

# Edge AI de Starkey intègre un réseau neuronal profond (DNN)



**Terence Betlehem, Ph.D., Parth Mishra, MSc., Jingjing Xu, Ph.D., Daniel Marquardt, Ph.D., Martin McKinney, Ph.D.**

---

## Introduction

Les techniques d'apprentissage automatique sont couramment utilisées dans le secteur des aides auditives, notamment pour l'identification des différents types d'environnements acoustiques, surtout en présence de parole. Ces algorithmes sont généralement entraînés sur un ordinateur externe, puis une version allégée est intégrée à l'aide auditive, avec pour seules limites les capacités de traitement et de mémoire de l'appareil. Les réseaux neuronaux profonds (DNN), une branche de l'apprentissage automatique, offrent la possibilité de créer des algorithmes plus sophistiqués et plus précis. Les DNN essaient d'imiter la manière dont le cerveau traite les informations, par le biais d'un maillage de nœuds et de couches capable de décoder des informations après avoir été entraîné de manière intensive. Le concept des réseaux neuronaux profonds est largement répandu, mais toutes les marques d'aides auditives n'en tirent pas forcément parti.

En 2018, Starkey a lancé Livio AI, la première aide auditive au monde intégrant un capteur de mouvement. Cette avancée technologique a permis d'intégrer d'autres formes de traitement intelligent dans les aides auditives, ce qui n'avait jamais été fait auparavant. L'intégration d'un capteur additionnel apportait en quelque sorte un sens supplémentaire et a permis de mieux comprendre l'environnement de l'auditeur. Cette avancée a été rapidement adoptée par le secteur des aides auditives. En lançant Genesis AI, Starkey a ouvert la voie avec un processeur avancé incorporant un accélérateur DNN assurant un traitement en temps réel, sans sacrifier la puissance de traitement ni la durée de vie de la batterie.

Les données de sortie de ce DNN ont permis de dresser une image plus précise de l'environnement acoustique de l'auditeur, qui a servi au développement du Mode Edge+ et a permis d'améliorer l'expérience d'écoute lors de déplacements dans une scène sonore en constante évolution. Pour Edge AI, le traitement DNN est inséré dans le trajet du signal de l'aide auditive, permettant une meilleure distinction entre les signaux de la parole et du bruit.

## Les défis en matière de réduction du bruit

Pour faire face au problème du bruit, les aides auditives actuelles ont recours à des systèmes de réduction numérique du bruit. La réduction numérique du bruit peut être définie comme un « procédé visant à réduire la sortie de l'appareil auditif en présence de bruit » (*Bentler, 2006*). Qu'est-ce que le « bruit » ? Tout simplement, le bruit est un son indésirable. Quels sons sont indésirables, cela dépend de la perception et des préférences de chacun. L'environnement, l'humeur et les circonstances de l'utilisateur de l'aide auditive influencent cette perception. Le secteur des aides auditives a toujours cherché à développer des procédés d'atténuation de ce qui est perçu comme des sons indésirables.

Un algorithme de réduction du bruit qui convient à un patient ou à un environnement particulier peut ne pas être adapté à un autre patient ou un autre environnement (*Bentler, 2006*). Il a été démontré que la réduction numérique du bruit, lorsqu'elle était associée au traitement de microphones directionnels, offrait des avantages tels qu'une diminution de l'effort d'écoute (*Chong & Jenstad, 2018 ; Desjardins, 2015, Iverson et. al., 2023*) et une amélioration du confort d'écoute (*Chong & Jenstad, 2018*).

