



This is the **published version** of the bachelor thesis:

Castano Martos, Ainhoa; Casanova Mohr, Raimon, tut. Algorismes per a processament del so en persones autistes amb hiperacusia. 2025. (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/308752>

under the terms of the  license

Algorismes per a processament del so en persones autistes amb hiperacusia

Ainhoa Castaño Martos

5 de febrer de 2025

Resum— Les persones dintre de l'espectre autista tenen tendència a tenir hipersensibilitat auditiva, el qual els hi causa una gran càrrega cognitiva. L'objectiu d'aquest projecte és desenvolupar i implementar una metodologia per investigar les diferències en càrrega cognitiva entre persones autistes i persones no autistes en un entorn virtual sonor d'una aula universitària i alhora veure els efectes que pot tenir la introducció d'un algorisme de reducció de soroll. Amb el paradigma experimental basat en la simulació dels sons d'una aula i realitzant mesures de pupil·lometria i electroencefalografia per valorar si l'ús d'algorismes reductors de soroll ambiental poden reduir la càrrega cognitiva, l'estrès i augmentar l'atenció en l'ambient educatiu. S'ha dut a terme una prova pilot obtenint resultats que indiquen que el sistema funciona i que es pot veure una millora amb l'algorisme de reducció de soroll ambient.

Paraules clau— Autisme, TEA, espectre autista, càrrega cognitiva, audició, algorismes de reducció del soroll, pupil·lometria, electroencefalografia, simulació d'aula, hipersensibilitat auditiva.

Abstract— Many people diagnosed with autism spectrum suffer from auditory hypersensitivity, which can lead to increased cognitive load. The goal of this project is to create a methodology to evaluate differences in cognitive load between autistic and non-autistic people in a virtual simulated sound classroom environment and assess the effects of a noise reduction algorithm. The goal of this project to develop and implement a paradigm that simulates the sounds of a classroom and records pupillometry and electroencephalography measurements to evaluate if the use of noise reduction algorithms can decrease the cognitive load, stress and increase attention in educational environment. A pilot test has been carried out with results indicating that the system works and that an improvement can be seen with using the noise reduction algorithm.

Keywords— Autism, ASD, cognitive load, hearing, noise reduction algorithms, pupillometry, electroencephalography classroom simulation, auditory hypersensitivity.



1 INTRODUCCIÓ

SEGONS el canal de salut de la Generalitat de Catalunya "Els trastorns de l'espectre autista (TEA) són alteracions del neurodesenvolupament que es caracteritzen per una afectació en la capacitat de comunicació i interacció social, la presència d'un patró de conductes repetitives (interessos restringits, estereotípies, etc.), rigidesa de pensament i una desregulació del perfil sensorial (la manera com processen la informació que capten amb els sentits)".

Dintre d'aquesta desregulació del perfil sensorial, hi ha un gran nombre de persones autistes amb hipersensibilitat auditiva. En aquest treball ens centrarem en les diferències auditives referents a la hipersensibilitat auditiva i el processament d'una conversa en un entorn amb soroll ambient de fons.

Els estudis de com les persones autistes processen la informació auditiva mostren que hi ha diferències diverses [1]. En aquest treball ens centrarem en les diferències auditives referents a la hipersensibilitat auditiva i el processament d'una conversa en un entorn amb soroll ambiental.

Per una altra banda, les persones dintre de l'espectre autista en un ambient amb un elevat soroll ambient experimenten més dificultats que persones neurotípiques a l'hora d'entendre una conversa, tant en l'àmbit de detecció del so com en l'àmbit del processament de la part lingüística. Aquestes dificultats poden variar segons com sigui el soroll de fons

• E-mail de contacte: ainhoacastmartos@gmail.com
 • Menció realitzada: Enginyeria de Computadors
 • Treball tutoritzat per: Waldo Nogueira i Raimon Casanova Mohr
 (Departament d'Arquitectura de Computadors i Sistemes Operatius)
 • Curs 2024/25

[2] [3] [4].

S'han realitzat investigacions similars amb persones amb afàsia [5]. En aquestes, es mesuraven els nivells d'estrès en un ambient sorollós on la persona amb afàsia havia de parlar. Es van comprovar els nivells d'estrès mitjançant una entrevista als participants. L'objectiu d'aquest Treball de Final de Grau és idear un experiment que permeti explorar la percepció auditiva de persones autistes en ambients de so realistes. A més s'investigarà si tècniques de reducció de soroll poden millorar l'enteniment de la parla i reduir la càrrega cognitiva en persones diagnosticades amb autisme. En aquest sentit, es farà servir uns algorismes reductors de soroll basat en el "Wiener Filter Noise Reduction" [6] i MetricGAN-OKD [7]. La idea d'aquest experiment és analitzar quin d'aquests dos algorismes de reducció de soroll ambient pot millorar la percepció de la parla i la càrrega cognitiva, això ho comprovarem desenvolupant un paradigma per avaluar l'atenció, càrrega cognitiva en persones amb i sense diagnòstic de trastorn de l'espectre autista així com investigar algorismes de reducció de soroll per millorar l'escolta i la càrrega cognitiva de persones amb diagnòstic d'autisme.

