Cortical tracking of prosodic and statistical regularities in artificial speech

Lorenzo Titone[1], Burkhard Maess[1] & Lars Meyer[1,2]

[1]*Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, DE*
[2]*Clinic for Phoniatrics and Pedaudiology, University Hospital Münster, DE*

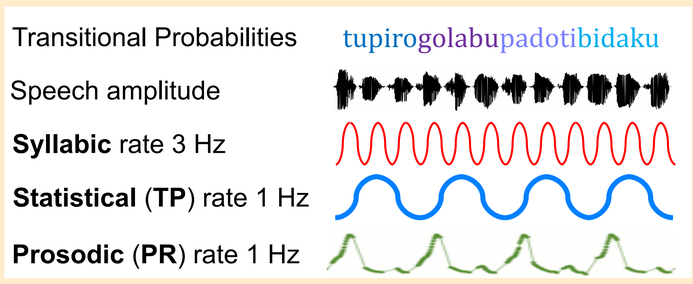
titone@cbs.mpg.de

Introduction

Rhythms of Speech and Language

- Prosody (PR): Exogenous acoustic cue for chunking speech [1].
- Statistical dependencies: Aid chunking and word learning [2].
- **What neural circuits track prosodic and statistical cues?**
- *Neural Frequency Tagging* (NFT): Tracking rhythmic patterns of transitional probabilities (TPs) across neighboring syllables [3-7].
- **Does prosody interact with TP tracking and learning?**

Neural Frequency Tagging



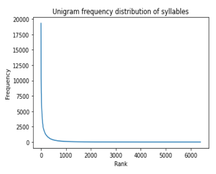
Hypotheses

- Tacking TPs enables word learning; prosody facilitates learning.
- Tracking at the chunk rate (1.1 Hz) increases with PR or TP.
- Tracking at the syllable rate (3.3 Hz) decreases with PR or TP.
- Chunk tracking index (CTI: 1.1 / 3.3 Hz) increases with time.
- Top-down modulations (TP) vs. bottom-up entrainment (PR).
- N400m amplitude increases in TP+ for part-words vs. words.

Methods

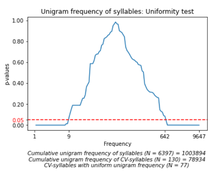
Artificial language with Rhythmicity Control (ARC)

Linguistic control: Frequency distributions

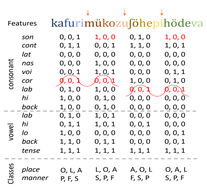


⚠ Language statistics bias new artificial word learning [6].

✔ **ARC** uses uni/bi/tri-grams w/ uniform probability in corpus

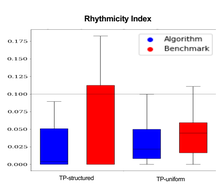


Phonological control: Rhythmicity index (RI)

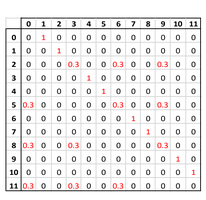


⚠ Phonological features can arrange rhythmically [3, 7].

✔ **ARC** controls phonotactics + RI of phonological features

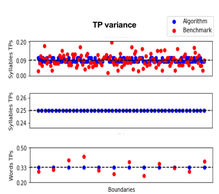


Statistical control: Pseudo-random-walk

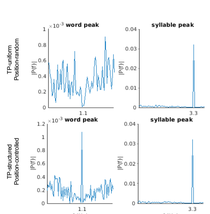


⚠ TP variance at boundaries can bias statistical learning [8].

✔ **ARC** generates streams w/ precise and stationary TPs.

