

Mémoire de Master 2 :
Relations entre Perception Globale
et Composition de Séquences Sonores

Aymeric DEVERGIE
Etudiant en Master 2 Informatique, spécialité SAR, parcours
ATIAM
Université Pierre et Marie Curie, Jussieu, Paris VI

Laboratoire d'accueil : IRCAM, Paris
Equipe : Perception et Design Sonore
Responsable : Patrick SUSINI

Mars / Septembre 2006

Résumé

Peu de travaux consacrés à l'étude perceptive de séquences de sons de l'environnement tiennent compte des effets de la mémoire auditive à court terme. Ce rapport présente une étude de l'influence de ces effets sur le jugement global de séquences sonores. Pour étudier ces effets, nous proposons aux sujets des séquences composées de trois catégories de sons : Trafic, Parc et Présence humaine. A la fin de chaque séquence, nous présentons un label correspondant à l'une des catégories. Le sujet doit répondre le plus rapidement possible à la question : "Est-ce que le mot évoque correctement la séquence?". Nous mesurons des temps de réponse. Le jugement global porte sur l'ensemble de la séquence. Il n'est pas demandé aux sujets de tenir compte uniquement des derniers éléments entendus mais de la totalité de la séquence pour construire leur jugement. En analysant les données, nous évaluons l'effet de la position de la catégorie cible sur le temps de réponse. Nous démontrons que le jugement global dépend de la récence de la catégorie cible. Plus la cible est récente, plus le taux de réponse "oui" est élevé et plus le temps de réponse est court. Ces résultats sont en accord avec les études menées sur la mémoire auditive des sons verbaux et non verbaux. Nous observons, d'autre part, des différences de taux et temps de réponse en fonction des catégories de sons. Les résultats montrent un net avantage de mémorisation pour les sons humains. Pour chaque catégorie, le temps de réponse évolue différemment en fonction de la position temporelle de celle-ci dans la séquence.

Remerciements

Je tiens à remercier :

Patrick SUSINI pour m'avoir accueilli dans son équipe, et également pour m'avoir proposé un sujet de stage aussi passionnant. Merci pour son dévouement et son soutien.

Olivier HOUIX et Nicolas MISDARIIS pour les conseils avisés, et leurs réflexions pertinentes.

Toute l'équipe Perception et Design Sonore pour sa bonne humeur intarissable, avec une mention spéciale pour Julien T.

Merci à tous pour votre sens de la communication et du partage. L'échange de point de vues au cours des réunions d'équipe s'est avéré extrêmement constructif et motivant.

Introduction

Le travail rapporté ici se déroule dans le cadre de recherches concernant la perception de séquences sonores menées dans l'équipe Perception et Design sonore à l'IRCAM. Cette équipe, créée en 1999, s'intéresse à la perception auditive de stimuli non verbaux. L'une de ses missions consiste à évaluer les processus cognitifs intervenant dans la perception auditive. Nous nous intéressons ici à l'influence de la composition temporelle sur le jugement global de séquences composées de sons de l'environnement. Cette thématique fédère deux axes de recherches. Le premier axe concerne la représentation cognitive des sons de l'environnement en termes de catégories sémantiques. Le second axe de recherche, qui est d'avantage le point central de cette étude, s'intéresse aux effets de la mémoire dans le cas de la perception globale d'une séquence sonore.

En effet, les travaux menés dans le domaine de la mémoire auditive avec des sons verbaux mettent en évidence une limitation temporelle à court terme concernant la rétention d'informations ; limitation aussi bien au niveau des mécanismes (boucle phonologique) qu'au niveau sensoriel (mémoire échoïque). Par conséquent, la courbe de rappel généralement obtenue pour des sons verbaux a une forme de "U" indiquant un bon rappel pour les premiers (effet de primauté) et les derniers (effet de récence) éléments de la liste. L'avantage pour les derniers éléments est un effet essentiellement observé lors d'une présentation auditive. En 2005, P. Guelton et al. ont mis en évidence des effets similaires pour les listes de sons de l'environnement, avec cependant une capacité de rétention accrue par rapport aux sons verbaux. D'autre part, P. Susini et al., en 2002, ont montré que le jugement global en sonie d'un son pur de 1 kHz non stationnaire sur quelques dizaines de secondes dépend de la position temporelle des variations du niveau sonore. Dans ce cas, les modèles classiques de mesure de la sonie échouent pour rendre compte de la sonie globale. Il apparaît donc naturel de s'interroger sur l'influence de la mémoire à court terme sur la perception d'une séquence temporelle. Notre objectif, à travers cette étude, est de prendre en compte les effets de mémoire auditive pour analyser les processus perceptifs mis en jeu lors d'une tâche consistant à évaluer l'impression globale d'une séquence sonore.

Nous dresserons, dans le chapitre 1, un état de l'art concernant la perception de séquences sonores, la structure et les mécanismes de la mémoire. Nous aborderons également les effets de la mémoire sur la perception auditive. Nous détaillerons, ensuite, dans le chapitre 2, l'expérience proprement dite. Nous discuterons à partir des résultats de

l'influence de l'organisation temporelle sur le jugement global de séquences de sons de l'environnement. Pour finir, nous aborderons, dans le chapitre 4 les perspectives envisagées pour poursuivre dans cet axe de recherche.

Table des matières

Introduction	iii
1 Etat de l'art	1
1.1 Perception de séquences sonores	1
1.1.1 Séquences de sons purs et sons verbaux	1
1.1.2 Séquences de sons de l'environnement	2
1.2 Effets de la mémoire sur la perception globale	4
1.2.1 Présentation succincte de la mémoire	4
1.2.2 Manifestations de la mémoire à court terme : Récence et Primauté	7
1.2.3 Effets inhérents au protocole expérimental	9
2 Expérience	11
2.1 Présentation	11
2.1.1 Rappel des objectifs	11
2.1.2 Protocole	11
2.1.3 Mesure du temps de réaction	12
2.2 Sélection et description du Matériau sonore	13
2.2.1 Extraction d'une base de sons existante	13
2.2.2 Catégorisation des sources sonores	14
2.2.3 Définition et ajustement du niveau sonore	18
2.3 Matériel	20
2.4 Stimuli	21
2.4.1 Sons utilisés	21
2.4.2 Organisation temporelle des séquences	21
2.4.3 Mots cibles	21
2.4.4 Bruit de fond	23
2.5 Procédure	23
2.6 Sujets	25
2.7 Résultats	25
2.7.1 Résultats concernant les séquences Test	25
2.7.2 Pourcentage de réponse pour chaque catégorie en fonction de la position temporelle	26
2.7.3 Temps de réponse pour chaque catégorie cible	27

2.7.4	Pourcentage et Temps de réponse pour la catégorie Traffic	28
2.7.5	Discussion des résultats	29
2.8	Discussion générale	32
2.8.1	Effet de la mémoire	32
2.8.2	Sons de l'environnement	33
3	Conclusion	36
4	Perspectives	37
4.1	Expérience de jugement global de séquences	37
4.2	Expériences de rappel de listes de sons environnementaux	37
4.3	Expérience de jugement pour chaque catégorie	38
4.4	Expérience de jugement d'agrément	38
A	Description du bruit de fond	39
A.1	Description du bruit de fond	39
A.1.1	Intérêt	39
A.1.2	Analyse Spectrale	39
A.1.3	Synthèse par Méthode LPC	40
B	Matériau sonore	43
B.1	Liste complète des sons extraits de la base de sons	43
C	Ajustement écologique en sonie	44

Table des figures

1.1	Modèle de la Mémoire Sémantique selon Collins et al.	6
1.2	Modèle de la Mémoire Sémantique par Comparaison d'attributs selon Smith et al. . . .	6
1.3	Récence et Primauté pour une liste d'items verbaux	7
1.4	Récence et Primauté pour une liste de sons de l'environnement	8
2.1	Déroulement des séquences au cours de l'expérience de Jugement global	11
2.2	Représentation arborée de la catégorisation des sons	16
2.3	Liste des sons du corpus définitif	17
2.4	Prototypes retenus par les sujets	18
2.5	Interface d'ajustement en Sonie sous PsiExp	19
2.6	Structure temporelle d'une séquence "type"	22
2.7	Structure temporelle d'une séquence "test"	22
2.8	Interface Utilisateur pour l'expérience de Jugement Global sous Max/MSP	24
2.9	Pourcentage de réponse en fonction du type de séquence	25
2.10	Effectif pour chaque catégorie en fonction de la position de la cible	26
2.11	Temps de réponse en fonction de la catégorie cible	27
2.12	% de l'effectif pour la catégorie Trafic en fonction des configurations	28
2.13	Temps de réponse pour la catégorie Trafic en fonction des configurations	29
A.1	Méthode LPC	40
A.2	Bruit de fond étudié par V. Maffiolo	41
A.3	Spectre et enveloppe spectrale du bruit de fond analysé	41
A.4	Spectre et Enveloppe spectrale du bruit de fond synthétisé	42
C.1	Niveaux écologique en dB SPL des sons de Présence Humaine	44
C.2	Niveaux écologique en dB SPL des sons de Parc	45
C.3	Niveaux écologique en dB SPL des sons de Trafic	45

Chapitre 1

Etat de l'art

Nous rapportons dans ce chapitre les travaux concernant les deux axes que nous avons dégagé dans l'introduction à savoir la représentation des sons de l'environnement en catégories sémantiques et les effets de la mémoire sur la perception de séquences sonores. Dans un premier temps, nous présentons les travaux concernant la perception de séquences de sons. Puis, nous abordons un chapitre concernant les effets de la mémoire sur la perception de séquences sonores.

1.1 Perception de séquences sonores

Les travaux mentionnés dans cette section révèlent des effets de la mémoire auditive. Historiquement, l'étude des séquences de sons a débuté avec des sons purs. Ces stimuli sont simples à produire et les paramètres acoustiques sont contrôlables. Par la suite, des études ont été menées sur la perception de séquences de sons verbaux. Enfin, nous présentons une étude concernant la perception de séquences de sons environnementaux. Ce type d'étude manque cependant de contrôle permettant de mettre en évidence des effets de la mémoire.

1.1.1 Séquences de sons purs et sons verbaux

Séquence de sons purs

En 2002, P. Susini et al. [SMS02] ont mené une expérience concernant l'estimation instantanée et globale de l'intensité d'un son pur variant en intensité au cours du temps. L'hypothèse qui a conduit à ces travaux est la suivante : les caractéristiques temporelles des variations de l'intensité modifient la perception globale d'intensité sonore. Le stimulus présenté au sujet est un son pur de 1 kHz dont l'intensité varie entre 60 et 90 dB SPL.

Deux jugements sont demandés aux sujets. Le premier est un jugement instantané suivi d'un jugement global en sonie. Le second jugement demandé est un jugement global en fin de présentation. La position temporelle des pics d'intensité conditionne le jugement global d'intensité. Les positions qui influencent le plus le jugement sont celles pour

lesquelles les effets de mémoire à court terme sont présents, c'est à dire au début (effet de primauté) et à la fin (effet de récence) de la séquence.

Séquence de voix parlée

L'étude de la voix se distingue de l'étude de sons purs. La voix est le vecteur d'informations sémantiques. Notre connaissance de la langue est impliquée dans l'écoute de la voix. Cela modifie profondément notre manière de percevoir une séquence de voix parlée.

L. Gros et al. [GC01] en 2001, ont étudié des séquences de voix parlée. Cette étude porte sur le jugement instantané et global de la qualité de la voix. Cette étude est menée pour appréhender les effets de perte de qualité de la voix dans les transmissions téléphoniques. Le stimulus utilisé dans l'expérience est un discours prononcé par une personne, numérisé et traité par informatique de manière à simuler les dégradations inhérentes à la transmission par téléphonie. La qualité de la voix évolue au cours du temps. Le jugement demandé est un jugement instantané, à donner pendant la diffusion de la séquence, suivi d'un jugement global à la fin de l'écoute.

Le jugement instantané suit parfaitement l'évolution de la qualité de la voix dans la séquence. Les sujets perçoivent plus rapidement les dégradations que les améliorations de la qualité. Le sujet crée, selon les auteurs de cette étude, des attentes lorsqu'il est confronté à des stimuli qui ont une signification. Plus la modification apparaît tard, plus le jugement global est affecté. Les sons purs ne sont pas des sons auxquels nous sommes exposés quotidiennement.

1.1.2 Séquences de sons de l'environnement

Les sons de l'environnement constituent une catégorie distincte à la fois des sons verbaux et des sons purs. Historiquement, les études de stimuli sonores ont concerné les sons purs, puis les sons verbaux. Les sons de l'environnement sont en fait tous les sons qui n'entrent pas dans ces deux catégories. Chaque son que nous percevons active une ou plusieurs connaissances ou concepts acquis antérieurement. Nous percevons et identifions les sources sonores qui nous entourent, nous en parlons. Nous sommes parfois amenés à décrire les qualités du son. Nous comparons les stimulations sensorielles avec nos représentations en mémoire. Ce processus d'identification se déroule avant tout processus de qualification du son. Les enquêtes de terrains et le recueil d'impressions des sujets permettent de constituer des corpus de sons pertinents. Ce point sera évoqué ici, en même tant que les paramètres proposés par J. Ballas concernant l'identifiabilité des sons environnementaux.

Facteurs influençant l'identification

En 1993, Ballas [A.B93] mène une étude permettant de définir les facteurs impliqués dans l'identification des sons brefs de l'environnement. Ballas définit la notion d'identifiabilité comme étant le temps nécessaire à l'identification des sons de l'environnement. Les facteurs qu'il met en évidence ne sont pas uniquement des facteurs acoustiques.

Le premier paramètre qui nous intéresse est la typicalité. Rosh définit, en 1975, la notion de typicalité. Un item est typique s'il est représentatif de la catégorie à laquelle il appartient. J. Ballas affirme que la typicalité influence le temps d'identification, ce qui corrobore les données recueillies par Collins et Quillian,[CQ69], pour expliquer l'organisation hiérarchique de nos connaissances. Nous reviendrons en détail sur ce modèle d'organisation. J. Ballas identifie comme second paramètre important le contexte. Un son présenté en dehors de son contexte est plus difficilement identifié.

Enquêtes de terrain

Pour recueillir l'impression des sujets confrontés à des environnements sonores, des enquêtes de terrain ont été menées. Ce type d'étude permet d'apprécier le jugement et la représentation d'espaces sonores selon le point de vue des sujets.

V. Maffiolo et al. [MVP⁺97] mènent, en 1997, une enquête auprès de la population parisienne. Celle-ci est menée dans le but de créer un corpus de sons de l'environnement tenant compte des impressions et jugements des individus. Il est demandé aux sujets de dessiner un environnement sonore. Par la suite, les sujets doivent évoquer un espace sonore de mémoire. Les représentations proposées par les sujets sont de 4 types, à savoir : des schémas abstraits, des cartographies, des lieux spécifiques ou encore des sources non organisées dans un site. Cette étude a permis d'identifier les lieux les plus souvent cités, et de montrer qu'il existe une structure dans l'appréhension d'un lieu, en terme de topologie, d'appréciation qualitative, du rapport au sujet qui en parle. La place que tient le sujet et l'activité qu'il peut avoir dans la scène sont aussi mentionnées dans leur description.

