Deep learning alapú agyi jel feldolgozás és beszédszintézis előkészítő munkálatai

Arthur Frigyes Viktor, Csapó Tamás Gábor

Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék

{arthur,csapot}@tmit.bme.hu

Kivonat Ebben a cikkben bemutatjuk a nem invazív, elektorenkefalográfián (EEG) alapuló kommunikációs agy-gép interfész (BCI) rendszerekkel kapcsolatos kutatásunk kezdeti eredményeit. Az agyi jelből akusztikus jelbe átalakító konverziós módszer célja, hogy az idegsejtek elektromos aktivitása alapján szintetizáljunk beszédet. A kutatás során a "Kara One" adatbázis beszélőivel végeztünk kísérleteket, melyből párhuzamos EEG és beszédfelvételeket használtunk fel. A cikkben részletesen bemutatjuk az agyi jel előfeldolgozásának lépéseit (szűrés, csatorna választás, zajok és műtermékek (artifact) eltávolítása, független komponens analízis, és szegmentálás). Az előfeldolgozott EEG jelből mély neuronhálóval becsüljük meg a beszéd mel-spektrális paramétereit. Végül egy neurális vokóderrel állítjuk elő a szintetizált beszédet. Az így szintetizált beszéd ugyan nem érthető, de emlékeztet az eredeti jelre, így a kezdeti eredményeket biztatónak tartjuk. A kutatás hosszú távú jelentősége, hogy a beszéd valós idejű szintézise közvetlenül a mért idegi aktivitásból lehetővé tenné a természetes beszédet, és jelentősen javítaná az életminőségét, különösen a kommunikációban súlyosan korlátozott személyek számára.

Kulcsszavak: beszédszintézis, agy-gép interfész (BCI), elektorenkefalográfia (EEG), némabeszéd-interfész (SSI)

1. Bevezetés

A beszéd az emberi kommunikáció elsődleges és legfontosabb eszköze. Sokan azonban elvesztették ezt a képességüket betegség vagy egészségkárosodás okán. Nincs sok olyan tanulmány, amely konkrét adatokat szolgáltatna ennek a fogyatékosságnak a gyakoriságáról. Dupré és Karjalainen (2003) arra a következtetésre jutnak, hogy az európai lakosság 0,4%-a szenved beszédkárosodásban. Egy későbbi, 2011-es felmérésben (Eurostat, 2011) arról számolnak be, hogy Európában az emberek 0,5%-a küzd kommunikációs nehézségekkel. Az általunk alkalmazott megközelítésben az agyi jelből akusztikus jelbe átalakító konverziós módszer célja, hogy az idegsejtek elektromos aktivitása alapján szintetizáljunk beszédet.

1.1. Agyi jelből akusztikus jelbe átalakítás elektroenkefalogram alapján

A kommunikációs agy-gép interfész (BCI) célja, hogy természetes vagy ahhoz közeli kommunikációs csatornát biztosítsanak olyan személyek számára, akik fizikai vagy neurológiai károsodás miatt nem tudnak beszélni. A beszéd valós idejű szintézise közvetlenül a mért idegi aktivitásból lehetővé tenné a természetes beszédet, és jelentősen javítaná az életminőségét, különösen a kommunikációban súlyosan korlátozott személyek számára. A jelen cikkben leírt megközelítésünkben non-invazív módon mintavételezett elektroenkefalográf (EEG) segítségével regisztrációra kerül a beszélő agyi elektromos aktivitása, mely jelből közvetlenül, a megfelelő átalakítási lépések után beszédet szintetizálunk. A konverziónak több lehetséges megoldása létezik.

Angrick és mtsai (2021) minimál invazív (sEEG) módon mért idegi aktivitásból dekódolási megközelítéssel a beszéd jó minőségű rekonstrukcióját tudták megvalósítani a közelmúltban. Sztereotaktikus mélységi elektródákkal beültetett résztvevővel megbízhatóan tudtak érthető beszédet generálni valós időben. Egy súlyos epilepsziás beteg (20 éves nő) orvosi kezelése során 11 sEEG-elektródaszárat ültettek be a bal féltekébe, 8–18 érintkezővel. Az elektródákat az epilepsziás fókuszok meghatározása és az agykérgi funkciók feltérképezése céljából ültették be, hogy azonosítsák azokat a kritikus területeket, amelyek reszekciója hosszú távú funkcionális hiányosságokat eredményezhet (Angrick és mtsai, 2021). A bemutatott eredmény impozáns, a beszéd-neuroprotézis széleskörű alkalmazhatóságának azonban korlátot szabhat, hogy ez a megoldás invazív műtéti beavatkozáshoz kötött.

