

Catégorisation asymétrique de séquences de hauteurs musicales

Titre courant : Catégorisation asymétrique

Charles Delbé
Robert M. French
Emmanuel Bigand
LEAD-CNRS, UMR 5022, Dijon, France.

Charles Delbé

Université de Bourgogne

LEAD CNRS UMR 5022

Pôle AAFE - Esplanade Erasme

21065 DIJON Cedex, France

charles.delbe@u-bourgogne.fr

Tél:+33(0)3.80.39.37.23.

Fax: +33 (0)3.80.39.57.67.

Résumé

Une asymétrie inattendue dans une tâche d'apprentissage de catégories visuelles chez de jeunes enfants a été observée par Quinn, Eimas & Rosenkrantz (1993). Une série de résultats expérimentaux et de simulations ont montré que cette asymétrie est due à un phénomène d'inclusion perceptive de la catégorie visuelle "chat" à l'intérieur de la catégorie "chien", qui trouve son origine dans la plus grande variabilité des distributions des attributs visuels de la catégorie des chiens par rapport à celles des chats (Mareschal & French, 1997; Mareschal, French, & Quinn, 2000; French, Mermillod, Quinn, & Mareschal, 2001; French, Mareschal, Mermillod, & Quinn, 2004). Dans la présente étude, nous avons cherché à savoir si ce phénomène de catégorisation asymétrique pouvait être répliqué dans le domaine auditif. Nous avons donc mis au point une série de stimuli auditifs séquentiels, analogues aux stimuli visuels de Quinn et al. Deux expériences portant sur des auditeurs adultes semblent démontrer la présence d'un effet de catégorisation asymétrique comparable dans la modalité auditive. De plus, des simulations connexionnistes confirment que des processus ascendants perceptifs, sont largement à l'origine de nos résultats comportementaux.

Asymmetric Categorization in the Sequential Auditory Domain

Abstract

An unusual visual category learning asymmetry in infants was observed by Quinn, Eimas, & Rosenkrantz (1993). A series of experiments and simulations seemed to show that this asymmetry was due the perceptual inclusion of the cat category within the dog category because of the greater perceptual variability of the distributions of the visual features of dogs compared to cats (Mareschal & French, 1997; Mareschal, French, & Quinn, 2000; French, Mermillod, Quinn, & Mareschal, 2001; French, Mareschal, Mermillod, & Quinn, 2004). In the present paper, we explore whether this asymmetric categorization phenomenon generalizes to the auditory domain. We developed a series of sequential auditory stimuli analogous to the visual stimuli in Quinn et al. Two experiments on adult listeners using these stimuli seem to demonstrate the presence of an identical asymmetric categorization effect in the sequential auditory domain. Furthermore, connectionist simulations confirmed that purely bottom-up processes were largely responsible for our behavioural results.

1 Introduction

1.1 L'effet d'asymétrie dans la formation de catégories visuelles

Quinn, Eimas, & Rosenkrantz (1993) et Eimas, Quinn, & Cowan (1994) ont démontré un effet d'asymétrie inattendu dans l'acquisition de catégories visuelles chez les jeunes enfants. Lorsque des nourrissons âgés de 3 à 4 mois étaient exposés à différentes photographies de chiens ou de chats, ils étaient capables de former des catégories perceptives de chacun de ces groupes d'images. Ainsi, les enfants auxquels on avait d'abord montré des images de chats, puis les images d'un chien ou d'un nouveau chat, ont porté plus d'attention au chien qu'au nouveau chat. Cette préférence fut interprétée en termes de formation d'une catégorie de chats qui exclut les chiens. Cependant, les enfants qui furent d'abord exposés à des images de chiens puis à l'image d'un chat ou d'un nouveau chien ne montrèrent de préférence pour aucune de ces deux photographies.

Ce résultat surprenant indiqua que les enfants avaient formé une catégorie "chiens" qui inclut les chats. En d'autres termes, ces enfants ont montré une exclusivité asymétrique lors du développement de certaines catégories perceptives. Ainsi, la catégorie "chiens" n'excluait pas les chats, alors que la catégorie de chats excluait les chiens.

Depuis 1997, un certain nombre de travaux expérimentaux et de modélisation ont tenté d'étendre et d'expliquer ces résultats (Mareschal & French, 1997; Mareschal, French, & Quinn, 2000; French, Mermillod, Quinn, & Mareschal, 2001; French, Mareschal, Mermillod, & Quinn, 2004). Ces études ont établi que l'origine de cet effet d'asymétrie provenait de la variabilité plus faible de la catégorie des chats comparée à celle des chiens, c'est-à-dire que, perceptivement, cette dernière catégorie englobait en grande partie la première. En effet, différentes caractéristiques visuelles des exemplaires de chiens et de chats présentés aux enfants, telles que la distance entre les yeux ou la longueur des oreilles, montraient des distributions de valeurs semblables à celles de la figure 1 ; par exemple, une faible proportion des exemplaires de chats présentait une longueur d'oreille fréquente dans les exemplaires de chiens, alors qu'une proportion plus importante de chiens montrait une longueur d'oreille fréquente dans les exemplaires de chats. En conséquence, des mécanismes perceptifs purement ascendants semblaient suffisants pour rendre compte de l'effet de catégorisation asymétrique observé par Quinn et al.

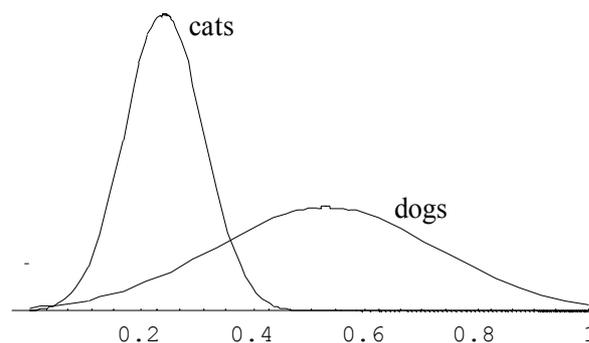


Figure 1. Inclusion et variabilités des distributions d'une caractéristique visuelle hypothétique des catégories *Chiens* et *Chats* dans l'étude de Quinn et al.(1993).

Figure 1. General inclusion and variability of the feature distributions for *Dogs* and *Cats* in Quinn et al. (1993).

Afin de confirmer que l'effet d'asymétrie observé n'était pas dû à une connaissance antérieure à l'expérience des catégories visuelles des chiens et des chats chez les jeunes enfants, French, Mareschal, Mermillod, & Quinn (2004) manipulèrent la variabilité de ces catégories en sélectionnant un jeu de différentes races de chiens, présentant peu de variabilité, ainsi qu'un jeu de photos de chats montrant une plus grande variabilité. De cette façon, bien que les catégories de haut-niveau (conceptuelles) correspondantes restèrent inchangées, leur variabilité perceptive de bas-niveau fut inversée. Le modèle connexionniste qui a été développé, et qui se base uniquement sur les distributions statistiques des caractéristiques visuelles des deux catégories, prédit une inversion de l'asymétrie initialement observée. Cette prédiction fut effectivement validée par les résultats comportementaux de l'étude. Ainsi, les auteurs conclurent que les jeunes enfants se basaient exclusivement sur les propriétés statistiques (extraits par des processus ascendants) des stimuli lors de la discrimination d'exemplaires des catégories.

1.2 Le domaine auditif

La capacité des jeunes enfants à discriminer deux catégories de stimuli visuels complexes de façon purement ascendante – c'est-à-dire en l'absence de connaissances conceptuelles concernant ces stimuli – nous a amené à nous demander si il pouvait exister un phénomène analogue dans la modalité auditive.

