

Présentation de MoreSound Intelligence™

MoreSound Intelligence (MSI) est la nouvelle fonction de « traitement et organisations » des scènes sonores inspirée par le cerveau embarquée dans les aides auditives Oticon More™. MSI se compose de différentes sous-fonctions et ce document technique vise à vous offrir une meilleure connaissance de ces dernières.

Tout d'abord, une vue d'ensemble complète de la fonction MSI vous sera présentée. Puis, nous nous pencherons sur les sous-fonctions individuelles, une par une, dans l'ordre dans lequel elles apparaissent dans le flux de traitement.

Voici quelques éléments importants :

- Virtual Outer Ear - le nouveau modèle réaliste du pavillon de l'oreille efficace dans les environnements faciles, avec trois réglages dans Oticon Genie 2 pour répondre aux préférences des utilisateurs
- Neural Clarity Processing - lors de sa phase de développement, ce Réseau neuronal profond est entraîné avec des scènes sonores réelles pour soutenir le cerveau de façon optimale. Il est intégralement intégré dans la conception de la plateforme Polaris™
- Sound Enhancer - gain dynamique principalement pour la parole, particulièrement utile dans les environnements complexes, avec trois réglages dans Oticon Genie 2 pour répondre aux préférences des utilisateurs

02	MoreSound Intelligence
02	Scanner et analyser
03	Spatial Clarity Processing
04	Virtual Outer Ear
05	Spatial Balancer
05	Neural Clarity Processing - Réseau Neuronal Profond
09	Sound Enhancer
10	Perspective
11	Références

AUTEURS DE CETTE PUBLICATION

Mette Brændgaard,
Spécialiste produits, Product Marketing Support, Oticon A/S

*Un remerciement particulier à **Brian Man Kai Loong**, Audiologiste en recherche clinique, pour sa contribution à la partie sur le Réseau Neuronal Profond*

Les scènes sonores sont dynamiques, complexes et imprévisibles et le rôle du cerveau est de gérer cette complexité ; d'entendre et de créer du sens à partir de cet ensemble. Le traitement de l'aide auditive doit fournir un signal de bonne qualité pour aider le cerveau à l'interpréter. Cependant, ce n'est pas en limitant la scène sonore par l'application de la réduction du bruit et de la directivité que cela se produit. Le cerveau a besoin d'accéder à plus d'informations provenant de son environnement immédiat pour être aidé dans sa façon naturelle de travailler. Plus de perspective intégrale des sons, pour mieux profiter de la vie. (O'Sullivan et al., 2019 ; Hausfeld et al., 2018 ; Puvvda & Simon, 2017)

MoreSound Intelligence d'Oticon réalise une avancée décisive dans le traitement de la scène sonore en donnant accès à la scène sonore intégrale avec un contraste et un équilibre clairs.

MoreSound Intelligence

MoreSound Intelligence (MSI) est une fonction avancée qui inclut plusieurs sous-fonctions. Dans les paragraphes suivants, ce document technique décrira les différentes sous-fonctions.

La figure 1 présente les différentes étapes et sous-fonctions de MSI. Tout d'abord, la scène sonore est scannée et analysée. En fonction de cette analyse, et en association avec les réglages du logiciel d'adaptation (Oticon Genie 2), la scène sonore est ensuite transmise au Spatial Clarity Processing et au Neural Clarity Processing. Elle suit l'une des deux trajectoires : la trajectoire pour les environnements faciles ou la trajectoire pour les environnements difficiles. À la sortie

finale du bloc MSI, on obtient un signal propre qui est prêt à une adaptation plus poussée par l'aide auditive (par ex. amplification).

Pour plus d'informations, veuillez consulter Brændgaard, M. 2020. The Polaris Platform. Oticon tech paper - on the complete processing flow for Oticon More, and Løve, S. 2020. Optimal Fitting of Oticon More. Oticon Whitepaper - on fitting Oticon More in Genie 2.

Tout au long de MSI, le traitement de l'entrée sonore est effectué dans 24 canaux. Par rapport aux précédents appareils auditifs haut de gamme d'Oticon, le nombre supplémentaire de canaux fournit deux fois plus de précision dans une plage de fréquences qui inclut des canaux de fréquence de 1,5 à 5 kHz (Brændgaard, 2020), les plus importants pour les sons de la parole.

En plus des canaux supplémentaires pour plus de précision, les 24 canaux sont aussi liés. Cela signifie qu'au sein du Réseau neuronal profond, tous les canaux peuvent voir le traitement qui se déroule dans les autres canaux. Cela minimise le risque d'artéfacts créés par un type de son ayant été mal catégorisé dans un seul canal. En minimisant les artéfacts, la qualité sonore est améliorée.

Scanner et analyser

Afin d'effectuer un traitement correct des différentes sources sonores, MSI a besoin de connaître les détails exacts de la scène sonore. Par ailleurs, les scènes sonores sont dynamiques et les sources sonores sont constamment en train de se déplacer et de changer. Pour veiller à ce que

