TECH PAPER

Introduzione a MoreSound Intelligence™

MoreSound Intelligence (MSI) è la nuova rivoluzionaria funzione "ispirata al funzionamento del cervello" presente negli apparecchi acustici Oticon More™. MSI è costituito da diverse funzionalità secondarie e questo documento tecnico mira a fornire una visione più approfondita di tutte le tecnologie presenti.

In primo luogo, ti verrà presentata una panoramica della funzionalità MSI completa, quindi approfondirai le singole funzionalità secondarie una per una nell'ordine in cui appaiono nel flusso di elaborazione del suono.

Alcuni punti salienti:

- Orecchio Virtuale Esterno – il nuovo modello pinna realistico del padiglione auricolare che fornisce supporto uditivo in ambienti semplici, con tre impostazioni presenti in Oticon Genie 2 per soddisfare le preferenze individuali degli utilizzatori
- Neural Clarity Processing – la rete neurale profonda integrata diret-tamente nella nuova piattaforma Polaris™, addestrata in fase di sviluppo con scene sonore reali per supportare il cervello in modo ottimale
- Sound Enhancer – guadagno dinamico principalmente per il parlato, fornito in ambienti complessi, con tre impostazioni in Oticon Genie 2 per soddisfare le preferenze degli utilizzatori

EDITORI DEL NUMERO

Mette Brændgaard,
Product Specialist, Product Marketing Support, Oticon A/S

Un particolare ringraziamento a **Brian Man Kai Loong,** *Audiologo della Ricerca Clinica, per il suo contributo alla parte relativa alla Rete Neurale Profonda*

oticon
life-changing **technology**

Gli ambienti sonori sono dinamici, complessi e imprevedibili, ed è compito del cervello gestire questa complessità; ascoltando e dando significato a tutto ciò. L'elaborazione effettuata dall'apparecchio acustico deve fornire un buon segnale sonoro per aiutare il cervello a interpretarlo. Tuttavia, ciò non avviene limitando la scena sonora tramite l'applicazione di sistemi di gestione del rumore e di direzionalità. Il cervello ha bisogno di accedere a tutte le informazioni nelle sue immediate vicinanze per aiutare il modo naturale di lavorare del cervello. Più prospettiva completa del panorama sonoro, al fine di ottenere di più dalla vita. (O'Sullivan et al., 2019; Hausfeld et al., 2018; Puvvda & Simon, 2017)

MoreSound Intelligence di Oticon introduce un salto di qualità nell'elaborazione del panorama sonoro dando accesso alla scena sonora completa in maniera chiara ed equilibrata, garantendo il corretto contrasto.

MoreSound Intelligence

MoreSound Intelligence (MSI) è una funzionalità avanzata che include al suo interno diverse funzioni secondarie. Nei paragrafi seguenti, questo documento tecnico descriverà tutte queste diverse funzioni.

La Figura 1 mostra i diversi passaggi e le funzioni secondarie presenti all'interno MSI. In primo luogo, la scena sonora viene scansionata e analizzata. Sulla base di tale analisi, insieme alle impostazioni nel software di adattamento (Oticon Genie 2), la scena sonora viene passata alle tecnologie Spatial Clarity Processing e Neural Clarity Processing. Il segnale può seguire due percorsi differenti a seconda della complessità dell'ambiente sonoro. Quello che otteniamo in uscita del blocco MSI è un segnale ripulito pronto per ulteriori regolazioni da parte dell'apparecchio acustico (ad es. Amplificazione). Per ulteriori informazioni consultare: Brændgaard, M. 2020. La Piattaforma Polaris. Oticon tech paper - sul flusso completo di elaborazione per Oticon More 2020 More, e Løve, S. 2020. Adattamento Ottimale di Oticon More. Oticon Whitepaper - sull'adattamento di Oticon More in Genie 2.

Durante tutto il processamento del segnale di MSI, l'elaborazione del suono in ingresso viene eseguita su 24 canali. Rispetto ai precedenti apparecchi acustici Oticon premium, il numero extra di canali offre una precisione doppia in una gamma di frequenze che include i canali di frequenza 1,5-5 kHz (Brændgaard, 2020), che sono i più importanti per i suoni del parlato.

Oltre ad avere più canali di elaborazione per una maggiore precisione, questi sono anche collegati. Ciò significa che durante l'elaborazione tramite la rete neurale profonda, tutti i canali possono vedere chec osa sta succedendo, e quindi che tipo di elaborazione. Si sta effettuando negli altri canali. Ciò riduce al minimo il rischio di artefatti creati da un tipo di suono classificato in modo errato in un solo canale. Riducendo al minimo gli artefatti si migliora la qualità del suono.

Scansione e Analisi

Per poter elaborare correttamente le diverse sorgenti sonore, MSI ha bisogno di conoscere i dettagli esatti della scena sonora. Inoltre, le scene sonore sono dinamiche e le sorgenti sonore si muovono e cambiano costantemente.