Les hipòtesis d'aquest projecte és que el grup de participants autistes tindran més càrrega cognitiva que el grup de participants neurotípic escoltant l'àudio sense reducció de soroll i, per una altra banda, ambdós reduiran la càrrega cognitiva amb els algorismes de reducció de soroll i tindran una càrrega cognitiva similar. Amb aquest objectiu es crea el paradigma d'aquest article que mitjançant la simulació dels sons d'una aula i realitzant mesures de pupil·lometria i electroencefalografia valorem si l'ús d'algorismes reductors de soroll ambiental poden reduir la càrrega cognitiva, l'estrès i augmentar l'atenció en l'ambient educatiu.

En aquest article primer introduïm la metodologia, explicant en quins grups es planteja fer l'estudi 2.1, el procediment 2.2 de les condicions d'escolta 2.3. Seguidament, es presentarà la simulació de l'aula 2.4 i els algorismes utilitzats 2.5, els aparells utilitzats per determinar els nivells de càrrega cognitiva, com són pupil·lometria 2.6, electroencefalografia 2.7 i com se sincronitzaran les dades entre si 2.8. Finalment, es presenta els resultats, la discussió i les conclusions.

2 METODOLOGIA

2.1 Subjectes

En un estudi futur es planteja dur a terme un experiment amb la metodologia que s'explica a continuació amb 10 persones autistes que expressin problemes de càrrega cognitiva en l'ambient universitari i 10 persones no autistes que faran de grup de control.

Actualment, s'ha dut a terme una prova pilot amb una persona no autista per validar la metodologia dissenyada i desenvolupada a aquesta tesi.

2.2 Procediment

Primer de tot, es presenta un àudio d'exemple per ajustar el volum a un volum fort però no molest per la persona.

Abans d'iniciar cada àudio, durant 10 s es farà un registre com a línia base, per comparar amb l'enregistrament amb

els àudios.

Es dividirà amb la mesura en fragments de dos minuts- corresponents a la duració dels àudios-, es processaran i compararan entre si.

Després d'escoltar cada fragment de l'àudio, es preguntarà de forma oral al participant com se sent i quins nivells de nerviosisme i ansietat nota. I es farà una pregunta d'opció múltiple per veure si ha prestat atenció o no.

Aquest procediment es realitzarà per un àudio amb soroll de fons sense modificar, dos àudios amb un filtre de soroll ambient diferent per cadascun.

Per facilitar la presentació dels àudios, s'ha realitzat un programa amb Matlab que realitza tot l'anteriorment explicat.



Figura 1: Imatge d'un subjecte amb el material per realitzar les mesures: eye-tracker, electroencefalograma i auriculars.

2.3 Condicions d'escolta

Les condicions d'escolta seran que els individus hauran d'estar a una sala amb il·luminació constant per no afectar a la mesura de la pupil·lometria, aïllada electromagnèticament per no afectar l'electroencefalografia i acústicament per assegurar-nos que només escolten els àudios proporcionats. Aquesta sala preferiblement seria una cabina on el subjecte a realitzar l'estudi estigui en solitud per evitar distraccions per moviment, petits sorolls, etc. que pugui provocar el fet de tenir una altra persona al costat.

Aquests àudios han sigut gravats en una sala aïllada acústicament i consisteixen en una gravació d'un professor donant una classe. Originalment, són àudios de 24 minuts els quals s'han processat per reduir qualsevol pausa major a 5 s, eliminar dintre del possible qualsevol crossa i reduïts a una longitud de 20 minuts per evitar qualsevol mena de disparitat de longitud.

En aquests àudios s'ha realitzat una simulació d'una aula on s'està fent una classe amb soroll ambiental mitjançant TASCAR (Toolbox for Acoustic Scene Creation and Rendering) [8].

Un cop tenim els àudios emulant una aula, farem un remostreig a 16 kHz perquè és la freqüència de mostreig que cobreix la veu humana tenint en compte el criteri de Nyquisti dos dels àudios els hi aplicarem un algorisme de reducció de soroll ambiental diferent. Un d'aquests àudios el processarem amb cap filtre. Finalment, retallarem cada àudio en segments de 2 minuts perquè, per una banda, el subjecte tingui un petit descans i per un altre, per a fer una pregunta d'opció múltiple per assegurar-nos que s'està atenent als àudios.

Per obtenir mesures tan objectives com sigui possible, cada individu escoltarà els àudios amb un ordre diferent i cada àudio amb algorismes de reducció de soroll diferents.

2.4 Simulació del aula

Per a realitzar la simulació de l'aula, utilitzarem l'eina TASCAR.

Amb l'ajut d'aquesta eina serem capaços de simular una aula d'una mida de 14 x 7 m amb parets d'una alçada de 2.5 m amb una tarima de dimensions 4 x 6 m, simulant una aula de l'Escola d'Enginyeria de la UAB.

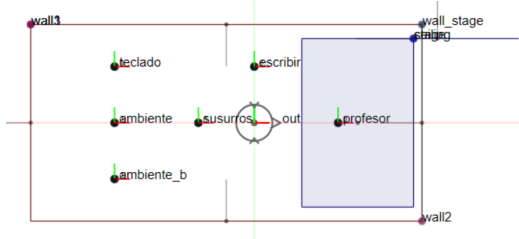


Figura 2: Simulació de l'aula feta amb TASCAR, 14 x 7 x 2.5 m. A sobre de l'escenari està situat el professor, just davant del receptor (out). Al voltant del receptor siulem sorolls de persones escrivint en paper i amb teclat i dues gravacions de fons que simulen un soroll ambiental d'una aula - xiuxiueigs, objectes que cauen a terra, portes que es tanquen, etc. Aquests sons s'emeten des dels punts indicats i TASCAR calcula com ho sentiria el receptor tenint en compte la reverberació de les parets.

