Predictive audiovisual speech processing with and without mouth cues in cochlear implant users

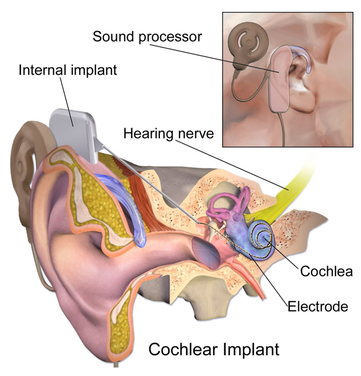
Simone Gastaldon[1,2], Noemi Bonfiglio[1], Francesco Vespignani[1,3], Flavia Gheller[1,3], Davide Brotto[4,3], Davide Bottari[5], Patrizia Trevisi[4,2,3], Alessandro Martini[4,2,3], Francesca Peressotti[1,2,3]

1. Dipartimento di Psicologia dello Sviluppo e della Socializzazione (DPSS), University of Padova, Italy **2.** Padova Neuroscience Center (PNC), University of Padova, Italy **3.** Centro Interdipartimentale di Ricerca "I-APPROVE – International Auditory Processing Project in Venice", University of Padova, Italy **4.** Dipartimento di Neuroscienze (DNS), UOC Otorinolaringoiatria, University of Padova, Italy **5.** Scuola IMT Alti Studi Lucca, Italy

s_gastaldon
simone.gastaldon@unipd.it

INTRODUCTION

Language comprehension is
PROACTIVE: top-down prediction of information **facilitates bottom-up processing** [1,2] especially in challenging situations [3]
MULTIMODAL: seeing the mouth of the speaker influences speech perception [4]
Comprehending speech is more than simply perceiving sounds. What happens when the sensory input for speech is chronically sub-optimal?



Cochlear implants (CI) are neuroprostheses that allow **deaf people** to perceive sounds. However, the **encoding of speech sounds is suboptimal** [5,6]. Therefore, for CI users, **visual mouth cues and predictability** may be particularly relevant to comprehend speech.

RESEARCH QUESTIONS

Does **visual information about the speaker's mouth** interact with **predictability**?
Are these sources of information **differently exploited by CI users** to compensate for a suboptimal speech input?

METHODS

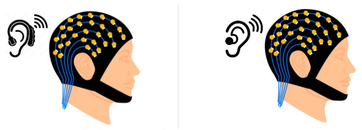
PARTICIPANTS

COCHLEAR IMPLANT (CI) USERS
N = 18
Age = 22.35 (10.8)
Age of implantation = 8.76 (11.5)
Early (≤ 3 y.o.)
N = 9
Age = 17.38 (5.13)
Age of implantation = 1.88 (0.64)
Late (≥ 5 y.o.)
N = 9
Age = 26.78 (12.79)
Age of implantation = 14.89 (13.21)
Side of CI
Unilateral right (7), unilateral left (3), bilateral (8)

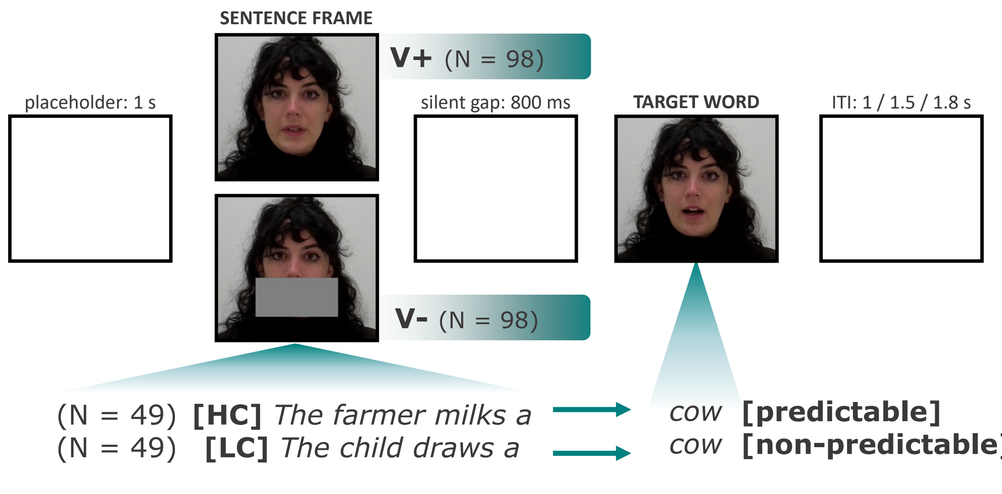
NORMAL HEARING (NH) CONTROLS
N = 18
Age = 21.39 (10.46)