Des recherches sur la perception musicale ont révélé des asymétries de jugement dans des tâches de catégorisation et discrimination de styles. Par exemple, Bigand & Barrouillet (1996) ont rapporté que, dans une tâche de détection d'intrus, des participants non-musiciens qui avaient été familiarisés avec une sélection d'extraits de musique baroque (e.g. Bach) discriminaient ensuite des nouvelles pièces baroques mélangées à une sélection d'extraits de musique tonale du début du XX^{ième} siècle (e.g. Debussy). Par contre, lorsque les participants étaient d'abord familiarisés avec le répertoire d'oeuvres du début du XX^{ième} siècle puis testés avec une nouvelle sélection d'oeuvres de cette période mélangées à des pièces baroques, ils ne discriminaient pas de façon supérieure au hasard les extraits d'œuvres de la catégorie d'exposition. Dans cette étude, l'asymétrie de discrimination peut être expliquée par l'inclusion de la catégorie Baroque dans la catégorie Moderne : en effet, ce dernier répertoire est statistiquement moins contraint en terme d'utilisation du matériau musical tonal (organisation horizontale et verticale des hauteurs, structures rythmiques,...). Ainsi, la variabilité des différentes dimensions musicales, plus large dans la catégorie Moderne que dans la catégorie Baroque, peut expliquer l'asymétrie de discrimination observée. Cependant, il est difficile de rejeter l'influence d'une connaissance conceptuelle des styles musicaux utilisés, antérieure à l'expérience.

Une étude plus récente apporte des éléments de réponses concernant l'origine de l'asymétrie perceptive des styles musicaux. Dalla Bella & Peretz (2005) observèrent un effet similaire chez des auditeurs non occidentaux, très peu exposés à cette variété de styles. Dans cette étude, des auditeurs musiciens et non-musiciens occidentaux, ainsi que des participants non-musiciens chinois, devaient évaluer la similarité d'extraits musicaux présentés par paire. Les auteurs rapportent que tous les participants jugèrent les extraits comme étant d'autant plus similaires que leurs styles compositionnels étaient historiquement proches. De plus, les styles musicaux étaient considérés par tous les participants comme étant davantage différents lorsqu'ils étaient présentés selon l'ordre historique, l'extrait de style plus ancien précédant l'extrait plus récent (e.g. un extrait Baroque suivi d'un extrait Romantique), que lorsque les extraits étaient présentés dans l'ordre inverse (e.g. Romantique suivi de Baroque). Là encore, le style Baroque semble être perceptivement inclus dans le style Romantique, mais pas l'inverse. Une analyse statistique *a posteriori* des stimuli révéla que, dans cette expérience,

l'effet de distance historique semblait lié à la dimension rythmique de la musique, i.e. à la variabilité des durées des notes.

A la lumière de ces résultats et de l'effet de catégorisation visuelle asymétrique observé chez les jeunes enfants, nous avons décidé d'explorer ce phénomène dans un environnement acoustique contrôlé, grâce à des stimuli construits artificiellement et présentés à des participants non-musiciens.

2 Expérience 1

La perception des hauteurs est fondamentale pour la mélodie en musique et la prosodie dans le langage. Cependant, lorsqu'ils mémorisent une mélodie, les auditeurs adultes ne représentent pas celle-ci en tant que série de notes indépendantes, mais traitent chaque note en fonction des autres. En fait, les relations entre hauteurs sont plus importantes que les valeurs de hauteurs absolues des notes utilisées. Ainsi, une mélodie sera reconnue facilement même si elle est transposée, c'est-à-dire si la hauteur de ses différentes notes est modifiée par une valeur constante, opération qui préserve la nature des intervalles (i.e. la relation entre deux notes successives) de la mélodie. La prévalence de l'intervalle dans le traitement et la mémorisation de séquences mélodiques est déjà présente chez le jeune enfant de 20 mois (Plantinga & Trainor, 2005).

L'objectif de cette première expérience était de montrer un effet de la distribution des intervalles musicaux dans la formation de catégories mélodiques. Dans ce but, nous avons comparé l'exclusivité de catégories séquentielles auditives formées par exposition à des exemplaires de courtes mélodies, statistiquement contrôlées en terme de distribution d'intervalles, et ne se conformant à aucun idiome musical particulier. Dans cette expérience nous avons tenté de reproduire l'inclusion catégorielle et les relations de distribution qui produisirent les résultats de catégorisation asymétrique de Quinn et al. (1993, cf. Figure 1).

2.1 Participants

51 étudiants en psychologie de l'université de Bourgogne, tous non-musiciens, prirent part à l'expérience.

2.2 Matériel

Pour chaque participant, deux jeux de mélodies furent créés, l'un obéissant à une distribution *étroite*, l'autre à une distribution *large* de la probabilité d'occurrence de 11 intervalles excluant l'octave et l'unisson (voir Figure 2)¹. Afin de contrebalancer l'effet potentiel de saillance de certains intervalles dans le processus de catégorisation mélodique, les intervalles spécifiques associés à chaque distribution de probabilités furent variés en fonction des participants ; ainsi, les intervalles associés aux indices 1 à 11 de la figure 2 étaient choisis aléatoirement pour chaque participant. Chaque mélodie fut construite à partir de cinq intervalles consécutifs (c'est-à-dire six notes), choisis aléatoirement (sans répétition consécutive) selon l'une des deux distributions. Chaque jeu de mélodies consistait en 72 exemplaires. La note de départ de ces mélodies fut contrebalancée entre les 12 classes de hauteurs (six mélodies furent donc construites à partir de chacune de ces 12 notes). Une des deux distributions (large vs étroite) fut attribuée aléatoirement à chaque participant².

¹ Dans le système chromatique, 13 intervalles mélodiques peuvent être définis à partir de chacune des notes. Parmi ceux-ci, l'unisson et l'octave ont un statut perceptif particulier, car ils sont perçus comme une forme de répétition.

² Des exemplaires des stimuli utilisés dans les expériences 1 et 2 sont téléchargeables à partir de : <http://leadserv.u-bourgogne.fr/~projetweb/spip/article728.html>

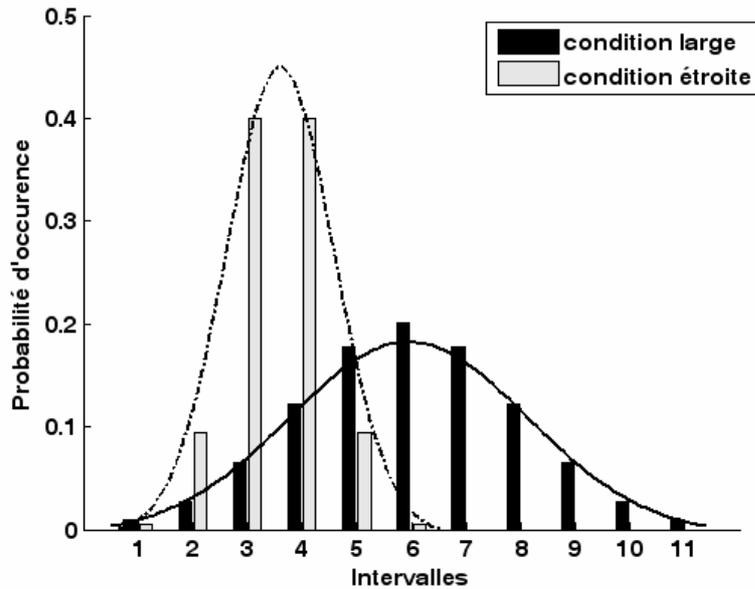


Figure 2. Probabilités d'occurrence des 11 intervalles musicaux sur l'ensemble d'un corpus d'apprentissage, pour les conditions *étroite* (courbe grise) et *large* (courbe noire).

Figure 2. Probabilities of occurrence of 11 musical intervals over a training corpus, for the *narrow* (gray) and *broad* (black) conditions.

Les mélodies, jouées au piano, furent générées grâce à un logiciel de synthèse MIDI Timidity++ (<http://timidity.sourceforge.net/>), dont le paramétrage correspondait à la qualité audionumérique la plus fréquente dans les études de la perception musicale (44.1 kHz, 16 bits). Afin d'éviter une lassitude des sujets, et d'augmenter la musicalité des stimuli, la durée de chaque note fut aléatoirement fixée à 250 ou 500 ms. Le contour, c'est-à-dire le pattern d'intervalles ascendants (plus aigu) et descendants (plus grave) entre deux notes consécutives, est une autre caractéristique d'une importance particulière dans la perception mélodique (Dowling & Harwood, 1986). Dans notre étude, le contour de chaque mélodie était aléatoire, et il n'y avait pas de répétition des patterns de contour à l'intérieur du même jeu de mélodies.

