10ème **C**ongrès **F**rançais d'**A**coustique
Lyon, 12-16 Avril 2010

Le liage de sons successifs par le système auditif

Laurent Demany

UMR CNRS 5227 (Mouvement, Adaptation, Cognition), Université de Bordeaux, 146 rue Leo-Saignat, F-33076 Bordeaux
laurent.demany@u-bordeaux2.fr

Un phénomène auditif surprenant a récemment été découvert : il est possible d'identifier la direction d'un changement de fréquence entre deux sons purs successifs dans des conditions où le premier des deux sons n'a pas été perçu consciemment. Ce phénomène semble indiquer que le système auditif contient des détecteurs automatiques de changements spectraux ("frequency-shift detectors", FSD). La fonction principale des FSD est hypothétiquement de créer un lien perceptif entre des sons successifs qui, bien que différents par le spectre, émanent probablement d'une même source acoustique. Ces détecteurs pourraient ainsi jouer un rôle important dans l'analyse des scènes auditives. Il existe à ce jour un assez grand nombre de données psychophysiques permettant de spécifier les propriétés des FSD, ainsi que celles de la mémoire auditive implicite qu'ils exploitent.

1 Introduction

« Emportés et modelés par l'universel changement, nous en sommes aussi les témoins parce que nous le percevons comme changement. Cette perception est possible dans la mesure où nous saisissons en une relative simultanéité plusieurs phases successives du changement, qui apparaissent ainsi liées. »

Ces phrases de Paul Fraisse [1] s'appliquent à la perception visuelle de même qu'à la perception auditive. Mais dans notre environnement quotidien, les changements visuels diffèrent en général des changements auditifs. Les premiers sont typiquement des déplacements continus, des mouvements. Les seconds peuvent être également continus (exemple : le spectrogramme d'une diphtongue), mais c'est typiquement à des **séquences** de sons, éventuellement séparés par des silences, que nous attribuons une signification dans le domaine acoustique. Il est donc particulièrement important pour le système auditif (au sens large) de lier des éléments sonores séparés dans le temps.

Ce liage, cependant, doit être sélectif. En effet, notre environnement sonore résulte rarement de l'activité d'une seule source acoustique. Généralement, plusieurs sources sont actives de façon concomitante, et leurs productions sont entremêlées. Il importe bien sûr que ne soient liés dans le temps que des éléments sonores issus d'une même source. C'est un aspect crucial de ce que l'on appelle, à la suite de Bregman [2], "l'analyse des scènes auditives".

L'analyse en question repose en grande partie sur des mécanismes neuronaux automatiques, qui organisent efficacement notre environnement sonore en prenant uniquement en compte les caractéristiques purement physiques des sons. Identifier ces mécanismes est aujourd'hui l'un des objectifs majeurs de la recherche sur l'Audition. A ce sujet, van Noorden [3] avait voici plus de 30 ans proposé une hypothèse qui a été largement ignorée depuis. Il avait suggéré que le système auditif contient des détecteurs automatiques de changements spectraux, ressemblant fonctionnellement aux détecteurs de mouvements spatiaux qui existent dans le système visuel. Ces derniers sont capables de détecter non seulement des mouvements continus mais aussi (sous certaines conditions) des déplacements discontinus. Leur fonction essentielle est d'établir un lien entre des stimuli successifs et de nous donner la capacité de les percevoir comme un seul et même objet [4]. Leurs cousins auditifs, imaginés par van Noorden, ont hypothétiquement une fonction comparable : en liant des éléments sonores successifs, ils pourraient nous rendre capables d'appréhender ces éléments en tant que parties d'un même "objet" ou "flux" sonore, émanant d'une seule source.

Depuis environ cinq ans, un certain nombre d'études psychoacoustiques ont montré que l'hypothèse de van Noorden devait être prise au sérieux. Elles ont fourni des arguments forts en faveur de l'existence de "frequency-shift detectors (FSD)" dans le système auditif. Elles permettent également de spécifier certaines de leurs propriétés. Je vais ici passer en revue les principaux résultats obtenus.

2 Un phénomène auditif paradoxal

Paradoxalement, un auditeur humain s'avère capable d'identifier la direction d'un changement de fréquence entre deux sons purs successifs (même séparés par un silence assez long) **sans avoir pourtant perçu consciemment l'un des deux sons**. Ce phénomène surprenant constitue un argument fort en faveur de l'existence de FSD.