EEG SESSION: AUDIOVISUAL SPEECH COMPREHENSION TASK

CI (N=18) **NH (N=18)**



(61 active channels)



LANGUAGE EVALUATION TASKS

PRODUCTION
- Semantic and phonological fluency
- Sentence generation

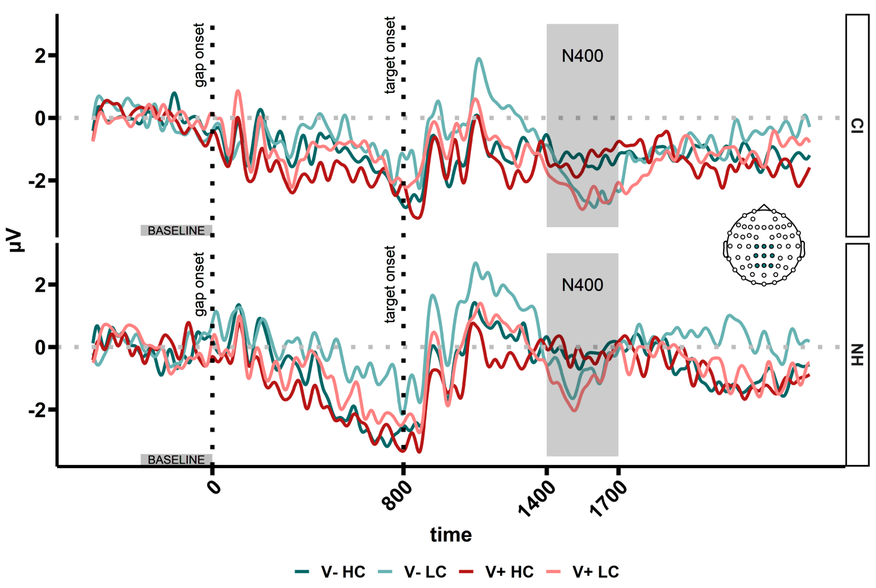
COMPREHENSION
- Lexical decision
- Sentence-picture matching