Aquesta eina també ens permet simular la reverberació dels materials. En cas de les parets, tenen una reverberació de 0.9 s a excepció de la paret que es troba en el costat de l'escenari, que en té de 0.8 s. La reverberació de l'escenari és de 0.85 s. La reverberació determina la persistència del so un cop ha cessat la font original i el temps de reverberació correspon al temps que triga la intensitat del so en reduir-se 60 dB respecte al valor original.

Com podem veure a la figura (2), la simulació consta de 5 sorolls ambientals diferents, l'oient (out) i el professor, que consistirà en una seqüència d'àudios de dos minuts de duració separats per un silenci.

Per a la simulació del professor s'han gravat 3 àudios corresponents a una classe de l'assignatura Processament d'Àudio i Parla. Aquests àudios han sigut editats, deixant totes les pauses amb una duració menor a 0.5 s a causa de l'algorisme que s'utilitzarà per processar les dades de l'experiment. A més s'ha remostrejar un cop afegit el soroll ambiental a 16 kHz.

TASCAR utilitza dos tècniques anomenades Image-Source Method i HRTF Interpolation. Un cop definida les dimensions, reverberació d'un espai emissors i receptors. En el cas que ens aplica, una aula amb un receptor i diverses fonts de sons. S'utilitza Image-Source Method, que és una tècnica geomètrica utilitzada en l'acústica per simular la resposta impulsional (RIR) i entendre com es comporta la reverberació en diferents entorns. Consisteix a realitzar una simulació de la reflexió especular del so entre l'emissor i el receptor, assumint que el so viatja amb una trajectòria en forma de línia recta i si arriba a una paret o objecte, la trajectòria sofreix una reflexió perfecta obtenint finalment una imatge amb tota la trajectòria dels sons.

Per una altra banda, HRTF Interpolation consisteix en un cop realitzat l'anterior, derivar la contribució del calculat mitjançant el mètode anterior a les coordenades del receptor, utilitzant 3-D head-related transfer function (HRTF) que consisteix en calcular l'elevació i l'angle de l'orientació sobre la superfície d'una esfera que en el nostre cas correspon al receptor. Amb aquestes dades es calcula la posició desitjada i es calcula com aquesta imatge es percebia des del punt que s'ha calculat.

2.5 Algorismes de reducció de soroll

Amb el propòsit de reduir la càrrega cognitiva mitjançant la reducció del soroll ambiental, compararem tres algorismes i seleccionarem els dos millors per comparar-los de forma experimental.

L'algorisme Wiener Noise Reduction té dues versions, la primera que utilitza una tècnica anomenada two-step noise reduction (TSNR) que el que fa és eliminar l'efecte de la reverberació, però, en el procés afegeix distorsió harmònica en la veu. Per contrarestar aquest problema, es fa servir la tècnica harmonic regeneration noise reduction (HRNR), sent una segona versió d'aquest algorisme. Aquests algorismes utilitzen xarxes neuronals profundes (DNN).[6]

Per una altra banda, l'algorisme de Matlab EnhanceSpeech, fa servir l'algorisme (Multi-Metric Optimization of MetricGAN via Online Knowledge Distillation for Speech Enhancement) MetricGAN-OKD, que consisteix en múltiples generadors i mètriques objectiu, que permet als generadors aprendre respecte a una única mètrica mentre a l'hora milloren la qualitat dels resultats respecte a altres mètriques, mitjançant la imitació dels altres generadors. [7]

Per poder comparar els algorismes, utilitzarem un àudio amb el soroll ambient generat amb la simulació de l'apartat previ, dividim els segments en àudios de 2 minuts, separarem els dos canals d'àudio, com que fem servir un format estèreo, i els comparem amb l'àudio original sense soroll ambiental mitjançant la mètrica Short-time objective intelligibility (STOI).

STOI [9] consisteix a comparar l'àudio sense soroll ambiental i el que en té. Cada àudio està dividit en N fragments, on N és el nombre de fragments, m indica el fragment actual, j com a subíndex indica que prèviament el senyal s'ha normalitzat a la banda d'un terç d'octava, Y' indica l'àudio amb soroll normalitzat i retallat, el mateix fragment sense soroll ambiental, X, fent servir la correlació creuada entre aquests i amb un previ remostreig a 10 kHz:

$$d_j(m) = \frac{\sum_n \left(x_j(n) - \frac{1}{N} \sum_l x_j(l) \right) \left(y'_j(n) - \frac{1}{N} \sum_l y'_j(l) \right)}{\sqrt{\sum_n \left(x_j(n) - \frac{1}{N} \sum_l x_j(l) \right)^2 \sum_n \left(y'_j(n) - \frac{1}{N} \sum_l y'_j(l) \right)^2}} \quad (1)$$

I després fer la mitja d'aquests segments per a obtenir el STOI general de l'àudio:

$$d = \frac{1}{JM} \sum_{j,m} d_j(m) \quad (2)$$

On $0 \leq d \leq 1$, sent 0 una molt mala intel·ligibilitat i 1 una molt bona.

2.6 Eye-Tracking

Per poder fer les mesures de pupil·lometria utilitzarem el dispositiu Pupil Core de la marca Pupil Labs [10] i el seu programari per a poder sincronitzar-nos amb l'aparell a través de LSL i processar les dades.