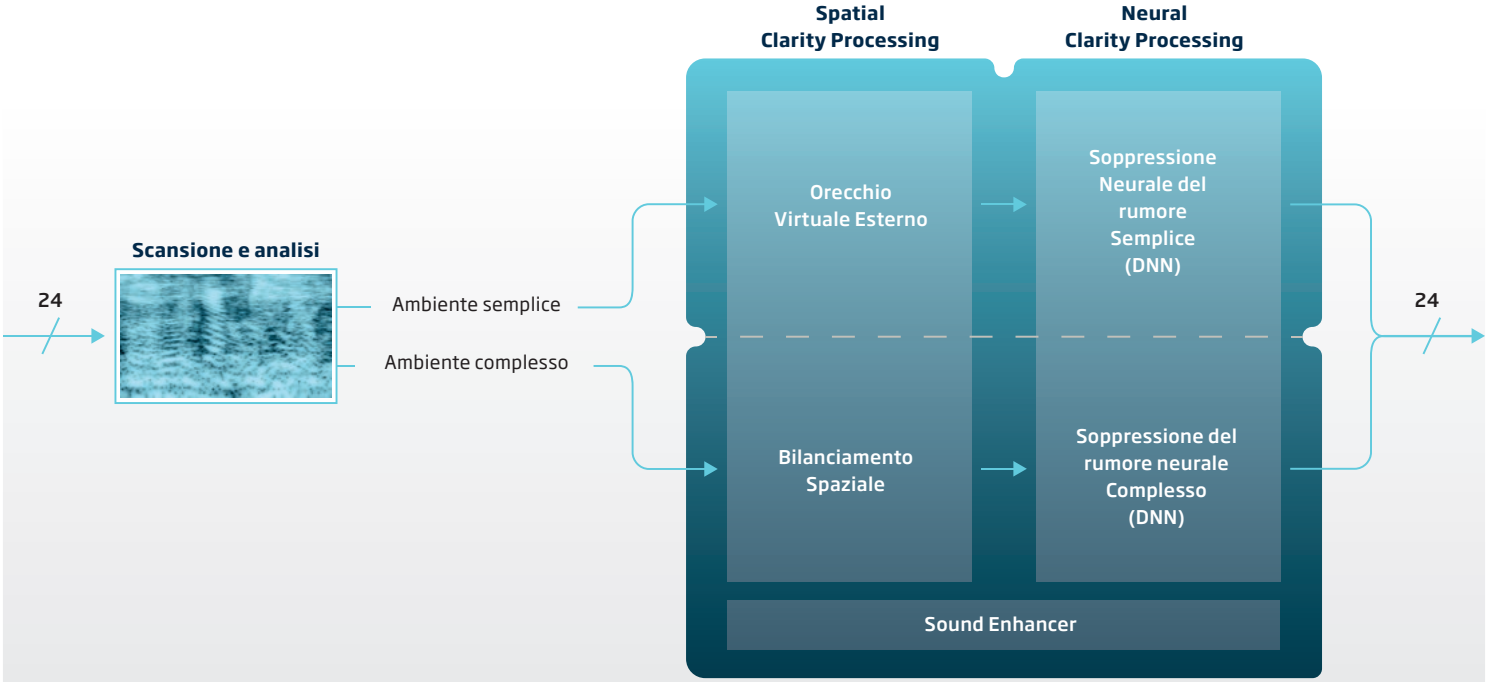

Figura 1: Lo schema di processamento di MoreSound Intelligence

Per garantire che tutti i dettagli vengano catturati, la scena sonora viene scansionata 500 volte al secondo, per mappare correttamente tutte le sorgenti. Sulla base di questa scansione e mappatura, MSI calcola il rapporto segnale/rumore (SNR) nonché i livelli di rumore.

Gli stimatori del SNR e del livello di rumore sono stati aggiornati in modo tale da funzionare su 24 canali, in più lo stimatore SNR lavora su un intervallo più ampio, e cioè da -10 a +15 dB SNR. L'SNR è il driver principale utilizzato per distinguere ambienti semplici e complessi, mentre il livello di supporto fornito dal sistema sarà determinato sia dall'SNR che dalle stime del livello di rumore. Il principio della relazione tra SNR, livello di rumore e aiuto fornito è rappresentato in figura 2.

Eventuali cambiamenti nella scena sonora verranno rilevati durante la scansione, ma solo i cambiamenti persistenti (più di 2 secondi) modificheranno la quantità di supporto fornito dall'apparecchio.

MoreSound Intelligence utilizza l'SNR per distinguere ambienti semplici e complessi, ma il dettaglio specifico in cui ciò avviene differisce da persona a persona. Il punto di confine tra ambienti semplici e complessi dipende dalle impostazioni del singolo paziente in Oticon Genie 2. Cioè, le impostazioni effettuate in Genie 2, che si basano sugli input del paziente durante la sessione di adattamento, determineranno a quale SNR deve presentarsi una scena sonora da elaborare come un ambiente facile o difficile.

Da un lato, se la scena sonora è classificata come un ambiente semplice, l'elaborazione segue il flusso (vedi figura 1) tramite l'Orecchio Virtuale Esterno e quindi Neural Noise Suppression - semplice. D'altra parte, se si determina che la scena sonora è un ambiente complesso, l'elaborazione segue il flusso che inizia con lo Spatial Balancer e si conclude con Neural Noise Suppression – complesso, con un ulteriore aiuto fornito dal Sound Enhancer.

L'analisi completa della scena sonora viene utilizzata anche dal Neural Clarity Processing per l'elaborazione con la rete neurale profonda (vedere più avanti in questo documento).

Spatial Clarity Processing

Essere in grado di collocare le sorgenti sonore nello spazio circostante è un'abilità molto importante che diventa più difficile quando è presente una perdita uditiva (Akeroyd, 2014).

Il padiglione auricolare ci aiuta a localizzare i suoni in tre dimensioni: distanza, verticale (su/giù) e orizzontale (davanti/dietro). A seconda che ci siano in gioco le differenze interaurali di tempo o le differenze interurali di livello, alcune frequenze sono più rilevanti rispetto ad altre per la localizzazione delle sorgenti. (Akeroyd, 2014).

Abbiamo tutti dimensioni dell'orecchio e forme del padiglione auricolare diverse e quindi, in base all'anatomia dell'orecchio, il suono subisce modifiche diverse quando entra nel condotto uditivo. Ad esempio, a causa della forma dell'orecchio esterno, alcune persone avranno più o meno focalizzazione frontale di altre.

Quando posizioniamo i microfoni dell'apparecchio acustico dietro l'orecchio, viene eliminata la capacità di utilizzare i segnali spaziali naturali forniti dal padiglione auricolare. Questa capacità deve essere ricreata dall'elaborazione del segnale nell'apparecchio acustico.