Appréciations subjectives de séquences sonores

V. Maffiolo et al. [MCD98] proposent, en 1998, une étude concernant le jugement de séquences réelles. Ces derniers démontrent que l'appréciation subjective et la stratégie d'écoute dépend du contenu de la séquence.

Les extraits sont choisis à partir de résultats d'enquêtes menées auprès des sujets, comme celles mentionnées plus haut.

Chaque scène extraite des ambiances, enregistrées in situ, dure entre 15 et 20 secondes. Deux types de scènes sont extraites : les scènes "événementielles", consistant en un ensemble d'événements discriminables et identifiables et les scènes "amorphes" composées d'un bruit de fond continu sans événements particuliers. Les auteurs démontrent que les scènes événementielles sont traitées sémantiquement et que les scènes amorphes le sont physiquement.

Le caractère événementiel ou amorphe a une grande influence sur l'évaluation de l'intensité perçue par les sujets. Pour une étude des sons de l'environnement le nombre et l'identification de sons discriminables est un critère pertinent. Les données verbales sont essentiellement constituées de deux types de descripteurs : les descripteurs de paramètres physiques et les descripteurs de sources. L'hypothèse est vérifiée.

Séquence de sons de l'environnement

A l'instar de J. Hellbrück, S. Namba et S. Kuwano [HKZ⁺01], R. Weber mène, en 1991 [Web91], une étude sur la perception de bruits de circulation. Cette étude concerne le jugement subjectif des variations continues du niveau sonore pour séquences enregistrées en situation réelle. Le niveau réel en dBA est mesuré simultanément. Cette étude révèle une dépendance non linéaire entre le jugement subjectif et niveau réel mesuré. Les auteurs notent aussi une grande variabilité entre les sujets. Les sujets détectent avec un décalage temporel les variations de niveau sonore. C'est en cela que la dépendance est non linéaire.

Les trois études concernant les séquences sonores (sons purs, sons verbaux et sons de l'environnement) montrent que les jugements d'intensité sonore pour les sons purs et de qualité pour la parole dépendent de l'organisation temporelle des séquences présentées. Nous étudierons des séquences de sons de l'environnement dont l'agencement temporel sera contrôlé. Cet agencement sera également sémantique. Pour savoir si le contenu sémantique des sons a une incidence sur la perception de séquences, nous organiserons les séquences en fonction de la catégorie à laquelle appartient chacun des événements sonores.

1.2 Effets de la mémoire sur la perception globale

Pour comprendre comment la mémoire peut influencer la perception globale nous présenterons succinctement l'organisation générale de la mémoire. Nous détaillerons davantage l'organisation de nos connaissances en mémoire. Ensuite, nous présenterons deux manifestations de la mémoire à court terme ; l'effet de récence et l'effet de primauté. Enfin, nous exposerons des effets que nous devrons intégrer pour mettre en place le protocole expérimental.

1.2.1 Présentation succincte de la mémoire

Structure de la mémoire

La structure de la mémoire que nous exposons ici repose sur un modèle proposé en 1968 par Atkinson et Shiffrin. Notre objectif ne consiste pas à étudier les modèles les plus récents proposés dans la littérature. Des améliorations de ce modèle sont proposées, mais dans le fond la distinction entre le court et le long terme reste d'actualité.

La mémoire sensorielle La mémoire sensorielle peut se définir comme la rétention temporaire des effets d'une stimulation sensorielle. C'est la trace mnésique la plus courte. Elle peut être maintenue sur une période de 300 à 500 ms. Elle est dite échoïque pour les stimuli auditifs, et iconique pour les stimuli visuels.

Elle a pour fonction de collecter les stimuli qui vont être traités. Lorsque le stimulus parvient au système auditif périphérique de manière intermittente, la mémoire échoïque permet l'illusion de continuité auditive.

La mémoire à court terme Dénommée mémoire à court terme, mémoire immédiate ou encore mémoire primaire, elle permet la restitution immédiate de l'information perçue. Les stimuli peuvent être maintenus pendant 30 à 90 secondes. La mémoire à court terme est sensible aux interférences. Elle est également sensible à l'interposition de stimuli parasite entre le test et le rappel de la liste. La mémoire à court terme est également sensible aux troubles attentionnels, comme la baisse de la concentration.

La mémoire à court terme, simple unité de stockage, est décrite par Baddeley et Hitch (1974), [Bad02], permet d'expliquer un lien entre MCT et MLT. La MCT devient la mémoire de travail et elle se divise alors en quatre sous systèmes. Cette représentation porte le nom de modèle modal. Les quatre sous systèmes sont : la boucle phonologique, le registre visuo-spatial, le buffer épisodique et le centre exécutif.

La mémoire à long terme La mémoire à long terme est la mémoire de stockage permanent. L'accès aux informations contenues dans cette mémoire est un accès lent comparé à l'accès dans la MCT. Il est courant de séparer au sein de la MLT la mémoire explicite ou déclarative de la mémoire implicite ou non déclarative. La mémoire explicite fait référence à tout ce que nous apprenons par association entre un stimulus et une étiquette, une représentation. Le langage constitue un apprentissage explicite. A l'opposé, l'apprentissage implicite n'établit pas de relation univoque et précise entre un stimulus et une représentation en mémoire. L'apprentissage de fonctions locomotrices, comme apprendre à faire du vélo, par exemple, est un apprentissage implicite.

Organisation des connaissances en mémoire sémantique

Modèle hiérarchique Le premier modèle est proposé par Collins et Quillian en 1969. Il s'agit du modèle hiérarchique (figure 1.1). Les connaissances des mots sont hiérarchisées sémantiquement. Chaque noeud représente un concept. Les concepts de bas niveau sont inclus dans les concepts de haut niveau. En haut de la hiérarchie se trouvent les concepts génériques, par exemple : véhicule. Au niveau de base, niveau intermédiaire, nous trouvons, voiture. Enfin, au niveau inférieur, nous trouvons les concepts détaillés, berline. Ceci est l'inclusion de concepts, tel que le définit Collins. [Gal05]. L'économie cognitive s'est réduite la quantité d'informations à mémoriser en "factorisant" les propriétés de chaque concept. Les concepts des noeuds inférieurs héritent des propriétés des niveaux supérieurs. Il n'y a pas de redondance des propriétés.

Modèle par comparaisons d'attributs Smith, Shoben et Rips proposent, en 1974, le modèle par "comparaison d'attributs". Le sens des mots est défini par ses attributs. (figure 1.2). Il y a deux types d'attributs : les attributs de définition et les attributs de caractérisation. Si un exemplaire ne présente pas les attributs de définition d'une catégorie, il n'appartient pas à cette catégorie. S'il ne présente pas un attribut de caractérisation, il peut tout de même appartenir à la catégorie. Lors de la présentation d'un stimulus, un processus de comparaison s'effectue en deux étapes, toujours d'après Smith et al. . La première étape consiste en la comparaison de tous les attributs, afin de juger du degré de similarité entre le test et le prédicat en mémoire. Si les deux exemplaires sont très

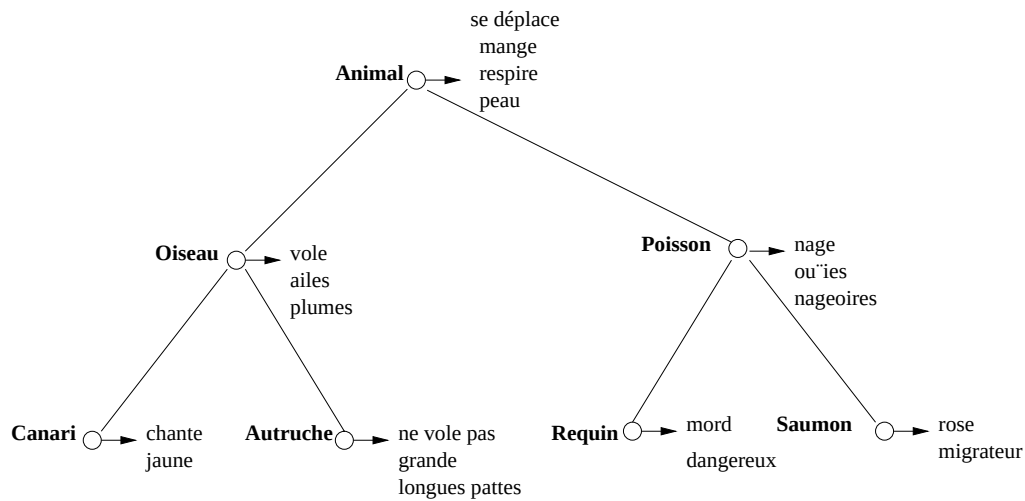


FIG. 1.1 – Modèle de la Mémoire Sémantique selon Collins et al.

semblables ou bien très dissemblables alors le processus de comparaison est terminé. Si la situation n'est pas aussi tranchée, une seconde comparaison est nécessaire. La comparaison porte sur les attributs de définition. Ce modèle permet d'expliquer les effets de typicalité. Des items similaires vont nécessiter une seule étape de comparaison. Ainsi une forte typicalité suscite des temps de réponse courts. Si les items sont peu similaires, les temps de réponses seront plus long.

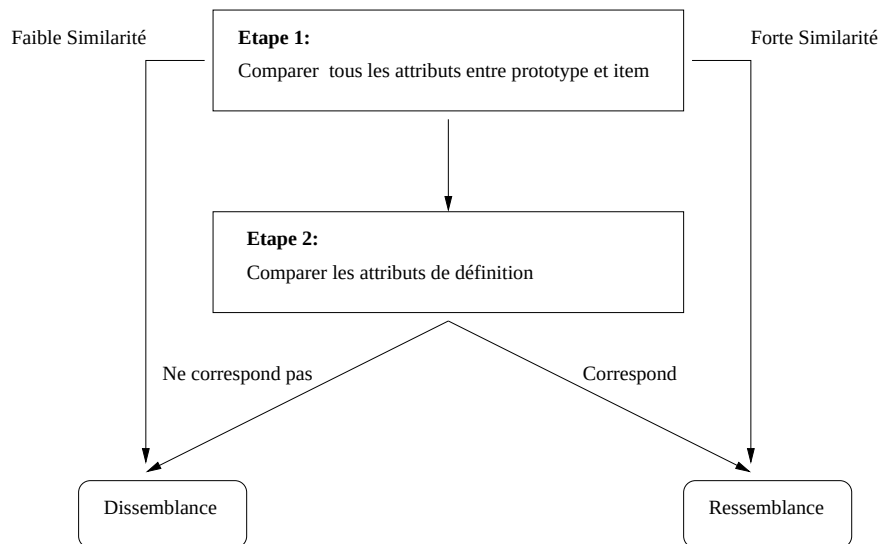


FIG. 1.2 – Modèle de la Mémoire Sémantique par Comparaison d'attributs selon Smith et al.

1.2.2 Manifestations de la mémoire à court terme : Récence et Primauté

Les effets connus sous le nom de récence et de primauté sont des manifestations claires de processus de notre mémoire. Glanzer et Cunitz (1966) montrent que l'effet de récence est dû à notre mémoire à court terme. Pour cela, ils proposent une tâche de rappel de liste de mots. L'effet de récence est réduit s'ils demandent aux sujets de compter à rebours jusqu'à 30 avant de rappeler les mots. Cette tâche intermédiaire affecte le contenu de la mémoire à court terme.

Les deux auteurs étudient également l'effet de primauté. Ils affirment que l'effet de primauté est dû à notre mémoire à long terme. Pour la même tâche de rappel libre, si le nombre de mots diminue, la disponibilité en mémoire est plus grande et l'effet de primauté observé est plus important. La figure 1.3 illustre ces effets dans une tâche de rappel d'une liste d'items verbaux.

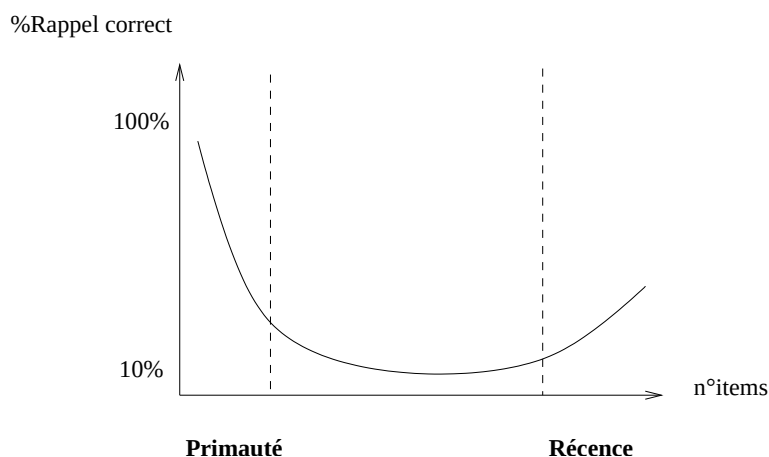


FIG. 1.3 – Récence et Primauté pour une liste d'items verbaux

Travaux antérieurs

Les travaux de P. Guelton et al. , en 2005 [Gue05], concernent la perception et la mémorisation des listes de sons de l'environnement.

Les travaux ont été menés afin d'étudier des tâches de rappel libre et sériel d'une liste de sons de l'environnement. L'hypothèse éprouvée dans cette étude est que les sons environnementaux, par leur complexité, sont moins bien rappelés que des sons verbaux. Les stimuli présentés aux sujets sont des sons environnementaux d'une durée moyenne comprise entre 1 et 4 secondes, échantillonnés à 44100 Hz, mixé en stéréo sur 32 bits. Les listes comprennent 14 sons. Il est demandé aux sujets différentes tâches. Pour un premier groupe, le rappel est libre. L'ordre dans lequel les sujets se rappellent n'est pas important. Seul compte la position du mot dans la liste. Un second groupe doit en revanche rappeler

les mots dans l'ordre dans lequel ils sont apparus. A la fin, les deux groupes sont amenés à identifier et nommer les sources présentées.

Les résultats de la figure 1.4 infirment l'hypothèse. Le taux de rappel moyen est meilleur que pour le rappel de sons verbaux. On peut donc supposer que le processus cognitif d'identification a permis de mieux mémoriser les stimuli présentés. Cette hypothèse est à vérifier.

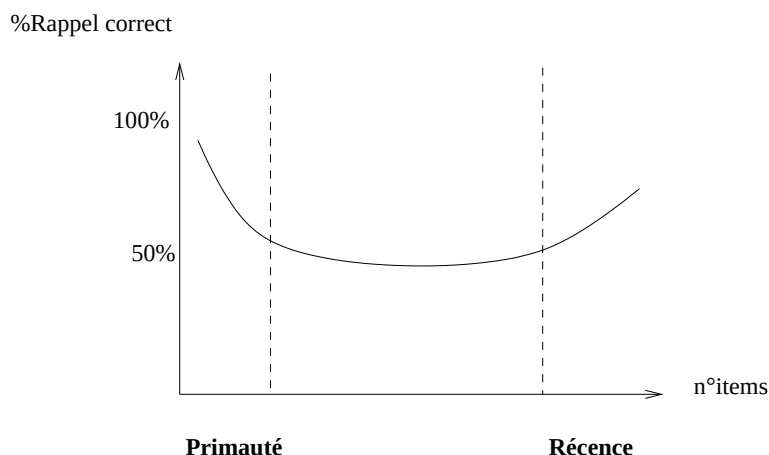


FIG. 1.4 – Récence et Primauté pour une liste de sons de l'environnement

D'autre part, les travaux mettent en évidence un phénomène de récence relative. Le phénomène de récence est dû à notre mémoire à court terme. Son empan est limité. Les travaux menés ici montrent que l'effet persiste alors que le nombre d'items concernés par cet effet est supérieur au nombre proposé par Miller. L'origine de cette prolongation de l'effet de récence reste peu comprise.