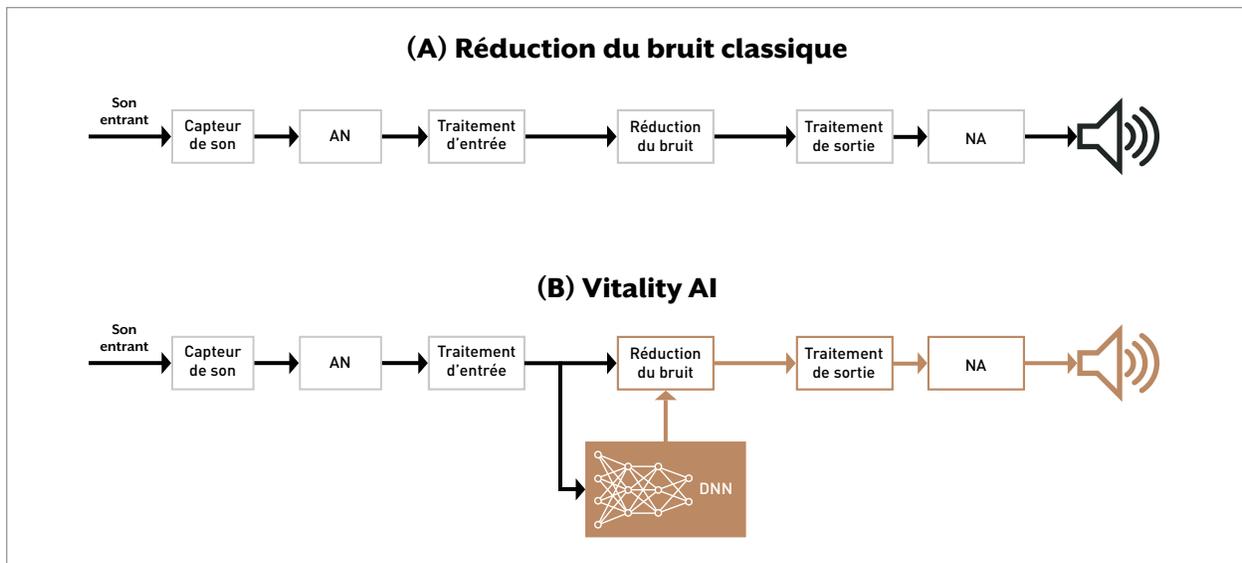


Figure 1 : Schéma comparant (a) la réduction de bruit conventionnelle et (b) Edge AI qui utilise un DNN pour guider la réduction de bruit.

La réduction numérique du bruit peut parfois s'avérer trop dynamique et ajouter des artefacts et de la distorsion. Les algorithmes classiques déterminent un masque à appliquer aux signaux vocaux bruités afin de séparer la parole du bruit. Pour réduire les artefacts, ce masque est soumis à un système de gestion de l'intensité. Ce système prévient les cas problématiques d'apparition des artefacts. Lorsqu'il détecte ces cas problématiques, le système diminue l'intensité de réduction du bruit afin d'éviter les artefacts et la distorsion indésirables.

Les approches classiques reposent sur une stratégie de minimisation des artefacts consistant à rendre moins dynamique la réduction du bruit en présence de parole. Starkey utilise cette stratégie depuis une dizaine d'années pour ses aides auditives. Bien que cette stratégie soit généralement efficace, plusieurs types de bruits rencontrés dans les situations d'écoute quotidiennes demeurent difficiles à gérer. La parole présente des fluctuations ou des modulations temporelles caractéristiques qui peuvent être utilisées pour la distinguer de nombreux types de bruit. Jusqu'à présent, la présence de la parole était détectée d'après des caractéristiques basées sur la modulation.

Malheureusement, il est difficile de distinguer la parole de certains types de bruits non stationnaires à partir de ces caractéristiques basées sur la modulation. Ces types de bruits comprennent le brouhaha, la frappe sur un clavier, les bruits de vaisselle et les aboiements de chiens. Ces bruits peuvent facilement être assimilés à de la parole et provoquer de mauvaises performances du système de gestion de l'intensité. Des bruits désagréables peuvent alors franchir le système de réduction du bruit et causer une gêne à l'utilisateur d'aides auditives.