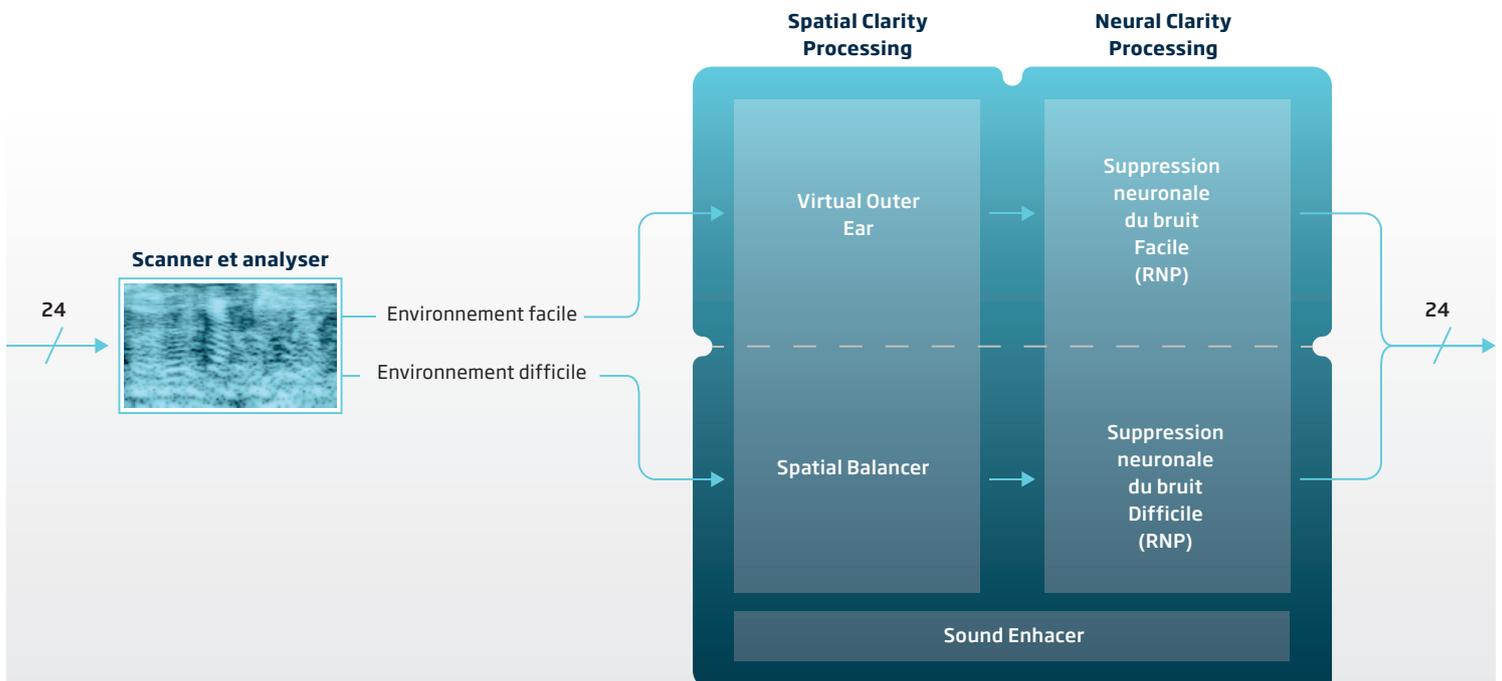


Figure 1 : Le flux de traitement de MoreSound Intelligence

tous les détails soient captés, la scène sonore est scannée 500 fois par seconde, afin de cartographier correctement les différentes sources sonores. En fonction de cette analyse et cartographie, MSI calcule le signal optimal pour le rapport signal/bruit ainsi que les niveaux de bruit.

Les estimateurs du rapport signal/bruit et les estimateurs de niveau ont été mis à jour de façon à s'exécuter désormais dans 24 canaux, et l'estimateur du rapport signal/bruit fonctionne à présent sur une plage plus large de -10 à +15 dB SNR. Le rapport signal/bruit est le principal élément utilisé pour distinguer les environnements faciles et difficiles, et le niveau d'aide fourni par le système sera déterminé grâce aux estimations du rapport signal/bruit et du niveau de bruit. Le principe de la relation entre le rapport signal/bruit, le niveau de bruit et l'aide fournie est indiqué dans la figure 2.

Tout changement dans la scène sonore sera détecté lors de l'analyse, mais seuls les changements persistants (plus de 2 secs) feront en sorte que l'aide auditive adapte le niveau d'assistance.

MoreSound Intelligence utilise le rapport signal/bruit pour distinguer les environnements faciles et difficiles, mais les détails spécifiques diffèrent d'une personne à l'autre. La distinction entre les environnements faciles et difficiles dépend des réglages individuels de l'utilisateur dans Oticon Genie 2. Autrement dit, les réglages effectués dans Genie 2, basés sur les données de l'utilisateur de la séance d'adaptation, détermineront le rapport signal/bruit qui doit être présent pour qu'une scène sonore soit traitée en tant qu'environnement facile ou difficile.

D'une part, lorsque la scène sonore est considérée comme un environnement facile, le traitement suit le flux (voir figure 1) qui commence avec le Virtual Outer Ear, puis la Suppression neuronale du bruit - Facile.

En revanche, lorsque la scène sonore est considérée comme un environnement difficile, le traitement suit le flux qui commence avec le Spatial Balancer, puis la Suppression neuronale du bruit - Difficile, avec l'aide sous-jacente du Sound Enhancer.

L'analyse complète de la scène sonore est également utilisée par le Neural Clarity Processing pour le traitement dans le Réseau Neuronal Profond (voir plus loin dans ce document).

Spatial Clarity Processing

Pouvoir situer les sources sonores dans l'environnement spatial constitue une capacité importante qui devient plus difficile en présence d'une perte auditive (Akeroyd, 2014).

Le pavillon de l'oreille nous aide à localiser les sons dans trois dimensions - distance, vertical (haut/bas) et horizontal (avant/arrière). Selon que la différence interaurale de temps, la différence d'intensité interaurale, ou l'effet du masque du crâne soient en œuvre, certaines fréquences sont plus pertinentes que d'autres pour la localisation (Akeroyd, 2014).

Nous avons tous des tailles d'oreilles et des formes de pavillons de l'oreille différentes, le son sera donc modifié de différentes façons lorsqu'il entre dans le conduit auditif, selon l'anatomie de l'oreille. Par exemple, en fonction de la forme de l'oreille externe, certaines personnes auront plus ou moins de focalisation frontale que d'autres.

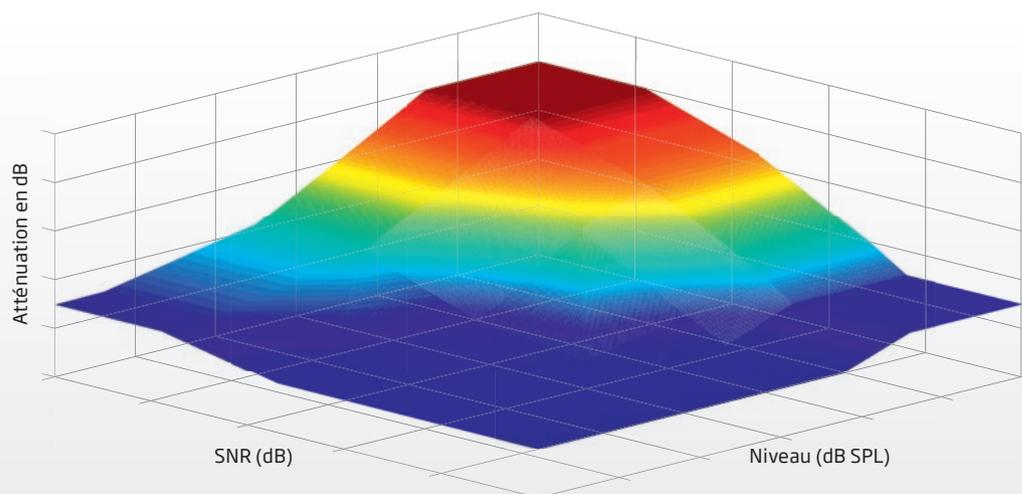


Figure 2. Principe de la relation entre le rapport signal/bruit, le niveau de bruit et la suppression du bruit. Plus on s'approche de la gauche de l'axe du rapport signal/bruit, plus le rapport signal/bruit est bas. Plus on s'approche de la droite de l'axe du niveau, plus le niveau est élevé. Plus le rapport signal/bruit est bas et plus le niveau est élevé, plus le système fournit de l'aide.

