

## RESEARCH ARTICLE

# Chunk boundaries disrupt dependency processing in an AG: Reconciling incremental processing and discrete sampling

Chia-Wen Lo<sup>1</sup> <sup>\*</sup>, Lars Meyer<sup>1,2</sup> 

**1** Research Group Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, **2** University Clinic Münster, Münster, Germany

 These authors contributed equally to this work.

\* [lo@cbs.mpg.de](mailto:lo@cbs.mpg.de)



## Abstract

Language is rooted in our ability to compose: We link words together, fusing their meanings. Links are not limited to neighboring words but often span intervening words. The ability to process these non-adjacent dependencies (NADs) conflicts with the brain's sampling of speech: We consume speech in chunks that are limited in time, containing only a limited number of words. It is unknown how we link words together that belong to separate chunks. Here, we report that we cannot—at least not so well. In our electroencephalography (EEG) study, 37 human listeners learned chunks and dependencies from an artificial grammar (AG) composed of syllables. Multi-syllable chunks to be learned were equal-sized, allowing us to employ a frequency-tagging approach. On top of chunks, syllable streams contained NADs that were either confined to a single chunk or crossed a chunk boundary. Frequency analyses of the EEG revealed a spectral peak at the chunk rate, showing that participants learned the chunks. NADs that cross boundaries were associated with smaller electrophysiological responses than within-chunk NADs. This shows that NADs are processed readily when they are confined to the same chunk, but not as well when crossing a chunk boundary. Our findings help to reconcile the classical notion that language is processed incrementally with recent evidence for discrete perceptual sampling of speech. This has implications for language acquisition and processing as well as for the general view of syntax in human language.

## OPEN ACCESS

**Citation:** Lo C-W, Meyer L (2024) Chunk boundaries disrupt dependency processing in an AG: Reconciling incremental processing and discrete sampling. PLoS ONE 19(6): e0305333. <https://doi.org/10.1371/journal.pone.0305333>

**Editor:** Hung Thanh Bui, Industrial University of Ho Chi Minh City, VIET NAM

**Received:** March 18, 2024

**Accepted:** May 29, 2024

**Published:** June 18, 2024

**Copyright:** © 2024 Lo, Meyer. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

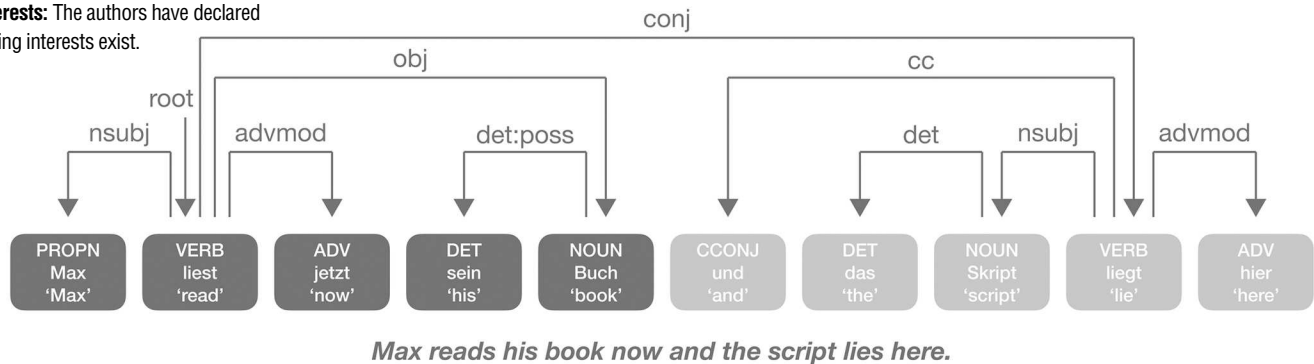
**Data Availability Statement:** Data cannot be shared publicly because of ethics regulations. Data are available from the MPI-CBS Institutional Data Access (MPI-CBS Data security co-ordinator Dr. Jöran Lepsien [[lepsien@cbs.mpg.de](mailto:lepsien@cbs.mpg.de)]) for researchers who meet the criteria for access to confidential data.

**Funding:** Research was supported by the Max Planck Society through the award of the Max Planck Research Group Language Cycles to Lars Meyer.

## Introduction

Language allows us to generate and process a huge, possibly unbounded number of word combinations. Not only can we link neighboring words, but also words that are separated by intervening words. Such non-adjacent dependencies (NADs) are attested across languages. In cognitive science, NAD processing is considered to be a working memory task: The first element of an NAD must be memorized until the second element arrives [1–4]. For example, in the sentence *John saw himself*, *John* must be held until the pronoun *himself*, which refers back to him. Behavioral and electrophysiological studies have shown that infants and adults learn

**Competing interests:** The authors have declared that no competing interests exist.



**Fig 1. A German example sentence annotated with dependency grammar and chunks.** Words with deep and light gray indicate different chunks. Most dependencies form locally within a chunk.

<https://doi.org/10.1371/journal.pone.0305333.g001>

NADs from statistical regularities—our brains monitor the co-occurrence of the first and second elements of NADs [5–11]. The tracking of all possible statistical regularities and the potential grouping of words may be affected by various cues such as transitional probabilities between elements [10, 12–16], prosodic properties of speech [17–20], distributional properties of elements [21–24], and other cues such as function words or morphology for constructing well-formed dependencies [18, 24, 25]. Both adults and infants can learn short NADs in the form of “AXB”, which forms an arbitrary dependency between A and B, interrupted by X (e.g. [7, 10, 23]). While the length of NADs in real languages is not limited to such a short dependency, it is widely accepted that intervening elements disrupt NAD processing. For example, Bock and Miller [26] found that the intervening noun phrase between subject and verb dependency can lead to more errors in a production task that adults tend to produce errors such as *\*The bridge to the islands were crowded*. Corpus analyses revealed that the dependency lengths are shorter than the random baseline across 37 languages [27] and a preference for minimization of the distance between two related syntactic elements was observed [28], suggesting that people have a strong bias toward short dependencies (see also [29]).

Our ability to learn and process NADs conflicts with evidence that our brain samples speech in larger chunks that contain only a limited number of words. Thus, in principle, chunk endings may be cutting into NADs, as exemplified by Fig 1. In an earlier study, [30] presented long sentences with numerous complex NADs (e.g. *boys who chase dogs see girls*.) to a connectionist network model. The model failed to capture the NADs when the whole sentence was presented at one time. However, when the sentences were presented within a limited processing window of 3–4 words or so chunk-by-chunk, the NAD patterns can be successfully captured by the model. Human chunk-by-chunk sampling is thought to arise from memory limitations: To counteract decay in working memory, we integrate information within limited time windows of up to 3 seconds [31–33]. For language specifically, memory may be restricted to about 2 seconds when articulating a word sequence [34]. In psycholinguistics, a processing time window of six words has been proposed [35], roughly equaling 2.5 seconds when assuming a speech rate of 150 words per minute [36]. Current neuroscientific work suggests that such temporal limitations may arise from a neurobiological constraint: the wavelength of slow-frequency neural oscillations. For instance, a recent study found electrophysiological activity in the delta band (< 4 Hz) to align to phrases and sentences [37]. Specifically, phase angles of delta-band activity predict the boundaries of multi-word chunks [38], particularly when these exhibit a duration of 2.7 seconds [39].

But if we sample and process one chunk at a time, how can we link words that belong to separate chunks [40]—or can we? In the following, we present EEG results that show the chunk boundaries appear to block NAD processing and possibly learning. We designed an AG, composed of syllable sequences with equal duration. This allowed us to employ the frequency-tagging paradigm to assess whether people sample continuous speech into the desired size of units. Participants learned 6-syllable chunks from sequences, based on transitional probabilities and an additional short pause and the chunk boundary. Syllable streams either contain the NADs within a 6-syllable chunk or across two chunks (see Fig 2). Spectral analyses of EEG indeed revealed a spectral peak at the chunk rate, suggesting that people are able to learn the chunks. Crucially, smaller electrophysiological responses were observed in NADs that traversed chunk boundaries, compared to the NADs within a chunk. This suggests that NADs across chunk boundaries are harder to process than NADs within a chunk. Our findings can help to bridge the gap between incremental processing of acoustic signals, such as speech; moreover, they strengthen previously proposed links between periodic neurobiological activity and the chunking of speech.

## Materials and methods

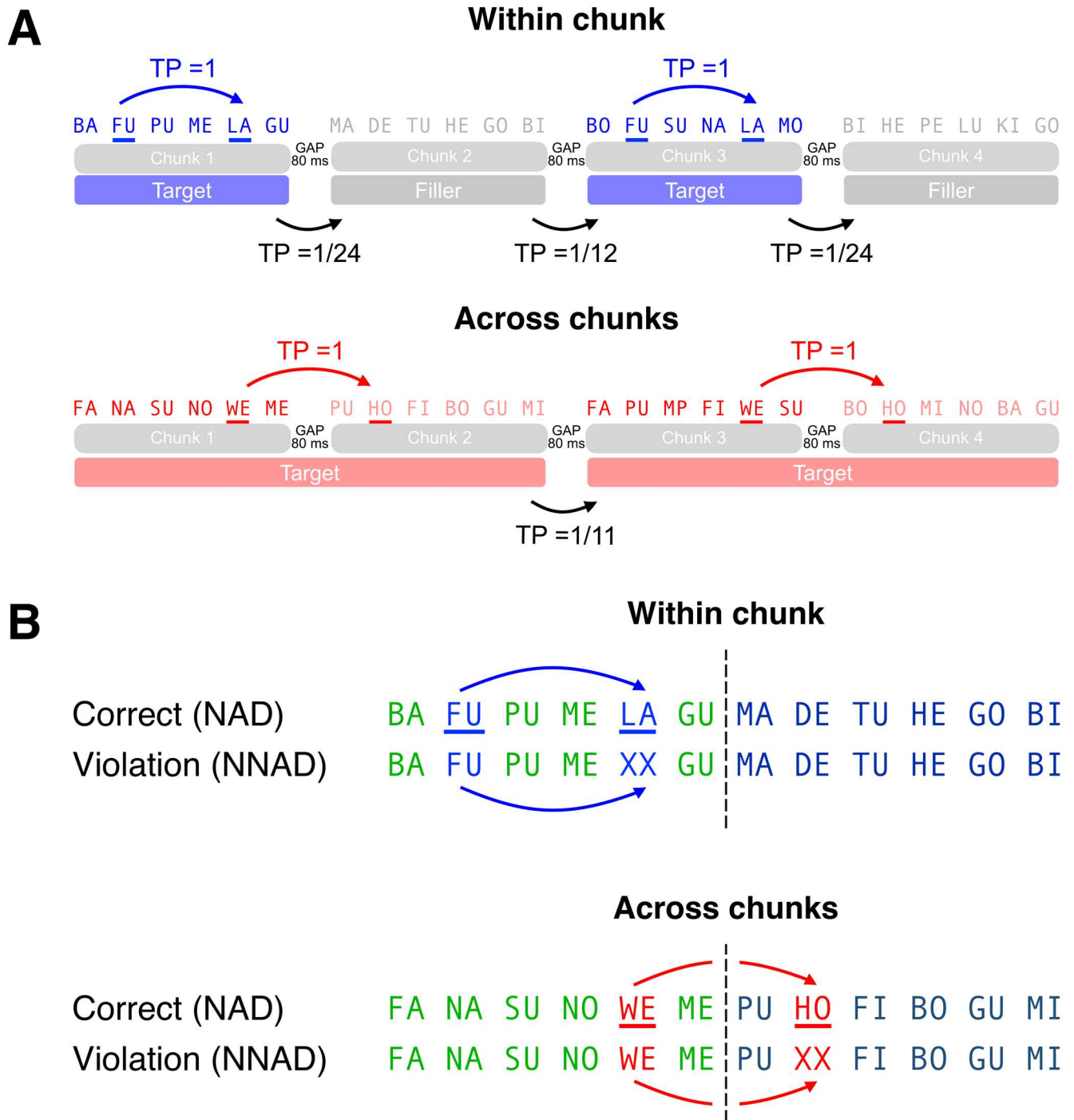
To test the hypothesis of whether an NAD across chunks would be harder to process, the current study applies an AG learning paradigm combined with frequency tagging (see Fig 3). By using frequency tagging, we can see whether participants acquire chunks during the learning phase. Thus the learning phase serves two purposes: first, participants learn NADs within and across chunks; second, participants learn six syllables as a chunk. If participants are able to learn six syllables as a chunk, a peak at the chunk frequency should be observed. Then during the test phase, participants listen to trials with either correct NADs or incorrect NADs in both within-chunk and across-chunk conditions. If NAD processing is indeed harder across chunk boundaries, we expect the difference of amplitude from correct and incorrect elements that complete NADs would be smaller in the across-chunk condition. The overall experimental procedure is demonstrated in Fig 4.

