

RAPPORT DE STAGE DE MASTER



SORBONNE
UNIVERSITÉ



Réduire le bruit grâce à la Réalité Augmentée Sonore - *Auditory Concealer*

Clara BOUKHEMIA

Encadrants : Nicolas MISDARIIS (STMS, IRCAM)

& Mathieu LAGRANGE (LS2N, CNRS)

Responsable pédagogique : Antoine GONOT

Rapporteur : Mitsuko ARAMAKI

1^{er} septembre 2024

M2 Acoustique & Musicologie
Année Universitaire 2023–2024

Résumé

Ce rapport présente le travail de 22 semaines de stage au sein de l'équipe Perception et Design Sonores du laboratoire Sciences et Technologies de la Musique et du Son (STMS) à l'Institut de Recherche et de Coordination Acoustique/Musique (IRCAM). Au lancement du projet *Reducing Noise with Augmented Reality (ReNAR)*, qui vise à créer un outil permettant de réduire en temps-réel l'impact cognitif des sons perçus comme désagréables ou gênants au sein d'environnements intérieurs, une première étude a été réalisée dans le but de valider la faisabilité et l'efficacité d'une nouvelle approche de masquage nommée *concealer*. L'hypothèse principale réside dans la possibilité que l'approche *concealer* fournisse de meilleurs résultats qu'une approche dite *masker* en termes d'agrément perçu. Des mixtures entre deux sources de bruit (ventilation) et cinq maquilleurs (sons d'eau) ont été générées à l'aide des deux approches selon plusieurs niveaux. L'évaluation de l'agrément perçu de ces mixtures a montré que l'approche *masker* reste meilleure que l'approche *concealer*, quelque soit la source de bruit, le son d'eau, ou le niveau utilisé.

Mots-clefs : *approche concealer, approche masker, agrément, bruit de ventilation, sons d'eau*

Remerciements

Je tiens en premier lieu à remercier mes deux encadrants, pour leur bienveillance, leur confiance et leurs précieux conseils tout au long de ce stage.

Merci à Nicolas Misdariis pour m'avoir formée au développement d'une expérience perceptive. Son insatiable curiosité, son optimisme et sa capacité à toujours s'émerveiller des nouvelles propositions artistiques me serviront d'exemple pour la suite de mon projet professionnel.

Merci à Mathieu Lagrange pour son aide et ses explications sur les aspects théoriques et pratiques du traitement du signal. Son ouverture d'esprit et son expérience de chercheur m'ont donné confiance en moi pour continuer dans cette voie.

Je remercie également toute l'équipe Perception et Design Sonores, qui m'ont accueillie dans une ambiance joviale, et qui ont fait de ce stage une période plus qu'agréable. Un grand merci pour ces pauses de midi soldées de points cultures cinématographiques, littéraires, et évidemment musicaux ! Je remercie particulièrement les doctorants, Matthieu, Michèle et Armand, mes voisins de bureau, pour leur aide et leur humour.

Comment ne pas dédier quelques mots à mes incroyables parents, pour les remercier de leur soutien sans faille. Merci de toujours croire en moi et d'être venu apporter un peu du soleil marseillais à Paris !

Un grand merci à mon super-copain, pour sa gentillesse sans limite et son soutien à toute épreuve. Merci d'être là.

Enfin, pour clore un chapitre et en ouvrir un nouveau, merci à toute l'équipe pédagogique du master Acoustique et Musicologie, et en particulier à Antoine Gonot, pour son écoute et son énergie. Merci à mes camarades et amis, avec qui j'ai passé deux années riches en joie et en rigolade !

Table des matières

1	Introduction	5
2	État de l’art	6
2.1	Qu’est-ce que le confort d’un point de vue sonore ?	6
2.2	Étudier la gêne pour améliorer le confort	7
2.3	Étudier l’agrément pour améliorer le confort	7
2.4	Le masquage pour traiter le bruit au niveau perceptif	9
2.5	<i>Best-Worst Scaling</i> : une méthode pour évaluer la perception sonore	10
2.5.1	Fonctionnement de la méthode <i>Best-Worst Scaling</i>	10
2.5.2	Algorithmes de notation : mise à jour des scores d’un grand nombre d’éléments	11
3	Construction d’une nouvelle approche de masquage	14
3.1	Objectifs et enjeux du projet <i>ReNAR</i>	14
3.2	Définition des approches <i>masker</i> et <i>concealer</i>	14
3.3	Choix de la source et du positif	16
3.3.1	Choix de la source	17
3.3.2	Choix du positif	19
3.4	Construction du signal <i>concealer</i>	20
3.4.1	Définition des méthodes de construction	20
3.4.2	Choix de la méthode	22
4	Expérience perceptive	24
4.1	Présentation de l’expérience	24
4.2	Méthodologie	24
4.2.1	Participants	24
4.2.2	Stimuli	25
4.2.3	Apparatus	28
4.2.4	Protocole	28
4.3	Résultats	31
4.3.1	Durée de l’expérience	31
4.3.2	Choix de l’algorithme de notation	31
4.3.3	Cohérence inter et intra-participants	31
4.3.4	Effet de l’approche sur le score	32
4.3.5	Effet de la source sur le score	33
4.3.6	Effet du niveau sur le score	34
4.3.7	Effet du positif sur le score	36
4.3.8	Visualisation des paramètres de l’expérience	37
4.3.9	Enquête sur les sons désirables et indésirables dans les bureaux ouverts	41
4.4	Discussion	43
5	Conclusion	45
	Annexes	48
	Consigne présentée aux participants (section 4)	48
	Feuille de consentement présentée aux participants (section 4)	50
	Questionnaire présenté aux participants à la fin de l’expérience (section 4)	52

Table des figures

1	Modèle de confort proposé par Vink et Hallbeck [30], repris et traduit par Matthieu Duroyon [8]	6
2	Schéma résumant le principe de l'approche masker	15
3	Schéma résumant le principe de l'approche concealer	16
4	Prise de son en champ proche du système de ventilation présent dans les bureaux du laboratoire STMS (équipe Perception et Design Sonores)	17
5	Prise de son du système de ventilation présent dans les bureaux du laboratoire STMS (équipe Perception et Design Sonores) au niveau d'un auditeur assis à son bureau	18
6	Spectre des deux bruits de ventilation	19
7	Schéma présentant l'amplitude du signal concealer en fonction de la fréquence (en vert), résultat de la soustraction des amplitudes du positif (en rouge) et de la source (en bleu)	21
8	Schéma présentant l'amplitude du signal concealer en fonction de la fréquence (en vert), résultat de la valeur absolue de la soustraction des amplitudes du positif (en rouge) et de la source (en bleu)	21
9	Schéma présentant l'amplitude du signal concealer en fonction de la fréquence (en vert), résultat de la soustraction non négative des amplitudes du positif (en rouge) et de la source (en bleu)	22
10	Spectres lissés des deux bruits de ventilation et des cinq sons d'eau pour une plage de fréquences allant de 0 à 20kHz	25
11	Spectres lissés des deux bruits de ventilation et des cinq sons d'eau pour une plage de fréquences allant de 0 à 1200Hz	26
12	Niveau de la mixture en fonction du niveau du positif pour le bruit de ventilation enregistré à l'IRCAM et le son de fontaine	27
13	Interface présentée aux participants lors de la première partie de l'expérience, visant à juger des groupes de quatre séquences en termes d'agrément	29
14	Interface présentée aux participants lors de la seconde partie de l'expérience, visant à étudier la reconnaissance de source des différents positifs	30
15	Cohérence inter-participants (en bleu) et intra-participants (en orange) pour chaque participant	32
16	Scores moyens des mixtures en fonction de l'approche utilisée (concealer en rond rose et masker en carré vert)	33
17	Scores moyens des mixtures en fonction de l'approche utilisée et des deux sources (ventil1 en rouge et ventil2 en bleu)	34
18	Scores moyens des mixtures en fonction de ΔL_{Aeq} pour les deux approches masker (carré vert) et concealer (rond rose)	35
19	Scores moyens des mixtures en fonction du type de son d'eau pour les deux approches masker (carré vert) et concealer (rond rose)	36
20	Scores des mixtures en fonction de ΔL_{Aeq} , de l'approche (ronds et carrés) et du positif utilisé (bleu, orange, vert, rouge et violet) pour les deux bruits de ventilation	38
21	Scores moyens des mixtures et indices spectraux pour les cinq sons d'eau	39
22	Scores moyens des mixtures en fonction des groupes de positifs pour la source "ventil1"	41
23	Occurrence des mots cités pour décrire les sons perçus comme gênants dans les bureaux ouverts	42
24	Occurrence des mots cités pour décrire les sons désirés dans les bureaux ouverts	43

1 Introduction

Défini par le Ministère de la Transition Écologique et de la Cohésion des Territoires comme “*un phénomène acoustique produisant une sensation auditive considérée comme désagréable ou gênante*”, le bruit représente un problème très présent dans le quotidien de beaucoup de personnes. Parfois constamment soumises à l’excès de bruit, ces personnes peuvent voir leur santé se détériorer d’un point de vue physiologique (endommagement du système auditif entraînant des lésions ou des pertes auditives), mais aussi d’un point de vue psychologique (apparition de symptômes tels que la perturbation du sommeil ou le stress). Ces nuisances sonores, résultat de la pollution sonore rencontrée dans plusieurs types d’environnements, soulèvent donc de réels enjeux de santé publique.

Pour tenter d’améliorer le confort des personnes au sein de leurs espaces de vie, le projet *Reducing Noise with Augmented Reality (ReNAR)* financé par l’Agence Nationale de la Recherche (ANR) propose de créer un outil permettant de réduire en temps-réel l’impact cognitif des sons perçus comme désagréables ou gênants au sein d’environnements intérieurs. Pour cela, deux approches de réalité augmentée sonore sont étudiées en parallèle au sein du laboratoire STMS de l’IRCAM à Paris, du LS2N à Nantes et du Loria à Nancy, grâce à une collaboration entre les chercheurs Nicolas MISDARIIS (Paris), Mathieu LAGRANGE (Nantes) et Romain SERIZEL (Nancy). Le travail autour des deux approches est mené selon deux axes. Le premier, intitulé *Auditory Concealer*, a pour but de déterminer quels sont les aspects importants de l’agrément ou du confort perçu des scènes acoustiques, et d’identifier les sources sonores gênantes ou désagréables présentes dans les environnements étudiés. Le deuxième, intitulé *Speech Intelligibility Reducer*, s’intéresse plus spécifiquement à la parole intelligible et vise à réduire l’impact négatif des conversations environnantes sur les personnes partageant un espace. Pour les deux axes, l’approche développée consiste à ajouter un signal à une scène sonore contenant un son indésirable, tout en minimisant l’augmentation du niveau sonore global de cette scène.

Le travail effectué dans le cadre du stage constitue un point d’entrée dans le projet *ReNAR* pour l’axe *Auditory Concealer*, avec une première étape dans le développement de l’approche *concealer* qui se focalise sur un type d’environnement bien précis : les bureaux ouverts. Dans la catégorie des environnements intérieurs, ce type d’espace de travail n’échappe pas au problème du bruit, et ce depuis sa conception dans les années 1950. Les nuisances sonores se traduisent dans ce cas par la présence de sources sonores saillantes, dérangeantes et indésirables qui peuvent polluer le paysage sonore global et le rendre désagréable, voire ennuyeux pour les personnes qui y travaillent. Partant du cas d’étude d’un système de ventilation, un questionnaire s’est posé sur la faisabilité et l’efficacité de l’approche *concealer* face à une approche de masquage nommée approche *masker*, souvent retrouvée dans la littérature. En particulier, il est intéressant de se demander si l’approche *concealer* se trouve être plus efficace que l’approche *masker* en termes d’agrément, et ce pour un niveau global proche de celui du bruit initial. Pour tenter de répondre à cette question, une revue de l’état de l’art sera tout d’abord explicitée en section 2. La construction de l’approche *concealer* sera ensuite détaillée en section 3, pour enfin la comparer en section 4 à l’approche *masker* à l’aide d’une expérience perceptive permettant d’évaluer l’agrément perçu. Cette dernière section présentera les premiers résultats de cette expérience, et une amorce de discussion conclura ce rapport pour ouvrir sur la suite du projet.

2 État de l’art

Avant d’entamer une réflexion en lien avec le projet *ReNAR*, cette partie reprendra les notions importantes citées dans l’état de l’art complet réalisé au début de la période de stage [3] afin de poser un cadre général sur ces notions et la problématique du bruit.

2.1 Qu’est-ce que le confort d’un point de vue sonore ?

Pour parler de confort en utilisant les termes “Confort/Inconfort”, il est important d’introduire une certaine contextualisation. En effet, les modèles de confort présentés dans la littérature établissent un lien étroit entre objet, environnement et utilisateur. Celui que proposent Vink et Hallbeck [30], et qui est repris par Matthieu Duroyon dans le cadre de sa thèse [8] permet de bien se rendre compte de ces interactions (Figure 1). Comme l’expliquent Duroyon et coll. par leur définition, “*le confort est une sensation subjective rendant un objet manufacturé agréable dans son utilisation et cohérent avec les attentes des utilisateurs*” [9]. Les attentes et la satisfaction des personnes concernant un produit entrent ainsi en compte dans l’évaluation du confort.

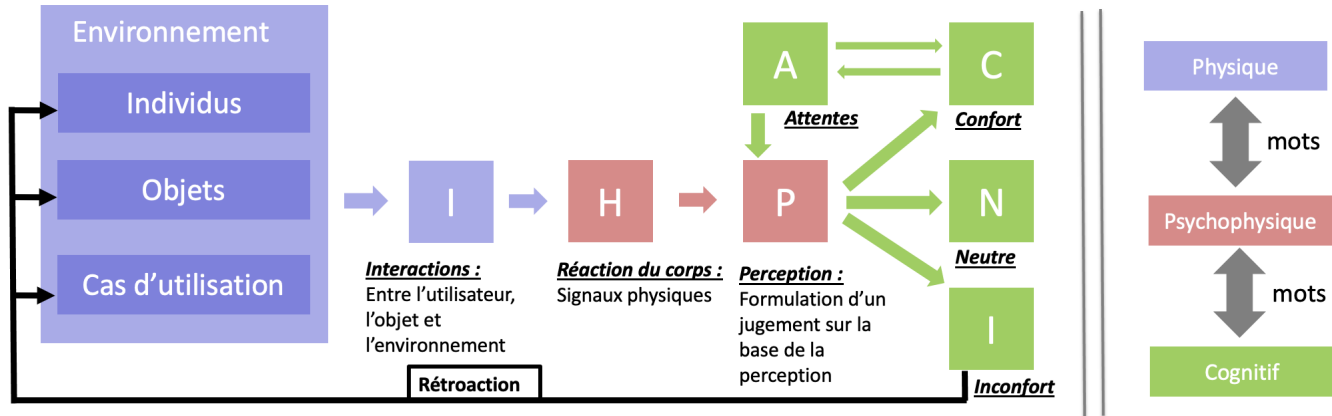


FIG. 1 – Modèle de confort proposé par Vink et Hallbeck [30], repris et traduit par Matthieu Duroyon [8]

Mais de quel confort parle-t-on ? S’il est clair que celui-ci concerne le son, il est possible de faire la différence entre plusieurs types de confort. Soucieux d’utiliser un vocabulaire adapté lors de sa recherche sur le confort au sein des véhicules électriques, Duroyon en différencie deux types liés au son [8]. D’une part le confort “acoustique”, qui s’approche d’un travail sur la “quantité sonore”, visant notamment à la réduire. D’autre part, le confort “sonore”, qui s’accorde davantage avec la “qualité sonore”, et suggère de s’intéresser à la perception des environnements.

Par ailleurs, le jugement d’un produit manufacturé impose un cadre industriel à l’évaluation du confort, qu’il soit acoustique ou sonore, mais il est possible de ramener ce jugement à des environnements entiers. Pour parler du confort à bord des trains ou des véhicules électriques par exemple, les adjectifs “Confortable” et “Inconfortable” sont utilisés pour décrire l’environnement comme un produit dont les personnes se servent au quotidien. Les études visant à évaluer le confort dans de tels environnements se déroulent la plupart du temps *in situ*, afin de mettre les usagers en conditions réelles, et de leur fournir un contexte qui se veut le plus réaliste possible. C’est ce que font notamment Mzali et coll. en 2000 [20] et Kossachka et coll. en 2006 [14] lors de leurs études sur le confort à bord des trains, utilisant une approche psycholinguistique. Il est également intéressant de citer la méthode d’écoute réactivée,

explicitée par Augoyard en 2001 dans son article *L’entretien sur écoute réactivée* [1], dont le but est de laisser les habitués d’un lieu expérimenter les séquences sonores typiques de ce lieu avant de mener un entretien pour récolter leurs ressentis et interprétations de l’environnement sonore.

2.2 Étudier la gêne pour améliorer le confort

En définissant le confort, il est apparu qu’une mesure ou une évaluation de ce dernier demanderait une contextualisation poussée, obtenue par la mise en place d’une expérience *in situ*. Une telle expérience étant difficile à mettre en œuvre, il est possible de s’intéresser à des notions étroitement liées à la notion de confort (si ces liaisons ne sont pas explicitement étudiées, elles paraissent assez intuitives dans la littérature qui traite du confort). C’est le cas de la notion de gêne, qui est plus étudiée dans la littérature et directement applicable au son grâce aux adjectifs “Gênant” et “Non gênant” (ce qui n’est pas le cas du confort, puisqu’un son est rarement évalué comme “confortable” ou “inconfortable”).

Plusieurs protocoles sont utilisés pour mesurer la gêne, qui permettent de récolter des données objectives et subjectives, mais qui semblent également requérir une certaine contextualisation. En effet, la notion de gêne est souvent liée à une action ou une tâche, que celle-ci soit réelle ou imaginée. Selon Guski et coll. [12], *“la gêne constitue une réaction à la perturbation de l’activité causée par des événements bruyants”*. D’après les conclusions de Misdariis et coll. [19] et de Marquis-Favre et coll. [17] lors de leurs études sur la gêne perçue, il est alors possible de simplement demander à des personnes d’effectuer, ou de s’imaginer effectuer une tâche pour évaluer la gêne, que ce soit lors d’une expérience *in situ* ou en laboratoire. Il est important de préciser gêne “perçue”, car celle-ci est évaluée de manière subjective, en demandant aux personnes qui effectuent la tâche de décrire à quel point ils se sentent gênés. Une autre mesure, cette fois objective, peut être envisagée en récoltant la performance à la tâche demandée, celle-ci donnant une indication sur la gêne induite par les sons. En étudiant l’évolution de cette performance, il est possible d’évaluer une gêne autre que perçue (une décroissance de performance indiquant un niveau de gêne élevé). C’est ce que mettent en place Brocolini et coll. [5] lors de leur étude visant à évaluer la gêne induite par la présence de différentes sources sonores dans les bureaux ouverts.

Finalement, s’il semble aller de soi qu’une réduction de la gêne due aux sons présents dans un environnement permettrait d’en améliorer le confort, quelques points posent tout de même question. Tout d’abord, puisque la gêne s’évalue en relation avec une action, il est nécessaire comme pour le confort de poser certains éléments de contexte. Aussi, il semble apparaître deux types de gêne : d’une part la gêne dite “spécifique”, ressentie par l’auditeur lorsque celui-ci est soumis au bruit d’une unique source sonore. Ce premier type de gêne a notamment été étudié par Sporer et coll. [27] dans le but d’identifier les sources sonores jugées les plus gênantes. D’autre part, la gêne dite “totale”, ressentie par l’auditeur lorsque celui-ci est soumis au bruit de plusieurs sources sonores combinées, pourrait permettre de se rapprocher un peu plus de la réalité d’une scène sonore complexe. Le lien entre gêne spécifique et gêne totale n’étant pas évident (la somme des gênes spécifiques ne permet pas de représenter la gêne totale a priori), il semble difficile d’imaginer se faire une idée de la gêne totale induite par l’ambiance sonore d’un environnement composé de différentes sources sonores en mesurant la gêne spécifique induite par chacune de ces sources sonores.

2.3 Étudier l’agrément pour améliorer le confort

La définition d’une tâche à réaliser étant nécessaire pour évaluer la gêne, il est possible de se tourner vers une dernière notion pour simplifier l’évaluation d’ambiances sonores : l’agrément. Désignée par les

adjectifs “Plaisant” et “Non plaisant”, ou encore “Agréable” et “Désagréable”, un lien intuitif peut être établi entre cette notion et celle de confort. En effet, il est peu probable qu’un environnement sonore confortable soit désagréable, et la notion d’agrément apparaît clairement dans la définition de Duroyon et coll. [9] citée en section 2.1. Une première réflexion tend à dire que l’agrément est une notion subjective relevant des goûts de chacun. C’est une sensation qui vient naturellement et qui n’est pas forcément explicable. Étudiée à l’aide de mesures subjectives d’agrément “perçu”, il est le plus souvent demandé aux participants d’expériences perceptives de juger s’ils trouvent un son ou un environnement sonore agréable ou désagréable, sans parler de tâche à réaliser comme pour la gêne.

