Corentin Puffay[1,2], Jonas Vanthornhout[1], Marlies Gillis[1], Bernd Accou[1,2], Hugo Van hamme[2], Tom Francart[1]
[1]KU Leuven, Dept. Neurosciences, ExpORL, Leuven, Belgium, [2]KU Leuven, Dept. of Electrical engineering, PSI, Leuven, Belgium

Relating linguistic features of speech with EEG using a dilated convolutional network

Background

State-of-the-art (SoA)

- Neural tracking of linguistics shown in linear models ([2, 3, 5])
- Linguistics processing can be a marker for speech understanding
- SoA studies: require control for acoustics to take into account correlation between speech features

Aim of the study

1. Measure neural tracking of linguistics in EEG using a deep learning model to overcome the linear assumptions made in SoA models
2. New method: control for lexical information to measure the semantic component of linguistic features
3. Measure the added value of linguistics over lexical features using a multi-input feature model [7]

Dataset

- Subjects: 60 young normal hearing native Flemish, screened with pure-tone audiometry and Flemish MATRIX test
- Stimulus (62 dBA, binaural): 10 children's stories (14 min each) without background noise, each narrated by a different single speaker.

Feature extraction

EEG was processed and linguistic and lexical features were extracted according to [5].

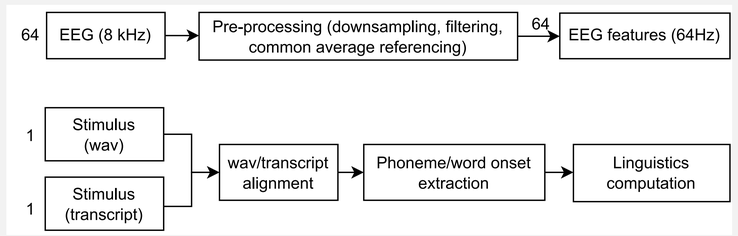
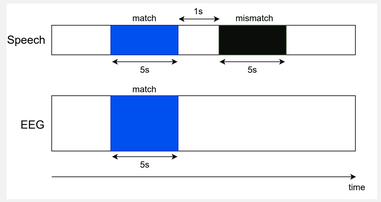


Figure: EEG and speech features extraction steps

- Lexical features: word onsets (WO), one-dimensional phoneme onsets (PO1D)
- Phoneme-based linguistics: phoneme surprisal (PS), cohort entropy (CE), for a phoneme i: $PS_i = -log_2(\frac{freq(cohort_i)}{freq(cohort_{i-1})})$, $CE_i = -\sum_{word}^{cohort_i} p(word) log_2(p(word))$
- Word-based linguistics: word frequency (WF), word surprisal (WS), for a word i: $WS_i = -log_{10}(p(w_i|w_{i-5}, .., w_{i-1}))$, $WF_i = -log_{10}(p(w_i))$

The match-mismatch task

We measure neural tracking of speech features in the EEG using the match-mismatch task [4, 6, 1].



Multi-input feature model [7]

We investigate the difference between the performance of the control models (2PO1D or 2WO, integrating twice onset information) and the linguistics model (PO1D+CE/PS or WO+WS/WF respectively).