## Participants

Thirty-seven German native speakers (18 females, 19 males) between the ages of 18 and 34 (mean = 24.5) participated in the experiment (recruitment period: 01/03/2022–31/07/2022). They were all right-handed and had normal hearing. They self-reported that they did not have any neurological disorders. They gave written informed consent before participation and were reimbursed for their time (9 Euros per hour). Data from four participants were excluded from the analysis due to technical recording issues and two were removed due to poor accuracy in the learning phase (< 50% accurate). Thus, data from 31 participants (16 female, 15 males) were included in the final analysis. The study was approved by the local ethics committee of the University of Leipzig (file 060/17-ek).

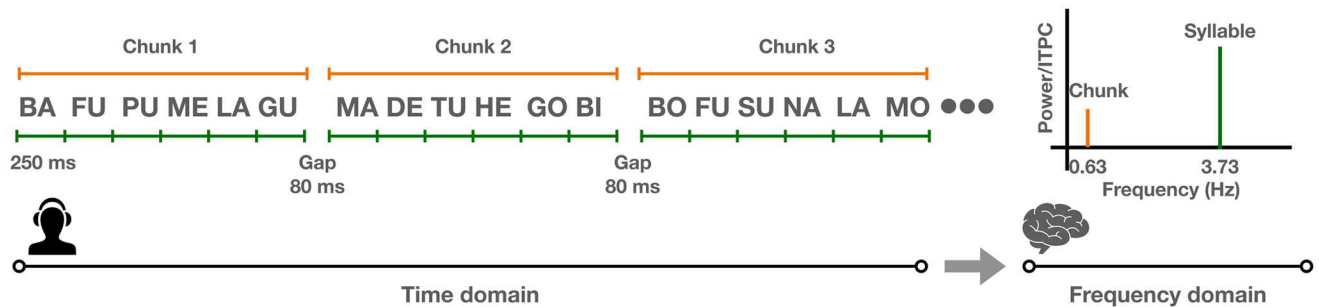
## Materials

German syllables (see Tables 1 and 2) were recorded individually as isochronous speech from Google Cloud Text-to-Speech (Male, de-DE-Wavenet-B). Complete stimuli can be found here: <https://github.com/chiawenl/NAD-exp>. Two native German speakers confirmed that syllables sounded naturally. Each syllable was adjusted to a duration of 240 ms and a 10-ms silence was appended to each syllable using the Praat vocal toolkit [41] in Praat [42] and customized Python scripts. Thus, each syllable lasts 250 ms. Six syllables were concatenated together to form a chunk. After each chunk, an extra 80-ms silence was appended to indicate a chunk



**Fig 2. Experimental conditions.** (A) Participants listened to both within-chunk and across-chunks in the learning phase. Target chunks in the within-chunk condition were intervened with the filler chunks to have the same amount of NADs as the across-chunk condition. The transitional probability between dependency is 1.; (B) Participants listened to both correct (NAD) and violation (NNAD) for both within-chunk and across-chunk conditions in the test phase.

<https://doi.org/10.1371/journal.pone.0305333.g002>

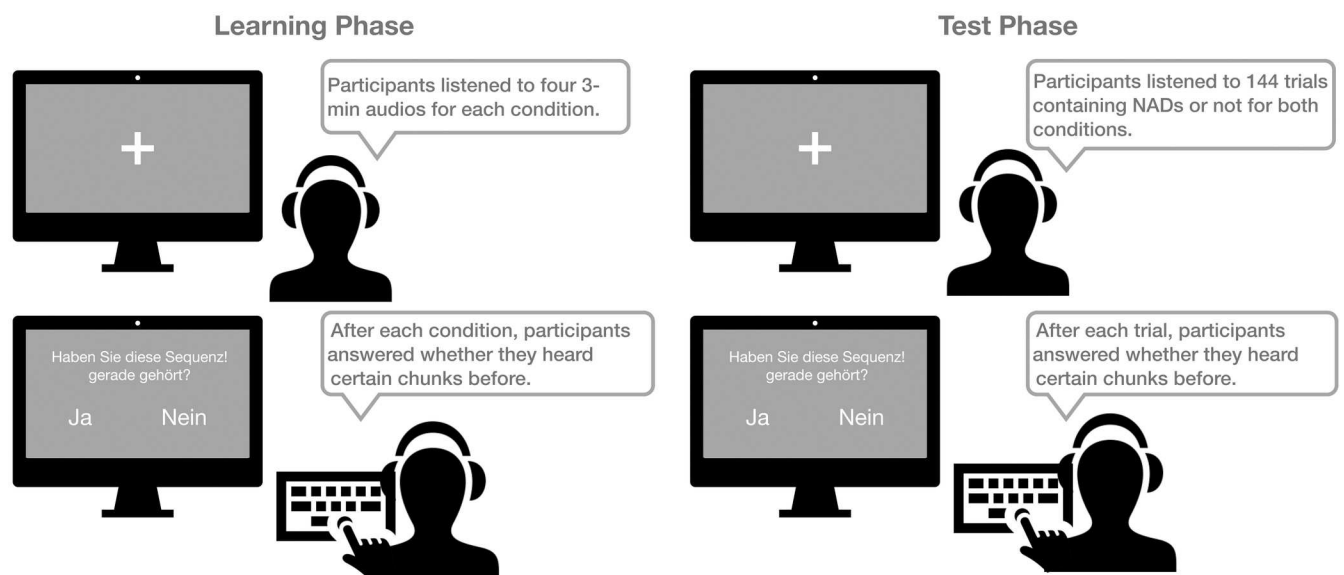


**Fig 3. Frequency-tagging paradigm.** Participants listened to a sequence of syllables. In the time domain (left), each syllable lasts 250 ms; each chunk consists of 6 syllables. An 80-ms silence was inserted between chunks. In the frequency domain (right), this leads to an according frequency of syllable occurrence of 3.73 Hz, while 6-syllable chunks occurred with a frequency of 0.63 Hz. When participants' brains track both syllables and chunks, we expect to observe peaks in the EEG power / ITPC (= inter-trial phase coherence) spectra at both syllable and chunk rates.

<https://doi.org/10.1371/journal.pone.0305333.g003>

boundary. These boundary silences were included to ensure that participants would succeed in learning the 6-syllabic chunks. It has been shown that prosodic cues aid both the formation of an NAD and the sampling of chunks [43, 44]. Critically, this intentional redundancy of transitional probability and pause duration would not affect the interpretability of any differences in NAD processing hypothesized to surface at the second element of the NAD (see Fig 2).

Twenty-four target chunks for each condition (12 for NAD and 12 for violation) in one list were created for each condition. To have equal length for both within-chunk and across-chunk conditions, 24 filler chunks constructed from filler syllables (Table 1) were created and inserted between target chunks in the within-condition. The same 12 target syllables and 12 filler syllables were used across participants. Four lists were created; each list includes two dependency pairs. One participant learned one dependency pair in the within-chunk condition and the other dependency pair in the across-chunk condition. Four dependency pairs



**Fig 4. Experimental procedure.** The whole experiment includes two phases: the learning phase and the test phase. During the learning phase, participants listen to four 3-minute audios for each condition and then answer 40 questions. During the test phase, participants answer a question after each trial.

<https://doi.org/10.1371/journal.pone.0305333.g004>

**Table 1. Syllables for target and filler chunks.**

Target syllable		Filler syllable	
1	BA	1	BI
2	BO	2	BU
3	FA	3	DE
4	FI	4	GO
5	GU	5	HE
6	ME	6	KI
7	MI	7	LE
8	MO	8	LI
9	NA	9	LU
10	NO	10	MA
11	PU	11	PE
12	SU	12	TU

<https://doi.org/10.1371/journal.pone.0305333.t001>

(Table 2) were balanced across within-chunk/across-chunk conditions. Thus, each syllable in the pairs appears in the second and fifth positions equally.

For the learning phase, two experimental conditions were created—within-chunk and across-chunk. An example stimulus is shown in Fig 2. As for the test phase, four conditions were made—whether the second element of a dependency pair forms the dependency in both within-chunk and across-chunk conditions. The elements of a dependency pair were situated in the second and the fifth positions for the within-chunk condition while they were situated in the fifth position of the first chunk and the second position of the next chunk for the across-chunk condition. This positioning served to address the difficulty of ruling out the possibility that participants of prior studies were merely sensitive to detecting the edge of a chunk, rather than segmenting continuous stream by the formation of NADs [21, 45]: In some prior studies, the second element of the NAD co-occurred with the (prosodic) chunk boundary. Endres and Mehler [21] directly test the edge difference between strings (AXYZB vs. XAYBZ). They found that participants are indeed sensitive to the NADs that occurred at the boundary, compared to the medial positions. The remaining syllables were distributed evenly from the 12 target syllables for each condition.

Four 3-minute audios were created for both conditions for the learning phase. The within-chunk condition included 72 target chunks and 72 filler chunks per audio while the across-chunk condition included 144 chunks that contain a dependency pair across two chunks. Seventy-two filler chunks were inserted between target chunks in the within-chunk condition to balance the amount of NADs for both conditions.

