composés de 6 harmoniques de même amplitude (expérience 2).

4.6.2.3 Procédure

Une consigne écrite était remise aux participants afin qu'ils prennent connaissance de la tâche (annexe C4). Ils passaient la condition expérimentale puis la condition contrôle qui étaient composées chacune de 32 essais et précédées par une phase de familiarisation. Un des trois intervalles de rétention leur était affecté pour toute la durée de l'expérience. La passation durait 30 minutes environ.

4.6.3 Résultats et discussion

4.6.3.1 Effet de la durée de l'intervalle de rétention sur les performances de post-reconnaissance de mélodies intercalées

La diminution de la durée de l'intervalle de temps séparant la mélodie cible intercalée avec la séquence distractrice et la mélodie test n'a pas affecté les performances de post-reconnaissance de la mélodie cible (figure 4.7).

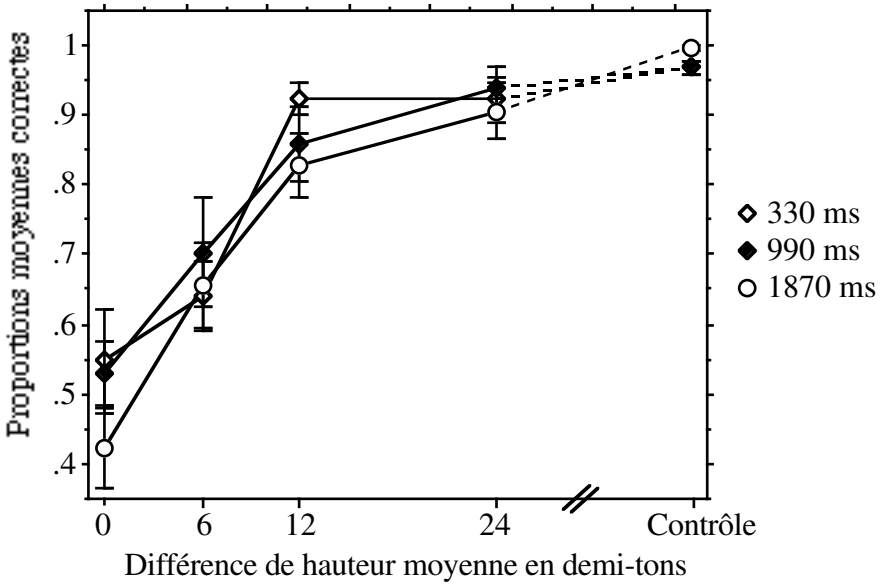


Figure 4.7. Proportions moyennes de réponses correctes pour les différentes conditions de séparation en hauteur moyenne entre la mélodie cible et la séquence distractrice (0, 6, 12 et 24 dt) et pour la condition contrôle (absence de distractrice) en fonction de la durée de l'intervalle séparant les deux séquences présentées (330, 990, et 1870 ms). Les barres verticales indiquent ± 1 erreur-type.

Une analyse de variance (plan d'analyse S8<IR3>*H5) évaluant l'effet de la durée de l'intervalle de rétention sur les proportions de reconnaissances correctes en fonction des différentes conditions de séparation en hauteur incluant la condition contrôle, n'a pas révélé d'effet de ce facteur, $F(2, 21) < 1$. Les performances augmentaient de façon significative avec l'accroissement de la différence de hauteur moyenne séparant la mélodie cible de la séquence distractrice, $F(4, 84) = 47.66$, $p < .0001$, mais elles augmentaient de la même façon quelle que soit la durée de l'intervalle inter-séquence, $F(8, 84) < 1$.

Le fait de pouvoir ou non se répéter la première séquence ne modifie donc pas les performances de post-reconnaissance de mélodies intercalées. Ce résultat suggère que ce paradigme permet de mesurer la représentation immédiate de la mélodie qui émerge de l'organisation perceptive de la séquence composite.

4.6.3.2 Effet de la présence de la séquence distractrice sur les performances de reconnaissance de la mélodie cible en fonction de la durée de l'intervalle de rétention

Nous avons constaté dans les expériences précédentes que les performances de reconnaissance de la mélodie cible étaient diminuées lorsque la séquence distractrice était présente, même si elle était éloignée en hauteur (expériences 1 et 2) ou distante dans l'espace de timbres (expérience 3). Une des hypothèses que nous avons émises pour expliquer cette différence de performances, était qu'elle pouvait provenir d'un phénomène d'interférence en mémoire (Deutsch, 1970). En effet, l'encodage simultané de la séquence distractrice et de la mélodie pourrait dégrader la trace mnésique de la mélodie cible par rapport à celle que les auditeurs se forment en l'absence de distractrice. Si cette trace mnésique est dégradée, elle est probablement moins robuste et pourrait par conséquent décliner plus rapidement que la trace non dégradée. Dans ce cas, la différence entre les performances de reconnaissance en présence et en l'absence de distractrice devrait être d'autant plus importante que l'intervalle inter-séquence s'allonge. Dans cette expérience, on constate effectivement que la différence de performances entre la condition de séparation maximale en hauteur (24 dt) et la condition contrôle est de 0.03 et 0.05 lorsque la durée du silence est respectivement de 990 et 330 ms alors qu'elle est de 0.09 lorsque la séquence composite et la mélodie test sont séparées par un silence de 1870 ms. Cependant une analyse de variance évaluant l'effet de la durée de l'intervalle inter-stimulus dans ces deux conditions 24 dt et contrôle (plan d'analyse S8<IR3>*H2), n'a pas révélé d'interaction significative entre ces deux facteurs, $F(2, 21) = 1{,}37$, N.S.

4.7 Discussion générale

4.7.1 Résumé des principaux résultats

Afin de mesurer de façon objective les processus primaires impliqués dans la formation des flux auditifs, nous avons proposé un nouveau paradigme consistant en une tâche de post-reconnaissance de mélodies intercalées dans laquelle la mélodie à reconnaître est présentée après la séquence à organiser.

Les trois premières expériences ont révélé que les performances de post-reconnaissance de la mélodie cible variaient en fonction de la différence de hauteur moyenne (expériences 1 et 2) et de la dissemblance de timbre (expérience 3) entre la mélodie cible et la séquence distractrice. Lorsque la mélodie et la séquence distractrice étaient présentées dans le même registre et jouées avec le même timbre, c'est-à-dire lorsque la séquence composite était organisée en un seul flux, les performances étaient aléatoires. Lorsque la différence de hauteur ou de timbre augmentait, les proportions moyennes de reconnaissance correcte ainsi que la sensibilité discriminative des auditeurs (d') s'amélioraient et atteignaient une valeur proche de celle obtenue dans une tâche de reconnaissance de mélodies présentées sans séquence distractrice.

Les mélodies étaient reconnues dans 75 % des cas lorsque les sons distracteurs étaient distants en moyenne de 10 dt correspondant au seuil de reconnaissance de mélodies intercalées (SRM). Cette valeur s'avère supérieure au seuil de fission d'une séquence cyclique dans laquelle alternent deux sons purs de fréquences différentes. Néanmoins à ce seuil évalué à 2-3 dt (Miller & Heise, 1950 ; van Noorden, 1975), les performances des auditeurs étaient juste supérieures au hasard suggérant que les mélodies s'étaient scindées partiellement. Ainsi on a pu montrer qu'au seuil de fission se produisait une fission mélodique partielle permettant d'établir un lien entre les résultats obtenus par des méthodes directes (Bregman, 1990 ; Miller & Heise, 1950 ; van Noorden, 1975) et ceux obtenus par des méthodes indirectes comme la reconnaissance de mélodies intercalées (Dowling, 1973 ; Dowling et al., 1987 ; Hartmann & Johnson, 1991 ; Iverson, 1995, expérience 2 ; Vliegen & Oxenham, 1999, expérience 2). Nous avons également pu établir qu'une dissemblance de timbre inférieure à une unité de distance dans l'espace de McAdams et al. (1995) conduisait à une fission mélodique partielle. L'effet de la dissemblance de timbre sur les performances de reconnaissance d'une mélodie intercalée confirme le fait que les jugements de dissemblance constituent de bons prédicteurs du degré de fission d'une séquence (Iverson, 1995, expérience 1).

La quatrième expérience n'a pas montré d'effet de la diminution de l'intervalle de rétention sur les performances de post-reconnaissance de la mélodie cible. Ce dernier résultat montre que le fait de pouvoir se répéter mentalement la première séquence n'affecte pas les performances dans cette tâche et suggère ainsi que ce paradigme permet de mesurer une

représentation immédiate de la mélodie qui émerge de l'organisation perceptive de la séquence composite.

4.7.2 Post-reconnaissance de mélodies intercalées : Une mesure des processus primaires d'organisation mais pas seulement...

Même si ce paradigme de post-reconnaissance permet de mesurer des processus primaires impliqués dans la formation des flux auditifs, il est toutefois probable que des connaissances implicites et explicites interviennent dans la réalisation de cette tâche. En outre, le fait d'accéder indirectement à l'organisation perceptive de la séquence composite fait intervenir d'autres composantes cognitives, attentionnelles et/ou mnésiques, n'ayant pas de lien direct avec les processus que l'on souhaite mesurer. L'implication d'autres processus que ceux mis en jeu dans la formation des flux auditifs invite à être prudent dans les inférences que l'on peut faire sur le nombre de flux perçus à partir des performances de reconnaissance d'une mélodie intercalée. Ces différents points sont discutés ci-dessous.

4.7.2.1 Intervention des connaissances

Ce paradigme de post-reconnaissance de mélodies intercalées a été conçu pour minimiser l'intervention des connaissances dans l'organisation perceptive de la séquence composite et sur les performances de reconnaissance de la mélodie cible. Les mélodies présentées ne sont pas familières aux auditeurs. Les tests sont composés de peu d'essais, limitant ainsi les facteurs d'apprentissage. Les mélodies sont présentées une à deux fois maximum (à l'exception de l'expérience 1). Et la mélodie à reconnaître est présentée après le mélange. Cependant certaines connaissances, élaborées pendant l'expérience ou liées à l'expertise des auditeurs, interviennent probablement dans cette tâche. En effet, certaines régularités propres à cette situation expérimentale sont apprises implicitement au cours de la passation, comme par exemple le fait que le registre de la mélodie cible et son timbre soient constants (seules les caractéristiques de la séquence distractrice fluctuent) et que dans les expériences où la différence de hauteur varie la mélodie soit toujours présentée dans le registre aigu et la séquence distractrice dans le registre plus grave. Dowling et al. (1987) ont en effet montré que des auditeurs pouvaient développer des attentes fréquentielles et temporelles qui amélioraient la reconnaissance d'une mélodie. Cette fenêtre d'attentes était d'autant plus étroite (circonscrite à une gamme fréquentielle et temporelle spécifique) que la mélodie était intercalée à des sons distracteurs. Ces connaissances que l'auditeur élabore au cours de l'expérience pourraient expliquer la tendance des performances à augmenter lors du deuxième test de l'expérience 2 malgré le faible nombre d'essais dans chaque test (32), tendance également observée par Dowling (1973, expérience 2) dans une tâche de reconnaissance de mélodies intercalées non familières. La supériorité des résultats des musiciens (expérience 2) témoigne également de l'effet de l'apprentissage dans cette tâche.

4.7.2.2 Implication de processus attentionnels et/ou mnésiques

La différence de performances observée entre la condition de séparation maximale (en hauteur ou en timbre) et la condition contrôle (expériences 1, 2 et 3), également constatée par Dowling et ses collaborateurs (Dowling, 1973 ; Dowling & Fujitani, 1971), suggère que les performances de post-reconnaissance de mélodies intercalées ne dépendent pas uniquement de l'organisation perceptive de la première séquence en deux flux. En effet, pour accomplir cette tâche, l'auditeur doit non seulement organiser la séquence composite en deux flux distincts mais également focaliser son attention sur le flux cible, accéder aux propriétés de la mélodie (le contour, les notes, le rythme...) et stocker ces informations en mémoire de travail. La dégradation des performances de reconnaissance de la mélodie cible en présence de la séquence distractrice présentée simultanément peut donc provenir : 1) d'une distractibilité attentionnelle, l'auditeur manifestant des difficultés à focaliser son attention de façon permanente sur un flux cible de sorte que la détection d'un changement dans la mélodie cible puisse lui échapper, et/ou 2) d'une interférence en mémoire causée par la présence de la séquence distractrice (Deutsch, 1970), l'encodage simultané de la séquence distractrice et de la mélodie pouvant en effet dégrader la représentation de la mélodie cible. Ce phénomène d'interférence révèle que ce paradigme fait intervenir en plus des processus primaires d'organisation d'autres composantes cognitives.

4.7.3 Performances de post-reconnaissance de mélodies intercalées et nombre de flux perçus, pas d'équivalence logique

La première expérience nous a permis d'établir qu'à l'exception des auditeurs ayant l'oreille absolue, lorsque la séquence composite était organisée en un seul flux les performances de post-reconnaissance de la mélodie cible intercalée étaient aléatoires. On peut donc affirmer que la perception d'un flux implique des réponses au hasard à cette tâche. En revanche, des performances aléatoires n'impliquent pas en retour que la séquence composite soit organisée en un seul flux. C'est une éventualité, mais nous ne pouvons pas l'inférer avec certitude en raison de l'intervention d'autres processus dans cette tâche. En effet, les auditeurs peuvent avoir séparé perceptivement la mélodie et la distractrice, mais manifester des troubles de focalisation attentionnelle ou bien encore des problèmes d'interférence en mémoire de travail qui altèrent leurs performances. De la même façon, constater que les performances sont aléatoires lorsque la séquence composite est organisée en un flux, nous permet d'en déduire que des performances supérieures au hasard impliquent que la mélodie ait été séparée au moins partiellement du mélange (à l'exception toujours des auditeurs ayant l'oreille absolue). En revanche, comme pour la proposition précédente, l'inverse n'est pas nécessairement vrai. L'organisation perceptive de la séquence composite en deux flux distincts ne garantit pas que les auditeurs

obtiennent des performances supérieures au hasard pour des raisons identiques à celles mentionnées ci-dessus.

Une remarque similaire a été formulée par Bregman (1978b, p. 394) à propos des jugements portant sur l'ordre des sons AB présentés dans une séquence cyclique ABXY dans laquelle la différence de hauteur entre les différents sons variait. A. S. Bregman avait constaté que lorsque les sons A et B appartenaient à des flux distincts, les auditeurs ne parvenaient pas à déterminer dans quel ordre les sons avaient été présentés. En revanche, des auditeurs pouvaient répondre au hasard mais avoir tout de même groupé les deux sons A et B dans un même flux en raison de la confusion entre les sons A et X et B et Y qui étaient proches en fréquence.

Ce constat signifie qu'il n'y a pas d'équivalence au sens logique entre le nombre de flux perçus et le fait que les performances de reconnaissance d'un pattern cible soient aléatoires ou non. La reconnaissance d'un pattern cible peut dépendre de l'organisation perceptive d'une séquence composite, comme c'est le cas de cette tâche de post-reconnaissance de mélodies intercalées. Néanmoins, l'analyse perceptive de cette séquence ne garantit pas pour autant la reconnaissance de ce pattern cible. Il convient donc d'être prudent dans les inférences que l'on peut faire sur l'organisation perceptive de la séquence composite à partir des performances de reconnaissance d'une mélodie intercalée, et sur l'explication que l'on peut donner aux échecs dans cette tâche. Des différences de performances induites par différents degrés d'expertise dans cette tâche (différence entre musiciens et non musiciens dans l'expérience 2) ne sont pas nécessairement le reflet de différences d'organisation. De même que l'échec à cette tâche peut provenir de déficits autres que des problèmes de formation des flux comme des déficits attentionnels et/ou mnésiques.

4.8 Conclusion

Les précédentes recherches ont permis d'établir les avantages et les inconvénients de l'utilisation des méthodes directes pour accéder à l'analyse primaire de la scène auditive (Bregman, 1990). L'apport de cette étude est de mettre en évidence les avantages et les inconvénients de l'utilisation de méthodes indirectes, comme la reconnaissance de mélodies intercalées, et ainsi de montrer la complémentarité de ces deux moyens d'investigation. Si les méthodes indirectes permettent de mesurer objectivement ces processus de formation des flux auditifs, elles font néanmoins intervenir d'autres composantes cognitives, invitant à être prudent dans les inférences que l'on peut faire sur l'organisation perceptive d'une séquence à partir des performances obtenues. Cet inconvénient peut être surmonté en recoupant les résultats obtenus avec des réponses recueillies par des méthodes directes. Et nous avons vu qu'il était possible d'établir un lien entre les résultats obtenus par ces deux types de mesure grâce aux notions de fission mélodique partielle et de seuil de reconnaissance de mélodies intercalées. Pour explorer de façon exhaustive ces mécanismes qui sous-tendent l'analyse de la scène auditive, il s'avère donc nécessaire de faire converger ces méthodes.

Chapitre 5

Reconnaissance de mélodies intercalées avec ou sans présentation préalable de la mélodie : Mise en évidence des processus d'analyse de scène auditive guidée par les connaissances

Selon A. S. Bregman, l'Analyse de la Scène Auditive serait sous-tendue par deux processus distincts : l'analyse primaire et l'analyse guidée par les schémas. Dans le chapitre précédent, une série d'expériences a été menée afin de mesurer les processus primaires impliqués dans la formation des flux auditifs. Dans ce chapitre, les résultats de trois expériences étudiant la contribution des connaissances dans l'analyse perceptive d'une séquence sonore, sont rapportés. Le paradigme de post-reconnaissance de mélodies intercalées est contrasté avec une tâche classique de reconnaissance de mélodies intercalées proposée par W. J. Dowling, dans laquelle la mélodie est présentée avant la séquence à organiser. La nature des schémas impliqués ainsi que le niveau de traitement auquel ces processus descendants interviennent, sont examinés.

5.1 Introduction

Quelle est la nature des processus impliqués dans la formation des flux auditifs ? Nous avons vu, dans le deuxième chapitre de ce document, que Bregman (1990, 1994) a émis l'hypothèse que l'analyse de la scène auditive était sous-tendue par deux processus psychologiques distincts : l'analyse primaire et l'analyse guidée par les connaissances.

L'analyse primaire de la scène auditive serait une méthode générale et automatique de partition de l'entrée sensorielle en événements sonores distincts, effectuée sur la base de régularités exploitées par le système auditif, comme par exemple le fait que des événements provenant d'une même source tendent à se modifier progressivement. L'analyse guidée par les connaissances, en revanche, serait un processus de sélection de l'information sur la base de l'activation de connaissances spécifiques sur les sons et les séquences signifiantes pour l'individu (musique, parole, bruits de machine, cris d'animaux, etc.) ; l'activation de ces connaissances pouvant être automatique ou volontaire.

Nous disposons actuellement de peu de données expérimentales sur le rôle des processus d'analyse guidée par les connaissances dans la construction de la scène auditive (Bregman, 1990, chap. 8). Van Noorden (1975) a montré que la fission perceptive d'une séquence cyclique dans laquelle deux sons aigu et grave alternaient, dépendait dans une certaine mesure de ce que l'auditeur cherchait à entendre. Dowling (1973, expérience 3) a révélé que des auditeurs, à qui on présentait des mélodies familières intercalées à des sons distracteurs présentés dans le même registre de fréquence, parvenaient à les reconnaître si le titre de la mélodie leur était donné préalablement.

Plusieurs questions théoriques se posent concernant la contribution des connaissances dans l'analyse de la scène auditive : De quelle nature est cette analyse de scène guidée par les connaissances, s'agit-il d'un mécanisme de partition sensorielle construisant une entité perceptive qui corresponde au schéma stocké en mémoire ou comme le postule A. S. Bregman d'un mécanisme de sélection d'information appariant les informations sensorielles aux connaissances activées ? L'analyse guidée par les connaissances et l'analyse primaire de la scène auditive interviennent-elles ainsi de façon différente dans la construction de la scène auditive (Bregman, 1990, 1994) ? Sont-elles indépendantes ou interagissent-elles ?

Les trois expériences relatées dans ce chapitre ont été menées dans le but d'apporter des éléments de réponse à ces questions concernant le rôle et la nature des processus d'analyse de scène auditive guidée par les connaissances. Nous avons étudié la contribution des connaissances dans la formation des flux auditifs en comparant les performances obtenues dans deux tâches expérimentales : la tâche de post-reconnaissance de mélodies intercalées élaborée précédemment dans laquelle la mélodie à comparer est présentée après une séquence à organiser, et la tâche classique de reconnaissance immédiate de mélodies non familières intercalées proposée par Dowling (1973, expérience 2) dans laquelle l'auditeur connaît précisément la mélodie qu'il doit extraire du "mélange" (expérience 5). Dans le but de faciliter l'identification de ces deux tâches, nous parlerons de la condition *pré* lorsque la mélodie à reconnaître est présentée *avant* la séquence composite, par contraste avec la situation de *post*-reconnaissance. L'amélioration des performances de reconnaissance de la mélodie cible constatée lorsque l'auditeur connaissait la mélodie qu'il devait extraire de la séquence composite, nous a conduit à examiner la nature des schémas qui avaient pu être impliqués dans

cette analyse perceptive (expérience 6). Des arguments expérimentaux ont été avancés en faveur de l'existence de deux mécanismes distincts impliqués dans ces deux types de tâches : un mécanisme de partition sensorielle (analyse primaire de scène auditive) et un mécanisme de sélection d'informations (analyse guidée par les connaissances), étayant la théorie de Bregman (1990, 1994) (expérience 7).

5.2 Expérience 5 : Reconnaissance de mélodies intercalées avec ou sans présentation préalable de la mélodie

5.2.1 Introduction

Cette première expérience a pour objectif d'étudier la contribution des connaissances dans la capacité de reconnaître une mélodie entremêlée avec des sons distracteurs. Deux paradigmes différents dans lesquels la mélodie à reconnaître était présentée préalablement ou non à la séquence à organiser ont été proposés respectivement à deux groupes de participants (conditions pré et post), ainsi qu'une tâche contrôle dans laquelle les mélodies étaient présentées isolément. Nous avons ainsi examiné si la connaissance précise de la mélodie permettait d'améliorer significativement les performances de reconnaissance de la mélodie cible intercalée, par rapport à une situation où les connaissances des auditeurs se réduisaient à la connaissance du registre (constant) dans lequel apparaissait la mélodie cible (voir le chapitre 4 pour une discussion des connaissances impliquées dans le paradigme de post-reconnaissance).