Demany et Ramos [5] ont mis en évidence le phénomène en question de la façon suivante. A chaque essai expérimental, une séquence de deux stimuli, durant chacun 300 ms et séparés par un silence de 500 ms, était présentée au sujet. Le premier stimulus était un "accord" de cinq sons purs synchrones, égaux en amplitude. Les fréquences des cinq composants de l'accord étaient tirées au hasard, mais de façon telle que deux composants ne soient jamais séparés par un intervalle inférieur à 6 demi-tons ; ce dernier point assurait que les composants soient toujours résolus (i.e., séparés tonotopiquement les uns des autres) dans la cochlée. Le second stimulus de la séquence était un simple son pur (*"T"*). Dans l'une des deux conditions expérimentales ("présent/absent", voir le panneau central de la Figure 1), *T* était ou bien identique à l'un des composants de l'accord (sélectionné au hasard parmi les trois composants intermédiaires), ou bien situé exactement à mi-chemin de deux composants (les distances fréquentielles étant mesurées sur une échelle logarithmique) ; le sujet devait indiquer si *T* était présent dans l'accord ou absent de

lui. Dans l'autre condition expérimentale ("up/down", voir le panneau gauche de la Figure 1), *T* était toujours positionné un demi-ton plus haut ou plus bas (au hasard) que l'un des composants de l'accord (sélectionné au hasard parmi les trois composants intermédiaires) ; le sujet devait identifier la direction de ce changement de fréquence. La Figure 2 (a) montre les performances des 11 sujets testés, en termes de l'indice *d'* [6]. On voit que chaque sujet a mieux réussi la tâche up/down que la tâche présent/absent, de façon extrêmement nette en général.

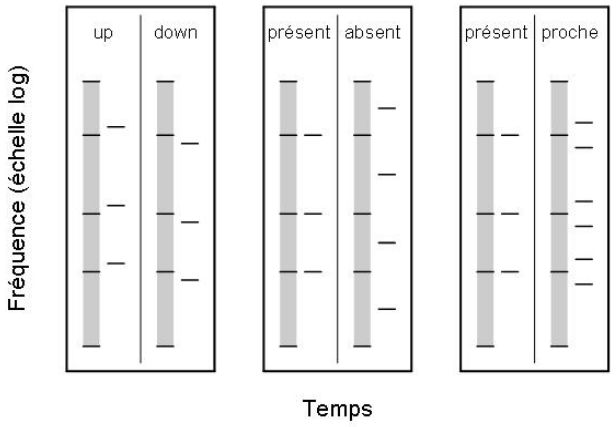

Figure 1. Illustration des trois conditions expérimentales utilisées par Demany et Ramos [5] : "up/down", "présent/absent", et "présent/proche". Chaque segment horizontal représente un son pur et les aires grisées représentent un accord possible. Le son pur unique suivant un accord (après un silence ignoré ici) pouvait prendre 6, 7, ou 9 fréquences possibles, selon la condition expérimentale.

Il était prévisible que, dans cette expérience, la tâche présent/absent serait mal réussie. Plusieurs études antérieures ont en effet montré qu'il est difficile de percevoir individuellement les éléments d'une somme de sons purs synchrones, même lorsque ces sons purs sont résolus dans la cochlée et n'ont pas de relations harmoniques ; les composants d'un tel accord exercent l'un sur l'autre un "masquage informationnel" [7]. Par contre, il était imprévisible que la tâche up/down serait mieux réussie que la tâche présent/absent : le modèle standard de la théorie de la détection du signal [6] prédisait l'issue inverse pour un auditeur idéal capable d'entendre individuellement les composants des accords.

Les sujets de l'expérience ont eu le sentiment qu'ils ne parvenaient pas à percevoir individuellement les composants des accords, et que ce n'était pas nécessaire pour réussir la tâche up/down. Dans la condition up/down, les sujets ont dit percevoir *T* comme l'aboutissement d'un mouvement mélodique dont le point de départ subjectif était l'accord tout entier plutôt que l'un de ses éléments. Cependant, il était permis d'imaginer que l'impression de ne pas entendre individuellement le composant de l'accord proche de *T* était illusoire et que les faibles performances obtenues dans la tâche présent/absent s'expliquaient en réalité par la nature de cette tâche, plus complexe que la tâche up/down du point de vue décisionnel selon la théorie de la détection du signal. Deux autres expériences ont permis d'écarter cette explication.

