

THÈSE de DOCTORAT de l'UNIVERSITÉ PARIS 6

ECOLE DOCTORALE : CERVEAU COGNITION COMPORTEMENT

Présentée par Clara Suied

Pour obtenir le grade de Docteur de l'Université Paris 6

DE L'URGENCE PERÇUE AU TEMPS DE REACTION :

Application aux alarmes sonores

Soutenance publique le 13 avril 2007 devant le jury composé de :

Georges Canévet (Rapporteur)

Anne Guillaume (Rapporteur)

Jean-Dominique Polack (Examineur)

Séverine Samson (Examineur)

Stephen McAdams (Directeur de la thèse)

Patrick Susini (Co-directeur de la thèse)

Sabine Langlois (Encadrante industrielle)

Remerciements

Je tiens à remercier en premier lieu Patrick Susini pour avoir encadré cette thèse. Merci pour son implication, tant au travers de ses compétences scientifiques que de ses encouragements amicaux.

Je remercie également Stephen McAdams pour avoir accepté de diriger cette thèse, pour son soutien et ses conseils.

Merci à Sabine Langlois pour le suivi de cette thèse, pour la confiance qu'elle m'a apporté et pour m'avoir laissé la liberté nécessaire à l'accomplissement de mes travaux, tout en gardant un oeil critique et avisé.

Merci à Hugues Vinet (STMS-IRCAM-CNRS, Paris) et Virginie Maillard (Renault, Guyancourt) pour m'avoir accueilli au sein de leurs équipes pendant cette thèse.

Merci à Nicolas Misdariis pour son aide efficace et pour la gestion de tous ces petits problèmes qui, sans lui, auraient largement perturbé le déroulement de cette thèse.

Merci à Gaël Guyader, qui, au travers de discussions stimulantes, m'a aidé à construire ma pensée.

Merci à Laurent Pruvost et Marie Magnin, dont les stages de fin d'études ont contribué à ce travail. Merci pour leur engouement et leur enthousiasme.

Merci à Guillaume Saint-Pierre pour l'analyse statistique du Chapitre 5. Merci pour son aide.

Merci à Georges Canévet d'avoir accepté d'être rapporteur de cette thèse. Merci pour sa lecture détaillée de chaque partie de cette thèse ainsi que des échanges constructifs qui ont suivi.

Merci à Anne Guillaume pour son intérêt et ses encouragements répétés tout au long de ces trois années, ainsi que d'avoir accepté d'être rapporteur de cette thèse.

Merci à Séverine Samson d'avoir enrichi la soutenance de cette thèse par son regard de neuropsychologue.

Merci à Jean-Dominique Polack, président du jury, d'avoir accepté de juger ce travail.

Merci à toute l'équipe Système, à Sylvie Benoit-Stanek et Didier Perini, pour leur efficacité et leur disponibilité.

Merci à tou(te)s celles et ceux qui ont partagé mon quotidien pendant ces années, aussi bien à l'Ircam que chez Renault. Merci pour leur aide, leur bonne humeur et leurs rires.

Merci à Isabelle Viaud-Delmon et Olivier Warusfel, pour m'ouvrir la voie de la recherche vers de nouveaux questionnements, ainsi pour leur incroyable générosité.

Merci à tou(te)s mes ami(e)s, qui, de près ou de loin, ont rendu plus facile la réalisation de ce travail, et grâce à qui la vie est plus lumineuse.

Enfin, merci à mes parents, pour leur soutien incomparable et leur sagacité bienveillante. Merci à ma mère pour le temps passé à la relecture de ce manuscrit. Merci à ma famille, si précieuse et irremplaçable.

A Daniel

Table des matières

Remerciements	iii
1 Introduction	1
I Contexte théorique	5
2 Perception et fonction d'un son	7
2.1 Introduction	7
2.2 Apparition d'un nouveau domaine	7
2.2.1 Pourquoi?	7
2.2.2 Point de vue historique	8
2.2.3 Champs d'application	9
2.2.4 Fondements et légitimations	10
2.2.5 Application de la perception auditive	12
2.3 Perception auditive	13
2.3.1 Théories	13
2.3.2 Méthodes	14
2.4 Psychologie de la perception auditive	15
2.4.1 Contexte	15
2.4.2 Étapes de traitement de l'information auditive	16
2.5 Perception des sons environnementaux	18
2.5.1 Définition des sons environnementaux	18
2.5.2 Acoustique écologique	20
2.5.3 Écoute de la source du son	20
2.5.4 Identification des sons environnementaux	22
2.6 Interfaces auditives	25

2.6.1	Introduction	26
2.6.2	Earcons	26
2.6.3	Icônes auditives	32
2.6.4	Lien entre les Earcons et les Icônes Auditives	35
2.7	Conclusion	38
3	Urgence perçue	39
3.1	Introduction	39
3.2	Typologie d'une alarme	39
3.2.1	Constat de l'existant	39
3.2.2	Proposition de conception d'une alarme	42
3.3	Paramètres acoustiques et urgence perçue	47
3.3.1	Influence des paramètres acoustiques sur l'urgence perçue	47
3.3.2	Quantification des effets	52
3.3.3	Limites de cette approche	54
3.4	Sons environnementaux et alarmes	55
3.5	Influence des facteurs cognitifs sur l'urgence perçue	57
3.6	Conclusion	59
4	Alarmes sonores et temps de réaction	61
4.1	Introduction	61
4.2	De l'urgence perçue au temps de réaction	61
4.2.1	Définition de l'urgence	61
4.2.2	Urgence objective de sons abstraits	62
4.2.3	Urgence objective de sons environnementaux	65
4.2.4	Limites de ces approches	67
4.3	Temps de réaction	68
4.3.1	Introduction	69
4.3.2	Modèle de Donders	69
4.3.3	Paradigmes de détection	71
4.4	Conclusion	72
4.5	Vers une nouvelle caractérisation des alarmes sonores	73

II	Travaux expérimentaux	75
5	Fonction perçue des IHM sonores	77
5.1	Introduction	77
5.2	Catégories fonctionnelles de l'automobile	79
5.3	Hypothèses expérimentales	80
5.3.1	Paramètres descriptifs de l'impulsion	81
5.3.2	Paramètres descriptifs du train d'impulsions : le rythme	84
5.3.3	Paramètre descriptif de l'alarme finale : la répétition	84
5.4	Méthode	85
5.4.1	Participants	85
5.4.2	Stimuli	85
5.4.3	Matériel	90
5.4.4	Procédure	90
5.5	Résultats	91
5.5.1	Analyses descriptives	91
5.5.2	Analyse des sujets	98
5.5.3	Arbres de discrimination	100
5.5.4	Régression logistique	103
5.5.5	Synthèse des résultats	115
5.6	Discussion	118
5.6.1	Influence du rythme	118
5.6.2	Influence de la F0 et du Timbre	120
5.6.3	Influence de la Modulation et de la Polyphonie	121
5.6.4	Sons <i>intentionnels</i> : limites et méthode	122
5.7	Conclusion	124
6	Evaluation de l'urgence par le TR	125
6.1	Introduction	125
6.1.1	What is called "urgency" ?	127
6.1.2	The Current Experiments	130
6.2	Experiment 1	131
6.2.1	Method	131
6.2.2	Results	133
6.2.3	Discussion	137

6.3	Experiment 2	140
6.3.1	Method	140
6.3.2	Results	141
6.3.3	Discussion	143
6.4	Experiment 3	146
6.4.1	Method	146
6.4.2	Results	147
6.4.3	Discussion	151
6.5	General Discussion	154
6.6	Conclusion	157
7	Sons synthétiques et sons naturels	159
7.1	Introduction	159
7.2	Expérience 1	161
7.2.1	Méthode	161
7.2.2	Résultats	164
7.2.3	Discussion	167
7.3	Expérience 2	171
7.3.1	Méthode	171
7.3.2	Résultats	175
7.3.3	Discussion	179
7.4	Conclusion	184
8	Conclusion	187
8.1	Synthèse des résultats obtenus	187
8.2	Création d'alarmes sonores	190
	Bibliographie	193
A	Compléments statistiques au Chapitre 5	205
A.1	Analyses descriptives	205
A.1.1	Tables de contingence	205
A.1.2	Tests d'indépendance	206
A.1.3	Coefficient d'association entre variables ordinales	207
A.2	Régression Logistique	208
A.2.1	Introduction	208

A.2.2	Régression logistique binomiale	209
A.2.3	Régression logistique multinomiale ou polytomique	209
A.2.4	Sélection du modèle logistique multinomial : critères de qualité du modèle et sélection de variables	210
A.2.5	Validation du modèle	211
A.2.6	Interprétation des résultats du modèle	212
	Résumé	215

Chapitre 1

Introduction

Le contraste entre la complexité et la diversité de notre environnement sonore quotidien et la simplicité avec laquelle nous naviguons dans ce monde sonore est saisissant. L'étude de la perception auditive humaine tente de lever le voile sur ce lien entre la description purement physique de notre univers sonore et la perception et l'appréhension que nous en avons. Les sources sonores auxquelles nous sommes confrontés peuvent se différencier - entre autres - selon la façon dont elles véhiculent une information. Essentiellement, il est possible d'opposer les sons intentionnels - *créés* avec la volonté de communiquer un message ; sirène, alarme, cloche, etc. - aux sons non intentionnels - sons intrinsèques au mécanisme d'un objet comme le son d'un aspirateur ou issus de l'excitation d'un objet, comme le frottement d'une plaque en métal.

Le thème de notre recherche concerne la première catégorie de sons, les sons intentionnels. L'émergence de ces sons est liée aux nouvelles possibilités technologiques permettant d'intégrer des systèmes sonores dans les produits ou les interfaces que nous manipulons au quotidien (distributeur de billets SNCF, molette de navigation des lecteurs de musique, par exemple). Ces nouveaux systèmes sonores miniatures embarqués, souvent appelés IHM (pour *Interfaces Homme-Machine*) sonores, permettent de produire des signaux sonores très riches. Cependant, même si le champ de possibilités sonores offertes par la technologie actuelle est infini, il n'en reste pas moins que les sons créés doivent répondre à des contraintes *fonctionnelles*.

Les travaux présentés dans cette thèse ont été effectués dans le cadre d'une Convention Industrielle de Formation par la Recherche (CIFRE) entre la Direction de la Recherche de Renault et l'équipe Perception et Design Sonores de l'IRCAM. Dans le contexte automobile, les IHM sonores sont de plus en plus nombreuses et permettent d'informer le

conducteur d'un dysfonctionnement du véhicule, d'un oubli (phares allumés, ceinture de sécurité non attachée) ou lui confirment simplement que tel ou tel système est bien activé (fermeture automatique du véhicule). Afin que ces IHM sonores puissent être comprises et que le conducteur y réagisse de manière adéquate, il était alors nécessaire de comprendre comment transmettre une information par la modalité auditive. Il s'agit alors de donner des recommandations fonctionnelles à la création d'IHM sonores porteuses de sens.

Les questions ergonomiques - quand est-il judicieux d'informer le conducteur par le biais d'une information sonore - et esthétiques - comment véhiculer une image de marque et de qualité par la dimension sonore - n'ont pas été abordées directement dans le cadre de cette thèse. Néanmoins, ces questions ont été traitées chez Renault en amont (pour la partie ergonomique) et en aval (pour le côté esthétique) de nos travaux¹.

Fonction d'un son Les premiers travaux menés afin d'étudier la fonction évoquée par un son donné, se sont inspirés principalement des domaines du design et de l'ergonomie (Blattner et coll., 1989 ; Brewster et coll., 1994). Elles s'inscrivent au sein d'un nouveau domaine de recherche pluridisciplinaire (AD, pour *Auditory Display*), dont l'un des objectifs principaux est de traduire des informations ou des processus à l'aide d'une représentation sonore.

Cette approche appliquée implique en fait une démarche "inverse" à celle classiquement adoptée dans l'étude de la perception auditive. Plutôt que de chercher à comprendre la perception auditive par l'étude de certains sons ou attributs auditifs, il s'agit ici d'identifier quels seront les composantes sonores les plus aptes à être perçues afin de satisfaire le cahier des charges fonctionnel. Ainsi, la volonté du constructeur permet de définir *a priori*, sous la forme de ce cahier des charges, la perception, et par là même la *réaction* que devront avoir les auditeurs à l'écoute de l'IHM sonore.

Urgence perçue Parmi les différentes fonctions et messages possibles, la notion d'urgence a été particulièrement développée, indépendamment du domaine des AD. L'étude de l'urgence perçue a été entreprise en utilisant des paradigmes définis par la psychophysique tels l'estimation de grandeur ou la production de rapports, par exemple (Edworthy et coll., 1991 ; Hellier et coll., 1993). Plusieurs facteurs interviennent pour nous permettre de juger de l'urgence d'une alarme sonore ; essentiellement, acoustiques et cognitifs (Guillaume et

¹La création d'IHM sonores pour un futur véhicule a été effectuée par le compositeur Andrea Cera en collaboration avec l'équipe Perception et Design Sonores de l'IRCAM

coll., 2003). Ce point peut en fait être reformulé dans une perspective plus générale, ces différents facteurs étant impliqués dans notre expérience humaine quotidienne du monde sonore (McAdams et Bigand, 1994).

Néanmoins, comment énoncer précisément ce que doit être une alarme sonore, au regard de ce modèle global de l'urgence perçue ? Comment traduire en recommandations fonctionnelles précises l'influence de facteurs cognitifs sur le jugement d'urgence ?

Urgence objective Considérons à nouveau la démarche "inverse" décrite précédemment. Contrairement à certaines IHM sonores qui apparaissent couramment dans notre vie quotidienne (clignotant, radar de recul, par exemple), certaines alarmes sonores ne sont que peu, voire jamais, entendue par les conducteurs, puisque signalant une situation anormale liée à un danger. La proposition de création sonore ne peut certes pas se contenter de règles arbitraires qui seront apprises et adoptées peu à peu, comme toute convention se développant dans le temps et par la pratique d'un groupe culturel donné. Comment définir certaines caractéristiques sonores qui aboutiront à la création d'une alarme sonore ajustée à l'intention souhaitée ? Davantage que la *perception* du degré d'urgence d'une alarme sonore, c'est la *réaction* de l'auditeur qui est attendue et voulue, puisque c'est bien de l'immédiateté de la réaction dont il s'agit dès lors que l'urgence d'un son est évoquée. Il devient alors crucial de questionner à nouveau, par un éclairage différent, la notion d'urgence.

Le lien entre l'urgence perçue et le temps de réaction a été étudié pour tenter de mieux appréhender les questions relatives aux alarmes (Haas et Casali, 1995 ; Burt et coll., 1995). Deux approches de l'urgence sont ainsi comparées : l'urgence *subjective*, ou perçue, et l'urgence *objective*, mesurée par un temps de réaction.

La compréhension de ce qui nous fait réagir le plus rapidement possible à l'écoute d'un son se doit de considérer un modèle de temps de réaction. L'observation de différents temps de réaction nous permettra alors d'inférer sur la vitesse de traitement de processus mentaux variés (Donders, 1868 - 1969) impliqués dans le processus perceptif des alarmes sonores.

Organisation de la thèse Les travaux rapportés ici abordent principalement cette question de l'édification d'une nouvelle démarche de création d'alarmes sonores efficaces et permettant de transmettre le degré d'urgence exigé.

Dans un premier temps, nous présenterons un état actuel des connaissances nécessaires à la création de sons non verbaux porteurs de sens, issues principalement des travaux relatifs au domaine des *Auditory Display*, en les resituant dans une perspective de perception et

cognition auditives (Chapitre 2). Puis, nous nous intéresserons à l'urgence des alarmes sonores sous deux angles bien distincts. Le premier concerne l'approche subjective, celle de l'urgence perçue (Chapitre 3). Les limites de cette démarche nous permettront alors de définir l'urgence objective comme deuxième angle d'approche. Elle est caractérisée par l'étude de la réaction, et plus précisément des temps de réaction (Chapitre 4).

Les chapitres suivants (Chapitres 5 à 7) décrivent les travaux expérimentaux menés dans le cadre de cette thèse. Le Chapitre 5 sera consacré à l'exposition de nouvelles données qui permettent de valider et d'étendre, dans le cadre de l'automobile, les résultats obtenus dans la littérature sur la création d'IHM sonores porteuses de sens. Les Chapitres 6 et 7 aborderont la question d'intérêt de cette thèse par le biais des temps de réaction. Le Chapitre 6 permettra également d'étudier cette question en contexte d'attention partagée, simulant la conduite automobile.

La nouvelle approche que nous proposerons - via les temps de réaction - pour la création d'alarmes sonores performantes apportera également certaines explications sur les processus sensoriels et cognitifs mis en jeu. Ces dernières indications paraîtront sans doute au lecteur un peu éloignées de la contrainte industrielle de départ et de la création sonore en tant que telle. Pourtant, nous sommes convaincus que c'est cette compréhension des phénomènes impliqués dans le traitement des alarmes qui pourra, à terme, donner lieu à des préconisations de création sonore robustes et complètes.

Première partie
Contexte théorique

Chapitre 2

Perception et fonction d'un son

2.1 Introduction

Ce chapitre présente un état actuel des connaissances nécessaires à la création de sons non verbaux porteurs de sens. Nous présenterons tout d'abord un nouveau domaine de recherche pluridisciplinaire, ICAD, dont un des objectifs est justement de chercher à traduire des données à l'aide de la modalité auditive (section 2.2). Nous avons choisi de traiter cette question de la création sonore par l'intermédiaire de la perception auditive, que nous exposerons alors très brièvement (section 2.3). Nous pourrons alors aborder plus en détail quelques aspects psychologiques de la perception auditive, tels que décrits dans une approche de type traitement de l'information (section 2.4). Puis, nous détaillerons l'analyse perceptive d'un type de sons particulier : les sons environnementaux (section 2.5). Ces considérations nous permettront alors de décrire les études majeures concernant les interfaces auditives en les situant dans la problématique de la perception et de la cognition auditive (section 2.6).

2.2 Apparition d'un nouveau domaine

2.2.1 Pourquoi ?

Dans notre vie quotidienne, nous sommes entourés par une multitude de sons qui nous permettent d'extraire des informations utiles sur le monde environnant. Partant de ce constat assez simple, certains concepteurs de systèmes informatiques ont proposé d'améliorer l'utilisation d'Interfaces Homme Machine (IHM) à l'aide de la modalité auditive, et

non plus seulement grâce à des icônes ou symboles visuels (Kramer, 1994). L'apparition récente de cette nouvelle modalité sensorielle dans les IHM est ainsi présentée comme un nouveau moyen d'interaction entre les utilisateurs et les outils à leur disposition. Notons ici que cet ajout de la modalité auditive dans les IHM va de pair avec la possibilité de synthétiser et d'intégrer du son dans les systèmes électroniques ou informatiques.

La question qui se pose alors est de savoir comment un son non verbal peut transmettre des informations aux auditeurs.

2.2.2 Point de vue historique

Un premier événement a constitué une étape importante dans le processus de création de sons non verbaux faciles à utiliser et à interpréter, et a ainsi contribué à la création d'un domaine et d'une communauté scientifique pluridisciplinaire : la première International Conference on Auditory Display (ICAD), en 1992, et qui a par la suite donné lieu à la publication d'un ouvrage (Kramer, 1994).

Outre l'intérêt de rassembler diverses personnes qui travaillaient sur les mêmes sujets, cette conférence a surtout permis de donner les fondations de certains mots clés du domaine, et de poser les questions qui ont inspiré la plupart des travaux jusqu'à ce jour. Comment représenter une information ? Est-il préférable d'utiliser des sons familiers, appris avec peu d'efforts, ou de créer un langage à partir de courtes séquences mélodiques, qui sont, de fait, plus difficiles à apprendre ? Dans le cas où des messages sonores et visuels sont utilisés conjointement, comment répartir les tâches ? Dans quels cas les phrases parlées deviendraient-elles plus efficaces ? Bien que cette dernière question soit justifiée et raisonnable, les études portant sur la parole ont été délibérément exclues du champ d'investigation d'ICAD, attendu qu'il existe déjà de très nombreuses études, journaux et conférences portant spécifiquement sur ce thème. Le terme *ICAD* sera utilisé dans la suite du document pour se référer au domaine de recherche, et non plus uniquement au congrès associé. Pour une définition précise du domaine, le lecteur pourra également se reporter au site <http://www.icad.org/websiteV2.0/References/nsf.html>.

Une réponse rigoureuse à ces interrogations n'est pas triviale, et diverses connaissances préalables sont nécessaires. En particulier, les recherches concernant la perception auditive sont généralement considérées (ou devraient l'être, du moins) comme des pré-requis nécessaires à l'élaboration de sons non verbaux dans des cadres applicatifs (informatiques par exemple). Nous commenterons ce point au cours du chapitre.

2.2.3 Champs d'application

Une kyrielle d'approches différentes a été suggérée pour imaginer comment les usagers de systèmes d'information pourront utiliser le son à bon escient. Ces problématiques et questions permettent alors de diviser la communauté ICAD en plusieurs sous domaines, et il nous a semblé utile de donner ici quelques définitions de ces champs d'applications. En effet, suivant les auteurs, des définitions parfois un peu différentes peuvent être rencontrées (comme dans tout nouveau domaine, il faut un certain temps pour aboutir à une définition consensuelle et partagée des termes employés). Ces définitions sont adaptées de l'ouvrage de Kramer (1994). Nous distinguons deux manières de coder des informations par le son : la traduction sonore de données quantitatives (la *sonification*) et la traduction sonore de différentes fonctions (les *interfaces auditives*).

La sonification est définie comme une cartographie sonore de données quantitatives (cours de la bourse, données sismiques, ...) et consiste à *contrôler* le son par les données. Considérons l'exemple des données sismiques (Hayward, 1994) pour illustrer cette définition. L'auteur défend la théorie suivant laquelle il devrait être possible de repérer et discriminer des tremblements de terre ou des explosions nucléaires autrement qu'avec des représentations graphiques : en "écoutant la terre chanter". En constatant que les équations d'onde du son dans l'air et des ondes sismiques dans la roche sont similaires et suivent les mêmes règles physiques, il "audifie" ainsi des enregistrements sismiques, en mettant à l'échelle les signaux sismiques pour qu'ils s'adaptent aux contraintes inhérentes à l'audition humaine.

De nombreux autres exemples de sonification ont été traités par la communauté ICAD (de 1992 à 2006), la sonification étant l'un de ses thèmes de recherche majeurs. Toutefois, comme il n'a pas de lien direct et évident avec notre sujet, nous ne le traiterons pas plus en détail. Le lecteur intéressé pourra se reporter - par exemple - à deux exemples d'application bien détaillés, celui de Neuhoff et coll. (2002) traitant de la sonification de données issues de cours de bourse, et celui d'Edworthy et coll. (2004) où des sons servant à la surveillance d'hélicoptères sont présentés au sein d'un cadre méthodologique plus global.

Le but d'une interface auditive est de transmettre un message, ou, en d'autres termes, la *fonction* du son. Dans le cadre de cette thèse, seules les études sur les interfaces auditives seront abordées. En effet, dans le cadre d'une application automobile, il s'agit d'informer le conducteur d'un non-fonctionnement du véhicule (défaillance du moteur), par exemple, et de l'inciter à agir (s'arrêter, boucler sa ceinture, etc.). Les interfaces auditives feront l'objet d'une section particulière, ci-après (section 2.6).

Cette volonté - et nécessité - de créer des sons informatifs est souvent assimilée aux sons d'alertes et aux diverses alarmes sonores (sirènes de pompier ou d'ambulance par exemple). Dans notre cas, la question de la fonction du son se pose également en termes d'urgence du son. Cependant, historiquement, les études portant sur les alarmes et l'urgence perçue n'ont pas été traitées par la communauté ICAD. Les études sur les interfaces auditives sont en fait centrées davantage sur des sons de rétroaction (pour *feedback*) informatique, ou sur des sons de *veille* (suivi de l'activité d'une machine en continu, dans une usine par exemple). Il semble que l'originalité d'ICAD soit justement d'aborder ce type d'informations, de nombreux travaux sur l'urgence perçue existant déjà par ailleurs (Patterson, 1982 ; Edworthy et coll., 1991 par exemple). Les alarmes sonores et l'urgence perçue feront l'objet du prochain chapitre.

2.2.4 Fondements et légitimations

Avant de se demander comment créer un son pour une fonction donnée, des questions bien légitimes se doivent d'être soulevées : pourquoi utiliser du son là où existe déjà du visuel ? Pourquoi ajouter des informations sonores, alors que notre environnement sonore quotidien est déjà considéré comme trop bruyant ? Ne devrait-on pas tendre plutôt vers un univers de silence ?

Audition et vision

Considérons tout d'abord le son par rapport à la vision. Bregman (1994) positionne cette question d'un point de vue global, en proposant de comprendre la différence entre le fonctionnement de ces deux modalités sensorielles, audition et vision. Lors de la résolution d'une tâche complexe, chaque sens est utilisé à des fins différentes. Il lui semble donc opportun de combiner vision et audition, et d'utiliser ainsi chaque modalité sensorielle pour ce qu'elle fait le mieux.

Le son peut alors être employé lorsque le visuel est saturé d'informations, afin de réduire la charge du système visuel, par exemple la quantité d'informations nécessaires à l'écran (Brewster, 2003).

Il peut également s'avérer nécessaire de réduire l'attention visuelle en ajoutant du son. En d'autres termes, parce que les utilisateurs ne peuvent pas garder leur regard rivé sur un écran d'ordinateur en continu, et parce qu'ils ne regardent pas toujours au bon moment au bon endroit (cas de contrôleurs aériens qui suivent l'activité de trois à cinq écrans

simultanément), la modalité auditive est alors ajoutée et permet de pallier ce type de problèmes (Kramer, 1994 ; Brewster, 2003) : nous pouvons fermer les yeux, pas les oreilles... Deux types d'informations peuvent alors être communiquées simultanément, une à l'écran, visuellement et une autre auditivement. Nous pouvons ainsi imaginer un exemple où le son apporterait une information supplémentaire à l'utilisateur d'une interface visuelle. Le cas de l'automobile est assez parlant : vous êtes en train de conduire, et regardez donc naturellement la route et les environs. Un dysfonctionnement du véhicule survient, un voyant s'allume dans un coin du tableau de bord pour l'indiquer. Cependant, rien ne vous amène à regarder le tableau de bord, et plus spécifiquement ce voyant, alors qu'une réaction immédiate serait nécessaire (s'arrêter). Il est donc nécessaire d'ajouter une information sonore pour attirer votre attention, et vous pousser à prendre connaissance du voyant allumé (reste à savoir quelle information sonore sera la plus adéquate pour remplir cette fonction, mais c'est justement le but de ce travail de thèse).

Enfin, Brewster (2003) cite également le cas des non-voyants, pour qui, bien évidemment, la composante sonore est indispensable à l'utilisation d'Interfaces Homme Machine.

Sons non verbaux *versus* voix humaine

A ce stade, il semble donc nécessaire de bénéficier d'IHM sonores pour une utilisation adéquate de systèmes d'informations. Pourquoi alors utiliser des sons non verbaux plutôt que des mots ? Gaver (1994) donne là encore quelques pistes pour expliquer ce choix. Suivant le principe qu'"un petit dessin vaut mieux qu'un long discours", il constate que, pour transmettre une même information, le message sera souvent plus long dans le cas de messages parlés. De plus, l'auditeur devra se concentrer du début à la fin s'il souhaite comprendre le message dans sa totalité. Kramer (1994) précise que lorsque les sons utilisés sont connus, la réaction à des sons non verbaux est plus rapide qu'à de la parole (cette affirmation mériterait un plus ample débat). L'hypothèse faite ici par Kramer repose essentiellement sur le fait que pour une même information, un son parlé a une durée totale supérieure à celle d'un simple "bip".

Il est également souvent avancé par les auteurs de la communauté ICAD qu'un son non verbal est, par définition, plus universel que de la parole, et plus apte à être compréhensible par tous.

Dans le cas de l'automobile, très pragmatiquement et en-dehors de toute validation expérimentale, nous pouvons simplement constater que les conducteurs n'aiment pas entendre leur voiture leur parler. De plus, les sons non verbaux sont plus faciles à mettre en

oeuvre industriellement : d'une part, l'exportation des voitures dans divers pays impliquerait une traduction des différents messages verbaux, et d'autre part, l'utilisation de la voix nécessite souvent des émetteurs plus sophistiqués, et de ce fait plus chers (contrairement à de simples buzzers employés dans le cas de sons non verbaux).

Nous avons ainsi résumé l'intérêt de la création d'un domaine tel qu'ICAD. Dès lors, la problématique de ce travail de thèse est justifiée, et il devient légitime de se demander quel type de sons sera le plus efficace pour transmettre un message donné. Comprendre les mécanismes (sensoriels et cognitifs) sous-jacents prend alors tout son sens.

2.2.5 Application de la perception auditive

Les travaux issus de la communauté ICAD ont tous un but applicatif (avec des applications parfois très diverses). De ce fait, l'objectif final n'est pas l'étude de la perception auditive en tant que telle. Cependant, comme le souligne Kramer (1994), certaines connaissances sont nécessaires à l'établissement de nouvelles règles de création d'interfaces auditives, par exemple. Nous pouvons reconnaître d'ores et déjà que le lien existant entre les études "pionnières" d'ICAD (issues de Kramer, 1994) et la perception auditive est essentiellement de l'ordre de la logique et du bon sens.

Considérons à titre d'exemple l'article de Williams (1994). L'auteur présente une revue large et détaillée des études de perception et de cognition auditives, tant sur les théories que sur les méthodes et paradigmes employés, en focalisant davantage son propos sur l'Analyse des Scènes Auditives (Bregman, 1990). La conclusion pratique de cette revue aboutit finalement au constat que, sur la base des connaissances actuelles (que ce soit en 1994 ou à ce jour), nous ne pouvons prédire *de manière complète et exhaustive* comment un signal acoustique sera interprété par un auditeur.

L'auteur en déduit cependant quelques principes fondamentaux à respecter pour un design efficace d'interfaces auditives ou de sonification : prendre en compte les connaissances sur le masquage (temporel et fréquentiel) avant de créer un son ; vérifier que des changements appliqués à certains paramètres du son (fréquentiels, d'intensité) seront perceptibles ; considérer la variabilité entre sujets, et la différence d'interprétations possibles des stimuli due aux expériences passées qui diffèrent suivant les individus ; etc. Bien que ces quelques clés de lecture soient importantes à rappeler, surtout dans un domaine pluridisciplinaire, il apparaît tout de même que ces recommandations sont extrêmement génériques et d'utilité réduite en ce qui concerne l'application en tant que telle.

Le point de vue que nous avons adopté tout au long de cette thèse est donc de redéfinir

plus précisément les objectifs d'une étude sur les interfaces auditives. Sous un angle applicatif, le but est de décrire précisément les sons les plus efficaces à transmettre les fonctions voulues (d'urgence, par exemple). Le Chapitre 5 met en avant la nécessité de déterminer et délimiter les fonctions que doivent remplir les IHM sonores, et insiste sur l'importance de repenser les méthodes et les paradigmes par lesquels les hypothèses seront testées. Les Chapitres 6 et 7 abordent indirectement une question au coeur du débat de la recherche (française), à savoir les relations parfois houleuses entre questionnements "fondamentaux" et "appliqués". Bien que ce débat sorte totalement du cadre de ce travail de thèse (si tant est qu'il ait un intérêt ainsi formulé), nous soulignons dès à présent l'intérêt et la nécessité de tenter de *comprendre* certains des mécanismes perceptifs et cognitifs qui sous-tendent les notions de sons fonctionnels et sons d'urgence. C'est cette compréhension qui pourra par la suite donner lieu à des créations sonores fonctionnellement adéquates au contexte d'usage défini par le *designer* ou le constructeur et aux attentes de l'utilisateur.

Avant de présenter plus en détail les travaux existants portant sur la création d'interfaces auditives, nous avons donc inclus dans ce chapitre de brefs rappels concernant la perception et la cognition auditives (section 2.3 et section 2.4), ainsi que des exemples d'études plus précises portant sur la perception de sons environnementaux (section 2.5). Ces rappels ne sont bien évidemment pas exhaustifs (loin de là); ils ont uniquement pour but de situer les travaux d'ICAD dans une problématique plus globale que de simples applications informatiques.

2.3 Perception auditive

2.3.1 Théories

La perception est définie comme la réaction du sujet à une stimulation extérieure se manifestant par divers mécanismes qui nous permettent d'appréhender le monde qui nous entoure. La cognition se définit comme l'ensemble des activités mentales et des processus qui se rapportent à la connaissance. Deux approches de la théorie de la perception sont traditionnellement proposées (et opposées).

L'approche la plus communément développée considère que toute activité "mentale" repose sur des mécanismes de traitement de l'information. Le rôle de la cognition est d'*interpréter* les signaux issus des informations sensorielles (vue, toucher, goût, audition, odorat) afin de réagir à l'objet (par exemple). Les "représentations mentales" servent de

base à ces interprétations. C'est une approche du type traitement de l'information.

Rejetant cette théorie, Gibson (1966) a proposé une nouvelle approche *écologique* de la perception. Selon lui, la perception du monde est directe et immédiate, et ne passe pas par une représentation intermédiaire. Dans le cadre de ce modèle, Gibson définit ainsi les *affordances* (du verbe anglais *to afford*, offrir, permettre), qui décrivent les possibilités d'interactions entre l'Homme et son environnement. Les *affordances* sont, en d'autres termes, les propriétés de l'objet qui seront utiles à l'Homme (ou à l'animal). Elles permettent alors de percevoir *directement* ce que l'on peut faire avec un objet. Par exemple, selon la théorie de Gibson, une poignée de porte offre toutes les caractéristiques pour être saisie sans pour autant donner lieu à une représentation mentale de la poignée ou de ses propriétés). Bien d'autres courants de recherches peuvent cependant être identifiés en psychologie.

Une grande partie des travaux connexes à la littérature d'ICAD s'inscrit - parfois de loin - dans le cadre de la théorie gibsonienne : d'une part, parce que la notion d'*affordance* s'est imposée comme un principe dans le monde informatique travaillant sur les IHM, monde connexe à celui d'ICAD, et d'autre part, parce que, historiquement, la théorie gibsonienne a recentré les études en perception auditive davantage sur la source et son contexte (domaine de l'acoustique écologique, voir Neuhoff, 2004) que sur le son (domaine de la psychoacoustique).

Les travaux expérimentaux que nous avons menés relèvent davantage d'une approche de la perception du type traitement de l'information, et utilisent des méthodes comportementales (en particulier les méthodes psychophysiques). Ce choix est justifié au cours du document, dans chacun des chapitres expérimentaux. L'objectif final de cette thèse n'est pas de valider l'une ou l'autre théorie.

2.3.2 Méthodes

La psychologie cognitive utilise en effet préférentiellement l'expérimentation et la mesure des performances comportementales (dont les méthodes psychophysiques). La tâche expérimentale est alors définie par des consignes verbales (chez l'Homme). Il existe également des méthodes neuropsychologiques et de neuroimagerie ; nous ne les aborderons pas dans le cadre de cette thèse.

Classiquement, la psychophysique est la discipline qui étudie quantitativement les relations qui existent entre un stimulus physique et la perception qu'on en a. Plus généralement, dans le cadre de la psychologie cognitive, cette discipline cherche donc à étudier par une approche comportementale les mécanismes de traitement de l'information. La psychoa-

oustique est la branche de la psychophysique qui s'intéresse au lien entre la perception que nous avons d'un son et sa description physique.

Les méthodes mesurant les performances comportementales sont variées, et permettent de répondre à de nombreuses questions (pour une revue des problématiques, protocoles et modèles auditifs associés, voir Moore, 2003). Notons ici qu'une des bases de l'expérimentation en psychologie cognitive repose sur l'approche de la *chronométrie mentale*, selon laquelle la mesure du temps de réaction (TR) permet d'inférer sur le temps de traitement d'un stimulus donné. Cette approche comportementale sera traitée plus en détail dans le Chapitre 4.

2.4 Psychologie de la perception auditive

2.4.1 Contexte

Classiquement, les études sur la perception auditive ont utilisé des sons que la communauté ICAD classerait comme "abstrait", i.e. des sons n'évoquant pas de sens particulier : des sons purs, ou des sons complexes musicaux. Bien qu'un son musical soit relié directement à une cause - l'instrument qui joue le son - il est d'usage (dans la communauté ICAD) de le considérer dans la catégorie des sons abstraits (peut-être parce que les sons dits musicaux ont souvent été réalisés par des synthétiseurs dont le rendu était peu réaliste).

Comme nous le verrons plus loin, les interfaces auditives "abstraites", appelées les *earcons*, n'ont pas été créées et pensées dans un cadre théorique précis (contrairement aux interfaces auditives iconiques, les *icônes auditives*, qui s'appuient en grande part sur l'acoustique écologique). De plus, les études d'*earcons* ne comportent en général que peu de validations expérimentales perceptives des propositions de design. Cependant, ces études s'inscrivent implicitement dans un schéma constructiviste de traitement de l'information auditive : les *earcons* sont décrits suivant une liste de paramètres acoustiques que l'auditeur est supposé extraire et reconnaître afin d'agir de manière appropriée.

Nous allons donc présenter ci-après les différentes étapes de traitement de l'information auditive comme proposée par la psychologie expérimentale. Le but ici est simplement de rappeler les grandes étapes de traitement des sons qui nous aideront par la suite à mieux appréhender les études sur les *earcons*. Ce rappel est donc, de fait, délibérément bref. Il s'inspire de l'ouvrage de McAdams et Bigand (1994).

2.4.2 Étapes de traitement de l'information auditive

La Figure 2.1 résume les différentes étapes de traitement. La principale tâche du système auditif est de réussir à décoder l'information qui lui arrive sous forme d'ondes sonores, pour aboutir, *in fine*, à l'identification de l'objet, qui conduira l'auditeur à réagir de la manière appropriée.

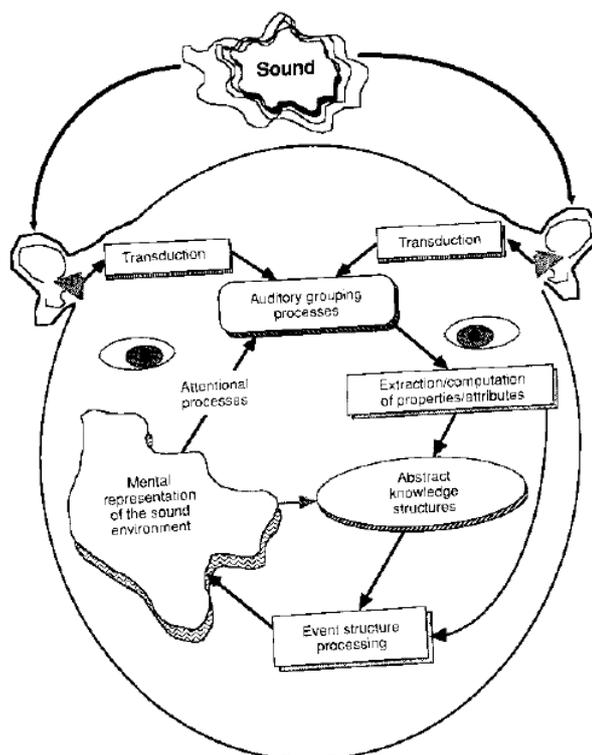


FIG. 2.1 – Étapes de traitement des informations auditives. D'après McAdams et Bigand, 1994.

La première étape de traitement est la transduction sensorielle du son : les signaux acoustiques sont encodés par l'oreille externe et moyenne, puis par la cochlée, au niveau de l'oreille interne.

Les étapes suivantes sont généralement admises comme modélisant des mécanismes plus centraux. Avant de reconnaître les sons, le système auditif doit différencier le mélange sonore qui s'offre à lui. Bregman (1990) a décrit ce phénomène comme la ségrégation de flux, appelée encore l'*Analyse de Scènes Auditives*. Schématiquement, Bregman distingue deux types d'analyse : l'analyse primitive, et l'analyse basée sur les schémas. L'analyse primitive se fonde sur des règles acoustiques générales, qui permettent une première séparation des

flux. Parmi ces règles, citons par exemple le fait que deux sources indépendantes n'ont que très peu de chances de commencer à émettre un son au même moment, ou à s'arrêter d'émettre au même moment. Par conséquent, si deux sons commencent exactement au même instant, le système auditif les considérera comme faisant partie d'un même son complexe. En plus de ce processus purement sensoriel, s'ajoute l'analyse basée sur les schémas, c'est-à-dire conditionnée par les connaissances antérieures et les attentes de l'auditeur. Par exemple, nous allons détecter et reconnaître immédiatement notre nom si quelqu'un nous appelle au milieu d'une salle d'attente bruyante. Enfin, Bregman différencie l'analyse simultanée (où les différents flux sont groupés au même instant) de l'analyse séquentielle (où les différents flux sont organisés suivant des contraintes temporelles).

A l'issue de cette séparation des sources, le système auditif est alors apte à extraire différents attributs de ces flux. Classiquement, les attributs auditifs sont la sonie, la hauteur, la durée, le timbre et la position spatiale. De nombreuses études se sont focalisées sur la psychophysique de l'un ou l'autre de ces attributs (pour une revue détaillée de l'étude de ces attributs ainsi que des modèles auditifs proposés pour expliquer leur perception, voir Moore, 2003 ; pour une étude approfondie du timbre, voir McAdams et coll., 1995). C'est également à ce moment que le système peut traiter la structure temporelle du son.

Les études des *earcons* considèrent essentiellement (même si cette considération est parfois implicite) ces étapes d'analyse des scènes auditives, d'extraction des attributs et d'analyse de la structure temporelle.

Les étapes suivantes concernent ce qui est souvent décrit comme un des buts principaux de notre système auditif, à savoir la reconnaissance des sources sonores. Cette étape interagit généralement avec les thèmes plus complexes de mémoire auditive et d'attention auditive. Toutefois, ces sujets sont encore peu consensuels dans la description et l'explication que nous pouvons en faire. Il n'existe toujours pas de modèle satisfaisant permettant d'expliquer comment le système auditif reconnaît des sources sonores différentes. Le thème de l'attention auditive sera traité un peu plus en détail dans le prochain chapitre, mais nous pouvons déjà remarquer que la définition elle-même de l'attention est parfois sujette à débat (Gallun, 2006).

Malgré la complexité de ces thèmes de reconnaissance, d'attention auditive et de mémoire auditive, il est probable qu'ils soient (ou qu'ils devraient être) en fait au centre des études sur les *earcons*. Toute création sonore passe nécessairement *a posteriori* par une phase d'identification, d'apprentissage, de mémorisation du nouveau son par les utilisateurs.

Ce point majeur de la reconnaissance et de l'apprentissage inévitable des sons abstraits a donc mené la communauté ICAD à s'interroger sur l'utilisation possible de sons déjà existants dans notre environnement, et par conséquent auxquels nous sommes quotidiennement confrontés.

2.5 Perception des sons environnementaux

Historiquement, la communauté ICAD s'est focalisée avec plus d'intérêt sur l'utilisation de sons environnementaux que sur celle de sons "abstraits" pour des applications pratiques et intuitives. Nous allons donc présenter en détail plusieurs études sur la perception des sons environnementaux qui seront nécessaires à une compréhension plus générale des propositions d'interfaces auditives de la littérature ICAD.

Nous allons voir au cours de cette section que l'étude des sons environnementaux a souvent été placée dans un cadre écologique (Gibson, 1966), en s'opposant ainsi en termes de théorie et de méthodes aux études menées en psychoacoustique. D'un point de vue bibliographique et historique, la distinction entre les études de sons environnementaux et celles des sons abstraits s'accompagne de ce fait d'une distinction théorique entre les approches du type traitement de l'information et écologique.

L'étude de sons environnementaux peut, bien entendu, se faire dans un tout autre cadre que celui de l'acoustique écologique. Citons à titre d'exemple l'étude de qualité sonore de Susini et coll. (2004) qui adopte une approche du type traitement de l'information et des méthodes psychophysiques pour étudier des sons de ventilateur. Plus récemment, les sons environnementaux ont été étudiés sous un angle nouveau, en imagerie cérébrale (Lewis et coll., 2005) ou en s'intéressant plus spécifiquement au temps de traitement de l'identification des sons grâce à des enregistrements électro-encéphalographiques (Murray et coll., 2006). Ces études sont éloignées de la problématique initiale de cette thèse, et ont donc été écartées de cette revue bibliographique. Elles seront cependant mentionnées et discutées au cours du Chapitre 7.

2.5.1 Définition des sons environnementaux

Un des constats communs à de nombreuses études d'ICAD et rappelé par Bregman (1994) oriente la problématique des interfaces auditives vers l'analyse et l'utilisation de sons plus réalistes. Il relève, en effet, que la plupart des études en perception auditive

se focalisent d'une part sur des sons très simples et qui ne portent pas, en général, de sens particulier (le si souvent cité son pur à 1 kHz), ou à l'autre extrême sur des stimuli complexes tels que la musique ou la voix humaine. Peu de recherches visent à comprendre la perception des sons "quotidiens", encore appelés sons *environnementaux*.

Vanderveer (1979) a proposé la première plusieurs critères de définition des sons environnementaux : 1) ils sont produits par des événements réels ; 2) de ce fait, ils sont porteurs de sens (celui de l'événement qui est la cause du son) ; 3) ils sont plus complexes que des "sons de laboratoire" tels des sons purs ; 4) ils ne font pas partie d'un système de communication comme la parole. De cette manière, les sons environnementaux sont définis par la négative : ils ne sont ni des stimuli "de laboratoire", ni de la voix, ni des sons musicaux.

Ballas et Howard (1987) n'ont gardé de cet ensemble de critères que les deux premiers, qui ne sont justement pas "exclusifs". Une définition centrée uniquement sur l'événement qui est à la cause du son et sur le sens de ce son s'inscrit *selon eux* dans une perspective écologique de la perception des sons environnementaux : la perception auditive a pour but de reconnaître et d'identifier des sources sonores (Schubert, 1974 cité dans Warren et Verbrugge, 1984).

Les sons environnementaux étudiés en général (Vanderveer, 1979 ; Ballas et Howard, 1987 ; Ballas, 1993) sont des sons quotidiens, regroupant des sons de l'environnement naturel (éléments naturels tels que l'eau, l'air, le vent, etc.), de l'environnement animal (vocalises animales) ou encore de l'environnement citadin (voiture, bruit de pas, ambiance de parc, etc.). Bien que les sons d'animaux fassent partie d'un système de communication et puissent s'apparenter d'un certain point de vue à la communication humaine, ils sont souvent inclus dans la catégories des sons environnementaux (peut-être parce que nous comprenons moins bien un cri de singe qu'une voix humaine).

Remarquons tout de même la variété, tant au niveau sémantique (i.e. regroupant des éléments de l'environnement naturel aussi bien que de l'environnement "civilisé", comme les voitures, par exemple) qu'acoustique de ces sons *environnementaux*. L'étude d'une catégorie "environnementale" en tant que telle pourrait être remise en cause, et il serait vraisemblablement plus intéressant de décomposer cette catégorie en sous parties moins hétéroclites (sons d'animaux, sons d'éléments naturels, etc.). Cependant, les études relatives à ces sons environnementaux ont tout de même mené à certaines conclusions intéressantes et ont été reprises maintes fois par la suite dans la littérature (notamment des interfaces auditives, voir section 2.6.3 page 32). C'est pourquoi, au moins dans cette partie bibliographique, nous traiterons des sons environnementaux au sens large et sans distinction.

2.5.2 Acoustique écologique

Parmi les études fondatrices d'ICAD qui se sont intéressées aux sons environnementaux, certaines ont tout particulièrement cherché à s'inscrire dans un cadre théorique avant de proposer un design effectif des interfaces auditives. Ce sont essentiellement celles de Gaver (1993) qui ont été par la suite maintes fois reprises et étendues.

En se basant sur la théorie de la perception écologique (Gibson, 1966), Gaver différencie l'écoute musicale de l'écoute quotidienne. L'écoute quotidienne est centrée sur la perception des événements à partir du son qu'ils produisent : les auditeurs perçoivent la source du son. A l'inverse, l'écoute musicale se fonde sur la perception de structures créées par les attributs du son lui-même : l'auditeur est alors attentif à la hauteur du son, sa sonie, sa durée, etc. La différence entre ces deux types d'écoute se fait au niveau de l'*expérience* que nous avons des sons, et non pas des sons en eux-mêmes.

Gaver s'intéresse uniquement à l'écoute quotidienne, centrée sur les événements et les actions qui produisent les sons (frottements, grattements, etc.) et développe ainsi ce qu'il appelle la psychologie de l'écoute quotidienne, d'après le cadre théorique gibsonien. Deux questions se posent alors : qu'est-ce que l'on écoute (Gaver, 1993a) - i.e. la source du son - et comment l'écoute-t-on (Gaver, 1993b) - i.e. quels sont les paramètres perceptifs et cognitifs qui ont du sens et dont nous nous servons pour identifier et discriminer des sons ? Ces problématiques seront traitées respectivement dans les sous-sections 2.5.3 et 2.5.4.

2.5.3 Écoute de la source du son

La grande diversité des sons environnementaux a amené de nombreux auteurs à proposer des théories spécifiques de traitement de cette information (comme souvent dans le domaine de l'audition, les nouvelles théories et hypothèses sont issues de la littérature en vision).

Le modèle de Gaver

Considérons tout d'abord la première question que pose Gaver (1993) : qu'écoute-t-on ? Comme nous venons de le voir, Gaver (1993a) souhaite proposer un nouveau modèle théorique de la classification des sons en fonction des *sources*.

Une définition de l'écoute quotidienne ne pouvant raisonnablement pas se baser sur une description des paramètres physiques uniquement, Gaver suggère de demander - tout simplement - à des sujets "ce qu'ils entendent". Cette approche s'inspire de l'étude de

Vanderveer (1979) qui a présenté à des sujets des enregistrements de sons environnementaux dans une tâche d'identification libre (décrire chaque son à l'aide d'une phrase). Le résultat principal - et souvent cité - de cette étude montre que les sujets identifient les sons en termes d'événements, d'objets et d'actions qui ont causés le son, et ne décrivent les attributs du son que lorsqu'ils ne peuvent pas l'identifier. L'étude de Gaver aboutit à peu près aux mêmes conclusions.

Nous pouvons remarquer ici que ce résultat ne valide pas réellement les hypothèses de l'acoustique écologique (i.e. perception *directe* du monde plutôt que construction d'un objet interne, comme décrit dans l'approche du type traitement de l'information). Il permet uniquement de confirmer qu'une des missions finales de notre système auditif est de reconnaître les objets sonores qui nous entourent.

L'analyse physique des sons combinée à l'interprétation des résultats de son expérience l'amène alors à poser les fondements d'un schéma décrivant les caractéristiques audibles d'un événement sonore. Son but, *in fine*, est d'organiser les caractéristiques de l'écoute quotidienne dans une forme relativement simple et générale.

Partant du constat pragmatique qu'il existe une quasi infinité de sons quotidiens, et qu'il ne serait pas pertinent de les classer en fonction de leurs contextes d'apparitions (Ballas et Howard, 1987) puisqu'un même son se retrouve dans divers contextes, Gaver propose alors une classification hiérarchique, tenant compte des dimensions et des caractéristiques des sons (en distinguant les sources dont les sons sont produits par des solides vibrants, celles dont les sons sont produits par des changements à la surface d'un liquide et celles dont les sons sont produits par des causes aérodynamiques dans l'atmosphère).

Cette notion d'écoute quotidienne et cette nouvelle classification sont à l'origine de la notion d'icônes auditives, type spécifique d'interface auditive qui sera décrit en détail plus loin (section 2.6.3).

Bien que cette nouvelle classification puisse apparaître séduisante à première vue, nous pouvons remarquer qu'elle n'a pas été validée dans la suite de la littérature (à notre connaissance). L'intérêt de cette étude réside davantage dans sa formulation nouvelle des problèmes posés par les interfaces auditives (constat valable en 1993) et dans les questionnements divers qui en sont issus.

Analyse grammaticale des sons environnementaux

En se fondant sur les principes de base de l'organisation visuelle, Ballas et Howard (1987) ont eux aussi proposé un nouveau modèle théorique pour appréhender la perception

des sons environnementaux.

Là encore, les hypothèses de départ sont très générales, et assimilent l'analyse des sons quotidiens à l'analyse grammaticale de la parole, en décrivant les règles de l'Analyse des Scènes Auditives (Bregman, 1990) comme des règles syntaxiques. Alors, d'après les auteurs, l'écoute des sons environnementaux devrait être soumise à deux processus de traitement, ascendant et descendant. Rappelons très brièvement que le processus ascendant est guidé par les stimuli, alors que le descendant est guidé par les connaissances (pour une discussion plus approfondie sur ce sujet, voir l'ouvrage de Bagot, 1996).

Il est envisageable que la trop grande généralité de ces hypothèses de départ puisse les rendre difficiles à valider ou à infirmer dans un cadre expérimental précis. En effet, la majorité des sons peuvent être traités suivant des processus ascendants et descendants, et il est, de plus, extrêmement difficile de différencier expérimentalement ces deux types de processus de manière claire.

Plusieurs conclusions utiles peuvent malgré tout être tirées de leurs résultats expérimentaux. Faisant suite aux résultats désormais "classiques" de Vanderveer (1979), les auteurs montrent que, dans la plupart des cas, les auditeurs identifient la source du son ainsi que l'action qui cause le son. Ils complètent également les travaux de Vanderveer (1979) et de Gaver (1993) en analysant plus finement les erreurs. Ils mettent ainsi à jour l'importance du contexte pour la reconnaissance du son. Cette explication semble assez intuitive : un son de voiture devrait être reconnu plus facilement si l'auditeur écoute avant un bruit ambiant urbain.

De plus, ils introduisent les notions d'*occurrence* d'un son, ainsi que celle de l'*homonymie* d'un son. Ainsi, un son plus commun (qu'un auditeur aura plus de chances d'avoir déjà entendu plusieurs fois) sera reconnu plus facilement et avec moins d'erreurs. En revanche, un son ayant de multiples homonymes (c'est-à-dire qu'un auditeur peut lui attribuer différentes causes possibles), surtout si rien ne l'informe sur le contexte d'apparition du son, aura plus de chances d'être mal, voire non, reconnu.

2.5.4 Identification des sons environnementaux

Une fois que les bases théoriques de l'écoute des sons environnementaux ont été posées (i.e. que la question "qu'écoute-t-on" a été traitée), Gaver (1993b) propose justement de s'intéresser à la manière précise dont nous écoutons les sons. A quelles propriétés acoustiques sommes-nous sensibles ? Quelle est l'importance des facteurs cognitifs ?

Influence des propriétés temporelles

Il reprend principalement et étend les travaux existants de Warren et Verbrugge (1984).

Warren et Verbrugge (1984) ont présenté un cas d'étude d'acoustique écologique, en prenant l'exemple de la perception auditive de sons de casse et de rebond. Partant du constat que les propriétés temporelles (variantes au cours du temps) sont capitales dans la reconnaissance de la parole ou de la musique, ils s'interrogent sur le rôle de ces propriétés dans l'identification de sons environnementaux bien précis : les sons de casse et de rebond. Leur hypothèse consiste à dire que le pattern *temporel* général de ces bruits est suffisant pour différencier la casse du rebond (les propriétés spectrales étant plus utiles à l'identification de la nature de l'objet en elle-même ; verre, bois, etc.).

Leur hypothèse est validée grâce à trois expériences de reconnaissance (choix forcé). La distinction entre un bruit de casse et un bruit de rebond est claire et sans équivoque, que ce soit avec des sons naturels enregistrés ou des sons synthétisés qui imitent les variations temporelles de la casse et du rebond, mais qui ont tous le même contenu spectral. Ils valident également que le premier bruit initial, lors de la casse, n'est pas responsable de la distinction perceptive entre casse et rebond ; ce constat est valable pour les sons naturels et pour les sons synthétisés. Ils valident ainsi - dans un cadre très restreint, certes - leur hypothèse de départ. Comme les auteurs le soulignent eux-mêmes, une tâche de catégorisation forcée à trois choix limite de fait la puissance des résultats, qui peuvent difficilement être transposables à d'autres études. Ce travail doit être considéré avant tout comme précurseur d'études de modélisation des mécanismes auditifs capables de détecter des invariants acoustiques.

Influence des propriétés spectrales

Alors que Warren et Verbrugge (1984) ont analysé les invariants transformationnels d'un son (i.e. l'information qui spécifie le type de changement en lui-même), Repp (1987) s'est davantage intéressé aux invariants structurels (les informations d'un son relatives au type de l'objet), en considérant comme cas d'étude l'applaudissement, et plus précisément le son de deux mains qui tapent. Tout comme Warren et Verbrugge (1984), l'étude est considérée sous un angle écologique, et comprend aussi bien l'étude acoustique de la production des sons que l'étude de la perception.

Il montre alors que les auditeurs sont capables de reconnaître leur propre applaudissement, reconnaissance due au rythme et à la vitesse du son, à sa sonie et à son contenu

spectral. En revanche, les auditeurs sont incapables d'inférer une information sur la taille des mains (donc, par voie de conséquence, sur le sexe de la personne qui applaudit, les femmes ayant en moyenne des mains plus petites que celles des hommes) d'après le son qu'ils entendent. Les auteurs constatent uniquement la force de certains préjugés : les applaudissements rapides et aigus sont catégorisés comme féminins, les graves et lents sont associés à des applaudissements d'homme (alors même que l'analyse acoustique des applaudissements ne montre aucune différence significative entre les applaudissements des hommes et ceux des femmes).

Ces deux cas d'études, bien que très limités dans leur généralisation possible, montrent l'influence précise de propriétés temporelles et spectrales sur l'identification de sons environnementaux, et pourront servir de point de départ à la création d'interfaces auditives.

Rôle des facteurs cognitifs

Toutefois, il est prévisible que des facteurs autres qu'acoustiques entrent en jeu dans la reconnaissance et l'identification de sons environnementaux.

Ballas (1993) étudie plus en avant ces quelques pistes d'interprétation, et développe plus précisément les notions d'occurrence et d'homonymie des sons environnementaux évoqués précédemment (Ballas et Howard, 1987). Il montre ainsi l'importance des facteurs cognitifs dans l'identification des sons environnementaux.

Nous retiendrons essentiellement de cette étude le lien établi entre l'incertitude causale d'un son (i.e. le fait d'identifier plusieurs sources possibles pour un son donné) à sa fréquence écologique. La fréquence écologique étant définie comme le nombre d'occurrences d'un son dans notre environnement, Ballas (1993) montre que plus un son est fréquent dans notre environnement quotidien, moins nous avons la possibilité de le confondre avec d'autres.

Fait intéressant à souligner, cette règle n'est pas respectée dans le cas des signaux d'alerte et des alarmes. Leur incertitude causale est faible, c'est-à-dire qu'il n'y a que peu de chances de les confondre avec d'autres, alors que leur fréquence écologique est très faible (comparativement aux autres sons environnementaux, nous n'entendons que très peu d'alarmes).

Ce résultat est interprété comme un effet de la *typicalité* du son (pour un exposé complet de la théorie de la catégorisation et de la notion de typicalité, voir Rosch, 1978) : une alarme est bien reconnue même si on ne l'entend que peu parce qu'elle est typique de sa catégorie (sons d'alerte) et qu'elle est de ce fait bien identifiable et reconnaissable. D'autres

raisons (d'ordre purement acoustique, par exemple) pourraient être invoquées ; cependant, les données expérimentales recueillies ne permettent pas de conclure de manière sûre et fiable sur l'influence comparée des différents facteurs (acoustiques, écologiques, perceptifs et cognitifs).