Lorsqu'on place les microphones des aides auditives derrière l'oreille, la capacité à utiliser les indices de localisation spatiale naturels fournis par le pavillon de l'oreille est éliminée. Cette capacité doit être recréée par le traitement du signal dans l'aide auditive.

Le Spatial Clarity Processing se compose de deux fonctions différentes, le Virtual Outer Ear et le Spatial Balancer, qui aident à recréer cette sensation spatiale, respectivement dans les environnements faciles et difficiles.

Virtual Outer Ear

Le Virtual Outer Ear (VOE) est constituée de trois modèles réalistes du pavillon de l'oreille qui peuvent être définis par l'audioprothésiste dans le logiciel d'adaptation en fonction des besoins sonores spatiaux de l'utilisateur d'aides auditives.

Le VOE aide l'utilisateur à recréer une conscience spatiale dans les environnements faciles.

Pour obtenir la meilleure compensation par traitement du signal du pavillon de l'oreille naturel, les fonctions de transfert liées à la tête (FTLT ou Head-Related Transfer Functions - HRTF) ont été mesurées à partir de différents angles, dans le plan horizontal, sur 130 oreilles humaines. L'Index de directivité (ID), qui est une mesure de la quantité de directivité des oreilles individuelles dans chaque bande de fréquence, peut être calculé à partir des FTLT. Un ID plus élevé équivaut à une oreille plus directionnelle.

Les résultats de l'ID sont indiqués dans la figure 3. La ligne magenta foncée représente la moyenne, tandis

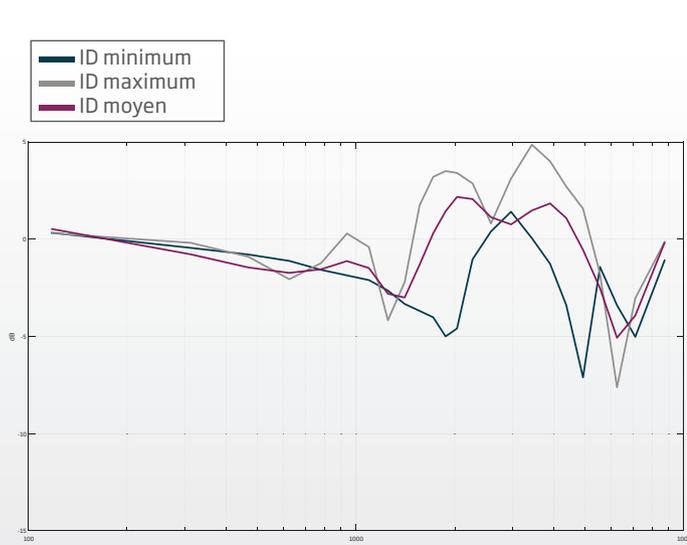


Figure 3. Indice de directivité pour les oreilles humaines présentant un ID moyen (magenta foncé), minimum (bleu) et maximum (gris).

que la ligne bleue correspond à l'oreille avec le plus petit ID calculé (et par conséquent le plus omni directionnel). La ligne grise représente l'oreille avec l'ID calculé le plus important (oreille avec le plus de focalisation frontale). La différence en termes d'ID entre les oreilles peut être principalement observée entre 2 et 5 kHz et il peut s'agir d'une différence relativement importante.

Les informations des FTLT peuvent également être présentées sous forme de diagramme polaire. Celui-ci montre la réponse pour un angle/une fréquence spécifiques atteints dans le VOE (figure 4). Encore une fois, la ligne magenta représente la moyenne, la bleue correspond au plus petit ID et la grise à l'ID le plus important.

La différence entre les oreilles varie autour d'un ID de 4 dB, de -2 dB à +2 dB, et la plupart des mesures se trouvent dans la zone de l'ID de 0,5 dB. Cela signifie que la plupart des personnes obtiennent une amplification naturelle de l'oreille externe d'environ 0,5-1 dB dans la zone 2-5 kHz.

Le VOE est donc basé sur ces trois mesures pour créer un modèle du pavillon de l'oreille aussi naturel et précis que possible.

Les mesures ont montré que l'effet du pavillon de l'oreille sur le son varie entre les oreilles. Cela signifie que la façon dont une personne est habituée à entendre le son différera selon l'anatomie de son oreille externe. En outre, il n'est pas possible de prendre une simple mesure de l'oreille externe, puis d'inférer la façon dont l'utilisateur « entend ». Cependant, étant donné que nous savons que cet aspect diffère entre les individus,

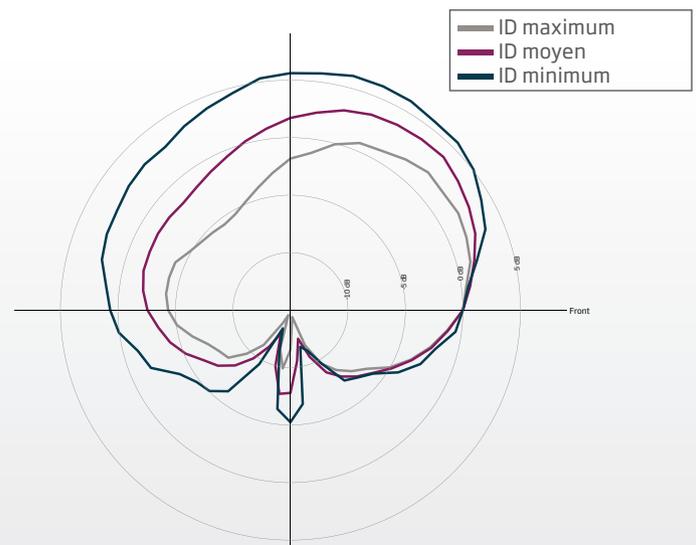


Figure 4. Diagramme polaire pour l'oreille gauche présentant la moyenne de la directivité des oreilles humaines calculée pour 2-5 kHz. Dans cet exemple, la personne fait face au côté droit.

le VOE dispose de trois réglages légèrement différents avec plus ou moins de focalisation frontale (voir la figure 5) et ceux-ci peuvent être définis dans le logiciel d'adaptation Oticon Genie 2 selon les préférences de l'utilisateur. La focalisation légèrement plus frontale est créée en laissant moins de son entrer en provenance de l'arrière.

Les trois réglages différents sont basés sur les mesures de l'ID suivantes :

- Les mesures moyennes sont utilisées pour le réglage Équilibré (par défaut) qui est conçu pour créer le meilleur rapport entre l'audibilité de tous les sons dans l'environnement de l'utilisateur tout en étant capable de se focaliser sur la parole provenant de l'avant
- Les mesures les plus élevées sont utilisées pour le réglage Restreint (plus de focalisation sur la parole provenant de l'avant)
- Les mesures les plus basses sont utilisées pour le réglage Elargi (accès à tout l'environnement)

Les différences entre les trois réglages sont d'un ID d'environ 0,5 dB, mesuré avec un IA pondéré*.