Dans la première [5], les performances obtenues par quatre sujets dans la tâche présent/absent ont été comparées aux performances de ces mêmes sujets dans une nouvelle tâche, baptisée "présent/proche" (voir le panneau droit de la Figure 1). Dans la tâche présent/proche, le son pur *T* suivant l'accord était ou bien identique à l'un des trois composants intermédiaires de l'accord, exactement comme dans la tâche présent/absent, ou bien positionné 1.5 demi-ton plus haut ou plus bas (au hasard) que l'un des trois composants intermédiaires (sélectionné au hasard) ; le sujet devait, comme dans la tâche présent/absent, juger seulement si *T* était présent ou non dans l'accord. Du point de vue de la théorie de la détection du signal – et en supposant crucialement que les composants des accords étaient perceptibles individuellement – cette tâche présent/proche était de même nature que la tâche présent/absent, mais plus difficile encore du fait que *T* n'était jamais éloigné en fréquence d'un composant de l'accord. Cependant, comme l'indique la Figure 2 (b), l'expérience a montré que la plus difficile des deux tâches était en réalité la tâche présent/absent. Nous verrons un peu plus loin comment un modèle simple de FSD peut rendre compte de ce constat, ainsi que des résultats décrits précédemment.

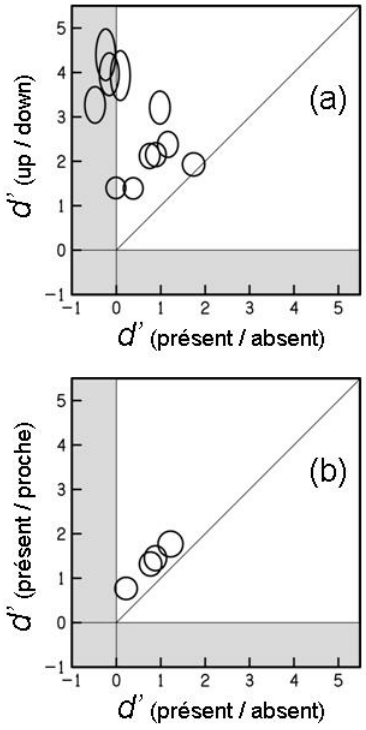

Figure 2. Performance des sujets dans deux expériences de Demany et Ramos [5]. La première (a) utilisait les tâches présent/absent et up/down. La seconde (b) utilisait les tâches présent/absent et présent/proche. Chaque sujet est représenté par une pastille dont le centre correspond à la valeur de *d'* mesurée dans les deux tâches et dont la surface délimite des intervalles de confiance à 95%. Les lignes diagonales indiquent où devraient se trouver les pastilles en cas d'égalité des performances dans les deux tâches.

Une autre expérience [8] a consisté à comparer les difficultés relatives de la tâche présent/absent et de la tâche up/down pour deux types de séquences sonores, où chaque fois un son pur (*T*) était suivi d'un ensemble (*E*) de cinq sons purs espacés de 5.5 demi-tons. Les éléments de *E* étaient soient synchrones (séquences du premier type) soient asynchrones (séquences du second type). Dans le second cas, les éléments de *E* se succédaient dans un ordre aléatoire, avec des écarts temporels ("stimulus onset

asynchrony", SOA) de 100 ms ou 250 ms. Quelle que soit la séquence, l'intervalle de temps séparant *T* du composant médian de *E* était de 1 s. La Figure 3 présente les perfrormances moyennes de cinq auditeurs. On voit que lorsque les éléments de *E* étaient synchrones (SOA = 0), la tâche up/down a été mieux réussie que la tâche présent/absent. Ce résultat est cohérent avec ceux exposés dans la Figure 2 (a). Notons cependant que l'avantage de la tâche up/down a été moins prononcé dans la nouvelle expérience. La raison en est probablement que, cette fois, *T* était présenté non pas pas après l'accord mais avant lui : cela permettait aux sujets de mieux percevoir individuellement un composant de l'accord identique à *T*.

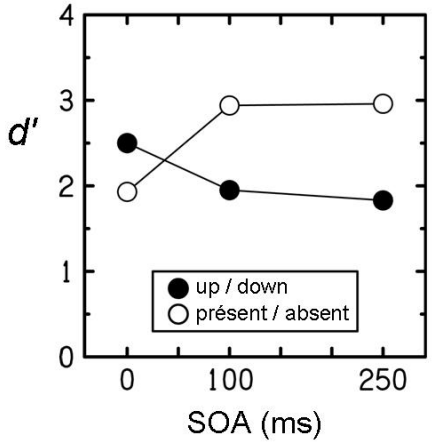

Figure 3. Performances moyennes des sujets dans une expérience de Demany, Semal, et Pressnitzer [8].