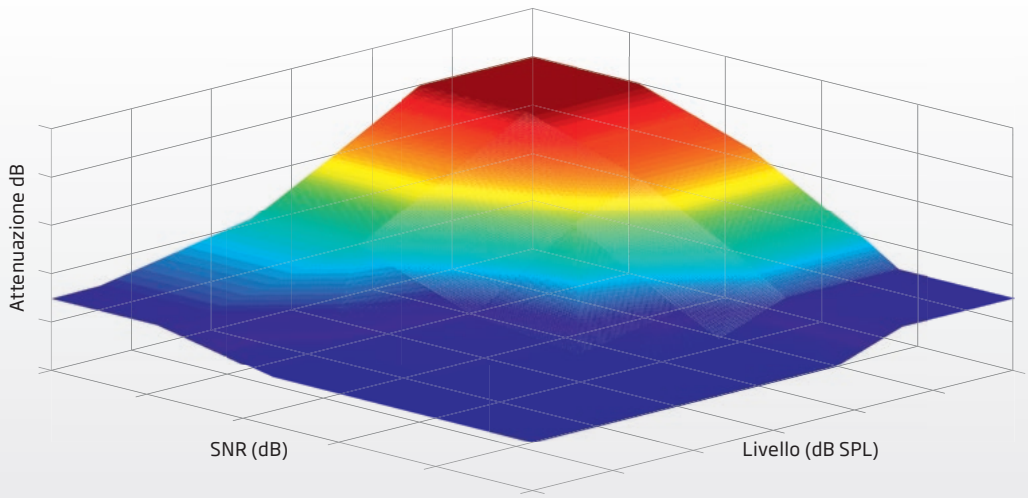

Figura 2. Principio di relazione tra SNR, livello di rumore e soppressione del rumore. Più basso è l'SNR e più alto è il livello di rumore, maggiore è l'aiuto fornito dal sistema.

Spatial Clarity Processing consiste in due diverse funzionalità, Virtual Outer Ear e Spatial Balancer, che aiutano a ricreare una corretta sensazione spaziale rispettivamente in ambienti semplici e complessi.

Orecchio Virtuale Esterno

L'Orecchio Virtuale Esterno (VOE: Virtual Outer Ear) sono tre diversi modelli di pinna reale che l'audioprotesista può impostare nel software di adattamento in base alle esigenze d'ascolto dell'utilizzatore.

VOE aiuta il paziente a ripristinare una corretta consapevolezza spaziale in ambienti d'ascolto semplici.

Per ottenere la migliore compensazione per il padiglione auricolare naturale tramite elaborazione del segnale, le funzioni di trasferimento correlate alla testa (HRTF) sono state misurate da diverse angolazioni sul piano orizzontale su 130 orecchie umane. L'indice di direttività (DI), che è una misura della quantità di direzionalità fornita delle singole orecchie in ciascuna banda di frequenza, può essere calcolato tramite l'HRTF. Un DI più alto equivale a un orecchio più direzionale.

I risultati di DI possono essere visti nella figura 3. La linea magenta scura è la media, mentre la linea blu è rappresentativa dell'orecchio con il DI più piccolo calcolato (e di conseguenza il più omnidirezionale). La linea grigia è invece l'orecchio con il DI calcolato più grande (orecchio con il focus più frontale). La differenza di DI tra le orecchie può essere vista principalmente tra 2 e 5 kHz e può essere una differenza piuttosto grande.

Le informazioni ottenuta dagli HRTF possono anche essere visualizzate come un diagramma polare, che mostra la risposta per uno specifico angolo/frequenza raggiunta dal VOE (figura 4). Di nuovo, la linea magenta scura è la media, quella blu è il DI più piccolo e quella grigia è il DI più grande.

La differenza tra le orecchie esaminate copre un range di termini di indice di direzionalità di 4 dB, da -2 dB a +2 dB, dove la maggior parte delle misurazioni si trova nell'area di 0,5 dB DI. Ciò significa che la maggior parte delle persone ottiene un'amplificazione naturale dall'orecchio esterno di circa 0,5-1 dB nell'area 2-5 kHz.

Di conseguenza, il VOE si basa su queste misurazioni per creare un modello di padiglione auricolare il più naturale e accurato possibile.

Le misurazioni hanno dimostrato che l'effetto del padiglione auricolare sul suono varia da un orecchio all'altro. Ciò significa che il modo in cui una persona è abituata a sentire i suoni sarà diverso, a seconda dell'anatomia dell'orecchio esterno. Inoltre, non è possibile effettuare una semplice misurazione dell'orecchio esterno e quindi dedurre come l'utilizzatore "sente". Tuttavia, poiché sappiamo che sarà diverso da individuo a individuo, il VOE ha tre diverse impostazioni con una messa a fuoco leggermente più o meno frontale (vedere la figura 5) e questa può essere impostata nel software di adattamento Oticon Genie 2 in base alle preferenze del paziente ipoacusico. La messa a fuoco leggermente più frontale viene creata facendo entrare meno suono proveniente da una direzione posteriore.

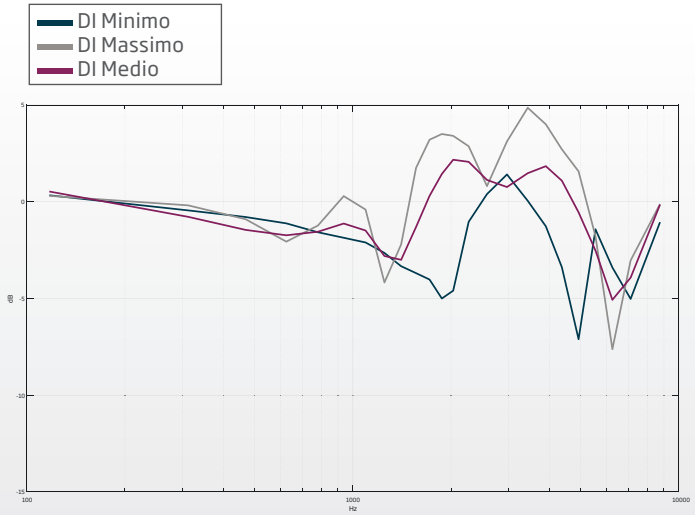

Figura 3. Indice di direzionalità per l'orecchio umano che mostra DI medio (magenta scuro), minimo (blu) e massimo (grigio).

