A beszéd artikulációs mozgásának predikciója agyi jel alapján – kezdeti eredmények

Csapó Tamás Gábor[1], Arthur Frigyes Viktor[1], Nagy Péter[2,3], Boncz Ádám[3]

[1]Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
[2]Budapesti Műszaki és Gazdaságtudományi Egyetem,
Méréstechnika és Információs Rendszerek Tanszék
[3]Természettudományi Kutatóközpont, Kognitív Idegtudományi és
Pszichológiai Intézet, Hang- és Beszédészlelési Kutatócsoport, Budapest
{csapot,arthur}@tmit.bme.hu,
nagy.peter.ssprg@ttk.hu, adam.boncz@gmail.com

Kivonat Az augmentatív és alternatív kommunikációs technológiák (pl. agy-gép interfész, BCI) közvetlenül olvashatják az agyi jeleket, hogy pótolják az elvesztett beszédképességet. Nemzetközi szinten végeztek már kezdeti kutatásokat agyi jel (pl. EEG, sEEG, ECoG) és beszéd alapú BCI kidolgozására, azonban hiányoznak azok a kombinált módszerek, amelyek a nem invazív EEG-t, az artikulációt és a beszédjeleket összevontan vizsgálnák, és elemeznék az agyban zajló tervezési folyamat, az artikulációs mozgás, és a keletkezett beszédjel kölcsönhatását. A jelen kutatásban ismertetett multimodális (EEG, nyelvultrahang és beszéd) analízis és szintézis révén túlmutatunk a legkorszerűbb nemzetközi trendeken. A beszéd közbeni agyi jelek elemzését nyelvultrahang-alapú artikulációs adatokkal bővítjük, hogy több összehasonlítható bioszignál álljon rendelkezésre. Az EEG-vel mért agyi jelből mély neuronhálóval az artikulációs mozgásra vonatkozó információt (nyelvultrahang képek) prediktálunk. Az eredmények szerint az EEG és nyelvultrahang közötti kapcsolat kimutatható. A jelen kutatás hosszútávú célja, hogy hozzájáruljunk a beszéd alapú agy-számítógép interfészekhez: az eredmények potenciálisan alkalmazhatók lehetnek pl. beszédsérülteknek szánt kommunikációs segédeszközként.
Kulcsszavak: beszédtechnológia, ultrahang, EEG, deep learning

1. Bevezetés

Az agy-számítógép interfészek (Brain-Computer Interface, BCI) lehetővé tehetik a számítógépek közvetlen, fizikai aktivitás nélküli vezérlését. Az augmentatív és alternatív kommunikációs (Augmentative and Alternative Communication, AAC) technológiák (pl. BCI) közvetlenül olvashatják az agyi jeleket, hogy pótolják az elvesztett beszédképességet (Chang és Anumanchipalli, 2020). A jövőben a beszédneuroprotézisek alkalmazása segíthet a neurológiai vagy beszédhibás betegeken.

Az agyi jel rögzítésére többféle technológia is rendelkezésre áll: például elektroenkefalográfia (EEG, McFarland és Wolpaw, 2017), sztereotaktikus mély elektródák (sEEG, Verwoert és mtsai, 2022), intrakraniális elektrokortikográfia (ECoG, Buzsáki és mtsai, 2012), magnetoenkefalográfia (MEG, Dash és mtsai, 2021), lokális mezőpotenciál (LFP, Buzsáki és mtsai, 2012). Az idegrendszeri jel rögzítési módok közül a BCI számára az EEG lehet a legmegfelelőbb, mivel elérhető árú, lényegesen kisebb kockázattal jár, mint az invazív módszerek, és hordozható is lehet (Casson, 2019). Nemzetközi szinten végeztek már kezdeti kutatásokat EEG és beszéd alapú BCI kidolgozására (Krishna és mtsai, 2020; Verwoert és mtsai, 2022; Arthur és Csapó, 2022a; Luo és mtsai, 2022), azonban ez még nem eredményezett jól érthető beszédet. Mivel az EEG csak a skalpon méri a jelet, ezért kevésbé pontosan lehet következtetni egyes agyi régiók aktivitására/tevékenységére, mintha közvetlenül az agyban mérnénk. Invazív módszerekkel már sikerült beszéd-szerű szintetizált beszédet létrehozni agyi jelek alapján, pl. ECoG (Herff és mtsai, 2015; Anumanchipalli és mtsai, 2019; Le Godais, 2022) és sEEG (Angrick és mtsai, 2021; Verwoert és mtsai, 2022; Arthur és Csapó, 2022b), de a fenti hátrány (invazív jelleg) miatt utóbbiak széles körű elterjedése nem várható.