L'auteur conclut sur l'intérêt de tenir compte de ces différents facteurs, et propose ainsi une théorie "hybride" de l'identification de sons environnementaux. Il met ainsi l'accent sur la variété et la diversité des sons environnementaux qui justifie et appuie selon lui l'importance de la théorie hybride.

Toutefois, comme évoqué par Ballas et Howard (1987), cette volonté de regrouper les sons "n'étant pas étudiés par ailleurs" est critiquable du fait justement d'une diversité trop grande des sons étudiés. En effet, les sons musicaux ou de parole ont des points communs entre eux, que ce soit à un niveau acoustique, sémantique ou cognitif. La proposition de regrouper des sons produits par des événements réels est intéressante du point de vue de l'acoustique écologique, mais n'a pas permis jusqu'ici de conclure, à notre connaissance, sur les facteurs responsables de l'identification des sons environnementaux. La multitude et la dispersion des sons étudiés n'aide peut-être pas à répondre à cette question. Bien qu'il soit probable que plusieurs facteurs soient impliqués (acoustiques, cognitifs), il serait intéressant de savoir dans quelle mesure et comment.

Dans une étude récente portant elle aussi sur les facteurs spectro-temporels dans l'identification des sons environnementaux, Gygi et coll. (2004) soulèvent ce point. De la même manière qu'avec des sons de parole où les connaissances antérieures du sujet aident à la détection et à la reconnaissance, les auteurs supposent qu'il devrait en être de même pour les sons quotidiens. Cependant, ils concluent, eux aussi, sur le manque de connaissances actuelles relatives à l'influence comparée de processus ascendant et descendant dans l'identification de sons environnementaux.

2.6 Interfaces auditives

Après cette présentation de la psychologie de l'audition et ces exemples d'études de sons environnementaux, nous pouvons maintenant aborder le coeur du sujet, à savoir la création d'interfaces auditives.

2.6.1 Introduction

Telles que définies par Kramer (1994), les Interfaces Auditives permettent de coder une fonction grâce à un son non verbal. Cette *fonction* désigne ce que le concepteur souhaite transmettre comme message par l'intermédiaire du son. Lorsqu'un utilisateur se trouve confronté à cette interface auditive, il doit déduire de ce son un certain nombre d'informations qui l'aideront à discriminer ce son des autres et à réagir de manière adéquate. C'est dans ce sens que nous pouvons considérer l'étude des Interfaces Auditives comme un exemple de psychologie cognitive appliquée : à partir de l'étude du comportement d'un auditeur, nous pouvons inférer des représentations et processus mentaux qui lui ont permis de différencier le stimulus entendu d'un autre, de lui associer un sens bien précis et d'y réagir de manière adéquate.

A la différence des études de perception et cognition auditives plus classiques, s'ajoute ici la contrainte de la *création* des sons. Il ne s'agit pas uniquement de comprendre comment nous acquérons des connaissances sur le monde qui nous entoure par l'intermédiaire de la perception de stimuli auditifs (par exemple). L'information finale à laquelle devra avoir accès l'auditeur est définie *a priori* par le concepteur des interfaces auditives ; à partir de la définition de cette fonction, il cherche alors à concevoir le son qui la transmettra au mieux. A la démarche d'*analyse* du son faite traditionnellement, s'ajoute donc une étape de *conception*, guidée par les attentes fonctionnelles associées au son.

Afin de transmettre une information à l'aide d'un son non verbal, deux grands types d'interfaces auditives ont été distingués dans la littérature ICAD : les *earcons*, sons abstraits, et les *icônes auditives*, sons représentatifs de notre environnement quotidien.

2.6.2 Earcons

Les *earcons*, analogie auditive des icônes visuelles

Définition La notion d'*earcons* a été tout d'abord introduite par Blattner et coll. (1989). Ce sont des sons abstraits, synthétiques, des sortes de motifs musicaux qui peuvent s'apparenter à de courtes mélodies. Ces motifs sont utilisés pour véhiculer une information et créer ainsi des messages sonores non verbaux. Il n'existe aucun lien intuitif entre le son et le message qu'il est censé véhiculer ; le sens du son doit être appris. Le son entendu lors de l'allumage d'un ordinateur est un exemple d'*earcon* : c'est une mélodie, sans lien avec la fonction qu'elle indique, à savoir le bon fonctionnement de l'ordinateur.

Par analogie avec le système visuel et les nombreuses IHM visuelles (des interfaces

informatiques, par exemple), Blattner et coll. (1989) proposent de construire une grammaire des sons abstraits. Ainsi, ils décrivent le nouveau langage des sons abstraits, d'après une sorte de syntaxe, où le sens du son abstrait devrait se construire à partir de différentes structures qu'il est possible d'agencer et/ou de transformer.

L'objectif de cette proposition est d'indiquer *comment* créer une interface auditive, et non de valider si ces sons seront effectivement compris par un utilisateur potentiel (et utilisés de ce fait à bon escient), ni même de comprendre pourquoi.

Analyse des icônes visuelles Les auteurs analysent alors en détails la construction des icônes visuelles, afin de calquer les principes de création de ces icônes à leurs earcons. Les icônes visuelles peuvent être classées en trois sous-groupes : les icônes représentatives (dessin ou images d'objets familiers, comme le dessin d'une corbeille à papier sur les ordinateurs Apple Macintosh, par exemple), les icônes abstraites (forme géométrique servant à dépeindre un objet ou une opération qu'il serait difficile de décrire à l'aide d'un dessin, comme un sens interdit dans le code de la route) et les icônes semi-abstraites (image représentative très simplifiée, comme certains dessins servant à représenter des appareils pouvant atteindre des hauts niveaux de pression, par exemple).

Quelle que soit l'icône considérée, elle peut être décrite par une structure générale : un élément unitaire, indécomposable, et un composé de ces différents éléments.

Description des earcons Bien que les earcons et les icônes visuelles fassent appel, par définition, à des sens différents, les contraintes de design et de communication sont similaires dans les deux cas. Les earcons sont alors décrits par analogie avec les icônes visuelles.

La première définition de l'earcon est en fait un peu plus générale que la définition retenue à ce jour et présentée ci-dessus : par analogie avec les icônes visuelles, elle comprend en effet trois types d'earcons, les représentatifs, les abstraits et les semi-abstraites. Toutefois, les auteurs ne s'intéressent qu'aux "earcons abstraits", en mentionnant uniquement les travaux de Gaver portant sur les "earcons représentatifs", appelés *icônes auditives*. Par habitude et par évolution du langage, les earcons font maintenant référence à ce qui a été appelé initialement "earcons abstraits", et les icônes auditives sont l'appellation commune de ce qui est désigné par Blattner et coll. (1989) comme "earcons représentatifs".

Les earcons (abstrait) sont alors définis de la même manière que les icônes visuelles, où les éléments unitaires sont décrits comme une note avec une hauteur donnée (ou un groupe de notes), pouvant alors s'assembler en motifs, qui forment donc des séquences de

notes de hauteurs différentes, aboutissant enfin, si nécessaire, à ce qui est appelé earcons composés, qui sont proposés pour transmettre des messages complexes.

Afin de créer des earcons complexes, cinq paramètres sont proposés. Le rythme est présenté comme étant la caractéristique dominante d'un motif donné, suivi par la hauteur, et enfin le timbre, le registre (au sens musical du terme, i.e. un changement d'octave d'un même motif) et la dynamique (ici, des changements d'intensité). Cette proposition tout comme la hiérarchie supposée des paramètres ne sont pas justifiées ni discutées plus en détails, d'un point de vue de la perception auditive, par exemple.

La dernière recommandation pratique (et toujours aussi peu justifiée) concerne la longueur d'un earcon : au sein d'un même motif, il faut utiliser de deux à quatre notes.

Les earcons "hiérarchiques" sont ceux qui ont été majoritairement étudiés par la suite. Chaque earcon est un noeud d'un arbre, et hérite ainsi des propriétés des earcons du niveau supérieur. Au n -ième niveau, l'earcon est une séquence de n éléments. Chaque niveau de l'arbre permet de modifier un paramètre sonore (un arbre ne peut donc avoir *au maximum* que cinq niveaux, chacun correspondant à un des cinq paramètres décrits di-dessus).

La Figure 2.2 montre un exemple introduit par les auteurs, et adapté par la suite par Brewster et coll. (1994). L'exemple est celui d'une interface informatique. Le premier niveau est un son représentant une erreur. C'est une séquence rythmique mais "sans hauteur" (i.e. des clicks, notés *unpitched sound* sur la figure). Au deuxième niveau, la structure rythmique est conservée, pour indiquer qu'il s'agit toujours d'une erreur, mais un second motif a été ajouté, joué avec une hauteur donnée cette fois (utilisation d'une onde sinusoïdale, noté *sine* sur la figure), pour indiquer le type d'erreur (erreur d'exécution). Notons que les auteurs considèrent le son pur comme un son "sans timbre"; cette affirmation pourrait être sujette à débat. Le troisième niveau est caractérisé par l'ajout d'un motif avec un timbre particulier (ici, une onde en dents de scie, noté *triangle* sur la figure).

Mémoire et apprentissage des earcons Ces questions de validation expérimentale et de compréhension des sons sont en fait abordées indirectement par le biais des problématiques d'apprentissage et de mémoire des earcons. Les auteurs précisent bien que le lien entre un earcon avec le message qu'il est censé véhiculer doit être appris. La définition et la création d'earcons comme ils la proposent est alors argumentée sous l'angle d'un nouveau type de communication : puisque les interfaces informatiques nécessitent un nouveau langage utilisant des sons non verbaux, il est alors intéressant de mettre en place une sorte de grammaire formelle pour définir leur construction.

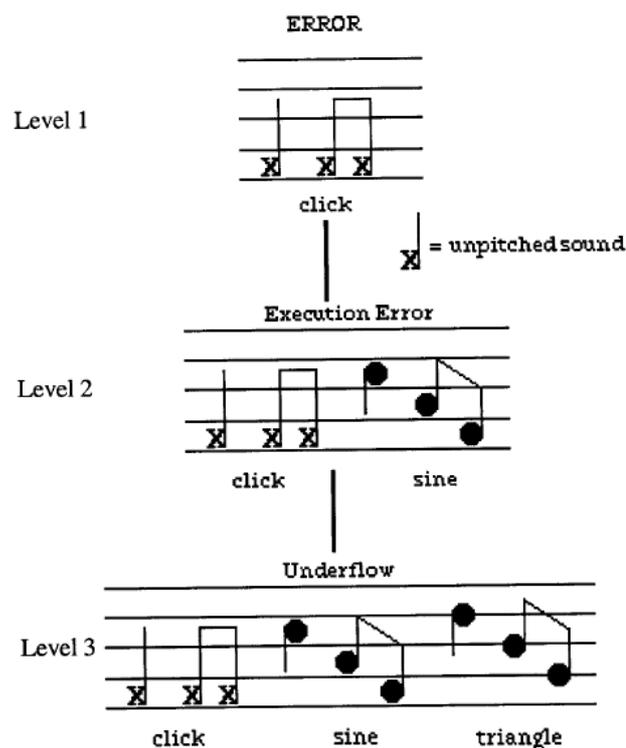


FIG. 2.2 – Earcons hiérarchiques caractérisant différents types d’erreurs d’une interface informatique. Plus le type d’erreur est détaillé et précis, plus l’earcon devient complexe (i.e. décrit par de plus en plus de paramètres sonores). Le premier niveau est caractérisé par un rythme, le deuxième par une hauteur, et le troisième par un timbre. D’après Brewster et coll., 1994

Certes ; mais nous pouvons tout de même nous demander, en poussant le raisonnement à l’extrême, quel est l’avantage de *ce* type de création plutôt qu’un autre. Puisqu’il est nécessaire de passer par une phase d’apprentissage, n’importe quel son (ou presque) pourrait être implémenté dans les interfaces informatiques, et il serait finalement appris par les usagers.

Validations expérimentales

Premières recommandations fonctionnelles Brewster et coll. (1994) ont alors cherché à savoir, par le biais d’une évaluation expérimentale, si les earcons étaient effectivement un moyen efficace de communiquer des informations. La validation expérimentale présentée est une validation que nous pouvons décrire comme ergonomique : elle vise à valider les

propositions concrètes de design de Blattner et coll. (1989), en utilisant les méthodes issues de la psychoacoustique comme un "outil". Il ne s'agit en aucune manière de tenter de comprendre d'un point de vue perceptif et cognitif les résultats obtenus. Les questions sont posées en termes de mémorisation et de reconnaissance des earcons complexes dans un contexte d'application informatique.

La principale expérience menée par Brewster et coll. (1994) a pour but de comparer des "mélodies sonores non structurées" à des earcons, et d'identifier les timbres les plus aptes à véhiculer une information. Les auteurs reprennent l'idée d'earcons hiérarchiques proposée initialement par Blattner et coll. (1989), et font varier les earcons suivant le rythme, la hauteur et le timbre.

Ils comparent alors trois ensembles de sons : des sons appelés "musicaux", qui sont en fait des sons de synthèse recréant les sonorités de différents instruments de musique ; des sons appelés "simples", qui regroupent des sons purs ou en dents de scie par exemple ; des sons de "contrôle", qui seront des sons sans information rythmique (voir la suite), mais de timbres similaires aux sons musicaux du premier groupe.

Trois groupes de sujets ont été testés pour chacun de ces trois groupes de sons. Ils étaient tout d'abord soumis à un apprentissage, à l'issue duquel les sujets devaient réussir à associer un son à une icône visuelle présente à l'écran (icône du type "écrire", "dessiner", etc.), et devaient connaître les relations de hiérarchie entre chaque icône. Ces connaissances étaient ensuite vérifiées, à l'aide des sons déjà entendus mais aussi à l'aide de sons nouveaux, que les sujets devaient associer au bon type et à la bonne famille d'icône d'après les principes issus de la phase d'apprentissage. Ces deux étapes (apprentissage et vérification) sont appelées la Phase I.

La Phase II est similaire à cette Phase I, mais concerne cette fois les earcons correspondant aux "menus informatiques" (par exemple "ouvrir", "créer", "sauver", "copier", etc.) et non plus aux icônes visuelles.

La Phase III est une redite de la Phase I, afin de voir si les sujets se souviennent des earcons appris précédemment, même après avoir été soumis à un autre corpus de sons.

La Phase IV combine les Phases I et II : deux sons, l'un issu d'un des deux ensembles appris précédemment et l'autre de l'autre ensemble, sont joués dans un ordre aléatoire ; les sujets doivent donner le sens de la séquence totale, composée alors de deux earcons différents.

Les résultats montrent que les timbres musicaux sont reconnus plus facilement que des sons "simples" et qu'un rythme "bien choisi" aide à la reconnaissance (il semble que d'après

les auteurs, un rythme "bien choisi" signifie suffisamment distinct des autres rythmes et facilement reconnaissables). En revanche, la hauteur ne semble pas aider les sujets à discriminer et reconnaître deux earcons différents (sauf éventuellement pour les sujets musiciens). Enfin, les sujets mémorisent correctement les earcons d'une phase à l'autre.

Les auteurs concluent alors sur quelques indications à respecter pour créer des earcons : utiliser des timbres instrumentaux, utiliser des changements de hauteur suffisamment grands (2 ou 3 octaves), et créer des rythmes bien distincts.

Une réelle validation fonctionnelle ? La conclusion de cette étude et les recommandations qui en découlent nous semblent toutefois un peu hâtives. En effet, Brewster et coll. (1994) valident ainsi - selon eux - l'intérêt des earcons musicaux comme moyen de communiquer une information. Le fait de mieux reconnaître des sons musicaux facilement identifiables (piano, marimba) de sons purs difficilement associables à une source pourrait s'expliquer assez simplement : les sujets peuvent nommer les sons musicaux, et les mémorisent de ce fait plus facilement.

De plus, en faisant varier des sons suivant un faible nombre de paramètres, et surtout très distincts (timbre, rythme), il était prévisible d'observer un effet de ces paramètres sur la réponse des sujets.

Enfin, l'effet quasiment non significatif de la hauteur s'explique assez bien aussi : nous reconnaissons difficilement des hauteurs absolues en tant que telles.

Pour finir, cette étude comporte une validation expérimentale centrée sur la mémorisation et la reconnaissance des earcons, mais n'étudie pas l'adéquation entre un earcon donné et le message que l'on souhaite transmettre. Quels sons pour quelles fonctions ? La question est toujours ouverte.

Les études ergonomiques plus récentes s'intéressant aux earcons ne répondent pas non plus complètement à cette question. Le cadre est donné en général par des interfaces visuelles où l'utilisateur doit agir en faisant dérouler des menus, en cliquant sur des boutons, etc. Beaudoin-Lafon et Conversy (1996) et Brewster (1998) ont montré par exemple que des earcons accompagnant ces interfaces visuelles permettaient de réduire le taux d'erreur, le temps nécessaire à accomplir la tâche et la charge mentale. Dans le même type d'études, Cockburn et Brewster (2005) ont montré qu'un feedback à l'aide d'un son non verbal aidait l'utilisateur dans une tâche psychomotrice (acquisition de petites cibles à l'écran). Dans une optique plus proche de la nôtre, i.e. donner des recommandations fonctionnelles pour la création de sons, Susini et coll. (2003) ont étudié les sons existants dans diverses interfaces

informatiques, et ont déduit quelques principes de design sonore à respecter à partir de ces sons existants. Cependant, les sons créés à partir de ces recommandations fonctionnelles n'ont pas pu être validés expérimentalement.

Toutefois, malgré ces protocoles expérimentaux de plus en plus rigoureux, et ces premières pistes de design sonore, la question initiale est toujours laissée sans réponse : quel son transmet le mieux la fonction pour laquelle il a été créé ?

2.6.3 Icônes auditives

Définition et description

Plutôt que de chercher à représenter une information en spécifiant des attributs sonores, Gaver (1989) a proposé de coder des informations grâce aux événements sonores en eux-mêmes. L'écoute quotidienne décrite plus haut a permis de donner des bases différentes à la création de nouvelles interfaces auditives : les *icônes auditives*. Ce sont des sons de tous les jours, resynthétisés et ainsi utilisés afin de transmettre un message sur certains événements des systèmes d'information, par analogie avec des sons issus de notre environnement quotidien qui véhiculent une information sur les événements qui les créent.

L'exemple le plus connu d'icône auditive est extrait des interfaces informatiques : lors de la suppression d'un fichier ou d'un dossier, un son accompagne l'action (i.e. faire glisser l'icône du fichier dans celle qui représente la corbeille). Ce son ressemble à celui d'un "vrai" papier jeté dans une "vraie" corbeille.

Le but des icônes auditives, complément auditif d'événements visuels plus "classiques", est de véhiculer une information riche à propos d'événements du système informatique, et d'être compris sans apprentissage préalable, pour écouter notre ordinateur de la même façon dont nous écoutons le monde environnant. De la même manière que certaines icônes visuelles liées à des objets quotidiens ont été conçues pour des interfaces informatiques (exemple de la corbeille), Gaver (1989) a alors proposé d'appliquer la même stratégie aux interfaces auditives.

A première vue, cette nouvelle stratégie de création d'interfaces auditives est beaucoup plus intuitive que pouvait l'être les earcons : il suffit de comparer le son correspondant à la suppression d'un fichier (cf. exemple ci-dessus) avec un earcon du type "bip". D'après l'approche écologique revue par Gaver (1993a ; 1993b), nous écoutons les sources d'un son et non pas ses attributs sonores en tant que tels. Il est donc plus intéressant d'utiliser des sons issus de notre environnement quotidien pour créer des interfaces auditives.

Ainsi, la phase d'apprentissage des sons et de leur association à une fonction qui était nécessaire avec les earcons devrait être inutile pour les icônes auditives.

Un autre argument souvent cité en faveur des icônes auditives vient de la richesse potentielle de l'information qu'il est possible de transmettre. D'après Gaver (1989), il est possible, à partir de l'écoute d'un son, d'inférer ses propriétés géométriques (taille, ...), structurelles (matériau) et l'interaction mise en jeu (par exemple, la force avec laquelle on tape sur une barre). De ce fait, une icône auditive pourrait alors véhiculer une information multidimensionnelle.

Nous pouvons toutefois modérer cette affirmation selon laquelle des auditeurs pourraient associer des propriétés géométriques à la simple écoute d'un son. Houix (2003) a en effet montré qu'il n'existe pas forcément de perception directe des propriétés géométriques des sources. En d'autres termes, nous pouvons nous tromper en estimant les dimensions réelles d'un objet juste à l'écoute du son produit qu'il produit après excitation (taper sur une barre, par exemple). Ce résultat a été étayé par ceux de Giordano et McAdams (2006), qui montrent que les auditeurs se trompent lorsqu'il s'agit d'identifier le matériau d'une plaque. En effet, ils associent une petite plaque à du verre ou du bois, et une grande à du métal ou du Plexiglas. Les auteurs interprètent ce résultat en expliquant que les auditeurs n'ont pas l'habitude dans le quotidien de manipuler des grandes plaques de verre ; nos attentes sont telles que ce qui est petit est en verre, et ce qui est grand est en métal. Cependant, cette erreur est surtout valable pour un jugement en absolu. En relatif, Lakatos et coll. (1997) montrent qu'il est possible de distinguer à l'oreille des sections de barres.

Exemple du SonicFinder

Gaver a alors exploité ces idées d'acoustique écologique et d'information multidimensionnelle contenue dans un son quotidien pour créer le SonicFinder. C'est une interface développée pour la société Apple ; elle est une extension du Finder, application existante qui permet d'organiser, de créer ou de supprimer des fichiers sur un Macintosh.

Elle offre des représentations auditives de certains objets et actions de l'interface informatique. Ainsi, de nombreuses actions sont accompagnées de sons : sélectionner un fichier, le faire glisser, le copier, le supprimer, ouvrir et fermer des dossiers, etc. Les fichiers sont associés à un son de bois, les applications à un son de métal et les dossiers à un son de papier. De plus, la plupart de ces sons sont paramétrables, et peuvent ainsi donner une information supplémentaire à l'action basique (copier, déplacer, ...). Par exemple, plus un objet est grand, plus le son associé est grave. L'exemple le plus connu et le plus détaillé du

SonicFinder est celui du fichier jeté à la corbeille, cité plus haut. La Figure 2.3 décrit, pour chaque étape de l'action (sélectionner un fichier, le déplacer, l'approcher de la corbeille et le supprimer), le son qui lui est associé.

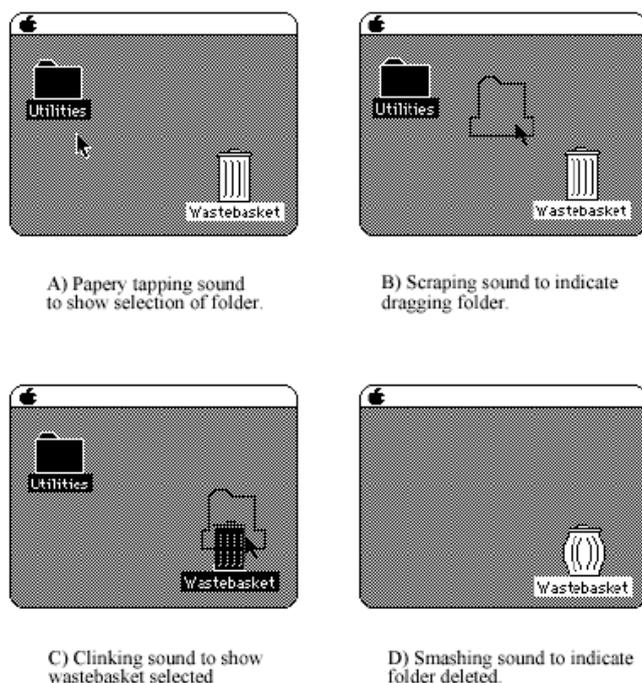


FIG. 2.3 – Exemple de la suppression d'un fichier du SonicFinder de Gaver, 1989. Figure extraite de Brewster, 2003

Cependant, un des problèmes majeurs posés par ce type d'interfaces auditives représentatives vient de l'abstraction de certains objets et actions du monde informatique qui n'ont pas, de fait, de représentations simples et évidentes dans le monde réel. L'exemple de la copie de fichier est cité par Gaver : il suggère d'associer un son de liquide qui coule pour indiquer cette action. Par extension à tous les autres exemples d'objets qui n'ont pas de représentation réelle claire, Gaver conseille alors d'utiliser ce type d'"effets sonores". Nous pouvons constater ici que, si l'effet sonore est mal choisi, l'icône auditive devient alors une interface abstraite, plus que représentative, et les avantages potentiels des icônes auditives sont perdus.

De plus, nous pouvons également nous interroger sur l'ambiguïté potentielle de certaines icônes auditives sorties d'un contexte naturel et placé dans un tout autre contexte.

Par exemple, une sonnerie de téléphone qui imiterait un cri d'alerte animale (e.g. un chien) pourrait laisser croire à la présence d'un vrai chien. Il faudrait alors réapprendre à associer des sons connus à d'autres types de fonctions ; là encore, l'icône auditive perdrait en intuitivité, et donc en intérêt.

Au final, Apple n'a jamais commercialisé le SonicFinder, essentiellement pour des raisons pratiques d'implémentation (taille de l'application trop importante), mais aussi parce que quelques utilisateurs (une faible proportion) ne trouvaient pas d'intérêt à cette nouvelle application, ou pire, la trouvait bruyante et agaçante (Buxton et coll., 1994).

Malgré ces quelques réticences, et bien que la validation expérimentale de ces icônes auditives ne soit pas présentée en tant que telle, le SonicFinder est souvent évoqué comme un exemple d'application d'icônes auditive efficace et réussi. De nombreuses autres interfaces informatiques utilisant des icônes auditives ont ensuite été développées. Buxton et coll. (1994) en donne un résumé assez clair et concis.

Le SonicFinder, comme les autres exemples d'application (voir Buxton et coll., 1994), n'a pas été testé expérimentalement de manière approfondie. Il semble que le fait d'utiliser des sons environnementaux et de s'appuyer sur un cadre théorique déjà développé par ailleurs soit une justification suffisante pour la plupart des auteurs pour ne pas avoir à tester plus en avant l'adéquation entre les sons qu'ils proposent et leurs fonctions associées. Toutefois, ne serait-ce que pour les "effets sonores" (cités plus haut), une validation au moins fonctionnelle s'avérerait nécessaire, puisque le lien entre le son et la fonction n'est plus direct.

2.6.4 Lien entre les Earcons et les Icônes Auditives

Préambule

A travers ces exemples, il apparaît alors clairement que la question de l'association entre un son et un événement est tout aussi fondamentale pour les icônes auditives que pour les earcons, malgré la volonté de se placer dans un cadre théorique différent.

A l'origine de la notion d'icône auditive, deux éléments s'imposent : l'utilisation de sons environnementaux paramétrables suivant leurs dimensions physiques, et l'emploi d'une sorte de cartographie intuitive entre le son et les événements du système d'information auxquels ils correspondent.

Il faut alors distinguer le type de son employé du type de lien (arbitraire ou intuitif) recherché. Il serait en effet possible d'utiliser un son environnemental pour une fonction

qui n'aurait aucun rapport avec la source du son ; mais ce ne serait alors peut-être plus une icône auditive.

Afin de formaliser ces différentes idées, nous allons détailler les différents liens, ou *mappings* qui peuvent se créer entre un son et sa fonction.

Types d'association entre son et fonction

Les différentes manières d'associer un son à une fonction reflètent en fait le degré de ressemblance et de proximité entre un son et ce qu'il est censé représenter dans le système d'information. Gaver (1989) distingue alors trois différents types de mapping : symbolique, métaphorique et iconique. Ces notions sont adaptées de recherche en sémiotique - science des signes (pour un ouvrage complet de ce domaine, voir Peirce, 1932).

Un mapping symbolique est arbitraire ; le sens donné au son l'est par rapport à une convention établie ou doit être appris. Un earcon est créé suivant un mapping symbolique. A l'autre extrême, un mapping iconique est basé sur des relations causales : il y a un lien physique de cause à effet entre le son et l'objet qu'il représente. C'est l'icône auditive. Entre les deux, le mapping métaphorique utilise des liens de similarité entre le son et ce qu'il doit représenter. Il pourrait être décrit comme une sorte de mapping arbitraire avec des contraintes d'ordre iconique. Suivant les cas, un earcon, comme une icône auditive peut suivre un mapping métaphorique.

Gaver précise que la définition de lien causal physique de la relation iconique n'implique pas forcément un enregistrement réel des sons pour l'utilisation des icônes auditives. Une icône auditive peut être aussi bien un enregistrement sonore d'un son environnemental qu'un son synthétisé imitant et caricaturant un événement quotidien. Ce qui importe est l'iconicité de la relation en elle-même : il doit exister un lien causal entre le son et sa fonction.

La Figure 2.4 montre un exemple (extrait de Gaver, 1989) de ces trois types de mapping pour des interfaces visuelles et auditives, dans le cas d'une suppression de fichier. Le lien symbolique est totalement arbitraire : l'interface auditive est un simple bip. Le lien iconique reprend le son d'un papier froissé et jeté à la corbeille : que le son soit enregistré ou resynthétisé, il existe bien une relation causale entre le "scrountch" de l'interface auditive (indiqué "crash" sur la Figure 2.4) et le fait de supprimer un document. Le lien métaphorique est une sorte d'analogie de la suppression dans le monde réel : l'interface auditive est un earcon qui diminue progressivement en amplitude, pour rappeler la notion de suppression. La diminution aurait alors très bien pu se faire suivant d'autres dimensions

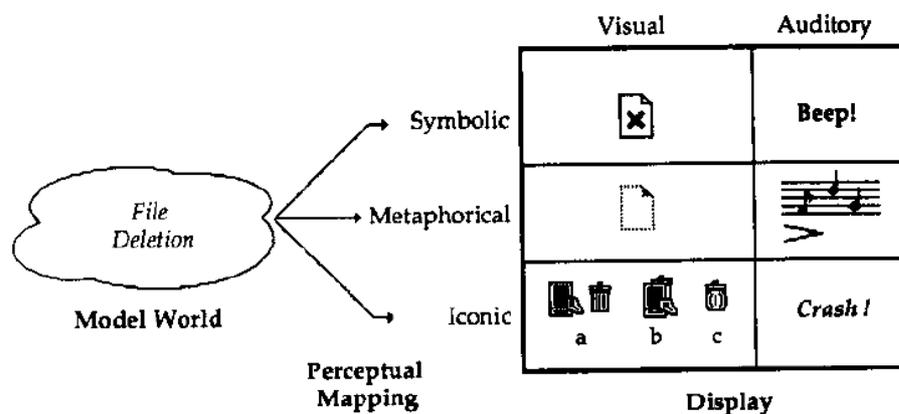


FIG. 2.4 – Exemple de la suppression d'un fichier. Trois types d'interfaces auditives et visuelles sont montrées, correspondant aux mappings iconiques, métaphoriques et symboliques entre un événement et un son. D'après Gaver, 1989

du son, la hauteur par exemple.

La hiérarchie entre relation iconique, métaphorique et arbitraire se fait donc suivant le degré d'abstraction. De ce fait, cette hiérarchie s'exprime également selon un apprentissage des sons. Plus le lien entre le son et sa fonction est abstrait, plus l'apprentissage est long. A l'inverse, une relation iconique ne nécessite a priori aucun apprentissage préalable, les règles sont déjà connues.

Au final, la dichotomie entre earcons et icônes auditives n'est plus si claire, et il serait sûrement plus juste de décrire un continuum entre earcon d'un côté et icône auditive de l'autre, et entre les deux, des sons qui imitent des sons naturels, des sons métaphoriques, des sons instrumentaux, etc.

Nous pouvons également noter que cette distinction entre earcons et icônes auditives n'a peut-être pas grand intérêt. En effet, ces deux types d'interfaces auditives sont différenciés d'une part selon la nature du son utilisé (environnemental ou abstrait) et d'autre part selon le type de lien voulu entre le son et la fonction (iconique, métaphorique ou arbitraire). L'amalgame entre ces deux dimensions - le son en lui-même et le rapport du son à la fonction - amène généralement à des définitions floues. Par exemple : un avertisseur sonore (klaxon) est-il un earcon, puisque le son du klaxon est un son abstrait et qu'il n'existe pas *a priori* de lien entre ce son et sa fonction, ou une icône auditive, puisque ce son fait partie de notre quotidien et que, de fait, nous savons précisément ce qu'il signale et ce à quoi il fait référence? Dans la suite de nos travaux expérimentaux (Chapitre 7), nous

nous attacherons à différencier les sons selon leur source (sons réels ou sons abstraits et synthétisés) plutôt que selon ce distinguo entre earcons et icônes auditives.

2.7 Conclusion

Les travaux présentés dans cette thèse concernent les interfaces auditives de l'automobile. Pour comprendre comment créer des sons efficaces et porteurs de sens, plusieurs étapes sont nécessaires. Traditionnellement, dans la littérature ICAD, il est d'usage de prendre comme point de départ le cahier des charges fonctionnel (un son qui indique une confirmation d'action, par exemple) pour en déduire - arbitrairement dans le cas des earcons, et de manière plus argumentée dans le cas des icônes auditives - des spécifications acoustiques qui vont guider la création sonore. Ensuite, vient - pour certains auteurs s'intéressant aux earcons, du moins - l'étape de validation des sons ainsi créés. Cependant, les validations expérimentales des interfaces auditives présentées dans ce chapitre manquent parfois de rigueur, et n'apportent pas toujours de réponses aux questions posées. Elles sont souvent centrées sur des problématiques de reconnaissance, d'identification et de mémorisation des earcons ou des icônes auditives proposés, et n'abordent en général pas d'un point de vue expérimental l'aspect fonctionnel d'un son. Nous avons également vu qu'une étape d'analyse perceptive de ces sons est manquante dans les travaux relatifs à ICAD, et nous avons resitué l'ensemble de ces travaux dans un cadre de perception et de cognition auditives. La première partie de nos travaux expérimentaux (Chapitre 5) reprendra une partie des résultats de la littérature ICAD, en cherchant à les mettre à l'épreuve - d'un point de vue fonctionnel - dans un cadre bien défini. Nous verrons alors que, parmi toutes les fonctions potentielles d'un son électronique de l'automobile, l'urgence apparaît comme une dimension importante, voire capitale. Il est en effet décisif d'informer le conducteur du degré d'urgence de l'information que l'on souhaite transmettre, afin qu'il puisse réagir dans les délais voulus. Le chapitre suivant présente les principaux travaux sur la perception de l'urgence.

Chapitre 3

Urgence perçue

3.1 Introduction

Comme décrit au chapitre précédent, les informations sonores peuvent être utilisées et modifiées pour transmettre certaines fonctions d'un système donné, dans notre cas, l'automobile. Parmi tous les messages différents qu'il est possible de véhiculer par un son, le degré d'urgence apparaît finalement comme une information cruciale. Ce chapitre présente tout d'abord les premiers travaux dans le domaine qui ont permis de définir les composantes basiques d'une alarme sonore (section 3.2). Après ces recommandations pratiques, nous décrirons les premières validations expérimentales de l'influence de paramètres acoustiques sur l'urgence perçue (section 3.3) ainsi qu'une application des travaux de Gaver (1989) sur les sons environnementaux appliqués aux alarmes (section 3.4). Les limites des deux approches précédentes nous permettront d'introduire en section 3.5 un nouveau cadre théorique de l'urgence perçue suggéré par Guillaume et coll. (2003). Enfin, nous situerons l'ensemble de ces travaux dans la problématique générale de la thèse pour justifier la proposition d'un nouveau paradigme associé à l'urgence d'un son (section 3.6).

3.2 Typologie d'une alarme

3.2.1 Constat de l'existant

Les alarmes auditives que nous connaissons et entendons dans notre environnement sonore quotidien prennent des formes très variées : klaxons, sirènes, buzzers, sons synthétisés, etc. Ces différents types d'alarmes se rencontrent dans des applications aussi diverses que

des appareils personnels (réveil, alarmes anti vol, ...), des moyens de transport (voitures, bus et aviation civile essentiellement), des applications militaires (aviation), ou encore des systèmes de surveillance (de centrale de contrôle, par exemple).

Ces applications diffèrent dans la fonction demandée aux alarmes, mais aussi et surtout dans leur contexte d'apparition. L'exemple de la voiture diffère radicalement de celui d'un réveil, dans le sens où plusieurs alarmes devront être utilisées dans l'automobile, à l'inverse d'un réveil personnel. La différence entre l'aviation civile et militaire se fait, quant à elle, davantage sur la tâche demandée à l'utilisateur, qui est, dans le cas militaire, beaucoup plus chargée attentionnellement. De plus, les alarmes militaires doivent informer le pilote de menaces extérieures aussi bien que d'une défaillance de l'avion en lui-même.

Ces exemples nous permettent alors de mettre en lumière plusieurs points. Tout d'abord, la première question concerne naturellement ce qu'il est possible de communiquer comme informations. D'après Stanton et Edworthy (1999), deux types d'informations peuvent être transmises par l'intermédiaire d'une alarme : d'une part, une indication sur la gravité du problème en cours, qui doit alors moduler la perception du degré d'urgence par l'utilisateur du système, et d'autre part, un renseignement sur la nature du problème en lui-même. Nous verrons dans la suite du document (Chapitre 5) que communiquer la cause du problème par une alerte sonore de manière précise est peut-être irréalisable, ou, du moins, n'est pas une question triviale.

Ensuite, d'un point de vue pragmatique, la première raison qui justifie l'apparition d'une alarme (visuelle ou sonore) est de réduire des accidents potentiels (Lehto et Papastavrou, 1998). Il est alors communément admis qu'une alarme sonore bien créée doit guider l'auditeur vers la réaction appropriée (Stanton et Edworthy, 1999 ; Morris et Montano, 1996 cité dans Guillaume et coll., 2003). Cependant, cette réaction attendue peut dépendre de facteurs contextuels, comme la charge attentionnelle demandée à l'utilisateur (cas de l'aviation militaire ou de l'automobile). Ce point sera traité plus en détail au cours du prochain chapitre.

Enfin, de la même manière que pour les interfaces auditives vues précédemment (Chapitre 2), se pose la question de la légitimité d'une alarme auditive par rapport à une alarme visuelle. Dans le cas présent, la justification semble plus aisée. En effet, comme souligné par Guillaume et coll. (2003), l'audition est un sens naturellement associé à la communication du danger. De plus, l'utilisation d'alarmes sonores est souvent justifiée par des comparaisons d'ordre physiologique entre les modalités visuelles et auditives. Nous retiendrons essentiellement que, dans certains cas, nous réagissons plus rapidement à des stimuli

auditifs qu'à des stimuli visuels. Kohfeld (1971) a montré, par exemple, que les temps de réaction à un son pur de 1 kHz ou à une lumière blanche produite par trois tubes fluorescents sont respectivement de 345 et 375 ms à 30 dB, 280 et 270 ms à 60 dB et 255 et 245 ms à 90 dB (respectivement 2×10^{-6} Pa pour le son et 10^{-10} Lambert pour la lumière). Des situations nécessitant une réaction rapide (en général, situations de danger) semblent donc bien indiquées pour être signalées par une composante sonore.

Pour finir, l'observation des cris d'alerte sonore du monde animal qui permettent de communiquer différents degrés d'urgence (et, de ce fait, la nécessité de fuir) ainsi que le type de prédateurs n'est sans doute pas anecdotique (Leavesley et Magrath, 2005 ; Manser et coll., 2002).

Suite à ces premières considérations utilitaires, les alarmes sonores sont devenues de plus en plus nombreuses, quelles que soient les applications. Cependant, la réaction des utilisateurs a souvent été inverse de celle attendue. Certains des problèmes découlant de cette expansion des alarmes dans nos systèmes d'information ont été extraits de récits de pilotes d'avion de chasse ou de médecins dans des services de réanimation, et résumés par Patterson (1990) ainsi que Sorkin et coll. (1988).

Le premier constat concerne le nombre d'alarmes qui peuvent apparaître au sein d'un même système (ou moyen de transport). Stanton et Edworthy (1999) constatent en effet que, au fur et à mesure de la complexification des systèmes, le nombre d'alarmes n'a cessé d'augmenter. Cependant, ces alarmes ont été ajoutées au fur et à mesure des besoins, sans une réelle vision d'ensemble. Il y a alors trop d'alarmes, dans le sens où il n'est plus possible de distinguer certaines alarmes entre elles (même avec un certain apprentissage).

Le deuxième problème résulte également de cette évolution historique : plus les alarmes sont nombreuses, plus la probabilité qu'elles apparaissent au même moment et soient masquées les unes par les autres est grande. De plus, ces alarmes peuvent être masquées également par le bruit de fond ambiant, très élevé dans le cas de l'aviation.

Du reste, aucune réflexion ergonomique n'ayant été menée au préalable, les pilotes (dans l'exemple de Patterson, 1990) se plaignent d'une part d'entendre les alarmes trop souvent, et d'autre part du niveau sonore parfois beaucoup trop élevé de ces alarmes. Les pilotes cherchent alors avant tout à couper l'alarme plutôt que d'y réagir d'une manière adéquate.

Enfin, Stanton et Edworthy (1999) remarquent que, bien que le but premier d'une alarme soit, de fait, d'attirer l'attention de l'utilisateur, elles peuvent aussi devenir une distraction néfaste au cours d'une tâche attentionnelle. Patterson (1990) met en garde contre une réaction de sursaut pouvant provenir d'alarmes mal conçues (i.e. avec un niveau

sonore trop élevé, par exemple), à la place de la réaction attendue.

Pour finir, les alarmes sont jugées généralement trop aiguës, et donc difficiles à localiser (Patterson, 1990).

Il peut paraître étonnant de présenter une analyse de l'existant datant d'il y a plusieurs années déjà. Cependant, aucune étude de la sorte n'a été menée dans le domaine automobile, et il semble alors utile de rappeler quelques éléments nécessaires à la conception d'une alarme. Bien qu'il ne soit pas toujours possible d'appliquer exactement les mêmes règles de création à des situations très différentes (aviation, équipement personnel, automobile, ...), quelques principes fondamentaux peuvent tout de même être avancés, sur la base des constats énoncés ci-dessus.

3.2.2 Proposition de conception d'une alarme

Patterson (1982) a alors proposé de résoudre ces problèmes d'adaptation des alarmes avec leur fonction (appelés encore l'approche "better safe than sorry"). Patterson et coll. (1986) précisent que les fonctions sont différenciées par rapport à la priorité de la situation. Ils distinguent l'*action immédiate*, la *prise de conscience immédiate* et la *simple information*.

Patterson (1982) suggère donc de créer des sons qui, d'une part, attirent l'attention sans faire sursauter, et d'autre part, permettent aux utilisateurs de communiquer entre eux pendant que l'alarme retentit, à l'opposé des alarmes existantes qui sont des sons continus - diffusés à des niveaux très élevés - et qui rendent, de ce fait, la communication difficile. De plus, il remarque que les différentes alarmes sont généralement confondues à cause de paramètres rythmiques trop proches ; il propose donc de créer les différentes alarmes d'un même système selon des patterns rythmiques bien distincts.

Ces principes sont issus des problèmes explicités ci-dessus (section 3.2.1), et peuvent être décrits plus en détail. Patterson (1982) a alors donné de grandes lignes directrices à suivre pour la création d'alertes sonores. Plus précisément, cette alternative aux sons d'alarmes classiques peut en fait être décomposée en deux grands thèmes : l'audibilité de l'alarme et sa compréhension.

Audibilité

L'audibilité d'une alarme doit certes prendre en compte les autres alarmes du système afin qu'elles ne se masquent pas entre elles, mais aussi, et surtout, le bruit de fond en-

vironnant. Patterson (1982) recommande alors que la limite basse d'une alarme soit au minimum 15 dB au-dessus du bruit de fond. Il fixe également un seuil maximum à 25 dB, pour éviter un niveau trop élevé des alarmes.

Une étude plus approfondie dans le cadre de l'aviation militaire (Patterson, 1990) lui permet de détailler ces recommandations, en exprimant ce seuil minimum d'une alarme (donc 15 dB au-dessus du bruit de fond) par bande de fréquences. Chaque composante fréquentielle d'une alarme doit être alors ajustée au centre d'un filtre auditif, le seuil étant ensuite déterminé en calculant la puissance à travers ce filtre. Le seuil minimum est alors exprimé en fonction de la fréquence. Il préconise ainsi de créer des sons élaborés suivant quatre composantes - au moins - respectant ce seuil et recouvrant une grande partie du spectre, afin d'éviter tout masquage auditif. Plus précisément, la fréquence fondamentale doit être comprise entre 150 et 1000 Hz, et la plage fréquentielle recommandée pour les harmoniques balaye les fréquences allant de 0.5 kHz à 4.0 kHz (Patterson et coll., 1986). Patterson (1990) précise l'importance de répartir les quatre composantes fréquentielles (au moins) à travers tout le spectre en s'appuyant sur certaines propriétés du système auditif relatives à la perception de hauteur. En effet, nous percevons une hauteur donnée à travers plusieurs composantes harmoniques reliées temporellement (la hauteur correspond alors à la fondamentale de cette série harmonique). Par conséquent, un son dont les harmoniques couvrent une grande partie du spectre aura bien moins de chance d'être masqué par un bruit de fond que si toute l'énergie du signal était concentrée à une harmonique donnée.

La problématique de l'audibilité de l'alarme ne sera pas traitée dans le cadre de cette thèse, les travaux existants étant suffisamment précis et détaillés. Des études internes Renault se chargeront de respecter le cahier des charges issu des travaux de Patterson (1982; 1990) et Patterson et coll. (1986) pour les futures IHM sonores, en l'adaptant au bruit de fond automobile. Le lecteur intéressé pourra également se reporter à l'étude de Laroche et coll. (1991) qui a mis en place un logiciel de détection d'alarmes dans un bruit de fond, basé lui aussi sur les filtres auditifs. Notons que les études de Patterson (1974; 1976) ainsi que celle de Patterson et Nimmo-Smith (1980) portant sur la forme des filtres auditifs sont à la base des recommandations et prédictions des travaux de Patterson sur l'audibilité des alarmes (Patterson, 1982; Patterson, 1990; Patterson et coll., 1986) alors que l'étude de Laroche et coll. (1991) se fonde sur les travaux de Zwicker et Scharf (1965) portant sur l'estimation de la sonie.

Hiérarchie structurelle

Une fois que les critères d'audibilité d'une alarme sont donnés, cette alarme doit être comprise et l'utilisateur doit lui associer le degré d'urgence adéquat. Cette compréhension passe par une phase d'apprentissage (cette assertion mériterait plus de détail, et sera discutée dans le Chapitre 8), par un faible nombre d'alarmes au global, et par un "bon design", qui permettra par la suite d'associer le degré d'urgence voulu à une alarme donnée.

Patterson (1982) propose alors de créer des alarmes suivant une hiérarchie structurelle, afin de véhiculer la fonction voulue. La construction se fait en trois étapes, décrites dans la Figure 3.1.

L'unité de base est une impulsion (pour *pulse*), d'une durée comprise entre 100 et 200 ms (notée t_0 sur la figure). Afin d'éviter les réactions de sursaut, Patterson recommande également d'avoir des temps de montée (pour *onset*) et de décroissance (pour *offset*) de 20 ms pour chaque impulsion (notés respectivement t_r et t_f sur la figure).

Chaque impulsion est ensuite répétée plusieurs fois. L'ensemble de ces impulsions forme un train d'impulsions (pour *burst*), défini par un nombre d'impulsions et un intervalle inter-impulsion. Patterson recommande d'utiliser au minimum cinq impulsions au sein d'un même train, de manière à pouvoir par la suite créer des rythmes suffisamment distincts les uns des autres. La figure indique un intervalle de temps t_s , analogue d'un *Inter Onset Interval* (IOI), qui comprend la durée d'une impulsion à laquelle s'ajoute l'intervalle inter-impulsion. Les différents trains d'impulsions durent approximativement $2s$, et ressemblent à des mélodies rythmiques.

Enfin, la combinaison de différents train d'impulsions forme l'alarme complète (noté *warning sound* sur la figure). Des plages de silence sont prévues entre chaque train d'impulsions, pour permettre la communication entre pilotes, comme expliqué précédemment. Cet intervalle de temps entre le début de deux trains d'impulsions successifs est appelé t_b .

Au sein d'un même système, six alarmes différentes au maximum peuvent être employées. En effet, Patterson (1982) remarque qu'il est difficile de mémoriser plus de cinq ou six sons d'alarmes distincts.

Perception et interprétation voulues

Au-delà de cette structure qui permet de définir les "fondations" d'une alarme, il est ensuite possible de modifier certains paramètres pour graduer le degré d'urgence perçue.

Ainsi, Patterson et coll. (1986) conseillent de faire varier les paramètres spectraux - dans

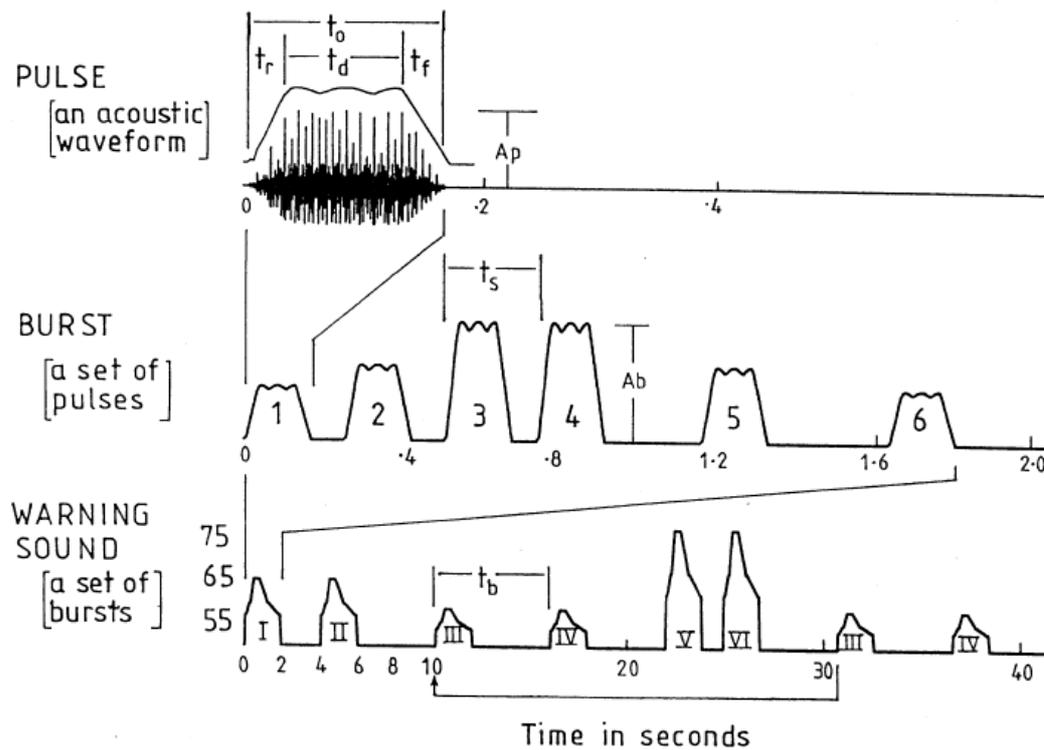


FIG. 3.1 – Structure d'une alarme suivant une hiérarchie structurale. Le premier élément est une impulsion, décrit par ses caractéristiques acoustiques. Le deuxième est appelé train d'impulsions; il est lui-même composé de plusieurs impulsions. La dernière étape consiste à agencer différents trains d'impulsions entre eux pour créer l'alarme finale. D'après Patterson, 1990

les limites des plages fréquentielles décrites ci-dessus - pour déterminer différents degrés d'urgence. Ces paramètres sont la fréquence fondamentale (ou F_0), le degré d'inharmonicité et le poids relatif entre les fréquences hautes et les fréquences basses.

Patterson et coll. (1986) introduisent également un paramètre temporel de l'impulsion qui permet de faire varier son urgence perçue : l'amplitude de modulation. Ce paramètre a également été proposé par Sanders et McCormick (1987, cité dans Stanton et Edworthy, 1999), où un son modulé devrait être associé à un son urgent.

De même que pour ces paramètres de l'impulsion, Patterson (1982) donne des recommandations concernant les paramètres du train d'impulsions. Ainsi, il suggère d'utiliser des intervalles inter-impulsion inférieurs à 150 ms pour des sons urgents, et supérieurs à 300 ms pour des non urgents.

Patterson et coll. (1986) définissent alors plus précisément le train d'impulsions suivant trois types de paramètres : le rythme, l'amplitude et la hauteur. Tout d'abord, aux préconisations sur les intervalles inter-impulsions s'ajoute ainsi un paramètre de régularité rythmique : un train d'impulsions régulier (défini par le paramètre t_s sur la Figure 3.1) doit être associé à un son moins urgent qu'un train d'impulsions irrégulier.

Concernant l'amplitude, un niveau sonore qui augmente au sein même d'un train d'impulsions devrait augmenter son urgence perçue. Nous pouvons remarquer ici que les études de Neuhoff (1998) d'une part et de Teghtsoonian et coll. (2005) d'autre part ont montré que la sonie d'un son croissant en intensité est plus élevée que celle d'un son décroissant. Bien que le lien entre ces études et notre sujet ne soit pas direct, il est tout de même intéressant de constater que la recommandation donnée en termes de niveau sonore pour véhiculer l'urgence est de faire varier ce niveau au cours du temps.

Enfin, l'utilisation de trains d'impulsions avec différents contours de hauteur (donc chaque impulsion a une hauteur différente) permet de rendre plusieurs alarmes bien distinctes les unes des autres. Un contour "montant" (donc avec des F_0 de plus en plus élevées) est décrit comme plus urgent qu'un contour "descendant".

La structure de l'alarme en elle-même peut également être paramétrée afin de transmettre le niveau d'urgence nécessaire. Patterson (1990) recommande de jouer en premier lieu les deux premiers trains d'impulsions au niveau d'urgence voulue. Le message étant ainsi communiqué, deux autres trains d'impulsions moins urgents (selon les paramètres de l'impulsion et du train d'impulsions que nous venons de décrire) sont alors joués. Si l'utilisateur (le pilote dans le cas présent) ne réagit pas, les deux premiers trains d'impulsions urgents sont joués à nouveau, et ce tant que l'utilisateur ne règle pas le problème indiqué

par l'alarme.

Malgré ces indications très précises, nous remarquons toutefois qu'aucune d'entre elles n'a fait l'objet d'une validation expérimentale en terme d'adéquation entre le son et la fonction. Ce sont des recommandations empiriques et utilitaires - certes de bon sens pour la plupart. Il serait alors intéressant d'étudier plus précisément le lien existant entre chacun des paramètres acoustiques décrits ci-dessus et une mesure subjective de l'urgence perçue.

3.3 Paramètres acoustiques et urgence perçue

Deux études majeures dans le domaine ont permis de définir les caractéristiques acoustiques de séquences sonores traduisant différents degrés d'urgence perçue (Edworthy et coll., 1991 ; Hellier et coll., 1993). Ces deux études se basent sur les recommandations de Patterson (1982) relatives à la construction d'une alarme suivant une hiérarchie structurale. Elles vont ainsi valider (ou infirmer) les préconisations de création d'alarmes sonores énoncées précédemment.

Ces études servant de référence à toute la littérature sur les alarmes sonores et l'urgence perçue, nous allons les présenter en détail. Une partie de ces résultats sera repris dans les travaux expérimentaux de la thèse (Chapitre 5 essentiellement). Nous commenterons également certaines limites inhérentes à cette approche psychophysique de l'urgence perçue en section 3.3.3.

3.3.1 Influence des paramètres acoustiques sur l'urgence perçue

Edworthy et coll. (1991) ont étudié l'urgence perçue grâce à deux paradigmes issus de la psychophysique : la comparaison multiple et l'estimation de grandeur. Les sujets doivent tout d'abord classer les sons en les comparant par groupe de quatre (choix du son le plus urgent, puis même procédure sur les trois sons restants, etc.). A la fin de cette première expérience, il leur est demandé d'évaluer l'urgence perçue de chaque son sur une échelle allant de 0 à 100. Notons tout de même ici que le terme d'*estimation de grandeur* est pour le moins un abus de langage, puisque l'échelle utilisée est bornée (entre 0 et 100). Les auteurs construisent donc une échelle d'intervalles (bien qu'ils prétendent par la suite obtenir une échelle de rapports). Malgré ces problèmes méthodologiques, cette étude servant de référence à toute la littérature de l'urgence et des alarmes sonores, nous allons la décrire ici en détail.

Trois groupes d'expériences sont menés : le premier sur les paramètres de l'impulsion, le deuxième sur ceux du train d'impulsions ; le dernier permet de vérifier les différentes hypothèses issues des premières expériences.

Paramètres de l'impulsion

Les auteurs choisissent de ne pas étudier l'intensité, Momtahan (1990) - cité dans Stanton et Edworthy (1999) - ayant déjà montré par ailleurs que l'intensité a une influence très forte sur l'urgence perçue d'une alarme : plus l'intensité est élevée plus le son est perçu comme urgent. De plus, comme le recommande Patterson (1982 ; 1990), l'intensité doit être ajustée pour s'assurer de l'audibilité du son ; les auteurs préconisent alors de modifier l'urgence perçue grâce à d'autres paramètres. De la même manière, ils choisissent de ne pas étudier les poids relatifs de chaque harmonique, puisque ces poids doivent être donnés par rapport au spectre du bruit de fond dans lequel sera diffusée l'alarme (voir section 3.2.2 page 42).

Au cours de deux expériences, les auteurs analysent alors plus en détail l'influence de la F0, de la série harmonique, de la forme de l'enveloppe temporelle et de la présence ou non d'harmoniques retardées. Deux ou trois valeurs de chaque paramètre sont croisées dans les deux expériences pour aboutir à des plans expérimentaux complets.

Les F0 étudiées prennent des valeurs comprises entre 150 et 530 Hz (expérience 1 : 150 Hz et 530 Hz ; expérience 2 : 200 Hz et 350 Hz). Les séries harmoniques sont régulières, irrégulières à 10% ou 50% (c'est-à-dire que les harmoniques paires sont 10% ou 50% plus élevées que leur valeur entière dans une série harmonique régulière) ou aléatoires (valeurs des harmoniques choisies aléatoirement). Les enveloppes temporelles sont *standards* (i.e. avec des temps de montée et de décroissance de 20 ms), avec *un temps de montée lent* (pas de définition plus précise donnée dans le texte, mais qui peut être estimé approximativement à 175 ms d'après une des figures de l'article) ou un *temps de décroissance lent*. Les harmoniques retardées sont *présentes* (les harmoniques élevées - à partir de la septième - sont retardées de 100 ms à partir du début de l'impulsion) ou *absentes* (i.e. que les harmoniques ne sont pas retardées).

Les sons sont présentés à un niveau maximum de 75 dB(A) sous la forme d'un train de six impulsions identiques de 200 ms, de rythme régulier. La première impulsion est atténuée de 5 dB par rapport aux autres pour éviter les réactions de sursaut. Chaque impulsion contient quinze harmoniques.

