

Chapitre 5

Tests exploratoires d'un découpage perceptif adapté à la parole

1. L'INTÉGRATEUR DE DENSITÉ SPECTRALE

1.1. LE DECOUPAGE IDS

Le découpage IDS tel que décrit par Émile Leipp dans son bulletin du GAM¹ émane de la question de « qualité sonore² » des instruments de musique posée par le « jury de concours de sonorité³ ». Cette notion de sonorité renvoie à celle du timbre musical que l'on définit généralement comme suit : le timbre est ce qui différencie, par exemple, une note jouée par un violon de cette même note jouée par un piano ou encore une flûte.

Les lois de la physique (acoustique musicale) sont assez bien connues de nos jours pour expliquer que la notion de timbre est liée à l'organisation fréquentielle d'un signal musical ou sonore. Or, si l'on peut aisément distinguer le son d'un violon de celui d'un piano, qu'en est-il du son d'un stradivarius par rapport à celui d'un violon plus commun ? Comment reconnaître celui qui a une meilleure sonorité ou un meilleur timbre comme le ferait un violoniste ou un luthier ? Ces derniers connaîtraient très certainement des difficultés pour expliquer cette différence à une personne non musicienne ou peu familière de tels objets.

¹ Émile LEIPP, *bulletin du Groupe d'Acoustique Musicale (GAM) n° 94*, Laboratoire d'Acoustique Musicale, Université Paris 6, décembre 1977.

² *Ibid*, p. 1.

³ *Ibid*, p. 1.

C'est donc en réponse à ce problème d'objectivité que Leipp propose en 1977 un moyen d'opérer une classification des sons instrumentaux en les cartographiant, en quelque sorte, grâce à l'intégrateur de densité spectrale (IDS).

Cet appareil, comme son nom l'indique, procède à l'intégration temporelle (ou moyenne dans le temps) des données concernant la répartition de l'énergie spectrale d'un son.

Cette énergie spectrale s'étend, comme nous le savons, de 20 Hz (basses fréquences, sons graves) à 20 kHz (hautes fréquences, sons aigus), correspondant à notre perception auditive.

L'IDS, proposé par Leipp et construit par Sapaly, propose de découper cet intervalle, en huit bandes de fréquences, tout comme l'arc-en-ciel décompose en huit bandes colorées la lumière⁴.

Pour définir les limites de ces huit bandes de fréquences, Leipp procède de la façon suivante :

« À un auditoire de 30 sujets, par exemple, on fait écouter la Grande Valse (ou telle autre oeuvre bien choisie). On procède alors systématiquement de la façon suivante. On coupe de temps à autre, en cours d'audition, la bande la plus grave, en réglant le filtre de réjection entre 50 et 75 Hz par exemple. Si personne ne réagit, on élargit cette bande, par exemple de 50 à 100, à 150, à 200 Hz, jusqu'au moment où tous les sujets (à peu près) réagissent : "il n'y a plus de basses" ! Si la fréquence de coupure aiguë était alors de 200 Hz par exemple, on appellera cette bande de 50 à 200 Hz la "bande des basses". »

De cette façon, en notant à chaque fois les fréquences marquant les changements perceptifs de transition, Leipp obtient le découpage suivant :

- la bande des "basses" : de 50 Hz à 200 Hz ;
- la bande "grave" : de 200 Hz à 400 Hz ;
- la bande "medium grave" : de 400 Hz à 800 Hz ;
- la bande "medium" : de 800 Hz à 1,2 kHz ;
- la bande "medium aigu" : de 1,2 kHz à 1,8 kHz ;
- la bande "aigu" : de 1,8 kHz à 3 kHz ;
- la bande "sur-aigu" : de 3 kHz à 6 kHz ;
- la bande "stridence" : de 6 kHz à 15 kHz.

⁴ exemple qu'utilise Leipp dans son *bulletin du GAM*, p. 5.

Ainsi construit, L'IDS proposé par Émile Leipp permet de caractériser n'importe quel son en lui attribuant un « portrait » de la répartition de son énergie (ou densité) à l'intérieur de chacune des huit bandes décrites plus haut.

Un instrument de musique aura son « portrait » le distinguant d'un autre. Mais cela peut s'étendre également à des œuvres musicales : ainsi la 9^e symphonie de Beethoven aura son « portrait » IDS caractéristique.

Enfin, et c'est le point qui nous importe particulièrement ici, l'IDS peut caractériser la voix humaine. On peut alors observer, comme Leipp l'a fait en 1977, que certaines bandes constituant le portrait IDS d'une personne sont immuables quelle que soit la modification que cette personne donne à sa voix.

