Perception et confort acoustiques des systèmes de traitement d'air

Antoine Minard

UNIVERSITÉ DE LA ROCHELLE
UNIVERSITÉ DE LIÈGE

École Doctorale Sciences Ingénierie en Matériaux, Mécanique, Énergétique et Aéronautique (SI-MMEA – N° 522)

Collège de doctorat en Électricité, électronique et informatique

THÈSE

pour obtenir le grade de :

DOCTEUR DE L'UNIVERSTITÉ DE LA ROCHELLE ET DE L'UNIVERSITÉ DE LIÈGE

Discipline : **GÉNIE CIVIL : ACOUSTIQUE**

Présentée et soutenue publiquement par :

ANTOINE MINARD

le 26 mars 2013

Perception et confort acoustiques des Systèmes de Traitement d'Air

Dirigée par **ANAS SAKOUT** & **JEAN-JACQUES EMBRECHTS**

JURY :

RAPPORTEURS :

S. MEUNIER	LMA, Marseille
E. PARIZET	INSA, Lyon

EXAMINATEURS :

J.-J. EMBRECHTS	Université de Liège
F. FOURNIER	CIAT, Culoz
A. SAKOUT	Université de La Rochelle
J. VERLY	Université de Liège

Perception et confort acoustiques des Systèmes de Traitement d'Air

Résumé :

Cette thèse a pour but d'étudier le confort acoustique ressenti par les usagers de systèmes de traitement d'air (STA), tels que les systèmes de climatisation installés en bureaux de différentes tailles, en prenant en compte les facteurs environnementaux liés au contexte particulier d'usage de ces produits. À l'origine s'est présenté le besoin industriel d'une norme fiable d'évaluation des sons émis par les STA, pour offrir une représentation plus fidèle de la perception que le seul standard existant aujourd'hui : le niveau acoustique émis exprimé en dBA. Ce contexte impose en premier lieu une étude approfondie de la perception des sons de STA tels qu'émis par les appareils. Ainsi, une démarche rigoureuse a été suivie allant du recueil d'un nombre conséquent d'enregistrements de sons de STA jusqu'à l'établissement d'une métrique robuste de prédiction de la qualité sonore perçue. Pour ce faire, ont tout d'abord été établies les différentes familles perceptives qui constituent l'ensemble des enregistrements de sons de STA effectués. De cet ensemble de familles a été extrait un corpus de travail permettant une complète représentativité des différents types de sons émis par les STA. Les attributs auditifs pertinents pour la description perceptive de ce corpus sonore ont ensuite été identifiés en appliquant les principes de l'étude du timbre musical, déjà adoptés avec succès pour la description d'autres types de sons de l'environnement. Ces attributs auditifs ont enfin servi de base descriptive pour expliquer les préférences des auditeurs vis-à-vis des sons de STA afin d'établir un prédicteur efficace de la qualité sonore à l'aide de descripteurs audio. Dans un second temps, afin de prendre en considération le contexte écologique des STA, l'influence de deux facteurs environnementaux sur l'évaluation de la qualité sonore des STA a été étudiée dans le but d'en évaluer l'importance sur le ressenti des usagers. D'une part, comme les STA étudiés sont des appareils exclusivement installés en intérieur (bureaux), l'effet de la réverbération sur l'évaluation de la qualité sonore a été étudié à l'aide d'un système d'auralisation permettant de reproduire virtuellement la réponse acoustique d'une salle. D'autre part, l'influence du contexte attentionnel des auditeurs sur la qualité sonore perçue a été évaluée à l'aide d'une étude comparative de différentes situations d'écoute. En effet, on observe que la perception du son émis par les STA se traduit typiquement par une forme de perturbation de leur occupation quotidienne. Il est donc apparu pertinent d'évaluer à quel point le degré d'attention portée sur le son influe sur l'évaluation de la qualité sonore par les auditeurs. Ainsi nous avons pu établir dans quelles mesures et selon quelles limites le prédicteur de qualité sonore établi peut représenter fidèlement le confort ressenti par les usagers dans un contexte offrant un meilleur degré de validité écologique que les conditions habituelles de laboratoire.

Mots-clés : Systèmes de traitement d'air, Évaluation de la qualité sonore, Design sonore, Étude du timbre, Réverbération, Contexte attentionnel.

Acoustic perception and comfort of Air-Treatment Systems

Abstract :

This thesis addresses the perceived acoustic comfort of Air-Treatment Systems (ATS), such as air-conditioners installed in offices, by taking into account the environmental factors related to the specific context of ATS usage. The only existing standard to evaluate the sounds emitted by ATS, which is the emitted sound level in dBA, is only loosely related to perception. Therefore, the need of manufacturers for a more reliable standard arises. This implies a thorough study of the perception of the sound of ATS as it is emitted. A precise methodology was then followed : it includes first collecting a high number of ATS sound recordings, up to finally developing a robust metrics to predict the perceived sound quality. For that purpose, different perceptual categories were first identified to constitute the recording database of ATS sounds. A corpus considered as fully representative of the different types of emitted sounds was then extracted from the recording database. Current principles of musical timbre description have already proved to be adequate to other types of environmental sounds ; by applying these principles, the relevant auditive attributes for the corpus perceptual description were identified. In order to develop an efficient sound quality predictor through audio features calculation, prominent features based on these auditive attributes were identified that explain the listeners' preferences among ATS sounds. The ecological context of ATS was examined in a second step. Two environmental factors were addressed in the context of ATS sound quality evaluation to ponder their importance in the listeners' perception. As the ATS under study are exclusively indoor systems designed for offices, the effect of reverberation on sound quality evaluation was first studied ; for that purpose, an auralization tool was used to simulate room acoustic response. The influence of listeners' attention context on perceived sound quality was then evaluated through a comparative study of various listening conditions. As a matter of fact, the sound emitted by ATS in real conditions is perceived as a perturbation of current activities. It is therefore relevant to evaluate how deeply the degree of attention related to the sound affects listeners as regards their perception of acoustic quality. Eventually, the relevance of the proposed sound quality predictor to comfort perception was assessed in conditions more ecologically representative than usual laboratory environment.

Keywords : Air-treatment systems, Sound quality evaluation, Sound design, Timbre study, Reverberation, Attention context.

Remerciements

Je tiens tout d'abord à remercier les membres du jury d'avoir accepté d'examiner cette thèse de doctorat. En particulier, je remercie sincèrement Jacques Verly d'avoir présider mon jury de doctorat. Également, je souhaiter remercier Sabine Meunier et Étienne Parizet d'avoir accepté de rapporter mes travaux de thèse. Je remercie chaleureusement Francette Fournier pour, en outre de sa participation au jury de doctorat, ses précieux conseils et les échanges enrichissants tout au long du déroulement de la thèse.

Je voudrais remercier tout particulièrement mes deux directeurs de thèses : Jean-Jacques Embrechts, pour son encadrement et pour m'avoir accueilli pendant plusieurs mois au sein de son laboratoire à Liège et permis de profiter des installations de ce dernier ; Anas Sakout, pour m'avoir proposé cette opportunité et pour, plus que son simple encadrement, son soutien permanent dans les aléas inhérent à trois années de thèse.

Je tiens également à exprimer ma gratitude envers le département Recherche et Innovations du groupe CIAT, et tout particulièrement Pierre-Jean Vialle et Benjamin Robin pour nos échanges et pour leurs conseils et leur disponibilité. J'adresse également mes remerciements à tous les membres de ce département pour leur accueil chaleureux lors de mes quelques visites à Culoz, et notamment aux techniciens pour avoir facilité toutes nos contraignantes campagnes de mesures, toujours avec bonne humeur et gentillesse.

J'adresse également de sincères remerciements à Francis Allard et Karim Ait-Mokhtar de m'avoir accueilli pendant ces quelques années au sein du LaSIE. Je remercie également l'ensemble du personnel, passé et présent, du laboratoire, dont j'ai pu apprécier l'accueil et l'ambiance agréable.

Je ressens également une grande gratitude envers Bertrand Goujard et Alexis Billon pour avoir « gravité » autour de ce projet, et pour leurs nombreux conseils avisés sur les points plus techniques de ce travail. Je les remercie tout particulièrement pour leur investissement lors des phases de rédaction d'articles et de la thèse.

Je remercie également toutes les paires d'oreilles qui se sont prêtées à mes expériences perceptives, sans qui, bien entendu, cette thèse n'aurait pas pu aboutir.

Sur un plan plus personnel, je voudrais adresser un grand merci à mes collègues anciens et actuels doctorants, post-doctorants et stagiaires, avec qui j'ai pu partager de quelques semaines à quelques années d'excellents moments qui me feront garder un souvenir fort agréable de mon passage à La Rochelle : Vincent B., Thomas, Antoine (pour sa bonne humeur infaillible), Maxime, Alexandra (nos galères concomitantes de fin de thèse), Adrien D., Mireille, Adrien G. (et tous ses gâteaux), Serge, Luc, Jean-Louis, Phu Tho, Salah, Hassan, Rémy, Axel, Ibrahim, Yacine, Nabil, Rabah, Kamilia, Kevin, Malek, Nissrine, Alice, Valérian, Laurent, Massi, Pierre (les séances hivernales d'entraînement au semi-marathon), Issa, Walid, Abdelkrim, Mahfoud, Nazir, Vincent N., Marie, Khaled, Benjamin... et tous ceux que j'oublie. A Liège, je remercie également Stéphane, Laurenz, Alexandre et tous les collègues

du laboratoire.

Je remercie enfin tous mes amis et ma famille, d'avoir simplement été là pendant ces quelques années, malgré l'éloignement et malgré ma récurrente indisponibilité : Guillaume, Henri, Pierre, Grégoire, Vicky, Aymeric... mes deux sœurs adorées, Anne-Laure et Aurélia, et surtout mes parents, qui m'ont mené là où je suis aujourd'hui, et à qui je dois tellement plus que cette thèse.

Table des matières

Table des figures

Liste des tableaux

Introduction

L'acoustique devient une priorité pour les industriels, car le choix des consommateurs passe, entre autres, par le confort acoustique lié au produit. En effet, les processus de fabrication devenant avec les progrès technologiques de plus en plus efficients et standardisés, les industriels cherchent de nouveaux moyens de se démarquer de la concurrence en améliorant leurs produits selon d'autres critères que la seule efficacité à remplir leur fonction. Ainsi, beaucoup se tournent aujourd'hui vers un processus de « design sonore » dans lequel l'évaluation du confort acoustique est une étape primordiale parce qu'elle permet de définir les spécifications acoustiques du produit. Les industriels, qui, pendant longtemps, se sont limités à rendre leur produit le plus silencieux possible, adoptent aujourd'hui d'autres démarches de recherche basées sur l'évaluation perceptive de la qualité sonore du produit, afin d'identifier les paramètres acoustiques influant sur le confort ressenti. Ils cherchent ainsi à définir dans quelle mesure les paramètres perceptifs influent sur le caractère gênant, désagréable ou intrusif du son produit par un objet conçu pour remplir une certaine fonction et non pour avoir un « beau » son, contrairement aux instruments de musique, par exemple.

C'est dans ce cadre que nous choisissons d'aborder la problématique du confort acoustique lié Systèmes de Traitement d'Air (STA). Les travaux de thèse ont été réalisés en cotutelle entre le LaSIE de l'Université de La Rochelle et l'institut Montefiore de l'Université de Liège. Ce travail s'inscrit dans le cadre du projet VAICTEUR AIR^2 [1], qui regroupe plusieurs partenaires, et porté par le groupe CIAT [2]. Le but de ce projet est le développement de technologies pour l'amélioration de la qualité des ambiances intérieures des bâtiments en diminuant la consommation d'énergie liée au chauffage, à la ventilation ou au conditionnement d'air. Cette amélioration de la qualité des ambiances intérieures passe également par l'optimisation du confort acoustique ressenti par les usagers dans les bâtiments. Dans ce contexte, et celui plus spécifique de l'étude du confort acoustique des STA, nous souhaitons ici adopter une démarche rigoureuse allant du recueil d'un nombre conséquent d'enregistrements du son de STA jusqu'à l'établissement d'une métrique robuste de prédiction de la qualité sonore perçue par les auditeurs. Nous cherchons donc à définir les paramètres acoustiques qui présentent une influence significative sur la perception et la qualité sonores des STA, qu'ils soient propres au son directement produit par les STA ou au contexte de perception.

Ce document s'articule en trois parties. La **partie I** précise la problématique générale et présente les choix méthodologiques effectués sur la base d'une étude bibliographique approfondie. La **partie II** présente les travaux réalisés afin d'établir une métrique de la qualité des sources sonores que sont les STA. Enfin, la **partie III** expose l'étude de l'influence de facteurs liés au contexte d'écoute sur la qualité sonore perçue.

1. Vie de l'homme dans les Ambiances Intérieures Contrôlées, Traitement par Énergies Utiles et Renouvelables, Apport d'Innovations Récurrentes et/ou de Rupture, `http://www.ciat.fr/medias/cp_vaicteur_air2_fra.pdf`. Ce projet est soutenu par les organismes OSEO et ADEME.

2. Compagnie Industrielle d'Applications Thermiques, `http://www.ciat.com/`.

Partie I

Problématique et positionnement méthodologique

Le problème industriel du son de STA soulève des questions de recherche fondamentale sur sa perception. Afin de développer ces questions et d'explorer les méthodologies existantes permettant d'y répondre, cette partie du document présente le positionnement de cette thèse et les orientations méthodologiques adoptées. Le **chapitre 1** introduit la problématique générale. Le **chapitre 2** expose l'état de l'art concernant l'évaluation de la perception et de la qualité des sons de l'environnement, dans le cadre desquelles se place l'étude du confort acoustique de STA. Enfin, le **chapitre 3** développe différents éléments bibliographiques importants afin d'aborder la problématique du contexte d'écoute lié aux STA.

Chapitre 1

Problématique et démarche générales

Ce chapitre a pour vocation, dans un premier temps, de présenter succinctement la problématique industrielle qui représente le point de départ de cette thèse (section 1.1). Il convient par la suite d'introduire le contexte dans lequel nous nous plaçons afin de répondre à cette problématique, ce qui implique la définition de quelques concepts essentiels pour la compréhension de ce document (section 1.2). Ces définitions permettent alors d'exposer la démarche globale adoptée dans le cadre de cette thèse (section 1.3).

1.1 Problématique industrielle

Le Système de Traitement d'Air (STA) représente un élément de l'environnement qui a tendance à se généraliser à l'intérieur des bâtiments, depuis de nombreuses années déjà. Si cette généralisation est bien évidemment rendue nécessaire par la volonté d'améliorer le confort thermique des usagers des bâtiments, elle s'accompagne également d'une détérioration du confort acoustique. En effet, les STA, par leur fonctionnement même, génèrent un bruit qui peut être perçu comme une source de nuisance. Ainsi, la problématique du confort acoustique dans les bâtiments, que ce soit d'un point de vue général ou dans le cas particulier des STA, a rencontré un fort intérêt ces dernières années, notamment des pouvoirs publics qui ont établi un certain nombre de normes acoustiques à respecter (bruit ambiant, isolations des sources extérieures, ...) afin de limiter les nuisances sonores.

En réponse à cela, et également dans le but de se démarquer de la concurrence, les industriels ont commencé à intégrer la notion de confort acoustique dans le processus de conception de leurs produits afin d'en améliorer la qualité sonore perçue. Dans un premier temps, il s'est agi de diminuer quantitativement le bruit émis par les appareils. Depuis longtemps, la seule grandeur utilisée afin d'en fournir une caractérisation acoustique est la mesure du niveau de puissance acoustique exprimé en dBA (voir section 2.3.3 pour plus de détail). C'est sur la base de cette seule métrique du son que sont fondées les normes existantes. Toutefois, cette métrique décrit le son d'une manière particulièrement réductrice comme une quantité d'énergie qu'il s'agit de diminuer autant que faire se peut, voire d'éliminer, afin d'obtenir un confort acoustique optimal. Toutefois, cette stratégie devient moins efficace à mesure que la réduction du niveau de puissance acoustique se rapproche d'une forme de limite physique, et les améliorations techniques supplémentaires pour réduire encore le niveau sonore émis deviennent coûteuses et/ou complexes à mettre en œuvre efficacement.

Ainsi, si les fabricants de STA sont parvenus à améliorer le confort ressenti par les usagers en réduisant quantitativement le bruit émis à la source, les limites physiques de cette approche amènent aujourd'hui à envisager de nouvelles stratégies d'amélioration. En effet, le manque de moyens de

quantifier de manière représentative la qualité sonore associée aux STA les a confrontés à la difficulté d'apporter une réponse efficace lorsque la mesure du niveau de puissance acoustique en dBA s'avère insuffisante pour décrire fidèlement le confort ressenti par les usagers. Ce constat représente tout l'enjeu de cette thèse. Nous souhaitons donc apporter un ensemble de connaissances du son de STA et d'outils permettant d'en quantifier les caractéristiques à partir du signal sonore. À terme, le but pour l'industriel est d'améliorer le confort acoustique ressenti par les usagers, en intégrant ces éléments au processus de conception de nouveaux produits.

1.2 Contexte scientifique de l'étude

Afin de définir précisément le contexte de cette thèse, il convient dans un premier temps de définir précisément certaines notions qui cadrent le travail dont ce document fait l'objet.

1.2.1 Les sons de l'environnement

Les sons de STA font partie d'un catégorie particulière de sons qu'il importe de préciser afin de comprendre la démarche mise en place : les sons de l'environnement. Nous entendons ici par « sons de l'environnement » l'objet de la définition de Vanderveer [155] :

> « ...any possible audible acoustic event which is caused by motions in the ordinary environment. [...] Besides having real events as their sources [...], [they] are usually more "complex" than laboratory sinusoids, [...], [they] are meaningful, in the sense that they specify events in the environment. [...], the sounds to be considered are not part of a communication system, or communication sounds, they are taken in their literal rather than signal or symbolic interpretation. [1] »

Cette définition exclut donc notamment les sons musicaux, c'est-à-dire conçus et émis spécifiquement dans un but musical, qu'ils proviennent ou non d'un instrument de musique traditionnel. Elle exclut également tous les sons « humains », et notamment la parole ou les imitations vocales.

1.2.2 Le design sonore

La définition précédente des sons de l'environnement a été introduite spécifiquement dans le cadre du *design sonore industriel – product sound design* – tel que défini par Blauert et Jekosch [27]. Le processus de design sonore industriel est schématisé sur la figure 1.1. L'étude des sons de l'environnement, et notamment l'évaluation de la qualité sonore, correspondent à la première étape d'*analyse* de l'existant, dont le but est d'obtenir un système de représentation des sons de l'environnement fidèle à leur perception. Ce système de représentation permet par la suite de produire des spécifications de conception, véritable cahier des charges et indispensables à tout projet de *création* de sons. Enfin, une dernière étape a pour but de *valider* l'*objet sonore* ainsi créé, ou, le cas échéant, d'affiner les spécifications avant une nouvelle étape de création.

1. ...un évènement acoustique audible quelconque causé par des mouvements dans l'environnement ordinaire. [...] Outre le fait d'avoir de réels évènements comme sources [...], [ils] sont habituellement plus "complexes" que des sinusoïdes de laboratoire, [...], [ils] sont porteurs d'information, en ce sens qu'ils permettent d'identifier les évènements dans l'environnement. [...], les sons considérés ne font pas partie d'un système de communication, ou des sons de communication, ils sont considérés en eux-mêmes plutôt qu'interprétés comme signaux ou comme symboles.

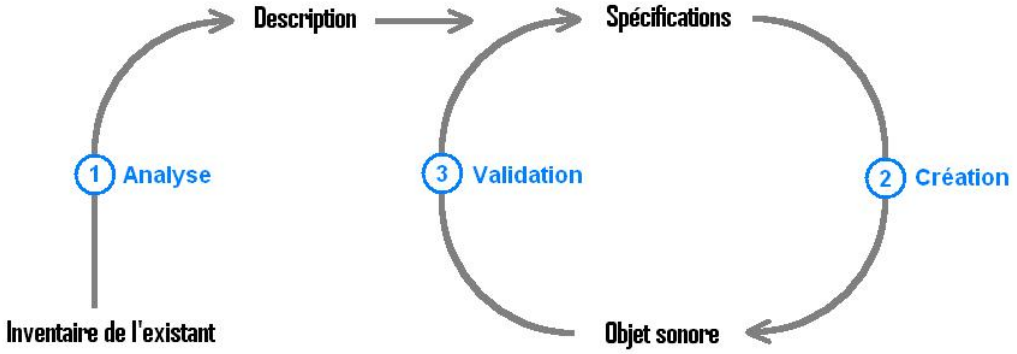


FIGURE 1.1 – Schéma des différentes étapes du *design sonore industriel.*

1.2.3 Qu'entend-on par « qualité sonore » ?

Dans le cadre de cette thèse, nous nous plaçons clairement au niveau de l'analyse de l'existant dans le schéma 1.1. Le travail présenté ici est donc très en amont du processus global de design sonore. Afin d'établir des spécifications techniques acoustiques, cette étape nécessite d'évaluer et d'expliciter la qualité sonore des éléments considérés, dans le cas général, les sons de l'environnement, et dans ce cas plus spécifique, les sons de STA. Il importe alors de définir précisément ce que l'on entend par « qualité sonore ».

En parcourant la littérature scientifique sur le sujet, on s'aperçoit rapidement que la définition de l'expression « qualité sonore » échappe au consensus. En conséquence, le but recherché par les industriels lorsqu'ils abordent cette question peut varier considérablement selon le produit concerné et sa fonction. On distingue typiquement trois cas de figure :

- le son considéré est la conséquence fortuite du fonctionnement normal d'un produit industriel, et n'a pas de raison d'être particulière (par exemple, un grincement de porte) ;
- le son considéré est la raison d'exister du produit, et a par conséquent une fonction particulière dans l'environnement (par exemple, les alarmes) ;
- le son est la conséquence du fonctionnement normal du produit, mais est également porteur d'information (par exemple, le son du moteur d'une voiture qui peut être porteur d'information pour le conducteur).

Dans ces trois cas, le but des industriels peut différer radicalement. Dans le premier cas, l'idéal serait de réduire le son produit au point qu'il ne soit plus perçu. En effet, on parle souvent alors de « gêne » ou de « désagrément ». Dans les deux autres cas, éliminer totalement le son produit n'est pas envisageable. De ce simple constat découle le fait que les stratégies adoptées pour aborder l'étude de la qualité sonore pour chacun de ces cas de figures peuvent être différentes. Mais cette différence se manifeste plutôt au travers de la définition du terme « qualité » et l'interprétation des résultats des expériences menées dans ce domaine, que du choix de la méthodologie expérimentale. Ainsi, si les recherches se portent souvent sur le caractère « hédonique » d'un son, il arrive que d'autres considérations doivent être prises en compte, afin de respecter l'aspect écologique du son considéré, c'est-à-dire sa place dans l'environnement. Le terme « qualité » n'est alors que l'expression générique de l'échelle que l'on tente d'étudier dans ces différents cas.

Si nous nous intéressons ici plus particulièrement au premier cas cité, c'est-à-dire la caractérisation hédonique du son, il est intéressant de mentionner les travaux menés dans le domaine du design

sonore sur la fonction du son. En effet, il est impossible de négliger l'identification de la source sonore lorsque l'on s'intéresse à la perception des sons de l'environnement (voir section 2.2). Il en découle que l'étude du rôle d'un son dans l'environnement est indispensable pour en estimer la « qualité ». De telles considérations se retrouve dans la thèse de Lemaitre [99] portant sur la perception acoustique des avertisseurs sonores automobiles, et plus particulièrement dans une étude de la « typicité » de tels sons [101]. Il entend par « typicité » le degré auquel les auditeurs considèrent qu'un enregistrement correspond bien à un son de klaxon. Il tente donc de construire une échelle permettant de quantifier la capacité d'un son de klaxon à remplir sa fonction d'avertisseur sonore. Cette échelle pourrait être considérée, dans ce cas précis, comme une échelle de qualité, puisque « avertir » est la raison d'être d'un klaxon. Ce rapprochement peut se retrouver dans d'autres travaux. On peut notamment citer une étude de Patsouras et al. [123, 124] abordant la « dieselitude » (« *dieselness* ») de sons de moteurs automobiles, ou les travaux sur le design de sons pour la navigation dans une interface complexe, comme un système d'exploitation informatique (*auditory displays*, voir [68, 41], entre autres).

Du point de vue de la caractérisation hédonique du son, de nombreuses études entreprennent d'aborder la problématique de l'évaluation de la qualité sonore, en la désignant par des termes différents. Parfois, ce ne sont que des synonymes, et parfois, ils sont porteurs de connotations sémantiques qui sont loin d'être anodines. Le terme le plus souvent utilisé est la *gêne – annoyance*. Lindvall et Radford définissent la gêne par « a feeling of displeasure associated with any agent or condition believed to affect adversely an individual or a group [2] » [102]. Berglund et al. ajoutent une nuance sémantique, et associent cette notion de gêne à un contexte environnemental donné : « Annoyance was defined as the nuisance aspect of the noise experienced in an imaginary situation phrased as : "After a hard day's work, you have just been comfortably seated in your chair and intend to read your newspaper" [3] » [19]. La définition précédente de la gêne, indépendante de tout contexte, est plus associée par Berglund et al. au terme de *bruyance*, ou caractère bruyant – *noisiness*. Toutefois la définition initiale de Lindvall et Radford de la gêne est plus souvent adoptée.

Plus récemment, Zwicker, trouvant cette définition de la gêne trop vague, préféra la préciser en introduisant la notion de « gêne non-biaisée » – *unbiaised annoyance* – incluant plusieurs conditions dont l'absence pourrait affecter la significativité de la gêne mesurée [167]. Cette définition impose notamment que l'auditeur n'ait pas de « relation » avec le son et la source sonore, en d'autres termes qu'il « subisse » le son (par exemple un motard ne jugera pas le son de moto de la même manière qu'un piéton). Elle exclut toute influence de la source de gêne liée aux autres modes sensoriels (vibrations, esthétique visuelle, odeurs associées, etc...). Enfin, dans l'expérience visant à mesurer la gêne, elle nécessite une reproduction fidèle des conditions réelles (type de champ, activité de l'auditeur, etc...).

D'autres termes, comme *désagrément – unpleasantness* – ou *dérangement – disturbance* – par exemple, peuvent également apparaître dans la littérature, mais apportent peu à la définition initiale. Paulsen a notamment observé, dans une expérience de mesure de gêne, de dérangement et de désagrément, que les différences observées dans ces mesures n'étaient pas significatives [126]. Toutefois, l'emploi d'un terme ou d'un autre dépend fortement des nuances sémantiques propres à la langue utilisée dans le cadre des expériences réalisées et du contexte d'étude particulier qui peut parfois imposer une terminologie spécifique. De manière plus générale, Guski [72] définit la gêne comme l'interférence du son sur une situation particulière, associée à un contexte environnemental spécifique,

2. « une sensation de déplaisir associée à tout agent ou condition perçus comme affectant négativement un individu ou un groupe »

3. « La gêne était présentée comme l'aspect nuisible du bruit ressenti dans une situation imaginaire décrite par : "Après une dure journée de travail, vous venez de vous installer confortablement dans votre fauteuil et comptez lire votre journal" »

avec parfois une tâche à réaliser, tandis que le désagrément correspond à une mesure où l'auditeur voit son attention focalisée sur le son.

1.3 Démarche adoptée

L'exploration en section 1.2.3 des différentes définitions théoriques de la « qualité sonore »permet de se faire une meilleure idée du percept des STA que nous cherchons à estimer. Nous avons d'ores et déjà mis de côté les définitions qui concernent des sons ayant une fonction, directe ou indirecte, dans leur environnement, car ce n'est pas le cas des sons de STA. En effet, si le STA a, lui, une fonction évidente, le son qu'il produit n'en a pas, et n'est que la conséquence inévitable de son fonctionnement [4]. Nous considérons donc principalement, dans le cadre de cette thèse, une caractérisation hédonique du son de STA, traduisant l'appréciation que les usagers en font.

Cette hypothèse fondamentale posée, ce que nous entendons par « qualité sonore » nécessite encore quelques précisions. Si nous tentons de résumer les éléments fournis en section 1.2.3 au sujet de la caractérisation hédonique des sons de l'environnement, nous pouvons considérer qu'il existe deux visions majeures de la notion de qualité sonore :

- La définition de *gêne* de Lindvall et Radford [102] et de *bruyance* de Berglund et al. [19] qui décrit le sentiment de déplaisir provoqué par le son indépendamment de tout contexte d'écoute.
- La définition de *gêne non-biaisée* de Zwicker [167] (initialement appelée simplement *gêne* par Berglund et al. [19]), qui replace la définition précédente par rapport à un contexte donné.

Dans le cas des STA, il semble que la seconde définition fournisse une description plus fidèle du confort ressenti par les usagers. Cependant, il convient de préciser que la prise en compte de l'environnement d'écoute n'est pas chose aisée. Certains éléments qui peuvent avoir une influence sur le confort ressenti sont difficilement quantifiables (l'activité des usagers lorsque ceux-ci perçoivent le son de STA, pour ne citer qu'un seul exemple). Par ailleurs, les différents éléments qui peuvent influer sur le confort ressenti sont difficilement dénombrables, et il semble impossible de tous les prendre en compte. Ainsi, si l'évaluation de la qualité sonore, selon la première définition, souffre en un sens d'un manque de représentativité, elle permet toutefois de la recentrer sur la source sonore qu'est le STA, élément sur lequel les fabricants peuvent agir afin d'améliorer le confort des usagers

Bien entendu, il convient, afin d'étudier la validité des résultats établis dans le cadre de l'évaluation de la qualité sonore selon la première définition, d'évaluer l'étendue de l'influence des paramètres liés au contexte d'écoute. Cela revient à aborder la seconde définition de la qualité sonore. Toutefois, compte tenu du grand nombre de paramètres potentiellement influents, il importe également de cibler plus précisément ceux qui nous semblent à priori les plus susceptibles d'avoir un réel impact sur le confort ressenti.

Ces deux définitions introduisent donc les deux orientations de recherche qui ont été suivies dans le cadre de cette thèse. En particulier, dans le contexte de cette partie du document, elles établissent la façon dont sont décomposés les deux chapitres suivants.

4. On pourrait ici objecter que le son de STA permet d'indiquer ou de confirmer à l'usager que l'air est bien traité dans la pièce où il se trouve, améliorant ainsi le confort ressenti d'un point de vue global. Mais cet aspect, si tant est qu'il ait un réel impact, nécessite une étude multisensorielle qui n'est pas le propos de cette thèse.

Chapitre 2

Qualité des sons de l'environnement

Ce chapitre présente l'état de l'art dans le domaine de l'évaluation de la qualité sonore. Les sons de STA font partie d'une catégorie de sons clairement identifiée que sont les sons de l'environnement (dont une définition précise est explicitée en section 1.2.1). Le but est donc ici de présenter les éléments de la littérature scientifique qui permettent de qualifier et de quantifier la perception humaine de ce type de sons, aussi bien d'un point de vue descriptif que d'un point de vue hédonique. Nous cherchons donc ici à explorer les différentes méthodologies qui ont été mises au point et utilisées afin d'établir des structures de description et de qualification des sons de l'environnement, valides vis-à-vis de leur perception. Nous souhaitons également étudier les éléments bibliographiques qui abordent la problématique concrète de l'association de ces structures perceptives à des éléments tangibles du « matériel sonore », c'est-à-dire le signal sonore.

Dans un premier temps, il convient de s'intéresser à ce qui définit et ce qui caractérise un son, sans jugement de valeur. Cette notion incite naturellement à aborder la question du *timbre*, véritable signature du son. Cet élément est ainsi l'objet de la section 2.1. La section 2.2 présente un ensemble de travaux portant sur l'identification de la source sonore, composante importante lorsque l'on considère plus particulièrement les sons de l'environnement. La section 2.3 aborde les stratégies et méthodologies employées afin de répondre à la problématique de l'évaluation de la qualité sonore. Enfin, la section 2.4 présente la démarche adoptée à la lumière de cette étude bibliographique dans le cadre de cette thèse.

2.1 Le timbre

2.1.1 Les sons musicaux

Le timbre a servi de point de départ à de nombreuses études sur la qualité sonore, car il définit une sorte de signature acoustique du son. Sa définition initiale vient de la tradition musicale occidentale, où il permet de distinguer intrinsèquement le son de chaque instrument, quelle que soit la note jouée. Avant de définir précisément le timbre, il convient donc de définir ce qui caractérise une note jouée et le son produit. On associe, dans un but de description et d'identification de chaque note jouée par un instrument, trois attributs élémentaires : l'intensité, c'est-à-dire grossièrement la force avec laquelle la note est jouée, la hauteur tonale, c'est-à-dire la note elle-même et la fréquence correspondante (dite *fréquence fondamentale*), et la durée. Ces trois paramètres sont les seuls dont on peut trouver une dénomination précise dans le langage courant et le langage musical, probablement car ce sont les plus faciles à appréhender. Ce sont également ceux que l'on peut trouver sur une partition musicale, qui peut être interprétée indépendamment par différents instruments. Ces trois paramètres ont été

longuement et exhaustivement étudiés et décrits par la psychoacoustique à l'aide notamment des fameuses lois de la psychophysique de Weber [158], Fechner [62] et Stevens [142].

Cependant, ces trois paramètres ne suffisent pas à décrire exhaustivement le son et à distinguer une même note jouée par deux instruments différents. Krumhansl [92] définit ainsi le timbre par « the way in which musical sounds differ once they have been equated for pitch, loudness and duration [1]. » Une définition similaire du timbre a par la suite été standardisée [1] : « Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar [2]. » Le timbre est donc tout ce qui constitue le son et qui n'est ni l'intensité, ni la hauteur tonale, ni la durée. Le timbre est ainsi défini par la négative, c'est-à-dire par ce qu'il n'est pas. Toute la difficulté pour les études menées sur le timbre musical, est d'en donner une définition plus classique, c'est-à-dire identifier l'ensemble des attributs perceptifs le constituant. Ces attributs doivent ainsi permettre de décrire exhaustivement la perception des sons égalisés en intensité, hauteur et durée. Le timbre revêt donc une forme multidimensionnelle et on lui associe souvent un espace, euclidien ou non, appelé *espace perceptif* ou *espace des timbres*, dont les axes constituent les attributs pertinents pour la perception. Ces attributs ont dans un premier temps été étudiés grâce à des sons « de laboratoire » ou choisis de sorte à ne varier que selon un paramètre donné. Ceci permet de maîtriser parfaitement l'attribut considéré, mais n'est pas représentatif de la réalité et ne prend pas en compte la nature multidimensionnelle du timbre.

Les différentiels sémantiques : Afin de remédier à ces critiques, l'exploration de l'aspect multidimensionnel du timbre a été entreprise grâce à l'utilisation d'une technique expérimentale dite de *différentiels sémantiques* introduites en acoustique par Solomon [140]. Le principe est simplement de demander à différents auditeurs d'évaluer les éléments d'un corpus sonore selon diverses échelles sémantiques censées représenter exhaustivement les différentes propriétés du son pouvant être perçues. Chacune de ces échelles est exprimée par un couple d'adjectifs de sens contraires et un curseur qu'il convient de positionner sur une règle graduée ou non à la position souhaitée entre les deux valeurs extrêmes associées à ces adjectifs. Le tableau 2.1 présente un exemple d'échelles de différentiels sémantiques, utilisés dans le cadre d'une étude interculturelle de Kuwano et al. sur la perception des avertisseurs sonores [97].

Cette méthode expérimentale permet donc d'extraire les attributs perceptifs d'un corpus sonore. Cependant, il est nécessaire de remédier aux différents problèmes découlant de l'utilisation d'une terminologie particulière. En effet, le choix des couples d'adjectifs n'est pas anodin. Il est notamment délicat d'obtenir une terminologie exhaustive et univoque, permettant ainsi de décrire la totalité des propriétés du son, indépendamment de l'interprétation des auditeurs. Le choix de la terminologie est donc critique et ne peut être fait qu'expérimentalement afin qu'il soit valide perceptivement. Pour cette raison, on associe souvent la procédure expérimentale des différentiels sémantiques à une étape expérimentale préalable de *verbalisation libre*. On y demande aux auditeurs de décrire avec leurs propres mots les éléments du corpus sonore et d'identifier les critères leur permettant d'effectuer leurs jugements. Une analyse sémantique des descriptions verbales obtenues permet d'extraire un ensemble de termes utilisés par un grand nombre d'auditeurs et dont l'interprétation ne prête pas à confusion. Cette terminologie, qui définit verbalement les attributs de la perception, permet alors d'identifier des échelles à évaluer lors de l'expérience de différentiels sémantiques. Ainsi, l'étape de verbalisation libre permet d'identifier et de valider perceptivement les attributs du son que les audi-

1. « ce en quoi les sons musicaux diffèrent, après avoir été égalisés en hauteur, volume et durée. »

2. « Le timbre est l'attribut de la sensation auditive en terme duquel un auditeur peut juger que deux sons présentés dans les mêmes conditions et ayant les mêmes volume et hauteur tonale sont dissemblables. »

adjective scales	
loud	soft
deep	shrill
frightening	not frightening
pleasant	unpleasant
dangerous	safe
hard	soft
calm	exciting
bright	dark
weak	powerful
busy	tranquil
conspicuous	inconspicuous
slow	fast
distinct	vague
weak	strong
tense	relaxed
pleasing	unpleasing

TABLE 2.1 – Exemple d'échelles de différentiels sémantiques (provenant de Kuwano et al. [97]).

teurs sont capables d'exprimer verbalement, et l'étape de différentiels sémantiques permet d'évaluer quantitativement ces attributs.

Cependant, la terminologie obtenue à l'issue de l'étape de verbalisation libre peut ne pas être exempte de redondance, certains termes employés par les auditeurs n'étant pas nécessairement indépendants entre eux. Il est donc nécessaire d'analyser les différentes échelles et de repérer d'éventuelles relations entre elles afin de définir un ensemble réduit de propriétés quantitatives indépendantes pouvant décrire exhaustivement le corpus sonore. Ceci est réalisé au moyen d'un outil statistique particulier : l'Analyse en Composantes Principales (PCA [98]). Le principe de celle-ci est d'établir, par combinaisons linéaires des échelles mesurées, un ensemble de variables indépendantes – les composantes principales – permettant d'expliquer au maximum la variance de ces échelles. Plus clairement, la technique PCA traite une matrice construite à partir des valeurs de chaque son sur chaque échelle (matrice NxP où N est le nombre de stimuli et P est le nombre d'échelles) et génère une matrice de composantes principales, combinaisons linéaires de ces échelles (matrice PxP) et une matrice des coordonnées des stimuli sur ces composantes (matrice NxP). Les composantes expliquent chacune une part décroissante de la variance des échelles (la 1re explique le maximum de variance, la 2e explique le maximum de la proportion restante, etc.). En pratique, on ne s'intéresse souvent qu'aux 2 ou 3 premières composantes. L'intérêt de cette méthode est de réduire la dimensionnalité de l'espace constitué par les différentes échelles en éliminant les redondances entre celles-ci tout en conservant un maximum de variance expliquée. On obtient donc un espace de faible dimensionnalité (donc aisément représentable) que l'on peut grossièrement associer à l'espace de timbre.

Pour résumer le principe général des différentiels sémantiques, la méthodologie se décompose de la manière suivante :

- étape préliminaire de *verbalisations libres* afin de définir verbalement un ensemble d'attributs sonores permettant de caractériser les éléments du corpus ;
- expérience de *différentiels sémantiques* afin de quantifier ces attributs sonores pour chacun des éléments du corpus ;

– *Analyse en Composantes Principales* afin de traiter les redondances et de dégager 2 ou 3 (le plus souvent) attributs généraux que l'on associe finalement à l'espace de timbre.

Toutefois, cette méthodologie présente quelques défauts. Il est possible, notamment, que certaines caractéristiques, pourtant importantes pour la perception, n'émergent pas d'une telle étude car elles font appel à des notions trop abstraites, ce qui entraîne souvent des disparités dans la façon de les exprimer ; elles peuvent alors être « oubliées » par l'analyse des verbalisations. En effet, les auditeurs non-entraînés utilisent le plus souvent un langage imagé, ou fait d'associations (« ce son ressemble au son produit par tel objet »), afin de décrire verbalement le son et les sensations qu'il provoque. S'il est aisé de trouver des éléments simples de langage permettant de décrire une image (couleur, forme, ...), même abstraite, le langage courant est plutôt pauvre en termes dédiés dés que l'on tente de décrire un son. Les rares éléments de terminologie propres au son sont bien souvent réservés à quelques initiés (musiciens, professionnels du son, ...). Le simple fait d'utiliser une méthodologie sémantique limite alors quelque peu la portée des résultats pouvant être obtenus, du fait de la subjectivité des images et des comparaisons employées.

L'analyse de proximités : Les progrès des outils d'analyse statistiques ont permis de remédier à ces inconvénients et de s'affranchir de l'utilisation d'une terminologie particulière. Ces techniques statistiques permettent l'analyse multidimensionnelle de données de proximité. Ainsi, au lieu de demander aux auditeurs d'évaluer les sons sur des échelles imposées, plus ou moins pertinentes, on leur demande d'évaluer le degré de similarité ou de dissemblance entre les sons – procédure dite de *mesure de similarités*. Ces notions de similarité ou de dissemblance des sons sont relativement concrètes et peu sujettes à des interprétations divergentes. Un des premiers exemples d'étude de timbre exploitant des jugements de proximité que l'on peut trouver dans la littérature est l'étude de Grey [70]. Au cours de ces travaux, il a été demandé à des auditeurs d'évaluer la proximité entre des paires de sons d'instrument de musique. Chaque paire de sons possible parmi un corpus constitué de 16 sons d'instrument de musique a été testée. Il en résulte, pour chaque auditeur, une matrice 16x16 de proximités. Une technique statistique – *MultiDimensional Scaling* (MDS), modèle MDSCAL (voir annexe A) – d'analyse de ces proximités mesurées permet d'obtenir une représentation des distances perceptives résultantes par un modèle euclidien classique. Cette analyse a abouti à un espace à trois dimensions. En observant la façon dont les sons étaient répartis le long de ces trois axes, Grey en a déduit les attributs perceptifs permettant d'expliquer ces variations.

Il n'est toutefois pas inintéressant de noter que les théoriciens distinguent les notions de *similarité* et de *proximité*, la première désignant la ressemblance entre des éléments (e.g. même couleur, dans le domaine de la vision), et la seconde désignant le rapprochement des éléments dans un espace (e.g. positions rapprochées des éléments dans le cas de la vision). Dans le cas du timbre, cette distinction peut être négligée, étant donné que l'on tente de modéliser par l'analyse MDS les distances perceptives – les *similarités* – par les distances dans l'espace des timbres – les *proximités*. On pourrait éventuellement parler de *similarité* lorsque l'on désigne la métrique de distance issue d'une expérience perceptive, et de *proximité* lorsque l'on se réfère à sa modélisation par l'analyse MDS. Par souci de clarté, on parlera plutôt de *distance mesurée* ou *distance perceptive* dans le premier cas, et de *distance modélisée* ou *distance prédite* dans le second.

L'analyse des proximités se résume donc à deux étapes principales :
– expérience de *mesure de similarités*, permettant d'établir le jeu de distances entre les éléments du corpus ;
– analyse *MDS* afin de modéliser ces distances dans un espace géométrique.

La finalité de cette méthode n'est toutefois pas éloignée de celle des *différentiels sémantiques*. Les différences majeures résident dans les points suivants :

- la nature des données recherchées, puis analysées, qui sont ici sous la forme d'un jeu de distances perceptives (matrice NxN, où N est le nombre d'éléments du corpus), contrairement à la méthode des différentiels sémantiques qui a pour but d'obtenir et analyser un ensemble de paramètres caractérisant ces éléments (matrice NxP, où N est le nombre d'éléments, et P le nombre de paramètres) ;
- l'outil statistique employé, qui a ici pour but de modéliser les distances mesurées à l'aide d'un modèle d'espace géométrique, tandis que l'outil de la méthode précédente tente d'établir un jeu de composantes permettant d'expliquer au maximum la variabilité des P paramètres sur le corpus étudié.

L'analyse subséquente n'est ainsi que peu différente pour les deux méthodes, puisque que l'on essaye alors de trouver une interprétation psychologique, d'un côté aux composantes principales pour les différentiels sémantiques, et de l'autre, aux dimensions de l'espace géométrique obtenu par l'analyse des proximités. Il est donc logique que la recherche de *descripteurs explicatifs* par corrélation et régression linéaire soit employée afin de définir précisément l'espace de timbre dans les deux cas.

Les descripteurs explicatifs : Après l'étape de caractérisation perceptive correspondant à l'identification des attributs perceptifs du timbre, que ce soit par la méthode des différentiels sémantiques ou par l'analyse de proximités, il convient de trouver une explication physique aux variations de ces attributs, afin d'établir le lien entre la perception et la physique. On entend ici par explication physique l'identification d'un paramètre calculable ou mesurable à partir du signal sonore [3], variant de manière proportionnelle à l'attribut perceptif considéré, et dont l'interprétation peut expliquer les variations de cet attribut. La relation entre paramètre calculable, que l'on nomme *descripteur*, et attribut perceptif mesuré expérimentalement s'établit aux moyens de la *corrélation* et de la *régression linéaire* (voir annexe B).

L'utilisation de ces outils statistiques a permis notamment à Grey [70] d'identifier les trois paramètres physiques supposés être à l'origine des variations des trois attributs perceptifs (matérialisés par les trois axes de l'espace des timbres obtenu), au travers du corpus de 16 sons instrumentaux étudiés dans son expérience. Il interprète la première dimension comme une échelle différenciant les sons en fonction de la répartition spectrale de leur énergie. Les deux autres dimensions semblaient être liées à des particularités des partiels harmoniques lors de l'attaque (synchronicité, fluctuations, inharmonicités, ...). On trouve également, dans la littérature, d'autres études du timbre des sons d'instruments de musique identifiant un espace perceptif à trois dimensions (notamment Krumhansl [92], McAdams et al. [108], et Marozeau et al. [104]). L'hypothèse sur laquelle repose la démarche adoptée dans le cadre de ces études est suggérée par McAdams [106], qui postule que la reconnaissance des sources sonores (en l'occurrence, les instruments de musique) s'effectue au travers d'un processus d'analyse, d'évaluation et d'extraction d'attributs auditifs liés aux paramètres acoustiques du signal. Usuellement, deux de ces trois dimensions sont expliquées par des paramètres liés respectivement à la répartition spectrale de l'énergie et au temps d'attaque. L'interprétation de la troisième dimension échappe en revanche au consensus, mais est toutefois généralement liée à un paramètre spectro-temporel ou à un paramètre décrivant la structure fine du spectre.

3. Il arrive toutefois également que l'on tente de relier les attributs perceptifs à des paramètres qui ne sont pas calculés sur le signal sonore, notamment lorsque l'on s'intéresse à des paramètres physiques des sources sonores (matériau, forme, dimensions, etc.).

L'étude du timbre des sons musicaux peut donc généralement se résumer en la succession de trois étapes :

- mesure de données perceptives sur les sons du corpus, soit par *verbalisations libres* suivi d'une procédure de *différentiels sémantiques*, soit par *analyse des proximités* ;
- analyse des données perceptives à l'aide d'un outil statistique (*PCA* ou *MDS*) visant à établir l'espace de timbre ;
- recherche de *descripteurs explicatifs* des dimensions de cet espace de timbre, afin de relier la perception à la physique des sons étudiés (cette étape étant toutefois souvent incluse dans l'analyse PCA dans le cas de différentiels sémantiques).

2.1.2 Les sons de l'environnement

Depuis plusieurs années, de nombreuses études ont tenté de transposer ce principe du timbre des sons musicaux à différents types de sons non-musicaux, correspondant à la production sonore induite par les éléments de notre environnement. En effet, l'intérêt croissant des industriels pour le design sonore et l'amélioration du confort acoustique lié à leurs produits a conduit à adopter d'autres démarches de description du son que le simple relevé de niveaux en dB(A). La notion de timbre, donc de description du son selon un continuum multidimensionnel, est apparue comme l'option la plus pertinente pour qualifier un son de produit industriel. Ainsi, la méthode des différentiels sémantiques suivie d'une analyse par la technique PCA, et la méthode de mesure de proximité associée à une analyse MDS (voir annexe A), ont été utilisées à de nombreuses reprises afin d'établir un espace de timbre correspondant à un type particulier de sons de l'environnement.

Susini et al. [145, 146] et McAdams et al. [107] ont étudié le timbre de sons d'habitacles automobiles en appliquant ces techniques de mesure de proximité et de MDS. Deux ensembles de sons étaient composés d'enregistrements stéréophoniques correspondant à deux régimes moteurs. La raison de cette décomposition est que, d'une part, de tels sons présentent une partie harmonique aisément audible, et d'autre part, que les différents régimes moteur vont induire des fréquences fondamentales différentes. Or, la définition du timbre (section 2.1.1) exclut toute variation liée à la hauteur tonale, fortement liée à la fréquence fondamentale d'un son harmonique. De plus, les variations de hauteur tonale, et donc de fréquence fondamentale, tendent à masquer les autres paramètres du timbre [110]. Par ailleurs, suite à une expérience préliminaire, il est apparu que les jugements de proximité étaient fortement liés aux variations d'intensité sonore. Comme la définition du timbre exclut également toute variation d'intensité sonore perceptive, les sons ont été expérimentalement égalisés en sonie au préalable. Les deux ensembles de sons étaient constitués chacun de 16 sons stéréophoniques égalisés en sonie. Une expérience de mesure de similarités a été réalisée pour chaque corpus avec 30 participants. Les deux matrices de proximité ainsi obtenues ont été analysées par la technique CLASCAL (voir annexe A). Cette technique permet de regrouper les participants en *classes latentes* en fonction des similarités de stratégies employées par chacun. Pour le premier corpus, une seule classe latente a été obtenue et un espace à trois dimensions a émergé. Ces trois dimensions ont été respectivement corrélées avec trois descripteurs pertinents : 1) un descripteur caractérisant le rapport d'énergie entre la partie harmonique et la partie bruitée du signal ; 2) le centre de gravité spectral prenant en compte une échelle de fréquence représentée en ERB (*Equivalent Rectangular Bandwidth*, Patterson [125] et Slaney [138]) ; 3) un descripteur quantifiant la décroissance spectrale de la partie harmonique du signal. Pour le second corpus, l'analyse CLASCAL a identifié une seule classe latente également et un espace à deux dimensions. Ces deux dimensions ont été, chacune,

corrélées avec un descripteur pertinent : 1) tout comme pour le premier corpus, un descripteur mesurant le rapport d'énergie entre la partie harmonique et la partie bruitée du signal ; 2) le centre de gravité spectral mesuré sur une échelle de fréquence en pondération C (voir section 2.3.3).

Suivant globalement la même méthode, Susini et al. [147] ont obtenu l'espace de timbre correspondant à un ensemble de sons d'unité de climatisation. Un première expérience de classification menée sur un ensemble initial de 43 sons (trop nombreux pour une expérience de mesure de similarités) a permis d'identifier trois classes de sons dont la discrimination était surtout fondée sur le niveau sonore. Afin d'empêcher la sonie de masquer l'influence d'autres paramètres plus fins, la catégorie de sonie moyenne, formée de 15 sons, a été sélectionnée pour la suite de l'étude. Une expérience préliminaire de mesure de similarités a été menée avec 5 participants afin d'avoir une idée du résultat de l'analyse MDS. Il est alors apparu que ce corpus sonore n'échantillonnait pas l'espace de manière uniforme. Quatre sons synthétiques, obtenus par un algorithme fondé sur l'interpolation linéaire dans l'espace préliminaire, ont été ajoutés au corpus pour compléter les zones « pauvres » de l'espace. Le corpus final est donc constitué de 19 sons monophoniques, non-égalisés en sonie mais correspondant à une gamme de valeurs réduite de ce paramètre. L'expérience principale de mesure de similarités a alors été conduite à l'aide de 50 participants. Une analyse CLASCAL des résultats a mis en évidence cinq classes latentes et un espace à trois dimensions. Ces trois dimensions ont été corrélées respectivement à : 1) un descripteur caractérisant le rapport d'énergie entre la partie harmonique et la partie bruitée du signal ; 2) le centre de gravité spectral de la partie bruitée du signal avec un échelle de fréquence en pondération B (voir section 2.3.3) ; 3) la sonie (voir section 2.3.3), ce qui confirme que même une faible variation de ce paramètre influe de manière importante sur la perception du timbre vis-à-vis des autres attributs perceptifs.

Il est possible de trouver d'autres études adoptant la même démarche associant mesure de proximité et analyse MDS. On peut notamment citer Parizet et al. [118] qui, avec 12 sons stéréophoniques de claquements de portière automobile, a obtenu un espace à trois dimensions avec le modèle INDSCAL (voir annexe A) dont deux ont été corrélées avec respectivement l'acuité définie par Aures (*sharpness* [13], paramètre similaire à un centre de gravité spectral, mais tenant compte d'un modèle auditif) et avec un indicateur de netteté calculé à partir d'une estimation court-terme de la sonie de Zwicker [165]. On peut également signaler l'étude de Lemaitre et al. [100] portant sur les sons de klaxons. Les auteurs, en appliquant la même procédure à un corpus de 22 sons monophoniques égalisés en sonie, ont obtenu un espace à trois dimensions, respectivement corrélées à la rugosité [48], au centre de gravité spectrale incluant un modèle auditif utilisant l'échelle des ERB (voir Marozeau et al. [104]), et à un paramètre lié à à la structure fine du spectre.

Il est important de remarquer que certaines constantes semblent se dégager de ces différentes études sur le timbre des sons de l'environnement. Un attribut semble notamment expliquer systématiquement en partie les jugements de proximité : il s'agit du centre de gravité spectral. Selon l'étude, il apparaît comme associé à une partie du signal et/ou à une description particulière de l'échelle des fréquences, prenant en compte une pondération particulière ou une modélisation plus sophistiquée de l'audition (échelle des Bark ou des ERB, ou acuité d'Aures – voir section 2.3.3 pour plus de détails). Les autres paramètres semblent être dictés par la typicité de certaines familles de sons de l'environnement. Ces observations sont résumées dans l'étude de Misdariis et al. [111] visant à généraliser les principes de description du timbre des sons de l'environnement au moyen d'une méthodologie « méta-descriptive » des corpus correspondant aux quatre études mentionnées précédemment, associant mesure de proximité et analyse MDS. Ces travaux ayant été menés séparément, l'idée de départ de cette étude était de comparer leurs résultats respectifs. De plus, si les descripteurs identifiés

semblent assez proches, il est important de noter que leur calcul fait intervenir de nombreux réglages de paramètres algorithmiques et ceux-ci varient d'une étude à l'autre. Il s'agit donc également de systématiser et standardiser les calculs de descripteur. Ainsi, dans les limites de représentativité des sons de l'environnement que constitue le regroupement des ensembles de sons de ces quatre études, trois espaces de timbres différents ont été identifiés, chacun correspondant à un type particulier de son :

La catégorie Moteur (*Motor*) : cette catégorie correspond au regroupement des sons des études sur les habitacles automobiles [107, 145] et sur les unités de climatisation [147]. L'espace commun obtenu est formé principalement par deux descripteurs :

- un paramètre lié au rapport d'énergie entre la partie harmonique et la partie bruitée du signal : l'*émergence harmonique*. Les parties harmonique et bruitée sont séparées grâce à l'algorithme de séparation de partiels *Pm2* [29], et le calcul consiste en le rapport de sonie (modèle de Zwicker [168]) entre les deux parties.
- un paramètre lié à la répartition spectrale de l'énergie des parties harmonique et bruitée : la *brillance complexe*, consistant en une combinaison linéaire des centres de gravité spectraux et étendues spectrales (incluant un modèle auditif) des deux parties.

La catégorie Pseudo-instrumentale (*Instrument-like*) : cette catégorie est constituée du corpus de l'étude des sons de klaxon [100]. L'espace est donc celui obtenu lors de l'étude originale, à l'exception de la troisième dimension, à présent associée à l'étendue spectrale incluant un modèle auditif.

La catégorie Impact (*Impact*) : cette catégorie correspond au corpus de l'étude sur les claquements de portière [118]. L'espace est similaire à celui de l'étude originale. Toutefois, les calculs de descripteurs ont été standardisés par rapport aux deux autres catégories.

2.2 Identification de la source sonore

L'observation des résultats des études sur le timbre des sons de l'environnement peut amener à s'interroger sur la pertinence d'un tel modèle de description lorsque l'on considère l'environnement dans son ensemble. Prenons un simple exemple issu du paragraphe précédent, afin d'illustrer cette problématique : le son d'habitacle automobile. Lorsqu'on le compare à d'autres sons d'habitacle automobile, l'hypothèse d'une structure de description continue multidimensionnelle semble convenir, et on peut faire émerger un ensemble réduit de descripteurs acoustiques permettant de caractériser ces sons et de les discriminer entre eux. Toutefois, on peut rapidement s'apercevoir que ces descripteurs ne sont pas les mêmes que ceux permettant de distinguer les sons de klaxon, par exemple. Il est donc évident que les espaces obtenus lors des études indépendantes de ces deux types de son ne sont pas compatibles. La question ainsi soulevée est : comment et sur la base de quel type de description commune peut-on comparer deux sons quelconques de notre environnement ?

En réalité, le jugement d'un son dépend grandement du contexte et de ce à quoi on le compare. Lorsque le son d'habitacle automobile est comparé à un autre son d'habitacle automobile, le jugement se fait simplement au travers d'un ensemble d'attributs descriptifs inhérents au son (mat / brillant, harmonique / bruité, ...). Mais lorsque ce son est comparé à un son de claquement de portière, par exemple, un processus cognitif de reconnaissance de la source sonore prend le relai : au lieu de formuler un jugement très descriptif du type « son assez basse fréquence et peu harmonique », par exemple, l'image de l'objet mis à contribution lors de la production du son va se former dans notre esprit, à tel point que la comparaison des deux sons revient à comparer les deux objets, ce qui ne semble pas compatible avec une structure de description continue. Dans notre exemple, le son d'habitacle va

nous faire penser au moteur ou à la voiture en mouvement, tandis que le son de claquement de portière appelle la portière elle-même et exclut, au contraire, tout mouvement de la voiture (les portières étant en principe actionnées à l'arrêt).

Ces observations mettent en évidence la nécessité de différencier différentes types d'écoute que nous utilisons en fonction du contexte. Schaeffer [134], par la suite repris par Chion [47], a proposé un système de description selon trois points de vue :

- l'écoute *causale*, liée à la reconnaissance de la source physique du son ;
- l'écoute *sémantique*, liée à la signification du son, à l'information qu'il fournit (par exemple, la fonction d'avertisseur sonore des klaxons) ;
- l'écoute *réduite*, liée aux caractéristiques inhérentes au son, indépendamment de sa cause et de sa signification.

On peut raisonnablement résumer l'écoute *réduite* au timbre des sons de l'environnement évoqué en section 2.1.2, bien que Schaeffer distingue également d'autres aspects superficiels du son. Par ailleurs, nous laissons volontairement de côté l'écoute *sémantique* qui a plus trait à l'organisation des connaissances qu'à la perception des sons de l'environnement à proprement parler (par exemple un son de moteur automobile et un son de claquement de portière seront regroupés dans une même catégorie « voiture »[4], alors que les sons produits sont radicalement différents d'un point de vue physique ou acoustique). Il convient à présent de s'intéresser à l'écoute dite *causale*.

La thèse de Vanderveer [155] est considérée comme une référence pour la description des sons de l'environnement et notamment le concept d'évènement sonore (*auditory event*). Ses travaux ont notamment indiqué que, lorsque les auditeurs étaient amenés à décrire verbalement un ensemble varié de sons de l'environnement, ils décrivaient rarement le son lui-même, mais plutôt l'évènement l'ayant produit et le contexte environnemental associé. Plus précisément, ils décrivaient 1) l'action, 2) l'objet de l'action et 3) le lieu où l'action intervient. Lors d'une autre expérimentation, elle demandait aux participants de dire s'ils « reconnaissaient » les sons présentés. Il est apparu que les quelques erreurs des participants (c'est-à-dire lorsqu'ils disaient reconnaître un son qu'ils n'étaient pas censés avoir déjà entendu) étaient dues à l'appartenance des sons correspondants à une catégorie commune. Ceci met en évidence l'idée d'une représentation catégorielle des évènements sonores.

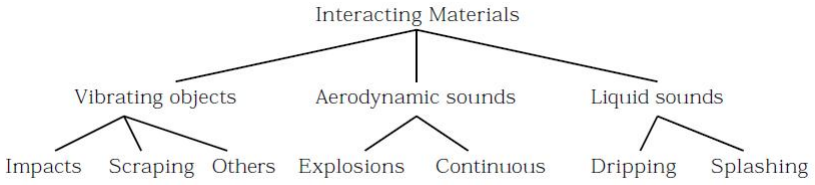


FIGURE 2.1 – Représentation hiérarchique de la taxinomie des évènements sonores proposée par Gaver [67].

Afin de répondre à cette problématique de la classification perceptive des sons de l'environnement, Gaver [67] a proposé une taxinomie des interactions sources de production sonore que l'on peut résumer à la structure hiérarchique en figure 2.1. Cette organisation est fondée sur l'hypothèse stipulant que l'écoute d'un son fournit l'information nécessaire à l'identification de la cause physique en termes d'interactions de matériaux. Cette taxinomie distingue les interactions de types solide, liquide et gazeux. Parmi chacun de ces types d'interactions, les évènements sonores sont regroupés

4. On parle dans ce cas de *méronymie*, c'est-à-dire que le moteur et la portière sont des parties – *méronymes* – de la voiture

en différents types d'actions, voire en différents types d'objets. Gaver considère également les éventuelles interactions entre ces trois catégories (bulles, etc...), et associe à chaque type d'interaction un ensemble de paramètres physiques pertinents (taille, matière, viscosité, force, dureté, ...) permettant d'identifier clairement chaque catégorie. Ces éléments, ainsi que la structure complète sont présentées en figure 2.2.

Ballas, dans une série d'études, a également abordé la problématique de la reconnaissance et de l'interprétation des sons de l'environnement en mettant en relief les similarités avec les principes du langage. Contrairement à Gaver, il considère que l'information fournie par l'écoute du son ne suffit pas à son identification. Selon lui, l'identification est réalisée conjointement par un processus montant (extraction de l'information pertinente du son et du contexte environnemental) et par un processus descendant (utilisation de notre connaissance du monde et de nos attentes) : « It is not only what we hear that tells us what we know ; what we know tells us what we hear [5] » [81]. Par une série d'expérimentation reportées dans [14], il démontre que la performance dans l'identification des évènements sonores dépend de différentes variables, comprenant des paramètres acoustiques, la *fréquence écologique* (la fréquence à laquelle le son a pu être entendu par le passé) et l'*incertitude causale* (correspondant au nombre de possibilités de cause physique du son). Les paramètres acoustiques n'expliquaient qu'environ 50 % de la variance des scores de performance d'identification.

Plus récemment, les travaux de Guyot [73] ont porté sur la classification de sons domestiques. La méthodologie employée dans ces travaux adopte une procédure expérimentale de catégorisation libre qui consiste à demander aux participants de classifier un ensemble de sons selon leurs propres critères et en autant de catégories qu'ils souhaitent. Il leur est également demandé d'expliciter les catégories formées en termes de caractéristiques spécifiques et/ou de critères utilisés. L'hypothèse posée dans ce type d'expérience est très différente de celle posée par les tests exposés précédemment. En effet, ceux-ci supposaient tous une échelle mesurée continue, tandis que ce test suppose une organisation catégorielle des sons. En effet, il a été observé [107] que lorsque les sons étudiés correspondent à des sources sonores facilement identifiables et différentiables, les auditeurs privilégient des critères cognitifs entraînant une classification naturelle des sources sonores, plutôt que la comparaison des sons sur un ensemble de dimensions perceptives continues. Dans un tel cas, une représentation du type espace multidimensionnel continu n'est pas appropriée. On préfère alors une représentation de type hiérarchique. Dans un premier temps, l'étude des verbalisations a montré que les participants adoptaient deux stratégies différentes pour effectuer leur classification. La première est fondée sur des critères facilement associables à des paramètres acoustiques, comme la hauteur tonale, l'évolution temporelle, tandis que la seconde se focalise sur le type d'excitation à l'origine de la production du son (mécanique, électronique, ...). Ceci rejoint les résultats obtenus dans les études précédemment citées et réaffirme l'importance de l'identification de la source sonore dans le cadre de la perception des sons de l'environnement. Les données individuelles de catégorisation prennent la forme d'une matrice dite de *co-occurrences* représentant combien de participants ont placé chaque paire de sons dans une même catégorie. L'analyse statistique des résultats de classification aboutit à une représentation en *arbre additif* [133] reporté sur la figure 2.3. Dans ce type de représentation, la distance perceptive [6] entre deux éléments (aux extrémités des branches) correspond à la longueur du chemin (i.e. la somme des longueurs des segments) allant de l'un à l'autre. Les arbres additifs uti-

5. « Ce n'est pas seulement ce que nous entendons qui nous dit ce que nous savons ; ce que nous savons nous dit ce que nous entendons »

6. Si on considère que la co-occurrence coo entre deux sons est égale au nombre de participants à l'expérience de catégorisation libre les ayant placés dans une même catégorie divisé par le nombre total de participants (donc $0 < coo < 1$), alors la distance perceptive d vaut $d = 1 - coo$.

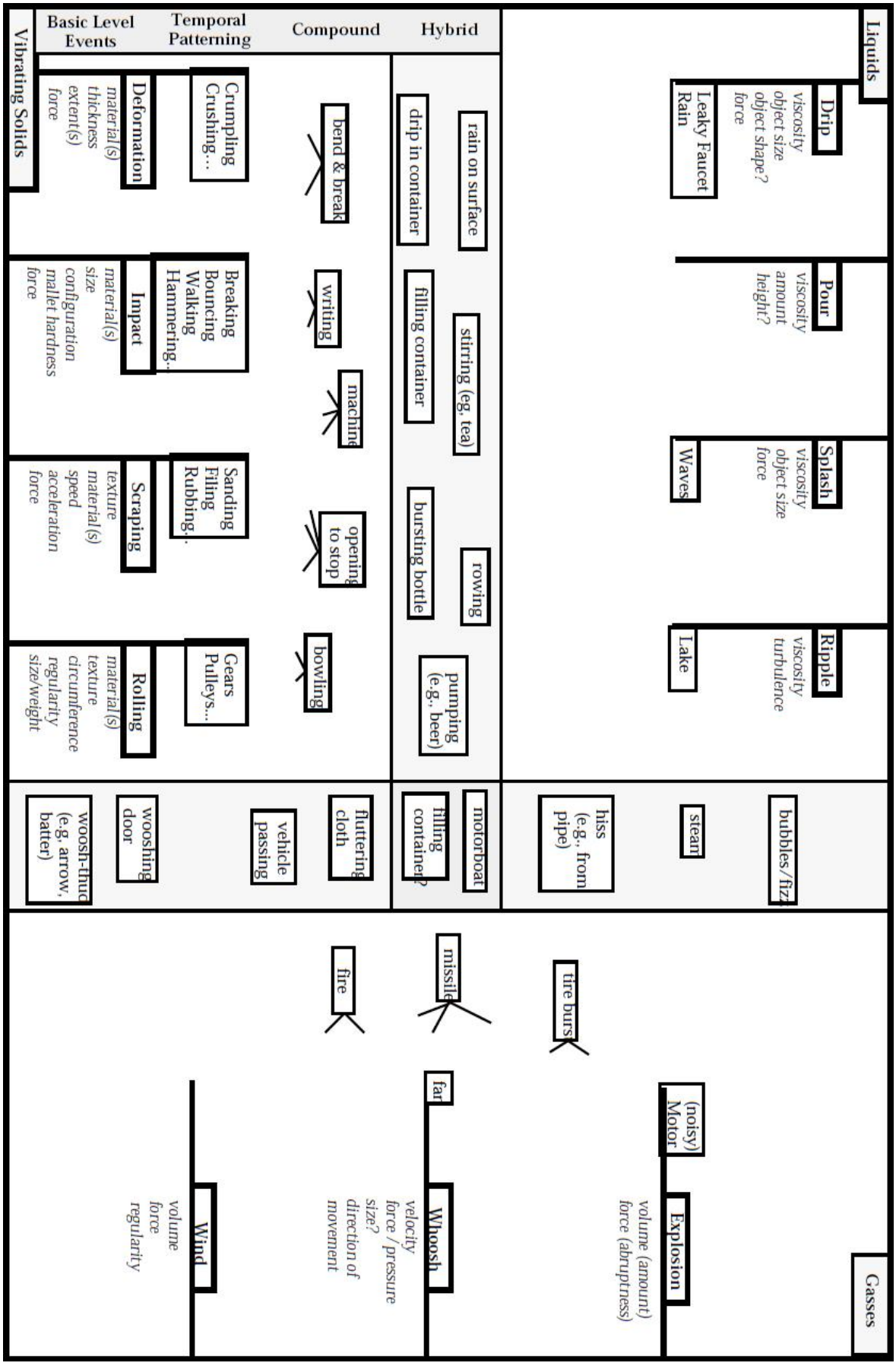


FIGURE 2.2 – Taxinomie des évènements sonores proposée par Gaver [67].

lisent également le concept de *prototype* (représentés par les couples lettre-chiffre sur la figure 2.3), introduit par Rosch [131] : un prototype est un élément abstrait d'une catégorie qui est à la fois le plus proche des éléments de cette catégorie et le plus éloigné des éléments des autres catégories. Les arbres additifs permettent donc de représenter à la fois l'organisation hiérarchique des catégories moyennes et la *typicité* des sons au sein d'une même catégorie.

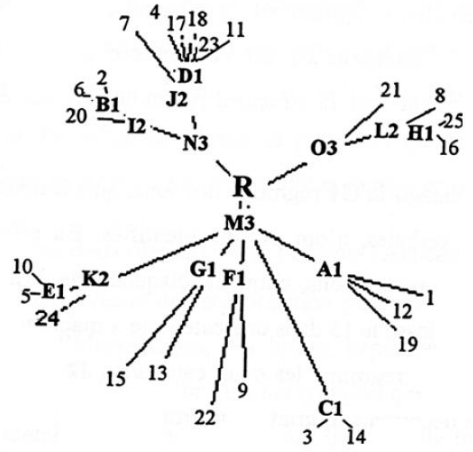


FIGURE 2.3 – Représentation en arbre additif des résultats de classification obtenus par Guyot [73]. Les numéros représentent les sons du corpus, et les lettres (suivies d'un numéro), les prototypes.

Enfin, l'étude déjà évoquée en section 2.1.2 de Misdariis et al. [111] expose une structure prédictive de description visant à associer dans une même organisation perceptive les notions d'identification et de classification en types de source sonore et l'aspect timbral du son (structure multidimensionnelle continue). L'idée majeure provient de l'hypothèse selon laquelle la discrimination des sons sur la base d'un espace multidimensionnel continu n'est perceptivement valide que si l'ensemble des sons considérés appartiennent à une même catégorie de sons de l'environnement et donc, en d'autres termes, au même type de source physique. La pertinence de cette hypothèse peut se retrouver dans une expérience de mesure de similarités de sons correspondant à des types de source très hétérogènes, menée par McAdams et al. [107] ; l'analyse MDS appliquée à ce corpus de sons a révélé une structure perceptive fortement catégorielle, montrant l'inadéquation de ce mode de représentation de la perception pour les sons étudiés. Outre les travaux de généralisation des espaces de timbre, déjà évoqués en section 2.1.2, les auteurs ont également mené une expérience de catégorisation libre sur l'ensemble des sons des études prises en compte. Les résultats ont été analysés par une méthode de partitionnement de données (*unweighted arithmetic average clustering – UPGMA* [98]) permettant d'obtenir une représentation du corpus sonore en *arbre ultramétrique* ou *dendrogramme* (représentation hiérarchique, non-additive). Misdariis et al. ont ainsi pu établir, dans les limites de la gamme de sons étudiés (habitacles automobiles, unités de climatisation, klaxons, claquements de portière automobile), une structure en deux niveaux :

1. un niveau catégoriel identifiant le type de source sonore (*Motor*, *Impact*, *Instrument-like*).
2. un niveau continu correspondant à l'espace de timbre obtenu pour chacune des ces 3 classes.