Si la connaissance préalable de la mélodie modifie l'organisation perceptive de la séquence composite, les auditeurs devraient parvenir à extraire la mélodie pour une différence de hauteur moyenne la séparant de la séquence distractrice plus petite dans la condition pré que dans la condition post (courbe psychométrique translatée vers la gauche). En revanche, si la préactivation de la mélodie améliore le contraste de celle-ci dans la séquence composite par un mécanisme d'appariement entre le schéma activé et la représentation sensorielle sans qu'elle soit pour autant séparée perceptivement de la séquence distractrice, les performances de reconnaissance devraient être globalement augmentées quel que soit le degré de séparation en hauteur (translation de la courbe vers le haut).

5.2.2 Méthode

La méthode générale employée (les stimuli, la procédure et le matériel) et les analyses statistiques effectuées, sont globalement identiques à celles décrites dans la série d'expériences précédentes (voir §4.2). Seules les spécificités de cette expérience sont donc mentionnées.

5.2.2.1 Conditions pré et post

Dans le but de mettre en évidence la contribution des connaissances dans la fission mélodique, deux conditions *pré* et *post* ont été contrastées (figure 5.1).

Dans la condition *post*, la mélodie à comparer est présentée *après* la mélodie cible intercalée à la séquence distractrice afin de minimiser les connaissances préalables dont dispose l'auditeur pour extraire la mélodie du "mélange", et ainsi faire intervenir majoritairement des processus primaires d'organisation. Cette condition correspond au paradigme de post-reconnaissance élaboré précédemment (voir chapitre 4).

Dans la condition *pré*, la mélodie à reconnaître est présentée *avant* la séquence composite. Les auditeurs disposent alors d'une représentation précise de la mélodie qu'ils doivent extraire du "mélange". Cette condition correspond au paradigme mis en place par Dowling (1973, expérience 2), et utilisé par Dowling et ses collaborateurs (1987) ainsi que Vliegen et Oxenham (1999, expérience 2).

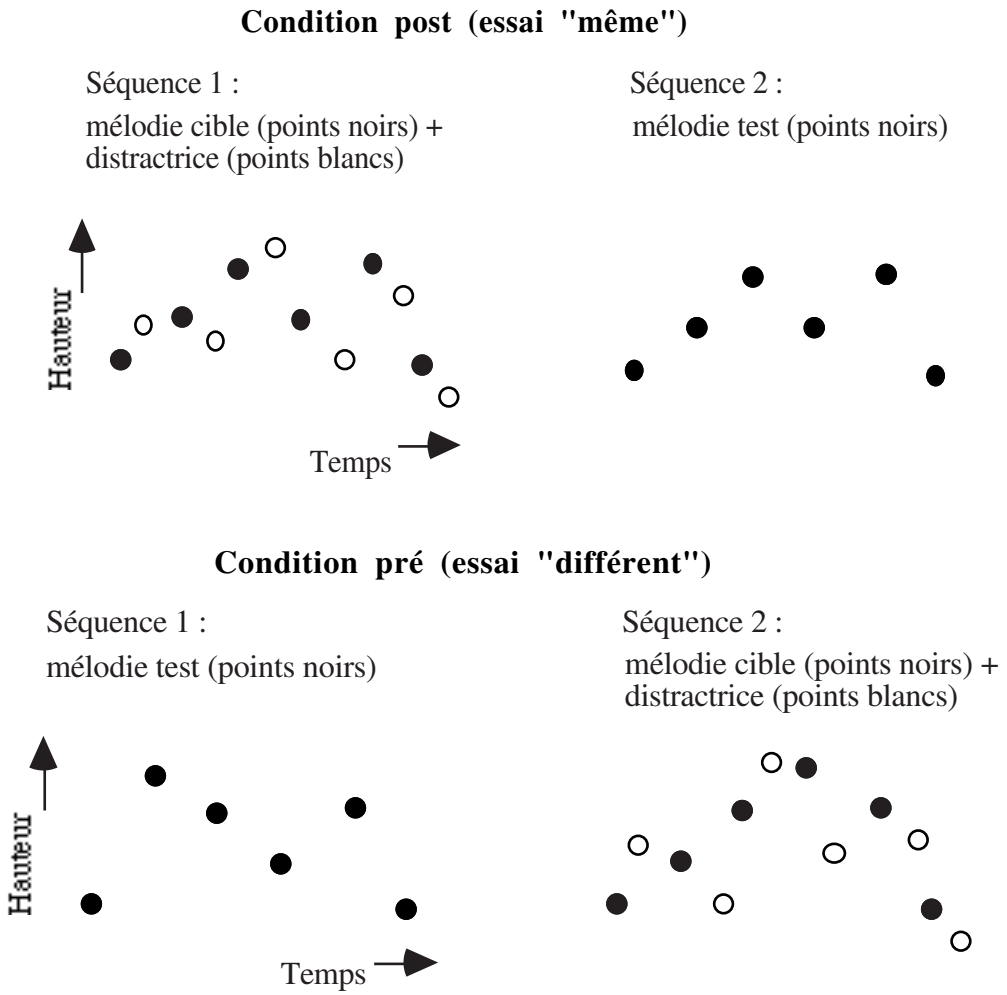


Figure 5.1. Illustration visuelle des deux types de conditions expérimentales. Dans la condition *post*, la mélodie test est présentée *après* la mélodie cible intercalée avec la séquence distractrice. Dans la condition *pré*, elle est présentée *avant*.

5.2.2.2 Stimuli

Les mélodies et les distractrices étaient identiques à celles utilisées dans les expériences précédentes. Les séquences étaient composées de sons purs.

5.2.2.3 Procédure

La séquence distractrice était présentée dans la même gamme de fréquences que la mélodie cible (0 dt de différence de hauteur moyenne entre la mélodie et la distractrice), ou éloignée de 1, 2, 3, 4, 6, 8, 12 ou 24 dt en moyenne vers les fréquences basses. Des exemples sonores de chacune de ces 9 conditions de séparation en hauteur figurent sur le disque, au niveau des plages 1 à 9 pour la condition post et des plages 37 à 45 pour la condition pré (annexe D).

La mélodie cible était toujours présentée dans le même registre de hauteur, seul celui de la distractrice variait. Dans chacune des 9 conditions de séparation en hauteur, 24 essais étaient proposés soit au total 216 essais dans la condition expérimentale (post ou pré). Pour la moitié de ces essais les mélodies étaient identiques, les deux versions originales étaient présentées dans six essais et les deux versions modifiées dans les six autres. Pour l'autre moitié les mélodies différaient par deux notes, la version originale suivie de la version modifiée ont été présentées dans six essais et l'inverse dans les six autres. La différence de hauteur moyenne entre la mélodie et la distractrice était affectée aléatoirement à chaque essai ainsi que le type d'essai "même" et "différent".

L'expérience se composait donc d'une condition expérimentale dans laquelle la mélodie était intercalée à une séquence distractrice, composée de 216 essais. Il s'agissait de la condition pré ou post selon le groupe d'auditeurs. Cette condition expérimentale était suivie d'une condition contrôle constituée de 24 essais, consistant en une simple tâche de reconnaissance de mélodies présentées sans séquence distractrice (la tâche contrôle est identique à celle illustrée sur la figure 4.1b). Chacun de ces tests était précédé de quelques essais de familiarisation, au nombre de 10 de difficulté croissante dans la condition expérimentale et de 4 dans la condition contrôle, au cours desquels la réponse correcte était donnée aux auditeurs. La passation durait environ 1 heure. La consigne écrite donnée à chacun des deux groupes de participants est reportée dans l'annexe C5.

5.2.2.4 Participants

Vingt-six auditeurs ont participé à l'expérience. Les résultats de six participants non musiciens, trois ayant été soumis à la condition post et trois à la condition pré, ont été écartés. Quatre auditeurs répondaient au hasard en présence de distractrice même lorsqu'elle était

éloignée en moyenne de 12 ou 24 dt (pourcentages de réponses correctes compris entre 50 et 67 %). Les deux autres avaient des performances extrêmement variables au travers les conditions, rendant peu fiables leurs données. Ils pouvaient par exemple répondre au hasard pour une différence de 8 dt, puis avoir 70 % de réponses correctes à 12 dt, et de nouveau avoir des performances aléatoires dans la condition 24 dt. Les performances de ces six participants dans la condition contrôle étaient par ailleurs comprises entre 88 et 100 % de réponses correctes.

Les résultats de vingt auditeurs ont donc été intégrés à l'analyse, dix dans la condition post et dix dans la condition pré. La composition des deux groupes était homogène : trois femmes et sept hommes composaient chacun des deux groupes, l'âge moyen des auditeurs ayant été soumis à la condition post était de 24 ans [19 ; 34], ceux ayant participé à la condition pré avaient en moyenne 25 ans [19 ; 28], et le nombre de musiciens et de non musiciens était équivalent dans chacun des groupes (8 musiciens et 2 non musiciens).

5.2.3 Résultats et discussion

5.2.3.1 Une différence de sensibilité

Sur la base de la théorie de la détection du signal (Green & Swets, 1974) appliquée au paradigme même/différent (Macmillan & Creelman, 1991), nous avons calculé la proportion de détections correctes (proportion de réponses "différent" lorsque les mélodies sont effectivement différentes) et de fausses alarmes (proportion de réponses "différent" lorsque les mélodies sont identiques). Nous avons ainsi déterminé le d' de chaque auditeur dans chacune des 9 conditions de séparation en hauteur puis moyenné les d' des 10 participants dans chacun des groupes et chacune des conditions de séparation en hauteur (figure 5.2).

Comme nous pouvons l'observer sur le graphique, les performances de reconnaissance de la mélodie cible intercalée sont accrues lorsque l'auditeur dispose d'une représentation précise de la mélodie qu'il doit extraire du mélange (pré), par rapport à la situation où les seules connaissances implicites disponibles sont la constance du registre de fréquence dans lequel la mélodie cible est présentée (post). Les performances obtenues dans la condition pré pour une différence de hauteur moyenne de 2 dt sont obtenues dans la condition post pour une différence de hauteur de 6 dt, soit un gain de 4 dt. Par ailleurs, les performances de reconnaissance de mélodies en l'absence de séquence distractrice (tâche contrôle) étaient identiques pour les deux groupes d'auditeurs, attestant de l'équivalence de ces groupes indépendants dans une tâche de discrimination mélodique.

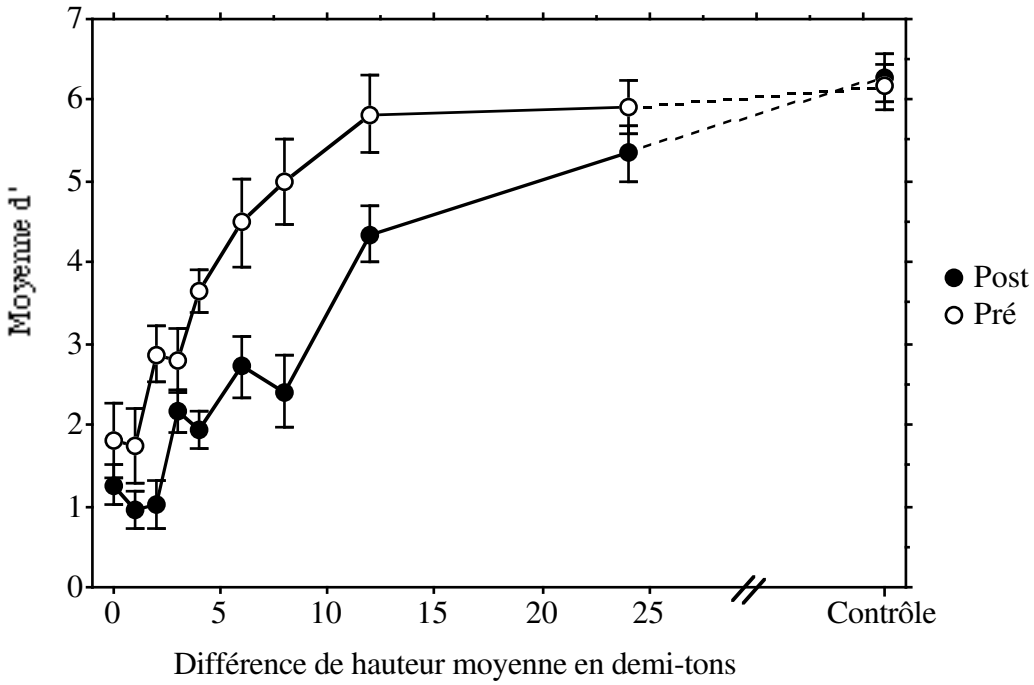


Figure 5.2. Moyenne des d' en fonction de la différence de hauteur moyenne entre la mélodie cible et la distractrice pour les conditions post (courbe avec les points noirs) et pré (courbe avec les points blancs). Afin de comparer les performances obtenues avec celles de la condition contrôle, celle-ci figure également sur l'axe des abscisses. Les barres verticales indiquent ± 1 erreur-type.

Une analyse de variance (plan d'analyse S10<G2>*H10) évaluant le rôle des différentes conditions, les 9 degrés de séparation en hauteur et la condition contrôle, en fonction du type de tâche pré ou post, a été réalisée sur la variable dépendante d'. Elle révèle que la différence entre ces deux types de tâches (pré/post) est statistiquement significative, $F(1, 18) = 13.67$, $p < .005$. Une analyse de variance multiple appliquant la correction de Bonferroni qui consiste à diviser le seuil de significativité (.05) par le nombre d'analyses réalisées (ici 10), a été effectuée pour comparer les d' obtenus dans les conditions pré et post pour chacune des dix conditions de séparation en hauteur. Cette analyse a révélé que les performances de reconnaissance étaient significativement différentes au seuil .005 dans les deux types de tâche pour une différence de hauteur moyenne entre la mélodie et la séquence distractrice de 2 dt, $F(1, 18) = 15.89$, $p < .001$, 4 dt, $F(1, 18) = 24.37$, $p < .0001$, et 8 dt, $F(1, 18) = 14.03$, $p < .005$. Pour les conditions de séparation en hauteur de 6 dt et 12 dt la différence entre les deux types de tâche n'était significative qu'à .05.

La courbe psychométrique obtenue dans la condition pré est non seulement translatée vers la gauche, signifiant que l'auditeur parvient à extraire la mélodie pour une différence de hauteur moyenne plus petite que dans la condition post, mais également vers le haut suggérant

que la connaissance préalable de la mélodie induit une augmentation globale des performances pour l'ensemble des degrés de séparation en hauteur.

Les auditeurs parviennent à extraire la mélodie cible du mélange pour une différence de hauteur moyenne avec les sons distracteurs plus petite dans la condition pré que dans la condition post. Les performances de reconnaissance s'élèvent à partir d'une différence de 2 demi-tons en moyenne lorsque l'auditeur connaît la mélodie, alors qu'une même augmentation n'est observée qu'à partir de 3 demi-tons en moyenne dans la condition post. Ce résultat suggère que la connaissance de la mélodie a modifié l'organisation de la séquence composite et permis la fission perceptive de la mélodie pour une différence de hauteur moyenne plus petite.

Ce résultat corrobore celui obtenu par van Noorden (1975) qui a montré que la différence de hauteur conduisant à la fission perceptive d'une séquence variait en fonction de ce que l'on cherchait à entendre. Elle était d'environ 2-3 dt lorsque les auditeurs cherchaient à séparer les sons aigus et graves d'une séquence cyclique (SF), et de 10-12 dt lorsqu'ils essayaient de les intégrer (SCT), pour un tempo de 6 sons/sec correspondant à celui des séquences présentées ici. Ce résultat converge également en partie avec celui obtenu par Dowling (1973, expérience 3). W. J. Dowling a en effet montré qu'un amorçage verbal congruent (titre de la mélodie présentée) permettait d'extraire une mélodie familière présentée dans un mélange au bout de 3.6 présentations en moyenne en l'absence même de différence de fréquence. Les résultats de cette étude ont été confirmés par la suite par Dowling et al. (1987, expérience 1), avec des auditeurs moins entraînés et une présentation unique de la séquence composite. Les auteurs ont montré que les auditeurs étaient capables de reconnaître 78 % des mélodies familières présentées dans un mélange en l'absence de différence de hauteur moyenne, si le titre leur était préalablement donné.

Néanmoins, si dans notre étude l'amorçage de la mélodie inédite a permis de l'extraire pour une différence de fréquence plus petite, il n'a pas pour autant permis de l'extraire pour une différence nulle. En effet, les performances obtenues dans la condition pré lorsque la mélodie était présentée dans le même registre que la séquence distractrice sont équivalentes à celles obtenues dans la condition post. Deux hypothèses peuvent être avancées pour rendre compte de cette différence entre les résultats de notre étude et de celles menées par W. J. Dowling (Dowling, 1973, expérience 3 ; Dowling et al., 1987, expérience 1). La première est que l'extraction de la mélodie dans une "mixture" sonore en l'absence de processus primaires d'organisation pourrait dépendre de la nature du schéma. En effet, Dowling et ses collaborateurs (Dowling, 1973, expérience 3 ; Dowling et al., 1987, expérience 1) utilisent des comptines stockées en mémoire à long terme, alors que dans notre étude il s'agit de mélodies non familières qui par conséquent sont juste stockées en mémoire à court terme. Or Dowling et ses coauteurs (Dowling, 1978 ; Dowling & Fujitani, 1971 ; Dowling & Harwood, 1986) ont montré que les auditeurs mémorisaient à court terme le contour de la mélodie, et à long terme la taille des intervalles (voir §3.1.2 pour le détail de ces études). Cette différence de nature des

indices encodés, liée à la différence de profondeur de traitement et donc de durée de stockage, pourrait expliquer la différence d'extraction d'une mélodie dans un mélange. La nature des schémas pourrait ainsi conditionner la puissance de ce mécanisme descendant. Une deuxième hypothèse pouvant être formulée pour lever cette apparente divergence, est que les mélodies familières intercalées utilisées par W. J. Dowling (Dowling, 1973, expérience 3 ; Dowling et al., 1987, expérience 1) pourraient, malgré une différence de fréquence moyenne nulle, présenter des différences de fréquences locales induisant une fission partielle de la mélodie. Cette fission partielle induite par des processus primaires d'analyse de scène, permettrait ainsi l'émergence de portions de chacune des deux mélodies suffisant à activer en mémoire le schéma de celle-ci. Si l'on prend l'exemple des mélodies familières Frères Jacques et Joyeux Anniversaire intercalées et présentées dans le même registre (figure 1.5), on peut remarquer que dans certains passages comme par exemple les mesures 5 et 6, les notes appartenant aux deux mélodies sont distantes en moyenne de 7 dt (différences entre les notes successives comprises dans une gamme allant de 3 à 10 dt). Cette différence de hauteur est supérieure au seuil de fission établi par van Noorden (1975) qui, nous l'avons vu dans le chapitre précédent, correspond à la fission partielle de la mélodie. Ceci laisse penser que les mélodies pourraient être partiellement séparées et parce qu'il s'agit de mélodies familières, ces quelques notes suffiraient à retrouver la trace de la mélodie en mémoire.

Les performances sont globalement plus élevées dans la condition pré que dans la condition post, pour l'ensemble des degrés de séparation en hauteur. Deux hypothèses peuvent être avancées pour expliquer ce gain général. La première est l'intervention des processus d'analyse de scène guidée par les connaissances, dont Bregman (1990, 1994) fait l'hypothèse qu'il s'agit de processus de sélection d'information et non de partition sensorielle. La présentation de la mélodie avant la séquence composite permettrait d'élaborer une représentation de celle-ci et de la préactiver. L'appariement entre ce schéma de la mélodie et la représentation sensorielle de la séquence composite, aurait pour conséquence de rendre la mélodie cible intercalée plus saillante sans pour autant la séparer perceptivement de la séquence distractrice. Une autre hypothèse formulée par Bregman (1997) à propos de ces résultats, est que la différence globale de performances entre ces deux tâches pourrait également être le reflet d'une asymétrie dans l'encodage en mémoire de la première mélodie entendue dans ces deux situations expérimentales. En effet, dans la condition post la première mélodie présentée, qui constitue la mélodie de référence, va être encodée avec la séquence distractrice alors que dans la condition pré cette mélodie de référence est encodée seule. La présence de la séquence distractrice dans la tâche de post-reconnaissance pourrait interférer dans l'encodage de la mélodie de référence et dégrader la représentation de celle-ci. La différence globale de performances constatée entre ces deux tâches, pourrait donc s'expliquer par la différence d'encodage en mémoire de la première mélodie entendue. Dans ce paradigme où l'on compare deux mélodies successives, l'encodage de la première mélodie est en effet probablement très importante. Un argument qui pourrait

conforter cette hypothèse de l'existence d'une interférence en mémoire plus importante dans la condition post que dans la condition pré, est le fait que les performances obtenues dans la condition de séparation maximale entre la distractrice et la mélodie (24 dt) sont inférieures à celles obtenues dans la condition contrôle en post mais pas en pré. Une analyse de variance comparant les d' obtenus dans chaque condition de séparation en hauteur a été réalisée pour chacune des tâches. L'effet de la différence de hauteur moyenne séparant la séquence distractrice de la mélodie cible était significatif dans les conditions post, $F(9, 81) = 44.96$, $p < .0001$ et pré, $F(9, 81) = 24.82$, $p < .0001$. Des contrastes planifiés issus de cette analyse montrent que la différence de performances observée entre les conditions 24 dt et contrôle est significative en post, $F(1, 81) = 5.69$, $p < .05$ mais pas en pré, $F(1, 81) < 1$.

