Title: **Neural harmonics reflect grammaticality**

Authors: Alessandro Tavano[1*], Stefan Blohm[2], Christine Knoop[2], R Muralikrishnan[1], Mathias Scharinger[2,3], Valentin Wagner[2], Dominik Thiele[1], Oded Ghitza[4], Nai Ding[5], Winfried Menninghaus[2], David Poeppel[1,6]

Affiliations:

1. Department of Neuroscience, Max Plank Institute for Empirical Aesthetics, 60322 - Frankfurt am Main, Germany
2. Department of Language and Literature, Max Plank Institute for Empirical Aesthetics, 60322 - Frankfurt am Main, Germany
3. University of Marburg, 35037 – Marburg, Germany
4. Department of Biomedical Engineering, Boston University, 02215 - Boston, USA
5. College of Biomedical Engineering and Instrument Sciences, Zhejiang University, 310027 - Hangzhou, Zhejiang Province, China
6. Department of Psychology, New York University, 10003 - New York City, New York, USA

* Corresponding author: alessandro.tavano@ae.mpg.de

Abstract

Can neural activity reveal syntactic structure building processes and their violations? To verify this, we recorded electroencephalographic and behavioral data as participants discriminated concatenated isochronous sentence chains containing only grammatical sentences (regular trials) from those containing ungrammatical sentences (irregular trials). We found that the repetition of abstract syntactic categories generates a harmonic structure of their period independently of stimulus rate, thereby separating endogenous from exogenous neural rhythms. Behavioral analyses confirmed this dissociation. Internal neural harmonics extracted from regular trials predicted participants' grammatical sensitivity better than harmonics extracted from irregular trials, suggesting a direct reflection of grammatical sensitivity. Instead, entraining to external stimulus rate scaled with task sensitivity only when extracted from irregular trials, reflecting attention-capture processing. Neural harmonics to repeated syntactic categories constitute the first behaviorally relevant, purely internal index of syntactic competence.

Keywords: Entrainment, brain rhythms, harmonics, syntax, syntactic competence, Grammaticality

Introduction

Speech and language are related but distinct dimensions of verbal communication [1–4]. While temporal regularities guide speech chunking in ways that are relevant for language production and comprehension [3, 5–13]), the core of language processing lies in the internal and tacit knowledge of syntactic rules, whose function is to determine how words combine into meaningful utterances [14]. Isolating a direct

neural correlate of such linguistic competence has proven an extremely difficult task. A growing body of literature has linked brain rhythms to various aspects of language comprehension. Recently, Ding and colleagues [15] tried to capture a neural signature of structure building in the frequency domain by having native speakers of English and Mandarin Chinese listen to continuous spoken sentences, concatenated as continuous trains of monosyllabic words presented at a fixed rate of 4 Hz. They tested sentences composed of two-word noun phrases (NP = a phrase headed by a noun, such as "My shoe") followed by two-word verb phrases (VP = "is wet", see Figure 1a). The lexical input therefore appeared at a rate of 4 Hz, phrasal units at 2 Hz, and sentences at 1 Hz. This design was selected to allow for ‘frequency tagging’ of sentence building processes, such that the neural data could be decomposed into linguistically relevant rhythmic response components. Remarkably, a Fast Fourier Transform (FFT) analysis highlighted significant peaks of spectral energy emerging not only at 4 Hz (word rate), but also at 2 Hz and 1 Hz. The authors deemed the neural responses as a direct window into the linguistic hierarchy: 2 Hz rhythms (2-word units) tracked the phrases, and 1 Hz rhythms (4-word units) tracked the sentences. Such linguistic rhythms emerged only if participants listened to their own language [16], and have been further interpreted by some as tracking the incremental combination of words into phrases, and phrases into sentences [17]. The findings have led to a number of follow-up studies, across languages.

The appeal of the interpretation notwithstanding, linguistic considerations question the isomorphic mapping of syntactic structure and brain rhythms. If a frequency analysis (e.g., an FFT) simply captures the rhythmic repetition of abstract linguistic categories, is it plausible that the spectral peak at 2 Hz reflects individual phrase processing? In fact, in such a frequency tagging design, each of the phrasal nodes -

NP and VP - repeats every second, not every half a second, so the repetition of a phrasal node, in and by itself, cannot generate a 2 Hz rhythm. For the 2 Hz phrasal rhythm interpretation to hold, one would have to hypothesize that: a) the human brain is insensitive to the difference between noun phrases and verb phrases; b) and, conversely, it is sensitive to a superordinate phrase category (which we term *sPhrase*), cycling periodically through all NPs and VPs (see Figures 1a,d, and 1b,e). Since in fact the human brain distinguishes between abstract phrasal categories [18–23], we suggest a simpler alternative model, based on categorical repetition, which does not require a direct mapping of structure building onto individual rhythms. We assume phrasal and sentence nodes repeat with the same period, so they jointly contribute to the same categorical rhythm: 1 Hz, in the example at hand. The 2 Hz peak would then be generated simply as a first harmonic consequence (see Figures 1a,d, and 1c,f). To test the '*sPhrase* response' model against the 'categorical harmonics' model by varying the size of phrasal units within sentences (see Figure 1a,d), ensuring that the period of phrases was not a multiple of sentence period, so as to be able to detect *sPhrase* rhythms in an FFT analysis. We recorded encephalographic (EEG) and behavioral data while 31 participants discriminated isochronous trials of ten sentences, half of them containing only grammatical sentences (*regular trials*), and half containing two randomly distributed ungrammatical sentences (*irregular trials* as defined by word order violations, see Figure 2).

The results are not consistent with the existence of a superordinate *sPhrase* rhythm, suggesting instead that the brain tracks the repetition of abstract phrasal categories - NPs, VPs and sentence nodes -, into a single neural rhythm, which generates its own harmonic structure. The higher harmonics are mathematically independent from

stimulus (word) rate, leading to a clear-cut distinction between internally constructed versus externally driven rhythms. Importantly, the median power of harmonics structures for regular trials was larger than that for irregular trials, suggesting that brain rhythms encode the distinction between grammatical and ungrammatical sentences, i.e. grammaticality. This was confirmed by the emergence of a functional double dissociation: (a) entraining to external stimuli was at ceiling for regular trials but scaled with task sensitivity for irregular trials, likely indexing lower-level attentive processes [24]; (b) neural harmonics extracted from regular trials better predicted individual grammaticality judgments than those extracted from irregular trials, linking for the first time brain rhythms to internalized linguistic knowledge governing behavioral decisions.

Results

Behavior. Monosyllabic stimuli formed sentences of either four (Conditions 1 and 2, see Figure 2a) or five words (Conditions 3 and 4, see Figure 2b). Disyllabic stimuli formed sentences of five words (Conditions 5 and 6, Figure 2c). For each condition, participants distinguished between trials containing solely grammatical sentences (regular trials), or trials in which two neighboring sentences were rendered ungrammatical by randomly shuffling word order (irregular trials).

Estimates of task sensitivity and response bias were obtained using a signal detection approach [25]. Discriminability index (d') measures of task sensitivity, showed that in all conditions participants successfully distinguished irregular from regular trials, suggesting that the grammatical nature of the task was well understood: all $ts_{(30)} \geq 7.94$, all $ps \leq 7.205^{-09}$, FDR-corrected (threshold $p = 0.008$. See Figure 3, left panel). There was a significant difference among conditions:

$F_{(30,185)} = 4.16$, $p = 0.001$, $\eta^2 = 0.04$. When FDR-corrected (threshold $p = 0.006$), only a difference between condition 1 and condition 4 survived, $t_{(30)} = -3.34$, $p = 0.001$, suggesting that sensitivity for condition 1 (mean = 1.83, SD = 1.24) was lower than for condition 4 (mean = 2.67, SD = 1.75). Response bias measures showed that participants adopted a stricter decision criterion in distinguishing between regular and irregular trials for conditions 1, 4, 5 and 6: all $ts_{(30)} \geq 3.65$, all $ps \leq 9.749^{-04}$, FDR-corrected (threshold $p = 0.021$. See Figure 3, right panel). This means that they were more likely to classify an irregular trial as regular. This suggests that the impoverished acoustic quality of stimuli following the pitch-flattening procedure did not drive participants' grammatical judgments. No significant difference in criterion estimates among conditions was found, suggesting that the stimulus generation procedure was bias-free: $F_{(30,185)} = 1.81$, $p = 0.115$.