Bien que fortement inspirés par la version IDS proposée par Leipp, la version de l'IDS que nous utilisons dans ce mémoire, diffère de ce que nous venons de décrire. Les raisons de ces modifications trouvent leur explication dans les limites du dispositif de 1977. Après avoir décrit ces limites, nous expliquons, dans ce qui suit, quelles sont ces modifications et ce qu'elles apportent.

1.2. IDS NUMERIQUE ET DECOUPAGE ETENDU

a) Les limites de l'IDS proposé par Emile Leipp

L'analyseur IDS proposé par Emile Leipp à la fin des années 70 repose sur la technologie analogique et est donc contraint par les possibilités techniques de l'époque concernant les dispositifs de captation, d'affichage et surtout de filtrage. Pour autant, le principe initial de l'analyse IDS a été conservé dans la version numérique, proposée par Laurent Millot^{5,6}.

⁵ Laurent Millot, Mohammed ElIiq et Gérard Pelé, Some clues to build a sound analysis relevant to hearing, *116th Convention of the Audio Engineering Society, Berlin*, Allemagne, 8-11 mai 2004, Paper 6041.

⁶ Laurent Millot et Gérard Pelé, An objective and subjective alternative audio sounds and scenes analysis: the IDS, *International Symposium on Musical Acoustics (ISMA 2007)*, Barcelone, Espagne, 9-12 Septembre 2007, ISBN : 978-84-934142-1-4.

Ce principe consiste à opérer un filtrage du signal analysé suivant un jeu de filtres passe-bande et à calculer l'énergie relative associée à chacune de ces sous-bandes : on obtient ainsi le poids relatif de chacune de ces sous-bandes comme rapport des énergies cumulées respectivement associées à la sous-bande étudiée et au signal analysé.

Ensuite, dans sa version originelle analogique, l'analyseur IDS propose par le biais d'indicateurs graphiques de suivre l'évolution au cours du temps du poids relatif de chacune des sous-bandes, donnée en pourcentages de l'énergie globale, pour aboutir à la répartition globale, ou balance spectrale globale, correspondant aux poids relatifs finaux pour chacune des bandes.

La version analogique compte 8 sous-bandes de fréquences de coupure à -6 dB : 50-200~Hz, 200-400 Hz, 400-800 Hz, 800-1200 Hz, 1.2-1.8 kHz, 1.8-3 kHz, 3-6 kHz et 6-15~kHz. On peut noter que ce découpage, obtenu à partir de tests utilisant des stimuli tels que de la musique, de la voix chantée ou de la voix parlée, et, réunissant des sujets ayant l'habitude d'écouter (musiciens, ingénieurs du son, acousticiens), n'est pas du tout régulier du point de vue de la progression d'une sous-bande à l'autre. Mais, ce découpage résulte d'expérimentations avec des stimuli « complexes » et est, donc, supposé assez bien relié avec la perception auditive « européenne » (du fait des stimuli musicaux et de voix chantée).

Les filtres utilisés sont des filtres de Butterworth analogiques avec une pente de -12 dB par octave.

b) L'extension de l'analyseur IDS numérique, proposée par Laurent Millot

L'extension proposée par Laurent Millot s'appuie sur les possibilités offertes par la technologie numérique que ce soit en termes de possibilité de filtrage, de captation ou encore de stockage au format numérique.

La première modification tient à la nature des filtres qui sont des filtres RIF à phase nulle, ce qui permet de s'affranchir de toute distorsion d'analyse. De plus, ces filtres comptent un nombre conséquent de coefficients (au moins 8191) et proposent donc une atténuation rapide supérieure à 80 dB.

La deuxième modification introduite concerne le découpage puisque les fréquences correspondent maintenant à des fréquences de coupure à -3 dB et que deux sous-bandes ont été ajoutées :

- la sous-bande 0-50 Hz car cette version numérique de l'analyseur a aussi été pensée pour analyser des sons et des phénomènes acoustiques au niveau voire à l'intérieur des sources (phénomène potentiellement très basse fréquence) ;
- la sous-bande allant de 15 kHz à la fréquence de Nyquist (moitié de la fréquence d'échantillonnage) pour garder une trace de ce qui peut exister potentiellement au-delà de 15 kHz dans les captations, voire pour certains instruments de musique (l'harmonica diatonique, par exemple, dont les harmoniques émergent jusqu'à 20 kHz suivant les notes jouées)⁷.

La troisième modification concerne la présentation des données puisque les poids relatifs ne sont plus donnés en pourcentages de l'énergie globale, mais, en décibels relatifs calculés comme le rapport de l'énergie cumulée pour la bande étudiée et de l'énergie cumulée pour le signal complet, ce qui revient à choisir l'énergie cumulée dans le signal comme référence à 0 dB.