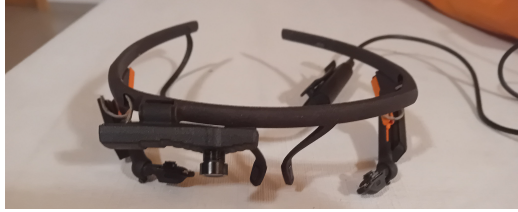


Figura 3: Imatge de Pupil Core de la marca Pupil Labs, del sistema d'Eye-Tracker. Consta de tres càmeres, una per gravar el que observa el participant, la qual no farem servir i dos càmeres que graven els ulls de qui les porta. Aquestes dues càmeres tenen una resolució de 92 x 192 px i operen a 200 Hz.

L'Eye-Tracker funciona mitjançant pupil·lometria, una per cada ull, que el que fan és mesurar la dilatació de les pupil·les, per poder tenir unes mesures correctes, s'ha de fer una calibració i ajustos inicials amb l'ajut del programari Pupil Capture - que també s'encarregarà de recollir les dades.

Durant l'experiment el participant haurà de mirar fixament una pantalla en negre amb una creu blanca enmig per poder mantenir la calibració.

Mitjançant aquest aparell s'estudiarà la dilatació i contracció de la pupil·la durant tot el temps que la persona estigui escoltant l'àudio, més dilatació de la pupil·la, més esforç per escoltar estarà fent el subjecte [11].

Per avaluar la dilatació les dades de la dilatació de les pupil·les, $p_{original}$, s'han computat de tal forma que quedin segmentades en instants de 5 s, de la següent forma obtenint $p(m)$:

$$p(m) = \frac{1}{N} \sum_{i=1}^M \sum_{l=T/M*i}^{T/M*(i+1)} p_{original}(l) \quad (3)$$

On M és el resultat de multiplicar la freqüència de mostreig per 5, l correspondria a la posició de cada mostra original, per veure quantes mostres hi ha cada 5 s i N és el nombre de mostres vàlides en aquest període, ja que quan el subjecte parpelleja, aquestes mostres no són vàlides. Aquestes mostres indicades com a no vàlides el mateix dispositiu les marca com NaN (Not-a-Number).

2.7 Electroencefalografia

Per a fer les mesures d'electroencefalografia es realitzaran mitjançant el sistema d'EEG Smarting de la marca mBrainTrain [12].



Figura 4: Imatge de Smarting de la marca mBrainTrain el sistema d'electroencefalografia utilitzat. Disposa de 24 elèctrodes que es connectem a un amplificador que envia el senyal al nostre dispositiu mitjançant un pendrive amb connexió Bluetooth.

Per poder utilitzar aquest dispositiu, s'ha de col·locar en el punt mitjà que hi ha entre el nas i l'occiput (punt entre les cel·les) i el final del crani). Això es realitza mesurant la distància entre els dos punts amb una cinta de costurer, dividint el valor entre dos i ficant el punt de referència de l'EEG en aquest punt.

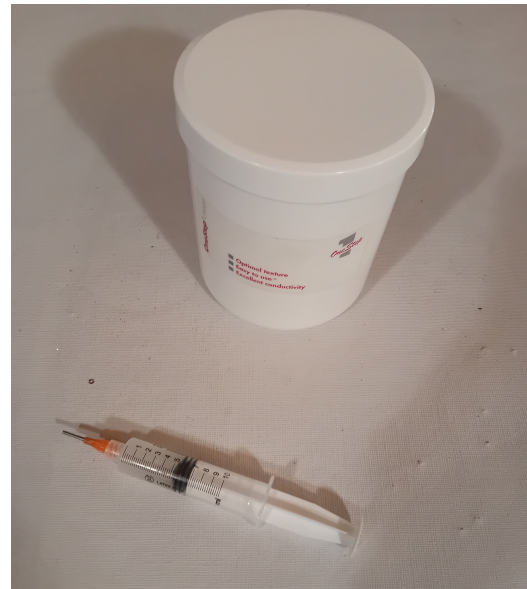


Figura 5: Imatge del líquid conductor utilitzat i xeringa utilitzada per aplicar gel als elèctrodes.

Per afavorir la conductivitat entre els elèctrodes i la pell del subjecte, utilitzem un gel per elèctrodes. Aquest gel anirà reduint la impedància entre la pell i l'elèctrode fins a arribar a una impedància inferior a 20 kOhms.

Usarem el programari proporcionat per Smarting per registrar les dades, que tindrà una freqüència de mostreig de 500 Hz.