Egy másik megközelítésben Herff és mtsai (2015) intrakraniális elektrokortikográfiás (ECoG) felvételekből állítottak elő szöveget (agyi jel – szöveg konverzió). Az eredmények szerint a szóhibaarány 50%, míg a fonéma hibaarány 25%-os. A szövegből egy TTS és neurális vokóder (Tacotron, Glow-TTS, SpeedySpeech, GAN-TTS) segítségével beszéd konvertálható. A viszonylag magas hibaarányok ugyan a valós életbeli felhasználást még nem teszik lehetővé, de mindenképp előremutatóak.

1.2. A jelen kutatás célja

A kutatás célja egy megvalósíthatósági tanulmány: azt teszteltük, hogy közvetlen agyi jelből, az idegsejtek elektromos aktivitásának regisztrátumából, érthető akusztikum közvetlenül előállítható-e mély tanulással.

2. Kísérlet

Először bemutatjuk röviden a mögöttes adatbázist, illetve áttekintjük a kísérlet alapjait.

2.1. Az adatbázis

Kutatásunkban a meglévő "Kara One Dataset" elnevezésű adatbázis került alkalmazásra, mely 12 résztvevő multimodális fiziológiai jeleiről készült felvételeit tartalmazza kiejtett beszéd, elképzelt beszéd és beszéd hallgatás közben (Zhao és Rudzicz, 2015). Minden résztvevő jobb kezes, 10 résztvevő angol anyanyelvű, míg 2 felsőfokon beszéli az angol nyelvet. Minden résztvevő legalább középfokú végzettséggel rendelkezik, valamint nem ismert, hogy lenne bármilyen neurológiai betegségük vagy álltak volna korábban kábítószer használat alatt. A kísérletben nyolc férfi valamint négy női beszélő vett részt, átlagéletkoruk 27,4 év ($\sigma = 5$) (Zhao és Rudzicz, 2015).

Az adatrögzítés helyszíne a Toronto Rehabilitációs Intézet egyik irodája volt. A videó és a beszéd rögzítésére a Microsoft Kinect kamera rendszert illetve szoftvert alkalmazták, míg az EEG jel rögzítésére 64 csatornás NeuroScan "Quickcap" sapkát használtak 10–20 csatorna kiosztási módban (Sharbrough és mtsai, 1991). Négy további elektródát helyeztek el az alanyokon: a bal szem alatt és felett, valamint egyet-egyet a jobb és bal szem mellett laterálisan, hogy később a horizontális és vertikális szemmozgás által okozott műtermékek (artifact) könnyen szűrhetőek legyenek az EEG jelből. A mintavételezési frekvencia 1000 Hz volt.

A kísérletet három elkülönített fázisra bontották, melyben az alanyokat arra kérték, hogy 7 fonémával és 4 szótaggal különböző módon járjanak el. Az első fázisban az alanyokat arra kérték, hogy hogy hallgassák azokat, a másodikban gondoljanak rá, míg a harmadikban, hogy hangosan mondják ki őket. A jelen kutatásban a fenti fázisokból csak azokat a részeket használtuk fel, amikor a kísérleti alany kimondta a fonémát vagy szótagot.

Fonémák	/iy/	Szótagok	pat
	/uw/		pot
	/piy/		knew
	/tiy/		gnaw
	/diy/		
	/m/		
	/n/		

1. táblázat. Rögzített promptok

A 4 szótag a Kent-féle, fonetikailag hasonló párok listájából származott (Kent és mtsai, 1989). A promptok (stimulusok) kiválasztásakor figyelembe vették, hogy viszonylag egyenlő számú nazális, zárhang és magánhangzó, valamint zöngés és zöngétlen fonéma legyen jelen.

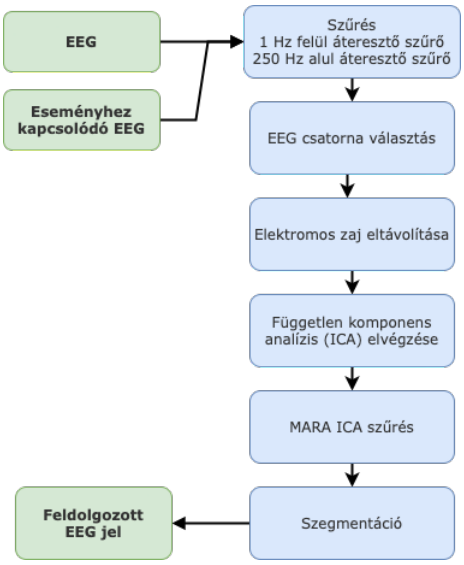
2.2. Az adatrögzítés módja

A résztvevők a következő feladatot hajtották végre.

