

Mémoire de fin d'études

Utilisation de l'Intégrateur de Densité Spectrale (IDS) pour
améliorer la compréhension de la parole dans le bruit pour les
personnes équipées d'aides auditives.

GAËTAN RICCIUTI

Jury :

Gérard PELÉ	Président du Jury
Mohammed ELLIQ	Rapporteur
Dr. Bernard AURIOL	Directeur externe
Laurent MILLOT	Directeur interne

Remerciements

Tout d'abord, je tiens à remercier Laurent Millot, Directeur Interne de ce mémoire, pour son soutien et son aide précieuse tout au long de mes travaux et le Docteur Bernard Auriol, Directeur Externe de ce mémoire, pour son accueil chaleureux et ses conseils avisés. Merci également à Gérard Pelé pour ses enseignements en méthodologie de la recherche et à Étienne Hendrickx pour son aide dans la réalisation des tests perceptifs.

Merci à Antonio Duarte, Mohammed Ellic et Taïeb Keraoun pour leur aide précieuse dans la réalisation électronique de ce mémoire. Merci également à Bernard Auriol, M. Sarda et Pascale Behe pour la collecte de prothèses auditives.

J'aimerais remercier l'ensemble des personnalités présentes autour du Docteur Bernard Auriol lors de notre rencontre à Toulouse, vos réflexions et conseils m'ont véritablement guidé dans ma démarche.

Enfin, un grand merci aux personnes ayant participé aux enregistrements et aux tests perceptifs réalisés au cours du projet.

Résumé

Dans ce mémoire nous proposons l'utilisation de l'Intégrateur de Densité Spectrale (IDS) comme solution alternative aux méthodes de réduction de bruit fréquentielles intégrées dans les prothèses auditives. Nous cherchons à questionner la pertinence et l'efficacité de cet outil de traitement du signal pour favoriser la compréhension de la parole dans le bruit pour les malentendants appareillés. Nous présentons tout d'abord la modélisation d'un prototype de prothèse auditive intégrant l'IDS et la conception d'un dispositif de captation basé sur le microphone d'une prothèse. Puis nous décrivons comment ce dispositif a permis de constituer un corpus de scènes sonores représentatives des situations d'écoute difficiles lors de conversations dans des environnements bruyants. Enfin, nous présentons la campagne de tests perceptifs menée à partir de ce corpus et auprès de malentendants, pour confronter l'IDS aux réalités perceptives des prothèses auditives.

Mots clés : réduction de bruit, analyse-resynthèse IDS, prothèses auditives, compréhension de la parole dans le bruit

Abstract

In this research memoir we study the use of Spectral Density Integrator (IDS) as an alternative solution for frequency noise reduction techniques used in hearing aids. First we present the modelling of a hearing aid prototype using IDS and the design of a recording tool using hearing aids microphone. Then we describe how a corpus was built with representatives audio scenes of difficult listening situations, using the recording tool designed. Finally we present how these audio scenes were used to conduct a test campaign with hearing-impaired persons to test the efficiency of IDS to improve the speech recognition in noise.

Key words : noise reduction, IDS analysis and synthesis, hearing aids, speech recognition in noise

Table des matières

1	Introduction générale	7
1.1	L'écoute	7
1.2	L'effet cocktail party	9
1.3	La réhabilitation auditive	10
1.4	Démarche	12
2	Les prothèses auditives aujourd'hui	13
2.1	Cadre général	13
2.2	La réduction de bruit	21
3	Une nouvelle approche pour la compréhension de la parole dans le bruit	27
3.1	L'Intégrateur de Densité Spectrale (IDS)	27
3.2	Le travail effectué par Titouan Ralle	35
3.3	Conclusion	41
4	Vers un prototype de prothèse auditive intégrant l'analyse-resynthèse IDS	43
4.1	Introduction	43
4.2	Dispositif de captation	44
4.3	Dispositif de diffusion	52
4.4	Traitement du signal en temps réel	54
4.5	L'interface utilisateur	57
4.6	Conclusion	62
5	Réalisation d'un corpus de scènes sonores à partir d'un prototype de prothèse auditive	63
5.1	Introduction	63

5.2	Les raisons de la constitution d'un corpus original	63
5.3	Composition du corpus	65
5.4	Dispositif d'enregistrement	71
5.5	Conclusion	73
6	Protocole du test perceptif auprès de malentendants	75
6.1	Introduction	75
6.2	Présentation générale du protocole de test	76
6.3	Améliorations de l' <i>IDS Speech Enhancer</i>	78
6.4	Conclusion	83
7	Résultats des tests et améliorations du dispositif	85
7.1	Résultats des tests	85
7.2	Améliorations du dispositif de test	94
8	Conclusion générale	97

1 Introduction générale

« L’ouïe est sans doute, avec la vue, l’un des sens qui apporte à l’être humain la plus grande quantité d’informations sur l’environnement qui l’entoure. »

Jéronimo Labranda, *El registro Sonoro*[1]

1.1 L’écoute

Parmi nos cinq sens, l’ouïe est certainement celui qui permet à l’être humain de réaliser l’analyse la plus complète de son environnement. La mise en vibration de nos deux tympans selon des infimes¹ variations de pression nous donne accès en permanence, grâce à un ensemble de phénomènes psychoacoustiques et cognitifs, à une multitude d’informations sur le monde sonore alentour. Nous pouvons non seulement distinguer les différents éléments qui composent notre environnement, mais aussi déterminer approximativement leur position, leur déplacement, leur dimension, leur nature, etc. L’analogie proposée par Albert Bregman et rapportée par Jean-Claude Risset est saisissante :

« Observons deux bouchons au bord d’un lac, mus par les ondes se propageant à la surface. Ne serait-ce pas un exploit d’en déduire les positions et les mouvements des poissons ou des autres êtres subaquatiques qui sont à la source de ces ondes ? »

[2, préface]

Si cet exploit est possible, c’est grâce au caractère actif de l’audition. Notre oreille n’est pas uniquement un capteur passif réagissant à des stimulations, comme le fait un microphone.

1. Au seuil de l’audition et à 1 000 Hz, les vibrations du tympan sont d’une amplitude de 10^{-11} m, soit de l’ordre du 1/10 du diamètre de l’atome d’hydrogène.

L'écoute est une action, bien qu'elle soit principalement inconsciente. Plusieurs phénomènes physiologiques en attestent, comme l'adaptation d'impédance² ou le réflexe stapédien³ qui interviennent au niveau de l'oreille moyenne. Mais surtout, l'audition est active car l'analyse qu'elle effectue s'appuie sur une hiérarchisation des informations sonores parvenant jusqu'à nos oreilles selon un jeu de préférences et de répulsions. Comme l'écrit le Docteur Bernard Auriol, cette hiérarchisation perceptive mobilise toutes les expériences antérieures d'exposition aux environnements sonores.

« Ce processus actif est évidemment présent lorsque tout l'être aiguise sa perception pour saisir tel signal précis dans la masse confuse des bruits extérieurs : la voiture de l'aimé, la sirène de police, la pub télévisée, etc. L'écoute est donc modelée par les expériences de toute la vie antérieure depuis que l'audition est fonctionnelle : environnement sonore familial, paysage sonore de la collectivité rurale ou urbaine. » [2, p.15]

Parmi les différents sons qui forment la masse confuse de bruits extérieurs, la voix attire tout particulièrement notre attention. *La parole* est le moyen de communiquer le plus utilisé par l'être humain. L'évolution de la communication verbale est d'ailleurs l'un des facteurs de l'un des tournants dans l'histoire de notre espèce, la révolution cognitive⁴. Si la parole joue un rôle si crucial dans notre système de communication, c'est qu'elle est à la fois porteuse de sens et d'émotion, le tout avec de nombreuses subtilités. Un simple « Allô » ou « Bonjour » au téléphone nous permet, malgré la piètre qualité⁵ du canal de transmission, de reconnaître la voix d'un proche ou de déterminer le sexe et l'âge d'un interlocuteur inconnu. Les différences de

2. Les deux muscles de l'oreille moyenne (muscle stapédien et muscle du marteau) se contractent plus ou moins en fonction de l'intensité des vibrations reçues et transmises par le tympan. Ces contractions musculaires correspondent à une adaptation d'impédance du signal acoustique.

3. Dans des cas extrêmes d'intensités sonores fortes et brèves, le réflexe stapédien correspond à une contraction soudaine de ces mêmes muscles pour protéger les cellules neuro-sensorielles de l'organe de Corti.

4. La révolution cognitive, qu'on estime s'être déroulée il y a entre 70 000 et 30 000 ans, a été marquée par l'apparition de nouvelles façon de penser et de communiquer. Dans *Homo Sapiens*, Yuval Noah Harari raconte que l'un des moteurs des évolutions linguistiques du Sapiens était le commérage, ce qui donnait au caractère social de notre espèce une nouvelle dimension [3].

5. La bande passante du téléphone est comprise entre 300 Hz et 3,4 kHz.

hauteur et de timbre donne à chaque voix un caractère unique, une identité, et une même voix peut refléter des émotions très différentes en fonction de l'état psychologique de son locuteur. Comme le disait Aristote, « *Les sons émis par la voix sont les symboles des états de l'âme.* » [4, p.89]

1.2 L'effet cocktail party

Le processus de hiérarchisation des informations sonores par notre système auditif est permanent, mais certaines situations nous montrent encore davantage sa performance. Prenons l'exemple d'une grande salle de réception remplie de gens, dans laquelle se mêlent nombre de conversations, une musique de fond ainsi que des bruits ponctuels en tout genre. Dans une telle situation, nous arrivons à suivre une discussion principale tout en étant attentifs à d'autres conversations secondaires, et ceci en restant en alerte en cas d'évènements sonores brefs pouvant signaler un danger. Cette capacité de focalisation de notre attention sur un contenu sonore (souvent la parole) dans des environnements bruyants a été théorisée par Colin Cherry sous le nom d'*effet cocktail party*⁶.

Ce processus principalement inconscient nous permet de nous adapter constamment aux différents environnements sonores que nous rencontrons, même les plus bruyants (café, métro, concerts), pour continuer à communiquer avec nos semblables. Notre audition fonctionne sur une analyse de l'ensemble des évènements sonores et une séparation entre l'information importante et le reste, que l'on considère dès lors comme du *bruit*. La notion de bruit est polysémique et il convient de la définir en fonction du contexte donné⁷. En physiologie de l'audition, et en particulier dans le cas de la compréhension de la parole, nous pouvons reprendre la définition donnée par Titouan Ralle dans son mémoire de fin d'études : « *Son ou ensemble de sons qui altèrent un signal vocal et provoquent une gêne pour sa compréhension* ». [6, p.15]

6. Le terme « cocktail party » apparaît pour la première fois dans l'article « Some experiments on the recognition of speech with one and two ears » rédigé par Colin Cherry en 1953 [5].

7. Les normes de l'acoustique définissent le bruit à la fois comme une « vibration acoustique erratique intermittente ou statistique aléatoire » (NF S 30-101) et comme « toute sensation auditive désagréable ou gênante » (NF S 30-105). En électronique, le bruit correspond à un signal aléatoire qui altère le signal utile.

L'effet cocktail party démontre la performance de notre audition qui mobilise des phénomènes psychoacoustiques aussi complexes que fragiles. La compréhension de la parole dans le bruit est l'une des principales problématiques de la réhabilitation auditive, à la fois parce que ce sont dans ces situations que les déficiences auditives sont les plus handicapantes et parce que cette séparation de la parole au sein de l'environnement sonore est complexe à réaliser en traitement du signal.

1.3 La réhabilitation auditive

Le déficit auditif concerne 10 %⁸ de la population française, soit environ 6 millions de personnes. Selon L'OMS, c'est environ 5 % soit 360 millions de la population mondiale qui serait concernée. Les personnes âgées de plus de 50 ans sont les plus touchées car l'une des causes principales du déficit auditif est un phénomène naturel de vieillissement de l'oreille, la presbycusie. Or, la population française est de plus en plus âgée, en raison de différents facteurs démographiques (baby boom, allongement de l'espérance de vie). Ces données font du déficit auditif un enjeu de santé publique majeur lorsqu'on sait son impact sur l'état de santé psycho-social des personnes présentant des déficiences.

« Si le déficit auditif se manifeste par des limitations fonctionnelles directes (difficulté de compréhension et de communication), il est également associé à une plus grande fréquence des troubles mentaux, du déclin cognitif, des chutes et même de la mortalité, indépendamment des effets de l'âge ou du sexe. Le déficit auditif peut donc précipiter les personnes âgées vers la dépendance. » [7, Synthèse]

Dans de nombreux cas de déficiences auditives, une réhabilitation de l'audition est envisageable grâce à l'usage de prothèses auditives. En 2015, on estimait à 3 millions le nombre de français étant éligibles aux prothèses auditives, alors que 2 millions d'entre eux en bénéficiait. Le

8. En France, la prévalence du déficit auditif est estimée entre 8,6 % et 11,2 % de la population en fonction des études [7, Synthèse].

principal frein pour les personnes qui ne sont pas encore appareillées est financier. Le prix moyen de l'appareillage pour une oreille est de 1535 € alors, qu'en moyenne, seul 8 % du coût est pris en charge par l'assurance maladie obligatoire et 30 % par l'assurance complémentaire, laissant 950 € à la charge du patient.

L'évolution technologique récente, tant du point de vue de la miniaturisation des composants électroniques que du développement des algorithmes de traitement du signal numérique, a permis aux prothèses auditives d'être de plus en plus performantes. Cependant, plusieurs études de satisfaction révèlent que les personnes appareillées se plaignent de leurs prothèses. Une enquête publiée en 2014 par *Audio 2000* et *Senior Strategic*, annonce que sur 1487 personnes utilisant des prothèses auditives, seulement 35 % en sont satisfaites [8]. Une autre, réalisée par *Que Choisir* en 2015 sur 2721 personnes, indique que : « *Malgré les aides auditives, 38 % des répondants se trouvent confrontés à des situations d'écoute difficiles* »[9]. Parmi ces situations d'écoute difficiles, les discussions dans des environnements bruyants sont celles qui sont les plus insatisfaisantes, puisque la réhabilitation prothétique ne permet qu'une compréhension partielle, voire faible, de la parole. Dans le cadre de ce mémoire nous avons recolté des témoignages de personnes malentendantes, l'extrait suivant nous semble représentatif des problématiques qu'ils rencontrent quotidiennement.

« La compréhension dans le bruit environnant est l'une des plus fortes attentes chez les malentendants. Autrement dit, faire en sorte que le malentendant, dans un environnement bruyant, ne soit ni coupé du monde, ni abruti par le bruit, qu'il puisse comprendre ceux qui s'adressent à lui à faible distance mais encore trop loin pour être compréhensible si d'autres parlent en même temps. »

1.4 Démarche

Dans ce mémoire de fin d'études, nous questionnerons le fonctionnement et la performance des traitements de réduction de bruit intégrés dans les prothèses auditives. En partant du constat que les situations d'écoute les plus exigeantes, celles de conversations dans des environnements bruyants, n'apportent qu'une satisfaction mitigée à leurs utilisateurs, nous proposerons un outil alternatif de traitement du signal dans le but d'améliorer la compréhension de la parole dans le bruit. Par cette démarche nous nous plaçons dans la continuité du mémoire de fin d'études de Titouan Ralle, diplômé de l'*École Nationale Supérieure Louis Lumière* en 2017. Son travail a permis de présenter l'analyse-resynthèse par Intégration de Densité Spectrale (IDS, outil que nous expliquerons plus largement par la suite, puisqu'il tient une place centrale dans notre recherche) comme une solution d'amélioration de la compréhension de la parole dans le bruit. Deux pistes de prolongements avaient été esquissées en conclusion de ses travaux, que nous avons voulu approfondir simultanément :

- confronter cet outil de traitement du signal aux contraintes technologiques réelles des prothèses auditives, et à la perception de sujets malentendants ;
- améliorer et reconduire les tests perceptifs permettant de tester la pertinence et l'efficacité de l'IDS pour améliorer la compréhension de la parole dans le bruit.

Tout d'abord nous présenterons le contexte général de la réhabilitation auditive et le fonctionnement particulier des prothèses auditives. Puis nous détaillerons l'historique et le fonctionnement de l'analyse-resynthèse IDS, ainsi que le travail effectué par Ralle dans son mémoire de fin d'études. Ensuite, nous aborderons la modélisation et la conception partielle d'un prototype de prothèse auditive intégrant l'analyse-resynthèse IDS. Puis, nous présenterons le corpus de scènes sonores constitué à partir d'un dispositif de captation original basé sur une prothèse auditive et la campagne de tests perceptifs menée auprès de personnes malentendantes à partir de ce corpus. Enfin, nous présenterons et analyserons les résultats de ces tests.

2 Les prothèses auditives aujourd’hui

2.1 Cadre général

2.1.1 La déficience auditive

La déficience auditive est un dysfonctionnement d’ordre psychoacoustique du système auditif. Ce n’est pas une maladie à proprement parler, bien que cette altération puisse être la conséquence d’une maladie ou d’un accident. La déficience ou *perte* auditive est définie par une élévation du seuil d’audition qui peut être mesurée grâce aux différentes techniques d’audiométrie et d’acoumétrie. Par exemple, l’audiométrie tonale liminaire mesure le seuil d’audition de sons purs à différentes fréquences. Les résultats sont reportés sur un *audiogramme* (Figure 1) qui permet de visualiser facilement, pour chacune des fréquences mesurées et pour chaque oreille, la perte auditive en décibels (dB)⁹.

Les déficiences auditives peuvent être classées selon plusieurs critères, que ce soit en fonction de l’importance de la perte, de la cause de l’altération (étiologie), des fonctions touchées (surdit  de transmission, cochl aire, nerveuses, mixtes). De plus, une s paration socio-culturelle importante se situe entre les personnes sourdes de naissance et les d ficients auditifs entendants   la naissance et devenus sourds plus ou moins tardivement.

9. Le d cibel est une unit  qui, en suivant une  chelle logarithmique plut t que lin aire, respecte davantage la sensation auditive.

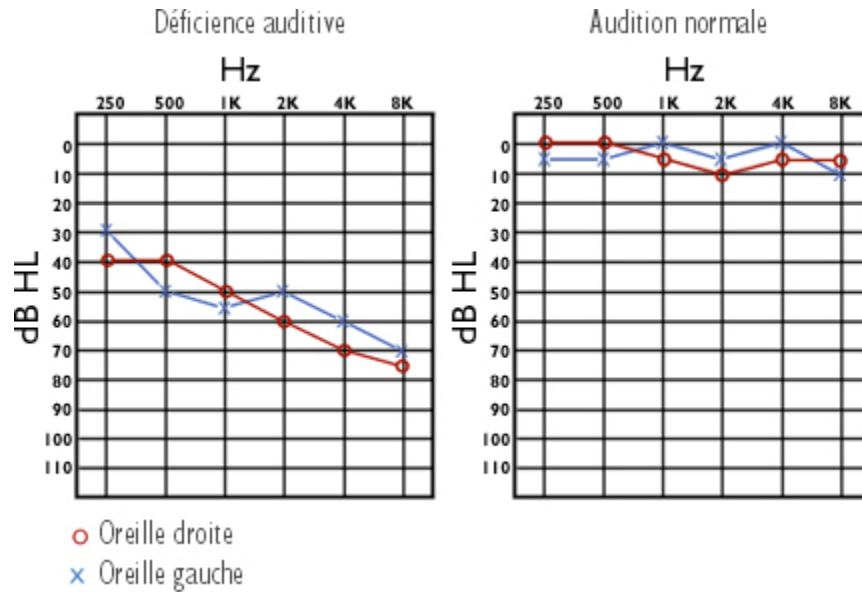


FIGURE 1 – Exemples d’audiogrammes respectivement associés à des sujets présentant une déficience auditive moyenne (à gauche) et une audition normale (à droite). D’après [10], p. 1.

En effet, dans un cas la réhabilitation auditive exige des moyens lourds, qui ne sont pas toujours souhaités par les personnes concernées (sourds sévères ou profonds) :

« Ces personnes utilisent volontiers la communication extra-auditive en particulier la communication visuelle et la langue des signes. Elles ne sont pas toujours intéressées par une réhabilitation auditive, il est vrai complexe, faisant par exemple appel à l’implantation d’électrodes intracochléaires suivie d’un soutien éducatif quasi-permanent. Nombreux de ces malentendants se réclament d’un « monde des sourds » avec son mode de vie, sa culture,... et n’acceptent qu’avec difficulté de calquer leur mode de communication et leur vie sociale sur celle des entendants. »

[11, p. 113]

Dans l’autre cas, les personnes ne s’identifient pas à des sourds de la même manière et une réhabilitation auditive satisfaisante peut être obtenue grâce à des moyens assez simples, comme les aides auditives.

La plus répandue des déficiences auditives est la *presbyacousie*, elle correspond au vieillissement « naturel » du système auditif. En effet, l’audition est l’une des modalités sensorielles qui vieillit le plus mal car les structures neuro-sensorielles (cellules ciliées, neurones de la cochlée) sont présentes en petit nombre¹⁰ et disparaissent peu à peu, entraînant une perte irrémédiable de l’audition avec l’âge.

« Dans les sociétés où l’espérance de vie ne cesse d’augmenter, la surdité liée à l’âge ou la presbyacousie est un problème de santé publique. En effet, 70 % des plus de 65 ans ont des problèmes auditifs. » [13]

Mais, surtout, la presbyacousie n’est pas indépendante des surdités acquises. Autrement dit ce phénomène de perte « normale » de l’audition peut être accéléré par des déficiences auditives supplémentaires, causées par exemple par la prise de médicaments ototoxiques¹¹ ou par des *traumatismes acoustiques*. La Figure 2 rapporte la comparaison faite sur le site du *Voyage au centre de l’audition* entre les audiogrammes représentant une presbyacousie « normale » et une presbyacousie « précoce » estimée pour des personnes ayant subi de nombreux traumatismes acoustiques.

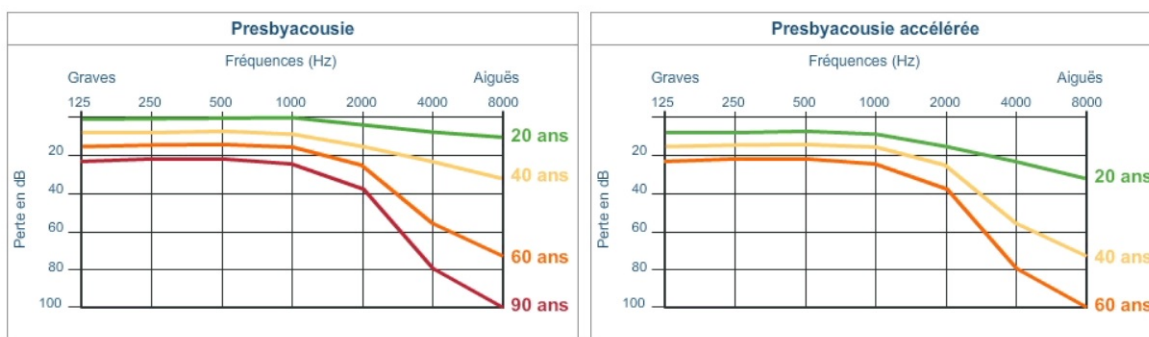


FIGURE 2 – Audiogrammes d’une presbyacousie « normale » à gauche, et d’une presbyacousie « précoce » à droite. D’après [13], p. 1.

10. Le canal cochléaire comporte 3500 cellules ciliées internes et 12 000 cellules ciliées externes. Ce sont ces dernières qui sont directement réceptrices, et donc indispensables à une écoute correcte [12, p. 36].

11. Une substance chimique est dite ototoxique lorsqu’elle est toxique pour les structures de l’oreille interne. L’ototoxicité est souvent un effet secondaire indésirable d’un médicament, elle en concerne actuellement une centaine.

Les traumatismes acoustiques sont la conséquence d'une exposition à un son trop fort ou durant une trop longue durée. Il en résulte une destruction irréversible de nos cellules sensorielles, ce qui provoquent des acouphènes et une surdité. Ces expositions à des intensités sonores trop fortes sont un enjeu de société considérable, elles concernent à la fois des problématiques d'habitudes d'écoutes (smartphone et écouteurs, concerts), de santé du travail, de nuisance sonore dans l'espace urbain, etc. Plusieurs organisations, telle que la *Semaine du son* et la *Journée Nationale de l'Audition* (JNA) en France, se sont emparées du sujet et sensibilisent le grand public à ces problématiques.

Mais cette sensibilisation semble loin d'être suffisante puisque des études démontrent une détérioration considérable de la santé auditive des français, notamment des plus jeunes. En effet, la dernière enquête JNA-Ifop¹² réalisée dans le cadre de la 22e *Journée Nationale de l'Audition* (le 14 Mars 2019), tire une sonnette d'alarme tant ses résultats sont inquiétants. Selon cette étude, bien que « 89 % des 15-17 ans qualifient leur audition de bonne ou très bonne », ils sont pourtant nombreux à rencontrer des gênes pour suivre des conversations dans des situations courantes de la vie : « 70 % dans les cafés, bars restaurants, 70 % dans leur établissement scolaire, 80 % dans l'espace public, 54 % dans les magasins, 76 % au téléphone, 74 % dans les transports en commun et 52 % pendant les repas de famille. » De plus, le nombre de jeunes français ayant des sensations d'acouphènes est conséquent, et a considérablement augmenté en une année (56 % en 2018 contre 65 % en 2019 chez les 15-17 ans et 49 % contre 59 % chez les 18-24 ans).

« Cette augmentation nous montre à quel point les plus jeunes sont sujets aux troubles de l'audition, probablement liés à des pratiques à risques (volume sonore trop élevé, exposition trop longue, etc.) Mais pas seulement (somme des expositions sonores sur la journée, peu de temps de repos auditif). Plus que la prévention des jeunes, la vigilance sanitaire effective est nécessaire pour endiguer l'augmentation de l'apparition des acouphènes chez les plus jeunes, symptômes ORL évitables. »[14]

12. L'institut français d'opinion publique (Ifop) s'associe chaque année avec la Journée Nationale de l'Audition (JNA) pour réaliser une enquête concernant la santé auditive des français. Cette année, elle a été effectuée au près de 1003 personnes âgées de 15 ans et plus.

Outre un accroissement des troubles auditifs chez les français, c'est un vieillissement de la population qui est à prévoir. En effet, l'*INSEE* estime qu'en 2060, « un français sur trois aura plus de 60 ans » [15] et qu'en 2070, « la population âgée de 75 ans ou plus serait deux fois plus nombreuse qu'en 2013 » [16]. Ces deux estimations amèneraient donc inévitablement à une forte augmentation dans les années à venir des personnes présentant des déficiences auditives.

2.1.2 La réhabilitation prothétique

La réhabilitation de l'audition peut-être envisagée en fonction du type de déficiences auditives et peut prendre plusieurs formes. Dans certains cas la restauration du système auditif est possible par voie chirurgicale ou thérapeutique, mais lorsqu'elle est impossible, la réhabilitation de la communication s'effectue par traitement du stimulus. La *réhabilitation prothétique* est le moyen le plus répandu pour rétablir la communication audio-verbale.

« Le principe est de transformer le stimulus, avec ou sans changement de nature, de telle sorte que, sous sa nouvelle forme, il soit susceptible de stimuler le récepteur défaillant. La réussite de la réhabilitation dépend de nombreux facteurs : choix du dispositif, adéquation du stimulus à la capacité auditive du sujet, plasticité cérébrale du malentendant, conservation du contenu informatif du stimulus, qualité de la rééducation sensorielle. » [11, p. 118]

Les différents dispositifs de réhabilitation prothétique comprennent tous :

- un microphone, transducteur transformant une énergie mécanique en signal électrique ;
- un dispositif de traitement du signal analogique ou numérique ¹³ ;
- un dispositif de stimulation qui délivre le stimulus sous forme d'une énergie acoustique, mécanique ou électrique.

13. L'arrivée en 1996 de traitements numériques du signal au sein des aides auditives a considérablement amélioré leur performance, de telle sorte que la quasi-totalité d'entre elles en sont actuellement équipées.

Les catégories de dispositif de réhabilitation prothétique dépendent de la nature de la stimulation effectuée, et donc de la nature des déficiences auditives :

- un stimulus acoustique peut être délivré au niveau de l'oreille externe (aides auditives ou *prothèses auditives*);
- un stimulus mécanique vibratoire peut être délivré pour stimuler la chaîne tympano-ossiculaire (prothèse de l'oreille moyenne) ou les parois de l'oreille interne (vibrateur en conduction osseuse);
- un stimulus électrique peut être utilisé pour exciter le système nerveux auditif (implants cochléaires).

Dans le cadre de ce mémoire nous nous intéressons en particulier aux prothèses auditives. C'est la solution de réhabilitation prothétique la plus répandue étant donné sa simplicité de mise en place et donc son coût¹⁴. Les prothèses de l'oreille moyenne ne sont envisagées que lorsque les prothèses auditives ne peuvent être utilisées (eczéma, otite externe, sténose,...) et les implants cochléaires seulement pour des surdités sévères (supérieures à 90 dB) pour lesquelles les prothèses auditives se sont avérées inefficaces. D'autres formes de réhabilitation auditive (implant du tronc cérébral, prothèse à ancrage osseux) peuvent être convoquées dans des cas très particuliers, nous ne les détaillerons pas davantage.

La figure 3 donne le fonctionnement général des prothèses auditives, qui peuvent être assimilées à des *amplificateurs correcteurs*. Le signal capté par le microphone est converti en information numérique et est traité avec un gain variable en fonction des fréquences, pour s'adapter à la déficience auditive du sujet et favoriser les fréquences permettant la compréhension de la voix. Le traitement du signal numérique mis en jeu ici faisant l'objet de ce mémoire, nous l'aborderons plus largement par la suite. Le réglage des amplificateurs correcteurs est laissé à la discrétion de l'audioprothésiste en fonction de l'audiogramme du malentendant, l'alimentation est quant à elle généralement assurée par une pile que l'utilisateur doit régulièrement remplacer.

14. En moyenne, une prothèse auditive coûte 1600 €, une prothèse de l'oreille moyenne 9000 € (6200 € pour l'opération et 2800 € pour le processeur externe), et 23000 € pour l'implant cochléaire [17].

Le signal une fois traité est diffusé via un écouteur qui s'insère dans le canal auditif externe.

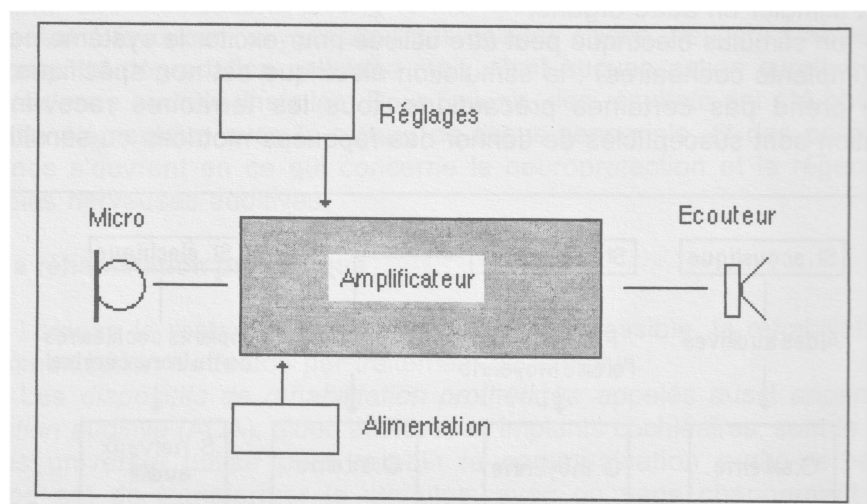


FIGURE 3 – Schéma de principe d'une aide auditive, ou *prothèse auditive*. D'après [11], p. 120.