Plusieurs études se sont intéressées à l’agrément d’environnements sonores, et l’évaluation de cette notion a permis de la mettre en relation avec certains facteurs. Par exemple, l’étude de Liebetrau et coll. a permis de faire des liens entre certains facteurs acoustiques comme la rugosité et la netteté et l’agrément perçu [16]. Des études de corrélations semblent en effet montrer que plus un son est net et rugueux, moins il est plaisant. D’autre part, l’étude de Guillén et coll., qui s’inscrit dans le contexte de la perception sonore en milieu urbain, met en avant un certain nombre de facteurs non-acoustiques apparemment corrélés avec la notion d’agrément [11]. Parmi eux se trouvent notamment l’aspect visuel des environnements et le comportement des participants vis-à-vis du bruit.

Cela étant dit, il est important de préciser que l’évaluation de l’agrément peut se faire de manière globale ou de manière continue. L’évaluation globale se fait généralement à l’aide d’une échelle sémantique, à la manière de Guillén et coll. [11], qui présentent aux participants une échelle de jugement allant de 1 pour “*Very Unpleasant*” à 7 pour “*Very pleasant*”. Ceux-ci expérimentent d’abord des séquences sonores, puis utilisent l’échelle pour les juger en termes d’agrément. L’évaluation continue, quant à elle, peut aussi se faire à l’aide d’une échelle sémantique, utilisée cette fois par les participants pendant l’écoute des sons ou des scènes sonores. Cette seconde méthode a notamment été utilisée par Aumond et coll. lors de leur étude visant à évaluer l’agrément général d’un environnement sonore le long d’un trajet urbain, qui montre que l’agrément ressenti global donné par les participants après avoir été immergés dans une scène sonore est principalement expliqué par la moyenne de toutes les valeurs recueillies [2]. Baptiste Bouvier utilise également une évaluation continue dans le cadre de sa thèse, qui traite de la saillance auditive dans le cadre de la perception de l’environnement sonore [4]. Cette évaluation concerne néanmoins le désagrément, et non l’agrément perçu. Le but étant d’étudier le potentiel lien entre saillance et désagrément, il justifie sa volonté de ne pas se restreindre à une mesure globale d’abord par souci de comparaison. En effet, la saillance a été mesurée de manière continue (une scène sonore peut contenir différents événements sonores saillants), il faut donc une mesure du même type pour le désagrément pour comparer les deux notions. Aussi, dans un cas où seul le désagrément serait étudié, Bouvier affirme qu’une restitution globale rétrospective est affectée par différents effets, notamment l’effet de récence et de primauté, décrit par Aumond et coll. [2] comme l’effet “*par lequel les jugements initiaux et finaux d’une séquence sont mieux mémorisés au moment où l’évaluation rétrospective est donnée*”. Il est donc selon lui plus souhaitable d’étudier le désagrément en continu pour se rapprocher de l’expérience des usagers immergés dans un paysage sonore.

Concernant l’agrément en tant que notion, regroupant à la fois ses aspects positif et négatif (agrément et désagrément), aucune définition ne semble donnée par la littérature. Les recherches sur l’agrément perçu dans un contexte sonore donnent néanmoins matière à réflexion sur les méthodes d’évaluation et les types de facteurs responsables de cette sensation subjective. Parmi ces derniers, des indicateurs acoustiques peuvent être évalués grâce à des corpus sonores représentatifs du paramètre étudié, et des indicateurs non-acoustiques peuvent être recueillis à l’aide de questionnaires. De plus, même si l’agrément

dépend des goûts de chacun, il est apparu dans la littérature que certains sons sont communément acceptés, c'est-à-dire qu'ils paraissent agréables à l'oreille humaine en général. C'est par exemple le cas des sons naturels. L'étude de Cai et coll., qui vise à évaluer l'effet du masquage faisant intervenir des sons d'eau sur la perception du bruit industriel, semble confirmer que les sons d'eau sont jugés de manière positive [6]. Il serait alors possible de dresser une typologie des sons agréables et désagréables, aidant à améliorer l'agrément perçu, et possiblement le confort par extension.

2.4 Le masquage pour traiter le bruit au niveau perceptif

Les mesures de gêne ou d'agrément représentent de bonnes méthodes pour faire un état des lieux de la perception d'environnements sonores, mais elles ne permettent pas d'en améliorer le confort. Dans le cas où ces environnements sont mal jugés par leurs usagers, il reste à trouver un moyen de les rendre plus agréables, ou moins gênants. Une technique fréquemment utilisée pour cela est le masquage, dont le but est de masquer, dissimuler ou encore habiller l'environnement sonore (ou le bruit induit par une seule source sonore) afin de le rendre plus supportable ou appréciable. Le terme "masquage" est souvent utilisé de manière très globale, mais il est possible d'identifier plusieurs approches de masquage selon l'objectif posé. Par exemple, si le but de l'opération de masquage est de couvrir un son, il est plus que pertinent de faire référence au livre *"Psychoacoustics: Facts and models"* de Zwicker et Fastl [31], qui présente les propriétés d'un masquage dit "énergétique" visant à couvrir un son pur avec plusieurs types de "masques" (bruit large-bande, son pur, ou encore son complexe).

Dans un contexte plus appliqué, le masquage peut être vu comme solution pour traiter les problématiques de bruit. C'est ce que tentent de faire Cai et coll. dans leur étude sur la gêne induite par le bruit industriel [6], dans laquelle aucune approche de masquage spécifique n'est explicitée, mais dont l'objectif est de tester l'efficacité de l'ajout de sons d'eau sur la gêne perçue. La volonté n'est donc pas forcément de couvrir le bruit, mais plutôt de l'"habiller", les sons d'eau s'apparentant plus à des "maquilleurs" qu'à des "masques". Il a ainsi été montré que la réduction de la gêne perçue dépendait du rapport de niveau entre le bruit et le maquilleur, et que cette réduction était maximale lorsque le son présenté aux participants était composé du bruit de soudeuse électrique et du son de fontaine superposés, tous les deux au même niveau. L'approche d'"habillage", ici mise en œuvre par combinaison de la source de bruit et des "maquilleurs", est d'autant plus intéressante qu'elle peut permettre de ne pas ajouter d'objet sonore à la scène considérée. En effet, en créant ou choisissant un "maquilleur" qui se fond avec le bruit, il serait possible de garder un environnement sonore homogène, qui ne capte pas l'attention de l'auditeur.

Au contraire, capter l'attention peut tout aussi bien constituer l'objectif d'une approche de masquage, alors appelée masquage "informationnel". Il s'agit alors de faire diversion avec le signal ajouté, pour que l'attention de l'auditeur se focalise sur autre chose que sur la source de bruit. Dans le cas de l'étude de Cai et coll. [6], la combinaison du bruit de soudeuse électrique et des sons d'eau peut faire en sorte que l'attention des auditeurs soient captée par les sons d'eau, et non plus par le bruit. Cette méthode peut en revanche s'avérer non pertinente dans certaines situations, comme celle de travailleurs qui se plaignent d'un bruit qui les empêche de se concentrer. Ajouter un autre son qui capte leur attention pourrait en effet amoindrir le problème de gêne, mais ne serait pas forcément favorable à la concentration.

Plusieurs approches de masquage comme celles décrites ici peuvent finalement se trouver couplées à l'évaluation du confort, de la gêne ou de l'agrément, et constituer de bonnes pistes pour traiter la problématique du bruit au niveau perceptif.

2.5 *Best-Worst Scaling* : une méthode pour évaluer la perception sonore

Pour mener des expériences dont l’objectif est d’évaluer la perception sonore, plusieurs méthodes ont été développées. L’une d’entre elles, nommée *Rating Scale (RS)*, est largement utilisée et propose une échelle sémantique permettant aux participants de juger les sons qui leur sont proposés selon un critère fixé. Si cette méthode permet de recueillir un certain nombre d’informations (le jugement des participants, mais aussi les valeurs moyennes obtenues pour un seul son ou encore la cohérence inter et intra-participants), elle présente certains désavantages. Pour que les participants se fassent une bonne représentation du corpus de sons et sachent bien les placer sur l’échelle sémantique, il est nécessaire de leur présenter les limites de ce corpus par rapport au critère étudié. Par exemple, si l’expérience a pour but d’étudier la perception de la rugosité, le fait de présenter aux participants le son le moins rugueux et le son le plus rugueux avant de commencer l’expérience permet d’assurer une bonne utilisation de l’échelle sémantique. Cependant, cela suggère que le critère étudié est défini de manière objective. Si l’évaluation concerne un critère subjectif, comme l’agrément ou la gêne perçue, il n’est pas possible pour l’expérimentateur de définir à l’avance deux sons représentant les limites du corpus, à moins de se baser sur sa propre perception. L’utilisation de l’échelle sémantique devient alors difficile à envisager, et le développement d’autres méthodes d’évaluation pour ce genre de critère semble nécessaire. Parmi les nouvelles méthodes de plus en plus utilisées, la méthode *Best-Worst Scaling (BWS)* représente une bonne alternative à la méthode *RS* pour l’évaluation des sons. Dans un contexte général (c’est-à-dire pas seulement sonore), cette méthode consiste à présenter plusieurs groupes d’éléments, ce qui peut se révéler très utile dans plusieurs domaines de recherche pour évaluer un grand nombre d’éléments.

2.5.1 Fonctionnement de la méthode *Best-Worst Scaling*

Lors de son travail visant à tester de nouveaux algorithmes capables de noter des centaines voire des milliers d’éléments, Geoff Hollis présente le fonctionnement de la méthode *BWS* [13] comme suit : pour un total de S éléments, les participants se voient présenter plusieurs groupes de N éléments. Pour chaque groupe, ceux-ci doivent juger tous les éléments et en choisir deux qu’ils considèrent comme le “meilleur” et le “pire” selon un critère fixé. Cette technique permet d’associer un rang à chaque élément (même ceux qui ne sont jamais jugés comme meilleur ou pire) à l’aide de différents algorithmes de notation, et de générer un classement des éléments selon le critère étudié. Par exemple, en prenant $N = 4$ (ce qui est souvent le cas en pratique) et en notant A , B , C et D les éléments d’un groupe, il est possible de tirer cinq informations du jugement effectué par un participant pour ce groupe. Si celui-ci choisit A comme étant le meilleur élément et D comme le pire, il est possible de déduire les relations suivantes entre les éléments (l’inégalité $A > B$ étant interprétée comme “A est meilleur que B”) :

$$\left\{ \begin{array}{l} A > B \\ A > C \\ A > D \\ B > D \\ C > D \end{array} \right.$$

Seule la relation entre B et C reste inconnue.

En comparaison avec la méthode *RS*, la méthode *BWS* permet d’évaluer un grand nombre d’éléments en seulement quelques essais (un essai correspondant à un jugement sur un groupe de N éléments). Le fait d’évaluer les éléments en groupe assure également le caractère relatif du jugement, ce qui peut s’avérer

compliqué avec la méthode *RS* si les éléments sont présentés un par un. Néanmoins, certaines contraintes sont à prendre en compte, notamment le fait que le nombre total d’éléments S doit être multiple de N (ou N diviseur de S) pour la constitution de groupes contenant le même nombre d’éléments. De plus, il n’est pas possible de faire de moyenne sur un élément avec la méthode *BWS*, car les données récoltées sont des valeurs non-numériques (l’élément jugé comme meilleur et celui jugé comme pire). Il est alors possible de faire appel à des méthodes de comptage, pour calculer par exemple la fréquence à laquelle chaque élément a été choisi comme le meilleur. Il reste néanmoins possible d’utiliser la méthode *BWS* dans beaucoup de domaines différents, et notamment pour le son. C’est d’ailleurs pour tester la validité d’une telle méthode sur des stimuli sonores que Rosi et coll. ont mené une expérience avec vingt participants, à qui il a été demandé d’évaluer un ensemble de $S = 100$ sons instrumentaux en termes de brillance [26]. Les deux méthodes *BWS* et *RS* ont été utilisées, et les résultats ont montré que la méthode *BWS* est équivalente à la méthode *RS* en termes de performance, et plus rapide que cette dernière. De plus, les participants ont préféré la méthode *BWS*, même s’il est selon eux plus difficile de faire un choix dans le cas où les sons présentés sont très similaires. La méthode *BWS* semble néanmoins un paradigme adapté pour l’étude de la perception sonore, et pourrait présenter une alternative à la méthode *RS* pour l’évaluation de critères subjectifs comme l’agrément.

2.5.2 Algorithmes de notation : mise à jour des scores d’un grand nombre d’éléments

Lors d’une expérience faisant appel à la méthode *BWS*, un certain nombre d’éléments sont comparés et jugés les uns par rapport aux autres selon un critère énoncé. Une fois les jugements des participants recueillis, plusieurs algorithmes peuvent être utilisés pour calculer le rang (ou le score) de chaque élément. Certains d’entre eux s’étant révélés coûteux en termes de temps de calcul, Hollis met en avant la nécessité de développer de nouveaux algorithmes qui puissent convertir les jugements de la méthode *BWS* en scores pour un grand nombre d’éléments. Il présente alors dans son étude le fonctionnement de différents algorithmes de notation, dont les performances sont validées empiriquement grâce à des simulations. Parmi ces algorithmes, trois ont révélé de bons résultats lors des différentes expériences [13].

Le premier est l’algorithme de notation des tournois développé par Arpad Elo [10], dont le fonctionnement s’apparente à un tournoi entre les éléments, qui sont vus comme autant de concurrents qui s’affrontent lors de matchs. Lorsqu’un participant est confronté à un groupe de N éléments et qu’il doit en choisir un meilleur et un pire, il est alors possible d’imaginer un tournoi entre les N éléments. Les relations entre les éléments représentées par les inégalités ci-dessus correspondent alors aux issues de certains matchs, et l’inégalité $A > B$ peut être interprétée comme “victoire du concurrent A face au concurrent B ”, ce qui correspond à “l’élément A a été jugé meilleur que l’élément B ”. Pour chacune de ces relations, le score des deux concurrents concernés est mis à jour en fonction de l’écart entre leurs scores avant le match : un concurrent avec un score élevé voit son score augmenter légèrement s’il gagne contre un concurrent avec un score bas, tandis qu’un concurrent avec un score bas voit son score augmenter beaucoup s’il gagne contre un concurrent avec un score élevé. Cette méthode prend donc en compte non seulement le résultat d’un jugement (ou d’un match), mais aussi l’écart de “compétence” entre les concurrents, c’est-à-dire la probabilité de chaque concurrent de gagner face à un autre. Ainsi, pour un concurrent A , son rang R_A est mis à jour après un match en fonction de l’écart entre son score réel S_A et son score attendu E_A , multiplié par un facteur K , comme présenté dans l’équation (1).

$$R'_A = R_A + K(S_A - E_A) \quad (1)$$

Pour déterminer le score attendu d’un concurrent, le système de notation Elo suppose que la compétence des concurrents suit une distribution normale avec un écart-type de 100 sur l’échelle Elo, et

que c'est cette compétence qui détermine principalement l'issue des matchs entre deux concurrents. Par conséquent, pour deux concurrents A et B de compétences égales, il est supposé que A et B gagnent l'un contre l'autre un nombre égal de fois. À l'inverse, si A possède un score plus élevé que B , alors A est plus susceptible de gagner. À partir de ces hypothèses, la probabilité attendue qu'un concurrent A gagne contre un concurrent B est définie par l'équation (2) [10] :

$$E_A = \frac{10^{\frac{R_A}{400}}}{10^{\frac{R_A}{400}} + 10^{\frac{R_B}{400}}} \quad (2)$$

La règle de mise à jour de l'algorithme Elo est sensible à la différence de rang entre les deux éléments concernés. De plus, la complexité pour les mises à jour des scores est constante, ce qui ne limite pas cet algorithme en termes d'efficacité computationnelle, et le rend pertinent lorsque de nombreux éléments doivent être notés.

Un deuxième algorithme, cette fois dit "discriminant", est basé sur le modèle de Rescorla-Wagner [25], qui met à jour des associations entre des indices et des événements. Lorsqu'un indice survient (par exemple un concurrent particulier) et est suivi d'un événement spécifique (par exemple ce concurrent gagne un match face à un autre), la force d'association V_x^n entre l'indice et l'événement est mise à jour pour refléter la relation observée grâce à l'équation (3) suivante :

$$V_x^{n+1} = V_x^n + \Delta V_x^{n+1} \quad (3)$$

Le degré de changement dû à chaque paire indice-événement ΔV_x^{n+1} dépend de plusieurs paramètres :

- la saillance de l'association observée notée α ,
- le taux d'apprentissage pour l'indice noté β ,
- la force d'association maximale pour l'événement notée λ ,
- la force d'association totale entre l'indice et tous les événements, notée V_{tot} .

Les équations (4) et (5) suivantes permettent respectivement d'exprimer ce degré de changement, ainsi que la saillance d'une association :

$$\Delta V_x^{n+1} = \alpha_x \beta (\lambda - V_{tot}) \quad (4)$$

$$\alpha = 1 - \frac{V_w(\text{meilleur})}{V_w(\text{meilleur}) + V_w(\text{pire})} \quad (5)$$

La règle de mise à jour de Rescorla-Wagner peut intégrer des informations comme celle de l'algorithme Elo sur la qualité d'un jugement via le terme de saillance (lorsqu'un élément A gagne contre un élément B avec un score beaucoup plus faible, le résultat est moins saillant que si B avait un score beaucoup plus élevé). L'équation (5) suppose que les événements inattendus sont saillants, en estimant la probabilité que A gagne contre B comme $\frac{V_w(A)}{V_w(A) + V_w(B)}$, avec $V_w(A)$ la force d'association entre A et l'événement " A gagne le match", et $V_w(B)$ la force d'association entre B et l'événement " A gagne le match". Dans le cas où aucun des éléments n'a d'association préalable avec l'événement cité, le terme de saillance est fixé à 0.5. En comparaison avec l'algorithme de notation Elo, qui donne des scores qui n'ont pas de limites supérieures ou inférieures (tant qu'un élément particulier gagne de manière cohérente, des distinctions de plus en plus grandes peuvent être faites entre lui et le reste des concurrents), les forces d'association du modèle de Rescorla-Wagner sont limitées par une force de conditionnement maximale

théorique possible entre un indice et un événement, spécifiée par le paramètre λ .

Enfin, un troisième algorithme de notation appelé *Value Learning* est introduit, qui consiste en l'apprentissage de l'issue attendue d'un match pour chaque concurrent. Le classement du tournoi peut ensuite être déterminé en ordonnant les concurrents en fonction des valeurs de leur score attendu lorsqu'ils participent à un match. La valeur du score du concurrent A , notée V_A , est calculée en mettant à jour le score estimé du concurrent en fonction des résultats observés lors des matchs auxquels il participe (représentés par le paramètre γ), comme présenté par l'équation (6).

$$V_A = V_A + \alpha\beta(\gamma - V_x) \quad (6)$$

La règle de mise à jour de l'algorithme *Value Learning* est similaire à celle de Rescorla-Wagner, car les scores sont déduits à partir de la différence entre ce qui était attendu et ce qui a réellement été observé. Cependant, *Value Learning* se concentre sur l'issue observée d'un événement (victoire ou défaite), plutôt que sur une association entre un indice et un événement. Les mêmes paramètres sont néanmoins retrouvés, à savoir le taux d'apprentissage β et la saillance α . Ainsi, les chances de victoire O_A du concurrent A sont calculées en considérant son score attendu comme une probabilité de victoire, et en la convertissant en une valeur de cote, comme le montre l'équation (7). La saillance de l'issue est ensuite calculée selon l'équation (8), qui assure que celle-ci soit entre 0 et 1, et qu'elle augmente à mesure que le résultat observé devient inattendu. Dans le cas où deux concurrents n'ont encore jamais gagné de match, la saillance est fixée à 0.5.

$$O_A = \frac{V_A}{1 - V_A} \quad (7)$$

$$\alpha = 1 - \frac{O_A}{O_A + O_B} \quad (8)$$

Une des limitations de cette méthode est que les valeurs des différentes issues des matchs (victoire, défaite, etc.) doivent être définies à l'avance. Toutefois, la valeur absolue de ces résultats n'est pas ce qui importe le plus; c'est la relation entre les différentes valeurs qui est essentielle : qu'une victoire vaille 1 ou 10 n'a pas d'importance tant que la défaite vaut 0 et qu'un match nul se situe quelque part entre les deux.

Finalement, chacun de ces trois algorithmes permet d'évaluer des scores pour un grand nombre d'éléments. Dans le contexte particulier d'une étude impliquant la méthode *BWS*, les simulations réalisées par Hollis pourront se révéler utiles pour le choix de l'algorithme de notation le plus efficace et le plus adapté. Une fois les scores calculés, ceux-ci peuvent être interprétés en relation avec les caractéristiques des éléments notés, et des tests statistiques peuvent être utilisés sur des groupes d'éléments pour confirmer ou réfuter certaines hypothèses de recherche.

3 Construction d’une nouvelle approche de masquage

Les notions qui serviront de socle à l’axe *Auditory Concealer* étant maintenant définies, la question est de savoir comment les mettre en œuvre dans le cadre du projet *ReNAR*. Cette partie détaillera toute la réflexion qui s’est menée autour de ces notions et présentera les aspects théoriques et pratiques de l’approche *concealer*, dans le but de la comparer à l’approche *masker*.