1. Nyugalmi állapot: (5 másodperc) A résztvevőknek azt az utasítást adták, hogy relaxáljanak és tisztítsák ki az elméjüket.
2. Stimulációs állapot: A képernyőn megjelent a fonémát vagy szótagot megjelenítő felirat, és a számítógép hangszóróin keresztül is szólt. Ezután a résztvevőket arra utasították, hogy vegyék fel az adott fonéma vagy szótag kiejtéséhez szükséges kiindulási pozíciót.
3. Elképzelt állapot: (5 másodperc) A résztvevők elképzelték, hogy kimondják az adott fonémát vagy szótagot, mindezt mozdulatlanul.
4. Beszélő állapot: A résztvevők hangosan kimondták a fonémát vagy szótagot. A Kinect kamera rendszer ebben a szakaszban mind a hangot, mind az arcmozgást rögzítette.

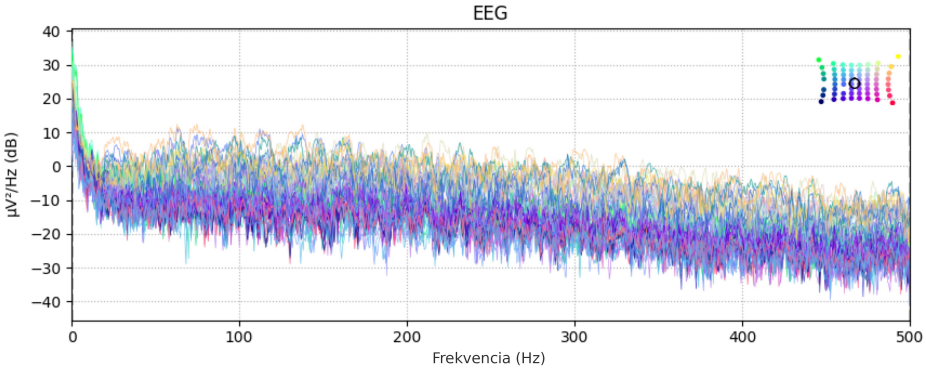
2.3. Az agyi jel előfeldolgozása

Saját kísérleteinkben az agyi jelek előfeldolgozása lazán követi a HAPPE (Harvard Automated Processing Pipeline for Electroencephalography) (Gabard-Durnam és mtsai, 2018) munkafolyamatát, lényegében a folytonos wavelet transzformációs szakaszig, mely helyett közvetlen szegmentációs szakasz következik (ld. 1. ábra).



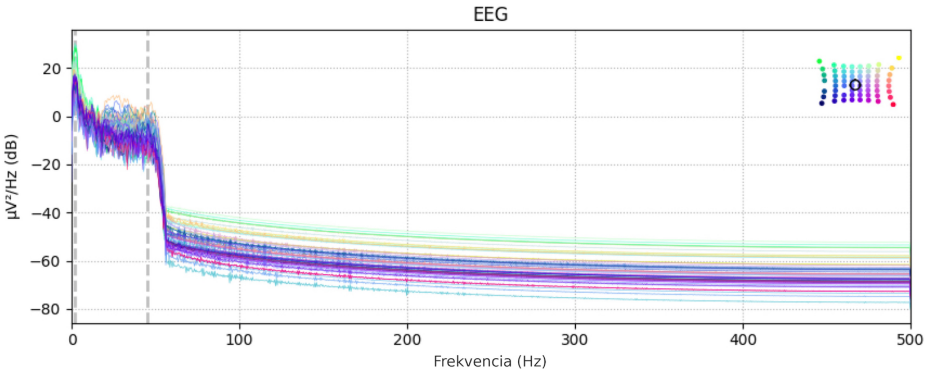
1. ábra: EEG jelfeldolgozási folyamat

Szűrés Az elektroenkefalogram feldolgozásának első lépése a szűrés. Egy 1 Hz-es felül-áteresztő szűrő eltávolítja nem stacionárius jelsodródást a felvételen. Ezt követően egy 250 Hz-es alul-áteresztő szűrőt alkalmaztunk.



2. ábra: A szűrést megelőző elektroenkefalogram teljesítménysűrűség spektruma - [MM05 beszélő]

Tápvezetéki zaj eltávolítása Az elektromos hálózat által keltett zaj a tápvezetéki zaj. Éles csúcsokból áll 50 Hz-en vagy 60 Hz-en, a földrajzi helytől függően. Tekintve, hogy a Kara One adatbázist Kanadában rögzítették, ahol az elektromos hálózat harmonikus frekvenciája 60 Hz, így ennek a frekvenciának a szűrését végeztük el. Ez egy egyszerű alul-áterező segítségével történt, melynek használata egyéb okból is indokolt, hiszen, elsősorban béta (12–40 Hz) agyhullámokkal dolgozunk. Az eredményre a 3. ábra mutat egy példát.