« Le signal de sortie est rayonné dans le conduit auditif externe par un écouteur. Ce stimulus possède des caractéristiques assez proches de celles du message sonore initial, et de ce fait, l'éducation prothétique est simplifiée et l'adaptation du malentendant à son appareil est le plus souvent rapide. » [11, p. 120]

Les différents composants constituant la prothèse peuvent soit être intégrés dans une coque moulée sur mesure et introduite à l'entrée du conduit auditif externe (ITE : *In the ear*, ITC : *in the canal*) ou dans le conduit auditif externe (CIC : *completely in the canal*), soit être rassemblés dans un boîtier externe placé derrière le pavillon (BTE : *Behind the ear*). Les vibrations acoustiques produites par l'écouteur dans le boîtier sont alors transmises jusqu'au conduit auditif grâce à un guide d'onde et un embout auriculaire. La figure 4 montre ces quatre formats de prothèses.



FIGURE 4 – Photographies des quatre formats de prothèses auditives, des plus visibles au plus discrètes : en haut : derrière l’oreille (à gauche), dans l’oreille (à droite) ; en bas : dans le canal (à gauche), complètement dans le canal (à droite). D’après [18], p. 1.

N’apparaît pas sur cette figure le RIC (*Receiver in the canal*), format le plus répandu dans les prothèses haut de gamme. En effet, la miniaturisation des composants permet, avec ce format, de loger l’écouteur non pas dans la prothèse (et de transmettre les vibrations via un guide d’onde jusqu’au canal auditif, comme pour les BTE), mais directement dans le canal auditif.

Si ces différents formats ont de l’importance, c’est parce qu’ils influent sur la discrétion de l’appareillage. Or l’utilisation de prothèses auditives est une démarche d’acceptation d’une déficience perceptive qui est difficile à vivre, souvent plus difficile que le port de lunettes de vue. La visibilité du dispositif de réhabilitation peut donc jouer un rôle majeur dans cette démarche d’acceptation.

2.2 La réduction de bruit

2.2.1 Les contraintes d'utilisation

Les corrections apportées lors du traitement du signal au sein d'une prothèse auditive ne consistent pas seulement à compenser les pertes auditives mesurées pour chaque fréquence caractéristique grâce à l'audiométrie tonale. Plusieurs outils matériels et logiciels sont mis en œuvre pour permettre une *réduction du bruit*. Comme nous l'avons vu en introduction, la notion de bruit est polysémique et doit être définie pour chaque contexte donné. Dans nos recherches, le bruit représente tous les sons perturbant l'intelligibilité de la parole.

Le développement des télécommunications mobiles a fait de la réduction de bruit un domaine de recherche de plus en plus important. Quelle que soit leur utilisation, la contrainte majeure des techniques de réduction de bruit est le compromis entre la performance de la réduction et la distorsion qui affecte le signal utile. Autrement dit, dans le cas des prothèses auditives, il faut à la fois isoler suffisamment la parole de l'environnement bruyant, sans induire de perturbations qui gêneraient la compréhension de cette parole.

L'autre contrainte qui s'applique au cas des prothèses auditives, comme à de nombreuses situations de télécommunications, est la notion de *temps réel*. En effet, toutes les opérations de traitement du signal effectuées entre sa captation et sa diffusion dans l'écouteur ne doivent pas induire des latences¹⁵ supérieures à 20 ms, car cela perturberait l'écoute globale et en particulier l'écoute de la parole.

2.2.2 Le rapport signal sur bruit (RSB)

Pour qu'une réduction du bruit soit possible, il faut pouvoir procéder à une séparation de l'ensemble du contenu sonore environnant entre l'information utile et le reste. Le *rapport signal sur bruit* (RSB ou R/B) permet de mesurer cette séparation en décibel. Un RSB de 0 dB correspond à un niveau équivalent entre le signal et le bruit, et plus le RSB est grand, plus la

15. Dans une chaîne de traitement du signal, une latence désigne un retard entre l'entrée et la sortie de la chaîne.

Perte auditive moyenne	Perte de rapport signal sur bruit
30 dB HL	4 dB
40 dB HL	5 dB
50 dB HL	6 dB
60 dB HL	7 dB
70 dB HL	9 dB
80 dB HL	12 dB*
90 dB HL	18 dB*

FIGURE 5 – Comparaison entre la perte auditive et la perte de RSB, d’après [19], p. 14.

sensation d’isolement du contenu sonore utile par rapport au bruit ambiant sera conséquente. Cet indicateur permet donc de mesurer la performance des différentes méthodes de réduction de bruit, et de les comparer entre elles. Il permet également de mettre en relation les pertes auditives avec la difficulté à effectuer cette séparation entre la parole et le bruit, comme le montre la Figure 5.

2.2.3 Les différentes méthodes de réduction de bruit

Pour améliorer la réduction de bruit, les prothèses auditives cumulent plusieurs approches qui travaillent chacune sur un critère psychoacoustique différent.

1. Le critère temporel part du principe que le bruit est stationnaire, à savoir qu’il présente des faibles variations d’amplitude au cours du temps. A l’inverse, on considère que la parole est un signal présentant des variations d’amplitude élevées. Dès lors, une analyse des variations d’amplitude permet de mesurer en continu l’évolution du RSB et donc de détecter la présence de parole. Cette détection permet ensuite d’appliquer une amplification qui va dépendre de la mesure du RSB (plus il est élevé, moins il faut amplifier) pour au final effectuer une réduction de bruit. Ces traitements basés sur l’évolution temporelle du signal peuvent se faire sur différents canaux fréquentiels (sous-bandes résultantes d’un filtrage préalable) pour améliorer davantage leur performance. Cependant, l’inconvénient principal de cette approche est qu’elle ne permet pas de traiter les bruits d’impacts.

En effet, certains évènements sonores perturbant la compréhension de la parole (comme les bruits de manipulations de couverts dans le contexte d'une salle de réception ou d'un restaurant) sont très brefs, ils présentent donc des variations d'amplitude importantes, ils seront donc considérés par cette méthode comme de l'information utile et seront amplifiés.

2. Le critère spatial est fondé sur la localisation du signal utile (la place de l'interlocuteur dans l'espace) face au bruit environnant qui, étant constitué d'une multitude de sources, n'a pas d'emplacement fixe. Des dispositifs de captation directifs mobilisant plusieurs microphones, souvent deux, permettent d'exploiter cette discrimination spatiale. Schématiquement, si on part du principe que le locuteur s'exprime face au sujet malentendant alors un microphone présent dans la prothèse et orienté vers l'avant permettra de capter un signal mêlant la parole et le bruit. A l'inverse, un microphone orienté vers l'arrière captera principalement du bruit. En créant un retard entre les deux microphones on met en opposition de phase les signaux avant et arrière, permettant de récupérer un signal dont le bruit a été considérablement diminué. Lorsqu'il sont gérés électriquement, le retard ainsi que la pondération des deux signaux fournis par les microphones peuvent être modifiés au niveau du processeur, afin de créer un nombre infini de diagrammes polaires. La directivité de ces systèmes à deux microphones peut donc être modifiée automatiquement au cours du temps et peut être différente en fonction du canal fréquentiel, ce qui permet une adaptabilité permanente en fonction des situations rencontrées.
3. De manière analogue aux deux autres critères, l'approche fréquentielle suppose une caractérisation spectrale distincte du signal utile par rapport au bruit. La plupart des méthodes de réduction de bruit fréquentiel sont basées sur le principe de l'*atténuation spectrale à court terme*. Son principe consiste à réaliser une analyse à court terme du signal original bruité, en considérant que la parole est quasi-stationnaire sur des trames de 20 à 40 ms, et d'en déduire une estimation du signal de parole utile. Les travaux de Cyril Plapous [20] nous permettent de présenter dans ses grandes lignes le fonctionnement de l'atténuation spectrale à court terme, ainsi que d'en présenter une schématisation en Figure 6.

- Le signal bruité est découpé en trames et chaque trame est fenêtrée. Une TFD (Transformée de Fourier Discrète) de chaque trame d’analyse est effectuée pour passer dans le domaine fréquentiel.
- Une estimation de la densité spectrale de puissance (DSP) du bruit à long terme est calculée à partir du module du signal (en se basant sur le caractère stationnaire du bruit).
- Le calcul du gain spectral évalué est effectué à partir de l’estimation de la DSP du bruit et du RSB.
- Le module du signal de parole estimé ainsi que la phase du signal original bruité sont utilisés pour revenir dans le domaine temporel au moyen d’une TFD inverse (TFDI).

Parmi les différentes approches permettant une réduction de bruit au sein des prothèses auditives, celle étant la plus performante actuellement est la discrimination spatiale basée sur une directivité adaptative, comme en atteste l’étude menée par trois audioprothésistes et publiée dans les *Cahiers de l’audition* en 2014 : « *Selon nos analyses, les microphones directionnels améliorent significativement le RSB en cas de bruit vocal ou non vocal. Ils semblent potentialiser l’action des débruiteurs* » [25].

Les approches temporelles et fréquentielles (puisque’elles sont majoritairement basées sur l’atténuation spectrale à court terme) restent limitées par la nécessité de considérer le bruit comme stationnaire, ce qui n’est pas le cas de toutes les sources sonores bruyantes. Pourtant, l’approche fréquentielle présente un potentiel inexploité, puisqu’elle pourrait permettre de caractériser fréquentiellement le signal utile (la parole) et le bruit, sans nécessairement s’appuyer sur leur stationnarité. C’est en tout cas l’axe de recherche que nous souhaitons explorer en mobilisant l’Intégrateur de Densité Spectrale.

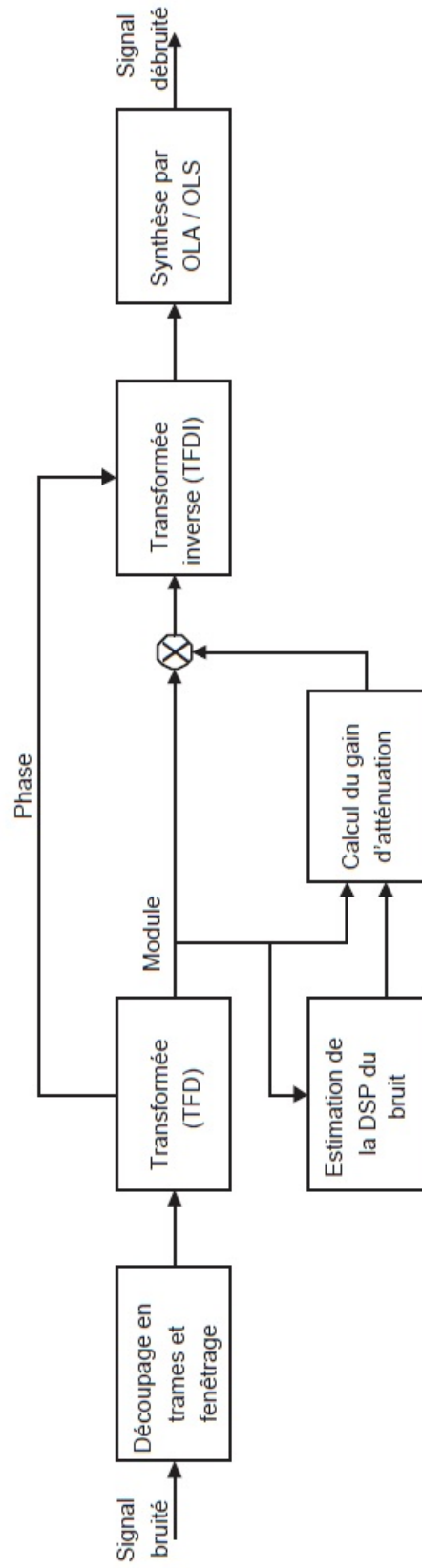


FIGURE 6 – Schéma de principe de fonctionnement des méthodes de réduction de bruit par atténuation spectrale à court terme, d'après [20], p. 22.

3 Une nouvelle approche pour la compréhension de la parole dans le bruit

3.1 L'Intégrateur de Densité Spectrale (IDS)

3.1.1 Historique

L'Intégrateur de Densité Spectrale (IDS) est un procédé d'analyse audio développé par le Professeur Émile Leipp, qu'il a présenté dans un bulletin du Groupe d'Acoustique Musicale (GAM) en décembre 1977. Ses recherches étaient motivées par un souci de caractérisation de la « sonorité »¹⁶ des instruments de musique.

L'arrivée en France du sonographe en 1962 a facilité considérablement ses recherches car il fournissait des « diagrammes fréquence – temps – intensité » d'un évènement sonore en temps réel. Cependant, le sonographe ne fournissant des informations que dans leur « contexte immédiat », son utilisation empêchait d'analyser les informations sur une longue durée, alors que cette durée est essentielle à notre appréciation sensible :

« Quand nous écoutons une longue séquence de musique, quand nous comparons deux violons jouant la même pièce dans les mêmes conditions, quand nous diffusons successivement le même disque sur deux chaînes d'écoute, l'expérience montre qu'à la fin de l'écoute nous sommes capables de formuler un jugement sur la qualité sonore globale, éventuellement comparative, entre deux évènements acoustiques. [...] Pour saisir cette information esthétique, cette "coloration" statistique particulière d'un message acoustique, notre système auditif exploite nécessairement les mémoires à long terme et réalise certainement des opérations mentales particulières, relevant de mécanismes intégratifs qui ont lieu dans notre cerveau. » [21]

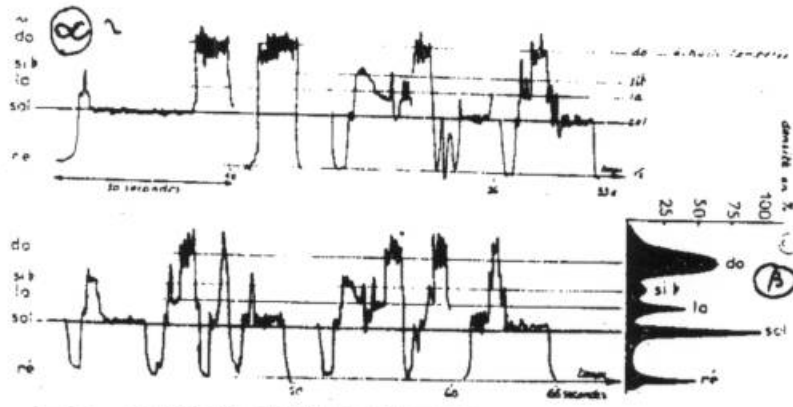
16. Dans son bulletin du GAM, Émile Leipp introduit la notion de sonorité comme synonyme à la qualité sonore.

La notion d'intégration, correspondant davantage à notre perception, devint dès lors essentielle. Leipp eut l'idée de relever la ligne mélodique d'une pièce musicale et de la reporter sur du papier millimétré (Figure 7). En additionnant sur chaque ligne horizontale les points d'intersection avec la ligne mélodique et en reportant le total sur un diagramme séparé, il obtint une courbe de « taux d'occurrence » de chaque fréquence. Cette courbe indiquant la répartition statistique de l'énergie au sein de la pièce musicale fut intitulée courbe de densité spectrale.

L'étape suivante consistait à définir les « bandes sensibles », à savoir le découpage du spectre des fréquences audibles¹⁷ en sous-bandes cohérentes avec les sensations auditives. Leipp a réalisé de nombreux tests d'écoute de pièces orchestrales qu'il filtrait en temps réel via un filtre réjecteur de bande, il élargissait progressivement la bande jusqu'à obtenir une réaction de l'auditoire remarquant l'absence de cette plage fréquentielle. Ses tests lui ont permis d'aboutir à un découpage du spectre audible en 8 bandes sensibles :

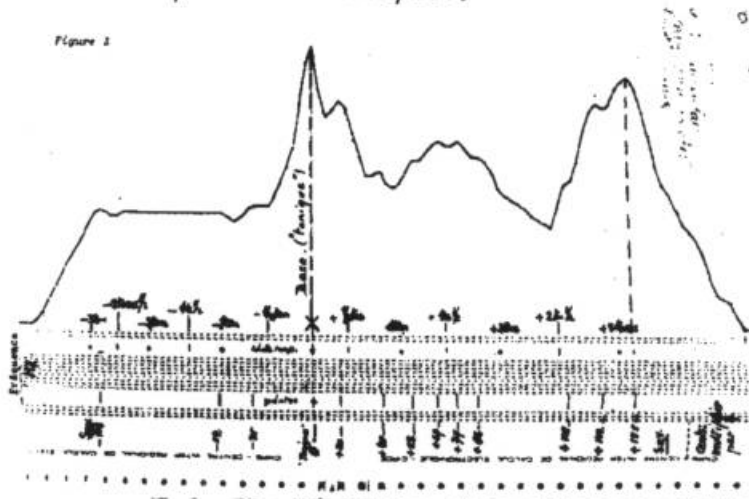
- 50 - 200 Hz : bande des basses ;
- 200 - 400 Hz : bande des graves ;
- 400 - 800 Hz : bande du médium grave ;
- 800 - 1200 Hz : bande du médium ;
- 1200 - 1800 Hz : bande du médium aigu ;
- 1800 - 3000 Hz : bande de l'aigu ;
- 3000 - 6000 Hz : bande du sur-aigu ;
- 6000 - 15000 Hz : bande de stridence.

17. Dans ce mémoire, nous considérons par « audibles » les fréquences qui sont audibles pour l'oreille humaine. Ce spectre fréquentiel est couramment défini par l'intervalle $[20 \text{ Hz}, 20000 \text{ Hz}]$, bien qu'il soit en réalité différent pour chaque être humain.



1a Densité spectrale et taux d'occurrence

La première idée que j'ai eue d'utiliser la technique de densité spectrale se rapportait à l'étude du taux d'occurrence des notes dans les musiques mélodiques. Ici une berceuse vietnamienne (mélodie relevée au sonographe). On additionne tous les points sur de très nombreuses lignes horizontales : les "notes" que la chanteuse a "encadrées" apparaissent avec évidence dans leur aspect statistique.



1b La même opération est faite à l'ordinateur (LIMSI-Mlouka) : ce diagramme correspond au diagramme (b) de la figure ci-dessus (en bas, à droite) : on voit émerger les "notes"

FIGURE 7 – Densité spectrale et taux d'occurrence, figures publiées par Émile Leipp dans le bulletin du GAM n°94. D'après [21], p. 3.

Les largeurs des sous-bandes n'étant pas régulières, on ne peut que constater que les résultats de son approche expérimentale démontrent une sensibilité fréquentielle de l'audition humaine loin d'être rationnelle :

« Cette anomalie de découpage fréquentiel n'est qu'apparente. En effet, comme nous partons de la réalité sonore d'un message conçu par l'homme pour le système auditif de l'homme, il ne peut y avoir aucune systématique dans les bandes sensibles pour la bonne raison que l'oreille, contrairement à ce que l'on pense et dit, n'est en réalité ni linéaire ni logarithmique lorsqu'il s'agit de messages informationnels. Par contre, tout le monde sait que l'oreille est beaucoup plus sensible dans la région de 1000 à 2000 Hz ; non seulement en fréquence, mais aussi en pouvoir séparateur temporel. C'est bien pourquoi les praticiens des messages sonores destinés à l'homme ont de tout temps exploité cette zone fréquentielle où il est possible de communiquer à l'homme la maximum d'information dans le minimum de temps. »

[21]

Une fois ce découpage défini, il fallait réemployer le filtre électronique paramétré cette fois en filtre « passe-bande » afin de laisser passer uniquement l'une des sous-bandes. En utilisant un intégrateur électronique on pouvait cumuler l'énergie contenue dans cette sous-bande durant la totalité de l'échantillon musical considéré. Le relevé des valeurs reporté sur un diagramme, avec pour abscisse les sous-bandes fréquentielles et pour ordonnée l'énergie, donnait alors le portrait IDS de l'échantillon.

Afin de rendre possible une comparaison des portraits IDS de différents contenus sonores, il ne fallait plus représenter l'énergie de chacune des sous-bandes fréquentielles de manière absolue mais comme un pourcentage de l'énergie totale. Ceci fait, le portrait IDS devenait un outil précieux pour pouvoir caractériser de nombreux événements sonores, quelles que soient leurs durées.

Bien que motivé au départ par l'étude de la qualité sonore des instruments de musique, Leipp employa l'outil qu'il avait développé pour de nombreuses autres applications, telles que la coloration de la voix humaine, la directionnalité d'une source sonore, l'acoustique des salles ou encore l'étude des bruits de fond et bruits de machine.

Pour préciser le fonctionnement de l'intégrateur de densité spectrale, on peut citer l'explication proposée par Titouan Ralle dans son mémoire de fin d'études :

« L'opération consistait à filtrer le signal analysé en suivant un jeu de filtres dont les fréquences de coupure correspondaient aux fréquences du découpage cité précédemment. Ces filtres étaient des filtres de Butterworth analogiques possédant une pente de -12 dB par octave.

Il s'agissait ensuite de calculer l'énergie cumulée des signaux de chaque sous-bande et l'énergie cumulée du signal total (signal étudié). On obtenait donc 9 valeurs d'énergie :

- 1 pour chaque sous-bande, appelons les « e_1 », « e_2 », . . . , « e_8 » ;
- 1 pour le signal étudié, appelons la « E » .

Enfin pour calculer l'énergie de toutes ces sous-bandes en pourcentage de l'énergie totale, il suffisait de faire le rapport e_n/E pour chacune d'entre elles. Par exemple e_1/E pour la bande des basses, e_2/E pour la bande des graves, e_3/E pour la bande du médium grave, etc. » [6], p.47

Depuis son invention par Émile Leipp, le principe de l'intégrateur de densité spectrale n'a été que très peu utilisé par d'autres chercheurs pour répondre aux problématiques d'analyse des signaux sonores. Laurent Millot l'a exhumé en en proposant une version étendue et numérique, profitant ainsi des nombreux avantages donnés par l'évolution de l'informatique.

3.1.2 Version numérique

C'est à partir des travaux de Leipp que Millot a développé un outil d'analyse des signaux reposant sur le principe de l'intégration de densité spectrale. Plusieurs modifications ont été apportées par rapport au principe original :

1. Le découpage fréquentiel a été étendu au nombre de 10 bandes sensibles, par ajout de deux bandes dans les extrémités du spectre (0 - 50 Hz pour les extrêmes graves fréquences et $15000 - F_e/2$ ¹⁸ pour les extrêmes aigus). Bien que ces plages fréquentielles soient très peu audibles à elles seules, elles ont certainement un impact sur la perception des transitoires et la spatialisation.

« Si a priori, on considère généralement que l'audition humaine concerne la bande des fréquences allant de 20 Hz à 20 kHz, il ne faut pas perdre de vue que cette information ne concerne que des tests effectués avec des tons purs (des sinusoïdes) mais pas des signaux complexes ou « réels » présentant une évolution temporelle s'opposant au caractère stationnaire des tons purs. Et, pour des signaux non stationnaires tels que de la parole ou de la musique, d'autres modalités perceptives sont mises en œuvre comme la perception des transitoires ou de la spatialisation des sons, même dans le cadre d'un système de diffusion stéréophonique. Or, les performances de ces autres modalités perceptives indiquent les limites de la perception auditives sont vraisemblablement supérieures à la gamme des fréquences 20-20 000 Hz. » [22], p. 107

De plus, ce découpage en sous-bandes devient un paramètre modifiable pour pouvoir s'adapter au corpus des signaux étudiés. Cette version de l'intégrateur de densité spectrale permet en effet de définir dans un premier temps le découpage employé (nombre de bandes et largeurs) avant de réaliser l'analyse de l'évènement sonore considéré.

18. F_e étant la fréquence d'échantillonnage, $F_e/2$ correspond à la fréquence maximale utilisable sans dégradation du signal suite à son échantillonnage, selon le théorème de Nyquist-Shannon.

2. Une fois le découpage fréquentiel déterminé et l'analyse IDS effectuée, on peut écouter chaque sous-bande séparément, pour évaluer subjectivement son rôle dans le signal étudié, ou une re-synthèse partielle à totale des sous-bandes, par exemple, pour déterminer la combinaison minimale pour avoir une approximation intéressante du signal original. C'est un avantage remarquable par rapport à de nombreuses analyses de signaux sonores qui ne donnent que des informations visuelles.
3. En plus de cette écoute des différentes sous-bandes possibles, on peut modifier le poids de chacune d'entre elles afin de modifier l'équilibre spectral global. De même qu'une console de mixage permet de modifier facilement le niveau des sources sonore entrantes, l'application de re-synthèse IDS développée par Millot permet de modifier le niveau sonore de chacune des sous-bandes fréquentielles, d'écouter instantanément le résultat, de muter l'une d'elles (mode *mute*) ou de n'en écouter qu'une seule (mode *solo*). Ces modifications conduisent à une re-synthèse partielle à totale du signal étudié.
4. Les portraits IDS de chacun des contenus sonores étudiés sont également disponibles pour une analyse graphique. L'énergie relative de chacune des sous-bandes n'est plus exprimée en pourcentage de l'énergie totale mais en décibels relatifs, afin de s'accorder davantage avec notre perception plus proche d'une échelle logarithmique et de faciliter les comparaisons entre deux portraits IDS par simple soustraction. De plus, un niveau moyen du signal étudié est calculé et affiché en dB FS.

Une fois de plus, nous nous appuyons sur les explications données par Titouan Ralle dans son mémoire de fin d'études de l'ENS Louis Lumière pour détailler le fonctionnement de l'analyse-resynthèse IDS dans sa version numérique :

« Les filtres utilisés pour l'analyse sont des filtres passe-bande RIF (Réponse Impulsionnelle Finie) à phase nulle, ce qui permet de s'affranchir de toute distorsion de phase mais implique un fonctionnement en temps différé. Ils comportent un nombre important de coefficients et proposent donc une atténuation rapide (minimum -80 dB FS en 5 Hz dans les bandes de transition) qui permet de limiter au maximum les recouvrements entre deux bandes consécutives.

Ces filtres sont calculés à partir de l'algorithme suivant.

- On part d'un filtre fréquentiel idéal, noté $\delta [n]$, qui possède un gain unitaire sur la gamme de fréquences $[0, F_N]$. On retranche à $\delta [n]$ la réponse impulsionnelle du filtre passe-bas à phase nulle dont la fréquence de coupure est égale à la limite inférieure de la dernière sous-bande (f_K pour un découpage utilisant K sous-bandes). Ce filtre passe-bas est noté h_{PB, f_K} . On obtient pour $h_K [n]$ le filtre de sous-bande K :

$$h_K [n] = \delta [n] - h_{PB, f_K} [n]$$

$h_k [n]$ correspond donc bien au filtre passe-bande ne contenant idéalement que la gamme de fréquences $[f_k, F_N]$ ¹⁹.

- Le même processus est itéré pour chacune des autres sous-bandes excepté la première. Ainsi pour la sous-bande $K - 1$ délimitée par la gamme de fréquences $[f_{K-1}, f_K]$, on calcule le filtre passe-bande :

$$h_{K-1} [n] = h_K [n] - h_{PB, f_{K-1}} [n].$$

- On a donc une dérivation itérative des filtres passe-bande. Pour le calcul du filtre de sous-bande $h_k [n]$ correspondant à la gamme de fréquences $[f_k, f_{k+1}]$, on obtient :

$$h_k [n] = h_{k+1} [n] - h_{PB, f_k} [n], \text{ excepté pour } k = 1.$$

- Pour la première sous-bande correspondant à la gamme de fréquences $[0, f_2]$, on a :

$$h_1 [n] = h_{PB, f_2} [n].$$

Comme les filtres utilisés sont complémentaires, il est possible de re-synthétiser le signal analysé par simple sommation de toutes les sous-bandes »²⁰. [6], p. 49

19. En reprenant le découpage de Leipp étendu cette gamme de fréquences correspond à la sous-bande $[15000 \text{ Hz}, F_N]$.

20. Du fait des contraintes imposées lors de la détermination des filtres (filtres RIF à phase nulle opérant en temps différé) cette re-synthèse donne une erreur théoriquement nulle, non audible en pratique.

3.1.3 Perspectives

L'intégrateur de densité spectrale tel qu'il a été étendu et « numérisé » par Millot élargit ses champs d'applications possibles. En plus de son intérêt dans l'analyse des signaux sonores, les étapes de modification de l'équilibre spectral et de re-synthèse développées en font un outil de correction de ces signaux. Cette correction fréquentielle n'est, dans la version actuelle de l'application de correction IDS, réalisable qu'en temps différé. Son utilisation en temps réel dépend de son optimisation algorithmique d'une part, et d'autre part de la puissance de calcul disponible en fonction des supports (ordinateurs, tablettes, smartphone).

Comme nous l'avons vu précédemment, la compréhension de la parole dans les environnements bruyants est une problématique d'autant plus actuelle que les réducteurs de bruits présents dans les prothèses auditives ne sont pas satisfaisants dans ces conditions extrêmes d'écoute. L'analyse et re-synthèse IDS permettant une correction du signal sonore qui soit adaptable par son utilisateur, elle devient dès lors une approche innovante pour constituer une solution à cette problématique.

3.2 Le travail effectué par Titouan Ralle

Dans son mémoire de fin d'études de l'*École Nationale Supérieure Louis Lumière*, soutenu en juin 2017, Titouan Ralle a travaillé sur l'usage de l'analyse IDS pour l'amélioration de la compréhension de la parole dans le bruit. L'ensemble de ses recherches et les résultats obtenus permettent désormais d'envisager l'analyse-resynthèse IDS comme un outil pertinent pour proposer une alternative aux réducteurs de bruits actuellement utilisés dans les prothèses auditives. Puisque nous nous inscrivons dans la suite des travaux effectués en 2017, nous en présentons les grands axes.

3.2.1 Constitution de corpus oraux représentatifs de la voix française

Comme nous l'avons évoqué précédemment, pour concevoir un réducteur de bruit adapté à la compréhension de la parole, il faut pouvoir déterminer des critères objectifs de séparation entre le signal utile (ici la parole) et le bruit. L'analyse IDS ne s'intéressant qu'à l'aspect fréquentiel de la réduction, le premier objectif était de déterminer les caractéristiques fréquentielles essentielles à la voix française. Cette étude se restreint donc, dans un premier temps, à une seule langue et sans prendre en compte la complexité de ses différents accents régionaux. Une fois le protocole mis en place pour la voix française, la généralisation aux accents, aux dialectes et aux autres langues pourrait être aisément mis en œuvre.

Les deux corpus réalisés par Ralle sont les suivants :

- un corpus de de voix parlées divisés en trois parties (liste de mots, lecture d'un texte, interview libre) ;

« Le premier corpus apparaît spontanément comme indispensable car chaque jour nous échangeons de manière vocale avec nos proches ou divers interlocuteurs. Cette parole réelle demeure encore aujourd'hui prédominante dans notre façon de communiquer. » [6], p. 28

- un corpus de voix télévisuelle.