3.1 Objectifs et enjeux du projet *ReNAR*

Au vu de l’objectif premier du projet *ReNAR*, qui est de proposer une solution à l’exposition au bruit des personnes dans leurs espaces de vie, la notion de confort a été identifiée comme notion centrale du projet, et guidera la réflexion autour du bruit dans les bureaux ouverts. Aussi, le projet n’ayant pas pour visée de réduire physiquement le bruit mais bien de modifier la perception des usagers, le mot “confort” sera utilisé dans la suite pour parler de “confort sonore”, tel que défini dans l’état de l’art (section 2).

Aussi, ayant choisi d’étudier le bruit dans les environnements spécifiques que sont les bureaux ouverts, il est important de garder à l’esprit que le confort est étroitement lié à différents éléments de contexte correspondant à ce type d’environnement. Dans ce cadre d’étude, la tâche effectuée par les usagers, les sources constituant l’ambiance sonore du lieu, ou encore la sensibilité au bruit de chacun sont autant d’éléments de contexte à prendre en compte et à ne pas négliger. Dans le modèle de confort de Vink et Hallbeck [30] repris par Duroyon [8] présenté en figure 1, une place importante est donnée aux attentes de l’utilisateur, car celles-ci influent sur son jugement du produit étudié (jugement séparé en trois cas dans le modèle : confort, inconfort, et neutre). De plus, ce modèle suggère l’étude de produits comme objets manufacturés, ce qui ne correspond pas au cadre du projet *ReNAR*, qui traite plutôt d’un environnement. Dans le cas particulier des bureaux ouverts, il est néanmoins possible d’assimiler un environnement à un produit, dont les usagers sont des travailleurs avec certaines attentes. Celles-ci pourraient par exemple se formuler de la manière suivante : “je m’attends à ce que mon environnement de travail soit propice à la concentration et au calme”, ou bien en ajoutant plus de contexte : “en temps que membre d’une équipe, je m’attends à ce que la communication soit facilitée sur mon lieu de travail”. L’évaluation du confort concerne alors l’ambiance sonore de l’environnement étudié dans son ensemble, et ce en cohérence avec le contexte dans lequel cette ambiance est soumise aux usagers de l’espace.

Cela étant dit, le premier axe *Auditory Concealer* du projet *ReNAR* se propose de développer une nouvelle approche de masquage, nommée approche *concealer*, dont le but est de modifier les environnements sonores perçus par les usagers de différents espaces en habillant les sons indésirables (c’est-à-dire jugés gênants ou désagréables). Cette approche de réalité augmentée sonore jouera pour cela sur les propriétés spectro-temporelles des sons, et devra s’efforcer de minimiser l’augmentation du niveau sonore global de la scène étudiée. En effet, le fait que le niveau sonore soit le premier facteur responsable de la gêne et du désagrément impose une attention particulière sur ce paramètre, qui constituera un facteur central lors de la construction de l’approche *concealer* dans les parties suivantes.

3.2 Définition des approches *masker* et *concealer*

Dans le cadre d’environnements de travail, il est intéressant de s’appuyer sur l’étude menée par Cai et coll., qui vise à évaluer l’effet du masquage sur la gêne induite par un bruit industriel [6]. L’approche de masquage utilisée vise à masquer une source de bruit en y ajoutant certains sons, afin de créer une

mixture qui se veut moins gênante, ou plus agréable. Une telle approche est en ce sens appelée approche *masker*, et se résume par le schéma présenté sur la figure 2 suivante :

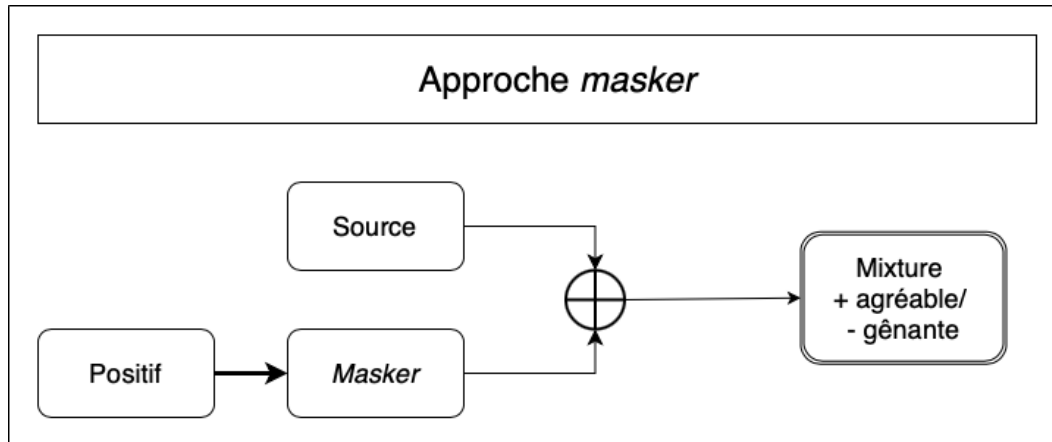


FIG. 2 – Schéma résumant le principe de l'approche masker

La scène sonore résultante de l'application de cette approche sur une source de bruit est alors une mixture entre ce bruit et un signal ajouté nommé *masker*. Ce dernier est construit à partir d'un son qualifié de "positif", c'est-à-dire considéré comme communément accepté voire apprécié. Dans le cas de l'étude de Cai et coll., la source choisie est un bruit de soudeuse électrique, et des sons d'eau (fontaine, pluie et cascade) sont utilisés comme *maskers*, car ceux-ci sont a priori considérés comme des sons positifs. L'addition se fait dans le domaine temporel, et il est intéressant de jouer sur le niveau sonore du *masker* pour trouver un compromis entre sa présence perçue dans la mixture et le niveau sonore global de la scène. Cai et coll. utilisent pour cela un rapport de niveau entre le bruit et le *masker*, appelé *Masking Sound to Noise Ratio (MSNR)*. Dans le cas de l'approche *masker*, le "*Masking Sound*" est le signal *masker*, et le "*Noise*" la source de bruit. De plus, le *MSNR* est exprimé en dB(A), de sorte qu'il soit positif lorsque le bruit est moins fort que le *masker*.

Pour l'approche *concealer*, le raisonnement est différent : l'idée est plutôt de compléter la source de bruit. Le signal ajouté, cette fois appelé *concealer*, est construit à partir du même positif que pour l'approche *masker*, mais aussi à partir de la source. Source et *concealer* sont ensuite ajoutés pour donner une mixture qui se veut elle aussi plus agréable et moins gênante. Cette nouvelle approche est présentée sur la figure 3 ci-dessous :

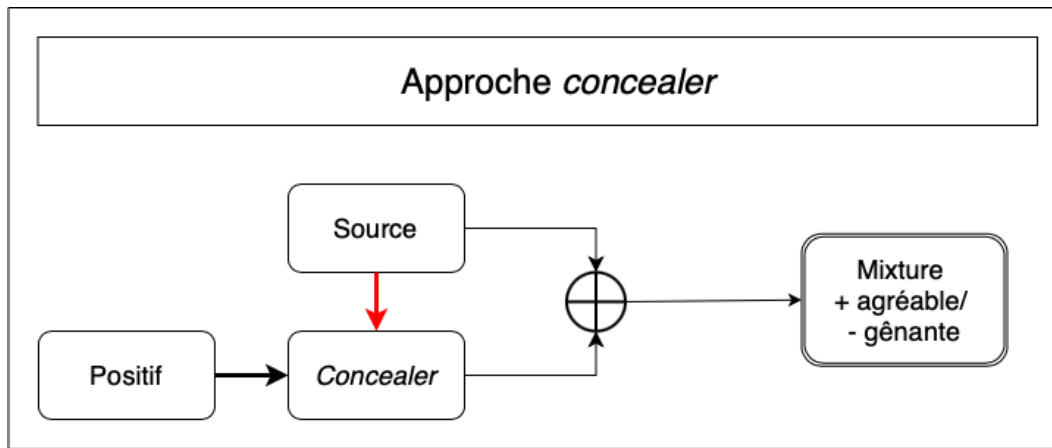


FIG. 3 – Schéma résumant le principe de l'approche *concealer*

Contrairement à l'approche *masker*, l'approche *concealer* permettrait de compléter une source pour garantir une mixture plus acceptée, voire agréable si le positif est bien choisi, mais aussi de travailler à des niveaux ajoutés moins importants. Puisque le signal *concealer* dépend de la source, il est également important de prendre en compte le contenu fréquentiel de celle-ci ainsi que celui du positif pour construire un bon *concealer*. Par exemple, si la source à habiller est le bruit d'un système de ventilation, il est difficilement concevable de choisir un chant d'oiseau composé uniquement de hautes fréquences comme positif, car le *concealer* se trouverait lui aussi composé de hautes fréquences, rendant la mixture entre lui et la source très hétérogène. En revanche, un son de vagues, dont le contenu fréquentiel est similaire à celui du son de ventilation, peut représenter une bonne option pour le positif, rendant une mixture plus homogène. À la création du *concealer*, les sons de ventilation et de vagues peuvent en effet se compléter l'un l'autre. La question du choix des positifs est donc primordiale pour que l'approche *concealer* permette d'améliorer le confort dans certains environnements, dont celui des bureaux ouverts.

Finalement, concernant les deux approches *masker* et *concealer*, le point essentiel est de construire des mixtures mieux perçues que le bruit. Cela requiert une bonne définition des signaux *masker* et *concealer*, tous les deux dépendant du positif. Une première étape consiste alors à s'assurer que le positif est effectivement perçu comme agréable, et ce dans un contexte bien défini. La perception d'un son peut en effet changer selon l'environnement et la situation dans lesquels il sont expérimentés. Une musique avec paroles peut par exemple paraître agréable dans un contexte de loisirs, mais être perçue comme moins agréable voire gênante dans un contexte de travail.

3.3 Choix de la source et du positif

Maintenant que les deux approches sont théoriquement bien définies, il faut être en mesure de les implémenter. Pour cela, il est nécessaire de connaître la source de bruit et le positif, qui permettront de créer les signaux *masker* et *concealer* (ceux-ci pourront être regroupés sous le terme de "maquilleur" pour parler du signal ajouté indépendamment des deux approches). Ce faisant, les différents signaux seront notés de la manière suivante dans la suite :

- s , la source de bruit;
- m , le *masker*;
- c , le *concealer*;
- p , le positif.

3.3.1 Choix de la source

Pour ce qui est de la source, il a été décidé d'étudier le bruit d'un système de ventilation. Dans le contexte des bureaux ouverts, cette source sonore est en effet très pertinente par sa présence quasiment permanente dans ce type d'espace. De plus, le caractère stationnaire de ce bruit a permis de réaliser de premiers essais assez simples avec les deux approches de masquage (un bruit présentant de grandes variations d'amplitude ou dans son contenu fréquentiel aurait en effet été beaucoup plus difficile à traiter dans un premier temps).

Deux bruits de ventilation ont été sélectionnés. Le premier est le bruit du système de ventilation présent dans les bureaux du laboratoire STMS (équipe Perception et Design Sonores) au sein de l'IRCAM. Celui-ci a été enregistré de différentes manières :

- en champ proche avec un *Zoom H4N* (Figure 4);
- en champ proche avec un microphone *Schoeps* muni d'une capsule *MK5* réglée en position cardioïde et d'un pré-ampli *CMC5* (Figure 4);
- à la position d'un auditeur avec deux microphones *Schoeps* présentant les mêmes caractéristiques décrites précédemment (ceux-ci étant placés au niveau des oreilles de l'auditeur si celui-ci était assis à un bureau) (Figure 5).



FIG. 4 – *Prise de son en champ proche du système de ventilation présent dans les bureaux du laboratoire STMS (équipe Perception et Design Sonores)*



((a)) *Vue de profil*



((b)) *Vue de face*

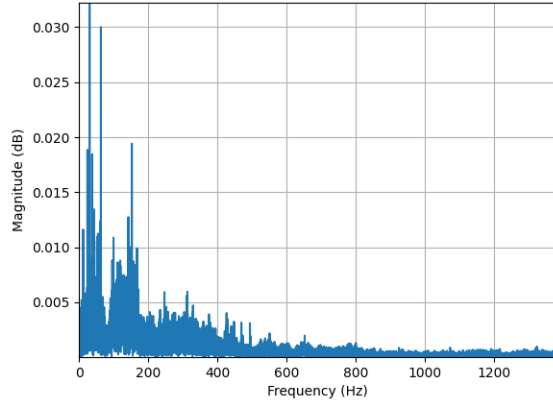
FIG. 5 – *Prise de son du système de ventilation présent dans les bureaux du laboratoire STMS (équipe Perception et Design Sonores) au niveau d’un auditeur assis à son bureau*

Une mesure au sonomètre a également été réalisée afin de caractériser le niveau auquel peuvent être soumis les usagers du bureau. Le sonomètre a pour cela été placé à la place d’un usager, et un niveau L_{Aeq} moyen de $33dB(A)$ a été recueilli sur deux minutes de mesure.

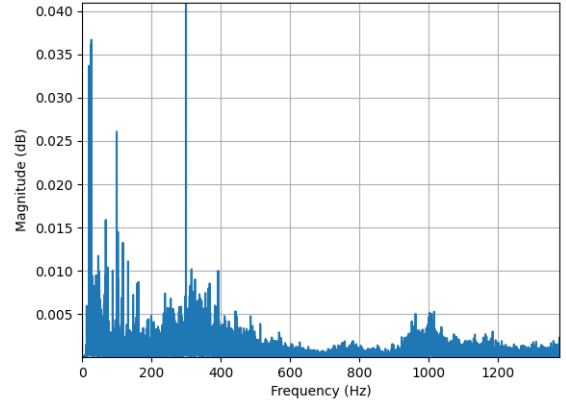
Par ailleurs, il a semblé que cette source en particulier n’était pas très représentative de l’aspect gênant que peuvent avoir certains bruits de ventilation. Il a donc été décidé d’en sélectionner une autre dans le corpus sonore utilisé dans l’étude de Susini et coll., qui vise à caractériser la qualité sonore des bruits de ventilation [28]. Ce deuxième bruit de ventilation présente une composante tonale continue en plus de la soufflerie basse fréquence, comme présenté sur la figure 6¹.

Le terme “source” évoqué dans les deux approches de masquage présentées précédemment (Figures 2 et 3) sera maintenant utilisé pour faire référence aux deux bruits de ventilation ainsi choisis. Pour les différencier, ceux-ci seront nommés “ventil1” pour le bruit enregistré dans le bureau de l’équipe Perception et Design Sonores, et “ventil2” pour le bruit provenant du corpus de Susini et coll. [28].

1. Les deux sons de ventilation sont contenus dans les fichiers *10s_ventil-ircam_65dB(A).wav* et *10s_c12dL_65dB(A).wav*, disponibles dans le dossier “sounds/ventilation_noises/”.



((a)) Spectre de “ventil1”



((b)) Spectre de “ventil2”

FIG. 6 – Spectre des deux bruits de ventilation

3.3.2 Choix du positif

Les sources ainsi définies, il a fallu déterminer le(s) positif(s) à utiliser pour la création des signaux *masker* et *concealer*. Pour cela, il a été décidé de s’inspirer de l’étude sur l’effet de masquage par les sons d’eau de Cai et coll. [6], qui fournit des résultats pour l’approche *masker*. Celle-ci a pu aiguiller la réflexion autour de la construction des mixtures avec les deux approches. Pour les sons de fontaine, de pluie et de cascade, il a en effet été montré que la mixture avec le bruit de soudeuse électrique était la moins gênante pour des valeurs de *MSNR* entre -3 et 0dB(A) , ce qui correspond à des niveaux de sons d’eau inférieurs à celui de la source de bruit [6]. Le fait que les sons d’eau n’aient pas besoin d’être à de hauts niveaux pour rendre la mixture moins gênante correspond bien à l’idée de minimiser le niveau de la mixture produite par l’approche *concealer*.

Par ailleurs, certains sons d’eau ont été classés comme communément acceptés par la littérature, et notamment par Maria Rådsten-Ekman lors de sa thèse et de plusieurs études visant à déterminer si les sons d’eau peuvent être utilisés pour améliorer la perception du bruit environnemental en milieu urbain [22] [21]. Il semble alors pertinent de les utiliser comme positifs, afin d’espérer créer des mixtures agréables, et de comparer les deux approches.

Cela étant dit, cinq sons d’eau ont été retenus après avoir parcouru les banques de sons disponibles à l’IRCAM. Parmi eux, des sons de pluie, de cours d’eau et de cascade (nommés *rain*, *stream* et *waterfall* dans la suite) ont été sélectionnés dans la banque *Sound Ideas* [24], et des sons de fontaine et de vagues (*fountain* et *waves*) ont été sélectionnés dans la banque *Blue Box* [23]². Il est également important de noter que le choix des sons d’eau a été fait dans l’optique d’une première expérience assez exploratoire. Une impasse a donc été faite sur le problème de congruence lié au contexte des bureaux ouverts. Comme précisé dans la section 4.1, il est essentiel de choisir des sons cohérents avec l’espace dont on veut améliorer le confort. Si les sons d’eau n’entrent pas forcément dans les attentes des usagers des bureaux ouverts, il reste tout de même intéressant de les étudier pour confirmer ou non leur statut de sons positifs dans ce contexte.

2. Les sons d’eau sont également disponibles dans le dossier “sounds/water_sounds/”, dans les fichiers *10s_fountain.wav*, *10s_rain.wav*, *10s_stream.wav*, *10s_waterfall.wav* et *10s_waves2.wav*.

Le choix des sons d'eau a finalement été motivé par leur diversité au niveau spectral. En plus de la comparaison globale entre les deux approches, les différents types de sons d'eau pourront en effet présenter un facteur secondaire influençant la perception des mixtures créées avec les deux approches. En particulier, cette diversité pourra fournir des informations plus précises quant aux conditions d'utilisation de l'approche *concealer*.

3.4 Construction du signal *concealer*

Les signaux *maskers* correspondant simplement aux positifs choisis ($m = p$), il est d'ores et déjà possible de mettre en œuvre l'approche *masker* en effectuant l'addition $s + m$ dans le domaine temporel. Mais concernant l'approche *concealer*, il reste à construire c , qui dépend à la fois du positif et de la source. Pour rendre la mixture résultante de l'addition $s + c$ agréable, l'idée est que celle-ci s'approche le plus possible du positif. Pour cela, plusieurs méthodes ont été implémentées, impliquant chacune une construction différente du contenu fréquentiel du *concealer* c .

3.4.1 Définition des méthodes de construction

L'idée est de trouver c tel que

$$s + c = \tilde{p}$$

avec \tilde{p} une estimation du positif p .

Pour obtenir \tilde{p} , il est possible d'utiliser les signaux s et p pour déduire simplement c en opérant la soustraction

$$c = p - s$$

En revanche, il ne faut pas oublier que ces signaux sont définis dans le domaine temporel. Or, pour tenter d'approcher p , il est nécessaire de jouer sur la structure spectro-temporelle de c , afin d'ajouter le contenu fréquentiel nécessaire à cette reconstruction. Les signaux s et p sont donc placés dans le domaine temps-fréquence ($s \rightarrow S$ et $p \rightarrow P$) pour déterminer l'amplitude et la phase du *concealer* C . Il suffira ensuite de repasser C dans le domaine temporel pour obtenir c et l'ajouter à la source.

Pour calculer l'amplitude m_C et la phase a_C du *concealer*, une transformée de Fourier à court-terme (TFCT) est calculée pour les signaux s et p grâce à la bibliothèque *librosa* disponible sous *Python* [18]. L'amplitude du *concealer* correspond alors à la soustraction entre les amplitudes du positif et de la source, et sa phase est fixée à la phase du positif, comme suit :

$$\begin{cases} m_C = m_P - m_S \\ a_C = a_P \end{cases}$$

avec m et a désignant respectivement l'amplitude (magnitude) et la phase (angle) des signaux d'une part, et S , P et C désignant respectivement les représentations spectrales des signaux source, positif, et *concealer* d'autre part.

Au vu des équations ci-dessus, un problème se présente au niveau des amplitudes : pour certaines fréquences, il est possible d'avoir $m_P > m_S$, et donc $m_C < 0$, ce qui est mathématiquement incorrect puisque l'amplitude est définie comme la valeur absolue du spectre d'un signal. Plusieurs méthodes ont été envisagées pour pallier ce problème, et construire une amplitude m_C positive.

