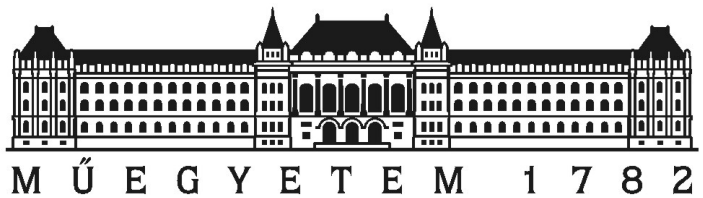
Agyi jel (iEEG) alapján beszédszintézis deep learning módszerekkel

Köller Donát Vastag Emese Vlaszov Artúr

Budapesti Műszaki és Gazdaságtudományi Egyetem



Témabemutatás

- Cél: kommunikációs eszköz biztosítása beszédben korlátozottak számára
- Agyi jelek átalakítása spektrogrammá, majd beszéd szintetizálás
- Agyi jelek rögzítése: esetünkben invazív módszerrel, intrakraniális EEG jelek

Háttér, korábbi megoldások

- Erősen kutatott téma, deep learning módszerekkel jelentős eredmények
- Miguel Angrick et al.: sEEG felvételek modellezése konvolúciós hálókkal [1]
- Gautam Krishna et al.: state-of-the-art eredmények RNN és GAN alapú rendszerekkel [2]
- Többségében invazív módszerek

Adatfeldolgozás

- Adathalmaz: SingleWordProductionDutch-iBIDS (Maxime Verwoert et al.) [3]
- 10 résztvevő, beszélt szöveg mellett 64 csatornás iEEG jelek rögzítése

Előfeldolgozási lépések

- Egy beszélős rendszer: eredeti cikkel megegyező
- Beszélőfüggetlen rendszer: magasabb dimenzióban iterált tanító-validációs-teszt halmaz szétválasztás alanyonként
  - Arány: 60-20-20

Egy beszélős rendszer

- 3 modell: Bottleneck FC DNN, FC DNN, BiGRU
- Rekonstrukció 5 iterációban
- Baseline tanítás
- Keras Tuner-rel optimizált tanítás

Beszélőfüggetlen rendszer

Kétféle megközelítés:

- Eredeti, magasabb dimenziós adattal
  - BiGRU modell
  - Konvolúciós háló
- Kisebb dimenzióba transzformált adattal
  - Finomhangolt AutoEncoder modell bottleneck rétege
  - Incremental PCA
  - Ezekre teljesen előrecsatolt hálózatok alkalmazása