Spatial Balancer

Le Spatial Balancer constitue une fonction plus puissante que l'VOE pour faire face aux environnements difficiles. Le Spatial Balancer équilibre rapidement les différentes sources sonores de l'environnement.

Le Spatial Balancer est alimenté avec un signal omnidirectionnel et un signal cardioïde arrière en provenance des deux microphones. Le signal omnidirectionnel fournit tous les sons de la scène sonore, y compris les sons provenant de l'avant, qui constituent souvent les signaux les plus importants pour l'utilisateur. Le signal cardioïde arrière fournit tous les sons de la scène sonore sauf les sons provenant de

* IA pondéré : Indice d'articulation pondéré. Cela signifie que dans le calcul de l'ID, les fréquences sont pondérées selon l'importance pour la compréhension de la parole. Les fréquences les plus basses et les plus hautes ont moins de poids que les moyennes fréquences.

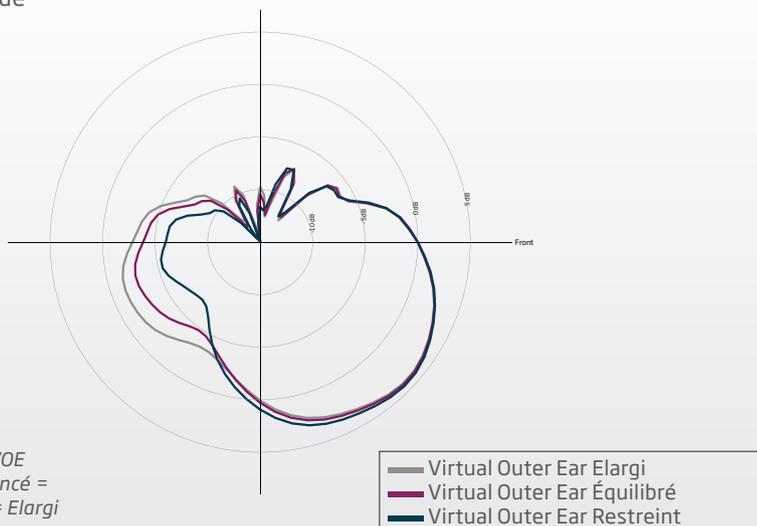
l'avant. Les deux signaux sont constamment comparés pour définir l'emplacement des sources de bruit. Le Spatial Balancer utilise un formateur de faisceaux à réponse sans distorsion de la variance minimale (MVDR) pour créer des diagrammes polaires effectuant l'équilibre le plus optimal pour une scène sonore donnée. Dans les diagrammes polaires, la direction nulle est placée vers les sources de bruit les plus dominantes pour atténuer le bruit et le maintenir dans le fond de la scène sonore. Le système peut créer des directions individuelles nulles pour chacun des 24 canaux de fréquence de l'aide auditive, ce qui permet, en principe, au Spatial Balancer de contrôler 24 sources sonores (48 au total autour de la tête). Chaque direction nulle est mise à jour 125 fois par seconde.

Le Spatial Balancer augmente le rapport signal/bruit en réduisant les sources de bruit individuelles, en les plaçant dans le fond et en créant ainsi une scène sonore équilibrée.

Par rapport aux produits Oticon antérieurs, le nombre de canaux supplémentaires, et donc le nombre de nuls possibles, rend le système plus précis en termes de ciblage de la source sonore. Le nul peut être plus profond, ce qui signifie que la source sonore peut être davantage poussée dans le fond si besoin.

Neural Clarity Processing - Réseau Neuronal Profond

C'est dans le Neural Clarity Processing que réside le développement unique dédié aux aides auditives Oticon More™ - le Réseau Neuronal Profond (RNP). Le RNP a appris à reconnaître, pour une scène sonore donnée, ce qui devrait être mis en avant (sons intéressants avec beaucoup d'informations) et ce qui devrait être mis en fond (sons moins intéressants avec moins d'informations). Il crée ainsi une meilleure clarté et un meilleur contraste entre les sources sonores. Cette section du document vous présentera les réseaux neuronaux et la mise en œuvre par Oticon d'un Réseau neuronal profond.



Vous n'y pensez peut-être pas, mais les réseaux neuronaux sont utilisés partout dans le monde pour faire face à de vastes questions complexes. En fonction de nos actions précédentes, ils peuvent suggérer une liste de chansons ou effectuer des recommandations d'achat. Ce type de problématiques multifactorielles complexes peut être plus facilement traité par un réseau neuronal que par des approches traditionnelles basées sur des règles pré-établies.

Avant de nous pencher sur le fonctionnement du RNP dans l'aide auditive, attardons-nous quelques instants sur la façon dont le RNP reproduit le fonctionnement du cerveau lors de ses apprentissages. Pour mieux illustrer notre propos, nous utiliserons un exemple visuel relatif à la distinction entre un chat et un chien.

Dans nos cerveaux, au sein du cortex, nous avons des zones spécialisées dans le traitement des signaux visuels, comme le cortex visuel primaire qui contient plus de cent millions de neurones et bien plus encore de connexions entre ceux-ci (Leuba G. & Kraftsik R, 1994). Au fil de nos vies en tant qu'êtres humains, nous sommes devenus particulièrement performants dans l'interprétation de ce que nous voyons mais avant de pouvoir le faire, le cerveau doit apprendre. La tâche qui consiste à distinguer un chat et un chien est normalement assez simple pour un adulte, mais il n'est pas si facile de déterminer comment nous la réalisons.

Quelles sont les règles que nous avons établies dans nos cerveaux qui nous rendent capables de gérer cette tâche ? Peut-être s'agit-il de quelque chose en lien avec le musée ou les oreilles, ou peut-être s'agit-il de quelque chose de complètement différent dont nous n'avons même pas conscience. Lorsque nous essayons de verbaliser ces règles, nous réalisons vite que nous sommes condamnés à nous perdre dans d'innombrables exceptions et restrictions. Il en va de même lorsque nous essayons de définir des règles pour reconnaître les sons, et c'est pourtant ce que les scientifiques font depuis longtemps lorsqu'ils s'intéressent à la réduction du bruit dans les aides auditives.

Gestion des sources de bruit

Jusqu'à présent, la réduction du bruit dans les aides auditives a été gérée par des algorithmes développés par l'homme qui définissaient ce qui relevait de la parole pertinente et ce qui relevait du bruit en appliquant des règles, par exemple basées sur la modulation du son et la direction d'origine de la source sonore. Cela impliquait souvent que seule la parole provenant de l'avant était considérée comme pertinente. Or, cette hypothèse ne se vérifie pas forcément dans tous les cas, puisqu'elle ignore l'importance de l'accès du cerveau à la scène sonore intégrale et limite notre capacité à localiser les sons en dehors de sa portée effective. (Pour plus d'informations sur l'importance de l'accès du cerveau à la scène sonore intégrale, veuillez consulter Man, B. et Ng, E. 2020. BrainHearing - La nouvelle perspective. Livre blanc Oticon.)