Edge AI, un produit Starkey à la pointe de la technologie, s'appuie sur des innovations dans le traitement DNN et intègre un prédicteur de présence de la parole avec un système de gestion du son exclusif mieux à même de distinguer les composantes de la parole et du bruit, un point crucial pour l'application du procédé de réduction du bruit approprié (Figure 1).

#### Les avantages du nouveau traitement DNN

La présence de la parole, facteur de mérite à la base du système de gestion de l'intensité, peut être prédite avec plus de précision grâce à un DNN. Cela permet une détection plus précise des sources de bruit non stationnaires et donc une meilleure réduction du niveau de bruit pour ces types de bruit.

Un prédicteur de présence de la parole plus précis permet d'augmenter l'intensité de réduction de l'algorithme sans dégrader la qualité de la parole. Autrement dit, nous pouvons ainsi adopter une approche moins conservatrice et exécuter l'algorithme de réduction du bruit à un niveau plus dynamique. Pour les patients, c'est une aubaine, car ils bénéficient d'une expérience d'écoute plus calme, moins parasitée par les bruits de fond. Grâce à la nouvelle technologie Edge AI basée sur le DNN, les patients profitent d'un meilleur confort d'écoute dans les environnements bruyants, par rapport à ce qui existait jusqu'à présent. Comme expliqué ci-après, les études cliniques révèlent une préférence pour cette nouvelle technologie, particulièrement marquée dans certains scénarios.

Cette fonctionnalité repose sur notre technologie DNN brevetée. Les sons présentent des caractéristiques temporelles qui décrivent la manière dont les signaux sonores évoluent dans le temps. La modélisation de ces caractéristiques est cruciale pour une amélioration efficace du signal de la parole. Nos réseaux neuronaux profonds (DNN) intègrent des composants adaptés au traitement séquentiel des données. Grâce à cela, nos aides auditives peuvent interpréter le contexte sonore au fil du temps.

Cette meilleure interprétation du contexte permet une distinction plus précise entre la parole et le bruit. Les anciennes approches basées sur la modulation sont efficaces pour détecter les bruits stationnaires, mais elles ne le sont pas pour les bruits non stationnaires. Les bruits stationnaires sont par exemple le souffle d'un climatiseur ou le bourdonnement constant d'un réfrigérateur. Les bruits non stationnaires comprennent le brouhaha ambiant dans les espaces où il y a beaucoup de monde et le cliquetis de la vaisselle, des casseroles et des poêles dans la cuisine.

L'amélioration de la prédiction de la présence de la parole est utile dans les environnements bruyants complexes. De nombreux environnements se composent de bruits stationnaires et non stationnaires. Le système de gestion de l'intensité de réduction prenant de meilleures décisions pour les bruits non stationnaires, il est plus efficace dans les environnements présentant un mélange de bruits stationnaires et non stationnaires.

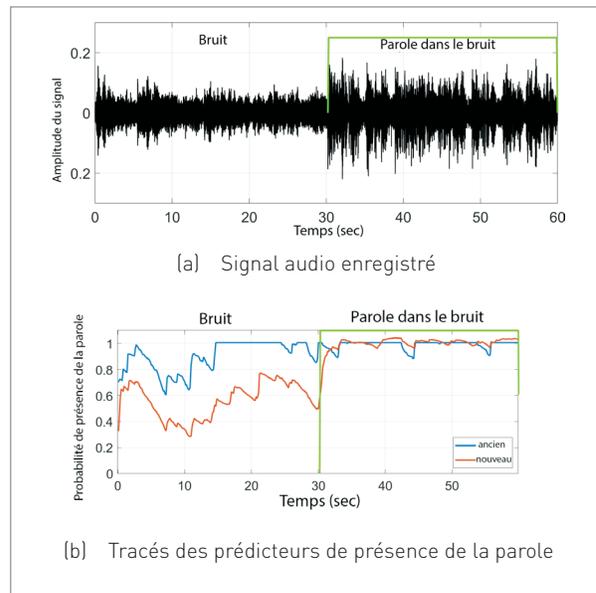


Figure 2 : Exemple du nouveau prédicteur de présence de la parole (rouge) par rapport à l'ancienne technologie (bleu). L'environnement d'écoute comporte plusieurs personnes parlant en arrière-plan dans une église avec réverbération, à 65 dB SPL et 0 dB de rapport signal/bruit. La personne ciblée commence à parler à partir de 30 secondes.