Optimalizálás: Keras Tuner-rel

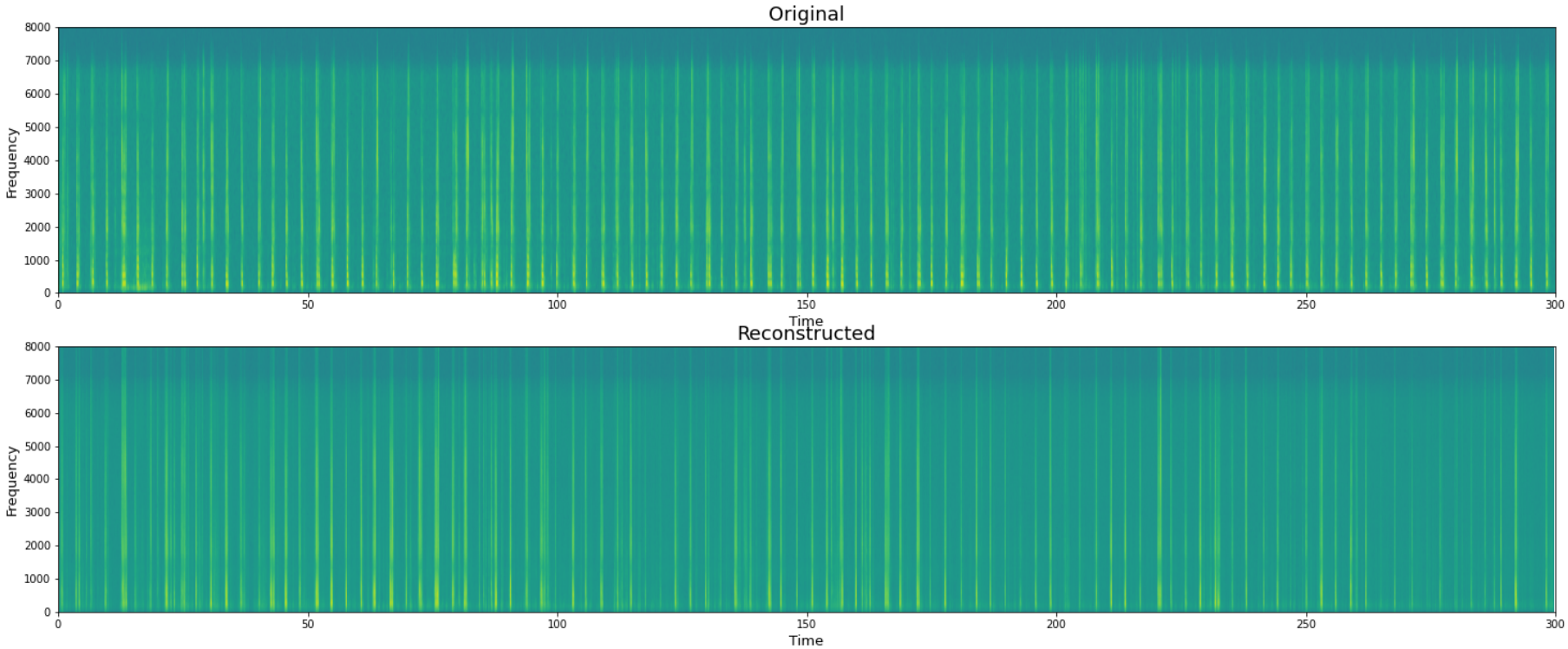
Kiértékelés 1.

- Egy beszélős rendszer: RMSE rekonstrukciós etaponként átlagolva
- Beszélőfüggetlen rendszer: MCD alanyonként átlagolva

		Metrikák		
Model típus	Hálózat	RMSE	Pearson korr.	MCD
Egy beszélős rendszer	Bottleneck	**1.5783**	**0.5670**	4.7027
	FC-DNN	1.5895	0.5661	7.8219
	BiGRU	1.6071	0.5627	**4.3637**
Beszélőfüggetlen rendszer	FC-DNN	1.4876	0.7329	**1.2362**
	BiGRU	**1.4536**	**0.7356**	1.3598
	Convolutional	1.4737	0.7273	1.6827

táblázat: Eredmények modellenként

Kiértékelés 2.



ábra: Az egy beszélős BiGRU modell eredménye

Összegzés

- Agyi jelek alapján rekonstruáltunk beszédet deep learning módszerekkel
- Habár a teljesítményt mérő metrikák értéke egészen jó, a modellek még nem használhatóak
- A rekonstruált hangfájlokban nem érthető a beszéd

További céljaink:

- Legjobb modellek kipróbálása a többi alanyon is
- További modellek alkalmazása: WaveGlow [4], HifiGAN [5]

Köszönjük a figyelmet!

Hivatkozások


Angrick, M., Ottenhoff, M., Goulis, S., Colon, A., Wagner, L., Krusienski, D., Kubben, P., Schultz, T., & Herff, C. (2021). Speech Synthesis from Stereotactic EEG using an Electrode Shaft Dependent Multi-Input Convolutional Neural Network Approach. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 6045-6048). doi:10.1109/EMBC46164.2021.9629711


Krishna, G., Tran, C., Carnahan, M., & Tewfik, A. H. (2021). Advancing Speech Synthesis using EEG. *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 199–204. doi:10.1109/NER49283.2021.9441306


Verwoert, M., Ottenhoff, M.C., Goulis, S. et al. (2022). Dataset of Speech Production in intracranial Electroencephalography. *Sci Data* **9**, 434. https://doi.org/10.1038/s41597-022-01542-9


Prenger, R., Valle, R., & Catanzaro, B. (2018). WaveGlow: A Flow-based Generative Network for Speech Synthesis. *ArXiv E-Prints*, arXiv:1811.00002. Retrieved from http://arxiv.org/abs/1811.00002


Kong, J., Kim, J., & Bae, J. (2020). HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis. *ArXiv E-Prints*, arXiv:2010.05646. Retrieved from http://arxiv.org/abs/2010.05646