Du propre aveu des auteurs, cette décomposition n'est valide que dans le cadre des types de sons étudiés et n'est pas représentative de l'ensemble des sons de l'environnement. En effet, elle n'est exhaustive ni en termes de catégories de sources sonores, ni en termes de précision de ces catégories,

en ce sens qu'elle ne fait pas apparaître le caractère hiérarchique de la classification perceptive des sources sonores (un seul niveau est représenté).

La formulation des différents types d'écoute de Schaeffer [134] mentionnée précédemment résume assez bien la problématique générale abordée dans cette section et qui concerne la description des sons de l'environnement dans son ensemble :

- Lorsque l'ensemble étudié constitue un corpus homogène de sons issus de sources physiques similaires (écoute *réduite* chez Schaeffer), une structure de description de type *multidimensionnel* et *continu* – tel que l'espace de timbre (voir section 2.1) – semble être adéquate. Elle convient donc parfaitement à la présente étude.
- En revanche, lorsque le corpus de sons considéré inclut des éléments correspondant à des types différents de cause physique du son, un processus de reconnaissance de source préempte la perception des attributs auditifs. Il en résulte qu'une structure de description *catégorielle* et *hiérarchique* permet de représenter plus fidèlement la perception du corpus (écoute *causale* chez Schaeffer).
- Enfin, lorsque les sons de l'ensemble considéré présentent des relations perceptives de type sémantique (écoute *sémantique* chez Schaeffer), nos connaissances et notre conception de l'environnement en général impliquent des structures perceptives de description plus complexes tels que les réseaux sémantiques développés dans le domaine de la linguistique (voir par exemple le réseau WordNet [63])

Dans le cadre de cette thèse, la description de type sémantique ne présente que peu d'intérêt, puisque le but est ici de relier la perception à des éléments physiques liés au signal ou à la source sonore. Par ailleurs, bien qu'il soit possible de relier la description perceptive de type causal à des éléments tangibles du signal ou de la source sonore – comme démontré, entre autres, par l'étude de Misdariis et al. [111] –, le fait que les sons considérés ici correspondent à un type de production sonore similaire incite à également laisser de côté ce type de description perceptive. Les méthodologies employées dans ce dernier cadre ne sont toutefois pas inintéressantes d'un point de vue pratique, car elle permettent d'identifier de manière pertinente des familles de sons parmi de grands ensembles. Nous nous intéresserons donc principalement, dans le cadre de cette thèse à la description liée à l'écoute *réduite*.

2.3 Évaluation de la qualité sonore

2.3.1 Procédures expérimentales de mesure de la qualité sonore

L'hypothèse de départ pour la caractérisation hédonique est que la qualité sonore peut être vue comme un attribut de la perception, au même titre que les attributs du timbre (section 2.1), mais diffère toutefois de ces derniers par son caractère complexe et subjectif. D'un point de vue théorique, il semble naturel à priori de considérer la qualité sonore comme une échelle unidimensionnelle continue, que l'on note U_i pour le son i, et qu'il convient de déterminer expérimentalement.

Il est également important de noter que les procédures expérimentales de mesure perceptive font intervenir des évaluations effectuées par un certain nombre d'auditeurs. Selon l'hypothèse d'une échelle de qualité sonore unique, chaque auditeur est considéré comme un instrument de mesure fluctuant. L'intérêt de faire intervenir un certain nombre d'entre eux est d'obtenir une estimation statistique des fluctuations (c'est-à-dire des variations entre auditeurs), et donc une mesure perceptive plus précise. Il convient alors de vérifier que ces fluctuations ne sont pas trop grandes par rapport

à la gamme de variation de l'échelle perceptive mesurée. Dans un cadre statistique plus rigoureux, cela consiste à évaluer le coefficient de concordance de Kendall et le coefficient de corrélation de rang de Spearman (voir section C.1 en annexe) qui permettront de conclure sur la validité statistique de la mesure effectuée. Cette méthode peut s'appliquer à l'ensemble des procédures expérimentales évoquées dans cette partie.

En pratique toutefois, il s'avère souvent que le caractère fortement subjectif de la qualité sonore empêche de faire émerger une échelle unique de description. En effet, les jugements de qualité sonore des auditeurs ne sont pas toujours influencés au même degré par les différents attributs auditifs associés au type de son étudié. Le cas échéant, ces différentes sensibilités pénalisent notablement les coefficients mentionnés ci-dessus, rejetant ainsi la validité statistique d'une échelle moyenne. Il convient alors d'analyser plus précisément la distribution des jugements des auditeurs, afin d'identifier les principales tendances parmi le panel de participants (un exemple de méthode d'analyse utilisée dans cette optique est présentée en section C.2 en annexe). Au lieu d'une échelle de qualité sonore unique, on tente alors d'obtenir autant d'échelles que l'on identifie de subdivisions du panel. Il est alors en principe possible de valider les échelles obtenues en évaluant de nouveau les coefficients de concordance de Kendall et de corrélation de rang de Spearman sur chacune des subdivisions du panel. La littérature montre ainsi plusieurs exemples où l'étude de la qualité sonore associée à un objet donné a abouti à une telle segmentation du panel d'auditeurs participant aux expériences (voir notamment Susini et al. [147] et Parizet et al. [119]).

Ordonnancement : La manière la plus simple d'aborder expérimentalement la problématique de la quantification de la qualité sonore est de demander à des auditeurs d'ordonner un ensemble de sons en fonction de la sensation de qualité qu'ils procurent. Cette méthode est celle employée par Fastl pour l'étude de la qualité sonore de rasoirs électriques [59], et par Patsouras et al. pour l'évaluation de la qualité de sons de moteurs automobiles [122]. L'échelle globale peut alors s'établir en observant la répartition des classements de chaque son sur l'ensemble des auditeurs, voire se calculer comme le classement moyen des sons. Mais la précision de l'échelle obtenue par une telle méthode n'est pas assurée.

Estimation de grandeur : Une autre méthode très simple est de demander à des auditeurs d'affecter à chaque son une valeur sur une échelle numérique, proportionnellement à leur sensation. Les valeurs de l'échelle de qualité U_i sont alors simplement ces estimations moyennées sur l'ensemble des auditeurs. À l'origine, cette méthode a surtout été utilisée pour l'estimation de la sonie (voir section 2.3.3). Dans certains cas, on choisit de fournir aux auditeurs un son de référence, auquel on attribue une valeur arbitraire. Ainsi, si un son procure une sensation deux fois plus forte que le son de référence, l'auditeur lui attribue la valeur double de la valeur de référence. Une échelle de rapport de sensation est ainsi obtenue. Si elle paraît très simple à mettre en œuvre, elle présente tout de même un fort inconvénient : le choix du son de référence peut apporter un biais aux résultats du test. En effet, il se peut, dans le cas où l'on s'intéresse à des attributs abstraits du son, comme la qualité, que le choix du son de référence focalise l'attention des auditeurs sur des caractéristiques propres à ce son. Cette méthode a toutefois été largement utilisée pour des études perceptives fondamentales (avec des sons de laboratoires parfaitement contrôlés) mais plus rarement adoptée lorsque l'on s'intéresse à des sons réels.

Évaluation absolue : Une variante de la méthode d'estimation de grandeur est l'évaluation absolue, dont la principale différence est que l'échelle proposée aux auditeurs n'est plus numérique mais

identifié grâce aux labels (souvent un couple d'adjectifs ou expressions sémantiquement opposés) placés aux extrémités de l'échelle. Dans le cas du caractère hédonique du son, on peut par exemple placer aux extrémités « pas du tout agréable » et « extrêmement agréable ». L'échelle peut également être discrète et présenter ainsi plusieurs paliers associés chacun à un label. L'échelle ainsi obtenue est donc absolue, contrairement à celle de l'estimation de grandeur de nature relative. Cela peut ne pas convenir pour certains ensembles de sons. Par exemple, si l'on s'intéresse à des sons correspondant au même type de source physique, suffisamment proches pour qu'il soit possible d'en établir un espace de timbre (voir section 2.1), on peut s'attendre à ce que tout ou partie des évaluations soient concentrées sur une faible portion de l'échelle, rendant les différences de jugements moyens entre sons non-significatives d'un point de vue statistique. Afin de pouvoir apprécier le caractère significatif des résultats, on utilise souvent un outil statistique dit d'*analyse de variance*, permettant de discriminer les cas où les différences sont porteuses de signification et ceux où elles sont plus probablement dues à la chance, et donc simplement à des fluctuations aléatoires de leur perception par les auditeurs.

Comparaison par paire : Le principe de cette méthode est fondé sur l'idée qu'il est plus facile de comparer deux sons et de choisir celui que l'on préfère que d'évaluer séparément les sons sur une échelle, numérique ou non, plus ou moins arbitraire, et qu'il est parfois difficile de définir précisément. On considère donc ici l'échelle unidimensionnelle de qualité comme une *échelle de préférence* qu'il convient d'établir pour le corpus de sons étudié. En pratique, on fait écouter aux auditeurs toutes les paires de sons différents qu'il est possible de constituer à partir du corpus de sons étudié. Pour tester une paire de sons, on présente à l'auditeur les deux sons diffusés successivement, et l'auditeur doit simplement choisir celui qu'il préfère. Il est également possible de ne pas limiter le nombre de réponses possibles à deux, et d'offrir des choix intermédiaires (par exemple, « les deux sons sont équivalents »). Dans le cas où l'on souhaite ne tester qu'une seule fois chaque paire de sons, on va donc tester $n(n-1)/2$ paires, pour un ensemble de n sons. Il arrive également que l'on teste chaque paire de sons dans les deux sens de présentation (son i puis son j, et son j puis son i), afin de vérifier que l'ordre de présentation des sons n'a pas d'influence. Dans ce cas, on doit présenter $n(n-1)$ sons. Quelques précautions doivent toutefois être prises. Bien entendu, l'ordre de présentation des paires de sons doit être en principe aléatoire. On prend toutefois garde à ce que les apparitions successives d'un même son ne soient pas trop rapprochées l'une de l'autre. En effet, si un même son était successivement confronté à plusieurs autres, il est possible que des caractéristiques propres à celui-ci prennent le pas sur d'autres, plus générales, et biaisent les jugements des auditeurs. Enfin, on veille également à ce que l'ordre de présentation des paires ne soit pas le même d'un auditeur à l'autre.

L'inconvénient majeur de cette méthode est d'ordre pratique. En effet, le nombre de jugements à effectuer par l'auditeur est proportionnel au carré du nombre de sons. Par conséquent, lorsque le corpus sonore devient d'une taille importante, la longueur de l'expérience peut devenir excessive. Ceci peut avoir pour conséquence de rapidement lasser, voire fatiguer, les auditeurs, dont les réponses peuvent devenir ainsi peu précises et peu cohérentes. En pratique, il est difficile d'appliquer cette méthode pour des ensembles de plus d'une quinzaine de sons (bien que cette limite dépende aussi de la longueur des sons et de l'ergonomie de l'interface de test).

Lors du traitement des résultats, les résultats de chaque auditeur sont placés dans une matrice NxN dans laquelle la case (i, j) indique la préférence de l'auditeur entre les sons i et j (0 si le son i est préféré, 1 si le son j est préféré, ou valeur intermédiaire si plus de deux réponses sont possibles). Bien entendu, on affecte également la valeur complémentaire dans la case (j, i), afin de compléter

la matrice. Il suffit alors de moyenner cette matrice sur l'ensemble des auditeurs pour obtenir un jeu de probabilité de préférence P_{ij} pour chaque paire de sons. Étant donné le jeu de probabilité de préférence P_{ij} entre les sons, il s'agit maintenant d'en déduire une valeur reflétant le potentiel de chaque son à être préféré aux autres. En conséquence, il est indispensable de poser l'hypothèse qu'il existe un continuum de sensation selon lequel on peut affecter à chaque son une valeur U_i représentant le degré de préférence. Les probabilités de préférence P_{ij} sont reliées à ces valeurs par une fonction f inconnue :

$$P_{ij} = f(U_i, U_j) \tag{2.1}$$

La fonction f doit bien entendue être définie. Plusieurs modèles existent. Parmi ceux-ci, 3 sont particulièrement utilisés dans ce type d'étude expérimentale : le modèle linéaire (somme des préférences pour chaque son), le modèle Thurstone V [150], utilisé notamment par Susini et McAdams pour étudier les jugements de préférence de sons d'habitacle automobile [107, 145] ou de sons d'unités de climatisations [147] et le modèle BTL (Bradley-Terry-Luce [40, 103]), utilisé par exemple par Zimmer et al. [163] pour établir une échelle de désagrément de divers sons de l'environnement.

Évaluation comparée : Cette méthode [46] associe évaluation directe et comparaison, et est inspirée des méthodes introduites indépendamment par Bodden et al. [28] et Maunder [105]. La tâche à effectuer par l'auditeur consiste toujours à évaluer les sons, mais l'auditeur a la possibilité de réécouter et réévaluer l'ensemble des sons tout au long de l'expérience, tout en visualisant en permanence l'ensemble des réglages effectués. En pratique les échelles d'évaluation correspondant aux différents sons sont présentées simultanément à l'écran, tel que sur la figure 2.4 par exemple. Il a été démontré [119] que cette méthode offre un bon compromis entre précision de l'échelle obtenue et durée de l'expérience. En effet, il apparaît que cette procédure nécessite une durée d'expérience plus courte qu'avec la méthode de comparaison par paire tout en donnant des résultats plus précis qu'avec l'évaluation absolue ou l'estimation de grandeur. Ceci est dû au fait qu'il est donné aux auditeurs la possibilité de corriger leurs évaluations initiales à la suite de l'écoute et de l'évaluation de l'ensemble du corpus sonore. En revanche, cette méthode peut parfois inciter les auditeurs à se contenter d'un ordonnancement des sons plutôt qu'à une évaluation précise. De plus, si le nombre de sons est élevé, l'interface peut s'en trouver d'autant plus surchargée. Elle laisse en revanche la possibilité d'opter pour une échelle absolue ou relative, en fonction des labels utilisés et des consignes fournies aux auditeurs. Le traitement se fait comme pour l'évaluation absolue ou l'estimation de grandeur, et s'avère donc moins « lourd » que pour la comparaison par paire.

2.3.2 Méthodes composées pour l'évaluation de la qualité sonore

La question qui se pose à présent est : « Quelle procédure expérimentale utiliser et comment associer la mesure de qualité obtenue à des descripteurs acoustiques ? » Pour le choix du type de test, tout dépend du contexte et des objectifs de l'expérience : grandeur à mesurer, nombre de stimuli, durée admissible de l'expérience, etc. Bien souvent, dans un but de précision, plusieurs procédures expérimentales sont associées entre elles. C'est en ce sens que l'on parle ici de « méthode composée ». La première a alors comme objectif de définir les attributs perceptifs pertinents pour les sons considérés. La seconde tente d'associer ces attributs à l'information recherchée.

On peut tout d'abord classer de nombreuses utilisations des différentiels sémantiques dans les

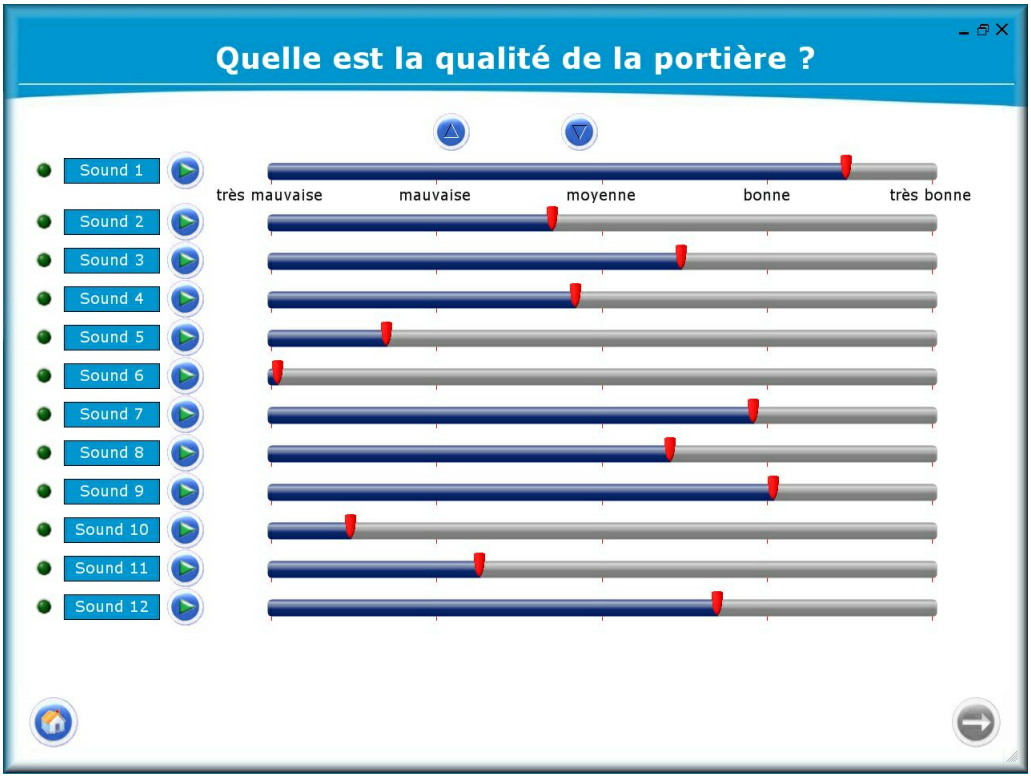


FIGURE 2.4 – Exemple d'interface pour un test d'évaluation comparée (provenant de Chevret et Parizet [46]).

méthodes composées. En effet, comme expliqué en section 2.1.1, cette méthode des différentiels sémantiques est souvent précédée d'une expérience de verbalisations libres afin de définir les échelles sémantiques à évaluer et qui représentent les attributs verbaux pertinents pour la perception. Par la suite, on conduit l'expérience de différentiels sémantiques en y associant également une échelle de qualité dont la dénomination varie d'une étude à l'autre (« gênant », « agréable », « désagréable », « plaisant », etc...). L'analyse par la technique PCA est alors appliquée à l'ensemble de ces échelles, échelle de qualité comprise, mais également à un ensemble de descripteurs acoustiques susceptibles d'être corrélés à certaines de ces échelles, et dont les expérimentateurs soupçonnent qu'ils expliquent les jugements de qualité. Les composantes principales issues de cette analyse permettent donc d'identifier les attributs verbaux, ainsi que les descripteurs explicatifs, associés à la qualité sonore. Toutefois, outre les inconvénients de la méthode des différentiels sémantiques, déjà évoqués en section 2.1.1 – nécessité de présupposer des attributs pertinents pour la perception des sons étudiés, et utilisation d'un vocabulaire dont l'exhaustivité et l'univocité ne sont pas assurées –, on peut également objecter que ce type de méthodologie mélange échelles descriptives et caractérisation hédonique (qualité sonore) d'une part, et évaluations perceptives et descripteurs du signal ou de la source sonores d'autre part. En effet, aucune distinction n'est faite entre ces échelles lors de l'Analyse en Composantes Principales, et elles sont considérées par l'algorithme comme des paramètres d'entrée de même nature. De plus, les résultats issus d'une telle méthodologie sont plus d'ordre informel, et ne permettent pas toujours d'établir un indicateur numérique fiable de la qualité sonore.

Par ailleurs, l'utilisation des techniques psychoacoustiques classiques de mesure de qualité sonore, évoquées en section 2.3.1, soulève les mêmes problèmes que dans le cas de l'étude du timbre. Par exemple, l'utilisation des procédures d'estimation de grandeur et d'évaluation absolue (qui se retrouve par ailleurs dans la méthode des différentiels sémantiques), voire d'évaluation comparée

passe également par une explication verbale aux auditeurs de la tâche à effectuer, et notamment une définition précise et univoque de l'attribut du son qu'ils doivent évaluer. Or, comme expliqué en section 1.2.3, la définition du terme « qualité » peut varier d'une étude à l'autre en fonction de l'objectif visé, et, de manière générale, échappe au consensus. De plus, on trouve dans la littérature de nombreux attributs perceptifs dont le lien avec la qualité perçue a été établi, bien souvent à l'aide de sons de laboratoire parfaitement maîtrisés mais non-représentatifs de l'ensemble des sons de l'environnement. Rien ne garantit que les indicateurs de qualité établis pour des sons de laboratoire, souvent synthétiques, soient aussi performants pour expliquer les jugements pour les sons de l'environnement. En effet, à trop limiter le nombre de paramètres acoustiques variant sur un corpus de sons synthétiques, on pose l'hypothèse que ces paramètres sont effectivement pertinents pour les jugements de qualité des auditeurs, et, dans le même temps, on incite involontairement les auditeurs à évaluer les sons sur les échelles perceptives correspondant à ces paramètres plutôt que sur celle de la qualité, ce qui ne permet pas de confirmer l'hypothèse posée.

Une autre approche a été développée afin de répondre à ces critiques en associant étude du timbre et étude des préférences. Elle a été souvent employée ces dernières années pour l'étude de la qualité sonore de produits industriels (Susini et McAdams [145, 107] sur les habitacles automobiles, Susini et al. [147] sur les unités de climatisation, Parizet et al. [118] sur les claquements de portières automobiles, ...). Cette technique est expliquée de manière plus générique par Susini et al. [146]. Son originalité provient de l'idée que l'étude de la qualité sonore passe par l'identification initiale des paramètres pertinents pour la perception des sons considérés. En effet, si l'on cherche à identifier les paramètres acoustiques pertinents pour expliquer la qualité perçue, il paraît logique de les trouver parmi ceux permettant de les différencier, dans un espace de timbre en l'occurrence. Si un paramètre s'avère avoir peu d'importance pour les jugements de similarité, il ne peut, en toute logique, pas avoir plus d'importance pour les jugements de qualité. Ainsi, les procédures expérimentales adoptés sont : mesure de similarités (voir section 2.1) et comparaison par paire.

Dans une étude de la qualité sonore, il importe, afin de relier la mesure de qualité effectuée à la description physique du son, d'identifier un ensemble de descripteurs acoustiques permettant d'expliquer les variations de qualité perçue. Or, les descripteurs existants ne manquent pas et la possibilité de tomber sur un descripteur corrélé de manière fortuite n'est pas négligeable. En d'autres termes, une bonne corrélation ne signifie pas nécessairement que le descripteur en question est un bon prédicteur de l'échelle mesurée, surtout si l'on a testé beaucoup de descripteurs – on peut facilement dépasser la centaine – sur un nombre faible de stimuli, souvent entre 10 et 20 pour une expérience de comparaison par paire. De plus, il est fort possible que l'échelle mesurée soit en théorie liée à plusieurs attributs perceptifs. En conséquence, la variance de l'échelle serait partiellement expliquée par chacun des descripteurs correspondant à ces attributs. Il est alors fort probable que le coefficient de corrélation de chaque descripteur ne soit pas significatif. Certaines méthodes statistiques, telle que la régression linéaire multiple (voir section 2.3.3), permettent d'expliquer une variable dépendante (ici l'échelle mesurée) par un ensemble de variables indépendantes (ici les descripteurs). Toutefois, il est nécessaire de nouveau d'avoir au préalable identifié les quelques descripteurs pertinents (2 ou 3 tout au plus), compte tenu du nombre de combinaisons possibles de descripteurs.

Pour résumer, l'étude du timbre (expérience de mesure des similarités) permet d'obtenir l'espace perceptif des sons du corpus étudié, et donc les attributs qui émergent de leur perception, et l'étude des préférences (expérience de comparaison par paire) relie les attributs de cet espace à la qualité sonore. Le gros avantage de cette méthode est que les procédures expérimentales utilisées sont très peu « verbeuses » : les seules questions posées aux auditeurs sont aussi simples et neutres

que « À quel point les deux sons sont-ils semblables ? » (pour l'étude du timbre) et « Quel son préférez-vous ? » (pour l'étude des préférences). Cette neutralité assure que les auditeurs ne seront pas influencés par la terminologie utilisée. De plus, les valeurs mesurées sont assez précises et la méthodologie est statistiquement valide, car on ne fait aucune hypothèse de départ sur les descripteurs audio associés à la qualité sonore.

En revanche, les deux procédures expérimentales utilisées présentent une certaine « lourdeur » de mise en œuvre pour deux raisons principales : d'une part, elles font intervenir des jugements par paire, ce qui rend les expériences à la fois longues et limitées en nombre de stimuli étudiés ; et d'autre part, cette simplicité/neutralité des questions posées rend ces procédures assez rébarbatives, ce qui peut entraîner un manque d'implication des auditeurs et indirectement allonger également la durée des expériences.

Enfin, une autre méthodologie provient, à l'origine, d'une technique souvent utilisée dans l'industrie agro-alimentaire : l'*analyse sensorielle* [109]. Cette technique a depuis peu été transposée au domaine de l'évaluation de la qualité sonore, notamment dans le domaine de l'industrie automobile. Comme les méthodes précédemment évoquées, elle implique 1) une étape visant à fournir une description de la perception d'un ensemble de produits : l'*analyse descriptive* aboutissant à un ensemble de *profils sensoriels* ; et 2) une étape dont le but est d'évaluer les préférences des utilisateurs sur cet ensemble de produits : l'*analyse hédonique*. L'originalité de cette méthode réside dans les procédures expérimentales particulières pour chacune de ces deux étapes.

Pour l'étape d'analyse descriptive, l'hypothèse posée est qu'il est possible d'obtenir une mesure de sensation objective à partir d'un panel d'auditeurs pouvant être considéré dans son ensemble comme un instrument de mesure. Cette hypothèse nécessite de rejeter tout jugement de préférence ou d'ordre hédonique, souvent subjectif, et implique que les auditeurs concernés soient « experts », c'est-à-dire entraînés par une phase préliminaire d'apprentissage. Cette phase d'apprentissage a pour but d'entraîner les auditeurs à exprimer verbalement leurs sensations de manière claire, et à éventuellement prêter attention à certaines sensations dont ils ne sont pas conscients naturellement. En revanche, cette « expertise » des auditeurs permet de réduire le panel d'auditeurs entre 10 et 20 environ pour obtenir les profils sensoriels.

Pour l'étape d'analyse hédonique, on s'intéresse cependant aux utilisateurs finaux. L'attitude vis-à-vis des auditeurs concernés est donc très différente. En effet, on préfère alors des auditeurs « naïfs », c'est-à-dire n'ayant pas bénéficié d'un apprentissage spécifique pour la tâche à accomplir. En conséquence, le panel d'auditeurs doit être bien plus nombreux (un minimum de 60 auditeurs est requis) et être représentatif de la population visée. Par ailleurs, les auditeurs de la première phase ne peuvent participer à cette phase, puisque, ayant bénéficié de l'apprentissage, ils ne sont plus représentatifs des utilisateurs finaux. Les auditeurs sont interrogés cette fois-ci uniquement sur leurs préférences.

Les procédures correspondant à ces deux étapes sont régies par des règles standardisées pour la sélection et l'apprentissage des participants [9] et pour le traitement de l'analyse descriptive et l'établissement des profils sensoriels [6]. Il existe en revanche peu de publications dans ce domaine, ce qui s'explique par l'orientation industrielle de l'application de cette méthode, et par les contraintes de confidentialité qui en découlent. On peut toutefois citer la thèse de Bézat [42] portant sur la perception des sons de claquement de portière automobile, la récente étude de Bergeron et al. [18] portant sur les sons émis par le roulement d'automobiles sur différents revêtements routiers, ou l'étude, plus ancienne, réalisée sur les sons d'unités de climatisation de Siekierski et al. [137], parallèlement à l'étude, évoquée précédemment, de ces mêmes sons par la méthode associant étude du timbre et étude des préférences de Susini et al. [147]. L'étude de Siekierski et al. a permis d'identifier un en-

semble de 12 descripteurs verbaux permettant de décrire exhaustivement les profils sensoriels correspondant aux sons étudiés qu'ils ont reliés aux jugements de préférence par une analyse en composantes principales. Par la suite, comme elles portaient sur le même corpus de sons, les résultats obtenus par les deux méthodes (analyse sensorielle [137], et étude du timbre et des préférences [147]) ont été comparés par Junker et al. [89]. Il est apparu que les résultats des deux études étaient similaires et que les deux méthodes étaient à peu près aussi efficaces. On constate tout de même que si la l'analyse sensorielle donne des informations plus complètes que l'étude du timbre et des préférences, les résultats semblent en revanche plus difficiles à interpréter.

En résumé, on peut distinguer trois grands types de méthodologie de la l'étude de la qualité sonore :

- une approche *descriptive* ou *verbale*, qui emploie souvent des procédures de verbalisations libres et de différentiels sémantiques, et des Analyses en Composantes Principales ;
- une approche *comparative*, qui emploie souvent des procédures de mesure de similarités et de comparaisons par paires, et des outils de MDS et de modélisation des préférences ;
- une approche d'*analyse sensorielle*, méthode intermédiaire des deux précédentes, de part le fait qu'elle fait intervenir des verbalisations mais dont la décomposition analyse descriptive / analyse hédonique ressemble plus à celle de l'approche *comparative*.

Bien entendu, certaines approches peuvent diverger de ces trois méthodologies, notamment dans les procédures et analyses employées, mais celles-ci correspondent plus à des grandes lignes directrices pour l'étude de la qualité sonore.

2.3.3 Descripteurs acoustiques associés à la qualité sonore

Nous ne considérons dans cette partie que les descripteurs acoustiques pertinents pour des sons de nature stationnaire et entretenue. Par conséquent, nous mettons volontairement de côté les paramètres liés à l'évolution lente du son, c'est-à-dire à l'enveloppe temporelle (e.g. temps d'attaque, durée effective, centre de gravité temporel, ...). Également, nous ne nous intéressons pas ici à l'influence de la hauteur tonale. En effet, celle-ci correspond au percept d'une « note de musique » d'une certaine fréquence. Cette notion est bien entendu établie pour des sons musicaux, mais semble moins pertinente pour les sons environnementaux. En effet, ceux-ci présentent peu ou pas d'élément qualifiable de musical (c'est-à-dire une partie harmonique). D'autres peuvent en présenter, mais il arrive souvent que la fréquence fondamentale correspondante, c'est-à-dire la raie harmonique de plus basse fréquence et/ou celle dont toutes les autres sont des multiples, soit trop basse pour être perçue en tant que hauteur tonale (le son est alors plutôt perçu comme inharmonique). Pour ces raisons, il ne semble pas pertinent de prendre en compte ce paramètre dans l'étude de la qualité acoustique des sons environnementaux, son influence n'étant, dans ce cas précis, pas continue. Seuls la sonie et les attributs principaux du timbre (brillance, émergence harmonique et rugosité/force de fluctuation, explicitées ci-dessous) sont donc en général abordés lors de l'étude de la qualité acoustique des sons environnementaux.

L'intensité sonore : L'intensité sonore constitue l'attribut du son le plus aisément appréhendable. Également désigné par « volume » ou « force » sonore, il est en revanche délicat de la quantifier par une mesure physique représentative de sa perception.

L'élément de mesure physique de base du son est la pression acoustique (instantanée) $p(t)$. Étant donné qu'il s'agit d'une grandeur oscillatoire, un moyen simple d'en estimer l'amplitude est d'en éva-

luer la valeur efficace p_{eff} exprimée en *Pascal* (Pa). Cette valeur correspond au calcul de la moyenne quadratique (ou valeur *RMS – Root Mean Square*) de la pression :

$$p_{eff}(T) = \sqrt{\frac{1}{T}\int_0^T p^2(t)dt} \quad \text{où } T \text{ est la période considérée} \tag{2.2}$$

Lorsque l'on parle de pression acoustique, on fait souvent référence à cette mesure plutôt qu'à la pression instantanée. Par ailleurs, la perception de l'oreille humaine couvre une gamme de pression acoustique très importante, en effet il est admis que la plus petite pression acoustique audible est de $p_0 = 2.10^{-5}$ Pa tandis que le seuil de douleur se trouve aux alentours d'une centaine de Pa. Cette gamme de variation « utile » met en évidence l'inadéquation de la pression acoustique en tant que mesure du percept d'intensité sonore. On préfère alors utiliser une échelle logarithmique du rapport de la pression mesurée p_{eff} sur la pression de référence p_0. Cette échelle est exprimée en Bel (B), et plus spécifiquement en acoustique, en *décibel SPL* (*Sound Pressure Level*, dB SPL), afin de définir le niveau acoustique :

$$L = 20\ \log\left(\frac{p_{eff}}{p_0}\right) \quad \text{en dB SPL} \tag{2.3}$$

Une autre métrique souvent évoquée est le niveau équivalent, noté L_{eq}. D'après sa définition, il s'agit du niveau de pression acoustique d'un son stationnaire ayant la même énergie acoustique qu'un son non-stationnaire, pendant un temps donné.

$$L_{eq} = 10\ \log\left(\frac{1}{T}\int_0^T \frac{p^2(t)}{p_0^2}dt\right) \quad \text{en dB SPL} \tag{2.4}$$

Néanmoins, il s'agit en réalité de la même métrique, obtenue par le même calcul, que le niveau acoustique évoqué au-dessus [7]. Seul le contexte impose l'utilisation d'une notation ou de l'autre : le niveau acoustique L est une mesure « instantanée » – incluant toutefois une intégration temporelle (voir équation 2.2), mais sur une courte période T où le signal de pression est supposé stationnaire – tandis que le niveau équivalent L_{eq} est utilisé pour estimer le niveau acoustique sur une certaine durée T plus longue d'un signal à priori non-stationnaire.

Toutefois, la sensibilité fréquentielle de l'oreille humaine n'est pas uniforme. Il est bien souvent admis que le spectre audible correspond à la bande de fréquences entre 20 et 20 000 Hz. De plus, entre ces limites, le système auditif ne réagit pas de la même manière sur toute la bande passante ; les fréquences extrêmes de cette plage sont notamment fortement atténuées. Les courbes de pondération fréquentielle ont été introduites pour permettre de décrire plus fidèlement cette sensibilité fréquentielle. Leur rôle consiste à pondérer le spectre du signal sonore de façon à lui donner grossièrement la forme compensatoire de la distorsion apportée par le système auditif. En pratique il existe plusieurs courbes de pondération (A, B, C et D – figure 2.5). Les pondérations A et C sont les plus couramment utilisées, et sont notamment prisées pour les normes de bruit. La pondération A permet de décrire la sensibilité spectrale des sons de faible niveau (inférieur à 55 dB SPL), tandis que la pondération C correspond à des sons de plus forte amplitude (supérieur à 85 dB SPL).

7. Cette formule s'obtient en effet en combinant les équations 2.2 et 2.3.

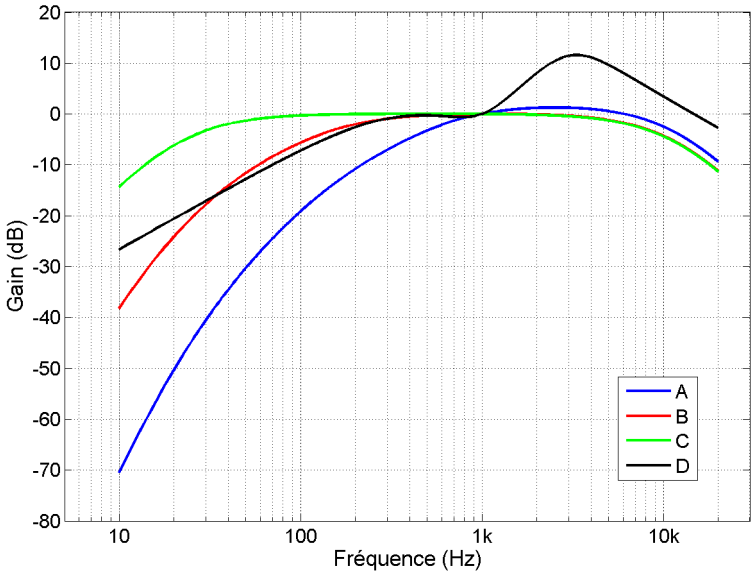


FIGURE 2.5 – Courbes de pondération fréquentielle A, B, C et D.

Ces courbes de pondérations sont fondées sur les *courbes d'isosonie* (Figure 2.6) décrivant, en fonction de la fréquence, le niveau acoustique nécessaire afin qu'il soit perçu comme équivalent à celui d'un son pur à 1000 Hz.

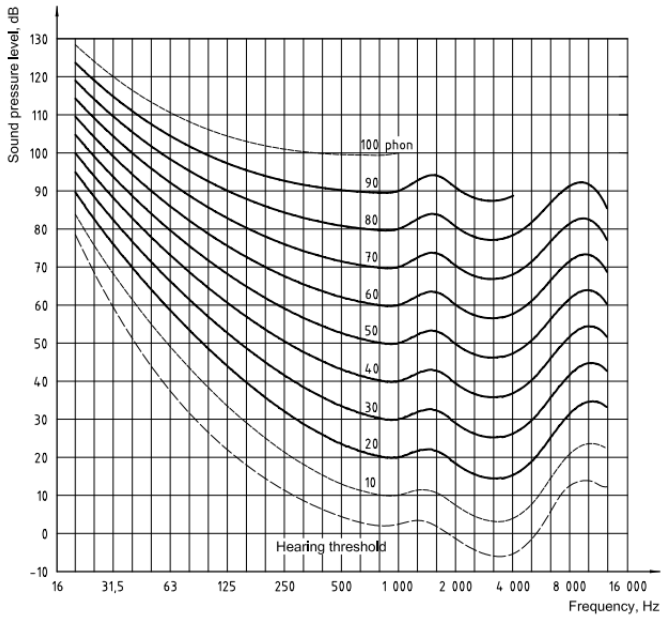


FIGURE 2.6 – Courbes d'isosonie [7] (norme révisée des courbes initialement publiées par Robinson et Dadson [130]).

Ces niveaux équivalents ont également permis de définir une métrique reflétant mieux la perception de l'intensité sonore, le *niveau d'isosonie*, exprimé en phones, noté P, et se définissant ainsi : la valeur de P phones correspond au niveau d'isosonie d'un son perçu comme aussi intense qu'un son pur de P dB SPL à 1000 Hz. Par ailleurs, il a été a démontré que, pour un son pur, une augmentation

de 10 dB induit un doublement de l'intensité perceptive. Le niveau d'isosonie répond donc aux deux définitions suivantes :

1. 40 phones est le niveau d'isosonie d'un son pur 40 dB SPL à 1000 Hz ;
2. un son de $P+10$ phones est perçu comme deux fois plus fort qu'un son de P phones.

On définit également la *sonie*, notée N et exprimée en *sones*, en apportant une légère modification à la seconde définition, de sorte à obtenir une relation linéaire entre sonie et sensation d'intensité sonore :

1. 1 sone correspond à l'intensité perceptive d'un son pur de 40 dB SPL à 1000 Hz ;
2. un son de $2N$ sones est perçu comme deux fois plus fort qu'un son de N sones.

Le niveau d'isosonie est donc l'expression logarithmique (à base 2) de la sonie, et leur relation s'exprime par :

$$N = 2^{\left(\frac{P-40}{10}\right)} \tag{2.5}$$

et la relation réciproque :

$$P = 10\ \log_2(N) + 40 \tag{2.6}$$

À partir de cette définition théorique, plusieurs modèles de calcul de la sonie ont été développés [168, 112]. La particularité de ces modèles est de fonder l'estimation de la sonie sur le codage de l'information sonore par le système auditif humain. Le descripteur psychoacoustique défini par Zwicker et Fastl [168], appelé également sonie [8] (*loudness*), a fait l'objet d'une norme internationale [8]. Il s'inspire du comportement mécanique de la membrane basilaire, entraînant notamment le masquage de certaines fréquences d'un son complexe par d'autres, les basses fréquences présentant un pouvoir masquant plus important que les hautes. Il se peut donc que certaines composantes du spectre fréquentiel ne participent pas à la sensation sonore. Un élément important de ce descripteur est lié à la notion de *bande critique*, formalisée notamment par Zwicker [164]. Ces bandes, au nombre de 24, exprimées en « Bark », correspondent à la largeur fréquentielle d'une bande de bruit [9] en-dessous de laquelle, à niveau sonore global constant, la sonie ne varie pas. Au-delà de cette largeur limite, la sonie augmente avec la largeur de bande. C'est ce que l'on appelle la *sommation de sonie*. L'oreille humaine se comporte donc comme un banc de filtres dont les bandes passantes correspondent aux bandes critiques. La correspondance entre Barks et bandes fréquentielles (fréquence centrale et largeur) est donnée dans le Tableau 2.2.

Zwicker et Fastl ont alors pu introduire la notion de sonie spécifique qui s'exprime en sone par Bark, représentant donc le niveau perçu dans chaque bande. Après avoir appliqué au signal un filtre modélisant les résonances de l'oreille externe et moyenne, la sonie spécifique N_z est calculée dans

8. Le terme *sonie* fait à la fois référence au modèle de calcul et au percept d'intensité sonore.
9. On parle souvent de *bruit à bande étroite*.

N° bande de Bark	Fréquence centrale (Hz)	largeur de bande (Hz)	N° bande de Bark	Fréquence centrale (Hz)	largeur de bande (Hz)
1	50	100	13	1850	280
2	150	100	14	2150	320
3	250	100	15	2500	380
4	350	100	16	2900	450
5	450	110	17	3400	550
6	570	120	18	4000	700
7	700	140	19	4800	900
8	840	150	20	5800	1100
9	1000	160	21	7000	1300
10	1170	190	22	8500	1800
11	1370	210	23	10500	2500
12	1600	240	24	13500	3500

TABLE 2.2 – Fréquences centrales et largeur des bandes de Bark.

chaque bande critique z (en Bark), comme une fonction de l'excitation E exprimée en unité de puissance. L'excitation est obtenue à partir du niveau de sortie des filtres de chaque bande critique, permettant ainsi de prendre en compte le masquage fréquentiel. La sonie globale N en sones est alors obtenue en sommant la sonie spécifique sur toutes les bandes critiques :

$$N = \sum_{1}^{24 \text{ barks}} N_z \tag{2.7}$$

Un second modèle de sonie est couramment utilisé. Il s'agit de celui de Moore et al. [112], normalisé en 2007 [2]. Il est proche de celui de Zwicker [8], et suit globalement le même schéma. Il diffère principalement par son filtre de l'oreille externe et moyenne, la décomposition des bandes de fréquence du spectre (où il introduit la notion d'ERB – *Equivalent Rectangular Bandwidth* –, remplaçant celle des bandes critiques), et le mode de calcul de l'excitation.

L'importance du percept de la sonie est évidente à la lecture de la littérature concernant l'évaluation de la qualité sonore. En effet, de nombreuses études ont mis en évidence que la sonie à elle seule expliquait une part considérable de la variance des jugements de qualité, de gêne ou de désagrément [19, 20, 166, 96], au point que certains auteurs assimilent mesure de sonie et mesure de gêne ou de désagrément [159].

L'émergence harmonique : L'émergence harmonique est un descripteur dont l'importance est souvent établie dans les études de timbre portant sur des sons de systèmes faisant intervenir un moteur. Les sons de ce type ont la particularité de présenter deux parties facilement discernables par l'oreille humaine :

- une partie harmonique, liée au fonctionnement cyclique du moteur, dont le spectre présente des raies harmoniques ou inharmoniques (non multiples de la fréquence fondamentale) et qui peut facilement être modélisée par une somme de sinusoïdes ;
- une partie bruitée pouvant être générée par exemple par des turbulences aériennes résultant du fonctionnement du moteur (dans l'exemple d'un son à l'intérieur d'un habitacle automobile, il s'agit du bruit de roulement et du bruit aérodynamique).

On trouve différentes études abordant la quantification de l'émergence harmonique, prenant souvent le nom de tonalité – *tonality* – ou de force tonale – *pitch strength* –, voire *Harmonic-to-Noise Ratio*, *Tone-to-Noise Ratio* ou *Prominence Ratio*. Comme l'indiquent certaines de ces dénominations, l'évaluation de ce percept passe par la séparation des parties harmonique et bruitée. Les algorithmes permettant cette opération peuvent être plus ou moins efficaces, et utilisent souvent des techniques de traitement du signal dites de « suivi de partiels » (*partial tracking*) relativement complexes ([29, 56] entre autres).

L'émergence harmonique vise alors à évaluer l'énergie relative de la partie harmonique et de la partie bruitée du signal. Elle se traduit souvent par un rapport d'énergie ou d'amplitude (noté ici HNR pour *Harmonic-to-Noise Ratio*). On utilise souvent le niveau RMS (*Root Mean Square*) du signal. Les courbes de pondérations (A, souvent) sont également utilisées pour renforcer la significativité perceptive du paramètre :

$$HNR = \frac{RMS_{A,h}}{RMS_{A,n}} \tag{2.8}$$

où $RMS_{A,h}$ et $RMS_{A,n}$ sont respectivement les niveaux RMS
en pondération A des parties harmonique et bruitée du signal.

Il est de nouveau possible d'augmenter la précision perceptive de ce descripteur en utilisant la sonie [168] comme estimateur d'amplitude :

$$HNR = \frac{N_h}{N_n} \tag{2.9}$$

où N_h et N_n sont les sonies des parties harmonique et bruitée du signal.

L'importance perceptive de ce paramètre a été observée à plusieurs reprises, notamment dans les études du timbre et les études de la qualité sonore et des préférences pour des sons de moteurs. Les études déjà évoquées portant sur les habitacles automobiles [107, 145] et sur les unités de climatisations [147] en sont les principaux exemples. Misdariis et al. [111] ont utilisé efficacement le dernier calcul cité pour expliquer la dimension du timbre correspondant à cet attribut perceptif pour trois groupes de sons de moteur de différents types.

La brillance : Ce que nous appelons « brillance » ici correspond au percept associé à l'équilibre entre basses et hautes fréquences. Elle décrit donc la manière selon laquelle l'énergie du signal sonore est répartie sur l'échelle des fréquences. Ce percept, bien que prépondérant pour la description des sons environnementaux et presque systématiquement identifié sous une forme ou sous une autre dans les études de timbre, est difficile à appréhender de prime abord. En conséquence, il est parfois nommé de différentes manières (« acuité », « clarté », ...etc) et associé à différents descripteurs. La plupart d'entre eux sont des descripteurs d'enveloppe spectrale. La forme la plus simple est le centre de gravité spectral SC (*Spectral Centroid*), calculé sur le spectre discret du signal :

$$SC = \frac{\sum_i f_i \cdot a_i}{\sum_i a_i} \tag{2.10}$$

avec f_i, les fréquences et a_i, les amplitudes du spectre à ces fréquences.

Le *SC* correspond donc au moment statistique d'ordre 1. Le moment d'ordre 2 peut également avoir une signification perceptive. Il correspond à « l'étendue spectrale » *SS* (*Spectral Spread*), et décrit d'une certaine façon la largeur du spectre :

$$SS = \frac{\sum_i (f_i - SC)^2 \cdot a_i}{\sum_i a_i} \tag{2.11}$$

Les moments d'ordres supérieurs existent également, mais sont moins pertinents pour la perception. Ces deux descripteurs sont souvent déclinés sous d'autres formes dans le but de décrire plus fidèlement la sensation. On leur associe souvent par exemple les pondérations fréquentielles (figure 2.5), appliquées sur les a_i, afin de prendre en compte la sélectivité fréquentielle de l'oreille humaine. Dans le but de rendre ces descripteurs les plus fidèles possibles, on retrouve également des versions utilisant la sonie spécifique définie par Zwicker et Fastl [168], *PSC* et *PSS* (*Perceptual Spectral Centroid/Spread*) :

$$PSC = \frac{\sum_z f_z \cdot N_z}{\sum_z N_z} \tag{2.12}$$

$$PSS = \frac{\sum_z (f_z - PSC)^2 \cdot N_z}{\sum_z N_z} \tag{2.13}$$

où les f_z sont les fréquences centrales des bandes de Bark,

et les N_z sont les sonies spécifiques, correspondant à ces bandes.

Ces calculs peuvent également prendre en compte séparément les parties harmoniques et bruitée du signal (voir paragraphe précédent) si cette séparation est pertinente pour les sons étudiés.

Une autre version sophistiquée, assez proche de la métrique *PSC*, également basée sur le calcul de sonie spécifique, est nommée « acuité » – *sharpness*. Ce paramètre a été développé par Von Bismarck [156], puis par Aures [13] et Zwicker et Fastl [168], et inclut également un modèle perceptif. L'acuité se note souvent *S* et s'exprime en *acum*. La valeur de référence de 1 acum correspond à l'acuité d'un son à bande étroite centrée autour de 1 kHz, à 60 dB SPL. Sa formulation est assez proche de celle du *PSC*, la principale différence provenant de l'utilisation d'une fonction de pondération g, traduisant l'influence plus importante des hautes fréquences :

$$S = 0,11 \cdot \frac{\sum_z z \cdot g(z) \cdot N_z}{\sum_z N_z} \tag{2.14}$$

Dans cette formulation de l'acuité selon Zwicker et Fastl [168], la forme de la fonction de pondération g est donnée en figure 2.7.

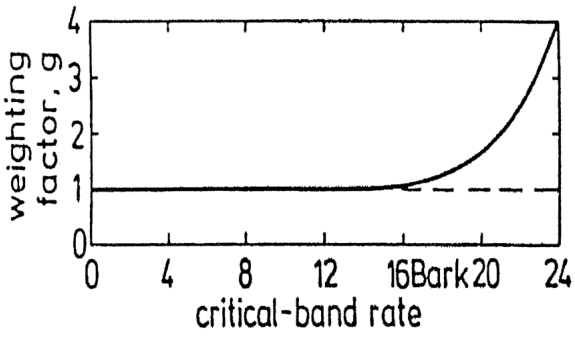


FIGURE 2.7 – Fonction de pondération g pour le calcul de l'acuité (provenant de Zwicker et Fastl [168]).

Un autre descripteur, ou plutôt groupe de descripteurs, a été développé afin de décrire fidèlement la perception de l'enveloppe spectrale d'un son. Il s'agit des MFCC (*Mel Frequency Cepstral Coefficients* [129]). Le principe de ce modèle est assez proche de celui de Zwicker et Fastl, c'est-à-dire un filtre de l'oreille externe et moyenne et un spectre divisé en bandes – les Mels – dont le nombre peut varier selon l'application, et dans lesquelles l'énergie du signal est estimée. En revanche, l'avantage de ce modèle est qu'il est beaucoup moins gourmand en calcul que celui de Zwicker, ce qui le rend très prisé dans le domaine de la reconnaissance de la parole.

Le percept de brillance semble être celui, après la sonie, qui se retrouve le plus souvent dans les études de qualité sonore comme critère de préférence. Comme déjà évoqué en section 2.1.2, cet attribut se retrouve presque systématiquement dans les études du timbre, quelle que soit la nature des sons considérés [111]. Il semble toutefois difficile d'en trouver un unique descripteur acoustique explicatif, comme le prouve la grande variété de descripteurs acoustiques de l'enveloppe spectrale que l'on peut trouver dans la littérature, selon les applications.

La rugosité et la force de fluctuation : La rugosité – *roughness* – et la force de fluctuation – *fluctuation strength* – définies par Terhardt [148] sont des descripteurs quantifiant les modulations d'amplitude du signal sonore. Lorsqu'un son présente une amplitude variant dans le temps de manière régulière, sur une dynamique importante, et à une fréquence relativement faible (inférieure à 100 Hz), ces variations ne sont pas perçues par l'oreille humaine comme un phénomène fréquentiel mais bien comme un phénomène temporel. Toutefois, ces variations peuvent produire deux percepts distincts. En dessous d'une dizaine de Hz de fréquence de modulation, l'oreille parvient à suivre les variations temporelles d'amplitude, et on parle alors de force de fluctuation. Au dessus, lorsque la fréquence de modulation atteint notamment quelques dizaines de Hz, l'oreille ne peut plus suivre les variations dans le temps, et le son prend alors un aspect « rugeux », tel que l'on obtient lorsque l'on frotte un objet contre un mur crépis, ou avec un sifflet à bille. On parle dans ce cas de rugosité.

La rugosité globale R est exprimée en *asper*. La valeur de 1 asper correspond à la rugosité d'un son pur de fréquence 1 kHz, modulé à 100 % (c'est-à-dire sur toute l'amplitude) à 70 Hz. Plusieurs modèles de calcul de la rugosité et de la force de fluctuation existent [13, 168, 48]. Souvent, la rugosité est calculée en estimant un indice de modulation à la sortie de chaque filtre auditif (les bandes de Bark par exemple). Cet indice de modulation permet donc d'établir ce que l'on appelle la rugosité

partielle – *partial roughness*. La rugosité se calcule souvent comme la somme des rugosités partielles R_i [10]. On évalue donc à la sortie de chaque filtre auditif la fréquence de modulation $f_{\text{mod } i}$ et la profondeur de modulation m_i en calculant l'enveloppe temporelle du signal. La rugosité partielle est proportionnelle au produit de la fréquence et de la profondeur de modulation $f_{\text{mod } i} \cdot m_i$:

$$\begin{aligned} R_i &= K \cdot f_{\text{mod } i} \cdot m_i \\ R &= \sum_i R_i \end{aligned} \tag{2.15}$$

où les K est le coefficient de proportionnalité.

La force de fluctuation est quant à elle le plus souvent estimée directement à partir du signal temporel (voir notamment le modèle de Zwicker et Fastl [168], fondé sur l'estimation de la profondeur de masquage temporel [11]). Elle se note souvent F et s'exprime en *vacil*. La valeur de 1 vacil correspond à la force de fluctuation d'un son pur de fréquence 1 kHz, modulé à 100 % à 4 Hz.

Ces deux paramètres sont identifiés comme critères perceptifs importants dans de nombreuses études du timbre et de la qualité sonore. Parmi celles-ci, on peut notamment citer l'étude du timbre des sons de klaxons de Lemaitre [99, 100] ou les travaux sur la « dieselitude » d'un moteur automobile de Patsouras et al. [122]. Il faut toutefois noter que le percept de rugosité rejoint des considérations d'inharmonicité (fréquences des harmoniques non multiples de la fréquence fondamentale) et de dissonance, qui n'ont de sens que lorsque l'on considère des sons au moins partiellement de nature harmonique, comme les sons de moteur, et dont la partie harmonique n'est pas trop masquée par le bruit généré par des turbulences aériennes par exemple.

Métriques composées de la qualité sonore : L'importance des descripteurs acoustiques mentionnés précédemment en tant qu'indicateurs pertinents est régulièrement établie dans les études de la qualité sonore. Cependant, à l'exception de la sonie, il est rare qu'ils expliquent à eux seuls toute la variabilité des jugements de préférence observés. Bien souvent, afin de tester la pertinence d'un descripteur en tant qu'indicateur d'une grandeur mesurée perceptivement, on utilise la régression linéaire entre le descripteur et la mesure. Le score de corrélation obtenu permet, en fonction de différentes considérations statistiques, de confirmer ou d'infirmer la pertinence du descripteur. En revanche, si l'on suppose que plusieurs attributs indépendants expliquent conjointement la variance de la mesure, il est probable que chaque attribut n'explique qu'une part de la variance observée. Ceci amènera irrémédiablement des scores de corrélation statistiquement non-significatifs avec les descripteurs correspondants. Ces descripteurs peuvent alors être écartés par erreur.

Il convient alors d'étudier d'une manière ou d'une autre la possibilité d'expliquer la variance de la mesure par plusieurs descripteurs à la fois. Ils sont donc souvent combinés par un procédé qui a pour but de modéliser l'influence conjointe de ces paramètres sur la qualité perçue : la régression linéaire multiple. Le principe de cette méthode est globalement le même que celui de la régression li-

10. Une notable exception est le modèle de Daniel et Weber [48], qui prend également en compte un coefficient correcteur liée à la corrélation entre bandes de fréquence.

11. Le masquage temporel correspond à la capacité d'un son à en masquer un autre intervenant juste avant ou juste après lui, sur des intervalles de temps très faibles (quelques dizaines de ms tout au plus).

néaire. Cette dernière tente de modéliser la relation mesure/descripteur par une fonction de la forme $f(x) = ax + b$. De manière similaire, la régression linéaire multiple vise à prédire la mesure par une combinaison linéaire d'un ensemble réduit de descripteurs, dont le nombre est intimement lié au nombre de points de mesures (au nombre de sons, en l'occurrence). Ceci aboutit à une relation du type $f(x) = \sum_i a_i x_i + a_0$, pour un ensemble de descripteur $x = \{x_i\}$. Il est alors également possible d'évaluer l'efficacité du modèle en calculant un coefficient de corrélation entre la mesure et le « méta-descripteur » $\sum_i a_i x_i$.

Du fait de sa simplicité de mise en œuvre, l'usage de cette méthode est souvent mentionné dans la littérature, ne serait-ce que pour le cas d'études préliminaires. En effet, elle permet souvent d'établir une première approximation de la relation mesure/descripteurs. En conséquence, cette méthode est celle employée par Khan et Dickson [91] pour l'étude la qualité sonore de chargeurs sur pneus. C'est également la méthode adoptée par Ellermeier et al. dans une série d'étude portant sur le désagrément provoqué par différents sons de l'environnement (naturels, industriels ou domestiques) [54, 55, 163] aboutissant à des relations linéaires entre la mesure et la sonie, l'acuité et la rugosité. Ces derniers ont notamment appliqué cette méthode à un ensemble de sons variés de notre environnement quotidien, et à ces mêmes sons traités par la méthode dite de « *spectral broadening* ». Cette méthode, développée par Fastl [60], consiste grossièrement en la reconstruction des signaux à partir uniquement de leurs enveloppes spectrale et temporelle, de sorte à leur ôter leur « identifiabilité ». La méthode donne de bons résultats pour les sons traités, pour lesquels la régression linéaire multiple sur la sonie, l'acuité et la rugosité permet d'expliquer 86 % de la variance des jugements de préférence. En revanche, la même méthode ne permet pas d'expliquer plus de 75 % de la variance pour les sons réels.

Cependant, cette méthode est relativement « dangereuse » si elle n'est pas appliquée avec précaution. En effet, le fait de multiplier le nombre de descripteurs utilisés pour la régression augmente rapidement la probabilité d'obtenir un bon score de corrélation de manière fortuite. Il est par exemple possible d'obtenir un bon score de corrélation en appliquant cette méthode afin de prédire un vecteur de valeurs aléatoires (et donc, par définition, imprédictible) par un ensemble suffisamment grand de descripteurs. En conséquence, l'obtention d'un bon score de corrélation par cette méthode ne garantit pas la significativité de la relation mesure/descripteurs. En pratique, on limite d'ailleurs le nombre de descripteurs à 2 ou 3, pour un nombre de points de mesure de quelques dizaines d'éléments. Ces limitations nécessitent parfois de tester la robustesse du modèle, en reproduisant par exemple l'expérience sur d'autres sons, et en observant la stabilité de la corrélation donnée par le modèle.

L'hypothèse de linéarité de la relation entre mesure de qualité/gêne/désagrément sonore et descripteurs n'étant en général pas vérifiée, d'autres modèles plus complexes de prédiction de la qualité sonore basée sur plusieurs descripteurs ont été développés. Parmi ceux-ci, le plus ancien est le modèle quadratique de Terhardt et Stoll [149], cité par Geissner et Parizet [69] dans une étude portant sur l'évaluation longue durée de sons de véhicule de livraison. Ce modèle est basé sur les calculs de rugosité de Terhardt [148] R et d'acuité de von Bismarck [156] S. Leur modèle de désagrément W^- est ainsi défini par :

$$W^- = S^2 + 0,16R^2 \tag{2.16}$$

Par la suite, ce modèle est enrichi d'un nouveau terme : T^- quantifiant le manque de caractère tonal (c'est-à-dire l'inverse de la tonalité/émergence harmonique évoquée plus haut), entre 0 et 1 :

$$W^- = \sqrt{S^2 + 0,25R^2 + 0,1(T^-)^2} \tag{2.17}$$

En revanche, La grandeur T^- n'est pas calculée par un modèle particulier mais a été établie expérimentalement.

Un autre modèle non-linéaire est le modèle de gêne non-biaisée défini par Zwicker [167] (*unbiased annoyance UBA*, voir section 1.2.3). Il a construit ce modèle en prenant en compte les attributs perceptifs que sont la sonie N, l'acuité S et la force de fluctuation F. Un terme correctif d, lié à la période de la journée, est aussi considéré, ce modèle ayant initialement été conçu pour quantifier les nuisances sonores dues au trafic routier. Selon Zwicker, la rugosité n'a pas d'influence prépondérante et l'effet de l'émergence harmonique est pris en compte dans la sonie. La sonie instantanée est calculée et la valeur retenue n'en est pas une moyenne, mais le seuil qui est dépassé 10 % du temps, la sonie au 10e centile : N_{10}. L'échelle ainsi mesurée à pour référence la gêne provoquée par un son pur de fréquence 1 kHz et à 40 dB SPL : 1 *au* pour *annoyance unit*. La gêne non-biaisée s'exprime alors par :

$$UBA = d \cdot (N_{10})^{1,3} \left(1 + 0,25(S-1) \cdot \log(N_{10}+10) + 0,3F \cdot \frac{1+N_{10}}{0,3+N_{10}}\right) \tag{2.18}$$

$$\text{avec } d = \begin{cases} 1 & \text{de 6 h à 20 h} \\ 1 + \left(\dfrac{N_{10}}{5}\right)^{0,5} & \text{de 20 h à 6 h} \end{cases}$$

On peut également citer le modèle d'agrément sensoriel – *sensory pleasantness* – mentionné par Zwicker et Fastl [168]. Dans ce modèle, une relation exponentielle est supposée entre l'agrément P et les attributs perceptifs de sonie N, d'acuité S, de rugosité R et d'émergence harmonique T. Par conséquent, par souci d'homogénéité d'équation, les grandeurs sont ici considérées comme relatives, c'est-à-dire par rapport à des valeurs de références, respectivement P_0, N_0, S_0, R_0 et T_0. Le modèle est alors établi par la relation suivante :

$$\frac{P}{P_0} = e^{-0,7\ R/R_0}\ e^{-1,08\ S/S_0}\ (1,24 - e^{-2,43\ T/T_0})\ e^{0,023\ N/N_0} \tag{2.19}$$

Le modèle plus récent, souvent repris dans la littérature est le modèle de gêne psychoacoustique – *Psychoacoustic Annoyance PA* – de Widmann [160], également cité dans les plus récentes versions de l'ouvrage de Zwicker et Fastl [168] et par Fastl dans l'ouvrage de référence de Blauert [61]. La gêne psychoacoustique PA se mesure, comme la gêne non-biaisée exposée précédemment, en *au* – annoyance unit. Elle est définie, en fonction de la sonie au 5e centile N_5 (valeur que la sonie instantanée dépasse 5 % du temps), de l'acuité S, de la force de fluctuation F et de la rugosité R, par la formule suivante :

$$PA = N_5 \left(1 + \sqrt{w_S^2 + w_{FR}^2}\right) \tag{2.20}$$

où w_S est la contribution de l'acuité :

$$w_S = \begin{cases} 0 & \text{pour } S \leq 1,75 acum \\ (S-1,75) \cdot 0,25 \log(N_5+10) & \text{pour } S > 1,75 acum \end{cases}$$

et w_{FR} la contribution de la force de fluctuation et de la rugosité :

$$w_{FR} = \frac{2,18}{(N_5)^{0,4}}\ (0,4F + 0,6R)$$

Ce modèle a notamment été utilisé par Ih et al. sur des sons d'aspirateurs [85] et, dans le domaine médical, par Nielsen et al. sur des sons de valve de cœur mécanique [116].

Susini et al. [146] ont entrepris d'établir le lien entre mesure de qualité et descripteurs acoustiques par une autre méthode, appliquée notamment aux sons d'habitacle automobiles [145, 107] et aux sons d'unité de climatisation [147] [12]. Cette méthode est fondée sur le recueil de préférences issues d'une expérience de comparaisons par paire et sur leur modélisation par le modèle Thurtsone V [150] . Ces données sont par la suite analysées par une technique développée par De Soete et Winsberg [139]. Dans leur modèle, la probabilité P_{ij} qu'un son i soit préféré à un son j est modélisée comme une fonction Φ de la différence d'*utilité* de chaque son, U_i et U_j, respectivement :

$$P_{ij} = \Phi(U_i - U_j) \tag{2.21}$$

Ainsi, Φ étant une fonction normale cumulative, plus l'utilité du son i est grande devant celle su son j, plus le son i à de chances d'être préféré au son j. Toute la difficulté consiste alors à relier la fonction d'utilité aux descripteurs acoustiques, souvent sélectionnés grâce à une étude de timbre préalable. Dans l'algorithme de De Soete et Winsberg, deux modèles existent. Le modèle additif suppose que les influences des descripteurs sont indépendantes et que l'utilité globale correspond à la somme des utilités des descripteurs. Si l'on considère deux descripteurs a et b, leur fonction d'utilité respective f et g, l'utilité globale s'exprime donc par :

$$U_i = f(a_i) + g(b_i) \tag{2.22}$$

Dans le modèle multivarié l'influence des deux descripteurs peut être conjointe. L'utilité globale s'exprime alors par :

$$U_i = f(a_i, b_i) \tag{2.23}$$

Les fonctions d'utilités sont modélisées comme des fonctions polynomiales par morceaux. Le nombre de « morceaux », et donc de nœuds, et l'ordre des polynômes sont des paramètres de l'algorithme permettant de modéliser plus finement la relation descripteur-utilité, en fonction de différentes considérations statistiques.

Les figures 2.8, 2.9 et 2.10 montrent, en guise d'exemple, les fonctions d'utilités obtenues par Susini et al. pour les sons d'unités de climatisation [147], à partir des trois descripteurs identifiés comme attributs du timbre (sur chaque graphique, les deux courbes correspondent aux fonctions d'utilité pour deux classes latentes différentes d'auditeurs, dont la discrimination provient de l'analyse statistique des résultats).

12. On pourrait également rapprocher cette technique de celle utilisée par Lemaitre et al. [101] pour l'étude de la typicité des sons de klaxons.

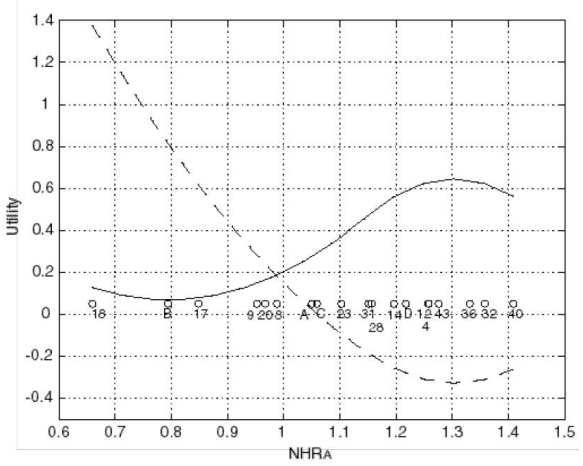


FIGURE 2.8 – Fonction d'utilité (spline) de l'émergence harmonique en pondération A (NHR_A) pour les sons d'unité de climatisation (provenant de Susini et al. [147]).

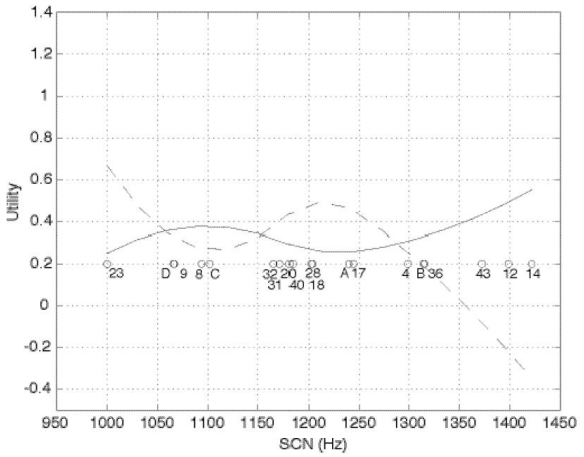


FIGURE 2.9 – Fonction d'utilité (spline) du centre de gravité spectrale de la partie bruitée (SCN) pour les sons d'unité de climatisation (provenant de Susini et al. [147]).

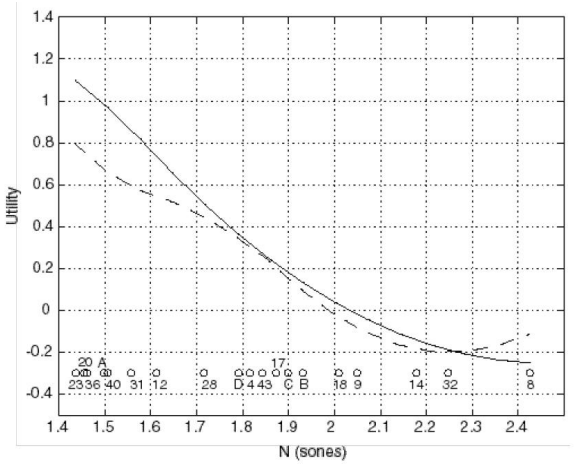


FIGURE 2.10 – Fonction d'utilité (spline) de la sonie (N) pour les sons d'unité de climatisation (provenant de Susini et al. [147]).

2.4 Démarche adoptée dans le cadre de cette thèse

Il convient à présent de définir précisément la démarche qui est adoptée dans le cadre de l'étude de la qualité sonore des STA, indépendamment de tout contexte. Les choix expérimentaux, et notamment les procédures et analyses envisagées, représentent le cœur de la démarche.

Tout d'abord, il semble peu pertinent d'aborder ici la question de l'identification de la source sonore, telle qu'introduite en section 2.2, En effet, il paraît raisonnable de penser que l'ensemble des sources sonores considérées (les STA) correspondent à un même type de cause physique, que l'on peut déjà placer dans le cadre de la catégorie *Moteur* identifiée par Misdariis et al. [111]. Cette catégorie regroupait selon les auteurs des sons incorporant une partie correspondant au son de fonctionnement du moteur (souvent sous forme de partiels harmoniques) et une partie correspondant aux turbulences aériennes produites (voir fin de la section 2.1.2). Il est difficile, en revanche, de placer ce type de source dans le cadre de la classification, plus générale que celle de Misdariis et al., des

sources sonores de Gaver [67]. En effet, cette classification (voir section 2.2) fait une distinction claire entre sons de solides en vibration (*vibrating objects* ou *vibrating solids*) et sons aériens (*aerodynamic sounds* ou *gasses*). Or, les sons de STA – et plus généralement les sons de *Moteur* – mélangent à priori les deux catégories citées.

Ainsi, comme nous considérons que les sons de STA appartiennent à une même catégorie de sons de l'environnement, nous pouvons d'ores et déjà nous intéresser au timbre et à la qualité perçue de ces sons. Pour sa précision et sa pertinence générale, l'approche *comparative*, telle qu'introduite en fin de section 2.3.2, a été adoptée. La première étape consiste en l'étude du timbre, qui, à la lumière des éléments bibliographiques présentés en section 2.1, a été réalisée par une méthodologie d'*analyse de proximités*. L'étude du timbre réalisée, incorporant expérience de *mesure de similarités*, *analyse MDS* et interprétation des dimensions perceptives, fait l'objet du chapitre 6.

Cette étude du timbre permettant d'identifier les attributs auditifs pertinents pour la perception des STA, nous pouvons nous intéresser ensuite à l'évaluation et la prédiction de la qualité sonore proprement dite. Ainsi, bien que l'approche *comparative* choisie inclue souvent une procédure de comparaison par paires, le chapitre 7 aborde cette étude en mesurant la qualité sonore par une procédure d'*évaluation comparée* (voir section 2.3.1). Ce choix a été fait pour le bon compromis de cette procédure entre précision et faisabilité pratique de l'expérience, sans travestir radicalement le principe de l'approche *comparative*.

Enfin, cette stratégie impose en premier lieu d'établir une base de travail, sous la forme d'un ensemble représentatif d'enregistrements sonores de STA. La méthode de prise de son et la description des STA enregistrés est exposées dans le chapitre 4. Par ailleurs, les types de procédure sélectionnés pour l'analyse de proximités et l'évaluation de la qualité sonore ont pour particularité de nécessiter un corpus réduit, ordinairement entre 10 et 20 éléments, pour conserver une forme de faisabilité pratique des expériences. Par conséquent, il est nécessaire d'extraire de l'ensemble des enregistrements effectués un corpus représentatifs de sa variété d'un point de vue perceptif. Le chapitre 5 présente le travail expérimental réalisé afin de réaliser cette sélection.