Avant de chercher à satisfaire la condition $m_C \geq 0$, une première méthode a été définie de sorte à construire le signal *concealer* avec une amplitude qui peut être négative. Si cela peut paraître absurde d'un point de vue mathématique, il a tout de même été intéressant d'écouter le résultat après retour dans le domaine temporel. Un exemple de la forme de l'amplitude d'un tel *concealer* est représentée en vert sur la figure 7 ci-dessous :

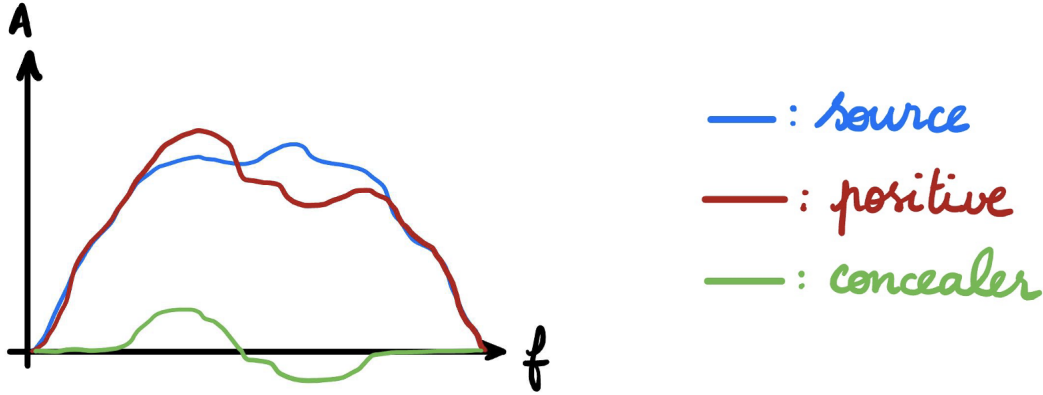


FIG. 7 – Schéma présentant l'amplitude du signal *concealer* en fonction de la fréquence (en vert), résultat de la soustraction des amplitudes du positif (en rouge) et de la source (en bleu)

Cette première méthode de construction a permis de réaliser de premiers essais et de déterminer m_C pour les sources et les positifs choisis en sections 3.3.1 et 3.3.2.

Afin d'éviter une amplitude négative, une deuxième méthode a cette fois consisté à appliquer une valeur absolue à la soustraction, de sorte que $m_C = |m_P - m_S|$, comme présenté sur la figure 8 suivante :

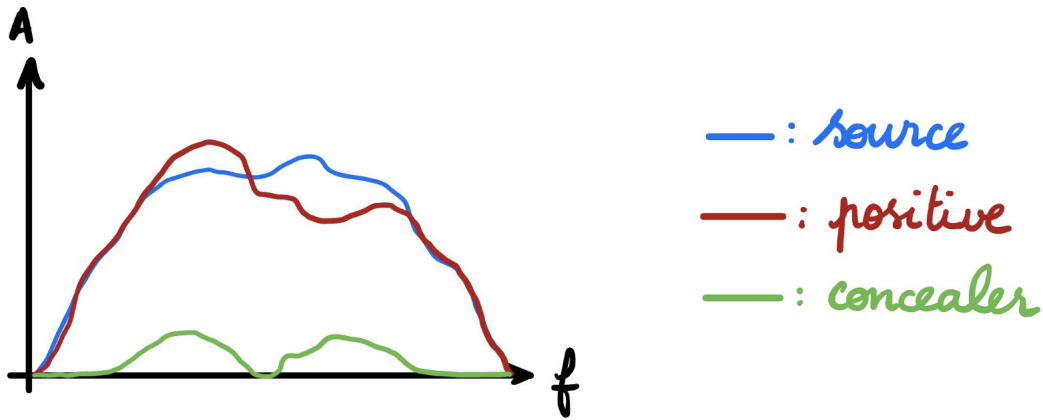


FIG. 8 – Schéma présentant l'amplitude du signal *concealer* en fonction de la fréquence (en vert), résultat de la valeur absolue de la soustraction des amplitudes du positif (en rouge) et de la source (en bleu)

Cette construction a permis de garantir une amplitude positive, mais ne représente pas réellement les dynamiques entre l'amplitude du positif et celle de la source. En effet, dans le cas où $m_P < m_S$, il y a moins d'énergie dans le spectre du positif que dans celui de la source. Il n'est donc pas possible d'approcher le spectre du positif en ajoutant du contenu fréquentiel à la source, comme le suggère la

figure 8 ci-dessus (le fait d'avoir $m_C > 0$ signifiant un ajout de contenu fréquentiel à la source).

Pour rectifier cela, une troisième et dernière méthode de construction du *concealer* a été envisagée, qui consiste à appliquer la soustraction $m_P - m_S$ seulement lorsque $m_P \geq m_S$. L'amplitude m_C est alors définie comme suit :

$$m_C = \begin{cases} m_P - m_S & \text{si } m_P \geq m_S \\ 0 & \text{sinon.} \end{cases}$$

Avec une telle méthode, une amplitude positive est garantie, et aucun ajout de contenu fréquentiel n'est effectué lorsque $m_P < m_S$, comme présenté sur la figure 9 ci-après.

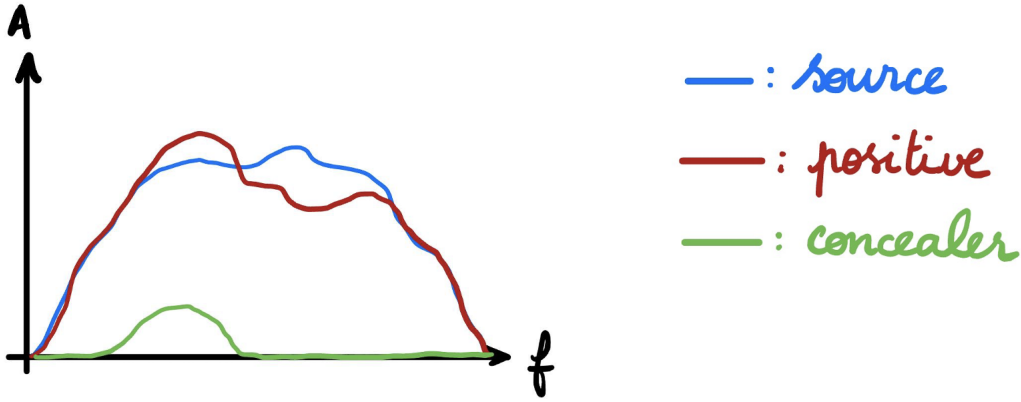


FIG. 9 – Schéma présentant l'amplitude du signal concealer en fonction de la fréquence (en vert), résultat de la soustraction non négative des amplitudes du positif (en rouge) et de la source (en bleu)

Un inconvénient à cette méthode reste le fait qu'en pratique, le spectre de la source et du positif n'est jamais aussi lisse que sur la figure 9. Si la différence $m_P - m_S$ change très souvent de signe, le spectre du signal *concealer* contiendra beaucoup de zéros.

Finalement, grâce aux trois méthodes présentées (notées *concealer-1*, *concealer-2* et *concealer-3* dans la suite), plusieurs options ont été considérées pour la construction du signal *concealer*. Il reste alors à reconstruire le signal *concealer*, à le repasser dans le domaine temporel, puis à l'additionner avec la source. La reconstruction C du signal *concealer* dans le domaine fréquentiel se calcule comme suit :

$$C = m_C * e^{i*a_C}$$

L'application d'une TFCT inverse (toujours grâce à *librosa*) sur C permettra ensuite de passer le signal dans le domaine temporel, pour finalement effectuer l'addition $s + c$, représentant la mixture résultante de l'approche *concealer*.

3.4.2 Choix de la méthode

Afin d'évaluer une seule version de l'approche *concealer*, il a fallu choisir l'une des méthodes de construction du signal *concealer* présentées en section 3.4.1. Pour cela, une série de mixtures a été générée grâce à un protocole proposé par la bibliothèque *doce*, disponible sous *Python* et développée par Mathieu Lagrange [15]. Celle-ci permet de définir des facteurs, puis de générer toutes les combinaisons

de facteurs grâce à une fonction appelée à chaque étape du protocole. Afin d'étudier les différentes méthodes de construction du signal *concealer*, les facteurs suivants ont été définis :

$$\left\{ \begin{array}{ll} s = & [\textit{ventil1}] \\ p = & [\textit{fountain}, \textit{rain}, \textit{stream}, \textit{waterfall}, \textit{waves}] \\ \textit{method} = & [\textit{concealer-1}, \textit{concealer-2}, \textit{concealer-3}, \textit{masker}] \\ \textit{msnr} = & [0, -1] \end{array} \right.$$

Ainsi, chaque mixture entre le bruit de ventilation enregistré à l'IRCAM et les sons d'eau sélectionnés a été générée pour toutes les méthodes (y compris l'approche *masker*, représentée par le facteur "masker") et les deux valeurs de *MSNR* calculés en dB LUFS³.

À l'écoute de ces mixtures, il est clair que la troisième méthode est meilleure que les méthodes *concealer-1* et *concealer-2*, pour lesquelles la ventilation apparaît beaucoup plus forte que le son d'eau. Ce dernier est en effet beaucoup plus audible avec la méthode *concealer-3*, créant des mixtures qui n'augmentent pas le niveau perçu de la ventilation. Cette différence entre les méthodes est également audible à l'écoute des *concealers* eux-mêmes. Pour les méthodes *concealer-1* et *concealer-2*, ceux-ci présentent une mixture entre son de ventilation et son d'eau, ce qui a pour effet d'ajouter une partie de la source lors de l'addition entre source et *concealer* et donc de rendre celle-ci prédominante sur le son d'eau. Pour la méthode *concealer-3* en revanche, le *concealer* ne présente que le son d'eau, ce qui implique un ajout de son positif uniquement lors de la mixture entre source et *concealer*.

Ce faisant, la méthode *concealer-3* a été retenue pour construire les différents signaux *concealers*. En plus de la comparaison entre méthodes, il est intéressant de comparer les mixtures créées avec la méthode *concealer-3* avec celles obtenues avec l'approche *masker*. En écoutant les différentes mixtures, il n'est apparu que très peu de différences entre les deux approches pour les niveaux étudiés, le seul contraste étant que le son d'eau paraissait plus brillant avec l'utilisation de l'approche *concealer*. Cette première comparaison entre les deux approches a permis de valider le potentiel de l'approche *concealer* face à l'approche *masker* pour les *MSNR* choisis.

3. Les mixtures générées ainsi que les *concealers* sont disponibles dans le dossier "*sounds/method_choice/*" et sont nommés de la manière suivante : *approach=1+msnr=0_fountain_stimuli.wav*.

4 Expérience perceptive

Maintenant que les deux approches sont fonctionnelles, il est possible de créer des mixtures entre source et *masker* (approche *masker*) et entre source et *concealer* (approche *concealer*). La prochaine et dernière étape du travail mené lors de ce stage est de les évaluer. Une expérience perceptive a été conduite, et cette dernière partie en résumera les objectifs et présentera la méthodologie employée ainsi que les premiers résultats obtenus.

4.1 Présentation de l'expérience

La première chose à définir pour une expérience perceptive est la dimension à évaluer. Si le projet *ReNAR* vise à améliorer le confort, il a été souligné dans l'état de l'art (section 2) que cette notion était difficile à évaluer. Une première option a donc été de tenter d'évaluer la gêne induite par les sources choisies. Cette notion de gêne semble d'autant plus pertinente à étudier dans le contexte d'un environnement de travail, car elle peut influencer sur le bien-être mais aussi sur la performance des usagers du lieu. En s'appuyant sur le protocole de l'étude de Brocolini et coll. [5], il aurait en effet été possible d'effectuer plusieurs mesures de performance, de charge de travail et de gêne perçue. En revanche, la définition d'une tâche de performance pertinente aurait nécessité plus ample préparation et le temps disposé pour réaliser l'expérience ne le permettait pas. De plus, une tâche de performance demande un temps d'exposition aux stimuli assez long, ce qui n'était pas vraiment compatible avec la diversité des paramètres prévus pour l'expérience, qui imposait un grand nombre de stimuli. Celle-ci aurait été trop longue, ou aurait nécessité une certaine organisation avec un grand nombre de participants ou un découpage en plusieurs sessions pour chaque participant.

Cela étant dit, il a été décidé de s'orienter vers la notion d'agrément, qui peut être évaluée plus rapidement avec une seule mesure d'agrément perçu. Cette notion semble être un bon compromis pour une première expérience exploratoire sur le thème de l'approche *concealer*.

Ainsi, l'objectif de l'expérience qui a été menée est de comparer les approches *masker* et *concealer* en termes d'agrément, afin d'évaluer le potentiel de l'approche *concealer* dans le traitement du bruit au niveau perceptif au sein d'un environnement intérieur tel que les bureaux ouverts. Une hypothèse générale selon laquelle l'approche *concealer* a des chances d'être meilleure que l'approche *masker* en termes d'agrément a donc été posée. Le facteur expérimental principal étant les deux approches, des facteurs secondaires comme le niveau, le type de source ou le type de positif ont été étudiés, afin d'espérer identifier certaines conditions sur ces facteurs pour lesquelles les mixtures générées grâce à l'approche *concealer* sont effectivement perçues comme plus agréables que celles générées par l'approche *masker*.

4.2 Méthodologie

Les objectifs de l'expérience étant posés, cette section présentera son déroulement, du recrutement des participants aux méthodes mises en œuvre.

4.2.1 Participants

Un total de 30 personnes a participé à cette expérience (9 femmes et 21 hommes), avec une distribution d'âge entre 22 et 57 ans, l'âge moyen étant de 30 ans et l'âge médian de 25 ans. Un questionnaire rempli à la fin de l'expérience a révélé des problèmes auditifs chez trois participants, à savoir des acouphènes pour deux d'entre eux et des pertes auditives au-dessus de 8kHz au niveau des deux oreilles pour

le dernier. Les résultats de ces participants n'ont pas été écartés. Les participants recrutés travaillaient pour la plupart à l'IRCAM, au sein du laboratoire STMS, ou encore dans les services de production, d'informatique, ou des ressources humaines. Quelques personnes extérieures à l'IRCAM ont aussi pris part à l'expérience.

4.2.2 Stimuli

Les stimuli ont été construits selon quatre facteurs indépendants : l'approche utilisée, le niveau des mixtures générées, le type de source, et le type de positif. Comme mentionné plus haut, l'approche utilisée (*masker* ou *concealer*) représente le facteur principal de l'expérience. Il reste donc à choisir les valeurs que prendront les trois facteurs secondaires avant de construire les stimuli.

Choix des types de source et de positif :

Ce sont les deux bruits de ventilation et les cinq sons d'eau sélectionnés en section 3.3 qui ont permis la création des stimuli suivant les deux approches de masquage⁴. Comme expliqué dans la section 3.3, les sons d'eau sont utilisés comme positifs pour les deux approches *masker* et *concealer*. Leur contenu spectral ainsi que ceux des deux sources sont présentés sur la figure 10 suivante.

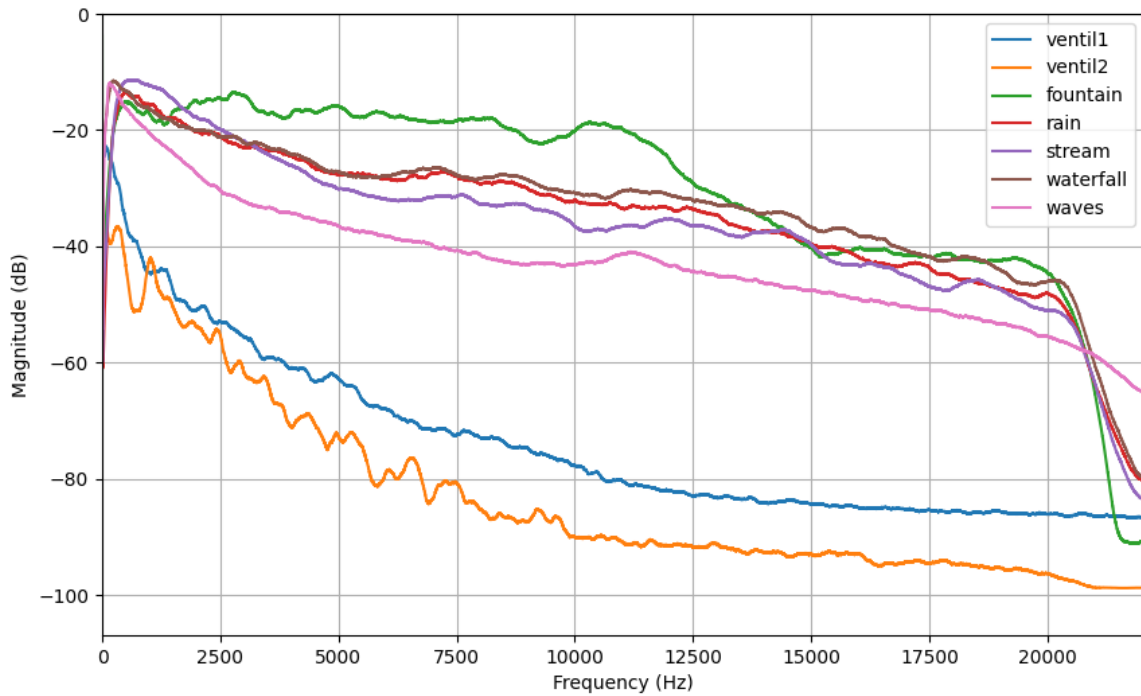


FIG. 10 – Spectres lissés des deux bruits de ventilation et des cinq sons d'eau pour une plage de fréquences allant de 0 à 20kHz

Les bruits de ventilation présentent de l'énergie principalement dans les basses fréquences, alors que celle des sons d'eau décroît petit à petit jusqu'à 20kHz. Pour détailler les endroits où tous les sons ont une énergie non négligeable, la figure 11 propose un zoom sur les basses fréquences.

4. Ces sons sont disponibles dans les dossiers *sounds/ventilation_noises/* et *sounds/water_sounds/*.

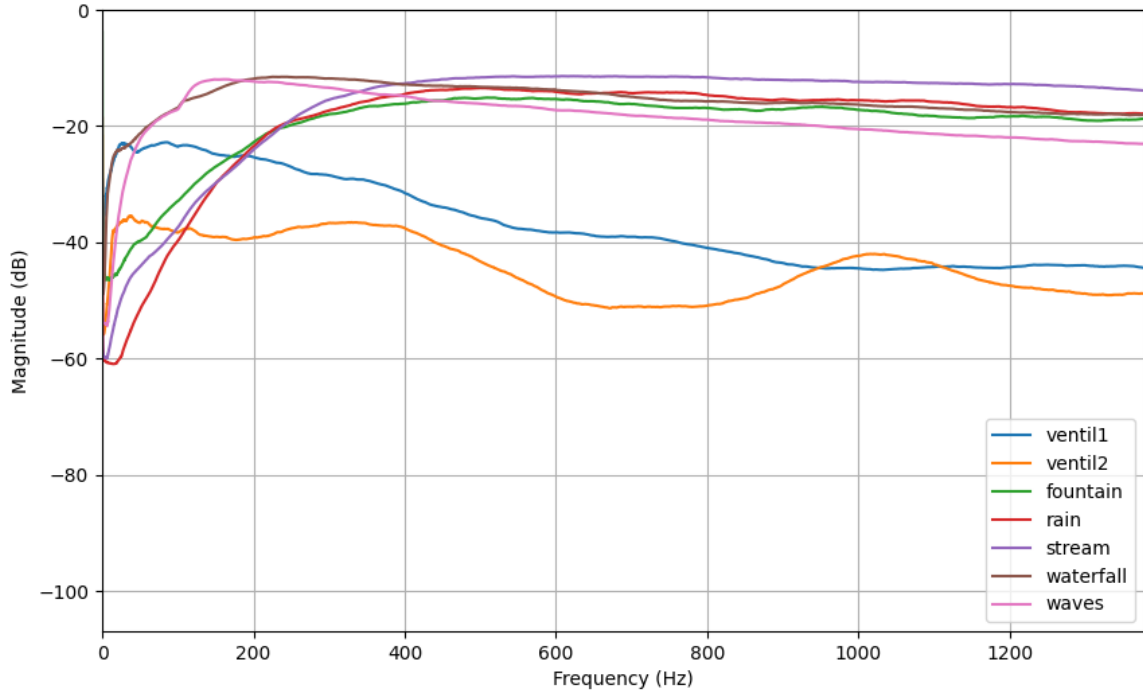


FIG. 11 – Spectres lissés des deux bruits de ventilation et des cinq sons d'eau pour une plage de fréquences allant de 0 à 1200Hz

Il est maintenant plus facile d'observer que les sons *fountain*, *rain* et *stream* n'ont que très peu d'énergie dans la plage couvrant les fréquences jusqu'à environ 200Hz, contrairement aux deux bruits de ventilation, pour lesquels l'énergie décroît au-delà de cette fréquence. Restent les sons *waterfall* et *waves*, dont l'énergie dans les très basses fréquences dépasse celle des bruits de ventilation.

Choix du niveau des stimuli :

Le niveau des stimuli représente un facteur secondaire mais essentiel pour cette étude, car le projet *ReNAR* repose sur un paradigme de minimisation de l'intensité sonore ajoutée. Il est alors nécessaire de s'assurer que le niveau de la mixture entre le bruit et le maquilleur ne se trouve pas nettement supérieur au niveau du bruit seul, et ce pour les deux approches.

Grâce aux premiers essais réalisés avec le paramètre *MSNR* lors du choix de la méthode de construction du signal *concealer* (section 3.4.2), une réflexion s'est entamée sur la manière de contrôler le niveau non pas du maquilleur utilisé, mais plutôt de la mixture générée par les deux approches. Pour assurer directement un niveau des mixtures qui ne soit pas nettement supérieur à celui de la source, il a alors été décidé de raisonner selon un autre paramètre de niveau, noté ΔL_{Aeq} , représentant la différence de niveau entre une mixture et la source utilisée pour cette mixture (en dB(A)). Contrairement au *MSNR*, ce nouveau paramètre est positif ou nul (il n'est pas possible que le niveau d'une mixture soit inférieur à celui de la source puisque celle-ci est définie par l'addition entre la source et un maquilleur), et permet de contrôler directement le niveau de la mixture qui sera présentée comme stimuli aux participants de l'expérience perceptive.

