



HAL
open science

Apprentissage perceptif de la hauteur musicale et de l'attention sélective auditive

Hadrien Jean

► **To cite this version:**

Hadrien Jean. Apprentissage perceptif de la hauteur musicale et de l'attention sélective auditive. Sciences cognitives. Université Paris sciences et lettres, 2018. Français. NNT : 2018PSLEE090 . tel-03407789

HAL Id: tel-03407789

<https://theses.hal.science/tel-03407789>

Submitted on 28 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée à l'École Normale Supérieure

**Apprentissage perceptif dans le domaine de la hauteur
et de l'attention sélective auditive**

Soutenue par

Hadrien JEAN

Le 20 décembre 2018

Ecole doctorale n° 540

**Ecole doctorale
transdisciplinaire
Lettres/Sciences**

Spécialité

Sciences cognitives



Composition du jury :

Christophe MICHEYL *Rapporteur*
Chercheur Principal, Starkey Hearing
Technologies

Bénédicte POULIN-CHARRONNAT *Rapporteuse*
Chargée de Recherche, Université de
Bourgogne Franche-Comté

Frédéric BEVILACQUA *Examineur*
Directeur de Recherche, IRCAM

Anne CACLIN *Examinatrice*
Chargée de Recherche, Centre de
Recherche en Neurosciences de Lyon

Laurent DEMANY *Examineur*
Directeur de Recherche émérite, Institut de
Neuroscience Cognitives et Intégratives
d'Aquitaine

Benoît MONTIGNÉ *Examineur*
Professeur, Ecole Nationale Supérieure des
Arts Décoratifs

Daniel PRESSNITZER *Directeur de
thèse*
Directeur de Recherche, Ecole Normale
Supérieure



Cognitive
Control of a
Hearing Aid



Thèse de doctorat de l'École Normale Supérieure
École Doctorale Transdisciplinaire lettres/sciences - ED540

Présentée par Hadrien Jean

Pour obtenir le grade de
Docteur de l'École Normale Supérieure

Apprentissage perceptif dans le domaine de la hauteur et de l'attention sélective auditive

Préparée au
Laboratoire des Systèmes Perceptifs, UMR 8248,
Département d'études cognitives, École Normale Supérieure

soutenue à Paris le 20 Décembre 2018 devant un jury composé de

Christophe Micheyl	Chercheur Principal, Starkey Hearing Technologies	Rapporteur
Bénédicte Poulin- Charronnat	Chargée de Recherche, Université de Bourgogne Franche-Comté	Rapporteuse
Frédéric Bevilacqua	Directeur de Recherche, IRCAM	Examinateur
Anne Caclin	Chargée de Recherche, Centre de Recherche en Neurosciences de Lyon	Examinatrice
Laurent Demany	Directeur de Recherche Émérite, Institut de Neuroscience Cognitives et Intégratives d'Aquitaine	Examinateur
Benoît Montigné	Professeur, Ecole Nationale Supérieure des Arts Décoratifs	Examinateur
Daniel Pressnitzer	Directeur de recherche, École Normale Supérieure	Directeur de thèse

Résumé

Pour s'adapter à un environnement changeant et parfois imprévisible, il est important que les systèmes perceptifs gardent la possibilité de se transformer et de faire l'objet d'apprentissage tout au long de la vie. Le but de cette thèse est l'étude de mécanismes d'apprentissage perceptif en audition, plus spécifiquement pour le cas de la perception de la hauteur et de l'attention sélective. Il a été suggéré que des tâches écologiques telles que la pratique musicale ou les jeux vidéo sont associées à des apprentissages perceptifs. Or, ces tâches sont caractérisées par la mise en jeu d'une boucle sensori-motrice. Dans une première série d'expériences, nous avons créé un paradigme sur tablette visant à utiliser une boucle audio-motrice en vue d'un apprentissage de la perception de la hauteur. Les résultats n'ont pas montré d'amélioration significative suite à l'entraînement. Dans une seconde série d'expériences, nous avons développé une seconde version de la tâche d'entraînement audio-moteur, et examiné son effet sur les performances à la fois pour des tâches de perception de la hauteur et des tâches de masquage informationnel, en faisant l'hypothèse que ces dernières testaient l'attention sélective basée sur la fréquence. Les résultats ont montré en tendance un effet spécifique de l'entraînement audio-moteur uniquement pour les participants avec les plus faibles performances initiales. Une amélioration rapide à la tâche de masquage informationnel non spécifique à l'entraînement audio-moteur a en outre été observée. Dans une troisième série d'expériences, nous avons pu confirmer cet effet d'apprentissage rapide dans les tâches de masquage informationnel, probablement lié à l'attention sélective. De plus, un fort effet de la fréquence de la cible sur l'attention sélective a été observé de manière inattendue. Pour cette raison, nous avons examiné cet effet de fréquence dans une quatrième série d'expériences, en contrôlant différents paramètres confondants comme les différences d'audibilité ou la distribution des nuages de tons. Les résultats montrent un effet robuste de la fréquence absolue sur les performances de masquage informationnel. Enfin, en vue

d'explorer cet effet et l'hypothèse basée sur l'attention sélective à l'aide d'indices électrophysiologiques, nous présentons des données préliminaires concernant les corrélats EEG de l'attention sélective dans des tâches de masquage informationnel. Les résultats suggèrent qu'il est possible d'utiliser une technique basée sur la reconstruction d'une enveloppe aléatoire imposée à la cible pour obtenir un corrélat EEG de l'attention. Pour conclure, cette thèse a démontré des effets d'apprentissage de l'attention sélective auditive observée dans des tâches de masquage informationnel. L'intérêt de l'utilisation de la boucle audio-motrice pour l'entraînement n'a pas pu être mise en évidence de façon claire, même si des tendances ont été observées. En outre, nous avons découvert un effet de la fréquence absolue sur la perception non encore documenté, un "audiogramme informationnel", qui ouvre de nouvelles questions sur ses mécanismes sous-jacents et ses possibilités d'entraînement.

Table des matières

1	CHAPITRE 1 : Partie théorique	5
1.1	Introduction	5
1.2	Apprentissage et perception de la fréquence	7
1.2.1	Protocoles de discrimination et d'identification de fréquence	7
1.2.2	Apprentissage perceptif : transfert et spécificité	11
1.2.3	Incertitude du stimulus	24
1.3	Apprentissage et masquage informationnel	31
1.3.1	Protocoles de masquage informationnel	32
1.3.2	Incertitude du stimulus	39
1.3.3	Similarité entre cible et masqueur	43
1.3.4	Apprentissage perceptif et masquage informationnel	45
1.4	Boucle audio-motrice	48
1.4.1	Pratique musicale	49
1.4.2	Jeux vidéo	52
1.4.3	Protocoles immersifs et jeux audio	56
2	CHAPITRE 2. Perception de la hauteur et boucle audio-motrice	63
2.1	Aperçu	63
2.2	Introduction	64
2.3	Expérience 1	66
2.3.1	Méthodes	66
2.3.2	Résultats	72
2.3.3	Discussion	78
2.4	Expérience 2	79
2.4.1	Méthodes	79
2.4.2	Résultats	80

2.4.3	Discussion	86
2.5	Expérience 3	88
2.5.1	Méthodes	88
2.5.2	Résultats	89
2.5.3	Discussion	92
2.6	Discussion générale	94
2.7	Conclusions	97
3	CHAPITRE 3. Effet d'un entraînement actif sur la perception de la fréquence et le masquage informationnel	99
3.1	Aperçu	99
3.2	Introduction	100
3.3	Méthodes	102
3.3.1	Participants	102
3.3.2	Stimuli	103
3.3.3	Matériel	106
3.3.4	Procédure	106
3.3.5	Analyses des données	113
3.4	Résultats	114
3.4.1	Pré-/post-test	114
3.4.2	Tâche audio-motrice	120
3.4.3	Tâche d'exposition	123
3.5	Discussion	123
3.6	Conclusions	126
4	CHAPITRE 4. Entraînement perceptif et masquage informationnel	127
4.1	Aperçu	127
4.2	Introduction	128

4.3	Méthodes	129
4.3.1	Participants	129
4.3.2	Stimuli	129
4.3.3	Matériel	132
4.3.4	Procédure	132
4.3.5	Analyse de données	134
4.4	Résultats	134
4.5	Discussion	136
4.6	Conclusions	138
5	CHAPITRE 5. Audiogramme informationnel : un effet de mas-	
	quage informationnel dépendant de la fréquence et sans lien avec	
	l'audiogramme de détection	141
5.1	Aperçu	141
5.2	Introduction	142
5.3	Expérience 1	143
5.3.1	Méthodes	143
5.3.2	Résultats	149
5.3.3	Discussion	151
5.4	Expérience 2	152
5.4.1	Méthodes	153
5.4.2	Résultats	156
5.4.3	Discussion	158
5.5	Discussion générale	159
5.6	Conclusions	163
6	CHAPITRE 6. Indices électrophysiologiques de l'attention sélec-	
	tive dans un paradigme de masquage informationnel	165
6.1	Aperçu	165
6.2	Introduction	166

6.3	Méthodes	168
6.3.1	Participants	168
6.3.2	Stimuli	169
6.3.3	Matériel	171
6.3.4	Procédure	171
6.3.5	Analyses de données	173
6.4	Résultats	175
6.4.1	Partie 1	175
6.4.2	Partie 2	180
6.5	Discussion	185
6.6	Conclusions	187
7	CHAPITRE 7 : Conclusions	189
7.1	Principaux résultats	189
7.1.1	Étude 1 : Perception de la hauteur et boucle audio-motrice . .	189
7.1.2	Étude 2 : Effet d'un entraînement actif sur la perception de la fréquence et le masquage informationnel	190
7.1.3	Étude 3 : Entraînement perceptif et masquage informationnel	190
7.1.4	Étude 4 : Audiogramme informationnel : un effet de masquage informationnel dépendant de la fréquence et sans lien avec l'audiogramme de détection	191
7.1.5	Étude 5 : Indices électrophysiologiques de l'attention sélective.	191
7.2	Perspectives	192
8	ANNEXES	195
8.1	Annexe 1. Notes sur l'implémentation technique des expériences . . .	195
8.2	Annexe 2. Collaboration avec le compositeur Fabien Touchard pour la mise en place de stimuli musicaux	203
	RÉFÉRENCES	209

1 CHAPITRE 1 : Partie théorique

1.1 Introduction

Il semble naturel que notre aptitude à réaliser une tâche s'améliore au fur et à mesure de la pratique. On peut penser au radiologue qui peut détecter un risque de tumeur avec une haute précision ou encore l'amateur de vin capable de reconnaître l'assemblage des cépages utilisé dans un cru. Des actions les plus fondamentales acquises précocement après la naissance, comme la capacité à saisir un objet, aux aptitudes les plus spécialisées, nous sommes en phase continue d'apprentissage. Beaucoup de ces tâches nécessitent, pour être réalisées, un concours des aptitudes perceptives et motrices. La saisie d'un objet implique sa localisation préalable. L'action motrice sollicite alors les informations récoltées par le biais de la perception visuelle pour être enclenchée et réalisée correctement. Cet aller-retour entre perception et action est perpétuel. Le sculpteur regarde le monde avec un œil altéré par sa pratique et son geste est la synthèse de ses perceptions. Cette altération des sens est nécessaire à l'adaptabilité des systèmes perceptifs à l'environnement. Cette adaptation naturelle du système sensoriel peut intervenir de manière implicite ou être volontairement favorisée par la pratique.

On parle d'apprentissage perceptif lorsque cette adaptation améliore une certaine capacité perceptive. De manière opérationnelle, l'apprentissage perceptif désigne l'amélioration par la pratique de la capacité à extraire et/ou utiliser des informations sensorielles de l'environnement. Ce phénomène s'observe pour les différentes modalités d'acquisition des informations sensorielles et peut concerner le toucher (Harrar et al., 2014; Sathian & Zangaladze, 1997), le goût (discrimination de bières : Peron & Allen, 1988), ou l'odorat (Rabin, 1988). La perception visuelle et auditive sont toutefois les domaines où la grande majorité des études a été menée, comme nous le verrons lors de la revue bibliographique.

Le but de cette thèse est d'étudier ces mécanismes d'apprentissage dans le cadre de l'audition. Deux capacités auditives principales ont été étudiées : la perception de la hauteur et l'attention sélective. La perception de la hauteur est un attribut majeur du langage ou de la musique. De faibles différences de fréquences sont importantes à détecter de manière à identifier les émotions véhiculées par la voix, ou les dessins mélodiques d'une pièce musicale. Cette capacité perceptive peut être particulièrement développée, comme dans le cas de participants musiciens, ou déficiente, comme dans le cas de patients malentendants, dont les aides auditives ne permettent pas de

récupérer cette aptitude de manière satisfaisante (McDermott, 2004). L'enjeu de mieux comprendre les phénomènes d'apprentissage associés est donc important.

La perception de la fréquence est également impliquée dans l'attention sélective auditive dans le sens où cet attribut constitue un indice permettant de séparer différentes sources sonores. L'attention sélective désigne la capacité à isoler une partie de l'environnement sonore (Bregman, 1994). Par exemple, suivre une conversation dans un café implique une capacité à filtrer les informations auditives d'intérêt. Mais est-il possible d'améliorer cette capacité sélective ? Le second but de ce travail est d'étudier les effets potentiels d'entraînement de cette aptitude.

Enfin, la question de la relation entre motricité et perception est au centre des premières études expérimentales de ce travail. De nombreux éléments de la littérature suggèrent que des activités écologiques telles que la pratique musicale ou la pratique des jeux vidéo sont associées à des capacités perceptives et motrices particulièrement développées (par exemple, Green & Bavelier, 2003; Kishon-Rabin et al., 2001; Micheyl et al., 2006). Une hypothèse est qu'un double effet de la boucle sensori-motrice permet de renforcer l'apprentissage : d'une part les informations perceptives servent à donner un retour sur la qualité de l'action qui a été réalisée et d'autre part, la stimulation sensorielle devient particulièrement saillante grâce à son utilité pour la tâche. Pour ces raisons, nous avons mis en place des protocoles d'apprentissage perceptif sur tablette mettant en jeu une boucle audio-motrice et visant l'amélioration des performances de perception de la fréquence et d'attention sélective.

Une revue sélective de la littérature sera tout d'abord proposée dans ce chapitre théorique. Cette revue concernera les effets d'apprentissage de la perception de la fréquence (1.2) et de l'attention sélective auditive, notamment dans les tâches de masquage informationnel (1.3), ainsi que l'influence de la boucle audio-motrice sur les mécanismes d'apprentissage perceptif (1.4). Suite à cette introduction théorique, les chapitres expérimentaux (du chapitre 2 au chapitre 6) exposeront les différents travaux expérimentaux réalisés pendant cette thèse. Dans le chapitre 2, un protocole d'apprentissage perceptif fondé sur la boucle audio-motrice et évalué par rapport à la perception de la fréquence sera présenté. Le chapitre 3 décrira une évolution de ce protocole dont l'efficacité a été évaluée sur la perception de la fréquence et sur la résistance au masquage informationnel. Dans le chapitre 4, la question spécifique de l'amélioration de l'attention sélective dans un protocole de masquage informationnel sera examinée. Dans le chapitre 5, nous examinerons l'effet de la fréquence de la cible dans un protocole de masquage informationnel, observé de manière inattendue

dans le chapitre 4. Enfin, le chapitre 6 présentera un travail préliminaire concernant les indices électrophysiologiques (EEG) de l'attention sélective auditive.

1.2 Apprentissage et perception de la fréquence

Une des questions examinées dans cette thèse concerne l'amélioration des capacités de perception de la fréquence. Dans cette partie, nous allons étudier le contexte théorique de cette problématique. Tout d'abord, nous allons présenter une synthèse illustrée des différentes tâches qui ont été utilisées pour évaluer la perception de la fréquence (1.2.1). Nous allons également voir qu'il est crucial de distinguer les protocoles évaluant la capacité à discriminer deux fréquences proches (tâches de *discrimination de fréquence*) et la capacité à identifier la direction d'un changement de fréquence (tâche d'*identification de fréquence*). Dans une seconde partie (1.2.2), la question de la spécificité de l'apprentissage perceptif sera mise en avant. Enfin, l'influence de l'incertitude du stimulus sur l'apprentissage perceptif sera questionnée (1.2.3).

1.2.1 Protocoles de discrimination et d'identification de fréquence

Une manière d'évaluer la capacité de perception de la hauteur associée à la fréquence de sons purs est de demander au participant de détecter si deux sons purs sont différents. Dans ces tâches, tout indice perceptif peut en théorie être utilisé par l'auditeur (par exemple, des différences de sonies associées au changement de fréquence), mais l'hypothèse implicite faite par la plupart des études est que l'indice principal est un indice de hauteur associé à la fréquence. Ces tâches seront désignées *tâches de discrimination de fréquence*. Il est également possible de présenter deux sons purs de fréquence légèrement différente et de demander au participant quel est celui qui est le plus aigu. Pour réaliser cette tâche, il est nécessaire de détecter la direction de la différence. Ces tâches seront par la suite de ce travail désignées *tâches d'identification de fréquence*¹. Il est important de réaliser cette distinction car les processus et l'apprentissage perceptif évalués par ces tâches pourraient être différents. Pour cette raison, il est intéressant de comparer les principaux protocoles utilisés dans la littérature sur l'apprentissage perceptif de la hauteur, et de déterminer s'il s'agit de tâches de discrimination ou d'identification de fréquence. Cette partie a

1. Nous désignerons ces tâches ainsi par simplicité, bien que l'identification concerne la direction d'un changement de fréquence et non la fréquence elle-même.

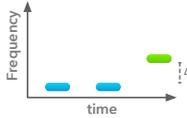
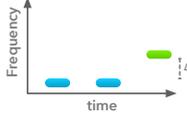
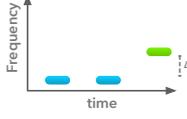
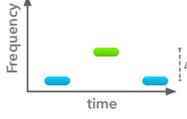
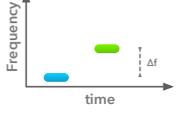
References	Stimulus	Procedure	Task	Discrimination or identification	Key in the text
Kishon-Rabin et al. (2001) Amitay et al. (2006) Amitay et al. (2012) Zhang et al. (2016)		3I-3AFC	Which tone is different?	Discr.	a
Moore & Peters (1992) Irvine et al. (2000) Amitay et al. (2006) Moore et al. (2008)		3I-3AFC	Which tone is higher?	Discr.	b
Hawkey et al. (2004) Amitay et al. (2006)		3I-2AFC: AXB	Was the middle tone the same pitch as the first or third tone?	Discr.	c
Demany & Semal (2002) Delhommeau et al. (2002) Delhommeau et al. (2005)		3I-2AFC: XAB	Which tone is different from the first?	Discr.	d
Jesteadt & Bilger (1974)		2I-2AFC	Same or different?	Discr.	e

Figure 1.1: Protocoles de discrimination et d'identification de fréquence. Généralement, la valeur Δf est variée de manière adaptative jusqu'à obtention du seuil perceptif. Les sons purs représentés en vert correspondent à l'intervalle qu'il faut sélectionner pour obtenir une réponse correcte. Le type de procédure ainsi que la tâche associées au stimuli sont précisés.

pour but de synthétiser les protocoles utilisés dans les études dont certaines seront décrites par la suite de cette introduction.

1.2.1.1 Discrimination de fréquence

Plusieurs types de protocoles peuvent être utilisés pour évaluer les capacités de discrimination de fréquence. Une possibilité est de présenter trois sons purs successifs, dont deux de fréquence identique. La consigne est alors d'identifier l'intervalle correspondant au son pur qui est différent (Figure 1.1 a; Kishon-Rabin et al., 2001; Amitay et al., 2006b, 2012; Zhang et al., 2016) ou bien au son pur le plus aigu (Figure 1.1 b; Irvine et al., 2000; Amitay et al., 2006a; Moore & Peters, 1992; Moore et al.,

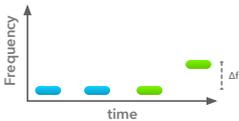
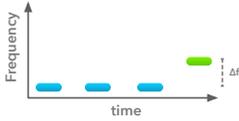
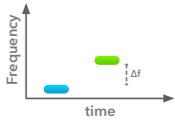
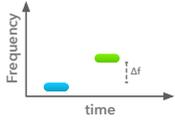
Reference	Stimulus	Procedure	Task	Discrimination or identification	Key in the text
Demany (1985) Semal & Demany (2006) Mathias et al. (2010)		4I-2AFC	In which pair the second tone is higher?	Discr.	f
Demany (1985) Semal & Demany (2006) Mathias et al. (2010)		4I-2AFC	In the 'different' pair, which tone is higher?	Id.	g
Kishon-Rabin et al. (2001) Hawkey et al. (2004) Amitay et al. (2005) Micheyl et al. (2006) Roth et al. (2006) Jones et al. (2013)		2I-2AFC	Which tone is higher?	Id.	h
Harris (1952) Jesteadt & Bilger (1974)		2I-2AFC	Is the second tone higher or lower?	Id.	i
Amitay et al. (2006) Roth et al. (2006) Roth et al. (2008) Jones et al. (2015)		2I-2AFC	Which tone is higher? (Impossible task)	-	j

Figure 1.2: Protocoles de discrimination et d'identification de fréquence (suite).

2008). Dans le cas des ces deux tâches de type 3I-3AFC (*3-Interval - 3-Alternative Forced-Choice*), il n'est pas nécessaire de percevoir la direction du changement car le son pur différent est toujours plus aigu. Dans un autre type de tâche, 3 sons purs sont également présentés, mais seulement deux choix de réponses sont proposés (3I-2AFC). Dans ce cas, on peut distinguer les tâches de type AXB dans lesquelles la consigne est de dire si le son pur central (X) est de fréquence identique au premier (A) ou au dernier son pur du triplet (B) (Figure 1.1 c; Hawkey et al., 2004; Amitay et al., 2006a), et les tâches de type XAB dans lesquelles le participant doit identifier le son pur qui est différent du premier (Figure 1.1 d; Demany & Semal, 2002; Delhommeau et al., 2005, 2002). On peut également noter l'utilisation de procédures *identique/différent* permettant d'évaluer les capacités de discrimination (Figure 1.1 e; Jesteadt & Bilger, 1974), où deux intervalles doivent être comparés et identifiés comme identiques ou différents. Enfin, la tâche à double paires, que nous utiliserons dans la première étude expérimentale (chapitre 2), consiste à présenter quatre sons purs successifs organisés en deux paires, dont l'une possède des sons purs différents et l'autre des sons purs identiques. La tâche est d'identifier la paire qui contient des sons purs différents (Figure 1.2 f; Demany, 1985; Mathias et al., 2010; Semal & Demany, 2006). Cette tâche a l'avantage de pouvoir évaluer également les seuils d'identification de fréquence avec les mêmes stimuli. La consigne est alors de choisir le son pur le plus aigu, dans la paire qui contient des sons purs de fréquences différentes (voir prochaine section).

1.2.1.2 Identification de fréquence

Une tâche classique d'identification de fréquence est de présenter deux sons purs au participant et de lui demander quel est celui qui est le plus aigu (Figure 1.2 h; Kishon-Rabin et al., 2001; Amitay et al., 2005; Hawkey et al., 2004; Jones et al., 2013; Micheyl et al., 2006; Roth et al., 2003, 2006). Contrairement aux tâches de discrimination, il faut identifier ici la direction du changement de fréquence pour répondre correctement. Une variante consiste à demander si le second son pur est plus grave ou plus aigu que le premier (Figure 1.2 i; Harris, 1952; Jesteadt & Bilger, 1974). Une troisième variante de cette tâche consiste à demander quel est le son pur le plus aigu tout en présentant des sons purs de fréquences identiques (Figure 1.2 j; Amitay et al., 2006b; Jones et al., 2015; Roth et al., 2008, 2006). Dans ce cas particulier, il ne s'agit donc pas d'une tâche permettant de mesurer les capacités du participants mais qui a été utilisée comme tâche d'entraînement perceptif. Enfin, il est possible d'utiliser la tâche à double paires (comme dans le cas de la discrimination de fréquence) mais en demandant quel est le son pur le plus aigu dans la paire qui

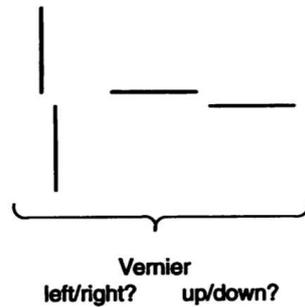


Figure 1.3: Issu de Poggio et al. (1992). Tâche de Vernier. Le but de la tâche est d’indiquer si la seconde barre est décalée vers la gauche ou vers la droite dans la condition verticale (gauche) ou vers le haut ou vers le bas dans la condition horizontale (droite).

contient les sons purs différents (Figure 1.2 g; Demany, 1985; Mathias et al., 2010; Semal & Demany, 2006).

En résumé, il existe de nombreux protocoles expérimentaux visant à évaluer la perception de la fréquence. Bien que tous ces protocoles aient superficiellement le même objectif, il est tout à fait possible que différents paramètres, comme la mémoire de travail, puissent influencer la difficulté de la tâche. On peut imaginer que les ressources cognitives nécessaires pour traiter 4 intervalles successifs sont supérieures que lorsqu’il n’y a que deux intervalles. Kishon-Rabin et al. (2001) ont montré que les seuils sont plus bas pour la tâche 3I-3AFC où les participants devaient choisir le son pur de fréquence différente, que pour la tâche 2I-2AFC où les participants devaient trouver le son pur le plus aigu. Les auteurs mettent en avant le rôle de la mémoire à court terme dans ces tâches, mais il est difficile de savoir si le gain de performances est lié au nombre de sons purs à garder en mémoire ou à la différence discrimination/identification. Ces considérations peuvent aussi être importantes pour les études visant à caractériser l’apprentissage de la perception de hauteur.

1.2.2 Apprentissage perceptif : transfert et spécificité

Une des découvertes capitales dans le domaine de l’apprentissage perceptif concerne la spécificité possible de l’apprentissage. La spécificité désigne le fait qu’une amélioration observée sur une tâche perceptive peut être spécifique aux stimuli ou à la tâche utilisée et ne se généralise pas à des tâches ou stimuli qui n’ont pas été rencontrés pendant l’apprentissage. Par exemple, dans un exemple classique du

domaine visuel, lorsqu'un participant est entraîné à discriminer un décalage entre deux barres verticales (tâche de Vernier ; Figure 1.3), ses performances augmentent avec la pratique, mais lorsqu'il doit ensuite discriminer un décalage entre deux barres horizontales, ses performances sont réduites, ce qui montre que l'apprentissage est spécifique et ne s'est pas généralisé d'une orientation à l'autre.

La question du transfert de l'apprentissage à d'autres tâches et stimuli est cruciale dans la littérature autant pour ses implications pratiques (rééducation perceptive, pédagogie) que pour le cadre théorique qu'elle suggère (les niveaux de traitement neuronal auxquels l'apprentissage se produit). La spécificité de l'apprentissage a en effet été utilisée comme argument d'un apprentissage perceptif impliquant les zones de bas niveau des systèmes sensoriels. Une amélioration des performances spécifique à une orientation par exemple, signifierait que l'apprentissage concerne une population de neurones spécifiquement associée au traitement de ces stimuli, probablement dans les aires corticales primaires.

Nous allons tout d'abord réaliser une revue des méthodologies utilisées pour évaluer la spécificité de l'apprentissage perceptif (1.2.2.1). Ensuite, l'implication de la spécificité de l'apprentissage en terme de processus neuronaux sera considérée (1.2.2.2). Le cas de la spécificité de la fréquence d'entraînement dans le cas de l'apprentissage perceptif sera ensuite examinée (1.2.2.3), étant au centre de cette revue. Les cas particuliers de l'apprentissage des caractéristiques non-pertinentes pour la tâche (*task-irrelevant learning*) et de conditions favorisant un apprentissage généralisable seront enfin considérés (1.2.2.4).

1.2.2.1 Méthodologies d'évaluation du transfert de l'apprentissage perceptif

D'un point de vue méthodologique, il est possible d'évaluer le transfert de l'apprentissage perceptif de différentes façons.

Une manière est de réaliser des pré-/post-tests sur une tâche/stimuli et de créer plusieurs groupes de participants entraînés sur des tâches/stimuli différents (par exemple, Demany, 1985). Un groupe contrôle qui ne réalise que les pré-/post-test peut également être ajouté (par exemple, Roth et al., 2003).

Il est aussi possible de mettre en place un apprentissage sur un type de tâche/stimuli et un pré-/post-test sur une variété de tâches/stimuli (par exemple, Irvine et al., 2000; Delhommeau et al., 2005). Par exemple, dans l'étude de Irvine et al. (2000), les participants réalisaient une tâche de discrimination de fréquence (Figure 1.1 b).

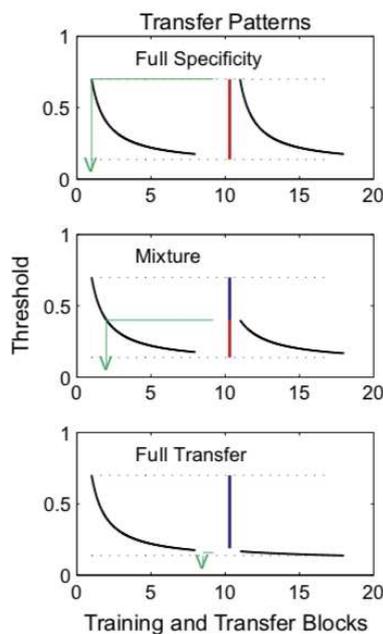


Figure 1.4: Issu de Jeter et al. (2010). Illustration de la spécificité de l'apprentissage perceptif.

Ils ont entendu les deux fréquences, 5000 et 8000 Hz, dans les pré-/post-tests, alors que la moitié des participants a été entraînée uniquement avec la fréquence 5000 Hz et l'autre moitié uniquement avec la fréquence 8000 Hz.

Une autre manière est de supprimer la phase de pré-test et d'évaluer les performances uniquement au post-test. Par exemple, Amitay et al. (2005) ont entraîné des participants sur une tâche d'identification de fréquence (Figure 1.2 h). L'entraînement utilisait des sons purs standard à 1000 Hz et les post-tests des sons purs standards à 500, 1000, 2000, 4000 Hz. Les différences entre les fréquences entraînées et non-entraînées ont été considérées. Cependant, sans le pré-test, il est difficile d'exclure la possibilité de différences de performances liées à la fréquence.

Une autre possibilité (par exemple, Demany & Semal, 2002; Jeter et al., 2010) est de comparer les courbes d'apprentissage dans deux conditions successives. Un apprentissage plus faible dans la seconde condition suggère un transfert, car une partie de l'apprentissage est alors attribué à la première condition. Cela est illustré dans la Figure 1.4. Dans cette étude, les 10 premiers blocs sont réalisés dans une condition et les 10 suivants dans une autre condition. Si les performances au début de la deuxième condition sont similaires à celles au début de la première condition

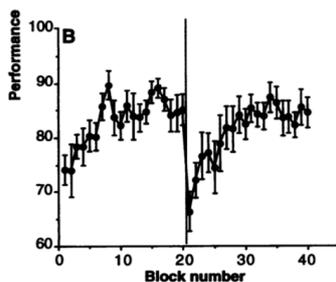


Figure 1.5: Issu de Poggio et al. (1992). Spécificité de l'orientation à une tâche de Vernier. Un changement d'orientation (de vertical à horizontal ou l'inverse) a eu lieu au bloc 20.

(haut), il n'y a pas de transfert. A l'inverse, si les performances au début de la seconde condition sont similaires à celles à la fin de la première condition (bas), il y a un transfert complet. Dans certaines situations, il peut y avoir un transfert partiel (milieu).

1.2.2.2 Processus de haut- ou bas-niveaux

Une question importante concerne les niveaux de traitement sous-jacents à l'apprentissage perceptif. Ces mécanismes pourraient en effet intervenir de manière précoce (où les caractéristiques basiques du stimulus, telle que l'orientation ou la fréquence spatiale dans le domaine visuel ou la fréquence du son pur dans le domaine auditif, sont encodées par des populations neuronales différentes) ou plus tardive (par le biais de processus attentionnels ou décisionnels, par exemple).

Nous allons maintenant reporter, à travers différents exemples du domaine de la perception visuelle et auditive, certains arguments en faveur de l'implication de processus haut- ou bas-niveaux dans l'apprentissage perceptif.

La spécificité de l'apprentissage a été utilisée comme argument en faveur d'un apprentissage perceptif concernant les niveaux précoces de traitement sensoriels (Gilbert, 1994) dans le sens où elle pourrait être associée à une plasticité de populations neuronales dont la réponse est spécifique à un aspect du stimulus (une orientation, une position rétinienne etc.). Selon cette perspective, deux tâches mettant en jeu des stimuli différents (par exemple, un changement d'orientation) recruteront des populations de neurones différentes dans le cortex visuel primaire, et aucun transfert ne sera observé. Il est toutefois difficile d'expliquer les situations où un transfert se produit (voir plus loin) selon cette perspective.

Il est ainsi important d'examiner dans quels cas un transfert est observé et dans quels cas l'apprentissage est spécifique. Dans le domaine visuel, la spécificité de l'apprentissage a été mise en avant dans de nombreux cas différents. Poggio et al. (1992) ont montré une absence de transfert entre deux orientations à une tâche de Vernier (Figure 1.3). La Figure 1.5 montre que les performances diminuent au bloc 20 lorsque la condition (horizontale ou verticale) change. Une spécificité est également observée concernant différents attribut du stimulus comme la direction du mouvement (Ball & Sekuler, 1987), l'orientation (Fiorentini & Berardi, 1980; Zhang et al., 2010a) et la fréquence spatiale (Fiorentini & Berardi, 1980), la position rétinale (Shiu & Pashler, 1992), l'œil de présentation (Karni & Sagi, 1991), ou encore le contraste (Yu et al., 2004).

1.2.2.3 Spécificité de la fréquence d'entraînement

Dans le domaine auditif, des exemples de spécificité de l'apprentissage perceptif ont également été observés. Dans le cas de la discrimination de fréquence, le transfert entre différentes fréquences est généralement observé, bien que souvent incomplet. Plusieurs études ont été réalisées pour évaluer la spécificité de la fréquence dans l'apprentissage. Demany (1985) a testé les capacités de discrimination de fréquence (Figure 1.2 f) à 200 Hz avant et après une phase d'entraînement. Chaque groupe était entraîné avec une fréquence différente (200, 360, 2500 ou 6000 Hz). Les résultats montrent une différence pré-/post-test équivalente pour les groupes 200, 360 et 2500 Hz, mais plus faible pour le groupe entraîné à 6000 Hz. Cela suggère un certain transfert des fréquences 360 et 2500 Hz à la fréquence 200 Hz. Cependant, ce transfert n'est pas total car il n'est pas observé de la fréquence 6000 Hz à la fréquence 200 Hz.

Dans une étude plus récente, Irvine et al. (2000) ont comparé les résultats de pré- et post-tests à une tâche de discrimination de fréquence (Figure 1.2 d). Les seuils ont été évalués à 5000 et 8000 Hz dans les phases pré-/post, et la moitié des participants a été entraîné à 5000 Hz et l'autre moitié à 8000 Hz. Les résultats montrent un apprentissage pour les deux fréquences, mais cet apprentissage est légèrement plus faible pour la fréquence n'ayant pas fait l'objet de l'entraînement. Les auteurs concluent ainsi à une spécificité partielle de l'entraînement.

Aussi, Delhommeau et al. (2005) ont entraîné plusieurs groupes avec un son pur standard de fréquence différente (750, 1500, 3000, ou 6000 Hz). L'ensemble des groupes a réalisé des pré-/post-tests comprenant l'ensemble des fréquences. Les résultats montrent une amélioration des performances entre les pré- et post-tests pour les fréquences entraînées et non-entraînées (Figure 1.6). Les auteurs notent

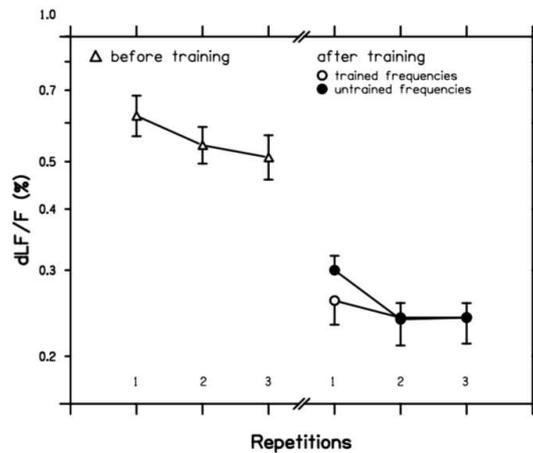


Figure 1.6: Issu de Delhommeau et al. (2005). Performances en fonction des sessions de pré- (triangles, gauche) et post-tests (cercles, droite). Pour les post-tests, la fréquence qui a été entraînée (cercles vides) a été distinguée des fréquences non-entraînées (cercles pleins). Les résultats montrent une diminution des seuils de discrimination de fréquence après l'entraînement pour l'ensemble des fréquences. Toutefois, la première session post-test montre de meilleures performances pour la fréquence entraînée. Cette différence disparaît pour les post-tests suivants.

que lorsqu'on considère uniquement le premier post-test, une différence légère mais significative entre les fréquences entraînées et non-entraînées est observée. L'hypothèse des auteurs est que l'entraînement est en partie spécifique à la fréquence des sons utilisés mais qu'une généralisation rapide se produit lorsque les auditeurs sont à nouveau exposés aux autre fréquences.

Ces différentes études se fondent sur la méthodologie pré-/post-test. Cela a l'avantage de donner une ligne de base des performances, mais a l'inconvénient de produire éventuellement un entraînement sur les stimuli dont le transfert est évalué lors de la phase de pré-test ou post-test (voir plus loin l'effet potentiel du pré-entraînement sur la généralisation de l'entraînement). La présence d'un groupe contrôle ne réalisant que les pré- et post-tests est un moyen d'évaluer une partie des effets d'apprentissage qui ne sont pas liés à l'entraînement.

Roth et al. (2003) utilisent ce type de méthode avec une comparaison pré-/post-test (à 1000, 1100 et 2000 Hz) et un groupe contrôle qui ne réalise pas d'entraînement, mais uniquement les pré-/post-tests. Les participants du groupe test ont été entraînés avec une fréquence de 1000 Hz. Selon les auteurs, le but du groupe contrôle est de prendre

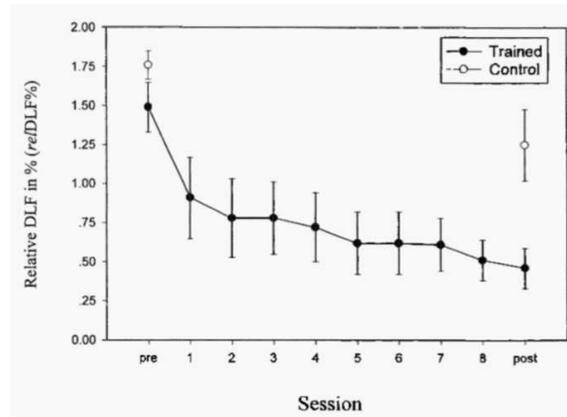


Figure 1.7: Issu de Roth et al. (2003). Performances du groupe contrôle (cercles vides) et du groupe test (cercles pleins). L'amélioration des performances du groupe contrôle est similaire à l'amélioration des performances du groupe test entre le pré-test et la première session d'entraînement.

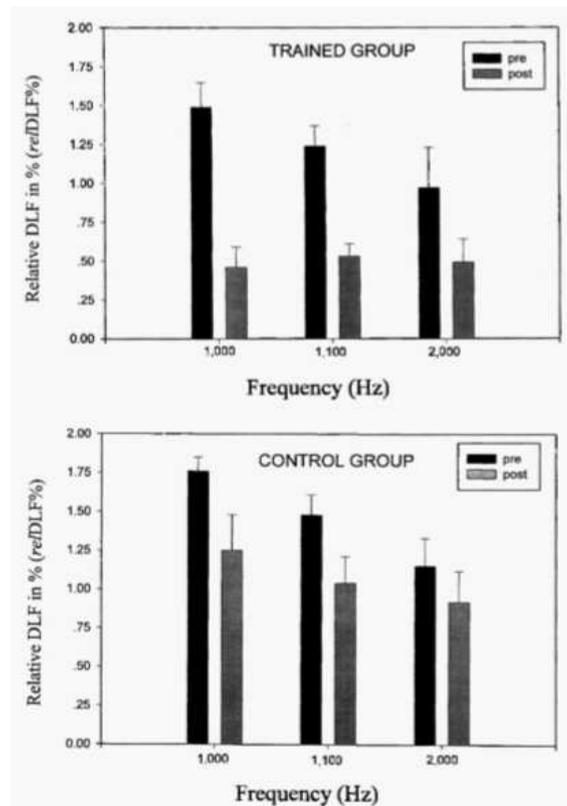


Figure 1.8: Issu de Roth et al. (2003). Transfert de l'apprentissage. Les seuils de fréquence sont représentés pour le groupe test (haut) et le groupe contrôle (bas) pour les pré- (noir) et post-test (gris) ainsi que pour les différentes fréquences.

en compte l'apprentissage procédural, c'est-à-dire l'apprentissage lié à l'exécution de la tâche. On peut préciser que l'apprentissage observé dans le pré-test pourrait également être perceptif (Hawkey et al., 2004), mais dans tous les cas, la présence du groupe contrôle est un apport méthodologique intéressant. Les résultats montrent un effet global du groupe (contrôle vs test), du test (pré vs post) et de la fréquence (1000, 1100, ou 2000 Hz). Les auteurs en concluent qu'une amélioration des performances s'est produite pour les deux groupes (Figure 1.7). Cependant, l'effet reporté est global et il pourrait donc être issu d'une amélioration d'un des groupes uniquement. Ils reportent également une absence de différence significative de performances entre les fréquences au post-test et en concluent que l'apprentissage s'est généralisé aux fréquences non-entraînées (Figure 1.8). Pour être sûr que l'apprentissage n'est pas issu des pré-tests (réalisés sur toutes les fréquences) mais bien de l'entraînement (réalisé avec les sons purs à 1000 Hz), il aurait fallu que l'interaction entre le groupe (contrôle vs test) et le test (pré vs post) soit reportée et qu'elle soit significative. On peut également noter que les pré-/post-tests du groupe test comportaient deux fois plus de mesures (tests sur les deux oreilles séparément) que ceux du groupe contrôle (tests sur une seule oreille), ce qui diminue l'intérêt du groupe contrôle.

Dans une autre étude, Amitay et al. (2005) ont utilisé une procédure entraînement - post-tests pour évaluer la spécificité de la fréquence dans l'apprentissage perceptif. Les auteurs ont mis en place plusieurs conditions dans lesquelles le son pur standard changeait d'un essai à l'autre (le fait de varier la valeur du stimulus est désigné par le terme *roving* ; voir section 1.2.3.1) ou était fixe (pas de *roving*). Dans un des groupes (sans *roving*), les participants étaient entraînés avec des sons purs à 1000 Hz et testés avec des sons purs à 500, 1000, 2000 et 4000 Hz. Les seuils mesurés étaient équivalents pour chaque fréquence suggérant un transfert de l'apprentissage. Cependant, comme le notent les auteurs, il n'y a pas eu de comparaison avec une ligne de base. On ne peut donc pas exclure l'hypothèse de seuils différents selon la fréquence, même avant l'entraînement, perturbant la comparaison.

Demany & Semal (2002) ont également évalué la spécificité de la fréquence d'entraînement dans une tâche de discrimination de fréquence (Figure 1.2 b), en combinant la comparaison de pré-/post-tests et la comparaison des courbes d'apprentissage succédant à une phase d'apprentissage. Dans leur procédure, des pré- et post-tests (nommés *mixed-blocks* et consistant en trois blocs entrelacés correspondant aux trois fréquences testées) comportant des fréquences de 1200, 3000 et 6500 Hz ont été utilisés. Entre ces pré- et post-tests, un entraînement avec une fréquence de 3000 Hz a été réalisé. La comparaison du pré-test et de la première session de post-test montre

une amélioration des performances similaire pour chaque fréquence, suggérant une généralisation de l'apprentissage. Cependant, si l'ensemble des post-tests (6 sessions) est considéré, les courbes d'apprentissage sont significativement différentes selon la fréquence, avec un apprentissage moins marqué pour la fréquence entraînée (3000 Hz). L'hypothèse est donc que qu'un apprentissage associé à la fréquence entraînée a déjà été réalisé et donc que la progression est ensuite moins grande. Ces résultats montrent également que le protocole utilisé pour évaluer le transfert est important et peut conduire à des résultats différents.

Pour résumer, ces différentes études suggèrent qu'il y a une généralisation d'une fréquence à l'autre de l'apprentissage perceptif fréquentiel en audition. Cependant, cette généralisation est souvent incomplète, comme le témoignent les différences de performances parfois observées entre fréquences entraînées et fréquences non-entraînées.

1.2.2.4 Conditions permettant une généralisation

Il semble que la généralisation à de nouveaux stimuli ou tâches dépend fortement du protocole utilisé. Plusieurs aspects expérimentaux semblent influencer la spécificité de l'apprentissage perceptif.

Pré-entraînement

Des travaux dans la modalité visuelle ont montré que la spécificité qui caractérise l'apprentissage perceptif dépend fortement du protocole utilisé. Schoups et al. (1995) ont entraîné des participants à discriminer l'orientation d'un stimulus élaboré à partir de bandes noires et blanches plus ou moins larges. Les résultats ont montré un apprentissage perceptif spécifique à l'orientation entraînée et à la position rétinienne du stimulus. Plus tard, Zhang et al. (2010b) ont repris un protocole identique et ont montré qu'il était possible d'induire un transfert total de l'apprentissage à des positions rétinienne non entraînées. La méthode a été d'ajouter un rapide pré-test (environ 200 essais) à la seconde position rétinienne (celle qui n'est pas entraînée). Cela suggère que la généralisation de l'apprentissage dépend fortement de la procédure utilisée.

Double entraînement

Une autre manière de favoriser une généralisation de l'apprentissage est l'utilisation d'un paradigme de double-entraînement. Xiao et al. (2008) ont mis en place deux

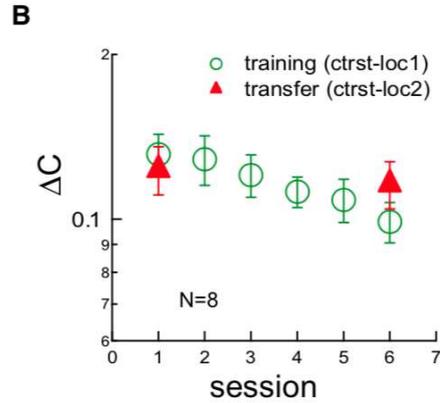


Figure 1.9: Issu de Xiao et al. (2008). Spécificité de la discrimination de contraste. Les participants ont été entraînés à discriminer des Gabors de contrastes différents à une position rétinienne (ctrst-loc1) et testés avant et après l'entraînement à la même tâche mais à une position rétinienne différente (ctrst-loc2). Les résultats suggèrent une spécificité de l'apprentissage.

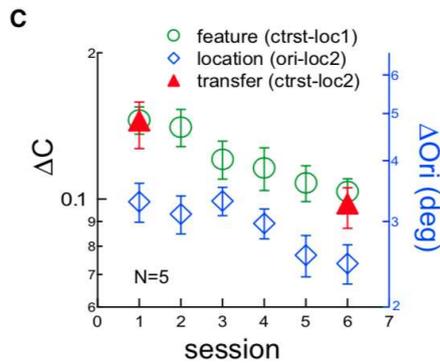


Figure 1.10: Issu de Xiao et al. (2008). Résultats du double entraînement. Comme dans la tâche contrôle (Figure 1.9), les participants ont été entraînés à discriminer des Gabors de contraste différent à une position rétinienne (ctrst-loc1) et testés avant et après l'entraînement à la même tâche mais à une position rétinienne différente (ctrst-loc2). En outre, une seconde tâche a été réalisée (bloc alternés) dans laquelle les participants devaient réaliser une discrimination d'orientation à la seconde position rétinienne (ori-loc2). Cette procédure de double entraînement a induit un transfert de l'apprentissage concernant la discrimination de contraste de la première à la seconde position rétinienne.

tâches dans lesquelles les participants devaient discriminer le contraste ou l'orientation de fréquence spatiales (patch de Gabor). La spécificité de la position rétinienne de l'apprentissage sur une tâche de discrimination de contraste (*quel Gabor a le contraste le plus élevé ?*) a tout d'abord été répliquée. Pour cela, les participants ont été testés à deux positions rétiniennes, entraînés à l'une des deux, puis re-testés aux deux. Les résultats montrent une amélioration à la position entraînée qui ne transfère pas à la position non-entraînée (Figure 1.9). Cependant, lorsque la même tâche est réalisée en alternant des blocs d'une tâche différente (discrimination d'orientation) positionnée à la seconde position rétinienne, la spécificité de l'apprentissage semble disparaître (Figure 1.10).

Durée de l'entraînement

Jeter et al. (2010) ont étudié la dynamique temporelle de l'apprentissage perceptif sur une tâche de discrimination d'orientation et ont montré que l'apprentissage était de plus en plus spécifique au fur et à mesure que l'apprentissage progresse. Au début, il y a transfert, mais celui-ci disparaît au profit d'une spécialisation. Cela est en concordance avec la Reverse Hierarchy Theory (Ahissar & Hochstein, 2004) qui propose que l'entraînement est tout d'abord associé aux niveaux élevés de traitements sensoriels et donc qui se généralise bien, puis aux niveaux plus bas associés à moins de généralisation.

Entraînement inter-tâches (Cross-task training)

Le principe d'un entraînement inter-tâches est d'alterner la pratique de plusieurs tâches différentes. Il semble que cette alternance a des effets sur la généralisation de l'entraînement.

Dans le domaine de l'audition, Wright et al. (2010) ont montré qu'une certaine quantité d'entraînement, insuffisante pour donner lieu à un apprentissage, pouvait être efficace lorsqu'une autre tâche, non directement reliée à l'entraînement mais utilisant les mêmes stimuli, était réalisée par les participants (Figure 1.11). Les auteurs ont comparé différents protocoles d'apprentissage perceptif, dont le nombre d'essais était insuffisant pour conduire à un apprentissage perceptif. Un groupe réalisait chaque jour une tâche d'identification de fréquence alternée avec une tâche contrôle (association de nombres et de symboles) dans le silence (*Freq-alternating-with-Silence*). Un second groupe alternait la tâche fréquence avec la tâche contrôle mais était en outre exposé aux stimuli utilisés dans la tâche fréquence (*Freq-alternating-with-Sound*). Un troisième groupe réalisait en alternance la tâche fréquence avec une tâche

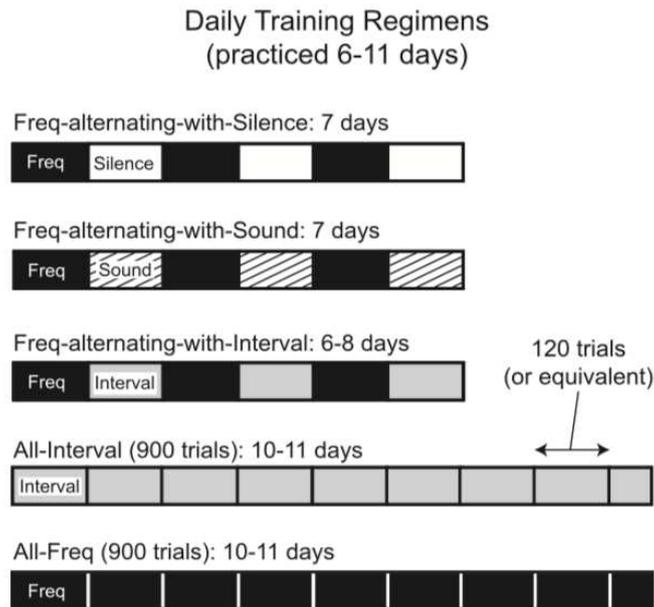


Figure 1.11: Issu de Wright et al. (2010). Illustration des différents groupes d'entraînement. Les blocs *Freq* consistent en une tâche d'identification de fréquence. On voit que le groupe *All-Freq* réalise un nombre plus important d'essais sur cette tâche.

de discrimination d'intervalles temporels (*Freq-alternating-with-Interval*). Un autre groupe ne réalisait que la tâche fréquence (*All-Freq*) avec un nombre d'essais plus important.

Les résultats sont présentés sur la Figure 1.12 et montrent que les protocoles *Freq-alternating-with-Sound*, *Freq-alternating-with-Interval* et *All-Freq* ont produit un apprentissage contrairement au protocole *Freq-alternating-with-Silence*. Les résultats du groupe *All-Freq* montrent simplement qu'un nombre d'essais plus important était nécessaire pour qu'il y ait apprentissage. Cependant, pour une même quantité d'essais d'entraînement, la différence de performances entre *Freq-alternating-with-Sound* et *Freq-alternating-with-Silence* suggère que l'ajout d'une tâche dans laquelle les stimuli sont entendus (mais non directement pertinents pour la tâche) permet de produire un apprentissage. De plus, les expériences suivantes montrent qu'il n'est pas nécessaire que les tâches soient alternées (mais organisées sous la forme de blocs). Cependant, lorsque les blocs correspondant à la seconde tâche ont été réalisés 4 heures après la tâche d'intérêt, aucun apprentissage n'a été observé. L'utilisation d'une fréquence différente pour les sons de la seconde tâche bloque également l'apprentissage, suggérant que l'exposition aux sons purs de la fréquence test est nécessaire. Enfin,

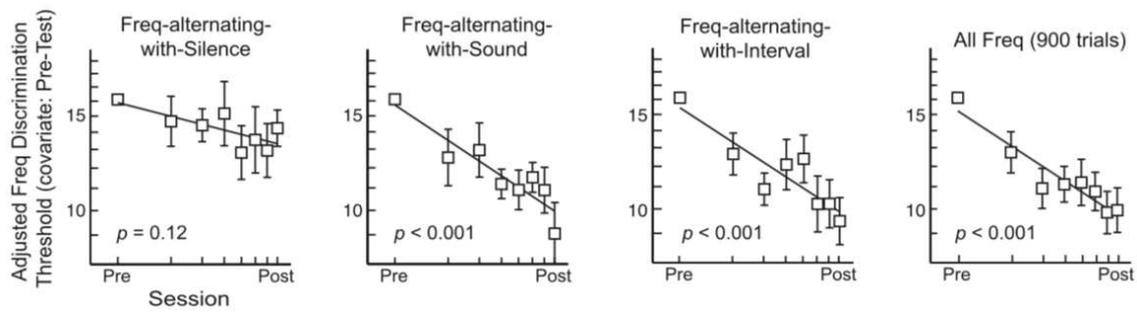


Figure 1.12: Issu de Wright et al. (2010). Présentation du seuil à la tâche d'identification de fréquence en fonction du bloc pour les groupes *Freq-alternating-with-Silence*, *Freq-alternating-with-Sound*, *Freq-alternating-with-Interval* et *All-Freq*.

l'exposition seule aux sons purs de la fréquence test ne suffit pas à produire un apprentissage. Les travaux de Wright et al. (2010) suggèrent donc qu'un apprentissage perceptif équivalent pourrait être obtenu en divisant par deux la pratique des participants grâce à l'ajout de phases de simple exposition.

L'effet d'imbriquer différentes tâches dans les protocoles d'apprentissage perceptif est également observé dans le domaine visuel. Par exemple, Szpiro et al. (2014) ont mis en place un protocole d'entraînement dans lequel les participants devaient réaliser une tâche de discrimination d'orientation et une tâche de discrimination de fréquence spatiale (sur des stimuli similaires). La durée de l'entraînement a été choisie assez faible pour ne pas produire d'apprentissage. Les auteurs montrent cependant qu'une alternance des deux tâches permettait d'obtenir un apprentissage. Il semble donc qu'une pratique continue de la tâche n'est pas nécessaire pour donner lieu à un apprentissage. Les auteurs expliquent ces résultats grâce à l'hypothèse de Wright et al. (2010). Wright et al. (2010) proposent que la tâche sensibilise le système sensoriel de manière à ce que la phase d'exposition aux mêmes stimuli soit efficace. Szpiro et al. (2014) ajoutent que dans leur cas, il est possible que la présence de stimuli identiques dans des contextes différents favorise l'apprentissage car l'utilisation des mêmes stimuli pour réaliser deux tâches différentes donnerait lieu à deux signaux d'erreur, et donc à une représentation sensorielle renforcée. On peut toutefois noter qu'une amélioration des performances ne concerne que la discrimination de l'orientation et non la discrimination de fréquence.

Outre ces facteurs influençant la spécificité de l'apprentissage, une généralisation d'une position rétinienne à une autre peut être observée lorsque l'attention exogène

est dirigée sur une position rétinienne non-entraînée (Donovan et al., 2015). Il semble également que la manière dont est répartie le nombre d’essais effectués a une influence sur l’apprentissage et sur le transfert (Aberg et al., 2009). Dans cette étude, tous les participants ont réalisé 1600 essais d’une tâche de chevrons. Le transfert est évalué avec la direction orthogonale.

- groupe 1 : 800 essais/session. 1 session par jour pendant 2 jours.
- groupe 2 : 400 essais/session. 1 session par jour pendant 4 jours.
- groupe 3 : 160 essais/session. 1 session par jour pendant 10 jours.
- groupe 4 : 160 essais/session. 1 session par semaine pendant 10 semaines.

Aucun apprentissage n’a été observé pour les groupes 3 et 4. Les groupes 1 et 2 ont amélioré leurs performances, mais un transfert à un chevron orthogonal n’a été observé que pour le groupe 2.

Ces effets complexes suggèrent que le transfert de l’apprentissage perceptif n’est pas systématique et que des paramètres expérimentaux multiples peuvent déterminer si, pour un même stimulus et des tâches très similaires, l’apprentissage sera spécifique ou généralisable.

1.2.3 Incertitude du stimulus

Un autre paramètre expérimental influençant le résultat de tâches d’apprentissage perceptif concerne la prédictabilité du stimulus par le participant. Par exemple, la fréquence des sons purs standards des tâches de discrimination et d’identification de fréquence présentées sur les Figures 1.1 et 1.2 peut être stable pendant toute l’expérience ou choisie aléatoirement à chaque essai. En cas de variation aléatoire, on parle alors de *roving* de la fréquence du son pur standard. Dans ce cas, le participant ne peut pas prévoir la fréquence du sons pur qui doit servir de base à la comparaison.

1.2.3.1 Effet du *roving* sur l’apprentissage perceptif fréquentiel

Amitay et al. (2005) ont évalué l’effet du *roving* sur l’apprentissage et le transfert d’une tâche d’identification de fréquence. Plusieurs groupes ont été créés dans lesquels la fréquence du son pur standard était soit fixe (1000 Hz : pas de *roving*), soit légèrement différente à chaque essai (*roving* étroit), soit très différente à chaque essai (*roving* large). Les participants ont été entraînés pendant 7 blocs.

Les résultats ont montré une amélioration des performances pour l’ensemble des groupes (Figure 1.13). De manière à prendre en compte la variabilité individuelle

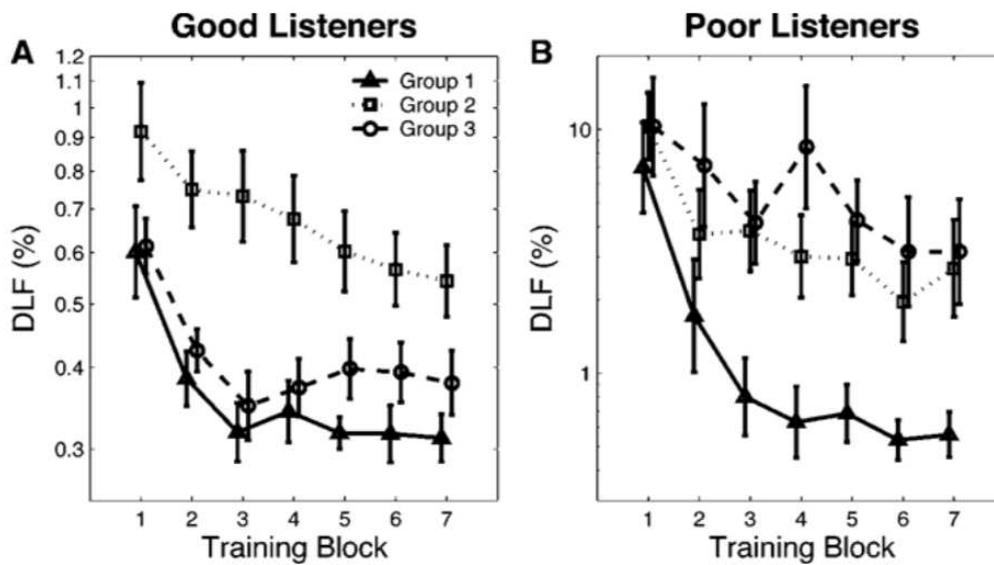


Figure 1.13: Issu de Amitay et al. (2005). Seuils d'identification de fréquence des participants en fonction du bloc d'entraînement et du groupe. Groupe 1 : sans *roving*, groupe 2 : *roving étroit* et groupe 3 : *roving large*. Les participants ont été analysés séparément en fonction de leurs performances de départ (bonnes à gauche et mauvaises à droite).

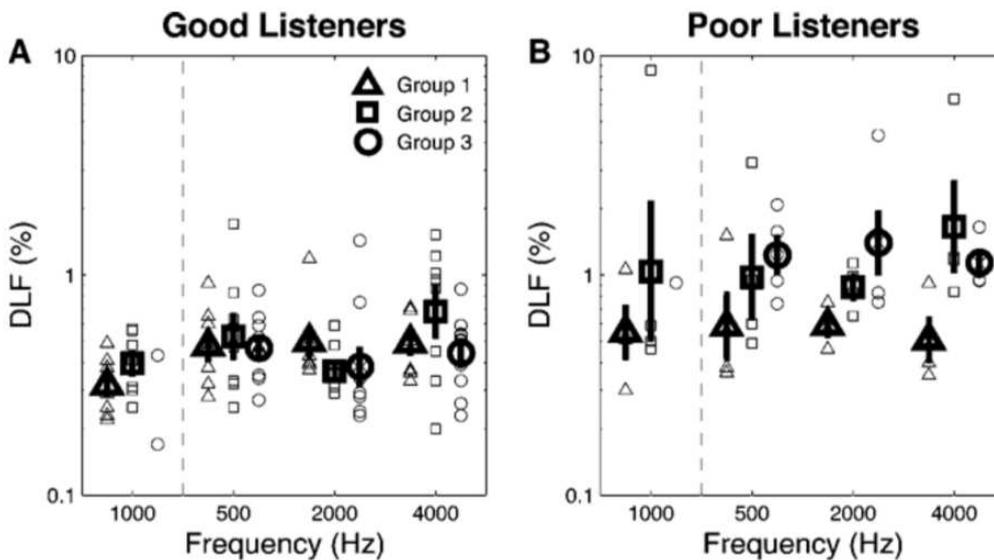


Figure 1.14: Issu de Amitay et al. (2005). Seuils d'identification de fréquence des participants au post-test pour différentes fréquences, entraînées (1000 Hz) et non-entraînées (500, 2000 et 4000 Hz) et pour les conditions sans *roving* (groupe 1, triangles), avec *roving étroit* (groupe 2, carrés) et avec *roving large* (groupe 3, cercles). Les participants ont été analysés séparément en fonction de leurs performances de départ (bonnes à gauche et mauvaises à droite).

dans les effets d'apprentissage et de transfert, deux groupes ont été créés selon la performance initiale des participants (deux tiers des participants avaient de bonnes performances). On peut voir que les participants du groupe sans *roving* (groupe 1) sont ceux dont l'entraînement a été le plus efficace. Dans le groupe des participants performants, le *roving* large (groupe 3) ne semble pas impacter l'apprentissage, contrairement au *roving* étroit (groupe 2) qui le réduit fortement. Cependant, pour les participants moins performants, les deux types de *roving* semblent avoir une influence néfaste sur l'apprentissage.

De manière à évaluer le transfert à des fréquences n'ayant pas fait l'objet de l'apprentissage, un post-test a été réalisé. La Figure 1.14 montre les résultats des seuils calculés à la fin de l'entraînement pour différentes fréquences (la fréquence de 1000 Hz est celle qui a été utilisée pendant l'entraînement). Un transfert aux fréquences non-entraînées caractérise le groupe des participants performants, pour l'ensemble des conditions (avec ou sans *roving*). Concernant le groupe des faibles participants, il semble que les seuils sont plus bas pour le groupe 1 (pas de *roving*), mais le seuil de significativité n'a pas été atteint. Ainsi, ces résultats suggèrent que le *roving* peut ralentir l'apprentissage (*roving* étroit pour les bons participants et *roving* étroit ou large pour les faibles participants), mais qu'un transfert à des fréquences non-entraînées est cependant possible.

Une chose étonnante cependant est que les seuils des participants sont parfois plus bas au post-test qu'à la dernière session d'entraînement, alors même que des performances asymptotiques semblent avoir été atteintes. Par exemple, les performances des faibles participants dans les conditions *roving* atteignent environ 3% contre 1% à 2% aux post-tests selon les fréquences. Il aurait pu être intéressant de tester statistiquement cette comparaison. On peut également noter que l'absence de pré-test ne permet pas d'éliminer l'hypothèse de différences intrinsèque de performances selon la fréquence.

Il est également intéressant de questionner l'effet néfaste plus important du *roving* étroit par rapport au *roving* large pour les bons participants. Selon la RHT, une explication est que dans le cas du *roving* large, les valeurs peuvent être clairement catégorisées comme différentes par les populations de neurones des aires haut-niveau. Les mécanismes *top-down* peuvent ainsi être mis en place plus efficacement que dans le cas du *roving* étroit. En outre, l'effet du *roving* dépend des performances de base des participants. Pour les bons participants, il y a ralentissement de l'apprentissage pour la condition *roving* étroit et pas d'effet sur le transfert. Pour les mauvais

participants, il y a diminution de l'apprentissage pour les deux conditions de *roving*, qui affecte en outre le transfert à d'autres fréquences mais pas à d'autres conditions.