Outre la cohérence des résultats entre eux (cohérence inter-sujets et cohérence des

résultats entre les deux méthodes de mesure subjective), les auteurs montrent plusieurs effets des paramètres étudiés :

- Une F0 élevée (comparativement à l'autre F0 du corpus considéré) est associée à un son urgent.
- Une enveloppe temporelle standard est plus urgente qu'une enveloppe avec un temps de montée lent, qui est elle-même plus urgente qu'une enveloppe avec un temps de décroissance lent.

Cet effet est expliqué grâce à une possible association entre un temps de montée lent et l'image d'un objet qui se rapproche et un temps de décroissance lent et un objet qui s'éloigne ; l'objet qui se rapproche est donc supposé plus dangereux. Nous pouvons rapprocher cet effet des résultats de Stecker et Hafter (2000) sur l'effet d'asymétrie temporelle sur la sonie. Ils montrent que des sons avec une attaque lente et une décroissance rapide sont jugés plus forts que des sons avec une attaque rapide et une décroissance lente, pour des stimuli d'énergie égale. Il est alors possible que l'effet d'enveloppe temporelle asymétrique sur l'urgence perçue puisse en réalité s'expliquer par des différences perceptives (ici, la sonie). Nous reviendrons sur ce point de manière plus globale en section 3.3.3 page 54.

- Plus la régularité des harmoniques est imprévisible, plus le son est urgent. Dans les autres cas, un son irrégulier harmoniquement est plus urgent qu'un son régulier.
- Ils montrent enfin un effet de l'interaction entre la F0 et la forme de l'enveloppe temporelle : la F0 a un effet sur l'urgence perçue lorsque l'impulsion a une enveloppe standard ou avec une décroissance lente, mais pas sur les stimuli avec un temps de montée lent.

Les auteurs expliquent cet effet en supposant que le temps de montée lent donne plus la possibilité d'entendre la structure harmonique de l'impulsion ; par comparaison, la F0 serait moins saillante. Nous remarquons ici que la description des saillances perceptives donnée dans cette interprétation mériterait - pour le moins - plus de détails.

Paramètres du train d'impulsions

Une fois ces paramètres de l'impulsion définis et étudiés, un deuxième groupe d'expériences s'attache à comprendre l'effet de différents paramètres du train d'impulsions sur l'urgence perçue. Cinq expériences sont menées ; la procédure est toujours la même que celle décrite précédemment.

Sept paramètres sont testés (deux paramètres par expérience) ; ils prennent chacun deux ou trois valeurs suivant l'expérience considérée. Ces paramètres sont la vitesse (i.e. le tempo), le rythme (ici, la régularité), le nombre de répétitions, le changement de vitesse, le contour de hauteur, la gamme de hauteurs utilisée et la structure musicale. La vitesse (qui peut encore s'exprimer comme un intervalle inter-impulsion) peut être lente, moyenne ou rapide (pas plus de détails donnés dans le texte). Le rythme est régulier ou syncopé (pas plus de détails non plus). Le nombre de répétitions est de 1, 2 ou 4 (c'est-à-dire qu'un groupe de quatre impulsions est répété 1, 2, ou 4 fois). Notons ici que ce nombre de répétitions devrait être classé comme un paramètre de l'alarme et non du train d'impulsions, d'après la hiérarchie définie par Patterson (1982) (voir section 3.2.2 page 44). Le changement de vitesse affecte l'intervalle inter-impulsion au sein d'un même train d'impulsions : il peut être régulier (intervalle inter-impulsion constant), ralenti (intervalle inter-impulsion qui décroît systématiquement du début à la fin du train d'impulsions) ou accéléré (intervalle inter-impulsion qui croît systématiquement). Le contour de hauteur peut être unidirectionnel (montant ou descendant, mais avec une seule direction de changement au sein d'un même train d'impulsions) ou aléatoire (i.e. avec autant de changements de direction que possible). La gamme de hauteurs peut être petite (3 demi-tons d'écart entre les hauteurs extrêmes utilisées dans le train d'impulsions), moyenne (5 demi-tons) ou grande (8 demi-tons). La structure musicale est résolue, non résolue ou atonale (ces trois termes étant employés dans leur sens musical).

Tous ces paramètres ont un effet significatif sur l'urgence perçue :

- Une vitesse rapide est plus urgente qu'une vitesse moyenne, qui est elle-même plus urgente qu'une vitesse lente.
- Un rythme régulier est plus urgent qu'un rythme syncopé.
Ce résultat est inattendu, et les auteurs ne l'expliquent pas réellement. Ils conseillent d'approfondir par la suite ce paramètre d'irrégularité. Nous proposerons une validation expérimentale un peu différente dans les travaux expérimentaux de la thèse (Chapitre 6).
- L'urgence croît avec le nombre de répétitions du train d'impulsions. Toutefois, les auteurs précisent que, dans le cas d'une application concrète de ce résultat, trop de répétitions peuvent devenir irritantes pour les utilisateurs.
- Un train d'impulsions accéléré est plus urgent qu'un train d'impulsions régulier qui est lui-même plus urgent qu'un train d'impulsions ralenti.
- Une grande gamme de hauteurs employée est plus urgente qu'une petite, elle-même

plus urgente qu'une gamme moyenne.

L'interprétation de ce résultat - peu intuitif - n'est pas exactement donnée dans l'article.

- Un contour de hauteur aléatoire est plus urgent qu'un contour unidirectionnel. Les auteurs n'expliquent pas vraiment cet effet. Ils suggèrent uniquement d'utiliser ce paramètre pour discriminer des alarmes entre elles plus que pour communiquer différents degrés d'urgence.
- Une structure musicale atonale est plus urgente qu'une structure non résolue, elle-même plus urgente qu'une structure résolue. D'après les auteurs, ce résultat concorde avec l'interprétation selon laquelle ce qui est imprévisible est urgent. En effet, dans un cadre de musique occidentale tonale classiquement entendue, une structure atonale est rare et inattendue.

Confirmation finale

Une fois ces résultats obtenus sur ces paramètres relatifs aux impulsions et trains d'impulsion, une dernière expérience de confirmation est menée afin de vérifier les différentes prédictions simultanément. Les stimuli employés pour cette expérience sont décrits en Figure 3.2. Les paramètres cités sont ceux qui ont été explicités précédemment. Puisqu'un plan expérimental complet était impossible (trop grand nombre de stimuli pour une même expérience), treize sons uniquement ont été créés. Ils sont classés dans l'ordre d'urgence prédit (du plus au moins urgent).

Le stimulus 8 mis à part, les résultats de cette expérience finale confirment les prédictions précédentes. Ce petit écart entre classement observé et classement prédit n'est malheureusement pas du tout commenté par les auteurs.

A ce détail près, il est intéressant de constater que les résultats obtenus indépendamment les uns des autres pour la plupart (attendu que chacune des expériences décrites précédemment ne fait varier que deux, voire trois paramètres simultanément) sont validés pour des stimuli plus complexes et basés sur un plus grand nombre de paramètres.

Malgré des interprétations parfois manquantes ou hâtives, comme soulevé au cours de l'énoncé des différents résultats, cette première étude permet tout de même de préciser expérimentalement les premières recommandations de Patterson (1982). Elle montre comment il est possible de manipuler les paramètres acoustiques d'une alarme sonore afin de moduler son degré d'urgence perçue.

Warning	Pulse Characteristics			Burst Characteristics			
	Envelope	Harmonic Regularity	Pulse-Pulse Interval	Rhythm	Average Pitch	Pitch Range	Pitch Contour
1	Standard	Random	150 ms	Regular	600 Hz	300 Hz	Random
2	Standard	10% Irregular	175 ms	Regular	585 Hz	230 Hz	Random
3	Slow onset	Random	200 ms	Speeding	510 Hz	280 Hz	Random
4	Slow onset	10% Irregular	225 ms	Regular	525 Hz	200 Hz	Up/down/up
5	Standard	50% Irregular	250 ms	Syncopated	500 Hz	275 Hz	Down/up × 2
6	Standard	10% Irregular	275 ms	Speeding	450 Hz	100 Hz	Up/down/up
7	Standard	50% Irregular	300 ms	Syncopated	400 Hz	125 Hz	Up/down
8	Slow offset	50% Irregular	325 ms	Regular	335 Hz	170 Hz	Down/up
9	Slow offset	Regular	350 ms	Syncopated	300 Hz	120 Hz	Up
10	Slow offset	Regular	375 ms	Regular	250 Hz	75 Hz	Down/up
11	Slow offset	Regular	400 ms	Syncopated	210 Hz	80 Hz	Down
12	Slow onset	Regular	450 ms	Slowing	175 Hz	50 Hz	Up/down
13	Slow onset	Regular	550 ms	Slowing	290 Hz	75 Hz	Down

FIG. 3.2 – Description des paramètres des impulsions et trains d'impulsions des alarmes testées par Edworthy et coll. (1991). D'après les prédictions des expériences précédentes, ces alarmes sont censées être classées de la plus à la moins urgente. Seul le stimulus 8 a été classé à une autre position que celle prédite (par conséquent, les stimuli 9 et 10 ont également été décalés juste au-dessus). D'après Edworthy et coll., 1991

Cependant, il apparaît également que tous les paramètres n'ont pas le même poids sur le jugement final. Quelle est alors l'importance relative des différents paramètres acoustiques sur le jugement d'urgence ?

3.3.2 Quantification des effets

La quantification des effets des paramètres acoustiques sur l'urgence perçue est justement le propos de l'étude d'Hellier et coll. (1993). Elle reprend les études de Patterson (1982) et Edworthy et coll. (1991), et y ajoute les questions suivantes :

- Quel est le paramètre le plus efficace pour augmenter le degré d'urgence perçue ?
- Quelle est la variation que doit suivre chaque paramètre pour accroître le degré d'urgence perçue ?

Pour y répondre, les auteurs ont choisi d'étudier la puissance des relations entre quatre paramètres - issus des expériences d'Edworthy et coll. (1991) - et l'urgence perçue. Ces quatre paramètres sont la vitesse de l'alarme, sa fréquence fondamentale, le nombre de répétitions (où l'élément unitaire répété est constitué lui-même de deux impulsions) et le contenu harmonique (défini comme le nombre d'harmoniques irrégulières parmi les quinze présents dans chaque impulsion).

Quatre expériences vont alors être menées, au cours desquelles un seul paramètre acoustique varie (suivant sept niveaux de variation). Ce sont des expériences d'appariement inter-modal entre la perception de l'urgence en fonction d'un paramètre acoustique et la sensation de longueur d'une ligne. En d'autres termes, il leur est demandé de dessiner une ligne dont la longueur doit être proportionnelle à leur jugement d'urgence. Les résultats seront exprimés sous la forme d'une loi de Stevens :

$$S = kO^m \quad (3.0)$$

S est la mesure subjective - ici le degré d'urgence - O la mesure objective - la caractéristique physique du signal - et k un coefficient multiplicateur. Ainsi, plus grand est l'exposant m , plus un changement dans la valeur du paramètre produit un changement dans le degré d'urgence perçue.

Toutefois, nous pouvons remarquer que ce type de protocole expérimental peut poser problème dans l'établissement d'une échelle de rapports d'urgence en fonction du paramètre acoustique (ce que les auteurs cherchent à faire). En effet, la longueur de la ligne est de fait limitée par la taille du support. Les jugements d'urgence risquent alors d'être écrêtés pour les valeurs d'urgence élevée. Cet effet avait déjà été observé par Kuwano et Namba (1990), lors d'une expérience d'appariement pour un jugement continu de sonie, avec une ligne projetée sur un écran de 1 m de longueur. Ce biais expérimental ne permet alors pas d'obtenir une fonction de puissance traduisant directement une échelle de rapports d'urgence en fonction du paramètre acoustique. Dans la suite des expériences présentées par Hellier et coll. (1993), ce problème méthodologique a été contourné dans l'exposé des données : elles sont exprimées comme des longueurs de ligne en fonction de paramètres acoustiques, et le lien avec l'urgence perçue est fait dans l'interprétation de leurs résultats. Ces interprétations sont donc à considérer avec précaution. Malgré cela, de la même manière que nous avons déjà présenté les travaux d'Edworthy et coll. (1991), nous allons détailler ici l'étude d'Hellier et coll. (1993), cette étude servant également de référence dans la littérature de l'urgence perçue.

Une impulsion basique est définie pour toutes ces expériences : $F_0 = 300$ Hz ; durée = 200 ms ; amplitude = 75 dB ; enveloppe d'amplitude régulière ; harmoniques régulières. A partir de là, des trains de six impulsions sont construits. Dans chaque expérience, un seul de ces paramètres est modifié.

Les auteurs prédisent donc par combien il faut multiplier chaque paramètre pour obtenir une *longueur de ligne* double ou triple, et montrent aussi que certains paramètres ont

plus de poids que d'autres (en comparant les exposants obtenus dans chaque expérience séparément). Ainsi, ils en concluent que le paramètre le plus "économique" pour communiquer un message d'urgence est la vitesse : un petit changement de vitesse implique un grand changement dans le degré d'urgence perçue. Viennent ensuite le nombre de répétitions, la fréquence fondamentale puis l'inharmonicité.

Afin de tester ces paramètres simultanément, une dernière expérience d'estimation de grandeur a été menée, où trois niveaux d'urgence (élevé, moyen et faible) ont été créés à partir de trois des quatre paramètres acoustiques précédents (l'inharmonicité n'a pas été gardée dans le plan expérimental, son effet étant trop faible).

Ainsi, vingt-sept stimuli sont testés à partir des paramètres de vitesse, fréquence fondamentale et nombre de répétitions. Pour chacun de ces stimuli, une prédiction de leur niveau d'urgence peut être faite. Par l'intermédiaire de cette dernière expérience de confirmation, se pose également la question de l'additivité des dimensions de l'urgence.

De manière générale, les résultats confirment les hypothèses issues des quatre premières expériences. Cependant, dix des vingt-sept stimuli ne sont pas classés dans l'ordre prédit. L'analyse de ces dix stimuli montre que la F0 contribue plus que prévu au jugement d'urgence final.

Les auteurs expliquent cet effet inattendu de la F0 en remarquant que c'est le paramètre qui a la plage de variation la plus grande des trois. En effet, d'après les premières expériences, la F0 était un des paramètres les moins influents sur le jugement d'urgence. De grands changements de F0 dans les stimuli de la dernière expérience étaient donc nécessaires pour communiquer un petit changement dans le degré d'urgence. De ce fait, ce paramètre a pu apparaître perceptivement plus saillant que les autres. Sachant que les premières expériences se focalisaient sur les paramètres acoustiques un par un, il était prévisible que la création de stimuli faisant varier trois paramètres simultanément ne donne pas exactement les résultats escomptés par les auteurs. Il est également possible que le biais méthodologique soulevé ci-dessus soit à l'origine de cette différence entre les prédictions issues des expériences d'appariement inter-modal et les résultats de l'expérience d'estimation de grandeur.

3.3.3 Limites de cette approche

Cet effet imprévu de la F0 dans l'étude d'Hellier et coll. (1993) nous permet de soulever un des problèmes posés par l'approche de l'urgence perçue ainsi décrite. En effet, des méthodes psychophysiques ont été adoptées pour étudier l'urgence perçue. Cependant, ce

n'est pas la *perception* élémentaire d'un son qui est analysée en fonction des caractéristiques physiques d'un stimulus (comme peuvent l'être la sonie ou la hauteur), mais l'*interprétation* donnée à ce stimulus. La notion d'urgence peut difficilement être assimilée à un percept élémentaire, et dépend très clairement de facteurs cognitifs tels les connaissances antérieures du sujet, par exemple. Il est alors prévisible qu'il y ait une confusion dans les résultats entre l'espace perceptif des stimuli et l'interprétation que les sujets donnent à ces stimuli selon un axe d'urgence. L'effet de la F0 s'explique-t-il par une saillance perceptive de ce paramètre par rapport aux autres (due aux modalités choisies pour chacun des paramètres) ou par une prédominance de ce paramètre sur un jugement d'urgence? Nous aborderons ce problème - c'est-à-dire l'impossibilité de trancher entre les deux hypothèses - lors de nos travaux expérimentaux (Chapitre 5).

Le biais expérimental entre facteurs acoustiques, perceptifs et cognitifs a également été souligné et démontré expérimentalement par Guillaume et coll. (2003). Nous détaillerons leurs travaux plus loin en section 3.5.

3.4 Sons environnementaux et alarmes

Les études que nous venons de présenter dans la section précédente proposent des premières pistes pour pallier les problèmes mis à jour par Patterson (1982). Elles traitent ainsi de l'urgence perçue en fonction de paramètres acoustiques permettant de créer des sons abstraits. Toutefois, deux problématiques restent encore en suspens.

Tout d'abord, comme souligné en section 3.2.1, Stanton et Edworthy (1999) suggèrent que les alarmes sonores doivent communiquer un degré d'urgence, mais aussi la nature de l'événement à la base de l'alarme. Ensuite, reste le problème de la limitation du nombre de sons possibles au sein d'un même système (cinq ou six selon Patterson, 1982).

Stanton et Edworthy (1998) proposent alors d'explorer la possibilité de créer des alarmes sonores grâce à des sons environnementaux. De fait, ces sons seront représentatifs de l'événement qui est la cause de l'alarme. Ils soulignent de plus le fait que ces alarmes devraient être mieux identifiées (i.e. avec moins d'ambiguïté) que des alarmes "classiques" telles les sons abstraits décrits dans la section précédente. Stanton et Edworthy (1999) remarquent également que l'usage de sons environnementaux devrait permettre d'augmenter le nombre possible d'alarmes sonores dans un même environnement. Cependant, ce postulat n'a pas été vérifié expérimentalement.

L'approche de Gaver (1989) - décrite au cours du Chapitre 2 - est donc développée par

Stanton et Edworthy (1998) et appliquée aux alarmes sonores d'un service de réanimation d'un hôpital. Nous retenons de cette étude l'expérience visant à tester les deux hypothèses suivantes ainsi formulées par les auteurs : la fonction d'un son environnemental devrait être identifiée plus facilement que celle d'un son abstrait ; l'interprétation de la fonction d'une alarme sonore dépend de l'expérience de chaque individu.

Ils comparent alors les sons existants - i.e. abstraits - à de "nouveaux" sons, représentatifs. Les sons existants sont enregistrés à partir des équipements déjà en place dans le service de réanimation. Les nouveaux sons sont classés selon les trois catégories de Gaver (1989), symbolique, métaphorique et iconique, et évoquent chacun, par un lien plus ou moins direct, le type de problème ou défaillance qu'ils signalent. Aucun des sons - abstraits ou représentatifs - ne sont décrits avec plus de précision dans l'article.

Ils démontrent ainsi que les "anciens" sons sont mieux reconnus (par rapport au problème qu'ils désignent) que les nouveaux par les membres du service de réanimation, alors que ce sont les nouveaux sons qui sont mieux reconnus par des auditeurs novices. Les auteurs concluent de ce premier résultat que, en dehors de tout apprentissage d'une association entre un son et sa fonction, les sons représentatifs sont plus efficaces que des sons abstraits.

Nous pouvons tout de même remarquer que ce résultat était prévisible : le son d'un coeur qui bat pour signaler un problème dans la surveillance du rythme cardiaque d'un patient semble plus évident à comprendre qu'un simple "bip bip". Toutefois, suivant l'application considérée (l'automobile par exemple), la fonction que le son doit désigner n'a pas toujours de référent direct dans le monde sonore (voir Chapitre 2 section 2.6.3). De plus, il est intéressant de mettre en évidence que, pour des auditeurs experts, malgré la représentativité des nouveaux sons, ce sont les anciens sons abstraits qui ont été mieux identifiés.

De plus, les auteurs ont également montré que l'association entre un degré d'urgence et le son était plus correcte dans le cas des sons abstraits, et ce, quel que soit le groupe d'auditeurs considéré.

Bien que les auteurs aient voulu insister sur l'avantage des sons représentatifs pour une meilleure utilisation des alarmes, nous conserverons essentiellement de cette étude l'effet de l'apprentissage sur la reconnaissance d'une alarme sonore et sur l'urgence perçue des sons. De notre point de vue, il a ainsi été démontré involontairement les limites d'une telle approche expérimentale de l'urgence perçue. Si les sons environnementaux (représentatifs) ont un apport pour la création d'alarmes sonores plus efficaces, une autre approche doit être proposée pour le démontrer. Nous aborderons expérimentalement cette question au

cours du Chapitre 7.

3.5 Influence des facteurs cognitifs sur l'urgence perçue

Le problème de l'articulation entre différents niveaux de traitements (perceptifs et cognitifs) de l'urgence perçue est particulièrement bien illustré par l'article de Guillaume et coll. (2003). A travers trois expériences, les auteurs montrent expérimentalement que l'urgence perçue n'est pas uniquement d'ordre psychophysologique et perceptif, mais relève aussi de facteurs cognitifs. Les auteurs suggèrent ainsi que l'urgence perçue d'une alarme est liée à la pertinence de la représentation mentale qu'elle évoque chez le sujet en fonction du contexte.

La première expérience menée reprend les stimuli d'Edworthy et coll. (1991) décrits en section 3.3.1 dans la Figure 3.2. Les auteurs proposent ainsi de tester à nouveau la validité de ces résultats. Comme nous l'avons souligné précédemment (section 3.3.3), une mesure psychophysique de l'urgence *perçue*, telle que décrite par Edworthy et coll. (1991) ou par Hellier et coll. (1993) implique une confusion possible entre la perception des propriétés acoustiques des séquences sonores et la perception de l'urgence de ces séquences.

Le nouveau paradigme expérimental proposé par les auteurs permet alors de contrer dans une certaine mesure ce problème. Deux groupes de participants sont identifiés, avec chacun une tâche différente. Le premier groupe doit juger de l'urgence perçue des 13 séquences sonores d'Edworthy et coll. (1991), alors que le deuxième groupe juge le degré de dissemblance acoustique, sans aucune référence à la notion d'urgence. Une analyse multidimensionnelle leur permet de mettre à jour deux espaces perceptifs, l'un lié aux données de dissimilarité acoustique et l'autre à celles d'urgence perçue. La comparaison de ces deux espaces confirme, par une approche nouvelle, les résultats d'Edworthy et coll. (1991). Les auteurs soulignent que ce résultat est valable pour ce corpus de sons bien précis.

Et, de fait, en appliquant exactement la même méthodologie aux sons existants de l'aviation militaire française, la deuxième expérience menée par les auteurs montre que de nouveaux facteurs semblent modifier le jugement d'urgence perçue. En effet, bien qu'il soit toujours possible d'observer une influence des paramètres acoustiques sur le jugement d'urgence, le rôle de facteurs cognitifs est souligné. Les auteurs expliquent ainsi le haut degré d'urgence associé à des séquences sonores alternant entre deux tons, qui peuvent faire penser à des sirènes de pompier ou d'ambulance.

La dernière expérience présentée dans cette étude permet alors de comparer les poids

relatifs des stimuli des deux expériences précédentes (ceux d'Edworthy et coll. (1991) et une partie des alarmes existantes). Les résultats confirment l'importance de la représentation mentale associée à certaines séquences sonores (par exemple, des sons évoquant les sirènes), et dont il faut alors tenir compte lors de la création d'alarmes sonores.

Nous pouvons également remarquer ici que ces résultats mettent indirectement à jour un des problèmes soulevés au Chapitre 2 (section 2.6.4 page 36). En effet, la différence entre les earcons et les icônes auditives est souvent faite par rapport au type de sons utilisés (abstrait *versus* environnemental), mais aussi par rapport à la connaissance que nous avons de ce son (les sons environnementaux étant généralement associés à des sons connus et les sons abstraits à des sons non connus). Les sirènes de pompier de l'exemple de Guillaume et coll. (2003) doivent-elles être alors considérées comme des earcons ou des icônes auditives ? Nous soulevons une fois de plus que cette distinction n'a peut-être pas beaucoup d'intérêt (voir section 2.6.4 page 36), du moins dans notre cadre, et par la suite, nous distinguerons uniquement les types de sons plutôt que l'association entre son et fonction en tant que telle.

A partir du modèle théorique de traitement de l'information auditive de McAdams et Bigand (1994) - que nous avons exposé au cours du Chapitre 2 (section 2.4.2) - Guillaume et coll. (2003) proposent un nouveau cadre théorique de la perception de l'urgence, repris dans la Figure 3.3. Alors que les premiers travaux d'Edworthy et coll. (1991) et Hellier et coll. (1993) se sont focalisés sur l'analyse des propriétés acoustiques, Guillaume et coll. (2003) suggèrent de considérer deux étapes supplémentaires : l'analyse du contexte et l'activation d'une représentation mentale. En d'autres termes, il faut tenir compte des connaissances antérieures du sujet, de ses attentes, du contexte culturel, et d'un certain apprentissage pour pouvoir prédire un niveau d'urgence perçue.

Guillaume et coll. (2003) proposent ainsi une approche bien plus générale de la perception de l'urgence que celles vues jusqu'à présent, en résolvant d'une certaine manière les problèmes que nous avons pointé en section 3.3.3.

Toutefois, ce cadre théorique permet certes d'expliquer des résultats expérimentaux *a posteriori*, mais n'aide pas à créer de nouvelles alarmes plus efficaces, et ce, quel que soit le contexte culturel et les connaissances d'un utilisateur donné.

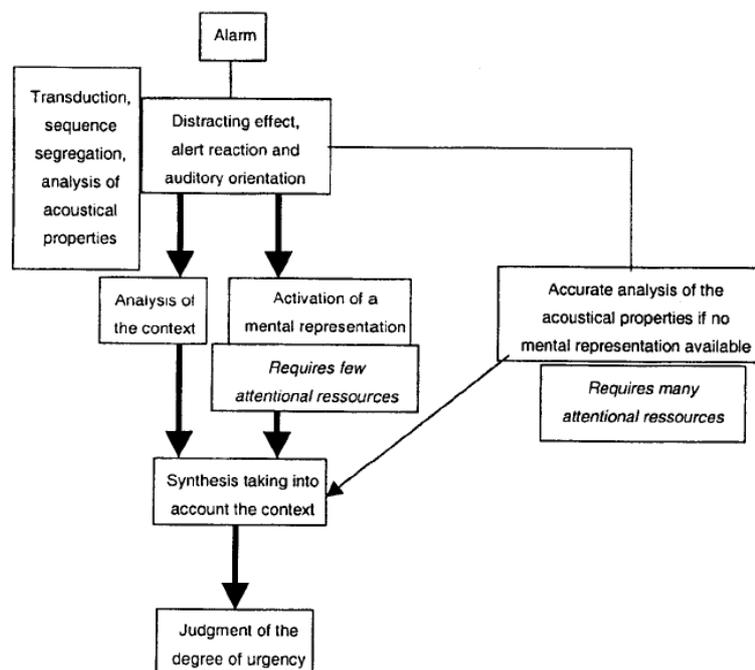


FIG. 3.3 – Modèle théorique de type traitement de l’information pour expliquer la perception de l’urgence. D’après Guillaume et coll., 2003.

3.6 Conclusion

Nous avons présenté dans ce chapitre les principaux travaux relatifs à l’urgence perçue de séquences sonores. Nous avons ainsi décrit l’influence de divers paramètres acoustiques sur l’urgence perçue : essentiellement, un tempo rapide et une fréquence fondamentale élevée sont associés à un fort degré d’urgence perçue. Cependant, l’approche psychophysique développée pour mettre à jour ces résultats comporte certaines limites (amalgame possible entre perception des propriétés acoustiques et degré d’urgence associé au son). Les travaux de Guillaume et coll. (2003) permettent de démontrer expérimentalement que des facteurs cognitifs entrent en jeu dans le jugement d’urgence, et qu’il faut alors, de fait, en tenir compte lors de la création d’alarmes efficaces et porteuses de sens. Il devient alors nécessaire de questionner à nouveau cette notion d’urgence, et ouvrir ainsi la voie à l’établissement d’un nouveau paradigme. Le chapitre suivant s’intéresse à cette question en se concentrant sur le temps de réaction en tant que mesure indirecte de l’urgence d’un son.

Chapitre 4

Alarmes sonores et temps de réaction

4.1 Introduction

Le chapitre précédent présentait les principaux travaux relatifs à l'urgence perçue d'alarmes sonores créées à partir de manipulation de paramètres acoustiques aussi bien que de sons réels. Nous avons également montré la diversité des facteurs impliqués dans la perception d'une alarme (Guillaume et coll., 2003). Ce constat étant posé, il reste alors à définir à nouveau des règles de création pour des alarmes sonores efficaces et performantes. Après une nouvelle définition de l'urgence, ce chapitre présente les premières études relatives au lien entre urgence perçue et temps de réaction (section 4.2). Les limites de ces études nous amèneront à exposer brièvement les théories et méthodes des temps de réaction (section 4.3). Enfin, après la conclusion de ce chapitre théorique (section 4.4), nous rappellerons la problématique des travaux expérimentaux décrits dans cette thèse (section 4.5).

4.2 De l'urgence perçue au temps de réaction

4.2.1 Définition de l'urgence

Nous avons montré dans le Chapitre 3 que l'étude de l'urgence évoquée par les alarmes sonores par le biais de jugements subjectifs comportait intrinsèquement certaines limites. En effet, les paradigmes utilisés, issus de la psychophysique, présupposent une caractérisation de l'urgence en tant que percept élémentaire, qui serait alors mesurable comme l'un des aspects fondamentaux de notre perception des sons. Bien qu'aucune définition de l'ur-

gence n'ait été proposée explicitement jusqu'alors, il semble manifestement qu'il ne s'agisse pas d'une qualité sensorielle "simple" (section 3.3.3 page 54). Le modèle hypothétique de la perception de l'urgence proposé par Guillaume et coll. (2003) met en avant la réalité complexe de cette notion, en tenant compte de facteurs acoustiques et cognitifs dans le jugement d'un degré d'urgence (Figure 3.3 page 59).

Cependant, comme expliqué à ce propos dans le Chapitre 3, la compréhension de ces différents phénomènes impliqués dans le jugement d'urgence perçue ne renseigne pas sur les règles de création sonore à suivre pour aboutir *in fine* à des alarmes sonores efficaces. Il nous a alors semblé utile à ce point de la présentation du contexte théorique de questionner à nouveau le concept d'urgence en le situant dans un cadre applicatif tel que le nôtre.

Par définition, l'urgence est le "caractère de ce qui est urgent, de ce qui requiert une action, une décision immédiate. L'urgence est la nécessité d'agir rapidement" (Dictionnaire *Trésor de la Langue Française*). Dans le cadre de l'automobile, les systèmes techniques sont de plus en plus performants et demandent des (ré)actions de plus en plus rapides lors de la réalisation de tâches critiques (dans le cas du régulateur de vitesse, par exemple, ce système additionnel prend en charge une partie de la tâche de conduite ; la reprise en main du véhicule par le conducteur dans certains cas critiques doit alors être la plus rapide possible). Il devient donc nécessaire de connaître la rapidité de certains processus mentaux et d'identifier les limitations temporelles des performances humaines. Ainsi, une alarme sonore urgente est une alarme qui provoque chez le conducteur la réaction la plus rapide possible, en général lors d'une tâche attentionnelle accrue. La question n'est alors plus de connaître directement le jugement d'un auditeur sur l'urgence perçue d'un son, mais de savoir si une alarme sonore donnée provoque une réaction plus ou moins rapide. La mesure indirecte de l'urgence par l'intermédiaire d'un temps de réaction est appelée l'*urgence objective*.

Nous présenterons au cours du chapitre l'intérêt d'une telle mesure comparativement au jugement d'urgence perçue. L'enregistrement d'indices comportementaux (i.e. temps de réaction) devrait permettre de fournir des informations essentielles sur l'analyse du déroulement temporel mis en jeu pour les différentes alarmes sonores testées et sur la possible variation du temps de traitement en fonction des paramètres acoustiques.

4.2.2 Urgence objective de sons abstraits

Nous ne sommes certes pas les premiers à proposer une mesure de temps de réaction pour évaluer l'urgence d'un son. En effet, les alarmes sonores sont diffusées afin d'attirer

l'attention du conducteur (ou du pilote) et de l'informer sur un état de fonctionnement - ou de dysfonctionnement - du système pour qu'il puisse réagir de la manière adéquate le plus rapidement possible. Deux grands thèmes semblent alors émerger : celui de l'attention, d'une part, et celui des temps de réaction, d'autre part. Ils peuvent être formalisés sous la notion de performance humaine. Les conducteurs sont soumis, de fait, à une certaine charge attentionnelle - qui peut être modélisée comme une performance continue - et doivent, de temps à autre, répondre à des alarmes - qui représentent une performance "discrète". Or, l'étude des alarmes sonores telle que nous l'avons présentée au Chapitre 3 (Edworthy et coll., 1991 par exemple) utilise le *judgement* comme mesure, et non la *performance*. Il devient alors nécessaire de proposer de nouveaux paradigmes simulant de manière plus complète la tâche réelle.

Haas et Casali (1995) ont mené une expérience en contexte attentionnel représentant l'environnement militaire (le *Criterion Task Set*, système développé par l'Armée de l'Air états-unienne, utilisé dans nombre d'études ayant trait à la performance humaine, et s'appuyant sur une batterie de tâches cognitives et psychomotrices) et en présence d'un masqueur (i.e. un bruit rose censé être représentatif du bruit ambiant dans le contexte réel). Les auteurs ont cherché d'une part à évaluer l'effet de certains paramètres acoustiques sur le temps de réponse, et d'autre part à examiner le lien entre l'urgence perçue et le temps de réponse. L'hypothèse expérimentale testée par les auteurs est en fait centrée sur l'urgence perçue des sons plus que sur le temps de réponse : les alarmes sonores les plus urgentes perceptivement devraient, du fait de cette urgence subjective, donner lieu à un temps de réponse plus rapide. Cette hypothèse est loin d'être évidente, et il n'est peut-être pas justifié, comme nous le verrons en section 4.2.4, de chercher une corrélation entre ces deux mesures bien différentes de l'urgence.

Trois paramètres acoustiques ont été étudiés : le "format" de l'impulsion unitaire, l'intervalle inter-impulsion et le niveau de l'impulsion. Le "format" peut être simultané (i.e. quatre sons purs à 500, 1000, 2000 et 3000 Hz présentés simultanément), séquentiel (cas où les mêmes quatre sons purs sont joués à la suite, chacun durant un quart de la durée d'une impulsion unitaire, pour que la durée totale des quatre sons purs soit la même que dans les autres cas) ou en dents de scie (un son pur à 500 Hz modulé en fréquence par une fonction en dents de scie, afin de produire un son variant de 500 Hz à 3000 Hz ; le texte ne donne pas plus de détails). L'intervalle inter-impulsion est défini comme le temps entre la fin d'une impulsion et le début de la suivante. Il peut être égal à 0, 150 ou 300 ms. Le niveau est fixé à 65 ou 79 dBC SPL ; les alarmes sont bien audibles dans le bruit

rose. Chaque stimulus est un train de huit impulsions, chaque impulsion ayant une durée de 350 ms, incluant 25 ms de temps d'attaque et de temps d'extinction. Remarquons que tous les stimuli créés n'ont pas la même durée totale, puisqu'ils varient selon l'intervalle inter impulsion.

L'expérience se déroulait en trois étapes : jugement d'urgence (estimation de grandeur) en présence du bruit rose, puis double tâche (temps de réaction et tâche attentionnelle) et jugement d'urgence à nouveau (comparaison par paire). Notons ici que lors de la double tâche, deux mesures - seulement - de temps de réaction par stimulus sont enregistrées (avec comme consigne de répondre le plus vite possible dès la diffusion du son). Nous reviendrons sur ce point en section 4.3.3 page 71.

Outre les effets des paramètres (ou d'interaction entre paramètres) sur l'urgence perçue, nous retiendrons essentiellement de cette étude l'analyse des temps de réaction. Seuls le format de l'impulsion et le niveau ont un effet significatif sur le temps de réaction. Or, les temps de réaction moyens mesurés sont de l'ordre de 450 ms. L'impulsion unitaire étant de 350 ms, les sujets répondent après avoir entendu une seule impulsion ; l'intervalle inter-impulsion ne peut donc pas, de fait, avoir d'effet sur le temps de réaction. De plus, les deux mesures de temps de réaction présentent une grande variabilité : pour un stimulus donné, l'écart-type mesuré représente 30% de la moyenne effectuée sur les deux temps de réaction. Nous reviendrons sur la variabilité intrinsèque aux mesures de temps de réaction en section 4.3.3 page 71 ; cependant, nous pouvons dire dès à présent qu'il est clair que cette grande variabilité observée dans les données de Haas et Casali (1995) amène à considérer leurs résultats avec beaucoup de précaution.

Burt et coll. (1995) sont partis d'un constat similaire à celui de Haas et Casali (1995) pour mener une étude psychophysique de l'urgence des alarmes sonores. Nous ne présenterons de cette étude que la partie comportementale relative à l'urgence perçue et au temps de réaction. L'hypothèse expérimentale liée à la mesure de temps de réaction est la même que dans le cas de Haas et Casali (1995) : les sons *perçus* comme plus urgents devraient induire les temps de réaction les plus rapides.

Nous pouvons également souligner dès à présent que, bien que l'application des travaux de Burt et coll. (1995) soit une application d'aviation militaire, la tâche attentionnelle utilisée par les auteurs (suivi de cible, i.e. coordination psychomotrice entre l'oeil et la main) représente une assez bonne approximation de la conduite automobile (De Waard, 1996). Deux conditions de suivi de cible ont été utilisées, une condition psychomotrice où le sujet devait suivre une cible à l'écran à l'aide d'un joystick, et une condition automatique

où le sujet devait uniquement regarder le suivi de cible effectué automatiquement par l'ordinateur. Nous reviendrons sur l'utilisation d'un suivi de cible au cours de nos travaux expérimentaux (Chapitre 6). Les sujets devaient donc accomplir une double tâche, à savoir le suivi de cible et la réponse aux alarmes.

Les sons utilisés ont été définis selon trois niveaux d'urgence supposés à partir des recommandations de Edworthy et coll. (1991). Deux paramètres ont été employés, la fréquence fondamentale et la série harmonique. Le son le plus urgent a une fréquence fondamentale de 146.8 Hz et une série harmonique aléatoire, le son d'urgence moyenne a une fréquence fondamentale de 523.3 Hz et une série harmonique irrégulière à 10%, et le son d'urgence faible a une fréquence fondamentale de 523.3 Hz et une série harmonique régulière (voir section 3.3.1 page 48 pour les définitions précises des paramètres employés). Nous pouvons tout de même remarquer que, d'après Edworthy et coll., le son censé être le plus urgent devrait avoir la fréquence fondamentale la plus élevée.

Outre une différence significative et importante des temps de réaction entre les deux tâches de suivi de cible (461 ms pour le suivi de cible automatique contre 828 ms pour le suivi manuel), les temps de réaction moyens ne diffèrent pas significativement entre les trois alarmes sonores utilisées. Cependant, les auteurs n'expliquent pas réellement ce résultat.

Les jugements d'urgence perçue effectués au préalable des mesures comportementales sont quant à eux significativement différents les uns des autres (selon le classement prévu, énoncé ci-dessus). Il apparaît donc qu'une différence d'urgence perçue n'est alors pas forcément traduite par une différence de temps de réaction.

Nous mettons principalement en avant ici que le lien supposé entre urgence perçue et temps de réaction n'est pas systématiquement validé, que ce soit dans l'étude de Burt et coll. (1995) ou dans celle de Haas et Casali (1995). Ce point sera précisé en section 4.2.4.

4.2.3 Urgence objective de sons environnementaux

En suivant toujours le même raisonnement, à savoir que la perception (d'urgence, par exemple) et la compréhension d'un son vont pouvoir moduler le temps de réaction, Graham (1999) propose d'étudier l'impact d'icônes auditives sur le temps de réaction dans un contexte automobile. L'expérience se déroule dans un simulateur de conduite automobile dans lequel une scène visuelle est projetée. L'alarme sonore sonne juste avant la collision ; il est alors demandé au sujet de freiner le plus vite possible.

Afin de vérifier l'hypothèse selon laquelle les sons environnementaux seraient mieux reconnus que les sons de parole ou que les sons abstraits, l'auteur compare alors quatre sons

différents, tous d'une durée de 700 ms. Le premier est une voix humaine disant "devant !" (*ahead*), le deuxième un son abstrait défini comme un son en dents de scie à 600 Hz et les deux derniers deux sons appelés environnementaux, à savoir un son d'avertisseur sonore et un crissement de pneu. Nous remarquons ici, une fois de plus, que la notion de son environnemental serait à redéfinir - à moins qu'elle n'ait pas réellement de sens : un son d'avertisseur sonore pourrait tout aussi bien être considéré comme un son abstrait par certains ; à l'inverse, des sons censés être abstraits pourraient faire partie de l'univers sonore quotidien de certains sujets (et il n'est pas possible d'avoir des données concrètes sur les connaissances antérieures des sujets sur tel ou tel type de son).

Trois mesures sont effectuées : le temps de réaction au freinage, le nombre de réponses inappropriées (i.e. le nombre de fois où le conducteur freine alors qu'il ne devrait pas et inversement) et un classement subjectif de ces alarmes en terme d'agrément (i.e. de préférence esthétique).

Les résultats obtenus semblent confirmer l'hypothèse de Graham : les sons environnementaux entraînent un temps de réaction plus court que les autres alarmes (en moyenne 730 ms pour le klaxon, 760 ms pour le crissement de pneu contre 810 ms pour le son abstrait et 850 ms pour la voix). Par ailleurs, les sons environnementaux entraînent plus de réponses inappropriées. Enfin, ces sons sont appréciés pour leur réalisme.

L'auteur propose alors une interprétation des résultats sous un angle "écologique" : la signification inhérente des sons environnementaux induit - selon lui - des temps de réaction plus courts ; l'alerte est interprétée plus rapidement et prépare à une action immédiate. Il remarque également que si une alerte n'est pas appréciée par l'utilisateur ou qu'il la considère comme inappropriée, cette alerte peut alors provoquer un énervement chez l'utilisateur et, de fait, n'entraînera pas la réaction appropriée.

Toutefois, des objections d'ordre méthodologique peuvent être formulées. Cinq mesures de temps de réaction sont enregistrées pour chaque condition expérimentale (type de son, type de collision), ce qui semble bien peu (voir section 4.3). Ensuite, ces temps de réaction sont mesurés au pied, ce qui augmente la variabilité due au temps moteur (la tâche précise consiste à enlever le pied de la pédale d'accélérateur et appuyer sur la pédale de frein). De plus, il paraît un peu hâtif de conclure que la différence de temps de réaction observée est due à la reconnaissance des sons environnementaux. D'une part, parce que les limitations méthodologiques que nous venons de formuler amènent à considérer les résultats avec précaution, et d'autre part, parce qu'il n'est pas évident que les auditeurs aient reconnu le son avant de répondre, puisqu'il leur était demandé de répondre le plus vite possible. Ce

point sera étayé plus en détail en section 4.3.2 page 69.

Notons qu'une étude semblable a été menée par Belz et coll. (1999) ; les auteurs concluent de manière similaire à un avantage des sons environnementaux comme alarme sonore.

4.2.4 Limites de ces approches

Comme nous l'avons souligné au fil du texte, les études que nous venons de présenter comportent certaines limites. Indépendamment des limites d'ordre méthodologique (nombre de mesures, etc.) qui seront abordées en section 4.3, une des hypothèses souvent posée concerne le lien entre urgence perçue et temps de réaction. Or, nous avons vu que, dans certains cas, un son perçu urgent ne donnait pas nécessairement lieu à une réaction plus rapide (travaux de Haas et Casali, 1995 et Burt et coll., 1995, présentés en section 4.2.2). Quand bien même une corrélation élevée aurait été observée systématiquement entre urgence perçue et temps de réaction, avec un temps de réaction qui décroît lorsque l'urgence perçue croît, les limites que nous opposons à cette hypothèse resteraient les mêmes.

En effet, l'urgence perçue se déduit du jugement des sujets. De fait, elle est subjective, et dépend du contexte culturel, des attentes de l'auditeur, de l'appréciation esthétique du son et de son agrément. Les temps de réaction reflètent les performances effectives des sujets, et surtout celles de notre système auditif (et moteur). Comme nous l'avons vu au chapitre précédent (section 3.5 page 57), Guillaume et coll. (2003) ont mis en évidence que de nombreux facteurs influaient sur le jugement d'urgence : facteurs acoustiques, cognitifs et attentionnels. Cependant, une fois que cette diversité de facteurs est ainsi mise à jour, il reste toujours nécessaire de les identifier plus précisément, et surtout, de savoir quels sons seront les plus efficaces en tant qu'alarmes sonores. Une approche comportementale telle que la mesure des temps de réaction semble alors adaptée pour obtenir une mesure *objective* de l'influence de certains paramètres. Comme nous l'avons évoqué au Chapitre 2 (section 2.3), les fonctions étudiées par la psychologie cognitive se définissent par l'assemblage d'éléments opérationnels nécessaires à la réalisation de tâches complexes (détection, identification, etc.). Afin de décomposer ces entités globales en opérations plus simples - les "briques" du système cognitif - il est courant d'utiliser la *chronométrie mentale*, i.e. la mesure du temps de réponse à un stimulus. Le temps de réaction mesuré dans les articles cités ci-dessus est un temps de réaction simple, i.e. incluant essentiellement une étape de *détection* (nous reviendrons plus précisément sur cette question par la suite en section 4.3.2). En d'autres termes, dès que le sujet détecte un son, il doit appuyer sur un bouton

le plus vite possible. Il n'y a alors pas de raison évidente pour que les différentes étapes - hypothétiques - intervenant dans le jugement d'urgence (analyse du contexte, comparaison avec des connaissances antérieures, activation d'une représentation mentale, reconnaissance ou analyse des propriétés acoustiques) soient traitées en amont de la simple détection du signal - voire même simultanément - pour pouvoir ainsi avoir une influence sur le temps de réponse *simple*. L'enregistrement d'un temps de réaction simple n'est censé refléter que deux étapes de traitement, l'étape de détection du signal par le système auditif et l'étape de temps purement moteur (voir sections 4.3.2 et 4.3.3), et ne peut vraisemblablement pas rendre compte à lui seul de la diversité des traitements effectués par notre cerveau lors de l'écoute d'une alarme sonore. Néanmoins, comme nous l'évoquerons au cours de nos travaux expérimentaux (section 7.3 page 171), il reste également envisageable que certains traitements (détection et reconnaissance de l'objet sonore) s'effectuent en parallèle.

Par ailleurs, ces limites peuvent en fait être abordées sous l'angle d'une problématique générale, à savoir la prise en compte de problèmes concrets et appliqués lors de la réalisation d'une expérience. Puisqu'il n'est pas possible de simuler de manière complète et précise la tâche réelle qu'aura à effectuer le conducteur (ou l'utilisateur plus généralement), il est alors nécessaire de la simplifier.

Cependant, outre cet aspect de simulation des aspects pertinents des tâches réelles, les paradigmes expérimentaux choisis doivent aussi servir d'étalon afin d'améliorer notre compréhension de certaines fonctions cognitives de base. Alors, un paradigme minimal (tel les temps de réaction simples décrits précédemment) devrait être un pont entre la recherche fondamentale et appliquée ; pont qui relierait la compréhension des mécanismes cérébraux à l'intérêt que porte la communauté des facteurs humains (*Human Factors*) pour l'analyse du comportement humain au cours de tâches complexes (comme la conduite automobile).

4.3 Temps de réaction

Plus qu'un simple paradigme expérimental ou qu'une simple variable dépendante dans un protocole, les temps de réaction définissent en réalité un domaine de recherche à part entière de la psychologie expérimentale. De très nombreuses recherches ont été menées afin de répondre à des questions variées autour de l'architecture cognitive et sur les algorithmes utilisés par le cerveau humain. Parmi les références incontournables dans ce domaine, citons l'ouvrage de Luce (1986). Le but ici n'est pas de fournir une revue exhaustive de tous ces travaux, mais d'en extraire quelques points qui nous ont semblé importants à

l'établissement d'un nouveau paradigme de l'urgence et à la compréhension des phénomènes qui seront ainsi mis en jeu, pour nous permettre d'enrichir les connaissances existantes sur les alarmes sonores (voir Chapitres 6 et 7).

4.3.1 Introduction

Toute réponse à une stimulation implique nécessairement un délai entre le début du stimulus et le moment de la réponse : le temps de réaction. En d'autres termes, quelle que soit la tâche que le sujet ait à effectuer, sa réalisation prend un certain temps. Le temps de réaction est donc communément défini comme le temps qui s'écoule entre le début d'un stimulus et le début de la réponse. Ce temps de réponse révèle des informations relatives à l'activité mentale. Cette idée a été introduite il y a plus d'un siècle, entre autres par Donders (1868) dont nous présenterons les principaux travaux en section 4.3.2. Dans le cadre d'une approche de type traitement de l'information (voir section 2.3 page 13), l'hypothèse générale peut être formulée ainsi : plus un traitement est complexe, plus il y aura d'étapes de traitements, et donc plus long sera le temps de réaction.

L'analyse du déroulement temporel des processus cognitifs soulève des questions fondamentales comme la décomposition d'une tâche cognitive en briques élémentaires, le temps nécessaire à chacune de ces étapes, ou encore l'influence des paramètres de la tâche sur le temps de traitement. Une autre question d'intérêt concerne l'ordre d'exécution des étapes, et le mode de traitement des informations (traitement parallèle ou séquentiel).

4.3.2 Modèle de Donders

Historiquement, après Helmholtz (1850) - cité dans Sanders (1998) - qui a cherché à déterminer la vitesse de la conduction nerveuse, Donders (1868 - 1969) s'est intéressé à la question de la mesure de la vitesse de la pensée, et a suggéré que les processus mentaux pouvaient être inférés à partir de l'observation de différents types de réaction. Nous ne traiterons dans cette section que des travaux de Donders, qui ont servi de référence aux théories cognitives de type traitement de l'information. Nous pouvons néanmoins noter ici que ses travaux ont été repris (et critiqués partiellement) par Wundt (1883) et Cattell (1886) - cité dans Sanders (1998) ; pour une revue détaillée de leurs travaux, le lecteur pourra se reporter à l'ouvrage de Luce (1986).

Ainsi, afin de mesurer la vitesse des opérations mentales, Donders a défini trois réactions prototypiques : le temps de réaction simple (TRS), le temps de discrimination et le temps

de réaction de choix (TRC). Dans les tâches de réaction simple, un stimulus est présenté au sujet ; la consigne est de répondre le plus rapidement possible dès qu'il a détecté le stimulus (la réponse étant généralement d'appuyer sur un bouton). Les tâches de temps de discrimination sont définies comme des "Go - No Go", dans lesquelles le stimulus prend deux valeurs ; le sujet ne doit répondre qu'à l'une d'elles le plus rapidement possible. Dans une tâche de temps de réaction de choix, le stimulus prend plusieurs valeurs, et le sujet a plusieurs réponses possibles (plusieurs boutons sur lesquels appuyer par exemple). Il est demandé au sujet de choisir le plus rapidement possible entre les différentes réponses suivant le stimulus.

Le modèle de Donders propose alors une simple additivité des étapes de traitement. Les étapes de traitement associées à chacune des tâches décrites ci-dessus sont détaillées dans le tableau 4.1.

Tâche	Étapes supposées de traitement	Modélisation des TR
Simple (TRS)	Détection sensorielle du stimulus (T_1) Étape motrice d'exécution (T_4)	$TR_{total} = T_1 + T_4$
Go - No Go	Détection sensorielle Discrimination perceptive (T_2) Étape motrice	$TR_{total} = T_1 + T_2 + T_4$
Choix (TRC)	Détection sensorielle Discrimination perceptive Choix (T_3) Étape motrice	$TR_{total} = T_1 + T_2 + T_3 + T_4$

TAB. 4.1 – Étapes supposées de traitement lors de la réalisation de trois tâches de temps de réaction. D'après le modèle de Donders (1868)

Il est alors possible, en comparant les temps mis dans deux conditions différentes, de mesurer la durée de l'étape de traitement supposé (celle de choix décisionnel par exemple) : cette méthode est appelée méthode soustractive.

La méthode soustractive et l'hypothèse ainsi faite de l'additivité des différentes étapes de traitement a été largement critiquée, par Cattell (1886) en premier lieu. Un des sujets de controverse depuis de nombreuses années concerne la constance de la durée de la phase d'exécution motrice et son indépendance relativement aux paramètres de la tâche. Il n'est en effet peut-être pas si intuitif que le temps moteur soit indépendant des paramètres

testés et ne varie pas, quelle que soit la tâche demandée (simple, go - no go ou choix). Néanmoins, Miller et Low (2001) ont pu récemment montrer, au moyen d'enregistrements électrophysiologiques, que cette hypothèse était bien vérifiée. Le temps moteur est constant quelle que soit la tâche, et estimé aux alentours de 100 ms.

Dans nos travaux expérimentaux (Chapitres 6 et 7), seul le TRS sera étudié, nous ne détaillerons donc pas plus les différents modèles proposés par la suite pour tenter d'expliquer l'organisation cérébrale. Le lecteur intéressé se reportera à l'ouvrage de Luce (1986) ou celui de Sanders (1998) pour plus de détails.

4.3.3 Paradigmes de détection

Le problème posé par le paradigme de détection (ou de TRS) est le suivant : à quelle vitesse peut-on détecter l'arrivée d'un signal et y répondre, sans nécessairement le reconnaître et l'identifier ? Afin de répondre correctement à cette question, certaines précautions sont nécessaires pour pouvoir mener à bien une expérience de temps de réaction. Là encore, de nombreuses données pourraient être rapportées ici. Nous les simplifions au maximum et ne gardons que ce qui sera nécessaire à nos travaux expérimentaux.

Les effets de l'intensité du signal (lumineux ou sonore) ont été très largement étudiés ; une courte revue des effets de l'intensité sonore sur le temps de réaction sera présentée au cours de nos travaux expérimentaux (Chapitre 6).

Parmi les trois principaux modes de réaction possibles (i.e. moteur, vocal ou par saccade oculaire), nous avons choisi - pour des raisons pratiques uniquement - la réponse motrice. Comme évoqué précédemment, le temps de réaction simple peut se décomposer en deux étapes de traitement : une liée à la détection sensorielle du stimulus et une liée à l'exécution de la tâche motrice. En général, c'est la détection sensorielle qui est étudiée, et le temps moteur est considéré comme une latence résiduelle. Nous ferons l'hypothèse dans nos travaux expérimentaux (Chapitres 6 et 7) que le temps moteur est constant pour un même sujet, suivant les différents facteurs expérimentaux testés. Comme expliqué en section 4.3.2, cette hypothèse se justifie - entre autres - par les travaux de Miller et Low (2001).

Toutefois, d'une mesure à l'autre, il est courant (et même systématique) d'observer des différences de temps de réaction. Cette variabilité intrinsèque à la mesure de temps de réaction est principalement due au temps moteur. Des consignes précises peuvent être données pour essayer de le minimiser (garder la main constamment appuyée sur le bouton réponse, par exemple). Comme nous l'avons remarqué précédemment (sections 4.2.2 et

4.2.3), une dizaine de mesures aboutit à une variabilité des données de l'ordre de 30% - pour les articles cités ici, du moins. Comme indiqué par Luce (1986), un grand nombre de mesures de temps de réaction permet de réduire la variabilité du temps moteur, et de pouvoir observer d'éventuelles différences entre conditions expérimentales qui seraient dues à l'étape de détection sensorielle du stimulus. Afin d'obtenir des distributions de temps de réaction acceptables, Chocholle (1940) fait autant de mesures que nécessaire pour qu'au final, l'écart-type observé soit inférieur à 10% de la moyenne. Suivant les tâches (double tâche, par exemple), il est possible de tolérer un peu plus de variabilité. Crabtree et Antrim (1988) recommandent d'effectuer au minimum 40 mesures de temps de réaction dans chaque condition.

Cependant, un grand nombre de répétitions implique inévitablement un effet d'apprentissage ou de fatigue. Afin d'éviter ces effets, il est possible de découper l'expérience en blocs expérimentaux et de vérifier, *a posteriori*, si le nombre de répétitions a eu un effet sur le temps de réaction.

Enfin, lors de l'observation des distributions de temps de réaction, il est également très classique de constater la présence de valeurs clairement différentes des autres ; ce sont les outliers. Ils reflètent en général des temps de réponse à des processus qui ne sont pas ceux étudiés : fausses alarmes, anticipations, inattention, prise de décision erronée, etc. Il est alors nécessaire de supprimer ces outliers des données "réelles". Cependant, séparer les outliers des "vraies" données est un problème complexe (Ratcliff, 1993). Cette question sera traitée dans chaque expérience lors de nos travaux expérimentaux (Chapitres 6 et 7).

4.4 Conclusion

Comme décrit au début de ce chapitre (section 4.2.1), l'urgence est définie comme la nécessité d'agir rapidement. Nous déduisons de cette définition très générale une autre plus précise en terme de paradigme expérimental. L'urgence, que nous appellerons *objective*, est alors définie comme un temps de réaction simple, soit principalement un temps de détection (bien qu'il ne soit pas possible d'omettre le temps moteur, ce n'est pas la variable d'intérêt dans la plupart des cas, et notamment dans le nôtre).

Malgré cette description qui semble assez intuitive, l'étude des alarmes sonores et de l'urgence s'est en général focalisée sur l'urgence perçue, et les études relatives au lien entre urgence et temps de réaction sont peu nombreuses. Elles comportent, de plus, certaines limites méthodologiques qui ne permettent pas réellement d'apporter de nouvelles pistes à

la création d'alarmes sonores performantes.

4.5 Vers une nouvelle caractérisation des alarmes sonores

Les paramètres issus des travaux sur l'urgence subjective et sur les IHM vont être alors à nouveau questionnés sous l'angle du nouveau paradigme expérimental proposé au cours de ce chapitre. Les questions se posent ainsi un peu différemment des résultats de la littérature obtenus jusqu'à présent :

- les paramètres rythmiques, exposés comme les plus importants dans un jugement d'urgence perçue, ont-ils une influence sur l'urgence objective, et dans quelle mesure ?
- si c'est le cas, comment serait-il possible d'interpréter ce résultat ?
- les sons dits "environnementaux", proposés comme une solution séduisante et intuitive dans de nombreuses IHM, ainsi que pour les alarmes sonores peuvent-ils influencer sur les performances en termes de temps de réaction ?
- si oui, ce résultat s'explique-t-il par le sens sous-jacent aux sons choisis ou par certaines caractéristiques acoustiques particulières ?

La problématique principale que nous allons aborder dans les travaux expérimentaux décrits dans les chapitres suivants est celle de l'influence de paramètres acoustiques sur l'urgence objective des alarmes sonores. L'ensemble des expériences conduites au cours de cette thèse ont été organisées comme suit : dans un premier temps, nous avons cherché à reprendre et à valider dans le cadre des IHM de l'automobile les résultats issus de la littérature relative au domaine des interfaces auditives et à celui concernant l'urgence perçue des alarmes (Chapitre 5). Cette expérience nous a permis de confirmer l'importance des paramètres rythmiques pour ordonner les sons abstraits dans les catégories d'IHM automobiles (définies entre autres selon une échelle d'urgence). Nous avons alors étudié l'effet de deux paramètres rythmiques sur le temps de réaction en contexte d'attention partagée, visant à simuler la conduite automobile (Chapitre 6). Puis, nous avons analysé plus en détail l'effet d'un paramètre temporel issu des expériences précédentes, et nous avons comparé l'influence de ce paramètre sur le temps de réaction à celle de sons d'animaux (Chapitre 7). L'ensemble de ces résultats permet d'apporter de nouvelles pistes prometteuses pour la création d'alarmes sonores performantes, tout en apportant certaines indications sur les processus sensoriels et cognitifs impliqués dans le traitement de ces alarmes.