« Le deuxième corpus a été constitué pour deux raisons. D'une part, pour observer s'il existe des différences notables dans les découpages fréquentiels obtenus à partir de voix provenant de sources sonores différentes. D'autre part la télévision est encore le média le plus consommé en France [23]. En outre, ce média est particulièrement utilisé chez les 50 ans et plus [23] qui sont nombreux à déclarer se sentir gênés au niveau de leur audition lorsqu'ils regardent la TV justement [24, p. 9]. Il paraît donc intéressant d'intégrer la parole télévisuelle à notre étude. » [6], p. 28

Les enregistrements ont été effectués avec deux microphones différents, l'un connu pour sa qualité audio lors de prise de son de voix (*Neumann TLM103*) et l'autre pour sa fiabilité comme outil de mesure (*Behringer ECM8000*). Le but de ces enregistrements doubles était d'évaluer l'influence du microphone sur l'analyse fréquentielle des signaux.

3.2.2 Découpage fréquentiel adapté à la voix française

Comme nous l'avons expliqué précédemment, l'intégrateur de densité spectrale permet d'analyser les signaux sonores selon leurs répartitions fréquentielles d'énergie par rapport à un découpage en sous-bandes (ou « bandes sensibles » dans les recherches de Leipp). L'un des avantages apporté par la version numérique de l'IDS est la possibilité d'adapter ce découpage en fonction de l'objet d'étude. Millot et Ralle ont donc cherché à déterminer quel était le découpage idéal pour analyser le corpus précédemment réalisé. Cette recherche s'appuie sur le calcul du spectre cumulé de corpus puis sur la détermination de minima locaux.

« L'idée pour la recherche d'un découpage fréquentiel adapté à un corpus donné consiste à calculer le spectre cumulé $S[f]$ pour chaque mesure puis le spectre cumulé d'amplitude $|S[f]|$ associé. [...] On fait ensuite la somme de tous ces spectres cumulés d'amplitude que l'on appelle spectres cumulés de corpus. Cette somme va faire émerger les creux spectraux qui sont pertinents à l'échelle du corpus et faire disparaître les petits creux qui ne sont valables que pour un nombre très limité de mesures (voire une seule mesure). C'est à partir de l'étude et la sélection des creux spectraux ayant résisté que l'on va pouvoir définir les limites des sous-bandes fréquentielles constituant le découpage fréquentiel : on va donc opérer une sélection dans les minima locaux du spectre cumulé de corpus. [...] Toutefois comme nous partons de spectres cumulés de corpus calculés pour 22050 fréquences, même en décidant de ne sélectionner que les minima locaux on aboutirait à un nombre trop conséquent de sous-bandes potentielles. Nous allons donc chercher à regrouper des sous-bandes potentielles consécutives afin de constituer des sous-bandes plus larges et ainsi simplifier la description spectrale des signaux étudiés. » [6], p. 56

Voici les découpages retenus pour les deux corpus :

Corpus de voix parlées :

- 0 - 40 Hz ;
- 40 - 83 Hz ;
- 83 - 229 Hz ;
- 229 - 625 Hz ;
- 625 - 867 Hz ;
- 867 - 2667 Hz ;
- 2667 - 7237 Hz ;
- 7237 - 16087 Hz ;
- 16087 - 22050 Hz.

Corpus de voix télévisuelles :

- 0 - 17 Hz ;
- 17 - 144 Hz ;
- 144 - 464 Hz ;
- 464 - 1189 Hz ;
- 1189 - 3240 Hz ;
- 3240 - 10038 Hz ;
- 10038 - 16565 Hz ;
- 16565 - 22050 Hz.

En plus de ces résultats, Ralle a observé que le dispositif d'enregistrement (le choix des microphones notamment) avait un impact non négligeable sur la détermination de ces découpages fréquentiels. Il conclut donc que le dispositif de captation doit se rapprocher au maximum de la situation réelle pour laquelle le découpage est réalisé. Dans le cadre d'une application de l'IDS aux prothèses auditives, il paraît indispensable d'avoir à réaliser des corpus représentatifs de la voix française qui soient enregistrés via la chaîne de captation propre aux prothèses.

3.2.3 Conception de l'IDS Speech Enhancer (IDSSE) et campagnes de tests perceptifs

Une fois les corpus vocaux réalisés et leurs découpages fréquentiels propres déterminés, Ralle a conçu l'*IDS Speech Enhancer* (IDSSE) en modifiant et en étendant les re-synthétiseurs IDS proposés par Millot. Cette application, fonctionnant sur le logiciel Pure Data, lui a permis de réaliser des tests perceptifs dans le but d'obtenir des filtres favorisant la compréhension de la parole dans un environnement bruyant.

L'IDSSE est un outil logiciel et matériel qui permet à un sujet d'écouter un enregistrement sonore mêlant une voix parlée et une ambiance sonore et d'en modifier en temps réel l'équilibre spectral (le poids relatif de chacune des sous-bandes fréquentielles) afin de déterminer le réglage favorisant la compréhension de ce qui est dit. Les voix parlées étaient issues de corpus préalablement réalisés, les ambiances sonores provenaient d'une sonothèque et correspondaient à des lieux communs et variés (café, gare, marché, ville, métro).

La problématique de réduction du bruit étant déjà pensée dans le contexte des prothèses auditives, ce dispositif cherchait à se rapprocher des conditions d'écoute des personnes appareillées. La diffusion des signaux se faisait via un casque audio pour se rapprocher d'une écoute de proximité telle que celle potentiellement proposée par le haut-parleur des prothèses. Le réglage de l'équilibre spectral se faisait via la manipulation d'une surface de contrôle MIDI²¹ proposant un fader linéaire par sous-bande.

Une campagne de tests sur 19 sujets a permis de tirer des conclusions :

- les deux sous-bandes les plus graves (0 - 40 Hz et 40 - 83 Hz) ne semblent pas contribuer à la compréhension de la parole ;
- une majorité des sujets agissent par soustraction dans la manipulation de l'IDSSE, ils préfèrent enlever du niveau qu'en rajouter ;
- les ambiances présentant un flux de parole important (café, marché) sont celles qui posent le plus de problèmes aux sujets, il semble qu'ils choisissent leur correction de manière aléatoire.

Mais l'échantillon de sujets étant trop petit pour obtenir des résultats plus complets, Ralle a proposé des améliorations possibles pour la suite de ses tests perceptifs :

- la diffusion pourrait être double : le casque audio pour le signal traité correspondant à celui diffusé par le haut-parleur d'une prothèse et une diffusion de l'enregistrement ori-

21. Le MIDI (Musical Instrument Digital Interface) est un protocole de communication permettant à un contrôleur d'interagir avec d'autres contrôleurs et des logiciels.

ginal via des enceintes, sans correction fréquentielle, correspondant à l'écoute naturelle qui persiste malgré le port d'une prothèse auditive ;

- réaliser un corpus d'ambiances sonores avec un protocole d'enregistrement uniforme ;
- modifier l'équilibre de niveau entre la voix et l'ambiance car la différence de -20 dB qui avait été définie au départ était peut-être trop faible pour les ambiances présentant un flux de parole important ;
- le corpus de voix parlée pourrait être amélioré en enregistrant soit la lecture d'un texte de la culture commune (qui soit donc connu du sujet, favorisant sûrement sa sensibilité à la compréhension de la parole dans le bruit), soit une histoire vécue racontée par la personne enregistrée (pour que son débit de parole soit plus fluide) ;
- guider les sujets dans la manipulation de la surface de contrôle en leur proposant de régler d'abord les niveaux des sous-bandes centrales avant de « préciser » la balance avec le réglage des niveaux des sous-bandes extrêmes ;
- modifier le découpage fréquentiel pour éliminer les sous-bandes extrêmes et permettre plus de précision dans la zone du « bas-aigu » ;

Voici la nouvelle proposition de découpage fréquentiel :

- 163 - 331 Hz ;
- 331 - 511 Hz ;
- 511 - 798 Hz ;
- 798 - 1048 Hz ;
- 1048 - 1558 Hz ;
- 1558 - 2208 Hz ;
- 2208 - 3388 Hz ;
- 3388 - 5205 Hz.

Enfin, une réflexion a été proposée concernant la possibilité d'adapter davantage les *stimuli* au sujet. L'enregistrement de voix et d'ambiances familières au sujet lui permettrait certainement de manipuler plus aisément l'IDSSE, puisqu'il les reconnaîtrait et pourrait donc évaluer plus précisément sa compréhension de la parole.

3.3 Conclusion

Comme nous l'avons vu dans ce chapitre, l'analyse IDS permet de réaliser un découpage fréquentiel d'un signal audio selon un nombre de bandes et des largeurs de bandes réglables en fonction de l'objet d'étude. Une fois ce filtrage effectué, la modification du niveau relatif de chaque bande permet une correction globale de la balance spectrale du signal. La re-synthèse IDS permet quant à elle d'écouter en temps réel le résultat des manipulations de la balance spectrale, et, dans sa version actuelle, de convoluer en temps différé le signal original avec le filtre réunissant les corrections souhaitées.

Puisque cette méthode de traitement fréquentiel du signal se base sur un découpage adaptable et sur une modification à la volée des corrections, elle devient une solution de réduction de bruit qui puisse répondre à l'insatisfaction des utilisateurs de prothèses auditives face à certaines situations d'écoute difficiles.

C'est en nous appuyant sur les travaux de Leipp, Millot et Ralle que nous souhaitons explorer les perspectives de l'usage de l'analyse-resynthèse IDS pour l'amélioration de la compréhension de la parole dans le bruit pour les malentendants.

4 Vers un prototype de prothèse auditive intégrant l'analyse-resynthèse IDS

4.1 Introduction

Dans ce mémoire, nous souhaitons confronter les recherches existantes concernant l'utilisation de l'analyse-resynthèse IDS pour l'amélioration de la compréhension de la parole dans le bruit à la réalité technologique et médicale des prothèses auditives. Dans cette partie, nous exposerons la modélisation et la conception partielle d'un prototype de prothèse intégrant l'analyse-resynthèse IDS comme solution de traitement fréquentiel du signal.

Comme nous l'avons vu précédemment, une prothèse auditive peut être schématisée par la captation de l'environnement sonore au moyen de microphones directifs, le traitement en temps réel du signal capté puis la diffusion du signal traité dans le canal auditif externe. Or, pour pouvoir concevoir et tester une alternative au niveau de l'étape du traitement du signal, il faudrait pouvoir conserver les conditions d'utilisation effectives lors des étapes de captation et de diffusion. Autrement dit, le modèle schématique d'un prototype de prothèse intégrant l'analyse-resynthèse IDS pourrait être décomposé en trois étapes fonctionnelles :

1. prélèvement du signal capté par un système microphonique identique à ceux présents dans les prothèses auditives ;
2. redirection du signal vers un processeur (FPGA²², station de travail informatique, tablette, smartphone) appliquant en temps réel les algorithmes de traitement du signal développés ;
3. réinjection du signal traité dans une chaîne de diffusion identique à celles présentes dans les prothèses auditives.

²². Le FPGA (*field-programmable gate array*) est l'un des types de circuits logiques programmables les plus courants.

Nous allons présenter par la suite les recherches mises en œuvre pour chacune de ces trois étapes de modélisation du prototype, dans l'ordre chronologique de notre démarche. Nous aborderons ensuite la problématique de l'interface utilisateur d'un tel prototype, puis nous conclurons avec les modifications apportées aux objectifs du mémoire suite aux difficultés rencontrées.

4.2 Dispositif de captation

L'une des conclusions principales de Ralle à l'issue de ses travaux était de constater que le dispositif de captation qu'il utilisait pour réaliser des corpus représentatifs de la voix parlée avait une influence non négligeable sur les caractéristiques fréquentielles des enregistrements et, par conséquent, sur les corrections apportées par les sujets lors de la manipulation de l'*IDS Speech Enhancer*. C'est en partant de ce constat que nous souhaitons concevoir un dispositif de captation qui soit, sinon identique, du moins aussi proche que possible de ceux présents dans les prothèses auditives.

4.2.1 Collecte de prothèses inutilisées et rétro-ingénierie

Tout d'abord, une phase de recherches bibliographiques et sitographiques a été menée afin de récolter des documentations précises sur le fonctionnement des différents modèles de prothèses auditives. Mais ces recherches ont été très peu fructueuses puisque la communication des différents fabricants est très opaque concernant le détail de leurs équipements. Ce manque de transparence est une problématique globale dans le domaine des prothèses auditives, et, certains audioprothésistes se plaignent d'un manque d'information concernant des outils qu'ils doivent eux-mêmes maîtriser suffisamment pour pouvoir les choisir, les conseiller aux malentendants, les régler. A titre d'exemple, nous pouvons citer ce commentaire fait par trois audioprothésistes en conclusion d'une étude comparative de la performance des différentes méthodes de réduction de bruit :

« L'audioprothésiste reste malheureusement à la merci d'une communication peu transparente des industriels qui au-delà de ce constat frustrant, nous proposent pour certains des innovations qui méritent le respect. » [25], p. 14

Ce manque d'informations a été un frein dans nos recherches pour pouvoir modéliser un système de captation proche de ceux conçus par les fabricants de prothèses²³. Étant donné les difficultés rencontrées, nous avons collecté des modèles de prothèses inutilisées auprès de malentendants et d'audioprothésistes pour nous inscrire dans une démarche de *rétro-ingénierie*²⁴. Six modèles de prothèse auditive ont été collectés, tous de marques différentes. Nous n'avons pas pu récupérer une paire de prothèses, ce qui nous a obligé par la suite à limiter notre démarche au cadre monaural. Nous estimons que cette démarche reste valable malgré cette restriction, et que les études menées concernant l'amélioration de la compréhension de la parole dans le bruit au moyen de l'analyse-resynthèse IDS en s'appuyant sur une seule prothèse devraient pouvoir être étendues au cas d'un appareillage binaural.

Nous avons démonté ces prothèses pour accéder aux composants et circuits électroniques, tout en veillant à ne pas les endommager pour pouvoir continuer à les utiliser, dans le cadre de nos recherches ou par la suite lorsque nous n'en aurons plus besoin. Ce démontage poursuivait trois objectifs :

- mieux comprendre le fonctionnement de ces prothèses en observant leurs composants et circuits intégrés ;
- comparer les modèles entre eux et esquisser des hypothèses quant au fonctionnement général des prothèses auditives ;
- évaluer la faisabilité d'une intervention électronique sur les composants et circuits imprimés pour pouvoir utiliser certaines prothèses comme base du prototype.

Pour faciliter cette recherche nous avons contacté certains des fabricants pour leur demander directement des documentations techniques concernant les modèles en question. L'un d'entre eux (*Unitron*) nous a communiqué le schéma électronique du modèle *Conversa.NT*. Ce docu-

23. Ce n'est que tardivement que nous avons trouvé quelques informations concernant le type de microphones utilisés (MEMS, *Micro-Electro-Mechanical System*) dans les prothèses les plus récentes, et l'un des fabricants de ces microphones (*Knowles*). [26]

24. La rétro-ingénierie, ou ingénierie inverse, consiste à étudier un objet pour en déterminer le fonctionnement interne ou la méthode de fabrication.

ment nous a été très précieux, bien qu'il laisse une grande part d'ombre sur le fonctionnement de la prothèse, puisque le processeur intitulé « Hybrid » reste une véritable boîte noire.

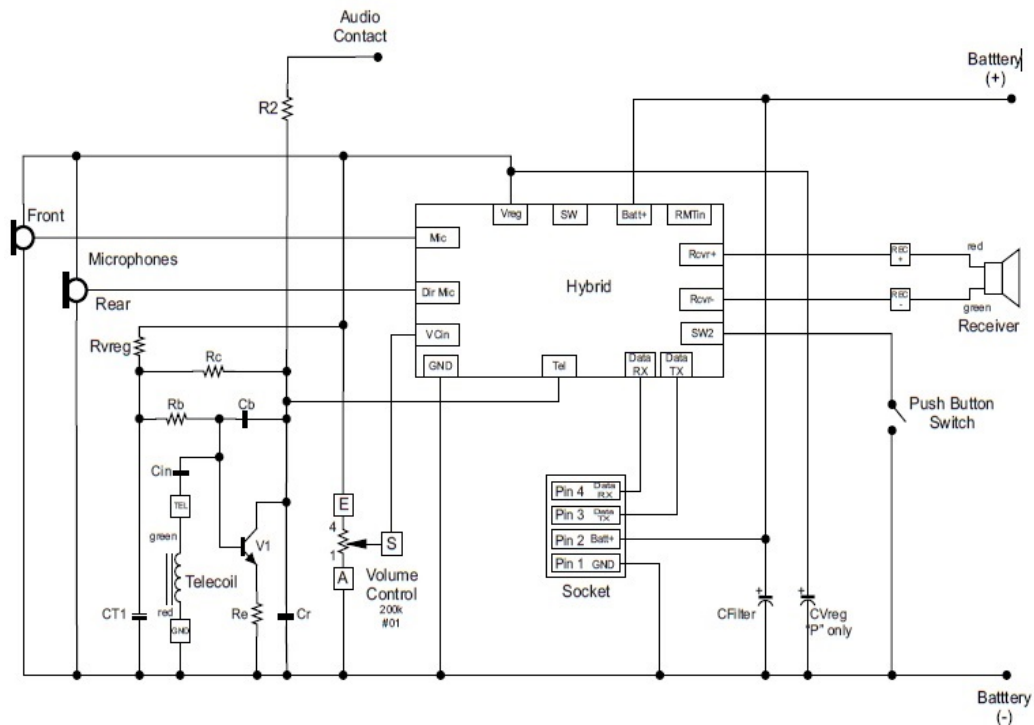


FIGURE 8 – Schéma électronique du modèle Conversa.NT [document original disponible en Annexe 1].

En comparant le schéma de la figure 7 et les prothèses démontées, nous pouvons repérer sur chaque modèle les différents composants réalisant les différentes étapes fonctionnelles de captation, traitement, alimentation et diffusion. Tous les modèles sont visiblement constitués de composants et d'une architecture matérielle similaires. Plusieurs observations peuvent être faites.

- Tous les modèles observés possèdent deux microphones pour la captation du signal. Le premier est situé à l'extrémité haute de la prothèse lorsqu'elle est placée sur l'oreille, il correspond à celui intitulé « *Front* » sur le schéma électronique de la figure 7 et doit donc être celui servant de microphone « avant » dans le dispositif de directivité adaptative décrite en partie 2.2.3. L'autre microphone est espacé d'environ 1 cm du premier et est situé plus bas dans le prolongement de la prothèse lorsqu'elle est portée, il correspondrait au micro arrière (intitulé « *Rear* » sur le schéma).

- Tous les modèles sont du type BTE (*Behind the ear*), avec l'écouteur (*Receiver*²⁵) présent dans le boîtier qui est porté à l'arrière de l'oreille, les vibrations acoustiques qu'il produit étant transmises via un guide d'onde jusqu'au canal auditif.
- Certains modèles (comme en atteste le schéma au niveau du « socket » pour le modèle *Conversa.NT*) ont un connecteur qui permet de paramétrer le processeur. En consultant des audioprothésistes, nous nous sommes assurés que ces connecteurs ne permettaient pas d'avoir accès au signal audio entrant ou sortant du processeur (si cela avait été le cas cela aurait grandement facilité la suite de notre démarche).
- La faisabilité d'une intervention électronique est variable en fonction des modèles. En effet, les modèles sont de tailles différentes en fonction de la miniaturisation plus ou moins poussée des composants et des circuits imprimés. De cette miniaturisation dépend l'accessibilité aux composants, au circuit imprimé et aux fils électriques les reliant. Dans certains cas les composants sont directement soudés sur un circuit imprimé flexible (*flex PCB*), ne permettant pas d'intervention électronique. Dans certains cas les composants et le circuit imprimé ont été recouverts de traitements chimiques, sûrement pour les protéger du milieu extérieur (poussière, humidité), ce qui empêche également d'accéder aux soudures.
- Dans tous les cas, les composants, fils électriques et soudures observées sont d'une taille très réduite, ce qui implique des interventions non sans risques pour le fonctionnement de la prothèse (risque de court-circuit notamment).

En raison des deux dernières observations, seuls deux des six modèles observés ont été retenus pour réaliser des tentatives d'intervention électronique. Le fait de retenir deux modèles présente deux avantages :

- en intervenant sur deux modèles nous nous laissons plus de chance que l'une des interventions fonctionne et que l'on puisse poursuivre notre démarche ;
- si les deux interventions réussissent nous pouvons envisager un prototype binaural, avec

25. Dans ce mémoire, nous traduisons le terme *Receiver* par « écouteur ».

deux prothèses certes de marques différentes mais permettant malgré tout une sensation d'utilisation proche d'un appareillage binaural.

4.2.2 Tentatives d'interventions électroniques

Ces interventions électroniques sur les prothèses collectées ont pour objectif de répondre à la première étape fonctionnelle de notre prototype, à savoir le prélèvement du signal capté par les microphones. Comme nous l'avons vu en théorie en partie 2.2.3 et observé pour ces deux modèles, la captation de l'environnement sonore est assurée par les deux microphones. Mais les signaux qu'ils fournissent sont ensuite traités par le processeur afin de déterminer la directivité la plus adaptée à la situation. Or, nous ne savons pas comment fonctionne le processeur et n'avons pas accès au signal résultant de cette étape d'adaptation de la directivité (puisque'il est interne au processeur, sous forme numérique). Nous avons donc du faire le choix de poursuivre nos recherches à partir du signal provenant d'un seul microphone, et sans prendre en compte l'apport (pourtant performant selon les études citées précédemment) de cette directivité adaptative. Cela dit, nous estimons que cette démarche reste valable à deux conditions :

- nous faisons le choix de nous intéresser au signal fourni par le microphone avant et uniquement dans une situation d'un locuteur parlant face à la personne utilisant le prototype ;
- notre étude ne s'intéresse qu'au caractère fréquentiel de la réduction de bruit, et, dans ce cas, le fait que le signal à traiter n'ait pas bénéficié d'une réduction de bruit grâce à la discrimination spatiale ne rend que plus pertinent la recherche d'une potentielle discrimination fréquentielle.

L'objectif suivant consiste donc à récupérer le signal provenant du microphone « avant » présent dans les deux modèles de prothèses retenus. L'étude du schéma de la figure 7 et l'observation des prothèses amènent à proposer plusieurs constats :

- chaque microphone est relié à trois points du circuit imprimé : l'un correspond à l'alimentation (elle-même fournie au processeur par la pile), l'autre à la masse, et le troisième au point « milieu » correspondant au signal de sortie du microphone ;
- puisque chaque microphone est alimenté par le reste du circuit et que nous ne connaissons pas les caractéristiques de cette alimentation, nous ne devons pas endommager le reste du circuit pour être certains que chaque microphone fonctionne ;
- pour pouvoir récupérer l'information sonore captée par le microphone il faut prélever le signal entre le point milieu et la masse ;
- il se peut que le signal prélevé au point milieu présente une composante continue qu'il faudra filtrer pour ne conserver que le signal alternatif.

Nous mesurons au moyen d'un voltmètre les tensions au niveau des trois fils qui relient le microphone au processeur pour déterminer à quels potentiels ils correspondent. Celui qui présente une variation de tension lorsqu'on génère un stimulus sonore proche du microphone devrait correspondre au point « milieu », celui qui nous intéresse. De cette étude découle l'hypothèse suivante, probablement confirmée par l'observation des autres liaisons électroniques entre les composants (en espérant qu'un code couleur cohérent soit respecté) :

- fil vert : masse ;
- fil rouge : potentiel positif de l'alimentation ;
- fil jaune : point milieu, signal de sortie du microphone.

A partir de cette hypothèse nous pouvons démarrer des tentatives de soudures pour se « greffer » sur les points de connexion des fils verts et jaunes (connexion en « Y »), ceci pour chacune des deux prothèses. Idéalement nous devrions intervenir en priorité du côté des microphones pour ne pas toucher au circuit imprimé, mais dans la pratique nous nous adaptons à l'accessibilité des points de soudure. Étant donné la complexité de certaines des soudures à réaliser, nous avons sollicité l'aide d'Antonio Duarte, formateur en électronique au *CRP Jacques Arnaud* de Bouffémont-Moiselles. Le reste des interventions a été effectué au laboratoire d'électronique de l'*ENS Louis Lumière*, avec l'aide de Mohammed Eллиq et de Taïeb Keraoun.

Nous avons rencontré deux difficultés majeures pour réaliser ces soudures :

1. les points de connexion étant de taille réduite et très rapprochés entre eux, la précision de la soudure doit être maximale pour éviter de créer un court-circuit ;
2. les soudures doivent être suffisamment solides pour pouvoir manipuler les fils que nous connectons au dispositif.

Malgré ces difficultés les soudures ont pu être réalisées, permettant la poursuite de la démarche.

4.2.3 Finalisation du dispositif de captation

Nous avons dans un premier temps testé le fonctionnement des interventions effectuées en reliant les fils électriques rajoutés aux prothèses à un étage d'amplification du signal, puis à une table de mixage audio pour pouvoir écouter le signal résultant au casque tout en générant un stimulus sonore à proximité des prothèses. Dans le cas où ce signal ne semble pas exploitable à l'écoute, nous pouvons passer par une étude plus complète de ses caractéristiques grâce à un oscilloscope.

Pour l'un des deux modèles étudiés, cette étude approfondie n'a pas été nécessaire puisque l'écoute au casque a permis de considérer que le signal récupéré était directement exploitable et potentiellement équivalent à une captation microphonique. Toujours en ne se basant que sur l'écoute subjective du signal, nous n'avons constaté la présence d'aucun parasite, d'aucune distorsion remarquable. Il est certain que nous n'avons aucune garantie concernant l'impact des interventions électroniques effectuées sur la qualité du signal et que des perturbations ont pu être induites (bruit de fond, distorsion, boucle de masse). Mais, puisque nous n'avons pas de moyen de comparaison avec le signal fourni à l'origine par le système de captation, nous ne pouvons que faire l'hypothèse que ces perturbations potentielles sont négligeables.

Pour l'autre modèle étudié, l'écoute au casque semble indiquer au contraire que les soudures effectuées ne permettent pas de récupérer le signal audio à la sortie du microphone. Nous n'obtenons qu'un bruit électronique aléatoire.

Nous aurions pu effectuer une étude plus poussée pour essayer de comprendre les raisons de ce résultat, mais nous avons préféré poursuivre notre démarche en ne conservant que le modèle qui a fonctionné directement, ceci pour plusieurs raisons.

- Le temps consacré à la collecte des modèles de prothèses inutilisées et aux interventions électroniques a été plus long que ce nous avons prévu, nous avons donc pensé qu'il était préférable de passer aux étapes suivantes de la conception du prototype.
- Même si nous avons pu récupérer le signal fourni par les microphones au sein des deux modèles manipulés, le fait qu'ils soient de marques différentes nous empêche en pratique de nous placer dans un cadre d'un appareillage binaural réel (avec deux modèles identiques).
- Le fait que l'un des modèles ait permis d'aboutir à la réalisation d'un dispositif de captation monophonique quasi identique à celui présent dans une prothèse auditive est une étape importante dans ce mémoire de fin d'études. Il est certain qu'un prototype binaural intégrant une captation stéréophonique aurait été préférable, mais l'ensemble des recherches effectuées sur une seule prothèse devraient pouvoir être étendues au cas d'un appareillage double.

Pour finaliser la conception de ce dispositif de captation, nous avons ajouté une fiche XLR mâle comme connectique de sortie pour correspondre au standard microphonique de prise de son professionnelle. De plus, nous avons autant que possible refermé le boîtier de la prothèse (les fils électriques rajoutés empêchent sa fermeture complète) pour permettre son utilisation dans des conditions classiques d'appareillage (en plaçant le boîtier derrière le pavillon, le conduit auditif dans le canal auditif externe, le câble et la connectique XLR sont suspendus au boîtier). Nous avons fait le choix d'une connectique XLR pour faciliter l'insertion de notre dispositif dans une chaîne de captation audio classique, mais nous pourrions la remplacer au besoin par une connectique Jack 3.5 mm, plus légère, par soucis de confort d'utilisation.

Ainsi, nous avons conçu un dispositif qui, lorsqu'on le porte à la manière d'une prothèse auditive et que l'on relie sa connectique externe à une station de travail audio, nous permet de récupérer le signal fourni par le microphone « avant » intégré à la prothèse *Conversa.NT* de la marque *Unitron*. Nous retiendrons que ce dispositif de captation n'est comparable à ceux présents dans les prothèses auditives que dans le cas d'un interlocuteur qui s'exprime face à l'utilisateur, puisque nous n'intégrons pas de directivité adaptative.

4.3 Dispositif de diffusion

A l'image des travaux effectués pour l'étape de captation, nous souhaitons mettre en œuvre un dispositif de diffusion qui soit comparable à ceux présents dans les prothèses auditives. Ainsi, nous espérons tester la pertinence d'une réduction de bruit via analyse-resynthèse IDS auprès d'un public malentendant et dans des conditions classiques d'appareillage. Nous présentons par la suite les réflexions menées quant à la modélisation de ce dispositif, puis les restrictions liées à la certification indispensable de ces recherches, certifications impossibles à obtenir dans le temps imparti pour le mémoire.

4.3.1 Modélisation

Pour rappel, le prototype permettant de confronter l'analyse-resynthèse IDS aux conditions technologiques et médicales des prothèses auditives doit présenter un dispositif de diffusion du signal après traitement qui soit comparable, voire identique, à ceux présents dans les modèles commercialisés actuellement. Autrement dit, nous voulons que l'écouteur présent dans le boîtier de la prothèse que nous manipulons, et dont les vibrations acoustiques produites sont transmises par un guide d'onde jusqu'au conduit auditif externe, puisse recevoir et transmettre un signal audio envoyé depuis l'extérieur de la prothèse.

De manière analogue à ce qui a été fait pour le dispositif de captation, nous avons pu observer sur les modèles collectés où se situent l'écouteur et évaluer la faisabilité d'une intervention électronique sur ses liaisons d'entrée. Sur certains modèles les deux fils reliant la sortie du processeur à l'entrée de l'écouteur sont visibles et nous pourrions envisager des soudures.

En effet, nous pouvons supposer qu'en analysant les caractéristiques de l'étage de sortie du dispositif actuel (impédance, puissance électrique), nous pourrions remplacer la liaison par une autre permettant d'envoyer un signal audio dans les mêmes conditions.

Ces hypothèses n'ont pu être confrontées à des expérimentations puisqu'elles impliquent des manipulations sur un composant de la prothèse, l'écouteur, qui présentent un risque non nul pour la santé des personnes qui seraient amenées à tester notre dispositif.

4.3.2 Restrictions pour la conception

Puisque notre démarche de conception d'un prototype de prothèse implique une manipulation électronique de l'écouteur, élément qui agit indirectement sur le tympan, nos recherches s'inscrivent dans la première catégorie de la loi n°2012-300 du 5 mars 2012, celle du 1° de l'article L. 1121-1 du code de la santé publique : « *Ce sont les recherches qui comportent une intervention sur les personnes non dénuée de risques pour celles-ci. [...] Elles ne peuvent être mises en œuvre qu'après autorisation de l'ANSM²⁶ et avis favorable d'un CPP²⁷ (article L. 1121-4 du CSP).* » [27]

C'est par le Docteur Auriol que nous avons appris, dans le courant du mois de mars, que l'on devait obtenir une certification pour les prothèses modifiées. Or, le délai indiqué ne serait-ce que pour la constitution d'un CPP une fois le dossier de certification déposé est de l'ordre du mois, temporalité impossible à concilier avec celle du mémoire. Dans ces conditions, nous avons décidé d'interrompre les recherches sur la réalisation du dispositif de diffusion par prothèse et de chercher une solution alternative ne nécessitant pas, *a priori*, de certification. Il faudra donc envisager une collaboration avec des audioprothésistes et, vraisemblablement, au moins un industriel avant de se lancer dans le processus de certification.