Sur un autre plan, D.L. Hintzman, [Hin03] étudie, en 2003, l'effet de l'identification des items présentés sur le jugement de récence. Pour les deux premières expériences, les stimuli sont des listes de 550 mots différents. Les mots sont répétés de 1 à 6 fois dans la liste présentée. La liste de mots est de longueur variable comprise entre 5 et 30 items. Une liste de mots est présentée sur un écran. Certains mots sont répétés. Le sujet doit dire si le mot qui apparaît est nouveau ou ancien. Si le mot est ancien, le sujet doit dire quelle est l'ancienneté du mot. Pour l'expérience suivante, le jugement de récence est fait sur une échelle de 1 à 6. Un enjeu de compétition avec les autres sujets est mis en place. Cela permet de renforcer la motivation du sujet.

Les jugements de récence sont meilleurs pour les catégories de mots reconnus. La trace en mémoire pour les mots anciens est plus forte et ce quelque soit la durée des listes. Le jugement de récence tient également compte de la modalité de présentation. Les jugements de récence et l'identification empreignent des chemins différents. La fréquence, la durée et la force de la trace en mémoire affecte prioritairement les jugements de récence.

1.2.3 Effets inhérents au protocole expérimental

Contexte de présentation

Nous étudions des sons l'environnement extrait de leur contexte d'origine. Snodgrass et Vandervart ont montré, en 1980, cf [Gér04], que des sons isolés peuvent ne pas être identifiés hors contexte à cause de l'incertitude causale.

Il existe, selon Smith et al., [Smi88], deux types de contexte : le contexte significatif et le contexte accidentel. Le contexte significatif est porteur de sens. Il peut s'agir de matériel sémantique ou verbal, comme les instructions fournies par l'expérimentateur. Ce matériel oriente directement les processus de sélection effectués par le sujet. En revanche, le contexte accidentel n'a aucun sens vis à vis des stimuli que se soit implicitement ou explicitement. Il n'influence donc pas les processus de sélection.

Importance de la consigne expérimentale

Tout indice verbal ou non verbal présenté avant le début du test amorce des processus cognitifs. Les indices donnés peuvent constituer un amorçage explicite, "Vous allez entendre des sons de voiture, d'oiseau, ou de foule", ou bien implicite : "Vous allez entendre des sons extraits de l'environnement urbain". Nous devons étudier ce problème avant d'envisager le protocole expérimental. Cette procédure d'amorçage peut faire partie intégrante de l'expérience, dans ce cas nous devons la définir précisément.

Nous indiquerons au sujet dans quel contexte il se situe. Ballas et Mullin prouvent en 1991, [BM91], que l'amorçage linguistique n'altère pas l'effet du contexte. Si nous proposons par un amorçage linguistique un contexte de "Port" et que le contexte évoque un environnement de "Ville", alors l'amorçage n'interférera pas avec l'effet de contexte.

Nous présentons un nombre important de séquences de sons de l'environnement. Le fait de répéter les sons d'une séquence à l'autre n'influencera pas les capacités de reconnaissance dû à un amorçage. En revanche, il ne faut pas exclure un phénomène d'apprentissage.

Identification et rappel

J. Ballas en 1993, [A.B93] montre que la rapidité d'identification est liée à la facilité de représenter un stimulus par une image mentale. La similarité entre le son et le référent prototypique associé en mémoire facilite l'identification. L'auteur montre également que des notions subjectives comme l'agrément ou l'agressivité d'un son altère la rapidité d'identification.

Le contexte dans lequel seront présentées les séquences de sons de l'environnement est donc un élément important. L'amorçage que nous proposerons aux sujets avant de débiter l'expérience sera lui aussi important. Nous devons donner des indices aux sujets, qui orientent leur écoute mais qui ne biaise pas l'expérience. Nous élaborerons le protocole expérimental en intégrant ces deux paramètres. Tous les sons que nous présenterons auront été reconnus. Il n'est pas nécessaire que les sujets soient en mesure de les nommer,

mais au moins qu'ils les associent aux catégories de sons que nous leur proposerons. Nous détaillerons ces points au cours de la présentation de l'expérience de catégorisation.

Les différents travaux conduits sur la perception de séquences sonores mettent en avant des effets de la mémoire sur le jugement global. Les effets de la mémoire que nous observons dans les tâches de rappel, qui interviennent également dans le jugement global sont des processus mnémoniques à court terme. Afin d'étudier plus précisément ces effets, nous allons contrôler l'organisation temporelle des séquences de sons de l'environnement.

Les modèles d'organisation de nos connaissances en mémoire sémantique reposent sur des relations hiérarchiques (modèle de Collins) entre des concepts imbriqués, ou bien sur des processus de comparaison d'attributs (modèle de Smith). Les différences de traitements en fonction des catégories de sons ne sont pas intégrées dans ces modèles.

A présent, nous présentons dans ce qui suit l'expérience proprement dite.

Chapitre 2

Expérience

2.1 Présentation

2.1.1 Rappel des objectifs

Nous souhaitons démontrer que la composition d'une séquence de sons de l'environnement a une influence sur le jugement global. Nous présentons trois catégories de sons. L'organisation temporelle de la séquence est la variable indépendante dans l'expérience. Son effet sur les réponses apportées par les sujets sera évalué à la fois sur le type de réponse fourni, i.e : OUI/NON et dans le cas de réponse positive nous examinerons le temps de réponse.

2.1.2 Protocole

Chaque sujet sera confronté à des séquences de sons de l'environnement urbain. Chaque une de ces séquences dure 1'24. A la fin de chaque séquence, un label cible sera affiché immédiatement sur un écran. Ce label sera appris au cours d'une pré-expérience. Dès que le label s'affiche le sujet répondra par OUI ou NON. Le temps mis pour répondre OUI sera mesuré. Le sujet décide quand déclencher la séquence suivante. Les séquences se succèdent dans un ordre aléatoire, figure 2.1.

Catégorie			Cible
1	2	3	1
1	1	1	1
2	1	3	3
2	2	2	3
...	...	...	...

FIG. 2.1 – Déroulement des séquences au cours de l'expérience de Jugement global

2.1.3 Mesure du temps de réaction

En plus de comptabiliser le nombre de réponses positives, et par complémentarité le nombre de réponses négatives, nous mesurons le temps de réponse dans le cas de réponses positives.

Contexte

Nous présentons une séquence de sons de l'environnement suivi d'une cible. Nous demandons au sujet d'appuyer sur la barre d'espace du clavier après la présentation de la cible. Le sujet peut choisir de ne pas appuyer sur la barre d'espace. Du point de vue de la procédure expérimentale, cela s'appelle un Go/NoGo. Le sujet ne doit pas faire de choix parmi plusieurs propositions.

Conséquences

Tous les temps introduits entre les événements sonores et la cible visuelle doivent être sciemment choisis.

Comme l'expérience dure environ une heure, il est important de maintenir l'attention du sujet au cours de l'expérience. Deux moyens ont été proposés par R.D. Lee, [Lee86]. La première solution est de diffuser un signal bref après la présentation du stimuli, appelé signal de réaction. Ce signal indique au sujet qu'il peut répondre. Ce signal de réaction doit durer pendant 500 ms tout au plus. Dans notre cas, nous substituons ce signal auditif par la cible visuelle. Elle durera 1,5 seconde avant de disparaître de l'écran.

La seconde option pour maintenir l'attention du sujet est de lui fournir un feedback sur ces performances au cours de l'expérience. On peut envisager, dans cette perspective, d'offrir une gratification si les réponses données sont correctes. Ceci présuppose qu'il y est une bonne ou une mauvaise réponse, ce qui n'est pas notre cas ici. Nous fournirons un feedback concernant l'état d'avancement de l'expérience.

Par ailleurs, plus la durée entre la fin de la séquence et l'apparition de la cible est grande plus le temps de réponse est long. Cet intervalle est dénommé foreperiod. Nous présenterons la cible immédiatement après la séquence. Les temps de réaction devront, par conséquent, être les plus courts possibles.

Enfin, le dernier facteur influant le temps de réaction est le niveau sonore et la durée de chaque son dans la séquence.

Mesure effectuée

Comme nous utilisons le paradigme Go/NoGo nous devons tenir compte des fausses alarmes. Celles-ci correspondent aux appuis accidentels ou volontaires mais incorrects sur la barre d'espace. Pour tenir compte de ces fausses alarmes, nous enregistrons les temps d'affichage du stimuli et de réponse du sujet par un chronomètre indépendant. La différence des temps donnés par cette horloge donnera une mesure du temps de réaction.

L'interface informatique que nous utilisons pendant l'expérience introduit un temps de latence dont il faut tenir compte pour mesurer le temps de réaction du sujet, cf

équation 2.1 avec T_R : Temps de réponse et L_r : latence du système. Le temps entre l'appui sur la barre d'espace et la réception du signal d'appui par le système est de 8 ms. Le temps d'intégration des données propre au système est de 26 ms. Cela dépend essentiellement de la taille des données échangées entre les différents composants de l'interface. Nous retirerons 34 ms aux temps affichés dans l'interface.

Le temps de réaction se décompose, selon Donders, cf [ML01], en trois temps : équation (2.2). T_a : Temps de détection, T_b : Temps de discrimination et T_c : Temps moteur. On suppose que le temps moteur est constant pour un individu quelque soit la tâche à effectuer. Cette décomposition du temps de réponse dépend de la tâche à effectuer. Dans notre expérience, le sujet doit choisir si oui ou non le mot évoque la séquence. La décomposition que nous proposons ici s'inscrit dans le cadre de notre expérience.

$$(2.1) \quad T_R^{mesure} = T_R + L_r$$

$$(2.2) \quad T_R = T_a + T_b + T_c$$

2.2 Sélection et description du Matériau sonore

Le choix des sons représente une étape importante. Il ne faut pas introduire de biais en utilisant des sons environnementaux trop hétérogènes en terme de qualité ou de durée.

2.2.1 Extraction d'une base de sons existante

Nous choisissons des sons présents dans un environnement de type environnement urbain. V.Shafiro et B.Gygi proposent dans l'étude mentionnée ici, [SG04], une méthodologie pour choisir correctement les sons utilisés comme stimuli. Ils ont montré que les sons doivent être retenus pour leur pertinence et leur adéquation avec la problématique. Les qualités acoustiques sont à considérer dans un second temps uniquement. Effectivement dans notre cas, l'approche repose essentiellement sur la reconnaissance et non le jugement de qualités acoustiques.

Base de sons

Pour un gain de temps évident, nous recherchons les extraits sonores dans les bases de sons existantes. Il existe différents types de sons référencés dans ce type de base de données. Vis à vis de notre problématique, nous choisirons les sons parmi un inventaire d'extraits de sources isolées et variées. Les sons ont été extraits de la base de sons *SoundIdeas* disponible sur le réseau interne de l'IRCAM. Les extraits sont échantillonnés à 44100 Hz sur 16 bits et mixés en stéréo.

Types de sons extraits

Nous étudions l'effet de la mémoire sémantique sur la perception de séquences de sons environnementaux. L'accès aux représentations en mémoire sémantique ne devra pas être dépendante du type de stimuli. Nous distinguons ici trois types de stimuli : les sources isolées, les sources étendues et les ambiances sonores. Le tableau figurant en Annexe B.1, dresse la liste des sons soumis aux participants lors de l'expérience de catégorisation.

2.2.2 Catégorisation des sources sonores

Pour valider le panel de sons que nous utiliserons dans l'expérience principale, nous effectuons un test de catégorisation orientée. Cette épreuve de catégorisation est importante. Le but est de valider le corpus, et de supprimer toutes les sources qui sont ambiguës. Notons que le groupe de sujets auquel nous soumettons notre panel de sons ne sera pas le groupe avec lequel nous travaillerons sur l'expérience proprement dite.

Catégories proposées

Pour définir les catégories précisément, mentionnons les travaux de V. Maffiolo, [Maf99]. V. Maffiolo démontre qu'un paysage sonore peut être de deux natures différentes. Il peut être amorphe. Dans ce cas, aucune source ou événement n'est discernable dans la scène auditive. La rumeur de trafic dans un environnement urbain peut être considéré comme l'un de ces paysages amorphes. La scène sonore peut également être événementielle. La scène sonore est alors constituée d'événements émergents et aisément identifiables.

Nous souhaitons constituer trois catégories de sons de l'environnement urbain. Nous devons contrôler de manière robuste la dénomination et la définition de ces catégories. V. Maffiolo montre que l'identité d'une scène sonore est définie en terme d'évènements, de bruits et d'activités propres à cette scène. Ainsi nous définirons les catégories de la manière suivante.

Catégorie Trafic Tout événement ou bruit se produisant dans un contexte de circulation en milieu urbain.

Catégorie Parc Tout événement, bruits ou activité humaine se produisant dans un espace vert, parc ou square de ville.

Catégorie Présence humaine Tout événement, bruits ou activité révélant la présence humaine dans un contexte urbain type zone piétonne ou centre ville.

Ces définitions tiennent lieu de postulat et seront évoquées lors de l'expérience de catégorisation.

Expérience de catégorisation

Sujets Les participants à cette expérience sont âgés de 18 à 48 ans . Le panel de sujet est composé de 11 femmes et 21 hommes. Les sujets sont tous normo-entendants.

Stimuli Les sons diffusés sont ceux de la base des sons retenus dans l'expérience de catégorisation. Ces sons ont une durée de 4 secondes. Ils sont échantillonnés sur 16 bits. La fréquence d'échantillonnage est de 44100 Hz.

Dispositif Les sons sont diffusés dans un casque. Ce système de diffusion a été retenu pour les raisons suivantes. La diffusion du son sur un système binaural permet de conserver une image sonore correctement spatialisée. L'immersion perceptive rend la scène sonore réaliste. L'interface est composée d'un écran et d'une souris. Le sujet écoute autant de fois qu'il le souhaite l'extrait sonore . A l'écran figure les trois catégories, à savoir : Trafic, Parc et Présence humaine. Des icônes symbolisant les sons sont disposés en bas de l'écran.

Protocole Dans un premier temps, le sujet écoute l'ensemble des sons en cliquant sur son icône. L'ordre d'écoute n'a aucune importance. Dès lors que le sujet a pris connaissance du corpus, il est invité à regrouper les sons en 4 catégories. Ces quatre catégories sont les 3 mentionnées plus haut, et une quatrième catégorie " Autre", où le sujet placera tous les sons qui n'entrent pas, selon lui, dans les trois autres. Il n'y a aucune contrainte concernant le nombre de représentants à placer par catégorie. Une fois cette organisation en groupe effectuée, le sujet choisi le représentant le plus prototypique de la catégorie.