For the test phase, 144 trials that include NADs or violations (NNADs) for both conditions were created. To avoid additional learning, trials with either correct NADs or violations for both conditions were randomly distributed during the test phase. Each list included 36 NAD

**Table 2. Syllables for dependency pairs.**

DEPENDENCY PAIRS		
1	FU	LA
2	TA	PI
3	KE	MÖ
4	WE	HO

<https://doi.org/10.1371/journal.pone.0305333.t002>

trials and 36 violated trials for both conditions. Four target chunks that either include correct NADs or violations were included in one trial. Hence, 4 target chunks and 4 filler chunks were included in the within-chunk condition while 4 target-across-chunk were included in the across-chunk condition. After each trial, a comprehension question about whether a specific chunk in the trial they just heard was assigned. The answers were balanced.

## Procedure

Participants sat comfortably in front of a computer screen in a quiet room. Stimuli were presented using Presentation (Neurobehavioral Systems, Inc., Albany, US). Before the main session, participants were fitted with an electrode cap. Electrolyte gel was applied to minimize impedance below 10 kOhms. The setup took approximately 30 minutes.

The main session included two phases—the learning phase and the test phase (see Fig 4). Participants were instructed to listen to the audio carefully and avoid unnecessary body movement and frequent eye blinking during recording. During the learning phase, participants listened to four 3-minute audios in one condition and answered 40 questions about whether they just heard a specific chunk in the previous audios. Then, participants continued the other three 3-minute audios and 40 questions for the other condition. The order of conditions was counterbalanced. The learning phase took roughly 30 minutes to finish.

During the test phase, participants listened to 144 trials (in six blocks) that included either the corrected NAD or the violations in both conditions. Participants could take a short break after each block. Before the main test, participants had four practice trials to become familiar with the task. After each trial, participants had to answer whether they heard a specific chunk in each trial. The test phase took roughly 40 minutes to finish. After the test phase, participants removed the cap and a debrief of the goal of the experiment was given.

## EEG recording and data analysis

EEG data were recorded at 500 Hz from 63 Ag/AgCl electrodes mounted in an elastic cap (ANT Neuro GmbH, Berlin, DE) with online reference to the left mastoid (A1). Horizontal and vertical eye movements were monitored by the bipolar electrodes placed above and below the right eye and the electrodes placed on the outer canthi. An electrode on the stratum served as the ground.

EEG pre-processing was done by applying the modified Harvard Automated Pre-processing Pipeline [46] with a combination of EEGLAB [47] and FieldTrip [48] functions executed in Matlab (The MathWorks, Inc., Natick, US). Line noise was removed by applying Zapline-plus [49]. Then data were re-referenced offline to the average of the left and the right mastoid electrodes. Bad channels were recognized by the normed joint probability of the average log power and rejected if they were above the threshold of 3 SD (mean number of removed channels = 4.13, SD = 2.62). A high-pass filter of 0.1 was applied (FIR, Hamming windowed, reversed filtering), and then the data were re-referenced to the common average of all electrodes excluding the channels that were marked as bad (see also [50]). To obtain optimal decomposition from independent component analysis (ICA, [51]), a wavelet-enhanced independent component analysis (W-ICA, [46, 52]) before applying ICA to remove large artifacts. ADJUST [53] was applied to detect artifact components based on a set of temporal and spatial features of each component (mean number of removed components = 15.1, SD = 7.92). Data in the learning phase were epoched to 12.64 seconds, resulting in 8 chunks in one trial. Thus 72 trials were analyzed for each condition. For the test phase, the elements of dependency pairs were epoched from 100 ms pre-stimulus to 300 ms post-stimulus and then baseline-corrected. After epoching, FASTER [54] was applied to each channel per epoch to detect artifacts

automatically. Channels were spline interpolated if it is contaminated within each epoch. Then channels that were recognized as bad previously were also interpolated by using surface spline interpolation [55].

As the goal of the spectral analysis during the learning phase was to see whether participants were able to derive 6-syllable chunks, we collapsed and analyzed neuronal synchrony across all trials in both conditions together. The target frequency for the syllable rate is around 3.73 Hz and for the chunk rate is 0.63 Hz. Neuronal synchrony was assessed from Evoked Power (EP) and Inter-trial Phase Coherence (ITPC), following the algorithm defined by [56]. EP (Eq (1)) reflects the power of EEG responses synchronized with speech stimuli in both phase and time.  $X_n(f)$  is the summation of complex-value Fourier coefficient of trials derived from the Discrete Fourier Transform. EP is obtained from  $X_n(f)$  averaged over the total number of trials  $N$ . We computed EP from 0.1 to 10 Hz in increments of 0.079 Hz. The 1/f noise in the power spectrum was normalized by dividing the value at the target frequency from the average of neighboring values within  $\pm 0.5$  Hz via the Eq (2) adopted from [56], where  $w$  represents the neighboring frequency around the target frequency  $f$ . There are many ways to normalize power (e.g. Irregular-resampling auto-spectral analysis (IRASA, [57]), fitting oscillations and one over  $f$  (FOOOF, [58])). The approach for normalization applied here has been shown that yield similar results by applying other normalization (see also [59]). ITPC reflects the phase consistency across trials. ITPC (Eq (3)) is obtained by averaging over the total number of trials  $N$  from the square root of summation of cosine and sine values of phase angles  $\theta_n$  of each complex-value Fourier coefficient. For statistical analysis, normalized EP and ITPC of target frequencies (chunk and syllable rate) were compared with the neighboring 4 frequency bins around the target frequencies.

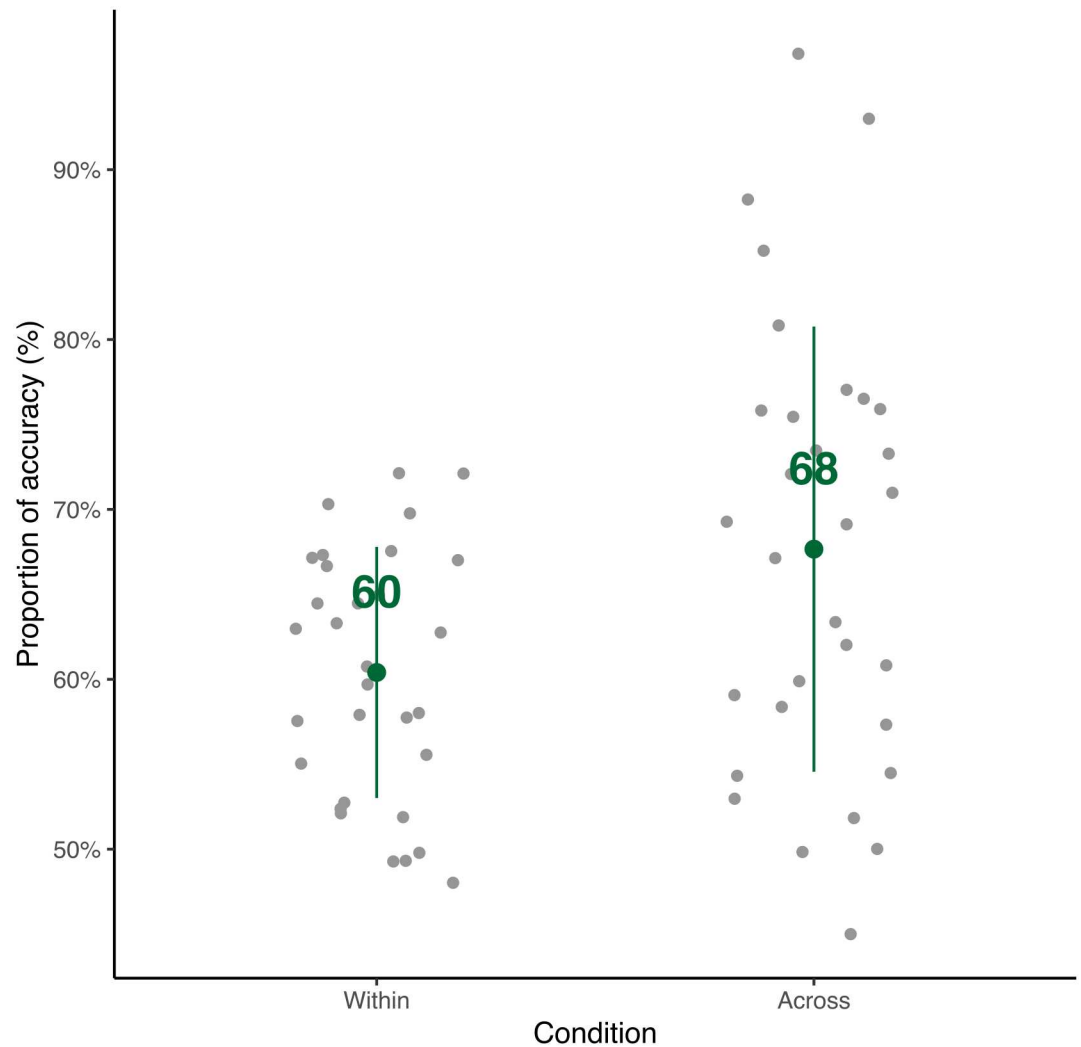
$$EP(f) = \frac{|\sum_n X_n(f)|^2}{N} \quad (1)$$

$$EPn(f) = \frac{E(f)}{\sum_w E(w)}, |w - f| < 0.5 \text{ Hz}, w \neq f \quad (2)$$

$$ITPC(f) = \frac{\sqrt{(\sum_n (\cos\theta_n))^2 + (\sum_n (\sin\theta_n))^2}}{N} \quad (3)$$

To examine whether the NAD across chunks is harder than the NAD within a chunk in the test phase, we analyzed the magnitude of the second element that forms or violates dependency in the event-related component (ERP) with time-locked to the syllable onset. A low-pass filter at 25 Hz (IIR, two-pass filtering, and Hamming windowed, default in Fieldtrip) was applied to the epoched data before group analysis. A non-parametric permutation test [60] was conducted to correct multiple comparisons across all electrodes time-locked to the second element, which forms a correct or incorrect dependency. For each condition, the permutation test was conducted by following these steps: (i) Dependent samples T-statistics were conducted at each time point and electrode, (ii) tests with  $p < 0.05$  were clustered based on spatial-temporal adjacency and their T-statistics were summed by using weighted cluster mass, a method that takes cluster size and intensity into account [61], (iii) Steps (i) and (ii) were repeated 10,000 times by randomly permuting the conditions for each subject, and (iv) clusters with summed statistics that surpassed at least 95% from the permutation test were kept as “statistically significant”.