Chapitre 3

Contexte environnemental et qualité perçue

Le chapitre précédent avait pour but de présenter les méthodologies couramment utilisées dans la littérature pour l'étude de la qualité des sons de l'environnement. Ces éléments ont toutefois permis, dés lors que l'on commence à définir les éléments de base d'une telle étude (voir section 1.2), de souligner l'importance du *contexte environnemental* dans les jugements effectués par des auditeurs. Une définition plus générale en est donnée dans le domaine de la *psychologie environnementale* qui « considère la relation de l'individu à l'espace qui l'entoure comme un système d'interdépendances complexes dont le rôle et la valeur sont notamment déterminés par la perception et l'évaluation subjective dont un lieu est l'objet » [86]. Force est de constater que cette définition est centrée exclusivement autour d'une relation entre l'individu et le lieu. Dans le cadre plus spécifique de notre étude, centrée autour d'une relation entre l'individu et la source sonore, il convient d'en ajuster la portée. Nous entendons donc ici par contexte environnemental l'ensemble des éléments qui ne sont pas directement liés aux sources sonores, mais qui ont potentiellement une influence sur les jugements, que l'effectivité de cette influence ait été démontrée ou non dans la littérature.

Il est évident cependant que les facteurs possibles de ce contexte environnemental sont presque innombrables dans le cas de l'étude des sons de l'environnement. On peut aisément penser à de nombreux exemples variés, parmi lesquels on peut citer :

- des éléments directement liés au son, comme la localisation de la source (le fait que le son provienne d'une direction particulière), ou dans le même ordre d'idées, le fait que le son soit liée à un mouvement particulier (effet Doppler par exemple, dans le cas de sons liés aux transports urbains), ...
- des éléments liés au contexte culturel, comme la *fréquence écologique* telle que définie par Howard et Ballas [81], c'est-à-dire la fréquence à laquelle le son est rencontré dans la vie quotidienne, ou le rapport au désagrément qui peut varier d'une société à l'autre, ...
- des éléments liés à la psychologie cognitive, comme notre connaissance du type de son considéré ou du son en général, le lien affectif avec le son (par exemple, la façon dont est perçue un son de moto est probablement différente entre le motard et un piéton), la fonction du son, et notamment le fait qu'il soit ou non la conséquence d'une action humaine (exemple du klaxon), ...
- ...

Il semble donc irréaliste ici de souhaiter prendre en compte l'ensemble des facteurs influant potentiellement sur la qualité sonore perçue. De plus, parmi cette multitude de facteurs possibles, il ne serait pas déraisonnable de considérer l'influence de certains comme négligeable, voire hors de propos, et de hiérarchiser dans une certaine mesure les facteurs susceptibles d'avoir un réel effet sur la perception des sons considérés. Nous souhaitons donc ici nous intéresser à une quantité raisonnable de facteurs potentiels, dont nous soupçonnons qu'ils jouent un rôle primordial dans la perception et plus particulièrement l'évaluation de la qualité des sons de STA.

Tout d'abord, il est à ce stade judicieux de prendre en compte un aspect environnemental particulier liés au STA considérés : le fait qu'ils sont, dans une très grande majorité des cas, utilisés dans le contexte intérieur d'un bâtiment, souvent dans des bureaux, donc plus généralement dans ce que nous nommerons des « espaces clos ». Ceci implique de s'intéresser aux lois qui gouvernent l'acoustique dans ces espaces clos, dont le domaine est plus communément appelé *acoustique des salles*. De plus, un autre aspect particulier des espaces clos, non sans rapport avec le précédent, est digne d'intérêt : l'*intelligibilité de la parole*. Bien que la parole ne soit pas à priori un type de signal sonore auquel nous nous intéressons ici, il s'avère que son intelligibilité est souvent reliée avec la qualité des salles considérées et avec le « bruit de fond » – où le son de STA peut jouer une part non-négligeable – qui peut y régner.

Par ailleurs, un autre élément du contexte environnemental, plus en relation avec la cognition, semble pertinent lorsque l'on s'intéresse aux sons de STA : le *contexte attentionnel*. En effet, il convient de préciser la situation comportementale à laquelle doit être associée la perception de ces sons : usuellement, l'« usager » effectue une activité particulière (dans le contexte d'un bureau par exemple) qui focalise son attention, et la détourne globalement de son environnement sonore. La plupart des méthodologies expérimentales présentées dans le chapitre 2 incluent des procédures qui impliquent une focalisation de l'attention des auditeurs sur le son, ce qui semble quelque peu contradictoire avec le contexte évoqué ici. Il importe donc d'explorer dans la littérature les études qui ont été menées en utilisant des procédures expérimentales prenant en compte le degré et l'objet de l'attention de l'auditeur.

Ainsi, la section 3.1 de ce chapitre présente tout d'abord les éléments de la littérature scientifique qui sont liés à l'acoustique des salles. La section 3.2 présente ensuite d'importantes notions liées à l'intelligibilité de la parole. La section 3.3 explore les études menées en prenant en considération le contexte attentionnel. Enfin, la section 3.4 présente la démarche adoptée dans le cadre de cette thèse.

3.1 Qualité sonore dans les espace clos : facteurs liés à l'acoustique des salles

3.1.1 Principes de base de la réverbération

Lorsqu'un son est émis au sein d'une salle, un récepteur (microphone ou auditeur) va capter un son constitué de l'onde sonore directe (correspondant au trajet direct émetteur-récepteur) et d'un ensemble d'ondes secondaires issues des réflexions sur les parois de la salle. Les réflexions de premier ordre, c'est-à-dire les ondes n'ayant rencontré une paroi qu'une seule fois avant de parvenir au récepteur, sont en principe les premières à suivre l'onde directe, puis viennent les réflexions d'ordre supérieur. L'énergie de chaque onde parvenant au récepteur décroit avec l'ordre de la réflexion (donc le nombre de parois rencontrées sur le trajet) à cause notamment de l'absorption plus ou moins grande

des parois. En revanche, le nombre d'ondes parvenant au récepteur par unité de temps augmente avec le carré du temps t écoulé depuis la réception de l'onde directe [95] :

$$\text{Nombre de réflexions par seconde} = \frac{4\pi c^3}{V} t^2 \tag{3.1}$$

où c est la vitesse du son et V le volume de la pièce.

Les réflexions sur les parois de la salle peuvent être spéculaires sur une paroi lisse, ou diffuses sur une paroi plus irrégulière (voir figure 3.1). Ces réflexions sont notamment responsables de la réverbération dans les salles de grand volume (églises, salles de concert, stades, ...). Toutefois, la réverbération est également perçue dans des environnements plus confinés, même sans que l'on en soit conscient. En réalité, l'absence totale de réverbération n'existe pas dans la nature. Même dans un environnement particulièrement dégagé, le sol produit des réflexions/diffusions venant s'ajouter à l'onde directe. Le concept de *champ libre*, c'est-à-dire l'absence totale de réflexion, est simulé et approximé dans les *chambres anéchoïques*, conçues pour présenter une absorption maximale des parois (plancher et plafond compris). L'impression peu naturelle, au point d'en être parfois oppressante, que l'on peut ressentir dans une telle salle démontre l'importance des phénomènes tels que la réverbération dans notre environnement.

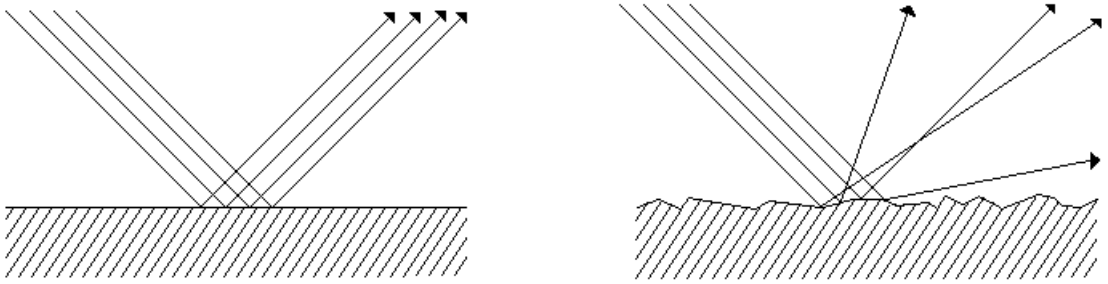


FIGURE 3.1 – Deux types de réflexions : réflexion spéculaire (gauche) et réflexion diffuse (droite).

On fait souvent l'hypothèse qu'une salle est un système linéaire, invariant dans le temps, et peut donc être caractérisée par sa réponse impulsionnelle, c'est-à-dire l'effet produit par les phénomènes de réflexions / absorption / diffusion des parois de la salle sur une impulsion acoustique unique émise par une source. En effet, il est possible de reproduire l'effet d'une salle sur un son enregistré en champ libre, en convoluant le signal sonore avec la réponse impulsionnelle, en tenant compte des positions respectives de la source et du récepteur. La réponse impulsionnelle se décompose usuellement en deux parties distinctes : la *partie précoce*, constituée de l'onde sonore directe (trajet direct de l'émetteur au récepteur) et des premières réflexions (trajets entre l'émetteur et le récepteur comprenant une ou deux réflexions sur des parois), et la *partie tardive*, comprenant l'ensemble des autres réflexions. Il est souvent admis que la partie précoce met en jeu des processus déterministes, qui peuvent donc être décrits de manière analytique. À l'inverse, compte tenu du nombre croissant de réflexions et de trajets d'onde possibles, les processus mis en jeu dans la partie tardive sont souvent considérés comme stochastiques, et une description d'ordre statistique est plus appropriée.

La question de l'instant de transition entre les parties précoce et tardive a été largement explorée dans la littérature. La plupart des auteurs semblent se rejoindre sur un temps de transition aux

alentours de 80 ms après réception de l'impulsion initiale (onde directe). Cependant, ce temps, qui correspond donc à la durée de la partie précoce, dépend de la fréquence et des dimensions de la salle. Ainsi, on peut trouver dans la littérature [17, 16, 94, 75, 141], plutôt qu'une valeur unique de durée, une gamme de variation de cette durée entre 50 et 200 ms. Ce temps de transition est souvent considéré comme constant jusqu'à une certaine taille de salle (quelques milliers de m^3) et varie au-delà proportionnellement à la racine carrée du volume ou au libre parcours moyen [23] (distance moyenne séparant deux impacts de l'onde acoustique sur des surfaces de la pièce [88]).

En termes perceptifs, outre les métriques permettant de mesurer le degré de réverbération ressentie, l'aspect spatial de la réverbération est décrit par deux attributs indépendants : la largeur apparente de source – *Apparent Source Width (ASW)* [15] – et l'enveloppement de l'auditeur – *Listener Envelopment (LEV)* [39]. Il est admis que la largeur de source apparente est liée à la partie précoce de la réponse impulsionnelle, tandis que l'enveloppement de l'auditeur est plutôt associé à la partie tardive.

3.1.2 Métriques liées à la réverbération

Le degré de réverbération : Ce que nous considérons par « degré de réverbération » est l'importance quantitative du percept de réverbération par rapport à la situation de référence qu'est le champ libre, indépendamment de sa distribution temporelle (précoce/tardive). Il existe de nombreux paramètres acoustiques permettant de mesurer le degré de réverbération d'une salle. Le plus répandu est le *temps de réverbération* (RT) égal au temps requis après l'arrivée du son direct pour que le niveau sonore du champ réverbéré décroisse jusqu'à une certaine valeur, souvent 60 dB en dessous du niveau du son direct. Le calcul se fonde toutefois souvent sur une partie réduite de l'enveloppe temporelle (entre -5 et -25 ou -35 dB par rapport au niveau du son direct).

Un autre paramètre permettant de quantifier la réverbération est l'EDT – *Early Decay Time* –, qui caractérise la décroissance de l'enveloppe temporelle sur les premières réflexions seulement (de 0 à -10 dB par rapport au niveau du son direct). Cette métrique est souvent considérée comme plus représentative du percept de réverbération que le temps de réverbération, notamment dans les salles de concert [11, 87].

Le rapport entre les énergies des parties précoce et tardive de la réponse impulsionnelle est également utilisé comme métrique (inverse) de la réverbération. Bien que le lien soit moins évident, il est aisé de concevoir qu'une partie tardive importante relativement à la partie précoce entraîne généralement un temps de réverbération long et donc une réverbération importante. Barron [16] parle d'ailleurs plutôt de « clarté » – *clarity* ou *early-to-late sound index* –, notée C_{80}, et décrite comme la capacité d'un auditeur à percevoir les détails sonores ou musicaux :

$$C_{80} = 10\log \frac{\int_0^{0.08} h_r^2(t)\,dt}{\int_{0.08}^{\infty} h_r^2(t)\,dt} \tag{3.2}$$

où $h_r(t)$ est la réponse impulsionnelle.

Pour les signaux de parole, on préfère associer les parties précoces et tardives de la réponse aux énergies arrivant respectivement avant et après 50 ms après l'arrivée du son direct [21]. C'est également le cas pour l'évaluation du paramètre appelé « définition » [95]– *Deutlichkeit* –, noté D_{50}, et qui permet notamment d'expliquer les variations de reconnaissance de la parole [136]. Il se calcule

comme le rapport entre l'énergie de la partie précoce (avant 50 ms) de la réponse impulsionnelle et son énergie totale, et s'exprime en pourcentage :

$$D_{50} = \frac{\int_0^{0.05} h_r^2(t)\,dt}{\int_0^{\infty} h_r^2(t)\,dt} \cdot 100\% \tag{3.3}$$

Bien que ces différentes grandeurs peuvent en théorie varier différemment l'une de l'autre, elles sont généralement fortement corrélées (voir par exemple les travaux de Bradley [34, 35]).

Largeur de source – ASW : On trouve dans la littérature différentes métriques visant à quantifier l'ASW. Barron [16], qui considère le temps de transition entre les partie précoce et tardive égal à 80 ms, lui associe la fraction latérale de l'énergie précoce sur les quatre octaves de 125 à 1000 Hz, *LF*. Celle-ci se calcule de la manière suivante :

$$LF = \frac{\int_{0,005}^{0,08} p_F^2(t)\,dt}{\int_0^{0,08} p^2(t)\,dt} \tag{3.4}$$

où $p(t)$ et $p_F(t)$ sont respectivement la réponse captée par un microphone omnidirectionnel, et par un microphone bidirectionnel avec le point nul [1] vers la source

Certains auteurs [75, 117, 132, 65, 66] préfèrent utiliser l'$IACC_E$ – *Inter-Aural Cross-correlation Coefficient* avec l'indice *E* pour *early* – pour caractériser l'ASW. Ce paramètre mesure la différence entre les sons parvenant aux deux oreilles à tout instant de la partie précoce de la réponse. Si les deux sons sont identiques, l'$IACC_E$ est nulle, s'ils sont complètement différents, l'$IACC_E$ vaut 1. L'$IACC_E$ est exprimée par :

$$IACC_E = \max_{\tau} \left(\frac{\int_0^{0,08} p_L(t)\,p_R(t+\tau)\,dt}{\int_0^{0,08} p_L^2(t)\,dt \int_0^{0,08} p_R^2(t)\,dt} \right) \tag{3.5}$$

pour τ compris entre -0,001 et 0,001 ms, et avec $p_L(t)$ et $p_R(t)$ correspondant au signal sonore arrivant respectivement dans l'oreille gauche et dans l'oreille droite.

Enveloppement de l'auditeur – LEV : Une métrique proposée par Bradley et Soulodre a permis d'expliquer efficacement la perception du LEV établie au travers d'expériences subjectives [39], et consiste à estimer la fraction d'énergie latérale de la partie tardive, notée *GLL*, également moyennée sur les octaves de 125 à 1000 Hz [38] :

1. Tout microphone est caractérisé, entre autres, par son *diagramme de directivité* qui représente sa sensibilité relative en fonction de l'azimut d'où provient l'onde sonore. Dans le cas d'un microphone bidirectionnel, ce diagramme est symétrique et en forme « huit », avec deux lobes sur ses côtés, et deux *point nuls* – à l'avant et à l'arrière – où la sensibilité est supposée nulle.

$$GLL = 10\log \frac{\int_{0,08}^{\infty} p_F^2(t)dt}{\int_0^{\infty} p_A^2(t)dt} \tag{3.6}$$

et $p_A(t)$ est le signal qui aurait été reçu à 10 m en champ libre pour la même source sonore

Afin de prendre en compte plus finement les aspects spatiaux de l'enveloppement de l'auditeur, un autre indicateur du LEV a été proposé par Hanyu et Kimura [74] le SBT_s – *Spatially Balanced Center Time* – qui est utilisé pour rendre compte de l'intelligibilité au sein d'une salle. Le SBT_s nécessite tout d'abord d'évaluer le temps central T_{si} pour chaque direction i :

$$T_{si} = \frac{\int_0^{\infty} t \cdot p_i^2(t)dt}{\int_0^{\infty} p^2(t)dt} \tag{3.7}$$

où $p_i(t)$ est la réponse impulsionnelle venant de la direction i

Le temps central est donc grossièrement un centre de gravité temporel dans une direction donnée. Le SBT_s est ensuite obtenu en intégrant les différentes directions du plan azimutal, tout en tenant compte de l'interaction entre les réponses des différentes directions de réflexion.

3.2 Qualité sonore dans les espaces clos : intelligibilité de la parole

Beaucoup de salles, notamment les lieux de travail ou d'études (bureaux, salles d'écoles, d'universités, ...), sont dédiées à des activités incluant beaucoup de communication orale entre les personnes s'y trouvant. Par conséquent, la qualité acoustique de telles salles passe par leur capacité à faciliter la communication orale. Les salles présentant un rendu acoustique pauvre peuvent entraîner deux problèmes. Premièrement, elles peuvent altérer l'efficacité d'apprentissage d'étudiants [83], par exemple, ou plus généralement la capacité à assimiler des informations ou des connaissances. Deuxièmement, elles peuvent entraîner fatigue, stress et divers problèmes de santé (maux de tête, de gorge), notamment pour les locuteurs qui doivent compenser ces mauvaises conditions acoustiques en élevant leur voix [153].

3.2.1 Lien entre réverbération, rapport signal-sur-bruit et intelligibilité

La propension d'une salle à faciliter la communication orale peut être quantifiée par une mesure dite d' « intelligibilité de la parole ». Cette métrique correspond grossièrement au pourcentage de parole correctement reconnu par les auditeurs. Hodgson postule, dans une étude recensant les caractéristiques acoustiques de diverses salles d'université [76], que cette métrique ne dépend que de deux facteurs :

– Le rapport signal-sur-bruit – *Signal-to-noise ratio* – qui correspond à la différence de niveau acoustique entre signal de parole et bruit de fond. Le signal de parole est le signal provenant du locuteur tel que reçu par l'auditeur. Le bruit de fond est ici associé aux autres sources sonores

qu'il est possible de trouver dans un tel environnement : bruit de ventilation, bruit d'activité des étudiants, bruit d'équipement particulier (projecteur, ...) et bruits extérieurs à la salle.
– La réverbération, dont le degré peut être quantifié par les différentes métriques évoquées en section 3.1.2. Dans le cas de l'étude de l'intelligibilité, on utilise souvent le rapport des énergies précoce et tardive, évaluées sur la réponse impulsionnelle (similaire au descripteur de *clarity* de Barron [16], voir section 3.1.2), voire directement sur un signal de parole [33] (dans le second cas, toutefois, on parle plutôt de parties utile et préjudiciable à l'intelligibilité – *useful/detrimental*).

Il est généralement admis que les relations entre ces paramètres et l'intelligibilité sont monotones. Ainsi, d'un côté, l'intelligibilité augmente lorsque le rapport signal-sur-bruit augmente, et de l'autre, elle diminue lorsque la réverbération augmente. La plupart des études expérimentales tendent à confirmer cette hypothèse en établissant que la condition théoriquement idéale de réverbération pour favoriser l'intelligibilité est le champ libre, correspondant notamment à un temps de réverbération nul. Parmi celles-ci, l'étude de Nábělek et Robinson [114], dans une approche perceptive de l'intelligibilité, a exploré différentes situations de réverbération en utilisant une méthode de MRT – *Modified Rhyme Test*[2] – à niveau de parole constant et sans bruit de fond. La mesure d'intelligibilité ainsi obtenue, sous la forme de pourcentage de reconnaissance correcte, augmente lorsque la réverbération diminue, ce qui semble confirmer l'hypothèse d'un temps de réverbération idéalement nul. D'autres études [113, 64] ont intégré le paramètre de rapport signal-sur-bruit dans le même type d'expérience et ont obtenu la même conclusion en termes de relation entre réverbération et intelligibilité.

Il existe toutefois une théorie selon laquelle la réverbération peut avoir un effet bénéfique sur l'intelligibilité car elle permet d'augmenter le niveau sonore global du signal de parole. La valeur de réverbération optimale serait alors non-nulle. Cette hypothèse est supportée par plusieurs études abordant la prédiction théorique de l'intelligibilité à partir, notamment, des métriques évoquées en section 3.1. Plomp et al. [128] ont ainsi trouvé un temps de réverbération optimal variant de 0,3 s à 1,7 s en fonction de la taille de la salle. Bistafa et Bradley [24] incorporent dans leur étude de l'intelligibilité le paramètre de rapport signal-sur-bruit et trouvent un temps de réverbération optimal de 0,1 s à 0,6 s en fonction également du volume de la salle. Ils justifient ce résultat en expliquant qu'une légère réverbération permet d'augmenter la partie précoce de l'énergie sonore, compensant ainsi la présence du bruit de fond.

Hodgson [77] tente d'aborder cette problématique et la controverse issue de la comparaison de ces études en posant l'hypothèse selon laquelle les paramètres de réverbération et de rapport signal-sur-bruit ne sont pas indépendants. Afin de tester cette hypothèse, il prend également en compte, au travers d'un modèle de prédiction théorique raffiné, les distances séparant l'auditeur des sources de parole et de bruit. Il observe ainsi que lorsque l'auditeur est plus proche du locuteur que de la source de bruit ou à égale distance, le temps de réverbération optimal est nul, tandis que, dans le cas inverse, il devient non-nul. Ceci peut s'expliquer par le fait que, lorsque la distance entre auditeur et locuteur est plus grande que celle entre auditeur et source de bruit, l'augmentation de la partie précoce de la réverbération du signal de parole – plus importante relativement au niveau émis que pour le bruit de fond – entraîne une augmentation du rapport signal-sur-bruit, qui peut compenser l'effet négatif inhérent à la réverbération. Il explique alors les résultats contradictoires des études précédemment mentionnées par certaines limitations pratiques des expériences qui y sont réalisées. En effet, les études expérimentales ayant abouti à des temps de réverbération optimaux nuls, par les

2. Méthode où les auditeurs doivent choisir le mot entendu parmi une liste de mots monosyllabiques consonants.

procédures expérimentales adoptées, ne prennent pas en compte la présence de bruit de fond, ou négligent l'influence de la réverbération sur les niveaux perçus des différentes sources.

Malgré tout, cette théorie n'est pas adoptée systématiquement dans la littérature et on considère souvent, dans une approximation raisonnable, que la valeur de réverbération maximisant l'intelligibilité est nulle ou très proche de 0. Par ailleurs, certaines études [32, 25] ont établi les conditions optimales de rapport signal-sur-bruit et de réverbération, notamment dans le cas de salles de classes. Selon les résultats de ces études, le temps de réverbération ne doit pas dépasser 0,7 secondes, tandis que le rapport signal-sur-bruit doit atteindre 15 dB(A) au minimum, ce qui implique un niveau de bruit de fond maximum de 35 dB(A), dans des conditions normales de paroles [127].

3.2.2 Descripteurs d'intelligibilité

Il existe un algorithme qui permet de modéliser l'influence conjointe des paramètres précédemment cités. La métrique obtenue est le STI – *Speech Transmission Index* – introduit par Houtgast et Steeneken [79]. Son calcul est normalisé [5], et est fondé sur le rapport signal-sur-bruit et la réverbération, mais prend également en compte certaines considérations liées à l'audition humaine, en particulier en termes de pondérations fréquentielles et de masquage. Cette méthode d'estimation consiste à émettre un signal de référence censé refléter les propriétés d'un signal de parole. Il s'agit d'un signal constitué de 7 porteuses, chacune constituée d'une bande d'une demi-octave de bruit blanc centrée sur l'une des 7 octaves entre 125 Hz à 8 kHz. Chacune de ces porteuses est modulée par des composantes sinusoïdales dont la profondeur de modulation vaut 100 %, et la fréquence de modulation varie de 0,63 à 12,5 Hz par tiers d'octave (14 composantes). L'idée générale est alors d'estimer la perte en profondeur de modulation pour chacune des composantes (porteuses et fréquence de modulation) liée à la réponse de la salle, en tenant compte du seuil d'audition selon la fréquence, des effets de masquage, et de pondérations propres à chaque bande et représentant l'importance relative de la bande pour l'intelligibilité de la parole.

Les valeurs du STI varient entre 0 (intelligibilité nulle) et 1 (parfaite intelligibilité)[3]. La pertinence perceptive de cette métrique a également été abordée en établissant une correspondance entre les valeurs de l'échelle de STI et une mesure perceptive d'intelligibilité, sous forme de pourcentage moyen de reconnaissance de mots monosyllabiques, ayant un sens sémantique ou non [10]. Le tableau 3.1 présente les différentes plages de valeurs du STI dont on admet communément qu'elles correspondent à différents degrés d'intelligibilité.

$0 \rightarrow 0,3$	*Bad*	« Inintelligible »
$0,3 \rightarrow 0,45$	*Poor*	« Faible »
$0,45 \rightarrow 0,6$	*Fair*	« Satisfaisant »
$0,6 \rightarrow 0,75$	*Good*	« Bon »
$0,75 \rightarrow 1$	*Excellent*	« Excellent »

TABLE 3.1 – Plages de valeurs du STI et qualifications correspondantes usuelles de l'intelligibilité, en anglais et en français.

Une méthode longtemps utilisée pour estimer plus simplement le STI est appelée RASTI – *RApid Speech Transmission Index* [80]. Cette méthode n'utilise plus que 2 porteuses de bruit rose d'une octave à 500 et 2000 Hz, modulées par un nombre réduit (respectivement 4 et 5) de fréquences de composante. Le RASTI est toutefois une méthode assez ancienne et a même été très récemment (2011)

3. Il convient de noter que le sens physique de ces bornes n'est pas précis ; en effet, les résultats du calcul sont tronqués au-delà de certaines valeurs extrêmes.

déclarée obsolète [5]. On lui préfère aujourd'hui la méthode STIPA [5] – *Speech Transmission Index for Public Address Systems* – développée, comme son nom l'indique, pour les système de diffusion acoustique dans les lieux publics, et qui y adjoint certains effets de dégradation du signal inhérents aux systèmes électro-acoustiques utilisés. Cette méthode utilise les mêmes porteuses que pour l'estimation du STI, mais seulement 2 fréquences de modulation pour chaque bande.

D'autres métriques conçues pour quantifier l'intelligibilité, et ayant été normalisées, sont mentionnées dans la littérature, notamment l'AI – *Articulation Index* [3] – et sa version plus récente, le SII – *Speech Intelligibility Index* [4]. Cependant, ces échelles sont peu différentes de celle de STI, si ce n'est qu'elles ne prennent pas en compte l'effet de la réverbération sur le signal de parole. On peut trouver une comparaison des trois métriques (AI, SII et STI) dans une étude de Bradley [36]. Il précise notamment que le STI est plus souvent utilisé en Europe, tandis que l'AI et le SII sont plutôt utilisés en Amérique du Nord.

3.2.3 Intelligibilité et confidentialité

Il est intéressant de signaler une utilisation dérivée de la métrique d'intelligibilité pour la qualification des environnements acoustiques des bureaux dits « paysagés », c'est-à-dire non cloisonnés – *open-plan offices*. La multiplication de ce type de configuration de bureaux et les procédures de développement et de design qui en découlent invitent à s'interroger sur les conditions de confort acoustique dans ce type d'environnement. Certaines considérations de design ont ainsi émergé, notamment par rapport à la problématique de la confidentialité. En effet, de nombreux usagers de ce type de bureaux peuvent être amenés dans le cadre de leur travail à avoir des conversations directes ou téléphoniques dont le contenu peut être accompagné d'un certain degré de confidentialité vis-à-vis des autres usagers. Outre leur caractère éventuellement confidentiel, ces conversations peuvent également provoquer un certain degré de gêne ressentie par les usagers situés sur les postes de travail adjacents, voire affecter leur efficacité au travail. Hongisto [78] présente notamment une compilation d'études confirmant l'effet préjudiciable d'un signal intrusif de parole sur la performance au travail. Le degré de confidentialité proposé par une configuration donnée de bureau paysagé pourrait donc être mesuré à l'aide d'une échelle inversée d'intelligibilité de la parole.

Il s'agit donc ici de jouer sur les deux paramètres influant sur l'intelligibilité de la parole : réverbération et rapport signal-sur-bruit. Il semble délicat de jouer sur la réverbération dans le contexte des bureaux paysagés, dont l'agencement et le caractère naturellement modulable n'autorise pas un contrôle efficace de ce paramètre. Il est en revanche intéressant de constater que le bruit de fond, quelle qu'en soit la nature, peut avoir ici un effet bénéfique, car il permet de masquer le signal de parole et en réduit ainsi les nuisances sonores. Cette théorie rejoint les résultats d'une étude de Khan [90] portant sur le confort acoustique dans les trains, et notamment sur le masquage des conversations téléphoniques par les différentes sources de bruit dans les rames. Augmenter le bruit de fond, et notamment le bruit de ventilation par exemple, permettrait donc d'améliorer le confort acoustique dans les bureaux paysagés. Bien entendu, cette assertion a ses limites, et une augmentation excessive du bruit de fond conduirait à une gêne accrue inhérente à l'augmentation du niveau global de l'environnement sonore.

Bradley et Gover [37] ont tenté d'évaluer la pertinence de cette théorie en abordant la relation entre une métrique d'intelligibilité – le SII, en l'occurrence – et le degré de distraction ressentie, vis-à-vis de tâches accaparantes d'édition de texte et de calcul simple. La performance des auditeurs dans la tâche à effectuer n'a pas été influencée par les différentes sources (bruit ou parole), probablement

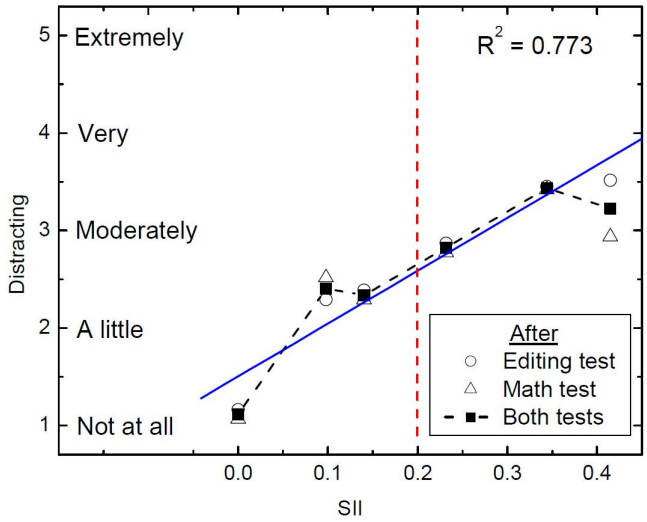


FIGURE 3.2 – Jugements de distraction en fonction de l'intelligibilité – SII (provenant de Bradley et Gover [37]).

du fait de la courte durée des tâches (1 à 2 minutes). La raison d'être de cette tâche accaparante est de rendre les conditions de test plus réalistes et de faire en sorte que les auditeurs ne se concentrent pas sur les aspects acoustiques. Les résultats ont montré que la mesure d'intelligibilité est un bon prédicteur de la distraction ressentie ($R^2 = 0.773$, voir figure 3.2). Une autre conclusion plus pratique de cette étude concerne le niveau optimal de bruit de fond permettant de minimiser la gêne provoquée par les conversations sur les postes de travail adjacents tout en limitant la gêne inhérente à la présence de ce bruit de fond. Ils ont ainsi établi qu'un bruit de fond à 45 dB(A) représente le meilleur compromis. Un niveau beaucoup plus faible réduira significativement la confidentialité tout en augmentant la gêne liée aux conversations adjacentes, et un niveau plus élevé (une limite de 48 dB(A) est indiquée) augmentera la gêne liée au bruit de fond, et pourra même inciter à parler plus fort, compensant ainsi tout gain de confidentialité.

Une autre étude de Bradley [36] a évalué l'effet de différentes modifications de design de bureaux paysagés (absorption et hauteur des panneaux entre postes de travail, taille du poste de travail, ...) sur les métriques d'intelligibilité (AI, SII et STI), et donc sur la confidentialité induite par ces conditions. Il est apparu que la hauteur de panneau est le paramètre ayant le plus d'influence sur l'intelligibilité.

3.3 Contexte attentionnel lié à l'environnement sonore

3.3.1 Effet « cocktail party »

Un aspect particulier de l'audition a fait l'objet de très nombreuses études depuis qu'il a été mis en évidence. Il s'agit de la capacité d'un auditeur à focaliser son attention sur un évènement ou message sonore – *auditory event* – particulier parmi plusieurs autres. Cette capacité est celle qui est mise à contribution lorsque l'on tente d'écouter un locuteur dans un environnement bruyant ou parmi d'autres locuteurs « interférents ». Cette particularité de l'audition est couramment nommée l'*effet cocktail party*. Cet effet a été pour la première fois mis en évidence par Cherry [44, 45], et est toujours étudié aujourd'hui. À l'origine, l'étude de ce phénomène avait pour motivation la résolution des problèmes rencontrés par les contrôleurs du trafic aérien dans les tours de contrôle. Ceux-ci recevaient alors les messages provenant de différents pilotes sur un seul haut-parleur. En conséquence, la tâche

consistant à distinguer les voix des pilotes s'avérait particulièrement difficile.

De nombreuses études ont mis en évidence que la détection d'un signal dans un environnement bruité est fortement liée au caractère binaural de notre audition : « One of the most striking facts about our ears is that we have two of them – and yet we hear one acoustic world ; only one voice per speaker [4] » [45]. Il s'avère que notre capacité de détection est bien plus élevée avec deux oreilles (présentation dichotique) qu'avec une seule (présentation monotique) [5]. Il a été montré que l'écoute binaurale peut améliorer le seuil de détection jusqu'à 25 dB BLMD – *Binaural Masking Level Difference* – dans des conditions idéales par rapport à l'écoute monaurale [51]. Cet état de fait permet d'expliquer en partie l'effet cocktail party, car la différence d'angle d'incidence entraîne un masquage moins important du signal cible par le signal interférant [26].

3.3.2 Attention et perception en contexte multi-tâche

L'observation des travaux effectués sur l'intelligibilité de la parole et le cas particulier de l'effet cocktail party ont révélé les particularités de la perception humaine lorsque l'on est face à plusieurs stimuli simultanés. Toutefois, dans le champ d'application abordé jusqu'ici, les stimuli, que l'on peut à ce stade différencier en cibles et perturbateurs, font appel à la même capacité sensorielle, l'audition. Par conséquent, si une partie du traitement de l'information est dévolue aux capacités cognitives de notre cerveau, certains aspects propres à la modalité auditive de notre perception entrent également en jeu, et permettent de fournir d'importants indices facilitant l'identification des sources (masquage binaural notamment).

Par ailleurs, l'étude de tels phénomènes soulève également la question des capacités attentionnelles humaines, c'est-à-dire en l'occurrence l'aptitude à focaliser son écoute sur une ou plusieurs sources sonores et faire abstraction des autres. Néanmoins, il est fréquent, dans un contexte environnemental réaliste, que l'on doive traiter des informations provenant de stimuli de différentes natures, et pas uniquement sonores. De plus, nous ne nous contentons pas d'être passifs vis-à-vis de la perception multi-sensorielle de notre environnement, mais nous interagissons avec lui, ajoutant par exemple la gestion de fonctions motrices aux tâches incombant à notre cortex cérébral déjà bien occupé.

Or, lors d'expériences auditives telles que celles introduites en section 2.3, on demande aux auditeurs d'émettre un jugement ponctuel sur un son, tandis qu'ils n'ont en principe aucune autre source de distraction. Ces conditions expérimentales particulières les incitent par conséquent à focaliser leur attention sur le son écouté. Ce phénomène est renforcé par la durée, souvent courte – quelques secondes tout au plus –, des sons, n'autorisant pas les auditeurs à laisser leur concentration dévier du son émis. Le contexte environnemental ainsi reproduit n'est pas représentatif de la réalité, où, d'une part, les sons – et plus particulièrement ceux étudiés ici – peuvent durer bien plus que quelques secondes, et, d'autre part, leur écoute est concomitante à la perception d'autres types de stimulus, sonores ou non, et à la réalisation de tâches plus ou moins accaparantes. Ces derniers éléments sont particulièrement importants lorsque l'on tente d'établir une mesure perceptive pertinente dans les conditions écologiques que sont celles de l'usage de STA. Il est donc spécialement important de s'intéresser à la problématique de la méthodologie expérimentale permettant d'obtenir des évaluations

4. « Un des aspects les plus frappants concernant nos oreilles est que nous en avons deux – et pourtant, nous entendons un seul monde acoustique ; une seule voix par locuteur »

5. L'écoute *monotique* ou monaurale consiste à présenter un signal à une seule oreille, tandis que l'écoute *dichotique* consiste à présenter des signaux différents aux deux oreilles, ce qui correspond à une condition d'écoute proche de la réalité. On parle aussi parfois d'écoute *diotique* où on présente aux deux oreilles le même signal. Cette condition expérimentale correspond dans la réalité au cas particulier où la source sonore est située dans le plan médian de la tête, et ne véhicule donc, comme l'écoute monotique, aucun indice spatial sur l'environnement sonore

perceptives représentatives du contexte environnemental des sons étudiés.

La notion d'attention est déjà en soi un sous-domaine de la psychologie cognitive ayant généré une littérature fournie (voir notamment les travaux de Pashler [120, 121], pour plus de détails). Dans le cadre de la problématique que nous nous posons, on peut trouver plus spécifiquement dans la littérature quelques études ambitionnant d'aborder le design sonore en général, en prenant en compte l'attention de l'auditeur dans la procédure expérimentale. On peut tout d'abord citer l'étude de Bradley et Gover [37], déjà évoquée en section 3.2.3. En effet, cette étude rapporte les résultats d'une expérience d'évaluation de la distraction provoquée par différentes conditions d'intelligibilité de la parole et de bruit de fond, vis-à-vis de tâches accaparantes d'édition de texte et de calcul simple. Ces conditions d'expérience représentent un premier exemple d'évaluation perceptive obtenue dans un contexte où l'attention de l'auditeur est détournée du son.

Nous pouvons également citer une étude similaire d'Ebissou et al. [52], beaucoup plus récente. En effet, cette étude aborde également la question de l'influence de l'intelligibilité sur la performance des auditeurs dans un contexte de bureau paysagé. Lors de l'expérience réalisée, les auditeurs devaient réaliser une tâche de mémorisation de séquences de chiffres. Pendant la réalisation de cette tâche, les auditeurs entendaient une séquence mélangeant les voix de plusieurs locuteurs (de sorte qu'aucun message ne soit compréhensible), dont on faisait varier par modélisation la valeur du STI (voir section 3.2.2). À la fin de chaque session de mémorisation, les auditeurs devaient évaluer, au travers de différentes questions, la difficulté qu'ils avaient éprouvée à réaliser cette tâche de mémorisation. Les résultats de cette expérience ont mis à jour une distinction de deux groupes d'auditeurs, qui manifestaient des tendances de performance différentes : dans le premier groupe, la performance des auditeurs ne semblait pas évoluer avec la valeur du STI, tandis que dans le second, elle diminuait quelque peu lorsque le STI augmentait. L'analyse des évaluations des auditeurs a également révélé que, d'une part ceux du premier groupe évaluaient globalement la tâche comme moins difficile que ceux du second, et d'autre part, pour les deux groupes, la difficulté de la tâche n'était jugée plus importante qu'entre des valeurs extrêmes de STI. Si la problématique de cette étude est assez éloignée de la notre, elle offre encore un excellent exemple d'étude de l'effet indirect de stimuli sonores sur la réalisation d'une tâche accaparant l'attention des auditeurs.

Les travaux de thèse de Suied sur l'urgence véhiculée par le son, et plus particulièrement une étude évaluant le temps de réaction en contexte attentionnel dévié [143], représentent également un exemple d'expérience incorporant une tâche accaparant l'attention des auditeurs. Si le but de cette étude n'est pas en rapport avec la présente problématique, la procédure expérimentale utilisée n'est pas dénuée d'intérêt. La tâche demandée aux auditeurs était de réagir le plus rapidement possible à différents stimuli censés véhiculer l'urgence, tout en ayant en permanence à suivre, avec la souris de l'ordinateur maniée de leur main non-dominante, un point se déplaçant sur l'écran. Outre la mesure du temps de réaction des auditeurs, les auteurs ont également observé le degré de performance de chacun d'eux à accomplir la tâche accaparant son attention, c'est-à-dire le suivi du point sur l'écran, sous forme d'une distance moyennée dans le temps entre le point et le curseur de la souris.

Une étude plus ancienne de Susini et McAdams [144] a également mis au point une procédure expérimentale permettant de comparer les résultats d'une expérience d'estimation de sonie globale de longues séquences sonores non-stationnaires dans différents contextes attentionnels. Le premier contexte attentionnel était la condition témoin, où l'on demande à l'auditeur d'évaluer en temps réel la sonie de la séquence au fur et à mesure que celle-ci est lue, avant d'estimer la sonie globale de la séquence. Cette condition était considérée comme la condition témoin, car elle forçait l'auditeur à focaliser son attention sur le son. Une deuxième condition consistait simplement à écouter chaque

séquence, sans tâche subsidiaire, et à en estimer la sonie globale à la fin de la lecture. Enfin, la dernière condition était celle où l'attention de l'auditeur était volontairement déviée du son : de manière similaire à l'étude précédemment citée de Ebissou et al. [52], l'auditeur devait, pendant la lecture du son, mémoriser puis taper au clavier une séquence de trois chiffres apparaissant successivement à l'écran. La conclusion principale de cette étude était que les évaluations de sonie semblaient plus élevées dans la dernière condition où l'attention des auditeurs était déviée du son.

Enfin, Gros et al. [71] ont également comparé des résultats obtenus dans deux contextes attentionnels différents. Leur étude présente une série d'expériences dont le but était d'évaluer, dans différentes situations, la qualité audio d'un signal de parole ponctuellement dégradé à différents degrés. La comparaison effectuée entre deux de ces situations paraît particulièrement intéressante dans le présent contexte. Dans la première situation, des paires d'auditeurs devaient converser par le biais d'un système de communication (dont la qualité était volontairement altérée par instant) selon des scénarii prédéfinis, puis évaluer en fin de conversation la qualité globale de la communication. Dans la seconde situation, d'autres auditeurs devait écouter des séquences de signal de parole (également dégradées ponctuellement), et évaluer la qualité du signal audio, d'une part en temps réel pendant la lecture de la séquence, et d'autre part, à la fin de sa lecture, d'un point de vue global cette fois-ci. Ainsi, si la seconde situation reproduit un contexte d'écoute attentive (imposé par le suivi en temps réel de la qualité perçue), la première situation reproduit bien un contexte d'écoute où l'attention des auditeurs est quelque peu déviée en leur imposant de participer à une conversation, même si celle-ci est prédéfinie. Les conclusions de la comparaison des jugements des auditeurs pour ces deux situations étaient que les jugements n'étaient que peu significativement différents entre les deux situations, mais que celle de conversation entraînait une gamme de valeurs de qualité légèrement plus faible et des écarts-types entre auditeurs plus importants. Si le type de procédure utilisée dans le but de modifier le contexte attentionnel au cours de cette étude n'a que peu de rapport avec notre problématique, les conclusions obtenues peuvent revêtir un certain intérêt dans notre cas, car elles portent sur des jugements de qualité sonore.

Dans un souci de synthèse, il convient de comparer les objectifs et les démarches des études précédemment citées. Nous constatons notamment que :

- Trois d'entre elles – Bradley et Gover [37], Ebissou et al. [52] et Suied et al. [143] – comparent une évaluation perceptive [6] en contexte d'*écoute distraite*, avec une mesure objective (la performance des auditeurs dans la tâche subsidiaire – mémorisation des séquences de chiffres pour Bradley et Gover et Ebissou et al., et suivi de point à l'écran pour Suied et al.).
- Deux autres – Susini et McAdams [144] et Gros et al. [71] – tentent notamment de comparer deux évaluations perceptives, l'une dans un contexte référence d'*écoute attentive* des stimuli, et l'autre dans un contexte d'*écoute distraite* où l'attention des auditeurs est centrée sur tâche subsidiaire.

Dans le cadre de notre problématique générale, il semble que la seconde comparaison citée représente l'objectif primordial, puisque notre questionnement concerne principalement la pertinence de l'évaluation perceptive en contexte d'*écoute attentive*. Cependant, dans le cadre de l'évaluation de la qualité sonore des STA, la mesure de la performance d'auditeurs en contexte d'*écoute distraite* n'est pas dénuée d'intérêt.

6. On pourrait toutefois objecter que le calcul du temps de réaction n'est pas une mesure perceptive à proprement parler.

3.4 Démarche adoptée dans le cadre de cette thèse

La large variété, évoquée en introduction de ce chapitre, de facteurs environnementaux ayant potentiellement une influence sur la qualité sonore perçue impose de les hiérarchiser lorsque l'on considère un type particulier de sons de l'environnement. Ainsi, le contenu de ce chapitre s'est focalisé sur trois éléments, que nous jugeons les plus importants dans le cadre de l'étude de la qualité sonore des STA : l'acoustique des salles, l'intelligibilité de la parole et le contexte attentionnel. La sélection de ces trois éléments découle du lieu typique d'installation des STA considérés, c'est-à-dire des bureaux de différentes tailles et configurations. Ce contexte particulier implique que les sons sont perçus en intérieur, où règne nécessairement une forme de réverbération, dont l'effet sur le son n'est bien entendu pas négligeable. Il nous amène également à nous intéresser à la relation qui existe entre l'environnement sonore, dont les STA font partie, et l'activité des usagers. Celle-ci peut inclure une communication orale sous différentes formes (téléphone, réunion, ...) entre ces derniers. Ce dernier point incite à considérer également la problématique de l'intelligibilité de la parole. En d'autres circonstances, l'activité des usagers peut aussi se traduire par la réalisation par les usagers de tâches d'un autre ordre, excluant au contraire toute forme de communication, et impliquant un certain degré de concentration que l'environnement sonore peut contrarier.

Dans le cadre de cette thèse, il semble toutefois difficile d'aborder de manière satisfaisante ces trois points en même temps. Bien que les éléments bibliographiques exposés au cours de ce chapitre montrent que ces trois facteurs ne sont pas sans rapport entre eux, il est nécessaire de pouvoir observer avec précision leur influence respective sur la qualité sonore perçue des STA, avant d'envisager de s'intéresser à une possible interaction de ces facteurs. Dans le cadre de cette thèse, nous avons pris le parti de les aborder séparément, l'étude de leur possible interaction faisant plutôt office de perspective. Ainsi, chacun d'eux implique une approche expérimentale différente. Il est donc nécessaire de considérer quels facteurs parmi les trois ciblés dans ce chapitre représentent les points d'intérêt les plus pertinents dans le présent contexte.

Nous considérons bien entendu qu'il est presque indispensable de commencer par évoquer la problématique de l'acoustique des salles. En effet, les travaux exposés dans les chapitres 5, 6 et 7 abordent l'évaluation de la qualité sonore à l'aide d'enregistrements anéchoïques. Cette particularité est rendue nécessaire par la grande variété de configurations de salle possibles dans lesquelles sont installés les STA. Ainsi, les études exposées dans ces chapitres sont centrées autour des sources sonores que sont les STA, afin d'éviter d'obtenir des résultats dont la validité ne dépasserait pas le cadre d'une configuration de salle choisie arbitrairement. Néanmoins, il importe également de considérer le son de STA tel qu'il parvient aux oreilles des auditeurs, en considérant les quelques configurations de salle les plus représentatives et les réverbérations produites, et d'observer à quel point la qualité sonore perçue s'en trouve modifiée. Cette étude fait l'objet du chapitre 8

Parmi les deux éléments restants du contexte environnemental, il pourrait être intéressant de lier la qualité sonore perçue des STA avec leur propension à perturber l'intelligibilité de la communication orale qui peut devoir intervenir typiquement dans un bureau. Toutefois, les travaux exposés en section 3.2.3 sur le lien entre intelligibilité et confidentialité montrent que la relation entre qualité sonore et intelligibilité, que l'on pourrait intuitivement supposer proportionnelle, est peut-être plus complexe. En effet, certains travaux y préconisent un niveau suffisant de bruit de fond, dans le cadre duquel entre le son de STA, afin d'assurer une faible intelligibilité, favorisant ainsi une forme de confidentialité. Cela signifie que la forme de confort sonore que reflète le degré de confidentialité est ponctuellement inversement proportionnelle à l'intelligibilité. Par ailleurs, il semble difficile de

parler d'intelligibilité sans considérer également la réverbération, ce qui va à l'encontre de notre volonté initiale de considérer séparément les différents facteurs susceptibles d'influer la qualité sonore perçue des STA. Pour toutes ces raisons, nous avons jugé que l'étude de la qualité sonore des STA vis-à-vis de l'intelligibilité ne représentait pas une priorité, elle n'a donc finalement pas été abordée dans le cadre de cette thèse.

Nous avons donc préféré nous concentrer sur le dernier point, qui concerne le contexte attentionnel. En effet, s'il existe de nombreuses études abordant la problématique de l'intelligibilité et plusieurs modèles dont le but est de quantifier cette dernière (voir sections 3.2.1 et 3.2.2 à ce sujet), la littérature est beaucoup moins fournie en ce qui concerne le lien entre qualité sonore perçue et contexte attentionnel des auditeurs. Seules quelques études mentionnées en section 3.3.2 mentionnent l'emploi de procédures expérimentales de mesure de perception sonore en contexte multi-tâche, mais le but des études correspondantes est souvent assez éloigné de la problématique de l'évaluation de la qualité sonore. Cependant, les méthodologies employées, et notamment les deux types de démarche mis en évidence en fin de section, ne sont pas dépourvues d'intérêt dans notre cas. La question de l'influence du contexte attentionnel représente donc, à nos yeux, une problématique pertinente et digne d'intérêt dans le cadre de l'évaluation de la qualité sonore des STA, et fait donc l'objet du chapitre 9.

Partie II

Perception des sons de STA : étude du timbre et prédiction des préférences

Cette partie du document expose les travaux réalisés dans le but d'établir une métrique de qualité sonore des STA. Le **chapitre 4** détaille le protocole d'enregistrement des échantillons sonores et les STA enregistrés, afin de construire une base de données sonores . Le **chapitre 5** décrit l'identification des différentes familles de sons de STA qui constitue la base de données sonores, afin d'établir un corpus sonore de travail. Le **chapitre 6** expose l'étude du timbre afin d'identifier les attributs auditifs pertinent pour les sons de STA. Enfin, le **chapitre 7** présente le travail réalisé dans l'optique d'établir une métrique de qualité sonore sur la base des descripteurs du timbre.

Chapitre 4

Enregistrement de sons de STA

Ce chapitre présente la méthodologie de prise de son employée pour enregistrer le son émis par différents STA. La section 4.1 présente le protocole général de prise de son qui a été utilisé pour l'ensemble des enregistrements. La section 4.2 présente les détails pratiques liés à l'enregistrement des différents types de STA fourni par le partenaire industriel (CIAT). Enfin la section 4.3 offre quelques points de discussion sur cet ensemble d'enregistrements vis-à-vis de la problématique de cette thèse.

4.1 Protocole de prise de son

4.1.1 Condition anéchoïque d'enregistrement

Tous les enregistrements ont été effectués dans une chambre semi-anéchoïque, c'est-à-dire où les parois sont traitées de telle sorte qu'elles absorbent au maximum l'énergie acoustique les atteignant, afin de réduire la réverbération au minimum (toutefois, le sol de la pièce n'était pas traité). La raison d'être d'une telle condition d'enregistrement est que les environnements spatiaux dans lesquels sont perçus les sons peuvent varier de manière significative, et enregistrer les sons dans une situation particulière invaliderait certainement les résultats obtenus lorsque l'on considère une autre situation. De plus, si la condition anéchoïque n'est certainement pas la plus fréquente dans la nature – d'autant plus lorsque l'on considère des sons de STA –, elle a l'avantage de procurer des signaux enregistrés neutres et permet de conserver la nature intacte du son tel qu'il est émis. De plus, les enregistrements obtenus sont plus facilement « manipulables » à posteriori. En effet, il est plus facile de passer par simulation d'un son enregistré en condition anéchoïque à un son représentant la façon dont il serait perçu dans une situation (non-anéchoïque) particulière – on parle d'*auralisation* (voir section 8.2) –, que de remplacer l'effet d'une condition particulière par une autre.

La salle semi-anéchoïque dans laquelle ont eu lieu les enregistrements était une cabine, isolée du sol, mesurant 3,4 m sur 4,2 m. La hauteur était approximativement 2,7 m. Cette cabine était équipée d'une porte, recouverte du même revêtement absorbant que les parois internes, au centre du côté le plus long, et d'une ouverture dans un coin du côté le moins long afin de faire passer les câbles, à la fois pour les STA et pour les microphones. Cette dernière ouverture était toutefois bouchée lors de l'utilisation par un morceau du même revêtement absorbant que celui recouvrant les parois. À partir de 200 Hz, le temps de réverbération se situe aux alentours de 0,1 s, voire moins, et, en dessous de cette fréquence, il ne dépasse pas 0,25. La fréquence de coupure de la salle (ou fréquence de Schroeder [135]) est de 160,5 Hz.

4.1.2 Matériel d'enregistrement utilisé

La photo en figure 4.1 présente la configuration d'enregistrement utilisée pour chaque enregistrement.

FIGURE 4.1 – Configuration des microphones pour l'enregistrement.

Pour chaque STA considéré, plusieurs types d'enregistrements ont été effectués :

– Deux prises monophoniques ont été enregistrées par des microphones à directivité cardioïde (afin de limiter l'influence des bruits extérieurs, l'isolation de la cabine n'étant pas parfaite, et de la réverbération résiduelle). Le modèle de microphone utilisé est le C451B de la marque AKG. Les deux prises effectuées avaient pour but d'enregistrer respectivement le son de *diffusion* d'air (où ce dernier est « éjecté » du STA) et le son de *reprise* d'air (où il est « aspiré »), le microphone pour ce dernier n'apparaissant pas sur la photo en figure 4.1. Ils ont été chacun placés en conséquence dans la pièce, sur un trépied, à une distance approximative de 1 m de la bouche de diffusion ou de reprise d'air. Le microphone pour la prise du son de diffusion d'air était systématiquement orienté à l'horizontale et placé à 1,5 m su sol, tandis que l'orientation et la position du second microphone dépendaient du type de système enregistré (voir détail en sections 4.2.1, 4.2.2 et 4.2.3). Ils étaient maintenus par un système de pince amortissante, limitant la transmission des vibrations entre le trépied et chaque microphone.

– Une prise stéréophonique a été enregistrée selon la technique *ORTF*. Cette technique met en jeu deux microphones cardioïdes placés de sorte qu'ils forment un angle de 110° et que les capsules des microphones soient distantes de 17 cm. Les mêmes modèles de microphone que pour les enregistrements monophoniques (AKG C451B) ont été utilisés. L'installation ORTF était placée juste en dessous du microphone monophonique enregistrant le son de diffusion d'air (voir figure 4.1). Ce système de microphones était placé à une hauteur de 1,4 m. Le son de reprise d'air n'a pas été enregistré avec cette technique. Par ailleurs, il était en pratique impossible d'utiliser la configuration ORTF avec les mêmes pinces amortissantes que pour les enregistre-

ments monophoniques, qui n'ont donc pas été utilisées.

– Un prise binaurale a également été réalisée à l'aide d'une tête artificielle du modèle Neumann KU100, placée sous l'installation ORTF (voir figure 4.1) afin d'enregistrer le son de diffusion d'air. La tête artificielle était placée face au système, à 1,2 m du sol. Le son de reprise d'air n'a pas été enregistré par cette technique. L'enregistrement binaural a pour but de prendre en compte lors de l'enregistrement l'influence de la présence de la tête d'une personne, voire de son torse (pour certains modèles de tête artificielle autres que celui utilisé ici), sur le son qu'il perçoit. En effet, lorsque l'on écoute directement la source sonore, le son qui parvient jusqu'à nos oreilles est modifié par ces éléments de notre corps, tandis que lorsque l'on écoute le son stéréophonique enregistré classiquement (technique ORTF ou autre) au travers d'un casque, celui-ci ne retranscrit pas l'influence de nos membres. Cela signifie que les enregistrements binauraux sont exclusivement réservés à une reproduction du son au casque. En effet, dans le cas d'une reproduction de tels enregistrements sur haut-parleur, l'influence des membres humains sera mise en jeu deux fois, une fois par la tête artificielle et une fois par la tête de l'auditeur. On préférera alors une technique d'enregistrement stéréophonique classique.

Tous les microphones positionnés pour enregistrer le son de diffusion d'air étaient légèrement décalés par rapport au flux d'air de sortie de l'appareil afin d'éviter les perturbations liées à l'écoulement d'air autour des capsules microphoniques. Par ailleurs, un morceau de moquette de 1,5 m sur 1,5 m environ était placé sur le sol sous les systèmes d'enregistrement, afin de limiter l'effet des réflexions les plus proches des microphones.

Enfin, tous les enregistrements, d'une dizaine de minutes environ, ont été effectués avec un ordinateur portable équipé du système d'exploitation Windows XP, et une interface audio externe RME Fireface 400 (voir figure 4.2), à une fréquence d'échantillonnage de 88,2 kHz, et 24 bits de quantification. Cette dernière ne propose que deux entrées analogiques préamplifiées, pour six autres entrées analogiques. Étant donné que tous les microphones utilisés (tête binaurale comprise) nécessitent une préamplification, un préamplificateur analogique RME Quadmic (voir figure 4.2) fournissant pour quatre canaux une préamplification similaire à celle de l'interface audio a également utilisé. Les entrées préamplifiées de la carte ont été utilisées pour les microphones de la tête binaurale, tandis que le préamplificateur externe a été utilisé pour les prises monophoniques et stéréophoniques ORTF. Ces éléments de prise de son ont été étalonnés à l'aide d'un signal de calibration fourni par un pistonphone (sinusoïde à 1000 Hz et 94 dB SPL).

FIGURE 4.2 – Photo de l'interface audio RME Fireface 400 (en bas) et du préamplificateur RME Quadmic (en haut).

4.2 Enregistrement pour les différents types de STA

Parmi l'ensemble des STA enregistrés, on peut identifier trois catégories :
- les systèmes *carrossés*, utilisés « en allège », c'est-à-dire fixés contre un mur ou contre le plafond ;
- les systèmes *gainables*, généralement installés en faux-plafond, raccordés à des interfaces de diffusion ou de reprise d'air par des gaines ;
- les systèmes type « *cassette* », semi-carrossés, le système étant installé dans le plafond directement derrière la grille de reprise et/ou de diffusion d'air.

Les trois sections suivantes détaillent les enregistrements pour ces trois types de système.

4.2.1 Systèmes carrossés, au sol

Quatre systèmes de ce type ont été enregistrés. Ces systèmes étaient posés au sol sur des supports en bois pour remédier à une irrégularité notable du sol, et parce que la prise d'air sur ce type de machines se fait par le dessous (voir figures 4.3 et 4.4). Les systèmes étaient disposés à distance des parois et en biais par rapport à la pièce (c'est-à-dire qu'ils n'étaient pas parallèles à un mur de la pièce), afin d'obtenir une densité d'énergie acoustique la plus homogène possible au sein de la salle anéchoïque. Le microphone pour le son de reprise d'air était placé à l'arrière du système à environ 50 cm du sol. Chacun de ces systèmes a été enregistré à 3 vitesses différentes de fonctionnement (voir tableau D.1 en annexe pour plus de détails).

4.2.2 Systèmes gainables, suspendus

Neuf systèmes de ce type ont été enregistrés. Ces systèmes, censés être installés dans les faux-plafonds, étaient suspendus à environ 1,8 m du sol sur une armature métallique utilisée pour l'occasion (voir figures 4.5, 4.6, 4.7 et 4.8). Les systèmes étaient disposés à distance des parois et en biais par rapport à la pièce (c'est-à-dire qu'ils n'étaient pas parallèles à un mur de la pièce), afin d'obtenir une densité d'énergie acoustique la plus homogène possible au sein de la salle anéchoïque. La reprise d'air étant faite latéralement pour ces systèmes, le microphone correspondant était placé à la même hauteur que ceux-ci (1,8 m) en face de la grille de reprise. Chacun de ces systèmes a été enregistré à 3 vitesses différentes de fonctionnement à l'exception du système **GNB2** qui n'a été enregistré qu'aux 2 vitesses les plus élevées, car les vitesses inférieures étaient trop silencieuses (voir tableau D.1 en annexe pour plus de détails).

4.2.3 Systèmes « cassettes », suspendus

Trois systèmes de ce type ont été enregistrés. Ces systèmes, censés être installés dans les faux-plafonds, étaient suspendus à environ 1,8 m du sol sur une armature métallique utilisée pour l'occasion (voir figures 4.9 et 4.10). Les systèmes étaient disposés à distance des parois et en biais par rapport à la pièce (c'est-à-dire qu'ils n'étaient pas parallèles à un mur de la pièce), afin d'obtenir une densité d'énergie acoustique la plus homogène possible au sein de la salle anéchoïque. Il est à noter que, contrairement aux deux autres types de systèmes, la grille de reprise d'air de ces STA était orientée vers le bas. En conséquence, le microphone dédié était placé en-dessous du système et orienté vers le haut. Chacun de ces systèmes a été enregistré à 3 vitesses différentes de fonctionnement à l'exception du système **CST3** qui n'a été enregistré qu'aux 2 vitesses les plus élevées, car les vitesses inférieures étaient trop silencieuses (voir tableau D.1 en annexe pour plus de détails).

FIGURE 4.3 – Photo de l'installation pour l'enregistrement du son du système **CRS1** (voir tableau D.1 en annexe).

FIGURE 4.4 – Photo de l'installation pour l'enregistrement du son du système **CRS3** (voir tableau D.1 en annexe).

FIGURE 4.5 – Photo de l'installation pour l'enregistrement du son du système **GNB1** (voir tableau D.1 en annexe).

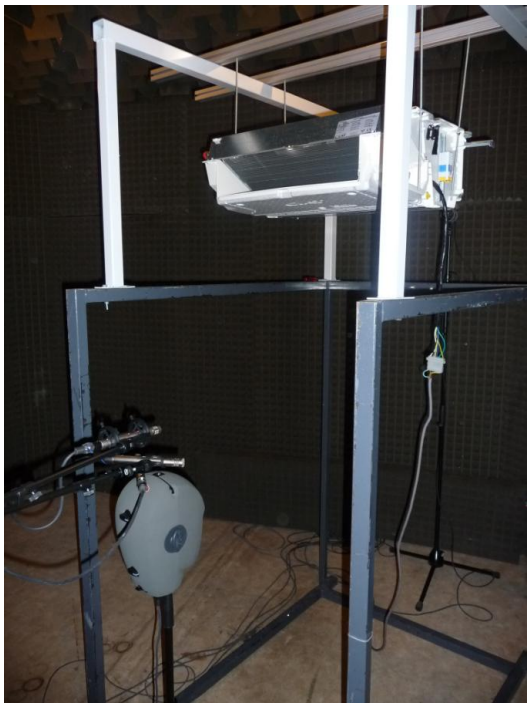


FIGURE 4.6 – Photo de l'installation pour l'enregistrement du son du système **GNB2** (voir tableau D.1 en annexe).

FIGURE 4.7 – Photo de l'installation pour l'enregistrement du son du système **GNB4** (voir tableau D.1 en annexe).

FIGURE 4.8 – Photo de l'installation pour l'enregistrement du son du système **GNB5** (voir tableau D.1 en annexe).

FIGURE 4.9 – Photo de l'installation pour l'enregistrement du son du système **CST1** (voir tableau D.1 en annexe).

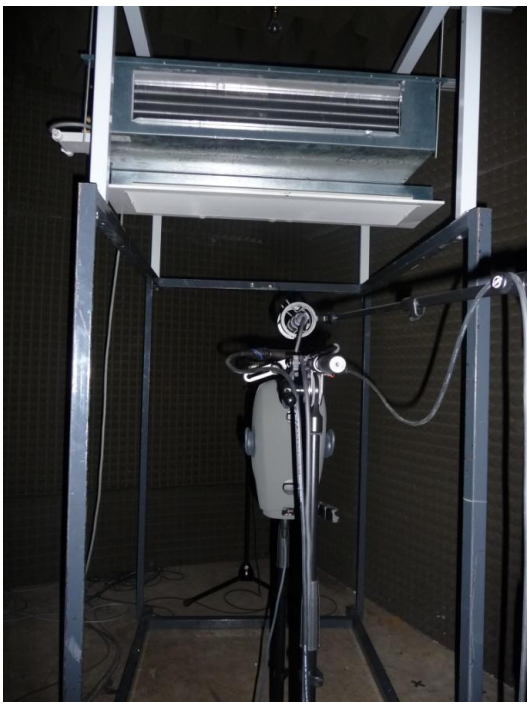


FIGURE 4.10 – Photo de l'installation pour l'enregistrement du son du système **CST2** (voir tableau D.1 en annexe).

4.3 Discussion

Cette session de mesure nous a donc permis d'enregistrer 16 STA différents. Chaque système a été enregistré à 2 ou 3 vitesses différentes, et, dans le cas des enregistrements monophoniques uniquement, à deux positions différentes (reprise et diffusion d'air). Si l'on considère séparément les enregistrements monophoniques, stéréophoniques et binauraux, nous avons trois corpus sonores composés respectivement de 92 sons monophoniques, 46 sons stéréophoniques et 46 sons binauraux (voir tableau D.1 en annexe pour plus de détails). Ces corpus sont assez représentatifs des types de STA actuellement disponibles sur le marché et des différentes conditions de fonctionnement possibles, même si les appareils enregistrés ici proviennent pour la plupart du même fabriquant.

Il est à noter toutefois que, dans un contexte réaliste d'installation du type particulier des STA gainables, les circuits aérauliques de reprise et de diffusion d'air (liés à l'utilisation de gaines) induisent un régime de fonctionnement différent de celui obtenu dans les conditions où les sons correspondants ont été enregistrés. Cependant, une telle installation était en pratique irréalisable dans le cadre de cette campagne de prises de sons, et nous avons souhaité garder une forme de cohérence avec les enregistrements réalisés pour les autres types de STA, notamment par rapport au contexte de salle (semi-)anéchoïque.

En outre, les enregistrements stéréophoniques sont théoriquement destinés à une reproduction sur haut-parleurs, quand les enregistrements binauraux sont eux censés être reproduits dans un casque audio. Étant donné qu'aucune expérience présentée dans cette thèse n'a été réalisée avec l'utilisation de haut-parleurs (à l'exception d'une expérience utilisant une auralisation sur haut-parleurs, mais nécessitant des enregistrements monophoniques, voir section 8.2), il n'était donc pas pertinent d'utiliser ces enregistrements. Leur réalisation était plutôt liée ici à la volonté de constituer une base de donnée d'enregistrements de STA à l'aide de différentes techniques.

Quant aux enregistrements binauraux, ils ont été utilisés uniquement lors d'une expérience perceptive d'évaluations du désagrément des sons de STA, en condition de sonie réelle (voir section 7.2). À cette exception près, l'ensemble des travaux présentés ici ont portés sur les enregistrements monophoniques, et ceci deux raisons principales :

- La très grande majorité des descripteurs acoustiques et psychoacoustiques existants, et notamment ceux présentés en section 2.3.3 que l'ont tente ici de relier à la qualité sonore perçue, sont définis sur un signal sonore monophonique. Ils ne prennent pas en compte l'éventualité d'un signal multicanal. Bien qu'il soit évidemment possible d'évaluer ces descripteurs pour les deux oreilles, il n'existe pas de règles unanimement définies concernant la façon d'intégrer les couples de descripteurs obtenus.
- Un des objectifs principaux est d'étudier l'influence de paramètres d'acoustique des salles sur la qualité sonore perçue. Dans le cadre de cette étude, il a été entrepris de comparer les résultats obtenus en condition anéchoïque – où donc, par définition, le même signal sonore doit être perçu par les deux oreilles, pourvu que l'on suppose la position de la source dans le plan médian de l'auditeur [1] – avec ceux obtenus en condition auralisée, sachant que l'algorithme d'auralisation utilisé pour générer les sons auralisés est applicable uniquement sur des sons monophoniques (voir chapitre 8 et la section 8.2 plus particulièrement).

1. La localisation de la source ne fait pas partie des problématiques abordées dans le cadre de cette thèse.

Chapitre 5

Identification des familles de sons de STA : constitution du corpus de travail

Ce chapitre présente l'étude réalisée afin, dans un premier temps, d'identifier les différentes familles de sons qui constituent l'ensemble des enregistrements réalisés, et décrit dans le chapitre 4, et dans un second temps, de sélectionner un corpus réduit de travail. La section 5.1 présente tout d'abord la problématique générale de cette étude. La section 5.2 présente le protocole expérimental utilisé afin de réaliser une expérience de catégorisation libre. La section 5.3 présente l'analyse des résultats de cette expérience en termes de classification hiérarchique des familles de sons. Enfin, la section 5.4 expose la discussion que soulèvent les résultats de cette étude, en termes de familles de sons ainsi identifiées et de corpus réduit sélectionné.

5.1 Problématique

La campagne de mesure décrite dans le chapitre 4 nous a permis d'obtenir un ensemble de 92 enregistrements monophoniques de sons de STA (les enregistrements stéréophoniques et binauraux ne sont pas considérés ici pour les raisons évoquées en section 4.3). Cet ensemble est relativement représentatif de la variété de STA et de conditions de fonctionnement que l'on peut rencontrer. Toutefois, au regard des études publiées dans la littérature sur l'évaluation de la qualité des sons de l'environnement (voir section 2.3), on s'aperçoit que les méthodologies expérimentales utilisées nécessitent, pour des raisons pratiques de réalisation des expériences, des effectifs bien inférieurs. La plupart d'entre elles sont conduites sur des corpus incluant entre 10 et 20 éléments.

Par ailleurs, les enregistrements réalisés incluent plusieurs prises (entre 4 et 6) de mêmes STA, à différentes positions et à différentes vitesses de fonctionnement (voir tableau D.1 en annexe). Il est donc fort probable que beaucoup d'entre eux présentent une forte similarité et que l'ensemble de sons ne soit pas exempt de redondances.

Pour ces raisons, il est nécessaire de réduire notre corpus à un échantillon d'une taille similaire à la taille des corpus utilisés dans la littérature et suffisamment représentatif de la variété de sons enregistrés. Il est importe également que la représentativité de l'échantillon recherché soit établie d'une manière qui soit valide perceptivement. Pour ce faire, nous devons donc utiliser une démarche d'identification des principales familles de sons constituant notre corpus de travail. Ceci peut être réalisé au moyen d'une expérience dite de catégorisation libre, qui consiste à regrouper les sons en familles (catégories) en fonction de leur similarité, les critères de similarité et la structure du regrou-

pement (nombre et taille des catégories) étant laissés libres au participant. Cette opération permettra à terme d'échantillonner les différentes catégories moyennes (sur l'ensemble des participants) afin de construire un corpus réduit représentatif de la variété de sons à disposition.