Résultats

Lecture de la représentation arborée Cette analyse se base sur la représentation arborée, cf figure 2.2. Nous pouvons noter une dissemblance maximum entre les sons associés à "Trafic" et les sons associés à "Parc" et "Présence humaine". La dissemblance entre ces deux dernières catégories est tout de même importante. Ce premier résultat confirme notre sélection des sons et la catégorisation que nous avons faite à priori. Pour s'assurer de la validité de la catégorisation, il faut vérifier que les sons n'ont pas été répartis dans les catégories de manière aléatoire.

Faisons l'hypothèse suivante : "tous les sons ont été catégorisés aléatoirement par l'ensemble des sujets". Pour vérifier cette hypothèse nous allons procéder au "Test du χ^2 ". Ce test permet de comparer la distribution des réponses données par les sujets avec la distribution théorique validant l'hypothèse. L'estimation de la valeur de χ^2 est comparée à une valeur seuil qui dépend à la fois du nombre de variables indépendantes, ici 4 et de l'erreur admissible permettant de valider ou non l'hypothèse. Nous tolérons une erreur de 1%. Dans notre cas cette valeur seuil vaut

$$\chi^2_{\alpha}(0.01) = 11,73$$

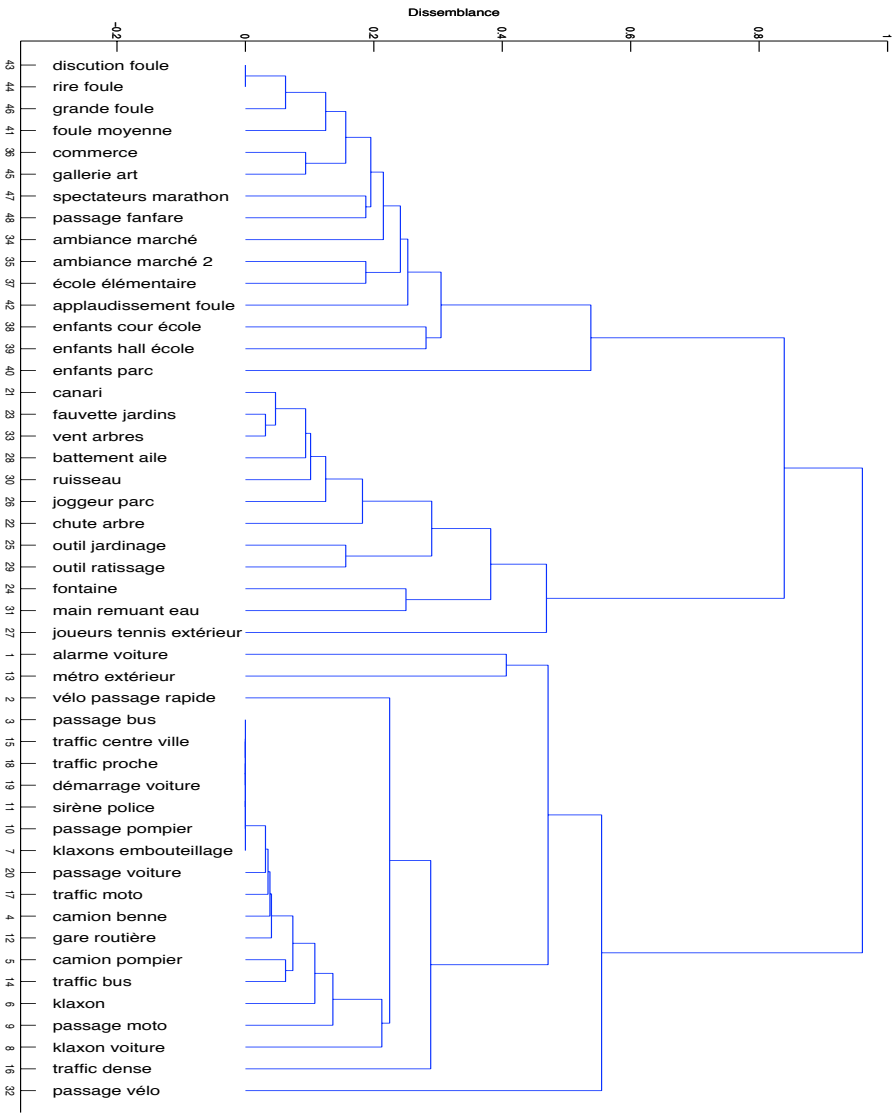


FIG. 2.2 – Représentation arborée de la catégorisation des sons

Selon cette hypothèse nous attendons une répartition équiprobable des sons dans les catégories. Sur un effectif total de 32 sujets, la répartition théorique est donc de 8 : 8 : 8 : 8. Connaissant cette espérance théorique et les données observées nous calculons le χ^2 de la manière suivante.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{O_i}$$

k : nb de catégorie

O_i : données observées par catégorie

E_i : données attendues par catégorie

Si l'estimation du χ^2 fournit une valeur supérieure à cette valeur seuil, alors la distribution observée infirme l'hypothèse d'une répartition aléatoire des réponses.

Tous les sons du panel infirme l'hypothèse, excepté le son "passage de vélo". Celui ci sera donc retirer de la base.

Réduction de la base des sons à l'issue de la catégorisation Dans la mesure du possible, nous conservons les sons qui sont le plus proche en terme de dissemblance. Dans un second temps, nous tacherons de sélectionner des sons diversifiés en terme de sources. Enfin, nous éliminerons du corpus les sons dont les qualités acoustiques sont différentes du reste du corpus. La qualité acoustique, dans notre cas, n'induit pas de difficulté d'identification car elle ne varie pas beaucoup.

Parc	Traffic	Présence Humaine
canari	passage de bus	grande foule
fauvette des jardins	camion benne	école élémentaire
battement d'aile	camion pompier	foule moyenne
ruisseau	klaxon de voiture	applaudissement foule
vent dans arbre	passage de moto	discussion foule
outils de jardinage	gare routière	rire de foule
outils de ratissage	trafic centre ville	spectateurs marathon
joueurs tennis extérieur	passage de voiture	passage parade
fontaine	trafic bus	enfants cour école

FIG. 2.3 – Liste des sons du corpus définitif

Sons prototypiques Chaque sujet a retenu un son représentatif pour chaque catégorie. Une proportion supérieure au hasard a élu comme prototypique les sons : "Passage de pompier" et "Jogeur dans un parc", pour les catégories "Trafic" et "Parc", respectivement. Pour la catégorie "Présence humaine" le choix du prototype n'est pas aussi saillant. 6 sujets ont retenu le son "rire de foule", 5 sujets "ambiance de marché" et 4 "grande foule". Comme il s'agit d'une répartition plus équilibré, nous retiendrons, le son

”Ambiance de marché”. Une différence de 1 voie n’est pas significative, et nous estimons que l’ambiance de marché est représentative de la catégorie ”Présence Humaine” au même titre que le son ”Rire de foule”.

Ces prototypes auront un rôle particulier pendant l’expérience principale. Ils ne seront pas intégrés aux séquences. Il serviront d’exemples sonores pour illustrer les labels cibles lors de l’expérience préliminaire au test principale. Cette expérience d’entraînement est capitale pour la suite. Le tableau indique également le nombre de sujets ayant désigné les sons suivants comme étant le plus représentatif des catégories.

Catégorie	Prototype	Nb de sujets
Traffic	passage de pompier	13
Parc	joggeur dans parc	15
Présence humaine	ambiance de marché / rire de foule	5 / 6

FIG. 2.4 – Prototypes retenus par les sujets

2.2.3 Définition et ajustement du niveau sonore

Pour mesurer le niveau en dB SPL de chaque source, nous nous référons à un son étalon. Dans un premier temps nous mesurons le niveau d’une sinusoïde à 1000Hz. A cette fréquence, il n’y a pas de correction due à la courbe de réponse du filtre auditif. Le niveau mesuré est le niveau absolu. Nous estimons la valeur efficace de ce signal à l’aide d’une routine sous *MATLAB*. Ce couple de valeurs représente l’étalon. Les valeurs efficaces de tous les sons du corpus sont estimées par cette méthode. Par proportionnalité, nous obtenons le niveau en dB SPL de tous les sons comparé au niveau de l’étalon.

Ajustement écologique

Pour contrôler le niveau sonore des sources, nous devons effectuer une expérience d’ajustement en sonie. Cette tâche réalisée par des sujets permet d’établir les niveaux d’intensité des sons de manière stable. Deux ajustements sont possibles. Le premier est un ajustement isosonique. Tous les sons sont ajustés au même niveau de sonie. Cet ajustement n’est pas écologique. Comme nous voulons créer des séquences réalistes, cet ajustement ne convient pas. Nous retenons l’ajustement écologique.

Les sources ont, dans leur contexte, un niveau caractéristique. Nous devons pour cela, décrire aux sujets dans quel contexte ils se trouvent. Nous demandons au sujet d’ajuster le niveau de la source tel qu’il le percevrai en contexte réel. Cela correspond à l’ajustement écologique. Par exemple, un son de marteau piqueur sera nécessairement plus fort qu’un son de canari. Le niveau sera différent si le sujet se situe à proximité ou non de la source. C’est pour cela que nous indiquons au sujet qu’il est situé à proximité de la source.

Sujets Les participants à cette expérience sont âgés de 18 à 48 ans Il y a 21 Hommes et 11 femmes.

Stimuli Les sons diffusés sont ceux de la base des sons retenus dans l'expérience de catégorisation. Ces sons ont une durée de 4 secondes. Ils sont échantillonnés sur 16 bits. La fréquence d'échantillonnage est de 44100 Hz. Le bruit de fond est un bruit de type environnement amorphe urbain. Ces propriétés spectro-temporelle sont présentées en Annexe A.1. Le niveau de ce bruit de fond est de 72 dB SPL.

Dispositif Les sons sont diffusés dans un casque. L'interface est composée d'un écran et d'une souris. Le sujet écoute autant de fois qu'il le souhaite l'extrait sonore couvert partiellement par le bruit de fond urbain. A l'écran figure, une interface d'ajustement, figure 2.5. L'interface est conçue de manière à ce que le niveau sonore ne sature jamais le système de diffusion.

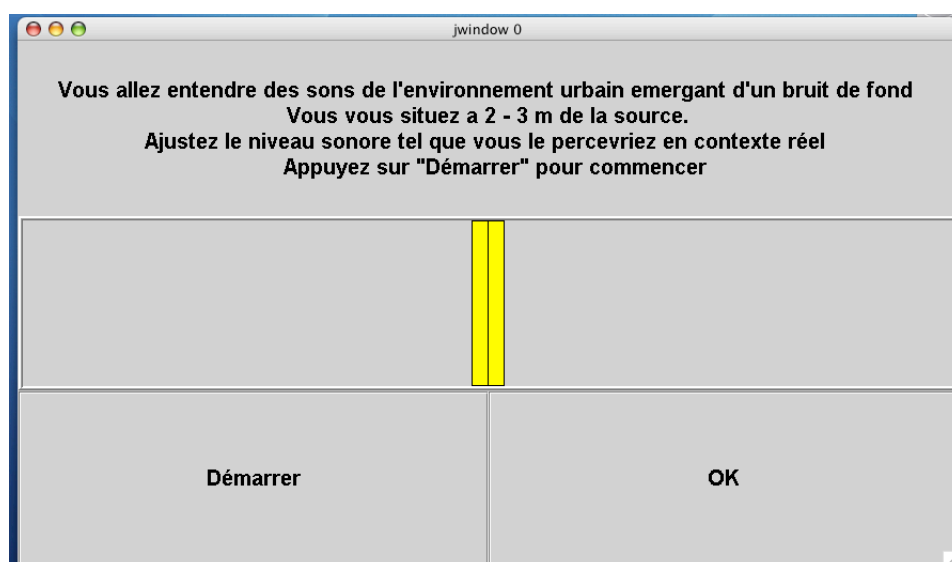


FIG. 2.5 – Interface d'ajustement en Sonie sous PsiExp

Protocole Pour chaque son, le sujet écoute le niveau initial de l'extrait. Il ajuste le niveau à sa convenance, ce qui provoque une nouvelle écoute de l'extrait. La consigne précise au sujet qu'il doit ajuster le niveau sonore tel qu'il percevrait étant situé à une distance approximative de 2 mètres de la source. Le sujet à la possibilité de réécouter l'extrait sans en modifier le volume. Enfin, lorsque le niveau est ajusté, il valide et passe à l'extrait suivant. L'expérience se termine quand tous les extraits ont été ajustés.

Résultats Les niveaux ajustés dans l'interface utilisateur *PsiExp* sont les coefficients multiplicateur de l'amplitude du signal. Le niveau initial en dB SPL est connu pour

chaque son. Ce niveau initial correspond à une valeur de 1 dans l'interface. Pour connaître la correction en dB SPL à apporter au niveau initial du son nous calculons le gain de la manière suivante.

$$N_{dB}^{init} = 20 \log(A)$$

Une valeur de 1 du coefficient ne modifie pas le niveau en dB SPL. Ce coefficient est celui que propose les sujets dans cette expérience d'ajustement en sonie. Ainsi, un coefficient proposé à 1,4, par exemple représentera un gain de 2.95 dB SPL :

$$N_{dB} = 20 \log(1,4 * A)$$

$$N_{dB} = 20 [\log(1,4) + \log(A)]$$

$$N_{dB} = 20 \log(1,4) + N_{dB}^{init}$$

$$N_{dB} = 2,95 + N_{dB}^{init}$$

Les sujets proposent un niveau écologiquement valide pour chacun des sons en comparaison avec le bruit de fond. Une moyenne arithmétique est calculée pour chaque son sur l'ensemble des valeurs fournies par les sujets. La valeur moyenne sera affectée définitivement au son correspondant. Lors de l'expérience finale, le niveau sonore de la source sera celui attribuée lors de cette expérience. Le tableau en Annexe C détaille le niveau réel mesuré en dB SPL puis le niveau écologique de chaque son.

Observations Tous les sons de la catégorie " Parc" sont en dessous de 70 dB SPL, mis à part les deux sons d'oiseaux et le son de la "main remuant l'eau". Les sons d'oiseaux sont ajustés plus fort, car ils émergent de la scène sonore. Les sons que nous retiendrons auront un niveau écologique proche de 68 dB SPL. En terme d'intensité subjective la catégorie "Parc" sera celle ayant le niveau le plus faible.

Les sons de la catégorie "Trafic" ont tous été augmentés de 4 à 5 dB SPL. Nous avons retenus des sons dont le niveau varie entre 75 dB SPL et 85 dB SPL. Du point de vue intensité subjective, les sons de la catégorie "Trafic" sont, cette fois ci les plus intenses.

Les sons de la catégorie "Présence humaine" sont eux aussi augmentés de 4 à 5 dB SPL. Le niveau moyen est compris entre 70 et 80 dB SPL.

Une fois les expériences préliminaires réalisées nous passons à la description de l'expérience principale. Nous présenterons, le matériel et les stimuli utilisés. Nous détaillerons également la procédure expérimentale.

2.3 Matériel

Cette section fournit les détails techniques du matériel utilisé pour réaliser les expériences.

Poste de travail Le poste de travail est situé dans une cabine insonorisée IAC. Le niveau du bruit de fond est de 63 dB SPL.

Station de travail Les stations de travail sont des *Mac G5* biprocesseurs équipés de cartes audio interne *RME hdsp9652*.