2.7.1 Neural Tracking

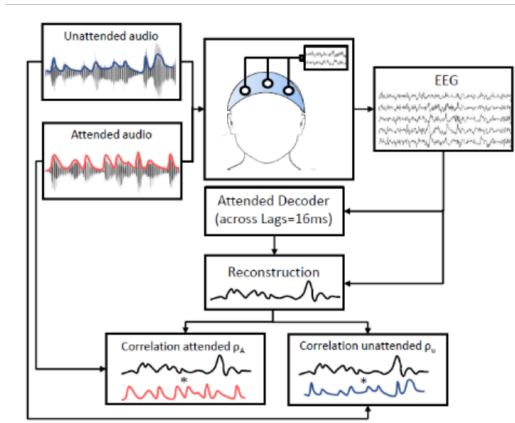


Figura 6: Diagrama de blocs mostrant el paradigma de Neural Tracking. Figura inspirada en les publicacions [13] [14] [15] [16]. Copyright de la imatge es Waldo Nogueira, Auditory Prosthetic Group, Hannover Medical School

Per mesurar els nivells d'atenció, s'utilitzarà la tècnica Neural Tracking [17] figura 6. El subjecte escolta un àudio amb sons que ha d'atendre i uns altres que no, en el nostre cas, "Unattended audio" serà el soroll de fons i "Attended Audio" serà la veu del professor, simultàniament es registra EEG. Es crea un descodificador, i es fa una reconstrucció dels àudios mitjançant aquest que es compararà amb els sons que el participant ha d'atendre ρ_a i els que no ρ_u , obtenint finalment un coeficient de correlació amb aquests sons.

Tant l'EEG com l'àudio es dividiran en segments de 60 s obtenint un total de K segments, en el nostre cas 10, llavors, $K-1$ segments s'utilitzaran per fer la reconstrucció i el restant per validar.

Amb aquesta equació obtindrem la reconstrucció dels sons que ha prestat atenció el participant mitjançant el seu EEG:

$$\hat{x}_{a,u}[k] = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} w_{n,l} \cdot y_n[k + \Delta + l], \quad (4)$$

On Δ representa la latència, N el nombre d'electrodes, en el nostre cas 24, L correspon a l'índex del vector de lags que probarem, k es el temps, $w_{n,l}$ correspon al descodificador i y_n a la resposta neuronal al temps de la mostra K .

Seguidament minimitzarem l'error quadràtic de l'envolvent extreta de l'àudio original, X_a i la reconstruïda de la següent forma:

$$J_{RLS}(W_a) = E \left[|x_a[k] - W_a^T y[k]|^2 \right] + \lambda W_a^T W_a$$

Sent λ un paràmetre de regularització. Si minimitzem l'equació anterior respecte als coeficients del descodificador, obtenim la següent solució:

$$W_a^T = (R_{xaY} + \lambda I)^{-1} \cdot R_{YY},$$

On R_{xaY} i R_{YY} estan definides de la següent forma:

$$R_{xaY} = \sum_{m=0}^K x_a[k] y[k + m],$$

$$R_{YY} = \sum_{m=0}^K y[k] y[k + m].$$

Una vegada tenim la reconstrucció, es calcula la correlació d'aquesta amb les envolupants dels dos àudios, obtenint ρ_a com a correlació del que ha escoltat el subjecte amb la part que havia d'escoltar, en el nostre cas, l'envolupant de la veu del professor i, per una altra banda, ρ_u correspon amb la correlació amb el que s'havia d'ignorar, en el nostre cas, el soroll ambiental. Es calculen correlacions per a cada desfasament (lag) entre l'àudio original, ja que no podem saber exactament quant triga el so a ser processat pel cervell.

Ens quedarem amb el lag de la mesura amb major correlació amb l'envolupant de la veu del professor, i agafarem la correlació corresponent en ambdós casos.

Per poder utilitzar correctament aquest algorisme, les gravacions no poden tenir més de 0.5 s de silenci de la veu principal.

2.7.2 Mesura d'engagement

Per una altra banda, per determinar els nivells de compromís amb la tasca farem servir la següent equació [18]:

$$e(m) = \frac{Beta(m)}{Alpha(m) + Theta(m)}$$

On M és el resultat de multiplicar la freqüència de mostreig per 5, l correspondria a la posició de cada mostra original, $Theta$ correspon a la potència del senyal de l'EEG que correspon a les bandes de 4- 7 Hz, Alpha als 7 -13 Hz i Beta als 13 - 25 Hz. Per obtenir aquests senyals, s'ha de filtrar l'EEG amb un filtre passa-banda amb les freqüències esmentades anteriorment, aquest filtratge es realitzarà processant les mesures d'electroencefalografia mitjançant l'eina EEGLAB [19] i es realitzarà la mitjana per cada instant de temps dels canals, a això li direm *senyal_{original}* per cada senyal en l'equació següent, computar la potència i computar aquesta de la següent forma, obtenint per cada senyal *senyal_m*, per a cada senyal de forma individual abans d'utilitzar-les a l'equació anterior:

$$senyal(m) = \sqrt{\sum_{i=n-i \cdot m}^{n \cdot M} \sum_{l=T/M \cdot i}^{T/M \cdot (i+1)} senyal_{original}^2(l)} \quad (5)$$

On M és el resultat de la multiplicació de la freqüència de mostreig per 5 s. Quan major sigui el valor de $e(t)$ aquest valor, més "engagement" n'hi ha.

2.8 Captació i sincronització de les dades

Per poder sincronitzar la reproducció de les gravacions, l'EEG i les mesures amb l'Eye-Traking utilitzarem Lab Streaming Layer (LSL) [20].

LSL és una capa de programari obert que permet enviar, rebre i sincronitzar fluxos de dades neuronals, fisiològiques i d'altres dispositius en temps real o marcadors com el descrit a la figura 7. Per al seu funcionament, en cas que no hi hagi

una instal·lació per defecte, s'ha d'instal·lar la llibreria compatible amb l'aplicació o dispositiu utilitzat. Aquesta llibreria enviarà les dades a la xarxa de forma pública, fent-ho disponible i descobrible per altres dispositius compatibles amb LSL que estiguin a la mateixa xarxa. Cada dispositiu emissor té un identificador únic i les dades es sincronitzen amb precisió d'un mil·lisegon mitjançant marques de temps basades en un rellotge comú utilitzant el protocol TCP/IP.