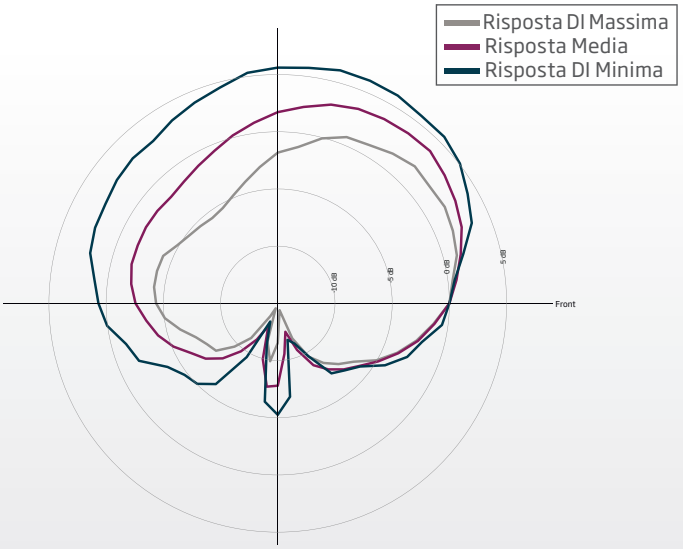

Figura 4. Diagramma polare per l'orecchio sinistro che mostra la direzionalità media dell'orecchio umano per le frequenze 2-5 kHz. In questo esempio, la persona è rivolta verso il lato destro.

Le tre diverse impostazioni si basano sulle seguenti misurazioni di indici di direzionalità (DI):

- Le misurazioni medie vengono utilizzate per l'impostazione Bilanciata (impostazione predefinita) che è progettata per creare il miglior rapporto tra l'udibilità di tutti i suoni nell'ambiente circostante e accesso alla voce frontale
- Le misurazioni più alte vengono utilizzate per l'impostazione Focalizzato (maggiore attenzione al parlato proveniente frontalmente)
- Le misurazioni più basse sono utilizzate per l'impostazione Consapevole (accesso a tutto l'ambiente circostante)

Le differenze tra le tre impostazioni sono di circa 0,5 dB DI*.

Spatial Balancer
Spatial Balancer è una funzionalità più potente del VOE e si attiva in ambienti complessi. Spatial Balancer bilancia rapidamente le sorgenti sonore presenti nell'ambiente.

Spatial Balancer è dotato di un diagramma polare omnidirezionale e di uno back-cardioide forniti dai due microfoni. Il microfono omnidirezionale raccoglie tutti i suoni della scena sonora inclusi quelli frontali, che spesso sono i più importanti per l'utilizzatore. Il microfono back-cardioide invece raccoglie tutti i suoni presenti nel panorama tranne quelli frontali. I due segnali vengono costantemente confrontati per definire il posizionamento delle sorgenti di rumore. Spatial Balancer utilizza un beamformer MVDR a variazione minima per creare diagrammi polari dedicati e ottenere il bilanciamento ottimale di una data scena sonora. La direzione dei punti di nullo nei diagrammi polari è posizionata verso le sorgenti di rumore più dominanti allo scopo di attenuare il rumore e mantenerlo sullo sfondo della scena sonora.

Il sistema può creare singoli punti di nullo per ciascuno dei 24 canali di frequenza in cui lavora l'apparecchio acustico, consentendo, in linea di principio, a Spatial Balancer di controllare simultaneamente 24 sorgenti sonore (48 in totale intorno alla testa). Ogni punto di nullo viene aggiornato 125 volte al secondo.

Spatial Balancer aumenta l'SNR sopprimendo le singole sorgenti di rumore, posizionandole sullo sfondo e creando così una scena sonora bilanciata.

L'aggiunta al numero di canali e quindi al numero di possibili punti di nullo rispetto ai precedenti prodotti Oticon rende il sistema più preciso nel "mirare" la sorgente sonora. Il punto di nullo può essere più profondo, il che significa che la sorgente di rumore può essere spostata più in secondo piano se necessario.

Neural Clarity Processing – Rete Neurale Profonda

Neural Clarity Processing è dove risiede la tecnologia principale degli apparecchi acustici Oticon More™: la Deep Neural Network (DNN) – rete neurale profonda. La DNN, per una data scena sonora, ha imparato a riconoscere cosa dovrebbe essere messo in primo piano (suoni di interesse con molte informazioni) e cosa invece lasciare in sottofondo (suoni di minore interesse con meno informazioni). In questo modo si crea una migliore chiarezza e un miglior contrasto tra le sorgenti sonore. Questa sezione del documento introdurrà alle reti neurali e all'implementazione di una rete neurale profonda negli apparecchi acustici da parte di Oticon.

Potresti non averci mai pensato, ma le reti neurali vengono utilizzate in tutto il mondo per affrontare grandi problematiche di tutti i giorni. In base alle tue azioni precedenti, una rete neurale implementata in un dispositivo elettronico potrebbe suggerire una playlist

** AI ponderato: Indice di Articolazione ponderato. Ciò significa che nel calcolo del DI, le frequenze vengono pesate in base all'impor-tanza della comprensione del parlato. Le frequenze più basse e più alte hanno un peso inferiore rispetto alle frequenze medie.*

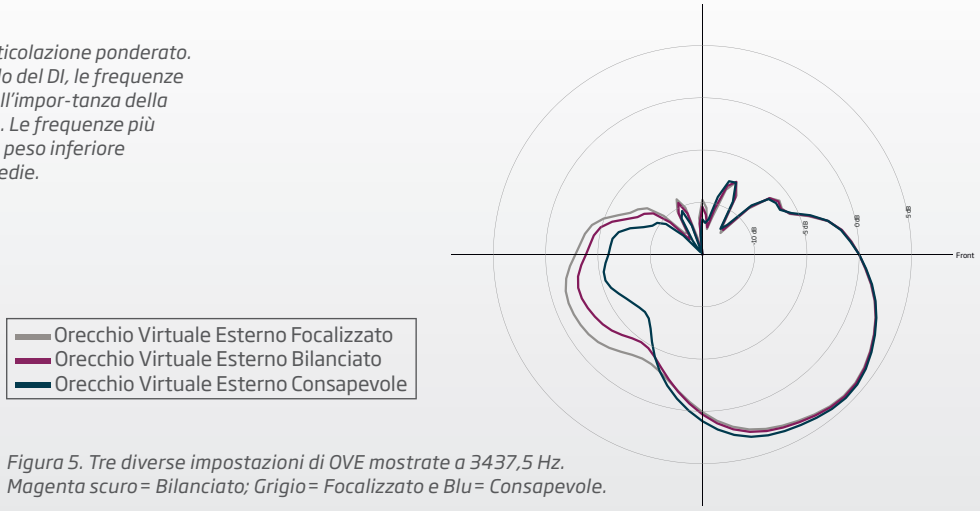

Figura 5. Tre diverse impostazioni di OVE mostrate a 3437,5 Hz. Magenta scuro = Bilanciato; Grigio = Focalizzato e Blu = Consapevole.

o consigliare cose da acquistare. Tali sfide complesse e multifattoriali possono essere affrontate più facilmente da una rete neurale che dagli approcci tradizionali che si basano sulla definizione di regole stabilite.
Prima di approfondire come funziona la DNN nell'apparecchio acustico, diamo una breve occhiata a ciò che è nel cervello e che il DNN cerca di imitare durante l'apprendimento. Useremo un esempio visivo di distinzione tra un gatto e un cane in quanto è di più facile concezione.