Un tel environnement peut par exemple comprendre à la fois un bruit de climatiseur (bruit stationnaire) et un bruit transitoire comme le passage d'une voiture sur un obstacle ou un tintement dans la cuisine.

Il a été constaté que cette nouvelle fonctionnalité améliorerait également l'écoute de la parole dans les mêmes environnements bruyants. La réduction du bruit basée sur le DNN est également supérieure à l'approche traditionnelle lorsque des paroles concurrentes sont présentes en arrière-plan, comme le brouhaha que l'on entend dans les espaces bondés. Dans de tels cas, l'algorithme basé sur le DNN permet de distinguer la parole cible du bruit de fond, ce qui accroît l'efficacité de la réduction du bruit. La figure 2 compare le nouveau prédicteur de présence de la parole à l'ancienne technologie. L'environnement sonore est l'intérieur d'une église : plusieurs personnes discutent en arrière-plan. En raison de l'acoustique de l'église, il y a beaucoup d'écho et de réverbération : le son rebondit sur les murs. Dans les 30 premières secondes, seul le bavardage en arrière-plan est présent. Le signal audio enregistré est représenté sur la figure 2(a). Le bavardage en arrière-plan est présent pendant toute la durée et la personne ciblée commence à parler au bout de 30 secondes. La figure 2(b) montre les tracés des prédicteurs de présence de la parole.

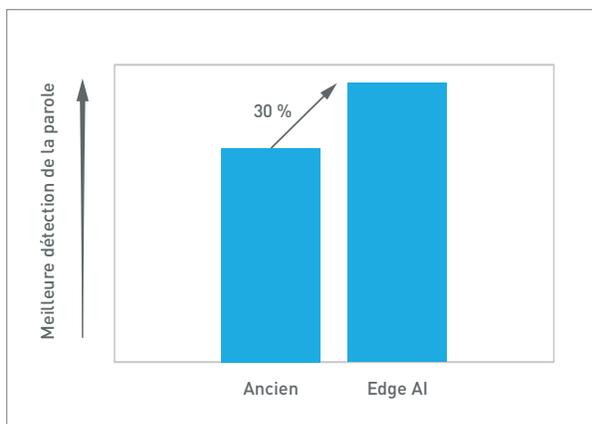


Figure 3 : L'algorithme amélioré par le DNN est 30 % plus performant que les approches classiques basées sur la modulation pour la parole dans le bruit, lorsqu'il est testé sur 12 types de bruits courants.

Pendant les 30 premières secondes, le nouveau prédicteur (tracé rouge) catégorise mieux le bavardage d'arrière-plan comme du bruit que l'ancien prédicteur (tracé bleu). À partir du début de la parole, après 30 secondes, le nouveau prédicteur catégorise plus nettement le son comme étant de la parole. Nous constatons que le prédicteur basé sur le DNN est plus précis, aussi bien pour le bavardage en arrière-plan, pour le bavardage sous forme de parole et pour la parole dans le bruit.

En moyenne, notre technologie DNN brevetée est plus performante en matière d'identification de la parole. La figure 3 montre qu'elle est 30 % plus performante que l'ancienne approche pour identifier la parole, pour une évaluation à 0 dB de rapport signal/bruit dans 12 environnements sonores difficiles (stationnaires et non stationnaires), dont un restaurant, un centre commercial, un aéroport, des enfants qui jouent et une grosse machine.