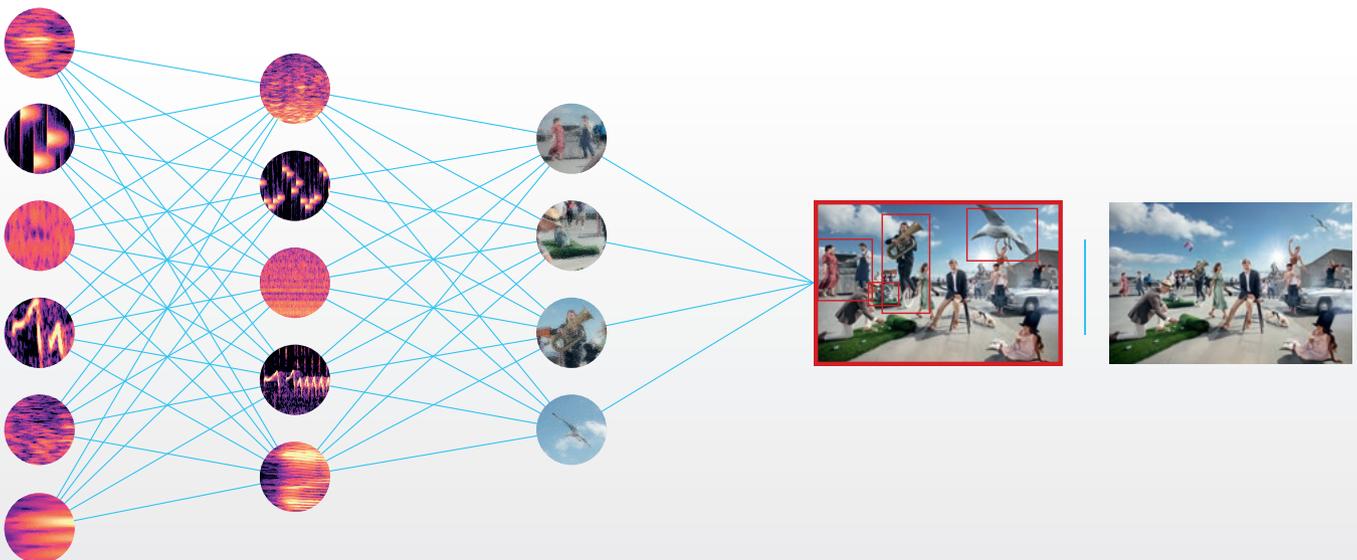


Figure 6. Illustration conceptuelle d'un réseau neuronal.

La distinction entre les objets est un problème multifactoriel. Notre cerveau exploite non seulement les caractéristiques et les structures que nous pouvons identifier à partir du signal, mais également la mémoire sémantique à long terme (Rönnberg et al., 2013). Après toute une vie à utiliser nos oreilles, notre cerveau a stocké une représentation d'un très grand nombre de sons différents. Notre cerveau a appris de façon efficace, par le biais d'erreurs et d'expériences, une approche méthodologique unique pour distinguer les sons pertinents et non pertinents. Les réseaux neuronaux suivent cette même approche. La structure d'un RNP est partiellement inspirée par la façon dont notre cerveau est organisé, à savoir les neurones et leurs synapses correspondantes. Le réseau neuronal utilise l'apprentissage itératif à partir d'une grande quantité de données du monde réel (voir plus loin pour plus d'informations sur le processus d'entraînement/apprentissage) pour établir des connaissances sur le son et la façon de le traiter. L'apprentissage itératif du RNP est appliqué au lieu de suivre un ensemble strict de règles préétablies basées sur des algorithmes développés par l'homme et généralisées aux nombreuses scènes sonores différentes et complexes du monde réel. Cette approche fait sortir le traitement du son et la gestion du bruit du laboratoire pour les emmener dans le monde réel.

Qu'est-ce qu'un réseau neuronal ?

Les réseaux neuronaux sont des catégories spécifiques d'algorithmes qui relèvent de la discipline plus générale de l'apprentissage automatique. L'idée de l'apprentissage automatique est de prendre une grande quantité de données, connues sous le nom d'échantillons d'apprentissage, puis de développer un système à partir duquel il devient possible d'apprendre. Le côté unique

des réseaux neuronaux provient de leur similitude architecturale avec le cerveau. Dans le contexte des réseaux neuronaux, il existe une unité de base nommée « neurone ». Le but d'un neurone, qui s'apparente à celui d'un neurone relié dans le cerveau, est de recevoir des informations, les stocker, et finalement les transmettre au neurone suivant. Un groupe de neurones forme une couche, et les différentes couches spécialisées et reliées entre elles forment le réseau neuronal qui se compose d'une couche d'entrée au début, de couches cachées au milieu et d'une couche de sortie à la fin. Cela constitue le modèle le plus basique de réseaux neuronaux (figure 6).

Le Réseau Neuronal Profond d'Oticon

En général, le développement d'un Réseau Neuronal Profond (RNP) requiert 3 étapes fondamentales : (1) La portée (2) L'entraînement et l'apprentissage, et (3) Le Test (voir figure 7).

1. La portée

Avant toute chose, nous déterminons la portée du problème et nous définissons ce que nous souhaitons fournir aux utilisateurs d'aides auditives. Dans notre cas, nous souhaitons améliorer l'accès des personnes malentendantes à la scène sonore intégrale, en créant un code neural de qualité pour mieux soutenir les stades d'orientation et de sélection dans le cerveau (Man & Ng, 2020).

Pour atteindre cet objectif, nous devons prendre en compte la nature des données (les scènes sonores). Quelles sont les caractéristiques des différents sons, comme la parole et le bruit ? La parole est dynamique et elle change en permanence, même lorsqu'elle provient d'une seule personne. La caractéristique clé d'un signal de parole est qu'il existe un degré de continuité. Par

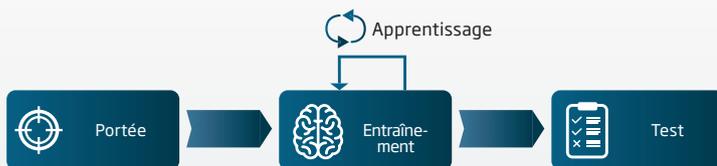


Figure 7. Étapes de développement d'un Réseau Neuronal Profond - 1 La Portée, 2 L'Entraînement et l'apprentissage, et 3 Le Test.