Nel nostro cervello abbiamo aree, all'interno della corteccia, specializzate nell'elaborazione di segnali visivi come la corteccia visiva primaria, che contiene oltre cento milioni di neuroni e ancor più connessioni tra di loro. (Leuba G. & Kraftsik R, 1994.) Noi come esseri umani, nel corso della nostra vita, siamo diventati sorprendentemente efficienti nel dare un senso a ciò che vediamo, ma il cervello prima di poterlo fare ha bisogno di imparare. Riuscire a distinguere tra un gatto e un cane è normalmente abbastanza facile per un adulto, ma non è necessariamente facile capire come lo facciamo.

Quali sono le regole che abbiamo stabilito nel nostro cervello che ci rendono capaci di gestire questo compito? Forse è qualcosa che riguarda il naso o le orecchie, o forse è qualcosa di completamente diverso di cui non siamo nemmeno consapevoli. Ci rendiamo subito conto che siamo destinati a perderci in un pantano di eccezioni nel tentativo di verbalizzare queste regole. Lo stesso vale quando si cerca di definire regole per il riconoscimento dei suoni, ma questo è ciò che gli scienziati hanno fatto per molto tempo quando si sono sviluppati sistemi di riduzione del rumore negli apparecchi acustici.

Gestione delle sorgenti rumorose
Fino ad ora la riduzione del rumore negli apparecchi acustici è stata gestita da algoritmi artificiali che definivano ciò che era rilevante, il parlato, e ciò che era rumore, applicando regole sulla modulazione del suono e sulla direzione da cui proveniva la sorgente sonora. Ciò potrebbe spesso significare che solo il discorso frontale era considerato rilevante. Una tale ipotesi potrebbe non essere vera in tutti i casi, poiché ignora l'importanza dell'accesso del cervello alla scena sonora completa e limita la nostra capacità di localizzare i suoni al di fuori del suo raggio d'azione effettivo. (Per ulteriori informazioni sull'importanza dell'accesso del cervello alla scena sonora completa, consultare Man, B. and Ng, E. 2020. BrainHearing - La nuova prospettiva. Whitepaper Oticon.)

La distinzione tra oggetti è un problema multifattoriale. Il nostro cervello fa leva non solo sulle caratteristiche identificabili del segnale su cui ci stiamo concentrando, ma anche sulla memoria semantica a lungo termine (Rönnberg et al., 2013). Dopo aver usato le nostre orecchie per tutta la vita, il nostro cervello ha immagazzinato una rappresentazione di molti, moltissimi suoni diversi. Il nostro cervello ha effettivamente acquisito, attraverso errori ed esperienza, un approccio metodologico unico per distinguere tra suoni rilevanti e non rilevanti. Le reti neurali hanno adottato questo approccio. La struttura di un DNN è in parte ispirata da come è organizzato il nostro cervello, ovvero i neuroni e le loro corrispondenti sinapsi. La rete neurale utilizza l'apprendimento iterativo dall'enorme quantità di dati del mondo reale (vedere più in basso per maggiori informazioni sul processo di addestramento) per stabilire la conoscenza del suono e come elaborarlo. Viene quindi applicato un apprendimento iterativo della DNN

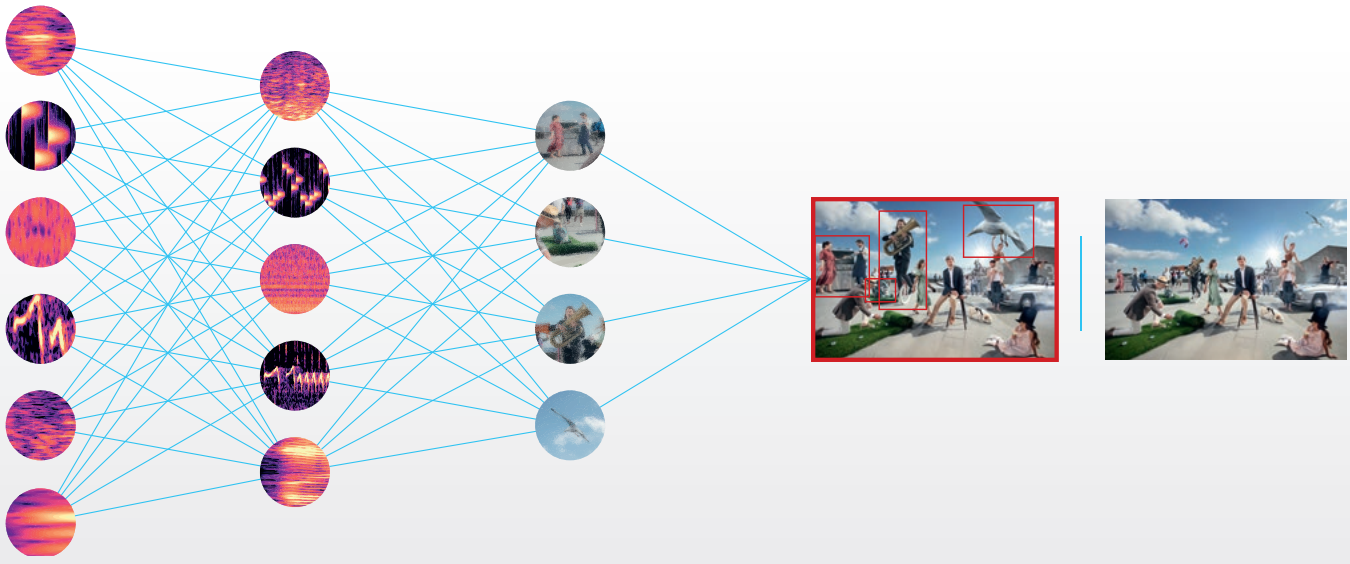

Figura 6. Illustrazione concettuale di una rete neurale.

invece di seguire un set rigoroso di regole prestabilite basate su algoritmi creati dall'uomo. Questo approccio porta l'elaborazione del suono e la gestione del rumore fuori dal laboratorio, nel mondo reale.