Si le nouvel algorithme donne de bons résultats dans les situations avec plusieurs interlocuteurs, il est exceptionnellement performant dans les scénarios avec un bruit de fond constant, par exemple une hotte de cuisine ou un climatiseur, et un bruit transitoire, par exemple le passage d'un véhicule sur un obstacle ou un tintement de casseroles dans la cuisine. Dans de tels scénarios, un grand nombre d'événements sonores transitoires sont catégorisés à tort comme de la parole par les anciens prédicteurs de présence de la parole, ce qui déclenche de façon intempestive l'algorithme de réduction du bruit qui se met alors à produire des niveaux élevés de bruit de fond.

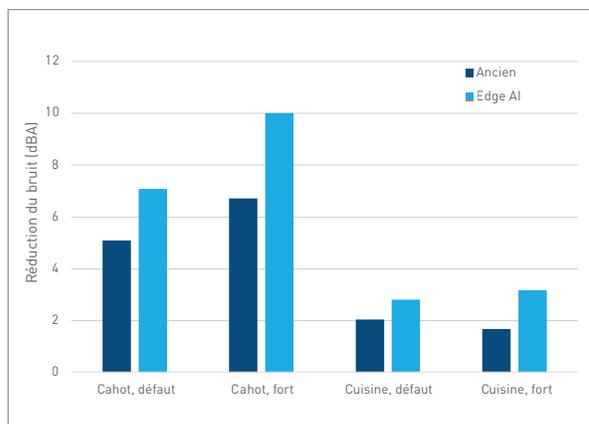


Figure 4 : Réduction du bruit par le nouvel algorithme basé sur le DNN par rapport à l'ancien algorithme basé sur la modulation. Les échantillons sont « Bruit routier avec cahots » et « Hotte de cuisine avec cliquetis de vaisselle ».

## Évaluation

Les comparaisons entre anciens et nouveaux paramètres de réduction du bruit montrent que la réduction du bruit basée sur le DNN offre des avantages par rapport aux anciens procédés.

### Réduction du bruit : résultats au banc d'essai

Comme expliqué précédemment, la réduction du bruit basée sur le DNN répond bien à certaines situations d'écoute difficiles. Analysons à présent le nouvel algorithme DNN dans ces situations en termes de réduction du bruit.

L'examen des enregistrements montre une meilleure réduction du bruit avec notre nouvel algorithme basé sur le DNN. La figure 4 illustre la réduction du bruit pour deux échantillons : « Bruit routier avec cahots » et « Hotte de cuisine avec cliquetis de vaisselle ». La réduction du bruit a été appliquée en utilisant à la fois le réglage par défaut et un réglage fort disponible dans les aides auditives Edge AI. Des améliorations jusqu'à 2 dBA et 3,5 dBA apparaissent lorsque la nouvelle approche est appliquée, pour le réglage par défaut et le réglage fort, respectivement. Une amélioration plus subtile de 1,5 dBA de la réduction du bruit est constatée pour l'environnement « Hotte de cuisine avec cliquetis de vaisselle ». Les résultats de la figure 4 ont été obtenus à l'aide d'audiogrammes N3 (Bisgaard, N. et al., 2010) sur des aides auditives portées par un mannequin KEMAR. Des niveaux similaires de réduction du bruit sont visibles pour les audiogrammes N2 et N4.