La présentation des balances spectrales, encore appelées portraits IDS en référence aux portraits des phases, avec des poids en dB relatifs, permet de mettre en évidence des influences potentiellement masquées avec une échelle linéaire (pourcentages) et d'envisager de comparer deux portraits IDS par simple soustraction des deux portraits.

La quatrième modification tient aux différentes possibilités de re-synthétiser le signal analysé. Il est possible, en effet, d'opérer la re-synthèse partielle du signal en choisissant d'écouter soit une seule sous-bande du signal (par exemple la 3^e sous-bande allant de la fréquence de 200 Hz à celle de 400 Hz), soit une sélection pertinente de plusieurs sous-bandes. De partielle, la re-synthèse peut ainsi peu à peu redevenir globale.

Enfin, on peut écouter les approximations ainsi réalisées, en basculant à la volée entre le signal complet, la re-synthèse partielle et les sous-bandes qui ne sont pas prises en compte pour la re-synthèse.

Ainsi, les analyses de sons ne sont plus limitées au tracé de graphiques puisqu'elles deviennent aussi sonores. L'ajout de l'écoute des éléments d'analyse permet ainsi d'étudier le

⁷ Il serait d'ailleurs intéressant d'étudier s'il est nécessaire de redécouper cette sous-bande lorsque l'on utilise des fréquences d'échantillonnage supérieures à 48 kHz.

poids pour l'audition des différentes sous-bandes, et, la comparaison de ces poids « auditifs » (ou perceptifs) avec les poids objectifs calculés lors de l'analyse, ce qui peut révéler des résultats inattendus⁸.

Une autre modification introduite est la possibilité d'enregistrer et de conserver les signaux de sous-bandes résultant de l'analyse, au cas où l'on souhaite pouvoir travailler sur la re-synthèse du son analysé.

Une autre modification concerne l'extension, pour le moment, de l'analyse à des scènes sonores ou des enregistrements stéréophoniques, dans l'optique d'étudier la composition de ces scènes, c'est-à-dire de déterminer le nombre de sources émergentes, leur position voire de les suivre, au cours du temps, dans leur mouvement⁹.

Enfin, compte tenu de l'algorithme utilisé pour faire les calculs, on peut envisager d'utiliser un découpage adapté à la situation testée.

Le découpage en octaves de fréquences centrales les octaves et sous-octaves de 1 kHz fait, par exemple, partie des découpages proposés dans la version actuelle.

Mais, d'autres découpages sont à l'étude pour, notamment, analyser des signaux issus de mesures pour des plateformes de posturologie, en lien avec un ostéopathe.

Ceci permet de souligner que l'analyseur IDS n'est en rien limité à une utilisation strictement dans le domaine audio avec un seul découpage en sous-bandes. D'ailleurs, l'extension au cas des images photographiques et aux vidéos est également à l'étude.

Par contre, l'évolution temporelle ou constitution progressive du portrait IDS n'est pas encore disponible car les choix, concernant les données à afficher pour une évolution temporelle, ne sont pas arrêtés, pas plus que la question de comment les afficher « ergonomiquement » et, dans le même temps de manière informative. Mais, l'idée d'observer l'évolution au cours du temps du profil IDS n'est pas abandonnée, elle fera vraisemblablement partie d'une version suivante de l'analyseur IDS.

⁸ sous-bande la plus énergétique non audible prise séparément mais pouvant modifier la perception suivant qu'elle est incluse ou non dans une re-synthèse ; sous-bande avec un poids objectif relativement faible qui fournit une perception synthétique incontournable lors d'une écoute, notamment.

⁹ Laurent Millot, Gérard Pelé et Mohammed Elli, Using perceptive subbands analysis to perform audio scenes cartography, *118th Convention of the Audio Engineering Society*, Barcelone, Espagne, 28-31 mai 2005, Paper 6340.

2. PROPOSITION D'UN PROTOCOLE DE TEST DE LATÉRALITÉ AUDITIVE : L'IDS-GRAMME

Comme nous l'avons vu, les résultats des tests perceptifs étudiant le phénomène de latéralité auditive sont difficilement interprétables, qu'il s'agisse du test d'écoute dichotique ou de celui de la boucle audio-phonatoire.

De plus, l'utilisation d'un test audiométrique tonal, utilisant des sons sinusoïdaux, en préambule ne nous paraît pas être le plus adapté compte tenu des stimuli, sons de parole, utilisés par la suite.