Les eines d'EEG i Eye-Tracking fetes servir són compatibles amb LSL, per sincronitzar la reproducció dels àudios amb les mesures, s'ha fet un codi en Matlab que cada cop que inicia i termina un àudio, envia un marcador al LSL.

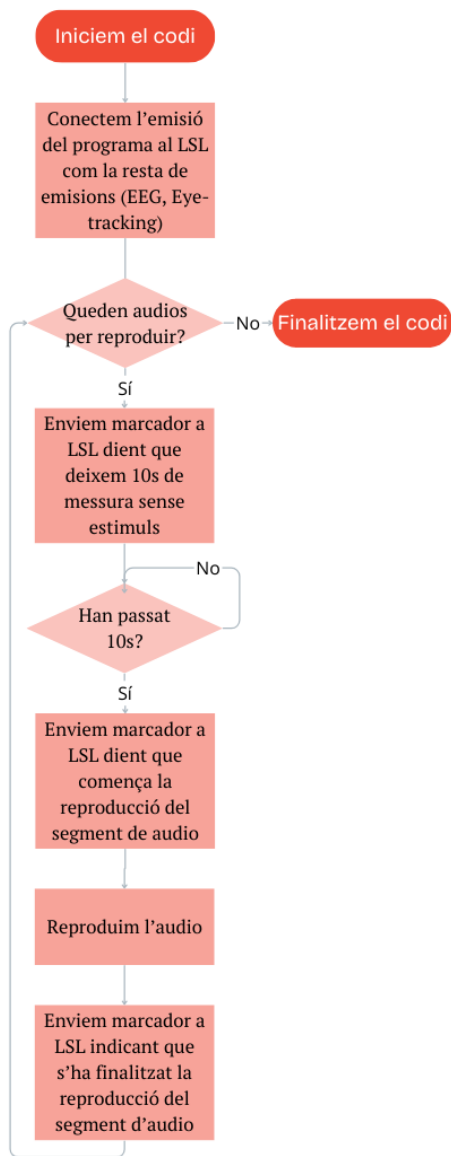


Figura 7: Diagrama de flux del codi.

Per poder fer servir aquesta eina, necessitem desconnectar l'ordinador on s'estiguin prenent les mesures de xarxes que no siguin locals, ja que com hem explicat, les dades s'emetten en broadcast i això pot comprometre la privacitat dels participants de l'estudi.

3 RESULTATS

3.1 Short-time objective intelligibility measure

Utilitzem la mètrica STOI per avaluar els diferents algorismes, obtenint les mesures mostrades a la figura 8

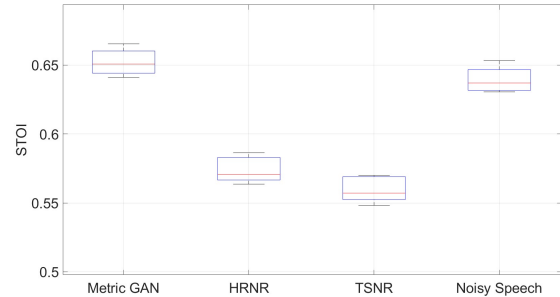


Figura 8: Comparativa dels algorismes metric GAN, HRNR i TSNR mitjançant STOI.

Com podem observar a la figura 8, l'algorisme EnanceSpeech és el que millors resultats genera amb una gran diferència, seguit de HRNR. L'algorisme TSNR queda descartat, ja que té el pitjor STOI. Tant HRNR i TSNR tenen pitjor STOI que l'àudio sense filtrar (condició Noisy Speech).

3.2 Electroencefalografia

Mitjançant les proves pilot, podem assegurar que tot el sistema enregistra de forma correcta les dades de l'EEG i que es poden processar mitjançant l'algorisme de Neural Tracking i la mesura per determinar els nivells de compromís.

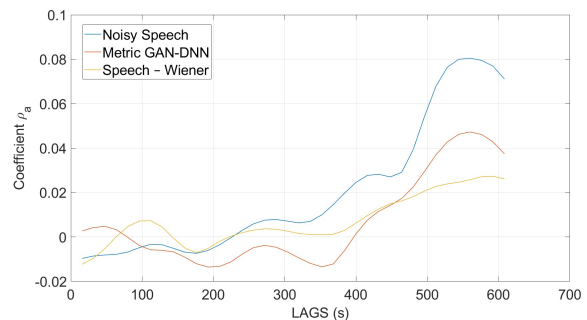


Figura 9: Resultats del experiment d'electroencefalografia basat en Neural Tracking per cada condició d'escolta: Noisy Speech, Speech - Wiener i Metric GAN-DNN. A l'eix "y" tenim ρ_a i a l'eix "x" Δ .

Com podem observar a la Figura 9, la reconstrucció amb la tècnica Neural Tracking obtenim un millor resultat per a Noisy Speech en comparació dels àudios que han passat pels filtres. Això pot estar ja que la veu principal sofreix certa distorsió, sent la més distorsionada la veu principal de Metric GAN-DNN, corresponent als resultats.