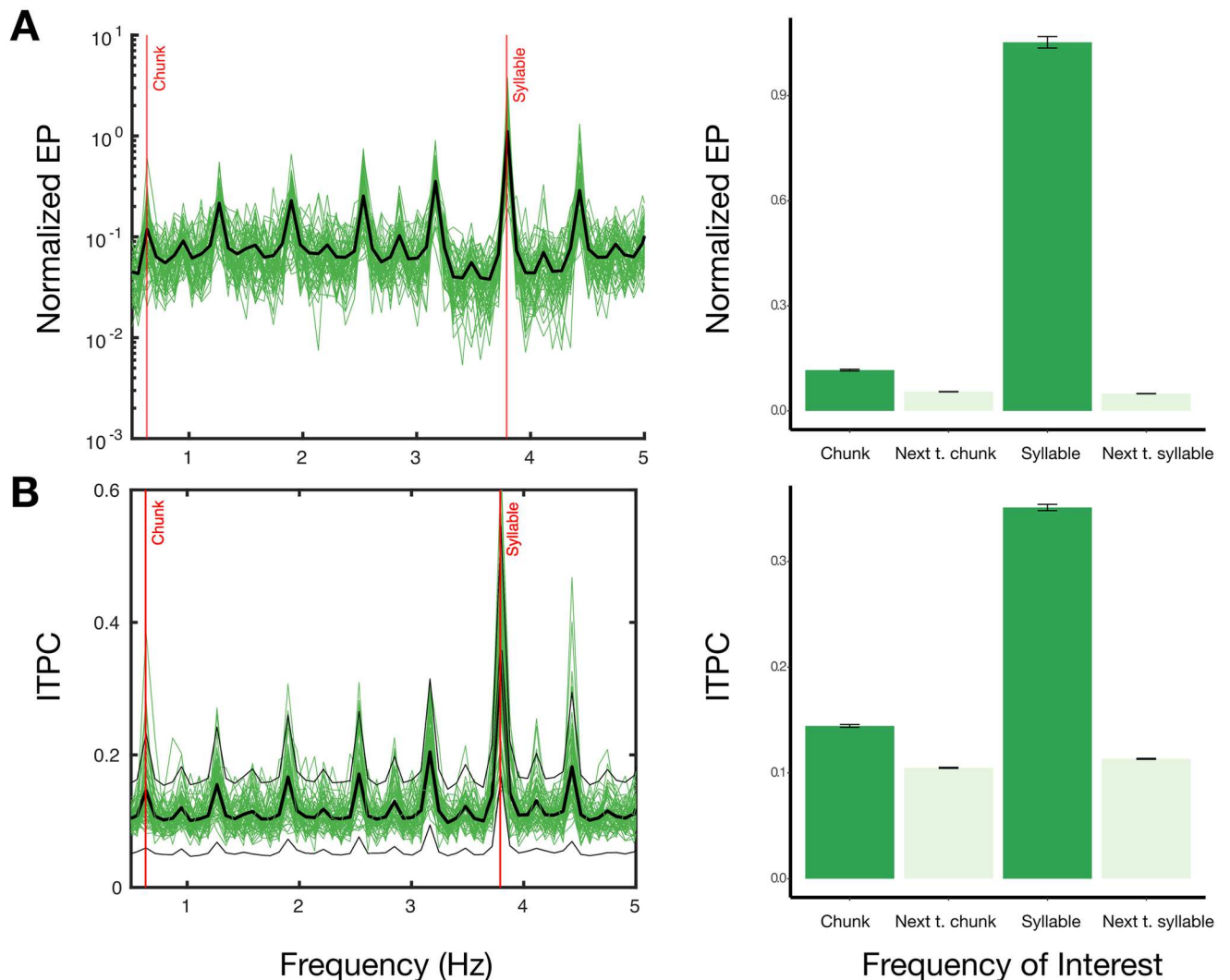
**Fig 5. Behavioral results in the learning phase.** The overall accuracy is 64%. The gray dot indicates the score for each participant. The green dot indicates the mean for each condition and the green line indicates the standard deviation.

<https://doi.org/10.1371/journal.pone.0305333.g005>

## Results

### Chunks are learned from syllable streams

The overall accuracy of the comprehension questions is 64% correct in the learning phase, indicating chunk learning. The accuracy for each condition is shown in Fig 5. The Paired  $t$ -test testing the difference between the two conditions (mean of within-chunk: 60%; mean of across-chunk = 68%) shows that the accuracy in the across-chunk condition is significantly higher than the accuracy in the within-chunk condition ( $t(30) = -2.9, p = 0.007$ ). The higher accuracy of the across-chunk condition might be due to the fillers inserted in the within-chunk condition. Since more novel syllables needed to be learned, hence increased noises and lowered accuracy in the within-chunk condition. To assess neural synchrony, normalized EP and ITPC were computed. Fig 6A shows the power spectrum during the learning phase. We used *lme4* [62] in R [63] to fit a linear mixed effect model (estimated using REML and nlopt-wrap) to test normalized EP and ITPC of target frequencies and neighboring frequencies



**Fig 6. Results of normalized EP and ITPC in the learning phase.** (A) Results of normalized EP: Peak at the syllable and chunk rates are marked as red in the power spectrum (left). Green lines indicate individual subjects. The black line indicates the average across subjects. The right bar graph shows the peak values from the chunk and syllable rate (deep green) and the average from the neighboring two frequencies around the chunk and syllable rate (light green). (B) Results of ITPC.

<https://doi.org/10.1371/journal.pone.0305333.g006>

(formula:  $EP_n \sim \text{type}$ ;  $ITPC \sim \text{type}$ ). The models include Subject as a random effect (formula:  $1 \mid \text{Subject}$ ). For normalized EP, the model's intercept, corresponding to neighboring frequencies around the chunk rate, is at 0.05 ( $t(15122) = 1.78, p = 0.076$ ). Within this model, the peak of averaged neighboring frequencies around syllable rate is statistically non-significant ( $\beta = -0.004, t(15122) = -0.33, p = 0.74$ ). The peak of chunk frequency is statistically significant ( $\beta = 0.07, t(15122) = 5.56, p < .001$ ). The peak of syllable frequency is also statistically significant ( $\beta = 1.05, t(15122) = 87.82, p < .001$ ). The post-hoc tests using Kenward-Roger methods in *lmerTest* [64] further confirm that there is a significant difference between the normalized EP of the chunk rate and the normalized EP of the neighboring frequencies around the chunk rate ( $t(15094) = -5.56, p < .001$ ). There is also a significant difference between the normalized EP of the syllable rate and the normalized EP of the neighboring frequencies around the syllable rate ( $t(15094) = -88.15, p < .001$ ). There is no significant

difference between the normalized EP of the neighboring frequencies around the chunk and the syllable rate ( $t(15094) = 0.33, p = 0.74$ ).

Fig 6B shows the results of ITPC. The model's intercept, corresponding to frequencies around the chunk rate, is at 0.10 ( $t(15122) = 17.33, p < .001$ ). Within this model, we found that the ITPC of the chunk rate is statistically significant ( $\beta = 0.04, t(15122) = 17.98, p < .001$ ). The ITPC of the syllable rate is also statistically significant ( $\beta = 0.25, t(15122) = 108.2, p < .001$ ). The ITPC of neighboring frequencies around the syllable rate is statistically significant ( $\beta = 0.01, t(15122) = 4.35, p < .001$ ). The post-hoc tests show that there is a significant difference between the ITPC of the chunk rate and the ITPC of the neighboring frequencies around the chunk rate ( $t(15094) = -17.98, p < .001$ ). A significant difference was also found in the comparison between the ITPC of the syllable rate and the ITPC of the neighboring frequencies around the syllable rate ( $t(15094) = -103.86, p < .001$ ).

Results from both EP and ITPC have shown that the peaks at the chunk and the syllable rate can be observed. The results suggest that participants' brains identify and represent six syllables. In line with previous work [43], this indicates that participants are able to learn chunks based on transitional probabilities and additional short pauses.

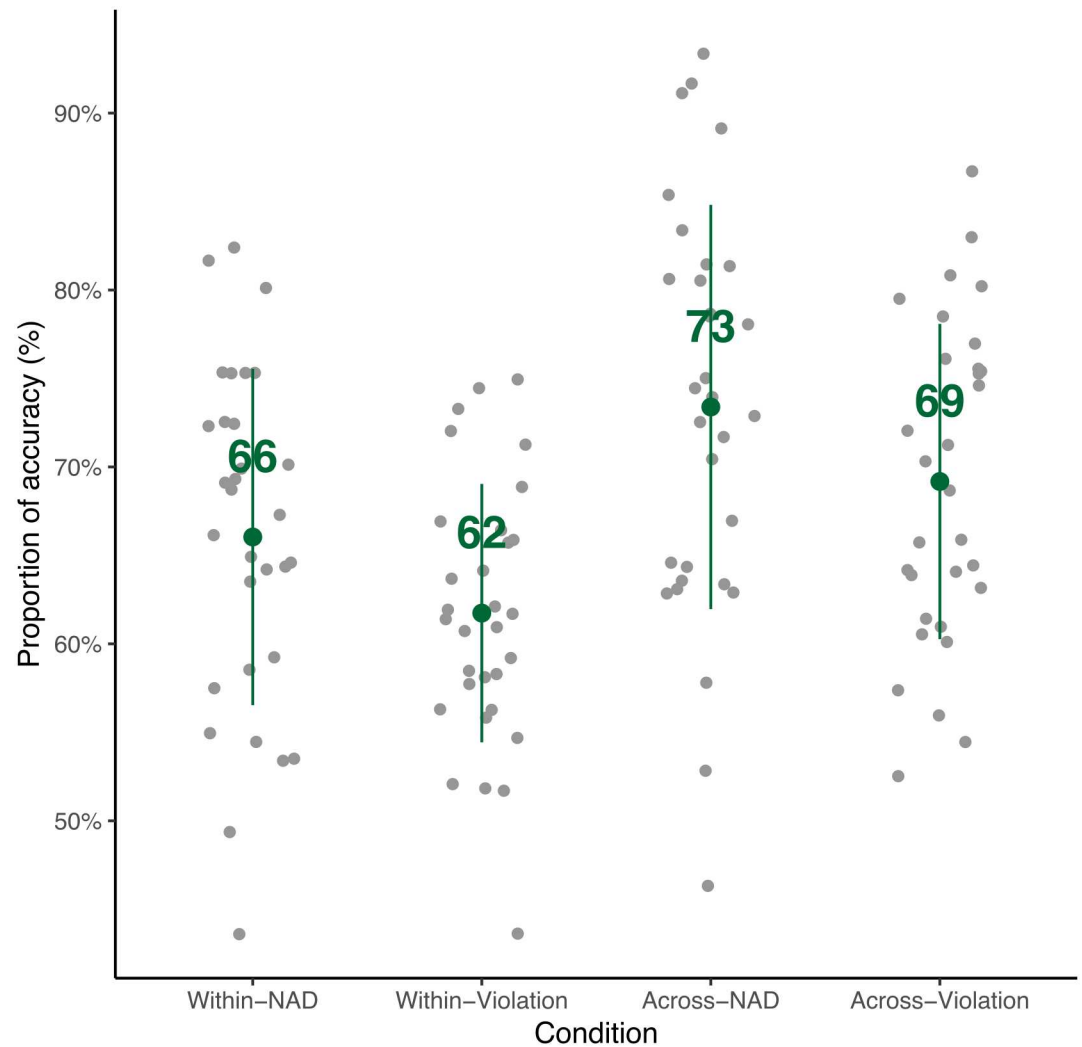
### Processing of within- but not across-chunk NADs

The overall accuracy of the behavioral responses is 68% correct in the test phase, indicating continued memory of the learned chunks. The accuracy across different conditions (Within-chunk vs. Across-chunk; NAD vs. Violation) is shown in Fig 7. We fitted a linear mixed model (estimated using REML and nloptwrap optimizer) to test the accuracy with the two conditions (within-chunk vs. across-chunk) and whether the dependency is correct (formula:  $\text{accuracy} \sim \text{within/across} * \text{dependency}$ ). The model included Subject as a random effect (formula:  $\sim 1 | \text{Subject}$ ). The effect of within/across is statistically significant ( $\beta = -0.07, t(118) = -3.55, p < .001$ ). The dependency effect is statistically significant ( $\beta = -0.04, t(118) = -2.04, p = 0.044$ ). The two main effects have no significant interaction ( $\beta = -0.009, t(118) = -0.03, p = 0.976$ ). Similar to the behavioral results in the learning phase, accuracy in the across-chunk condition is higher than the one in the within-chunk condition. Within each condition, the accuracy of the correct dependency is significantly higher than the accuracy of the violation (Post-hoc Within-NAD vs. Within-violation:  $t(90) = 2.08, p < .05$ ; Across-NAD vs. Across-violation:  $t(90) = 2.04, p < .05$ ).