Tout comme pour le choix de la méthode de construction du *concealer*, un essai a été réalisé grâce au protocole proposé par la bibliothèque *doce* [15] avec les facteurs suivants :

$$\begin{cases} s = & [\textit{ventil1}, \textit{ventil2}] \\ p = & [\textit{fountain}, \textit{rain}, \textit{stream}, \textit{waterfall}, \textit{waves}] \\ \textit{method} = & [\textit{concealer-3}, \textit{masker}] \\ \Delta L_{Aeq} = & [0, 1, 2, 3, 4, 5, 6] \end{cases}$$

Pour cet essai, la source a été normalisée à 65dB(A), car le niveau enregistré à 33dB(A) a paru trop faible pour se placer dans le contexte d’une source sonore désagréable. En effet, les études menées sur le masquage ont présenté des niveaux beaucoup plus élevés (notamment celle de Cai et coll., qui utilisent un bruit de soudeuse électrique à 80dB(A) [6]). Ainsi, un ΔL_{Aeq} de 0 impose un niveau de 65dB(A) pour les mixtures générées avec les deux approches. Plus précisément, il a fallu trouver le niveau du positif tel que le niveau de la mixture soit égal à $65 \pm 0.1\text{dB(A)}$ ⁵, et ce pour les deux approches.

Cette relation entre niveau du positif et niveau de la mixture a été étudiée et sauvegardée pour chaque son d’eau, comme présenté sur la figure 12 dans le cas du son de fontaine.

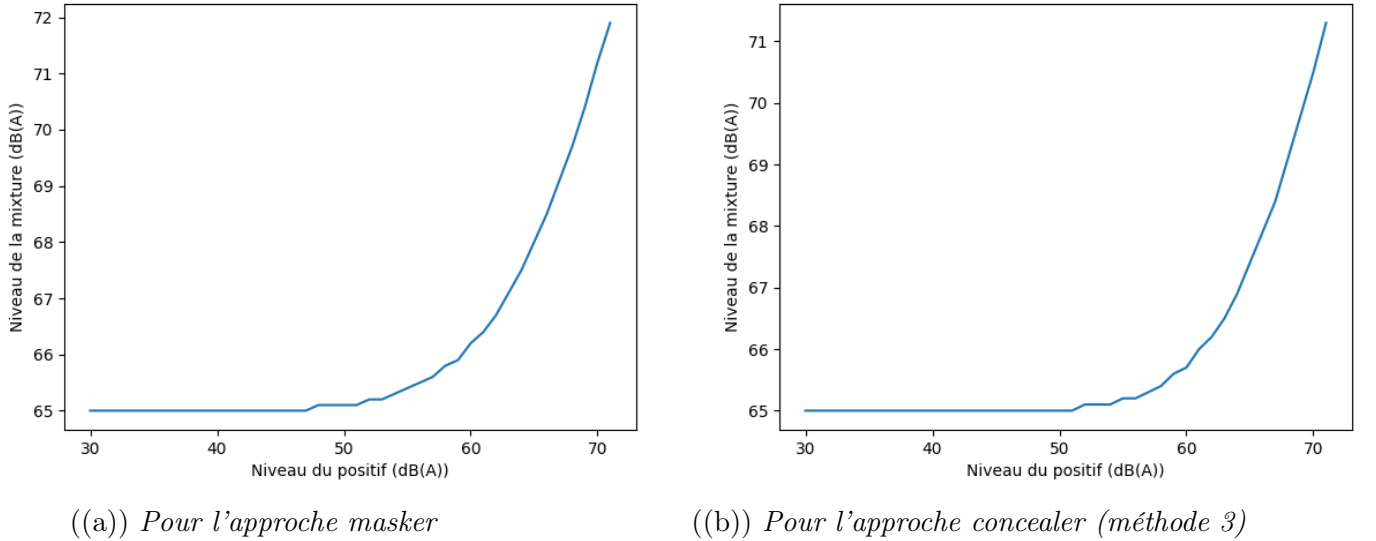


FIG. 12 – Niveau de la mixture en fonction du niveau du positif pour le bruit de ventilation enregistré à l’IRCAM et le son de fontaine

Pour déterminer le niveau du positif tel que le niveau de la mixture soit égal à $65 + \Delta L_{Aeq} \pm 0.1\text{dB(A)}$ (le niveau cible), une technique de *Grid Search* a été utilisée. Ce type de technique est un processus qui effectue une recherche exhaustive à travers un sous-ensemble spécifié manuellement. Un ensemble de niveaux de positifs a donc été défini avant de lancer une recherche telle que : pour chaque niveau de cet ensemble, une mixture est calculée et son niveau est comparé au niveau cible dépendant de la valeur de ΔL_{Aeq} . Si le niveau de la mixture est égal au niveau cible à 0.1dB(A) près, alors la recherche s’arrête.

Ainsi, le protocole *doce* et la recherche de type *Grid Search* ont permis d’une part de calculer les niveaux des positifs nécessaires à la création de mixtures dont le niveau $N_{mixture}$ est égal à $N_{source} + \Delta L_{Aeq}$,

5. Une erreur maximale de 0.1dB(A) a été autorisée sachant que le seuil de discrimination du niveau sonore est de l’ordre de 1dB(A), donc bien supérieur à 0.1dB(A).

et ce pour les deux sources et les cinq positifs retenus. D’autre part, ces mixtures ainsi que les *concealers* utilisés pour l’approche *concealer* ont été générés et enregistrés au format *wav*.

À l’écoute des mixtures générées⁶, il est intéressant de remarquer de grandes différences entre les valeurs de ΔL_{Aeq} . Par exemple, le son d’eau n’est quasiment pas perceptible pour $\Delta L_{Aeq} = 0\text{dB(A)}$, mais le devient assez clairement pour $\Delta L_{Aeq} = 1\text{dB(A)}$, surtout pour les sons *fountain* et *rain*. Aussi, plus ΔL_{Aeq} augmente, plus les mixtures se différencient selon les deux approches. En effet, pour des valeurs de $\Delta L_{Aeq} \geq 4\text{dB(A)}$, l’approche *concealer* produit des mixtures pour lesquelles le son d’eau semble beaucoup plus brillant que pour l’approche *masker*. Finalement, les niveaux obtenus pour les mixtures étant considérés comme trop forts pour $\Delta L_{Aeq} \geq 4\text{dB(A)}$, il a été décidé de garder des valeurs de ΔL_{Aeq} allant de 0 à 3dB(A), par pas de 0.5dB(A) afin d’affiner la perception du positif.

Construction des stimuli :

Les facteurs de l’expérience ainsi fixés, les stimuli ont pu être construits. Le niveau sonore équivalent pondéré A des sources de bruit a été fixé à 65dB(A), et la construction des stimuli a été entreprise grâce aux fonctionnalités de la bibliothèque *doce* [15] sous *Python*, qui ont permis de générer des mixtures entre les sources et les maquilleurs pour les deux approches. La moitié des stimuli a donc été construite selon l’approche *masker*, et l’autre moitié selon l’approche *concealer*.

De plus, au cours de la génération des mixtures, ces dernières ont été déclinées en plusieurs niveaux selon le paramètre ΔL_{Aeq} . Une valeur de $\Delta L_{Aeq} = 2.5\text{dB(A)}$ signifie donc que le niveau de la mixture générée est fixé à 67.5dB(A), soit 2.5dB(A) de plus que la source seule. Chaque paire constituée d’une source et d’un maquilleur a ainsi donné sept mixtures différentes, créant des stimuli de 10s enregistrés au format *wav* avec une fréquence d’échantillonnage de 44.1kHz.

Au final, un total de 140 stimuli a été généré selon les facteurs suivants :

$$\begin{cases} s = & [\textit{ventil1}, \textit{ventil2}] \\ p = & [\textit{fountain}, \textit{rain}, \textit{stream}, \textit{waterfall}, \textit{waves}] \\ \textit{method} = & [\textit{concealer-3}, \textit{masker}] \\ \Delta L_{Aeq} = & [0, 0.5, 1, 1.5, 2, 2.5, 3] \end{cases}$$

4.2.3 Apparatus

L’expérience a eu lieu dans une cabine audiométrique dans les locaux de l’équipe Perception et Design Sonores à l’IRCAM. Les stimuli ont été présentés grâce à un casque ouvert *Beyerdynamic DT 990 Pro* avec une impédance de 250Ω, connecté à un *MacMini*. Avant d’accueillir les participants, le casque a été calibré grâce à un sonomètre. Le réglage du niveau a été effectué à l’aide des boutons de volume de l’ordinateur, avec un volume retenu de 7 points, de telle sorte que les deux sources normalisées à 65dB(A) soient à un niveau d’environ 50dB(A) dans le casque (un niveau supérieur ayant été jugé trop élevé pour les participants).

4.2.4 Protocole

Avant de prendre place dans la cabine, chaque participant a pris connaissance du déroulement de l’expérience (fiche “Consigne” en Annexe 5), et a signé une “Feuille de Consentement” (également dis-

6. Celles-ci sont disponibles dans le dossier *sounds/level_choice/*, avec une nomenclature de la forme “*approach=3+delta_laeq=0+source=ventil1_fountain_stimulus.wav*”.

ponible en Annexe 5). Une fois installés dans la cabine, il a bien été spécifié aux participants de ne pas toucher au volume de l'ordinateur. L'expérience leur a ensuite été présentée grâce à une interface web codée en *JavaScript* à l'aide de la bibliothèque *jsPsych* [7].

Dans un premier temps, la partie principale de l'expérience faisant appel à la méthode *Best-Worst Scaling (BWS)* a été présentée. Cette première partie d'expérience consistait à demander aux participants d'évaluer des groupes de 4 "séquences sonores" (appelés 4-tuples dans la suite) en sélectionnant celle qu'ils jugeaient comme la plus agréable et celle qu'ils jugeaient comme la moins agréable. Ce type de protocole, ainsi que la manière d'en analyser les résultats, a fortement été inspiré de l'étude menée par Matthieu Duroyon dans le cadre de sa thèse sur le confort sonore au sein des véhicules électriques [8]. La figure 13 présente l'interface proposée aux participants pour effectuer leur jugement sur un 4-tuple. Plus précisément, la première partie s'est déroulée en trois temps :

- **Phase d'entraînement** : deux premiers 4-tuples ont d'abord été présentés pour permettre aux participants de se familiariser avec l'interface et les stimuli (les données n'ont pas été enregistrées pour ces deux 4-tuples, qui étaient les mêmes pour tous les participants).

- **Phase principale** : 35 4-tuples ont suivi cette phase d'entraînement, constituant le coeur de l'expérience. Ceux-ci ont été constitués avec les 140 stimuli, de sorte que chaque participant entende une seule fois tous les stimuli, et que les 4-tuples qui lui sont présentés soient uniques. Ces deux contraintes ont été assurées grâce à un script *Python* fourni par Matthieu Duroyon, qui a permis de construire les groupes de stimuli de chaque participant avant la passation.

- **Phase de *re-test*** : enfin, quatre 4-tuples supplémentaires ont été présentés pour vérifier la cohérence intra-participants. Ceux-ci ont pour chaque participant été sélectionnés de manière aléatoire parmi les 4-tuples présentés durant la phase principale, et n'ont pas été séparés des autres afin de ne pas prévenir les participants de cette nouvelle phase.

Imaginez-vous dans un espace que vous côtoyez au quotidien (par exemple votre espace de travail),
et considérez que chaque séquence sonore que vous allez entendre constitue **l'ambiance sonore générale** de cet espace.

Appuyez sur la touche , , ou du pavé numérique pour écouter les différentes séquences.

Quelle est la séquence sonore la **plus** agréable ?

Quelle est la séquence sonore la **moins** agréable ?

-

+

☐ Séquence 1
☒ Séquence 2
☐ Séquence 3
☐ Séquence 4

Continuer

FIG. 13 – Interface présentée aux participants lors de la première partie de l'expérience, visant à juger des groupes de quatre séquences en termes d'agrément

Comme présenté sur la figure 13 ci-dessus, une mise en contexte a aussi été implémentée. Ce choix a résulté du test pilote réalisé pour vérifier le bon fonctionnement de l'expérience. À ce moment là, l'absence de contexte et l'utilisation du mot "Son" à la place de "Séquence" a paru manquer à l'expérience. En effet, le fait de juger simplement des sons de 10s ou d'imaginer que ces sons constituent l'ambiance sonore d'un lieu ne conditionne pas du tout le même état d'esprit. Il a donc été jugé nécessaire de formuler une légère mise en contexte qui oriente les participants vers leurs propres espaces de travail.

À la suite de la première partie de l'expérience, une seconde partie a été développée autour d'une tâche de verbalisation. Au vu des différents types de positifs (les cinq sons d'eau) utilisés dans la création des stimuli, il a en effet paru intéressant de relever la proportion de participants qui ont reconnu les sources sonores produisant ces sons d'eau. Au cours de la seconde partie, les cinq positifs ont alors été présentés aux participants, et il leur a été demandé de décrire la source du son proposé, comme exposé sur la figure 14 ci-dessous. Les cinq sons ont été présentés l'un après l'autre, et dans un ordre aléatoire.

Son A

Appuyez sur la touche 0 du pavé numérique pour écouter le son.

Quelle est la source qui produit le son selon vous ?

Continuer

FIG. 14 – *Interface présentée aux participants lors de la seconde partie de l'expérience, visant à étudier la reconnaissance de source des différents positifs*

La seconde partie concluant l'expérience, la récolte des données se déclenche grâce au bouton "Terminer l'expérience" présenté sur l'interface. Toutes les réponses des participants pour les deux parties ont ainsi été sauvegardées au format *csv* dans des fichiers nommés "*results(numéro du participant).csv*". La durée de l'expérience, estimée à 30min, a également été enregistrée pour chaque participant.

Pour terminer, il a été demandé aux participants de compléter le questionnaire présenté en Annexe 5, afin de recueillir quelques informations supplémentaires sur leur point de vue et leur ressenti par

rapport au problème de bruit, et notamment sur leur lieu de travail. Quelques remarques intéressantes de certains participants seront présentées dans la section 4.3 suivante.

4.3 Résultats

Les deux parties de l'expérience ont permis de fournir plusieurs types de résultats, qui seront présentés dans cette partie, et grâce auxquels les hypothèses de départ ont été confirmées ou réfutées.

4.3.1 Durée de l'expérience

Tout d'abord, la durée de l'expérience a été récoltée pour chaque participant, afin de s'assurer que l'estimation de 30min était bonne, et pour tenter de déceler des personnes très rapides ou très longues par rapport à l'ensemble des participants, ces personnes pouvant être reconnues comme *outliers*. Cette durée correspond plus précisément au temps écoulé entre le moment où le participant appuie sur le bouton "Commencer l'expérience" et le moment où il appuie sur le bouton "Terminer l'expérience", tous deux présentés sur l'interface graphique permettant le déroulement de l'expérience. La durée moyenne calculée grâce au passage des 30 participants s'élève à 36min (durée arrondie à l'unité). La durée la plus courte enregistrée est de 16min et la plus longue de 66min. La médiane étant de 35min, l'estimation d'une durée d'expérience de 30min semble néanmoins correcte, et aucun *outlier* n'a été détecté.

4.3.2 Choix de l'algorithme de notation

La première partie de l'expérience a fait appel à la méthode *BWS*, et les jugements pour chaque 4-tuples ont été enregistrés pour chaque participant. Après un formatage des données, un script *Python* a permis de calculer les scores de chaque stimuli, et de les classer en termes d'agrément. Les scores ont été calculés grâce à l'algorithme *Value Learning* présenté en section 2.5 de l'état de l'art. Cet algorithme a été retenu sur la base des résultats des simulations menées par Hollis pour tester la performance de plusieurs algorithmes de notation [13]. En effet, même si les algorithmes Elo et Rescorla-Wagner sont performants, c'est *Value Learning* qui s'est trouvé le moins affecté par le bruit du à la variabilité inter et intra-participants. Puisque l'expérience menée dans le cadre de cette étude implique des jugements humains à l'origine de bruit dans les données, cet algorithme a semblé le plus adapté pour calculer les scores des stimuli sonores utilisés.

4.3.3 Cohérence inter et intra-participants

Grâce à des fonctions de conformité, des valeurs de cohérence ont été calculées, afin de savoir à quel point chaque participant est cohérent avec les autres et avec lui-même. Le principe de ces fonctions, implémentées en *Python*, est de vérifier pour chaque 4-tuple si les relations entre les sons suggérées par le choix du meilleur et du pire se retrouvent bien en termes de score. En effet, lorsque qu'un participant choisit un meilleur et un pire parmi les quatre stimuli qui lui sont présentés, cinq relations peuvent être déduites, comme expliqué dans la section 2.5. Pour chaque participant et chaque 4-tuple qui lui a été proposé, il est possible de récupérer ces cinq relations, et de vérifier si elle est respectée en termes de scores. En comparant les relations choisies par un participant aux scores déduits de l'ensemble de l'expérience (avec tous les participants), il est ainsi possible de calculer une cohérence inter-participants. Si ces relations sont comparées aux scores déduits de ce même participant, il est possible d'obtenir une cohérence intra-participants. Cela étant dit, la figure 15 ci-dessous montre les résultats obtenus.

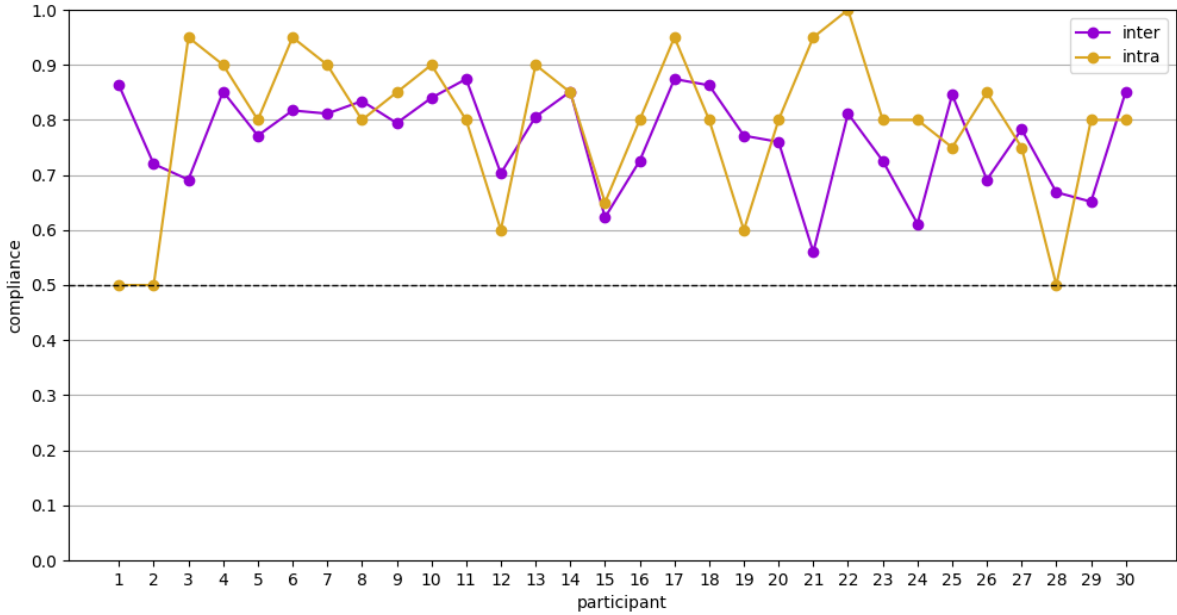


FIG. 15 – Cohérence inter-participants (en bleu) et intra-participants (en orange) pour chaque participant

La ligne en pointillés marquant la valeur 0.5 décrit la valeur minimale que peuvent prendre les valeurs de cohérence. Ce minimum pourrait être assimilé au “hasard”, puisqu’il suggère que la moitié des jugements d’un participant ne correspondent pas aux scores obtenus pour l’ensemble des participants, ou pour ce même participant. En observant la figure 15, la cohérence inter-participants est supérieure à 0.7 pour la plupart des participants, mis à part les participants 15, 21, 24, 28 et 29 (seul le participant 21 a obtenu une valeur inférieure à 0.6). Par rapport à la valeur minimale 0.5, les participants sont donc assez cohérents entre eux. Concernant la cohérence intra-participants, elle est également assez élevée pour la plupart des participants, dépassant souvent 0.8. Cela signifie que ces participants sont très cohérents avec eux-mêmes, à l’inverse des participants 1, 2 et 28, qui obtiennent des valeurs de 0.5. Ces résultats assez bas étaient attendus car certains participants ont signifié à la fin de l’expérience qu’il avait été difficile pour eux de savoir ce qui leur plaisait réellement, et que les réponses données au début de l’expérience ne seraient sûrement pas cohérentes avec les réponses données à la fin de celle-ci.

4.3.4 Effet de l’approche sur le score

Une fois les scores calculés pour tous les stimuli, il est possible de les visualiser en fonction des paramètres de l’expérience. Pour tenter de répondre à la problématique principale, les données ont tout d’abord été agrégées selon les deux approches. La figure 16 ci-dessous montre néanmoins que la différence entre les scores obtenus avec les approches *masker* et *concealer* n’est pas significative. Il n’est donc pas possible à ce stade de prétendre que l’approche *concealer* donne de meilleurs résultats que l’approche *masker*, la figure 16 suggérant l’inverse.