Entrainment to word rate. To avoid low-level confounds on syntactic chunking, we first asked if speech sound statistics acted as implicit cues to phrasal boundaries. There emerged no significant difference between phonemic transition probabilities within and between phrases (see Supplemental Information, Figure S1a,b). We found significant differences in word frequency (Figure S1c), reflecting the known asymmetry in the distribution of monosyllabic and disyllabic words in natural languages (for the German language, see [26]).

As for power peaks, we then used the same analysis approach for both acoustic stimulus sequences and neural data, in order to make their results directly comparable: A Fast Fourier Transform analysis was run on individual trials, normalized power estimates were extracted using the complex modulus, and signal-to-noise ratio estimates (SNR) were calculated for each subject and condition. This

approach highlights the concentration of auditory or neural activity across frequencies (See Materials and Methods). Acoustic data showed a significant word rate entrainment effect: all $ts_{(30)} \geq 55.04$, all ps $\leq 1.080^{-31}$, FDR-corrected (threshold p = 0.001, Figure S2a,b,c). Neural frequency tagging effects in the EEG data were analyzed separately for regular and irregular trials. We compared activity at target frequencies to their own floor noise level, defined as the average of the samples two positions to the left and right side of each target frequency. For regular trials, we found significant peaks of spectral energy at word rhythm in all conditions, thereby verifying that the brain robustly entrained to both monosyllabic and disyllabic rhythms: all $ts_{(30)} \geq 7.97$, all ps $\leq 6.648^{-09}$, FDR-corrected (threshold p = 0.008). The same held for irregular trials: all $ts_{(30)} \geq 5.95$, all ps $\leq 1.564^{-06}$, FDR-corrected (See Figures 4, 5, and S3). We then tested the effect of grammaticality (irregular vs. regular trials) and internal syntactic structure (Conditions 1 vs. 2, with four words running at 4 Hz; conditions 3 vs. 4, with five words running at 4 Hz; conditions 5 vs. 6, with five words running at 2 Hz) in a series of two-way rmANOVAs, but found no significant main effect or interaction: all $Fs_{(1,30)} \leq 6.35$, all ps ≥ 0.017, FDR corrected (threshold p = 0.0083). Hence, neural entrainment to word rate did not reflect the judgments of grammaticality required to distinguish regular from irregular trials.

Internal linguistic rhythms. Next, we focused on the evidence for *sPhrase* vs categorical models. First, there was no evidence for linguistic rhythms in the auditory signal (see Figure S2, Supplemental Information). Second, we hypothesized that either significant *sPhrase* rhythms emerge across conditions, or a robust harmonic structure emerges across conditions. We tested these hypotheses on regular trial data. Condition 1 replicates the main condition in Ding et al., 2016). As for *sPhrase*

rhythms, condition 1 is effectively uninformative, as a putative *sPhrase* rhythm would overlap with the first harmonic to sentence rhythm. For condition 2, a *sPhrase* rhythm for the VP should have had a peak at 1.33 Hz, but we could not detect it: $t_{(30)} = -1.08$, $p = 0.28$ (Figure 4a). For conditions 3 and 4, which mirror each other in the size of NPs and VPs, *sPhrase* rhythms should have emerged at 1.33 Hz for three-word phrasal units, and 2 Hz for two-word phrasal units (Figure 2b). Again, no evidence could be found for those rhythms being different from floor noise: all $ts_{(30)} \leq -1.71$, all $ps \geq 0.09$ (See Figure 4b). Finally, for conditions 5 and 6, two *sPhrase* rhythms should emerge: 0.66 Hz for three-word units, and 1 Hz for two-word units (Figure 2c). Again and conclusively, no evidence supporting their emergence from noise could be found: all $ts(30) \leq 1.29$, all $ps \geq 0.20$ (See Figure 4c). Conversely, we found strong evidence for the presence of categorical rhythms, contributed by both sentence and phrasal nodes, and their harmonic structures. Categorical rhythms were predicted at 1 Hz for conditions 1 and 2, 0.8 Hz for conditions 3 and 4, and 0.4 for conditions 5 and 6. Significant peaks for abstract category rhythms were found in all conditions relative to floor noise, except condition 6: all $ts_{(30)} \geq 4.03$, all $ps \leq 3.471^{-04}$, FDR-corrected (threshold $p = 0.015$). Condition 6 approached significance: $t_{(30)} = 1.88$, $p = 0.068$. The first harmonic process was significant in all conditions: all $ts_{(30)} \geq 2.57$, all $ps \leq 0.015$, FDR-corrected. The second harmonic process was significant in all conditions: all $ts_{(30)} \geq 2.56$, all $ps \leq 0.015$, FDR-corrected. Importantly, our experimental design for conditions 3 to 6 ensure that first and second harmonic processes cannot be mathematically derived as subharmonics of stimulus rate [27]: for example, in conditions 3 and 4 there exists no integer factor which when multiplied by the first harmonic at 1.6 Hz would yield the 4 Hz word rate. This simple

fact grants complete functional independence to harmonic structures, suggesting the existence of purely endogenous neural generators.

Grammaticality in the frequency domain. We then turned to irregular trials, in order to assess the extent to which internalized language knowledge affects frequency domain responses. Category rhythm peaks were largely suppressed when participants attended to irregular trials: We found significant activity only in conditions 2, 3 and 4: all $ts_{(30)} \geq 2.68$, all $ps \leq 0.011$, FDR-corrected (threshold $p = 0.017$. Figure 5abc, left panel). Conditions 5 and 6 approached significance: $ts_{(30)} = 3.33$ and 1.91, $ps = 0.017$ and 0.065, respectively. Importantly, no significant harmonic process was found in any experimental condition: all $ts_{(30)} \leq 2.76$, all $ps \geq 0.009$, FDR-corrected (threshold $p = 0.008$. See Figure 5abc, left panel). We directly compared regular against irregular trials in an rmANOVA with factors Grammaticality (regular vs. irregular) and Condition for each harmonic structure peak (all comparisons FDR-corrected). There was no significant Grammaticality x Condition interaction (all $Fs_{(1,30)} \leq 4.31$, all $ps \geq 0.0465$, FDR-corrected with threshold $p = 0.005$). There was a significant Condition effect for the first harmonic of conditions 3 and 4: $F_{(1,30)} = 9.25$, $p = 0.004$, $\eta^2 = 0.28$. Activity for condition 3 (mean = 1.20 μV^2, SD = 0.30) was reduced relative to condition 4 (mean = 1.48 μV^2, SD = 0.44). More importantly, we found a significant Grammaticality effect for all category, first harmonic and second harmonic processes (all $Fs_{(1,30)} \geq 5.55$, all $ps \leq 0.025$, FDR-corrected with threshold $p = 0.025$), except for category peaks in conditions 5 and 6, and first harmonics in conditions 3 and 4 (Figure 5abc, right panel). This suggests that higher harmonics encode perceived grammaticality more reliably than lower ones. More generally, we

conclude that task sensitivity, driven by perceived grammaticality, modulates the strength of neural harmonic response.