26. Agence Nationale de Sécurité du Médicament

27. Comité de Protection de la Personne

Avant d'expliquer les modifications apportées aux objectifs de ce mémoire de fin d'études suite aux difficultés rencontrées dans la conception du prototype, nous présentons les réflexions menées pour les deux autres étapes fonctionnelles de ce prototype, le traitement du signal en temps réel et l'interface utilisateur.

4.4 Traitement du signal en temps réel

Si nous avions pu mettre en œuvre un dispositif de captation qui soit aussi fonctionnel que celui de captation, nous aurions alors pu confronter l'analyse-resynthèse IDS aux conditions réelles de son intégration dans une prothèse auditive. Comme nous l'avons vu, la contrainte principale de cette utilisation est un traitement en temps réel (2.2.1), ce qui n'est pas encore le cas de cet outil (3.1.3). En nous inspirant des pistes de réflexions données par Ralle et en nous appuyant sur les recherches les plus récentes de Millot concernant les applications qu'il a développées, nous présentons dans cette partie les évolutions des algorithmes à mettre en œuvre pour pouvoir atteindre des latences de traitement les plus faibles possibles. Nous abordons cette réflexion en deux temps, l'analyse IDS et la re-synthèse IDS, les enjeux n'étant pas les mêmes.

4.4.1 Analyse IDS

Si nous reprenons l'explication de Ralle concernant le fonctionnement de l'analyse IDS (3.1.2), nous voyons qu'elle est basée sur un filtrage successif du signal pour le décomposer en sous-bandes fréquentielles. Ce filtrage s'effectue dans la version actuelle au moyen de filtres « passe-bande » RIF (Réponse Impulsionnelle Finie) à phase nulle, ayant une atténuation très rapide (-80 dB FS en 5 Hz). Deux avantages majeurs justifient leur usage :

- les filtres RIF à phase nulle n'induisent pas de distorsions de phase ;
- une atténuation rapide (ou plus graphiquement, une pente raide) permet de limiter au maximum les recouvrements entre deux sous-bandes successives.

On obtient donc avec ce filtrage un découpage du spectre fréquentiel en sous-bandes tout en conservant une qualité audio optimale et une séparation précise de chacune des bandes, mais cela implique des temps de calcul longs. Pour pouvoir alléger ces temps de calcul, la solution envisagée consiste à réduire le nombre de coefficients des filtres. Puisque ce nombre est directement relié à la raideur de leur pente d'atténuation, il faudrait estimer le bon compromis entre la latence induite et la gêne occasionnée éventuellement par le recoupement entre deux bandes successives. La démarche que nous pensons la plus efficace consisterait à réduire le nombre de coefficients afin d'avoir des temps de calculs minimaux, puis, tester ensuite l'impact du recoupement du point de vue de l'écoute. Pour obtenir une analyse en temps réel, il faudrait que la latence totale, associée au filtrage de toutes les sous-bandes, soit inférieure ou égale à 20 ms, durée que l'on considère comme inaudible pour une oreille humaine.

De plus, les temps de traitement dépendent de la puissance de calcul disponible pour effectuer cette tâche. Or, les algorithmes développés pour l'instant fonctionnent sur des ordinateurs portables ayant une puissance de calcul certainement plus grande que celle des processeurs présents dans les prothèses auditives actuelles. Il faudrait donc pouvoir évaluer la puissance de calcul de ces processeurs pour pouvoir effectuer les tests évoqués précédemment dans des conditions technologiques cohérentes.

Puisqu'il faudra réussir à traiter des flux stéréophoniques en temps réel, avec de l'ordre de 8 filtrages pour traiter toutes les sous-bandes, il faudra étudier si une architecture embarquée permet de réaliser ces traitements à l'aide des convolutions temporelles par l'algorithme des tâches ou s'il convient de mettre en œuvre ces traitements en utilisant des transformées de Fourier rapides (FFT) qui permettent de traiter des convolutions de signaux stéréophoniques si l'algorithme de FFT utilisé est basé sur une utilisation des nombres complexes.

4.4.2 Re-synthèse IDS

Une fois l'analyse IDS effectuée, les manipulations de l'équilibre des sous-bandes fréquentielles n'induisent *a priori* pas de latences supplémentaires. En revanche, l'application, idéalement en temps réel, de la correction définie aux signaux captés par les prothèses implique à nouveau la mise en œuvre de calculs lourds.

En effet, il faut convoluer les signaux captés par les prothèses par le filtre de correction, correspondant à la somme pondérée par les poids de correction des filtres de sous-bandes. L'étude de l'utilisation de l'algorithme des tâches est en cours pour étudier si elle pourrait être réalisée en temps réel avec une architecture embarquée. Mais, s'il s'avérait que cette mise en œuvre n'est pas possible, pour le moment, il faudrait opter pour une utilisation des FFT en utilisant idéalement *BruteFIR*²⁸ ou *FTW*²⁹, ou plus vraisemblablement, une implémentation analogue adaptée à l'architecture embarquée choisie.

Comme ces optimisations algorithmiques nécessitent des connaissances approfondies en programmation informatique appliquée à l'audio, en langage *C* et un travail dans une optique temps réel, il serait vraisemblablement nécessaire de faire appel à des soutiens externes. Aussi, nous avons préféré laisser ces problématiques de traitement du signal à l'état de réflexions générales, pour plusieurs raisons :

- ne pouvant pas mettre en œuvre le dispositif de diffusion dans le temps imparti pour ce mémoire, la conception du prototype de prothèse n'est plus envisageable dans sa totalité et la nécessité de résoudre les problèmes de traitement du signal en temps réel est moins importante à court terme ;
- étant donné le calendrier très restreint et le manque de contacts auprès d'informaticiens spécialisés, poursuivre cette recherche ne nous a pas semblé susceptible de donner des résultats dans le cadre de ce mémoire de fin d'études ;
- il nous a semblé plus pertinent et enrichissant de consacrer le temps restant à des axes de recherches dans lesquels nous sommes davantage autonomes.

28. *BruteFIR* est un outil logiciel *open source* permettant de réaliser du filtrage RIF, potentiellement en temps réel, qui s'appuie sur la librairie *FTW* [28].

29. La *Fastest Fourier Transform in the West*, ou *FTW* est une librairie en C permettant de réaliser des *Discrete Fourier Transform* avec des latences très faibles. Elle a été développée au *Massachusetts Institute of Technology (MIT)* par Matteo Frigo et Steven G. Johnson [29].

4.5 L'interface utilisateur

4.5.1 Les raisons d'une interface

La plupart des prothèses auditives actuelles fonctionnent sur une réduction de bruit figée ou automatisée. Autrement dit, une partie des corrections fréquentielles faites par la prothèse est réglée par l'audioprothésiste (pour l'adapter à l'audition du malentendant) et reste ensuite fixe jusqu'à un nouveau réglage de sa part, et l'autre partie est potentiellement gérée de manière totalement automatique en fonction des situations rencontrées par le malentendant. A part pour les modèles les plus récents (qui proposent une communication entre une application sur smartphone et les prothèses), l'utilisateur ne peut pas interagir avec la prothèse pour optimiser son réglage. Nous faisons l'hypothèse que les corrections automatiques sont l'une des raisons de l'insatisfaction des malentendants, puisqu'elles ne prennent pas en compte la particularité des déficiences auditives de chaque individu, ni de leur évolution au cours du temps. Or, c'est ce que pourrait permettre un prototype de prothèse utilisant l'analyse-resynthèse IDS.

La question est donc de savoir de quelle manière l'utilisateur du prototype pourrait agir sur la réduction de bruit pour améliorer sa compréhension de la parole, et ce que cela implique en terme de traitement du signal. Nous présentons trois approches, la troisième étant celle qui nous paraît constituer le meilleur compromis entre simplicité d'utilisation et faisabilité technologique.

4.5.2 Version détaillée

La première version que nous avons envisagée pour cette interface permet à l'utilisateur de régler précisément la correction fréquentielle qui est appliquée, selon plusieurs moyens :

- accès aux sous-bandes fréquentielles fournies par l'analyse IDS et écoute en mode *solo* possible pour déterminer l'impact de chaque bande sur la compréhension de la parole ;
- possibilité de modifier la contribution relative de chacune des sous-bandes pour pouvoir établir la *balance spectrale* optimale ;

- sauvegarde de différents *presets* (réglages favoris de la balance spectrale en fonction des lieux fréquentés, des personnes rencontrées, de l'état de fatigue) ;
- rappel des *presets*, pour retrouver directement un réglage satisfaisant dans une situation type.

L'ensemble de ces manipulations pourrait se faire à partir d'une application sur smartphone ou tablette tactile qui communiquerait avec les prothèses. Ainsi, l'utilisateur pourrait à tout moment ouvrir cette application, régler la balance spectrale issue de l'analyse IDS, écouter en direct le résultat de ses manipulations et enregistrer le réglage s'il souhaite pouvoir le réutiliser dans une situation d'écoute similaire.

Cette version présente comme avantage majeur sa simplicité technologique. En effet, contrairement aux autres versions envisagées, celle-ci nécessite une réponse instantanée aux manipulations de l'utilisateur³⁰, mais ne fait pas appel à une logique d'apprentissage automatique.

En revanche, une telle interface serait certainement très peu satisfaisante en raison de la complexité de son utilisation. Nous avons donc par la suite envisagé des interfaces qui soient plus simples du point de vue de l'utilisateur.

4.5.3 Version simplifiée

Dans une logique totalement opposée à celle décrite précédemment, nous pouvons imaginer une interface qui soit la plus simple d'utilisation possible, tout en conservant une interaction entre le malentendant et la prothèse pour permettre une amélioration de ses réglages. L'une des suggestions faites par le Docteur Bernard Auriol, Directeur Externe de ce mémoire, serait de ne laisser à la disposition de l'utilisateur que deux boutons, fonctionnant selon deux logiques possibles :

30. Plus précisément, un tel prototype pourrait certainement supporter une latence de mise en application des corrections fréquentielles voulues qui soit audible, à condition que l'information audio originale soit, elle, bien transmise en temps réel.

- évaluation absolue de la qualité du réglage : un bouton indique que le réglage proposé par la prothèse est « bon » selon l'utilisateur, et l'autre indique qu'il n'est « pas bon » ;
- évaluation comparative de la qualité du réglage : un bouton indique que le réglage proposé par la prothèse est « mieux » que le précédent, l'autre qu'il est « moins bien ».

Dans les deux cas, un système d'interaction aussi élémentaire suppose que la prothèse puisse « proposer » des réglages, ce qui implique une forme d'intelligence artificielle. Bien que les notions de *machine learning*³¹ et de *deep learning*³² dépassent complètement le cadre de ce mémoire de fin d'études, nous pouvons esquisser très théoriquement leurs implications dans nos recherches.

Pour qu'une prothèse auditive puisse faire différentes propositions de corrections fréquentielles, il faudrait dans un premier temps mener des études statistiques à partir d'un nombre conséquent de données sur des corrections réalisées par des sujets malentendants. En appliquant ces études à notre démarche, il faudrait mener un nombre très conséquent de campagnes de tests basés sur l'utilisation de l'analyse-resynthèse IDS pour déterminer quels sont les balances spectrales permettant d'améliorer la compréhension de la parole dans le bruit. Ce n'est qu'après avoir accumulé de nombreux résultats, et à condition que des tendances statistiques puissent être dégagés, que nous pourrions proposer des filtres génériques de correction pertinents.

Il est probable qu'une interface aussi simplifiée serait d'une grande satisfaction auprès des malentendants si elle permettait à la prothèse d'améliorer progressivement la finesse de ses réglages, et si elle pouvait s'adapter rapidement aux situations rencontrées (par exemple en détectant une situation d'écoute qui soit comparable à une précédente et pour laquelle l'utilisateur aurait apprécié le réglage proposé). Mais elle suppose une intelligence artificielle capable de prévoir les attentes de l'utilisateur à partir de peu d'informations.

31. L'apprentissage automatique (*machine learning*) est un processus d'entraînement des machines basées sur des méthodes statistiques ([30], p. 9).

32. L'apprentissage profond (*deep learning*) repose sur un réseau de neurones artificiels, à l'image du fonctionnement de notre cerveau, pour pouvoir extraire des données en s'affranchissant de l'expertise humaine ([31], p. 1).

4.5.4 Version équilibrée

Finalement, nous pouvons imaginer une interface qui soit simple d'utilisation, basée également sur une intelligence artificielle, mais dont l'apprentissage automatique serait facilité par les interactions avec l'utilisateur.

De même que pour la première version, une application sur smartphone ou tablette donnerait accès à davantage d'interactions possibles. On pourrait imaginer deux modes de complexité d'utilisation.

1. Un mode basique, ressemblant à la version décrite en 4.5.3, qui proposerait un nombre restreint de corrections fréquentielles possibles et qui permettrait à l'utilisateur de les comparer à l'écoute et d'éliminer la moins satisfaisante. Par élimination successive, l'apprentissage automatique se baserait sur les caractéristiques des propositions rejetées pour interpréter, en fonction de la situation d'écoute, les critères qui perturbent la compréhension de la parole dans le bruit.
2. Un mode complet qui rajouterait la possibilité d'enregistrer et de rappeler des *presets* pour permettre à l'utilisateur de retrouver directement des réglages satisfaisant dans des conditions familières d'écoute.

Les modes d'interaction avec le dispositif pourraient être de natures différentes (boutons, écran tactile, voire commande vocale si cela est possible pour un utilisateur malentendant) pour pouvoir s'adapter aux préférences de chaque utilisateur. Cette adaptabilité de l'outil nous semble d'autant plus indispensable qu'il s'adresse à des personnes pouvant présenter des réticences face à l'utilisation des nouvelles technologies.

Une fois encore, pour qu'une telle interface puisse fonctionner il faudrait mener des recherches approfondies en terme de caractérisation statistique des balances spectrales favorisant la compréhension de la parole, puis entraîner l'algorithme de réduction de bruit à s'adapter aux différents environnements sonores rencontrés et aux informations de satisfaction envoyés par l'utilisateur. Ces perspectives sont ambitieuses, mais l'évolution très rapide des technologies d'intelligence artificielle pourrait les rendre rapidement abordables.

Lorsqu'une logique d'intelligence artificielle pourra être envisagée pour un tel prototype de prothèse auditive, plusieurs approches d'apprentissage automatique pourront être combinées. Non seulement l'apprentissage devra porter sur l'estimation des balances spectrales les plus adaptées à chaque situation d'écoute, mais il pourrait également étudier les stratégies mises en œuvre par chaque utilisateur pour modifier ces balances lorsque l'environnement sonore évolue. Il s'agirait, lors d'une première séance avec la personne malentendante à équiper, de déterminer quelle type identifié de stratégie de réglage d'une correction elle met en œuvre sur la base d'un test pour plusieurs ambiances usuelles. En fonction de l'identification du type de stratégie, le réglage du cœur d'intelligence artificielle serait ajusté afin de privilégier le type de correction identifié. Évidemment, il faudrait faire une étude préalable sur un nombre suffisant de sujets malentendants pour être en mesure de déterminer les stratégies typiques de correction. Et, on pourrait envisager de garder un historique des stratégies mises en œuvre, entre deux visites, pour définir de nouvelles corrections afin d'affiner la connaissance de la stratégie de la personne mais aussi d'enrichir la base des types de stratégies.

4.6 Conclusion

Dans ce chapitre nous avons vu comment un dispositif de captation semblable à ceux présents dans les prothèses auditives a été conçu. Puisqu'il n'intègre pas de directivité microphonique adaptative, il n'est valable pour les problématiques de réduction de bruit que dans une situation type d'un interlocuteur qui s'exprime face à la personne utilisant le dispositif.

En raison de la nécessité d'obtenir une autorisation médicale préalable s'agissant des recherches concernant le dispositif de diffusion avant de les mettre en œuvre, nous n'avons pu poursuivre la conception d'un prototype de prothèse qui utilise comme traitement du signal l'analyse-synthèse IDS. Nous avons cependant exposé nos réflexions concernant les enjeux d'un traitement en temps réel et d'une interface permettant à l'utilisateur d'interagir avec la prothèse pour affiner ses réglages selon ses propres critères.

Il nous faut désormais revoir les objectifs de ce mémoire, en nous appuyant sur le dispositif de captation réalisé. En effet, ce dernier nous permet de réaliser des enregistrements sonores dans des conditions similaires à la captation faite en permanence par une prothèse auditive. Nous pouvons dès lors reprendre les recommandations données par Ralle pour poursuivre ces recherches :

« Par exemple dans le cadre de notre étude, il ne faudrait donc pas utiliser les microphones Behringer ou TLM103 pour les enregistrements des corpus mais plutôt des microphones directement embarqués dans le réducteur de bruit final (prothèse auditive notamment). » [6], p. 60

5 Réalisation d'un corpus de scènes sonores à partir d'un prototype de prothèse auditive

5.1 Introduction

Le dispositif de captation que nous avons conçu, basé sur une prothèse auditive et utilisant l'un de ses microphones, nous permet de réaliser des enregistrements représentatifs de la captation sonore faite en permanence par les prothèses. Nous pouvons donc poursuivre nos recherches concernant l'utilisation de l'analyse-resynthèse IDS pour améliorer la compréhension de la parole dans le bruit, non plus dans une logique de temps réel comme le nécessiterait l'utilisation véritable d'un appareillage, mais en temps différé. En effet, à la manière des méthodes d'audiométrie vocale utilisées lors du diagnostic des déficiences auditives, nous pouvons nous appuyer sur des scènes sonores préalablement enregistrées pour évaluer par la suite la compréhension de la parole lors de l'écoute de ces scènes.

Nous présentons dans ce chapitre les raisons de la constitution d'un nouveau corpus, les scènes sonores qui le composent ainsi que le dispositif d'enregistrement mobilisé.

5.2 Les raisons de la constitution d'un corpus original

5.2.1 Protocole uniforme et qualité sonore

Comme l'avait expliqué Ralle dans son mémoire, les corpus oraux représentatifs de la voix française existants servent à des recherches sur la sémantique de la parole plus que sur ses caractéristiques psychoacoustiques, ils présentent donc des inconvénients pour notre étude. D'une part la qualité sonore des enregistrements n'est pas suffisamment bonne pour pouvoir mener des analyses fréquentielles précises, mais surtout les conditions d'enregistrement ne sont pas uniformes, ce qui risque de fausser les résultats d'une analyse comparative.

C'est pourquoi Ralle a constitué un corpus original avec du matériel professionnel en respectant un protocole uniforme pour chaque enregistrement, protocole qu'il a d'ailleurs précisément documenté [6], p. 34. Grâce à ce corpus il a pu mener une étude sur trois découpages fréquentiels distincts : un découpage obtenu en considérant toutes les voix du corpus, un découpage associé à une voix féminine choisie ainsi qu'un découpage associé à une voix masculine choisie.

Les découpages obtenus sont des résultats *a priori* intéressants pour toute recherche qui porterait sur les caractéristiques fréquentielles de la voix humaine parlant la langue française, et en particulier sur la réduction de bruit.

Cependant, plusieurs raisons justifient la nécessité de réaliser un nouveau corpus, non plus pour la recherche de découpages fréquentiels, mais pour tester la pertinence de l'analyse-resynthèse IDS pour l'amélioration de la compréhension de la parole dans le bruit pour les malentendants.

5.2.2 Des conditions cohérentes avec l'objet d'étude

Pour une recherche appliquée aux prothèses auditives, deux critiques principales peuvent être faites au corpus réalisé par Ralle, dont la première est déjà explicitée dans sa conclusion.

- Étant donné l'impact du dispositif de captation sur l'analyse IDS, il faut se placer dans des conditions technologiques similaires, voire identiques, à l'objet d'étude. Autrement dit, des enregistrements effectués avec des microphones de mesure ou de prise de son radiophonique ne permettent pas de tester l'amélioration de la compréhension de la parole dans le bruit via l'utilisation de l'analyse-resynthèse IDS, puisque la sensation même d'écoute va être différente par rapport à celle proposée par une prothèse auditive.
- Lors de la conception de l'IDS Speech Enhancer, Ralle a utilisé les enregistrements oraux issus de son corpus et les a « mélangé » à des *ambiances sonores*³³ issues d'une sonothèque pour constituer des scènes sonores imitant une conversation dans un environnement bruyant. Cette étape de mixage des deux enregistrements distincts pour former

33. Dans le domaine de la production audiovisuelle, les ambiances sonores correspondent à des enregistrements effectués dans des lieux pour capter leur ambiance caractéristique, souvent avec un dispositif microphonique stéréophonique ou multicanal.

un nouveau contenu sonore nous semble approximative. D'une part, le fait même que les enregistrements d'ambiance aient été effectués avec un protocole non documenté et non uniforme ne permet pas le contrôle des caractéristiques fréquentielles des scènes sonores qui sont testées. D'autre part, le choix arbitraire de l'écart de niveau sonore entre la voix et l'ambiance lors du mixage rend la scène sonore résultante peu comparable à une situation d'écoute naturelle où tous les sons sont mêlés.

Enfin, nous pouvons qualifier le travail effectué par Ralle pour la constitution d'un corpus oral représentatif de la voix française de recherche prospective concernant les caractéristiques fréquentielles de la voix humaine. Mais dans le cadre de notre mémoire nous souhaitons davantage mener une recherche appliquée au contexte particulier de la réduction de bruit introduit par les prothèses auditives. Nous devons donc concevoir un corpus représentatif de situations où la compréhension de la parole est perturbée par un environnement bruyant, en utilisant le dispositif de captation que nous avons conçu.

5.3 Composition du corpus

5.3.1 Contenu des scènes sonores

L'objectif de ce corpus est de fournir des *stimuli* pour des tests perceptifs permettant d'étudier la pertinence de l'analyse-resynthèse IDS lors de sa manipulation par des malentendants. Puisque nous souhaitons tester l'amélioration de la compréhension de la parole dans le bruit, il faut imaginer des scènes sonores qui placent l'auditeur dans une situation d'écoute comparable à celles pour lesquelles il rencontre des difficultés, à savoir des discussions dans des environnements bruyants. Nous avons donc enregistré des scènes donnant à entendre la voix d'un locuteur s'exprimant dans un espace bruyant, comme s'il s'adressait à l'auditeur.

Concernant le contenu de la voix parlée, nous avons plusieurs possibilités.

- Des listes de mots, à l’image de listes cochléaires du professeur Lafon, sélectionnés judicieusement pour que l’ensemble des phonèmes de la langue étudiée soit prononcés³⁴. Mais la lecture d’une telle liste donnerait à entendre une parole saccadée, sans lien sémantique entre les différents mots. Or, pour notre étude, nous allons demander aux malentendants de manipuler la correction fréquentielle tout en écoutant le résultat de ces manipulations, il nous semble donc important que le débit de parole soit plus fluide.
- Une interview : en posant des questions au locuteur, on espère obtenir par ses réponses une parole très naturelle, proche de celle des conversations du quotidien. Cependant, comme l’a constaté Ralle, la parole d’une personne répondant à des questions risque d’être également peu fluide, que ce soit en raison des questions qui entrecoupent le propos ou des temps de réflexions ou des hésitations.
- La lecture d’un texte : c’est le meilleur moyen d’enregistrer une parole qui soit fluide, en s’assurant que le texte soit connu du locuteur pour favoriser un débit continu. L’inconvénient principal de la lecture, que nous essayerons de minimiser grâce aux indications de « jeu » données au locuteur, est que la voix peut être trop monotone, sans les intonations d’une conversation vivante et improvisée.
- Une histoire personnelle du locuteur qu’il raconte en improvisant : cette suggestion de Ralle semble être un bon compromis entre la recherche d’une parole à la fois vivante (comme pour l’interview) et fluide (comme pour la lecture), puisque l’histoire est connue du locuteur et non interrompue par des questions.

Dans un premier temps, nous avons retenu l’enregistrement de la lecture d’un texte et des histoires personnelles racontées par les locuteurs.

Pour la lecture du texte, nous faisons le choix de conserver celui sélectionné par Ralle pour la constitution de son corpus, à savoir un extrait de la nouvelle de l’écrivain Didier Daeninckx : *Coupe-Coupe* [33]. En effet, sa sélection a été faite en respectant des critères précis pour constituer un enregistrement oral représentatif de la parole :

34. Les listes cochléaires ont été constituées par le professeur Lafon pour renseigner sur les déformations acoustiques provoquées par la cochlée [32].

« D'autre part, le texte ne devait pas employer de manière excessive des figures de style comme la répétition, l'allitération ou l'assonance. Il fallait éviter que certaines sonorités prennent explicitement le dessus dans l'ensemble de la lecture, toujours dans le but d'avoir un corpus représentatif de la voix française. » [6], p.

31

De plus, en conservant le même texte on se place dans une poursuite de recherches qui pourrait permettre une comparaison avec les résultats obtenus par Ralle dans son mémoire. Bien que l'intérêt d'une étude comparative soit à nuancer au vu des différences de protocole d'enregistrement entre les deux corpus, conserver la lecture du même texte nous laisse une chance d'obtenir des résultats comparables.

Une retranscription du texte est disponible en Annexe 2.

Concernant les espaces dans lesquels nous souhaitons placer le locuteur, nous en avons sélectionné trois :

- un marché couvert (celui de Saint-Denis), lieu rythmé par les nombreuses rumeurs des passants se bousculant entre les étals et les cris des vendeurs voulant attirer la clientèle ;
- un hall de gare (hall 1 de la gare de Lyon), grand espace donnant à entendre une rumeur de conversations indistinctes et lointaines mais très réverbérées, une musique improvisée sur un piano, des passages de valises à roulette ;
- un intérieur de métro (ligne 14 du métro parisien) à une heure de pointe, espace bruyant à la fois en raison des vibrations et frottements de la rame lorsqu'elle circule à pleine vitesse et des conversations perturbantes à proximité du locuteur.

Après avoir défini les contenus de parole et d'espaces bruyants que nous souhaitons enregistrer, il restait à déterminer la manière de constituer les scènes sonores mêlant les deux.

5.3.2 Expérimentations

Dans un premier temps, nous pensions enregistrer séparément la voix et les espaces bruyants pour pouvoir multiplier les combinaisons de scènes sonores possibles, tout en facilitant la mise en œuvre des enregistrements. L'autre avantage d'un enregistrement séparé était de pouvoir réaliser des scènes sonores avec un niveau de difficulté de compréhension de la parole plus ou moins élevé, en fonction de l'écart de niveau sonore entre la parole et l'ambiance. Afin de résoudre la problématique des niveaux sonores auxquels « placer » les deux enregistrements lors de leur mixage pour former la scène sonore finale, nous pensions réaliser un enregistrement témoin dans chacun des espaces, avec un locuteur lisant quelques lignes du texte de Didier Daeninckx. Cet enregistrement était destiné à servir de référence pour pouvoir définir les niveaux sonores auxquels « placer » la voix et l'ambiance pour procurer une sensation d'écoute identique.

Nous avons donc réalisé une première série d'enregistrements avec deux locuteurs sélectionnés pour leur pratiques théâtrales. En effet, même dans le cas de la lecture du texte nous demandons aux locuteurs de s'exprimer de manière aussi proche que possible de celle d'une conversation naturelle, en s'adressant au dispositif d'enregistrement comme s'il s'agissait de leur interlocuteur. Pour que cette demande puisse être respectée, et pour que la parole prononcée soit bien articulée, nous nous appuyons sur l'expérience du jeu théâtral ou cinématographique de nos deux locuteurs.

L'enregistrement des voix a été fait dans un studio insonorisé de l'ENS Louis Lumière, pour isoler autant que possible la parole de tout bruit extérieur. Par ailleurs, nous avons effectué l'enregistrement d'une ambiance sonore du marché couvert de Saint-Denis, ainsi qu'un enregistrement témoin comme expliqué précédemment.

Suite à ces premières mises en application du protocole envisagé, nous en avons constaté rapidement les limites. La comparaison de l'enregistrement témoin avec une sommation des deux enregistrements « voix + ambiance » ne permettait jamais de retrouver une sensation d'écoute comparable, quel que soit le niveau sonore des sources mixées entre elles.

Comme nous pouvions nous y attendre, comparée avec un enregistrement d'une voix naturellement située dans un environnement bruyant, la sommation des deux stimuli enregistrés séparément paraît bien plus artificielle. Bien que nous ayons demandé aux locuteurs de s'exprimer en « jouant » comme s'ils se trouvaient dans un environnement bruyant, le résultat de cet enregistrement superposé à celui d'un espace bruyant donne une sensation de décorrélation entre la voix et l'ambiance, s'éloignant ainsi d'une sensation d'écoute naturelle.

Puisque pour notre étude nous voulions constituer un corpus composé de scènes sonores qui se rapprochent autant que possible des sensations d'écoutes des malentendants, nous avons modifié notre approche.

5.3.3 Composition finale

Finalement, nous avons fait le choix d'enregistrer des scènes sonores composées d'une voix parlée qui soit naturellement située dans un espace bruyant. Nous avons enregistré nos locuteurs, non pas dans un studio insonorisé, mais directement dans les lieux que nous avons retenus (marché couvert, hall de gare, intérieur de métro). Ainsi, nous nous sommes placés dans le cadre le plus cohérent pour pouvoir étudier des corrections fréquentielles sur des stimuli qui correspondent aux conditions d'utilisation des prothèses auditives, puisque dans la pratique la voix des interlocuteurs d'une personne appareillée est captée en même temps que l'environnement sonore. Cependant, ce choix présente des inconvénients logistiques qui restreignent le nombre d'enregistrements réalisables. Le fait d'enregistrer les locuteurs dans différents endroits est une contrainte relativement lourde, nous devons donc limiter le nombre d'espaces et la durée des enregistrements. Nous avons décidé également de ne conserver que la lecture du texte comme contenu de parole. Cependant, chaque fois que cela a été possible, nous avons demandé au locuteur de lire le texte avec deux niveaux de voix :

- un niveau normal : avec lequel il s'exprimerait normalement pour s'adresser à un interlocuteur proche de lui ;
- un niveau fort : avec lequel il s'exprimerait s'il sentait que son interlocuteur avait des difficultés pour le comprendre.

Ces deux niveaux devraient nous permettre de conserver l'idée de scènes sonores ayant deux niveaux de difficulté en terme de compréhension de la parole, puisque nous supposons que la scène avec un niveau de voix fort devrait être plus intelligible que celle avec un niveau de voix normal. Encore une fois, nous nous sommes appuyés sur l'expérience théâtrale des locuteurs enregistrés pour que la gestion de ces deux niveaux de voix soit maîtrisée, sans que cela paraisse caricatural.