Deuxième partie

Travaux expérimentaux

Chapitre 5

Fonction perçue des IHM sonores

5.1 Introduction

Lorsque nous nous retrouvons au volant d'une voiture, nous devons faire face à de multiples tâches (suivi de la route, prise en compte d'événements extérieurs, etc.) et notre attention se retrouve alors sollicitée dans des situations très diverses. A la tâche de conduite "classique", s'ajoute désormais une interaction de plus en plus grande entre le conducteur et sa voiture par le biais de nouvelles fonctionnalités électroniques et d'assistance à la conduite au sens large. Ce sont ces Interfaces Homme Machine (IHM) - visuelles et sonores - qui permettent de communiquer au conducteur certaines informations relatives à l'état du véhicule dans son ensemble, sur le plan électronique comme mécanique. Dans le cadre de cette thèse, seules les informations *sonores*, et *non verbales* ont été étudiées.

Outre un aspect "esthétique" recherché par le constructeur, afin de donner une image de qualité du véhicule, le message communiqué par les informations sonores doit être également parfaitement décodé par le conducteur. Le son doit être utilisé au moment adéquat, et aider à la transmission de l'information. Il aide le conducteur à comprendre, décider et agir. Dans l'immense majorité des cas, les IHM sonores sont accompagnées de signaux visuels, afin d'explicitier le message véhiculé. Les IHM sonores doivent donc transmettre un message "simple" (qui sera étoffé par la suite grâce à l'information visuelle) et être appréhendées rapidement et sans ambiguïté.

Cependant, une étude préliminaire (Suied et coll., 2005) a montré que les IHM sonores existantes (en 2003) ne communiquaient pas l'information pour laquelle elles avaient été conçues (par exemple : "oubli de fermeture de porte", "liquide frein défaillant", etc.). Trente-huit sons différents, enregistrés dans dix-huit véhicules Renault et de la concurrence

ont été étudiés. Le but de cette pré-expérience était de classer les sons dans trois catégories pré-définies suivant le niveau de priorité. La catégorie la plus urgente était définie comme "Avertir d'un danger"; venait ensuite "Signaler un oubli sans risque"; enfin, "Informé (confirmer, assister)". Globalement, les sons existants dans l'automobile ne respectaient pas du tout le cahier des charges fonctionnel, tous constructeurs et gammes de véhicules confondus. Plus spécifiquement, les sons de danger ont été les plus mal classés : 71% des sons de danger ont été classés comme sons d'oubli ou comme sons d'information. La connaissance de la littérature des sons d'urgence non verbaux aurait pu permettre d'éviter, au moins en partie, ces écueils.

Les études que nous avons regroupées sous le nom d'approche subjective de l'urgence (Chapitre 3) ont étudié l'influence de certains paramètres acoustiques sur l'urgence perçue (Patterson, 1982; Edworthy et coll., 1991; Hellier et coll., 1993). Une vitesse - ou cadence - rapide (qui peut s'exprimer comme un intervalle inter-impulsion faible), une fréquence fondamentale (F0) élevée et des harmoniques irrégulières sont reliées à une grande urgence perçue (Edworthy et coll., 1991). Hellier et coll. (1993) ont montré par la suite que la vitesse de l'alarme est le paramètre prédominant de l'urgence perçue, suivie par la F0. A l'issue de ces résultats, des recommandations précises peuvent donc être données quant au design d'alarmes efficaces et capables de transmettre différents niveaux d'urgence.

La communauté ICAD, s'est, pour sa part, intéressée à définir des principes de création d'"earcons" (Chapitre 2, section 2.6.2 page 26), en calquant les règles et principes de la création d'icônes visuelles pour une interface informatique (Blattner et coll., 1989). Le rythme est spécifié comme la caractéristique dominante du motif sonore, suivi des paramètres "timbre" et "hauteur". Toutefois, comme nous l'avons déjà souligné au cours du Chapitre 2, la validation expérimentale de ces propositions de design sonore est souvent peu précise, et concerne principalement la reconnaissance et la mémorisation des earcons (Brewster et coll., 1994). En d'autres termes, les propositions de design sonore de ces auteurs, bien que validées d'un point de vue de la reconnaissance, restent assez intuitives d'un point de vue fonctionnel (i.e. quel son pour une fonction donnée?).

Nous allons étudier ici l'influence de six paramètres acoustiques sur la classification des sons dans quatre catégories fonctionnelles de l'automobile, que nous décrirons dans la section suivante. Ces paramètres sont extraits de la bibliographie de l'urgence perçue et de la communauté ICAD et sont détaillés et justifiés par la suite dans la présentation de l'expérience proprement dite.

La problématique de cette expérience s'inscrit dans un cadre pragmatique et concret.

Les attendus en terme de design sonore sont définis par rapport à ces quatre catégories. Autrement dit, cette expérience doit aboutir à des recommandations de design sonore concrètes et fondées sur la pratique. Le propos ici est donc de tester expérimentalement certaines hypothèses issues de la littérature sans chercher à comprendre *pourquoi*, par exemple, certains sons évoquent l'urgence plus que d'autres.

5.2 Catégories fonctionnelles de l'automobile

Les études issues de la bibliographie sont circonscrites d'une part à des applications dans l'aviation civile ou militaire et hospitalières (études d'urgence perçue), et d'autre part à des applications informatiques, pour la plupart (études de la communauté ICAD). Le contexte automobile diffère de ces recherches, au moins quant aux *fonctions* relatives aux IHM sonores.

Une étude interne Renault a permis de spécifier différentes catégories fonctionnelles, en accord avec les attentes des clients, en se restreignant aux informations et aux alertes directement liées à la conduite. Les catégories fonctionnelles ainsi définies par Renault devront nous servir de support pour l'expérience décrite dans ce chapitre.

Ces catégories se déclinent suivant deux "échelles", ou deux axes : le premier qui permet de définir la "raison du son" - "pourquoi ce son survient-il?" - (oubli, défaillance technique, action du conducteur), et le second qui définit la gravité ou l'urgence du son, sa fonction (alerter, signaler ou informer). La "cause du son" peut être imputable au conducteur (rappel, oubli du conducteur) ou non (défaillance du véhicule), ou encore résultante d'une action volontaire du conducteur. Les trois fonctions liées à l'urgence ont également été définies précisément par le groupe de travail Renault. Sur une échelle allant du danger extrême à l'absence totale de danger, on distingue "Alerter d'un danger" (avertir en cas de danger pour que des mesures soient prises), Signaler (faire remarquer ou connaître en attirant l'attention et en mettant en garde – appel à la prudence), et Informer (mettre au courant, faire part, faire savoir). Quatre catégories, issues d'une "combinaison" de ces deux axes, ont dû être utilisées dans le cadre de l'Expérience 1. Par ordre d'urgence décroissante, ces quatre catégories sont les suivantes :

- Alerter d'un danger lié à un oubli du conducteur (*Catégorie 1*),
 - *ex.* : oubli porte ouverte en roulant,
- Alerter d'un danger lié à une défaillance technique du véhicule (*Catégorie 2*),
 - *ex.* : surchauffe moteur,

- Signaler un oubli sans risque (*Catégorie 3*),
 - *ex.* : oubli des phares allumés,
- Informer (*Catégorie 4*),
 - *ex.* : confirmation d'action.

Il est intéressant de remarquer que la catégorie "Alerter d'un danger lié à un oubli du conducteur" est considérée comme plus urgente que la catégorie "Alerter d'un danger lié à une défaillance technique du véhicule". Cette hiérarchie s'établit en considérant le comportement attendu du conducteur : une réaction plus rapide est attendue d'un son de la catégorie "Alerter d'un danger lié à un oubli du conducteur" que d'un son de la catégorie "Alerter d'un danger lié à une défaillance technique du véhicule".

Ces quatre fonctions peuvent être décrites comme un agencement, ou une association des fonctions déjà étudiées dans les deux grands domaines de la littérature - urgence perçue et ICAD. Quatre niveaux d'urgence sont spécifiés, s'apparentant ainsi à une étude de l'urgence perçue, et explicités par l'item "cause du son", se rattachant ainsi davantage à des fonctions utilisées dans les interfaces informatiques.

Ce sont ces quatre fonctions qui servent de base à l'Expérience décrite ci-après.

5.3 Hypothèses expérimentales

Nous soulignons ici que, bien que les finalités de ce premier travail expérimental soient multiples, nous allons chercher à y répondre en une seule et même expérience. Le propos n'est pas d'analyser tous les paramètres possibles un à un ; ce travail a déjà été mené par Edworthy et coll. (1991), et validé avec succès maintes fois (Hellier et coll., 1993 ; Guillaume et coll. 2003). Il convient davantage ici de compiler des résultats expérimentaux ainsi que des premières propositions de design sonore, puis de les vérifier, *ensemble*, dans un cadre empirique et pratique rationnel, tout en restant en accord avec l'environnement automobile voulu. En d'autres termes, nous devons valider et étendre des propositions de design sonore existantes pour les quatre catégories fonctionnelles de l'automobile.

Les stimuli de l'expérience décrite ci-après sont décrits selon le modèle de Patterson (1982). Nous allons donc spécifier tout d'abord les paramètres relatifs à l'élément unitaire de l'alarme, l'impulsion, puis ceux concernant le train d'impulsions, et enfin les paramètres de l'alarme finale (composée de plusieurs trains d'impulsions).

Les six paramètres exposés ci-après ont été déterminés afin de vérifier ou d'infirmier certaines hypothèses expérimentales issues des quelques études majeures dans les domaines

de l'urgence perçue et de la communauté ICAD. Nous cherchons à estimer l'influence de ces six paramètres *conjointement* sur le classement en quatre catégories d'IHM. Dès lors, nous présumons que deux types d'effets peuvent être observés : les effets principaux des paramètres, et des interactions entre eux.

5.3.1 Paramètres descriptifs de l'impulsion

Quatre paramètres spécifiques à l'impulsion ont été choisis : le timbre, la F0, la modulation et la polyphonie.

Fréquence fondamentale (F0)

La F0 a été étudiée sous deux angles bien distincts, adoptés d'une part par la communauté ICAD et d'autre part par celui de l'urgence perçue. Notons ici que nous avons regroupé sous la même rubrique la *F0* étudiée comme paramètre potentiel des alarmes (Edworthy et coll., 1991 ; Hellier et coll., 1993), et la *hauteur*, étudiée par la communauté ICAD. En effet, dans le cas de sons complexes périodiques, la hauteur est considérée comme l'attribut perceptif lié à la F0 - malgré quelques exceptions à cette règle simple (Moore, 2003). Nous choisissons ici de présenter ces deux paramètres simultanément, tout en gardant présent à l'esprit la différence entre paramètre physique du son et attribut perceptif correspondant.

Brewster et coll. (1994) recommandent d'utiliser de "grandes différences" de hauteur entre les sons (2 octaves minimum) afin de reconnaître et séparer véritablement les *earcons* entre eux.

Edworthy et coll. (1991) ont montré que plus la F0 est élevée, plus l'urgence perçue est grande (avec des F0 de 150 Hz, 200 Hz, 350 Hz et 530 Hz). Par la suite, Hellier et coll. (1993) ont quantifié l'importance relative de la F0, qui apparaît alors comme le deuxième paramètre le plus important dans le jugement d'urgence, après la vitesse de l'alarme.

L'objectif principal de notre expérience est de comparer l'importance de la F0 relativement aux autres paramètres de l'expérience. Si la proposition de Brewster et coll. (1994) est vérifiée, la F0 ne devrait avoir que peu d'influence sur le classement des sons. A l'inverse, si l'hypothèse d'Hellier et coll. (1993) est avérée, la F0 devrait être le deuxième paramètre conditionnant le classement final (puisque les quatre catégories fonctionnelles suivent une échelle d'urgence). D'autre part, nous devrions observer un classement des sons tel que l'urgence perçue augmente avec la F0 (Edworthy et coll., 1991) : les F0 faibles devraient

être associées aux catégories de non urgence ou de faible urgence (catégorie 3 et 4), alors que les F0 élevées devraient être associées aux catégories d'urgence (1 et 2).

Timbre

Le *timbre* est proposé par Blattner et coll. (1989) comme étant la deuxième "dimension" suivant laquelle la hiérarchie des *earcons* doit être créée (la première "dimension" étant rythmique). Brewster et coll. (1994) ont montré que les timbres musicaux facilitent la reconnaissance des *earcons*, et recommandent d'utiliser préférentiellement des timbres musicaux, ou du moins complexes (comparativement à des sons purs).

Edworthy et coll. (1991) ont, pour leur part, testé deux paramètres que nous pouvons rattacher à la définition de "timbre" : la série harmonique et la forme de l'enveloppe. Hellier et coll. (1993) ont repris le paramètre de série harmonique, et ont ainsi montré que ce paramètre était secondaire voire négligeable dans l'urgence perçue des alarmes.

Constatons ici la polysémie du mot "timbre" et l'amalgame possible entre ces différentes utilisations. Une définition communément acceptée du timbre le propose comme un attribut perceptif n'étant ni la hauteur, ni la sonie, et qui permet à un sujet de juger que deux sons présentés de façon similaire sont semblables (ANSI, 1973). McAdams et coll. (1995) ont étudié les différentes dimensions perceptives du timbre, et ont caractérisé le timbre des sons complexes suivant trois descripteurs : le centre de gravité spectral (CGS), le logarithme du temps d'attaque (LTA) et le flux spectral. La série harmonique définie par Edworthy et coll. (1991) rappelle principalement le CGS, et la forme de l'enveloppe évoque en partie le LTA.

Il semblait utopique - et peu utile à première vue - de vouloir tester tous les descripteurs du timbre, leurs importances relatives et leurs impacts sur le classement des sons au sein d'une même expérience. Nous avons donc privilégié dans notre étude la comparaison d'un paramètre "générique" appelé *timbre* aux cinq autres paramètres de l'expérience.

Les timbres créés pour cette expérience respectent en partie les recommandations de Brewster et coll. (1994). Ce sont bien des sons complexes, mais non "musicaux". La reconnaissance d'un instrument de musique spécifique n'est pas recherchée, au contraire. En effet, il est probable que la reconnaissance de la source d'un son perturbe le jugement demandé aux sujets, et converge davantage vers un classement des sons suivant les sources qui les créent que suivant la fonction transmise par ce son (voir Guyot, 1996). Le détail des sons sera donné en section 5.4.2.

Si l'hypothèse de Brewster et coll. (1989) est validée, le classement des différentes IHM

sonores devrait s'effectuer grâce aux paramètres rythmiques (cf. ci-après), et en deuxième lieu, suivant des changements de timbre. Dans le cas contraire, si l'hypothèse d'Hellier et coll. (1993) est vérifiée, le timbre ne devrait pas apparaître comme paramètre prédominant dans le classement des sons.

Modulation d'amplitude

La modulation d'amplitude, et plus précisément la fréquence de modulation d'un son modulé sinusoïdalement en amplitude, est un paramètre suggéré par Patterson et coll. (1986).

Nous avons conscience de l'assimilation probable entre une variation physique imposée au signal comme la modulation et un attribut perceptif comme le timbre. Par surcroît, la rugosité - influencée grandement par la fréquence de modulation (Pressnitzer, 1998) - peut également être vue comme une dimension du timbre.

L'objectif premier de cette expérience est d'étudier - *simultanément* - l'effet de plusieurs composantes issues de littératures très différentes sur un classement fonctionnel des sons. La compilation de ces études hétéroclites engendre inévitablement des ambiguïtés dans les termes employés. Nous cherchons à explorer d'une part l'influence comparée du timbre (relativement aux autres paramètres) sur le classement des sons, et d'autre part, l'effet potentiel de la modulation d'amplitude.

Polyphonie

La polyphonie est un paramètre issu de la littérature ICAD, et proposé par Brewster et coll. (1994) sous l'appellation d'"*earcons* multi-timbres". Il en déduit des solutions tout aussi pragmatiques que les précédentes : utiliser de "grandes différences" dans les multi-timbres employés afin de faciliter la reconnaissance des earcons.

Kramer (1994), pour sa part, suggère d'utiliser la polyphonie pour rendre les sons plus "compréhensibles" et transmettre ainsi plusieurs informations simultanément.

Nous nous proposons donc d'introduire ce paramètre dans notre plan expérimental afin de tester ces postulats intuitifs. Ont-ils une confirmation expérimentale? Les sons polyphoniques permettent-ils un classement clair et sans équivoque?

5.3.2 Paramètres descriptifs du train d'impulsions : le rythme

Le "rythme" a été avancé par Blattner et coll. (1989) et Brewster et coll. (1994) comme paramètre dominant pour la création d'earcons.

En parallèle, divers paramètres rythmiques ont été testés par Patterson (1982), Edworthy et coll. (1991) et Hellier et coll. (1993). Ils ont ainsi démontré qu'une urgence élevée était associée à une durée d'impulsion de 100 - 200 ms, à un minimum de cinq impulsions par train d'impulsions, et à une vitesse élevée (intervalle inter-impulsion inférieure à 150 ms). Si l'intervalle inter-impulsion excède les 300 ms, l'urgence associée à l'alarme diminue. De plus, il a été montré que l'intervalle inter-impulsion d'une alarme est le paramètre prédominant dans le jugement d'urgence perçue.

Nous avons cherché dans cette expérience à évaluer l'influence respective de chacun de ces paramètres entre eux, et à les comparer aux autres paramètres de l'expérience. L'intervalle inter-impulsion est-il bien le premier critère pour définir l'urgence d'une alarme ? La durée de l'impulsion est-elle réellement importante en soi ? Quelle est la portée du nombre d'impulsions ?

5.3.3 Paramètre descriptif de l'alarme finale : la répétition

Il est également recommandé de répéter un train d'impulsions quatre fois pour atteindre un niveau d'urgence élevé (Edworthy et coll., 1991 ; Hellier et coll., 1993).

Nous cherchons ici à tester des paramètres initialement introduits comme paramètres du train d'impulsions. En effet, cette dissociation des paramètres en impulsion, train d'impulsions et alarme (Patterson, 1982) est arbitraire, et semble concorder principalement à des recommandations *pratiques* de design.

Ainsi, nous proposons de tester d'une part l'effet de l'intervalle inter-répétition (i.e. intervalle de temps entre deux trains d'impulsions), et d'autre part l'effet d'un changement de vitesse au sein de l'alarme (i.e. une accélération de la répétition des trains).

Si la vitesse de l'alarme (qu'elle soit régulière ou avec des changements de rythme) a un effet sur la perception des alarmes, nous devrions observer cet effet quelle que soit la manière d'introduire des changements de vitesse.

5.4 Méthode

5.4.1 Participants

Soixante trois sujets (30 femmes et 33 hommes, hors personnel Renault et hors personnel IRCAM) ont pris part à cette expérience. Les auditeurs ont déclaré n'avoir aucun problème d'audition. La plupart d'entre eux ont été recrutés à partir d'une base de données de volontaires et n'avaient jamais pris part à une expérience de psychoacoustique. Tous ont été rémunérés. Les auditeurs étaient également sélectionnés s'ils avaient une pratique régulière de la conduite.

5.4.2 Stimuli

Nous présentons ici les six paramètres acoustiques issus des hypothèses expérimentales. Les quatre paramètres de l'impulsion prennent chacun deux valeurs ; les deux paramètres qui définissent le train d'impulsions et l'alarme prennent chacun quatre valeurs. Afin de tester les stimuli produits par les croisements des modalités de ces paramètres, nous avons utilisé un plan expérimental bien particulier détaillé ci-après. En effet, l'utilisation d'un plan expérimental complet - soit 256 sons - est impossible à vérifier expérimentalement. Les 32 stimuli dérivés de ce plan expérimental sont décrits dans le tableau 5.2.

Paramètres acoustiques

Nous détaillons pour chaque paramètre les différentes modalités utilisées, ainsi que les notations qui seront utilisées dans la suite du document.

Nous remarquons ici, une fois de plus, que les termes employés peuvent prêter à confusion, du fait de l'amalgame - sémantique - entre attribut physique du son et attribut perceptif. Nous considérons ces notations comme purement "symboliques", et ces abus de langage sont employés par commodité : pour pouvoir parler simplement des résultats, tout d'abord ; et de surcroît, pour tenter de représenter un espace de sons résumant le mieux possible des travaux très différents. Nous avons parfaitement conscience que cet exercice est bien délicat. Ces aspects seront débattus plus longuement dans la Discussion.

Paramètres de l'impulsion

- Fréquence fondamentale (F_0)
 - Modalité $F_0=110\text{Hz}$, notée $A1$ (suivant la notation anglo-saxonne)

- Modalité F0=440Hz, notée *A3*
- Timbre (*T*)

Les "sons de base" utilisés dans cette expérience ont été créés grâce au logiciel Modalys (Adrien, 1991), qui permet de créer des sons à partir de modèles physiques, basés sur la représentation modale d'objets vibrants. Les sons produits par ce type de synthèse sonore sont généralement réalistes, puisque définis à partir de contraintes mécaniques réelles. Nous avons cependant utilisé un des modules de Modalys étendant les règles de création à des instruments appelés "hybrides", et qui ne sont donc plus tout à fait réels d'un point de vue physique. L'avantage était de pouvoir créer assez simplement des sons d'une grande complexité, sans qu'ils rappellent pour autant spécifiquement un instrument pré défini (voir page 82 ci-dessus).

 - Timbre créé à partir d'un modèle physique de corde, noté *String*
 - Timbre créé à partir d'un modèle physique de plaque, noté *Plate*
- Polyphonie (*P*)

Les hypothèses de la littérature concernant ce paramètre étant assez légères, nous avons choisi de comparer simplement deux modalités totalement opposées de ce paramètre, i.e. présent ou non. S'il s'avère par la suite que la polyphonie a bien un effet, nous pourrions alors étudier plus en détail différentes modalités de ce paramètre.

 - Pas de polyphonie - une seule note, qui correspond à la F0. Notée *No*
 - Accord de 3 notes (A1 - C2 - E2 ou A3 - C4 - E4). Noté *Yes*
- Modulation (*M*)

Nous pouvons faire ici les mêmes commentaires que pour la Polyphonie, quant à l'utilisation de deux modalités bien distinctes.

 - Pas de modulation. Notée *No*
 - Modulation avec une profondeur de modulation de 0.8 et une fréquence de modulation de 60 Hz. Notée *Yes*.

Paramètre du train d'impulsion

- Paramètre rythmique (noté *Signal.Silence* ou *SS* par la suite)

Ce paramètre rythmique comprend trois sous paramètres : durée de l'impulsion (DI), intervalle inter-impulsion (III) et nombre d'impulsions (NI). D'après les hypothèses expérimentales présentées ci-dessus, nous pouvons ordonner les modalités de ce paramètre sur une échelle d'urgence. Cet ordre devra être validé grâce à l'expérience. De la moins à la plus urgente, les modalités ordonnées sont donc les suivantes :

- NI = 1, DI = 200 ms. Notée *1S* (pour 1 Short)
- NI = 1, DI = 2000 ms. Notée *1L* (pour 1 Long)
- NI = 5, DI = 200 ms, III = 400 ms. Notée *5S* (pour 5 Slow)
- NI = 5, DI = 200ms, III = 50ms. Notée *5F* (pour 5 Fast)

Paramètre de l'alarme

- Répétition (*R*)

Ce paramètre concerne la répétition du train d'impulsions. Les modalités de ce paramètre peuvent être considérées comme ordonnées, l'ordre intuitif étant celui d'une répétition de plus en plus rapide, de la même façon qu'un train d'impulsions dont les impulsions sont répétées de plus en plus vite est perçu comme urgent. Les modalités ordonnées sont les suivantes :

- Modalité sans répétition. Notée *No*
- 4 motifs répétés toutes les 2,5 s. Notée *2.5s*
- 4 motifs répétés en accélérant. Les intervalles inter train sont les suivants : 2,5 s / 1,2 s / 0,6 s. Notée *accel*
- 4 motifs répétés toutes les 0,6 s. Notée *0,6s*

Plan d'expérience

Le plan d'expérience utilisé est de type Hyper Gréco Latin Géométrique (HGLG), appelé HGLG26. Sa construction est détaillée par Benoist et coll. (1994). Ce plan est défini par 4 variables à 2 niveaux et 2 variables à 4 niveaux, qui aboutissent à 32 essais. Ici, les facteurs du plan sont les paramètres acoustiques (Signal.Silence, Répétition, etc.) ; les essais sont donc les 32 sons utilisés. Un plan de résolution 4 permet l'analyse de toutes les variables et d'une partie des interactions d'ordre 2 (i.e. interactions entre 2 facteurs). Un plan de résolution 5 aurait permis l'analyse de toutes les interactions d'ordre 2. Cependant, ce type de plan est très coûteux en temps, voire impossible à exploiter dans un cadre expérimental, du fait de la croissance extrêmement rapide du nombre de sons si l'on souhaite prendre en considération toutes les interactions entre paramètres.

Certaines interactions sont confondues, et il est utile de préciser lesquelles afin d'en tenir compte dans la mise au point des modèles statistiques, par la suite. L'interaction du paramètre *x* avec le paramètre *y* est notée *x×y*. Les interactions confondues sont récapitulées dans le tableau 5.1, en respectant la notation des paramètres acoustiques introduite à la section précédente.

L'interaction	Est confondue avec les interactions suivantes
SS×R	T×F0/T×P/T×M/F0×P/F0×M/P×M
SS×T	R×F0/R×P/R×M
SS×F0	R×T/R×P/R×M
SS×P	R×T/R×F0/R×M
SS×M	R×T/R×F0/R×P
R×T	SS×F0/SS×P/SS×M
R×F0	SS×T/SS×P/SS×M
R×P	SS×T/SS×F0/SS×M
R×M	SS×T/SS×F0/SS×P
T×F0	P×M/SS×R
T×P	F0×M/SS×R
T×M	F0×P/SS×R
F0×P	SS×R/T×M
F0×M	SS×R/T×P
P×M	SS×R/T×F0

TAB. 5.1 – Tableau des interactions confondues dans le plan d'expérience utilisé. SS désigne le paramètre Signal.Silence, R la Répétition, T le Timbre, P la Polyphonie et M la Modulation

Il apparaît ainsi que seules certaines interactions peuvent être prises en compte dans les modèles. En effet, l'orthogonalité de ce plan d'expérience permet d'observer tous les effets principaux ainsi que toutes les interactions de la colonne de gauche. Il n'est toutefois pas possible de mettre au point un modèle contenant à la fois une interaction de la colonne de gauche, et une autre de la colonne de droite. Par exemple, l'interaction $F0 \times T$, ne peut pas être incluse dans un modèle contenant déjà l'interaction $SS \times R$. L'implication pratique de ces interactions confondues sera détaillée dans la suite du chapitre.

Stimuli utilisés

Les 32 stimuli qui découlent de ce plan d'expérience sont présentés dans la Table 5.2.

	Signal.Silence	Répétition	Timbre	F0	Polyphonie	Modulation
1	5S	no	String	A1	no	no
2	5S	no	Plate	A3	yes	yes
3	5S	2.5s	String	A3	no	yes
4	5S	2.5s	Plate	A1	yes	no
5	5S	0.6s	String	A1	yes	yes
6	5S	0.6s	Plate	A3	no	no
7	5S	accel	String	A3	yes	no
8	5S	accel	Plate	A1	no	yes
9	5F	no	String	A1	yes	yes
10	5F	no	Plate	A3	no	no
11	5F	2.5s	String	A3	yes	no
12	5F	2.5s	Plate	A1	no	yes
13	5F	0.6s	String	A1	no	no
14	5F	0.6s	Plate	A3	yes	yes
15	5F	accel	String	A3	no	yes
16	5F	accel	Plate	A1	yes	no
17	1S	no	String	A3	no	yes
18	1S	no	Plate	A1	yes	no
19	1S	2.5s	String	A1	no	no
20	1S	2.5s	Plate	A3	yes	yes
21	1S	0.6s	String	A3	yes	no
22	1S	0.6s	Plate	A1	no	yes
23	1S	accel	String	A1	yes	yes
24	1S	accel	Plate	A3	no	no
25	1L	no	String	A3	yes	no
26	1L	no	Plate	A1	no	yes
27	1L	2.5s	String	A1	yes	yes
28	1L	2.5s	Plate	A3	no	no
29	1L	0.6s	String	A3	no	yes
30	1L	0.6s	Plate	A1	yes	no
31	1L	accel	String	A1	no	no
32	1L	accel	Plate	A3	yes	yes

TAB. 5.2 – Définition des 32 stimuli utilisés dans l'expérience selon les paramètres Signal.Silence, Répétition, Timbre, F0, Polyphonie et Modulation

Expérience préliminaire : égalisation en sonie

Au préalable, les sons ont été égalisés en sonie, lors d'une expérience d'ajustement. Étant donné la diversité et la complexité des sons dont nous disposons, les modèles computationnels classiques de sonie étaient assez peu adaptés ; il était donc nécessaire d'effectuer cette égalisation à l'aide d'une expérience.

Le matériel utilisé pour cette pré-expérience est le même que pour l'expérience principale. Il était demandé aux sujets d'ajuster chacun des 32 sons par rapport à un son de référence, choisi hors corpus (pour une justification du choix d'un son hors corpus comme référence, voir Berglund et coll., 1975). Le son de référence créé est défini par une impulsion de durée d'1s, de timbre *String*, sans polyphonie et sans modulation). Il est diffusé à 80 dB.

A l'issue de cette expérience, chacun des signaux est alors multiplié par un coefficient (moyenne des coefficients extraits des réponses de chacun des sujets) qui lui est propre et qui ajuste son amplitude de manière à ce que la sonie soit identique pour les 32 sons.

5.4.3 Matériel

Les 32 sons étaient reproduits à travers une carte RME HDSP9652, des convertisseurs analogique / numérique RME ADI-8 PRO et ont été amplifiés par un amplificateur stéréo Yamaha P2075. Ils étaient présentés aux sujets pour une écoute binaurale à travers un casque Sennheiser HD 250 linear II. Le test s'est déroulé dans la cabine audiométrique insonorisée de l'Ircam. La conduite de l'expérience ainsi que la collecte des données par l'intermédiaire d'une interface graphique étaient assurées par un programme écrit dans l'environnement PsiExp sur un ordinateur Apple Macintosh (Smith, 1995).

5.4.4 Procédure

Les auditeurs étaient informés qu'ils allaient devoir ordonner des sons suivant quatre catégories pré-définies. Les catégories fonctionnelles étaient alors définies grâce à une consigne écrite et explicitées avec quelques exemples (non sonores) de fonctions connues, extraites de l'univers automobile : "oubli d'extinction des phares", "oubli de la ceinture de sécurité", "liquide frein défaillant", etc. La hiérarchie de ces catégories selon leur degré d'urgence était bien précisée. Une fois que les auditeurs déclaraient avoir compris précisément chacune des catégories, l'expérience pouvait débuter. Il était demandé aux sujets d'écouter

une première fois l'ensemble des stimuli, pour qu'ils aient connaissance de l'espace perceptif qu'ils auront à juger. Les sons étaient présentés dans un ordre aléatoire. Les auditeurs devaient alors classer les sons dans la catégorie qui leur semblait la plus adéquate. Avant de répondre, les sujets pouvaient, s'ils le désiraient, réécouter les stimuli autant que nécessaire. Ils pouvaient également, au besoin, modifier le classement d'un son *a posteriori*.

L'expérience totale durait approximativement 1h.

5.5 Résultats

Avant d'utiliser des méthodes complexes pour modéliser les résultats de l'expérience, il est primordial d'effectuer quelques analyses statistiques simples afin de décrire le jeu de données dont on dispose (section 5.5.1), et de s'assurer du bon accord des sujets entre eux (section 5.5.2 page 98). Ensuite, des méthodes dites de segmentation seront étudiées et leurs résultats analysés en détail (section 5.5.3 page 100). Ces méthodes descriptives, de fait, constituent des approches décisionnelles mais participent peu à la compréhension du phénomène étudié. La mise au point d'un modèle de régression logistique a donc été privilégiée et sera enfin exposée dans la présentation de ces résultats (section 5.5.4 page 103).

5.5.1 Analyses descriptives

Analyse graphique

En première analyse, il est possible d'observer la distribution des réponses données par les sujets. La catégorie 4 (non danger) n'est que peu représentée, alors que la catégorie 2 (danger lié à une défaillance) est la réponse la plus souvent donnée par les sujets. En effet, sur les 2016 réponses totales - i.e. 32 sons classés par 63 sujets - 24% d'entre elles sont imputables à la catégorie 1 (danger extrême, lié à un oubli), 32% à la catégorie 2 (danger lié à une défaillance), 27% à la catégorie 3 (signaler un oubli) et 17% à la catégorie 4 (non danger).

Le consensus sur les réponses peut également être évalué (ici, de manière globale). Les 32 sons sont examinés en fonction du classement déterminé par les sujets pour chacun d'entre eux. La figure 5.1 donne un exemple d'histogrammes des classes obtenues pour les sons 9 à 16. Certains sons semblent plus "compréhensibles" que d'autres. En effet, pour certains, le classement obtenu est sans équivoque, avec une catégorie très fortement

représentée. Le son 14 par exemple, est classé par plus de 80% des sujets dans une seule catégorie. A l'inverse, d'autres sons sont répartis presque uniformément entre les quatre catégories (sons 9 et 10, par exemple).

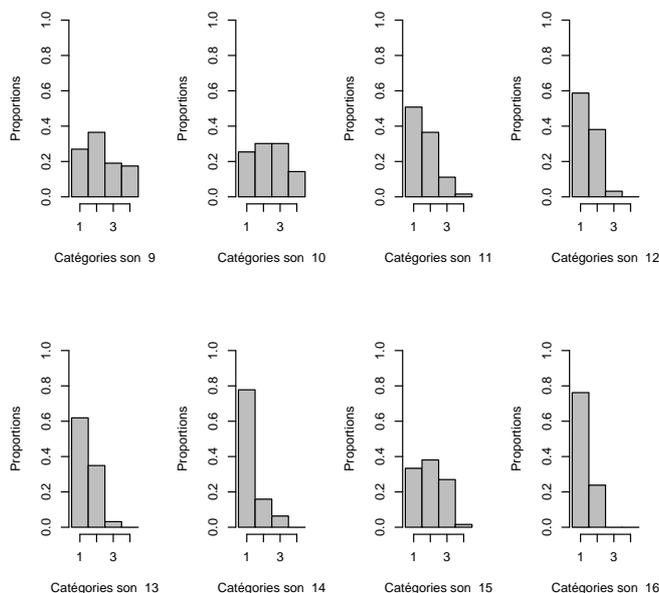


FIG. 5.1 – Histogramme de la classification des sons n°9 à 16

Tables de contingence et tests d'indépendance pour variables qualitatives

Quelques notions complémentaires sur les tables de contingence et les mesures d'association entre variables sont rappelées en Annexe A.

Tables de contingence Une méthode simple et efficace pour résumer une partie de l'information contenue dans les données est de construire des tables de contingence. On construit donc des tables à deux entrées résumant le lien entre la réponse catégorielle et la variable considérée.

Le jeu de données considéré est constitué des 32 sons classés par les 63 sujets séparément. L'*individu* au sens statistique du terme correspond donc à la réponse de l'individu i au son j . Il y a donc 63×32 (2016) individus statistiques. Les 32×63 réponses sont alors utilisées pour construire ces tables.

Classification	Variable Signal.Silence			
	1S	1L	5S	5F
1	23	65	136	259
2	101	149	232	160
3	204	164	118	63
4	176	124	18	22

TAB. 5.3 – Table de contingence : répartition des sons en fonction des modalités du paramètre Signal.Silence

Classification	Variable Répétition			
	no	2.5s	accel	0.6s
1	70	125	121	167
2	115	159	182	186
3	105	175	145	124
4	214	45	56	27

TAB. 5.4 – Table de contingence : répartition des sons en fonction des modalités du paramètre Répétition

Les tableaux 5.3 et 5.4 présentent les tables de contingence croisant la réponse et les variables Signal.Silence et Répétition respectivement. En examinant pour chaque modalité de la réponse la modalité de la variable la plus représentée, nous pouvons constater un lien très clair. Dans le cas de la variable Signal.Silence, la classe 1 est majoritairement représentée pour la modalité *5F* (cinq sons rapides), tandis que la classe 4 l'est davantage pour la modalité *1S* (un son court). De même, dans le cas de la variable Répétition, les modalités *No* et *2.5s* sont associées principalement aux catégories 4 et 3 respectivement, *a contrario* de la modalité *0.6s* majoritaire pour la catégorie 1. La répartition est moins claire dans le cas de la catégorie 2, malgré un maximum pour les modalités *accel* et *0.6s*. Précisons que ce ne sont que de premières observations, qui se doivent d'être validées ou infirmées par des calculs plus précis (section suivante).

Les tables de contingence des variables F0, Timbre, Polyphonie et Modulation sont présentées dans les Tableaux 5.5, 5.6, 5.7 et 5.8.

La variable F0 semble également avoir une influence sur la réponse donnée par les

Classification	F0	
	A1	A3
1	280	203
2	360	282
3	228	321
4	140	202

TAB. 5.5 – Table de contingence : répartition des sons en fonction des modalités du paramètre F0.

Classification	Timbre	
	String	Plate
1	188	295
2	349	293
3	282	267
4	189	153

TAB. 5.6 – Table de contingence : répartition des sons en fonction des modalités du paramètre Timbre.

sujets : les catégories 1 et 2 sont plus représentées pour la modalité *A1* de ce paramètre, alors que les catégories 3 et 4 sont plus souvent associées à la modalité *A3*.

La variable Timbre semble avoir moins d'effet sur le classement des sons, les différences dans la répartition des réponses suivant les modalités de la variable sont plus faibles que précédemment. Toutefois, la modalité *Plate* apparaît davantage pour la catégorie 1. De même, malgré un effet faible de la variable Polyphonie, nous pouvons relever qu'un son polyphonique est plus souvent classé dans la catégorie 1. De manière encore moins flagrante, un son est plus souvent classé dans la catégorie 1 s'il est modulé.

Ayant ainsi obtenu les tables de contingence, il devient possible d'appliquer des tests statistiques plus sophistiqués afin de préciser et de valider (ou d'infirmer) les commentaires précédents.

Tests d'indépendance Les deux tests d'indépendance utilisés dans cette section (test du khi-deux et test du rapport de vraisemblance) sont détaillés en Annexe A.

Classification	Polyphonie	
	oui	non
1	283	200
2	300	342
3	259	290
4	166	176

TAB. 5.7 – Table de contingence : répartition des sons en fonction des modalités du paramètre Polyphonie.

Classification	Modulation	
	oui	non
1	213	270
2	349	293
3	279	270
4	167	175

TAB. 5.8 – Table de contingence : répartition des sons en fonction des modalités du paramètre Modulation.

Les résultats de ces deux tests calculés pour chaque table de contingence sont présentés dans le tableau 5.9.

La p-value indiquée dans cette table représente la probabilité qu'une variable suivant une loi de χ^2 avec d degrés de liberté dépasse la valeur calculée par le test. Cette valeur est comprise entre 0 et 1. Une variable associée à une grande p-value signifie que la statistique calculée a de grandes chances de suivre effectivement une loi de χ^2 . Étant donné que cette loi est de type χ^2 uniquement sous l'hypothèse d'indépendance, une grande p-value confirmera une hypothèse d'indépendance des variables, tandis qu'une faible p-value nous entraînera à refuser cette hypothèse. Plus la p-value est petite, plus il est raisonnable d'accorder du poids à la variable associée, sans aucune certitude pour autant.

Les deux statistiques calculées suivent des lois de χ^2 sous l'hypothèse d'indépendance. Ces deux tests donnent des valeurs similaires et confirment donc mutuellement leurs résultats. Il devient ainsi très clair que toutes les variables sont significatives, même si elles n'ont pas toutes la même importance.

Variable	Test du χ^2	Test du rapport de Vraisemblance	Degrés de liberté	p-value pour le test du χ^2
Répétition	341.23	306.89	9	$< 2.2e-16$
Signal.Silence	614.77	645.90	9	$< 2.2e-16$
F0	48.75	48.96	3	$1.48e-10$
Modulation	11.95	11.97	3	0.008
Timbre	32.79	32.99	3	$3.6e-07$
Polyphonie	19.05	19.13	3	0.0003

TAB. 5.9 – Tests du khi-deux et du rapport de vraisemblance pour chaque table de contingence

Ainsi, les futures sélections de modèles statistiques accorderont de l'importance aux variables Signal.Silence et Répétition (p-value très faibles), un peu moins aux variables F0 et Timbre, et une importance faible à la Polyphonie et à la Modulation. La variable Modulation est la moins influente, et il peut être envisageable de l'exclure des modèles, un compromis étant toujours nécessaire entre complexité d'un modèle et qualité de l'explication fournie.

Deux variables apparaissent comme étant fortement reliées à la réponse catégorielle : Signal.Silence et Répétition. Ce résultat était déjà apparent dans les tables de contingence ; les tests effectués le confirment. Il est vraisemblable que ces deux variables soient capables d'expliquer une grande part de la variabilité. Une autre mesure d'association va alors permettre de mesurer la force du lien.

Coefficient d'association entre variables ordinales Le coefficient le plus souvent utilisé pour mesurer des dépendances monotones entre les lignes et les colonnes d'un tableau de contingence est le γ de Goodman et Kruskal. Ce coefficient est compris entre -1 et 1. Si les variables sont indépendantes, alors $\gamma = 0$. Plus γ est grand, plus il y a un lien entre les deux variables qualitatives. Si γ est négatif, alors la liaison est négative et le lien est de type linéaire décroissant. En d'autres termes, plus les modalités de la variable augmentent, plus celles de la réponse diminuent.

Dans le cas présent, nous obtenons $\gamma = -0.58$ pour Signal.Silence et $\gamma = -0.32$ pour Répétition. Le classement présupposé sur une échelle d'urgence des modalités de Signal.Silence et Répétition (page 86) est étayé par ces liaisons négatives : on passe de la catégorie 1 à

la catégorie 4 lorsque Signal.Silence passe des modalités *5F* à *1S* et Répétition de *0.6s* à *No*. Enfin, nous pouvons constater que la force du lien est plus élevée pour la variable Signal.Silence que pour la variable Répétition.

Coefficient kappa de Cohen La dernière mesure d'association traitée est le kappa de Cohen. Ce coefficient s'applique uniquement dans un contexte bien particulier. Supposons que plusieurs juges ou observateurs doivent classer N objets. Ce classement se fait, au moins partiellement, sur la base de critères subjectifs. Ce qui implique que les juges pourront classer les objets différemment. Le coefficient κ mesure maintenant le degré d'accord entre les juges. Il est défini d'une façon générale par :

$$\kappa = \frac{p_A - p_E}{1 - p_E} \quad (5.0)$$

où p_A est la probabilité observée (extraite des données) que les sujets émettent le même jugement, et p_E est la probabilité que les juges soient d'accord au cas où ils jugeraient au hasard. Le coefficient κ est compris entre 0 et 1. En cas d'accord parfait entre les juges, κ égal 1. S'ils jugent de façon aléatoire, κ est égal à 0. Dans notre cas, les juges sont les 63 sujets, et le coefficient κ vaut 0.42.

Ce résultat montre que les jugements effectués par les sujets sont homogènes. D'après la littérature, une telle valeur est associée à un assez bon accord dans les réponses, malgré quelques divergences. En effet, les histogrammes des classes des sons montrent que, dans la majorité des cas, les catégories étaient toutes représentées pour un son donné. De ce fait, la probabilité de ne pas être d'accord entre deux sujets est assez grande, considérant les 4 catégories. Le coefficient kappa trouvé ici est donc très correct.

Résumé des premières analyses

- Les catégories extrêmes 1 et 4 sont bien comprises et discriminées par les sujets, *a contrario* des catégories 2 et 3 qui semblent peu claires ;
- Les variables Signal.Silence et Répétition apparaissent comme les plus importantes, suivies par F0 et Timbre, et enfin par Polyphonie et Modulation qui semblent les moins influentes ;
- Dans le cas de la variable Signal.Silence : l'ordre décroissant des modalités (de *5F* à *1S*) est associé à un ordre de catégorisation croissant (de 1 à 4). En d'autres termes, *5F* est associée à la catégorie la plus urgente, alors que *1S* est associée à la catégorie de non-urgence ;

- Dans le cas de la variable Répétition : l'ordre décroissant des modalités (de *0.6s* à *No*) est associé à un ordre de catégorisation croissant (de 1 à 4). En d'autres termes, *0.6s* est associée à la catégorie la plus urgente, alors que *No* (pas de répétition) est associée à la catégorie de non-urgence ;
- Il existe un assez bon accord entre les sujets.

Toutefois, avant d'aller plus loin dans l'analyse des données, et bien que le coefficient κ observé soit correct, il semble important de tester l'existence potentielle de groupes de sujets.

5.5.2 Analyse des sujets

En effet, la réponse demandée aux sujets ici peut dépendre de facteurs autres que purement perceptif. Dès lors que la question posée relève de la cognition, au sens où un jugement global est demandé, faisant intervenir l'expérience antérieure des sujets, leurs habitudes, un certain apprentissage, etc., il est probable que tous les individus ne réagissent pas de façon identique à un stimulus donné. Il semble alors nécessaire, à ce stade de l'analyse, d'examiner l'existence possible de plusieurs comportements types parmi l'ensemble des sujets. On parlera aussi de typologie des individus.

Précédemment (pour établir des tables de contingence, par exemple ; voir page 92), nous avons considéré l'individu au sens statistique du terme : les 63*32 (2016) individus statistiques.

L'optique choisie pour faire une typologie des sujets est différente. En effet, il s'agit là d'étudier le sujet i en tenant compte de ses 32 réponses, et d'évaluer de manière objective s'il est possible de trouver des groupes de réponses semblables. Pour ce faire, il faut considérer des vecteurs X_i composés des 32 réponses X_{ij} pour j allant de 1 à 32. Les valeurs X_{ij} sont donc les classes dans lesquelles le sujet i a classé le son j .

Plusieurs approches sont possibles afin de classer les réponses des individus. Dans un premier temps, après avoir fait le choix d'une distance appropriée, la dissimilarité entre les réponses a été obtenue. La technique des arbres de classification (et non de discrimination) nous a ensuite permis de classer les distances obtenues à l'aide d'un arbre appelé *dendrogramme*.

Notion de distance

Afin de pouvoir classer de manière objective les réponses des individus, il est nécessaire de définir une distance entre les vecteurs de réponses. Les coordonnées de ces vecteurs sont qualitatives à plusieurs modalités. Une notion de distance adaptée à ce type de données doit prendre en compte la contribution de chaque coordonnée à la mesure de dissimilarité globale. Cette mesure globale de dissimilarité entre deux individus sera la somme (pondérée ou non) des contributions de chaque coordonnée (i.e. de chaque dissimilarité de réponse).

Il existe plusieurs façons de mesurer l'accord entre deux réponses qualitatives, mais il est d'usage de considérer que les dissimilarités calculées doivent être comprises entre 0 et 1.

Dans le cas où l'on dispose de réponses ordinales, nous pouvons définir :

$$\delta_{ik} = \frac{1}{32} \sum_{j=1:32} \frac{|X_{ij} - X_{kj}|}{3}. \quad (5.0)$$

Cette distance sera dénommée Distance 1. Le calcul de la norme $|X_{ik} - X_{kj}|$ donne une valeur comprise entre 0 et 3 (le maximum étant de $4 - 1 = 3$ et le minimum de $1 - 1 = 0$, où 4 et 1 correspondent aux catégories extrêmes). Elle est donc bien comprise entre 0 et 1 (car division par 3) et, contrairement à d'autres mesures qui accordent un poids similaire quelle que soit la nature du désaccord, elle permet d'accorder un poids plus grand aux grandes erreurs commises. La contribution maximale d'un désaccord de jugement est de 1.

Arbres de classification

Une manière de classer les distances est fournie par la théorie des arbres de classification (ou *dendrogrammes*). Les dendrogrammes permettent de représenter les dissemblances entre sujets.

La figure 5.2 montre le dendrogramme associé à la Distance 1 (en utilisant la technique d'analyse "average"). En ordonnée est représentée la mesure de dissemblance entre les sujets. Chaque noeud de l'arbre représente le point où deux classes (au sens statistique) se rejoignent. Pour chaque noeud une mesure de dissemblance est associée. Par exemple, les sujets 12 et 57 (à gauche de l'arbre) sont dissemblants d'une valeur comprise entre 0.3 et 0.4. De la même manière, on peut estimer la dissemblance entre deux sujets, même très éloignés dans l'arbre, en regardant la valeur de dissemblance là où ils ont un noeud en commun. Les sujets 29 (tout à gauche de l'arbre) et 27 (tout à droite) ont donc une dissemblance comprise entre 0.6 et 0.7.

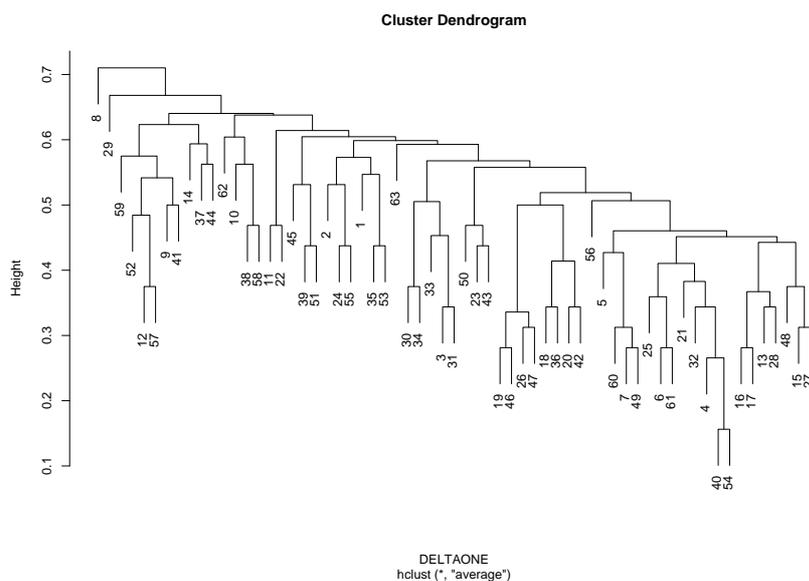


FIG. 5.2 – Arbre de classification des sujets pour la distance 1

Visiblement, les données présentent une telle continuité dans les dissemblances qu'il ne semble pas possible de découper les données en plusieurs groupes de sujets. S'il y avait eu des groupes de sujets, deux par exemple, nous aurions dû observer deux groupes bien distincts tout en bas de l'arbre, qui se regroupent eux-mêmes bien plus haut dans l'arbre. L'aspect de cet arbre de classification indique donc clairement qu'il n'y a pas de groupes de sujets.

Cette analyse complémentaire associée au Kappa de Cohen permet de conclure que pour cette expérience, les sujets ont classé les sons de manière similaire, et qu'il n'existe pas de groupe de sujets répondant de manière différente de la majorité.

Il est donc possible de poursuivre notre analyse sur les 63 sujets au global.

5.5.3 Arbres de discrimination

Définition

Les arbres binaires de discrimination sont complémentaires de méthodes statistiques plus classiques (analyse des correspondances ou régression linéaire). Les solutions obtenues sont présentées sous une forme graphique simple à interpréter, et constituent une aide efficace à la décision. Ces méthodes sont particulièrement adaptées au cas où les variables explicatives sont nombreuses. Le principe de construction d'un arbre binaire consiste à

déterminer une séquence de noeuds.

- Un noeud est défini par le choix conjoint d'une variable parmi les explicatives et d'une *division* qui induit une partition en deux classes. Implicitement, à chaque noeud correspond donc un sous-ensemble de l'échantillon auquel est appliquée une dichotomie ;
- une division est elle-même définie par une valeur seuil de la variable quantitative sélectionnée ou un partage en deux groupes des modalités si la variable est qualitative ;
- à la racine ou noeud initial correspond l'ensemble de l'échantillon ; la procédure est ensuite itérée sur chacun des sous-ensembles.

L'algorithme considéré nécessite :

- la définition d'un critère permettant de sélectionner la "meilleure" division parmi toutes celles *admissibles* pour les différentes variables ;
- une règle permettant de décider qu'un noeud est terminal : il devient ainsi une *feuille* ;
- l'affectation de chaque feuille à l'une des classes ou à une valeur de la variable à expliquer.

Arbre global

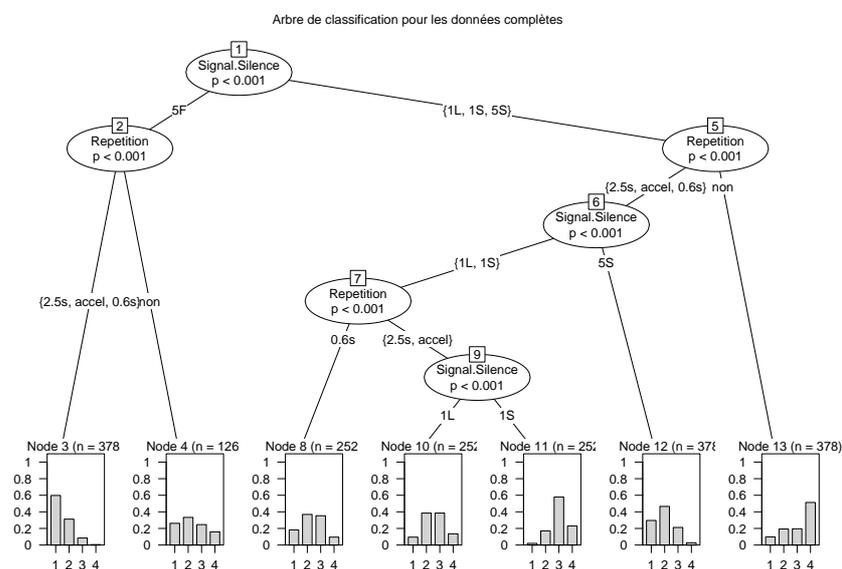


FIG. 5.3 – Arbre obtenu pour le modèle contenant toutes les variables.

L'arbre obtenu pour l'ensemble des données en tenant compte de toutes les variables

est présenté à la figure 5.3.

Comme expliqué précédemment, chaque noeud de l'arbre permet de découper les données en deux sous-groupes. Les variables indiquées dans chaque bulle de l'arbre sont celles qui découpent le mieux les données. Puis, chaque sous-groupe est défini par les valeurs de ces variables indiquées sur les branches de l'arbre. Par exemple, le premier découpage de l'arbre se fait suivant la variable `Signal.Silence`, et permet de séparer les données qui seront décrites par la valeur $5F$ de toutes les autres ($1L$, $1S$ et $5S$). Le processus de division se répète pour chaque sous-groupe, jusqu'à aboutir à un noeud terminal. Les feuilles terminales de l'arbre sont ici décrites chacune par un histogramme des données correspondantes à la feuille. Classiquement, les feuilles terminales des arbres de discrimination n'indiquent que la catégorie présente majoritairement dans les données de la feuille. Ces histogrammes permettent d'être plus précis dans l'interprétation des résultats, puisque nous savons non seulement à quelle catégorie est majoritairement associée une feuille, mais nous avons aussi accès à la répartition précise des données associées à la feuille dans chacune des autres catégories. Par exemple, la feuille 3 (tout à gauche de l'arbre) permet de voir que la feuille est associée majoritairement à la catégorie 1, et que la catégorie 4 n'est pas du tout représentée, alors que la feuille 4 (à droite de la feuille 3 précédente) montre une répartition quasi uniforme des données dans les quatre catégories. Dans ce dernier cas, il serait plus hasardeux d'associer la feuille à la catégorie 2, qui est la plus présente, mais de peu.

Plusieurs informations intéressantes peuvent être extraites de cet arbre.

Les variables `Signal.Silence` et `Répétition` sont les seules sélectionnées. Elles apparaissent donc comme étant les plus à même de discriminer les quatre catégories de réponse.

La première coupure de l'arbre met en évidence un comportement particulier pour les sons de type $SS=5F$. La seconde coupure de cette branche se fait au niveau de la variable `Répétition`, ce qui permet d'isoler de manière nette la catégorie 1 correspondant aux sons $5F$ répétés (les sons $5F$ non répétés sont quant à eux répartis uniformément). La catégorie 4 est elle aussi bien repérée comme étant associée aux sons de type $1L$, $1S$ et $5S$ non répétés. La catégorie 3 est elle plutôt associée aux sons $1S$ répétés $2.5s$ ou *accel*, bien qu'elle soit représentée dans plusieurs autres feuilles de l'arbre. Nous pouvons également constater qu'aucune feuille de l'arbre ne correspond nettement à la catégorie 2. Elle peut néanmoins être associée aux sons de type $5S$ répétés (quelle que soit la répétition).

Toutefois, comme mentionné précédemment, les arbres de discrimination sont des analyses descriptives, et ne permettent pas de modéliser précisément (donc de quantifier) les effets observés. De plus, bien que les variables `Signal.Silence` et `Répétition` soient confirmées

comme étant prédominantes, rien n'indique pour autant de manière claire que les quatre autres variables soient négligeables. En effet, des variables significatives mais secondaires peuvent ne pas apparaître dans des arbres de discrimination.

La modélisation précise des données de l'expérience a donc été effectuée grâce à un modèle de régression logistique.

5.5.4 Régression logistique

Les concepts de base et quelques-uns des aspects principaux de la régression logistique nécessaires à l'interprétation des résultats sont exposés dans l'Annexe A.

Remarque : La notation utilisée par la suite pour les modèles statistiques étudiés est la suivante :

$M_{SS.R.F0}$ désigne le modèle $\text{logit}[\pi(x)] = \beta_0 + \beta_1 SS + \beta_2 R + \beta_3 F0$.

$M_{SSR.F0}$ désigne le modèle $\text{logit}[\pi(x)] = \beta_0 + \beta_1 SS \times R + \beta_2 F0$.

C'est à dire que SS.R désigne un effet additif tandis que SSR (ou $SS \times R$) désigne l'interaction.

Notons aussi que l'interaction $SS \times R$ inclut les effets additifs de SS et de R.

Tests de significativité des variables explicatives

Les tables de contingences et les tests d'indépendance de type χ^2 permettaient déjà de comparer, en première approche, le poids des différentes variables susceptibles d'entrer dans le modèle (voir page 92). Cependant, il est également possible d'utiliser plusieurs régressions logistiques pour chaque variable susceptible d'intervenir dans le modèle. Il suffit ensuite de les comparer au modèle ne faisant intervenir que les constantes. Ainsi, en observant le critère des déviations et le critère AIC pour les modèles simples ne faisant intervenir que les variables explicatives prises séparément, une mesure fiable de l'influence des variables sur la classification est obtenue. Le tableau 5.10 présente le critère des déviations ainsi que le critère AIC, pour les modèles simples ne faisant intervenir que les variables explicatives prises séparément. Les valeurs correspondent exactement au test d'indépendance basé sur la vraisemblance présenté précédemment. Les deux variables les moins influentes sont donc la Modulation et la Polyphonie.

