



UNIVERSITATEA POLITEHNICA
DIN BUCUREȘTI



Școala Doctorală de Electronică, Telecomunicații
și Tehnologia Informației

Decizie nr. 972 din 08-12-2022

REZUMAT TEZĂ DE DOCTORAT

Ing. Ana-Luiza RUSNAC

Recunoașterea vorbirii imaginate prin analiza semnalelor
EEG

Imaginary speech recognition by EEG signal analysis

COMISIA DE DOCTORAT

Prof. dr. ing. Ion MARGHESCU Univ. Politehnica din București	Președinte
Prof. Dr. Ing. Ovidiu GRIGORE Univ. Politehnica din București	Conducător de doctorat
Prof. dr. ing. Daniela TĂRNICERIU Univ. Tehnică „Gh. Asachi” din Iași	Referent
Prof. dr. ing. Ioan LIȚĂ Univ. din Pitești	Referent
Conf. dr. ing. Anamaria RĂDOI Univ. Politehnica din București	Referent

BUCUREȘTI 2022

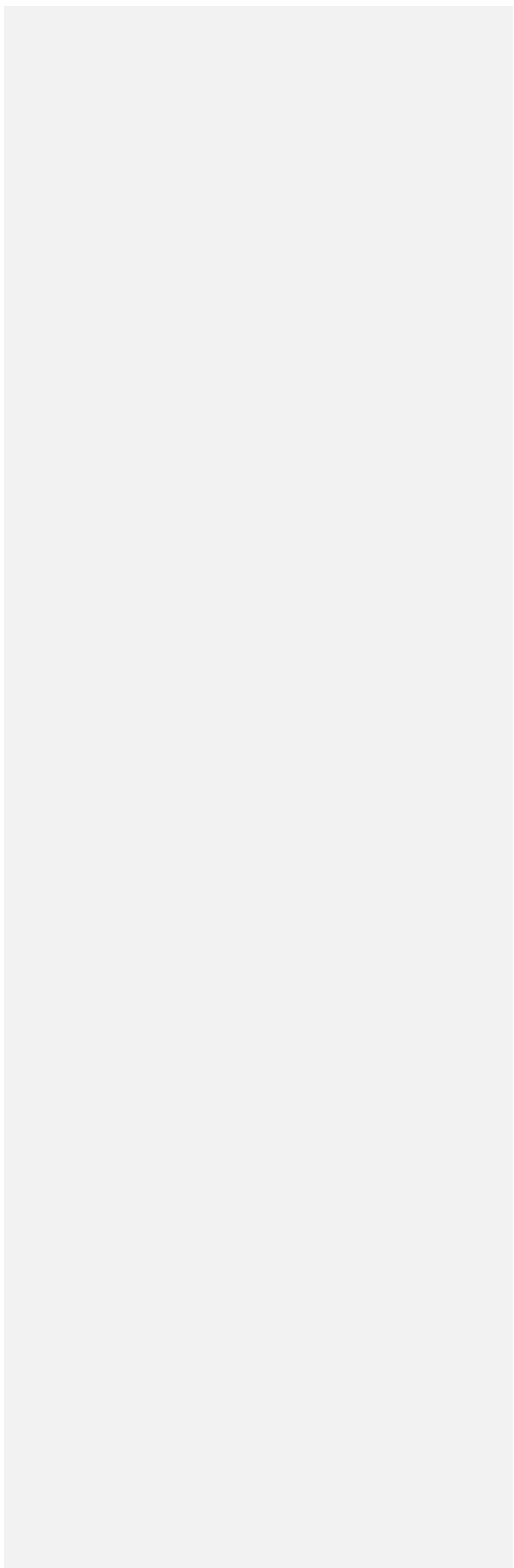
Cuprins

Cuprins

Capitolul 1 Introducere	1
1.1 Prezentarea domeniului tezei de doctorat	1
1.2 Scopul tezei de doctorat	1
1.3 Conținutul tezei de doctorat	2
Capitolul 2 Baze teoretice	4
2.1 Producerea semnalului vocal	4
2.2 Producerea, modelarea și analizarea activității electrice a creierului	4
2.3 Înregistrarea activității electrice a creierului	5
2.3.1 Electroencefalografia (EEG)	5
2.3.2 Electrococtigrafia (ECoG)	5
2.3.3 Stereoelectroencefalografia (sEEG)	6
Capitolul 3 Mecanisme de pronunție	7
3.1 Rostirea vocalelor	7
3.2 Rostirea consoanelor	8
Capitolul 4 Stadiul actual al sistemelor de recunoaștere a discursului imaginat	9
Capitolul 5 Baze de date	29
5.1 Baza de Date Kara One (KODB)	29
Capitolul 6 Preprocesarea bazei de date KODB	30
Capitolul 7 Eliminarea artefactelor oculare [15]	11
7.1 Filtru adapiv	11
7.2 PCA	11
7.3 Rezultate	12
7.4 Concluzii	12
Capitolul 8 Sistem de recunoaștere a mecanismelor diferite de pronunție [19]	14
8.1 Coeficienții MFCC	14
8.2 LPC	14
8.3 Augmentarea datelor	15

8.4 Clasificarea	15
8.5 Rezultate	15
8.6 Concluzii	16
Capitolul 9 Recunoașterea fonemelor utilizând MFCC și CNN [23]	17
9.1 Rezultate	17
9.2 Concluzii	17
Capitolul 10 Analizarea și clasificarea discursului imaginat utilizând rețeaua SOM.....	19
10.1 Covariația încrucișată în domeniul timp	19
10.2 Covariația încrucișată în domeniul frecvență	19
10.3 Reprezentarea spațiului trăsăturilor semnalelor EEG utilizând coeficienții MFCC și rețeaua SOM	20
10.4 Reprezentarea spațiului trăsăturilor semnalelor EEG utilizând covariația încrucișată în domeniul timp și rețeaua SOM	20
10.5 Reprezentarea spațiului trăsăturilor utilizând covariația încrucișată în domeniul frecvență și rețeaua SOM.....	20
10.6 Clasificarea datelor de intrare utilizând rețeaua neuronală SOM	20
10.7 Concluzii.....	21
Capitolul 11 Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB utilizând CNN [25]	23
11.1 Clasificarea semnalelor.....	25
11.2 Metrice de evaluare a sistemului	26
11.3 Rezultate	26
11.3.1 Analizarea funcțiilor de activare: Tanh și ReLU	26
11.3.2 Studiul trăsăturilor: Timp vs Frecvență	26
11.3.3 Analiza lungimilor de prelucrare: 0.25, 0.5 și 1s.....	27
11.3.4 Comparatie filtru de mediere: B0, B3 și B5	27
11.3.5 Metrice de evaluare a performanțelor sistemului	27
11.3.6 Măsurători de complexitate și memorie.....	27
11.3.7 Comparatie între semnalele neprocesate și semnalele procesate utilizând PCA pentru eliminarea artefactelor oculare	27
11.4 Discuție	27
11.5 Concluzii	28

Capitolul 12 Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB utilizând CNNLSTM [27].....	55
12.1 Calcularea trăsăturilor.....	55
12.2 Clasificarea utilizând CNNLSTM	55
12.3 Rezultate	56
12.3.1 CNNLSTM vs CNN	56
12.3.2 Analiza regiunilor craniene.....	56
12.3.3 Analiza complexității de calcul și a memoriei.....	56
12.3.4 Analiza semnalelor neprocesate și semnalelor procesate utilizând PCA.....	56
12.4 Discuții.....	57
12.5 Concluzii	57
Capitolul 13 Concluzii.....	58
13.1 Rezultate	59
13.1.1 Capitolul 7: Eliminarea artefactelor oculare	59
13.1.2 Capitolul 8: Sistem de recunoaștere a mecanismelor diferite de pronunție...	60
13.1.3 Capitolul 9: Recunoașterea fonemelor utilizând MFCC și CNN	60
13.1.4 Capitolul 10: Analizarea și clasificarea discursului imaginat utilizând rețeaua SOM.....	60
13.1.5 Capitolul 11: Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB	61
13.1.6 Capitolul 12: Sisteme de recunoaștere a fonemelor și cuvintelor din baza de date KODB utilizând CNNLSTM	62
13.2 Contribuții originale.....	62
13.3 Lista lucrărilor originale	63
Bibliografie	65



|

Capitolul 1

Introducere

1.1 Prezentarea domeniului tezei de doctorat

Sistemele de interfață creier-mașină au ocupat în ultimii ani un rol important printre subiectele de interes ale cercetătorilor din toată lumea. Acestea au pornit de la sarcini mai ușor de realizat, precum mișcarea unui scaun cu roțile în direcțiile față-spate și stânga-dreapta și au tins către ținte mai dificile, precum mișcarea unei proteze de mână ce realizează motricitatea fină a falangelor sau sinteza vocală.

Acest domeniu este într-o continuă creștere datorită tendinței actuale de a se implementa cât mai multe dispozitive portabile ce au ca scop îmbunătățirea calității vieții utilizatorilor. Un dispozitiv de recunoaștere automată a discursului imaginat aduce o valoare considerabilă pacienților ce suferă de afecțiuni care afectează abilitatea de comunicare, precum atacul cerebral, sindromul lock-down, etc., comunicarea fiind un element foarte important în viața cotidiană.

1.2 Scopul tezei de doctorat

Această lucrare reprezintă un studiu al sistemelor inteligente de interfață creier-mașină (BCI – Brain Computer Interface) ce au ca scop decodarea semnalelor preluate de la nivelul scalpului în timpul discursului imaginat. Discursul imaginat se referă la procesul de gândire a realizării vorbirii, fără mișcarea propriu-zisă a mușchilor implicați în producerea vorbirii. Discursul imaginat se referă la imaginarea mișcării tractului vocal în scopul pregătirii acestuia pentru articularea unui fonem sau cuvânt, fără a realiza mișcarea propriu-zisă a componentelor implicate în vorbire. Sistemele de recunoaștere a discursului imaginat dezvoltate în cadrul acestei teze sunt neinvazive, folosind semnale electroencefalografice (EEG), cu scopul de a recunoaște în timp real o serie de foneme și cuvinte atingând o acuratețe cât mai mare.

Formatted: Font color: Auto, Not Highlight

1.41.3 Conținutul tezei de doctorat

1.41.3 Conținutul tezei de doctorat

În prima parte a lucrării, *Capitolul 1*, sunt prezentate noțiuni generale cu privire la această teză de doctorat, cuprinzând o scurtă introducere în sistemele inteligente de recunoaștere a discursului imaginat împreună cu prezentarea domeniului de studiu și a obiectivelor acesteia.

Capitolul 2 reprezintă o introducere în elementele teoretice din domeniul fiziologiei necesare pentru studierea și aprofundarea mecanismului complex de pronunție, pornind de la intenția de producere a semnalului vocal la nivel cortical până la rostirea sunetelor componente discursului dorit. În *Capitolul 3* s-a realizat o scurtă descriere a diferitelor mecanisme de pronunție și modul de articulare al acestora.

În *Capitolul 4* au fost studiate sistemele deja existente cu privire la recunoașterea discursului imaginat din semnale corticale și s-a urmărit evoluția acestora în decursul ultimilor ani.

Următoarele două capitole, *Capitolul 5* și *Capitolul 6*, s-au axat pe descrierea bazei de date utilizate și metoda de preprocesare a acesteia.

În *Capitolul 7* au fost prezentate două metode de eliminare a artefactelor oculare: prima metodă bazându-se pe implementarea unui filtru adaptiv, iar a doua metodă pe eliminarea surselor ce conțin artefacte utilizând Analiza Componentelor Principale (PCA), fiind realizat un studiu comparativ între acestea.

În *Capitolul 8* s-a realizat un studiu al diferitelor mecanisme de pronunție, bazate pe (a) rostirea fonemului /iy/ (/iy/, /piy/, /tiy/, /diy/); (b) rostirea fonemului /uw/; și (c) rostirea consoanelor (/m/ și /n/) prezente în baza de date Kara One. Pentru modelarea celor trei mecanisme de pronunție s-a încercat utilizarea unor tehnici aplicate pe scară largă în recunoașterea automată a vorbirii (ASR – Automatic Speech Recognition): coeficienții Mel-Cepstrali (MFCC) și algoritmul de predicție liniară (LPC). Studiul comparativ a fost detaliat în acest capitol utilizând rețele neuronale convoluționale (CNN – Convolutional Neural Networks) în etapa de clasificare.

Capitolul 9 introduce primul sistem din procesul de cercetare urmat în această teză de recunoaștere automată a fonemelor din baza de date Kara One, bazându-se pe coeficienții MFCC și rețeaua neuronală CNN.

Capitolul 10 introduce o analiză a trăsăturilor extrase pe baza covariației încrucișate în domeniul timp comparativ cu domeniul frecvență, precum și a coeficienților MFCC, utilizând rețeaua neuronală nesupervizată Kohonen (SOM – Self Organizing Map). În acest capitol este introdusă pentru prima dată în recunoașterea cuvintelor din discursul imaginat covariația încrucișată între canale în domeniul frecvență.

Capitolul 1 - Introducere

Capitolul 11 a urmărit îmbunătățirea sistemului de recunoaștere automată a discursului imaginat, proiectarea, implementarea și analiza făcându-se pentru toate fonemele și cuvintele bazei de date Kara One. Capitolul prezintă o analiză detaliată a arhitecturilor și hiperparametrilor rețelei CNN utilizați pentru obținerea celor mai bune rezultate. Următorul capitol, **Capitolul 12**, prezintă sistemul propus pentru recunoașterea celor 11 clase din baza de date Kara One. Acesta folosește în etapa de extragere de trăsături covariația în domeniul frecvență împreună cu rețeaua neuronală CNNLSTM obținându-se cele mai bune rezultate ale studiului.

Ultimul capitol, **Capitolul 13**, conține concluziile finale ale întregii lucrări împreună cu contribuțiile originale aduse în decursul studiului și perspectivele viitoare ale acestuia.

Capitolul 2

Baze teoretice

2.1 Producerea semnalului vocal

În urma unor studii realizate [1], [2], s-a confirmat importanța deosebită pe care o are lobul temporal în reprezentarea lingvistică și înțelegerea conceptelor. În consecință, mecanismul vorbirii pornește din lobul temporal urmând apoi un traseu care permite transpunerea gândurilor în cuvinte rostite. Următorul element specific lanțului vorbirii este aria Broca, aceasta având rol major în planificarea, inițierea și modificarea ~~articulațiilor-articulatorilor~~ necesarie în ~~procesul de vorbire~~rostirii. Pe lângă aria Broca, în planificarea poziționării elementelor tractului vocal pentru rostire participă insula anterioară, pentru inițierea mișcărilor articulare participă aria motorie secundară, iar pentru execuția mișcărilor organelor efectuate participă cortexul motor primar facial și cortexul premotor. În finalizarea actului vorbirii se mai implică și ganglionul bazal și cerebelul, care se activează pentru modificarea frecvenței fundamentale, a volumului și a ritmului vorbirii [2].

După planificarea și transmiterea semnalelor motorii de la nivelul creierului, acestea ajung la organele efectoare. Aceste organe sunt flexibile, iar forma și mărimea lor se modifică în funcție de semnalul primit de la centrul nervos responsabil de articularea sunetelor. Sunetul se formează începând din plămâni. Plămânii oferă forța de aer necesară pentru a genera rostirea sub formă acustică. În continuare, aerul trece prin tractul vocal, coardele vocale, glotă, epiglotă și alte organe pentru a ajunge mai departe în cavitatea bucală sub formă de undă acustică [3].