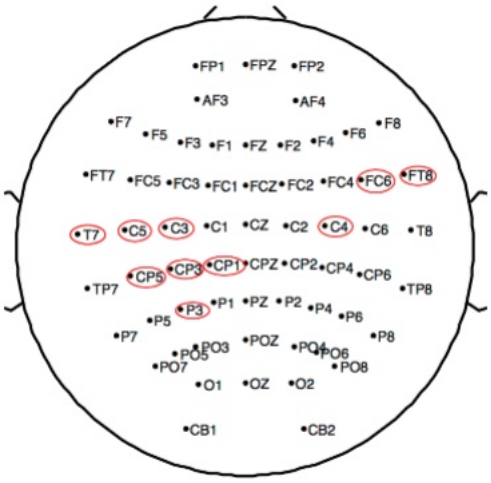


3. ábra: Tápvezetéki zaj eltávolítása után az elektroenkefalogram teljesítménysűrűség-spektruma [MM05 beszélő]

EEG csatorna választás A csatorna választás során első lépésben azokat a csatornákat szűrjük ki, melyek esetleg egyértelmű elektróda hibával rendelkeznek. Majd azokat, melyek regisztrátumai nem korrelálnak a megfigyelt agyi tevékeny-

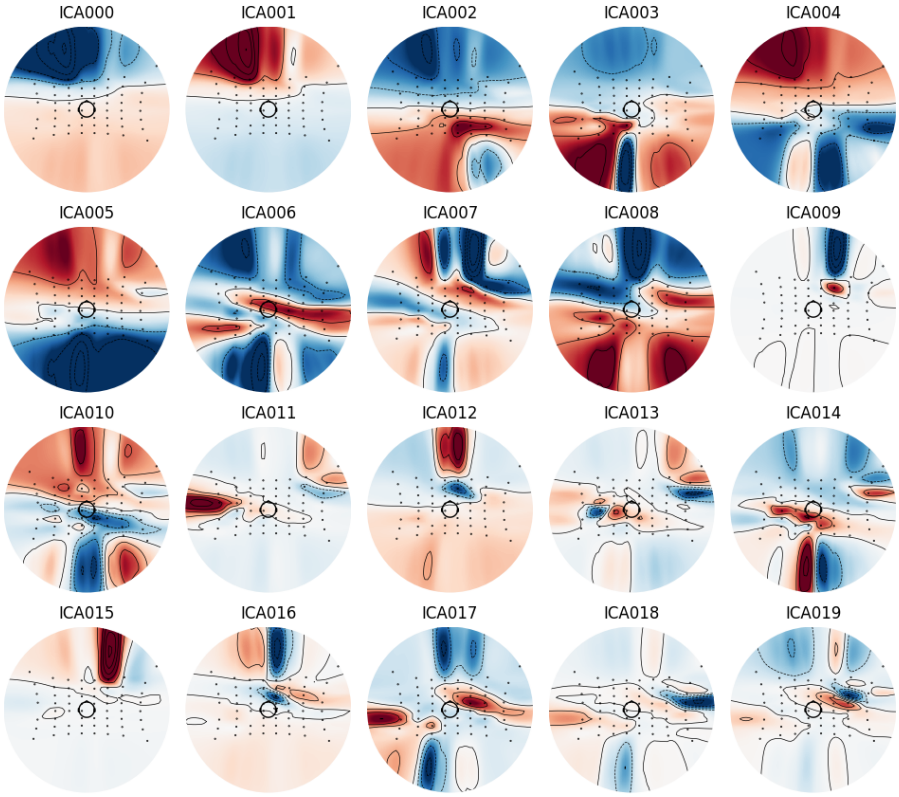
séggel. Feltételezhető beszéd közben, hogy a kísérletben résztvevő alanyok első és másodlagos motoros agykérgei, valamint Broca és Wernicke agyi területei felett mérhető akkumulált elektromos aktivitás releváns a beszéd szempontjából, míg például a vizuális kéreg, szomatoszenzoros kéreg felett mérhető adatok kevésbé relevánsak. Ez könnyen ellenőrizhető lineáris korreláció vizsgálattal ahol a 64 EEG csatornák jeleit mint normál eloszlású változókat Pearson-féle korreláció (Benesty és mtsai, 2009) vizsgálattal megvizsgáljuk. A legalább $avg(r) \geq 0,36$ korrelációval rendelkező EEG csatornákat a 4. ábrán mutatjuk (a többi csatorna lényegesen kisebb korrelációt mutat).

Ez az információ a későbbi független komponens analízis elvégzésekor segítségünkre volt, hiszen azok a komponensek, melyek a beszéd szempontjából irrelevánsnak bizonyulnak, a releváns agyi területeket könnyebben azonosíthatóvá tette.