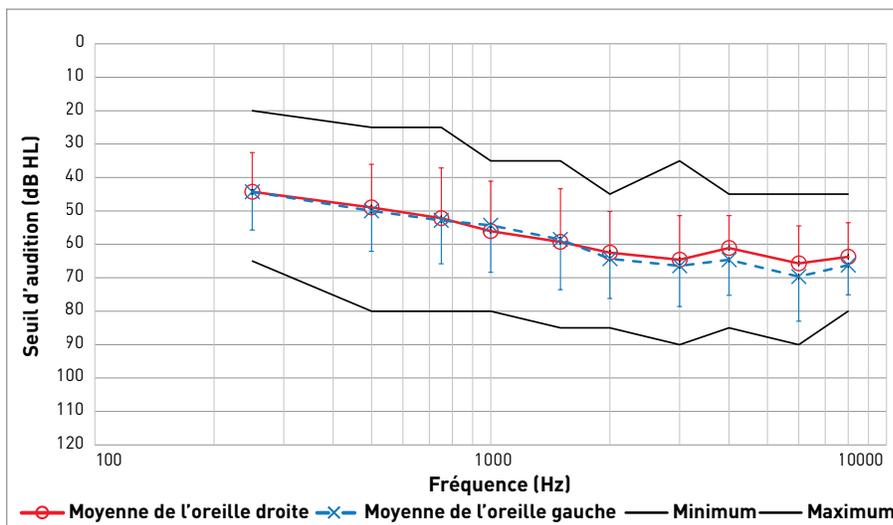


Figure 5 : Audiogramme moyen des 14 participants évalués. Les tracés et symboles rouges et bleus indiquent les seuils auditifs moyens pour l'oreille droite et l'oreille gauche, respectivement. L'écart-type est représenté en fonction de la fréquence pour les oreilles droite et gauche. Les tracés noirs indiquent les seuils d'audition minimum et maximum.

### Réduction du bruit : données perceptives

Pour évaluer de manière perceptive l'efficacité de la réduction du bruit basée sur le DNN, nous avons invité 14 utilisateurs d'aides auditives expérimentés (6 hommes et 8 femmes) à participer à une étude en laboratoire. Les participants étaient âgés de 52 à 85 ans, avec une moyenne de 71,6 ans et un écart-type de 9,3 ans. Ils ont été appareillés avec des aides auditives Edge AI utilisant e-STAT 2.0. Leurs adaptations ont été vérifiées par des mesures de l'oreille réelle. Les statistiques d'audiogramme sont présentées dans la figure 5.

Les algorithmes de réduction du bruit ont été évalués au moyen de tests de comparaison par paires. Les algorithmes de réduction du bruit nécessitant un court délai pour s'adapter à un environnement bruyant, il n'a pas été possible d'évaluer les algorithmes à l'aide d'essais en conditions réelles. Au lieu de cela, des stimuli d'essai ont été préenregistrés et présentés aux participants en laboratoire via un casque. Les échantillons audio ont été préenregistrés sur un mannequin KEMAR portant une paire d'aides auditives Edge AI, pour des scénarios sonores spatiaux reproduits à l'aide d'un réseau circulaire de 8 haut-parleurs. Les aides auditives étaient adaptées à la perte auditive de chaque participant malentendant.

Plusieurs scénarios ne comprenant que du bruit ont été examinés par comparaison par paires. Les préférences ont été calculées à partir des 42 réponses obtenues. Les résultats de cette étude portant sur la réduction du bruit pour les réglages par défaut et de forte intensité sont présentés à la figure 6.