L'interface utilisateur Les deux logiciels principaux sont *Max/Msp* et *Psiexp*. *Max/Msp* logiciel de traitement de flux de données en temps réel, nous sert comme interface de lecture des fichiers sons principalement. le second logiciel, *Psiexp*, permet de concevoir les interfaces utilisateurs. Ce logiciel a été développé par B. Smith à l'IRCAM.

Système de diffusion Les sorties de la carte audio sont connectées à un ampli *Lake* dédié pour une diffusion dans un casque *Sennheiser HD II*.

2.4 Stimuli

Nous détaillons dans cette section les stimuli utilisés dans notre expérience.

2.4.1 Sons utilisés

Les sons utilisés sont référencés dans la section 2.2.1. Le son dure 5 secondes. Lors de sa diffusion, le son émerge progressivement du bruit de fond par une rampe en intensité qui dure 2 secondes. Ensuite, le son est stable pendant 1 seconde, puis décroît selon une rampe de 2 secondes également.

2.4.2 Organisation temporelle des séquences

Les séquences "types" sont composées de 12 sons, figure fig :seq. Trois groupes de 4 sons se succèdent. Chaque groupe est constitué de sons appartenant à la même catégorie, cf section 2.2.2. L'ordre des catégories est déterminés aléatoirement par Max/Msp. Cela représente 6 permutations possibles des catégories. L'ordre de diffusion des sons dans un groupe catégoriel est aléatoire. L'organisation temporelle de la séquence est sauvegardée dans un fichier texte.

Les séquences "tests" sont composées de 9 sons d'une même catégorie, 2.7.

Chaque son dure 4 secondes. Entre chaque son, il y a une pause de 2 secondes. Ces valeurs ont été ajustées pour préserver un rythme de présentation correct. Ces ajustements nous permettront de tester toute la combinatoire possible des séquences sans prolonger démesurément le temps de l'expérience. Le rythme de présentation des sons est réaliste.

2.4.3 Mots cibles

Nous proposons aux sujets 4 mots cible. Ces mots sont les noms des catégories que nous avons choisi, à savoir : Trafic, Parc et Présence humaine. Une quatrième cible est introduite. Nous choisissons le label "Port". Il a une fonction distractive. Le sujet

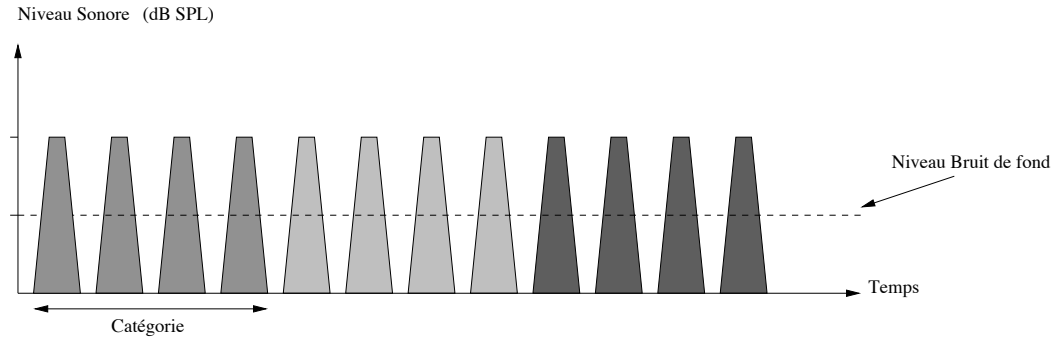


FIG. 2.6 – Structure temporelle d’une séquence ”type”

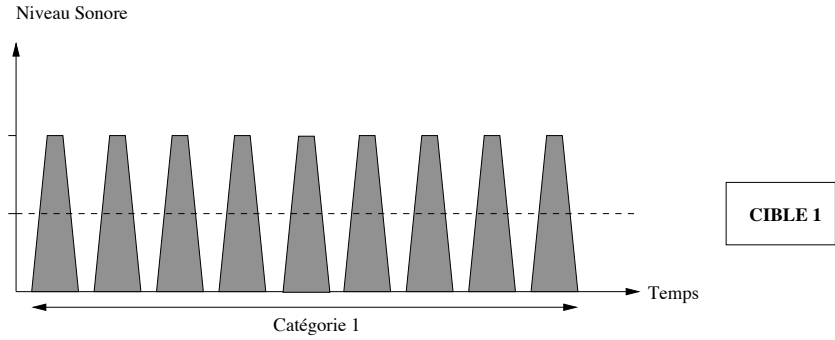


FIG. 2.7 – Structure temporelle d’une séquence ”test”

devra rester attentif à la cible sachant qu'il peut, s'il n'y prête pas assez d'attention, répondre de manière erronée. Combiné aux 6 permutations possibles, cela représente 24 séquences différentes. Nous introduisons également des séquences test. Ces séquences sont composées uniquement de sons d'une même catégorie. Pour les trois séquences test nous présentons le label de la catégorie correspondante et un label autre. Cela représente donc 6 nouvelles configurations. Cela représente donc un ensemble de 30 séquences.

2.4.4 Bruit de fond

Le bruit de fond est celui décrit en annexe A. Son niveau sonore est de 72 dB SPL. Il débute par une rampe progressive en amplitude sur 200 ms. Il dure le temps de la diffusion des 12 sons. Enfin, il décroît progressivement sur une durée de 2 secondes. Les rampes en amplitude sont utilisées pour éviter les clic. La rampe de sortie est plus longue, car nous percevons plus rapidement une diminution du niveau sonore. Pour que les rampes d'entrée et de sortie est une durée subjective identique nous avons retenu ces valeurs de manière empirique.

2.5 Procédure

L'expérience se déroule en deux temps. Tout d'abord, le sujet est invité à manipuler une interface présentant tous les couples [son prototypique - label associé]. Cela va permettre au sujet d'être plus efficace pendant l'expérience principale. Le sujet prend connaissance de tous les labels auxquels il sera confronté plus tard. L'interface à l'écran est composé de 4 boutons. En cliquant sur chacun d'eux le son prototypique associé est joué. Cette étude dure 1 à 2 minutes.

L'expérience en elle même peut débiter. Le sujet reçoit au début de l'expérience un résumé des consignes énoncées par l'expérimentateur. A la suite de cette présentation la première séquence débute. A la fin de celle-ci le label cible est affiché pendant 1 seconde. A partir de cet instant, le sujet est invité à répondre dans les 3 secondes. Un compteur affiche le décompte. Ceci est fait pour conditionner le sujet. Le sujet appui sur la barre espace s'il estime que le label évoque correctement la séquence. Comme sa réponse doit être la plus rapide possible, le jugement qu'il porte sur la séquence est basé sur ses souvenirs à court terme. Si le sujet estime que le label n'évoque pas correctement la séquence il n'appuie pas sur la barre. Son appui est pris en compte et indiqué par un témoin qui change de couleur à l'écran. Chaque appui sur la barre est stocké dans un fichier de résultats. A la fin de cette première phase, le sujet décide quand débiter la séquence suivante.

Ce dernier est invité à faire une pause après les 10 ou 15 premières séquences.

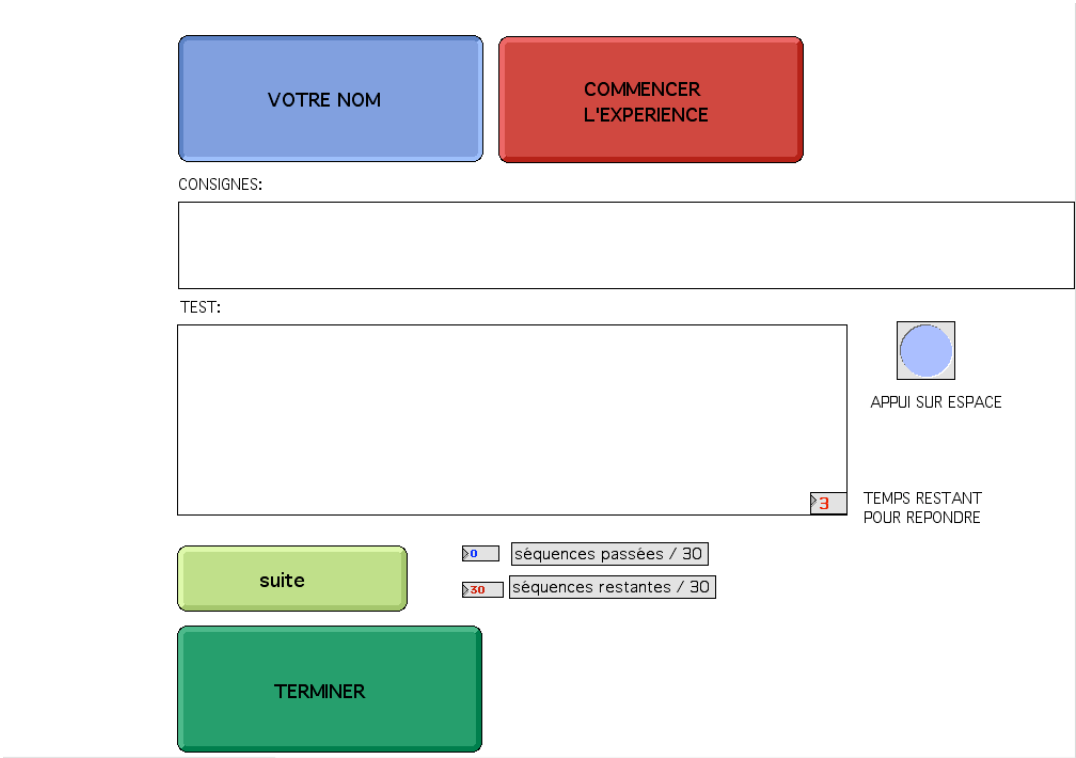


FIG. 2.8 – Interface Utilisateur pour l'expérience de Jugement Global sous Max/MSP

2.6 Sujets

30 sujets ont participé à l'expérience de jugement global. Tous sont normo entendants. Les sujets sont âgés de 18 à 47 ans. Le panel est composé de 6 femmes et 24 hommes.

2.7 Résultats

Nous rapportons dans cette section les résultats de l'expérience de jugement global. Nous présentons les résultats obtenus lorsque les sujets ont été confrontés aux séquences "test". Ensuite, nous observons l'évolution du taux de réponse et du temps de réponse en fonction de la position temporelle de la catégorie cible. Puis, nous mettons en avant les différences entre les catégories. Enfin, nous présentons l'évolution du taux et du temps de réponse en fonction de la configuration temporelle des séquences pour la catégorie Trafic.

2.7.1 Résultats concernant les séquences Test

Nous avons présenté aux sujets des séquences composées de 9 sons appartenant à une catégorie unique. Chacune des 3 séquences est suivie du label de la catégorie et d'un label différent. Nous les appellerons, respectivement, séquence "Test" et séquence "Piège".

Données

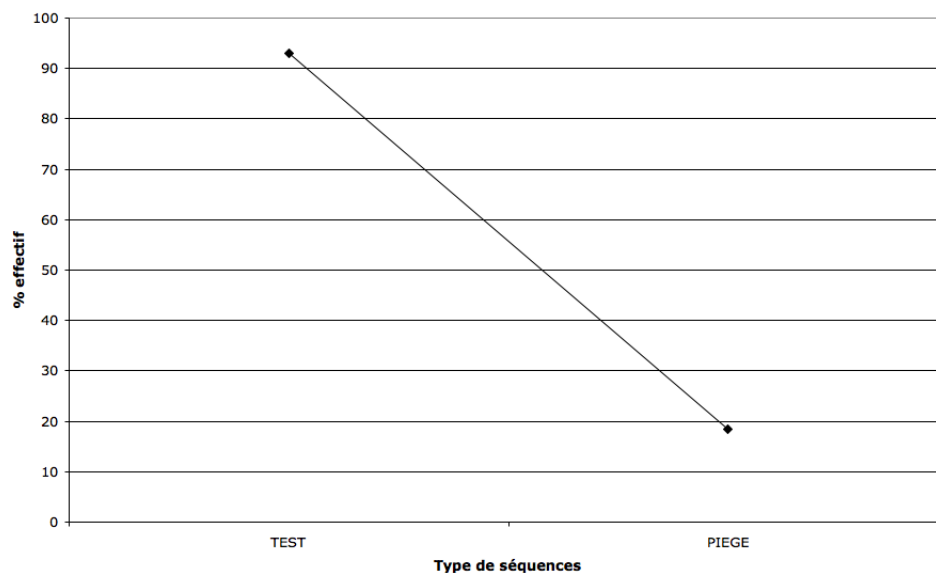


FIG. 2.9 – Pourcentage de réponse en fonction du type de séquence

Figure 2.9 93% des sujets ont répondu dans le cas de séquence "Test". 18% des sujets ont répondu dans le cas de séquences "Piège".

Temps de réponse moyen Le temps de réponse pour la catégorie Trafic est de 689 ms. Le temps de réponse pour la catégorie Parc est de 666 ms. Le temps de réponse pour la catégorie Présence humaine est de 885 ms.

2.7.2 Pourcentage de réponse pour chaque catégorie en fonction de la position temporelle

La question posée est "Est-ce que le mot évoque correctement la séquence?". Pour chaque catégorie nous calculons le nombre de sujets ayant répondu oui. L'effectif et le temps de réponse sont calculés pour chaque position de la catégorie dans la séquence.

Données

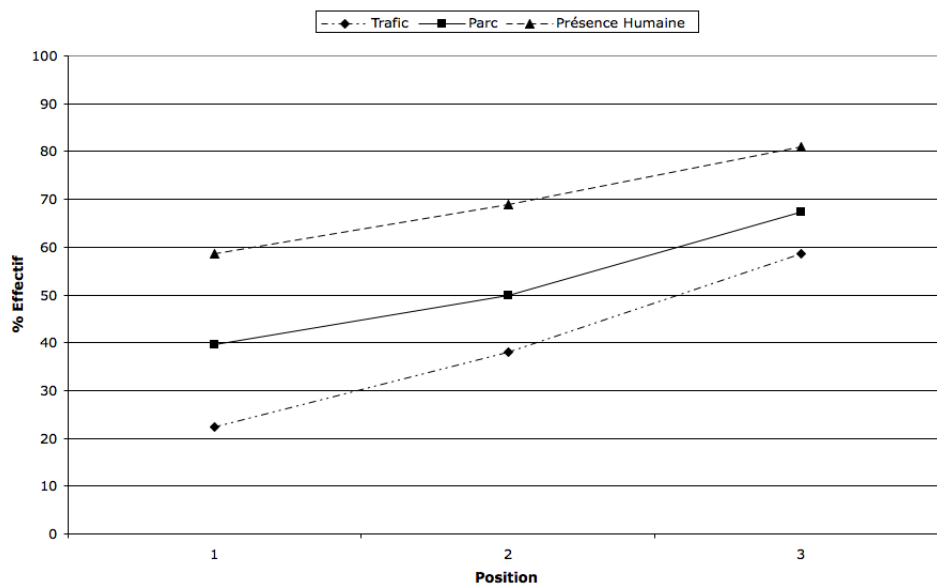


FIG. 2.10 – Effectif pour chaque catégorie en fonction de la position de la cible

Figure 2.10 22% des sujets ont répondu lorsque la cible Trafic est en position initiale. 38% des sujets ont répondu lorsque la cible est en position intermédiaire et 59% des sujets lorsque la cible est en position finale. Le pourcentage d'effectif cumulé pour la catégorie Trafic est de 40%.