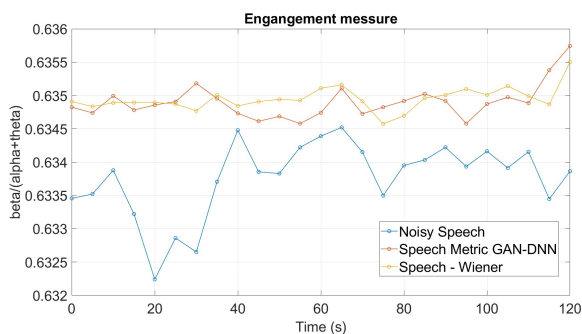


Figura 10: Resultats de quantitat d'engagemment per cada condició d'escolta: Noisy Speech, Speech - Wiener i Metric GAN-DNN.

Per una altra banda, quant al nivell d'engatjament, que podem veure a la figura 10, els dos àudios filtrats tenen uns valors similars que superen als del Noisy Speech, el qual té com a observació adicional, molta més variació que els àudios filtrats.

3.3 Eye-tracking

Mitjançant les proves pilot, podem assegurar que tot el sistema enregistra de forma correcta les dades de l'Eye-Tracking i es poden visualitzar mitjançant MatLab.

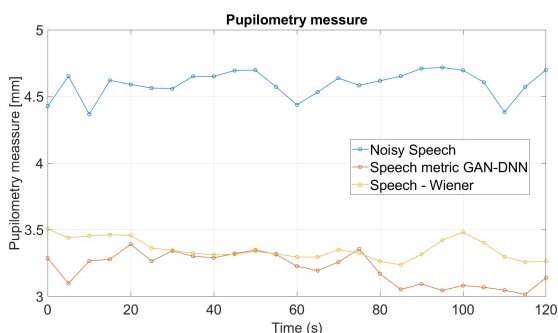


Figura 11: Diàmetre de la pupil·la [mm] per cada condició d'escolta: Noisy Speech, Speech - Wiener i Metric GAN-DNN

Podem observar a la figura 11 que de nou, la dilatació de la pupil·la per als àudios processats amb reductors de soroll tenen unes mesures similars, en aquest cas tenint una dilatació menor de mitjana que l'àudio amb soroll ambient, això ens indicaria que el subjecte està fent més esforços per escoltar Noisy Speech que els àudios filtrats.

3.4 Percepció dels àudios

A l'hora de preguntar al subjecte amb el qual s'ha dut a terme una prova pilot, ha mencionat un punt interessant. En utilitzar els algorismes de reducció de soroll, la veu principal es torna "robòtica", sent l'algorisme HRNR el que més es nota la diferència, això també ens ho indica el STOI, però, per contra, és en el que més se sent concentrat perquè els sorolls li semblaven menys distractors.

4 DISCUSSIÓ

Les hipòtesis d'aquest projecte és que el grup de participants autistes tindran més càrrega cognitiva que el grup de

participants neurotípic escoltant l'àudio sense reducció de soroll i, per una altra banda, ambdós reduiran la càrrega cognitiva amb els algorismes de reducció de soroll i tindran una càrrega cognitiva similar. Amb aquest objectiu es crea el paradigma d'aquest article que mitjançant la simulació dels sons d'una aula i realitzant mesures de pupil·lometria i electroencefalografia valorem si l'ús d'algorismes reductors de soroll ambiental poden reduir la càrrega cognitiva, l'estrès i augmentar l'atenció en l'ambient educatiu. És aviat per extreure conclusions, però sí que podem veure indicis a la pupil·lometria que utilitzant tècniques de filtratge de soroll ambiental redueix l'esforç necessari per necessitar prestar atenció.

Respecte al STOI, s'esperaven millors resultats, sobretot amb l'algorisme Wiener. S'ha de valorar si és per la robotització de la veu i quan pot afectar aquesta a la comprensió de la parla.

Pel que fa a les mesures obtingudes mitjançant la tècnica de Neural Tracking, la forma observada no es similar a la esperada, en estudis anteriors s'observen dos crestes en compres d'una i el valor de pic es molt tard, normalment observant-se a valors de lags de 200-400 ms [17]. Això pot ser degut a un retard degut a l'ús de LSL, per a futurs estudis s'ha de veure si es algo comú o ha sigut un fet puntual a causa de les condicions no-ideals d'estudi. En tractar-se d'una prova pilot, les mesures no s'han realitzat en una habitació aïllada tant acústicament com electromagnèticament, lo qual ha pogut afectar els resultats. Per a futurs estudis seguint aquesta metodologia, s'haurà de fer en les condicions esmentades, a causa d'això i que només s'ha fet una prova pilot.

Continuant amb l'engagemment, aquesta mètrica sí que mostra que hi ha més engagemment als àudios sense filtrar, però realment la diferència de mitjanes entre tots és de l'ordre de les mil·lèsimes, llavors no podem assegurar que la utilització de filtratge de soroll tinguin un impacte significatiu en persones no autistes. S'ha de repetir amb més subjectes tant autistes com no autistes.

Finalment, respecte a la pupil·lometria, sí que s'observa un impacte en quant al diàmetre de la pupil·la si l'àudio està filtrat o no i s'observa un decaïment en el seu diàmetre en el el temps com a estudis anteriors de temàtica similar [18]. La dilatació de la pupil·la ens indica els esforços per entendre la veu, en aquest cas, el subjecte ha fet més esforços per escoltar l'àudio amb soroll, tenint aproximadament una dilatació de la pupil·la 1 mm superior de mitjana a la mesurada amb els àudios amb soroll ambient, el qual sí que implica una millora amb la utilització d'àudios amb filtratge de soroll ambient.