Pour nommer les enregistrements constituant le corpus nous adoptons la convention suivante :

- en premier vient la dénomination du lieu : « Marche », « Gare » ou « Metro » ;
- vient ensuite un numéro permettant de distinguer le nombre de scènes sonores faites dans un même lieu ;
- on indique ensuite la nature du contenu de parole : dans notre cas, ce sera toujours « txt » puisque finalement nous n'avons retenu que la lecture du texte, mais nous gardons cette convention pour pouvoir étendre le corpus à des contenus de parole d'une autre nature ;
- viennent après des informations sur le locuteur : sexe (« H » pour homme, « F » pour femme), âge, et nationalité³⁵, comme pour la nature du contenu nous indiquons la nationalité pour que cette convention puisse englober l'enregistrement de locuteurs d'une autre nationalité que française ;
- vient enfin une lettre indiquant si le niveau de voix est normal (« n ») ou fort (« f ») ;
- tous ces éléments sont séparés par un *underscore* : « _ ».

Par exemple, si le fichier correspond à la première scène sonore enregistrée au marché avec un locuteur masculin français de 26 ans et que le texte est lu avec un niveau de voix fort, on lui attribuera le nom : « Marche_1_txt_H26fra_f ».

35. Pour indiquer la nationalité du locuteur nous adoptons le système d'abréviation utilisé lors des compétitions sportives. Pour la nationalité française, l'abréviation est « FRA ».

5.4 Dispositif d'enregistrement

5.4.1 Dispositif de captation basé sur une prothèse auditive

Le corpus que nous avons constitué a comme axe fondamental l'utilisation du dispositif de captation que nous avons conçu à partir du modèle de prothèse auditive *Conversa.NT* de la marque *Unitron*. Il conditionne donc totalement le reste de la chaîne de captation, les méthodes d'enregistrement, ainsi que le dispositif de test perceptif développé à partir de ce corpus.

1. Comme nous l'avons expliqué dans le chapitre consacré à la conception de ce prototype de prothèse auditive, nous n'avons pu conclure les manipulations qu'avec un seul modèle, ce qui implique un enregistrement monophonique.
2. De plus, nous devons retenir que la captation microphonique n'est pertinente que dans le cas où le locuteur se tient face à la personne utilisant le dispositif.
3. Enfin, puisque le dispositif se porte à la manière d'une prothèse auditive, il est certain que l'individu qui le porte a une influence sur la qualité et l'homogénéité des enregistrements. Je suis donc l'unique personne à avoir porté le dispositif lors des enregistrements, dispositif que je plaçais toujours sur la même oreille (la droite). Je veillais à me tenir toujours face au locuteur et à ne pas effectuer de mouvements de tête susceptibles de perturber la captation.

Le locuteur comme la personne portant le dispositif de captation étaient en position debout lors des enregistrements, leur visages étant séparés de 50 cm.

5.4.2 Compléments de la configuration technique

Le signal fourni par le microphone intégré à la prothèse auditive étant de faible amplitude, nous avons dû le préamplifier avant de l'enregistrer. L'étape de préamplification était faite au moyen d'une « mixette » 442 de la marque *Sounddevices*, connue dans la production audiovisuelle pour la qualité audio de ses préamplificateurs. L'enregistrement était effectué via un enregistreur mobile *DR100-MkII* de la marque *Tascam*. Afin de contrôler le bon déroulement de l'enregistrement, j'étais équipé d'un casque audio, dont je n'utilisais que l'écouteur gauche pour ne pas perturber le prototype de prothèse présent sur mon oreille droite, ce qui ne posait de problème *a priori* puisque l'enregistrement était monophonique.

Les documentations techniques des équipements cités sont disponibles en Annexe 3 et 4.

Une fois tous les enregistrements terminés, nous avons procédé à une normalisation de leurs niveaux à -9 dB LU³⁶, valeur moyenne de leurs niveaux avant la normalisation.

³⁶. Normalisation faite à -9 dB LU (Loudness Unit), mesuré à court terme sur les passages de niveaux moyens.

5.5 Conclusion

Dans ce chapitre, nous avons exposé la méthode employée pour constituer un corpus de voix françaises parlant dans des environnements bruyants. Les enregistrements ont été effectués avec un dispositif de captation réutilisant le microphone « avant » d'une prothèse auditive. Le contenu de la parole est la lecture d'une nouvelle de Didier Daeninckx intitulé « *Coupe-Coupe* » et les différents espaces sont au nombre de trois : marché couvert, hall de gare, intérieur d'une rame de métro. La convention adoptée pour nommer les éléments du corpus permet de l'étendre à d'autres contenus de parole, d'autres lieux et à des locuteurs de nationalité différente. Elle est aussi adaptée à l'étape d'analyse comme de re-synthèse IDS et elle est susceptible d'être réutilisée pour un traitement semi-automatique, au minimum, d'un lot conséquent d'enregistrements. Dans notre recherche, le corpus constitué a pour but d'être utilisé lors de tests perceptifs permettant à des malentendants d'évaluer la pertinence de l'analyse-resynthèse IDS comme outil de réduction de bruit pour améliorer leur compréhension de la parole dans le bruit.

6 Protocole du test perceptif auprès de malentendants

Dans ce chapitre, nous présentons et justifions les choix opérés pour arrêter le protocole du test perceptif qui a été mené de manière exploratoire. Nous nous intéressons d'abord au choix des sujets en partant des choix opérés par Ralle, puis, nous indiquons les objectifs associés au choix de sujets malentendants, et, enfin, nous exposons les modifications introduites dans l'application utilisée lors des tests, modifications logicielles mais aussi des données et paramètres utilisés en pratique.

6.1 Introduction

Puisque la conception d'un prototype de prothèse basé sur l'analyse-resynthèse IDS n'a pas pu être mise en œuvre dans sa totalité, nous ne pouvons pas le faire tester par des malentendants dans des conditions d'appareillage traditionnelles. Nous avons donc imaginé un protocole de test perceptif, inspiré de ceux effectués par Ralle pour son mémoire, qui soit mieux adapté aux contraintes réelles des prothèses auditives.

En effet, plusieurs critiques peuvent être formulées sur les tests perceptifs qu'il a menés :

- Sur les 19 sujets participant au test, seuls deux d'entre eux étaient malentendants, et leur surdité était légère (pertes comprises principalement entre -20 et - 40 dB). Même si cela n'est pas indiqué dans le mémoire, il est probable que les sujets en question n'étaient pas appareillés. Or, pour que l'étude de la pertinence d'un outil de traitement du signal soit valable en vue de son utilisation au sein de prothèses auditives, il nous semble indispensable de le faire tester par des personnes appareillées.
- Pour les mêmes raisons de cohérence avec l'objet d'étude, des *stimuli* enregistrés avec un dispositif de captation basé sur une prothèse auditive seront certainement plus à même de procurer aux sujets une sensation d'écoute analogue à celle proposée par ces prothèses.

- Si l’objectif du test est d’évaluer l’intérêt d’une manipulation de l’analyse-resynthèse IDS par le sujet, il nous paraît nécessaire de mettre en place un protocole qui permette d’évaluer la variabilité intra-sujet. Autrement dit, il faut que le sujet soit amené lors du test à écouter à deux reprises le même *stimulus* (ou un *stimulus* au contenu très proche) pour observer si ses réglages sont cohérents.
- Si nous permettons au sujet de donner une appréciation de chacun de ses réglages en terme de compréhension de la parole, il faut analyser cette donnée et faire émerger les résultats de cette étude.

Nous avons donc fait le choix de reconduire des tests perceptifs en s’appuyant sur ces remarques pour améliorer le protocole. La première contrainte liée aux améliorations souhaitées est le fait de faire passer le test uniquement à des personnes malentendantes et appareillées. Cela limite le nombre de sujets potentiels et complique la mise en place technique et logistique du test. En effet, pour avoir plus de chance de trouver des sujets volontaires, nous avons élaboré un dispositif de test mobile, pour pouvoir se déplacer au domicile des sujets ou dans des locaux d’associations. Ce choix a des implications sur les conditions technologiques du test et sur le nombre de sujets que nous avons pu rencontrer, puisque les temps de déplacement, d’installation et de rangement sont plus conséquents que lorsque les tests sont menés dans un lieu unique.

6.2 Présentation générale du protocole de test

Puisque ce test perceptif n’est destiné qu’à des sujets malentendants et appareillés, nous ne considérons pas qu’il soit nécessaire de leur faire passer un examen audiométrique. En revanche, lorsque c’est possible nous leur demanderons de nous transmettre un exemplaire des résultats de celui qui a été fait par leur audioprothésiste. Ces documents (l’audiogramme notamment) nous permettront de comparer les déficiences auditives du sujet aux résultats des corrections mises en œuvre lors de l’analyse-resynthèse IDS.

Le test se déroulera de la façon suivante.

- Le sujet s’assoit face à un ordinateur et renseigne certaines informations personnelles (âge et sexe) dans un formulaire affiché à l’écran. Ces informations nous permettront de distinguer les différents sujets en leur donnant un identifiant unique.
- On demande ensuite au sujet de retirer ses prothèses auditives et de s’équiper d’une paire d’écouteurs. Il va écouter 6 scènes sonores dans lesquelles un locuteur parle au sein d’un environnement bruyant.
- A l’aide d’une surface de contrôle muni de *faders*³⁷, le sujet peut modifier la balance spectrale, à savoir la contribution en niveau sonore des différentes bandes fréquentielles. On lui demande de modifier la balance spectrale jusqu’à ce que sa compréhension de la voix soit optimale pour la scène sonore qu’il écoute.
- Le sujet peut à tout moment comparer son réglage avec la scène sonore originale, pour estimer l’amélioration apportée par ses manipulations.
- Lorsque le sujet semble avoir obtenu un réglage optimal, on lui demande d’attribuer une note de satisfaction sur les corrections apportées, toujours par rapport à sa compréhension de la voix au sein de la scène sonore. Il répète ensuite ce processus pour les autres scènes.

L’objectif du test est double :

- évaluer la pertinence de l’analyse-resynthèse IDS comme outil permettant l’amélioration de la compréhension de la parole dans le bruit, en se basant sur la satisfaction des sujets par rapport aux corrections apportées et sur l’étude de la variabilité intra-sujets ;
- analyser les corrections apportées sur la balance spectrale pour chacune des scènes sonores et observer si d’éventuelles tendances de correction émergent pour un même sujet, ou pour différents sujets par rapport à une même scène sonore, puisque, si des tendances se dégagent, elles pourraient permettre de définir des filtres favorisant la compréhension de la parole dans le bruit pour un sujet donné ou pour une situation d’écoute donnée.

³⁷. Les *faders* sont des actionneurs sous forme de tirettes qui agissent en général sur le volume d’une source sonore.

6.3 Améliorations de l'*IDS Speech Enhancer*

Pour mener ces tests perceptifs nous avons d'abord repris les différentes étapes qui constituent le dispositif de test. Puisque nous nous inspirons largement de celui mis en œuvre par Ralle, nous avons conservé le nom donné à l'application utilisée : *IDS Speech Enhancer* (IDSSE). Ainsi, ce nom laisse entendre que le dispositif permet de « réhausser » la parole pour mieux la comprendre.

6.3.1 Diffusion

La diffusion des scènes sonores durant les tests doit répondre à deux préoccupations :

- procurer une sensation d'écoute aussi proche que possible de celle de l'utilisation de prothèses auditives ;
- diffuser un signal dans des conditions techniques (bande passante³⁸ notamment) qui soient semblables à celles des prothèses auditives.

Concernant la sensation d'écoute, il nous semble que les écouteurs intra-auriculaires sont les dispositifs les plus proches de ceux utilisés dans les prothèses auditives, que ce soit les modèles BTE (*behind the ear*) ou RIC (*receiver in the canal*). Puisque ces écouteurs proposent une bande passante plus large que celles des prothèses auditives³⁹, ils répondent également à la contrainte de la cohérence technique. Plus précisément, le fait que leur bande passante soit plus large pourrait nous contraindre à mettre en place un filtrage supplémentaire lors de la diffusion, mais cette restriction fréquentielle est déjà assurée à l'étape du découpage fréquentiel que nous allons aborder par la suite.

38. En télécommunication audio, la bande passante correspond à l'intervalle des fréquences qui sont transmises par un dispositif.

39. En observant plusieurs documentations techniques de prothèses on constate qu'en moyenne leurs bande passante est comprise entre 100 et 8000 Hz [34].

Parmi les différents modèles d'écouteurs intra auriculaires que nous avons à notre disposition, nous avons sélectionné celui ayant l'effet d'occlusion⁴⁰ le moins important, à la fois pour favoriser le confort d'écoute et pour limiter les sensations de résonance au sein du conduit auditif.

Bien qu'un casque audio puisse rendre le test plus confortable, on ajouterait une incertitude dans la diffusion en raison du filtrage induit par le couplage « haut-parleur/pavillon », d'autant plus incontrôlable qu'il dépend de l'anatomie de chaque sujet. Ralle avait également suggéré d'améliorer le dispositif en diffusant simultanément la scène sonore sans corrections au moyen d'enceintes, pour correspondre à l'information sonore « résiduelle » non corrigée par la prothèse. Nous n'avons pas pu approfondir ces réflexions puisque notre dispositif devait être mobile et qu'une diffusion en champs diffus n'a de sens pour des tests perceptifs que si le lieu où ils se déroulent reste le même.

Enfin, le niveau de diffusion sera le même pour tous les sujets au départ du test. Ils pourront ajuster le volume général de la scène sonore via l'interface, ce réglage faisant partie des modifications potentielles pour favoriser la compréhension de la parole dans le bruit.

6.3.2 *Stimuli*

Les six *stimuli* utilisés pour ces tests sont issus du corpus présenté précédemment (5), constitués d'enregistrements de locuteurs français lisant un texte dans un environnement bruyant. Nous avons apporté quelques contraintes supplémentaires à l'utilisation de ces scènes sonores pour répondre aux objectifs du test.

- Pour pouvoir tester la variabilité intra-sujets, nous souhaitons que pour certaines scènes sonores testées, le sujet en écoute une autre de même nature. Pour ce faire, nous avons scindé chaque scène en deux nouvelles, plus courtes, pour qu'au cours du test le sujet soit amené à écouter chacune des deux, mais pas à la suite l'une de l'autre. Ainsi, nous

40. « Cet effet se produit lorsqu'un embout ou une coque de l'appareil est inséré dans l'oreille. La sensation d'occlusion provoque une sensation d'oreille bouchée ainsi qu'une modification de la sensation de sa propre voix » [35].

pourrons observer si les corrections fréquentielles apportées par le sujet sont cohérentes pour chaque moitié de scène sonore. Le fait de rendre les scènes sonores deux fois plus courtes ne perturbera pas le test puisque dans tous les cas elles pourront être écoutées en boucle jusqu'à ce que le sujet aboutisse à un réglage satisfaisant pour lui.

- Pour pouvoir tester l'apport d'un niveau de voix plus fort lors de l'enregistrement du locuteur, correspondant à un potentiel niveau de compréhension plus favorable, nous ferons en sorte que parmi les 6 scènes sonores écoutées, deux d'entre elles correspondent au même locuteur dans le même espace, mais une des scènes avec un niveau de voix normal et l'autre avec un niveau de voix fort.
- Parmi les 6 scènes sonores testées par chaque sujet, il faut que les trois espaces bruyants et qu'un locuteur féminin ainsi que masculin aient été entendus. Cela permettra d'étudier les variations de corrections apportées en fonction des espaces et du sexe du locuteur, pour pouvoir observer d'éventuelles tendances dans ces corrections.

Nous ne pouvons pas savoir à l'avance combien de temps durera le réglage de chaque scène sonore car il sera certainement variable d'un sujet à l'autre. Dans tous les cas, nous considérons que 6 scènes sonores constitue un nombre raisonnable permettant de ne pas trop fatiguer les sujets.

6.3.3 Découpages fréquentiels

L'*IDS Speech Enhancer* permet à son utilisateur de manipuler la balance spectrale de la scène sonore qu'il est en train d'écouter. Pour ce faire, chaque *fader* qu'il manipule agit sur le niveau d'une des sous-bandes fréquentielles obtenues après l'analyse IDS. Cette analyse IDS a été faite au préalable, pour un découpage fréquentiel choisi. Comme l'a fait Ralle pour ses tests, nous pourrions prendre en compte les changements obtenus selon que l'on utilise un découpage général pour un corpus de voix ou un découpage adapté à une voix, féminine ou masculine, particulière. Seulement, le choix entre trois découpages ajouterait une variable supplémentaire à étudier lors des tests, ce qui obligerait à augmenter le nombre de scènes sonores testées par chaque sujet où à procéder à une organisation des tests en deux groupes. Nous ne souhaitons pas allonger la durée de chaque session, car une sollicitation trop importante du

sujet pourrait rendre ses résultats moins voire non exploitables. N'ayant pas assez de temps et de contacts auprès de malentendants pour organiser de nombreux tests, les organiser en deux groupes risquerait d'aboutir à des résultats trop peu nombreux pour pouvoir mener des analyses comparatives.

Nous avons donc fait le choix de ne conserver qu'un seul découpage fréquentiel, celui défini par Ralle à la fin de son mémoire de fin d'études et présenté précédemment (3.2.3).

Comme nous l'avons évoqué concernant la diffusion des scènes sonores, ce découpage limite la bande passante du dispositif à l'intervalle [163 Hz ; 5205 Hz], ce qui restreint fréquemment le signal transmis d'une manière comparable à la diffusion réalisée avec les prothèses auditives.

6.3.4 Outil logiciel

Contrairement à la première version de l'IDSSE qui avait été développée sur le logiciel *Pure Data*, nous avons développé cette version avec le logiciel *Max* en version 7.

L'IDS Speech Enhancer est une adaptation du patch de re-synthèse IDS développé par Laurent Millot et Romain Vuillet pour accompagner le livre *Traitement du signal : Applications avec Pure Data* [17]. C'est à partir de la dernière version de ce patch, intitulé IDS Enhancer et portée par Millot sur Max 7 que nous avons mis au point l'outil logiciel permettant d'effectuer ce test.

Voici un résumé du fonctionnement du patch dans sa version originale, nous présenterons ensuite les modifications apportées pour correspondre aux objectifs du test.

Plusieurs sections différentes composent le patch, et donc son interface utilisateur.

1. Une section permet de sélectionner quel fichier audio on souhaite écouter et une barre de « transport » permet de contrôler la lecture du fichier.

Lorsqu'on sélectionne un fichier, le programme recherche à partir de son nom le dossier permettant d'accéder aux sous-bandes issues de l'analyse IDS. En effet, l'analyse IDS qui a été effectué au préalable (au moyen d'une application indépendante également programmée par Millot en *Python 3*) permet de générer à partir du fichier original 10 nouveaux fichiers (si le découpage comprend 10 sous-bandes), chacun étant le signal

audio correspondant uniquement à une bande fréquentielle. Ainsi, lorsqu'on lance la lecture du fichier, ce sont 10 lecteurs synchronisés qui gèrent chacun une sous-bande, et qui permettent d'écouter directement une re-synthèse des sous-bandes.

2. Une section permet de manipuler la balance spectrale, en agissant sur le niveau sonore de chaque lecteur au moyen de 10 *faders*. Deux boutons permettent de remettre tous les niveaux des sous-bande à une contribution nulle (« Zero ») ou à la contribution naturelle (« Reset »).
3. Une section permet d'enregistrer la balance effectuée en indiquant les informations sur le sujet (âge, sexe). Les informations sur le sujet, la date de l'enregistrement ainsi que la valeur de niveau sonore retenue pour chacune des sous-bandes sont inscrites dans un fichier texte. Ce fichier nous permet par la suite d'analyser les corrections apportées ou de charger la balance enregistrée.
4. Une section permet de charger une balance préalablement enregistrée.

Voici les modifications apportées au patch pour permettre de mener les tests.

1. La sélection du fichier à lire n'est plus disponible pour l'utilisateur. Il lance le test après avoir fourni des informations personnelles, les scènes sonores sont alors choisies aléatoirement dans le dossier de travail (en respectant les contraintes concernant le choix de ces *stimuli*) et sont lues automatiquement à la suite.
2. La section permettant de manipuler la balance spectrale a été réduite à 8 sous-bandes, pour correspondre au découpage fréquentiel défini et aux limitations de la surface de contrôle MIDI⁴¹. Cette surface de contrôle permet de faciliter la manipulation de la balance spectrale au moyen de ses 8 *faders* motorisés, et du dispositif global puisque d'autres boutons sont assignés aux fonctions « Bypass », « Mute », « Solo ». La fonction « Bypass » ajoutée permet de comparer les corrections fréquentielles apportées par rapport à la scène sonore originale, pour mesurer l'amélioration de la compréhension de la parole.

41. Nous utilisons une surface de contrôle MIDI *BCF2000* de la marque *Behringer*.

3. Une section a été rajoutée sur l'interface à l'écran pour pouvoir noter (sur 5 échelons de valeur) l'amélioration de la compréhension de la parole permise par les corrections effectuées.
4. L'enregistrement de la balance spectrale se fait automatiquement, après que le sujet ait validé la balance, lui ait attribué une note, et soit passé à la scène sonore suivante.

6.4 Conclusion

Dans ce chapitre nous avons présenté le protocole de tests perceptifs que nous avons mis au point pour permettre à des sujets malentendants de tester l'utilisation de l'analyse-resynthèse IDS pour améliorer leur compréhension de la parole dans le bruit. Nous avons apporté plusieurs modifications à l'*IDS Speech Enhancer* tel qu'il avait développé par Millot et Ralle, pour pouvoir l'ancrer davantage dans le contexte de l'utilisation de prothèses auditives.

Dans le chapitre suivant nous présentons les résultats des tests exploratoires que nous avons menés.

7 Résultats des tests et améliorations du dispositif

Dans ce chapitre nous présentons les résultats des tests exploratoires qui ont été menés ainsi que les modifications à apporter au protocole qui sont apparues à l'issue de la réalisation et du dépouillement de ces tests.

7.1 Résultats des tests

Une campagne de tests perceptifs a été menée auprès de 6 personnes malentendantes et appareillées, vivant dans la région parisienne, âgées de 51 à 80 ans. Les tests se déroulaient selon le protocole présenté précédemment, et duraient de 30 à 45 minutes en fonction des sujets.

Pour rappel, l'objectif était à la fois d'évaluer la pertinence de l'*IDS Speech Enhancer* dans sa nouvelle version, et de pouvoir, à partir des corrections fréquentielles choisies par le sujets lors de son utilisation, esquisser des filtres permettant l'amélioration de la compréhension de la parole dans le bruit, pour un sujet donné, ou pour une situation d'écoute donnée.

Ces tests nous ont apporté des résultats de deux natures différentes :

- les observations des sujets lors de la séance, dans leur manière d'accueillir notre démarche, de manipuler l'outil, de réagir durant les tests, constituent des témoignages précieux de personnes totalement concernées par notre objet d'étude pour critiquer notre travail.
- les résultats concrets des tests forment des données qui, après études statistiques, permettent de répondre aux objectifs fixés.

Étant donné le petit nombre de personnes ayant pu participer aux tests, l'étude statistique des résultats des tests reste à un stade préliminaire, et devrait être confirmée par la reconduction d'autres tests pour pouvoir confirmer les conclusions esquissées.

7.1.1 Enthousiasme et satisfaction des sujets

L'un des aspects les plus encourageant des différentes rencontres effectuées auprès de ces malentendants est l'accueil dont ils ont fait preuve vis-à-vis de notre démarche. Ils se sont tous

montrés très enthousiastes à l'idée qu'une telle recherche puisse aboutir à une amélioration de leurs prothèses auditives pour pouvoir « entendre dans le bruit ». En effet, nombre d'entre eux racontent que dans des environnements bruyants ils préfèrent « débrancher leur prothèses » pour être « tranquille ». Or cette démarche les rend encore moins habitués à leurs prothèses, ce qui aggrave leur difficulté à comprendre un interlocuteur lors d'une discussion au restaurant ou dans les transports en commun.

Tous se sont montrés volontaires pour manipuler le dispositif de correction fréquentielle qui était mis à leur disposition et se sont félicités de pouvoir grâce à lui « baisser le bruit » lors de l'écoute des scènes sonores proposées. La compréhension de l'intérêt des différents actionneurs sur l'interface était variable mais paraissait toujours suffisante pour que le dispositif puisse être maîtrisé. Cependant, il semble que les deux premières scènes sonores écoutées étaient l'occasion pour plusieurs sujets de s'habituer à la manipulation de l'outil, et formaient donc une phase d'entraînement dont les résultats seraient moins valables.

Le questionnaire concernant la pertinence de l'outil proposé peut déjà être conforté par la satisfaction dont les sujets ont souvent fait preuve, puisque les corrections qu'ils ont pu apporter leur ont donné l'impression d'améliorer la compréhension⁴² de la voix au sein de la scène sonore.

7.1.2 Variabilité intra-sujets

Afin d'étudier la pertinence de l'IDSSE selon une autre approche que l'observation du ressenti des sujets lors de son utilisation, nous avons étudié le critère de variabilité intra-sujets. Parmi les 6 scènes sonores qui étaient diffusées durant la séance de tests, deux d'entre elles étaient des *stimuli* quasi identiques. Elles correspondaient aux deux moitiés d'un même enregistrement, avec un même locuteur et dans un même environnement. Pour ce faire, chaque scène sonore enregistrée a été divisée en deux pour former deux fichiers. On retrouve donc dans le répertoire des *stimuli* deux fichiers dont le nom est presque identique, à la différence du « 01 » ou « 02 »

42. Cette compréhension restait toujours insuffisante selon eux, aspect sur lequel nous reviendrons par la suite.

correspondant à la première ou deuxième moitié de la scène sonore.

Ainsi, on a cherché à étudier les différences de corrections apportées par le sujet entre ces deux moitiés d'une même scène sonore. Plus l'outil proposé est performant, plus on devrait observer une cohérence dans les corrections apportées pour ces deux *stimuli* quasi identiques. Pour pouvoir étudier ces différences de corrections, on utilise le gradient IDS, un outil développé par Millot permettant de visualiser les différences de poids des sous-bandes de deux balances spectrales respectant un même découpage fréquentielle. A l'image des portraits IDS qui permettent une étude absolue de la répartition énergétique d'un enregistrement audio selon un découpage fréquentielle donné, les gradients IDS permettent de mener cette étude de manière comparative, entre deux fichiers.

Prenons l'exemple de l'étude de la variabilité intra-sujets pour le sujet n°3 et pour la scène sonore Gare_2_txt_F22fra_f. On étudie donc le gradient IDS (cf Figure 9) entre les corrections apportées à Gare_2_txt_F22fra_f_01 et Gare_2_txt_F22fra_f_02 par ce sujet n°3.

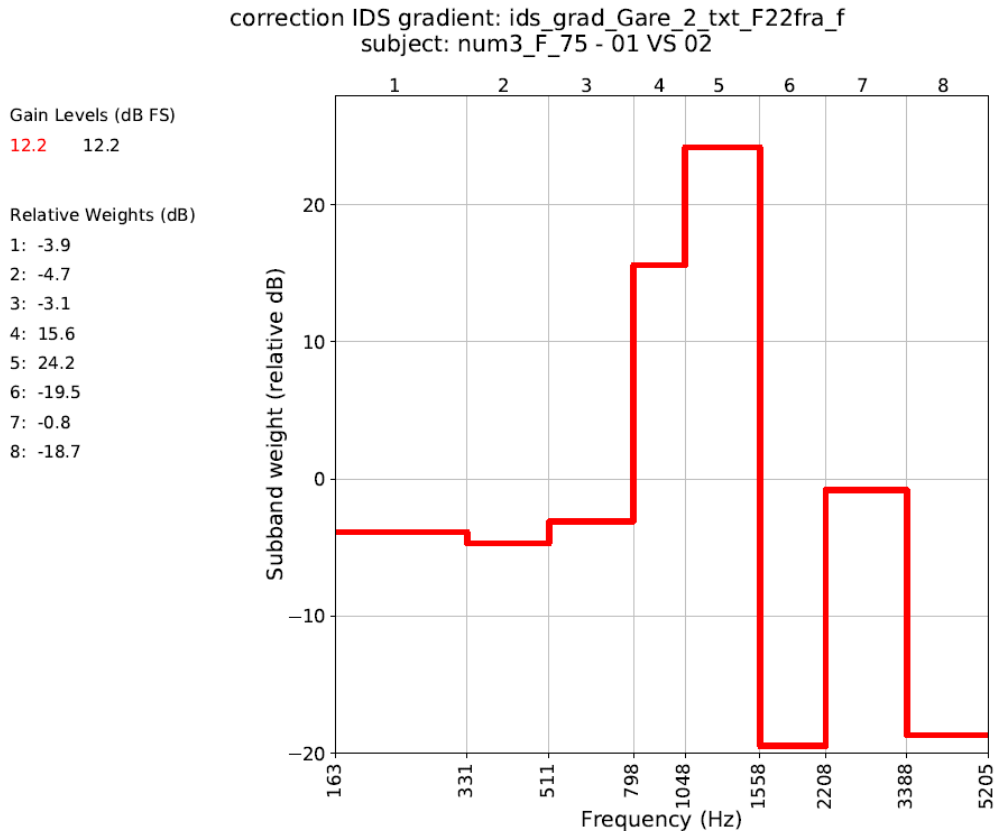


FIGURE 9 – Gradient IDS des corrections du sujet n°3 entre la première et deuxième moitié de la scène sonore Gare_2_txt_F22fra_f.

On peut observer, par lecture graphique ou en se reportant au détail des poids relatifs de chacune des sous-bandes (section *Relative Weights*), que les différences de corrections sont très importantes sur certaines bandes fréquentielles, atteignant jusqu'à 24.2 dB pour la bande n°5.

Avant d'interpréter ces résultats, il faut prendre en compte les variations de la répartition énergétique au sein des deux fichiers audios comparés. Bien que les deux *stimuli* soient issus d'une même scène sonore, des évènements ponctuels ont pu faire varier le profil IDS de chaque fichier, ce qui expliquerait pour partie les variations observées sur le gradient IDS de la figure 9.

On étudie donc dans un second temps le gradient IDS représenté sur la figure 10 des deux fichiers en question, Gare_2_txt_F22fra_f_01 et Gare_2_txt_F22fra_f_02.