2.2 Producerea, modelarea și analizarea activității electrice a creierului

Sistemul nervos poate fi divizat în sistem nervos central și sistem nervos periferic. Sistemul nervos central este alcătuit din encefal și măduva spinării și are rolul de a interpreta informațiile senzoriale și de a transmite instrucțiuni către organele efectuate pe baza informațiilor din experiențele anterioare [4].

Țesutul nervos este alcătuit din neuroni, celule gliale și celule endoteliale. Rolul funcțional al acestuia este de a primi și de a transmite mai departe impulsurile electrice care comunică mesaje cu privire la stimuli senzoriali, motori sau informații cognitive.

Neuronii sunt alcătuiți din corp (sau soma), axon și dendrite. Axonul are rolul de a transmite informația primită unidirecțional către alte celule nervoase. În timp ce dendritele primesc informația de la alți neuroni prin intermediul sinapselor. Cu alte cuvinte informația se transmite prin neuron pornind de la dendrite, trecând prin corpul neuronului și ajungând la axon. Acest schimb de informații care se realizează prin intermediul neuronului poate fi înregistrat ca semnal electric de către electroencefalograf (EEG) [5].

2.3 Înregistrarea activității electrice a creierului

Transmiterea informațiilor la nivelul sistemului nervos se realizează prin intermediul unor impulsuri electrice generate de procese (electro) biochimice. Această activitate poate fi înregistrată neinvaziv, prin poziționarea unui set de electrozi pe suprafața scalpului, metodă numită electroencefalografie (EEG) sau invaziv, care se poate realiza în două moduri: prin poziționarea unui set de electrozi direct pe suprafața cortexului, metodă ce poartă numele de electrocorticografie (ECoG) sau prin introducerea unor electrozi profund în țesutul creierului, metodă numită stereoelectroencefalografie (sEEG).

2.3.1 Electroencefalografia (EEG)

Electroencefalografia permite observarea proceselor electrice ale creierului ce au loc la suprafața cortexului. Mai exact, este o măsură a câmpului electric produs de un număr mare de neuroni activați simultan ca funcție de timp. Câmpul electric măsurat pe suprafața scalpului se datorează excitației a zeci de mii de neuroni prezenți în regiunea corticală respectivă. Un element important în ceea ce privește achiziționarea de semnale EEG, a fost dezvoltarea unui sistem prin care să poată exista o repetabilitate a înregistrărilor. Așa a luat naștere sistemul 10-20 de poziționare a electrozilor pe scalp [6]. Denumirea de 10-20 provine de la distanțele proporționale, reprezentate în procente, raportat la reperele craniene stabilite ca standard.

2.3.2 Elocorticografia (ECoG)

ECoG reprezintă o metodă invazivă de achiziționare a activității creierului direct de pe suprafața scalpului în sala de operație. Această metodă achiziționează semnale într-o manieră asemănătoare cu EEG, cu mențiunea că se elimină atenuarea dată de cutia craniană și scalp.

În Figura 2.9 este prezentat un exemplu de achiziționare a semnalelor ECoG pentru un subiect în sala de operații. Figura 2.9 (A) prezintă o imagine laterală radiografică unde se pot observa electrozii poziționați pe cortex, în timp ce imaginea (B) este o imagine preluată din sala de operații și permite vizualizarea electrozilor direct pe țesutul cortical. Figura 2.9 (C) prezintă răspunsul activității electrice înregistrat de electrozi în urma

Recunoașterea vorbirii imaginare prin analiza semnalelor EEG

aplicării unor stimuli precum: voce, tonuri și limbaj. Figura 2.9 (D) reprezintă același răspuns al creierului la aplicarea stimulilor exemplificat pe un model 3D al cortexului.

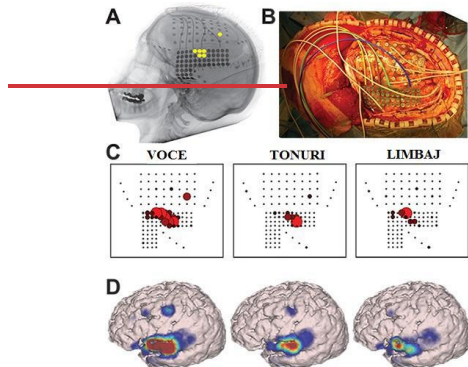


Figura 2.9 Exemplu de achiziționare a semnalelor corticale utilizând ECoG și răspunsul oferit în timpul aplicării unor stimuli: (A) Imagine radiografică ce prezintă cca 120 de electrozi (B) Imagine preluată din sala de operație având electrozii poziționați pe cortex (C) Răspunsul obținut de electrozi la aplicarea diferiților stimuli și (D) Răspunsul obținut după aplicarea stimulilor pe un model 3D al scalpului

2.3.62.3.3 Stereoelectroencefalografia (sEEG)

Descoperirea ECoG a condus la descoperirea posibilității de a utiliza electrozi de profunzime poziționați stereotactic. Aceștia sunt utilizați în special pentru determinarea precisă a focarului epileptic.

Capitolul 3

Mecanisme de pronunție

Limbajul reprezintă un sistem de comunicare alcătuit din sunete articulate, specifice oamenilor, prin care aceștia comunică. Modul în care aceste sunete sunt articulate pentru a exprima elemente specifice limbajului reprezintă mecanismul de pronunție.

Sunetele limbii engleze sunt, în general, împărțite în două categorii de bază: sunete segmentale și suprasegmentale. Sunetele segmentale cuprind **consoanele** și **vocalele** în timp ce sunetele suprasegmentale sunt descrise de o serie de parametrii fonetici precum: tonalitate (frecvența fundamentală), intonația și accentul.

Vocalele sunt sunete care se rostesc fără obstrucții majore ale tractului vocal, astfel încât aerul care pleacă din plămâni trece prin aparatul fonator destul de ușor în timpul rostirii. Exemple de astfel de sunete sunt: /iy/, /uw/, /ah/, /oh/.

Consoanele, spre deosebire de vocale, implică obstrucții sau constricții ale tractului vocal, elementele acestuia modificându-și poziția pentru restricționarea aerului conform rostirii dorite. De exemplu, pentru rostirea consoanei /p/, buzele sunt închise, prevenind aerului să iasă în timpul rostirii [7].

3.1 Rostirea vocalelor

Rostirea vocalelor presupune o deschidere mai mare a tractului vocal spre deosebire de rostirea consoanelor. Există două elemente primare implicate în mecanismul rostirii: forma și poziția limbii și forma buzelor.

În continuare vor fi analizate în detaliu cele trei mecanisme de articulare a vocalelor: *deschiderea* (gradul de deschidere a cavității orale), *locul de articulare* și *gradul de rotunjire a buzelor*.

Deschiderea sau gradul de deschidere a cavității orale presupune modul în care mecanismul de rostire implică deschiderea buzelor în momentul în care vocala este articulată. Acest element oferă informații și despre frecvența în rostirea vocalelor, mai exact, vocalele închise prezintă o frecvență mai înaltă în timp ce vocalele deschise prezintă o frecvență mai joasă.

Locul de articulare se referă la poziționarea limbii în momentul rostirii vocalei. Acesta poate fi anterior (/æ/) sau posterior (/ɑ/). Pentru a face diferența dintre cele două tipuri de rostiri, se poate încerca articularea cuvintelor din limba engleză pan (/pæn/) și

Formatted: Indent: First line: 0.5"

Formatted: Font color: Auto

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

palm (/palm/). În acest caz, în timpul rostirii cuvântului pan, limba se află în partea anterioară a cavității bucale, spre deosebire de palm, unde limba este poziționată în partea posterioară a cavității bucale pentru rostirea vocalei.

Gradul de rotunjire a buzelor se referă la modul de poziționare a buzelor într-o formă rotunjită sau nu în timpul rostirii vocalei. Acestea se mai numesc și labializate (/o/, /u/) sau nelabializate (/a/, /e/, /i/).

3.2 Rostirea consoanelor

Consoanele sunt sunete care se creează prin restricția parțială sau totală a trecerii aerului prin tractul vocal. Această restricționare se obține prin mișcarea a cel puțin unei componente către cealaltă astfel încât acestea să se atingă sau să fie foarte apropiate. Partea mobilă se numește componentă activă și este reprezentată de articulatorii inferiori, în timp ce partea fixă se numește componentă pasivă și este dată de articulatorii superiori.

Locul de articulare al consoanelor

În rostirea consoanelor un element foarte important îl constituie **locul de articulare**. Acesta reprezintă combinarea unui articulator activ cu unul pasiv.

Fonația

Pe lângă rolul de articulator, coardele vocale sunt utilizate de asemenea și pentru a controla trecerea aerului prin tractul vocal. Există cazuri în care la poziționarea coardelor vocale într-o manieră specifică, trecerea aerului prin glotă permite vibrarea acestora.

Când se obține vibrarea coardelor vocale în momentul rostirii unei consoane sau vocale, rostirea poartă denumirea de rostire **sonoră** sau, în caz contrar, când nu are loc vibrarea coardelor vocale se numește rostire **nesonoră**.

Modul de articulare al consoanelor

Consoanele pot fi de asemenea clasificate în funcție de modul de articulare, care se referă la modul prin care aerul circulă prin tractul vocal, pe baza mărimii și formei de comprimare a aerului dintre articulatori.

Procesele fiziologice implicate în articulare sunt complexe, iar toate aceste procese sunt controlate de către creier. Modul în care mușchii sunt comandați pentru rostirea fiecărui mecanism conduce la o activitate complexă a creierului în timpul realizării mișcărilor. Din acest motiv, se poate considera că activitatea electrică a creierului înregistrată de semnalele EEG în timpul rostirilor conține informații cu privire la aceste procese.

Formatted: Font color: Auto

Capitolul 4

State of the art — Literatura de specialitate Stadiul actual al sistemelor de recunoaștere a discursului imaginat

Brain-Computer Interface (BCI) reprezintă un sistem bazat pe calculator care măsoară activitatea neuronală a sistemului nervos central (SNC) și o decodează într-o ieșire artificială comandă capabilă să înlocuiască, restaureze, îmbunătățească sau să suplimenteze funcția motorie naturală a SNC modificând astfel interacțiunea primară dintre SNC și mediul extern [8].

Cele mai populare sisteme BCI sunt reprezentate de sistemele în care achiziționarea datelor se realizează neinvaziv, cu ajutorul electroencefalografului (EEG) de suprafață. Aceste dispozitive permit măsurarea activității neuronale prin amplificarea diferențelor de potențial dintre electrozii plasați pe scalp și câmpul electric emis de neuronii piramidali din cortexul cerebral [9].

Pentru o mai bună rezoluție spațială și temporală, în aplicațiile BCI se pot utiliza și semnale electrocorticografice (ECoG), semnale care oferă un raport SNR (sinal-to-noise ratio) foarte mare. Cu toate acestea, marele dezavantaj al acestei metode de achiziționare a semnalelor este reprezentată de caracterul invaziv, fiind necesar contactul direct dintre electrozi și cortexul cerebral.

BCI pentru sinteză vocală reprezintă un sistem care achiziționează semnale de la nivelul creierului, le procesează și le codifică mai departe în semnale vocale, comenzi care acționează diferite dispozitive sau le transformă în afișaje text. Principiul care stă la baza acestor tipuri de BCI pornește de la ideea că pentru producerea unui cuvânt, creierul trebuie să transmită informații specifice către elementele motorii ale tractului vocal precum limbă, maxilar, buze, laringe etc, în aceeași manieră în care creierul trimite semnale pentru mișcarea altor elemente motorii precum mâna sau piciorul.

Astfel că, în anul 2014, în studiul [10] s-a încercat diferențierea vocalelor „a”, „e”, „i”, „o” și „u” utilizând semnale EEG. Cercetătorii au calculat media, varianța, deviația standard și puterea medie a semnalului pentru diferențierea celor 5 vocale. Clasificarea semnalelor a fost realizată cu ajutorul unei rețele neuronale de tipul Multilayer Perceptron

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

(MLP) pentru fiecare subiect și s-au obținut următoarele rezultate: 44% acuratețe pentru primul subiect și 32% pentru subiecții 2 și 3.

Recunoașterea vocalelor din semnale EEG au fost studiate în continuare și în anul 2016 în Columbia de către Diego A. Rojas, Olga L. Ramos și Jorge E. Saby în studiul [11]. Aceștia au utilizat EMOTIV Epoch pentru achiziționarea semnalelor, un dispozitiv simplu și ușor de purtat. Au fost achiziționate semnale în timpul rostirii a două vocale: „a” și „e”. Pentru extragerea trăsăturilor și selectarea acestora, cercetătorii au utilizat Symbolic Aggregate Aproximation (SAA), iar pentru clasificare a fost utilizat algoritmul Support Vector Machine (SVM). Rezultatele obținute au depășit 84% pentru diferențierea dintre vocalele „a”, „e” și semnal neutru, în care nu a fost rostită nicio vocală. Rezultatele acestui studiu încurajează decodarea semnalelor din înregistrări EEG, însă este necesar de menționat că diferențierea s-a făcut doar între două vocale, fiind de așteptat ca acuratețea să scadă semnificativ odată cu introducerea mai multor vocale/consoane.

Dezvoltarea semnificativă a acestui domeniu în ultimii ani a condus la încurajarea cercetătorilor de a se implica în crearea sistemelor BCI pentru sinteză vocală prin punerea la dispoziție a mai multor baze de date pentru aceste aplicații. Este cunoscut faptul că achiziționarea datelor reprezintă o mare provocare a acestui domeniu din mai multe puncte de vedere: este necesar un echipament profesional, sunt necesare cunoștințe de specialitate pentru poziționarea corectă a electrozilor, iar semnalele sunt greu de achiziționate deoarece este necesar un context special, în care subiectul să se poată concentra specific pe aplicație, iar acesta trebuie să fie odihnit și să se facă pauze regulate pentru refacere.

În momentul de față există mai multe baze de date cu acces public care pot fi utilizate pentru această aplicație. Una dintre bazele de date care pot fi utilizate este oferită de Chuong H. Nguyen, George K. Karavas și Panagiotis Artemiadis în lucrarea [12]. Aceasta conține semnale achiziționate de la 15 subiecți sănătoși (11 bărbați și 4 femei) în timpul rostirii a trei grupuri de cuvinte: cuvinte scurte („in”, „out” și „up”), cuvinte lungi („cooperate” și „independent”) și vocale („a”, „i” și „u”). Fiecare cuvânt a fost rostit de 100 de ori într-o etapă de înregistrare.

O altă bază de date populară în acest domeniu a fost dezvoltată la Universitatea din Toronto, de către cercetătorii Shunan Zhao și Frank Rudzicz [13]. Această bază de date conține semnale colectate de la 12 subiecți (8 bărbați și 4 femei) în timpul rostirii a 7 foneme („iy”, „uw”, „piy”, „tiy”, „diy”, „m” și „n”) și 4 cuvinte scurte („pat”, „pot”, „knew” și „gnaw”). Setul de cuvinte și foneme a fost rostit de 12 ori de fiecare subiect, ajungând la 144 de rostiri pentru fiecare fonem și cuvânt.