Aussi, nous nous sommes posés la question de la mise en place d'un nouveau protocole, utilisant des stimuli de type parole ainsi que le découpage IDS, correspondant à un découpage potentiellement relié à des bandes sensibles pour l'oreille humaine. Nous appelons, dans la suite, cet appareil l' « IDS-gramme ».

2.1. DESCRIPTION DU PROTOCOLE

a) Test du seuil d'audition et d'intelligibilité

L'étude envisagée ici concernant la perception de la parole, nous pensons utiliser comme stimuli des textes lus ou des monologues faisant sens¹⁰, potentiellement enregistrés avec une voix féminine et une voix masculine.

La principale caractéristique de ce protocole réside dans l'interactivité entre le sujet et le stimulus. C'est, en effet, suivant la façon dont le sujet influe sur le stimulus que se dégageront les caractéristiques significatives de sa perception auditive. La méthode employée consiste à permettre à l'auditeur de modifier les niveaux d'intensité sonore de chacune des dix sous-bandes de l'IDS-gramme.

Afin de rendre la description qui va suivre la plus claire possible, nous procédons à un abus de langage en désignant par « équilibrage spectral » cet ajustement des poids relatifs de chacune des dix sous-bandes de l'IDS-gramme. On veillera donc à ne pas confondre ce terme

¹⁰ Utiliser un monologue correspondant à une « scénette » pourrait s'avérer plus représentatif d'une situation d'écoute réelle, empruntée à la vie courante, qu'un texte lu avec le jeu sur les intonations et le rythme propre à la déclamation.

avec celui de la « balance spectrale » qui désignera l'état figé des différents niveaux des sous-bandes de l'IDS-gramme. Le premier terme correspond donc à la correction du second.

Pour le cas incluant le travail sur l'équilibrage spectral, une interface de contrôle ergonomique et simple devra être élaborée, par exemple en s'inspirant de l'interface d'un égaliseur graphique avec un réglage de niveau de diffusion. Il pourrait être pertinent d'inclure, dans le protocole de cet examen, une phase d'explication, potentiellement illustrée d'un point de vue sonore de manière évidente, ou « d'apprentissage » du dispositif de contrôle.

En s'inspirant des méthodes traditionnelles utilisées pour déterminer la sensibilité auditive du sujet et vérifier le degré de dissymétrie de cette sensibilité, la tâche à réaliser consisterait, pour le sujet :

- à ajuster le niveau de diffusion, si le stimulus utilisé est une voix naturelle, jusqu'à ce que la voix soit parfaitement intelligible ;
- à ajuster le niveau de diffusion global, jusqu'à ce qu'un signal soit perçu, puis à procéder à l'équilibrage spectral jusqu'à ce que la re-synthèse obtenue lui paraisse complètement intelligible.

Ainsi seraient mises en évidence la (les) sous-bande(s) pertinente(s) pour laquelle (lesquelles) le sujet dépasse la perception d'un son de parole pour accéder à son intelligibilité.

Un troisième ajustement pourrait également être envisagé permettant au sujet de passer de l'équilibrage spectral caractérisant l'intelligibilité de la voix à celui représentatif, pour lui, d'une voix naturelle.

Une fois le niveau global de diffusion réglé sur le seuil d'audibilité par le sujet, il pourrait être intéressant, dans le cas du travail sur l'équilibrage spectral, de partir de stimuli pour lesquels la balance spectrale n'a pas été modifiée et, au contraire, de re-synthèses pour lesquelles la balance spectrale a été sévèrement modifiée de façon à altérer, de manière importante, l'intelligibilité des stimuli.

En observant les différences entre les corrections introduites lors de l'altération initiale des stimuli et celles introduites par le sujet, on pourrait vraisemblablement accéder à la sensibilité, sous-bande par sous-bande, du sujet testé¹¹.

Il faudrait alors étudier les réglages du seuil d'audibilité, les corrections de la balance spectrale voire les modifications du niveau global de diffusion, après corrections, pour des tranches de populations bien définies afin d'obtenir des informations analogues à celles sur les pertes d'audition mises en évidence par les audiogrammes suivant la catégorie d'âge du sujet.

Enfin, ce test serait réalisé pour chaque oreille voire avec une diffusion sur les deux oreilles, avec la question de la pertinence de cette diffusion puisque le signal perçu correspond alors à une source fantôme.