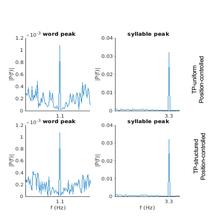


Acoustic control: Envelope spectra

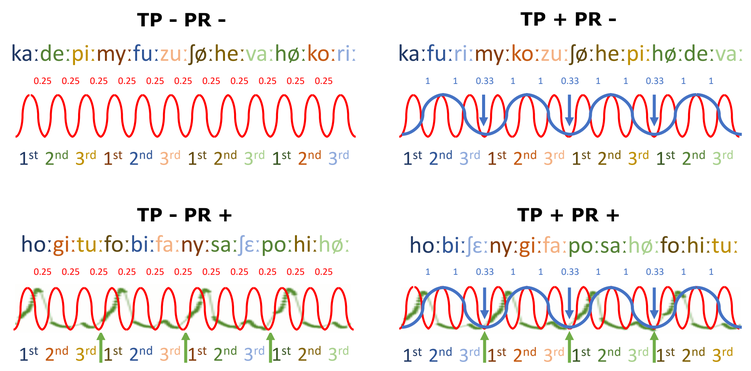


⚠ Acoustic regularities may create rhythms [5, 7].

✔ **ARC** removes spectral differences (position control)



Exposure Phase



- 2 × 2 within subj design (N = 32), orthogonal manipulation of:
  - Statistical rhythm: TP rhythmic (blue) [+] vs. uniform [-]
  - Prosodic rhythm: PR: rhythmic (green) [+] vs. absent [-]
- 2 × 3 min exposure / condition (balanced presentation order)
- 2 lexicons (counterbalanced for PR+ and PR-)

Analyses

- Power: Normalize spectra using neighboring frequencies as baseline to identify significant peaks
- Inter-trial phase coherence (ITPC):
  - ITPC: 1 trial = 1 trisyllabic chunk;
  - Computed over both exposure blocks or by block (3 min.)
- Chunk tracking index (CTI) :
  - CTI = ITPC (1.1 HZ) / ITPC (3.3 Hz)

Testing phase

- 2-alternative forced choice (2-AFC): explicit learning (n=48)
  - Purple: example of artificial word from lexicon inventory.
  - Orange: example of part-word "violating" artificial word.

(Pre-)propocessing

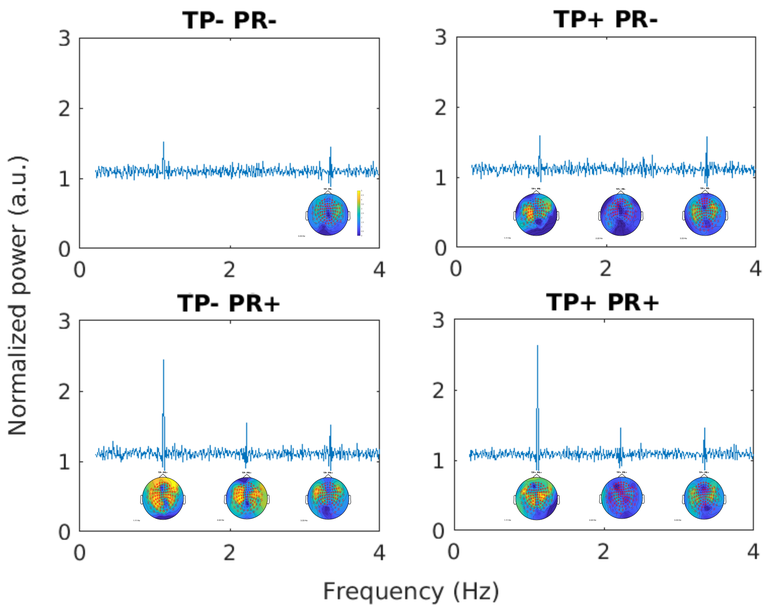
- Bandpass filter (0.1-30 Hz) → Gradiometers transformation: 510 magnetometers → Artifacts rejections: PCA-ICA + visual
- 2 × 2 ANOVA → one-tail cluster-based permutations → main effects (TP, PR) and interaction effects (TP × PR)
- TP x PR = [(TP+ PR+) - (TP+ PR-)] - [(TP- PR+) - (TP- PR-)]