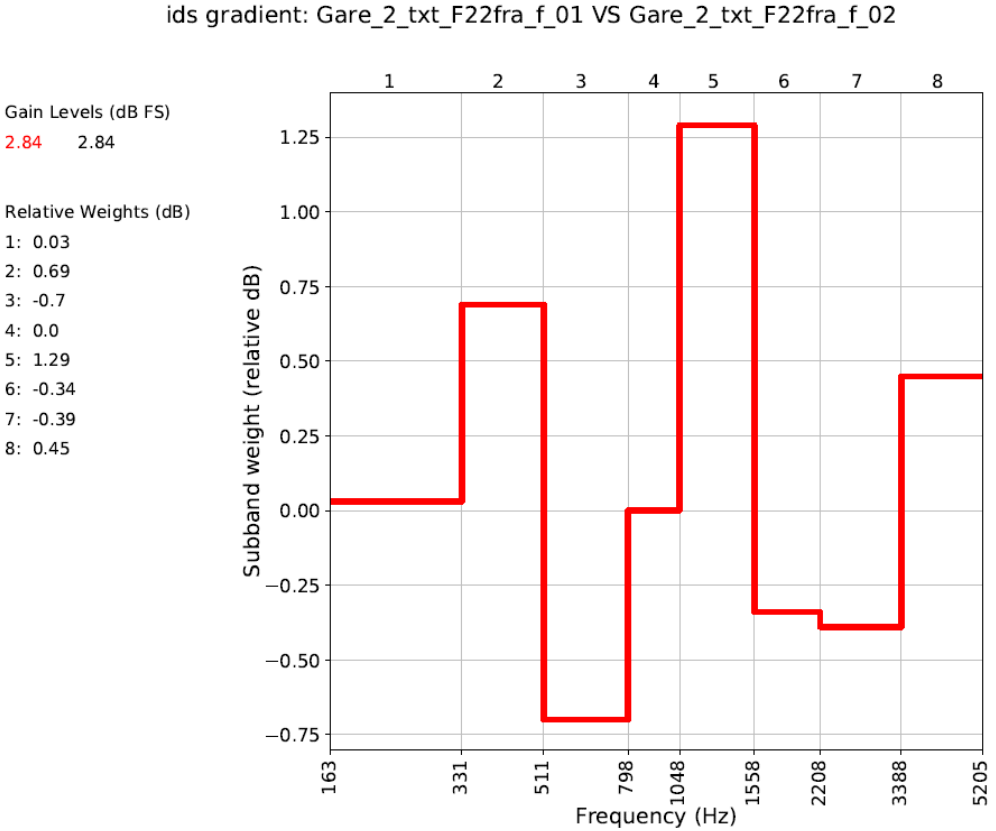


FIGURE 10 – Gradient IDS entre la première et deuxième moitié de la scène sonore Gare_2_txt_F22fra_f.

Sur ce gradient IDS, on observe que les variations entre les deux portraits IDS étudiés sont de l'ordre d'un décibel, avec au maximum 1.29 dB de différence pour la bande n°5. Donc, ces variations sont négligeables par rapport à celles observées précédemment.

Afin de conclure sur cette étude de la variabilité du sujet n°3 quand il est confronté à la scène sonore Gare_2_txt_F22fra_f, on peut établir un dernier gradient IDS, intitulé *IDS global variations*, qui donne les différences de corrections apportées par le sujet pour les deux moitiés de la même scène sonore, auxquelles on a retranché la variations de la répartition énergétique au sein des deux enregistrements en question. Autrement dit, on a fait la différence entre le gradient IDS présenté en figure 9 et celui présenté en figure 10.

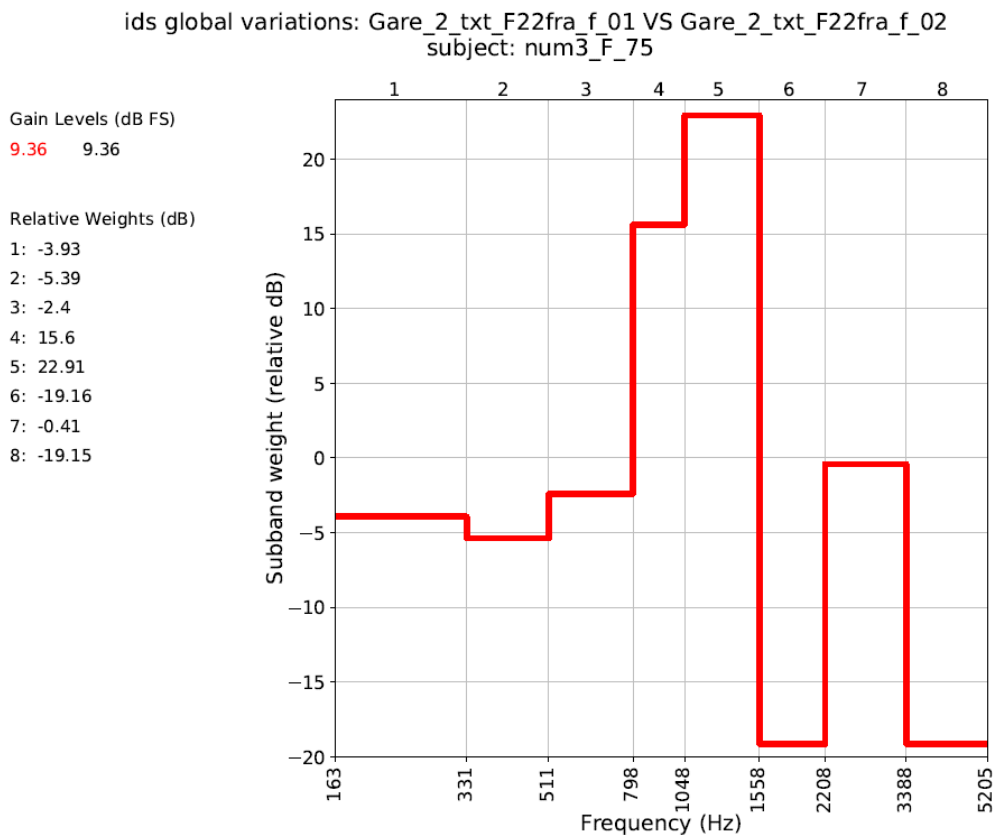


FIGURE 11 – Gradient IDS des variations globales de corrections du sujet n°3 entre la première et deuxième moitié de la scène sonore Gare_2_txt_F22fra_f.

A l'aide de ce gradient IDS des variations globales, on peut conclure s'agissant de l'étude de la cohérence d'un même sujet par rapport aux corrections qu'il a apporté pour deux *stimuli* quasi identiques. Dans notre exemple, on observe donc des variations de corrections conséquentes,

allant jusqu'à 22.91 dB pour la bande n°5. Nous avons observé cette tendance de variabilité intra-sujets conséquente pour l'ensemble des sujets testés. Nous présentons en annexe 5 les autres gradients IDS de variations globales étudiés. Cette observation laisse à penser que le dispositif de test mis au point n'est pas concluant, puisque les sujets n'aboutissent pas à des corrections cohérentes pour des *stimuli* semblables.

L'une des raisons qui pourraient expliquer ce résultat est le lien entre la nature des scènes sonores et les méthodes de manipulation de l'*IDSSE* que l'on a pu observer. Dans la plupart des enregistrements effectués, des événements ponctuels viennent perturber l'environnement sonore, le rendant non stationnaire. Or, lors de la modification de la balance spectrale par les sujets, il est fréquent que ces événements viennent influencer sur le réglage du poids de l'une ou l'autre sous-bandes fréquentielles, laissant penser qu'il y a une forme de confusion entre l'impact de ces réglages sur l'écoute de la scène sonore et le changement intrinsèque propre à l'enregistrement.

7.1.3 Tendances de correction

Outre l'étude de la variabilité intra-sujets, les tests perceptifs réalisés nous permettent de comparer les corrections apportés par différents sujets pour une même scène sonore. Par cette étude comparative nous espérons pouvoir esquisser des tendances de corrections pour améliorer la compréhension de la parole dans le bruit dans une situation donnée. Pour ce faire, nous étudions un graphique montrant plusieurs profils IDS relatifs à un même enregistrement, à l'image de l'exemple présenté en figure 12.

IDS corrected profiles - audio scene: Marche_1_txt_F22fra_f_02

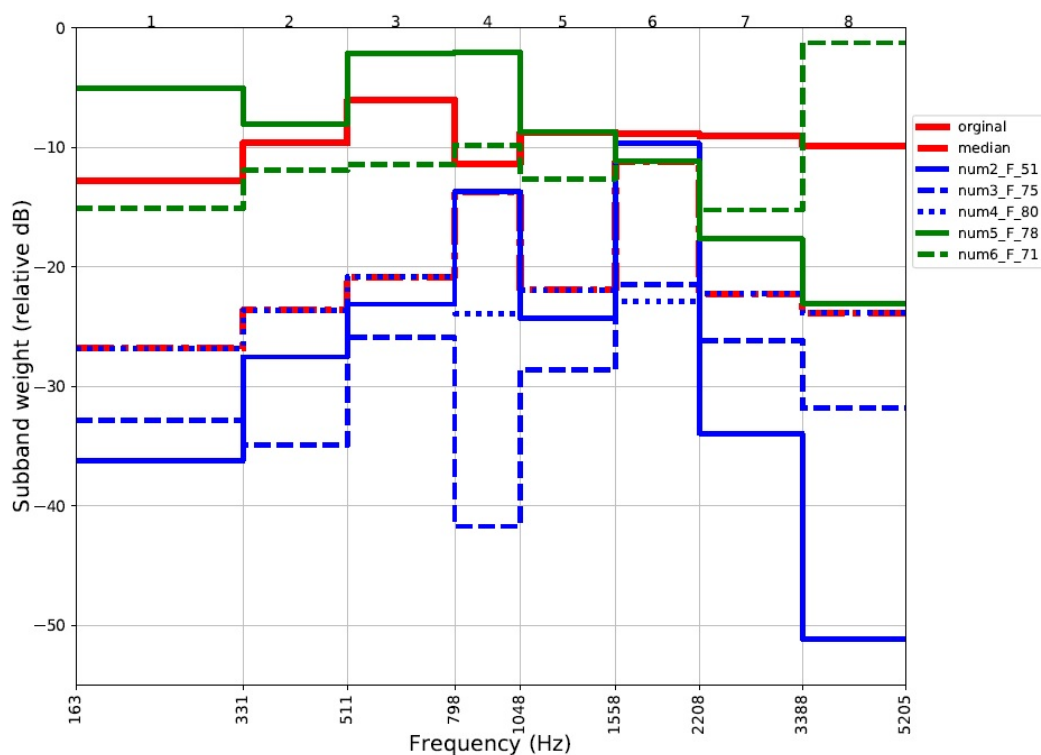


FIGURE 12 – Comparaisons des profils IDS des corrections apportées par 5 sujets pour la scène sonore Marche_1_txt_F22fra_f_02.

Sur cette figure, nous pouvons comparer :

- le profil IDS original de la scène sonore Marche_1_txt_F22fra_f_02 ;
- les profils IDS de cette même scène mais en prenant en compte les corrections apportées par les sujets 2,3,4,5 et 6 ;
- le profil IDS médian, correspondant pour chaque sous-bande à la valeur médiane des 5 courbes cités précédemment.

Comme nous pouvons le voir, il est très difficile d’aboutir à des tendances de corrections, puisque les différentes courbes représentent des résultats chaotiques. Cette grande variabilité des courbes de corrections se retrouvent sur l’ensemble des enregistrements testés, comme en attestent les autres graphiques IDS présentés en Annexe 6.

L’absence de tendance de corrections observables peut s’expliquer par le nombre trop faible de sujets testés. Mais cependant, nous pouvons émettre l’hypothèse que même avec un plus

grand nombre de tests ces tendances n'apparaissent pas, puisque les corrections de la balance spectrale dépendent de l'audition de chacun des sujets. Or, et ce d'autant plus pour un public ayant des déficiences auditives très marquées, chaque sujet possédant une audition très spécifique va corriger la balance spectrale pour améliorer sa compréhension de la parole de manière très différente des autres. Cette recherche de tendances de corrections et de filtres génériques aurait alors potentiellement moins de valeur que celle d'une plus grande adaptabilité du réglage des aides auditives à chaque malentendant.

7.1.4 Méthodes de manipulation

L'observation des méthodes de manipulation employées par les sujets durant les tests nous permettent à la fois d'évaluer la pertinence de l'outil mis en place, d'améliorer le dispositif de test, mais également d'esquisser des tendances de manipulations, qui pourraient permettre à très long terme de mettre en œuvre des algorithmes d'apprentissage automatique.

Comme nous l'avons évoqué précédemment, nous avons constaté que certains réglages du poids des sous-bandes étaient liés à des événements ponctuels intervenant dans la scène sonore. Or, la plupart des sujets ne font qu'une seule « passe » de réglage, c'est-à-dire qu'ils ne modifient qu'une seule fois chacun des 8 faders, en allant toujours des graves aux aigus. Ces deux observations amènent à penser que les corrections introduites par les sujets sont parfois faussées par des variations intrinsèques à l'enregistrement. A moins de demander aux sujets de revenir plusieurs fois sur les réglages pour les confirmer à des instants différents de la scène sonore, il faudrait revoir le contenu de ces scènes, comme nous le proposons par la suite.

Globalement, les sujets préfèrent retirer du niveau qu'en rajouter. Cette observation était d'autant plus flagrante lorsque les scènes sonores présentaient des environnements très bruyants. Parmi les trois environnements choisis, le métro et le marché sont les scènes qui ont posé le plus de problèmes de compréhension du locuteur car il s'agissant d'ambiances sonores plus fortes que celle du hall de gare. Dans ces cas d'environnement trop forts, les sujets avaient tendance à baisser toutes les bandes fréquentielles pour « baisser le bruit », quitte à perdre complètement l'intelligibilité de la voix.

Les observations faites concernant les méthodes de manipulation dépendent certainement de la disposition initiale des 8 faders et des éventuels conseils d'utilisation données par le superviseur du test. Dans notre cas, les 8 faders étaient au départ au niveau 0 dB, correspondant à la contribution naturelle de chaque sous-bande dans la balance spectrale. De plus, leur répartition de gauche à droite correspondait à l'évolution des fréquences les plus graves aux plus aigües. Enfin, aucun conseil d'utilisation n'était donné aux sujets.

Afin de vérifier les hypothèses faites précédemment, il faudrait proposer des dispositions initiales de l'interface qui soient différentes, par exemple avec tous les faders au niveau minimal (aucune contribution des sous-bandes) ou avec une répartition entre les sous-bandes graves et aigües qui soient aléatoires. On pourrait également conserver une répartition traditionnelle avec les fréquences graves à gauche et les fréquences aigües à droite, mais en guidant le sujet en lui conseillant de commencer par les sous-bandes centrales.

7.2 Améliorations du dispositif de test

Nous avons constaté plusieurs défauts dans le dispositif de test mis en œuvre initialement, certains en menant les tests et d'autres en étudiant les résultats. Nous avons donc identifié plusieurs améliorations possibles, dont certaines sont déjà intégrées dans la version actuelle de l'*IDS Speech Enhancer*.

7.2.1 Dispositif d'écoute

Le dispositif d'écoute de l'*IDDSE* doit être adapté à un public malentendant, ce qui suppose d'avoir des niveaux de puissance acoustique très conséquents, puisque nous faisons passer les tests aux sujets sans leurs prothèses auditives. Or, dans le cadre de notre campagne de test nous n'avions pas d'autre matériel que celui utilisé habituellement dans la production musicale et audiovisuelle. Nous avons pu régler en partie le problème en rajoutant un étage d'amplification analogique en sortie de notre dispositif, mais nous arrivons cependant à la limite de ce qu'un casque audio classique peut supporter comme niveau électrique avant d'induire des distorsions harmoniques. Pour poursuivre les tests avec davantage de marge de manœuvre s'agissant du niveau d'écoute souhaité, il faudrait s'équiper d'un dispositif d'écoute tel que ceux utilisés par les audioprothésistes.

De plus, nous avons ajouté à notre programme la possibilité de régler des niveaux d'écoute différents entre le canal droit et gauche. Cette amélioration permet de s'adapter davantage aux malentendants rencontrés, notamment pour ceux qui sont appareillés d'un côté et implantés de l'autre (qui ont un implant cochléaire). Puisque l'implant cochléaire ne peut pas se retirer, il faut pouvoir régler avec des écarts de niveaux très conséquents le son diffusé à chacune des deux oreilles du sujet.

7.2.2 Écoute comparative et notation

Lorsque nous présentions le test en début de chaque séance, nous expliquions la fonction « bypass » permettant de comparer les réglages effectués avec l'enregistrement original. Malgré cette explication, avec démonstration, nous avons constaté que durant les tests rares sont les sujets à avoir utilisé cette fonction. Or elle nous semble essentielle, surtout lorsqu'on pense

avoir trouvé un réglage satisfaisant et qu'on donne une appréciation notée de l'amélioration de la compréhension de la parole. Mais nous pensons qu'elle a été peu employée car cela rajoute une manipulation supplémentaire, avec un outil déjà complexe. C'est pourquoi nous pensons qu'il faudrait séparer les séances de tests en deux temps :

- séance de manipulation de la balance spectrale et d'enregistrement des corrections qui semblent satisfaisantes ;
- séance d'écoute comparative de différentes corrections proposées : scène originale sans correction, correction faite par le sujet précédemment, autre correction provenant d'un répertoire.

Ainsi, c'est lors de la deuxième séance que le sujet pourrait se concentrer sur la comparaison de différentes balances spectrales, sans savoir laquelle correspond aux réglages qu'il avait arrêtés, pour pouvoir déterminer de manière plus efficace quelle correction lui apparaît la plus satisfaisante.

L'autre élément important pour que cette fonction « bypass » soit fonctionnelle, c'est de tenir compte du filtrage induit par le découpage fréquentiel voulu. Ce découpage, en 8 sous-bandes dans cette version de l'IDSSE, signifie que même si les 8 faders sont au niveau nominal (celui de leur contribution à la balance spectrale), un filtrage passe-bande modifie sensiblement la scène sonore, puisque les fréquences inférieures à la bande n°1 et supérieures à la bande n°8 ne sont pas diffusées. Concrètement, dans notre cas et avec le découpage fréquentiel défini, il faut que lorsque l'on active la fonction « bypass » :, le programme nous fasse entendre non pas l'enregistrement original mais le résultat de cet enregistrement convolué par un filtre passe-bande d'intervalle [163 Hz ; 5205 Hz]. Cette amélioration va être apportée dans une nouvelle version de l'IDSSE mais il faudra ne pas oublier d'en tenir compte lorsque le programme permettra un choix entre plusieurs découpages fréquentiels différents.

7.2.3 *Stimuli*

La dernière amélioration à apporter au dispositif de test concerne les scènes sonores enregistrées. En effet, nous avons choisi trois environnements considérablement bruyants, puisqu'on ciblait les situations d'écoutes le plus extrêmes. Mais, comme nous l'avons dit, face à ces

situations plusieurs sujets se sont retrouvés à baisser le niveau de toutes les sous-bandes fréquentielles, baissant ainsi le bruit mais aussi la voix du locuteur par la même occasion. Ces scènes sonores restent valables selon nous, puisque tout l'enjeu est de pouvoir déterminer quel est l'équilibre optimal pour conserver l'intelligibilité de la voix tout en baissant le bruit environnant, même pour ces enregistrements. Mais il faudrait certainement proposer différents niveaux de difficulté pour les *stimuli* proposés lors du test. On pourrait commencer par enregistrer les locuteurs dans des lieux calmes, comme un salon ou parc, puis dans des environnements légèrement bruyants. Ainsi, on pourrait créer un corpus avec au moins deux niveaux intermédiaires avant d'arriver à des situations d'écoutes aussi délicates que celles déjà enregistrées.

Cette démarche de difficulté variable permettrait à chaque sujet de s'habituer progressivement à l'IDSSE, et également d'adapter le niveau de complexité des situations d'écoute en fonction des sujets. On pourrait alors s'inspirer des méthodes de réhabilitation de l'audition pour les personnes implantées : après leur opération, une équipe de médecins les accompagne pour « réapprendre à entendre », puisque leur audition a été considérablement modifiée par la pose de l'implant.

La difficulté des scènes sonores pourraient être liées à la fois à l'environnement plus ou moins bruyant, mais également à l'articulation plus ou moins marquée des locuteurs. En effet, on a constaté que les enregistrements pour lesquels on avait demandé aux locuteurs d'articuler davantage et de parler plus fort ont souvent été plus appréciés par les sujets en terme de compréhension de la voix. Mais cette accentuation de l'articulation et du niveau sonore lors de la lecture pourrait être encore renforcée, pour créer des situations d'écoutes plus « simples ».

Enfin, le contenu de la lecture pourrait être modifié, pour privilégier un texte de la culture populaire (fable, chanson, hymne national), afin que tous les sujets puissent reconnaître dans la lecture un contenu familier, et ainsi évaluer plus précisément l'amélioration de la compréhension de la voix dans le bruit.

8 Conclusion générale

A travers ce mémoire nous avons cherché à confronter l'analyse-resynthèse IDS aux contraintes technologiques et médicales des prothèses auditives, pour étudier si on peut en faire un outil de traitement du signal qui puisse favoriser la compréhension de la parole dans des environnements bruyants pour les malentendants.

Tout d'abord nous avons présenté le contexte général des déficiences auditives et des prothèses auditives, qui forment la solution de réhabilitation prothétique la plus répandue. Le déficit auditif étant l'un des enjeux de santé et de société actuel des plus majeurs, nous avons souhaité contribuer aux recherches permettant d'améliorer des outils qui, bien qu'ils ne permettent pas de « rendre » l'audition, ont le mérite de limiter l'isolement psychologique et social dont peuvent souffrir les malentendants.

Dans un deuxième temps nous avons décrit le fonctionnement de l'Intégrateur de Densité Spectrale (IDS) dans sa version analogique développée par Leipp, puis dans sa version numérique développée par Millot. Puisque nos travaux s'inscrivaient dans la suite logique de ceux menés par Ralle dans son mémoire de fin d'études de l'*ENS Louis Lumière*, nous en avons présenté les grands axes. Tout au long de notre démarche nous nous sommes appuyés sur l'idée que ses travaux devaient être approfondis par une recherche davantage appliquée au cas concret des prothèses auditives, dont les contraintes technologiques, perceptives, médicales, modifient l'approche.

Ensuite nous avons expliqué comment un prototype de prothèse auditive basé sur l'analyse-resynthèse IDS a été modélisé et partiellement conçu. La conception d'un dispositif de captation utilisant le microphone « avant » d'un modèle de prothèse préexistant a été réalisée, permettant de réaliser des enregistrements dans des conditions similaires à notre objet d'étude, à condition que le locuteur s'exprime face à la personne utilisant le dispositif. Nous avons par la suite présenté les optimisations à apporter aux algorithmes de l'analyse-resynthèse IDS pour qu'ils puissent fonctionner en temps réel, condition indispensable à leur intégration dans des prothèses. Enfin, nous avons détaillé les réflexions menées concernant l'interface qui permettrait à l'utilisateur d'optimiser les réglages du prototype par rapport à ses critères d'écoute et en fonction de chaque situation rencontrée, et les implications d'une telle interface dans

l'optique d'une programmation par apprentissage automatique (ou *machine learning*).

En utilisant le dispositif de captation basé sur une prothèse auditive nous avons constitué un corpus de scènes sonores, dont chacune donne à entendre un locuteur s'exprimant dans un environnement bruyant. A partir de ces scènes sonores nous avons mis au point un protocole de tests perceptifs permettant à des sujets malentendants d'évaluer la pertinence de l'analyse-resynthèse IDS comme outil de correction fréquentielle pour l'amélioration de leur compréhension de la parole dans le bruit. Le dispositif *IDS Speech Enhancer* qui avait été conçu par Ralle à partir des applications développées par Millot et Vuillet a été repris et modifié, pour en faire une version optimisée qui soit plus appliquée aux conditions de captation des prothèses et aux sensations d'écoute des malentendants appareillés.

Une campagne de tests perceptifs a été menée auprès de quelques personnes malentendantes et appareillées. Bien que le nombre de sujets soit trop faible pour tirer des conclusions définitives, il semblerait que le dispositif mis en œuvre permette aux malentendants d'améliorer leur compréhension de la parole lors de l'écoute des scènes sonores. Mais plusieurs améliorations peuvent être apportées, dont certaines nécessitent des partenariats avec des audioprothésistes et des fabricants de prothèses. Un dispositif d'écoute identique à ceux utilisées par les audioprothésistes lors du réglage des prothèses permettrait d'éviter les risques de distorsions induits par les équipements d'écoute traditionnels dans la production audiovisuelle. Une plus grande variété des *stimuli* testés, possédant des environnements plus ou moins bruyants, permettrait de s'adapter davantage aux sujets en leur proposant un niveau de difficulté du test qui soit plus progressif. Enfin, les tests pourraient être séparés en deux séances distinctes : l'une serait dédiée à la manipulation des corrections des balances spectrales pour chercher celles qui soient les plus favorables à la compréhension de la parole dans le bruit pour le sujet, l'autre consisterait à faire des écoutes comparatives pour définir, après plusieurs jours de repos, quels corrections sont effectivement plus efficaces.

Sur un plus long terme, l'objectif serait de reconduire des tests avec un prototype de prothèse complet intégrant l'analyse-resynthèse IDS. Le partenariat avec des industriels devrait permettre de progresser notamment concernant la mise en œuvre de la correction proposée en temps réel. Outre la pertinence de l'analyse-resynthèse IDS pour traiter la réduction de bruit, ces tests nous ont permis d'observer de très nets différences de corrections apportées

par chaque sujet. Si ces observations se vérifiaient sur un plus grand échantillon de personnes, on pourrait conclure quant à l'impossibilité de déterminer des filtres génériques favorisant la compréhension de la parole. Il faudrait alors poursuivre la piste d'une prothèse auditive totalement adaptée à l'audition de chaque malentendant et à ses préférences pour chaque situation d'écoute rencontrée.

Mais, il faudra au préalable se pencher sur la question des compromis pouvant être introduits dans le design des filtres d'analyse et, par suite, du filtre de correction pour chacune des prothèses. Il s'agira de déterminer grâce à des tests d'écoute et en fonction des capacités des architectures de traitement utilisables (tablette, téléphone, périphérique embarqué, ...) quelle atténuation minimale et quelle largeur de bande de transition maximale sont possibles. Il faudra probablement se contenter d'une implémentation de convolution utilisant une bibliothèque de FFT dans un premier temps, peut-être avec des filtres d'analyse, et donc de correction, légèrement moins performants que les filtres minimaux nécessaires.

Il reste aussi en suspens la question du réglage sur le terrain consistant à modifier une correction pré-définie afin de l'adapter à une situation où les corrections pré-définies ne sont pas satisfaisantes. Pour réussir à rendre utilisable cette option à tous les utilisateurs, dont notamment ceux qui sont plutôt peu à l'aise avec la technologie, il y aura un travail conséquent de design de l'interface à proposer pour piloter le réglage d'une nouvelle correction. S'agissant de la mise à jour du filtre de correction, pendant la phase d'adaptation d'une correction à une nouvelle situation, une solution pourrait consister à mettre en œuvre une modification partielle du filtre de correction en augmentant (ou en diminuant) un peu le niveau de la sous-bande sélectionnée, c'est-à-dire à ajouter, au filtre de correction actuel, le filtre de la sous-bande sélectionnée multiplié par la modification envisagée pour le poids relatif effectif de la sous-bande étudiée. On pourrait d'ailleurs étudier la pertinence de permettre un réglage de poids de sous-bande par pas fixe ou par pas fin et variable. *A priori*, il faudra mener plus de tests avec des sujets malentendants pour être en mesure de proposer un ou plusieurs algorithmes de correction progressive automatique, et, par suite, se résoudre à proposer une correction manuelle dans un premier temps.

Enfin, comme nous n'avons pas pu le faire dans le cadre du mémoire, il serait intéressant de procéder à des campagnes de tests mettant en œuvre trois découpages fréquentiel : un découpage adapté à un corpus de voix de référence ; un découpage adapté à une voix féminine choisie ; un découpage adapté à une voix masculine choisie. Cette démarche permettrait de voir s'il est ou non intéressant de proposer la détermination de découpages spécifiques adaptées par exemple aux voix des interlocuteurs de l'utilisateur s'avérant les plus difficiles à comprendre ou à un corpus d'enregistrements faisant intervenir ces proches. Ainsi, c'est la question de l'adaptation du ou des découpages fréquents à chaque utilisateur qui se pose puisque rien ne permet de supposer que les modalités de perception, qui plus est chez les malentendants, soient « universelles ». Ceci étant, rien n'empêcherait de commencer par proposer dans un premier temps, avec des prothèses auditives fonctionnelles utilisant l'analyse-resynthèse IDS en temps réel (perceptif), un jeu limité de découpages fréquentielles considérés comme potentiellement « universels ». voire d'envisager que la proposition de découpages fréquents adaptés relèvent d'une prestation complémentaire, ne serait-ce que là aussi dans un premier temps.

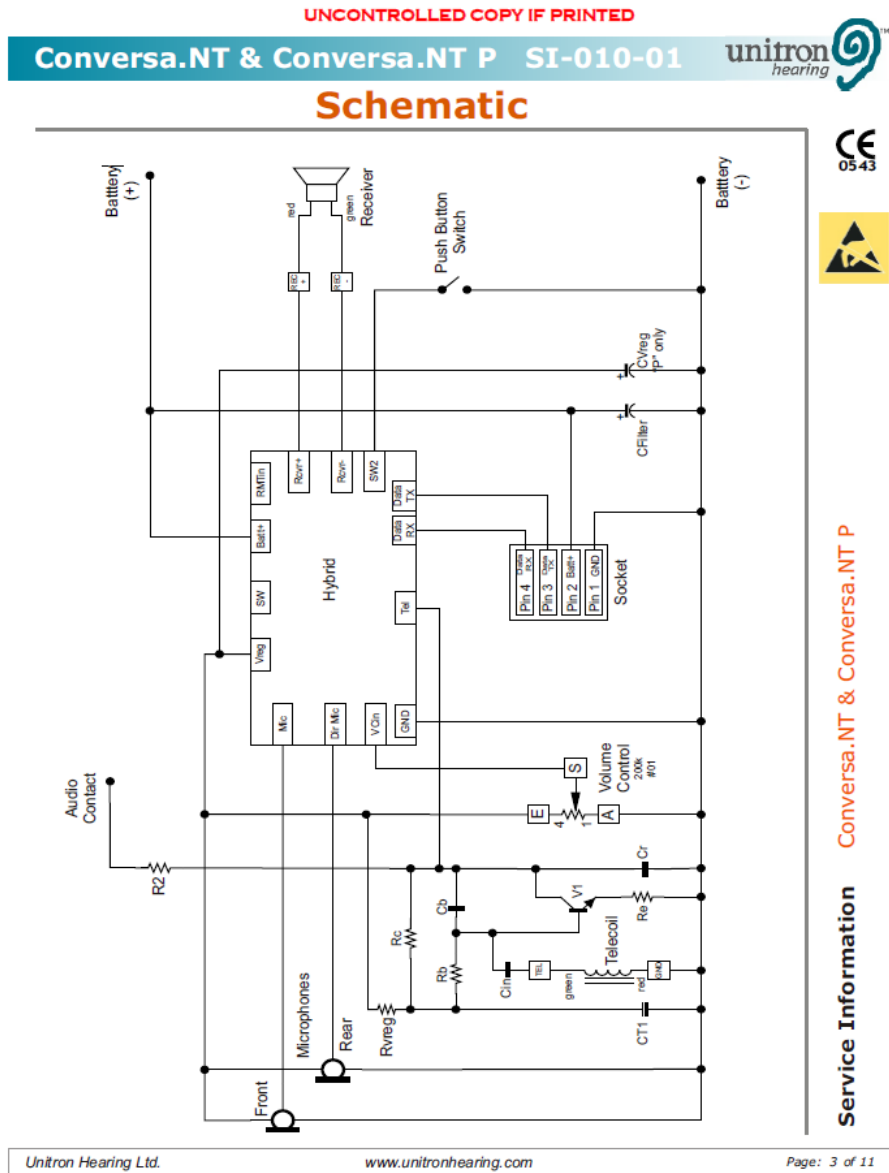
Références

- [1] LABRANDA J., *El Registro Sonoro*, San Antonio de los Baños, Ed. EICTV, 2009.
- [2] AURIOL B., *La Clef des sons*, préface de J-C Risset, Toulouse, Erès, 1991.
- [3] HARARI Y. N., *Sapiens : Une brève histoire de l'humanité* (Sapiens : A brief history of the Humankind, 2014), traduit de l'anglais par Pierre-Emmanuel Dauzat, Paris, Albin Michel.
- [4] ARISTOTE, *De l'Interprétation*, Organon II, tr. J. Tricot, Paris, Vrin, 1994.
- [5] CHERRY C., « Some experiments on the recognition of speech with one and two ears », *The Journal of the Acousical Society of America*, Vol. 25, n°5, septembre 1953, p.975-979.
- [6] RALLE T., *Usage de l'analyse IDS pour une meilleure compréhension de la parole dans le bruit*, Mémoire de fin d'études de l'ENS Louis Lumière sous la direction de Laurent Millot et Bernard Auriol, 2017.
- [7] DE KERVASDOUE J., HARTMANN L., « Revue de la littérature scientifique 2005-2015, Impact Économique du Déficit Auditif en France et dans les Pays Développés », Unsaf, 2016, [http ://www.unsaf.org/doc/Impact_Economique_Deficit_Auditif_en_France.pdf](http://www.unsaf.org/doc/Impact_Economique_Deficit_Auditif_en_France.pdf), consulté le 28 février 2019.
- [8] SERRIÈRE F., « Résultat du baromètre de l'audition Audio 2000 et Senior Strategic », Acuité, portail des décideurs de l'optique, mars 2014. [http ://www.acuite.fr.](http://www.acuite.fr.), consulté le 3 mars 2019.
- [9] « Prothèse auditive, audioprothèse (Infographie). Vous et votre audition », Que Choisir, septembre 2015, [https ://www.quechoisir.org](https://www.quechoisir.org), consulté le 27 mars 2019.
- [10] « Comprendre son audiogramme », La maison de l'appareil auditif, [https ://www.maison-appareil-auditif.com/comprendre-son-audiogramme/](https://www.maison-appareil-auditif.com/comprendre-son-audiogramme/), consulté le 25 mars 2019.
- [11] GELIS C., *Biophysique de l'environnement sonore*, Paris, Ellipses, 2002.
- [12] AURIOL B., *La Clef des sons*, préface de J-C Risset, Toulouse, Eres, 1991.
- [13] « Presbyacousie », Voyage au centre de l'audition, 2016, [http ://www.cochlea.eu/pathologie/presbyacousie](http://www.cochlea.eu/pathologie/presbyacousie), consulté le 2 avril 2019.