2.3 Procédure

Chaque participant fut exposé à 84 séquences de notes. Les 60 premières séquences – qui constituaient la phase d'apprentissage – étaient des exemplaires d'une des deux distributions possibles. Les 24 autres mélodies – qui constituaient la phase de test – étaient composées de 12 nouvelles mélodies suivant la distribution d'entraînement, et de 12 autres séquences suivant l'autre distribution. Ces 24 dernières séquences étaient ordonnées aléatoirement.

Un programme de présentation a été écrit dans le langage de programmation MATLAB. Les participants entendaient les mélodies grâce à des écouteurs. A la fin de la présentation de chaque séquence, les participants devaient indiquer si ils pensaient avoir déjà entendu cette séquence au cours de la session. L'intervalle inter-essai entre la réponse du participant et le début de la mélodie suivante était de 2 sec.

2.4 Résultats

Pour chaque participant, le taux d'acceptation (i.e. le pourcentage de réponses positives) de chacun des types de séquences de test fut calculé. La figure 3 illustre les taux d'acceptations associés aux nouvelles séquences *étroites* et *larges*, en fonction de la condition d'apprentissage. Une ANOVA mixte à 2 facteurs (condition d'entraînement étroite/large x séquences de test étroites/larges) révéla un effet principal significatif de la distribution d'apprentissage ($F(1,49) = 4.51$; $MSE = 1363.6$; $p < .05$), ainsi qu'un effet tendanciel du type de séquences de test ($F(1,49) = 2.87$; $MSE = 834.01$; $p = .09$), et une interaction tendancielle entre ces facteurs ($F(1,49) = 2.76$; $MSE = 801.93$; $p = .10$). Cependant, des tests *t* appariés bilatéraux révélèrent que seuls les participants de la condition étroite ont accepté significativement plus de séquences de leur distribution d'apprentissage (51.9%) que de l'autre (40.7%), $t(25) = 2.41$, $p = .023$. En d'autres termes, de façon similaire aux études de Quinn et al. (1993), les participants familiarisés à des stimuli issus de la condition étroite (i.e. de faible variabilité) ont exclu plus de séquences contrastantes (dans ce cas, des séquences issues de la distribution large), alors que les participants familiarisés à des stimuli de la condition large (i.e. de grande variabilité) ne rejetèrent pas les séquences étroites de manière supérieure au hasard. Cela suggère que le groupe exposé à des mélodies de variabilité étroite a appris des aspects des statistiques d'intervalles de ces séquences, lui permettant de reconnaître de nouvelles mélodies formées à partir de cette distribution, ce qui réplique dans le domaine auditif les relations d'inclusion et d'exclusion observées par Quinn et al. (1993).

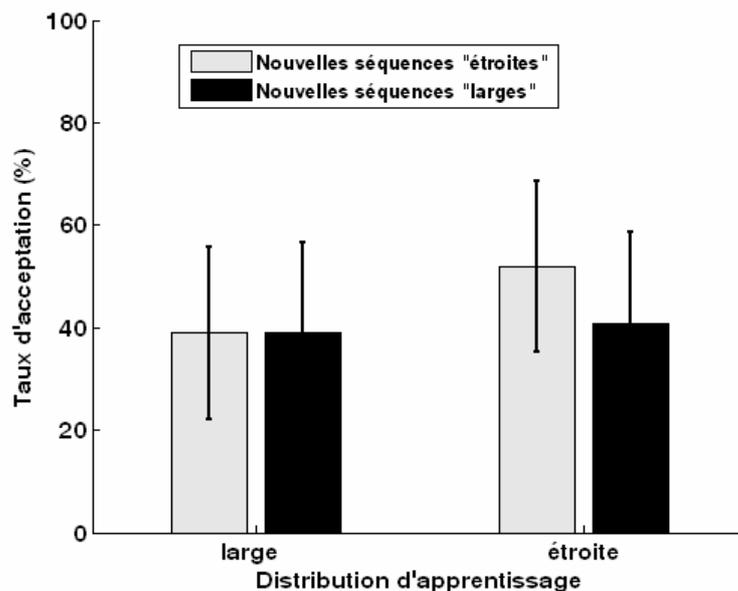


Figure 3. Taux d'acceptation moyens et écarts-types pendant la phase de test pour les groupes *large* (gauche) et *étroit* (droite).

Figure 3. Mean endorsement rate during test phase for the *broad* group (left) and the *narrow* group (right). Error bars represent standard errors.

3 Expérience 2

A la lumière des résultats de l'expérience 1, nous avons décidé de renforcer les contraintes des propriétés statistiques des séquences de notes formant les mélodies des catégories large et étroite. De plus, pendant la phase d'entraînement, une tâche de

mémorisation plus incidente fut utilisée, ainsi qu'un paradigme à choix forcé à deux alternatives lors de la phase de test, afin d'éviter des biais de réponses négatives ou positives.

3.1 Participants

24 étudiants en Psychologie de l'université de Bourgogne, tous non-musiciens et n'ayant pas participé à l'expérience 1 prirent part à cette expérience.

3.2 Matériel

Les mélodies furent générées par un processus de Markov, produisant des séquences hautement contraintes, dans lesquelles la probabilité d'un événement particulier i dépend de l'occurrence d'un événement précédent. Nous avons utilisé un modèle de Markov d'ordre 1, qui peut être représenté par une matrice de probabilités transitionnelles. Plus spécifiquement, la probabilité d'occurrence d'un événement particulier dépend uniquement de l'évènement précédant dans la séquence.

Tableau I. Probabilités transitionnelles entre les hauteurs, pour la condition large (Figure 4, courbe noire).

Table I. Transitional probabilities between pitches, for the *broad* distribution (in black in Figure 4).

		Vers la note						
		fin	1	2	3	4	5	6
De la note	début	.1	.25	.15	.15	.2	.2	.1
	1	.2	.15	.15	.1	.25	.1	.2
	2	.1	.1	.15	.15	.2	.2	.25
	3	.1	.2	.15	.15	.2	.1	.25
	4	.1	.2	.2	.15	.15	.25	.1
	5	.1	.25	.15	.1	.15	.2	.2
	6	.1	.1	.25	.2	.15	.2	.15

Une séquence est construite en sélectionnant une première note selon les probabilités de la ligne "début" du tableau I, puis en sélectionnant une seconde note en fonction de sa probabilité d'occurrence après cette première note, et en répétant ce processus jusqu'à ce que la colonne "fin" soit atteinte. Les probabilités transitionnelles entre une note et les six possibles suivantes (plus le code "fin") suivaient soit une distribution "large", soit une distribution "étroite" (Figure 4)

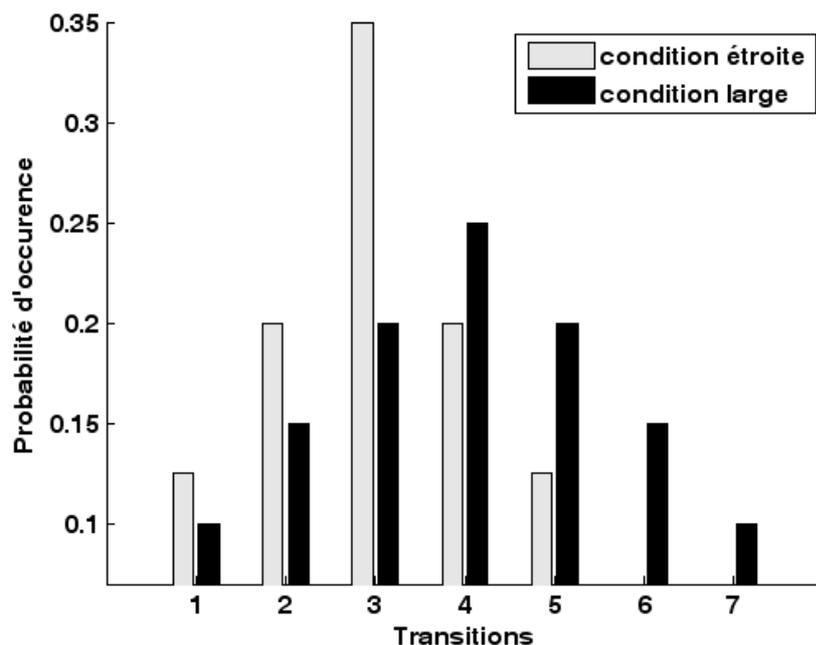


Figure 4. Probabilités d'occurrence des 7 transitions (i.e. une ligne du tableau I), pour la condition *étroite* (courbe grise) et la condition *large* (courbe noire).