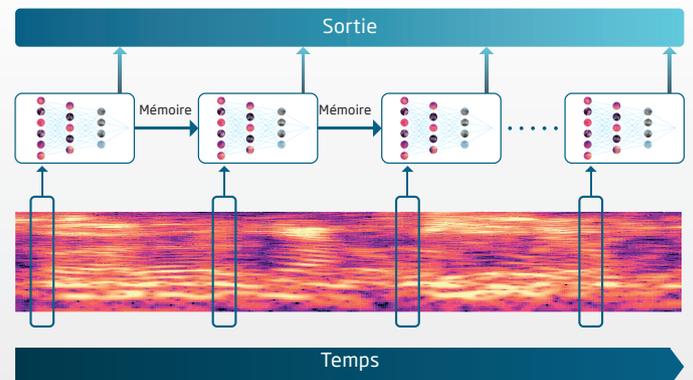


Figure 8. Le principe d'un réseau neuronal à mémoire à court terme et long terme (LSTM) ou un réseau de neurones récurrents fermés (GRU) - relier le réseau neuronal au fil du temps avec le réseau qui se transmet des informations à lui-même.

exemple, si vous écoutez la voix de votre ami, il est peu probable que sa voix change rapidement de rythme ou de hauteur. En revanche, le bruit ambiant peut être un mélange de verres qui s'entrechoquent, de voix de différentes personnes en arrière-plan. Tous ces sons différents, non seulement en termes de hauteur et de rythme, mais également dans leur façon de varier au fil du temps. C'est pourquoi nous avons créé un réseau neuronal spécialisé dans la gestion de ces signaux dynamiques - un réseau de neurones récurrents fermé (Gated Recurrent Unit - GRU), qui est une variante des réseaux de neurones récurrents à mémoire court terme et long terme (Long Short-Term Memory - LSTM).

Les réseaux de neurones récurrents à mémoire court terme et long terme (LSTM), comme leur nom l'indique, sont liés à la mémoire, à court et à long terme.

En psychologie, la mémoire est définie comme la faculté à encoder, stocker et récupérer des informations (Squire, 2009). Comme nous le savons, la mémoire s'apparente à un système de stockage des informations qui peut être préservé au fil du temps. Par ailleurs, on peut y accéder pour effectuer des prises de décisions plus efficaces. Les LSTM et les GRU fonctionnent sur ce principe. L'idée est de « relier » le réseau neuronal au fil du temps, avec le réseau qui se transmet des informations à lui-même (figure 8).

Il en résulte un algorithme qui, non seulement, reconnaît les différentes caractéristiques des sons à un moment précis, mais également la façon dont ces caractéristiques sonores varient au fil du temps. La capacité à intégrer des informations au fil du temps constitue précisément

l'attribut unique dont nous avons besoin pour analyser un signal dynamique comme les sons.

Le RNP se compose d'une couche d'entrée, de couches cachées au sein desquelles le traitement n'est pas visible, et d'une couche de sortie avec le résultat que nous pouvons entendre. Les couches d'entrée et de sortie ont 24 neurones qui correspondent aux 24 canaux de traitement.

2. Entraînement

Le but de cette étape est d'entraîner le RNP sur des scènes sonores jusqu'à ce qu'il puisse résoudre la tâche pour laquelle il a été conçu. De nombreuses données sont nécessaires pour l'entraînement. Ces données sont enregistrées dans différentes scènes sonores parmi un large éventail d'environnements d'écoute auxquels les personnes sont exposées dans leur vie quotidienne. Nous avons utilisé un microphone sphérique spécialisé, capable de capter les sons à 360 degrés, pour fournir au RNP une scène sonore spatialement précise et détaillée et l'entraîner sur la scène sonore intégrale.

Certaines des données recueillies étaient destinées à l'entraînement du RNP et d'autres au test. Les données pour le test ne seront pas utilisées jusqu'à l'étape du test. Les données d'entraînement ont été fournies au RNP en tant qu'entrée pour développer un apprentissage sur les scènes sonores.

Le processus d'entraînement peut être subdivisé en 4 étapes qui se déroulent en boucle : (A) Entrée (B) Propagation avant (C) Sortie (D) Propagation arrière (figure 9).

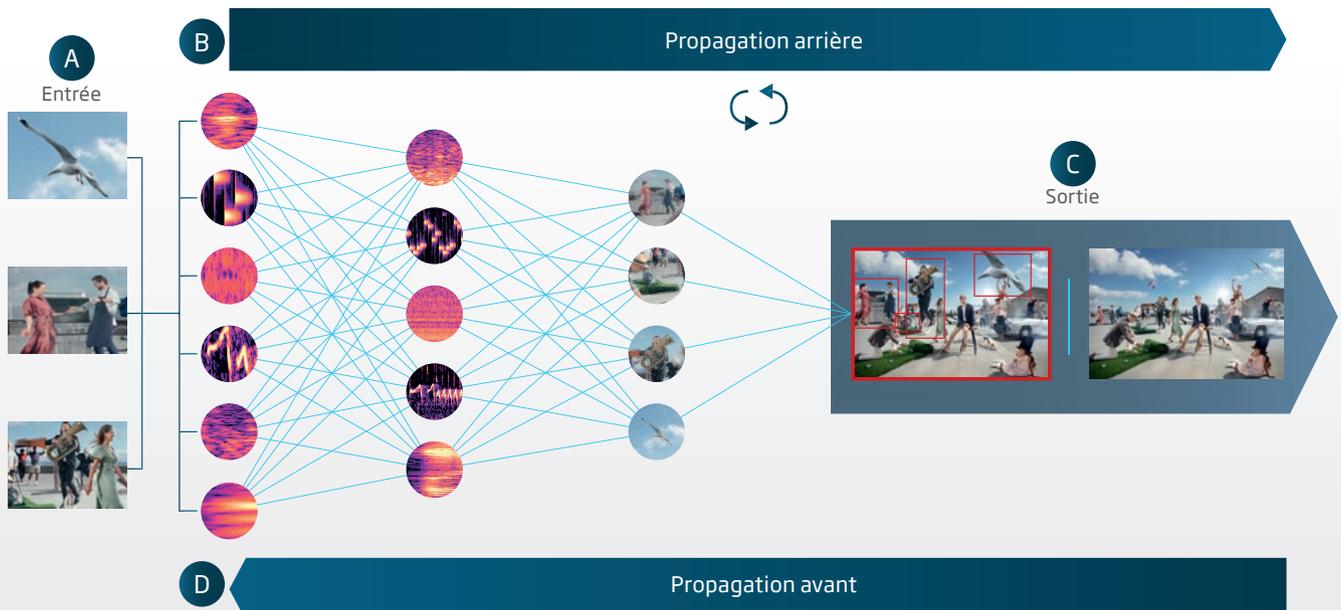


Figure 9. Voir le texte pour l'explication.