Cos'è una rete nuerale?

Le reti neurali sono una classe specifica di algoritmi nell'ambito della disciplina più generale dell'apprendimento automatico. L'idea dell'apprendimento automatico è quella di prendere una grande quantità di dati, noti come campioni di addestramento, per poi sviluppare un sistema in grado di apprendere da essi. L'unicità delle reti neurali deriva dalla loro somiglianza architettonica con il cervello. Nel contesto delle reti neurali, esiste un'unità di base chiamata neurone. Lo scopo di un neurone, proprio come un neurone relè nel cervello, è ricevere informazioni, memorizzarle e infine trasmetterle al neurone successivo. Un gruppo di neuroni forma uno strato e, più strati specializzati e interconnessi, formano la rete neurale costituita da un layer di input all'inizio, layer nascosti al centro e layer di output alla fine. Questa forma la classe più elementare di reti neurali (figura 6).

La Rete Neurale Profonda di Oticon

Generalmente, sviluppare una Rete Neurale Profonda (DNN) richiede 3 fasi fondamentali: (1) Scopo, (2) Addestramento e Apprendimento e (3) Prova (vedi figura 7).

1. Scopo

Innanzitutto, esaminiamo il problema e definiamo ciò che vogliamo offrire agli utilizzatori di apparecchi acustici. Nel nostro caso, vogliamo migliorare l'accesso degli ascoltatori con problemi di udito alla scena sonora completa, creando un buon codice neurale per supportare al meglio le fasi di Orientamento e Focus nel cervello (Man & Ng, 2020).

Per raggiungere tale scopo, dobbiamo considerare la natura dei dati (le scene sonore). Quali sono le caratteristiche di suoni diversi come la voce e il rumore.

La voce è dinamica e cambia continuamente anche quando proviene da una sola persona. La caratteristica fondamentale di un segnale vocale è che esiste un certo grado di continuità. Ad esempio, se hai ascoltato la voce del tuo amico, è improbabile che il tono della sua voce cambi improvvisamente. Al contrario, il rumore circostante può essere un misto di bicchieri tintinnanti, voci di persone diverse sullo sfondo che differiscono tutte non solo per intonazione, ma anche per come variano nel tempo. Questo è il motivo per cui abbiamo progettato una rete neurale specializzata nella gestione di tali segnali dinamici - A Gated Recurrent Unit (GRU), che è una variante di reti neurali Long-Short Term Memory (LSTM).

Le reti neurali Long-Short Term Memory (LSTM), come suggerisce il nome, hanno qualcosa a che fare con la memoria, sia a breve che a lungo termine.

La memoria è definita in psicologia come la facoltà di codificare, memorizzare e recuperare informazioni (Squire, 2009). Come sappiamo, la memoria serve come sistema di archiviazione delle informazioni che può essere preservata nel tempo. Inoltre, vi si può accedere per facilitare un processo decisionale più efficiente. LSTM e GRU operano in base a questo principio. L'idea è di "collegare" la rete neurale nel tempo consentendo alla rete di trasmettere le informazioni a se stessa (figura 8).

Ne risulta un algoritmo che non solo riconosce le diverse caratteristiche che i suoni hanno in un singolo momento, ma anche come queste caratteristiche del suono variano nel tempo. La capacità di incorporare le informazioni nel tempo è precisamente l'attributo unico di cui abbiamo bisogno per analizzare un segnale dinamico come i suoni. La DNN è formata da un later di input, layer nascosti in cui l'elaborazione non è visibile e un layer di output con il risultato finale dell'elaborazione che possiamo sentire. I layer di input e output hanno 24 neuroni corrispondenti ai 24 canali di elaborazione.

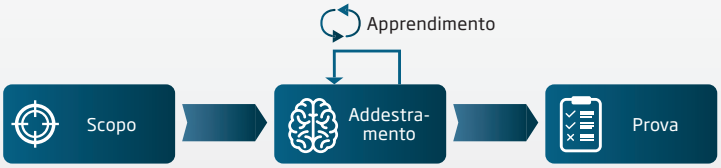

Figura 7. Fasi di sviluppo di una rete neurale - 1 Scopo, 2 Addestramento e Apprendimento e 3 Prova.

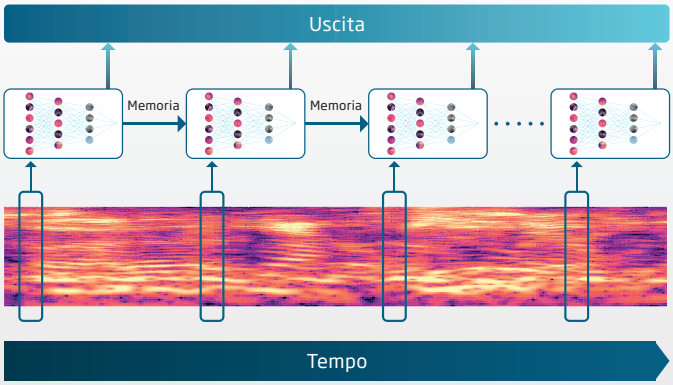

Figura 8. Il principio di una rete neurale LSTM (Long-Short Term Memory) o Gated Recurrent Unit (GRU).

2. Addestramento
L'obiettivo di questa fase è addestrare il DNN sulle scene sonore in modo che possa svolgere il compito per cui è stato progettato. Sono necessari molti dati per la formazione. Questi dati sono stati registrati in diverse scene sonore in un'ampia gamma di ambienti di ascolto a cui gli ascoltatori potrebbero essere esposti nella loro vita quotidiana. Abbiamo utilizzato un microfono sferico specializzato, in grado di catturare suoni a 360 gradi per fornire al DNN una scena sonora spazialmente precisa e dettagliata e addestrarlo sulla scena sonora completa. Alcuni dei dati raccolti sono stati utilizzati con lo scopo di addestrare il DNN e altri per la fase di test. I dati per la fase di test non verranno utilizzati fino alla fase di verifica finale. I dati di addestramento sono stati forniti al DNN come input per conoscere le scene sonore.