40% des sujets ont répondu lorsque la cible Parc est en position initiale. 50% sujets ont répondu lorsque la cible est en position intermédiaire et 67% sujets lorsque la cible est en position finale. Le pourcentage d'effectif cumulé pour la catégorie Parc est de 52 % .

59% des sujets ont répondu lorsque la cible Présence Humaine est en position initiale. 69% des sujets ont répondu lorsque la cible est en position intermédiaire et 81% des sujets lorsque la cible est en position finale. Le pourcentage d'effectif cumulé pour la catégorie Présence Humaine est de 70%.

2.7.3 Temps de réponse pour chaque catégorie cible

Les mesures précédentes suggèrent une différence de temps de réponse moyen quelque soit la position de la catégorie dans la séquence.

Données

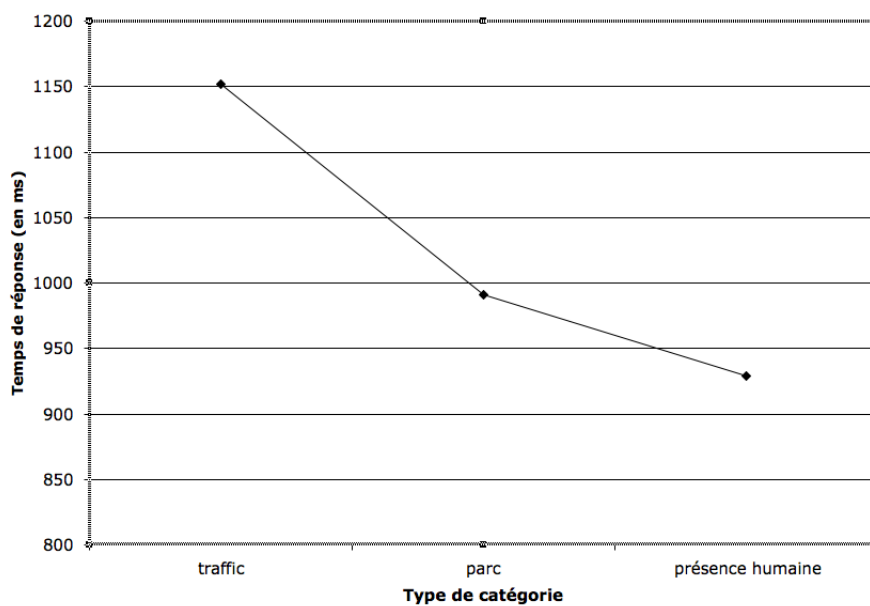


FIG. 2.11 – Temps de réponse en fonction de la catégorie cible

Figure 2.11 Le temps de réponse pour la catégorie Trafic est de 1150 ms. Le temps de réponse pour la catégorie Parc est de 990 ms. Le temps moyen pour la catégorie Présence humaine est de 930 ms.

2.7.4 Pourcentage et Temps de réponse pour la catégorie Traffic

Nous étudions dans cette section, l'évolution de l'effectif et du temps de réponse pour la catégorie Traffic en fonction de la configuration temporelle des séquences.

Données

Chaque configuration temporelle de la séquence est représentée par une série de chiffres. La configuration [1 2 3] signifie Trafic/Parc/Présence humaine. Les figures présentent toutes les configurations ordonnées par position de la catégorie cible dans la séquence. Dans les deux premières configurations, la cible est en position 1, dans les deux suivantes, en position 2, et enfin la cible est en position 3. Ces figures fournissent des données complémentaires à celles rapportées dans les figures de la section 2.7.2.

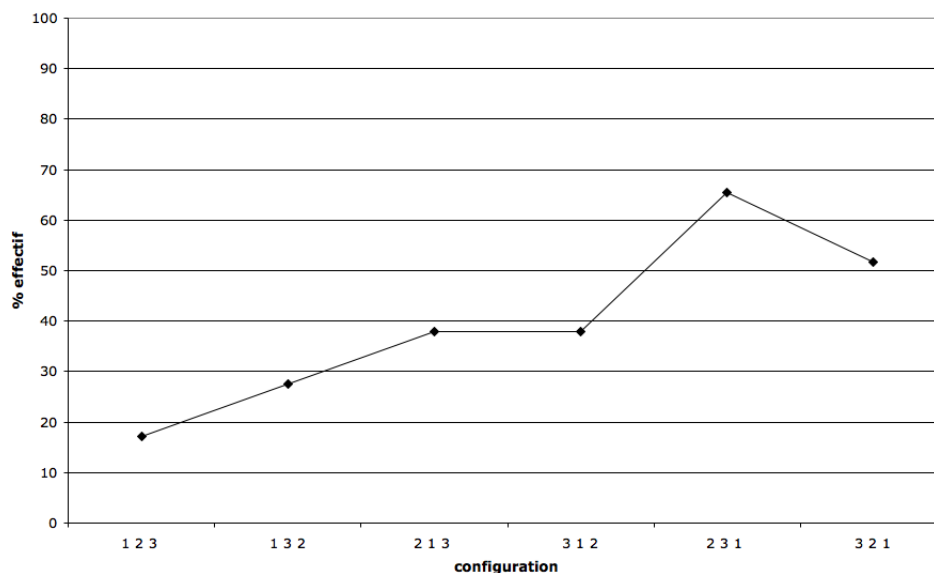


FIG. 2.12 – % de l'effectif pour la catégorie Traffic en fonction des configurations

Figure 2.12 En position initiale, le score est de 17% pour la configuration [1 2 3], et de 28% pour la configuration [1 3 2]. Pour la cible en position intermédiaire, il n'y a pas de différence. En position finale, le score est de 66% pour la configuration [2 3 1], et de 52% pour la configuration [3 2 1].

Figure 2.13 Pour la cible placée en début de séquence, le temps de réponse est de 1624 ms pour la configuration [1 2 3] et de 1272 ms pour la configuration [1 3 2]. Pour une cible en position intermédiaire, le temps de réponse est de 935 ms pour la configuration [2 1 3]

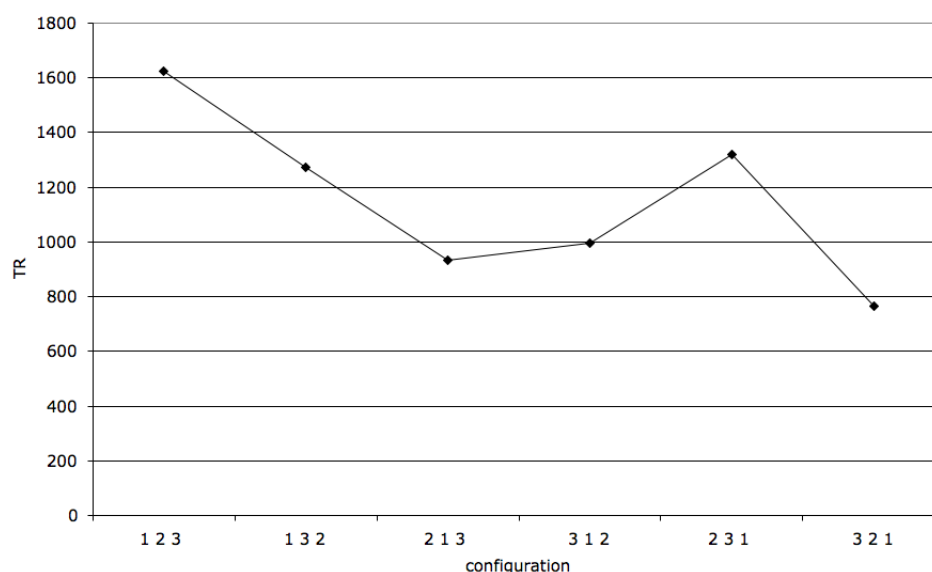


FIG. 2.13 – Temps de réponse pour la catégorie Trafic en fonction des configurations

3] et de 996 ms pour la configuration [3 1 2]. Pour la cible en position finale, le temps de réponse est de 1318 ms pour la configuration [2 3 1], et de 764 ms pour la configuration [3 2 1].

Ces résultats ne seront pas analysés dans le détail, car nous ne disposons pas de données suffisantes pour conclure. Il apparaît cependant un autre effet de l'organisation temporelle. Les catégories situées avant et/ou après la catégorie cible influence le jugement des sujets.

2.7.5 Discussion des résultats

Nous examinons les données recueillies dans les sections précédentes. Nous présentons les ressemblances et dissemblances qui apparaissent en comparant à la fois les taux de réponse mais aussi les temps de réponse. Les temps de réponse observés varient entre 700 et 1500 ms.

Temps minimum de décision

Lorsque les séquences sont composées de sons appartenant à la même catégorie 93% des sujets estiment que le mot évoque la séquence. 19% des sujets estiment que le mot évoque la séquence alors que celui-ci n'est, en principe, pas représentatif de la séquence (cf section 2.7.1). Il peut également s'agir de fausses alarmes.

Les temps de réponse mesurés pour les séquences "Test" sont compris entre 690 et 890 ms selon la catégorie. L'intégralité de la séquence évoque le label présenté à la fin. Si l'on admet que le temps de détection pour un stimulus visuel est de 120 ms, selon **ref** alors le temps de choix + le temps moteur, tel que le définit R.D Lee (cf [Lee86]) est compris entre 570 et 770 ms.

Evolution du taux de réponse avec la récence de la cible

Pour chaque catégorie, le pourcentage de réponse augmente avec la récence de la cible. Cela est dû, en premier lieu, au fait que les derniers événements entendus sont favorisés car récents en mémoire à court terme.

En comparant les pourcentages moyens pour chaque catégorie, section 2.7.2, nous observons 30% de réponse pour la catégorie Trafic contre 70% pour Présence humaine. Plus de sujets répondent lorsque nous demandons si la cible "Présence humaine" évoque correctement la séquence. Les sons humains ont un poids plus important dans le jugement que portent les sujets sur la séquence.

Pour la catégorie Trafic, le taux de réponse est de 20% pour la position initiale. Les sujets estiment que le label Trafic n'est pas représentatif de la séquence. Lorsque la cible est en position intermédiaire, le taux de réponse reste en faveur du non avec seulement 40 % de oui. Si la cible est en position finale, le label recueille 60% de réponse. Ce qui est peu comparativement aux autres catégories. Nous notons une progression de 40% du taux de réponse en fonction de la position, ce qui est important.

Pour la catégorie Présence humaine, les résultats sont différents. Pour une cible en position initiale, le score est déjà en faveur du oui avec 60 % de réponse. Même si la cible est ancienne, plus de la moitié des sujets estime le label Présence Humaine est représentatif de la séquence. Les sons humains laissent une empreinte plus importante que les sons de Trafic. En position finale, le taux de réponse est de 80%. Comparé au 60% que recueille le label Trafic pour la même position, le label Présence humaine est plus souvent associé à la séquence quelque soit le contenu de celle-ci. La progression du taux de réponse est de 20%. Une fois encore, les sons humains se révèlent plus influant.

Pour la catégorie Parc, nous pouvons noter un résultat intéressant. Lorsque la cible est en position intermédiaire, le score est de 50 %. A la lumière de ce que nous avons dit plus haut, cela s'explique. La cible est nécessairement suivi de la catégorie Trafic ou Présence humaine. Nous avons vu que les scores pour ces deux catégories sont plus élevés.

Différences de jugement selon les catégories

Le jugement global porté sur chaque catégorie est différent selon la cible présentée. La figure présentée en section 2.7.3 illustre ce point de manière synthétique. Le temps moyen pour répondre lorsque que la cible est la catégorie trafic est de 1150 ms. Lorsqu'il s'agit de la catégorie Parc, le temps est de 990 ms. Enfin, lorsque la catégorie est Présence humaine, le temps est de 930 ms.

Synthèse des résultats

Nous ne pouvons affirmer que les différences observées entre les temps de réponse soient significatives. Nous constatons qu'il existe des différences entre les catégories. Expliquer les raisons pour lesquelles il existe ces différences et en connaître les causes est une démarche encore prématurée. Cependant, ces résultats expérimentaux fournissent des données capitales concernant l'étude de ce type de stimuli. L'étude du jugement global doit tenir compte de bien plus de paramètres que l'organisation temporelle.

Pour résumé, notons les résultats les plus marquants :

1. Le jugement global d'une séquence porte sur l'ensemble des événements perçus.
2. La récence de la cible a un effet sur le taux et le temps de réponse.
3. Le temps minimum de réponse pour juger du lien entre un label et une séquence est compris entre 700 et 900 ms dans le cas de séquences homogènes, et compris entre 820 et 1450 ms dans le cas de séquences hétérogènes.
4. Le taux de réponse pour la catégorie cible Présence humaine est plus élevé que pour les catégories Parc et Trafic. Le temps de réponse est, en revanche, plus court.
5. Les temps de réponse sont différents, quelque soit la catégorie, entre une cible placée au début de la séquence et une cible placée au milieu ou à la fin.

2.8 Discussion générale

Dans cette discussion, nous aborderons le point central de notre étude ; l'effet de l'organisation temporelle sur la perception globale d'une séquence sonore composée de sons de l'environnement urbain. Nous discuterons de la présence de l'effet de récence et de l'absence de l'effet de primauté. Puis, nous aborderons la question du jugement global. Dans un second temps, nous présenterons les raisons probables à l'origine des différences observées entre les catégories.

2.8.1 Effet de la mémoire

Effet de récence

La récence de la catégorie cible influence le temps de réponse. Il apparaît donc que le jugement d'une séquence de sons de l'environnement soit également propice à l'effet de récence. Notre expérience diffère de celle menée par P. Guelton en 2005, [Gue05]. L'effet de récence est avéré si l'on observe le taux de réponse en fonction de la position de la cible dans la séquence. Cet effet de récence doit être étudié pour chacune des catégories présentées dans notre protocole.

Effet de primauté

Nous n'observons pas d'effet de primauté. Ni le taux de réponse, ni le temps de réponse semblent être améliorés lorsque la catégorie cible est en position initiale. L'étude rapportée ici diffère des travaux de P. Guelton. Il n'est pas explicitement demandé aux auditeurs de se rappeler de la séquence. Les processus de boucle phonologique ne sont probablement pas déclenchés. Il serait intéressant de reproduire l'expérience de P. Guelton sans indiquer explicitement aux auditeurs qu'ils doivent se rappeler de la liste. Ce protocole serait alors similaire au notre dans la tâche demandée aux sujets.

Jugement global de séquences de sons

Le jugement global dépend de l'organisation temporelle des séquences. Susini et al., [SMS02], ont montré que le jugement global du niveau sonore dépend de l'organisation temporelle des séquences. Les stimuli utilisés dans leur étude sont des sons purs. Hellbrück et al., [HKZ⁺01], ont reproduit ces résultats pour des séquences de sons de l'environnement. Nous poursuivons dans cette voie, et démontrons dans notre étude que l'organisation temporelle a également une influence sur des jugements concernant d'autres paramètres que des paramètres acoustiques. Les effets de la mémoire se retrouvent sur différents plans. Nous avons orienté les sujets de notre expérience vers un jugement du contenu sémantique. Le jugement global reste conditionné par la consigne que reçoit le sujet. Comme le suggère V. Maffiolo, [MCD98], les séquences "événementielles" sont analysées sémantiquement alors que des séquences amorphes le sont qualitativement.