1.2.3.2 Organisation temporelle du stimulus

Un des arguments en défaveur de mécanismes bas-niveau associés à l'apprentissage perceptif est l'influence de l'incertitude du stimulus, qui a été observée dans le domaine visuel (Adini et al., 2004; Yu et al., 2004) ou auditif (discrimination de mots dans le bruit : Nahum et al., 2010). Dans ces cas-là, il y a apprentissage lorsque les essais sont organisés par blocs mais pas quand les conditions sont en ordre aléatoire. Le *roving* des valeurs du stimulus inhibe ainsi l'apprentissage perceptif. Par exemple, aucun apprentissage perceptif n'est observé dans une tâche de discrimination de contraste lorsque les valeurs de référence changent aléatoirement d'un essai à l'autre (Kuai et al., 2005; Yu et al., 2004). Il a cependant été montré qu'il peut y avoir apprentissage malgré le *roving* lorsqu'une séquence temporelle régulière des valeurs du stimulus est utilisée (Kuai et al., 2005). Cette étude a montré que l'effet néfaste du *roving* sur l'apprentissage pouvait être annulé lorsque les valeurs du stimulus étaient présentées de manière régulière et non aléatoire. Les auteurs comparent par exemple l'entraînement à une tâche de discrimination de contraste dans deux conditions. Dans la première, les valeurs de contraste (0.2, 0.3, 0.47 ou 0.63) étaient choisies de manière aléatoire à chaque essai. Dans la deuxième condition, une séquence était mise en place : 0.2, 0.47, 0.3, 0.63. Un apprentissage perceptif n'a été observé que lorsque les valeurs de contraste sont organisées temporellement, et non lorsqu'elles sont aléatoires. Selon la RHT, l'organisation temporelle permettrait d'étiqueter les différentes valeurs possibles et ainsi d'améliorer les processus *top-down*.

Toujours selon la RHT, les populations neurales impliquées pour réaliser une tâche perceptive sont en premier lieu de haut-niveau. Lorsque la tâche nécessite une résolution importante, comme dans le cas d'une discrimination de fréquences très proches par exemple, des processus *top-down* sont initiés de manière à recruter les populations neurales plus bas niveau dont la spécificité permet une représentation du stimulus dans des détails suffisant pour que la tâche soit réalisée. Dans ce cas, l'apprentissage perceptif consisterait en une plus grande accessibilité de ces populations. Ainsi, l'incertitude du stimulus rendrait l'apprentissage plus difficile par l'obstruction de ces processus *top-down*. Déterminer quelles populations sont informatives pour la tâche nécessiterait un certain nombre de répétitions consécutives. Dans ce cas, les populations bas-niveau ne pourraient pas être sollicitées et la tâche devra être réalisée à partir des populations haut-niveau. La théorie prédit alors une

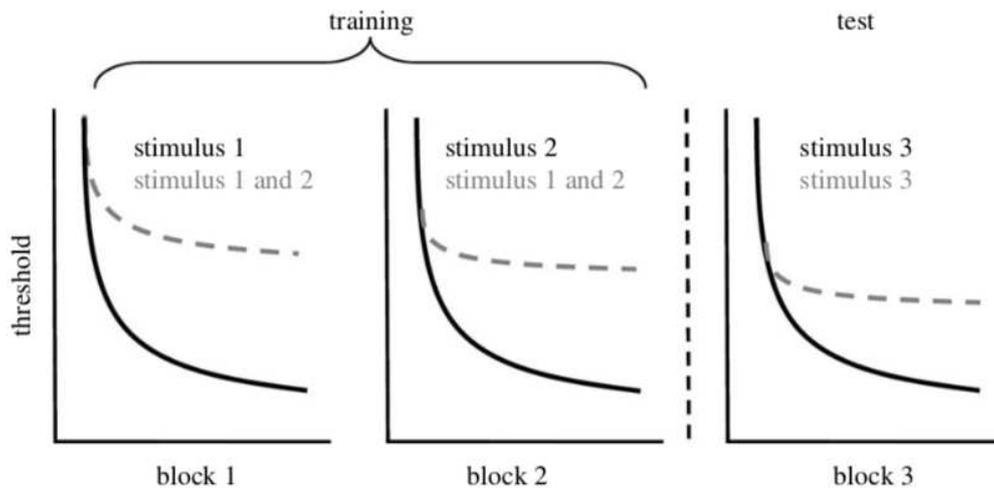


Figure 1.15: Issu de Ahissar et al. (2009). Illustration de la spécificité de l'apprentissage. Un entraînement perceptif est réalisé dans les blocs 1 et 2. Les lignes complètes représentent un entraînement par blocs : le stimulus 1 est utilisé dans le bloc 1 puis le stimulus 2 est utilisé dans le bloc 2. Les lignes pointillées représente un entraînement avec *roving* : les stimuli 1 et 2 sont utilisés dans les blocs 1 et 2 de manière entrelacée. Lorsqu'un test est réalisé avec un stimulus qui n'a pas été utilisé dans l'entraînement (stimulus 3), l'entraînement avec *roving* donne lieu à un transfert contrairement à l'entraînement par blocs.

amélioration limitée des performances sur la tâche en question, mais également une plus grande généralisation (Figure 1.15).

Une hypothèse proposée par Zhang et al. (2016) concernant l'effet du *roving* est que les sons purs doivent être stockés en mémoire et mis à jour à chaque nouvel essai, impliquant des ressources de mémoire de travail plus importantes. De manière à vérifier le rôle de la mémoire de travail dans ces processus, les auteurs ont comparé l'effet d'un entraînement de discrimination de fréquence avec ou sans *roving* (restreint malgré tout aux fréquences : 900, 950, 1000, 1050, et 1100 Hz, ce qui constitue un *roving* étroit avec un nombre limité de fréquences possibles) sur l'amélioration des performances de mémoire de travail. Pour cela, un groupe a été entraîné avec une tâche de discrimination de fréquence dans une condition fixe (sans *roving* : FDF) et un autre groupe dans une condition *roving* (FDR). Un groupe contrôle a également réalisé les pré-/post-tests mais sans entraînement. Les résultats ont tout d'abord confirmé de meilleures performances dans la condition sans *roving* (Figure 1.16, fixe : **a**, *roving* : **b**). Ensuite, la Figure 1.16, **c**. montre qu'une amélioration entre pré- et post-tests est visible à la tâche de mémoire de travail uniquement pour le groupe entraîné dans la condition *roving*. Les auteurs suggèrent que cela met en évidence le rôle de la mémoire de travail dans l'amélioration des performances de perception de la fréquence dans des situations d'incertitude du stimulus.

Pour conclure, on peut noter l'analogie énoncée par Sagi (2011). La réalisation d'une tâche perceptive peut être considérée comme un problème de modélisation statistique dans lequel le but des observateurs est de construire un modèle à partir de l'expérience permettant une prise de décision efficace. L'absence de transfert est considéré comme un *overfitting* : il y a une amélioration des performances avec les stimuli en question mais pas de généralisation. Cela est causé par le fait que des particularités du stimulus utilisé dans l'entraînement sont apprises alors qu'elles ne sont pas utiles pour d'autres conditions. L'utilisation de stimuli plus variés est ainsi l'équivalent de l'ajout de données à un modèle qui est un moyen d'éviter l'*overfitting*.

Pour résumer, plusieurs facteurs reportés dans les études citées plus haut semblent avoir un impact sur l'apprentissage perceptif :

- La durée de l'entraînement : un entraînement court semble mieux se généraliser qu'un entraînement long.
- L'entraînement inter-tâche, le pré-entraînement ou le double entraînement sont des procédures qui semblent favoriser la généralisation de l'apprentissage.

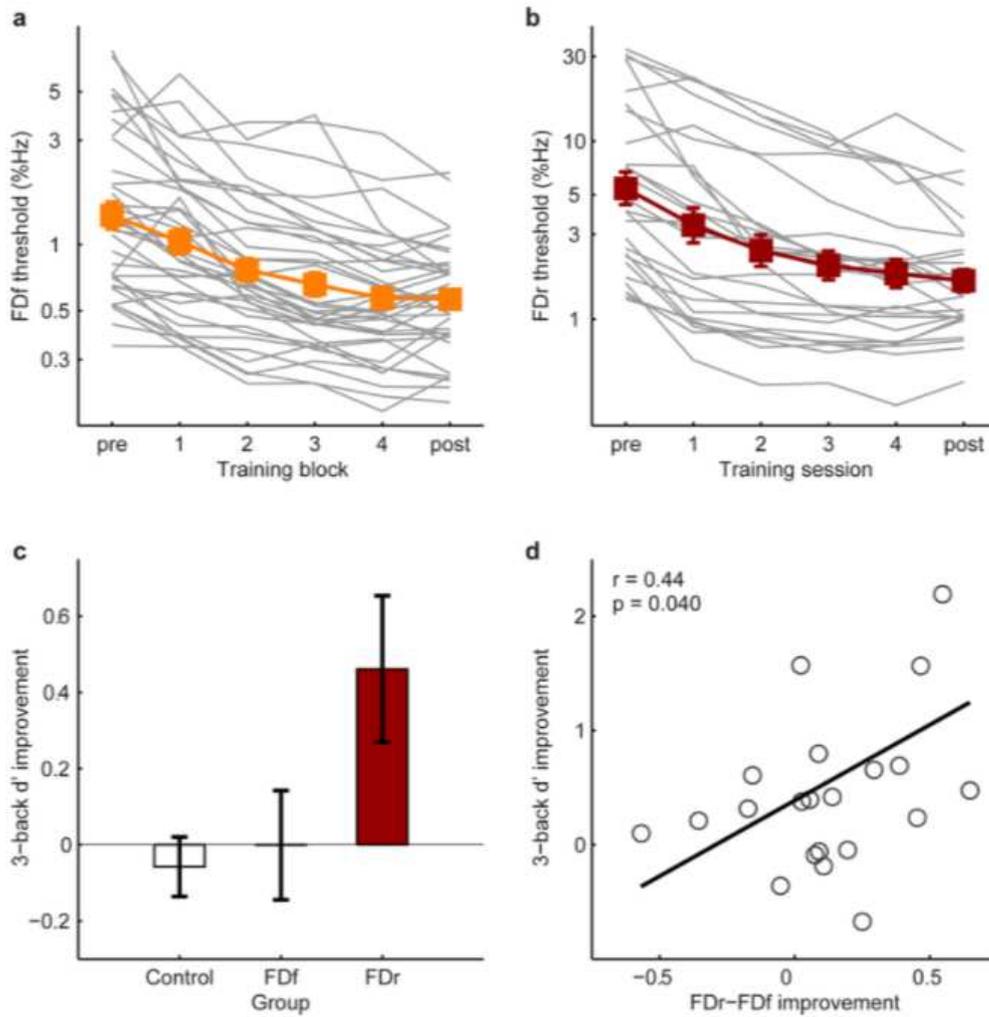


Figure 1.16: Issu de Zhang et al. (2016). Performances à la tâche de discrimination de fréquence dans la condition fixe (FDf, sans *roving*, **a**) et avec *roving* (FDr, **b**) aux pré- et post-tests et sessions d'entraînement. **c.** L'amélioration à la tâche de mémoire de travail est représentée pour les groupes contrôle, fixe et *roving*. **d.** Différence d'amélioration des participants du groupe *roving* entre la tâche *roving* et fixe [(pré-test FDr - post-test FDr) - (pré-test FDf - post-test FDf)] en fonction de l'amélioration à la tâche de mémoire de travail.

- Le *roving* semble empêcher l'apprentissage dans certaines études (perception du contraste) alors que cela n'est pas le cas dans d'autres études (perception de la hauteur).
- L'organisation temporelle du stimulus : la prédictabilité (lorsque la séquence des valeurs de stimuli est prédictible) permet l'apprentissage perceptif dans les cas où l'incertitude (*roving*) empêche l'apprentissage (études sur la perception du contraste).

1.3 Apprentissage et masquage informationnel

Dans cette section, nous allons réaliser une revue de la littérature sur l'amélioration de la capacité d'attention sélective auditive, plus particulièrement dans les tâches de masquage informationnel. Dans un premier temps, une revue illustrée des différents stimuli utilisés dans le cadre des tâches de masquage informationnel sera présentée (1.3.1). Dans un second temps, et comme pour l'apprentissage perceptif fréquentiel, l'effet de l'incertitude du stimulus sera questionné (1.3.2). Ensuite, la question de la similarité entre cible et masqueur et l'effet sur la quantité de masquage observé sera examinée (1.3.3). Enfin, les éléments de la littérature concernant le phénomène d'apprentissage dans ces tâches seront mis en avant (1.3.4).

Attention sélective

L'attention sélective désigne la capacité à focaliser son attention sur une source sonore tout en ignorant le reste de l'environnement. Cette aptitude est cruciale dans la vie de tous les jours car elle permet d'ignorer, dans une certaine mesure, la présence de sons parasites. Par exemple pour suivre une conversation dans un environnement bruyant, il est nécessaire d'isoler le signal d'intérêt (la voix de la personne) et d'ignorer les autres sons. Dans une tentative d'expliquer le fonctionnement de cette aptitude à filtrer le monde sonore, Broadbent (1958) a le premier proposé l'idée selon laquelle l'environnement serait intégralement filtré sur la base des caractéristiques acoustiques d'intérêt du son. Cela implique que tous les événements sonores hors de ce filtre attentionnel sont supprimés. On a découvert par la suite que des stimuli hors du champs attentionnel peuvent être traités malgré tout, comme dans le cas où le nom du participant est prononcé (Moray, 1959). La théorie évolue et on parle de filtre atténuateur plutôt que supprimeurs (Treisman, 1960). Plus tard, des études abordent l'attention sélective auditive à partir de la perception de la fréquence en utilisant une méthode *signal-probe* (Greenberg & Larkin, 1968; Scharf et al., 1987).

Cette méthode consiste en une tâche de détection d'un son pur d'une fréquence attendue par le participant (*signal*). Cependant, dans certains essais, une fréquence différente est utilisée (*probe*), et les performances sont comparées pour les différentes fréquences. Les résultats montrent des gains attentionnels liés à la présence d'une amorce uniquement dans une région fréquentielle avoisinant l'amorce. Il semble donc que des mécanismes attentionnels permettent une pondération plus importante des fréquences d'intérêt.

Masquage informationnel

Une manière d'aborder l'attention sélective de manière expérimentale est d'utiliser les tâches dites de masquage informationnel. La première apparition du terme *masquage informationnel* est généralement attribuée à Pollack (1975). Ce type de masquage est souvent décrit par la négative, comme étant un masquage qui n'est pas la forme classique du masquage auditif, désigné ici sous le terme d'énergétique. Au niveau de la cochlée, les sons sont encodés selon une représentation spectro-temporelle. On parle de masquage énergétique lorsqu'il y a chevauchement de la cible et du masqueur en temps et en fréquence dans ces étapes périphériques de traitement. Le terme de masquage informationnel a été utilisé pour désigner les pertes de performances qui ne sont pas liées au masquage énergétique, par exemple dans le cas où masqueurs et cible sont suffisamment éloignés en fréquence. Cet effet est visible malgré une cible présentée à un niveau au-dessus du seuil de détection dans le silence. Au niveau expérimental, le masqueur contient généralement une région fréquentielle protégée, de manière à s'assurer que cibles et masqueurs dépendent de filtres cochléaires différents, et donc à minimiser l'effet de masquage énergétique. Il est possible de considérer que le masquage total correspond à la somme des masquages énergétique et informationnels (Alexander & Lutfi, 2004). En contraste avec cette définition par la négative du masquage informationnel, pour certains auteurs, il désigne l'élévation du seuil perceptif causée par une incertitude du stimulus.

1.3.1 Protocoles de masquage informationnel

Comme pour les protocoles visant à évaluer la perception de la fréquence, il semble intéressant de recenser les différents protocoles expérimentaux utilisés pour évaluer l'attention sélective par le biais du masquage informationnel. De nombreuses variantes, utilisées pour examiner le phénomène ont été utilisées, mais certains critères sont communs à l'ensemble des études que nous allons décrire ici. Les tâches sont généralement de détecter une cible présentée au-dessus du seuil de détection dans le

silence, cachée dans un masqueur plus ou moins complexe. Une région protégée est présente pour les fréquences autour de la cible de manière à minimiser le masquage énergétique. Nous distinguerons deux catégories de stimuli : les complexes multi-tons et les nuages de tons, principalement différents par la composante temporelle aléatoire associée aux nuages de tons.

Le but de ce paragraphe n'est pas de réaliser un relevé exhaustif des études sur le masquage informationnel, mais plutôt de mettre en avant les stimuli utilisés dans les travaux principaux à titre comparatif. Le but est également d'illustrer les travaux rapportés dans la suite de cette section. Par exemple, seuls les travaux qui utilisent des sons purs en guise de stimuli ont été sélectionnés (pour, par exemple, masquage informationnel avec des stimuli de parole, voir Brungart, 2001).

Les Figures 1.17, 1.18 et 1.19 montrent différents stimuli multi-tons utilisés dans les travaux sur le masquage informationnel. Le type de multi-tons peut-être le plus répandu est construit à partir d'une superposition de sons purs de fréquences aléatoires et de durées identiques (références A, B, C, D, F, L, M, N, P, Q). Un de ces sons purs est la cible à détecter (représentée en gras sur les figures) et une région fréquentielle autour de la fréquence de la cible est *protégée* (représentée en bleu sur les figures), c'est-à-dire qui ne contient pas de son pur masquant. Le nombre de composants du masqueur (c'est-à-dire le nombre de sons purs superposés en plus de la cible), l'ambitus des fréquences utilisées, la durée des sons purs, la fréquence de la cible ou la largeur de la région protégée sont des paramètres dont les valeurs diffèrent selon les études présentées.

Une variante de ces stimuli multi-tons est le *multi-burst* qui consiste à présenter une répétition d'occurrences plus courtes de multi-tons (références E, G, H, K, O). Les paramètres énoncés plus haut comme le nombre de composants, la fréquence de la cible, la taille de la région protégée etc. sont également utilisés avec ce type de stimuli. On peut distinguer les stimuli dont les fréquences des sons purs masquants sont identiques à chaque répétition (SB pour *single-burst* et MBS pour *multi-burst same* de E, MBS de G, SB de H, MBS de K) et ceux dont elles sont différentes (MBD pour *multi-burst different* de E, MBD de G, MBD de H, MBD de K, O. Dans certains cas plus rares, la cible n'est pas de fréquence constante au sein d'un essai (J ou O).

Enfin, un autre type de stimuli utilisé pour évaluer les effets de masquage informationnel est le nuage de tons. Il s'agit de stimuli proche de ceux précédemment décrits mais dont l'aspect temporel est également aléatoire. Ainsi, alors que la cible est un son pur répété à un rythme régulier et de fréquence constante, les éléments

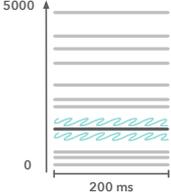
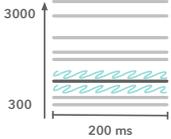
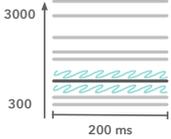
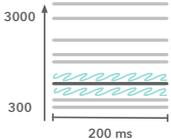
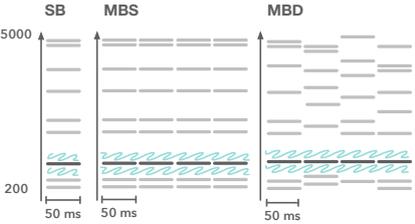
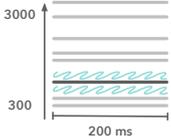
Article	Stimulus	Masker components	Targets Freq. (Hz)	Protected region (octave)	Ref.
Neff & Green, 1987		1-100	250, 1000, or 4000	0.0036 - 0.058 (10 Hz)	A
Neff & Callaghan, 1988		2-200	1000	0.25	B
Neff et al., 1993		2, 6, 10, 30, 100	1000	0.25	C
Neff et al., 1993 and Neff, 1995		2-100	1000	0.25	D
Kidd et al., 1994		1, 2, 4 or 8	1000	0.47	E
Neff & Dethlefs, 1995		2, 5, 6, 10, 20, 50, 100	1000	0.25	F

Figure 1.17: Comparaison des protocoles de masquage informationnel qui utilisent des stimuli multi-tons. Une illustration des stimuli est présentée ainsi que le nombre de composants présents dans le masqueur, la fréquence de la cible à détecter et la taille de la région protégée en fréquence. Il s'agit des paramètres principaux qui diffèrent dans les études.

Article	Stimulus	Masker components	Targets Freq. (Hz)	Protected region (octave)	Ref.
Kidd et al., 2002		2	1000	0.56	G
Kidd et al., 2003		8	1000	0.47	H
Durlach et al., 2003 - EXP 1		8	1000	0.64	I
Durlach et al., 2003 - EXP 2		8	820-1220	0.64	J
Oxenham et al., 2003		8	1000	0.57	K

Figure 1.18: Suite des travaux comportant des stimuli multi-tons.

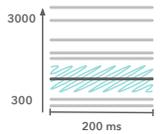
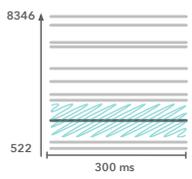
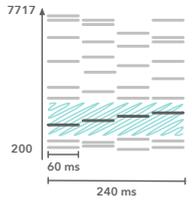
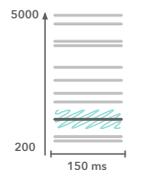
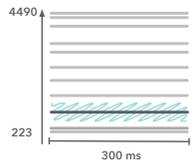
Article	Stimulus	Masker components	Targets Freq. (Hz)	Protected region (octave)	Ref.
Richards & Neff, 2004		2, 6 or 10	1000	0.23	L
Richards et al., 2004 - EXP 1		2, 6 or 10	1000	0.35	M
Alexander et al., 2004		10	2000	1.33	N
Kidd et al., 2011		8	438, 880, 1767, 3549	1.57	O
Cao & Richards, 2012		6	1000	0.44	P
Jones et al., 2014		30	1000	0.33	Q

Figure 1.19: Suite des travaux comportant des stimuli multi-tons.

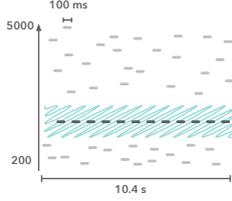
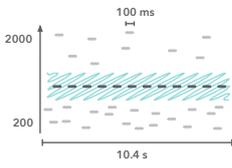
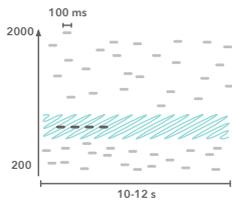
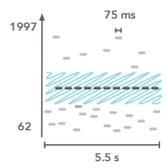
Article	Stimulus	Masker components	Targets Freq. (Hz)	Protected region (octave)	Target/ masker level diff. (dB)	Target SOA (ms)
Gutschalk et al., 2008 and Hausfeld et al., 2017		18 bands	489, 699, 1000, 1,430, 2045, 2924	1.31 - 1.38	-18 or 0	800
Königs & Gutschalk, 2012		18 bands	344, 451, 591, 775, 1016, 1332	0.91 - 0.99	0	800
Wiegand & Gutschalk, 2012		36 bands	836, 1000, 1196	0.57 - 0.59	-6	400
Akram et al., 2014		-	250 - 500	0.33, 0.66, 1 or 1.3	-	100 - 500

Figure 1.20: Comparaison des protocoles de masquage informationnel qui utilisent des nuages de tons. Une illustration des stimuli est présentée ainsi que le nombre de composants présents dans le masqueur, la fréquence de la cible à détecter, la taille de la région protégée en fréquence, la différence de niveau entre la cible et les sons purs du masqueur et le SOA de la cible. Ces paramètres sont ceux qui font l'objet des principales différences dans les études présentées.

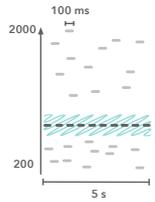
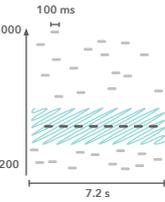
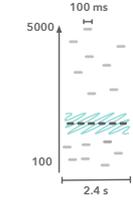
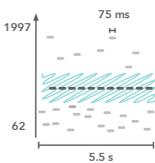
Article	Stimulus	Masker components	Targets Freq. (Hz)	Protected region (octave)	Target/ masker level diff. (dB)	Target SOA (ms)
Dykstra & Gutschalk, 2015		24	589, 619, 902, 947, 1334, 1400, 1931, 2028	0.18 - 0.21	-	500
Dykstra & Gutschalk, 2016		-	489, 699, 1000, 1,430, 2045, 2924	1.31 - 1.38	0	800
Prilop et al., 2017		2-12	846, 950, 1066	0.5	0	400
Akram et al., 2014		-	250 - 500	0.33, 0.66 or 1	0	250

Figure 1.21: Suite des travaux qui utilisent des nuages de tons.

du masqueur sont des sons purs aléatoires en temps et en fréquence. Ces stimuli ont également une bande protégée autour de la fréquence de la cible dans laquelle les sons purs du nuage ne peuvent pas être choisis. L'ambitus des fréquences utilisées, la durée de l'essai, la durée des sons purs, le nombre de bandes (correspondant au nombre de couches fréquentielles du nuage de tons), la fréquence de la cible, la taille de la région protégée, le niveau de la cible par rapport aux éléments individuels du nuage de tons ou la fréquence de répétition de la cible sont autant de paramètres permettant de modifier la difficulté de la tâche. Ces paramètres sont synthétisés dans les Figures 1.20 et 1.21.

De manière générale, ces protocoles de masquage informationnel permettent d'évaluer la capacité à isoler une cible au sein de distracteurs similaires grâce à ses attributs de fréquence. On peut toutefois remarquer que des différences importantes en terme de quantité de masquage sont observées en fonction des paramètres et du protocole utilisé. Principalement, un masquage important est observé lorsque le nombre de composants est important et que la largeur de la zone protégée est réduite. En outre, nous allons voir plus en détail deux caractéristiques principales qui semblent impacter le masquage informationnel : l'incertitude du stimulus et la similarité entre cible et masqueur.

1.3.2 Incertitude du stimulus

1.3.2.1 Effets de l'incertitude

L'incertitude du stimulus est une notion importante à considérer dans le cadre du masquage informationnel. Il s'agit même de la caractéristique déterminante de ce que l'on nomme masquage informationnel pour de nombreux chercheurs (Alexander & Lutfi, 2004 ; Neff, 1995 ; Neff et al., 1993; Oh & Lutfi, 2000). Par exemple, pour Neff (1995), « le masquage informationnel est défini de manière opérationnelle comme étant l'augmentation du seuil produit par l'incertitude du stimulus. » (« *Informational masking is operationally defined as the elevation in threshold produced by stimulus uncertainty* »).

Dans l'étude de Neff & Green (1987), la fréquence et le niveau des éléments du masqueurs étaient modifiés d'un intervalle à l'autre au sein d'un même essai. De manière à évaluer l'effet de cette incertitude sur la quantité de masquage, Neff & Callaghan (1988) (expérience 2) ont comparé les performances de participants dans les quatre combinaisons d'incertitude impliquant une fréquence et une intensité des composants du masqueur fixe ou changeante (Figure 1.17, B). Il faut noter que dans

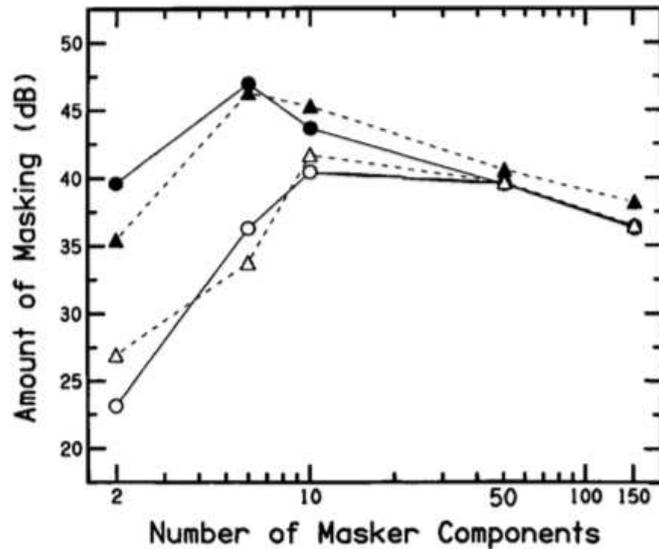


Figure 1.22: Issu de Neff & Callaghan (1988). Quantité de masquage en fonction du nombre de composants du masqueur dans les conditions où la fréquence était aléatoire d'un intervalle à l'autre (symboles pleins) ou fixe (symboles vides) et où l'intensité était aléatoire (triangles) ou fixe (cercles).

la condition fréquence fixe, la fréquence était identique entre les deux intervalles, mais de nouvelles fréquences étaient choisies à chaque essai. La Figure 1.22 synthétise les résultats de ces travaux. On peut voir que la variation de l'amplitude des composants du masqueur ne semble pas impacter la quantité de masquage (différence de masquage entre les symboles triangle ou cercle). En revanche, une différence importante est observée selon la variation de la fréquence d'un intervalle à l'autre pour les masqueurs entre 2 et 10 composants (différence entre symboles pleins ou vides). Par exemple, la différence est d'environ 15 dB pour un masqueur à deux composants. Toutefois, cette différence disparaît lorsqu'un nombre plus important de composants est utilisé. Il est possible que les composants soient plus difficiles à identifier individuellement lorsqu'ils sont présentés en nombre important, conduisant à un gain de performance moins important associé à la prédictibilité des fréquences utilisées.

Cela suggère que l'incertitude en terme de fréquence des composants du masqueur a un impact important sur la quantité de masquage, ce qui ne semble pas être le cas pour l'intensité. Une explication est que l'extraction de la cible est réalisé sur la base d'indices fréquentiels et donc que l'intensité est plus facilement ignorée.

1.3.2.2 Réduction de l'incertitude par l'amorçage

Lors de l'analyse de scènes auditives, l'attention peut se focaliser sur un objet sonore grâce à la sélection des attributs d'intérêt (fréquence, timbre, rythmes, etc.). Cet effet attentionnel *top-down* est également en jeu lorsque la cible est amorcée : l'amorce est définie comme un indice concernant un attribut particulier indiquant où l'attention doit se porter. Par exemple, dans le cas de stimuli complexes (langage), des effets d'amorçage de la position spatiale peuvent être observés (Kidd et al., 2005). Dans le cas de stimuli plus simples, l'amorçage peut également avoir un impact positif sur les performances. On peut toutefois noter que la présence d'une amorce ne suffit pas à éliminer les effets de masquage informationnel, qui peuvent être observés même lorsque la cible est présentée seule avant chaque essai et dans le cas de participants très entraînés (Neff et al., 1993).

Au vu de ces effets complexes, on peut se demander quels mécanismes sont mis en jeu. Est-ce que l'amorce agit sur la réduction de l'incertitude liée au stimulus ? Ou bien est-ce une manière d'orienter l'attention de manière adéquate ? Richards & Neff (2004) (Figure 1.19 L) ont mis en avant le rôle de l'amorçage, même lorsque le masqueur et non la cible sert d'amorce. Ils montrent que le fait de présenter un exemplaire de la cible ou du masqueur avant l'essai est associé à une diminution du seuil de 14.4 dB (amorçage de la cible) ou de 13.9 dB (amorçage du masqueur). Une hypothèse émise par les auteurs est que l'amorce permet une diminution de l'incertitude liée au stimuli et donc une diminution de l'effet de masquage informationnel. Cependant, cela est en contradiction avec le constat qu'une présentation de l'ensemble du stimulus (cible + masqueur) sous forme d'amorce donne lieu à des performances moins bonnes qu'avec une amorce masqueur seul (Richards & Huang, 2003).

Richards et al. (2004) répliquent ce constat (d' supérieur d'environ 1.5 dans la condition amorce masqueur seul par rapport à la condition amorce cible + masqueur ; Figure 1.23). Les auteurs s'assurent également que le gain lié à l'amorçage du masqueur n'est pas expliqué par une comparaison des intensités des stimuli. En effet, l'avantage de l'amorce masqueur seul pourrait venir de la différence d'intensité entre l'amorce et le stimulus (qui lui, est constitué du masqueur seul ou du masqueur et de la cible), indice qui n'est pas présent dans le cas de l'amorce signal + masqueur. Les résultats montrent que l'avantage de l'amorce masqueur seul par rapport à l'amorce masqueur + cible est toujours présent lorsque les indices de niveaux sont supprimés. En outre, les auteurs ont testé l'hypothèse selon laquelle l'avantage serait issu de la détection de l'ajout d'un composant avec l'amorce masqueur seul. Pour cela, les participants devaient comparer deux intervalles contenant un masqueur de N

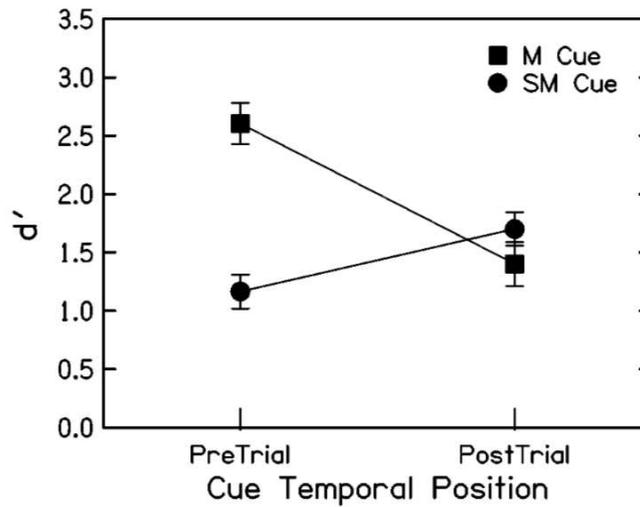


Figure 1.23: Issu de Richards et al. (2004). Résultat de l'expérience 1. Performances moyenne pour une amorce précédant (*PreTrial*) ou suivant (*PostTrial*) l'essai, dans le cas d'une amorce masqueur seul (carrés) ou masqueur + cible (cercles).

composants et un autre de $N + L$ composants. La seule différence avec l'expérience précédente (lorsque $L = 1$) était que le composant supplémentaire était aléatoire et non fixé à 1000 Hz comme la cible de l'expérience 1. En outre, il est intéressant de noter que lorsque la fréquence de la cible est aléatoirement choisie à chaque essai, les résultats sont similaires et l'avantage de l'amorce masqueur seul par rapport à l'amorce masqueur + cible disparaît.

Ces résultats impliquent que l'hypothèse d'une diminution de l'incertitude liée au stimulus ne peut pas, à elle seule, expliquer l'effet de l'amorçage. Une hypothèse alternative est que l'amorçage de la cible permet une attention focalisée sur la bonne bande de fréquence. Toutefois, cela n'explique pas les gains associés à l'amorçage du masqueur. En outre, l'hypothèse d'une adaptation au masqueur qui pourrait expliquer une cible plus saillante dans l'essai est également contredite par Richards et al. (2004) qui montrent que l'avantage lié à l'amorce du masqueur disparaît dans des conditions d'incertitude minimales (fréquences du masqueur identiques pour l'ensemble des essais).

Dans une autre étude, l'efficacité de présenter un exemplaire du masqueur seul avant l'essai est confirmé (Kidd et al., 2011). La tâche est légèrement différente de celles de Richards et al. (2004). Les participants devaient identifier un motif

fréquentiel de quatre sons purs parmi 6 possibilités en présence d'un masqueur multi-tons (Figure 1.19 O). Les résultats montrent une amélioration des performances lorsque le masqueur seul est amorcé.

Cet avantage lié à l'amorçage du masqueur peut être mis en relation avec le concept d'augmentation auditive (*enhancement*). Ce phénomène est par exemple observé lorsqu'un son complexe est présenté deux fois successivement et qu'un composant supplémentaire est présent dans la deuxième présentation. La saillance de ce composant est alors augmentée. Feng & Oxenham (2015) montrent que les performances à une tâche de discrimination ou d'identification de fréquence en présence d'un masqueur multi-tons étaient améliorées par la présence d'un précurseur présentant le masqueur seul. Cet effet dépend du décours temporel de la présentation du précurseur et de l'essai et pourrait être expliqué par le principe du groupement perceptif : le précurseur et le masqueur sont groupés grâce à leurs caractéristiques spectrales et temporelles, permettant un détachement de la cible comme objet différent. À un autre niveau, une explication parallèle (non exclusive de la première) est celle de l'adaptation neurale, selon laquelle le masqueur serait affaibli par la présence de l'amorce.

1.3.3 Similarité entre cible et masqueur

Outre l'incertitude du stimulus, la similarité entre la cible et le masqueur semble constituer un des paramètres les plus influents sur la quantité de masquage informationnel. Plus cible et masqueur sont similaires et plus le masquage informationnel est fort. Dans le cadre de stimulus écologiques comme la parole dans le bruit, Brungart (2001) montre que la difficulté dépend fortement de la similarité acoustique entre parole cible et masquante (sexes identiques ou différents, par exemple). Nous allons voir plusieurs éléments montrant l'importance de la similarité entre cible et masqueur dans les procédures classiques de masquage informationnel.

1.3.3.1 Composantes spectrale et temporelle

Cet effet de similarité a été observé par Kidd et al. (1994). Les auteurs ont comparé deux types de masqueurs dans une tâche de masquage informationnel avec des stimuli multi-tons (Figure 1.17, E). La cible consistait en quatre répétitions d'un son pur de 1000 Hz. Dans le cas du masqueur *multi-burst same* (MBS), chaque composant était de fréquence constante au sein d'un essai. Dans le cas du masqueur *multi-burst different* (MBD), la fréquence des composants du masqueur changeait à chaque

occurrence. Les résultats ont montré que le masqueur MBS est plus masquant, ce qui est interprété par le fait qu'il est plus similaire à la cible. Cet effet MBS/MBD est observé dans plusieurs études utilisant des stimuli similaires (Durlach et al., 2003b; Oxenham et al., 2003).

Kidd et al. (1994) montrent également que l'organisation temporelle peut être un indice permettant la discrimination. Les auteurs ont élaboré une condition MBS dans laquelle la cible n'intervient pas à toutes les occurrences. Les résultats montrent que les performances sont meilleures que dans la condition MBS classique, ce qui suggère de nouveau l'importance de la similarité entre cible et masqueur.

1.3.3.2 Position spatiale

Les éventuelles différences interaurales entre cible et masqueur peuvent être utilisées comme des indices permettant de détecter la cible. Kidd et al. (1994) (Figure 1.17, E) ont montré que pour une présentation de la cible dans une oreille, le masquage était moins important lorsque le masqueur était présenté dans les deux oreilles plutôt que dans une seule.

Cela a également été montré par Durlach et al. (2003b) : un masquage moins important est observé lorsque la cible est présentée dans une seule oreille et le masqueur dans les deux par rapport à la situation où la cible est également présentée dans les deux oreilles. En outre, Kidd et al. (1998b) ont montré que la quantité de masquage informationnel diminuait lorsque la cible et le masqueur sont séparés spatialement. Cette différence est plus grande dans le cas d'un masqueur informationnel (multi-tons avec une région protégée autour de la fréquence de la cible) que dans le cas d'un masqueur énergétique (bruit blanc).

Pour conclure, un certain nombre des effets de similarité influençant la quantité de masquage ont été reproduits par Durlach et al. (2003b). L'ensemble des expériences présentées par les auteurs montre la forte dépendance de ce phénomène à la similarité entre cible et masqueur. Cette similarité peut concerner la durée du stimulus (Figure 1.18, I) et la quantité de masquage diminue lorsque la cible est plus courte que le masqueur (elle commence après le masqueur). Elle peut aussi être liée à la fréquence (Figure 1.18, J) : lorsque la cible est un son pur modulé en fréquence (*sweep*) vers le haut ou vers le bas, une direction opposée à celle des sons purs du masqueurs (également modulés en fréquence dans une direction unique) est associé à un masquage moins important. La similarité interaurale est également un facteur impactant la quantité de masquage informationnel (voir plus haut). Enfin, les résultats obtenus

par Kidd et al. (1994) sont répliqués et montrent qu'un protocole de type MBS est associé à plus de masquage qu'un protocole MBD.

1.3.4 Apprentissage perceptif et masquage informationnel

Une question centrale aux travaux expérimentaux qui ont été réalisés dans le cadre de cette thèse est celle de l'amélioration des performances suite à l'entraînement, et cette question se pose aussi pour des tâches de masquage informationnel. En particulier, est-il possible d'améliorer son aptitude à focaliser son attention sur un objet sonore particulier ? Est-il possible de réduire l'effet néfaste de sons non pertinents sur la perception de sons d'intérêt pour une tâche donnée ? D'un point de vue expérimental, ces questions sont abordées ici dans le cadre des tâches de masquage informationnel. Autrement dit, en faisant l'hypothèse que l'attention sélective à la cible est déterminante dans le paradigme de masquage informationnel, peut-on observer une amélioration des performances avec l'entraînement sur une tâche de masquage informationnel ? Bien que peu nombreux, certains éléments bibliographiques peuvent être récoltés de manière à fournir un début de réponse.

On peut tout d'abord noter que les effets d'entraînement dans les tâches de masquage informationnel sont généralement des effets que les chercheurs souhaitent supprimer car il peut s'agir d'une variable confondante. Pour cette raison, dans un certain nombre de travaux, les participants sont pré-entraînés avant le début de l'expérience (Alexander & Lutfi, 2004; Kidd et al., 1994; Neff & Callaghan, 1988; Neff & Dethlefs, 1995; Richards et al., 2004). Les données de pré-entraînement n'étant pas présentées, il est difficile de savoir s'il y a eu des effets d'apprentissage pendant cette période. Par exemple, Alexander & Lutfi (2004) utilisent une tâche multi-tons et précisent qu'il y a peu d'amélioration des performances sur les 5 blocs de l'expérience principale. Toutefois, cela ne correspond qu'à environ 45 minutes d'expérience et surtout, des blocs de pré-entraînement dans toutes les conditions ont été réalisés.

Malgré un pré-entraînement, l'étude de Neff & Dethlefs (1995) (Figure 1.17, F) permet d'observer des effets d'apprentissage pendant l'expérience principale. Le but premier de l'étude est d'observer les différences inter-individuelles en terme de masquage informationnel ainsi que l'effet de différents paramètres expérimentaux tels que la randomisation des fréquences du masqueur. Une amélioration des performances est observée pour 11 des 49 participants avec une réduction des seuils de 12 dB en moyenne. Comme mentionné plus haut, les participants étaient fortement pré-entraînés et avaient réalisé 600 essais avant le début de l'expérience (elle-même

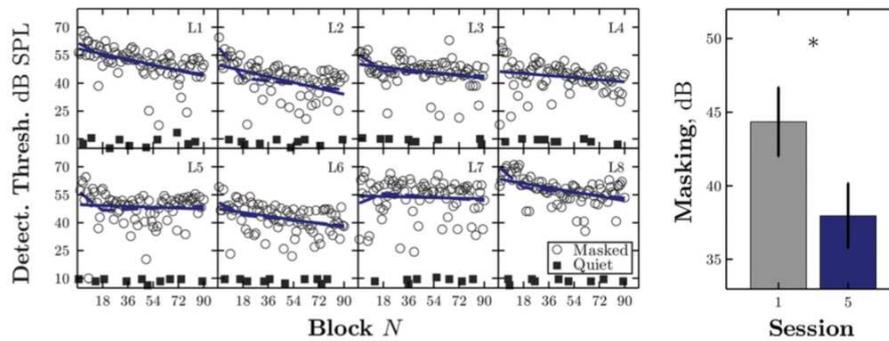


Figure 1.24: Issu de Jones et al. (2014). Gauche : Évolution des seuils dans la condition masquage (cercles vides) et silence (carrés pleins) en fonction du bloc. Droite : Moyennes des quantités de masquage (seuil dans la condition masquage - seuil dans la condition silence) dans le premier (gris) et le dernier bloc (bleu). Les barres d'erreur représentent l'erreur standard.

constituée de 600 à 700 essais). Il reste donc difficile de conclure sur l'effet total de l'entraînement à ce type de tâche.

De manière à évaluer plus précisément ces effets, Jones et al. (2014) ont entraîné des participants sur une tâche de masquage informationnel (Figure 1.19, Q). Il s'agissait d'une tâche à choix forcé à deux alternatives et deux intervalles dans laquelle les participants devaient choisir l'intervalle contenant la cible. Les participants ont été entraînés pendant 45 minutes par jour pendant 5 jours. Des mesures de seuils dans le bruit (masquage informationnel) et dans le silence ont été réalisés chaque jour (16 blocs avec masqueur et 2 blocs dans le silence). Les résultats montrent une amélioration des seuils dans la tâche de masquage de 6.3 dB en moyenne, mais pas de changement des seuils dans le silence (Figure 1.24). Des analyses supplémentaires ont montré que l'effet d'apprentissage n'est pas significatif lorsque les 600 premiers essais sont retirés (pour simuler le pré-entraînement de Neff & Dethlefs, 1995). Cela suggère que cet effet d'entraînement intervient principalement pendant les 600 premiers essais, bien qu'il se poursuive avec une pratique plus étendue.

Le but de l'étude était en outre de distinguer une amélioration des performances liée à une diminution du bruit interne ou à une attention plus sélective. L'hypothèse est que l'attention sélective est opérationnalisée comme la pondération des différentes fréquences réalisée par le participant. D'un point de vue attentionnel, il est possible d'expliquer l'effet de masquage informationnel comme la largeur fréquentielle d'une fenêtre attentionnelle du sujet. Si cette fenêtre est large, la cible et des éléments

du masqueur interagissent et du masquage est observé. Si celle-ci est plus étroite, la cible peut être isolée améliorant les performances de détection. Dans le cadre de cette hypothèse, l'amélioration des performances par entraînement viendrait d'une réduction de la largeur de la fenêtre attentionnelle. Par exemple, dans l'étude de Jones et al. (2014), la cible était toujours de 1000 Hz. Une pondération idéale serait donc d'attribuer un poids maximum à une étroite bande de fréquence autour de 1000 Hz et d'atténuer les poids des autres bandes de fréquence. Les auteurs utilisent la méthode de la corrélation inverse pour évaluer la pondération fréquentielle qui est faite par les participants pour prendre leur décision perceptive concernant la présence ou non de la cible. Le principe est de mettre en relation la réponse du sujet (intervalle 1 ou intervalle 2) et la différence de niveau sonore dans chaque bande de fréquence entre les deux intervalles. Les résultats montrent une évolution de la pondération spectrale entre le premier et le dernier bloc, comme prévu par l'hypothèse d'amélioration de l'attention sélective.

Une autre interprétation serait que l'entraînement permet une diminution de la variabilité décisionnelle du participant. Jones et al. (2014) ont également testé cette deuxième hypothèse en utilisant les pondérations calculées à partir des corrélations inverses de manière à évaluer la variabilité dans la réponse du sujet. Les résultats montrent que l'index de variabilité décisionnelle reste stable. Ces résultats suggèrent que l'amélioration des performances est réalisée presque exclusivement par le biais d'un gain au niveau de l'attention sélective plutôt que par une réduction de la variabilité décisionnelle du participant.

Un autre type de donnée supportant l'idée d'un apprentissage dans le cadre des tâches de masquage informationnel est l'étude de Oxenham et al. (2003) dans laquelle les auteurs comparent un groupe de musiciens et de non-musiciens. Les résultats montrent un avantage des musiciens à la tâche de masquage informationnel utilisée (Figure 1.18, K), suggérant ainsi un apprentissage à long terme de cette population, potentiellement lié à l'attention sélective. Cependant, aucune amélioration n'est observé malgré la pratique d'au moins 400 essais par condition (total : au moins 800 essais) excepté pour les musiciens dans la condition MBD (*multi-burst different*, MBD, voir Figure 1.18, K). Plus de détails sur cette étude seront présentés page 51.

1.3.4.1 Conclusions

Pour conclure, les travaux présentés dans cette section suggèrent que la capacité de résistance au masquage informationnel n'est pas une aptitude figée mais qu'une amélioration des performances peut être observée dans certains cas. Cependant,

l'influence de certains facteurs tels que l'incertitude du stimulus ou la similarité entre cible et masqueur sur la quantité et la qualité de cet apprentissage restent encore difficiles à évaluer du fait du peu d'études disponibles. On peut par exemple se demander si l'influence néfaste de l'incertitude du stimulus (*roving*) observée dans le cas de l'amélioration de la perception de la fréquence est également présente dans le cas du masquage informationnel. On peut également se demander si l'apprentissage observé par Jones et al. (2014) est propre aux stimuli utilisés (multi-tons) ou bien si ce phénomène serait également visible avec d'autres stimuli (tel que les nuages de tons). Il semble, en tout état de cause, que l'utilisation de ces protocoles de masquage informationnel peut fournir des réponses intéressantes à ces questions. Nous utiliserons un tel paradigme d'entraînement pour du masquage informationnel avec nuage de tons au chapitre 3.

1.4 Boucle audio-motrice

Dans cette dernière section bibliographique, nous allons mettre en avant les éléments de la littérature concernant le rôle de la boucle audio-motrice sur l'amélioration des performances perceptives. L'utilisation de la boucle audio-motrice est l'une des motivations des études expérimentales présentées aux chapitres 2 et 3. Plus particulièrement, nous allons considérer ici les cas de la pratique musicale, de la pratique des jeux vidéo et l'utilisation des protocoles immersifs et audio-moteurs.

Les précédentes parties ont mis en avant la question récurrente de la généralisation de l'apprentissage perceptif. Il s'agit d'une question importante autant pour l'aspect théorique (implications des traitements neuronaux mis en jeu) que pour l'aspect pratique (élaboration de protocoles de rééducation sensorielle pour populations cliniques). La question est compliquée par la multitude, et parfois la contradiction, des observations réalisées dans la littérature comme évoqué dans la partie 1.2. Cependant, la littérature a montré que la pratique musicale et la pratique des jeux vidéo étaient associées à un apprentissage perceptif efficace et généralisable.

Une hypothèse séduisante est que la présence d'une boucle perceptivo-motrice pourrait être un moyen d'obtenir un apprentissage perceptif plus efficace. Un aller-retour continu entre action motrice et perception auditive caractérise la pratique musicale. Par exemple, de légères modifications de la position de la main gauche sur un manche de violoncelle réalisées en temps réel permet à l'instrumentiste une correction de la justesse de la note. Dans le cas des jeux vidéo d'action, cette boucle est visuo-motrice dans le sens où une action doit être réalisée en fonction de la

stimulation visuelle (déplacement, présence d'ennemis...). Nous allons finalement voir en plus de détails les quelques exemples d'implémentation de ces caractéristiques à des protocoles d'apprentissage perceptif en audition.

1.4.1 Pratique musicale

1.4.1.1 Pratique musicale et perception de hauteur

La pratique musicale implique une perception précise des différentes caractéristiques du son, telles que l'organisation temporelle, la fréquence ou le timbre. Témoin de la plasticité cérébrale associée à des tâches écologiques, l'effet de la pratique musicale a été montré sur diverses tâches auditives. Nous nous focaliserons dans un premier temps sur l'effet de la pratique musicale sur la perception de hauteur. La hauteur joue un rôle majeur dans la perception musicale et il ne semblerait donc pas étonnant que la pratique musicale soit associée à un développement accru de cette aptitude perceptive.

Cette hypothèse semble supportée par diverses études expérimentales. Spiegel & Watson (1984) ont mesuré les capacités de perception de la fréquence sur des sujets musiciens et non-musiciens, et ont montré une différence entre les deux groupes (médiane des seuils 3 fois plus bas pour les musiciens). Plus tard, Kishon-Rabin et al. (2001) ont également comparé musiciens et non-musiciens sur une tâche d'identification de fréquence et montré des seuils deux fois plus bas pour les musiciens que pour les non-musiciens, confirmant ainsi les travaux de Spiegel & Watson (1984). Plusieurs répétitions de la mesure de seuil ont été effectuées et une amélioration a été constatée entre ces différentes répétitions, pour les groupes de musiciens et de non-musiciens. En outre, il n'y avait pas de différence significative entre les seuils des non-musiciens à la troisième mesure et les seuils des musiciens à la première mesure, ce qui suggère que les gains associés à la pratique musicale peuvent être rapidement compensés par une pratique de la tâche. Cependant, l'amélioration du groupe de musiciens suggère plutôt des mécanismes d'apprentissage distincts entre la pratique musicale et la tâche de mesure de seuil, plutôt de nature procédurale (c'est-à-dire lié à un apprentissage des modalités d'exécution de la tâche). Dans ce cas, les performances équivalentes entre la troisième mesure de seuil des non-musiciens et la première des musiciens aurait pour origine deux sources différentes : la pratique musicale permettant un meilleur seuil de départ pour les musiciens, et la pratique de la tâche pour le groupe de non-musiciens. Une autre possibilité est que les seuils nécessaires et utiles pour la pratique musicale soient plus hauts que ceux requis pour

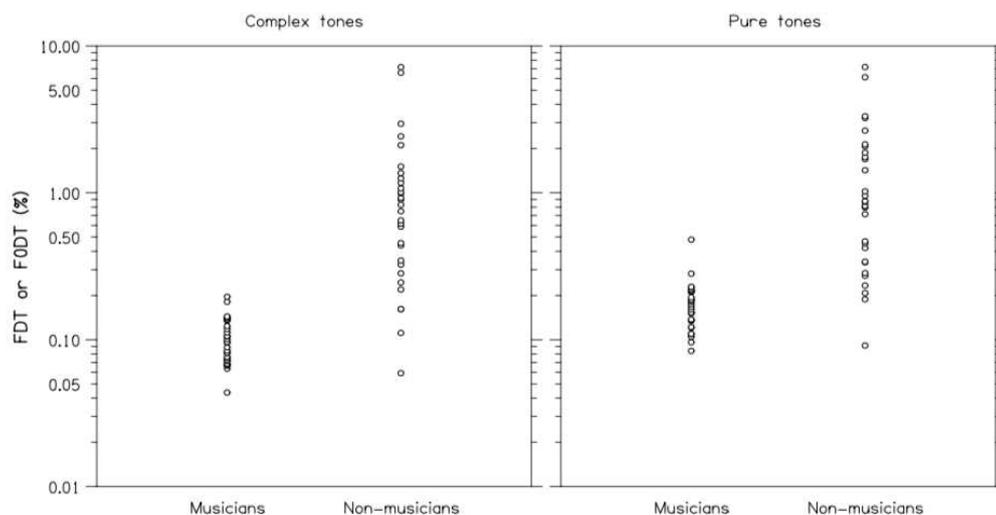


Figure 1.25: Issu de Michey et al. (2006). Comparaison des seuils d’identification de fréquence de musiciens et non-musiciens. pour des sons purs et complexes.

réaliser la tâche d’identification de fréquence, et que donc la pratique musicale n’ait que peu d’effet pour ces tâches au seuil.

Michey et al. (2006) s’étonnent qu’une heure de pratique (total des mesures de seuils de Kishon-Rabin et al., 2001) permette aux non-musiciens d’obtenir des performances similaires aux musiciens qui ont une pratique régulière depuis plusieurs années. Ils émettent l’hypothèse selon laquelle la méthode de séparation des groupes musiciens/non-musiciens n’a pas permis de mettre en avant la différence maximale entre ces groupes. Plus de la moitié des musiciens avaient un parcours moderne ou jazz, et les résultats montrent une différence significative de performance selon le groupe de musiciens (moderne vs classique). Il est également possible que l’expérience principale relativement courte n’ait pas pu mettre en évidence les performances les plus élevées des sujets une fois l’apprentissage procédural terminé pour tous les groupes de participants. Pour tester ces hypothèses, Michey et al. (2006) ont réalisé une étude fondée sur une distinction stricte entre les groupes de musiciens et de non-musiciens (musiciens classique avec plus de 10 ans de pratique) et sur un entraînement assez long sur la tâche pour que les seuils puissent atteindre une valeur asymptotique. Les résultats ont montré des performances initiales environ 6 fois meilleures des musiciens par rapport aux non-musiciens (Figure 1.25), contrairement aux études précédentes où la différence était de facteur compris entre 2 ou 3. En outre, contrairement à Kishon-Rabin et al. (2001) qui montrent qu’une heure de pratique suffit à égaliser les

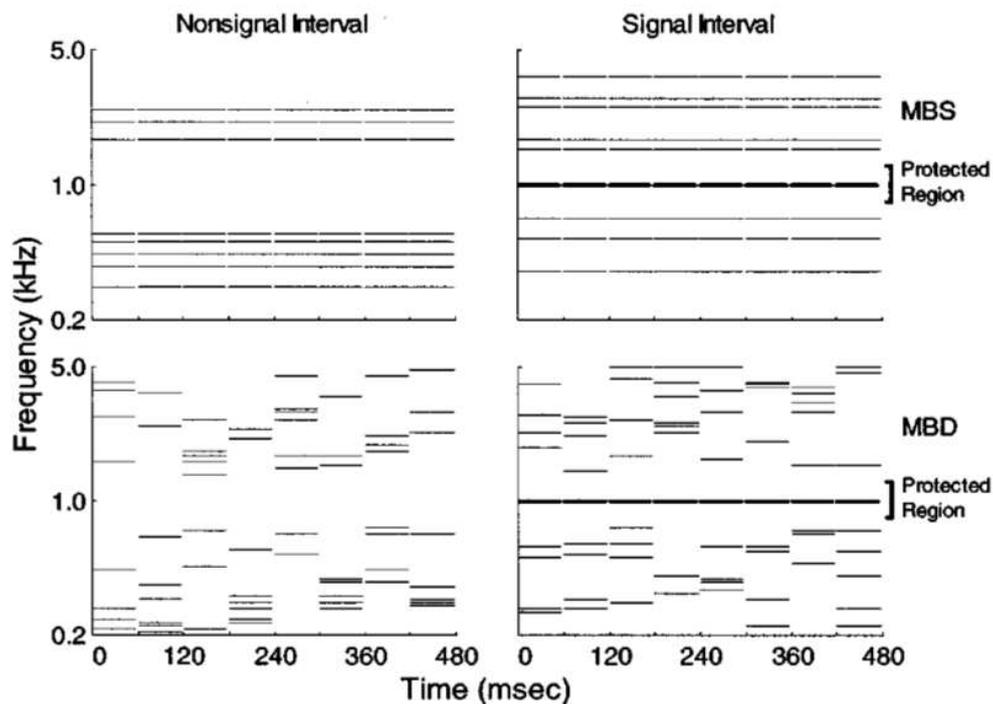


Figure 1.26: Issu de Oxenham et al. (2003). Représentation temps-fréquence des tâches de masquage informationnel. Les participants devaient choisir parmi deux intervalles lequel comportait la cible. On voit également la région protégée en fréquence autour de la fréquence de la cible (1000 Hz) dans l'ensemble des conditions, y compris lorsque la cible n'est pas présente. Des exemples d'essais sans cible (gauche) et avec cible (droite) sont présentés. Dans la condition *Multi-Burst Same* (MBS, haut), les fréquences des composants du masqueur sont identiques au sein de l'essai. Dans la condition *Multi-Burst Different* (MBD, bas), les fréquences des composants du masqueur varient aléatoirement au sein de l'essai.

performances des deux groupes, un entraînement d'une durée de 4 à 8 heures a été nécessaire pour que les seuils des non-musiciens soient similaires à ceux des musiciens. Une explication fournie par les auteurs est l'impact des critères de sélection des participants du groupe de musiciens.

1.4.1.2 Pratique musicale et masquage informationnel

Des gains associés à la pratique musicale ont également été observés en terme de résistance au masquage informationnel (Oxenham et al., 2003). Comme nous l'avons vu précédemment, le masquage informationnel peut être défini comme une

réduction des performances de détection d'une cible en présence d'un ensemble de sons interférents. Contrairement au masquage énergétique où les sons interférents dépendent des mêmes filtres cochléaires que la cible, le masquage informationnel est observé alors même que cible et masqueur sont résolus fréquemment au niveau de la cochlée. L'hypothèse de Oxenham et al. (2003) est que les musiciens ont entraîné cette aptitude car ils sont habitués à isoler une partie de l'environnement musical (par exemple, isoler une mélodie dans une musique polyphonique). Nous avons en effet vu que l'apprentissage perceptif dans les tâches de masquage informationnel semblait provenir d'une amélioration de l'attention sélective (Jones et al., 2014). Oxenham et al. (2003) ont ainsi comparé les capacités de musiciens et de non-musiciens dans une tâche de masquage informationnel. La tâche est issue de Kidd et al. (1994) (Figure 1.26). Les participants devaient choisir parmi deux intervalles lequel contenait la cible. La cible était un son pur répété de manière régulière et de fréquence fixe (1000 Hz). Le masqueur était un complexe multi-tons dont les fréquences des composants étaient choisies aléatoirement à chaque essai. Dans la condition *Multi-Burst Same* (MBS), les fréquences de ces composants étaient identiques au sein de l'essai. Dans la condition *Multi-Burst Different* (MBD), elles variaient au sein de l'essai. Les performances des participants étaient évaluées en changeant le rapport signal sur bruit (niveau de la cible sur niveau du masqueur) de manière adaptative. Les résultats n'ont pas montré de différences entre musiciens et non-musiciens dans la condition MBD. Selon les auteurs, cette condition est associée à un masquage informationnel moins fort que dans la condition MBS car la moins grande similarité entre cible et masqueur permet une ségrégation plus facile. Ils avancent que dans cette condition, le masquage est plutôt de nature énergétique. Cela est discutable du fait de la présence d'une région protégée, mais si l'on considère que le masquage total correspond au masquage énergétique ajouté au masquage informationnel, la proportion de masquage informationnel est effectivement réduite dans la condition MBD. En revanche, une différence forte entre musiciens et non-musiciens a été observée dans la condition MBD, interprété comme montrant que la pratique musicale a permis de développer les capacités d'attention sélective des participants.

1.4.2 Jeux vidéo

Depuis le début des années 2000, l'étude de l'impact de la pratique des jeux vidéo sur la perception visuelle a connu un intérêt croissant, notamment dans le cadre de la réhabilitation clinique (amblyopie par exemple). Il semble en effet que la pratique

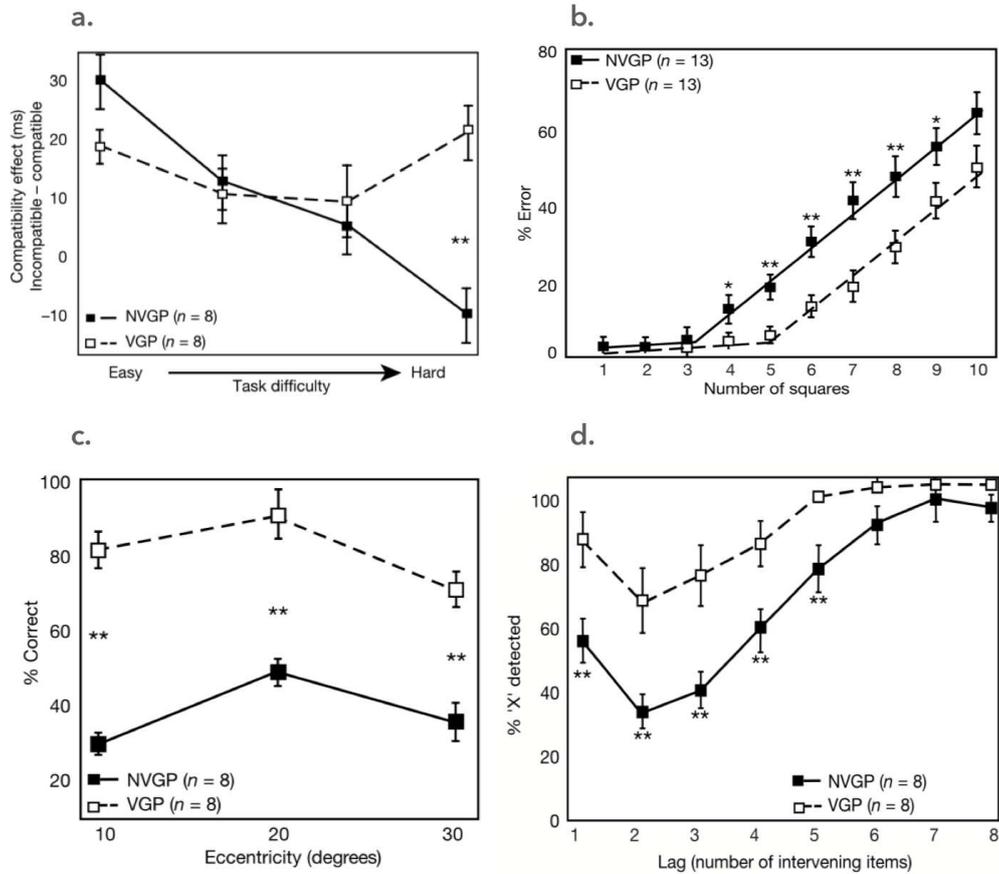


Figure 1.27: Issu de Green & Bavelier (2003). Performances des joueurs (carrés vides) et des non-joueurs (carrés pleins) à la tâche *flanker compatibility* (a.), la tâche d'énumération (b.), la tâche *useful field of view* (c.) et la tâche *attentional blink* (d.).

des jeux vidéo constitue un entraînement perceptif efficace et qui se généralise à des tâches distinctes (pour une revue synthétique, voir Green & Bavelier, 2012).