Analyses

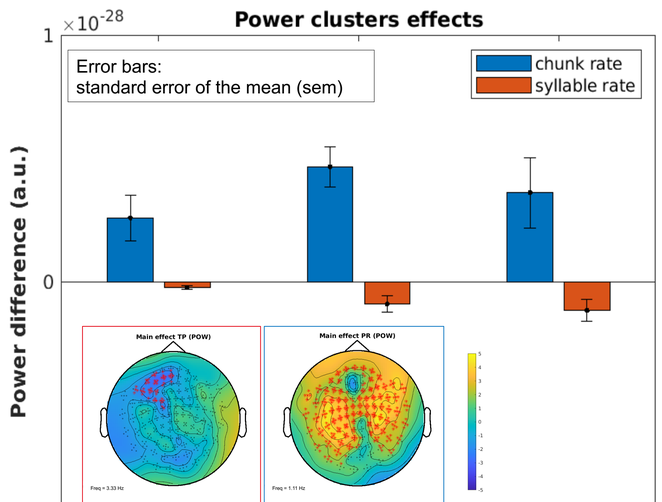
- Event Related Field (ERF):
  - N400m amplitude (part-words – words) difference.
- Behavioral accuracy (2-AFC task, n = 48 trials)
  - One sample t-tests vs chance (50%) within condition
  - Paired t-tests across conditions
  - Model comparison (mixed effect models)

Results

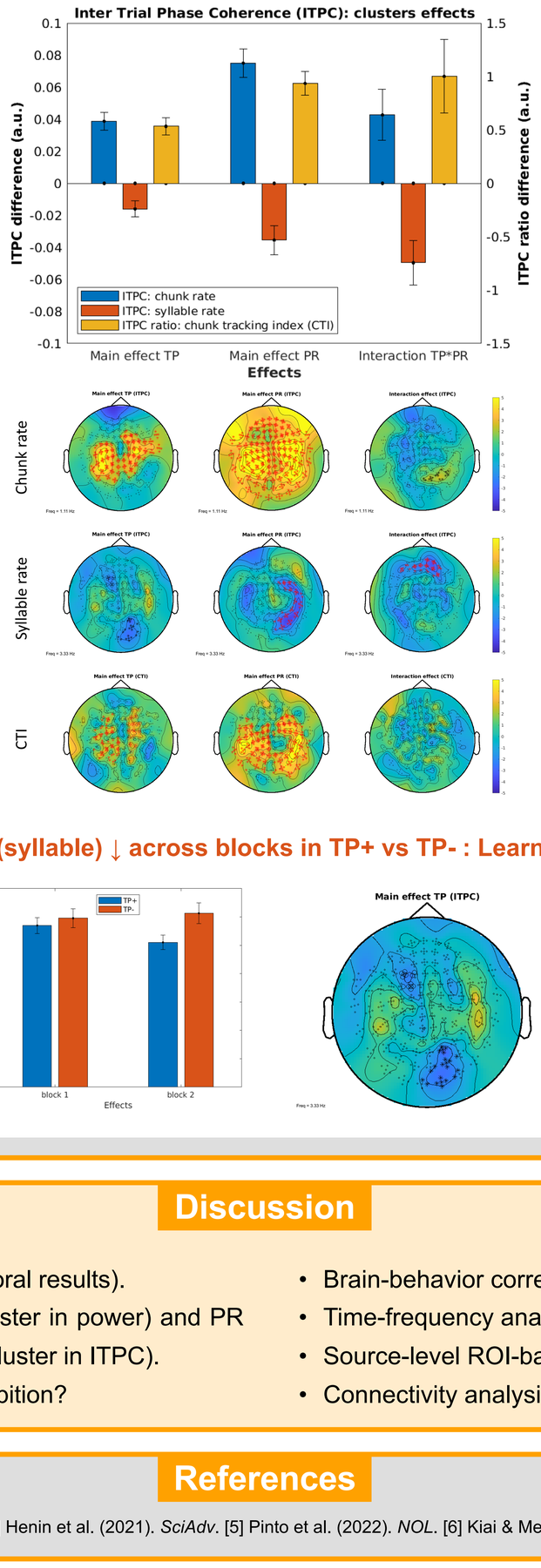
Power spectra: tracking of syllables and chunks