Une autre façon de repérer les variables qu'il est possible d'exclure du modèle final est de regarder la statistique de Wald associée aux coefficients estimés pour celles-ci. En d'autres termes, il s'agit d'estimer les coefficients d'une régression multinomiale pour un modèle

Modèle	Deviance	$\Lambda (M_{const}) - \Lambda (M_i)$	Degrés de liberté	P-value	AIC
M_{const}	5491.23				5497.236
M_{SS}	4843.79	647.44	9	0	4867.795
M_R	5184.34	306.89	9	0	5208.345
M_{F0}	5442.27	48.96	3	1.33e-10	5454.276
M_T	5458.24	32.99	3	3.22e-07	5470.237
M_P	5472.11	19.13	3	0.000257	5484.109
M_M	5479.27	11.97	3	0.007493	5491.268

TAB. 5.10 – Sélection des variables à l'aide des déviations.

additif complet faisant intervenir toutes les variables. Les éventuelles interactions ne sont donc pas prises en compte. Le modèle estimé est le suivant : **Réponse = Signal.Silence + Répétition + F0 + Timbre + Polyphonie + Modulation**, c'est-à-dire un modèle prenant en compte des effets *additifs* de toutes les variables. La catégorie prise comme référence est la catégorie 4. Les résultats sont présentés dans le tableau 5.11.

Le coefficient de Wald est obtenu en calculant le carré du rapport coefficient/erreur standardisée, et suit une loi de khi-deux à 1 degré de liberté lorsque le coefficient associé est nul. De grandes p-value correspondent donc aux coefficients que l'on peut considérer comme nuls et susceptibles d'être exclus du modèle final. L'examen des valeurs de la p-value conduit à penser que la variable Modulation a un effet négligeable, sa p-value n'étant significative que pour deux des trois catégories. De toutes les variables étudiées, la Modulation est la seule à suivre un tel comportement. Elle est confirmée comme étant la plus susceptible d'être exclue du modèle.

Cas particulier de la variable Modulation Il est alors intéressant d'analyser précisément l'effet de cette variable, en comparant la différence des Déviations entre les modèles $M_{F0.T.P.SSR}$ et $M_{F0.T.P.M.SSR}$.

Nous constatons que la variable Modulation n'apporte un gain de déviance que de 11.239 par rapport au modèle $M_{F0.T.P.SSR}$. Cependant, ceci reste significatif pour une loi de χ^2 à 3 degrés de liberté, avec une p-value de 0.01.

Toutefois, lors de l'estimation du modèle ne contenant pas la variable Modulation (i.e. le modèle **Réponse = Signal.Silence + Répétition + F0 + Timbre + Polyphonie**), l'impact sur l'estimation des coefficients du modèle **Réponse = Signal.Silence +**

	Variables	Coefficient	Erreur Standardisée	Statistique de Wald	p-value
3	Intercept	-1.232	0.227	29.557	0
	Sig.Sil=1L	0.411	0.189	4.732	0.03
	Sig.Sil=5S	2.626	0.307	73.205	0
	Sig.Sil=5F	1.876	0.305	37.777	0
	Rep=2.5s	2.580	0.229	127.231	0
	Rep=accel	2.177	0.221	96.636	0
	Rep=0.6s	3.015	0.284	112.635	0
	F0=A3	-0.451	0.172	6.893	0.009
	Timbre=String	-0.489	0.170	8.282	0.004
	Polyphonie=oui	0.138	0.166	0.690	0.406
	Modulation=oui	-0.065	0.165	0.156	0.693
2	Intercept	-2.209	0.253	76.339	0
	Sig.Sil=1L	1.120	0.211	28.170	0
	Sig.Sil=5S	4.266	0.319	178.495	0
	Sig.Sil=5F	3.823	0.311	151.064	0
	Rep=2.5s	2.891	0.252	131.784	0
	Rep=accel	2.816	0.241	136.219	0
	Rep=0.6s	3.937	0.303	168.316	0
	F0=A3	-1.247	0.182	46.990	0
	Timbre=String	-0.579	0.180	10.314	0.001
	Polyphonie=oui	0.225	0.175	1.655	0.198
	Modulation=oui	0.000	0.174	0.000	0.999
1	Intercept	-3.802	0.227	127.001	0
	Sig.Sil=1L	1.679	0.189	31.151	0
	Sig.Sil=5S	5.261	0.307	196.998	0
	Sig.Sil=5F	5.927	0.305	267.318	0
	Rep=2.5s	3.286	0.229	135.916	0
	Rep=accel	3.149	0.221	130.638	0
	Rep=0.6s	4.599	0.284	197.077	0
	F0=A3	-1.435	0.172	51.382	0
	Timbre=String	-1.296	0.170	42.317	0
	Polyphonie=oui	0.762	0.166	15.518	0
	Modulation=oui	-0.416	0.165	4.656	0.031

TAB. 5.11 – Coefficients de la régression logistique obtenus avec le modèle additif complet, détaillant les coefficients de Wald associés. Sig.Sil correspond au paramètre Signal.Silence ; Rep correspond à Répétition. Intercept représente la constante du modèle

Répétition + F0 + Timbre + Polyphonie + Modulation est très faible. La seule modification des coefficients concerne les probabilités de classement dans la catégorie 1, et ce de manière assez faible. Les coefficients estimés pour la variable Modulation sont de plus très faibles, ce qui implique une absence d'effet.

En vertu du principe de parcimonie et de l'analyse qui précède, la variable Modulation a donc été écartée du modèle final.¹

Recherche d'interactions

Afin de détecter les interactions susceptibles d'avoir de l'influence, il est d'usage de comparer la déviance du modèle avec interaction avec la déviance du modèle additif associé. Dans la mesure où il s'agit de modèles emboîtés, la différence des déviances suit là encore une loi du χ^2 . De manière plus générale, la différence des déviances permet de tester l'apport explicatif d'une suite de modèles emboîtés. Un modèle ne modifiant que très peu la déviance n'apportera que peu d'information mais aura pour conséquence d'ajouter des variables, ce qui peut nuire à la qualité d'estimation des paramètres. Il est important de choisir le modèle le plus parcimonieux possible, faisant un compromis entre la part d'information expliquée et la complexité.

Plusieurs interactions ont été testées, de deux manières différentes et complémentaires. Une première approche consiste à comparer le gain de la déviance (ou, ce qui revient au même à un facteur 2 près, en comparant les vraisemblances) entre le modèle avec interaction et le modèle additif associé. Les résultats apparaissent dans le Tableau 5.12.

Une autre manière d'évaluer l'importance d'une interaction est de comparer la déviance du modèle considérant uniquement les effets moyens (c'est-à-dire en ne supposant que des effets additifs et linéaires), avec le modèle considérant les effets moyens auxquels s'ajoute l'interaction souhaitée. De la même façon que précédemment, la différence des déviances suit aussi une loi du χ^2 . Les résultats apparaissent dans le Tableau 5.13.

Les seules interactions suffisamment significatives et qui apparaissent comme des composantes essentielles du modèle sous-jacentes aux données sont SS*R et F0*T.

¹Les analyses effectuées nous ont malgré tout permis d'évaluer son influence (légère). En effet, nous avons vu précédemment qu'un son modulé a plutôt tendance à être classé dans les catégories 2 ou 3. L'analyse de régression logistique nous indique de plus que le seul coefficient non nul de la variable modulation est celui associé à $\log\left(\frac{\pi_1(x)}{\pi_4(x)}\right)$ et qui vaut -0.46. L'influence très légère de la Modulation tend donc à associer un son modulé à la catégorie 4 plutôt qu'à la catégorie 1. Même si, toutes choses égales par ailleurs, un son modulé a peu de chance d'être classé dans la catégorie 1, ce sont surtout les autres variables qui font la différence.

Variabes	déviante Modèle additif	déviante Modèle avec interaction	Différence	Degrés de liberté	P-value
SS et R	4443.924	4346.039	97.88502	27	5.711591e-10
F0 et T	5409.084	5336.821	72.26281	3	1.44329e-15
SS et T	4803.636	4793.605	10.03055	9	0.3480173
SS et F0	4778.255	4768.609	9.6466	9	0.3798437
R et T	5149.853	5131.101	18.7521	9	0.02738528
R et F0	4830.988	4819.403	11.58537	9	0.2377058
SS et M	5131.599	5115.417	16.18245	9	0.0632
SS et P	4820.914	4808.153	12.76099	9	0.1737226

TAB. 5.12 – Recherche d'interactions à l'aide du critère de déviance.

Modèle	Deviance	Différence	Degrés de liberté	P-value	AIC
$M_{SS.R.F0.T.P.M.}$	4282.20				4348.2
$M_{SS.R.F0.T.P.M.SSR}$	4190.45	91.74	27	5.595e-09	4310.456
$M_{SS.R.F0.T.P.M.F0T}$	4261.87	20.32	3	0.000145	4333.876
$M_{SS.R.F0.T.P.M.SST}$	4272.12	10.07	9	0.344	4356.127
$M_{SS.R.F0.T.P.M.SSF0}$	4275.32	6.88	9	0.649	4359.319
$M_{SS.R.F0.T.P.M.RT}$	4272.64	9.55	9	0.388	4356.646
$M_{SS.R.F0.T.P.M.RF0}$	4271.64	10.55	9	0.307	4355.647
$M_{SS.R.F0.T.P.M.SSM}$	4272.63	9.56	9	0.387	4356.636
$M_{SS.R.F0.T.P.M.SSP}$	4275.43	6.76	9	0.662	4359.437

TAB. 5.13 – Recherche d'interactions à l'aide du critère de déviance.

Cependant, il s'avère que le plan d'expérience utilisé confond les deux interactions précédentes. Ces deux interactions étant confondues, nous retiendrons uniquement l'interaction SS*R dans le modèle final. Ce choix se justifie essentiellement par les résultats existants de la littérature, qui accordent un poids prépondérant aux paramètres de type rythmique, comparativement à la F0 ou à des paramètres relatifs au timbre des sons (Edworthy et coll., 1991 ; Hellier et coll., 1993).

Choix du modèle final

D'après tout ce qui précède, le modèle choisi pour expliquer nos données est le modèle **Signal.Silence*Répétition + F0 + Polyphonie + Timbre**, i.e. un modèle prenant en compte l'interaction entre Signal.Silence et Répétition à laquelle s'ajoute l'effet additif de F0, Polyphonie et Timbre.

Validation du modèle

Cette procédure de validation a été implémentée afin de tester la robustesse de notre modèle. Comme expliqué précédemment, il faut s'attendre à un fort taux d'erreur dans les tests de validation. En effet, en comparant les classes majoritaires des sons avec les réponses des sujets contenues dans les données, l'accord maximal n'est que de 54%, qui est donc le meilleur taux de bon classement possible.

Dans un premier temps, les analyses d'influence des variables et des interactions ont été à nouveau effectuées sur des échantillons déterminés au hasard de 70% des données complètes. Ces analyses ont montré des résultats similaires à ceux obtenus sur l'ensemble des données, et il a donc été décidé de garder le modèle choisi.

Pour la procédure de validation proprement dite, 70% des données ont été tirées au hasard afin d'estimer le modèle choisi, puis la classe prédite par le modèle estimé a été comparée aux réponses fournies par les sujets des 30% de données restantes. Cette approche a été effectuée 100 fois.

Le meilleur modèle, c'est-à-dire celui ayant le plus fort taux de bon classement, est bien évidemment celui comportant le plus grand nombre de variables : le modèle SSR.F0.T.P.M. Néanmoins, le modèle choisi par notre procédure statistique est celui ne tenant pas compte de la variable Modulation. Outre le fait que son critère AIC est très bon comparé aux autres choix, son taux de bon classement est lui aussi un des meilleurs. Remarquons donc ici, une fois de plus, l'intérêt de choisir un modèle simple, tout en étant suffisamment explicatif.

Dans un troisième temps, les proportions estimées par le modèle construit sur les données complètes ont été comparées aux proportions contenues dans les données. Nous avons donc comparé les histogrammes de chaque son présentés précédemment aux proportions prédites par le modèle : la qualité d'explication du modèle choisi est très bonne. L'écart entre les proportions observées et celles prédites est en effet très faible, et paraît négligeable. Un exemple pour les sons 1 à 8 est donné en Figure 5.4.

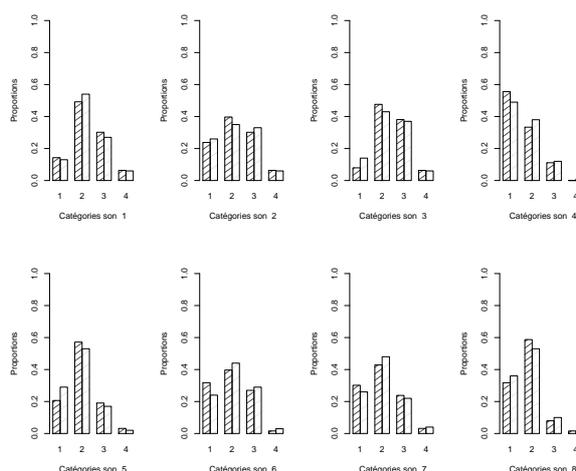


FIG. 5.4 – Histogramme de la classification des sons n°1 à 8 (barres hachurées), comparés aux probabilités prédites par le modèle (barres non hachurées).

Interprétation des résultats du modèle *SSR.F0.T.P*

Dans un premier temps, nous avons calculé les probabilités de classement pour les différentes combinaisons des variables Signal.Silence et Repetition, en considérant que les quatre autres variables (F0, T, P, M) sont dans leur état de "référence" (F0=A1, T=Plate, Polyphonie=No et Modulation=No). Les résultats sont présentés dans la Figure 5.5.

De la même manière, les probabilités associées aux autres modalités de F0 et du Timbre ont été calculées et sont présentées dans les Figures 5.6, 5.7 et 5.8.

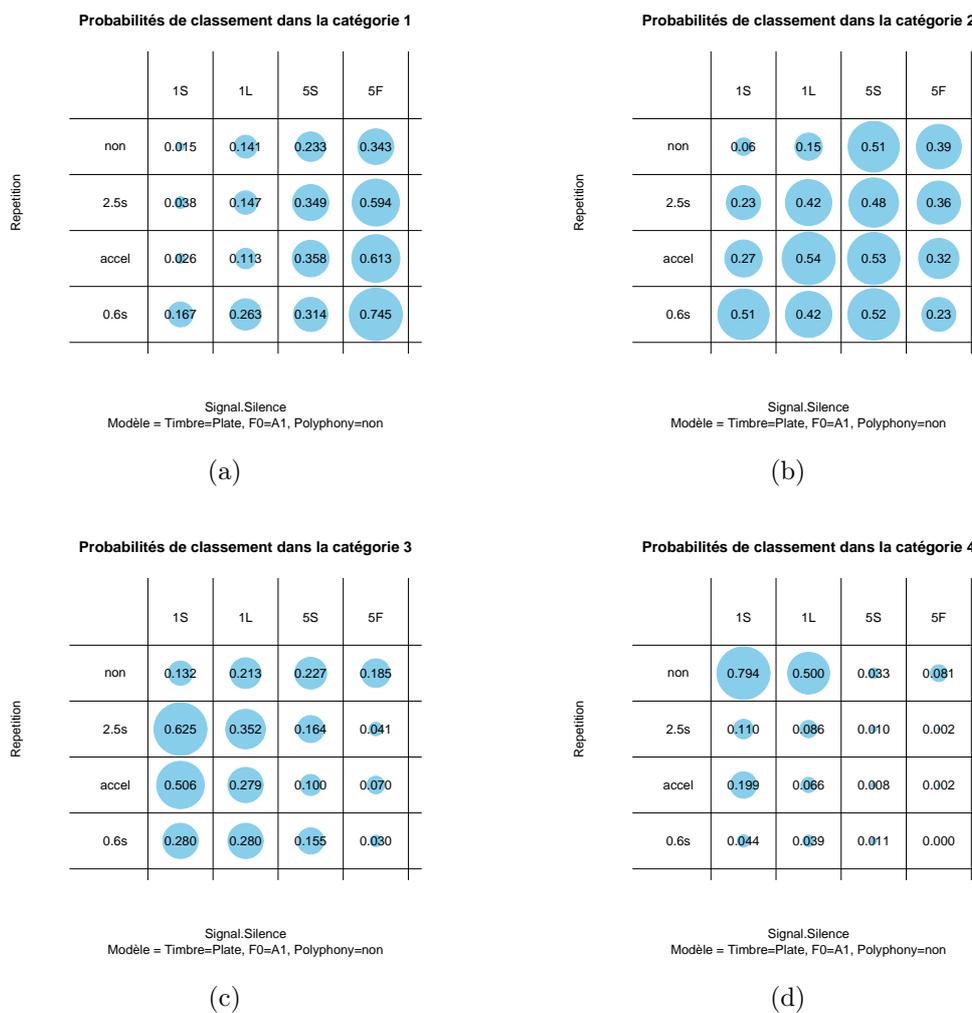


FIG. 5.5 – Probabilités de classement dans chaque catégorie pour le modèle SSR.F0.T.P avec Timbre=Plate, F0=A1, et Polyphony=No

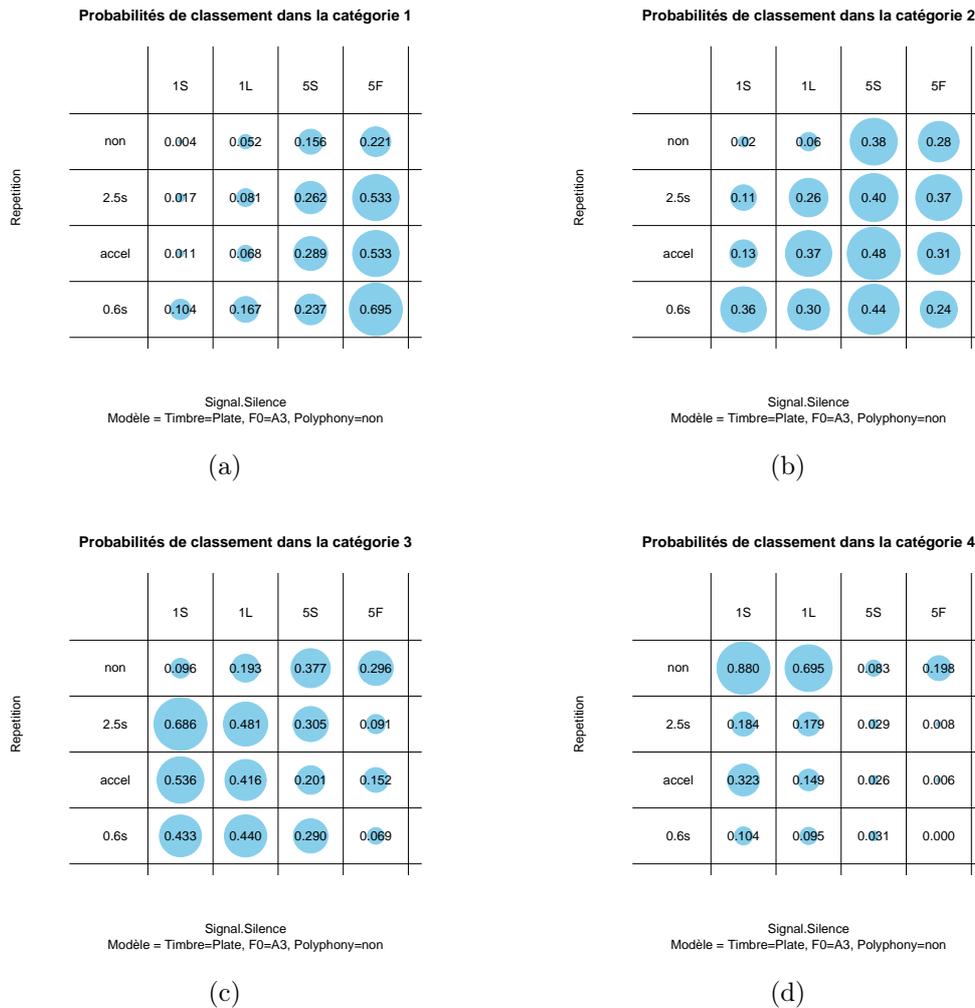


FIG. 5.6 – Probabilités de classement dans chaque catégorie pour le modèle SSR.F0.T.P avec Timbre=*Plate*, F0=*A3*, et Polyphony=*No*

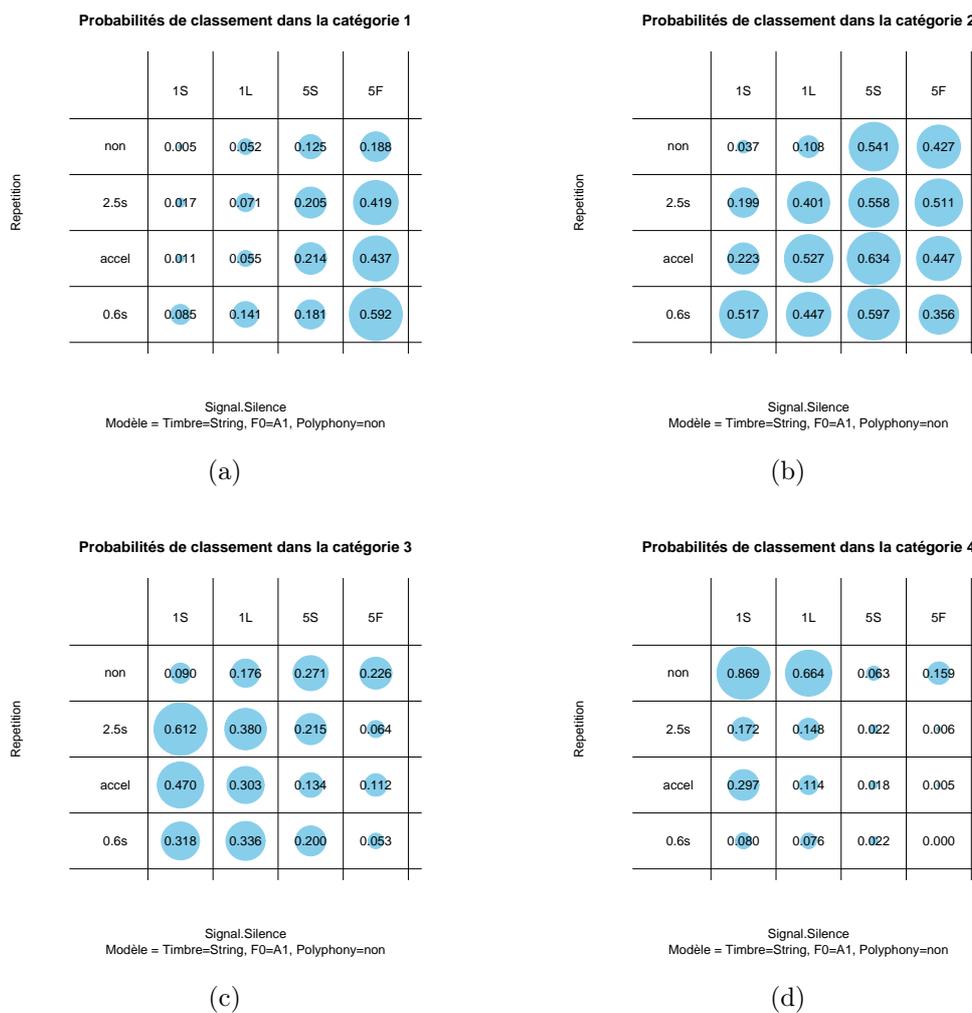


FIG. 5.7 – Probabilités de classement dans chaque catégorie pour le modèle SSR.F0.T.P avec Timbre=*String*, F0=*A1*, et Polyphony=*No*

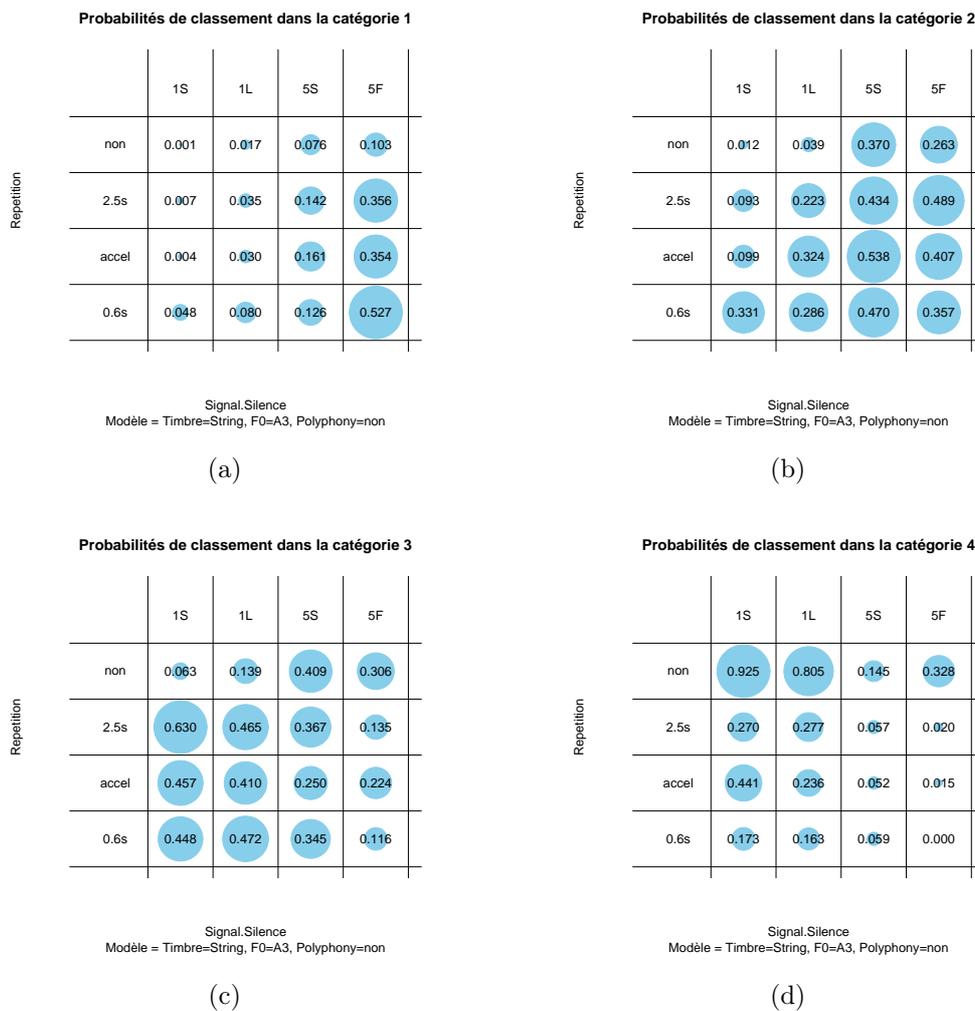


FIG. 5.8 – Probabilités de classement dans chaque catégorie pour le modèle SSR.F0.T.P avec Timbre=String, F0=A3, et Polyphony=No

De ces tables, nous pouvons déduire quelques conclusions intéressantes :

- De manière globale, les sons les plus classés dans la catégorie 1 sont ceux associés à la modalité *5F* de Signal.Silence, avec une préférence pour les sons répétés (*2.5s*, *accel* ou *0.6s*).
- De même, les sons classés dans la catégorie 4 sont clairement les sons *1S* et *1L* non répétés. Ce sont quasiment les seuls, avec, dans une bien moindre mesure, les sons de type *1S* avec $R=accel$.
- Les sons de la catégorie 2 sont beaucoup plus disparates. Cette catégorie est mal comprise et draine l'essentiel des sons de type *5S*. S'y trouvent aussi les sons $SS=1S / R=0.6s$, $SS=1L / R=accel$ ou *0.6s*, ainsi que les sons *5F* non répétés.
- La catégorie 3 est composée quant à elle des sons $SS=1S$ ou *1L* (lorsque $F0=A\beta$), répétés (quelle que soit la répétition).
- La modalité $F0=A\beta$ augmente la probabilité d'être classé dans la catégorie 4 (coefficients de la régression négatifs).
- De même, la modalité *Timbre=String* augmente la probabilité de classer un son dans la catégorie 4.
- L'influence des paramètres *Timbre* et $F0$ est faible relativement à la répartition des sons en fonction de SS et de R .

Analyse complémentaire : durée totale du son

Remarquons ici, avant toute conclusion finale, que, bien que la durée totale du son ne soit pas une variable explicite présente dans les données, elle est une variable sous-jacente puisqu'elle dépend des variables "Signal.Silence" et "Répétition". Il ne serait pas aberrant de considérer que la durée du son ait une influence sur le classement des sons.

Les durées totales des 32 sons (calculées à partir des variables Signal.Silence et Répétition) sont exposées dans les tableaux 5.14 et 5.15. Afin de savoir si une relation entre le classement des sons et la durée de ceux-ci existe, nous avons observé les probabilités de classement obtenues par le modèle logistique en fonction de la durée totale des sons. Les graphiques obtenus sont présentés en Figure 5.9. Ils invalident l'existence d'une influence forte de la longueur du son en présence des variables Signal.Silence et Répétition. Même si une tendance peut apparaître comme dans le cas de la catégorie 4, celle-ci est artificielle. En effet, nous avons vu qu'un son à 1 impulsion non répétée était associé majoritairement à la catégorie 4 (de non danger). Ces sons étant les plus courts, nous observons donc un lien entre son court et la catégorie 4. Toutefois, même si une liaison existe, elle est faible,

et il semble plus pertinent de considérer séparément, comme nous l'avons fait, les effets de Signal.Silence et Répétition.

Durée (en s)	0.2	1.2	2.0	2.6	2.6	5.1	6.6	8.3	9.1	9.8
Signal.Silence	1S	5F	1L	1S	5S	1S	5F	1S	5F	1L
Repetition	non	non	non	0.6s	non	accel	0.6s	2.5s	accel	0.6s

TAB. 5.14 – Durée du son exprimée en fonction des paramètres Signal.Silence et Répétition

Résumé des résultats issus de la régression logistique

Les résultats de la régression logistique correspondent aux intuitions fournies par l'analyse descriptive et par les résultats des arbres de discrimination ; elle permet cependant d'aller bien plus en finesse dans les résultats obtenus.

La dominance des paramètres Signal.Silence et Répétition est ainsi confirmée. Les sons *5S* et *5F* sont respectivement fortement associés aux catégories 1 et 2.

L'ensemble des sons *1S* et *1L* est majoritairement classé dans les catégories 3 ou 4 (bien que les sons *1S* et *1L*, avec $R=0.6s$ soient à part).

De plus, l'influence des paramètres secondaires Timbre et F0 est détaillée : les modalités Timbre=*String* et F0=*A3* augmentent les probabilités de classement dans les catégories 3 et 4.

5.5.5 Synthèse des résultats

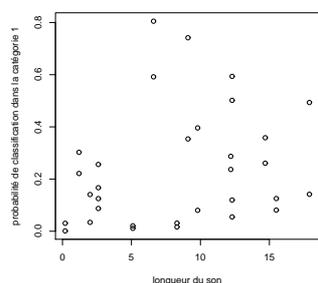
Nous pouvons alors dresser une "carte acoustique" de chaque catégorie.

Catégorie 1 :

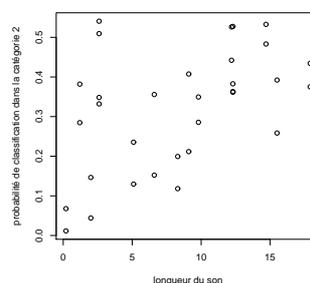
- Sons de type *5F* répétés (quelle que soit la répétition)
- Timbre=*Plate*, F0=*A1*

Durée (en s)	12.2	12.3	12.3	14.7	15.5	17.9
Signal.Silence	5S	5F	1L	5S	1L	5S
Repetition	0.6s	2.5s	accel	accel	2.5s	2.5s

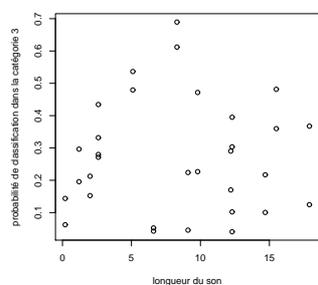
TAB. 5.15 – Durée du son exprimée en fonction des paramètres Signal.Silence et Répétition (suite)



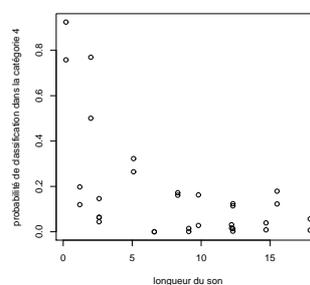
(a) Durée du son comparé à la probabilité d'être classé dans la catégorie 1.



(b) Durée du son comparé à la probabilité d'être classé dans la catégorie 2.



(c) Durée du son comparé à la probabilité d'être classé dans la catégorie 3.



(d) Durée du son comparé à la probabilité d'être classé dans la catégorie 4.

FIG. 5.9 – Probabilités de classement dans chaque catégorie en fonction de la durée des sons.

Catégorie 2 :

- Sons de type $5S$
- S'y ajoutent les sons $[SS=1L, R=2.5s, accel, 0.6s]$, $[SS=5F, R=no, 2.5s, accel]$ et $[SS=1S, R=0.6s]$

Catégorie 3 :

- Sons de type $1S$ ou $1L$ avec $R= 2.5s$ ou $accel$
- Timbre= $String$, $F0=A3$

Catégorie 4 :

- Sons $1S$ et $1L$ non répétés
- Timbre= $String$, $F0=A3$

En reprenant les hypothèses expérimentales décrites précédemment (section 5.3), nous pouvons également présenter les résultats paramètre par paramètre :

– **Fréquence Fondamentale (F0)**

La F0 a été mise en avant comme le deuxième paramètre important dans le classement des sons, après les paramètres rythmiques Signal.Silence et Répétition. Les hypothèses d'Hellier et coll. (1993) sont alors vérifiées. Cependant, l'effet de la F0 est inverse de celui montré habituellement dans la littérature (Edworthy et coll., 1991 ; Hellier et coll., 1993) : les F0 faibles sont associées aux catégories d'urgence les plus élevées. Ce point sera traité au cours de la Discussion.

– **Timbre**

Le Timbre apparaît ici comme un paramètre secondaire (après la F0 ; voir les tests d'indépendance en page 92). Là aussi, c'est l'hypothèse d'Hellier et coll. (1993) qui est vérifiée.

– **Modulation**

Comparativement aux autres paramètres de l'expérience, la Modulation n'a pas d'influence réellement significative sur le classement des sons. Il est probable que ce résultat ne soit pas généralisable pour d'autres corpus de sons (la comparaison d'un son pur et d'un son pur modulé à 60 Hz donnerait vraisemblablement une différence de jugement d'urgence).

– **Polyphonie**

Bien que ce paramètre ait été conservé dans le modèle logistique final, il n'a qu'une influence très faible. L'utilisation de sons polyphoniques ne facilite pas réellement le classement des sons, et ne permet pas de mieux discriminer certaines catégories. Là encore, ce résultat n'est valable que pour ce corpus de sons, bien évidemment. L'observation approfondie du modèle logistique permet tout de même de mettre à jour un faible effet des sons polyphoniques sur la catégorie d'urgence extrême (catégorie 1) comparativement aux trois autres.

– **Signal.Silence**

Les hypothèses de Brewster et coll. (1994) ainsi que celles d'Edworthy et coll. (1991) et Hellier et coll. (1993) sont vérifiées. Le paramètre rythmique est bien dominant. Le nombre d'impulsions permet de faire une première discrimination entre des sons urgents et des sons non urgents ; l'intervalle inter-impulsion permet, quant à lui, de séparer deux catégories d'urgence plus fines (les catégories 1 et 2). Telle que nous l'avons introduite dans le plan expérimental, la durée d'impulsion n'a pas vraiment de poids sur la catégorisation des sons.

– Répétition

En tant que paramètre rythmique, la Répétition apparaît également comme paramètre dominant, les hypothèses expérimentales sont bien vérifiées. Toutefois, les différentes répétitions de train d’impulsions utilisées (rapide, lente ou accélérée) n’ont pas été déterminantes dans le classement des sons ; ce qui importe est le fait de répéter le train d’impulsions, quelle que soit la répétition.

5.6 Discussion

5.6.1 Influence du rythme

La discrimination des sons dans les quatre catégories d’IHM automobiles a été déterminée principalement par les paramètres rythmiques : les variables Signal.Silence et Répétition sont les plus influentes sur le classement et permettent aux sujets de distinguer clairement les classes de sons.

Modélisation des catégories automobiles selon la dimension rythmique

Outre une bonne discrimination des catégories selon la dimension rythmique, la modélisation logistique a permis de préciser les premiers résultats, et d’apporter ainsi des compléments aux études se focalisant sur l’urgence perçue (Edworthy et coll., 1991 ; Hellier et coll., 1993). En effet, il est envisageable, au vu des résultats, que les sujets aient utilisé avant tout l’échelle d’urgence sous-jacente aux quatre catégories, classant ainsi les stimuli de manière claire et consensuelle dans les deux catégories extrêmes, et en regroupant un peu plus confusément dans les deux catégories centrales les sons dont la dimension fonctionnelle leur apparaissait moins ostensiblement. Le distinguo entre catégories automobiles et urgence perçue sera discuté plus loin (section 5.6.4 page 122).

Remarquons ici que la distinction entre Signal.Silence et Répétition n’est pas toujours limpide. En effet, prenons l’exemple d’un son de type *5S* non répété ; il est très proche perceptivement d’un son de type *1S* répété toutes les *0.6s*. Ce rapprochement - entre autres - a pu contribuer à une certaine assimilation des deux paramètres.

Une vitesse rapide, avec plusieurs répétitions du motif est la condition *sine qua non* de l’urgence maximale. Une vitesse lente ne peut être associée à un son d’urgence. Nous confirmons ainsi les résultats d’Hellier et coll. (1993), à l’aide d’une validation expérimentale différente : ce résultat est bien robuste.

Le nombre d'impulsions semble, quant à lui, permettre de discriminer des catégories d'urgence bien distinctes : les sons à 1 impulsion non répétée sont très clairement associés à une catégorie de non urgence, alors que des sons à 5 impulsions répétées définissent la catégorie la plus urgente. Nous avons démontré que ce n'est pas la longueur du son implicitement codée dans nos données qui est à l'origine de ces observations (section 5.5.4 page 114). Enfin, la durée de l'impulsion ne semble pas avoir d'influence sur le classement des sons, du moins pour les sons à 1 impulsion.

Effet du contexte culturel

Comme pressenti également par Blattner et coll. (1989) et Brewster et coll. (1994), le rythme a donc bien une importance majeure sur la distinction des différentes IHM. Les limites de la validation expérimentale proposée par ces auteurs ont été soulevées précédemment. Toutefois, la spécificité de la communauté ICAD est de s'inscrire davantage dans un cadre de design - auquel peuvent se juxtaposer certaines validations ergonomiques - que dans le domaine de la psychophysique. Une démarche de design s'amorce généralement par un inventaire et une analyse de la situation existante. Bien que cette étape ne soit pas explicitement commentée par les auteurs cités ci-dessus, nous pouvons interpréter une partie de ces travaux comme une première étape de design sonore. De ce fait, il apparaît intuitivement que les préconisations de Brewster et coll. (1994) se fondent sur les sons électroniques actuels (en 1994, du moins).

Ainsi, la prévalence des paramètres rythmiques s'explique par la simple écoute des sons électroniques environnants (ce constat est du reste encore valable), et le classement des sons suivant cette dimension rythmique peut alors être commenté sous cet angle. Lors de notre expérience, les sujets établissaient un classement essentiellement suivant les paramètres Signal.Silence et Répétition, en comparant les stimuli de l'expérience aux sons connus dans leur environnement quotidien et, plus particulièrement, à ceux qu'ils ont l'habitude d'entendre dans leur automobile. C'est l'apprentissage que chacun avait des IHM en général, et de l'automobile en particulier, qui a pu guider le classement des sons.

A l'inverse, indépendamment de tout apprentissage spécifique de sons automobiles et électroniques, il est également possible que cette dimension temporelle soit une dimension prédominante dans notre perception des informations sonores, en général. Cependant, cette expérience ne peut permettre de conclure sur ce point, ni même d'appuyer ce postulat d'une quelconque façon.

5.6.2 Influence de la F0 et du Timbre

Le Timbre apparaît comme un paramètre secondaire dans la catégorisation des sons. Bien que ce résultat confirme d'une certaine manière l'hypothèse d'Hellier et coll. (1993) - en considérant la différence d'inharmonicité des timbres employés - il ne nous permet pas d'étendre les préconisations de Brewster et coll. (1994) à une catégorisation suivant une échelle d'urgence.

En effet, malgré des timbres parfaitement différenciables, ce paramètre ne ressort pas de l'analyse par arbre de discrimination, et n'est défini que comme un paramètre très secondaire dans le modèle logistique. Il est possible que les modalités du paramètre, voire le paramètre de Timbre en lui-même tel que défini dans cette expérience, aient été "écrasées" par d'autres paramètres bien plus influents sur un jugement d'urgence (comme le rythme). Tel que proposé par la communauté ICAD, et plus particulièrement par Gaver (1993), le timbre est une caractéristique fine qui peut permettre de distinguer des nuances au sein même d'une catégorie fonctionnelle. Dans notre expérience, l'urgence perçue est mêlée à la notion de catégorie fonctionnelle ; de ce fait, les paramètres rythmiques ont pu prédominer sur le timbre.

Une autre explication possible pourrait être, une fois de plus, issue de l'observation des sons de l'environnement automobile et des sons électroniques au sens large. Du fait de l'utilisation de systèmes de restitution tels des buzzers, la différenciation des sons fonctionnels ne se fait que très rarement selon une dimension telle que le timbre. Les sons connus des auditeurs sont des "bips", qui varient suivant le rythme, et suivant la fréquence fondamentale.

Cependant, la F0 n'apparaît pas non plus ici comme un paramètre dominant, contrairement aux hypothèses expérimentales issues de la littérature ICAD, et à l'étude d'Hellier et coll. (1993) sur l'urgence perçue. De plus, les premières analyses, tout comme la modélisation logistique montre de manière claire un effet inverse à celui attendu : l'urgence perçue augmente lorsque la F0 diminue.

Outre le jugement demandé aux sujets, qui diffère entre notre étude et celles de la littérature d'urgence, les stimuli utilisés sont également très différents. Il est probable que cet effet inverse trouve son explication dans le type de stimuli utilisés. Considérons des sons purs, ou des sons complexes composés d'une dizaine d'harmoniques, créés par synthèse additive. Ce sont les stimuli employés par Edworthy et coll. (1991) et Hellier et coll. (1993) ; ils ressemblent considérablement aux sons existants dans notre environnement sonore électronique. Selon l'interprétation du contexte culturel déjà proposée précédemment, il est

alors envisageable qu'une variation de F0 entraîne un jugement du type "un son aigu est un son urgent".

Les sons utilisés dans notre expérience sont très différents de notre référentiel sonore quotidien et classique ; ils ne ressemblent à rien de connu ou déjà entendu par ailleurs. Une variation de F0 appliquée à de tels sons n'engendre alors pas l'effet prédit par la littérature. Bien que le plan d'expérience utilisé ne permette pas de distinguer l'effet de l'interaction entre Signal.Silence et Répétition de celle entre Timbre et F0, nous pouvons supposer que cette dernière interaction ait en fait une influence sur un jugement d'urgence. Notre expérience ne nous permet pas de conclure sur ce point.

De plus, l'interprétation de l'effet de l'interaction F0*Timbre sur le jugement d'urgence peut être faussée par un effet potentiel de cette interaction à un niveau uniquement perceptif (pour les détails de l'influence de la F0 sur le Timbre, voir Marozeau et coll., 2003). Nous reviendrons sur ce point de la confusion possible entre influence perceptive et/ou cognitive plus en détails en section 5.6.4 page 123.

5.6.3 Influence de la Modulation et de la Polyphonie

Enfin, il apparaît clairement dans nos données que la Modulation n'a que peu d'influence. De plus, même en analysant son effet sur la catégorisation précisément et indépendamment des autres variables, il en ressort qu'un son modulé est davantage associé à la catégorie 4 (donc de non urgence), ce qui semble contradictoire avec les préconisations de Patterson (1986).

Toutefois, comme nous l'avons soulevé précédemment (section 5.3 page 83), la modulation a une influence sur la perception du timbre, et il est possible que ce biais implicite obscurcisse un effet potentiel de la modulation sur le jugement d'urgence.

De la même manière, la Polyphonie a une influence très faible dans le modèle final, et il semble difficile d'étendre les recommandations d'ICAD concernant la polyphonie. Les catégories 2, 3 et 4 ne se différencient pas selon cette dimension polyphonique ; seule la catégorie 1 opposée aux trois autres permet de distinguer les sons polyphoniques des autres. L'association d'un son polyphonique à une catégorie d'urgence s'explique peut-être plus par une métrique psychoacoustique que par ce paramètre de Polyphonie en tant que tel. Toutefois, les données de notre expérience ne permettent pas d'en dire beaucoup plus.

5.6.4 Sons *intentionnels* : limites et méthode

Les questions mises à jour par l'analyse des paramètres acoustiques sont en fait un peu plus générales.

Catégorisation et catégories automobiles

Les expériences d'urgence perçue (Edworthy et coll., 1991 ; Hellier et coll., 1993) se fondent sur des méthodes subjectives de comparaison multiple et d'estimation de grandeur. Les fonctions des IHM automobiles étant plus diverses qu'un simple niveau d'urgence, quatre catégories fonctionnelles ont été choisies ; de manière évidente, nous avons donc décidé de nous placer dans le cadre d'une expérience de catégorisation orientée afin de valider nos hypothèses expérimentales.

Outre certaines conclusions utiles de cette expérience concernant l'influence des différents paramètres sur le jugement demandé aux sujets, il apparaît également que le classement des sons suivant ces quatre catégories d'IHM est loin d'être consensuel. Dès les premières analyses descriptives, la catégorie 2 (alerter d'un danger lié à une défaillance) émerge comme une sorte de catégorie "poubelle" : c'est la réponse donnée le plus couramment par les sujets (voir section 5.5.1), tout en étant la moins bien décrite par les paramètres acoustiques introduits. Bien que nous ayons pu extraire des données une définition de cette catégorie à l'aide de la régression logistique, il est tout de même intéressant de noter qu'aucune conclusion n'avait pu être tirée à partir de la simple observation des arbres de discrimination. A l'opposé, les catégories 1 et 4 (les catégories extrêmes), elles, sont bien représentées et définies clairement d'un point de vue acoustique.

Comme souligné précédemment, il semble que les sujets aient utilisé uniquement l'échelle d'urgence sous-jacente aux quatre catégories. Bien qu'une certaine confusion puisse être due aux stimuli ou au trop grand nombre de paramètres utilisés, il ressort tout de même de cette expérience que les catégories d'IHM utilisées dans l'automobile ne peuvent être testées ainsi directement. Les sujets ne semblent pas utiliser les dimensions sémantiques introduites dans ces catégories, mais plutôt un simple niveau d'urgence. Pire, l'introduction de catégories libellées clairement par rapport à leur fonction automobile a obscurci les réponses des sujets : c'est le cas de la catégorie 2.

Une nuance pourrait être apportée à ce constat, en observant que les sons d'oubli ont été définis majoritairement comme des sons répétés, quelle que soit la répétition. Ce profil semble cohérent avec les sons d'oubli classiques de notre environnement quotidien (oubli de

carte bancaire par exemple) qui sont des "bips" répétés. Cependant, ce profil est également cohérent avec une échelle d'urgence décroissante, où une catégorie d'urgence moyenne est définie comme un son répété lentement (i.e. une vitesse faible).

Nous pouvons également noter que l'expérience présentée dans ce chapitre ne peut apporter plus d'informations que ce pour quoi elle existe : les stimuli de l'expérience sont comparés aux sons existants dans l'univers sonore des sujets, et sont ainsi jugés. Comment proposer alors des IHM sonores nouvelles et potentiellement plus efficaces dans la transmission du message ? Cette question sera traitée dans la suite de nos travaux expérimentaux (Chapitres 6 et 7).

Problème inverse

La problématique posée par les sons *intentionnels*, i.e. porteur d'une intention ou d'un message, peut en effet être vue comme la réciproque d'études plus "traditionnelles". Les sons qui nous entourent sont par essence porteurs d'information, et la question est généralement de savoir quel type d'informations est véhiculé par ces sons, et sur quoi ces sons nous renseignent (sur la source du son, par exemple). Dans le cas de sons intentionnels, l'avènement de sons électroniques porteurs de messages impose alors une problématique inverse : une fonction (i.e. un message) étant définie, quelle information sonore est le mieux à même de transmettre cette fonction ? Une partie de la littérature citée ici a favorisé la fonction d'urgence, par exemple.

Cependant, comme souligné en section 3.3.3 page 54, lorsque l'effet d'un paramètre acoustique sur un jugement d'urgence est étudié, il n'est pas possible, en analysant les résultats, de faire la différence entre l'influence de la *perception* de chacun des paramètres (ainsi que celle de possibles interactions entre eux) et l'influence de ces paramètres sur le *jugement d'urgence* en lui-même. En d'autres termes, deux niveaux de traitements auditif et/ou cognitif sont à distinguer : l'extraction d'attributs auditifs et une première analyse perceptive du son, suivies d'un traitement d'ordre cognitif qui nous permet de juger de l'urgence potentielle d'un son. En demandant aux sujets uniquement un jugement d'urgence (que ce soit dans notre expérience ou dans la littérature relative à l'urgence perçue), il n'est pas possible de discriminer dans l'analyse des réponses ce qui pourrait potentiellement être attribuable à des facteurs perceptifs (interactions perceptives entre paramètres ou un écrasement perceptif d'une dimension par rapport à une autre, par exemple). En d'autres termes, le rythme est-il dominant parce qu'il pourrait émerger de l'espace perceptif de nos stimuli (suite à une expérience de dissemblance, par exemple) ou parce qu'il est associé à

la notion d'urgence, "intrinsèquement" ?

5.7 Conclusion

L'influence de six paramètres acoustiques sur une catégorisation fonctionnelle automobile a été étudiée expérimentalement. Nous avons démontré que les paramètres rythmiques (Signal.Silence et Répétition) ont une influence forte sur la discrimination et la modélisation de chacune des catégories d'IHM. Des trains d'impulsions composés de plusieurs impulsions rapides et répétées sont associés à un niveau d'urgence élevé, à l'opposé d'une impulsion unique non répétée qui est quant à elle associée à un son non urgent. L'échelle d'urgence se fait ensuite selon le nombre d'impulsions d'abord, et la vitesse du train d'impulsions ensuite. Ce résultat confirme ceux obtenus par Edworthy et coll. (1991), Hellier et coll. (1993) et Brewster et coll. (1994). Il a également été mis en évidence que le Timbre et la F0 avaient une influence secondaire sur la catégorisation fonctionnelle des sons, avec un effet de la F0 inverse à celui de la littérature : une F0 élevée est associée dans le cas présent à une urgence faible, et non l'inverse. L'interprétation de ces résultats met en évidence des effets de contexte culturel, et montre les limites d'une telle expérience.

De plus, la catégorisation demandée aux sujets n'était pas toujours claire et consensuelle. Toutefois, bien que la dimension sémantique introduite dans les catégories d'IHM ne semble pas avoir été comprise par les sujets, il s'avère nécessaire de questionner à nouveau la notion d'urgence perçue. Le jugement subjectif demandé aux sujets ne concerne pas un percept élémentaire (comme la hauteur ou la rugosité par exemple), mais un paramètre plus "haut niveau", au sens où il fait appel à une comparaison du stimulus jugé aux connaissances antérieures du sujet, au contexte culturel, et à une multitude d'autres facteurs tout aussi insaisissables. Le jugement d'urgence est, de plus, largement influencé par le corpus de sons testés. Une plus ample définition de l'urgence perçue sera exposée au cours du chapitre 6. Nous proposerons ainsi un nouveau paradigme de mesure *objective* de l'urgence.

Chapitre 6

Evaluation de l'urgence par le TR

Ce Chapitre décrit une première série d'expérience visant à évaluer l'urgence des alarmes sonores par le biais des temps de réaction en contexte attentionnel. L'objectif de cette étude était triple ; il s'agissait :

- tout d'abord, de valider notre nouvelle proposition de paradigme de l'urgence, à savoir les temps de réaction, en utilisant des alarmes sonores variant selon une dimension rythmique ;
- ensuite, de tester ce paradigme dans deux conditions attentionnelles différentes, l'une d'elles devant simuler la conduite automobile, afin de tester la validité des paramètres des alarmes sonores dans plusieurs contextes réalistes ;
- enfin, de comparer les résultats obtenus dans le cadre du nouveau paradigme de temps de réaction à un jugement d'urgence perçue, pour pouvoir comparer nos résultats à ceux obtenus dans la littérature jusqu'alors.

Cette première étude est décrite dans le texte d'un article ayant pour titre "Evaluating warning sounds urgency with reaction times", soumis pour publication dans le *Journal of Experimental Psychology : Applied*. Ce texte est reproduit ci-après.

6.1 Introduction

The use of sounds to present information is becoming increasingly common in a diverse array of settings that range from hospital equipment to high performance aircraft. As noticed by Stanton & Edworthy (1999), it is not so surprising that hearing appears as a natural warning sense, because we cannot "shut our ears in the same way that we can [close] our eyes" (p.7). However, until a few years ago, warning signals were badly designed, and

Patterson first reported some of the problems typically associated with auditory displays from aviation or medical environments (Patterson et al. 1986 and Patterson 1990). One of the main problem he highlighted is that warning signals can become a distraction during times of high workload, instead of just attracting listener's attention and providing relevant information about a potential problem or sudden emergency.

On the basis of these established facts, Patterson (1990) proposed a pragmatic methodology in four steps to design properly warning sounds in an industrial and high workload environment. These four steps include determining the appropriate level of loudness, designing a small pulse of sound, incorporating the pulse into a longer burst of sounds, and forming a complete auditory warning signal using bursts of sounds followed by shorts periods of silence. His work is mainly focused on the first step, i.e. taking into account the background noise in which the sounds are to be heard. The three final steps concerned the interpretation of the warning signal (design of pulse, burst and warning). The author proposed that once the 'structure' of the auditory warning is designed, the perceived urgency communicated by the warnings may be altered by adjusting the pitch, the intensity and the speed of the burst.

In order to give precise recommendations on "urgency matching", i.e. to associate the appropriate warning sound to the appropriate urgent situation, knowledge of the effect of sound parameters on perceived urgency is required.

Edworthy and her colleagues thus conducted an extensive series of studies (Edworthy et al. 1991) showing that some pulse and burst parameters had clear and consistent effects on perceived urgency of a warning sound. Thus, the authors could define a series of pulse parameters such as fundamental frequency, harmonic series, amplitude envelope shape, delayed harmonics, as well as temporal and melodic parameters such as speed, rhythm, pitch range, and melodic structure to describe precisely what a warning sound would have to be. The subjective judgments indicated that the faster the rate, the higher the pitch, and the more randomly irregular the harmonics, the greater the perceived urgency.

These results have been confirmed in different studies (e.g. Hellier et al. 1993; Hellier & Edworthy 1999). Hellier et al.'s (1993) results revealed the power of the relationship between acoustic parameters (speed, fundamental frequency, repetition number and harmonic content) and perceived urgency. The authors showed that some parameters contribute more to perceived urgency than do others, with speed being the most efficient parameter to communicate urgency.

Edworthy and her colleagues (1991) also studied the effect of unpredictable events

(temporal or harmonic) on perceived urgency. They first found that an *irregular* rhythm is perceived as slightly less urgent than a regular one. However, in the irregular rhythm used by the authors, pulses were syncopated. Actually, a syncopated rhythm is a structure where one can perceive regularities. By definition (Fraisse 1982), arrhythmia is a sequence of continuous sounds where no temporal organization is perceptible. So, there is a question that still needs an answer : do temporal unpredictable events might be more attention-getting than others? Second, the authors found that a random harmonic irregularity corresponds to the most urgent warning. This finding was confirmed by other studies (Hellier et al. 1993, Burt et al. 1995). Hellier et al. (1993) however qualified this result showing that inharmonicity had a weak effect on perceived urgency if compared to other parameters such as fundamental frequency, speed or repetition.

In summary, the Edworthy et al.'s (1991) alarm-design principles appear to be relatively robust. For instance, their validity was tested again under the same conditions and verified by Hellier et al. (1993) and Guillaume et al. (2003). These studies were performed in order to resolve the so-called mismatch between perceived and situational urgency (Edworthy et al., 1991 ; Momtahan et al., 1993 ; Burt et al., 1995 ; Haas & Casali, 1995). Perceived urgency of a warning was defined as "its implicit urgency as a function of its sound parameters" and situational urgency as "the degree of urgency that the operator has learnt to associate with the warning as a function of the situation itself" (Edworthy et al., 1991). On the basis of the Patterson's four steps and these studies, urgency matching is thus possible : the most urgent warning sound could be associated with the most urgent situation.

6.1.1 What is called "urgency" ?

As noticed by Guillaume et al. (2003), this psychophysical approach addresses only "low-level" aspects of the problem. She conducted three experiments highlighting the role of cognitive factors in perceived urgency. The authors showed that although Edworthy et al.'s (1991) results were replicated with synthesized alarms, they were not entirely valid when they applied the same methodology to real alarms, recorded from alarms used in military aircraft in France. Some real alarms did not follow the predicted pattern, because of a possible highly learnt association between an alarm and its meaning. The authors concluded that the design of alarms should take into consideration the acquisition of a "mental representation". In other words, perceived urgency should be considered as a complex process, including a stimulus-driven level in which sensory data is transformed and combined into concepts of perception and attention, and a concept-driven level, focusing

on an individual's concepts and involving the deployment of existing knowledge that would generate an expectation.

However, even if we consider a more global approach of perceived urgency, it is not obvious that the final stage of the model should be a judgment of urgency. For instance, urgency ratings as predicted by Edworthy et al.'s studies (1991) were disturbed when presented in a task involving a high workload (Burt et al., 1995). All the same, in realistic conditions, an urgent situation often occurs under high workload. Auditory warnings thus signal potentially dangerous conditions or equipment malfunctions. As a result, the warning attracts attention, and the listener has to *react* immediately.

Whereas the "perceived urgency" approach has provided several useful experimental results, it is not clearly established which warning sound leads to a faster reaction when people are engrossed in an urgent situation. What is "urgency", and what do we really need to study? In the common sense, urgency is what requires immediate action or attention. Therefore, a warning signal is efficient when it increases the probability of an appropriate reaction under emergency conditions (Guillaume et al., 2003), which is usually associated with decreased reaction times (Bliss et al., 1995). Thus, the most relevant questions are : do we react more rapidly to some sounds than to others? If we do, which are the parameters that most improve our reaction time (in terms of a decrease in reaction time)? An answer to these questions should lead to a more appropriate understanding of the processes involved in urgent situations and to more useful recommendations of how to design powerful warning sounds.

To avoid possible differences between participants and studies due to the role of cognitive factors in urgency judgments, measures that can reflect the degree of perceptual and cognitive processing are needed. Instead of using behavioral studies in which a listener is asked to make some judgment about the sounds that are played (e.g. which of the two sounds is the more urgent), we also need an objective measurement, such as reaction time (RT). Since the work of Donders (1969), chronometric analyses of mental processes have been performed using the simple reaction-time paradigm. RTs were considered as an external source of information about an internal mental structure. RT data have played an important role in testing hypotheses about processes and structures. Thus, we could infer, perhaps, which parameter differences lead to different RT patterns. These differences reflected in response time will inform us about different paths through mental structure (for a complete review, see Luce, 1986).