Ainsi, cette partie du test nécessiterait comme matériel :

- un ordinateur ;
- une paire d'écouteurs ;
- une interface de contrôle, idéalement distincte de l'ordinateur, se présentant principalement sous la forme d'un égaliseur graphique où chaque potentiomètre serait affecté à la modification du poids relatif d'une sous-bande, avec un potentiomètre dédié au réglage du niveau de diffusion, chacun des potentiomètres étant dénué de toute indication de nature à limiter la liberté des réglages par le sujet ;
- un dispositif permettant d'enregistrer le seuil d'audibilité (appui sur un premier bouton), puis les corrections introduites et le seuil de diffusion finale (appui sur un deuxième bouton), ainsi qu'un dispositif permettant de passer à la situation suivante (nouveau stimulus, nouvelle altération de la balance spectrale, remise à zéro de chacun des potentiomètres de sous-bande et à « moins l'infini » pour le niveau de diffusion).

¹¹ Travailler sur le « ré-équilibrage » d'une balance spectrale très perturbée semble, en effet, faire plus sens que de chercher à connaître le seuil de perception pour chacune des sous-bandes, car leur contenu peut être de nature à déstabiliser le sujet.

b) Question de l'émergence de la voix au travers d'une ambiance sonore

L'audiogramme se limite à tester les seuils d'audition pour un jeu de fréquences limité, tandis que l' "Audioscan" réalise une exploration plus fine, mais reposant toujours sur la diffusion de sinusoïdes, afin de déterminer les encoches dans l'audiogramme.

Or, l'écoute de sinusoïdes, même en multipliant le nombre, ne correspond pas vraiment à une situation d'écoute naturelle et, plus particulièrement aux situations pour lesquelles les pertes d'audition sont les plus évidentes pour les personnes.

En effet, ces situations de gênes auditives sont, d'abord quand on interroge les personnes concernées ne serait-ce que dans notre entourage, associées à une préoccupation centrale : réussir à suivre une conversation avec une personne en vis-à-vis, dans les ambiances sonores rencontrées dans la vie quotidienne. On peut notamment citer le cas de la discussion dans les transports publics (trains, bus, métro et RER pour les gens en région parisienne), dans la rue, au restaurant ou dans un café, pendant une réunion de groupe, voire dans des lieux tels que les boîtes de nuit. On peut aussi citer le cas du suivi de la conversation lors de repas de famille réunissant des grands-parents, voire arrière-grands-parents, aux petits enfants, ou encore d'un programme télévisuel qui entraîne vite un réglage important du volume sonore, devenant agressif pour les proches non affectés par un problème d'audition similaire. Et, il serait aussi intéressant de penser au cas de l'écoute téléphonique, qu'il s'agisse d'un téléphone fixe ou portable utilisant l'un comme l'autre des signaux diffusés modifiés par des algorithmes de compression de données.

Il convient également de garder en mémoire la question de l'adaptation des prothèses auditives et de leur réglage quand on entend, par exemple, des personnes âgées, appareillées, déclarer qu'elles n'entendent rien ou qu'elles entendent un sifflement. Il ne paraît pas normal de les voir laisser dans un tiroir de leur table de chevet ces prothèses déclarant (ou parce qu'elles le décident ?) que ces dernières ne leur servent à rien puisqu'elles ne parviennent pas à entendre, suivre et comprendre ce qui se dit. L'étape suivante étant souvent un enfermement progressif dans un silence constant voire une « évasion » dans un monde intérieur, synonyme de rupture du lien avec les proches.

En considérant ces problèmes et les doutes quant à la validité de l'audiogramme par rapport à ces situations problématiques, nous proposons l'élaboration d'un protocole

d'évaluation de l'audition centré sur le problème de l'émergence du signal de parole au sein d'une ambiance sonore fortement perturbante.

Ainsi, nous focaliserons notre proposition sur le cas des ambiances potentiellement les plus gênantes, nécessitant le plus d'effort et donc de nature à fatiguer le plus la personne éprouvant des difficultés à suivre, sans avoir recours à la lecture sur les lèvres, le discours de son vis-à-vis. Ceci signifie que nous laissons notamment de côté le cas de l'écoute d'un programme télévisuel.

Le protocole proposé va s'appuyer sur deux phases successives, mettant en scène la situation d'écoute de ce que dit une personne en vis-à-vis, dans le cadre d'une écoute au casque même si l'on s'écarte sensiblement de la situation d'écoute naturelle. Ce protocole reposerait sur des re-synthèses IDS de stimuli réels enregistrés.

La première phase consisterait en un travail dans le cas où aucune ambiance sonore ne vient perturber l'écoute de ce que dit le vis-à-vis. Les stimuli utilisés seraient des monologues, « vivants », correspondant à une situation où la personne en vis-à-vis monopolise la parole le temps de raconter une histoire qui lui est arrivée ou non, sur un sujet emprunté au quotidien.