Enfin, à la lecture de ces deux courbes on peut constater que la progression des performances de reconnaissance de la mélodie intercalée diffère dans les deux tâches. En effet, si le d' augmente progressivement à mesure que la différence de hauteur moyenne entre la mélodie et la distractrice s'accroît dans la condition pré, il n'en est pas de même dans la condition post où la fonction présente des discontinuités (des paliers et des sauts de performances). L'analyse de variance a révélé que l'interaction entre l'effet du type de tâche et les conditions de séparation en hauteur était significatif, $F(9, 162) = 3.48$, $p < .005$. Ces deux progressions différentes pourraient être le reflet de l'implication de deux processus distincts dans ces deux types de tâche. En effet, les discontinuités observées dans la tâche de post-reconnaissance pourraient être liées à la fission des sons distracteurs ; un palier s'observerait lorsqu'un nombre constant de sons distracteurs se sépare perceptivement de la séquence composite et un saut de performances correspondrait à des sons supplémentaires qui se séparent laissant émerger des liens entre les notes de la mélodie cible. Si ces discontinuités sont effectivement liées au nombre de sons distracteurs qui se séparent du mélange à mesure que la différence de hauteur moyenne augmente, cela signifierait que les performances de post-reconnaissance dépendent directement des processus primaires d'organisation, c'est-à-dire de la fission pré-attentive des sons de la distractrice. Par conséquent, l'absence de discontinuités dans la condition pré pourrait être le reflet de l'implication d'un processus supplémentaire de sélection d'information qui "lisserait" cette progression. Ce résultat étayerait la théorie de Bregman (1990, 1994) qui postule l'existence d'un processus d'analyse de scène guidée par les connaissances de nature différente du processus primaire d'analyse de scène auditive.

5.2.3.2 Une différence de jugement

A une différence de sensibilité observée entre les conditions pré et post, s'ajoute une différence dans les jugements opérés par les auditeurs. En effet, les proportions moyennes de détections correctes et de fausses alarmes sont globalement plus élevées dans la condition pré que dans la condition post, surtout pour les degrés de séparation faibles (figures 5.3 et 5.4, le lecteur devra se focaliser uniquement sur les deux courbes correspondant aux conditions pré et post de cette expérience 5).

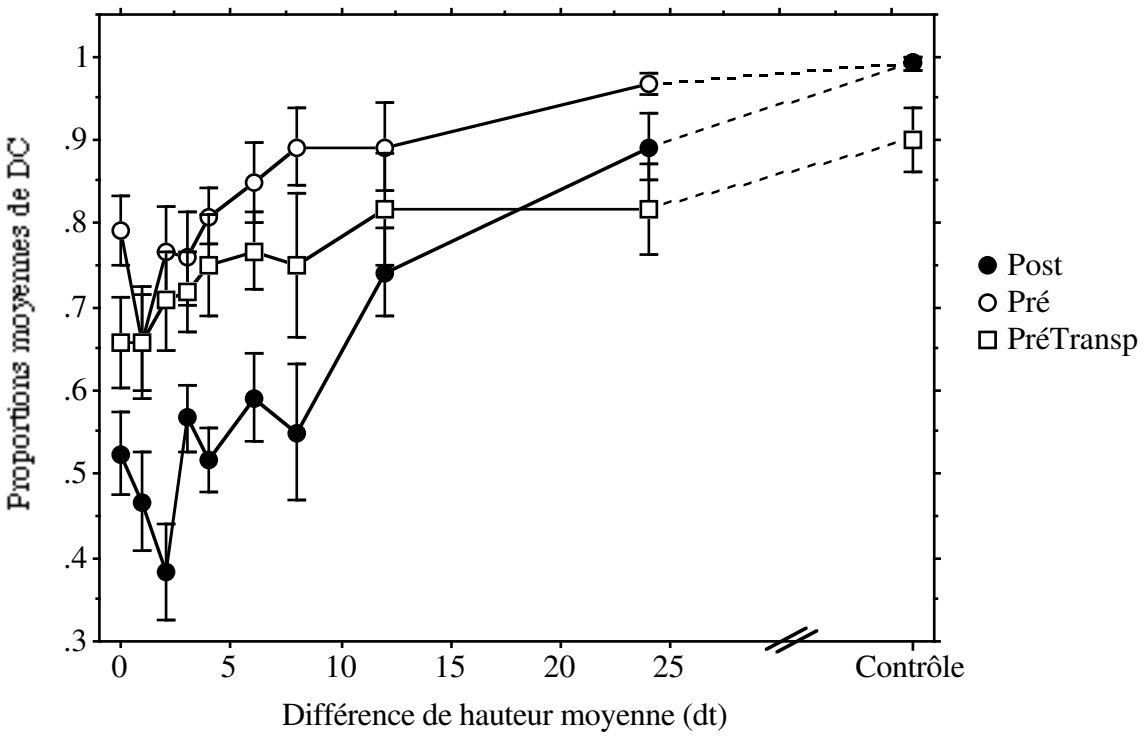


Figure 5.3. Proportions moyennes de détections correctes (DC) obtenues par 10 participants (8 musiciens et 2 non musiciens) dans les conditions Post et Pré (expérience 5), ainsi que dans une tâche de reconnaissance de mélodies intercalées dans laquelle la mélodie présentée préalablement est transposée de +12, 13 ou 14 dt (PréTransp) (expérience 6). Ceci pour chaque condition de séparation en hauteur moyenne (0, 1, 2, 3, 4, 6, 8, 12 et 24 dt), ainsi que pour la condition contrôle (pour l'expérience PréTransp il s'agit de la condition contrôle avec transposition). Les barres verticales indiquent ± 1 erreur-type.

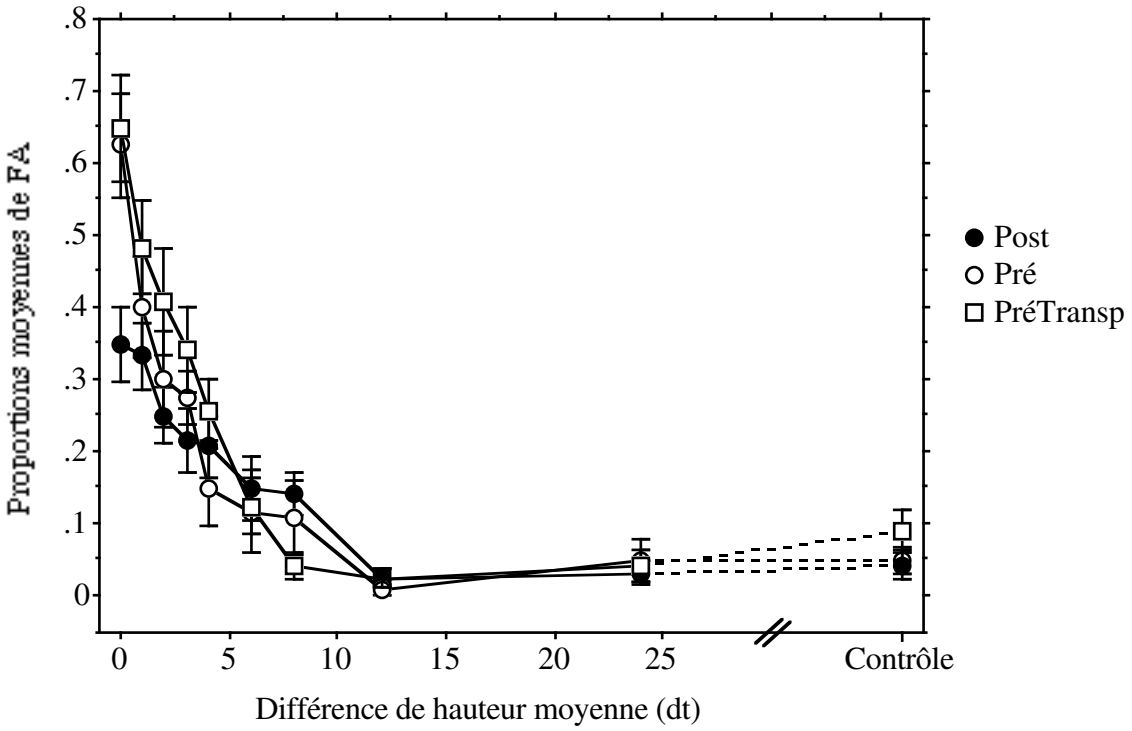


Figure 5.4. Proportions moyennes de fausses alarmes (FA) obtenues par 10 participants (8 musiciens et 2 non musiciens) dans les conditions Post et Pré (expérience 5), ainsi que dans une tâche de reconnaissance de mélodies intercalées dans laquelle la mélodie présentée préalablement est transposée de +12, 13 ou 14 dt (PréTransp) (expérience 6). Ceci pour chaque condition de séparation en hauteur moyenne (0, 1, 2, 3, 4, 6, 8, 12 et 24 dt), ainsi que pour la condition contrôle (pour l'expérience PréTransp il s'agit de la condition contrôle avec transposition). Les barres verticales indiquent ± 1 erreur-type.

Lorsque la différence de hauteur moyenne entre la mélodie cible et la séquence distractrice était comprise entre 0 et 2 dt, la mélodie cible était difficile voire impossible à extraire du mélange. Les auditeurs qui commençaient par entendre la séquence composite (post) avaient par conséquent une représentation très approximative voire inexistante de cette mélodie de référence qu'ils devaient comparer à la mélodie qui suivait. En revanche, les auditeurs pour qui la mélodie de référence était présentée isolément (pré) pouvaient élaborer une représentation précise de celle-ci. Il se trouve que ces derniers ont répondu plus souvent que les deux mélodies présentées étaient différentes. Cette différence de critère de réponse, pourrait découler de la différence de précision de la représentation de la mélodie de référence induite par la présence ou non de la séquence distractrice. On peut émettre l'hypothèse en effet que la distance perceptive entre les deux mélodies à comparer, de laquelle dépend le jugement même/différent, serait liée à la précision de la représentation de la première mélodie entendue. Pour une même différence perceptive, on aurait plus tendance à juger que les mélodies sont différentes si on a une représentation précise de la mélodie de référence et donc des attentes spécifiques sur la mélodie qui suit, que dans le cas contraire.

En résumé, cette expérience montre que lorsque l'auditeur connaît la mélodie qu'il doit sélectionner dans une "mixture" sonore, ses performances de reconnaissance sont accrues par rapport à une situation où il n'a pas entendu préalablement cette mélodie. Cette différence de performances se caractérise à la fois par une différence de sensibilité et par un biais de réponse. La mélodie est extraite pour une différence de hauteur moyenne la séparant des sons distracteurs plus petite et les performances sont globalement augmentées, suggérant que les connaissances ont non seulement modifié la partition sensorielle mais aussi amélioré de façon globale l'extraction de la mélodie cible. En outre, les auditeurs répondent plus souvent que les mélodies présentées sont différentes lorsqu'ils connaissent précisément la mélodie qu'ils ont à comparer. Ces différences de sensibilité et de critère de réponse pourraient être le résultat de l'implication des processus d'analyse de scène guidée par les connaissances, mais aussi la conséquence de la différence d'encodage en mémoire de la première mélodie dans les deux types de tâches. La réalité se trouve probablement dans la combinaison de ces deux effets.

L'objectif des deux expériences suivantes est donc de préciser l'interprétation que l'on peut donner à la différence de performances observée dans ces deux tâches. L'expérience 6 examine la nature des connaissances impliquées dans l'amélioration de ces performances. L'expérience 7 a pour objectif de déterminer si les discontinuités observées dans la tâche de post-reconnaissance sont liées à la fission perceptive des sons distracteurs, ceci afin de déterminer la nature des processus mis en jeu dans ces deux tâches.

5.3 Expérience 6 : Reconnaissance de mélodies intercalées lorsque la mélodie présentée préalablement est transposée de +12, 13 ou 14 demi-tons

5.3.1 Introduction

Cette expérience se propose de déterminer la nature des schémas qui ont permis l'amélioration des performances observée dans la condition pré. Ce gain de performances est-il le résultat d'un amorçage sensoriel fréquentiel, ou de l'amorçage perceptif du lien entre les sons (contour et/ou intervalles mélodiques), ou bien les deux ?

Pour examiner cette question, nous avons proposé à des auditeurs une tâche identique à la condition pré proposée dans l'expérience précédente, à la différence que la première mélodie présentée était transposée de façon aléatoire de +12, 13 ou 14 dt (prétransposé). Si le gain de performances observé dans la condition pré est lié à un amorçage fréquentiel, les performances obtenues lorsque la première mélodie est transposée devraient être inférieures à celles observées sans transposition. Et si les fréquences des notes constituent les seules informations qui permettent d'améliorer la reconnaissance de la mélodie cible intercalée, les performances devraient même être comparables à celles obtenues dans la condition post. En revanche, si le

contour et/ou les intervalles (ces deux indices n'étant pas dissociés dans cette expérience) sont responsables de l'augmentation des performances observée dans la condition pré, on ne devrait pas observer de différence entre les performances obtenues avec et sans transposition, puisque ces informations sont toujours présentes. Néanmoins, si l'on n'observait pas de différence de performances dans ces deux tâches, cela pourrait également signifier que le gain constaté dans la condition pré n'est pas lié à l'implication de processus d'analyse de scène guidée par les connaissances, mais pourrait par exemple être le seul reflet d'une différence dans les processus mnésiques mis en jeu. Enfin, si les deux informations, les fréquences et le lien entre les fréquences (contour et/ou intervalles) ont contribué à ce gain, les performances obtenues dans cette expérience où la mélodie présentée préalablement est transposée devraient être intermédiaires à celles recueillies dans les conditions pré et post.

5.3.2 Méthode

La méthode utilisée était globalement identique à celle de l'expérience 5 pour la condition pré, seules les caractéristiques spécifiques à cette expérience sont donc signalées.

5.3.2.1 Stimuli

Nous avons choisi de transposer la première mélodie d'une valeur supérieure à l'ambitus maximal des mélodies employées (soit une valeur supérieure à 11 dt), afin de s'assurer qu'il n'y ait pas de recouvrement fréquentiel entre les deux mélodies présentées, et donc plus d'amorçage fréquentiel possible. Ce degré de transposition variait aléatoirement entre trois valeurs : +12, +13 ou +14 demi-tons. La fluctuation du degré de transposition à chaque essai a été mise en place dans le but d'éviter que l'auditeur ne développe une éventuelle stratégie de calcul lui permettant d'imaginer les notes de la deuxième mélodie à partir de la première, et ainsi de préactiver les fréquences de la mélodie cible. Enfin, la première mélodie a été transposée vers les fréquences élevées pour ne pas amorcer les fréquences des sons de la séquence distractrice qui dans cette tâche doivent être ignorés des auditeurs.

Des exemples sonores de cette tâche reconnaissance de mélodies intercalées avec la première mélodie transposée aléatoirement de +12, 13 ou 14 dt, se trouvent au niveau des plages 46 à 54 du disque pour les différentes conditions de séparation en hauteur entre la mélodie et la séquence distractrice (0, 1, 2, 3, 4, 6, 8, 12 et 24 dt) (voir annexe D pour la structure précise des essais présentés). Les plages 55 à 57 présentent des extraits de la condition contrôle dans laquelle la première mélodie est transposée respectivement de +12, 13 ou 14 dt.

5.3.2.2 Procédure

Nous avons soumis aux auditeurs cette tâche reconnaissance de mélodies intercalées avec la première mélodie transposée de +12, 13 ou 14 demi-tons. Cette condition expérimentale était composée de 216 essais (9 degrés de séparation en hauteur * 24 observations par condition) et précédée de 10 essais de familiarisation de difficulté croissante. Elle était suivie de deux conditions contrôles : une tâche de reconnaissance de mélodies avec la première mélodie transposée de +12, 13 ou 14 demi-tons et une tâche de reconnaissance de mélodies dans laquelle les deux mélodies étaient présentées dans le même registre de fréquence. Ces deux conditions contrôles étaient chacune composées de 24 essais et précédées de 4 essais de familiarisation. Leur ordre de présentation a été contrebalancé sur l'ensemble des participants. La durée totale de la passation était d'environ 1h 15. La consigne donnée aux participants figure dans l'annexe C6.

5.3.2.3 Participants

Quatorze auditeurs ont participé à l'expérience, quatre femmes et dix hommes d'un âge moyen de 25 ans [21 ; 30] dont huit avaient reçu une formation musicale. Les auditeurs étaient indemnisés pour leur participation.

5.3.3 Résultats et discussion

5.3.3.1 Rôle des informations fréquentielles dans la reconnaissance immédiate de mélodies non familières

Les performances de reconnaissance de la mélodie cible intercalée obtenues par les auditeurs musiciens et non musiciens, sont représentées sur la figure 5.5. Elles augmentent avec la différence de hauteur moyenne séparant la mélodie de la séquence distractrice. Pour une différence supérieure ou égale à 6 dt, les participants ayant bénéficié d'une formation musicale ont des performances supérieures à celles des non musiciens. Pour les deux groupes d'auditeurs, les performances obtenues pour une différence de 24 dt sont équivalentes à celles obtenues dans la condition contrôle dans laquelle la première mélodie a également été transposée. Cependant, elles s'avèrent inférieures à celles obtenues dans la condition contrôle sans transposition. Cette différence est en moyenne de 2 points sur l'échelle du d', 1.5 chez les musiciens et 2.5 chez les non musiciens. Les musiciens et les non musiciens ont par ailleurs des performances de reconnaissance équivalentes dans la condition contrôle sans transposition.

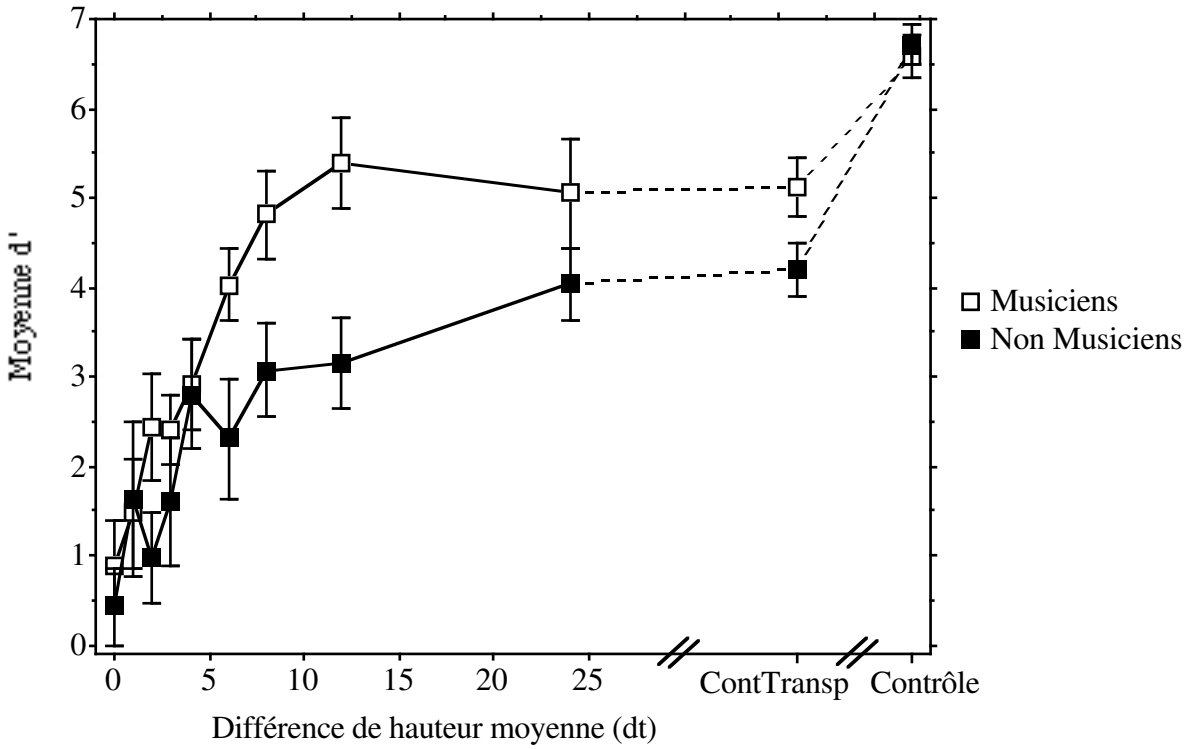


Figure 5.5. Moyenne des d' obtenus par 14 auditeurs (8 musiciens et 6 non musiciens) dans la tâche de reconnaissance de mélodies intercalées lorsque la mélodie présentée préalablement est transposée de +12, 13 ou 14 dt pour les différentes conditions de séparation en hauteur (0, 1, 2, 3, 4, 6, 8, 12 et 24 dt) et les deux conditions contrôles avec et sans transposition (respectivement ContTransp et Contrôle). Les barres verticales indiquent ± 1 erreur-type.

Une analyse de variance (Plan d'analyse S<M2>*H11) évaluant l'effet des 11 conditions, les 9 degrés de séparation en hauteur entre la mélodie et la séquence distractrice et les deux conditions contrôle respectivement avec et sans transposition, en fonction de la formation musicale des auditeurs, a été réalisée sur la variable dépendante d'. L'effet de ces différentes conditions était statistiquement significatif, $F(10, 120) = 45.1$, $p < .0001$. Les contrastes planifiés n'ont pas révélé de différence significative entre les performances obtenues pour une différence de hauteur moyenne de 24 dt et celles obtenues dans la condition contrôle dans laquelle la première mélodie était transposée, $F(1, 120) < 1$, comme c'était le cas dans la condition pré de l'expérience précédente où la mélodie présentée préalablement n'était pas transposée. Cependant, les performances obtenues dans la condition contrôle avec transposition s'avèrent significativement inférieures à celles obtenues sans transposition, $F(1, 120) = 29.12$, $p < .0005$. Ce résultat est cohérent avec celui obtenu par Dowling et Fujitani (1971, expérience 1) qui montrent que les performances de reconnaissance immédiate d'une mélodie sont altérées si les deux mélodies ne sont pas présentées dans le même registre, suggérant que l'information fréquentielle est importante pour la rétention de mélodies non familières.

L'effet de la formation musicale n'était globalement pas significatif, $F(1, 12) = 2.92$, N.S, mais l'interaction entre la formation musicale et les conditions de séparation en hauteur est juste en-dessous du seuil de significativité, $F(10, 120) = 2.43$, $p = .052$. Une analyse multiple appliquant la correction de Bonferroni au seuil de significativité, ne révèle pas de comparaison significative à .005 entre les performances des musiciens et des non musiciens pour chaque degré de séparation en hauteur. Cependant, les musiciens ont tendance à être plus performants dans cette tâche pour les conditions de séparation en hauteur moyenne de 6 dt, $F(1, 12) = 5.54$, $p < .05$, 8 dt, $F(1, 12) = 5.69$, $p < .05$ et 12 dt $F(1, 12) = 9.37$, $p < .01$.