Mais le résultat le plus important de l'expérience est le renversement de tendance obtenu lorsque les composants de *E* sont devenus asynchrones : on voit dans la Figure 3 que cette asynchronie a rendu la tâche présent/absent plus facile que la tâche up/down. Cela invalide l'idée que la tâche présent/absent est intrinsèquement difficile du point de vue décisionnel. Subjectivement, une asynchronie des composants de *E* rendait ces composants plus faciles à percevoir individuellement. Telle est très probablement l'origine du renversement de tendance.

3 Les propriétés des FSD

3.1 Un modèle qualitatif

On peut rendre compte des résultats paradoxaux décrits ci-dessus à l'aide d'un modèle schématique de FSD qui tient dans les trois hypothèses suivantes.

1. Le système auditif contient deux sous-ensembles de FSD, qui ont des préférences directionnelles opposées : l'un répond préférentiellement à des augmentations de fréquence, l'autre à des baisses.

2. Chacun des deux sous-ensembles de FSD répond maximalement à des **petits** changements de fréquence.

3. Lorsqu'une séquence sonore active simultanément les deux sous-ensembles de FSD, la direction perçue du changement est celle que préfère le sous-ensemble dont l'activation est la plus forte.

Selon ce modèle, chaque fois que, dans nos expériences, le son pur *T* devait être mis en relation avec un accord, les deux sous-ensembles de FSD étaient activés simultanément. Dans la condition up/down, cependant, chaque séquence présentée au sujet devait (en vertu de l'hypothèse 2) activer l'un des deux sous-ensembles plus fortement que l'autre ; cette asymétrie indiquait au sujet la réponse correcte. Dans la condition présent/absent, par contre, chaque séquence devait selon le modèle activer à peu près au même degré les deux sous-ensembles, quelle que soit la réponse correcte ; les FSD fournissaient donc moins d'information utile, ce qui permet de comprendre la faiblesse des performances enregistrées dans cette condition. De meilleures performances étaient prédites dans la condition présent/proche puisque cette fois l'activation des FSD devait être symétrique quand la réponse correcte était "présent" mais asymétrique dans le cas contraire.

Nous supposons par ailleurs que les FSD sont activés beaucoup plus fortement par deux sons purs immédiatement consécutifs (ou séparés par un silence) que par deux sons purs entre lesquels vient s'interposer temporellement un troisième son pur. Cette supposition permet de rendre compte de l'effet d'asynchronie dépeint dans la Figure 3. Elle est également justifiée par d'autres données expérimentales [8]. Il apparaît cependant que si, dans la tâche up/down, c'est un bruit large-bande qui vient s'interposer entre un accord de sons purs et le son pur *T*, ce bruit dégrade à peine la performance. Nous l'avons constaté pour des bruits roses de même sonie que les accords [8]. Ainsi, les FSD semblent insensibles au bruit.

3.2 La taille optimale des changements de fréquence

Le modèle qualitatif exposé dans la section précédente veut que les FSD répondent maximalement à des "petits" changements de fréquence. Quelle est plus précisément la taille des changements évoquant une réponse maximale des FSD ? Nous l'avons estimée dans une étude [9] où, une fois de plus, des auditeurs avaient à effectuer la tâche up/down en réponse à des séquences sonores constituées d'un accord de sons purs suivi d'un seul son pur (*T*). Chaque accord était constitué de six sons purs, espacés en fréquence de 650 "cents" (i.e., 6.5 demi-tons) pour certaines séquences et 1000 cents pour d'autres. Outre cet intervalle de fréquence (*I*), nous avons manipulé la durée des stimuli, la durée du silence séparant l'accord de *T*, et surtout la taille de l'intervalle de fréquence (Δ, en cents) séparant *T* du composant de l'accord le plus proche en fréquence (celui-ci pouvait être n'importe lequel des six composants). Notons que Δ était toujours largement inférieur à *I*/2, de sorte que la tâche n'était jamais objectivement ambiguë.