Figure 4. Probabilities of occurrence of the 7 transitions (i.e. one row of table I), in *narrow* (gray) and *broad* (black) conditions.

Les notes correspondant aux indices 1 à 6 du tableau I furent associées à quatre différents jeux de six notes : C4/D4/D#4/F4/G#4/A#4, D#4/F4/F#4/G#4/B4/C#4, A3/B3/C4/D4/F4/G4 ou F#4/G#4/A4/B4/D4/E4, afin d'empêcher des effets spécifiques dus à l'un ou l'autre des jeux de hauteurs. Notez que les intervalles musicaux entre deux notes adjacentes sont identiques pour les quatre jeux de hauteurs.

80 mélodies, dont le nombre de notes variait de 4 à 7 ($M_{\text{étroit}} : 5$, $M_{\text{large}} : 4.7$), furent générées pour chacune des deux distributions. La durée des notes et le contour des mélodies furent contrôlés de la même manière que pour l'expérience 1. 60 séquences furent utilisées comme exemplaires pendant la phase d'apprentissage. Pendant la phase de test, 20 paires de mélodies furent présentées aux participants, composées des 20 séquences restantes de chacune des distributions. Chaque paire de test était composée d'une mélodie de chaque distribution, sélectionnée aléatoirement, l'ordre des séquences à l'intérieur des paires étant contrebalancé.

3.3 Procédure

Les conditions de passation étaient identiques à celles de l'expérience 1. Dans une première phase, les participants devaient écouter chaque mélodie, et indiquer le nombre total de notes dans la séquence. Un retour était donné après chaque réponse, indiquant si celle-ci est correcte, ainsi que le nombre exact de hauteurs en cas d'erreur. L'intervalle inter-essai entre la réponse du sujet et le début de la mélodie suivante était de 2 sec. Dans une seconde phase, les participants entendirent des paires de mélodies. Ils durent alors sélectionner, pour chaque paire, la mélodie la plus similaire à celles qu'ils avaient entendues dans la première phase. Les deux groupes étaient composés de 12 participants.

3.4 Résultats

Les données obtenues lors de la phase d'apprentissage furent tout d'abord analysées. Le nombre moyen de réponses correctes pour chaque participant était élevé pour les 2 groupes (large : 88 %, écart-type : 7.5; étroit : 83.3 %, écart-type : 11.1). La tâche de dénombrement de notes était donc relativement facile pour les participants. Leur stratégie semble généralement avoir consisté en un rappel de la mélodie, note par note, après son audition. Un test t avec le groupe en variable indépendante et le pourcentage de réponses correctes pendant la phase d'apprentissage comme variable dépendante ne montre aucun effet significatif du groupe, $t(22) = 1.21$, ns . Ainsi, les mélodies des deux conditions ne furent pas traitées différemment, en terme de dénombrement de notes.

Les analyses suivantes portèrent sur la phase de test. Là encore, nos résultats ressemblent fortement à ceux des expériences de catégorisation chez les enfants de l'expérience de Quinn et al. (1993). La performance du groupe large (49.2 %) n'était pas significativement différente du hasard, $t(11) = -0.3$, ns , alors que celle du groupe étroit (57.5 %) l'était, $t(11) = 3.95$, $p = .023$ (Figure 5).

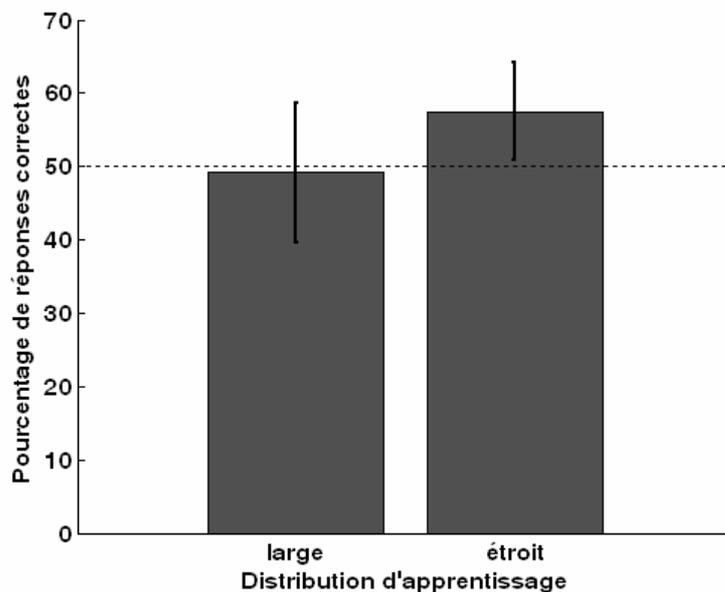


Figure 5. Pourcentages de réponses correctes et écarts-types pendant la phase de test pour les groupes *étroit* et *large*.

Figure 5. Percentages of correct responses during test phase for the *broad* group and the *narrow* group. Error bars represent standard errors.

Ces résultats suggèrent que les participants du groupe étroit ont internalisé la distribution statistique des intervalles des séquences, ce qui leur a permis de reconnaître les nouvelles mélodies créées à partir de cette distribution, alors que les participants de la condition large n'ont pas montré de préférence pour l'un ou l'autre type de mélodies de test.

3.5 Discussion

L'expérience 2 réplique l'effet observé dans l'expérience 1. Il semble peu probable que celui-ci soit la conséquence d'une connaissance de type syntaxique antérieure à l'expérience, car bien que le matériel utilisé dans ces expériences était construit à partir d'une dimension psychoacoustique très familière, la hauteur, l'organisation de ces notes en séquences ne suivait aucun idiome musical particulier. Cependant, afin de confirmer que ces résultats ne

proviennent pas de l'influence d'une connaissance musicale, nous avons cherché à produire cet effet d'asymétrie avec un modèle connexionniste du traitement de séquence, opérant sur des représentations sensorielles calculées par un modèle du système auditif périphérique, et ne possédant aucune connaissance de type syntaxique.

4 Simulations

Le but de la présente simulation est de montrer qu'un modèle cognitif traitant l'information de hauteur de façon purement ascendante, et ne possédant aucune connaissance séquentielle, peut expliquer les résultats de nos expériences.

Le choix de la représentation des stimuli est une question particulièrement non triviale dans le cas de la modélisation de la cognition musicale. La représentation classique des notes, basée sur les 12 classes de hauteurs chromatiques occidentales, semble peu adaptée au regard des données expérimentales concernant la perception de séquences mélodiques. Par exemple, Schellenberg & Trehub (1996) ont montré que les performances de bébés de 6 et 9 mois dans une tâche de détection de changement dans une mélodie sont meilleures lorsque les intervalles utilisés forment des rapports simples de fréquence entre les notes successives (e.g., 2:1, 3:2), que lorsque ces rapports sont plus complexes (e.g., 45:32). De la même façon, les similarités de timbre et de registre ont un effet sur le type de régularités séquentielles extraites par des auditeurs adultes dans des tâches de segmentation de flux continu de hauteurs (Creel, Newport, & Aslin, 2004; Tillmann & McAdams, 2004). Ces exemples de données comportementales soulignent l'importance des propriétés perceptives des unités séquentielles mélodiques (les hauteurs et intervalles), propriétés qui ne sont pas présentes dans les représentations localistes (orthogonales) de ces unités. Ainsi, une approche écologique de la modélisation de la perception des séquences mélodiques a été choisie dans cette étude, grâce à l'utilisation d'un modèle auditif de la perception de la hauteur.