We are not the first to propose such a measurement in the context of warning sounds.

To determine the degree to which a participant's response is due to these different processes, Burt et al. (1995) added physiological measures and reaction times to subjective assessment. One of the assumptions tested by the authors was that faster reaction times would occur in response to the most urgent warning. However, reaction times did not differ significantly in response to the urgency levels tested (parameters tested were fundamental frequency and harmonic series, as recommended by Edworthy et al., 1991). The authors supposed that this lack of urgency effect on reaction time was due to the warnings used in the experiment : they were rather homogeneous with respect to their range of spectral sound parameters. However, Edworthy et al. (1991) did find (weak) differences in urgency judgments with these same parameters. The distinction between these two studies thus may be due to differences in experimental paradigm. Consequently, prediction about a possible link between higher-urgency alarms and improvement in alarm reaction time was not really verified.

At the same time, Haas & Casali (1995) conducted a study investigating the effect of pulse format, pulse level and inter-pulse interval on *subjective* perceived urgency and *objective* reaction time. They found that only pulse format (sequential pure tones, simultaneous pure tones, or frequency-modulated tones) and pulse level significantly influenced RT. However, each pulse of the total signal had a duration of 350 ms and they found a mean reaction time of approximately 450 ms. So inter-pulse interval couldn't have an effect on reaction time. The authors did not comment this lack of effect of the inter-pulse interval on reaction time. In addition, the coefficient of variation (the ratio of the standard deviation to the mean) was approximately 0.3 (almost the same in all conditions). A usually good approximation estimates this value to be around 0.1 to 0.2 (Luce, 1986). The larger value found by Haas & Casali (1995) may have been due to the small number of measurements carried out (only two RT measurements for each signal). It is generally accepted that simple reaction time can be partitioned into at least two additive processes : perception time and motor time (Luce, 1986). The main problem is that overall RT measures thus contain a motor component which obscures the relation between stimulus and the actual "detection" component. A great range of variation in RT measurements is often due to a variability of the motor component (Luce, 1986). To avoid this problem, Crabtree and Antrim (1988) recommended repeating a measure at least forty times, for each experimental condition. The variation in RT measurement should then be minimized and potential differences between two experimental conditions could be interpreted as differences at the detection stage.

Finally, the last purpose of Haas & Casali's study (1995) was to investigate the rela-

tionship between perceived urgency and reaction times. The authors found a correlation between the two metrics : as perceived urgency increases, response time to the signal decreases. However, as just mentioned, inter-pulse interval had no effect on reaction time, but did have an effect on perceived urgency (an inter-pulse interval of 0 ms sounding more urgent than those of 150 ms or 300 ms). As in the Burt et al. (1995) study , this result could be seen as a counterexample of the supposed link between subjective and objective urgency. Indeed, these studies proposed an hypothesis concerning the influence of alarm urgency on response behavior : they assumed that higher-urgency alarms would appear more important to the listener, thus leading to faster reaction times. It was obviously not the case, at least for one of the parameters tested. There is consequently no obvious reason that can explain why perceived urgency judgments should produce a decrease in an objective measurement such as simple reaction time.

In industrial environments, listeners are often engrossed in alternative tasks that require attention. For this reason, a dual task paradigm is often used to determine the effect of the alarm and to evaluate attention and mental workload (Sorkin, 1988 ; Bliss et al., 1995). To show that performance on reaction time should improve when the alarm is well designed is not sufficient. It appears important to study the effect of different alarms in a realistic context, because real alarms are embedded within systems. Two studies (Burt et al., 1995 ; Haas & Casali, 1995) used a cognitive loading task which imposes attentional demands, as in military or industrial environments. They thus established that the Edworthy alarm-design principles resist high cognitive load. Burt et al. (1995) studied attentional engagement by manipulating two conditions of a (visual) tracking task : a manual and automated. They could thus show that participants provide slower reaction times during the manual tracking condition. They assumed that this result is due to the cognitive and/or physical interference of the tracking task with the push-button response. Indeed, attentional limitations were often observed in several studies in dual-task performance (for a review, see Pashler and Johnston, 1998).

6.1.2 The Current Experiments

In the current study, we examined some of the questions regarding the influence of acoustic parameters emphasized in the classical urgency literature on reaction time under high workload conditions.

In Experiment 1, we addressed the issue introduced by Haas & Casali (1995) and investigated the influence of InterOnset Interval (IOI) on reaction time. As highlighted before,

the authors did not find an effect of the IOI on reaction time. However, we hypothesize that this lack of effect is due to the pulse duration (350 ms); thus, the rhythm could not be heard before the listener reacted. In the present experiment, listeners were presented with a burst of pulses that changes in IOI, and were asked to press a button as soon as they heard the sound. At the same time, listeners were asked to perform simultaneously a primary task involving continuous tracking. The tracking task was designed to maintain a high level of attention at all times.

Experiment 2 was partly suggested by results obtained in Experiment 1. Stimuli of Experiment 1 were equalized in loudness in order to examine the effect of IOI removing any possible effect of loudness on reaction time.

In Experiment 3, we continued the examination of the influence of temporal parameters on reaction time and introduced temporal irregularity. As noticed previously, temporally unpredictable events might be more attention-getting than predictable ones.

At the end of each experiment, participants also performed a magnitude estimation task on the two different warnings (two different IOIs and two different temporal regularities). An urgency rating scale was used.

The first issue addressed in this research concerns how IOI and temporal irregularity affect participants' reaction times under high workload conditions. The second issue concerns a comparison between subjective and objective measurements.

6.2 Experiment 1

6.2.1 Method

Participants

Thirteen participants (nine male, four female) were recruited for this experiment. They were paid for their participation. None of them reported having hearing problems. All of them were right-handed and reported normal or corrected-to-normal vision.

Stimuli

The template for the two different stimuli was an isochronous sequence of short pulses. Each pulse of the burst was a 1-kHz pure tone, 20 ms in duration, and included 5-ms, linear on and off ramps. Stimuli only varied along a single dimension, the InterOnset

Interval (IOI). IOI is the time elapsed from the onset of one pulse to the onset of the next. The two IOIs were 100 ms and 300 ms. The total duration of each burst was 920 ms.

It should be noted here that an IOI variation leads to a variation in the number of pulses in a burst. As it is necessary in a RT task to keep the same total duration of stimuli, we cannot study IOI and number of pulses as independent factors. We accepted this inevitable fact, and will henceforth only use the IOI terminology.

Apparatus

The sound samples were generated with a 44.1-kHz sampling rate under the control of a PC, using Matlab v.7 software. The sound samples were amplified by a Yamaha P2075 stereo amplifier and presented binaurally over a Sennheiser HD 250 linear II headphone. Stimuli were presented at 76.5 dB SPL. The experimental sessions were run using a Max/MSP interface on an Apple computer. Participants responded by using the space bar of the computer keyboard placed on a table in front of them. The responses were recorded by Max/MSP, with a temporal precision for stimulus presentation and responses around 1 ms. The tracking task was created using Jitter, the graphical part of Max/MSP. Performance on the tracking task was recorded every 10 ms by calculating the distance between the target and the pointer. All data were collected in the computer memory for further off-line analysis. The experiments took place in a double-walled IAC sound booth.

Procedure

A 4-factor (two IOI \times two tracking conditions \times three blocks \times 25 repetitions), full-factorial, within-participants (repeated measures) design was used. Participants were treated as a random-effect variable. The remaining variables were treated as fixed-effect variables.

The tracking task consisted in psychomotor and visual conditions. During the psychomotor condition, participants were required to track a circular target manually, in order to keep it within a circular boundary that moved at a constant speed and a random trajectory on the screen. Participants had to perform it with their non-dominant hand. This enabled the primary task to be continuously demanding for each participant. Participants were instructed to optimize response speed *and* accuracy of the psychomotor task. It was necessary for the primary task to be challenging, so that the effort expended by the participants to perform it (and respond to the alarms) would approximate the effort experienced in many

complex task situations (de Waard, 1996). During the visual condition, participants were required to monitor computer tracking of the circular target and to perform only reaction time task.

Throughout the tracking task, one replication for each of the two auditory warnings was presented in random order. Following a standard RT procedure, participants had to respond as soon as they detected the sound by pressing the space bar as fast as possible. They were asked to keep their finger in contact with the space bar between trials. The intertrial interval was randomly fixed between 3 and 8 seconds.

Prior to data collection, a short practice period of approximately 3 minutes was provided. The two conditions (psychomotor and visual) were presented three times each in a counterbalanced order. Each participant thus participated in six experimental sessions, each session consisting of 50 stimuli. The stimuli of different IOI were randomly intermixed. The number of stimuli of different IOI was equal (25 each). RT scores were calculated for each IOI value in each block, attentional condition and repetition. We thus obtained 300 RT scores from each participant.

During this first part of experiment, no reference was made at any time to the concept of urgency or to alarms. After the first part of the experiment, subjects were informed about the goal of the study and were asked to perform the subjective task. They had to provide subjective urgency ratings by estimating the urgency of each warnings on a rating scale from 0 to 1 labelled 'not urgent' (0) to 'very urgent' (1). The entire experimental session lasted about 1 hour.

6.2.2 Results

RT task

Errors Errors included anticipations (RTs less than 100 ms), misses (failures to respond), and RTs greater than 1000 ms. Errors accounted for 0.5% of the trials on average. No anticipations and no misses were observed for the psychomotor task. There was only one miss for the visual task. Although there was only one RT greater than 1000 ms during the visual task, 19 were revealed in the psychomotor task. We assumed that number of errors was more salient when participants' processing "ressources" were more devoted to the tracking task. Nevertheless, RTs greater than 1000 ms did not follow any systematic pattern : the number of RTs greater than 1000 ms was 10 for IOI = 100 ms and 9 for IOI = 300 ms.

Reaction time data RT was defined as the time interval between the onset of a stimulus and the onset of a response. Only RTs of correct responses were analyzed. For each participant, we removed outliers that were more than two standard deviations below or above the mean (e.g. Penel and Jones, 2005 ; Arieh and Marks, 2003). Outliers accounted for 4.4% of the error-free trials. All these missing data were replaced by the mean of the corresponding condition for each participant.

Reliability of participant's response In order to study the variability of each participant's response, we also calculated the coefficient of variation (for the error and outlier-free trials). As noted by Woodworth and Scholberg (cited in Kohfeld et al. 1981), the coefficient of variation of a highly practiced participant under the most ideal RT conditions is typically 0.1 to 0.13. Green and Luce (1971) found a coefficient of variation somewhat larger, 0.2. This small difference between the two studies may reflect a difference between experimental designs, especially between fixed and random foreperiods. (In classical RT studies, foreperiods are defined as the time elapsed between the "alert signal" and the onset of the "reaction signal" ; the alert signal makes the participant aware that it is time to attend to the reaction signal. In our case, we did not have any alert signal, and we randomized the time between two successive reaction signals). The coefficient of variation of each participant ranged from 0.09 to 0.2 in the visual task, with a mean of 0.14, and from 0.12 to 0.24 in the psychomotor task, with a mean of 0.18. The difference between the two tasks could be due to the dual-task interference. These human dual-task performance limitations as well as the experimental design proposed by Luce (1986) could also explain the change of the coefficient of variation from 0.1 to 0.2. Since the variability in our RT measures is small and follows well what is found in most simple RT experiments, we can assume that any mean differences among the conditions in our study are reliable.

The data were then analyzed as a function of IOI, attentional condition, block and repetition. Statistical analyses were carried out using a repeated-measures analyses of variance. p-values were corrected using the Huynh-Feldt correction.

A 2 (IOI) \times 2 (attentional condition) \times 3 (block) \times 25 (repetition) ANOVA was performed. Results are shown in Table 6.1.

Main effect of IOI First, the data show that mean RT decreases when IOI decreases. Figure 6.1 depicts mean RT as a function of IOI in the psychomotor and visual attention condition. Each bar represents the mean of 975 RT trials (75 per participant).

Source	df	SS	MS	F	ϵ	p
IOI	1	68905	68905	37.09	1	0.00005
Repetition	24	225477	9395	2.85	0.8	0.01
Attentional condition	1	6784423	6784423	114.53	1	0.00000017
Block*Task	2	89062	44531	13.72	0.8	0.00009

TAB. 6.1 – ANOVA table for Experiment 1. IOI : Inter Onset Interval. SS : sum of squares, MS : mean square, F : F-values, ϵ : Huynh-Feldt correction factor applied to the degrees of freedom, p : p-values. Only effects significant at the $p < 0.05$ level are presented

This tendency is common to the two attentional conditions. For the psychomotor condition, the decrease is 11 ms. When IOI = 300 ms, the mean RT is 384 ms, and the standard deviation is 91 ms. When IOI = 100 ms, the mean RT is 373 ms, and the standard deviation is 85 ms. For the visual condition, the decrease is 6 ms (298 ms ; 59 ms for IOI = 300 ms and 292 ms ; 56 ms for IOI = 100 ms).

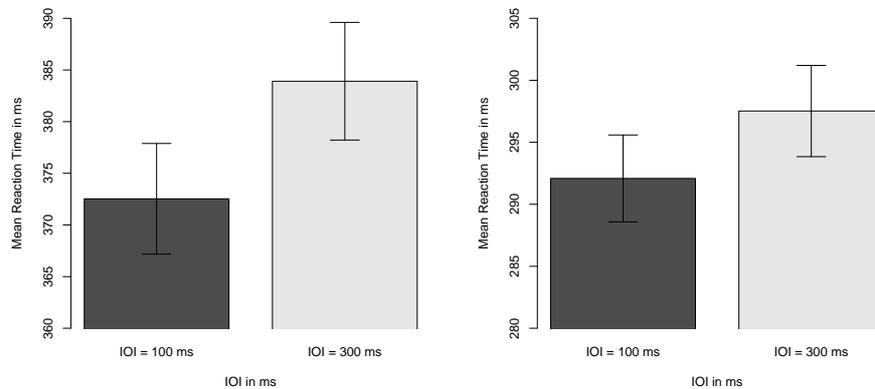


FIG. 6.1 – Mean RT as a function of IOI. On the left : results for the psychomotor task. On the right : results for the visual task. The vertical bar represent 95% confidence intervals.

Main effect of attentional condition Participants had significantly longer reaction times with the psychomotor condition (378 ms ; 88 ms) than with the visual one (295 ms ; 57 ms). The finding that mean RT increased when attention became higher is consistent with previous studies (Burt et al., 1995 ; Bliss et al., 1995).

Main effect of repetition It appears difficult for the participants to always maintain a high level of attention. However, the non significant IOI \times Repetition interaction indicates that the shape of the RT-IOI function did not change markedly during the experimental session. The same comment could be done for all the other interactions with the Repetition factor. Moreover, the non significant main effect of the block indicates that the fatigue factor could be neglected in the main interpretation of the IOI effect, which is, in our opinion, the most interesting finding concerning this Repetition factor.

Block by task interaction In the visual condition, mean RT is almost constant over the three blocks (Block 1 : 293 ms ; Block 2 : 296 ms ; Block 3 : 295 ms), whereas it decreases in the psychomotor condition (Block 1 : 389 ms ; Block 2 : 378 ms ; Block 3 : 368 ms). This interaction shows that participants' performance on the RT task was improved during the psychomotor condition.

Performance on the psychomotor tracking task

In supplementary steps, we analyzed a possible heightened effect of the primary task through the three blocks. Each 10 ms, the position of the target and the pointer was saved. We calculated the distance between the target and the pointer in the three blocks each 10 ms. For each block, the mean and the standard deviation of all these distances were calculated. We thus performed a 1-way ANOVA to investigate the effect of a possible improvement through blocks. The data show a nonsignificant effect of the blocks on the mean of the distance between target and pointer ($p = 0.15$), and a significant effect of the standard deviation ($p = 0.05$). A somewhat learning effect could be seen in the significant improvement of the standard deviation : performance on the tracking task was more stable and continuous in the third block.

Urgency judgment

Participants rated signal with an IOI of 300 ms (mean magnitude estimation of 0.2, standard deviation of 0.1) as sounding less urgent than the one with an IOI of 100 ms (0.7 ; 0.2). A t-test showed a significant effect of the IOI ($p \leq 0.0001$). This result was as expected. Previous findings (Edworthy et al., 1991 ; Hellier et al., 1993) are supported once again.

6.2.3 Discussion

One of the issues of this study concerned the comparison between subjective and objective measurement. In their study, Haas and Casali (1995) examined the relationship between perceived urgency and reaction time to warning signals. Their results suggested that there was a correlation between the two measurements : as perceived urgency increased, reaction time decreased. This conclusion was not verified for the inter-pulse interval factor (similar to the IOI factor used in our experiment), because pulse duration was too long compared to mean RT. The outcome of Experiment 1 thus extends these previous results to the IOI factor. Furthermore, urgency rating performance presumably depends on the corpus of sounds investigated. It is possible that the fast-rate sound was rated as the most urgent because "classical" and existing warning sounds were already varying along a "speed" dimension. Therefore, whereas one could see a link between perceived urgency and reaction time, this correlation cannot explain why RT to the fast rate is lower than to the slow rate.

The second aim of this experiment concerned how IOI would affect participants' reaction times under high workload conditions. The outcome of Experiment 1 first showed that there is a straightforward relation between attention and RT : specifically, RTs in the psychomotor condition were longer than RTs in the visual attention condition. This result provided support for previous findings (Burt et al., 1995; Bliss et al., 1995). From a more theoretical point of view, Pashler and Johnston (1998) reviewed different models of attentional limitations based on empirical evidence. When attention increases, reaction times also generally increase. As already said, this finding strongly suggests the existence of capacity limits in perceptual analysis. Roughly speaking, it means that a portion of the capacity is allocated for tracking task performance, and we could suppose that the more difficult the primary task, the higher the mean reaction time. Of primary interest here is also the finding that we did not observe a significant interaction between task and IOI. Whatever the attentional condition, the IOI effect was observed. IOI effect should therefore be discussed separately, as apparently reflecting a general process that can occur in different attentional conditions.

Actually, this study addressed the relationship between IOI and reaction time in a *new experimental paradigm*. Haas & Casali (1995) previously investigated the effect of inter-pulse interval on reaction time. However, they did not find a significant effect of the inter-pulse interval on reaction time. As was mentioned in the Introduction, inter-pulse interval couldn't have an effect on reaction time because the pulse duration used was as large as

the mean reaction times obtained. We thus designed our experiment (with more reaction times measurements and smaller IOI) in order to provide an answer to this unresolved question concerning the IOI effect. Results obtained in Experiment 1 indicate that there is a substantial effect on RT when IOI is manipulated, that is to say that RT is lower for the higher IOI value. One could explain this result by a specific IOI effect.

There is at least one other possibility that could however explain these results. Stimuli were designed to vary in terms of IOI and were presented at a level of 76.5 dB SPL. When listening to these two stimuli, it is easy to hear that although the sound level fluctuates, one can still perceive an overall loudness of the sound. Based on recent studies concerning the loudness of modulated sounds, the two different time-varying sounds presented during the experiment at the same maximum level may have different loudnesses (for reviews, data and models on the loudness of amplitude-modulated sinusoidal carriers, see Moore et al., 1999; Glasberg and Moore, 2002). Figure 6.2 shows main results of the predictions of Glasberg and Moore's time-varying model. The short-term loudness for the fast rate (IOI = 100 ms) is slightly higher in response to the second pulse (69.4 phons) than in response to the first (68.3 phons) : a form of temporal integration occurs. This effect is negligible for the slow rate (IOI = 300 ms), for which the loudness was 68.3 phons for the first and the second pulse. There is a very weak, if still noticeable, loudness difference between the two sounds. The second pulse of the slow rate sound and the fourth pulse of the fast rate sound both occurred at $t = 320$ ms. As can be seen in Figure 6.2, short-term loudness at $t = 320$ ms for the fast rate is higher (69.5 phons) than the one for the slow rate (68.3 phons). The same values are observed at $t = 620$ ms and $t = 920$ ms.

Early experiments (Chocholle, 1940) showed that simple reaction time decreased monotonically when sound intensity increased, and that equally loud stimuli produced equal RTs regardless of stimulus frequency. Chocholle's data indicate that RT depends on loudness and not on physical intensity. However, as noted by Chocholle himself, some of the listeners heard clicks at the signal onset (due to the equipment of his day), and that may have had an influence on the results. Kohfeld et al. (1981) partially replicated Chocholle's experiments, and found a less clear relationship between RT and loudness : they were sometimes imperfectly correlated. They showed that equally loud tones produced equal RTs at high levels (60 and 80 phons) but not at low levels (20 and 40 phons). The equal-loudness with equal-RT relation was restored only when signal intensity increased. These data should be interpreted with caution, because the equal loudness contours reported differ a great deal from those obtained in most other studies (Luce, 1986). Buus et al. (1982) investigated the

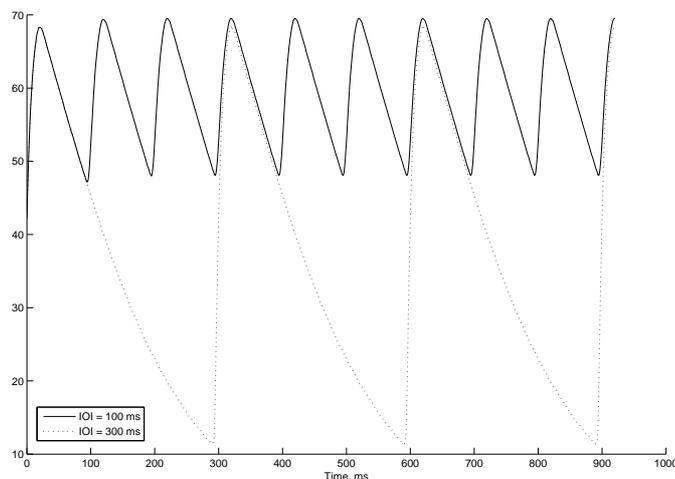


FIG. 6.2 – Time-varying model (Glasberg & Moore, 2002) showing short-term loudness, plotted as a function of time, in response to the two stimuli.

effect of the loudness-matching procedure used by Kohfeld et al. (1981) and showed that reaction times to equally loud tones were approximately equal if the loudness-matching procedure is a more "classical" one, such as the method of adjustment.

Wagner et al. (2004) also evaluated the relation between loudness and reaction time, but with narrowband and broadband noises. These stimuli differ from those in the majority of the literature which have almost always focused on pure tones. Their result showed, once more, that RT is closely related to loudness.

One way to conceive the "loudness-reaction time" results is in terms of Grice's (1968) model. He proposed an accumulation model to account for the effect of intensity on simple reaction time. The model assumes that the onset of a stimulus initiates a sequence of hypothetical neural events. He suggests that this information accumulates over time in the sensory pathways. When this information reaches a certain criterion level (enabling the system to distinguish external stimulation from internal noise), the system responds. Because information accumulates more rapidly when intensity is higher, stronger signals reach the response criterion sooner than weaker signals, leading to quicker responses. Thus, in RT experiments, the latency of a system response is a measure of the time required by these neural events to reach the criterion value. The rate at which the events accumulate is determined by stimulus intensity. This model was applied (Kohfeld, 1971), and fit his data

well. This decision model provides one way to account for the effects of signal intensity in auditory RT conditions. Even though the model appears quite simple, it is still valuable in accounting for psychophysical data. In other words, a simplistic model of intensity coding is to consider that intensity is encoded in the same manner for all frequency channels, and that the higher the intensity, the higher the firing rate of fibers in the auditory nerve (Plack and Carlyon, 1995). Loudness may thus be considered as depending on a central response derived from a weighted sum of spikes in the auditory nerve (Florentine et al., 2003).

In summary, it might be possible that the results of Experiment 1 could be explained by a loudness effect more than an IOI effect. In our view, a reasonable explanation for our results is based more on a specific effect of IOI, even if loudness might be responsible for a part of the results. Indeed, as mentioned previously, the loudness difference between the two sounds might be too small to explain the effect observed on the RT. Still, the problem remains that our stimuli were not equalized for global loudness. Experiment 2 was performed in order to investigate this possibility.

6.3 Experiment 2

6.3.1 Method

Participants

Thirteen other participants (eight male, five female) were recruited for this experiment. They were paid for their participation. None of the participants reported having hearing problems. All of them were right-handed and reported normal, or corrected-to-normal vision.

Stimuli

A loudness equalization was performed on the two stimuli described in Experiment 1. A group of fourteen listeners participated in this preliminary experiment. Loudness matches between the fast- and slow-rate sounds were obtained with an adjustment procedure. The listener was asked to adjust the comparison stimulus until it seemed equal in loudness to the standard stimulus. The fast-rate sound was used as the comparison stimulus and the slow-rate one as the standard. Level for the slow-rate stimulus was fixed to 76.5 dB, as in Experiment 1. The mean level difference at which the slow-rate and fast-rate sounds were

judged equal in loudness was 5.4 dB SPL. As explained before, Buus et al. (1982) showed that tones matched in loudness by this conventional method yielded nearly identical RTs.

Apparatus and procedure

These were the same as in Experiment 1, except that we did not include the urgency judgment. The slow-rate and fast-rate stimuli were presented at 76.5 dB SPL and 71.1 dB SPL respectively.

6.3.2 Results

RT task

Errors Errors were defined as in Experiment 1. Errors accounted for 0.3% of the trials on average. No anticipations, three misses and no RTs greater than 1000 ms were observed for the visual task. Although there were no anticipations and no misses during the psychomotor task, nine RTs greater than 1000 ms were obtained in the psychomotor task. As in Experiment 1, the number of errors was more salient when participants' processing "ressources" were more devoted to the tracking task.

Reaction time data Data were obtained as in Experiment 1. Only RTs of correct responses were analyzed. Outliers accounted for 4.6% of the error-free trials. All missing data (outliers and errors) were replaced by the mean of the corresponding condition for each participant.

As explained before, we calculated the coefficient of variation for each participant in order to test response reliability. The coefficient of variation of each participant range from 0.09 to 0.17 in the visual task, with a mean of 0.14, and from 0.1 to 0.27 in the psychomotor task, with a mean of 0.16. As noticed in the first experiment, the difference between the two tasks could be due to dual-task interference. Since the variability in our RT measures was small and followed well what is found in most simple RT experiments, we can assume once again that any mean differences among the conditions in our study are reliable.

A 2 (IOI) \times 2 (attentional condition) \times 3 (block) \times 25 (repetition) ANOVA was performed. Results are shown in Table 6.2.

Main effect of IOI First, and as in Experiment 1, the data show that the mean RT decreases when the IOI decreases. Figure 6.3 depicts mean RT as a function of IOI in the psychomotor and visual attentional conditions. Each bar represents the mean of 975 RT

Source	df	SS	MS	F	ϵ	p
IOI	1	25253	25253	8.61	1	0.01
Repetition	24	119500	4979	2.41	0.5	0.03
Attentional condition	1	2976939	2976939	47.9	1	0.000016

TAB. 6.2 – ANOVA table for Experiment 2. IOI : Inter Onset Interval. SS : sum of squares, MS : mean square, F : F-values, ϵ : Huynh-Feldt correction factor applied to the degrees of freedom, p : p-values. Only effects significant at the $p < 0.05$ level are shown.

trials (75 per participant). For the psychomotor condition, the decrease is 8 ms. When IOI = 300 ms, the mean RT is 327 ms, and the standard deviation is 67 ms. When IOI = 100 ms, the mean RT is 319 ms, and the standard deviation is 63 ms. For the visual condition, the decrease is 2 ms (269 ms ; 49 ms for IOI = 300 ms and 267 ms ; 49 ms for IOI = 100 ms).

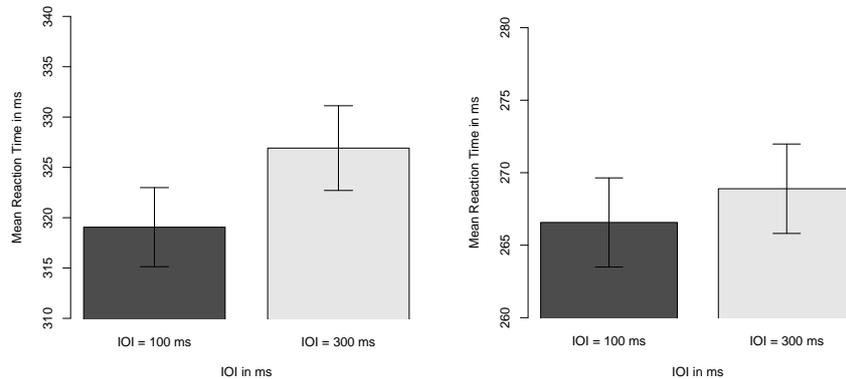


FIG. 6.3 – Mean RT as a function of IOI. On the left : results for the psychomotor task. On the right : results for the visual task. The vertical bars show the 95% confidence intervals.

Main effect of Repetition Interpretation of this Repetition effect is exactly the same as in Experiment 1 : it seems to be difficult for the participants to always maintain a high level of attention. The nonsignificant main effect of the block as well as the nonsignificant IOI \times Repetition interaction indicates that the fatigue factor can be neglected in the main interpretation of the IOI effect, as in Experiment 1.

Main effect of Attentional Task Participants had significantly longer reaction times with the psychomotor condition (323 ms; 65 ms) than with the visual one (268 ms; 49 ms). As in Experiment 1, and consistently with previous studies (Burt et al., 1995; Bliss et al., 1995), mean RT increased when attentional demands were increased.

Performance on the psychomotor task

As explained in Experiment 1, we performed a 1-way ANOVA to investigate the effect of learning over blocks. The data show a nonsignificant effect of the blocks on the mean of the distance ($p=0.35$) and on the standard deviation ($p=0.37$) between target and pointer. Participants' performance was stable and continuous.

6.3.3 Discussion

The major purpose of this experiment was to examine whether RT differences observed in Experiment 1 were attributable to a loudness or an IOI effect. The approach used in the current study was still a dual-task methodology. The procedure and the methodology were exactly the same as in Experiment 1 except that the stimuli used (IOI = 300 ms and IOI = 100 ms) were presented at equal-loudness levels.

As noticed before, RT experiments have been suggested to provide an indirect estimate of loudness (Wagner et al., 2004; Arieh and Marks, 2003). Good fits between equal-loudness and equal-RT data have been indeed found in studies using pure-tone stimuli (Chocholle, 1940; Pfingst et al., 1975), and narrow or broadband noises (Wagner et al. 2004). All these data support the fact that loudness and RT are closely related. If loudness was the only factor responsible for the RT effect observed in Experiment 1, no RT differences should be observed in the current experiment where stimuli are equalized in loudness.

Nevertheless, the most significant feature of this set of results is that they showed similar RT patterns to those of Experiment 1 : RTs are smaller for IOI = 100 ms than for IOI = 300 ms, and this tendency is common to the two attentional conditions (Experiment 1, psychomotor condition : 373 ms and 384 ms, respectively; Experiment 2, psychomotor condition : 319 ms and 327 ms, respectively). Overall, this experiment provides evidence that reaction times are sensitive to an IOI effect.

However, the weight of this evidence can be questioned on at least one respect. When compared to the preceding results of Experiment 1, we observed that the size of the IOI effect became relatively smaller with sounds equalized in loudness. When the two stimuli

were presented at equal-loudness levels (Experiment 2), mean RT decreased by 8 ms when IOI decreased whereas the decrease was 11 ms for Experiment 1 (results for the psychomotor condition). For the visual condition, the decrease was 6 ms for Experiment 1 and only 2 ms for the current experiment. This finding suggests that IOI has a real and consistent effect on RT and supports the idea that loudness and RT are closely related, even if the loudness effect observed is weak. The most likely explanation for results obtained actually lies on a possible combined effect of loudness and IOI on RT. In Experiment 1, the RT difference observed between the two sounds was possibly due both, whereas in Experiment 2 this difference could only be due to an IOI effect. In regard to the literature on RT and loudness (e.g. Chocholle, 1940; Pfingst et al. 1975; Wagner et al. 2004; Arieh and Marks, 2003), it may be argued that this smaller RT difference observed between the two stimuli in Experiment 2 compared to Experiment 1 could be due to a loudness effect more than inter-individual differences.

Even though this loudness effect has to be taken into account, data obtained in Experiment 2 still highlighted an IOI effect on RT. This effect can be viewed as an indirect measure of temporal response of the auditory system to time-varying changes in acoustic parameters. The psychoacoustic literature has seen a great amount of data and theory in order to describe and interpret temporal phenomena. However, the time constants measured indirectly in these studies ranged from 0.25 ms to 200 ms (Eddins and Green, 1995).

Major conflict between these time constants comes from a fundamental difference between temporal *integration* and temporal *resolution* in the auditory system and between the experimental tasks used. As temporal integration concerns the link between signal duration and its detectability, temporal resolution is defined as "the ability of the auditory system to respond to rapid changes in the envelope of a sound over time" (Viemeister and Plack, 1993). Temporal integration can be seen, for example, in the growth of loudness with duration (e.g. Florentine et al., 1988), and can be modelled by a long time window (e.g. Zwislocki, 1960; Zwislocki, 1969), with a maximum duration of the temporal integrator at about 100 - 200 ms. However, the auditory system is also capable of fine temporal resolution, as proved by temporal gap detections, for instance (Plomp, 1964; for a review of all the studies that have confirmed his results, see Eddins and Green, 1995). Some aspects of temporal resolution are relevant to this study, in the sense that we ask whether the detection difference between the two IOI stimuli could be due to a sort of temporal resolution.

Evidence supports the view that results obtained from temporal acuity studies (Plomp,

1964; Viemeister, 1979) are a consequence of temporal summation processes in the central auditory nervous system (Moore et al., 1998; Plack and Moore, 1990). The basic idea of these two influential studies was that internal representation of auditory signals is smoothed in time by a sliding temporal window for each frequency channel. Output of this temporal integrator leads to a "temporal excitation pattern" (TEP).

This temporal window model is one part of a four-stage model classically used to describe temporal resolution of the auditory system. The first two stages are auditory filters (cochlear processing implemented as a gammatone filterbank, see Irino and Patterson, 1996), followed by a nonlinearity simulating basilar membrane-based compression (Plack and Oxenham, 1998). The two final stages (temporal window and decision device) are assumed to model more central mechanisms.

In order to characterize the window function, Moore et al. (1988) introduced the equivalent rectangular duration (ERD), which is thus a measure of the temporal resolution of the auditory system (Moore et al., 1993; Peters et al., 1995). Moore et al. (1996) found that the optimal duration of the ERD is around 7 ms at all frequencies. To summarize, the low-pass filtering stage, described by an ERD, was intended to simulate the temporal resolution limit by attenuating rapid changes in the envelope of the signal.

One of the positions often adopted in the literature to account for the "resolution - integration" paradox is to assume that different time constants coexist in the auditory system, and that the auditory system "chooses" one of these different modes of processing according to the task. In order to resolve this unsatisfactory paradox, Viemeister and Wakefield (1991) proposed a unified model : the "multiple-look" model. They argued that, rather than a single integration function with a particular time constant (a short one or a long one), the auditory system takes multiple "looks" at the ongoing signal and combined the information across "looks" to improve performance. Their "multiple-look" model is a good attempt to unify the perspectives of integration and resolution, as shown by their own experiments involving the detection of two separate tone bursts.

Our data could then be more easily explained and interpreted within this framework. According to this model, the output of a first stage, let say the auditory filters, is sampled at a high rate. These samples ("looks") are stored in memory. Then, looks are combined to reach a decision statistic. Because the looks are based on a short-time constant (around 5 ms), this model seems to explain, at least qualitatively, our results : at each pulse, the looks are independent and convey information that is combined in an "intelligent" fashion. The different looks are used and *selected* in the best manner in order to optimize the task. The

fast-rate sound seemed to improve performance by repeatedly sampling. As the number of independent samples increased, the variance decreased and performance increased. We can remark that, in our case, the Moore et al.'s (1988) model could also well explain the data of Experiment 2.

6.4 Experiment 3

Despite the fact that a small difference between Experiment 1 and Experiment 2 on RTs obtained were probably due to the loudness difference, we still observed an IOI effect. This first temporal parameter seemed to have a clear and consistent effect on RT, and extended previous results of the urgency literature (Edworthy et al., 1991 ; Hellier et al., 1993). In Experiment 3, we repeated the same paradigm as before with sounds varying along a temporal irregularity dimension. Unpredictable events such as irregularity have indeed been suggested to be more attention-getting than the others. However, a clear experimental design and additional results are still needed to support this hypothesis.

6.4.1 Method

Participants

Thirteen other participants (seven male, six female) were recruited for this experiment. They were paid for their participation. None of them reported having hearing problems. All participants were right-handed and reported normal or corrected-to-normal vision.

Stimuli

The template for the two different stimuli was a sequence of five short pulses, with a total duration of 540 ms. Each pulse of the burst was a 1-kHz pure tone, 20 ms in duration, and included 5-ms, linear on and off ramps. Stimuli only varied along a single dimension : temporal regularity. Two temporal patterns were chosen, a regular and an irregular one. The regular one had an IOI of 130 ms. The irregular one was constructed on the basis of this basic sequence, by displacing the second, third and fourth pulses of 40 ms, which resulted in a lengthening or shortening of the intervals in the sequence. The IOI sequence was : {90 ; 210 ; 50 ; 170} ms. The two stimuli are presented in Figure 6.4, where the left figure is the regular one and the right represent the irregular one.

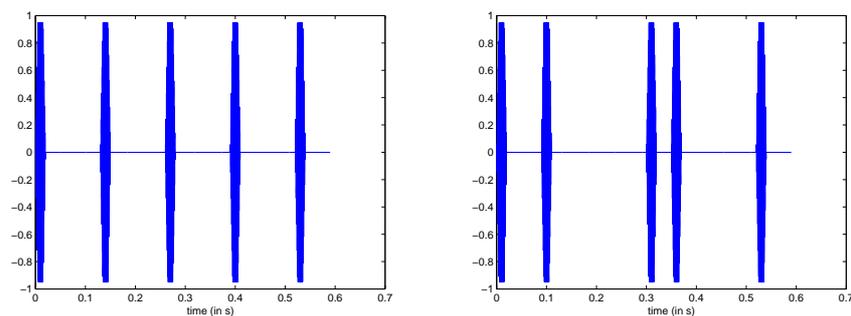


FIG. 6.4 – Waveforms of the stimuli used in Experiment 3. Left panel : the regular sound. Right panel : the irregular sound.

Apparatus and Procedure

These were the same as in Experiment 1.

6.4.2 Results

RT task

Errors Errors were defined as in Experiment 1. Errors accounted for 0.7% of the trials on average. No anticipations, three misses and one RT greater than 1000 ms were observed for the visual task. Although there were no anticipation during the psychomotor task, five misses and twenty RTs greater than 1000 ms were obtained. As in Experiments 1 and 2, the number of errors was greater when participants' processing "ressources" were more devoted to the tracking task.

However, RTs greater than 1000 ms seemed to follow a systematic pattern : the number of RTs greater than 1000 ms was 14 for the regular stimulus and only six for the irregular one. The process that generates errors like RT greater than 1000 ms can be due to participant's inattention, or guesses based on participant's failure to reach a decision. These errors could be eliminated of the data of real interest only if they do not change markedly the potential significance of the real processes under study. During this experiment, 70% of RTs greater than 1000 ms resulted from the regular sound. This result may suggest that temporal irregularity capture a listener's attention and tended to trigger responses.

Outliers Outliers were defined as RTs that were more than two standard deviations below or above the mean, as in Experiments 1 and 2. As for errors studied before, outliers are reaction times generated by processes that are not the ones being generally studied. In a classical RT study, it is desirable to eliminate outliers from the data. However, to do so, distributions of RT with and without outliers should be similar. If not, it means that some of outliers follow a systematic pattern along the dimension under study, and thus convey information that should not be eliminated. Outliers accounted for 5.2% of the error-free trials, i.e. 204 measurements. 107 of them were obtained during the psychomotor condition and 97 during the visual condition. There was also a difference, even weak, of the number of outliers between the two attentional conditions. For the psychomotor condition, 61 of the outliers obtained were in response to the regular sound, and 46 of them in response to the irregular one. In terms of percentage, 57% of the outliers were due to the regular sound. For the visual condition, 54 of the outliers were due to the regular sound (56%), and 43 to the irregular one. This observation seems to corroborate the suggestion concerning errors presented above. Outliers and errors should convey important information. We thus decided to analyse the data in more detail by comparing RT data *with* and *without* outliers and errors.

Reaction time data without outliers and errors These data were obtained as in previous experiments. All missing data (outliers and errors) were replaced by the mean of the corresponding condition, for each participant. As explained in Experiment 1, we calculated the coefficient of variation for each participant in order to test response reliability. The coefficient of variation of each participant ranged from 0.11 to 0.17 in the visual task, with a mean of 0.14, and from 0.11 to 0.25 in the psychomotor task, with a mean of 0.17. As noticed in the first experiment, the difference between the two tasks could be due to the dual-task interference. Since the variability in our RT measures is small and well follows what is found in most simple RT experiments, we can assume once again that any mean differences among the conditions in our study are reliable.

A 2 (temporal irregularity) \times 2 (attentional condition) \times 3 (block) \times 25 (repetition) ANOVA was performed. Results are shown in Table 6.3.

Main effect of attentional condition As in Experiments 1 and 2, participants had significantly longer reaction times with the psychomotor condition (381 ms ; 94 ms) than with the visual one (285 ms ; 60 ms). This result was as expected.

Source	df	SS	MS	F	ϵ	p
attentional condition	1	8952798	8952798	66.84	1	0.000003
repetition	24	225136	9381	3.06	0.7	0.007
block*task	2	46614	23307	7.57	0.5	0.01

TAB. 6.3 – ANOVA table for Experiment 3. RT data without errors and outliers. SS : sum of squares, MS : mean square, F : F-values, ϵ : Huynh-Feldt correction factor applied to the degrees of freedom, p : p-values. Only effects significant at the $p < 0.05$ level are shown.

Main effect of repetition The conclusions are similar to those of Experiments 1 and 2.

Block \times task interaction In the visual condition, the mean RT slightly increased over the two first blocks (Block 1 : 281 ms ; 61 ms - Block 2 : 288 ms ; 59 ms - Block 3 : 286 ms ; 59 ms), whereas it decreased over the three blocks in the psychomotor condition (Block 1 : 387 ms ; 114 ms - Block 2 : 379 ms ; 87 ms - Block 3 : 378 ms ; 78 ms). As in Experiment 1, this interaction shows that participants' performance on the RT task improved during the psychomotor condition compared to the visual condition.

As showed in Figure 6.5 on the left, it is not clear how the pattern of RTs as a function of temporal irregularity differed in the two cases ; the ANOVA did not show any significant effect for the temporal regularity factor.

Therefore, as explained before, errors followed a systematic pattern. Results of this experiment thus couldn't be analyzed as in classical RT experiments : errors and outliers should be included in the RT data, as they potentially conveyed some information.

Reaction time data including errors and outliers A 2 (temporal irregularity) \times 2 (attentional condition) \times 3 (block) \times 25 (repetition) ANOVA was performed. Results are shown in Table 6.4.

Temporal irregularity effect Of special interest here was this effect. Including errors and outliers, temporal irregularity had a significant effect on RT. The data showed that RT are faster for temporal irregularity. Figure 6.5 on the right depicts mean RT as a function of temporal irregularity in the psychomotor attentional condition. Each bar represents the mean of 975 RT trials (75 per participant). This tendency is common to

Source	df	SS	MS	F	ϵ	p
attentional condition	1	12448080	12448080	45.79	1	0.00002
repetition	24	981328	40889	2.94	0.5	0.01
temporal irregularity	1	60172	60172	8.98	1	0.01

TAB. 6.4 – ANOVA table for Experiment 3. RT data with outliers and errors. SS : sum of squares, MS : mean square, F : F-values, ϵ : Huynh-Feldt correction factor applied to the degrees of freedom, p : p-values. Only effects significant at the $p < 0.05$ level are shown.

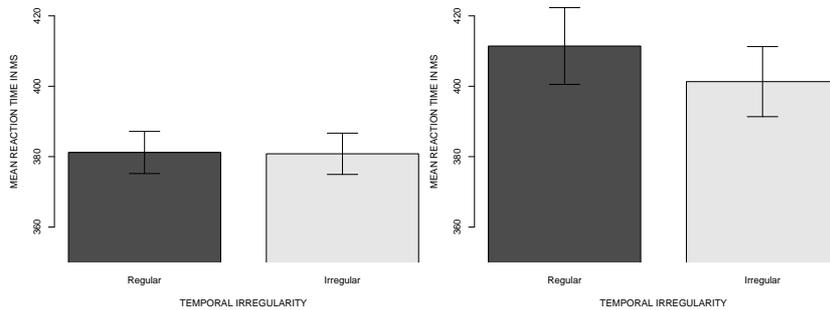


FIG. 6.5 – Mean RT as a function of temporal differences. On the left, RTs are calculated without errors and outliers. On the right, all data are included.

the two attentional conditions. For the psychomotor condition, RT are 10 ms faster for the temporal irregularity (411 ms; 173 ms for the regular sound and 401 ms; 158 ms for the irregular one). For the visual condition, RT are 6 ms faster for the temporal irregularity (296 ms; 81 ms for the regular sound and 290 ms; 72 ms for the irregular one). This significant effect of the temporal irregularity in this special case (i.e. including errors and outliers) confirmed the "arousal" effect of temporal irregularity evoked previously.

Other effects Interpretations of attentional condition and repetition effects are the same as previously : on one hand, a primary task effect, and on the other hand, an improvement in participants' performance.

Performance on the psychomotor tracking task

As explained in Experiment 1, we performed a 1-way ANOVA to investigate the effect of learning through blocks. The data show a nonsignificant effect of the blocks on the mean of the distance between target and pointer ($p=0.1$), and a marginally significant effect of

the standard deviation ($p=0.08$). There was a small effect of learning, and still, participants performance was stable and continuous.

Urgency judgment

Participants rated regular sounds (mean magnitude estimate of 0.5 standard deviation of 0.2) as sounding as urgent as irregular ones (0.5, 0.2). This result was confirmed by a t-test which showed a non significant effect of the temporal regularity ($p=0.8$).

6.4.3 Discussion

This experiment was conducted to gain further insight into the effects of temporal irregularities on mean RT during time of high workload. The current study extend previous studies of urgency judgments of sound based on temporal differences. The effect of an irregular rhythm on subjective urgency was investigated in Edworthy et al.'s (1991) study. They found that an irregular rhythm is perceived as less urgent than a regular one. However, as mentioned in the Introduction, the stimulus used was a syncopated rhythm, i.e. a metric rhythm. Metric rhythms are defined as displaying multiple levels of periodic temporal structuring. In other words, the temporal irregularity factor was partly unexploited in Edworthy et al.'s study (1991). As a matter of fact, metric rhythms are easier to remember and reproduce than nonmetric rhythms (Essens and Povel, 1985). Therefore, it seems interesting to investigate the structure of metric vs non metric rhythms in participants while they listened to regular or irregular sound. Moreover, the effect of temporal irregularities on objective measurements such as RT has never been studied.

The first outcome of the third experiment was the study of RT in response to temporal irregularity-based stimuli. We asked whether a significant increase or decrease in RT would be observed in a detection task in which the temporal irregularity is manipulated.

The most readily explainable pattern in our results is that of errors and outliers, which highlighted the idea of "capture" by temporally perturbed events. Errors and outliers showed a kind of "arousal effect" of temporal irregularities. We found that temporal perturbations tended to trigger responses : RTs greater than 1000 ms and outliers occurred more often in response to regular than to irregular stimuli. This result suggests that when participant's attention was *possibly* lower than it should have been, temporal irregularities aroused the participant more than a regular sound did.

The Dynamic Attending Theory (Jones, 1976 ; Jones and Boltz, 1989) proposes a theo-

retical framework of how listeners respond to and use tempo and time hierarchies. Jones claims that time periods play an intrinsic role in the relational structure of everyday events. Jones suggests that the degree of rhythmic regularity of an event time structure affects how effectively we attend to it. This theory is based on the principle of synchrony. The author assumes that people perceive a rhythm in relation to the activity of a small system of internal oscillations. Attending is thus described as the synchronization of these internal attending periodicities with the external rhythm of an auditory sequence. The relationship between this system of internal oscillations and the external rhythm controls listeners' perception of temporal intervals. This internal system uses a type of continuous "beat" as an extrapolation of a stable period even in the absence of a "driving rhythm" (i.e. an external rhythm, a characteristic flow pattern produced by a sound-making event).

Jones (2004) considers two aspects of attending : anticipatory attending and reactive attending. First, the listener can anticipate when the next time interval will begin : there is a shift of attending in order to coincide with expected sound onsets. Such anticipations are called temporal expectancies. Perception of temporal regularity is based on temporal expectancies that adapt in response to temporally fluctuating inputs. The second aspect involves expectancy violations. Any deviation observed by the listener from an expected timing is an expectancy violation. This expectancy violation could lead to a potential surprise. Reactive attending involves rapid attentional shifts associated with unexpected sound onsets. The relationship between internal oscillations and external rhythm is the basis of listeners' responses to temporal fluctuations as deviations from categorical expectations. With a rhythmically irregular event, temporal expectancies are imprecise, and lead to a poor attentional synchrony. Snyder and Large (2005) provide evidence for early auditory processing underlying metrical expectancy. This is consistent with the notion of an attentional process that primes perceptual responses for temporally predictable stimulus contexts (Large and Jones, 1999).

The general process by which any attentional oscillator becomes locked into a stimulus periodicity is called "attunement". It means that the internal oscillator adapts its own period to the external one. This external driving rhythm is identified as a referent level. Whereas a referent period refers to subjective tempo, a referent level refers to a salient stimulus time level. Obviously, we do not attend solely to elements occurring at a referent level (Jones and Boltz, 1989). Listeners spontaneously focus on events occurring at an intermediate rate (the referent level), and they then may shift attention to events occurring over longer (future attending) or shorter (analytic attending) periodicities. For instance,

we may focus attention on low-level details, or upon higher-level characteristics such as a whole movement in a complex auditory sequence.

Error and outlier patterns are consistent with Jones' suggestion that a temporally perturbed event may be more salient than temporally expected ones in a very regular context. More deviants occurred in response to a regular sound. For a listener attending to a rhythmic sound pattern, a temporally unexpected sound is considered as an expectancy violation. This "surprise" can contribute to a tuning of a listener's attention : the listener is engrossed in reactive attending.

However, responses to both signals would have some elements of reactive attending because there is little opportunity to anticipate the onset of either sound. They were presented as isolated 540 ms patterns at unexpected times throughout the session with equal probability. On some trials, participants heard an irregular burst and then on others a regular one. Therefore, the possibility that irregular signals produced somewhat faster RT is probably due to the fact that its rhythm had deceptive spacing and captured attention more than the regular one. This could exaggerate differences between the two because the irregular would seem more irregular given the reference of the regular burst.

Another finding of the present study that requires comment in light of the literature on perceived urgency is the finding of a nonsignificant effect of the temporal irregularity on perceived urgency. Edworthy et al. (1991) suggested that there may be a weak relationship between a syncopated rhythm and perceived urgency. With a nonmetric rhythm, we did not observe any difference on perceived urgency between the two sounds. This kind of sound construction is quite unusual in a "real-alarm world". Guillaume et al. (2003) argued that some learned adequacy with common warning signals leads to a particular and well defined urgency level. The influence of our cultural environment is thus highlighted. Therefore, to the contrary, it may be possible that participants could not make urgency judgments on these uncommon sounds. This observation confirms the fact that parameters other than physical properties may interact in the judgment of urgency. Because of these other factors, another reliable and objective urgency measurement is needed. The RT procedure provides interesting insights. We know in which cases participants react faster to temporal irregularity.

6.5 General Discussion

Until now, the most dominant approach in the urgency studies (e.g. Edworthy et al. 1991; Hellier et al. 1993) has been to consider only the perceived urgency of alarm sounds related to their acoustical properties such as rate, pitch or timbre. Showing that the ranking of alarms on an urgency scale could be modified by "high-level" parameters, Guillaume et al. (2003) extended these studies by adding an "activation of a mental representation" stage. Even though, the final purpose was still the estimation of a *judgment* of the degree of urgency.

However, whereas such approaches have provided powerful experimental results and have been applied in several industrial or military environments, we claimed that the urgency issue should be considered in a more *objective* way. First problems were described in the Introduction. The main results of these papers were obtained in various experiments using psychoacoustic paradigms (e.g. dissimilarity judgments or ranking) designed to study a cognitive and somewhat global factor such as perceived urgency. Perceived urgency judgments could be associated to cognitive factors (Guillaume et al. 2003). Judgments are indeed strongly modified by some learning effects, cultural influences, participants strategies and the set of stimuli under study, for instance. We are more interested here in studying the "real" urgency conveyed by warning sounds rather than urgency plus "peripheral" factors (such as learning, cultural influences, ...). We thus reshaped the urgency paradigm to fit a speeded detection task in order to measure reaction times and errors rather than only "perceived urgency" per se. RT is closely correlated with urgency, decreasing monotonically as urgency increases. So RT may in a sense serve as a surrogate measure of urgency. The current study hence partially replicates and extends previous findings on urgency.

We cannot assume, however, that the processes underlying changes in urgency in paradigms of magnitude judgment or paired comparison are the same ones as those underlying changes in RT. Several models have been proposed to account for auditory processing (McAdams and Bigand, 1994). This theoretical framework could help us to comprehend why the several methodological differences between the present study and classical urgency experiments could lead to an understanding of different processes involved in urgency studies. When an incoming sound reaches the auditory system, the first stages of processing occur in the periphery and cochlea. The following stages model more central mechanisms, and could be studied as auditory grouping processes (Bregman 1990), as well as temporal integration, for example. Then, it is assumed that a stage involving extraction and compu-

tation of perceptual properties follows. The representation of the signal is then compared with existing knowledge structures in order to be interpreted. The final stages of event structure processing and mental representation of sound end the whole process.

Previous urgency studies were centered essentially on the perceived relations among sound events that can affect the urgency judgment, and on the mental representation of warning sounds that are experienced by the listener. The adopted point of view is that the distinction between the observed phenomenon and the process that accounts for the phenomenon is important.

The present study first focused on more "low-level" stages of the urgency process (Experiments 1 and 2). The observed behavioral performance observed apparently reflects processes occurring before the computation of perceptual properties or the mental representation stage. Experiment 3 completed this global study showing that attentional processes may be affected by abstract knowledge structures as well as perceived regularities in the stimulus structure. The current findings thus reflect both stimulus-driven and expectancy-based representations of alarm sounds.

Results of Experiment 1 and 2 reflects a clear and consistent effect of IOI on RT.

It has been proposed that the two necessary components in the RT response are signal detection and response initiation (Green and Luce, 71). Thus, the overall RT measures contain a motor component which somewhat hides the relation between the stimulus parameter under study here (IOI) and the detection ("sensory") component (Kohfeld, 1971). However, the motor-execution stage has been proved to be relatively invariant across changes in perceptual and cognitive task requirements (Miller and Low, 2001). We can thus assume that RT differences observed during the three experiments presented here appeared to be attributable to differences in the durations of sensory (or cognitive) processes rather than to differences in motor time. Temporal "integration" of some sort seems fundamental to models of detection and RT.

Experiment 1 stimuli varied (weakly) along the loudness dimension. Although some investigators have suggested that RT does not assess loudness (Kohfeld et al. 1981), the majority of studies indicated that RT may be used as an indirect measure of loudness for tones or noises (Chocholle, 1940; Pfingst et al., 1975; Wagner et al., 2004). Because of this, we concluded that this IOI effect observed in Experiment 1 was partially or perhaps entirely due to the loudness difference between the two sounds.

In Experiment 2, we asked whether the IOI effect would be observed in a design in which stimuli were equalized in loudness. Results of Experiment 2 still support the idea of

a clear effect of IOI on RT, even though small RT differences between Experiments 1 and 2 can be considered as resulting specifically from a loudness effect. The "multiple-look" model proposed by Viemeister and Wakefield (1991) to account for time constants of the auditory system seems to explain our data. "Samples" at the output of initial processing stages in the auditory system are used and selected in order to maximize the goal in that task. Stimuli with lower IOI had faster RT. If we assume that an improvement in performance of the auditory system leads to faster RT, results can then be interpreted within this framework.

In Experiment 3, we found that, in some cases, RTs tended to be slower with regular than irregular sequences. There are several possible explanations. First, the implied rate of a sequence may effectively modulate RT in such a way that people get into the "rhythm" and continue it such that the RT is based on some multiple of the implied beat. If so, then irregular rhythms may induce a different implied beat than the regular rhythms when the IOI is fixed. Depending on the irregularity, this could lead to a faster RT for irregular than regular rhythms. Second, people might "tune out" irregularly placed elements whereas they become locked into successive sounds via entrainment in regular sequences. Then they might initiate sooner response in irregular than regular ones. This explanation suggests that people cannot phase lock to individual elements in an irregular sequence as well as they can in a regular one. Third, rhythmic expectancies generated by hearing regular sequences within the same session might lead to expectancy violations that can capture attending - what is termed "reactive attending" (Jones, 2004). In this case, the faster RT is, the greater the temporal deviation from an expected rhythmic frame. This means that, instead of "tuning out" deviant elements, one of them captures attention. This capturing process is somewhat of an expectancy violation, and thus influenced response speed.

Another noteworthy result highlighted in the three experiments is the comparison of RT between two levels of attentional conditions. For the visual task condition, participants only needed to press the space bar as quickly as possible when they heard a sound, whereas on the psychomotor attentional condition, participants not only needed to respond to the sound as soon as possible but also to track the target and keep the pointer in its center. When participants were asked to perform simultaneously the RT and tracking tasks, their mean reaction times were slower than when they performed the RT task in the visual condition. This result is consistent with literature on dual-task paradigms (Pashler and Johnston, 1998). When a task is performed as a switch trial, there is a sizable decrement in performance (increased reaction time). This decrement is called the shift cost (for a

review, see Hsieh, 2002). Even though attentional tasks had a noticeable effect on RT, it is important to highlight the fact that no interactions between stimulus parameters and attentional conditions were observed. In other words, IOI as well as temporal irregularity had an effect on RT, and RT differences between the two sounds under study did not change markedly along an attentional condition dimension. The high workload condition just "added" an offset to the RT data independently of the acoustic parameter.

A comparison of the urgency judgments and the RT data in Experiments 1 and 3 plainly demonstrates the potential benefit of an objective measurement, even in a practical and applicative framework. The new experimental paradigm proposed in this study leads to various levels of comprehension of the urgency process. Alarms can be designed by studying the effect of parameters on the first stages of auditory system processing. Then, we claim that efficient alarms are those that are detected and processed more rapidly at the output of the auditory filters. A careful study of outliers and errors also reflects different and higher-level processes than those highlighted by pure RT results, and can thus be useful for the design of alarms in an urgency situation needing an arousal effect, for instance. Moreover, we have designed experiments intended to test our urgency hypothesis under high workload conditions. From a practical point of view, it may be very useful to note that the effect of temporal parameters was the same during high or low workload conditions.