Le panneau (a) de la Figure 4 montre les résultats obtenus chez sept sujets pour des séquences assez lentes (durée des stimuli : 300 ms; durée du silence inter-stimuli : 500 ms) avec *I* = 650 cents. On voit que chez chaque sujet, quand Δ a varié de 50 à 250 cents, *d′* a d'abord augmenté puis diminué. Nous avons ajusté aux données moyennes, représentées par des cercles dans le panneau (b) de la Figure, une fonction continue de la forme

$$d' = a \cdot \Delta^{b} \cdot \exp(-c\Delta), \qquad (1)$$

où a, b et c constituaient trois paramètres. La courbe tracée dans le panneau (b) est le meilleur ajustement trouvé. Cette courbe atteint son maximum pour Δ = 117 cents.

Le panneau (c) de la Figure 4 montre les résultats obtenus chez quatre sujets pour des séquences nettement plus rapides (durée des stimuli : 100 ms ; durée du silence inter-stimuli : 100 ms) avec *I* = 1000 cents. Une fonction du type défini par l'équation (1) a de nouveau été ajustée aux données moyennes, et le meilleur ajustement trouvé est tracé dans le panneau (d). La courbe atteint son maximum

pour Δ = 122 cents, ce qui ne diffère pas significativement du résultat précédent.

Dans d'autres conditions expérimentales, nous avons à nouveau trouvé que la valeur optimale de Δ pour la tâche up/down est environ 120 cents. Au total, l'ensemble des données permet donc de penser que les FSD répondent maximalement à des changements de fréquence d'environ 120 cents (soit 0.1 octave), au moins dans des séquences lentes ou modérément rapides. Ceci n'est vrai, cependant, que si chaque sous-ensemble de FSD ne répond pas (ou répond très peu) aux changements de fréquence évoquant une réponse maximale de l'autre sous-ensemble.

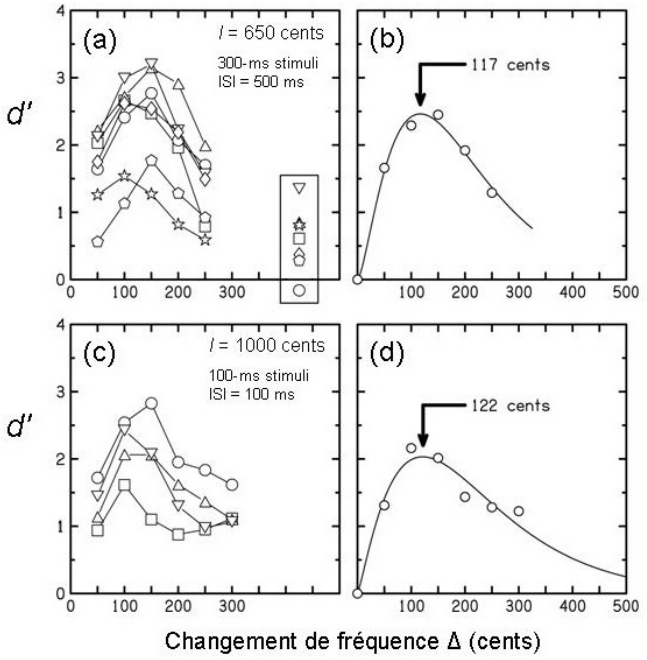

Figure 4. Résultats d'une recherche de Demany, Pressnitzer, et Semal [9]. Voir le texte pour des commentaires. Dans le panneau (a), les sept symboles non connectés et encadrés par un rectangle représentent les performances des sujets dans la tâche présent/absent.

3.3 La mémoire auditive exploitée par les FSD

Il parait probable que les FSD se situent dans le cortex auditif. En tout cas, leur emplacement n'est certainement pas très périphérique. En effet, lorsque la tâche up/down est effectuée avec un accord de sons purs présenté seulement à l'oreille gauche et un son pur *T* présenté seulement à l'oreille droite (ou vice versa), les performances sont à peine inférieures à celles trouvées quand l'accord et *T* sont présentés à la même oreille [10, 11]. Par ailleurs, si *T* n'est plus un son pur mais est remplacé par du bruit large-bande à chaque oreille, en manipulant les relations de phase inter-aurales de façon à ce que le stimulus évoque malgré tout, grâce au système binaural, une sensation de hauteur tonale similaire à celle évoquée par un son pur, les FSD semblent encore capables d'intervenir : la tâche up/down se révèle à nouveau plus facile que la tâche présent/absent [11].