4. ábra: A korreláló csatornákat megjelenítő skalp topográfiai térkép (12 óra pozícióban az orr, 3 óra pozícióban a jobb míg 9 óra pozícióban a bal fül található)

Független komponens analízis Független komponens analízis (independent component analysis, ICA) segítségével (Hyvärinen és Oja, 2000) az elektroenkefalogramot először lineárisan független komponenseikre bontjuk. Majd azonosítjuk a pislogás, szemmozgás valamint izomaktivitás eredetű zaj-komponenseket, azok teljesítményűrűség-spektruma, vizuális megjelenése és az agykérgen való térbeli eloszlása alapján (Gabard-Durnam és mtsai, 2018). Ezen komponensek eltávolítása után a megmaradt idegi aktivitáshoz kötődő komponenseket inverz ICA-transzformációval alakítottuk vissza az eredeti, 64-csatornás kiterjesztésre (Rácz, 2019). Az eredményül kapott 20 független komponenst az 5. ábra mutatja.



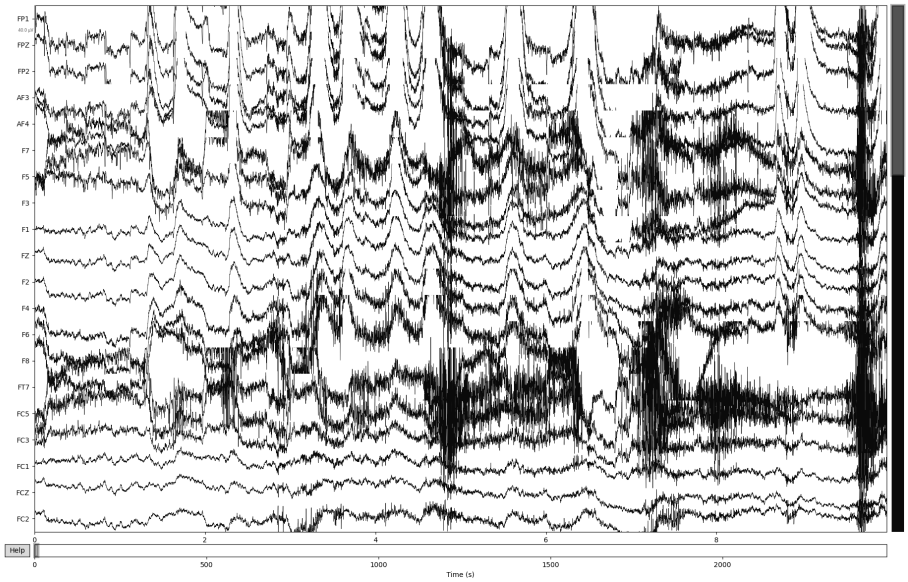
5. ábra: 20 független komponenst megjelenítő skalp topográfiai térkép (12 óra pozícióban az orr, 3 óra pozícióban a jobb míg 9 óra pozícióban a bal fül található) - [MM05 beszélő]

MARA ICA szűrés A Multiple Artifact Rejection Algorithm (MARA) egy gépi tanulási algoritmus, amely kiértékeli az ICA-ból származó összetevőket (Winkler és mtsai, 2014). Bár más algoritmusok is léteznek a műtermék bizonyos kategóriáinak automatikus észlelésére (pl. szemmozgási műtermékek és jelek folytonossági zavarai) (Mognon és mtsai, 2011), a MARA-t manuálisan kialakított komponens kategóriák felismerésére tanították, így viszonylag alacsony domain tudás nélkül is automatikusan képes a műtermékek széles skáláját szűrni.

Szegmentáció A szegmentációs lépés lényegében annotálja és epochokhoz sorolja a folytonos elektroenkefalogramot, így kialakítva fragmentumokat. Beszélőnként 131 illetve 132 próbát különböztetünk meg. A szegmentáció HAPPE megközelítésben csak egy opcionális lépés, viszont számunkra elengedhetetlen volt, tekintve, hogy minden egyes epoch egy fonémához vagy szótaghoz tartozó

elektroenkefalogram szakaszt különböztet meg, melyből becsülni szándékoztuk az akusztikumot, pontosabban annak mel-spektrális paraméterit.

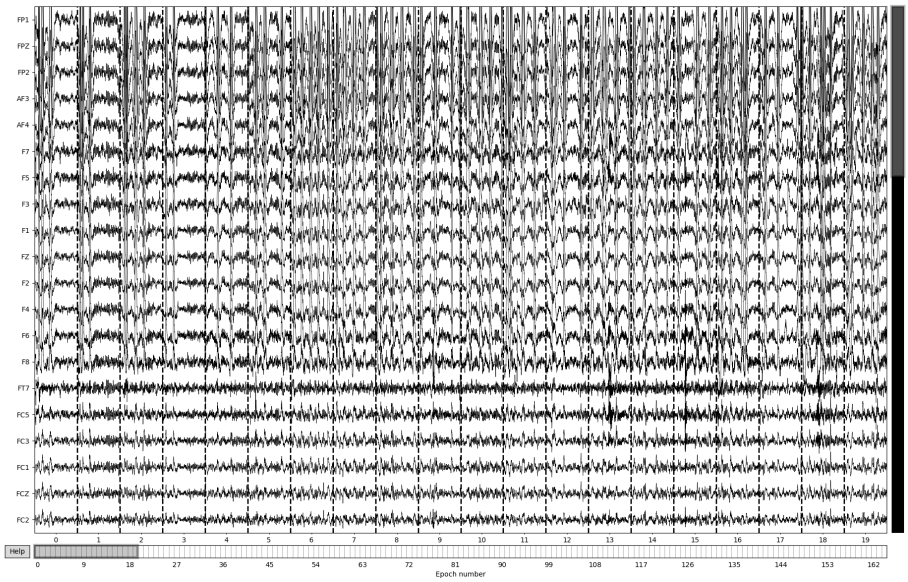
Az 6. ábrán látható egy példa nyers elektroenkefalogramra, majd a feldolgozási lépések elvégzése után a 7. ábrán látható feldolgozott elektroenkefalogramot kapjuk az MM05 azonosítójú beszélőn.