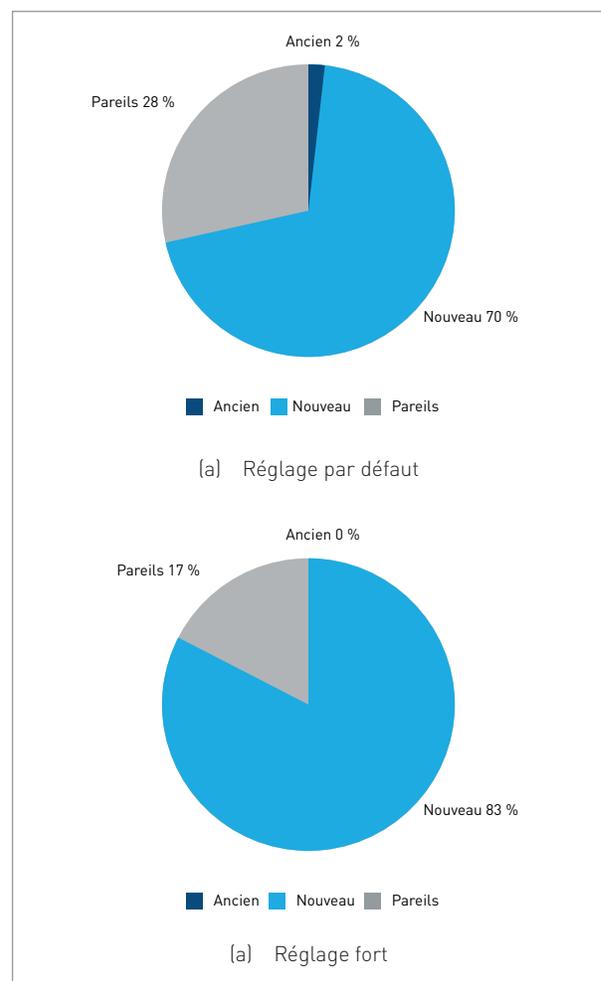


Figure 6 : Diagrammes circulaires représentant l'approche préférée des participants en matière de réduction du bruit pour (a) le réglage par défaut et (b) le réglage fort, dans le cadre d'un test d'écoute au casque avec comparaison par paires.

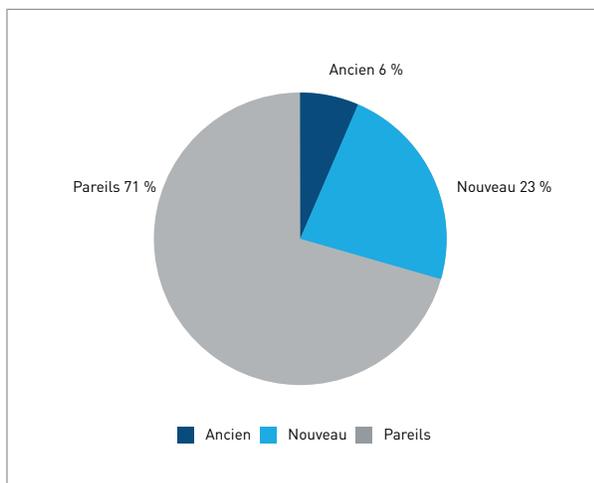


Figure 7 : Préférence des participants concernant les échantillons de parole dans le bruit dans le cadre d'un test d'écoute au casque avec comparaison par paires pour le réglage par défaut.

Les échantillons utilisés pour évaluer les performances étaient comme précédemment « Bruit routier avec cahots » et « Hotte de cuisine avec cliquetis de vaisselle », ainsi qu'un échantillon supplémentaire de « Tintement de vaisselle ». Nous constatons une forte préférence pour le nouvel algorithme basé sur le DNN en termes de réduction du bruit perçue.

Diverses conditions de parole dans le bruit ont également été étudiées avec du brouhaha (restaurant et bar) et des bruits de voiture présentés dans des conditions sonores difficiles de -5 dB et 0 dB de rapport signal/bruit. Les résultats présentés dans la figure 7 montrent une préférence globale des utilisateurs pour le nouvel algorithme par rapport à l'ancien. Cela suggère que l'ancien algorithme apporte un bénéfice suffisant dans des conditions de parole dans le bruit pour certains, tandis que d'autres ont bénéficié des nouvelles améliorations. Les préférences individuelles entrent également en ligne de compte. L'étude a montré que deux participants dont les moyennes en sons purs étaient similaires (55 dB HL et 57 dB HL pour des moyennes à 500, 1000, 2000 et 4000 Hz) présentaient des préférences différentes dans des conditions différentes de parole dans le bruit. Cette observation est similaire à celle d'une étude précédente dans laquelle Jaekel & Xu (2024) ont rapporté des différences de bénéfice en matière de rapport signal/bruit entre deux participants présentant des pertes auditives similaires, dans le cadre de l'évaluation des aides auditives dans un contexte de parole dans le bruit.

## Conclusion

En résumé, les données perceptives et les résultats obtenus au banc d'essai montrent des améliorations de la réduction du bruit et pour les scénarios de parole dans le bruit avec le nouvel algorithme basé sur le DNN, dans des environnements bruyants réalistes.