Une autre raison de penser que les FSD sont corticaux est qu'ils apparaissent capables de relier deux sons purs séparés par plusieurs secondes, ce qui fait nécessairement intervenir une forme élaborée de mémoire. La Figure 5 montre les performances de quatre sujets dans une tâche up/down pour diverses durées du silence suivant l'accord et précédant *T* (0.5, 1, 2, 4, et 8 s). Sont également représentées ici, pour les mêmes sujets et les mêmes accords, les performances obtenues dans la tâche présent/absent quand le silence durait 4 s (la seule durée testée pour la tâche présent/absent, dans cette expérience). On voit que les performances dans la tâche up/down ont décliné quand la durée du silence augmentait, mais que cette tâche est demeurée plus facile que la tâche présent/absent pour un silence de 4 s.

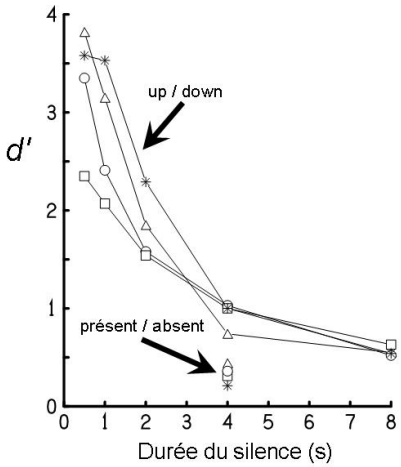

Figure 5. Effet de la durée du silence séparant l'accord du son *T* subséquent dans une tâche up/down [5] : données individuelles de quatre auditeurs. La Figure montre également la performance des mêmes auditeurs dans la tâche présent/absent lorsque le silence durait 4 s.

La mémoire sensorielle auditive est multiforme [12]. C'est une forme **implicite** de mémoire auditive que doivent exploiter les FSD puisque cette mémoire se manifeste par la rétention de stimuli pourtant non perçus consciemment. La mémoire en question doit cependant différer de la mémoire auditive implicite qui se manifeste dans des effets perceptifs dits "d'amorçage" ("priming") [13], et elle doit également différer de la mémoire implicite qui se traduit physiologiquement par ce qu'on appelle de "l'adaptation" neuronale [14]. En effet, les conséquences d'un stimulus-amorce ou d'un stimulus adaptateur sur l'encodage d'un stimulus subséquent sont maximales lorsque ce dernier est identique à l'amorce ou au stimulus adaptateur ; l'effet d'amorçage ou d'adaptation est d'autant plus fort que les deux stimuli se ressemblent. Or les FSD sont censés répondre optimalement à de légères différences entre stimuli. Un phénomène du type amorçage ou adaptation devrait logiquement favoriser la tâche présent/absent sans apporter de bénéfice pour la tâche up/down, alors que paradoxalement cette seconde tâche s'avère plus facile que la première.

La mémoire auditive exploitée par les FSD paraît puissante non seulement par sa durée de rétention, mais aussi par sa capacité − i.e., la quantité d'éléments qu'elle peut simultanément retenir. En fait, sa capacité semble essentiellement illimitée. Telle est la conclusion suggérée par une étude [15] dans laquelle la tâche up/down a été effectuée avec des accords composés d'un nombre variable (*N*) de sons purs. Outre *N* (varié de 1 à 12), nous avons manipulé dans cette étude la durée (*D*) du silence séparant l'accord du son pur *T* qui le suivait : *D* était varié de 0 à 2000 ms. Comme on pouvait s'y attendre, la performance des sujets a été d'autant meilleure que *N* était petit et que *D*, lui aussi, était petit. Mais le résultat le plus important est

que l'effet de *D* sur la performance n'a pas été plus grand quand *N* était grand que quand *N* était petit. Autrement dit, la rétention mnésique des composants de l'accord n'a pas été plus faible quand ces composants étaient nombreux que quand ils étaient peu nombreux.

Le résultat que je viens de mentionner est surprenant au regard de ce qu'on sait sur la mémoire visuelle. On s'accorde en effet à penser que celle-ci est incapable de retenir en détail une image complexe plus de 100 ms, et que seules des images très simples peuvent être fidèlement retenues plus longuement [16, 17]. En fait, de grandes différences semblent exister entre vision et audition pour ce qui concerne la détection de changement [18].

3.4 Changements spectraux versus changements de périodicité

L'existence de FSD n'est pas suggérée seulement par le phénomène auditif paradoxal que traduit l'avantage de la tâche up/down sur la tâche présent/absent. Elle est également suggérée par des résultats obtenus tout récemment dans des expériences très différentes [19].