1.1. Az agyi jelek és artikulációs mozgás kapcsolata

Az artikulációs mozgást még nem vizsgálták az agyi jelekkel párhuzamosan. A legtöbb kapcsolódó kutatás csak származtatott adatokat használ, azaz a beszédjelből visszakövetkeztetett artikulációs információt veszik figyelembe (Carey és mtsai, 2017; Anumanchipalli és mtsai, 2019; Favero és mtsai, 2022; Le Godais, 2022). Carey és mtsai (2017) kutatásában például ugyanazon beszélőkkel vettek fel MRI-t az artikulációs csatornáról, és funkcionális MRI-t az agyról. Mivel a két jel egyszerre nem rögzíthető, ezért a beszélők ugyanazt a stimulust többször megismételték a két modalitáshoz, így a beszédjelen „keresztül" vizsgálható az agyi jel és artikuláció kapcsolata. Anumanchipalli és mtsai (2019) az artikulációs csatornára vonatkozó kinematikus információt (pl. ajakmozgás, nyelvmozgás, és állkapocs állása), valamint az egyéb fiziológiai jellemzőket (pl. artikuláció mechanizmusa) a beszédjelből becsüli, akusztikum-artikuláció inverziós (Acoustic-to-Articulatory Inversion, AAI) módszerekkel. Az AAI modell betanítására a MOCHA-TIMIT adatbázist használták (az agyi jeles felvételektől független beszélőkkel), melyben elektromágneses artikulográffal rögzítették az artikulációs mozgást (Wrench, 2000). Favero és mtsai (2022) szintén AAI jellegű módszereket alkalmaztak, de itt az artikulációs információ nem valós mérésen alapul, hanem a beszédjelből számított ún. TADA jellemzőkön (Nam és mtsai, 2004). Le Godais (2022) kortikális aktivitásból beszéd dekódolása során lineáris módszerekkel próbálkozott az artikulációs információ hozzáadásával. Az artikulációt beszédből származtatta, dinamikus idővetemítés alapján, független beszélők artikulációs mozgását felhasználva.

A fenti tanulmányok konklúziója és távlati célja, hogy olyan betegek számára, akiknél az artikuláció agykérgi feldolgozása még érintetlen, a beszédalapú BCI dekóder intuitívabb lehet és könnyebben megtanulható a használata.

1.2. A jelen kutatás célja

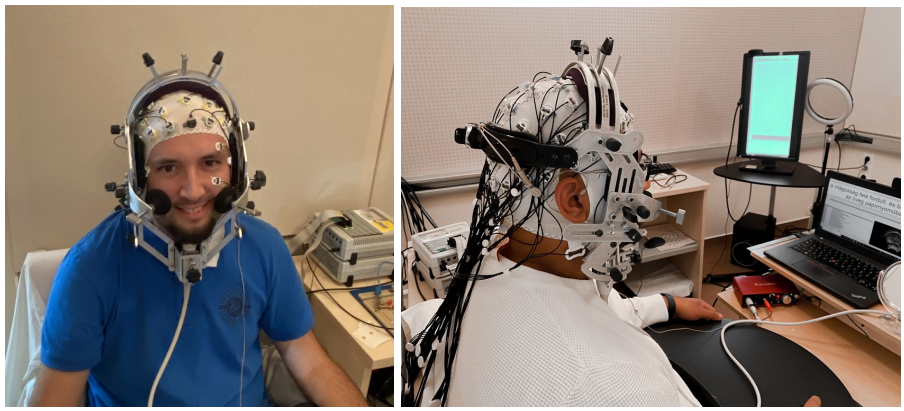
A jelen fejezet áttekintése szerint egyelőre hiányoznak azok a kombinált módszerek, amelyek a nem invazív EEG-t, az artikulációt és a beszédjeleket összevontan vizsgálnák, és elemeznék az agyban zajló tervezési folyamat, az artikulációs mozgás, és a keletkezett beszédjel kölcsönhatását. A jelen kutatásban a beszéd közbeni agyi jelek elemzését nyelvultrahang-alapú artikulációs adatokkal bővítjük, hogy több összehasonlítható bioszignál álljon rendelkezésre. Az EEG-vel mért agyi jelből mély neuronhálóval az artikulációs mozgásra vonatkozó információt prediktálunk, nyelvultrahang képek formájában.

2. Módszerek

2.1. Felvételek

A felvételek az ELKH TTK egyik csendes szobájában készültek. Az EEG jelet 64 csatornás Brain Products actiCHamp típusú erősítővel rögzítettük, actiCAP aktív elektródák felhasználásával. Négy csatorna a horizontális és vertikális szemmozgás követését szolgálta. Az elektródákat a nemzetközi 10-20-as elrendezés szerint helyeztük el (Klem és mtsai, 1999). Az elektródák impedanciáját 15 kOhm alatt tartottuk. A felvétel során az FCz elektróda töltötte be a referencia elektróda szerepét. A jelet 1000 Hz frekvencián mintavételeztük.