Une étude classique témoignant de l'effet de la pratique des jeux vidéo sur l'attention visuelle est celle de Green & Bavelier (2003). L'hypothèse est que les tâches variées inhérentes à la pratique des jeux vidéo d'action sont associées à une amélioration des capacités d'attention visuelle. Les auteurs ont comparé un groupe de joueurs et un groupe de non-joueurs, et ont montré que les joueurs ont de meilleures performances que les non-joueurs sur plusieurs tâches attentionnelles. Dans la première tâche (*flanker compatibility task*), les participants devaient détecter une cible (carré ou losange) en présence d'un distracteur compatible (identique à la forme à détecter) ou incompatible (différent de la forme à détecter). Le principe est que l'effet du type

de distracteur est prédit comme fort lorsque les ressources attentionnelles du participant sont grandes. La logique est que la tâche avec distracteur incompatible nécessite peu de ressources attentionnelles, alors que la tâche avec distracteur compatible en demande plus. Si l'ensemble des ressources du participant est dévoué à la tâche principale de détection, alors l'effet de compatibilité (différence de temps de réaction entre conditions de masqueur) sera faible. En revanche, s'il reste des ressources disponibles, elles pourront être mobilisées pour le distracteur incompatible. Qui plus est, des versions faciles et difficiles de la tâche principale (selon le nombre de distracteurs) ont été testées. La taille de l'effet de compatibilité donne ainsi une mesure des ressources attentionnelles disponibles. Les résultats montrent que les joueurs ont un fort effet de compatibilité dans la condition difficile contrairement aux non-joueurs, suggérant une plus grande quantité de ressources attentionnelles disponibles (Figure 1.27, a.). Cet avantage a été confirmé dans une tâche d'énumération dans laquelle les participants devaient reporter le nombre d'éléments présentés simultanément pendant un court instant. Lorsque peu d'items sont présentés, ils peuvent être comptés par le biais d'un processus automatique, précis, rapide et le temps de réponse ne dépend pas du nombre d'éléments. A partir d'un certain nombre d'éléments, un processus sériel et plus lent doit être mis en œuvre. Le nombre maximal d'éléments quantifiés par le premier processus donne une estimation du nombre d'éléments qui peuvent faire l'objet de l'attention en même temps. Les résultats montrent que ce nombre est plus élevé chez les joueurs que chez les non-joueurs (Figure 1.27, b.). Dans une troisième tâche, les auteurs ont évalué la répartition de l'attention spatiale des participants dans une tâche de détection de cible parmi des distracteurs (*useful field of view*). De manière à tester si les performances des joueurs étaient également meilleures à des positions spatiales non-entraînées (excentricité plus grande), trois positions ont été comparées (Figure 1.27, c., 30° correspondant à la position non entraînée). Les résultats montrent un avantage des joueurs pour l'ensemble des positions spatiales. Enfin, une tâche de clignement attentionnel (*attentional blink*) a été utilisée. Ce phénomène correspond à une baisse de l'attention quelques centaines de millisecondes après la détection d'une cible. Dans cette version de la tâche, des lettres étaient présentées successivement aux participants. Lorsqu'une cible apparaissait (lettre blanche), ils devaient détecter la présence d'un 'X' à la place de l'une des 8 lettres suivant la cible. Cette tâche d'identification/détection a été utilisée pour son aspect amodal. Cela permettait de vérifier le transfert du gain attentionnel des joueurs à une tâche qui n'est pas spécifiquement visuelle (bien que présentée dans la modalité visuelle). Les résultats montrent à nouveau un avantage du groupe de joueurs (Figure 1.27, d.). Enfin, de manière à vérifier que l'avantage des joueurs n'était pas lié à

des différences intrinsèques entre les deux populations, les auteurs ont réalisé une seconde expérience dans laquelle des participants ont été entraînés avec un jeu vidéo d'action, comme dans la première expérience, ou un jeu vidéo classique (Tetris). Les résultats montrent à nouveau un avantage pour le groupe entraîné avec le jeu d'action. Cet effet de la pratique de jeux vidéo qui semble pouvoir se généraliser à d'autres types de tâches et de stimuli suggère un apprentissage global, par exemple lié à la capacité à apprendre de nouvelles tâches (Green & Bavelier, 2012).

D'autres études ont montré un effet de la pratique des jeux vidéo sur la sensibilité au contraste (Li et al., 2009). Les auteurs ont comparé des joueurs et des non-joueurs puis ont également réalisé un entraînement sur des jeux vidéo d'actions (un groupe contrôle étant entraîné avec un jeu vidéo qui n'implique pas les processus des jeux vidéo d'action). Les résultats montrent une plus grande sensibilité au contraste pour les joueurs, et une amélioration de cette sensibilité chez des participants entraînés avec les jeux vidéo d'action en comparaison du groupe contrôle.

D'autres aptitudes, telles que l'intégration multi-sensorielle (stimuli auditif et visuel synchronisés ou légèrement décalés : Donohue et al., 2010) ou le contrôle cognitif (multi-tâches : Anguera et al., 2013) sont des aptitudes qui semblent pouvoir être améliorées par la pratique de jeux vidéo. L'impact des jeux vidéo sur des troubles perceptifs tels que l'amblyopie (différence d'acuité visuelle entre les deux yeux sans cause périphérique) a aussi été rapportée (Li et al., 2011).

Deveau et al. (2014) ont testé le transfert des bénéfices lié à l'apprentissage perceptif visuel à une tâche perceptive plus écologique : la pratique du *baseball*. Différentes caractéristiques issues des jeux vidéo, comme l'utilisation de stimuli variés, l'optimisation de la présentation du stimulus, l'utilisation de plusieurs modalités sensorielles, le renforcement positif lié à l'exécution de la tâche ont été implémentées dans le protocole d'apprentissage perceptif. En plus d'une meilleure acuité visuelle et sensibilité au contraste, les joueurs de *baseball* ayant réalisé l'entraînement ont témoigné de meilleures performances sportives en comparaison d'un groupe contrôle n'ayant pas été entraîné.

Au niveau expérimental, on peut noter l'émergence de critiques associées au fait de ne pas avoir un recrutement aveugle des participants (Boot et al., 2011). Un biais potentiel est que les participants hautement spécialisés comme les musiciens ou les joueurs de jeux vidéo savent qu'ils sont recrutés pour cette caractéristique et sont plus motivés à réaliser la tâche demandée. Cependant, certaines études utilisent un recrutement aveugle (par exemple, classification des participants après l'étude : Dye et al., 2009; Donohue et al., 2010). Green & Bavelier (2012) notent cependant la

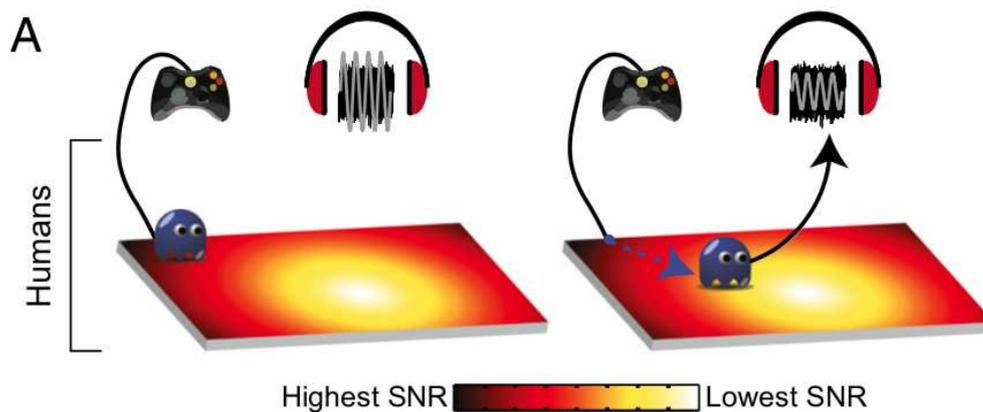


Figure 1.28: Issu de Whitton et al. (2014). Protocole d'entraînement perceptif. Les participants contrôlaient les déplacements d'un avatar sur une surface en deux dimensions. La tâche était de trouver la position associée au plus faible rapport signal sur bruit (ton pur de 500 Hz dans un bruit blanc large bande).

complexité d'un recrutement aveugle dans le cas de populations dont la distribution est faible. Combien de participants devrait-on recruter avant de pouvoir constituer un groupe de musiciens professionnels si cela ne fait pas partie des critères d'inclusion ?

1.4.3 Protocoles immersifs et jeux audio

Qu'il s'agisse de la pratique musicale ou des jeux vidéo, la diversité des situations rencontrées, la nouveauté perpétuelle des actions à produire et la complexité de ces actions en font un apprentissage qui ne peut jamais être complètement automatisé (Bavelier et al., 2018). De plus, le participant est actif et la boucle sensori-motrice recrutée en permanence. Il semble donc attrayant d'utiliser certains critères issus de la pratique musicale ou des jeux vidéo de manière à améliorer les protocoles d'apprentissage perceptif existant.

Whitton et al. (2014) ont mis en place un protocole d'entraînement perceptif audio-moteur à boucle fermée qui comprenait un certain nombre d'aspects présents dans la pratique musicale ou des jeux vidéo. Les participants contrôlaient un avatar sur une surface en deux dimensions dont la position était associée à une stimulation auditive spécifique (Figure 1.28), constituée d'une cible (ton pur à 500 Hz) et de bruit masquant (bruit blanc large bande). Alors que le niveau du bruit était fixe, le niveau de la cible était modifié en temps réel en fonction de la position de l'avatar. Le but du participant était de trouver la position associée au rapport signal sur bruit

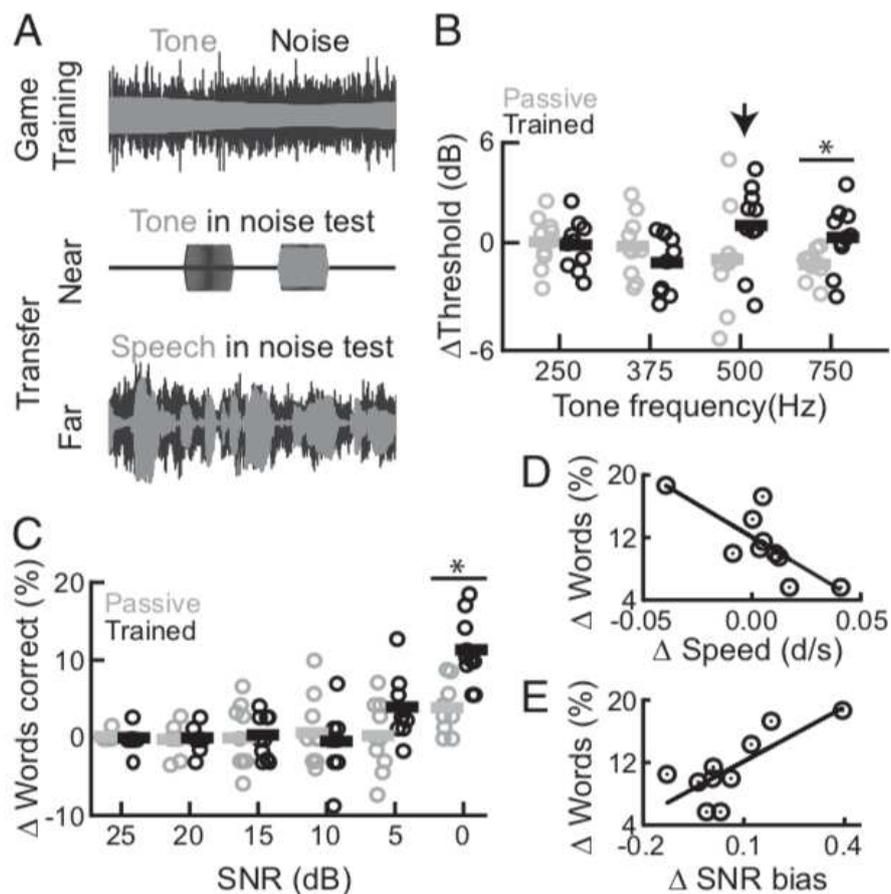


Figure 1.29: Issu de Whitton et al. (2014). A. Illustration des tâches à transfert proche (milieu) et éloigné (bas). B. Résultats de la tâche de détection de sons purs dans le bruit. La différence de seuil (post - pré) est illustrée en fonction de la fréquence du son pur à détecter pour les groupes contrôle (gris) et test (noir). C. Résultats de la tâche de perception de la parole dans le bruit. La différence de mots correctement rapportés (post - pré) est illustrée en fonction du rapport signal sur bruit (SNR) pour les groupes contrôle (gris) et test (noir).

le plus faible (en blanc sur la Figure 1.28). Cette position changeait à chaque essai et le rapport signal sur bruit était le seul indice qui pouvait être utilisé pour réaliser la tâche.

Les participants ont été entraînés à faire cette tâche pendant 30 minutes par jour pendant un mois. Un groupe contrôle a été inclus dans lequel les participants devaient écouter les sons produits par des participants du groupe test. Il s'agissait donc d'une exposition passive et les participants pouvaient réaliser des tâches silencieuses pendant cette exposition.

Des tests perceptifs ont été réalisés avant et après l'entraînement ou l'exposition passive. Une tâche de détection de sons purs dans le bruit a tout d'abord été réalisée (tâche à transfert proche : Figure 1.29 A, milieu). Parmi deux intervalles contenant un bruit blanc mais dont un seul contenait un son pur, les participants devaient indiquer celui qui contenait le son pur. Les seuils de détection de sons purs à 250, 375, 500 et 750 Hz ont été ainsi mesurés. Les résultats ne montrent pas d'amélioration des performances entre les pré- et post-tests pour les groupes test et contrôle. Dans une seconde tâche de perception de la parole dans le bruit (tâche à transfert éloigné : Figure 1.29 A, bas) a été réalisée dans laquelle les participants devaient écouter des phrases en présence d'un bruit de parole (*4-talker babble*) et devaient reporter oralement la phrase entendue. Une amélioration du groupe test par rapport au groupe contrôle a été observée dans la condition la plus difficile de rapport signal sur bruit (0 dB, Figure 1.29 C). Les auteurs proposent que cette généralisation de l'entraînement perceptif est issue de la nature dynamique de l'entraînement. Le protocole se rapproche effectivement des paradigmes où il y a *roving* dans le sens où il n'y a pas de stimulus standard qui peut servir de référence. Une autre possibilité est que l'aspect sensori-moteur et immersif de la tâche d'entraînement est à la source de cette généralisation.

Plus récemment, Whitton et al. (2017) ont mis en place une étude avec groupe placebo menée en double aveugle inspirée par ces résultats. Un groupe test a réalisé une tâche d'entraînement fondée sur la boucle audio-motrice, comme cela était le cas de leur précédente étude (Whitton et al., 2014). Comme représenté sur la Figure 1.30 C, les participants devaient discriminer de faibles changements d'un paramètre du son (niveau, fréquence, fréquence de modulation d'un son répété ou d'un bruit modulé en amplitude) en présence d'un son de parole masquant (*speech babble*). Plus les performances des participants s'amélioraient, plus le niveau des sons masquants était élevé. Les changements du paramètre du son étaient entendus de manière continue et mis à jour en temps réel. Ils dépendaient de la position du doigt ou du stylet sur l'écran

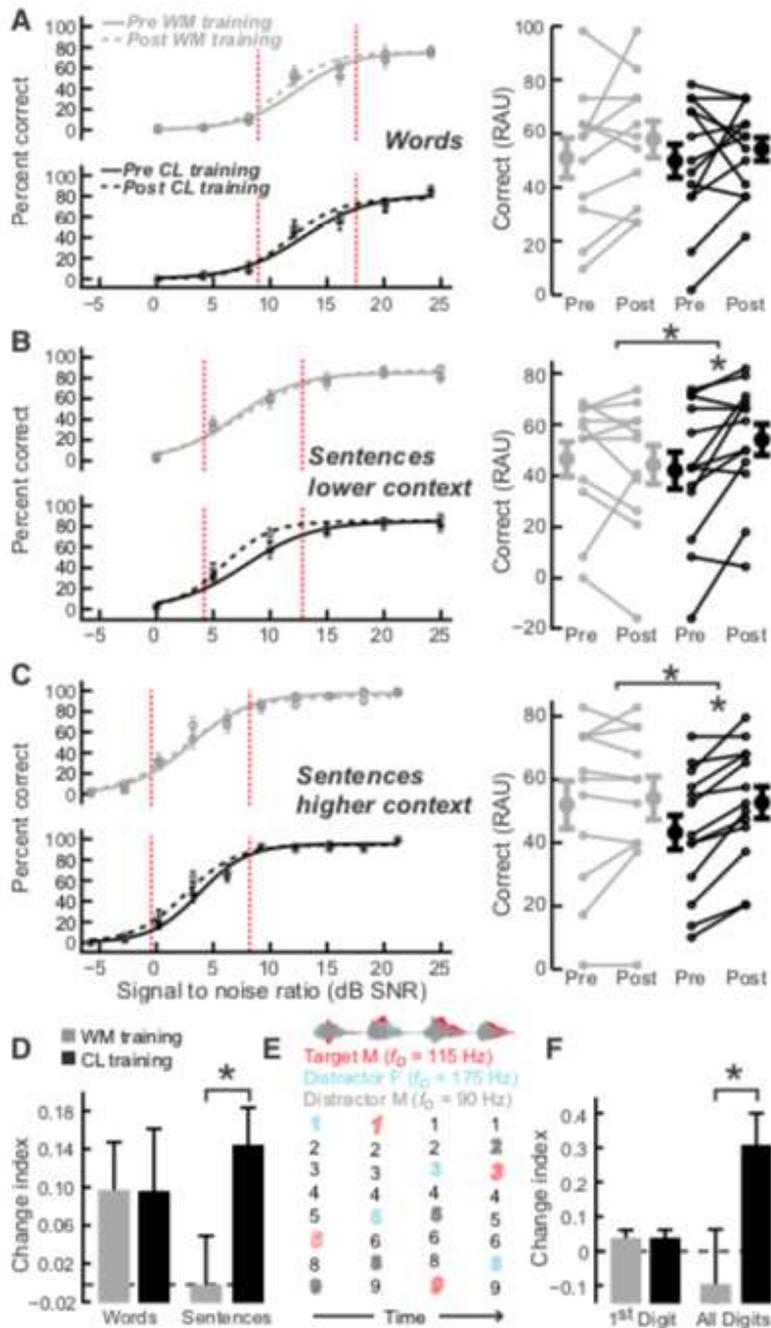


Figure 1.31: Issu de Whitton et al. (2017). Résultats des tâches de perception de la parole dans le bruit. A. WIN (*Words In Noise*). B. QuickSIN (*Quick Speech In Noise*). C. BKBSIN (*BKB Speech In Noise*). Le groupe mémoire de travail est représenté en gris et le groupe audio-moteur en noir.

et le but était de dessiner le contour d'une pièce de puzzle grâce à ces informations. Le groupe placebo a réalisé une tâche de mémoire de travail auditive (Figure 1.30, B). Des personnes âgées et malentendantes ont été aléatoirement réparties et entraînées environ 3.5 heures par semaines pendant 8 semaines. Des tests de perception de la parole dans le bruit (WIN, QuickSIN et BKBSIN; Wilson et al., 2007) ont été réalisés avant et après l'entraînement. Les résultats montrent une amélioration des performances du groupe audio-moteur en comparaison du groupe mémoire de travail pour la perception de phrases dans le bruit (QuickSIN et BKBSIN, Figure 1.31, B et C) mais pas des mots dans le bruit (Figure 1.31, D) Les analyses se sont focalisées sur les SNR difficiles mais pas impossibles (compris entre les lignes rouges sur la Figure 1.31 A, B et C), et les gains étaient d'environ 25% de mots reconnus en plus. Enfin, une tâche d'identification de chiffres a été réalisée dans laquelle les participants devaient suivre la voix d'un des locuteurs et ignorer celles des autres. Le but était de relever les chiffres prononcés par ce locuteur (Figure 1.31 E). Il s'agit donc d'une tâche de perception de la parole sans indice sémantique. De manière à comparer cette tâche aux tâches de perception de la parole dans le bruit qui ont été utilisées, les réponses ont été analysées en fonction de l'identification correcte du premier chiffre de la séquence (correspondant à la tâche WIN de perception de mots dans le bruit) et en fonction de l'identification correcte de l'ensemble des chiffres (correspondant aux tâches d'identification de phrases dans le bruit, QuickSIN et BKBSIN). Les résultats montrent une amélioration des participants du groupe audio-moteur en comparaison du groupe placebo uniquement lorsque l'ensemble des chiffres sont pris en compte (Figure 1.31 F). Par ailleurs, d'autres tâches ont été réalisées avant et après l'entraînement ou le placebo. Il s'agissait de versions auditive et visuelle du test de Stroop, d'une tâche de mémoire de travail (séquence lettres-chiffres) et d'une tâche d'identification de modulations de fréquence. Aucune différence n'a été observée sur ces tests. Pour résumer, les résultats montrent un transfert de l'entraînement perceptif à des tâches haut-niveau telles que la perception des phrases dans le bruit. Par ailleurs, des tâches plus bas-niveau telle que la perception de mots dans le bruit ou la détection du premier chiffre d'une séquence n'ont pas montré d'amélioration.

1.4.3.1 Conclusions

Qu'il s'agisse de la pratique musicale ou de la pratique des jeux vidéo, l'aspect immersif de l'expérience est en partie lié au fait que la stimulation sensorielle n'est pas passive mais s'intègre dans une boucle motrice active. La mise en place de la boucle audio-motrice dans le cadre de l'apprentissage perceptif est une des motivations des recherches expérimentales qui ont été menées dans cette thèse. Notamment, la

question de son rôle sur l'apprentissage perceptif de la hauteur et dans des tâches de masquage informationnel seront considérées dans les chapitres 2 et 3.

2 CHAPITRE 2. Perception de la hauteur et boucle audio-motrice

Une partie de ce travail a été présentée en tant que :

Jean, H., Demany, L. & Pressnitzer, P. (2016). Auditory Pitch Perception and the Audio-Motor Loop. 39th MidWinter Meeting of the Association for Research in Otolaryngology, San Diego, USA (Poster).

2.1 Aperçu

La hauteur musicale est un attribut perceptif jouant un rôle majeur dans la musique ou le langage. Même si les capacités de perception de la hauteur semblent pouvoir être améliorées par des protocoles d'entraînement spécifiques, les conditions permettant l'efficacité maximale d'un tel entraînement restent méconnues. D'autre part, des tâches écologiques comme la pratique musicale sont associées à une amélioration à long terme de la perception de la hauteur. Nous testons ici l'efficacité d'un entraînement perceptif rapide (approximativement 30 minutes) élaboré de manière à mettre en action la boucle audio-motrice. Dans cette tâche audio-motrice sur tablette, les participants devaient contrôler la fréquence d'un son en temps réel. Des mesures de seuils de discrimination et d'identification de fréquence ont été réalisées avant et après l'entraînement. L'ensemble des tâches a été réalisée avec (expérience 1) et sans (expérience 2) *roving* de la fréquence. En outre, un sous-groupe de participants a réalisé une série de tests/entraînement plus long, approximativement 5h, sur plusieurs jours (expérience 3). Les résultats confirment un effet délétère du *roving* sur les performances ainsi qu'un avantage de la condition discrimination par rapport à la condition identification. Cependant, aucune amélioration significative des seuils de discrimination ou d'identification n'a été observée, même avec l'entraînement prolongé.

2.2 Introduction

La perception de hauteur est cruciale dans de nombreuses situations d'écoute. Elle est nécessaire pour percevoir une mélodie ou les nuances des émotions de la voix. Pourtant, elle est mise en difficulté et témoigne d'une grande variabilité chez les malentendants et les utilisateurs d'aides auditives ou d'implants cochléaires (Moore & Carlyon, 2005). Les possibilités de rééducation constituent donc un enjeu majeur.

La capacité à dire que deux fréquences proches sont différentes (discrimination) et la capacité à déterminer la direction de cette différence (identification) constituent deux aptitudes qui semblent au moins partiellement distinctes. Johnsrude et al. (2000) et Tramo et al. (2002) ont en effet montré que des patients possédant des lésions cérébrales présentaient un déficit à percevoir la direction d'un changement qui a lui-même été détecté. De façon complémentaire, Demany & Ramos (2005) ont montré qu'il était possible d'identifier la direction d'un changement de hauteur alors que les fréquences responsables du changement n'étaient pas détectables. Il est donc important de distinguer les tâches qui évaluent la discrimination d'un changement de fréquence (désignées tâches de discrimination ci-après) et celles qui évaluent les capacités d'identification de la direction du changement de fréquence (désignées tâche d'identification ci-après).

S'appuyant sur les résultats neuropsychologiques de Johnsrude et al. (2000) et Tramo et al. (2002), Semal & Demany (2006) ont testé les capacités de discrimination et d'identification de participants normo-entendants. Ils ont mis en place une tâche dans laquelle deux paires de sons purs étaient jouées à chaque essai. Une des paires contenait des sons purs de mêmes fréquences et l'autre paire des sons purs de fréquences différentes. La tâche était de déterminer la paire dans laquelle les sons purs étaient différents (discrimination) ou bien la direction du changement dans la paire contenant les sons purs différents (identification). Les résultats montrent clairement que certains participants ne peuvent pas déterminer la direction d'un changement de fréquence pourtant aisément détecté, suggérant à nouveau une dissociation entre capacités de discrimination et d'identification de fréquence mais cette fois chez des sujets sains normo-entendants.

Une caractéristique importante de cette étude de Semal & Demany (2006) est la présence de *roving* de la fréquence, c'est-à-dire que la fréquence du son pur standard était choisie aléatoirement à chaque essai au lieu d'être la même dans tous les essais. En utilisant le même protocole, Mathias et al. (2010) ont montré que la différence discrimination/identification était largement influencée par la présence du

roving et que cet effet est réduit lorsque le son pur standard est fixe. Étayant les travaux de Jesteadt & Bilger (1974), les résultats suggèrent donc que la différence discrimination/identification n'est fortement présente que lorsque le *roving* est assez important.

Comme nous l'avons vu dans la partie 1.2, les études sur l'apprentissage perceptif et la perception de la hauteur musicale ont montré qu'une diminution des seuils de discrimination (Delhommeau et al., 2005; Demany, 1985; Demany & Semal, 2002; Irvine et al., 2000) et d'identification (Michey et al., 2006) de fréquences peut être observée avec la pratique, et que cet apprentissage peut être rapide. Ces études ne présentent pas de *roving* de la fréquence, c'est-à-dire que la fréquence du son pur standard est la même d'un essai à l'autre. Or, dans les études citées plus haut qui présentent un *roving* de la fréquence, les auteurs ne constatent pas d'apprentissage perceptif dans les tâches de discrimination et d'identification (Semal & Demany (2006) : plus de 2000 essais; Mathias et al. (2010) : environ 3 heures de pratique; avec procédure adaptative et *feedback* dans les deux cas). On peut également noter que la différence de performance entre les tâches de discrimination et d'identification n'est pas réduite après la pratique. Certains travaux présentés dans la partie théorique ont en effet montré un effet délétère du *roving* sur les seuils de discrimination et d'identification de fréquence (Amitay et al., 2005; Demany & Semal, 2005; Harris, 1952; Jesteadt & Bilger, 1974).

Nous avons également vu (1.4) que des travaux suggèrent l'efficacité de tâches écologiques comme les jeux vidéos d'action (Green & Bavelier, 2003) ou la pratique musicale (Kishon-Rabin et al., 2001; Michey et al., 2006; Swaminathan et al., 2015) sur l'amélioration de performances perceptives. On peut en outre noter la présence dans ces tâches d'une boucle fermée entre perception sensorielle et action motrice (Whitton et al., 2017).

Le but de cette série d'expériences est de mettre en place un protocole d'apprentissage perceptif rapide fondé sur la boucle audio-motrice. Nous avons créé une tâche d'entraînement dans laquelle les participants devaient contrôler la hauteur d'un son pur continu en déplaçant leur doigt sur une tablette, et trouver la position correspondant à la fréquence la plus haute. Cette position changeait d'un essai à l'autre et les participants recevaient un score dépendant de leur précision et de leur rapidité. Pour évaluer d'éventuels bénéfices de l'entraînement, des mesures des seuils de discrimination et d'identification de fréquence (tâche à double paires issue de Semal & Demany (2006)) ont été administrées aux participants avant (pré-test) et après (post-test) l'entraînement. Dans l'expérience 1, la fréquence standard des

pré-/post-test et la fréquence cible de la tâche audio-motrice (la fréquence la plus aiguë à retrouver) étaient différentes à chaque essai (*roving*). Dans l'expérience 2, l'ensemble des stimuli (pré-/post-tests et tâche d'entraînement audio-motrice) ne comportait pas de *roving* de fréquence. La fréquence du son pur standard dans les pré-/post-tests, ainsi que la fréquence cible dans la tâche audio-motrice étaient donc identiques d'un essai à l'autre. Enfin, dans l'expérience 3, un entraînement plus long a été réalisé sur 5 jours avec des paramètres identiques à ceux de l'expérience 1 (*roving*) de manière à étudier les effets d'apprentissage avec un entraînement plus conséquent.

2.3 Expérience 1

2.3.1 Méthodes

2.3.1.1 Participants

Quatorze participants (9 femmes ; age $M = 25.29$, $SD = 4.79$) normo-entendant ont participé à cette étude. Ils étaient indemnisés 10 euros par heure. Un audiogramme préalable permettait de s'assurer que tous les participants avaient une audition normale, c'est à dire des seuils auditifs de 20 dB HL ou moins pour les fréquences testées (125, 250, 500, 1000, 2000, 3000, 4000, et 8000 Hz). Tous les participants étaient non-musiciens (moins de deux ans de pratique musicale au moment de l'étude).

2.3.1.2 Stimuli

Pré et post-test. Discrimination et identification de fréquences

Tous les sons utilisés étaient des sons purs de 250 ms (incluant des rampes ascendantes et descendantes de 10 ms chacune). Les deux sons purs de chaque paire étaient séparés par un silence de 250 ms et les deux paires par un silence de 750 ms. Le son pur test (dans la paire contenant des sons purs différents) était soit plus aigu soit plus grave que le son pur standard (aléatoirement avec une probabilité de 50%). Le premier son pur de chaque paire était choisi aléatoirement à chaque essai entre 400 et 2400 Hz dans une distribution uniforme. La tâche sera décrite en détail ci-dessous, dans la section 2.3.1.4.

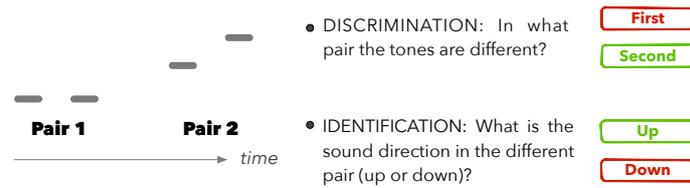


Figure 2.2: Procédure. Paradigme de discrimination et d’identification de fréquence à double paires. La même tâche est utilisée pour les conditions discrimination et identification mais la consigne est différente. Dans la condition discrimination, les participants doivent dire dans quelle paire les sons purs sont différents. Dans la condition identification, ils doivent donner la direction du changement de fréquence dans la paire contenant des sons purs différents.

Une des paires présentait des sons purs de même fréquence, et l’autre des sons purs de fréquences différentes. Dans la condition discrimination, la tâche était de répondre dans quelle paire les sons purs étaient différents. Cette tâche ne nécessitait donc pas d’identifier la direction du changement de fréquence. Au contraire, dans la condition identification, la tâche était de donner la direction du changement de fréquence dans la paire présentant les sons purs différents. Un retour sur leurs performances (*feedback*) était donné aux participants après chaque essai. Une procédure adaptative a été utilisée dans laquelle la différence de fréquence (Δf) était augmentée à chaque réponse incorrecte et diminuée à chaque réponse correcte. Jusqu’au quatrième *reversal*, le Δf était multiplié par 2.25 ou divisé par $\sqrt{2.25}$. Après le quatrième *reversal*, le facteur multiplicateur était de 1.5 au lieu de 2.25. Cela conduit à une estimation de 75% de réponses correctes sur la courbe psychométrique (Kaernbach, 1991). La tâche se terminait une fois que 14 changements de direction (*reversals*) avaient été effectués. Le seuil était calculé à partir de la moyenne des 10 derniers *reversals*.

Tâche audio-motrice

Les participants contrôlaient la fréquence d’un son pur en déplaçant leur doigt sur une tablette. Chaque position était associée à une fréquence du son pur contrôlé. Les fréquences étaient réparties sur l’axe vertical et un déplacement horizontal ne changeait pas la fréquence du son pur contrôlé. Le but de la tâche était de trouver la position spatiale associée à la fréquence la plus élevée (Figure 2.3). Cette position était choisie aléatoirement à chaque essai. Aucun indice visuel n’était donné pour réaliser la tâche et les participants devaient donc écouter les changements de fréquence du son pur contrôlé en temps réel. Ils étaient encouragés à répondre le plus rapidement



Figure 2.3: Procédure. Interface de la tâche audio-motrice avec présentation du *feedback*. La ligne centrale correspond aux positions associées au son pur le plus aigu. Le dégradé de couleur illustre le changement de fréquence associé à la position (hautes fréquences en vert). Le score total, le score de précision et le coefficient multiplicateur calculé à partir de la durée de l'essai (cf. méthodes) sont aussi indiqués.

et le plus précisément possible et il leur était indiqué que le nombre de points obtenus dans l'essai dépendait de ces deux critères.

Le but des participants était de récolter le plus de points possible sur l'ensemble des essais de la phase d'entraînement (30 minutes). Le nombre de points remportés à chaque essai dépendait de la précision (rapport entre la fréquence associée à la position choisie et la fréquence cible) et de la rapidité, de manière à favoriser une stratégie où la précision est recherchée en premier lieu mais où la rapidité est également valorisée. La formule utilisée pour calculer le nombre de points gagnés dans un essai était la suivante :

$$\left(\frac{1}{\left(\frac{duration}{100} + 1 \right)^2} + 1 \right) \times 600 \times \left(\frac{accuracy}{100} \right)^6$$

La durée (*duration*) correspond à la durée totale de l'essai (en secondes) et la précision (*accuracy*) correspond à la proximité entre la fréquence cible et la fréquence choisie par le participant (en pourcentage). La Figure 2.4 illustre l'équation de calcul du score. La durée de l'essai donne lieu à un facteur multiplicateur compris entre 1 pour un essai très long et 2 pour un essai très court. Le score de précision est

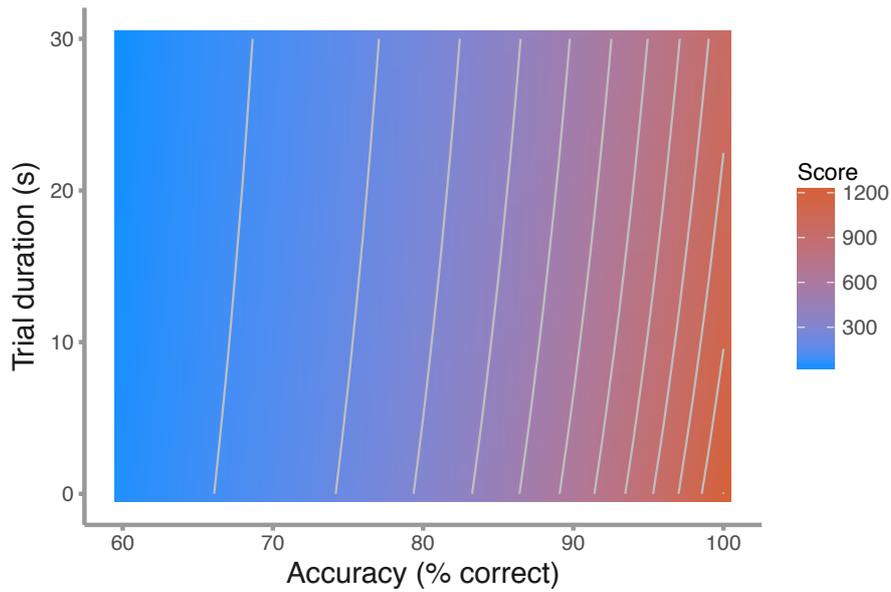


Figure 2.4: Procédure. Représentation du score en fonction de la précision et de la durée de l’essai. Le score augmente quand la durée diminue et la précision augmente.

compris entre 0 et 600 points. La fonction puissance permet une diminution rapide du nombre de points gagnés avec une moins bonne précision de manière à récompenser les essais très précis. Notons que cette formule a été choisie de façon heuristique, sans que nous ayons d’éléments expérimentaux permettant de motiver les choix réalisés. Cette limitation est commune à la plupart des tâches d’entraînement visant à se rapprocher de situations écologiques comme le jeu vidéo (Whitton et al., 2017).

A la fin de chaque essai, les participants recevaient un *feedback* visuel (Figure 2.3 : ligne représentant la position sur l’axe y correspondant à la fréquence la plus aiguë et un dégradé de couleur correspondant à l’éloignement progressif en fréquence) et auditif (deux sons purs à la fréquence choisie par le participant puis un son pur à la fréquence cible étaient entendus après que la réponse ait été donnée).

2.3.1.5 Analyse des données

Dans l’ensemble des analyses, les Δf en cents ont été convertis en pourcentages $(2^{\frac{cent}{1200}} - 1) \times 100$.

Discrimination et identification de fréquence

L'analyse des performances des tâches de discrimination et d'identification de fréquence a été réalisée en comparant les seuils obtenus aux pré- et post-tests. Trois effets d'apprentissage potentiels ont été évalués :

- L'entraînement auditif en comparant (PRE1+PRE2) avec (POST1+POST2).
- L'aspect procédural de l'apprentissage en comparant PRE1 avec PRE2, dans le sens où les participants ont été familiarisé avec la procédure de la tâche dans PRE1.
- L'habituation à la tâche en comparant (PRE1+POST1) avec (PRE2+POST2). Cette comparaison permet d'observer l'amélioration au sein des phases PRE ou POST qui n'est pas dépendante de la phase d'entraînement.

Des test statistiques (Wilcoxon signed ranks tests) ont été réalisés pour évaluer la significativité des résultats. Les valeurs de p ont été corrigées de manière adéquate en fonction du nombre de comparaisons réalisées (correction de Bonferroni). Comme dans certains travaux (Delhommeau et al., 2005, 2002; Irvine et al., 2000; Micheyl et al., 2006) une transformation log a été appliquée avant de réaliser les analyses statistiques, aidant à corriger le fait que la variabilité des seuils fréquentiels augmente avec leur magnitude.

Les graphiques représentent donc les valeurs de seuils d'identification et de discrimination de fréquence en pourcentage et sur une échelle logarithmique.

Tâche audio-motrice

L'intégralité des coordonnées x et y correspondantes à la position du doigt sur la tablette et leur décours temporels a été enregistrée pour chaque essai. Les analyses ont été réalisées en utilisant ces données ainsi que la précision (en %) et la durée de chaque essai (en s).

Pour évaluer les performances au fur et à mesure de la phase d'entraînement et pour utiliser les données de l'ensemble des participants, les essais ont été séparés en tranches de 10% du nombre total d'essais réalisé. Les analyses comparent alors chaque tranche de 10% sachant que cela ne correspond pas nécessairement au même nombre d'essais pour chaque participant.

Des régressions linéaires ont été réalisées pour analyser la relation entre la durée de l'entraînement et les performances.

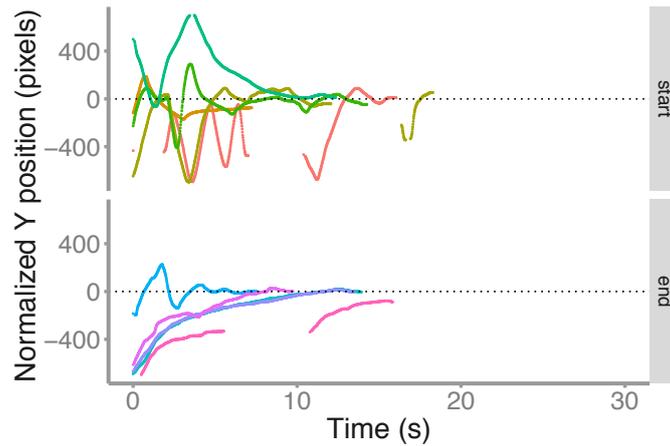


Figure 2.5: Expérience 1 - Tâche audio-motrice. Exemple des 5 premiers (haut) et des 5 derniers essais (bas) de la phase d’entraînement d’un participant. Les essais sont normalisés de manière à avoir la position correspondante à la fréquence la plus aiguë égale à 0.

2.3.2 Résultats

2.3.2.1 Tâche audio-motrice

Les Figures 2.5 et 2.6 représentent des exemples de parcours réalisés par un participant pour des essais en début et en fin d’entraînement. Sur la Figure 2.5, les coordonnées y ont été normalisées de manière à ce que la position correspondante à la fréquence la plus élevée (différente d’un essai à l’autre) soit égale à 0 ($y_{norm} = y - t_{pos}$, avec t_{pos} la position du son le plus aigu). Cette position normalisée est représentée en fonction du temps. Sur la Figure 2.6, les coordonnées x et y sont représentées et le temps est figuré par la couleur. On voit que les premiers essais sont associés à un parcours sinueux et que le son pur cible est trouvé plus rapidement et avec moins d’hésitation dans un essai en fin d’entraînement.

La Figure 2.7 montre l’évolution de la performance (% d’erreur) à la tâche audio-motrice au cours de la pratique. Les seuils de discrimination et d’identification (moyennés entre pré- et post-tests) sont également représentés sur la Figure. On voit que les performances sont globalement meilleures à la tâche audio-motrice par rapport aux tâches d’identification et de discrimination de fréquence.

Pour mettre en avant d’éventuelles différences inter-individuelles, la Figure 2.8 montre l’évolution des performances au fur et à mesure de l’entraînement pour chaque participant ainsi que leurs seuils de discrimination et d’identification (moyennes

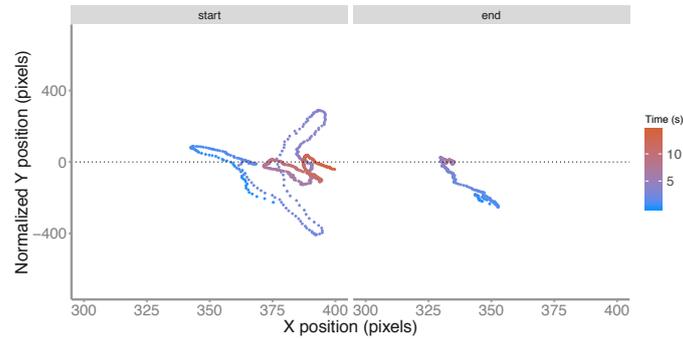


Figure 2.6: Expérience 1 - Tâche audio-motrice. Exemple de parcours (coordonnées x et y) réalisés par un participant dans un essai au début (gauche) et à la fin (droite) de la phase d'entraînement. Les coordonnées y ont été normalisées de manière à avoir la position correspondante au son pur le plus aigu centrée en 0.

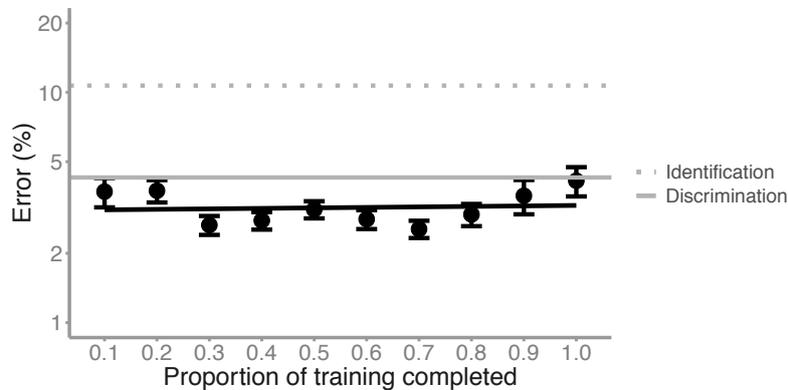


Figure 2.7: Expérience 1 - Tâche audio-motrice. Évolution des performances (% d'erreur) au fur et à mesure de l'entraînement, en moyenne pour tous les participants ($N=14$). Chaque point représente la moyenne des performances des participants pour les 10% d'essais affichés sur l'axe x (cf. Analyse des données). Les barres d'erreur correspondent à l'erreur standard. La droite obtenue par régression est également affichée. Les moyennes des seuils de discrimination et d'identification calculées à partir des tâches de pré- et post-test sont également représentées (lignes en gris, pleine et en pointillé).

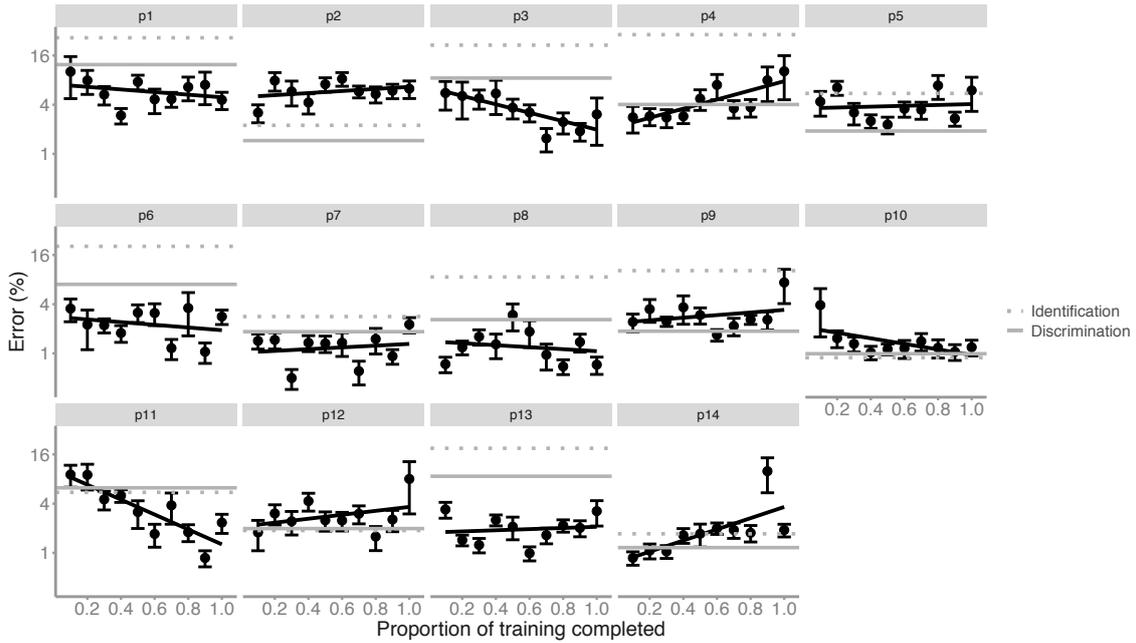


Figure 2.8: Expérience 1 - Tâche audio-motrice. Évolution des performances (% d'erreur) au fur et à mesure de l'entraînement pour chaque participant. Format identique à la Figure 2.7.

pré- et post-tests). Comme le montraient les données moyennées, on voit que les performances sont généralement similaires ou meilleures à la tâche audio-motrice par rapport aux tâches d'identification et de discrimination (à l'exception de $p2$ et éventuellement $p10$).

Concernant la durée des essais, on voit sur la Figure 2.9 que les valeurs sont relativement stables avec une légère tendance à la diminution de la durée. Les données individuelles (Figure 2.10) montrent en outre que certains participants sont rapides et présentent peu de variabilité dans le temps passé à chaque essai ($p10$, $p13$, $p14$).

On voit également que certains participants semblent progresser en terme de précision ($p3$, $p10$ et $p11$; Figure 2.8). Cependant, il ne semble pas y avoir une amélioration globale au fur et à mesure de l'entraînement (Figure 2.7). Pour évaluer plus précisément les effets d'entraînement, nous avons comparé les premiers 25% d'essais avec les derniers 25% de la tâche audio-motrice pour chaque individu. La Figure 2.11 montre les performances (la précision en pourcentage d'erreur), le score (en nombre de points : cf. Procédure, et la durée des essais au début (axe x) et à la fin (axe y) de l'entraînement. Les points au-dessus de la diagonale correspondent

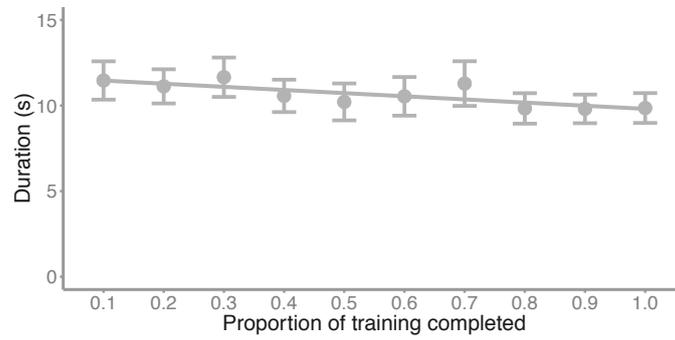


Figure 2.9: Expérience 1 - Tâche audio-motrice. Évolution de la durée des essais au fur et à mesure de l'entraînement. Chaque point représente la moyenne des participants, les barres d'erreur correspondent à l'erreur standard et la droite a été obtenue par régression.

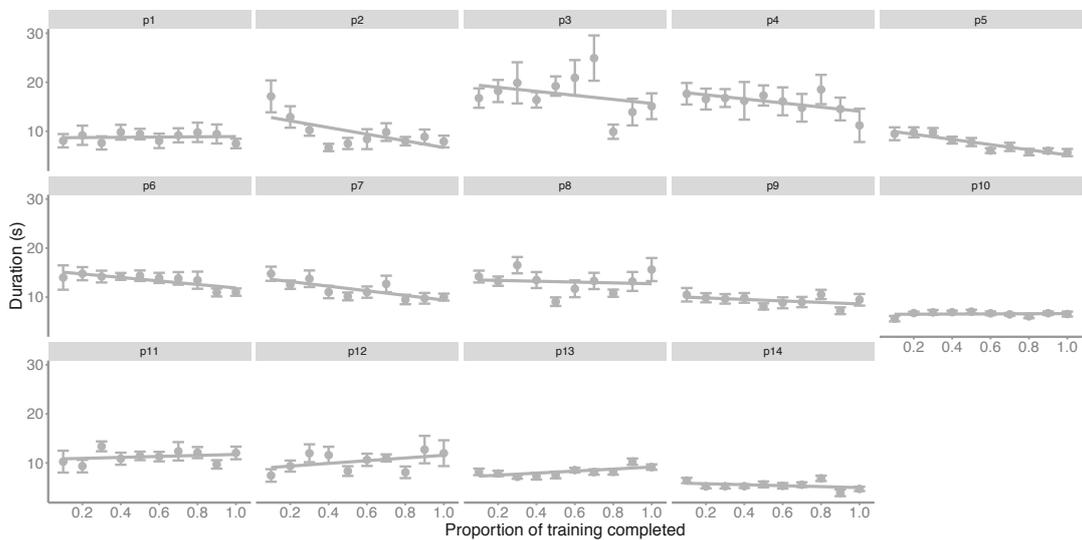


Figure 2.10: Expérience 1 - Tâche audio-motrice. Évolution des durée d'essais (en s) au fur et à mesure de l'entraînement pour chaque participant. Format identique à la Figure 2.9.

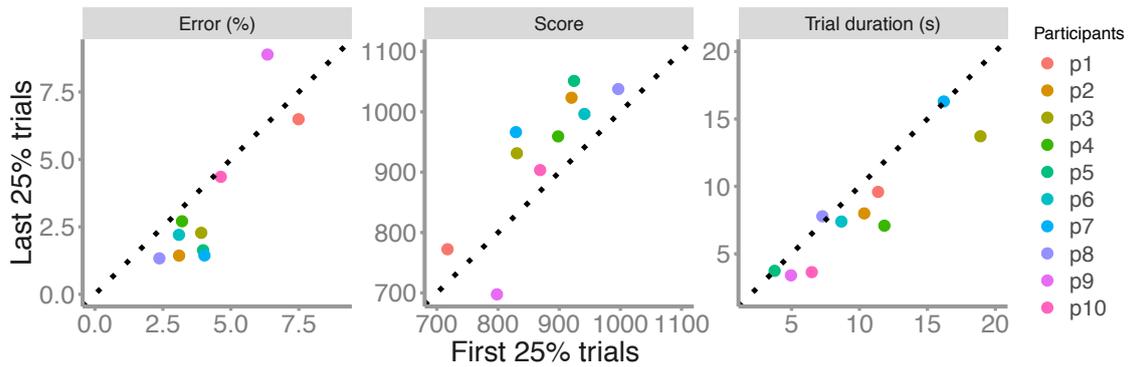


Figure 2.11: Expérience 1 - Tâche audio-motrice. Comparaison de la précision, de la durée de l’essai, et du score au début (premiers 25% des essais) et à la fin (derniers 25% des essais) de l’entraînement. L’échelle des axes est linéaire et non logarithmique.

donc aux participants qui ont de meilleurs résultats à la fin par rapport au début de l’entraînement pour les performances et le score, et à ceux qui ont de moins bons résultats à la fin pour la durée.

Les tests statistiques comparant les premiers 25% et les derniers 25% d’essais n’ont pas montré de différences de performances ou de durée d’essais à la fin de l’entraînement par rapport au début (tous les $p > 0.05$).

2.3.2.2 Pré- et post-tests

La Figure 2.12 montre la différence entre les seuils de discrimination et d’identification pour les pré-tests uniquement de manière à comparer discrimination et identification avant l’apprentissage. Il faut toutefois noter que cette comparaison est réalisée à titre indicatif car il est délicat de comparer les performances à des tâches différentes. On constate que de nombreux participants ont plus de difficulté à faire la tâche d’identification que la tâche de discrimination (seuils d’identification jusqu’à plus de 7 fois plus élevés que les seuils de discrimination). Les seuils sont globalement plus bas pour la tâche de discrimination ($M = 5.01\%$) en comparaison de la tâche d’identification ($M = 11.90\%$). Cela a été confirmé par les tests statistiques ($Z = 7$, $p < 0.01$). En outre, cet effet semble être conservé dans le post-test ($Z = 6$, $p < 0.01$) où les seuils de discrimination ($M = 3.67\%$) sont également plus bas que les seuils d’identification ($M = 10.4\%$).

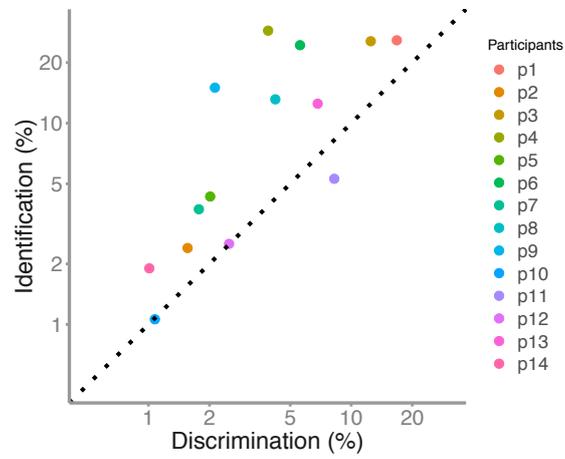


Figure 2.12: Expérience 1 - Pré-/post-tests. Seuils de discrimination (axe x) en fonction des seuils d'identification (axe y) pour le pré-test uniquement.

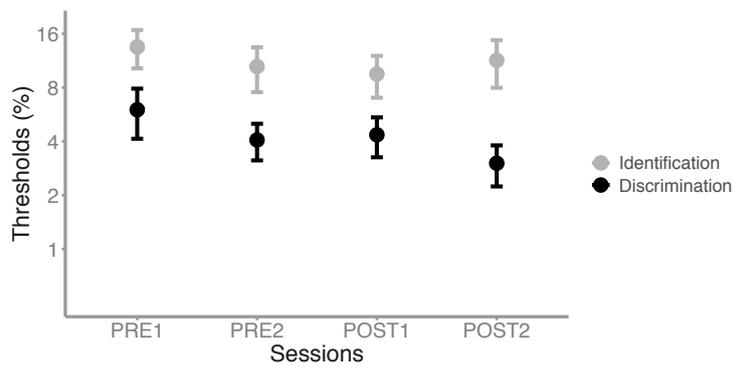


Figure 2.13: Expérience 1 - Pré-/post-tests. Seuils de discrimination et d'identification pour les tests PRE1, PRE2, POST1 et POST2.

Enfin, la Figure 2.13 compare les seuils de discrimination et d'identification de fréquence pour les pré- et post-tests. On observe une légère diminution des seuils de discrimination entre pré- ($M = 5.05\%$) et post-test ($M = 3.67\%$). Cependant, les tests statistiques n'ont pas mis en avant d'amélioration de performance, ni procédural, ni d'habituation à la tâche ou d'entraînement (tous les $p > 0.05$).

2.3.3 Discussion

Tout d'abord, aucune amélioration des performances n'a été observée sur l'une ou l'autre de ces tâches. En outre, les résultats n'ont pas montré d'amélioration des performances à la tâche audio-motrice au fur et à mesure de l'entraînement.

Les participants ont réalisé la tâche audio-motrice avec une précision dépassant celle obtenue à la tâche de discrimination de fréquence. Cependant, il est à nouveau délicat de comparer ces tâches car leurs implications sont différentes. Par exemple, on peut supposer que la quantité d'information obtenue par le biais de la tâche audio-motrice est plus importante (notamment grâce à la durée de l'essai). On peut également noter que la charge mnésique associée aux diverses tâches est différente. Par exemple, Kishon-Rabin et al. (2001) observent des seuils significativement plus élevés avec une procédure 3I-2AFC par rapport à une procédure 2I-2AFC. Ils avancent l'hypothèse selon laquelle la mémoire joue un rôle important dans les tâches d'identification de fréquence et que pour cette raison, la présence de 3 intervalles plutôt que 2 rendrait la tâche plus difficile. Il serait ainsi nécessaire de modéliser et de quantifier les informations sensorielles disponibles dans chacune de ces tâches pour être en mesure de les comparer correctement.

Par ailleurs, les résultats n'ont pas montré d'amélioration significative à la tâche audio-motrice. Il est possible que l'entraînement de 30 minutes n'ait pas été assez long pour produire une amélioration des performances.

De meilleures performances à la condition discrimination par rapport à la condition identification ont également été observées. Cela réplique les résultats de Semal & Demany (2006). Les auteurs montrent en effet que certains participants sont incapables de dire la direction d'un changement de fréquence qu'ils ont pourtant détecté. Dans notre étude, comme dans celle de Semal & Demany (2006), la fréquence standard changeait à chaque essai (*roving*) ce qui est susceptible d'avoir eu une influence sur cet effet (Mathias et al., 2010).

Enfin, aucune amélioration des seuils de discrimination ou d'identification (pré/post-tests) n'a été observée grâce à l'entraînement audio-moteur. Cependant, l'amé-

lioration des performances entre PRE1 et POST2 dans la condition discrimination laisse penser que le fait de faire la tâche pendant un certain temps peut conduire à un apprentissage perceptif. Cela ne semble pas être le cas dans la condition identification. Dans notre première étude exploratoire, les tâches pré-/post-tests n'ont été réalisées que pendant 40 minutes au total (20 minutes pour la condition discrimination et 20 minutes pour la condition identification), ce qui est loin de la durée d'entraînement typique des travaux sur l'apprentissage perceptif.

Il est possible que la présence du *roving* de la fréquence dans les tâches pré-/post-tests et dans la tâche audio-motrice d'entraînement ait bloqué l'apprentissage (Mathias et al., 2010; Semal & Demany, 2006), malgré notre entraînement audio-moteur. De manière à étudier l'effet du *roving* sur cet absence d'apprentissage, une seconde expérience a été mise en place dans laquelle il n'y a pas de *roving* de la fréquence. Le but est ainsi de voir si c'est le *roving* qui a eu un effet distracteur sur l'apprentissage et si la mise en action de la boucle audio-motrice peut conduire à une amélioration rapide des seuils de discrimination et/ou d'identification de la fréquence lorsqu'il n'y a pas de *roving*.

2.4 Expérience 2

Dans l'expérience 1, nous n'avons pas observé d'apprentissage perceptif de la perception de la hauteur lié à l'entraînement audio-moteur. Une explication potentielle est la présence du *roving* de fréquence. En effet, comme exposé lors de la revue bibliographique, les changements de paramètres d'essai en essai peuvent limiter les effets de l'apprentissage perceptif (Amitay et al., 2005). Pour vérifier l'hypothèse d'un effet du *roving* de la fréquence sur l'apprentissage perceptif, nous avons réalisé une expérience similaire à l'expérience 1, mais sans *roving* de la fréquence. Mis à part ce paramètre, la procédure, le matériel et les stimuli sont identiques à ceux de l'expérience 1.

2.4.1 Méthodes

2.4.1.1 Participants

Dix participants (6 femmes ; age en années $M = 24.2$, $SD = 4.13$) normo-entendants n'ayant pas participé à l'expérience 1 ont participé à cette étude. Ils ont été indemnisés 10 euros de l'heure. Un audiogramme préalable a permis de s'assurer que tous les participants recrutés pour l'expérience avaient une audition normale,

c'est à dire des seuils auditifs de 20 dB HL ou moins pour les fréquences testées (250, 500, 1000, 2000, 3000, 4000, et 8000 Hz). Tous les participants étaient non-musiciens (moins de deux ans de pratique musicale au moment de l'étude).

2.4.1.2 Stimuli

Discrimination et identification de fréquences

Les stimuli sont identiques à ceux de la tâche discrimination/identification de fréquence de l'expérience 1 mis à part qu'il n'y a pas de *roving* de fréquence. Le premier son pur de chaque paire était fixé à 979.8 Hz pour tous les essais réalisés, ce qui correspond au centre de l'ambitus du *roving* de l'expérience 1 (400-2400 Hz).

Tâche audio-motrice

Les stimuli étaient identiques à ceux de la tâche audio-motrice de l'expérience 1 mis-à-part que la fréquence du son pur cible (le plus aigu) était de 979.8 Hz pour tous les essais. La position associée à cette fréquence était toutefois choisie aléatoirement, comme dans l'expérience 1.

2.4.1.3 Matériel, procédure et analyse de données

L'ensemble du matériel et de la procédure de l'expérience 1 a été utilisé pour l'expérience 2. L'analyse des données est également identique.

2.4.2 Résultats

2.4.2.1 Tâche audio-motrice

Comme dans l'expérience 1, les Figures 2.14 et 2.15 montrent des exemples de parcours réalisés par les participants. Des parcours moins sinueux caractérisent les essais réalisés à la fin en comparaison des essais réalisés au début de l'entraînement.

La Figure 2.16 montre l'évolution du pourcentage d'erreur à la tâche audio-motrice au cours de la pratique. Les seuils de discrimination et d'identification (moyennés entre pré- et post-tests) sont également représentés. On constate que la précision est dans le même ordre de grandeur que le seuil d'identification tout en restant légèrement plus élevée. Pour tester un effet potentiel du *roving* sur les performances à la tâche audio-motrice, nous avons comparé les performances des participants (moyennées pour l'ensemble des essais) entre l'expérience 1 et l'expérience 2. Les tests

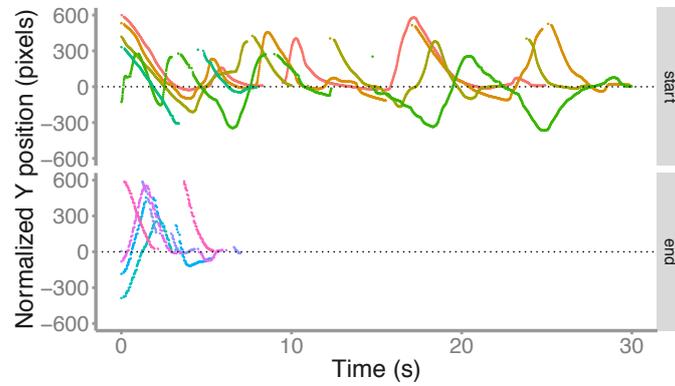


Figure 2.14: Expérience 2 - Tâche audio-motrice. Exemple des 5 premiers essais (haut) et des 5 derniers essais (bas) de la phase d'entraînement d'un participant. Les essais ont été normalisés de manière à avoir la position correspondante à la fréquence la plus aiguë égale à 0.

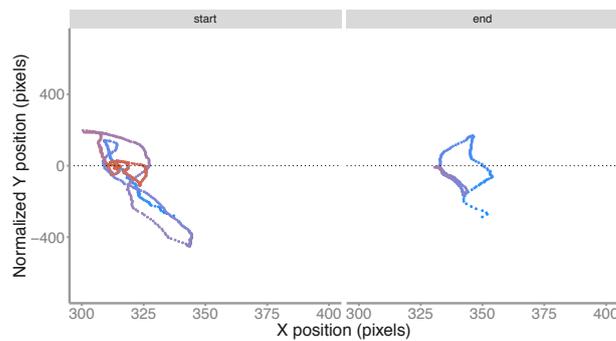


Figure 2.15: Expérience 2 - Tâche audio-motrice. Exemple de parcours (coordonnées x et y) réalisés par un participant dans un essai au début (gauche) et à la fin (droite) de la phase d'entraînement. Les coordonnées y ont été normalisées de manière à avoir la position correspondante au son pur le plus aigu centrée en 0.

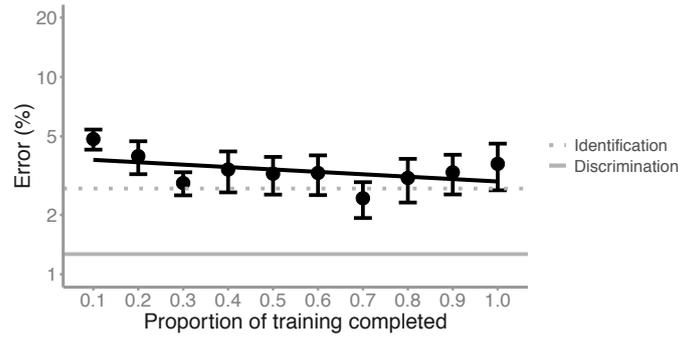


Figure 2.16: Expérience 2 - Tâche audio-motrice. Évolution des performances (% d'erreur) au fur et à mesure de l'entraînement. Chaque point représente la moyenne des participants, les barres d'erreur correspondent à l'erreur standard et la droite de régression est également représentée. Les seuils de discrimination et d'identification moyens calculés à partir des tâches de pré- et post-test sont aussi représentés (en gris).

statistiques n'ont pas montré de différences significatives entre les deux expériences ($Z = 70$, $p > 0.05$).

Comme dans l'expérience 1, nous avons représenté les données individuelles des participants montrant l'évolution des performances au fur et à mesure de l'entraînement sur la Figure 2.17. Ces données montrent que certains participants ont eu beaucoup plus de difficulté à réaliser la tâche audio-motrice que les tâches de discrimination et d'identification de fréquence ($p4$, $p9$, $p10$). Ce n'était pas le cas dans l'expérience 1, principalement par le fait de seuils de discrimination et d'identification plus bas dans cette expérience (voir section pré-/post-tests plus bas).

On voit également une légère diminution de l'erreur au fur et à mesure de l'entraînement (Figure 2.16). Pour vérifier cela, nous avons comparé les essais réalisés au début et à la fin de l'entraînement. Comme pour l'expérience 1, la Figure 2.20 montre l'erreur (en pourcentage), le score (en nombre de points; calculé à partir de la précision et de la durée de l'essai), et la durée des essais, au début (axe x) et à la fin (axe y) de l'entraînement. Cette amélioration a été confirmée par les tests statistiques comparant la durée des essais ($Z = 50$; $p = 0.02$) et les scores ($Z = 6$; $p = 0.03$) pour les premiers 25% et les derniers 25% des essais. Cependant, la comparaison de la précision n'a pas atteint notre seuil de significativité ($Z = 50$; $p = 0.06$).