Cette stratégie peut paraître contradictoire avec certaines conclusions de l'étude bibliographique réalisée qui concerne l'identification de la source sonore (voir section 2.2). Ces conclusions stipulent notamment qu'une structure catégorielle de description des sons, telle que l'on obtient communément à l'issue d'une expérience de catégorisation libre, convient pour décrire l'identification du type de source sonore, lorsque les sons du corpus considéré correspondent à différentes causes physiques. À l'inverse, lorsque les sons correspondent à un type unique de source physique, ce qui est le cas ici, une structure de description continue et multidimensionnelle est adéquate. Cependant, outre le fait que ce type de description représente également un objectif de cette thèse et fait l'objet du chapitre 6, il est obtenu le plus souvent au travers d'une étude du timbre (voir section 2.1), incluant par exemple une expérience de mesure de similarités. Ce type d'étude, tout comme pour l'étude de la qualité sonore – dont elle fait souvent partie –, implique l'obtention d'un corpus de travail dont l'effectif se situe d'habitude également entre 10 et 20 sons. Ainsi, il est, d'un point de vue pratique, indispensable d'identifier les familles de sons constituant notre corpus pour réaliser les travaux subséquents.

5.2 Protocole expérimental de catégorisation libre

5.2.1 Stimuli

Nous disposons de 92 sons monophoniques de STA (voir section 4.3 et tableau D.1 en annexe). Toutefois, bien qu'une expérience de catégorisation libre puisse être conduite sur un nombre plus élevé de sons que beaucoup d'autres types d'expérience (et plus particulièrement ceux décrits dans les chapitres suivants), cet effectif est encore trop important. Il serait raisonnable de le réduire de moitié, soit une cinquantaine d'éléments.

Toutefois, compte tenu de la présence d'un léger bruit de fond dans les enregistrements effectués à l'aide des microphones AKG (voir section 4.2), une quinzaine de sons environ présentaient un rapport signal-sur-bruit trop faible, au point de rendre le bruit de fond audible au travers du son de STA (notamment à basses vitesses). Ils ont par conséquent été retirés du corpus.

Parmi les sons restants, nous devons en sélectionner une cinquantaine. L'idée étant principalement de comparer les différents STA (au nombre 16), il nous a semblé raisonnable de sélectionner trois enregistrements différents pour chaque STA, pour porter le total à 48 sons. La sélection a été effectuée de sorte à avoir pour chaque STA au moins un enregistrement à chacune des deux positions (reprise et diffusion d'air), et au moins deux vitesses différentes de fonctionnement. De plus, nous souhaitions avoir au total une représentation équivalente des enregistrements à une position par rapport à l'autre.

Enfin, étant donné que la sonie est l'attribut auditif prépondérant de la perception comme le montrent les études de qualité sonore mentionnées en section 2.3, il est fort probable que les regroupements de sons seront principalement dirigés par le percept de sonie si les sons manifestent ne serait-ce que de faibles variations de ce paramètre. Compte tenu des différences de niveau sonore émis par les STA enregistrés, un tel phénomène est très susceptible de se produire avec les sons obtenus.

Or, il est important ici que les regroupements de sons ne se fassent pas sur la base du niveau sonore ; on s'intéresse plutôt aux différentes catégories de timbre. Le corpus de sons a donc subi une

égalisation en sonie, c'est-à-dire qu'un facteur multiplicatif a été appliqué à chaque signal sonore de sorte à ce que tous les sons soient perçus avec le même niveau sonore. Toutefois, compte tenu de la taille du corpus en question, il semble difficilement envisageable d'obtenir ce jeu de facteurs à l'aide d'une égalisation en sonie par la procédure expérimentale décrite en section 8.3.3, où on demande aux participants de régler le niveau de chaque son afin qu'il soit perçu comme aussi « fort » qu'un son de référence [1]. En conséquence, le corpus à été égalisé en sonie par rapport au modèle de sonie de Zwicker et Fastl [168]. Cela signifie que la sonie n'est pas ici « mesurée » par des auditeurs, mais calculée à partir du modèle.

La sonie étant un paramètre dont la relation avec le signal sonore est notablement non-linéaire, il est presque impossible de trouver analytiquement le facteur multiplicatif entre deux signaux audio permettant de les égaliser en sonie. Une procédure itérative a donc été mise au point afin de faire converger la sonie – au sens du modèle de Zwicker et Fastl – d'un son à égaliser N_{egal} vers celle d'un son de référence N_{ref}. Le son de référence choisi – **GNB2 v1 dif**, voir tableau 5.1 – est celui dont le maximum de la valeur absolue du signal était le plus faible, afin d'éviter tout problème de *clipping*[2] lors de l'application de facteurs multiplicatifs. À chaque pas d'itération, le signal audio du son à égaliser est multiplié par le rapport N_{ref}/N_{egal}. L'itération s'arrête lorsque l'on parvient par cette méthode à un ratio de sonie tel que $|1 - N_{ref}/N_{egal}| \leq 0.01$. Il a par ailleurs été vérifié par l'expérimentateur à la suite de cette procédure qu'aucun des sons du corpus ne présentait de différence de sonie manifeste à l'oreille avec les autres.

Compte tenu de la nature relativement stationnaire des sons enregistrés, la durée des échantillons est de 5 secondes. Le détail du corpus égalisé en sonie est exposé dans le tableau 5.1. Il s'agit donc de 48 sons monophoniques égalisés en sonie, d'une durée fixe de 5 secondes, et dont le niveau acoustique varie de 38,1 à 40,4 dBA après égalisation en sonie.

5.2.2 Participants

21 auditeurs volontaires (15 hommes, 6 femmes, entre 20 et 29 ans), n'ayant pas fait mention d'un problème majeur d'audition, et n'ayant pas participé à une autre expérience réalisée dans le cadre de cette thèse, ont pris part à cette expérience de catégorisation libre.

5.2.3 Matériel

L'expérience a été réalisée grâce à une interface graphique programmée en MATLAB programmée par Vincent Rioux à l'IRCAM (figures 5.1 et 5.2). Les sons ont été diffusés par une interface RME Fireface 400 dans un casque ouvert Sennheiser HD650 [3]. Le mode de reproduction sonore sur casque a été employé car c'est le seul qui garantit d'obtenir une écoute diotique (signal identique dans les deux oreilles). L'expérience a eu lieu dans une salle traitée acoustiquement pour être isolée des bruits extérieurs.

1. En effet, les participants auraient alors à effectuer 47 réglages (un son est pris comme référence et n'est donc pas à égaliser) ce qui représente une tâche assez lourde pour une expérience que l'on peut qualifier de subsidiaire.

2. Saturation du signal quand les valeurs de celui-ci dépassent la gamme de valeurs autorisée par son format ($\{-1;1\}$ dans la cas d'un fichier wave)

3. Un casque ouvert est un casque conçu pour laisser passer l'air, par opposition au casque fermé qui est étanche à l'air et isole l'auditeur des sources de bruit extérieures. Dans une même gamme de produits, les casques ouverts sont souvent considérés comme plus précis.

	nom du système	vitesse	position	identifiant
carrossés	**CRS1**	1	diffusion	**CRS1 v1 dif**
	CRS1	2	diffusion	**CRS1 v2 dif**
	CRS1	2	reprise	**CRS1 v2 rep**
	CRS2	1	reprise	**CRS2 v1 rep**
	CRS2	2	diffusion	**CRS2 v2 dif**
	CRS2	5	reprise	**CRS2 v5 rep**
	CRS3	1	diffusion	**CRS3 v1 dif**
	CRS3	1	reprise	**CRS3 v1 rep**
	CRS3	3	reprise	**CRS3 v3 rep**
	CRS4	2	diffusion	**CRS4 v2 dif**
	CRS4	2	reprise	**CRS4 v2 rep**
	CRS4	3	diffusion	**CRS4 v3 dif**
gainables	**GNB1**	3	reprise	**GNB1 v3 dif**
	GNB1	3	diffusion	**GNB1 v3 rep**
	GNB1	6	reprise	**GNB1 v6 dif**
	GNB2	1	diffusion	**GNB2 v1 dif**
	GNB2	1	reprise	**GNB2 v1 rep**
	GNB2	2	reprise	**GNB2 v2 rep**
	GNB3	2	diffusion	**GNB3 v2 dif**
	GNB3	2	reprise	**GNB3 v2 rep**
	GNB3	3	diffusion	**GNB3 v3 dif**
	GNB4	1	diffusion	**GNB4 v1 dif**
	GNB4	2	diffusion	**GNB4 v2 dif**
	GNB4	2	reprise	**GNB4 v2 rep**
	GNB5	2	diffusion	**GNB5 v2 dif**
	GNB5	2	reprise	**GNB5 v2 rep**
	GNB5	3	reprise	**GNB5 v3 rep**
	GNB6	1	diffusion	**GNB6 v1 dif**
	GNB6	1	reprise	**GNB6 v1 rep**
	GNB6	2	diffusion	**GNB6 v2 dif**
	GNB7	2	reprise	**GNB7 v2 rep**
	GNB7	4	diffusion	**GNB7 v4 dif**
	GNB7	4	reprise	**GNB7 v4 rep**
	GNB8	2	diffusion	**GNB8 v2 dif**
	GNB8	2	reprise	**GNB8 v2 rep**
	GNB8	3	reprise	**GNB8 v3 rep**
	GNB9	2	diffusion	**GNB9 v2 dif**
	GNB9	4	diffusion	**GNB9 v4 dif**
	GNB9	4	reprise	**GNB9 v4 rep**
cassette	**CST1**	2	diffusion	**CST1 v2 dif**
	CST1	3	diffusion	**CST1 v3 dif**
	CST1	3	reprise	**CST1 v3 rep**
	CST2	1	diffusion	**CST2 v1 dif**
	CST2	1	reprise	**CST2 v1 rep**
	CST2	2	reprise	**CST2 v2 rep**
	CST3	1	diffusion	**CST3 v1 dif**
	CST3	1	reprise	**CST3 v1 rep**
	CST3	2	reprise	**CST3 v2 rep**

TABLE 5.1 – Corpus sonore utilisée dans l'expérience de catégorisation libre (voir aussi tableau D.1 en annexe).

5.2.4 Procédure

Au début de l'expérience, les participants ont reçu une consigne écrite, retranscrite en section E.1 en annexe, leur expliquant le contexte de l'étude, la nature des sons étudiés et la tâche à accomplir. Cette dernière consistait à regrouper les sons en fonction de leur ressemblance acoustique. Sur l'interface de cette expérience (voir figure 5.1), les 48 sons sont représentés par des points rouges disposés et numérotés de manière aléatoire de 1 à 48 (la numérotation aléatoire étant toutefois commune à tous les participants). Le participant peut à sa guise écouter les sons dans l'ordre qu'il souhaite et autant de fois qu'il le désire en double-cliquant sur les points rouges. La tâche à réaliser est alors de regrouper graphiquement les points rouges en « paquets » représentant les différentes familles de sons identifiées, et de déclarer chacune d'elle afin qu'elles apparaissent dans la liste sur la gauche de l'interface (voir figure 5.2 en guise d'aperçu de l'interface en fin d'expérience).

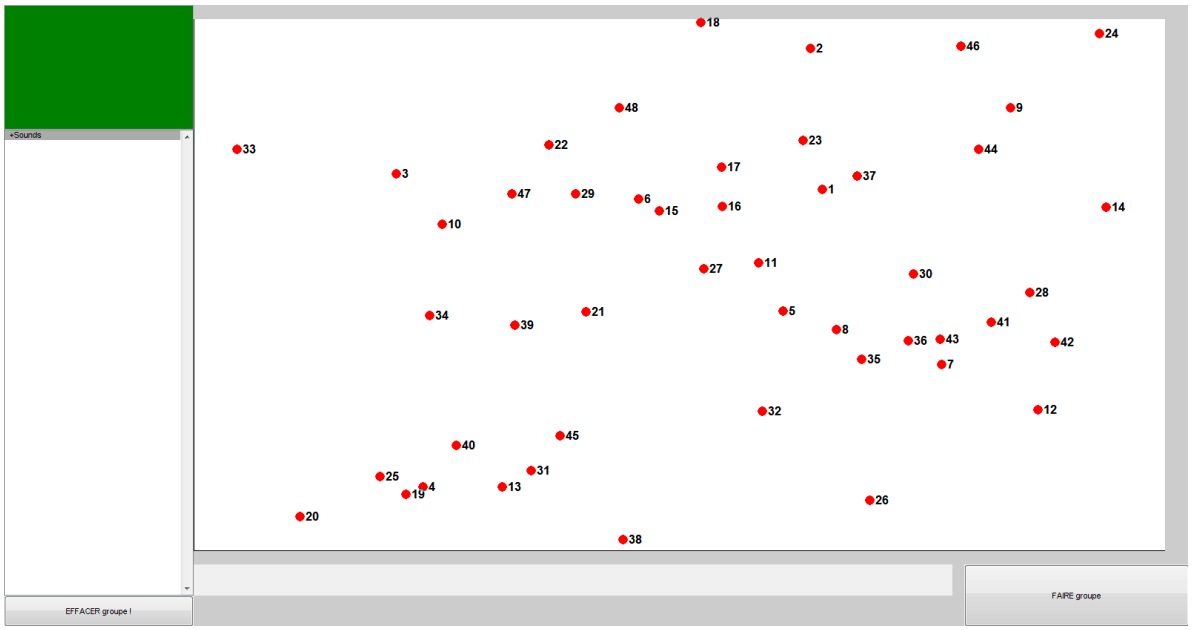


FIGURE 5.1 – Interface graphique de l'expérience de catégorisation libre dans son état initial. Les points rouges cliquables et déplaçables représentent les 48 sons. Les boutons en bas à droite et à gauche permettent respectivement de déclarer un groupe et d'en annuler la sélection.

Il doit être précisé ici que chaque son (point rouge) doit être placé dans une famille (« paquet ») et une seule. Cela signifie que la classification demandée à chaque participant n'est constituée que d'un seul niveau hiérarchique (pas de « sous-familles » de sons pour un même participant), bien que la consigne ne le mentionne pas en ces termes par souci de clarté. En d'autres termes, on demande à chaque participant une « liste » de familles de sons, non-hiérarchisée. La hiérarchie qui sous-tend la structure perceptive des familles de sons apparaîtra grâce à l'analyse subséquente des résultats de l'ensemble du panel de participants.

La durée moyenne de la procédure dans le cas de cette expérience a été de 33 minutes.

5.3 Résultats et analyse

Les données de sortie de l'expérience sont, pour chaque participant, une liste de regroupements de sons du corpus. Étant donné que tous les participants, non seulement n'ont pas nécessairement

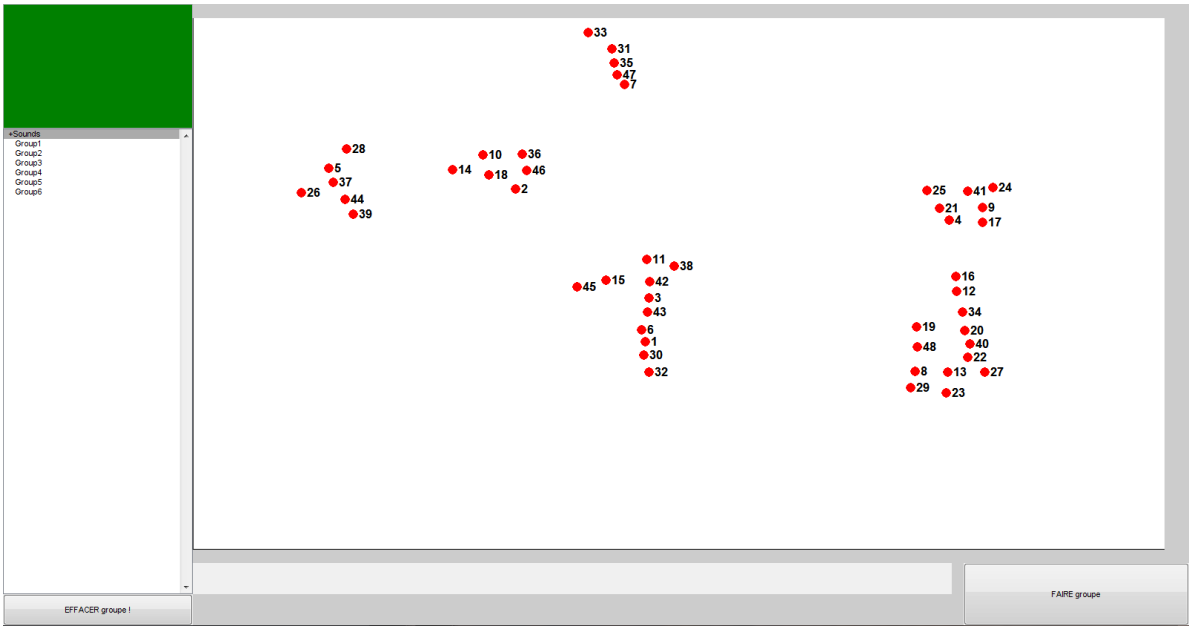


FIGURE 5.2 – Interface graphique de l'expérience de catégorisation libre dans son état final pour un des participants. Les points rouges ont été regroupés en paquets à l'écran et la partie gauche affiche désormais la liste des familles identifiées.

identifié les mêmes catégories, mais n'ont pas nécessairement non plus identifié le même nombre de catégories, il est tout d'abord indispensable d'écrire ces données sous une forme plus exploitable et plus globalisable sur l'ensemble des participants. Pour ce faire, les regroupements sont traités par paire de sons : pour chaque paire de sons i et j, on place dans une matrice M_{coo} de taille 48x48 dans la case (i, j) (et dans la case (j, i)) un 1 si les deux sons ont été placés dans le même groupe par le participant, un 0 sinon. Ainsi, pour chaque participant, nous obtenons une matrice binaire symétrique, appelée *matrice de co-occurrences*. Les matrices individuelles sont alors moyennées sur l'ensemble des participants pour obtenir une matrice de co-ocurrences moyenne M_{coo_m} dont on calcule finalement la matrice complémentaire : $1 - M_{coo_m}$. La raison d'être de cette dernière opération est que cela nous permet de construire une matrice de « distances » $M_{dis} = 1 - M_{coo_m}$, la distance entre deux sons étant alors définie comme égale au nombre de fois qu'ils ont été placés dans deux catégories distinctes par un participant divisé par le nombre total de participants.

5.3.1 Cohérence inter-participant

L'analyse qui suit est faite sur une matrice issue du moyennage sur le panel de participants. Avant cette opération, il convient de comparer les résultats des participants entre eux. En effet il importe de vérifier à quel point les résultats individuels sont compatibles entre eux, sans quoi le moyennage pourrait aboutir à des résultats peu représentatifs. Lorsque l'on souhaite analyser la cohérence inter-participant, on envisage naturellement l'usage du coefficient de corrélation de Bravais-Pearson (voir section B.1 en annexe) entre les résultats (ici les jeux individuels de co-occurrences entre sons) de chaque paire de participants.

Toutefois, compte tenu du fait que les données individuelles sont binaires ('1' si les sons ont été placés dans le même groupe, '0' sinon), le coefficient de corrélation de Bravais-Pearson est très mal adapté, et les valeurs que l'on va obtenir seront très faibles. Afin de comparer plusieurs jeux de classification d'objets, il convient plutôt d'utiliser l'*Indice de Rand RI*. Celui-ci se définit, entre deux matrices

individuelles de co-ocurrences M_{coo_1} et M_{coo_2}, comme suit :

$$RI = \frac{a+b}{a+b+c+d} \tag{5.1}$$

avec a le nombre de paires d'objets classés ensemble par les deux participants ('1' dans M_{coo_1} et dans M_{coo_2}), b le nombre de paires d'objets classés séparément par les deux participants ('0' dans M_{coo_1} et dans M_{coo_2}), c le nombre de paires d'objets classés ensemble par le participant 1 et séparément par le participant 2 ('1' dans M_{coo_1} et '0' dans M_{coo_2}), et d le nombre de paires d'objets classés séparément par le participant 1 et ensemble par le participant 2 ('0' dans M_{coo_1} et '1' dans M_{coo_2}). La somme $a+b$ représente donc le nombre d'accords entre les deux participants, tandis que la somme $c+d$ représente le nombre de désaccords. De plus, la somme $a+b+c+d$ correspond au nombre total de paires d'objets. La valeur de l'Indice de Rand se situe donc entre '0', si les deux participants sont en désaccord sur chaque paire d'objets , et '1', si les deux classifications sont parfaitement identiques.

	RI
moyenne μ	0,76
écart-type σ	0,05
min	0,62
max	0,81

TABLE 5.2 – Statistiques de l'Indice de Rand RI pour l'ensemble des paires de participants.

L'Indice de Rand a donc été évalué pour chaque paire de participants de l'expérience de catégorisation libre exposée ici. Le tableau 5.2 présente les statistiques obtenues de ce coefficient, montrant une concordance générale modérée. On peut toutefois s'interroger sur le comportement de ce coefficient vis-à-vis de la compatibilité hiérarchique entre les classifications de deux participants, c'est-à-dire le fait que celles-ci peuvent être différentes, si elles correspondent à deux niveaux de précision différents, sans pour autant être incompatibles. Prenons un exemple extrême pour illustrer ce propos. Supposons que deux participants voient globalement la même organisation dans le corpus étudié, mais ne se placent pas au même niveau hiérarchique ; les catégories identifiées par l'un sont des regroupements des catégories identifiées par l'autre. La figure 5.3 présente un exemple simple de ce type de situations, sur 8 éléments numérotés de 1 à 8. Le premier participant a regroupé les sons en 4 catégories (1 avec 2, 3 avec 4, etc ...), tandis que le second a regroupé les sons en 2 catégories correspondant aux regroupements des catégories du premier participant (1, 2, 3 et 4 ; puis 5, 6, 7 et 8).

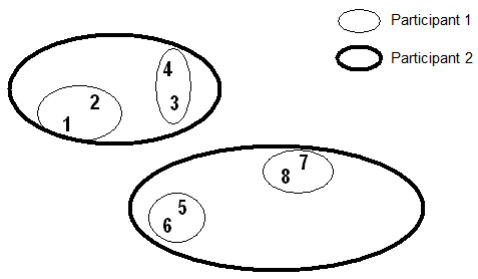


FIGURE 5.3 – Exemple de classification à des niveaux hiérarchiques différents.

Ces deux classifications peuvent être vues comme assez proches, puisqu'elles sont parfaitement compatibles en termes de hiérarchie. Cependant, lorsque l'on regarde la façon dont ces classifications

se traduisent dans les matrices de co-occurrences, ces dernières semblent très différentes (premier participant à gauche, second à droite) :

	1	2	3	4	5	6	7	8
1		1	0	0	0	0	0	0
2			0	0	0	0	0	0
3				1	0	0	0	0
4					0	0	0	0
5						1	0	0
6							0	0
7								1
8								

	1	2	3	4	5	6	7	8
1		1	1	1	0	0	0	0
2			1	1	0	0	0	0
3				1	0	0	0	0
4					0	0	0	0
5						1	1	1
6							1	1
7								1
8								

Seuls les '1' placés dans les cases (1,2), (3,4), (5,6) et (7,8) dans chacune des deux matrices traduisent la compatibilité hiérarchique entre les deux classifications. Ils indiquent que les paires correspondantes d'éléments ont bien été associées indépendamment par les deux participants. Ceci se confirme lorsque l'on évalue l'Indice de Rand entre ces deux classifications : on obtient $RI \simeq 0,71$, valeur peu élevée qui ne traduit donc pas leur compatibilité totale. Afin de rendre compte de la compatibilité hiérarchique, il importe donc de comparer le nombre de '1' communs aux deux matrices (a dans la définition de RI) au nombre de '1' placés dans la matrice du participant dont la classification est la plus détaillée (c'est-à-dire celle du participant qui a identifié le plus grand nombre de familles). Dans l'exemple donné, il s'agit de la classification du participant 1, et on observe que le nombre de '1' dans sa matrice de co-occurences et le nombre de '1' communs aux deux matrices sont identiques.

Afin de pouvoir généraliser et quantifier cette comparaison, un coefficient, que nous appellerons « coefficient de compatibilité hiérarchique » c_{ch}, a été mis au point afin de prendre en compte la compatibilité hiérarchique des résultats des participants deux à deux. Pour chaque paire de participants et donc chaque paire de matrices (triangulaires sans la diagonale) de co-occurrences, voici comment il est défini :

$$c_{ch} = \frac{n_{1\&2}}{\min(n_1, n_2)} \tag{5.2}$$

avec $n_{1\&2}$ le nombre de '1' en commun dans les deux matrices (identique au a de la définition de RI), et n_1 et n_2 les nombres de '1' apparaissant dans chacune des deux matrices. Le fait de diviser ici par le nombre minimum de '1' entre les deux participants permet d'identifier laquelle des deux classifications est la plus détaillée.

La valeur maximale '1' est atteinte lorsque les deux classifications sont parfaitement compatibles (notamment pour l'exemple ci-dessus), et pas uniquement lorsqu'elles sont totalement identiques – comme c'est le cas pour l'Indice de Rand. En revanche, ce coefficient baissera rapidement dés que les familles identifiées par les deux participants se « chevauchent », c'est-à-dire lorsque les sons ne sont pas placés dans les mêmes familles au niveau de classification le plus détaillé. La valeur minimale est '0' lorsque les deux participants ne présentent aucune paire en commun de sons classés ensembles.

Pour chaque participant, ce coefficient est moyenné sur l'ensemble des autres participants. Les statistiques du coefficient de compatibilité hiérarchique sont résumées dans le tableau 5.3. Les valeurs sont donc globalement plus élevées que celles de l'Indice de Rand, ce qui signifie qu'un certain nombre de classifications individuelles, bien que différentes entre elles, restent néanmoins compa-

tibles hiérarchiquement. Le coefficient de corrélation de Bravais-Pearson entre les valeurs de RI et de c_{ch} pour l'ensemble des participants est $r(ddl = 19) = 0,70$. Celui-ci montre que, si le lien entre ces deux coefficients est significatif ($p < 0,01$), la prise en compte de la compatibilité hiérarchique modifie d'une manière non-négligeable l'estimation réalisée de la cohérence inter-participant. La valeur moyenne $\mu = 0,90$ de c_{ch} est assez proche de la valeur idéale '1', et, sur l'ensemble des participants, les résultats semblent très proches ($\sigma = 0,02$). On remarque notamment qu'aucun participant n'obtient un coefficient significativement inférieur aux autres, et en particulier, inférieur à $\mu - 1,5\sigma$, ce qui correspond à un seuil de détection d'*outlier* souvent utilisé.

	c_{ch}
moyenne μ	0,90
écart-type σ	0,02
min	0,87
max	0,95

TABLE 5.3 – Statistiques du coefficient de compatibilité hiérarchique c_{ch} pour l'ensemble des participants.

En conséquence, l'ensemble des résultats des participants a été conservé pour la suite de l'analyse.

5.3.2 Analyse de cluster

La matrice de distances est alors traitée par une *analyse de cluster* (voir [98] pour plus de détails), dont le but est de construire une représentation hiérarchique du corpus. Les différents niveaux de hiérarchie sont obtenus grâce à la variabilité du nombre de catégories identifiées par chaque participant. Ce dernier point justifie alors le choix de laisser les participants libres de décider du nombre de catégories de sons qu'ils peuvent définir parmi le corpus de 48 sons. Il existe plusieurs méthodes de traitement de la matrice de distances M_{dis} permettant d'obtenir la représentation hiérarchique. Bien que ce ne soit pas celle qui a été utilisée, la méthode « simple » – *single linkage clustering* – est celle qui permet de comprendre le plus facilement comment la représentation hiérarchique est construite. Pour cette méthode, on procède comme suit : un premier cluster est formé par les deux éléments présentant la distance la plus faible (de toute la matrice de distances), puis on procède de même pour la distance la plus faible parmi les distances restantes, ... etc. Lorsqu'une distance met en jeu un nouvel élément et un élément déjà incorporé dans un cluster, le nouvel élément est lié au cluster en question pour formé un nouveau cluster avec un niveau de hiérarchie supplémentaire. On continue de cette manière jusqu'à ce que tous les éléments apparaissent dans la hiérarchie. Cela signifie bien souvent qu'une partie de la matrice de distances n'est pas utilisée (les distances les plus élevées). Au final, la distance séparant deux clusters (contenant un ou plusieurs éléments) est la distance entre les deux plus proches voisins, c'est-à-dire ceux qui présentent la distance la plus faible. D'autres méthodes existent. Par exemple, la méthode « complète » – *complete linkage clustering* – définit la distance entre deux clusters comme la distance entre les deux éléments les plus éloignés des deux clusters. La méthode qui a été utilisée ici est une méthode intermédiaire – *unweighted arithmetic average clustering (UPGMA*[4]*)* – qui définit la distance entre deux clusters comme la moyenne des distances entre chaque paire d'éléments des deux clusters. Cette méthode a l'avantage d'utiliser l'ensemble des données de la matrice M_{dis}. En outre, cette distance moyenne porte le nom de *distance cophénétique*, et modélise

4. Le sigle UPGMA signifie en réalité "Unweighted Pair-Group Method using Arithmetic averages", plus souvent dénommée "unweighted arithmetic average clustering", voire "average clustering" ou "average linkage clustering'.'

ainsi la distance originale – dans la matrice M_{dis} introduite au début de cette section – entre chacun des éléments des deux clusters.

5.3.3 Représentation en dendrogramme

Il est possible de représenter graphiquement la structure hiérarchique obtenue sous forme d'un *dendrogramme*. Dans le cas de l'exemple considéré précédemment, le dendrogramme serait tel qu'affiché en figure 5.4. Celui-ci peut être vu comme un « arbre » dont les feuilles représentent les 8 éléments de l'exemple (à l'extrémité gauche de l'arbre). Sur cet arbre, la distance cophénétique entre deux éléments (c'est-à-dire entre les clusters auxquels ils appartiennent) est donc la distance entre les feuilles correspondantes et le nœud où les branches issues de ces feuilles se rejoignent, également appelée *hauteur de fusion*. On voit alors clairement apparaître la hiérarchie que l'on devinait sur la figure 5.3 :

- les paires $(1,2)$, $(3,4)$, $(5,6)$ et $(7,8)$ forment chacune un cluster, et la distance cophénétique entre les deux éléments de chaque paire est nulle ;
- les clusters $(1,2)$ et $(3,4)$ sont séparés par une distance cophénétique de $0,5$ et forment ensemble un nouveau cluster d'un niveau hiérarchique supérieur ;
- il en est de même pour les paires $(5,6)$ et $(7,8)$;
- enfin, les clusters $(1,2,3,4)$ et $(5,6,7,8)$ sont séparés par une distance cophénétique de 1.

Il convient de préciser que dans cet exemple, volontairement simpliste, les valeurs des distances originales et de leur modélisation par la distance cophénétique sont identiques, ce qui n'est pas nécessairement vrai dans le cas général.

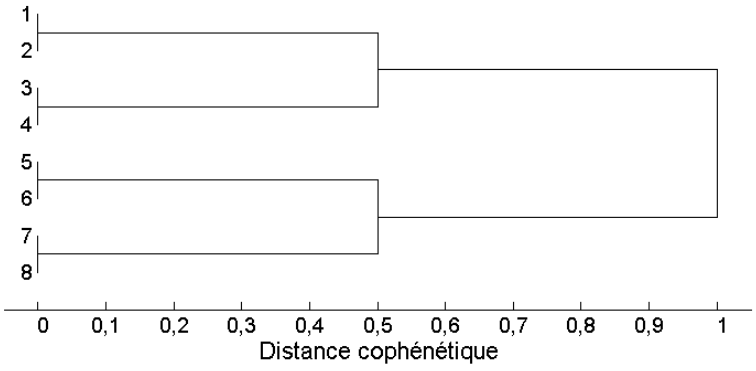


FIGURE 5.4 – Dendrogramme correspondant à l'exemple fictif à 8 éléments et 2 participants.

Dans le cas des résultats de l'expérience de catégorisation libre réalisée, le dendrogramme est affiché en figure 5.5. On peut interpréter cette figure de la même manière que pour l'exemple de la figure 5.4, et ainsi considérer globalement qu'un son est proche d'un autre s'ils sont voisins sur le dendrogramme, tout en tenant compte des hauteurs de fusion des clusters auxquels ils appartiennent.

5.3.4 Adéquation de la structure hiérarchique aux données

Il importe à présent d'évaluer à quel point le type de représentation obtenue par la méthode d'analyse choisie est bien en adéquation avec les données de départ. Pour ce faire, il est de coutume d'évaluer différents critères :

- *Coefficient de corrélation cophénétique :*
  La première façon, assez naturelle, d'estimer à quel point la structure hiérarchique est en adé-

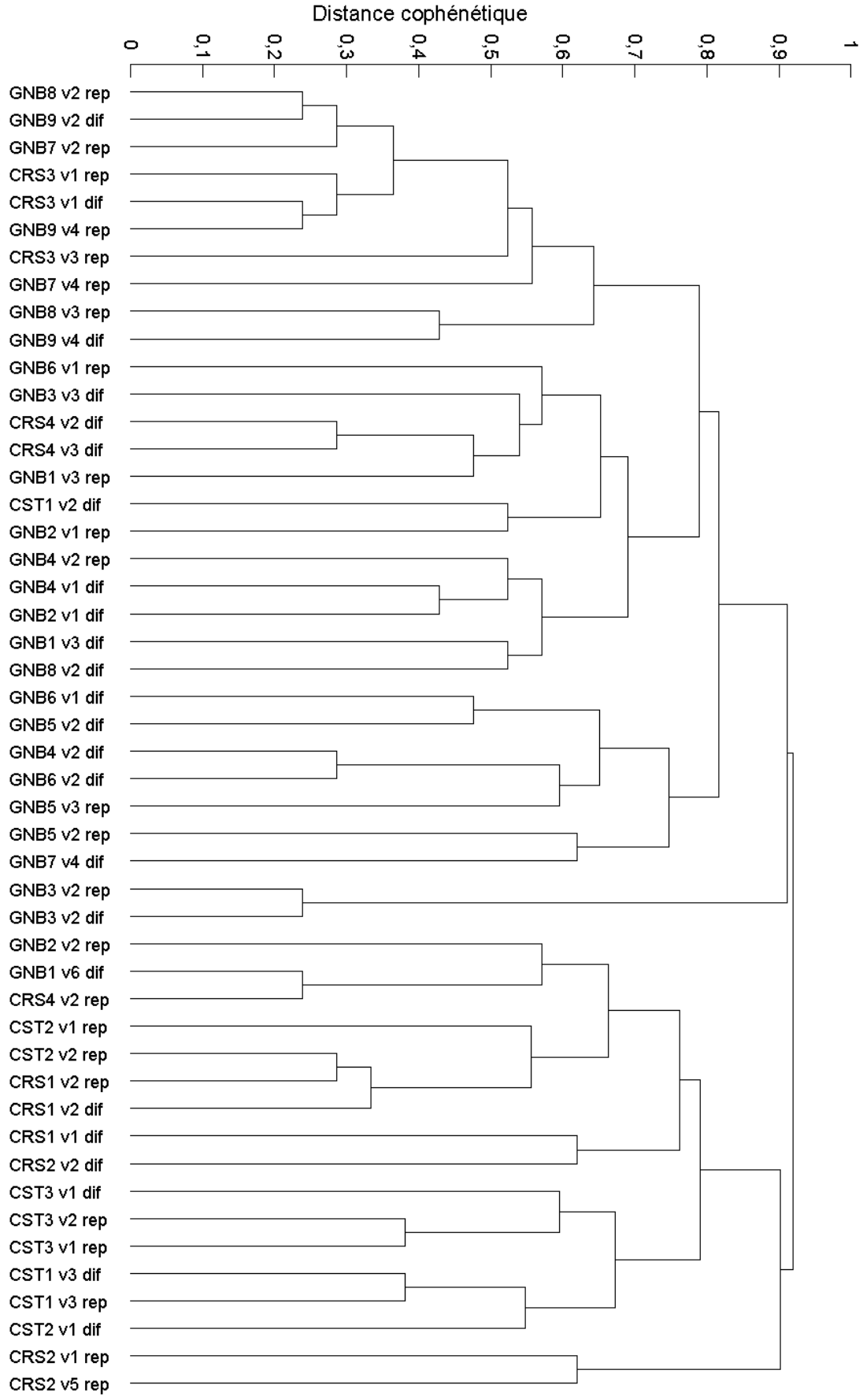


FIGURE 5.5 – Dendrogramme issu de l'expérience de catégorisation libre.

quation avec les données de départ de l'analyse de cluster est d'évaluer le coefficient de corrélation de Bravais-Pearson (annexe B.1) entre les distances originales (celle de la matrice M_{dis}) et les distances cophénétiques. Cette corrélation est alors appelée *corrélation cophénétique*. Pour la présente expérience, ce coefficient est de 0,82. La figure 5.6 présente le *diagramme de type Shepard* (nom donné au graphe affichant les valeurs d'une distance modélisée en fonction d'une distance originale). On remarque sur cette figure que la répartition des points est partitionnée sur les deux axes, ce qui était à attendre, puisque les distances cophénétiques présentent 47 valeurs différentes possibles (soit le nombre d'éléments moins un), et les distance originales présentent 22 valeurs différentes possibles (soit le nombre de participants plus un). D'un point de vue statistique rigoureux, on ne peut pas utiliser la table B.1 en annexe B.1 afin d'établir la significativité statistique de la corrélation, étant donné que les deux variables (distance originale et distance cophénétique) ne peuvent pas être supposées à priori indépendantes, puisque l'une est utilisée pour calculer l'autre. Toutefois, elle permet d'évaluer, au moins qualitativement, la modélisation des distances originales par la distance cophénétique.

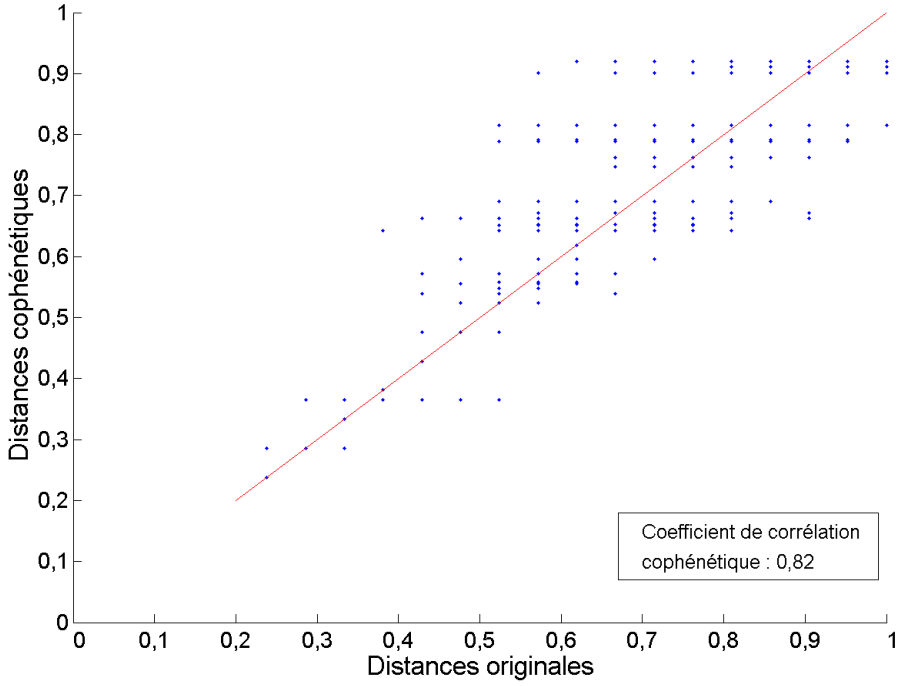


FIGURE 5.6 – Diagramme de type Shepard – Représentation des distances cophénétiques en fonction des distances originales. La droite rouge représente la modélisation idéale (distances cophénétiques identiques aux distances originales.

– *Inégalité triangulaire :*
Étant donné que la matrice M_{dis} est considérée comme une matrice de distances, il convient de vérifier que son jeu de valeurs vérifie bien les propriétés élémentaires d'une distance. Les deux premières propriétés (*symétrie* : $M_{dis}(i,j) = M_{dis}(j,i)$ et *séparation* $M_{dis}(i,j) = 0 \Leftrightarrow i = j$) semblent triviales compte tenu de la façon dont la matrice M_{dis} est construite. La troisième propriété d'inégalité triangulaire est moins évidente à priori. Voici comment elle se traduit dans notre cas :

$$\forall i,j,k \qquad M_{dis}(i,k) \leq M_{dis}(i,j) + M_{dis}(j,k) \tag{5.3}$$

Si l'on tente d'interpréter géométriquement cette inégalité, on comprend qu'il serait impossible de représenter les éléments i, j et k comme des points dans un espace géométrique si elle n'est pas vérifiée. En effet, elle se traduit simplement par le fait que ces trois points forment un triangle, ou dans le cas critique (égalité) une droite. On comprend ainsi son appellation et on peut difficilement imaginer de parler de distances si elle n'est pas vérifiée.
Sur l'ensemble des triplets possibles parmi les 48 éléments (103776 triplets en tenant compte de l'ordre des éléments), 100 % d'entre eux vérifient cette inégalité. On en conclut que la matrice M_{dis} est bien une matrice de distances et que la validité de cette inégalité découle sans aucun doute de la façon dont cette matrice est construite.

– *Inégalité ultramétrique :*
Le type particulier de représentation que l'on tente d'appliquer au jeu des distances originales de la matrice M_{dis} implique également de s'intéresser à une autre propriété qui justifie l'emploi d'une telle méthode. L'inégalité ultramétrique, lorsqu'elle est respectée par une distance, en fait ce que l'on nomme une *distance ultramétrique*. Elle s'énonce de la manière suivante, toujours avec la notation propre à notre analyse :

$$\forall i, j, k \qquad M_{dis}(i,k) \leq max(M_{dis}(i,j), M_{dis}(j,k)) \tag{5.4}$$

Une rapide inspection permet, tout d'abord de constater que l'inégalité ultramétrique implique l'inégalité triangulaire, mais la réciproque n'est pas vraie [5]. L'inégalité ultramétrique est donc bien une propriété supplémentaire aux propriétés élémentaires d'une distance. L'intérêt particulier de l'inégalité réside dans l'interprétation de la propriété correspondante.
Si l'on choisit trois éléments i, j et k au hasard et que l'on teste cette inégalité sur les différentes permutations des 3 seules distances mises en jeu (compte tenu de la propriété de symétrie), on observe que, sauf cas particulier où certaines sont égales, l'inégalité sera vérifiée 2 fois sur 3, étant donné qu'une des trois distances est forcément plus grande que les deux autres.
À partir de cette observation, on peut se poser la question de savoir dans quel(s) cas particulier(s) cette inégalité sera respectée pour toutes les permutations possibles des éléments i, j et k, et donc des distances mises en jeu. On s'aperçoit alors que la seule possibilité est que deux des trois distances soient égales et qu'elles soient supérieures à la troisième. Ceci peut se traduire par le fait que, dans un espace ultramétrique – où toutes les distances respectent l'inégalité ultramétrique – tous les triangles formés par ces distances sont isocèles et voient leur base être plus petite que les côtés adjacents. Ce constat permet de comprendre l'importance de la propriété lorsque l'on considère une représentation hiérarchique, comme les dendrogrammes, d'un jeu de distances. En guise d'illustration de ce propos, la figure 5.7 présente à gauche un triangle vérifiant l'inégalité ultramétrique, et à droite la représentation que l'on tente d'appliquer pour ce triplet d'éléments. Si cette représentation s'adapte parfaitement aux données, il est normal que les deux éléments en bas soient séparés d'une distance moindre que celles les séparant du troisième élément, et soient à égale distance de ce dernier.
Au final, si l'on teste la propriété de l'inégalité ultramétrique sur la totalité des triplets que l'on peut former à partir des 48 éléments (103776 triplets en tenant compte de l'ordre des éléments), celle-ci est vérifiée dans 76,6 % des cas. Ce score, bien que significativement supérieur au score de base de 66,7 % (2 sur 3), n'est pas aussi élevé que l'on pourrait souhaiter pour ce type d'analyse. Il faut toutefois garder à l'esprit l'interprétation géométrique de cette inégalité

5. En effet, prenons en guise de contre-exemple les valeurs fictives 5, 4 et 3 pour respectivement $M_{dis}(i,k)$, $M_{dis}(i,j)$ et $M_{dis}(j,k)$. L'inégalité triangulaire est bien vérifiée : $5 \leq 4+3$, mais l'inégalité ultramétrique ne l'est pas : $5 > \max(4,3)$.

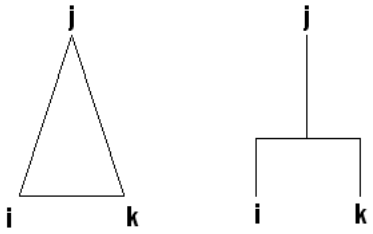


FIGURE 5.7 – Illustration de l'inégalité ultramétrique entre des points fictifs i, j et k.

telle qu'illustrée en figure 5.7, et il semble notamment difficile dans le cas de données d'expérience que, systématiquement pour chaque triplet, deux des trois distances soient égales. De plus, les sons étudiés ici sont d'une nature très proche, ce qui justifierait en principe l'emploi d'un autre type de représentation, notamment continu (espace de timbre). La principale raison d'être de cette expérience et de l'analyse qui s'en est suivie est plutôt d'ordre pratique, puisque l'on souhaite réduire le corpus sonore à une taille raisonnable permettant de conduire d'autres expériences, (notamment celle permettant d'obtenir une représentation continue comme l'espace de timbre), et parce que ce type de méthodologie est le seul permettant de réaliser cette opération en se fondant sur la perception.

5.4 Discussion – Sélection du corpus réduit

Maintenant que nous sommes parvenus à identifier la hiérarchie des catégories de sons qui composent le corpus de 48 sons, il importe d'identifier les catégories principales. Cela signifie en fait qu'il faut fixer un seuil de distance cophénétique, en dessous duquel on considèrera que les sons sont dans la même catégorie. Cela revient à fixer un niveau critique de précision dans la hiérarchie. Une valeur usuelle de seuil de distance cophénétique est 0,7. Sur la figure 5.8, la ligne verticale rouge en pointillés identifie le seuil de distance cophénétique. Les différentes catégories de sons correspondent donc aux hiérarchies à la gauche de cette ligne en pointillés. Sur la figure, ces « sous-hiérarchies » sont identifiées par des couleurs différentes. Nous obtenons donc 9 classes de sons dont les éléments sont séparés sur la figure par les lignes bleues en pointillés.

L'écoute des sons au sein de ces différentes catégories permet d'en identifier certaines spécificités susceptibles d'expliquer cette décomposition du corpus. Ces spécificités sont indiquées dans le tableau 5.4. Il importe de conserver à l'esprit le fait qu'elles revêtissent un certain degré de subjectivité car elles ont été identifiées par l'expérimentateur et non par les auditeurs ayant participé à l'expérience de catégorisation libre. Elles permettent toutefois de donner une idée du contenu de chaque catégorie. Le tableau 5.4 liste également, pour chaque catégorie, les modèles de STA les plus représentés (au moins deux des trois représentants initiaux de chacun des 16 modèles – voir tableau 5.1). Nous pouvons ainsi constater qu'aucun des 16 modèles ne présente trois représentants placés dans trois catégories différentes. En revanche, seuls les modèles **CRS3** et **CST3** présentent leurs trois représentants dans une même catégorie. Cela signifie, dans une certaine mesure, que les sons provenant d'un même modèle de STA ont été jugés comme relativement proches par les auditeurs.

La question qui se pose à présent est de savoir quels sons choisir parmi les différentes catégories identifiées, afin de réduire et d'échantillonner le corpus de 48 sons d'une manière qui soit représentative de la variété des sons, c'est-à-dire des catégories. Il est également important de conserver la

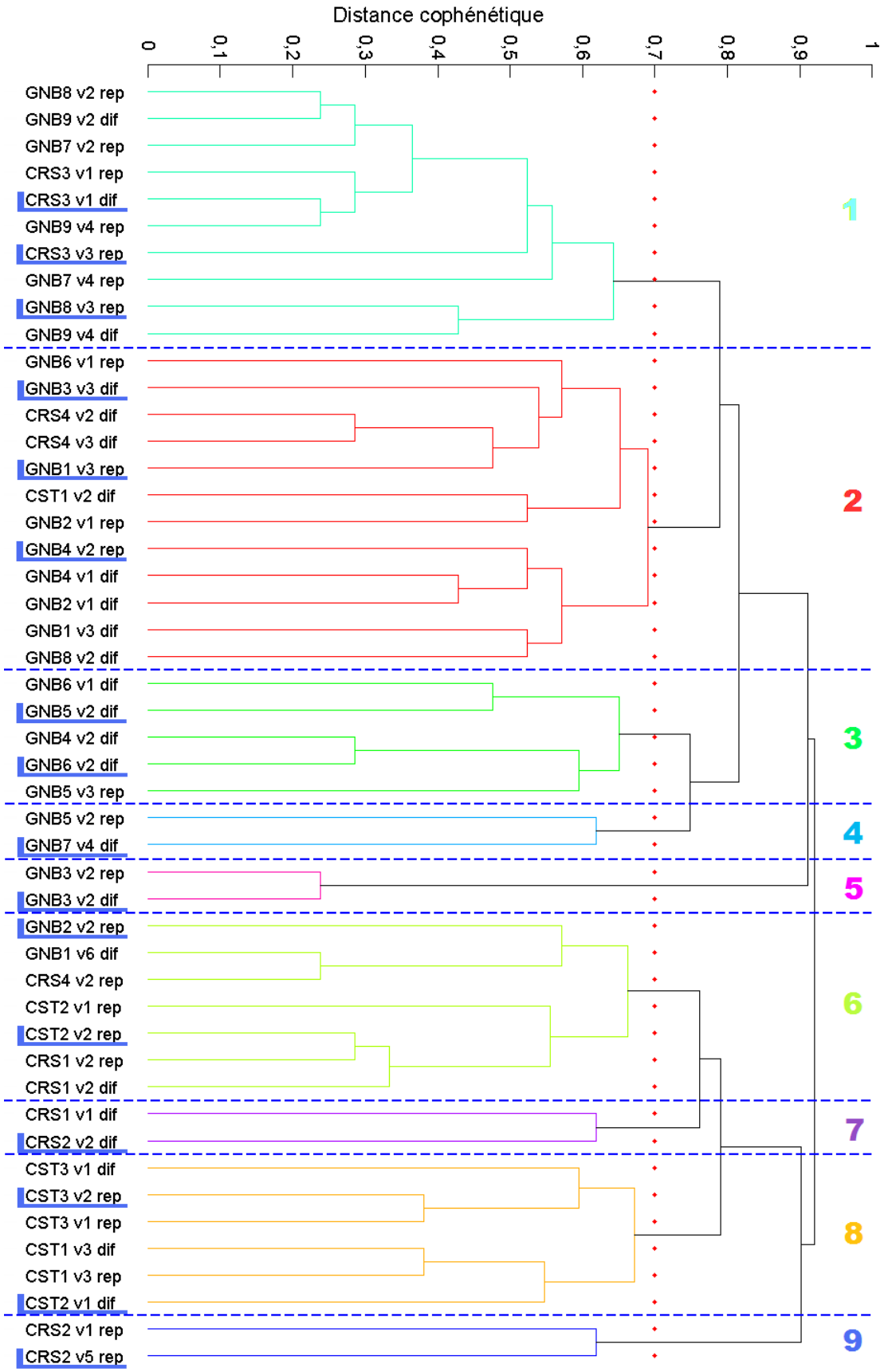


FIGURE 5.8 – Dendrogramme issu de l'expérience de catégorisation libre, avec les catégories identifiées (séparées par les lignes en pointillés bleus) et les *exemplaires représentatifs* initiaux (soulignés en bleu).

Numéro (couleur) sur la figure 5.8	Spécificités évoquées	STA représentés
1 (cyan)	large-bande, évoquant un important flux d'air	**CRS3** **GNB7** **GNB8** **GNB9**
2 (rouge)	bande plus centrée vers les moyennes/hautes fréquences	**CRS4** **GNB1** **GNB2** **GNB4**
3 (vert)	bande plus centrée vers les basses/moyennes fréquences	**GNB5** **GNB6**
4 (bleu clair)	proches de la catégorie 3, peu de sons, caractérisation difficile	–
5 (rose)	sifflement en hautes fréquences	**GNB3**
6 (vert clair)	basses fréquences, présence de fluctuations	**CRS1** **CST2**
7 (violet)	basses fréquences, sans fluctuation	–
8 (orange)	proches de la catégorie 6, fluctuations moins prononcées	**CST1** **CST3**
9 (bleu foncé)	forte émergence harmonique dans les basses fréquences	**CRS2**

TABLE 5.4 – Description des 9 catégories identifiées en termes de spécificités (telles qu'identifiées par l'expérimentateur) caractérisant les sons de chacune d'elles, et de modèles de STA majoritairement représentés (au moins 2 représentants sur 3).

distribution des sons, en termes d'effectif, dans les différentes classes. Par ailleurs, comme expliqué en section 5.1, nous avons besoin d'un corpus constitué de 10 à 20 sons. Or nous avons enregistré 16 STA différents, censés représenter la variété de STA considérés, et on observe globalement que les représentants des différents modèles apparaissent regroupés dans des même clusters ou dans des clusters proches. Il semble par conséquent naturel de partir simplement sur une sélection de 16 sons, et de tenter d'obtenir un représentant de chaque STA enregistré.

En terme d'effectif, quatre des neuf catégories sont constituées de 2 sons, les autres sont constituées respectivement de 5, 6, 7, 10 et 12 sons. Afin de conserver la représentativité de cette répartition de l'effectif, et en visant un total de 16 sons, nous avons fixé les nombres de sons à sélectionner à 1 pour les quatre catégories avec 2 éléments, à 2 pour les trois catégories avec 5, 6 et 7 éléments, et à 3 pour les deux catégories avec 10 et 12 éléments.

Pour la sélection dans chaque catégorie, la stratégie qui a été adoptée est fondée sur la définition du *prototype* de Rosch [131]. Pour elle, le prototype d'un groupe est l'élément central, c'est-à-dire l'élément qui est à la fois le plus proche en moyenne des autres éléments du groupe, et le plus éloigné en moyenne des éléments des autres groupes. Sur la base de cette définition, un algorithme a été mis au point afin de calculer pour chaque son de chaque catégorie un critère de sélection c_{sel}. Pour un son i appartenant la catégorie G, ce score correspond à la différence entre sa distance moyenne (au sens de la matrice M_{dis}) avec les sons j des autres catégories ($j \notin G$) et sa distance moyenne avec les sons du groupe auquel il appartient ($j \in G$) :

$$c_{sel}(i) = moyenne_{j \notin G}\,(d_{i,j}) - moyenne_{j \in G}\,(d_{i,j}) \tag{5.5}$$

La sélection se fait simplement en identifiant l'élément de chaque groupe maximisant ce coefficient. Toutefois, un problème se pose lorsque l'on souhaite sélectionner plusieurs sons dans les classes présentant un effectif plus important. Dans un tel cas, les distances entre sons sélectionnés d'une même catégorie doivent être également prises en compte, l'idée majeure étant de trouver une combinaison de sons de la catégorie qui soit représentative de la variété de sons à l'intérieur de la catégorie. Par conséquent, on cherche également, au sein d'une même catégorie, à maximiser les distances des sons sélectionnés entre eux. Ainsi, pour chaque catégorie G, un nouveau critère de sélection est évalué pour chaque combinaison P ($P \subset G$) possible de p (2 ou 3) sons à sélectionner parmi l'effectif de la catégorie.

La formule du critère de sélection pour la combinaison P est alors ainsi enrichie :

$$c_{sel}(P) = moyenne_{i\in P,\ j\notin G}\ (d_{i,j}) - moyenne_{i\in P,\ j\in G\setminus P}\ (d_{i,j}) + moyenne_{i\in P,\ j\in P,\ j\neq i}\ (d_{i,j}) \qquad (5.6)$$

où $G \setminus P$ signifie la catégorie G privée des éléments de la combinaison P. On cherche toujours à maximiser le critère c_{sel}. On cherche donc à la fois à maximiser les distances des exemplaires sélectionnés avec les éléments des autres catégories, à minimiser leurs distances avec les éléments non-sélectionnés de la catégorie considérée, et à maximiser leurs distances entre eux.

Il est à noter que ce raffinement du critère est quelque peu contradictoire avec la définition de *prototype* de Rosch. En effet, celle-ci conçoit les prototypes comme les exemplaires *centraux* d'un groupe, c'est-à-dire qu'il doivent regrouper toutes les caractéristiques propres aux éléments du groupe et s'opposer à celles des autres groupes. Selon cette définition, ces exemplaires devraient, au contraire, être proches les uns des autres, ce qui n'est ici pas souhaitable, les sons sélectionnés étant censés représenter la gamme de variation des sons de la catégorie. Par conséquent, plutôt que de parler de *prototypes*, nous préférerons parler d'*exemplaires représentatifs*.

Les exemplaires représentatifs ainsi sélectionnés sont indiqués sur la figure 5.8 (soulignés en bleu). En voici un inventaire :

1. **CRS2 v2 dif**
2. **CRS2 v5 rep**
3. **CRS3 v1 dif**
4. **CRS3 v3 rep**

5. **GNB1 v3 rep**
6. **GNB2 v2 rep**
7. **GNB3 v2 dif**
8. **GNB3 v3 dif**
9. **GNB4 v2 rep**
10. **GNB5 v2 dif**
11. **GNB6 v2 dif**
12. **GNB7 v4 dif**
13. **GNB8 v3 rep**

14. **CST2 v1 dif**
15. **CST2 v2 rep**
16. **CST3 v2 rep**

Plusieurs observations peuvent être faites sur cette sélection. Tout d'abord, lorsque l'on regarde la figure 5.8, il semble que la sélection des exemplaires représentatifs échantillonne chaque catégorie à grand effectif d'une manière satisfaisante, puisqu'un élément est généralement sélectionné dans chaque partie de la hiérarchie interne à chaque catégorie. On remarque également que les prises de sons de diffusion (**dif**) et de reprise (**rep**) sont représentées de manière équivalente. Il en est globalement de même lorsque l'on compare les enregistrements à faible vitesse (**v3**, **v4** et **v5**) et à haute vitesse (**v1** et **v2**). En revanche, on observe que seuls 12 des 16 différents STA enregistrés sont finalement représentés, 4 d'entre eux ayant deux exemplaires sélectionnés. Nous avons préféré ajuster la sélection afin de corriger ce dernier point, et avoir un représentant de chacun des 16 STA enregistrés. Cet ajustement a été fait en s'attachant dans la mesure du possible à respecter la structure hiérarchique de l'expérience et à conserver la bonne représentativité évoquée ci-dessus des différentes conditions d'enregistrement. Il a également été vérifié que les éléments ajoutés à la sélection (ou les combinaisons d'exemplaires résultant de ces ajouts) conservent un score au critère de sélection c_{sel} parmi les plus élevés de chaque catégorie concernée. La nouvelle sélection des exemplaires représentatifs, qui correspond donc au corpus réduit qui sera utilisé pour les expériences subséquentes, est indiquée en figure 5.9, et est inventoriée dans le tableau 5.5.

1.	**CRS1 v1 dif**
2.	**CRS2 v1 rep**
3.	**CRS3 v3 rep**
4.	**CRS4 v2 dif**
5.	**GNB1 v3 dif**
6.	**GNB2 v2 rep**
7.	**GNB3 v2 dif**
8.	**GNB4 v2 rep**
9.	**GNB5 v2 dif**
10.	**GNB6 v2 dif**
11.	**GNB7 v4 dif**
12.	**GNB8 v3 rep**
13.	**GNB9 v4 rep**
14.	**CST1 v3 rep**
15.	**CST2 v2 rep**
16.	**CST3 v2 rep**

TABLE 5.5 – Corpus final sélectionné à l'issue de l'expérience de catégorisation libre (voir également le tableau 5.1 pour plus de détails).

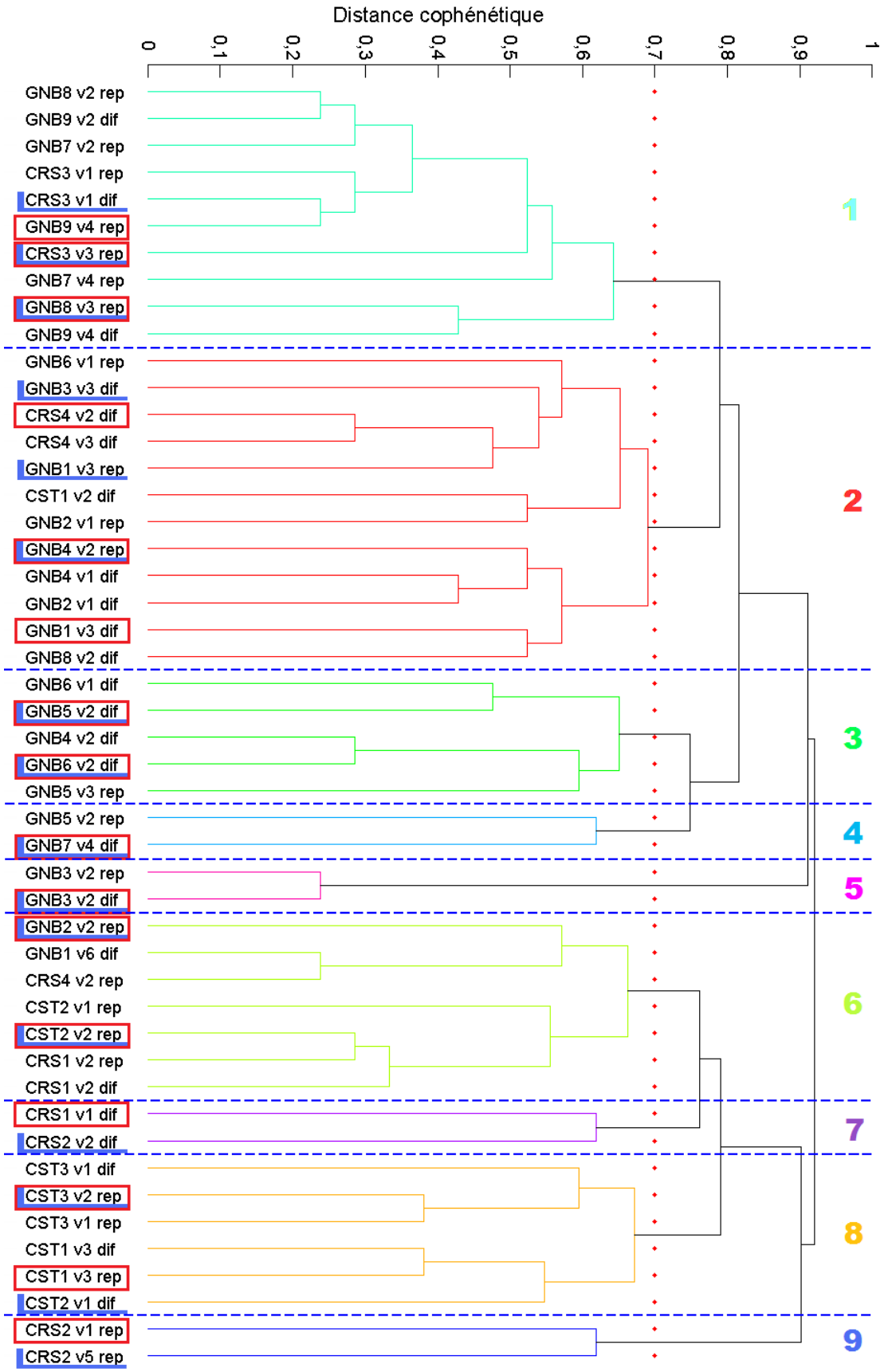


FIGURE 5.9 – Dendrogramme issu de l'expérience de catégorisation libre, avec les catégories identifiées (séparées par les lignes en pointillés bleus) et les *exemplaires représentatifs* finaux (dans les cadres rouges).

Chapitre 6

Identification des attributs auditifs des STA : étude du timbre

Ce chapitre présente l'étude du timbre des sons de STA, réalisée dans le but d'identifier les attributs auditifs pertinents pour la description des sons de STA. Dans cette optique, la section 6.1 présente tout d'abord la problématique générale de l'étude du timbre. La section 6.2 présente le protocole expérimental utilisé afin de réaliser une expérience de mesure de similarités. La section 6.3 présente l'analyse des résultats de cette expérience en termes d'espaces perceptifs et de descripteurs explicatifs. Enfin, la section 6.4 expose la discussion que soulèvent les résultats de cette étude.

6.1 Problématique

Le chapitre 5 nous a permis d'extraire de l'ensemble d'enregistrements effectués un corpus sonore réduit. Dans le cadre de l'évaluation de la qualité sonore des STA, il importe d'identifier les paramètres acoustiques des éléments de ce corpus qui permettent d'expliquer les préférences des auditeurs. Or, la recherche de ces paramètres nous amène à nous interroger sur les caractéristiques des sons qui permettent en premier lieu de les distinguer entre eux. En effet, il est raisonnable de supposer que, si un paramètre acoustique explique la distinction perceptive entre les différentes paires de sons, ce même paramètre devrait également expliquer les préférences des auditeurs. À l'inverse, si une caractéristique n'induit aucune différence perceptive entre les sons, il semblerait peu cohérent qu'elle puisse avoir une quelconque influence sur la qualité sonore perçue. L'enjeu de cette étude est donc d'identifier les attributs auditifs pertinents pour la perception des sons de STA.

Les sons qui constituent le corpus peuvent théoriquement varier selon une multitude de paramètres, y compris ceux déjà identifiés en section 2.3.3. Toutefois, compte tenu de la nature similaire des sons considérés, qui sont tous des enregistrements de STA, il est fort probable que beaucoup de ces paramètres ne varient que peu sur l'échelle absolue des valeurs qui peuvent théoriquement leur être associées. De plus, un point plus important encore est que tous les paramètres n'ont très certainement pas la même importance dans la description perceptive des éléments du corpus considéré. Il est au contraire logique de supposer que seulement un nombre réduit d'entre eux permet de distinguer les sons du corpus de manière relativement exhaustive. Cela ne signifie pas que les variations d'autres facteurs ne sont pas perçues, mais que dans le contexte de recherche de la structure générale de description sous-tendant la perception du corpus considéré, ces facteurs apportent peu et influent de manière négligeable sur la ressemblance acoustique qu'un auditeur peut formuler entre deux éléments.

Il convient alors d'identifier les paramètres sonores qui permettent d'expliquer les ressemblances acoustiques perçues entre chaque paire de sons qu'il est possible de former à partir du corpus de travail. La littérature nous offre différents exemples d'étude (Susini et al. [145, 146], McAdams et al. [107], Susini et al. [147], Parizet et al. [118], Lemaitre et al. [100] et Misdariis et al. [111], entre autres) répondant à cette problématique pour différents types de sons. Ces études détaillées en section 2.1.2 emploient toutes une méthodologie commune dite d'*étude du timbre*. Cette méthodologie consiste principalement en l'exploitation des résultats d'une expérience de *mesure de similarités*, dans laquelle on demande aux participants d'évaluer pour chaque paire de sons à quel point ils se ressemblent. L'analyse des résultats permet d'extraire un nombre réduit de dimensions perceptives permettant d'expliquer les jugements de similarité.

La dernière étape de cette méthodologie consiste à donner une interprétation à ces dimensions perceptives, notamment en termes de descripteurs explicatifs. Il importe de préciser la démarche adoptée dans le cadre de la recherche des descripteurs explicatifs. Bien entendu, il serait aisé de tester leur capacité à expliquer les dimensions de l'espace de timbre un à un, et de sélectionner ceux qui offre la plus forte corrélation (au sens du coefficient de corrélation de Bravais-Pearson, voir section B.1 en annexe). Toutefois, cette stratégie est contestable, car, compte tenu de la grande variété de descripteurs acoustiques et psychoacoustiques existants, il est fort possible de trouver des descripteurs fortement corrélés avec les dimensions perceptives identifiés de manière tout-à-fait fortuite. Or une forte corrélation n'implique pas la causalité, et ne garantit pas que les descripteurs identifiés aient un réel lien avec les dimensions. Il est indispensable de tenter auparavant d'identifier par « introspection » les attributs auditifs susceptibles d'expliquer les dimensions perceptifs, et de cibler en conséquence les descripteurs possibles à associer à ces attributs auditifs. Pour cette raison, la démarche suivie ici – mais également dans le cadre des études mentionnées ci-dessus – pour la recherche des descripteurs explicatifs est la suivante :

- Identification, à l'écoute des sons selon les valeurs de chaque dimension perceptive mise à jour, du percept qui pourraient, intuitivement, expliquer ces valeurs ;
- Recherche de descripteurs acoustiques ou psychoacoustiques usuellement associé à ce percept ;
- Évaluation des coefficients de corrélation de Bravais-Pearson et des régressions linéaires (voir annexe B) entre les valeurs de la dimension perceptive et celles des descripteurs ainsi sélectionnés, afin de confirmer l'intuition initiale.

Étant donné que notre écoute des sons peut être « biaisée », compte tenu de notre connaissance de ceux-ci, mais aussi de nos attentes, il est possible que pour une même dimension, il soit nécessaire de répéter cette démarche plusieurs fois en cas de corrélations non-significatives. Il nous semble, malgré tout, que le fait de fixer cette stratégie en amont permet de nous prémunir de conclusions erronées.

6.2 Protocole expérimental de mesure de similarités

6.2.1 Stimuli

Le corpus sonore utilisé lors de cette expérience est celui identifié à l'issue du chapitre 5 et énuméré dans le tableau 5.5. Il s'agit donc de 16 sons monophoniques égalisés en sonie [1] d'une durée de

1. L'égalisation en sonie automatique fondée sur le modèle de Zwicker explicitée en section 5.2.1 s'est avérée satisfaisante à l'oreille, et il n'a pas été jugé nécessaire de la réaliser expérimentalement, comme c'est le cas en section 8.3.3.

4 secondes [2], et dont le niveau acoustique varie de 38,1 à 40,1 dBA.

6.2.2 Participants

24 auditeurs volontaires (18 hommes, 6 femmes, entre 22 et 25 ans), n'ayant pas fait mention d'un problème majeur d'audition, et n'ayant pas participé à une autre expérience réalisée dans le cadre de cette thèse, ont pris part à cette expérience.

6.2.3 Matériel

Une interface graphique spécifique, assurant la lecture des sons sur l'interface audio et l'enregistrement des réponses des participants, a été programmée en LabVIEW 2010 (figure 6.1) pour cette expérience. Les sons ont été diffusés par une interface RME Fireface 800 dans un casque ouvert Sennheiser HD650. Le mode de reproduction sonore sur casque a été employé car c'est le seul qui garantit d'obtenir une écoute diotique (signal identique dans les deux oreilles). L'expérience a eu lieu dans une cabine audiométrique IAC à double paroi.

FIGURE 6.1 – Interface graphique de l'expérience de mesure de similarités.

6.2.4 Procédure

Au début de l'expérience, les participants ont reçu une consigne écrite, retranscrite en section E.2 en annexe, leur expliquant le contexte de l'étude, la nature des sons étudiés et la tâche à accomplir.

2. Bien que le corpus de référence ait été établi à l'aide de sons d'une durée de 5 secondes, cette dernière à été réduite ici à 4 secondes par souci de cohérence avec le corpus utilisé dans le cadre de l'étude exposée au chapitre 8. En effet, la première seconde des sons auralisés utilisés dans le cadre de cette étude a été supprimée car le champ réverbéré issu de l'auralisation n'y est encore que peu présent.

Cette dernière consistait à évaluer pour chaque paire présentée la similarité des deux sons à l'aide d'un curseur se déplaçant sur une glissière allant de « très semblables » à « très dissemblables » (voir figure 6.1). Les participants avaient, pour chaque paire, la possibilité d'écouter chaque son autant de fois qu'ils le souhaitaient. Une fois satisfaits de leur réponse, ils validaient à l'aide du bouton correspondant afin de passer à la paire suivante.

Toutes les paires (i, j) possibles parmi les 16 sons du corpus ont été présentées à chaque participant, en sachant que :

- les paires (i, i) n'étaient pas présentées ;
- chaque paire était présentée dans un ordre unique ((i, j) ou (j, i) exclusivement).

Ainsi pour ce corpus de $N = 16$ sons, $N(N-1)/2 = 120$ paires ont été présentées. L'ordre de présentation des paires était rendu aléatoire pour chaque participant, avec toutefois quelques contraintes :

- chaque présentation successive d'un même son dans une paire devait être séparée par au moins deux autres paires où ce son n'intervenait pas, afin d'éviter tout « effet d'ancrage ». En effet, lorsqu'un son est confronté successivement aux autres, les jugements peuvent être trop influencés par des caractéristiques propres à celui-ci [49] ;
- chaque son devait être présenté autant de fois en tant que premier que deuxième son, afin de limiter l'influence de cette position du son sur les jugements.