6.6 Conclusion

We carried out three experiments in order to extend existing knowledge on urgency. We have confirmed and extended the importance of temporal factors (IOI and temporal irregularity) by the use of an objective measurement. This RT paradigm not only leads to concrete recommendations concerning the design of powerful alarms, but also provide a better understanding of the processes involved. A decrease in IOI leads to a decrease in RT probably because of sensory coding differences between the two sounds being tested. We also broaden the possible applications of temporal factors in warning sound design with the temporal irregularity parameter showing an "arousal" effect on participants. This effect should be considered in a more high-level framework than the one involved in the IOI studies.

Chapitre 7

Sons synthétiques et sons naturels

7.1 Introduction

L'étude des temps de réaction simple en réponse à des changements selon la dimension temporelle des stimuli (modification de l'intervalle inter impulsion, appelé par la suite IOI, pour *InterOnset Interval*, par souci de cohérence avec le Chapitre 6) nous a permis de conclure qu'il existait un effet de l'IOI sur nos performances dans différents contextes attentionnels, avec une diminution des temps de réaction observés pour la valeur de l'IOI la plus faible (parmi les deux valeurs utilisées, 100 ms et 300 ms). La question qui nous intéresse dans ce chapitre, suite à ces premiers résultats, est de savoir si cet effet de l'IOI est plus global et peut alors se généraliser à d'autres valeurs d'IOI. Les différences de performances observées s'expliquent-elles par un effet bien spécifique de l'IOI? Auquel cas, quelle est l'influence précise de l'IOI sur le temps de réaction? Existe-t-il un seuil? Précisons ici que nous ne parlerons dans la suite de ce chapitre que de seuil "bas", puisque le seuil "haut" est défini, de fait, par le paradigme utilisé de TR : il serait en effet hors sujet de tester des sons dont l'IOI serait supérieur aux temps de réaction moyens, puisque l'IOI ne pourrait alors pas être perçu avant que le sujet réagisse. La valeur de l'IOI de 300 ms testée dans les Expériences 1 et 2 du Chapitre 6 peut être considérée comme ce seuil haut, attendu que les temps de réaction correspondants se produisaient pour la plupart après une seule et unique impulsion ; l'effet de l'IOI en tant que tel ne pouvait alors pas être contrôlé. Enfin, si l'effet de l'IOI existe bien, il serait intéressant de voir dans quelle mesure le modèle du "multiple look" (appelé dans la suite du texte par son nom anglais, par cohérence avec le Chapitre 6) de Viemeister et Wakefield (1991) présenté au Chapitre 6 est une interprétation plausible d'un effet spécifique de l'IOI.

Outre l'importance de la dimension temporelle dans la création de sons d'urgence, nous avons soulevé en section 2.6.3 page 32 l'intérêt que pouvait présenter l'étude de sons environnementaux. La comparaison de sons abstraits variant selon l'IOI à des sons environnementaux dans un paradigme de temps de réaction simple apporterait des premières pistes de réponse à la question de la plus value potentielle des sons environnementaux. En effet, la plupart des travaux traitant des sons environnementaux s'intéressent à leur identification et leur reconnaissance (voir section 2.5 page 18). Les nouvelles données de l'Expérience présentée ici devraient permettre de compléter les résultats existants, en considérant l'étape supplémentaire de la *détection* des sons. Le distinguo entre détection et reconnaissance est fait ici dans le cadre du modèle de Donders (1969) présenté en section 4.3.2 page 69.

Comme nous l'avons remarqué en section 2.5.1 page 18, le regroupement de sons disparates au sein d'une même catégorie de "sons environnementaux" ne se justifie pas forcément. Pour cette raison, nous avons choisi de restreindre l'étude des sons environnementaux à une catégorie bien définie. Les sons d'animaux (et plus particulièrement les sons de félins) ont alors été choisis. Ce choix est arbitraire, il a été guidé essentiellement par l'intuition d'un traitement particulier de certains sons d'animaux, puisqu'un son de félin est censé véhiculer une image de danger. Précisons bien que ce dernier point n'est en aucun cas une justification précise du choix des sons d'animaux.

D'un point de vue pratique, nous avons également cherché à étendre l'ensemble des recommandations de création d'alarmes sonores efficaces. L'étude de ces deux questions issues de nos précédents travaux expérimentaux ainsi que de la littérature devraient également aboutir à de nouvelles règles de design sonore.

Le protocole expérimental utilisé dans le Chapitre 6 a été simplifié. Nous avons en effet supprimé le contexte attentionnel, qu'il soit visuel ou psychomoteur. La tâche à effectuer est alors une tâche de temps de réaction simple. Le facteur Condition Attentionnelle n'a alors plus lieu d'être. Les autres sont identiques au chapitre précédent. Cette simplification nous permet de tester davantage de stimuli. De plus, nous n'avons montré au cours du Chapitre 6 aucune interaction entre la tâche attentionnelle et le type de son ; l'ajout d'un contexte attentionnel ralentit les performances, mais ne change pas significativement l'effet spécifique de l'IOI.

Nous étudierons tout d'abord en section 7.2 l'influence de différents IOI sur le temps de réaction. La section 7.3 page 171 sera consacrée à la comparaison des sons abstraits (i.e. variant selon l'IOI) et des sons d'animaux.

7.2 Expérience 1

7.2.1 Méthode

Participants

Douze auditeurs ont participé à l'expérience. Ils étaient rémunérés pour leur participation. Le groupe était constitué de cinq femmes et sept hommes. Les auditeurs ont déclaré n'avoir aucun problème d'audition. La plupart d'entre eux ont été recrutés à partir d'une base de données de volontaires et n'avaient jamais participé à une expérience psychoacoustique.

Stimuli

Quatre sons ont été utilisés. Les quatre sons ont été construits à partir d'une même impulsion unitaire (identique à celle utilisée dans les expériences du Chapitre 6) d'une durée de 20 ms, dont 5 ms de temps de montée et de décroissance linéaires, de fréquence 1 kHz. Pour chaque stimulus, seul l'IOI (InterOnset Interval) variait. Les quatre IOI utilisés étaient de 100 ms, 50 ms, 33 ms et 25 ms (Les IOI utilisés dans les Expériences 1 et 2 du Chapitre 6 étaient de 100 et 300 ms). La durée totale de chaque stimulus était de 220 ms. La Figure 7.1 présente les quatre stimuli employés.

Au préalable, les sons ont été égalisés en sonie, lors d'une expérience d'ajustement. Cinq sujets (trois hommes, deux femmes) ont participé à cette pré-expérience. Ces sujets faisaient partie du laboratoire et étaient habitués aux expériences psychoacoustiques. Ils ont déclaré n'avoir aucun problème d'audition. Le matériel utilisé pour cette pré-expérience est le même que pour l'expérience principale. Le son de référence est celui pour lequel IOI = 100 ms, diffusé à 76.5 dB dans chaque écouteur. Il était demandé aux sujets d'ajuster les trois autres sons par rapport à ce son de référence. A chaque essai, deux sons sont diffusés consécutivement, le premier étant le son de référence et le deuxième le son dont le niveau est à ajuster. Les sujets disposent d'un curseur permettant d'ajuster le niveau sonore du deuxième son, de manière à ce qu'ils le perçoivent à la même intensité que le premier. Pour chaque essai, les sujets peuvent réécouter les sons autant qu'il leur semble nécessaire. Pour chaque son, l'atténuation moyenne est calculée et appliquée. Elle est de 4.2 dB pour le son avec IOI = 50 ms, de 2.8 dB pour IOI = 33 ms, et de 5.6 dB pour IOI = 25 ms.

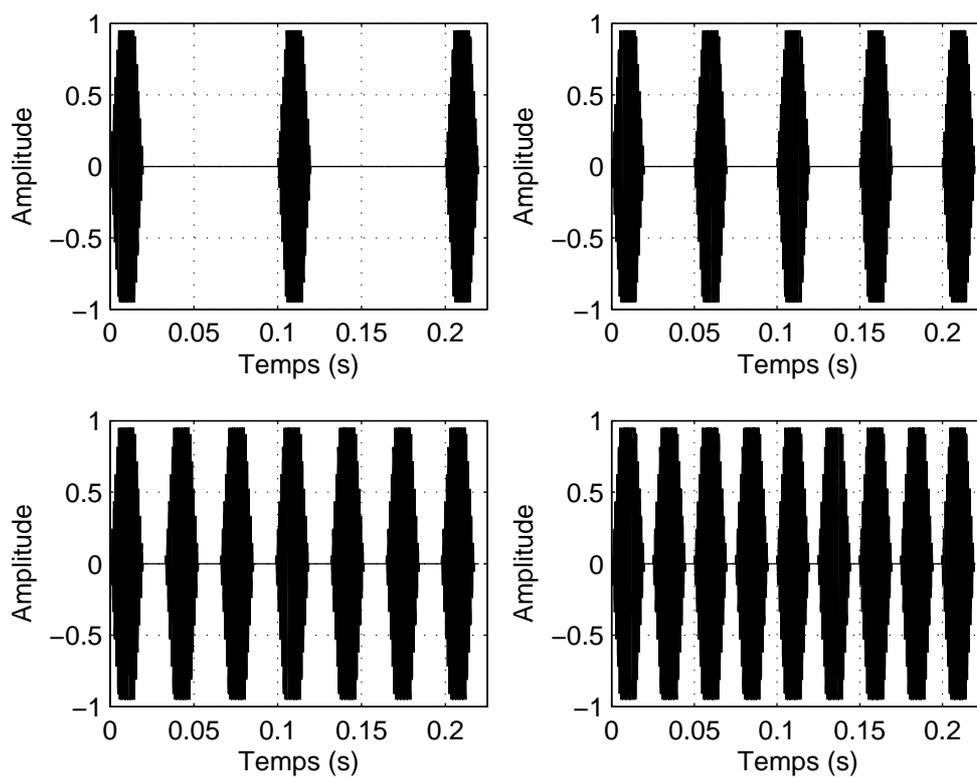


FIG. 7.1 – Formes d’onde temporelle des quatre stimuli utilisés. En haut à gauche, IOI = 100 ms ; en haut à droite, IOI = 50 ms ; en bas à gauche, IOI = 33 ms ; en bas à droite, IOI = 25 ms.

Matériel

Les stimuli ont été générés numériquement avec une résolution de 16 bits et un échantillonnage de 44.1 kHz. Les sons étaient reproduits à travers une carte RME HDSP9652, des convertisseurs analogique / numérique RME ADI-8 PRO et ont été amplifiés par un amplificateur stéréo Yamaha P2075. Ils étaient présentés aux sujets pour une écoute binaurale à travers un casque Sennheiser HD 250 linear II. Le test s'est déroulé dans la cabine audiométrique IAC de l'Ircam. La conduite de l'expérience ainsi que la collecte des données étaient assurées par un programme écrit dans l'environnement Max/MSP sur un ordinateur Apple Macintosh. Les réponses étaient enregistrées avec une précision proche de 1 ms.

Procédure

Les auditeurs étaient informés qu'ils allaient devoir appuyer sur la barre espace du clavier le plus vite possible dès qu'ils entendaient un son. Il leur était demandé de garder constamment les doigts posés sur la barre espace, afin de minimiser le temps moteur. L'expérience comportait 3 blocs de 100 sons chacun, i.e. 25 répétitions des quatre stimuli testés dans chaque bloc. Les sons étaient diffusés à des moments aléatoires (fixés entre 3 s et 8 s) et en ordre aléatoire. Chaque bloc durait approximativement 15 minutes. Aucune référence au concept d'urgence n'est faite au préalable, le but de l'étude n'est expliqué aux sujets qu'à la fin de l'expérience.

Analyses statistiques

Les analyses ont été réalisées avec le logiciel R. Pour l'analyse statistique des données, de la même manière que pour les expériences du Chapitre 6, les erreurs ont été éliminées. Aucun temps de réaction supérieur à 1s n'a été observé. Onze oublis ont été relevés, tous sujets et tous blocs confondus. Les oublis ont été remplacés par la moyenne des autres temps de réaction au sein d'un même bloc afin de ne pas avoir de données manquantes. Ceci concernait cinq sujets. Les outliers (définis de la même manière que pour les expériences du Chapitre 6, i.e. comme les temps de réaction supérieurs ou inférieurs à la moyenne des temps de réaction plus ou moins deux fois l'écart-type) ont également été éliminés et remplacés par la moyenne des TR restants au sein d'un bloc. Ceci concernait 3.5% de l'ensemble des données (les 300 mesures de TR pour les 12 sujets). Nous avons analysé les données dans une ANOVA à mesures répétées à trois facteurs intra-sujets : IOI (4

niveaux) \times Bloc (3 niveaux) \times Répétition (25 niveaux). Les résultats de l'ANOVA à mesures répétées ont été corrigés selon la méthode de Huynh-Feldt. Keppel (1982) explique qu'il est inutile d'appliquer une correction si la valeur du p non modifié rapporte déjà un effet non significatif. Par conséquent, les valeurs de p présentées ci-après sont les valeurs corrigées, avec les ϵ correspondants uniquement dans les cas où les résultats non corrigés de l'ANOVA rapportaient des résultats significatifs. Dans les autres cas, ce sont les valeurs des p non modifiés (et donc sans valeur correspondante d' ϵ). Les effets significatifs ont été explorés avec des tests post-hoc (Tukey HSD).

7.2.2 Résultats

Analyse des temps de réaction

Les temps de réaction moyens pour les 12 sujets sont présentés Figure 7.2.

L'analyse des temps de réaction pour les différents IOI a pour but de révéler de potentielles différences de temps de traitement entre les quatre valeurs de ce facteur IOI (100 ms, 50 ms, 33 ms, 25 ms). Les résultats des ANOVA comparant les temps de réaction dans les différentes conditions expérimentales sont présentées dans le tableau 7.1.

Source	df	SS	MS	F	ϵ	p
IOI	3	41328	13776	11.5	0.92	0.0002
Répétition	24	86867	3619	1.65	0.7	0.14
Bloc	2	256954	128477	6.86	1.08	0.006
Répétition \times IOI	72	96580	1341	0.98	-	0.51
Répétition \times Bloc	48	67435	1405	1.03	-	0.41
IOI \times Bloc	6	6869	1145	0.84	-	0.54
Répétition \times IOI \times Bloc	144	169888	1180	0.87	-	0.87

TAB. 7.1 – Résultats de l'ANOVA de l'Expérience 1. IOI : InterOnset Interval. SS : Somme des carrés. MS : carré moyen.

Effet de l'IOI Tout d'abord, les données font ressortir un effet significatif de l'IOI ($F = 11.5$, $p < 0.001$). La Figure 7.2 montre les temps de réaction moyen en fonction de l'IOI. Nous pouvons observer de prime abord que le temps de réaction décroît avec l'IOI. Chaque point représente la moyenne de 900 mesures de TR (75 par sujet). Pour IOI = 100 ms, le

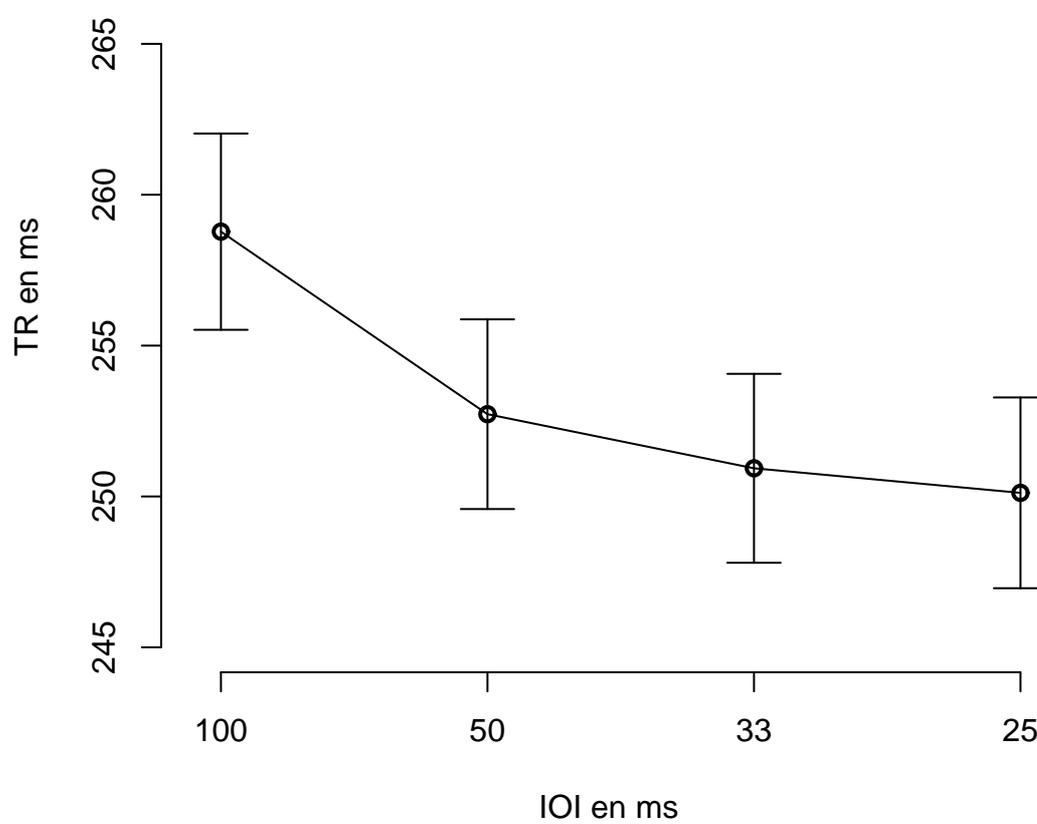


FIG. 7.2 – Expérience 1 : Temps de réaction (moyenne des 12 sujets par IOI). Les barres verticales représentent les intervalles de confiance à 95%.

TR moyen est de 259 ms avec un écart-type de 50 ms. Pour IOI = 50 ms, les valeurs de TR et d'écart-type sont respectivement de 253 ms et de 48 ms. Pour IOI = 33ms, elles sont de 251 ms et de 48 ms. Pour IOI = 25 ms, elles sont de 250 ms et de 48 ms.

Un test post-hoc (Tukey HSD) a été effectué afin d'explorer l'effet de ce facteur. L'effet du facteur IOI est dû à la fois à une différence entre la valeur IOI = 100 ms et les trois autres ($p < 0.004$), à une différence entre IOI = 25 ms et IOI = 50 ms ($p < 0.07$), et à une différence (au seuil $p < 0.1$) entre IOI = 50 ms et IOI = 33 ms. Ces différents résultats semblent mettre à jour un effet de seuil pour la valeur IOI = 33 ms. Pour des valeurs d'IOI supérieures à 33 ms, les temps de réaction décroissent lorsque l'IOI décroît ; pour des valeurs inférieures à 33 ms, nous n'observons plus d'effet significatif de l'IOI sur les temps de réaction.

Effet de la fatigue Nous observons également un effet significatif du facteur Bloc sur les temps de réaction ($F = 6.86$, $p \leq 0.006$). Les sujets réagissent en moyenne moins vite à la fin de l'expérience qu'au début. Le TR moyen pour le Bloc 1 est de 242 ms avec un écart-type de 45 ms, pour le Bloc 2, les valeurs obtenues sont de 255 ms et de 49 ms, et pour le Bloc 3 de 263 ms et de 50 ms. Cet effet est dû à des temps de réaction plus rapides dans le Bloc 1 que dans les deux autres ($p \leq 0.02$). La comparaison entre les deux derniers blocs n'est pas significative ($p \geq 0.19$). Ces résultats mettent en avant un léger effet de fatigue chez les sujets au cours de l'expérience.

Analyse des outliers

Nous avons également cherché à vérifier que les outliers étaient bien répartis uniformément en fonction du type de son, c'est-à-dire que le nombre et la valeur des temps de réaction outliers ne devrait pas dépendre des valeurs de l'IOI. En d'autres termes, les outliers ne devraient refléter que l'inattention des sujets, et peuvent de ce fait être éliminés des données que nous cherchons réellement à analyser. Il n'est pas possible d'effectuer une analyse de variance uniquement sur les temps de réaction outliers, puisqu'ils ne sont pas en nombre égal pour un type d'IOI donné et un Bloc donné. De plus, les outliers étant par définition des données marginales et extrêmes, elles comportent une grande variabilité et les hypothèses de l'ANOVA ne peuvent que difficilement s'adapter à ce type de données.

Nous avons donc analysé les temps de réaction sans supprimer au préalable les outliers. L'effet de l'IOI est significatif ($F = 14.81$, $\epsilon = 0.9$, $p < 0.0001$). Cet effet significatif est dû à des temps de réaction plus lents pour IOI = 100 ms que pour les trois autres valeurs

($p < 0.002$) ainsi qu'à des temps de réaction plus lents pour IOI = 25 ms que pour IOI = 50 ms ($p < 0.02$) et enfin, des temps de réaction plus lents pour IOI = 33 ms que pour IOI = 50 ms ($p < 0.11$). Ces résultats sont similaires à ceux trouvés pour les données sans outliers, avec cependant un effet de seuil à IOI = 33 ms un peu moins clair (effet marginalement significatif de la comparaison entre IOI = 50 ms et IOI = 33 ms).

L'effet du facteur Bloc est également significatif ($F = 5.9$, $\epsilon = 1.09$, $p \leq 0.01$). Cet effet est dû à des temps de réaction plus rapides dans le Bloc 1 que dans les deux autres ($p \leq 0.02$). Cet effet est exactement le même que précédemment pour les données sans outliers. Contrairement aux données sans outliers, l'effet du facteur Répétition apparaît aussi comme significatif ($F = 2.65$, $\epsilon = 0.6$, $p \leq 0.02$). Toutefois, la faible valeur de l' ϵ obtenue nous incite à considérer ce résultat avec précaution. En effet, un faible ϵ peut amener à considérer comme significatif un effet qui ne l'est peut-être pas (à un seuil différent, du moins). Plus les valeurs du facteur Répétition sont élevées plus les temps de réaction sont lents. En d'autres termes, cet effet de la Répétition met en avant la fatigue des sujets au sein même d'un bloc. Enfin, les différentes interactions entre facteurs ne sont pas significatives.

Au global, l'analyse des données avec outliers montre que ces outliers sont bien répartis uniformément quelle que soit la valeur de l'IOI considérée.

7.2.3 Discussion

Urgence objective

L'expérience décrite ici dans laquelle les temps de réaction sont plus rapides lorsque l'IOI décroît permet d'étendre les résultats de l'Expérience 2 du Chapitre 6 à des valeurs plus faibles d'IOI et dans un contexte non attentionnel (i.e. ni visuel ni psychomoteur). La comparaison des valeurs d'IOI choisies pour le Chapitre 6 (300 ms et 100 ms) aux temps de réaction moyens obtenus en contexte attentionnel visuel (268 ms) pouvait être interprétée non pas comme un effet spécifique de l'IOI mais comme un effet dû à la différence entre un stimulus composé de plusieurs impulsions avec un stimulus composé d'une unique impulsion. En effet, dans le cas où IOI = 300 ms, les temps de réaction moyens (donc 268 ms) apparaissaient avant la deuxième impulsion du stimulus. L'expérience décrite dans ce chapitre permet alors de conclure qu'il existe bien un effet spécifique de l'IOI sur le temps de réaction; cet effet est robuste et valable dans des contextes attentionnels différents. Cependant, bien que nous ayons ainsi pu éliminer l'interprétation du type "1 impulsion *versus* plusieurs impulsions", il n'est toutefois pas possible de distinguer le facteur IOI du

facteur nombre d'impulsions, les stimuli utilisés ayant la même durée totale. Ce point avait déjà été souligné dans le Chapitre 6.

Nous avons donc pu mettre à jour un nouveau paramètre permettant de modifier l'urgence objective d'un son selon plusieurs niveaux de manière efficace et robuste. Ainsi, il est possible d'imaginer de futures alarmes sonores véhiculant différents degrés d'urgence suivant des IOI plus ou moins grands. De plus, d'un point de vue pragmatique, il est intéressant de constater qu'il semble exister un seuil tel que $IOI = 33$ ms en deçà duquel le temps de réaction des auditeurs ne change plus de manière significative. En effet, un effet non significatif d'un paramètre ou de certaines valeurs d'un paramètre permet alors de laisser une certaine latitude de création sonore au concepteur des alarmes. Le cahier des charges d'une alarme sonore s'exprime alors en terme de paramètres (ou valeurs de paramètres) à respecter pour transmettre un niveau d'urgence voulue et, en parallèle, en terme de paramètres selon lesquels il est possible d'agir à des fins "esthétiques" qui véhiculent l'image de marque de la voiture ou du système en question (par exemple, pour deux alarmes données indiquant le même degré d'urgence, il serait possible de les différencier selon leur IOI, à condition de le choisir inférieur à 33 ms dans les deux cas).

Il est également intéressant de noter que pour les IOI de 50 ms, 33 ms et 25 ms, le rythme n'est plus perceptible ; les différentes impulsions ne sont pas séparables perceptivement. Les conclusions des travaux de Brewster et coll. (1994) d'une part et d'Edworthy et coll. (1991) et Hellier et coll. (1993) d'autre part selon lesquelles le *rythme* est une dimension perceptive primordiale dans le jugement d'urgence doivent alors être étendues aux facteurs d'ordre temporel, comme l'IOI. Nous mettons une fois de plus en avant ici l'intérêt d'utiliser une mesure objective de l'urgence confrontée à des paramètres physiques, et non perceptifs (comme la notion de rythme telle que Brewster et coll. (1994) l'emploient).

Modèle du *multiple look*

Dans le Chapitre 6 (Expérience 2), nous avons proposé d'expliquer l'effet spécifique de l'IOI par le modèle du "multiple look" de Viemeister et Wakefield (1991). Ce modèle a initialement été proposé pour résoudre le paradoxe souvent cité (de Boer, 1985) entre les modèles d'intégration temporelle (Green, 1960) et de résolution temporelle (Viemeister (1979) pour la détection de modulation ; Moore et coll. (1988) pour les aspects temporels de masquage). Viemeister et Wakefield (1991) insistent sur la distinction entre le *phénomène* d'intégration temporelle en lui-même (i.e. diminution des seuils d'audibilité lorsque la durée du signal augmente) et les *processus* qui sont à la base de cette intégration. Comme

nous l'avons expliqué dans le Chapitre 6, le modèle qui est alors proposé par Viemeister et Wakefield (1991) considère des sortes d'"échantillons" (les *looks*) avec de très petites constantes de temps (de l'ordre de quelques millisecondes) qui sont stockés en mémoire et qui peuvent être traités sélectivement en fonction de la tâche demandée. Notons toutefois que le modèle n'explique pas *comment* les différents looks sont combinés, ni ce qu'est réellement un *look*.

Ce modèle semble bien expliquer les données de l'Expérience 1 décrite dans ce chapitre, de la même manière qu'il expliquait les résultats de l'Expérience 2 du Chapitre 6. En effet, plus l'IOI est petit, plus l'étape de décision finale du modèle qui s'effectue à partir des looks sélectionnés se fait rapidement. Les looks étant basés sur des constantes de temps très courtes (5 ms), si l'intervalle de séparation entre deux impulsions est supérieur à cette constante de temps, les looks de chaque impulsion sont indépendants et l'information issue de ces looks peut alors être combinée de manière optimale. Le seuil que nous avons observé pour IOI = 33 ms en deçà duquel les temps de réaction moyen ne sont plus significativement différents peut alors être expliqué sous cet angle. Nous pouvons faire l'hypothèse que pour un IOI inférieur à 33 ms, les looks de chaque impulsion ne sont plus indépendants, et les performances, de ce fait, ne s'améliorent plus (ce qui s'exprime donc comme des temps de réaction moyens entre IOI = 33 ms et IOI = 25 ms quasi égaux).

Dans le cadre de l'Expérience 1 décrite dans ce chapitre, le modèle de fenêtre temporelle (Moore et coll., 1996; Moore et coll., 1988; Plack et Moore, 1990) permet également d'expliquer nos données. Notons que le modèle de multiple look n'est pas incompatible avec celui de fenêtre temporelle. Il permet simplement de l'étendre dans certains cas, et, de manière plus générale, d'expliquer deux types de données qui étaient initialement opposées (intégration et résolution temporelles). L'idée de base du modèle de fenêtre temporelle est de considérer que les représentations des stimuli dans le système auditif périphérique sont "lissées" et moyennées par une fenêtre temporelle qui "glisse" temporellement au fur et à mesure. Cette fenêtre temporelle est donc une fonction de pondération qui s'applique à l'énergie du signal. Moore et coll. (1996) ont alors proposé une mesure appelée ERD (pour *Equivalent Rectangular Duration*) pour caractériser la longueur effective de cette fenêtre temporelle. Ils ont montré que, pour une large gamme de fréquences et de niveaux sonores, l'ERD optimum était aux alentours de 7 ms. Cet ordre de grandeur nous permet alors d'expliquer le seuil observé pour IOI = 33 ms (l'impulsion utilisée dans nos stimuli ayant une durée de 20 ms, l'intervalle de silence entre deux impulsions est de 13 ms pour IOI = 33 ms et de 5 ms pour IOI = 25 ms).

Valeur seuil de 30 ms

Il est également intéressant de constater que le seuil observé dans nos données pour un IOI de 33 ms correspond à l'apparition du percept de hauteur, estimé aux alentours de 30 Hz (Krumbholz et coll., 2000 ; Pressnitzer et coll., 2001). Cette limite inférieure du percept de hauteur a été définie par ces auteurs comme le taux de répétition le plus faible qui évoque la sensation de hauteur, pour des sons complexes harmoniques. Ils ont pu montrer expérimentalement que cette limite inférieure était due à des mécanismes d'ordre temporel (et non à une limite de la résolution spectrale des harmoniques). Plus précisément, les auteurs expliquent que ces mécanismes analysent les intervalles de temps entre les pics dans le pattern d'activité neuronale, avec une limitation de ces intervalles aux alentours de 30 ms.

Wang et coll. (2003) ont proposé de faire le lien entre ces données psychophysiques et certains résultats d'une série d'études électrophysiologiques - enregistrements unitaires chez le singe éveillé - traitant des représentations de modulations temporelles dans le cortex auditif (Lu et Wang, 2000 ; Lu et coll., 2001 ; Liang et coll., 2002). Deux types de codage ont ainsi été mis en évidence. Pour des fréquences de modulation inférieures à 30 Hz, un codage temporel *explicite* est effectué, c'est-à-dire que les décharges neuronales sont synchrones avec le stimulus considéré. En revanche, pour des fréquences de modulation supérieures à 30 Hz, le codage est *implicite*, et s'effectue par le biais d'un taux de décharge neuronale. En d'autres termes, ces études suggèrent que le traitement des stimuli auditifs est basé sur une fenêtre d'intégration temporelle de l'ordre de 30 ms.

Bien que l'Expérience 1 décrite dans ce chapitre diffère des études présentées brièvement dans ce paragraphe, aussi bien en termes de stimuli utilisés que de méthodes et même d'hypothèses expérimentales, il n'est sans doute pas anecdotique que les constantes de temps observées soient les mêmes (30 ms), et reflètent ainsi deux types de traitement des stimuli auditifs (codage implicite ou explicite des modulations temporelles).

7.3 Expérience 2

7.3.1 Méthode

Participants

Douze auditeurs ont participé à l'expérience. Ils étaient rémunérés pour leur participation. Le groupe était constitué de huit femmes et quatre hommes. Les auditeurs ont déclaré n'avoir aucun problème d'audition. Ils étaient sélectionnés de la même façon que pour l'Expérience 1.

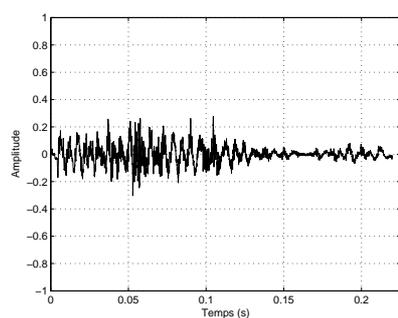
Stimuli

Huit stimuli ont été utilisés. Quatre d'entre eux étaient exactement les mêmes que ceux de l'Expérience 1, i.e. quatre stimuli d'une durée totale de 220 ms avec un IOI de 100 ms, 50 ms, 33 ms ou 25 ms. Les quatre autres stimuli utilisés sont des sons d'animaux (à 44.1 kHz, 16 bits) extraits de la base de données sonore *SoundIdeas*, et modifiés avec le logiciel Adobe Audition, de façon à présenter la même durée que les quatre stimuli variant selon l'IOI. Les quatre sons d'animaux sont des sons de félins : jaguar, lion, et deux sons de léopard différents (appelés par la suite léopard1 et léopard2). Les formes d'onde temporelle de ces sons sont représentées en Figure 7.3, et leurs sonagrammes en Figure 7.4.

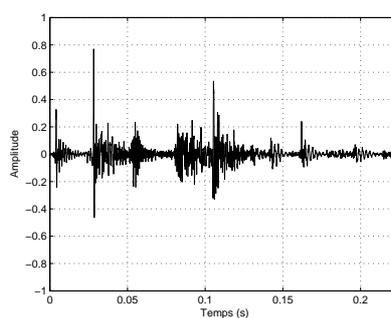
Un rapport signal sur bruit moyen a été calculé pour les sons d'animaux, grâce à l'algorithme HNR (pour *Harmonic to Noise Ratio*) du logiciel Praat (<http://fonsg3.let.uva.nl/praat>). Cette mesure du HNR a été introduite récemment comme une méthode pour quantifier et comparer certaines propriétés acoustiques des sons environnementaux (Lewis et coll., 2005 ; Murray et coll., 2006). Elle permet d'obtenir une mesure du degré de périodicité d'un signal $x(t)$, en faisant le rapport de l'énergie présente dans la partie harmonique du son sur l'énergie du signal "restant", appelé bruit, qui est donc non harmonique et irrégulier. Plus précisément, le HNR s'exprime comme :

$$HNR \text{ (en dB)} = 10 * \log_{10} \frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})}. \quad (7.0)$$

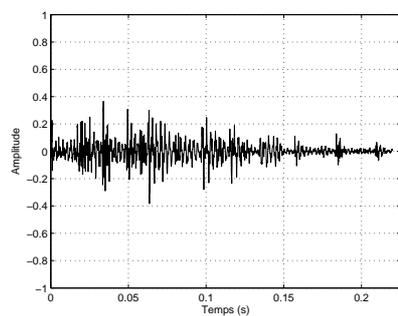
où $r'_x(\tau_{\max})$ est le maximum d'autocorrélation du signal $x(t)$, et où le retard temporel (τ) est supérieur à zéro. A titre d'exemple, un bruit blanc de 2 s donne un HNR de -2.6 dB alors qu'un sinus à 1 kHz de 2 s donne un HNR de 71.2 dB (en fixant les paramètres à leur valeur standard : pas temporel = 10 ms ; hauteur minimum = 75 Hz ; 1 période par fenêtre). En fixant le pas temporel à 10 ms, la hauteur minimum à 30 Hz et le nombre



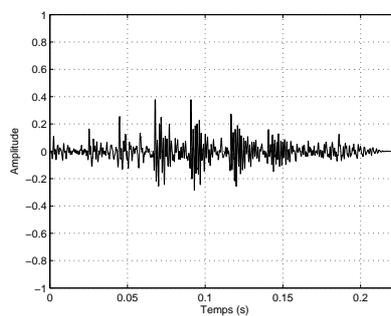
(a)



(b)



(c)



(d)

FIG. 7.3 – Formes d'onde temporelle des quatre sons d'animaux : jaguar (a), léopard1 (b), léopard2 (c), lion (d).

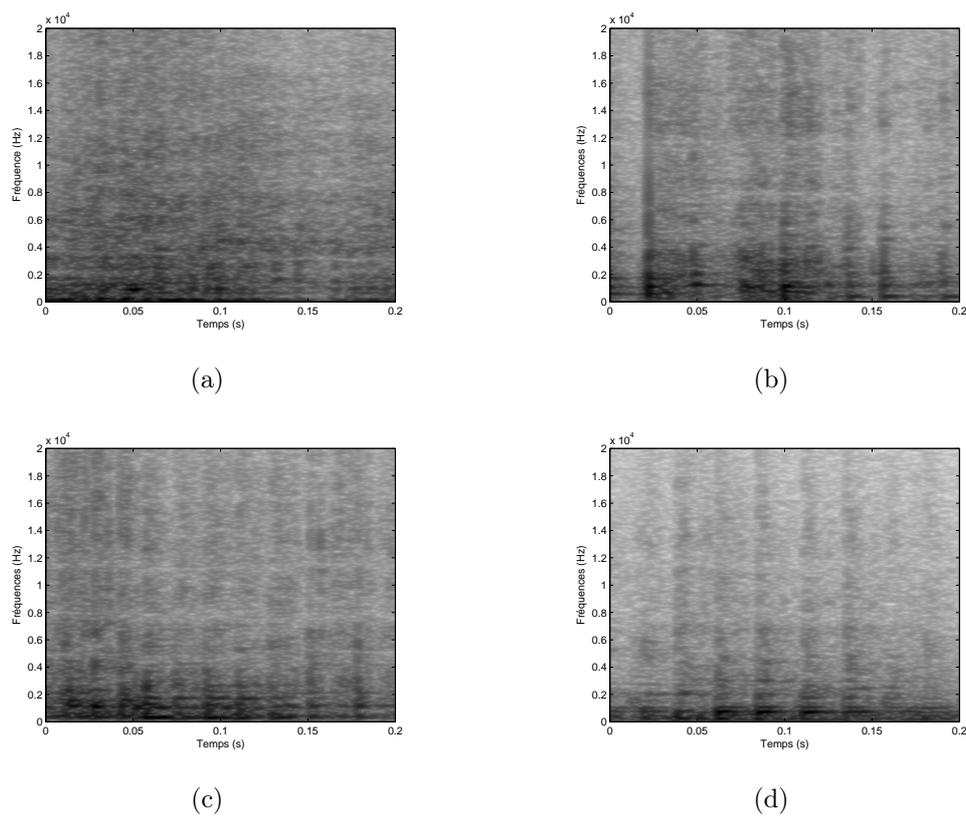


FIG. 7.4 – Spectrogrammes des quatre sons d’animaux : jaguar (a), léopard1 (b), léopard2 (c), lion (d). Le nombre de points utilisés pour la transformée de Fourier rapide est de 2048. Une fenêtre de Blackman a été utilisée, avec un taux de recouvrement défini sur 480 points.

de période par fenêtre à 1, les HNR obtenus pour les sons d'animaux sont alors de -0.21 dB pour le jaguar, de 2.91 dB pour le lion, de -1.14 dB pour le léopard1 et de 0.01 dB pour le léopard2. Ces valeurs de HNR obtenues nous permettent de constater que les sons d'animaux utilisés sont peu harmoniques. Comparativement aux autres, le son de lion est celui qui est le plus harmonique.

Nous avons également cherché à décrire les sons d'animaux comme des sons modulés en amplitude, en identifiant leur fréquence de modulation f_m . Ce descripteur est basé sur un algorithme de détection des pics dans le spectre d'amplitude, ce spectre étant calculé à partir de l'enveloppe d'énergie du signal (Peeters, 2003). Pour le son de lion, une seule fréquence de modulation a été bien identifiée, à 40 Hz. Pour le son de jaguar, aucune fréquence de modulation ne peut être extraite du signal. Pour le son de léopard1, trois fréquences de modulation ont été identifiées, à 14 Hz, 37 Hz et 52 Hz (par ordre d'importance). Pour le son de léopard2, une seule fréquence de modulation a été mise à jour, à 25 Hz.

Un test informel nous a permis de tester l'identification des sons d'animaux. Précisons tout d'abord que l'identification de l'animal avec précision (jaguar, léopard, lion) serait très peu probable (voire impossible), étant donné le peu de différence dans les représentations que nous nous faisons de sons de différents félins. Suivant les sujets, l'identification d'un son d'animal est plus ou moins claire. Notons cependant qu'étant donnée la durée très courte des sons testés, il était difficile d'obtenir une identification parfaite des sons en tant que sons d'animaux.

Au préalable, les sons ont été égalisés en sonie lors d'une expérience d'ajustement. Neuf sujets ont participé à cette pré-expérience. Ils ont déclaré n'avoir aucun problème d'audition. Le matériel utilisé est le même que pour l'Expérience 1. Le son de référence est celui pour lequel IOI = 100 ms, diffusé à 76.5 dB dans chaque écouteur. La procédure était la même que pour l'expérience d'égalisation en sonie de l'Expérience 1. Ainsi, l'atténuation moyenne de chaque son est obtenue : elle est de 2.7 dB pour IOI = 33 ms, de 4 dB pour IOI = 50 ms, de 5.2 dB pour IOI = 25 ms, de 6.4 dB pour le son de lion, de 8.0 dB pour le son de léopard1, de 12.6 dB pour le son de léopard2, de 12.1 dB pour le son de jaguar.

Matériel

Le matériel utilisé est le même que pour l'Expérience 1.

Procédure

La tâche demandée aux sujets est la même que lors de l'Expérience 1 : appuyer sur la barre espace du clavier le plus vite possible dès qu'ils entendent un son. L'expérience comportait six blocs de 96 sons chacun ; 12 répétitions des huit stimuli testés dans chaque bloc. Les sons étaient diffusés à des moments aléatoires (fixés entre 1 s et 7 s) et en ordre aléatoire. Chaque bloc durait approximativement 8 minutes. Aucune référence au concept d'urgence ni à la nature des sons n'est faite au préalable, le but de l'étude ainsi qu'une description des sons utilisés sont expliqués à la fin de l'expérience.

Analyses statistiques

Les analyses statistiques ont été réalisées avec le logiciel R. De la même manière que précédemment, les erreurs ont été éliminées. Un seul temps de réaction supérieur à 1 s a été observé. Quatre oublis ont été relevés, tous sujets et tous blocs confondus. De même que dans l'Expérience 1, les oublis ont été remplacés par la moyenne des autres temps de réaction au sein d'un même bloc afin de ne pas avoir de données manquantes. Ceci concernait trois sujets. Les outliers ont également été éliminés et remplacés par la moyenne des TR restants au sein d'un bloc. Ceci concernait 2.1% de l'ensemble des données (les 576 mesures de TR pour les 12 sujets). Nous avons analysé les données dans une ANOVA à mesures répétées à trois facteurs intra-sujets : Son (8 niveaux) \times Bloc (6 niveaux) \times Répétition (12 niveaux). Les résultats de l'ANOVA à mesures répétées ont été corrigés selon la méthode de Huynh-Feldt. Pour les mêmes raisons que dans l'Expérience 1, les valeurs de p rapportées sont les valeurs corrigées, avec les ϵ correspondants dans les cas où les résultats non corrigés de l'ANOVA rapportaient des résultats significatifs. Dans les autres cas, ce sont les valeurs des p non modifiés (et donc sans valeur correspondante d' ϵ). Les effets significatifs ont été explorés avec des tests post-hoc (Tukey HSD).

7.3.2 Résultats

Analyse des temps de réaction

Les temps de réaction moyens pour les 12 sujets sont présentés Figure 7.5.

L'analyse des temps de réaction pour les différents sons de l'Expérience 2 a pour but de confirmer d'une part les résultats de l'Expérience 1 sur les différences de traitement des quatre stimuli variant suivant l'IOI et d'autre part de révéler de potentielles différences de

temps de traitement entre les deux types de sons considérés (abstraites et animaux) ou entre les différents sons d'animaux. Les résultats des ANOVA comparant les temps de réaction dans les différentes conditions expérimentales sont présentés dans le tableau 7.2.

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	ϵ	<i>p</i>
Son	7	366586	52369	19.59	0.8	<0.0000001
Répétition	11	43707	3973	1.5	-	0.14
Bloc	5	207217	41443	0.75	-	0.59
Répétition \times Son	77	128184	1665	1.11	-	0.23
Répétition \times Bloc	55	176002	3200	2.13	0.07	0.08
Son \times Bloc	35	28877	825	0.55	-	0.98
Répétition \times Son \times Bloc	385	494628	1285	0.86	-	0.97

TAB. 7.2 – Résultats de l'ANOVA de l'Expérience 2. *SS* : Somme des carrés. *MS* : carré moyen.

Effet du type de son sur les temps de réaction moyens Le facteur Son a clairement une influence sur les temps de réaction moyens ($F = 19.59$, $p < 0.0001$). La Figure 7.5 montre les temps de réaction moyens en fonction du type de son considéré. A première vue, nous pouvons observer d'une part que, pour les sons abstraits, de manière similaire à l'Expérience 1, le temps de réaction décroît avec l'IOI, et d'autre part, que le temps de réaction est en moyenne plus court pour les sons d'animaux que pour les sons variant selon l'IOI. Chaque point de la figure représente la moyenne de 864 mesures de TR (72 par sujet). Pour IOI = 100 ms, le TR moyen est de 258 ms avec un écart-type de 56 ms. Pour IOI = 50 ms, les valeurs de TR et d'écart-type sont respectivement de 256 ms et de 58 ms. Pour IOI = 33 ms, elles sont de 252 ms et de 54 ms. Pour IOI = 25 ms, elles sont de 250 ms et de 53 ms. Pour le son de lion, elles sont de 246 ms et de 53 ms ; pour le son de léopard1 de 241 ms et de 54 ms ; pour le son de léopard2 de 239 ms et de 52 ms ; pour le son de jaguar de 238 ms et de 55 ms.

Un test post hoc (Tukey HSD) a été effectué afin d'explorer plus précisément l'effet de ce facteur. Il révèle que l'effet de l'IOI est dû à une différence entre les sons d'animaux et les sons "abstraites" ($p < 0.0001$), à une différence entre le son de lion et les trois autres sons d'animaux ($p \leq 0.007$), ainsi qu'au même effet de seuil à IOI = 33 ms que dans l'Expérience précédente ($p \leq 0.1$). Enfin, nous remarquons que les temps de réaction les plus longs pour

les sons d'animaux (le son de lion) sont tout de même plus courts que les temps de réaction les plus courts pour les sons abstraits (IOI = 25 ms) ($p < 0.03$).

Interaction Répétition × Bloc La valeur du p pour l'interaction Répétition × Bloc est à considérer avec précaution, étant donné la très faible valeur de ϵ . Quoiqu'il en soit, l'interaction Répétition × Bloc est un effet marginal.

Analyse des outliers

De la même manière que dans l'Expérience 1, nous avons cherché à savoir si les outliers étaient bien répartis uniformément en fonction du type de son. L'ANOVA effectuée sur les données avec outliers montre que le facteur Son est toujours significatif ($F = 26$; $\epsilon = 0.6$; $p < 0.00001$). Cet effet est dû à des temps de réaction plus longs pour les sons abstraits que pour les sons d'animaux ($p \leq 0.00002$), à un effet de seuil à IOI = 33 ms similaire aux données sans outliers ($p \leq 0.09$), ainsi qu'à des temps de réaction plus longs pour le son de lion que pour les trois autres ($p \leq 0.03$). Nous remarquons également, de manière similaire aux données sans outliers, que les temps de réaction au son d'IOI = 25 ms sont plus longs que pour le son de lion ($p \leq 0.04$). Le facteur Répétition apparaît également comme significatif ($F = 2.8$; $\epsilon = 0.6$; $p \leq 0.04$), contrairement aux données sans outliers. Comme pour l'Expérience 1, plus les valeurs du facteur Répétition sont élevées plus les temps de réaction sont lents. En d'autres termes, cet effet de la Répétition met en avant la fatigue des sujets au sein même d'un bloc.

Au global, l'analyse des données avec outliers montre que ces outliers sont bien répartis uniformément quelle que soit la valeur de l'IOI considérée.

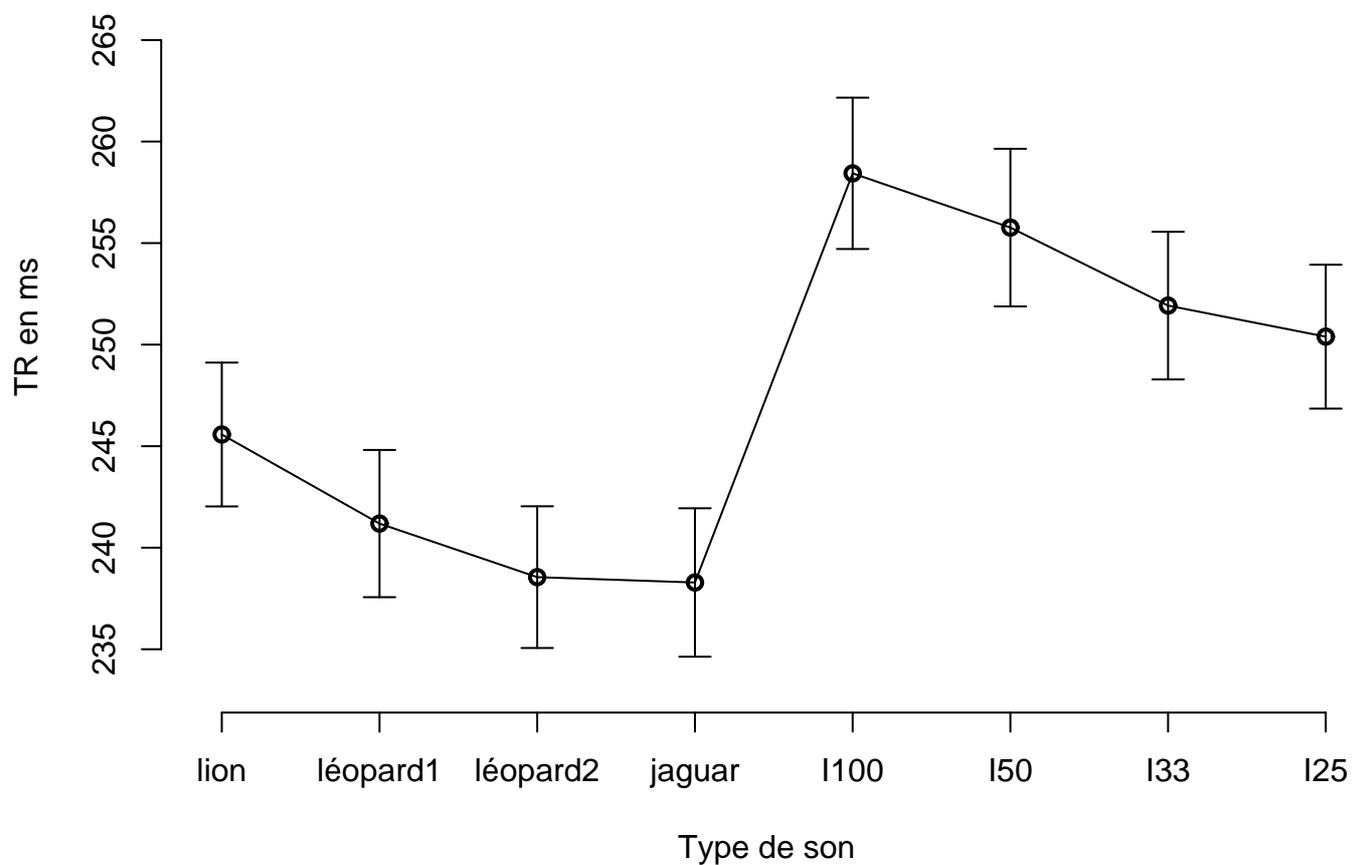


FIG. 7.5 – Expérience 2 : Temps de réaction (moyenne des 12 sujets par son). La notation I100 correspond au stimulus dont l'IOI est de 100 ms ; de même pour I50, I33 et I25. Les barres verticales représentent les intervalles de confiance à 95%.

7.3.3 Discussion

Confirmation d'un seuil à 30 ms

Les résultats décrits dans l'Expérience 1 de ce chapitre nous ont permis de mettre en évidence un effet spécifique de l'IOI : les temps de réaction moyens sont plus courts lorsque l'IOI diminue. Nous avons également observé un seuil pour une valeur de l'IOI de 33 ms, en deçà de laquelle les performances ne s'améliorent plus. Les résultats peuvent s'expliquer par le modèle du "multiple look" de Viemeister et Wakefield (1991) ou par le modèle de fenêtre temporelle de Moore et coll. (1988). Ici, nous observons à nouveau cet effet spécifique de l'IOI avec un seuil à 33 ms. Ce résultat semble donc robuste. En outre, il est intéressant de constater que les valeurs des temps de réaction moyens obtenus sont les mêmes (à 1 ms près) dans les deux expériences.

Toutefois, les performances comportementales ne peuvent être expliquées uniquement par la fréquence de modulation des sons. Si seule la fréquence de modulation avait une influence sur les temps de réaction, le son de lion qui module à 40 Hz ne serait pas traité différemment du son dont l'IOI est de 25 ms. Le même effet est observé pour le son de léopard2 qui module à 25 Hz : il devrait être traité de manière similaire ou proche des sons dont l'IOI est de 50 ms ou 33 ms.

Reconnaissance ou hypothèse spectrale ?

Outre la confirmation des conclusions de l'Expérience 1 décrite dans ce chapitre, les résultats rapportés ici mettent également en avant un effet très clair et bien marqué des sons d'animaux dans leur ensemble, pour lesquels les temps de réaction obtenus sont systématiquement plus courts que les temps de réaction pour les stimuli variant selon l'IOI. Deux grands types d'explications sont possibles. En effet, les sons d'animaux diffèrent des sons "abstrait" de cette expérience selon deux aspects. Tout d'abord, les sons d'animaux sont des sons réels, et peuvent, de ce fait avoir une représentation en mémoire à long terme, contrairement aux sons variant selon l'IOI qui sont, par définition, abstraits, et ne font donc référence à aucune source sonore identifiable et nommable. Ensuite, les sons d'animaux sont des sons complexes (voir section 7.3.1 pour une représentation temporelle et spectro-temporelle des sons d'animaux), contrairement aux sons de différents IOI dont l'impulsion unitaire est un son pur à 1 kHz.

Un avantage des sons d'animaux Considérons en premier lieu l'hypothèse selon laquelle c'est l'identification des sons d'animaux en tant que tels qui induit un temps de traitement plus rapide. En effet, un des buts de notre système auditif est d'extraire des informations des stimuli pour aboutir à une certaine connaissance relative à une source sonore, dans un contexte donné (voir McAdams et Bigand, 1994). Citons à titre d'exemple quelques études qui ont traité cette question sous l'angle de l'identification de propriétés de matériaux par des auditeurs (Giordano et McAdams, 2006) ou de l'identification de sons environnementaux (Ballas, 1993; Gygi et coll., 2004) - pour plus de détails, voir Chapitre 2, sections 2.4.2, 2.6.3 et 2.5.4.

Bien que l'objectif de l'Expérience 2 n'ait pas été formulé en termes de temps d'identification, cette question ne peut être éludée de l'interprétation des résultats obtenus. D'après le modèle de Donders (1969), chaque étape de traitement (détection, identification, etc.) s'effectue de manière sérielle, et chacune de ces étapes peut être déterminée par un paradigme de temps de réaction différent : simple, go / no-go ou avec choix (pour plus de détails, voir Chapitre 4, section 4.3.2 page 69). Cependant, il est également possible que certaines de ces étapes s'effectuent en parallèle. Nous pouvons considérer la possibilité d'une identification très rapide de certains sons, qui se traduirait alors par une différence de performance, même au sein d'une tâche de temps de réaction simple. S'il n'existe que peu (voire pas) de données psychophysiques permettant d'expliquer précisément cette différence de performance observée entre les deux types de sons de notre expérience, il existe cependant diverses études de neuroscience et d'imagerie cérébrale qui peuvent apporter des premières pistes d'explication pour nos données.

En premier lieu, il est intéressant de remarquer que parmi les sons qui sont habituellement classés sous une même dénomination de "sons environnementaux", il existe des différences notables, en terme de localisation tout d'abord, et qu'il n'est peut-être pas justifié de considérer un ensemble vaste et disparate de sons environnementaux tel que proposé par Ballas et Howard (1987), par exemple. Cette question de la définition précise des sons environnementaux et de la légitimité de regrouper dans un même ensemble certains types de sons a été soulevée au cours du Chapitre 2 (section 2.5.1 page 18). En comparant des sons d'animaux à des sons d'objets manufacturés, Lewis et coll. (2005) ont mis à jour différentes régions cérébrales impliquées dans la différenciation de ces deux catégories de sons.

Outre ces données qui permettent de localiser des zones du cerveau plus ou moins activées selon les sons environnementaux entendus, il est également intéressant de savoir

quand la reconnaissance a lieu, à quel moment les différentes régions cérébrales deviennent actives, et d'identifier ainsi le temps nécessaire au traitement de ces différents sons pour mieux comprendre, *in fine*, le type de mécanismes neuronaux impliqués. En confrontant les sons du vivant (sons d'animaux par exemple) à des sons d'objets, Murray et coll. (2006) ont alors pu montrer l'existence d'une latence plus précoce de l'activité électroencéphalographique dans le cas des sons d'objets. Les auteurs expliquent ce pic d'activité en comparant les représentations mentales des deux types de sons : nous utilisons et manipulons des objets manufacturés au quotidien, nous aurions donc de ce fait une représentation abstraite plus précise des sons associés à ces objets, ce qui mène alors à une discrimination plus facile de ces sons. Malgré les précautions nécessaires à prendre en mettant en parallèle les approches comportementales (comme celle de chronométrie mentale utilisée dans ce chapitre) et l'étude des fondements cérébraux de la perception (comme l'étude de Murray et coll., 2006), nous pouvons simplement retenir de cette étude que, suivant la représentation abstraite que nous avons des sons et selon la force du degré d'association de ce son à un objet ou une source sonore, le temps de traitement de l'information sonore varie, avec un avantage pour les sources connues et bien identifiées. Il est alors possible que les sons d'animaux utilisés ici ayant, par définition, une représentation plus claire que les sons variant selon l'IOI, une différence de temps de traitement entre ces deux types de sons à un niveau cérébral ait été reflétée par une différence au niveau comportemental.

Une dernière hypothèse pertinente d'un point de vue écologique a été proposée par Kaiser et coll. (2002). En comparant des sons d'animaux (abolements de chiens) à des sons de voix synthétiques (voyelles) ainsi qu'à des sons de chiens modifiés (i.e. créés à partir des sons réels de chiens mais non reconnaissables, car modifiés à l'aide d'un algorithme du logiciel CoolEdit), les auteurs ont pu mettre en avant une activité précoce des enregistrements magnéto-encéphalographiques pour les sons de chiens. Ils suggèrent alors qu'il est possible qu'un traitement parallèle des informations ayant trait à la peur existe ; ce traitement serait sous-cortical. Cette hypothèse d'un traitement sous cortical des signaux auditifs liés à la peur a également été montré (entre autres) par Armony et coll. (1997). Il serait alors possible que les sons de félins de notre expérience aient bénéficié d'une sorte d'"avantage biologique" et aient ainsi été traités plus rapidement. Cette dernière explication mériterait d'être approfondie par une autre étude, étant donné que les sons de l'expérience présentée ici ne provoquent pas réellement une impression de peur.

Un avantage des sons complexes Toutefois, malgré l'attrait de ces explications privilégiant une compétence perceptive complexe telle que la reconnaissance, la différence des contenus spectraux des deux types de sons (IOI et animaux) ne peut être négligée dans l'interprétation des résultats. En effet, les sons d'animaux, dont le contenu spectral est de fait beaucoup plus riche que les sons variant selon l'IOI (créés à partir d'un son pur unitaire à 1 kHz), passent par davantage de canaux fréquentiels issus de la cochlée. Nous pouvons alors faire l'hypothèse que ce codage plus riche des sons d'animaux est à l'origine de l'avantage des sons d'animaux sur les sons abstraits, créés à partir d'un son pur.

