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

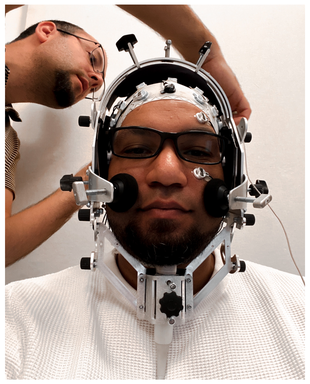
{csapot, arthur}@tmit.bme.hu, nagy.peter.ssprg@ttk.hu, adam.boncz@gmail.com

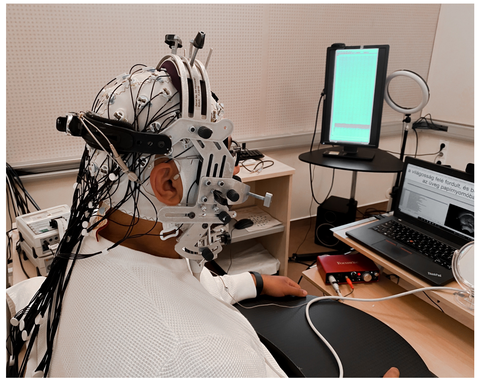
1. Az agyi jelek és artikulációs mozgás kapcsolata

1.1 A jelen kutatás célja

- Agy-gép interfészek fejlesztéséhez igyekszünk hozzájárulni, hogy hosszú távon pótolhassuk az elvesztett beszédképességet
- Multimodális analízist alkalmazunk az agy-beszéd-artikulációs interakciók megértéséhez
- Hosszú távú célunk: hozzájárulni a beszéd alapú BCI fejlesztéséhez, beszédsérülteknek szánt kommunikációs segédeszközök fejlesztéséhez
- Az EEG-t tartjuk a jelenleg elérhető legmegfelelőbb módszernek az agyi jelek rögzítésére, mivel elérhető árú, nem invazív és hordozható
- Az artikulációs mozgás és az agyi jelek közötti kapcsolat kevésbé volt kutatva

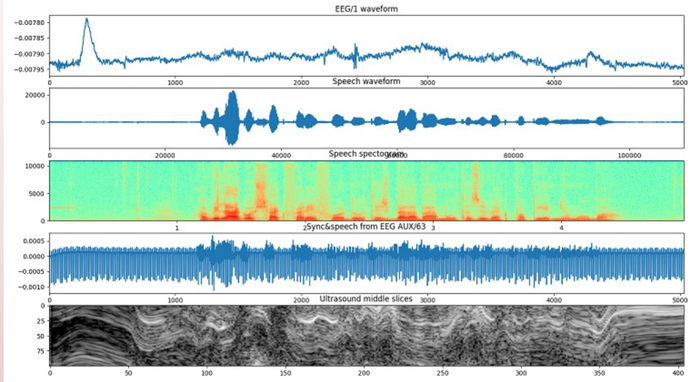
2. Módszerek





2.1 Felvételek

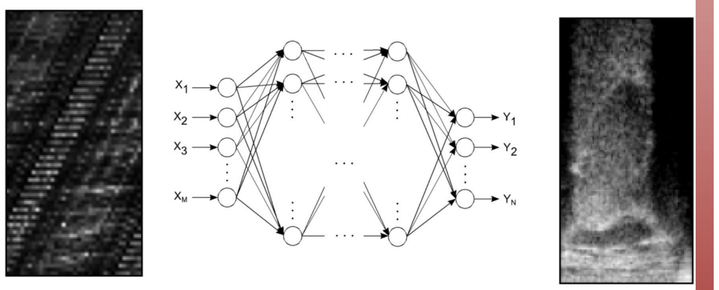
- A felvételeket az ELKH TTK csendes szobájában végeztük.
- Az EEG jeleket egy 64 csatornás Brain Products actiCHamp erősítővel és actiCAP aktív elektródákkal rögzítettük.
- Négy csatornát használtunk a szemmozgások követésére
- Az elektródákat a nemzetközi 10-20-as elrendezésnek megfelelően helyeztük el.
- Az impedanciát 15 kOhm alatt tartottuk.
- Az FCz elektródát használtuk referencia elektródaként.
- A jelet 1000 Hz-en mintavételeztük
- A nyelv ultrahangját a "Micro" rendszerrel rögzítettük 2-4 MHz-es frekvenciájú, 64 elemes, 20 mm átmérőjű, domború ultrahangszondával 81,67 fps sebességgel.
- A beszédet 44,100 Hz-es frekvencián, egy omnidirekcionális kondenzátor mikrofonnal rögzítettük.
- A beszélő arc- és szájmozgását webkamerával rögzítettük.
- 3 férfi beszélő