Il processo di addestramento può essere ulteriormente suddiviso in 4 fasi di loop: (A) Input (B) Propagazione in avanti (C) Output (D) Propagazione all'indietro (figura 9).

Durante la fase di input (A), la rete neurale profonda sperimenta esattamente ciò che è stato menzionato in precedenza. I neuroni ricevono le informazioni di una scena sonora e le memorizzano. Successivamente, la propagazione in avanti (B) utilizza i dati dall'input in cui ogni neurone passa le informazioni al livello successivo.

Fondamentalmente, la quantità di informazioni trasmesse dipende dalla forza delle connessioni che ciascuno dei neuroni ha con gli altri presenti all'interno della rete. Questo porta alla previsione dell'output (C) degli oggetti che il DNN pensa di dover migliorare e sopprimere nella scena sonora. Tuttavia, proprio come chiunque apprenda una nuova abilità per la prima volta, il DNN commette degli errori. Ad esempio, potrebbe enfatizzare eccessivamente il gabbiano (figura 9, prima immagine sotto C).

Poiché si tratta di una forma di apprendimento supervisionato, istruiamo il DNN se ha commesso un errore e che deve cambiare decisione. Questa azione guida il processo di propagazione all'indietro (D), in cui il DNN modifica le singole connessioni tra ciascun neurone per sopprimere meglio il gabbiano la prossima volta che si ripresenta un suono simile. Questo processo viene quindi ripetuto per tutte le scene sonore e, di conseguenza, la rete neurale profonda inizia a identificare le caratteristiche di ciascun oggetto per distinguerle meglio. Allo stesso tempo, gli scienziati regoleranno anche le caratteristiche specifiche del DNN, ad esempio la velocità con cui apprende la rete neurale profonda. Questo forma una simbiosi naturale tra l'incredibile adattabilità del DNN e la profonda conoscenza degli scienziati per raggiungere al meglio il nostro obiettivo di fornire un buon codice neurale. Nel corso del tempo, mentre i 4 stadi ripetono tutti i 12 milioni di scene sonore che abbiamo inserito nel DNN, la sua capacità di enfatizzare e sopprimere i rispettivi oggetti significativi e non significativi migliora fino a quando gli scienziati non sono soddisfatti. Ciò è dimostrato dalle prestazioni sempre più accurate mostrate dalle immagini in uscita (figura 9 (C)).

3. Prova
Quando il DNN è stato completamente sviluppato, è il momento di testare come si comporta con i dati a cui non è stato esposto prima durante la fase di addestramento. Questo passaggio è fondamentale poiché alcuni DNN potrebbero non funzionare bene. Sebbene le reti neurali funzionino eccezionalmente bene nel completare l'attività

Figura 9. Leggi il testo per maggiori chiarimenti.

in questione, il DNN potrebbe essere troppo rigido e specifico per i dati su cui è stato addestrato. Ci imbattiamo quindi nel problema del sovrallenamento, dove il DNN è addestrato a diventare molto preciso, ma non "ad adattarsi" al mondo reale. In altre parole, sa solo come trattare i dati su cui è stato addestrato, ma non è all'altezza se esposto a nuove sfide. D'altra parte, potrebbe anche non funzionare bene con i dati su cui è stato addestrato, rendendolo troppo ambiguo (figura 10).

Al fine di evitare questi problemi, la fase di test è fondamentale per fornirci un certo grado di fiducia su come il DNN si comporterebbe nel mondo reale, dove i suoni sono altamente variabili e le possibilità illimitate. Durante la fase di test, esaminiamo il DNN con i suddetti dati di test. Idealmente, il DNN funziona bene anche se testato con dati sconosciuti. Se non va a buon fine, è necessario modificare la rete neurale profonda o progettare un altra DNN. Nel nostro caso, noi di Oticon abbiamo progettato diverse versioni del DNN, ognuna con le proprie caratteristiche uniche e abbiamo selezionato quella attuale in base alle sue prestazioni sia in fase di addestramento che di test.

L'ultimo test che il DNN attraversa è quando viene implementato nell'apparecchio acustico come parte del MSI in fase di test sulle orecchie degli ascoltatori con problemi

di udito. I risultati di alcuni di questi test saranno descritti in un whitepaper che sarà presto disponibile.

Una rete neurale profonda consente di gestire i suoni del mondo reale in maniera precisa e automatica. Questo ottimizza il modo in cui Oticon More rende i suoni più distinti e funziona perfettamente in diversi ambienti di ascolto. Con questa intelligenza integrata, Oticon More ha imparato a riconoscere tutti i tipi di suoni, i loro dettagli e il modo in cui dovrebbero suonare idealmente, il tutto per supportare in modo ottimale il cervello.

Sound Enhancer

Normalmente il massimo effetto di un sistema di soppressione del rumore deve essere un compromesso che soddisfi tutti i pazienti, anche se alcuni pazienti avrebbero preferito rimuovere più suono e per alcuni pazienti è stato rimosso troppo suono. L'elaborazione del suono nell'apparecchio acustico deve garantire che il paziente sia in grado di gestire l'ambiente e allo stesso tempo ottenere la giusta percezione della scena sonora.

Sound Enhancer fornisce i dettagli dinamici del suono quando la soppressione del rumore è attiva, principalmente in ambienti complessi, il che consente di personalizzare l'uscita.

L'impostazione Comfort può essere scelta per il pieno effetto del sistema di soppressione del rumore, mentre l'impostazione Dettaglio può essere scelta per un forte livello di connessione con l'ambiente circostante e con gli oratori presenti. Ciò fornisce un'opzione di personalizzazione in base alle preferenze dell'utilizzatore.