5.3.3.2 Amorçage fréquentiel

Compte tenu de cette tendance, nous avons choisi de comparer les trois tâches, celles de reconnaissance de mélodies intercalées lorsque la mélodie est présentée préalablement et transposée (prétransposé) ou non (pré) et la tâche de post-reconnaissance (post), en veillant à ce que la proportion de musiciens et de non musiciens soit équivalente dans les trois groupes. Dans les conditions pré et post de l'expérience précédente, les groupes étaient respectivement composés de huit musiciens et deux non musiciens. Nous avons donc inclus dans le groupe d'auditeurs ayant passé cette tâche de reconnaissance de mélodies intercalées avec transposition, l'ensemble des musiciens ayant passé le test (au nombre de huit) et seulement deux non musiciens choisis au hasard parmi les six. Les résultats de ces trois groupes d'auditeurs sont portés sur le même graphique (figure 5.6).

La comparaison des performances obtenues dans ces trois types de tâches, révèle que lorsque la première mélodie est présentée isolément mais transposée (prétransposé), les performances de reconnaissance de la mélodie cible intercalée sont inférieures à celles observées en l'absence de transposition (pré), mais supérieures à celles obtenues dans une tâche où l'auditeur ne connaît pas précisément la mélodie qu'il doit extraire du mélange (post).

Une analyse de variance (Plan d'analyse S<G3>*H10) a été effectuée sur ces trois groupes indépendants composés de 10 auditeurs ayant réalisés 10 mesures répétées (9 degrés de séparation en hauteur et la condition contrôle qui est transposée dans le groupe ayant passé le test prétransposé) en conservant le d' comme variable dépendante[1] . L'effet du type de tâche sur les performances de reconnaissance était statistiquement significatif, $F(2, 27) = 5.27$, $p < .05$. L'analyse planifiée des contrastes a révélé une différence significative entre les performances de reconnaissance de mélodies intercalées obtenues avec et sans transposition,

[1] Les analyses de variance réalisées sur ce sous-groupe de 10 auditeurs (8 musiciens et 2 non musiciens) ayant participé à l'expérience de reconnaissance de mélodies intercalées avec transposition, ont également été effectuées sur l'ensemble des auditeurs (soient 14 auditeurs dont 6 non musiciens). La significativité statistique des effets décrits ici était identique.

$F(1, 27) = 4.87$, $p < .05$ mais pas entre les performances de post-reconnaissance et celles de reconnaissance de mélodies intercalées avec transposition, $F(1, 27) < 1$.

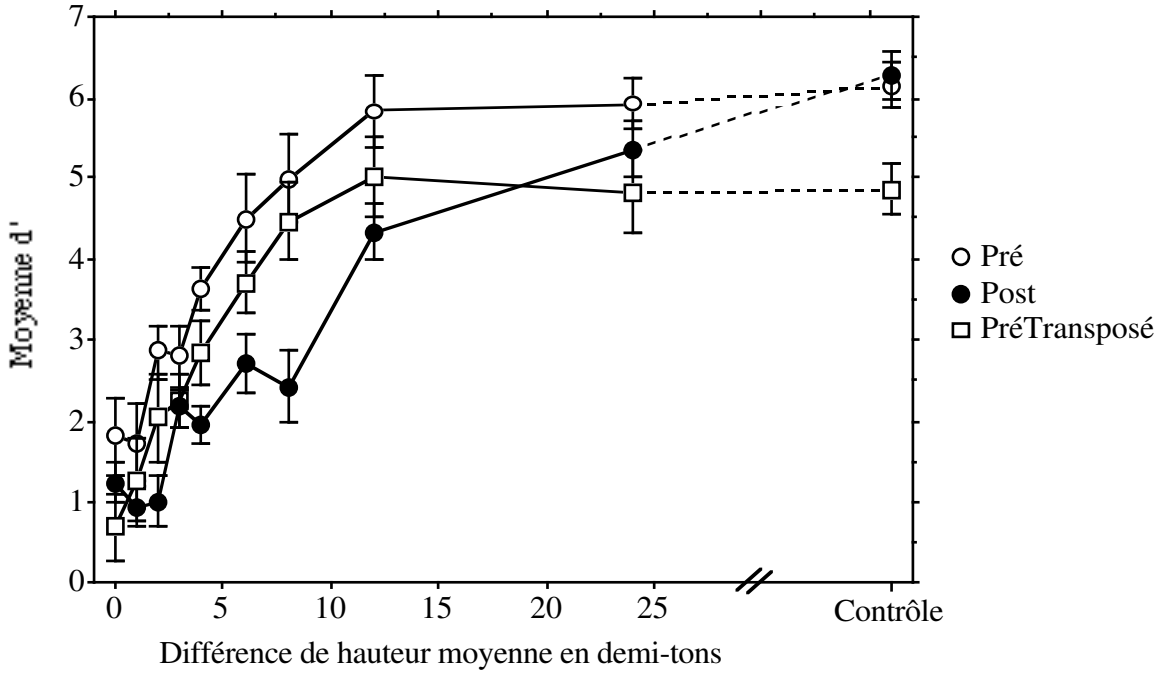


Figure 5.6. Moyenne des d' en fonction de la différence de hauteur moyenne entre la mélodie cible et la distractrice pour les 10 participants (8 musiciens et 2 non musiciens) composant chacun des 3 groupes ayant passé respectivement la tâche de post-reconnaissance (Post, courbe avec les points noirs), et celles de reconnaissance de mélodies intercalées lorsque la mélodie était présentée préalablement dans le même registre (Pré, courbe avec les points blancs) ou transposée de +12, 13 ou 14 dt (PréTransposé, courbe avec les carrés blancs). La condition contrôle figure également sur l'axe des abscisses pour comparaison. Pour l'expérience prétransposé, il s'agit de la condition contrôle avec transposition. Les barres verticales indiquent ± 1 erreur-type.

La dégradation des performances de reconnaissance de mélodies intercalées lorsque la première mélodie est transposée, montre que le gain observé dans l'expérience précédente entre les conditions pré et post était en partie dû à un amorçage sensoriel fréquentiel. Ce résultat conforte ainsi l'hypothèse de l'implication d'un processus d'analyse de scène guidée par les connaissances dans la tâche de reconnaissance de mélodies intercalées lorsque la mélodie est présentée avant la séquence à organiser. En outre, le fait que pour une différence de hauteur moyenne inférieure à 12 dt les performances dans l'expérience prétransposé demeurent supérieures à celles observées en post-reconnaissance, laisse penser que d'autres connaissances comme le contour mélodique et/ou les intervalles contribuent probablement au gain observé entre les conditions pré/post. Les informations qui ont permis d'extraire la mélodie du mélange et améliorer ainsi sa reconnaissance sont donc non seulement la fréquence des notes, mais aussi

le contour et/ou les intervalles. Nous ne pouvons pas distinguer la contribution du contour et des intervalles ici car ils n'ont pas été testés indépendamment. Cependant, puisqu'il s'agit de mélodies non familières, nous pouvons faire l'hypothèse que les auditeurs se sont basés sur le contour plutôt que les intervalles (Dowling & Harwood, 1986).

Les performances augmentent progressivement à mesure que la différence de hauteur s'accroît. La continuité de cette progression s'apparente à celle constatée dans l'expérience précédente pour la condition pré, et contraste avec les paliers et sauts de performances observés en post-reconnaissance.

L'analyse de variance a révélé que l'interaction entre le type de tâche et les conditions de séparation en hauteur était globalement significative, $F(18, 243) = 3.02$, $p < .0005$. Elle était significative entre les expériences prétransposé et post, $F(9, 243) = 5.17$, $p < .0001$ mais pas entre les deux expériences pré avec et sans transposition, $F(9, 243) < 1$. Ceci suggère que la progression des performances observée dans la tâche de reconnaissance de mélodies intercalées avec présentation préalable de la mélodie transposée, diffère de celle observée en post-reconnaissance, tout comme la progression des performances recueillies dans la tâche homologue sans transposition (pré) différait. Il semble donc que cette augmentation progressive du d' en fonction de la différence de hauteur moyenne séparant la mélodie cible de la séquence distractrice, soit caractéristique de cette tâche dans laquelle la mélodie à reconnaître est présentée avant le mélange, par opposition aux discontinuités observées en post-reconnaissance.

5.3.3.3 Un profil de jugement spécifique

Les proportions moyennes de détections correctes et de fausses alarmes ont un profil similaire dans cette expérience prétransposé, à celui obtenu dans la condition pré de l'expérience précédente (Figures 5.3 et 5.4). Ce profil de jugement semble constituer une sorte de "signature" de ce type de tâche dans laquelle la mélodie est présentée avant la séquence à organiser. La présentation préalable de la mélodie conduirait les auditeurs à faire plus de fausses alarmes que lorsqu'ils n'ont pas connaissance de celle-ci. Ce résultat constitue un élément supplémentaire allant dans le sens de l'hypothèse formulée précédemment selon laquelle cette différence de jugement induite par le fait de présenter préalablement ou non la mélodie, pourrait être liée à une différence de précision de la représentation de la première mélodie entendue dans ces deux types de paradigmes.

5.4 Expérience 7 : Nombre de sons séparés et progression des performances de post-reconnaissance de mélodies intercalées

5.4.1 Introduction

L'expérience précédente a montré que l'amélioration des performances de reconnaissance de la mélodie cible intercalée lorsque la mélodie test était présentée préalablement, pouvait s'expliquer par le fait que l'auditeur connaissait la mélodie qu'il devait extraire du mélange, et en particulier ses notes et son contour. Elle a ainsi permis d'identifier les schémas activés. Cette présente expérience a pour objectif de déterminer de quelle façon ces schémas sont utilisés dans l'analyse perceptive d'une séquence composite, et ainsi de préciser la nature des processus impliqués dans ces deux types de tâches (pré/post).

Cette expérience teste l'hypothèse formulée précédemment selon laquelle la progression discontinue des performances observée dans la tâche de post-reconnaissance, serait liée à la fission perceptive des sons distracteurs. Les paliers observés correspondraient à un nombre constant de sons séparés et les sauts de performances entre les conditions 2 et 3 dt, 4 et 6 dt et 8 et 12 dt à un changement du nombre de sons qui se séparent perceptivement de la séquence composite. Si cette hypothèse est confirmée, nous pourrons en déduire que dans la mesure où la progression des performances est au contraire continue dans la condition pré, cette tâche fait intervenir non seulement des processus primaires d'organisation (processus de partition sensorielle) mais aussi des processus de nature différente de sélection de l'information, étayant la théorie de Bregman (1990, 1994).

Nous avons donc présenté à des auditeurs l'ensemble des mélodies présentées dans les expériences précédentes, intercalées avec une séquence distractrice choisie aléatoirement, dans chacune des neuf conditions de séparation en hauteur (0, 1, 2, 3, 4, 6, 8, 12 et 24 dt). Nous leur avons demandé de compter le nombre de sons qui se séparaient perceptivement de la séquence composite dans les basses fréquences. Ce nombre pouvait être égal à 0, 1, 2, 3, 4, 5 ou 6 puisque la séquence distractrice est composée de 6 notes. Si notre hypothèse est valide, les jugements concernant le nombre de sons qui se séparent devraient être identiques pour les degrés de séparation en hauteur pour lesquels on observe des performances constantes (paliers), et changer pour les degrés de séparation en hauteur pour lesquels on a observé des sauts de performances.

5.4.2 Méthode

5.4.2.1 Stimuli

Les caractéristiques des séquences étaient identiques à celles présentées précédemment. Seule la séquence composée de la mélodie intercalée avec une séquence distractrice était présentée à chaque essai.

5.4.2.2 Procédure

L'expérience s'est déroulée en deux sessions composées chacune de 324 essais. Elle comprenait donc 648 essais au total, correspondant à la combinaison de l'ensemble des 72 mélodies (36 versions originales et 36 versions modifiées) et des 9 degrés de séparation en hauteur (0, 1, 2, 3, 4, 6, 8, 12 et 24 dt). Une séquence distractrice parmi les 5 associées à chacune des 36 mélodies était choisie aléatoirement à chaque essai. La combinaison entre l'ensemble des mélodies et des degrés de séparation, était présentée à chacun des auditeurs. Cependant, du fait du choix aléatoire de la séquence distractrice, les participants n'entendaient pas nécessairement le même mélange pour une mélodie donnée à une différence de hauteur moyenne fixée (ce qui était également le cas dans les expériences précédentes). L'ensemble des mélodies a été partagé en deux blocs : les mélodies n°1 à 18 dans leurs versions originales et modifiées constituaient le premier bloc, et les mélodies n°19 à 36 dans leurs versions originales et modifiées le second. L'ordre de présentation de ces deux blocs de mélodies lors des deux sessions expérimentales correspondantes, a été contrebalancé sur l'ensemble des participants.

La tâche consistait à compter le nombre de sons qui se séparaient de la séquence composite dans les basses fréquences (annexe C7). Au début de l'expérience, 9 essais de démonstration correspondant aux 9 degrés de séparation en hauteur entre la mélodie et la séquence distractrice présentés au cours de l'expérience, étaient proposés afin de familiariser l'auditeur à la tâche (CD plages 58 à 66, annexe D). Les auditeurs répondaient à l'aide du clavier numérique un nombre compris entre 0 et 6, puis validaient leur réponse. La réponse 0 correspondait à la perception d'un flux de 12 notes (la séquence composite formée de la mélodie intercalée à la séquence distractrice), la réponse 6 à la perception de deux flux de 6 notes (la mélodie et la distractrice perçues distinctement), et les réponses intermédiaires à la fission partielle de sons distracteurs. Chaque session expérimentale durait environ 1h15. Deux pauses étaient aménagées respectivement après le premier et le second tiers des essais composant chaque bloc.

5.4.2.3 Participants

Cinq auditeurs ont participé à l'expérience, quatre collègues et moi-même. Il y avait au total deux femmes et trois hommes dont l'âge était compris entre 25 et 46 ans. Parmi eux, trois avaient reçu une formation musicale.

5.4.3 Résultats et discussion

Chaque participant a donc effectué 72 jugements de fission pour chaque degré de séparation en hauteur, correspondant aux 72 mélodies présentées. Au total sur les cinq participants, 360 jugements ont donc été réalisés par condition. La distribution du nombre de réponses données dans chaque catégorie de réponse (0, 1, 2, 3, 4, 5 ou 6 sons séparés) en fonction de la condition de séparation en hauteur (0, 1, 2, 3, 4, 6, 8, 12 et 24 dt) est portée sur l'histogramme ci-dessous (Figure 5.7).

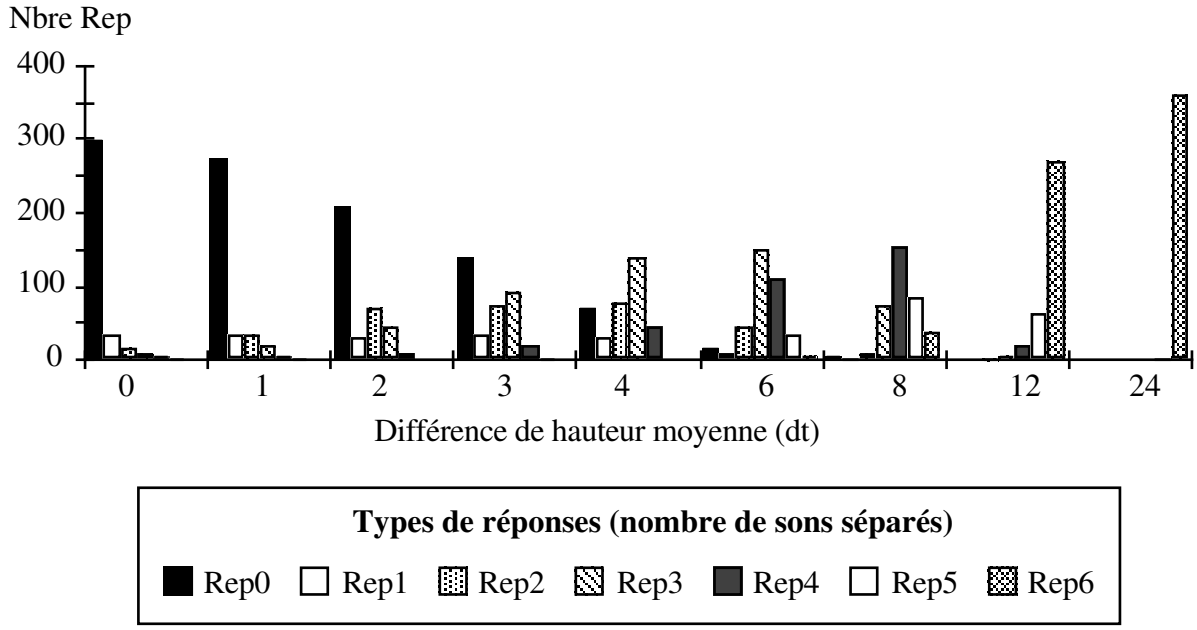


Figure 5.7. Distribution du nombre total de réponses "0" "1" "2" "3" "4" "5" et "6" données par les 5 auditeurs en fonction de la différence de hauteur moyenne séparant la mélodie et la séquence distractrice (0, 1, 2, 3, 4, 6, 8, 12 ou 24 dt).

D'après l'hypothèse que nous avons formulée, si la progression discontinue des performances observée dans la tâche de post-reconnaissance est liée à l'émergence des contours et intervalles de la mélodie suite à la fission perceptive des sons distracteurs, nous devrions observer que les jugements concernant le nombre de sons qui se séparent du "mélange" soient constants pour les différences de hauteur moyenne comprises entre 0 et 2 dt, 3 et 4 dt, 6 et 8 dt et 12 et 24 dt (correspondant aux paliers observés). Au contraire, ces jugements devraient être différents entre les conditions 2 et 3 dt, 4 et 6 dt et 8 et 12 dt (correspondant aux 3 sauts

de performances). Or si l'on regarde le mode de chaque distribution, c'est-à-dire les jugements les plus fréquemment donnés par condition de séparation, il apparaît que pour une différence de hauteur moyenne comprise entre 0 et 3 dt, les auditeurs répondent qu'aucun son ne se sépare, pour une différence comprise entre 4 et 6 dt que 3 sons se séparent, pour une différence de 8 dt que 4 sons se séparent et enfin pour une différence de 12 et 24 dt que 6 sons se séparent du mélange. Donc le changement de jugement ne se fait pas entre 2 et 3 dt mais entre 3 et 4 dt, pas non plus entre 4 et 6 dt mais entre 6 et 8 dt, en revanche on a bien comme cela était attendu un changement entre 8 et 12 dt.

Néanmoins, si l'on regarde de plus près les distributions, on constate que pour les réponses intermédiaires les jugements sont souvent partagés entre deux catégories de réponses adjacentes. Par exemple, pour une différence de hauteur moyenne de 3 dt, sur les 360 réponses 140 sont des réponses "0", mais il y a également 73 réponses "2" et 91 réponses "3" soit au total 164 réponses "2" et "3". Le fait que les distributions soient bimodales dans certaines conditions de séparation en hauteur, pourrait s'expliquer : 1) par le fait que le nombre de sons distracteurs qui se scindent de la séquence composite pour un degré de séparation en hauteur donné, varie en fonction des mélodies présentées puisqu'elles ont des ambitus différents, et/ou 2) par l'incertitude des jugements, la séquence composite n'étant présentée qu'une seule fois. Par conséquent, afin de prendre en compte ce recouvrement dans les catégories de réponses intermédiaires, nous avons construit une distribution de réponses cumulées. Nous avons sommé l'effectif affecté aux réponses connexes pour l'ensemble des réponses intermédiaires et l'ensemble des conditions de séparation en hauteur, afin de ne pas avoir d'a priori sur les réponses à cumuler (Figure 5.8).

Les modes des différentes distributions observées pour chacune des conditions de séparation en hauteur sont alors respectivement égal à "0" dans les conditions 0, 1 et 2 dt, "2 et 3" dans les conditions 3 et 4 dt (on observe également un nombre assez important de réponses "0" dans la condition 2 dt), "3 et 4" dans la condition 6 dt, "3 et 4" et "4 et 5" dans la condition 8 dt, et enfin "6" dans les conditions 12 et 24 dt.

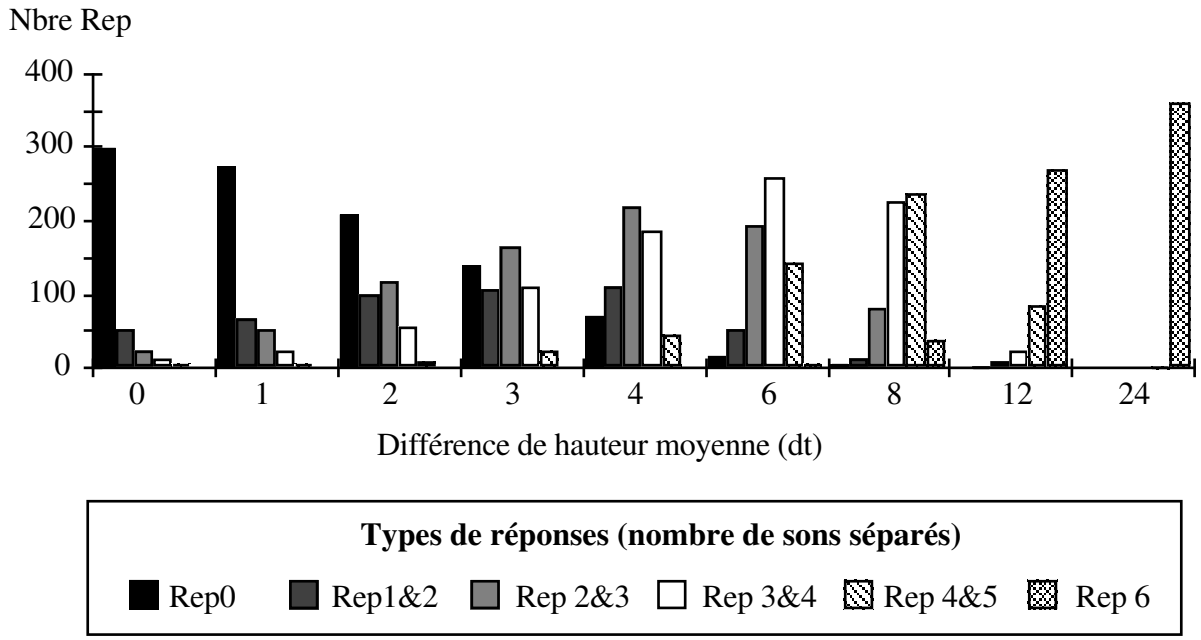


Figure 5.8. Distribution du nombre de réponses "0", "1" et "2" cumulées, "3" et "4" cumulées, "4" et "5" cumulées et "6" en fonction de la différence de hauteur moyenne séparant la mélodie et la séquence distractrice (0, 1, 2, 3, 4, 6, 8, 12 ou 24 dt).