5 CONCLUSIONS

Concluïm da la prova pilot de l'experiment impementat que és una forma adequada per a mesurar la càrrega cognitiva, estrès i atenció en un entorn sonor virtual d'una classe. Els programes permeten presentar de forma correcta els àudios i es comuniquen correctament amb LSL. L'enregistrament de les dades és correcte, podent ser aquestes processades.

Donada la percepció del participat durant a la prova pilot, deixem per al futur afegir una comparació entre si es millora l'atenció disminuint els sorolls ambientals a costa de tenir una veu principal més robòtica.

D'acord amb els resultats vists durant el desenvolupament concloem que es podria iniciar un assaig clínic amb les especificacions donades a la metodologia.

AGRAÏMENTS

Agraeixo al meu tutor, Waldo Nogueira i al meu co-tutor Raimon Casanova per tot el suport donat.

Agraeixo al departament de Microelectrònica de la UAB i Hannover Medical School APG Auditory Prosthetic Group, en especial a la Hanna Dolhopiatenko, a Lourdes Encinas Donaldo pel suport a la part ètica i a UAB Open Labs per facilitar-nos els seus espais per a poder fer les gravacions.

I finalment, a la meva família, amics i parella pel suport donat durant aquest projecte.

REFERÈNCIES

- [1] K. O'Connor, "Auditory processing in autism spectrum disorder: a review," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 2, pp. 836–854, 2012.
- [2] S. E. Rotschafer, "Auditory discrimination in autism spectrum disorder," *Frontiers in Neuroscience*, vol. 15, p. 651209, 2021.
- [3] W. A. Teder-Sälejärvi, K. L. Pierce, E. Courchesne, and S. A. Hillyard, "Auditory spatial localization and attention deficits in autistic adults," *Cognitive Brain Research*, vol. 23, no. 2-3, pp. 221–234, 2005.
- [4] H. Fujihira, C. Itoi, S. Furukawa, N. Kato, and M. Kashino, "Sensitivity to interaural level and time differences in individuals with autism spectrum disorder," *Scientific reports*, vol. 12, no. 1, p. 19142, 2022.
- [5] T. G. Harmon, R. Hegewald, and C. Dromey, "“competing noises”: How background noise impacts the communication experiences of people with mild-to-moderate aphasia," *American Journal of Speech-Language Pathology*, vol. 33, no. 6S, pp. 3393–3409, 2024.
- [6] P. Scalart, "Wiener filter for noise reduction and speech enhancement (<https://www.mathworks.com/matlabcentral/fileexchange/24462-wiener-filter-for-noise-reduction-and-speech-enhancement>)," *American Journal of Speech-Language Pathology*, 2025.
- [7] W. Shin, B. H. Lee, J. S. Kim, H. J. Park, and S. W. Han, "Metricgan-okd: multi-metric optimization of metricgan via online knowledge distillation for speech enhancement," in *International Conference on Machine Learning*, pp. 31521–31538, PMLR, 2023.
- [8] G. Grimm, "Tascar, <https://www.tascar.org/>."
- [9] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *2010 IEEE international conference on acoustics, speech and signal processing*, pp. 4214–4217, IEEE, 2010.
- [10] "Pupil labs, <https://pupil-labs.com/products/core>."
- [11] T. Seifi Ala, C. Graversen, D. Wendt, E. Alickovic, W. M. Whitmer, and T. Lunner, "An exploratory study of eeg alpha oscillation and pupil dilation in hearing-aid users during effortful listening to continuous speech," *Plos one*, vol. 15, no. 7, p. e0235782, 2020.
- [12] "Smarting, <https://mbraintrain.com/smarting-wireless-eeg/>."
- [13] W. Nogueira and H. Dolhopiatenko, "Predicting speech intelligibility from a selective attention decoding paradigm in cochlear implant users," *Journal of Neural Engineering*, vol. 19, no. 2, p. 026037, 2022.
- [14] W. Nogueira, G. Cosatti, I. Schierholz, M. Egger, B. Mirkovic, and A. Büchner, "Toward decoding selective attention from single-trial eeg data in cochlear implant users," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 38–49, 2019.
- [15] N. Aldag, A. Büchner, T. Lenarz, and W. Nogueira, "Towards decoding selective attention through cochlear implant electrodes as sensors in subjects with contralateral acoustic hearing," *Journal of Neural Engineering*, vol. 19, no. 1, p. 016023, 2022.
- [16] H. Dolhopiatenko and W. Nogueira, "Selective attention decoding in bimodal cochlear implant users," *Frontiers in Neuroscience*, vol. 16, p. 1057605, 2023.
- [17] W. Nogueira, H. Dolhopiatenko, I. Schierholz, A. Büchner, B. Mirkovic, M. G. Bleichner, and S. Debener, "Decoding selective attention in normal hearing listeners and bilateral cochlear implant users with concealed ear eeg," *Frontiers in neuroscience*, vol. 13, p. 720, 2019.
- [18] A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biological psychology*, vol. 40, no. 1-2, pp. 187–195, 1995.
- [19] "Eeglab, <https://scn.ucsd.edu/eeglab/>."
- [20] "Lab streaming layer, <https://github.com/scn/labstreaminglayer>."