Lors de l'étape d'entrée (A), le Réseau Neuronal Profond effectue exactement ce que nous avons mentionné plus haut. Les neurones reçoivent des informations d'une scène sonore et les stockent. Ensuite, la propagation avant (B) utilise les données de l'entrée où chaque neurone transmet des informations à la couche suivante. La quantité d'informations transmises dépend de la force des connexions que les neurones entretiennent les uns avec les autres. Cela conduit à la prédiction de sortie (C) des objets que le RNP pense devoir mettre en valeur et supprimer de la scène sonore. Cependant, tout comme une personne apprenant une nouvelle compétence pour la première fois, le RNP fait des erreurs. Il peut, par exemple, accorder trop d'importance à la mouette (figure 9, première image sous le C).

Étant donné qu'il s'agit d'une forme d'apprentissage supervisé, nous informons le RNP qu'il a commis une erreur et que des changements doivent être effectués. Cette action dirige le processus de propagation arrière (D), au cours duquel le RNP ajuste les connexions individuelles entre chaque neurone pour mieux supprimer la mouette la prochaine fois. Ce processus est ensuite répété pour toutes les scènes sonores et par conséquent, le Réseau neuronal profond commence à identifier les caractéristiques de chaque objet pour mieux les distinguer. Dans le même temps, les scientifiques ajusteront également les caractéristiques spécifiques du RNP, par exemple la vitesse à laquelle le Réseau neuronal profond apprend. Cela forme une symbiose naturelle entre l'incroyable adaptabilité du RNP et les connaissances approfondies des scientifiques pour mieux atteindre notre objectif de fournir un code neural de bonne qualité. Au fil du temps, tandis que les 4 étapes

se répètent sur l'ensemble des 12 millions de scènes sonores que nous avons apportées au RNP, sa capacité à mettre en valeur les objets significatifs et non significatifs s'améliore jusqu'à ce que les scientifiques soient satisfaits. Cela est démontré par la performance de plus en plus précise dont font part les images en sortie (figure 9 (C)).

3. Test

Une fois que le RNP a été pleinement développé, il est temps de tester ses performances avec des données auxquelles il n'a pas été exposé auparavant, lors de l'étape d'entraînement. Cette étape est critique puisqu'il est possible que certains RNP ne présentent pas de bonnes performances.

Les réseaux neuronaux peuvent avoir des performances exceptionnelles lorsqu'il s'agit de réaliser la tâche en cours mais le RNP peut être trop strict et spécifique aux données pour lesquelles il a été entraîné. Par exemple, comme indiqué ci-dessus, le RNP a appris à quel point supprimer la mouette. Et s'il y avait un pigeon ? À ce moment précis, cela n'est sans doute pas significatif pour la personne qui écoute, mais le Réseau neuronal profond peut décider de mettre en valeur le pigeon, même s'il supprime la mouette. Nous nous confrontons alors au problème du surentraînement, lorsque le RNP est entraîné pour devenir très précis, mais qu'il ne parvient pas à généraliser au monde réel. En d'autres termes, il sait uniquement gérer les données avec lesquelles il a été entraîné mais il n'est pas à la hauteur lorsqu'il est exposé à de nouveaux défis. Par ailleurs, il est également possible qu'il ne fonctionne pas correctement sur les données avec lesquelles il a été entraîné, en les rendant trop ambiguës (figure 10).

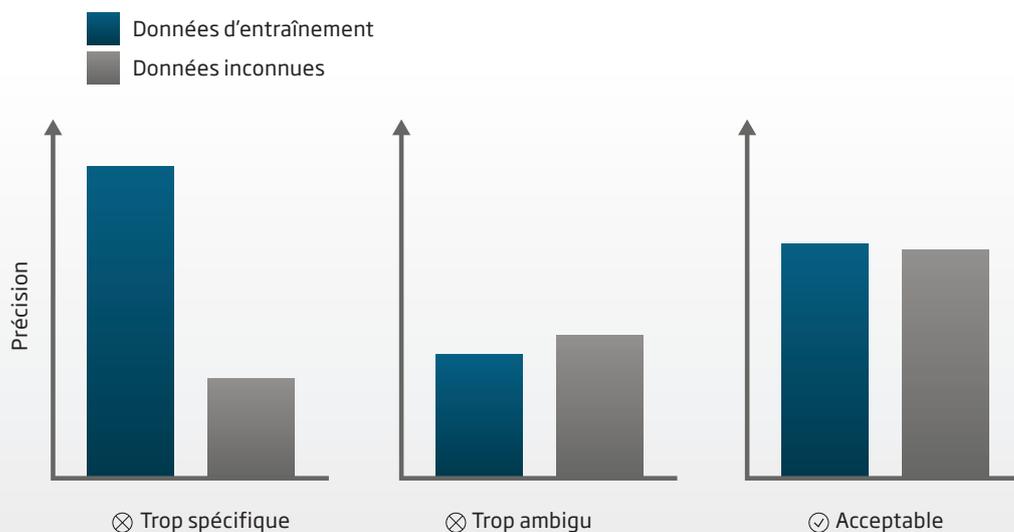


Figure 10. Précision du RNP à l'égard des données d'entraînement et des données inconnues. Haute précision sur les données d'entraînement mais faible sur les données inconnues - le RNP est trop spécifique. Faible précision sur les données d'entraînement et les données inconnues - le RNP est trop ambigu. Haute précision sur les données d'entraînement et les données inconnues - le RNP est acceptable.

Afin d'éviter ces problèmes, l'étape de test s'avère cruciale pour nous permettre d'avoir confiance dans les performances du RNP dans le monde réel, où les sons sont extrêmement variables et les possibilités illimitées. Au cours de cette étape de test, nous examinons le RNP avec les données de test susmentionnées. Dans l'idéal, le RNP présente de bonnes performances, même s'il est testé avec des données inconnues. Si ce n'est pas le cas, le Réseau neuronal profond doit être ajusté ou un autre RNP doit être conçu. Dans notre cas, chez Oticon, nous avons conçu plusieurs versions du RNP, chacune avec ses attributs propres, et nous avons sélectionné la version actuelle en fonction de sa performance au cours des étapes d'entraînement et de test.

Le dernier test auquel est soumis le RNP se déroule au moment où il est intégré dans l'aide auditive comme partie intégrante de MSI et testé sur les oreilles des personnes malentendantes. Les résultats de certains de ces tests seront décrits dans un livre blanc qui est disponible «L'audiologie d'Oticon More, 2020».

Un Réseau neuronal profond permet aux sons du monde d'être gérés précisément et automatiquement. Cela optimise la façon dont Oticon More rend les sons plus distincts et fonctionne parfaitement dans des environnements d'écoute variés. Avec son intelligence intégrée, Oticon More a appris à reconnaître tous types de sons, leurs détails et ce à quoi ils doivent ressembler dans l'idéal – dans le but de soutenir le cerveau de façon optimale.