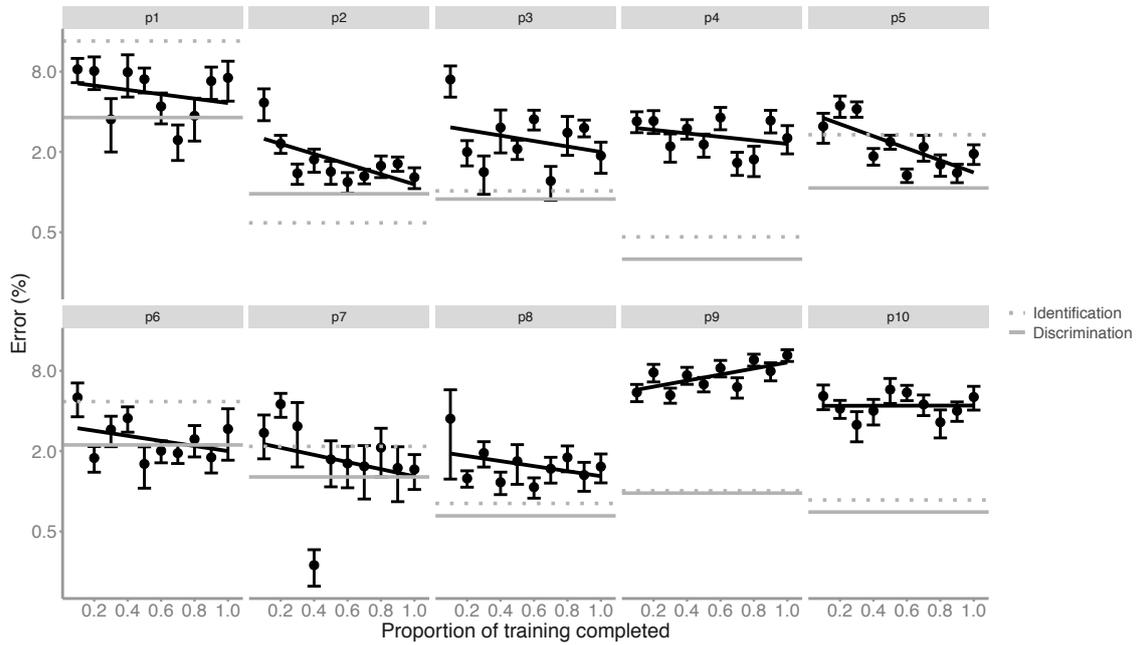


Figure 2.17: Expérience 2 - Tâche audio-motrice. Évolution des performances (% d'erreur) au fur et à mesure de l'entraînement pour chaque participant. Les seuils de discrimination et d'identification moyens calculés à partir des tâches de pré- et post-test sont également représentés (en gris).

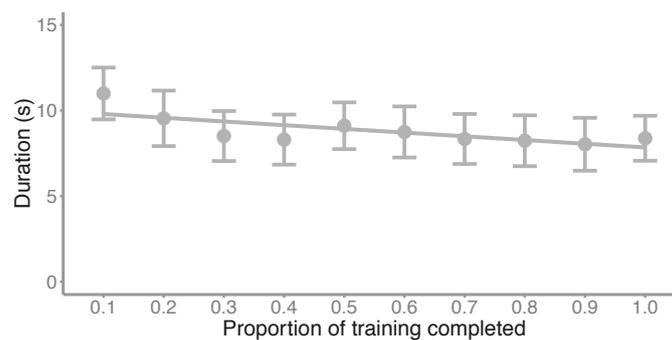


Figure 2.18: Expérience 2 - Tâche audio-motrice. Évolution des durée d'essais (en s) au fur et à mesure de l'entraînement. Chaque point représente la moyenne des participants, les barres d'erreur correspondent à l'erreur standard et la droite correspond à la droite de régression.

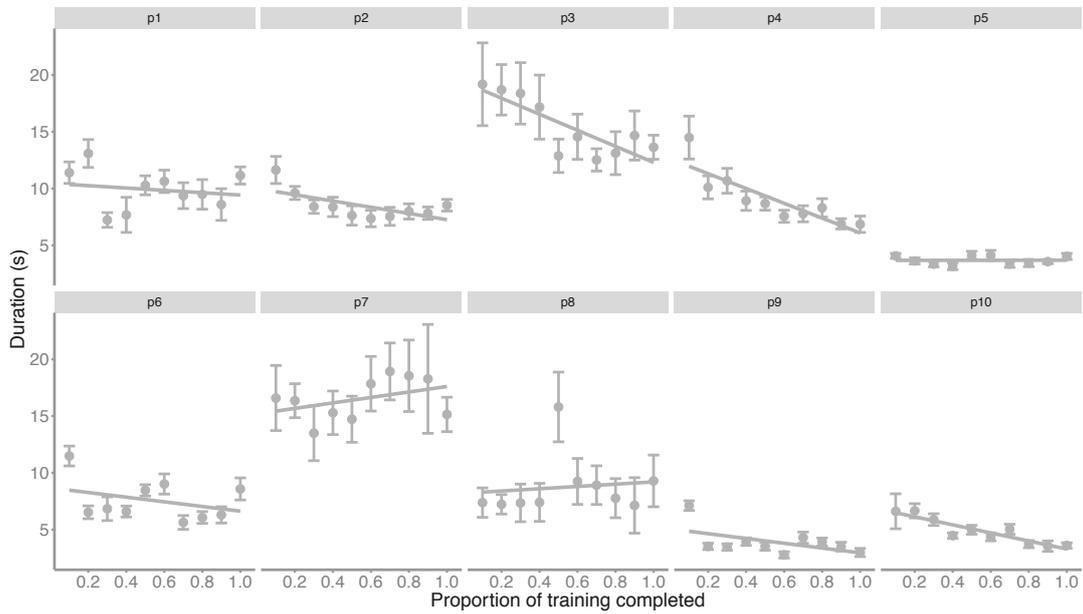


Figure 2.19: Expérience 2 - Tâche audio-motrice. Évolution des durée d'essais (en s) au fur et à mesure de l'entraînement pour chaque participant.

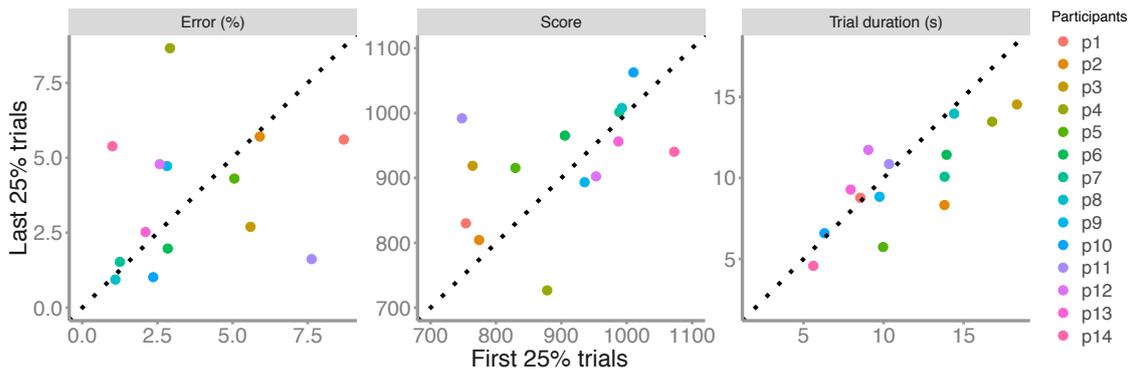


Figure 2.20: Expérience 2 - Tâche audio-motrice. Comparaison de la précision, de la durée de l'essai, et du score au début (premiers 25% des essais) et à la fin (derniers 25% des essais) de l'entraînement. L'échelle des axes est linéaire et non logarithmique.

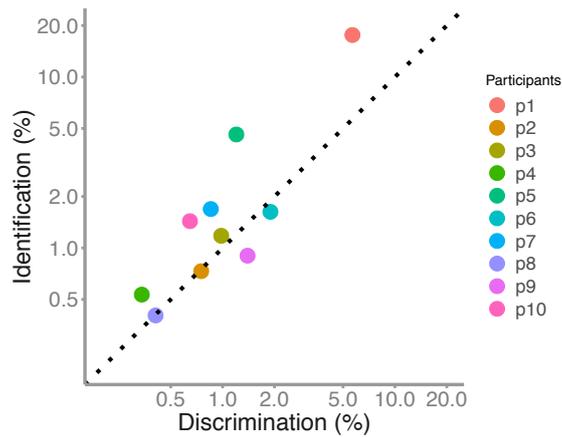


Figure 2.21: Expérience 2 - Pré-/post-tests. Seuils de discrimination (axe x) en fonction des seuils d'identification (axe y) pour le pré-test uniquement.

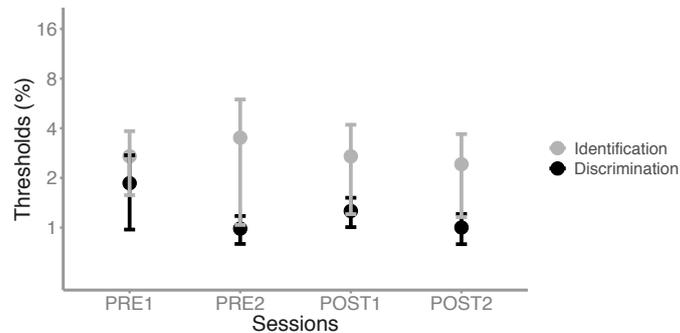


Figure 2.22: Expérience 2 - Pré-/post-tests. Seuils de discrimination et d'identification de fréquence moyens pour PRE1, PRE2, POST1 et POST2.

2.4.2.2 Pré- et post-tests

La Figure 2.21 montre les seuils de discrimination et d'identification de chaque participant au pré-test. On constate que la majorité des points sont autour de la diagonale ce qui implique que les scores de discrimination sont similaires aux scores d'identification. Les tests statistiques n'ont pas montré de différences de performances significatives entre discrimination ($M = 1.27\%$) et identification ($M = 2.72\%$; $Z = 6$, $p > 0.05$).

La Figure 2.22 montre les seuils de discrimination et d'identification de fréquence pour les pré- et post-tests. On voit que les seuils sont relativement stables dans les différentes conditions. Les tests statistiques n'ont pas mis en avant d'amélioration

de performance, ni procédurale, ni d’habituatation à la tâche, ni auditive (tous les $p > 0.05$).

De manière à évaluer l’effet du *roving* sur les performances dans les tâches de discrimination et d’identification, les résultats des expériences 1 et 2 ont été comparés. Un effet délétère significatif du *roving* a été observé pour les tâches de discrimination ($M = 1.3\%$ sans *roving* et $M = 4.4\%$ avec *roving* ; $Z = 20$, $p = 0.001$) et d’identification ($M = 2.8\%$ sans *roving* et $M = 11\%$ avec *roving* ; $Z = 20$, $p = 0.001$).

La différence de performance entre discrimination et identification a été comparée dans les expériences 1 (avec *roving*) et 2 (sans *roving*). Malgré un écart important (différence de 6.9 points dans la condition avec *roving* et différence de 1.6 points dans la condition sans *roving*), les tests statistiques n’ont pas révélés d’écart significatif ($Z = 40$, $p > 0.05$).

2.4.3 Discussion

Tout d’abord, les seuils calculés à partir des tâches de discrimination et d’identification de fréquence sont largement affectés par le *roving* et les performances sont ainsi améliorées par rapport à l’expérience 1. Cela est en accord avec de précédentes études (Amitay et al., 2005; Mathias et al., 2010) qui montrent un effet du *roving*.

Ensuite, nous n’avons pas observé d’amélioration des performances avec la pratique, contrairement à Micheyl et al. (2006) ou Kishon-Rabin et al. (2001) par exemple, qui ont montré que les seuils d’identification pouvaient être améliorés de manière significative. Une explication possible est la courte durée de la pratique dans notre cas (4 conditions : PRE1, PRE2, POST1, POST2, environ 10 minutes par condition) et un entraînement de 30 minutes.

Les seuils d’identification que nous avons observé ($M = 2.72\%$) sont un peu plus élevés que dans des études précédentes (par exemple, Kishon-Rabin et al. (2001) : 1.153% ; Micheyl et al. (2006) : environ 1.5% pour le groupe de non-musiciens, première session d’entraînement et sons purs purs). Dans notre cas et comme dans l’expérience 1, des seuils plus élevés pourraient être expliqués par la présence de 4 sons purs à stocker en mémoire ainsi que de la courte durée du test.

Nous avons ensuite observé une différence de performance entre les conditions discrimination et identification. Cependant, contrairement à Mathias et al. (2010), cette différence n’était pas significative. Cela peut s’expliquer par la manière de sélectionner les participants et par la grande variabilité inter-individuelle qui caractérise

cet effet (Amitay et al., 2005). Mathias et al. (2010) sélectionnent les participants sensibles à la différence de condition avec *roving* et les testent à nouveau sans *roving*. Dans notre cas, il se peut que les participants qui ne sont pas sensibles à cette différence aient réduit l'écart entre les conditions avec et sans *roving*.

On peut noter que Amitay et al. (2005) et Matthews & Stewart (2008) ont observé un effet assez faible du *roving* lorsque les fréquences utilisées couvraient une large bande de fréquence comme c'était le cas dans notre expérience 1. Pourquoi, dans notre cas, l'effet du *roving* est-il plus important ? Une différence majeure est que, dans les travaux de Amitay et al. (2005) et Matthews & Stewart (2008), le *roving* consistait à utiliser plusieurs valeurs fixes de fréquences et non une fréquence aléatoire choisie à chaque essai comme c'était le cas dans notre étude et dans celle de Semal & Demany (2006). Or, Amitay et al. (2005) observent un effet distracteur plus important pour un *roving* restreint en fréquence. Il est possible qu'avec des fréquences éloignées, et donc qui dépendent de bandes attentionnelles différentes, les participants aient pu surveiller les différents sons purs standards plus facilement que dans le cas où les fréquences dépendaient de la même bande attentionnelle. Cela n'est pas possible dans le cas où les fréquences sont choisies aléatoirement à chaque essai, comme dans notre expérience.

Concernant la tâche audio-motrice, les résultats sont comparables à ceux de l'expérience 1. Cela suggère donc que le *roving* n'a pas d'effet distracteur dans cette tâche (ou en tout cas qu'un effet distracteur ne peut être mis en évidence avec nos analyses). Dans le cas des tâches à choix forcé sans *roving*, la fréquence standard est la même d'un essai à l'autre et le son pur test peut être comparé à une référence plus stable. En revanche, la tâche audio-motrice peut être réalisée sans qu'une référence soit stockée en mémoire, mais par la discrimination du son pur le plus aigu. Il est donc possible que cette différence explique l'avantage lié à l'absence de *roving* uniquement dans les tâches à choix forcé.

On peut également noter que dans cette expérience sans *roving*, la précision est plus grande dans les tâches pré-/post-tests que dans la tâche audio-motrice. Encore une fois, on peut se questionner sur cette différence avec les réserves liées à la comparaison de tâches aux implication perceptives différentes notées plus haut. Wier et al. (1976) ont montré que les seuils d'identification de fréquence étaient plus bas (d'un facteur allant jusqu'à 2) lorsqu'ils étaient évalués avec une méthode d'ajustement (ressemblant par certains aspects à notre tâche). On aurait également pu attendre de meilleures performances dans notre tâche car elle ne nécessite pas de garder un son cible en mémoire (mais simplement de trouver le son pur le plus

aigu). Une explication possible est que la prise en compte de la durée de l'essai dans le score a poussé les participants à être rapide au détriment de la précision.

Les données individuelles à la tâche audio-motrice ont également montré que certains participants (*p4*, *p9* et *p10*) présentaient de bons seuils aux tâches de discrimination et d'identification mais de mauvaises performances à la tâche audio-motrice. Il est donc possible que ces participants aient rencontré une difficulté particulière dans la tâche audio-motrice (par exemple liée à l'interface, à la motricité, ou à la stratégie), qui ne soit pas dépendante de leur capacité à percevoir des différences de fréquence.

Pour résumer, il semble que l'entraînement audio-moteur réalisé pendant 30 minutes n'a pas eu d'impact sur les capacités de discrimination et d'identification de fréquence, dans les conditions avec (expérience 1) et sans *roving* (expérience 2). Une hypothèse est que la durée des expériences n'a pas été suffisante pour qu'un apprentissage perceptif se produise. Le but de l'expérience 3 est de tester cette hypothèse grâce à un entraînement similaire à celui de l'expérience 1, mais sur 5 jours consécutifs.

2.5 Expérience 3

Une explication possible de l'absence d'effet d'entraînement pour l'expérience 1 pourrait être la courte durée de l'entraînement (30 minutes). Le but de ces expériences courtes était d'évaluer de manière pilote l'efficacité et la rapidité potentielle de l'entraînement audio-moteur. Les études sur l'apprentissage perceptif proposent généralement un entraînement plus conséquent (par exemple, Micheyl et al., 2006 : jusqu'à 14h ; Deveau et al., 2014 : environ 15h réparti sur 8 semaines). Pour cette raison, 4 participants de l'expérience 1 ont été rappelés pour réaliser à nouveau l'ensemble des tâches chaque jour pendant 4 jours supplémentaires (5 jours au total). Le but est de tester si une amélioration des seuils de discrimination et d'identification de la fréquence pouvait être observée malgré le *roving* de la fréquence.

2.5.1 Méthodes

2.5.1.1 Participants

Quatre participants ayant réalisé l'expérience 1 ont également participé à cette expérience (3 femmes ; age $M = 24$, $SD = 4.3$).

2.5.1.2 Stimuli

Les stimuli des tâches de discrimination/identification de fréquence et de la tâche audio-motrice sont identiques à ceux de l'expérience 1 (présence d'un *roving* de la fréquence).

2.5.1.3 Matériel, procédure et analyse de données

L'ensemble du matériel et de la procédure de l'expérience 1 a été utilisé pour l'expérience 3, mis à part que la session d'une heure comprenant PRE1, PRE2, entraînement, POST1 et POST2 (Figure 2.1) a été réalisée chaque jour pendant 5 jours. L'analyse des données est également identique aux expériences précédentes. Cependant, aucun test statistique n'a été réalisé à cause du faible nombre de participants.

La première session (PRE1, PRE2, tâche audio-motrice, POST1, POST2) correspond à celle réalisée dans l'expérience 1 concaténée à 4 autres sessions.

2.5.2 Résultats

2.5.2.1 Tâche audio-motrice

La Figure 2.23 montre l'évolution du pourcentage d'erreur à la tâche audio-motrice au cours de la pratique sur les 5 jours. Les seuils de discrimination et d'identification (moyennés entre pré- et post-tests) sont également représentés. Elle montre qu'il n'y a pas d'amélioration des performances au fur et à mesure de l'entraînement. La tendance va même dans le sens d'une baisse des performances, potentiellement expliquée par des effets de lassitude à réaliser la tâche. De plus, on constate que les performances à la tâche audio-motrice sont similaires aux seuils de discrimination. Les données individuelles (Figure 2.24) montrent que les participants ont des performances assez stables mais qu'une tendance à l'amélioration des performances semble se produire pour $p4$.

La Figure 2.25 montre les durées des essais au fur et à mesure de l'entraînement. Il semble y avoir une légère baisse de la durée des essais (probablement issue principalement de $p1$ et $p3$; Figure 2.26).

Comme dans les expériences précédentes, nous avons évalué l'évolution des performances des participants en comparant les premiers 25% des essais réalisés avec les derniers 25%. La Figure 2.27 représente ainsi la précision, le score et la durée des essais au début (axe x) et à la fin (axe y de l'entraînement).

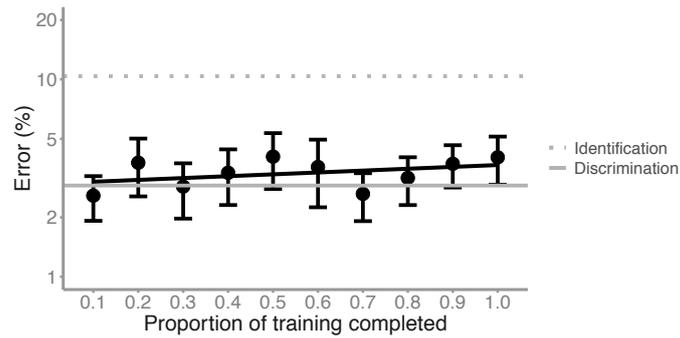


Figure 2.23: Expérience 3 - Tâche audio-motrice. Évolution des performances (% d'erreur) au fur et à mesure de l'entraînement. Chaque point représente la moyenne des participants, les barres d'erreur correspondent à l'erreur standard et la droite correspond à la droite de régression. Soulignons que l'abscisse est toujours en proportion de l'entraînement réalisé, mais maintenant cette échelle varie de 0 à approximativement 150 minutes au lieu de 0 à 30 minutes dans les Figures précédentes. Les seuils de discrimination et d'identification moyens calculés à partir des tâches de pré- et post-test sont également représentés (en gris).

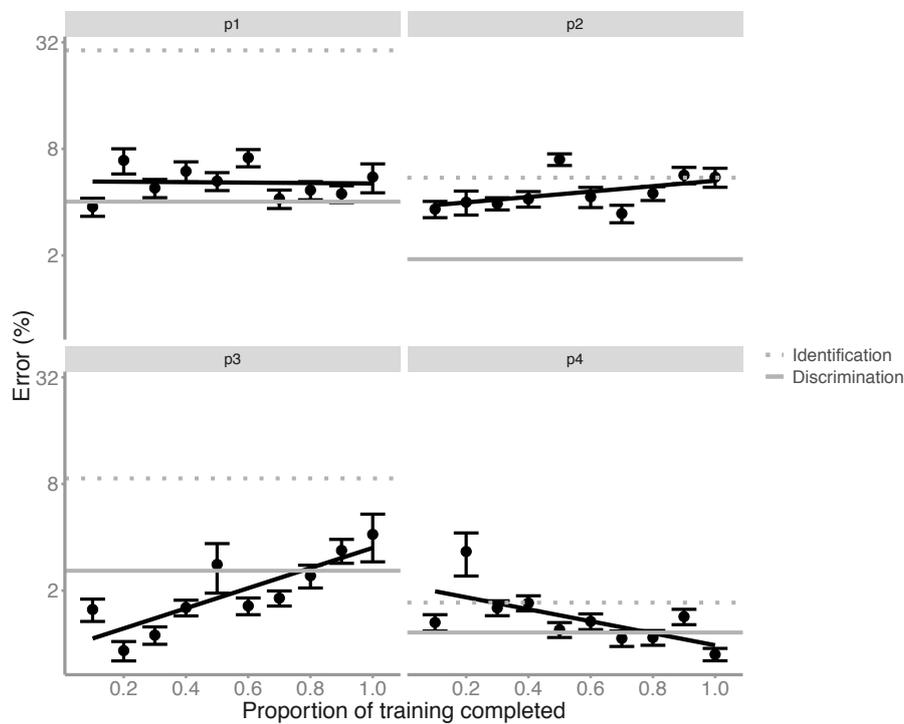


Figure 2.24: Expérience 3 - Tâche audio-motrice. Évolution des performances (% d'erreur) au fur et à mesure de l'entraînement pour chaque participant. Format identique à la Figure 2.23.

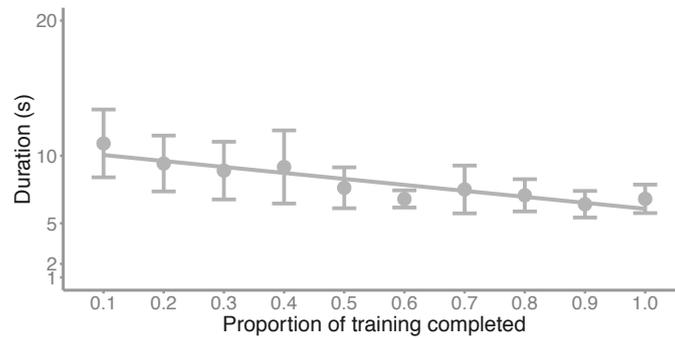


Figure 2.25: Expérience 3 - Tâche audio-motrice. Évolution des durée d'essais (en s) au fur et à mesure de l'entraînement. Chaque point représente la moyenne des participants, les barres d'erreur correspondent à l'erreur standard et la droite correspond à la droite de régression.

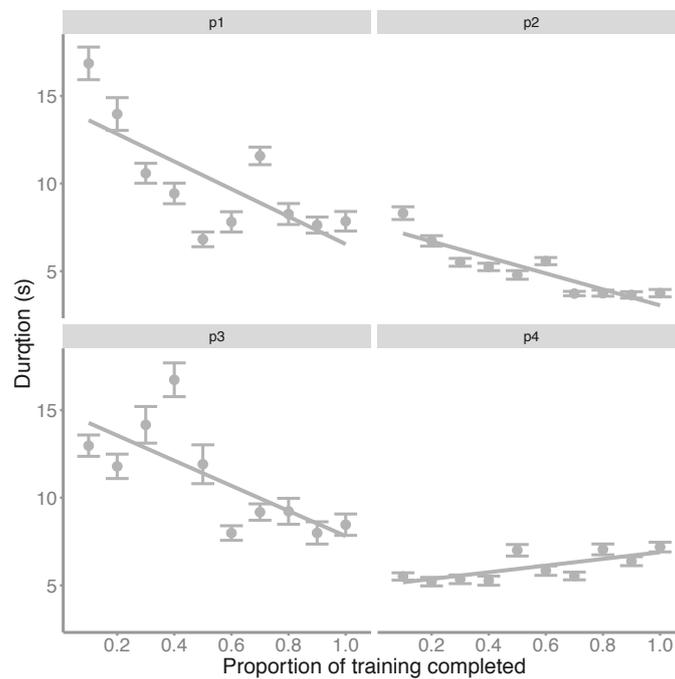


Figure 2.26: Expérience 3 - Tâche audio-motrice. Évolution des durée d'essais (en s) au fur et à mesure de l'entraînement pour chaque participant. Format identique à la Figure 2.25.

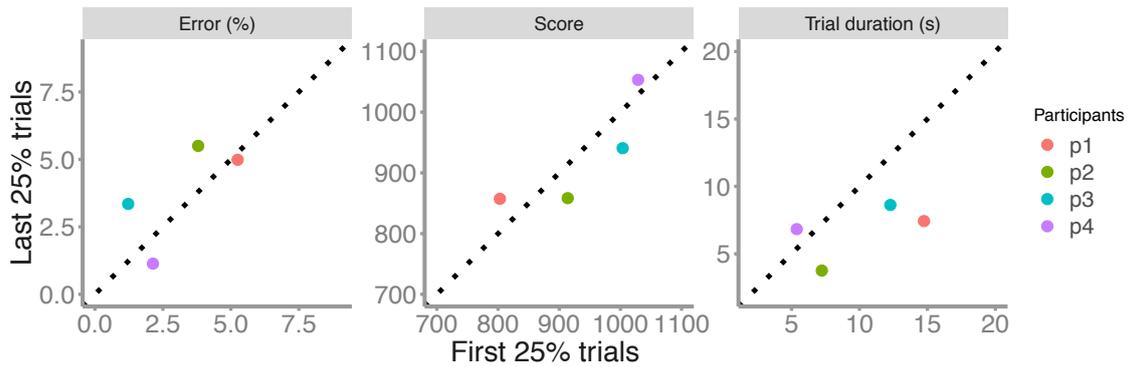


Figure 2.27: Expérience 3 - Tâche audio-motrice. Comparaison de la précision, de la durée de l'essai, et du score au début (premiers 25% des essais) et à la fin (derniers 25% des essais) de l'entraînement. L'échelle des axes est linéaire et non logarithmique.

2.5.2.2 Pré- et post-tests

La Figure 2.28 montre les seuils d'identification et de discrimination pour toutes les sessions. Quatre tâches étaient réalisées chaque jour (2 PRE et 2 POST) conduisant à un total de 20 mesures de seuils de discrimination et d'identification. Le graphique suggère qu'il n'y a pas de changement de seuil au cours de la pratique. Nous avons également représenté les données individuelles (Figure 2.29) qui montrent une tendance à la diminution des seuils de discrimination pour $p3$ (moyenne PRE1 et PRE2 : 4.22% ; moyenne POST9 et POST10 : 1.97%) et $p4$ (moyenne PRE1 et PRE2 : 1.01% ; moyenne POST9 et POST10 : 0.64%) et d'identification pour $p3$ (moyenne PRE1 et PRE2 : 13.37% ; moyenne POST9 et POST10 : 3.68%) et $p4$ (moyenne PRE1 et PRE2 : 1.90% ; moyenne POST9 et POST10 : 0.91%).

La Figure 2.30 montre les seuils de discrimination en fonction des seuils d'identification. On voit que, comme dans l'expérience 1, les participants sont plus performants à la tâche de discrimination par rapport à la tâche d'identification.

2.5.3 Discussion

Cette expérience se distingue de l'expérience 1 par le nombre de répétitions réalisées par les participants (une heure comprenant pré-/post-tests et entraînement une fois

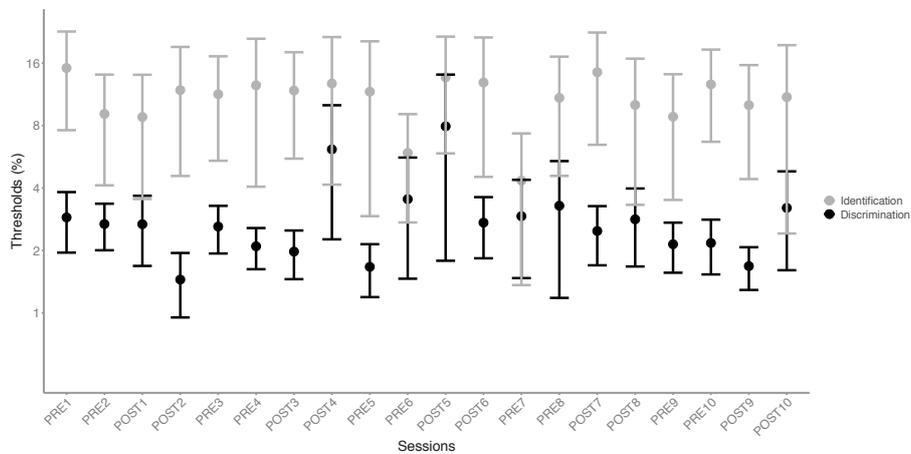


Figure 2.28: Expérience 3 - Pré-/post-tests. Seuils de discrimination et d'identification de fréquence moyens pour chaque session de PRE1 (jour 1) à POST10 (jour 5).

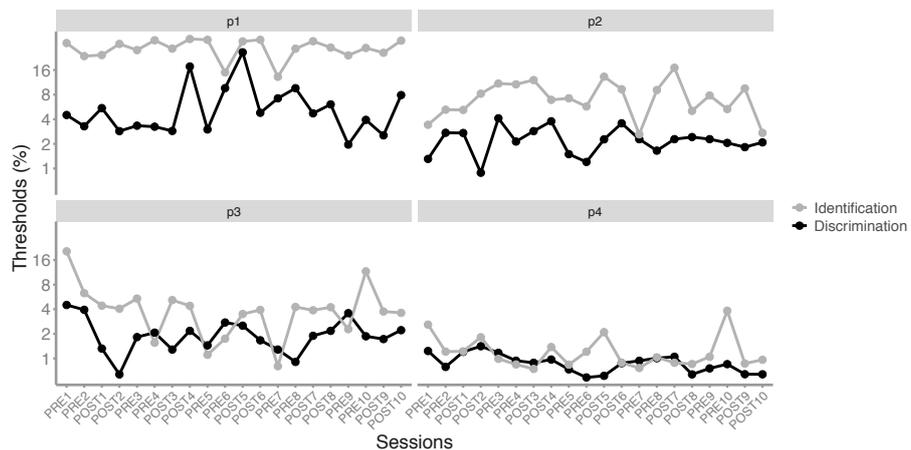


Figure 2.29: Expérience 3 - Pré-/post-tests. Seuils de discrimination et d'identification de fréquence moyens pour chaque session de PRE1 (jour 1) à POST10 (jour 5) pour chaque participant.

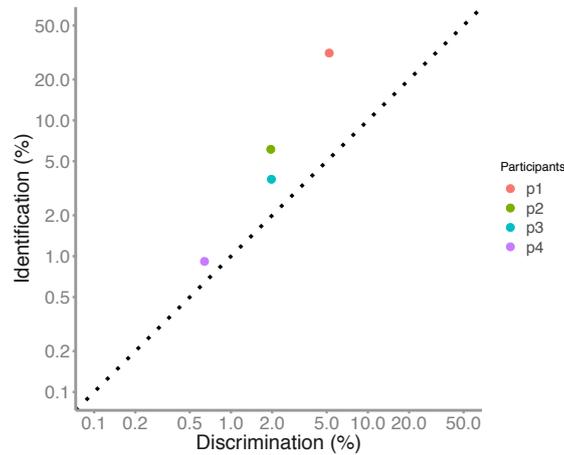


Figure 2.30: Expérience 3 - Pré-/post-tests. Seuils de discrimination (axe x) en fonction des seuils d'identification (axe y) pour PRE1 et PRE2 uniquement.

par jour pendant 5 jours). A part cela, les paramètres utilisés étaient similaires à ceux de l'expérience 1.

La précision obtenue à la tâche audio-motrice est semblable à celle observée dans l'expérience 1. En outre, dans les deux cas, la précision à la tâche audio-motrice est similaire au seuil de discrimination.

En outre, comme dans l'expérience 1, certains participants semblent présenter plus de difficultés dans la tâche d'identification ($p1$ et $p3$).

Il semble enfin que les performances de discrimination et d'identification de fréquence n'aient pas été améliorées par la pratique ou l'entraînement audio-moteur malgré la répétition de la procédure sur 5 jours. Les participants ont réalisé environ 1000 essais de la tâche à double-paires sur l'ensemble des 5 jours. Cela va donc dans le sens de l'étude de Semal & Demany (2006) où les auteurs n'observent pas d'apprentissage perceptif dans cette tâche.

2.6 Discussion générale

L'effet d'un entraînement perceptif mettant en action la boucle audio-motrice a été examiné sur l'amélioration de la perception de la hauteur musicale avec et sans *roving* de la fréquence. Pour évaluer ce protocole d'entraînement perceptif, nous avons utilisé une tâche permettant de mesurer les seuils de discrimination et d'identification de fréquence et dans lequel l'apprentissage perceptif semblait difficile (Semal & Demany,

2006). Trois expériences (avec *roving* - court, sans *roving* - court et avec *roving* - long) ont été mises en place de manière à étudier les causes d'un non-apprentissage contrairement aux études classiques montrant une amélioration des performances de perception de hauteur (Michey et al., 2006). Dans l'ensemble, aucun apprentissage n'a été observé dans ces trois expériences.

L'expérience 1 a montré que l'ensemble du protocole réalisé avec un *roving* de la fréquence n'était pas associé à un apprentissage significatif. Il est possible que la tâche audio-motrice n'ait pas visé la source de difficulté réelle rencontrée par les sujets dans les tâches de discrimination et d'identification de fréquences. Une manière d'expliquer l'effet du *roving* dans la tâche à double-paires est de considérer que la difficulté est augmentée par les changements de fréquence n'étant pas associés à la tâche (entre le dernier son pur d'un essai et le premier son pur de l'essai suivant, ou bien entre le deuxième son pur de la première paire avec le premier son pur de la seconde ; Mathias et al., 2011). Dans ce cas, il est possible qu'un protocole d'apprentissage visant à rendre les participants plus aptes à ignorer la variabilité qui n'est pas associée à la tâche elle-même soit plus efficace.

Les résultats à la tâche audio-motrice ont montré des performances similaires avec et sans le *roving* de fréquence. Cela laisse en effet penser que le changement de la fréquence cible à chaque essai n'est pas une source de variabilité interférente et que l'entraînement n'a peut-être pas particulièrement favorisé la résistance à l'incertitude du stimulus. Par ailleurs, on peut noter que sans *roving*, la tâche audio-motrice peut être réalisée en mémorisant le son pur cible et en le cherchant sur la tablette. Cela se rapproche d'une tâche de *pitch matching* dans laquelle la direction du changement de fréquence n'est pas nécessaire à percevoir. Le fait que cette tâche n'a pas conduit à de meilleures performances laisse penser que sans *roving*, la meilleure stratégie n'est pas de chercher le son pur cible mais d'écouter les changements de direction de fréquence jusqu'à ce que le pic soit identifié.

Pour évaluer si l'absence d'apprentissage observé dans l'expérience 1 était lié au *roving* de la fréquence, nous avons réalisé une seconde expérience dans laquelle la fréquence des sons purs standard était fixe dans les pré-/post-tests et dans la tâche audio-motrice. Les résultats ont montré que les seuils de discrimination et d'identification sont plus bas dans cette deuxième condition sans *roving* de la fréquence, confirmant des travaux antérieurs (par exemple, Amitay et al. (2005); Demany & Semal (2005); Mathias et al. (2010)). Comme pour la perception de l'intensité (Durlach & Braida, 1969), cette différence peut être expliquée par le fait que dans la condition sans *roving*, le son pur test peut être comparé à une

représentation interne construite à partir de l'ensemble des essais passés alors que dans la condition avec *roving*, le son pur test doit être comparé au son pur standard entendu une seule fois (Mathias et al., 2010).

Les résultats de l'expérience 2 n'ont pas révélé d'effet d'amélioration des performances liée à l'entraînement perceptif malgré l'absence de *roving*. Il semble donc qu'un apprentissage rapide n'ait pas été produit par la phase d'apprentissage audio-moteur. On aurait pu attendre une amélioration liée à la répétition des tâches pré-/post-tests, cependant, contrairement aux études où de tels effets sont observés, l'ensemble de la tâche a été réalisée sur une courte durée. Une prédiction est qu'une réduction des seuils de discrimination et d'identification de fréquence aurait pu se produire avec un plus grand nombre de répétitions des tâches pré-/post-tests.

En outre, si l'expérience 1 a montré qu'un apprentissage rapide n'avait pas été produit par notre protocole d'entraînement audio-moteur, on peut se demander si un entraînement plus long aurait permis une amélioration des seuils perceptifs malgré le *roving*. Les résultats de l'expérience 3 suggèrent que ce n'est pas le cas. Un apprentissage équivalent à celui de l'expérience 1 (donc avec *roving*) a été mis en place sur 5 jours consécutifs (tâches pré-/post-tests et tâche audio-motrice). Comme pour l'expérience 1, aucune amélioration des seuils n'a été observée. Pourtant, sur l'ensemble des pré-/post-tests, les participants ont réalisé un total d'environ 1000 essais (si l'on additionne les essais issus des tâches de discrimination et d'identification). Or, dans l'expérience 1 de Micheyl et al. (2006), les auteurs montrent une amélioration des seuils très rapide, déjà entre la première et la deuxième session (pour le groupe de non-musiciens). Une de leurs sessions consiste en 12 conditions (environ 40 essais par condition), un apprentissage est donc observé alors que les participants n'ont réalisé que $12 \times 40 = 480$ essais. Cependant, notre absence d'apprentissage réplique Mathias et al. (2010) et Semal & Demany (2006) qui n'observent pas d'effets d'entraînement malgré le nombre d'essais réalisé. On pourrait penser que la présence du *roving* de la fréquence a bloqué l'apprentissage, mais, les participants de notre expérience 2 ont réalisé ces mêmes tâches sans *roving*, et aucune amélioration des performances n'a été observée. Il est possible que le nombre d'essais (entre 200 et 300) ait été insuffisant, bien que peu éloigné des 480 essais de Micheyl et al. (2006).

Une explication alternative est que la tâche à double-paire est à l'origine de l'absence d'apprentissage. Encore une fois, Semal & Demany (2006) et Mathias et al. (2010) utilisent cette même tâche. Cependant, Demany (1985) utilise aussi une tâche de discrimination à double-paire sans *roving* similaire et observent des effets d'apprentissage rapides (700 essais) ce qui va à l'encontre de cette hypothèse.

2.7 Conclusions

Les résultats obtenus par Mathias et al. (2010) suggèrent que certains participants ne sont pas capables d'ignorer les variations du stimulus qui ne sont pas importants pour la tâche. Les auteurs mettent en avant le lien entre cette aptitude et les performances aux tâches de masquage informationnel dans lesquelles un son pur doit généralement être détecté en présence d'un masqueur dans une région fréquentielle différente. Comme pour la tâche d'identification de fréquence, de grandes variabilités sont observées, et un lien entre ces deux aptitudes pourrait peut-être exister.

3 CHAPITRE 3. Effet d'un entraînement actif sur la perception de la fréquence et le masquage informationnel

Une partie de ce travail a été présentée en tant que :

Jean, H., Demany, L. & Pressnitzer, P. (2017). Effect of active perceptual training on frequency discrimination and auditory informational masking. *Neurosciences of Music VI*, Boston, USA (Poster).

3.1 Aperçu

Le but de cette étude est d'examiner l'effet d'un entraînement audio-moteur conséquent (5h sur 5 jours) sur la résistance au masquage informationnel et sur la perception de la hauteur, avec un groupe expérimental important (N=48 participants répartis en 3 groupes). Comme dans les études préliminaires décrites au chapitre précédent, des tâches de discrimination et d'identification de fréquence ont été réalisées par les participants avant et après un entraînement audio-moteur. De plus, ici nous introduisons une tâche de masquage informationnel, avec l'hypothèse qu'un entraînement audio-moteur pourrait améliorer les capacités d'attention sélective à la fréquence. L'entraînement audio-moteur consistait pour les participants à contrôler, de façon tactile, la hauteur d'un son pur en temps réel pour trouver la position associée à la fréquence la plus élevée, comme dans le chapitre précédent. De plus, dans certains cas, un nuage de sons purs distracteur plus ou moins dense était présent en plus du son contrôlé par les participants. Trois groupes ont été constitué. Le premier groupe, principal, réalisait les pré- et post-test avant et après l'entraînement audio-moteur actif. Dans un second groupe, les participants ont été exposés aux essais du premier groupe de manière à distinguer la stimulation auditive passive de l'utilisation active de la boucle audio-motrice. Enfin, un dernier groupe contrôle a réalisé uniquement les pré- / et post-tests. Les résultats ont montré une amélioration des performances aux trois tâches pré-/post-tests statistiquement équivalente pour les trois groupes, suggérant un apprentissage perceptif associé aux tâches elles-mêmes et non à l'entraînement audio-moteur, avec toutefois une tendance à l'amélioration spécifique plus importante pour le groupe actif si l'on considère les participants avec les plus faibles performances initiales.

3.2 Introduction

Le but de cette étude est d'évaluer l'efficacité d'un entraînement audio-moteur à la fois sur la discrimination et l'identification de fréquence, comme dans le chapitre précédent, avec un groupe expérimental plus large (3 sous-groupes de 16 participants chacun) et une durée d'entraînement conséquente (5h sur 5 jours successifs). De plus, nous testons aussi l'effet d'un entraînement audio-moteur sur une tâche qui pourrait recruter l'attention sélective, le masquage informationnel. L'hypothèse est que la production sonore, opposée à la perception passive, peut être utilisée pour produire un apprentissage perceptif performant.

Qu'il s'agisse d'écouter une conversation dans un environnement bruyant ou d'isoler une mélodie dans une pièce musicale, il est capital de pouvoir focaliser son attention sur une source d'intérêt tout en ignorant le reste de l'environnement sonore. Dans le cas de patients malentendants appareillés ou porteurs d'implants cochléaires, cette aptitude peut faire défaut et rendre la perception auditive difficile dès que l'environnement n'est pas tout à fait silencieux. Les enjeux de pouvoir améliorer cette aptitude sont donc importants.

Dans le cadre de tâche psychophysiques, un paradigme qui semble recruter l'attention sélective est celui du masquage informationnel. Dans ces tâches, les participants doivent détecter une cible (généralement un son pur ou une séquence de sons purs) en présence d'autres sons distracteurs, le masqueur. Le masquage informationnel se distingue du masquage énergétique car le masqueur comporte une région fréquentielle protégée autour de la cible. La cible et le masqueur sont ainsi traités par des filtres cochléaires différents. Dans ce cas, l'observation d'une élévation de seuil ne peut être attribuée au masquage énergétique classique.

De grandes différences inter-individuelles sont observées dans les performances à ce type de tâche de masquage informationnel (Durlach et al., 2003a, 2003b; Kidd et al., 2003; Neff & Dethlefs, 1995; Richards & Neff, 2004). Mais est-il possible d'améliorer cette capacité d'attention sélective chez les participants qui ont des difficultés ?

Peu de travaux ont systématiquement étudié les effets d'apprentissage de l'attention sélective auditive. Dans les études où l'apprentissage n'est pas le but premier, il est difficile de tirer des conclusions car les participants sont souvent largement pré-entraînés. Kidd et al. (1994) utilisent 4800 (expérience 1) ou 2400 (expérience 2) essais d'entraînement avant le début de l'expérience. Neff & Callaghan (1988) ont regardé l'évolution des performances des participants mais un entraînement préalable d'environ 10 heures a été réalisé (masquage avec bruit blanc ou sinusoïde). Les travaux

de Neff & Dethlefs (1995) pourraient laisser penser qu'il y a peu d'apprentissage (Kidd et al., 2008; Oxenham et al., 2003), mais les sujets étaient également entraînés avant de commencer la tâche (600 essais). Enfin, Richards et al. (2004) entraînent les sujets pendant environ 2 ou 3 heures avant le début de l'expérience et les sujets font 300 essais avant le début de chaque condition. Il est donc difficile d'examiner les effets d'apprentissage dans ces études.

Comme nous l'avons vu dans la partie théorique (1.3.4), Jones et al. (2014) ont directement évalué les effets d'entraînement sur une tâche de masquage informationnel similaire à celle de Neff & Callaghan (1988), mais sans que les participants soient pré-entraînés. Une amélioration significative des performances a été observée et cet effet disparaît lorsque l'on enlève les 600 premiers essais (équivalent au pré-entraînement de Neff & Callaghan, 1988). Ceci suggère qu'un apprentissage rapide de l'attention sélective se produit.

Par ailleurs, les résultats obtenus par Oxenham et al. (2003), qui ont comparé musiciens et non-musiciens sur une tâche de masquage informationnel, suggèrent un avantage des musiciens. Une hypothèse est que les musiciens, par leur pratique, sont entraînés à isoler des voix mélodiques au sein de musiques polyphoniques.

Dans ce chapitre, nous présentons un protocole d'entraînement élaboré à partir de la méthode présentée dans le chapitre précédent. Les caractéristiques voulues de ce protocole sont l'utilisation de la boucle audio-motrice, une difficulté croissante, un renforcement régulier par le biais de points, une implication supplémentaire grâce à une limite de temps. Nous avons choisi une tâche sur tablette dans laquelle les participants déplaçaient leur doigt sur l'écran pour contrôler la fréquence d'un son pur. Leur but était de trouver la position associée au son le plus aigu. Les participants devaient donc garder leur attention sur les changements de fréquences associés à leur action motrice. La tâche était organisée sous forme de niveaux de difficulté croissante. Un renforcement permettant une motivation supplémentaire a également été mise en place sous forme de points à remporter dans chaque niveau. Enfin, la pression temporelle étant une caractéristique importante de la pratique musicale ou des jeux vidéo, une limite de temps a été instaurée.

Trois groupes ont été définis pour tester si un tel entraînement était associé à une amélioration des performances. Les participants du groupe MOTEUR ont réalisé les tâches de pré-test (discrimination de fréquence, identification de fréquence et masquage informationnel), puis l'entraînement actif décrit ci-dessus, puis le post-test (identique au pré-test). De manière à distinguer la stimulation auditive passive de l'utilisation de la boucle audio-motrice, les participants du groupe EXPOSITION ont

été passivement exposés aux essais réalisés par les participants du groupe MOTEUR. Ils ont également réalisé les pré- et post-tests. Enfin, les participants du groupe CONTRÔLE n'ont réalisé que les pré- et post-tests.

De manière à évaluer d'éventuelles améliorations de la perception de la fréquence, les seuils de discrimination et d'identification de fréquence des participants ont été mesurés à l'aide de la tâche à double paire décrite dans le chapitre précédent (Mathias et al., 2010; Semal & Demany, 2006).

Pour évaluer la capacité des participants à focaliser leur attention sur la bande de fréquence du stimulus cible et à ignorer les autres fréquences, une tâche de masquage informationnel (Gutschalk et al., 2008) a été utilisée. Le but était de détecter un son pur répété régulièrement et présenté au sein d'un nuage de sons purs aléatoire en temps et en fréquence.

Nous avons vu dans la partie théorique que l'apprentissage perceptif peut être influencé par la présence d'un *roving* des valeurs du stimulus. Par exemple, le *roving* semble avoir un effet délétère sur l'apprentissage de la perception de la hauteur (Amitay et al., 2005). En outre, dans l'expérience 2 du chapitre précédent, nous n'avons pas observé d'apprentissage malgré l'absence de *roving*. Dans ce chapitre, les fréquences standard utilisées ne changent pas d'un essai à l'autre (pas de *roving*). Cependant, les participants réalisant un entraînement (groupes MOTEUR et EXPOSITION) ont été entraînés soit avec une fréquence standard de 515 Hz soit avec une fréquence standard de 1099 Hz. De manière à évaluer le transfert de l'apprentissage perceptif d'une fréquence à l'autre, les participants ont été testés avec les deux fréquences standards (515 Hz et 1099 Hz) pour chacune des trois tâches des pré-/post-tests.

3.3 Méthodes

3.3.1 Participants

Quarante-huit auditeurs normo-entendants ont participé à l'expérience, répartis en 3 groupes expérimentaux de 16 participants chacun (8 femmes et 8 hommes dans chaque groupe ; âge en années pour chaque groupe respectivement $M = 26.4$, $SD = 3.54$; $M = 25.8$, $SD = 4.30$; $M = 26$, $SD = 3.27$). Ils étaient indemnisés 10 euros de l'heure pour participer à cette étude. Un audiogramme préalable permettait de s'assurer que tous les participants avaient une audition normale, c'est à dire des seuils auditifs de 20 dB HL ou moins pour les fréquences testées (125, 250, 500, 1000,

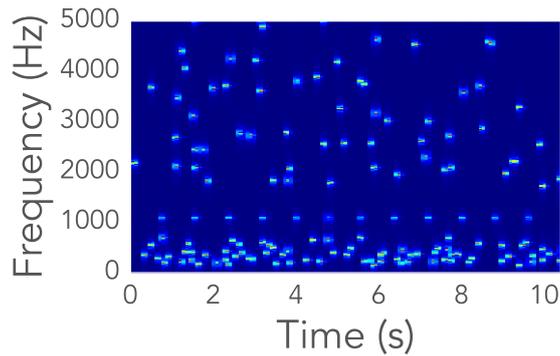


Figure 3.1: Procédure - Pré/-post-tests - Masquage informationnel. Spectrogramme d'un essai avec cible de la tâche de masquage informationnel. Dans cet exemple, la fréquence de la cible est de 1099 Hz. On peut noter la présence de la région protégée autour de la fréquence de la cible.

2000, 3000, 4000, et 8000 Hz). Tous les participants étaient non-musiciens (moins de deux ans de pratique musicale au moment de l'étude).

3.3.2 Stimuli

3.3.2.1 Discrimination et identification de fréquence

Tous les sons utilisés étaient des sons purs de 250 ms (incluant des rampes ascendante et descendante de 10 ms chacune) joués avec un intervalle inter-stimulus (ISI) de 250 ms. Le son pur test était toujours de fréquence plus élevée que le son pur standard. La différence de fréquence entre test et standard (Δf en cents) était modifiée de manière adaptative.

3.3.2.2 Masquage informationnel

La tâche de masquage informationnel a été inspiré par Gutschalk et al. (2008). Les essais contenaient un nuage de tons en guise de masqueur (en gris sur la Figure 3.2). Une cible (en noir sur la Figure 3.2) était également présente dans 75% des essais (Gutschalk et al., 2008). La Figure 3.1 montre un exemple de spectrogramme d'un essai avec cible. Bien que la cible soit clairement visible sur le spectrogramme, la tâche est assez difficile du fait du masquage informationnel. La présence du nuage de tons rend la cible difficile à détecter malgré un niveau de cette cible bien au-dessus du seuil de détection dans le silence.

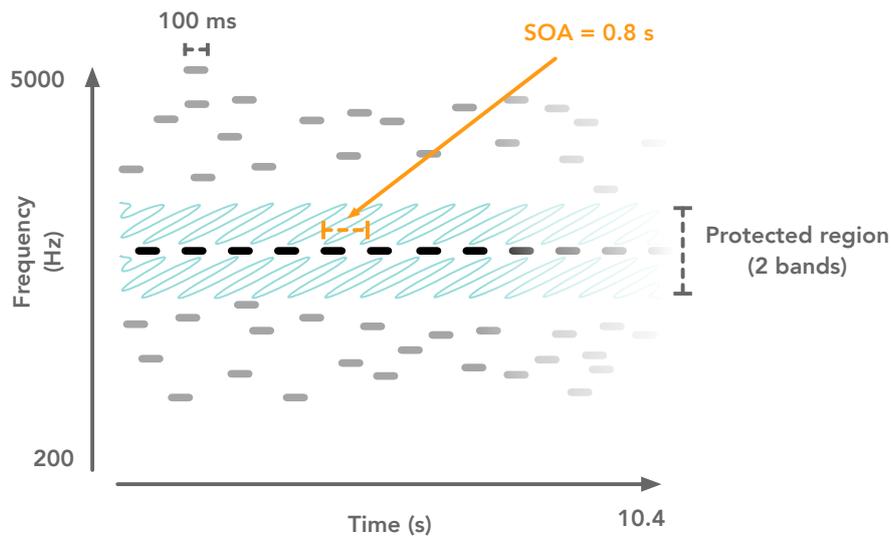


Figure 3.2: Procédure - Pré/-post-tests - Masquage informationnel. Représentation de la cible (en noir) et des sons purs du nuage de tons (en gris). La région protégée en fréquence est illustrée en bleu.

Cibles

Les cibles étaient des sons purs de 100 ms (incluant des rampes ascendantes et descendantes de 10 ms chacune) répétés avec un SOA (durée entre le début d'un son pur et le début du suivant) de 800 ms et commençant 800 ms après le début du masqueur. Un exemple de cible est représenté en noir sur la Figure 3.2.

Masqueurs

Les masqueurs étaient des nuages de tons construits à partir d'un ensemble de sons purs de 100 ms avec des caractéristiques spectrales et temporelles aléatoires et organisés sous la forme de 13 voix simultanées (ou composants). Pour chaque composant, la fréquence de chaque son pur était choisie à partir de 18 fréquences centrales réparties régulièrement sur une échelle logarithmique de 200 Hz à 5000 Hz (fréquences centrales: 200, 242, 292, 353, 427, 515, 623, 753, 910, 1099, 1328, 1605, 1940, 2344, 2833, 3424, 4138, 5000 Hz). Deux fréquences centrales au-dessous et deux au-dessus de la fréquence de la cible étaient exclues donnant lieu à une région protégée autour de la fréquence de la cible (illustrée en bleu sur la Figure 3.2). Pour chaque son pur du nuage, une fréquence centrale était choisie parmi les 13 possibilités restantes. Ainsi, pour la condition dans laquelle la cible était de 515 Hz, les fréquences centrales possibles étaient 200, 242, 292, 910, 1099, 1328, 1605, 1940, 2344, 2833,

3424, 4138, ou 5000 Hz. Pour la condition où la cible était de 1099 Hz, elles étaient de 200, 242, 292, 353, 427, 515, 623, 1940, 2344, 2833, 3424, 4138, ou 5000 Hz. La fréquence exacte de chaque son pur était ensuite choisie aléatoirement à partir d'une distribution uniforme de la largeur de deux ERB (*Equivalent Rectangular Bandwidth*; Moore, 1995) autour de la fréquence centrale.

$$ERB = 24.7 \times (4.37 \times f_c + 1)$$

avec f_c en kHz. Le SOA était également choisi aléatoirement à chaque son pur dans une distribution uniforme entre 100 ms et 1600 ms. Une représentation schématique d'un nuage de tons obtenu selon cette méthode est montré en gris sur la Figure 3.2.

3.3.2.3 Tâche audio-motrice

Entre les pré- et post-tests, les participants du groupe MOTEUR ont réalisé une tâche audio-motrice. Comme dans celle du chapitre 1, les participants devaient contrôler un son pur (*ton contrôlé*) changeant de fréquence selon la position du doigt sur la tablette. Dans certaines conditions, un nuage de tons était également présent, présentant les mêmes caractéristiques que les masqueurs de la tâche des pré-/post-tests, mais de densité variable selon le niveau de difficulté.

Les stimuli utilisés étaient différents selon le niveau de difficulté. Dans les niveaux 0 à 4, les participants contrôlaient la fréquence d'un son pur continu. Dans les niveaux 5 à 15, le son contrôlé était pulsé. Il s'agissait de sons purs de 100 ms (incluant des rampes ascendantes et descendantes de 10 ms chacune) répétés avec un SOA constant de 800 ms, comme la cible dans les pré- et post-tests.

Les niveaux 0 et 5 ne comportaient pas de masqueur, permettant ainsi au sujet de se familiariser avec la tâche utilisant des sons continus et pulsés. Dans les autres niveaux, des nuages de tons du même type que ceux décrits dans la tâche de masquage informationnel ont été utilisés. La combinaison de deux paramètres a été utilisée pour moduler la difficulté de la tâche. Tout d'abord, la densité du nuage modifiée en changeant le nombre de composants (4, 8, 13 ou 17 composants, sachant que le nuage de tons de la tâche de masquage comportait 13 composants). L'intensité totale du nuage de tons était pondérée de manière à obtenir la même intensité moyenne du masqueur dans les différentes conditions ($gain_{total} = \frac{gain}{n_{comp}}$, avec n_{comp} le nombre de composants). Ensuite, la cible pouvait prendre différents niveaux relatifs au gain du masqueur (-8 dB, -16 dB, -24 dB).

La position du son pur cible sur l'axe y , c'est-à-dire le son le plus aigu (515 ou 1099 Hz selon la condition) était choisie aléatoirement sur l'ensemble de l'aire de jeu en excluant une zone fixe de manière à assurer la présence d'un sommet d'où la fréquence décroît de chaque côté.

3.3.2.4 Tâche exposition

L'ensemble des paramètres des essais des participants du groupe MOTEUR a été enregistré (position du doigt sur la tablette, stimuli auditifs associés, déplacement visuel du curseur sur l'écran. . .). Pour le groupe EXPOSITION, les essais du groupe MOTEUR ont été rejoués intégralement. Ainsi, les participants du groupe EXPOSITION ont eu la même stimulation visuelle et auditive que ceux du groupe MOTEUR. De plus, une tâche de discrimination de bruits blancs a été ajoutée pour s'assurer de l'attention auditive des participants tout au long de l'exposition. Des bruits blancs d'une durée de 200 ms étaient filtrés autour de 1000 Hz avec un filtre passe-bande bi-quadratique de facteur $Q = 0,4083$. Les bruits blancs *ascendants* possédaient une rampe ascendante de 150 ms et une rampe descendante de 50 ms. Inversement, les bruits blancs *descendants* possédaient une rampe ascendante de 50 ms et une rampe descendante de 150 ms. L'ensemble des rampes utilisées étaient de forme \cos^2 .

3.3.3 Matériel

Les stimuli ont été générés en JavaScript par les tablettes utilisées (iPad Air 2 et Nexus 9; voir annexe 1, page 195) et présentés par le biais de casques audio (Beyerdynamic, DT 770 PRO). Tous les participants ont été testés dans une cabine insonorisée à double parois (Industrial Acoustics).

3.3.4 Procédure

L'expérience s'est déroulée sur 6 jours consécutifs (Figure 3.3) avec le pré-test le jour 1, une session d'entraînement par jour du jour 1 au jour 5 (une heure par session) et le post-test le jour 6. Les sujets ont été aléatoirement répartis en 3 groupes : CONTRÔLE, EXPOSITION et MOTEUR. Le groupe CONTRÔLE n'a réalisé que les phases pré- et post-tests. Dans le groupe EXPOSITION, les participants ont en plus réalisé une phase d'exposition aux stimuli des participants du groupe MOTEUR. Dans le groupe MOTEUR, ils ont réalisé les pré-tests, une phase d'entraînement audio-moteur puis les post-tests.

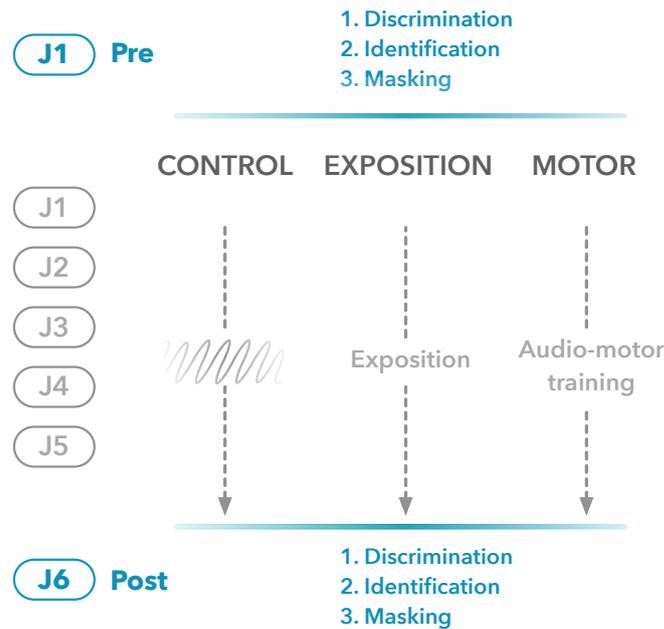


Figure 3.3: Procédure. Déroulement de l'expérience. Les tâches de pré-tests et une première session d'entraînement ont été réalisées le premier jour. Quatre autres sessions d'entraînement ont été réalisées les jours 2, 3, 4 et 5. Le post-test s'est déroulé le jour 6.

Pour tous les groupes, une procédure identique a été réalisée en pré- et post-test, à 5 jours d'intervalle. Cette phase consistait en trois tâches visant à évaluer les seuils de discrimination et d'identification de fréquence ainsi que les performances à une tâche de masquage informationnel. Ces trois tâches ont été réalisées par tous les participants dans un ordre fixe : 1. discrimination de fréquence, 2. identification de fréquence, 3. masquage informationnel. Avant chaque tâche (en pré- et post-test) quelques essais non-inclus dans les analyses ont été réalisés de manière à familiariser les sujets avec la tâche et s'assurer qu'elle a été comprise. Pour la tâche de masquage informationnel, les cibles de chaque condition (515 Hz et 1099 Hz) et les nuages de tons (1099 Hz) ont été joués seuls pour montrer aux participants ce qui devait être détecté et ignoré. Chaque participant a réalisé chacune de ces trois tâches deux fois successivement avec une fréquence standard de 515 Hz et de 1099 Hz (ordre aléatoire). En outre, les tâches de discrimination et d'identification de fréquence ont été répétées trois fois de manière à augmenter la précision de la mesure. Ainsi, chaque participant a réalisé 3 répétitions \times 2 fréquences résultant en 6 tâches de discrimination et 6 tâches d'identification.

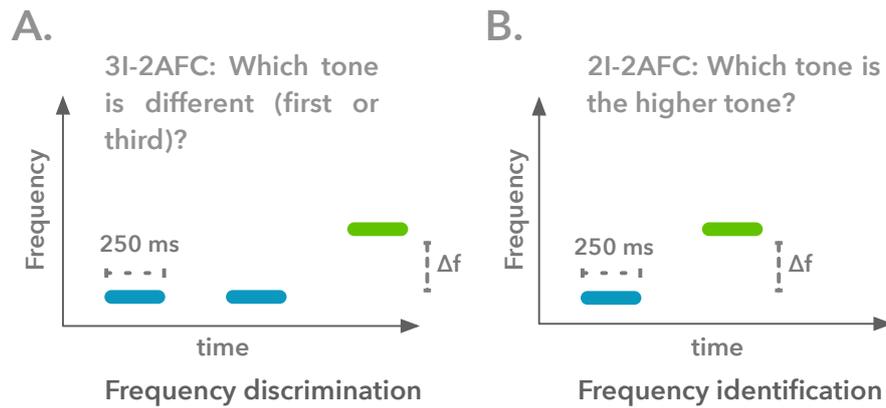


Figure 3.4: Procédure - Pré-/post-tests. Tâche de discrimination et d'identification de fréquence. **A.** Dans la tâche de discrimination de fréquence, les participants devaient choisir parmi le premier ou le dernier ton, lequel était différent du son pur central. **B.** Dans la tâche d'identification de fréquence, les participants devaient choisir le son le plus aigu parmi les deux intervalles.

3.3.4.1 Pré-/post-tests

Seuils de discrimination et d'identification de fréquence

Les tâches de discrimination et d'identification de fréquence utilisées dans ce chapitre diffèrent de la tâche à double paires du chapitre précédent. Il ne s'agit pas ici de comparer discrimination et identification, et la tâche à double paires ne semblait donc pas nécessaire. Toutefois, le principe de ces tâches est similaire et vise à évaluer les seuils de discrimination et d'identification, mais cette fois avec deux tâches indépendantes.

La Figure 3.4 montre la procédure des tâches de discrimination et d'identification. Pour ces deux tâches, la valeur correspondant à la différence de fréquence (Δf) entre les intervalles standard (en bleu) et test (en vert) a été modifiée de manière adaptative. La valeur initiale de Δf était de 100 cents (un demi-ton). Une procédure adaptative en escalier (*2-down, 1-up* visant 70.7% de réponses correctes, Levitt, 1971) a été utilisée. Jusqu'au 4^{ème} changement de direction (*reversal*) le facteur utilisé pour augmenter ou diminuer le Δf était de 2 sauf pour le 2^{ème} bloc de fréquence de la tâche de discrimination où il était de 1.25 due à un problème expérimental. Après le 4^{ème} *reversal*, ce facteur était égal à $\sqrt{2}$. La tâche s'arrêtait lorsque 9 *reversals* avaient été réalisés (au lieu de 10, également suite à une erreur expérimentale), et le seuil a été calculé en prenant la moyenne des quatre derniers *reversals*.

Pour calculer les seuils de discrimination de fréquence, une tâche de type 3I-2AFC (AXB) a été utilisée (Figure 3.4, A.). Trois sons successifs étaient présentés aux participants qui devaient indiquer si le son central (X) était identique au premier (A) ou au dernier (B). Par exemple, sur la figure, la bonne réponse correspond au dernier ton.