A functional double dissociation in the frequency domain. An important test of the function of neural responses in the frequency domain is the extent to which they predict behavior. We created two simple neural predictors: an internal one, by calculating for each participant the median power across harmonics and conditions, separately for regular and irregular trials; and an external one, by averaging the strength (in SNR) to which each participant entrained to external stimulus rate across conditions, again separately for regular and irregular trials. We then correlated each neural predictor with standardized task sensitivity values, averaged across conditions. A functionally relevant double dissociation emerged. As for the internal index, power estimates extracted from both regular and irregular trials significantly predicted participants' ability to decide whether a trial was regular or irregular: regular trials, $r_{(29)} = 0.69$, $p < 0.001$; irregular trials, $r_{(29)} = 0.42$, $p = 0.017$. A comparison between correlation coefficients using the Steiger test [28] determined that the fit for regular trials was superior to that for irregular trials: $Z = 1.77$, $p = 0.038$ (Figure 6a). This suggests that the extraction of abstract category rhythms was boosted when linguistic knowledge was employed for grammaticality judgments. As for the external index, we previously showed that entrainment to word rates did not differ between conditions or by grammaticality (Figure S3). However, power estimates extracted from irregular trials predicted behavior, with higher power indicating increased attention to irregular stimuli: $r(29) = 0.40$, $p = 0.022$. Locking to stimulus rate within regular trials was at ceiling regardless of behavioral performance, confirming that entrainment per se does not reflect internal language

knowledge: $r(29) = 0.02$, $p = 0.89$ (Figure 6b). A Steiger test comparison confirmed this conclusion: $Z = 1.95$, $p = 0.025$.

Discussion

We show that the perceived repetition, at constant rates, of linguistic categories generates a harmonic structure of categorical rhythms, which directly reflects the dynamics of syntactic structure building, and its violations. We varied the syntactic complexity of phrases across six conditions, and demonstrated that linguistic hierarchy, i.e. the assumption that words combine into phrases and phrases combine to form sentences, is not isomorphically reflected in the hierarchy of energy peaks in the frequency domain [15]. Rather, the iso-periodic repetition of nominal and verbal phrases, as well as sentences, feeds into one categorical peak reflecting categorical repetition, which in turn generates its own harmonic structure and correlates with task sensitivity. Intriguingly, the categorical model closely reflects certain syntactic theories which conceive of sentence level simply as a phrasal node, reflecting a functional category on a pair with other phrases [29]. More broadly, because frequency tagging is so widely used as a design, these results provide insights into correctly interpreting the cognitive/linguistic mechanisms that underwrite rhythmic brain activity.

Behavioral analyses document that participants not only were generally highly proficient in distinguishing regular from irregular trials but that they also took a conservative stance in deeming a trial as irregular, suggesting awareness about their own grammaticality judgments. Importantly, our experimental manipulations document that the harmonic structures generated by the category-driven syntactic rhythms are independent from mere entrainment to external stimulus rate. For

example, in conditions 3 and 4, the first and second harmonics at 1.6 and 2.4 Hz cannot be derived as sub-harmonic processes of the 4 hz word rate. Furthermore, we document a double dissociation between internally generated and externally elicited rhythms. The median power of internal harmonics extracted from regular trials predicts grammatical judgments (i.e., task sensitivity) better than power extracted from irregular trials. However, the median power of word rate extracted from regular trials is at ceiling regardless of task sensitivity, suggesting that the two measures are unrelated. Word rates extracted from irregular trials correlated with task sensitivity. We interpret these findings as evidence that category-driven harmonic structures reflect internal, top-down generators of syntactic knowledge, whereas the strength of entrainment to word sequences mediates the deployment of task-relevant attention [24], with more attention to stimuli leading to better detection of irregular trials.

From a functional viewpoint, the linguistic representations and computations that can be captured by our experimental setup cannot provide decisive evidence for incremental buildup processes combining words into phrases, and phrases into sentences [17], or for lexical buildup processes [30]. This is because the brain rhythms - resulting from an FFT analysis - do not depict the temporal unfolding of linguistic structure; rather, they depict the end result of linguistic parsing, i.e. the effect of the repeated recognition of phrasal and sentential nodes within trials in the form of a harmonic structure. The findings indicate that the strength of harmonic structures predicts the degree to which native German speakers are sensitive to word order violations, i.e., to the grammaticality manipulation we selected for our experiment. The online comprehension of German relies substantially on word order within and across phrases [31]. However, within the Germanic language family,

English, sensibly, assigns word order more weight for comprehension than German does [32]. We therefore predict that the sensitivity to word order violations, as reflected by the strength of neural harmonics, should be larger for languages which strongly depend on word order for comprehension, such as English, and smaller for morphologically richer languages, such as Italian, in which comprehension is locally bound by rules of declension. This perspective opens new, exciting avenues of cross-linguistic research for testing different types of grammaticality constraints in the frequency domain.

The data functionally dissociate between internal and external rhythmic neural modulations within and below the delta band range (all categorical rhythms ≤ 1 Hz). Delta band (0.5-4 Hz) and infra-slow (< 0.5 Hz) oscillations are natural spectral niches for linguistic operations because their extended time windows match the ranges of spontaneous language segment duration [33, 34], allowing for the incremental integration of statistically driven linguistic information [35]. In congruence with this stance, slow oscillations have been found to convey top-down knowledge across brain areas [36]. However, delta rhythms also robustly entrain to salient events in continuous speech [37], as well as to meaningless word lists [15]. Hence, rhythmic activity in the delta and sub-delta bands in response to spoken language likely reflects a mixture of neural processes pertaining to attentive speech tracking as well language knowledge, possibly running in a concurrent and interactive fashion. To partition their contribution, linguistically balanced and controlled designs are of the essence.

Of note, the strength of harmonic processes (first, second and higher order) is predictable only up to a point [27]. For example, it is unclear from our data why some harmonic processes are larger than others (see the difference between conditions 3

and 4 as far as the first harmonic is concerned). In our view, what is missing at present is a good theory of signal-to-noise distribution for harmonic process. This also suggests caution in interpreting the lack of a significant peak as an argument against a harmonic interpretation.

Conclusions

By using carefully designed linguistic structures, we conclusively distinsguished internally constructed neural rhythms from externally neural rhythms induced via entrainment. The latter likely convey differences in stimulus-driven attention capture, while the former reflect the perception of regular and violated syntactic structure rules.

Materials and methods

Participants. Thirty-two young adults participated in the Experiment (Age range: 19-30, 8 males). The EEG data from one participant were not faithfully recorded, and were thus discarded. All analyses were run on the remaining participants (N = 31): they were all German native speakers, right-handed (Oldfield Test), and self-reported normal hearing, normal-to-corrected vision, no medical history of treatments affecting the central nervous system, and no psychiatric disturbances. They were compensated with 10 euros per hour (~3 hours, including instruction, EEG montage and cleaning). All experimental procedures were approved by the Ethics Committee of the Max Planck Society, and were undertaken with a written informed consent signed by each participant.

Stimuli. Stimuli were words composing different sentence types. To reduce implicit word- and phrase-level prosodic cues, stimuli were individually generated using the MacinTalk Text-to-Speech Synthesizer (OSX El Capitan, voice Anna, f0 set at 220 Hz), and then pitch-flattened to 220 Hz across all time points (fundamental frequency estimated using an autocorrelation approach, Praat, v. 6.0.16, www.praat.org). If an original stimulus was longer than the relative SOA, its duration was compressed to fit the stimulation window while preserving pitch.

Individual trials were composed of ten sentences, randomly selected from predefined lists of 50 sentences, one list per sentence type condition (see Procedure, and Appendix). There were two types of trial: regular trials, which contained only grammatically correct sentences; and irregular trials, which contained also two ungrammatical sentences, obtained by randomly shuffling word positions across two successive sentences. There were 50 trials per condition, half were regular and half irregular, thereby yielding unbiased estimates of accuracy in discriminating regular from irregular trials, a task akin to classic grammaticality judgments.

Procedure. Participants sat comfortably in a IAC 40a sound attenuating and electrically shielded recording booth (IAC Acoustics), approximately 1m from an LCD computer screen. They were instructed to fixate a cross at the center of the screen and listen attentively to each trial; When a trial ended, they were asked to determine whether it was regular or irregular. They were asked to be as accurate as possible, with no time limit. Participants used their right hand to press on two pre-defined buttons (arrow up regular, arrow down irregular) on a keyboard. Stimuli were delivered diotically at 75 dBs SPL via loudspeakers positioned at circa 1.2 m from participants, 25 cm to the left and right of the LCD screen. On presentation, stimuli

were further attenuated using a fixed -20 dB SPL step, resulting in a comfortable and perceptually controlled environment. Brief rest periods between blocks were self-determined by each participant.