A nyelv középvonalának (szagittális) mozgását a „Micro” rendszerrel rögzítettük (AAA v220.02 szoftver, Articulate Instruments Ltd.) egy 2–4 MHz frekvenciájú, 64 elemű, 20 mm sugarú konvex ultrahang-vizsgálófejjel, 81,67 fps sebességgel, és rögzítő sisakot is alkalmaztunk (Csapó és mtsai, 2017a). A fém sisakot az EEG érzékelők fölé helyeztük el úgy, hogy az eszközök lehetőleg ne zavarják egymást. A felvételi elrendezésre az 1. ábra mutat példákat.

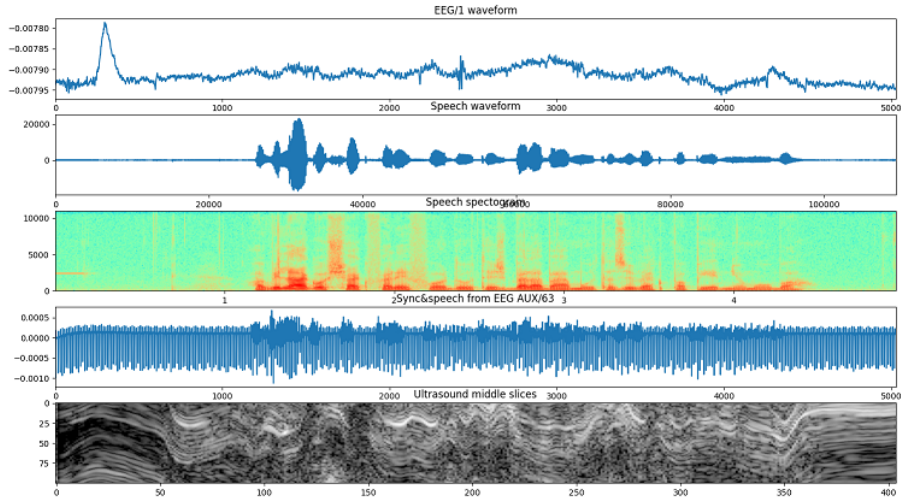


1. ábra: Felvételi elrendezés: EEG, nyelvultrahang, mikrofon és webkamera.

A beszédet egy Beyerdynamic TG H56c tan omnidirekcionális kondenzátor mikrofonnal vettük fel, és M-Audio M-Track 2x2 / FocusRite Scarlett 2i2 USB-s külső hangkártyával digitalizáltuk, 44 100 Hz-en. A beszélő arc- és szájmozgását egy Logitech C925e webkamerával, 1920x1080 pixel felbontásban rögzítettük.

A hangkártya kimenetét (amely a „Micro” ultrahang „frame sync” szinkronizáló jelét és a mikrofonból származó beszédjelet együttesen tartalmazza) rákötöttük az EEG AUX csatornájára – így az agyi és artikulációs jeleket ugyan külön számítógépeken rögzítettük, de mégis tudjuk utólag szinkronizálni az adatokat (ld. 2. ábra). Az EEG jelet folyamatosan vettük fel, míg a nyelvultrahangot és beszédet mondatonként. Mivel a beszédjel és a nyelvultrahang szinkronizációs jel (így az adott mondat felvételének kezdete és vége) megjelenik az EEG egyik csatornáján is, ezért a jeleket automatikusan szinkronizálni tudjuk utólag.

Az cikk beküldésének idejéig három magyar anyanyelvű férfi beszélővel rögzítettünk beszélőnként kb. 10–10 percnyi felvételt (a PPBA adatbázisból származó mondatokat (Olaszy, 2013)), melyet a későbbiekben további beszélőkkel bővítünk.