Într-un studiu recent asupra recunoașterii discursului imaginat efectuat în Rusia [14], s-a reușit obținerea celei mai mari baze de date fiind achiziționate semnale de la 268 de subiecți pentru opt cuvinte rusești diferite: „înainte”, „înapoi”, „sus”, „jos”, „ajutor”, „ia”, „stop” și „eliberează”. În urma studiului, cercetătorii au susținut că este mai fezabilă realizarea unui sistem dependent de subiect care să prezinte o acuratețe mai mare, spre

Capitolul 4 - Stadiul actual al sistemelor de recunoaștere a discursului imaginat |

deosebire de crearea unui sistem generalizat utilizând semnale achiziționate de la un număr mare de subiecți diferiți.

Capitolul 5

Baze de date

În momentul de față, achiziționarea datelor reprezintă în continuare o provocare pentru cercetătorii din acest domeniu. Este cunoscut faptul că achiziționarea datelor necesită un efort semnificativ din mai multe puncte de vedere: este necesar un echipament profesional, sunt necesare cunoștințe de specialitate pentru poziționarea corectă a electrozilor, iar semnalele sunt greu de achiziționat deoarece este nevoie de un context special, în care subiectul să se poată concentra specific pe aplicație, de asemenea, acesta trebuie să fie odihnit și să se facă pauze regulate pentru refacerea energiei.

5.1 Baza de Date Kara One (KODB)

Kara One este o bază de date dezvoltată de o echipă de cercetători de la Universitatea din Toronto [13] în anul 2015. Această bază de date conține semnale colectate de la patru femei și opt bărbați, având media de vârstă de 27,4 ani. Toți participanții care au luat parte la studiu sunt dreptaci, urmează studii superioare, nu prezintă probleme vizuale, auditive sau motorii și nu au istoric de probleme neurologice sau consum de substanțe narcotice.

Subiecții au fost instruiți să urmărească monitorul instalat și să nu se miște. O sesiune de înregistrare a durat între 30 și 40 de minute, timp în care pe ecran se puteau observa unul dintre cele 7 foneme utilizate pentru recunoaștere: „iy”, „uw”, „piy”, „tiy”, „diy”, „m”, „n” sau unul dintre cuvintele: „pat”, „pot”, „knew”, „gnaw”.

Fiecare experiment a constat din patru stări succesive: (1) o perioadă de 5s de pauză, în care participanții au fost instruiți să se relaxeze și să nu se gândească la nimic; (2) o perioadă de stimulare, în care pe ecran apărea un text ce conținea un fonem sau un cuvânt, împreună cu un stimul auditiv corespunzător stimulului de pe ecran. După apariția stimulului auditiv, a urmat o perioadă de 2 secunde în care subiectul era instruit să își miște articulațiile în poziția necesară începerii pronunției stimulului vizual; (3) o perioadă de 5s în care fiecare participant era instruit să își imagineze rostirea cuvântului; (4) o perioadă în care subiectul rostea cu voce tare cuvântul, iar senzorul Kinect înregistra atât semnalul vocal cât și trăsăturile faciale.

Fiecare stimul vizual a fost prezentat de 12 ori, obținându-se astfel 132 de experimente. În final, 4 dintre cei 12 subiecți au fost eliminați din studiu deoarece prezentau electrozi detașați, iar doi dintre subiecți au adormit în timpul înregistrărilor.

Capitolul 6

Preprocesarea bazei de date KODB

Scopul ~~acestei lucrări~~ cercetării făcute de autoarea acestei lucrări a constat din identificarea cuvintelor ~~din semnale~~ pe baza semnalelor EEG, cuvinte rostite în timpul discursului imaginat. Din acest motiv, pentru experimentele care s-au realizat în acest scop, din înregistrările bazei de date Kara One au fost segmentate pentru utilizări ulterioare doar semnalele de 5s corespunzătoare rostirii în gând a fonemelor și cuvintelor. Pentru a elimina tranzițiile de la o stare la alta, în continuare din cele 5s de semnal s-au eliminat primele și ultimile 0.5s, obținând în final pentru fiecare stimul un semnal EEG de 4s.

Semnalele obținute în final au fost în continuare trecute printr-o etapă de curățare a datelor. În prima etapă a analizării vizuale, s-a descoperit că șase din cele 14 sesiuni de înregistrări prezentau zgomote foarte mari sau electrozi care nu erau bine atașați, oferind un semnal care nu putea fi utilizat în recunoașterea cuvintelor. Din acest motiv, toți acești subiecți au fost eliminați din studiu.

A urmat analizarea vizuală a tuturor semnalelor EEG corespunzătoare discursului imaginat și au fost eliminate din studiu cele care conțineau zgomote puternice datorate, în general, mișcării subiectului.

În urma acestui proces, de analizare vizuală a semnalelor, am obținut în final baza de date având în total 993 de semnale care au fost folosite mai departe în decursul studiului.

Semnalele din baza de date Kara One au fost filtrate utilizând un filtru Notch de 60Hz pentru a elimina artefactele provenite de la frecvența de rețea, împreună cu toți multiplii acesteia, mai mici decât frecvența Nyquist pentru eliminarea armonicilor.

Capitolul 7

Eliminarea artefactelor oculare [15]

Semnalele EEG sunt semnale de amplitudine mică, având un domeniu de variație cuprins între 5 și 200 μV [16]. Din această cauză, aceste semnale pot fi cu ușurință contaminate de alte semnale biologice precum semnale electrocardiografice, semnale electromiografice, semnale electrooculografice etc. În acest capitol s-a urmărit compararea a două metode de eliminare a artefactelor oculare: filtrul adaptiv și metoda PCA, artefactele oculare având cea mai mare influență asupra semnalelor EEG.

7.1 Filtru adaptiv

O metodă des utilizată de cercetători pentru a îmbunătăți calitatea semnalelor EEG constă din implementarea filtrelor adaptive. Marele avantaj al acestei metode de filtrare constă din adaptarea coeficienților progresiv, ținând cont de statistica semnalului la fiecare moment de timp. În studiul de față s-a utilizat un filtru adaptiv de dimensiune 400 coeficienți utilizând eroarea medie pătratică ca algoritm de antrenare. Deoarece acest algoritm este unul care performează bine în timp, cele 4 secunde de semnal al fiecărei înregistrări rostite din baza de date KODB au fost extinse la 60s prin concatenarea succesivă a aceleași înregistrări de mai multe ori.

7.2 PCA

Algoritmul își propune să găsească o matrice de coeficienți care să conțină sursele necorelate din semnal. Primul pas al algoritmului a fost să calculăm matricea de covarianță a semnalului EEG înregistrat în raport cu trăsăturile. În continuare, după calcularea matricei de covarianță se vor calcula cu ușurință vectorii proprii și valorile proprii ale matricei, vectori ce corespund componentelor principale ale semnalului.

7.3

Formatted: Caption

7.47.3 Rezultate

După aplicarea filtrului adaptiv, se poate observa o îmbunătățire a calităților semnalului EEG prin atenuarea componentei oculare din canalele frontale. Acest lucru este evidențiat în **Figura 7.3**.

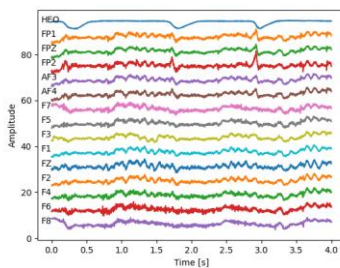


Figura 7.3 Semnalul EEG după aplicarea filtrului adaptiv – canalele FP1-F8 și HEO

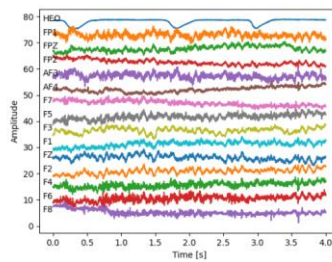


Figura 7.5 Semnalele EEG după eliminarea primelor două componente principale – canalele FP1-F8 și HEO

A doua metodă de eliminare a artefactelor oculare s-a bazat pe calcularea componentelor principale. În urma obținerii componentelor principale, s-a putut observa vizual faptul că primele două componente sunt foarte asemănătoare semnalului HEO. După eliminarea celor două componente, se poate vedea cu ușurință, conform **Figurii 7.5**, faptul că artefactele oferite de mișcările oculare au fost eliminate.

Semnalele canalelor frontale sunt cele mai afectate de aceste mișcări oculare. Corelațiile acestor canale cu semnalul HEO sunt foarte mari, atingând valori ce depășesc 0.8. După aplicarea filtrului adaptiv, această valoare a corelației a scăzut la aprox. 0.1, iar în cazul filtrului PCA la 0.2. În continuare a fost realizată o analiză cantitativă bazată pe corelația inter-clasă. S-a putut observa că pentru PCA, corelațiile dintre înregistrările aceleiași clase crește pentru majoritatea fonemelor și cuvintelor, în timp ce filtrul adaptiv decorează aceste înregistrări.

7.57.4 Concluzii

În acest capitol s-a urmărit compararea a două metode diferite de filtrare a artefactelor oculare: filtrare adaptivă și PCA pentru utilizarea mai departe a semnalelor filtrate în aplicații de recunoaștere a discursului imaginat. În urma acestui studiu, am concluzionat faptul că ambele metode atenuază artefactele oculare prin decorelarea semnalelor cu

Capitolul 7 - Eliminarea artefactelor oculare

HEO. Inițial s-a pornit de la o corelație foarte mare dintre semnalele achiziționate de pe electrozii frontali și HEO, având valori ce depășesc 0.8 și ajungând la valori de aproximativ 0.1 pentru filtru adaptiv și aproximativ 0.2 după filtrarea cu PCA. S-a observat că semnalele după ce au fost filtrate utilizând metoda PCA au prezentat o corelație mai mare inter-clasă, spre deosebire de semnalele obținute după aplicarea filtrului adaptiv, unde a apărut o decorelare între acestea.

Capitolul 8

Sistem de recunoaștere a **trei** **mecanismemecansimelor diferite** de pronunție [19]

Metoda propusă în acest capitol este dezvoltarea unui sistem BCI de recunoaștere a fonemelor din baza de date KODB grupate în trei categorii: C1, foneme ce conțin vocala /iy/, C2, fonemul /uw/ și C3, consoanele /m/ și /n/. Pentru realizarea scopului propus, au fost utilizate semnalele din KODB preprocesate utilizând metoda prezentată detaliat anterior în *Capitolul 6*. În etapa de extragere a trăsăturilor, s-a urmărit compararea a patru modele de caracteristici bazate pe calcularea coeficienților MFCC și LPC. Cele patru metode analizate în acest capitol al lucrării sunt următoarele: (1) coeficienți MFCC: MFCC (dimensiune 62x62); (2) coeficienți LPC: LPC (dimensiune 62x62); (3) concatenare MFCC și LPC într-o matrice de dimensiune 62x36 peste care a fost calculată covarianța: MFCC+LPC V1 (dimensiune 62x62); (4) concatenare MFCC și LPC într-o matrice tridimensională: MFCC + LPC V2 (dimensiune 62x62x2).

8) —

Formatted: Normal, No bullets or numbering

8.28.1 Coeficienții MFCC

Semnalele de intrare, corespunzătoare semnalelor EEG achiziționate în timpul discursului imaginat, sunt transformate utilizând FFT în domeniul frecvență. Peste spectrul semnalului, se aplică un banc de filtre triunghiulare a căror lungimi de bandă sunt calculate utilizând scara Mel. Pentru fiecare filtru triunghiular, se calculează energia spectrală, ca sumă de eșantioane ridicate la pătrat. În final, se vor obține coeficienții MFCC prin conversia logaritmului coeficienților calculați anterior din benzile Mel în domeniul timp, utilizând iFFT.

8.38.2 LPC

Algoritmul de predicție liniară este o tehnică foarte cunoscută în domeniul semnalelor **biologie-vocale**, deoarece această metodă oferă informații importante atât în domeniul timp cât și în domeniul frecvență [20]. Utilizarea pe scară largă a algoritmului LPC în

Capitolul 8 - Sistem de recunoaștere a mecansimelor diferite de pronunție

recunoașterea semnalelor vocale se bazează pe capacitatea acestei metode de a extrage informațiile esențiale din semnal și de a oferi un număr mic de parametri care descriu destul de bine configurația tractului vocal din timpul rostirii [21]. În lucrarea dezvoltată, au fost calculați 18 coeficienți LPC pentru decodarea informațiilor EEG a șapte foneme imaginate grupate în trei clase.

8.48.3 Augmentarea datelor

În urma grupării fonemelor în cele trei clase, a apărut problema dezechilibrului numărului de vectori din fiecare clasă. Pentru a rezolva această problemă, a fost introdusă o etapă de augmentare a datelor. În această etapă au fost generați artificial vectori din clusterelor C2 și C3 pentru a obține în final 200 de observații din fiecare clasă, în timp ce pentru clusterul C1 au fost eliminați aleator vectori pentru a ajunge la același număr egal cu 200 și din această categorie. Generarea artificială de date de intrare pentru clusterelor C2 și C3 a fost realizată utilizând distribuții Gausiene corespunzătoare fiecărei trăsături și fiecărui cluster.

8.58.4 Clasificarea

Pentru acest studiu a fost utilizată în etapa de clasificare o rețea CNN ce conține două straturi convoluționale 2D având numărul de filtre 64 și respectiv 32 și trei straturi complet conectate de dimensiune 32, 16 și 3 neuroni în ultimul strat echivalentul celor trei clusterelor. După fiecare strat convoluțional a fost introdus un strat de normalizare a lotului de antrenare. Funcția de activare utilizată a fost tangenta hiperbolică (tanh) pentru toate straturile cu excepția ultimului strat, unde funcția de activare utilizată a fost softmax.

8.68.5 Rezultate

În urma rezultatelor obținute, se poate susține că în comparație cu LPC, coeficienții MFCC au oferit un rezultat mai bun atingând valoarea medie a acurateței de 0.39, fiind mai adecvați pentru clasificarea discursului imaginar. De asemenea, se poate observa că au fost obținute rezultate mai bune prin combinarea celor două trăsături într-o matrice tridimensională cu două canale (MFCC+LPC V2) în comparație cu metoda MFCC + LPC V1. Utilizând un procesor i7-3537U având 6GB memorie RAM și 2.5GHz frecvența de ceas, am reușit să obținem o medie pentru recunoaștere fonemelor utilizând trăsăturile MFCC de 5.28s și 105.53 pentru LPC.

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

8.88.6 Concluzii

În urma rezultatelor obținute, am concluzionat că trăsăturile MFCC oferă o înțelegere mai bună a discursului imaginat, oferind cele mai bune rezultate de acuratețe (0.39) comparativ cu celelalte trăsături. De asemenea, s-a putut observa că în urma concatenării celor două tipuri de trăsături, MFCC și LPC, într-o matrice tridimensională rezultatele au fost îmbunătățite, acuratețea atingând o valoare medie de 0.38.

Capitolul 9

Recunoașterea fonemelor utilizând MFCC și CNN [23]

În acest capitol s-a urmărit diferențierea celor șapte foneme înregistrate în timpul discursului imaginat din baza de date KODB. S-a proiectat un sistem inteligent generalizat, independent de subiect, bazat pe calcularea a 18 coeficienți MFCC în etapa de extragere de trăsături și o rețea neuronală CNN în etapa de clasificare. Procesarea acesteia a urmat procedura detaliată în *Capitolul 6*. Pe lângă aceste etape de preprocesare, în acest studiu a fost introdus și un filtru trece bandă având lărgimea de bandă cuprinsă între 0.5 și 100 Hz. În această etapă a studiului recunoașterii discursului imaginat, s-au eliminat canalele din zona occipitală, fiind localizate în zona cortexului vizual. În plus, au fost eliminate și canalele din zona frontală deoarece aceste canale sunt în general puternic afectate de mișcările oculare. Astfel, după eliminarea canalelor s-au utilizat mai departe trăsăturile extrase pentru 45 din cele 62 de canale. Aceste trăsături au fost transmise mai departe către o rețea neuronală de tipul CNN pentru clasificarea fonemelor.