En outre, dix paires aléatoires étaient également présentées en début d'expérience en guise de phase d'entraînement. Les réponses pour cette phase n'ont pas été prises en compte. Les participants pouvaient ainsi se familiariser avec l'interface et avoir une idée de la variété de sons qu'ils allaient rencontrer au cours de l'expérience. Enfin, une pause était proposée aux participants après 70 paires (entraînement compris).

La durée moyenne de la procédure dans le cas de cette expérience a été de 31 minutes.

6.3 Résultats et analyse

Les données issues d'une telle expérience prennent la forme de matrices symétriques NxN – N est le nombre de stimuli étudiés, 16 dans le cas présent – de distances ou similarités perceptives [3] mesurées pour chaque participant. Ces matrices de données ont été analysées par une technique de *Multi-Dimensional Scaling* (*MDS*, voir annexe A) dont le but est d'établir un espace géométrique constitué de points représentant les stimuli, et où les distances perceptives mesurées sont modélisées par les distances géométriques entre les points. Le modèle utilisé ici est le modèle INDSCAL (voir section A.2 en annexe) dont l'intérêt majeur est de permettre une interprétation psychologique pertinente des dimensions obtenues. En effet, ce modèle prend en compte l'importance relative que chaque participant attribue à chacune des caractéristiques sonores qu'il a utilisées pour effectuer ses jugements de similarité. Cette importance relative prend la forme dans le modèle de poids propres à chaque participant appliqués à chacune des dimensions de l'espace géométrique. Ces poids permettent de « fixer « les axes de l'espace obtenu, c'est-à-dire d'en identifier le seul jeu d'orientations orthogonales qui en autorise l'interprétation perceptive. On dit alors que les poids empêchent l'*invariance rotationnelle* de l'espace. Le logiciel IBM SPSS Statistics 20 [84] permet d'utiliser ce modèle (toutefois désigné par « Positionnement multidimensionnel – PROXSCAL/modèle Euclidien pondéré »).

3. Si les valeurs de distance ou de similarité prennent des valeurs entre 0 et 1, on a la relation : $distance = 1 - similarité$.

6.3.1 Dimensionalité de l'espace perceptif recherché

Il est important de préciser que le nombre de dimensions de l'espace recherché est une inconnue du problème, et est considéré dans l'implémentation du modèle dans le logiciel SPSS comme un paramètre d'entrée. Un espace à $N-1$ dimensions permet de représenter géométriquement un jeu de distances entre N éléments sans la moindre erreur. Bien entendu, dans le cas d'un corpus de 16 sons, un espace à 15 dimensions est tout-à-fait inutilisable et peu pertinent. Il est en revanche évident que lorsque l'on diminue la dimensionalité de l'espace, on augmente l'erreur entre les distances prédites par le modèle (c'est-à-dire les distances géométriques dans l'espace) et les distances perceptives mesurées. Il s'agit donc de trouver un compromis entre faible erreur de représentation et pertinence perceptive des dimensions obtenues.

Par conséquent, l'analyse a été réalisée pour différentes dimensionalités. L'adéquation aux données de ce nombre de dimensions peut être évaluée à l'aide de différents coefficients. Parmi ceux-ci, le *coefficient de congruence de Tucker* [154] r_c peut, dans une certaine mesure, se comparer à un coefficient de corrélation de Bravais-Pearson (voir section B.1 en annexe), notamment au travers de la gamme de valeurs qu'il peut théoriquement prendre, de '-1' à '1'. La différence majeure réside dans le fait que les variables testées ne sont pas centrées lors du calcul de ce coefficient. Il s'exprime pour deux variables $X = \{x_i\}$ et $Y = \{y_i\}$ par :

$$r_c = \frac{\sum_i x_i y_i}{\left[\sum_i x_i^2 \sum_i y_i^2\right]^{1/2}} \tag{6.1}$$

Ici, ce coefficient est estimé entre les distances perceptives mesurées et les distances modélisées dans l'espace obtenu. Ces distances étant, par définition, positives, il s'avère, contrairement au coefficient de corrélation de Bravais-Pearson, que ce coefficient, d'une part, ne peut pas être nul ou négatif. En conséquence, sa valeur approche '0' lorsque le modèle ne parvient absolument pas à représenter le jeu de distances originales, et, à l'inverse, est d'autant plus proche de '1' que l'erreur entre les distances modélisées et les distances originales est faible.

À partir du coefficient de congruence de Tucker, on peut également évaluer facilement la *dispersion expliquée – Dispersion Accounted For, DAF* – par le modèle, c'est-à-dire la quantité en pourcentage d'information des données originales que le modèle est capable de représenter :

$$DAF(\%) = {r_c}^2 \times 100 \tag{6.2}$$

La figure 6.2 affiche les valeurs du coefficient de congruence de Tucker r_c et de la dispersion expliquée DAF par le modèle. Selon la forme des deux courbes, il semble qu'une dimensionalité de 2 soit le compromis optimal. Afin d'améliorer l'efficacité du modèle de prédiction utilisé en section 7.6.2, le modèle à 3 dimensions sera également considéré par la suite.

6.3.2 Espaces perceptifs

La figure 6.3 permet d'observer l'espace perceptif à 2 dimensions (« espace 2D ») qui a été obtenu. Les figures 6.4 et 6.5 permettent, quant à elle, d'observer l'espace perceptif à 3 dimensions (« espace 3D »), et plus précisément la projection de cet espace, respectivement, sur les dimensions 1 et 2, et sur les dimensions 1 et 3. Enfin, les tableaux 6.1 et 6.2 montrent les pondérations des dimensions pour chacun des participants, respectivement pour l'espace 2D et pour l'espace 3D.

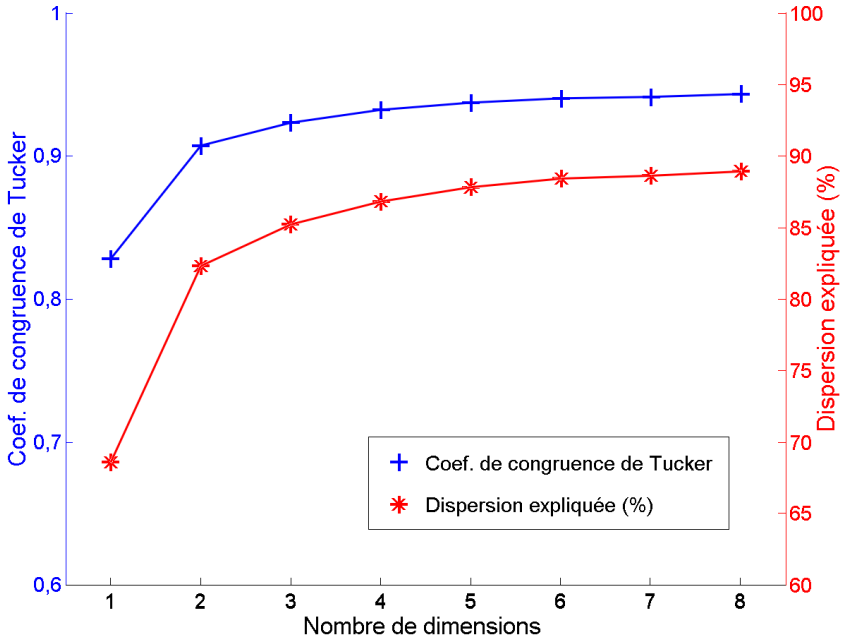


FIGURE 6.2 – Coefficient de congruence de Tucker et dispersion expliquée par le modèle en fonction du nombre de dimensions choisi.

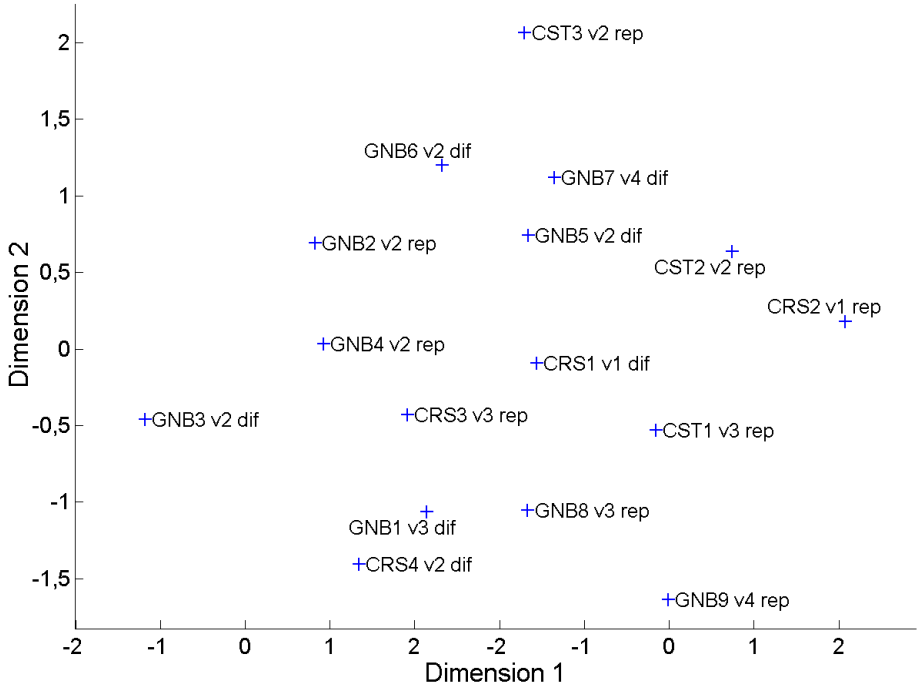


FIGURE 6.3 – Espace perceptif à 2 dimensions.

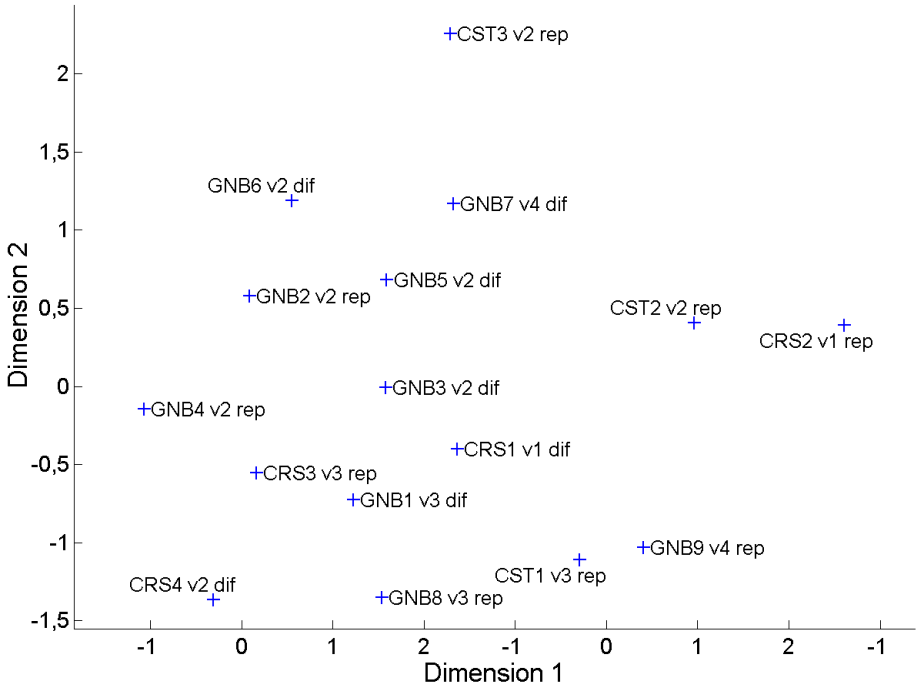


FIGURE 6.4 – Espace perceptif à 3 dimensions – 1ère et 2ème dimensions.

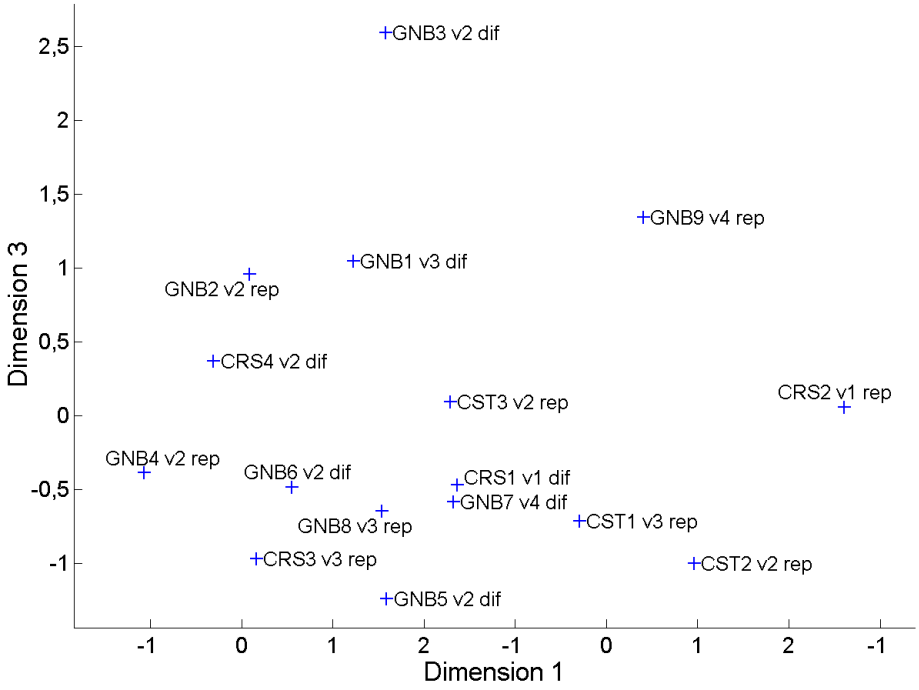


FIGURE 6.5 – Espace perceptif à 3 dimensions – 1ère et 3ème dimensions.

Index du participant	Dimension 1	2
1	0,472	0,393
2	0,282	0,263
3	0,435	0,445
4	0,371	0,299
5	0,433	0,454
6	0,344	0,298
7	0,459	0,440
8	0,310	0,335
9	0,504	0,413
10	0,338	0,238
11	0,470	0,409
12	0,339	0,270
13	0,418	0,425
14	0,296	0,226
15	0,441	0,465
16	0,333	0,345
17	0,497	0,347
18	0,382	0,260
19	0,464	0,444
20	0,411	0,292
21	0,459	0,444
22	0,361	0,280
23	0,390	0,473
24	0,297	0,364

TABLE 6.1 – Pondérations individuelles des dimensions pour l'espace 2D.

Index du participant	Dimension 1	2	3
1	0,458	0,332	0,282
2	0,263	0,333	0,385
3	0,371	0,299	0,352
4	0,401	0,344	0,298
5	0,444	0,363	0,310
6	0,335	0,419	0,367
7	0,338	0,238	0,402
8	0,349	0,339	0,270
9	0,364	0,389	0,296
10	0,226	0,380	0,398
11	0,333	0,345	0,393
12	0,274	0,382	0,260
13	0,350	0,362	0,411
14	0,292	0,379	0,392
15	0,361	0,280	0,371
16	0,399	0,297	0,364
17	0,362	0,414	0,359
18	0,332	0,358	0,386
19	0,374	0,327	0,397
20	0,393	0,307	0,234
21	0,349	0,336	0,321
22	0,266	0,307	0,333
23	0,322	0,237	0,365
24	0,397	0,354	0,333

TABLE 6.2 – Pondérations individuelles des dimensions pour l'espace 3D.

6.3.3 Interprétation de l'espace à 2 dimensions

Nous avons appliqué la démarche explicitée en fin de section 6.1 afin de rechercher les descripteurs explicatifs des dimensions de l'espace 2D. Cette section présente les résultats finaux obtenus.

À l'écoute, la dimension 1 de l'espace 2D semble discriminer un aspect particulier des sons. En effet, à une extrémité, on y trouve des sons comme **CRS2 v1 rep** présentant un contenu en basses fréquences très important, à cause de la forte audibilité du son produit par le groupe moto-ventilateur dans ce cas précis. Tandis que les sons à l'opposé de cette dimension voient leur contenu en basses fréquences bien amoindri. Il est alors cohérent de constater que cette dimension est fortement corrélée avec la sonie spécifique (modèle de Zwicker et Fastl [167, 8]) dans les 2 premières bandes de Bark N_1 et N_2 correspondant à la bande de fréquences se situant entre 0 et 200 Hz. Le coefficient de corrélation de Bravais-Pearson (voir section B.1 en annexe) le plus élevé est obtenu quand il est calculé entre les valeurs pour chaque son de la somme de la sonie spécifique dans ces 2 bandes, et leurs valeurs sur la dimension 1 : $r(ddl = 14) = 0,93$ ($p < 0,01$).

Il est aisé d'interpréter la seconde dimension à l'écoute. En effet, il est rapidement apparu que

cette dimension décrit de manière relativement continue le caractère *brillant*, allant de sons assez « mats », comme **CST3 v2 rep**, vers des sons plus « cinglants », comme **GNB1 v3 dif**. Il semble donc que cette dimension corresponde à la position générale du contenu spectral des sons sur l'échelle des fréquences. Il est par conséquent logique de constater des coefficients de corrélation élevés pour des descripteurs d'enveloppe spectrale tels que le centre de gravité spectral PSC (tel que défini en section 2.3.3, équation 2.12), souvent corrélé lui-même avec l'étendue spectrale PSS (telle que définie en section 2.3.3, équation 2.13) pour les sons de l'environnement. Le coefficient de corrélation le plus élevé est obtenu pour le calcul de l'acuité (telle que définie en section 2.3.3, équation 2.14) : $r(ddl = 14) = -0,94$ ($p < 0,01$).

Le tableau 6.3 résume les coefficients de corrélation de Bravais-Pearson (voir section B.1 en annexe) obtenus. Les figures 6.6 et 6.7 montrent le résultat de la régression linéaire entre les descripteurs psychoacoustiques et les dimensions correspondantes.

Descripteurs psychoacoustiques	Dimension 1	Dimension 2
N_1	$r(ddl = 14) =$ **0,84****	$r(ddl = 14) = 0,21$
N_2	**0,87****	0,36
$N_1 + N_2$	**0,93****	0,32
PSC	-0,39	**-0,91****
PSS	0,07	**-0,89****
Acuité	-0,28	**-0,94****

TABLE 6.3 – Coefficients de corrélation entre les descripteurs et les dimensions de l'espace perceptif 2D (** $p < 0,01$).

6.3.4 Interprétation de l'espace à 3 dimensions

Nous avons appliqué la démarche explicitée en fin de section 6.1 afin de rechercher les descripteurs explicatifs des dimensions de l'espace 3D. Cette section présente les résultats finaux obtenus.

Toutefois, il importe, tout d'abord, de comparer les espaces à 2 et à 3 dimensions afin de savoir à quel point les dimensions correspondantes diffèrent. Le tableau 6.4 répertorie les coefficient de corrélation des dimensions de l'espace 3D avec celles de l'espace 2D. Il apparaît clairement que les 2$^{\text{èmes}}$ dimensions des 2 espaces sont les mêmes, de même que les 1$^{\text{ères}}$ dans une moindre mesure. En revanche, la 3$^{\text{ème}}$ dimension n'est significativement liée à aucune des 2 dimensions de l'espace 2D, ce qui est facilement explicable par l'orthogonalité des dimensions d'un même espace. Cette dimension apporte donc un raffinement à l'espace 2D, à priori de moindre importance perceptive que les 2 premières.

		Espace 2D	
		dimension 1	dimension 2
Espace 3D	dimension 1	$r(ddl = 14) =$**0,85****	$r(ddl = 14) = 0,03$
	dimension 2	0,04	**0,96****
	dimension 3	-0,46	-0,34

TABLE 6.4 – Coefficients de corrélation entre les dimensions des espaces 2D et 3D (** $p < 0,01$).

Il est alors logique de constater à l'écoute que les deux premières dimensions de l'espace 3D ressemblent fortement aux deux dimensions de l'espace 2D. Les coefficients de corrélation obtenus avec

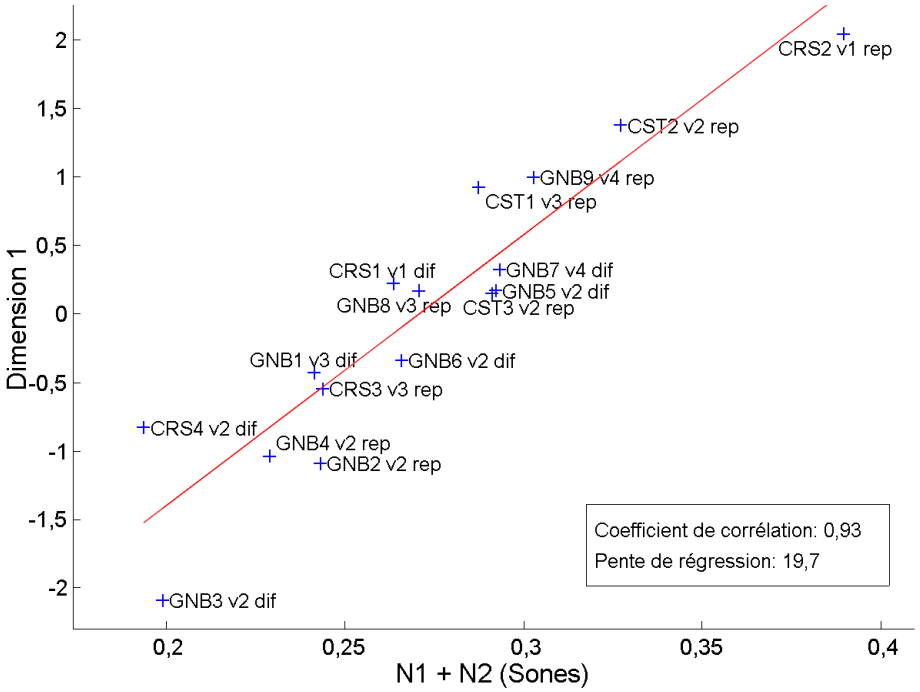


FIGURE 6.6 – Régression linéaire entre la somme des sonies spécifiques dans les 2 premières bandes de Bark et la 1e dimension de l'espace 2D.

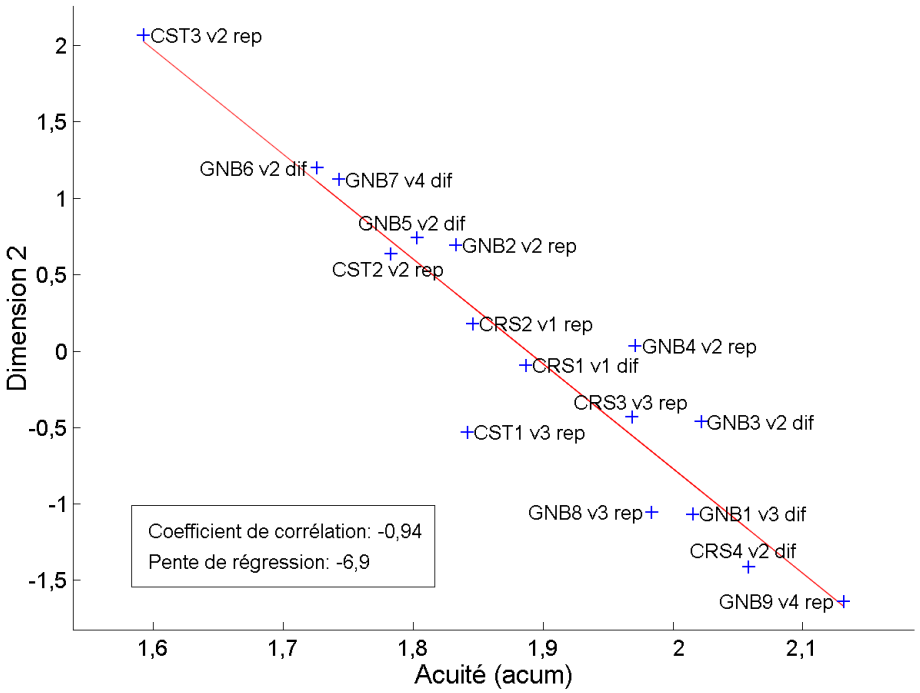


FIGURE 6.7 – Régression linéaire entre l'acuité (*Sharpness*) et la 2e dimension de l'espace 2D.

les 2 descripteurs psychoacoustiques expliquant les deux dimensions de l'espace 2D permettent en conséquence d'obtenir des coefficients de corrélation important pour les deux premières dimensions de l'espace 3D également. En effet, la somme de la sonie spécifique dans les deux premières bandes de Bark N_1 et N_2 est corrélée avec la dimension 1 à hauteur de $r(ddl = 14) = 0,85$, et l'acuité est corrélée avec la dimension 2 à hauteur de $r(ddl = 14) = -0,86$. On remarque toutefois que les coefficients de corrélations sont globalement moins élevés que pour l'espace 2D, tout en restant statistiquement significatifs ($p < 0,01$). Par ailleurs, il n'a pas été possible de mettre à jour de meilleures explications qualitatives (à l'écoute) ou quantitatives (en termes de coefficient de corrélation) de ces deux dimensions.

L'interprétation de la 3e dimension est celle qui a été le plus difficile. On constate toutefois à l'écoute que les sons présentant une valeur élevée sur cette dimension présent un caractère « sifflant » plus ou moins prononcé, et d'autant plus flagrant pour le son **GNB3 v2 dif** qui se détache clairement du reste du corpus sur cette dimension (voir figure 6.5). Un descripteur a été utilisé afin de refléter cette caractéristique, qui permet de se rapprocher de la notion d'*émergence harmonique* introduite en section 2.3.3. Le descripteur mis au point sera d'ailleurs par la suite nommé par ces mêmes termes, « Émergence harmonique ». Cette dernière est significativement corrélée ($p < 0,01$) avec la dimension 3 de l'espace à hauteur de $r(ddl = 14) = -0,87$.

Les figure 6.8, 6.9 et 6.10 montrent les régressions linéaires entre les 3 dimensions de l'espace et, respectivement, $N_1 + N_2$, l'Acuité, et l'Émergence harmonique. Le tableau 6.5 résume les coefficients de corrélation de Bravais-Pearson (voir section B.1 en annexe) entre les 3 dimensions de l'espace et les descripteurs mentionnés ci-dessus.

Descripteurs psychoacoustiques	Dimension 1	Dimension 2	Dimension 3
N_1	$r(ddl = 14) =$ **0,79****	$r(ddl = 14) =$ 0,20	$r(ddl = 14) =$ -0,28
N_2	**0,82****	0,39	-0,35
$N_1 + N_2$	**0,86****	0,31	-0,37
PSC	-0,27	**-0,83****	0,47
PSS	-0,13	**-0,83****	0,35
Acuité	-0,17	**-0,86****	0,46
Émergence harmonique	0,04	0,13	**0,87****

TABLE 6.5 – Coefficients de corrélation entre les descripteurs et les dimensions de l'espace perceptif 3D (** $p < 0,01$).

6.4 Discussion

Cette étude a permis d'identifier l'espace de timbre correspondant au corpus de sons étudié. Plus exactement, deux espaces ont été identifiés. Le premier, l'*espace 2D* composé donc de deux dimensions, permet d'expliquer une majeure partie des jugements de dissimilarité des auditeurs, et représente un compromis optimal entre faible erreur de prédiction des dimensions perceptives et dimensionalité réduite. Un second espace à trois dimensions, l'*espace 3D*, a également été mis en évidence, principalement pour des raisons pratiques liées à l'efficacité de prédiction de la qualité sonore (voir

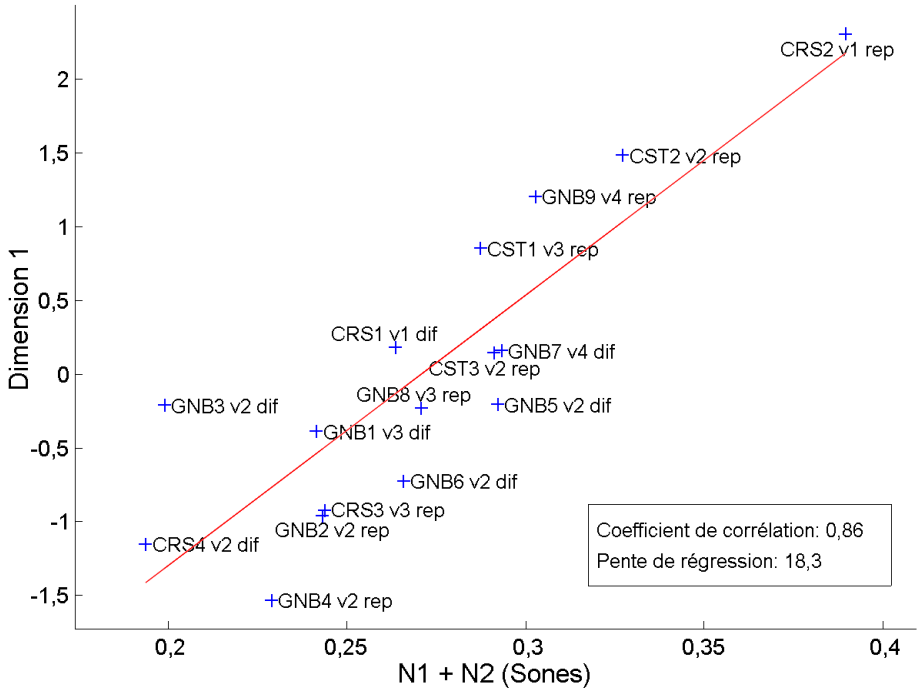


FIGURE 6.8 – Régression linéaire entre la somme des sonies spécifiques dans les 2 premières bandes de Bark et la 1e dimension de l'espace 3D.

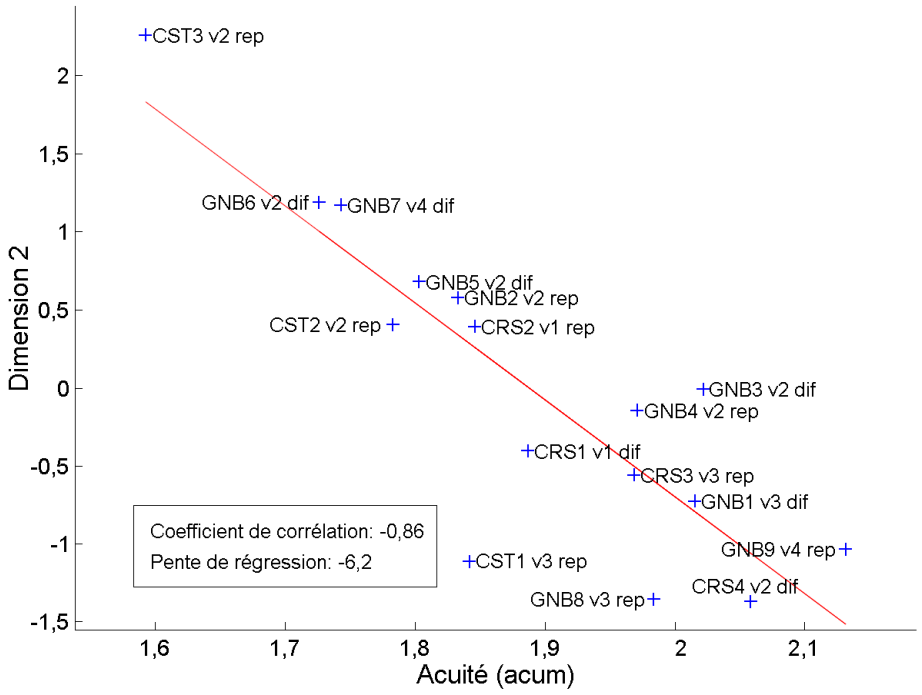


FIGURE 6.9 – Régression linéaire entre l'acuité (*Sharpness*) et la 2e dimension de l'espace 3D.

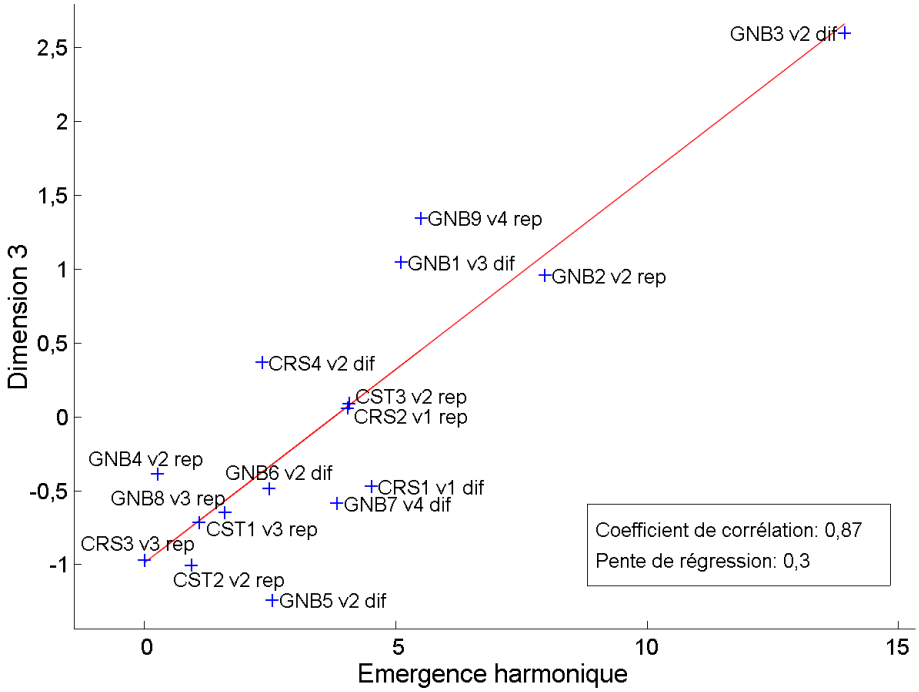


FIGURE 6.10 – Régression linéaire entre l'Émergence harmonique (défini sur une échelle arbitraire) et la 3e dimension de l'espace 3D.

section 7.6.2), mais également par analogie avec certaines études portant sur des sons similaires (notamment plusieurs études de Susini et al. [145, 146, 147]). Cet espace n'est toutefois pas contradictoire par rapport à l'espace 2D, car deux des trois dimensions sont fortement corrélées avec les deux dimensions de l'espace 2D, comme l'indique le tableau 6.4. L'espace 3D apporte donc principalement un raffinement par rapport à l'espace 2D, en adjoignant aux 2 dimensions déjà identifiées, une troisième dont l'importance perceptive est à priori moindre.

L'interprétation de ces espaces et la recherche de descripteurs explicatifs des dimensions ont permis de donner une signification perceptive plus concrète des espaces. Ainsi, il s'avère que les deux dimensions perceptives principales correspondent respectivement à la perception de l'intensité sonore dans les très basses fréquences (sonie dans les deux premières bandes de Bark, c'est-à-dire entre les fréquences 0 et 200 Hz du spectre, $N_1 + N_2$), et à la répartition de l'énergie acoustique sur l'échelle des fréquences (dans le cadre du modèle perceptif de l'*Acuité* et dans celui du centre de gravité spectral perceptif *PSC*). La troisième dimension de l'espace 3D correspond à une forme d'émergence harmonique traduisant principalement l'apparition d'une sorte de « sifflement », particulièrement notable pour le son **GNB3 v2 dif**.

Les descripteurs explicatifs identifiés ici semblent globalement cohérents avec la littérature sur le timbre des sons de l'environnement. En effet, une des conclusions de la partie de l'étude bibliographique qui concerne les descripteurs du timbre (voir section 2.1.2 et notamment les études de Susini et al. [145, 146], McAdams et al. [107], Susini et al. [147], Parizet et al. [118], Lemaitre et al. [100]), est que le percept de *brillance* – notamment expliqué par des descripteurs tels que l'acuité et le centre de gravité spectral – émerge systématiquement des études du timbre des sons de l'environnement, quelle que soit la catégorie de source sonore considérée. Cette conclusion est également renforcée par l'étude plus globale de différentes catégories de sons de l'environnement de Misdariis et al. [111].

L'*émergence harmonique* apparaît également dans la littérature lorsque l'on s'intéresse aux sons d'appareils incluant un moteur, bien que dans le cas présent elle désigne plutôt des émergences spécifiques de type « sifflement » dans des fréquences assez élevées, et non le son généré par le fonctionnement même du moteur, plutôt centré autour des basses fréquences. En revanche, nous pouvons observer qu'un percept souvent mis en évidence dans les études du timbre, n'est pas apparu dans notre étude : la *rugosité*. Ceci peut aisément s'expliquer par le fait que le type de son ayant permis l'émergence de ce paramètre présentait généralement une importante partie harmonique, et notamment une ou plusieurs fréquences fondamentales aisément détectables (voir notamment l'étude de Lemaitre et al. sur les sons de klaxons [100]), ce qui n'est pas le cas des sons de STA étudiés ici. Enfin, les résultats que nous avons obtenus rejoignent également ceux de la littérature concernant la dimensionnalité de l'espace de timbre, évaluée selon les études à 2 ou à 3.

Le principal point de divergence des résultats présentés dans ce chapitre avec ceux de la littérature concerne surtout la première dimension des espaces 2D et 3D : la sonie dans les deux premières bandes de Bark. Par ailleurs, la sonie spécifique, notamment dans cette partie du spectre, est un paramètre qui est également pris en compte dans le calcul de l'acuité, tout comme dans celui du PSC (voir notamment équation 2.12 en section 2.3.3). Il n'est pas illogique de supposer donc que les paramètres $N_1 + N_2$ et *Acuité* ne sont pas indépendants. Ainsi, il peut sembler étonnant, à première vue, de constater que ces deux paramètres sont corrélés avec deux dimensions différentes, compte tenu de l'orthogonalité qu'implique la définition des espaces obtenus par l'analyse MDS. Toutefois, il importe de garder à l'esprit deux éléments importants : d'une part les deux premières bandes de Bark ne représentent qu'une partie minoritaire du spectre, qui s'étend selon cette échelle sur 24 bandes ; et d'autre part, la partie du spectre à laquelle l'oreille humaine est la plus sensible – donc la partie du spectre à priori la plus importante afin de juger du contenu spectral global, ce que représente le percept d'acuité – se situe globalement de quelques centaines à quelques milliers de Hz, donc au-dessus de ces deux bandes. Ceci se confirme lorsque l'on évalue le coefficient de corrélation de Bravais-Pearson entre ces deux paramètres : $r(ddl = 14) = -0,43$, non-significatif à $p < 0,05$.

Il semble donc que, pour les sons de STA, la partie très basse fréquence du spectre et le reste de celui-ci soient ressentis par les auditeurs comme deux percepts distincts. Cela pourrait s'expliquer par la nature des sources sonores en question qui correspondent le plus souvent au mélange de différentes causes physiques. Les STA étudiés incluent un système de moto-ventilateur, dont le son produit se situe le plus souvent principalement dans les basses fréquences. Cela ne se traduit pas sous forme d'*émergence harmonique* comme dans le cas des études mentionnées en section 2.3.3, probablement parce que cette partie est moins audible dans les sons étudiés ici. En revanche, le son produit par le flux d'air est plus large-bande, et centré autour de fréquences plus élevées. On peut alors supposer que la différenciation des deux percepts vient d'une forme de distinction des deux types de cause physique du son par les auditeurs. Ce type d'hypothèse peut revêtir une certaine importance dans le cadre de la conception et de la mise au point de nouveaux produits, car l'identification plus formelle des causes physiques possibles de ces deux percepts pourrait alors fournir des pistes afin d'améliorer le ressenti des usagers.

Chapitre 7

Évaluation et prédiction de la qualité sonore des STA

Ce chapitre présente l'étude de la qualité sonore des STA. L'idée majeure est donc ici d'appliquer une méthodologie d'évaluation de la qualité sonore afin d'établir une échelle de préférence des auditeurs pour le corpus de sons de STA considéré. À terme, l'objectif est d'établir une métrique, sur la base des descripteurs du timbre, prédisant la qualité sonore perçue. Dans cette optique, la section 7.1 présente tout d'abord la problématique générale de l'évaluation et la prédiction de la qualité sonore. Les sections 7.2 et 7.3 présentent respectivement le protocole expérimental utilisé afin de réaliser une expérience d'évaluation comparée en condition de sonie réelle (voir problématique en section 7.1 pour plus de détails sur ce que l'on entend par « sonie réelle »), et les résultats obtenus. Les sections 7.4 et 7.5 présentent respectivement le protocole similaire utilisé en condition de sonie égalisée (voir problématique en section 7.1) et les résultats obtenus. La section 7.6 présente les résultats de la prédiction des échelles de qualité sonore. Enfin, la section 7.7 expose la discussion que soulèvent les résultats de cette étude.

7.1 Problématique

Cette étude représente d'un point de vue pragmatique le cœur de cette thèse. En effet, si on se place dans le cadre industriel qui caractérise ces travaux, il importe que l'on puisse estimer facilement la qualité sonore de tout nouveau STA qui se présenterait. Les expériences perceptives telles que celles présentées dans cette thèse représentent une importante mise en œuvre de différents types de « ressources » afin d'établir une échelle de qualité sonore : enregistrements sonores de nombreux STA, mise en place de protocoles expérimentaux, recrutement de participants, évaluation de descripteurs acoustiques et psychoacoustiques, analyses statistiques des réponses des auditeurs, ... Il est évident que ce type de procédure ne représente pas un moyen « facile » d'estimer la qualité sonore d'un STA. Du point de vue industriel, on souhaite donc pouvoir obtenir rapidement une mesure fiable de la qualité sonore sans recourir à de telles procédures à chaque nouveau STA à évaluer.

Lorsque l'on souhaite évaluer un nouveau STA, il importe donc d'être capable de « simuler » l'évaluation de la qualité sonore qui aurait été obtenue dans le cadre d'une expérience perceptive telle que celle présentée dans la suite de ce chapitre. Cela signifie qu'il faut pouvoir disposer d'une métrique fondée sur des éléments aisément mesurables, c'est-à-dire calculables à partir du signal sonore, et reflétant la qualité sonore ressentie par les usagers. Ceci souligne le caractère essentiel des différentes étapes qui constituent l'étude de la qualité sonore : la mesure perceptive de la qualité sonore sur un

ensemble représentatif, et sa prédiction par les descripteurs acoustiques et/ou psychoacoustiques pertinents pour la description du timbre. Ce travail doit donc être réalisé en amont dans le but d'établir des spécifications pour la conception et l'évaluation de nouveaux produits.

La première étape consiste donc ici à mesurer perceptivement la qualité sonore du corpus de sons étudié. Il s'agit alors de définir l'échelle de mesure sur laquelle celle-ci va être décrite. La principale hypothèse, assez naturelle, est que la qualité sonore peut être décrite par une échelle *unidimensionnelle*. En effet, si plusieurs attributs auditifs entrent en jeu lors du jugement perceptif de la qualité sonore, le résultat de ce jugement correspond à une notion unique, par rapport à laquelle les sons sont comparés de manière unilatérale : un son A est dit de « meilleure qualité sonore » qu'un son B, ce qui ne serait pas possible dans le cas d'un mode multidimensionnel de représentation. Par ailleurs, la littérature scientifique nous a amené à poser également une autre hypothèse, celle d'une échelle *continue*. En effet, cette propriété de l'échelle de qualité sonore est implicitement admise par la plupart des procédures expérimentales décrites en section 2.3.1 et la plupart des métriques de qualité sonore proposées et inventoriées en en fin de section 2.3.3.

De ces hypothèses et de la tendance naturelle à comparer deux sons en termes de qualité sonore découle la nature de l'échelle que nous allons mesurer : l'*échelle de préférence*. Afin d'identifier cette dernière dans le cas des sons considérés, la littérature scientifique nous offre une certaine variété de procédures expérimentales qui sont explicitées en section 2.3.1. Elles présentent chacune avantages et inconvénients et impliquent bien entendu des traitements statistiques différents des résultats. Pour son bon compromis entre précision et faisabilité pratique de l'expérience pour un corpus d'une taille raisonnable, la procédure d'*évaluation comparée* a été choisie.

La seconde étape consiste à relier cette mesure de la qualité sonore aux attributs auditifs pertinents pour la description du timbre. Toutefois, un élément important doit être pris en compte : l'importance perceptive de la sonie. Nous avons empêché l'influence de ce paramètre dans le cadre des chapitres 5 et 6, car le but des études correspondantes était d'identifier les familles et l'espace de timbre, excluant, selon la définition du timbre donnée en section 2.1, toute variation de sonie. En revanche, il est naturel de supposer que cet attribut auditif joue un rôle prépondérant dans la qualité sonore perçue par les auditeurs. Il est donc indispensable de prendre en compte ce paramètre et donc de réaliser l'expérience d'évaluation comparée dans des conditions où les sons sont diffusés au niveau sonore auquel ils ont été enregistrés. L'intérêt de cette condition d'expérience est double : d'une part confirmer notre intuition concernant l'importance du percept de sonie vis-à-vis des autres attributs auditifs, et d'autre part confronter l'efficacité de la mesure de niveau sonore en dBA pour la prédiction des préférences à d'autres descripteurs plus raffinés de niveau perçu.

Toutefois, il est évident que nous souhaitons également estimer l'influence des attributs auditifs pertinents identifiés dans le cadre de l'étude du timbre des sons de STA (voir chapitre 6). En effet, si l'on peut supposer que la sonie explique à elle seule une grande partie de la qualité sonore perçue, il est indispensable de savoir comment cette dernière varie quand plusieurs sons présentent des valeurs de sonie équivalentes. En effet, il est dans ce cas toujours possible de les distinguer entre eux, et rien n'indique que les auditeurs ne soient pas capables de formuler des préférences. Il est donc indispensable de considérer également, pour l'évaluation de la qualité sonore, le corpus de sons, égalisé en sonie, pour lequel nous avons établi l'espace de timbre dans le chapitre 6.

L'expérience d'évaluation comparée a donc été réalisée dans deux conditions distinctes : en condition de *sonie réelle*, et en condition de *sonie égalisée*.

7.2 Protocole expérimental d'évaluation comparée - sonie réelle

7.2.1 Stimuli

Il est important de noter que cette expérience faisait partie d'une étude préliminaire réalisée en amont de toutes les autres expériences perceptives présentées dans cette thèse. Ainsi elle a notamment été réalisée avant l'étude des familles de sons de STA, ayant abouti à la sélection du corpus de travail décrit en section 5.3. Il est évident qu'il aurait été souhaitable que toutes les expériences aient été conduites sur un même corpus sonore de travail, mais devant le caractère probant des résultats de l'expérience présentée ici, associés à ceux de la littérature, il n'a pas été jugé nécessaire de la reproduire sur le corpus de travail. De plus, il faut garder à l'esprit que le corpus de travail utilisé pour les autres expériences a été établi dans le cadre de l'identification des différentes familles de timbre qui constituaient l'ensemble des enregistrements en notre possession (voir chapitre 5). Ce cadre excluait toute variation du percept de sonie, ce qui est incompatible avec les objectif de cette expérience.

Dix sons ont été sélectionnés parmi les enregistrements binauraux décrits en section 4.2 et ont été utilisés pour cette expérience. Seuls les enregistrements correspondant à la première vitesse de fonctionnement (la plus élevée) ont été considérés. Parmi les 16 sons binauraux restants, les 10 sons ont été sélectionnés pour prendre en compte les différents types de STA enregistrés (carrossés, gainables et cassettes). La liste des sons, d'une durée de 5 secondes et dont le niveau sonore se situe entre 41,6 et 50,6 dBA, constituant le corpus de cette expérience, ainsi que l'identifiant utilisé pour chacun d'eux dans la suite de l'étude, est présentée dans le tableau 7.1.

nom du système	vitesse	position	identifiant
CRS2	1	diffusion	**CRS2 v1 dif**
CRS4	1	diffusion	**CRS4 v1 dif**
GNB1	1	diffusion	**GNB1 v1 dif**
GNB3	1	diffusion	**GNB3 v1 dif**
GNB4	1	diffusion	**GNB4 v1 dif**
GNB6	1	diffusion	**GNB6 v1 dif**
GNB8	1	diffusion	**GNB8 v1 dif**
GNB9	1	diffusion	**GNB9 v1 dif**
CST2	1	diffusion	**CST2 v1 dif**
CST3	1	diffusion	**CST3 v1 dif**

TABLE 7.1 – Corpus sonore utilisé dans l'expérience d'évaluation comparée - sonie réelle (voir aussi tableau D.1 en annexe).

7.2.2 Participants

29 auditeurs volontaires (20 hommes, 9 femmes, entre 23 et 50 ans), n'ayant pas participé à une autre expérience réalisée dans le cadre de cette thèse, ont pris part à cette expérience. Aucun d'entre eux n'a fait mention d'un problème majeur d'audition.

7.2.3 Matériel

Une interface graphique spécifique, assurant la lecture des sons sur l'interface audio et l'enregistrement des réponses des participants, a été programmée en LabVIEW 7.0 pour cette expérience.

Les sons ont été diffusés par une interface RME Fireface 400 dans un casque fermé [1] BeyerDynamic DT 770 Pro en écoute dichotique (signaux différents dans les deux oreilles). Ce mode de reproduction sonore se justifie par l'emploi d'enregistrements binauraux en guise de stimuli. L'expérience a eu lieu dans une salle non traitée acoustiquement, mais relativement isolée de toute source de bruit extérieur (ce qui justifiait par ailleurs l'emploi d'un casque de type fermé).

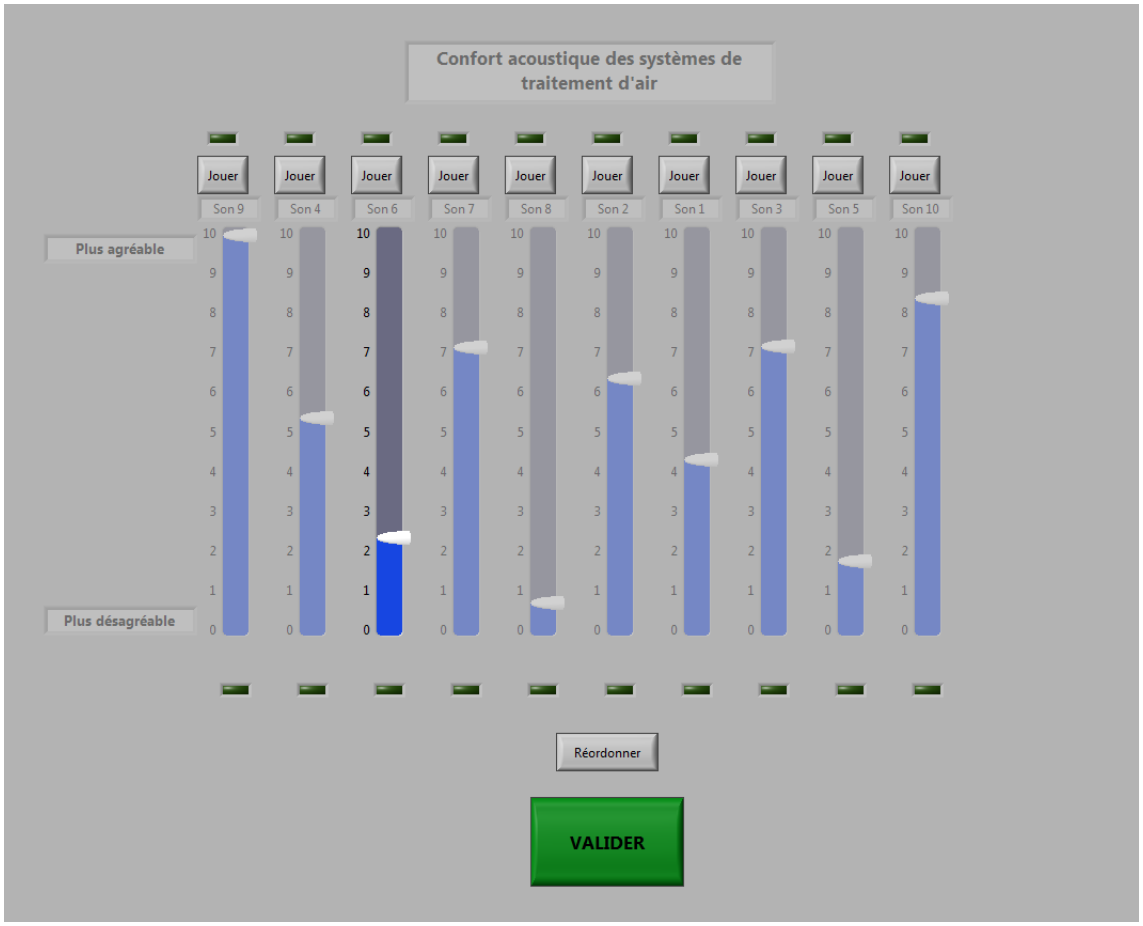


FIGURE 7.1 – Interface graphique de l'expérience d'évaluation comparée - sonie réelle.

7.2.4 Procédure

La procédure expérimentale utilisée ici est celle d'*évaluation comparée* (voir section 2.3.1), offrant un compromis entre les procédures d'évaluation absolue et de comparaison par paires. Au début de l'expérience, les participants reçoivent une consigne écrite retranscrite en section E.3 en annexe, présentant le contexte de l'étude et expliquant la tâche à accomplir. Une méthode d'évaluation comparée (voir 2.3.1) est utilisée. Sur l'interface associée à cette procédure (voir figure 7.1), les participants sont amenés à écouter chaque son du corpus, et à en évaluer sur une échelle allant de 0 à 10 le degré de désagrément (du plus désagréable au plus agréable). Une forte autonomie est laissée aux participants qui peuvent choisir la stratégie d'écoute et d'évaluation qui leur convient le mieux, puisqu'ils peuvent réécouter chaque son et revoir leurs évaluations autant de fois qu'ils le souhaitent et dans l'ordre de leur préférence. Un bouton permet par ailleurs de faciliter les comparaisons des sons en classant ceux-ci en fonction des évaluations déjà effectuées. Lorsque les participants sont satisfaits

1. Un casque fermé, par opposition à un casque ouvert, isole l'auditeur des sources de bruits extérieurs.

de leurs évaluations, ils valident mettant ainsi fin à l'expérience (le bouton correspondant n'est bien entendu accessible qu'une fois chaque son écouté et évalué au moins une fois). La durée moyenne de la procédure dans le cas de cette expérience a été de 9 minutes.

7.3 Résultats et analyse – sonie réelle

Les données brutes issues de cette expérience prennent la forme d'un jeu d'évaluations des sons pour chaque participant. L'analyse naturelle de ces résultats consiste à moyenner pour chaque son les évaluations de l'ensemble des participants, permettant ainsi d'obtenir une évaluation globale de la qualité sonore des éléments du corpus. La figure 7.2 permet d'observer les évaluations moyennes et l'écart-type des évaluations individuelles pour chaque son. On peut alors constater que les écarts-types sont relativement faibles devant la dynamique de l'échelle moyenne. En effet, la moyenne de ceux-ci et_m vaut 1,67, quand l'écart-type des valeurs moyennes de l'échelle m_{et}, traduisant la dynamique de celle-ci, vaut 2,58.

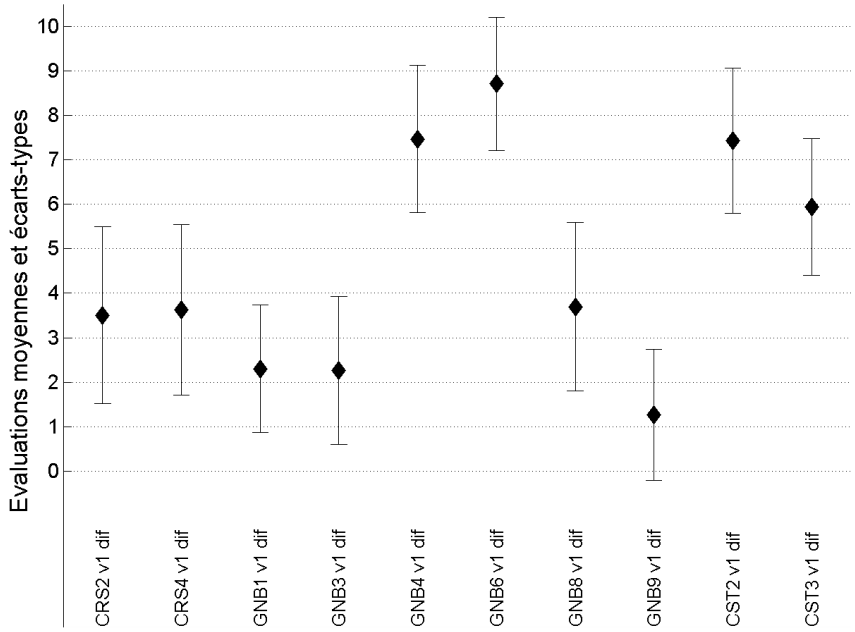


FIGURE 7.2 – Évaluations moyennes et écarts-types du désagrément pour l'expérience d'évaluation comparée. L'échelle va de '0' pour les plus désagréables à '10' pour les plus agréables.

Il convient également, par souci de cohérence avec l'analyse des évaluations des expériences exposées dans les parties subséquentes de ce document, d'observer à quel point l'échelle moyenne constitue un consensus des participants. À cet effet, le coefficient de concordance de Kendall W et le coefficient de corrélation de rang de Spearman $\bar{r}_s$, qui traduisent de manière plus globale la cohérence inter-participants des résultats, ont été évalués. Ces deux coefficients permettent de quantifier le degré de concordance, parmi les participants, de l'ordonnancement (ou rang) des sons sur l'échelle des évaluations. L'idée majeure est, grossièrement, de constater, pour l'ensemble des participants, si les sons évalués comme les plus désagréables et ceux évalués comme les moins désagréables sont bien les mêmes, ou non (voir section C.1 en annexe pour le détail du calcul de ces coefficients). Les valeurs possibles pour ces deux coefficients sont dans la gamme $\{-1;1\}$. Pour chacun d'eux, la valeur '1' est atteinte lorsque les résultats sont parfaitement concordants, et la valeur '-1', quand ils sont in-

versement concordant [2], tandis que la valeur '0' signifie qu'il n'y a aucun lien entre les résultats des participants. Par ailleurs, contrairement au coefficient de concordance de Kendall, la valeur du coefficient de corrélation de rang de Spearman peut être comparée aux valeurs critiques du coefficient de corrélation de Bravais-Pearson (voir tableau B.1 en annexe B).

Pour l'expérience réalisée, ces deux coefficients valent respectivement $W = 0,84$ et $\bar{r_s} = 0,83$. Les valeurs de ces deux coefficients étant relativement proches de '1', Les résultats individuels semblent globalement concordants. Le tableau 7.2 résume les différents éléments statistiques inter-participants.

et_m	1,67
m_{et}	2,58
W	0,84
$\bar{r_s}$	0,83

TABLE 7.2 – Statistiques inter-participants de l'expérience d'évaluation comparée en condition de sonie réelle (écart-type moyen entre les participants et_m, écart-type entre les évaluations moyennes des sons m_{et}, coefficient de concordance de Kendall W et coefficient de corrélation de rang de Spearman $\bar{r_s}$).

Toujours dans un objectif de cohérence d'analyse avec les expériences présentées dans la suite de ce document, il convient également d'évaluer la divergence ou la concordance des résultats des participants les uns par rapport aux autres à l'aide d'une *analyse de cluster* appliquée aux données individuelles de l'expérience (voir section C.2 en annexe). Le dendrogramme traduisant cette représentation est affiché en figure 7.3.

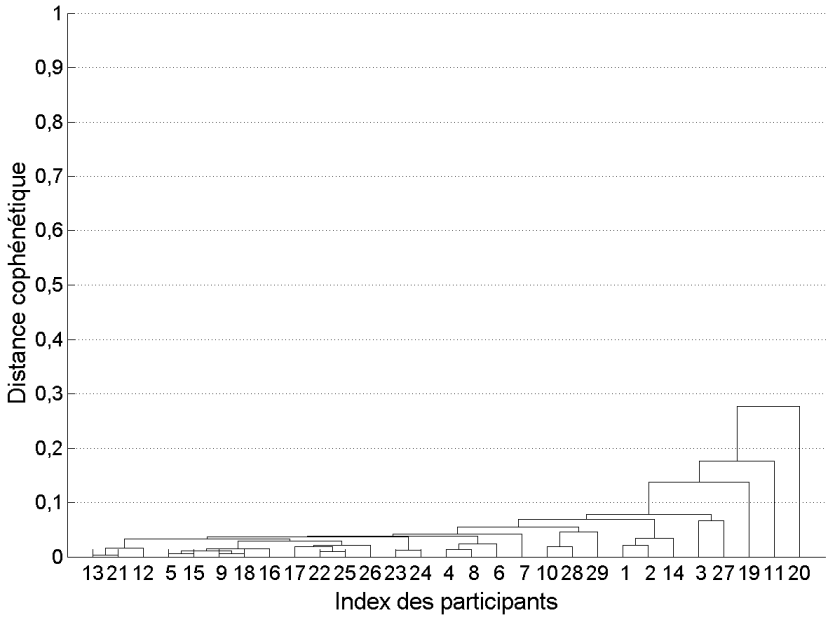


FIGURE 7.3 – Dendrogramme issu des corrélations inter-participant pour l'expérience d'évaluation comparée en condition de sonie réelle.

Ainsi, le participant 20 à l'extrême droite du dendrogramme (et dans une moindre mesure les participants 11 et 19) semble se détacher des autres participants. Toutefois, la hauteur du nœud où ce

2. Ce cas de figure semble toutefois impossible lorsque l'on compare les résultats de plus de 2 participants, comme dans le cas présent.

participant et le cluster formé par l'ensemble des autres participants se rejoignent (*distance cophénétique*) reste toutefois assez basse (moins de 0,3).

L'ensemble de ces éléments statistiques a conduit à ce qu'aucun des participants ne soit retiré du panel. Ainsi, l'échelle moyenne affichée en figure 7.2 représente l'échelle de qualité sonore retenue dans l'optique de sa prédiction sur la base de descripteurs audio (voir section 7.6).

7.4 Protocole expérimental d'évaluation comparée - sonie égalisée

L'idée directrice de cette expérience est d'établir une échelle de qualité sonore, non pas lorsque les sons sont diffusés à sonie réelle comme c'est le cas dans l'expérience décrite en section 7.2, mais lorsque l'ensemble des sons sont perçus à sonie égale. En effet, comme l'indique la littérature, le percept de sonie est le principal facteur influençant les préférences des auditeurs (ce qui est de plus confirmé par les résultats de régression linéaire exposés en section 7.6). Ainsi, le but est ici de reproduire une procédure expérimentale similaire à celle exposée en section 7.2, mais avec un corpus sonore égalisé en sonie.

7.4.1 Stimuli

Le corpus sonore utilisé lors de cette expérience est le même que celui utilisé pour l'expérience de mesure de similarités décrite en section 6.2, c'est-à-dire celui identifié à l'issue du chapitre 5 et énuméré dans le tableau 5.5. Il s'agit donc de 16 sons monophoniques égalisé en sonie [3] d'une durée de 4 secondes, et dont le niveau acoustique varie de 38,1 à 40,1 dBA.

7.4.2 Participants

19 auditeurs volontaires (11 hommes, 8 femmes, entre 22 et 25 ans), n'ayant pas participé à une autre expérience réalisée dans le cadre de cette thèse, ont pris part à cette expérience. Aucun d'entre eux n'a fait mention d'un problème majeur d'audition.

7.4.3 Matériel

Une interface graphique spécifique, assurant la lecture des sons sur l'interface audio et l'enregistrement des réponses des participants, a été programmée en LabVIEW 2010 (figure 7.4) pour cette expérience. Les sons ont été diffusés par une interface RME Fireface 800 dans un casque ouvert Sennheiser HD650. Le mode de reproduction sonore sur casque a été employé car c'est le seul qui garantit d'obtenir une écoute diotique (signal identique dans les deux oreilles). L'expérience a eu lieu dans une cabine audiométrique IAC à double paroi.

7.4.4 Procédure

La procédure expérimentale utilisée ici est la même procédure d'*évaluation comparée* que celle exposée en section 7.2. En résumé, après avoir reçu une consigne écrite de l'expérience (voir section E.3 en annexe), les participants doivent évaluer chaque son du corpus sur une échelle allant de 0 à 10 le degré de désagrément (du plus désagréable au plus agréable), tout en pouvant réécouter et modifier leurs évaluations tout au long de l'expérience (voir figure 7.4).

3. L'égalisation en sonie automatique fondée sur le modèle de Zwicker explicitée en section 5.2.1 s'est avérée satisfaisante à l'oreille, et il n'a pas été jugé nécessaire d'effectuer expérimentalement une égalisation en sonie, comme c'est le cas en section 8.3.3.

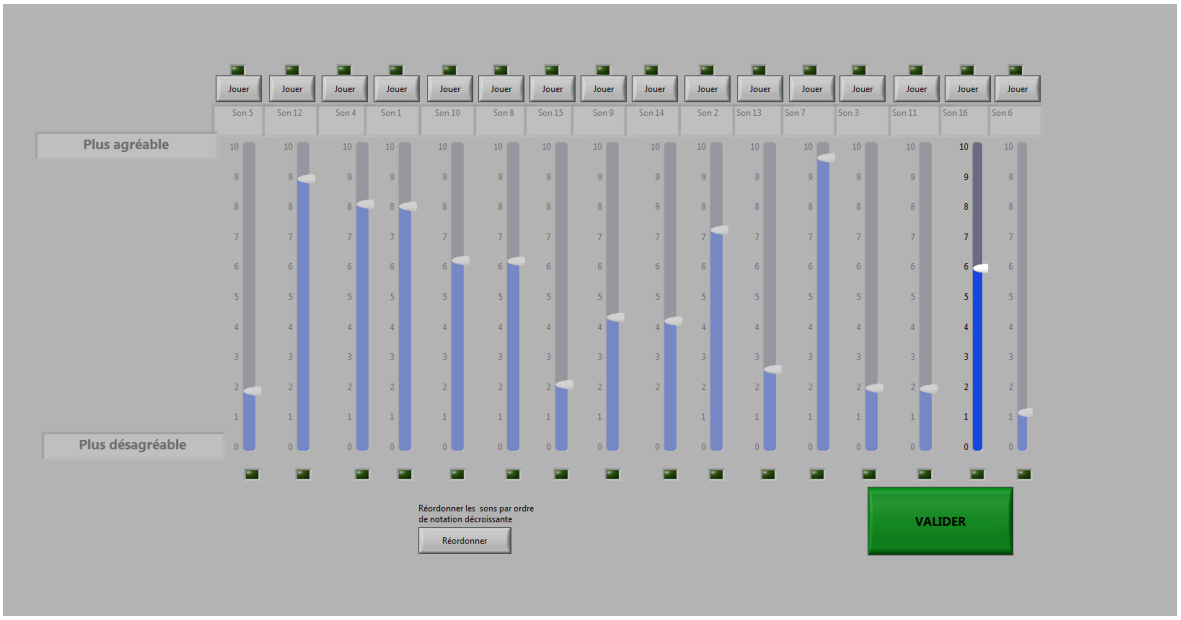


FIGURE 7.4 – Interface graphique de l'expérience d'évaluation comparée - sonie égalisée.

La durée moyenne de la procédure dans le cas de cette expérience a été de 14 minutes.

7.5 Résultats et analyse – sonie égalisée

Tout comme pour l'expérience en sonie réelle (section 7.3), les données brutes issues de cette expérience prennent la forme d'un jeu d'évaluations des sons pour chaque participant. Les mêmes analyses ont été appliquées ici. Tout d'abord, les évaluations moyennes et leurs écarts-types sont affichés en figure 7.5. À la différence des résultats obtenus en condition de sonie réelle (voir figure 7.2), les écarts-types sont ici très importants. En effet, la moyenne de ceux-ci et_m vaut 2,67, quand l'écart-type des valeurs moyennes de l'échelle m_{et}, traduisant la dynamique de celle-ci, vaut 1,70. De même, les valeurs du coefficient de concordance de Kendall et du coefficient de corrélation de rang de Spearman, qui traduisent globalement le degré de concordance de l'ordonnancement des sons sur l'échelle des évaluations par les différent participants (voir sections 7.3, et C.1 en annexe, pour plus de détails) valent respectivement $W = 0,29$ et $\bar{r}_s = 0,25$, et sont par conséquent bien inférieurs à ceux obtenus en condition de sonie réelle, et très éloignés de la valeur idéale '1'. Ces éléments statistiques inter-participants sont résumés dans le tableau 7.3.

et_m	2,67
m_{et}	1,70
W	0,29
$\bar{r}_s$	0,25

TABLE 7.3 – Statistiques inter-participants de l'expérience d'évaluation comparée en condition de sonie égalisée (écart-type moyen entre les participants et_m, écart-type entre les évaluations moyennes des sons m_{et}, coefficient de concordance de Kendall W et coefficient de corrélation de rang de Spearman $\bar{r}_s$).

Une *analyse de cluster* a été appliquée aux données individuelles de l'expérience (voir section C.2 en annexe) afin d'observer les divergences entre participants qui expliquent ces statistiques. Le den-

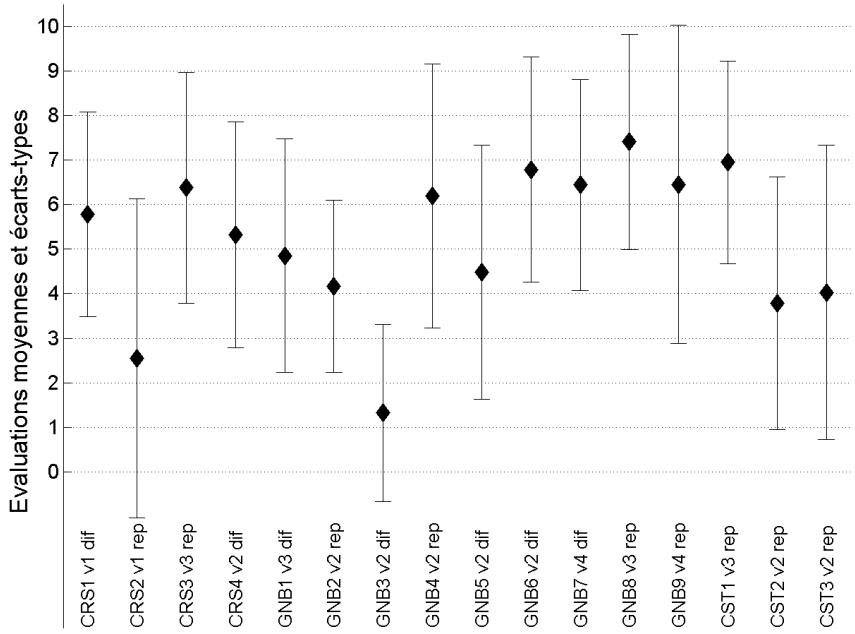


FIGURE 7.5 – Évaluations moyennes et écarts-types du désagrément pour l'expérience d'évaluation comparée en condition de sonie égalisée. L'échelle va de '0' pour les plus désagréables à '10' pour les plus agréables.

drogramme traduisant la représentation hiérarchique du panel de participants pour cette expérience est affiché en figure 7.6.

Ce dendrogramme présente tout d'abord une allure bien différente de celle du dendrogramme de l'expérience en condition de sonie réelle (voir figure 7.3). En effet, il apparaît que les hauteurs de fusions des différents *cluster* sont globalement plus élevées que celles du dendrogramme sus-cité. Ceci est cohérent avec les statistiques observées sur la figure 7.5 et dans le tableau 7.3. Par ailleurs, un autre phénomène semble évident à la vue du dendrogramme : un groupe de 4 participants (participants 9, 12, 14 et 4, à l'extrême droite du dendrogramme) présente des résultats modérément similaires, mais manifestement très différents de ceux de l'ensemble des autres participants. En effet, la distance cophénétique entre le cluster formé par ces 4 participants et celui formé par tous les autres (aux alentours de 0,55, ce qui correspond à un coefficient de corrélation moyen de -0,1 – voir formule C.3 en annexe) est largement supérieure aux plus grandes distances cophénétiques entre participants de l'un ou l'autre de ces deux clusters.