6. ábra: Nyers elektroenkefalogram [MM05 beszélő]

Neurális vokóder A WaveNet 2016-os bevezetése óta, melyet van den Oord és mtsai (2016) mutattak be, a neurális vokóderek egy izgalmas új módja a beszéd nyers mintáinak generálására a szöveg-beszéd szintézis (Text-To-Speech, TTS) során. A korai WaveNet-szerű modellek egyik hátránya, hogy azok rendkívül számításigényesek voltak. Jelenleg a legmodernebb TTS modellek parametrikus neurális hálózatokon alapulnak, amelyek WaveNet-szerű neurális vokóderek továbbfejlesztett változatait használják. A TTS-szintézis jellemzően két lépésben történik: az első lépésben a szövegben lévő karakterekhez rendelünk hangitőtartamokat majd átalakítjuk a beszédet leíró jellemzőkké, például mel-spektrogrammá; a második lépés ezeket a spektrális jellemzőket alakítja át a beszéddé. Azaz a második lépés során, a becsült spektrális paramétereket felhasználva beszédet generálhatunk.

A neurális vokóderek egyik legújabb típusa, a WaveGlow (Prenger és mtsai, 2018) egy áramlásalapú hálózat (flow-based network), amely képes mel-spektrogramokból jó minőségű beszédet generálni. A WaveGlow modell előnye,

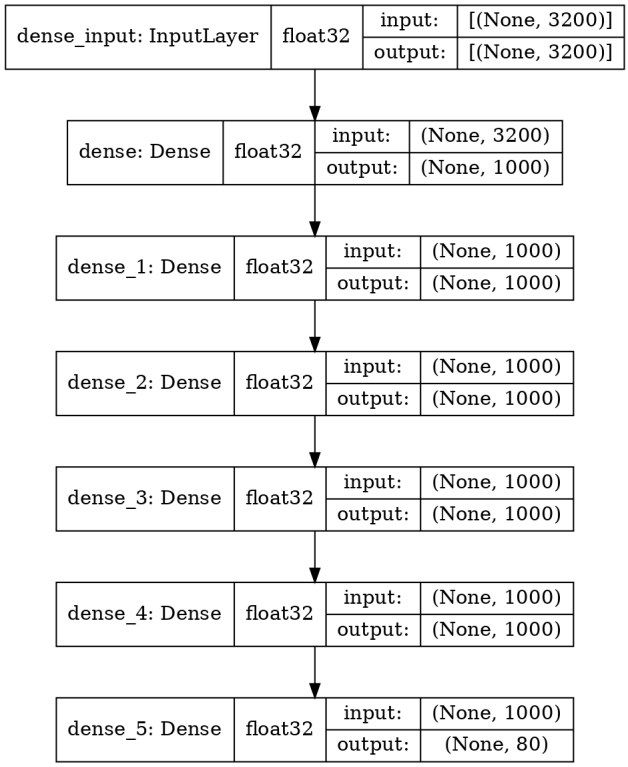


7. ábra: Feldolgozott elektroenkefalogram [MM05 beszélő]

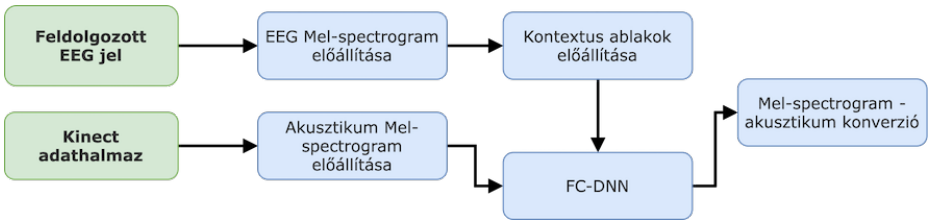
hogy viszonylag egyszerű, a szintézis valós időben elvégezhető, mégis jó természetességet eredményez. Képes beszédet generálni mintavételezéssel egy mel-spektrogrammal kondicionált eloszlásból.