9.1 Rezultate

Scopul acestui capitol a fost de a diferenția cele șapte foneme prezente în baza de date Kara One. Pentru fiecare dintre semnalele din baza de date s-au calculat coeficienții MFCC pentru 45 de canale rezultând o matrice de dimensiune [45 x 18] pentru fiecare rostire. Trăsăturile extrase au fost introduse într-o rețea neuronală de tipul CNN având trei straturi convoluționale 2D urmate de max-pooling și două straturi complet conectate. Utilizând această arhitectură, cele mai bune rezultate obținute au fost de 24.19% acuratețe pentru setul de testare.

9.2 Concluzii

În acest capitol s-a urmărit dezvoltarea unui sistem inteligent de recunoaștere a fonemelor din baza de date KODB. Sistemul dezvoltat este unul generalizat subiecților din baza de date, fiind independent de subiect. Acesta s-a bazat pe calcularea coeficienților MFCC în

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

etapa de extragere de trăsături, împreună cu o rețea neuronală de tipul CNN în etapa de clasificare.

Studiul a arătat că semnalele EEG conțin informații ascunse profund în ceea ce privește discursul imaginat, iar caracteristicile utilizate în mod obișnuit în recunoașterea automată a vorbirii, precum MFCC, conțin markeri esențiali și în ceea ce privește vorbirea imaginată.

Utilizând 18 coeficienți MFCC și o rețea CNN având trei straturi convoluționale de dimensiune 64, 32 și respectiv 32, fiecare dintre ele urmate de un strat max-pooling, și două straturi complet conectate de dimensiune 16 și respectiv 7, s-a obținut o acuratețe pe setul de test de 24.19%.

Rezultatele obținute arată că utilizarea rețelei neuronale CNN îmbunătățește performanțele sistemului, prin comparație cu studiul [24], în care acuratețea maximă a atins doar 20.80% utilizând aceeași bază de date și trăsături bazate pe coeficienți MFCC, însă utilizând în etapa de clasificare o rețea SVM. De menționat faptul că studiul este specific fiecărui subiect, dar au fost utilizate în analiză toate fonemele și cuvintele din baza de date KODB, diferențiind în final 11 clase.

În urma analizării matricei de confuzie, am putut observa că sistemul a reușit să recunoască cel mai bine /iy/, /uw/, /m/ și /n/, acestea având mecanisme de pronunție diferite. O mare parte dintre fonemele care nu au putut fi recunoscute, precum /tiy/ și /diy/ au fost confundate cu fonemul /iy/, având mecanisme de pronunție asemănătoare. Un element din matricea de confuzie care a prezentat interes a fost, de asemenea, și eroarea rețelei între fonemele /piy/ și /m/. Analizând detaliat mecanismul de pronunție a celor două foneme ce au fost confundate, se poate argumenta faptul că această confuzie este din cauza faptului că ambele consoane, /p/ și /m/, se rostesc prin închideri tranzitorii ale buzelor.

Semnalele preprocesate utilizând filtrul PCA pentru eliminarea artefactelor oculare au condus la un răspuns mai slab al rețelei neuronale, fiind obținută o acuratețe de 14.05%. În urma celor două studii de clasificare a semnalelor EEG preprocesate utilizând metoda PCA care au condus la o performanță mai slabă a sistemului putem concluziona faptul că eliminarea componentelor conduce la o scădere a informației ascunsă în semnalele EEG.

Capitolul 10

Analizarea și clasificarea discursului imaginat utilizând rețeaua SOM

Scopul utilizării rețelei neuronale SOM a fost de a realiza o analiză comparativă a mai multor tipuri de trăsături în vederea observării comportamentului acestora relativ la diferențierea rostirilor imaginate prezente în baza de date. Trăsăturile analizate au fost: (1) Coeficienții MFCC; (2) Covariația încrucișată în domeniul timp; (3) Covariația încrucișată în domeniul frecvență. Studiul realizat în acest capitol a urmărit de asemenea și analiza răspunsului rețelei SOM la adăugarea în etapa de preprocesare a eliminării semnalelor oculare pe baza algoritmului PCA. În a doua etapă a studiului, această rețea a fost utilizată pentru clasificarea datelor de intrare prin crearea unor hărți bidimensionale de clasificare bazate pe răspunsul majorității neuronilor câștigători.

Semnalele preprocesate au fost segmentate în ferestre de 0.25s fără suprapunere, iar 50% dintre ferestrele unei înregistrări au fost distribuite aleator în setul de antrenare, restul de 50% fiind distribuite în setul de test.

10.1 Covariația încrucișată în domeniul timp

Se consideră covariația încrucișată între două canale, $c1$ și $c2$, ca fiind descrisă de ecuația:

$$Cov(X^{c1}(t), X^{c2}(t)) = E[[X^{c1}(t) - E(X^{c1}(t))][X^{c2}(t) - E(X^{c2}(t))]], \quad (10.1)$$

Unde $X^{c1}(t)$ reprezintă semnalul EEG achiziționat pentru canalul $c1$, $X^{c2}(t)$ reprezintă semnalul achiziționat pentru canalul $c2$ și $E[X^{ch}(t)]$ reprezintă media canalului ch (care poate fi $c1$ sau $c2$).

10.2 Covariația încrucișată în domeniul frecvență

Se consideră transformata FFT a unui canal:

$$FX^{ch}(f) = \sum_{t=0}^{n-1} X_t^{ch} e^{-\frac{j2\pi ft}{n}} \quad (10.2)$$

Covariația încrucișată în domeniul frecvență se calculează conform ecuației:

$$Cov(FX^{c1}(t), FX^{c2}(t)) = E[[FX^{c1}(t) - E(FX^{c1}(t))][FX^{c2}(t) - E(FX^{c2}(t))]], \quad (10.3)$$

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

10.3 Reprezentarea spațiului trăsăturilor semnalelor EEG utilizând coeficienții MFCC și rețeaua SOM

În urma analizei vizuale a spațiului trăsăturilor rezultat după introducerea semnalelor în rețeaua SOM s-a observat faptul că trăsăturile MFCC, în acest caz, nu oferă o separabilitate bună a claselor, neuronii majoritari câștigători fiind distribuiți pe întreaga suprafață a hărții SOM. Rezultatele nu au putut fi îmbunătățite nici după mărirea spațiului de intrare de la (31, 31) la (62, 62). De asemenea, nu se pot observa diferențe majore între semnalele neprocesate și semnalele procesate utilizând metoda PCA.

10.4 Reprezentarea spațiului trăsăturilor semnalelor EEG utilizând covariația încrucișată în domeniul timp și rețeaua SOM

Rezultatele calitative observate în acest subcapitol arată că nici în acest caz rețeaua neuronală Kohonen nu reușește să separe cele unsprezece clase ale bazei de date în spațiul de ieșire, neuronii câștigători fiind suprapuși pe întreaga suprafață a hărții rezultate. Conform hărților obținute, putem susține faptul că nu există diferențe majore între reprezentarea spațiului trăsăturilor utilizând semnalele neprocesate față de cele procesate. Clasele rezultate sunt în continuare distribuite pe întreaga suprafață a hărții.

10.610.5 Modificarea Reprezentarea spațiului trăsăturilor utilizând covariația încrucișată în domeniul frecvență și rețeaua SOM

În urma analizei realizate în acest capitol s-a putut observa că există zone de grupare a fonemelor și zone de grupare a cuvintelor, însă nu se pot identifica regiuni specifice fiecărei clase. Se poate observa dispersia claselor pe întreaga hartă a rețelei neuronale, fără posibilitatea de delimitare specifică a regiunilor corespunzătoare diferitelor clase. Creșterea spațiului de ieșire nu a adus îmbunătățiri semnificative în ceea ce privește cartografierea datelor de intrare.

10.910.6 Clasificarea datelor de intrare utilizând rețeaua neuronală SOM

Cele mai bune rezultate au fost obținute după 100 000 iterații utilizând o rețea neuronală SOM de dimensiune (31, 31), împreună cu covariația încrucișată în domeniul frecvență utilizând semnale procesate cu ajutorul metodei PCA, acuratețea atingând o valoare de 28.49%.

10.1010.7 Concluzii

S-a observat în urma analizei calitative că trăsăturile calculate în domeniul frecvență au prezentat o așezare mai bună pe hartă din punct de vedere al diferențierii între clase, însă nu au existat zone diferențiate clar pentru fiecare fonem sau cuvânt din baza de date nici în cazul acestor trăsături. În ceea ce privește clasificarea, se poate considera că datele de intrare corespunzătoare covariației încrucișate în domeniul frecvență au oferit o clasificare mai bună decât cele în domeniul timp, ridicând acuratețea la o valoare de 0.25. Procesând semnalele utilizând metoda PCA, performanța sistemului a crescut la 0.28.

Cu toate acestea rezultatele clasificării utilizând rețeaua neuronală SOM nu depășesc rezultatele obținute în capitolul anterior Capitolul 11, în care s-a utilizat rețeaua neuronală CNN ca formă de clasificare. Acest lucru poate fi explicat de faptul că rețeaua neuronală realizează la rândul său prin straturile convoluționale o extragere de trăsături mai bine adaptată a semnalelor de intrare reușind să modifice spațiul trăsăturilor într-un spațiu ce prezintă o separabilitate mai bună a elementelor din setul de date.

Formatted: Highlight

Formatted: Heading 1

Capitolul 12 Capitolul 11

Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB utilizând CNN [25]

Acest studiu a fost efectuat pe opt subiecți diferiți și a fost proiectat ca un sistem comun subiecților din baza de date. Unul dintre scopurile sistemului a fost de a compara două tipuri diferite de extragere de trăsături: covariația încrucișată în domeniul timp și domeniul frecvență. O altă direcție în care ne-am îndreptat atenția pe parcursul dezvoltării studiului a fost testarea diferitelor lungimi de fereastră de analiză: 0.25s, 0.5s și 1s. În a doua parte a lucrării ne-am focusat pe testarea diferitelor arhitecturi ale rețelei CNN utilizată pentru clasificarea trăsăturilor extrase pentru a determina care dintre ele oferă cea mai bună performanță a sistemului.

12.1 Covariația încrucișată — domeniul timp

Se consideră covariația încrucișată între două canale, $c1$ și $c2$, ea fiind descrisă de ecuația:

$$\text{Cov}(X^{c1}(t), X^{c2}(t)) = E[(X^{c1}(t) - E(X^{c1}(t)))(X^{c2}(t) - E(X^{c2}(t)))], \quad (10.1)$$

Unde $X^{c1}(t)$ reprezintă semnalul EEG achiziționat pentru canalul $c1$, $X^{c2}(t)$ reprezintă semnalul achiziționat pentru canalul $c2$ și $E[X^{ch}(t)]$ reprezintă media canalului ch (care poate fi $c1$ sau $c2$) și se calculează astfel:

$$E(X^{ch}(t)) = \frac{1}{W} \sum_{i=0}^{W-1} x_i^{ch} \quad (10.2)$$

Unde variabila W corespunde ferestrei de analiză a domeniului timp pentru care sunt calculate trăsăturile.

În Figura 9.1 este prezentat un exemplu de imagine 2D corespunzătoare covariației încrucișate în domeniul timp pentru fiecare fonem și cuvânt din baza de date KODB pentru o fereastră de dimensiune 0.25s.

Formatted: Space Before: 12 pt, After: 12 pt

Formatted: Space Before: 12 pt, After: 12 pt

Recunoașterea vorbirii imaginare prin analiza semnalelor EEG

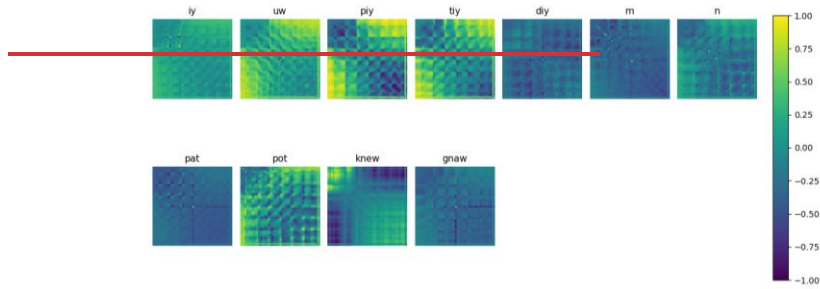


Figura 10.1 Exemplu de imagine 2D corespunzătoare covariației încrucișate calculată în domeniul timp pentru o fereastră de dimensiune 0.25s

12.15 Covariația încrucișată — domeniul frecvență

A doua metodă de extragere a trăsăturilor analizată în această lucrare implică transformarea din domeniul timp a semnalelor de intrare în domeniul frecvență utilizând FFT. Transformata Fourier este o metodă utilizată pentru descompunerea semnalului în sinusuri și cosinusuri.

Fiind dat un semnal discret în timp $x = (x_0, x_1, \dots, x_{n-1})$ care are n eșantioane, transformata Fourier discretă a semnalului poate fi calculată după ecuația:

$$y(\omega) = \sum_{t=0}^{n-1} x_t e^{-j\omega t} \quad (10.3)$$

Formatted: Space Before: 12 pt, After: 12 pt

Unde ω este frecvența unghiulară și este calculată utilizând următoarea ecuație:

$$\omega = \frac{2\pi f}{n} \quad (10.4)$$

Formatted: Space Before: 12 pt, After: 12 pt

Considerând reprezentarea anterioară a frecvenței unghiulare, ω , ecuația (9.3) poate fi înlocuită cu:

$$y(f) = \sum_{t=0}^{n-1} x_t e^{-\frac{j2\pi f t}{n}} \quad (10.5)$$

Formatted: Space Before: 12 pt, After: 12 pt

Revenind la semnalele EEG, transformata FFT a unui canal este calculată utilizând următoarea ecuație:

După transformarea semnalelor în domeniul frecvență utilizând ecuația (9.6), se calculează în continuare covariația încrucișată pe rezultatul transformatei Fourier a canalelor utilizând ecuația (9.1) modificată astfel:

$$\begin{aligned} \text{Cov}(FX^{e1}(t), FX^{e2}(t)) \\ = E\{[FX^{e1}(t) - E(FX^{e1}(t))][FX^{e2}(t) - E(FX^{e2}(t))]\}, \end{aligned} \quad (10.7)$$

În Figura 9.2 este prezentat un exemplu de imagine 2D corespunzătoare covariației încrucișate în domeniul frecvență pentru fiecare fonem și cuvânt din baza de date KODB pentru o fereastră de dimensiune 0.25s.