Pour calculer les seuils d'identification de fréquence, une tâche de type 2I-2AFC a été utilisée (Figure 3.4, B.). Deux sons de fréquences différentes étaient présentés aux participants dont la tâche était de donner l'intervalle correspondant au son le plus aigu. Par exemple, sur la figure, la bonne réponse correspond au deuxième ton.

Masquage informationnel

Pour chaque essai, la tâche des participants était de répondre à l'aide d'un bouton *cible* dès qu'ils entendaient la cible. L'essai suivant était alors automatiquement déclenché. Pour fournir une réponse négative (pas de cible), le sujet devait attendre la fin de l'essai. La longueur maximum des essais était de 10.4 s dans le cas où aucune réponse n'était fournie. Comme dans Gutschalk et al. (2008) le masqueur était présent dans la totalité des essais et la cible dans 75% des essais. Tous les participants ont été testés dans deux conditions (ordre aléatoire) où la fréquence de la cible était de 515 Hz ou de 1099 Hz organisées en blocs. Contrairement aux tâches de discrimination et d'identification de fréquence, la méthode des stimuli constants a été utilisée pour la tâche de masquage informationnel. L'intensité moyenne du masqueur était constante d'un essai à l'autre et l'intensité de la cible pouvait prendre 5 niveaux par rapport au masqueur : -8, -18, -28, -38 et -48 dB. Le gain de la cible était calculé ainsi à partir du SNR :

$$gain_{cible} = 10^{\frac{SNR}{20}} \times gain_{masqueur}$$

SNR étant le rapport de l'intensité du masqueur et de la cible et *gain_{masqueur}* le gain moyen du nuage de tons.

Au moins 30 essais par SNR et par condition (515 ou 1099 Hz) ont été réalisés par les participants (parfois plus à cause d'une erreur technique) pour un total de 300 essais en pré-test et 300 essais en post-test.

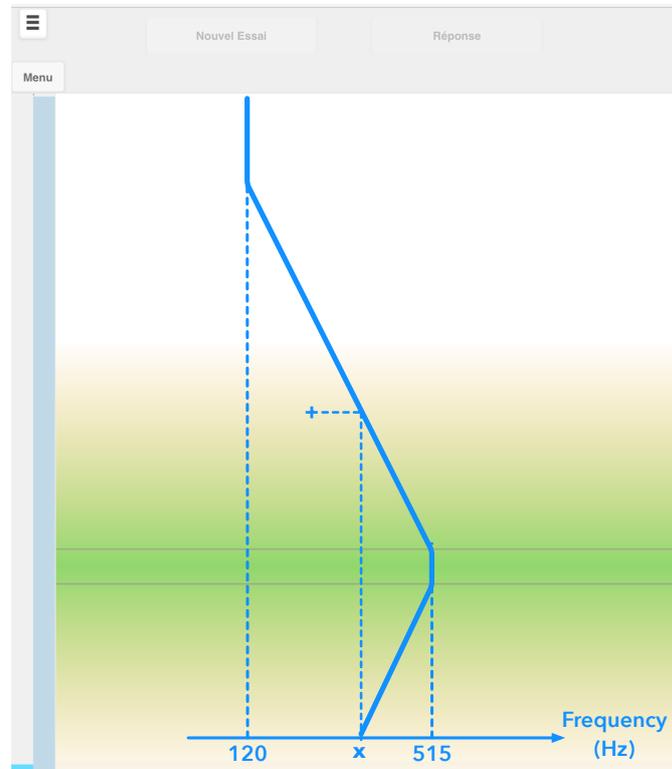


Figure 3.5: Procédure - Tâche audio-motrice. Relation entre la position spatiale sur l'écran et la fréquence pour un essai de la condition 515 Hz. Toutes les lignes bleues ont été ajoutées pour le schéma. L'image correspond à l'interface de la tâche audio-motrice après que la réponse ait été donnée. Le dégradé vert et les lignes grises correspondent au *feedback* visuel donné aux participants. Par exemple, la position de la croix bleue est associée à une fréquence de x Hz. Tous les points d'une ligne horizontale correspondent donc à la même fréquence. Seul un mouvement sur l'axe y permet un changement de fréquence. L'échelle des fréquences est linéaire car il a semblé plus aisé au cours des pilotes de contrôler précisément la hauteur de cette manière plutôt que sur une échelle logarithmique. La répartition des fréquences présente en outre un sommet plat. La réponse est considérée comme correcte lorsque la valeur y est sur ce sommet

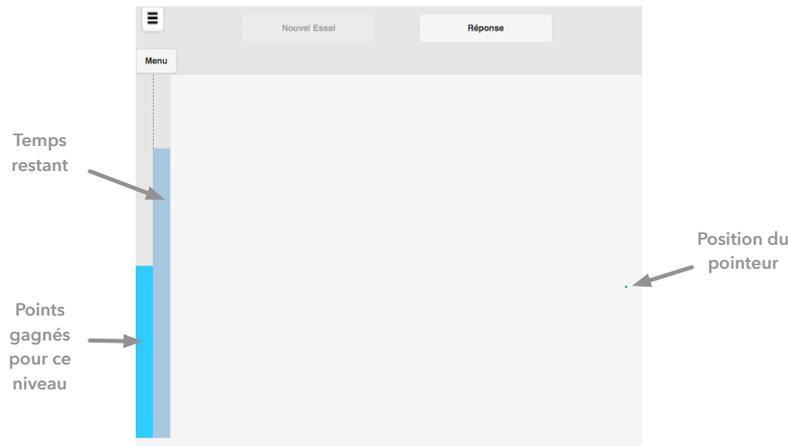


Figure 3.6: Procédure - Tâche audio-motrice. Interface de la tâche d'entraînement. Un point indique la position du curseur et les deux barres verticales correspondent à la jauge de temps (diminuant au fur et à mesure de l'essai) et de points (augmentant à chaque essai).

3.3.4.2 Entraînement audio-moteur

Les 16 participants du groupe MOTEUR ont suivi un entraînement de 1h/jour en laboratoire pendant 5 jours consécutifs. Un protocole sur tablette a été réalisé dans lequel les participants contrôlaient la fréquence d'un son pur continu en déplaçant leur doigt sur l'écran. A chaque essai, ils devaient trouver la position sur l'écran associée à la fréquence la plus élevée. Cette fréquence maximum était constante pour l'ensemble des essais de l'entraînement. Alors que l'ensemble des participants a réalisé les tâches pré-/post-tests dans les deux conditions (515 et 1099 Hz), seulement la moitié d'entre eux a réalisé l'entraînement dans la condition 515 Hz et l'autre moitié dans la condition 1099 Hz.

Cette tâche d'entraînement était organisée sous forme d'essais d'une durée maximale de 10 secondes. Le temps restant était indiqué aux participants grâce à une barre verticale descendante sur le côté gauche de l'écran. Un son n'était produit et le temps décompté que lorsque l'écran était touché. Ainsi, l'essai était mis en pause lorsque le participant enlevait le doigt de l'écran. La réponse pouvait être donnée à l'aide d'un bouton *Réponse* en haut de l'écran (Figure 3.6). Si aucune réponse n'était donnée au terme des 10 secondes, l'essai se terminait automatiquement et la réponse était la dernière position enregistrée.

Une fois l'essai terminé, un *feedback* visuel était donné aux participants sous la forme d'un gradient linéaire de couleur représentant la fréquence du son (Figure 3.5).

Pour chaque pixel, la différence entre la fréquence correspondante à la position du pixel et la fréquence cible (la fréquence la plus haute : 515 ou 1099 Hz selon la condition) était convertie en couleur à l'intérieur de ce gradient. En outre, deux lignes horizontales étaient dessinées déterminant les bornes de la réponse considérée comme correcte.

Une réponse correcte rendait le bouton *Essai suivant* actif permettant au sujet de passer à l'essai suivant. Dans le cas d'une réponse incorrecte, le sujet devait déplacer son doigt depuis la position de sa réponse jusqu'à une position correcte à l'aide du *feedback* visuel pour rendre le bouton *Essai suivant* actif. Le but était que le participant entende la différence de fréquence entre sa réponse et la réponse correcte.

Cette phase d'entraînement était constituée de 16 niveaux de difficulté croissante, chacun composé d'un nombre d'essais dépendant de ses performances. Lorsque les 16 niveaux étaient réussis, le participant recommençait au niveau 1. Pour terminer chaque niveau, un score de 100000 points devaient être récoltés. Cela correspondait à remplir entièrement la barre de points (Figure 3.6). Dans le cas d'un essai incorrect (réponse en-dehors des deux lignes horizontales), un minimum de 1000 points était tout de même remporté, pour ne pas décourager le participant. Dans le cas d'un essai correct, le nombre de points gagnés dépendait de la quantité de temps restant à raison de 1000 points par seconde restante.

3.3.4.3 Tâche d'exposition

Chacun des 16 participants du groupe EXPOSITION a été exposé aux essais réalisés par un participant du groupe MOTEUR. Un essai rejoué correspond à la reproduction intégrale des déplacements du curseur sur la tablette par le participant ainsi que le son produit par ce déplacement. Les nuages de tons ont également été reproduit dans les niveaux utilisant ces masqueurs de manière à obtenir une stimulation sonore identique à celle d'un participant du groupe MOTEUR. L'association par paires des participants des groupes MOTEUR et EXPOSITION a été réalisée aléatoirement.

Pour s'assurer que les participants maintenaient une attention sur la stimulation sonore même pendant la phase d'exposition passive, une tâche auditive non reliée à la fréquence ou à l'attention sélective a été introduite. Il s'agissant d'une tâche de discrimination de l'enveloppe temporelle de bruits blancs, ces bruits étant présentés en plus des sons d'exposition. Dans certains essais, un bruit blanc caractérisé par une rampe d'amplitude ascendante ou descendante était entendu en plus des stimuli issus du participant du groupe MOTEUR. Les sujets devaient fournir une réponse

à la fin de chaque essai de manière à indiquer la présence et le type de bruit blanc entendu (*bruit blanc ascendant, bruit blanc descendant, pas de bruit blanc*).

3.3.5 Analyses des données

3.3.5.1 Discrimination et identification de fréquences

Les seuils ont été calculés en cents à partir des 4 derniers *reversals* pour chaque sujet, pour chaque test (pré et post) et pour chaque condition (fréquence identique ou différente de l'entraînement). Ils ont ensuite été convertis en pourcentages :

$$seuil_{pourcentage} = (2^{\frac{seuil_{cents}}{1200}} - 1) \times 100$$

Des tests statistiques ont été réalisés sur les seuils en pourcentage (convertis sur une échelle logarithmique) à l'aide d'ANOVA à modèles mixtes avec le facteur pré-/post (pré ou post) comme facteur intra-participant et le facteur groupe (CONTRÔLE, EXPOSITION ou MOTEUR) comme facteur inter-participants. L'homogénéité des variances a été vérifiée et corrigée lorsque nécessaire à l'aide de tests de Levene.

3.3.5.2 Masquage informationnel

Les réponses ont été converties en pourcentages de détections correctes et de fausses alarmes et l'index de sensibilité d' a été calculé pour chaque participant, pour chaque test (pré ou post), pour chaque fréquence de la cible (515 Hz ou 1099 Hz) et pour chaque SNR (-8, -18, -28, -38 et -48 dB). Lorsqu'une valeur de détections correctes ou de fausses alarmes était de 0% ou de 100%, 0.5 essai a été ajouté ou retiré de manière à rendre possible le calcul du d' . Des tests statistiques ont été réalisés sur les valeurs de d' à l'aide d'ANOVA à modèles mixtes avec le facteur pré-/post (pré ou post) comme facteur intra-participant et le facteur groupe (CONTRÔLE, EXPOSITION ou MOTEUR) comme facteur inter-participants. L'homogénéité des variances a été vérifiée lorsque nécessaire à l'aide de tests de Levene.

Lorsque deux conditions ont été comparées, des tests t de Student ont été utilisés si l'hypothèse de normalité de chaque échantillon était respectée. Dans le cas contraire, des tests de Wilcoxon (*Wilcoxon signed ranks tests*) ont été utilisés à la place.

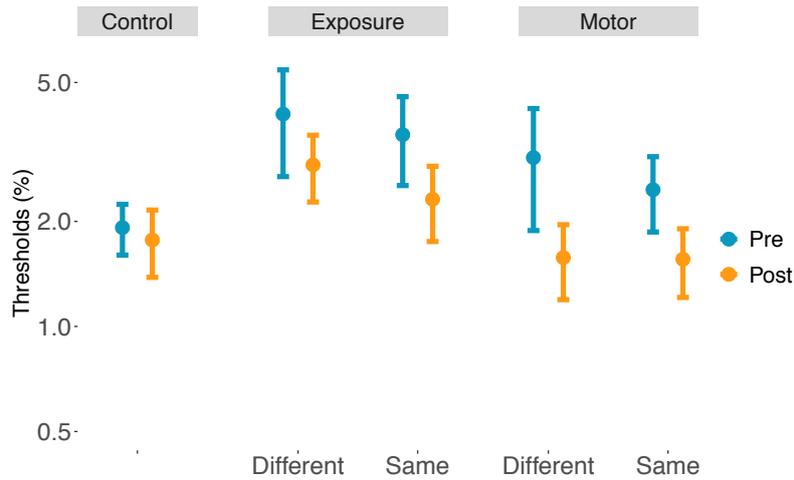


Figure 3.7: Pré/-post-tests - Discrimination de fréquence. Seuils de discrimination de fréquence aux pré- (en bleu) et post-tests (en orange) pour les groupes CONTRÔLE, EXPOSITION et MOTEUR. Les performances pour les fréquences identiques et différentes entre entraînement et test sont représentées. Les barres d'erreur représentent l'erreur standard.

3.4 Résultats

3.4.1 Pré-/post-test

3.4.1.1 Discrimination de fréquence

La Figure 3.7 montre les seuils de discrimination de fréquence pour les pré- et post-tests des participants des groupes CONTRÔLE, EXPOSITION et MOTEUR. Pour les groupes EXPOSITION et MOTEUR, les résultats pour la fréquence identique à celle utilisée dans l'entraînement ou l'exposition (*same*) et pour la fréquence opposée (*different*) sont présentés. Le groupe CONTRÔLE ne comporte pas la distinction entre fréquences identiques et différentes car les participants ont uniquement réalisé les pré- et post-tests.

Il ne semble pas qu'il y ait de différence de performance entre les conditions *same* et *different*. Les tests statistiques n'ont en effet pas révélé d'effet de ce facteur [$F(1, 30) = 0.20371, p > 0.05, \eta_g^2 = 0.00035597$]. Les données ont ainsi été moyennées pour la suite des analyses.

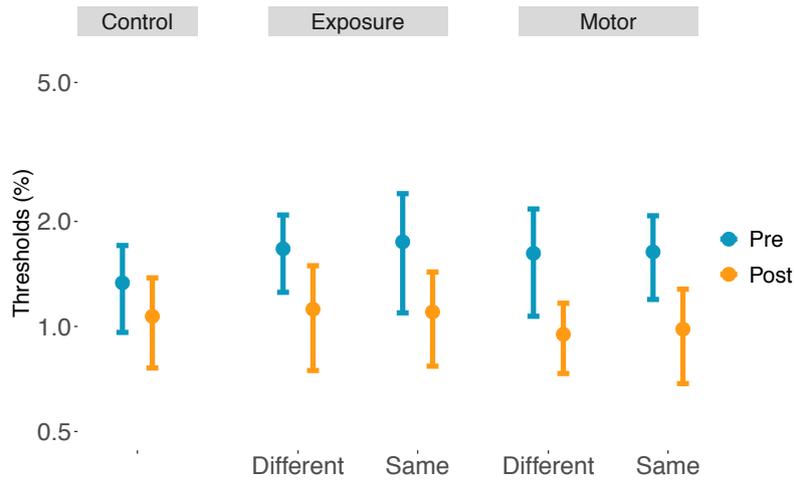


Figure 3.8: Pré-/post-tests - Identification de fréquence. Seuils d'identification de fréquence aux pré- (en bleu) et post-tests (en orange) pour les groupes CONTRÔLE, EXPOSITION et MOTEUR. Les performances pour les fréquences identiques et différentes entre entraînement et test sont représentées. Les barres d'erreur représentent l'erreur standard.

On voit que les seuils post-tests sont plus bas que les seuils pré-tests pour les trois groupes. Il semble également que cette différence est moins grande pour le groupe CONTRÔLE que pour les groupes EXPOSITION et MOTEUR. Les tests statistiques ont confirmé un effet global du facteur pré-/post-tests [$F(1, 45) = 16.8976, p < 0.001, \eta_g^2 = 0.0258464$], mais pas d'influence du groupe sur cet effet [$F(2, 45) = 0.2648, p > 0.05, \eta_g^2 = 0.0008308$].

3.4.1.2 Identification de fréquence

La Figure 3.8 présente les seuils d'identification de fréquence pour les pré- et post-tests des participants des trois groupes.

Comme pour la tâche de discrimination, il ne semble pas qu'il y ait de différence de performance entre les conditions *same* et *different*. Les tests statistiques n'ont en effet pas révélé d'effet de ce facteur [$F(1, 30) = 0.08267, p > 0.05, \eta_g^2 = 0.00019369$]. Les données ont ainsi été moyennées pour la suite des analyses.

Ensuite, les seuils sont généralement plus bas aux post-tests par rapport aux pré-tests. Les tests statistiques ont à nouveau confirmé un effet du facteur pré-/post-tests

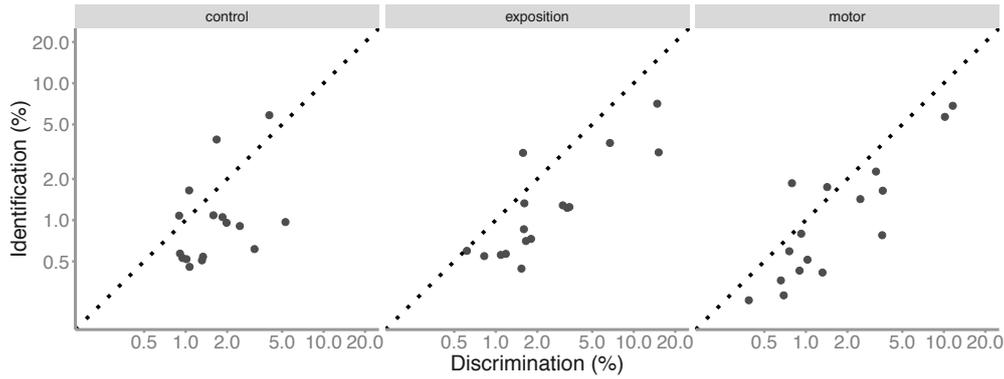


Figure 3.9: Pré/-post-tests - Discrimination vs. identification de fréquence. Comparaison des seuils de discrimination (axe x) et d'identification (axe y) de fréquence pour les trois groupes de participants.

$[F(1, 45) = 40.41187, p < 0.001, \eta_g^2 = 0.058835]$, mais pas d'influence du groupe sur cet effet $[F(2, 45) = 0.38416, p > 0.05, \eta_g^2 = 0.001187]$.

3.4.1.3 Comparaison entre discrimination et identification de fréquence

La Figure 3.9 montre les seuils d'identification de fréquence en fonction des seuils de discrimination de fréquence pour les participants des groupes MOTEUR, EXPOSITION et CONTRÔLE. Les points sous la ligne diagonale correspondent à de meilleures performances d'identification. On voit donc que les performances sont généralement meilleures pour la condition identification par rapport à la discrimination de fréquence. Un test t de Student a révélé une différence significative $[t(47) = 8.4, p < 0.001]$ entre les performances de discrimination (2.383%) et d'identification (1.293%) de fréquence. Cela s'oppose à ce qui observé dans la littérature et dans le chapitre précédent et pourrait s'expliquer par l'ordre fixe des tâches réalisée (cf. discussion).

3.4.1.4 Masquage informationnel

La Figure 3.10 montre la sensibilité (en d') aux pré- et post-tests des participants des groupes EXPOSITION et MOTEUR pour des fréquences identiques ou différentes de la fréquence d'entraînement. Les participants du groupe CONTRÔLE ne sont pas inclus, car n'ayant réalisé que les pré- et post-tests, il n'avaient pas de fréquence entraînée. La figure suggère qu'il n'y pas d'effet de la fréquence d'entraînement sur la différence pré-/post-tests (l'écart entre pré- et post-test semble équivalent

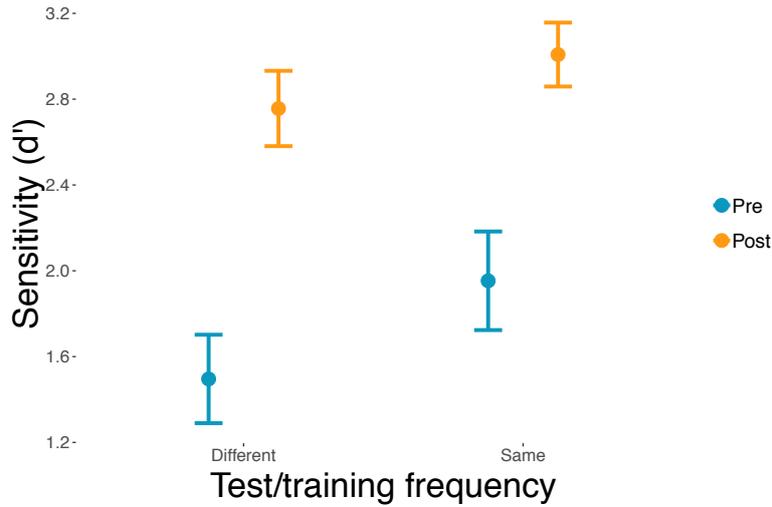


Figure 3.10: Pré-/post-tests - Masquage informationnel. Comparaison des performances à la tâche de masquage informationnel pour les pré-/post-tests en fonction de la fréquence d’entraînement (identique ou différente de la fréquence test). Seuls les groupes EXPOSTION et MOTEUR pour lesquels un entraînement a été réalisé sont représentés et leurs performances ont été moyennées. Les barres d’erreurs représentent l’erreur standard.

dans les deux conditions). Des tests statistiques ont confirmé l’absence d’interaction significative entre les facteurs pré-/post-tests et la fréquence d’entraînement identique ou différente de la fréquence test [$F(1, 31) = 0.77, p > 0.05, \eta_g^2 = 0.0021$]. Les conditions fréquences test/entraînement identiques et différentes ont donc été agrégées pour la suite des analyses.

On voit sur la Figure 3.11 les valeurs de d' moyennées selon le groupe (CONTRÔLE, EXPOSITION ou MOTEUR), le test (pré ou post) et le SNR (-8, -18, -28, -38 ou -48 dB). Comme prévu, la sensibilité est plus élevée lorsque le niveau de la cible était plus élevé (haut SNR). La condition -48 est associée à des performances proche du niveau du hasard pour l’ensemble des groupes et tests. On voit également que de meilleures performances caractérisent le post-test en comparaison du pré-test, et cela pour l’ensemble des groupes expérimentaux. Les tests statistiques ont révélé un effet du facteur pré-/post-tests [$F(1, 45) = 78.126, p < 0.001, \eta_g^2 = 0.20$] et un effet du groupe [$F(2, 45) = 3.523, p < 0.038, \eta_g^2 = 0.12$] mais pas d’interaction entre les deux facteurs [$F(2, 45) = 2.251, p > 0.05, \eta_g^2 = 0.014$]. Cette interaction est le test critique pour déterminer un avantage spécifique de certaines groupes par rapport à d’autres.

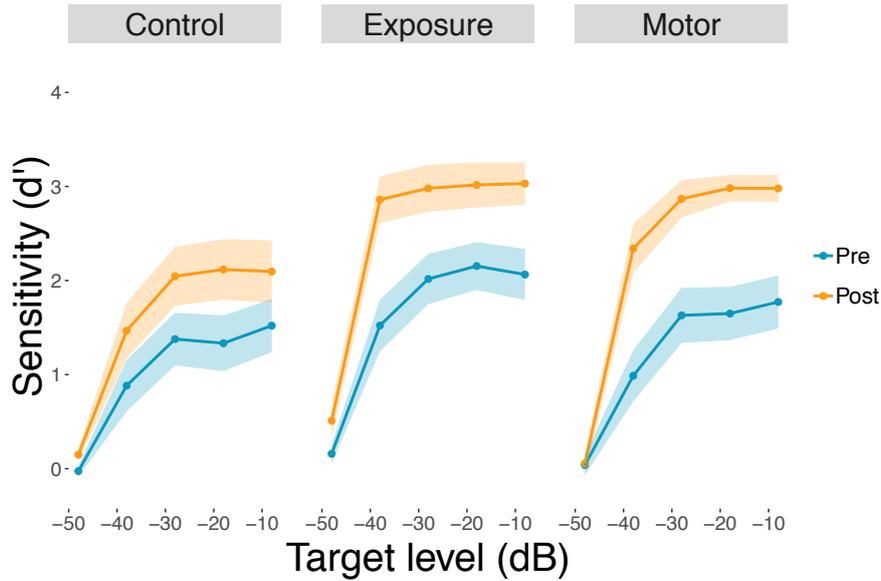


Figure 3.11: Pré-/post-tests - Masquage informationnel. Sensibilité moyenne (en d') des trois groupes ($N=16$ par groupe) à la tâche de masquage informationnel pour les pré- et post-tests. Les rubans représentent l'erreur standard.

Participants aux performances faibles

Comme souligné par d'autres auteurs, une mesure des effets d'entraînement peut être rendue moins sensible par des participants démontrant de bonnes performances aux pré-tests, avant tout entraînement (Amitay et al., 2005). Nous avons donc répété les analyses précédente sur un sous-groupe de participants ayant de faibles performances au pré-tests. Nous avons sélectionné le tiers des participants présentant la sensibilité la plus faible au pré-test de la tâche de masquage informationnel, tous groupes confondus (CONTRÔLE : 7, EXPOSITION : 4, MOTEUR : 5). Les résultats sont illustrés sur la Figure 3.12. On voit une différence plus marquée entre pré- et post-tests selon le groupe, avec un écart pré-/post plus grand pour le groupe MOTEUR. Des tests statistiques ont montré un effet pré-/post significatif [$F(1, 13) = 27.759$, $p < 0.001$, $\eta_g^2 = 0.5048$]. Un effet d'interaction significatif a cette fois été observé [$F(2, 13) = 4.007$, $p = 0.044$, $\eta_g^2 = 0.2274$], mais cet effet disparaît après correction de la valeur p (multiplié par deux car il s'agit du second test statistique réalisé sur ces participants).

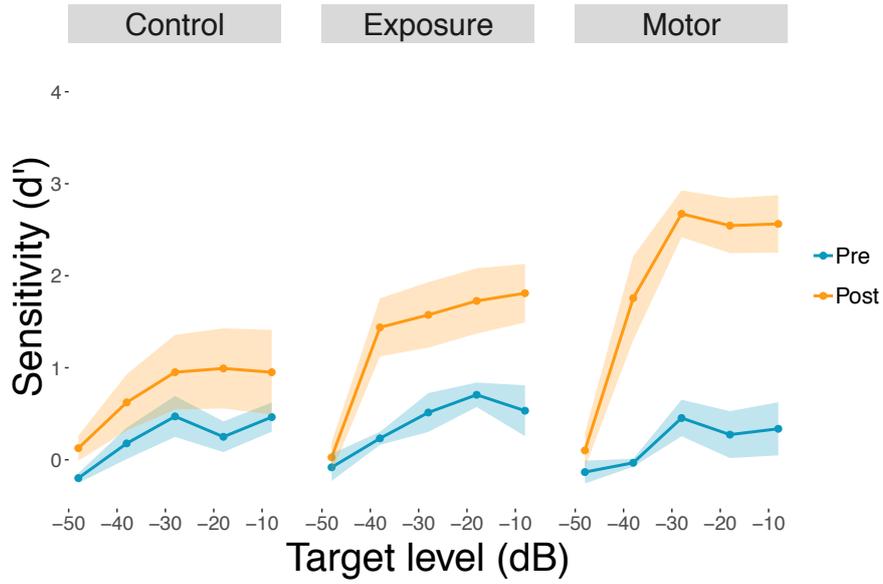


Figure 3.12: Pré/-post-tests - Masquage informationnel. Sensibilité moyenne (en d') des trois groupes à la tâche de masquage informationnel pour les pré- et post-tests uniquement pour les participants aux performances faibles. Les rubans représentent l'erreur standard.

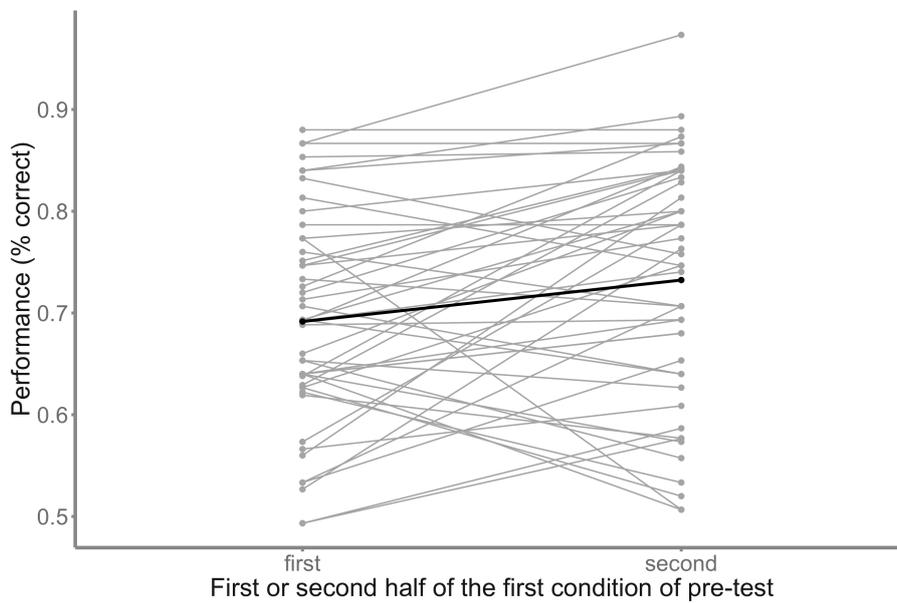


Figure 3.13: Pré/-post-tests - Masquage informationnel. Performance (% correct) à la première et deuxième moitié de la première condition (515 ou 1099 Hz) du pré-test pour chaque participant.

Performance à la première condition du pré-test

De manière à évaluer l'apprentissage lié à la réalisation de la tâche de masquage informationnel elle-même, les performances de la première moitié de la première condition (515 Hz ou 1099 Hz) du pré-test ont été comparées avec les performances de la seconde moitié de la première condition du pré-test. Cela revient à comparer les 75 premiers essais avec les 75 suivants pour chaque participant. La Figure 3.13 illustre cette analyse en montrant les performances (% correct) pour la première et la deuxième moitié de la première condition du pré-test pour chaque participant (gris) et en moyenne (noir). Les tests statistiques ont montré une différence de performance légère mais significative [$t(47) = 2.7, p = 0.01$] entre la première (69.16%) et la deuxième moitié (73.24%) de la tâche de pré-test.

Corrélations entre les tâches des pré-/post-tests

Le calcul des corrélations entre les performances des différentes tâches a montré que les tâches de discrimination et d'identification de fréquence sont fortement corrélées ($r = 0.78$). Cependant, la tâche de masquage informationnel témoigne d'une corrélation faible avec la tâche d'identification ($r = 0.23$) et la tâche de discrimination ($r = 0.21$).

Détection de la cible après sa première occurrence

La Figure 3.14 montre la distribution des temps de réponse des participants et les occurrences de la cible (lignes verticales). On voit que la majorité des réponses ont été effectuées après que la première occurrence de la cible ait été entendue (première ligne verticale à 0.8 seconde). Un second pic est visible après la deuxième occurrence de la cible (deuxième ligne verticale à 1.6 seconde).

3.4.2 Tâche audio-motrice

La Figure 3.15 montre les performances des participants du groupe MOTEUR à la tâche d'entraînement audio-motrice. Les performances ont été calculées pour chaque niveau et chaque participant en prenant la proportion d'essais réalisés (cela correspond donc potentiellement à un nombre d'essais différent selon les conditions et participants). On voit une tendance à la baisse malgré un éventuel effet plafond pour les niveaux les plus faciles. Les tests statistiques (*Wilcoxon signed ranks tests* car la distribution des données n'était pas normale) ont confirmé une différence significative entre les performances pour la première moitié de l'entraînement (3.551%)

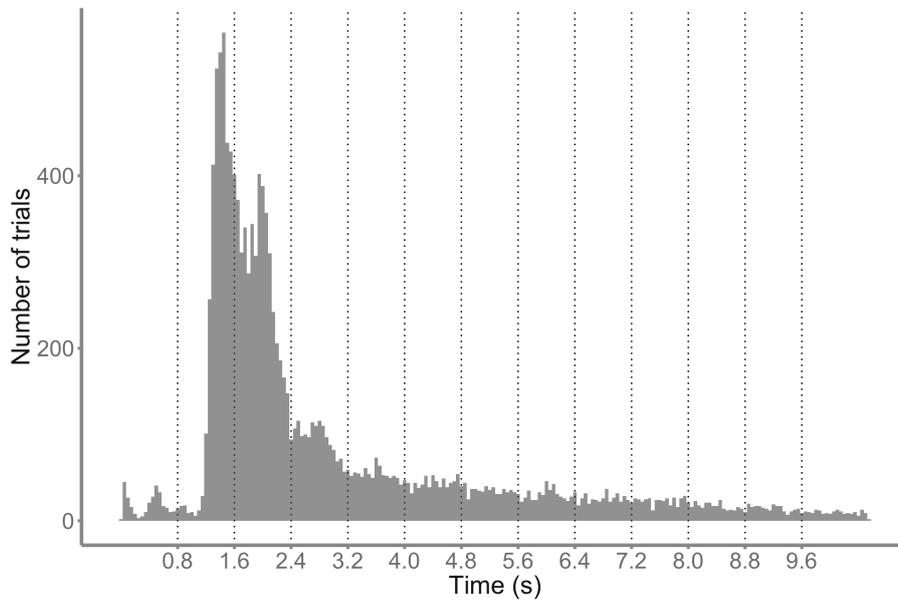


Figure 3.14: Pré/-post-tests - Masquage informationnel. Distribution des temps de réponse pour l'ensemble des participants pour les essais comportant une cible détectée. Les lignes verticales correspondent à chaque répétition du son pur cible.

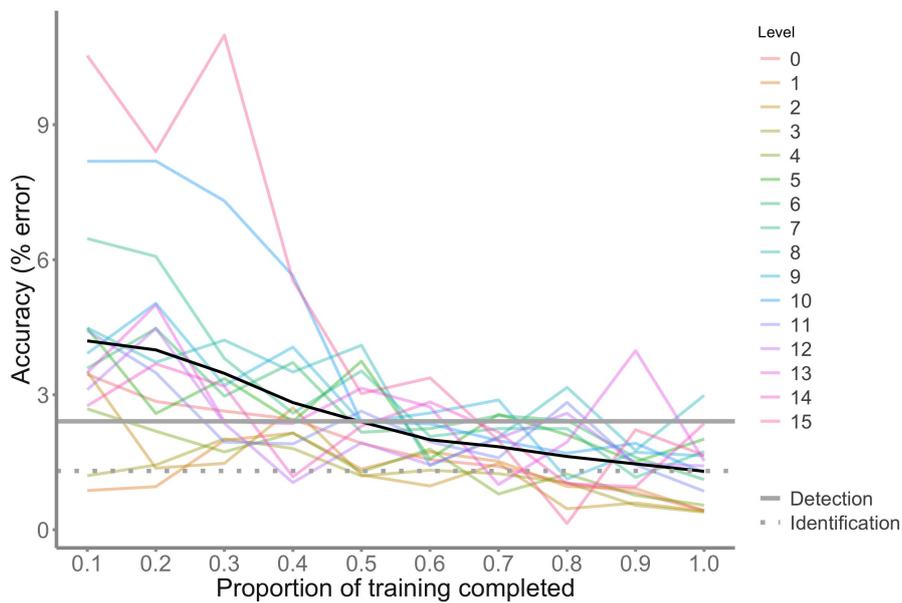


Figure 3.15: Tâche audio-motrice. Performance en % d'erreur au fur et à mesure de l'entraînement pour chaque niveau de difficulté (la moyenne est représentée en noir). Les seuils moyens (pour l'ensemble des participants du groupe MOTEUR, les pré- et post-tests, et les deux fréquences) de discrimination et d'identification de fréquence obtenus aux tâches de pré-/post-tests sont également représentés.

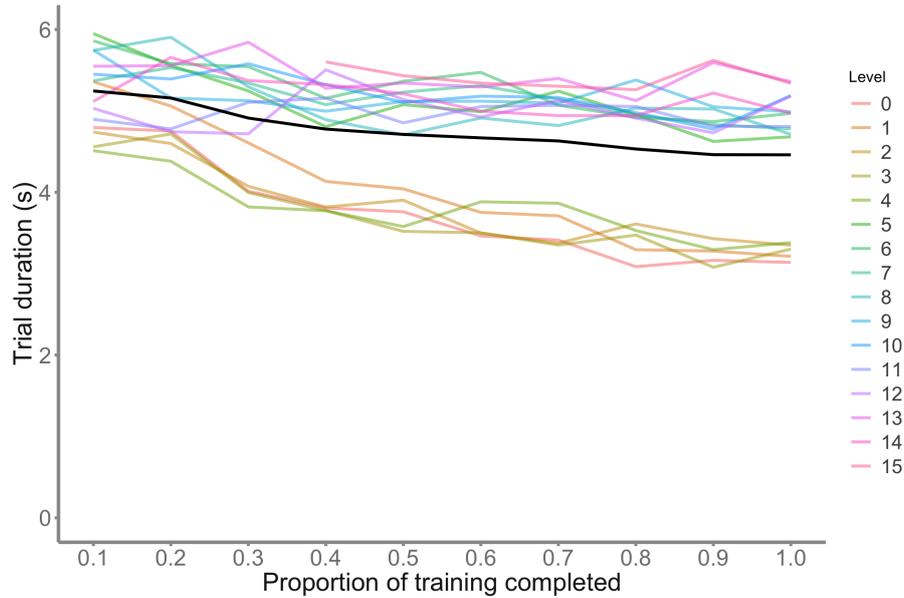


Figure 3.16: Tâche audio-motrice. Durée des essais au fur et à mesure de l’entraînement pour chaque niveau de difficulté (la moyenne est représentée en noir).

en comparaison des performances pour la seconde moitié de l’entraînement (1.701%) tous niveaux confondus ($Z = 140$, $p < 0.001$). On voit également que les performances atteignent celles obtenues à la tâche d’identification de fréquence.

La Figure 3.16 montre la durée des essais en fonction de la proportion de l’entraînement effectuée pour chaque niveau. Comme pour la Figure 3.15, cette proportion ne correspond pas nécessairement au même nombre d’essais d’un participant et d’un niveau à l’autre. On voit qu’après environ 30% de l’entraînement réalisé, les niveaux se distinguent en deux groupes distincts. Les niveaux 0 à 4 sont associés à des durées d’essais plus courtes (3.831 s) que les niveaux 5 à 15 (5.208 s). En outre, il semble que les durées d’essais soient légèrement plus basses à la fin de l’entraînement (5.509 s) par rapport au début de l’entraînement (5.007 s). Des tests t de Student ont confirmé la différence entre les performances des niveaux 0 à 4 en comparaison des niveaux 5 à 15 [$t(15) = 5.8$, $p < 0.001$], ainsi que la différence entre les durées d’essais pour la première moitié de l’entraînement en comparaison des durées d’essais pour la seconde moitié de l’entraînement tous niveaux confondus [$t(15) = 4$, $p = 0.001$].

3.4.3 Tâche d'exposition

Les participants du groupe EXPOSITION ont eu de bonnes performances à la tâche de discrimination de bruits blancs. En moyenne, les participants ont détecté la présence du bruit blanc dans 97% des essais et ont réussi à discriminer les bruits blancs ascendants des bruits blancs descendants dans 81% des essais.

3.5 Discussion

Ces résultats confirment que les capacités de perception de la fréquence (Micheyl et al., 2006) et d'attention sélective (Jones et al., 2014) peuvent être améliorées avec la pratique, car les participants se sont en général améliorés entre les pré- et post-tests pour les tâches de discrimination et identification de fréquence, ainsi que pour la tâche de masquage informationnel. Cependant, cette expérience n'a pas permis de mettre en évidence un effet robuste de l'entraînement audio-moteur en comparaison d'une simple exposition sonore ou même de la seule répétition des tâches tests, même si des tendances sont observées notamment pour les participants aux performances initiales les plus faibles. Une possibilité est que l'exposition au stimuli a également produit un apprentissage. Comme nous l'avons vu dans la partie théorique, (1.2.2.4, dans la sous-section *Entraînement inter-tâches*), les travaux de Wright et al. (2010) ont montré qu'un apprentissage perceptif pouvait être observé lorsqu'une phase d'exposition aux stimuli de la tâche était présente. Il semble donc que l'exposition aux stimuli ait un rôle non négligeable. Cependant, dans cette étude, l'exposition n'est efficace que lorsqu'elle est associée à une pratique de la tâche.

Les résultats ont montré que les performances aux tâches de discrimination et d'identification de fréquence étaient meilleures aux post-tests par rapport aux pré-tests. Cette différence ne dépendait pas du groupe ce qui suggère que c'est le fait de répéter les tâches de pré-/post-tests qui a permis cette amélioration. Cela confirme des travaux passés montrant une diminution des seuils de discrimination (Delhommeau et al., 2005; Demany, 1985; Demany & Semal, 2002; Irvine et al., 2000) et d'identification de fréquence (Micheyl et al., 2006). Cependant, ce n'est pas le cas de l'étude de Semal & Demany (2006) où un apprentissage n'a pas pu être observé. Une différence majeure entre l'étude de Semal & Demany (2006) et notre étude ou celles citées précédemment est que Semal & Demany (2006) utilisent un roving de la fréquence, c'est-à-dire que la fréquence du son pur standard change d'un essai à l'autre. Ce n'est pas le cas ici où la fréquence standard était identique sur l'ensemble du bloc (soit 515 Hz soit 1099 Hz). Cela suggère donc que le roving est un élément perturbant

l'apprentissage perceptif dans le cas de la perception de la fréquence. Une explication alternative est que la tâche utilisée par Semal & Demany (2006) utilise 4 sons purs à chaque essais ce qui peut impliquer des processus mnésiques différents des tâches 2I-2AFC et 3I-2AFC utilisées ici.

Ensuite, de meilleures performances à la tâche d'identification par rapport à la tâche de discrimination ont été observées. Semal & Demany (2006) et Mathias et al. (2010) ont montré à l'inverse que certains participants normo-entendants étaient incapables de déterminer la direction d'un changement de fréquence détectable. Cependant, Mathias et al. (2010) ont montré que cet effet est moindre lorsqu'il n'y a pas de roving de fréquence. Ici, les participants ont réalisé les tâches pré-/post-tests avec une fréquence standard de 515 Hz et de 1099 Hz sous forme de blocs ce qui pourrait expliquer que les seuils d'identification de fréquence ne sont pas plus élevés que les seuils de discrimination. Cependant, cela n'explique toujours pas les meilleures performances à la tâche d'identification. Une origine possible à cet effet dans notre protocole est l'ordre fixe de passation des tâches de discrimination et d'identification. Le but étant la comparaison des performances selon le groupe (CONTRÔLE, EXPOSITION ou MOTEUR) et le test (pré- ou post-test), ce facteur n'a pas été randomisé. Un effet d'apprentissage et de transfert entre la tâche de discrimination de fréquence passée en premier et la tâche d'identification passée en second pourrait peut-être expliquer les différences de performances. De plus, Hawkey et al. (2004) ont montré un transfert de l'apprentissage perceptif réalisé sur une tâche identique à notre discrimination de fréquence à une tâche identique à notre identification de fréquence. Il semble donc possible qu'un tel transfert a été opérant ici, fournissant un autre facteur pour les performances plus fortes sur la tâche d'identification.

Concernant la tâche de masquage informationnel, les résultats s'organisent selon un motif similaire. Pour l'analyse concernant l'ensemble des participants, une amélioration globale a été observée entre pré- et post-tests, mais aucune interaction significative avec le facteur groupe n'a été observée. Cela suggère donc que la phase d'entraînement audio-motrice n'est pas à l'origine de l'amélioration des performances et donc qu'une simple exposition aux stimuli ou même la simple répétition des tâches pré-/post-tests a suffit pour produire l'apprentissage. Pourtant, l'inspection visuelle des données et les statistiques descriptives ont montré une différence entre pré- et post-tests moins importante pour le groupe CONTRÔLE (0.558 d') que pour le groupe EXPOSITION (0.896 d') ou le groupe MOTEUR (1.03 d'). Il est possible que la forte variabilité des performances entre les groupes de sujets (performances

au pré-tests de 1.02, 1.58 et 1.22 d' respectivement pour les groupes CONTRÔLE, EXPOSITION et MOTEUR) aient rendu les comparaisons délicates. La variabilité inter-individuelle dans les tâches de masquage informationnel a été mise en avant dans de nombreuses études (Durlach et al., 2003a, 2003b; Kidd et al., 2003; Neff & Dethlefs, 1995; Richards & Neff, 2004).

En outre, les différences inter-individuelles sont importantes à prendre en compte dans l'influence qu'elles peuvent avoir sur l'apprentissage perceptif (Amitay et al., 2005). Par exemple, il peut sembler cohérent que les participants aux bonnes performances au pré-tests ont moins de marge de progression et il est possible que l'entraînement perceptif ne soit pas très efficace pour ces participants. De manière à vérifier l'effet de notre entraînement audio-moteur sur les participants les plus faibles à la tâche de masquage informationnel en pré-test, une seconde analyse n'incluant que le tiers le plus faible des participants a été réalisé. Malgré une amélioration moyenne plus importante pour le groupe MOTEUR (1.75 d') que pour le groupe EXPOSITION (0.935 d') et que le groupe CONTRÔLE (0.496 d'), les tests statistiques n'ont pas montré d'effet significatif après correction des valeurs de p ($p = 0.044$ avant correction). Étant donné le faible nombre de participants restant (16, répartis de manière inégale : CONTRÔLE : 7, EXPOSITION : 4, MOTEUR : 5), il est possible que l'analyse statistique ait simplement manqué de puissance.

Les résultats ont également montré des corrélations entre les tâches du pré-test. Il n'est pas étonnant que les performances entre les tâches de discrimination et d'identification de fréquence soient corrélées étant donné que la perception de la fréquence est mise en jeu dans le cas de ces deux tâches. En outre, une corrélation faible a été observée entre les tâches de discrimination et d'identification de fréquence et la tâche de masquage informationnel. Les tâches impliquent vraisemblablement des aptitudes et des mécanismes différents, avec une implication de l'attention sélective plus forte pour le masquage informationnel, et il semble donc cohérent que ces corrélations soient moins marquées.

En ce qui concerne les stratégies utilisées pour réaliser la tâche, la distribution des temps de réponse a montré l'utilisation d'une seule occurrence de la cible pour la plupart des essais. Cela suggère que la tâche a été réalisée en utilisant la hauteur absolue de la cible et non grâce à la détection d'une répétition d'un des sons purs du stimulus. Comme les conditions étaient organisées en blocs, la fréquence de la cible était identique pour l'ensemble du bloc, permettant une mémorisation de sa hauteur absolue. Cela est différent de la tâche de (Gutschalk et al., 2008) où il y a un roving de la fréquence de la cible, c'est-à-dire qu'elle prend une valeur différente à chaque

essai. Dans ce cas, la détection nécessite obligatoirement l'intégration temporelle de deux sons purs au minimum car la hauteur absolue ne suffit pas à prédire la présence de la cible.

Enfin, les résultats ont montré un apprentissage significatif dès le premier bloc du pré-test (515 ou 1099 Hz), ce qui correspond à seulement 150 essais. On peut noter que l'apprentissage observé par Jones et al. (2014) est aussi assez rapide et intervient principalement dans les 600 premiers essais. Dans la tâche utilisée par les auteurs, les stimuli sont d'une durée de 300 ms à chaque essai ce qui est beaucoup plus court que les essais de notre tâche pouvant aller jusqu'à 10 secondes.

3.6 Conclusions

Cette étude a confirmé que des capacités telles que la perception de faibles différences de fréquences, ou la capacité à isoler une cible sonore dans un bruit aux caractéristiques non-prédictibles peuvent être améliorées par la pratique. Cependant, nous n'avons pas observé d'apprentissage nettement plus efficace dans le cas d'un entraînement audio-moteur par rapport à une simple répétition des tâches pré- et post-test (groupe CONTRÔLE) ou à une stimulation auditive passive (groupe EXPOSITION), même si des tendances suggèrent que l'hypothèse mériterait d'être approfondie.

En revanche, un apprentissage rapide à la tâche de masquage informationnel révélé dès la phase de pré-test a été observé. Dans notre cas, cet effet a pu réduire d'éventuelles différences entre les groupes et rendre l'apprentissage lié à la tâche audio-motrice plus difficile à observer. Les mécanismes mis en jeu par cette amélioration de performances restent à éclaircir. Dans le chapitre qui suit, nous étudierons l'effet du roving de la fréquence cible sur ces effets d'apprentissage.

4 CHAPITRE 4. Entraînement perceptif et masquage informationnel

4.1 Aperçu

Le chapitre précédent a démontré un effet d'apprentissage rapide dans la tâche de masquage informationnel, même si seule la répétition de la tâche pendant le pré-test était prise en compte. Au cours des chapitres précédents, l'apprentissage perceptif a été considéré sur des tâches différentes des tâches tests (pré-/post-tests). À l'inverse, l'étude est ici focalisée sur l'amélioration des performances de la tâche elle-même, et sans l'implication de la boucle audio-motrice. Le but principal est de caractériser plus précisément la nature de l'apprentissage perceptif dans la tâche de masquage informationnel, notamment en contrôlant (1) l'effet d'une présentation de cibles dont la fréquence varie d'un essai à l'autre (*roving*), (2) des caractéristiques temporelles du stimulus (SOA de la cible) et (3) du niveau (différence de niveau entre la cible et le nuage de tons). Les résultats ont montré une amélioration des performances au cours de l'entraînement malgré la présence du *roving* et un contrôle des paramètres temporel et de niveaux, indiquant que la tâche de masquage informationnel semble réellement susceptible à une amélioration par la pratique. En outre, un effet inattendu de la fréquence de la cible sur la quantité de masquage a été observé, avant et après entraînement.

4.2 Introduction

Dans le chapitre précédent, deux fréquences ont été utilisées en blocs de 150 essais et les résultats ont montré un apprentissage perceptif rapide. On peut donc se demander si cet apprentissage perceptif était dépendant de la présentation par bloc, et donc si un apprentissage pourrait être observé lorsque la fréquence change à chaque essai. Pour cela, dans ce chapitre, la fréquence de la cible était différente d'un essai à l'autre.

Nous avons en effet vu dans la partie théorique que de nombreux travaux montrent l'influence de l'ordre temporel des valeurs du stimulus (*roving*) sur l'apprentissage perceptif. Par exemple, dans le domaine visuel, un apprentissage de différentes valeurs de contraste est observé uniquement lorsqu'il n'y a pas de *roving* (Yu et al., 2004). Un apprentissage peut toutefois être observé lorsque qu'une séquence temporelle fixe non-ordonnée des valeurs du stimulus est utilisée (Kuai et al., 2005).

Dans le cadre du masquage informationnel auditif, il semble également que le *roving* de la fréquence ait un effet délétère sur les performances (Richards et al., 2004, expérience IVa). A l'inverse, le fait de garder fixe les fréquences de la cible et des composants du masqueur améliore les performances des participants (Richards et al., 2004, expérience IVb). Cependant les performances ne peuvent pas être directement comparées à celle où le masqueur est différent à chaque essai (expérience I) car la difficulté a été ajustée pour maintenir une performance d'environ 75% (mais les auteurs reportent que le niveau du signal nécessaire pour atteindre cette performance était plus bas).

Ensuite, le but de cette série d'expériences est de contrôler le SOA de la cible, qui était similaire pour tous les essais, et aurait ainsi pu faire l'objet d'apprentissage. Pour cette raison, le SOA des cibles était aléatoire d'un essai à l'autre (mais constant à l'intérieur de chaque essai) dans ce chapitre.

En outre, dans la tâche de masquage informationnel du chapitre précédent, le niveau de la cible (SNR) était aléatoire à chaque essai mais toujours différent du niveau des sons purs individuels du masqueur (différence entre -8 et -48 dB). De manière à éviter que l'apprentissage ne concerne la perception d'une différence d'intensité, cibles et masqueurs ont été fixés à niveaux identiques. Cela diffère donc également des travaux où une procédure adaptative sur le niveau de la cible est utilisée (par exemple, Oxenham et al., 2003), et notamment où un apprentissage a été observé (Jones et al., 2014).

Enfin, une amorce a été utilisée dans certains blocs expérimentaux de ce chapitre. La présence d'une amorce distingue les études de Oxenham et al. (2003) et Jones et al. (2014) qui obtiennent des résultats contradictoires concernant l'évolution des performances des participants à des tâches de masquage informationnel. Alors que Jones et al. (2014) observent une amélioration des performances, Oxenham et al. (2003) montrent un apprentissage uniquement dans une des conditions (MBD, *MultiBurst Different*), et pour un des groupes (musiciens). Une différence notable entre les deux procédures est l'amorce, soit l'écoute d'une cible seule avant chaque bloc dans l'étude de Jones et al. (2014). Le but de cette amorce était également de réduire l'incertitude du stimulus et donc de diminuer la variabilité inter-individuelle, sans éliminer l'effet de masquage informationnel (Durlach et al., 2003a; Neff et al., 1993).

4.3 Méthodes

4.3.1 Participants

Dix participants (7 femmes ; âge en années $M = 20.9$, $SD = 2.807$) normo-entendant ont participé à cette étude. Ils étaient indemnisés 10 euros par heure. Un audiogramme préalable permettait de s'assurer que tous les participants avaient une audition normale, c'est à dire des seuils auditifs de 20 dB HL ou moins pour les fréquences testées (125, 250, 500, 1000, 2000, 3000, 4000, et 8000 Hz). Tous les participants étaient non-musiciens (moins de deux ans de pratique musicale au moment de l'étude).

4.3.2 Stimuli

4.3.2.1 Masqueurs

Les masqueurs étaient des ensembles de sons purs de 100 ms formant des nuages de tons aléatoires en temps et en fréquence et organisés sous la forme de plusieurs composants simultanés dont le nombre a été adapté aux performances du participant (pendant la première phase de l'expérience, voir Figure 4.3). La méthode développée pour construire les nuages de tons est similaire à celle décrite dans l'étude de Gutschalk et al. (2008). Cependant, les composants n'étaient pas contraints à l'une des bandes de fréquence, mais une nouvelle bande était choisie aléatoirement à chaque nouvelle occurrence. Le but était de pouvoir utiliser un nombre de composants qui n'était pas nécessairement un multiple du nombre de bandes de fréquence.

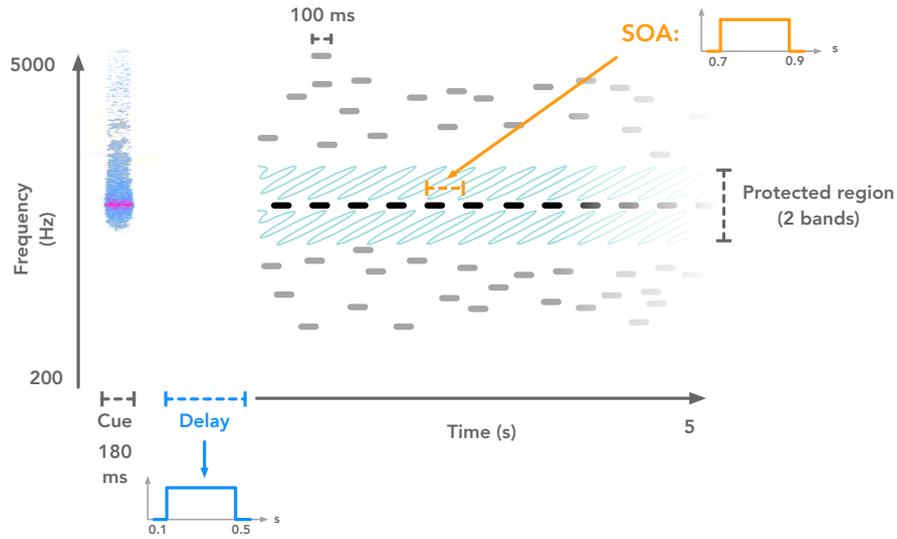


Figure 4.1: Procédure. Représentation de la cible (en noir) et des sons purs du nuage de tons (en gris). La région fréquentielle protégée est illustrée en bleu.

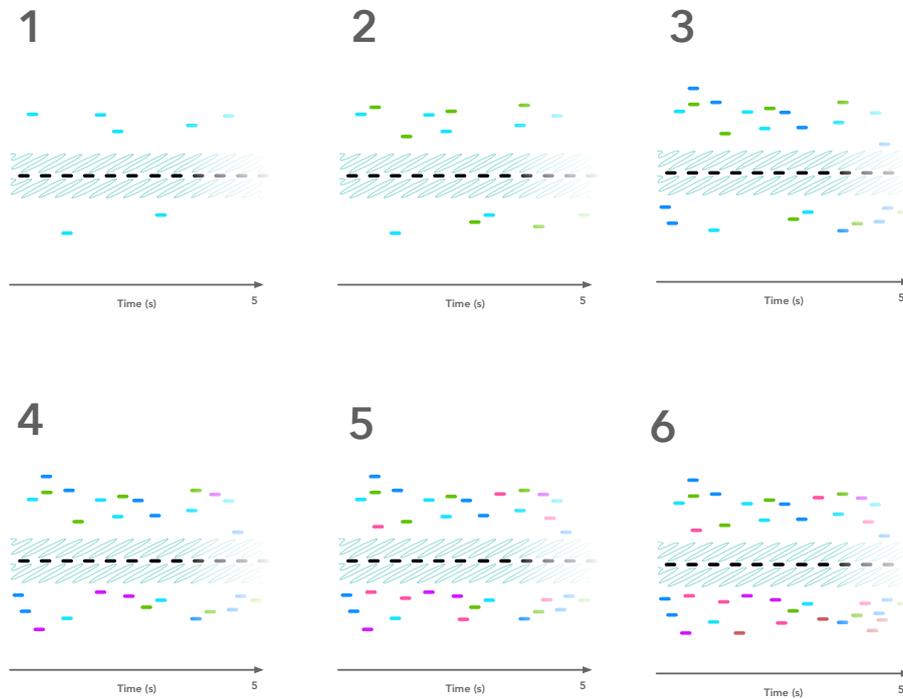


Figure 4.2: Procédure. Processus de création des nuages de tons. Des nuages de tons comportant de 1 à 6 composants sont représentés.

La Figure 4.2 illustre le processus de création des nuages de tons en montrant des nuages de tons comprenant un nombre de composants de 1 à 6. Pour chaque composant, la fréquence de chaque son pur était choisie aléatoirement. Dix-huit fréquences centrales réparties régulièrement sur une échelle logarithmique de 200 Hz à 5000 Hz (239, 286, 342, 409, **489**, 585, **699**, 836, **1000**, 1196, **1430**, 1710, **2045**, 2445, **2924**, 3497, 4181, 5000 Hz ; les cibles possibles sont en gras) étaient utilisées. Deux fréquences centrales au-dessous et deux au-dessus de la fréquence de la cible étaient exclues donnant lieu à une région protégée autour de la fréquence de la cible (illustrée en bleu sur la Figure 4.1). Pour chaque son pur du nuage, une fréquence centrale était choisie parmi les 13 possibilités restantes. Par exemple, pour une cible à 1000 Hz, les fréquences centrales possibles étaient 239, 286, 342, 409, 489, 585, 1710, 2045, 2445, 2924, 3497, 4181, 5000 Hz. La fréquence exacte de chaque son pur était ensuite choisie aléatoirement à partir d'une distribution uniforme de la largeur d'un ERB (*Equivalent Rectangular Bandwidth*) autour de la fréquence centrale.

$$ERB = 24.7 \times (4.37 \times f_c + 1)$$

avec f_c en kHz (Moore, 1995). Pour chaque composant, un SOA, correspondant à la durée entre un son pur et le suivant, était choisi à chaque occurrence dans une distribution uniforme entre 100 ms et $SOA_{cible} \times 2$ (SOA_{cible} étant le SOA de la cible pour l'essai en question). Une représentation d'un nuage de tons est montré en gris sur la Figure 4.1.

Le paramètre que nous désignons sous le terme de densité correspond au nombre de composants, c'est-à-dire au nombre maximal théorique de sons purs qui peuvent être joués simultanément. En fait, du fait du caractère aléatoire du SOA, la plupart des sons purs ne sont pas simultanés.

Les sons purs du nuage étaient individuellement présentés à un niveau de 55 à 60 dB SPL (selon la réponse du casque, différente en fonction de la fréquence).

4.3.2.2 Cibles

Les cibles étaient des sons purs de 100 ms (incluant des rampes ascendantes et descendantes de 10 ms chacune) répétés de façon isochrone, c'est à dire avec un SOA (durée entre le début d'un son pur et le début du suivant) constant à l'intérieur d'un essai. Elles commençaient après le début du masqueur avec un délai de $1000 - SOA_{tc}$ avec SOA_{tc} correspondant au SOA du premier son pur du nuage. Comme SOA_{tc} était compris entre 0.1 et $SOA_{cible} \times 2$ pour cet essai, le délai était

compris entre $1000 - (SOA_{cible} \times 2)$ et 900 ms. Un exemple de cible est représenté en noir sur la Figure 4.1. A chaque essai, la fréquence de la cible était choisie de manière pseudo-aléatoire parmi les valeurs suivantes : 489, 699, 1000, 1430, 2045, 2924 Hz (probabilités égales). Le SOA des sons purs de la cible (SOA_{cible}) était choisi aléatoirement dans une distribution uniforme de 700 à 900 ms. Cette variation a été utilisée de manière à s'assurer que l'apprentissage ne dépend pas du déroulement temporel du stimulus.

Les cibles étaient de même niveau que les sons purs du nuage de manière à les rendre indiscernables par le biais de l'attribut d'intensité. Elles étaient donc présentées entre 55 et 60 dB SPL (en fonction de la réponse du casque selon la fréquence). De cette manière, la détection de la cible ne pouvait pas être réalisée (et entraînée) grâce à une différence de niveau entre les sons purs de la cible et les sons purs du nuage de tons.

4.3.2.3 Amorces

Dans l'ensemble des essais, des amorces étaient présentes avant le début de l'essai. Les amorces étaient des bruits blancs, filtrés autour de la fréquence à amorcer par un filtre bi-quadratique avec $Q = 69.25$ (correspondant à une largeur de bande de 1/8 de ton à -3 dB) d'une durée de 180 ms.

4.3.3 Matériel

Les participants ont été testés individuellement dans une cabine insonorisée à double paroi (Industrial Acoustics). Les stimuli ont été joués de manière diotique par le biais d'une carte son RME Fireface UC à une fréquence d'échantillonnage de 44.1 kHz et de résolution égale à 16-bit. Les sons ont été présentés par un casque audio Sennheiser HD 600.

4.3.4 Procédure

Avant de commencer l'expérience, les participants ont été familiarisés avec les stimuli et ont entendu trois exemples de cibles (489, 1000 et 2045 Hz), d'abord seules puis avec masqueur.

La Figure 4.3 montre le déroulement de l'expérience. La tâche dans les parties 1 et 2 était de détecter des cibles masquées par un nuage de tons. Après chaque essai de

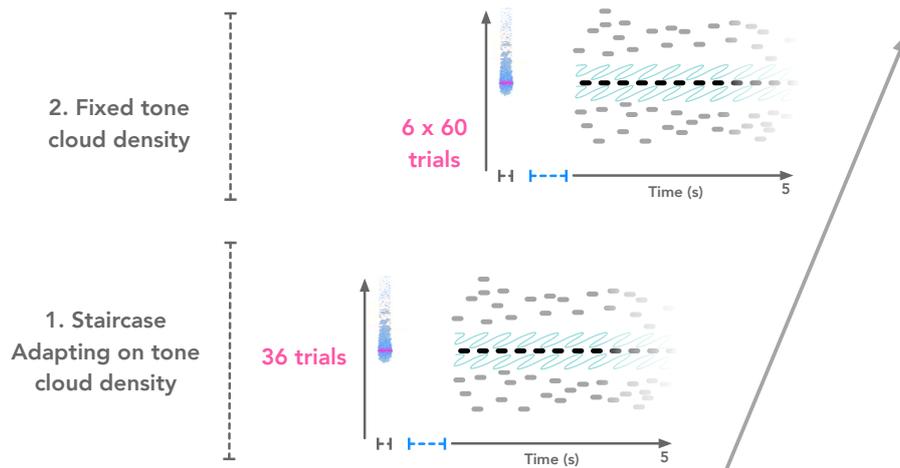


Figure 4.3: Procédure. Déroulement de l’expérience. Dans une première phase (1), les participants ont réalisé la tâche avec une densité du nuage adaptée selon une procédure en escalier. Dans la phase principale (2), ils ont réalisés 6 sessions de 60 essais chacune avec la densité résultant de la première phase.

5 secondes, les participants devaient répondre si une cible était présente dans l’essai ou non. La réponse déclenchait le passage à l’essai suivant.

De manière à éviter les effets plafond et plancher liées aux fortes variabilités inter-individuelles dans les paradigmes de masquage informationnel (Durlach et al., 2003a, 2003b; Kidd et al., 2003; Neff & Dethlefs, 1995; Richards & Neff, 2004), que nous avons retrouvés lors de données pilotes, la difficulté de la tâche a été adaptée pour chaque participant. Ceci a été fait en adaptant la densité du nuage de tons pour chaque participant. Dans cette première partie de l’expérience, les participants ont été testés sur la tâche avec un retour sur leurs performances (*feedback*), et avec un nuage de tons adapté en densité. La procédure adaptative était une procédure en escalier de type *2-up/1-down* visant une performance de 70.7% (Levitt, 1971). Une réponse correcte conduisait à une augmentation de la densité du nuage de tons (du nombre de composants) et une réponse incorrecte conduisait à une diminution de la densité. La valeur de départ était de 5 composants et était restreinte entre 5 et 50 composants. Le pas utilisé était de 1.5 jusqu’à ce que 3 *reversals* aient été réalisés puis de 1.25 ensuite. Le nombre de composants était multiplié ou divisé par le pas et arrondis de manière à obtenir un entier. Cela implique que dans certains cas, la difficulté n’était pas modifiée. On peut noter que toutefois, les procédures ont convergé de manière satisfaisante. Cette partie se terminait lorsque les participants avaient réalisé 36 essais (de manière à obtenir un pré-apprentissage limité et avec un

nombre d'essais identique pour les différents participants). Le seuil était calculé à partir des derniers *reversals* (3 si le nombre total de *reversals* était compris entre 3 et 5 et 5 si le nombre total de *reversals* était supérieur à 5).