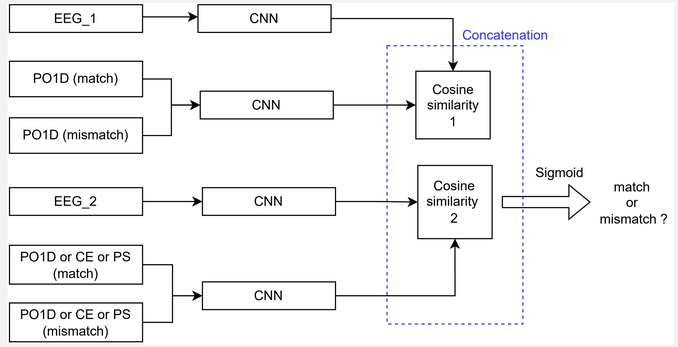


Figure: Multi-input feature model

Result 1: Word-level linguistic features

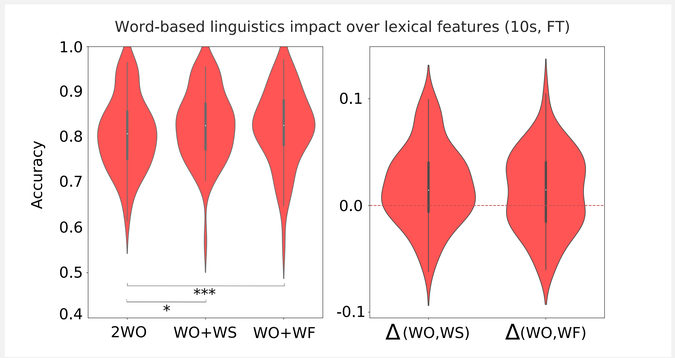


Figure: Impact of world-level linguistic features over word onsets. *(*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$)*

Result 2: Phoneme-level linguistic features

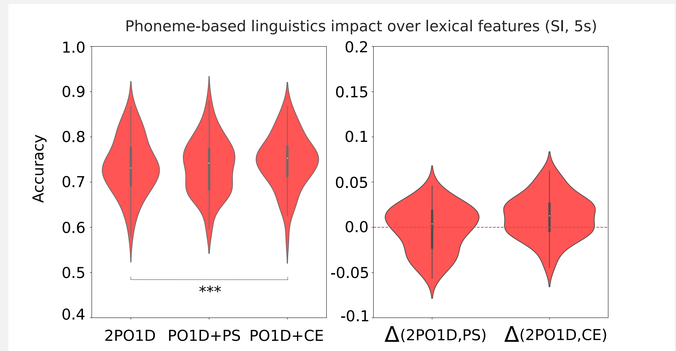


Figure: Impact of phoneme-level linguistic features over phoneme onsets. *(*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$)*

Discussion and Conclusion

- Phoneme-based linguistics models (PO1D+CE) have an added value over control models (2PO1D)
- Word-based linguistics models (WO+WS, WO+WF) have an added value over control models (2WO)
- We measured neural tracking of the semantic component of linguistics features
- We will next use this model to measure speech understanding and correlate results with behavioural measurements

Acknowledgments

Funding was provided by the KU Leuven Special Research Fund C24/18/099 (C2 project to Tom Francart and Hugo Van hamme), FWO research project G0D6720N, and an FWO postdoctoral fellowship to Jonas Vanthornhout (1290821N).

Bibliography

[1] B Accou, M Jalilpour-Monesi, J Montoya-Martínez, HV hamme, and T Francart. "Modeling the relationship between acoustic stimulus and EEG with a dilated convolutional neural network". In: *2020 28th European Signal Processing Conference (EUSIPCO)* (2021), pp. 1175–1179.

[2] C Brodbeck, LE Hong, and JZ Simon. "Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech". In: *Current Biology* 28.24 (2018), 3976–3983.e5.

[3] MP Broderick, AJ Anderson, GM Di Liberto, MJ Crosse, and EC Lalor. "Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech". In: *Current Biology* 28.5 (2018), 803–809.e3.

[4] A de Cheveigné, M Slaney, SA Fuglsang, and J Hjortkjaer. "Auditory stimulus-response modeling with a match-mismatch task". In: *Journal of Neural Engineering* 18.4 (2021), p. 046040.

[5] M Gillis, J Vanthornhout, JZ Simon, T Francart, and C Brodbeck. "Neural markers of speech comprehension: measuring EEG tracking of linguistic speech representations, controlling the speech acoustics". In: *Journal of Neuroscience* (2021). eprint: https://www.jneurosci.org/content/early/2021/10/29/JNEUROSCI.0812-21.2021.full.pdf.

[6] MJ Monesi, B Accou, J Montoya-Martinez, T Francart, and HV Hamme. "An LSTM Based Architecture to Relate Speech Stimulus to Eeg". In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* 2020-May.637424 (2020), pp. 941–945.

[7] C Puffay, J Van Canneyt, J Vanthornhout, H Van Hamme, and T Francart. *Relating the fundamental frequency of speech with EEG using a dilated convolutional network*. 2022.