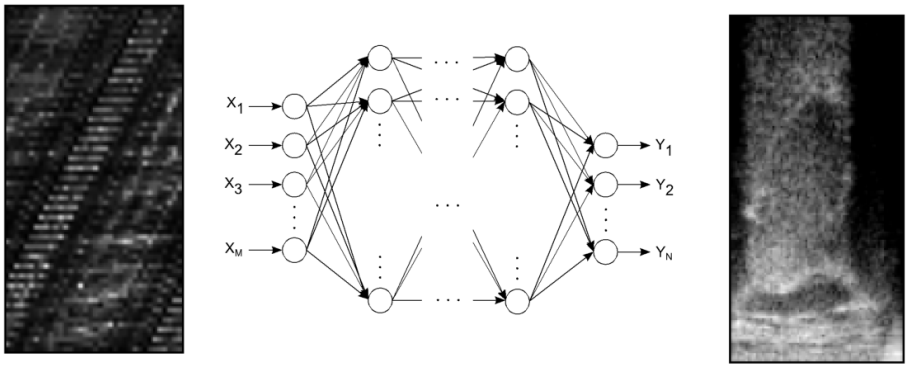


2. ábra: Példa a szinkronizált EEG, beszéd, és nyelvultrahang felvételre. a) EEG / 1. csatorna, b) beszédjel, c) beszéd spektrogram, d) ultrahang szinkornizálójel és beszédjel (EEG AUX-on), e) nyelvultrahang képek középső vonalának időbeli változása.

2.2. Az adatok előfeldolgozása

Az EEG jel előfeldolgozását Verwoert és mtsai (2022) alapján végeztük, a rendelkezésre álló szkriptekből kiindulva (`https://github.com/neuralinterfacinglab/SingleWordProductionDutch/blob/main/extract_features.py`). Az EEG jel minden csatornájára (az EEG AUX kivételével) kiszámítjuk a Hilbert burkolót négy frekvenciasávban: 1–50 Hz, 51–100 Hz, 101–150 Hz, és 151–200Hz. A burkolót 50 ms-onként átlagoltuk, és 12 ms-os eltolással számoltuk, hogy a nyelvultrahanggal összhangban legyen (melynek sebessége 81,67 fps volt). Ahhoz, hogy az időbeli információt is figyelembe vegyük, 4 megelőző és 4 követő blokkot is felhasználtunk a Hilbert-transzformált EEG jelekből. Az így előállított bemeneti jelre a 3. ábra bal oldala mutat példát.

A nyelvultrahang képeket 8 bites szürkeárnyalatos pixelekként használtuk fel, a „Micro" rendszer nyers ultrahang formájában. Az eredetileg 64x842 pixeles képeket átméreteztünk 64x128 pixelre (3. ábra, jobb oldal), mivel ez nem okoz jelentős információ veszteséget (Csapó és mtsai, 2022), de így kevesebb a feldolgozandó adat mennyisége.



3. ábra: A neuronháló bemenete (bal oldalon) és kimenete (jobb oldalon).

2.3. Artikuláció predikciója EEG bemenetből

A kutatás kezdeti fázisában egy egyszerű kísérletet végeztünk: teljesen kapcsolt (fully connected, FC-DNN) mély „egyenirányított" (rectifier) neurális hálózatot (Glorot és mtsai, 2011) tanítottunk, melynek során a nyelvultrahang képeket predikáltuk, a Hilbert-transzformált EEG bemenetből (3. ábra). A tanítás során az átlagos négyzetes hibafüggvényt (MSE) alkalmaztuk. Kísérleteink során egy 5 rejtett réteges, rétegenként 1000 neuront tartalmazó neuronháló struktúrát használtunk, ReLU aktivációval és lineáris kimeneti réteggel (hasonlóan az első nyelvultrahang-beszéd szintézis tanulmányunkhoz, (Csapó és mtsai, 2017b)). A

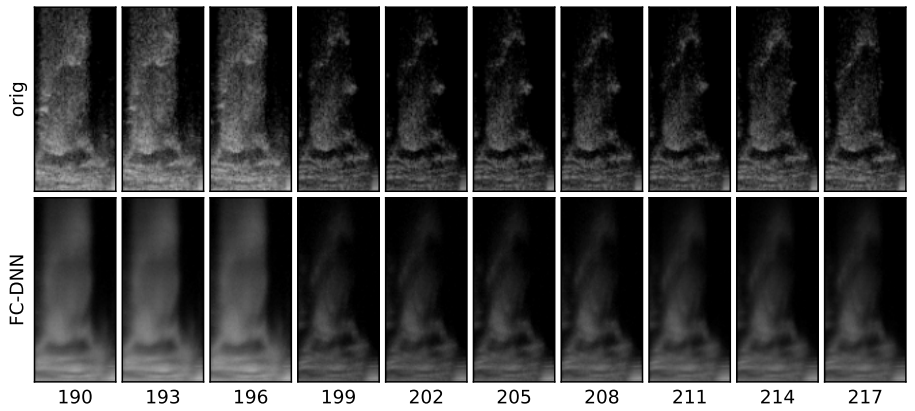
bemeneti EEG értékeket és a kimeneti ultrahang pixeleket a tanítás előtt 0–1 közé normalizáltuk. 100 epochig tanítottunk, de korai leállást alkalmaztunk, azaz ha a validációs hiba nem csökkent 3 epoch-on keresztül, a tanítást leállítottuk.