Pour cette phase, on proposerait à la personne suivant l'examen de régler le niveau de diffusion pour que le signal de voix soit audible et, ensuite, de procéder à la modification des poids relatifs de chacune des sous-bandes afin d'obtenir un signal parfaitement intelligible, avec éventuellement la possibilité d'ajuster le niveau de diffusion global pendant et/ou après l'opération d'équilibrage spectral.

Cette première phase permettrait à la fois de recueillir des données en situation d'écoute facilitée afin de réaliser des comparaisons avec les résultats des audiogrammes de références, mais, aussi de phase d'apprentissage pour le sujet à la fois du fonctionnement de l'interface de contrôle et de familiarisation avec la ou les voix utilisées lors du test.

L'idée de familiariser le sujet avec les voix utilisées dans le cadre du test repose sur le constat que l'on discute plus souvent avec des personnes que l'on connaît, et dont la voix a été entendue dans des ambiances et situations variées, y compris potentiellement calmes. Nous introduisons donc l'hypothèse selon laquelle le fait de faire « connaissance » avec les voix constituant la banque de référence pour le test est de nature à augmenter la proximité de la situation de test avec une situation réelle.

Il serait ainsi intéressant de discuter l'idée de conserver, pour un même sujet, les voix ayant servi au(x) test(s) précédent(s) tout en disposant d'une banque d'histoires suffisamment

importante pour que l'effet de mémorisation ne puisse pas jouer. Une idée à tester consisterait à choisir des histoires banales, comme on entend sans arrêt, afin de limiter les éléments permettant une mémorisation du patient. Une autre idée, potentiellement plus difficile à mettre en œuvre, consisterait à demander à des proches du patient de pré-enregistrer des histoires banales en amont de l'examen. Une autre idée, pourrait consister à envisager que l'examen se fasse non plus avec des voix pré-enregistrées mais directement avec les proches ayant accepté de venir accompagner la personne pendant son examen, examen qui nécessiterait la mise à disposition d'une version temps réel de l'analyseur IDS incluant jusqu'à la re-synthèse à la volée. Une étape, ultérieure, consisterait à envisager que les deux personnes passent le test et dialoguent réellement.

Mais, à chaque fois, la consigne serait de procéder au réglage du niveau de diffusion et à l'équilibrage spectral afin qu'au final le signal du vis-à-vis soit aussi intelligible que possible.

La deuxième phase consisterait à répéter l'expérience en ajoutant une ambiance sonore particulièrement gênante correspondant à l'une des situations d'écoute les plus difficiles : transport, rue avec du trafic routier, café ou café-restaurant au moment du « coup de feu », marché, grande gare, réunion mondaine (cocktail, réunion publique animée).

On utiliserait à nouveau les mêmes voix mais avec des histoires différentes, ou les voix des proches en temps réel à condition de disposer d'un convolveur en temps réel à associer avec l'analyseur IDS temps réel ! La situation où les deux personnes passeraient le test nécessiterait évidemment que l'on double le matériel d'analyse (interfaces de contrôle, microphone, ordinateur gérant l'analyseur IDS ainsi que la convolution, casque).

On pourrait ainsi obtenir des données à comparer avec la situation d'écoute « facilitée », sans ambiance (mais avec une acoustique potentiellement neutre). Et, cela permettrait de voir si la nature de l'ambiance influence les corrections apportées sur la balance spectrale.

Ces données pourraient servir à concevoir des prothèses intégrant ces corrections voire disposant de « presets » associés aux ambiances sonores gênantes caractéristiques. Ces prothèses auditives devraient donc intégrer un analyseur IDS temps réel capable de tenir compte des corrections de balance spectrale.

Evidemment, ce protocole de test nécessite d'être expérimenté afin de vérifier sa pertinence et de procéder à tous les amendements pouvant apparaître lors de la phase d'expérimentation.

On peut déjà recenser plusieurs questions ou points à discuter afin d'affiner et d'améliorer la proposition que nous faisons :