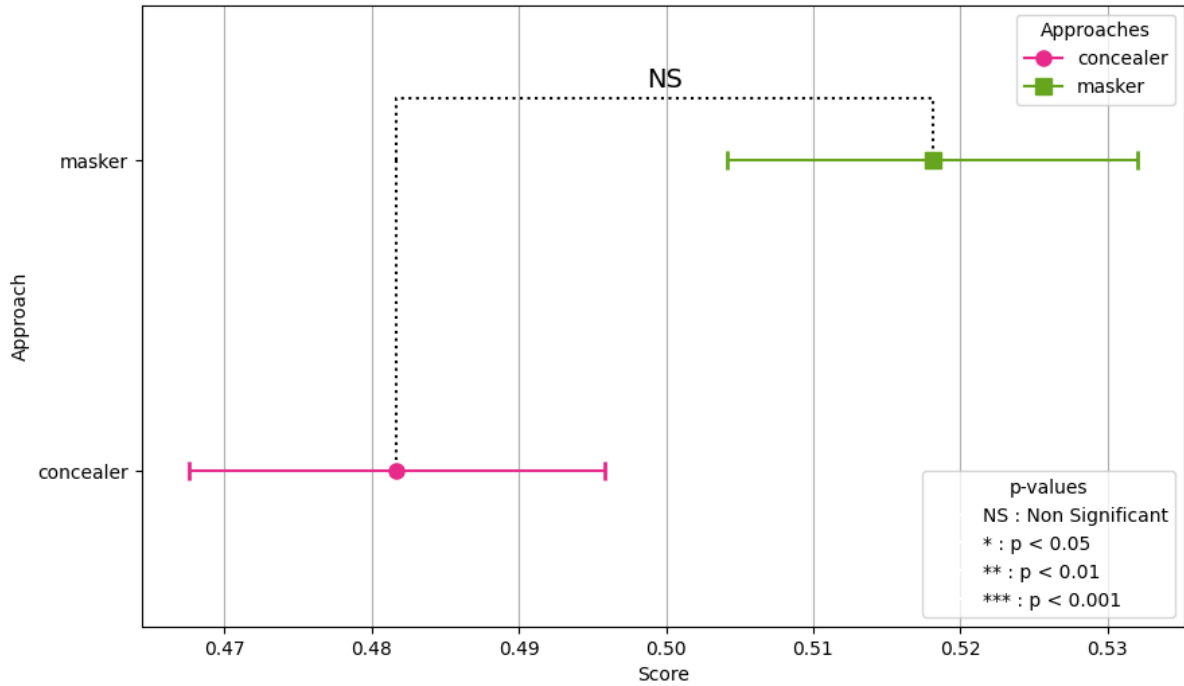


FIG. 16 – Scores moyens des mixtures en fonction de l’approche utilisée (concealer en rond rose et masker en carré vert)

L’approche *masker* est globalement meilleure que l’approche *concealer* en termes de score, et l’utilisation d’un *t-test* fournissant une *p-value* $p = 0.07$ montre que la différence des scores moyens observés pour les deux approches n’est pas significative. Les participants ont donc globalement jugé les mixtures générées grâce à l’approche *masker* comme plus agréables que celles générées grâce à l’approche *concealer*.

4.3.5 Effet de la source sur le score

Afin de comprendre les effets des paramètres secondaires sur les scores et d’espérer trouver certains paramètres pour lesquels l’approche *concealer* fournit de meilleurs résultats que l’approche *masker*, il est possible de séparer les données obtenues selon les différents paramètres.

En premier lieu, il est pertinent d’étudier l’effet du type de source sur le score. La figure 17 suivante présente les scores moyens recueillis pour les deux sources selon les deux approches.

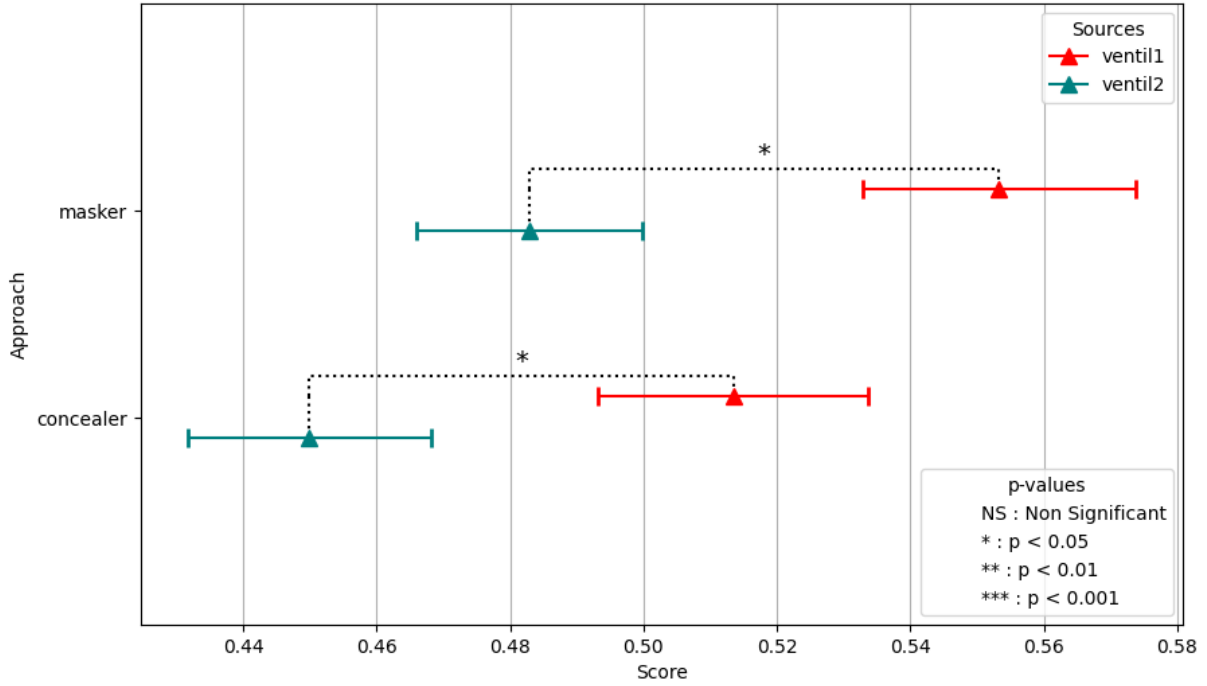
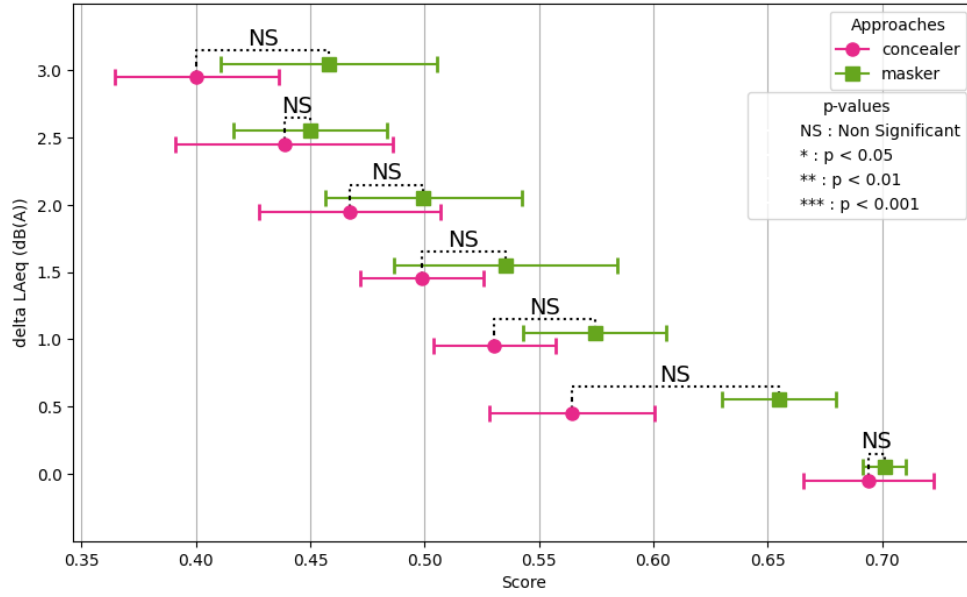


FIG. 17 – Scores moyens des mixtures en fonction de l’approche utilisée et des deux sources (ventil1 en rouge et ventil2 en bleu)

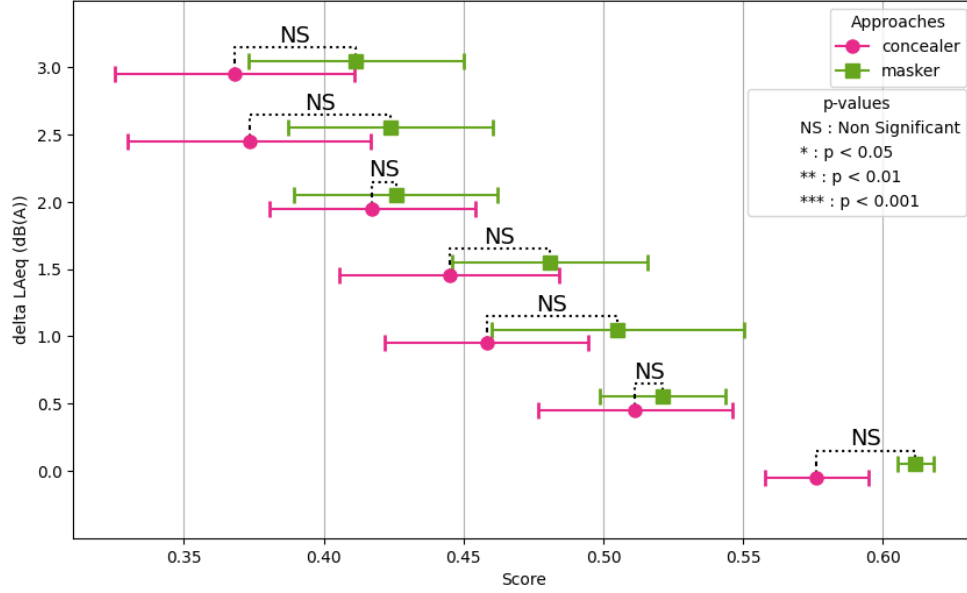
Cette figure et les *t-tests* réalisés montrent que la différence entre les deux sources est significative, et ce pour les deux approches. La source *ventil1*, qui représente le bruit de ventilation enregistré à l’IRCAM, est mieux perçu que la source *ventil2*. Ce résultat était attendu puisque la source *ventil2* a été choisi pour son caractère plus gênant dû à sa composante tonale. Aussi, puisque les deux sources ont des effets différents sur le score, et donc sur l’agrément, les résultats les concernant ne seront pas agrégés dans la suite. Une analyse en parallèle sera donc effectuée pour séparer les scores des mixtures contenant *ventil1*, et ceux des mixtures contenant *ventil2*.

4.3.6 Effet du niveau sur le score

L’effet du niveau sur le score est considérable, et la figure 18 montre une tendance linéaire entre le score et le paramètre ΔL_{Aeq} , et ce indépendamment de la source utilisée pour la génération des mixtures.



((a)) Scores pour "ventil1"



((b)) Scores pour "ventil2"

FIG. 18 – Scores moyens des mixtures en fonction de ΔL_{Aeq} pour les deux approches masker (carré vert) et concealer (rond rose)

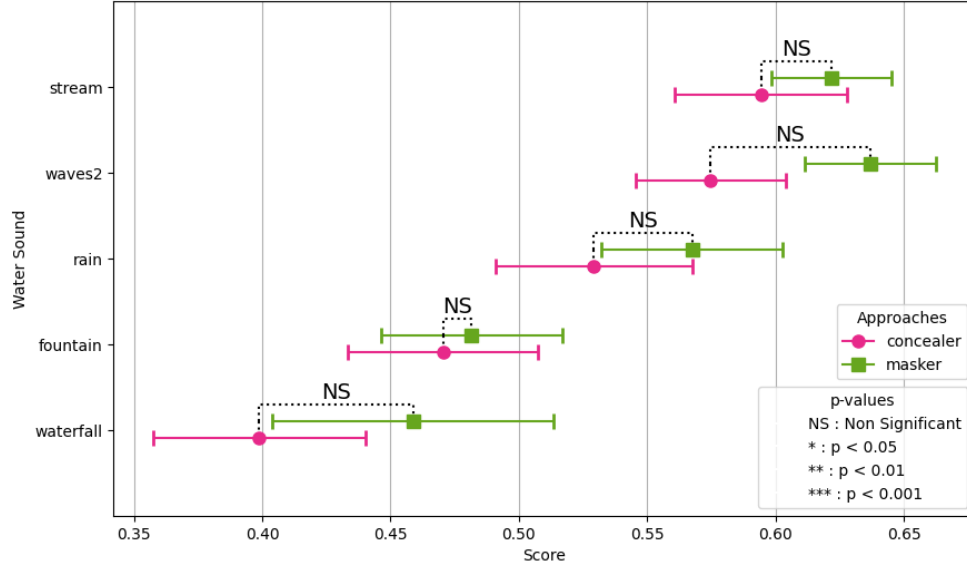
Les figures 18(a) et 18(b) montrent une tendance décroissante du score lorsque le paramètre ΔL_{Aeq} augmente, et suggèrent une relation presque linéaire. Pour les deux sources de bruit, ce sont les mixtures générées pour $\Delta L_{Aeq} = 0\text{dB(A)}$ qui sont les mieux perçues.

De plus, l'approche *masker* reste meilleure que l'approche *concealer* pour toutes les valeurs de ΔL_{Aeq} en termes de score, et donc en termes d'agrément. Il arrive même que l'approche *masker* fournisse de meilleurs scores que l'approche *concealer* pour des valeurs plus élevées de ΔL_{Aeq} . Par exemple, les mixtures créées grâce à l'approche *masker* obtiennent de meilleurs scores pour $\Delta L_{Aeq} = 1.5\text{dB(A)}$ que celles créées à l'aide de l'approche *concealer* pour $\Delta L_{Aeq} = 1\text{dB(A)}$. Cependant, la différence entre les

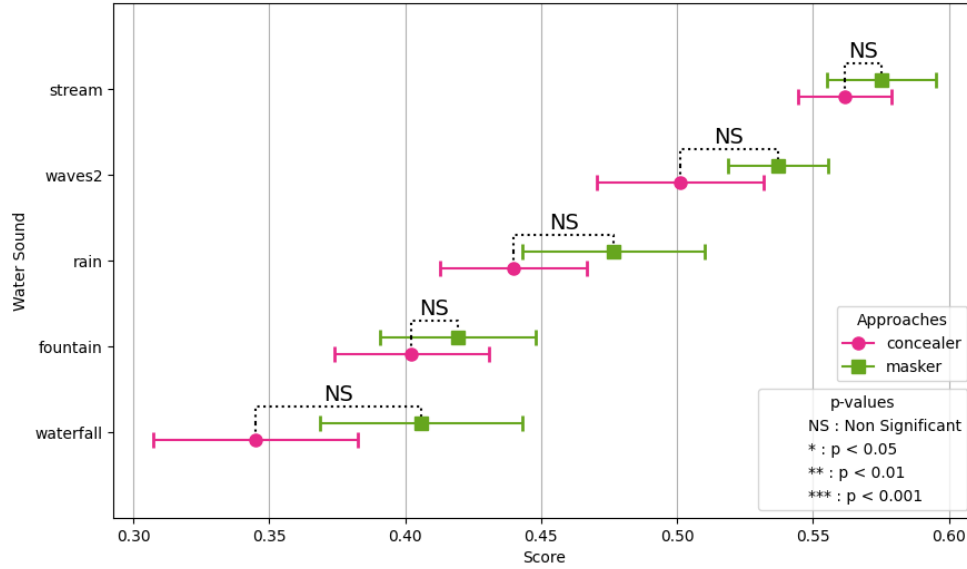
deux approches n'est toujours pas significative, et ceux pour les deux sources de bruit.

4.3.7 Effet du positif sur le score

Le dernier paramètre, le type de positif, semble lui aussi avoir un effet significatif sur le score. La figure 19 fait émerger un ordre parmi les positifs, plaçant les mixtures générées avec le son *stream* comme les mieux perçues, tandis que celles générées avec le son *waterfall* sont les moins bien perçues.



((a)) Scores pour "ventil1"



((b)) Scores pour "ventil2"

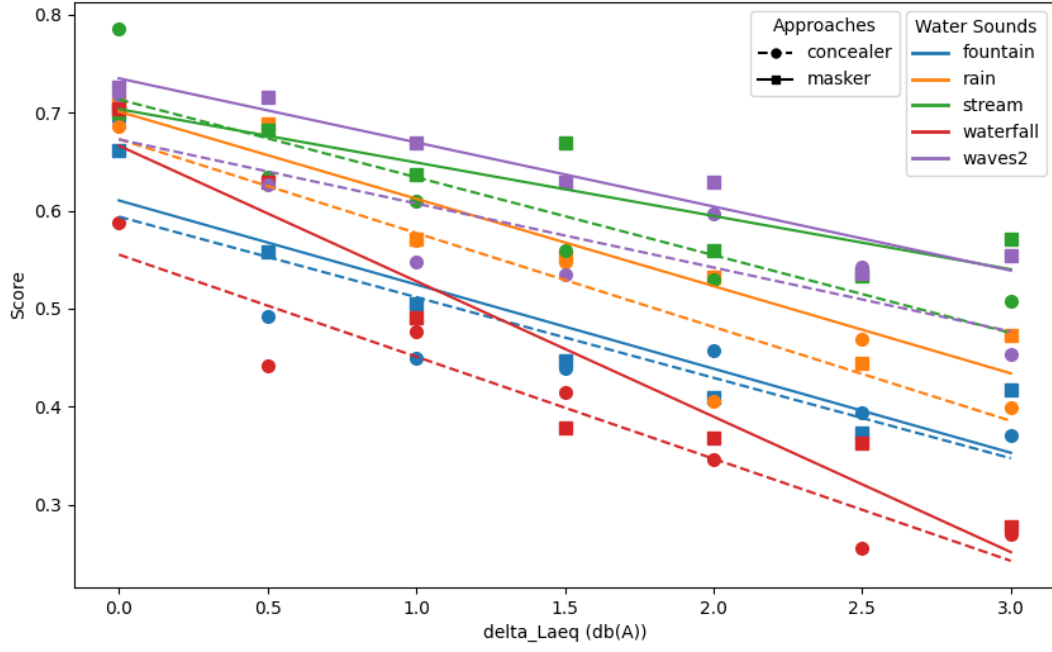
FIG. 19 – Scores moyens des mixtures en fonction du type de son d'eau pour les deux approches masker (carré vert) et concealer (rond rose)

Tout comme pour le paramètre de niveau des mixtures ΔL_{Aeq} , les deux figures 19(a) et 19(b) confirment que pour chaque positif, l’approche *masker* fournit de meilleurs résultats en termes de score que l’approche *concealer*. Aucune différence significative n’est relevée entre les scores moyens obtenus pour chaque son d’eau avec les deux approches.

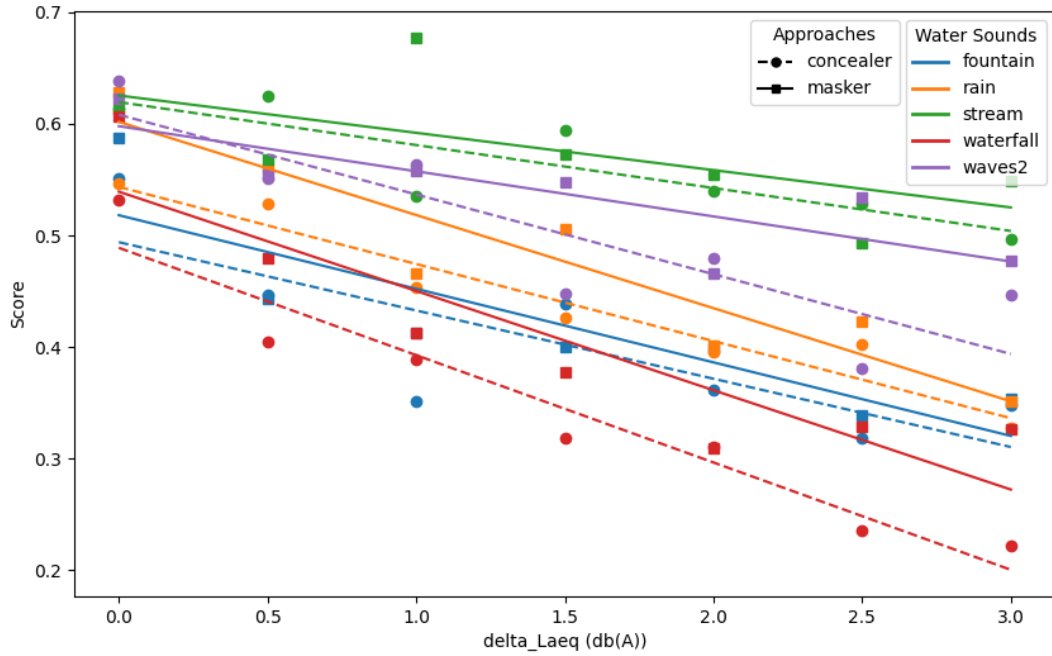
4.3.8 Visualisation des paramètres de l’expérience

Les figures précédentes, présentées sous forme d’*error bars*, ont principalement appuyé le fait que la différence entre les deux approches n’est pas significative, et ce pour tous les paramètres de l’expérience. Seul l’effet de la source sur le score a pu être démontré par l’utilisation d’un *t-test* puisque seulement deux sources ont été utilisées pour la génération des mixtures. Pour le type de positif et le niveau des mixtures, qui présentent un plus grand nombre de conditions, il n’a pas été possible d’employer le même type de test statistique, et le format des *error bars* semble moins pertinent. C’est en ce sens que la figure 20 ci-après présente tous les paramètres utilisés, et permet de se rendre compte de l’effet du paramètre ΔL_{Aeq} et du type de positif sur le score, et donc sur l’agrément.