3. Kísérletek és eredmények

A rendelkezésre álló felvételekből beszélőfüggő tanításokat végeztünk, az adatok 80%-át használtunk a neuronháló tanítására, 10%-ot validációra, a maradék 10%-ot pedig tesztelésre (FF1 beszélő esetén: 25 600 / 3200 / 3200 mintapont).

3.1. Demonstrációs minták

A DNN tanítás után a teszthalmazon EEG-ből nyelvultrahang predikciót végeztünk. A 4. ábra az FF1 beszélőtől mutat néhány eredeti és EEG alapján becsült nyelvultrahang képet, az ultrahang gép „nyers" reprezentációjában. A nyelvultrahang kontúrja az eredeti képeken sem minden esetben látszik jól – ennek oka a nyelvultrahang beszélőfüggősége (Csapó, 2022). Az EEG alapján becsült képeken a nyelv kontúrja elkent, és a nyelv helyzetének képkockát közötti változása is nehezen kivehető - azaz a DNN az általános nyelvalakot (az átlagos képet) meg tudta tanulni, de a nyelvmozgásra vonatkozó finom részletek nem kivehetőek.



4. ábra: Demonstrációs minta: eredeti (felül) és EEG alapján becsült (alul) nyelvultrahang képek az FF1 beszélőtől, „nyers" ultrahang reprezentációban.

Ugyanezt a képsorozatot mutatja az 5. ábra, „szétterített" reprezentációban (a transzformációt az `https://github.com/UltraSuite/ultrasuite-tools` eszközzel végeztük). A szürkeárnyalatos képeken hasonló tendencia vehető észre, mint a 4. ábrán: az eredeti nyelvultrahang képeken még sejthető a nyelv felső

határvonala, de az EEG alapján becsült képeken már elmosódnak az ultrahang pixelek, és a nyelv kontúrja nem látható. Viszont a 196 és 199 időpillanat között a fényerősség váltása a DNN-predikált esetben is észrevehető.

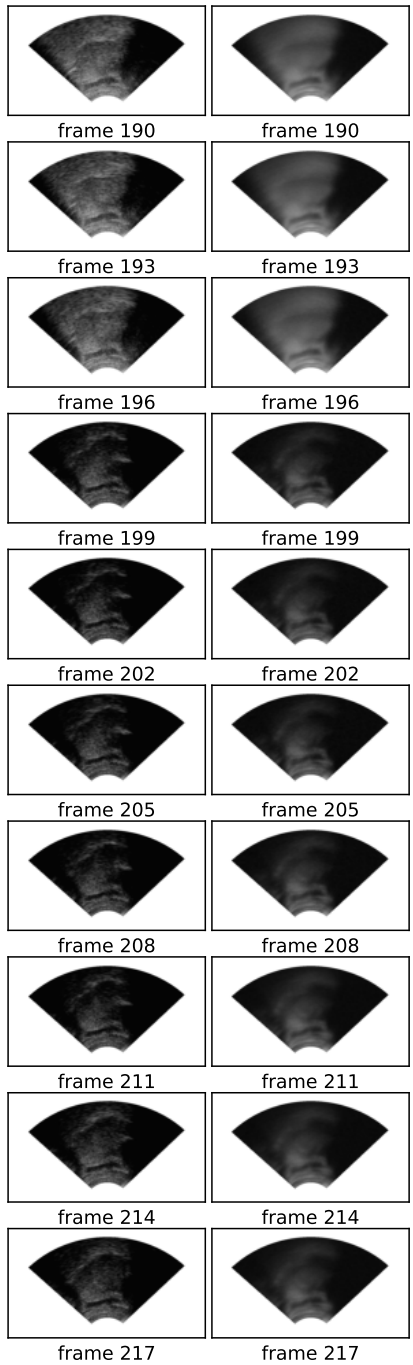
Ha képenként nézzük az eredményeket (mint a 4. és 5. ábrákon), akkor kevésbé látszik a hosszabb időtávlatú tendencia. Emiatt egy másik elrendezésben is ábrázoljuk az eredményeket: minden nyelvultrahang képből kivágtuk a középső függőleges vonalat (kb. ez megfelel a nyelv közepének), és ezen vonal időbeli változását ábrázoltuk, egy spektrogramhoz hasonlóan. A 6. ábra ennek eredményét mutatja: felül a beszédhez tartozó spektogram, középen az ugyanehhez bemondáshoz tartozó nyelvultrahang középvonal időbeli változása, alul pedig a DNN által prediktált nyelvultrahang középvonal látható. Az a) mel-spektrogram és b) artikulációs mozgás közötti hasonlóság egyértelműen észrevehető: a beszédben lévő formánsmozgások, és a nyelv függőleges mozgása nagyjából kivehető az ábrákon. A c) DNN-predikált nyelvultrahang középvonalon viszont a nyelv mozgása nem látszik, azaz a DNN nem tudta megtanulni az EEG és a nyelvultrahang közötti összefüggést. Ugyanakkor valamilyen információ mégis látszik a DNN-predikált képeken: a 195. időpillanat végén az egyik mondatnak vége van, és a következő elkezdődik, ami a b) eredeti nyelvultrahangon jól kivehető, és a c) becsült nyelvultrahangon is látszik.