Le modèle d'apprentissage statistique présenté ici est justifié par les propriétés somatotopiques de la modalité auditive, ainsi que par l'aspect séquentiel des stimuli utilisés dans les précédentes expériences. La somatotopie est une propriété commune à différentes aires du cortex cérébral. Le cortex auditif est un excellent exemple d'aire cérébrale de traitement somatotopique, car il s'organise de façon tonotopique (dépendent de la fréquence) à différents niveaux de traitement. Ainsi, les fibres et neurones auditifs sont organisés de telle façon que des neurones spatialement proches répondront à des bandes de fréquences physiquement proches. Dans le but de modéliser les cartes somatotopiques observées dans le cortex cérébral, Kohonen (1984) a développé les Cartes Auto-Organisées (SOM). Cependant, bien que celles-ci aient produit de bons résultats avec des stimuli statiques, ce modèle original n'est pas conçu pour le traitement dynamique de séquences. Or, la différence principale avec les stimuli des expériences de Quinn et al. réside dans l'aspect séquentiel de notre matériel auditif. Toute tentative de modélisation de l'effet d'asymétrie observé dans les expériences 1 et 2 doit donc prendre en compte les propriétés séquentielles de ce matériel expérimental. De plus, comme noté plus haut, les interactions entre les notes d'une mélodie (et l'interprétation cognitive de ces événements musicaux) sont relationnelles, dans le sens que ces hauteurs musicales acquièrent une signification qui dépend des notes précédentes et suivantes, i.e. du contexte. Des extensions de la carte auto-organisée capables d'extraire et de représenter des dépendances temporelles ont été proposées (Voegtlin, 2002; Strickert & Hammer, 2004). Ces modèles ont en commun le traitement d'une information temporelle séquentielle. En plus des connexions d'excitation et d'inhibition purement spatiales de la carte auto-organisée originale, ils utilisent des connexions récurrentes. Ainsi, ces modèles montrent des capacités de mémorisation des états dus à des stimulations passées, et mettent en oeuvre des processus de reconnaissance de contexte et d'apprentissage auto-organisé, ce qui en fait de bons candidats

pour la modélisation des processus impliqués dans la perception auditive en général, et dans la perception musicale en particulier.

La suite de cet article introduit le modèle du système auditif périphérique puis le modèle de carte auto-organisée temporelle utilisés dans nos simulations.

4.1 Le modèle

4.1.1 Le modèle du système auditif périphérique

La perception mélodique repose sur le traitement précis de hauteurs musicales isolées. Le traitement neuronal de l'information basique de hauteur, indépendamment du contour musical ou de la transposition mélodique, est effectué par le cortex auditif. Il existe deux concepts fondamentaux concernant la façon dont l'information de hauteur d'un signal musical peut être représentée par des patterns de décharges neuronales : le codage par des patterns d'excitation spatialement organisés (la tonotopie), et le codage par les patterns temporels des trains de potentiels d'actions. Les représentations spatialement distribuées sont validées par l'observation que des sous populations particulières des neurones auditifs ne réagissent qu'à des bandes de fréquences particulières. Le taux moyen de décharge des neurones de ces zones tonotopiques fournit une représentation neuronale du spectre fréquentiel du stimulus. Parallèlement, un codage temporel de la hauteur est supporté par de nombreuses études physiologiques, chez l'animal (Palmer, 1992; Bendor & Wang, 2005) comme chez l'homme (Cariani & Delgutte, 1996a, 1996b; Langner, Schreiner, & Biebel, 1998). La représentation temporelle est basée sur les distributions des intervalles de temps séparant les potentiels d'action émis par les neurones auditifs, en réponse au déplacement de la membrane basilaire. De fortes correspondances existent entre les statistiques d'intervalles entre potentiels d'action sur les fibres du nerf auditif et les hauteurs subjectives produites par différents types de sons complexes (Cariani & Delgutte, 1996a). De plus, différents aspects de la perception des hauteurs, tels que le phénomène de fréquence fondamentale absente³, l'équivalence de hauteur de stimuli avec des spectres (timbres) très différents et les hauteurs produites par des harmoniques non résolues par le système auditif périphérique, peuvent être expliqués en terme de distributions d'intervalles au niveau du nerf auditif. Un codage temporel de la hauteur, basé sur la fonction d'auto-corrélation⁴ appliquée à une estimation des probabilités de décharge d'un modèle des neurones auditifs, est employé dans nos simulations.

³ Un son complexe harmonique dont la fréquence fondamentale F0 est physiquement absente du signal acoustique garde pourtant son identité perceptive de hauteur.

⁴ La fonction d'auto-corrélation consiste à comparer un signal avec des versions temporellement décalées de lui-même, afin de déterminer le décalage qui produit la valeur de corrélation la plus importante entre ces deux versions du signal. C'est un indicateur des périodicités, ou répétitions, présentes dans le signal

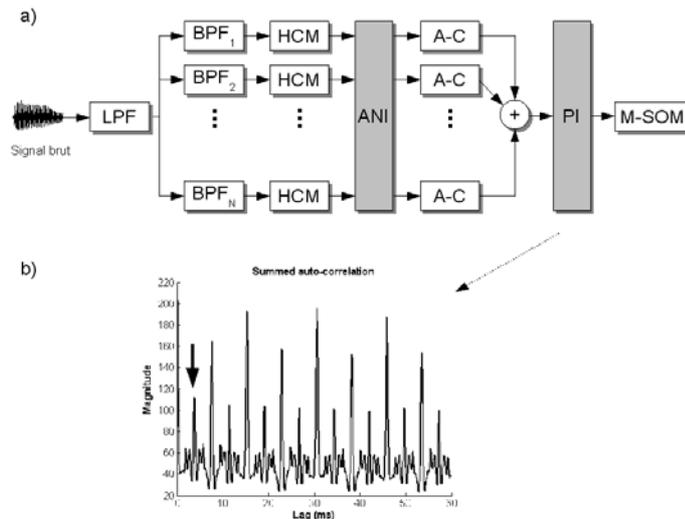


Figure 6. a) schéma du modèle du système auditif périphérique utilisé dans la présente simulation. (LPF = filtre passe-bas, BPF = filtre passe-bande, HCM = modèle des cellules ciliées, ANI = image du nerf auditif, A-C = auto-corrélation, PI = image des périodicités, M-SOM=Merge-SOM). b) exemple de pattern d'auto-corrélation calculé par le modèle en réponse à un son complexe dont la fréquence fondamentale est manquante (dont la période est indiquée par la flèche).

Figure 6. a) schematic diagram of the peripheral auditory model used in the present simulation. (LPF = low-pass filter, BPF = band-pass filter, HCM = Hair cell model, ANI = auditory nerve image, A-C = auto-correlation function, PI = pitch image, M-SOM=Merge-SOM). b) example of an auto-correlation pattern in response to an harmonic complex sound without the fundamental frequency (whose period is shown by the arrow).

Le modèle auditif utilisé dans nos simulations est une adaptation par Leman (2000) du modèle de Van Immerseel & Martens (1992) (Figure 6, pour les détails techniques, voir (Leman)). Différentes étapes de formation et de traitement de représentations des ondes acoustiques sont mises en oeuvre dans ce modèle. Ces signaux sont tout d'abord transformés en représentations neuronales (*rate coding*) par un modèle du système auditif périphérique. Cette première étape de traitement transforme le signal acoustique en un pattern de décharges neuronales dans un certain nombre de fibres auditives, en simulant le filtrage de l'oreille externe et moyenne, la résonance de la membrane basilaire de l'oreille interne et la dynamique des cellules ciliées. L'image auditive résultante est alors traitée par un modèle de la perception des hauteurs des sons complexes.

