Does cortical tracking signify linguistic hierarchical structures in the visual domain?

Likan Zhan [1*] Yanan Xue [1] Peng Zhou [2] Yuan Xie [3]

[1]Beijing Language and Culture University, [*]zhanlikan@hotmail.com [2]Zhejiang University [3]Tsinghua University

Introduction

Pioneered by Ding et al. (2016), the frequency tagging paradigms have been widely used to study the hierarchical structure of human language: Participants' electrophysiological responses are recorded while they are listening to a continuous series of syllables. The frequency of the occurrence of the complex linguistic units are found to have a corresponding peak in the frequency domain of the recorded M/EEG signals. These observations are regarded as the evidence that the hierarchical structure of human language could be encoded by the peaks in the spectrum of the electrophysiological signals.

In addition to the auditory domain, human language can exemplify its hierarchical structure in other domains, such as writing system, or sign language. It remains unclear whether these tagging relations observed in the auditory domain could be extended to other domains. To address this question, the present study investigated whether the tagging relation observed in the auditory domain in prior research could also be observed in the visual domain.

Experiment

We replicated Ding et al. (2016)'s first experiment using visually presented Chinese characters. Participants' electrophysiological responses were recorded with EEG while they were watching a series of Chinese characters. In line with Ding et al. (2016), three kinds of stimuli (Figure 1), N2V2, V2N2, and V1N3, were created that had the following three properties:

(a). In all three conditions, each character occurred with the frequency of 4Hz, as each character was presented with a length of 250ms;

(b). In all three conditions, each phrase occurred with a frequency of 1Hz, as every four characters formed a phrase with the length of 1s (= 250ms X 4). The four characters in N2V2 merged into a full sentence, while the four characters in the latter two conditions only merged into a verb phrase.

(c). In conditions N2V2 and V2N2, each two-syllable word (500ms = 250ms X 2) occurred with the frequency of 2Hz, whereas no such word occurred in condition V1N3.

Results

The EEG spectrums of forty healthy native Chinese speakers were shown in Figure 2. A 4Hz peak ($p < .001$) was observed in all three conditions, which is in consistent with the results observed in the auditory domain in previous research, confirming the validity of our results. Excluding the 4Hz peak corresponding to the frequency of the stimuli presentation, only one peak corresponding to the mental merging of the presented stimuli was observed in each condition.

(a). In condition N2V2 (Figure 2, panel A), the only peak was 1Hz ($p < .01$), corresponding to the frequency of the sentences. In condition V1N3, the only peak was also 1Hz ($p < .01$), corresponding to the frequency of the verb phrases.

(b). In condition V2N2 (Figure 2, panel B), the only peak was 2Hz, corresponding to the frequency of the two-syllable words. Different from the effects observed in the auditory domain, the 2Hz peak was not observed in N2V2, and the 1Hz peak was not observed in V2N2 condition.

Discussion

Our results cannot be fully explained by the viewpoint that the peaks observed in the M/EEG spectrum are tagged to the abstract hierarchical structure of human language. More specifically, in N2V2 condition, the two-syllable words were presented in the frequency of 2Hz, but no such peak was found in the EEG spectrum; and in V2N2, the verb phrases were presented in the frequency of 1Hz, but no such peak was observed in the EEG spectrum.

One possible explanation for the discrepancy between the auditory domain and the visual domain could be that in Chinese writing system but not in the auditory domain punctuation markers are typically used to facilitate the construction of hierarchical structure. We wish to note, however, that no punctuation markers were used in our study, which excluded this possibility due to the effect of punctuation markers. A straightforward way to explain our results is that the participants simply chunked the stimuli in the most memory-efficient way as they processed visual stimuli, leading to one peak per condition.

Reference

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016).Cortical tracking of hierarchical linguistic structures in connected speech. Nature Neuroscience, 19(1), 158–164. https://doi.org/10.1038/nn.4186

Figure 1. Test stimuli

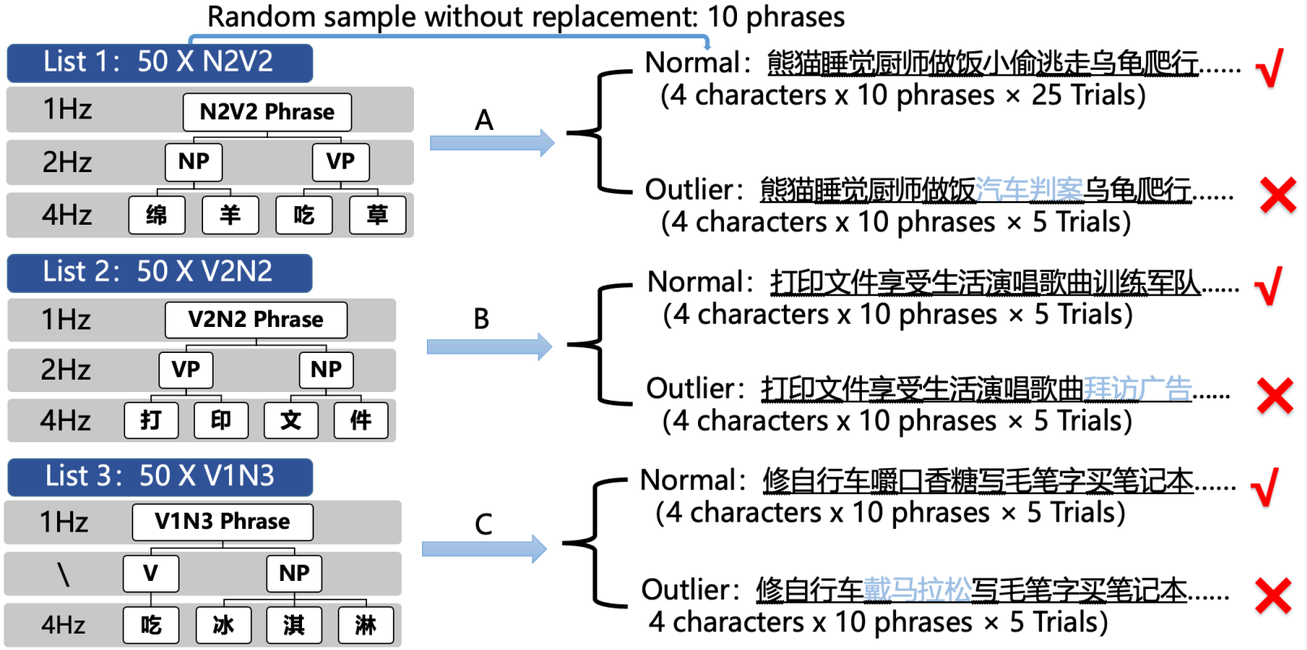

Figure 2. Ding et al. (2016)'s results (left) and our results (right)