Dans la seconde partie de l'expérience, les participants ont réalisé 6 sessions de 60 essais chacune. Aucun *feedback* n'était donné aux participants, et la densité du nuage de tons était fixe et correspondait à la valeur de seuil calculée lors de la première phase. Les essais étaient d'une durée de 5 secondes et 50% d'entre eux comportaient une cible.

4.3.5 Analyse de données

Les réponses ont été converties en pourcentages de détections correctes et de fausses alarmes. L'index de sensibilité d' a été calculé pour chaque participant, chaque session (1, 2, 3, 4, 5 ou 6) et chaque fréquence de la cible (489, 699, 1000, 1430, 2045 ou 2924 Hz). Lorsqu'une valeur de détections correctes ou de fausses alarmes était de 0% ou de 100%, 0.5 essai a été ajouté ou retiré de manière à rendre possible le calcul du d' . Des tests statistiques ont été réalisés sur les valeurs de d' à l'aide d'ANOVA à mesures répétées avec les facteurs session et fréquence comme facteurs intra-participants. L'homogénéité des variances entre les différentes conditions (sphéricité) a été vérifiée et des corrections de Greenhouse-Geisser ont été appliquées lorsque nécessaire.

4.4 Résultats

La Figure 4.4 montre la densité du nuage de tons en nombre de composants qui a été utilisée pour chaque participant. On voit qu'une densité d'environ 10 composants ($M = 11.1$) a généralement été utilisée sauf pour $p2$ qui a eu des performances particulièrement bonnes à la première partie de l'expérience, et où la procédure adaptative sur la densité a convergé vers un nuage de tons comportant 29 composants.

La Figure 4.5 représente l'évolution des performances au fur et à mesure des 6 sessions réalisées par les participants. On peut noter des performances globalement assez faibles. On voit une tendance à l'augmentation des performances avec la pratique, la session 1 ($0.69 d'$) étant associée à de moins bonnes performances que la session 6 ($1.36 d'$). Les tests statistiques ont confirmé un effet de la session sur les performances [$F(5, 45) = 4.46, p = 0.0022, \eta_g^2 = 0.169$].

La Figure 4.6 représente les performances en fonction de la fréquence de la cible. On voit que les cibles de basses fréquences sont associées à de meilleures performances

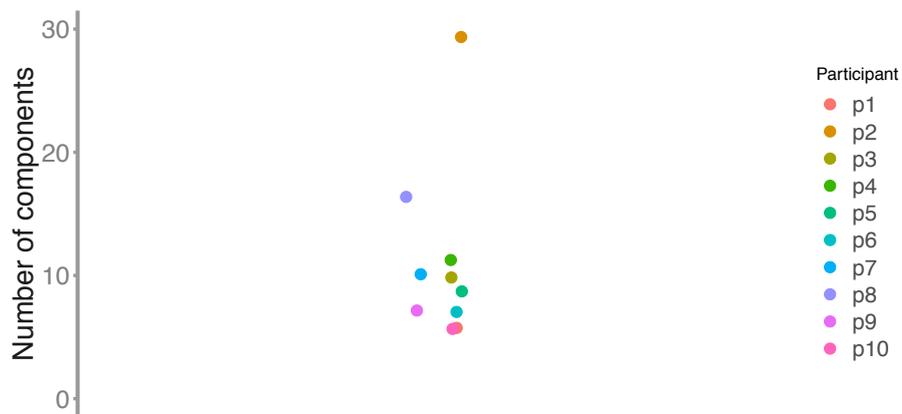


Figure 4.4: Densité des nuages de tons. Les résultats de la première partie de l'expérience ont donné lieu à un nombre de composants différent pour chaque participant.

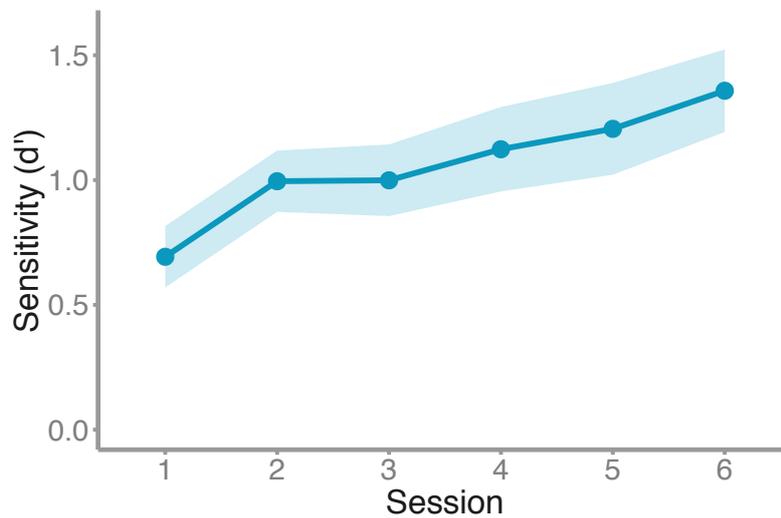


Figure 4.5: Effet de la session. Les performances (mesurées en d') moyennées entre les participants sont représentées en fonction de la session. Le ruban représente l'erreur standard.

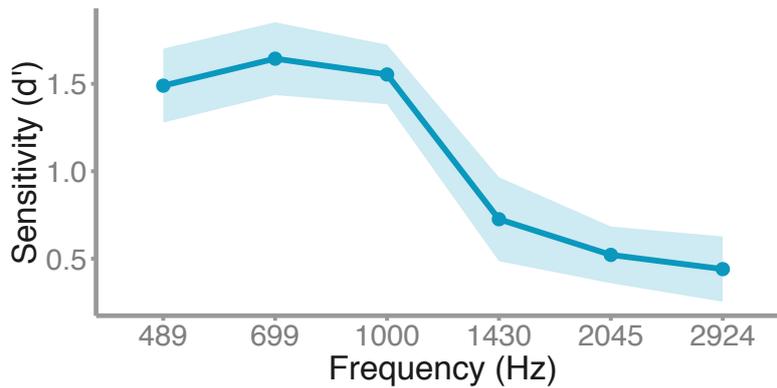


Figure 4.6: Effet de la fréquence. Les performances (mesurées en d') sont représentées en fonction de la fréquence de la cible (toutes sessions moyennées). L'échelle de l'axe x est logarithmique.

(1.49, 1.64 et 1.55 d' respectivement pour les cibles à 489, 699 et 1000 Hz) que les cibles de hautes fréquences (0.73, 0.52 et 0.44 respectivement pour les cibles à 1430, 2045 et 2924 Hz). Les tests statistiques ont confirmé un effet de la fréquence sur les performances [$F(5, 45) = 10.18, p < 0.001, \eta_g^2 = 0.423$].

La Figure 4.7 montre les performances en fonction de la fréquence de la cible (facteur représenté par la couleur) et de la session. L'augmentation des performances au fur et à mesure des sessions semble similaire pour les différentes fréquences. On peut tout de même noter que la fréquence 2924 Hz, qui est associée aux moins bonnes performances, ne semble pas faire l'objet de progrès. Les tests statistiques ne révèlent cependant aucune interaction significative entre les facteurs fréquence et session [$F(25, 225) = 0.74, p > 0.05, \eta_g^2 = 0.0419$].

La Figure 4.8 montre les performances individuelles des participants. L'effet de fréquence semble concerner la majorité des participants, bien qu'une forte variabilité caractérise les performances moyennes de chaque participant. Le point d'inflexion de la baisse de performances est en outre relativement stable (1430 Hz).

4.5 Discussion

Les résultats ont montré un effet d'apprentissage dans la tâche de masquage informationnel. Cela confirme les travaux utilisant des complexes multi-tons à la place des nuages de tons (Jones et al., 2014), ainsi que les résultats observés dans le chapitre précédent. Il semble donc que cet apprentissage est possible malgré le

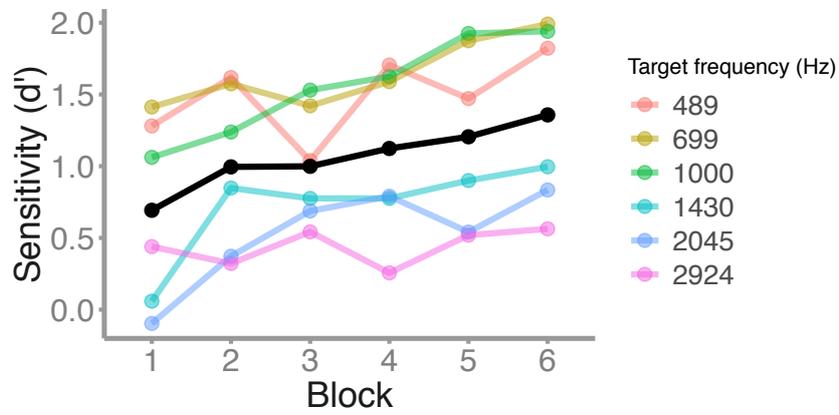


Figure 4.7: Interaction entre la fréquence et la session. Les performances (mesurées en d') sont représentées en fonction de la fréquence de la cible (couleurs) et la session (axe x).

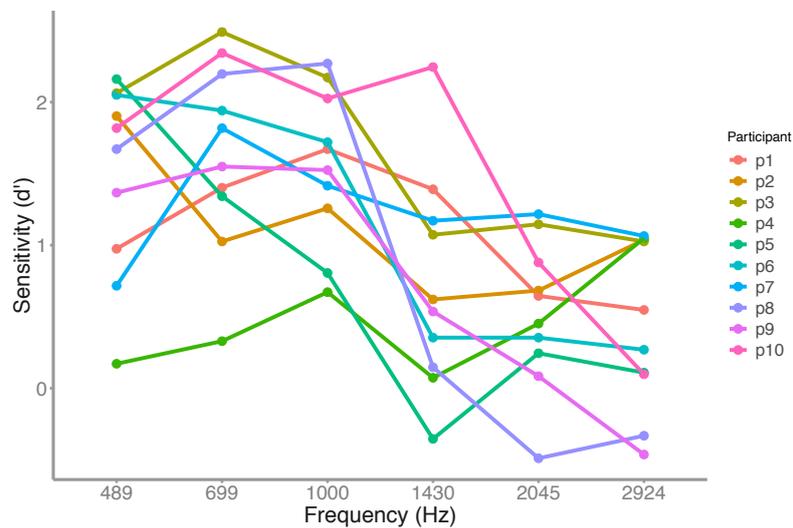


Figure 4.8: Performances individuelles des participants en fonction de la fréquence de la cible. L'échelle de l'axe x est logarithmique.

roving de la fréquence de la cible d'un essai à l'autre. De plus, le fait que les SOA de la cible soient aléatoires d'un essai à l'autre (bien que constant à l'intérieur d'un essai) élimine l'hypothèse d'un apprentissage fondé sur l'aspect temporel du stimulus. Enfin, les niveaux identiques entre tons des masqueurs et des cibles suggèrent que l'apprentissage n'est pas non plus fondé sur la perception de l'intensité. En effet, dans le chapitre précédent, les sons purs de la cible avaient un niveau différent des sons purs individuels du nuage de tons. Ce n'était pas le cas dans l'expérience de ce chapitre, rendant la détection de la cible fondée sur une différence de niveau impossible. Une hypothèse, en concordance avec les résultats de Jones et al. (2014), suggérant que l'amélioration des performances dans une tâche de masquage informationnel est liée à une amélioration de l'attention sélective, est que l'attention sélective s'est améliorée par le biais d'une meilleure extraction des caractéristiques fréquentielles de la cible permettant d'élaborer un flux auditif à partir de la répétition d'un son pur de même fréquence, et ainsi de l'extraire des éléments du nuage de tons.

De manière à éviter les effets plafond et plancher observés dans ce type de tâche de masquage informationnel, les participants ont réalisé une session courte où la densité du nuage était adaptée. Cependant, il est possible que cette manière de moduler la difficulté de la tâche ne soit pas la plus efficace. Il a été montré que l'augmentation de la difficulté avec le nombre de composants n'est pas linéaire, mais atteint un pic puis redescend (Neff & Green, 1987). Bien que les stimuli utilisés par Neff & Green (1987) étaient des multi-tons qui ne comportent pas d'aspect temporel aléatoire comme c'est le cas des nuages de tons, il est possible que cet effet soit également présent. Le but de cette phase de l'expérience est utilisée uniquement pour trouver une zone approximative de performance permettant d'éviter les performances extrêmes.

Enfin, les résultats ont montré un effet inattendu et robuste de la fréquence sur la capacité à percevoir la cible dans le bruit. Cet effet n'est pas documenté à notre connaissance. La majorité des études de masquage informationnel utilisent des cibles à 1000 Hz ou bien ne comparent pas les performances selon la fréquence. Il semble également que cet effet ne soit pas réduit par la pratique car aucun effet d'interaction significatif entre fréquence et session n'a été observé.

4.6 Conclusions

On peut se demander si cet effet de fréquence pourrait provenir des seuils auditifs qui ne sont pas identiques selon la fréquence, et qui pourraient donc faire que certaines cibles aient un niveau subjectif plus élevées que d'autres même si leur

niveau acoustique est identique ? Pour cela, une série d'expériences a été réalisée dans le chapitre suivant, dans lesquelles l'ensemble des stimuli - les cibles et tous les sons purs des nuages de tons - a été pondéré selon l'audiogramme de chaque participant.

5 CHAPITRE 5. Audiogramme informationnel : un effet de masquage informationnel dépendant de la fréquence et sans lien avec l'audiogramme de détection

5.1 Aperçu

Dans le chapitre précédent, un effet de la fréquence de la cible sur la quantité de masquage informationnel a été observé. Cet effet, non documenté jusqu'à présent, pourrait au moins en partie être dû à des différences de détectabilité des cibles, liées à l'audiogramme classique. Pour tester cette interprétation, le même protocole de masquage informationnel a été repris mais en pondérant chaque fréquence utilisée dans les stimuli par l'audiogramme du participant. Les résultats ont à nouveau montré une amélioration des performances et un effet de fréquence. Une autre interprétation possible pourrait être que la position relative de la cible par rapport aux bords du nuage de ton servant de masqueur informationnel a une influence sur les performances. Une seconde expérience a contrôlé ce paramètre et utilisé des fréquences absolues de cibles identiques pour différentes formes de masqueurs. Les résultats montrent que l'effet de fréquence est absolu, et non lié aux caractéristiques fines du masqueur. L'effet est de plus non monotone par rapport à la fréquence, rejetant une interprétation basée sur la largeur des filtres auditifs qui eux s'élargissent de façon monotone. Ces résultats suggèrent un réel effet de la fréquence sur le masquage informationnel, non attribuable aux seuils de détections dans le silence, ni à la largeur des filtres auditifs.

5.2 Introduction

La plupart des études sur le masquage informationnel utilisent des cibles à 1000 Hz (voir les protocoles de masquage informationnel dans 1.3.1) et ne questionnent donc pas le rôle de la fréquence de la cible sur ce phénomène. Même si aucune étude à notre connaissance n'a spécifiquement examiné l'effet de la fréquence sur l'attention sélective, des observations secondaires peuvent être récoltées, mais mènent cependant à des conclusions contradictoires. Certaines études suggèrent que la fréquence a un effet sur l'attention sélective (Dykstra & Gutschalk, 2015; Giani et al., 2015; Gutschalk et al., 2008; Kidd et al., 2011; Neff & Green, 1987; Wiegand & Gutschalk, 2012) et d'autres que ce n'est pas le cas (Cao & Richards, 2012, table 1; Richards & Neff, 2004, exp 3).

Un effet de la fréquence de la cible dans une tâche de masquage informationnel a été observé dans le chapitre précédent. Plus précisément, les cibles aiguës ont été plus difficilement perçues lorsque masquées informationnellement. Le but de ce chapitre est de répliquer ces résultats et de tester deux hypothèses qui pourraient expliquer ces résultats.

La première concerne les différences d'audibilité en fonction de la fréquence. En effet, malgré la présentation des cibles au-dessus du seuil de détection dans le silence, on peut penser que leur saillance relative dépend de leur niveau au dessus du seuil de détection. De manière à tester cette hypothèse simple, l'ensemble des stimuli, cibles et nuages de tons, ont été présenté à niveau dépendant de l'audiogramme de chaque participant.

Le but de cette première expérience était également d'évaluer l'hypothèse d'un effet de fréquence dont l'origine est liée à l'incertitude de la fréquence de la cible. Pour cela, des conditions avec et sans amorce ont été réalisées. Si cet effet provenait d'un biais attentionnel naturel vers telle ou telle zone fréquentielle, la condition avec amorce devrait réduire l'effet de fréquence par rapport à la condition sans amorce car ce biais serait en partie compensé par l'indication de fréquence. Ensuite, la condition avec amorce a été réalisée par les participants entre deux sessions sans amorce. Cela a permis d'évaluer l'effet d'entraînement perceptif associé à la condition avec amorce en comparant les sessions sans amorces avant et après les sessions avec amorce. Une interaction avec l'effet de fréquence a été spécifiquement recherchée de manière à tester l'influence de l'apprentissage sur cet effet de fréquence.

Une seconde hypothèse qui pourrait expliquer l'effet de fréquence est que les cibles aiguës ont une position relative au nuages de tons qui diffère par rapport aux cibles

grave ou médium. Pour tester cette hypothèse, une seconde expérience a été mise en place dans ce chapitre, dans laquelle deux types de nuages de tons ont été utilisés pour des cibles de fréquences identiques. Les bornes grave et aiguë de l'un des nuages de tons ont été déplacées. Cela permet ainsi de comparer pour chaque cible les performances entre les deux types de nuages de tons.

Une deuxième extension permise par l'expérience 2 était d'examiner les performances pour des cibles plus graves, pour déterminer si l'effet de fréquence est passe-bas ou passe-bande. Les cibles utilisées dans les expériences précédentes utilisaient une fréquence de 489 Hz minimum. Les cibles testées dans l'expérience 2 ont inclus des fréquences allant de 340 Hz à 4574 Hz.

5.3 Expérience 1

Le but de cette expérience est d'évaluer l'effet de fréquence lorsque l'ensemble des stimuli, cibles et nuages de tons, sont pondérés en fonction de l'audiogramme du participant. Ensuite, l'effet de donner aux participants un indice exogène concernant la région fréquentielle de la cible avant l'essai (une amorce) a été évalué.

De manière pratique, les stimuli ont été optimisés et certains paramètres ont été modifiés par rapport à ceux du chapitre précédent. Les essais ont été raccourcis pour accélérer la procédure. En outre, une adaptation de la largeur de la région de fréquence protégée, connue pour influencer les performances de détection (Elhilali et al., 2009; Micheyl et al., 2007; Neff et al., 1993), a été utilisée à la place de l'adaptation sur la densité du nuage de tons. Cela permet d'éviter l'évolution non-linéaire de la quantité de masquage en fonction du nombre de composants.

5.3.1 Méthodes

5.3.1.1 Participants

Dix participants (5 femmes ; âge en années $M=24.5$, $SD=2.42$) normo-entendant ont participé à cette étude. Ils étaient indemnisés 10 euros par heure. Un audiogramme préalable permettait de s'assurer que tous les participants avaient une audition normale, c'est à dire des seuils auditifs de 20 dB HL ou moins pour les fréquences testées (125, 250, 500, 1000, 2000, 3000, 4000, et 8000 Hz). Tous les participants étaient non-musiciens (moins de deux ans de pratique musicale au moment de l'étude).

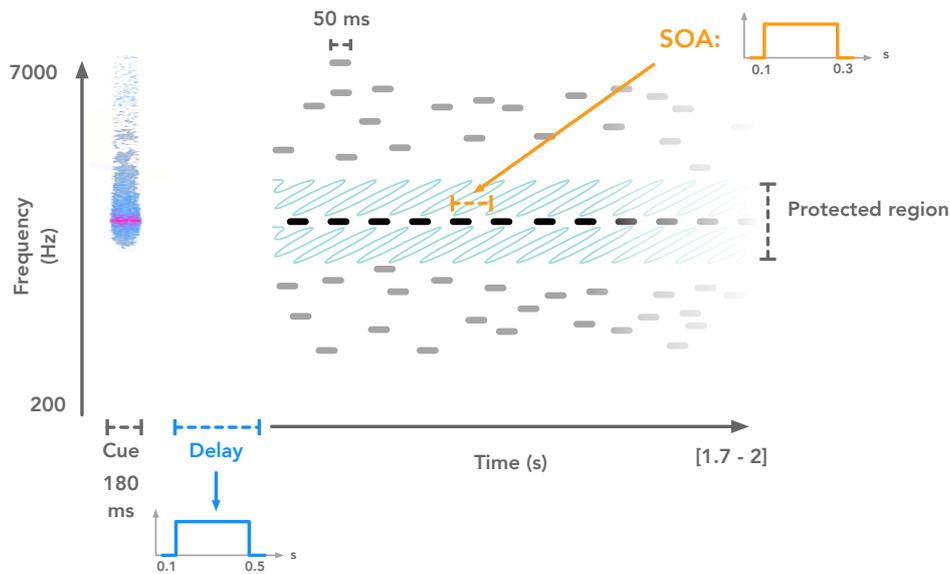


Figure 5.1: Expérience 1 - Cible et masqueur. Un essai de la condition amorce est représenté.

5.3.1.2 Stimuli

Les stimuli sont similaires à ceux de l'expérience du chapitre précédent.

La Figure 5.1 représente un essai avec amorce. Les différences avec les stimuli du chapitre précédent sont précisées dans ce qui suit.

Masqueurs

Les tons des nuages de tons étaient d'une durée de 50 ms. Ils étaient choisis de manière aléatoire dans une distribution logarithmique entre 200 et 7000 Hz. La largeur de la région protégée autour de la fréquence de la cible était choisie grâce à une procédure adaptative initiale.

De manière à pouvoir choisir n'importe quelle valeur pour la taille de la région protégée, les nuages de tons ont été élaborées différemment par rapport au chapitre précédent, où les bandes de fréquence autour de la cible étaient exclues. La densité spectro-temporelle était contrôlée en divisant chaque essai en *cellules temporelles*. La durée des cellules temporelles correspondait au SOA de la cible pour cet essai (choisi aléatoirement dans une distribution uniforme entre 100 et 300 ms ; voir plus bas). Les essais étaient d'une durée maximum de 2 secondes et contenaient autant de cellules temporelles que possible à l'intérieur de ces 2 secondes, ce qui a donné

lieu à des durées d'essais entre 1.7 et 2 secondes. Treize tons étaient joués dans chacune des cellules temporelles, dont la fréquence était aléatoirement choisie dans une distribution uniforme sur une échelle logarithmique. Leur positions temporelles exactes étaient choisies aléatoirement dans une distribution uniforme ayant pour borne le début et la fin de la cellule.

Cibles

Les cibles étaient des tons purs répétés d'une durée de 50 ms et de fréquence et de SOA constant à l'intérieur de l'essai. Le SOA était choisi aléatoirement dans une distribution uniforme entre 100 et 300 ms à chaque essai mais restait constant au sein d'un essai.

Amorces

Dans la condition AMORCE (voir procédure), une amorce identique à celles utilisées dans l'expérience du chapitre précédent était présente au-début de l'essai.

5.3.1.3 Matériel

Le matériel est identique à celui du chapitre précédent.

5.3.1.4 Procédure

La procédure est similaire à celle du chapitre précédent. La Figure 5.2 représente les différentes phases de l'expérience 1.

Phase audiogramme

La première étape réalisée par les participants était un audiogramme aux fréquences extrêmes du masqueur (200 et 7000 Hz) et aux fréquences des cibles (489, 699, 1000, 1430, 2045, 2924 Hz). Les participants entendaient deux intervalles successifs visuellement distinguables et devaient choisir lequel des deux contenait un son pur. Une procédure adaptative, indépendante pour chaque fréquence (ordre aléatoire) de type *2-down/1-up* visant 70.7% de réponses correctes (Levitt, 1971), a été utilisée sur le niveau du ton. Le pas initial était fixé à 6 dB, puis 3 dB après le troisième *reversal*. La procédure en escalier se terminait après le dixième *reversal* et le seuil était calculé comme étant la moyenne des 6 derniers *reversals*. La courbe d'audibilité a été obtenue en réalisant une interpolation linéaire sur des échelle fréquence et gain

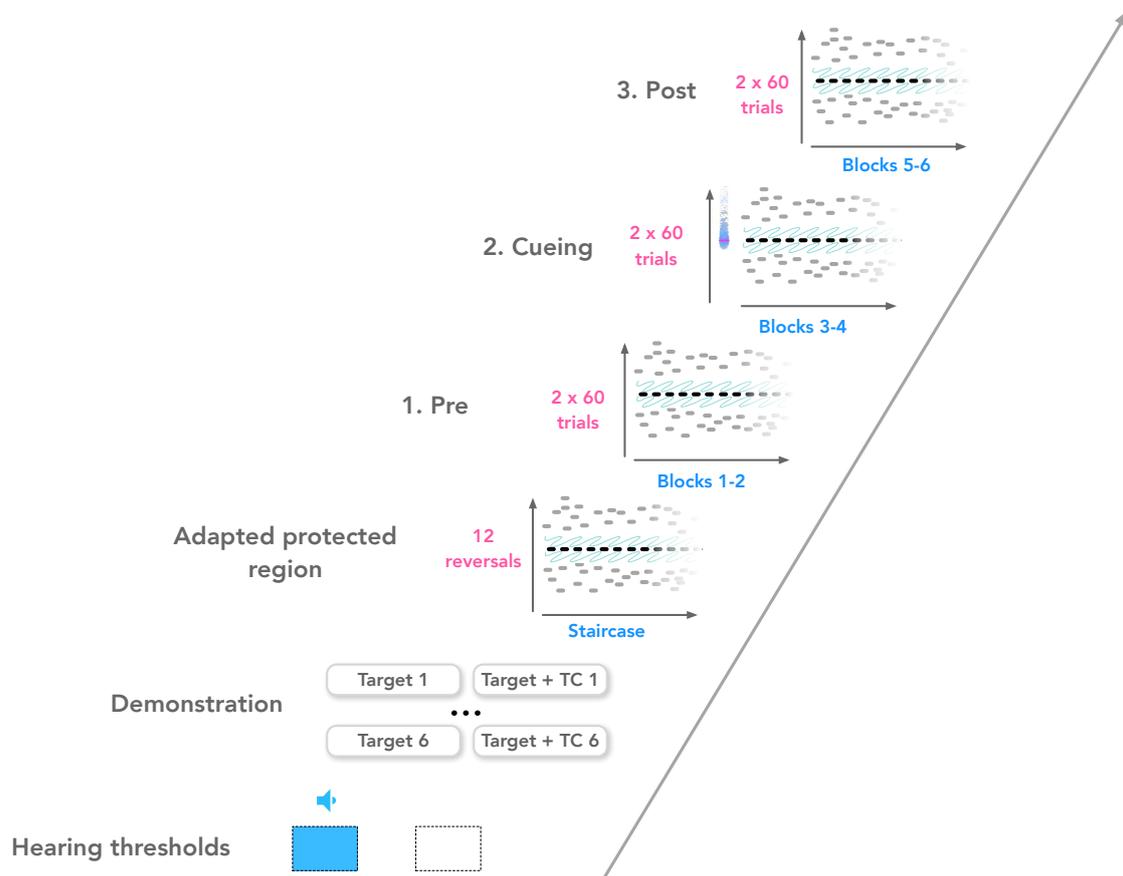


Figure 5.2: Expérience 1 - Procédure. Les participants ont tout d'abord réalisé un audiogramme visant à évaluer leurs seuils auditifs pour différentes fréquences. Ensuite, des exemples de cibles seules puis avec masqueurs leur ont été présentés. Puis une procédure adaptative a été utilisée de manière à choisir la largeur de la région protégée. Enfin, les participants ont réalisé la tâche principale séparée en 3 conditions contenant 2 sessions de 60 essais chacune.

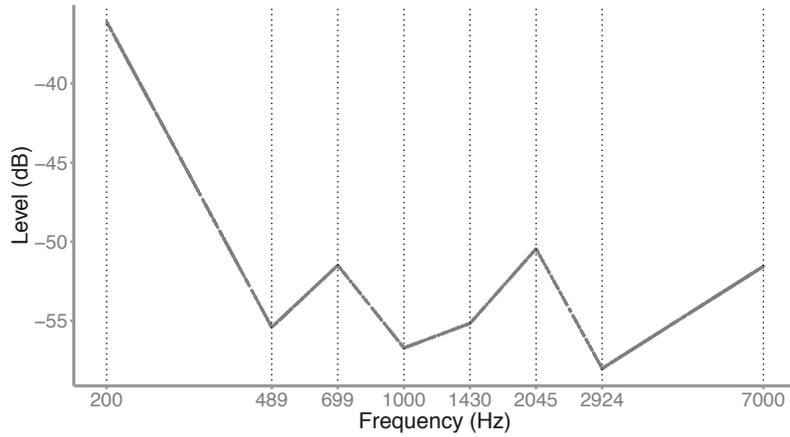


Figure 5.3: Expérience 1 - Exemple de courbe audiométrique d'un participant montrant les seuils de détection d'un son pur à 200, 489, 699, 1000, 1430, 2045, 2924 et 7000 Hz. La pondération du niveau des stimuli pour les phases ultérieures de l'expérience s'effectue selon la courbe présentée : un gain élevé correspond à un seuil élevé. On voit par exemple que le seuil était le plus élevé à 200 Hz et le plus bas à 2924 Hz.

logarithmiques pour les valeurs qui n'ont pas été testées. La courbe obtenue pour chaque participant a ensuite été utilisée pour pondérer les niveaux des tons dans les étapes suivantes de l'expérience. Un exemple de courbe audiométrique obtenue par cette méthode est présentée sur la Figure 5.3. L'ensemble des stimuli des phases suivantes a été présenté à un niveau de 45 dB au dessus du seuil (45 dB SL, *Sensation Level*).

Phase démonstration

De manière à familiariser les participants avec les stimuli, ils ont entendu chacune des 6 cibles possibles, seules puis en présence d'un nuage de tons. L'ordre des cibles était aléatoire. Une région protégée de 1800 cents (1.5 octave) a été utilisée. À l'issue de cette phase, l'expérimentateur s'est assuré que la tâche était comprise et que le participant n'avait pas de question.

Adaptation de la région protégée

De manière à compenser la variabilité inter-individuelle attendue dans les tâches de masquage informationnel, et pour éviter les performances extrêmes au plancher ou au plafond, la largeur de la bande protégée nécessaire pour que les performances

atteignent 70.7% de réponse correctes (Levitt, 1971) a été estimée pour chaque participant. Une procédure adaptative *2-down/1-up* a été appliquée à la largeur de bande. La largeur initiale était de 1200 cents (1 octave) et le pas initial était un facteur 1.5 puis 1.25 après que le troisième *reversal* ait été atteint. La largeur de bande était contrainte entre 0 et 2400 cents (2 octaves). La procédure prenait fin lorsque 10 *reversals* avaient été atteints. Les seuils ont été calculés comme étant la moyenne des 6 derniers *reversals*. La fréquence de la cible était aléatoirement choisie à chaque essai.

Masquage informationnel

La tâche de masquage informationnel était similaire à la tâche de l'expérience du chapitre précédent. La largeur de la région protégée correspondait à la valeur seuil obtenue lors de la procédure adaptative et le niveau de chaque son pur (de la cible et des nuages de tons) était pondéré selon la courbe audiométrique du participant. Les deux premières sessions sont désignés PRÉ et les deux dernières POST et ne présentaient pas d'amorçage. Les deux sessions centrales désignées AMORCE présentaient une amorce en début d'essai.

5.3.1.5 Analyse de données

Les réponses ont été converties en pourcentages de détections correctes et de fausses alarmes et l'index de sensibilité d' a été calculé pour chaque participant, chaque condition (PRÉ, AMORCE ou POST) et chaque fréquence de la cible (489, 699, 1000, 1430, 2045 ou 2924 Hz). Lorsqu'une valeur de détections correctes ou de fausses alarmes était de 0% ou de 100%, 0.5 essai à été ajouté ou retiré de manière à rendre possible le calcul du d' . Des tests statistiques ont été réalisés sur les valeurs de d' à l'aide d'ANOVA à mesures répétées avec les facteurs "condition" (PRÉ, AMORCE ou POST) et "fréquence" (489, 699, 1000, 1430, 2045 ou 2924 Hz) comme facteurs intra-participants. L'homogénéité des variances entre les différentes conditions (sphéricité) a été vérifiée et des corrections de Greenhouse-Geisser ont été appliquées lorsque nécessaire. Les *p-values* des t-tests réalisés pour les comparaisons *post-hoc* ont été corrigées par la méthode de Holm-Bonferroni.

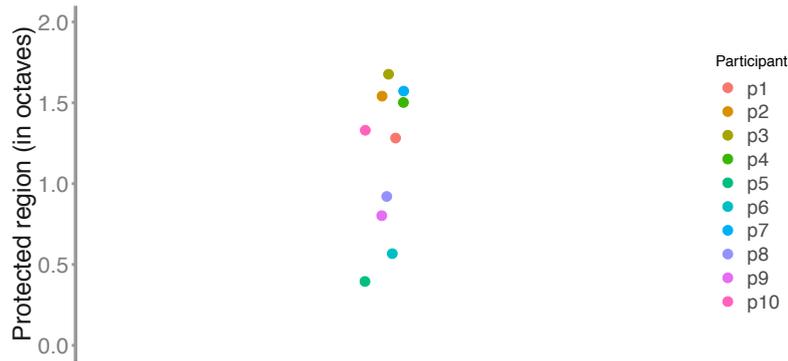


Figure 5.4: Expérience 1 - Régions protégées des participants. Seuils (en octave) obtenus à la tâche *Adaptation de la région protégée*.

5.3.2 Résultats

La Figure 5.4 montre le seuil de chaque participant obtenu à la phase *Adaptation de la région protégée*. Ils indiquent la largeur de la région protégée donnant lieu approximativement à 70.7% de réponses correctes.

La Figure 5.5 présente les performances moyennes des participants (en d') en fonction de la fréquence de la cible, pour les trois conditions successives. On peut voir que l'effet de la fréquence observé dans l'expérience du chapitre précédent semble être présent. On voit par exemple un d' égal à 1.96 dans la condition PRÉ pour les cibles à 699 Hz, et égal 0.619 dans la condition PRÉ pour les cibles à 2924 Hz. Les tests statistiques ont montré un effet de la fréquence de la cible sur les performances [$F(5, 45) = 7.987, p < 0.001, \eta_g^2 = 0.201$]. On voit également que les conditions AMORCE puis POST sont associées à de meilleures performances globales mais que l'effet de fréquence reste similaire. Cela est appuyé par l'absence d'interaction significative entre les facteurs condition (PRÉ, AMORE ou POST) et fréquence de la cible [$F(10, 90) = 0.687, p > 0.05, \eta_g^2 = 0.0153$].

La Figure 5.6 montre les performances individuelles des participants en fonction de la fréquence. Une certaine variabilité caractérise les résultats bien que dans l'ensemble, les participants semblent avoir de moins bonnes performances dans les hautes fréquences.

La Figure 5.7 montre les performances des sessions PRÉ et POST, en moyennant les résultats à travers les fréquences. Les performances sont meilleures en POST (2.04 d') qu'en PRÉ (1.41 d'). Des tests statistiques ont confirmé un effet de la condition (PRÉ,

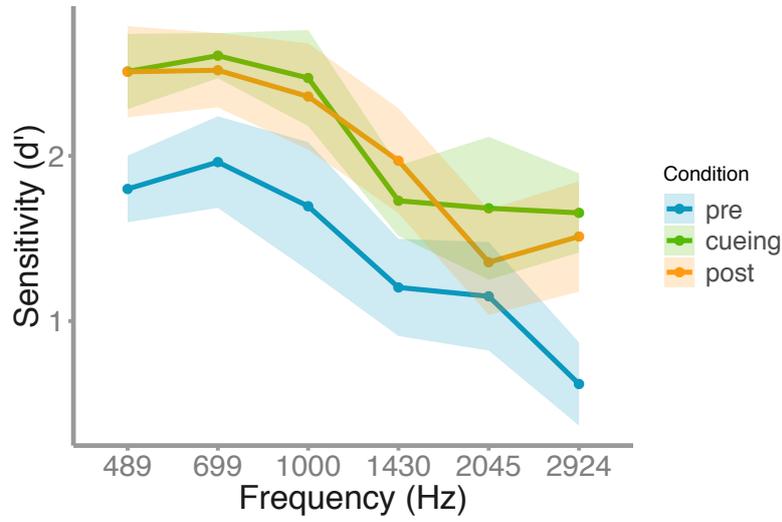


Figure 5.5: Expérience 1 - Effet de la fréquence sur les performances. Les performances (en d') sont représentées en fonction de la fréquence de la cible pour les conditions PRÉ (bleu), AMORCE (vert) et POST (orange). L'échelle de l'axe x est logarithmique.

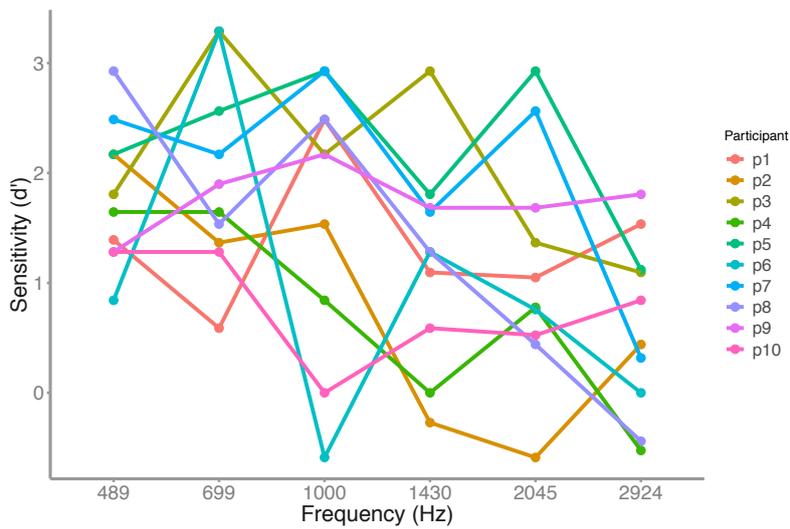


Figure 5.6: Expérience 1 - Performances individuelles des participants en fonction de la fréquence de la cible. L'échelle de l'axe x est logarithmique.

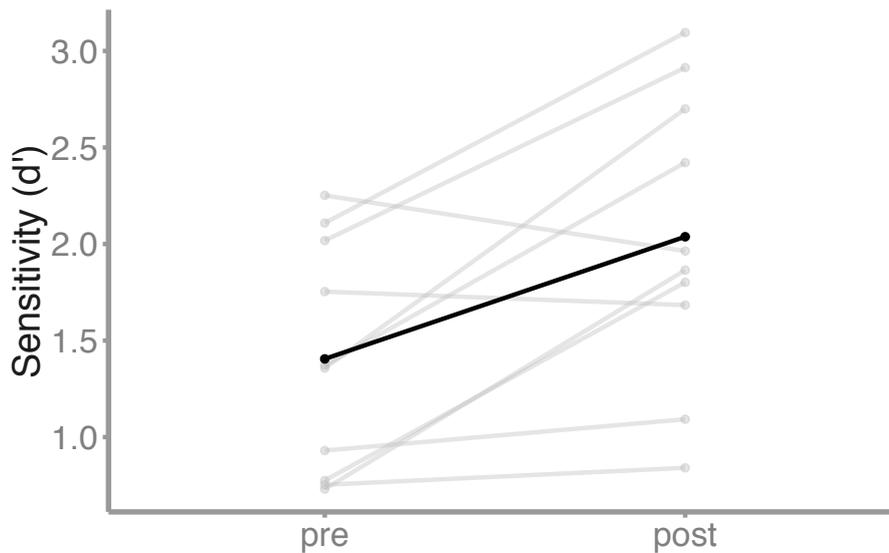


Figure 5.7: Expérience 1 - Effet de la session sur les performances. Les performances (en d') sont représentées en fonction de la session (PRÉ ou POST).

AMORCE et POST) sur les performances [$F(2, 18) = 13.54, p < 0.001, \eta_g^2 = 0.117$]. Des test post-hoc (t-test) ont été réalisés et ont montré une différence entre PRÉ et POST [$t(9) = -3.388, p = 0.024$] et entre PRÉ et AMORCE [$t(9) = -8.5654, p < 0.001$] mais pas entre POST et AMORCE [$t(9) = 0.453, p > 0.05$].

5.3.3 Discussion

Cette expérience a permis de montrer que les différences liées à l'audiogramme des participants n'était pas à l'origine de l'effet de la fréquence de la cible observé dans le chapitre précédent. Cet effet a été ainsi observé malgré une pondération de l'ensemble des stimuli selon l'audiogramme. Cette pondération, désignée en général sous le nom d'échelle SL (*sensation level*), devrait en première approximation égaliser les niveaux subjectifs des différents sons. On peut toutefois noter que cette égalisation SL n'est pas nécessairement la meilleure pondération, dans le sens où l'ajout d'une valeur identique (dans notre cas 45 dB) ne conduira pas nécessairement à la même sensation selon la fréquence. Une pondération selon les courbes d'isophonie aurait également pu être utilisée. Néanmoins, des différences de sonie ne semblent pas pouvoir expliquer

l'effet : les fréquences graves sont moins bien détectées que les médium, alors que l'inverse serait prédit par le recrutement de sonie pour les fréquences graves.

Un autre résultat inattendu est que la présence de l'amorce n'a pas permis de réduire cet effet ni dans la condition POST ni dans la condition AMORCE elle-même. Cela suggère que cet effet n'est pas lié à l'incertitude associée à la cible. Une interprétation liée à l'incertitude aurait été que chaque auditeur porte spontanément son attention sur une région fréquentielle, qui serait donc favorisée dans la tâche de masquage informationnel. Il serait alors plausible qu'une amorce indiquant la zone fréquentielle de la cible réduise voire élimine l'effet de fréquence. L'amorce n'ayant en fait aucun impact sur l'effet de fréquence, cette interprétation est donc largement rejetée : soit l'effet de fréquence ne dépend pas d'une attention portée "par défaut" sur une certaine zone de fréquence, soit l'amorce n'a pas permis de focaliser l'attention sur la zone de la cible. Notons que la comparaison des conditions avec et sans amorce a montré que les performances sont globalement meilleures avec amorce. Il semble donc que l'amorce ait eu une influence sur l'attention, mais que les bénéfices liés à la présence de l'amorce sont équivalents pour l'ensemble des fréquences utilisées. On ne peut exclure que l'utilisation d'une amorce donnant une indication encore plus précise de la fréquence de la cible (par exemple, une occurrence exacte de la cible) aurait permis des performances encore meilleures, mais des données pilotes indiquent que ceci pose des problèmes expérimentaux en termes d'effets de plafond dans certaines conditions.

La Figure 5.8 montre que les performances globales des participants étaient meilleures dans l'expérience 1 par rapport au chapitre précédant. Une différence majeure entre les deux protocoles est que l'adaptation a été réalisée sur la densité du nuage dans le chapitre précédent et sur la taille de la région protégée dans cette expérience. Il est donc possible que cette différence explique la différence de performances. On voit cependant que le point d'inflexion de la baisse de performance en fonction de la fréquence semble correspondre à 1430 Hz dans les deux cas. On peut ainsi noter que cet effet et son ampleur ne semble pas dépendre des performances de base des participants.

5.4 Expérience 2

Une explication alternative à l'effet de fréquence est que certaines cibles sont plus difficiles à percevoir en raison de leur position relative à la distribution du nuage de tons. Par exemple, il se pourrait que les cibles positionnées de manière centrale

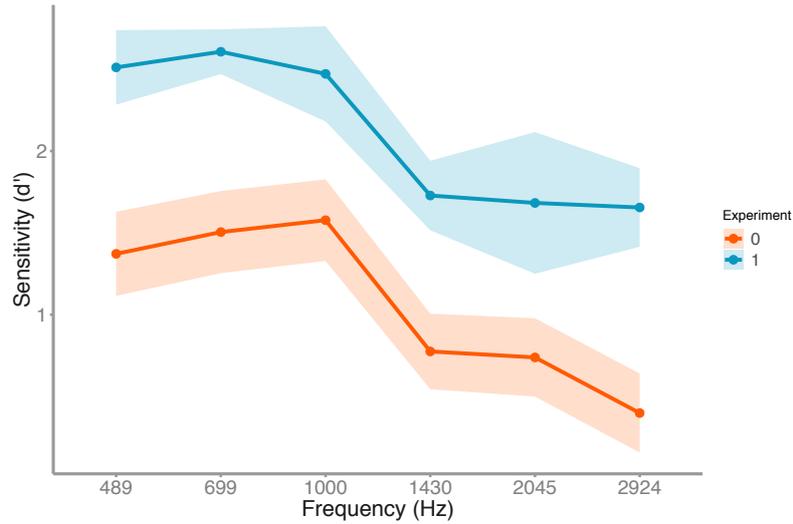


Figure 5.8: Expérience 1 et chapitre précédent. Résultats de l'expérience 1 reproduits avec ceux de l'expérience comparable du chapitre précédent. Comparaison de l'effet de la fréquence sur les performances des participants (en d') entre les sessions 2 et 3 de l'expérience du chapitre précédent et les sessions AMORCE de l'expérience 1. L'échelle de l'axe x est logarithmique.

par rapport à la distribution du nuage de tons soient plus aisément isolées. Une autre explication pourrait être qu'il est plus difficile d'ignorer les éléments de basses fréquences du nuage de tons. Pour les cibles aiguës, la majorité des éléments du masqueur est inférieure à la fréquence de la cible. La fréquence moyenne du nuage de tons est également plus basse que pour les cibles graves car la région protégée est plus aiguë. De manière à tester ce type d'interprétation, une deuxième expérience a été mise en place dans laquelle deux distributions de nuages de tons ont été comparées pour des cibles de fréquences absolues identiques.

5.4.1 Méthodes

5.4.1.1 Participants

Dix participants (9 femmes ; âge en années $M=26.1$, $SD=4.89$) normo-entendants ont été payés 10 euros de l'heure pour participer à cette étude. Ils ont été sélectionnés de manière à avoir des seuils auditifs de 20 dB HL ou moins pour les fréquences testées (250, 500, 1000, 2000, 3000, 4000, et 8000 Hz). Tous les participants étaient non-musiciens (moins de deux ans de pratique musicale au moment de l'étude).

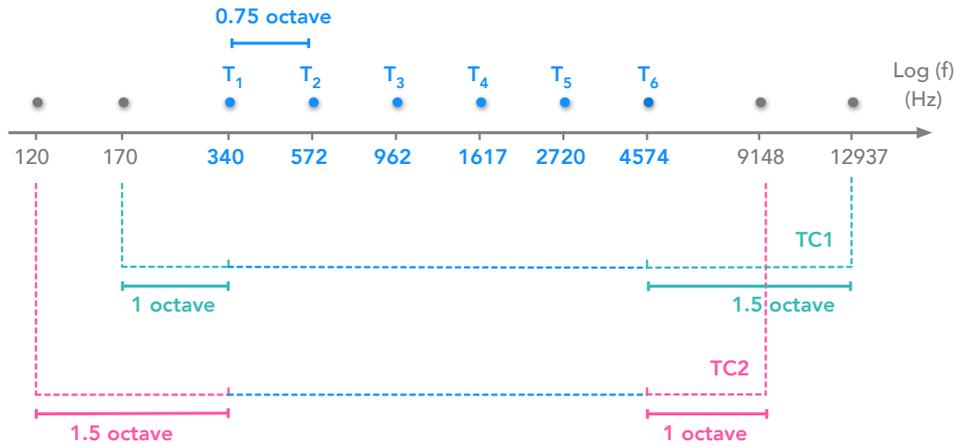


Figure 5.9: Expérience 2. Illustration des stimuli et des distributions des nuages de tons. Les cibles (en bleu) étaient séparées par 0.75 octaves et de mêmes fréquences pour les deux conditions TC1 et TC2. Dans la première condition (TC1), les fréquences des nuages de tons étaient choisies dans une distribution uniforme entre 170 et 12937 Hz. Dans une seconde condition (TC2), elles étaient décalées de 0.5 octave, donc choisies dans une distribution uniforme entre 120 et 9148 Hz.

5.4.1.2 Stimuli

Mis à part les différences décrites dans cette section, les stimuli sont similaires à ceux de l'expérience 1.

Masqueurs

Les nuages de tons étaient similaires à ceux utilisés dans l'expérience 1 mais les distributions utilisées étaient différentes. Dans une première condition (TC1), les fréquences des tons des nuages étaient comprises entre 170 et 12937 Hz. Dans une seconde condition (TC2), elles étaient comprises entre 120 et 9148 Hz. La Figure 5.9 illustre la différence entre TC1 et TC2. On voit que les cibles sont identiques pour TC1 et TC2 mais que TC1 s'étend d'une octave inférieure à la cible la plus grave à 1.5 octave supérieure à la cible la plus aiguë alors que TC2 s'étend d'une octave et demi inférieure à la cible la plus grave à une octave supérieure à la cible la plus aiguë. Cette différence de 0.5 octave, assez faible, a été choisie comme compromis pour avoir des cibles de fréquences assez basses.

Cibles

Les cibles étaient de fréquences identiques dans les deux conditions TC1 et TC2. Elles étaient espacées régulièrement sur une échelle logarithmique et séparées par 0.75 octave. Leurs fréquences étaient les suivantes : 340, 572, 962, 1617, 2720 et 4574 Hz.

5.4.1.3 Matériel et procédure

Le matériel était identique à celui de l'expérience 1. Les tâches *Audiogramme*, *Démonstration* et *Adaptation de la région protégée* étaient également identiques.

Dans la tâche *Audiogramme*, les fréquences utilisées étaient celles des cibles (340, 572, 962, 1617, 2720, 4574 Hz). Les fréquences extrêmes des nuages de tons TC1 et TC2 ont également été testées (120 et 12937 Hz).

Dans la tâche principale *Masquage Informationnel*, les participants ont réalisé 6 sessions de 60 essais chacune. La tâche était identique aux PRÉ ou POST de l'expérience 2 et aucune amorce n'a été utilisée. Les conditions TC1 et TC2 ont été alternées à chaque session. Pour la session 1, TC1 a été utilisé pour les participants ayant réalisé l'*Adaptation de la région protégée* avec TC2 et inversement, TC2 a été utilisé pour les participants ayant réalisé l'*Adaptation de la région protégée* avec TC1.

5.4.1.4 Analyse de données

Les réponses ont été converties en pourcentages de détections correctes et de fausses alarmes et l'index de sensibilité d' a été calculé pour chaque participant, chaque nuage de tons (TC1 ou TC2) et chaque fréquence (489, 699, 1000, 1430, 2045 ou 2924 Hz). Lorsqu'une valeur de détections correctes ou de fausses alarmes était de 0% ou de 100%, 0.5 essai a été ajouté ou retiré de manière à rendre possible le calcul du d' . Des tests statistiques ont été réalisés sur les valeurs de d' à l'aide d'ANOVA à mesures répétées avec les facteurs distribution et fréquence comme facteur intra-participants. L'homogénéité des variances entre les différentes conditions (sphéricité) a été vérifiée et des corrections de Greenhouse-Geisser ont été appliquées lorsque nécessaire.

De manière à analyser l'effet de la session (1, 2, 3, 4, 5 ou 6) sur les performances, les pourcentages de réponses correctes ont été utilisés de manière à éviter de calculer les d' avec un nombre trop faible de valeurs (si on utilise les facteurs nuages de tons, fréquence, participants et sessions). Des tests statistiques ont également été réalisés sur ces pourcentages à l'aide d'ANOVA à mesures répétées avec le facteur session comme facteur intra-participants. L'homogénéité des variances entre les différentes



Figure 5.10: Expérience 2 - Régions protégées des participants. Seuils obtenus à la tâche *Adaptation de la région protégée*.

conditions (sphéricité) a également été vérifiée et des corrections de Greenhouse-Geisser ont été appliquées lorsque nécessaire.

5.4.2 Résultats

La Figure 5.10 montre le seuil de chaque participant obtenu à la tâche *Adaptation de la région protégée*.

La Figure 5.11 montre les performances des participants à la tâche principale en fonction de la fréquence de la cible pour les conditions TC1 et TC2. Les fréquences sont différentes de celles de l'expérience 1 mais on voit à nouveau un effet de la fréquence. Cela a été confirmé par les tests statistiques [$F(5, 45) = 6.493, p < 0.001, \eta_g^2 = 0.206$].

La Figure 5.12 montre les performances individuelles des participants en fonction de la fréquence. On remarque une certaine variabilité, notamment dans les fréquences intermédiaires, bien que l'effet semble être similaire pour la plupart des participants dans le cas des fréquences extrêmes en comparaison des fréquences intermédiaires.

La Figure 5.13 montre l'évolution des performances au fur et à mesure des sessions. On voit une tendance à l'augmentation des performances entre la session 1 (65.67% de réponses correctes) et la session 6 (78.5% de réponses correctes). Cette tendance a été confirmée par des tests statistiques [$F(5, 45) = 9.627, p < 0.001, \eta_g^2 = 0.198$].

Enfin, il semble que les deux distributions de nuages de tons ont donné lieu à des performances similaires (Figure 5.11). En effet, les tests statistiques n'ont pas mis en

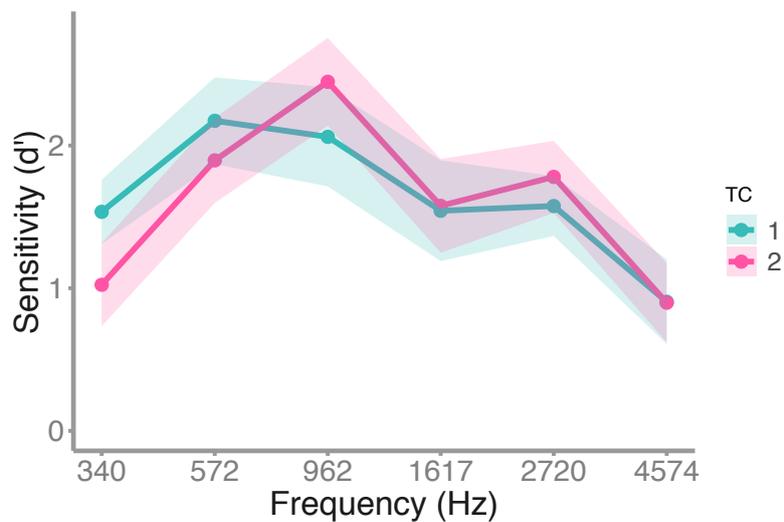


Figure 5.11: Expérience 2 - Effet de la fréquence sur les performances. Les performances (en d') sont représentées en fonction de la fréquence de la cible. L'échelle de l'axe x est logarithmique.

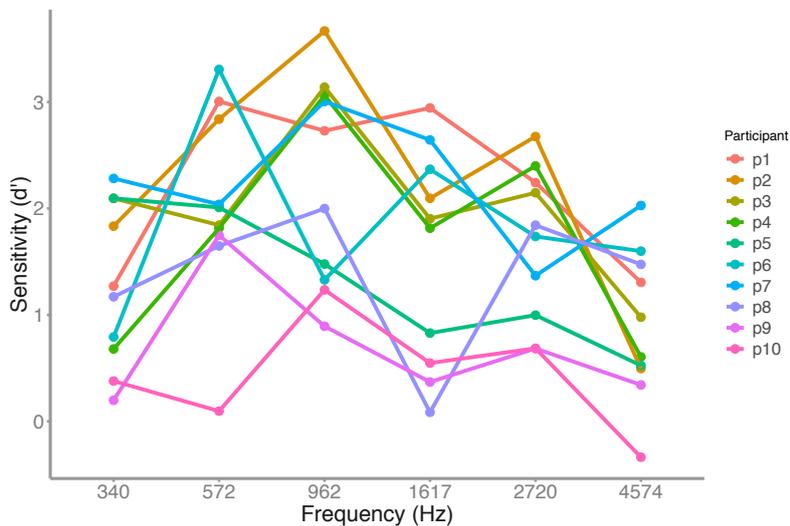


Figure 5.12: Expérience 2 - Performances individuelles des participants en fonction de la fréquence de la cible. L'échelle de l'axe x est logarithmique.

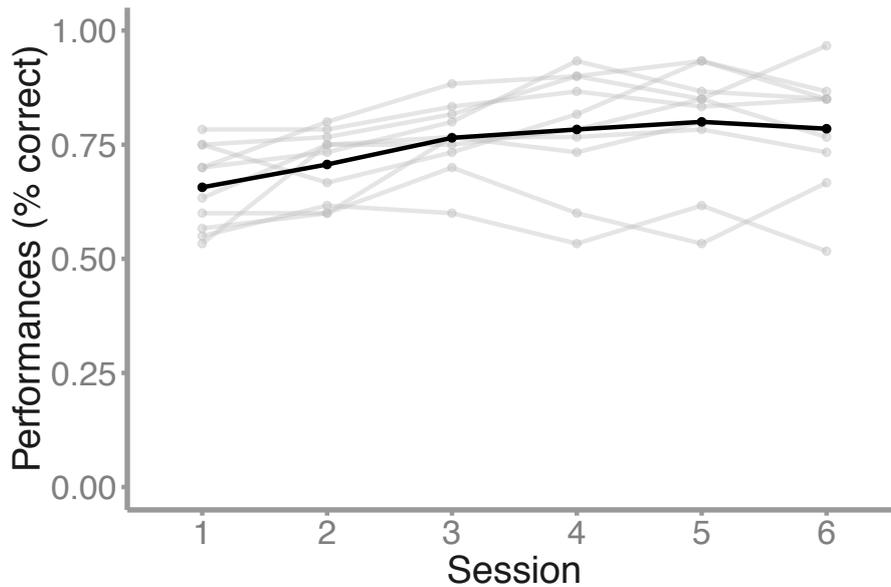


Figure 5.13: Expérience 2 - Effet de la session sur les performances. Les performances (en % de réponses correctes) sont représentées en fonction de la session (1, 2, 3, 4, 5 ou 6). Les données individuelles sont représentée en gris.

avant un effet significatif de la distribution [$F(1, 9) = 0.0671, p > 0.05, \eta_g^2 = 0.000262$] ni d'interaction significative entre la fréquence de la cible et la distribution du nuage de tons [$F(5, 45) = 2.202, p = 0.07, \eta_g^2 = 0.0274$].

5.4.3 Discussion

Cette expérience a confirmé les observations de l'expérience du chapitre précédent et les résultats de l'expérience 1 concernant l'effet de la fréquence de la cible sur le masquage informationnel. Elle a montré en outre que les performances liées aux cibles de basses fréquences étaient également dégradées par rapport aux fréquences intermédiaires. Il semble donc que les performances sont maximales pour les cibles entre 500 et 1000 Hz mais décroissent au-dessous de 500 Hz et au-dessus de 1000 Hz. On peut se demander si les performances seraient encore plus basses pour des fréquences plus extrêmes mais cela est délicat à mesurer. En effet, dans cette dernière expérience, les cibles utilisées impliquaient des sons purs du masqueurs allant de 120 à 12937 Hz et il est difficile d'utiliser des valeurs plus extrêmes tout en ajustant les niveaux de sensation (SL).

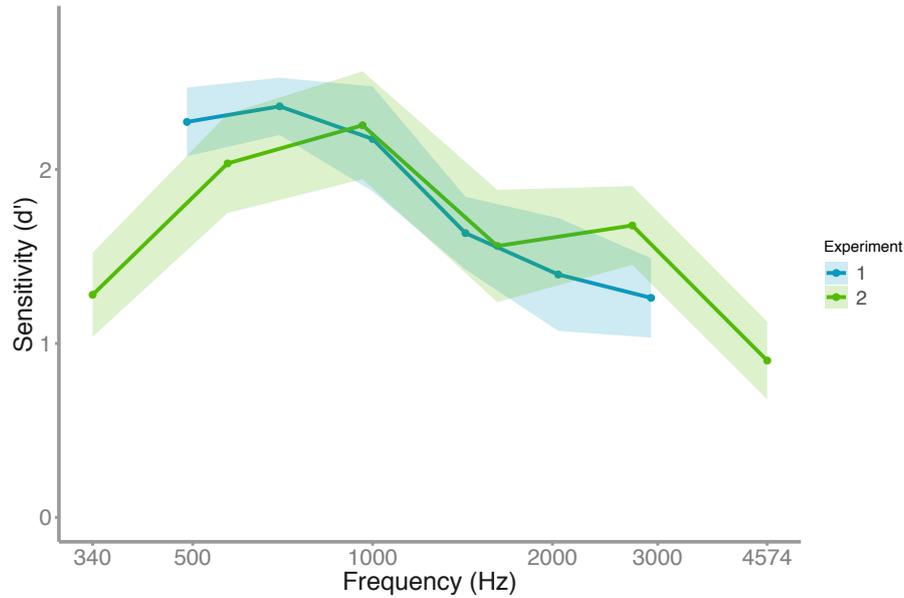


Figure 5.14: Expériences 1 et 2 - Comparaison des effets de la fréquence de la cible entre les expériences 1 et 2. L'ensemble des sessions des expériences a été prise en compte (avec/sans amorce moyenné pour l'expérience 1 et TC1/TC2 moyenné pour l'expérience 2). L'échelle de l'axe x est logarithmique.

La comparaison des résultats des expériences 1 et 2 (Figure 5.14), montre également un motif similaire.

5.5 Discussion générale

Un point important mis en avant par notre série d'expériences est que la quantité de masquage informationnel est particulièrement dépendante de la fréquence de la cible. Nous avons montré que cet effet robuste était répliqué dans les deux expériences présentées. Dans l'expérience 1, nous avons rejeté l'hypothèse selon laquelle cet effet de fréquence proviendrait d'une différence d'audibilité. L'ensemble des stimuli (cibles et nuages de tons) ont été présenté à un niveau de sensation équivalent, et l'effet de fréquence a malgré tout été observé. Dans l'expérience 2, nous avons testé l'hypothèse d'une influence de la distribution du nuage de tons en fonction des fréquences de la cible. Pour cela, deux types de nuages de tons possédant des bornes différentes ont été utilisés avec des cibles de fréquences absolues identiques. L'absence d'effet suggère, avec les réserves liées à l'interprétation de résultats négatifs, que ce n'est pas la distribution du nuage de tons qui est à l'origine de cet effet de fréquence.

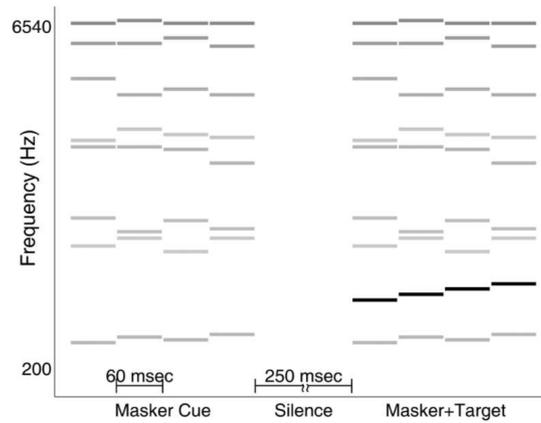


Figure 5.15: Discussion - Issu de Kidd et al. (2011). Dans cette tâche de masquage informationnel, les participants devaient identifier un motif parmi 6 possibilités. Le motif était présenté en présence de nuages de tons.

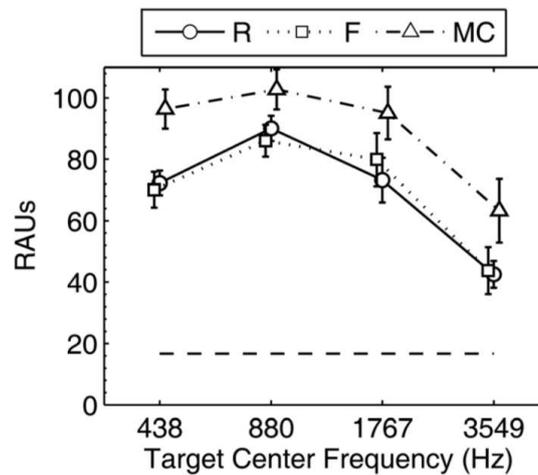


Figure 5.16: Discussion - Issu de Kidd et al. (2011). Performance (*rationalized arcsine units*) en fonction de la fréquence centrale de la cible dans une tâche d'identification de séquences mélodiques avec masquage informationnel (complexes multi-tons). On peut voir une baisse des performances pour les fréquences plus hautes, similaire à ce que nous avons observé dans ce chapitre.

On peut également noter la proximité des résultats avec ceux de Kidd et al. (2011). La tâche des participants était d'identifier des motifs mélodiques en présence de masqueurs multi-tons (Figure 5.15). Les résultats ont montré des performances dégradées pour les hautes fréquences (Figure 5.16). Selon les auteurs, ces résultats s'expliquent par la nature de la tâche de détection de motifs mélodiques. Dans une étude précédente (Kidd et al., 1998a), ils ont montré que les performances à une tâche de détection de séquences mélodiques dépendaient de l'ambitus de la séquence. Une séquence mélodique avec des intervalles plus larges, et donc d'un ambitus globale plus élevée, conduit à de meilleures performances. Ainsi, la même séquence présentée dans une bande de fréquences élevées aurait un ambitus plus faible que dans une bande de fréquences basses, relativement à la bande de fréquence centrale. Cependant, nos résultats vont à l'encontre de cette explication, étant donné que cet effet est observé dans notre tâche qui n'implique pas la discrimination de motif mélodique. Aux vues de nos résultats, les résultats de Kidd et al. (2011) suggèrent que l'effet de fréquence n'est pas propre à l'utilisation de nuages de tons, mais pourrait être observé de manière similaire avec des complexes multi-tons.