Sound Enhancer viene applicato nello schema di elaborazione dopo Spatial Clarity e Neural Clarity Processing. Sound Enhancer esamina la soppressione dinamica eseguita da Spatial Balancer e dalla DNN, quindi calcola la quantità di suono da aggiungere al segnale in base alle impostazioni in Oticon Genie 2. Sound Enhancer

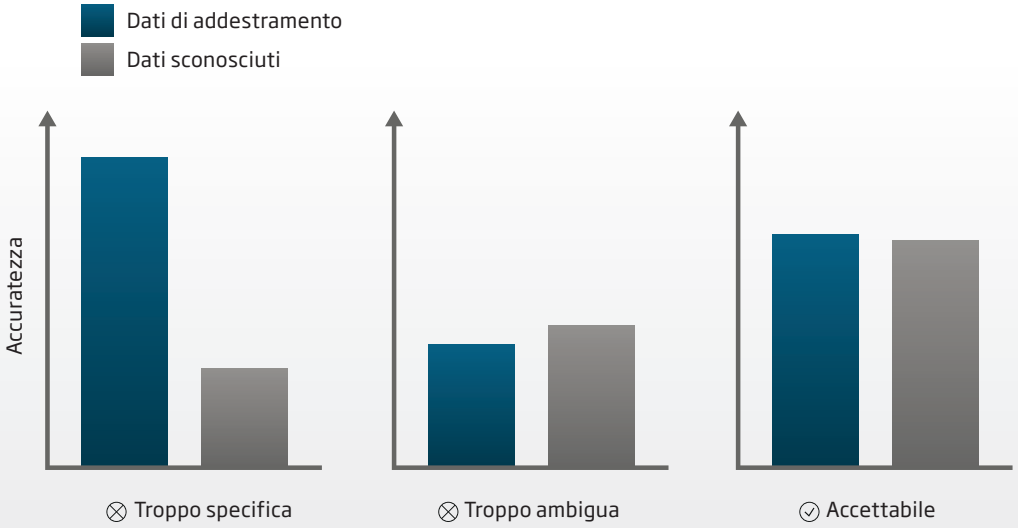

Figura 10. Precisione del DNN sui dati di allenamento e sui dati sconosciuti. Elevata precisione sui dati di addestramento ma scarsa sui dati sconosciuti: DNN è troppo specifica. Bassa precisione sia sui dati di addestramento che sui dati sconosciuti: DNN è troppo ambigua. Elevata precisione sia sull'addestramento che sui dati sconosciuti: DNN accettabile.

segue la soppressione adattativa complessiva eseguita da Spatial Balancer e DNN. Non si adatterà sempre a piccoli cambiamenti, ma a grandi cambiamenti generali nell'ambiente sonoro.

I dettagli aggiunti vengono forniti principalmente nell'area 1-4 kHz. Ciò significa che migliorerà principalmente i segnali vocali. Il compromesso però è che migliora anche altri tipi di suono. La Figura 11 mostra la maggiore enfasi sulla regione del parlato con una transizione verso le aree delle basse e delle alte frequenze.

Questo modellamento della frequenza significa che le frequenze medie (1-4 kHz) con la maggior parte dei segnali vocali avranno più peso rispetto alle frequenze basse e alte. A causa di questo modellamento della frequenza e del fatto che il rumore è spesso a bassa frequenza e i segnali vocali sono a media frequenza, verrà creato un contrasto leggermente migliore tra voce e rumore quando il suono viene aggiunto al segnale.

Bilanciato è l'impostazione predefinita e funziona per la maggior parte delle persone. Dettaglio è l'impostazione che fornisce la maggior parte del suono con il massimo livello di dettaglio per i pazienti che preferiscono avere un forte livello di connessione con l'ambiente circostante e con chi parla. Comfort è l'impostazione con meno suono e fornisce più comfort per i pazienti che preferiscono ridurre leggermente lo sforzo di ascolto complessivo attenuando leggermente l'ambiente circostante rispetto agli interlocutori presenti.

La Figura 12 mostra l'output di Oticon More rispetto a Oticon Opn S. Lo scopo dell'illustrazione è mostrare la relazione tra le diverse impostazioni.

La prospettiva

MoreSound Intelligence esegue la scansione dell'intera scena sonora, applica le impostazioni personali per il singolo paziente, organizza con precisione i suoni circostanti e utilizza il DNN per creare contrasto tra i suoni identificati. Tutto questo viene fatto molto velocemente e con precisione.

MoreSound Intelligence si basa sulla filosofia BrainHearing™ di Oticon. Dà accesso alla scena sonora completa, dove i singoli suoni sono perfettamente bilanciati e forniscono al cervello un codice neurale integro, fondamentale per una corretta comprensione e consapevolezza.

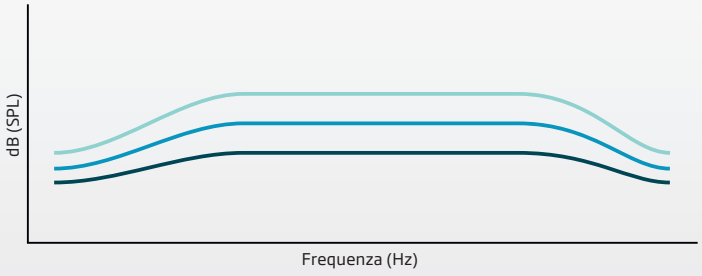

Figura 11. Modellazione della frequenza per i suoni aggiunti da Sound Enhancer.

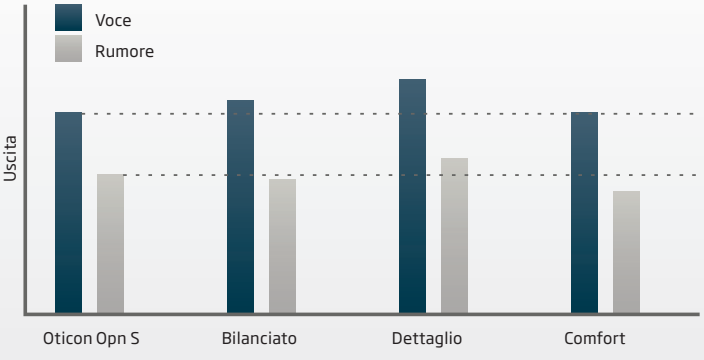

Figura 12. Relazione tra le tre diverse impostazioni in Sound Enhancer rispetto a Oticon Opn S.

Fonti

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. The Polaris Platform. Oticon tech paper

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. Optimal Fitting of Oticon More. Oticon Whitepaper

Man, B. and Ng, E. 2020. BrainHearing – The new perspective. Oticon Whitepaper

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.

72034IT / 2021.01.27 / v1

www.oticon.global

oticon
life-changing **technology**