2.4. Neurális hálózat

Teljesen kapcsolt mély neurális hálózati architektúrát (Fully Connected Deep Neural Network, FC-DNN) használtunk, mely 5 rejtett rétegből áll, amelyek mindegyike 1000 neuront tartalmaz. ReLU aktivációs függvényt alkalmaztunk. A bemeneti réteg 3200 neuronból áll (a bemeneti dimenzió megegyezik az EEG jelből képzett mel-spektogram, valamint a visszatekintési ablak tizedének a szorzatával). Visszatekintési ablakot azért alkalmaztunk, hogy a hálózat tudja kezelni az EEG és beszédjel közötti időben távolabbi összefüggéseket is. Az EEG jel feldolgozása során 400 ms hosszú ablakot alkalmaztunk. A beszéd mintavételezése 22 kHz-en történt, amiből 80-dimenziós mel-spektrogramot számítottunk. A kimeneti réteg 80, a mel-spektrális együtthatók darabszámával megegyező neuront tartalmaz. Az FC-DNN hálózat felépítését a 8. ábra mutatja.



8. ábra: FC-DNN hálózati architektúra



9. ábra: A beszédszintézis folyamata

2.5. A beszédszintézis folyamata

Az EEG jel előfeldolgozását követően előállítjuk annak mel-spektrogramját melyet a 'librosa' könyvtár segítségével teszünk meg. Ezzel párhuzamosan feldolgozzuk a 132 Microsoft Kinect rendszer által rögzített audio állományt. Az EEG

hullámhoz hasonlóan ennek is előállítjuk a mel-spektrogramját. Mielőtt a FC-DNN hálózat tanítását megkezdenénk az EEG jelet ablakozzuk, vagyis egy 400 ms hosszú hullámszakasszal bővítjük az előzőleg már szegmentált jelet. Erre azért van szükség, mert megfigyeléseink szerint az agykéregben mért idegi aktivitás és a rögzített beszéd között késleltetés is megjelenhet. Ez következhet az észlelés és a cselekvés közötti effektusfüggő motoros adaptációjából. Ez az érték beszélőnként eltérhet, de a kutatás ezen szakaszában ezzel az általános értékkel számoltunk. A tanítóhalmaz 72%-át teszi ki a teljes adathalmaznak, a validációs halmaz 8%-át, míg a teszt halmaz 20%-át. Ez beszélőnként 95, 11 illetve 26 regisztrátumot jelent. Következő lépésben betanítjuk a neurális hálózatot, mely bemenetként ablakozott EEG mel-spektogram fragmentumokat kap, kimenetként pedig 80 mel-spektrális együtthatót kapunk, melyből a WaveGlow modellt használva inverz STFT transzformációval megkapjuk a becsült beszédet.

3. Eredmények

Objektív mérések Annak ellenőrzése céljából, hogy a tervezett modell reprodukálni tudja-e az eredeti beszéd jellemzőit, kiértékeltük a spektrális torzulásokat a természetes beszéd és a szintetizált beszéd között. Erre a "Mel-Cepstral Distortion" (MCD) mérőszámot alkalmaztuk:

$$MCD = \frac{10\sqrt{2}}{\ln 10} \frac{1}{T} \sum_{t=1}^{T} \sqrt{\sum_{i} \left(C_{ti} - \hat{C}_{ti} \right)^2}, \tag{1}$$

ahol C_{ti} és $\hat{C}_{ti}$ a mel-kepsztrum az eredeti valamint a szintetizált beszéd esetén.

A 2. táblázat a validációs adatokon számolt MCD értékeket valamint az eredeti és becsült mel-spektrális együtthatókon mért átlagos négyzetes hibát (MSE) jeleníti meg beszélőnként. A beszélők azonosítására a Kara One adatbázisban is használt azonosítók kerültek alkalmazásra. Látható, hogy az eredmények erősen függenek a beszélőtől – pl. MM12 férfi beszélővel és MM19 női beszélővel lényegesen jobb MCD eredményeket sikerült elérnünk, mint a többiekkel.

Szubjektív meghallgatásos teszt elvégzését ebben a korai szakaszban még nem eszközöltünk. A teszthalmazon becsült mel-spektrális paraméterekből visszaalakított beszéd ugyan emlékeztet az eredeti akusztikumra, de zajos és nehezen érthető.

Az MCD értékek abban a tartományban vannak (3–5 dB), ahol a beszédszintetizátor rendszerek általában már jól érthető beszédet generálnak, de meghallgatva a mintákat itt ez nem teljesül. Valószínűleg a felvételekben lévő sok szünet szakasz eredményezte a viszonylag alacsony hiba értékeket, ami viszont itt nem korrelál a szubjektív eredménnyel. A jelen kutatásban nem vizsgáltuk, hogy a szintetizálni kívánt beszédhangok sorozatától mennyire függött az MCD értéke.