There were six sentence type conditions. Trial presentation was blocked by sentence type. Stimuli within a trial were presented at a constant Stimulus Onset Asynchrony (isochronous SOA): 250 ms for monosyllabic words, 500 ms for disyllabic words. As in Ding et al. [15], there was no gap between sentences within a trial: Participants listened to a continuous flow of words equally spaced in time [38], further reducing prosodic cues at phrasal and sentence level.

The main experimental manipulation pertained to the size - in number of words - of Noun Phrases (NP) and Verb Phrases (VP) composing each sentence. An NP is a phrase with a noun as head (e.g., "The girls"), while a VP has a verb as head (e.g., "play rugby"). NPs usually perform the grammatical functions of verb subject or verb object. In our experiments, all grammatical sentences contained an NP functioning as verb subject, and a VP, which either included a second NP functioning as a verb object or a different phrasal component. The Appendix lists all token sentences. See Figure 2 for an exemplary analysis of a sentence used in each condition. In the main text, NP always refers to a Noun Phrase functioning as subject, and VP refers to the whole Verb Phrase, regardless of its internal composition.

Stimulus sequences were created using custom scripts written in Matlab (R2015b, 64 bit, mathworks.com). Sequence delivery was controlled by Psychophysics Toolbox Version 3 (PTB-3, psychtoolbox.org) for Matlab, running on a Windows 7 computer (ASIO sound card for optimal stimulus latency control).

Data Recording and Analysis. Behavioral accuracy data were subject to a signal detection theory approach, obtaining measures of task sensitivity (d' = zscore(Hits) minus zscore(False Alarms)) and response bias (criterion = -0.5*(zscore(Hits)+zscore(False Alarms))) for each participant and condition [25]. To record EEG data, we used an actiCAP 64-channel, active electrode set (10-10 system, Brain Vision Recorder, Brain Products, brainproducts.com) to record electroencephalographic (EEG) activity at a sampling rate of 1KHz, with a 0.1 Hz online filter (12 dB/octave roll-off). Additionally, electrocardiographic (ECG) and electrooculographic (EOG) signals were recorded using a standard bipolar montage. All impedances were kept below 10 kOhm. Data were recorded with an on-line reference to FCz channel, and offline re-referenced to the average activity of all scalp channels, and downsampled to 250 Hz. Using the EEGLAB toolbox for Matlab [39] (sccn.ucsd.edu), continuous EEG data were first visually inspected to remove large non-stereotypical artifacts (e.g., sudden head movements, chewing), digitally filtered at 35 Hz low-pass (kaiser window, beta = 5.65326, filter order 93, transition bandwidth 10 Hz), high-pass filtered at 1 Hz (filter order 455, transition bandwidth 2 Hz) to ensure data stationarity, submitted to an automatic artifact rejection based on spectrum thresholding (threshold = 10 standard deviations, 1-35 Hz), and then decomposed into Independent spatial components using the Infomax algorithm, which allows for optimal source separation [40]. The resulting Independent Components (ICs) were tested using the SASICA toolbox for EEGLAB [41]: ICs reflecting blinks/vertical eye movements, lateral eye movements and heart-beat, were detected by means of a correlation threshold (0.7) with bipolar Vertical, Horizontal EOG, and ECG channels, respectively, and found to be present in all participants (range: 1-3 ICs for vertical eye movements, 1-2 ICs for horizontal, 1 IC

for heart beat). To reduce inter-trial signal variability, ICs reflecting muscle artifacts were identified via autocorrelation (threshold = 0, lag = 20 s), while ICs reflecting focal topography was used for bad electrodes (= 7 standard deviations threshold relative to the mean across electrodes). ICA results were copied back to the original EEG datasets (0.1 Hz high-pass), and ICs marked as artifactual were rejected before trial epoching. Each ten-sentence trial varied in duration between 10 seconds (4 words per sentence, running at 4 Hz) and 25 seconds (5 words per sentence, running at 2 Hz). Individual trial power estimates were extracted for each electrode from Hann-windowed and standardized data using the complex modulus of the Fast Fourier Transform, correcting for the Hann window loss of power (sqrt(1.5)). Signal-to-noise ratio estimates (SNR) were calculated for each peak of interest using the average of two samples before and two samples after.

For all hypothesized peaks of energy, we created a surrogate noise floor condition by averaging the SNR values of two samples below and two samples above the peak sample. Each floor condition was used as the term of comparison in a T-test to verify the peak's significance. To verify the significance of syntactic manipulations, repeated measures analysis of variance (rmANOVAs) were run on each condition pair with the same stimulation rate and number of words in a sentence. False-discovery-rate (FDR) correction with Q value = 0.05 was applied in all cases of multiple comparisons [42].

Appendices

See Table 1

Acknowledgements

The authors would like to thank Cornelius Abel, Jana Gessert, Freya Materne, Claudia Lehr, Alexander Lindau, Georg-Friederich Paasch for help with stimulus creation and data collection, and Johanna Rimmele, Andrea Martin and Nina Kazanina for critical feedback on data analysis.

References

1. Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 2004; 92: 67–99.

2. Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007;8: 393–402.

3. Giraud AL, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 2012;15(4) :511–517.

4. Tavano A, Scharinger M. Prediction in speech and language processing. Cortex 2015; 68: 1–7.

5. Ding N, Patel AD, Chene L, Butler H, Luo C, Poeppel D. Temporal modulations in speech and music. Neurosci Biobehav Rev. 2017;81: 181–187.

6. Ghitza O, Greenberg S. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica 2009;66: 11326.

7. Ghitza O. Acoustic-driven delta rhythms as prosodic markers. Lang Cogn Neurosci. 2017;32: 54561.

8. Assaneo F, Poeppel D. The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. Sci Adv. 2018;4(2): eaao3842.

9. Steinhauer K, Alter K, Friederici AD. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat Neurosci. 1999;2: 191—196.

10. Hwang H, Steinhauer K. Phrase length matters: The interplay between implicit prosody and syntax in Korean garden path sentences. J Cogn Neurosci. 2017;23(11): 295– 300.

11. Tavano A, Pesarin A, Murino V, Cristani M. Automatic conversational scene analysis. low-level acoustic feature classification of speech activity in children with Asperger syndrome and typically developing peers. PLoS One 2014;9(1): e85819.

12. Scharinger M, Steinberg J, Tavano A. Integrating speech in time depends on temporal expectancies and attention. Cortex 2017;93: 28–40.

13. Keitel A, Gross J, Kayser C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biology 2018;16(3): e2004473.

14. Chomsky N. Language architecture and its import for evolution. Neurosci Biobehav Rev. 2017;81: 295-300.

15. Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci. 2016;19:15864.

16. Ding N, Pan X, Luo C, Su N, Zhang W, Zhang J. Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. J Neurosci. 2017; 38(5): 1178-1188.

17. Martin AE, Doumas LA. A mechanism for the cortical computation of hierarchical linguistic structure. PLoS Biology 2017;15(3): e2000663.

18. Pallier C, Devauchelle AD, Dehaene S. Cortical representation of the constituent structure of sentences. PNAS USA 2011;108: 2522–2527.

19. Pylkknen, Liina B, Bemis DK, Elorrieta EB. Building phrases in language production: an Meg study of simple composition. Cognition 2014;133(2): 371–384.

20. Fruchter J, Linzen T, Westerlund M, Marantz A. Lexical preactivation in basic linguistic phrases. J Cogn Neurosci. 2015;27(10): 1912–1935.

21. Pylkknen L. The neural basis of combinatory syntax and semantics. Science 2019;366(6461): 62–66.

22. Zaccarella E, Mayer L, Makuuchi M, Friederici AD. Building by syntax: The neural basis of minimal linguistic structures. Cereb Cortex 2017;27(1): 411–421.

23. Nelson MJ, El Karoui I, Giber K, Yang X, Cohen L, Koopman H et al. Neurophysiological dynamics of phrase-structure building during sentence processing. PNAS USA 2017;114: E3669–E3678.