- [14] « Nouvelle enquête Ifop-JNA : Les oreilles des français sous pression. Quels impacts sur la santé », Journée Nationale de l’Audition, <https://www.journee-audition.org/les-actualites-de-l-association.html>, consulté le 10 avril 2019.
- [15] « Projections de population à l’horizon 2070 - Deux fois plus de personnes de 75 ans ou plus qu’en 2013 », Insee Première n° 1619, novembre 2016.
- [16] « Projections de population à l’horizon 2060 - Un tiers de la population âgé de plus de 60 ans », Insee Première n° 1320, octobre 2010.
- [17] BOUCCARA D., AVAN P., MOSNIER I., GRAYELI A. B., FERRARY É. & STERKERS O., « Réhabilitation auditive ». *M/S : médecine sciences*, 21, (2), 190–197, 2005.
- [18] DI CRISTINA J., « Introduction to hearing aids and important design considerations », Application Note 4691, Maxim, 2010.
- [19] KILLION C., « SNR Loss : "I Can Hear What People Say But I Can't *Understand* Them" », *The Hearing Review*, Vol.4, N°12, p.8,10,12&14, December 1997.
- [20] PLAPOUS C., *Traitements pour la réduction de bruit. Application à la communication parlée*. Signal and Image processing. Université Rennes 1, 2005.
- [21] LEIPP É., « L’intégrateur de densité spectrale (IDS) et ses applications en acoustique », *Bulletin du Groupe d’Acoustique Musicale (GAM)* n94, Laboratoire d’Acoustique Musicale, Université Paris 6, décembre 1977.
- [22] MILLOT L., *Traitement du signal audiovisuel*, Paris, Dunod, 2008.
- [23] « La télévision, média le plus consommé en France », Le Monde, mars 2013, <http://www.lemonde.fr>, consulté le 15 février 2019.
- [24] « Les seniors et l’audition », Ipsos, mars 2013, <http://www.ipsos.fr>, consulté le 15 février 2019.
- [25] DELERCE X., GAULT A., DELANDE J-B., « Technologies et traitement du signal Gestion de la parole en milieu bruyant par les aides auditives », *Les cahiers de l’audition*, N°2, p.9-15, 2014.
- [26] STAAB W., « *Hearing Aid Evolution : II* », Hearing Health & Technology Matters, 2014, <https://hearinghealthmatters.org/waynesworld/2014/hearing-aid-evolution-ii/>, consulté le 5 avril 2019.

- [27] « Recherches impliquant la personne humaine », Ministère des solidarités et de la Santé, <https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/recherche-et-innovation/recherches-impliquant-la-personne-humaine/>, consulté le 14 mars 2019.
- [28] « What is it? », BruteFIR, <https://www.ludd.ltu.se/~torger/brutefir.html#whatis>, consulté le 21 Avril 2019.
- [29] « Introduction », FFTW, <http://fftw.org/>, consulté le 16 Avril 2019.
- [30] « Faut-il avoir peur de l'intelligence artificielle », *Le Drenche*, N°16, p.8-11, 2019.
- [31] LAMBERT R., « Différence entre Intelligence Artificielle, Machine Learning et Deep Learning », Pensée Artificielle, Décembre 2018, <http://penseeartificielle.fr/difference-intelligence-artificielle-machine-learning-deep-learning/>, consulté le 28 avril 2019.
- [32] LAFON J.-C., *Le test phonétique et la mesure de l'audition*, Paris, Dunod, 1964.
- [33] DAENINCKX D., L'espoir en contrebande, *Coupe-Coupe*, Paris, Gallimard, 2013.
- [34] Phonak Audéo M, Technical Data,
https://www.phonakpro.com/content/datasheet_audeo_marvel_r.pdf, consulté le 5 Mai 2019.
- [35] GENY S., « Effet d'occlusion », Blog audioprothésiste, <https://www.blog-audioprothesiste.fr/effet-docclusion/>, consulté le 5 Mai 2019.

Annexe 1 : Schéma électronique du modèle de prothèse auditive Conversa.NT



Annexe 2 : Texte lu lors des scènes sonores constituant le corpus

La légende prétend qu'aussitôt décapité, Denis, premier évêque de Paris, prit sa tête dans ses mains, traversa ce qui allait devenir le Quartier latin, escalada la colline de Montmartre, puis redescendit droit sur un village qui bien des temps plus tard fut baptisé du nom de ce martyr chrétien : Saint-Denis. C'est à l'endroit même où la vie le quitta que fut édifiée une orgueilleuse basilique. Elle renferme les dépouilles des rois de France dont celle, ironie de l'histoire, d'un autre décapité célèbre, Louis le seizième. Si l'on consulte un plan du secteur, on constate que Denis, lesté de son chef, a foulé les terres du Cornillon et qu'il est probablement passé sur l'ellipse aujourd'hui dévolue au Stade de France. Peut-être même que pour reprendre, sinon son souffle, du moins ses forces, il a un instant posé sa tête dans l'herbe et que, mille ans plus tard, une puissance obscure a conduit géomètres et architectes à faire de cet emplacement le point d'engagement de la finale de la Coupe du monde ! On sait que les dernières décennies ne furent pas tendres pour le paysage de ma ville natale. Tous les minerais de France, tous les produits chimiques convergeaient sur ce coin de banlieue. Toutes les misères aussi. Un ciel lourd naissait en permanence des cheminées qui hérissaient le quartier Pleyel, le quartier du Landy. En lieu et place du stade s'élevaient d'autres enceintes arrondies, celles des gazomètres. Les voies du secteur avaient puisé leur nom dans leur ombre : rue du Gaz, impasse du Gaz. . . Le monde entier fournissait des bras aux industries dionysiennes. Au début du siècle, c'est par milliers que les jeunes désertaient les villages du Morbihan, du Morvan et de l'Aubrac pour venir peupler les ateliers des usines automobiles Delaunay- Belleville ou Hotchkiss. Les Bretons étaient les plus nombreux. Les plus rejetés aussi à cause de leur langue, étrangère d'apparence. Les registres policiers gardent la trace des chasses au faciès dont ils étaient victimes, le samedi soir, et un nom fut même inventé par un fonctionnaire pour les désigner : les bretonnades .

Plus tard, ce furent les Italiens qui prirent la relève, puis les cohortes espagnoles et républicaines, vaincues par les armées noires de Franco. Le quartier se transforma en petite Espagne où résonnèrent longtemps les accords de guitare rageurs de Paco Ibañez. D'autres exilés, affamés de pain ou de liberté, trouvèrent le chemin à leur tour. Portugais, Algériens, Maliens, Yougoslaves, s'installèrent par vagues successives dans les pauvres maisons laissées vacantes par ceux qui avaient enfin réussi à gagner le cœur des villes. Aujourd'hui, c'est d'un pas volontaire et le sourire aux lèvres que de tous les pays du monde on se presse vers Saint-Denis pour encourager les dieux du Stade de France, pour applaudir les vainqueurs de la Coupe. Toutes les couleurs, tous les drapeaux, toutes les musiques. Comme si les hommes s'étaient enfin aperçus qu'avant d'être le nom d'un martyr, Denis désignait le dieu grec du vin et de la vigne, le dieu de la fête : Dionysos.

DAENINCKX, Didier. L'espoir en contrebande, *Coupe-Coupe* , p. 111-113, Paris, Gallimard, 2013.

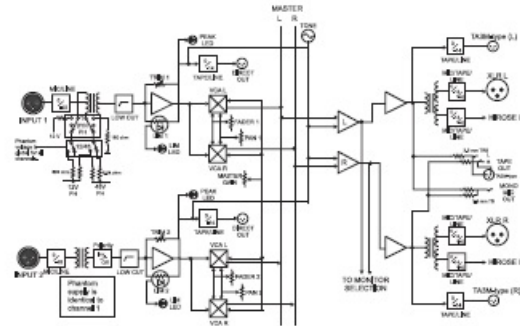
Annexe 3 : Documentation technique de la « mixette » *Sounddevices 442*



[Input Panel]



[Output Panel]



[Input/Output Block Diagram]

Specifications

Frequency Response:

- 20 Hz - 30 kHz, +0.2, -0.5 dB,
- -1 dB @ 5 Hz and 50 kHz typical

Equivalent Input Noise:

- -126 dBu (-128 dBV) maximum (22 Hz - 22 kHz bandwidth, flat filter, trim control fully up)

Dynamic Range:

- 115 dB minimum (trim fully down)

THD + Noise:

- 0.007% typical (1 kHz, +4 dBu at line out)
- 0.09% max (50 Hz - 20 kHz, +18 dBu at line out, fader fully up)

Output Noise:

- -100 dBu (-102 dBV) maximum (22 Hz - 22 kHz bandwidth, flat filter, master gain fully up, faders fully down)

High-Pass Filters:

- Sweepable 80 Hz to 240 Hz, 12 dB/oct @ 80 Hz, 6 dB/octave at 240 Hz

Powering:

- Internal ± 16 V (bi-polar) regulated audio rails
- 3.2-8-volt range internal batteries
- Isolated (floating) external DC input jack, 5-18-volt
- Locking 4-pin Hirose connector, pin-4 = (+), pin-1 = (-),

Metering:

- 40 segment, sunlight-viewable
- Selectable Peak, VU, or Peak (with or without peak hold) + VU ballistics

Input Clipping Level:

- 0 dBu minimum (trim control fully down)

Line Output Clipping Level: (1% THD):

- 20 dBu minimum
- 18 dBu minimum w/600 ohm load

Input Type:

- XLR mic - 2k ohm Z, transformer-balanced, 0 dBu max input
- XLR line - 16k ohm Z, transformer-balanced, +40 dBu max input

Output Type:

- XLR & multi-pin @ line - 85 ohm Z, transformer balanced, +20 dBu max output
- XLR & multi-pin @ mic - 150 ohm Z, transformer-balanced, +6 dBu max output
- TA3 line - 1k ohm Z, impedance-balanced, +20 dBu max output

Common Mode Rejection Ratio:

- 120 dB minimum at 80 Hz, mic input,
- 100 dB minimum at 10 kHz, mic input

L/R Gain Matching:

- Mic In to Line Out: ± 0.1 dB

Mic Powering (each mic selectable):

- No power applied, 12-volt phantom, 48-volt, 12-volt T-power

Input Limiters:

- Affects the output of the mic preamps only ("Trim" stage), +18 dBu threshold, 20:1 limiting ratio, 1 mS attack time, 200 mS release time.

Output Limiters:

- Threshold selectable from +4 dBu to +20 dBu, 1 dB steps, 20:1 limiting ratio, 1 mS attack time, 200 mS release time.

Temperature:

- Operating: -20°C to 60°C, 0 to 95% relative humidity; (non-condensing).
- Storage: -40°C to 85°C

Dimensions:

- 53 mm x 165 mm x 279 mm (H x W x D)
- 2.1" x 6.5 x 11"

Weight:

- 2.0 kg, (4.5 lbs) unpackaged with 4-AA batteries

Description Continued

The high-efficiency power circuitry runs the mixer from either four internal AA batteries or external 5-18 VDC.

The 442 and 442N, like all Sound Devices professional audio products, are designed to withstand the physical and environmental extremes of field production. Its compact mechanic construction strikes the perfect balance between access to all functions, and portability.

Discover more about Sound Devices products at www.sounddevices.com

Features, nomenclature, and specifications subject to change.
© 2009, Sound Devices, LLC

Annexe 4 : Documentation technique de l'enregistreur *Tascam* DR-100MkII

Données

Supports d'enregistrement

Cartes SD (64 Mo à 2 Go) et cartes SDHC (4 Go à 32 Go)

Formats d'enregistrement/lecture

WAV : 44,1/48/96 (mode HS) kHz, 16/24 bits
MP3 : 44,1/48 kHz, 32/64/96/128/192/256/320 kb/s

Nombre de canaux

2 canaux (stéréo)

Valeurs d'entrée/sortie

Entrée/sortie audio analogique

Connecteurs XLR (MIC/LINE 1) IN

(compatible avec l'alimentation fantôme)

Connecteurs : XLR-3-31

(1 : masse, 2 : point chaud, 3 : point froid)

Quand le sélecteur MIC/LINE 1 est réglé sur MIC

Impédance d'entrée : 2,1 k Ω

- Quand GAIN est réglé sur HIGH
Niveau d'entrée nominal : -70 dBu
Niveau d'entrée maximal : -54 dBu
- Quand GAIN est réglé sur MID
Niveau d'entrée nominal : -42 dBu
Niveau d'entrée maximal : -26 dBu
- Quand GAIN est réglé sur LOW
Niveau d'entrée nominal : -14 dBu
Niveau d'entrée maximal : +2 dBu

Quand le sélecteur MIC/LINE 1 est réglé sur LINE 1

Impédance d'entrée : 2,5 k Ω

Niveau d'entrée nominal : +4 dBu
Niveau d'entrée maximal : +24 dBu

Prise LINE 2 IN

Connecteur : mini-jack 3,5 mm stéréo

Impédance d'entrée : 22 k Ω ou plus

Niveau d'entrée nominal : -10 dBV

Niveau d'entrée maximal : +6 dBV

Prise LINE OUT

Connecteur : mini-jack 3,5 mm stéréo

Impédance de sortie : 200 Ω

XLR LINE 1 IN :

Niveau de sortie nominal : -14 dBV

Niveau de sortie maximal : +6 dBV

Autres :

Niveau de sortie nominal : -10 dBV

Niveau de sortie maximal : +6 dBV

Prise Ω

Connecteur : mini-jack 3,5 mm stéréo

Puissance de sortie maximale : 25 mW + 25 mW
(avec casque d'impédance 32 Ω)

Haut-parleur intégré

Sortie : 0,4 W (mono)

Données d'entrée audio numérique

Prise d'entrée DIGITAL IN

Connecteur : mini-jack (2,5 mm) 3 points (TRS) (en utilisant le câble de conversion prévu pour cette unité)

Format : IEC60958-3 (S/PDIF)

Valeurs d'entrée/sortie de commande

Port USB

Connecteur : type mini-B

Format : USB 2.0 HIGH SPEED, classe de périphériques de stockage de masse

Prise REMOTE

Connecteur : mini-jack (2,5 mm) 3 points (TRS)

Performances audio

Réponse en fréquence

20 Hz-20 kHz, +1/-3 dB

(EXT IN vers LINE OUT, F ϕ 44,1 kHz, JEITA)

20 Hz-22 kHz, +1/-3 dB

(EXT IN vers LINE OUT, F ϕ 48 kHz, JEITA)

20 Hz-40 kHz, +1/-3 dB

(EXT IN vers LINE OUT, F ϕ 96 kHz, JEITA)

Distorsion

Moins de 0,01 %

(XLR LINE 1 IN vers LINE OUT, F ϕ 44,1 kHz, entrée +24 dBu, JEITA)

Rapport signal/bruit

92 dB ou plus

(XLR LINE 1 IN vers LINE OUT, F ϕ 44,1 kHz, entrée +24 dBu, JEITA)

Note) JEITA : conformément à la norme JEITA CP-2150

Configurations informatiques requises pour les ordinateurs connectés

Recherchez sur le site web TASCAM (<http://tascam.com/>) des informations récentes sur la compatibilité des derniers systèmes d'exploitation.

Windows

Pentium 300 MHz ou plus rapide

128 Mo ou plus de mémoire

Port USB (USB 2.0)

Macintosh

Power PC 266 MHz, iMac, G3, G4 ou mieux

64 Mo ou plus de mémoire

Annexe 5 : Gradients IDS des variations globales de corrections

ids global variations: Gare_2_txt_F22fra_f_01 VS Gare_2_txt_F22fra_f_02
subject: num4_F_80

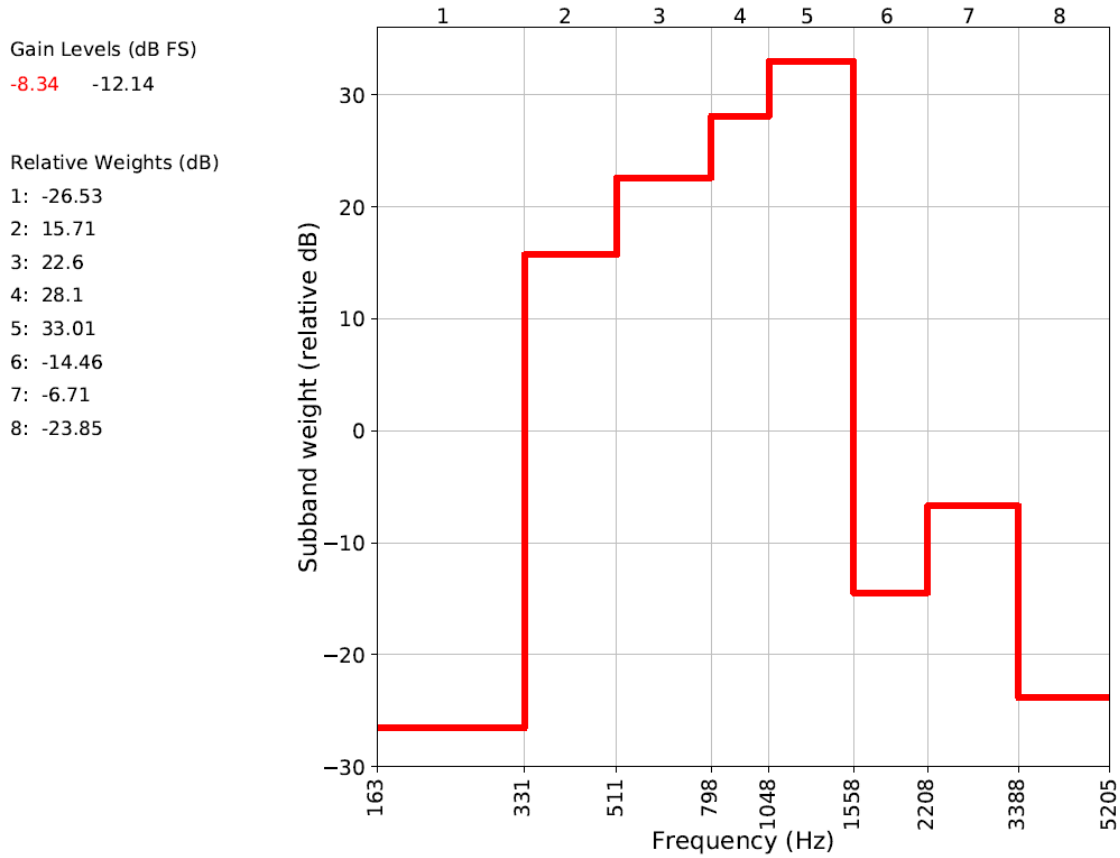


FIGURE 13 – Gradient IDS des variations globales de corrections du sujet n°4 entre la première et deuxième moitié de la scène sonore Gare_2_txt_F22fra_f.

ids global variations: Gare_2_txt_F22fra_f_01 VS Gare_2_txt_F22fra_f_02
 subject: num5_F_78

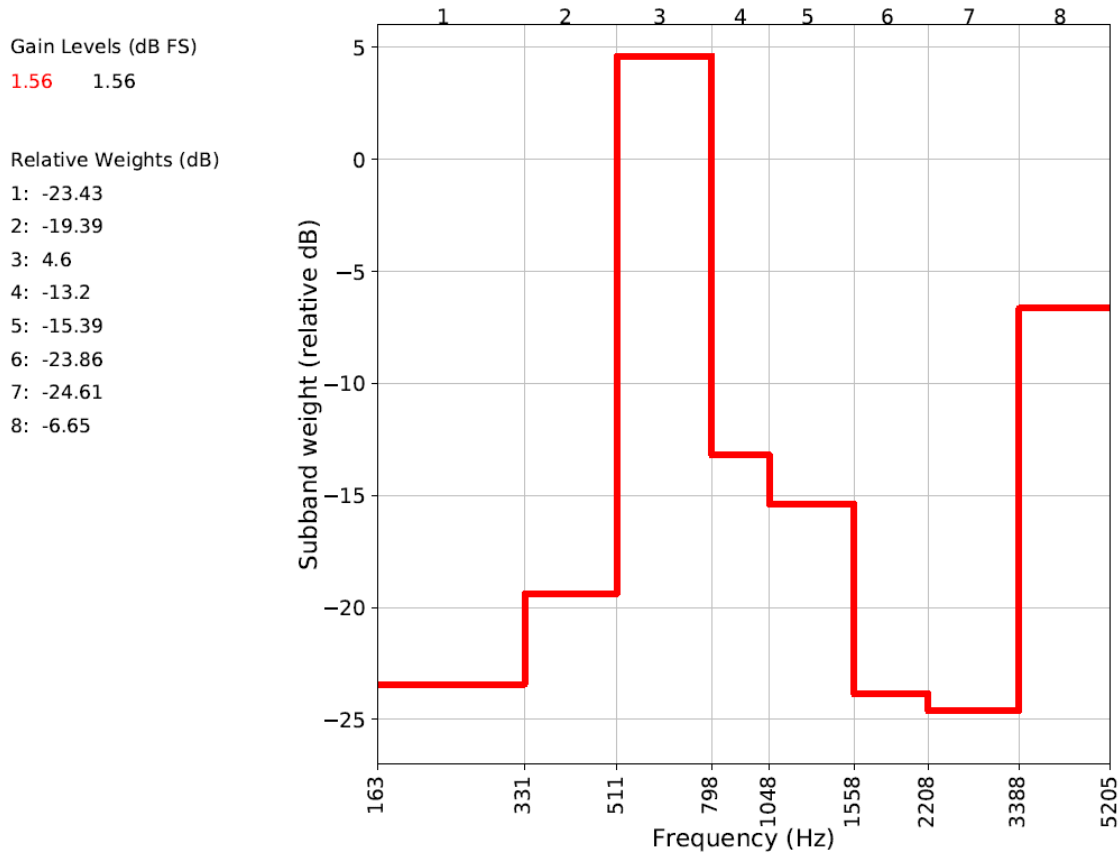


FIGURE 14 – Gradient IDS des variations globales de corrections du sujet n°5 entre la première et deuxième moitié de la scène sonore Gare_2_txt_F22fra_f.

ids global variations: Gare_2_txt_F22fra_f_01 VS Gare_2_txt_F22fra_f_02
 subject: num6_F_71

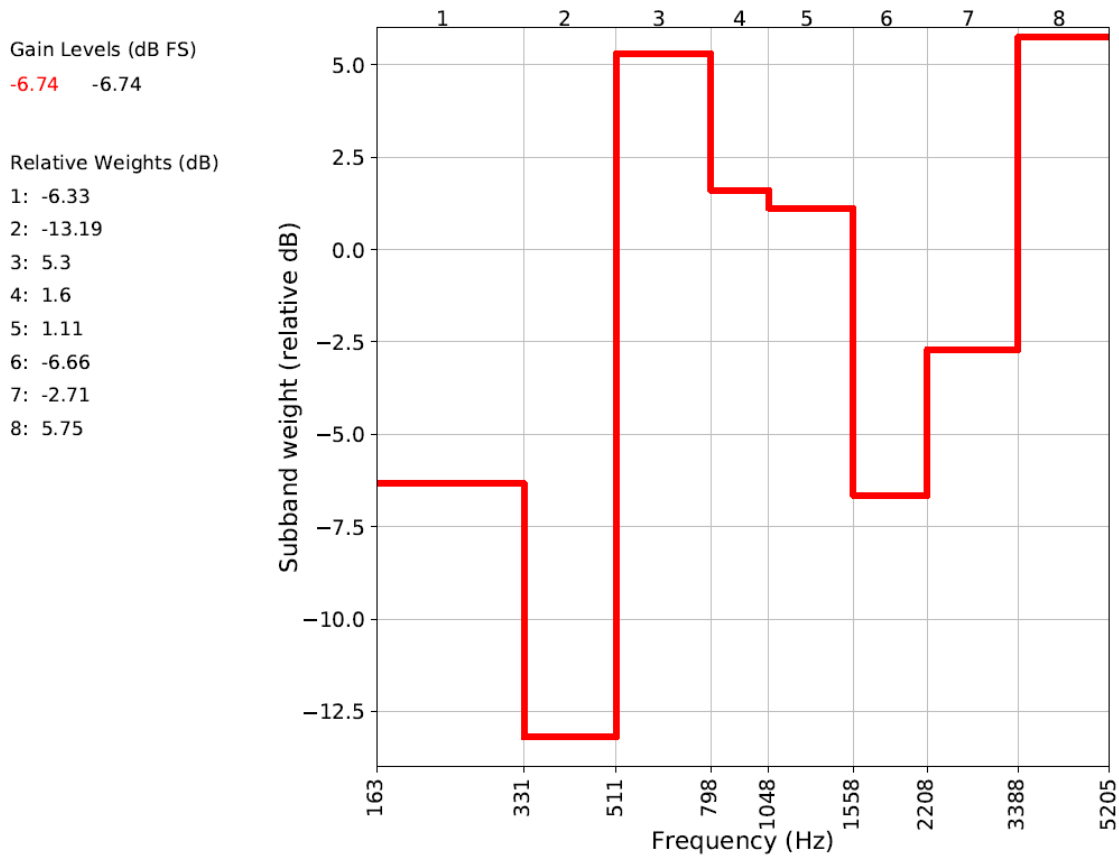


FIGURE 15 – Gradient IDS des variations globales de corrections du sujet n°6 entre la première et deuxième moitié de la scène sonore Gare_2_txt_F22fra_f.

ids global variations: Metro_1_txt_F22fra_n_01 VS Metro_1_txt_F22fra_n_02
 subject: num2_F_51

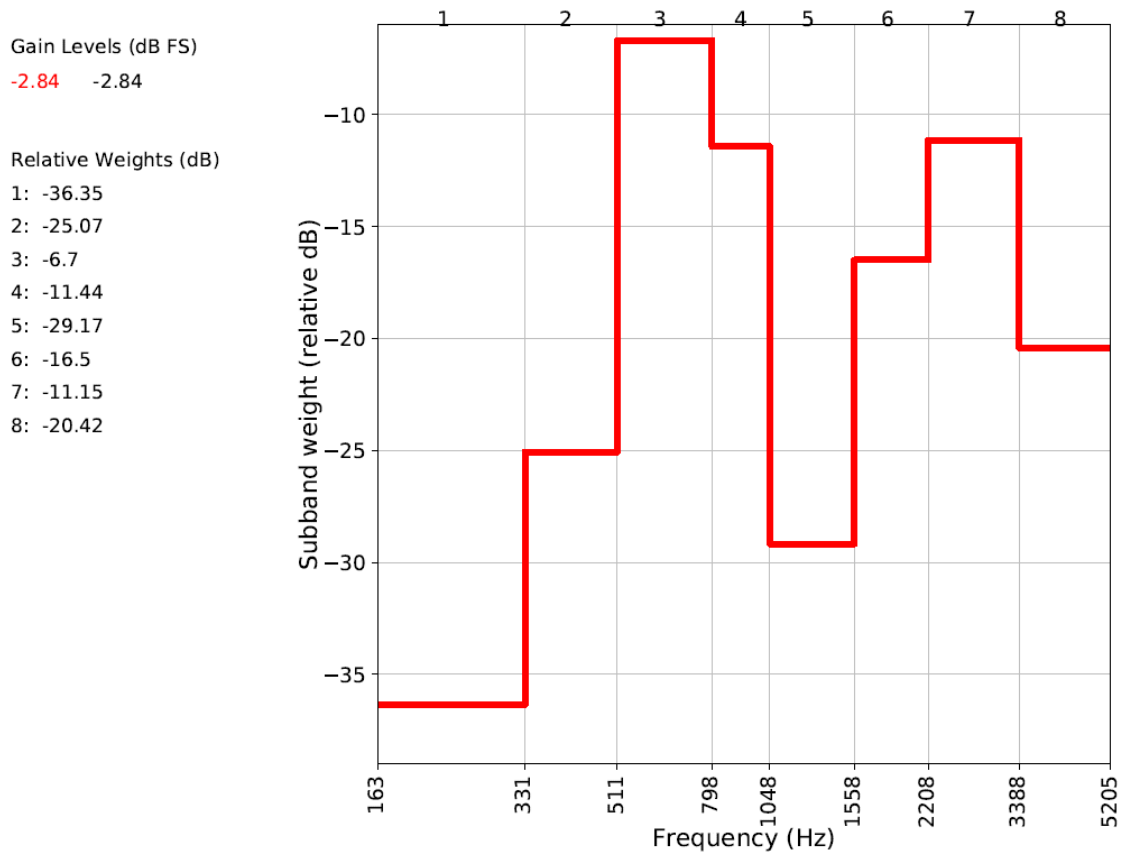


FIGURE 16 – Gradient IDS des variations globales de corrections du sujet n°2 entre la première et deuxième moitié de la scène sonore Metro_1_txt_F22fra_n.