3.2. Objektív mérések

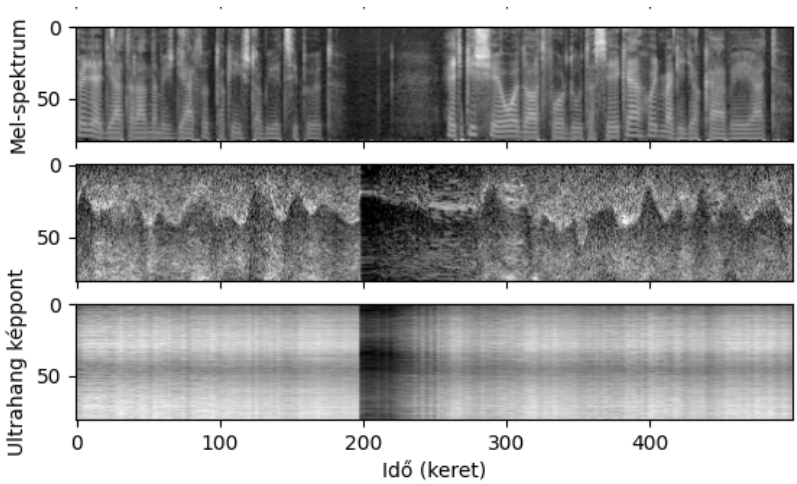
A fenti FC-DNN hálózattal elért átlagos négyzetes hiba (MSE) értékek az FF1 beszélő esetén: validációs hiba: 0,0053, teszthiba: 0,0055. Az érték önmagában nehezen értelmezhető, és az sem egyértelmű, milyen jóságot ír le, de pl. korábbi akusztikum-artikuláció inverziós kísérleteinkben (melynek során a beszédjelből becsültünk nyelvultrahang képeket (Porras és mtsai, 2019; Csapó és Sepúlveda, 2021)) a kapott NMSE validációs hiba értékek 0,0053–0,0088 nagyságrendben voltak; és ez esetben a beszédből generált nyelvulrahang kép közelítette az eredeti artikulációt. Ebből is látható, hogy az MSE érték a jelen kutatásban nem elegendő az eredmények jóságának megítélésére, mindenképp szükség van a vizuális vizsgálatra. A korábbi ultrahangos kutatásaink esetén kísérleteztünk más hibamértékek vizsgálatával is, úgymint Structural Similarity Index (SSIM) (Wang és mtsai, 2004), és Complex Wavelet Structural Similarity (CW-SSIM) (Sampat és mtsai, 2009), ultrahangon (Xu és mtsai, 2016; Csapó és mtsai, 2020; Csapó és Sepúlveda, 2021). Azonban a fenti vizuálisan gyenge eredmények miatt az SSIM-et és CW-SSIM-et itt az EEG-ből kiinduló nyelvultrahang becslés esetén egyelőre nem vizsgáltuk.

4. Diszkusszió és következtetések

A kutatás jelen kezdeti fázisában ismertetett multimodális (agy, beszéd és artikuláció) analízis és szintézis révén túlmutatunk a legkorszerűbb nemzetközi trendeken.



5. ábra: Demonstrációs minta: eredeti (balra) és EEG alapján becsült (jobbra) nyelvultrahang képek az FF1 beszélőtől, „szétterített” ultrahang reprezentációban.



6. ábra: Demonstrációs minta: a) eredeti beszédminta 80-dimenziós mel-spektrogramja, b) eredeti nyelvultrahang felvétel középvonalának időbeli változása, c) a becsült nyelvultrahang középvonalának időbeli változása.

A szakirodalomban számos korábbi kísérletet láttunk EEG (vagy más eszközzel mért agyi jel) alapján készült kezdeti beszéd BCI kutatásokra (Herff és mtsai, 2015; Anumanchipalli és mtsai, 2019; Krishna és mtsai, 2020; Arthur és Csapó, 2022a,b), azonban egyelőre az agyi jel alapú beszédszintézis esetén nem sikerült jól érthető beszédet létrehozni. Kézenfekvő megoldásnak tűnik az artikuláció, mint az agyi jel és végső beszéd közti köztes reprezentáció vizsgálata, mellyel a jelen cikkben is foglalkoztunk. Korábbi kutatásokban az artikulációs adatokat csak származtatott módon, nem közvetlenül mérve tudták felhasználni az agyi jel és beszéd vizsgálata során (Carey és mtsai, 2017; Anumanchipalli és mtsai, 2019; Favero és mtsai, 2022; Le Godais, 2022). Bár ez a közvetett artikulációs információs is segített az eredmények pontosításában, az artikuláció valós eszközökkel történő mérése további javulást eredményezhet.