Sur cette figure 20, les valeurs des scores des différentes mixtures sont représentées par les points, et des régressions linéaires (représentées par les droites) ont été calculées pour chaque approche et chaque type de positif. Cette visualisation permet dans un premier temps d’observer une tendance générale sur le niveau des mixtures, et de supposer une relation linéaire entre le score et le paramètre ΔL_{Aeq} . En effet, comme spécifié plus haut, le score décroît lorsque la valeur de ΔL_{Aeq} augmente. Il est alors intéressant de remarquer que c’est une augmentation du niveau du positif qui implique une augmentation de ΔL_{Aeq} (puisque le niveau de la source est constant). Ces résultats traduisent donc le fait que l’ajout de positif à la source n’est pas apprécié, que ce soit avec l’utilisation de l’approche *masker* ou de l’approche *concealer*. Les participants semblent préférer les mixtures telles que $\Delta L_{Aeq} = 0\text{dB(A)}$, qui sont perceptivement proches des deux sources. De la même manière, les décroissances de score lorsque ΔL_{Aeq} augmente prouvent que plus la part de son d’eau est forte dans les mixtures, moins ces dernières sont jugées agréables, et ce pour tous les types de sons d’eau.



((a)) Scores pour "ventil1"



((b)) Scores pour "ventil2"

FIG. 20 – Scores des mixtures en fonction de ΔL_{Aeq} , de l'approche (ronds et carrés) et du positif utilisé (bleu, orange, vert, rouge et violet) pour les deux bruits de ventilation

D'autre part, cette figure suggère un classement des mixtures selon le positif utilisé. En notant $S_{positif}$ le score moyen des mixtures créées à l'aide du son d'eau indiqué, la relation suivante peut être supposée :

$$S_{stream} > S_{waves} > S_{rain} > S_{fountain} > S_{waterfall}$$

Pour tenter de comprendre cette relation, plusieurs informations ont été récoltées sur les différents sons d'eau. Plusieurs indices spectraux ont d'abord été étudiés. Plus précisément, un indice d'entropie spectrale a été calculé grâce à la bibliothèque *scikit-maad* [29] sous *Python*. La bibliothèque *librosa* [18] a également permis de calculer le centroïde spectral, la bande passante, le contraste spectral et la planéité spectrale de chaque son d'eau. La figure 21 ci-dessous présente les valeurs de ces indices ainsi que les scores obtenus pour chaque son d'eau :

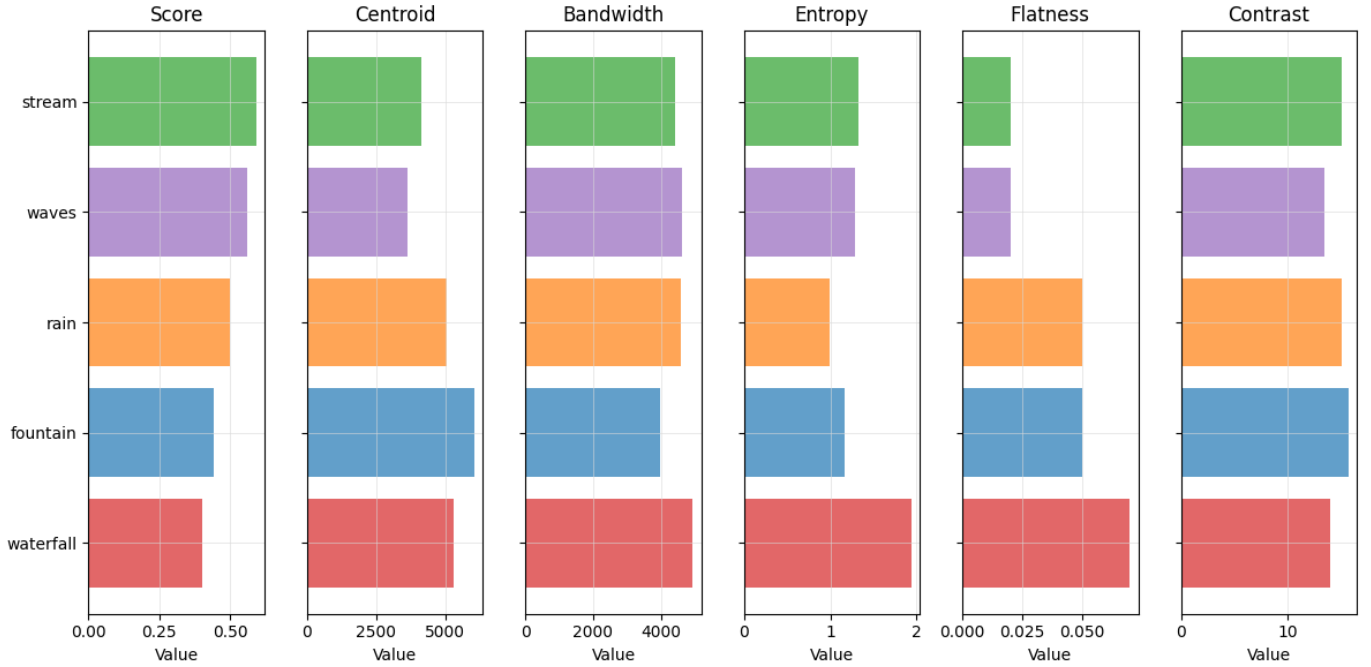


FIG. 21 – Scores moyens des mixtures et indices spectraux pour les cinq sons d'eau

Cette figure 21 permet d'observer les variations des indices calculés en fonction du type de son d'eau. En particulier, les variations de l'indice de planéité spectrale (*Flatness* sur la figure) semble suivre une tendance inverse à celle du score. Sachant que les scores des mixtures générées à l'aide des différents sons d'eau représentent l'agrément perçu, et que la planéité spectrale quantifie la ressemblance d'un son avec le bruit par opposition à sa tonalité, une relation entre agrément et diversité spectrale peut être supposée. Les mixtures les mieux perçues sont celles constituées des sons *stream* et *waves*, qui ont les valeurs les plus faibles de planéité spectrale, et donc les moins proches spectralement du bruit blanc (c'est-à-dire les plus tonaux). À l'inverse, le son *waterfall* présente une valeur élevée de planéité spectrale, et se rapproche donc du bruit blanc. Ce sont les mixtures contenant ce son qui sont les moins bien perçues.

En plus des valeurs physiques que fournissent les indices spectraux, la deuxième partie de l'expérience a permis de recueillir des informations sur la capacité des participants à reconnaître la source des sons d'eau. Les réponses des participants ont été étudiées et un pourcentage de reconnaissance a été initialisé pour chaque son d'eau. Si le participant définit clairement la source du son en mentionnant les noms des sons d'eau (fontaine, pluie, cours d'eau, cascade et vagues), le score de la source reconnue est augmenté

de 1. Pour les sons *stream* et *waves*, d’autres mots ont été acceptés, tant que le participant utilise bien ces mots pour décrire la source du son, et pas seulement un décor. Par exemple, les mots “rivière”, “ruisseau”, “torrent” et “fleuve” sont acceptés pour parler du son *stream*, et le mot “mer” est accepté pour parler du son *waves*. En revanche, “le vent dans les arbres à l’arrière d’une plage” n’est pas accepté, car la réponse suggère que la source du son est le vent, et non les vagues. Enfin, plusieurs participants ont ajouté du contexte dans leurs réponses, en donnant des descriptions assez précises de l’environnement qu’ils ont imaginé autour de la source du son. Cela ne changeant pas la réponse principale sur la source des sons, tous les contextes ont été acceptés. Ainsi, “pluie sur un toit” est autant accepté que “la pluie qui tombe et qui tape sur les carreaux”, ou simplement “pluie”, et “les vagues de la mer par temps calme” est autant accepté que “l’Atlantique par gros temps”. Finalement, cette partie a révélé que le son *rain* est le plus identifiable, suivi des sons *stream* et *waves*, suivis du son *waterfall*, et enfin du son *fountain*, que seulement quatre participants ont reconnu. Les résultats ont été relevés grâce à des pourcentages de reconnaissance des sources, présentés dans le tableau 1 suivant :

	Type de son d’eau				
	<i>fountain</i>	<i>rain</i>	<i>stream</i>	<i>waterfall</i>	<i>waves</i>
Reconnaissance de la source (en %)	13.3	83.3	60	26.7	60

TAB. 1 – *Reconnaissance de la source des sons d’eau utilisés pour la création des mixtures*

La reconnaissance des sons d’eau ne semble pas être le facteur principal responsable de l’agrément perçu dans le cadre des mixtures créées avec les deux approches, car les pourcentages de reconnaissance ne croient pas avec le score moyen des sons d’eau. En revanche, il est important de remarquer que les trois sons d’eau les mieux perçus sont aussi les plus reconnus. En ce sens, il peut être intéressant d’étudier les données en constituant deux groupes de sons d’eau. Le premier, le groupe 1, est constitué des sons *rain*, *stream* et *waves*, tandis que le groupe 2 est constitué des sons *fountain* et *waterfall*. La figure 22 ci-dessous montre les scores obtenus pour les deux groupes (seules les mixtures générées avec la source “ventill” ont été retenues) :

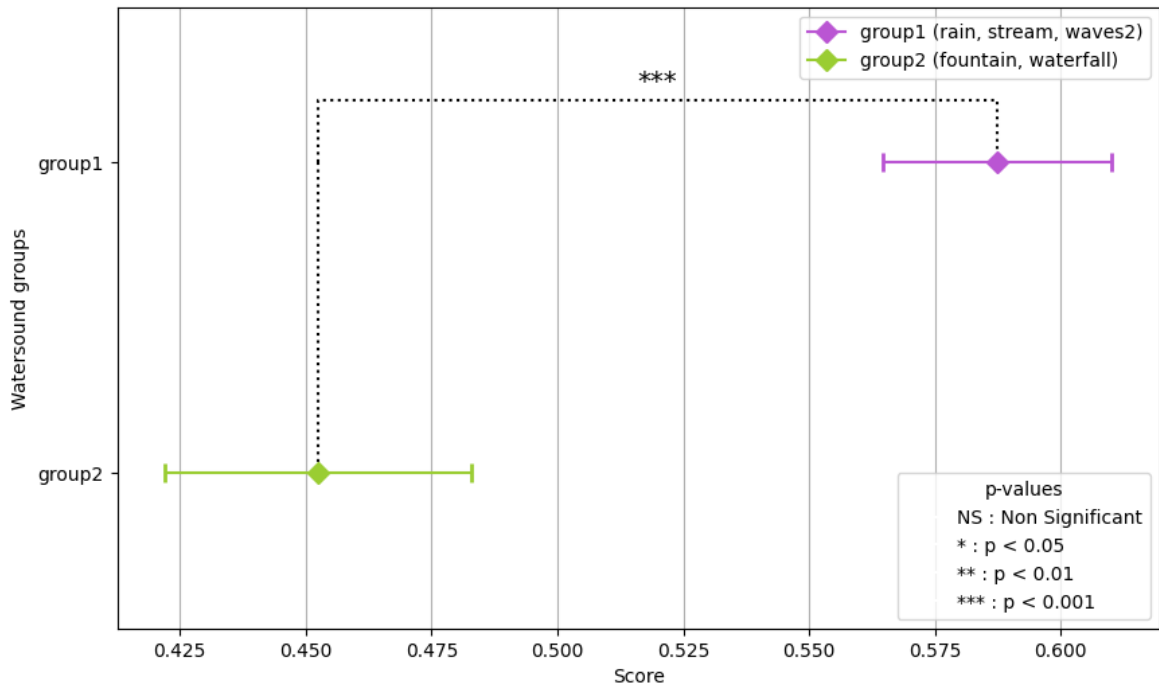


FIG. 22 – *Scores moyens des mixtures en fonction des groupes de positifs pour la source “ventil1”*

Une différence significative a été relevée entre les scores moyens des deux groupes, ce qui confirme l’effet significatif des positifs sur le score. Ces résultats peuvent laisser penser que la reconnaissance des sources pour les sons d’eau a également un effet sur le score.

4.3.9 Enquête sur les sons désirables et indésirables dans les bureaux ouverts

Le questionnaire proposé à la fin de l’expérience a finalement révélé des informations sur les sons rencontrés par les participants sur leur lieu de travail. En particulier, il a été demandé aux participants d’indiquer les sons perçus comme gênants, ainsi que les sons désirés, c’est-à-dire ceux qu’ils aimeraient entendre dans un contexte de travail quotidien. La figure 23 présente les mots cités pour décrire les sons perçus comme gênants par les participants. Ces mots sont séparés en deux catégories : ceux qui décrivent une source sonore, et ceux qui décrivent plutôt un ensemble de sons avec une caractéristique spécifique.

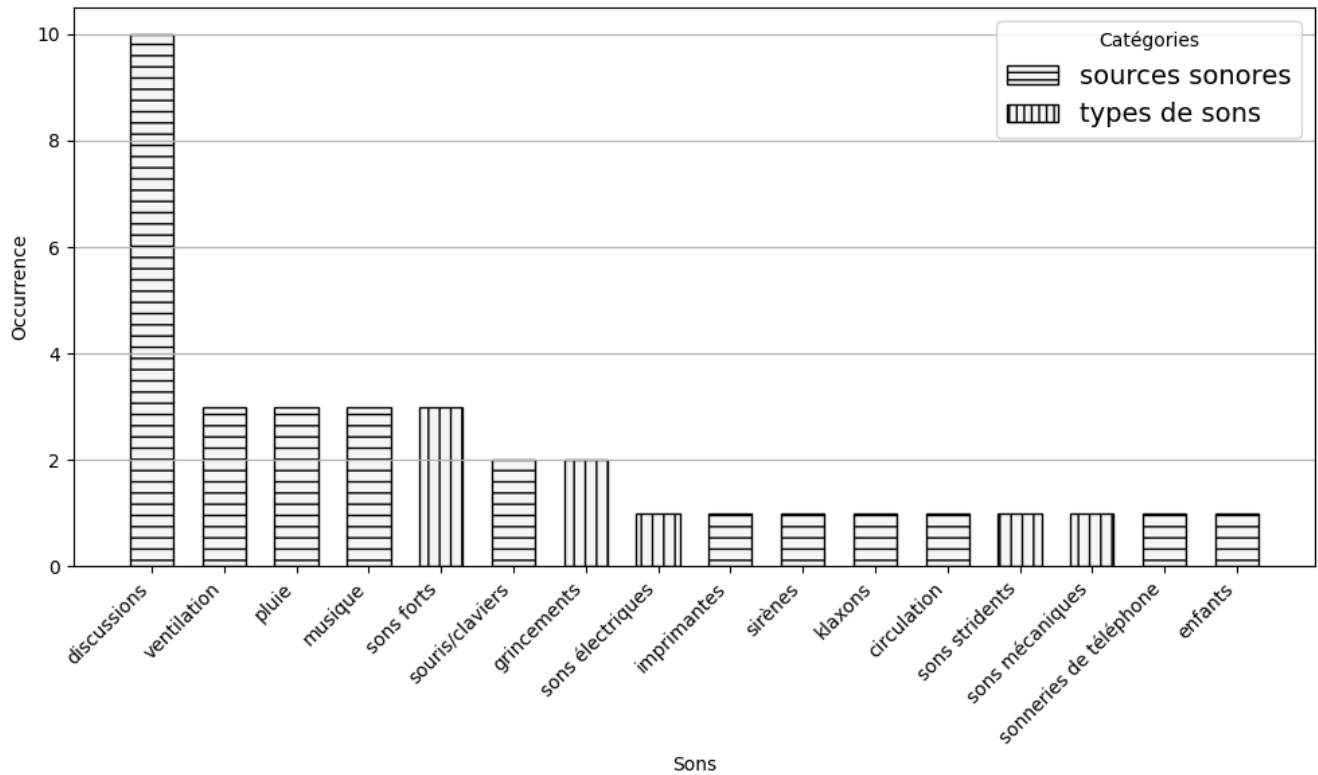


FIG. 23 – Occurrence des mots cités pour décrire les sons perçus comme gênants dans les bureaux ouverts

Parmi les sons gênants, des sources sonores spécifiques mais aussi des types de sons sont cités. Par exemple, le bruit des discussions est le plus souvent évoqué. Certaines caractéristiques sont aussi souvent considérées comme gênantes. C'est notamment le cas des sons stridents, des sons forts ou des sons mécaniques. Par ailleurs, il est possible de distinguer les bruits intérieurs, comme les bruits d'imprimantes et la ventilation, des bruits extérieurs comme les sirènes, les klaxons et le bruit des enfants.

Concernant les sons désirés, c'est-à-dire les sons que les participants aimeraient entendre dans leur environnement de travail, plus d'une vingtaine de sources sonores ont été proposées, dont certaines peuvent se regrouper dans des catégories, comme le montre la figure 24 suivante.

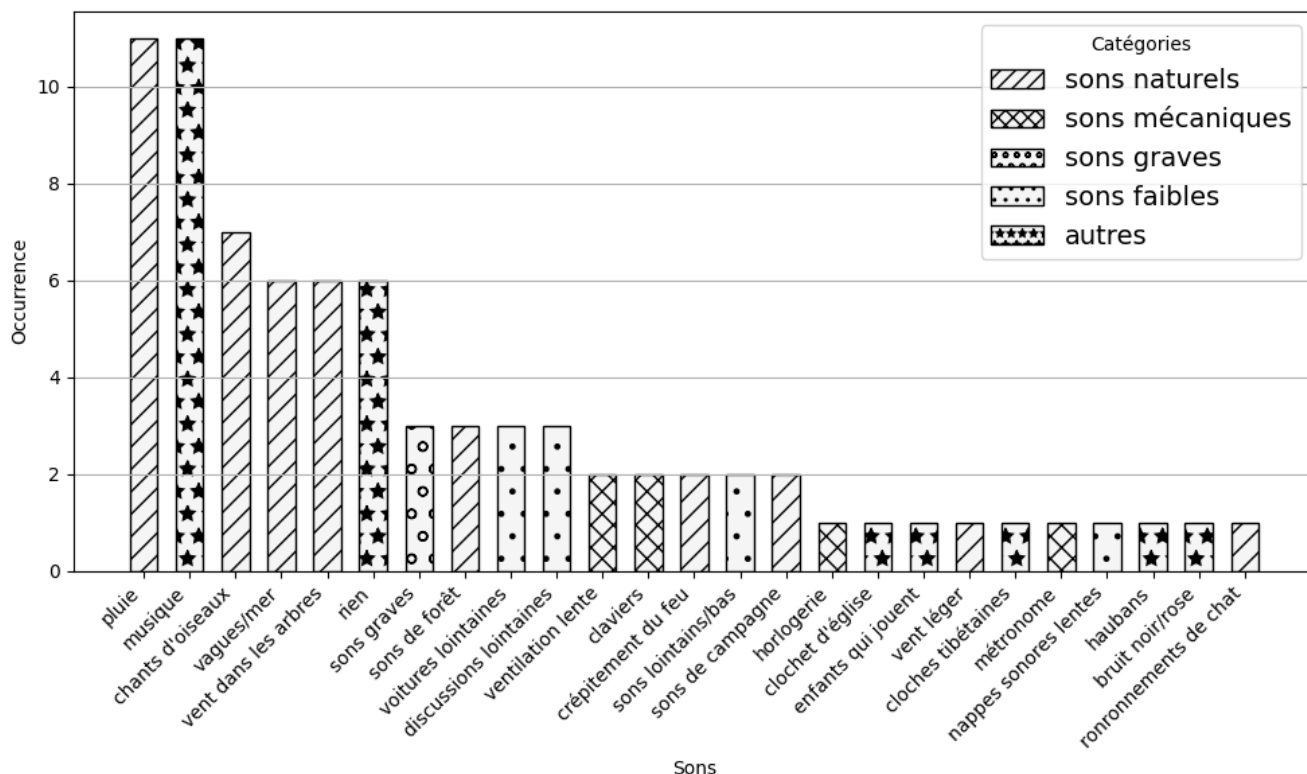


FIG. 24 – Occurrence des mots cités pour décrire les sons désirés dans les bureaux ouverts

Les deux sons les plus désirés sont le son de la pluie et la musique. De plus, beaucoup de sons naturels ont été cités (les vagues, le vent dans les arbres, le chant des oiseaux ou encore le crépitement du feu). Certains sons mécaniques semblent aussi susciter l'agrément de certains participants, comme les bruits d'horlogerie, de claviers ou de métronome. Certains sons se retrouvent également dans leur aspect grave, comme pour la ventilation, ou faible en termes de niveau, comme pour les discussions ou les voitures "lointaines". Le nombre de mots cités relate bien du caractère subjectif de l'agrément, et il est important de noter que certains participants ont décrit les sons qu'ils aimeraient entendre avec beaucoup de précision. Par exemple, le mot 'musique', qui est apparu chez 11 participants, était toujours accompagné d'adjectifs, comme "personnelle", "calme", "douce", "pas trop fort", "sons d'ambiance", "sans paroles" ou encore "prévisible".