- Quels doivent-êtrre la diversité et le niveau réel des ambiances sonores composant la sonothèque ?
- Quel doit-êtrre le positionnement de la voix sachant que l'on se tourne naturellement vers la source, ce qui conduirait à travailler, dans le cas de la diffusion au casque, avec des sources fantômes ? C'est une question dont l'importance pourrait diminuer si le test était réalisé avec des personnes physiques plutôt que des voix enregistrées ;
- Doit-on avoir recours à des proches, en temps réel, plutôt qu'à la personne faisant passer le test ? Celle-ci pouvant être éventuellement plus à même de mener une discussion avec un rythme suffisant à condition d'être « entraînée » à cette tâche ;
- L'hypothèse de ne pas utiliser de sources diffusées soit à gauche soit à droite uniquement, mais positionnées au centre (bien qu'il s'agisse de sources fantômes) est-elle recevable ? Car, en effet, quand on cherche à suivre ce que dit quelqu'un, en général on décide de lui faire face ;
- La proposition de régler le poids relatif pour chacune des oreilles indépendamment n'aboutira-t-elle pas à une interface de nature à faire peur au sujet car comportant trop de boutons (un réglage par sous-bande et un réglage de niveau par canal ?) ?
- Est-il pertinent de partir d'un pré-réglage correspondant à ce qui a été obtenue, suivant la voix utilisée, lors de la première phase ? Ceci dans l'idée que cela pourrait aider la personne à trouver plus rapidement les corrections à introduire, mais aussi, pour voir les corrections relatives à apporter potentiellement par rapport à la situation d'écoute « facile » ;
- Quel doit-êtrre le cahier des charges pour les ambiances ? Il s'agirait, ici, du dispositif d'enregistrement (système de captation, format audionumérique utilisé tant en termes de fréquences d'échantillonnage que de dynamique de codage, niveau d'enregistrement), du niveau de diffusion principalement ;
- Serait-il pertinence d'utiliser une interface de contrôle semblable au premier test avec reset de la configuration, passage à la nouvelle ambiance ?
- Etc.

2.2. ETUDE DE LA PERTINENCE DU DECOUPAGE DIT DE « LEIPP ETENDU »

Jusqu'à présent, la question de la pertinence du découpage n'a pas été évoquée et on a considéré principalement l'hypothèse de l'utilisation du découpage dit de « Leipp étendu ».

Il serait donc intéressant de vérifier sa pertinence ainsi que celle des extensions proposées, même si certains constats associée à la sous-bande 0-50 Hz sont de nature à accorder du crédit à l'ajout de cette sous-bande. Cette question de la pertinence du découpage n'est pas anodine car on peut se demander si, par exemple, les habitudes d'écoute comme les ambiances sonores auxquelles nous sommes exposés dans le cadre de notre quotidien n'ont pas évolué et donc changé les caractéristiques de l'audition.

On peut envisager de reprendre le protocole utilisé par Emile Leipp en utilisant les avantages proposés par la technologie audionumérique et en l'adaptant un peu.

Le cœur du protocole résiderait toujours sur la détermination des fréquences de coupure, mais à -3 dB, limitant chacune des sous-bandes, avec une question supplémentaire liée à la coupure des filtres puisque dans la version numérique de l'analyseur IDS les filtres ont une coupure importante, permettant de diminuer de manière conséquente la présence du contenu des sous-bandes adjacentes. Il serait donc nécessaire d'étudier aussi l'influence de la coupure des filtres sachant que cela introduit un paramètre supplémentaire à gérer pour le sujet (la longueur des filtres étant relié au choix de la méthode de calcul de la réponse impulsionnelle des filtres, à l'atténuation et à la raideur des filtres).

Par contre, du fait de l'utilisation de filtres numériques, il faudrait disposer d'un dispositif de mise à jour des filtres, *a priori* RIF à phase linéaire, rapide permettant d'opérer les changements en temps réel. Et, puisque les filtres seraient amenés à voir leurs caractéristiques évoluer pendant la phase de détermination d'un paramètre, il faudrait disposer d'une version temps réel de l'algorithme de convolution temporelle ou « taches-algorithm »¹².

¹² Laurent Millot et Gérard Pelé, An alternative approach for the convolution in time-domain: the taches-algorithm, *124th Convention of the Audio Engineering Society*, Amsterdam, Pays-Bas, 17-20 mai 2008, Paper 7412.

Une autre question à étudier concerne la gamme de fréquence où commencer la recherche du découpage. En effet, si Leipp commence par la sous-bande des « graves » (au final la sous-bande 2, 50-200 Hz), on pourrait envisager de commencer par la sous-bande 7 (1.8-3 kHz) qui semble pouvoir constituer un résumé perceptif du signal ou bien alors commencer par la sous-bande 3 (200-400 Hz) ou 4 (400-800 Hz) qui sont associés aux fréquences fondamentales pour les instruments de musique ou assez corrélés au premier formant au moins dans le cas de la voix. Mais, une alternative pourrait consister à laisser le sujet libre de commencer là où il le souhaite, tout en notant la stratégie suivie pour construire le découpage.

Si dans son protocole Leipp, détermine les sous-bandes les unes après les autres sans, *a priori*, revenir sur les sous-bandes déjà déterminées, il faudrait étudier la pertinence de laisser la possibilité de revenir sur les réglages des premières sous-bandes déjà trouvées par le sujet, afin de lui laisser la possibilité d'affiner son découpage. Mais il faudra vérifier si cela ne conduit pas le sujet à faire des aller-retour nombreux entre les réglages des différentes sous-bandes et, ainsi, augmenter la durée de la procédure de détermination du découpage.