Dans ces expériences, le sujet devait à chaque essai juger si deux séquences sonores présentées successivement étaient identiques ou non. Chaque séquence comprenait *N* = 1, 2, 4, ou 8 éléments successifs. Dans la première des deux séquences présentées, chacun des éléments était choisi au hasard parmi deux stimuli possibles (A et B), qui étaient deux sons complexes harmoniques. Lorsque la seconde séquence différait de la première, cette différence se limitait à un seul élément, choisi au hasard ; celui-ci devenait B plutôt que A, ou vice versa. La tâche n'était pas triviale car la différence entre A et B était faible. Cette différence portait soit sur leur intensité, soit sur leur fréquence fondamentale (et donc leur périodicité). En outre, les deux sons pouvaient être formés soit d'harmoniques résolus dans la cochlée, soit d'harmoniques non résolus. Lorsque leurs harmoniques n'étaient pas résolus, A et B étaient indiscriminables sur la base d'indices spectraux ; ils ne pouvaient être différenciés que du point de vue de la périodicité globale, sur la base d'indices purement temporels. Dans tous les cas, la taille de la différence physique entre A et B était initialement ajustée de façon à ce que les deux sons aient un degré fixe de discriminabilité (correspondant à un *d′* d'environ 2).

Le panneau gauche de la Figure 6 montre l'effet qu'a eu *N* sur la discrimination des séquences dans trois conditions expérimentales. Dans la condition "INT", A et B différaient par l'intensité. Dans la condition "F0/R", A et B différaient par la fréquence fondamentale et leurs composantes spectrales étaient résolues. Dans la condition "F0/NR", A et B différaient encore par la fréquence fondamentale mais leurs composantes spectrales n'étaient pas résolues. On voit que les résultats obtenus dans les conditions INT et F0/NR ont été similaires : dans ces deux conditions, la performance des sujets a décliné régulièrement quand *N* a varié de 1 à 8. Dans la condition F0/R, par contre, *d′* est resté quasi-constant tant que *N* n'excédait pas 4, et n'a diminué que quand *N* a atteint sa valeur maximale, 8.

Le panneau droit de la Figure 6 reproduit les résultats obtenus dans les conditions INT (aire gris clair) et F0/R (aire gris sombre), en y ajoutant (traits noirs et cercles) la prédiction d'un modèle simple concernant l'effet de *N* sur la performance. Ce modèle supposait que la discrimination des séquences était limitée seulement par la discrimination entre A et B, et que les éléments des séquences étaient traités indépendamment les uns des autres. On voit que la discrimination des séquences par les sujets a été moins bonne que prévu par le modèle dans la condition INT (c'était aussi le cas dans la condition F0/NR), mais que dans la condition F0/R les performances des sujets ont dépassé celles prévues par le modèle.

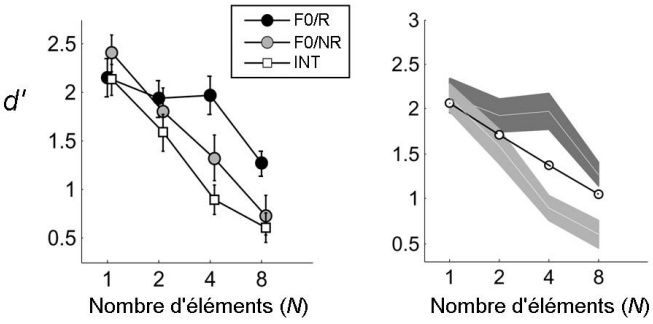

Figure 6. Résultats d'une expérience de Cousineau, Demany, et Pressnitzer [19]. Voir le texte pour des commentaires.

Les FSD étant conçus comme des détecteurs de changements **spectraux**, une seule des trois conditions de l'expérience leur permettait de jouer un rôle − et d'avoir un effet bénéfique − dans la discrimination des séquences ; c'était la condition F0/R. Le fait que les performances des sujets aient été meilleures dans cette condition que dans les deux autres donne donc un argument en faveur de l'existence de FSD, conçus spécifiquement comme des détecteurs de changements spectraux et non pas, plus généralement, des détecteurs de changements de périodicité. Les FSD traitant par définition des **relations** entre composantes sonores successives, on peut comprendre que les performances obtenues dans la condition F0/R aient dépassé les prédictions d'un modèle supposant que les éléments des séquences étaient traités indépendamment les uns des autres.