L'algorithme d'extraction de la hauteur effectue une analyse des périodicités des représentations neuronales de l'image du nerf auditif. Une fonction d'auto-corrélation est appliquée sur chacun des canaux de l'image du nerf auditif. Les patterns de périodicités résultants sont alors additionnés, ce qui produit un pattern d'auto-corrélation unique pour chaque fenêtre temporelle d'analyse. Ainsi, pour chaque note d'une mélodie donnée, un vecteur de périodicités est calculé, vecteur utilisé dans la suite de cette simulation comme représentation d'entrée d'un modèle connexionniste de l'apprentissage de séquences.

4.1.2 Merge-SOM

Les représentations créées par le précédent modèle du système auditif périphérique ont ensuite servi à entraîner un système connexionniste du traitement de séquences. Dans ce but, une modification temporelle de la carte auto-organisée de Kohonen (1984), proposée par Strickert & Hammer (2004), a été utilisée.

Une carte auto-organisée est constituée d'une couche d'entrée et d'une couche topologique, elle-même composée de N neurones positionnés sur une grille rectangulaire, habituellement à 1 ou 2 dimensions (Figure 7). Ceux-ci sont connectés aux neurones adjacents selon une relation de voisinage qui dicte la structure topographique de la carte. Dans le cas de 2 dimensions, les neurones de la carte peuvent être organisés selon un treillis rectangulaire ou hexagonal. De plus, si les côtés de la carte sont connectés, la forme globale de celle-ci devient cylindrique ou toroïdale.

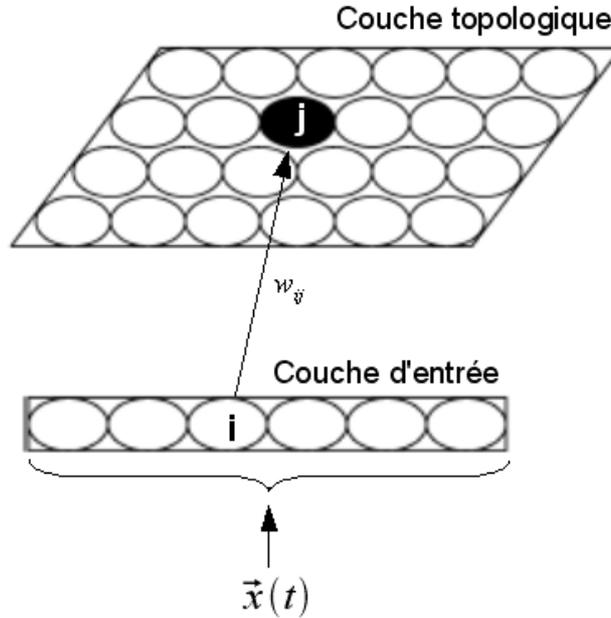


Figure 7. Schéma d'une carte auto-organisée de Kohonen (1984).

Figure 7. Diagram of the Self-Organizing Map (Kohonen, 1984)

Dans l'algorithme original de la carte auto-organisée (Kohonen, 1984), chaque neurone $j \in 1 \dots N$ compare son vecteur-prototype \vec{w}_j à un vecteur d'entrée $\vec{x}(t)$, où t est un index temporel. Cette comparaison est généralement basée sur la distance euclidienne. Pour un vecteur d'entrée donné, l'erreur de représentation du neurone j est définie par :

$$err_j = \sum_i \|x_i(t) - w_{ij}\|_2 \quad (1)$$

Le neurone vainqueur v , c'est-à-dire le neurone de la couche topologique qui représente le mieux le vecteur d'entrée actuel, est celui qui minimise cette erreur :

$$err_v = \min_{j \in 1 \dots N} err_j \quad (2)$$

L'algorithme d'apprentissage modifie alors les poids des connexions du neurone vainqueur ainsi que ceux de ses voisins, en direction du vecteur d'entrée :

$$\Delta w_{ij} = \eta \cdot h_{jv}^{(t)} \cdot (x_i^{(t)} - w_{ij}) \quad (3)$$

où v est l'index du neurone vainqueur, η un taux d'apprentissage et h_{jv} une fonction de voisinage, qui décroît avec la distance entre les unités j et v . Cette fonction de voisinage, qui définit la force des connexions entre neurones adjacents, est classiquement gaussienne :

$$h_{jv}^{(t)} = \exp\left(-\frac{d(j,v)^2}{\sigma(t)^2}\right) \quad (4)$$

où $d(j,v)$ correspond à la distance des neurones j et v sur la carte, et σ est l'écart, qui définit la taille de la fonction de voisinage. Plus le neurone j est spatialement proche de v , plus ce dernier aura d'influence sur le neurone j . Classiquement, la valeur de l'écart est une fonction décroissante de t . La réduction de ce paramètre, nécessaire pour que le réseau s'organise de façon topologique, peut être mise en relation avec la diminution de la plasticité cérébrale au cours du développement.

Dans la Merge-SOM (Strickert & Hammer, 2004), chaque neurone j de la carte auto-organisée est représenté par, en plus du vecteur-prototype \vec{w}_j , un vecteur-contexte \vec{c}_j .

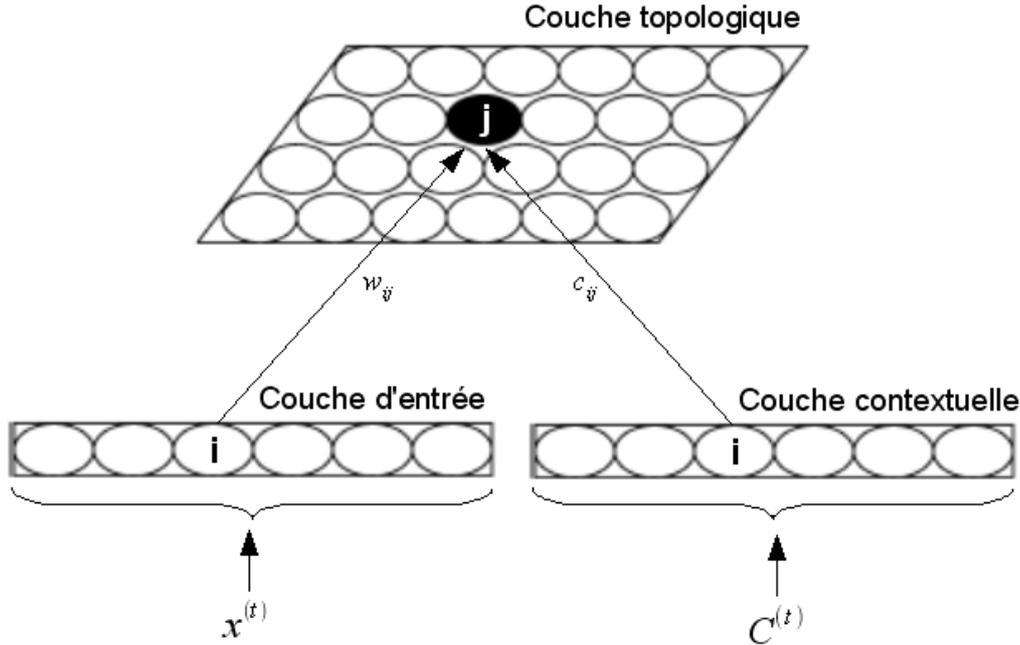


Figure 8. Schéma d'une Merge-SOM (Strickert & Hammer, 2004)

Figure 8. Diagram of the Merge-SOM (Strickert & Hammer, 2004)

Le descripteur de contexte au temps t , désigné par $\vec{C}^{(t)}$, fait référence à la fusion de deux propriétés caractérisant le neurone vainqueur à $t-1$: le vecteur-prototype et le vecteur-contexte du dernier neurone vainqueur v sont associés par une combinaison linéaire pondérée⁵ :

$$C_i^{(t)} = (1 - \beta) \cdot w_{iv}^{(t-1)} + \beta \cdot c_{iv}^{(t-1)} \quad (5)$$

Le paramètre β contrebalance l'influence réciproque des propriétés (i.e. le vecteur-prototype et le vecteur-contexte) de la cellule vainqueur au temps $t-1$ dans le calcul du contexte temporel de la carte.