Je voudrais faire deux remarques à propos de cette distribution de réponses cumulées. La première concerne le fait que pour une différence de 3 dt les auditeurs répondent majoritairement que 2 et 3 sons se séparent de la séquence composite, mais également pour une quantité d'essais non négligeable qu'aucun son ne se scinde (140 réponses "0" et 164 réponses "2 et 3"). Cette bimodalité peut s'expliquer par le fait qu'une différence de 3 dt en moyenne à un tempo de 6 sons/sec correspond à la zone ambiguë définie par van Noorden (1975) dans laquelle il est possible de séparer ou bien d'intégrer les sons d'une séquence. Or la tâche de compter le nombre de sons qui se séparent dans les basses fréquences, même s'il s'agit bien d'une tâche de fission comme la tâche de fission mélodique, constitue une tâche moins sélective au niveau attentionnel que la tâche de focaliser sur la mélodie. En effet, la focalisation sélective est moins aisée dans les basses fréquences car le registre sur lequel on doit focaliser notre attention varie à chaque essai (le registre de la séquence distractrice fluctuant à chaque essai). En revanche, lorsque l'on écoute sélectivement la mélodie, le registre dans lequel elle apparaît est fixe et facilite ainsi la focalisation. Les auditeurs ont d'ailleurs reporté avoir eu parfois une écoute globale pour écouter les sons qui se scindaient. La deuxième remarque concerne le léger déplacement de mode observé entre 6 et 8 dt (on passe de réponses "3 et 4" à des réponses "3 et 4" et "4 et 5"). Pourtant pour ces degrés de séparation en hauteur, les performances de post-reconnaissance (figure 5.2) demeurent constantes. Ce léger déplacement de mode passant de 3.5 sons séparés en moyenne à 4 n'est probablement pas suffisant pour induire un changement de performances, inférieur à 1 en moyenne il n'y a pas de lien supplémentaire entre les sons de la mélodie qui émerge.

On peut donc considérer que les principaux changements de mode se produisent entre les conditions 2 et 3 dt (on passe d'une réponse "0" à une réponse "2 et 3" sons séparés), 4 et 6 dt (on passe d'une réponse "2 et 3" à une réponse en moyenne de 4 sons séparés), et 8 et 12 dt (on passe de 4 sons en moyenne à 6 sons séparés), ce qui correspond aux sauts de performances observés dans la tâche de post-reconnaissance de mélodies intercalées (figure 5.9).

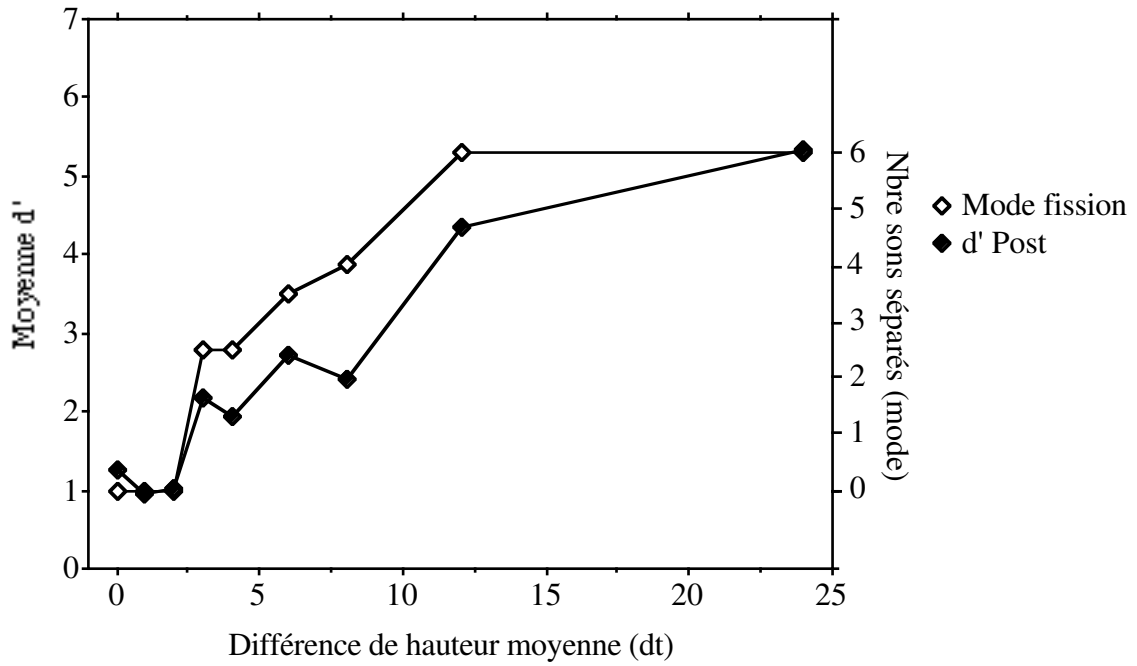


Figure 5.9. Représentation de la progression des performances recueillies dans la condition post de l'expérience 5 (moyenne des d' obtenus par 10 auditeurs, courbe avec les losanges pleins), et du nombre de sons jugés se séparer dans les basses fréquences (mode de la distribution des réponses cumulées obtenues par cinq auditeurs, courbe avec les losanges vides) en fonction de la différence de hauteur moyenne séparant la mélodie cible de la séquence distractrice.

Les discontinuités observées dans la progression des performances de post-reconnaissance de mélodies intercalées semblent donc être liées au nombre de sons distracteurs qui se séparent de la séquence composite. La fission des sons distracteurs ferait émerger les contours et intervalles de la mélodie, et augmenterait ainsi la probabilité de détecter un changement (les mélodies qui diffèrent ayant au moins un changement de contour).

5.5 Discussion générale

5.5.1 Résumé des principaux résultats

Afin d'examiner la contribution des connaissances dans l'organisation perceptive d'une séquence sonore, nous avons comparé les performances de reconnaissance d'une mélodie cible intercalée à des sons distracteurs obtenues dans deux situations expérimentales différentes (expérience 5). Dans la première situation, la mélodie à reconnaître était présentée après la séquence à organiser, il s'agissait d'une tâche de post-reconnaissance de mélodies intercalées. Dans la deuxième situation (condition pré), la mélodie de référence était présentée avant la séquence composite de sorte que l'auditeur connaissait la mélodie qu'il devait extraire du "mélange". Les résultats ont révélé que dans ces deux tâches, l'aptitude des auditeurs à reconnaître la mélodie cible entremêlée ainsi que le critère sur lequel ils ont basé leurs jugements, étaient différents.

Lorsque les auditeurs disposaient d'une représentation précise de la mélodie qu'ils devaient rechercher dans le mélange, les performances de reconnaissance étaient accrues par rapport à celles obtenues par des auditeurs dont les connaissances préalables se limitaient à la connaissance du registre dans lequel cette mélodie était présentée. Cette amélioration des performances s'est traduite à la fois par la possibilité d'extraire la mélodie pour une différence de hauteur moyenne plus petite, et par une augmentation globale des performances pour l'ensemble des degrés de séparation en hauteur (à l'exception du degré 0). Cette augmentation globale des performances suggère l'intervention d'un autre mécanisme, différent du processus de partition sensorielle. Deux hypothèses ont été formulées sur la nature de ce processus. Il pourrait s'agir d'un processus d'analyse de scène guidée par les connaissances (Bregman, 1990, 1994). La préactivation de la mélodie permettrait de sélectionner les notes de celle-ci au niveau de la représentation sensorielle de la séquence composite, améliorant ainsi son contraste sans pour autant la séparer perceptivement du mélange. Néanmoins, l'amélioration des performances pourrait également s'expliquer par des différences mnésiques inhérentes à ces deux situations expérimentales (Bregman, 1997). En effet, alors que dans la condition pré la première mélodie présentée est encodée seule, dans la condition post elle est encodée avec la séquence distractrice. L'encodage simultané de la mélodie et de la séquence distractrice en mémoire de travail pourrait dégrader la représentation que les auditeurs se construisent de la mélodie (Deutsch, 1970 ; Dowling et al., 1995), et ainsi baisser les performances de reconnaissance de façon générale.

Les auditeurs qui connaissaient précisément la mélodie à extraire de la séquence composite (pré) ont eu tendance à répondre plus souvent que les mélodies étaient différentes, surtout lorsque la mélodie n'était pas séparée de la séquence distractrice. Cette différence de critère de réponse pourrait être liée à la différence de précision du codage de la première mélodie entendue dans ces deux tâches. En effet, ces deux tâches consistent à comparer une première

mélodie, qui constitue la mélodie de référence, à une deuxième mélodie présentée ultérieurement. Dans la condition pré, l'auditeur peut se construire une représentation précise de cette mélodie de référence (notes, contour...), et ainsi générer des attentes fréquentielles et temporelles sur la mélodie qui est présentée ensuite (Dowling et al., 1987). La confrontation entre ses attentes précises et la représentation "brouillée" de cette deuxième mélodie (du fait qu'elle soit présentée simultanément avec une autre séquence avec laquelle elle est entremêlée), pourrait le conduire à surestimer la différence existant entre la première et la deuxième mélodie. En revanche, dans la condition post l'auditeur commence par entendre la mélodie intercalée aux sons distracteurs. Il se construit donc une représentation plus floue de la mélodie de référence, et la fenêtre d'attente est alors probablement plus large. Ce doute pourrait le conduire à ne pas rejeter systématiquement la deuxième mélodie présentée comme étant différente de la première.

Un argument en faveur de l'hypothèse selon laquelle la tâche de reconnaissance de mélodies intercalées avec présentation préalable de la mélodie ferait intervenir des processus d'analyse guidée par les connaissances, a été fourni par l'expérience suivante. Dans l'expérience 6, une tâche semblable mais dans laquelle la première mélodie avait été transposée, a été proposée aux auditeurs. Les performances de reconnaissance de la mélodie cible intercalée se sont avérées intermédiaires à celles obtenues précédemment dans les conditions pré et post. Ce résultat a ainsi révélé que le gain de performances observé dans la condition pré pouvait s'expliquer en partie par un amorçage fréquentiel. La préactivation de la fréquence de chacun des sons composant la mélodie a permis d'améliorer sa reconnaissance dans un mélange. Cependant, les performances obtenues dans cette tâche de reconnaissance de mélodies intercalées avec transposition, demeuraient supérieures à celles obtenues en post-reconnaissance. Deux hypothèses ont été avancées pour expliquer ce gain résiduel. Il pourrait s'expliquer par l'intervention d'autres connaissances, probablement liées au contour (Dowling & Fujitani, 1971), qui faciliteraient la reconnaissance de la mélodie par rapport à la tâche de post-reconnaissance. Il pourrait également être lié à la différence de précision de l'encodage de la mélodie de référence dans les deux situations expérimentales, comme celle constatée entre les conditions pré/post. D'ailleurs, le profil des jugements des auditeurs dans les deux tâches dans lesquelles la mélodie était présentée avant le mélange (pré et prétransposé), était similaire, constituant une sorte de "signature" de cette situation expérimentale et apparentant ces deux tâches en tout point.

Enfin, la dernière expérience a permis d'émettre une hypothèse sur la façon dont ces connaissances avaient pu intervenir dans l'analyse perceptive de la séquence composite. Les résultats de l'expérience 7 révèlent en effet que les discontinuités observées dans la progression des performances de post-reconnaissance semblent correspondre au nombre de sons distracteurs qui se séparent de la séquence composite. Ainsi, à mesure que la différence de hauteur moyenne s'accroît entre la mélodie et la séquence distractrice, certains sons se scindent

faisant émerger le contour et les intervalles de la mélodie, et mettant ainsi à la disposition de l'auditeur un indice supplémentaire pour pouvoir accomplir la tâche. Compte tenu de l'absence de discontinuités dans la progression des performances de reconnaissance lorsque la mélodie est présentée avant la séquence composite qu'elle soit ou non transposée, ce résultat suggère que les processus d'analyse de scène guidée par les connaissances qui conduisent au moins en partie à l'amélioration des performances ne sont pas des processus de fission mais probablement des processus de sélection de l'information (Bregman, 1990, 1994). Le rapprochement des résultats obtenus dans les expériences 5 et 7 conforte l'idée que ces processus permettraient d'accéder à la mélodie cible sans pour autant la séparer perceptivement de la séquence distractrice. En effet, dans l'expérience 7, les auditeurs ont jugé que pour une différence de hauteur moyenne de 2 dt entre la mélodie et la séquence distractrice, globalement aucun son ne se séparait du mélange. Pour cette même différence de 2 dt, alors que les auditeurs ne parvenaient pas à reconnaître la mélodie entremêlée dans la condition post (les proportions moyennes de reconnaissance correcte correspondant à un d' égal à 1 étaient de .57, inférieures à .58 qui est la valeur atypique d'une distribution binomiale à .01 pour n = 240), ils réussissaient à reconnaître 73 % des mélodies dans la condition pré (d' proche de 3).

Les résultats de cette série d'expériences constituent des arguments expérimentaux qui étayent la théorie de Bregman (1990, 1994) concernant l'existence de deux processus distincts : un processus primaire d'analyse de scène auditive qui partitionne l'entrée sensorielle, et un processus d'analyse de scène guidée par les connaissances qui sélectionne l'information associée au schéma activé. En outre, ces résultats soulèvent des questions quant aux relations que ces deux mécanismes entretiennent, et conduisent ainsi à formuler de nouvelles hypothèses sur le fonctionnement de l'analyse de la scène auditive.

5.5.2 Nouvelles hypothèses théoriques sur le fonctionnement de l'analyse de la scène auditive

Selon Bregman (1990, 1994), l'analyse primaire et l'analyse guidée par les schémas opéreraient conjointement dans la construction de la scène auditive. Mais comment le mécanisme de partition sensorielle qui construit des entités perceptives (les sons et les flux auditifs), et le mécanisme de sélection de l'information qui apparie les schémas activés à la représentation sensorielle, fonctionnent-ils ensemble ? Opèrent-ils de façon indépendante sur la représentation sensorielle ou interagissent-ils pour construire les flux auditifs ?

A la lumière des résultats de ces expériences, nous pouvons formuler certaines hypothèses sur les relations qu'entretiennent ces deux mécanismes.

Tout d'abord, il semble que l'analyse primaire de la scène auditive, qui est le mécanisme même de construction des entités perceptives indépendantes que sont les flux auditifs, ne soit

pas totalement "encapsulée" (Fodor, 1986). On aurait pourtant pu imaginer que ce mécanisme pré-attentif (Sussman et al., 1999), automatique, et précâblé (Demany, 1982 ; McAdams & Bertoncini, 1997) fonctionne comme un module autonome. Cependant, nous avons pu voir (expérience 5), que la préactivation de la mélodie avait pour effet de modifier la fission mélodique. Ce résultat avait également été mis en évidence par van Noorden (1975) dont les recherches ont révélé l'existence d'une zone ambiguë dans laquelle le nombre de flux que l'auditeur percevait dépendait de ce qu'il cherchait à entendre. Cette zone de flexibilité, dans laquelle les processus d'analyse guidée par les connaissances peuvent intervenir et modifier l'analyse primaire, présente cependant des limites. Il existe en effet des conditions (différence de fréquences et tempo) pour lesquelles ce que l'auditeur entend ne dépend pas de ce qu'il cherche à entendre. Par exemple, pour une différence de fréquence inférieure à 2 dt, même si l'auditeur essaie de séparer perceptivement les sons aigu et grave composant une séquence alternante, il n'y parviendra pas (van Noorden, 1975, expériences 1 et 2). Nous avons également constaté une limite dans l'intervention des connaissances sur la fission mélodique (expérience 5). En effet, les performances obtenues dans les conditions pré et post étaient équivalentes lorsque la mélodie cible et la séquence distractrice étaient présentées dans le même registre de hauteur (différence de hauteur moyenne nulle), ou distante de 1 dt. Si ce résultat est en accord avec l'existence de limites postulées sur la base des résultats de l'étude de van Noorden (1975), il s'avère cependant différent de celui trouvé par Dowling (Dowling, 1973, expérience 3 ; Dowling et al., 1987, expérience 1). En effet, W. J. Dowling a constaté que des mélodies familières intercalées pouvaient être reconnues même lorsqu'elles étaient présentées dans le même registre de hauteur. Ceci signifie que le processus d'analyse de scène guidée par les connaissances pourrait modifier l'analyse primaire effectuée pour une différence de hauteur moyenne nulle. Nous avons discuté ce résultat dans la section 5.2.3 et formulé l'hypothèse que la puissance de ce mécanisme descendant pourrait dépendre de la nature des schémas activés. Des informations stockées en mémoire à long terme, comme les intervalles, pourraient si elles sont activées permettre d'accéder à la mélodie même lorsque la séquence composite est perçue comme un seul flux. En revanche, les notes et/ou le contour stockés en mémoire à court terme ne permettraient pas cette analyse. Cependant, cette hypothèse nécessite de s'assurer auparavant que dans ces séquences formées par des mélodies familières intercalées, aucun son ne se sépare du "mélange". En effet, il faut contrôler qu'aucune partie de la mélodie ne puisse émerger de la séquence composite qui pourrait permettre à l'auditeur d'activer en mémoire cet extrait musical.

Les connaissances peuvent donc modifier l'analyse primaire de scène auditive mais elles peuvent également intervenir par le biais d'un mécanisme d'une autre nature en sélectionnant l'information sensorielle associée au schéma activé, processus que Bregman (1990, 1994) appelle l'analyse de scène auditive guidée par les connaissances. La contribution de ce processus pourrait varier en fonction de l'analyse primaire réalisée par le système auditif. En effet, l'expérience 5 montre que la différence de performances obtenues dans les conditions pré et post est d'autant plus importante que la différence de hauteur moyenne s'accroît (jusqu'à une

différence de 12 dt, car ensuite les performances dans la condition pré plafonnent). L'amplification de ce gain suggère que ces deux processus interagissent dans l'analyse de la scène auditive. Ainsi, nous pouvons faire l'hypothèse que la contribution des processus d'analyse guidée par les connaissances serait fonction du degré d'appariement entre le schéma et la représentation sensorielle. Ce degré d'appariement étant d'autant plus important que la mélodie émerge de la séquence composite, ce processus descendant aurait pour effet d'amplifier la ségrégation perceptive opérée par le processus ascendant.

5.6 Conclusion

Cette série d'expériences révèle l'importance des processus d'analyse de scène auditive guidée par les connaissances dans la formation des flux auditifs. Le système auditif est doté d'un mécanisme général, issu du couplage entre notre espèce et son environnement, lui permettant de construire des descriptions distinctes des différentes sources sonores de l'environnement (Bregman, 1990 ; Shepard, 1981). La combinaison entre cette analyse primaire et les processus d'analyse guidée par les schémas confère au système cognitif une adaptation optimale à son milieu. L'analyse guidée par les connaissances apporte un gain en termes de lisibilité de la scène sonore, mais probablement également dans la rapidité avec laquelle l'auditeur est capable de reconnaître et identifier les objets de son environnement. Elle permet également de modifier dans une certaine mesure cette analyse primaire, et ainsi réorienter l'interprétation de notre environnement sonore en faveur de la construction d'"objets" familiers, souvent rencontrés par l'individu.

En conclusion, je voudrais souligner que les résultats obtenus en contrastant ces deux conditions nommées pré et post, dépassent probablement le cadre de l'étude des processus d'organisation auditive. En effet, la mise en évidence de l'implication de processus distincts et de stratégies différentes dans des situations expérimentales qui sont similaires en tout point où seul l'ordre de présentation des deux intervalles est modifié, pourrait avoir des conséquences dans d'autres études de perception auditive utilisant des paradigmes de discrimination ou de détection avec deux intervalles. Ce résultat met l'accent sur l'importance de la séquentialité des événements dans la perception auditive.

Chapitre 6

Déficit de la reconnaissance de mélodies intercalées : Etude de patients cérébro-lésés

Les bases neurales des mécanismes impliqués dans l'analyse de la scène auditive sont actuellement totalement inconnues. Nous avons tenté d'apporter des éléments de réponse à cette question encore peu étudiée, en examinant la capacité de patients cérébro-lésés à reconnaître une mélodie intercalée à des sons distracteurs. Deux études exploratoires ont été menées et révèlent : 1) la possible existence d'un phénomène d'interférence accrû en écoute polyphonique chez des patients épileptiques présentant une atrophie hippocampique droite ou gauche, et 2) une dissociation dans les tâches de reconnaissance et post-reconnaissance de mélodies intercalées chez un patient présentant une lésion temporale droite. Ces premiers résultats ouvrent des perspectives intéressantes pour l'exploration neuropsychologique des processus de formation des flux auditifs.

6.1 Introduction

Les corrélats neuronaux des mécanismes impliqués dans la formation des flux auditifs sont actuellement totalement inconnus. A ma connaissance, aucune étude d'imagerie cérébrale fonctionnelle n'a encore été menée, et seules deux recherches se sont intéressées aux déficits éventuels de patients cérébro-lésés dans l'analyse de la scène auditive. Ces études ont été décrites dans la section 2.3 de ce document, nous en donnons ici un bref résumé.