Main effects of power for TP (↓ syllable) and PR (↑ chunk)

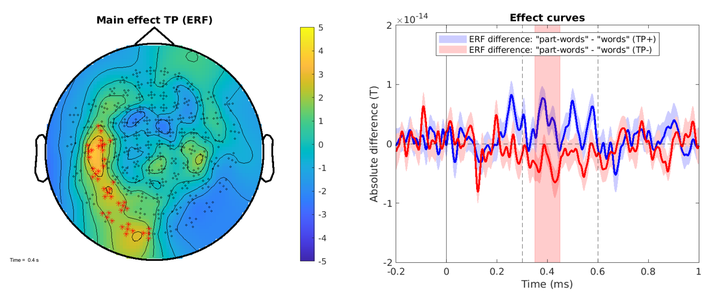


Distinct ITPC clusters show effects of TP and PR

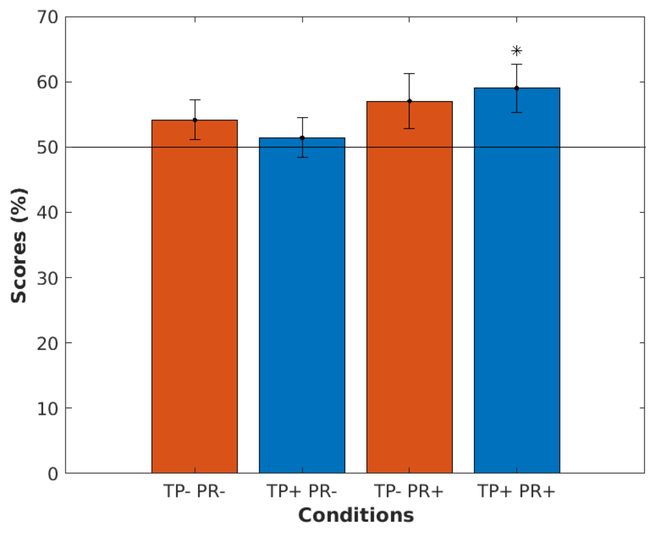


ITPC (syllable) ↓ across blocks in TP+ vs TP- : Learning?

Main effect of TP: N400m amplitude ↓ (part-words - words)



Behavioral accuracy: explicit learning in TP+ PR+



Binomial mixed effect models

- Main effect of PR; no TP; no PR × TP.
- Condition order contributes to explaining performance.
  - Three-way interaction (PR × TP × Condition order).
  - ↑ facilitatory effect of PR on learning when the TP+ PR+ condition was presented in the first vs. last block.

Discussion

Summary

- Word learning relies on both statistical regularities and prosody (behavioral results).
- Distinct networks show reduced syllable tracking for TP (left frontal cluster in power) and PR (right temporal cluster in ITPC) and for the TP × PR interaction (frontal cluster in ITPC).
- ITPC at the syllable rate decreases over blocks in TP+: learning via inhibition?

Future directions

- Brain-behavior correlation: slope of ITPC over time ↔ N400m reduction ↔ explicit learning?
- Time-frequency analysis of evoked responses: post-stim differences in the alpha-beta range?
- Source-level ROI-based analyses: left fronto-temporal network (TP); right temporal (PR)?
- Connectivity analysis for deeper investigation of TP × PR effect: frontal source of inhibition?

References

[1] Rimmele et al. (2021). *Eneuro.* [2] Saffran et al. (1996). *Science*. [3] Batterink et al. (2017). *Cortex*. [4] Henin et al. (2021). *SciAdv*. [5] Pinto et al. (2022). *NOL*. [6] Kiai & Melloni (2021). *bioRxiv*. [7] Chen et al. (2020). *Neuropsychologia*. [8] Hasson (2017). *Phil Trans R Soc B.*