24. Andersen SK, Müller MM. Driving steady-state visual evoked potentials at arbitrary frequencies using temporal interpolation of stimulus presentation. BMC Neurosci. 2015;16: 95.

25. Macmillan NA, Creelman CD. Detection theory: A user's guide. 2nd ed. New York: Lawrence Erlbaum Associates Publishers; 2005.

26. Kupietz, M. BCKH, Witt A. The German reference corpus Dereko: A primordial sample for linguistic research. In: Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, et al., editors. Proceedings of the 7th conference on international language resources and evaluation (lrec 2010). Valletta, Malta: European language resources association (Elra); 2010. pp. 1848-1854.

27. Zhou H, Melloni L, Poeppel D, Ding N. Interpretations of frequency domain analyses of neural entrainment: Periodicity, fundamental frequency, and harmonics. Front Hum Neurosci. 2016;10: 274.

28. Steiger JH. Tests for comparing elements of a correlation matrix. Psychological Bulletin 1980;87: 245–251.

29. Rizzi L, Cinque G. Functional categories and syntactic theory. Ann Rev Ling. 2016;2(1): 139–163.

30. Frank SL, Yang J. Lexical representation explains cortical entrainment during speech comprehension. PLoS ONE 2018;13(5): e0197304.

31. Weyerts H, Penke M, Münte TF, Heinze HJ, Clahsen H. Word order in sentence processing: An experimental study of verb placement in German. J Psycholing Res. 2002;31(3): 21168.

32. Koster J. Word order in sentence processing: An experimental study of verb placement in German. In: Abraham W, editor. Groninger arbeiten zur germanistischen linguistik. Groningen, The Netherlands: University of Groningen; 1999. pp. 1-42.

33. Bieber D. Variation across speech and writing. Cambridge: Cambridge University Press; 1988.

34. Greenberg S, Carvey H, Hitchcock L, Chang S. Temporal properties of spontaneous speech - a syllable-centric perspective. J Phonetics 2003;31(3): 465–485.

35. Meyer L, Henry ML, Gaston P, Schmuck N, Friederici AD. Linguistic bias modulates interpretation of speech via neural delta-band oscillations. Cereb Cortex 2017;27(9): 4293–4302.

36. Park H, Ince RAA, Schyns PG, Thut G, Gross J. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr Biol. 2015;25: 1649–1653.

37. Keitel A, Robin AA Ince, Joachim Gross CK. Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. Neuroimage 2017;147: 32–42.

38. Rossion B. Understanding face perception by means of human electrophysiology. Trends Cogn Sci. 2014;18(6): 310–318.

39. Delorme A, Makeig S. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics. J Neurosci Methods 2004;134: 9–21.

40. Bell A, Sejnowski T. An information-maximization approach to blind separation and blind deconvolution. Neural Computation 1995;7: 1129–1159.

41. Chaumon, M. BD, Busch N. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. J Neurosci Methods 2015;250: 47–63.

42. Genovese, Christopher R. LN, Nichols T. Thrsholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 2002;15: 870–878.

43. Baayen, H PR, Gulikers L. The Celex lexical database [cd-rom]; 1995. Philadelphia, PA: Linguistic data consortium, University of Pennsylvania.

Figure legends

Figure 1, Neural response models: a. A syntactic description of a sample sentence from condition 1, replicating in German the structure used by Ding et al. [15] in English and Mandarin Chinese. b. According to the *sPhrase* model, linguistic brain rhythms emerge by cycling through all nodes of the same syntactic level, regardless of their distinct categorical status. A Fast Fourier Transform (FFT) analysis captures the *sPhrase* model results. Notice that the model predicts also the absence of a 3 Hz higher harmonic. c. The categorical model holds that the brain distinguishes between NPs and VPs as separate abstract categories: their individual period equals that of the sentence node, as phrases also repeat every second. The FFT analysis of brain data should therefore highlight a 1 Hz peak of energy and its harmonics, reflecting the contribution from both phrase and sentence levels of syntactic competence. d. A sample sentence from condition 2 with the same number of words, but different phrasal organization. e. *sPhrase* model results: The NP - with one word running at 4 Hz - would generate a 4 Hz rhythm and thus be indistinguishable from external word rhythm, while the VP - with three words running at 4 Hz - would yield a distinct peak of energy at 1.33 Hz in the resulting FFT model. f. A categorical analysis of condition 2 assumes that each phrase type cycles independently, thus effectively overlapping in predictions with condition 1. The resulting neural response should not differ from that of condition 1. See text for specific model predictions in each condition.

Figure 2, Syntactic variability across conditions: a. Condition 1 and 2 have four monosyllabic words running at a constant speed of 4 Hz, with sentences in condition 1 composed of a two-word NP and a two-word VP - replicating Ding et al. [15] -,

while sentences in condition 2 are composed by a one-word NP followed by a three-word VP. b. Conditions 3 and 4 have five monosyllabic words running at a constant speed of 4 Hz: Sentences in condition 3 are composed of a three-word NP followed by a two-word VP, while condition 4 presents the reverse pattern. c. Conditions 5 and 6 replicate the structure in conditions 3 and 4, only with disyllabic words running at 2 Hz. For each condition, a syntactic tree of an exemplary sentence and the corresponding waveform are provided. Red and green color codes identify even and odd conditions from each pair (1-2, 3-4, 5-6, respectively), here and throughout the text. See Supplemental Information for a complete list of sentence materials.

Figure 3, Behavioral accuracy: a. In all conditions, participants were able to discriminate between regular and irregular trials. A significant difference between conditions 1 and 4 suggests that discriminating irregular from regular trials in condition 1, the original condition of Ding et al. [15], was more difficult than in condition 4. b. Positive response bias measures for conditions 1, 4, 5 and 6 suggest that participants adopted a conservative criterion when operating grammatical judgments: They were more likely to respond that a given trial was regular, rather than irregular.

Figure 4, Neural linguistic rhythms in regular trials. The left panel plots the FFT results for regular trials across condition pairs: 1-2, 3-4, 5-6. Word rate is highlighted, along with the positional distribution of identifiable *sPhrase* rhythms (vertical hyphenated bars). No significant peaks were found for any of the hypothesized *sPhrase* rhythms. Conversely, a significant harmonic structure was found in all conditions: the right panel reports individual peak values for category rhythms, as

well as first and second harmonic process, plotted against floor noise estimates. Significance is marked as follows: ** $\leq$ 0.01, * $\leq$ 0.05, ns = non-significant. Notice that all probability values, here and in the text, are objectively reported, and deemed significant or not within an FDR approach.

Figure 5, Neural linguistic rhythms in irregular trials: The left panel plots the FFT results for all irregular trials in all condition pairs. A generalized suppression of category rhythms and their harmonic structure is clear. Right panel: for each condition pair, SNR estimates are plotted - with horizontal bar indicating a sample's median – according to whether data were extracted from regular trials (REG), or irregular trials (IRR). The significance of each comparison is marked as follows: ** $\leq$ 0.01, * $\leq$ 0.05, ns = non significant (FDR-corrected).

Figure 6, Brain/behavior fit: a. The median power of internal harmonics extracted from regular trials across conditions predicts grammatical judgments (black thick line) better than power extracted from irregular trial harmonics (ochre thick line); b. Entrainment to stimulus rate from regular trials is at ceiling, unmodulated, and thus not predictive of task sensitivity (black thick line). Entrainment values extracted from irregular trials (ochre thick line) significantly increase with task sensitivity. We take this finding to suggest that attention to stimuli modulated the ability to detect ungrammatical sentences within irregular trials. Thin, dotted lines indicate each correlation's 95% confidence interval. Significance: *** $\leq$ 0.001, * $\leq$ 0.05, ns = non significant.