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-jeleket a Verwoert et al. (2022) módszerével előfeldolgoztuk, és négy frekvenciasávban kiszámítottuk a Hilbert-burkolót.
- A burkológörbét az ultrahangadatokkal összehangoltuk 50 ms-os ablakok 12 ms-os eltolásokkal történő átlagolásával.
- A Hilbert-transzformált EEG-jelek 4 megelőző és 4 követő blokkját használtuk az időbeli információk beépítéséhez.
- A nyelv ultrahangjának 8 bites szürkeárnyalatos képeit használtuk, és a könnyebb feldolgozás érdekében 64x128 pixelre méreteztük át őket.

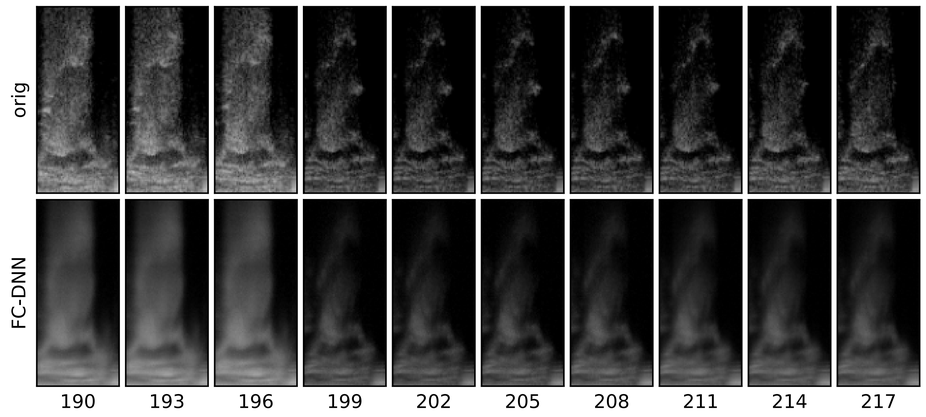


A neuronháló bemenete (bal oldalon) és kimenete (jobb oldalon).

2.3. Artikuláció predikciója EEG bemenetből

- A kutatás kezdeti szakaszában egy egyszerű kísérletet végeztünk: Egy teljesen összekapcsolt mély neurális hálózatot (Glorot et al., 2011) tanítottunk beszédultrahang-képek prediktációjára Hilbert-transzformált EEG-bemenet segítségével.
- A tanítás során az átlagos négyzetes hiba függvényt (MSE) használtuk.
- Kísérleteinkben 5 rétegű neurális hálózati struktúrát használtunk, rétegenként 1000 neuronnal, ReLU aktiválással és lineáris kimeneti réteggel (hasonlóan az első beszédultrahang-szintézis vizsgálatunkhoz, (Csapó et al., 2017b)).
- A bemeneti EEG értékeket és a kimeneti ultrahang pixeleket a tanítás előtt 0-1 között normalizáltuk.

3. Kísérletek és eredmények



4. Összefoglalás

- Tanulmány a multimodális (EEG, UTI, beszéd) elemzés és beszédszintetizáció kezdeti szakaszának lefolytatására
- Teljesen összekapcsolt DNN az artikuláció becslésére az EEG-bemenetből
- Az eredmények azt mutatják, hogy a generatív artikuláció még mindig messze van az eredetitől
- További kutatás szükséges a tiszta beszédgeneráláshoz
  - Lendület pályázat beadása folyamatban

Hivatkozások

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

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

Köszönetnyilvánítás

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal OTKA programja támogatta (FK~142163 projekt) OTKA FK-22 / "Az artikuláció és az agyi jelek elemzése beszéd alapú agy-gép interfészhez". 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.

Forráskód