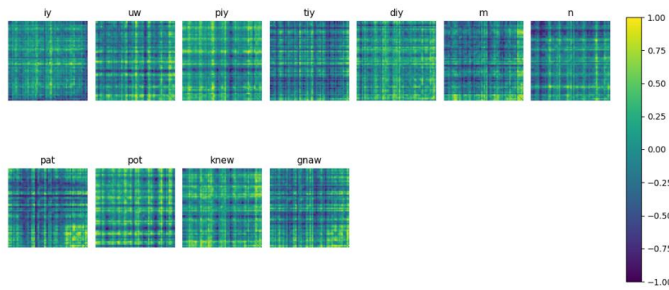


Figura 10.2 Exemplu de imagine 2D corespunzătoare covariației încrucișate calculată în domeniul frecvență pentru o fereastră de dimensiune 0.25s

12.4311.1 Clasificarea semnalelor

În cercetarea realizată în această lucrare, am testat diferite arhitecturi ale rețelelor CNN cu scopul de a găsi arhitectura care oferă cea mai bună performanță, ținând cont însă și de complexitate, memorie și timpul de rulare. Am pornit de la o arhitectură având o complexitate mică, un strat convoluțional și unul complet conectat (fără stratul de ieșire) și am crescut complexitatea până la trei straturi convoluționale și unul complet conectat având un număr mai mare de filtre și neuroni. În acest punct, am considerat că performanțele sistemului nu se îmbunătățesc, în schimb vor fi afectate memoria și timpul de rulare.

În cercetarea realizată în această lucrare, am testat diferite arhitecturi ale rețelelor CNN cu scopul de a găsi arhitectura care oferă cea mai bună performanță, ținând cont însă și de complexitate, memorie și timpul de rulare. Am pornit de la o arhitectură având o complexitate mică, un strat convoluțional și unul complet conectat (fără stratul de ieșire) și am crescut complexitatea până la trei straturi convoluționale și unul complet conectat având un număr mai mare de filtre și neuroni. În acest punct, am considerat că performanțele sistemului nu se îmbunătățesc, în schimb vor fi afectate memoria și timpul de rulare.

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

11.2 Metrice de evaluare a sistemului

Pentru evaluarea performanțelor sistemului au fost urmărite o serie de metrice precum acuratețea, acuratețea echilibrată, precizia și sensibilitatea care pot oferi informații cantitative cu privire la gradul de recunoaștere a fonemelor și cuvintelor prezente în baza de date.

12.4411.3 Rezultate

12.44.111.3.1 Analizarea funcțiilor de activare: Tanh și ReLU

Rezultatele obținute pe setul de testare utilizând diferite arhitecturi ale rețelei CNN și diferite funcții de activare pentru straturile convoluționale: tangentă hiperbolică versus unitate liniară rectificată (ReLU – Rectified Linear Unit) arată că funcția de activare ReLU oferă rezultate semnificativ mai bune, ridicând acuratețea maximă de la 0.3169 (tanh) la 0.3758 (ReLU). Rezultatele s-au obținut utilizând ca metodă de extrageri de trăsături covariația încrucișată în domeniul timp pe o fereastră de 0.25s.

12.44.211.3.2 Studiul trăsăturilor: Timp vs Frecvență

În continuare, studiul a urmărit compararea diferențelor dintre trăsăturile calculate în domeniul timp și în domeniul frecvență. Un studiu al diferitelor arhitecturi ale rețelei CNN arată că utilizarea a două straturi convoluționale având numărul de filtre 64 și 128 conectate la un strat complet conectat de dimensiune 64 de neuroni funcționează cel mai bine pentru caracteristicile calculate în domeniul frecvență (caracteristicile care au oferit, de asemenea și cea mai bună acuratețe), obținându-se performanța de 37% acuratețe. În ceea ce privește domeniul timp, cele mai bune rezultate au fost obținute utilizând arhitecturi mai puțin complexe, cele mai bune performanțe ale sistemului fiind surprinse de o rețea cu un singur strat convoluțional având 64 de filtre și un strat complet conectat cu 64 de neuroni.

12.44.311.3.3 Analiza lungimilor de prelucrare: 0.25, 0.5 și 1s

Următorul pas a fost testarea rețelei cu diferite lungimi ale ferestrei de analiză aplicată pe datele de intrare: 0.25s, 0.5s și 1s. În urma studiului s-a putut observa că rezultatele obținute pentru fereastra de analiză de 0.25s sunt cele mai bune, atingând o acuratețe de 37%.

12.44.411.3.4 Comparație filtru de mediere: B0, B3 și B5

Un alt studiu urmărit în această etapă a lucrării s-a axat pe aplicarea unui filtru de mediere pe spectru înainte de calcularea matricei de covariație pe canale având nucleu de diferite mărimi: trei și cinci eșantioane. Rezultatele studiului nu au arătat îmbunătățiri în ceea ce privește acuratețea rețelei

pentru niciuna dintre dimensiunile nucleului. Valoarea maximă în cazul utilizării filtrului cu nucleul egal cu trei eșantioane fiind 0.2886, iar pentru nucleul de cinci eșantioane 0.2863.

12.44.511.3.5 Metrici de evaluare a performanțelor sistemului

Pentru o mai bună înțelegere a rezultatelor înregistrate precum și a performanțelor sistemului, s-a introdus o serie de metrici noi: acuratețea echilibrată, precizia și senzitivitatea, calculate conform articolului [26]. Rezultatele arată că nu există un dezechilibru semnificativ al semnalelor din baza de date.

12.44.611.3.6 Măsurători de complexitate și memorie

Utilizând un sistem AMD Ryzen 7 4800HS CPU având 16 GB de memorie RAM și 2.9 GHz frecvența de ceas, s-a obținut un timp mediu de recunoaștere a unui vector de intrare de 1.8×10^{-3} s pornind de la etapa de extragere a trăsăturilor până la luarea deciziei. Timpul a fost estimat utilizând covariația încrucișată în domeniul frecvență pe o fereastră de 0.25s introdusă între rețea CNN având arhitectura C64-128/D64.

11.3.7 Comparatie între semnalele neprocesate și semnalele procesate utilizând PCA pentru eliminarea artefactelor oculare

În urma analizării rezultatelor s-a observat că utilizând analiza PCA pentru eliminarea artefactelor oculare performanțele rețelei CNN scad de la 0.37 la 0.35, rezultat diferit față de cel obținut utilizând rețeaua neuronală SOM.

12.4511.4 Discuție

În urma comparației dintre cele două metode de extragere de trăsături, covariația încrucișată în domeniul timp și frecvență s-a putut observa că în domeniul frecvență acuratețea crește cu aproximativ 16% atingând o valoare de 0.37 comparativ cu 0.21, valoare obținută în urma utilizării trăsăturilor în domeniul timp. Un alt element pe care l-a urmărit acest studiu a fost compararea diferitelor dimensiuni ale ferestrelor de analiză cu scopul de a observa statistica semnalului pentru intervale diferite de timp. În urma acestei comparații, s-a putut observa că cea mai bună fereastră de analiză este de 0.25s. Acuratețea obținută pentru această lungime a ferestrei este semnificativ mai mare, atingând o valoare de 0.37, comparativ cu 0.29 rezultat obținut în urma utilizării unei ferestre de 1s. Diferența dintre performanțele obținute pentru o fereastră de 0.25s și 0.5s nu este semnificativă, aceasta scade cu doar 1%. În timpul cercetărilor noastre pentru îmbunătățirea sistemului, am încercat de asemenea aplicarea unui filtru de mediere asupra spectrului utilizând un nucleu de trei și cinci eșantioane. S-a observat în acest studiu că aplicând un filtru asupra spectrului nu a fost crescută acuratețea sistemului, ci din potrivă, aceasta a scăzut cu aproximativ 9% când au fost utilizate filtrele de dimensiune trei și cinci eșantioane. Studiul final s-a bazat pe testarea diferitelor arhitecturi ale rețelei CNN pentru a observa performanțele sistemului și pentru a modela caracteristicile finale ale acestuia. Am concluzionat că în cazul trăsăturilor extrase în domeniul frecvență (trăsăturile care au oferit și cea mai bună performanță a sistemului) cea mai bună arhitectură folosită conține doar două straturi convoluționale 2D având 64 și 128 de filtre conectate cu un strat complet conectat ce conține 64 de neuroni.

12.4611.5 Concluzii

Studiul realizat în acest capitol și a propus analizarea semnalelor EEG pentru recunoașterea rostirii imaginate a șapte foneme și patru cuvinte. Pentru a se îndeplini scopul propus, am dezvoltat un sistem inteligent partajat de subiecți utilizând un lanț de procesare aplicat pe semnalele din baza de date KODB [38]. Lanțul de analiză a debutat cu preprocesarea semnalelor de intrare având scopul de a îmbunătăți calitatea acestora. Mai departe, în etapa de extragere de trăsături, au fost comparate rezultatele obținute după calcularea covariației încrucișate pe canale în domeniul timp și în domeniul frecvență. Studiul a continuat cu analiza diferitelor lungimi ale ferestrei de analiză: 0.25, 0.5 și 1s pentru a găsi fereastra în care semnalul devine evasi staționar sau aproape evasi staționar, dar care conține de asemenea și informații cu privire la rostire. A fost studiată și comportarea sistemului după aplicarea unui filtru de mediere având nucleul egal cu trei sau cinci eșantioane, presupunând că analizarea mai multor eșantioane adiacente vor descrie mai bine spectrul rostirilor claselor diferite. În final, în etapa de clasificare, am testat arhitecturi multiple ale rețelei CNN pentru a vedea care dintre ele oferă performanțele cele mai bune.

Cele mai bune rezultate au fost obținute utilizând covariația încrucișată pe canale în domeniul frecvență utilizând o fereastră de analiză de 0.25s. Cea mai bună performanță a sistemului a fost obținută utilizând o rețea CNN cu două straturi convoluționale 2D având 64 și 128 de filtre și un strat complet conectat având numărul de neuroni egal cu 64. Utilizând aceste caracteristici ale sistemului, acuratețea maximă a sistemului a atins o valoare de 0.37. Am arătat de asemenea și că o fereastră de analiză mai mică oferă o înțelegere mai bună a discursului imaginat. În final, putem susține că sistemul propus se poate implementa pe un dispozitiv portabil de cost redus având resurse limitate pentru a lua deciziile cu privire la rostirea imaginată a fonemelor sau cuvintelor.

~~Capitolul 13~~

~~Analizarea și clasificarea discursului imaginat utilizând rețeaua SOM~~

Formatted: Heading 1

~~În acest capitol al lucrării s-a urmărit
maparea trăsăturilor utilizând rețeaua
neuronă Self Organizing Map
(SOM), denumită și Kohonen după
profesorul finlandez Tuevo Kohonen
care în anul 1980 a introdus această
rețea neuronală [72] ce are ca scop
organizarea datelor de intrare în hărți
de dimensiune redusă care grupează
datele pe baza similitudinii acestora,
asemănător funcționării cortexului
vizual.~~

~~Scopul inițial al utilizării rețelei neuronale SOM a fost de a analiza trăsăturile setului de date bazate pe covariația încrucișată în vederea obținerii unui spațiu bidimensional care să descrie cât mai bine fonemele și cuvintele analizate pentru recunoașterea discursului imaginat. În a doua etapă a studiului, această rețea a fost utilizată pentru clasificarea datelor de intrare prin crearea unei mape bidimensionale de clasificare bazată pe răspunsul majorității neuronilor câștigători.~~

~~Setul de date pentru care s-au obținut rezultatele este KODB descris detaliat în Capitolul 4.~~

Formatted: Heading 1, Indent: First line: 0"

~~20.0 Maparea semnalelor EEG bazate pe covariația încrucișată în domeniul timp utilizând SOM~~

~~Reteaua neuronală SOM are capacitatea de a grupa datele de intrare conform similitudinii acestora. Inițial, studiul din acest capitol s-a axat pe analizarea trăsăturilor bazate pe covariația încrucișată în domeniul timp. Pentru realizarea scopului propus, pentru semnalele din baza de date a fost calculată covariația în domeniul timp conform metodei detaliate în *Subcapitolul 9.1*.~~

Formatted: Heading 1

~~Trăsăturile extrase au fost în continuare liniarizate pentru crearea unui spațiu unidimensional și introduse într-o rețea neuronală de tipul SOM. În Figura 10.1 (a) sunt reprezentați neuronii câștigători din spațiul bidimensional de ieșire al rețelei SOM de dimensiune (31, 31) pentru fiecare vector de intrare, fiecare etichetă fiind reprezentată cu o culoare diferită. În Figura 10.2 sunt reprezentații vectorii de intrare după antrenarea unei rețele neuronale SOM având dimensiunea la ieșire de (62, 62).~~

Formatted: Heading 1, Indent: First line: 0"

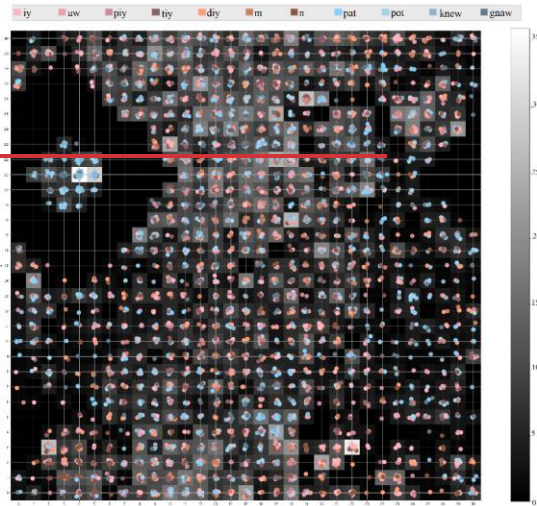


Figura 11.1 ~~Reprezentarea covariației încrucișate în domeniul timp în spațiul de ieșire SOM, având dimensiunea (31, 31) după 10000 iterații~~

Formatted: Heading 1, Left, Indent: First line: 0", Don't keep with next

Formatted: Heading 1

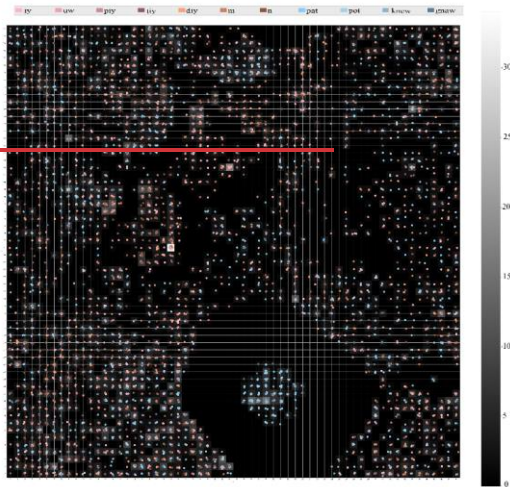


Figura 11.2 ~~Reprezentarea covariației încrucișate în domeniul timp în spațiul de ieșire SOM, având dimensiunea (62, 62) după 10000 iterații~~

Formatted: Heading 1, Left, Don't keep with next

Formatted: Heading 1

~~Rezultatele obținute în figurile Figura 10.1 și Figura 10.2 arată că rețeaua neuronală Kohonen nu reușește să separe cele unsprezece clase ale bazei de date în spațiul de ieșire, neuronii câștigători fiind suprapuși pe întreaga suprafață a hărții rezultate. Acest rezultat trimite la concluzionarea faptului că trăsăturile bazate pe covariația în domeniul timp nu reușese să extragă informațiile ascunse din semnalele EEG cu privire la recunoașterea discursului imaginat~~

~~28.0 Maparea trăsăturilor bazate pe covariația încrucișată în domeniul frecvență utilizând SOM~~

Formatted: Heading 1, Indent: First line: 0"

Formatted: Heading 1

~~Semnalele bazei de date KODB au fost trecute din nou printr-o etapă de procesare, mai exact de extragere a trăsăturilor, necesară pentru a extrage informațiile asunse din semnal. În această etapă a fost utilizată covariația încrucișată în domeniul frecvență. Această metodă de extragere a trăsăturilor prezintă marele avantaj de reducere a cantității de informație din semnal, precum și înglobarea informațiilor referitoare la conexiunile dintre electrozi prezente în timpul transmiterii impulsului nervos. Realizarea covariației pe canale în domeniul frecvență avantajează procesul datorită eliminării posibilelor întârzieri pe electrozi ce pot apărea în~~