Lors du calcul de l'erreur d'un neurone j , les deux vecteurs \vec{w}_j et \vec{c}_j sont respectivement comparés au vecteur d'entrée $\vec{x}^{(t)}$ et au descripteur de contexte $\vec{C}^{(t)}$ du réseau. L'équation 1 devient alors :

$$err_j = (1 - \alpha) \cdot \sum_i \|x_i^{(t)} - w_{ij}\|_2 + \alpha \cdot \sum_i \|C_i^{(t)} - c_{ij}\|_2 \quad (6)$$

⁵ Notez que la couche contextuelle C n'est pas connectée à l'extérieur du modèle.

Le paramètre α contrebalance l'effet respectif des deux valeurs de distances dans la détermination du vainqueur. Ainsi, l'erreur d'un neurone en réponse à une entrée donnée est à la fois fonction de l'adéquation du symbole d'entrée avec le vecteur-prototype et du contexte actuel de la carte avec le vecteur-contexte de ce neurone. Comme dans l'algorithme original de la SOM, la cellule vainqueur v pour l'élément $\bar{x}^{(t)}$ correspond au neurone qui minimise l'erreur err de l'équation 6.

Dans les simulations suivantes, l'erreur moyenne $E(x)$ associée à une séquence donnée x correspondait à la moyenne des erreurs des neurones vainqueurs associés à chaque élément t de cette séquence :

$$E(x) = \frac{\sum_t err_v(x^{(t)})}{|x|} \quad (7)$$

où v est l'index du neurone vainqueur pour l'élément t de la séquence x . Enfin, les poids des connexions du neurone j sont mis à jour selon l'équation 8, où η est un taux d'apprentissage :

$$\begin{aligned} \Delta w_{ij} &= \eta \cdot h_{jv}^{(t)} \cdot (x_i^{(t)} - w_{ij}) \\ \Delta c_{ij} &= \eta \cdot h_{jv}^{(t)} \cdot (C_i^{(t)} - c_{ij}) \end{aligned} \quad (8)$$

4.1.3 Calibration des réseaux

Les participants adultes des expériences présentées précédemment ont vraisemblablement été très fréquemment exposés à des exemplaires de notes musicales, et en ont donc formé des représentations mentales, dont la tonotopie est un des aspects majeurs, comme souligné plus haut. Le but de cette phase de calibration était donc d'initialiser les réseaux MSOM dans l'espace des représentations d'entrée, c'est-à-dire des vecteurs de périodicités calculés par le modèle auditif en réponse à des notes musicales.

Dans toutes nos simulations, les stimuli étaient des signaux acoustiques, numérisés à 16kHz, avec une résolution de 16 bits. Après traitement par le modèle auditif, un événement musical (une note) était représenté par un vecteur de 109 dimensions, correspondant à la magnitude de différentes périodes estimées par le modèle auditif, moyennée sur la durée du signal musical (750 ms par note). la couche d'entrée des réseaux était constituée de 109 cellules, correspondant au nombre de dimensions des vecteurs produits par le modèle auditif en réponse à un stimulus. La couche topologique formait une grille de 20 x 20 cellules, organisées de façon toroïdale. Les poids des connexions de chaque réseau furent initialisés aléatoirement. Les paramètres α et β des équations 5 et 6 furent tous deux fixés à 0.5. Le taux d'apprentissage η était de .05 pour toutes les simulations. L'écart σ de l'équation 4, définissant la taille de la fonction de voisinage, décroissait au cours de la phase de calibration (de 10 à 0.5), mais restait fixée à 0.5 pendant les simulations des expériences 1 et 2.

Pour calibrer les réseaux, les vecteurs de périodicités correspondant aux notes de piano E2 (F0=82.4 Hz) à D#5 (F0=622.3 Hz) ont été utilisés. 400 blocs de ces 36 notes ordonnées aléatoirement furent présentés aux modèles. L'activité des unités contextuelles était réinitialisée entre chaque note, ce qui empêchait l'apprentissage d'une quelconque connaissance de type séquentielle avant la simulation des expériences 1 et 2.

Afin d'évaluer la qualité des représentations des exemplaires de notes musicales par les réseaux pré-entraînés, nous avons déterminé pour chaque carte la zone qui répond spécifiquement à chacune de ces notes. En particulier, l'erreur de l'ensemble des neurones en

7). Pour chaque réseau, T et k étaient ajustés manuellement afin de donner un même nombre de réponses « oui » que de réponses « non » lors de la phase de test. Cette fonction fut appliquée à l'erreur obtenue pour chacune des séquences de test (étroite ou large), et une probabilité $p(\text{"oui"})$ supérieure à .5 fut considérée comme une réponse positive.

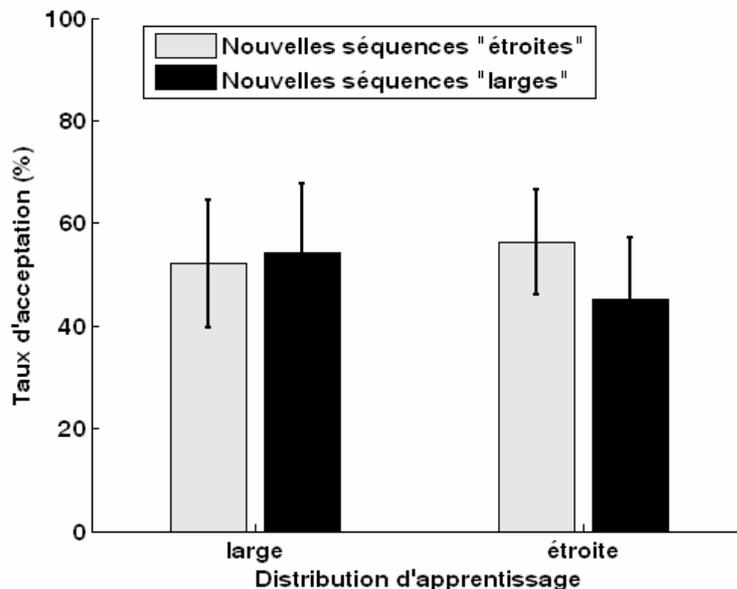


Figure 10. Taux d'acceptation moyens pendant la phase de test pour les réseaux entraînés dans les conditions *étroite* et *large* de l'expérience 1. Les barres d'erreurs représentent les écart-types.

Figure 10. Mean endorsement rate during test phase for the *broad* and *narrow* trained-networks. Error bars represent standard errors.

L'analyse des taux d'acceptation des nouvelles séquences pendant la phase de test révèle un effet principal des items de test ($F(1,48) = 5.44$; $MSE = 836.22$; $p < .05$), ainsi qu'une interaction tendancielle entre la condition d'apprentissage et le type de séquence de test ($F(1,48) = 3.69$; $MSE = 567.12$; $p = .06$). Cependant, de façon similaire aux résultats expérimentaux de l'expérience 1, des test t appariés montrent que seuls les réseaux exposés à une distribution étroite ont accepté significativement plus de séquences de leur distribution d'apprentissage que de l'autre ($t(24) = 2.41$, $p < .01$, Figure 10).

4.3 Simulation de l'expérience 2

Deux nouveaux groupes de 12 réseaux pré-calibrés furent entraînés avec un matériel donc les contraintes statistiques étaient strictement identiques à celles de l'expérience 2. Là encore, il n'y avait pas de différence de paramétrage entre la phase d'exposition et la phase de test.