Efron et al. (1983) ont mis en évidence la difficulté de patients ayant subi une lobectomie temporale antérieure (droite ou gauche) épargnant les gyri de Heschl, à extraire et localiser un son de l'environnement familier présenté controlatéralement à la résection, lorsque celui-ci était diffusé simultanément avec quatre autres sons. Les résultats de cette étude ne permettent malheureusement pas de distinguer un déficit lié à l'extraction de sons dans une "mixture" sonore, de celui lié à la localisation et à l'identification de ces sons. Dans le cadre d'une recherche examinant les déficits de la perception auditive non verbale de patients cérébro-lésés, Clarke et ses collaborateurs (1996) ont procédé à une évaluation sommaire des processus d'organisation auditive à la fois simultanée et séquentielle. Sur les vingt patients étudiés, dont le site et l'origine des lésions étaient très variés, huit présentaient un déficit dans la capacité de séparer perceptivement les sons. Parmi ces huit patients, six avaient une lésion incluant le lobe temporal gauche. Néanmoins, les auteurs ont souligné la difficulté clinique d'évaluer les processus d'organisation auditive de façon directe, en demandant à l'auditeur de signaler le nombre de sons ou de flux perçus. Ces deux études ne permettent donc pas d'établir avec certitude l'existence d'une déficience dans la capacité de séparer les émissions des différentes sources sonores de l'environnement. En outre, les structures cérébrales nécessaires au fonctionnement de ces mécanismes n'apparaissent pas clairement.

Afin de rechercher d'éventuels déficits de la construction des flux auditifs suite à une atteinte cérébrale, nous avons soumis à des patients cérébro-lésés les paradigmes de reconnaissance et/ou post-reconnaissance de mélodies intercalées. Deux études exploratoires ont été menées.

La première avait pour objectif d'examiner les performances de post-reconnaissance de mélodies intercalées d'un groupe de patients épileptiques avant et après l'opération chirurgicale consistant en une résection d'une partie plus ou moins importante du lobe temporal. Les patients présentant des lésions ou résections au niveau des régions temporales, impliquées dans le traitement de l'information auditive (Peretz, 1994 ; Zatorre, 1988a, pour des revues), se plaignent en effet souvent d'avoir des difficultés à suivre une conversation dans du bruit. Une cause possible à ce trouble pourrait provenir d'une difficulté à séparer les sons sur la base de leur différence de hauteur ou de timbre, donc d'un déficit dans la formation des flux auditifs. Malheureusement, nous n'avons pas pu tester les patients après cette résection. Nous présenterons donc uniquement les résultats obtenus par des patients épileptiques testés avant l'opération qui présentaient une atrophie hippocampique droite ou gauche et un hypométabolisme dans les régions temporales et frontales adjacentes. Cette étude a été menée en collaboration avec le professeur S. Samson (Hôpital de la Salpêtrière, Paris).

La deuxième étude a été conduite auprès d'un patient, L.B, dont l'accident cérébro-vasculaire avait provoqué des lésions temporales droites touchant les aires auditives associatives et probablement les circuits reliant l'aire primaire. Nous lui avons soumis deux tâches de reconnaissance de mélodies intercalées avec et sans présentation préalable de la mélodie à

reconnaître (conditions pré et post, voir chap. 5). Cette étude s'intégrait dans une évaluation plus globale des troubles centraux de ce patient, menée par C. Lorenzi (Laboratoire de Psychologie Expérimentale, Université Paris V) en collaboration avec B. Frachet et C. Belin (Hôpital Avicenne, Bobigny).

6.2 Post-reconnaissance de mélodies intercalées : Étude d'un groupe de patients cérébro-lésés

6.2.1 Introduction

Cette première étude a donc exploré la capacité de patients épileptiques, présentant des dommages cérébraux au niveau de l'hippocampe et des structures temporales et frontales adjacentes, à séparer les émissions des différentes sources sonores de l'environnement.

Une tâche de post-reconnaissance de mélodies intercalées dans laquelle la différence de hauteur moyenne entre la mélodie et la séquence distractrice variait, a été proposée à deux groupes de patients présentant respectivement une atrophie hippocampique droite ou gauche. La passation de témoins appariés suivant des critères d'âge et d'éducation étant actuellement en cours, nous avons comparé les performances obtenues à celles des participants de l'expérience 2 qui ne sont pas appariés selon ces critères (voir §4.4). Cette comparaison constitue une première approximation et ne nous permet pas de conclure avec certitude. Cependant, il faut noter que les performances obtenues dans ce type de tâche sont apparues très stables dans l'ensemble des expériences menées (expériences préliminaires et celles reportées dans la thèse) incluant une population relativement variée.

6.2.2 Méthode

6.2.2.1 Patients

Dix-huit patients atteints depuis l'enfance d'une épilepsie rebelle à tout traitement médicamenteux, ont participé à cette étude. Ces patients présentaient une atrophie hippocampique associée à un hypométabolisme dans les structures adjacentes temporales et parfois frontales. Neuf patients présentaient une atrophie hippocampique droite (AHD), 7 femmes et 2 hommes, d'un âge moyen de 39 ans, ayant en moyenne 12 années d'éducation et dont le QI était compris entre 80 et 108. Les neuf autres avaient une atrophie hippocampique gauche (AHG), 5 femmes et 4 hommes, d'un âge moyen de 37 ans, ayant en moyenne 10 années d'éducation et dont le QI était compris entre 80 et 114 (tableau 6.1). Ils ne souffraient pas d'aphasie, ni d'hémiplégie. L'ensemble de ces patients avaient le langage localisé dans l'hémisphère gauche d'après le test de l'injection intracarotidienne d'amytal sodique (Wada &

Rasmussen, 1960), et les réponses à un questionnaire de préférence manuelle. Tous étaient non musiciens.

Tous les patients ont donné leur consentement éclairé par écrit, et cette étude réalisée sous le contrôle médical du neurologue S. Bakchine a été approuvée par le comité d'éthique.

Patients	Nombre	Sexe	Age	Latéralité	Éducation	QI
AHD	9	7 F et 2 H	39 ans [32 ; 56]	Droitier	12 années [5 ; 17]	94 [80 ; 108]
AHG	9	5 F et 4 H	37 ans [17 ; 52]	Droitier	10 années [5 ; 15]	89 [80 ; 114]

Tableau 6.1. Caractéristiques des deux groupes de patients épileptiques, les patients présentant une atrophie hippocampique droite (AHD) et ceux présentant une atrophie hippocampique gauche (AHG). Pour l'âge, le nombre d'années d'éducation et le QI (WAIS-R) (Wechsler, 1981), la moyenne et la gamme de variation sont mentionnées pour chaque groupe.

6.2.2.2 Matériel et procédure

Une tâche de post-reconnaissance de mélodies intercalées dans laquelle la différence de hauteur moyenne entre la mélodie cible et la séquence distractrice variait, a été proposée aux deux groupes de patients. Cette tâche était identique à celle présentée dans l'expérience 2 (voir §4.2 pour des détails sur la méthode). Quatre degrés de séparation en hauteur moyenne ont été présentés. La distractrice était présentée dans le même registre que la mélodie cible, ou s'en éloignait en moyenne de 6, 12 ou 24 demi-tons (dt) vers les fréquences basses. Dans chacune des quatre conditions de séparation en hauteur 32 essais étaient proposés, soit au total 128 essais. Cette condition expérimentale était suivie d'une condition contrôle composée de 32 essais. Dans chacun de ces deux tests, la moitié des mélodies était identique, l'autre moitié différait par deux notes. La différence de hauteur moyenne entre la mélodie et la distractrice était distribuée aléatoirement sur l'ensemble des essais, ainsi que le type d'essai "même" et "différent". Les séquences étaient composées de sons purs.

Les patients ont donc été soumis à deux tests successifs : la tâche de post-reconnaissance de mélodies intercalées (condition expérimentale) et la tâche de reconnaissance de mélodies (condition contrôle). Chacun de ces tests était précédé de quelques essais de familiarisation, au nombre de 20 de difficulté croissante dans la condition expérimentale et de 8 dans la condition contrôle, au cours desquels la réponse correcte était donnée aux auditeurs une fois leur jugement effectué. Les tests ont été enregistrés sur une cassette DAT, et diffusés sous casque dans un endroit calme. La passation s'est déroulée dans l'enceinte de l'Hôpital de la Salpêtrière, elle durait environ 1 heure. La consigne orale donnée aux patients est reportée dans l'annexe C8.

6.2.3 Résultats et discussion

Sur les dix-huit patients ayant participé à cette étude, six ne sont pas parvenus à accomplir cette tâche de post-reconnaissance de mélodies intercalées. Parmi ces patients, quatre présentaient une AHG et deux une AHD. Il se trouve qu'ils répondaient au hasard également dans la condition contrôle qui consistait en une tâche de reconnaissance immédiate de mélodies présentées sans séquence distractrice. Ce résultat suggère que ces patients manifestaient des difficultés à discriminer des mélodies non familières, indépendamment du fait qu'elles soient présentées ou non dans un "mélange". En outre, parmi ces six patients, quatre avaient un empan mnésique faible (empan digital égal à 4), suggérant que leurs difficultés pourraient provenir d'un déficit général de la mémoire à court terme.

Nous avons donc écarté les résultats de ces six patients, et pris en compte les performances des douze autres. Les proportions moyennes de réponses correctes obtenues par 7 patients présentant une AHD (7 patients * 32 observations par condition de séparation), et 5 ayant une AHG (5 patients * 32 observations par condition), sont donc reportées sur la figure 6.1. Les performances des participants de l'expérience 2 sont également portées sur le graphique pour comparaison (16 participants * 8 observations par condition).

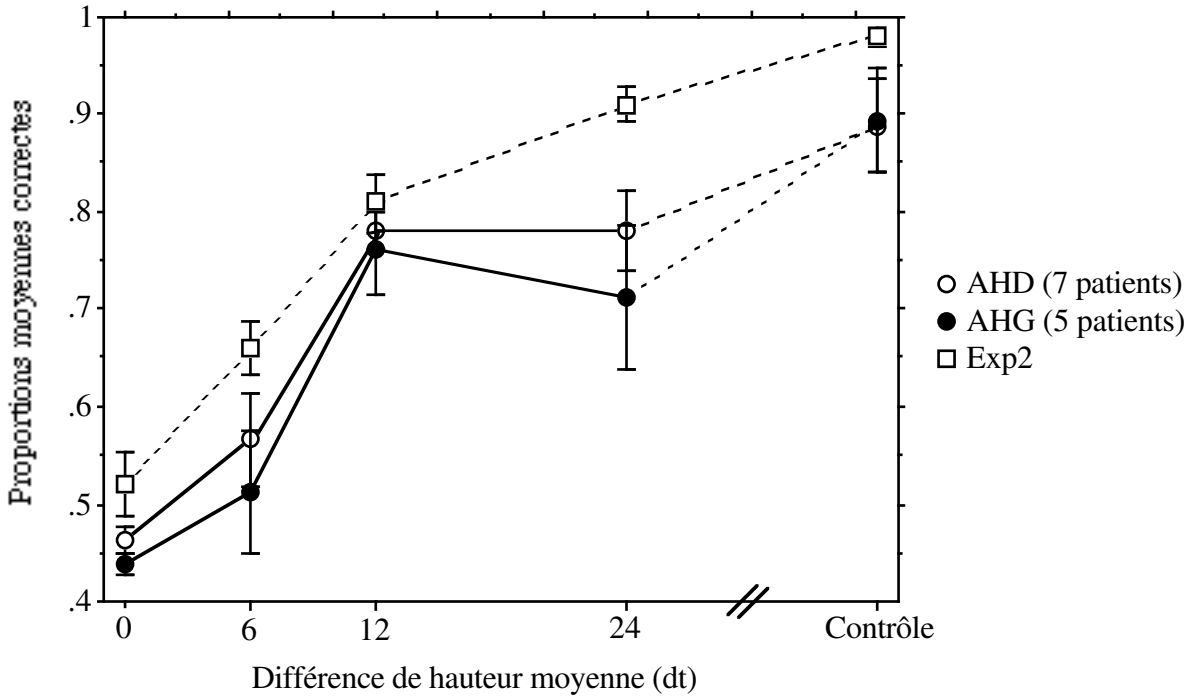


Figure 6.1. Proportions moyennes de réponses correctes obtenues par 7 patients présentant une AHD et 5 patients ayant une AHG dans la tâche de post-reconnaissance de mélodies intercalées, pour les différentes conditions de séparation en hauteur (0, 6, 12 ou 24 dt) et la condition contrôle. Les performances des participants de l'expérience 2 sont portées sur le graphique pour comparaison (ligne brisée). Les barres verticales représentent ± 1 erreur type.

Globalement, les performances de post-reconnaissance de mélodies intercalées obtenues par les patients présentant une atrophie hippocampique droite ou gauche étaient équivalentes, et elles étaient comparables à celles obtenues par les participants de l'expérience 2. Elles augmentaient avec la différence de hauteur moyenne séparant la mélodie cible et la séquence distractrice, suggérant que les patients parvenaient à extraire la mélodie du mélange et à réaliser cette tâche de discrimination mélodique. Cependant, alors que les auditeurs ne présentant pas de troubles neurologiques avaient des performances qui continuaient à progresser au-delà d'une différence moyenne de 12 dt, celles des patients plafonnaient. Cette différence de performances entre les patients et les participants de l'expérience 2 pour un degré de séparation de 24 dt, demande à être confirmée avec un groupe témoin apparié. Elle pourrait être liée à une différence de degré d'expertise. En effet, parmi les seize participants de l'expérience 2, dix avaient reçu une formation musicale, alors que les patients étaient tous non musiciens. Or nous avons vu dans l'expérience 2, que les performances de post-reconnaissance de mélodies intercalées obtenues par les musiciens étaient significativement supérieures à celles des non musiciens. Elle pourrait également provenir de facteurs socioculturels, ce groupe n'étant pas apparié aux patients sur les critères d'éducation et de QI.

Pour un degré de séparation de 24 dt, les performances obtenues par les deux groupes de patients étaient inférieures à celles recueillies dans la condition contrôle. Cette altération des performances de reconnaissance de la mélodie cible en présence de la séquence distractrice, même lorsque l'auditeur organise la séquence composite en deux flux distincts, a été constatée également chez les auditeurs sains pour l'ensemble des expériences de post-reconnaissance menées dans le cadre de cette thèse (voir les résultats des expériences 1 à 5). Cependant, l'interférence induite par la présence de la distractrice sur la reconnaissance de la mélodie cible, pourrait être plus importante chez les patients présentant une atrophie hippocampique droite ou gauche. En effet, alors que la différence de performances obtenue dans la condition de séparation de 24 dt et dans la condition contrôle est de 0.07 (± 0.09) pour les participants de l'expérience 2, elle est de 0.11 (± 0.17) chez les patients ayant une AHD et de 0.18 (± 0.15) chez les patients ayant une AHG.

Deux hypothèses ont été formulées dans les précédents chapitres pour rendre compte de cet effet d'interférence. Il pourrait s'expliquer par une distractibilité attentionnelle. En présence de la distractrice, l'auditeur ne parviendrait pas à maintenir sont attention de façon permanente sur le flux cible. Cette hypothèse d'un déficit attentionnel est cependant affaiblie par le fait que l'effet d'interférence se retrouve pour l'ensemble des auditeurs. La deuxième hypothèse qui a été formulée postulait que l'interférence pourrait être de nature mnésique. Or, Zatorre (1985) mentionne que des résections temporales droite ou gauche touchant des structures telles que l'hippocampe et l'amygdale, s'accompagent de troubles mnésiques comme l'attestent les déficits dans des tâches de mémorisation incidente de mélodies. Par conséquent, ces patients présentant des lésions au niveau de l'hippocampe et des structures temporales adjacentes pourraient

manifester un déficit mnésique spécifique à l'écoute polyphonique où deux séquences sont encodées simultanément. Plusieurs études ont révélé l'existence d'une interférence en mémoire de travail lorsque des sons interférents étaient présentés durant la période de rétention (Deutsch, 1970, 1972 ; Semal & Demany, 1991, 1993 ; Semal, Demany, Ueda & Hallé, 1996). L'interférence en mémoire constatée dans cette tâche de post-reconnaissance de mélodies intercalées, ne se produirait pas durant le stockage, mais durant l'encodage de l'information (voir la section 3.3.2 pour une discussion de cet aspect). Cette piste de recherche demande à être explorée sur un échantillon plus important.

Les patients présentant une atrophie hippocampique gauche apparaissent un peu plus affectés dans cette tâche que les patients dont la lésion se situe dans l'hémisphère droit (en particulier pour les conditions de séparation 6 et 24 dt). En outre, les proportions de détections correctes et de fausses alarmes calculées sur ces deux groupes de patients, révèlent qu'ils ont également des profils de jugement différents (figures 6.2 et 6.3). Lorsque la mélodie et la séquence distractrice ne sont pas séparées perceptivement (différence de hauteur moyenne nulle), les patients présentant une AHD répondent plus souvent que les deux mélodies présentées sont différentes, que les patients qui ont une AHG.

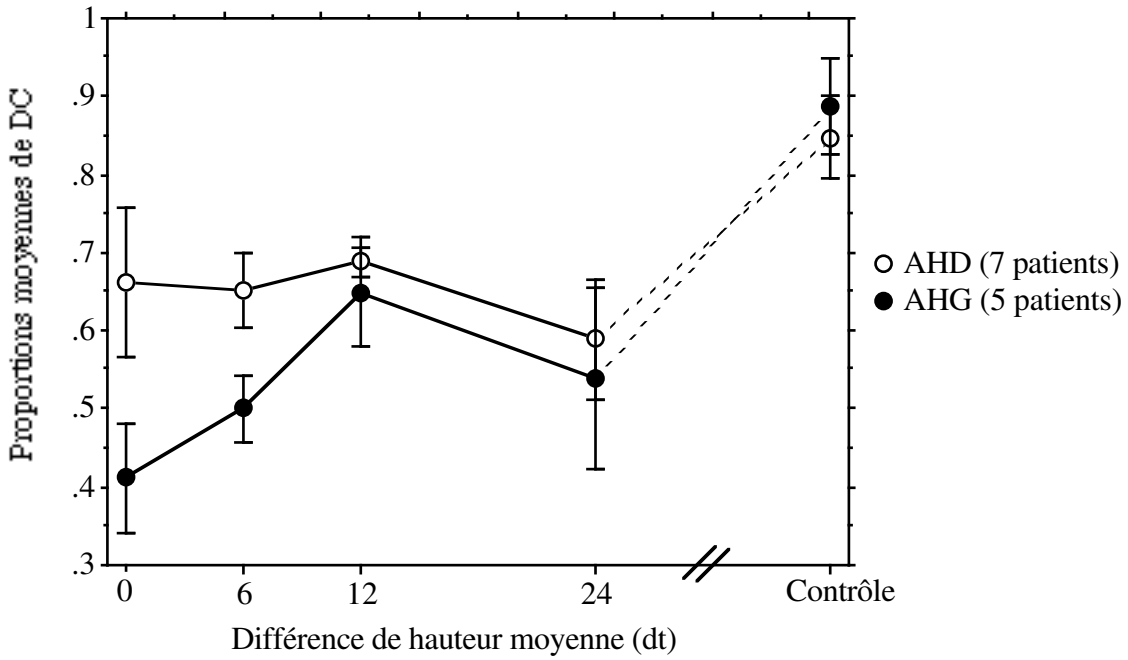


Figure 6.2. Proportions moyennes de détections correctes (DC) obtenues par 7 patients présentant une AHD et 5 patients ayant une AHG dans la tâche de post-reconnaissance de mélodies intercalées, pour les différentes conditions de séparation en hauteur (0, 6, 12 ou 24 dt) et pour la condition contrôle. Les barres verticales indiquent ± 1 erreur-type.

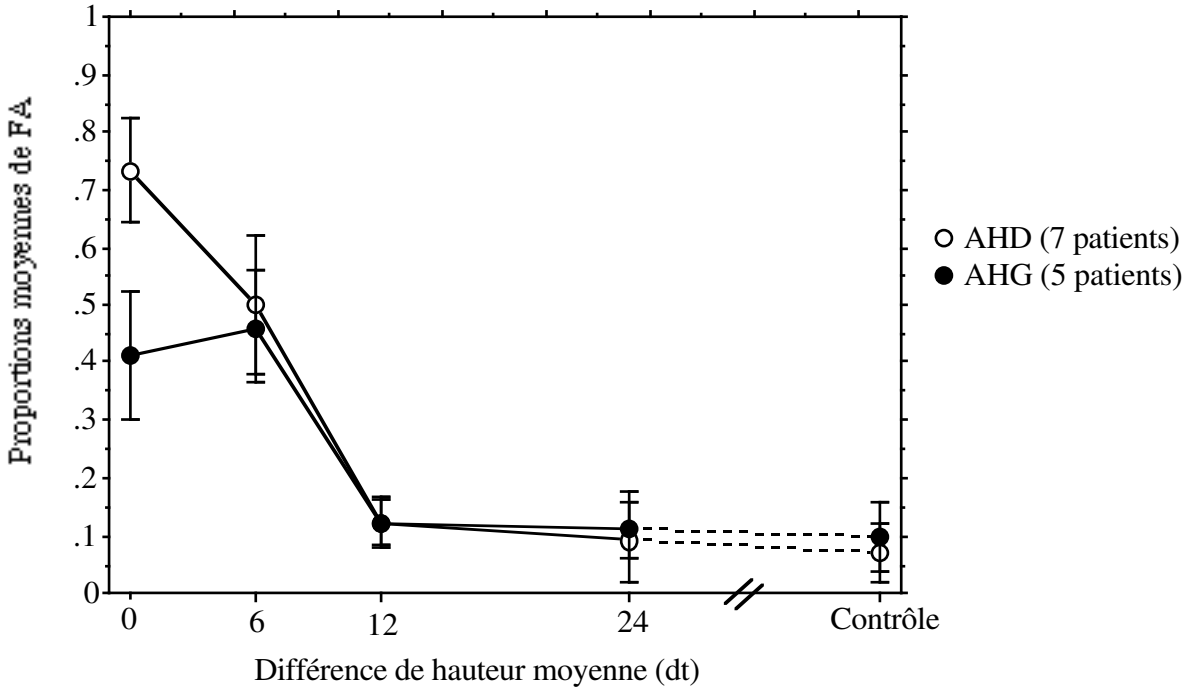


Figure 6.3. Proportions moyennes de fausses alarmes (FA) obtenues par 7 patients présentant une AHD et 5 patients ayant une AHG dans la tâche de post-reconnaissance de mélodies intercalées, pour les différentes conditions de séparation en hauteur (0, 6, 12 ou 24 dt) et pour la condition contrôle. Les barres verticales indiquent ± 1 erreur-type.