A jelen kutatásban az EEG-vel mért agyi jel és mikrofonnal rögzített beszéd vizsgálatát egészítettük ki nyelvultrahanggal mért artikulációs felvételekkel. Mély neuronháót (FC-DNN) taníottunk EEG bemenet alapján nyelvultrahang becslésére. Az eredmények szerint a generált nyelvultrahang képsorozat még távol van az eredeti nyelvultrahangtól, de az EEG és a nyelvultrahang közötti kapcsolat egyértelműen kimutatható.

A jelen kutatás hosszútávú célja, hogy hozzájáruljunk a beszéd alapú agy-számítógép interfészekhez. Az eredmények potenciálisan alkalmazhatók lehetnek rehabilitáció során, pl. kommunikációs segédeszközként.

A fent bemutatott DNN-es kísérletek keras implementációja a következő címen érhető el: `https://github.com/BME-SmartLab/EEG-to-UTI`

Köszönetnyilvánítás

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal OTKA programja támogatta (FK 142163 projekt). Csapó Tamás Gábor kutatásait az MTA Bolyai János kutatói ösztöndíja, valamint az Új Nemzeti Kiválóság Program Bolyai+ (ÚNKP-22-5-BME-316) pályázata támogatta.

Szeretnénk köszönetet mondani Béres Lucának és Várkonyi Emesének az EEG felvételekben nyújtott segítségért, az ELKH TTK-nak az EEG eszközök biztosításáért, valamint az MTA-ELTE „Lendület" Lingvális Artikuláció Kutatócsoportnak a nyelvultrahang eszközök rendelkezésre bocsátásáért.

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

Anumanchipalli, G.K., Chartier, J., Chang, E.F.: Speech synthesis from neural decoding of spoken sentences. Nature 568(7753), 493–498 (apr 2019), https://www.nature.com/articles/s41586-019-1119-1

Arthur, F.V., Csapó, T.G.: Deep learning alapú agyi jel feldolgozás és beszédszintézis előkészítő munkálatai. In: MSZNY 2022. pp. 185–198. online (2022a)

Arthur, F.V., Csapó, T.G.: Speech synthesis from intracranial stereotactic Electroencephalography using a neural vocoder. In: submitted (2022b)

Buzsáki, G., Anastassiou, C.A., Koch, C.: The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nature Reviews Neuroscience 13(6), 407–420 (may 2012), https://www.nature.com/articles/nrn3241

Carey, D., Miquel, M.E., Evans, B.G., Adank, P., Mcgettigan, C.: Vocal Tract Images Reveal Neural Representations of Sensorimotor Transformation During Speech Imitation. Cerebral Cortex 27(5), 3064–3079 (2017)

Casson, A.J.: Wearable EEG and beyond. Biomedical engineering letters 9(1), 53–71 (feb 2019), https://pubmed.ncbi.nlm.nih.gov/30956880/

Chang, E.F., Anumanchipalli, G.K.: Toward a Speech Neuroprosthesis. JAMA 323(5), 413–414 (feb 2020), https://jamanetwork.com/journals/jama/fullarticle/2758116

Csapó, T.G.: A nyelvmozgás ultrahangos vizsgálata és az automatikus elemzés alkalmazási lehetőségei a beszédtechnológiában, pp. 197–219 (2022), https://m2.mtmt.hu/api/publication/33232043

Csapó, T.G., Deme, A., Gráczi, T.E., Markó, A., Varjasi, G.: Szinkronizált beszéd- és nyelvultrahang-felvételek a SonoSpeech rendszerrel. In: MSZNY 2017. pp. 339–346. Szeged (2017a)

Csapó, T.G., Gosztolya, G., Tóth, L., Shandiz, A.H., Markó, A.: Optimizing the Ultrasound Tongue Image Representation for Residual Network-Based Articulatory-to-Acoustic Mapping. Sensors 22 (2022)

Csapó, T.G., Grósz, T., Tóth, L., Markó, A.: Beszédszintézis ultrahangos artikulációs felvételekből mély neuronhálók segítségével. In: MSZNY 2017. pp. 181–192 (2017b)

Csapó, T.G., Sepúlveda, A.: Ultrasound Tongue Image Generation for Acoustic-to-Articulatory Inversion using Convolutional and Recurrent Deep Neural Networks. submitted to Multimedia Tools and Applications (2021)