ids global variations: Metro_1_txt_F22fra_n_01 VS Metro_1_txt_F22fra_n_02
 subject: num3_F_75

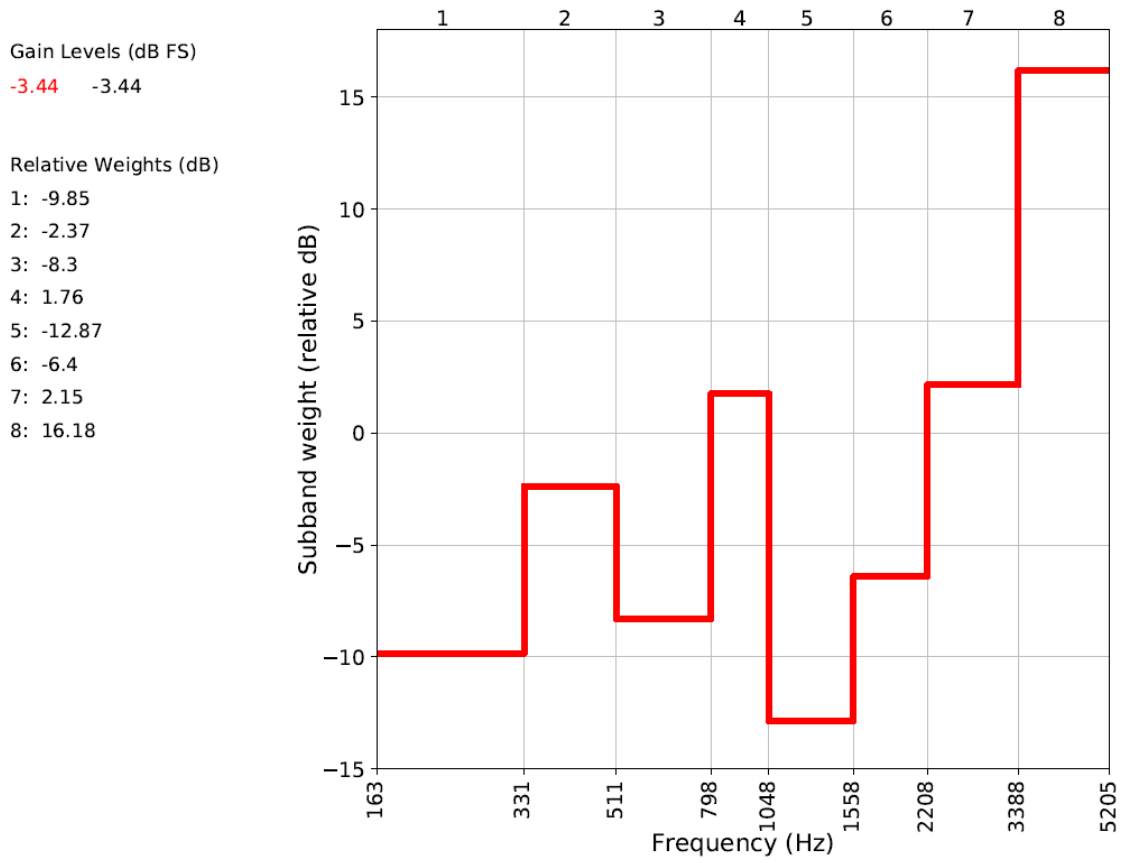


FIGURE 17 – Gradient IDS des variations globales de corrections du sujet n°3 entre la première et deuxième moitié de la scène sonore Metro_1_txt_F22fra_n.

ids global variations: Metro_1_txt_F22fra_n_01 VS Metro_1_txt_F22fra_n_02
 subject: num5_F_78

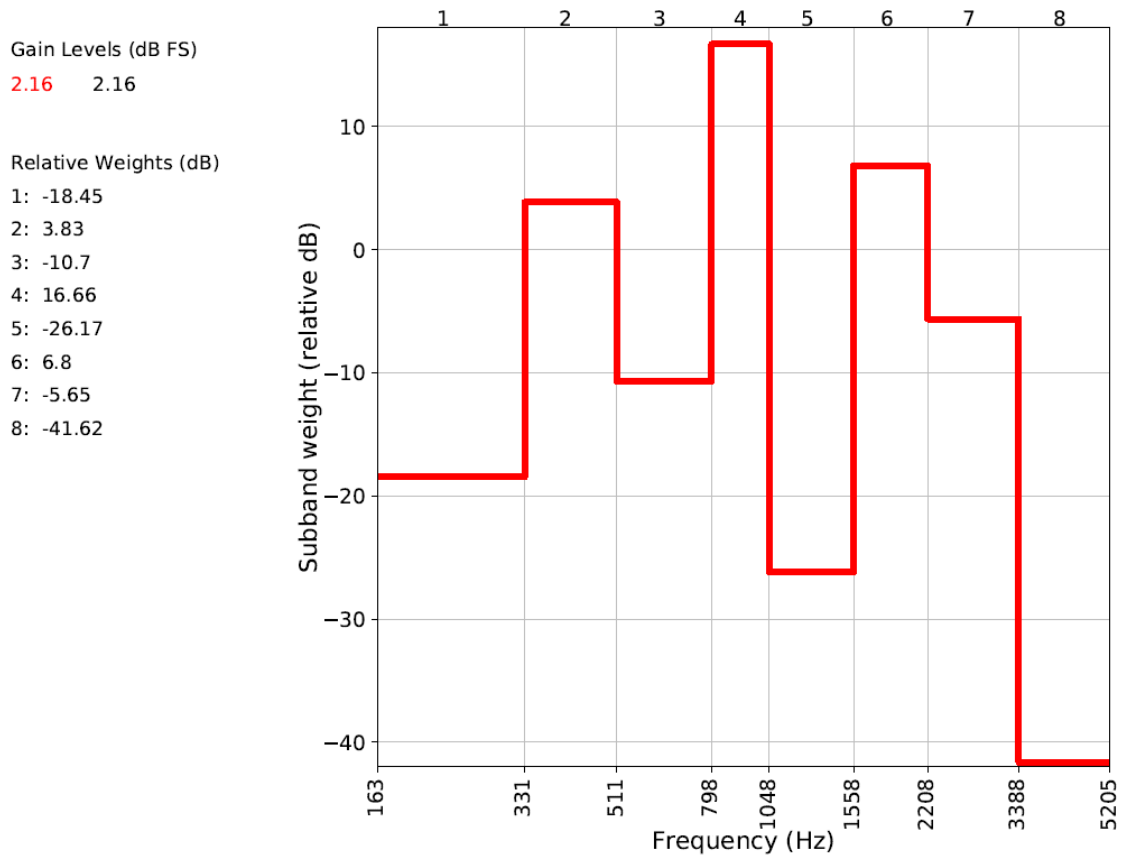


FIGURE 18 – Gradient IDS des variations globales de corrections du sujet n°5 entre la première et deuxième moitié de la scène sonore Metro_1_txt_F22fra_n.

Annexe 6 : Profils IDS des comparaisons

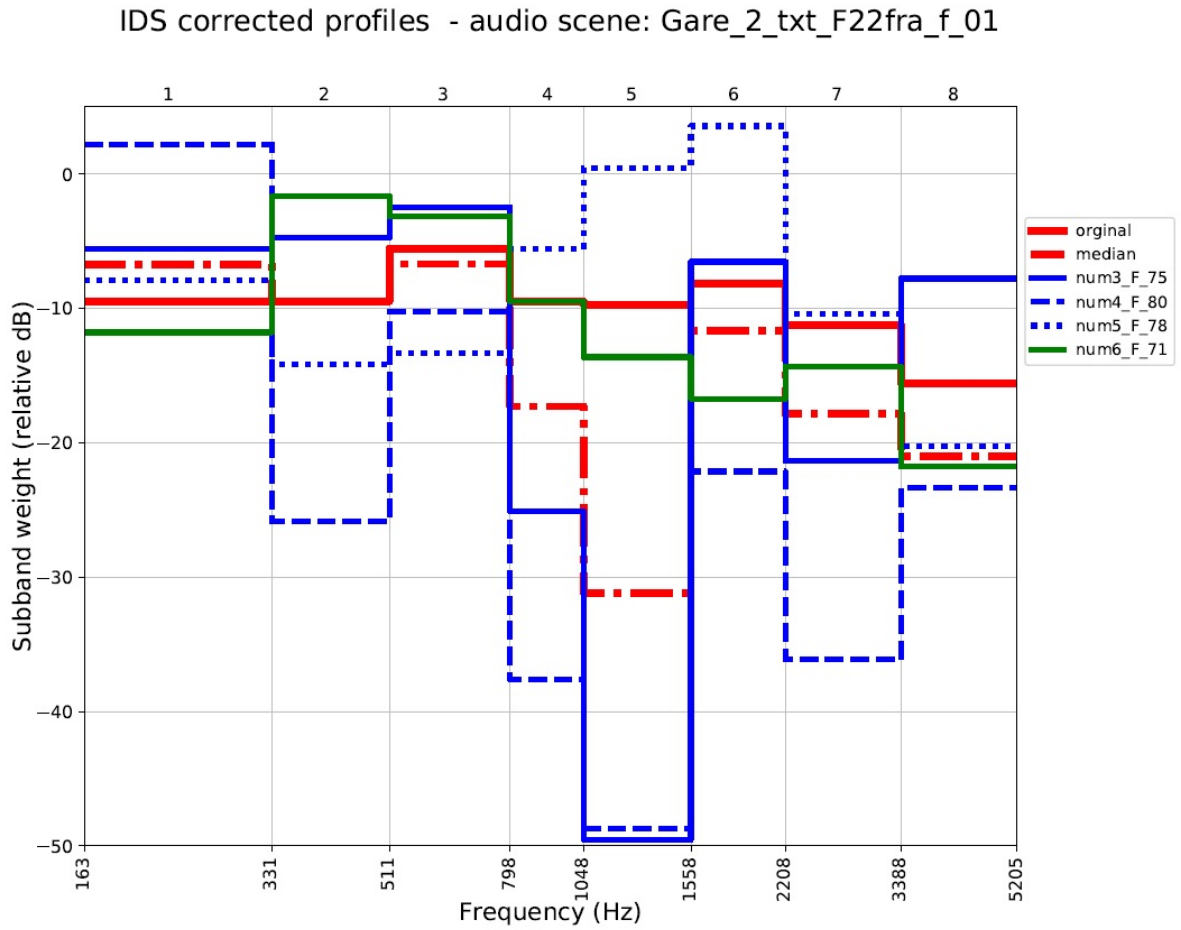


FIGURE 19 – Comparaisons des profils IDS des corrections apportées par 4 sujets pour la scène sonore Gare_2_txt_F22fra_f_01.

IDS corrected profiles - audio scene: Gare_2_txt_F22fra_f_02

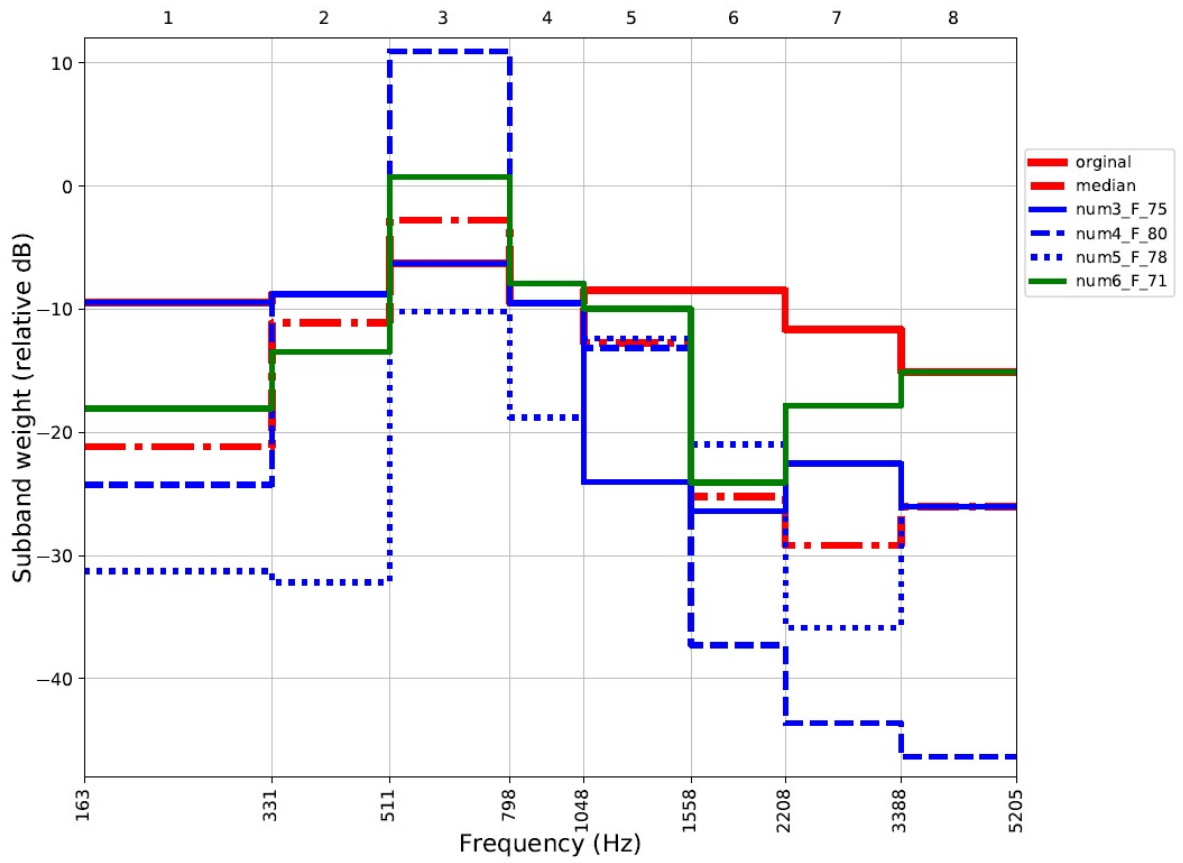


FIGURE 20 – Comparaisons des profils IDS des corrections apportées par 4 sujets pour la scène sonore Gare_2_txt_F22fra_f_02.

IDS corrected profiles - audio scene: Gare_3_txt_H24fra_n_02

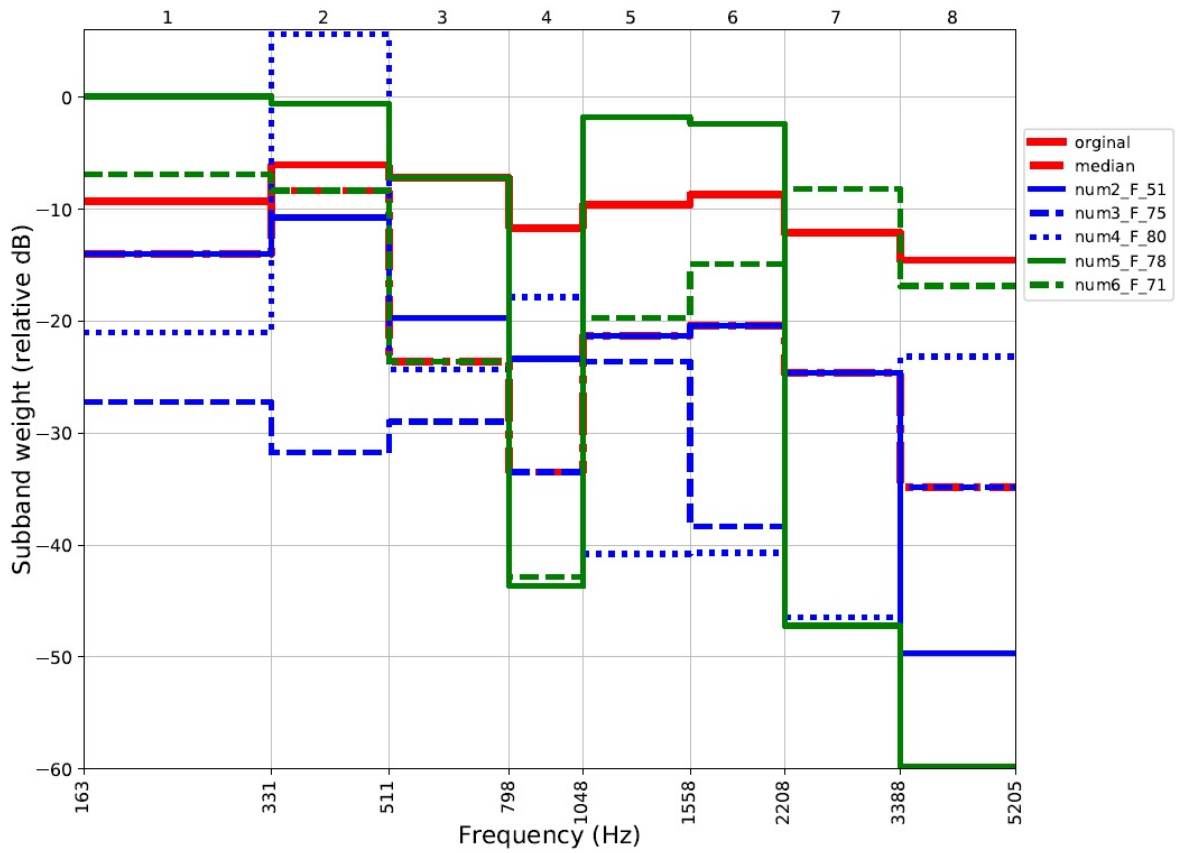


FIGURE 21 – Comparaisons des profils IDS des corrections apportées par 5 sujets pour la scène sonore Gare_3_txt_H24fra_n_02.

IDS corrected profiles - audio scene: Gare_4_txt_H24fra_f_01

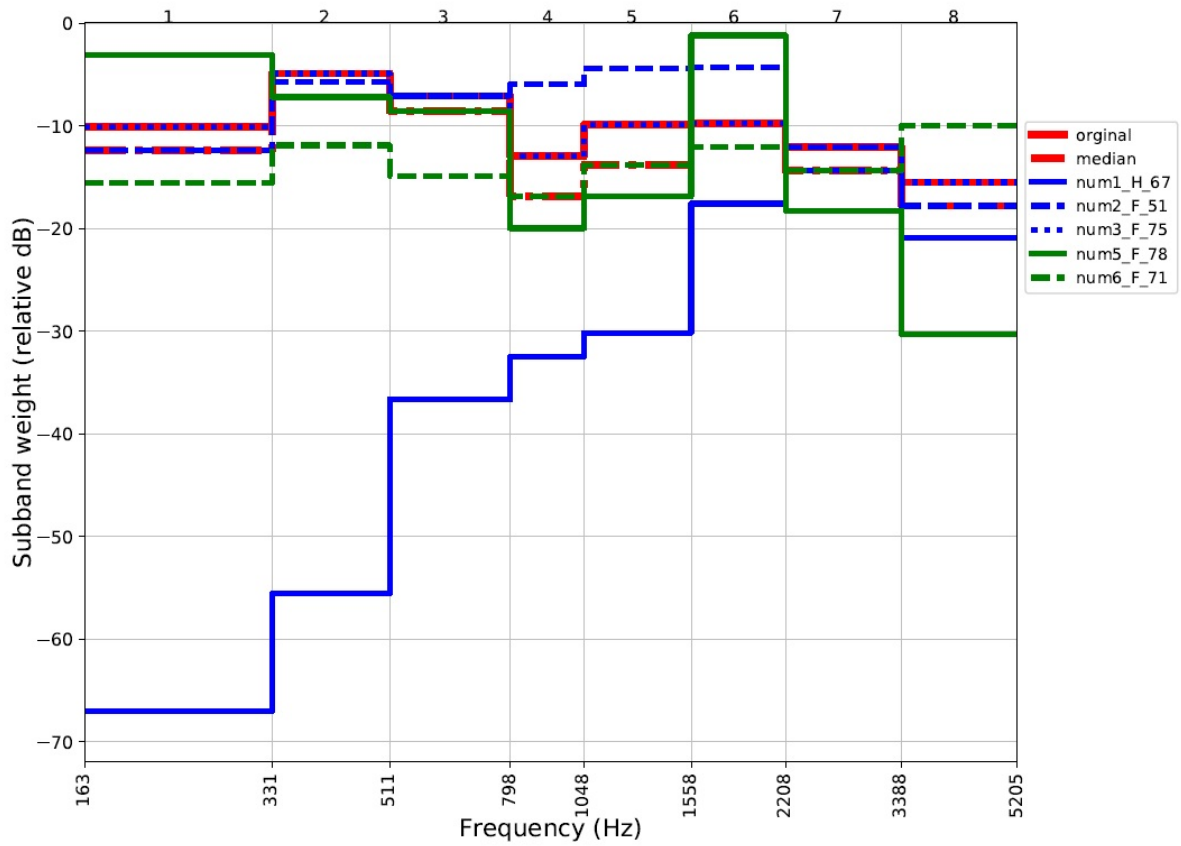


FIGURE 22 – Comparaisons des profils IDS des corrections apportées par 5 sujets pour la scène sonore Gare_4_txt_H24fra_f_01.

IDS corrected profiles - audio scene: Marche_1_txt_F22fra_n_01

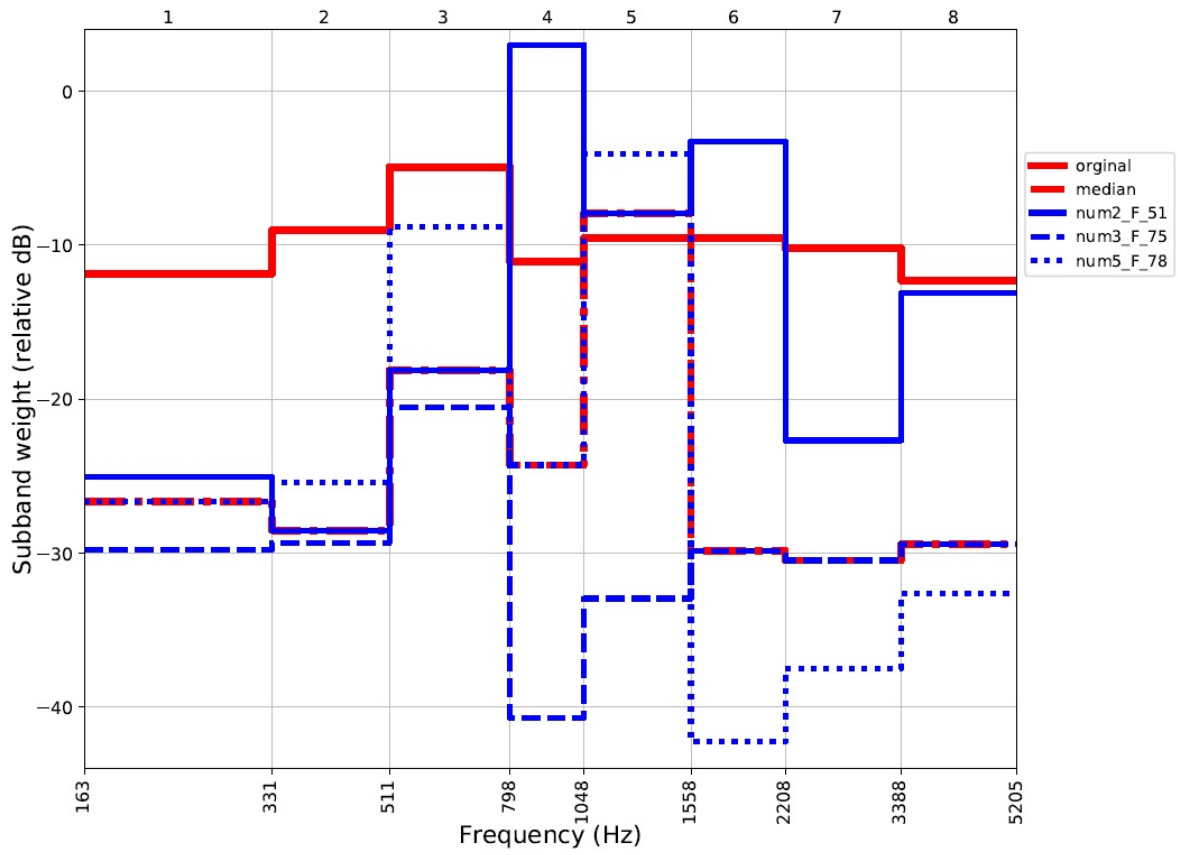


FIGURE 23 – Comparaisons des profils IDS des corrections apportées par 3 sujets pour la scène sonore Marche_1_txt_F22fra_n_01.

IDS corrected profiles - audio scene: Metro_1_txt_F22fra_n_01

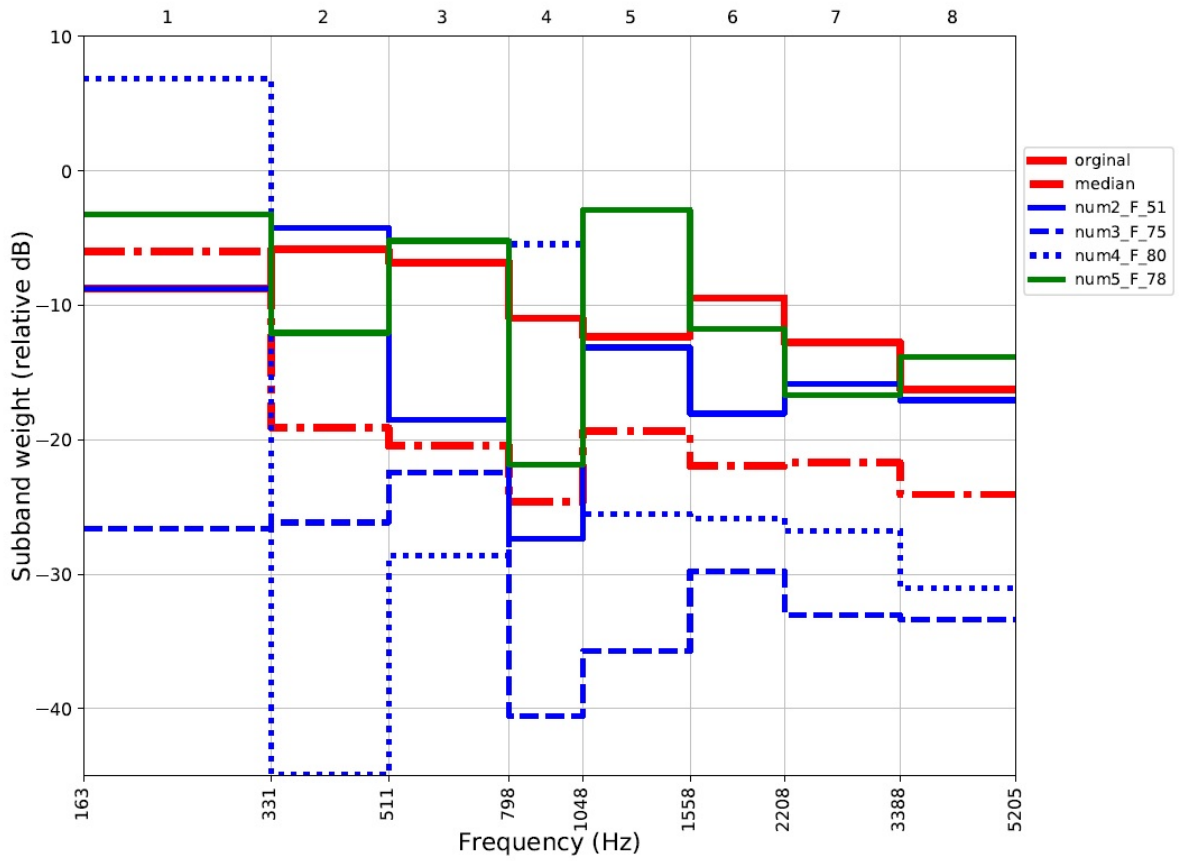


FIGURE 24 – Comparaisons des profils IDS des corrections apportées par 4 sujets pour la scène sonore Metro_1_txt_F22fra_n_01.

IDS corrected profiles - audio scene: Metro_1_txt_F22fra_n_02

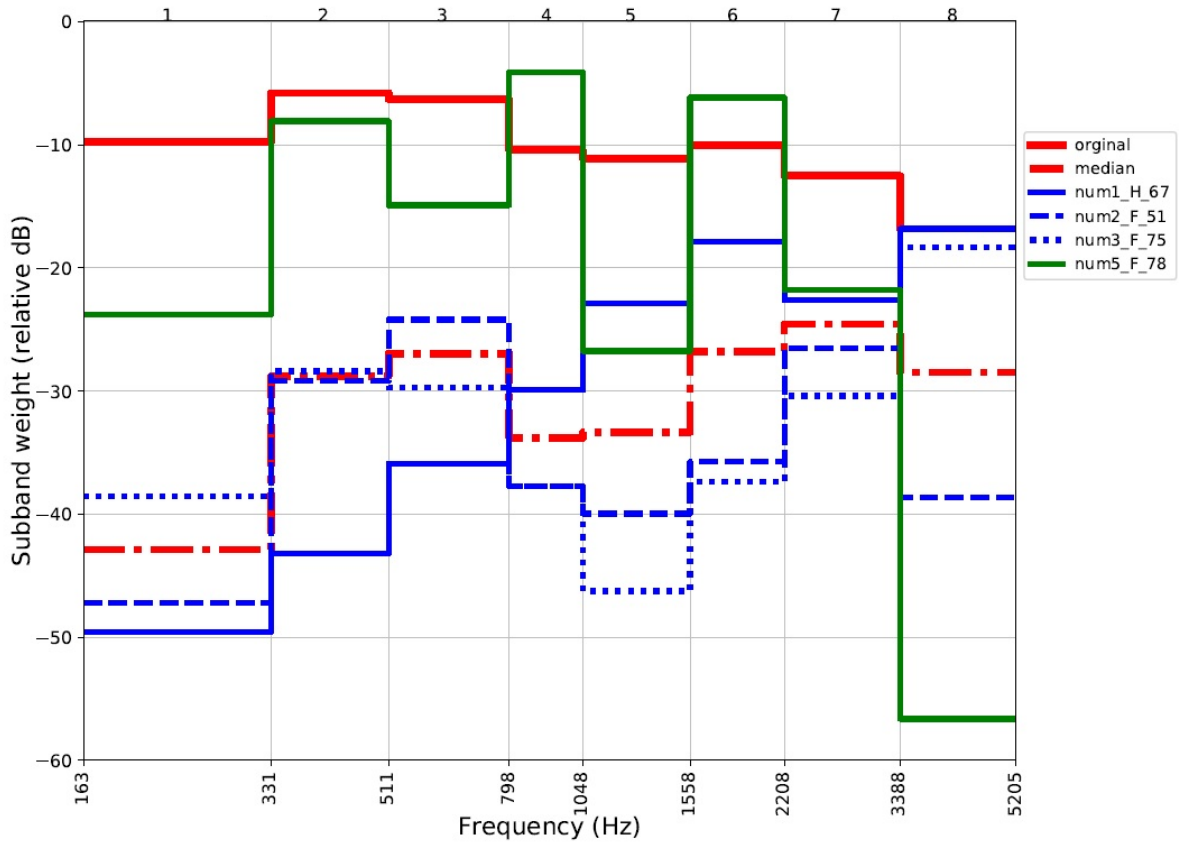


FIGURE 25 – Comparaisons des profils IDS des corrections apportées par 4 sujets pour la scène sonore Metro_1_txt_F22fra_n_02.