~~domeniul timp la transmiterea impulsului.~~

~~Metoda utilizată pentru calcularea covariației încrucișate pe canale în domeniul frecvență este descrisă pe larg în capitolul anterior, *Subcapitolul 9.2.*~~

~~Trăsăturile extrase au fost trimise mai departe către o rețea neuronală SOM de dimensiune de (31, 31) și o rețea neuronală de dimensiune (62, 62) neuroni dispuși bidimensional.~~

Formatted: Heading 1, Indent: First line: 0"

~~În Figura 10.3 sunt reprezentați vectorii de intrare în spațiul bidimensional obținut după antrenarea rețelei SOM de dimensiune (31, 31) pentru un număr de 10000 de iterații. Se poate observa în imaginea rezultată că există zone în spațiul bidimensional care răspund la stimulii de intrare și sunt neuroni care nu sunt câștigători pentru niciuna dintre valorile date la intrare (zonele marcate cu negru). Se poate observa de asemenea că există zone de grupare a fonemelor și zone de grupare a cuvintelor, însă nu se pot identifica regiuni specifice fiecărei clase.~~

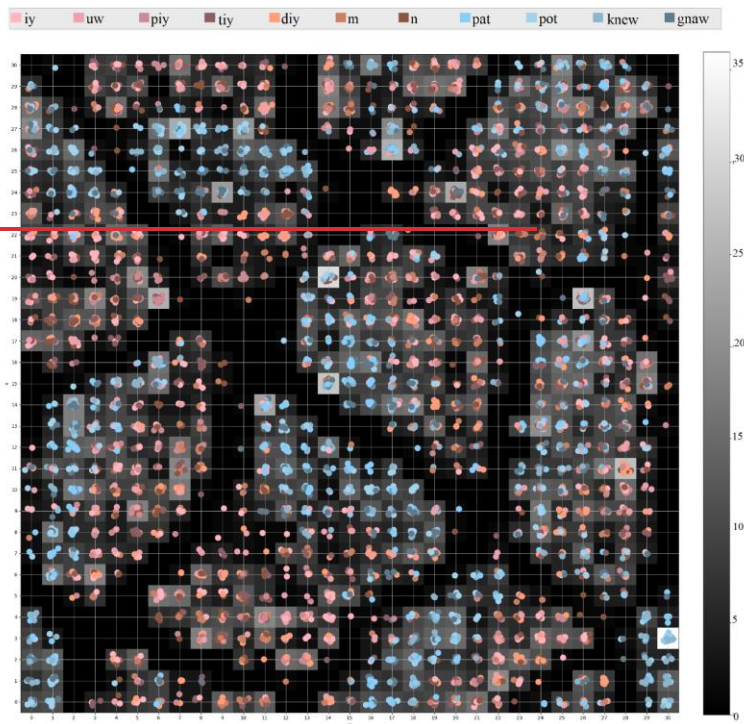


Figura 11.3 ~~Reprezentarea covariației încrucișate în domeniul frecvență în spațiul de ieșire SOM, având dimensiunea (31, 31) după 10000 iterații~~

Formatted: Heading 1, Left, Don't keep with next

Formatted: Heading 1

~~Zonele corespunzătoare răspunsului neuronilor pentru fiecare clasă pot fi observate în Figura 10.4. În urma analizării acestei figuri, putem observa tot dispersia claselor pe întreaga mapă a rețelei neuronale, fără posibilitatea de delimitare specifică a regiunilor corespunzătoare diferitelor clase.~~

~~După 100000 de iterații, se poate vedea o îmbunătățire a așezării vectorilor de intrare pe hartă, însă din nou nu se pot delimita clar zone corespunzătoare fiecărei clase. Rezultatul obținut este prezentat în Figura 10.5 și Figura 10.6.~~

Formatted: Heading 1, Indent: First line: 0"

Formatted: Heading 1, Indent: First line: 0", Space After: 0 pt, Line spacing: single

~~Creșterea spațiului de ieșire nu a adus îmbunătățiri semnificative în ceea ce privește maparea datelor de intrare. Se poate observa în Figura 10.8 (rezultatul obținut după 10000 iterații), dar și în Figura 10.6 (rezultatul obținut după 100000 iterații) că și în acest caz există zone specifice cuvintelor și zone specifice fonemelor, însă există confuzie între cuvinte și între foneme.~~

Formatted: Heading 1, Indent: First line: 0"

Capitolul 1 -

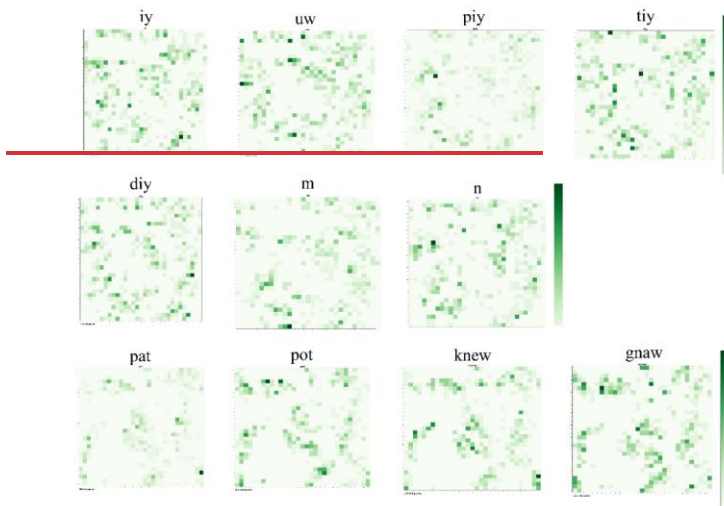


Figura 11.4 Răspunsul neuronilor pentru fiecare clasă la primirea stimulilor de intrare de antrenare; trăsături bazate pe covariația încrucișată în domeniul frecvență; 10000 iterații; dimensiune spațiu (31, 31)

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

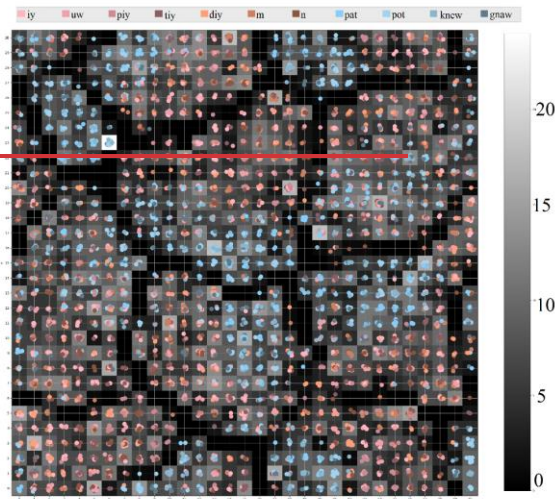


Figura 11.5 ~~Reprezentarea covariației încrucișate în domeniul frecvență în spațiul de ieșire SOM, având dimensiunea (31, 31) după 100000 iterații~~

Formatted: Heading 1, Left, Space After: 0 pt, Line spacing: single, Don't keep with next

Formatted: Heading 1

Capitolul 1 -

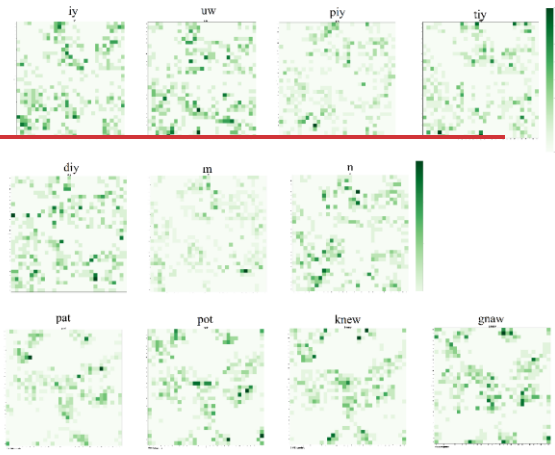


Figura 11.6 Răspunsul neuronilor pentru fiecare clasă la primirea stimulilor de intrare de antrenare; trăsături bazate pe covariația încrucișată în domeniul frecvență; 100000 iterații; dimensiune (31, 31)

Formatted: Heading 1, Left, Don't keep with next

Formatted: Heading 1

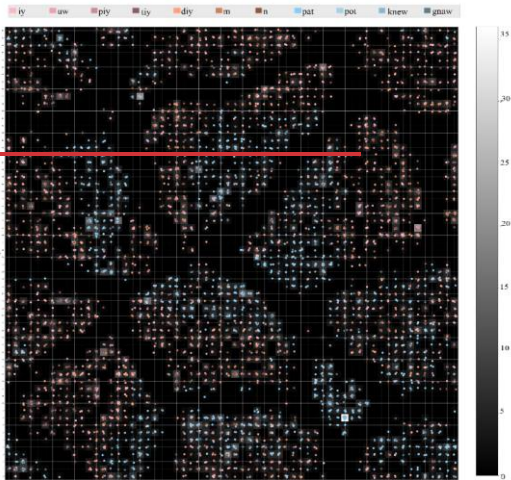


Figura 11.7 ~~Reprezentarea covariației~~
~~înerucișate în domeniul frecvență în~~
~~spațiul de ieșire SOM, având~~
~~dimensiunea (62, 62) după 10000~~
~~iterații~~

Formatted: Heading 1, Left, Indent: First line: 0", Don't keep with next

Formatted: Heading 1

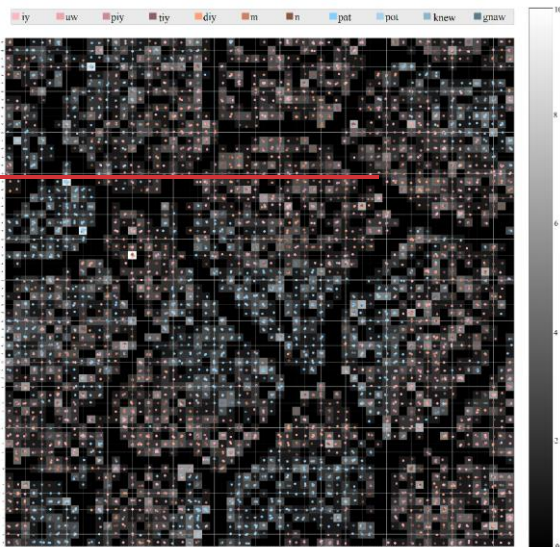


Figura 11.8 ~~Reprezentarea covariației încrucișate în domeniul frecvență în spațiul de ieșire SOM, având dimensiunea (62, 62) după 100000 iterații~~

~~51.0~~ Clasificarea datelor de intrare utilizând rețeaua neuronală SOM

Formatted: Heading 1, Left, Don't keep with next

Formatted: Heading 1

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

**Tabelul 11.1 Rezultatele clasificării
semnalelor din setul de test utilizând
rețeaua SOM**

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

Capitolul 82 Capitolul 12

Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB utilizând CNNLSTM [27]

Capitolul 82 Capitolul 12

Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB utilizând CNNLSTM [27]

~~S-a observat că trăsăturile calculate în domeniul frecvență au prezentat o așezare mai bună pe hartă din punct de vedere al diferențierii între clase, însă nu au existat zone diferite clare pentru fiecare fonem sau cuvânt din baza de date nici în cazul acestor trăsături. În schimb, în domeniul frecvență s-au putut vedea grupări ale fonemelor și ale cuvintelor în zone diferite ale hărții, făcându-se o diferențiere mai bună între aceste două clase.~~

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

Capitolul 82 Capitolul 12

Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB utilizând CNNLSTM [27]

~~Capitolul 82~~ Capitolul 12

Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB utilizând CNNLSTM [27]

În acest capitol s-a urmărit comportarea sistemului de recunoaștere a discursului imaginat utilizând în etapa de clasificare o Rețea Neuronală Convoluțională ce înglobează în straturile convoluționale celule recurente de tipul Long-Short Term Memory (CNNLSTM).

Acest studiu a pus de asemenea în evidență faptul că utilizând în analiză semnale provenite de la zonele anatomice recunoscute pentru implicațiile în producerea discursului: aria Broca, cortexul motor primar și cortexul motor secundar, se păstrează aproximativ 93% din informația obținută de la toți electrozii pentru care au fost achiziționate semnalele.

82.1 12.1 **Calcularea trăsăturilor**

Trăsăturile calculate pentru dezvoltarea sistemului au fost alese conform rezultatelor studiului anterior. Astfel că, în acest studiu s-a urmărit doar calcularea covariației încrucișate în domeniul frecvență pentru segmente de 0.25s. Pentru a oferi variația în timp rețelei CNNLSTM, segmentele de 0.25ms au fost împărțite la rândul lor în ferestre de 0.1s cu suprapunere de 50%.

În continuare, s-au investigat calitativ trăsăturile calculate utilizând LDA pentru reducerea numărului de trăsături la un spațiu bidimensional astfel încât să se poată realiza inspecția vizuală a acestora. După analiza vizuală și cantitativă a trăsăturilor amintite anterior, am concluzionat că trăsăturile calculate utilizând covariația încrucișată în domeniul frecvență împarte cel mai bine spațiul trăsăturilor în ceea ce privește discursul imaginat și doar acestea au fost utilizate mai departe în studiu.

82.2 12.2 **Clasificarea utilizând CNNLSTM**

Arhitectura rețelei neuronale se bazează de asemenea pe rezultatele studiilor obținute anterior în *Capitolul 11*. Cele mai bune rezultate obținute, după testarea mai multor arhitecturi și hiperparametrii s-au obținut utilizând două straturi convoluționale de dimensiune 64 și 128 conectate la un strat complet conectat ce conține 64 de neuron, stratul

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

de ieșire având 11 neuroni corespunzători numărului de clase. În final rețeaua neuronală a fost antrenată utilizând optimizatorul Adam având o rată de învățare de 0.0001 și folosind ca funcție eroare entropia încrucișată.

82.312.3 Rezultate

82.3.112.3.1 CNNLSTM vs CNN

Acest studiu are ca scop evidențierea avantajelor utilizării rețelei neuronale CNNLSTM pentru recunoașterea fonemelor și cuvintelor rostite în timpul discursului imaginat utilizând în etapa de extragere de trăsături covariația încrucișată pe canale. Rezultatele obținute în această etapă arată o îmbunătățire față de rețeaua CNN, acuratețea crescând de la 37% la 43%.

82.3.212.3.2 Analiza regiunilor craniene

Următorul studiu realizat în acest capitol evaluatează performanțele sistemului relativ la reducerea numărului de semnale astfel încât să se analizeze doar semnalele preluate de la anumite regiuni craniene.

Regiunile au fost selectate inițial pe baza regiunilor majore ce definesc și pozițiile electrozilor, regiunile: Frontală, Centrală și Occipitală. Apoi s-au selectat electrozii corespunzători zonelor anatomice implicate în realizarea discursului, pornind de la conceptualizare și planuirea mișcărilor articulare până la inițierea și coordonarea neuronilor implicați în transmiterea stimulului electric trimis către efectori. Rezultatele obținute pentru fiecare zonă a creierului și combinațiile realizate între aceste zone analizate în acest studiu au arătat că cea mai bună rată de recunoaștere o au electrozii poziționați în regiunile anatomice specifice discursului, atingând o valoare de 0.4027.