Il est possible que ces 4 participants représentent une tendance différente quant à la perception des sons et de la qualité associée, mais compte tenu de leur nombre, il est impossible de les considérer en tant que tel. Par conséquent, il ont été considérés comme des *outliers*, dont les résultats ne sont pas compatibles avec la tendance générale, et ont donc été retirés du panel de participant.

La figure 7.7 montre les évaluations moyennes et les écarts-types inter-participants obtenus en ne prenant pas en compte les résultats de ces 4 participants. En conséquence, la moyenne des écarts-types entre les évaluations des participants restants et_m devient 2,25, quand l'écart-type des valeurs moyennes de l'échelle m_{et} augmente à 2,19. De même les valeurs du coefficient de concordance de Kendall et du coefficient de corrélation de rang de Spearman augmentent nettement jusqu'à respectivement $W = 0,50$ et $\bar{r}_s = 0,47$. Sans être entièrement satisfaisants, ces éléments statistiques inter-participants confirment une nette amélioration de la représentativité de l'échelle de qualité sonore obtenue. Ils sont résumés dans le tableau 7.4.

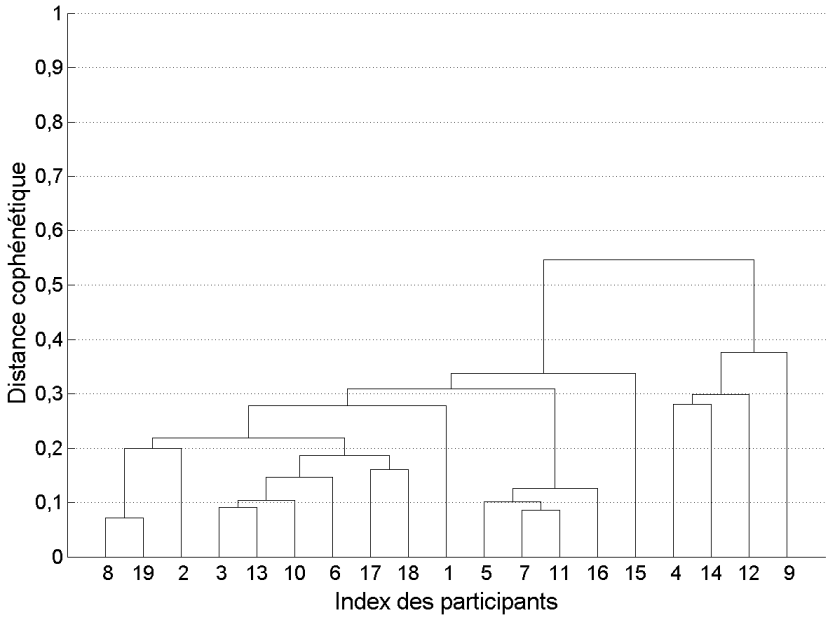


FIGURE 7.6 – Dendrogramme issu des corrélations inter-participant pour l'expérience d'évaluation comparée en condition de sonie égalisée.

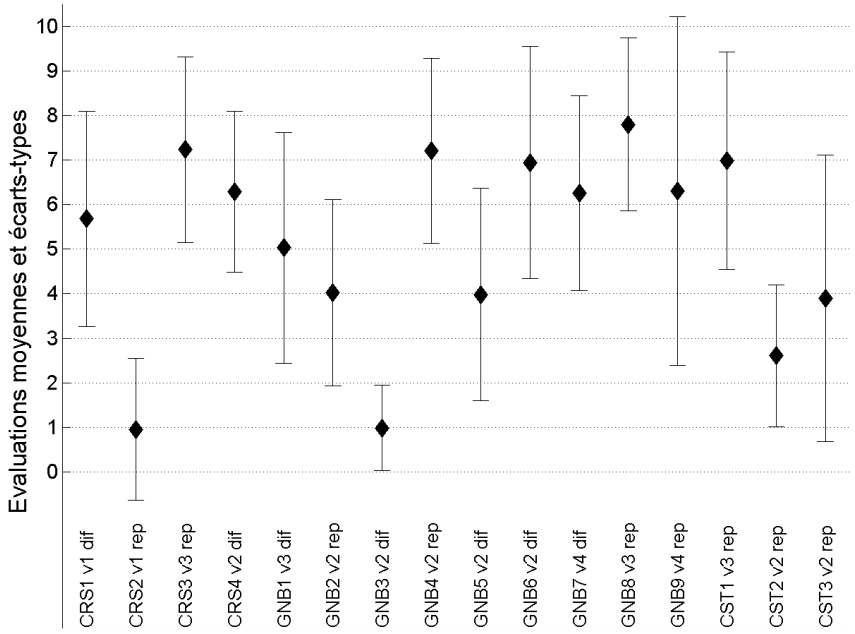


FIGURE 7.7 – Évaluations moyennes et écarts-types du désagrément pour l'expérience d'évaluation comparée en condition de sonie égalisée (sans les résultats des 4 *ouliers*). L'échelle va de '0' pour les plus désagréables à '10' pour les plus agréables.

7.6 Prédiction des échelles de qualité sonore

7.6.1 Échelle de qualité sonore en condition de sonie réelle

Lorsque les sons sont diffusés à sonie réelle, leur écoute permet de facilement distinguer d'évidentes variations d'intensité sonore perçue. Il paraît alors naturel de tenter d'effectuer une régression linéaire avec un descripteur psychoacoustique censé quantifier cet attribut auditif. Dans cette

et_m	2,25
m_{et}	2,19
W	0,50
$\bar{r}_s$	0,47

TABLE 7.4 – Statistiques inter-participants de l'expérience d'évaluation comparée en condition de sonie égalisée, sans les résultats des 4 *ouliers* (écart-type moyen entre les participants et_m, écart-type entre les évaluations moyennes des sons m_{et}, coefficient de concordance de Kendall W et coefficient de corrélation de rang de Spearman $\bar{r}_s$).

optique, une régression linéaire (voir section B.2 en annexe) a été effectuée entre l'échelle des évaluations moyennes et différents descripteurs d'intensité sonore : le calcul du niveau acoustique en dB SPL L, du niveau acoustique pondéré en dBA L_A et le résultat d'un calcul de sonie fondé sur le modèle de Zwicker et Fastl [8] N (voir section 2.3.3, notamment l'équation 2.3 et la figure 2.5, et l'équations 2.7, respectivement).

Les coefficients de corrélation de Bravais-Pearson (voir section B.1 en annexe) obtenus sont affichés dans le tableau 7.5. Ils établissent un lien particulièrement fort entre l'échelle de qualité sonore mesurée et le niveau acoustique mesuré en dBA d'une part et le descripteur de sonie d'autre part. En effet, si la corrélation avec le niveau acoustique exprimé en dB SPL n'est pas significative ($r_L(ddl = 8) = -0,34$ ($p > 0,1$)), le coefficient de corrélation de Bravais-Pearson obtenu lorsque le niveau est exprimé en dBA est, d'un point de vue statistique, aussi significatif que celui obtenu avec la sonie N : $r_{L_A}(ddl = 8) = -0,97$ ($p < 0,01$) et $r_N(ddl = 8) = -0,98$ ($p < 0,01$). Ces éléments ne démontrent pas d'apport significatif de la sonie par rapport au niveau acoustique en dBA, en termes d'efficacité de prédiction de la qualité sonore pour le corpus de sons étudié. Cela ne signifie pas que le modèle de sonie de Zwicker et Fastl n'est pas un meilleur descripteur du percept d'intensité sonore que le niveau acoustique mesuré en dBA dans le cas général, mais simplement qu'il apporte peu dans le cas de la qualité sonore des sons étudiés ici, probablement à cause de la nature large-bande des sons considérés.

Descripteur	Coefficient de corrélation de Bravais-Pearson
L	$r(ddl = 14)$ = -0,34
L_A	**-0,97****
Sonie (ISO532) N	**-0,98****

TABLE 7.5 – Coefficients de corrélation de Bravais-Pearson pour l'échelle de qualité sonore en condition de sonie réelle (** $p < 0,01$).

Enfin, les figures 7.8 et 7.9 montrent les valeurs de l'échelle de qualité sonore mesurée en fonction de, respectivement, le niveau acoustique en dBA L_A et la sonie N, ainsi que, dans les deux cas, la droite de régression obtenue (en rouge). Si les coefficients de corrélation n'ont pas fait apparaître d'amélioration significative statistiquement entre le niveau acoustique en dBA et la sonie, il semble toutefois que les points représentant certains sons sur ces figures se trouvent légèrement rapprochés de la droite de régression dans le cas de la sonie (pour les valeurs élevées de ces descripteurs tout du moins).

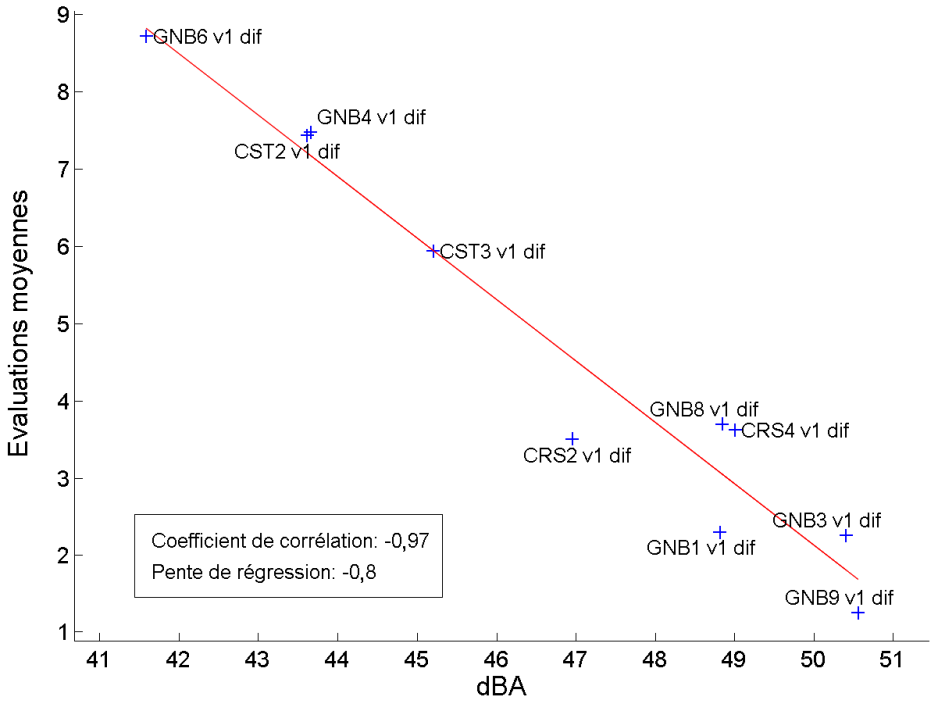


FIGURE 7.8 – Régression linéaire entre l'échelle des évaluations moyennes et le niveau acoustique en dBA (expérience en condition de sonie réelle).

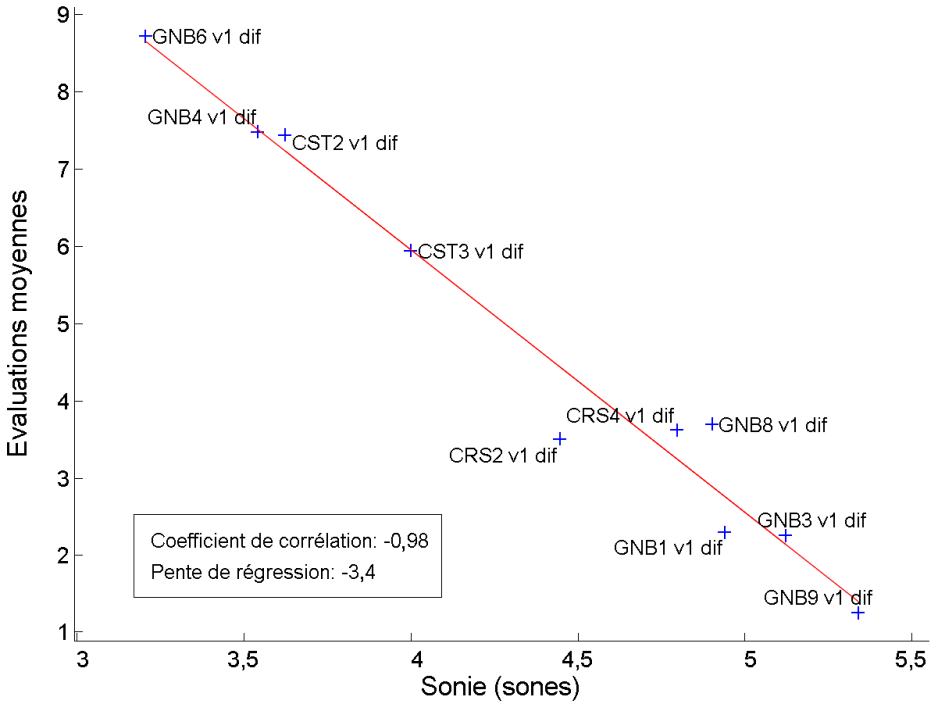


FIGURE 7.9 – Régression linéaire entre l'échelle des évaluations moyennes et la sonie (expérience en condition de sonie réelle).

7.6.2 Échelle de qualité sonore en condition de sonie égalisée

Il importe ici de relier une variable dépendante continue – l'échelle de qualité sonore établie en section 7.5 – à un jeu de trois variables indépendantes continues – les descripteurs du timbre identifiés au chapitre 6. Une régression linéaire multiple a donc été appliquée entre les descripteurs du timbre et l'échelle de qualité sonore. Le modèle à appliquer est simplement :

$$Q = a_0 + a_1 \cdot d_1 + a_2 \cdot d_2 \ (+a_3 \cdot d_3) \tag{7.1}$$

où Q est la valeur sur l'échelle de qualité sonore, d_1, d_2 et d_3 sont les descripteurs, et les $\{a_i\}$ sont les constantes du modèle à rechercher [4].

Le tableau 7.6 montre les coefficients de corrélation entre l'échelle de qualité sonore et chacun des 3 descripteurs identifiés ($N_1 + N_2$, Acuité et Émergence harmonique, voir section 6.3). On peut déjà constater qu'il ne semble pas y avoir de lien direct manifeste entre les descripteurs $N_1 + N_2$ et Acuité, et l'échelle de préférence, et que, en revanche, il existe un lien modérément significatif ($p < 0,02$) entre le descripteur Émergence harmonique et cette même échelle. Cela s'explique de nouveau par le son **GNB3 v1 dif** qui présente un sifflement particulièrement audible. En effet, il était manifeste que ce sifflement pénalisait fortement l'appréciation de ce son par les participants, et celui-ci offrait ainsi le meilleur consensus. Il en découle que l'échelle de préférence est fortement dirigée par cette spécificité. Or cette spécificité a déjà été identifiée comme la principale explication de la 3e dimension de l'espace 3D exposé en section 6.3. Il est donc logique de trouver une corrélation significative entre le descripteur expliquant le mieux cette dimension et l'échelle de préférence.

Descripteur	Coefficient de corrélation de Bravais-Pearson
$N_1 + N_2$	$r(ddl = 14)$ = -0,32
Acuité	0,20
Émergence harmonique	**-0,59***

TABLE 7.6 – Coefficients de corrélation entre l'échelle de préférence et les descripteurs $N_1 + N_2$, Acuité et Émergence harmonique (*$p < 0,02$).

Les figures 7.10 et 7.11 affichent la correspondance entre les évaluations moyennes de la qualité sonore et le résultat de la régression linéaire multiple à l'aide des jeux de respectivement 2 et 3 descripteurs précédemment identifiés. La figure 7.11 montre également la droite de régression obtenue. Comme attendu, il apparaît que la prise en compte de l'Émergence harmonique améliore considérablement le résultat de la régression, puisque l'on passe d'un coefficient de corrélation de $r(ddl = 13) = 0,33$ pour la régression à l'aide des deux descripteurs de l'espace 2D ($N_1 + N_2$ et Acuité), à $r(ddl = 12) = 0,78$, statistiquement significatif à $p < 0,01$, lorsque l'Émergence harmonique est également incluse dans la régression (le degré de liberté en moins étant lié à l'usage d'une constante supplémentaire dans le modèle). Enfin, le tableau 7.7 affiche, pour chaque son, les valeurs des descripteurs du timbre et des prédicteurs Q obtenus respectivement avec 2 et 3 descripteurs.

Compte tenu de l'importance manifeste du descripteur d'Émergence harmonique dans la prédiction des jugements de qualité sonore, une *régression multiple pas à pas* (voir l'ouvrage de Draper et Smith [50]) a été appliquée à ces données. Le principe est de sélectionner les variables indépendantes – les descripteurs du timbre en l'occurrence – en fonction de leur propension à expliquer en

4. La constante a_0 n'a pas d'intérêt particulier vis-à-vis de l'efficacité relative du prédicteur Q, elle sert juste à calquer, autant que faire se peut, la dynamique de ce dernier sur celle de l'échelle de qualité sonore (de '0,9' à '7,8' en l'occurrence).

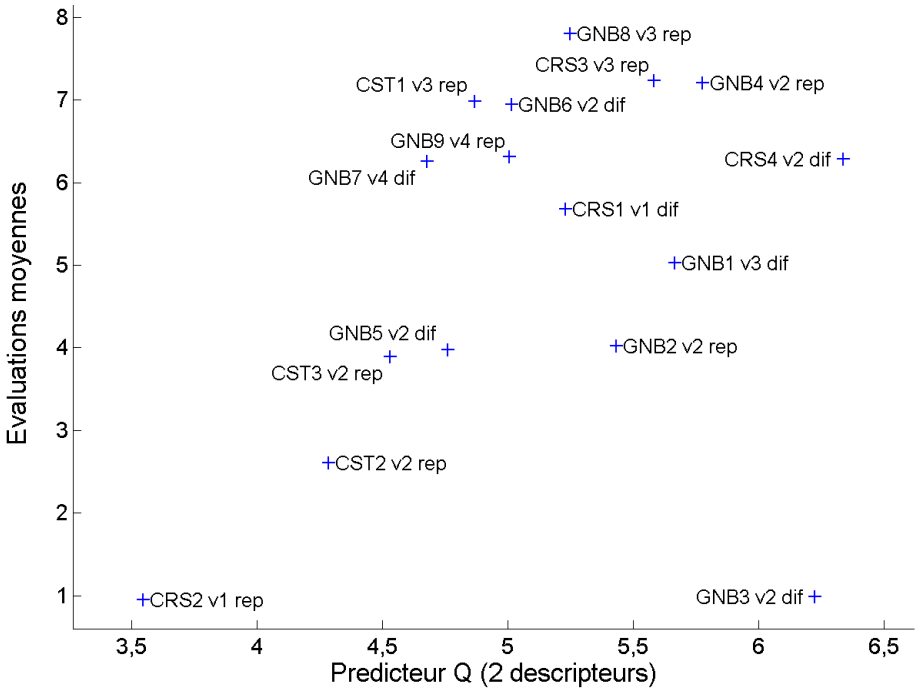


FIGURE 7.10 – Régression linéaire multiple entre l'échelle de préférence et les descripteurs $N_1 + N_2$ et Acuité (expérience en condition de sonie égalisée).

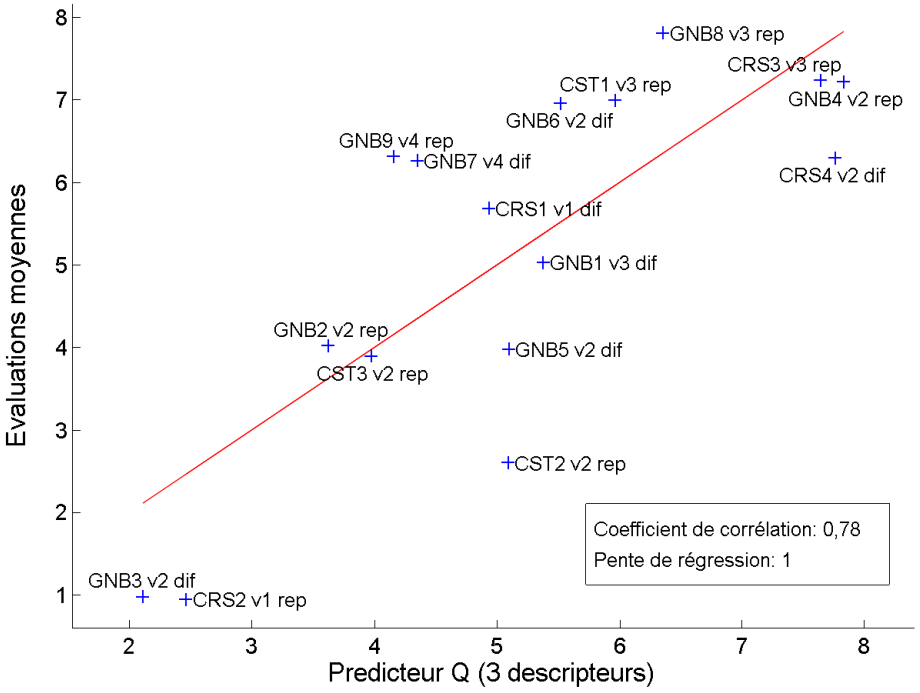


FIGURE 7.11 – Régression linéaire multiple entre l'échelle de préférence et les descripteurs $N_1 + N_2$, Acuité et Émergence harmonique (expérience en condition de sonie égalisée).

Son	$N_1 + N_2$ (sones)	Acuité (acum)	Émergence harmonique	Q (2D)	Q (3D)
CRS1 v1 dif	0,26	1,9	4,5	5,2	4,9
CRS2 v1 rep	0,39	1,8	4,0	3,5	2,5
CRS3 v3 rep	0,24	2,0	0,0	5,6	7,6
CRS4 v2 dif	0,19	2,1	2,3	6,3	7,7
GNB1 v3 dif	0,24	2,0	5,1	5,7	5,4
GNB2 v2 rep	0,24	1,8	8,0	5,4	3,6
GNB3 v2 dif	0,20	2,0	13,9	6,2	2,1
GNB4 v2 rep	0,23	2,0	0,3	5,8	7,8
GNB5 v2 dif	0,29	1,8	2,5	4,8	5,1
GNB6 v2 dif	0,27	1,7	2,5	5,0	5,5
GNB7 v4 dif	0,29	1,7	3,8	4,7	4,3
GNB8 v3 rep	0,27	2,0	1,6	5,2	6,3
GNB9 v4 rep	0,30	2,1	5,5	5,0	4,2
CST1 v3 rep	0,29	1,8	1,1	4,9	6,0
CST2 v2 rep	0,33	1,8	0,9	4,3	5,1
CST3 v2 rep	0,29	1,6	4,1	4,5	4,0

TABLE 7.7 – Valeurs, pour chaque son, des 3 descripteurs du timbre, et des prédicteurs de qualité sonore à l'aide respectivement des 2 premiers descripteurs et des 3 descripteurs.

partie la variance de la variable dépendante. À l'exécution de l'algorithme, sont successivement ajoutées au modèle les variables Émergence harmonique (significative à $p < 0,01$) et $N_1 + N_2$ (significative à $p < 0,02$), tandis que la variable Acuité est rejetée. En effet, une régression linéaire multiple avec pour variables explicatives l'Émergence harmonique et $N_1 + N_2$ permet d'obtenir un coefficient de corrélation de Bravais-Pearson quasiment identique à celui obtenu à l'aide des trois descripteurs : $r(ddl = 13) = 0,77$ ($p < 0,01$). Il semble donc que seuls les descripteurs Émergence harmonique et $N_1 + N_2$ soient utiles pour la description des jugements de qualité sonore, bien que l'acuité soit l'attribut du timbre des sons de l'environnement qui ressort le plus dans la littérature (voir section 2.1.2).

7.7 Discussion

Cette étude nous a permis d'identifier l'échelle de préférence des auditeurs parmi les sons de STA considérés, traduisant ainsi la perception de la qualité sonore de ce type de sons. La première conclusion, quoique assez prévisible, à tirer de l'analyse des résultats de l'expérience menée en condition de sonie réelle est qu'un descripteur de sonie tel que celui de Zwicker et Fastl [8] est un parfait prédicteur de l'évaluation par les auditeurs de la qualité sonore des STA. Toutefois, il ne l'est pas plus que le niveau sonore exprimé en dBA. Cela ne signifie pas que les autres paramètres acoustiques mentionnés en section 2.3.3 n'influent pas théoriquement sur la perception des auditeurs. On constate simplement que la sonie est l'attribut auditif prépondérant lorsque l'on s'intéresse à la perception de la qualité sonore. Compte tenu de la grande dynamique des valeurs de sonie pour le corpus utilisé, ce paramètre a probablement masqué l'effet plus subtil d'autres attributs.

C'est en effet ce que confirme, dans une certaine mesure, l'analyse des résultats de l'expérience réalisée en condition de sonie égalisée. Le fait d'avoir fixé ce paramètre a permis, dans le cadre de l'étude du timbre exposée dans le chapitre 6, de faire émerger un ensemble d'attributs auditifs et de descripteurs explicatifs associés confirmant l'hypothèse stipulant que les auditeurs perçoivent des

différences parmi les sons de STA qui ne sont pas expliquées par les variations de sonie. Ainsi, une méthode de régression linéaire multiple a permis de mettre au point un prédicteur de la qualité sonore des STA dans cette condition particulière, sur la base des descripteurs psychoacoustiques expliquant les dimensions du timbre : la sonie dans les deux premières bandes de Bark, l'acuité, et un descripteur d'émergence harmonique.

Toutefois, nous avons observé que, bien que les deux premiers cités représentent les deux plus importantes dimensions dans la perception (d'un point de vue descriptif) des sons de STA – ces deux dimensions sont en effet les deux seules qui sont communes à l'espace 2D et à l'espace 3D (voir section 6.3) –, ils s'avèrent insuffisants lorsqu'il s'agit d'expliquer les jugements de la qualité sonore. La troisième dimension, expliquée par un descripteur d'émergence harmonique, semble être primordiale pour expliquer ces jugements comme l'indiquent les coefficients de corrélations des descripteurs avec l'échelle de qualité sonore mesurée, affichés dans le tableau 7.5. Ainsi, l'ajout de ce descripteur dans le modèle de prédiction a permis d'obtenir un coefficient de corrélation modéré mais significatif : $r(ddl = 12) = 0,78$, tandis que l'acuité ne semble pas avoir d'apport significatif au modèle.

Le caractère modéré de ce coefficient de corrélation doit néanmoins être considéré en gardant à l'esprit la forte variabilité des réponses des auditeurs lors de l'expérience en condition de sonie égalisée. En effet, les grands écarts-types qui peuvent être observés sur la figure 7.7 incitent à considérer les évaluations moyennes, et donc les valeurs des sons sur l'échelle de qualité sonore considérée, avec prudence. L'écoute approfondie des sons et les impressions générales des participants de l'expérience [5] semblent indiquer que les sons du corpus sont peu différenciés lorsqu'il s'agit d'en évaluer le degré d'agrément/désagrément. Ce problème n'est pas apparu dans le cadre de l'étude du timbre car la méthode d'analyse utilisée (MDS – modèle INDSCAL, voir sections 6.3 et A.2 en annexe) prend en compte la variabilité inter-participants afin de déterminer les différentes stratégies individuelles et de fixer les axes des espaces perceptifs. La procédure expérimentale choisie ne semble pas à mettre en cause, car elle a déjà été utilisée avec succès dans la littérature (voir notamment [119]). L'explication est probablement à chercher plutôt dans les sons eux-mêmes, qui semblent globalement très proches les uns des autres, et ne génèrent pas de consensus quant à la qualité sonore perçue.

D'un point de vue pragmatique, les résultats de cette étude ont fourni des outils qui permettent d'estimer la qualité sonore d'un STA. Cette estimation peut être appliquée même si le son en question n'a pas été pris en compte dans les expériences réalisées ici. En effet, tout l'intérêt de ces travaux est de proposer des métriques évaluées directement sur le signal sonore, permettant d'obtenir une estimation de la qualité sonore des STA, et donc du confort acoustique ressenti par les usagers. L'idée générale est que la sonie, ou le niveau sonore exprimé en dBA, évalués sur l'enregistrement audio d'un STA, permettent d'obtenir une première mesure de qualité sonore, qui pourra dans la majorité des cas le situer vis-à-vis des STA qui ont été évalués au cours de cette étude. Dans le cas où la valeur de cette première mesure ne permettrait pas de le distinguer suffisamment d'un autre (ou de plusieurs autres) STA, la qualité sonore de ces deux (ou plus) STA peut être nuancée grâce à la seconde métrique, fondée sur les calculs de la sonie dans les deux premières bandes de Bark, de l'acuité, et de l'émergence harmonique des signaux enregistrés.

5. D'un point de vue informel, plusieurs participants ont fait part à la fin de l'expérience de leur manque de conviction vis-à-vis de leurs évaluations et ont expliqué avoir plusieurs fois changer d'avis, voire de stratégie, avant de valider leurs évaluations en fin de procédure.

Partie III

Influence du contexte d'écoute sur la qualité sonore perçue des STA

Cette partie du document présente les travaux réalisés dans le but d'étudier l'influence de différents paramètres liés au contexte d'écoute des STA. Le **chapitre 8** détaille l'étude de l'influence de l'acoustique des salles sur la qualité sonore des STA. Le **chapitre 9** expose le travail effectué dans le but de confronter la situation classique d'écoute attentive pour l'évaluation de la qualité sonore à une situation d'écoute plus réaliste et représentative des conditions d'usage des STA.

Chapitre 8

Influence de l'acoustique des salles

L'idée directrice du chapitre est de présenter l'étude de l'influence de paramètres liés à l'acoustique des salles en comparant la qualité sonore perçue en différentes conditions de réverbération et de spatialisation du son. Dans cette optique, la section 8.1 présente tout d'abord la problématique générale de l'étude de l'influence de l'acoustique des salles. Cette problématique implique l'utilisation d'un outil particulier de spatialisation du son, présenté en section 8.2 : l'*auralisation*. La section 8.3 présente le protocole expérimental d'évaluation comparée utilisé dans le cadre de cette étude. La section 8.4 présente l'analyse des résultats de cette expérience. Enfin, la section 8.5 expose la discussion que soulèvent les résultats de cette étude.

8.1 Problématique

Les travaux réalisés dans le cadre des études présentées dans les chapitres 5, 6 et 7 ont été réalisés à l'aide de sons correspondant à une condition bien particulière : ils ont été enregistrés dans une salle *(semi-)anéchoïque*, dont le but est de supprimer une majeure partie du champ réverbéré. En effet, nous souhaitions alors décrire la perception des sources sonores que sont les STA, indépendamment de tout facteur environnemental. Toutefois, dans des conditions réalistes d'usage de STA, ceux-ci sont installés la plupart du temps dans des bureaux, mais dans tous les cas dans des espaces clos. Cette situation particulière, contrairement à la condition anéchoïque d'enregistrement, génère de la *réverbération* qui modifie le son perçu par les auditeurs d'une manière non-négligeable.

En effet, le son est une onde qui, par conséquent, interagit avec son environnement, et, à l'exception d'une situation bien particulière (*champ libre*), ne se propage pas à l'infini. Lorsqu'il rencontre un obstacle, une partie de son énergie est absorbée par celui-ci (voire transmise à travers lui), et la partie restante est réfléchie. La proportion d'énergie qui est réfléchie dépend notamment de la bande de fréquence considérée et de la nature de l'obstacle (forme, rugosité de la surface, matériau, ...). Ainsi, entre le moment où le son est émis (par le STA) et le moment où il est arrive « à destination » (jusqu'à l'auditeur), il a subi un nombre plus ou moins grand de réflexions qui se traduisent par la réverbération modifiant de manière certaine le son entre ces deux instants.

Ainsi, il n'est pas garanti que les résultats des chapitres précédents, obtenus avec des sons correspondant à des enregistrements anéchoïques, soient les mêmes que ceux qui auraient été obtenus si les sons avaient été enregistrés dans une salle plus représentative des conditions (non-anéchoïques) dans lesquelles sont habituellement installés les STA. Il est donc indispensable de comparer les résultats d'évaluation de la qualité sonore obtenus dans le chapitre 7 avec ceux qui pourraient être obtenus

dans des conditions de réverbération réelles ou réalistes. Il est possible de simuler l'effet d'une salle particulière sur le son tel qu'il serait reçu à une position plus ou moins éloignée de sa source (le STA). Cette simulation est réalisée à l'aide de l'auralisation (voir section 8.2). Cette outil permet de générer, à partir d'enregistrements anéchoïques, des sons incorporant l'aspect temporel (« persistance » du son) de la réverbération, et, dans le cas de l'outil utilisé, son aspect spatial (lié à la localisation de la source notamment), dans la salle considérée.

Dans le cadre des expériences perceptives réalisées ici, l'utilisation d'un outil d'auralisation présente un avantage majeur, qui est de pouvoir comparer plusieurs (deux, en l'occurrence) salles différentes. Ceci aurait été difficile si l'on avait souhaité effectuer cette comparaison dans des salles réelles. En effet, les sons auralisés par les différents types de salles peuvent ici être directement comparés par les auditeurs, sans avoir à « changer » de salle (ce qui aurait compliqué les comparaisons à cause de notre mémoire auditive peu efficace).

Le principe de l'étude présentée dans ce chapitre est donc, dans un premier temps, de reproduire le protocole expérimental de mesure perceptive de la qualité sonore, utilisé dans le chapitre 7, avec un corpus composé de sons auralisés, ainsi enrichis d'un champ réverbéré. Dans un second temps, il importe de comparer l'échelle de qualité sonore ainsi obtenue avec celle décrite en section 7.5 [1], afin de déterminer si les jugements des auditeurs sont significativement influencés par la présence de réverbération.

8.2 Présentation de l'outil d'auralisation et des simulations utilisées

Un système d'auralisation est un système dont le but est de reproduire la dimension temporelle de la réverbération d'une salle particulière (voir [157]). Dans certains cas, il permet également de reproduire la dimension spatiale de la réverbération, lorsque le mode de reproduction utilisé est un casque en écoute dichotique (signal différent dans les deux oreilles) ou un système de haut-parleurs dans un studio immersif. Cette reproduction passe par l'estimation de la réponse impulsionnelle de la salle considérée – *Room Impulse Response (RIR)*, qui est ensuite convoluée avec le signal anéchoïque. La réponse impulsionnelle peut être obtenue par mesure directe dans la salle dont on souhaite reproduire la réverbération, ou par simulation. Ainsi, dans le cas d'une simulation, il est possible de générer, à partir des caractéristiques géométriques de la salle et des matériaux des éléments composants ses parois, un échogramme, c'est-à-dire la suite d'impulsions que capterait un microphone placé dans la salle. Chacune de ces impulsions correspond à l'impulsion initialement émise par la source et ayant subi un plus ou moins grand nombre de réflexions sur les parois de la pièce.

On distingue souvent les premières impulsions, correspondant à l'impulsion initiale n'ayant subi que peu de réflexions, des impulsions suivantes provenant d'un plus grand nombre de réflexions successives. Les premières, qui fournissent à elles-seules une grande partie de l'information spatiale, notamment les indices de localisation de la source, peuvent être modélisées analytiquement (sauf dans le cas où les parois réfléchissent de manière diffuse). Les dernières ont tout de même une importance perceptive non-négligeable, mais sont de plus faible amplitude et beaucoup plus rapprochées dans le temps – donc beaucoup plus nombreuses par unité de temps. Elles sont par conséquent beaucoup plus difficiles à décrire de manière déterministe. Elles sont plutôt obtenues à partir de modèles stochastiques, ou sont même souvent exclues de la modélisation.

1. L'échelle considérée est donc celle obtenue en condition de sonie égalisée. En effet, l'influence du percept de sonie est de nouveau mise de côté, afin qu'elle ne masque pas l'éventuel effet de l'auralisation. Cette étude s'accompagne par conséquent également d'une *égalisation en sonie* (voir section 8.3.3).

8.2.1 Système d'auralisation *AURALIAS*

Le système utilisé pour l'auralisation est celui développé dans le cadre du projet *AURALIAS* [12] et porte le même nom (voir aussi Bos et Embrechts [30, 31] à ce sujet). Ce système utilise le programme *Salrev* (voir Billon et Embrechts [22] pour plus de détails) pour simuler la réponse impulsionnelle d'une salle. Le programme *Salrev* incorpore deux types de modélisation différents pour pouvoir générer aussi bien les premières réflexions que les réflexions d'ordre supérieur. La première modélisation utilise la méthode dite des *sources-images*, qui consiste simplement à évaluer les différents trajets possibles entre la source et le récepteur, comme on le ferait pour un rayon optique se réfléchissant sur des miroirs. La seconde utilise la méthode de *tir de rayons* qui consiste à simuler un très grand nombre de rayons (c'est-à-dire d'ondes acoustiques directionnelles) partant de la source dans des directions aléatoires, et à observer les rayons passant au travers d'une sphère autour du point de réception (voir [57, 58] pour plus de détails sur l'algorithme de tracé de rayons utilisé).

Un exemple tiré de Bos et Embrechts [30, 31] d'échogrammes directionnels ainsi obtenus est montré en figure 8.1. Un échogramme directionnel correspond, pour une bande d'octave donnée, à la suite d'ondes acoustiques reçues provenant d'une direction particulière. Ces échogrammes directionnels sont transformés en réponses impulsionnelles directives, qui permettent alors, au travers d'une convolution temporelle avec le signal anéchoïque à auraliser, de générer un son auralisé, qui prend la forme d'un signal stéréophonique dans le cas de la reproduction au casque, ou d'un signal à 6 canaux pour la reproduction en studio immersif. Dans le cas de la reproduction au casque, l'auralisation comprend un filtre *HRTF – Head Related Transfer Function*, fonction de transfert d'une tête humaine – censé simulé l'effet du torse, de la tête et du pavillon auditif d'un individu sur le signal sonore (on parle alors plutôt de *signal binaural*). Dans le cas de la reproduction sur haut-parleurs, le studio immersif est une pièce traitée acoustiquement pour l'auralisation et dans laquelle 6 haut-parleurs ont été disposés à des positions bien précises dans un plan horizontal à une hauteur d'environ 1,5 m. Un plan de ce studio immersif est montré en figure 8.2 (tiré de Bos et Embrechts [30]) La reproduction spatiale est donc une reproduction 2D. À l'avenir 2 haut-parleurs seront ajoutés en hauteur de sorte à avoir une reproduction spatiale en 3D.

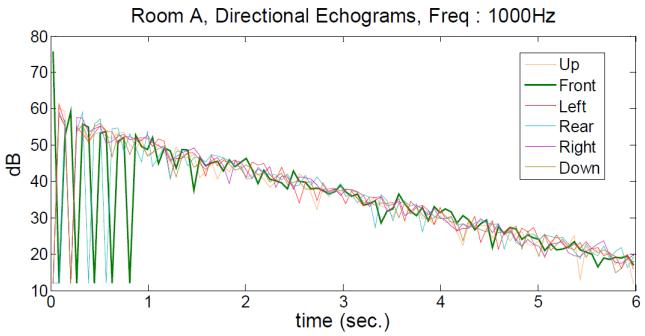


FIGURE 8.1 – Exemple d'échogrammes directionnels, tiré de Bos et Embrechts [30, 31], provenant des directions « haut », « devant », « gauche », « derrière », « droite » et « bas », respectivement.

8.2.2 Simulations effectuées

Deux simulations ont été effectuées avec le programme Salrev, dans le but d'effectuer les expériences dans deux types différents de salle virtuelle. La première simule un bureau individuel de 4 m par 4 m et de 3 m de haut, la seconde simule une salle de réunion de 20 m par 20 m et 3 m de hauteur

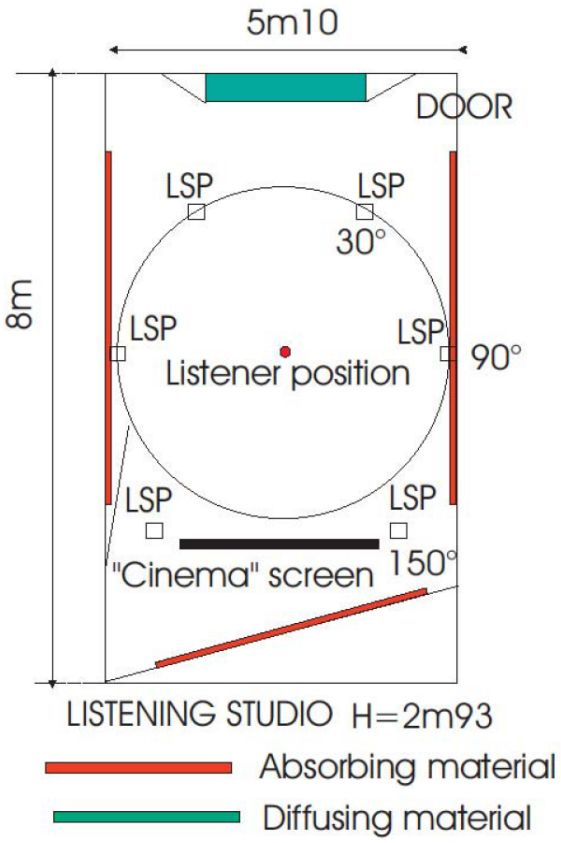


FIGURE 8.2 – Plan du studio immersif, tiré de Bos et Embrechts [30]. « LSP » signifie *Loudspeaker Position* – position du haut-parleur.

également. Une table est placée à 0,9 m de hauteur dans chaque pièce (2 x 1 m pour le bureau individuel, 13 x 13 m pour la salle de réunion). Les deux salles présentent également chacune un mur avec une grande baie vitrée de 2 m de haut sur toute la longueur du mur. Les matériaux utilisés dans les deux cas sont :

- plaques de plâtre peintes pour les murs
- double-vitrage pour la baie vitrée
- absorbant fibreux pour le plafond
- moquette sur béton pour le sol
- bois pour la table

Les figures 8.3 et 8.4 affichent une perspective en 3 dimensions des modèles géométriques des 2 simulations de salle, sur laquelle la source est repérée par l'étoile rouge, et le récepteur, par l'étoile bleue (la baie vitrée correspond à la partie en bleue, et la table est identifiée en beige). La source est modélisée dans le programme comme une source ponctuelle.

Le tableau 8.1 décrit les propriétés acoustiques des matériaux utilisés dans les simulations en termes de coefficient d'absorption α [115, 53]. Les tableaux 8.2 et 8.3 affichent les caractéristiques acoustiques des simulations du bureau individuel et de la salle de réunion, respectivement (voir section 3.1.2 pour la définition et la signification de ces paramètres). Les valeurs de RASTI pour ces deux simulations sont respectivement 0,76 et 0,81, qui, dans les deux cas, correspondent à une « excellente » intelligibilité (voir tableau 3.1 en section 3.2.2).

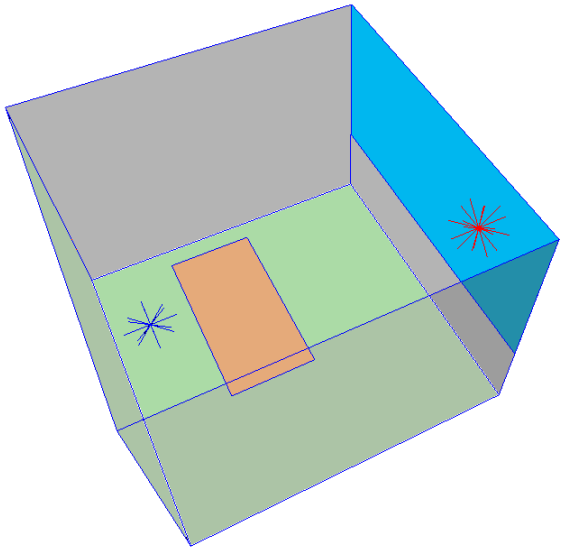


FIGURE 8.3 – Perspective en 3D de la simulation de bureau individuel. Les étoiles rouges et bleus représentent respectivement la position de la source sonore et celle du récepteur.

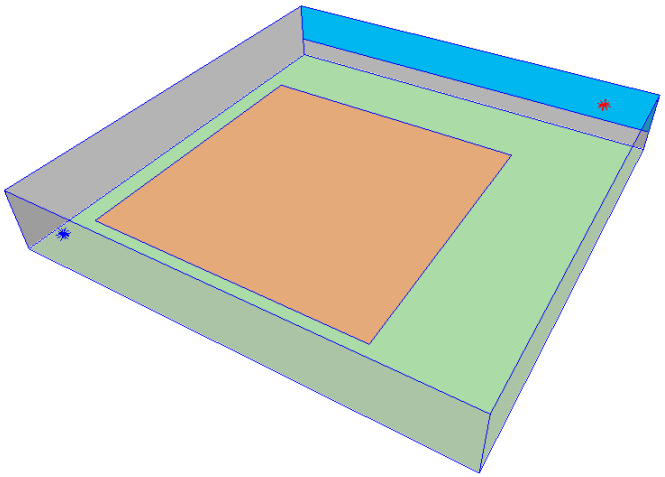


FIGURE 8.4 – Perspective en 3D de la simulation de salle de réunion. Les étoiles rouges et bleus représentent respectivement la position de la source sonore et celle du récepteur.

8.3 Protocole expérimental d'évaluation comparée avec auralisation

8.3.1 Principe

Le principe de cette expérience est globalement de reproduire la procédure expérimentale établie pour mesurer les préférences des auditeurs (sections 7.2 et 7.4), dans différentes conditions d'auralisation. Cette auralisation est réalisée grâce à l'auralisateur introduit en section 8.2 utilisant un système de reproduction spatiale 2D du son à l'aide de six haut-parleurs, diffusant chacun un signal monophonique modifié afin de reproduire le champ réverbéré correspondant aux simulations décrites en section 8.2. Nous souhaitons également comparer l'effet de l'auralisation lorsqu'elle est réalisée au casque. L'auralisateur permet également de générer à partir d'un enregistrement anéchoïque, un signal stéréophonique reproduisant le champ réverbéré correspondant aux simulations, que l'on peut

Élément	62,5 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Murs	0,4	0,35	0,22	0,15	0,1	0,07	0,07	0,07
Baie vitrée	0,17	0,14	0,07	0,04	0,03	0,02	0,02	0,02
Plafond	0,3	0,4	0,5	0,55	0,65	0,75	0,70	0,65
Sol	0,09	0,09	0,08	0,21	0,26	0,27	0,37	0,45
Table	0,03	0,04	0,07	0,09	0,10	0,08	0,06	0,04

TABLE 8.1 – Coefficients d'absorption α des matériaux utilisés dans les simulations.

Métrique	62,5 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
$RT(s)$	0,27	0,28	0,37	0,44	0,55	0,68	0,70	0,55
$EDT(s)$	0,30	0,30	0,38	0,42	0,45	0,50	0,49	0,40
$C_{80}(dB)$	16,5	16	12,3	10,9	9,9	9	9,1	11,1
$D_{50}(\%)$	94	93,5	88,2	85,7	83,5	81,4	81,9	86,8
$SBT_s(ms)$	18,7	19,3	25,1	27,7	30,2	32,8	32,2	26,1

TABLE 8.2 – Caractéristiques acoustiques de la simulation de bureau individuel.

Métrique	62,5 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
$RT(s)$	0,78	0,84	1,20	1,56	1,93	2,21	1,93	1,04
$EDT(s)$	0,30	0,26	0,25	0,24	0,23	0,23	0,19	0,17
$C_{80}(dB)$	11,8	12,2	11,3	10,9	10,4	10	10,5	13,3
$D_{50}(\%)$	91,4	92,6	91,7	91,7	91,3	90,8	91,7	95,4
$SBT_s(ms)$	20	18,6	20,9	21,6	23,3	24,8	21,8	13,1

TABLE 8.3 – Caractéristiques acoustiques de la simulation de salle de réunion.

alors diffuser au casque.

Il est par ailleurs nécessaire d'avoir un point de référence auquel comparer les résultats obtenus avec auralisation. Ainsi, la mesure des préférences décrite en section 7.5, réalisée au casque en écoute diotique avec des enregistrements anéchoïques, fera office de condition témoin, sans auralisation ou effet de salle quelconques.

8.3.2 Stimuli

Bien que le corpus sonore établi grâce à l'expérience de catégorisation libre décrite au chapitre 5 soit constitué de 16 sons, deux simulations de salles – donc deux auralisations – sont comparées dans cette expérience. Or, il est nécessaire, afin de ne pas influencer les résultats des expériences, et pour en conserver la faisabilité pratique, de conserver le même nombre de stimuli présentés à chaque participant. En conséquence, seule une moitié du corpus établi en section 5.3 sera utilisée dans ces conditions d'expérience.

De nouveau la sélection de ces 8 sons a été faite de sorte à conserver, d'une part la représentativité des familles identifiées au chapitre 5, et d'autre part celle des différents types de systèmes enregistrés (carrossés, gainables , et cassettes, voir section 4.2). La figure 8.5 présente la sélection des 8 sons (flèches rouges), vis-à-vis de la sélection initiale de 16 sons et de la représentation en dendrogramme des familles de sons exposées au cours du chapitre 5. En voici l'inventaire :

1. **CRS1 v1 dif**
4. **CRS4 v2 dif**

7. **GNB3 v2 dif**
8. **GNB4 v2 rep**
10. **GNB6 v2 dif**
13. **GNB9 v4 rep**

14. **CST1 v3 rep**
15. **CST2 v2 rep**

Ces 8 enregistrements, monophoniques à l'origine, ont été auralisés à l'aide des deux simulations décrites en section 8.2 portent donc le total de stimuli évalués en condition auralisée à 16 sons.

8.3.3 Égalisation en sonie

Comme lors de plusieurs des expériences précédemment exposées, il est nécessaire que les jugements des auditeurs ne soient pas influencés par les variations d'intensité sonore perçue. Par conséquent, les sons doivent être égalisés en sonie. Compte tenu de la nature multicanale des stimuli, aussi bien pour la condition au casque que pour celle en studio, il est impossible de réaliser cette opération de manière automatique à l'aide de la méthode décrite en section 5.2.1. Il est par conséquent nécessaire de réaliser cette égalisation en sonie de manière expérimentale. Le principe est de choisir un son du corpus dont la sonie va servir de valeur référence, et de présenter successivement au participant chaque paire constituée de ce son et de l'un des autres du corpus à égaliser. La tâche du participant consiste alors, pour chaque paire son de référence / son à égaliser, à régler, à l'aide d'un curseur, le niveau sonore du second de sorte à ce qu'il soit perçu comme aussi « fort » que le premier.

En général, peu de participants (moins de 10) sont nécessaires pour cette procédure, car un relatif consensus est le plus souvent rapidement atteint. De plus, pour des corpus comprenant entre 10 et 20 éléments, la durée de cette procédure dépasse rarement les 10 minutes. Enfin, puisque l'on cherche à mesurer un paramètre audiologique, peu sujet à une interprétation subjective, la connaissance des sons ou du travail effectué (caractère *naïf* / *expert* du participant) n'est pas un paramètre important dans le choix des participants. Pour toutes ces raisons, elle a été réalisée à l'aide de collègues volontaires du laboratoire (9 pour l'auralisation au casque, 7 pour l'auralisation sur haut-parleurs).

Une interface graphique spécifique a été programmée en LabVIEW 7.0 (figure 8.6) pour l'expérience d'égalisation en sonie. La procédure est la suivante :

- Les paires de sons – son de référence / son à égaliser – sont jouées une par une.
- Pour chacune d'elles, le participant doit modifier la position du curseur pour régler le « niveau » du son à égaliser de sorte à ce qu'il soit perçu comme identique à celui du son de référence.
- Le participant a la possibilité de rejouer chaque paire de sons à tout moment tant qu'il n'a pas validé son réglage (une réécoute au moins est imposée après déplacement du curseur avant de pouvoir valider).
- L'expérience se termine lorsque tous les sons ont été égalisés.

Enfin, sur cette interface, le curseur est placé initialement au milieu de l'échelle, lorsque le facteur multiplicatif su signal audio est égal à 1. L'échelle étant linéaire, on en déduit que l'échelle des facteurs utilisables s'étend de 0 à 2.

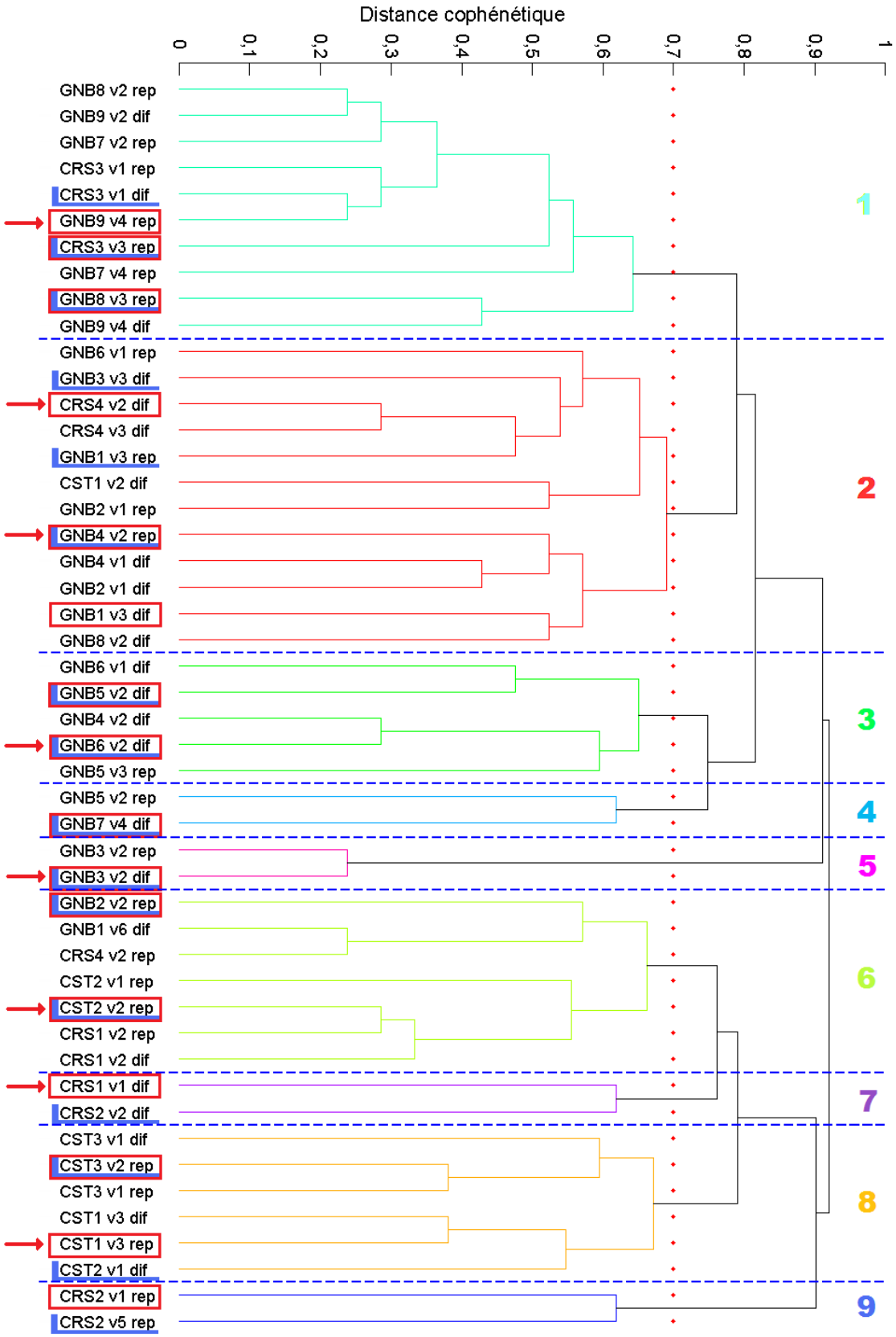


FIGURE 8.5 – Dendrogramme issu de l'expérience de catégorisation libre, avec les catégories identifiées (séparées par les lignes en pointillés bleus), les 16 *exemplaires représentatifs* (dans les cadres rouges), et, parmi ceux-là, les 8 sons sélectionnés pour l'utilisation de l'auralisation (flèches rouges).

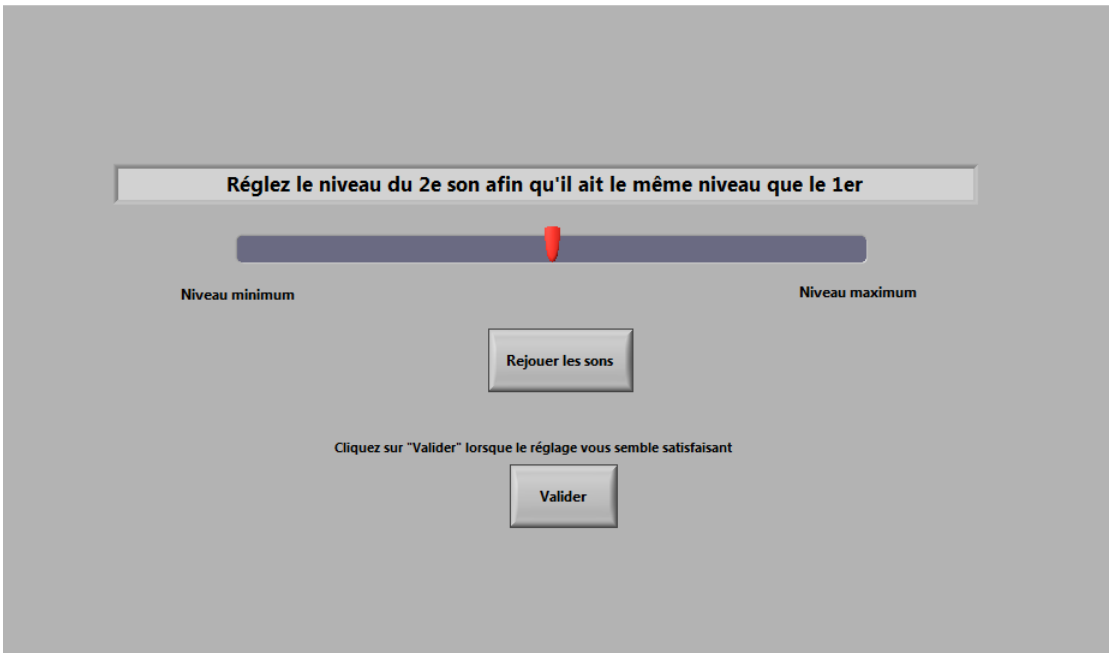


FIGURE 8.6 – Interface graphique pour l'égalisation en sonie.

Il est à noter que l'égalisation en sonie à été réalisée séparément pour les stimuli stéréophoniques et pour les stimuli à 6 canaux, puisque ces stimuli sont utilisés dans deux expériences distinctes. Les résultats de cette expérience d'égalisation en sonie, sous forme de facteurs multiplicatifs moyens et d'écarts-types, sont affichés en figure 8.7 pour la condition au casque, et en figure 8.8 pour la condition en studio.

Il convient de préciser un point important qui diverge pour la réalisation de l'expérience dans ces 2 conditions :

- Pour la condition au casque, le son de référence, dont la sonie sert de point de comparaison pour les autres sons, était la version anéchoïque du son **CST1 v3 rep**. En effet, les sons étant diffusés au casque, il a semblé judicieux qu'ils aient sensiblement la même sonie que dans l'expérience de référence en condition anéchoïque (voir section 7.4), également réalisée au casque. Il importe toutefois de noter que l'opération d'auralisation a fortement modifié la sonie des sons. Un ajustement manuel préalable du niveau sonore a donc été effectué, afin de rendre utilisable l'échelle de l'interface, compte tenu de la limitation des facteurs multiplicatifs entre 0 et 2. Lors de cet ajustement, un facteur commun a été appliqué aux sons résultant de l'auralisation pour la simulation de bureau individuel, et un autre a été appliqué à ceux résultant de l'auralisation pour celle de la salle de réunion.
- Pour la condition en studio, la comparaison à un son anéchoïque n'est évidemment pas judicieuse et difficilement réalisable de manière cohérente. Par conséquent, la référence choisie est le son **CST1 v3 rep** auralisé avec la simulation du bureau. Pour cette raison, les niveaux émis sont tels qu'obtenus après auralisation, après toutefois un ajustement général – commun à tous les sons – afin de rendre les sons aisément audibles, tout en conservant un forme de validité écologique.

Ces éléments permettent d'expliquer que dans les cas de l'auralisation au casque, les égalisations sont similaires entre les deux simulations, tandis que dans le cas de l'auralisation en studio, il existe une différence manifeste entre les facteurs obtenus pour les 2 simulations. En effet, l'ajustement préa-

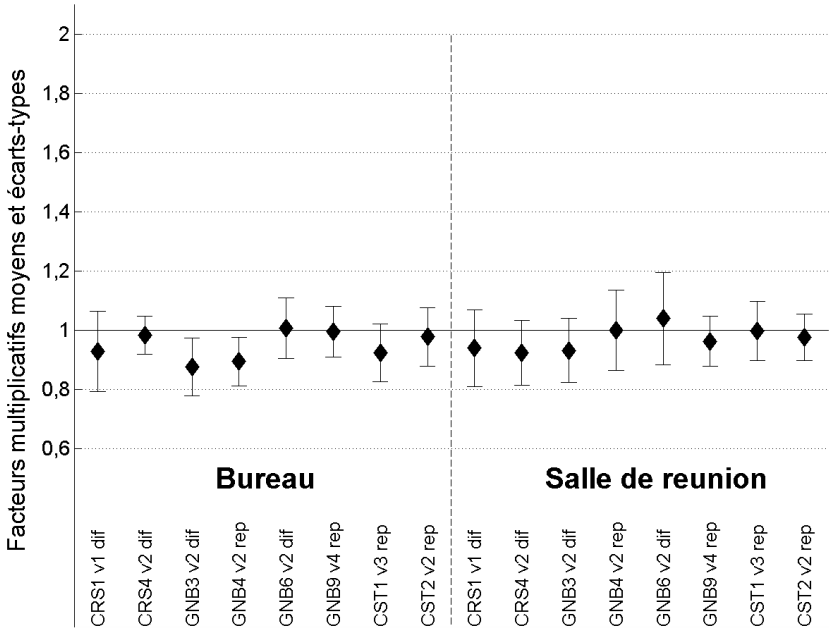


FIGURE 8.7 – Résultats de l'égalisation en sonie des sons auralisés au casque.

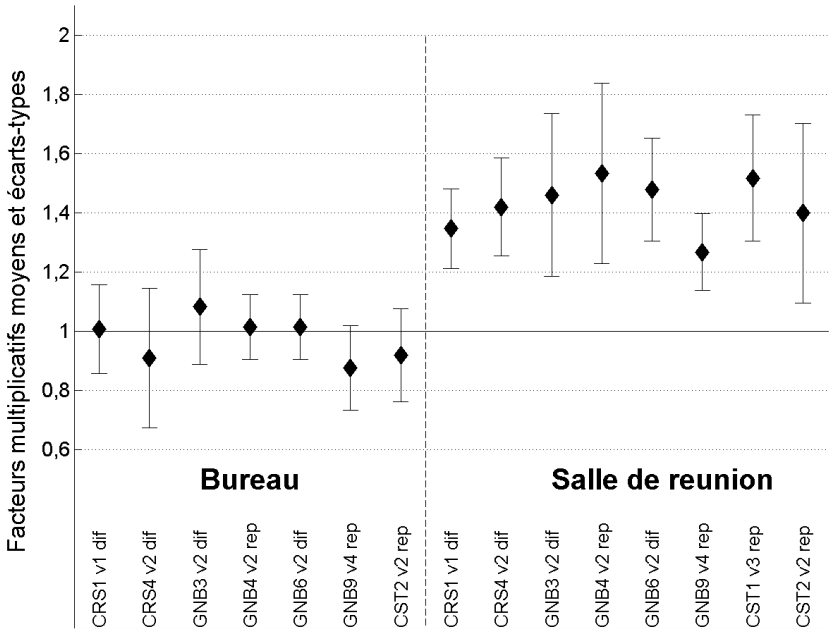


FIGURE 8.8 – Résultats de l'égalisation en sonie des sons auralisés sur haut-parleurs.

lable, qui diffère entre les deux simulations, dans la condition au casque, implique que les sonies sont initialement plus proches du son de référence que dans la condition en studio.

Par ailleurs, on observe également que pour une même simulation – aussi bien pour la condition au casque que pour celle en studio – que les facteurs multiplicatifs moyens sont tous très proches. Ce phénomène trouve son explication dans les sons utilisés pour cette expérience : les simulations d'auralisation ont été appliquées sur les sons issus de l'expérience de catégorisation (voir section 5.3), qui avait déjà subi une égalisation en sonie automatique. Or, si l'auralisation modifie bien évidemment la sonie des sons dans l'absolu – notamment à cause de la distance source / récepteur dans les simulations –, d'un point de vue relatif, l'auralisation a en pratique peu d'influence sur la sonie des sons pour une même simulation. En conséquence, il est parfaitement logique d'observer que les différences entre les facteurs multiplicatifs moyens pour une même simulation soient faibles en pratique, et que les sons après auralisation restent très proches en sonie.

Deux corpus distincts sont donc finalement utilisés dans les deux conditions d'auralisation :

– Auralisation au casque : 16 sons stéréophoniques, égalisés en sonie, et d'une durée de 4 secondes.
– Auralisation en studio (sur haut-parleurs) : 16 sons à 6 canaux, égalisés en sonie, et d'une durée de 4 secondes.

8.3.4 Matériel

Les deux expériences (au casque et en studio) ont été réalisées à l'aide de la même interface graphique programmée en LabVIEW 2010 (identique à la figure 7.4), assurant la lecture des sons sur l'interface audio et l'enregistrement des réponses des participants. Les sons ont été diffusés :

– pour la condition d'auralisation au casque, par une interface RME Fireface 800 au travers d'un casque ouvert Sennheiser HD650 en écoute dichotique (signaux différents dans les deux oreilles, les sons auralisés au casque étant stéréophoniques) dans une cabine audiométrique IAC à double paroi ;
– pour la condition d'auralisation en studio, par une interface RME Fireface 400 dans le studio immersif décrit en section 8.2.1 au travers de 6 haut-parleurs FAR XM6D.

8.3.5 Participants

Dix-neuf participants se sont portés volontaires pour l'expérience dans chacune des deux conditions d'auralisation (participants différents pour chacune). Pour la condition au casque, 11 hommes et 8 femmes (entre 23 et 28 ans) de l'Université de La Rochelle ont participé, tandis que 15 hommes et 4 femmes (entre 21 et 25 ans) de l'Université de Liège ont participé pour la condition en studio. Aucun d'entre eux n'a fait mention d'un problème majeur d'audition, et aucun n'a participé à une autre expérience réalisée dans le cadre de cette thèse.

8.3.6 Procédure

La procédure expérimentale utilisée ici est la même procédure d'*évaluation comparée* que celle exposée en section 7.4. En résumé, après avoir reçu une consigne écrite de l'expérience (voir section E.3 en annexe), les participants doivent évaluer chaque son du corpus sur une échelle allant de

0 à 10 le degré de désagrément (du plus désagréable au plus agréable), tout en pouvant réécouter et modifier leurs évaluations tout au long de l'expérience (voir figure 7.4).

La durée moyenne de la procédure a été de 14 minutes dans le cas de l'expérience sur casque, et de 15 minutes dans le cas de l'expérience sur haut-parleurs.

8.4 Résultats et Analyse

Les résultats bruts de cette expérience dans chacune des conditions prennent la forme d'un jeu d'évaluations de tous les sons pour chaque participant. Par conséquent, comme c'était le cas en sections 7.3 et 7.5, les évaluations moyennes et les écart-types associés ont été évalués pour chaque son, afin de quantifier la qualité sonore mesurée. Les figures 8.9 et 8.10 affichent ces éléments respectivement pour l'auralisation au casque et pour l'auralisation sur haut-parleurs.

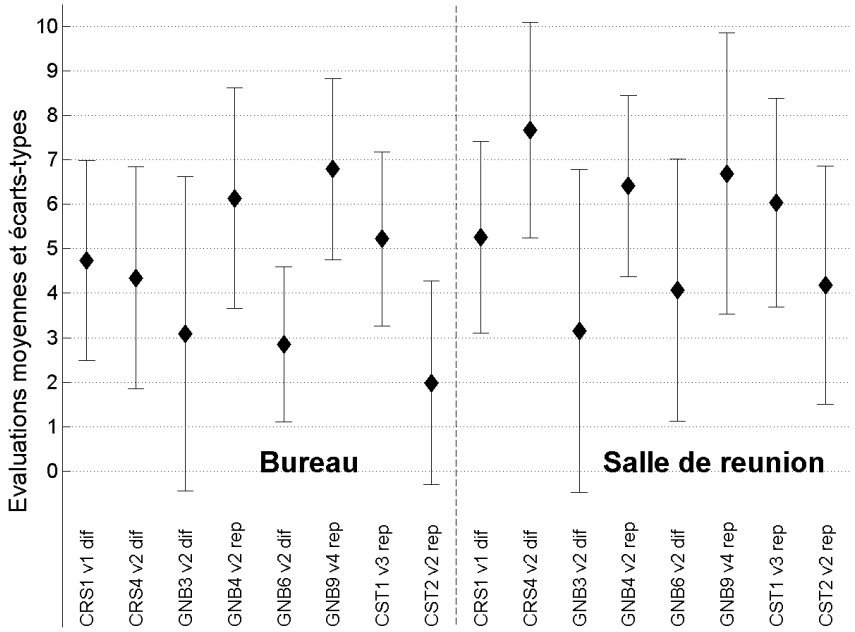


FIGURE 8.9 – Évaluations moyennes et écarts-types du désagrément pour l'expérience d'évaluation comparée en condition d'auralisation au casque. L'échelle va de '0' pour les plus désagréables à '10' pour les plus agréables.

De nouveau, on peut observer de fortes variations parmi les participants, compte tenu des écarts-types importants, que l'auralisation soit faite au casque ou sur haut-parleurs. Dans le cas de l'auralisation au casque, la moyenne des écarts-types et_m vaut 2,51, et l'écart-type des valeurs moyennes de l'échelle m_{et}, traduisant la dynamique de celle-ci, vaut 1,64. Le coefficient de concordance de Kendall et le coefficient de corrélation de rang de Spearman, qui traduisent le degré de concordance de l'ordonnancement des sons sur l'échelle des évaluations par les différent participants (voir sections 7.3, et C.1 en annexe, pour plus de détails) valent respectivement $W = 0.31$ et $\bar{r}_s = 0,27$, et sont par conséquent très faibles, et éloignés de la valeur idéale '1'. Pour l'auralisation sur haut-parleurs, ces éléments statistiques semblent légèrement meilleurs mais sans être pour autant satisfaisants. Les coefficients et_m, m_{et}, W et $\bar{r}_s$ valent respectivement 2,34, 2,00, 0,42 et 0,38. Ils sont résumés dans le tableau 8.4.

Une analyse de cluster a été appliqué aux panels de participants de ces deux expériences (voir setion C.2 en annexe pour plus de détails) afin d'explorer les possibles causes de cette forte variabilité des réponses. Les dendrogrammes correspondant aux deux conditions d'auralisation (casque

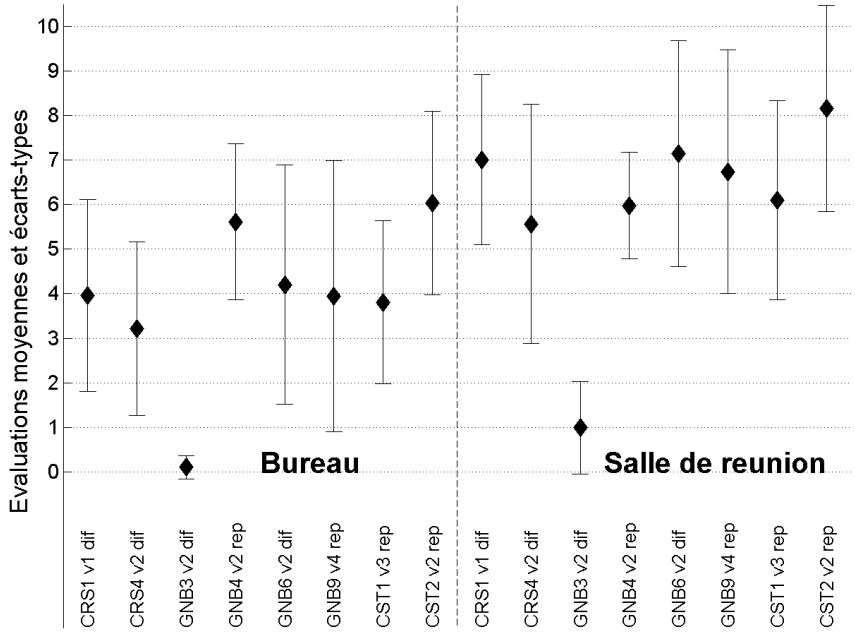


FIGURE 8.10 – Évaluations moyennes et écarts-types du désagrément pour l'expérience d'évaluation comparée en condition d'auralisation sur haut-parleurs. L'échelle va de '0' pour les plus désagréables à '10' pour les plus agréables.

	casque	haut-parleurs
et_m	2,51	2,02
m_{et}	1,64	2,21
W	0,31	0,53
$\bar{r}_s$	0,27	0,50

TABLE 8.4 – Statistiques des expériences d'évaluation comparée en condition auralisée au casque et sur haut-parleurs (écart-type moyen entre les participants et_m, écart-type entre les évaluations moyennes des sons m_{et}, coefficient de concordance de Kendall W et coefficient de corrélation de rang de Spearman $\bar{r}_s$).

et haut-parleurs) sont exposés en figures 8.11 [2] et 8.12. Pour rappel, celui obtenu pour la condition référence (enregistrements anéchoïques) est également affiché en figure 8.13. Il apparaît clairement sur ces 3 figures que des groupes restreints de participants ont des résultats divergeant sensiblement de ceux des autres participants.