Les items de test furent présentés par paire, l'ordre des items à l'intérieur d'une paire étant aléatoire. L'activité des unités contextuelles des réseaux était réinitialisée à 0 entre les paires de test, mais pas entre les items d'une même paire, afin de refléter la procédure de choix forcé à deux alternatives employée dans l'expérience 2. A l'intérieur de chaque paire de test, l'erreur de représentation moyenne fut calculée pour chaque item. La séquence présentant l'erreur la plus faible correspondait au choix du réseau. Nous avons ensuite calculé pour chaque réseau le pourcentage de réponses correctes.

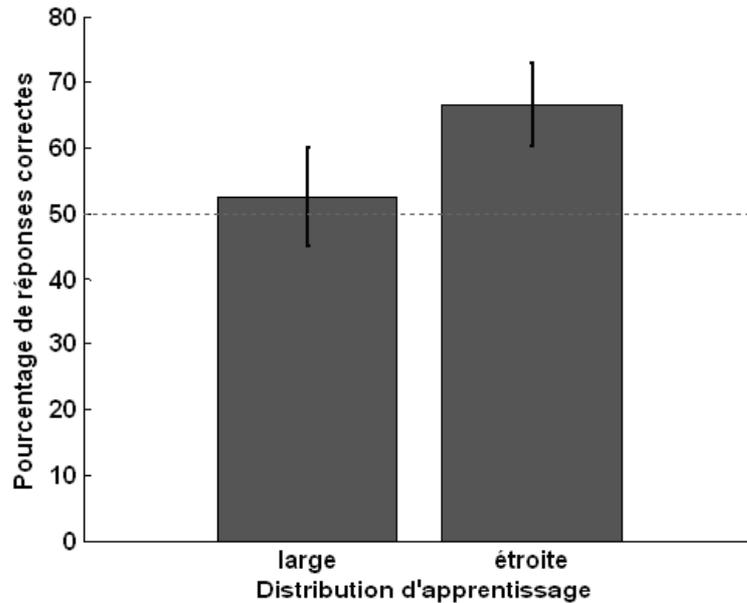


Figure 11. Pourcentages moyens de réponses correctes pendant la phase de test pour les réseaux entraînés dans la condition *étroite* et *large* de l'expérience 2. Les barres d'erreurs représentent les écart-types.

Figure 11. Mean scores during test phase for the *broad* and *narrow* trained-networks. Error bars represent standard errors.

Les scores moyens des réseaux, illustrés figure 11, ressemblent fortement à ceux de l'expérience 2; les réseaux entraînés avec des exemplaires tirés de la catégorie large ne distinguent pas les nouveaux exemplaires de chaque catégorie de manière significativement supérieure au hasard ($t(11) = 1.48, ns.$), alors que les réseaux entraînés avec des items de la catégorie étroite présentent une erreur inférieure pour les séquences étroites, associée à un taux de reconnaissance correcte plus important ($t(11) = 11.7, p < .001$).

Les réseaux ne possédant aucune connaissance *a priori* du matériel que nous avons utilisé, nous pouvons conclure qu'un modèle connexionniste simple, qui traite les stimuli de façon purement ascendante, est suffisant pour rendre compte de nos résultats comportementaux.

5 Conclusions

Ces résultats préliminaires suggèrent que la catégorisation asymétrique observée chez les jeunes enfants par Quinn et al. (1993) n'est pas limitée au domaine visuel, il est probable que ce phénomène s'applique aussi à la perception auditive. En effet, les simulations présentées, si elles n'apportent pas la preuve irréfutable que des processus ascendants étaient à l'oeuvre dans nos expériences, soutiennent l'hypothèse que l'effet d'asymétrie observé trouve son origine dans les distributions des intervalles des mélodies employées. Nos résultats soulignent donc l'importance des traitements ascendants (i.e. statistiques) dans la perception et la catégorisation de stimuli séquentiels auditifs par des non musiciens. De plus, ces simulations montrent que les représentations catégorielles qui présentent ces propriétés d'inclusion et d'exclusion émergent dans un système simple de traitement de séquences de hauteurs, sans recours à des schémas mentaux stockés en mémoire à long-terme.

L'aspect écologique de la modélisation auditive employée dans cette étude permettra d'établir des prédictions quant à la formation de catégories mélodiques basées sur d'autres systèmes de relations de hauteurs que l'échelle chromatique, tels que la musique micro-tonale ou d'autres systèmes d'intonation (non tempérés).

Les résultats comportementaux rapportés ici se limitent à une seule dimension auditive (les hauteurs). Il sera aussi intéressant d'investiguer l'influence des nombreuses autres dimensions de la perception musicale (e.g., la durée, le rythme, le timbre, la spatialisation, etc.) ainsi que les interactions entre ces différentes dimensions dans les processus de formation de catégories musicales.

Remerciements

Cette étude a été partiellement financée par une bourse de la Commission Européenne FP6 NEST « PathFinder » n°516542 attribuée au 2^{ème} auteur.

Bibliographie

- Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054), 1161-1165.
- Bigand, E., & Barrouillet, P. (1996). Processi di classificazione degli stili nei bambini e negli adulti. *Quaderni della SIEM. semestrale di ricerca e didattica musicale*, 10(1), 81-93.
- Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 73(3).
- Cariani, P. A., & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. *Journal of Neurophysiology*, 76(3).
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1119-1130.
- Dalla Bella, S., & Peretz, I. (2005). Differentiation of classical music requires little learning but rhythm. *Cognition*, 96, 65-78.
- Dienes, Z. (1992). Connectionist and Memory-Array Models of Artificial Grammar Learning. *Cognitive Science*, 16, 41-79.
- Dowling, J., & Harwood, D. (1986). *Music Cognition*. Orlando, FL: Academic Press.
- Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology*, 58, 418-431.
- French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The Role of Bottom-up Processing in Perceptual Categorization by 3- to 4-month-old Infants: Simulations and Data. *Journal of Experimental Psychology-General*, 133(3), 382-397.
- French, R. M., Mermillod, M., Quinn, P., & Mareschal, D. (2001). *Reversing Category Exclusivities in Infant Perceptual Categorization: Simulations and Data*. Paper presented at the Proceedings of the 23rd Annual Conference of the Cognitive Science Society, NJ.
- Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science*, 298, 2167-2170.
- Kohonen, T. (1984). *Self-Organizing Maps*. Berlin: Springer-Verlag.
- Langner, G., Schreiner, C. E., & Biele, U. W. (1998). Functional implications of frequency and periodicity coding in auditory midbrain. In A. R. Palmer, A. Rees, A. Q. Summerfield & R. Meddis (Eds.), *Psychophysical and Physiological Advances in Hearing* (pp. 277-285). London: Whurr.
- Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, 17(4), 435-463.
- Mareschal, D., & French, R. M. (1997). *A connectionist account of interference effects in early infant memory and categorization*. Paper presented at the Proceedings of the 19th Annual Cognitive Science Society Conference, New Jersey.
- Mareschal, D., French, R. M., & Quinn, P. (2000). A Connectionist Account of Asymmetric Category Learning in Early Infancy. *Developmental Psychology*, 36, 635-645.
- Palmer, A. R. (1992). Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. In M. E. H. Schouten (Ed.), *The Auditory Processing of Speech* (pp. 115-124). Berlin: Mouton de Gruyter.
- Plantinga, J., & Trainor, L. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, 98, 1-11.

- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*, 463-475.
- Schellenberg, E. G., & Trehub, S. (1996). Natural musical intervals: evidence from infant listeners. *Psychological Science, 7*, 272-277.
- Strickert, M., & Hammer, B. (2004). *Self-Organizing Context Learning*. Paper presented at the European Symposium on Artificial Neural Networks (ESANN 2004).
- Tillmann, B., & McAdams, S. (2004). Implicit Learning of musical timbre sequences : statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, memory and cognition, 30*(5), 1131-1142.
- Van Immerseel, L., & Martens, J. P. (1992). Pitch and voiced/unvoiced determination with an auditory model. *Journal of the Acoustical Society of America, 91*, 3511-3526.
- Voegtlin, T. (2002). Recursive Self-Organizing Maps. *Neural Networks, 15*(8-9), 979-991.