Csapó, T.G., Xu, K., Deme, A., Gráczi, T.E., Markó, A.: Transducer Misalignment in Ultrasound Tongue Imaging. In: 12th International Seminar on Speech Production (2020)

Dash, D., Ferrari, P., Babajani-Feremi, A., Borna, A., Schwindt, P.D., Wang, J.: Magnetometers vs Gradiometers for Neural Speech Decoding. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference 2021, 6543–6546 (nov 2021), https://pubmed.ncbi.nlm.nih.gov/34892608/

Favero, P., Berezutskaya, J., Ramsey, N.F., Nazarov, A., Freudenburg, Z.V.: Mapping Acoustics to Articulatory Gestures in Dutch: Relating Speech Gestures, Acoustics and Neural Data. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 802–806 (2022)

Glorot, X., Bordes, A., Bengio, Y.: Deep Sparse Rectifier Neural Networks. In: Gordon, G.J., Dunson, D.B. (szerk.) Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS). vol. 15, pp. 315–323. Journal of Machine Learning Research - Workshop and Conference Proceedings, Ft. Lauderdale, FL, USA (2011), http://www.jmlr.org/proceedings/papers/v15/glorot11a/glorot11a.pdf

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

Klem, G.H., Lüders, H.O., Jasper, H.H., Elger, C.: The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. Electroencephalography and clinical neurophysiology. Supplement (jan 1999), https://www.scienceopen.com/document?vid=5960cfa8-7fde-441c-8592-35fdb9841499

Krishna, G., Tran, C., Han, Y., Carnahan, M., Tewfik, A.H.: Speech Synthesis Using EEG. In: Proc. ICASSP. pp. 1235–1238. online (2020)

Le Godais, G.: Decoding speech from brain activity using linear methods. Ph.D.-értekezés, Université Grenoble Alpes (2022), https://tel.archives-ouvertes.fr/tel-03852448

Luo, S., Rabbani, Q., Nathan, ·., Crone, E.: Brain-Computer Interface: Applications to Speech Decoding and Synthesis to Augment Communication. Neurotherapeutics 2022 1, 1–11 (jan 2022), https://link.springer.com/article/10.1007/s13311-022-01190-2

McFarland, D.J., Wolpaw, J.R.: EEG-based brain–computer interfaces. Current Opinion in Biomedical Engineering 4, 194–200 (dec 2017)

Nam, H., Goldstein, L., Saltzman, E., Byrd, D.: TADA: An enhanced, portable Task Dynamics model in MATLAB. The Journal of the Acoustical Society of America 115(5), 2430 (apr 2004), https://asa.scitation.org/doi/abs/10.1121/1.4781490

Olaszy, G.: Precíziós, párhuzamos magyar beszédadatbázis fejlesztése és szolgáltatásai [Development and services of a Hungarian precisely labeled and segmented, parallel speech database] (in Hungarian). Beszédkutatás 2013 [Speech Research 2013] pp. 261–270 (2013)

Porras, D., Sepúlveda-Sepúlveda, A., Csapó, T.G.: DNN-based Acoustic-to-Articulatory Inversion using Ultrasound Tongue Imaging. In: International Joint Conference on Neural Networks. pp. N–19221. Budapest, Hungary (2019), http://arxiv.org/abs/1904.06083

Sampat, M.P., Wang, Z., Gupta, S., Bovik, A.C., Markey, M.K.: Complex Wavelet Structural Similarity: A New Image Similarity Index. IEEE Transactions on Image Processing 18(11), 2385–2401 (nov 2009), http://ieeexplore.ieee.org/document/5109651/

Verwoert, M., Ottenhoff, M.C., Goulis, S., Colon, A.J., Wagner, L., Tousseyn, S., van Dijk, J.P., Kubben, P.L., Herff, C.: Dataset of Speech Production in intracranial Electroencephalography. Scientific Data 2022 9:1 9(1), 1–9 (jul 2022), https://www.nature.com/articles/s41597-022-01542-9

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing 13(4), 600–612 (apr 2004), http://ieeexplore.ieee.org/document/1284395/

Wrench, A.A.: A Multichannel Articulatory Database and its Application for Automatic Speech Recognition. In: Proc. 5th Seminar on Speech Production: Models and Data. pp. 305–308. Kloster Seeon, Bavaria, Germany (2000), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.195.1846

Xu, K., Csapó, T.G., Roussel, P., Denby, B.: A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization. The Journal of the Acoustical Society of America 139(5), EL154–EL160 (may 2016), http://scitation.aip.org/content/asa/journal/jasa/139/5/10.1121/1.4951024