82.3.312.3.3 Analiza complexității de calcul și a memoriei

Această secțiune a capitolului s-a axat pe studiul complexității și memoriei sistemului propus. Complexitatea unui sistem inteligent este dată în general de rețeaua neuronală. În cazul de față, complexitatea maximă este dată de cel de-al doilea strat convoluțional și este de forma $O(4 \times 4 \times (2(N+3)^2 \times 64) \times \log(2(N+3)^2 \times 64) \times 128)$. În ceea ce privește timpul de execuție, măsurat utilizând un procesor AMD Ryzen 7 4800HS cu 16GB memorie RAM și 2.9 GHz frecvența de ceas, valoarea medie pe vectorii din setul de date este de 81.9ms.

12.3.4 Analiza semnalelor neprocesate și semnalelor procesate utilizând PCA

A fost introdusă în acest capitol o analiză cu privire la performanțele sistemului utilizând semnale neprocesate și semnale procesate în vederea eliminării artefactelor oculare utilizând PCA. Rezultatele obținute oferă informații asemănătoare cu cele din capitolul anterior, în care a fost utilizată rețeaua CNN pentru clasificare. Scăderea performanței sistemului la utilizarea semnalelor procesate poate fi explicată de același proces în care rețelele neuronale cu învățare profundă folosesc propriile filtre pentru a extrage din semnal informația esențială clasificării și pentru a elimina artefactele sau elementele ce pot perturba procesul de învățare.

Capitolul 12 - Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB utilizând CNNLSTM

82.412.4 Discuții

Avantajele majore ale unei rețele LSTM este capacitatea de memorare pe termen lung a trăsăturilor introduse la intrare. Această abilitate are o valoare semnificativă în momentul în care se analizează semnale netaționare invariante în timp, precum sunt semnalele EEG. Această legătură timp-spațiu a ajutat rețeaua neuronală să ridice acuratețea de la 0.37 la 0.43 utilizând arhitecturi și parametrii similari. Matricea de confuzie medie pentru cele 4-fold-uri arată că există o distincție destul de clară între foneme și cuvinte, acestea fiind foarte rar încurcate între ele, confuzia rețelei făcându-se între foneme și între cuvinte.

Scopul final al sistemului dezvoltat a fost de a obține o acuratețe cât mai mare a acestuia cu posibilitatea implementării într-un dispozitiv portabil având resurse limitate. În continuare am studiat comportarea sistemului utilizând un număr redus de electrozi localizați în arii specifice. Acest lucru ajută la portabilitatea dispozitivului și reduce resursele necesare pentru dezvoltare, însă având dezavantajul unei ușoare descreșteri ale acurateței sistemului. Prin reducerea numărului de canale, acuratețea a scăzut de asemenea, fiind un fenomen anticipat. Cu toate acestea, utilizând electrozii din zonele anatomice specifice implicate în realizarea discursului, acuratețea sistemului a atins o valoare de 0.40, fiind o scădere de doar 3% în comparație cu utilizarea tuturor canalelor. Acest lucru înseamnă că 93% din informația discursului imaginat este concentrat în aceste canale și doar 7% din informație este distribuită regiunilor parietale și occipitale.

Un aspect important care trebuie luat în considerare în momentul dezvoltării unui sistem de recunoaștere a discursului imaginat este complexitatea și memoria utilizată. În general, cel mai mare consumator al resurselor este rețeaua neuronală. Cel mai mare număr de operații computaționale este dat de cel le-al doilea strat al rețelei CNNLSTM și este de ordinul aprox. $O(6.3 \times 10^9)$. Cu toate acestea, timpul de execuție pentru a lua o decizie în ceea ce privește un stimul primit la intrare de către rețea este sub 100ms, chiar și utilizând toate canalele în procesul computațional. Aceste valori indică faptul că sistemul poate fi implementat în continuare în timp real. În ceea ce privește memoria utilizată, sistemul prezintă o limitare deoarece sunt necesari minim 2GB pentru a reține doar ponderile rețelei neuronale. Acest lucru se datorează memoriei pe termen lung a rețelei utilizate.

82.512.5 Concluzii

Acest capitol a arătat o îmbunătățire a performanței sistemului la utilizarea rețelei CNNLSTM în comparație cu rețeaua CNN. Acuratețea a crescut de la 37% la 43% la utilizarea rețelei CNNLSTM fără modificări în lanțul de preprocesare sau extragere de trăsături. Sistemul dezvoltat și-a propus să considere în proiectare și posibilitatea implementării în timp real pe un dispozitiv portabil cu resurse limitate. Astfel că, s-a realizat și un studiu în ceea ce privește reducerea numărului de electrozi din componența sistemului. Am concluzionat în urma studiului că 93% din informație este concentrată în regiunile anatomice specifice producerii discursului, obținând o acuratețe de 40% pentru utilizarea a 29 de electrozi, comparativ cu 62, acesta fiind numărul inițial al acestora.

Capitolul 13

Concluzii

În această lucrare s-a urmărit dezvoltarea unui sistem inteligent de recunoaștere automată a discursului imaginat. Pentru atingerea scopului propus, inițial a fost realizat un studiu al mecanismului de pronunție a sunetelor pornind de la intenția rostirii ce apare la nivel cortical până la transmiterea impulsului electric către organele efectuate implicate în procesul rostirii.

Această teză a utilizat pentru dezvoltarea sistemului baza de date Kara One achiziționată în timpul colaborării Universității din Toronto cu Centrul de Reabilitare din Toronto. Baza de date conține semnale preluate în timpul discursului imaginat a șapte cuvinte și patru foneme. În continuare semnalele au fost analizate și preprocesate în vederea creșterii calității acestora pentru a putea fi utilizate mai departe în studiu. Preprocesarea a constat din analizarea vizuală a acestora de către un expert și eliminarea epocilor ce conțin semnale zgomotoase sau electrozi cu conectivitate slabă, urmând să fie filtrate eliminând zgomotul de rețea. Principalele artefacte ale semnalelor EEG sunt date de mișcarea oculară, deoarece se înregistrează activitatea electrică a mușchilor de la nivelul ochilor care au o amplitudine mai mare decât semnalul EEG, motiv pentru care au fost implementate și două metode de eliminare a artefactelor oculare. Prima metodă a constat din filtrarea acestora utilizând un filtru adaptiv, iar a doua metodă s-a bazat pe eliminarea surselor semnalului care conțin aceste mișcări prin separarea în componente principale utilizând algoritmul PCA.

Următorul studiu realizat a urmărit posibilitatea de recunoaștere a trei tipuri diferite de mecanisme fonetice: (a) rostirea fonemului /iy/ (/iy/, /piy/, /tiy/, /diy/) (b) rostirea fonemului /uw/ (/uw/) și (c) rostirea consoanelor (/m/ și /n/). Acest studiu a demonstrat că există makeri descriptivi pentru diferite mecanisme de pronunție la analizarea semnalelor EEG. Mergând mai departe, a fost creat un sistem de diferențiere a tuturor fonemelor (șapte clase) din baza de date. În acest punct canalele occipitale au fost eliminate pentru a nu influența răspunsul sistemului, ținând cont de faptul că stimulul a fost aplicat vizual. În etapa de extragere a trăsăturilor s-a utilizat rețeaua neuronală CNN pentru considerarea în clasificare și a legăturilor spațiale ale electrozilor.

În capitolul următor, s-a încercat cu ajutorul rețelei neuronale SOM reprezentarea trăsăturilor setului de date în spațiul bidimensional cu scopul de a modifica spațiul trăsăturilor într-un spațiu cu separabilitate mai mare între clase. Studiul este unul

Formatted: Indent: First line: 0.5"

comparativ între trăsăturile: (1) MFCC, (2) covariația încrucișată în domeniul timp și (3) covariația încrucișată în domeniul frecvență.

Studiul realizat în *Capitolul 11* s-a axat pe clasificarea tuturor fonemelor și cuvintelor prezente în baza de date KODB. În acest capitol au fost urmărite mai multe obiective, printre care: (a) influența hiperparametrilor CNN; (b) modificarea arhitecturii rețelei; (c) impactul diferitelor funcții de activare utilizate pentru straturile rețelei CNN; (d) trăsături diferite capabile să decodifice informațiile ascunse din semnalele EEG prin calcularea covarianței încrucișate în domeniul timp și în domeniul frecvență; (e) dimensiuni diferite ale ferestrelor de analiză pentru metodele de extragere de trăsături; (f) aplicarea unui filtru de mediere având nucleul de trei (B3) și cinci (B5) eșantioane aplicat pe spectrul semnalelor.

~~În capitolul următor, s-a încercat cu ajutorul rețelei neuronale SOM mapearea trăsăturilor setului de date cu scopul de a modifica spațiul trăsăturilor într-un spațiu cu separabilitate mai mare între clase. Studiul este unul comparativ între trăsăturile calculate pe baza covariației încrucișate în domeniul timp și în domeniul frecvență.~~

Ultimul studiu a inclus implementarea unui dispozitiv inteligent de recunoaștere a fonemelor și cuvintelor din baza de date Kara One utilizând covariația încrucișată în domeniul frecvență și rețeaua convoluțională CNLSTM. În timpul dezvoltării sistemului, a fost studiată și performanța sistemului la analiza diferitelor regiuni craniene: frontală, centrală și occipitală pentru jumătatea stângă (S) și dreaptă (D) a acestora, precum și combinații între regiunile menționate. Au fost selectate și regiunile anatomice implicate în procesul de producere a vorbirii. În final, s-a urmărit și analiza complexității și memoriei sistemului pentru posibilitatea implementării pe un dispozitiv portabil cu resurse limitate.

83.113.1 Rezultate

83.1.113.1.1 Capitolul 7: Eliminarea artefactelor oculare

Acest capitol a urmărit eliminarea artefactelor oculare din semnalele EEG pentru a îmbunătăți calitatea acestora. Pentru a realiza acest lucru am testat două metode diferite de filtrare, una bazată pe filtru adaptiv, iar cealaltă pe calcularea componentelor principale.

În urma acestui studiu, am concluzionat faptul că ambele metode atenuază artefactele oculare prin decorelarea semnalelor cu HEO. S-a putut observa că inițial s-a pornit de la o corelație foarte mare dintre semnalele achiziționate de pe electrozii frontali și HEO, având o valori ce depășesc 0.8 și ajungând la valori de aproximativ 0.1 pentru filtru adaptiv și aproximativ 0.2 după filtrarea cu PCA. Am observat, în continuare, că semnalele după ce au fost filtrate utilizând metoda PCA au prezentat o corelație mai mare atunci când s-a urmărit relația dintre înregistrările aceleiași clase, spre deosebire de semnalele obținute după aplicarea filtrului adaptiv, unde a apărut o decorelare între acestea.

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

83.1.213.1.2 Capitolul 8: Sistem de recunoaștere a mecanismelor diferite de pronunție

Acest studiu a avut ca scop confirmarea posibilității de a diferenția trei mecanisme de pronunție diferite prin analizarea doar a semnalelor EEG achiziționate în timpul discursului imaginat. De asemenea, am urmărit comportarea trăsăturilor utilizate în mod obișnuit în recunoașterea automată a vorbirii transpuse în recunoașterea automată a vorbirii imaginate prin analizarea semnalelor EEG: coeficienții MFCC și LPC.

În urma rezultatelor obținute, am concluzionat că trăsăturile MFCC oferă o înțelegere mai bună a discursului imaginat, oferind rezultate de acuratețe și precizie mai mari decât LPC. De asemenea, s-a putut observa că în urma concatenării celor două tipuri de trăsături, MFCC și LPC, într-o matrice tridimensională rezultatele au fost îmbunătățite, acuratețea atingând o valoare de 0.38. Acest studiu a arătat de asemenea și că trăsăturile utilizate în recunoașterea automată a vorbirii pot diferenția diferitele mecanisme de pronunție atunci când se urmărește clasificarea semnalelor preluate în timpul discursului silențios.

83.1.313.1.3 Capitolul 9: Recunoașterea fonemelor utilizând MFCC și CNN

Studiul prezentat în acest capitol a avut ca scop testarea coeficienților MFCC pentru un sistem automat de recunoaștere a discursului imaginat. Acești coeficienți au fost combinați cu o rețea neuronală CNN capabilă să găsească legături spațiale între coeficienții calculați pentru fiecare canal. Cele mai bune rezultate obținute au fost de 24.19% acuratețe pentru setul de testare.

13.1.4 Capitolul 10: Analizarea și clasificarea discursului imaginat utilizând rețeaua SOM

În acest capitol s-a urmărit analiza a trei tipuri de trăsături: MFCC, covariația încrucișată în domeniul timp și covariația încrucișată în domeniul frecvență utilizând rețeaua neuronală SOM nesupervizată pentru cartografierea spațiului trăsăturilor într-un spațiu bidimensional cu scopul de a obține o transformare a spațiului care să îmbunătățească separabilitatea datelor. S-a observat că trăsăturile calculate în domeniul frecvență au prezentat o așezare mai bună pe hartă din punct de vedere al diferențierii între clase, însă nu au existat zone diferite clare pentru fiecare fonem sau cuvânt din baza de date nici în cazul acestor trăsături. În schimb, în domeniul frecvență s-au putut vedea grupări ale fonemelor și ale cuvintelor în zone diferite ale hărții, făcându-se o diferențiere mai bună între aceste două clase. În ceea ce privește clasificarea, se poate trage aceeași concluzie și anume datele de intrare corespunzătoare covariației încrucișate în domeniul frecvență au oferit o clasificare mai bună decât cele în domeniul timp, ridicând acuratețea la o valoare de 0.25. Aplicând și tehnica PCA pentru eliminarea artefactelor oculare, acuratețea a atins o valoare de 0.28.

83.1.413.1.5 Capitolul 11: Sistem de recunoaștere a cuvintelor și fonemelor din baza de date KODB

Studiul realizat în acest capitol și-a propus analiza semnalelor EEG pentru recunoașterea rostirii imaginate a șapte foneme și patru cuvinte. Pentru a se îndeplini scopul propus, a fost dezvoltat un sistem inteligent comun subiecților din baza de date utilizând un lanț de procesare aplicat pe semnalele din baza de date KODB. În etapa de extragere de trăsături, au fost comparate rezultatele obținute după calcularea covariației încrucișate pe canale în domeniul timp și în domeniul frecvență. A fost realizată și o analiză a diferitelor lungimi ale ferestrelor: 0.25, 0.5 și 1s pentru a găsi fereastra în care semnalul devine cvasi-staționar sau aproape cvasi-staționar, dar care conține de asemenea și informații cu privire la rostire. În final, în etapa de clasificare, au fost testate arhitecturi multiple ale rețelei CNN pentru a vedea care dintre ele oferă performanțele cele mai bune.

Cele mai bune rezultate au fost obținute utilizând covariația încrucișată pe canale în domeniul frecvență utilizând o fereastră de analiză de 0.25s. Cea mai bună performanță a sistemului a fost obținută utilizând o rețea CNN cu două straturi convoluționale 2D având 64 și 128 de filtre și un strat complet conectat având numărul de neuroni egal cu 64. Utilizând aceste caracteristici ale sistemului, s-a obținut o acuratețe de 37%, o îmbunătățire semnificativă în comparație cu utilizarea coeficienților MFCC, unde acuratețea maximă înregistrată a fost de 20.80% utilizând SVM în etapa de clasificare [24] și 24.19% utilizând un clasificator CNN [23].