Event-related potentials (ERP) were computed to assess whether NADs across chunks are harder to process. ERPs at the second elements that either complete an NAD (NAD) or violate it (NNAD) were compared within each condition. Fig 8 illustrates the results. The difference between NAD and NNAD syllables is significantly larger around 0-0.05 seconds in the within-chunk condition ( $p = 0.0017$ ) while there is no significant difference in the NAD and violation in the across-chunk condition ( $p = 0.08$ ). The difference found in the within-chunk condition but not in the across-chunk condition suggests that NADs across chunks are indeed harder to process.

### Discussion

Our results dissociate and link the active segmentation of continuous speech into multi-word memory chunks and the incremental construction of compositional meaning through NADs. This may help to reconcile disparate aspects of human language comprehension: On the one hand, in the face of memory limitations, the human brain needs to sample speech in short chunks; on the other hand, our brains link words incrementally, often requiring the formation of NADs. The findings in the current study serve as the initial electrophysiological evidence

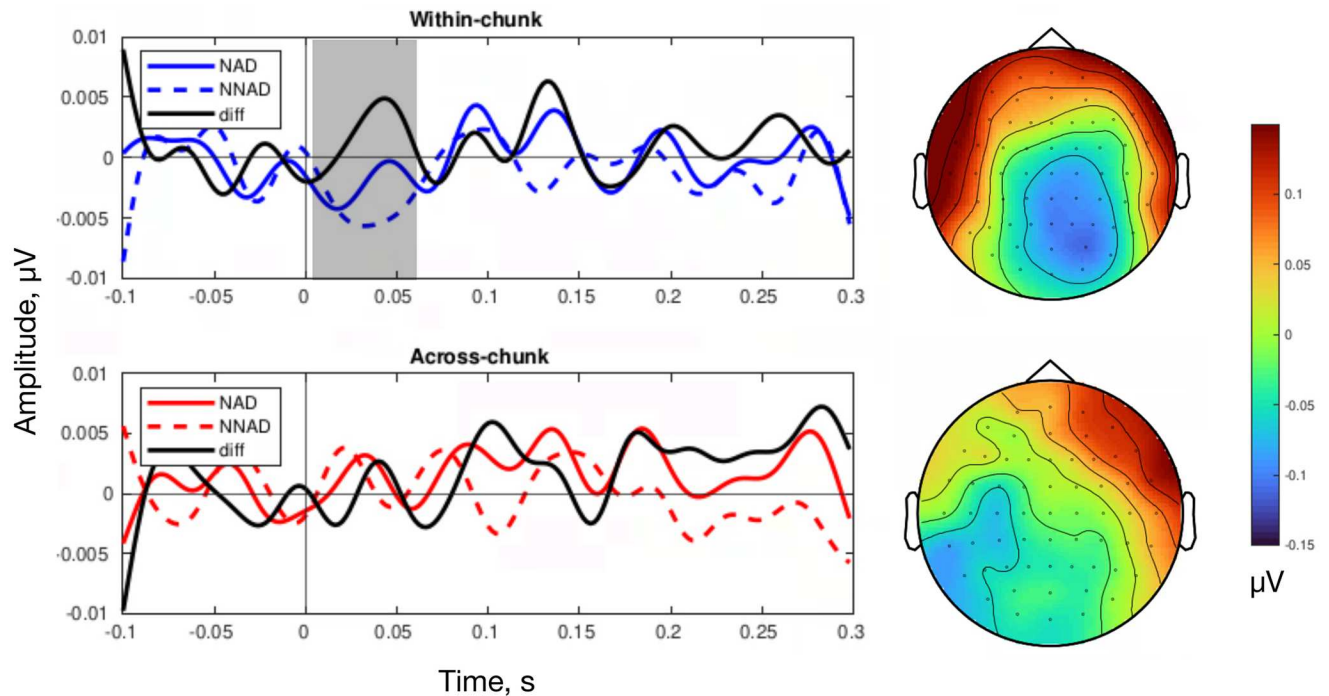


**Fig 7. Behavioral results in the test phase.** The overall accuracy is 68%. The gray dot indicates the score for each participant. The green dot indicates the mean for each condition and the green line indicates the standard deviation.

<https://doi.org/10.1371/journal.pone.0305333.g007>

that NAD processing is restricted to the current memory chunk, which had been previously only shown by computational modeling [30] and behavioral studies [20, 25]. Frequency-tagging results converge on prior work that suggests that the brain actively segments continuous speech into discrete chunks with the help of low-frequency activity [37–39]. ERP results indicate that NADs are harder to process once the dependencies cross the chunk boundary. Critically, our results overcome ambiguities in prior research, which mostly defined chunk boundaries by NADs themselves, making it difficult to dissociate the processing of segmentation and the formation of dependencies.

The current results further indicate that comprehension involves the sampling of chunks in a shallow manner, in line with previous psycholinguistic work [65–67] and neuroscientific studies [68]. In the background of building dependencies amongst incoming words incrementally in real-time, speech is sampled superficially in coarse second-long chunks; here, subjects were able to learn these based on both statistical and prosodic cues. Structure building, as operationalized here through the statistical co-occurrence of the first and second elements of

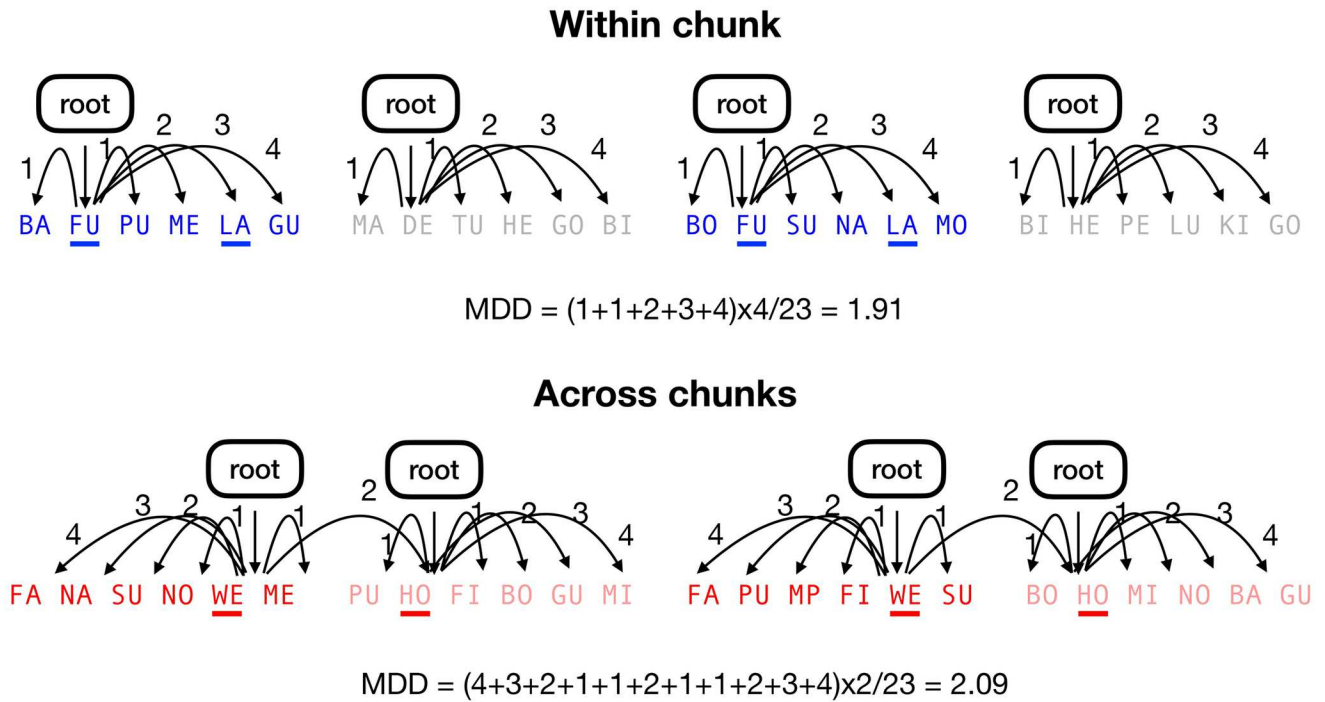


**Fig 8. ERP results in the test phase.** ERPs for the second element of correct NAD and violation (NNAD) for each condition. There is a statistical difference between the NAD and the NNAD in the within-chunk condition (the gray area), but not in the across-chunk condition. The black line indicates the difference between the correct and incorrect elements. Topographies show the difference between the NAD and the NNAD between 0-50 ms for each condition.

<https://doi.org/10.1371/journal.pone.0305333.g008>

NADs, seems incremental, yet constrained by the boundaries of memory units. This is consistent with the chunk-and-pass model of language processing, which adopts an incremental approach and requires people to integrate information as quickly as possible [40]. Our results are aligned with earlier processing models such as the sausage machine [35] and provide a link with incremental approaches [69].