2. táblázat. A különböző beszélőkkel tanított neuronhálókkal elért átlagos hibák.

Beszélő	MCD (dB)	MSE
MM05	5,54612	1,25103
MM08	5,73413	1,27839
MM09	5,69279	1,07313
MM10	7,49676	1,23934
MM12	3,83136	1,00725
MM14	4,22095	1,29970
MM18	4,61194	1,29009
MM21	3,75126	1,26096
Átlag (férfi beszélők)	**5,11066**	**1,21248**
MM11	4,31737	1,59456
MM15	5,09933	1,20212
MM16	4,63597	1,36154
MM19	3,80825	1,53233
MM20	4,62845	1,48949
Átlag (női beszélők)	**4,49787**	**1,43600**
Átlag	**4,87497**	**1,29845**

4. Összefoglalás

Ebben a cikkben részletesen bemutattuk az agyi jel előfeldolgozásának lépéseit (szűrés, csatorna választás, zajok és műtermékek eltávolítása, független komponens analízis, és szegmentálás). Egy kezdetleges FC-DNN hálózat tanítását és predikcióját végeztük el az agyi jelből beszédbe szintetizálás feladat során. A kezdeti értékelésünk azt mutatja, hogy az eljárás megvalósítható, ugyanakkor az eredmény még nem kielégítő.

Az eljárást a továbbiakban a neurális hálózati architektúra finomításával szándékozzuk javítani. Megfigyelhető továbbá, hogy a tanító halmaz méretét növelve javítható az eredmény még a jelenlegi architektúra mellett is, így a jövőben szándékozunk saját adatbázist rögzíteni, mely fonémák számában, mind pedig elemszámban bővebb.

Köszönetnyilvánítás

A kutatást részben a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal OTKA programja támogatta (FK 124584 és PD 127915 projektek). Csapó Tamás Gábor kutatásait az MTA Bolyai János kutatói ösztöndíja, valamint az Új Nemzeti Kiválóság Program Bolyai+ (ÚNKP-21-5-BME-2060) pályázata támogatta.

Hivatkozások

Angrick, M., Ottenhoff, M.C., Diener, L., Ivucic, D., Ivucic, G., Goulis, S., Saal, J., Colon, A.J., Wagner, L., Krusienski, D.J., Kubben, P.L., Schultz, T., Herff, C.: Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. Communications Biology 4(1), 1055 (2021), https://doi.org/10.1038/s42003-021-02578-0

Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson correlation coefficient. In: Noise reduction in speech processing, pp. 37–40. Springer (2009)

Dupré, D., Karjalainen, A.: Employment of disabled people in europe in 2002. Statistics in focus pp. 3–26 (2003)

Eurostat: Eurostat - data explorer - population by type of disability, sex, age and labour status (2011), https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlth_dlm040&lang=en

Gabard-Durnam, L.J., Mendez Leal, A.S., Wilkinson, C.L., Levin, A.R.: The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. Frontiers in Neuroscience 12, 97 (2018), https://www.frontiersin.org/article/10.3389/fnins.2018.00097

Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G., Schultz, T.: Brain-to-text: decoding spoken phrases from phone representations in the brain. Frontiers in Neuroscience 9, 217 (2015), https://www.frontiersin.org/article/10.3389/fnins.2015.00217

Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Networks 13(4), 411–430 (2000), https://www.sciencedirect.com/science/article/pii/S0893608000000265

Kent, R.D., Weismer, G., Kent, J.F., Rosenbek, J.C.: Toward phonetic intelligibility testing in dysarthria. Journal of Speech and Hearing Disorders 54(4), 482–499 (1989), https://pubs.asha.org/doi/abs/10.1044/jshd.5404.482

Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M.: ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 48(2), 229–240 (2011), https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8986.2010.01061.x

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A.W., Kavukcuoglu, K.: WaveNet: A generative model for raw audio. CoRR abs/1609.03499 (2016), http://arxiv.org/abs/1609.03499

Prenger, R., Valle, R., Catanzaro, B.: WaveGlow: A flow-based generative network for speech synthesis. CoRR abs/1811.00002 (2018), http://arxiv.org/abs/1811.00002

Rácz, F.S.: A nyugalmi agyi konnektivitás multifraktális dinamikája (2019), http://repo.lib.semmelweis.hu//handle/123456789/8389

Sharbrough, F., Chatrian, G., Lesser, R., Luders, H., Nuwer, M., Picton, T.: American electroencephalographic society guidelines for standard electrode position nomenclature. Clinical Neurophysiology 8, 200–202 (01 1991)

Winkler, I., Brandl, S., Horn, F., Waldburger, E., Allefeld, C., Tangermann, M.: Robust artifactual independent component classification for BCI practitioners. Journal of Neural Engineering 11(3), 035013 (2014)

Zhao, S., Rudzicz, F.: Classifying phonological categories in imagined and articulated speech. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 992–996 (2015)