Enfin, il semble intéressant de faire le rapprochement entre les résultats présentés dans ce chapitre et les travaux sur les bandes attentionnelles. Le concept de bandes critiques ou filtres auditifs (Fletcher, 1940) explique que des sons de fréquences différentes vont être traités de manière différente au niveau de la cochlée. Cela explique également que des sons de fréquences proches se masqueront de manière plus importante que des sons de fréquences éloignées. Au niveau comportemental, la forme de ces filtres auditifs a été mise en avant par la méthode *notched-noise* (Patterson, 1976). Cette méthode consiste à présenter une cible en présence d'un bruit blanc. Le bruit blanc est filtré de manière à ce qu'une encoche soit présente autour de la fréquence de la cible. Les seuils sont calculés pour différentes largeurs de la région protégée, et on constate que l'effet masquant du bruit diminue lorsque la région protégée augmente.

Par ailleurs, la sensibilité du système auditif peut être modulée par des mécanismes attentionnels. Ainsi, la méthode *probe-signal* (Greenberg & Larkin, 1968) a été utilisée pour montrer que lorsque l'attention est portée à une certaine fréquence, les seuils de détection de fréquences éloignées augmentent. Dans cette procédure, les participants entendent la cible seule (amorçage) suivie de deux intervalles : un bruit blanc seul dans l'un et un bruit blanc avec cible dans l'autre. Les performances diminuent lorsque la fréquence de la cible s'éloigne de celle de l'amorçage, suggérant une zone fréquentielle attentionnelle. Cette méthode a été utilisée pour montrer que les bandes

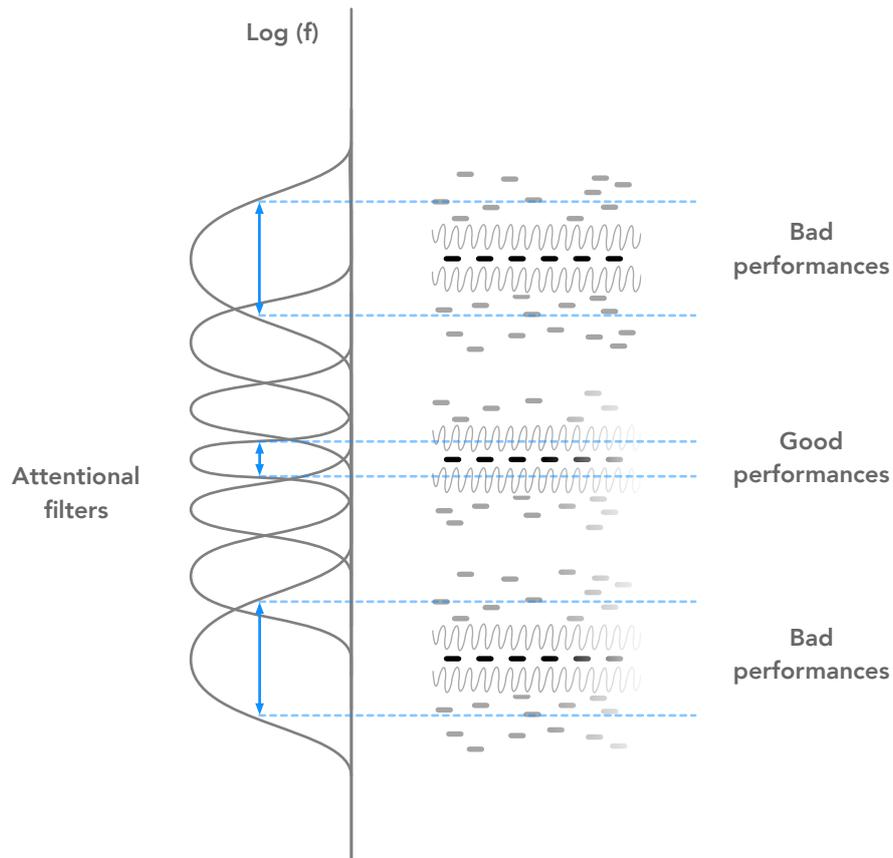


Figure 5.17: Discussion - Illustration schématique de la relation entre le concept de filtres attentionnelles et la procédure de masquage informationnel.

attentionnelles étaient de formes similaires aux filtres auditifs à 1000 Hz (Scharf et al., 1987) puis pour différentes fréquences entre 250 et 4000 Hz (Dai et al., 1991).

Cependant, les travaux sur le masquage informationnel ont préalablement montré que des sons purs qui dépendaient de filtres auditifs différents de ceux de la cible à détecter pouvaient malgré tout avoir un effet masquant. Ces constats sont donc en contradiction avec les travaux sur les bandes attentionnelles qui suggèrent que leurs formes sont équivalentes à celles des filtres auditifs. En effet, si des processus attentionnels peuvent augmenter la saillance des sons purs qui dépendent du même filtre auditif que la cible et diminuer celle des autres, les éléments du masqueurs hors de ces filtres auditifs devraient avoir peu d'effets masquants. La Figure 5.17 illustre cette interprétation de la relation entre la tâche de masquage informationnel dans des nuages de tons et le concept de filtre attentionnel. Les performances sont dégradées lorsque des éléments du masqueur dépendent du même filtre attentionnel que la cible. Or nous avons montré dans ce chapitre que les performances dépendaient fortement de

la fréquence de la cible, avec les meilleurs résultats entre 500 et 1000 Hz. Cela suggère que les bandes attentionnelles ont des formes qui ne suivent pas nécessairement celles des bandes critiques (illustrée sur la Figure 5.17).

5.6 Conclusions

Pour conclure, des travaux supplémentaires seront nécessaires pour clarifier les mécanismes sous-jacents à cet effet de fréquence. Tout d'abord, on peut se demander si cet effet concerne l'ensemble des tâches de masquage informationnel. Les résultats de Kidd et al. (2011) suggèrent que c'est le cas pour les stimuli multi-tons. Ensuite, s'agit-il d'un effet propre au masquage informationnel, ou bien s'agit-il d'une caractéristique fondamentale de l'attention sélective auditive ? Une possibilité serait qu'il y a une pondération naturelle des différentes fréquences, qui est non sensible à l'attention ou à l'entraînement perceptif, et qui induit une plus grande saillance des fréquences médiums. Enfin, la mise en perspective de nos résultats avec les travaux sur les filtres attentionnels pourrait remettre en question l'idée de bandes attentionnelles de formes similaires à celles des bandes auditives.

6 CHAPITRE 6. Indices électrophysiologiques de l'attention sélective dans un paradigme de masquage informationnel

Une partie de ce travail a été présentée en tant que :

Jean, H., Pressnitzer, P., Di Liberto, G. (2018). Fast decoding of auditory steady-state responses in an informational masking paradigm. CuttingEEG, Paris, France (Poster).

6.1 Aperçu

Les deux chapitres précédents ont montré que les performances dans une tâche de masquage informationnel pouvaient être améliorées par apprentissage perceptif, ce qui a été attribué à un effet sur l'attention sélective. Dans ce chapitre, nous présentons des données pilotes utilisant l'électroencéphalographie (EEG) pour évaluer l'attention sélective à la cible pendant une tâche de masquage informationnel. Deux indices ont été testés pour évaluer l'attention sélective auditive : la réponse auditive soutenue (*auditory Steady-State Response* ; aSSR) et les techniques de reconstruction du stimulus (RS) à partir de données EEG. Pour ce faire, de nouveaux stimuli comportant une double modulation d'amplitude ont été élaborés. Les participants ont réalisé une tâche de détection de silence et devaient se focaliser sur une cible continue en présence d'un nuage de tons (partie 1) puis entendaient deux cibles continues simultanées et devaient se focaliser sur l'une ou sur l'autre (partie 2). Des décodeurs ont été construits (à partir de l'aSSR, de la RS ou des deux) visant à déterminer la cible continue qui faisait l'objet de l'attention. Les résultats ont confirmé qu'il est possible d'utiliser les stimuli à double modulation pour utiliser aSSR et RS de manière congruente. De plus, nous avons observé de bonnes performances du décodeur RS (jusqu'à 90% de classification correcte) mais de faibles résultats du décodeur aSSR. L'ajout de cet indice n'a donc pas permis d'améliorer le décodeur RS. Ces données pilotes (4 participants) suggèrent qu'il est possible d'utiliser les techniques de reconstruction du stimulus à partir d'enveloppe synthétiques, ce qui ouvre la possibilité d'optimiser les caractéristiques de l'enveloppe. De plus, la méthode explorée ici permet d'envisager l'élaboration de techniques de type "neurofeedback", pour lesquelles les indices EEG de l'attention seraient présentés aux auditeurs afin d'améliorer l'efficacité de protocoles d'entraînement perceptif.

6.2 Introduction

L'attention sélective auditive désigne la capacité se focaliser sur une source sonore et à ignorer les autres sons. Une manière de l'évaluer est d'utiliser des protocoles dans lesquels une cible doit être détectée en présence d'autres sons distracteurs. Dans les chapitres précédents, nous avons utilisé une tâche de masquage informationnel où les participants devaient détecter des sons purs répétés cachés dans des nuages de tons. On parle de masquage informationnel car une région fréquentielle est protégée et ne comporte pas d'éléments du masqueur. Cela limite le masquage énergétique qui a lieu lorsque cibles et masqueurs dépendent des mêmes filtres cochléaires, et permet donc de mettre en avant d'autres processus comme l'attention sélective. Dans le chapitre précédent, nous avons observé un apprentissage perceptif pour ce type de tâche, et fait l'hypothèse que cet apprentissage agissait sur l'attention sélective. Dans ce chapitre, nous présentons une nouvelle méthode et des données pilotes pour évaluer l'attention sélective dans une tâche de masquage informationnel.

Différents indices électrophysiologiques ont été utilisés pour évaluer les processus d'attention sélective auditive. L'*auditory Steady-State Response* (aSSR) a notamment été utilisée. Il s'agit d'une réponse observée en électroencéphalographie (EEG) ou magnétoencéphalographie (MEG) associée à un stimulus stationnaire, comme des sons purs modulés en fréquence ou en amplitude par exemple (Picton et al., 2003a). Cette réponse cérébrale présente une oscillation dominante à la fréquence de modulation acoustique du signal entendu. Il est en outre possible d'examiner l'aSSR correspondante à des stimuli présentés simultanément et modulés à des fréquences différentes (John et al., 1998 , 2002). Par ailleurs, une réduction de son amplitude est observée dans des cas de masquage, même d'origine centrale, comme c'est le cas pour un masqueur présenté de manière contra-latérale (Kawase et al., 2012 ; Ross et al., 2012). Il semble toutefois que l'aSSR puisse être observée malgré la présence d'un nuage de tons (Gutschalk et al., 2008, expérience 2).

L'effet de l'attention sélective sur l'amplitude de l'aSSR fait l'objet de résultats contradictoires. Certaines études suggèrent que l'attention sélective auditive ne permet pas de moduler l'amplitude de l'aSSR (Linden et al., 1987). Dans cette étude, les participants écoutaient des séquences de tons pulsés à une fréquence entre 37 et 41 Hz. Les auteurs ont comparé l'aSSR produite lorsque les participants réalisaient une tâche n'impliquant pas l'attention au stimulus (lire un livre) et lorsque la tâche concernait le stimulus (compter le nombre de fois où une augmentation de l'intensité était perçue). Aucune différence de l'amplitude de l'aSSR n'a été observée

entre ces deux conditions. Pourtant, d'autres études ont montré que l'attention pouvait moduler l'amplitude de l'aSSR (Skosnik et al., 2007), notamment dans des paradigmes dans lesquels plusieurs flux sont présentés simultanément (*intracranial EEG* : Bidet-Caulet et al., 2007 ; EEG : Mahajan et al., 2014). Dans l'étude de Bidet-Caulet et al. (2007), les patients entendaient des flux simultanés modulés en amplitude à 21 ou 29 Hz et devaient focaliser leur attention sur l'un ou sur l'autre. Les résultats ont montré que l'aSSR correspondant au flux faisant l'objet de l'attention était d'amplitude plus importante. Dans l'étude de Mahajan et al. (2014), les participants entendaient des flux de bruits blancs modulés en amplitude à une fréquence de 16, 23.5, 32.5 ou 40 Hz et présentés par paire, chacun à une oreille. Les résultats montrent que l'effet de l'attention sur l'aSSR n'est pas observé dans toutes les conditions, suggérant un effet de la fréquence de modulation et de l'activation (contralatérale ou ipsilatérale).

La reconstruction du stimulus (RS) est une autre technique d'analyse utilisée pour sa sensibilité à l'attention sélective auditive. Le principe de cette méthode est de reconstruire, à partir de l'activité cérébrale évoquée, un signal le plus proche possible de l'enveloppe du stimulus entendu. Dans le cas le plus simple, un modèle linéaire prédisant l'enveloppe du stimulus à partir des signaux physiologiques est appris par un algorithme minimisant par exemple l'erreur entre prédiction et stimulus sur une partie des données (Crosse et al., 2016). Il a été montré que la qualité de la reconstruction dépend de l'attention sélective, et cette approche a donc pu être utilisée pour étudier la capacité à isoler une voix parmi d'autres à partir de MEG (Ding & Simon, 2012) ou d'EEG (O'Sullivan et al., 2014). L'effet attentionnel sur la qualité de la reconstruction du stimulus à partir du signal neural permet un décodage de l'attention performant à l'échelle d'essais uniques (d'environ 60 secondes) (O'Sullivan et al., 2014).

Le but de ce chapitre est de considérer la possibilité d'utiliser les deux indices électrophysiologiques évoqués plus haut (aSSR et RS) à partir d'enregistrements EEG et ceci de manière congruente. Si cela est possible, on peut penser que l'évaluation de l'attention sélective à partir du signal EEG pourra être plus performante, et par exemple être ainsi suivie tout au long d'un paradigme d'apprentissage perceptif, pour confirmer ou infirmer l'hypothèse que cet apprentissage joue sur l'attention. Pour cela, un nouveau type de stimuli auditifs a été élaboré. Il s'agit de cibles continues construites à partir de sons purs de fréquence porteuse constante modulés par deux enveloppes d'amplitude différentes. Une première enveloppe est sinusoïdale et à pour but de produire une aSSR à la fréquence de cette sinusoïde (ici, 36 ou 44 Hz). Une

seconde enveloppe est aléatoire et plus lente (15 Hz en moyenne), et a pour but de fournir un signal unique à chaque fréquence porteuse, adapté aux algorithmes utilisés pour la reconstruction du stimulus dans le cas de la parole notamment.

Dans une première partie, la possibilité d'utiliser l'aSSR et la reconstruction du stimulus (RS) de manière congruente à partir des stimuli dans une tâche de masquage informationnel est évaluée. Dans cette étape, les participants devaient focaliser leur attention sur la cible continue comportant la double modulation. Une tâche de détection de silence introduit au hasard dans le stimulus a été demandée aux participants, pour mesurer leur attention de manière comportementale. Des conditions avec et sans nuage de tons (masquage informationnel) ont été testées pour vérifier que l'aSSR et la reconstruction pouvaient être réalisées, malgré la présence du masqueur. Autrement dit, est-ce qu'il est possible d'observer une aSSR malgré la seconde modulation d'amplitude, et est-ce qu'il est possible d'utiliser la reconstruction du stimulus malgré la modulation d'amplitude sinusoïdale ? Dans une seconde partie, nous avons tenté d'évaluer les gains éventuels liés à l'utilisation des deux méthodes d'analyse en congruence sur une tâche de décodage attentionnel. Les participants entendaient cette fois deux cibles continues à double modulation et devaient se focaliser sur l'une ou sur l'autre. De manière à évaluer les performances des méthodes d'analyse (aSSR et RS), une procédure de décodage a été mise en place. Le but des décodeurs était de déterminer pour chaque essai, laquelle des cibles continues faisait l'objet de l'attention. Trois décodeurs ont été comparés : un construit à partir des corrélations obtenues par la RS, un à partir de l'aSSR et un combinant ces deux indices. La précision des décodeurs a été évaluée à partir d'une procédure de validation croisée (*cross-validation*). Précisons que ce chapitre ne présente que des données pilotes, sur 4 participants, et vise principalement à motiver une étude plus exhaustive de la méthode dans le futur.

6.3 Méthodes

6.3.1 Participants

Quatre participants (1 femme ; âge en années $M=23.0$, $SD=5.79$) normo-entendants ont participé à cette étude. Ils ont été sélectionnés de manière à avoir des seuils auditifs de 20 dB HL ou moins pour les fréquences testées (250, 500, 1000, 2000, 3000, 4000, et 8000 Hz).

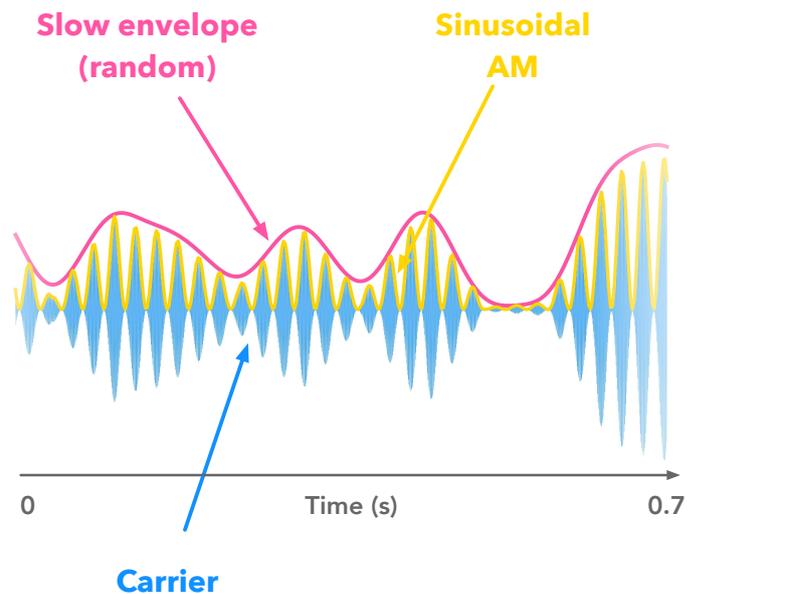


Figure 6.1: Stimuli. Cibles continues à doubles modulations. La porteuse (en bleu) est modulée par une première enveloppe lente et aléatoire (en rose) et par une seconde enveloppe sinusoïdale (en jaune).

6.3.2 Stimuli

6.3.2.1 Cibles continues

Les cibles étaient des sons purs de fréquence porteuse fixe, d'une durée correspondante à la durée de l'essai (60 secondes), et modulés simultanément par deux enveloppes d'amplitude. Une première enveloppe était sinusoïdale et a été utilisée pour induire une aSSR (en jaune sur la Figure 6.1). Une seconde modulation était appliquée au même stimulus et était lente et aléatoire (en rose sur la Figure 6.1). Elle était destinée à être reconstruite par la méthode RS. Elle a été générée en choisissant aléatoirement 900 valeurs de gain entre 0 et 1 (distribution uniforme) par essai de 60 secondes. Une enveloppe d'amplitude a alors été construite à partir de ces valeurs, après application d'un filtre passe-bas à 15 Hz. Cette valeur a été choisie plus élevée que dans le cas des enveloppes de son de parole (entre 1 et 10 Hz ; Ding & Simon, 2014), généralement utilisées dans les travaux de reconstruction du stimulus. L'hypothèse est qu'une plus grande quantité d'information serait disponible pour reconstruire le stimulus en un temps donné avec une enveloppe un peu plus rapide que la parole.

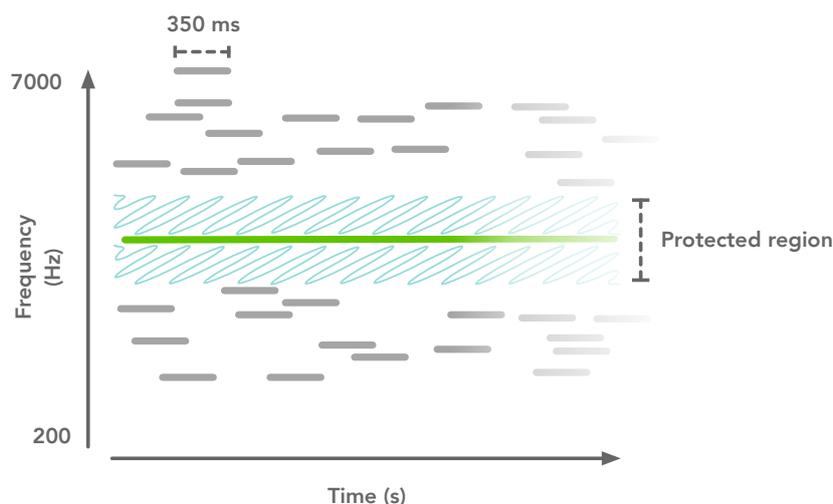


Figure 6.2: Stimuli - Cibles continues et nuages de tons. Les composants du nuage de tons (gris) sont répartis aléatoirement en temps et en fréquence. Une région fréquentielle protégée est représentée (bleu). La cible continue à double modulation (vert) est au centre de la région protégée.

Deux types de cibles continues ont été élaborées : une qui présentait une porteuse à 700 Hz et dont l'enveloppe sinusoïdale était de fréquence égale à 36 Hz (désignée CIBLE36 ci-après et représentée en bleu sur les figures) et une autre avec une porteuse à 1599 Hz et dont l'enveloppe sinusoïdale était de 44 Hz (désignée CIBLE44 ci-après et représentée en vert sur les figures). Dans les deux cas, la modulation d'amplitude aléatoire était présente et différente à chaque essai.

6.3.2.2 Nuages de tons

Les nuages de tons ont été construits d'une manière similaire à celle présentée dans le chapitre 5. Cependant, des sons purs de 350 ms (incluant des rampes ascendantes et descendantes de 15 ms chacune) ont été utilisés, et leurs caractéristiques spectrales et temporelles ont été contrôlées. Les fréquences des sons purs étaient choisies aléatoirement dans une distribution uniforme sur une échelle logarithmique entre 200 et 7000 Hz. Une région fréquentielle de 1 octave était protégée autour de la fréquence de la cible (Figure 6.2) en excluant que des sons purs soient dans cette bande de fréquence. De manière à contrôler la densité temporelle, les essais étaient séparés en *cellules temporelles*. La durée de chaque cellule temporelle était de 0.2 seconde donnant lieu à 300 cellules par essai. Deux sons purs étaient joués dans chaque cellule et leur position temporelle était choisie dans une distribution uniforme ayant pour

bornes le début et la fin de la cellule temporelle. Ces paramètres ont été choisis de manière à avoir une tâche de difficulté intermédiaire malgré la présence d'une cible continue (contrairement aux cibles discrètes utilisées dans les chapitres précédents).

6.3.3 Matériel

Les participants ont été testés individuellement dans une cabine insonorisée à double paroi (Industrial Acoustics). Les stimuli ont été joués de manière diotique par le biais d'une carte son RME Fireface UC à une fréquence d'échantillonnage de 44.1 kHz et de résolution égale à 16-bit. Les sons ont été écoutés par un casque audio Sennheiser HD 600. Les données EEG ont été enregistrées avec une fréquence d'échantillonnage de 512 Hz pendant la durée totale de l'expérience pour les 4 participants par le biais d'un appareil Biosemi Active Two avec 64 électrodes (ainsi que 2 électrodes mastoïdes) positionnées selon le système 10/20.

6.3.4 Procédure

Dans les deux parties de l'expérience, la tâche était de détecter des silences de 300 ms insérés dans les cibles continues. La présence des deux enveloppes, notamment l'enveloppe aléatoire qui pouvait passer par 0 rendait la tâche non triviale. Il y avait 4, 5 ou 6 silences dans chaque essai (probabilités équiprobables). Les participants devaient presser un bouton à chaque fois qu'ils détectaient un silence. La détection correcte des silences a permis de s'assurer que les participants pouvaient faire la tâche, et dans la deuxième partie qui comportait deux cibles concurrentes, qu'ils portaient bien leur attention à la cible continue qui était indiquée.

Dans la première partie de l'expérience, 4 blocs de 10 essais ont été réalisés, donnant lieu à environ 40 minutes de test (Figure 6.3). Les cibles continues ont été entendues séparément avec et sans masqueur. Dans tous les essais, la cible commençait 2 s avant le masqueur, ce qui permettait aux participants de focaliser leur attention sur cette cible avant l'arrivée du masqueur éventuel. Cela a donné lieu à 4 conditions expérimentales réalisées dans cet ordre :

1. CIBLE36 sans masqueur (Figure 6.3, a.)
2. CIBLE36 avec masqueur (Figure 6.3, b.)
3. CIBLE44 sans masqueur (Figure 6.3, c.)
4. CIBLE44 avec masqueur (Figure 6.3, d.)

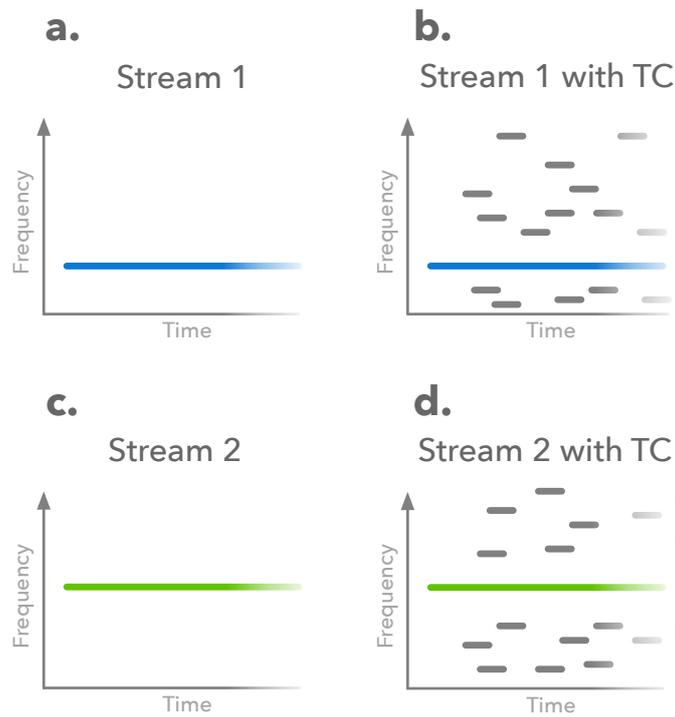


Figure 6.3: Procédure - Partie 1. Conditions expérimentales de la première tâche. Quatre blocs de 10 essais chacun ont été réalisés par les participants. a. Dans le premier bloc les CIBLE36 (en bleu) étaient entendues seules. b. Dans le bloc 2, les CIBLE36 étaient entendues en présence d'un nuage de tons (*tone cloud*, TC). c. Dans le bloc 3, les CIBLE44 (en vert) étaient entendues seules. d. Dans le bloc 4, les CIBLE44 étaient entendues en présence d'un nuage de tons.

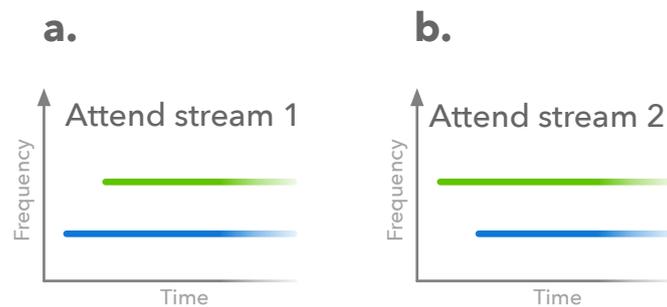


Figure 6.4: Procédure - Partie 2. Conditions expérimentales de la deuxième tâche. Les participants entendaient simultanément les CIBLE36 (en bleu) et CIBLE44 (en vert). a. Dans les blocs 1 et 2, ils devaient se focaliser sur la CIBLE36. b. Dans les blocs 3 et 4, ils devaient se focaliser sur la CIBLE44.

Dans la seconde partie, les participants ont réalisé 4 blocs de 10 essais menant à une durée d'environ 40 minutes également. Les cibles continues ont été toujours entendues simultanément et sans masqueur. Dans la première condition (blocs 1 et 2), les participants devaient prêter attention à la CIBLE36 (Figure 6.4, a.) et dans la seconde (blocs 3 et 4) à la CIBLE44 (Figure 6.4, b.). La cible à écouter commençait 2 s avant celle à ignorer. Des silences étaient présents sur les deux cibles continues mais seuls ceux présents sur la cible suivie par les participants devaient faire l'objet de réponses. Les silences des deux cibles étaient répartis pseudo-aléatoirement dans l'essai de manière à éviter qu'ils soient présents sur les deux cibles en même temps ou trop rapprochés les uns des autres (marge temporelle de 0.8 seconde minimum).

6.3.5 Analyses de données

6.3.5.1 Analyses comportementales

L'index de sensibilité d' a été calculé à partir des pourcentages de détections correctes et de fausses alarmes pour chaque essai de chaque participant. On peut noter que cela a donné lieu à un petit nombre de valeurs utilisées pour calculer un d' (entre 4 et 6 observations selon le nombre de silences dans l'essai), et que d'autres métriques seraient probablement souhaitées pour une expérience fondée sur ce travail pilote. Une réponse intervenant entre 0.1 et 1 seconde après un silence à détecter était comptabilisée comme une détection correcte. Sinon, la réponse était comptabilisée comme une fausse alarme. Un demi essai a été ajouté ou retiré pour les cas où 0% ou 100% de détections correctes ou de fausses alarmes a été observé. Seuls les essais pour lesquels le d' était supérieur à 0.5 ont été utilisés dans les analyses.

6.3.5.2 EEG

Les données EEG ont été séparées en époques correspondant à chaque essai d'une minute, excluant les deux premières secondes de l'essai, pour éviter de contaminer les analyses par les réponses classiques évoquées par la transition silence-son, et la période éventuellement sans masqueur dans les cas avec nuage de tons.

aSSR

Les données ont été filtrées avec un filtre de Chebyshev passe-bande de type II entre 1 et 100 Hz puis ont été référencées par rapport à la moyenne des deux électrodes mastoïdes. Les analyses ont été réalisées à partir des 64 électrodes moyennées. Les aSSRs ont été calculées à partir de la transformée de Fourier (FFT) des essais

moyennés dans le domaine temporel (Picton et al., 2003b) pour chaque condition. Les valeurs de pics pour 36 et 44 Hz ont été évaluées en prenant le rapport de la puissance pour des échantillons correspondant à la fréquence du pic à évaluer (moyenne des 5 échantillons inférieurs et des 5 échantillons supérieurs, correspondant environ à 0.083 Hz au-dessous et 0.083 Hz au-dessus) avec la puissance moyenne des échantillons environnant (15 échantillons inférieurs et 15 échantillons supérieurs, correspondant environ à 0.25 Hz au-dessous et 0.25 Hz au-dessus).

Reconstruction du stimulus

Les données ont été filtrées avec un filtre Chebyshev passe-bande de type II entre 1 et 15 Hz puis ont été référencées par rapport à la moyenne des deux électrodes mastoïdes. Elles ont ensuite été sous-échantillonnées de 512 Hz à 64 Hz. Les analyses ont été réalisées à partir des 64 électrodes moyennées. Le but de la reconstruction du stimulus est de trouver une estimation de l'enveloppe du stimulus qui a été entendu à partir de la réponse neurale, ici par l'intermédiaire d'un modèle de reconstruction linéaire (O'Sullivan et al., 2014). Un modèle par participant a été calculé à partir de l'ensemble des essais, par le biais de la *toolbox* mTRF (Crosse et al., 2016). Les corrélations de Pearson ont ensuite été calculées entre l'enveloppe reconstruite et l'enveloppe du stimulus réel (calculée comme détaillé plus loin). Dans la première partie, comme une seule cible a été présentée (CIBLE36 ou CIBLE44), les corrélations ont été calculées entre l'enveloppe reconstruite et l'enveloppe de cette cible. Dans la seconde partie, les corrélations entre l'enveloppe reconstruite et l'enveloppe de la cible continue qui a fait l'objet de l'attention ont été comparées, à chaque essai, aux corrélations entre l'enveloppe reconstruite et l'enveloppe de la cible continue qui devait être ignorée.

6.3.5.3 Traitement audio

L'enveloppe de chaque cible continue pour chaque essai a été extraite en utilisant le module de la transformée de Hilbert. Ces enveloppes ont ensuite été filtrées avec un passe-bas en dessous de 15 Hz pour éviter qu'elles ne suivent les modulations sinusoïdales à 36 ou 44 Hz, mais uniquement l'enveloppe aléatoire plus lente. Les enveloppes ont ensuite été sous-échantillonnées à 64 Hz pour être sur la même échelle temporelle que les données EEG.

6.3.5.4 Classification

Des algorithmes de classification utilisant la technique de machines à vecteur de support (*Support Vector Machine* ; SVM) ont été utilisés dans la partie 2 pour déterminer quelle cible continue a fait l'objet de l'attention dans chaque essai (CIBLE36 ou CIBLE44). Pour le décodeur construit à partir de l'aSSR, les données fournies aux SVM étaient les valeurs des pics d'amplitude du spectre à 36 ou 44 Hz. Pour le décodeur construit à partir de la reconstruction du stimulus, les données fournies étaient les valeurs des corrélations entre l'enveloppe reconstruite et l'enveloppe qui a fait l'objet d'attention (qu'il s'agisse de CIBLE36 ou CIBLE44) et les valeurs des corrélations entre l'enveloppe reconstruite et l'enveloppe ignorée. Enfin, pour évaluer la complémentarité des deux types d'indices, un troisième décodeur a été construit qui prend en compte l'ensemble des caractéristiques précédentes.

Pour chaque décodeur, les SVM ont été évalués par le biais d'une procédure de validation croisée (*5-fold cross-validation*). Cela signifie que l'ensemble des données a été divisé en 5 parts égales. Les SVM ont alors été entraînés sur 4 des 5 parties et testés sur la dernière, ce qui produit 5 valeurs de performances (une valeur de performance pour chacune des 5 parties utilisée comme test ; proportion d'essais correctement classifiés dans la partie test) qui sont ensuite moyennées. Les hyperparamètres des SVM ont été fixés arbitrairement et non-optimisés (paramètre de marge $C = 1$ et noyau linéaire). De manière à évaluer la précision du décodeur en fonction de la quantité de données, la précision a été calculée selon la procédure décrite ici en prenant en compte des durées d'essais différentes (de 1 à 60 secondes).

6.3.5.5 Analyses statistiques

Étant donné la nature pilote de cette expérience et du faible nombre de participants et de points de mesures, aucune analyse statistique n'a été réalisée.

6.4 Résultats

6.4.1 Partie 1

6.4.1.1 Données comportementales

La Figure 6.5 montre les performances des participants pour chaque essai de la partie 1. Les essais 1 à 10 correspondent au bloc 1 (CIBLE36 sans masqueur), les essais 10 à 20 au bloc 2 (CIBLE36 avec masqueur), les essais 20 à 30 au bloc 3

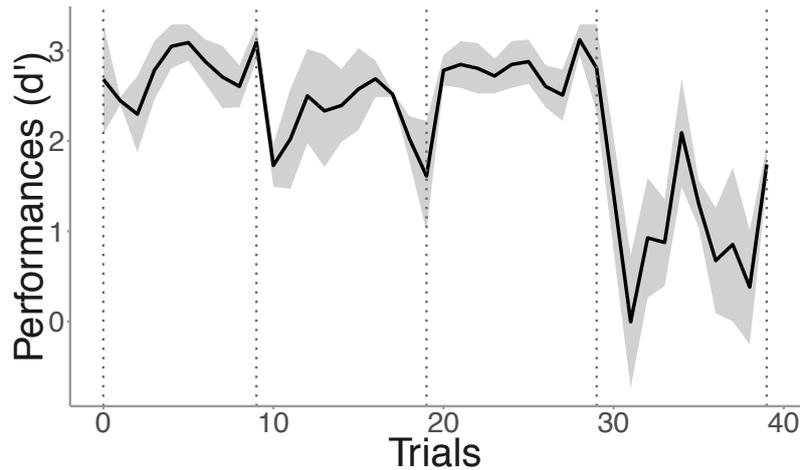


Figure 6.5: Partie 1. Données comportementales. Les performances moyennées des participants en d' sont montrées pour chaque essai des blocs 1 (essais 0 à 10), 2 (essais 10 à 20), 3 (essais 20 à 30) et 4 (essais 30 à 40). Le ruban représente l'erreur standard.

(CIBLE44 sans masqueur), et les essais 30 à 40 au bloc 4 (CIBLE44 avec masqueur). On voit que les performances sont bonnes mais qu'elles chutent dans le dernier bloc.

La Figure 6.6 montre les performances des participants moyennées pour les essais avec et sans masqueur et pour les CIBLE36 et les CIBLE44. On voit que les essais avec masqueur sont globalement associés à de moins bonnes performances que les essais sans masqueur. On voit également une interaction entre le type de cible (CIBLE36 ou CIBLE44) et la présence du masqueur. Il semble qu'une forte différence de performances entre CIBLE36 et CIBLE44 caractérise les essais avec masqueurs mais pas les essais sans masqueurs. Rappelons que du fait du faible nombre de participants et de points de mesure, aucune analyse statistique n'est réalisée sur ces données.

La suite des analyses ne comporte que les essais où le d' était supérieur à 0.5 donnant lieu à l'exclusion de 12 essais sur les $4 \times 40 = 160$ (40 essais par participant pour 4 participants).

6.4.1.2 Analyse de l'aSSR

La Figure 6.7 illustre le contenu fréquentiel des réponses EEG. Les spectres sont présentés pour les différentes conditions (10 essais d'une minute moyennés pour chaque condition). On voit qu'un pic correspondant à la fréquence de modulation

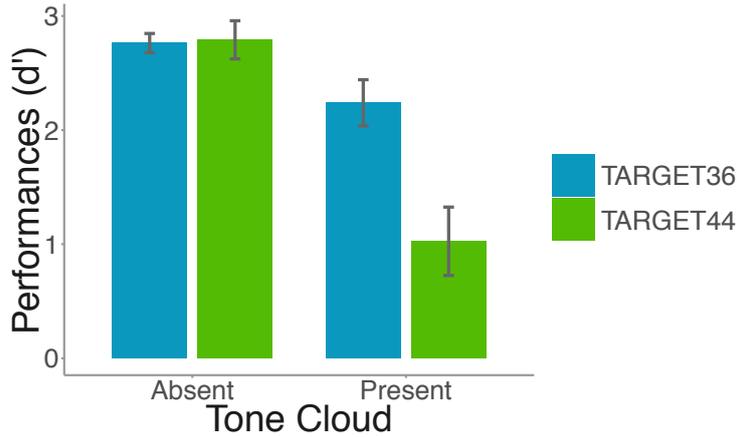


Figure 6.6: Partie 1. Données comportementales. Les performances en d' sont montrées pour les essais avec et sans masqueurs pour les CIBLE36 et les CIBLE44. Les barres d'erreur représentent l'erreur standard.

(36 Hz ou 44 Hz) est bien présent dans toutes les conditions. On peut en outre noter que l'amplitude des pics est plus importante dans la condition sans nuage de tons en comparaison de la condition avec nuages de tons, et cela pour les deux cibles (CIBLE36 et CIBLE44). Il semble également que l'amplitude des pics soit légèrement plus grande pour les CIBLE44 que pour les CIBLE36.

6.4.1.3 Reconstruction du stimulus

La Figure 6.8 compare les corrélations entre stimulus reconstruit et stimulus réel (axe x) avec les corrélations entre stimulus reconstruit et le stimulus correspondant à un autre essai, utilisé comme ligne de base correspondant au hasard (axe y). Les points sur la diagonale correspondent donc à des corrélations identiques dans les deux cas. On voit que la majorité des points sont au-dessous de la diagonale montrant que la corrélation est généralement la plus forte entre le stimulus reconstruit et le stimulus réel que dans la ligne de base. On peut remarquer que les valeurs de corrélations sont faibles. Cela vient du fait que le signal est très bruité est cela est donc attendu avec la méthode de reconstruction utilisée. Toutefois, les corrélations que nous avons observé sont légèrement plus faibles que dans des études précédentes (O'Sullivan et al., 2014).

La Figure 6.9 compare les corrélations obtenues entre l'enveloppe reconstruite et l'enveloppe du stimulus réel pour des essais sans et avec nuages de tons. On voit que les corrélations sont globalement plus basses pour les essais avec nuage de tons.

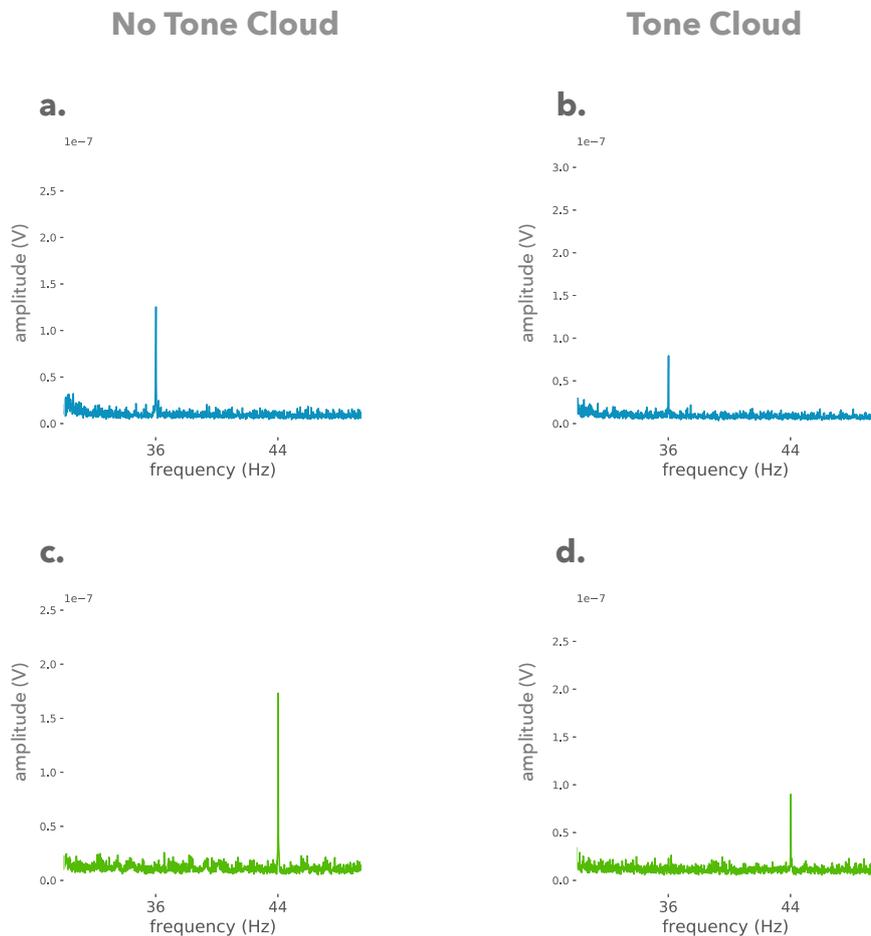


Figure 6.7: Partie 1. aSSR. Amplitude en fonction de la fréquence du signal dans le domaine fréquentiel montrant les aSSRs correspondant aux essais des différentes cibles (CIBLE36 : haut, CIBLE44 : bas) et différentes conditions (sans nuages de tons : gauche, avec nuage de tons : droite). a. CIBLE36 (modulation à 36 Hz) sans nuage de tons. b. CIBLE36 (modulation à 36 Hz) avec nuage de tons. c. CIBLE44 (modulation à 44 Hz) sans nuage de tons. d. CIBLE44 (modulation à 44 Hz) avec nuage de tons.

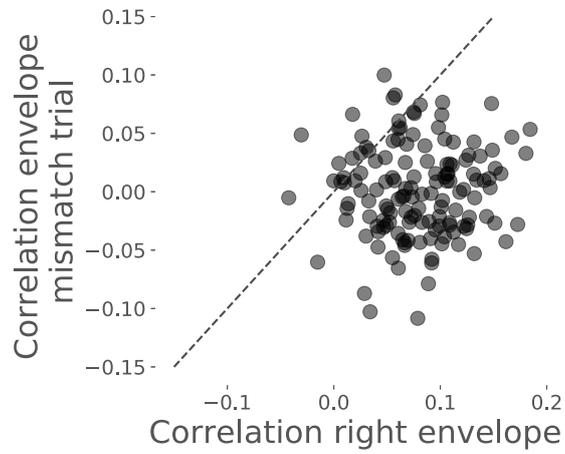


Figure 6.8: Partie 1. Reconstruction du stimulus. Comparaison des corrélations entre le stimulus reconstruit et le stimulus réel (x) et entre le stimulus reconstruit et le stimulus d'un autre essai (y).

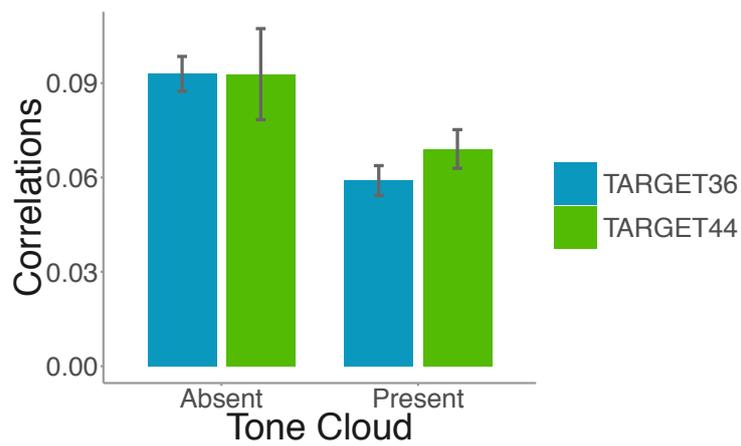


Figure 6.9: Partie 1. Reconstruction du stimulus. Comparaison des corrélations moyennes entre le stimulus reconstruit et le stimulus réel pour les essais sans et avec nuages de tons.

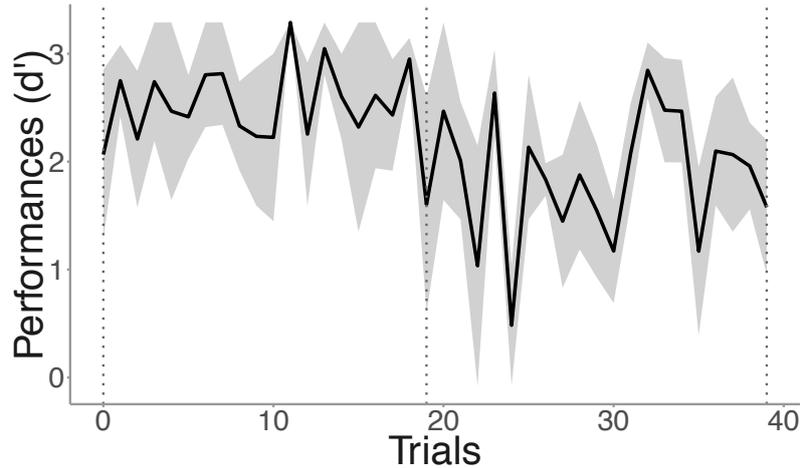


Figure 6.10: Partie 2. Données comportementales. Les performances moyennées des participants en d' sont montrées pour chaque essai des blocs 1 et 2 (essais 0 à 20 ; attention portée aux CIBLE36), et 3 et 4 (essais 20 à 40 ; attention portée aux CIBLE44). Le ruban représente l'erreur standard.

Contrairement aux données comportementales, il semble que les corrélations soient légèrement plus fortes pour la condition CIBLE44 lorsqu'un masqueur est présent. Cette différence n'existe pas pour les essais sans masqueur.

6.4.2 Partie 2

6.4.2.1 Données comportementales

La Figure 6.10 montre les performances des participants pour chaque essai de la partie 2. Les essais 1 à 20 correspondent au blocs 1 et 2 (attention portée sur la CIBLE36) et les essais 20 à 40 au bloc 3 et 4 (attention portée à la CIBLE44). Les performances semblent légèrement plus faibles que dans la partie 1 et meilleures dans la condition CIBLE36 ($d' = 2.51$) par rapport à la condition CIBLE44 ($d' = 1.87$). Cela est également montré par la Figure 6.11 qui représente les performances globales en d' pour les essais où la CIBLE36 devait être écoutée et pour les essais où la CIBLE44 devait être écoutée. On voit en effet un léger avantage pour les CIBLE36.

La suite des analyses des signaux EEG comporte uniquement les essais où le d' était supérieur à 0.5 donnant lieu à l'exclusion de 16 essais sur les $4 \times 40 = 160$ (40 essais par participant et 4 participants).

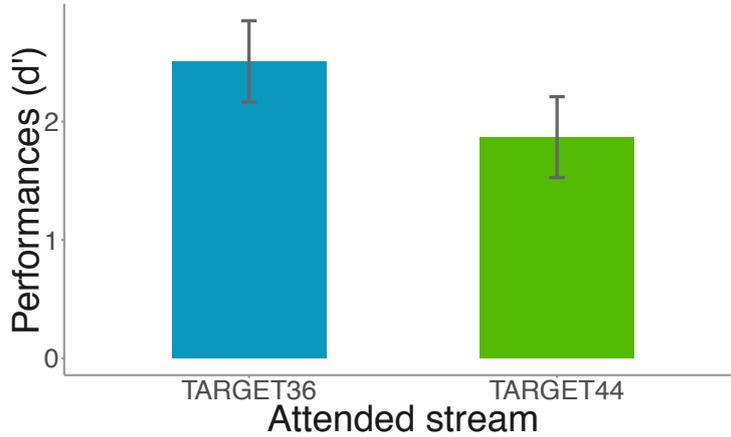


Figure 6.11: Partie 2. Données comportementales. Les performances en d' sont montrées pour les CIBLE36 et les CIBLE44. Les barres d'erreur représentent l'erreur standard.

6.4.2.2 Analyse de l'aSSR

Les Figures 6.12 et 6.13 montrent l'amplitude du signal après l'application de la transformation de Fourier pour les essais où la CIBLE36 (Figure 6.12) et la CIBLE44 (Figure 6.13) a été l'objet de l'attention. On voit que l'amplitude du pic à 36 Hz était plus large lorsque que c'est la CIBLE36 qui a été écoutée et inversement que l'amplitude de pic à 44 Hz était plus large lorsque c'est la CIBLE44 qui a été écoutée.

6.4.2.3 Reconstruction du stimulus

La Figure 6.14 montre les corrélations entre l'enveloppe reconstruite et les enveloppes de chaque cible pour les essais où les participants devaient focaliser leur attention sur l'une et l'autre des cibles continues. On voit une séparation nette entre les essais, ce qui montre que les corrélations sont plus fortes entre l'enveloppe reconstruite et la cible continue qui a été écoutée en comparaison des corrélations entre l'enveloppe reconstruite et la cible continue qui devait être ignorée.

La Figure 6.15 montre le déroulé temporel des corrélations entre l'enveloppe reconstruite et l'enveloppe du stimulus (CIBLE36 en bleu et CIBLE44 en vert) pour les essais où les participants devaient porter leur attention sur la CIBLE36 ou sur la CIBLE44. On voit que les corrélations entre l'enveloppe reconstruite et l'enveloppe de la cible qui a fait l'objet de l'attention (bleu à gauche et vert à droite) sont plus forte que celles entre l'enveloppe reconstruite et l'enveloppe de la cible qui devait

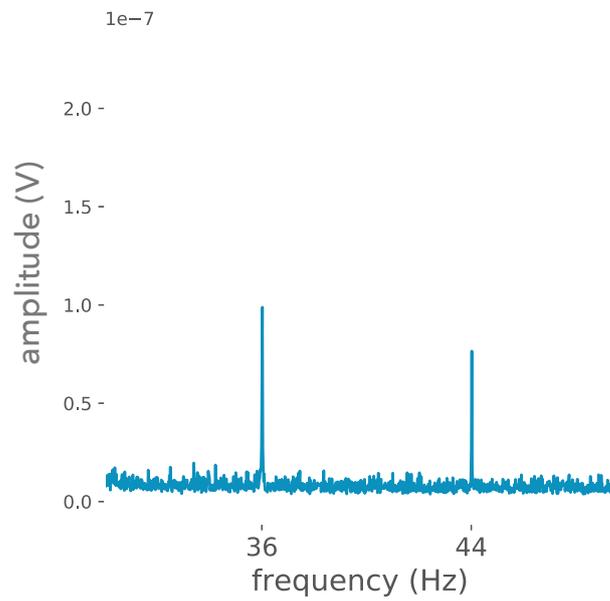


Figure 6.12: Partie 2. aSSR. aSSR correspondant aux essais où l'attention était portée à la CIBLE36 (modulation à 36 Hz).

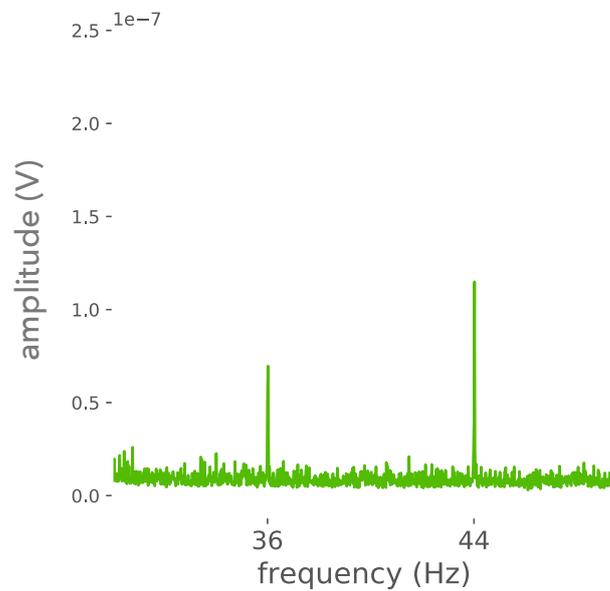


Figure 6.13: Partie 2. aSSR. aSSR correspondant aux essais où l'attention était portée à la CIBLE44 (modulation à 44 Hz).

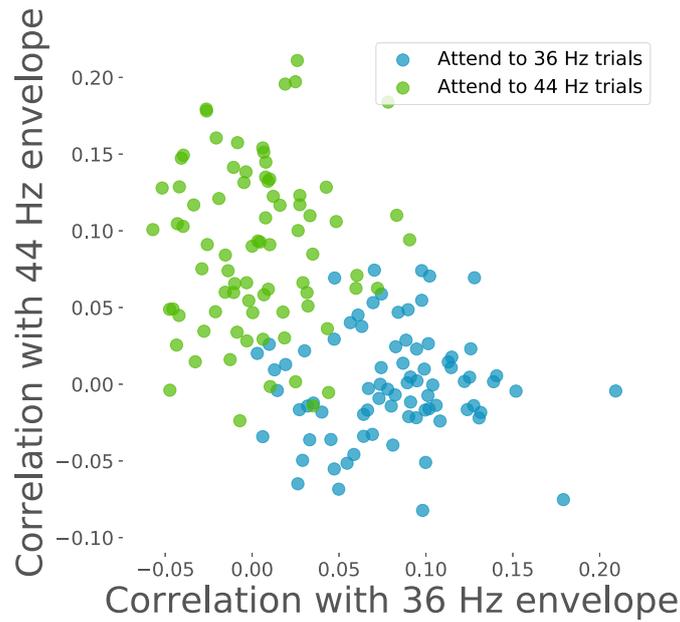


Figure 6.14: Partie 2. Reconstruction du stimulus. Corrélations entre l'enveloppe reconstruite et l'enveloppe de la CIBLE36 sur l'axe x et corrélations entre l'enveloppe reconstruite et l'enveloppe de CIBLE44 sur l'axe y . Les essais où le participant devait focaliser son attention sur la CIBLE36 (en bleu) ou sur la CIBLE44 (en vert) sont représentés.

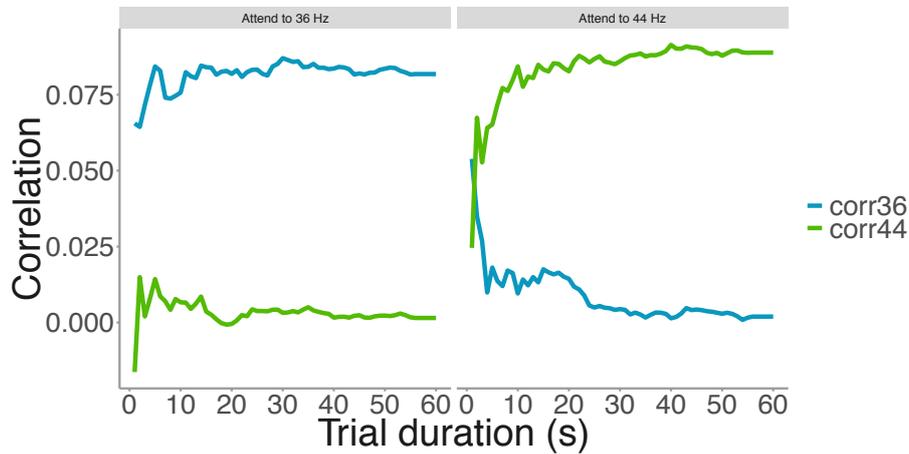


Figure 6.15: Partie 2. Reconstruction du stimulus. Corrélations entre l'enveloppe reconstruite et l'enveloppe du stimulus en fonction de la durée de l'essai pris en compte. Chaque point représente un essai où le participant devait focaliser son attention sur la CIBLE36 (gauche) ou sur la CIBLE44 (droite) et les corrélations avec l'enveloppe de la CIBLE36 (bleu) et avec la CIBLE44 (vert).

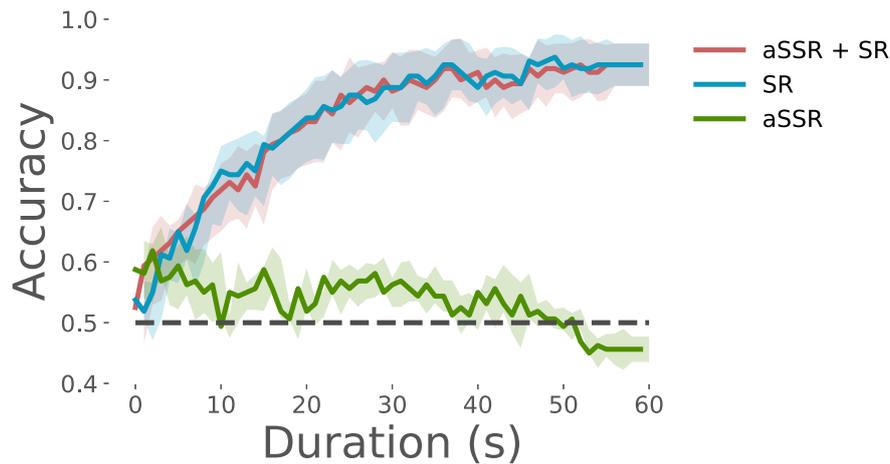


Figure 6.16: Partie 2. Décodage de l’attention. Performances des décodeurs construits à partir des corrélations issues de la reconstruction du stimulus (bleu), à partir de l’aSSR (vert) ou à partir des deux méthodes (rouge). Les performances correspondent à la proportion d’essais correctement classifiés selon la procédure de validation croisée (*5-fold cross-validation*). Les rubans représentent l’erreur standard.

être ignorée (vert à gauche et bleu à droite). Ces corrélations ont été calculées en prenant en compte différentes durées d’essais de 1 à 60 secondes.

6.4.2.4 Décodage

La Figure 6.16 montre les performances des trois décodeurs. On voit que les décodeurs construits à partir des corrélations issues de la reconstruction du stimulus (reconstruction seule ou reconstruction et aSSR) donnent les meilleurs résultats, et atteignent 90% de classification correcte en environ 30 secondes. Cependant, les décodeurs ne semblent pas bénéficier de l’ajout de l’amplitude de l’aSSR car le décodeur utilisant les deux indices n’est pas plus performant que le décodeur utilisant uniquement la reconstruction du stimulus. Cela est confirmé par le fait que le décodeur utilisant uniquement l’aSSR ne donne pas des performances supérieures au hasard.

6.5 Discussion

Ce travail a permis de mettre en avant divers éléments méthodologiques et théoriques. Il s'agit d'un travail préliminaire dans le sens où seulement 4 participants ont été testés, et les résultats doivent être considérés avec précaution.

Nous avons vu dans les chapitres précédents que lorsqu'une cible est présentée avec un masqueur simultané, les performances de détection sont impactées, même lorsque le masqueur exclut une région protégée dans la région fréquentielle de la cible. Dans la première partie de cette expérience, les performances comportementales ont été largement réduites par la présence du nuage de tons, confirmant cet effet de masquage informationnel. En outre, les deux méthodes d'analyse ont montré que l'aSSR était réduite et la reconstruction de l'enveloppe moins efficace en présence du nuage de tons.

Nous avons observé dans la première partie de l'expérience un large avantage comportemental pour les CIBLE36 par rapport aux CIBLE44, uniquement en présence du nuage de tons. Il est possible que cette différence provienne de la fréquence de modulation ou de la fréquence de la porteuse de la cible. Nous avons vu dans le chapitre précédent que les performances à une tâche de détection de cible répétées dans des nuages de tons étaient fortement dépendantes de la fréquence. Notamment, de meilleures performances étaient associées aux cibles entre 500 Hz et 1000 Hz par rapport aux cibles de fréquences supérieures. Cela correspond à ce qui est observé ici (meilleure performance pour la CIBLE36 de porteuse 700 Hz comparée à la CIBLE44 de porteuse 1599 Hz). Nous avons également constaté que la différence entre les CIBLE36 et les CIBLE44 n'était pas observée lorsqu'il n'y avait pas de masqueur ce qui suggère que l'effet de masquage est plus important pour la CIBLE44 et non que ces cibles sont intrinsèquement plus difficiles à percevoir. Ces observations seraient tout à fait cohérentes avec une hypothèse basée sur l'attention sélective pour les résultats des chapitres précédents, et notamment sur l'effet de la fréquence sur le masquage informationnel.

Toutefois, dans la partie 2 où les cibles étaient présentées simultanément mais sans masqueur, les CIBLE44 ont également été associées à des performances légèrement inférieures à celles des CIBLES36. Une explication potentielle est que, dans cette deuxième partie, la seconde cible a eu un effet masquant (bien que moins fort que l'effet du nuage de tons de la partie 1).

Cependant, cet avantage pour les CIBLE36 dans la partie 1 ne semble pas être visible par le biais de l'aSSR ou de la reconstruction du stimulus. Il semble même que

la tendance est opposée et que l'amplitude de l'aSSR était légèrement plus grande et que les corrélations entre l'enveloppe reconstruite et l'enveloppe du stimulus légèrement plus fortes pour les CIBLE44 que pour les CIBLE36. Les différences étant faibles, il serait nécessaire de confirmer cela avec plus de participants.

Les performances du décodeur construit sur les deux indices électrophysiologiques n'ont pas été meilleures que celles où seule la reconstruction du stimulus a été utilisée. Cela provient de la faible précision de classification du décodeur aSSR. Une explication est que l'effet de l'attention sur l'aSSR dans ce type de tâche est trop faible pour être observé à l'échelle d'un essai unique. Cet avantage du décodeur construit sur la base de la reconstruction du stimulus confirme les travaux de Horton et al. (2014) dans lesquels les participants devaient focaliser leur attention sur un flux de parole parmi deux. L'attention a été décodée à partir de l'enveloppe (cross-corrélation entre l'enveloppe du stimulus et le signal EEG) et de la modulation d'amplitude sinusoïdale qui était appliquée aux flux de parole. Les résultats ont montré un meilleur décodage issu de l'enveloppe.

Finalement, dans la partie 2, les performances du décodeur construit à partir de la reconstruction de l'enveloppe ont atteint presque 90% de classification correctes. Cela est comparable avec les travaux antérieurs utilisant la méthode de reconstruction du stimulus (O'Sullivan et al., 2014 : 82%-89%) et suggère qu'une enveloppe artificielle peut être utilisée comme source du décodage plutôt que l'enveloppe naturelle des stimuli de paroles généralement utilisée. Cela suggère donc que ces techniques pourraient être utilisées avec des stimulus non langagiers. Un avantage majeur d'une enveloppe artificielle et qu'il est possible de contrôler ses caractéristiques. Il est également possible que les performances des décodeurs puissent être optimisées en modifiant les caractéristiques de l'enveloppe artificielle. Par exemple, nous avons utilisé une enveloppe modulée à une fréquence moyenne de 15 Hz contrairement à la parole qui est modulée entre 1 et 10 Hz (Ding & Simon, 2014), avec l'idée que cela fournirait plus de données au modèle linéaire décrivant les liens entre le signal EEG et l'enveloppe entendue. Il est possible que ce ne soit pas la fréquence moyenne idéale. Une manière de vérifier cela serait de varier expérimentalement ce paramètre et de comparer les performances de décodage de l'attention. Il serait également possible d'étudier les limites hautes de cette fréquence de modulation moyenne. Il serait encore possible de mettre en place une enveloppe plus saillante, par exemple par le biais d'une interpolation logarithmique plutôt que linéaire.

6.6 Conclusions

Pour conclure, l'effet de la fréquence de la cible observé au chapitre précédent pourrait être examiné sous l'angle neurophysiologique en comparant la réponse cérébrale selon ces différents indices pour des fréquences porteuse différentes. Il serait alors nécessaire que les conditions avec et sans masqueur et les associations entre fréquence de la porteuse et fréquence de la modulation d'amplitude soient contrebalancées.

De manière plus générale, les expériences décrites ici ont montré qu'il était possible de reconstruire les enveloppes des stimuli à double modulation d'amplitude avec des résultats similaires à ceux obtenus à partir de stimuli de parole.

Pour conclure, ce dernier chapitre constitue une ouverture plutôt qu'une conclusion du travail présenté dans les chapitres précédents. Il a été réalisé dans le cadre du projet CoCoHA (*COgnitive COntrol of a Hearing Aid*), et d'un point de vue personnel, il a également constitué une opportunité de me former aux méthodes et analyses de l'EEG.

7 CHAPITRE 7 : Conclusions

Cette thèse a présenté une série d'expériences visant initialement à caractériser le rôle de la boucle audio-motrice sur l'apprentissage perceptif. Nous avons élaboré de nouveaux outils permettant de mettre en place un lien entre motricité et audition sur tablette. La modification des caractéristiques du son réalisée en temps réel selon l'action tactile du participant a permis d'élaborer un protocole audio-moteur d'entraînement perceptif. Ce protocole a été utilisé pour examiner l'apprentissage de la perception de la fréquence et de l'attention sélective auditive. Excepté sur les tâches de masquage informationnel et pour les participants aux performances les plus faibles, cet entraînement audio-moteur n'a pas conduit à une amélioration robuste des performances. Cependant, nous avons pu observer une amélioration rapide des performances de masquage informationnel et constater un effet fort de la fréquence sur ces tâches. Ensuite, un travail préliminaire permettant d'évaluer les performances de masquage informationnel grâce à des indices électroencéphalographiques (EEG) a été proposé. L'indice *auditory steady-state response* (aSSR) ainsi que des techniques de reconstruction du stimulus ont été utilisées pour examiner les processus associés au masquage informationnel. Nous allons maintenant résumer les principaux résultats des différentes études, puis mettre ces résultats en perspective et proposer d'éventuelles pistes pour donner suite et concrétiser certains aspects de ce travail.