The spectral peak observed in the learning phase provides supporting evidence that neural activity in the delta band underlies chunking, consistent with recent findings [37–39]. This might entail that the cognitive units of language comprehension are clusters held together by local transitional probabilities. The current memory chunk may thereby allow for establishing all dependencies required for understanding the current chunk. An exploratory analysis, using the metric of mean dependency distance (MDD; see Eq (4)), following corpus simulations from [70], suggests that this would indeed be an effective manner of processing. MDD quantifies syntactic complexity by measuring the distance between each word and chunk. In the Eq (4),  $n$  refers to the number of element/word and  $k$  refers to the number of chunks;  $cdd$  refers to the distance within the  $i$ th chunk and  $ldd$  refers to the distance between chunks. Indeed, MDD is higher for across-chunk dependencies, an example demonstrated in Fig 9. In line with the rarity of crossing dependencies found in natural languages, our results show that syntactic complexity may be reduced due to the proper local arrangement of units to be linked (in our AG: syllables; in natural language: words/morphemes). This is consistent with the chunk-and-pass model of language processing [40], evidence for dependency length minimization in the world's languages [27, 29], and the dualism of segmentation and parsing in classical psycholinguistic frameworks of sentence processing [35, 71]. Yet, our findings suggest that NAD length



**Fig 9. An example of MDDs in both conditions.** The MDD is higher in the across-chunk condition than the MDD in the within-chunk condition.

<https://doi.org/10.1371/journal.pone.0305333.g009>

is not only limited by memory constraints and flexible otherwise— apparently, our cognitive flexibility to form NADs is also constrained by the boundaries of the memory units segmented from speech. The rarity of crossing dependencies may result from the duality of language [70]. The encoding of sound structure and the composition of meaningful words or morphemes happen rapidly, incrementally, and concurrently; thus, information can be integrated locally and the possibility of crossing dependencies may be reduced. Hence, our memory capacity would not be overloaded due to this multi-level parallel processing and thus achieve successful language comprehension and production. Beyond the corpus simulation results in [70], we further show that the reduction of syntactic complexity may be observed on the neural level. Future directions addressing how different boundaries (i.e., syntactic boundary vs. prosodic boundary) may elicit similar interference and how different types of syntactic dependencies (e.g. nested dependencies in German and crossed dependencies in Dutch, [72]) are disrupted will be key to further disentangling how chunk boundaries interfere with the formation of NADs.

$$MDD = \frac{1}{n-1} \sum_{n=1}^k (|cdd_i| + |ldd_i|) \tag{4}$$

It must be debated whether our findings can be taken to fully generalize to natural language processing. As criticized by previous research, most AGs lack syntactic word categories (parts of speech) and the hierarchical rules that many researchers assume to be found in natural language; both may limit the generalization to natural language processing [73]. However, previous studies have also shown that the transitional probabilities that define AGs capture at least

some aspects of the cognitive form of linguistic knowledge. For instance, phonotactic statistical knowledge of a language stored in our long-term memory can indeed influence how we recall non-words [74–76]. That is, non-words are better recalled when the transitional probabilities of non-words are consistent with the pattern in natural language (see [77] for further discussion on memory). Therefore, we do suggest that the AG paradigm here captures the dualism of chunking and NADs also found in natural language (processing) reasonably well. We should also note that, for natural languages, we are still able to link words across chunks. Yet, it remains to be tested whether the current experimental design, transposed to natural language, will similarly show that this ability is less articulate than NAD processing within chunks. For future research, as suggested by [73], a combination of an AG with pseudo-words could be used. Results may be more comparable with natural language processing.

In addition to transitional probabilities, our experimental setup ensured the formation of chunks through a short pause and the chunk boundary. Pauses certainly support chunking. For instance, low-frequency periodicity was observed in speech sequences defined by intonation units, defined, amongst other factors, by pauses [78, 79]. Neural responses track both overt and covert prosodic boundaries [80] and chunk boundaries defined by intonation units [81]. Neural activities in the delta band may be affected by both prosodic information and syntactic structure simultaneously and the processes of these two kinds of information are overlapped strongly [82, 83]. Future research is required to see how prosody and syntax play a role in defining an optimal memory chunk formed by local dependencies. In any case, the combined marking of chunk boundaries by transitional probabilities and a short pause in the current study leaves the NAD-blocking effect untouched.

Our ERP results during the test phase reveal an unexpectedly early effect, inconsistent with the previous findings that would let us expect a modulation of the N100 component, indexing form-based processing of words during the first 80–120 ms [84–86]. However, this early effect is plausible for linguistic processing. For example, Herrmann and colleagues [87] found a greater activation between 40–80 ms for the incorrect phrases during an MEG recording of an auditory oddball paradigm for syntactic stimuli (see also [88]). Another study also shows an early effect around 20–100 ms for the violation of grammatical categories during sentence listening [89]. As mentioned by Herrmann and colleagues [87], this earlier effect is possibly associated with the P50 component, which has been associated with preferential attention to sensory inputs and general auditory arousal [90], stimulus onset perception [91], and phonemic encoding [92]. In line with the assumption of the relationship between the P50 and the early grammatical effect in [87], it seems plausible that the dependency processing can be modulated by preferential attention to sensory inputs in our finding. Critically, the current NADs could be processed on their acoustic/auditory form alone (i.e., subjects needed to learn syllable form), which overlaps in time with the current result [93–97]. Still, more future research is needed to investigate how this early effect is associated with dependency processing and how perception interacts with comprehension of higher-level information at the very early stages of language comprehension.

Future work should also extend the current findings to language development. It has been shown that infants store large chunks in early development [98, 99]. Infants are initially sensitive to slow prosodic information, which marks phrase or clause boundaries—chunk boundaries. Slowness facilitates the neural tracking of prosody in infants [100, 101]. In addition, infants 8 months of age track statistical regularities in speech and exploit transitional probability to segment continuous speech [14]. The detection of NADs can also be observed at a very young age. A study has shown that 3-month-old infants can detect the violation of AXB, accompanied by their auditory processing capacities [102]. Other studies have also shown that 8-month-old infants can learn the NADs from an AG [103, 104]. At roughly 16 months,

children learning English can recognize the grammatical dependency between auxiliary and inflectional morphology [24] (e.g. *Everybody is always baking bread.*). Similar evidence was also observed in children across different languages (German: [105], French: [25, 106], Dutch: [107]). Remarkably, a study has shown that 17-month-old infants can track NADs even when two elements are across different phonological units [25]. This indeed leaves us asking how infants integrate phonological and syntactic information during oscillatory chunk sampling. Infants start from sampling in larger chunks and then recognize the complex structures such as NADs in real language late. How children deal with the coarse units and the NAD processing remains unknown. If children sample in larger units and those units are able to include NADs involving longer distances, would the distance of NAD show a reduced effect in children, compared to adults? In addition, children's strategies for segmentation may change across development; that is, children may rely on acoustic information (e.g., stress) initially and later adapt their strategy to statistical cues to segment continuous speech [17]. How this dynamic adaptation to different strategies for speech segmentation across development interacts with the processing and learning of NADs remains a major puzzle for language acquisition research. These possible directions would be fruitful for future research and could add a valuable dimension to early language development.

## Conclusion

We show that learners segment continuous sequences into chunks, possibly with the help of delta-band oscillations. Cognitive chunk boundaries then block NAD processing. Chunk-wise sampling and NAD processing go hand in hand, with NAD processing operating within the current memory chunk. Results link the segmentation of speech by our memory-limited brains and the formation of NADs that are the combinatorial basis of human language. For further application of the current research, this can be possibly applied to language teaching. For example, students might acquire number/gender agreement or filler-gap dependency more easily if they can identify the plausible multi-word chunks in a language. Teachers can start teaching NADs within small plausible multi-word chunks when teaching a foreign language. After students gain sensitivity to the NADs within a chunk, they can gradually extend the chunk size and then form NADs across chunks. In future directions, it will be worth investigating how different types of chunk boundaries, also in natural language, (i.e., syntactic boundary vs. prosodic boundary) may drive similar types of interference and whether different types of NADs could be disrupted equally by chunk boundaries. Additionally, as children acquire larger chunks in early language development, examining how different segmentation strategies interact with the formation of NADs in language acquisition would be promising.

## Acknowledgments

We thank Chih Yeh for stimuli creation and Johannes Gereons and Paula Baer for data collection.

## Author Contributions

**Conceptualization:** Chia-Wen Lo, Lars Meyer.

**Data curation:** Chia-Wen Lo.

**Formal analysis:** Chia-Wen Lo, Lars Meyer.

**Funding acquisition:** Lars Meyer.

**Investigation:** Chia-Wen Lo, Lars Meyer.



**Methodology:** Chia-Wen Lo, Lars Meyer.

**Project administration:** Lars Meyer.

**Resources:** Lars Meyer.

**Supervision:** Lars Meyer.

**Validation:** Lars Meyer.

**Visualization:** Chia-Wen Lo, Lars Meyer.

**Writing – original draft:** Chia-Wen Lo, Lars Meyer.

**Writing – review & editing:** Chia-Wen Lo, Lars Meyer.

## References

1. Fiebach CJ, Schlesewsky M, Friederici AD. Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *Journal of Memory and Language*. 2002; 47(2):250–272. [https://doi.org/10.1016/S0749-596X\(02\)00004-9](https://doi.org/10.1016/S0749-596X(02)00004-9)
2. Gibson E. The dependency locality theory: a distance-based theory of linguistic complexity. In: Marantz A, Miyashita Y, O'Neil W, editors. *Image, Language, Brain: Papers From the First Mind Articulation Project Symposium*. The MIT Press; 2000. p. 94–126.
3. Lewis RL, Vasishth S, Dyke JAV. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*. 2006; 10(10):447–454. <https://doi.org/10.1016/j.tics.2006.08.007> PMID: 16949330
4. Phillips C, Kazanina N, Abada SH. ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*. 2005; 22(3):407–428. <https://doi.org/10.1016/j.cogbrainres.2004.09.012> PMID: 15722211
5. Friederici AD, Mueller JL, Oberecker R. Precursors to Natural Grammar Learning: Preliminary Evidence from 4-Month-Old Infants. *PLoS ONE*. 2011; 6(3):e17920. <https://doi.org/10.1371/journal.pone.0017920> PMID: 21445341
6. Frost RLA, Monaghan P. Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*. 2016; 147:70–74. <https://doi.org/10.1016/j.cognition.2015.11.010> PMID: 26638049
7. Gómez RL. Variability and detection of invariant structure. *Psychological Science*. 2002; 13(5):431–6. <https://doi.org/10.1111/1467-9280.00476> PMID: 12219809
8. Gómez RL, Gerken L. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*. 1999; 70(2):109–135. [https://doi.org/10.1016/S0010-0277\(99\)00003-7](https://doi.org/10.1016/S0010-0277(99)00003-7) PMID: 10349760
9. Marchetto E, Bonatti LL. Finding words and word structure in artificial speech: the development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language*. 2015; 42(4):873–902. <https://doi.org/10.1017/S0305000914000452> PMID: 25300736
10. Peña M, Bonatti LL, Nespor M, Mehler J. Signal-Driven Computations in Speech Processing. *Science*. 2002; 298(5593):604–607. <https://doi.org/10.1126/science.1072901> PMID: 12202684
11. Perruchet P, Tyler MD, Galland N, Peereman R. Learning Nonadjacent Dependencies: No Need for Algebraic-Like Computations. *Journal of Experimental Psychology: General*. 2004; 133(4):573–583. <https://doi.org/10.1037/0096-3445.133.4.573> PMID: 15584807
12. Aslin RN, Saffran JR, Newport EL. Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*. 1998; 9(4):321–324. <https://doi.org/10.1111/1467-9280.00063>
13. Getz H, Ding N, Newport EL, Poeppel D. Cortical tracking of constituent structure in language acquisition. *Cognition*. 2018; 181:135–140. <https://doi.org/10.1016/j.cognition.2018.08.019> PMID: 30195135
14. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274:1926–1928. <https://doi.org/10.1126/science.274.5294.1926> PMID: 8943209
15. Saffran JR, Newport EL, Aslin RN. Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*. 1996; 35(4):606–621. <https://doi.org/10.1006/jmla.1996.0032>
16. Thiessen ED, Saffran JR. When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*. 2003; 39(4):706–16. <https://doi.org/10.1037/0012-1649.39.4.706> PMID: 12859124