Pour la condition auralisée au casque, le cluster formé par les 3 participants les plus à droite du dendrogramme – index 9, 10 et 15 – sont reliés aux autres avec une distance cophénétique aux alentours de 0,55 (ce qui correspondrait à un coefficient de corrélation de -0,1). Ces 3 participants, qui semblent percevoir les sons avec une sensibilité différente des autres [3], ont par conséquent été retirés du panel de participants.

Pour la condition auralisée sur haut-parleurs, le même phénomène peut être observé avec 2 participants, dans une moindre mesure toutefois. En effet, les participants 12 et 16 sont reliés aux autres

2. Sur ce dendrogramme (figure 8.11), seuls 18 participants sont affichés car les résultats d'un des participants se sont avérés incohérents avec les instructions données. Celui-ci a donc été directement retiré du panel.

3. Cela ne signifie pas que leurs résultats sont dénués d'informations, mais simplement qu'ils représentent une classe minoritaire de participants qui jugent le désagrément des sons d'une manière différente. En revanche, leur effectif réduit empêche toute étude approfondie de leurs résultats et des raisons expliquant ce phénomène.

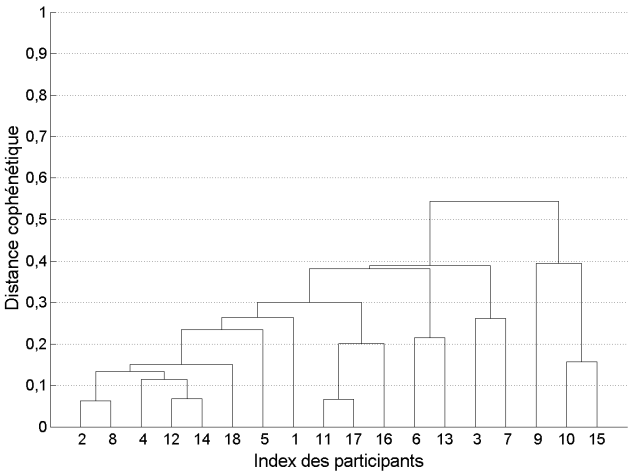


FIGURE 8.11 – Dendrogramme issu des corrélations inter-participant pour l'expérience en condition auralisée et au casque.

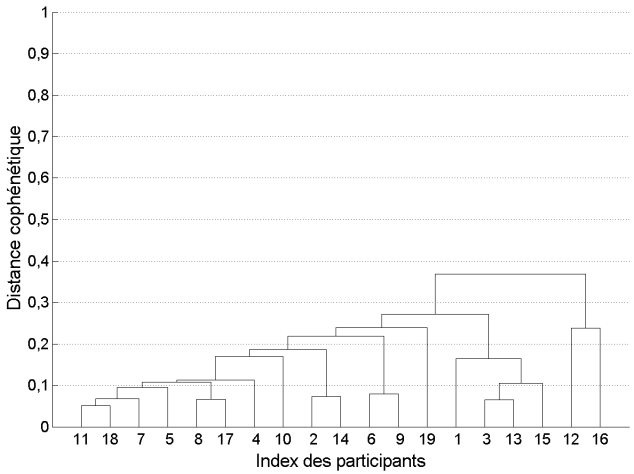


FIGURE 8.12 – Dendrogramme issu des corrélations inter-participant pour l'expérience en condition auralisée et sur haut-parleurs.

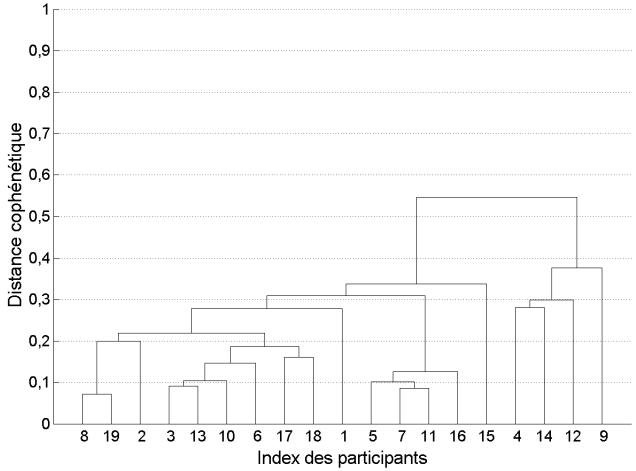


FIGURE 8.13 – Dendrogramme issu des corrélations inter-participant pour l'expérience en condition anéchoïque et au casque.

avec une distance cophénétique au-delà de 0,35 (ce qui correspondrait à un coefficient de corrélation de 0,3). De même ces 2 participants ont été retirés du panel.

Pour rappel, ce phénomène apparaissait déjà dans la condition référence, où 4 participants (index 4, 9, 12 et 14) étaient reliés aux autres avec une distance cophénétique aux alentours de 0,55. Ils avaient eux aussi été retirés du panel de participants.

Après avoir mis de côté les résultats de tous les participants qui semblaient incompatibles avec ceux de la majorité des autres participants, il est maintenant possible d'observer les évaluations moyennées sur ces panels de participants. Les figures 8.14, 8.15 et 8.16 montrent les résultats de l'expérience respectivement pour les conditions anéchoïque, auralisée au casque et auralisée sur haut-parleurs, sous la forme d'évaluation moyenne et écart-type pour chaque son. Afin de faciliter la comparaison des graphes, les stimuli sont présentés selon l'ordre croissant des évaluations moyennes pour l'expérience en condition anéchoïque. Les éléments statistiques inter-participants sont ainsi améliorés : et_m, m_{et}, W et $\bar{r}_s$ passent à respectivement 2,34, 2,00, 0,42 et 0,38 pour la condition d'auralisation au casque, et à 1,87, 2,32, 0,59 et 0,57 pour la condition d'auralisation sur haut-parleurs. Sans être entièrement satisfaisants, ces éléments statistiques inter-participants confirment une nette amélioration de la représentativité des échelles de qualité sonore obtenues. Ils sont résumés dans le tableau 8.5.

	casque	haut-parleurs
et_m	2,34	1,87
m_{et}	2,00	2,32
W	0,42	0,59
$\bar{r}_s$	0,38	0,57

TABLE 8.5 – Statistiques inter-participants des expériences d'évaluation comparée en condition d'auralisation au casque et sur haut-parleurs, après retrait des résultats des *outliers* (écart-type moyen entre les participants et_m, écart-type entre les évaluations moyennes des sons m_{et}, coefficient de concordance de Kendall W et coefficient de corrélation de rang de Spearman $\bar{r}_s$).

8.5 Discussion

L'étude présentée dans ce chapitre a permis de comparer une échelle de préférence obtenue à l'aide d'enregistrements anéchoïques, avec celle obtenue dans des conditions plus réalistes, c'est-à-dire en prenant en compte la réverbération de la salle. Nous avons donc reproduit le protocole expérimental utilisé dans le chapitre 7, c'est-à-dire dans une *condition anéchoïque* avec une diffusion diotique au casque de sons monophoniques, dans deux *conditions auralisées – au casque* et *sur haut-parleurs*, voir section 8.3 – permettant de recréer les aspects temporels et spatiaux du son liés à la réverbération.

La première observation que l'on peut faire à partir des résultats présentés en section 8.4 est que les évaluations des participants présentent une forte variabilité, même après retrait des *outliers*. Ce phénomène est confirmé par les statistiques inter-participants du tableau 8.5. Toutefois, cette forte variabilité est également présente dans le cas de la condition anéchoïque. En effet, les statistiques du tableau 7.4, dans le chapitre précédent, caractérisant les résultats dans cette condition de référence, sont relativement comparables à celles obtenues ici. Il est par conséquent probable que la forte variabilité des réponses dans les deux conditions auralisées trouve son explication principalement dans

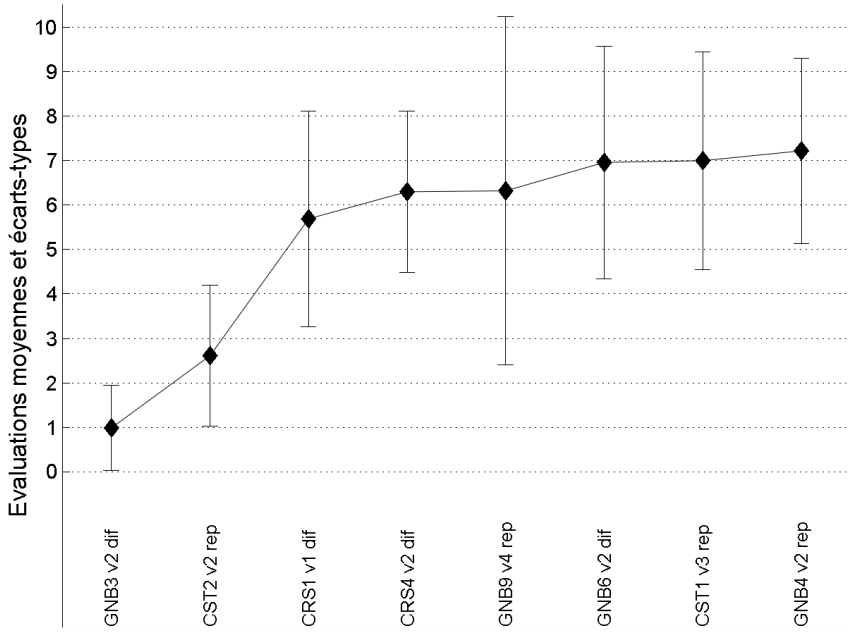


FIGURE 8.14 – Échelle moyenne et écarts-types pour les enregistrements anéchoïques diffusés au casque. L'échelle va de '0' pour les plus désagréables à '10' pour les plus agréables.

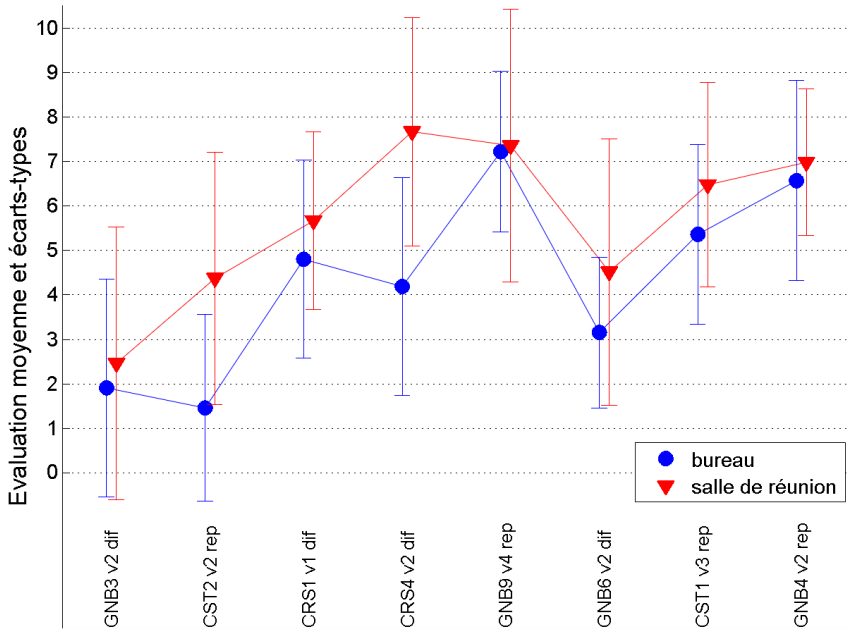


FIGURE 8.15 – Échelle moyenne et écarts-types pour les stimuli auralisés au casque. L'échelle va de '0' pour les plus désagréables à '10' pour les plus agréables.

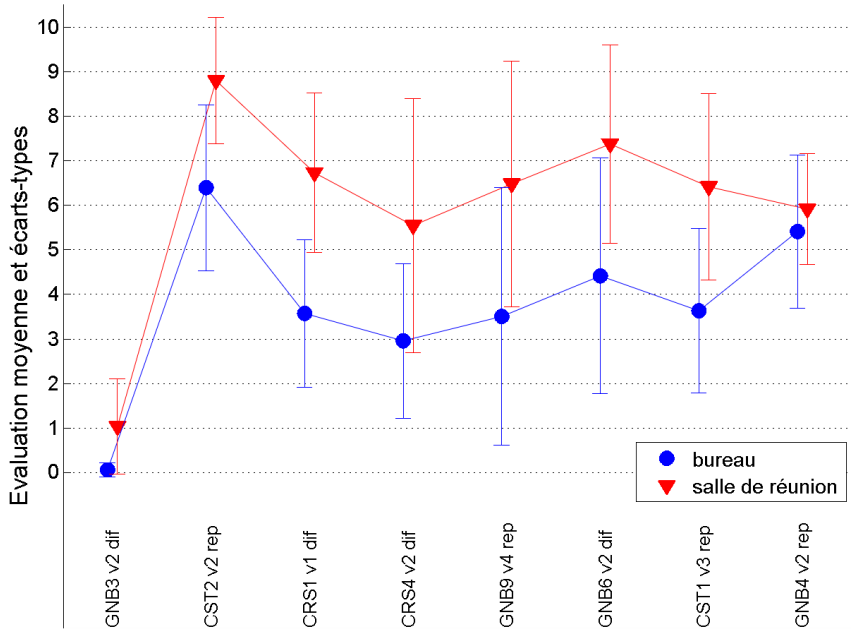


FIGURE 8.16 – Échelle moyenne et écarts-types pour les stimuli auralisés sur haut-parleurs. L'échelle va de '0' pour les plus désagréables à '10' pour les plus agréables.

le même phénomène que pour la condition anéchoïque. En effet, comme mentionné en section 7.7 pour la condition anéchoïque, il semble que la proximité générale des sons en termes de qualité sonore perçue ne permette pas de générer un consensus parmi les participants. Il est alors raisonnable de penser que cette proximité initiale (i.e. avant auralisation) des sons explique également en grande partie la forte variabilité observée dans le cas des conditions auralisées.

Il convient à présent de s'intéresser à la comparaison des évaluations moyennes entre la condition anéchoïque et les conditions auralisées. Tout d'abord, il est nécessaire de préciser que les valeurs de l'échelle de qualité sonore ne sont pas comparables dans l'absolu entre deux conditions différentes (c'est-à-dire qu'une valeur de '7', par exemple, dans le cas d'une condition n'est pas nécessairement une meilleure évaluation qu'une valeur de '6' dans une autre condition). En effet, chacune des conditions correspond à une expérience différente, où la consigne donnée aux participants précisait à chaque fois d'utiliser l'échelle présentée dans son ensemble. Ainsi, chacune des échelles des trois conditions sont des échelles relatives, qui n'ont donc à priori pas la même gamme de valeurs dans l'absolu. En revanche, dans le cadre de chacune des deux conditions auralisées, deux simulations de salle (voir section 8.2) ont été utilisées dans la même expérience, et il est donc possible d'en comparer les évaluations moyennes dans l'absolu.

Par ailleurs, puisque les trois conditions considérées utilisent 8 mêmes enregistrements de STA initiaux, il est intéressant de comparer l'ordonnancement de ces 8 STA sur chacune des échelles, afin d'observer l'éventuelle influence de la réverbération sur l'évaluation de la qualité sonore. Il est aisé de comparer cet ordonnancement en observant chaque courbe des figures 8.14, 8.15 et 8.16.

En comparant les deux premières citées, on observe que, si l'allure est globalement similaire (les courbes de la figure 8.15 identifient principalement les mêmes extrêmes), certains enregistrements de STA voient leur évaluation modifiée par l'auralisation d'une manière plus importante que les autres. Ce phénomène est assez significatif sur le son **GNB6 v2 dif** dont l'évaluation dans la condition aurali-

sée au casque est fortement pénalisée, vis-à-vis des autres enregistrements, par la réverbération [4]. Par ailleurs, l'ordonnancement semble assez similaire entre les deux simulations dans cette condition, puisque les deux courbes correspondantes (i.e. les deux courbes de la figure 8.15) ont globalement la même forme. Il apparaît toutefois que la simulation du bureau semble pénaliser davantage les évaluations que celle de la salle de réunion. Ceci peut sembler contradictoire à première vue lorsque l'on observe les valeurs des caractéristiques acoustiques des deux salles simulées (voir tableaux 8.2 et 8.3) qui nous indiquent que la simulation du bureau offre une qualité acoustique légèrement meilleure. Toutefois, la qualité acoustique que reflètent ces caractéristiques correspond plus à la propension de la salle à favoriser l'intelligibilité de la parole, ce qui est une problématique assez différente de celle abordée dans cette étude. En effet, nous cherchons ici à évaluer la qualité sonore de sons de STA dans chaque salle, et non la qualité acoustique de chacune d'elle. Par ailleurs, les évaluations globalement meilleures dans le cas de la salle de réunion pourraient trouver une forme d'explication dans la capacité à localiser la provenance du son dans les deux simulations de salle. En effet, une rapide écoute des sons permet de constater que s'il est possible d'identifier grossièrement la direction dans laquelle se trouve la source sonore dans le cas de la simulation de salle de réunion, la provenance du son est beaucoup moins nette dans le cas de la simulation de bureau individuel. La localisation plus aisée de la source pour la salle de réunion simulée pourrait donc avoir un effet bénéfique par rapport à une situation ou l'enveloppement de l'auditeur est plus important.

La comparaison des figures 8.14 et 8.16 semble indiquer une plus forte influence de l'auralisation sur les jugements des participants. En effet, on observe cette fois-ci que l'ordonnancement des stimuli est assez différent entre la condition anéchoïque et la condition auralisée sur haut-parleurs, et ceci pour les deux simulations, à la notable exception du son **GNB3 v2 dif** qui est perçu de manière consensuelle comme le plus désagréable. Toutefois, cette impression s'explique principalement par le son **CST2 v2 rep**, perçu comme l'un des plus désagréable en condition anéchoïque, et comme le moins désagréable en condition auralisée sur haut-parleurs. À l'écoute, l'enregistrement anéchoïque correspondant semble présenter une forme de « cliquetis », qui n'est plus audible une fois le son auralisé. Cet élément pourrait expliquer cette divergence. En dehors donc des sons **GNB3 v2 dif** et **CST2 v2 rep**, même si l'ordonnancement des six autres stimuli est différent, ils présentent des évaluations similaires compte tenu des écarts-types, aussi bien en condition anéchoïque qu'en condition auralisée. Par ailleurs, on observe dans le cas de l'auralisation sur haut-parleurs la même tendance que dans le cas de l'auralisation au casque entre les deux simulations de salle, c'est-à-dire le fait que la simulation de bureau individuel pénalise davantage les évaluations.

Enfin, les formes des courbes sont assez différentes entre la condition auralisée au casque (figure 8.15) et la condition auralisée sur haut-parleurs (figure 8.16). Cette observation peut paraître à première vue assez surprenante. Néanmoins, il semble que cette divergence soit principalement due aux deux stimuli mentionnés précédemment pour avoir généré des évaluations relativement différentes par rapport à la condition anéchoïque – **GNB6 v2 dif** pour l'auralisation au casque et **CST2 v2 rep** pour l'auralisation sur haut-parleurs. Par ailleurs, il importe également de garder à l'esprit que les deux types de diffusion du son correspondant à ces deux conditions auralisées sont très différentes. Notamment, l'auralisation au casque inclut un *filtre HRTF – Head Related Transfer Function*, fonction de transfert d'une tête humaine « standard » – qui n'est pas sans conséquence sur le timbre des sons perçus. Dans le cas de l'auralisation sur haut-parleurs, le son subit une modification similaire, chaque participant possédant sa propre *HRTF* réelle, qui peut être différente de la fonction *HRTF* standard

4. Nous pourrions également évoquer les sons **CRS4 v2 dif** et **GNB9 v4 rep** dont les évaluations sont à l'inverse légèrement meilleures en condition auralisée au casque, mais cette différence n'est pas significative à $p < 0,05$ (test de *Student*, échantillons appariés [82, 98], entre ces sons et les sons **CST1 v3 rep** et **GNB4 v2 rep** pour la salle de réunion).

utilisée pour l'auralisation au casque. Ceci pourrait expliquer en partie les différences observées.

D'un point de vue plus global, il est nécessaire d'extraire de cette étude des éléments permettant de juger de la pertinence des conditions expérimentales de l'évaluation de la qualité sonore, vis-à-vis de la problématique de la réverbération dans les lieux où sont usuellement installés les STA. Les points soulevés précédemment indiquent que la réverbération, dans le cadre de l'auralisation entourant ce travail expérimental, ne modifie pas de manière drastique la perception de la qualité sonore des STA relativement entre eux. Il est toutefois possible que la qualité sonore perçue de certains STA soit ponctuellement altérée par la réverbération, au point d'en modifier le classement vis-à-vis des autres STA.

Ainsi, les résultats obtenus dans un cadre expérimental incluant des stimuli correspondant à des enregistrements anéchoïques, tel que celui de l'étude présentée dans le chapitre 7, permettent d'obtenir une évaluation de la qualité sonore des STA globalement fiable. Cependant, il importe de prendre également en compte la réverbération afin d'obtenir une estimation précise du ressenti des usagers. Dans le contexte industriel encadrant ces travaux, ceci pourrait être réalisé en incorporant des métriques caractérisant la réverbération, telles que celles présentées en section 3.1.2, dans la prédiction de la qualité sonore. Toutefois, dans le cas présent, seules deux simulations différentes, et donc deux valeurs uniquement pour chacune de ces différentes métriques, étaient considérées. Ceci exclut ici la possibilité de lier de manière précise les mesures perceptives aux caractéristiques acoustiques des salles par des méthodes de régression.

Chapitre 9

Influence du contexte attentionnel

Le but de l'étude présentée dans ce chapitre est d'évaluer l'influence du contexte attentionnel sur la qualité sonore. On souhaite donc évaluer ici à quel point le degré de focalisation sur le son de l'attention des auditeurs modifie les jugements de ces derniers. Dans cette optique, la section 9.1 présente tout d'abord la problématique générale de cette étude. La section 9.2 présente le protocole expérimental spécialement conçu pour évaluer l'influence d'un contexte multi-tâche sur la qualité sonore perçue par les auditeurs. La section 9.3 présente l'analyse des résultats de cette expérience. Enfin, la section 9.4 expose la discussion que soulèvent les résultats de cette étude.

9.1 Problématique

L'idée se cachant derrière notre volonté d'incorporer, dans le chapitre 8, l'acoustique des salles dans l'évaluation de la qualité sonore des STA est d'établir les échelles perceptives au travers de conditions expérimentales écologiquement valides, c'est-à-dire dans des situations représentatives des conditions réelles dans lesquelles sont perçus les sons de STA. Dans ce cadre, un autre aspect de la démarche expérimentale utilisée pour l'évaluation de la qualité sonore est digne d'intérêt, l'influence du contexte attentionnel.

En effet, dans le cadre des méthodologies d'évaluation de la qualité sonore rencontrées dans la littérature – et dont les travaux exposés dans les chapitres 5, 6 et 7 sont fortement inspirés –, les protocoles expérimentaux mettent en jeu des procédures où la seule tâche demandée aux participants est d'évaluer les sons selon une certaine consigne. Ce contexte incite volontairement les participants à écouter attentivement les sons afin d'établir leurs jugements. C'est ce que nous appelons l'*écoute attentive*. Or, dans une situation réaliste où le son de STA est typiquement perçu, les auditeurs ne sont ni passifs, ni totalement concentrés sur le son produit. Le plus souvent, ils perçoivent le son tandis qu'ils sont occupés à effectuer une autre activité, et le considèrent plutôt comme une source d'intrusion dans leur vie quotidienne. Ainsi les usagers n'écoutent pas attentivement le son de STA mais focalisent leur attention sur une autre tâche, quelle qu'elle soit, ce qui a tendance à rendre la perception de ce type de sons « intermittente », voir complètement masquée par l'activité qui les occupe.

Ainsi, la condition d'*écoute attentive*, couramment adoptée dans le cadre de l'évaluation de la qualité sonore, n'est pas entièrement fidèle à la perception réelle des sons de STA. Sur la base de cette observation, deux hypothèses peuvent être posées :

– soit le contexte attentionnel d'écoute n'a aucune influence significative sur les jugements des auditeurs, en ce sens que leurs réponses obtenues dans le cadre d'une méthodologie expéri-

mentale telle que celle décrite dans les chapitres 5, 6 et 7 ne seraient pas différentes dans des conditions plus réalistes ;
– soit ce contexte à un effet significatif – qu'il convient alors de préciser – sur les jugements, et les réponses obtenues ne sont pas représentatives du ressenti réel des usagers.

Le but du travail décrit dans ce chapitre est d'explorer ces deux possibilité et donc de répondre à la question suivante : la qualité sonore mesurée des STA serait-elle différente si les enregistrements étaient présentés aux participants alors que ces derniers doivent effectuer une autre tâche accaparant leur attention ? Plus exactement, à quel point les jugements des auditeurs diffèrent entre une situation d'écoute attentive du son de STA et une situation où leur attention est détournée du son, que nous appelons situation d'*écoute distraite* ? L'idée majeure est donc de mettre au point une procédure expérimentale de mesure de qualité sonore incluant une tâche subsidiaire dont le but est de focaliser l'attention des participants de sorte à ce qu'ils ne prêtent plus attention au contexte sonore. Comme mentionné en section 3.3.2, il est possible de trouver des études exposant des travaux expérimentaux mettant en jeu des types de procédure similaires dans la littérature, mais leur existence y reste toutefois sporadique, et leurs objectifs respectifs sont souvent bien différents du nôtre.

La procédure mise au point, et décrite en section 9.2.4, inclut donc une tâche subsidiaire fortement inspirée par l'étude de Susini et McAdams [144] évoquée en section 3.3.2, et peut être également rapprochée de la récente étude d'Ebissou et al. [52], portant sur l'influence de différents degrés d'intelligibilité (i.e. différentes valeurs de STI – voir section 3.2.2) sur la performance des auditeurs. Le choix de la tâche subsidiaire, qui consiste à mémoriser une séquence de chiffres, représente un bon compromis entre simplicité, afin qu'elle soit compréhensible et réalisable par tout participant, et charge cognitive importante, afin de garantir un maximum d'attention focalisée sur cette tâche. L'évaluation de chaque son est effectuée à posteriori, lorsque l'on demande aux auditeurs, une fois la tâche de mémorisation et la lecture du son terminées, à quel point le son les a perturbés dans la réalisation de cette tâche.

À terme, le but de cette étude est de comparer le degré auquel les différents sons perturbent la réalisation de la tâche de mémorisation avec l'échelle de qualité sonore de référence, décrite en section 7.5 et obtenue dans un contexte d'écoute attentive du son. Le degré de perturbation provoquée par les sons pourra être évalué au travers de deux paramètres : d'une part le ressenti des auditeurs, tel qu'ils l'expriment lorsque cela leur est demandé, et, d'autre part, leur performance dans la réalisation de la tâche de mémorisation. Ainsi, il sera possible de déterminer si les jugements des auditeurs sont significativement influencés par le contexte attentionnel.

9.2 Protocole expérimental d'évaluation en contexte multi-tâche

9.2.1 Stimuli

Le corpus sonore pour cette expérience est constitué de 5 sons provenant du corpus de travail établi en section 5.3. Ces 5 sons font partie également de la sélection de 8 sons effectuée pour l'étude de l'influence de la réverbération (voir section 8.3.2). Les raisons de cette réduction de taille du corpus sonore reposent sur un compromis nécessaire entre durée d'expérience et nombre d'évaluations requises par la procédure décrite en section 9.2.4.

En voici la liste :

4. **CRS4 v2 dif**

7. **GNB3 v2 dif**
8. **GNB4 v2 rep**
13. **GNB9 v4 rep**

14. **CST1 v3 rep**

La durée des sons utilisés était ici de 2 minutes, bien que leur durée effective dépendait de la durée de chaque session de mémorisation (voir section 9.2.4). En effet, leur lecture était interrompue avant la fin, lorsque la session de mémorisation était terminée. La durée de la session de mémorisation se situait usuellement entre 50 et 70 secondes, en fonction de la promptitude du participant à entrer les chiffres de la séquence.

La question de l'égalisation en sonie des 5 sons du corpus peut être soulevée ici. Cependant, il a été démontré que, si la nature du son a une réelle influence sur la performance des auditeurs dans une tâche de mémorisation de séquences de chiffres, cette performance ne dépend pas du niveau sonore des stimuli [152]. Par conséquent, nous avons choisi, par souci de cohérence par rapport aux expériences exposées dans les chapitres précédents, d'égaliser en sonie les sons du corpus. Compte tenu de la stationnarité des sons considérés, cette opération a été réalisée en appliquant les facteurs obtenus pour les versions courtes de ces mêmes sons au cours de l'égalisation en sonie automatique décrite en section 5.2.1.

Le corpus est donc constitué de 5 sons monophoniques égalisés en sonie d'une durée de 2 minutes (bien qu'ils étaient interrompus avant d'atteindre cette durée, comme expliquer précédemment), et dont le niveau acoustique varie de 38,1 à 40,1 dBA.

9.2.2 Matériel

Une interface graphique spécifique (figures 9.1, 9.2 et 9.3), assurant la lecture des sons sur l'interface audio, l'affichage et la saisie des séquences de chiffres par l'utilisateur, et l'enregistrement des évaluations des participants, a été programmée en LabVIEW 2010 (figure 7.4) pour cette expérience. Les sons ont été diffusés par une interface RME Fireface 800 dans un casque ouvert Sennheiser HD650. Le mode de reproduction sonore sur casque a été employé car c'est le seul qui garantit d'obtenir une écoute diotique (signal identique dans les deux oreilles). L'expérience a eu lieu dans une cabine audiométrique IAC à double paroi.

9.2.3 Participants

Vingt-neuf participants (20 hommes et 9 femmes de l'Université de La Rochelle, âgés entre 20 et 25 ans), n'ayant pris part à aucune autre expérience réalisée dans le cadre de cette thèse, se sont portés volontaires pour cette expérience. Aucun d'entre eux n'a fait mention d'un problème majeur d'audition. Toutefois, 2 d'entre eux ont été rapidement retirés du panel, leurs résultats s'avérant incohérents avec les instructions données. Le panel est donc finalement constitué de 27 participants.

9.2.4 Procédure

Au début de l'expérience, les participants ont reçu une consigne écrite, retranscrite en section E.4 en annexe, présentant le contexte de l'étude et leur expliquant la tâche à accomplir. Il devait remplir

2 tâches successives :

– Dans la première, des chiffres sont présentés un à un sur l'écran (voir figure 9.1) formant ainsi une séquence à mémoriser et à entrer au clavier (voir figure 9.2) à la fin de l'affichage des chiffres. Après validation, une autre séquence à mémoriser est présentée de la même manière, et ainsi de suite jusqu'à la fin du temps imparti. Ce dernier est de 20 secondes, et n'inclut pas le temps d'affichage des chiffres ; seul le temps d'entrée au clavier est pris en compte. Le temps global d'affichage des chiffres est variable car il est lié au nombre de séquences affichées, et donc dépend de la promptitude des participants à entrer les séquences de chiffres. La durée globale (temps d'affichage et temps de réponse compris) d'une tâche de mémorisation varie entre 50 et 70 secondes.

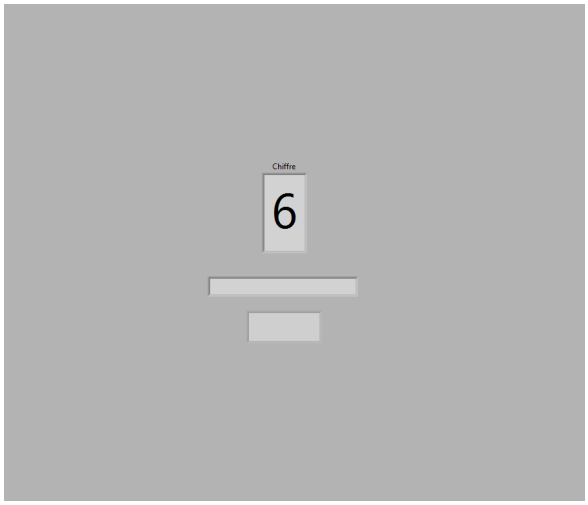


FIGURE 9.1 – Interface graphique de la tâche de mémorisation - affichage des chiffres de la séquence.

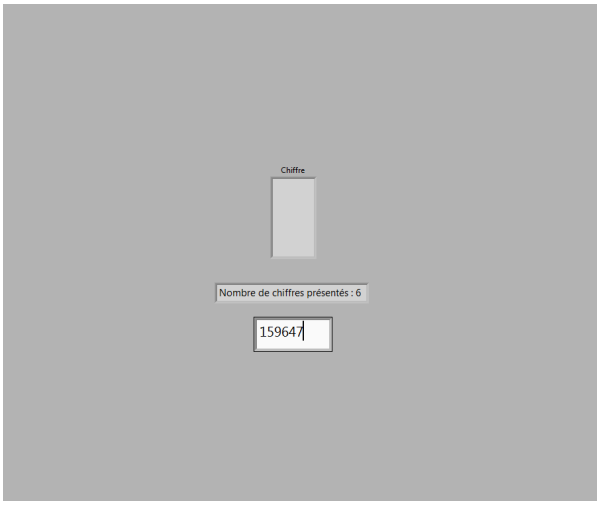


FIGURE 9.2 – Interface graphique de la tâche de mémorisation - entrée au clavier de la séquence de chiffres.

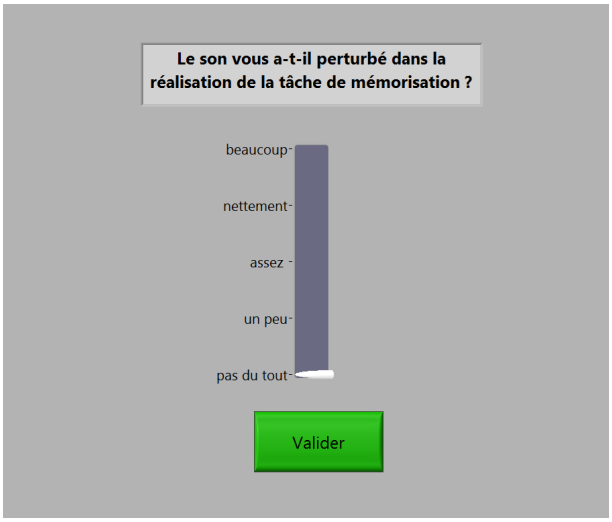


FIGURE 9.3 – Interface graphique de la tâche d'évaluation.

La première séquence est constituée de 4 chiffres, et ce nombre est incrémenté chaque fois que le participant répond correctement à 2 séquences de même longueur. Un des 5 sons du corpus

est joué au casque pendant toute la durée de cette phase, à l'exception d'une condition « silence » de référence où aucun son n'est joué (par opposition aux conditions « sonores », lorsque un son est joué). Cette condition silence est utilisée afin d'observer une éventuelle influence de la présence de son sur la performance des participants dans la réalisation de la tâche de mémorisation.

– A la fin du temps imparti, la seconde tâche consiste à évaluer à quel point le son a interféré avec la réalisation de la tâche de mémorisation sur une échelle catégorielle à 5 niveaux (voir figure 9.3). Les étiquettes associées aux 5 niveaux sont : « pas du tout », « un peu », « assez », « nettement », et « beaucoup ». Il est important de noter qu'aucun son n'est joué pendant cette tâche, sa lecture ayant été interrompu à la fin de la tâche de mémorisation. Les participants n'ont à ce stade aucun moyen d'écouter de nouveau le son joué. Par ailleurs, cette tâche ne doit être effectuée que dans le cas où un son a été joué pendant la phase de mémorisation, ce qui n'est pas le cas dans la condition silence mentionnée ci-dessus.

La succession de ces 2 tâches est répétée 3 fois pour chacun des 5 sons du corpus dans un ordre aléatoire. Cependant, chaque répétition d'une tâche de mémorisation avec le même son est nécessairement séparée par une tâche de mémorisation avec un autre son du corpus (ou sans son). La condition silence est quant à elle répétée 5 fois. Ainsi, 20 successions de couple de tâche mémorisation/évaluation (sauf pour la condition silence, où seule la tâche de mémorisation est à accomplir) sont à présenter aux participants au cours de l'expérience.

La durée moyenne de la procédure dans le cas de cette expérience a été de 26 minutes.

9.3 Résultats et analyse

Pour chaque participant et chaque condition, sonore ou silence, il importe de s'intéresser à deux résultats principaux. Premièrement, son évaluation du degré auquel le son a interféré avec la réalisation de la tâche de mémorisation (pour les conditions avec son joué uniquement) est moyenné sur les 3 présentations de la même condition sonore. Deuxièmement, la performance du participant dans la réalisation de la tâche de mémorisation peut être estimée comme le nombre de séquences correctement mémorisées dans le temps imparti, moyenné sur les 3 ou 5 présentations de la même condition sonore.

9.3.1 Évaluation du degré de gêne

Concernant le premier point d'intérêt, c'est-à dire le degré de gêne qu'ont ressenti les participants, la figure 9.4 montre les évaluations moyennées sur le panel de participants et leur écart-type pour chacun des 5 sons. Sur cette figure, les valeurs sur l'axe des ordonnées représentent les étiquettes de l'échelle allant de '0' pour « pas du tout » à '4' pour « beaucoup ». On peut y observer de fortes variations parmi les participants.

Compte tenu des écarts-types important des évaluations, il semble alors impossible de faire apparaître suffisamment de différences significatives entre les évaluations moyennes de chaque son. Selon un test de *Student* (échantillons appariés) [82, 98], seul le son **GNB3 v2 dif** diffère de manière significative des autres à $p < 0,05$. L'écart-type moyen et_m vaut 0,90, tandis que l'écart-type des évaluations moyennes m_{et} vaut 0,29. Les coefficients de concordance de Kendall et de corrélation de rang

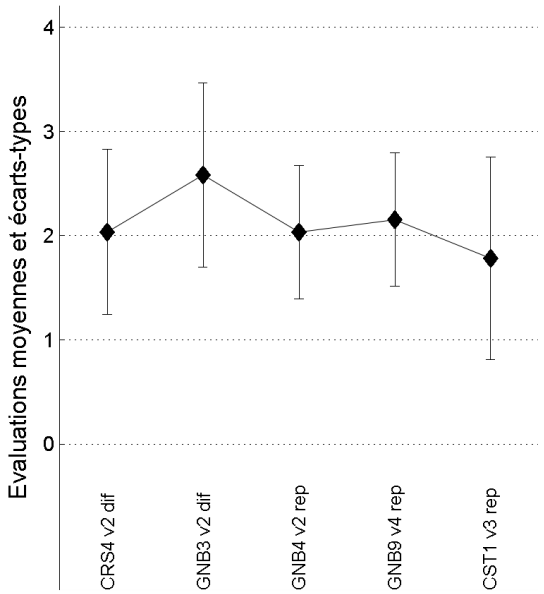


FIGURE 9.4 – Échelle moyenne et écarts-types des évaluations du degré de gêne pour l'ensemble du panel de participants. Les 5 valeurs de '0' à '4' correspondent respectivement aux étiquettes des 5 niveaux de l'échelle : « pas du tout », « un peu », « assez », « nettement », et « beaucoup ».

de Spearman, qui traduisent le degré de concordance de l'ordonnancement des sons sur l'échelle des évaluations par les différent participants (voir sections 7.3, et C.1 en annexe, pour plus de détails) sont particulièrement faibles, et n'atteignent que, respectivement, $W = 0,10$ et $\bar{r}_s = 0,07$. Ces éléments statistiques inter-participants sont résumés dans la première colonne du tableau 9.1 et confirment que les différences entre les évaluations moyennes des 5 sons sont trop peu significatives.

En conséquence, il est nécessaire de s'intéresser à la répartition des évaluations individuelles autour des valeurs moyennes. Dans cette optique, une analyse de cluster a été appliquée aux résultats individuels (voir section C.2 en annexe). La représentation hiérarchique résultant de cette analyse est affichée en figure 9.5 .

À la lumière de ce dendrogramme, il semble qu'il y ait une importante divergence dans les résultats de deux groupes de participants (mise en évidence par la ligne verticale rouge en pointillés), qui explique probablement en partie les forts écarts-types et la faible différenciation des évaluations moyennes des 5 sons. En effet, le nœud où se rejoignent les clusters correspondant à ces deux groupes se situe à une distance cophénétique aux alentours de $0,65$, ce qui correspond à un coefficient de corrélation de $-0,3$. Le groupe 1 est constitué de 18 participants, tandis que le groupe 2 est constitué de 9 participants. Il n'a pas été mis en évidence de lien apparent entre cette séparation et les « caractéristiques » des participants (âge, sexe, etc.).

Le groupe 2, plus petit des deux, étant ici d'une taille significative, il ne s'agit pas seulement d'en considérer les éléments comme des *outliers*, mais bien de le considérer comme une tendance différente bien réelle. En conséquence, plutôt que de retirer ces participants du panel, les 2 groupes ont été considérés séparément dans les analyses subséquentes. Il convient donc tout d'abord de représenter, comme en figure 9.4, les évaluations moyennes et leurs écarts-types pour les 2 groupes de participants. Les figures 9.6 et 9.7 permettent d'observer ces éléments respectivement pour le groupe 1 et pour le groupe 2. Sur ces figures, les valeurs sur l'axe des ordonnées représentent les étiquettes de l'échelle allant de '0' pour « pas du tout » à '4' pour « beaucoup ». Les valeurs des éléments statistiques inter-participants sont légèrement améliorés pour les deux groupes, et valent respectivement

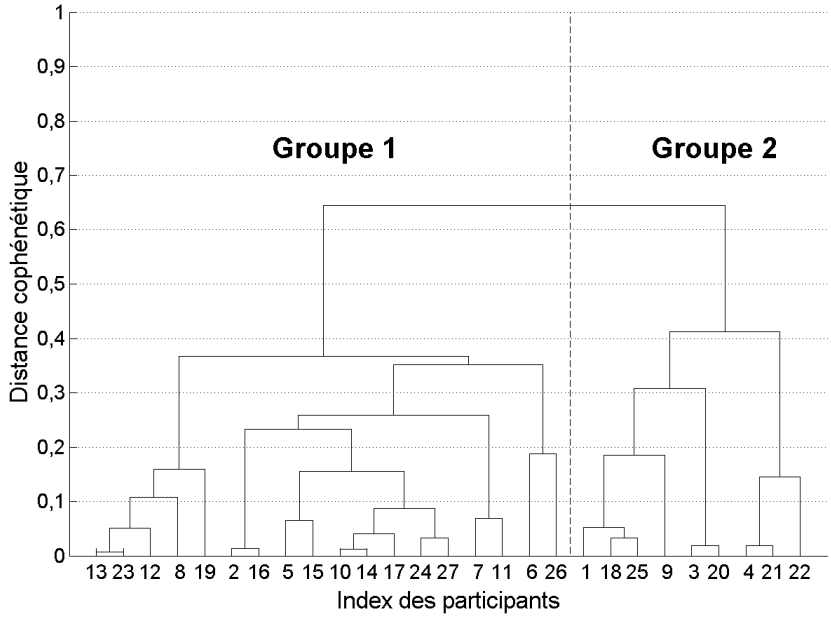


FIGURE 9.5 – Dendrogramme issu des corrélations inter-participant des évaluations du degré de gêne.

$et_{m_1} = 0,76$, $m_{et_1} = 0,53$, $W_1 = 0,36$ et $\bar{r_{s_1}} = 0,33$, et $et_{m_2} = 0,86$, $m_{et_2} = 0,62$, $W_2 = 0,40$ et $\bar{r_{s_2}} = 0,33$. Ils sont résumés dans les deuxième et troisième colonnes du tableau 9.1. Bien qu'ils ne soient que peu satisfaisants, on peut tout de même distinguer, à la vue des formes respectives des deux courbes sur les figures 9.6 et 9.7, que l'ordonnancement des sons sur l'échelle des évaluations moyennes diffère entre les deux groupes.

	Panel	Groupe 1	Groupe 2
et_m	0,90	0,76	0,86
m_{et}	0,29	0,53	0,62
W	0,10	0,36	0,40
$\bar{r_s}$	0,07	0,33	0,33

TABLE 9.1 – Statistiques inter-participants des évaluations du degré de gêne pour, respectivement, l'ensemble du panel de participants, le groupe 1 et le groupe 2 (écart-type moyen entre les participants et_m, écart-type entre les évaluations moyennes des sons m_{et}, coefficient de concordance de Kendall W et coefficient de corrélation de rang de Spearman $\bar{r_s}$).

Les résultats affichés sur les figures 9.6 et 9.7 doivent être considérés vis-à-vis d'une évaluation comparable des sons dans un contexte classique où les sons sont jugés sans tâche subsidiaire, c'est-à-dire en contexte d'écoute « attentive ». La figure 9.8 reprend les résultats exposés en section 7.5, correspondant à une expérience réalisée en condition d'écoute attentive. L'échelle a été spécialement adaptée afin d'adapter sa dynamique à celle des résultats présentés ici. En effet, lors de cette expérience, l'échelle mesurée représentait la qualité sonore allant de '0' pour le « plus désagréable » à '10' pour le « plus agréable ». Compte tenu de l'échelle utilisée ici, la transformation linéaire (et inversée) de l'échelle p_1 alors obtenue en une échelle comparable ici p_2 consiste en l'opération suivante :

$$p_2 = \left(1 - \frac{p_1}{10}\right) \cdot 4 \tag{9.1}$$

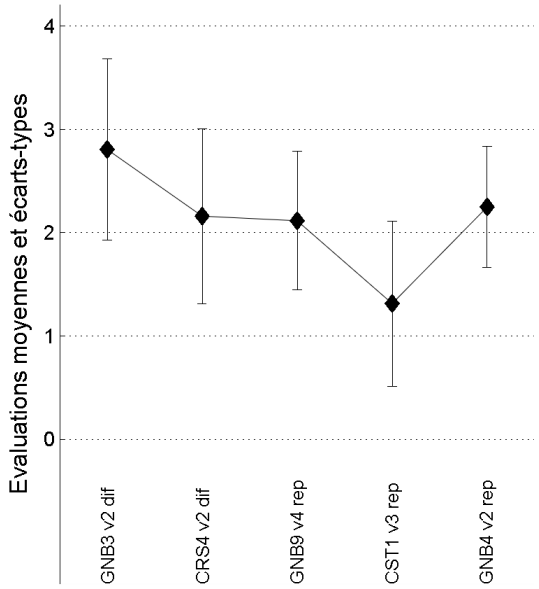


FIGURE 9.6 – Échelle moyenne et écarts-types des évaluations du degré de gêne pour le premier groupe de participants. Les 5 valeurs de '0' à '4' correspondent respectivement aux étiquettes des 5 niveaux de l'échelle : « pas du tout », « un peu », « assez », « nettement », et « beaucoup ».

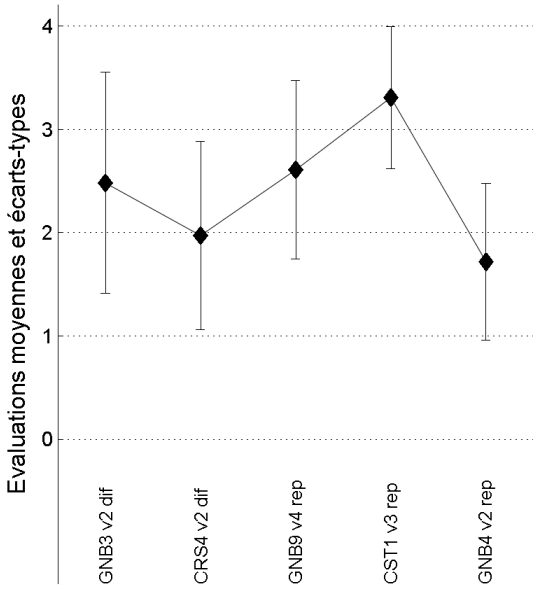


FIGURE 9.7 – Échelle moyenne et écarts-types des évaluations du degré de gêne pour le second groupe de participants. Les 5 valeurs de '0' à '4' correspondent respectivement aux étiquettes des 5 niveaux de l'échelle : « pas du tout », « un peu », « assez », « nettement », et « beaucoup ».

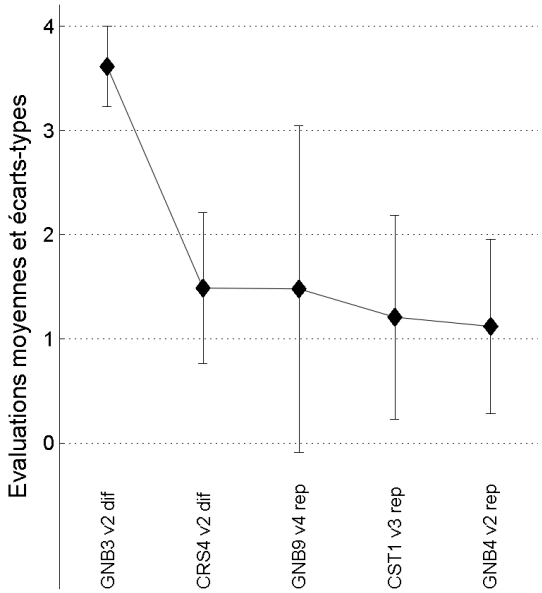


FIGURE 9.8 – Échelle moyenne et écarts-types des évaluations du caractère agréable/désagréable obtenus dans un contexte expérimental monotâche. L'échelle de '0' à '4' correspondent à l'échelle de qualité sonore mesurée au cours de l'expérience décrite en section 7.5, ayant été inversée (de '10' à '0') à l'aide de la formule 9.1.

L'observation des résultats de l'expérience en contexte d'écoute attentive permet de également de préciser les tendances aperçues sur les figures 9.6 et 9.7. Il semble que les résultats des participants du groupe 1 rejoignent ceux obtenus en condition d'écoute attentive en ce sens que l'ordonnancement des sons sur l'échelle est similaire (bien que plus différencié pour la condition d'écoute attentive). À l'inverse, les résultats du groupe 2 montrent une divergence relativement forte avec ceux de la condition d'écoute attentive. Notamment, le son **GNB3 v2 dif**, qui émerge clairement des autres comme un son particulièrement désagréable en condition d'écoute attentive, s'avère bien moins dérangeant en contexte multi-tâche pour ce groupe de participants, et n'apparaît pas comme celui du corpus qui les a le plus perturbés. Le phénomène inverse se produit pour le son **CST1 v3 rep**.

Si tous les participants avaient manifesté la même tendance que celle du groupe 1, on pourrait conclure que le contexte attentionnel dans la procédure expérimentale pour la mesure de la gêne n'a pas d'influence sur les résultats. Mais en l'occurrence, l'identification d'un groupe minoritaire de participants montrant une certaine divergence avec la majorité empêche une telle conclusion. Il semble au contraire que le contexte attentionnel ait une influence significative sur les jugements de certains auditeurs, au moins dans le cas des sons de STA.

Par ailleurs, nous pourrions être tentés de comparer dans l'absolu les évaluations obtenues en condition d'écoute attentive et celles obtenues au cours de cette expérience, en condition d'écoute distraite. On remarque effectivement que, pour les deux groupes, à l'exception du son **GNB3 v2 dif**, les évaluations des quatre autres sons sont globalement plus élevées que les évaluations (telles que transformées par la formule 9.1) obtenues en condition d'écoute attentive. On pourrait ainsi conclure que la condition d'écoute attentive sous-estime la gêne réellement ressentie par les auditeurs. Toutefois, la nature des échelles d'évaluation proposées aux participants dans les deux expériences sont très différentes. Si, dans le cas de l'écoute attentive, l'échelle était relative et la consigne donnée aux participants leur demandait explicitement d'utiliser l'ensemble des valeurs possibles, l'échelle utilisée ici peut être considérée comme absolue (de « pas du tout » à « beaucoup »), et la consigne donnée aux participants ne leur imposait pas d'utiliser l'ensemble de la gamme de valeurs. Par conséquent, il est impossible de comparer d'un point de vue global et absolu les évaluations dans les deux conditions.

9.3.2 Performance dans la tâche de mémorisation

Concernant le second point d'intérêt, c'est-à-dire la performance des participants dans la tâche de mémorisation, les données prennent la forme d'un nombre de séquences – par la suite nommé « score » – correctement mémorisées dans le temps imparti, pour chaque condition sonore ou silence présentée. Ce score est moyenné sur les 3 ou 5 présentations de la même condition. Nous obtenons donc pour chacun des participants un jeu de 6 scores correspondant à leur performance dans la condition silence et dans les 5 conditions où un son leur a été présenté. La figure 9.9 montre les scores moyennés sur l'ensemble des 27 participants qui ont été pris en compte, et les écarts-types associés. Selon un test de *Student* (échantillons appariés) [82, 98], les seules différences significatives à $p < 0,05$ sont entre les scores moyens de la condition silence et les conditions **GNB3 v2 dif** et **CST1 v3 rep**, et entre les conditions **CST1 v3 rep** et **GNB9 v4 rep**. La condition silence ne semble donc que peu favoriser la performance des participants par rapport aux conditions où un des sons est présenté. Par ailleurs, aucun participant n'a présenté un score en moyenne nettement supérieur (c'est-à-dire au moins 2 séquences correctement mémorisées supplémentaires) dans la condition silence par rapport à sa moyenne sur les 5 conditions sonores, certains participants présentant même un score moyen inférieur pour la condition silence.

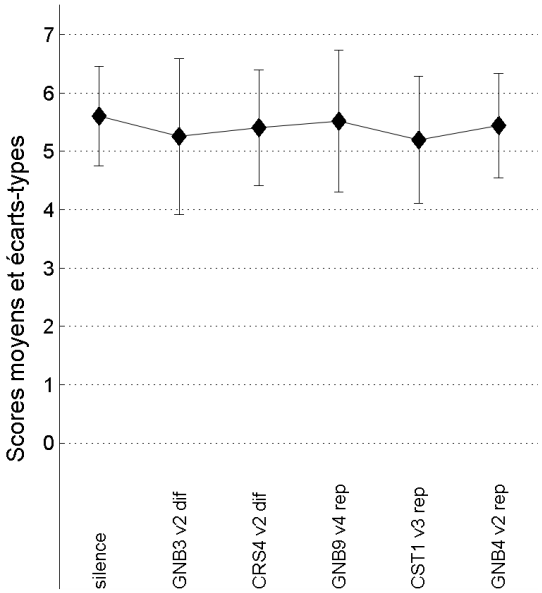


FIGURE 9.9 – Scores (nombre de séquences correctement mémorisées) moyens obtenus par le panel de 27 participants dans la condition silence et dans les 5 conditions sonores.

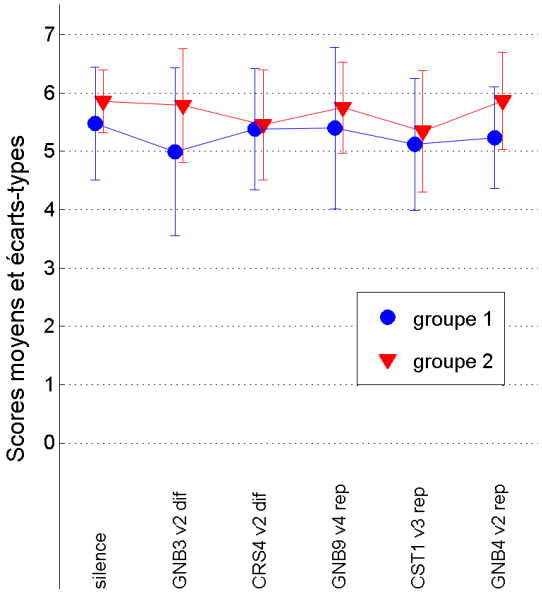


FIGURE 9.10 – Scores (nombre de séquences correctement mémorisées) moyens obtenus par chacun des 2 groupes de participants dans la condition silence et dans les 5 conditions sonores.

En revanche, il n'est pas dénué d'intérêt de s'intéresser en particulier aux scores moyens obtenus par chacun des 2 groupes de participants identifiés ci-dessus (voir figure 9.5). La figure 9.10 présente les scores moyens obtenus par chacun de ces 2 groupes, ainsi que les écarts-types associés. Bien qu'il ne soit de nouveau pas possible de mettre en évidence, à l'aide d'un test de *Student* [82, 98], de différence significative à $p < 0,05$, aussi bien entre les conditions sonores pour un groupe ou l'autre (échantillons appariés) qu'entre les 2 groupes (échantillons indépendants), il apparaît une légère amélioration des scores moyens pour le groupe 2. Cette différence est plus importante pour 2 des sons, mais le score moyen de ce groupe est égal ou supérieur pour les 6 conditions présentées. Si les conclusions que l'on pourrait tirer de ces observations souffrent de la non-significativité statistique

de ces différences, cette divergence de performance pourrait être une explication de la séparation du panel de participants en 2 groupes distincts. Il est en effet possible qu'elle traduise une différence d'implication et de concentration des participants dans la réalisation de la tâche de mémorisation :

- Les participants du groupe 2, peut-être plus absorbés par la tâche de mémorisation, ont eu une écoute moins attentive des sons. S'en est traduit des jugements de perturbation différents de ceux obtenus en condition d'écoute attentive. Ceci pourrait expliquer les formes de courbe différentes observées sur les figures 9.8 et 9.7.
- Les participants du groupe 1, en revanche, se sont moins impliqués dans la réalisation de la tâche de mémorisation. Leur attention a alors pu être plus facilement captée par le son diffusé, et les conditions d'expérience se sont alors rapprochées de celles réalisées en écoute attentive. Ceci expliquerait alors la ressemblance entre les courbes des figures 9.8 et 9.7.

9.4 Discussion

L'idée majeure de cette étude était de comparer les évaluations de la qualité sonore obtenus dans deux conditions différentes, reflétant deux degrés bien différents de focalisation de l'attention des auditeurs sur le son. Ainsi, la procédure expérimentale spécifiquement conçue dans cette optique, et décrite en section 9.2.4, nous a permis d'obtenir une évaluation de la qualité sonore en situation d'écoute distraite, contexte attentionnel plus proche de la condition réelle de perception des sons de STA qu'une condition d'écoute attentive.

La première conclusion que l'on peut extraire des résultats de cette expérience est que la forte variabilité des réponses, déjà observée au cours des précédentes procédures d'évaluation de la qualité sonore (chapitres 7 et 8), est toujours présente avec la procédure expérimentale mise en place ici. Elle semble même plus importante si l'on compare les statistiques inter-participants obtenues (voir tableau 9.1) avec celle des précédentes expériences (voir tableaux 7.4 et 8.5). Cette comparaison et celle des figures 9.6, 9.7 et 9.8 permet également de constater que la gamme de valeurs obtenues dans cette condition d'écoute distraite est plus réduite que dans le cas de l'expérience de référence, en écoute attentive. Ces constats ne sont pas sans rappeler les résultats obtenus dans l'étude de Gros et al. [71] évoquée en section 3.3.2. En effet, cette étude avait mis à jour les mêmes phénomènes dans le cadre de jugements de qualité audio de signaux de parole. Pour les deux contextes attentionnels qui y étaient considérés, la gamme de valeurs moyennes était néanmoins plus importante que dans notre cas, tandis que les écarts-types étaient plus faibles, .

Il est également nécessaire de garder à l'esprit que les évaluations ont été obtenues ici d'une manière sensiblement différente dans les deux conditions d'écoute attentive et d'écoute distraite . En effet, dans le cas de l'expérience exposée dans ce chapitre, l'échelle présentée aux participants, ainsi que la question qui leur était posée au moment d'évaluer les sons, ont subi, par rapport aux précédentes expériences, une importante modification qu'imposait le contexte de présentation des sons. Les participants, n'ayant pas la possibilité d'écouter les sons au moment de les évaluer, pouvaient éprouver des difficultés à se souvenir précisément du son s'ils avaient réalisé la tâche de mémorisation avec suffisamment d'application. Ceci pourrait expliquer en partie qu'il est plus difficile d'extraire des évaluations moyennes consensuelles dans le cas présent.

La seconde information, déjà évoquée en section 9.3, que l'on obtient à la lumière de ces résultats est que tous les participants n'adoptent pas la même stratégie vis-à-vis des deux tâches demandées. Cela se traduit, malgré la grande variabilité globale évoquée précédemment, par une forte séparation

entre deux groupes de participants sur la figure 9.5, et par un ordonnancement très différent des évaluations moyennes des deux groupes sur les figures 9.6 et 9.7.

Nous pouvons identifier deux particularités de la procédure expérimentale qui sont susceptibles d'avoir une responsabilité dans ce phénomène. Tout d'abord, il est possible que la consigne fournie aux participants n'ait pas assez insisté sur le fait qu'ils devaient tenter d'être le plus performant possible dans la tâche de mémorisation, afin de maximiser leur concentration pendant cette phase. Ainsi, il se peut que certains auditeurs, sachant qu'ils devaient après chaque session de mémorisation évaluer le son qu'ils avaient entendu, ont été inconsciemment incités à prêter trop attention à celui-ci, et à ainsi négliger la tâche de mémorisation. Le second élément qui pourrait expliquer cette divergence entre les deux groupes de participants est la courte durée des sessions de mémorisation. Cette durée répondait dans le cas présent à un compromis entre nombre de conditions sonores à présenter aux auditeurs et durée globale d'expérience. Il est probable que cette courte durée ait entraîné des fluctuations de concentration et d'attention d'une condition sonore à l'autre, tandis qu'une durée plus longue aurait peut-être permis d'avoir un degré de concentration plus stable entre les différentes conditions sonores. En effet, si la session est plus longue, ces fluctuations serait très certainement intégrées et moyennées au sein d'une même condition sonore.

Par ailleurs, cette courte durée de session de mémorisation est également à relier aux faibles variations de performances entre les différentes conditions sonores (voir figure 9.9). En effet, une conclusion similaire avait été observée dans l'étude de Bradley et Gover [37], également évoquée en section 3.3.2, abordant la problématique de l'évaluation du degré de distraction provoquée par différentes conditions d'intelligibilité et de bruit de fond. Dans cette étude, les auteurs n'avaient pas observé de variations majeures de performance, et avaient expliqué ce fait par la durée des tâches accaparant l'attention (1 ou 2 minutes selon la tâche, donc très similaire à la durée de la tâche accaparante que nous avons utilisée). Dans le cas de notre étude, seule une légère divergence de performance a pu être mise à jour entre les deux groupes de participants, mais elle s'avère non-significative d'un point de vue statistique rigoureux (voir figure 9.10). Une durée plus longue de la tâche de mémorisation aurait probablement laissé apparaître des différences de performance plus prononcées entre chaque condition sonore.

En outre, il est intéressant d'observer également que lors d'une étude portant sur l'effet de différents degrés d'intelligibilité sur la performance d'auditeurs dans la réalisation d'une tâche semblable de mémorisation de séquences de chiffres [52], Ebissou et al. avaient observé un phénomène similaire de séparation en deux groupes d'auditeurs. Toutefois, dans le cas de cette étude, cette séparation était liée aux performances et non aux évaluations de la difficulté qu'ils avaient ressentie (ce qui représente une évaluation indirecte des stimuli). Dans notre cas, l'étude de la performance des auditeurs n'a pas mis à jour différentes tendances parmi les participants, en dehors de celle, très légère, évoquée précédemment. Il convient de préciser, que dans le cas de l'étude d'Ebissou et al., la durée de session de mémorisation était de 10 minutes, donc beaucoup plus longue que dans notre cas.

Dans le contexte industriel qui caractérise cette thèse, il est à ce stade difficile d'intégrer les résultats présentés dans ce chapitre, par rapport à la problématique de départ qui est d'établir une métrique de qualité sonore. En effet, devant le peu de différenciations significatives des mesures perceptives réalisées ici, et le nombre réduit de stimuli considérés, il semblerait incongru de tenter de relier les échelles à des paramètres tangibles du signal sonore tout en y intégrant des paramètres reflétant le contexte attentionnel – si tant est qu'il soit possible de définir ces derniers.

En revanche, cette approche expérimentale de l'évaluation de la qualité sonore semble indiquer

qu'il est nécessaire prendre en compte le contexte attentionnel lorsque l'on souhaite obtenir une mesure perceptive reflétant de manière réaliste le ressenti des auditeurs. Il semble effectivement que le degré de focalisation de l'attention des auditeurs sur le son puisse modifier dans une certaine mesure leurs jugements. Cette tendance nécessiterait probablement d'être confirmée plus formellement, et sur une plus grande variété de sons de l'environnement. Il importe toutefois de conserver à l'esprit que cette démarche expérimentale a été adoptée en réponse au contexte usuel d'écoute des sons de STA, et ce contexte ne vaut pas nécessairement pour tous les types de sons de l'environnement.

Conclusions

Le travail exposé dans ce document aborde la problématique de la perception des sons de Systèmes de Traitement d'Air (STA), et le confort acoustique ressenti par les usagers. D'un point de vue industriel, le but de ce travail est de fournir un ensemble d'outils permettant de caractériser acoustiquement un STA à partir d'un enregistrement du son produit, d'une manière qui soit représentative de sa perception réelle. Afin d'atteindre cet objectif, deux axes de recherches ont été suivis : en premier lieu la caractérisation, aussi bien descriptive qu'hédonique, des sources sonores que sont les STA ; et dans un second temps, l'étude de l'influence sur cette caractérisation de différents éléments liés au contexte d'écoute, afin de se rapprocher d'une définition de la qualité sonore représentative du confort acoustique ressenti par les usagers en situation réelle.

Démarche suivie

La problématique de départ est avant tout celle de la qualité sonore des STA. L'étude de la littérature permet d'extraire un ensemble de concepts liés à la perception de ce type de sons (identification de la source, timbre, qualité sonore, ...), et les méthodologies pertinentes pour aborder ces concepts. Dans le cas des STA, il est rapidement apparu que leur qualité sonore ne pouvait se traduire que par une caractérisation hédonique du son, c'est-à-dire un jugement d'agrément, d'appréciation affective, associé à l'écoute du son. En effet, le son produit n'est qu'une conséquence inévitable du fonctionnement de l'appareil, et n'a pas de fonction ou de raison d'être particulière dans son environnement. Toutefois, la seule caractérisation des sources sonores n'autorise pas de représenter fidèlement le ressenti des usagers, et il est nécessaire de prendre en compte le contexte d'écoute de ces sons.

Ainsi, l'étude perceptive de la qualité des sources sonores a été réalisée à l'aide d'une approche psychophysique incluant, l'établissement d'une base de données d'enregistrements de STA, l'identification des principales familles de sons la constituant, la détermination de l'espace de timbre des STA, et l'évaluation et la prédiction des préférences des auditeurs.

Concernant la problématique du contexte d'écoute, nous nous sommes concentrés sur deux facteurs, que nous considérons parmi les plus importants quant à leur influence sur la perception du son de STA. Le premier est l'influence de l'acoustique des salles, et plus précisément du phénomène de réverbération qui caractérise les lieux habituels d'installation des STA. Il a alors été mis au point une stratégie permettant de comparer la perception de sons anéchoïques de STA avec celle de sons incluant, par simulation, la réverbération que ces mêmes sons provoqueraient dans une salle typique. Le second point d'intérêt concerne le contexte attentionnel des auditeurs. Afin d'évaluer la qualité sonore d'une manière plus représentative de l'aspect intrusif du son de STA vis-à-vis d'un activité donnée, nous avons mis au point une méthodologie originale de mesure perceptive du degré de gêne, et comparé les résultats avec ceux d'une situation plus classique d'écoute attentive.

Résultats obtenus

La première partie de ce travail a été de constituer un corpus sonore de travail pour l'étude de la perception des sons de STA. Il a donc été tout d'abord nécessaire d'établir une base de données sonores représentative, en effectuant un ensemble conséquent d'enregistrements, afin de prendre en compte une large variété de types et de tailles d'appareils. Plusieurs techniques d'enregistrement ont été utilisées (monophonique, stéréophonique et binaurale), bien que la suite du travail ait rapidement imposé l'utilisation des enregistrements monophoniques dans la majorité des cas. Parce que les analyses qui ont suivi exigeaient une taille de corpus exploitable (de 10 à 20 éléments), et afin de conserver la représentativité de la base initiale d'enregistrements, une expérience de catégorisation libre a été conduite dans le but d'identifier les principales familles de sons de STA. Á la lumière des résultats de cette expérience, il semble que les neuf familles de sons identifiées soient principalement gouvernées par le contenu spectral du son, ce qui s'explique aisément par la nature globalement stationnaire des sons (à l'exception particulière de certaines familles présentant des fluctuations plus ou moins prononcées). À l'aide de la mise au point d'un critère de sélection, correspondant initialement à la notion de *prototype*, il a alors été possible d'extraire un corpus réduit permettant de représenter fidèlement ces différentes familles de sons.

L'étape suivante du travail a été d'identifier les attributs auditifs pertinents pour la perception et l'évaluation de la qualité des sons de STA. Pour ce faire, une expérience de mesure de similarités a été réalisée, et l'analyse des résultats a révélé dans un premier temps un espace de timbre de deux dimensions. Il s'avère que ces deux dimensions représentent respectivement l'énergie acoustique dans les très basses fréquences, et la brillance du son, c'est-à-dire la répartition de l'ensemble de l'énergie acoustique sur l'échelle des fréquences. Par la suite, nous avons abordé la question de la qualité sonore à proprement parler, c'est-à-dire les préférences des auditeurs. L'étude de ces préférences à l'aide d'une expérience d'évaluation comparée nous a contraints à nous intéresser à une troisième dimension de l'espace de timbre qui s'est avérée primordiale pour expliquer les jugements de préférence. Ainsi, si cette troisième dimension, qui représente une forme d'émergence harmonique, semble moins importante pour la définition d'un son d'un point de vue purement descriptif, elle est grandement mise à contribution lorsqu'il s'agit d'en évaluer la qualité perçue. D'un point de vue plus pragmatique, cette partie du travail a permis de mettre au point une métrique, calculée sur le signal sonore, qui permet d'estimer la qualité sonore telle qu'elle est ressentie par les auditeurs.

Enfin, la dernière partie a consisté à aborder l'influence du contexte d'écoute sur la qualité sonore perçue. Plus précisément, deux éléments du contexte d'écoute ont été étudiés. Dans un premier temps, nous nous sommes intéressés à l'influence de la réverbération sur les jugements de préférence des auditeurs. L'idée était de comparer les résultats obtenus à l'aide d'échantillons anéchoïques avec ceux obtenus lorsque le son a subi l'effet de la réverbération. Pour ce faire, un outil d'auralisation, permettant de reproduire les aspects temporels et spatiaux de la réverbération provoquée dans deux types de salles représentatifs du contexte d'installation habituel des STA, a été utilisé. Les résultats obtenus semblent montrer que la réverbération a une influence sur la qualité perçue, certains stimuli présentant une qualité perçue très différente entre une condition anéchoïque et une condition auralisée. Toutefois, cette influence semble relativement ponctuelle, et ne modifie pas radicalement l'allure globale de l'échelle de qualité. Dans un second temps, la question du contexte attentionnel d'écoute a été abordée. Afin de refléter plus fidèlement les conditions d'écoute des STA, une procédure expérimentale spécifique recréant une condition d'*écoute* dite *distraite* a été mise au point et utilisée afin de détourner l'attention des participants du son. La comparaison des évaluations des sons obtenues à l'aide de cette procédure avec celles obtenues dans le cas d'une expérience en contexte

habituel d'*écoute attentive* a permis de mettre à jour, parmi le panel de participants, deux tendances différentes. La première, majoritaire, semble similaire aux résultats obtenus en condition d'écoute attentive, et la seconde est différente. Ceci tendrait à montrer que les résultats obtenus dans le cadre d'une procédure avec un contexte d'écoute attentive permettent d'obtenir des évaluations relativement pertinentes dans la majorité des cas, mais que, ponctuellement, l'attention des auditeurs peut influer de manière importante sur leur ressenti.