Organisation de la mémoire sémantique

Plus que la récence, c'est l'organisation de notre mémoire sémantique qui est responsable de la variabilité du temps de réponse en fonction de la catégorie cible, et en second de la position de celle-ci dans la séquence. La mémoire sémantique est organisée selon Collins et al. en un arbre hiérarchique de concepts. Les concepts partagent des attributs. Cette organisation ne rend pas compte des effets de typicalité. Comme nous le suggérons plus loin, l'effet de typicalité du prototype présenté en pré expérience à un effet sur le jugement sémantique global. Ceci constitue un argument supplémentaire en défaveur du modèle de Collins. Le modèle par comparaison d'attribut n'est pas plus intéressant. La comparaison d'attributs ce fait par paire. Ce modèle ne tient pas compte de différences du point de vue du temps de traitement des sons. Peut être que la comparaison d'attributs pour des sons de Présence humaine s'effectue plus rapidement que pour les autres sons. Ceci ne reste qu'une hypothèse. Nous pourrions envisager une expérience. En proposant plusieurs prototypes de sons humains et non humains, nous observerons les temps mis pour dire si des sons test sont similaires ou différents des prototypes. Aucun des modèles que nous avons mentionné plus haut, n'intègre cette distinction entre les différentes catégories de sons. Le seul qui suggère une distinction, ne propose qu'une séparation entre matériel verbal et non verbal. Il s'agit du modèle d'organisation fonctionnelle proposé par Barlett. Notre étude suggère une autre distinction sur le plan fonctionnel justement. Il ne faut plus distinguer le matériel verbal du matériel non verbal, mais distinguer tout ce qui attire à l'humain et tout le reste.

2.8.2 Sons de l'environnement

Sons Humains / Sons Non Humains

Les travaux de V. Maffiolo, [Maf99] montrent que les auditeurs décrivent les ambiances sonores urbaines selon trois catégories d'événements. La première distinction se situe entre les sons humains et les sons non humains. Puis, parmi les sons non humains, les sujets distinguent les sons naturels des sons mécaniques. Les résultats des travaux présentés dans ce rapport montrent que le traitement des séquences sonores en terme de temps de réaction dépend de la catégorie de sons considérée : sons humains, sons naturels ou sons manufacturés. Une explication possible est que les sons humains sont séparés des sons non humains dans l'organisation des représentations en mémoire comme le suggère Bower et Holyoak, [BH73].

Reconnaissance des sons de l'environnement

J.A. Ballas affirme, en 1993, [A.B93], que le temps d'identification est fonction de trois paramètres qui, dans notre cas, peuvent s'avérer variable selon la catégorie de sons. La variabilité inter sujet observée pour les mesures de temps de réponse est probablement liée à l'un de ces paramètres. Le plus marquant est la typicalité. Si parmi les séquences présentées un exemplaire ou plusieurs ne sont pas pour un sujet donné, typique de la catégorie cible, alors le jugement porté sur la séquence peut varier d'un sujet à l'autre.

Nous avons, cependant mené une expérience de catégorisation forcée. Seulement cette expérience à été réalisée auprès de 30 sujets qui n'ont pas participé à l'expérience de jugement global. Il sera donc intéressant, d'évaluer la typicalité des sources avant de poursuivre l'étude sur le jugement de séquences. L'incertitude causale est peut être responsable de cette variabilité dans les temps de réponse. Elle l'est au moins pour le taux de réponse non nul pour les séquences "Piège". En effet, le son prototypique proposé pour la catégorie "Port" comportait un bruit de moteur de fort niveau mêlé à un bruit d'eau remuant sous l'effet de la propulsion. Ce son prototypique a pu être associé, par certains sujets, à la catégorie cible Trafic, en raison de ce bruit de moteur. L'incertitude causale doit être testée pour l'ensemble des sons du corpus. Le troisième paramètre, à savoir le contexte, s'est avéré pertinent d'un point de vue subjectif. J.A. Ballas suggère également que l'agrément et l'agressivité des sons tient un rôle important dans le processus d'identification, mais surtout dans le temps de rappel. Cet élément permet de comprendre pourquoi, lorsque la catégorie cible est située au début de la séquence le temps de réponse est variable.

Le dernier point concernant la reconnaissance des sons de l'environnement est l'effet de contexte temporel ou sémantique évoqué par K.A. Klein en 2004,[KCS04]. Les différences de temps de réponse pour la catégorie Trafic suggère un effet du contexte sémantique sur le temps de réponse. Il sera intéressant d'étudier ce type d'effet pour évaluer l'importance du contexte temporel ou sémantique autour de la catégorie cible et de mesurer son impact sur le jugement global d'une séquence de sons de l'environnement.

Etiquetage sémantique

L'étiquetage sémantique des sons prototypiques lors de la pré expérience s'est avéré nécessaire pour améliorer les performances des sujets alors qu'ils ne connaissaient pas a priori les sons auxquels ils allaient être confronté. Cet étiquetage linguistique comme le propose J. Barlett en 1977,[Bar77], conditionne les performances de reconnaissance et d'identification.

Ambiguïté des catégories Il se peut que les sons de Trafic où de Parc évoquent une présence humaine indirecte. Un son comme un son de klaxon peut être interprété comme étant une conséquence d'une présence humaine. Pour estimer cet effet nous pouvons regarder le pourcentage de sujets ayant répondu Oui lorsque la séquence est uniquement composée de son de Parc, et que le label Présence humaine a été présenté. 20% des sujets répondent dans ces conditions. Ce pourcentage est non négligeable. La raison la plus plausible est l'ambiguïté sémantique entre les définitions des catégories et les sons. Pour résoudre ce problème nous referons l'expérience avec un apprentissage des sons du corpus. Nous observerons, dans ces conditions si les différences entre les temps de réponse sont conservées.

Distances par rapport aux prototypes Les sujets jugent les séquences sur la base des exemples fournis lors de la pré expérience. Les prototypes retenus évoquent un ensemble de sons que les sujets peuvent associer à l'étiquette. Le sujet peut associer un

ensemble plus ou moins large de sons selon le prototype et sa typicalité. Si le son est un son de bruit d'eau, le sujet pourra associer à Parc des sons naturels très variés. Si le prototype est un bruit de tourniquet, par exemple, le sujet associera que des sons qu'il est susceptible d'entendre dans un parc de ville. Le prototype doit être choisi précisément. Nous pouvons à la suite d'une expérience de catégorisation comme nous l'avons fait, évaluer la typicalité des prototypes.

Chapitre 3

Conclusion

Nous avons étudié l'influence de l'organisation temporelle de séquences de sons de l'environnement sur le jugement global. Nous souhaitons mettre en évidence les effets liés à la mémoire auditive à court terme et à l'organisation des séquences en catégories. En conduisant une expérience de jugement global de séquences de sons de l'environnement organisées en catégories de sons, nous observons les résultats suivants.

Le jugement global dépend de la position de la catégorie cible. Plus la catégorie cible est récente plus elle influence le jugement global. Ce résultat est observé, quelque soit la catégorie, sur le taux de réponse. Cependant, nous constatons que quelque soit la position de la catégorie cible, le temps de réponse varie selon la catégorie considérée. Il apparaît une nette amélioration du temps de réponse pour les sons de présence humaine sur les autres sons. De plus, nous observons des différences de temps de réponse entre chaque catégorie. Nous n'observons pas d'effet de primauté. La consigne donnée aux sujets en est probablement la cause. Il n'est pas demandé aux sujets de se rappeler de tous les sons de la séquence. Par conséquent, ils ne déclenchent pas de processus de boucle phonologique. Aucun des modèles de mémoire sémantique n'a été retenu pour expliquer les phénomènes que nous observons dans notre expérience.

Il apparaît judicieux de distinguer différentes catégories de sons de l'environnement. Parmi les catégories de sons que nous proposons, certaines sont plus sensibles au positionnement dans la séquence que les autres. Les sons de Trafic sont ceux pour lesquels le jugement global est le plus sensible à la position de la catégorie cible dans la séquence. Au contraire, le jugement global pour les sons de présence humaine est le moins influencé par la position de la catégorie.

La méthode, proposée dans notre étude permettant de rendre compte de l'impression globale d'une séquence, nous a permis de mettre en évidence les différences de jugement en fonction de la catégorie de sons. La méthode en elle-même peut être appliquée pour étudier des séquences composées de différentes catégories, sous réserve de faire apprendre le corpus des sons avant de mener l'expérience.

Chapitre 4

Perspectives

Les perspectives que nous envisageons ici, sont en premier lieu des perspectives à court terme qui nous permettront de confirmer les hypothèses que nous avons émises dans la discussion générale, section 2.8. Ensuite, nous proposons de poursuivre les expériences de P.Guelton. Enfin, nous proposons d'étudier d'autres paramètres susceptibles de varier dans une séquence de sons comme le niveau sonore.

4.1 Expérience de jugement global de séquences

Nous étudierons l'effet de l'apprentissage du corpus sur l'association du label à la séquence. Nous pourrions présenter l'ensemble des sons du corpus lors d'une expérience préliminaire afin de contrôler les paramètres tel que l'incertitude causale, la typicalité ou encore l'étiquetage sémantique. Cet apprentissage préliminaire permettra de lever l'ambiguïté des étiquettes et des sons présentés.

Il sera intéressant de demander aux sujets d'évaluer les séquences de sons du point de vue de l'agrément ou du désagrément suscité. Comme le contenu sémantique sera contrôlé, nous pourrions construire un nouveau corpus pour chaque catégorie en distinguant parmi elles des sous-catégories. Cette piste reste ouverte.

4.2 Expériences de rappel de listes de sons environnementaux

L'étude que nous avons menée ici démontre qu'il est important de séparer chaque catégorie pour étudier les phénomènes de récence et de primauté pour les sons environnementaux. L'étude conduite par P.Guelton portait sur des listes hétérogènes de sons de l'environnement. Nous pourrions étendre l'étude des phénomènes de récence et primauté à des listes de sons respectant la séparation proposée, à savoir, sons humains, sons non humains. Nous pourrions évaluer le taux de rappel pour chaque catégorie d'une cible en fonction de la position dans la liste.

Dans un second temps, nous pourrions évaluer le jugement de récence d'un élément d'une liste homogène et mesurer le temps de réponse. Ceci en reprenant le paradigme

de Sternberg. Au lieu de présenter un label, nous présenterons un des sons appartenant à la séquence. Nous demanderons si ce son appartient ou non à la séquence. Cette mesure sera davantage contrôlée par rapport au protocole que nous avons proposé, et permettra d'observer d'éventuelles différences dans le traitement des stimuli humains et non humains. En étudiant des listes de sons présentées comme dans l'étude de P.Guelton nous pourrions mesurer l'effet d'un apprentissage des sons du corpus sur le jugement de récence, ou bien estimer l'effet de l'amorçage sémantique sur le taux de rappel.

Toujours en reprenant les stimuli utilisés par P. Guelton, nous pourrions comparer les résultats de son étude avec les résultats que nous obtiendrons si nous ne demandons pas explicitement aux sujets de se rappeler de la liste des sons. Nous pourrions comparer les résultats dans ces conditions avec les résultats que nous obtenons dans notre étude.

4.3 Expérience de jugement pour chaque catégorie

Enfin, comme le contenu des séquences sera regroupé par catégorie, nous pourrions évaluer d'autres paramètres. Les effets ne seront pas corrélés avec une variabilité du contenu des séquences. Nous pourrions, par exemple, évaluer l'impact de la durée des événements ou encore de la durée des séquences. Nous pourrions ainsi, selon la durée des séquences provoquer des processus à court terme ou bien à long terme. Nous pourrions étudier les effets de récence relative évoqués par B. McElree, [Elr06], et P.Guelton,[Gue05].

Il sera également possible de mesurer l'effet du niveau sonore sur le jugement global des séquences. Dans notre expérience, nous avons proposé un ajustement écologique pour ne pas que l'attention des sujets soit focalisée sur un événement en particulier. Si l'un des sons de la séquence avait eu un niveau non écologique, le son aurait influencé le jugement global de la séquence. Nous pourrions donc évaluer l'influence du niveau sonore sur le jugement global de séquences.

4.4 Expérience de jugement d'agrément

La notion d'agrément est à considérer pour étudier le jugement global de séquences hétérogènes comprenant des sons agréables comme des sons de Parc ou de nature et des sons désagréables comme des sons de voitures ou de Trafic. Les jugements d'agrément semblent influencer les processus de mémorisation. Il sera intéressant d'approfondir la question.

Annexe A

Description du bruit de fond

A.1 Description du bruit de fond

A.1.1 Intérêt

Nous diffusons les événements sonores urbains dans un bruit de fond caractéristique d'un environnement urbain. En l'absence de bruit de fond, le séquençage des événements rend la scène peu cohérente. La présence d'un bruit de fond de type urbain induit, chez le sujet, un effet de contexte. D'autre part, nous souhaitons maîtriser l'organisation temporelle de la séquence. Il était alors difficilement envisageable d'enregistrer des séquences in situ et d'en contrôler l'organisation temporelle. Ainsi, l'alternative que nous proposons ici est satisfaisante d'un point de vue subjectif.

Pour résumer, l'intérêt du bruit de fond est double. Il rend la séquence cohérente temporellement et induit un effet de contexte qui facilite l'identification des sources par le sujet.

A.1.2 Analyse Spectrale

Les travaux de C. Guastavino [Gua03] concernent la caractérisation acoustique des bruits de fonds urbains. Ces travaux nous permettent de connaître les caractéristiques fréquentielles de ces bruits de fond.

A partir de bruits de fond répertoriés comme prototypique par un panel de sujets, C. Guastavino établit le contenu spectral de bruit de fond urbain de type circulation.

Le niveau sonore moyen est de 80 dB SPL. En dessous de 200 Hz, le niveau atteint 90 dB SPL pour son maximum. Le spectre décroît exponentiellement de 100 à 10000 Hz entre 80 et 20 dB SPL. Le bruit de fond urbain est caractérisé par un spectre dense en basses fréquences. La densité du spectre explique, du point de vue perceptif, la rugosité perçue, cf. section A.1.3.

D'autre part, V. Maffiolo [Maf99] mène une expérience auprès de citoyens. Cette étude a permis de classer subjectivement différentes ambiances urbaines. Parmi celles-ci,

nous retrouvons des ambiances "amorphes" où peu de sources émergent du bruit de fond ambiant.

Nous avons isolé du panel utilisé dans son étude, un court extrait d'une ambiance amorphe. L'extrait dure 1 seconde est échantillonné à 22100 Hz. Dans cette ambiance prédomine le son basse fréquence caractéristique du bruit de circulation des automobiles et poids lourds. Pour analyser ce bruit nous appliquons la méthode LPC. Cette méthode nous permet, à partir d'un extrait enregistré in situ, de synthétiser un bruit de fond qui sera très proche spectralement du bruit analysé.