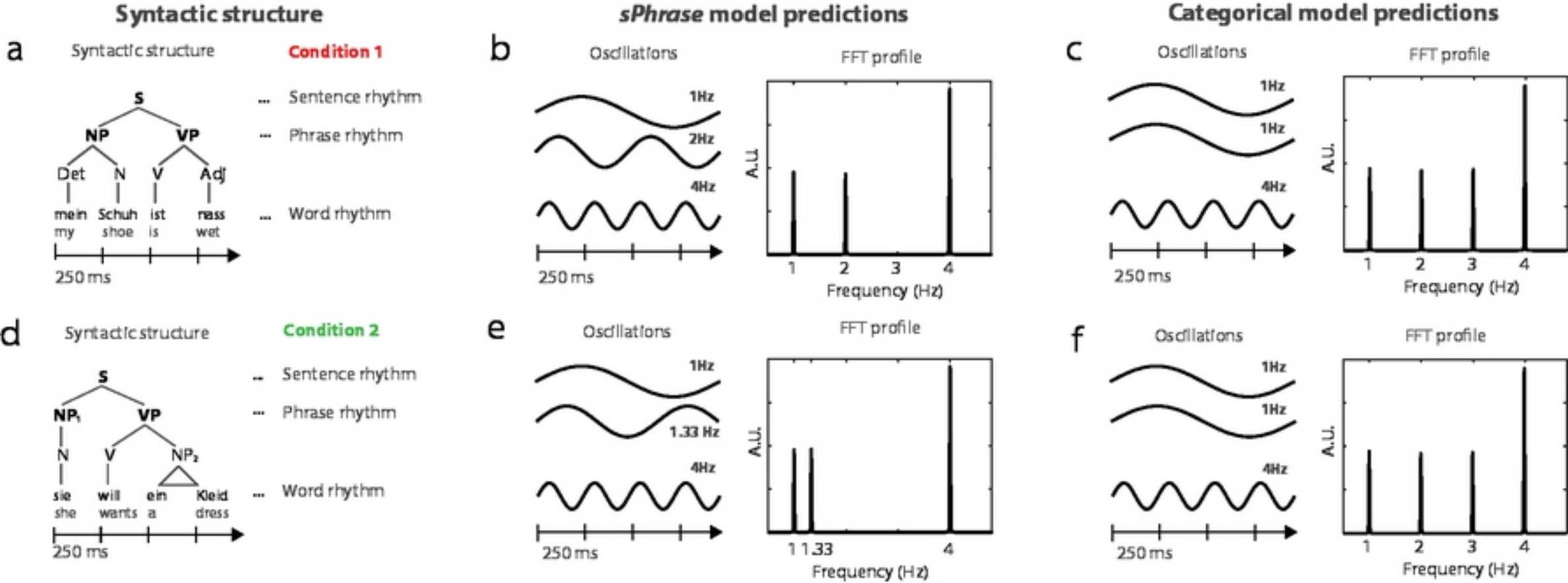
Syntactic structure
sPhrase model predictions
Categorical model predictions
a
Syntactic structure
Condition 1
S
NP
VP
Det
N
V
Adj
mein
Schuh
ist
nass
my
shoe
is
wet
250 ms
Sentence rhythm
Phrase rhythm
Word rhythm
b
Oscillations
FFT profile
1Hz
2Hz
4Hz
A.U.
1
2
3
4
Frequency (Hz)
c
Oscillations
FFT profile
1Hz
1Hz
4Hz
d
Syntactic structure
Condition 2
NP1
VP
N
V
NP2
sie
will
ein
Kleid
she
wants
a
dress
e
Oscillations
FFT profile
1.33 Hz
1 1.33
f

Figure 1

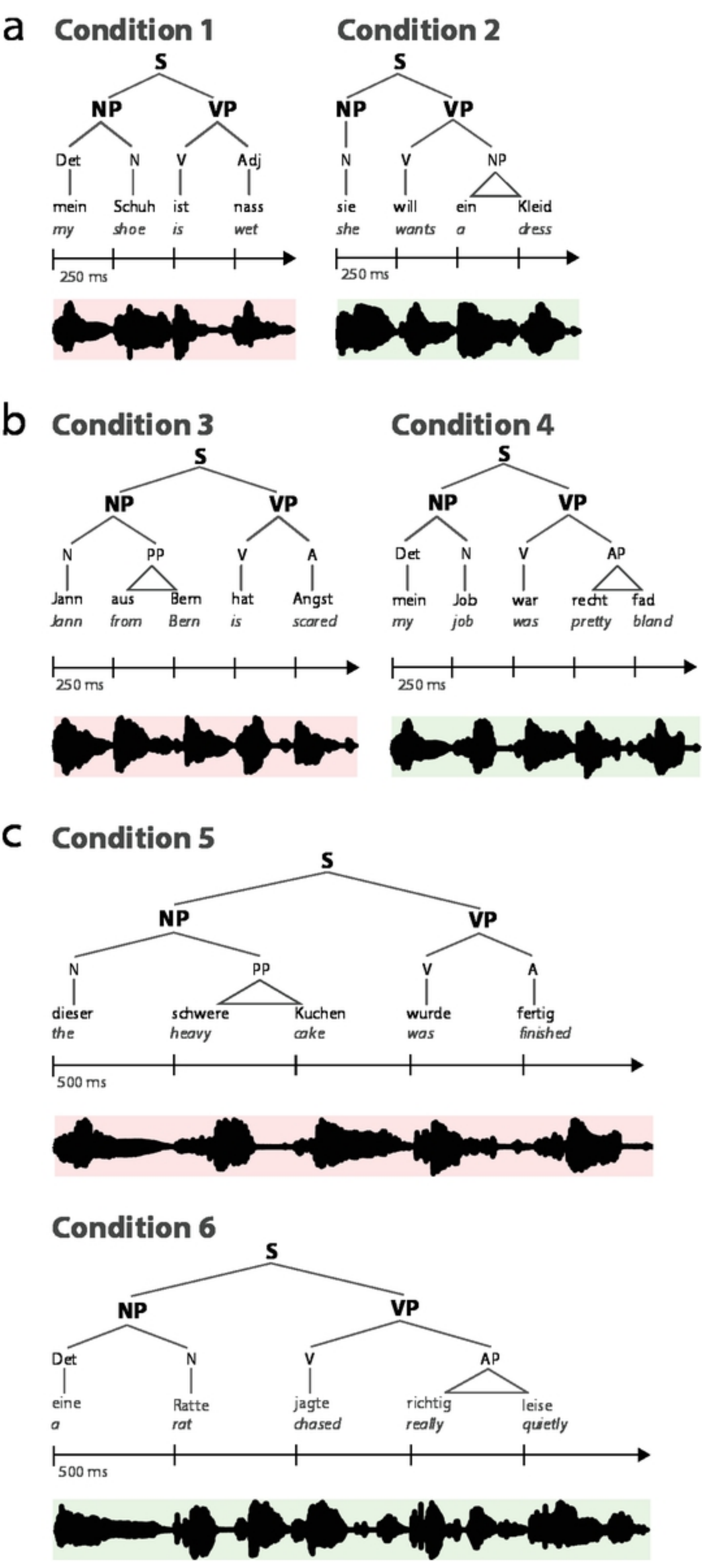
a
Condition 1
S
NP
VP
Det
N
V
Adj
mein
Schuh
ist
nass
my
shoe
is
wet
250 ms
Condition 2
S
NP
VP
N
V
NP
sie
will
ein
Kleid
she
wants
a
dress
250 ms
b
Condition 3
S
NP
VP
N
PP
V
A
Jann
aus
Bern
hat
Angst
Jann
from
Bern
is
scared
250 ms
Condition 4
S
NP
VP
Det
N
V
AP
mein
Job
war
recht
fad
my
job
was
pretty
bland
250 ms
c
Condition 5
S
NP
VP
N
PP
V
A
dieser
schwere
Kuchen
wurde
fertig
the
heavy
cake
was
finished
500 ms
Condition 6
S
NP
VP
Det
N
V
AP
eine
Ratte
jagte
richtig
leise
a
rat
chased
really
quietly
500 ms