Finalement, il est intéressant de remarquer que certains sons se trouvent à la fois dans les sons gênants et dans les sons désirés. Les bruits de claviers, des voix et des discussions environnantes, ainsi que des enfants qui jouent sont perçus très différemment. Par ailleurs, deux des sons d'eau utilisés pour l'expérience ont été cités dans les sons désirés, à savoir le son de pluie et celui de vagues. Cela confirme en partie leur statut de "positifs", car certains participants précisent que les sons d'eau ne pas sont agréables pour eux de manière générale.

4.4 Discussion

Le but de cette expérience est d'évaluer l'impact des approches *masker* et *concealer* sur l'agrément perçu, lorsque celles-ci sont utilisées sur des bruits de ventilation avec l'aide de différents sons d'eau supposés acceptés. Les résultats obtenus ne sont pas ceux espérés, car ils montrent que l'approche *concealer*

ne permet pas de générer des mixtures mieux perçues que celles obtenus avec l’approche *masker* en termes d’agrément.

D’autre part, l’étude du niveau grâce au paramètre ΔL_{Aeq} semble suggérer que l’ajout des sons d’eau ne permet pas d’améliorer l’agrément perçu des deux sons de ventilation, et ce pour les deux approches utilisées. Ces résultats diffèrent de ceux obtenus par Cai et coll. pour l’approche *masker*, qui ont trouvé une relation quadratique entre niveau (représenté par le paramètre *MSNR*) et réduction de la gêne perçue [6]. Cette différence peut être due à plusieurs éléments, notamment le type et le niveau de la source choisie (un bruit de soudeuse électrique à $80dB(A)$ pour l’étude de Cai et coll., contre un bruit de ventilation à $50dB(A)$ ici), mais aussi le paramètre étudié (la gêne et l’agrément perçus sont deux paramètres bien distincts). Dans le cadre de cette étude, l’ajout de complexité à la source de bruit n’est pas désiré, du moins en termes d’agrément.

En revanche, au vu de l’effet du type de source et du type de positif, il serait intéressant de construire le signal *concealer* avec d’autres sources et d’autres positifs, car les scores en termes d’agrément semblent très dépendants de ces paramètres pour les deux approches. Les résultats obtenus lors de cette étude peuvent par ailleurs suggérer que le cadre choisi n’est pas assez représentatif pour voir les effets des traitements proposés. Il serait par exemple possible de traiter des sources beaucoup plus gênantes et désagréables, et d’étudier les sons désirables dans les bureaux ouverts pour mieux choisir les positifs. C’est pourquoi il est important de mieux comprendre le ressenti des participants quant aux stimuli présentés. D’une part, les discussions post-expérience ont révélé que les sons ont plus souvent été évalués en termes de désagrément qu’en termes d’agrément. Beaucoup de participants ont en effet précisé qu’aucun des sons ne leur paraissait agréable, et que le choix du “meilleur” stimulus en termes d’agrément correspondait en fait au moins pire. Le statut de “positifs” de sons d’eau a été remis en cause dans le contexte étudié, car ceux-ci ont souvent été perçus comme trop “forts”, trop “proches”, ou encore non cohérents avec un contexte de travail, certains participants mentionnant même une sensation d’inconfort à l’écoute des stimuli. Cela ne semble cependant pas unanime, puisque quelques participants ont affirmé être sortis relaxés de l’expérience. D’autre part, les participants ont identifié certaines caractéristiques des sons comme des facteurs responsables de l’agrément ou du désagrément perçu. La sonie est restée le premier facteur responsable du désagrément, ce qui rejoint la littérature et les résultats présentés plus haut. De manière générale, les participants ont dit préférer les sons graves et homogènes. En parlant plus précisément de la création de mixtures grâce aux sons de ventilation et d’eau, celles-ci ont été perçues plus agréables lorsque le son d’eau était dans la même gamme de fréquences que le bruit de ventilation. Un “équilibre” entre les sons a aussi été mentionné pour caractériser le fait que la mixture entre deux sons est la plus agréable lorsque ceux-ci sont perçus au même niveau.

Finalement, l’aspect subjectif de la notion d’agrément peut imposer un choix personnalisé des positifs. Cette idée de personnalisation est d’ailleurs confirmée par l’un des participants qui mentionne qu’il préfère contrôler son environnement sonore plutôt que de laisser place à l’imprévu. L’idée d’imposer des sources sonores spécifiques, même si elles sont communément acceptées, mérite d’être plus amplement étudiée pour s’assurer de son efficacité pour un traitement visant à réduire perceptivement le bruit. Des études supplémentaires sont nécessaires pour espérer identifier des typologies de sons agréables et désagréables, et ainsi trouver de bons couples source-positif pour appliquer l’approche *concealer* efficacement en termes d’agrément.

5 Conclusion

Afin de proposer une solution de traitement perceptif du bruit dans les environnements intérieurs, une nouvelle approche nommée *concealer* a été développée. La particularité de cette approche est la construction d'un signal *concealer*, qui une fois ajouté à la source de bruit agit comme "maquilleur" et permet de créer une mixture qui soit mieux perçue que le bruit initial. Pour s'assurer de cela, le signal *concealer* dépend d'un son dit positif, mais aussi de la source de bruit elle-même. Cette méthode de construction du maquilleur diffère de celle utilisée pour l'approche *masker*, qui consiste simplement à définir un son positif comme maquilleur. L'idée portée par l'approche *concealer* est alors de compléter la source de bruit tout en minimisant l'augmentation du niveau sonore global.

Après avoir mis en œuvre cette approche en utilisant des bruits de ventilation comme sources et des sons d'eau comme positifs, une expérience perceptive a été menée pour tenter d'évaluer l'efficacité de cette approche face à l'approche *masker* en termes d'agrément. Plusieurs hypothèses ont été émises, et l'expérience a été conçue en espérant que l'approche *concealer* soit meilleure que l'approche *masker*, et que l'utilisation des deux approches ait un effet positif sur l'agrément. De plus, il était attendu que l'utilisation de différents types de sons d'eau ait un impact sur l'agrément perçu. Cette dernière hypothèse s'est révélée vérifiée, car les mixtures impliquant les sons *stream* et *waves* ont été mieux perçus que les sons *rain*, *fountain* et *waterfall*. En revanche, les résultats n'ont pas permis de valider l'hypothèse principale, car l'approche *masker* s'est révélée meilleure que l'approche *concealer*.

Par ailleurs, le rapport entre le niveau ajouté et l'agrément perçu a une tendance linéaire, c'est-à-dire que plus le niveau du son positif est élevé, plus le niveau des mixtures est élevé, et moins celles-ci sont agréables. Les mixtures générées par l'utilisation des deux approches sont donc moins bien perçues que la source de bruit mixée à un très faible de niveau de positif.

Enfin, une enquête sur les sons jugés désirables dans un contexte de travail a révélé une grande diversité de sources sonores. La divergence des goûts des participants ne rend pas évidente l'amélioration de l'agrément en créant des mixtures composées de sources sonores spécifiques. Par extension, il serait intéressant de mener une enquête plus détaillée sur les sources sonores ou les caractéristiques des sons désirés dans le contexte des bureaux ouverts.

Pour conclure, un examen plus approfondi de cette étude pourrait apporter des informations supplémentaires aux résultats présentés dans ce document. Par exemple, l'étude d'*outliers* chez les participants pourrait conduire à des analyses partielles complémentaires. Une fois cette étude terminée, plusieurs pistes sont envisageables, notamment la réévaluation des hypothèses de travail et la redéfinition du domaine d'application de l'approche *concealer*. Plutôt que de travailler sur un paramètre de niveau tel que le paramètre ΔL_{Aeq} , une nouvelle idée serait de définir un cadre pour le niveau sonore de la source de bruit. Par exemple, l'approche *concealer* pourrait se trouver couplée à des techniques de réduction de bruit, qui ne permettent pas de supprimer complètement le bruit ambiant. L'approche *concealer* trouverait alors son utilité pour des sources avec de très bas niveaux, en tentant de modifier la perception des résidus de bruit. Les mixtures générées pourraient davantage s'apparenter aux sons positifs choisis dans ce nouveau cadre. Un tel couplage ne serait en revanche pas pertinent dans le contexte des bureaux ouverts, car cela imposerait le port d'écouteurs ou de casques munis d'une réduction du bruit. Il est alors possible de s'intéresser à d'autres terrains d'application, comme les usines dans lesquelles les travailleurs sont soumis à de hauts niveaux de bruits industriels.

Références

- [1] Jean-François Augoyard. “L’entretien sur écoute réactivée”. In: *L’espace urbain en méthodes* (2001), pp. 127–152.
- [2] Pierre Aumond et al. “Global and continuous pleasantness estimation of the soundscape perceived during walking trips through urban environments”. In: *Applied Sciences* 7.2 (2017), p. 144.
- [3] Clara Boukhemia. “Réduire le bruit grâce à la Réalité Augmentée Sonore - *Auditory Concealer* : État de l’art”. État de l’art rédigé en 2024 (stage de master 2).
- [4] Baptiste Bouvier. “Saillance auditive: de la caractérisation psychoacoustique à la perception de l’environnement sonore”. PhD thesis. Sorbonne Université, 2024.
- [5] Laurent Brocolini, Etienne Parizet, and Patrick Chevret. “Effect of masking noise on cognitive performance and annoyance in open plan offices”. In: *Applied acoustics* 114 (2016), pp. 44–55.
- [6] Jun Cai et al. “Effect of water sound masking on perception of the industrial noise”. In: *Applied Acoustics* 150 (2019), pp. 307–312.
- [7] Joshua De Leeuw, Rebecca Gilbert, and Björn Luchterhandt. “jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments”. In: *Journal of Open Source Software*, 8(85), 5351 (2023). URL: <https://joss.theoj.org/papers/10.21105/joss.05351>.
- [8] Matthieu Duroyon. “Du confort acoustique au confort sonore – Comprendre et améliorer le confort dans l’habitable grâce à l’acoustique active”. Soutenance de thèse prévue en 2025.
- [9] Matthieu Duroyon et al. “How do Automotive Acousticians talk about Sounds and Comfort in Electric Vehicle?” In: *10th Convention of the European Acoustics Association Forum Acusticum 2023*. European Acoustics Association. 2023, pp. 6335–6342.
- [10] Arpad E Elo. “The international chess federation rating system”. In: Chess. 1973.
- [11] José Domingo Guillén and Isabel López Barrio. “Importance of personal, attitudinal and contextual variables in the assessment of pleasantness of the urban sound environment”. In: (2007).
- [12] Rainer Guski, U Felscher-Suhr, and R Schuemer. “The concept of noise annoyance: How international experts see it”. In: *Journal of sound and vibration* 223.4 (1999), pp. 513–527.
- [13] Geoff Hollis. “Scoring best-worst data in unbalanced many-item designs, with applications to crowdsourcing semantic judgments”. In: *Behavior research methods* 50.2 (2018), pp. 711–729.
- [14] Mila Kossachka et al. “Une approche psycholinguistique et cognitive du sentiment de «confort»: de l’analyse linguistique à la physique”. In: *Dénomination, désignation et catégories* (2006), p. 65.
- [15] Mathieu Lagrange. *doce: Design Of Computational Experiments*. URL: <https://github.com/mathieulagrange/doce?tab=readme-ov-file>.
- [16] Judith Liebetrau et al. “Quantifying auditory perception: Underlying dimensions of pleasantness and unpleasantness”. In: (2015).
- [17] Catherine Marquis-Favre. “Laboratory methods with imaginary and simulated contexts to assess noise annoyance: A comparison in terms of annoyance model testing”. In: *Acta Acustica united with Acustica* 105.6 (2019), pp. 1078–1085.
- [18] Brian McFee et al. “librosa: Audio and music signal analysis in python”. In: *Proceedings of the 14th python in science conference*. Vol. 8. 2015.
- [19] Nicolas Misdariis et al. “Community Noise Quality Assessment (Annoyance) by means of a Virtual Audio Environment”. In: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Vol. 259. 9. Institute of Noise Control Engineering. 2019, pp. 460–471.
- [20] M Mzali et al. “Auditory comfort on board of trains: passenger point of view”. In: *Proceedings of the 2000 internoise congress. Nice*. 2000.

- [21] Maria Rådsten Ekman, Peter Lundén, and Mats E Nilsson. “Similarity and pleasantness assessments of water-fountain sounds recorded in urban public spaces”. In: *The Journal of the Acoustical Society of America* 138.5 (2015), pp. 3043–3052.
- [22] Maria Rådsten-Ekman. *MAY NOISY SOUND ENVIRONMENTS BE IMPROVED BY ADDING PLEASANT WATER SOUNDS?* 2010.
- [23] Institut de Recherche et de Coordination Acoustique/Musique (IRCAM). *Blue Box*. Accessed: 04-2024.
- [24] Institut de Recherche et de Coordination Acoustique/Musique (IRCAM). *Sound Ideas*. Accessed: 04-2024.
- [25] Robert A Rescorla. “A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement”. In: *Classical conditioning, Current research and theory* 2 (1972), pp. 64–69.
- [26] Victor Rosi et al. “Best-worst scaling, an alternative method to assess perceptual sound qualities”. In: *JASA Express Letters* 2.6 (2022).
- [27] Thomas Sporer et al. “Distracting Noise”. In: *Audio Engineering Society Convention 140*. Audio Engineering Society. 2016.
- [28] Patrick Susini et al. “Characterizing the sound quality of air-conditioning noise”. In: *Applied acoustics* 65.8 (2004), pp. 763–790.
- [29] Juan Sebastián Ulloa et al. “scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python”. In: *Methods in Ecology and Evolution* 12.12 (2021), pp. 2334–2340.
- [30] Peter Vink and Susan Hallbeck. *Comfort and discomfort studies demonstrate the need for a new model*. 2012.
- [31] Eberhard Zwicker and Hugo Fastl. *Psychoacoustics: Facts and models*. Vol. 22. Springer Science & Business Media, 2013.

Annexes

Consigne présentée aux participants de l'expérience (section [4](#))

CONSIGNE

Présentation de l'expérience

Le but de cette expérience est d'étudier la perception de sons et l'agrément qu'ils suscitent.

L'expérience se déroulera en deux parties :

- Durant la première, 39 groupes de 4 séquences sonores vous seront présentées. Pour chaque groupe, vous devrez alors choisir la séquence qui vous paraîtra **la plus agréable**, et celle qui vous paraîtra **la moins agréable**.
- La seconde partie consistera à écouter 5 sons, et de les décrire grâce à un **processus de verbalisation**.

Déroulement de l'expérience

L'expérience se déroulera dans une cabine audiométrique.

Pour la première partie de l'expérience, un groupe de 4 séquences sonores vous sera présenté à chaque étape. En vous imaginant dans un espace que vous côtoyez au quotidien (par exemple votre espace de travail), vous devrez considérer chaque séquence sonore proposée comme constituant l'ambiance sonore de cet espace. Après quelques secondes d'écoute, vous devrez indiquer la séquence qui vous semble **la moins agréable** en la sélectionnant dans la colonne « - », et celle qui vous semble **la plus agréable** en la sélectionnant dans la colonne « + ». Vous pourrez écouter les séquences autant de fois que vous le souhaitez en appuyant sur les touches du clavier indiquées.

Cette partie est composée de 39 groupes. Avant de commencer, une phase d'entraînement constituée de 2 groupes vous sera proposée pour vous familiariser avec les séquences sonores et l'interface.

Concernant la seconde partie, 5 sons vous seront présentés. Vous devrez **décrire ces sons avec vos propres mots**, en utilisant l'encadré prévu à cet effet.

Fin de l'expérience

Une fois l'expérience terminée, **il vous faudra impérativement appuyer sur le bouton « Terminer l'expérience » pour enregistrer vos résultats.** Attention : ceux-ci ne seront pas enregistrés si vous fermez la page sans avoir appuyé sur ce bouton ! Si vous n'êtes pas sûrs d'avoir correctement enregistré vos résultats, vous pouvez laisser la page de l'expérience ouverte et demander à l'expérimentateur de vérifier.



L'expérience est maintenant terminée.

Merci de votre participation !

Appuyez sur le bouton ci-dessous pour finir l'expérience et enregistrer vos résultats.

Terminer l'expérience

Feuille de consentement présentée aux participants de l'expérience (section [4](#))

FEUILLE DE CONSENTEMENT

Bienvenue à l'Institut de Recherche et Coordination Acoustique/Musique (IRCAM), Laboratoire Sciences et Technologies de la Musique et du Son (STMS). Les chercheurs qui conduisent l'étude à laquelle vous allez participer sont : Nicolas Misdariis, nicolas.misdariis@ircam.fr, et Clara Boukhemia, clara.boukhemia@ircam.fr.

Dans cette étude, vous devrez : évaluer des séquences sons qui vous seront présentées par groupes de 4. L'évaluation consistera à choisir la séquence qui vous semblera **la plus agréable**, ainsi que celle qui vous semblera **la moins agréable**. Les séquences sonores présentées ont pour vocation à améliorer le jugement perçu, et ne seront donc pas trop désagréables ou trop fatigantes a priori. Le temps estimé nécessaire pour finir l'étude est de 30 minutes.

Nos études sont à visées académiques, et les résultats seront dévoilés dans des publications scientifiques. Nous ne réalisons pas d'études pour le compte d'entreprises privées. Il n'existe aucun risque lié à cette étude autre que ceux de la vie de tous les jours. Nous tenons néanmoins à préciser que l'étude se déroulera dans une cabine audiométrique, conçue pour vous isoler de tout bruit extérieur. Si vous ressentez un quelconque malaise en raison de cela, merci de bien vouloir nous en informer.

Si vous terminez cette étude, vous pourrez recevoir des informations sur ses conclusions si vous le souhaitez, ainsi que des références concernant le type de recherche auquel vous avez participé. Cependant, parce que vos réponses sont anonymes, nous ne pourrions vous renseigner que sur les résultats agrégés de l'étude, et non sur votre performance ou les performances de n'importe qui d'autre ayant participé.

Les données recueillies lors de cette étude seront placées sous la responsabilité d'un Référent pour la Protection des Données, qui en assurera la confidentialité et l'intégrité.

Votre participation à l'étude doit être entièrement volontaire, et vous avez la possibilité de vous retirer de l'étude à tout moment sans aucune pénalité.

Je déclare avoir l'âge légal de consentement, et ayant lu et parfaitement compris les paragraphes ci-dessus, accepte de mon plein gré de participer à cette étude.

DATE :

NOM et PRENOM :

SIGNATURE :

Questionnaire présenté aux participants à la fin de l'expérience (section [4](#))

QUESTIONNAIRE

Numéro du participant :

Date de l'expérience :

I – Informations personnelles :

Nom* :

Prénom* :

Adresse mail* :

Âge :

Genre : *Femme / Homme / Ne se prononce pas*

Activité professionnelle* :

Problèmes auditifs éventuels (pertes auditives, acouphènes, gênes dues à certains sons, ...) :

**champs optionnels*

II – Exposition et sensibilité au bruit :

- 1) Vous sentez-vous exposé(e) au bruit dans votre vie personnelle (lieu de vie, transports, loisirs, ...) ?

<i>Pas du tout</i>		<i>Modérément</i>		<i>Toujours</i>
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>

- 2) Utilisez-vous des écouteurs au quotidien ?

<i>Pas du tout</i>		<i>Modérément</i>		<i>Toujours</i>
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>

- | <i>Pas du tout</i> | | <i>Modérément</i> | <i>Toujours</i> | |
|--------------------|----------|-------------------|-----------------|----------|
| <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> |

- | <i>Pas du tout</i> | | <i>Modérément</i> | <i>Toujours</i> | |
|--------------------|----------|-------------------|-----------------|----------|
| <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> |

1) Dans quel environnement travaillez-vous (extérieur, bureau individuel, bureau collectif, télétravail, ...) ? Pourriez-vous décrire en quelques mots cet environnement ?

- | <i>Pas du tout</i> | | <i>Modérément</i> | <i>Toujours</i> | |
|--------------------|----------|-------------------|-----------------|----------|
| <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> |

- | <i>Pas du tout</i> | | <i>Modérément</i> | <i>Toujours</i> | |
|--------------------|----------|-------------------|-----------------|----------|
| <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> |

- 4) Imaginez-vous en train de travailler comme vous le faites généralement. Quels sont les sons que vous entendez pendant votre travail ?

- 5) Parmi ces sons, quels sont ceux qui vous gênent le plus (par exemple parce qu'ils vous empêchent de travailler, ou parce que vous les trouvez désagréables) ?

- 6) Pourriez-vous donner au moins trois idées de sons que vous aimeriez entendre si vous pouviez contrôler l'environnement sonore de votre lieu de travail ?