4 Remarques finales

L'ensemble des résultats présentés ici donne de bonnes raisons de penser que le système auditif dispose d'outils permettant de lier automatiquement des sons successifs différant par le spectre. Cependant, le substrat physiologique de ces outils − les FSD − reste à découvrir. Par ailleurs, des questions fondamentales subsistent quant à leur rôle exact dans l'analyse des scènes auditives. Il reste à déterminer, en particulier, si les FSD constituent des outils de ségrégation entre flux sonores concomitants ou bien s'ils n'interviennent qu'une fois cette ségrégation réalisée. Un autre problème est le suivant. On a coutume de penser qu'une séquence sonore est perçue comme d'autant plus cohérente que ses éléments sont similaires, et donc que la cohérence perceptive maximale est obtenue lorsque les éléments sont identiques. Si cela est exact, et si la fonction essentielle des FSD est de créer de la cohérence perceptive, comment comprendre que les FSD paraissent répondre de façon maximale à des **différences** de fréquence ? Il faut hélas admettre que la psychophysique est plus à même de révéler des phénomènes intriguants que d'en fournir l'explication.

Remerciements

Christophe Ramos, Catherine Semal, Wiebke Trost, Maja Serman, Jean-René Cazalets et Samuele Carcagno ont activement contribué aux recherches décrites ici ; je leur en suis très reconnaissant. Ma gratitude va également à Daniel Pressnitzer et Marion Cousineau (UMR CNRS 8158, Ecole Normale Supérieure et Université Paris-Descartes), avec qui j'entretiens une précieuse collaboration.

Références

[1] Fraisse P. *Psychologie du temps*, PUF, Paris (1967).

[2] Bregman A.S. *Auditory scene analysis*, MIT Press, Cambridge, MA (1990).

[3] van Noorden L.P.A.S. *Temporal coherence in the perception of tone sequences*, Doctoral thesis, University of Eindhoven (1975).

[4] Ullman S. "Two-dimensionality of the correspondence process in apparent motion", *Perception* 7, 683-693 (1978).

[5] Demany L., Ramos C. "On the binding of successive sounds: Perceiving shifts in nonperceived pitches", *J. Acoust. Soc. Am*. 117, 833-841 (2005).

[6] Green D.M., Swets J.A. *Signal detection theory and psychophysics,* Krieger, New York (1974).

[7] Kidd G.D., Mason C.R., Richards V.M., Gallun F.J., Durlach N.I. "Informational masking", in *Auditory perception of sound sources* (editors: Yost W.A., Popper A.N., Fay R.R.), Springer, New York (2008).

[8] Demany L., Semal C., Pressnitzer D. "Implicit versus explicit frequency comparisons: Two mechanisms of auditory change detection", *J. Exp. Psychol. Human Percept. Perform.* (in revision).

[9] Demany L., Pressnitzer D., Semal C. "Tuning properties of the auditory frequency-shift detectors", *J. Acoust. Soc. Am*. 126, 1342-1348 (2009).

[10] Demany L., Ramos C. "A paradoxical aspect of auditory change detection", in *Hearing − From sensory processing to perception* (editors: Kollmeier B. et al.), Springer, Heidelberg (2007).

[11] Carcagno S., Demany L. Manuscrit en préparation.

[12] Demany L., Semal C. "The role of memory in auditory perception", in *Auditory perception of sound sources* (editors: Yost W.A., Popper A.N., Fay R.R.), Springer, New York (2008).

[13] Schacter D.L. "Priming and multiple memory systems: perceptual mechanisms of implicit memory", in *Memory systems, 1994* (editors: Schacter D.L., Tulving E.), MIT Press, Cambridge, MA (1994).

[14] Ulanovsky N., Las L., Nelken I. "Processing of low-probability sounds by cortical neurons", *Nat. Neurosci*. 6, 391-398 (2003).

[15] Demany L., Trost W., Serman M., Semal C. "Auditory change detection: Simple sounds are not memorized better than complex sounds", *Psych. Sci.* 19, 85-91 (2008).

[16] Phillips W.A. "On the distinction between sensory storage and short-term visual memory", *Percept. Psychophys*. 16, 283-290 (1974).

[17] Luck S.J., Vogel E.K. "The capacity of visual working memory for features and conjunctions", *Nature* 390, 279-281 (1997).

[18] Demany L., Semal C., Cazalets J.R., Pressnitzer D. "Fundamental differences between change detection in vision and audition", *Exp. Brain Res.* (in revision).

[19] Cousineau M., Demany L., Pressnitzer D. "What makes a melody: The perceptual singularity of pitch sequences", *J. Acoust. Soc. Am.* 126, 3179-3187 (2009).