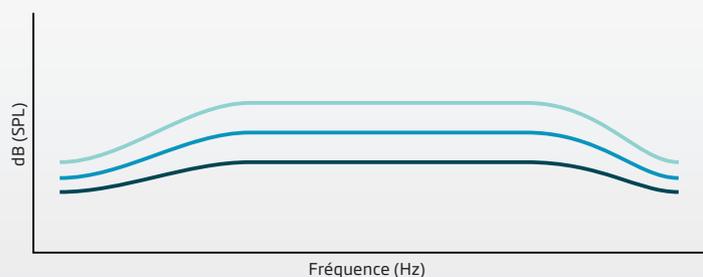


Figure 11. Égalisation pour les sons ajoutés par le Sound Enhancer. Ligne du haut : Réglage Riche. Ligne du milieu : Réglage Équilibré. Ligne du bas : Réglage Confort.

Sound Enhancer

Normalement, l'effet maximum d'un système de suppression de bruit doit être un compromis qui fonctionne pour tous les utilisateurs, même si certains utilisateurs auraient préféré que plus de bruit soit réduit et que d'autres utilisateurs trouvent que trop de son a été réduit. Le traitement du son dans les aides auditives doit veiller à ce que l'utilisateur soit capable de gérer l'environnement et qu'en même temps il dispose d'une bonne perception de la scène sonore pour profiter de l'ambiance.

Le Sound Enhancer fournit des détails sonores dynamiques lorsque la suppression du bruit est active – surtout dans les environnements difficiles – ce qui permet à la sortie d'être individualisée.

Le réglage Confort peut être choisi pour profiter de l'effet complet du système de suppression du bruit, tandis que le réglage Riche peut être choisi pour un fort niveau de connexion à l'environnement et aux interlocuteurs à écouter. Il s'agit d'une option à personnaliser selon les préférences de l'utilisateur.

Dans le plan de traitement, le Sound Enhancer est appliqué après le Spatial Clarity Processing et le Neural Clarity Processing. Le Sound Enhancer analyse la suppression dynamique effectuée par le Spatial Balancer et le RNP, puis il calcule la quantité de son à ajouter au signal selon les réglages d'Oticon Genie 2. Le Sound Enhancer suit la suppression adaptative globale effectuée par le Spatial Balancer et le RNP. Il ne s'adapte pas en permanence aux changements mineurs, mais aux changements généraux plus importants dans l'environnement sonore.

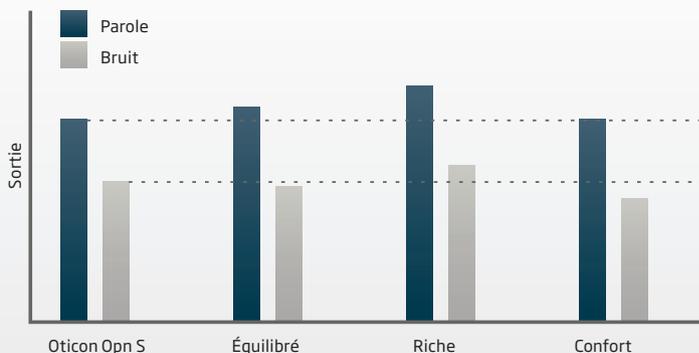


Figure 12. Relation entre les trois différents réglages du Sound Enhancer par rapport à Oticon Opn S.

Les détails ajoutés sont principalement fournis dans la zone 1-4 kHz. Cela signifie qu'il va principalement améliorer les sons de la parole. Cependant, le compromis réside dans le fait qu'il améliore également les autres types de sons. La figure 11 montre une plus grande emphase sur la région de la parole avec une transition vers les zones des basses et des hautes fréquences.

Cette égalisation signifie que les moyennes fréquences (1-4 kHz) avec le plus d'indices de parole auront plus de poids que les basses et les hautes fréquences. En raison de cette égalisation et le fait que le bruit relève souvent des basses fréquences et les indices de parole à des moyennes fréquences, un contraste légèrement meilleur sera créé entre la parole et le bruit lorsque le son est ajouté au signal.

Équilibré constitue le réglage par défaut qui convient à la plupart des personnes. Riche est le réglage qui fournit le plus de son avec le meilleur niveau de détail pour les utilisateurs qui préfèrent avoir un fort niveau de connexion à l'environnement et aux interlocuteurs à écouter. Confort constitue le réglage avec le moins de son et fournit plus de confort pour les utilisateurs qui préfèrent voir leur effort d'écoute légèrement réduit en soumettant subtilement l'environnement aux locuteurs à écouter.

La figure 12 présente la sortie d'Oticon More comparée à celle d'Oticon Opn S. Le but de cette illustration est de montrer la relation entre les différents réglages, mais les barres ne peuvent pas être traduites automatiquement en quantités familières puisque la fonction est dynamique.

Perspective

MoreSound Intelligence analyse la scène sonore intégrale, applique des réglages personnels pour chaque utilisateur, organise précisément les sons autour de l'utilisateur, et utilise le RNP pour créer du contraste entre les sons identifiés. Tout cela est effectué de façon très rapide et précise.

MoreSound Intelligence est basé sur la philosophie BrainHearing™ d'Oticon. Il donne accès à la scène sonore intégrale, au sein de laquelle les sons individuels se démarquent clairement les uns des autres et fournit au cerveau le code neural de qualité dont il a besoin.

Références

Akeroyd, M. A. (2014). An Overview of the Major Phenomena of the Localization of Sound Sources by Normal-Hearing, Hearing-Impaired, and Aided Listeners. *Trends in Hearing Vol. 18*, pp. 1-7.

Brændgaard, M. 2020. La plateforme Polaris. Document technique Oticon

Hausfeld, L., Riecke, L., Valente, G., & Formisano, E. 2018. Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *NeuroImage, 181*, 617-626.

Leuba G; Kraftsik R (1994). "Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age". *Anatomy and Embryology, 190* (4): 351-366. doi:10.1007/BF00187293. PMID 7840422. S2CID 28320951.

Løve, S. 2020. Adaptation optimale d'Oticon More. Livre blanc Oticon

Man, B. and Ng, E. 2020. BrainHearing – La nouvelle perspective. Livre blanc Oticon

O'Sullivan, J., Herrero, J., Smith, E., Schevon, C., McKhann, G. M., Sheth, S. A., ... & Mesgarani, N. 2019. Hierarchical Encoding of Attended Auditory Objects in Multi-talker Speech Perception. *Neuron, 104*(6), 1195-1209.

Puvvada, K. C., & Simon, J. Z. 2017. Cortical representations of speech in a multitalker auditory scene. *Journal of Neuroscience, 37*(38), 9189-9196.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in systems neuroscience, 7*, 31.

Squire, L. R. (2009). Memory and brain systems: 1969-2009. *Journal of Neuroscience, 29*(41), 12711-12716.