La question des conditions d'écoute est aussi à poser. En effet, par rapport aux séances en groupe de Leipp, il pourrait s'avérer intéressant de tester ce que donne une procédure individuelle et libre. Evidemment, en procédant de la sorte, on perd le principe de la bascule en aveugle entre la version filtrée et la version originale, qui ne semble pas inintéressante ne serait-ce que pour aboutir plus rapidement à un découpage satisfaisant globalement l'ensemble du groupe testé.

Mais il convient quand même de remarquer que Leipp fige une sous-bande quand la totalité (ou presque) du groupe est d'accord pour indiquer qu'il y a quelque chose qui manque lorsque l'on bascule entre l'original et la version filtrée. Or, cela signifie que les limites sont, d'une certaine manière, déterminées par la ou les personnes les moins sensibles du groupe, ce qui peut conférer au découpage trouvé une plus grande représentativité allant au-delà du panel des sujets participant à l'expérience.

Si l'on décide de reprendre l'idée d'une détermination individuelle du découpage, il faudra discuter du choix de travailler au casque, ce qui permet de s'affranchir de la salle. Ceci pourrait se révéler intéressant dans le cas où l'étude du découpage ne se ferait pas dans un lieu unique mais reposerait sur une diffusion large du matériel, ou des spécifications du protocole, accompagnée d'un outil de collecte (site web dédié) des découpages obtenus par les personnes

acceptant de participer à cette recherche. Dans cette optique, le travail à partir d'une diffusion par une enceinte, pour éviter les sources fantômes, serait potentiellement ignoré. Par contre, l'idée de travailler sur des sources ou scènes sonores stéréophoniques semble intéressante car de nature à faire intervenir dans l'obtention du découpage une prise en compte des effets de spatialisation et de masquages/démasquages des constituant d'une scène sonore.

La question des stimuli est aussi à poser, et en suivant plusieurs axes de réflexion.

En effet, dans le cadre d'une procédure individuelle, il semble intéressant de proposer au sujet la bascule à la volée d'un stimulus à l'autre pour une banque de stimuli représentative des scènes sonores que l'on veut étudier.

Il faudrait aussi potentiellement conserver le principe de bascule à la volée entre l'original et sa version filtrée.

Si Leipp suppose, plus ou moins implicitement, l'universalité de son découpage. Il semble pertinent de vérifier que ce découpage est universel et qu'il ne faut pas plutôt envisager plusieurs découpages. En effet, le découpage obtenu par Leipp a été déterminé en utilisant principalement de la musique classique et de la voix chantée. Or, il pourrait s'avérer intéressant de vérifier s'il faut ou non disposer de plusieurs découpages en considérant, par exemple, les cas de figure suivants, même si l'unicité de l'appareil auditif peut être un argument pour penser que ce découpage est unique et universel :

- variation du découpage suivant le genre musical ce qui conduirait à des découpages potentiellement différentes ;
- influence de la langue et de la culture des sujets ce qui pourrait conduire à l'émergence de découpages spécifiques à des zones ou à des continents, bien que la mondialisation puisse gommer ces différences, du moins dans les pays dits « développés » ;
- travail sur des ambiances sonores gênantes pour voir si la nature de la source ou son caractère intrusif est de nature à changer les stratégies ou réflexes d'analyse.

On peut alors essayer de dresser une liste non exhaustive des outils nécessaires et esquisser le cahier des charges :

- fonction ou outil logiciel pour modifier les caractéristiques du filtre en temps réel ;
- filtrage en temporel par algorithme des taches¹³ ;
- application pure data voire compilée ;

¹³ *Ibid.*

- réutilisation des limites des sous-bandes déjà réglées ;
- question de la pertinence de laisser retoucher aux limites des sous-bandes déjà réglées quand on travaille sur une autre bande ;
- écoute à la volée entre l'originale et la version filtrée par la sous-bande en cours d'étude ;
- question de savoir si l'on doit avoir à l'écran un ajout au fur et à mesure de lignes avec le positionnement des limites des bandes déjà trouvées (présentation avec une échelle logarithmique) ou si l'on doit les faire figurer sur une seule voire deux lignes (une pour ce qui est réglée, l'autre pour la sous-bande en cours de détermination) ;
- chargement d'une sélection de stimuli adaptée à une utilisation ou à une culture/langue données avec possibilité de basculer à la volée d'un stimulus à l'autre pour tester la pertinence de la sous-bande en cours de détermination sur un nombre suffisant de stimuli variés.