En résumé, ces premiers résultats portant sur douze patients épileptiques présentant une atrophie hippocampique droite ou gauche et un hypométabolisme dans les structures adjacentes temporales et parfois frontales, montrent que dans l'ensemble leurs performances de post-reconnaissance de mélodies intercalées sur la base d'une différence de hauteur sont comparables à celles obtenues par des auditeurs sains. Ces patients ne semblent donc pas présenter de déficit dans la capacité de séparer perceptivement une mélodie présentée dans une "mixture" sonore. Cependant, le fait que les performances des patients plafonnent dans la condition de séparation 24 dt pourrait s'expliquer par une interférence en mémoire plus importante chez ces patients quelle que soit la latéralisation du dysfonctionnement cérébral (droit ou gauche). Néanmoins, leurs performances étant également plus faibles dans la condition contrôle, l'hypothèse de l'implication de l'hippocampe dans l'encodage en mémoire de plusieurs informations simultanées demande à être plus amplement testée.

6.3 Reconnaissance de mélodies intercalées avec ou sans présentation préalable de la mélodie : Un cas de dissociation

6.3.1 Introduction

Nous avons également été amenés à tester un patient, L.B, qui suite à un accident vasculaire cérébrale se plaignait de troubles dans l'appréciation musicale. Par ailleurs, ce patient ne présentait pas de déficit dans la compréhension, ni dans la production du langage. Les audiogrammes tonaux et les potentiels évoqués auditifs précoces étaient également normaux, attestant de l'intégrité du système auditif périphérique et des premiers relais centraux.

Les plaintes de L.B concernaient à la fois le changement de la qualité des sons et la difficulté d'extraire un son dans une "mixture". Un extrait de ses commentaires au sujet des modifications observées dans son environnement sonore, est reporté dans l'encadré 6.1.

La dernière plainte, rendant compte des difficultés du patient à suivre une conversation dans du bruit, sauf lorsqu'il sait qui parle ou d'où ça vient, a attiré notre attention. Elle suggérait en effet que L.B manifestait des difficultés à séparer les émissions des différentes sources sonores à l'exception des situations dans lesquelles il connaissait leurs propriétés, lui permettant alors de les rechercher activement. Nous avons donc testé l'hypothèse d'une éventuelle dissociation entre les deux mécanismes impliqués dans l'analyse d'une scène auditive complexe : l'analyse primaire et l'analyse guidée par les connaissances. La série d'expériences précédentes (voir chap. 5) a montré que la comparaison des performances obtenues dans les tâches de reconnaissance de mélodies intercalées lorsque la mélodie à reconnaître était ou non présentée préalablement, permettait de mettre en évidence l'implication de ces deux processus. Nous avons donc proposé ces deux tâches à L.B ainsi qu'une tâche contrôle de simple reconnaissance de mélodies non familières.

Remarques générales

"Les sons sont plus aigus. Ils sont devenus bruts, sans couleur et sans relief. Ils sont confus. Ils semblent amortis".

Les sons de l'environnement

"Ils sont comme filtrés, très aigus. Les sons de foule sont insupportables. J'ai des problèmes lorsqu'il y a des réflexions".

Les sons musicaux

"Les percussions sont très bien entendues, mais d'un niveau sonore plus important que le reste de la mélodie. Il y a un déséquilibre entre les percussions et les instruments à cordes. La trompette est entendue assez bien, mais légèrement voilée, comme à travers un voile de coton. Tous les instruments semblent filtrés. Dans une mélodie, certains instruments sont mieux entendus que d'autres. Les lignes mélodiques sont mal perçues pour les instruments aigus. La musique est nasillarde.

Les sons restent agréables si je me les joue mentalement. Je n'ai pas de problème pour reconnaître des airs connus, ni pour les chanter. Pas de problème pour localiser les sons".

La parole

"Ma propre voix est moins grave. De façon générale, toutes les voix sont plus aiguës. J'ai des difficultés au téléphone pour reconnaître les voix, elles sont tassées, aplaties, sans timbre, sans relief. Le timbre de ma propre voix a changé, je parle dans une chambre avec du coton autour.

"J'ai des difficultés à suivre une conversation dans du bruit mais c'est plus facile lorsque je sais qui parle, d'où ça vient".

Encadré 6.1. Commentaires de L.B sur les modifications de sa perception auditive suite à un accident cérébro-vasculaire causant une lésion située entre les deux premières circonvolutions temporales droites (T1 et T2).

6.3.2 Méthode

6.3.2.1 Présentation du patient et des témoins appariés

L. B. est un homme de 56 ans, droitier, mélomane. Un accident cérébro-vasculaire survenu en août 1997 a causé une lésion temporale postérieure droite. Cette lésion est cortico-sous-corticale. Elle est relativement focale et siège sur les deux premières circonvolutions

temporales (versant inférieur de T1 et supérieur de T2), touchant les aires auditives associatives de l'hémisphère droit et probablement les circuits reliant l'aire primaire (figure 6.4). Les plaintes concernant l'appréciation musicale persistaient toujours au moment de la passation des tests de reconnaissance de mélodies intercalées, qui ont été proposés en octobre 1998. L.B a donné son consentement éclairé par écrit, et cette étude réalisée sous le contrôle médical de C. Belin et B. Frachet a été approuvée par le comité d'éthique.

Trois participants n'ayant pas de troubles neurologiques, et appariés au patient sur les critères d'âge, sexe, niveau d'éducation, formation musicale et latéralité manuelle, ont également participé à l'étude (tableau 6.2).

Patient	Age	Sexe	Profession/ Éducation	Formation musicale	Latéralité manuelle	Audition
L.B	56	M	Ingénieur Bâtiment et Travaux Publics (Bac + 4)	Mélomane	Droitier (90% Test d'Edimbourg)	Audiogramme normal
Sujets témoins						
Pierre. A	46	M	Gérant de studio (Bac)	Musicien	Droitier (92%)	ne rapporte pas de troubles d'audition
Jean-Marc. P	46	M	Bibliothécaire d'orchestre	Musicien (niveau fin d'étude)	Droitier (92%)	ne rapporte pas de troubles d'audition
Jean-François. M	51	M	Informaticien (Bac + 5)	Musicien amateur	Droitier (39%)	rapporte un début de presbyacousie

Tableau 6.2. Caractéristiques du patient L.B et des trois sujets témoins. Ces derniers ont répondu à un questionnaire de latéralité manuelle adapté de Oldfield (1971), présenté dans l'annexe C11.

a)

b)

Figure 6.4. Radiographie du cerveau de L.B en coupe coronale **(a)** et axiale **(b)**. L'hémisphère droit se trouve à gauche sur l'IRM et l'hémisphère gauche à droite, comme si le cerveau était vu de face pour la coupe coronale et de dessous pour la coupe axiale. La lésion (masse blanche) se situe à cheval entre la première et la deuxième circonvolution temporale.

6.3.2.2 Matériel et procédure

Des conditions pré et post identiques à celles proposées dans l'expérience 5 (voir §5.2.2 pour des détails sur la méthode), ainsi que la tâche contrôle de reconnaissance immédiate de mélodies non familières, ont été soumises à L.B et aux trois témoins. Quatre degrés de séparation en hauteur ont été présentés. La séquence distractrice était présentée dans le même registre que la mélodie cible, ou s'en éloignait en moyenne de 6, 12 ou 24 demi-tons vers les fréquences basses. Dans chacune des quatre conditions de séparation en hauteur et pour les deux types de tâche, 16 essais étaient proposés, soit au total 64 essais par test. La condition contrôle était également composée de 16 essais. Un nombre égal d'essais identiques et différents étaient présentés. La différence de hauteur moyenne entre la mélodie et la distractrice était affectée aléatoirement à chaque essai, ainsi que le type d'essai "même" et "différent". Les séquences étaient composées de sons purs.

La passation s'est déroulée à l'Université Paris V à la fois pour L.B et les trois témoins, dans des conditions similaires à celles offertes aux participants des expériences relatées dans les chapitres précédents. La réalisation des deux tâches expérimentales et de la tâche contrôle nécessitait environ 1 heure. Les auditeurs étaient placés dans une salle insonorisée et les séquences étaient diffusées sous écouteurs. Ils donnaient leur réponse en appuyant sur deux touches du clavier, "m" s'il s'agissait des mêmes mélodies et "d" si elles leur semblaient différentes. L'essai suivant était lancé une fois leur réponse donnée.

En plus de ces trois tâches, L.B a passé une série de cinq petits tests visant à dépister d'éventuels troubles de la reconnaissance de mélodies familières (annexe C9), mais il n'a manifesté aucune difficulté à reconnaître les comptines présentées. La tâche contrôle et les deux conditions post et pré, constituaient respectivement les tests 3, 4 et 5 dans cette série de huit tests. La condition contrôle a ainsi été passée avant les conditions expérimentales pour le patient. En revanche, les témoins ont passé d'abord les conditions post et pré, puis la tâche contrôle.

6.3.3 Résultats et discussion

Les performances obtenues par L.B dans les tâches de reconnaissance de mélodies intercalées avec et sans présentation préalable de la mélodie (conditions pré et post) pour les différents degrés de séparation en hauteur (16 essais par condition de séparation), ainsi que dans la tâche contrôle, sont reportées sur la figure 6.5. Les proportions moyennes de reconnaissances correctes recueillies auprès des trois témoins appariés dans ces trois tâches (3 participants * 16 essais par degré de séparation en hauteur), figurent également sur le graphique pour comparaison.

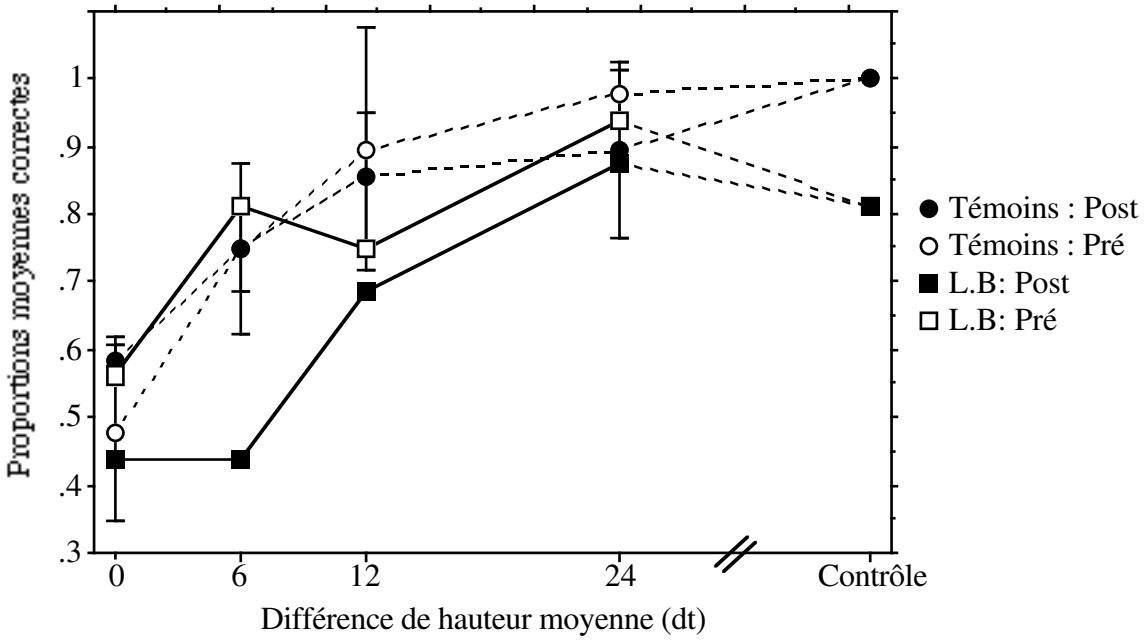


Figure 6.5. Proportions moyennes de réponses correctes obtenues par L.B (courbes en trait plein avec les carrés) et par les trois témoins (courbes en pointillés avec les cercles) pour les différents degrés de séparation en hauteur (0, 6, 12 et 24 dt), dans les conditions Pré et Post (représentées respectivement par les symboles vides et pleins) ainsi que dans une simple tâche de discrimination de mélodies non familières (tâche contrôle). Les barres verticales représentent ± 1 erreur type.

Les résultats obtenus dans la tâche de post-reconnaissance montrent que la différence de hauteur entre la mélodie et la séquence distractrice nécessaire pour reconnaître environ 75 % des mélodies intercalées présentées (SRM), est en moyenne de 6 dt chez les témoins (en accord avec les résultats obtenus dans les expériences précédentes) alors qu'elle est supérieure à 12 dt pour L. B. Pour une différence de hauteur moyenne de 6 dt entre la mélodie et la distractrice, la proportion moyenne de reconnaissances correctes est seulement de 0.44 pour L.B, soit une performance inférieure de plus de deux écart-types à la moyenne obtenue par les témoins. Les performances obtenues par le patient et les témoins sont par ailleurs équivalentes pour une différence de 24 dt. Néanmoins, les performances de L.B obtenues dans la condition contrôle s'avèrent inférieures de 20 % à celles recueillies auprès des témoins, révélant un léger déficit dans cette tâche de discrimination de mélodies non familières. Ce dernier résultat est en accord avec plusieurs études qui ont montré qu'une lésion temporale droite entraînait des déficits dans la discrimination de mélodies non familières (Liégeois-Chauvel, Peretz, Babaï, Laguitton & Chauvel, 1998 ; Peretz, 1990 ; Samson & Zatorre, 1988 ; Zatorre, 1984, 1985, 1988a). Pourtant, de nombreux indices sont mis à la disposition du patient pour accomplir cette tâche. En effet, lorsque les mélodies à comparer diffèrent, deux notes sont modifiées sur les six, le contour, les intervalles et la tonalité sont également changés (annexe A2). Il est par conséquent possible que L.B présente un déficit plus important dans le traitement d'un de ces indices

mélodiques, par exemple le contour et/ou les notes qui constituent des indices prégnants dans la mémorisation immédiate de mélodies (Dowling & Harwood, 1986). Cette possibilité reste à explorer.

Globalement, les résultats à ce test de post-reconnaissance suggèrent que L.B parvient à séparer perceptivement une mélodie intercalée avec des sons distracteurs, mais son seuil de reconnaissance de mélodies intercalées est augmenté par rapport aux témoins. Une différence de hauteur moyenne plus importante entre les événements lui est donc nécessaire pour extraire un pattern dans une "mixture" sonore.

Cette différence de performances observée dans la tâche de post-reconnaissance entre le patient et les témoins, s'estompe lorsque la mélodie est présentée avant le mélange (condition pré). En effet, lorsque L.B connaît précisément la mélodie qu'il doit extraire du "mélange", ses performances de reconnaissance de la mélodie cible intercalée pour une séparation en hauteur de 6 dt, sont comparables à celles obtenues par les témoins[2]. L'absence de déficit dans cette tâche révèle une dissociation entre la capacité d'extraire une mélodie dans un "mélange" lorsque celle-ci est ou n'est pas présentée préalablement.

6.4 Discussion générale

6.4.1 Résumé des principaux résultats

Les résultats de deux études exploratoires examinant la capacité de patients cérébro-lésés à extraire et reconnaître une mélodie inédite intercalée à des sons distracteurs, ont été rapportés dans ce chapitre. La première étude a porté sur un groupe de patients épileptiques présentant une atrophie hippocampique droite ou gauche. Elle révèle que ces patients ne présentent pas de déficit dans la capacité d'extraire une mélodie d'un "mélange". La deuxième étude a examiné l'aptitude d'un patient L.B, présentant des lésions temporales postérieures droites suite à un accident cérébro-vasculaire, à reconnaître une mélodie intercalée à des sons distracteurs lorsque celle-ci était ou non présentée préalablement. La dissociation de ses performances, conforte l'hypothèse de l'implication de mécanismes distincts dans ces deux tâches.

[2]Par ailleurs, on peut noter que le gain constaté dans la condition pré de l'expérience 5 (figure 5.2) ne s'observe pas chez les témoins qui ont des résultats relativement équivalents dans les conditions pré et post. Ce résultat pourrait s'expliquer par le faible nombre d'auditeurs (3) et par leur expertise musicale qui rend leurs performances déjà très élevées dans la condition post. En effet, le gain de performances constaté lorsque l'auditeur entend préalablement la mélodie qu'il doit extraire du "mélange" semble robuste puisque des résultats similaires à ceux obtenus dans l'expérience 5 ont été recueillis dans une expérience préliminaire* conduite auprès de 32 participants (16 par groupe). *Bey, C., & McAdams, S. (1997). Implication des processus descendants dans la formation des flux auditifs : une mise en évidence expérimentale. Actes des Journées Internationales d'Orsay sur les Sciences Cognitives.

6.4.2 Interférence en mémoire due à l'encodage de plusieurs informations simultanées

L'exploration d'une éventuelle amplification du phénomène d'interférence lors de l'écoute polyphonique chez des patients présentant des lésions hippocampiques pourrait conforter l'existence de cette interférence mnésique liée à l'encodage de plusieurs informations simultanées (expériences 1 à 5), et mettre en évidence un déficit mnésique spécifique.

L'encodage en mémoire d'une mélodie présentée simultanément avec une autre séquence dégraderait la trace mnésique de cette mélodie. Il serait intéressant d'examiner les conditions de cette dégradation. En effet, concernant l'interférence en mémoire de travail provoquée par la présentation d'événements durant l'intervalle de rétention (Deutsch, 1970), Semal et Demany (1991, 1993) ont montré qu'elle dépendait de la proximité en hauteur entre les sons interférents et les sons tests. Il serait intéressant d'étudier si l'interférence en mémoire due à l'encodage simultané de plusieurs informations auditives présente les mêmes propriétés.

Les structures hippocampiques et/ou temporales inférieures pourraient ainsi jouer un rôle dans l'encodage en mémoire de plusieurs séquences présentées simultanément. Cependant, cette hypothèse ne va pas dans le sens des résultats obtenus par Zatorre et Samson (1991). Ces auteurs ont en effet montré que la mémoire tonale des patients présentant une résection temporale, frontale ou fronto-temporale droite, était affectée par la présence de sons interférents pendant l'intervalle de rétention. Ce résultat suggère que le maintien de l'information tonale à court terme serait plutôt latéralisé dans l'hémisphère droit, et impliquerait les structures antérieures. En outre, une plus grande étendue de l'excision des structures limbiques n'entraînait pas de déficits supplémentaires. Une hypothèse qui pourrait rendre compte de cette différence est que l'interférence en mémoire induite pendant l'encodage de l'information auditive et celle provoquée pendant leur stockage, seraient deux phénomènes distincts et ne feraient pas intervenir les mêmes structures cérébrales.

6.4.3 Dissociation entre les mécanismes impliqués dans les conditions pré et post

La dissociation des performances obtenues dans les conditions pré et post chez le patient L.B présentant une lésion au niveau des structures temporales postérieures droites, pourrait être le reflet d'une dissociation entre les deux types d'analyses de la scène auditive distinguées par Bregman (1990, 1994) : l'analyse primaire et l'analyse guidée par les connaissances. Les résultats des expériences relatés dans les chapitres précédents, suggèrent en effet que la tâche de post-reconnaissance de mélodies intercalées implique majoritairement des processus primaires d'analyse de scène auditive (chap. 4), et que le fait de présenter la mélodie à reconnaître avant la séquence composite fait vraisemblablement intervenir en plus des processus d'analyse guidée par les connaissances (chap. 5).

Nous avons vu que le seuil de post-reconnaissance de mélodies intercalées était augmenté chez L.B, ce qui pourrait être dû à une altération des mécanismes impliqués dans l'analyse primaire. L'étude menée par Efron et al. (1983) a en effet suggéré que des patients ayant subi une résection temporale droite ou gauche touchant essentiellement les aires auditives associatives, manifestaient des difficultés à extraire un son dans une "mixture". Les performances de L.B étaient par ailleurs tout à fait comparables à celles obtenues par les témoins dans la condition pré, suggérant que les processus d'analyse guidée par les connaissances demeuraient opérationnels et ont permis de compenser le déficit observé dans la condition post. La légère infériorité de ses performances dans la tâche de reconnaissance de mélodies présentées sans séquence distractrice (tâche contrôle), laisse penser qu'il pourrait manifester des difficultés dans le traitement d'indices mélodiques, comme les notes et/ou le contour. Cependant, le fait qu'il puisse améliorer ses performances dans la condition pré, suppose qu'il ait tout de même encodé certaines propriétés de la première mélodie présentée. Des tests supplémentaires sont donc nécessaires pour déterminer un éventuel déficit dans le traitement de certains indices mélodiques, et ainsi mettre en évidence les connaissances qui lui ont permis d'extraire la mélodie du "mélange". Les résultats de l'expérience 6 ont révélé qu'un amorçage fréquentiel pouvait expliquer en partie l'amélioration de performances constatée dans la condition pré. Or dans l'étude menée par Peretz (1990), nous avons souligné (voir §3.2.1) que malgré un déficit dans la discrimination de mélodies non familières, les patients présentant une lésion dans l'hémisphère droit avaient des performances augmentées si les deux mélodies étaient présentées dans le même registre de hauteur, par rapport à une condition où elles étaient transposées. Ce résultat laisse penser que ces patients sont encore capables d'extraire la hauteur des notes, et de se servir de cette information pour détecter un changement. Nous pouvons donc faire l'hypothèse que L.B a pu se servir de cette information pour améliorer ses performances dans la condition où la mélodie était présentée préalablement.