A.1.3 Synthèse par Méthode LPC

Cette méthode, illustrée en figure A.1, utilise une modélisation de signal par prédiction linéaire. L'extrait à analyser est composé d'un bruit de fond stationnaire, dont les résonances fréquentielles sont prédictibles, et d'événements singuliers non prédictibles, figure A.2 et A.3. La partie stationnaire du signal, i.e : le bruit de fond, est la partie qui nous intéresse. L'analyse LPC appliquée à cet extrait fournit une estimation du signal stationnaire. On calcule ensuite la transformée de Fourier de ce signal. Connaissant le spectre et l'enveloppe spectrale du bruit de fond analysé, nous pouvons reconstruire un bruit de fond synthétique. L'enveloppe spectrale est appliquée comme filtre d'un bruit blanc. Le signal obtenu possède la même enveloppe spectrale que le bruit de fond initial, figure A.4.

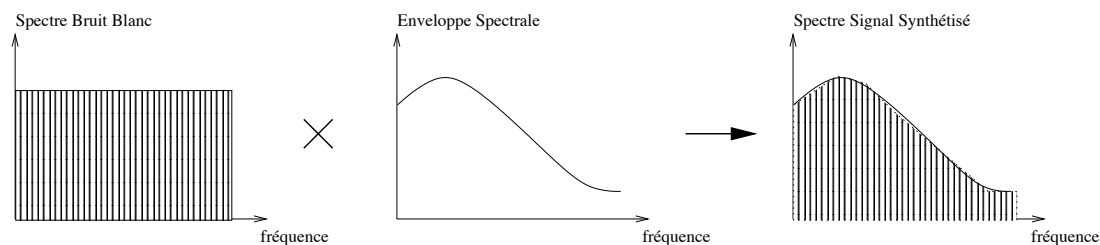


FIG. A.1 – Méthode LPC

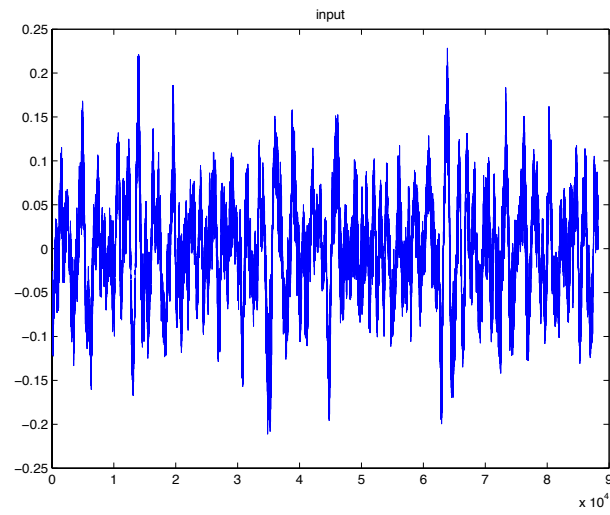


FIG. A.2 – Bruit de fond étudié par V. Maffiolo

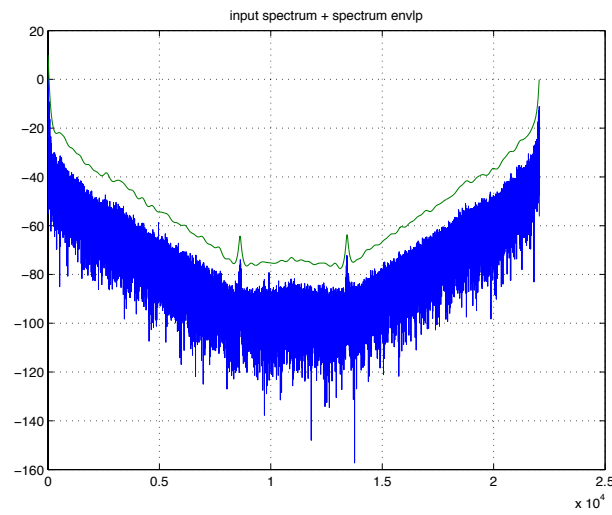


FIG. A.3 – Spectre et enveloppe spectrale du bruit de fond analysé

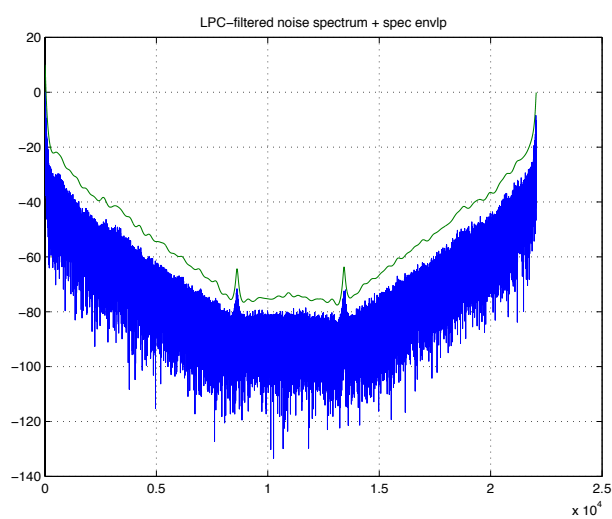


FIG. A.4 – Spectre et Enveloppe spectrale du bruit de fond synthétisé

Annexe B

Matériau sonore

B.1 Liste complète des sons extraits de la base de sons

Sources isolées	Sources multiples	Ambiances
alarme voiture	enfants cours d'école	marché
vélo passage rapide	enfants hall d'école	marché 2
passage d'un bus	enfants parc	commerce
camion benne	foule moyen	école élémentaire
camion pompier	applaudissement de foule	galerie d'art
canari	discussion foule	joggeur dans parc
chute d'arbre	rire de foule	spectateur de marathon
fauvette des jardins	grande foule	passage de fanfare
fontaine	klaxon dans embouteillage	gare routière
outil de jardinage	passage de camion pompier	trafic avec bus
joueurs de tennis en extérieur		trafic en centre ville
klaxon		trafic dense
klaxon voiture		trafic avec moto
passage de moto		trafic proche
battement d'ailes		vent dans les arbres
outil de ratissage		
ruisseau		
sirène de police		
main remuant l'eau		
métro en extérieur		
passage de vélo		
démarrage de voiture		
passage de voiture		

TAB. B.1 – Liste des sources extraites de *SoundIdeas*

Annexe C

Ajustement écologique en sonie

Son	Niveau initial	Niveau écologique
Présence Humaine		
ambiance de marché	66,2	71,2
école élémentaire	77,4	82
foule moyenne	62,5	68,9
applaudissement foule	68,3	73,7
discussion foule	69,4	74,9
rire de foule	71,6	73,7
spectateur marathon	70,1	76
passage parade	72,5	80,4
enfants cour d'école	67,8	73,8
enfants hall d'école	65,6	71,7

FIG. C.1 – Niveaux écologique en dB SPL des sons de Présence Humaine

Son	Niveau initial	Niveau écologique
Parc		
joggeur dans parc	61	65,8
canari	69,7	72,2
fauvette des jardins	75,1	79,7
battement d'aile	64,9	65,5
ruisseau	66,8	66,3
vent dans arbres	62,8	66,1
outils de jardinage	60,7	60,9
outils de ratissage	67,6	66,5
joueur de tennis extérieur	62,2	67,7
fontaine	64,9	62,6

FIG. C.2 – Niveaux écologique en dB SPL des sons de Parc

Son	Niveau initial	Niveau écologique
Trafic		
passage de pompier	72,6	79,7
passage de bus	75,8	81,6
camion benne	66,8	73,4
camion pompier	69,2	74
klaxon voiture	66,8	73,5
passage de moto	71,4	76,1
gare routière	72,5	77,6
trafic centre ville	68,8	74,6
passage de voiture	69,7	73,5
trafic proche	76,1	80,6

FIG. C.3 – Niveaux écologique en dB SPL des sons de Trafic

Bibliographie

- [A.B93] James A.Ballas. *Common factors in the identification of an assortment of brief everyday sound*, volume 19. Journal of experimental Psychology : Human perception and performance, 1993.
- [AH87] James A.Ballas and James H.Howard. *Interpreting the language of environmental sounds*, volume 19. Environment and behavior, Department of Psychology, Washington DC, 1987.
- [Bad00] Alan Baddeley. The episodic buffer : a new component of working memory ? In *Trends in Cognitive Siences*, volume 4. Elsevier Science Ltd, 2000.
- [Bad02] Alan D. Baddeley. *Theoretical Background*, chapter The Psychology of Memory, pages 1–15. Department of Experimental Psychology, University of Bristol, UK, john wiley and sons, ltd edition, June 2002.
- [Bar77] J.C. Barlett. Remembering environmental sounds : The role of verbalisation at input. *Memory and Cognition*, 5 :404–414, 1977.
- [BH73] G.H. Bower and K. Holyoak. Encoding and recognition memory for naturalistic sounds. *Journal of Environmental Psychology*, 101(2) :360–366, 1973.
- [BM91] J.A Ballas and T. Mullin. Effects of context on the identification of everyday sounds. *Human Performance*, 4(3) :199–219, 1991.
- [Chi00] C.Y. Chiu. Specificity of auditory implicit and explicit memory : is perceptual priming for environmental sound exemplar specific ? *Memory and Cognition*, 28(7) :1126–1139, 2000.
- [CQ69] A.M. Collins and M.R. Quillian. Retrieval time from semantic memory. *Journal of verbal learning and verbal behaviour*, (8) :240–247, 1969.
- [CS95] C.Y. Chiu and D.L. Schacter. Auditory priming for nonverbal information : Implicit and explicit memory for environmental sounds. *Consciousness and Cognition*, 4 :440–458, 1995.
- [Eic85] J.E. Eich. Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 11(4) :764–770, 1985.
- [Elr06] Brian Mc Elree. *Accessing Recent Events*, volume 46. Elsevier Inc, 2006.
- [Gal05] Galotti. *Cognition : Semantic Memory*, chapter Memory for General Knowledge. Number Psyc 3705. November 2005.

- [GB80] R.E. Geiselman and R.A. Bjork. Primary versus secondary rehearsal in imagined voices : differential effects on recognition. *Cognitive Psychology*, 12 :188–205, 1980.
- [GC01] L. Gros and N. Chateau. Instantaneous and overall judgements for time-varying speech quality : Assessments and relationships. In *ACUSTICA*, volume 87, pages 367–377. France telecom R et D, Technopole Anticipa, France, 2001.
- [Gér04] Yannick Gérard. *Mémoire sémantique et sons de l'environnement*. PhD thesis, Université de Bourgogne, France, décembre 2004.
- [Gua03] C. Guastavino. *Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain*. PhD thesis, Université Paris IV, Paris, 2003.
- [Gue05] Pierre Guelton. Effets de mémoire auditive liés à la perception d'une séquence de sons environnementaux. Institut de Recherche et de Coordination Acoustique/Musique (IRCAM) , Paris, FRANCE, 2005.
- [HB80] J.H. Howard and J.A. Ballas. Syntactic and semantic factors in the classification of nonspeech transient patterns. *Perception et Psychophysics*, 28 :431–439, 1980.
- [Hin03] Douglas L. Hintzman. *Judgments of recency and their relation to recognition memory*, volume 31, pages 26–34. Synchronomic Society, Inc, University of Oregon, Eugene, Oregon, 2003.
- [HKZ⁺01] J. Hellbrück, T. Kato, A. Zeitler, S. Kuwano A. Schick, and S. Namba. Loudness scaling of traffic noise : Perceptual and cognitive factors. *Recent trends in basic psychophysics and their applications to acoustics*, 2001.
- [KCS04] Krystal A. Klein, Amy Criss, and Richard Shiffrin. Recency judgments and list context. Department of Psychology, Indiana University, USA, 2004.
- [KP85] C. Krafka and S. Penrod. Reinstatement of context in a field experiment on eyewitness identification. *Journal of personality and Social Psychology*, 49 :58–69, 1985.
- [LC78] D.M. Lawrence and N.J. Cobb. Cross-modal utilization of information : Recognition memory for environmental stimuli. *Perceptual and Motor skills*, 47 :1203–1206, 78.
- [Lee86] R. Duncan Lee. *Response Times - Their role in inferring elementary mental organization*, chapter 2. Simple Reaction Time Basic Data. Oxford Science Publications, 1986.
- [Maf99] Valérie Maffiolo. *De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain*. PhD thesis, Université Paris VI, 1999.
- [MCD98] Valérie Maffiolo, Michèle Castellengo, and Danièle Dubois. Qualité sonore de l'environnement urbain : sémantique et intensité, 98.
- [MD81] R.S. Malpass and P.G. Devine. Guided memory in eyewitness identification. *Journal of applied Psychology*, 66 :343–350, 1981.

- [ML01] Jeff O. Miller and Kathy Low. Motor processes in simple, go/nogo and choice reaction time tasks : A psychophysiological analysis. In *Journal of Experimental Psychology : Human Perception and Performance*, volume 27, pages 266–289. 2001.
- [MVP⁺97] V. Maffiolo, C. Vogel, J-D. Polack, M. Castellengo, D. Dubois, and S. David. Ambiances sonores représentatives d’une ville : le cas de paris. In teknea, editor, *Actes du 4ième congrès français d’acoustique*, volume 1, pages 303–306, Marseille, France, avril 1997. Société française d’acoustique.
- [NK81] S. Nixon and N.J. Kanak. The interactive-effects of instructional set and environmental context changes on the serial position effect. *Bulletin of the psychonomic Society*, 18 :237–240, 1981.
- [PBG⁺00] Maura Pilotti, Erik T. Bergman, David A. Gallo, Mitchell Sommers, and Henry L. Roediger III. Direct comparison of auditory implicit memory tests. In *Psychonomic Bulletin and Review*, volume 7, pages 347–353. Psychonomic Society, Inc, Washington University, Saint Louis, Missouri, 2000.
- [PPR75] A. Paivio, R. Philipchalk, and E.J. Rowe. Free and serial recall of pictures, sounds and words. *Memory and Cognition*, 3 :586–590, 1975.
- [SG04] Valeriy Shafiro and Brian Gygi. How to select stimuli for environmental sound research and where to find them? *Behavior Research Methods, Instruments and Computers*, 36(4) :590–598, 2004.
- [SJ95] G.P. Stuart and D.M. Jones. Priming the identification of environmental sounds. *Quarterly Journal of Experimental Psychology*, 38a :741–761, 1995.
- [Smi88] Steven M. Smith. *Environmental Context-Dependent Memory*, chapter 2, pages 13–33. John Wiley & Sons Ltd, 1988.
- [SMS02] Patrick Susini, Stephen McAdams, and Bennet K. Smith. Global and continuous loudness estimation of time-varying levels. In *ACUSTICA*, volume 88, pages 536 – 548. Institut de Recherche et Coordination Acoustique/Musique, Paris, France, 2002.
- [Sur01] Aimée M. Surprenant. *Distinctiveness and serial position effects in tonal sequences*, volume 63. Purdue University, Indiana, 2001.
- [SWH05] Daniela B. Sellner, Michael R. Waldmann, and Keith J. Holyoak. Accessing causal relations in semantic memory. Department of Psychology, University of Göttingen, Germany, 2005.
- [TS90] E. Tulving and D.L. Schacter. Priming and human memory systems. *Science*, 247(4940) :301–306, 1990.
- [Web91] Reinhard Weber. The continuous loudness judgement of temporally variable sounds with an analog procedure. In *Contributions to Psychological Acoustics*, pages 267–293. University of Oldenburg, Departement of Physics, Acoustic Section, 1991.