Figure 2

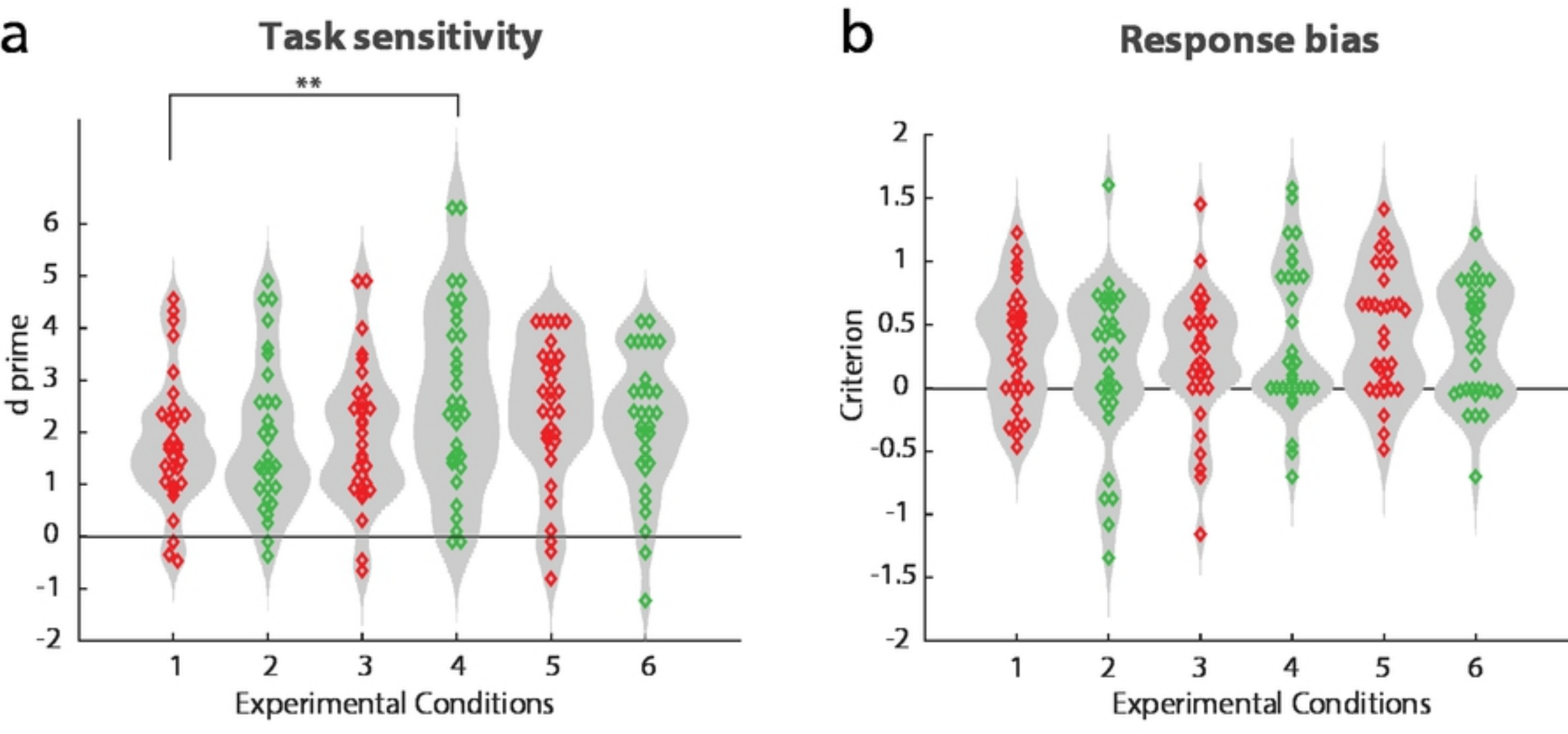
a
Task sensitivity
**
d prime
6
5
4
3
2
1
0
-1
-2
1
2
3
4
5
6
Experimental Conditions
b
Response bias
Criterion
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1
2
3
4
5
6
Experimental Conditions

Figure 3

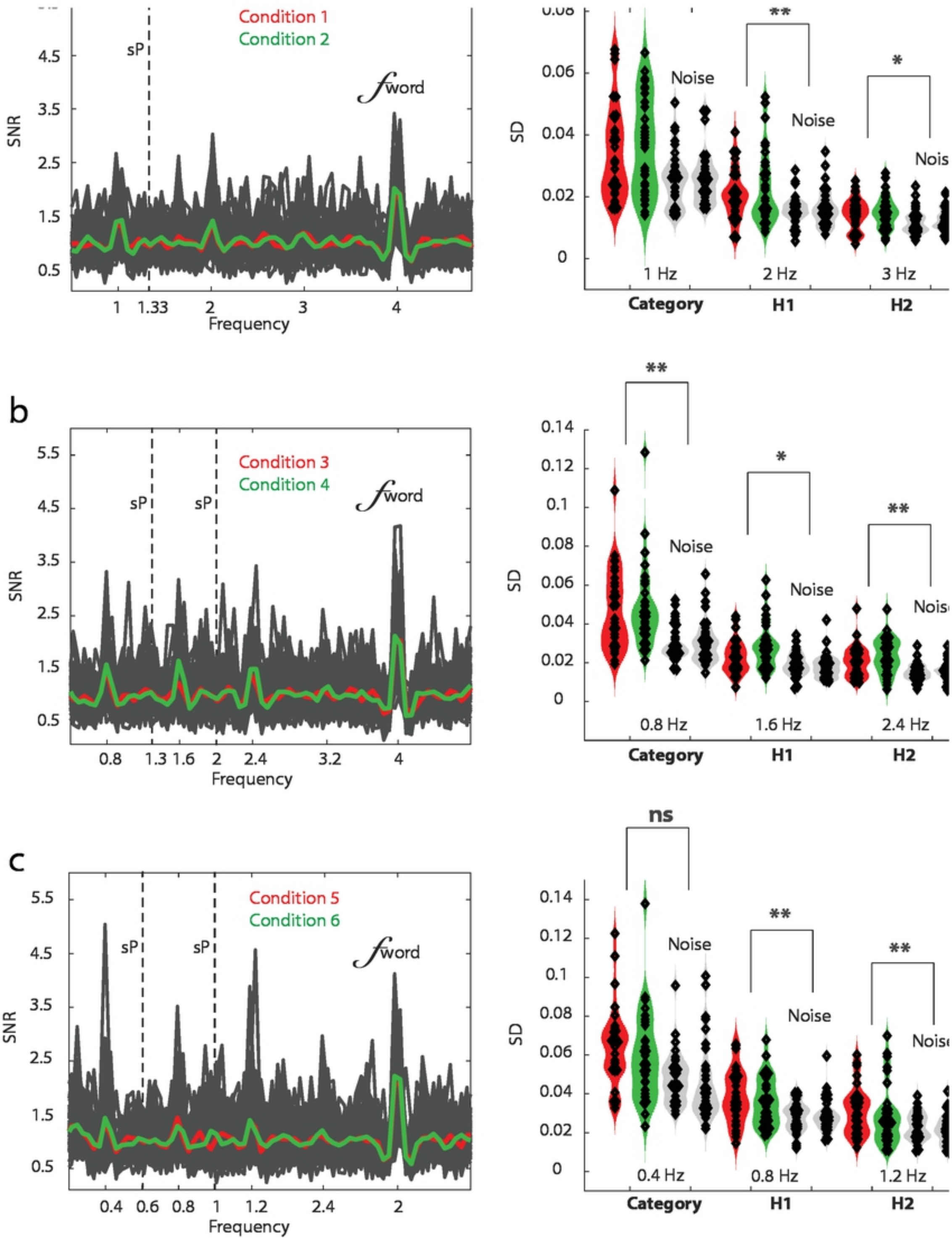
Condition 1
Condition 2
sP
fword
SNR
Frequency
SD
Noise
1 Hz
2 Hz
3 Hz
Category
H1
H2
b
Condition 3
Condition 4
0.8 Hz
1.6 Hz
2.4 Hz
c
Condition 5
Condition 6
0.4 Hz
0.8 Hz
1.2 Hz
ns