La plainte formulée par ce patient concernant ses difficultés à suivre une conversation dans du bruit sauf lorsqu'il connaît la voix de son interlocuteur et la localisation de la source, converge donc tout à fait avec ce pattern de résultats. L.B a également formulé d'autres plaintes concernant notamment le changement de qualité sonore des événements (encadré 6.1). Il serait par conséquent intéressant de lui proposer ces mêmes tâches de reconnaissance de mélodies intercalées avec comme seul indice de ségrégation, le timbre. Ceci nous permettrait de déterminer s'il présente un déficit dans l'analyse perceptive d'une séquence composite dont les événements se différencient sur des caractéristiques timbrales.

6.5 Conclusion

En conclusion, je voudrais encore une fois souligner le fait que les deux études présentées dans ce chapitre sont exploratoires et que ces premiers résultats demandent à être complétés. Néanmoins, j'ai souhaité que ces premières observations apparaissent dans ce document parce qu'elles permettent de formuler des hypothèses intéressantes par rapport aux questions soulevées à savoir, celle de l'implication de processus distincts dans les conditions pré et post, et celle de l'existence d'un phénomène d'interférence en mémoire lors de l'encodage de plusieurs séquences sonores présentées simultanément.

Peu d'études en neuropsychologie ont exploré les mécanismes impliqués dans la formation des flux auditifs. Ce fait peut être attribué en partie à la difficulté de mesurer ces processus en milieu clinique. Des tests comportant une consigne simple et peu d'essais pouvant faciliter l'évaluation de ces processus dans un contexte hospitalier, ont fait défaut jusqu'à présent. Les paradigmes de reconnaissance et post-reconnaissance de mélodies intercalées fournissent donc à la communauté scientifique et médicale, un outil d'investigation permettant d'explorer ces processus de construction des flux auditifs.

Nous avons vu qu'en raison de l'implication d'autres composantes cognitives, l'échec dans ce type de tâche indirecte ne nous permettait pas d'inférer avec certitude l'existence d'un déficit dans l'analyse perceptive d'une séquence composite (voir §4.7.3 pour une discussion de la non équivalence logique entre le nombre de flux perçus et les performances de reconnaissance d'un pattern cible). Néanmoins, la convergence de plusieurs tâches directe et indirecte, ou bien la passation des deux conditions pré et post ainsi que la condition contrôle, permettent d'outrepasser cette difficulté. En effet, l'étude du patient L.B a montré que ses performances obtenues dans la condition pré constituaient une référence nous permettant d'interpréter son déficit en post-reconnaissance. L'introduction de ces tests en milieu clinique pour des études à plus grande échelle, offre donc des perspectives dans le dépistage d'éventuels déficits des mécanismes impliqués dans l'analyse de la scène auditive.

Conclusion et perspectives

Les étapes de traitement dans la perception de mélodies intercalées

Les recherches menées dans le cadre de cette thèse permettent d'émettre des hypothèses sur les opérations mentales mises en jeu dans la reconnaissance d'une mélodie intercalée. Ces étapes de traitement : l'organisation perceptive de la séquence composite, l'encodage en mémoire de la première mélodie présentée et la comparaison entre la représentation de cette mélodie de référence et la deuxième mélodie proposée, sont présentées ci-après. La connaissance préalable ou non de la mélodie modifie chacune de ces étapes.

Organisation perceptive de la séquence composite

A l'écoute d'une séquence composée d'une mélodie intercalée avec des sons distracteurs, l'auditeur va organiser perceptivement la séquence composite en flux auditifs. Cette analyse perceptive se fait de façon pré-attentive (Bregman, 1990 ; Sussman et al., 1999) sur la base de la similarité en hauteur, timbre, intensité et localisation des événements successifs (Bregman, 1990, chap. 2). Les performances de reconnaissance de mélodies intercalées augmentent en effet avec la différence de hauteur séparant la mélodie cible de la distractrice (expériences 1, 2 et 5), et leur dissemblance de timbre (expérience 3). Cette analyse perceptive conduit à la construction d'un seul flux si l'ensemble des événements présentent des caractéristiques communes, de deux flux distincts si les sons de la mélodie se distinguent de ceux de la distractrice, ou bien à la fission partielle de la mélodie si quelques sons seulement de la mélodie sont suffisamment dissemblables de ceux de la distractrice pour se séparer perceptivement du "mélange". Nous avons vu que les performances à cette tâche étaient aléatoires lorsque la séquence composite était organisée en un seul flux, et s'approchaient de celles obtenues en l'absence de distractrice lorsque l'auditeur percevait deux flux distincts (expérience 1). Elles étaient juste supérieures au hasard pour un faible degré de séparation en hauteur ou en timbre, suggérant que certains sons de la mélodie se scindaient et laissaient émerger certains indices permettant d'accomplir la tâche. L'expérience 7 a conforté l'idée que la progression des performances était liée au nombre de sons qui se séparaient perceptivement de la séquence composite.

Cette analyse primaire de la scène auditive opère conjointement et peut-être en interaction avec une analyse guidée par les connaissances (Bregman, 1994, chap. 4). Dans la situation où la mélodie à reconnaître est présentée après la séquence à organiser (post), les schémas qui interviennent dans l'organisation perceptive de la séquence sont probablement réduits à la connaissance du registre dans lequel la mélodie est présentée. En effet, dans ce paradigme l'auditeur ne connaît pas la mélodie qu'il doit extraire du mélange mais la tendance des performances à augmenter au cours de la passation (expérience 2), a suggéré qu'il devait extraire certaines régularités propres à cette situation expérimentale, comme par exemple le fait que celle-ci apparaisse toujours dans le même registre (voir §4.7.2.1 pour une discussion de cet aspect). En revanche, lorsque la mélodie à reconnaître est présentée avant la séquence composite (pré) (Dowling, 1973, expérience 2 ; Vliegen & Oxenham, 1999, expérience 2), l'auditeur a préalablement encodé les propriétés de cette mélodie inédite, les notes et probablement le contour (Dowling & Fujitani, 1971) et va apparier ce schéma élaboré à la représentation sensorielle de la séquence composite. L'amélioration des performances de reconnaissance, réduite si la mélodie préalablement présentée est transposée, atteste de l'implication des connaissances de l'auditeur dans cette situation, en particulier la fréquence des notes et peut-être le contour puisque les auditeurs demeurent plus performants qu'en post-reconnaissance (expériences 5 et 6). De l'action conjointe de ces deux processus d'analyse de scène auditive, vont émerger les propriétés de la mélodie cible.

Dans le cas de la reconnaissance d'une mélodie familière intercalée (Dowling, 1973, expériences 1 et 3 ; Dowling, Lung & Herrbold, 1987, expérience 1 ; Hartmann & Johnson, 1991), les processus sont probablement comparables à ceux décrits ci-dessus dans la situation où la mélodie à reconnaître est présentée avant la séquence à organiser (pré). La différence réside dans la nature du schéma activé. Dans le cas d'une mélodie stockée en mémoire à long terme, les intervalles plutôt que les notes et le contour sont sûrement préactivés (Dowling, 1978 ; Dowling & Fujitani, 1971 ; Dowling & Harwood, 1986) et guident l'analyse perceptive. Ceci pourrait expliquer la différence observée entre les résultats de l'expérience 5 qui montrent que même si l'auditeur a préalablement entendu la mélodie il ne parvient pas à l'extraire du mélange, et ceux des études menées par W. J. Dowling (Dowling, 1973 ; Dowling et al., 1987) qui suggèrent l'inverse pour des mélodies familières. En raison de la différence de nature du schéma activé, l'appariement entre le schéma de la mélodie et la représentation sensorielle de la séquence composite est peut être différent. En effet, du fait du stockage de la mélodie en mémoire à long terme, l'émergence d'une petite partie de la mélodie cible du mélange suffit à l'auditeur pour évaluer s'il s'agit ou non de la mélodie recherchée. L'activation du schéma de la mélodie familière se fait, avant la présentation du mélange lorsque l'auditeur est informé de la mélodie qui va être présentée (Dowling, 1973, expérience 3 ; Dowling et al., 1987, expérience 1), ou au moment de la présentation de la séquence composite en l'absence d'amorçage verbal (Dowling, 1973, expérience 1 ; Hartmann & Johnson, 1991). Le type d'amorçage, verbal ou mélodique, a également sûrement une influence sur la nature des schémas activés.

Encodage en mémoire de la première mélodie présentée

Les propriétés de la première mélodie présentée sont encodées en mémoire de travail par l'auditeur. Elle constitue la mélodie de référence sur la base de laquelle la deuxième mélodie va être comparée. Dans la tâche de post-reconnaissance, la mélodie de référence est encodée simultanément avec la séquence distractrice. La présence de cette autre séquence va alors interférer dans l'élaboration de la trace mnésique de la mélodie de référence. Dans l'ensemble des expériences que nous avons menées, les performances de post-reconnaissance de la mélodie cible intercalée obtenues pour un degré de séparation maximal, demeuraient toujours inférieures à celles recueillies lorsqu'elle était présentée isolément (condition contrôle). Cette différence de performances ne s'observait pas en revanche lorsque la première mélodie entendue était encodée seule (condition pré, expérience 5), suggérant que la trace mnésique élaborée était alors plus précise. Dans le cas de la présentation de mélodies familières intercalées (Dowling, 1973 ; Dowling et al., 1987 ; Hartmann & Johnson, 1991), le problème de l'interférence en mémoire due à l'encodage simultané de la mélodie et des sons distracteurs ne se pose pas. La mélodie est familière par conséquent, elle a déjà été encodée préalablement et stockée.

Comparaison entre les mélodies

La représentation de la mélodie de référence ainsi élaborée, est comparée à la deuxième mélodie présentée. Sur la base de la similarité/dissemblance de ces deux représentations, l'auditeur met en oeuvre des processus décisionnels qui vont l'amener à juger si les deux mélodies présentées sont identiques ou différentes. Dans le cas où l'auditeur dispose d'une représentation précise de la mélodie de référence (pré), il va développer des attentes fréquentielles et temporelles spécifiques concernant la deuxième mélodie (Dowling et al., 1987). Si celle-ci apparaît entremêlée aux sons distracteurs dans un même flux, il va avoir tendance à rejeter la possibilité qu'il s'agisse de la même mélodie. Les participants ayant été soumis à la condition pré (expérience 5) ont en effet majoritairement répondu que les mélodies présentées étaient différentes (proportion importante de fausses alarmes). Ce biais de réponse ne s'observait pas en revanche dans la condition post. Ce résultat laisse penser que lorsque la représentation de la mélodie de référence est au contraire imprécise, à l'écoute de la deuxième mélodie présentée isolément les auditeurs dans le doute acceptent la possibilité que cette même mélodie ait pu être présentée préalablement. Un profil de jugement similaire à celui observé dans la condition pré a été constaté dans une situation homologue dans laquelle la mélodie était transposée (expérience 6). Ceci nous amène à la conclusion que la différence d'encodage de la mélodie de référence dans ces deux types de situations où la mélodie à reconnaître est présentée avant ou après la séquence composite, pourrait être à l'origine de ce changement de critère décisionnel.

L'ensemble des opérations mentales mises en évidence dans ces différentes tâches expérimentales, peuvent être généralisées à ce qui se passe dans une situation naturelle d'écoute polyphonique. Lorsqu'un auditeur écoute une fugue de Bach, il organise perceptivement les événements sonores successifs en voix mélodiques distinctes afin de pouvoir suivre une voix puis une autre. Les mécanismes mis en évidence dans la perception de mélodies intercalées à savoir, l'intervention conjointe des processus d'analyse primaire et d'analyse de scène guidée par les connaissances et l'encodage en mémoire de plusieurs informations auditives simultanées, s'appliquent donc à l'écoute de plusieurs voix mélodiques parallèles. Plus généralement, les mécanismes examinés dans cette étude concernent l'ensemble des situations dans lesquelles l'auditeur suit l'émission d'une source sonore alors que plusieurs autres sources émettent simultanément. Il peut s'agir de parole, de musique ou de sons de l'environnement.

Apport dans l'étude de la mémoire auditive

L'analyse fonctionnelle des processus impliqués dans la perception de mélodies intercalées fournit des éléments intéressants pour notre connaissance des mécanismes impliqués dans la formation des flux auditifs, mais aussi sur le fonctionnement de la mémoire auditive. En effet, ces recherches ont révélé que les performances de reconnaissance d'une mélodie étaient affectées par la présence d'une autre séquence présentée simultanément. Nous avons émis l'hypothèse que cette altération des performances pouvait s'expliquer par une dégradation de la trace mnésique de la mélodie lorsque celle-ci était encodée simultanément avec une autre séquence. L'existence d'un phénomène d'interférence en mémoire de travail a été mis en évidence par Deutsch (1970). L'auteur a en effet montré que la mémoire de la hauteur tonale de sons isolés était affectée par la présence de sons distracteurs diffusés dans l'intervalle de temps qui séparait la présentation des deux sons. L'étude de Dowling et ses collaborateurs (1995) a révélé que la mémoire des mélodies pouvait également être dégradée par la présence d'items interférents pendant l'intervalle de rétention. La différence entre le phénomène d'interférence décrit dans la littérature (Deutsch, 1970, 1972 ; Dowling et al., 1995 ; Semal & Demany, 1991, 1993 ; Semal, Demany, Ueda & Hallé, 1996) et celui rapporté ici, réside dans le fait qu'ils ne se produisent pas au même moment dans le processus de mémorisation. Dans notre étude, les sons distracteurs ne sont pas présentés durant l'étape de stockage, mais au moment de l'encodage de la mélodie. Les études portant sur la perception de plusieurs mélodies présentées simultanément s'avèrent donc très intéressantes pour examiner le fonctionnement de la mémoire auditive, en particulier durant cette étape d'encodage.

De plus, l'existence ou non d'un phénomène d'interférence en mémoire de travail offre un moyen de tester l'indépendance de certains traitements. Semal et ses collaborateurs (Semal & Demany, 1991, 1993 ; Semal, Demany, Ueda & Hallé, 1996) ont ainsi montré, en utilisant le paradigme employé par Deutsch (1970), que la mémoire tonale ne dépendait pas de la nature des

sons. Elle était altérée par la présentation durant l'intervalle de rétention de sons interférents de hauteurs proches, qu'ils aient ou non le même timbre, et que ce soit ou non de la parole. La mise en évidence d'un phénomène d'interférence en mémoire durant l'encodage de plusieurs informations auditives, offre par conséquent des perspectives de recherche pour mieux comprendre les conditions de ce phénomène, et également approfondir notre connaissance de la mémoire des mélodies dans des situations musicales réelles (Dowling & Harwood, 1986).

Apport dans la façon d'envisager la perception auditive

Enfin, les résultats des études qui ont examiné l'implication des processus descendants dans l'organisation perceptive d'une séquence sonore (chapitre 5), nous conduisent à une réflexion générale sur la façon d'envisager le fonctionnement du système cognitif. Actuellement, les modèles de la cognition auditive envisagent le traitement des informations sensorielles comme une série d'étapes séquentielles et hiérarchiques. Cette approche admet que l'"étape" d'organisation perceptive, qui consiste à regrouper les informations sensorielles pour construire des "objets", précède l'extraction des propriétés des objets ainsi formés, leur reconnaissance et leur identification. Elle privilégie ainsi la voie "ascendante" du traitement. S'intéresser à la contribution des connaissances dans la construction des "objets", c'est se poser la question de la séquentialité de ces étapes, et celle apparemment paradoxale de savoir si la connaissance préalable des propriétés d'un objet intervient dans la construction de cet objet et donc dans l'émergence de ces propriétés.

Dans le cadre de cette thèse, nous avons examiné si la connaissance préalable des propriétés d'une mélodie pouvait améliorer leur émergence lorsque la mélodie était intercalée à des sons distracteurs. Les résultats obtenus laissent penser qu'effectivement l'émergence des propriétés de la mélodie a été favorisée par la connaissance de ses éléments constitutifs (amorçage fréquentiel) et peut-être également du lien entre ses éléments (contour) (expérience 6). Le rôle des connaissances dans la construction des flux auditifs demande à être plus amplement examiné. Cette question apportera des éléments intéressants dans notre façon de concevoir le fonctionnement de la perception auditive et plus généralement du système cognitif.

Perspectives

L'ensemble de ces recherches nous amènent à nous poser plusieurs questions concernant la façon dont le système auditif procède pour organiser perceptivement les événements sonores qui lui parviennent. Les deux types d'analyses distinguées par Bregman (1990, 1994), l'analyse primaire et l'analyse guidée par les schémas, opèrent probablement conjointement dans la construction de la scène auditive. Néanmoins, la question est de savoir comment le mécanisme de partition sensorielle qui construit des entités perceptives (les sons et les flux auditifs), et le mécanisme de sélection de l'information qui apparie les schémas activés à

la représentation sensorielle, fonctionnent-ils ensemble ? Opèrent-ils de façon indépendante sur la représentation sensorielle ou interagissent-ils pour construire les flux auditifs ? Si, le mécanisme descendant semble pouvoir amplifier la fission d'une séquence lorsque les deux types d'analyses conduisent à la même interprétation perceptive (expérience 5), qu'en est-il lorsque ces deux processus sont en conflit ? En d'autres termes, quelle est la puissance de ce mécanisme ? Lorsque l'auditeur entendait préalablement la mélodie inédite qu'il devait sélectionner dans le "mélange", nous avons constaté qu'il ne parvenait pas à l'extraire si elle était intégrée avec les autres sons distracteurs (expérience 5). Cependant, les études menées par W. J. Dowling (Dowling, 1973 ; Dowling et al., 1987) ont suggéré le contraire pour des mélodies familières. Cela signifie-t-il que la puissance de l'analyse guidée par les connaissances dépend de la nature des schémas activés ? Et si ce mécanisme descendant permet de sélectionner un sous-ensemble de sons à l'intérieur d'un flux auditif, parvient-il en revanche à regrouper des sons qui ont été assignés à des entités perceptives distinctes ?

Les futures recherches seront conduites dans le but d'apporter des éléments de réponse à ces questions concernant les relations qu'entretiennent les processus impliqués dans l'analyse de la scène auditive. Des études menées à la fois chez des auditeurs sains et des patients cérébro-lésés contribueront à une meilleure compréhension de ces mécanismes clés de la perception auditive.

Annexes

ANNEXE A

1-Les mélodies

36 mélodies de 6 notes ont été construites, chacune présente une version originale (o) et une version modifiée (m) soit au total 72 mélodies différentes. Dans la version modifiée, deux notes ont été changées dans une gamme de ± 4 demi-tons, la deuxième et la quatrième ou la troisième et la cinquième. La note moyenne autour de laquelle les mélodies ont été présentées était le la4 (fréquence fondamentale de 880 Hz, note MIDI 81) pour les expériences où la différence de hauteur variait (l'ensemble des expériences sauf l'expérience 3) et le mib3 (fréquence fondamentale de 311 Hz, note MIDI 63) lorsque le timbre variait (expérience 3). Cette note moyenne fluctuait à chaque essai dans une gamme allant de -3 à +2 dt. Les 6 notes de chacune des mélodies reportées ci-dessous sont donc exprimées en intervalles (dt) par rapport à la note moyenne.

N° mélodie	**Version**	**Notes (intervalles par rapport à la note moyenne en dt)**						**Version**	**Notes (intervalles par rapport à la note moyenne en dt)**					
1	o	2	4	2	1	-1	-3	m	2	0	2	-2	-1	-3
2	o	-3	-1	1	2	4	2	m	-3	-1	0	2	1	2
3	o	-2	0	-2	0	2	3	m	-2	0	-1	0	5	3
4	o	-2	0	2	3	2	3	m	-2	0	5	3	0	3
5	o	3	5	3	1	0	-4	m	3	1	3	-3	0	-4
6	o	-2	0	2	3	2	-2	m	-2	0	1	3	-2	-2
7	o	-3	-1	1	-1	1	4	m	-3	3	1	-2	1	4
8	o	3	5	3	1	-2	-4	m	3	2	3	-2	-2	-4
9	o	-2	0	2	3	0	-2	m	-2	0	4	3	-4	-2
10	o	-3	-1	-3	-1	3	4	m	-3	0	-3	3	3	4
11	o	-3	-1	1	-1	3	4	m	-3	-1	-3	-1	5	4
12	o	-1	1	3	1	-2	-1	m	-1	2	3	-3	-2	-1

N° mélodie	Version	Notes (intervalles par rapport à la note moyenne en dt)						Version	Notes (intervalles par rapport à la note moyenne en dt)					
13	o	3	5	3	0	-2	-4	m	3	5	1	0	-6	-4
14	o	-4	-2	0	3	5	3	m	-4	-2	3	3	2	3
15	o	-3	-1	-3	1	3	4	m	-3	-1	0	1	7	4
16	o	1	3	1	-2	0	1	m	1	0	1	-4	0	1
17	o	-1	1	3	-1	-2	-1	m	-1	5	3	1	-2	-1
18	o	-3	-1	1	4	3	4	m	-3	-1	3	4	6	4
19	o	1	3	5	1	0	-4	m	1	3	2	1	-3	-4
20	o	3	5	1	0	-2	-4	m	3	5	3	0	-6	-4
21	o	-2	0	3	2	0	-2	m	-2	2	3	-2	0	-2
22	o	-4	-2	2	3	5	3	m	-4	-2	6	3	7	3
23	o	-3	-1	3	4	3	4	m	-3	3	3	2	3	4
24	o	-4	-2	1	3	5	1	m	-4	-2	-1	3	2	1
25	o	-1	1	-3	-1	3	4	m	-1	1	0	-1	1	4
26	o	2	4	1	0	-3	-2	m	2	4	-1	0	0	-2
27	o	-3	-1	2	4	1	2	m	-3	-1	6	4	3	2
28	o	-2	2	3	2	0	-2	m	-2	2	1	2	-4	-2
29	o	-4	0	2	3	5	3	m	-4	-2	2	7	5	3
30	o	-3	1	-1	1	3	4	m	-3	-2	-1	2	3	4
31	o	-1	3	1	-1	-2	-1	m	-1	3	-3	-1	0	-1
32	o	-3	1	3	4	3	4	m	-3	1	6	4	5	4
33	o	-4	0	1	3	5	1	m	-4	3	1	2	5	1
34	o	-3	1	2	4	1	2	m	-3	4	2	0	1	2
35	o	-4	0	1	5	3	1	m	-4	0	5	5	2	1
36	o	-2	2	3	0	2	3	m	-2	2	4	0	5	3