Une étude récente (Lakatos et coll., 2005) tend à prouver que la vitesse de transmission sensorielle au sein du système auditif pourrait dépendre de la complexité du stimulus auditif (chez le macaque). Les auteurs ont en effet montré un avantage des sons complexes (bruit large bande) sur les sons purs, en terme de rapidité de traitement. Il serait alors possible que les sons soient traités suivant différents chemins parallèles, selon leur degré de complexité. Là encore, bien que le pas entre cette étude et nos données comportementales chez l'Humain puisse être délicat, nous pouvons avancer l'hypothèse que ces différents temps de traitement au niveau des transmissions sensorielles pourraient se répercuter à des niveaux de traitement plus élevés et avoir ainsi une influence sur les performances comportementales.

Enfin, nous pouvons également remarquer que parmi les quatre sons d'animaux, le son de lion se détache des autres : les temps de réaction moyens observés pour le son de lion sont plus longs que pour les trois autres. Le fait que ce même son de lion soit justement le son le plus harmonique (celui dont le HNR est le plus élevé), n'est peut-être pas une coïncidence. Bien qu'aucune conclusion définitive ne puisse être dégagée de cette unique expérience, il serait intéressant de tester le lien entre le degré d'harmonicité d'un son et les différences potentielles de performances comportementales engendrées par ce son.

Pour finir, nous pouvons également noter que l'étude acoustique d'une grande diversité de cris d'alertes du monde animal (mammifères et oiseaux) permet de conclure que les sons d'urgence faible ont tendance à être des sons harmoniques, alors que les sons signalant une grande urgence (prédateur proche, etc.) ont tendance à être "rugueux et bruités" (Manser et coll., 2002). Dans une perspective de théorie de l'évolution, ce n'est sans doute pas un hasard que notre système auditif nous permette de réagir plus rapidement à des sons signalant une urgence élevée.

Perspectives

Afin de savoir si ce sont les facteurs cognitifs (reconnaissance de l'animal) ou les facteurs acoustiques (contenu fréquentiel) qui sont à l'origine de cet avantage des sons d'animaux, et afin d'essayer de comprendre ce qui rend ces sons spéciaux, il serait intéressant de mener une autre expérience qui permettrait de comparer les sons d'animaux réels de l'Expérience 2 décrite dans ce chapitre à des sons d'animaux modifiés, avec le même spectre à long terme que les sons réels mais une structure fine temporelle différente, par exemple, ce qui rendrait les sons d'animaux non reconnaissables.

Cette idée de créer des sons avec le contenu spectral de sons réels, mais modifiés de sorte à les rendre non reconnaissables, a déjà été utilisée, avec des sons de voix humaine (Belin et coll., 2000; Belin et coll., 2002) ou des sons d'animaux (Kaiser et coll., 2002; Lenz et coll., 2006). La méthode utilisée pour créer les sons modifiés est similaire dans les deux cas. Elle consiste à effectuer une transformée de Fourier rapide sur le signal temporel, et de mélanger au sein de chacune des fenêtres (définies pour la transformée de Fourier) les composantes d'amplitude (et également de phase, dans le cas de Belin et coll., 2002). La transformée de Fourier inverse est alors appliquée sur ce signal modifié. Ainsi, les valeurs en amplitude sont les mêmes dans les signaux réels et les signaux modifiés, l'énergie totale est préservée; mais chaque valeur d'amplitude est affectée à une fréquence différente de celle initiale, ce qui rend le son non reconnaissable.

Belin et coll. (2000; 2002) ont ainsi montré une activation plus importante dans certaines régions cérébrales en réponse à des sons de voix humaines, comparativement à des sons environnementaux. Là encore, la "pertinence écologique" de la distinction entre voix humaine et tous les autres types de sons est mise en avant, dans le sens où, en tant qu'humain, nous traiterions différemment les sons de notre propre espèce. De manière similaire, un avantage des sons d'animaux a également été mis en avant par Kaiser et coll. (2002), comme nous l'avons évoqué précédemment.

Si l'hypothèse de pertinence écologique est vérifiée, il semblerait alors logique, dans notre cas, d'étudier également des sons issus de voix humaines. Toutefois, nous pouvons noter, pour finir, qu'un argument contre cette hypothèse de codage spécifique de sons de notre propre espèce a été apporté récemment par Schnupp et coll. (2006). Plus précisément, l'hypothèse testée était celle de l'existence de neurones chez une espèce donnée répondant spécifiquement aux vocalises de sa propre espèce. En entraînant un animal à réagir aux vocalises d'un autre animal, les auteurs ont en effet montré que le codage temporel (avec une fenêtre de 30 ms) permettait de représenter les vocalises animales, alors même qu'il

s'agissait de vocalises d'une autre espèce. Mais le débat est loin d'être terminé.

7.4 Conclusion

Les performances comportementales liées à des stimuli variant selon différents IOI, de 25 à 100 ms ont été étudiées. Les résultats étendent ceux obtenus au cours du Chapitre 6, et confirment un effet spécifique de l'IOI sur le temps de réaction, avec un effet de seuil pour une valeur d'IOI de 33 ms (i.e. un intervalle entre deux impulsions de 13 ms), en deçà de laquelle il n'y a plus d'amélioration de la performance. Cet effet peut être interprété aussi bien par le modèle de fenêtre temporel de Moore et coll. (1988), dans lequel l'énergie présente dans le signal auditif est pondérée par une fenêtre d'une durée d'environ 7 ms, que par le modèle du "multiple look" de Viemeister et Wakefield (1991) dans lequel un processus de boucles multiples permet au système auditif de traiter avec une résolution temporelle fine différents stimuli. En deçà d'une valeur *de l'ordre de* 5 ms, ces boucles ne sont plus indépendantes, et les performances du système auditif ne s'améliorent plus. Le seuil observé à 30 ms a également été comparé à des résultats psychophysiques ou électrophysiologiques mettant en lumière deux types de traitement des stimuli auditifs délimité par une fenêtre de l'ordre de 30 ms (ou une fréquence de modulation de 30 Hz) : le codage se fait de manière explicite et synchrone pour des intervalles de temps inférieurs à 30 ms, et de manière implicite - par un taux de décharge neuronale - pour des intervalles supérieurs à 30 ms (Wang et coll., 2003) .

La comparaison de ces stimuli abstraits à des sons d'animaux selon le même protocole expérimental a permis de mettre en avant un avantage clair des sons d'animaux. Les performances sont améliorées pour les sons d'animaux. Cet effet peut être expliqué sous l'angle d'une hypothèse cognitive qui semblerait écologiquement pertinente : la reconnaissance précoce d'un son d'animal associé au danger influe sur le temps de traitement, et est ainsi répercutée au niveau comportemental. Il est également possible de l'expliquer en comparant simplement le contenu fréquentiel des sons abstraits - créés à partir d'un son pur à 1 kHz - à celui des sons d'animaux, sons complexes et peu harmoniques. Une étude de sons d'animaux modifiés - i.e. avec le même contenu spectral mais non reconnaissables - devrait permettre d'apporter de nouveaux arguments à l'une ou l'autre hypothèse.

Enfin, ces deux expériences nous ont permis d'étendre les préconisations pour la création de sons d'urgence efficaces. Les facteurs temporels peuvent être manipulés selon le degré d'urgence voulue : un son modulé à 30 Hz devrait produire une urgence maximale.

Cependant, l'utilisation de sons riches au niveau fréquentiel et préférentiellement non harmoniques semble bien plus efficace encore.

Chapitre 8

Conclusion

Les expériences qui ont été présentées au cours de cette thèse avaient pour but la création d'alarmes sonores. Plusieurs aspects ont été envisagés : l'influence de plusieurs paramètres acoustiques sur la fonction associée à un son a été traitée dans le cas des IHM sonores de l'automobile ; l'évaluation de l'urgence d'alarmes sonores en contexte d'attention partagée a été abordée par le biais des temps de réaction ; l'extension des résultats obtenus a été considérée au travers de sons synthétiques et de sons naturels. Pour conclure, nous allons résumer les principaux résultats obtenus, et mettre en lumière les perspectives ouvertes par cette nouvelle approche de l'urgence dans le cadre de recherches sur la perception auditive et ses applications.

8.1 Synthèse des résultats obtenus

L'étude des alarmes sonores peut être appréhendée selon deux grandes approches, que nous avons appelées *subjective* et *objective*. L'approche subjective regroupe la question des interfaces auditives (Blattner et coll., 1989 ; Brewster et coll., 1994) et celle de l'urgence perçue (Patterson, 1982 ; Edworthy et coll., 1991 ; Guillaume et coll., 2003). L'approche objective a été abordée par l'analyse de la réaction des auditeurs (Haas et Casali, 1995). Elle est surtout liée à la thématique des temps de réaction (Donders, 1868-1969 ; Luce, 1986).

L'approche subjective a été examinée au cours du Chapitre 5. Nous avons cherché à évaluer l'influence de six paramètres acoustiques sur le classement de sons complexes dans les catégories d'IHM automobiles, définies, entre autres, selon un axe d'urgence. Il a été démontré que les paramètres rythmiques sont dominants sur un jugement d'urgence. Des

trains d'impulsions rapides et répétés sont associés à un fort niveau d'urgence, à l'opposé d'une impulsion unique, associée à un son non urgent. Ce résultat confirme ceux obtenus par différents auteurs (Edworthy et coll., 1991 ; Hellier et coll., 1993 ; Brewster et coll., 1994). Il a également été mis en évidence que le timbre des sons et leurs fréquences fondamentales (ou F0) avaient une influence secondaire. L'effet de la F0 est contraire à celui prédit par la littérature : dans le cas présent, des F0 élevées sont associées à un faible niveau d'urgence. Néanmoins, le choix d'un jugement subjectif - dans le cas d'une étude de l'urgence perçue - comporte certaines limites, notamment dans l'interprétation et l'explication des résultats obtenus. Pourquoi observons-nous cet effet inverse de la F0 ? Quant aux paramètres rythmiques, il est possible que leur prédominance soit due à un apprentissage, dans un contexte culturel donné (les IHM sonores de notre environnement quotidien sont pour la plupart créées selon la dimension rythmique). Mais alors, ces paramètres ont-ils une importance sur un *jugement* d'urgence ou sur l'urgence en elle-même ? Ces résultats nous ont alors conduit à une révision de l'approche des IHM sonores, et plus particulièrement des alarmes.

Une nouvelle définition de l'urgence a alors été proposée : l'urgence *objective*. Le domaine des temps de réaction a été choisi comme cadre théorique, en précisant qu'une réaction rapide est associée à un degré d'urgence élevé. Les résultats de l'Expérience 3 du Chapitre 6 ont permis d'accentuer l'intérêt de ce nouveau paradigme de l'urgence : une différence de temps de réaction est observée entre les deux stimuli utilisés (régulier et irrégulier rythmiquement), contrairement au jugement d'urgence perçue qui est le même dans les deux cas. Les résultats des trois expériences du Chapitre 6 mettent également en avant la consistance du paradigme : le contexte attentionnel ne change pas l'effet possible d'un paramètre acoustique. Les temps de réaction sont plus longs dans le cas où l'effort attentionnel est élevé, mais l'influence des paramètres acoustiques testés reste la même dans les deux contextes attentionnels (psychomoteur et visuel).

L'influence de différents paramètres acoustiques sur les temps de réaction a pu être mise à jour grâce aux expériences menées dans les Chapitres 6 et 7. L'Expérience 1 du Chapitre 6 a montré que l'IOI (pour *InterOnset Interval*, soit l'intervalle entre deux impulsions dans un train d'impulsions) avait un effet sur le temps de réaction : un grand IOI est associé à des temps de réaction plus longs qu'un IOI plus faible. Un effet spécifique de l'IOI peut expliquer ce résultat. Toutefois, ces résultats peuvent également s'expliquer par une différence de sonie entre les deux sons utilisés (Chocholle, 1940 ; Buus et coll., 1982 ; Wagner et coll., 2004). L'Expérience 2 a permis de trancher entre ces deux hypothèses, en

égalisant au préalable les sons en sonie : l'effet de l'IOI est confirmé. Des modèles auditifs d'intégration et de résolution temporelle (Moore et coll., 1988 ; Viemeister et Wakefield, 1991) peuvent être pris en compte pour expliquer cet effet. Afin d'explorer plus en détail les conclusions de ces deux expériences, l'étude de l'IOI a été approfondie au cours de l'Expérience 1 du Chapitre 7 (sans contexte attentionnel). Précisons que les IOI utilisés pour cette nouvelle expérience varient entre 25 et 100 ms, et ne sont donc plus (pour trois d'entre eux) perçus comme des séquences rythmiques. L'effet de l'IOI est robuste : une diminution des temps de réaction est observée, à mesure que l'IOI décroît. Un seuil a été observé pour une valeur d'IOI autour de 30 ms : les performances ne s'améliorent plus pour des valeurs d'IOI en-deçà de ce seuil. Les modèles auditifs proposés précédemment (Moore et coll., 1988 ; Viemeister et Wakefield, 1991) sont toujours adéquats pour expliquer ces résultats. Ce seuil autour de 30 ms a été mis en parallèle avec l'apparition du percept de hauteur à 30 Hz (Krumbholz et coll., 2000 ; Pressnitzer et coll., 2001). Il est enfin intéressant de constater que des études électrophysiologiques chez le singe éveillé ont également suggéré que le traitement des stimuli auditifs est basé sur une fenêtre d'intégration temporelle de l'ordre de 30 ms (Lu et Wang, 2000 ; Lu et coll., 2001 ; Liang et coll., 2002 ; Wang et coll., 2003). Le nouveau paradigme de l'urgence que nous avons proposé a, de plus, permis de mettre en évidence l'influence d'un facteur cognitif : un stimulus irrégulier apparaît comme plus saillant qu'un stimulus régulier, et induit de ce fait un effet attractif qui amène les sujets à réagir plus rapidement (dans certains cas). Ce résultat peut être interprété dans le cadre théorique proposé par Jones (2004), qui expose comment des auditeurs utilisent (et répondent à) certaines structures rythmiques.

Les sons synthétiques (construits selon les différents IOI de l'Expérience 1 du Chapitre 7) ont été comparés à des sons naturels (sons d'animaux, de félins plus exactement) au cours de l'Expérience 2 du Chapitre 7. L'avantage des sons d'animaux sur les sons synthétiques est remarquable : les temps de réaction observés sont systématiquement plus courts pour les sons d'animaux. Ce résultat peut être expliqué sous deux angles bien distincts : le contenu spectral très riche des sons d'animaux active plus de canaux fréquentiels, menant alors à un traitement plus rapide ; ou la reconnaissance de l'animal en tant que tel est codée de manière précoce et accélère ainsi la réaction.

8.2 Création d'alarmes sonores

Les résultats classiques de la littérature de l'urgence (Edworthy et coll., 1991 ; Hellier et coll., 1993) ont ainsi été étayés par une nouvelle approche expérimentale. La prévalence des paramètres *rythmiques* sur l'urgence perçue, telle que présentée dans la littérature et au cours du Chapitre 5 a pu également être formulée en des termes plus généraux. En effet, les IOI utilisés ne sont pas perçus en tant que séquence rythmique, et il devient alors plus juste de parler de l'influence de paramètres *temporels* sur le temps de réaction. Outre la volonté d'employer les termes adéquats, ce distinguo a son importance en ce sens où il ouvre la voie à de nouvelles explorations en terme de création sonore (modulation d'amplitude, par exemple).

Parallèlement, il a également été démontré au cours de l'Expérience 3 du Chapitre 6 qu'un rythme irrégulier peut provoquer, dans certains cas, une réaction plus rapide. Cette constatation mériterait cependant d'être approfondie. L'étude de différentes irrégularités rythmiques sur le temps de réaction, et plus généralement d'événements imprévisibles permettrait sans doute d'apporter de nouveaux éléments. Il pourrait également être intéressant de complexifier et de caractériser en détail la tâche attentionnelle afin d'identifier plus précisément le contexte dans lequel l'effet de l'irrégularité peut être observé. Il serait alors envisageable d'utiliser des sons irréguliers pour signaler un oubli (ceinture, phares), par exemple.

Enfin, l'Expérience 2 du Chapitre 7 soulève indirectement une question sous-jacente à toute problématique appliquée : nous l'avions présentée dans l'Introduction comme un "problème inverse". Dans une étude de perception et/ou de cognition auditives, le but, *in fine*, est de comprendre certains aspects de notre perception, en analysant alors les caractéristiques physiques de la stimulation (ici, le son). Ici, l'approche est contraire. Un cahier des charges fonctionnel est défini : on attend telle réaction ou telle perception d'une IHM sonore. Reste alors à définir les composantes du son qui permettront de répondre au mieux aux attentes du constructeur (il n'est d'ailleurs pas évident que les réponses existent). Nous avons démontré tout au long de cette thèse qu'il est indispensable de choisir un paradigme adéquat à la question posée, ici celle de l'urgence. Cependant, le choix des sons testés est généralement arbitraire. Comme nous l'avions évoqué en section 2.6.3 page 32, l'apparition des icônes auditives coïncide avec cette volonté de proposer des IHM sonores selon des critères argumentés. Bien que la plupart des travaux relatifs aux sons environnementaux comporte certaines limites (voir section 2.5 page 18), méthodologiques

entre autres, l'idée initiale n'en est pas moins intéressante.

En se focalisant sur des sons d'animaux (et plus précisément des sons de félins), nous avons pu montrer que ces sons semblaient être traités différemment des sons synthétiques étudiés jusqu'alors. Quelle que soit la raison de ce traitement spécifique (hypothèse spectrale ou reconnaissance précoce), l'utilisation de ces sons naturels a en fait permis d'élargir et d'améliorer l'espace des possibles en terme de création sonore d'IHM. De nombreuses autres questions se posent alors : comment seraient traités des sons de voix humaine (pho-nèmes, par exemple) ? La spécificité de la voix humaine (Belin et coll., 2000 ; Belin et coll., 2002) pourrait-elle également être traduite par un traitement temporel plus rapide ? De même, la particularité des sons d'objets manufacturés que nous manipulons au quotidien (Murray et coll., 2006) pourrait-elle induire des réactions différentes ?

Bibliographie

- Adrien, J.-M. (1991). The missing link : modal synthesis. In *Representations of musical signals*, pages 269–298. MIT Press, Cambridge, MA, USA.
- Arieh, Y. and Marks, L. E. (2003). Recalibrating the auditory system : a speed-accuracy analysis of intensity perception. *Journal of Experimental Psychology : Human Perception and Performance*, 29 :523–536.
- Armony, J. L., Servan-Schreiber, D., Romanski, L. M., Cohen, J. D., and LeDoux, J. E. (1997). Stimulus generalization of fear responses : effects of auditory cortex lesions in a computational model and in rats. *Cerebral Cortex*, 7 :157–165.
- Bagot, J. D. (1996). *Information, sensation et perception*. Armand Colin.
- Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology : Human Perception and Performance*, 19 :250–267.
- Ballas, J. A. and Howard-Jr., J. H. (1987). Interpreting the language of environmental sounds. *Environment and Behavior*, 19 :91–114.
- Beaudoin-Lafon, M. and Conversy, S. (1996). Auditory illusions for audio feedback. In Tauber, editor, *Chi'96 Conference Companion*, pages 299–300. ACM Press, Addison-Wesley, Vancouver, Canada.
- Belin, P., Zatorre, R. J., and Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13 :17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 43 :309–312.

- Belz, S. M., Robinson, G. S., and Casali, J. G. (1999). A new class of auditory warning signals for complex systems : auditory icons. *Human Factors*, 41 :608–618.
- Benoist, D., Tourbier, S., and Tourbier, Y. (1994). *Plans d'expérience : construction et analyse*. TEC and DOC Lavoisier.
- Berglund, B., Berglund, U., and Lindvall, T. (1975). Scaling loudness, noisiness, and annoyance of aircraft noise. *Journal of the Acoustical Society of America*, 57 :930–934.
- Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. (1989). Earcons and icons : Their structure and common design principles. *Human-Computer Interaction*, 4 :11–44.
- Bliss, Gilson, and Deaton (1995). Human probability matching behaviour in response to alarms of varying reliability. In *Ergonomics vol.38 : Warnings special issue*, pages 49–88. Taylor & Francis publishing.
- Boer, E. d. (1985). Auditory time constants : A paradox ? In Michelsen, A., editor, *Time Resolution in Auditory Systems*, pages 141–158. Springer-Verlag, Berlin.
- Bregman, A. (1990). *Auditory Scene Analysis*. MIT Press, Cambridge, Massachusetts.
- Bregman, A. (1994). Foreword. In Kramer, G., editor, *Auditory display : Sonification, audification and auditory interfaces*. Addison-Wesley Longman Publishing Co., Boston, MA.
- Brewster, S. A. (1998). The design of sonically-enhanced widgets. *Interacting with Computers*, 11 :211–235.
- Brewster, S. A. (2003). Nonspeech auditory output. In J. A. Jacko, J. A. and Sears, A., editors, *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Associates.
- Brewster, S. A., Wright, P. C., and Edwards, A. D. N. (1994). A detailed investigation into the effectiveness of earcons. In Kramer, G., editor, *Auditory display : Sonification, audification and auditory interfaces*, pages 471–498. Addison-Wesley Longman Publishing Co., Boston, MA.
- Burt, J. L., Bartolome, D. S., Burdette, D. W., and Comstock, J. R. J. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. *Ergonomics*, 38 :2327–2340.

- Buus, S., Greenbaum, H., and Scharf, B. (1982). Measurements of equal loudness and reaction times. *Journal of the Acoustical Society of America*, 72 :S94. (Abstract).
- Buxton, W., Gaver, W., and Bly, S. (1994). Auditory interfaces. In *Auditory Interfaces : The Use of Non-Speech Audio at the Interface*. Unpublished Manuscript.
- Chocholle, R. (1940). Variations des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. *L'Année Psychologique*, 41 :65–124.
- Cockburn, A. and Brewster, S. (2005). Multimodal feedback for the acquisition of small targets. *Ergonomics*, 48 :1129–1150.
- Crabtree, D. A. and Antrim, L. R. (1988). Guidelines for measuring reaction time. *Perceptual and Motor Skills*, 66 :363–370.
- De Waard, D. (1996). *The measurement of drivers' mental workload*. PhD thesis, University of Groningen.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30 :412–431. Translation of : Die Schnelligkeit psychischer Prozesse, first published in 1868.
- Eddins, D. A. and Green, D. M. (1995). Temporal integration and temporal resolution. In Moore, B. C. J., editor, *Hearing*. Academic Press, fifth edition.
- Edworthy, J., Hellier, E., Aldrich, K., and Loxley, S. (2004). Designing trend-monitoring sounds for helicopters : methodological issues and an application. *Journal of Experimental Psychology : Applied*, 10 :203–218.
- Edworthy, J., Loxley, S., and Dennis, I. (1991). Improving Auditory Warning Design : Relationship between Warning Sound Parameters and Perceived Urgency. *Human Factors*, 33 :205–231.
- Essens, P. J. and Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception and Psychophysics*, 37 :1–7.
- Florentine, M., Fastl, H., and Buus, S. (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *Journal of the Acoustical Society of America*, 84 :195–203.

- Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D., editor, *The psychology of music*, pages 149–180. Academic Press, New York.
- Gallun, F. J. (2006). Auditory Attention : one term for several concepts? *Journal of the Acoustical Society of America*, 119 :3415. (Abstract).
- Gaver, W. W. (1989). The sonicfinder : An interface that uses auditory icons. *Human-Computer Interaction*, 4 :67–94.
- Gaver, W. W. (1993a). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, 5 :285–313.
- Gaver, W. W. (1993b). What do we hear in the world? An ecological approach to auditory event perception. *Ecological Psychology*, 5 :1–29.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton-Mifflin, Boston, MA, USA.
- Giordano, B. L. and McAdams, S. (2006). Material identification of real impact sounds : effects of size variation in steel, glass, wood, and plexiglass plates. *Journal of the Acoustical Society of America*, 119 :1171–1181.
- Glasberg, B. R. and Moore, B. C. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50 :331–342.
- Graham, R. (1999). Use of auditory icons as emergency warnings : evaluation within a vehicle collision avoidance application. *Ergonomics*, 42 :1233–1248.
- Green, D. M. (1960). Auditory detection of a noise signal. *Journal of the Acoustical Society of America*, 32 :121–131.
- Green, D. M. and Luce, R. D. (1971). Detection of auditory signals presented at random times : III. *Perception and Psychophysics*, 9 :257–268.
- Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review*, 75 :359–373.
- Guillaume, A., Pellieux, L., Chastres, V., and Drake, C. (2003). Judging the urgency of nonvocal auditory warning signals : perceptual and cognitive processes. *Journal of Experimental Psychology : Applied*, 9 :196–212.

- Guyot, F. (1996). *Etude de la perception sonore en terme de reconnaissance et d'appréciation qualitative : une approche par la catégorisation*. PhD thesis, Université du Maine.
- Gygi, B., Kidd, G. R., and Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America*, 115 :1252–1265.
- Haas, E. C. and Casali, J. G. (1995). Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. *Ergonomics*, 38 :2313–2326.
- Hayward, C. (1994). Listening to the earth sing. In Kramer, G., editor, *Auditory display : Sonification, audification and auditory interfaces*. Addison-Wesley Longman Publishing Co., Boston, MA.
- Hellier, E. J., Edworthy, J., and Dennis, I. (1993). Improving auditory warning design : quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors*, 35 :693–706.
- Houix, O. (2003). *Catégorisation auditive des sources sonores*. PhD thesis, Université du Maine.
- Hsieh, S. (2002). Task shifting in dual-task settings. *Perceptual and Motor Skills*, 94 :407–414.
- Irino, T. and Patterson, R. D. (1996). Temporal asymmetry in the auditory system. *Journal of the Acoustical Society of America*, 99 :2316–2331.
- Jones, M. R. (1976). Time, our lost dimension : toward a new theory of perception, attention, and memory. *Psychological Review*, 83 :323–355.
- Jones, M. R. (2004). Attention and timing. In Neuhoff, J., editor, *Ecological Psychoacoustics*, pages 49–88. Elsevier Academic Press, San Diego.
- Jones, M. R. and Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96 :459–491.
- Kaiser, J., Lutzenberger, W., Ackermann, H., and Birbaumer, N. (2002). Dynamics of gamma-band activity induced by auditory pattern changes in humans. *Cerebral Cortex*, 12 :212–221.

- Keppel, G. (1982). *Design and Analysis : A Researcher's Handbook*. Englewood Cliffs, NJ : Prentice-Hall.
- Kohfeld, D. L. (1971). Simple reaction time as a function of stimulus intensity in decibels of light and sound. *Journal of Experimental Psychology*, 88 :251–257.
- Kohfeld, D. L., Santee, J. L., and Wallace, N. D. (1981a). Loudness and reaction time : I. *Perception and Psychophysics*, 29 :535–549.
- Kohfeld, D. L., Santee, J. L., and Wallace, N. D. (1981b). Loudness and reaction time : II. Identification of detection components at different intensities and frequencies. *Perception and Psychophysics*, 29 :550–562.
- Kramer, G. (1994a). *Auditory display : Sonification, audification and auditory interfaces*. Addison-Wesley Longman Publishing Co., Boston, MA.
- Kramer, G. (1994b). Some organizing principles for representing data with sounds. In Kramer, G., editor, *Auditory display : Sonification, audification and auditory interfaces*. Addison-Wesley Longman Publishing Co., Boston, MA.
- Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *Journal of the Acoustical Society of America*, 108 :1170–1180.
- Kuwano, S. and Namba, S. (2001). Temporal change of timbre of helicopter noise. In *Proceedings of the International Conference on Noise Control Engineering, Goteborg, Sweden, Aug. 13-15, 2001*.
- Lakatos, P., Pincze, Z., Fu, K.-M. G., Javitt, D. C., Karmos, G., and Schroeder, C. E. (2005). Timing of pure tone and noise-evoked responses in macaque auditory cortex. *NeuroReport*, 16 :933–937.
- Lakatos, S., McAdams, S., and Causse, R. (1997). The representation of auditory source characteristics : simple geometric sources. *Perception and psychophysics*, 59 :1180–1190.
- Large, E. W. and Jones, M. R. (1999). The dynamics of attending : How people track time-varying events. *Psychological Review*, 106 :119–159.

- Laroche, C., Tran Quoc, H., Hétu, R., and McDuff, S. (1991). Detectsound : A computerised model for predicting the detectability of warning signals in noisy workplaces. *Applied Acoustics*, 32 :193–214.
- Leavesley, A. and Magrath, R. D. (2005). Communicating about danger : urgency alarm calling in a bird. *Animal Behaviour*, 70 :365–373.
- Lehto, M. R. and Papastavrou, J. D. (1998). A signal-detection theory based perspective on design of warnings. *Perceptual and Motor Skills*, 86 :720–722.
- Lenz, D., Schadow, J., Thaerig, S., Busch, N. A., and Herrmann, C. S. (2006). What's that sound ? Matches with auditory long-term memory induce gamma activity in human EEG. *International Journal of Psychophysiology*.
- Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., and DeYoe, E. A. (2005). Distinct cortical pathways for processing tool versus animal sounds. *The Journal of Neuroscience*, 25 :5148–5158.
- Liang, L., Lu, T., , and Wang, X. (2002). Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *Journal of Neurophysiology*, 87 :2237–2261.
- Lu, T., Liang, L., and Wang, X. (2001). Neural representation of temporally asymmetric stimuli in the auditory cortex of awake primates. *Journal of Neurophysiology*, 85 :2364–2380.
- Lu, T. and Wang, X. (2000). Temporal discharge patterns evoked by rapid sequences of wide- and narrow-band clicks in the primary auditory cortex of cat. *Journal of Neurophysiology*, 84 :236–246.
- Luce, R. D. (1986). *Response Times : Their Role in Inferring Elementary Mental Organization*. Oxford University Press, New York.
- Manser, M. B., Seyfarth, R. M., and Cheney, D. L. (2002). Suricate alarm calls signal predator class and urgency. *Trends in Cognitive Sciences*, 6 :55–57.
- Marozeau, J., de Cheveigne, A., McAdams, S., and Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, 114 :2946–2957.

- McAdams, S. and Bigand, E., editors (1994). *Penser les sons. Psychologie cognitive de l'audition*. Presses Universitaires de France.
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres : common dimensions, specificities, and latent subject classes. *Psychological Research*, 58 :177–192.
- Miller, J. O. and Low, K. (2001). Motor processes in simple, go/no-go, and choice reaction time tasks : a psychophysiological analysis. *Journal of Experimental Psychology : Human Perception and Performance*, 27 :266–289.
- Momtahan, K., Hetu, R., and Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics*, 36 :1159–1176.
- Moore, B. C., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (1988). The shape of the ear's temporal window. *Journal of the Acoustical Society of America*, 83 :1102–1116.
- Moore, B. C., Peters, R. W., and Glasberg, B. R. (1993). Detection of temporal gaps in sinusoids : effects of frequency and level. *Journal of the Acoustical Society of America*, 93 :1563–1570.
- Moore, B. C., Peters, R. W., and Glasberg, B. R. (1996). Detection of decrements and increments in sinusoids at high overall levels. *Journal of the Acoustical Society of America*, 99 :3669–3677.
- Moore, B. C., Vickers, D. A., Baer, T., and Launer, S. (1999). Factors affecting the loudness of modulated sounds. *Journal of the Acoustical Society of America*, 105 :2757–2772.
- Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. Academic Press, San Diego, USA, fifth edition.
- Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., and Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *The Journal of Neuroscience*, 26 :1293–1302.
- Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, 395 :123–124.
- Neuhoff, J. G. (2004). *Ecological psychoacoustics*. Elsevier Academic Press.
- Neuhoff, J. G., Kramer, G., and Wayand, J. (2002). Pitch and loudness interact in auditory displays : can the data get lost in the map? *Journal of Experimental Psychology : Applied*, 8 :17–25.

- Pashler, H. E. and Johnston, J. C. (1998). Attentional limitations in dual-task performance. In Pashler, H. E., editor, *Attention*, page 155–189. Psychology Press, Hove, UK.
- Patterson, R. D. (1974). Auditory filter shape. *Journal of the Acoustical Society of America*, 55 :802–809.
- Patterson, R. D. (1976). Estimating auditory filter shape for tones masked by a notched noise. *Journal of the Acoustical Society of America*, 56(S36).
- Patterson, R. D. (1982). Guidelines for auditory warning systems on civil aircraft. *Civil Aviation Authority Paper 82017*.
- Patterson, R. D. (1990). Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 327(1241) :485–492.
- Patterson, R. D., Edworthy, J., Shailer, M., Lower, M., and Wheeler, P. (1986). Alarm sounds for medical equipment in intensive care areas and operating theatres. Technical report, Institute of sound and vibration research, University of Southampton.
- Patterson, R. D. and Nimmo-Smith, I. (1980). Off-frequency listening and auditory-filter asymmetry. *Journal of the Acoustical Society of America*, 67 :229–245.
- Peeters, G. (2003). A large set of audio features for sound description (similarity and classification) in the cuidado project. Technical report, Projet CUIDADO.
- Peirce, C. S. (1931-1935). *Collected Papers of C. S. Peirce*. Harvard University Press, Cambridge.
- Penel, A. and Jones, M. R. (2005). Speeded detection of a tone embedded in a quasi-isochronous sequence : effects of a task irrelevant temporal regularity. *Music Perception*, 22 :371–388.
- Peters, R. W., Moore, B. C., and Glasberg, B. R. (1995). Effects of level and frequency on the detection of decrements and increments in sinusoids. *Journal of the Acoustical Society of America*, 97 :3791–3799.
- Pfingst, B. E., Hienz, R., and Miller, J. (1975). Reaction-time procedure for measurement of hearing. II. Threshold functions. *Journal of the Acoustical Society of America*, 57 :431–436.

- Plack, C. J. and Moore, B. C. (1990). Temporal window shape as a function of frequency and level. *Journal of the Acoustical Society of America*, 87 :2178–2187.
- Plack, C. J. and Oxenham, A. J. (1998). Basilar-membrane nonlinearity and the growth of forward masking. *Journal of the Acoustical Society of America*, 103 :1598–1608.
- Plomp, R. (1964). Rate of Decay of Auditory Sensation. *Journal of the Acoustical Society of America*, 36 :277–282.
- Pressnitzer, D. (1998). *Perception de rugosité psychoacoustique : d'un attribut élémentaire de l'audition à l'écoute musicale*. PhD thesis, Université Paris 6.
- Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (2001). The lower limit of melodic pitch. *Journal of the Acoustical Society of America*, 109 :2074–2084.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114 :510–532.
- Repp, B. H. (1987). The sound of two hands clapping : an exploratory study. *Journal of the Acoustical Society of America*, 81 :1100–1109.
- Rosch, E. (1978). Principles of categorization. In Rosch, E. and Lloyd, B. B., editors, *Cognition and Categorization*, chapter 2, pages 27–48. Lawrence Erlbaum Associates.
- Schnupp, J. W., Hall, T. M., Kokelaar, R. F., and Ahmed, B. (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *The Journal of Neuroscience*, 26 :4785–95.
- Smith, B. (1995). PsiExp : an environment for psychoacoustic experimentation using the IRCAM musical workstation. In *Society for Music Perception and Cognition Conference '95*. University of Berkeley, California.
- Snyder, J. S. and Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Brain Research Cognitive Brain Research*, 24 :117–126.
- Sorkin, R. D., Kantowitz, B. H., and Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors*, 30 :445–459.
- Stanton, N. and Edworthy, J. (1998). Auditory affordances in the intensive treatment unit. *Applied Ergonomics*, 29 :389–394.

- Stanton, N. A. and Edworthy, J. (1999). Auditory warnings and displays : an overview. In *Human Factors in Auditory Warnings*. Ashgate Publishing Ltd.
- Stecker, G. C. and Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *Journal of the Acoustical Society of America*, 107 :3358–3368.
- Suied, C., Susini, P., Misdariis, N., Langlois, S., Smith, B. K., and McAdams, S. (2005). Toward a sound design methodology : application to electronic automotive sounds. In *Proceedings of the 2005 International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005*.
- Susini, P., Gaudibert, P., Deruty, E., and Dandrel, L. (2003). Perceptive study and recommendations for sonification categories. In *Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003*, pages 292–295.
- Susini, P., McAdams, S., Winsberg, S., Perry, I., Vieillard, S., and Rodet, X. (2004). Characterizing the sound quality of air-conditioning noise. *Applied Acoustics*, 65 :763–790.
- Teghtsoonian, R., Teghtsoonian, M., and Canevet, G. (2005). Sweep-induced acceleration in loudness change and the "bias for rising intensities". *Perception and Psychophysics*, 67 :699–712.
- Vanderveer, N. J. (1979). *Ecological acoustics : human perception of environmental sounds*. PhD thesis, Cornell University.
- Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *Journal of the Acoustical Society of America*, 66 :1364–1380.
- Viemeister, N. F. and Plack, C. J. (1993). Time analysis. In Yost, W. A., Popper, A. N., and Fay, R. R., editors, *Human Psychophysics*, volume 3, pages 116–154. Springer-Verlag, New York.
- Viemeister, N. F. and Wakefield, G. H. (1991). Temporal integration and multiple looks. *Journal of the Acoustical Society of America*, 90 :858–865.
- Wagner, E., Florentine, M., Buus, S., and McCormack, J. (2004). Spectral loudness summation and simple reaction time. *Journal of the Acoustical Society of America*, 116 :1681–1686.

- Wang, X., Lu, T., and L, L. (2003). Cortical processing of temporal modulations. *Speech Communication*, 41 :107–121.
- Warren, W. H. and Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events : a case study in ecological acoustics. *Journal of Experimental Psychology : Human perception and performance*, 10 :704–712.
- Williams, S. M. (1994). Perceptual principles in sound grouping. In Kramer, G., editor, *Auditory display : Sonification, audification and auditory interfaces*. Addison-Wesley Longman Publishing Co., Boston, MA.
- Zwicker, E. and Scharf, B. (1965). A model of Loudness Summation. *Psychological Review*, 72 :3–26.
- Zwislocki, J. (1960). Theory of Temporal Auditory Summation. *Journal of the Acoustical Society of America*, 32 :1046–1060.
- Zwislocki, J. (1969). Temporal Summation of Loudness : An Analysis. *Journal of the Acoustical Society of America*, 46 :431–441.

Annexe A

Compléments statistiques au Chapitre 5

A.1 Analyses descriptives

A.1.1 Tables de contingence

Soient X et Y deux variables catégorielles. Supposons que X prenne des valeurs à I niveaux et Y à J niveaux. Il n'existe pas nécessairement un ordre dans les différentes modalités de X et de Y , c'est pourquoi X et Y seront traitées comme des variables nominales ou, autrement dit, polytomiques. Les modalités de X sont codées par les chiffres de 1 à I et les modalités de Y par 1 à J . La distribution conjointe de la variable (X, Y) est caractérisée par :

$$\pi_{ij} = P(X = i, Y = j) \quad \forall 1 \leq i \leq I \quad \forall 1 \leq j \leq J.$$

La distribution marginale de X est alors donnée par :

$$\pi_{i.} = P(X = i) = \sum_{j=1}^J \pi_{ij}$$

pour $1 \leq i \leq I$. On a une expression identique pour la loi marginale de Y .

Supposons maintenant que Y soit la variable à expliquer, et X la variable explicative, il est intéressant d'étudier la loi de Y , étant donné le niveau de X . Cette loi conditionnelle est donnée pour chaque i par :

$$\pi_{j|i} = P(Y = j | X = i) = \frac{\pi_{ij}}{\pi_{i.}} \quad \forall 1 \leq j \leq J.$$

En pratique bien entendu la distribution jointe, les distributions marginales et conditionnelles doivent être estimées à partir des réalisations des variables (X, Y) . Notons n le

nombre total d'observations et N_{ij} le nombre d'observations pour lesquelles $X = i$ et $Y = j$. Les réalisations de N_{ij} , notées par n_{ij} , peuvent être mises dans un tableau avec I lignes et J colonnes. Ce tableau est appelé un **tableau de contingence** $I \times J$, de dimension 2 ou à deux entrées. La distribution conjointe est estimée d'une façon naturelle par les fréquences :

$$p_{ij} = \frac{N_{ij}}{n}.$$

La **distribution marginale** de X s'estime par les fréquences marginales :

$$p_{i.} = \sum_{j=1}^J p_{ij} \quad \forall 1 \leq i \leq I.$$

Et il en est de même pour la loi marginale de Y . On utilisera fréquemment les notations suivantes :

$$n_{i.} = \sum_{j=1}^J n_{ij} \quad n_{.j} = \sum_{i=1}^I n_{ij}$$

qui permettent d'écrire

$$p_{i.} = \frac{n_{i.}}{n} \quad \text{et} \quad p_{.j} = \frac{n_{.j}}{n}.$$

A partir du tableau de contingence, nous pouvons estimer facilement la distribution conditionnelle de Y , étant donné $X = i$ de la manière suivante :

$$p_{j|i} = \frac{n_{ij}}{n_{i.}}.$$

Les notions et définitions présentées ci dessus se généralisent facilement au cas d'un tableau de contingence à plusieurs entrées, et serviront dans l'exposé détaillé des tests d'indépendances. En effet, de l'usage des tables de contingence découle la mise au point de critères statistiques permettant de juger de l'importance de la liaison entre deux variables.

A.1.2 Tests d'indépendance

Ici, il s'agit de tester l'indépendance de deux variables qualitatives X (J modalités) et Y (I modalités), ce qui revient à tester l'hypothèse nulle suivante :

$$H_0 : \pi_{ij} = \pi_{i.} \pi_{.j} \quad \forall 1 \leq i \leq I, 1 \leq j \leq J$$

où à tester

$$H_0 : \pi_{j|i} = \pi_{.j} \quad \forall 1 \leq i \leq I, 1 \leq j \leq J.$$

L'hypothèse d'indépendance de X et Y est donc équivalente à l'hypothèse que les distributions conditionnelles soient identiques aux distributions marginales. Dans le langage non mathématique, on dit parfois que l'on teste l'indépendance entre les lignes et les colonnes.

Test du khi-deux

La statistique de test habituellement utilisée pour tester cette H_0 est bien évidemment la statistique du khi-deux introduite en 1900 par Pearson, et définie par :

$$X_P^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}},$$

où les n_{ij} sont les différents éléments du tableau de contingence et $\hat{\mu}_{ij} = \frac{n_{i.}n_{.j}}{n}$. La valeur de $\hat{\mu}_{ij}$ est parfois considérée comme la valeur attendue de N_{ij} sous l'hypothèse d'indépendance. La loi asymptotique de X_P^2 sous H_0 et pour $n \rightarrow \infty$, est bien connue comme étant une loi du khi-deux avec $df = (I - 1)(J - 1)$ degrés de liberté. Nous rejetons donc H_0 si la valeur de X_P^2 est plus grande que $\chi_{1-\alpha}^2(df)$, le quantile $(1 - \alpha)$ d'une distribution $\chi^2(df)$.

Le test du χ^2 est un test statistique qui aide simplement à prendre une décision quant à la probabilité que les variables soient indépendantes l'une de l'autre ou non dans la population. La valeur du χ^2 n'est pas elle-même un bon indice de la taille ou de l'importance de l'association entre les traits mesurés.

Test du rapport de vraisemblance

Le second test calculé à partir des tables de contingence est celui du rapport de Vraisemblance. Sa statistique de test est :

$$G^2 = 2 \sum_{i=1}^I \sum_{j=1}^J n_{ij} (\log n_{ij} - \log \hat{\mu}_{ij}).$$

Sous l'hypothèse d'indépendance, la statistique calculée suit elle aussi une loi de χ^2 avec le même nombre de degré de liberté que précédemment.

A.1.3 Coefficient d'association entre variables ordinales

Si les variables X et Y sont ordinales, il existe des coefficients d'association particuliers. Supposons que les numéros des lignes et des colonnes correspondent à l'ordre "naturel", c'est à dire que si $i < i'$, la modalité i vaut "moins" que la modalité i' . Le coefficient

introduit dans ce paragraphe mesure si de "grandes" valeurs de X ont tendance à être réalisés pour de "grandes" valeurs de Y . C'est donc une mesure de monotonie pour des variables qualitatives ordinales.

Le coefficient le plus souvent utilisé pour mesurer des dépendances monotones entre les lignes et les colonnes d'un tableau de contingence est le γ de Goodman et Kruskal. Prenons deux couples (X_1, X_2) et (Y_1, Y_2) indépendants et ayant la même distribution que (X, Y) . On appelle C la probabilité que les deux couples soient en accord positif, et D la probabilité qu'ils soient en accord négatif. On définit

$$C = P(X_1 > X_2 \text{ et } Y_1 > Y_2) + P(X_2 > X_1 \text{ et } Y_2 > Y_1)$$

et

$$D = P(X_1 < X_2 \text{ et } Y_1 > Y_2) + P(X_2 > X_1 \text{ et } Y_2 < Y_1).$$

Le coefficient γ est donné par

$$\gamma = \frac{C - D}{C + D}.$$

Ce coefficient est compris entre -1 et 1 et si X et Y sont indépendantes, alors $\gamma = 0$. Plus γ est grand, plus il y a un lien entre les 2 variables qualitatives. Si γ est négatif, alors la liaison est négative.

A.2 Régression Logistique

A.2.1 Introduction

Le modèle linéaire généralisé constitue un cadre général permettant d'englober une grande quantité des modèles disponibles. Cette famille de modèles permet d'étudier la liaison entre une variable dépendante ou **réponse** Y et un ensemble de variables explicatives ou **prédicteurs** X_1, \dots, X_K . Plus précisément, un modèle linéaire généralisé spécifie comment l'espérance mathématique de Y , notée μ , est liée au prédicteur linéaire construit à partir des variables explicatives. On peut modéliser une fonction monotone $g(\mu)$ de l'espérance. On a alors :

$$g(\mu) = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K. \quad (\text{A.0})$$

La fonction g est appelée fonction de lien.

La fonction de lien $g(\mu) = \log \frac{\mu}{1-\mu}$ modélise le logarithme du rapport des chances. Elle est appelée logit et est adaptée au cas où μ est compris entre 0 et 1. C'est ce que l'on appelle la régression logistique.

La régression logistique est une des méthodes les plus utilisées dans le cas où la variable réponse Y est dichotomique. Lorsqu'elle est polytomique (comme c'est le cas ici), on parlera plutôt de régression logistique multinomiale mais le principe est identique. Dans son acception classique du terme, la régression logistique fait référence au cas où toutes les variables explicatives sont qualitatives. Il est cependant possible d'utiliser la régression logistique lorsque les variables explicatives sont qualitatives et quantitatives.

La fonction logistique est bien adaptée à la modélisation de probabilités, car elle prend ses valeurs entre 0 et 1 selon une courbe en S.

A.2.2 Régression logistique binomiale

En présence d'une variable réponse binaire Y (à valeur 0 ou 1) et d'une variable explicative X , on note $\pi(x) = P(Y = 1 | X = x) = 1 - P(Y = 0 | X = x)$. Le modèle de régression logistique consiste à poser

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}. \quad (\text{A.0})$$

De manière équivalente, on suppose que le log du rapport des côtes (odds ratios dans la littérature anglaise), aussi appelé logit, respecte la relation linéaire suivante :

$$\text{logit}[\pi(x)] = \log \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x. \quad (\text{A.0})$$

Plusieurs prédicteurs peuvent être pris en compte. On notera dans ce cas x_i la valeur (ou la modalité) prise par le prédicteur X_i .

A.2.3 Régression logistique multinomiale ou polytomique

Lorsque Y représente une réponse qualitative à J catégories (avec $J > 2$), la régression logistique multinomiale est utilisée. Elle est également appelée régression logistique polytomique. Celle-ci consiste à effectuer $J-1$ régressions logistiques binomiales correspondant aux combinaisons de la catégorie de référence avec les $J-1$ autres catégories (Agresti, 2002 ; Celeux et Nakache, 1994).

Dans le cas présent, la catégorie de référence est celle correspondant à l'absence d'urgence (catégorie 4). Nous effectuerons donc trois régressions logistiques binomiales diffé-

rentes :

$$\begin{aligned}\log\left(\frac{P(Y=1|X=x)}{P(Y=4|X=x)}\right) &= \log\left(\frac{\pi_1(x)}{\pi_4(x)}\right) = \beta_{01} + \beta_{11}x. \\ \log\left(\frac{P(Y=2|X=x)}{P(Y=4|X=x)}\right) &= \log\left(\frac{\pi_2(x)}{\pi_4(x)}\right) = \beta_{02} + \beta_{12}x. \\ \log\left(\frac{P(Y=3|X=x)}{P(Y=4|X=x)}\right) &= \log\left(\frac{\pi_3(x)}{\pi_4(x)}\right) = \beta_{03} + \beta_{13}x.\end{aligned}$$

Lorsqu'il y a K prédicteurs, on note $x = (x_1, \dots, x_K)$ une valeur de $X = (X_1, \dots, X_K)$. Le modèle logistique a alors pour expression :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_K x_K)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_K x_K)} \quad (\text{A.0})$$

Estimer un modèle de régression revient donc à estimer les coefficients β de ce modèle.

L'analyse précise des résultats d'un modèle logistique se déroule en trois grandes étapes : la sélection du "bon" modèle, la validation de ce modèle, et l'interprétation des résultats proprement dite.

A.2.4 Sélection du modèle logistique multinomial : critères de qualité du modèle et sélection de variables

Dans un cadre de régression logistique multinomiale, le nombre de coefficients à contrôler est très important, et le choix du "bon" modèle l'est d'autant plus.

La sélection du modèle commence par une comparaison deux à deux de suites de modèles dits *emboîtés*. Un modèle M_1 est emboîté dans un autre modèle M_2 lorsque toutes les variables prises en compte dans M_1 se retrouvent aussi dans M_2 . On peut aussi dire que le modèle M_1 est "plus petit" que le modèle M_2 , ou qu'il est moins complexe (i.e. avec moins de variables).

Plusieurs indicateurs permettent d'obtenir des renseignements sur le poids des variables explicatives et sur la qualité du modèle choisi. De manière globale, la qualité d'un modèle est mesurée par la vraisemblance : plus celle-ci est grande, plus le modèle est adapté aux données.

Dans le cadre de la régression logistique, il est d'usage d'utiliser à la place la quantité appelée déviance, souvent notée Λ :

$$\Lambda = -2\text{Log}(L), \quad (\text{A.0})$$

où L est la vraisemblance (Likelihood en anglais). Sous l'hypothèse que deux modèles sont emboîtés, la différence entre la déviance d'un modèle M_2 et celle d'un modèle M_1 "plus petit" est donc une valeur positive, qui suit une loi du χ^2 dont le degré de liberté est la différence entre le nombre de paramètres du modèle M_2 et celui du modèle M_1 . Lorsque le gain en terme de déviance (i.e. le gain en explication des données obtenu avec le modèle le plus complet par rapport au modèle le "plus petit") est faible, la p-value est élevée. En effet, plus la p-value est élevée, plus la différence des déviances est susceptible de suivre une loi de χ^2 , et plus les modèles sont proches. Les modèles acceptables sont donc ceux qui apportent un gain non négligeable en terme de déviance, relativement au nombre de paramètres utilisés (c'est à dire des modèles dont la p-value est faible).

Une autre mesure de qualité de modèle que nous utiliserons est le critère AIC pour "Akaike Information Criterion". Celui-ci est défini de la manière suivante :

$$AIC = -2 \log(L) + 2k, \quad (\text{A.0})$$

où L est la Vraisemblance du modèle, et k le nombre de paramètres de celui-ci. Ce critère permet de construire un classement de modèles statistiques tenant compte du principe de parcimonie. Les meilleurs ajustements correspondent aux plus faibles valeurs.

Lors de la sélection d'un modèle pertinent, il est important de repérer les variables explicatives qui ont une forte influence sur la variable à étudier (ici la classification des sons). Il est possible, tout d'abord, d'étudier le gain en Déviance (qui peut se ramener dans certains cas aux tests du maximum de vraisemblance appliqués aux tables de contingence).

Un autre critère communément utilisé est celui du test de Wald. Lors de l'estimation d'un modèle logistique additif incluant toutes les variables explicatives, nous pouvons mesurer une quantité égale au carré du rapport du coefficient estimé sur l'erreur commise. Cette quantité suit un $\chi^2(1)$ lorsque le coefficient considéré est égal à 0. L'usage de ce test est donc de sélectionner les variables susceptibles d'être exclues du modèle final, correspondant à des p-values élevées, c'est à dire les variables avec une statistique de Wald faible (et donc une erreur de mesure importante). Ce test est à utiliser en combinaison avec le critère des Déviances car il peut induire des erreurs d'appréciations.

A.2.5 Validation du modèle

Une fois que le modèle est sélectionné, il reste à le valider. Afin de valider un modèle logistique, il est nécessaire de s'assurer que celui-ci est robuste, c'est à dire qu'il n'est pas

trop dépendant des données utilisées pour le construire. Pour ce faire il est d'usage de re-estimer le modèle uniquement sur une partie des données. En général, une proportion de 70% des données est suffisante. Il ne reste alors plus qu'à obtenir les prédictions de classes fournies par le modèle, et tester leur validité sur les 30% de données restantes.

Dans notre cas, sur les réponses fournies sur les 30% de données restantes, le même son est classé plusieurs fois, par plusieurs personnes différentes. Or, pour un son donné, le modèle cherche à prédire la répartition des classes représentée dans les réponses. Donc, lors de la prédiction de la catégorie d'un nouveau son, celle-ci correspond à la classe majoritairement représentée dans les prédictions du modèle (i.e. la classe ayant la probabilité la plus élevée). En testant les prédictions du modèle avec un grand nombre de sujets, il n'y aura donc accord que lorsque le sujet donne la classe majoritairement prédite par le modèle. De ce fait, il faut s'attendre à un fort taux d'erreur dans les tests de validation.

Néanmoins, il existe un autre moyen d'étudier encore plus précisément quels sont les sons pour lesquels l'adéquation est la moins bonne : les résidus de Pearson. Lorsque ces coefficients sont compris entre -2 et 2, il est d'usage de considérer que le modèle est bon et qu'aucune erreur grave n'a été commise.

A.2.6 Interprétation des résultats du modèle

La régression logistique utilise des variables indicatrices, c'est à dire des variables valant 0 ou 1 selon que la modalité associée à la variable indiquée est présente dans le son étudié ou non.

Dans notre cas, on note z_{SS-1L} , z_{SS-5S} , z_{SS-5F} pour les variables indicatrices représentant la variable Signal.Silence et $z_{R-2.5s}$, $z_{R-accel}$, $z_{R-0.6s}$ pour les variables indicatrices représentant la variable Répétition. On utilisera z_{F0} pour la fréquence fondamentale, z_T pour le Timbre, z_P pour la Polyphonie, et $z_{1L2.5S}$, $z_{5S2.5S}$, $z_{5F2.5S}$, $z_{1Laccel}$, $z_{5Saccel}$, $z_{5Faccel}$, $z_{1L0.6s}$, $z_{5S0.6s}$, $z_{5F0.6s}$ pour l'interaction SS×R. Avec l'aide de ces variables indicatrices et du modèle logistique estimé, il devient possible de calculer le log du rapport des probabilités $\frac{\pi_1}{\pi_4}$, $\frac{\pi_2}{\pi_4}$ et $\frac{\pi_3}{\pi_4}$, π_4 étant la probabilité de classement dans la catégorie 4 servant de référence (le choix de la catégorie de référence est arbitraire ; l'interprétation des résultats sera la même par la suite, quelle que soit la catégorie choisie).

De manière équivalente, il est possible de calculer tous les autres rapports de probabilités de classement avec l'aide de formules provenant de propriétés évidentes du logarithme.

Ainsi, pour calculer $\log\left(\frac{\pi_1(x)}{\pi_2(x)}\right)$ on utilisera l'équation suivante :

$$\log\left(\frac{\pi_1(x)}{\pi_2(x)}\right) = \log\left(\frac{\pi_1(x)}{\pi_4(x)}\right) - \log\left(\frac{\pi_2(x)}{\pi_4(x)}\right). \quad (\text{A.0})$$

Nous pouvons ainsi calculer toutes les probabilités de classement possibles. Lorsque le son considéré fait intervenir la valeur d'une variable qui constitue la modalité de référence (par exemple la modalité *A1* pour la variable *F0*), la variable indicatrice associée est égale à 0. Il ne faut donc pas tenir compte du coefficient associé (par exemple, lorsque le son utilise la modalité *A1*, $z_{F0} = 0$).

Remarque : La modalité *1S* étant la catégorie de référence pour la variable Signal.Silence, tous les termes de l'interaction *SS*×*R* comportant la modalité *1S* sont aussi considérés comme étant des catégories de référence, et n'ont donc pas de coefficients associés.

Pour des raisons d'interprétation évidentes, il est nécessaire de revenir aux probabilités π_1, π_2, π_3 et π_4 de classement dans chaque catégorie pour un son déterminé (et donc de s'affranchir du logarithme). La formule suivante permet ce passage :

$$\pi_j(x) = \frac{\exp(\alpha_j + \beta_j x)}{1 + \sum_{k=1}^{J-1} \exp(\alpha_k + \beta_k x)} \quad (\text{A.0})$$

avec $\alpha_J = 0$ et $\beta_J = 0$.

De cette façon, il est possible de calculer les probabilités de classement dans chaque catégorie pour toutes les combinaisons des variables constituant les sons.

Résumé

Le thème de cette thèse est l'étude expérimentale des alarmes sonores de l'automobile. Deux approches sont confrontées : la perception de l'urgence et le temps de réaction. L'ensemble des recommandations (issues de la littérature) permettant de créer des alarmes sonores efficaces a été questionné à nouveau dans le cadre d'une expérience appliquée aux sons fonctionnels de l'automobile. Les résultats confirment la prédominance des paramètres rythmiques, mais mettent également en évidence des effets inverses à ceux prédits par la littérature. L'interprétation de ces derniers résultats conduisent à proposer une nouvelle mesure objective de l'urgence. Un paradigme de temps de réaction (TR) a été adopté. Trois expériences ont alors testé l'influence de paramètres rythmiques sur le TR en contexte d'attention partagée (simulant la conduite automobile), et ont permis de comparer les deux approches expérimentales d'urgence perçue et de TR. Il apparaît que le TR décroît lorsque le tempo croît. Un modèle auditif combinant les approches de résolution et d'intégration temporelles est proposé pour expliquer ce résultat. D'autre part, l'irrégularité rythmique d'un son capte l'attention des auditeurs plus qu'une régularité. Ce résultat s'intègre dans un cadre théorique qui expose comment des auditeurs utilisent les structures temporelles des stimuli sonores. Ces deux résultats sont valables dans les deux contextes attentionnels testés. Deux nouvelles expériences permettent de confronter des sons de félins à des sons synthétiques, créés par manipulation d'un intervalle inter-impulsion. Tout d'abord, en comparant les sons synthétiques entre eux, le TR est minimum pour un intervalle inter-impulsion de 33 ms. Ce résultat est mis en parallèle avec différentes études mettant en évidence cette valeur seuil de 30 ms. En outre, la comparaison des deux types de sons montre que le TR est systématiquement plus faible pour les sons de félins. Deux interprétations sont alors possibles pour expliquer l'avantage de ces sons d'animaux : soit la reconnaissance de l'animal en lui-même est en jeu, soit le contenu spectral plus riche des sons de félins permet de réagir plus rapidement. La compréhension de ces phénomènes sous-jacents ouvre la voie à une création d'alarmes sonores précises et efficaces.