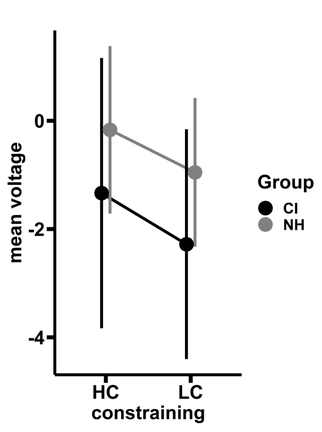
EXPLORATORY ERP RESULTS

CI *vs* NH

MAIN EFFECT OF CONSTRAINT
β = 0.43, t = -4.09, p < 0.001

MAIN EFFECT OF GROUP
β = -0.62, t = -2.15, p < 0.039

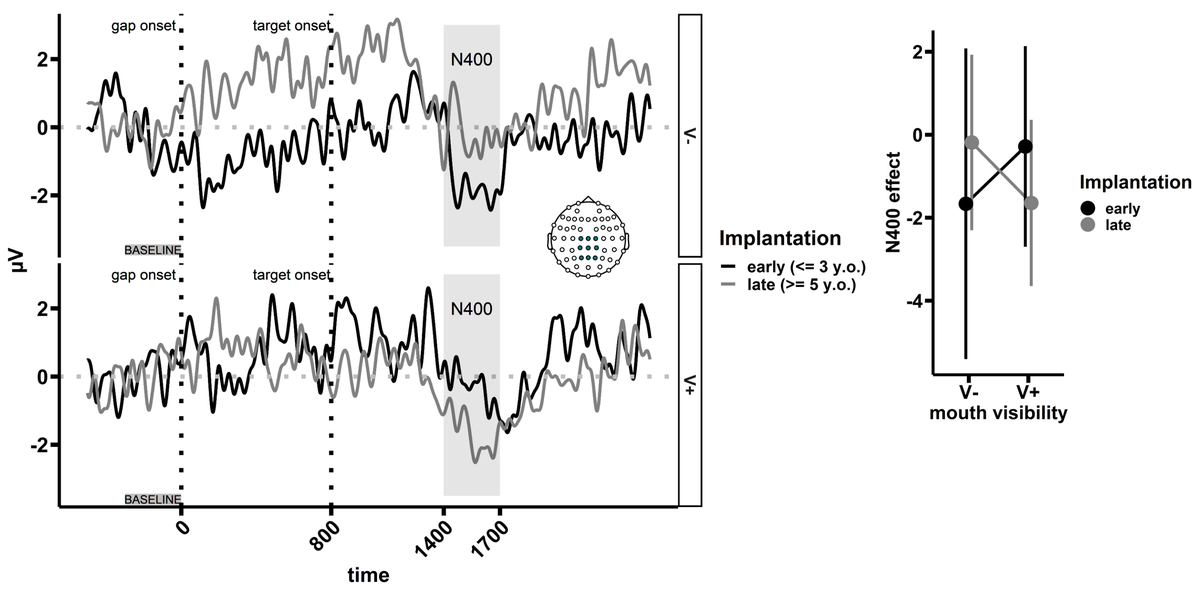




(voltage averaged across mouth visibility conditions)

N400 EFFECT IN CI USERS: EARLY *vs* LATE IMPLANTATION, V+ *vs* V-

INTERACTION VISIBILITY × STAGE OF IMPLANTATION
β = -0.71, t = -2.27, p = 0.039, 95%CI [-1.32 -0.1]



PRELIMINARY OBSERVATIONS

1. Consistent effect of constraint (more variability in CI than in NH)
2. No observable effect of mouth visibility
3. No observable differences between CI and NH
4. The time-window of the effect in CI seems more extended than NH: can we quantify N400 effects with other measures instead of mean amplitude to unveil possible group differences?
5. Processes reflected in the N400 may have different latencies in CI vs NH
6. Interaction between mouth visibility and early/late implantation in CI users: in late implanted the effect emerges only when the mouth was visible (during the previous sentence, not the word!), while it is the opposite in late implanted CI.

SOME ISSUES AND NEXT STEPS

1. Heterogeneous CI group
2. Late N400 window: great variability of speech and mouth onset in the video stimuli.
3. Baseline selection: differences between conditions during the gap (no good baseline), but distant baselines lead to slow fluctuations later in time.
4. CI users have fewer channels: TP-P-PO channels are too close to CI
5. ERP at the gap: slow anticipatory wave
6. Time-frequency during the gap: modulations of alpha-beta bands during prediction.
7. Source estimation of frequency-specific activity in regions of interest. Problem: effect on estimation of having fewer channels in one group. How can we avoid introducing biases when comparing groups?

REFERENCES

[1] Kuperberg, G. R., & Jaeger, T. F. (2016). *Language, Cognition and Neuroscience* 31(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299
[2] Pickering, M. J., & Gambi, C. (2018). *Psychological Bulletin*, 144(10), 1002–1044. https://doi.org/10.1037/bul0000158
[3] Sohoglu, E., & Davis, M. H. (2016). *PNAS* 113(12), E1747–E1756. https://doi.org/10.1073/pnas.1523266113
[4] Peelle, J. E., & Sommers, M. S. (2015). *Cortex* 68, 169–181. https://doi.org/10.1016/J.CORTEX.2015.03.006
[5] Pisoni, D. B. (2005). In D. B. Pisoni & R. E. Remez (Eds.), The Handbook of Speech Perception (pp. 494–523). Blackwell Publishing.
[6] Macherey, O., & Carlyon, R. P. (2014). *Current Biology* 24(18), R878–R884. https://doi.org/10.1016/J.CUB.2014.06.053

FUNDING

This research has been funded by the Italian Ministry of University and Research, PRIN project n. 20177894ZH, "*The role of cochlear implantation and bimodal bilingualism in early deafness: a window into the neurofunctional mechanisms of human language*" to FP (local PI).