## Sources

1. Bentler, R. and Chiou, L.-K. (2006). Digital Noise Reduction: An Overview, *Trends in Hearing*, 10(2), 67-82.
2. Chong, F. Y., & Jenstad, L. M. (2018). A critical review of hearing-aid single-microphone noise-reduction studies in adults and children. *Disability and Rehabilitation: Assistive Technology*, 13(6), 600-608.
3. Desjardins J. L. (2016). The Effects of Hearing Aid Directional Microphone and Noise Reduction Processing on Listening Effort in Older Adults with Hearing Loss. *Journal of the American Academy of Audiology*, 27(1), 29-41. <https://doi.org/10.3766/jaaa.15030>
4. Iverson, S. R., Olson, M. M., Waite, B. (2023) Understanding Speech in Noise: A new beginning for an age-old problem. Starkey whitepaper [Link](#)
5. Bisgaard N, Vlaming MS, Dahlquist M. (2010) Standard audiograms for the IEC 60118-15 measurement procedure. *Trends Amplif*, 14(2):113-20. doi: 10.1177/1084713810379609. PMID: 20724358; PMCID: PMC4111352.
6. Jaekel, B.N. & Xu, J. (2024). Edge Mode+: On-demand processing improves speech recognition and listening effort in hearing aid users. *Audiology Practices*, 15(4), 8-16.

## Biographies des auteurs



**Terence Betlehem, Ph.D.**, est titulaire d'une licence en sciences, d'une licence en ingénierie et d'un doctorat de l'Australian National University. Sa thèse de doctorat était intitulée « Traitement des signaux acoustiques pour les environnements réverbérants ». Il a travaillé comme ingénieur de recherche chez Callaghan Innovation (anciennement Industrial Research Ltd) à Wellington, en Nouvelle-Zélande, et comme ingénieur de recherche principal chez Samsung Electronics. Depuis 2018, il est ingénieur de recherche principal chez Starkey.



**Parth Mishra, MSc.**, est titulaire d'un Master of Science en génie électrique de l'Université du Texas à Dallas, avec une spécialisation dans le traitement des signaux audio. Fort d'une vaste expérience du secteur, il a occupé, de 2018 à 2022, diverses fonctions chez BOSE Corporation. Il est actuellement ingénieur de recherche en traitement des signaux chez Starkey, et s'attache à perfectionner les algorithmes de traitement des signaux pour les appareils auditifs.



**Jingjing Xu, Ph.D.**, est chercheur au département de recherche clinique et audiolgique de Starkey. Avant de rejoindre Starkey en 2016, il était professeur assistant de recherche en audiologie à l'Université de Memphis. Il a obtenu une maîtrise en ingénierie acoustique à l'Université technique du Danemark et un doctorat en sciences et troubles de la communication à l'Université de Memphis. Ses recherches portent sur l'acoustique, la reconnaissance vocale, les mesures des performances des aides auditives et l'évaluation écologique momentanée.



**Daniel Marquardt, Ph.D.**, a obtenu un diplôme d'ingénieur en technologie des médias à l'université de technologie d'Ilmenau (Allemagne) et un diplôme d'ingénieur en traitement des signaux de la parole à l'université d'Oldenburg (Allemagne), en 2010 et 2015 respectivement. De 2015 à 2017, il a été chercheur postdoctoral à l'université d'Oldenburg, en Allemagne. Il est ingénieur de recherche en traitement numérique du signal (DSP) chez Starkey depuis 2018.



**Martin McKinney, Ph.D.**, est titulaire d'une licence en génie électrique de l'université de Tufts, d'une maîtrise en musique électroacoustique du Dartmouth College et d'un doctorat en sciences de la parole et de l'audition du Massachusetts Institute of Technology. Il travaille actuellement comme directeur des algorithmes et de la technologie des données chez Starkey, à Eden Prairie, Minnesota.