17. Johnson EK, Jusczyk PW. Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*. 2001; 44(4):548–567. <https://doi.org/10.1006/jmla.2000.2755>
18. Morgan JL, Meier RP, Newport EL. Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*. 1987; 19(4):498–550. [https://doi.org/10.1016/0010-0285\(87\)90017-X](https://doi.org/10.1016/0010-0285(87)90017-X) PMID: 3677585
19. Morgan JL, Saffran JR. Emerging Integration of Sequential and Suprasegmental Information in Pre-verbal Speech Segmentation. *Child Development*. 1995; 66(4):911–936. <https://doi.org/10.2307/1131789> PMID: 7671658
20. Shukla M, Nespors M, Mehler J. An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*. 2007; 54:1–32. <https://doi.org/10.1016/j.cogpsych.2006.04.002> PMID: 16782083
21. Endress AD, Mehler J. Primitive computations in speech processing. *Quarterly Journal of Experimental Psychology*. 2009; 62(11):2187–209. <https://doi.org/10.1080/17470210902783646> PMID: 19418378
22. Newport EL, Aslin RN. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*. 2004; 48:127–162. [https://doi.org/10.1016/S0010-0285\(03\)00128-2](https://doi.org/10.1016/S0010-0285(03)00128-2) PMID: 14732409
23. Onnis L, Monaghan P, Christiansen MH, Chater N. Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2004; 26:1047–1052.
24. Santelmann LM, Jusczyk PW. Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition*. 1998; 69(2):105–134. [https://doi.org/10.1016/S0010-0277\(98\)00060-2](https://doi.org/10.1016/S0010-0277(98)00060-2) PMID: 9894402
25. van Heugten M, Shi R. Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *The Journal of the Acoustical Society of America*. 2010; 128(5):EL223–EL228. <https://doi.org/10.1121/1.3486197> PMID: 21110530
26. Bock K, Miller CA. Broken agreement. *Cognitive Psychology*. 1991; 23(1):45–93. [https://doi.org/10.1016/0010-0285\(91\)90003-7](https://doi.org/10.1016/0010-0285(91)90003-7) PMID: 2001615
27. Futrell R, Mahowald K, Gibson E. Large-scale evidence of dependency length minimization in 37 languages. *PNAS*. 2015; 112(33):10336–10341. <https://doi.org/10.1073/pnas.1502134112> PMID: 26240370
28. Ferrer i Cancho R. Euclidean distance between syntactically linked words. *Physical Review E*. 2004; 70:056135. <https://doi.org/10.1103/PhysRevE.70.056135> PMID: 15600720
29. Fedzechkina M, Chu B, F Jaeger T. Human Information Processing Shapes Language Change. *Psychological Science*. 2018; 29(1):72–82. <https://doi.org/10.1177/0956797617728726> PMID: 29192836
30. Elman JL. Learning and development in neural networks: the importance of starting small. *Cognition*. 1993; 48(1):71–99. [https://doi.org/10.1016/0010-0277\(93\)90058-4](https://doi.org/10.1016/0010-0277(93)90058-4) PMID: 8403835
31. Pöppel E. A hierarchical model of temporal perception. *Trends in Cognitive Sciences*. 1997; 1(2):56–61. [https://doi.org/10.1016/S1364-6613\(97\)01008-5](https://doi.org/10.1016/S1364-6613(97)01008-5) PMID: 21223864
32. Pöppel E. Oscillations as Possible Basis for Time Perception. In: Fraser JT, Haber FC, Müller GH, editors. *The Study of Time*. Springer, Berlin, Heidelberg; 1972. p. 565–589.
33. Elbert T, Ulrich R, Rockstroh B, Lutzenberger W. The processing of temporal intervals reflected by CNV-like brain potentials. *Psychophysiology*. 1991; 28(6):648–655. <https://doi.org/10.1111/j.1469-8986.1991.tb01009.x> PMID: 1816592
34. Baddeley AD, Thomson N, Buchanan M. Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*. 1975; 14(6):575–589. [https://doi.org/10.1016/S0022-5371\(75\)80045-4](https://doi.org/10.1016/S0022-5371(75)80045-4)
35. Frazier L, Fodor JD. The sausage machine: A new two-stage parsing model. *Cognition*. 1978; 6(4):291–325. [https://doi.org/10.1016/0010-0277\(78\)90002-1](https://doi.org/10.1016/0010-0277(78)90002-1)
36. Tauroza S, Allison D. Speech Rates in British English. *Applied linguistics*. 1990; 11(1):90–105. <https://doi.org/10.1093/applin/11.1.90>
37. Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*. 2016; 19:158–164. <https://doi.org/10.1038/nn.4186> PMID: 26642090
38. Meyer L, Henry MJ, Gaston P, Schmuck N, Friederici AD. Linguistic Bias Modulates Interpretation of Speech via Neural Delta-Band Oscillations. *Cerebral Cortex*. 2016; 27(9):4293–4302.
39. Henke L, Meyer L. Endogenous Oscillations Time-Constrain Linguistic Segmentation: Cycling the Garden Path. *Cerebral Cortex*. 2021; 31(9):4289–4299. <https://doi.org/10.1093/cercor/bhab086> PMID: 33949654

40. Christiansen MH, Chater N. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*. 2016; 39:e62. <https://doi.org/10.1017/S0140525X1500031X> PMID: 25869618
41. Corrette R. Praat Vocal Toolkit; 2020. <http://www.praatvocaltoolkit.com>.
42. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.2.09, retrieved 15 February 2022; 2022. <http://www.praat.org>.
43. Buiatti M, Peña M, Dehaene-Lambertz G. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage*. 2009; 44(2):509–19. <https://doi.org/10.1016/j.neuroimage.2008.09.015> PMID: 18929668
44. Mueller JL, Bahlmann J, Friederici AD. The role of pause cues in language learning: The emergence of event-related potentials related to sequence processing. *Journal of Cognitive Neuroscience*. 2008; 20(5):892–905. <https://doi.org/10.1162/jocn.2008.20511> PMID: 18201121
45. Wilson B, Spierings M, Ravnani A, Mueller JL, Mintz TH, Wijnen F, et al. Non-adjacent Dependency Learning in Humans and Other Animals. *Topics in Cognitive Science*. 2020; 12:843–858. <https://doi.org/10.1111/tops.12381> PMID: 32729673
46. Gabard-Durnam LJ, M Leal AS, Wilkinson CL, Levin AR. The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data. *Frontiers in Neuroscience*. 2018; 12:97. <https://doi.org/10.3389/fnins.2018.00097> PMID: 29535597
47. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. 2004; 134(1):9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009> PMID: 15102499
48. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*. 2011; 2011:156869. <https://doi.org/10.1155/2011/156869> PMID: 21253357
49. Klug M, Kloosterman NA. Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*. 2022; 43(9):2743–2758. <https://doi.org/10.1002/hbm.25832> PMID: 35278015
50. Dong L, Li F, Liu Q, Wen X, Lai Y, Xu P, et al. MATLAB Toolboxes for Reference Electrode Standardization Technique (REST) of Scalp EEG. *Frontiers in Neuroscience*. 2017; 11:601. <https://doi.org/10.3389/fnins.2017.00601> PMID: 29163006
51. Makeig S, Bell AJ, Jung TP, Sejnowski TJ. Independent component analysis of electroencephalographic data. In: Touretzky DS, Mozer MC, Hasselmo ME, editors. *Advances in Neural Information Processing Systems 8*. Cambridge, MA: MIT Press; 1995. p. 145–151.
52. You RY, Zhong C. Blind source separation of multichannel electroencephalogram based on wavelet transform and ICA. *Chinese Physics*. 2005; 14(11):2176–2180. <https://doi.org/10.1088/1009-1963/14/11/006>
53. Mogron A, Jovicich J, Bruzzone L, Buiatti M. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*. 2011; 48(2):229–40. <https://doi.org/10.1111/j.1469-8986.2010.01061.x> PMID: 20636297
54. Nolan H, Whelan R, Reilly RB. FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*. 2010; 192(1):152–162. <https://doi.org/10.1016/j.jneumeth.2010.07.015> PMID: 20654646
55. Perrin F, Pernier J, Bertrand O, Echallier JF. Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*. 1989; 72(2):184–187. [https://doi.org/10.1016/0013-4694\(89\)90180-6](https://doi.org/10.1016/0013-4694(89)90180-6) PMID: 2464490
56. Ding N, Melloni L, Yang A, Wang Y, Zhang W, Poeppel D. Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). *Frontiers in Human Neuroscience*. 2017; 11:481. <https://doi.org/10.3389/fnhum.2017.00481> PMID: 29033809
57. Wen H, Liu Z. Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal. *Brain Topography*. 2015; 29:13–26. <https://doi.org/10.1007/s10548-015-0448-0> PMID: 26318848
58. Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, et al. Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*. 2020; 23:1655–1665. <https://doi.org/10.1038/s41593-020-00744-x> PMID: 33230329
59. Lo CW, Tung TY, Ke AH, Brennan JR. Hierarchy, Not Lexical Regularity, Modulates Low-Frequency Neural Synchrony During Language Comprehension. *Neurobiology of Language*. 2022; 3(4):538–555. [https://doi.org/10.1162/nol\\_a\\_00077](https://doi.org/10.1162/nol_a_00077) PMID: 37215342

60. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*. 2007; 164(1):177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024> PMID: [17517438](https://pubmed.ncbi.nlm.nih.gov/17517438/)
61. Hayasaka S, Nichols TE. Combining voxel intensity and cluster extent with permutation test framework. *NeuroImage*. 2004; 23(1):54–63. <https://doi.org/10.1016/j.neuroimage.2004.04.035> PMID: [15325352](https://pubmed.ncbi.nlm.nih.gov/15325352/)
62. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015; 67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>
63. R Core Team. R: A Language and Environment for Statistical Computing; 2022. Available from: <https://www.R-project.org/>.
64. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*. 2017; 82(13):1–26. <https://doi.org/10.18637/jss.v082.i13>
65. Marslen-Wilson WD. Sentence Perception as an Interactive Parallel Process. *Science*. 1975; 189(4198):226–228. <https://doi.org/10.1126/science.189.4198.226> PMID: [17733889](https://pubmed.ncbi.nlm.nih.gov/17733889/)
66. Sturt P, Lombardo V. Processing Coordinated Structures: Incrementality and Connectedness. *Cognitive Science*. 2005; 29:291–305. [https://doi.org/10.1207/s15516709cog0000\\_8](https://doi.org/10.1207/s15516709cog0000_8) PMID: [21702775](https://pubmed.ncbi.nlm.nih.gov/21702775/)
67. Swets B, Desmet T, Clifton C, Ferreira F. Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*. 2008; 36(1):201–216. <https://doi.org/10.3758/MC.36.1.201> PMID: [18323075](https://pubmed.ncbi.nlm.nih.gov/18323075/)
68. Brennan JR, Pykkänen L. MEG Evidence for Incremental Sentence Composition in the Anterior Temporal Lobe. *Cognitive Science*. 2017; 41(S6):1515–1531. <https://doi.org/10.1111/cogs.12445> PMID: [27813182](https://pubmed.ncbi.nlm.nih.gov/27813182/)
69. Hale JT. What a Rational Parser Would Do. *Cognitive Science*. 2011; 35(3):399–443. <https://doi.org/10.1111/j.1551-6709.2010.01145.x>
70. Lu Q, Xu C, Liu H. Can chunking reduce syntactic complexity of natural languages? *Complexity*. 2016; 21(S2):33–41. <https://doi.org/10.1002/cplx.21779>
71. Fodor JD. Learning to parse? *Journal of psycholinguistic research*. 1998; 27:285–319. <https://doi.org/10.1023/A:1023258301588>
72. de Vries MH, Christiansen MH, Petersson KM. Learning Recursion: Multiple Nested and Crossed Dependencies. *Biolinguistics*. 2011; 5(1–2):10–35. <https://doi.org/10.5964/bioling.8825>
73. Uddén J, Männel C. Artificial Grammar Learning and Its Neurobiology in Relation to Language Processing and Development. In: Rueschemeyer SA, Gaskell MG, editors. *The Oxford Handbook of Psycholinguistics* (2 ed.). Oxford University Press; 2018. p. 1–36.
74. Gathercole SE, Frankish CR, Pickering SJ, Peaker S. Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1999; 25(1):84–95. PMID: [9949710](https://pubmed.ncbi.nlm.nih.gov/9949710/)
75. Majerus S, der Linden MV, Mulder L, Meulemans T, Peters F. Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language*. 2004; 51(2):297–306. <https://doi.org/10.1016/j.jml.2004.05.002>
76. Thorn ASC, Frankish CR. Long-Term Knowledge Effects on Serial Recall of Nonwords Are Not Exclusively Lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31(4):729–735. PMID: [16060776](https://pubmed.ncbi.nlm.nih.gov/16060776/)
77. Schwering SC, MacDonald MC. Verbal Working Memory as Emergent from Language Comprehension and Production. *Frontiers in Human Neuroscience*. 2020; 14:68. <https://doi.org/10.3389/fnhum.2020.00068> PMID: [32226368](https://pubmed.ncbi.nlm.nih.gov/32226368/)
78. Inbar M, Grossman E, Landau AN. Sequences of Intonation Units form a ~ 1 Hz rhythm. *Scientific Reports*. 2020; 10:15846. <https://doi.org/10.1038/s41598-020-72739-4> PMID: [32985572](https://pubmed.ncbi.nlm.nih.gov/32985572/)
79. Stehwien S, Meyer L. Short-Term Periodicity of Prosodic Phrasing: Corpus-based Evidence. In: *Proc. Speech Prosody 2022*; 2022. p. 693–698.
80. Glushko A, Poeppel D, Steinhauer K. Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Scientific Reports*. 2022; 12:14759. <https://doi.org/10.1038/s41598-022-18162-3> PMID: [36042220](https://pubmed.ncbi.nlm.nih.gov/36042220/)
81. Inbar M, Genzer S, Perry A, Grossman E, Landau AN. Intonation Units in Spontaneous Speech Evoke a Neural Response. *Journal of Neuroscience*. 2023; 43(48):8189–8200. <https://doi.org/10.1523/JNEUROSCI.0235-23.2023> PMID: [37793909](https://pubmed.ncbi.nlm.nih.gov/37793909/)
82. Truckenbrodt H. On the relation between syntactic phrases and phonological phrases. *Linguistic inquiry*. 1999; 30(2):219–255. <https://doi.org/10.1162/002438999554048>

83. Wagner M, Watson DG. Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*. 2010; 25(7-9):905–945. <https://doi.org/10.1080/01690961003589492> PMID: 22096264
84. Sanders LD, Newport EL, Neville HJ. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*. 2002; 5:700–703. <https://doi.org/10.1038/nn873> PMID: 12068301
85. Cunillera T, Toro JM, Sebastián-Gallés N, Rodríguez-Fornells A. The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*. 2006; 1123(1):168–178. <https://doi.org/10.1016/j.brainres.2006.09.046> PMID: 17064672
86. Cunillera T, Càmara E, Toro JM, Marco-Pallares J, Sebastián-Gallés N, Ortiz H, et al. Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*. 2009; 48(3):541–553. <https://doi.org/10.1016/j.neuroimage.2009.06.069> PMID: 19580874
87. Herrmann B, Maess B, Hasting AS, Friederici AD. Localization of the syntactic mismatch negativity in the temporal cortex: An MEG study. *NeuroImage*. 2009; 48(3):590–600. <https://doi.org/10.1016/j.neuroimage.2009.06.082> PMID: 19595773
88. Herrmann B, Maess B, Hahne A, Schröger E, Friederici AD. Syntactic and auditory spatial processing in the human temporal cortex: An MEG study. *NeuroImage*. 2011; 57(2):624–633. <https://doi.org/10.1016/j.neuroimage.2011.04.034> PMID: 21554964
89. Herrmann CS, Oertel CU, Wang Y, Maess B, Friederici AD. Noise affects auditory and linguistic processing differently: an MEG study. *NeuroReport*. 2000; 11(2):227–9. <https://doi.org/10.1097/00001756-200002070-00001> PMID: 10674460
90. Key APF, Dove GO, Maguire MJ. Linking brainwaves to the brain: an ERP primer. *Developmental Neuropsychology*. 2005; 27(2):183–215. [https://doi.org/10.1207/s15326942dn2702\\_1](https://doi.org/10.1207/s15326942dn2702_1) PMID: 15753046
91. Pratt H, Starr A, Michalewski HJ, Bleich N, Mittelman N. The auditory P50 component to onset and offset of sound. *Clinical Neurophysiology*. 2008; 119(2):376–387. <https://doi.org/10.1016/j.clinph.2007.10.016> PMID: 18055255
92. Khalighinejad B, C da Silva G, Mesgarani N. Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech. *Journal of Neuroscience*. 2017; 37(8):2176–2185. <https://doi.org/10.1523/JNEUROSCI.2383-16.2017> PMID: 28119400
93. Palva S, Palva JM, Shtyrov Y, Kujala T, Ilmoniemi RJ, Kaila K, et al. Distinct Gamma-Band Evoked Responses to Speech and Non-Speech Sounds in Humans. *The Journal of Neuroscience*. 2002; 22(4):RC211. <https://doi.org/10.1523/JNEUROSCI.22-04-j0003.2002> PMID: 11844845
94. Shtyrov Y, Lenzen M. First-pass neocortical processing of spoken language takes only 30 msec: Electrophysiological evidence. *Cognitive Neuroscience*. 2017; 8(1):24–38. <https://doi.org/10.1080/17588928.2016.1156663> PMID: 26919206
95. Shtyrov YY, Stroganova TA. When ultrarapid is ultrarapid: on importance of temporal precision in neuroscience of language. *Frontiers in Human Neuroscience*. 2015; 9:576. <https://doi.org/10.3389/fnhum.2015.00576> PMID: 26539098
96. Thönnessen H, Boers F, Dammers J, Chen YH, Norra C, Mathiak K. Early sensory encoding of affective prosody: Neuromagnetic tomography of emotional category changes. *NeuroImage*. 2010; 50(1):250–259. <https://doi.org/10.1016/j.neuroimage.2009.11.082> PMID: 19969096
97. Zora H, Wester J, Csépe V. Predictions about prosody facilitate lexical access: Evidence from P50/N100 and MMN components. *International Journal of Psychophysiology*. 2023; 194:112262. <https://doi.org/10.1016/j.ijpsycho.2023.112262> PMID: 37924955
98. Grimm R, Cassani G, Gillis S, Daelemans W. Facilitatory Effects of Multi-Word Units in Lexical Processing and Word Learning: A Computational Investigation. *Frontier in Psychology*. 2017; 8:555. <https://doi.org/10.3389/fpsyg.2017.00555> PMID: 28450842
99. Peters AM. *The Units of Language Acquisition*. NY: Cambridge University Press; 1983.
100. Menn KH, Michel C, Meyer L, Hoehl S, Männel C. Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*. 2022; 251(1):118991. <https://doi.org/10.1016/j.neuroimage.2022.118991> PMID: 35158023
101. Menn KH, Ward EK, Braukmann R, van den Boomen C, Buitelaar J, Hunnius S, et al. Neural Tracking in Infancy Predicts Language Development in Children With and Without Family History of Autism. *Neurobiology of Language*. 2022; 3(3):495–514. [https://doi.org/10.1162/nol\\_a\\_00074](https://doi.org/10.1162/nol_a_00074) PMID: 37216063
102. Mueller JL, Friederici AD, Männel C. Auditory perception at the root of language learning. *PNAS*. 2012; 109(39):15953–15958. <https://doi.org/10.1073/pnas.1204319109> PMID: 23019379

103. Kabdebon C, Pena M, Buiatti M, Dehaene-Lambertz G. Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*. 2015; 148:25–36. <https://doi.org/10.1016/j.bandl.2015.03.005> PMID: 25865749
104. Gerken L. Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition*. 2006; 98(3):B67–B74. <https://doi.org/10.1016/j.cognition.2005.03.003> PMID: 15992791
105. Höhle B, Schmitz M, Santelmann LM, Weissenborn J. The Recognition of Discontinuous Verbal Dependencies by German 19-Month-Olds: Evidence for Lexical and Structural Influences on Children's Early Processing Capacities. *Language Learning and Development*. 2006; 4:277–300.
106. Culbertson J, Koulaguina E, Gonzalez-Gomez N, Legendre G, Nazzi T. Developing Knowledge of Nonadjacent Dependencies. *Developmental Psychology*. 2016; 52(12):2174–2183. <https://doi.org/10.1037/dev0000246> PMID: 27893252
107. van Heugten M, Johnson EK. Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory and Language*. 2010; 63(2):197–209. <https://doi.org/10.1016/j.jml.2010.04.001>