Figure 4

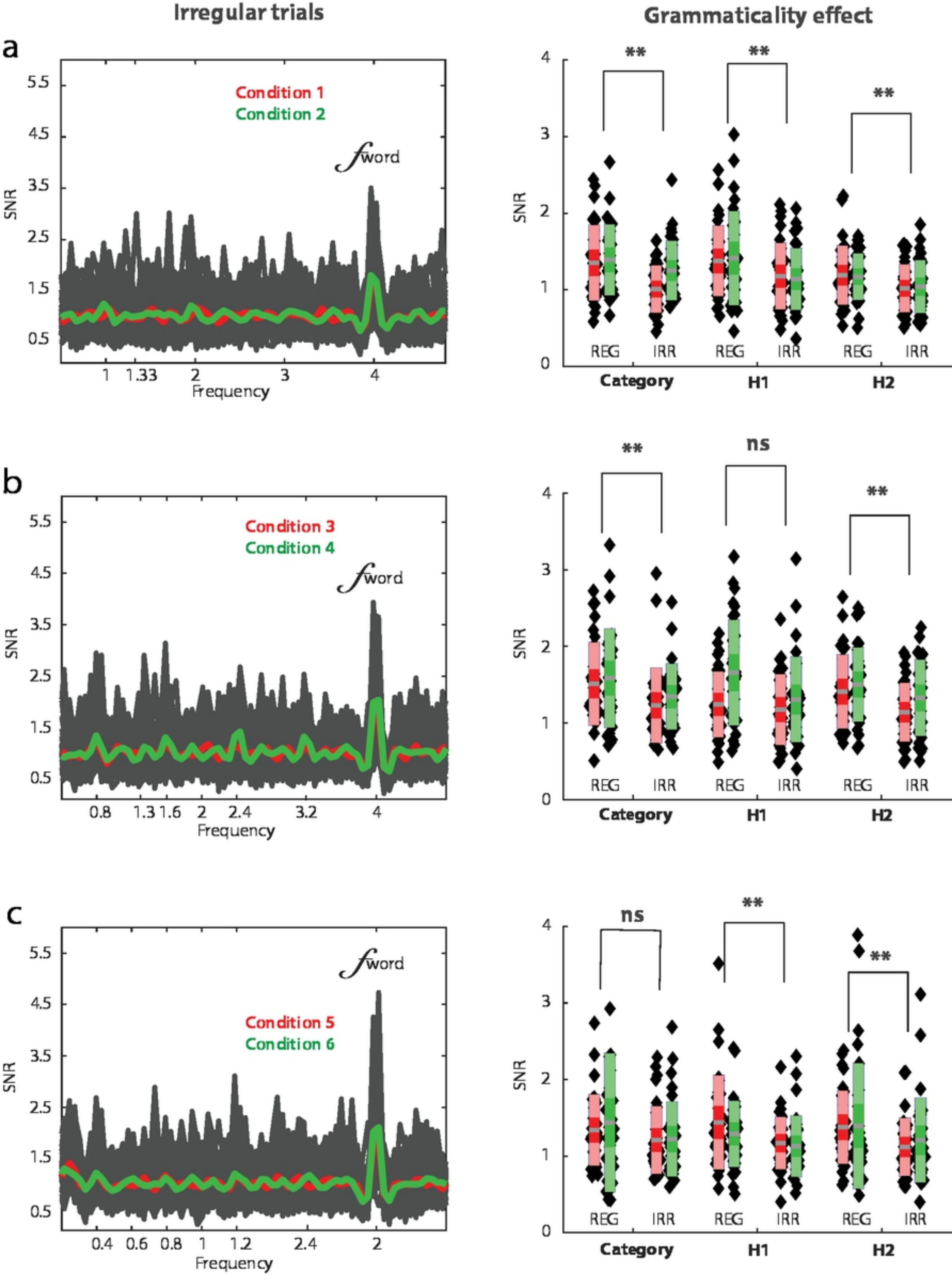
Irregular trials
Grammaticality effect
a
Condition 1
Condition 2
fword
SNR
Frequency
REG
IRR
Category
H1
H2
**
b
Condition 3
Condition 4
ns
c
Condition 5
Condition 6

Figure 5

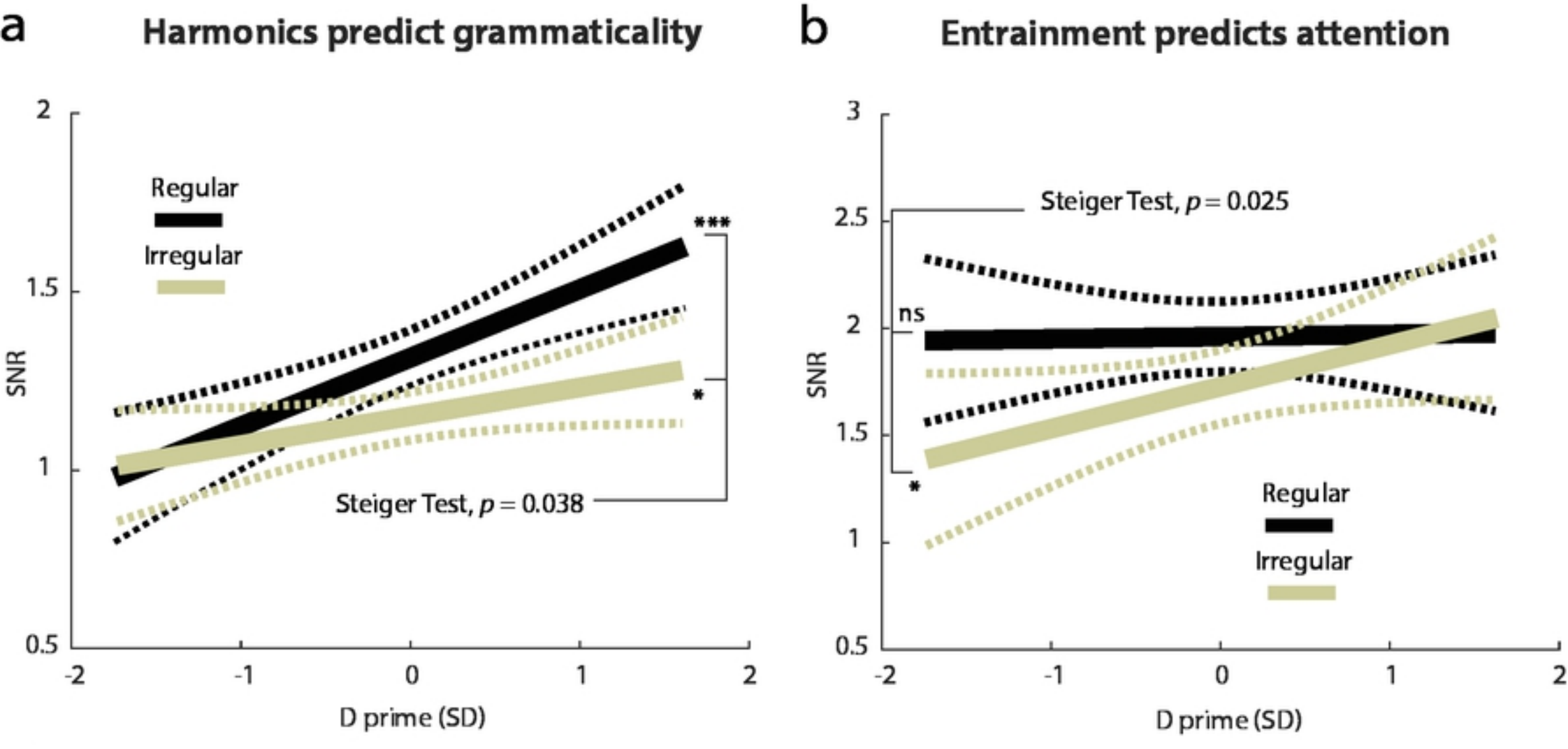
a
Harmonics predict grammaticality
Regular
Irregular
***
*
Steiger Test, p = 0.038
SNR
D prime (SD)
b
Entrainment predicts attention
Steiger Test, p = 0.025
ns
*
Regular
Irregular
SNR
D prime (SD)

Figure 6