Perspectives

Le premier élément qui doit naturellement être abordé à la suite de ce travail est l'intégration des résultats obtenus ici, et notamment l'identification des attributs auditifs du timbre et la métrique de qualité sonore, dans le processus de conception des STA. Si la prise en compte du niveau sonore perçu, permettant d'obtenir une première approximation de la qualité sonore (notamment au travers de l'utilisation de l'échelle des dBA), est déjà couramment adoptée, les autres attributs identifiés ici doivent être intégrés. En particulier, il est nécessaire de s'intéresser aux éléments de la conception qui ont potentiellement une influence sur les descripteurs explicatifs mis en lumière pour la description du timbre. Le fait d'identifier ces éléments (qu'ils soient liés aux matériaux utilisés, au dimensionnement des pièces, ou à la conception plus générale des appareils) permettrait de mettre au point des solutions techniques afin de modifier les valeurs de ces descripteurs en accord avec la métrique de qualité. L'idée générale est donc de traduire les spécifications acoustiques que fournissent les résultats obtenus ici en spécifications techniques de conception des STA.

Un autre point, qui se place également dans le cadre industriel qui entoure ces travaux, doit être abordé. Ce point répond à une particularité du corpus de sons considérés déjà évoquée dans la discussion en section 4.3 qui suit la description des enregistrements effectués. En effet, un type particulier de STA qui a été pris en compte correspond à des systèmes « gainables », généralement installés dans les faux plafonds. Le contexte pratique de prise de son, et notamment l'utilisation d'une salle semi-anéchoïque, empêchaient de prendre en compte ce contexte spécifique d'installation. Dans une situation réaliste d'utilisation des STA gainables, la présence de circuits aérauliques de reprise et de diffusion d'air implique des régimes de fonctionnement différents pour ces appareils. Il est donc nécessaire d'évaluer à quel point ces éléments en modifient la perception sonore. Dans un premier temps, il conviendrait d'observer sur quelle étendue les descripteurs des dimensions de l'espace de timbre et la métrique de qualité sont modifiés si ces appareils sont enregistrés dans un contexte plus réaliste. Ces données permettraient de conclure quant à la validité des résultats obtenus dans ces travaux pour les STA gainables. Il est toutefois nécessaire de conserver à l'esprit qu'il est difficile de prendre en compte, lors de l'enregistrement, l'effet du contexte particulier d'installation de ces STA sans également faire intervenir d'autres paramètres, comme la réverbération par exemple.

D'un point de vue plus général, un autre élément nous semble mériter d'être abordé. Il concerne la portée des résultats obtenus en termes de métrique de la qualité sonore des STA. En effet, un problème majeur que nous avons rencontré est la forte variabilité des jugements de préférence des auditeurs. Cette forte variabilité, en s'accompagnant dans certains cas, et pour certains groupes de stimuli, d'une gamme de valeurs moyennes peu étendue, empêche d'extraire une hiérarchisation claire et nette des STA en termes de qualité sonore. Seuls les STA dont le son produit se distingue clairement des autres ont permis de générer une forme de consensus et des valeurs moyennes significativement différentes du reste du corpus. Ceci peut être dû au fait que les modèles considérés correspondent globalement à des STA actuellement présents sur le marché, et dont les sons produits sont déjà le ré-

sultat d'optimisations empiriques ayant pour but d'en améliorer la qualité. Par ailleurs, la très grande majorité d'entre eux proviennent du même fabricant, et résultent donc probablement d'une même stratégie d'amélioration du rendu sonore. Compte tenu de contraintes pratiques, notamment liées au partenariat avec un même partenaire industriel, il n'a pas été possible d'élargir la variété de modèles pris en compte. Il est possible que l'intégration de modèles bien différents de STA permettrait d'étendre la portée de l'échelle de qualité sonore obtenue.

Par ailleurs, les résultats obtenus dans le cadre des études présentées dans la partie III de ce document (chapitres 8 et 9) semblent indiquer que le contexte d'écoute n'est pas sans importance sur la perception des auditeurs. Cette observation semble rejoindre la problématique plus générale de la comparaison des études « in situ » et « en laboratoire ». Si les premières offrent plus de représentativité, elles ne permettent pas le contrôle de l'ensemble des paramètres à prendre en considération. Ce débat n'est pas nouveau, et les travaux réalisés ici n'avaient pas pour ambition de le résoudre. Il nous semble cependant que, bien qu'ils se placent résolument dans le cadre d'une étude « en laboratoire », la démarche adoptée ici représente une tentative d'aborder la problématique de l'évaluation de la qualité sonore d'une manière intermédiaire, et de prendre en considération les paramètres du contexte d'écoute. Néanmoins, les conclusions obtenues ici à ce sujet ne peuvent s'étendre qu'au cas des STA, et il serait intéressant d'appliquer cette méthodologie dans une certaine mesure à d'autres types de sons de l'environnement.

Dans le prolongement de la problématique évoquée au paragraphe précédent, il conviendrait également de tenter d'intégrer les éléments du contexte d'écoute abordés dans le cadre de cette thèse dans une méthodologie plus globale. Il s'agirait de considérer ces éléments dans le cadre d'une même expérience, afin de prendre en compte de possibles interactions dans l'effet des facteurs correspondants sur la perception des auditeurs. L'idée majeure serait d'aborder le contexte d'écoute dans les bureaux d'un point de vue plus général, en considérant les différents types de sources (son de STA, discussion dans les postes voisins, activité générale dans le cadre d'un bureau, ...) que l'on peut percevoir dans ce type d'environnement. Cette question rejoint notamment la problématique de l'intelligibilité de la parole, et son importance dans le contexte des bureaux paysagés, évoquées en section 3.2. Ainsi, dans une procédure expérimentale plus complexe et plus ambitieuse, dont le but serait d'évaluer le contexte de perception dans sa globalité, il serait intéressant de faire se croiser la problématique de l'effet du son de STA sur l'intelligibilité et celle de l'effet de la réverbération, aussi bien au travers de l'expression du ressenti des auditeurs que de leur performance dans une tâche donnée. Néanmoins, il semble difficile à ce stade de considérer réalisable d'un point de vue pratique ce type d'expérience, compte tenu la complexité de mise en place d'une telle procédure, qui nécessiterait une durée d'expérience importante et la mise au point de scénarii sophistiqués à reproduire par les auditeurs. De plus, compte tenu de la difficulté de générer des consensus, dans le cadre, plus simple, des expériences réalisées au cours de cette thèse, le caractère conclusif des résultats que l'on pourrait obtenir dans le cadre d'une telle expérience semble très incertain. Cette perspective, et la possibilité de l'aborder expérimentalement, nous semble néanmoins être une question qui mérite d'être considérée.

Bibliographie

[1] ANSI S1.1-1994 (R2004). Acoustical terminology. Acoustical Society of America, New York, NY, United States, 1994.

[2] ANSI S3.4-2007. American national standard procedure for the computation of loudness of steady sound. Acoustical Society of America, New York, NY, United States, 2007.

[3] ANSI S3.5-1969. American national standard methods for the calculation of the Articulation Index. Acoustical Society of America, New York, NY, United States, 1969.

[4] ANSI S3.5-1997 (R2007). Methods for calculation of the Speech Intelligibility Index. Acoustical Society of America, New York, NY, United States, 1997.

[5] IEC 60268-16 :2011(E). Sound system equipment – part 16 : Objective rating of speech intelligibility by speech transmission index. International Electrotechnical Commission, Geneva, Switzerland, 2011.

[6] ISO 11035. Sensory Analysis – Identification and selection of descriptors for establishing a sensory profile by a multidimensional approach. International Organization for Standardization, Geneva, 1994.

[7] ISO 226 :2003. Acoustics – Normal equal-loudness-level contours. International Organization for Standardization, Geneva, 1975.

[8] ISO 532. Acoustics – Methods for calculating loudness level. International Organization for Standardization, Geneva, 1975.

[9] ISO 8586. Sensory Analysis – General guidance for the selection, training and monitoring of assessors. International Organization for Standardization, Geneva, Switzerland,1993.

[10] B. W. Anderson and J. T. Kalb. English verification of the STI method for estimating speech intelligibility of a communications channel. *Journal of the Acoustical Society of America*, 81(6) :1982–1985, 1987.

[11] B. S. Atal, M. R. Schroeder, and G. M. Sessler. Subjective reverberation time and its relation to sound decay. In *Proceedings of the 5th International Congress on Acoustics (ICA)*, Liège, Belgium, 1965.

[12] The AURALIAS project. `http://www.auralias.be`.

[13] W. Aures. The sensory euphony as a function of auditory sensations. *Acustica*, 58(5) :282–290, 1985.

[14] J. A. Ballas. Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology : Human Perception and Performance*, 19(2) :250–267, 1993.

[15] M. Barron. The subjective effects of first reflections in concert halls : The need of lateral reflections. *Journal of Sound and Vibration*, 15(4) :475–494, 1971.

[16] M. Barron. *Auditorium acoustics and architectural design*. E & FN Spon, London, United Kingdom, 1993.

[17] L. Beranek. Concert hall acoustics–1992. *Journal of the Acoustical Society of America*, 92(1) :1–39, 1992.

[18] F. Bergeron, C. Astruc, A. Berry, and P. Masson. Sound quality assessment of internal automotive road noise using sensory science. *Acta Acustica united with Acustica*, 96(3) :580–588, 2010.

[19] B. Berglund, U. Berglund, and T. Lindvall. Scaling loudness, noisiness, and annoyance of aircraft noise. *Journal of the Acoustical Society of America*, 57(4) :930–934, 1975.

[20] B. Berglund, U. Berglund, and T. Lindvall. Scaling loudness, noisiness, and annoyance of community noise. *Journal of the Acoustical Society of America*, 60(5) :1119–1125, 1976.

[21] D. A. Bies and C. H. Hansen. *Engineering noise control : theory and practice*. E & FN Spon, London, United Kingdom, 1988.

[22] A. Billon and J. J. Embrechts. Objective study of spatial attributes in the room impulse response's late part and their relevance for auralization. In *Proceedings of Euronoise 2009*, Edinburgh, Scotland, 2009.

[23] A. Billon and J. J. Embrechts. Numerical evidence of mixing in rooms using the free path temporal distribution. *Journal of the Acoustical Society of America*, 130(3) :1381–1389, 2011.

[24] S. R. Bistafa and J. S. Bradley. Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics. *Journal of the Acoustical Society of America*, 107(2) :861–875, 2000.

[25] S. R. Bistafa and J. S. Bradley. Optimum acoustical conditions for speech intelligibility in classrooms. *Noise Notes*, 1(4) :11–17, 2009.

[26] J. Blauert. *Spatial Hearing : The Psychophysics of Human Sound Localization*. MIT Press, London, United Kingdom, 1997.

[27] J. Blauert and U. Jekosch. Sound-quality evaluation – A multi-layered problem. *Acta Acustica united with Acustica*, 83(5) :747–753, 1997.

[28] M. Bodden, R. Heinrichs, and A. Linow. Sound quality evaluation of interior vehicle noise using an efficient psychoacoustic method. In *Proceedings of Euronoise 98*, Munich, Germany, 1998.

[29] N. Bogaards, A. Röbel, and X. Rodet. Sound analysis and processing with Audiosculpt 2. In *Proceedings of International Computer Music Conference (ICMC)*, Miami, FL, USA, 2004.

[30] L. Bos and J. J. Embrechts. An interactive and real time 3d auralization system for room acoustics (the auralias project). In *3D Media*, Liège, Belgium, 2009.

[31] L. Bos and J. J. Embrechts. An interactive and real-time based auralization system for room acoustics, implementing directional impulse responses and multiple audio reproduction modules for spatialization (the auralias project). In *DAGA 2009*, Rotterdam, The Netherlands, 2009.

[32] J. S. Bradley. Uniform derivation of optimum conditions for speech in rooms. Technical Report BRN 239, National Research Council Canada, 1985.

[33] J. S. Bradley. Predictors of speech intelligibility in rooms. *Journal of the Acoustical Society of America*, 80(3) :837–845, 1986.

[34] J. S. Bradley. Speech intelligibility studies in classrooms. *Journal of the Acoustical Society of America*, 80(3) :846–854, 1986.

[35] J. S. Bradley. Relationships among measures of speech intelligibility in rooms. *Journal of the Audio Engineering Society*, 46(5) :396–405, 1998.

[36] J. S. Bradley. Designing and assessing speech privacy in open-plan offices. In *Proceedings of the 19th International Congress on Acoustics (ICA)*, Madrid, Spain, 2007.

[37] J. S. Bradley and B. N. Gover. Criteria for acoustic comfort in open-plan offices. In *Proceedings of Internoise 2004*, Prague, Czech Republic, 2004.

[38] J. S. Bradley and G. A. Soulodre. The influence of the late-arriving energy on spatial impression. *Journal of the Acoustical Society of America*, 97(4) :2263–2271, 1995.

[39] J. S. Bradley and G. A. Soulodre. Objective measures of listener envelopment. *Journal of the Acoustical Society of America*, 97(5) :2590–2597, 1995.

[40] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 39 :324–345, 1952.

[41] E. Brazil and M. Fernström. Where's that sound ? Exploring arbitrary user classification of sounds for audio management. In *Proceedings of the 2003 International Conference on Auditory Displays*, Boston, MA, USA, 2003.

[42] M.-C. Bézat. *Perception des bruits d'impact – Applicationau bruit de fermeture de porte automobile*. PhD thesis, Université de Provence – Aix-Marseille I, 2007.

[43] J. D. Carroll and J. J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3) :283–319, 1970.

[44] E. C. Cherry. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25(5) :975–979, 1953.

[45] E. C. Cherry and W. K. Taylor. Some further experiments upon the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 26(4) :554–559, 1954.

[46] P. Chevret and E. Parizet. An efficient alternative to the paired comparison method for the subjective evaluation of a large set of sounds. In *Proceedings of the 19th International Congress on Acoustics (ICA)*, Madrid, Spain, 2007.

[47] M. Chion. *Guide des objets sonores : Pierre Schaeffer et la recherche musicale*. Buchet/Chastel, Paris, France, 1983.

[48] P. Daniel and R. Weber. Psychoacoustical roughness : Implementation of an optimized model. *Acustica united with Acta Acustica*, 83(1) :113–123, 1997.

[49] H. A. David. *The method of paired comparison*. Oxford University Press, New York, NY, United States, 1988.

[50] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley-Interscience, New York, NY, United States, 3rd edition, 1998.

[51] N. I. Durlach and H. S. Colburn. *Handbook of Perception – Volume IV : Hearing*, chapter Binaural Phenomena. Academic Press, 1978.

[52] A. Ebissou, P. Chevret, and E. Parizet. Objective and subjective assessment of disturbance by office noise - relevance of the use of the speech transmission index. In *Proceedings of Acoustics 2012*, pages 2485–2490, Nantes, France, 2012.

[53] M. D. Egan. *Concepts in architectural acoustics*. McGraw-Hill, New York, NY, United States, 1972.

[54] W. Ellermeier, M. Mader, and P. Daniel. Scaling the unpleasantness of sounds according to the BTL model : Ratio-scale representation and psychoacoustical analysis. *Acta Acustica united with Acustica*, 90(1) :101–107, 2004.

[55] W. Ellermeier, A. Zeitler, and H. Fastl. Predicting annoyance judgments from psychoacoustic metrics : Identifiable versus neutralized sounds. In *Proceedings of Internoise 2004*, Prague, Czech Republic, 2004.

[56] D. P. W. Ellis. Sinewave and sinusoid + noise analysis/synthesis in Matlab. `http://labrosa.ee.columbia.edu/matlab/sinemodel/`.

[57] J. J. Embrechts. Sound field distribution using randomly traced sound ray techniques. *Acustica*, 51(6) :288–295, 1982.

[58] J. J. Embrechts. Broad spectrum diffusion model for room acoustics ray-tracing algorithms. *Journal of the Acoustical Society of America*, 107(4) :2068–2081, 2000.

[59] H. Fastl. Sound quality of electric razors – Effects of loudness. In *Proceedings of Internoise 2000*, Nice, France, 2000.

[60] H. Fastl. Neutralizing the meaning of sound for sound quality evaluations. In *Proceedings of the 18th International Congress on Acoustics (ICA)*, Kyoto, Japan, 2001.

[61] H. Fastl. *Communication acoustics*, chapter Psychoacoustics and sound quality, pages 139–162. Springer, New York, NY, United States, 2005.

[62] G. Fechner. *Readings in the history of psychology. Century psychology series*, chapter Elements of psychophysics, pages 206–213. Appleton-Century-Crofts, East Norwalk, CT, United States, 1860.

[63] C. Fellbaum. *Encyclopedia of Language and Linguistics*, chapter WordNet(s), pages 665–670. Elsevier, Oxford, United Kingdom, 2006.

[64] T. Finitzo-Hieber and T. W. Tillman. Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. *Journal of Speech and Hearing Research*, 21(3) :440–458, 1978.

[65] K. Fujii, J. Atagi, and Y. Ando. Temporal and spatial factors of traffic noise and its annoyance. *Journal of Temporal Design in Architecture and the Environment*, 2(1) :33–41, 2002.

[66] K. Fujii, T. Hotehama, K. Kato, R. Shimokura, Y. Okamoto, and Y. Ando. Spatial distribution in concert halls : Comparison of different scattered reflexions. *Journal of Temporal Design in Architecture and the Environment*, 4(1) :59–68, 2004.

[67] W. W. Gaver. How do we hear in the world ? Explorations in ecological acoustics. *Ecological Psychology*, 5(4) :285–313, 1993.

[68] W. W. Gaver. *Handbook of human-computer interaction*, chapter Auditory interfaces, pages 1003–1041. Elsevier, Amsterdam, The Netherlands, 1997.

[69] E. Geissner and E. Parizet. Continuous assessment of the unpleasantness of a sound. *Acta Acustica united with Acustica*, 93(3) :469–476, 2007.

[70] J. M. Grey. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5) :1270–1277, 1977.

[71] L. Gros, N. Chateau, and S. Busson. Effects of context on the subjective assessment of time-varying speech quality : Listening / conversation, laboratory / real environment. *Acta Acustica united with Acustica*, 90(6) :1037–1051, 2004.

[72] R. Guski. Psychological methods for evaluating sound quality and assessing acoustic information. *Acta Acustica united with Acustica*, 83(5) :765–774, 1997.

[73] F. Guyot. *Etude de la perception sonore en termes de reconnaissance et d'appréciation qualitative : Une approche par la catégorisation.* PhD thesis, Université du Maine, 1996.

[74] T. Hanyu and S. Kimura. A new objective measure for evaluation of listener envelopment focusing on the spatial balance of reflections. *Applied Acoustics*, 62(2) :155–184, 2001.

[75] T. Hidaka, L. Beranek, and T. Okano. Interaural cross correlation, lateral fraction, and low and high-frequency sound levels as measure of acoustical quality in concert halls. *Journal of the Acoustical Society of America*, 98(2) :988–1007, 1995.

[76] M. Hodgson. Experimental investigation of the acoustical characteristics of university classrooms. *Journal of the Acoustical Society of America*, 106(4) :1810–1819, 1999.

[77] M. Hodgson. Effet of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms. *Journal of the Acoustical Society of America*, 111(2) :931–939, 2002.

[78] V. Hongisto. A model predicting the effect of speech of varying intelligibility on work performance. *Indoor Air*, 15(6) :458–468, 2005.

[79] T. Houtgast and H. J. M. Steeneken. A physical method for measuring speech transmission quality. *Journal of the Acoustical Society of America*, 67(1) :318–326, 1980.

[80] T. Houtgast and H. J. M. Steeneken. A multi-language evaluation of the rasti method for estimating speech intelligibility in auditoria. *Acta Acustica united with Acustica*, 54(4) :185–199, February 1984.

[81] J. H. Howard and J. A. Ballas. Syntactic and semantic factors in the classification of nonspeech transient patterns. *Perception and Psychophysics*, 28(5) :431–439, 1980.

[82] D. C. Howell. *Statistics methods for psychology*. Wadsworth, Cengage learning, Stamford, CT, United States, 7th edition, 1992.

[83] R. Hétu, C. Truchon-Gagnon, and S. Bilodeau. Problems of noise in school settings : a review of the literature and the results of an exploratory study. *Journal of Speech-Language Pathology and Audiology*, 14(3) :31–39, 1990.

[84] IBM SPSS Statistics. `http://www-01.ibm.com/software/analytics/spss/products/statistics/`.

[85] J. G. Ih, D. H. Lim, S. H. Shin, and Y. Park. Experimental design and assessment of product sound quality : Application to a vacuum cleaner. *Noise Control Engineering Journal*, 51(4) :244–252, 2003.

[86] W. H. Ittelson. Environmental perception and urban experience. *Environment and Behavior*, 10(2) :193–213, 1978.

[87] V. L. Jordan. Acoustical criteria for auditoriums and their relation to model techniques. *Journal of the Acoustical Society of America*, 47(2A) :408–412, 1970.

[88] W. B. Joyce. Sabine's reverberation time and ergodic auditoriums. *Journal of the Acoustical Society of America*, 58(3) :643–655, 1975.

[89] F. Junker, P. Susini, and P. Cellard. Sensory evaluation of air-conditioning noise : Comparative analysis of two methods. In *Proceedings of the 17th International Congress on Acoustics (ICA)*, pages 342–343, Rome, Italy, 2001.

[90] M. S. Khan. Effects of masking sound on train passenger aboard activities and on other interior annoying noises. *Acta Acustica united with Acustica*, 89(4) :711–717, 2003.

[91] M. S. Khan and C. Dickson. Evaluation of sound quality of wheel loaders using a human subject for binaural recording. *Noise Control Engineering Journal*, 50(4) :117–126, 2002.

[92] C. L. Krumhansl. *Structure and perception of electroacoustic sound and music*, chapter Why is musical timbre so hard to understand ?, pages 43–53. Elsevier (Excerpta Medica 846), Amsterdam, The Netherlands, 1989.

[93] J. B. Kruskal and M. Wish. *Multidimensional scaling*. Quantitative Application in the Social Sciences. Sage Publications, Beverly Hills, CA, United States, 1977.

[94] H. Kuttruff. Auralisation of impulse responses modeled on the basis of ray-tracing results. *Journal of the Audio Engineering Society*, 41(11) :876–880, 1993.

[95] H. Kuttruff. *Room acoustics*. Spon Press, London, United Kingdom, 4th edition, 2000.

[96] S. Kuwano, S. Namba, and H. Miura. Asvantages and disadvantages of A-weighted sounds pressure level in relation to subjective impression of environmental noises. *Noise Control Engineering Journal*, 33(3) :107–115, 1989.

[97] S. Kuwano, S. Namba, A. Schick, H. Hoege, H. Fastl, T. Filippou, M. Florentine, and H. Muesch. The timbre and annoyance of auditory warning signals in different countries. In *Proceedings of Internoise 2000*, Nice, France, 2000.

[98] P. Legendre and L. Legendre. *Numerical ecology. Development in environmental modelling*. Elsevier, Amsterdam, The Netherlands, 2nd edition, 1998.

[99] G. Lemaitre. *Étude perceptive et acoustique de nouveaux avertisseurs sonores automobiles*. PhD thesis, Université du Maine, 2004.

[100] G. Lemaitre, P. Susini, S. Winsberg, S. McAdams, and B. Letinturier. The sound quality of car horns : A psychoacoustical study of timbre. *Acta Acustica united with Acustica*, 93(3) :457–468, 2007.

[101] G. Lemaitre, P. Susini, S. Winsberg, S. McAdams, and B. Letinturier. The sound quality of car horns : Designing new representative sounds. *Acta Acustica united with Acustica*, 95(2) :356–372, 2009.

[102] T. Lindvall and T. P. Radford. Measurement of annoyance due to exposure to environmental factors. *Environmental Research*, 6(1) :1–36, 1973.

[103] R. D. Luce. *Individual choice behavior : A theoretical analysis*. Wiley, New York, NY, United States, 1959.

[104] J. Marozeau, A. de Cheveigne, S. McAdams, and S. Winsberg. The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, 114(5) :2946–2957, 2003.

[105] R. Maunder. An interactive subjective assessment method for recorded sound. In *Proceedings of European Conference on Vehicle Noise and Vibration*, pages 345–354, London, United Kingdom, 1998. Institute of Mechanical Engineers (IMechE).

[106] S. McAdams. *Thinking in sound : The cognitive psychology of human audition*, chapter Recognition of auditory sound sources and events, pages 157–213. Oxford University Press, Oxford, United Kingdom, 1993.

[107] S. McAdams, P. Susini, N. Misdariis, and S. Winsberg. Multidimensional characterisation of perceptual and preference judgements of vehicle and environmental noises. In *Proceedings of Euronoise 98*, Munich, Germany, 1998.

[108] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. Perceptual scaling of synthesized musical timbres : Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58 :177–192, 1995.

[109] M. C. Meilgaard, B. T. Carr, and G. V. Civille. *Sensory evaluation techniques*. CRC Press, New York, NY, United States, 4th edition, 1991.

[110] J. R. Miller and C. Carterette. Perceptual space for musical structures. *Journal of the Acoustical Society of America*, 58 :711–720, 1975.

[111] N. Misdariis, A. Minard, P. Susini, G. Lemaitre, S. McAdams, and E. Parizet. Environmental sound perception : meta-description and modeling based on independent primary studies. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010. `http://www.hindawi.com/journals/asmp/2010/362013.html`.

[112] B. C. J. Moore, B. G. Glasberg, and T. Baer. A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4) :224–240, 1997.

[113] A. K. Nábělek and J. M. Pickett. Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners. *Journal of Speech and Hearing Research*, 17(4) :724–739, 1974.

[114] A. K. Nábělek and P. K. Robinson. Monaural and binaural speech perception in reverberation for listeners of various ages. *Journal of the Acoustical Society of America*, 71(5) :1242–1248, 1982.

[115] P. Navi. *Propriétés acoustiques des matériaux. Propagation des ondes planes harmoniques*. Presses polytechniques et universitaires romandes, Lausanne, Switzerland, 2006.

[116] T. Nielsen, T. V Nielsen, P. Johansen, J. M. Hasenkam, and H. Nygaard. Psychoacoustic quantification of mechanical heart valve noise. *Journal of Heart Valve Disease*, 14(1) :89–95, 2005.

[117] T. Okano, L. Beranek, and T. Hidaka. Relations among interaural cross-correlation coefficient ($IACC_E$), lateral energy fraction (LF_E) and apparent source width (ASW) in concert halls. *Journal of the Acoustical Society of America*, 104(1) :255–265, 1998.

[118] E. Parizet, E. Guyader, and V. Nosulenko. Analysis of car door closing sound quality. *Applied Acoustics*, 69(1) :12–22, 2006.

[119] E. Parizet, N. Hamzaoui, and G. Sabatié. Comparison of some listening test methods : A case study. *Acta Acustica united with Acustica*, 91(2) :356–364, 2005.

[120] H. Pashler. *Attention*. Taylor & Francis Press, Philadelphia, PA, United States, 1998.

[121] H. Pashler. *The psychology of attention*. MIT Press, Cambridge, MA, United States, 1998.

[122] C. Patsouras, H. Fastl, D. Patsouras, and K. Pfaffelhuber. Psychoacoustic sensation magnitudes and sound quality ratings of upper middle class cars' idling noise. In *Proceedings of the International Conference on Acoustics (ICA)*, Rome, Italy, 2001.

[123] C. Patsouras, H. Fastl, D. Patsouras, and K. Pfaffelhuber. Subjective evaluation of loudness reduction and sound quality ratings obtained with simulations of acoustic materials for noise control. In *Proceedings of Euronoise 2001*, Patras, Greece, 2001.

[124] C. Patsouras, H. Fastl, D. Patsouras, and K. Pfaffelhuber. How far is the quality of a diesel powered car away from that of a gasoline one ? In *Proceedings of the Forum Acusticum*, Sevilla, Spain, 2002.

[125] R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand. *Auditory Physiology and Perception*, chapter Complex sounds and auditory images, pages 429–446. Pergamon Press, Oxford, United Kingdom, 1992.

[126] R. Paulsen. On the influence of the stimulus duration on psychophysical judgement of environmental noises takin in the laboratory. In *Proceedings of Internoise 1997*, volume 3, Budapest, Hungary, 1997.

[127] K. S. Pearsons, R. L. Bennett, and S. Fidell. Speech levels in various noise environments. Technical Report EPA-600/1-77-025, U.S. Environmental Protection Agency,, 1977.

[128] R. Plomp, H. J. M. Steeneken, and T. Houtgast. Predicting speech intelligibility in rooms from the modulation transfer function ii. mirror image computer model applied to rectangular rooms. *Acustica*, 46 :73–81, 1980.

[129] L. Rabiner and B. Juang. *Fundamentals of speech recognition*. Prentice Hall, Upper Saddle River, NJ, United States, 1993.

[130] D. W. Robinson and R. S. Dadson. A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7(5) :166–181, 1956.

[131] E. Rosch. *Cognition and categorization*, chapter Principles of categorization, pages 27–48. Lawrence Erlbaum Associates, Hillsdale, NJ, United States, 1978.

[132] S. Sato and Y. Ando. On the apparent source width (ASW) for bandpass noises related to the IACC and the width of the interaural cross-correlation function (W_{IACC}). In *Proceedings of the 137th ASA / 2nd EAA / 25th DAGA*, Berlin, Germany, 1998.

[133] S. Sattah and A. Tversky. Additive similarity trees. *Psychometrika*, 42(3) :319–345, 1977.

[134] P. Schaeffer. *Traité des objets musicaux*. Seuil, Paris, France, 1966.

[135] M. R. Schroeder and K. H. Kuttruff. On frequency response curves in rooms. comparison of experimental, theoretical, and monte carlo results for the average frequency spacing between maxima. *Journal of the Acoustical Society of America*, 34(1) :76–80, 1962.

[136] A. Sehr, E. Habets, R. Maas, and W. Kellermann. Towards a better understanding of the effect of reverberation on speech recognition performance. In *Proceedings of IEEE International Workshop on Acoustic Echo and Noise Control*, Tel Aviv, Israel, 2010.

[137] E. Siekierski, C. Derquenne, and N. Martin. Sensory evaluation of air-conditioning noise : Sensory profiles and hedonic tests. In *Proceedings of the 17th International Congress on Acoustics (ICA)*, pages 170–171, Rome, Italy, 2001.

[138] M. Slaney. Auditory toolbox, version 2. Technical Report 1998-010, Interval Research Corporation, 1998. `http://cobweb.ecn.pursue.edu/~malcolm/interval/1998-010/`.

[139] G. De Soete and S. Winsberg. A thurstonian pairwise choice model with univariate and multivariate spline transformation. *Psychometrika*, 58(2) :233–256, 1993.

[140] L. N. Solomon. Semantic approach to the perception of complex sounds. *Journal of the Acoustical Society of America*, 30(5) :421–425, 1958.

[141] G. Soulodre. Can reproduced sound be evaluated using measures designed for concert halls ? In *Proceeding of Spatial Audio and Sensory Evaluation Techniques*, Guilford, United Kingdom, 2006.

[142] S. S. Stevens. On the psychophysical law. *Psychological Review*, 64(3) :153–181, 1957.

[143] C. Suied, P. Susini, and S. McAdams. Evaluating warning sound urgency with reaction times. *Journal of Experimental Psychology : Applied*, 14(3) :201–212, 2008.

[144] P. Susini and S. McAdams. Influence of sound-directed attentional focus on overall loudness ratings. In *Proceedings of EAA Congress*, Sevilla, Spain, 2002.

[145] P. Susini, S. McAdams, and S. Winsberg. Caractérisation perceptive des bruits de véhicules. In *Proceedings of 4ème Congrès Français d'Acoustique*, Marseille, France, 1997.

[146] P. Susini, S. McAdams, and S. Winsberg. A multidimensional technique for sound quality assessment. *Acta Acustica united with Acustica*, 85(5) :650–656, 1999.

[147] P. Susini, S. McAdams, S. Winsberg, I. Perry, S. Vieillard, and X. Rodet. Characterizing the sound quality of air conditioning noise. *Applied Acoustics*, 65(8) :763–790, 2004.

[148] E. Terhardt. Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, 55(5) :1061–1069, 1974.

[149] E. Terhardt and G. Stoll. Skalierung des Wohlklangs on 17 Umweltschallen und Untersuchung des beteiligten Hörparameter. *Acustica*, 48(4) :247–253, 1981.

[150] L. L. Thurstone. A law of comparative judgement. *Psychological Review*, 34(4) :273–286, 1927.

[151] W. S. Togerson. *Theory and methods of scaling*. Wiley, New York, NY, United States, 1958.

[152] S. Tremblay and D. M. Jones. Change of intensity fails to produce an irrelevant sound effect : implications for the representation of unattended sound. *Journal of Experimental Psychology : Human perception and performance*, 25(4) :1005–1015, 1999.

[153] C. Truchon-Gagnon and R. Hétu. Noise in day-care centres for children. *Noise Control Engineering Journal*, 30(2) :57–64, 1988.

[154] L. R. Tucker. A method for the synthesis of factor analysis studies. Technical Report Personnel Research Section Report No. 984, Washington : Department of the Army., 1951.

[155] N. Vanderveer. *Ecological acoustics : Human perception of environmental sounds*. PhD thesis, Cornell University, 1979.

[156] G. von Bismarck. Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30(3) :159–172, 1974.

[157] M. Vorländer. *Auralization*. Springer-Verlag, Berlin, Germany, 2008.

[158] E. H. Weber. *Handwörterbuch der Physiologie*, chapter Tastsinn und Gemeingefühl, pages 481–588. W. Engelmann, Leipzig, Germany, 1846.

[159] R. Weber. The continuous loudness judgement of temporally variable sounds with an "analog" category procedure. In *Proceedings of the 5th Oldenburg Symposium on Psychological Acoustics*, Oldenburg, Germany, 1991.

[160] U. Widmann. A psychoacoustic annoyance concept for application in sound quality. *Journal of the Acoustical Society of America*, 100(5) :3078, 1997.

[161] S. Winsberg and J. D. Carroll. A quasi non-metric method for multidimensional scaling via an extended Euclidian model. *Psychometrika*, 54(2) :217–229, 1989.

[162] S. Winsberg and G. De Soete. A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika*, 58(2) :315–330, 1993.

[163] K. Zimmer, W. Ellermeier, and C. Schmid. Using probalistic choice models to investigate auditory unpleasantness. *Acta Acustica united with Acustica*, 90(6) :1019–1028, 2004.

[164] E. Zwicker. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *Journal of the Acoustical Society of America*, 33(2) :248–248, 1961.

[165] E. Zwicker. Procedure for calculating loudness of temporally variable sounds. *Journal of the Acoustical Society of America*, 58(5) :282–290, 1977.

[166] E. Zwicker. Meaningful noise measurement and effective noise reduction. *Noise Control Engineering Journal*, 29(3) :66–76, 1987.

[167] E. Zwicker. A proposal for defining and calculating the unbiaised annoyance. In *Proceedings of the 5th Oldenburg Symposium on Psychological Acoustics*, pages 187–202, Oldenburg, Germany, 1991.

[168] E. Zwicker and H. Fastl. *Psychoacoustics : Facts and models*. Springer, New York, NY, United States, 1990.

Annexe A

Techniques statistiques d'analyse multidimensionnelle

L'analyse multidimensionnelle – *MultiDimensional Scaling (MDS)* [151, 93] – a pour but de réduire raisonnablement la dimensionnalité d'un espace de variables. En pratique, on dispose d'un ensemble de variables qui ne sont pas nécessairement indépendantes. En effet, elles peuvent présenter une certaine redondance, comme c'est le cas lorsque l'on considère par exemple un ensemble de descripteurs acoustiques, qui peuvent inclure plusieurs déclinaisons d'un attribut donné (usage ou non d'un modèle perceptif, d'une courbe de pondération fréquentielle, etc....). La MDS vise donc à obtenir un espace de représentation continu, orthogonal et à dimensionnalité réduite (souvent 2 ou 3 dimensions). Les données de départ sont sous la forme d'une matrice de dissemblances, assimilées à des distances perceptives, où figure la dissemblance mesurée entre chaque paire d'éléments (de sons en l'occurrence). Le résultat attendu correspond aux coordonnées des éléments dans l'espace identifié. Ceci est réalisé en modélisant les dissemblances par une distance euclidienne ou pseudo-euclidienne, et en minimisant l'erreur moyenne entre les distances prédites et les distances mesurées. Plusieurs modèles de distance existent. En voici les principaux exemples employés dans la littérature sur l'étude du timbre.

A.1 Modèle simple – MDSCAL

Dans le modèle MDSCAL, la distance perceptive entre les sons i et j, obtenue en moyennant les jugements de tous les participants, est représentée par une distance euclidienne $\widehat{d}_{ij}$ dans un espace cartésien à R dimensions :

$$\widehat{d}_{ij} = \sqrt{\sum_{r=1}^{R} \left(x_{ir} - x_{jr}\right)^2} \tag{A.1}$$

avec x_{ir} et x_{jr} les coordonnées des sons i et j sur la dimension r.

Toutefois, cette technique présuppose que les sons sont jugés et discriminés sur un ensemble exhaustif de dimensions continues communes à tous les participants, et partagées par l'ensemble des sons. Il est difficile d'affirmer avec certitude que cette hypothèse est valide dans tous les cas. Il est notamment possible que les participants n'associent pas à une même dimension la même importance

pour établir leurs jugements. Il en résulte une propriété particulière : l'invariance rotationnelle de l'espace obtenu par le modèle MDSCAL, ce qui signifie que le fait de faire pivoter les axes de l'espace dans quelque sens que ce soit ne changera pas la structure intrinsèque de l'espace tant que les axes restent orthogonaux. Ceci rend l'interprétation des axes difficile.

A.2 Modèle pondéré – INDSCAL

Afin de faciliter l'interprétation des axes, un raffinement particulier a été apporté au modèle de base dans la variante INDSCAL – *individual differences scaling model* – proposée par Carroll et Chang [43]. Cette variante prend en compte la possibilité que les participants attribuent une importance différente aux dimensions de l'espace, que le modèle traduit par la pondération des axes. La distance modélisée $\widehat{d}_{ijn}$ entre les sons i et j pour le participant n s'exprime alors par :

$$\widehat{d}_{ijn} = \sqrt{\sum_{r=1}^{R} w_{nr} \cdot \left(x_{ir} - x_{jr}\right)^2} \tag{A.2}$$

avec w_{nr} le poids donné à la dimension r par le participant n.

La présence de ces poids de dimension différents empêche l'invariance rotationnelle de l'espace obtenu, car les axes sont fixés par la présence de ces poids. Il est supposé dans ce modèle que les dimensions obtenues sont perceptivement pertinentes, ce qui en facilite l'interprétation.

A.3 Modèle à spécificités – EXSCAL

Un autre modèle de MDS permet de prendre en compte la possibilité que les sons soient également discriminés sur la base de traits spécifiques, appelés *spécificités*, qui ne sons pas partagés par l'ensemble des sons : le modèle EXSCAL – *extended two-way euclidian model with common and specific dimensions* – développé par Winsberg et Carroll [161]. La distance modélisée $\widehat{d}_{ij}$ entre les sons i et j s'exprime alors par :

$$\widehat{d}_{ij} = \sqrt{\sum_{r=1}^{R} \left(x_{ir} - x_{jr}\right)^2 + s_i + s_j} \tag{A.3}$$

où s_i et s_j peuvent être vus comme les degrés de spécificité respectifs des sons i et j.

A.4 Modèle à classes latentes – CLASCAL

Une autre variante notable de modèle MDS allie les avantages des modèles INDSCAL et EXSCAL, c'est-à-dire l'usage, d'une part, de pondérations des dimensions différentes entre les participants et, d'autre part, de spécificités modélisant les éventuels traits caractéristiques propres à chaque son. Il s'agit du modèle CLASCAL développé par Winsberg et De Soete [162]. Ce modèle adopte ce que l'on

nomme l'*approche en classes latentes*. Cette approche a pour but d'observer les différentes stratégies naturellement employées par les participants au travers de la pondération des dimensions. Ces pondérations ne sont donc plus individuelles, et les participants sont regroupés en *classes latentes* de participants présentant des pondérations communes. Cette approche permet une modélisation plus réaliste du concept de « stratégie de jugement » souvent adoptée par plusieurs participants à la fois. La distance modélisée $\widehat{d}_{ijt}$ entre les sons i et j pour la classe latente t s'exprime alors par :

$$\widehat{d}_{ijt} = \sqrt{\sum_{r=1}^{R} w_{tr} \cdot \left(x_{ir} - x_{jr}\right)^2 + v_t \cdot (s_i + s_j)} \tag{A.4}$$

avec w_{tr} le poids donné à la dimension r pour la classe latente t

et v_t le poids donné aux spécificités pour la classe latente t

Il convient de noter que le modèle CLASCAL empêche également l'invariance rotationnelle de l'espace obtenu, à cause de la présence de pondérations des axes différentes d'une classe à l'autre.

Annexe B

Corrélation et régression linéaire

La corrélation a pour but d'évaluer le degré de lien existant entre deux variables X et Y, tandis que la régression linéaire a pour but de prédire les valeurs Y à partir de celles de X selon une relation affine du type $Y = aX + b$. Les deux techniques sont donc fortement liées, puisqu'un haut degré de lien permet de justifier la prédiction d'une variable par l'autre, et elles sont souvent mentionnée conjointement dans la littérature [1].

B.1 Coefficient de corrélation de Bravais-Pearson

La corrélation s'évalue grâce au coefficient de corrélation de Bravais-Pearson. Celui-ci est lié au calcul de la covariance cov_{XY} des deux variables X et Y à N observations :

$$cov_{XY} = \frac{1}{N-1} \sum (X - \bar{X})(Y - \bar{Y}) \tag{B.1}$$

Le coefficient de corrélation de Bravais-Pearson r est alors obtenu en divisant la covariance par le produit des écarts-types s_X et s_Y des variables X et Y :

$$r = \frac{cov_{XY}}{s_X s_Y} \tag{B.2}$$

Le résultat de ce calcul est borné entre -1 et 1. Ces deux valeurs particulières correspondent à une parfaite corrélation entre les deux variables, le signe n'indiquant que le sens de la corrélation. Plus généralement, le lien entre les variables est d'autant plus manifeste que le coefficient r est proche de l'une de ces deux valeurs extrêmes, et d'autant plus contestable qu'il est proche de 0.

Ce coefficient est toutefois considéré comme un estimateur biaisé de la corrélation qu'il peut exister entre les deux variables. En effet, il prend en compte un nombre fini N d'observations des variables, qui n'est pas nécessairement suffisamment grand pour rendre l'échantillonnage des variables représentatif de leurs distributions réelles. Pour plus de rigueur, un ajustement doit être apporté à ce calcul, afin d'obtenir un estimateur non-biaisé r' de la corrélation :

$$r' = \sqrt{1 - \frac{(1 - r^2)(N-1)}{N-2}} \tag{B.3}$$

1. Certains statisticiens font toutefois une distinction plus radicale des deux techniques (voir [82]).

Ce calcul modifie la valeur du coefficient de corrélation, qui est d'autant plus réduite que N est faible. Toutefois, la corrélation est tout de même souvent évaluée à l'aide du coefficient r, sans pour autant négliger le biais lié à la taille de l'échantillonnage. Ce dernier peut en effet être pris en compte d'une autre manière pour tempérer l'interprétation d'un coefficient de corrélation proche de 1 ou de -1.

Dans un cadre statistique rigoureux, il convient de fixer un seuil du coefficient de corrélation à partir duquel on peut considérer que la valeur obtenue est significative, c'est-à-dire qu'il y a un lien manifeste entre les deux variables. Ceci rentre dans le cadre du test d'hypothèse. En l'occurrence, on cherche à savoir si l'on peut raisonnablement rejeter l'hypothèse nulle qui stipule que les deux variables sont linéairement indépendantes. Cette opération est réalisable en comparant le coefficient de corrélation calculé r aux valeurs critiques du coefficient de corrélation de Bravais-Pearson que l'on peut lire dans le tableau B.1.

$ddl = N-2$	α			$ddl = N-2$	α		
	0,05	0,02	0,01		0,05	0,02	0,01
1	1,00	1,00	1,00	21	0,41	0,48	0,53
2	0,95	0,98	0,99	22	0,40	0,47	0,52
3	0,88	0,93	0,96	23	0,40	0,46	0,51
4	0,81	0,88	0,92	24	0,39	0,45	0,50
5	0,75	0,83	0,87	25	0,38	0,45	0,49
6	0,71	0,79	0,83	26	0,38	0,44	0,48
7	0,67	0,75	0,80	27	0,37	0,43	0,47
8	0,63	0,72	0,77	28	0,36	0,42	0,46
9	0,60	0,69	0,74	29	0,35	0,42	0,46
10	0,58	0,66	0,71	30	0,35	0,41	0,45
11	0,55	0,63	0,68	35	0,32	0,38	0,42
12	0,53	0,61	0,66	40	0,30	0,36	0,39
13	0,51	0,59	0,64	45	0,29	0,34	0,37
14	0,50	0,57	0,62	50	0,27	0,32	0,35
15	0,48	0,56	0,61	60	0,25	0,30	0,33
16	0,47	0,54	0,59	70	0,23	0,27	0,30
17	0,46	0,53	0,58	80	0,22	0,26	0,28
18	0,44	0,52	0,56	100	0,20	0,23	0,25
19	0,43	0,50	0,55	150	0,16	0,19	0,21
20	0,42	0,49	0,54	200	0,14	0,16	0,18

TABLE B.1 – Valeurs critiques du coefficient de corrélation de Bravais-Pearson.

Le nombre de degrés de liberté ddl se calcule simplement par $N-2$. Ensuite, il convient de fixer le seuil α (dont l'interprétation est expliquée dans l'exemple suivant) que l'on souhaite atteindre. Si la valeur absolue du coefficient de corrélation obtenu $|r|$ est supérieure à la valeur indiquée dans le tableau au seuil souhaité, alors on peut raisonnablement rejeter l'hypothèse nulle, c'est-à dire que les deux variables sont indépendantes. Par exemple, prenons un coefficient $r = 0,75$ pour 12 observations. Le nombre de degrés de liberté est donc $ddl = 10$, et si l'on souhaite un seuil $\alpha = 0,01$, on peut lire dans le tableau la valeur critique $0,71$. Notre valeur du coefficient de corrélation étant supérieure, nous avons donc moins de 1 % ($\alpha = 0,01$) de chances de nous tromper si nous rejetons l'hypothèse nulle, c'est-à-dire si nous affirmons que les variables ne sont pas indépendantes. En pratique, cette condition est souvent notée $p < \alpha$, donc dans l'exemple précédent $p < 0,01$.

B.2 Régression linéaire

Le coefficient de corrélation de Bravais-Pearson nous donne un moyen d'établir un lien statistique entre deux variables. Toutefois, son interprétation reste quelque peu abstraite. La *régression linéaire* est un outil statistique dont l'interprétation graphique permet de se faire une meilleure idée de ce que représente ce coefficient de corrélation. Tout d'abord, il convient de tracer ce que l'on nomme le *diagramme de dispersion*, qui consiste à représenter dans un repère cartésien les points dont les coordonnées correspondent aux valeurs des variables X et Y. La figure B.1 en montre un exemple. Sur ce graphique, on peut également tracer la droite de régression (en rouge sur la figure B.1). Elle est établie en tentant de minimiser les distances la séparant des points du diagramme. On cherche donc à modéliser l'éventuelle relation liant X et Y par une fonction affine, c'est-à-dire à estimer les valeurs de Y par la relation : $\widehat{Y} = aX + b$. Il s'agit donc de minimiser l'erreur de prédiction qui est la somme quadratique des écarts entre Y et son estimation $\widehat{Y}$: $\sum(\widehat{Y} - Y)^2$.

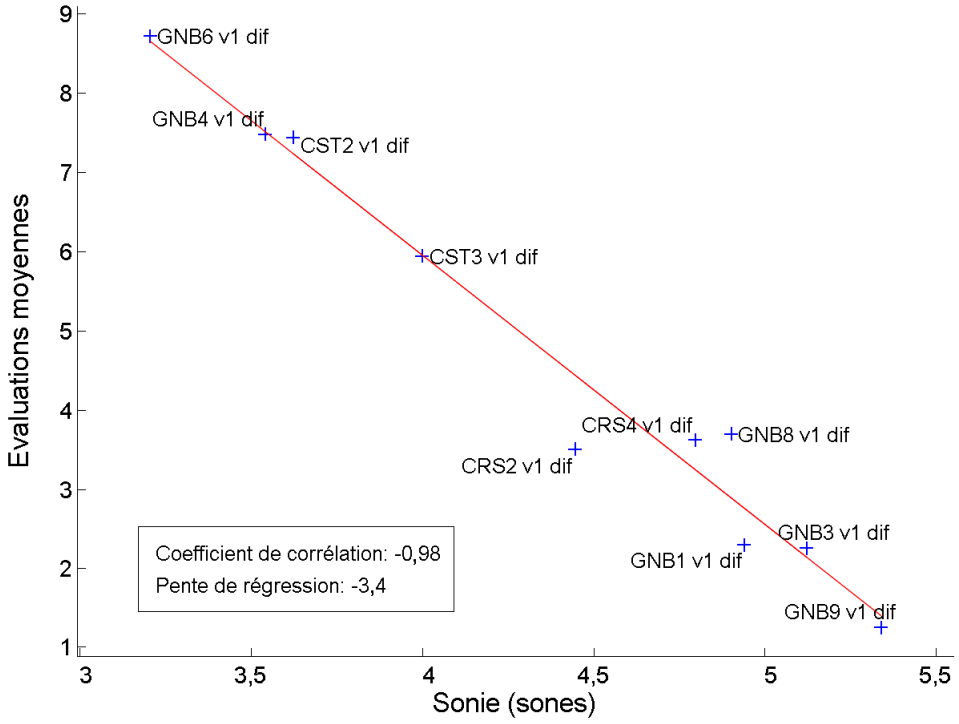


FIGURE B.1 – Exemple de diagramme de dispersion et de droite de régression (provenant de la section 7.6.1).

L'observation de la répartition du nuage de points autour de la droite de régression permet de mieux apprécier la qualité de la prédiction. Lorsque la prédiction est efficace, on doit constater que le nuage de points forme grossièrement une ellipse autour de la droite de régression. Cela va de paire avec un coefficient de corrélation r élevé en valeur absolue. Dans l'exemple de la figure B.1, r vaut -0,98, ce qui correspond à une forte corrélation, confirmée par la répartition des points autour de la droite de régression.

Annexe C

Méthodes d'analyse de variabilité des réponses

La particularité des expériences perceptives est de faire appel à des participants volontaires, qui ont, à priori, leur sensibilité propre vis-à-vis des sons étudiés. Il s'en traduit des jugements – dont la forme dépend du type d'expérience – reflétant parfois un caractère très subjectif. La conséquence possible lors d'une expérience perceptive est une forte variabilité des réponses recueillies. Cette variabilité peut s'avérer très problématique lorsque le but de l'expérience est d'extraire des données – sous quelque forme que ce soit – moyennées sur le panel de participant. En effet, le cas échéant, la significativité statistique des différences observées sur les données moyennes peut alors être difficile à démontrer, empêchant ainsi de faire émerger de manière concluante certaines tendances. Il convient alors de comparer les résultats des participants entre eux et de vérifier la concordance de leurs résultats.

Cette annexe présente deux méthodes permettant d'analyser et de comparer entre eux les résultats individuels des participants d'une expérience. La section C.1 expose l'évaluation de deux coefficients – toutefois liés l'un à l'autre – permettant d'obtenir une mesure globale de la variabilité des réponses des participants : le coefficient de concordance de Kendall et le coefficient de corrélation de rang de Spearman. La section C.2 présente la méthode d'*analyse de cluster* appliquée aux résultats d'un panel de participant afin d'observer la dispersion des réponses des participants autour de la tendance moyenne.

C.1 Concordance de Kendall et corrélation de rang de Spearman

Afin de comparer les résultats d'un ensemble de participants dans le cadre d'une expérience perceptive, il convient de comparer l'ordonnancement des échelles mesurées pour chaque participant et de tenter de quantifier le degré de concordance entre les ordonnancements obtenus. Ceci peut se faire au moyen du coefficient de concordance de Kendall et du coefficient de corrélation de rang de Spearman (voir [82] pour plus de détails).

C.1.1 Coefficient de concordance de Kendall

Pour calculer le coefficient de concordance de Kendall, il convient en premier lieu de sommer les classements successifs de chaque son. Par exemple, si un son a été classé successivement 1er, 2^{e}et 5^{e}, la somme des classements vaut 8. Le coefficient se calcule alors comme le quotient de la variance des sommes ainsi calculées sur la variance maximale possible de ces sommes pour ce nombre de sons et

pour ce nombre de participants. Comme nous travaillons ici avec des rangs (de 1 à N pour N sons), nous connaissons la variance maximale possible des sommes de classement. Il est ainsi possible de définir la formule du coefficient de concordance de Kendall par :

$$W = \frac{12\sum_j T_j^2}{k^2 N(N^2-1)} - \frac{3(N+1)}{N-1} \tag{C.1}$$

avec T_j la somme des classements du son j,
N le nombre de sons, et k le nombre de participants.

C.1.2 Coefficient de corrélation de rang de Spearman

Le coefficient de concordance de Kendall n'est pas un coefficient de corrélation à proprement parler, et son interprétation n'est pas triviale. En revanche, il est possible de le relier au coefficient de corrélation de rang de Spearman r_s (voir [82]). Ce dernier est adéquat lorsque l'on souhaite évaluer la concordance entre les classements effectués par deux participants seulement. Il est alors possible de calculer ce coefficient pour chaque paire de participants parmi les k participants considérés pour en évaluer la moyenne $\bar{r}_s$, mais ceci peut s'avérer assez fastidieux. Toutefois, il existe une formule reliant $\bar{r}_s$ au coefficient de concordance de Kendall :

$$\bar{r}_s = \frac{kW-1}{k-1} \tag{C.2}$$

avec k le nombre de participants.

Il est alors possible de tester la significativité statistique du coefficient obtenu comme pour le coefficient de corrélation de Bravais-Pearson (voir annexe B), afin de conclure sur la concordance globale entre les réponses des participants. Il est à noter que cette technique peut être appliquée à chacune des méthodes exposées en section 2.3.1, si tant est que l'on ait transformé les données obtenues en données de classement.

C.2 Analyse de cluster appliquée à un panel de participants

Les coefficients présentés en section C.1 permettent d'obtenir une mesure globale de la concordance des réponses des participants d'une expérience perceptive. Cependant, ils ne permettent pas d'observer individuellement le comportement des échelles de chaque participant. Plusieurs cas de figure peuvent résulter en des coefficients de concordance de Kendall et de corrélation de rang de Spearman faibles.

Une possibilité est par exemple de voir un nombre réduit de participants – par rapport à l'effectif total – présenter des résultats radicalement différents de ceux de la majorité et entraîner une forte variabilité globale, et de faibles valeurs pour ces coefficients. On parle alors d'*outliers* pour désigner les résultats de chacun des participants de ce nombre réduit, et par abus de langage, les participants eux-mêmes.

Une autre possibilité est de voir les participants se regrouper en sous-groupes présentant des résultats proches, et représenter ainsi chacun une tendance particulière qu'il convient de considérer

séparément des autres. Toutefois, ce deuxième cas de figure, bien différent du premier, peut très bien résulter en des valeurs similaires de coefficients de concordance de Kendall et de corrélation de rang de Spearman.

Ces deux coefficients ne permettent donc pas de distinguer ces deux cas de figure, tandis que ces derniers requièrent chacun un traitement particulier de l'ensemble des résultats des participants. Dans le premier cas, il convient de retirer les résultats considérés comme *outliers*, quand on préférera séparer le panel de participants en deux (ou plus le cas échéant) sous-panels pour lesquels l'analyse désirée sera reproduite à chaque fois.

Un moyen de remédier à ce problème, et donc de pouvoir distinguer, par exemple, les deux cas de figure envisagés ci-dessus, est d'appliquer une *analyse de cluster* (voir [98] et section 5.3 pour plus de détails) aux résultats individuels. Ce type d'analyse est habituellement utilisée afin d'extraire une structure hiérarchique décrivant les distances ou co-occurrences entre un ensemble d'éléments (de sons dans le cas de la section 5.3). Mais il est également possible de considérer le panel de participants comme cet ensemble d'éléments dont on souhaite extraire une représentation des proximités.

Pour ce faire, la matrice de corrélation inter-participant permet d'observer les divergences et les similitudes des évaluations des participants deux à deux (coefficient de corrélation proche de 1 pour des jeux d'évaluations très proches, et proche de 0, ou négatif, pour des jeux très différents). Cette matrice est donc particulièrement informative quant à la variabilité des réponses autour des valeurs moyennes. Elle est toutefois difficilement lisible, car centrée sur des comparaisons deux à deux, et parce qu'elle ne donne pas une vision globale du comportement du panel de participants.

En revanche, l'analyse de cluster, appliquée à cette matrice de corrélation, va permettre d'obtenir une représentation globale des proximités entre les résultats des participants. Toutefois, l'analyse de cluster doit s'appliquer sur une matrice de données assimilées à des distances, ce qui n'est pas le cas d'une matrice de corrélation. Afin de rendre les données compatibles avec ce type d'analyse, la matrice de corrélation est linéairement transformée en matrice de distance $D = \{d_{ij}\}$, dont la valeur minimale pour une paire de participants (i, j) est '0' pour un coefficient de corrélation $r_{ij} = 1$, et la valeur maximale est 1 pour un coefficient de corrélation de $r_{ij} = -1$:

$$d_{ij} = \frac{1 - r_{ij}}{2} \tag{C.3}$$

L'analyse de cluster, utilisant le même algorithme qu'en section 5.3 (*unweighted arithmetic average clustering – UPGMA*), est ensuite appliquée sur cette matrice de distance afin d'obtenir une représentation hiérarchique du panel de participants. Un exemple de dendrogramme, tiré de la section 7.5, et traduisant cette représentation est affiché en figure C.1. Sur celui-ci les « feuilles » en bas de la figure représente ici les participants, et la distance cophénétique, modélisant la distance d_{ij}, correspond à la hauteur de fusion des deux feuilles correspondantes, c'est-à-dire la hauteur à laquelle leurs « branches » se rejoignent. Cette représentation, en fonction de la répartition des participants dans les clusters, permet alors de visualiser si une éventuelle forte variabilité est liée par exemple à des *outliers* ou à différentes tendances parmi le panel.

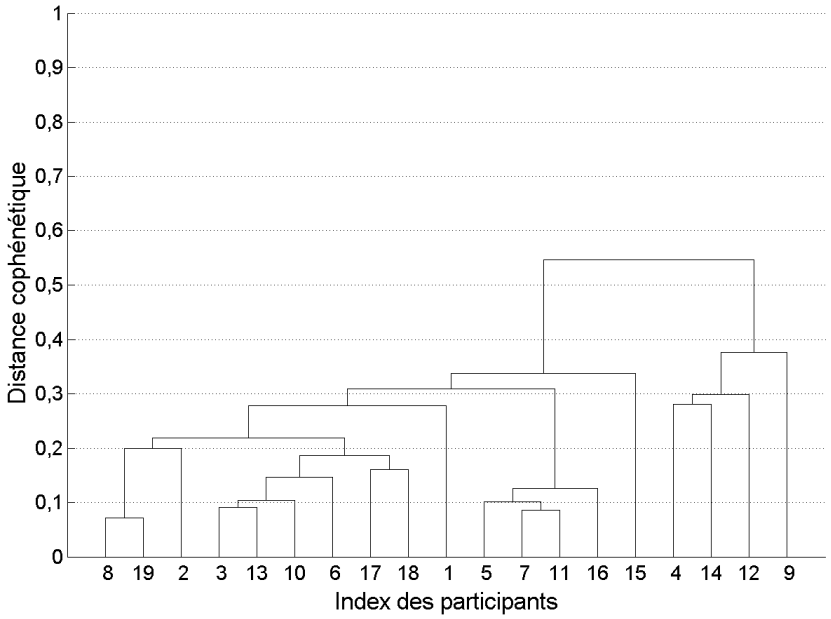


FIGURE C.1 – Exemple de dendrogramme, tiré de la section 7.5, et représentant les proximités entre les résultats des participants d'une expérience.

Annexe D

Liste des enregistrements de STA effectués

	nom complet du système	vitesses enregistrées	positions	types d'enregistrements
carrossés	**CRS1**	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	CRS2	1,2,5	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	CRS3	1,3,4	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	CRS4	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
gainables	**GNB1**	1,3,6	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	GNB2	1,2	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	GNB3	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural

	nom complet du système	vitesses enregistrées	positions	types d'enregistrements
gainables	**GNB4**	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	GNB5	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	GNB6	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	GNB7	1,2,4	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	GNB8	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	GNB9	1,2,4	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
cassette	**CST1**	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	CST2	1,2,3	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural
	CST3	1,2	diffusion reprise diffusion diffusion	monophonique monophonique stéréophonique ORTF binaural

TABLE D.1 – Liste des enregistrements effectués.

Annexe E

Consignes écrites pour les expériences perceptives

E.1 Expérience de catégorisation libre

Consigne de l'expérience
Catégorisation de sons de climatiseur

Nous nous intéressons aux sons de climatiseurs. Le but de cette expérience est de mettre à jour les différentes familles de ce type de sons.
Vous allez donc entendre des sons de climatiseur et vous aurez à les classer dans différentes catégories.

Voici le descriptif de l'expérience:

- Ecoutez les sons en double-cliquant sur les points rouges de l'écran dans l'ordre que vous souhaitez et autant de fois que vous le désirez. Chaque point rouge correspond à un son de climatiseur.
- Regroupez les points rouges correspondant aux sons selon leur ressemblance acoustique en les déplaçant sur l'écran avec la souris, et en prenant soin de bien distinguer les groupes par leur disposition à l'écran.
- Une fois les différents groupes constitués, sélectionnez, pour chacun d'eux, les différents sons qui le compose (les points sélectionnés doivent devenir verts), puis cliquez sur [**Faire un groupe]** puis cliquez sur **[OK]**. Cliquez sur le groupe que vous venez de créer dans la partie gauche de l'écran (liste des groupes créés) sous le cadre vert, afin de vérifier qu'aucun son n'a été oublié ou ajouté par erreur à la sélection. Pour afficher de nouveau l'ensemble des sons, cliquez sur **[Sounds]** en haut de la partie gauche de l'écran (sous le cadre vert).
- Répétez l'étape précédente pour chacun des groupes que vous avez identifiés, en veillant, à chaque fois, à ce que tous les sons du groupe soient sélectionnés (points verts) et à ce qu'aucun son des autres groupes ne le soit.
- Lorsque vous êtes satisfait de votre classification, et que vous avez déclaré tous les groupes, l'expérience est terminée, vous pouvez appeler l'expérimentateur.

Remarque:

- Du fait du grand nombre de sons, il est conseillé d'être méthodique dans l'écoute des sons et de commencer à les classer en les répartissant à l'écran au fur et à mesure de l'écoute.

E.2 Expérience de mesure de similarités

Similarités des sons de systèmes de traitement d'air

Consigne de l'expérience perceptive :

Objectif du test :

Nous nous intéressons aux sons émis par les systèmes de traitement d'air. Des sons de ce type de systèmes vont vous être présentés au casque par paire. Pour chacune d'elles, vous aurez à évaluer à quel point les deux sons sont similaires.

Interface :

Pour chaque paire de son présentée, vous devez :

1. **Ecouter** chacun des deux sons à l'aide du **bouton** correspondant
2. Evaluer à **quel point ils sont similaires** – entre « très semblable » et « très dissemblables » – à l'aide du **curseur sur la glissière** situé en dessous des deux boutons
3. Lorsque celui-ci apparaît (après avoir déplacé le curseur), cliquer sur le **bouton valider** pour passer à la paire suivante

Pour chaque paire, vous pouvez réécouter chaque son et modifier la position du curseur à guise, tant que vous n'avez pas cliqué sur le bouton valider.

S'il n'est pas demandé d'utiliser toute l'échelle de similarité proposée (positions possibles du curseur), il est toutefois nécessaire de **suffisamment différencier vos jugements**.

Les **premières paires** présentées serviront d'**entraînement**, pour vous familiariser avec l'interface et avec la diversité des sons proposés (vos réponses ne sont alors pas prises en compte). L'interface vous indiquera clairement lorsque la phase d'entraînement sera terminée.

A environ la moitié du test, il vous sera proposé de faire une **courte pause** afin de vous détendre et de vous aérer l'esprit !

Lorsque l'interface vous l'indique, l'expérience est terminée.

E.3 Expérience d'évaluation comparée

Désagrément lié aux sons des systèmes de traitement d'air

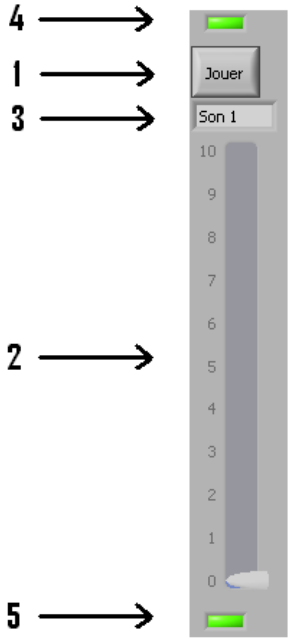
Consigne de l'expérience perceptive :

Objectif du test :

Nous nous intéressons aux sons émis par les systèmes de traitement d'air. Des sons de ce type de systèmes vont vous être présentés au casque. Pour chacun d'eux, vous aurez à donner une « note » qualifiant son caractère agréable ou désagréable. Vous aurez à guise, tout au long de l'expérience, la possibilité de réécouter les sons et de modifier vos réponses.

Interface :

Pour chaque son, l'interface de test se présente comme suit :



- *Éléments interactifs pour chaque son :*
    1. Bouton permettant de jouer le son (diffusé au casque)
    2. Curseur permettant d'ajuster la note, allant de **0** pour le ***plus désagréable*** à **10** pour le ***plus agréable***. Le curseur n'est **accessible que lorsque le son correspondant vient d'être joué** (1. Bouton)

- *Indicateurs (non-interactifs) pour chaque son:*
    3. Index du son
    4. LED indiquant si le son à déjà été joué (allumée au départ, éteinte dés lors que le son a été joué au moins une fois)
    5. LED indiquant si la position du curseur a déjà été modifiée (allumée au départ, éteinte dés lors que le curseur a été déplacé au moins une fois).
- *Éléments interactifs globaux :*

  L'interface présente également deux autres boutons (qui ne sont pas propres à un son particulier) :
    - Bouton « *Réordonner* » : Permet de modifier l'ordre des sons sur l'interface en fonction de la position du curseur. Les sons réordonnés selon un ordre décroissant des notes.
    - Bouton « *VALIDER* » **(masqué au début de l'expérience)** : Permet de valider l'ensemble des notations effectuées et met fin au test. Ce bouton **n'apparaîtra que lorsque la position de tous les curseurs a été modifiée au moins une fois.**

Tâche à accomplir :

En résumé, la tâche suit la logique de l'interface et se décompose de la manière suivante :

- Écoute d'un son (au choix)
- Notation via le curseur
- Écoute d'un nouveau son
- Notation via le curseur
- … Répétition sur tous les sons, voire également sur les sons déjà écoutés/notés
- Un réordonnancement peut être fait à tout moment (même avant d'avoir écouté/noté tous les sons)
- Lorsque tous les sons ont été écoutés/notés, validation des réponses

L'interface vous permet de réécouter les sons **autant de fois que vous le souhaitez**, et de **modifier à tout moment les notations** (toutefois, la position d'un curseur ne peut être modifiée qu'après avoir écouté le son correspondant). Bien que, en général, on se fasse rapidement une idée du degré de gêne d'un son, n'hésitez donc pas à **écouter plusieurs fois les sons** et à r**evenir sur vos notations après avoir évalué une première fois tous les sons**.

Par ailleurs, votre tâche consiste à **évaluer les sons relativement aux autres** et non par rapport à une représentation que vous vous faîtes de ce qu'est un son agréable ou désagréable dans l'absolu. Cela ne signifie pas qu'il faille se contenter de « classer » les sons, la précision de la notation est importante. Il faut en revanche, dans la mesure du possible, **utiliser au maximum la dynamique de l'échelle** (gamme de valeurs du curseur – de 0 à 10).

E.4 Expérience d'évaluation de gêne en contexte multi-tâche

Influence des sons de systèmes de traitement d'air sur une tâche de mémorisation

Consigne de l'expérience perceptive :

Objectif du test :

Nous nous intéressons aux sons émis par les systèmes de traitement d'air et plus particulièrement à leur influence sur la réalisation d'une tâche de mémorisation. L'expérience consistera à mémoriser des séquences de chiffres et à évaluer ensuite le caractère perturbant du fond sonore

Interface :

L'interface est explicite et conçue pour vous guider dans les opérations à réaliser. Ce qui suit n'en est qu'une description succincte. Afin de vous familiariser avec celle-ci, une session d'entraînement vous sera proposée par l'expérimentateur. Les phases de mémorisation sont alors plus courtes, et effectuées sans fond sonore, mais vos réponses ne sont pas enregistrées.

Après avoir entré votre nom, il vous sera indiqué sur l'interface la marche à suivre :

a. **Phase de mémorisation :**
   - Des chiffres vont vous être alors présentés un par un, formant ainsi une séquence à mémoriser.
   - Une fois tous les chiffres affichés, entrez dans le champ adéquat la suite de chiffres que vous avez mémorisée, et validez en appuyant sur la touche *Entrée* du pavé numérique (pas celui de la partie « alphabétique » du clavier)

Les séquences seront de plus en plus longues : 4 chiffres affichés, puis 5, puis 6 …etc. La longueur augmentera à chaque fois que vous aurez mémorisé et retranscrit 2 séquences correctement. La performance (nombre de séquence correctement retranscrites) est prise en compte.
Toutefois **le temps imparti est limité.** La limite de temps est globale sur UNE seule phase de mémorisation (c'est-à-dire sur plusieurs séquences de chiffres avec le même fond sonore). De plus seul le temps de saisie au clavier est pris en compte.
Par ailleurs, cette phase sera **parfois** à réaliser **sans fond sonore.** Ne vous en étonnez pas.

b. Phase d'évaluation du caractère perturbant :

- Lorsque le temps imparti est écoulé, il vous sera demandé d'évaluer à quel point le fond sonore vous a perturbé dans la tâche de mémorisation, à l'aide du curseur affiché.
- Après validation, une nouvelle tâche de mémorisation commencera comme la précédente, avec un autre fond sonore…

Cette étape n'est bien entendu pas à effectuer si aucun fond sonore n'était présent pendant la phase de mémorisation. **Si cette phase vous est proposée alors qu'aucun son n'a été diffusé dans le casque pendant la phase de mémorisation, LE SIGNALER A L'EXPERIMENTATEUR.**

<u>REMARQUES IMPORTANTES</u> :

- Bien que la stratégie pour la mémorisation est laissée libre, il peut être utile pour les séquences plus longues (6 ou 7 chiffres, ou plus) d'essayer de les mémoriser en associant les chiffres par 2 ou 3.

- S'il n'y pas de consigne particulière quant à la phase d'évaluation (positions possibles du curseur), il est toutefois nécessaire de **suffisamment différencier vos jugements** (par exemple, placer le curseur à la même position pour tous les sons ne présente pas d'intérêt).

- A environ la moitié de l'expérience, il vous sera proposé de faire une **courte pause** afin de vous détendre et de vous aérer l'esprit !