7.1 Principaux résultats

7.1.1 Étude 1 : Perception de la hauteur et boucle audio-motrice

Un apprentissage rapide et généralisable peut-il être induit par un entraînement perceptif avec roving de la fréquence et grâce à l'utilisation de la boucle audio-motrice ?

Nous avons examiné dans cette étude l'effet de la boucle audio-motrice sur l'amélioration des performances de discrimination et d'identification de la fréquence. Nous avons mis en place une tâche qui nécessitait l'association de l'audition et de la motricité pour être réalisée. Les participants contrôlaient la fréquence d'un ton pur continu en déplaçant leur doigt sur la tablette et devaient trouver la position associée au ton le plus aigu. L'efficacité de ce protocole a été testée sur la perception de la fréquence dans des conditions avec et sans *roving*. Dans les tâches avec *roving*, la fréquence du ton standard dans les pré-/post-tests et du ton cible (le plus aigu) dans la tâche audio-motrice d'entraînement changeaient d'un essai à l'autre. En outre, un

effet néfaste du *roving* sur les performances a également été constaté. Cependant, aucun apprentissage associé à l'entraînement audio-moteur n'a été observé, avec ou sans *roving*.

7.1.2 Étude 2 : Effet d'un entraînement actif sur la perception de la fréquence et le masquage informationnel

Un entraînement basé sur la boucle audio-motrice peut-il être utilisé pour améliorer les capacités de perception de la hauteur et la résistance au masquage informationnel ?

Dans cette deuxième étude, nous avons mis en place un protocole d'entraînement perceptif inspiré de la première étude, avec des modifications visant un entraînement plus efficace (niveaux de difficulté croissante, présence de nuage de tons pour entraîner l'attention sélective, durée des essais limitée etc.). De manière à distinguer la perception auditive passive de l'utilisation de la boucle audio-motrice, nous avons exposé un second groupe de participants aux essais réalisés par ceux du groupe actif. Un groupe contrôle ne réalisant que les pré- et post-tests a aussi été mis en place. Il semble que l'entraînement audio-moteur n'a pas été significativement plus efficace que l'exposition ou même que la simple répétition des tâches test, même si une tendance en ce sens a été observée pour les participants aux performances initiales faibles. Par ailleurs, une amélioration des performances à la tâche de masquage a été observée entre les pré- et post-tests pour l'ensemble des groupes. Cela suggère que les performances dans des tâches de masquage informationnel peuvent être améliorées par la pratique.

7.1.3 Étude 3 : Entraînement perceptif et masquage informationnel

Une amélioration rapide et généralisable des performances de masquage informationnel peut-elle être observée avec un roving de la fréquence de la cible ?

On peut se demander si l'apprentissage observé dans l'Étude 2 est dépendant du fait qu'il n'y avait pas de *roving* de la fréquence, c'est-à-dire que la fréquence de la cible à détecter était la même pour l'intégralité du bloc, ce qui poserait des questions sur la généralisation possible d'un tel entraînement aux fréquences cibles non entraînées. Dans l'Étude 3, nous avons repris un protocole de masquage informationnel similaire à celui de l'Étude 2, mais avec la fréquence de la cible variée d'un essai à l'autre. Les résultats ont montré une amélioration claire des performances au cours de la pratique de la tâche, avec variation de la fréquence cible et avec ou sans amorce indiquant

la fréquence cible. De façon inattendue, nous avons constaté que les performances à cette tâche dépendaient fortement de la fréquence de la cible, et que cet effet ne semblait pas être réduit par la pratique.

7.1.4 Étude 4 : Audiogramme informationnel : un effet de masquage informationnel dépendant de la fréquence et sans lien avec l'audiogramme de détection

Le masquage informationnel dépend-il de la fréquence de la cible, lorsque d'autres interprétations comme le niveau de la cible sont écartées ?

Nous avons cherché à examiner l'effet de fréquence qui caractérisait le masquage informationnel observé dans l'Étude 3. Nous avons tout d'abord cherché à savoir si cet effet provenait de différences d'audibilité en pondérant le niveau des stimuli selon l'audiogramme. Nous avons aussi cherché à évaluer l'effet de l'amorçage en présentant des blocs sans amorce, suivis de blocs avec amorce puis à nouveau de blocs sans amorce. Les résultats ont montré un effet de la fréquence malgré la pondération des stimuli. Ensuite, la comparaison des premiers et derniers blocs a montré un effet d'apprentissage mais pas réduction de l'effet de la fréquence sur les performances. Nous avons enfin cherché à savoir si cet effet de fréquence pourrait être issu d'une différence dans la position relative de la fréquence de la cible et des bornes du nuage de tons. Les résultats ont montré que l'effet de masquage était le plus faible dans les fréquences médium, et que les performances diminuaient pour les basses ou les hautes fréquences. L'absence de différence significative entre des nuages de tons décalés en fréquence (pour des cibles de fréquences identiques) suggère que l'effet de la fréquence est un effet de la fréquence absolue, et ne provient pas de la distribution spécifique du nuage de tons.

7.1.5 Étude 5 : Indices électrophysiologiques de l'attention sélective.

Quels indices EEG sont utilisables pour obtenir un corrélat neuronal de l'attention sélective dans les tâches de masquage informationnel ?

Une hypothèse permettant d'expliquer à la fois l'effet d'apprentissage et l'effet de fréquence sur le masquage informationnel se base sur l'attention sélective. Il serait donc utile, pour tester une telle supposition, de disposer d'un corrélat objectif de l'attention sélective. Dans cette étude préliminaire, nous avons évalué l'utilisation de deux techniques d'analyse EEG pour obtenir un corrélat neuronal objectif de

l'attention sélective : l'aSSR (*auditory Steady State Response*) et la reconstruction du stimulus. Pour cela, un stimulus possédant une double modulation d'amplitude a été élaboré. La première modulation était sinusoidale et a été utilisée de manière à induire une aSSR à la fréquence de la modulation. La deuxième modulation était aléatoire et plus lente de manière à utiliser les méthodes de décodage du signal acoustique à partir du signal EEG, généralement utilisées avec des stimuli de parole aux variations d'amplitude complexes et lentes. Les résultats ont montré la possibilité d'observer une aSSR produite en réponse de la modulation sinusoidale en parallèle de la reconstruction de l'enveloppe lente. Des effets de l'attention sélective (qu'est-ce qui était écouté ou ignoré) ont été montrés sur l'aSSR et sur la qualité de la reconstruction. Des décodeurs ont été construits à partir de l'aSSR, de la reconstruction du stimulus ou des deux indices. Les résultats des décodages ont montré que le décodeur basé sur la reconstruction était le plus performant, et qu'il n'y avait pas d'avantage à ajouter l'aSSR. Ces résultats préliminaires suggèrent que ce type de stimuli pourrait être utilisé pour examiner l'attention sélective auditive dans les tâches de masquage informationnel.

7.2 Perspectives

Outils de développement

Tout d'abord, de nombreux outils ont été développés permettant la mise en place de tâches psychoacoustiques sur des supports comme les tablettes. Cela démontre le potentiel des technologies récentes du web pour mettre en place, par exemple, des tâches en ligne, de manière complémentaire à des expériences en laboratoire. De plus, le développement accru de ces appareils a donné lieu à la banalisation de capteurs tels que les accéléromètres (présents dans n'importe quel téléphone mobile), pouvant être utilisés en guise de contrôleur à partir des tâches que nous avons mis en place. Cela pourrait donner lieu à un contrôle du son par le biais de mouvements appliqués à l'appareil, ce qui serait probablement plus immersif que le déplacement du doigt sur la tablette que nous avons utilisé dans nos expériences. Cela pourrait donner lieu à des interfaces audio-motrices plus proches du jeu vidéo ou de la pratique d'un instrument de musique, avec éventuellement une efficacité plus grande que celle - modeste, voire absente - que nous avons observée dans cette thèse.

Stimuli musicaux

Enfin, une collaboration avec le compositeur Fabien Touchard a permis d'initier un travail avec des stimuli à mi-chemin entre des stimuli naturels (musicaux) et des stimuli contrôlés. Cette collaboration n'est pas décrite dans le corps de la thèse mais fait l'objet d'une Annexe (Annexe 2, page 203).

L'écoute musicale est caractérisée par une étape importante qui est la sélection des informations sonores. L'attention peut par exemple souligner les relations harmoniques entre telle et telle voix d'une œuvre polyphonique. L'écoute et la pratique musicale représentent donc un domaine particulièrement favorable au déploiement de l'attention sélective auditive. Or, si un apprentissage rapide peut être observé dans des tâches de masquage informationnel utilisant des nuages de tons, il semblerait intéressant de questionner ces effets d'apprentissage dans le cadre de stimuli plus naturels. Nous avons donc proposé au compositeur Fabien Touchard de composer de tels stimuli musicaux. Ce premier aperçu d'une transposition de nos tâches de laboratoire à des situations plus musicales a donné lieu la création par le compositeur d'une pièce, à la fois œuvre à part entière et matériau sonore dont certains aspects sont contrôlés. Cela ouvre la possibilité de mettre en place des protocoles d'entraînement perceptif avec des stimuli naturels, et de tester les effets de masquage informationnel dans un contexte musical. Dans le cadre de cette collaboration, une application web est en développement, mettant en jeu une tâche de détection de motifs musicaux, à partir du matériau composé.

Neurofeedback

Certaines études de cette thèse pourraient être mises à contribution dans le cadre d'une forme d'entraînement actif, n'impliquant pas la boucle audio-motrice mais plutôt l'utilisation d'indices objectifs de performance pendant la tâche sous forme de *neurofeedback*. L'idée principale serait ainsi d'utiliser (1) l'ensemble du dispositif technique qui a été élaboré, (2) les tâches de masquage informationnel qui ont été développées et dont les protocoles ont été affinés, et (3) l'idée que la reconstruction du stimulus permet de fournir un indice de l'attention sélective en temps réel, même dans le cas de stimuli synthétiques. Tout d'abord, concernant le dispositif technique, une première ébauche a été mise en place. Nous avons testé, du point de vue technique, l'idée d'utiliser les données EEG enregistrées en temps réel et de les déployer sous forme de retour visuel. La restitution de cette piste méthodologique (mettre en place un protocole de *neurofeedback* dans le cadre d'une application web) est évoquée à la fin de l'Annexe 1 (page 195).

Audiogramme informationnel

Nous avons observé un effet inattendu robuste de la fréquence de la cible sur le masquage informationnel. Mais s'agit-il d'un effet propre à l'utilisation de nuages de tons, ou bien caractérise-t-il d'autres tâches connues pour évaluer le masquage informationnel ? Il serait intéressant de vérifier cet effet avec des complexes multi-tons par exemple. On peut également se demander les liens entretenus entre le phénomène de masquage et le concept de bandes attentionnelles. Il serait intéressant d'évaluer la forme des bandes attentionnelles à partir d'une tâche de masquage informationnel et de comparer aux bandes attentionnelles évaluées à partir des tâches classiques du type *signal-probe*. L'effet de fréquence suggère que la forme des bandes attentionnelles mesurées à partir du masquage informationnel serait différente de la forme des bandes critiques. Cela va à l'encontre des travaux sur les bandes attentionnelles et poserait la question de la nécessité de clarifier cette notion. Il semble bien toutefois que ces résultats aient permis de révéler un aspect encore inconnu de l'audition au dessus du seuil, qui pourrait être désigné sous le terme d'« audiogramme informationnel ».

8 ANNEXES

8.1 Annexe 1. Notes sur l'implémentation technique des expériences

Cette annexe décrit les procédés techniques utilisés pour la réalisation des expériences présentées dans les chapitres de cette thèse. Le but des premières expériences de ce travail était de recruter la boucle audio-motrice pour des protocoles d'entraînement perceptif. Il a donc été choisi d'utiliser des tablettes tactiles comme outil simple de recrutement de cette boucle audio-motrice (par exemple, Whitton et al., 2017). De plus, en vue d'une éventuelle utilisation hors du laboratoire (qui n'a finalement pas été testée au cours de la thèse), nous avons choisi le principe d'expériences en ligne, pouvant être réalisée à la fois en laboratoire et chez les participants. Une fois l'architecture technique réalisée, elle a été conservée pour toutes les expériences, même celles plus classiques n'utilisant pas d'apprentissage audio-moteur. Le laboratoire d'accueil n'ayant pas d'expérience préalable avec ces contraintes, j'ai donc d'abord dû faire des choix techniques et les implémenter.

Le format application web a été choisi car cela permet à l'ensemble du protocole expérimental d'être utilisé sur des appareils variés (ordinateurs, tablettes, téléphone). À l'inverse, une application programmée nativement (c'est-à-dire avec le langage utilisé par l'appareil) serait spécifique à un système d'exploitation. Par exemple, un langage différent est utilisé pour les appareils mobiles IOS (Objective-C) et pour les appareils Android (Java ; qui malgré le nom est un langage indépendant de Javascript). Ce n'est pas le cas d'une application web qui peut être utilisée sur n'importe quel appareil qui possède un navigateur web. On peut toutefois noter d'autres types de difficultés liées à l'utilisation d'engin Javascript différent par les différents navigateurs, donnant parfois lieu à des incompatibilités.

Lorsqu'un navigateur se connecte à un site Internet (par exemple, par le biais d'une URL spécifique), une requête est envoyée depuis l'appareil de l'utilisateur (nommé le *client*) jusqu'au serveur (Figure 8.1). Selon les caractéristiques de l'URL (sous-domaines, paramètres etc.), le serveur envoie en retour un contenu particulier. Généralement ce contenu est une page web (écrite en HTML), des fichiers statiques (fichiers CSS pour le style de la page, images, sons etc.) et éventuellement des fichiers de code écrits en Javascript. Dans le cas d'une application web, c'est dans ce code que l'ensemble du programme est présenté. Ce code est étroitement lié au fichier HTML, et par exemple, des fonctions peuvent être déclenchées lorsqu'une interaction

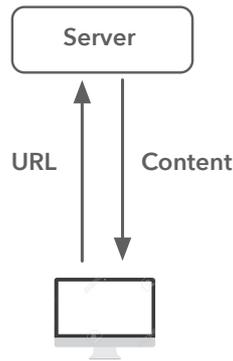


Figure 8.1: Illustration d’une requête de type GET. L’ordinateur envoie une requête au serveur par le biais du navigateur et reçoit en échange les fichiers correspondant à sa requête (page web, images, feuilles de styles, fichiers Javascript etc.).

est réalisée avec une partie de la page (un bouton, un déplacement de souris etc.). Ce système permet de réaliser des applications interactives.

Ainsi, une fois que les fichiers statiques sont chargés (envoyé par le serveur et reçus par le client) la latence éventuelle ne dépend pas de la connexion internet. L’application peut même être élaborée de manière à pouvoir être utilisée hors-ligne étant donné que tous les fichiers ont été chargés par le biais de la requête initiale.

La Figure 8.2 représente l’ensemble des composantes de la structure des applications créées pour les expériences de cette thèse. L’utilisateur utilise un navigateur web pour accéder au contenu de l’application stocké sur un serveur hébergé par Heroku (qui permet le déploiement de l’application : *Platform as a Service* ou PaaS). Le serveur web est écrit avec Django qui est un *framework* web de Python. Il est ainsi possible de réaliser des opérations en Python associées à une requête particulière du client. Par ailleurs, de manière à stocker les données de l’utilisateur (réponses etc.), la base de données CouchDB a été utilisée. Elle est hébergé sur un serveur virtuel DigitalOcean (*cloud computing platform*). En outre, une base de donnée locale (c’est-à-dire présente sur l’appareil de l’utilisateur) a été utilisée (PouchDB). Le principe de PouchDB est de stocker localement les données dans une base organisée de manière identique à la base distante (CouchDB) et de synchroniser les bases lorsqu’une connexion internet est détectée. Des informations stockées sur la base de données distante peuvent également être récupérées de manière à retrouver l’état de l’application lorsque le participant se connecte. Les différents composants seront décrit plus en détail plus bas dans cette annexe.

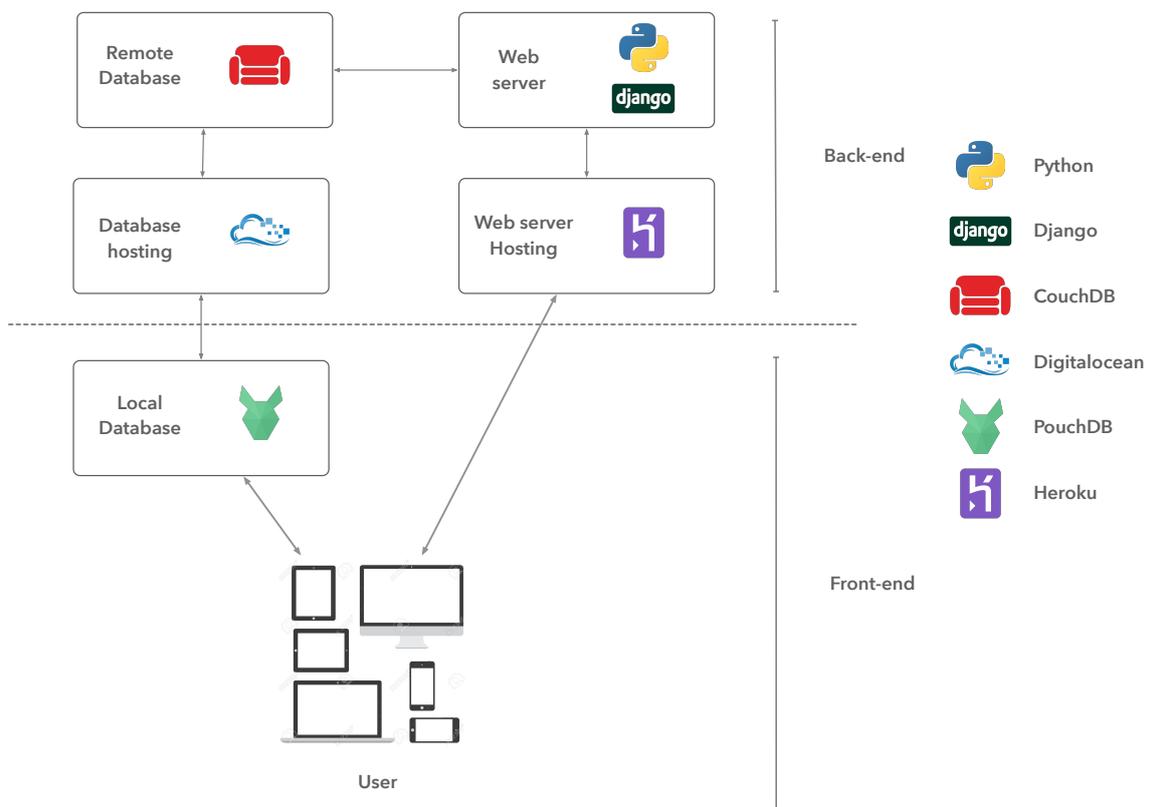


Figure 8.2: Structure type des applications web créées dans le cadre de cette thèse.

La partie *front-end* concerne tout le code qui est présent localement sur l'appareil de l'utilisateur et la partie *back-end* concerne les serveurs distants sur lesquels sont stockés les codes permettant de gérer les requêtes envoyés par le *front-end* (serveur web) et le stockage externe des données (base de données).

Front-end

Le *front-end* est écrit en Javascript et contient le code qui permet à l'application de jouer les stimuli, gérer les données etc. Il est donc spécifique à chaque tâche. Différentes applications ont été créées utilisant un *front-end* différent mais avec la même structure *back-end* (voir plus loin).

Création des sons - Web Audio API

La création d'expérience portant sur l'audition en Javascript est possible grâce à la Web Audio API, ajoutée au langage Javascript en 2012 et en constante évolution, qui permet la création de sons. Une nouvelle horloge beaucoup plus précise que l'horloge Javascript de base (utilisée par exemple dans les fonctions `setTimeout()` ou `setInterval()`) a notamment été ajoutée. L'horloge classique ne peut pas être utilisée pour créer des sons car elle est sur le même flux (*thread*) que d'autres événements tels que les requêtes, les interactions, callbacks etc. Il est donc possible que des événements sonores planifiés soient retardés. A l'inverse, l'horloge de la Web Audio API fonctionne sur un flux indépendant et ne possède pas ce problème. En outre, elle a une précision permettant une planification à un échantillon près, même pour des fréquences d'échantillonnage élevées.

Base de données locale - PouchDB

PouchDB est une bibliothèque Javascript permettant de stocker des données au niveau du client. En effet, les navigateurs web prévoient un espace particulier, isolé du reste de l'ordinateur pour des raisons de sécurité, où des données peuvent être stockées par les applications présente sur Internet. Deux bases de données locales principales peuvent être utilisées : IndexedDB et WebSQL. Il s'agit d'espaces différents des cookies ou des localStorage et sessionStorage qui sont plutôt destinés à stocker de faibles quantités de données.

De manière à créer une application qui peut fonctionner hors-ligne, un stockage temporaire des données utilisateur (réponse, conditions etc.) a été utilisé. L'ap-

plication implémente une synchronisation de la base locale avec la base distante lorsqu'une connexion Internet est détectée, mettant ainsi à jour les données récoltés pendant la phase hors-ligne. Le but n'est donc pas nécessairement de permettre un fonctionnement intégral hors-ligne mais plutôt d'augmenter la résilience à des coupure de connexions (qui peuvent être fréquente dans des cabines insonorisées, par exemple).

Back-end

Serveur web - Django/Python

Les opérations réalisées par le serveur web (codé avec le *framework* Django) sont utilisées pour répondre aux requêtes du client. Par exemple, lorsqu'une URL est demandée, une fonction est déclenchée qui permet de renvoyer une page spécifique. La requête peut déclencher n'importe quel code écrit en Python dont le résultat pourra être renvoyé au client.

Un système d'authentification a été mis en place de manière à créer un espace indépendant pour chaque participant permettant par exemple de reprendre la tâche où elle a été laissée pour les sessions multiples d'entraînement perceptif. A la création du compte utilisateur, le serveur Django interagit avec le serveur Couchdb de manière à créer une nouvelle base pour le participant.

Base de données distante - CouchDB

La structure consiste à avoir une base de donnée CouchDB par participant. Cette base contient les données permettant de définir l'état de l'application (combien de niveaux ont été réalisés, etc.). Ces données sont donc utilisées par le client lorsque l'application est démarrée pour retrouver l'état précédent. En outre, les données comportementales (réponses, stimuli utilisés, déplacement sur la tablette etc.) sont stockées de manière à être ensuite analysées.

Contrairement aux bases de données de type SQL, CouchDB organise les données sous forme de documents hiérarchiques similaires au format JSON. Un document était créé pour chaque essai avec les données souhaitées. Par exemple, voici un fragment de document :

```
{
  "_id": "dataTrial_1_104",
```

```

    "_rev": "1-2cdb847b5320fe994a38772ad8e16f95",
    "condition": "yAxis",
    "curXY": [
      [
        248.91900634765625,
        130.27398681640625,
        0
      ],
      ...
      [
        227.92599487304688,
        467.905029296875,
        3152
      ]
    ],
    "flatTopWidth": 20,
    "masker": false,
    "level": 5,
    "maskerCompNumber": 1,
    "maskerGain": 1,
    "mountainSlope": 1,
    "points": 1000,
    "result": 98.1896629141855,
    "isRoving": false,
    "isTargetContinuous": false,
    "score": null,
    "screenHeight": 707,
    "screenWidth": 768,
    ...
    "trialNumYAxis": 105,
    "levelPoints": 69461
  }

```

Sécurité

Pour l'ensemble des connexions entre le serveur web, le serveur de la base de donnée et le client, un protocole SSL (*Secure Sockets Layer*) a été mis en place de manière à assurer un échange crypté et sécurisé des données.

Exemple de mise en place technique d'un protocole de *neurofeedback*

Le dispositif technique mis en place dans les expériences de cette thèse pourrait également être utilisé dans le cadre d'un protocole de *neurofeedback* à partir d'enregistrements électroencéphalographique (EEG). Le but serait alors de présenter en temps réel des informations extraites à partir du signal EEG dans l'application web. Un travail technique préliminaire a été réalisé, et il semblait donc intéressant de reporter brièvement les différentes étapes ici.

La première étape est de mettre en place le passage des données depuis l'appareil EEG jusqu'à l'ordinateur d'acquisition. Dans notre cas, le logiciel Actiview fourni avec l'appareil Biosemi a été utilisé. Il est ensuite possible de partager les données acquises en temps réel par le biais d'un flux TCP/IP, à partir du logiciel Actiview. Ce flux peut alors être récupéré en python depuis le *back-end* de l'application web. C'est à cette étape que les traitements appliqués au signal EEG peuvent être réalisés en Python (par exemple, l'extraction d'un pic à une certaine fréquence). Ensuite, une méthode appelée *Web-sockets* permet d'avoir un flux d'échange entre le *back-end* et le *front-end* sans qu'une requête ne doive être réalisée à chaque fois. Cela permet de garder une communication continue et sécurisée entre le *back-end* et le *front-end*. Dans notre cas, cela permet d'envoyer en flux (par l'intermédiaire d'un *buffer*) le résultat des traitements réalisés par le *back-end*. Une fois les données sur le *front-end*, Javascript peut être utilisé pour créer un *feedback*, visuel par exemple.

Pour résumer, les différentes étapes sont les suivantes :

- Acquisition des données par Actiview (un enregistrement peut être réalisé en parallèle).
- *Streaming* des données par paquets depuis actiview en TCP/IP.
- Récupération du flux TCP/IP en python depuis le *back-end*.
- Traitements du signal EEG en python (par exemple, extraction de la puissance dans une certaine bande de fréquence).

- Utilisation de *web sockets* pour ouvrir une connexion continue entre *back-end* et *front-end*. Le résultats du traitement réalisé en Python peut être envoyé au *front-end* en continu.
- Utilisation des données en Javascript pour créer le *feedback*.

8.2 Annexe 2. Collaboration avec le compositeur Fabien Touchard pour la mise en place de stimuli musicaux

Une partie de ce travail a été présentée en tant que :

Jean, H., Touchard, F. (2018). Attention sélective et musique. CRR de Boulogne-Billancourt, France (Présentation orale, 12 novembre 2018).

Introduction

Un des buts de cette thèse a été d'étudier la capacité à isoler une source auditive dans un environnement sonore, par le biais de tâches de masquage informationnel. Dans le cadre musical, la capacité à se focaliser sur une partie de l'environnement sonore est crucial pour, par exemple, isoler un instrument dans une musique polyphonique. Il est en effet important de noter que cette thèse a été réalisé dans le cadre du programme doctoral SACRe (Sciences, Arts, Création, Recherche), promouvant les recherches interdisciplinaire entre art et science.

L'utilisation des stimuli du type des nuages de tons est justifiée par le fait qu'ils sont contrôlés. Cela permet de réduire l'impact de variables confondantes sur les résultats observés. Cependant, ils sont de ce fait très éloignés de situations d'écoute réelles et écologiques, comme l'écoute de la musique. Il peut donc être parfois délicat de transférer les conclusions scientifiques vers une application concrète. En outre, il peut être fastidieux - du point de vue du participant - de réaliser une tâche peu ludique, notamment sur de longues durées, ce qui est particulièrement pertinent pour le type d'applications qui pourrait découler de mon travail, comme des aides à la réhabilitations auditive.

Pour ces raisons, j'ai exploré la possibilité de mettre en place un protocole à partir de stimuli musicaux dont certains éléments sont contrôlés. Une collaboration avec un compositeur a permis la création d'un ensemble de pièces musicales aux contraintes compositionnelles liées aux nécessités expérimentales. Ces pièces sont tout à la fois des œuvres musicales à part entière, mais également de possibles stimuli expérimentaux possédant donc certaines propriétés spécifiques et contrôlées.

Le but serait de mettre en place un protocole d'entraînement auditif, qui n'a pas été testé pour le moment, qui serait d'une part plus stimulant à utiliser et d'autre part directement lié à une tâche écologique : l'attention sélective musicale. Les implications en terme de pédagogie musicale pourront également être questionnées.

(* les registres pour ce passage sont différents)

- 8 -

www.schott-music.com · SKK 46

Figure 8.3: Une page de la partition de la pièce composée par Fabien Touchard.

Le but de cette annexe est de présenter les fondements et motivations de cette collaboration.

Tâche

La tâche mise en place est une tâche de détection de motif. Il s'agit d'écouter un motif mélodique seul, puis de répondre à l'aide d'un bouton lorsque ce motif est entendu dans un extrait musical. Pour chaque motif cible, un extrait de la pièce d'une durée de 30 à 60 secondes et qui contient ce motif, est sélectionné. Le début et la fin de l'extrait sont choisis de manière à ce que la position du motif dans l'extrait soit aléatoire.

Le but est de détecter le motif dans différents niveaux de difficulté générés par la combinaison des paramètres suivants : niveau de la voix comportant le motif par rapport aux niveaux des autres voix, durée de l'extrait à écouter, présence du motif à une voix identique ou différente du motif entendu seul.

Stimuli

Les stimuli utilisés sont issus du matériau musical composé par Fabien Touchard. La pièce comporte des parties libres, sans contraintes spécifiques, et des parties contrôlées, nommé plus bas *passages cibles*, lorsque le motif cible est joué.

L'œuvre réalisée est une pièce pour piano seul. Les passages cibles sont écrits comme une superposition de 5 voix mélodiques distinctes dans laquelle les voix ne peuvent pas se croiser et où une région protégée est présente entre chaque voix. Cela permet d'éviter trop de masquage énergétique du motif à détecter. Il faut toutefois noter qu'il s'agit de sons complexes (piano) et donc que des harmoniques seront nécessairement présentes dans les régions protégées également. Concernant les parties libres, les voix n'ont plus ces contraintes et l'écriture peut sortir du cadre des 5 voix. Un des enjeux, pour le compositeur, était d'éviter un changement perceptible entre les passages cibles et les passages libres.

Pour les passages cibles, les voix étaient organisées en 5 bandes d'une octave chacune, permettant une transposition aisée des motifs d'une voix à l'autre. Une zone protégée était également présente autour de chacune des voix. La largeur de ces voix excluant les zones protégées était d'une sixte majeure (une seconde majeure soustraite à l'octave de chaque côté de la bande).

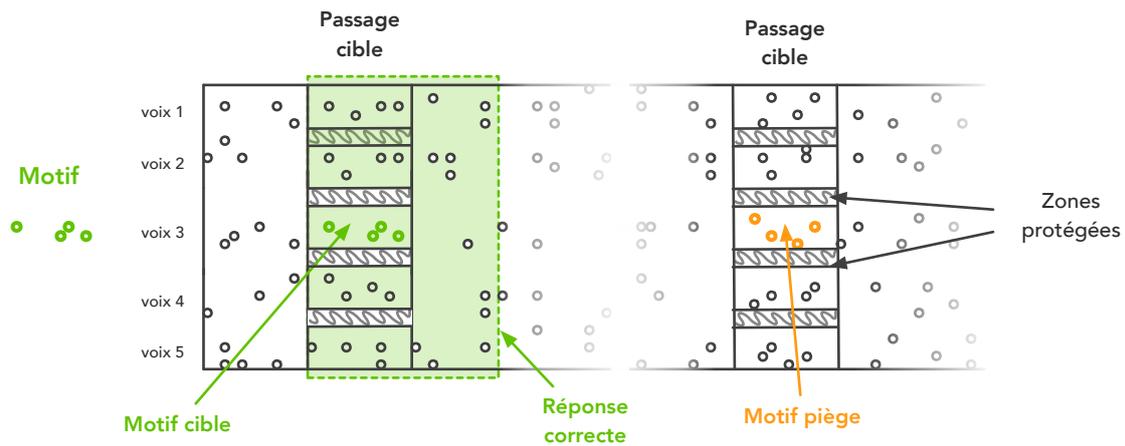


Figure 8.4: Illustration des stimuli et de la structure de la tâche. La réponse était considérée comme correcte lorsqu'elle était donnée dans la zone représentée en vert. Les passages cibles comportaient les contraintes énoncées dans le texte. Certains passages cibles présentaient un motif piège ne devant pas être associé à une réponse du participant.

En outre, de manière à contrôler la densité spectro-temporelle des passages cibles, les voix qui ne présentaient pas le motif avaient une densité équivalente à celle qui le contenait. Cela est visualisé sur la Figure 8.4 où un découpage temporel et fréquentiel (correspondant aux différentes voix) est présenté. On voit que dans les colonnes comportant le motif, les autres voix présentent un nombre de notes identique dans chaque cellule (quatre dans l'exemple). Une note est considérée dans une cellule lorsqu'elle commence dans la cellule même si elle finit dans la suivante.

Pour chaque motif cible, un *motif piège* (en orange sur la Figure 8.4) a également été composé. Il s'agissait d'un motif différent de celui à détecter, mais identique en terme de densité spectro-temporelle. Cela permettait d'éviter que le sujet réponde parce qu'il a distingué un changement de la densité ou des modalités de composition.

De manière à augmenter le matériau utilisable dans le cadre d'une tâche perceptive, une autre contrainte de composition était de rendre possible l'interversion des voix intérieures. Cependant, les deux voix extrêmes n'étaient pas impliquées dans les interversions. Elles restaient identiques dans toutes les configurations du passage cible, et étaient toujours présentes. Ainsi, pour chaque passage cible, 3 voix intermédiaires étaient choisies parmi les pistes suivante :

- Une voix cible qui présente le motif cible.
- Une voix piège qui présente le motif piège.

voix 1	-----	-----		-----
voix 2	X	X	...	Cible
voix 3	Y	Cible	...	X
voix 4	Piège	Y	...	Y
voix 5	-----	-----		-----

Figure 8.5: Exemple d'interversions entre les motifs cible ou piège, motif X et motif Y. Toutes les combinaisons des motifs cible, X et Y ou des motifs piège, X et Y étaient possibles.

- Une voix avec un motif ‘de remplissage’, que l’on peut nommer X.
- Une voix avec un motif ‘de remplissage’, que l’on peut nommer Y.

Chacune des combinaisons de trois voix parmi ces 4 était possible, excepté la présence de la voix cible et de la voix piège. Il était donc possible de présenter le motif seul à une voix puis à une autre (à une autre octave) dans l'extrait.

La Figure 8.5 montre quelques exemples d'interversion possible des différents motifs. Toutes les combinaisons étaient possibles. Il s'agissait donc là d'une difficulté majeure de composition.

Enregistrement et mixage

Les pièces musicales ont été enregistrées au Studio de Meudon par le pianiste Philippe Hattat. Un travail de mixage particulier a donné lieu à une version fixe et définitive de l'œuvre, et à une version *pistes séparées* réalisée pour notre tâche.

RÉFÉRENCES

Aberg, K. C., Tartaglia, E. M., & Herzog, M. H. (2009). Perceptual learning with Chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep. *Vision Research*, *49*(16), 2087–2094. doi:[10.1016/j.visres.2009.05.020](https://doi.org/10.1016/j.visres.2009.05.020)

Adini, Y., Wilkonsky, A., Haspel, R., Tsodyks, M., & Sagi, D. (2004). Perceptual learning in contrast discrimination: The effect of contrast uncertainty. *Journal of Vision*, *4*(12), 2–2. doi:[10.1167/4.12.2](https://doi.org/10.1167/4.12.2)

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. doi:[10.1016/j.tics.2004.08.011](https://doi.org/10.1016/j.tics.2004.08.011)

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 285–299. doi:[10.1098/rstb.2008.0253](https://doi.org/10.1098/rstb.2008.0253)

Alexander, J. M., & Lutfi, R. A. (2004). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *The Journal of the Acoustical Society of America*, *116*(4), 2234–2247. doi:[10.1121/1.1784437](https://doi.org/10.1121/1.1784437)

Amitay, S., Hawkey, D. J. C., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception & Psychophysics*, *67*(4), 691–698. doi:[10.3758/BF03193525](https://doi.org/10.3758/BF03193525)

Amitay, S., Irwin, A., Hawkey, D. J. C., Cowan, J. A., & Moore, D. R. (2006a). A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners. *The Journal of the Acoustical Society of America*, *119*(3), 1616–1625. doi:[10.1121/1.2164988](https://doi.org/10.1121/1.2164988)

Amitay, S., Irwin, A., & Moore, D. R. (2006b). Discrimination learning induced by training with identical stimuli. *Nature Neuroscience*, *9*(11), 1446–1448. doi:[10.1038/nn1787](https://doi.org/10.1038/nn1787)

Amitay, S., Zhang, Y.-X., & Moore, D. R. (2012). Asymmetric Transfer of Auditory Perceptual Learning. *Frontiers in Psychology*, *3*. doi:[10.3389/fpsyg.2012.00508](https://doi.org/10.3389/fpsyg.2012.00508)

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., ... Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, *501*(7465), 97–101. doi:[10.1038/nature12486](https://doi.org/10.1038/nature12486)

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, *27*(6), 953–965. doi:[10.1016/0042-6989\(87\)90011-3](https://doi.org/10.1016/0042-6989(87)90011-3)

Bavelier, D., Bediou, B., & Green, C. S. (2018). Expertise and generalization: Lessons from action video games. *Current Opinion in Behavioral Sciences*, *20*, 169–173. doi:[10.1016/j.cobeha.2018.01.012](https://doi.org/10.1016/j.cobeha.2018.01.012)

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M.-H., & Bertrand, O. (2007). Effects of Selective Attention on the Electrophysiological Representation of Concurrent Sounds in the Human Auditory Cortex. *Journal of Neuroscience*, *27*(35), 9252–9261. doi:[10.1523/JNEUROSCI.1402-07.2007](https://doi.org/10.1523/JNEUROSCI.1402-07.2007)

Boot, W. R., Blakely, D. P., & Simons, D. J. (2011). Do Action Video Games Improve Perception and Cognition? *Frontiers in Psychology*, *2*. doi:[10.3389/fpsyg.2011.00226](https://doi.org/10.3389/fpsyg.2011.00226)

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Broadbent, D. (1958). *Perception and Communication*. Pergamon Press.

Brungart. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. doi:[10.1121/1.1345696](https://doi.org/10.1121/1.1345696)

Cao, X., & Richards, V. M. (2012). Enhancement in Informational Masking. *Journal of Speech, Language, and Hearing Research*, *55*(4), 1135–1147. doi:[10.1044/1092-4388\(2011/09-0149\)](https://doi.org/10.1044/1092-4388(2011/09-0149)

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*. doi:[10.3389/fnhum.2016.00604](https://doi.org/10.3389/fnhum.2016.00604)

Dai, H., Scharf, B., & Buus, S. (1991). Effective attenuation of signals in noise under focused attention. *The Journal of the Acoustical Society of America*, *89*(6), 2837–2842. doi:[10.1121/1.400721](https://doi.org/10.1121/1.400721)

Delhommeau, K., Micheyl, C., & Jouvent, R. (2005). Generalization of Frequency Discrimination Learning Across Frequencies and Ears: Implications for Underlying Neural Mechanisms in Humans. *Journal of the Association for Research in Otolaryngology*, *6*(2), 171–179. doi:[10.1007/s10162-005-5055-4](https://doi.org/10.1007/s10162-005-5055-4)

Delhommeau, K., Micheyl, C., Jouvent, R., & Collet, L. (2002). Transfer of learning across durations and ears in auditory frequency discrimination. *Perception & Psychophysics*, *64*(3), 426–436. doi:[10.3758/BF03194715](https://doi.org/10.3758/BF03194715)

Demany, L. (1985). Perceptual learning in frequency discrimination. *The Journal of the Acoustical Society of America*, *78*(3), 1118–1120. doi:[10.1121/1.393034](https://doi.org/10.1121/1.393034)

Demany, L., & Ramos, C. (2005). On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *The Journal of the Acoustical Society of America*, *117*(2), 833–841. doi:[10.1121/1.1850209](https://doi.org/10.1121/1.1850209)

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *The Journal of the Acoustical Society of America*, *111*(3), 1377–1388. doi:[10.1121/1.1445791](https://doi.org/10.1121/1.1445791)

Demany, L., & Semal, C. (2005). The slow formation of a pitch percept beyond the ending time of a short tone burst. *Perception & Psychophysics*, *67*(8), 1376–1383. doi:[10.3758/BF03193642](https://doi.org/10.3758/BF03193642)

Deveau, J., Ozer, D. J., & Seitz, A. R. (2014). Improved vision and on-field performance in baseball through perceptual learning. *Current Biology*, *24*(4), R146–R147. doi:[10.1016/j.cub.2014.01.004](https://doi.org/10.1016/j.cub.2014.01.004)

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*(29), 11854–11859. doi:[10.1073/pnas.1205381109](https://doi.org/10.1073/pnas.1205381109)

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*. doi:[10.3389/fnhum.2014.00311](https://doi.org/10.3389/fnhum.2014.00311)

Donohue, S. E., Woldorff, M. G., & Mitroff, S. R. (2010). Video game players show more precise multisensory temporal processing abilities. *Attention, Perception, & Psychophysics*, *72*(4), 1120–1129. doi:[10.3758/APP.72.4.1120](https://doi.org/10.3758/APP.72.4.1120)

Donovan, I., Szpiro, S., & Carrasco, M. (2015). Exogenous attention facilitates location transfer of perceptual learning. *Journal of Vision*, *15*(10), 11–11. doi:[10.1167/15.10.11](https://doi.org/10.1167/15.10.11)

Durlach, N. I., & Braida, L. D. (1969). Intensity Perception. I. Preliminary Theory of Intensity Resolution. *The Journal of the Acoustical Society of America*, *46*(2B), 372–383. doi:[10.1121/1.1911699](https://doi.org/10.1121/1.1911699)

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003a). Note on informational masking (L). *The Journal of the Acoustical Society of America*, *113*(6), 2984–2987. doi:[10.1121/1.1570435](https://doi.org/10.1121/1.1570435)

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Jr, G. K. (2003b). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *The Journal of the Acoustical Society of America*, *114*(1), 368–379. doi:[10.1121/1.1577562](https://doi.org/10.1121/1.1577562)

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). The development of attention skills in action video game players. *Neuropsychologia*, *47*(8), 1780–1789. doi:[10.1016/j.neuropsychologia.2009.02.002](https://doi.org/10.1016/j.neuropsychologia.2009.02.002)

Dykstra, A. R., & Gutschalk, A. (2015). Does the mismatch negativity operate on a consciously accessible memory trace? *Science Advances*, *1*(10), e1500677. doi:[10.1126/sciadv.1500677](https://doi.org/10.1126/sciadv.1500677)

Elhilali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between Attention and Bottom-Up Saliency Mediates the Representation of Foreground and Background in an Auditory Scene. *PLOS Biology*, *7*(6), e1000129. doi:[10.1371/journal.pbio.1000129](https://doi.org/10.1371/journal.pbio.1000129)

Feng, L., & Oxenham, A. J. (2015). New perspectives on the measurement and time course of auditory enhancement. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(6), 1696–1708. doi:[10.1037/xhp0000115](https://doi.org/10.1037/xhp0000115)

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*(5777), 43–44. doi:[10.1038/287043a0](https://doi.org/10.1038/287043a0)

Fletcher, H. (1940). Auditory Patterns. *Reviews of Modern Physics*, *12*(1), 47–65. doi:[10.1103/RevModPhys.12.47](https://doi.org/10.1103/RevModPhys.12.47)

Giani, A. S., Belardinelli, P., Ortiz, E., Kleiner, M., & Noppeney, U. (2015). Detecting tones in complex auditory scenes. *NeuroImage*, *122*, 203–213. doi:[10.1016/j.neuroimage.2015.07.001](https://doi.org/10.1016/j.neuroimage.2015.07.001)

Gilbert, C. D. (1994). Early perceptual learning. *Proceedings of the National Academy of Sciences*, *91*(4), 1195–1197. doi:[10.1073/pnas.91.4.1195](https://doi.org/10.1073/pnas.91.4.1195)

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. doi:[10.1038/nature01647](https://doi.org/10.1038/nature01647)

Green, C. S., & Bavelier, D. (2012). Learning, Attentional Control, and Action Video Games. *Current Biology*, *22*(6), R197–R206. doi:[10.1016/j.cub.2012.02.012](https://doi.org/10.1016/j.cub.2012.02.012)

Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-Response Characteristic of Auditory Observers Detecting Signals of a Single Frequency in Noise: The Probe-Signal Method. *The Journal of the Acoustical Society of America*, *44*(6), 1513–1523. doi:[10.1121/1.1911290](https://doi.org/10.1121/1.1911290)

Gutschalk, A., Micheyl, C., & Oxenham, A. J. (2008). Neural Correlates of Auditory Perceptual Awareness under Informational Masking. *PLOS Biol*, *6*(6), e138. doi:[10.1371/journal.pbio.0060138](https://doi.org/10.1371/journal.pbio.0060138)

Harrar, V., Spence, C., & Makin, T. R. (2014). Topographic generalization of tactile perceptual learning. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 15.

Harris, J. D. (1952). The decline of pitch discrimination with time. *Journal of Experimental Psychology*, *43*(2), 96–99. doi:[10.1037/h0057373](https://doi.org/10.1037/h0057373)

Hawkey, D. J., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience*. Retrieved from <http://psycnet.apa.org/psycinfo/2004-19317-008>

Horton, C., Srinivasan, R., & D’Zmura, M. (2014). Envelope responses in single-trial EEG indicate attended speaker in a “cocktail party”. *Journal of Neural Engineering*, *11*(4), 046015. doi:[10.1088/1741-2560/11/4/046015](https://doi.org/10.1088/1741-2560/11/4/046015)

Irvine, D. R. F., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *The Journal of the Acoustical Society of America*, *108*(6), 2964–2968. doi:[10.1121/1.1323465](https://doi.org/10.1121/1.1323465)

Jesteadt, W., & Bilger, R. C. (1974). Intensity and frequency discrimination in one- and two-interval paradigms. *The Journal of the Acoustical Society of America*, *55*(6), 1266–1276. doi:[10.1121/1.1914696](https://doi.org/10.1121/1.1914696)

Jeter, P. E., Doshier, B. A., Liu, S.-H., & Lu, Z.-L. (2010). Specificity of perceptual learning increases with increased training. *Vision Research*, *50*(19), 1928–1940. doi:[10.1016/j.visres.2010.06.016](https://doi.org/10.1016/j.visres.2010.06.016)

John, M. S., Lins, O. G., Boucher, B. L., & Picton, T. W. (1998). Multiple Auditory Steady-state Responses (MASTER): Stimulus and Recording Parameters. *Audiology*, *37*(2), 59–82. doi:[10.3109/00206099809072962](https://doi.org/10.3109/00206099809072962)

John, M. S., Purcell, D. W., Dimitrijevic, A., & Picton, T. W. (2002, May). Advantages and Caveats When Recording Steady-State Responses to Multiple Simultaneous Stimuli [Text]. Retrieved August 29, 2018, from <https://www.ingentaconnect.com/content/aaa/jaaa/2002/00000013/00000005/art00004>

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*(1), 155–163. doi:[10.1093/brain/123.1.155](https://doi.org/10.1093/brain/123.1.155)

Jones, P. R., Moore, D. R., Amitay, S., & Shub, D. E. (2013). Reduction of internal noise in auditory perceptual learning. *The Journal of the Acoustical Society of America*, *133*(2), 970–981. doi:[10.1121/1.4773864](https://doi.org/10.1121/1.4773864)

Jones, P. R., Moore, D. R., Shub, D. E., & Amitay, S. (2014). Learning to detect a tone in unpredictable noise. *The Journal of the Acoustical Society of America*, *135*(3), EL128–133. doi:[10.1121/1.4865267](https://doi.org/10.1121/1.4865267)

Jones, P. R., Moore, D. R., Shub, D. E., & Amitay, S. (2015). The role of response bias in perceptual learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1456–1470. doi:[10.1037/xlm0000111](https://doi.org/10.1037/xlm0000111)

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*(3), 227–229. doi:[10.3758/BF03214307](https://doi.org/10.3758/BF03214307)

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, *88*(11), 4966–4970. doi:[10.1073/pnas.88.11.4966](https://doi.org/10.1073/pnas.88.11.4966)

Kawase, T., Maki, A., Kanno, A., Nakasato, N., Sato, M., & Kobayashi, T. (2012). Contralateral white noise attenuates 40-Hz auditory steady-state fields but not N100m in auditory evoked fields. *NeuroImage*, *59*(2), 1037–1042. doi:[10.1016/j.neuroimage.2011.08.108](https://doi.org/10.1016/j.neuroimage.2011.08.108)

Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America*, *118*(6), 3804–3815. doi:[10.1121/1.2109187](https://doi.org/10.1121/1.2109187)

Kidd, G. J., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory Perception of Sound Sources* (pp. 143–189). Springer US. doi:[10.1007/978-0-387-71305-2_6](https://doi.org/10.1007/978-0-387-71305-2_6)

Kidd, G., Mason, C. R., & Chiu, C. P. (1998a). Identification of brief auditory patterns. *The Journal of the Acoustical Society of America*, *103*(5), 3019–3020. doi:[10.1121/1.422519](https://doi.org/10.1121/1.422519)

Kidd, G., Mason, C. R., & Richards, V. M. (2003). Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *The Journal of the Acoustical Society of America*, *114*(5), 2835–2845. doi:[10.1121/1.1621864](https://doi.org/10.1121/1.1621864)

Kidd, G., Mason, C. R., Rohtla, T. L., & Deliwala, P. S. (1998b). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *104*(1), 422–431. doi:[10.1121/1.423246](https://doi.org/10.1121/1.423246)

Kidd, G., Richards, V. M., Streeter, T., Mason, C. R., & Huang, R. (2011). Contextual effects in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *130*(6), 3926–3938. doi:[10.1121/1.3658442](https://doi.org/10.1121/1.3658442)

Kidd, J. G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *The Journal of the Acoustical Society of America*, *95*(6), 3475–3480. doi:[10.1121/1.410023](https://doi.org/10.1121/1.410023)

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology*, *12*(2 Suppl), 125–143.

Kuai, S.-G., Zhang, J.-Y., Klein, S. A., Levi, D. M., & Yu, C. (2005). The essential role of stimulus temporal patterning in enabling perceptual learning. *Nature Neuroscience*, *8*(11), 1497–1499. doi:[10.1038/nn1546](https://doi.org/10.1038/nn1546)

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. doi:[10.1121/1.1912375](https://doi.org/10.1121/1.1912375)

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, *12*(5), 549–551. doi:[10.1038/nn.2296](https://doi.org/10.1038/nn.2296)

Li, R. W., Ngo, C., Nguyen, J., & Levi, D. M. (2011). Video-Game Play Induces Plasticity in the Visual System of Adults with Amblyopia. *PLOS Biology*, *9*(8), e1001135. doi:[10.1371/journal.pbio.1001135](https://doi.org/10.1371/journal.pbio.1001135)

Linden, R. D., Picton, T. W., Hamel, G., & Campbell, K. B. (1987). Human auditory steady-state evoked potentials during selective attention. *Electroencephalography and Clinical Neurophysiology*, *66*(2), 145–159. doi:[10.1016/0013-4694\(87\)90184-2](https://doi.org/10.1016/0013-4694(87)90184-2)

Mahajan, Y., Davis, C., & Kim, J. (2014). Attentional Modulation of Auditory Steady-State Responses. *PLoS ONE*, *9*(10), e110902. doi:[10.1371/journal.pone.0110902](https://doi.org/10.1371/journal.pone.0110902)

Mathias, S. R., Bailey, P. J., Semal, C., & Demany, L. (2011). A note about insensitivity to pitch-change direction. *The Journal of the Acoustical Society of America*, *130*(4), EL129–EL134. doi:[10.1121/1.3629139](https://doi.org/10.1121/1.3629139)

Mathias, S. R., Micheyl, C., & Bailey, P. J. (2010). Stimulus uncertainty and insensitivity to pitch-change direction. *The Journal of the Acoustical Society of America*, *127*(5), 3026–3037. doi:[10.1121/1.3365252](https://doi.org/10.1121/1.3365252)

Matthews, W. J., & Stewart, N. (2008). The effect of stimulus range on two-interval frequency discrimination. *The Journal of the Acoustical Society of America*, *123*(4), EL45–EL51. doi:[10.1121/1.2884084](https://doi.org/10.1121/1.2884084)

McDermott, H. J. (2004). Music Perception with Cochlear Implants: A Review. *Trends in Amplification*, *8*(2), 49–82. doi:[10.1177/108471380400800203](https://doi.org/10.1177/108471380400800203)

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1–2), 36–47. doi:[10.1016/j.heares.2006.05.004](https://doi.org/10.1016/j.heares.2006.05.004)

Micheyl, C., Shamma, S. A., & Oxenham, A. J. (2007). Hearing Out Repeating Elements in Randomly Varying Multitone Sequences: A Case of Streaming? In P. D. B. Kollmeier, P. D. G. Klump, D. V. Hohmann, D. U. Langemann, D. M. Mauermann, D. S. Uppenkamp, & D. J. Verhey (Eds.), *Hearing – From Sensory Processing to Perception* (pp. 267–274). Springer Berlin Heidelberg. doi:[10.1007/978-3-540-73009-5_29](https://doi.org/10.1007/978-3-540-73009-5_29)

Moore, B. C. J. (1995). *Hearing*. Academic Press.

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch* (pp. 234–277). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/0-387-28958-5_7

Moore, B. C. J., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *The Journal of the Acoustical Society of America*, *91*(5), 2881–2893. doi:[10.1121/1.402925](https://doi.org/10.1121/1.402925)

Moore, D. R., Ferguson, M. A., Halliday, L. F., & Riley, A. (2008). Frequency discrimination in children: Perception, learning and attention. *Hearing Research*, *238*(1–2), 147–154. doi:[10.1016/j.heares.2007.11.013](https://doi.org/10.1016/j.heares.2007.11.013)

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. doi:[10.1080/17470215908416289](https://doi.org/10.1080/17470215908416289)

Nahum, M., Nelken, I., & Ahissar, M. (2010). Stimulus uncertainty and perceptual learning: Similar principles govern auditory and visual learning. *Vision Research*, *50*(4), 391–401. doi:[10.1016/j.visres.2009.09.004](https://doi.org/10.1016/j.visres.2009.09.004)

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *The Journal of the Acoustical Society of America*, *98*(4), 1909–1920. doi:[10.1121/1.414458](https://doi.org/10.1121/1.414458)

Neff, D. L., & Callaghan, B. P. (1988). Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *The Journal of the Acoustical Society of America*, *83*(5), 1833–1838. doi:[10.1121/1.396518](https://doi.org/10.1121/1.396518)

Neff, D. L., & Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *The Journal of the Acoustical Society of America*, *98*(1), 125–134. doi:[10.1121/1.413748](https://doi.org/10.1121/1.413748)

Neff, D. L., Dethlefs, T. M., & Jesteadt, W. (1993). Informational masking for multicomponent maskers with spectral gaps. *The Journal of the Acoustical Society of America*, *94*(6), 3112–3126. doi:[10.1121/1.407217](https://doi.org/10.1121/1.407217)

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception & Psychophysics*, *41*(5), 409–415. doi:[10.3758/BF03203033](https://doi.org/10.3758/BF03203033)

Oh, E. L., & Lutfi, R. A. (2000). Effect of masker harmonicity on informational masking. *The Journal of the Acoustical Society of America*, *108*(2), 706–709. doi:[10.1121/1.429603](https://doi.org/10.1121/1.429603)

O’Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., . . . Lalor, E. C. (2014). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex (New York, N.Y.: 1991)*. doi:[10.1093/cercor/bht355](https://doi.org/10.1093/cercor/bht355)

Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America*, *114*(3), 1543–1549. doi:[10.1121/1.1598197](https://doi.org/10.1121/1.1598197)

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, *59*(3), 640–654. doi:[10.1121/1.380914](https://doi.org/10.1121/1.380914)

Peron, R. M., & Allen, G. L. (1988). Attempts to train novices for beer flavor discrimination: A matter of taste. *The Journal of General Psychology*, *115*(4), 403–418.

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003a). Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. *International Journal of Audiology*, *42*(4), 177–219. doi:[10.3109/14992020309101316](https://doi.org/10.3109/14992020309101316)

Picton, T. W., John, M. S., Purcell, D. W., & Plourde, G. (2003b). Human Auditory Steady-State Responses: The Effects of Recording Technique and State of Arousal. *Anesthesia & Analgesia*, *97*(5), 1396. doi:[10.1213/01.ANE.0000082994.22466.DD](https://doi.org/10.1213/01.ANE.0000082994.22466.DD)

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, *256*(5059), 1018–1021. doi:[10.1126/science.1589770](https://doi.org/10.1126/science.1589770)

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S1), S5–S5. doi:[10.1121/1.1995329](https://doi.org/10.1121/1.1995329)

Rabin, M. D. (1988). Experience facilitates olfactory quality discrimination. *Perception & Psychophysics*, *44*(6), 532–540. doi:[10.3758/BF03207487](https://doi.org/10.3758/BF03207487)

Richards, V. M., & Huang, R. (2003). Criteria placement in staircase procedures. *The Journal of the Acoustical Society of America*, *113*(4), 2287–2287. doi:[10.1121/1.4780607](https://doi.org/10.1121/1.4780607)

Richards, V. M., Huang, R., & Kidd, G. (2004). Masker-first advantage for cues in informational masking. *The Journal of the Acoustical Society of America*, *116*(4), 2278–2288. doi:[10.1121/1.1784433](https://doi.org/10.1121/1.1784433)

Richards, V. M., & Neff, D. L. (2004). Cuing effects for informational masking. *The Journal of the Acoustical Society of America*, *115*(1), 289–300. doi:[10.1121/1.1631942](https://doi.org/10.1121/1.1631942)

Ross, B., Miyazaki, T., & Fujioka, T. (2012). Interference in dichotic listening: The effect of contralateral noise on oscillatory brain networks. *European Journal of Neuroscience*, *35*(1), 106–118. doi:[10.1111/j.1460-9568.2011.07935.x](https://doi.org/10.1111/j.1460-9568.2011.07935.x)

Roth, D. A.-E., Amir, O., Alaluf, L., Buchsenspanner, S., & Kishon-Rabin, L. (2003). The Effect of Training on Frequency Discrimination: Generalization to Untrained Frequencies and to the Untrained Ear. *Journal of Basic and Clinical Physiology and Pharmacology*, *14*(2), 137–150. doi:[10.1515/JBCPP.2003.14.2.137](https://doi.org/10.1515/JBCPP.2003.14.2.137)

Roth, D. A.-E., Appelbaum, M., Milo, C., & Kishon-Rabin, L. (2008). Generalization to Untrained Conditions Following Training with Identical Stimuli. *Journal of Basic and Clinical Physiology and Pharmacology*, *19*(3-4). doi:[10.1515/JBCPP.2008.19.3-4.223](https://doi.org/10.1515/JBCPP.2008.19.3-4.223)

Roth, D. A.-e., Refael-Taub, R., Sharvit, R., & Kishon-Rabin, L. (2006). Can learning a frequency discrimination task occur without discrimination? *Journal of Basic and Clinical Physiology and Pharmacology*, *17*(3), 159–171.

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision Research*, *51*(13), 1552–1566. doi:[10.1016/j.visres.2010.10.019](https://doi.org/10.1016/j.visres.2010.10.019)

Sathian, K., & Zangaladze, A. (1997). Tactile learning is task specific but transfers between fingers. *Perception & Psychophysics*, *59*(1), 119–128. doi:[10.3758/BF03206854](https://doi.org/10.3758/BF03206854)

Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Perception & Psychophysics*, *42*(3), 215–223. doi:[10.3758/BF03203073](https://doi.org/10.3758/BF03203073)

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocular. *The Journal of Physiology*, *483*(3), 797–810. doi:[10.1113/jphysiol.1995.sp020623](https://doi.org/10.1113/jphysiol.1995.sp020623)

Semal, C., & Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *The Journal of the Acoustical Society of America*, *120*(6), 3907–3915. doi:[10.1121/1.2357708](https://doi.org/10.1121/1.2357708)

Shiu, L.-P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics*, *52*(5), 582–588. doi:[10.3758/BF03206720](https://doi.org/10.3758/BF03206720)

Skosnik, P. D., Krishnan, G. P., & O'Donnell, B. F. (2007). The effect of selective attention on the gamma-band auditory steady-state response. *Neuroscience Letters*, *420*(3), 223–228. doi:[10.1016/j.neulet.2007.04.072](https://doi.org/10.1016/j.neulet.2007.04.072)

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *76*(6), 1690–1695. Retrieved from <http://cat.inist.fr/?aModele=afficheN&cpsidt=9185083>

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, J. G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, *5*. doi:[10.1038/srep11628](https://doi.org/10.1038/srep11628)

Szpiro, S. F. A., Wright, B. A., & Carrasco, M. (2014). Learning one task by interleaving practice with another task. *Vision Research*, *101*(Supplement C), 118–124. doi:[10.1016/j.visres.2014.06.004](https://doi.org/10.1016/j.visres.2014.06.004)

Tramo, M. J., Shah, G. D., & Braida, L. D. (2002). Functional Role of Auditory Cortex in Frequency Processing and Pitch Perception. *Journal of Neurophysiology*, *87*(1), 122–139. doi:[10.1152/jn.00104.1999](https://doi.org/10.1152/jn.00104.1999)

Treisman, A. M. (1960). Contextual Cues in Selective Listening , Contextual Cues in Selective Listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248. doi:[10.1080/17470216008416732](https://doi.org/10.1080/17470216008416732)

Whitton, J. P., Hancock, K. E., & Polley, D. B. (2014). Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. *Proceedings of the National Academy of Sciences*, 201322184. doi:[10.1073/pnas.1322184111](https://doi.org/10.1073/pnas.1322184111)

Whitton, J. P., Hancock, K. E., Shannon, J. M., & Polley, D. B. (2017). Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise. *Current Biology*, 0(0). doi:[10.1016/j.cub.2017.09.014](https://doi.org/10.1016/j.cub.2017.09.014)

Wiegand, K., & Gutschalk, A. (2012). Correlates of perceptual awareness in human primary auditory cortex revealed by an informational masking experiment. *NeuroImage*, 61(1), 62–69. doi:[10.1016/j.neuroimage.2012.02.067](https://doi.org/10.1016/j.neuroimage.2012.02.067)

Wier, C. C., Jesteadt, W., & Green, D. M. (1976). A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination. *Perception & Psychophysics*, 19(1), 75–79. doi:[10.3758/BF03199389](https://doi.org/10.3758/BF03199389)

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. *Journal of Speech Language and Hearing Research*, 50(4), 844. doi:[10.1044/1092-4388\(2007/059](https://doi.org/10.1044/1092-4388(2007/059)

Wright, B. A., Sabin, A. T., Zhang, Y., Marrone, N., & Fitzgerald, M. B. (2010). Enhancing Perceptual Learning by Combining Practice with Periods of Additional Sensory Stimulation. *The Journal of Neuroscience*, 30(38), 12868–12877. doi:[10.1523/JNEUROSCI.0487-10.2010](https://doi.org/10.1523/JNEUROSCI.0487-10.2010)

Xiao, L.-Q., Zhang, J.-Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete Transfer of Perceptual Learning across Retinal Locations Enabled by Double Training. *Current Biology*, 18(24), 1922–1926. doi:[10.1016/j.cub.2008.10.030](https://doi.org/10.1016/j.cub.2008.10.030)

Yu, C., Klein, S. A., & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision*, 4(3), 4–4. doi:[10.1167/4.3.4](https://doi.org/10.1167/4.3.4)

Zhang, J.-Y., Zhang, G.-L., Xiao, L.-Q., Klein, S. A., Levi, D. M., & Yu, C. (2010a). Rule-Based Learning Explains Visual Perceptual Learning and Its Specificity and Transfer. *The Journal of Neuroscience*, 30(37), 12323–12328. doi:[10.1523/JNEUROSCI.0704-10.2010](https://doi.org/10.1523/JNEUROSCI.0704-10.2010)

Zhang, T., Xiao, L.-Q., Klein, S. A., Levi, D. M., & Yu, C. (2010b). Decoupling location specificity from perceptual learning of orientation discrimination. *Vision Research*, 50(4), 368–374. doi:[10.1016/j.visres.2009.08.024](https://doi.org/10.1016/j.visres.2009.08.024)

Zhang, Y.-X., Moore, D. R., Guiraud, J., Molloy, K., Yan, T.-T., & Amitay, S. (2016). Auditory Discrimination Learning: Role of Working Memory. *PLOS ONE*, *11*(1), e0147320. doi:[10.1371/journal.pone.0147320](https://doi.org/10.1371/journal.pone.0147320)

RÉSUMÉ

Pour s'adapter à un environnement changeant, il est important que les systèmes perceptifs gardent la possibilité de se transformer et de faire l'objet d'apprentissages tout au long de la vie. Dans cette thèse, nous avons étudié des mécanismes d'apprentissage auditif dans le domaine de la perception de la hauteur et de l'attention sélective. Nous avons développé des protocoles comportementaux sur tablette permettant la mise en action de la boucle audio-motrice, étudié leur pertinence dans des tâches de perception de hauteur, fait de même pour des tâches de masquage informationnel, et suggéré une méthode électrophysiologique (EEG) visant à caractériser l'attention sélective. Les résultats montrent que des effets d'apprentissage peuvent être observés dans les tâches de masquage informationnel, sans spécificité évidente pour l'entraînement audio-moteur, mais avec un effet de la zone fréquentielle inconnu jusqu'ici.

MOTS CLÉS

Entraînement perceptif, Boucle audio-motrice, Attention sélective auditive, Perception de la hauteur, masquage informationnel

ABSTRACT

To adapt to an ever-changing environment, it is important that perceptual systems are able to transform themselves and exhibit learning throughout life. In this thesis, we have studied auditory perceptual learning of pitch and selective attention. We have developed behavioral training protocols on tablets requiring the use of the audio-motor loop and evaluated them on pitch perception and informational masking tasks. We also suggested an electrophysiological method (EEG) aiming at characterizing auditory selective attention. The main results showed that rapid perceptual learning can be observed in informational masking tasks, but without any obvious advantage for audio-motor training. In addition, a previously unknown effect of frequency on informational masking was discovered, suggesting the existence of an "informational audiogram".

KEYWORDS

Perceptual learning, Audio-motor loop, Auditory selective attention, Pitch perception, informational masking