~~83.1.5 Capitolul 10: Analizarea și clasificarea discursului imaginat utilizând rețeaua SOM~~

~~În acest capitol s-a urmărit analiza a două tipuri de trăsături: covariația încrucișată în domeniul timp și în domeniul frecvență utilizând rețeaua neuronală SOM nesupervizată pentru maparea spațiului trăsăturilor într-un spațiu bidimensional cu scopul de a obține o transformare a spațiului care să îmbunătățească separabilitatea datelor. S-a observat că trăsăturile calculate în domeniul frecvență au prezentat o așezare mai bună pe hartă din punct de vedere al diferențierii între clase, însă nu au existat zone diferite clare pentru fiecare fonem sau cuvânt din baza de date nici în cazul acestor trăsături. În schimb, în domeniul frecvență s-au putut vedea grupări ale fonemelor și ale cuvintelor în zone diferite ale hărții, făcându-se o diferențiere mai bună între aceste două clase.~~

~~În ceea ce privește clasificarea, se poate trage aceeași concluzie și anume datele de intrare corespunzătoare covariației încrucișate în domeniul frecvență au oferit o clasificare mai bună decât cele în domeniul timp, ridicând acuratețea la o valoare de 0.24. Cu toate~~

Formatted: No bullets or numbering

Formatted: Heading 3

Formatted: Heading 3, Indent: First line: 0"

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

~~acestea rezultatele clasificării utilizând rețeaua neuronală SOM nu depășesc rezultatele obținute în capitolul anterior, în care s-a utilizat rețeaua neuronală CNN ca formă de clasificare.~~

83.1.613.1.6 Capitolul 12: Sisteme de recunoaștere a fonemelor și cuvintelor din baza de date KODB utilizând CNNLSTM

Acest capitol și-a propus dezvoltarea unui sistem comun subiecților din baza de date pentru recunoașterea a șapte foneme și patru cuvinte achiziționate în timpul discursului imaginat la Centrul de Reabilitare din Toronto. Capitolul curent a arătat o îmbunătățire a performanței sistemului pentru utilizarea rețelei CNNLSTM în comparație cu rețeaua CNN. Acuratețea a crescut de la 37% la 43% la utilizarea rețelei CNNLSTM fără modificări în lanțul de preprocesare sau extragere de trăsături. Sistemul dezvoltat și-a propus să considere în proiectare și posibilitatea implementării în timp real pe un dispozitiv portabil cu resurse limitate. Astfel că, s-a realizat și un studiu în ceea ce privește reducerea numărului de electrozi din componența sistemului. S-a concluzionat în urma studiului că 93% din informație este concentrată în regiunile anatomice specifice producerii discursului, obținând o acuratețe de 40% pentru utilizarea a 29 de electrozi, comparativ cu 62, acesta fiind numărul inițial de canale.

83.213.2 Contribuții originale

Contribuțiile originale aduse pe durata dezvoltării tezei de doctorat sunt:

- Implementarea unor algoritmi de eliminare a artefactelor oculare cu aplicații în recunoașterea discursului imaginativ, prezentați în detaliu în lucrarea (C2). Au fost testați comparativ două tipuri de algoritmi: filtru adaptiv și PCA. Metoda PCA a oferit rezultate mai bune, decorelând semnalul HEO de canalele frontale afectate (FP1, FP2, FPZ) și îmbunătățind corelațiile pentru semnalele din aceeași clasă.
- Evidențierea prezenței unor markeri descriptivi pentru diferite mecanisme de pronunție la analizarea semnalelor EEG în studiul (C1). Mecanismele de pronunție analizate au fost: fonemul /iy/ (conținând vocalele: /iy/, /piy/ /tiy/ și /diy/), fonemul /uw/ și consoane (/m/ și /n/).
- Implementarea unui sistem de recunoaștere a șapte foneme din baza de date Kara One în lucrarea (C5). Studiul folosește coeficienții MFCC utilizând ecuația de transformare pe scara mel adaptată pentru frecvențele EEG, în etapa de extragere de trăsături. Combinând coeficienții Mel-Cepstrali cu rețeaua neuronală CNN, sistemul a depășit rezultatele prezentate în literatura de specialitate.
- Studiu asupra influenței hiperparametrelor și a diferitelor arhitecturi ale rețelei neuronale CNN în recunoașterea discursului imaginat (J1).
- Studiu asupra ferestrelor de analiză în realizarea unui sistem de recunoaștere a discursului imaginat (J1). S-au analizat ferestre de dimensiuni diferite: 0.25s, 0.5s

Capitolul 13 - Concluzii

și Is. Această analiză a arătat că dimensiunea ferestrei afectează performanța sistemului, având un rol important în găsirea dimensiunii optime pentru rostirea fonemelor și cuvintelor.

- Introducerea trăsăturilor de covariație încrucișată pe canale în domeniul recunoașterii automate a discursului imaginat din semnale EEG (J1). S-a dovedit că această metodă de extragere de trăsături îmbunătățește semnificativ performanțele sistemului.
- Realizarea unui sistem inteligent de recunoaștere a discursului imaginat care poate fi implementat în timp real pe un dispozitiv portabil de cost redus (J1), (J2). acest lucru s-a evidențiat prin calcularea complexității, a memoriei și a timpului de execuție al acestuia.
- Analizarea diferitelor regiuni craniene în dezvoltarea sistemului de recunoaștere a discursului imaginat (J2). Această analiză a arătat că regiunile anatomice implicate în procesul de preparare și execuție a discursului oferă cele mai bune rezultate de performanță a sistemului.

83.313.3 Lista lucrărilor originale

În această secțiune a lucrării sunt prezentate toate lucrările publicate în decursul realizării acestei teze, împreună cu cele care sunt trimise către revizuire.

Articole publicate în jurnale științifice internaționale:

(J1) **A.L. Rusnac**, O. Grigore, “*CNN Architectures and Feature Extraction Methods for EEG Imaginary Speech Recognition*”, *Sensors*, 22(13), p. 4679, 2022, WOS:000822119000001, Articol Q2, Factor de Impact: 3.847, eISSN:1424-8220, DOI: 10.3390/s22134679

(J2) **A.L.Rusnac**, O. Grigore, “*Imaginary speech recognition using a convolutional network with long-short term memory*”, *Applied Science*, 2022, WOS:000887061100001, Articol Q2, Factor de Impact: 2.838, DOI: 10.3390/app122211873

Lucrări în volumele manifestărilor științifice internaționale (conferințe, simpozioane) indexate ISI:

(C1) **A. L. Rusnac**, O. Grigore, “*Convolutional Neural Network applied in EEG imagined phoneme recognition system*”, 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE), MAR 25-27, 2021, București, ROMÂNIA, pp: 1-4, ISSN: 1843-8571, ISBN: 978-1-6654-1878-2, DOI: 10.1109/ATEE52255.2021.9425217, WOS:000676164800094

Recunoașterea vorbirii imaginate prin analiza semnalelor EEG

(C2) **A. L. Rusnac**, O. Grigore, “*EEG Preprocessing Methods for BCI Imagined Speech Signals*”, 9th IEEE International Conference on e-Health and Bioengineering (EHB), NOI 18-19, 2021, Grigore T Popa Univ Med & Pharmacy, ELECTR NETWORK, pp: 1-4, ISSN: 2575-5137, ISBN: 978-1-6654-4000-4, DOI: 10.1109/EHB52898.2021.9657563, WOS:000802227900027

(C3) **A. L. Rusnac**, O. Grigore, “*Development of an Intelligent Seizure Prediction System*”, 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), MAR 28-30, 2019, București, ROMANIA, pp: 1-5, ISSN: 1843-8571, ISBN: 978-1-4799-7514-3, WOS:000475904500102

(C4) **A. L. Rusnac**, O. Grigore, “*Intelligent Seizure Prediction System Based on Spectral Entropy*”, 14th International Symposium on Signals, Circuits and Systems (ISSCS), IUL 11-12, 2019, Iași, ROMANIA, pp: 1-4, ISBN: 978-1-7281-3896-1, WOS:000503459500070

Lucrări în volumele manifestărilor științifice internaționale (conferințe, simpozioane) indexate BDI:

(C5) A. L. Rusnac, O. Grigore, “*Generalized Brain Computer Interface System for EEG Imaginary Speech Recognition*”, 24th International Conference on Circuits, Systems, Communications and Computers (CSCC), IUL 19-22, 2020, Chania, GRECIA, p. 184-188, eISBN: 978-1-7281-6503-5, ISBN: 978-1-7281-6504-2, DOI: 10.1109/CSCC49995.2020.00040

Rapoarte de cercetare științifică:

(R1) **A. L. Rusnac**, Coordonator: O. Grigore, “*Studiu asupra metodelor actuale de recunoaștere a discursului imaginativ din semnale EEG*”

(R2) **A. L. Rusnac**, Coordonator: O. Grigore, “*Analiza utilizării entropiei spectrale a semnalelor EEG în aplicații de recunoaștere a cuvintelor rostite imaginativ*”

Bibliografie

- [1] J. T. Crinion, M. A. Lambon-Ralph, E. A. Warburton, D. Howard, and R. J. S. Wise, "Temporal lobe regions engaged during normal speech comprehension," *Brain*, vol. 126, no. 5, pp. 1193–1201, May 2003, doi: 10.1093/brain/awg104.
- [2] G. Mobus, "Part 4. The Neuroscience of Sapience," in *A THEORY OF SAPIENCE: Using Systems Science to Understand the Nature of Wisdom and the Human Mind*, Millennium Alliance for Humanity and the Biosphere, 1992, pp. 192–258. Accessed: Feb. 08, 2022. [Online]. Available: <https://mahb.stanford.edu/library-item/theory-sapience-using-systems-science-understand-nature-wisdom-human-mind/>
- [3] H. C. Mahendru, "Quick review of human speech production mechanism," *International Journal of Engineering Research and Development*, vol. 9, no. 10, pp. 48–54, 2014.
- [4] E. N. Marieb, S. J. Mitchell, L. A. Smith, and P. Z. Zao, *Human anatomy & physiology laboratory manual*, Eleventh edition (cat version). Boston: Pearson, 2014.
- [5] T. H. Wideman *et al.*, "Brain, Tissue," in *Encyclopedia of Behavioral Medicine*, M. D. Gellman and J. R. Turner, Eds. New York, NY: Springer New York, 2013, pp. 262–263. doi: 10.1007/978-1-4419-1005-9_1105.
- [6] M. Teplan, "FUNDAMENTALS OF EEG MEASUREMENT," 2002.
- [7] H. Rogers, *The sounds of language: an introduction to phonetics*. Harrow, England ; New York: Longman, 2000.
- [8] N. J. Hill and J. R. Wolpaw, "Brain–Computer Interface☆," in *Reference Module in Biomedical Sciences*, Elsevier, 2016, p. B978012801238399322X. doi: 10.1016/B978-0-12-801238-3.99322-X.
- [9] M. E. Saab, "Basic Concepts of Surface Electroencephalography and Signal Processing as Applied to the Practice of Biofeedback," 2009.
- [10] K. Ravi, R. Rajkumar, M. M. Raj, and S. S. Devi, "Imagined Speech Classification using EEG," *Advances in Biomedical Science and Engineering*, vol. 1, pp. 20–32, Dec. 2014.
- [11] D. A. Rojas, O. L. Ramos, and J. E. Saby, "Recognition of Spanish Vowels through Imagined Speech by using Spectral Analysis and SVM," *J. Inf. Hiding Multim. Signal Process.*, vol. 7, pp. 889–897, 2016.
- [12] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features," *J. Neural Eng.*, vol. 15, no. 1, p. 016002, Feb. 2018, doi: 10.1088/1741-2552/aa8235.
- [13] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, Queensland, Australia, Apr. 2015, pp. 992–996. doi: 10.1109/ICASSP.2015.7178118.
- [14] D. Vorontsova *et al.*, "Silent EEG-Speech Recognition Using Convolutional and Recurrent Neural Network with 85% Accuracy of 9 Words Classification," *Sensors*, vol. 21, no. 20, p. 6744, Oct. 2021, doi: 10.3390/s21206744.

- [15] A.-L. Rusnac and O. Grigore, "EEG Preprocessing Methods for BCI Imagined Speech Signals," in *2021 International Conference on e-Health and Bioengineering (EHB)*, Iasi, Romania, Nov. 2021, pp. 1–4. doi: 10.1109/EHB52898.2021.9657563.
- [16] R. Srinivasan and P. L. Nunez, "Electroencephalography," in *Encyclopedia of Human Behavior 2nd Edition*, vol. 2, Academic Press, 2012, pp. 15–23.
- [17] R. Kher and R. Gandhi, "Adaptive filtering based artifact removal from electroencephalogram (EEG) signals," in *2016 International Conference on Communication and Signal Processing (ICCSP)*, Melmaruvathur, Tamilnadu, India, Apr. 2016, pp. 0561–0564. doi: 10.1109/ICCSP.2016.7754202.
- [18] L. Tan and J. Jiang, "Chapter 9: Adaptive filters and applications," in *Digital signal processing: Fundamentals and applications*, New Mexico: Academic Press, 2018, pp. 421–462.
- [19] A.-L. Rusnac and O. Grigore, "Convolutional Neural Network applied in EEG imagined phoneme recognition system," in *2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Bucharest, Romania, Mar. 2021, pp. 1–4. doi: 10.1109/ATEE52255.2021.9425217.
- [20] M. M. Azmy Gad, "Classification of mental tasks using support vector machine based on linear predictive coding and new mother wavelet transform," in *2015 International Conference on Biomedical Engineering and Computational Technologies (SIBIRCON)*, Novosibirsk, Russia, Oct. 2015, pp. 156–159. doi: 10.1109/SIBIRCON.2015.7361873.
- [21] D. O'Shaughnessy, "Linear predictive coding," *IEEE Potentials*, vol. 7, no. 1, pp. 29–32, Feb. 1988, doi: 10.1109/45.1890.
- [22] Y. Padmasai, K. SubbaRao, V. Malini, and C. R. Rao, "Linear Prediction Modelling for the Analysis of the Epileptic EEG," in *2010 International Conference on Advances in Computer Engineering*, Bangalore, Karnataka, India, Jun. 2010, pp. 6–9. doi: 10.1109/ACE.2010.20.
- [23] A.-L. Rusnac and O. Grigore, "Generalized Brain Computer Interface System for EEG Imaginary Speech Recognition," in *2020 24th International Conference on Circuits, Systems, Communications and Computers (CSCC)*, Chania, Greece, Jul. 2020, pp. 184–188. doi: 10.1109/CSCC49995.2020.00040.
- [24] C. Cooney, R. Folli, and D. Coyle, "Mel Frequency Cepstral Coefficients Enhance Imagined Speech Decoding Accuracy from EEG," in *2018 29th Irish Signals and Systems Conference (ISSC)*, Belfast, Jun. 2018, pp. 1–7. doi: 10.1109/ISSC.2018.8585291.
- [25] A.-L. Rusnac and O. Grigore, "CNN Architectures and Feature Extraction Methods for EEG Imaginary Speech Recognition," *Sensors*, vol. 22, no. 13, p. 4679, Jun. 2022, doi: 10.3390/s22134679.
- [26] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview." arXiv, Aug. 13, 2020. Accessed: May 24, 2022. [Online]. Available: <http://arxiv.org/abs/2008.05756>
- [27] A.-L. Rusnac and O. Grigore, "Imaginary speech recognition using a convolutional network with long-short term memory," *Applied Sciences*, 2022.